



Análisis multivariante: **Aplicaciones con SPSS**

Elnor Crespín Elías

UFG Editores
www.ufg.edu.sv

Instituto de Ciencia, Tecnología e Innovación

Análisis multivariante:

Aplicaciones con SPSS

Misión

La formación de profesionales competentes, innovadores, emprendedores y éticos, mediante la aplicación de un proceso académico de calidad que les permita desarrollarse en un mundo globalizado.

Visión

Ser la mejor Universidad salvadoreña, reconocida regionalmente, que se caracteriza por la calidad de sus graduados, de su investigación y su responsabilidad social.

Consejo Directivo

Presidenta:	Lcda. Rosario Melgar de Varela, MEd.
Vicepresidenta:	Dra. Leticia Andino de Rivera
Secretaria General:	Lcda. Teresa de Jesús González de Mendoza, MEd.
Primer Vocal:	Dr. e Ing. Mario Antonio Ruiz Ramírez
Segundo Vocal:	Dr. Juan Portillo Hidalgo

Rector

Dr. e Ing. Mario Antonio Ruiz Ramírez

Vicerrectora

Dra. Leticia Andino de Rivera

Secretaria General

Lcda. Teresa de Jesús González de Mendoza, MEd.

Fiscal

Dr. Juan Portillo Hidalgo

Dirección y contacto

Universidad Francisco Gavidia: Calle El Progreso No. 2748, Edificio de Rectoría,
San Salvador, El Salvador: Tel. (503) 2249-2700
www.ufg.edu.sv

Análisis multivariante:

Aplicaciones con SPSS

Elnor Crespín Elías

UFG Editores
www.ufg.edu.sv

Instituto de Ciencia, Tecnología e Innovación



Misión

Diseñar, promover y acompañar iniciativas, políticas, programas y proyectos académicos empresariales para el desarrollo de la Ciencia, la Tecnología y la Innovación que impacten en la productividad y competitividad de El Salvador.

Visión

Ser el instituto científico líder en El Salvador en el desarrollo de la ciencia, la tecnología y la innovación.

Director

Dr. Óscar Picardo Joao, MEd., DEA

UFG EDITORES

Coordinación

Lcda. Claudia Meyer

Corrector de estilo

Lic. Carlos Alberto Saz

Diseñador

Talleres Gráficos UCA

DIRECCIÓN Y CONTACTO

Universidad Francisco Gavidia: Calle El Progreso No. 2748, Edificio de Rectoría, San Salvador, El Salvador.
Tel.: (503) 2249-2700 y 503) 2249-2716
Correo electrónico: editores@ufg.edu.sv
www.ufg.edu.sv

DE LA EDICIÓN

Título: Análisis multivariante:

Aplicaciones con SPSS

Autor: Elnor Crespín Elías

Colección: Tecnología

Ficha catalográfica

519.535	Crespín Elías, Elnor Osmín, 1968-
C921a	Análisis multivariante : aplicaciones con SPSS /
sv	Elnor Crespín Elías. --1ª ed. -- San Salvador, El Salv. : UFG Editores, 2016. 304 p. ; 22 cm. ISBN 978-99923-47-54-6 1. Análisis estadístico multivariable. 2. Decisiones estadísticas. 3. Análisis multivariado. I. Título.

Primera edición

©Instituto de Ciencia, Tecnología e Innovación (ICTI), 2016

ISBN: 978-99923-47-54-6

El contenido y opiniones vertidas en la publicación son responsabilidad exclusiva del autor. Se prohíbe la reproducción total o parcial de este documento, sin previa autorización de los editores.

Hecho el depósito que dicta la ley.

Edición de 500 ejemplares.

Impreso en Talleres Gráficos UCA.
Abril de 2016, San Salvador, República de El Salvador, América Central.

Índice

Prefacio	7
Nota introductoria	9
1. INTRODUCCIÓN AL ANÁLISIS MULTIVARIANTE.	13
1.1. Tipos de técnicas multivariantes.	15
1.2. Análisis exploratorio multivariante de datos	17
1.3. La matriz de datos.	18
1.4. Medidas de centralización: el vector de medias	19
1.5. La matriz de varianzas y covarianzas.	20
1.6. La varianza generalizada.	21
1.7. Coeficientes de asimetría y kurtosis.....	22
1.8. Aplicaciones en SPSS.....	23
1.9. Análisis gráfico y datos atípicos: histogramas y diagramas de dispersión.	28
1.10. Ejercicios	35
2. ANÁLISIS DE REGRESIÓN LINEAL MÚLTIPLE (RLM)	37
2.1. Metodología.	38
2.2. Aplicación del análisis de regresión múltiple.....	44
2.3. Supuestos del modelo de regresión lineal múltiple.....	48
2.4. Métodos para aplicar RLM.....	55
2.5. Análisis de regresión por pasos. Regresión STEP WISE-pasos sucesivos.....	57
2.6. Ejercicios.	60
3. ANÁLISIS DE COMPONENTES PRINCIPALES (ACP)	61
3.1. Planteamiento del problema	62
3.2. Enfoque descriptivo.....	62
3.3. Enfoque estadístico.....	64
3.4. Enfoque geométrico	64
3.5. Cálculo del primero y segundo componentes	65
3.6. Propiedades de los componentes	72
3.7. Análisis normado o con correlaciones.....	74
3.8. Interpretación de los componentes	76
3.9. Aplicación en SPSS.	77
3.10. Ejercicios	87

4. ANÁLISIS FACTORIAL.....	89
4.1. El modelo factorial	90
4.2. Hipótesis previas	91
4.3. Importancia de las aplicaciones de análisis factorial.....	92
4.4. Metodología del análisis factorial.....	93
4.5. Aplicación del método en SPSS.....	94
4.6. Ejercicios	125
5. ANÁLISIS SIMPLE DE CORRESPONDENCIAS.....	129
5.1. Métodos de reducción: Análisis de correspondencias.....	129
5.2. Metodología del Análisis de correspondencias	133
5.3. Aplicaciones del Análisis de correspondencias	139
5.4. Análisis de correspondencias simple (ANACOR).....	141
5.5. Aplicación mediante SPSS (ANACOR).....	141
5.6. Ejercicios	156
6. ANÁLISIS MÚLTIPLE DE CORRESPONDENCIAS (HOMALS)	159
6.1. Escalamiento óptimo.....	161
6.2. Aplicación mediante SPSS (HOMALS)	162
6.3. Ejercicios	174
7. ANÁLISIS DE CORRELACIÓN CANÓNICA NO LINEAL.....	177
7.1. Aplicación en SPSS.	179
7.2. Ponderaciones, cuantificaciones y saturaciones.....	185
7.3. Cuantificaciones	189
7.4. Ejercicios	192
8. ANÁLISIS DE CORRELACIÓN CANÓNICA LINEAL	195
8.1. Fundamentación matemática del Análisis de correlación canónica	196
8.2. El análisis de relaciones mediante la correlación canónica	200
8.3. Metodología del Análisis de correlación canónica	201
8.4. Supuestos y términos clave.....	206
8.5. Aplicación en SPSS	208
8.6. Ejercicios	219
9. Bibliografía.....	221
10. Anexos	223

Prefacio

En relación con la obra que ahora nos ocupa, no podemos sino celebrar su aparición en un momento en el que se carece de textos apropiados para el analista cuantitativo que se encara a una masa de datos de los que debe extraer la información más precisa posible para caracterizarlos, analizarlos, extraer conclusiones, y tomar decisiones.

La investigación económica en general -en sus diferentes vertientes, como la de mercados, de opinión pública, análisis de encuestas de hogares, etc.- demanda cada vez más la utilización de un instrumental estadístico y analítico riguroso que con mucha frecuencia se encuentra solo en textos de autores foráneos que los presentan con un lenguaje que no es familiar para un gran número de analistas.

Pero la obra del Dr. Elnor Crespín Elías no es así. Al contrario, comienza introduciendo al estudioso con algunos de los temas más básicos del análisis multivariante como en el caso de la regresión múltiple, para trascender posteriormente a una presentación de los instrumentos necesarios para un análisis más demandante y exigente, de manera consistente con la necesidad de llevar a cabo investigaciones más completas, conforme lo impone el estudio de los fenómenos sociales cada vez más complejos que se manifiestan en las sociedades modernas.

Y lo hace atendiendo dos aspectos fundamentales en un texto introductorio de métodos estadísticos avanzados: por un lado, recurriendo al aparato matemático necesario pero mínimo; por otro lado, ilustrándolo con una variedad de ejercicios que permiten al lector hacerse una idea clara de la riqueza y potencial analítico de los métodos multivariantes. El empleo del software SPSS, relativamente conocido, lo hace todavía más asequible al lector.

Adicionalmente, es bien sabido que el país mantiene una brecha amplia en el manejo de aspectos cuantitativos para la investigación social y económica respecto de otras regiones en desarrollo como el este asiático. Por esto, la obra del Dr. Crespín es doblemente bienvenida, ya que contribuye a reducir esta brecha. Definitivamente, se trata de una obra que se recomienda a todo aquel estudioso que necesite llevar a cabo su investigación con propiedad, autoridad y rigor analíticos.

Mauricio González Orellana
Economista e investigador

Nota introductoria

Generalmente se nos ha enseñado en las aulas universitarias el proceso de análisis estadístico a partir de un análisis en particular: Recolectar la data a través de diferentes mecanismos o herramientas; realizar un análisis estadístico, aplicando herramientas de estadística descriptiva, que es lo más usual, presentar la información en una forma comprensible, y muy pocas veces aplicamos estadística inferencial.

El investigador o profesional calcula frecuencias, hace gráficas, encuentra medidas de tendencia central, calcula la dispersión e identifica valores atípicos en la distribución de los datos. Se aplica estadística descriptiva, que, como su nombre lo indica, describe la data en general; mientras que la estadística inferencial no solo describe los datos, sino que aborda el problema de hacer generalizaciones más amplias o inferencias de una muestra de datos hacia la población; sin embargo, la estadística como herramienta tiene un amplio abanico de opciones para tratar y analizar datos que van más allá de las tradicionales dos variables.

Un ejemplo muy clásico es el analizar la situación de salud o educación de una comunidad, un instrumento de psicología que explora comportamientos o habilidades en los individuos; un análisis de mercado donde se exploran nichos o niveles de satisfacción de los usuarios; un análisis climático, donde se analizan datos del clima de una región o de una situación particular; una base de datos con indicadores sociales y económicos de una población en particular; un censo agropecuario, de hogares o de vivienda, que contiene decenas o centenas de variables. Estos y otros ejemplos parecidos denotan una gran cantidad de variables que requieren métodos estadísticos avanzados para poder interpretarlos y tomar decisiones efectivas.

Generalmente, los estudiantes o profesionales que se introducen en el campo de la investigación científica, solamente utilizan frecuencias simples o variables univariadas, donde ejemplifican porcentajes específicos; en otros casos, se hace un análisis bivariado en el cual se toman de dos a dos las variables de un problema en particular; no obstante, estos análisis son limitados y su tratamiento no es útil para tomar decisiones ya que se necesita de analizar categorías y determinados perfiles, lo cual requiere tomar más variables.

En la vida cotidiana encontramos situaciones donde se precisa dar respuesta a diferentes preguntas como las siguientes: ¿Cuáles son los atributos de un producto que más influye en las decisiones de compra de un conjunto de consumidores?

¿Cuáles son los atributos que un consumidor utiliza para distinguir las marcas de auto, de ropa, zapatos, etc? ¿Cuáles son los factores que subyacen en la personalidad de un grupo de estudiantes de educación secundaria? ¿Cuáles son los factores que determinan la situación de salud o educación de una comunidad en particular? ¿Cómo la inteligencia predice el rendimiento académico entre estudiantes de secundaria? Solamente a través del análisis multivariante podemos dar respuesta a estas interrogantes.

El análisis multivariante tiene múltiples aplicaciones, desde crear perfiles de grupos de individuos, de niveles de satisfacción, realizar tratamiento y análisis de encuestas de satisfacción (políticas, económicas), encuestas de opinión pública; utilizando escalas de percepciones, identificar ciudades que sean comparables para abordar políticas (sociales y económicas) similares. El objetivo del análisis multivariante es propiciar el estudio conjunto de múltiples variables (cualitativas y cuantitativas), procurando obtener información útil y de apoyo a la toma de decisiones; para ello, reduce el número de variables y sintetiza grandes conjuntos de datos, buscando una mejor interpretación y, por lo tanto, una mayor comprensión de los fenómenos que se estudian.

Este libro es producto de la experiencia de un curso diseñado e impartido para estudiantes universitarios, y puede ser utilizado por investigadores y profesionales que desean retomar la estadística como una herramienta para la investigación científica. El material ha sido cuidadosamente preparado con una metodología paso a paso, con aplicaciones y ejercicios propuestos, en los cuales se utiliza el software SPSS versión 22.¹

Cada capítulo contiene una serie de archivos en formato SPSS² (*.sav), los cuales son presentados en los anexos, y constituyen el insumo para comprender la herramienta multivariante. El lector podrá utilizar estos archivos y ejecutarlos en el programa SPSS e interpretar los resultados, que se detallan en el apartado de aplicaciones de cada capítulo.

Este libro tiene la característica que utiliza un lenguaje sencillo, de que puede ser entendido por cualquier profesional, y no es necesario ser matemático para entender y utilizar la herramienta. Tiene una fundamentación matemática sencilla que explica la metodología, una base teórica para entender los resultados que generan los programas estadísticos cuando se ejecuta el procedimiento correspondiente, aplicaciones de casos particulares y una lista de ejercicios propuestos para que el lector aplique la técnica.

1 Lo cual no implica que pueda utilizarse otro software como STATA o R; que genera resultados similares, pero con sintaxis diferente, propia de cada programa estadístico.

2 Estos archivos se presentan en el Anexo del libro, con el correspondiente diccionario de datos.

Esperamos que esta obra sea utilizada por profesionales y estudiantes interesados en ampliar sus conocimientos y fortalecer el análisis de grandes bases de datos que se generan en los diferentes campos profesionales y, por supuesto, que sea utilizado por investigadores dentro de las etapas del método científico.

El libro está constituido por ocho capítulos; haciendo una introducción al Análisis Multivariante se explora como tema introductorio el Análisis de Regresión Lineal Múltiple (RLM), continuando con el Análisis de Componentes Principales, el Análisis Factorial, el Análisis Simple de Correspondencias y Múltiple, y el Análisis de Correlación Canónica Lineal y no Lineal.

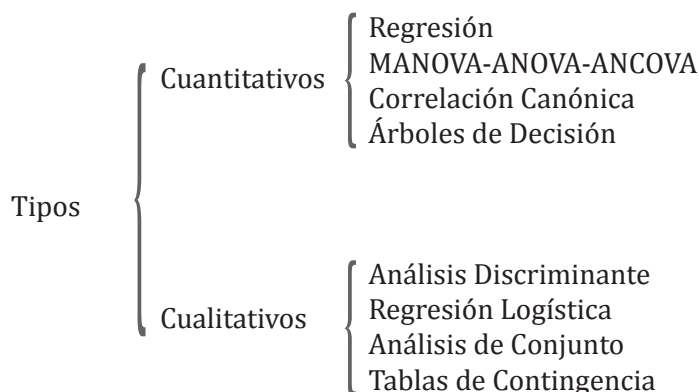
Clasificación del Análisis Multivariante

Las técnicas multivariantes dependen del tipo y la escala de medida de las variables que pretendemos analizar. Actualmente existen muchas técnicas multivariantes contenidas en los usuales programas estadísticos; y algunas de ellas, de muy reciente surgimiento. De manera general estas técnicas se pueden clasificar como sigue:

1. Métodos de dependencia

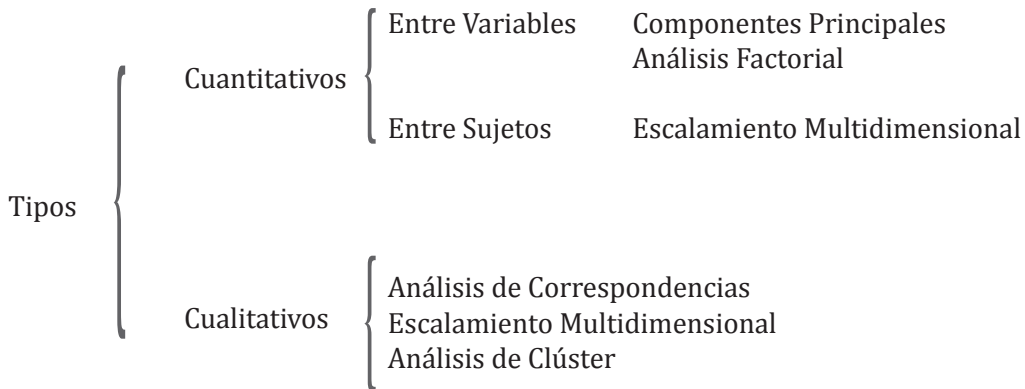
En esta clasificación las variables a analizar se dividen en dos grupos: las variables dependientes y las variables independientes, y los métodos tienen como objetivo determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma lo hace; es decir, se estima o predice una variable específica en función de otras variables.

Los tipos de herramientas más conocidos se pueden calificar de la siguiente manera:



2. Métodos de interdependencia

En esta clasificación no hay variables con el rol de dependiente e independiente, lo único que sabemos es que existe una posible interrelación entre ellas. El objetivo de estos métodos es identificar qué variables están relacionadas, y cómo lo están.



3. Métodos estructurales

Estos métodos suponen que las variables están divididas en dos grupos: variables dependientes y variables independientes, y tienen como objetivo analizar no solo cómo las variables independientes afectan a las variables dependientes, sino también cómo están relacionadas entre sí las variables de los dos grupos. Estas herramientas son un insumo de técnicas recientes, donde hay mucha investigación, como lo es la Minería de datos³ y *Data Warehouse*,⁴ donde se almacenan y se analizan grandes cantidades de datos.

3 La Minería de datos comprende la búsqueda de patrones dentro de grandes bases de datos, utilizando para ello métodos estadísticos e inteligencia artificial. En realidad es la combinación de varias áreas: estadística, inteligencia artificial, computación gráfica, bases de datos y procesamiento masivo, principalmente usando como materia prima las bases de datos. Fuente: Mariela Yanina, Universidad Nacional del Nordeste Facultad de Ciencias Exactas, Naturales y Agrimensura, República de Argentina.

4 Un sistema de *Data Warehouse* contiene datos refinados, históricos, resumidos y no volátiles, y ofrece un entorno integrado de información organizada de acuerdo a los requisitos establecidas por una organización. Es un almacenamiento no volátil de datos, transacciones y eventos.

1. Introducción al análisis multivariante

¿Qué es el análisis multivariante?

La tecnología informática disponible hoy en día, casi inimaginable, hace sólo dos décadas, ha hecho posible avances extraordinarios en el análisis de datos psicológicos, sociológicos y de otro tipo de datos referidos al comportamiento humano. Este impacto es más evidente en la relativa facilidad con la que las computadoras pueden analizar enormes cantidades de datos complejos. Casi cualquier problema se puede analizar fácilmente hoy en día por un número grande de programas estadísticos, incluso en computadoras personales. Además, los efectos del progreso tecnológico han extendido aún más la capacidad de manipular datos, liberando a los investigadores de las restricciones del pasado y permitiéndoles así abordar investigaciones más sustantivas y ensayar sus modelos teóricos. Las limitaciones metodológicas no son ya un asunto crítico para el investigador interesado en la búsqueda de evidencia empírica. Gran parte de esta creciente comprensión y pericia en el análisis de datos ha venido a través del estudio de la estadística y de la inferencia estadística. Igualmente importante, sin embargo, ha sido el dilatado conocimiento y aplicación de un grupo de técnicas estadísticas conocidas como análisis multivariante.

Las técnicas del análisis multivariante están siendo ampliamente aplicadas a la industria, administración y centros de investigación. Por otra parte, pocos campos de investigación o estudio han fracasado en integrar las técnicas multivariantes en su caja analítica de herramientas. Para atender este creciente interés se han publicado numerosos libros y artículos sobre aspectos teóricos y matemáticos de estas herramientas, y han aparecido textos introductorios en casi todos los campos. Sin embargo, se han escrito pocos libros para el investigador, que traten las aplicaciones de la estadística multivariante, así como que suministren un tratamiento conceptual de los métodos estadísticos.

Revolución informática

Es poco menos imposible discutir la aplicación de las técnicas multivariantes sin una mención al impacto de la informática. El amplio desarrollo de la aplicación

de las computadoras para procesar grandes y complejas bases de datos, ha estimulado de manera impresionante el uso de los métodos de estadística multivariante. Toda la estadística teórica de las técnicas multivariantes actuales fue desarrollada mucho antes de la aparición de las computadoras; pero sólo cuando estuvo disponible el poder de la informática, particularmente en las computadoras personales, han puesto a disposición de cualquier investigador interesado el acceso a todos los recursos necesarios como para resolver un problema multivariante de casi cualquier dimensión. En la comunidad académica, disciplinas de todos los campos del saber han adoptado técnicas multivariantes, y los académicos deben estar cada vez más versados en las técnicas multivariantes apropiadas para sus investigaciones empíricas. Incluso para personas con sólida preparación cuantitativa, la disponibilidad de programas preparados para el análisis multivariante ha facilitado la compleja manipulación de matrices de datos que durante mucho tiempo ha retrasado el crecimiento de técnicas multivariantes.

Todos los programas (*software*) de estadística (por ej., SPSS, SAS, STATA, R, etc.) contienen rutinas o algoritmos que tratan el análisis multivariante. Estos programas especializados incluyen todo tipo de análisis multivariante, incluyendo los de escala multidimensional, modelos de ecuaciones simultáneas/estructurales y análisis conjunto. Se están desarrollando sistemas expertos dirigidos incluso a temas tales como la selección de una técnica estadística o diseñar un plan de muestreo que asegure los objetivos prácticos y estadísticos deseados.

Definición del análisis multivariante

En un sentido amplio, se refiere a todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación. Cualquier análisis simultáneo de más de dos variables puede ser considerado aproximadamente como un análisis multivariante. En sentido estricto, muchas técnicas multivariantes son extensiones del análisis univariante (análisis de distribuciones de una sola variable) y del análisis bivalente (clasificación cruzada, correlación, análisis de varianza y regresiones simples utilizadas para analizar dos variables). Por ejemplo, una regresión simple-RLS (con una variable predictor) se extiende al caso multivariante para incluir varias variables predictor. De la misma forma, la variable dependiente que se encuentra en el análisis de la varianza se extiende para incluir múltiples variables dependientes en el análisis multivariante de la varianza. Otras técnicas multivariantes, sin embargo, están diseñadas exclusivamente para tratar con problemas multivariantes, tales como el análisis factorial que sirve para identificar la estructura subyacente de un conjunto de variables o el análisis discriminante que sirve para diferenciar entre grupos basados en un conjunto de variables.

Para ser considerado un análisis multivariante, todas las variables deben ser aleatorias y estar interrelacionadas de tal forma que sus diferentes efectos no puedan ser interpretados separadamente con algún sentido. Algunos autores afirman que el propósito del análisis multivariante es medir, explicar y predecir el grado de relación de los valores teóricos (combinaciones ponderadas de variables). Por tanto, el carácter multivariante reside en los múltiples valores teóricos (combinaciones múltiples de variables) y no sólo en el número de variables u observaciones.

1.1. Tipos de técnicas multivariantes

El análisis multivariante es un conjunto de técnicas de análisis de datos en expansión. Entre las técnicas más conocidas expuestas en este texto tenemos regresión múltiple, componentes principales, análisis factorial, análisis de correspondencia y correlación canónica.⁵ A continuación describiremos cada una de las técnicas multivariantes, definiendo brevemente la técnica y el objetivo de su aplicación.

a) Componentes principales y Análisis factorial

El análisis factorial, que incluye variaciones tales como el análisis de componentes, es una aproximación estadística que puede usarse para analizar interrelaciones entre un gran número de variables y explicar estas variables en términos de sus dimensiones subyacentes comunes (factores). El objetivo es encontrar un modo de condensar la información contenida en un número de variables originales en un conjunto más pequeño de variables (factores) con una pérdida mínima de información.

b) Regresión múltiple

La regresión múltiple es el método de análisis apropiado cuando el problema del investigador incluye una única variable métrica dependiente que se supone está relacionada con una o más variables métricas independientes. El objetivo del análisis de la regresión múltiple es predecir los cambios en la variable dependiente en respuesta a cambios en varias de las variables independientes. Este objetivo se consigue muy a menudo a través del método de los mínimos cuadrados.

La regresión múltiple es útil cuando se está interesado en predecir la cantidad o magnitud de la variable dependiente. Por ejemplo, predecir las ventas de una compañía a partir de información sobre sus gastos en publicidad, el número de vendedores y el número de tiendas que distribuyen sus productos.

⁵ Existen otras técnicas que no están incluidas en este libro, pero son herramientas valiosas en el análisis multivariado: análisis discriminante, análisis de clúster, análisis conjunto, regresión logística y probit, modelos de ecuaciones estructurales, entre otras.

c) Análisis de correspondencias

El análisis de correspondencia es una técnica de interdependencia que facilita tanto la reducción dimensional de una clasificación de objetos (por ejemplo, productos, personas, etc.) sobre un conjunto de atributos y el mapa perceptual de objetos relativos a estos atributos. Los investigadores se enfrentan constantemente a la necesidad de cuantificar datos cualitativos que se encuentran en variables nominales. El análisis de correspondencia difiere de otras técnicas de interdependencia en su capacidad para acomodar tanto datos no métricos como relaciones no lineales.

En su forma más básica el análisis de correspondencias emplea una tabla de contingencia, que es la tabulación cruzada de dos variables categóricas. A continuación transforman los datos no métricos en un nivel métrico y realiza una reducción dimensional (similar al análisis factorial) y un mapa perceptual. A modo de ejemplo, las preferencias por una marca de los encuestados pueden ser tabuladas de forma cruzada con variables demográficas (por ejemplo, género, categorías de renta, ocupación) indicando cuánta gente que prefiere cada una de las marcas entra en cada categoría de las variables demográficas. A través del análisis de correspondencias, la asociación o correspondencia de marcas y las características distintivas de aquellos que prefieren cada marca se muestran en un mapa bi o tri-dimensional, tanto de marcas como características de los encuestados. Las marcas percibidas como similares están localizadas en una cercana proximidad unas de otras. De la misma forma, las características más distintivas de los encuestados que prefieren cada marca están determinadas también por la proximidad de las categorías de las variables demográficas respecto de la posición de la marca. El análisis de correspondencia proporciona una representación multivariante de la interdependencia de datos no métricos que no es posible realizar con otros métodos.

d) Correlación canónica

El análisis de correlación canónica puede verse como una extensión de un análisis de regresión múltiple. Recordemos que el análisis de regresión múltiple implica una única variable dependiente métrica y varias variables métricas independientes. Con el análisis canónico el objetivo es correlacionar simultáneamente varias variables dependientes métricas y varias variables métricas independientes. Mientras que la regresión múltiple implica una única variable dependiente, la correlación canónica implica múltiples variables dependientes. El principio subyacente es desarrollar una combinación lineal de cada conjunto de variables (tanto independientes como dependientes) para maximizar la correlación entre los dos conjuntos. O dicho de otra forma, el procedimiento implica obtener un conjunto de ponderaciones para las variables dependientes e independientes

que proporcione la correlación única máxima entre el conjunto de variables dependientes y el conjunto de variables independientes.

e) Análisis multidimensional

En el análisis multidimensional el objetivo es transformar los juicios de los consumidores de similitud o preferencia (por ejemplo, preferencias por tiendas o marcas comerciales) en distancias representadas en un espacio multidimensional. Si los objetos A y B son en opinión de los encuestados más similares que el resto de los pares posibles de objetos, las técnicas de análisis multidimensional situarán a los objetos A y B de tal forma que la distancia entre ellos en un espacio multidimensional es menor que la distancia entre cualquier otro par de objetos. Los mapas perceptuales resultantes muestran el posicionamiento relativo entre los objetos; pero es necesario un análisis adicional para evaluar qué atributos predicen la posición de cada objeto.

1.2. Análisis exploratorio multivariante de datos

A continuación se hace una introducción de cómo describir un conjunto de datos multivariantes. Supondremos que hemos observado una muestra de variables en una muestra de elementos de una población y presentaremos métodos para resumir los valores de las variables y describir su estructura de dependencia. Completaremos además, el análisis descriptivo analizando cómo representar los datos gráficamente y elegir transformaciones de las variables originales que conduzcan a una descripción más simple. También comentaremos el problema de limpiar los datos de valores atípicos, que son observaciones debidas a errores de medida u otras causas de heterogeneidad.

El análisis descriptivo debe siempre aplicarse como primer paso para comprender la estructura de los datos y extraer la información que contienen, antes de pasar a los métodos más complejos de las secciones siguientes.

Las herramientas simples que se describen, cuando el interés se centra en la relación entre las variables o en la comparación de dos conjuntos de datos, pueden ser de gran ayuda antes de emprender estudios más complejos.

Datos multivariantes

La información de partida para los métodos estudiados en esta sección puede ser de varios tipos. La más habitual es una tabla donde aparecen los valores de p variables observadas sobre n elementos. Las variables pueden ser cuantitativas, cuando su valor se exprese numéricamente, como la edad de una persona, su estatura o sus ingresos; o cualitativas, cuando su valor sea un atributo o categoría, como el género, el color de los ojos o el municipio de nacimiento. Las variables

cuantitativas pueden, a su vez, clasificarse en continuas o de intervalo, cuando pueden tomar cualquier valor real en un intervalo, como la estatura; o discretas, cuando sólo toman valores enteros, como el número de hermanos. Las variables cualitativas pueden clasificarse en binarias, cuando toman dos valores posibles, como el género (mujer, hombre); o generales, cuando toman muchos valores posibles, como el municipio de residencia.

Las variables cualitativas pueden también codificarse numéricamente, pero requieren un tratamiento distinto. Si los valores de las categorías no tienen relación entre sí, la forma más útil de codificarlas es convertirlas en variables binarias. Por ejemplo, supongamos la variable color de los ojos, CO; y para simplificar, supongamos que las categorías posibles son azules (A), verdes (V), castaños (C) y negros (N). Tenemos $p=4$ categorías, que podemos representar con $p-1=3$ variables binarias definidas como:

- a) $X_1=1$ si $CO=A$, $X_1=0$ en otro caso.
- b) $X_2=1$ si $CO=V$, $X_2=0$ en otro caso.
- c) $X_3=1$ si $CO=C$, $X_3=0$ en otro caso.

La siguiente tabla presenta la codificación de la variable atributo CO en las tres variables binarias cuantitativas, x_1 , x_2 , x_3 .

Tabla 1: Codificación de variables categóricas

CO	X1	X2	X3
A	1	0	0
V	0	1	0
C	0	0	1
N	0	0	0

1.3. La matriz de datos

Supondremos, en adelante, que hemos observado p variables numéricas en un conjunto de n elementos. Cada una de estas variables se denomina variable escalar o univariante, y el conjunto de las p variables forman una variable vectorial o multivariante. Los valores de las p variables escalares en cada uno de los n elementos pueden representarse en una matriz $\mathbf{X}$, de dimensiones $(n \times p)$, que llamaremos matriz de datos. Denotaremos por x_{ij} al elemento genérico de esta matriz, que representa el valor de la variable escalar j sobre el individuo i . Es decir:

$\mathbf{X} = \{x_{ij}\}$ donde $i=1, \dots, n$ representa el individuo; y $j=1, \dots, p$ representa la variable.

La matriz de datos $\mathbf{X}$, puede representarse de dos formas distintas: por filas, como:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} = \begin{bmatrix} X'_1 \\ X'_2 \\ \dots \\ X'_p \end{bmatrix}$$

Donde cada variable x'_i es un vector fila, $px1$, que representa los valores de las p variables sobre el individuo i . Alternativamente, podemos representar la matriz $\mathbf{x}$ por columnas:

$$\mathbf{X} = [\mathbf{x}_{(1)} \dots \mathbf{x}_{(p)}],$$

Donde ahora cada variable $\mathbf{x}_{(j)}$ es un vector columna, $nx1$, que representa la variable escalar x_j medida en los n elementos de la población.

1.4. Medidas de centralización: el vector de medias

La medida de centralización más utilizada para describir datos multivariantes es el vector de medias, que es un vector de dimensión p , cuyos componentes son las medias de cada una de las p variables. Puede calcularse, como el caso escalar, promediando las medidas de cada elemento, que ahora son vectores:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_p \end{bmatrix}$$

Su expresión a partir de la matriz de datos es la siguiente:

$$\bar{\mathbf{x}} = \frac{1}{n} \mathbf{X}' \mathbf{1}$$

Donde $\mathbf{1}$ representará siempre un vector de números unos, de la dimensión adecuada. Escribiendo la matriz $\mathbf{X}$ en términos de sus vectores fila, que son vectores de dimensión $1xp$ que contienen los valores de las p variables en cada elemento de la muestra, estos vectores son las columnas de $\mathbf{X}'$, y tendremos que:

$$\bar{\mathbf{x}} = \frac{1}{n} \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

El vector de medias se encuentra en el centro de los datos, en el sentido de hacer cero la suma de las desviaciones:

$$\sum_{i=1}^N (x_i - \bar{x}) = 0$$

Las medidas de centralización escalares, basadas en el orden de las observaciones, no pueden generalizarse fácilmente al caso multivariante. Por ejemplo, podemos calcular el vector de medianas; pero este punto no tiene necesariamente una situación como centro de los datos. Esta dificultad proviene de la falta de un orden natural de los datos multivariantes.

1.5. La matriz de varianzas y covarianzas

Como hemos comentado, para variables escalares la variabilidad respecto a la media se mide habitualmente por la varianza, o su raíz cuadrada, la desviación típica. Además, la relación lineal entre dos variables x_j y x_k se mide por la covarianza, y se calcula por:

$$S_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k),$$

y mide su dependencia lineal.

Para una variable multivariante se define la matriz de varianzas y covarianzas como:

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

La matriz de varianzas y covarianzas, que llamaremos, para simplificar, matriz de covarianzas, es la matriz simétrica de orden p con forma:

$$S = \begin{bmatrix} s_{11}^2 & \cdots & s_{1p} \\ \vdots & & \vdots \\ s_{p1} & \cdots & s_{pp}^2 \end{bmatrix}$$

Cálculo a partir de la matriz de datos centrados

La matriz S puede obtenerse directamente a partir de la matriz de datos centrados $\tilde{X}$, que se define como la matriz resultado de restar a cada dato su media:

$$\tilde{X} = X - 1\bar{X}'$$

Sustituyendo el vector de medias, tenemos:

$$\tilde{X} = PX$$

Donde la matriz cuadrada P está definida por:

$$p = 1 - \frac{1}{n} \mathbf{1}\mathbf{1}'$$

Entonces, la matriz S puede escribirse:

$$S = \frac{1}{n} \bar{X}' \bar{X} = \frac{1}{n} X' P X$$

1.6. La varianza generalizada

Una medida mejor que la variabilidad global es la varianza generalizada, que es el determinante de la matriz de varianzas y covarianzas, es decir:

$$VG = |S|$$

Su raíz cuadrada se denomina desviación típica generalizada, y tiene las propiedades siguientes:

- Está bien definida, ya que el determinante de la matriz de varianzas y covarianzas es siempre no negativo.
- Es una medida del área (para $p=2$), volumen (para $p=3$) o hipervolumen (para $p>3$), ocupado por el conjunto de datos.

Supongamos el caso $p=2$. Entonces, S puede escribirse:

$$S = \begin{bmatrix} s_x^2 & r s_x s_y \\ r s_x s_y & s_y^2 \end{bmatrix}$$

Y la desviación típica generalizada es:

$$|S|^{1/2} = s_x s_y \sqrt{1-r^2}$$

La variabilidad promedio

Peña y Rodríguez (2000) han propuesto como medida global de variabilidad, la variabilidad promedio, dada por:

$$VP = |S|^{1/p}$$

Que tiene la ventaja de que, cuando todas las variables van en las mismas dimensiones esta medida tiene las unidades de la varianza. Para matrices diagonales esta medida es simplemente la media geométrica de las varianzas. Observemos que, como el determinante es el producto de los valores propios, la variabilidad promedio es la media geométrica de los valores propios de la matriz S , que por ser semidefinida positiva, serán siempre no negativos.

Análogamente podemos definir la desviación promedio mediante:

$$DP = |S|^{1/2 \ p}$$

1.7. Coeficientes de asimetría y kurtosis

La generalización de los coeficientes de asimetría y kurtosis al caso multivariante no es inmediata. Una de la propuestas más utilizadas es debida a Mardia (1970), que propone calcular las distancias de Mahalanobis para cada par de elementos muestrales (i,j) :

$$d_{ij} = (x_i - \bar{x})' S^{-1} (x_j - \bar{x})$$

Y define el coeficiente de asimetría multivariante en la distribución conjunta de las p variables como:

$$A_p = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^3$$

Y el de kurtosis:

$$K_p = \frac{1}{n} \sum_{i=1}^n d_{ii}^2$$

Estos coeficientes tienen las propiedades siguientes:

1. Para variables escalares $K_p = A^2$. En efecto, entonces:

$$A_p = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x})^3 (x_j - \bar{x})^3 / s^6 = \frac{(\sum_{i=1}^n (x_i - \bar{x})^3)^2}{n^2 s^6} A^2$$

2. El coeficiente de asimetría es no negativo y será cero si los datos están distribuidos homogéneamente en una esfera.

3. Para variables escalares $K = K_p$, el resultado es inmediato porque entonces:

$$A_p d_{ii}^2 = (x_i - \bar{x})^4 / s^4.$$
4. Los coeficientes son invariantes ante transformaciones lineales de los datos. Si $y = AX + b$, los coeficientes de asimetría y kurtosis de “x” y “y” son idénticos.

1.8. Aplicaciones en SPSS

Los datos del fichero *acciones.sav* incluyen tres medidas de rentabilidad de 34 acciones en bolsa durante un período de tiempo. La primera, x1, es la rentabilidad efectiva por dividendos (dividendos repartidos por acción y dividendos por precio de la acción); x2, es la proporción de beneficios que va a dividendos (beneficios repartidos en dividendos sobre beneficios totales); y x3 es el coeficiente entre precio por acción y beneficios. La tabla siguiente presenta dicha información, así:

Se solicita:

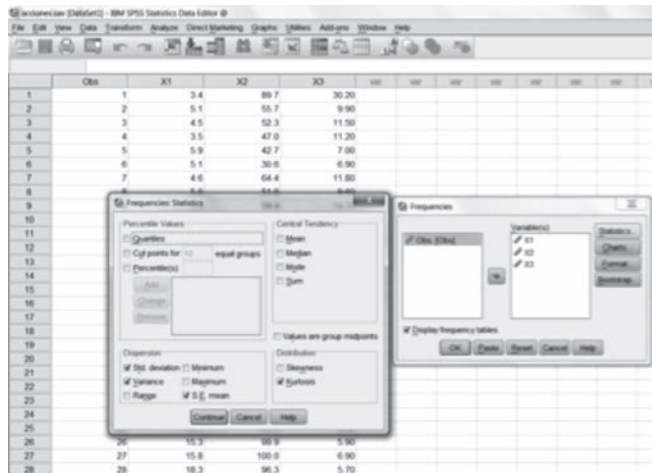
- a) Realizar un análisis descriptivo para las tres variables en estudio. Calculando en el software SPSS las medias, desviación típica, coeficiente de asimetría y coeficiente de kurtosis. Realice un análisis estadístico a su criterio.
- b) Calcular la matriz de covarianzas de las acciones, luego analizar su resultado y concluir.

Tabla 2. Datos que representan medidas de rentabilidad de 34 acciones en bolsa

Obs.	X1	X2	X3	Obs	X1	X2	X3
1	3.4	89.7	30.2	18	4.4	58.5	12.1
2	5.1	55.7	9.9	19	7.8	84.3	11.0
3	4.5	52.3	11.5	20	16.0	96.5	6.0
4	3.5	47.0	11.2	21	16.7	100.0	6.8
5	5.9	42.7	7.0	22	15.2	92.3	5.2
6	5.1	30.6	6.9	23	17.5	99.9	6.8
7	4.6	64.4	11.8	24	16.2	93.5	6.1
8	5.0	51.0	9.6	25	14.7	100.0	6.6
9	3.2	54.4	14.7	26	15.3	99.9	5.9
10	3.4	45.7	13.2	27	15.8	100.0	6.9
11	6.5	39.9	5.2	28	18.3	96.3	5.7
12	4.4	40.3	13.7	29	15.9	100.0	6.9
13	5.1	52.4	11.0	30	16.1	92.5	6.1
14	5.8	43.9	8.0	31	9.7	87.6	7.7
15	4.6	52.8	14.4	32	6.9	53.6	6.6
16	7.2	65.8	7.8	33	14.4	87.8	5.2
17	7.2	58.1	7.7	34	14.9	34.5	4.69

Procedimiento:

- En archivos – abrir- seleccionamos el anexo acciones.sav.
- En analizar – Frecuencias – seleccionamos las variables X1, X2 y X3 y los estadísticos descriptivos, como se muestra en la siguiente imagen:



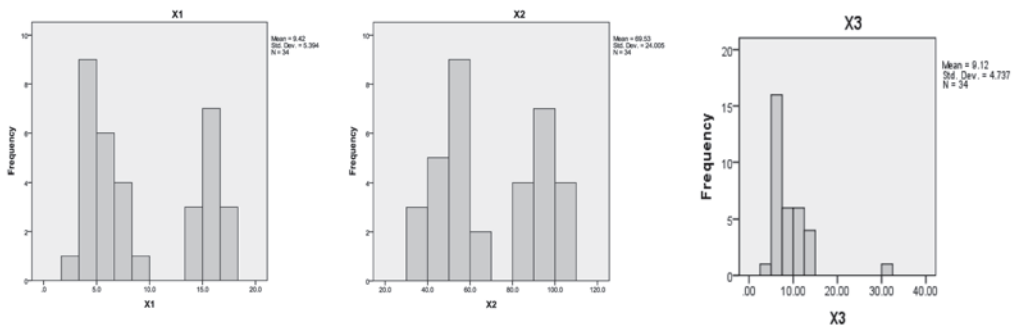
- Las estadísticas descriptivas se muestran a continuación.

Statistics			
	X1	X2	X3
N Valid	34	34	34
Missing	0	0	0
Mean	9.421	69.526	9.1203
Std. Error of Mean	.9251	4.1168	.81240
Median	7.050	61.450	7.3500
Mode	5.1	100.0	5.20 ^a
Std. Deviation	5.3940	24.0048	4.73704
Variance	29.096	576.229	22.440
Skewness	.390	.049	2.847
Std. Error of Skewness	.403	.403	.403
Kurtosis	-1.683	-1.663	11.287
Std. Error of Kurtosis	.788	.788	.788

a. Multiple modes exist. The smallest value is shown

Las medidas descriptivas como kurtosis indican un alejamiento de la distribución normal para las tres variables: las dos primeras tienen valores muy bajos de la kurtosis, lo que indica alta heterogeneidad, posiblemente por la presencia de dos grupos de datos distintos, y la tercera tiene alta kurtosis, lo que sugiere la presencia de valores atípicos.

Estas características son muy claras en los histogramas de las variables. La primera variable, rentabilidad efectiva por dividendos, x1, muestra dos grupos de acciones con comportamiento distinto. El histograma de la segunda variable, x2, muestra también dos grupos de acciones. Finalmente, la distribución de la tercera variable es muy asimétrica, con un valor atípico muy destacado. La evidencia disponible indica que las acciones pueden probablemente dividirse en dos grupos más homogéneos. Sin embargo, vamos a ilustrar el análisis de todos los datos.

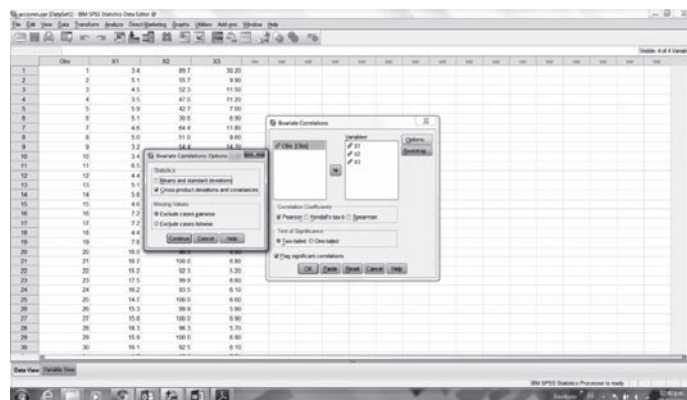


a) Histogramas de la rentabilidad por dividendos.

b) Histograma del precio por acción con relación a los beneficios.

Para obtener la matriz de varianzas en SPSS, sigamos los siguientes pasos:

- En el menú Analizar-Correlaciones-Bivariadas arrastramos todas las variables al campo de lista.
- Luego, en el botón de Opciones seleccionamos desviaciones de productos cruzados y covarianzas. Así como se muestra en la figura siguiente:



- El resultado de la matriz de varianzas y covarianzas de estas tres variables se presenta en la siguiente imagen (resultado de SPSS):

Correlations				
		X1	X2	X3
X1	Pearson Correlation	1	.776**	-.608**
	Sig. (2-tailed)		.000	.000
	Sum of Squares and Cross-products	960.156	3314.541	-513.009
	Covariance	29.096	100.441	-15.546
	N	34	34	34
X2	Pearson Correlation	.776**	1	-.157
	Sig. (2-tailed)	.000		.376
	Sum of Squares and Cross-products	3314.541	19015.566	-587.648
	Covariance	100.441	576.229	-17.808
	N	34	34	34
X3	Pearson Correlation	-.608**	-.157	1
	Sig. (2-tailed)	.000	.376	
	Sum of Squares and Cross-products	-513.009	-587.648	740.504
	Covariance	-15.546	-17.808	22.440
	N	34	34	34

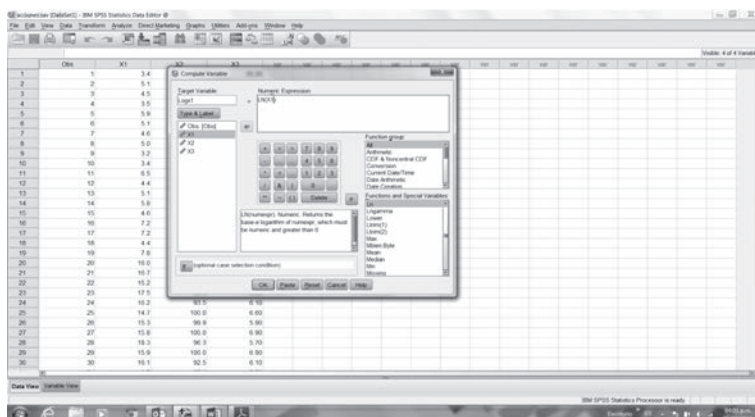
** Correlation is significant at the 0.01 level (2-tailed).

Los elementos diagonales de esta matriz son los cuadrados de las desviaciones típicas de la tabla donde se muestran los resultados de las estadísticas descriptivas. Como las dimensiones de las variables son distintas, no tiene sentido calcular medidas promedio.

Los histogramas de las tres variables han mostrado una clara ausencia de normalidad. Una posibilidad es transformar las variables para facilitar su interpretación. Tomando logaritmos y calculando nuevamente la matriz de covarianzas, realicemos en SPSS lo siguiente:

Pasos:

- Para transformar las tres variables primero tenemos que realizar un cálculo, para ello seleccionemos en el menú Transformación – Compute variable –; luego nombramos la nueva variable como $\text{Logx1}=\ln(x1)$, tal como se observa en la imagen siguiente:



- Realizamos esto para las tres variables, y repetimos los pasos, como en el procedimiento anterior, para calcular la matriz de covarianzas. Así tenemos como resultado:

Correlations		Logx1	Logx2	Logx3
Logx1	Pearson Correlation	1	.686**	-.791**
	Sig. (2-tailed)		.000	.000
	Sum of Squares and Cross-products	11.970	5.011	-6.303
	Covariance	.363	.152	-.191
	N	34	34	34
Logx2	Pearson Correlation	.686**	1	-.205
	Sig. (2-tailed)	.000		.244
	Sum of Squares and Cross-products	5.011	4.460	-.998
	Covariance	.152	.135	-.030
	N	34	34	34
Logx3	Pearson Correlation	-.791**	-.205	1
	Sig. (2-tailed)	.000	.244	
	Sum of Squares and Cross-products	-6.303	-.998	5.304
	Covariance	-.191	-.030	.161
	N	34	34	34

** . Correlation is significant at the 0.01 level (2-tailed).

1. Se observa que los logaritmos modifican los resultados. Los datos ahora son más homogéneos y la variable que tenía mayor varianza, ahora pasa a ser la menor varianza (el logaritmo de la proporción de beneficios); mientras que la variable que tenía la menor varianza, ahora ocupa el segunda lugar (el coeficiente entre precio por acción y beneficios). Además, la relación entre el logaritmo del ratio/beneficio (x3) y la rentabilidad efectiva (x1) es negativa; mientras que las otras relaciones son débiles.

Una ventaja adicional de los logaritmos es que hace las variables independientes de la escala de medida. Si multiplicamos las variables por una constante al tomar logaritmos, esto es equivalente a sumar una cantidad, y sumar una constante a los datos no altera su variabilidad. Por tanto, al tomar logaritmos en las variables, las varianzas pueden compararse aunque los datos tengan distintas dimensiones. La varianza media de las tres variables viene dada por:

$$Var = \frac{0.36 + 0.13 + 0.16}{3} = 0.213$$

La cual parece describir razonablemente la variabilidad de las variables.

1.9. Análisis gráfico y datos atípicos: histogramas y diagramas de dispersión

En esta sección vamos a continuar la descripción de datos multivariantes, estudiando su representación gráfica y las posibles transformaciones de las variables que conduzcan a una descripción más simple de los datos. También introduciremos un análisis inicial de la homogeneidad de la muestra, mediante el estudio de los posibles valores atípicos, debido a errores de medida u otras causas de heterogeneidad.

Recordemos que las correlaciones miden las relaciones lineales entre las variables y pueden ser mal interpretadas cuando las relaciones no son lineales. Por esa razón se intenta transformar las variables para que las variables transformadas tengan relaciones aproximadamente lineales. Por último, los datos multivariantes contienen con frecuencia observaciones que son heterogéneas con el resto y que pueden alterar completamente el análisis descriptivo de las variables.

El primer paso de cualquier análisis multivariante consiste en representar en forma gráfica las variables individualmente, mediante un histograma o un diagrama de caja. Estas representaciones son muy útiles para detectar asimetrías, heterogeneidad, datos atípicos, etc. En segundo lugar, conviene construir los diagramas de dispersión de las variables por pares, y esta posibilidad se incluye ya en muchos programas de computadoras como el SPSS. Con p variables existen $p(p-1)/2$ gráficos posibles, que pueden disponer en forma de matriz para entender el tipo de relación existente entre pares de variables e identificar puntos atípicos en la relación bivalente. En particular, estos gráficos son importantes para apreciar si existen relaciones no lineales, en cuyo caso la matriz de covarianzas puede no ser un buen resumen de la dependencia entre las variables.

Otros ejemplos:

1. En el Software SPSS generar⁶ el anexo *datosnormales.sav*, posteriormente realizar lo siguiente, de acuerdo a las sentencias en capturas de pantalla, como se muestra a continuación:
 - a) Graphs-chartbuilder-histogram (“click” dos veces). Luego arrastrar z_1 frente a z_2 . Seguido Bar.

⁶ Para generar números aleatorios y almacenarlos en una nueva variable, utilizar la expresión RV.NORMAL(0,1). Ver literal b), de este mismo ejercicio.

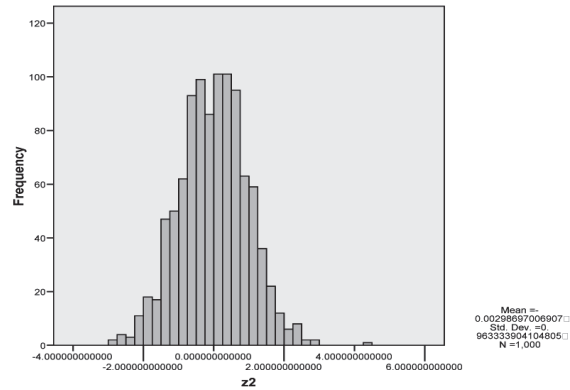
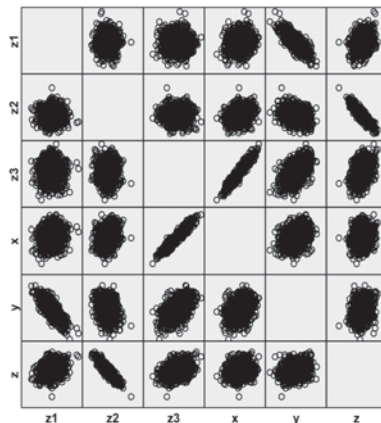
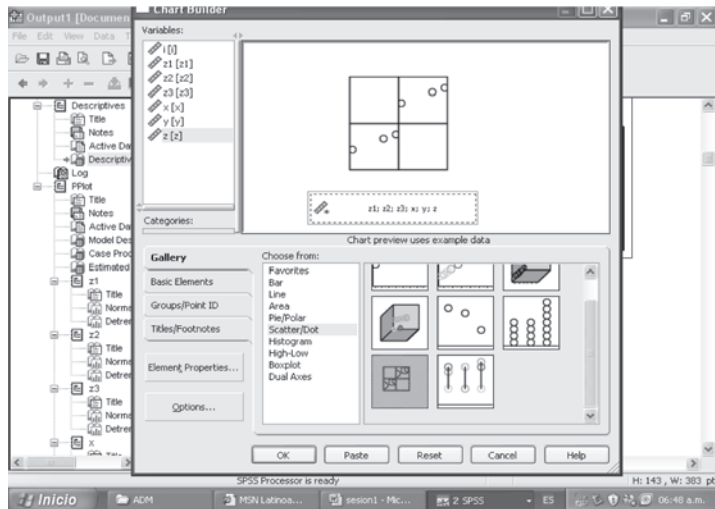
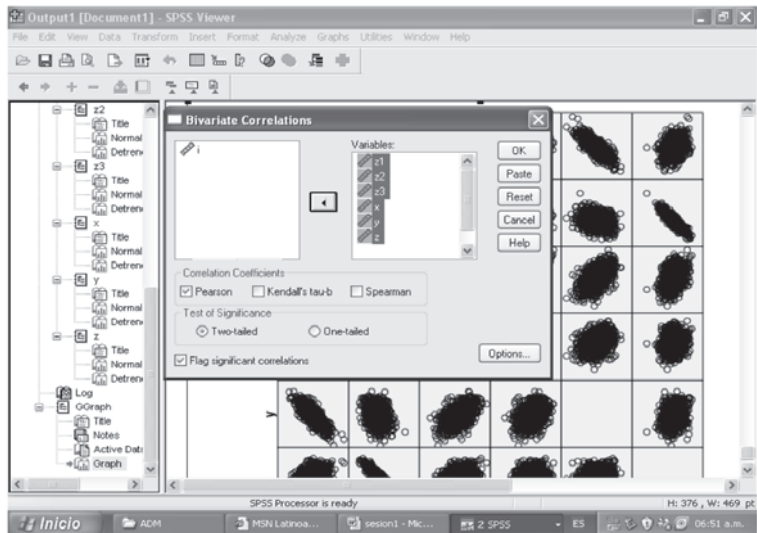


Gráfico Nube de puntos. (/Scatter/Dot Scatterplots Matrix).
(Puede aplicar otras técnicas como Gráfico de probabilidad PP. Normalidad)



Las variables z1 con z son independientes.

Ahora, obtener las correlaciones (Analizar/Correlate/Bivariate).

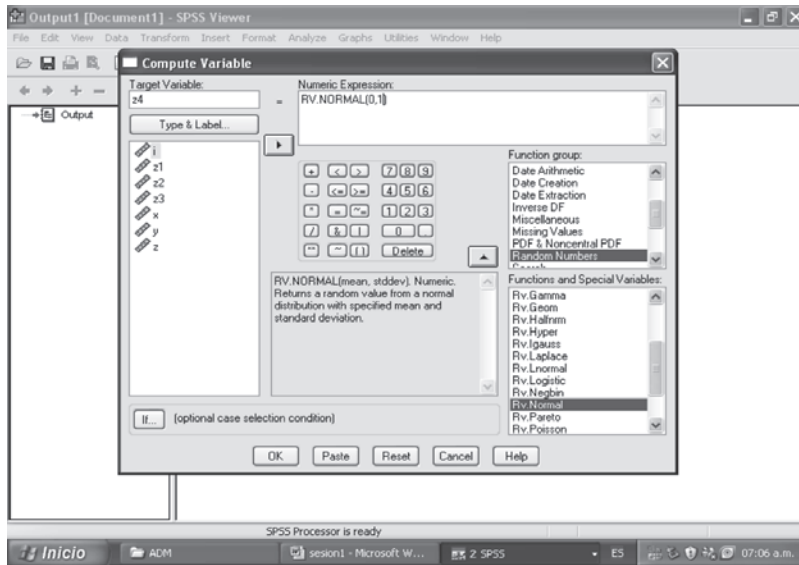


		z1	z2	z3	x	y	z
z1	Pearson Correlation	1	-.045	.011	.157**	-.808**	.324**
	Sig. (2-tailed)		.157	.727	.000	.000	.000
	N	1000	1000	1000	1000	1000	1000
z2	Pearson Correlation	-.045	1	-.003	.249**	-.236**	-.898**
	Sig. (2-tailed)	.157		.932	.000	.000	.000
	N	1000	1000	1000	1000	1000	1000
z3	Pearson Correlation	.011	-.003	1	.955**	.515**	.342**
	Sig. (2-tailed)	.727	.932		.000	.000	.000
	N	1000	1000	1000	1000	1000	1000
x	Pearson Correlation	.157**	.249**	.955**	1	.302**	.145**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	1000	1000	1000	1000	1000	1000
y	Pearson Correlation	-.808**	-.236**	.515**	.302**	1	.156**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	1000	1000	1000	1000	1000	1000
z	Pearson Correlation	.324**	-.898**	.342**	.145**	.156**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	1000	1000	1000	1000	1000	1000

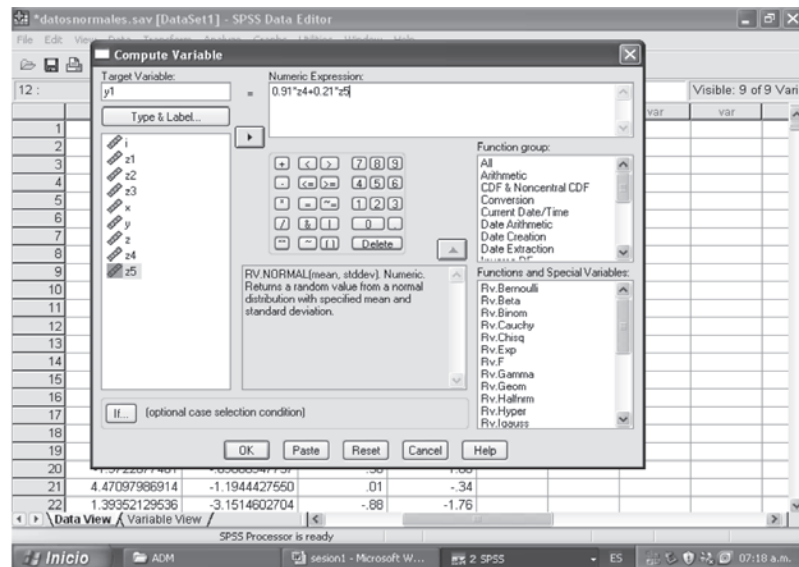
** . Correlation is significant at the 0.01 level (2-tailed).

Se observa que existe relación lineal entre la variable z3 con x.

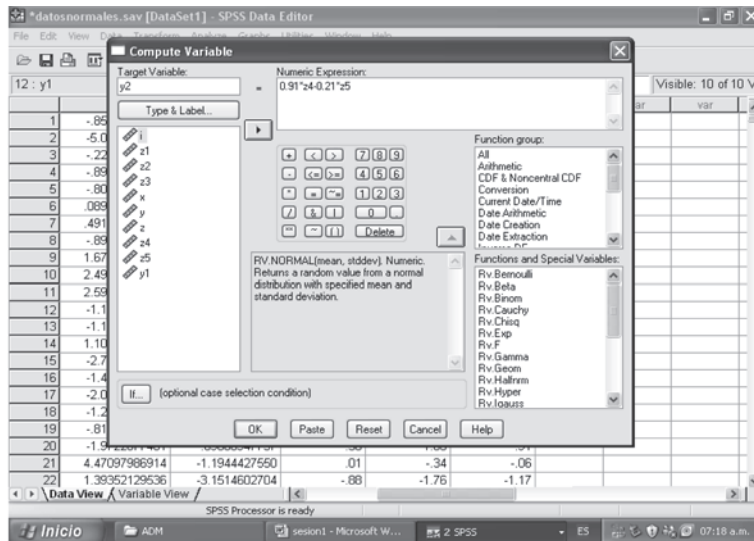
- Generar dos variables aleatorias (z4, z5) normales (0,1)
- Obtener las variables (y1, y2), según se indica en el cuadro de diálogo.
¿Cuál es el resultado?



Calculando y1.



Calculando y2.



Ahora, calculamos Matriz de correlaciones de (y1, y2).

Correlations

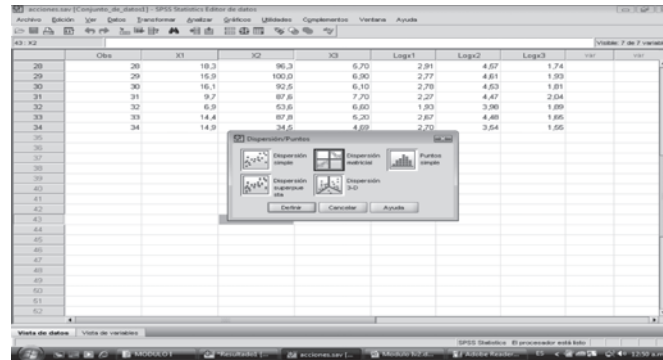
		y1	y2
y1	Pearson Correlation	1	.891**
	Sig. (2-tailed)		.000
	N	1000	1000
y2	Pearson Correlation	.891**	1
	Sig. (2-tailed)	.000	
	N	1000	1000

**. Correlation is significant at the 0.01 level

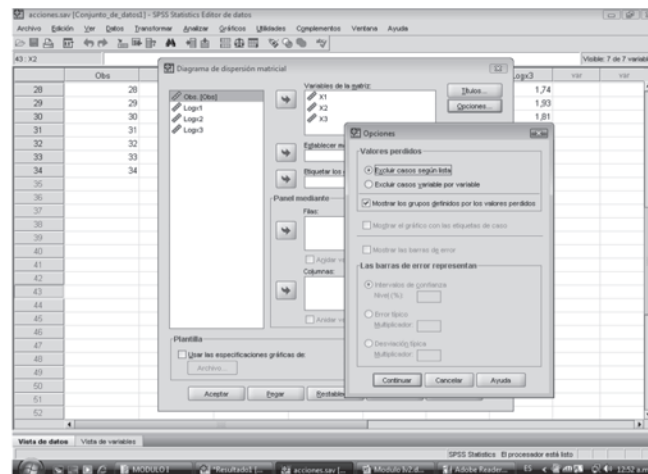
- Utilizando el archivo *acciones.sav*, generar el gráfico de dispersión para las tres variables X1, X2, X3, luego realizar el análisis respectivo.

Para poder generar el gráfico de dispersión, utilicemos el SPSS y sigamos los siguientes pasos:

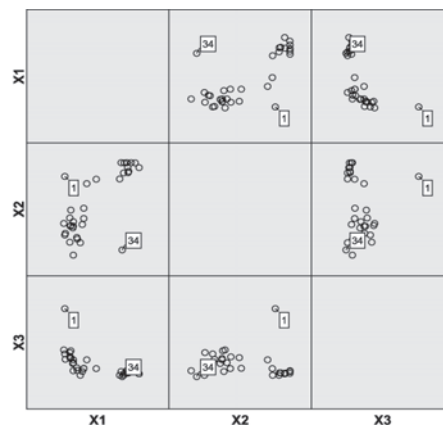
- En Gráficos, seleccionar cuadro de diálogos antiguos, y en dispersión/puntos, seleccionar dispersión matricial, así como se muestra en la figura siguiente:



- Una vez seleccionado dispersión matricial, arrastrar las variables X1, X2 y X3 a la lista de variables. Veamos la siguiente imagen:



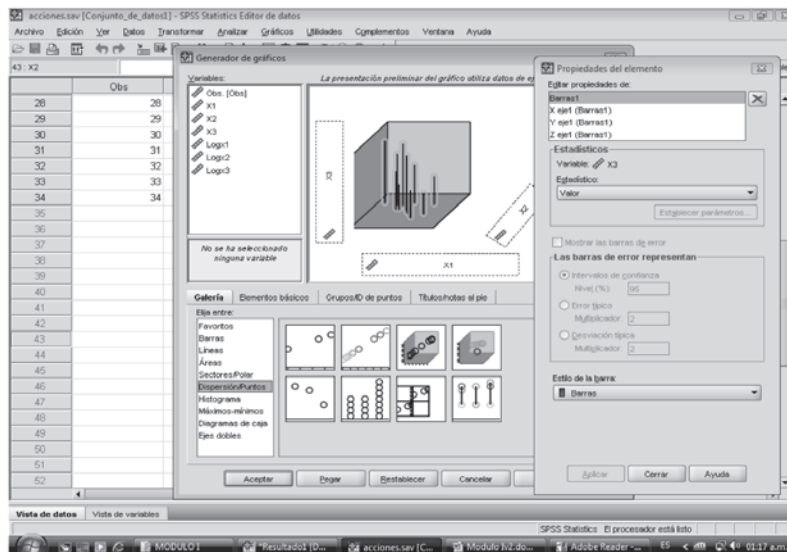
- Al hacer “click” en Continuar se genera el siguiente gráfico de dispersión matricial para las tres variables.



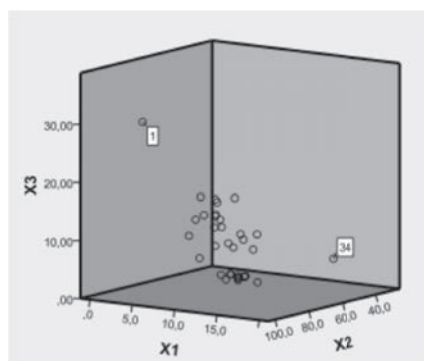
En el gráfico de dispersión se observa cómo la primera observación (1) aparece como un valor atípico en todos los diagramas de dispersión. En los gráficos con la tercera variable este punto es un valor muy extremo en la relación, mientras en los gráficos de las otras dos variables aparece como atípico; pero no parece muy influyente en la relación lineal entre ambas variables. El dato 34 aparece, al igual que la observación 1, como una observación aislada heterogénea con el resto.

En este caso, podemos hacer una representación en tres dimensiones de las tres variables. Para ello, utilicemos el SPSS y sigamos los pasos siguientes:

- En generador de gráficos, seleccionar dispersión/puntos y ubicar el cursor en 3-D (Simple), luego arrastrar las variables X1, X2 y X3 en cada eje, según se muestra en la figura siguiente:



- Luego de seguir las sentencias anteriores, se tiene el gráfico de dispersión 3D para los tres indicadores económicos.



En la figura anterior se observa que la relación entre las variables X_1 y X_2 depende del nivel de la X_3 . Esta figura ilustra también con claridad el carácter claramente atípico de las observaciones 1 y 34, que aparecen muy separadas del resto. Por otro lado, se observa en el gráfico tridimensional que las observaciones se agrupan en dos conjuntos distintos. Esta característica, que se apunta en los gráficos bidimensionales, aparece claramente de manifiesto en la representación tridimensional, ilustrando las ventajas de construir estas representaciones cuando sea posible.

1.10. Ejercicios

Ejercicio 1

Los datos de la tabla 3 corresponden a restaurantes construidos por diez promotoras que operan a lo largo de la costa salvadoreña.

Tabla 3. Diez promotoras de la costa salvadoreña

Promotora	X_1 = Duración media hipoteca (años)	X_2 = Precio medio (miles de dólares)	X_3 = Superficie media (m^2) de cocina
1	8.7	0.3	3.1
2	14.3	1.8	9.0
3	18.9	1.8	9.0
4	19.0	0.8	9.4
5	20.5	0.9	8.3
6	14.7	1.1	7.6
7	18.8	2.5	12.6
8	37.3	2.7	18.1
9	12.6	1.3	5.9
10	25.7	3.4	15.9

Se solicita:

- Para X_1 y X_2 calcúlense, respectivamente, las medias muestrales $\bar{x}_1$ y $\bar{x}_2$, las varianzas muestrales s_{11} y s_{22} , la correlación entre ambas, r_{12} . Interpretese el valor obtenido de r_{12} .
- Utilizando la matriz de datos X y la de centrado P , definida anteriormente, calcúlense el vector de medias muestrales $\bar{x}_1$ y la matriz de covarianzas muestrales S . A partir de ésta obténgase la matriz de correlaciones R .⁷

⁷ Según las medidas de dependencia lineal, la matriz de correlación viene dada por la siguiente fórmula:
 $R = D^{-1/2} S D^{-1/2}$, $S = D^{1/2} R D^{1/2}$.

Ejercicio 2

Partiendo de la matriz de covarianzas S para los logaritmos de las acciones (anexo *acciones.sav*), del ejercicio anterior, calcular:

- a) La varianza generalizada.
- b) La varianza efectiva.
- c) La desviación efectiva.

Ejercicio 3

A partir de los tres indicadores económicos X_1, X_2, X_3 del ejemplo 1, se construyen dos nuevos indicadores:

$$y_1 = (1/3)x_1 + (1/3)x_2 + (1/3)x_3$$

$$y_2 = x_1 - 0.5x_2 - 0.5x_3$$

Se solicita:

Calcular el vector de medias para $y' = (y_1, y_2)$, su matriz de varianzas y covarianzas, la matriz de correlación y la varianza generalizada.

Ejercicio 4

Realizar el mismo ejercicio anterior para los logaritmos de las variables de los tres indicadores económicos; luego, analizar dichos resultados y compararlos con el ejemplo sin la transformación debida.

Ejercicio 5

Utilizar el anexo *mundodes.sav* y construir los diagramas de dispersión mediante el software SPSS. Realizar el análisis respectivo. El anexo *mundoes.sav* consta de 91 observaciones y 6 variables. Las observaciones corresponden a 91 países (las variables son indicadores de desarrollo humano). Las seis variables son estas:

- Tasa Nat: Ratio de natalidad por 1000 habitantes.
- Tasa Mort: Ratio de mortalidad por 1000 habitantes.
- Mort. Inf: Mortalidad infantil (por debajo de un año).
- Esp.Hom: Esperanza de vida en hombres.
- Esp.Muj: Esperanza de vida en mujeres.
- PNB: Producto Nacional Bruto *per cápita*.

2. Análisis de regresión lineal múltiple (RLM)

RLM es el método explicativo de análisis multivariante más conocido. Consiste en explicar una variable, llamada variable dependiente, mediante un conjunto de variables explicativas (llamadas independientes o covariables).

- ¿Cuándo se debería utilizar RLM?
- ¿Puede explicarse una variación de las ventas de un producto en términos de la variación de precios y en la fuerza de ventas?
- ¿Están determinadas las percepciones de los consumidores (sobre la calidad del producto) por su percepción de los precios, de la imagen de marca y del servicio postventa?
- ¿Cuánto contribuye el presupuesto destinado a promociones a explicar la variación de las ventas del producto, manteniendo constantes los niveles de precio y el nivel de inversión publicitaria?

Por ejemplo, el director de mercadeo de un hotel de montaña puede estar interesado en determinar cuáles son las variables que mejor explican que un turista visite sus instalaciones (n=217 clientes).

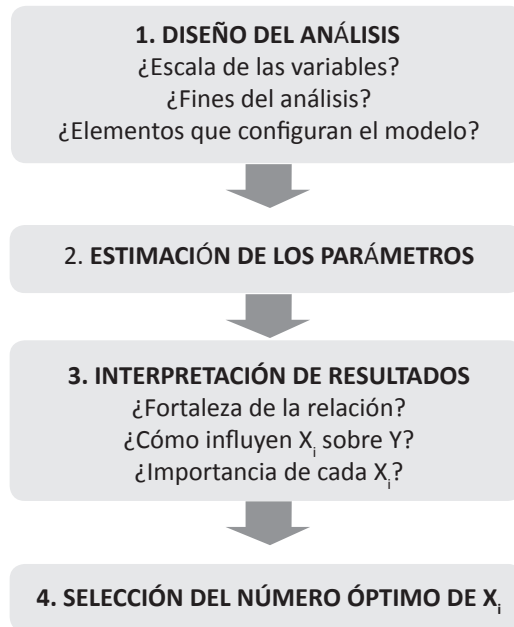
- número de días visitados
- edad
- años de hacer turismo
- gasto medio por día durante la estancia
- satisfacción general (de 0 a 10)
- número de personas con las que visita el hotel

¿Cuáles son los fines del RLM?

- Determinar la estructura o la forma de la relación (la ecuación matemática que relaciona las variables explicativas con la variable a explicar);

- Determinar si las X_i explican una variación significativa de la variable a explicar (dependiente); determinar si la relación existe;
- Determinar la importancia de la relación de asociación entre las variables explicativas y la variable a explicar; determinar la fuerza de la relación.

2.1. Metodología



1. Diseño del análisis

- La variable dependiente (VD) y las variables independientes (VI) deben ser cuantitativas.
- En un modelo RLM se distinguen: variables (Y , X_p), parámetros o coeficientes de regresión (valores que relacionan las X_i con Y ; β_p)⁸, término residual (recogen las variaciones de las X_i no cubiertas por el modelo, las especificaciones incorrectas de la forma funcional del modelo, los errores en la medición de las variables, el comportamiento cambiante e imprevisible de los sujetos entrevistados) y ecuación ($Y = B_0 + B_1X_1 + \dots + B_pX_p + e$).

⁸ B_0 es el término constante, indica el valor esperado de Y cuando todas las X_i son iguales a cero. El valor B_p ($p \neq 0$) expresa el cambio esperado en la variable Y , debido a un cambio de una unidad en X_p cuando el resto de variables explicativas X_{p-1} se mantienen constantes.

• Supuestos:

- Linealidad. La relación entre las X_i y Y es lineal.⁹
- Ausencia de multicolinealidad. No debe existir correlación entre las X_i , si es el caso, no podrían interpretarse los parámetros de la regresión.
- Homocedasticidad. El término residual debe seguir una distribución normal con media igual a cero y con una varianza constante.

2. Estimación de los parámetros B_p

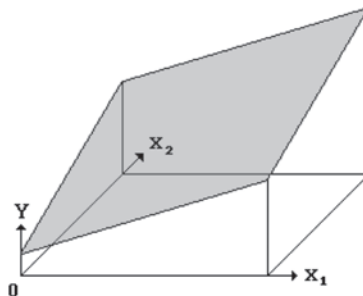
- Esto permite obtener la ecuación siguiente:¹⁰

$$Y_i = E(Y_i/X_{1i}, X_{2i}, \dots, X_{pi}) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_p X_{pi}$$

$$Y_i = \hat{Y}_i + e$$

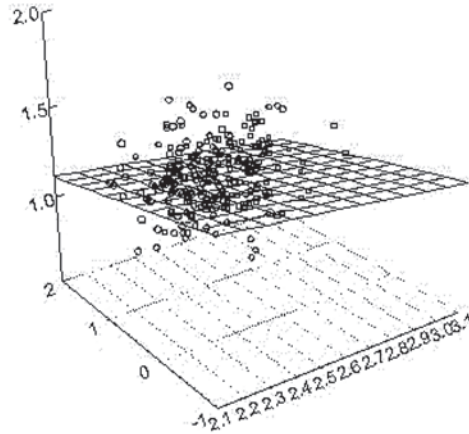
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

- En la Regresión Lineal Simple (RLS), la estimación de los parámetros significaba, desde el punto de vista geométrico, la obtención de la línea recta que mejor ajustaba la nube de puntos; en la RLM la obtención del vector de parámetros permitirá ajustar un plano a la nube de puntos (en el caso de dos variables explicativas), un *hiperplano* en el caso de más de dos variables explicativas.



⁹ Los cambios en Y motivados por la variación en una de las X_i son constantes y dependen únicamente de los valores de dicha variable explicativa. Si la relación no es lineal se debe recurrir a otros métodos.

¹⁰ Observar que los $\hat{\beta}_i$ son los mejores estimadores de β_i .



- El método más común de estimación de los parámetros es la estimación por mínimos cuadrados, cuyo procedimiento se basa en la minimización del término residual. Este procedimiento utiliza la descomposición de la variación total de Y en dos fuentes de variación: la procedente de las X_i (la variación explicada por el modelo) y la variación del término residual (la variación no explicada por el modelo).

$$SC_y = SC_{reg} + SC_{res}$$

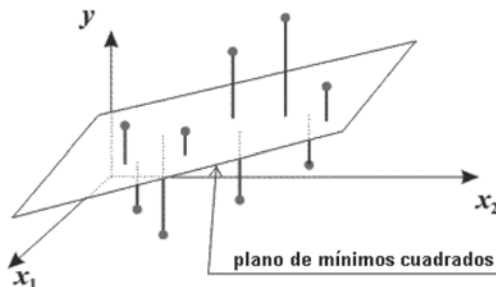
Siendo:

$$SC_y = \sum_{i=1}^n (y_i - \hat{y})^2$$

$$SC_{reg} = \sum_{i=1}^n (\hat{y}_i - \hat{y})^2$$

$$SC_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

En el caso de dos variables explicativas, los términos residuales pueden representarse como:



La estimación de los parámetros por mínimos cuadrado nos lleva a la minimización de SC_{res}

$$\min \sum_{i=1}^n e_i^2 = \min \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Obteniendo la siguiente estimación de los parámetros:

$$[\hat{\beta}] = [X'X]^{-1} [X'] [Y]$$

Donde:

[X] Es la matriz que contiene los distintos valores de las variables explicativas.
[Y] Es la matriz que contiene los distintos valores de la variable a explicar.

Requerimientos:

1. El tamaño de la muestra n no debe ser inferior o igual al número de variables explicativas P .
2. Una o algunas X_i están altamente correlacionadas con otra X_i (multicolinealidad). Si esto sucede, los parámetros estimados no son estables y su interpretación es poco realista. Una forma para detectar la multicolinealidad consiste en analizar las correlaciones entre las X_i . Si, por ejemplo, X_1 y X_2 están muy correlacionadas ($R > 0.7$), bastara con utilizar en el modelo de regresión una de las dos variables.¹¹
3. Interpretación de los resultados. Para validar las estimaciones introducimos las variables al procedimiento estadístico de SPSS y encontramos los valores de los coeficientes B_i .

Para una correcta interpretación de resultados se deben tomar en cuenta estos pasos:

1. Existencia y fortaleza de la relación. Se refiere a la relación de asociación entre las Y y X_i , dado por el cuadrado del coeficiente de correlación múltiple (R^2), que indica la proporción de la variación de Y que viene explicada por la variación de las X_i .

El coeficiente R^2 tiene en cuenta las dos fuentes de variación:

$$R^2 = \frac{SC_{reg}}{SC_y} = \frac{SC_y - SC_{res}}{SC_y}$$

11 Para eliminar problemas de multicolinealidad se puede aplicar a las X_i iniciales un análisis factorial de componentes principales, cuyo resultado será la creación de unas nuevas variables incorrelacionadas.

Si, por ejemplo, se obtuviera un valor de R^2 igual a 0.70141, implicaría que aproximadamente un 70% de la variación de Y viene explicada por las X_i . En la práctica se considera que los dos conjuntos de variables están relacionadas cuando el valor de R^2 es mayor o igual a 0.7.

En el caso de modelos con muchas variables explicativas, R^2 aumentará a medida que se vayan agregando más variables explicativas en el modelo. Para medir la fortaleza de la relación es preferible utilizar un R^2_{ajustado} (Adjusted R square), el cual se calcula en función del número de X_i y del tamaño de la muestra, a través de la fórmula siguiente:

$$R^2_{\text{ajustado}} = R^2 - (p(1 - R^2))/(n - p - 1),$$

Donde p = número de X_i ; n = tamaño de la muestra; R^2 = cuadrado del coeficiente de correlación múltiple.

2. Interpretación de los coeficientes. Indica la dirección de la relación entre la variable Y y las X_i . El signo asociado permite establecer en qué sentido varía la variable Y ante una variación de la variable X_p correspondiente.

Un $B_p > 0$ significa que un incremento, en una unidad, de la variable explicativa asociada X_p , implica un incremento de la variable a explicar Y en B_p unidades, siempre y cuando las demás variables explicativas se mantengan constantes.

Si $B_p < 0$ significa que un incremento, en una unidad, de la variable explicativa asociada X_p , implica una disminución de la variable a explicar Y en B_p unidades, siempre y cuando las demás variables explicativas se mantengan constantes.

Importancia de cada X_i . Considerando que las X_i pueden consistir de medidas de escalas distintas (años, dólares, número de personas, etc.), los coeficientes B_p obtenidos no son comparables entre sí. Para poder compararlos y así evaluar la importancia de cada X_i en la explicación de las variaciones de la variable a explicar, se utilizan los coeficientes normalizados NB_p , los cuales se obtienen a partir de las X_i normalizadas, es decir transformadas de tal forma que tengan el mismo valor medio (igual a cero) y la misma varianza (igual a 1).

Siendo S_{xp} y S_y las desviaciones típicas de Y y X_p respectivamente.

El peso real o importancia de cada X_i en la explicación de las variaciones de la variable a explicar, se obtiene a partir de:

$$IX = \frac{NB_p}{\sum_{p=1}^p |NB_p|} * 100$$

3. Pruebas de significación

- i. Prueba para determinar si cada variable por sí sola influye significativamente sobre la variable a explicar.

La hipótesis planteada es esta:

$$H_0: \beta_p = 0, \forall p = 1, \dots, P$$

$$H_1: \beta_p \neq 0$$

La prueba utilizada es la prueba t, cuyo resultado se presenta en los resultados que genera el programa estadístico SPSS (Estadístico t con n-p-1 g.l.):

$$t = \frac{\beta_p}{SE_{\beta_p}}$$

$$SE_{\beta_p} = \frac{S_{\beta_p}}{\sqrt{n}}$$

Siendo SE_{β_p} el error estándar de β_p ; S_{β_p} la desviación típica de β_p .

- ii. Prueba de significación global de la relación entre todas las Y y X_i . Esto es, indagar si globalmente el modelo de regresión planteado es significativo, equivale a decir:

$$H_0: \beta_1 = \beta_2 = \dots \beta_p = 0.$$

$$H_1: \text{como mínimo } \beta_1 \text{ es diferente de cero.}$$

Para ello se utiliza la prueba F, cuyo estadístico es:

$$F = \frac{\frac{SC_{reg}}{P}}{\frac{SC_{res}}{(n-P-1)}}$$

Con P y (n-P-1) grados de libertad.

Si el resultado de la prueba F indica que la relación entre los dos conjuntos de variables es significativa ($\text{Sig} < .000$), implicará que tenemos una probabilidad nula de equivocarnos si rechazamos la hipótesis nula.

4. Selección del número óptimo de variables explicativas

Para establecer la ecuación de regresión definitiva es necesario eliminar de la misma las variables explicativas cuyos parámetros asociados no son significativos y volver a efectuar el proceso de cálculo con el objetivo de obtener las estimaciones correctas de los parámetros. Para ello están disponibles procedimientos para seleccionar el número óptimo de variables explicativas a incluir en la ecuación de regresión.

Los métodos para seleccionar el número óptimo de variables explicativas tienen como objetivo seleccionar, entre un amplio número de variables explicativas, las que mejor expliquen las variaciones en la variable a explicar. Teniendo en cuenta lo siguiente:

- El coeficiente de correlación múltiple R^2 aumenta en la medida en que se adicionan variables a la ecuación; pero a partir de cierto punto, el incremento de R^2 para cada nueva variable que se añade es insignificante.
- Un buen modelo no debe presentar demasiadas variables ni debe olvidar las que sean verdaderamente relevantes.

Existen tres procedimientos para seleccionar el número óptimo de variables explicativas, así:

1. Método *forward selection*.
2. Método *backward elimination*.
3. Método *stepwise*.

Antes de explorar estos métodos, iniciaremos ejecutando un procedimiento básico a través del método tradicional (Enter) y luego realizaremos una verificación de los supuestos del modelo.

2.2. Aplicación del análisis de regresión múltiple

El modelo de dos variables es limitado e inadecuado en la práctica. En la realidad hay más variables que afectan a la variable dependiente. La lógica es la misma; la variable dependiente o regresada Y depende de dos o más variables explicativas o regresoras.

Se trata de modelos de regresión lineal múltiple, modelos lineales en los parámetros que pueden o no ser lineales en las variables.

El modelo hereda las propiedades del modelo de dos variables, los coeficientes de regresión se obtienen a través del método MCO (Mínimos Cuadrados Ordinarios).

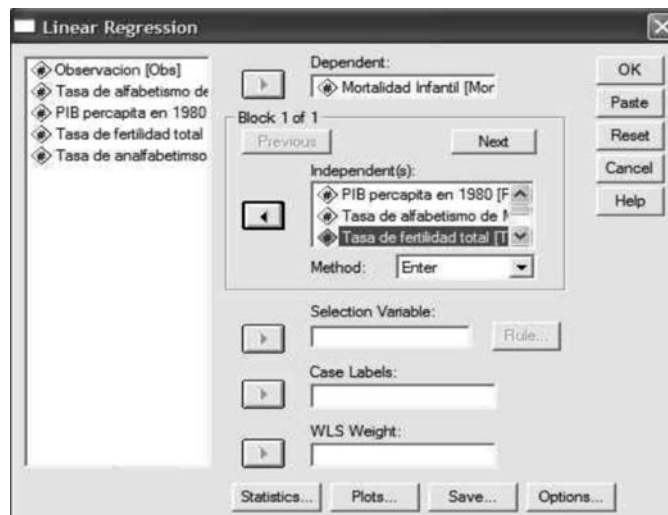
R^2 es la proporción de la variación en Y , explicada por las variables regresoras X_i conjuntamente (de manera similar, mide la bondad del ajuste de la ecuación de regresión, es decir, la proporción de la variación total en la variable dependiente Y , explicada por las variables explicativas X_i). Si R^2 es 1 la recta de regresión ajustada explica el 100% de la variación en Y ; si es cero, el modelo no explica nada de la variación en Y (el modelo es mejor entre más se acerca a 1).

R es una medida del grado de asociación entre Y y todas las X_i conjuntamente.

Retomemos un archivo de datos, el cual contiene información referida a la mortalidad infantil, tasa de alfabetismo de mujeres, el PIB *per cápita* en 1980 y la tasa de fertilidad total de 64 países. Usaremos dicha tabla para mostrar paso a paso la forma de encontrar un modelo que se ajuste a los datos y veremos la forma de utilizar el programa estadístico SPSS e interpretar los resultados.

Primero recuperemos el archivo en el editor de datos (el anexo Mortalidad_Infantil_RLM.sav). Ejecutemos el procedimiento de regresión lineal y definamos como variable dependiente la mortalidad infantil y como variables independientes el PIB *per cápita*, la tasa de alfabetismo y la tasa de fertilidad total.

Cuando se introducen dichas variables en el modelo, se requiere eliminar la influencia neta de cada regresora (se requiere estimar los coeficientes de regresión parcial de cada regresora).



La tabla “model summary” se refiere a la calidad del modelo de regresión (bondad del ajuste): tomadas juntas las 3 variables independientes incluidas, explican un 74% de la variabilidad en la mortalidad infantil (variable dependiente).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.860 ^a	.740	.727	39.72990

a. Predictors: (Constant), Tasa de fertilidad total, PIB *per cápita* en 1980, Tasa de alfabetismo de Mujeres-en porcentaje

La tabla ANOVA muestra el estadístico F, el cual contrasta la Ho, de que el valor poblacional de R=0 y, por tanto, permite decidir si existe relación lineal significativa entre la mortalidad infantil y el conjunto de variables independientes tomadas juntas.

El valor del nivel crítico (Sig.=0.0000), puesto que es menor que 0.05, indica que sí existe relación lineal significativa.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	268970.08	3	89656.695	56.800	.000 ^a
	Residual	94707.915	60	1578.465		
	Total	363678.00	63			

a. Predictors: (Constant), Tasa de fertilidad total, PIB *per cápita* en 1980, Tasa de alfabetismo de Mujeres-en porcentaje
b. Dependent Variable: Mortalidad Infantil

La tabla de coeficientes contiene toda la información necesaria para construir la ecuación de regresión.

Coefficients³

Model	unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerancia	VIF
1 (Constant)	163.241	33.211		4.915	.000		
Tasa de alfabetismo de Mujeres-en porcentaje	-1.722	.250	-.593	-6.889	.000	.587	1.704
PIB <i>per cápita</i>	-.006	.002	-.199	-2.912	.005	.927	1.078
Tasa de fertilidad total	13.327	4.244	.265	3.140	.003	.611	1.638

a. Dependet Variable: Mortalidad infantil

$$MortInfantli = b_1 + b_2 PIBP_i + b_3 TAM_i + b_3 TFT_i + u_i$$

$$MInf_i = 163.241 - 0.006 PIBP_i - 1.722 TAM_i + 13.327 TFT_i$$

Dado que MInf (Mortalidad Infantil) es el número de muertes de niños menores de cinco años por cada 1000 nacimientos vivos, ¿cómo se interprete la ecuación de regresión?

El coeficiente de regresión parcial del PIB *per cápita* (-0.006) indica que si la influencia de la TAM y la TFT se mantiene constante, conforme el PIBP se incrementa, digamos en \$1, en promedio, la mortalidad infantil disminuye en 0.006 unidades; en otras palabras, si el PIBP se incrementa en \$1000, en promedio, el número de muertes de niños menores de cinco años se reduciría a seis por cada 1000 nacimientos vivos.

El coeficiente -1.722 señala que si la influencia del PIBP y la TFT se mantienen constantes, el número de muertes de niños menores de cinco años disminuirá, en promedio, 1.7 por cada 1000 nacimientos vivos, en tanto que la tasa de alfabetización en las mujeres subiría un punto porcentual.

El coeficiente 13.33 indica que si la TAM y el PIBP se mantienen constantes, la tasa de mortalidad infantil aumentaría, en promedio, en 13.33 por cada 1000 nacimientos vivos, en tanto que la tasa de fertilidad total subiría en un punto porcentual.

Los coeficientes de la ecuación de regresión indican la cantidad de cambio que se producirá en la variable dependiente por cada cambio de una unidad en la correspondiente variable independiente (manteniendo constantes el resto de variables independientes).

Los coeficientes señalan la importancia relativa de cada variable independiente en la ecuación de regresión.

Una variable tiene tanto más peso (importancia) en la ecuación de regresión, cuanto mayor (en valor absoluto) es su coeficiente de regresión.

La prueba t y sus niveles críticos sirven para contrastar la H_0 , de que un coeficiente de regresión vale cero en la población (un coeficiente de cero indica ausencia de relación lineal, de modo que los coeficientes significativamente distintos de cero informan sobre qué variables son relevantes en la ecuación de regresión).

Niveles críticos (Sig.) muy pequeños (generalmente menores a 0.05) indican que esa hipótesis debe ser rechazada.

En el ejemplo en mención, observando el nivel crítico asociado a cada prueba t, puede verse que las tres variables utilizadas poseen coeficientes significativamente distintos de cero. Las tres variables contribuyen de forma significativa al ajuste del modelo.

2.3. Supuestos del modelo de regresión lineal múltiple

Son una serie de condiciones que deben darse para garantizar la validez del modelo, así:

1. Análisis de residuos.
2. Linealidad. La variable dependiente es la suma de un conjunto de elementos (origen de recta, combinación lineal de variables independientes, residuos); el incumplimiento de linealidad se llama error de especificación.
3. Independencia. Los residuos del modelo son independientes entre sí, son variables aleatorias.
4. Homoscedasticidad. Para cada valor de X_i , la varianza de los residuos es constante.
5. Normalidad. Para cada valor de X_i , los residuos se distribuyen normalmente con media igual a cero, $E(u_i)=0$.
6. No colinealidad. No existe relación lineal exacta entre las variables independientes.

Cada uno de estos supuestos puede ser verificado a través del programa estadístico SPSS.

Análisis de residuos

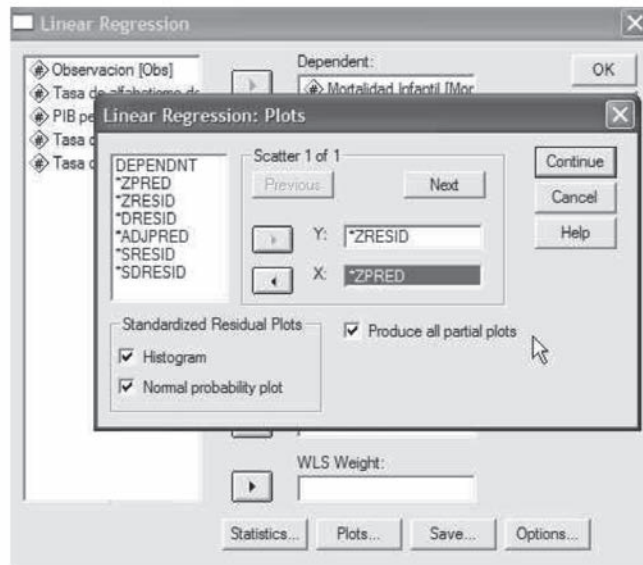
Los residuos son muy importantes en los modelos de regresión lineal, ya que el método MCO se basa en ellos (seleccionar los valores de los parámetros o estimar los coeficientes de tal forma que la suma de los residuos al cuadrado sea tan pequeña como sea posible).

Los residuos informan sobre el grado de exactitud de los pronósticos (cuanto más pequeño es el error típico de los residuos, mejores son los pronósticos; es decir, mejor se ajusta la recta de regresión a los puntos del diagrama de dispersión).

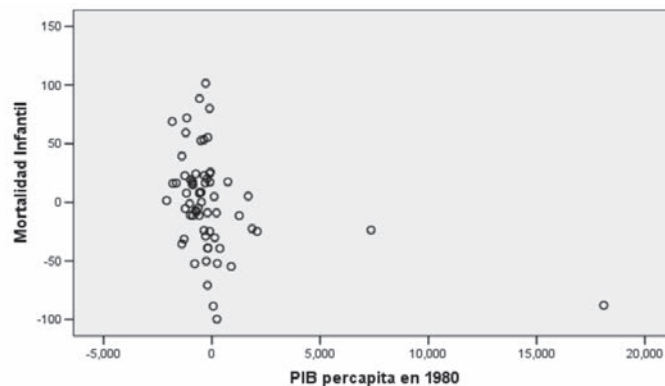
Linealidad

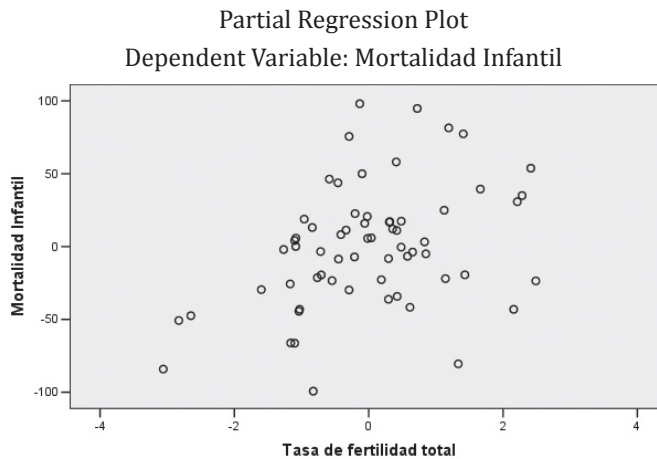
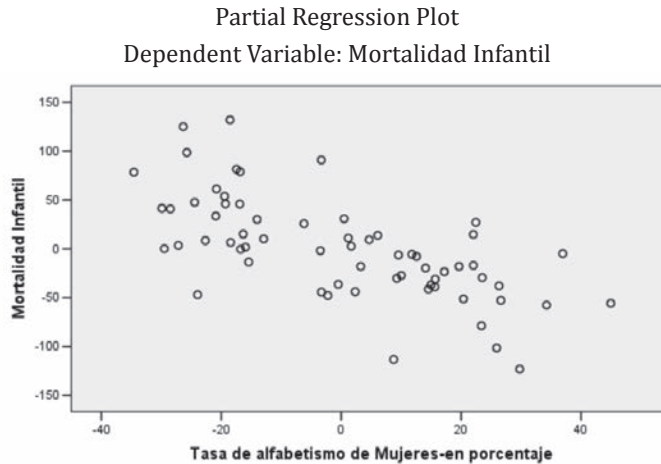
El programa SPSS permite generar un tipo particular de diagrama de dispersión llamado diagramas de regresión parcial. Estos ayudan a formarse una idea rápida sobre la forma que adopta una relación (genera tantos gráficos parciales como variables independientes incluya el análisis).

La opción está en el botón <plots> ... Produce all partial plots.



Partial Regression Plot
Dependent Variable: Mortalidad Infantil

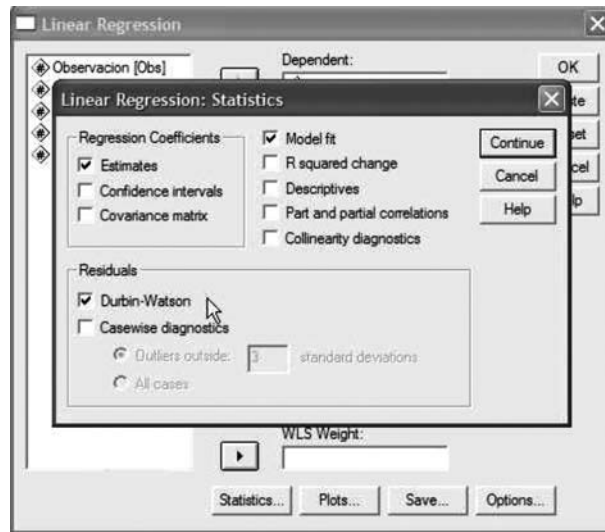




Estos gráficos permiten ver la relación existente entre la variable dependiente y cada una de las variables independientes por separado (tras eliminar de ellas el efecto del resto de variables independientes incluida en el análisis). No están basados en las puntuaciones originales de las tres variables representadas, sino en los residuos obtenidos al llevar a cabo un análisis de regresión con el resto de variables independientes.

Independencia

El programa SPSS utiliza un coeficiente (estadístico Durbin-Watson--DW) para verificar la independencia de los residuos. Un valor de $DW=2$ indica que son independientes completamente (en realidad, si $1.5 \leq DW \leq 2.5$ asegura la independencia).



La opción está en el botón <Statistics>.

El Visor de SPSS muestra un DW de 2.1., lo cual asegura una independencia entre los residuos del modelo.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.860 ^a	.740	.727	39.72990	2.143

a. Predictors: (Constant), Tasa de fertilidad total, PIB percapita en 1980, Tasa de alfabetismo de Mujeres-en porcentaje

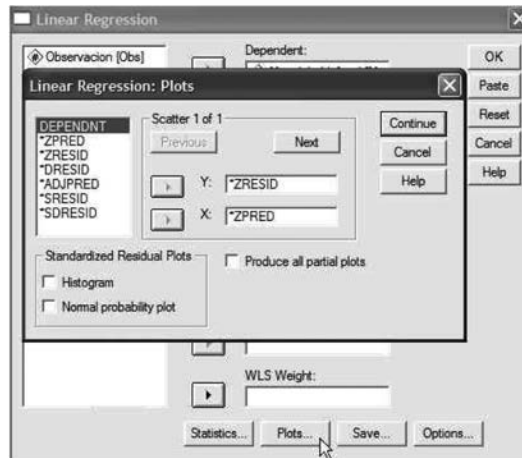
b. Dependent Variable: Mortalidad Infantil

Homoscedasticidad

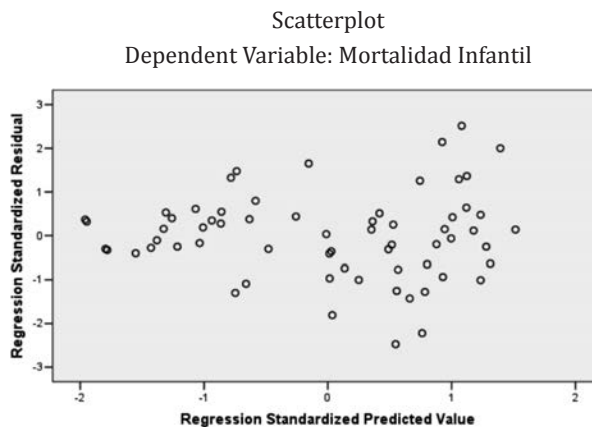
Homoscedasticidad implica igual varianza de los residuos. Dado el valor de X_i , la varianza de u_i es la misma para todas las observaciones (de otra manera, la variación alrededor de la recta de regresión es la misma para los valores de X_i ; ni aumenta ni disminuye conforme la variable independiente varia).

Para comprobar la homoscedasticidad no es necesario realizar complejos cálculos; en SPSS hay dos variables (ya creadas por el programa) cuyo diagrama de dispersión informa sobre el supuesto de homoscedasticidad ZPRED y ZRESID; el primero se grafica sobre el eje X y el segundo sobre el eje Y, e indican los pronósticos tipificados y los residuos tipificados, respectivamente.

El supuesto de igualdad de varianzas implica que la variación de los residuos debe ser uniforme en todo el rango de los valores pronosticados. La prueba implica que el diagrama de dispersión de ZPRED y ZRESID no debe mostrar ninguna pauta de asociación (la nube de puntos no sigue ninguna pauta de asociación clara, ni lineal, ni de otro tipo).



La opción está en el botón <Plots>.

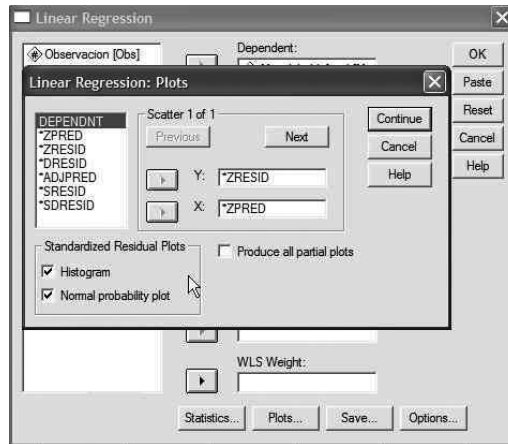


La gráfica muestra que los residuos y los pronósticos parecen ser independientes.

Normalidad

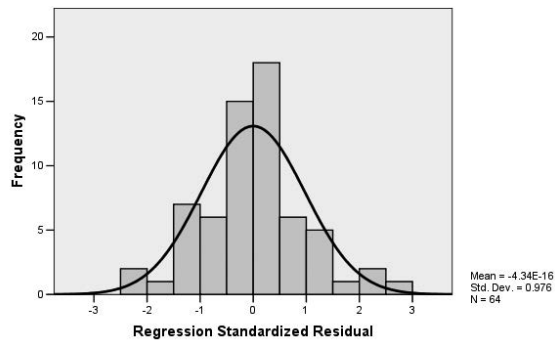
SPSS proporciona gráficas que informan sobre el grado en el que los residuos tipificados se aproximan a una distribución normal. El histograma de los residuos tipificados con una curva normal superpuesta informa si los residuos siguen el modelo de probabilidad normal.

La opción está en el botón <plots> ... Gráfica de residuos tipificados.



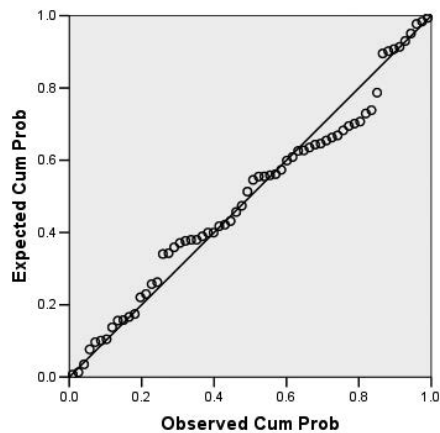
Histogram

Dependent Variable: Mortalidad Infantil



Normal P-Plot of Regression Standardized Residual

Dependent Variable: Mortalidad Infantil



El gráfico de probabilidad normal permite obtener un diagrama de probabilidad normal. Cuando los residuos se distribuyen normalmente, la nube de puntos se encuentra alineada sobre la diagonal del gráfico.

Colinealidad

Existe colinealidad perfecta cuando una de las variables independientes se relaciona de forma perfectamente lineal con una o más del resto de variables independientes de la ecuación.

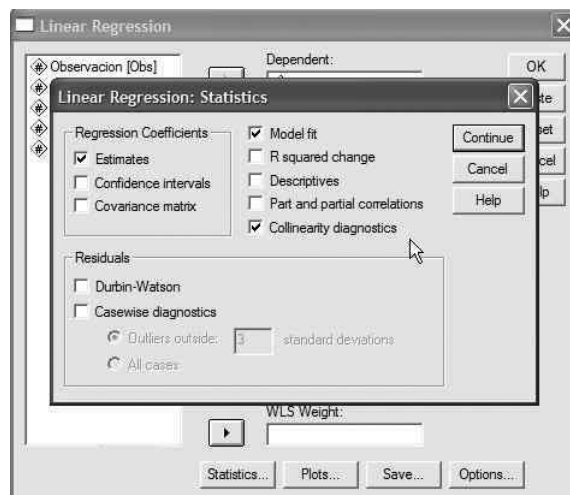
La colinealidad parcial se da cuando entre las X_i de una ecuación existen correlaciones altas (se podrían utilizar muchos indicadores económicos para construir una ecuación de regresión). Cuantas más variables hay en una ecuación, más fácil es que exista colinealidad.

Una observación importante es que r^2 no se altera por la presencia de colinealidad; pero, ¿cuál es el grado máximo de relación permisible entre las variables independientes? No existe un consenso generalizado.

Algunos indicios que pueden apuntar a la colinealidad en un análisis de regresión son estos:

- El estadístico F que evalúa el ajuste general de la ecuación de regresión es significativo, pero no lo es ninguno de los coeficientes de regresión parcial.
- Existen valores de tolerancia pequeños (próximos a 0.01). La tolerancia de una X_i es la proporción de varianza de esa variable que no está asociada (que no depende) del resto de variables X_i incluidas en la ecuación.

SPSS ayuda a detectar colinealidad.



La opción se encuentra en el botón <Statistics> ---Collinearity Diagnosis.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	163.241	33.211		4.915	.000		
	Tasa de alfabetismo de Mujeres-en porcentaje	-1.722	.250	-.593	-6.889	.000	.587	1.704
	PIB <i>per cápita</i>	-.006	.002	-.199	-2.912	.005	.927	1.078
	Tasa de fertilidad total	13.327	4.244	.265	3.140	.003	.611	1.638

a. Dependent Variable: Mortalidad Infantil

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	Tasa de alfabetismo de Mujeres-en porcentaje	PIB <i>per cápita</i>	Tasa de fertilidad total
1	1	3.071	1.000	.00	.01	.03	.00
	2	.710	2.080	.00	.00	.87	.01
	3	.205	3.870	.00	.38	.10	.06
	4	.014	14.861	.99	.61	.00	.93

a. Dependent Variable: Mortalidad Infantil

El nivel de tolerancia de una X_i se obtiene restando a 1 el r^2 resultante de regresar esa variable sobre el resto de X_i (valores de tolerancia muy pequeños indica que esa variable puede ser explicada por una combinación lineal del resto de variables, lo cual significa que existe colinealidad).

Las proporciones de varianzas recogen la proporción de varianza de cada coeficiente de regresión parcial que está explicada por cada dimensión o factor.

Si se detecta colinealidad hay que aplicar algún remedio: aumentar el tamaño de la muestra, excluir variables redundantes, entre otras.

2.4 Métodos para aplicar RLM

Método *Forward selection*

Se inicia el proceso sin ninguna variable en la ecuación. La primera variable considerada para entrar en la ecuación será la que presente la correlación más alta con la variable a explicar. La segunda variable a introducir en la ecuación será la que, de las restantes, presente la correlación más alta con la variable a explicar. El proceso queda paralizado cuando ninguna de las variables que aún no han sido seleccionadas satisfaga el criterio de selección.

Generalmente suele considerarse como criterio de selección de una variable explicativa, el hecho de que la estimación del parámetro β a dicha variable asociada sea distinta de cero con un nivel de significación $\alpha < .05$

En la primera etapa entra en la ecuación la variable explicativa que más influye en la variación de la variable a explicar, es decir, la variable con un $N\beta_p$ más alto.

Para las variables que todavía no han sido seleccionadas, obtenemos la siguiente información:

- Beta entra. Estimación de los parámetros $N\beta_p$ si las variables explicativas asociadas se hubiesen incluido en la ecuación.
- *Partial Correlation*. Para una variable explicativa X_p es el coeficiente de correlación parcial entre Y y X_p al eliminar los efectos de las otras variables explicativas.
- *Minimun Tolerance*. Puede considerarse como un buen indicador de la existencia del multicolinalidad entre variables explicativas (algunas variables explicativas son combinaciones lineales de otras variables explicativas). Un valor cercano a 1 para X_p significa que X_p no es una función lineal de las variables explicativas contenidas en la ecuación (no existe colinealidad).

El proceso finaliza cuando ninguna de las variables aun por incluir en la ecuación cumple el criterio de selección.

Método *Backward elimination*

El proceso inicia con todas las variables explicativas en la ecuación. En cada etapa se van eliminando sucesivamente las variables que satisfacen el criterio de eliminación del modelo. El proceso finaliza cuando ninguna de las variables de la ecuación satisface dicho criterio.

El criterio de eliminación de una variable explicativa es en general menos restrictivo que el criterio de selección. Una variable quedará fuera de la ecuación si la estimación del parámetro β asociado es igual a cero con un nivel de significación $\alpha < 0.10$.

Método *Stepwise*

El proceso se inicia sin ninguna variable en la ecuación de regresión. En cada etapa se introduce o se elimina una variable. El proceso finaliza cuando no queda ninguna variable fuera de la ecuación que satisfaga el criterio de selección, ni tampoco queda ninguna variable en la ecuación que satisfaga el criterio de eliminación. Al combinar los dos métodos anteriores, es probablemente el más utilizado.

Aplicaremos este método en el análisis de regresión por pasos.

2.5 Análisis de regresión por pasos. Regresión STEP WISE-pasos sucesivos

Hay situaciones en las que se tiene un elevado número de posibles X_i , y no existe teoría o un trabajo previo que oriente al analista en la elección de las variables relevantes.

El control sobre las X_i que ha de formar parte de la ecuación de regresión pasa de las manos o del criterio del analista a una regla de decisión basada en criterios estadísticos.

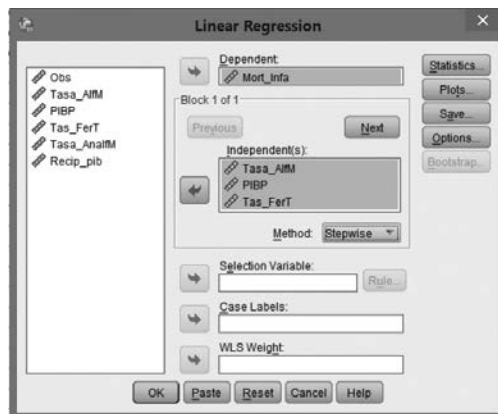
Hasta el momento hemos analizado criterios de selección como el r^2 , coeficientes de regresión parcial entre cada variable independiente y la variable dependiente. Todo lo visto hasta ahora, ayuda a maximizar el ajuste del modelo de regresión utilizando el mínimo número de posibles variables.

Los métodos por pasos basan la selección de variables en el criterio estadístico referido a la significación. Este criterio de significación implica que solo se incorporan al modelo aquellas variables que contribuyen al ajuste de forma significativa (se establece contrastando a partir del R, la hipótesis de independencia entre esa variable y la variable dependiente). Así, para decidir si se rechaza/acepta la hipótesis de independencia, se utiliza la prueba F y la prueba t.

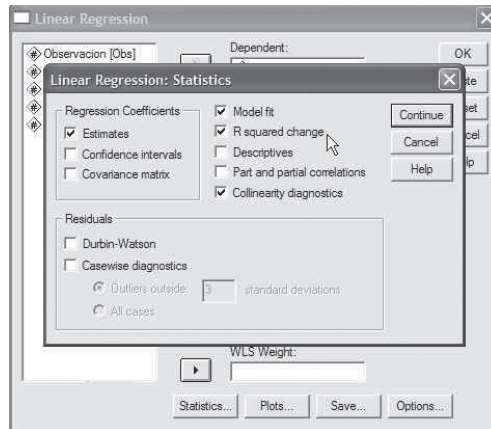
Una variable se incluye en el modelo, si el nivel crítico asociado a su R al contrastar la hipótesis de independencia es menor que 0.05 y queda fuera del modelo de regresión si el nivel crítico es mayor que 0.10.

Una forma intuitiva de comprender y valorar el efecto resultante de aplicar estos criterios de selección, consiste en observar el cambio que se va produciendo en r^2 a medida que se van incorporando (o eliminando) variables en el modelo de regresión.

Para aplicar regresión por pasos se selecciona el método de la siguiente forma:



Luego, se utiliza el botón <Statistic> ---/R squared Change.



Utilicemos el ejemplo de Mortalidad Infantil y agreguemos una variable independiente: la tasa de analfabetismo

El cuadro siguiente muestra las variables introducidas al modelo e indica los niveles de significación del estadístico F.

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Tasa de analfabetismo de Mujeres en porcentaje		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Tasa de fertilidad total		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	PIB percapita		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Mortalidad Infantil

El R^2 se observa que ha aumentado a medida que se han incluido las variables en el modelo.¹²

12 La tabla resumen del modelo muestra estadísticos del cambio experimentado por R^2 , el estadístico F y su significación. Este estadístico permite contrastar la hipótesis de que el cambio en R^2 vale cero en la población. Si la probabilidad asociada es menor que .05 se podrá afirmar que la proporción de varianza explicada por la variable introducida (Xi) en un paso específico, es significativamente distinta de cero.

Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.811 ^a	.658	.653	44.78822	.658	119.297	1	62	.000	2.143
2	.838 ^b	.703	.693	42.09602	.045	9.184	1	61	.004	
3	.860 ^c	.740	.727	39.72990	.037	8.482	1	60	.005	

a. Predictors: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje

b. Predictors: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje, Tasa de fertilidad total

c. Predictors: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje, Tasa de fertilidad total, PIB *per cápita*

d. Dependent Variable: Mortalidad Infantil

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	239306.958	1	239306.958	119.297	.000 ^b
	Residual	124371.042	62	2005.985		
	Total	363678.000	63			
2	Regression	255581.449	2	127790.724	72.114	.000 ^c
	Residual	108096.551	61	1772.075		
	Total	363678.000	63			
3	Regression	268970.085	3	89656.695	56.800	.000 ^d
	Residual	94707.915	60	1578.465		
	Total	363678.000	63			

a. Dependent Variable: Mortalidad Infantil

b. Predictors: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje

c. Predictors: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje, Tasa de fertilidad total

d. Predictors: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje, Tasa de fertilidad total, PIB *per cápita*

La tabla ANOVA muestra el valor del estadístico F obtenido al contrastar la hipótesis de que el valor poblacional de R^2 en cada paso es cero (se evalúa el valor de R^2 en cada paso).

La tabla de coeficientes de regresión parcial presenta los coeficientes correspondientes a cada uno de los X_i incluidos en el modelo de regresión múltiple. Estos coeficientes sirven para construir la ecuación de regresión en cada paso.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	26.027	11.963		2.176	.033		
	Tasa de analfabetismo de Mujeres-en porcentaje	2.358	.216	.811	10.922	.000	1.000	1.000
2	(Constant)	-25.541	20.396		-1.252	.215		
	Tasa de analfabetismo de Mujeres-en porcentaje	1.867	.260	.642	7.193	.000	.611	1.637
	Tasa de fertilidad total	13.623	4.495	.271	3.030	.004	.611	1.637
3	(Constant)	-9.005	20.069		-.449	.655		
	Tasa de analfabetismo de Mujeres-en porcentaje	1.722	.250	.593	6.889	.000	.587	1.704
	Tasa de fertilidad total	13.327	4.244	.265	3.140	.003	.611	1.638
	PIB <i>per cápita</i>	-.006	.002	-.199	-2.912	.005	.927	1.078

a. Dependent Variable: Mortalidad Infantil

La tabla presenta los coeficientes de regresión parcial estandarizados (Beta), los cuales informan sobre la importancia relativa de cada Xi dentro de la ecuación. Las dos últimas columnas contrastan la hipótesis de que los coeficientes de regresión parcial valen cero en la población. Si el valor crítico es menor de .05 implica que la variable contribuye significativamente a mejorar el ajuste del modelo.

La tabla de variables excluidas muestra los coeficientes de regresión parcial de las variables no seleccionadas para formar parte de la ecuación de regresión en cada paso.

En la tabla de variables excluidas se seleccionan las variables, aquellas que tienen valores de correlación R más altos (en valor absoluto) con la variable dependiente.

Excluded Variables^a

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Tasa de analfabetismo de Mujeres-en porcentaje	. ^b	.	.	.	.000	.	.000
	PIB per cápita	-.204 ^b	-2.792	.007	-.337	.928	1.078	.928
	Tasa de fertilidad total	.271 ^b	3.030	.004	.362	.611	1.637	.611
2	Tasa de analfabetismo de Mujeres-en porcentaje	. ^c	.	.	.	.000	.	.000
	PIB per cápita	-.199 ^c	-2.912	.005	-.352	.927	1.078	.587
3	Tasa de analfabetismo de Mujeres-en porcentaje	. ^d	.	.	.	.000	.	.000

a. Dependent Variable: Mortalidad Infantil

b. Predictors in the Model: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje

c. Predictors in the Model: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje, Tasa de fertilidad total

d. Predictors in the Model: (Constant), Tasa de analfabetismo de Mujeres-en porcentaje, Tasa de fertilidad total, PIB per cápita

¿Qué variables se deben incluir en la ecuación de regresión? El método de selección por pasos ha llevado a construir una ecuación de regresión con tres variables (porque poseen coeficientes de regresión significativos). Esta ecuación con tres variables es la mejor de las posibles con el mínimo número de variables

2.6. Ejercicios

Ejercicio 1

Aplicar los tres métodos de RLM al anexo DB_Empleados.sav

Ejercicio 2

Utilice los anexos Deserción_d_extorsión y Deserción_D_homicidios para aplicar el RLM.

3. Análisis de componentes principales (ACP)

Introducción

Un problema central en el análisis de datos multivariantes es la reducción de la dimensionalidad. Si es posible describir con precisión los valores de p variables por un pequeño subconjunto $r < p$ de ellas, se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información.

El análisis de componentes principales tiene ese objetivo: Con n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales. Por ejemplo, con variables con alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20% de las originales) expliquen la mayor parte (más del 80%) de la variabilidad original.

La técnica de componentes principales es debida a Hotelling (1933), aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson (1901). Su utilidad es doble:

1. Permite representar óptimamente en un espacio de dimensión pequeña, observaciones de un espacio general p -dimensional. En este sentido la técnica de componentes principales es el primer paso para identificar posibles variables “latentes” o no observables, que están generando la variabilidad de los datos.
2. Permite transformar las variables originales, en general correladas, en nuevas variables incorreladas, facilitando la interpretación de los datos.

En este apartado presentamos únicamente esta técnica como una herramienta exploratoria para facilitar la descripción y la interpretación de los datos. El problema de inferir si las propiedades de reducción de la dimensión encontradas en los datos pueden extenderse a una población, se estudiará en el capítulo de análisis factorial.

3.1. Planteamiento del problema

Supongamos que se dispone de los valores de p -variables en n elementos de una población, dispuestos en una matriz X de dimensiones $n \times p$, donde las columnas contienen las variables, y las filas los elementos. Supongamos que previamente hemos restado a cada variable su media, de manera que las variables de la matriz X tienen media cero y su matriz de covarianzas vendrá dada por $(1/n) X'X$.

El problema que se desea resolver es cómo encontrar un espacio de dimensión más reducida que represente adecuadamente los datos. El problema puede abordarse desde tres perspectivas equivalentes.

3.2. Enfoque descriptivo

Se desea encontrar un subespacio de dimensión menor que p , tal que al proyectar sobre él los puntos conserven su estructura con la menor distorsión posible. Veamos cómo convertir esta noción intuitiva en un criterio matemático operativo. Consideremos primero un subespacio de dimensión uno, una recta. Se desea que las proyecciones de los puntos sobre esta recta mantengan, lo más posible, sus posiciones relativas. Para concretar, consideremos el caso de dos dimensiones ($p=2$). La figura 2.1. indica el diagrama de dispersión y una recta que, intuitivamente, proporciona un buen resumen de los datos, ya que las proyecciones de los puntos sobre ella indican aproximadamente la situación de los puntos en el plano. La representación es buena porque la recta pasa cerca de todos los puntos y estos se deforman poco al proyectarlos. Esta propiedad puede concretarse exigiendo que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo más pequeñas posibles. En consecuencia, si consideramos un punto x_i y una dirección $a_1 = (a_{11}, \dots, a_{1p})'$ definida por un vector a_1 de norma igual a 1, la proyección del punto x_i sobre esta dirección es el escalar $z_i = (a_{11} x_{i1} + \dots + a_{1p} x_{ip}) = a_1' x_i$

Y el vector que representa esta proyección es la distancia entre el punto, y su proyección sobre la dirección, este criterio implica:

$$\text{minimizar } \sum_{i=1}^n r_i^2 = \sum_{i=1}^n |x_i - z_i a_1|^2$$

Donde es la norma euclídea o módulo del vector u .

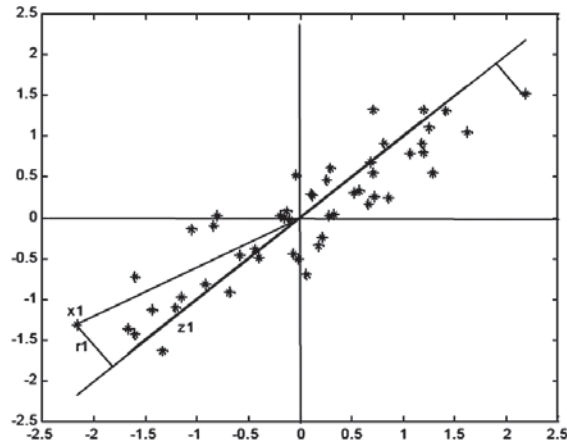


Figura 2.1. Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a ella.

Dicha figura muestra que al proyectar cada punto sobre la recta se forma un triángulo rectángulo, donde la hipotenusa es la distancia al origen del punto $(x'_i, x_i)^{1/2}$, y los catetos son la proyección del punto sobre la recta (z_i) y la distancia entre el punto y su proyección (r_i). Por el teorema de Pitágoras, podemos escribir:

$$x'_i x_i = z_i^2 + r_i^2$$

Y sumando esta expresión para todos los puntos, se obtiene:

$$\sum_{i=1}^n x'_i x_i = \sum_{i=1}^n z_i^2 + \sum_{i=1}^n r_i^2$$

Como el primer miembro es constante, minimizar $\sum_{i=1}^n r_i^2$, la suma de las distancias a la recta de todos los puntos es equivalente a maximizar $\sum_{i=1}^n z_i^2$, que es la suma al cuadrado de los valores de las proyecciones. Como las proyecciones z_i son variables de media cero, maximizar la suma de sus cuadrados equivale a maximizar su varianza. Este resultado es intuitivo: la recta de la figura 2.1. parece adecuada porque conserva lo más posible la variabilidad original de los puntos.

El objetivo de proyectar los puntos con mínima deformación puede abordarse desde otro punto de vista que conduce al mismo resultado final. El espacio de p -dimensiones, es lo característico de la nube de puntos en sus distancias relativas. Tratemos de encontrar un subespacio de dimensión uno, es decir, una recta tal que los puntos proyectados conserven lo más posible sus distancias relativas. Si

llamamos $d_{ij}^2 = x_i' x_j$ a los cuadrados de las distancias originales entre los puntos y $\hat{d}_{ij}^2 = (z_i - z_j)^2$ a las distancias entre los puntos proyectados sobre una recta, deseamos que $D = \sum_i \sum_j (d_{ij}^2 - \hat{d}_{ij}^2)$ sea mínima. Como la suma de las distancias originales es fija, minimizar D requiere maximizar $\sum_i \sum_j \hat{d}_{ij}^2$, las distancias entre los puntos proyectados.

3.3. Enfoque estadístico

Representar puntos p dimensionales con la mínima pérdida de información en un espacio de dimensión uno es equivalente a sustituir las p variables originales por una nueva variable, z_p que resuma óptimamente la información. Esto supone que la nueva variable debe tener globalmente máxima correlación con las originales o en otros términos, debe permitir prever las variables originales con la máxima precisión. Esto no será posible si la nueva variable toma un valor semejante en todos los elementos.

En la figura 2.1. que se analizaba, la recta indicada no es la línea de regresión de ninguna de las variables con respecto a la otra, que se obtienen minimizando las distancias verticales u horizontales, sino que al minimizar las distancias ortogonales o de proyección se encuentra entre ambas rectas de regresión.

Este enfoque puede extenderse para obtener el mejor subespacio resumen de los datos de dimensión dos. Para ello calcularemos el plano que mejor se aproxima a los puntos. El problema se reduce a encontrar una nueva dirección definida por un vector unitario, a_2 , que, sin pérdida de generalidad, puede tomarse ortogonal a a_1 , y que verifique la condición de que la proyección de un punto sobre este eje maximice las distancias entre los puntos proyectados. Estadísticamente esto equivale a encontrar una segunda variable z_2 , incorrelada con la anterior, y que tenga varianza máxima. En general, la componente z_r ($r < p$) tendrá varianza máxima entre todas las combinaciones lineales de las p variables X originales, con la condición de estar incorrelada con las $z_1, \dots, z_{r-1}$ previamente obtenidas.

3.4. Enfoque geométrico

El problema puede abordarse desde un punto de vista geométrico con el mismo resultado final. Si consideramos la nube de puntos de la figura 2.1, vemos que los puntos se sitúan siguiendo una elipse y podemos describir su orientación dando la dirección del eje mayor de la elipse y la posición de los puntos por su proyección sobre esta dirección. Puede demostrarse que este eje es la recta que minimiza las distancias ortogonales. En varias dimensiones tendremos elipsoides y la mejor

aproximación a los datos es la proporcionada por el eje mayor del elipsoide. Considerar los ejes del elipsoide como nuevas variables originales supone pasar de variables correladas a variables ortogonales, como se verá a continuación.

3.5. Cálculo del primero y segundo componente

Componente 1

El primer componente principal será la combinación lineal de las variables originales que tenga varianza máxima. Los valores de este primer componente en los n individuos se representarán por un vector z_1 dado por:

$$z_1 = Xa_1$$

Como las variables originales tienen media cero también z_1 tendrá media nula. Su varianza será:

$$Var(z_1) = \frac{1}{n} z_1' z_1 = \frac{1}{n} a_1' X' X a_1 = a_1' S a_1$$

Donde S es la matriz de varianzas y covarianzas de las observaciones. Es obvio que podemos maximizar la varianza sin límite aumentando el módulo del vector a_1 . Para que la maximización de la expresión anterior tenga una solución debemos imponer una restricción al módulo del vector a_1 , y sin pérdida de generalidad, impondremos que $a_1' a_1 = 1$. Introduciremos esta restricción mediante el multiplicador de Lagrange:

$$M = a_1' S a_1 - \lambda(a_1' a_1 - 1)$$

Y maximizaremos esta expresión de la forma habitual derivando respecto a los componentes de a_1 e igualando a cero. Entonces:

$$\frac{\partial M}{\partial a_1} = 2S a_1 - 2\lambda a_1 = 0$$

Cuya solución es:

$$S a_1 = \lambda a_1$$

Que implica que es un vector propio de la matriz S , y su correspondiente *valor propio*. Para determinar qué valor propio de S es la solución de la ecuación anterior, tendremos en cuenta que, multiplicando esta ecuación por a_1' :

$$a_1' S a_1 = \lambda a_1' a_1 = \lambda$$

Y concluimos en que es la varianza. Como ésta es la cantidad que queremos maximizar, será el mayor valor propio de la matriz S . Su vector asociado, define los coeficientes de cada variable en el primer componente principal.

Por ejemplo, si retomamos el ejemplo de acciones.sav (los programas estadísticos generan automáticamente los valores de los componentes principales), sabemos que la matriz de covarianzas viene dada por:

$$S = \begin{bmatrix} 0.35 & 0.15 & -0.19 \\ 0.15 & 0.13 & -0.03 \\ -0.19 & -0.03 & 0.16 \end{bmatrix}$$

Para el cálculo de los autovalores tenemos que calcular las raíces de la ecuación:

$$\begin{aligned} 0 &= |S - \lambda I| = \\ &= \left| \begin{bmatrix} 0.35 & 0.15 & -0.19 \\ 0.15 & 0.13 & -0.03 \\ -0.19 & -0.03 & 0.16 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \right| = \\ &= -0,000382 - 0,0628\lambda + 0,64\lambda^2 - \lambda^3 \end{aligned}$$

Al utilizar un software estadístico podemos encontrar las raíces del polinomio anterior, $\lambda_1 = 0.521$, $\lambda_2 = 0.113$, $\lambda_3 = 6.51 \times 10^{-3}$

El autovector asociado λ_1 nos da los pesos de la primera componente principal. Para calcular el primer autovector debemos resolver el sistema:

$$S a_1 = \lambda_1 a_1$$

Que conduce a:

$$\begin{aligned} \begin{bmatrix} 0.35 & 0.15 & -0.19 \\ 0.15 & 0.13 & -0.03 \\ 0.19 & -0.03 & 0.16 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} &= 0.521 \times \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} \\ -0.17a_{11} + 0.15a_{12} - 0.19a_{13} &= 0 \\ 0.15a_{11} - 0.391a_{12} - 0.03a_{13} &= 0 \\ -0.19a_{11} - 0.03a_{12} - 0.361a_{13} &= 0 \end{aligned}$$

El sistema anterior es compatible indeterminado, y para encontrar una de las infinitas soluciones se debe considerar la primera variable como parámetro, x , resolviendo el sistema en función de la x . La solución es:

$$\{a_{11} = x, a_{12} = 0.427x, a_{13} = -0.562x\}$$

El valor de x se obtiene imponiendo que el vector tenga norma unidad, con lo que resulta:

$$a_1 \begin{bmatrix} -0.817 \\ -0.349 \\ 0.459 \end{bmatrix}$$

Y el primer componente es:

$$Z_1 = -0.817X_1 - 0.349X_2 + 0.459X_3$$

Donde X_1 , X_2 y X_3 son las variables en logaritmos.

Componente 2

Vamos a obtener el mejor plano de proyección de las variables X_i . Lo calcularemos estableciendo como función objetivo que la suma de las varianzas de $z_1 = Xa_1$ y $z_2 = Xa_2$ sea máxima, donde a_1 y a_2 son los vectores que definen el plano. La función objetivo será esta:

$$\phi = a_1' S a_1 + a_2' S a_2 - \lambda_1 (a_1' a_1 - 1) - \lambda_2 (a_2' a_2 - 1) \text{ (Ecuación 3.5)}$$

Que incorpora las restricciones de que las direcciones deben tener módulo unitario ($a_i' a_i = 1$, $i = 1, 2$). Derivando e igualando a cero, así:

$$\frac{\partial \phi}{\partial a_1} = 2S a_1 - 2\lambda_1 a_1 = 0$$

$$\frac{\partial \phi}{\partial a_2} = 2S a_2 - 2\lambda_2 a_2 = 0$$

La solución de este sistema es:

$$S a_1 = \lambda_1 a_1$$

$$S a_2 = \lambda_2 a_2$$

Que indica que a_1 y a_2 y deben ser vectores propios de S . Tomando los vectores propios de norma uno y sustituyéndolo en la Ecuación 3.5, se obtiene que, en el máximo, la función objetivo es:

$$\phi = \lambda_1 + \lambda_2$$

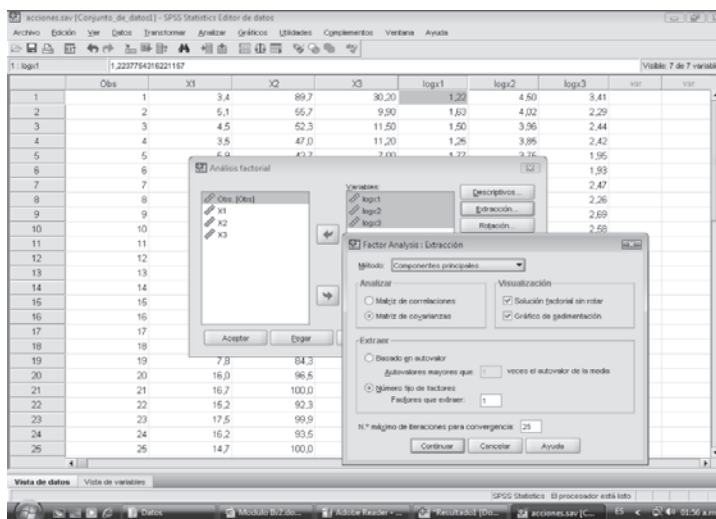
Es claro que λ_1 y λ_2 deben ser los dos autovalores mayores de la matriz S , y a_1 y a_2 sus correspondientes autovectores. Observemos que la covarianza entre z_1 y z_2 , dada por $a_1' S a_2$, es cero, ya que $a_1' a_2 = 0$, y las variables z_1 y z_2 estarán incorreladas.

Ejemplo 1

Calculemos la primera componente principal utilizando el software SPSS con los datos de los logaritmos de las acciones (anexo acciones.sav). Este paquete proporciona directamente los componentes principales.

Pasos para realizar el ejemplo en SPSS:

- En el menú Analizar-seleccionamos Reducción de la dimensión-Factor.
- Seleccionar las variables logx1, logx2 y logx3.
- Dentro del subcuadro de diálogo Factor, seleccionar el botón Extracción y luego seleccionar las opciones que se muestran en la figura.



- Seleccionamos el método de componentes principales; luego, el número fijo de factores a extraer, en este caso, 1 después en el botón Puntajes seleccionar matriz de coeficientes y puntajes componentes, y así obtenemos el vector de componentes:

Component Score Coefficient Matrix^a

	Component
	1
Logx1	.678
Logx2	.176
Logx3	-.248

Extraction Method: Principal. Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Component Scores.

a. Coefficients are standardized

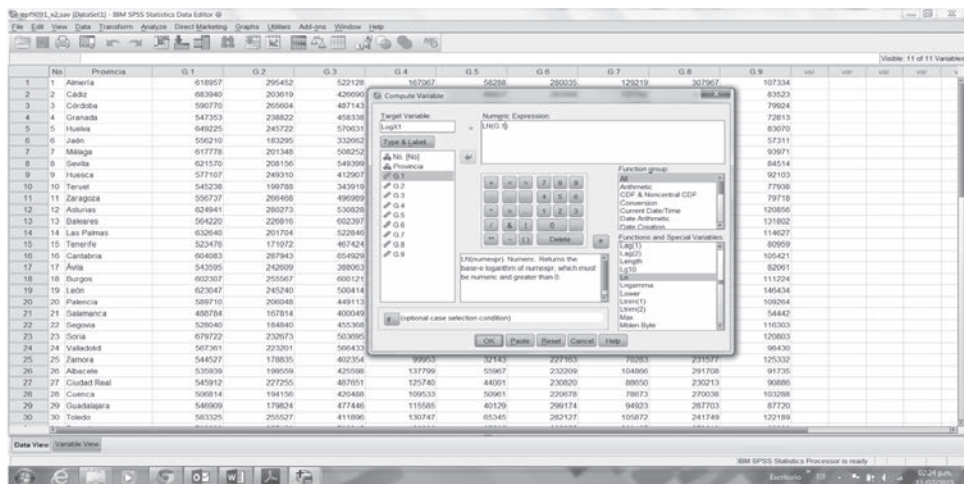
Ejemplo 2

La encuesta de presupuestos familiares en España (anexo epf9091_v2.sav) presenta los gastos medios de las familias españolas en nueve variables: X1=alimentación, X2=vestido y calzado, X3=vivienda, X4=mobiliario doméstico, X5=gastos sanitarios, X6=transportes, X7=enseñanza y cultura, X8=turismo y ocio, X9=otros gastos, para las 51 provincias españolas. La matriz de covarianzas resume la variabilidad de estas variables en los elementos observados.

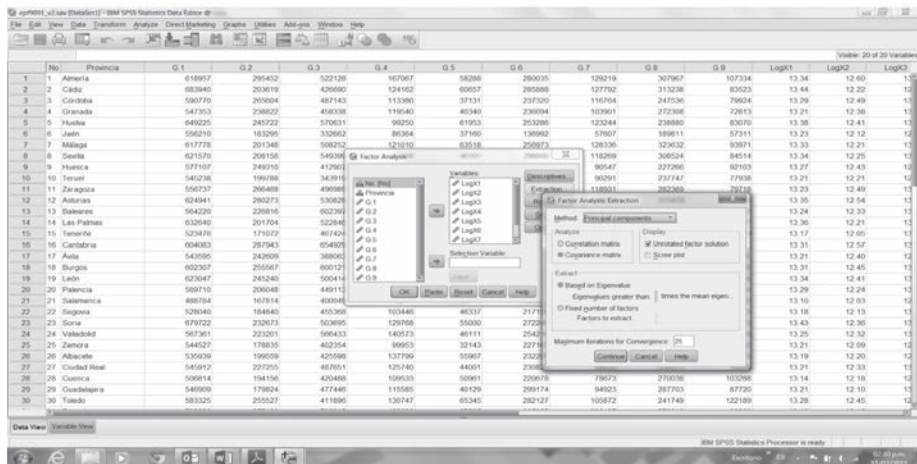
Esta base de datos contiene 51 observaciones y nueve variables, y corresponde a información de gastos de las familias españolas, que es una encuesta de presupuestos familiares.

Pasos:

- Hacer un análisis gráfico de las variables y observar si hay simetría o asimetría en los datos.
- Como las distribuciones de los gastos son muy asimétricas, las variables deben expresarse en logaritmos. Hay que realizar en SPSS una transformación para crear dichas variables de la siguiente manera, vista en la imagen siguiente:



- Luego, en el menú Analizar-Reducción de Dimensión, agregamos la lista de variables y seleccionamos la matriz de varianza o covarianza, así como se muestra en la siguiente imagen:



- Y observamos los resultados siguientes: la matriz de covarianza total explicada.

Total Variance Explained							
Component		Initial Eigenvalues ^a			Extraction Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Raw	1	.348	74.124	74.124	.348	74.124	74.124
	2	.032	6.830	80.954			
	3	.027	5.743	86.697			
	4	.018	3.730	90.427			
	5	.013	2.684	93.112			
	6	.011	2.288	95.399			
	7	.010	2.216	97.615			
	8	.006	1.265	98.880			
	9	.005	1.120	100.000			
Rescaled	1	.348	74.124	74.124	6.124	68.042	68.042
	2	.032	6.830	80.954			
	3	.027	5.743	86.697			
	4	.018	3.730	90.427			
	5	.013	2.684	93.112			
	6	.011	2.288	95.399			
	7	.010	2.216	97.615			
	8	.006	1.265	98.880			
	9	.005	1.120	100.000			

Extraction Method: Principal Component Analysis.

a. When analyzing a covariance matrix, the initial eigenvalues are the same across the raw and rescaled solution.

- El vector propio asociado al mayor valor propio, 0.348, define la siguiente componente:

	Raw
	Component
	1
LogX1	.068
LogX2	.108
LogX3	.176
LogX4	.181
LogX5	.271
LogX6	.201
logX7	.294
LogX8	.181
LogX9	.185

El cual se puede expresar en notación matemática de la siguiente manera:

$$z_1 = 0.07x_1 + 0.11x_2 + 0.18x_3 + 0.18x_4 + 0.27x_5 + 0.20x_6 + 0.29x_7 + 0.18x_8 + 0.19x_9$$

Se observa que es una suma ponderada de todos los gastos con mayor peso en los gastos en enseñanza y cultura (X7) y gastos sanitarios (X5). El menor peso lo tiene en gastos de alimentación (X1).

Si calculamos las coordenadas para las provincias españolas y las ordenamos por esta nueva variable, las provincias quedan prácticamente ordenadas por su renta. La primera componente principal tiene en este caso una explicación inmediata: redescubre la renta de cada provincia.

El segundo componente principal para las variables de gastos de la referida encuesta es el asociado al segundo valor propio mayor, que es 0.032. El vector propio asociado a este valor propio define la nueva variable.

En SPSS entrar nuevamente al cuadro de diálogo de Componentes Principales/ Botón Extracción (ingresar y número dos indicando el número de componentes a extraer), obteniendo el resultado siguiente:

Component Matrix^a

	Raw		Rescaled	
	Component		Component	
	1	2	1	2
LogX1	.068	.010	.611	.088
LogX2	.108	.029	.664	.179
LogX3	.176	-.031	.858	-.150
LogX4	.181	.013	.886	.062
LogX5	.271	-.038	.882	-.124
LogX6	.201	.052	.872	.224
LogX7	.294	-.072	.937	-.230
LogX8	.181	-.031	.882	-.149
LogX9	.185	.141	.771	.585

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Si lo expresamos en forma matemática, se tiene que:

$$z_2 = 0.01x_1 + 0.03x_2 - 0.03x_3 + 0.013x_4 - 0.04x_5 + 0.05x_6 - 0.07x_7 - 0.03x_8 + 0.14x_9 =$$

$$z_2 = (0.01x_1 + 0.03x_2 + 0.013x_4 + 0.05x_6 + 0.14x_9 - (0.03x_3 + 0.04x_5 + 0.07x_7 + 0.03x_8))$$

Esta variable puede verse como la diferencia entre dos medias ponderadas de los gastos. La primera asigna un peso a *otros gastos* (X9) y *transporte* (X6). En la variable *otros gastos* están incluidas las transferencias fuera de la provincia a miembros de la familia mayores de 14 años que no residan en ella; podemos conjeturar que esta variable separa las provincias que reciben transferencias de las que las envían. Es también significativo que estas provincias tienen altos gastos en *transporte* (x6). La primera media ponderada puede considerarse un indicador de cómo esta provincia envía recursos a otras. La segunda media ponderada asigna mayor peso a las variables *enseñanza y cultura* (X7) y *gastos sanitarios* (X5).

3.6. Propiedades de los componentes

Los componentes principales como nuevas variables tienen las propiedades siguientes:

1. Conservan la variabilidad inicial: la suma de las varianzas de los componentes es igual a la suma de las varianzas de las variables originales, y la varianza generalizada de los componentes es igual a la original.
2. La proporción de variabilidad explicada por un componente es el cociente entre su varianza, el valor propio asociado al vector propio que lo define y la suma de los valores propios de la matriz.

En efecto, como la varianza del componente h es λ_h , el valor propio que define el componente y la suma de todas las varianzas de las variables originales es $\sum_{i=1}^p \lambda_i$, igual a la suma de las varianzas de los componentes. La proporción de variabilidad total explicada por el componente h es $\lambda_h / \sum \lambda_i$.

3. Las covarianzas entre cada componente principal y las variables X vienen dadas por el producto de las coordenadas del vector propio que define el componente por el valor propio:

$$\text{Cov}(z_i; x_1, \dots, x_p) = \lambda_i a_i = (\lambda_i a_{i1}, \dots, \lambda_i a_{ip})$$

Donde a_i es el vector de coeficientes de la componente z_i .

4. Las correlaciones entre un componente principal y una variable X es proporcional al coeficiente de esa variable en la definición del componente, y el coeficiente de proporcionalidad es el cociente entre la desviación típica del componente y la desviación típica de la variable.
5. Las r componentes principales ($r < p$) proporcionan la predicción lineal óptima con r variables del conjunto de variables X .

Esta afirmación puede expresarse de dos formas. La primera, demostrando que la mejor predicción lineal con r variables de las variables originales se obtiene utilizando las r primeras componentes principales. La segunda, demostrando que la mejor aproximación de la matriz de datos que puede construirse con una matriz de rango r se obtiene construyendo esta matriz con los valores de los r primeros componentes principales.

6. Si estandarizamos los componentes principales, dividiendo cada uno por su desviación típica, se obtiene la estandarización multivariante de los datos originales. Por tanto, la estandarización multivariante puede interpretarse como:

(1) Obtener los componentes principales.

(2) Estandarizarlos para que tengan todos la misma varianza.

Esta relación se presenta en la figura que sigue. La transformación mediante componentes principales conduce a variables incorreladas, pero con distinta varianza, puede interpretarse como rotar los ejes de la elipse que definen los puntos para que coincidan con sus ejes naturales. La estandarización multivariante produce variables incorreladas con varianza unidad, lo que supone buscar los ejes naturales y luego estandarizarlos. En consecuencia, si estandarizamos los componentes se obtienen las variables estandarizadas de forma multivariante.

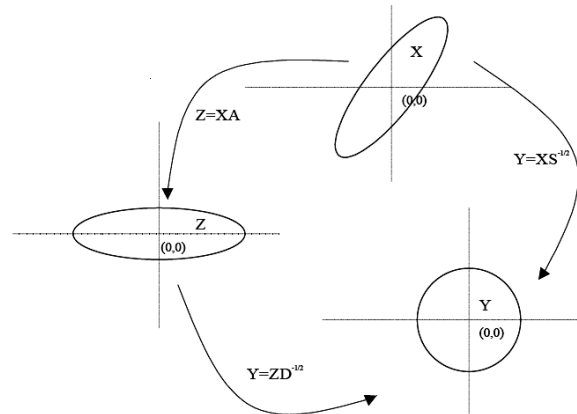


Figura 2.2. Representación gráfica de la relación entre componentes principales y estandarización multivariante

3.7. Análisis normado o con correlaciones

Los componentes principales normados se obtienen calculando los vectores y valores propios de la **Matriz R**, de coeficientes de correlación. Llamando λ_p^R a las raíces características de esa matriz, que suponemos no singular, se verifica que:

$$\sum_{i=1}^p \lambda_i^R = \text{traza}(R) = p$$

Las propiedades de los componentes extraídos de R son estos:

$$\frac{\lambda_p^R}{p}$$

1. La proporción de variación explicada por λ_p^R será la siguiente:
2. Las correlaciones entre cada componente z_j y las variables X_i originales vienen dadas directamente por $a'_i \sqrt{\lambda_j}$ siendo $z_j = Xa_j$.

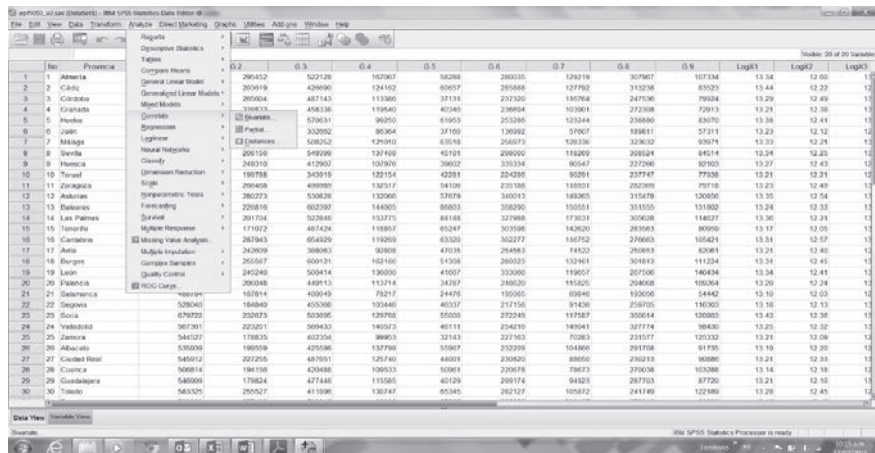
Cuando las variables X originales están en distintas unidades conviene aplicar el análisis de la matriz de correlaciones o análisis normado. Cuando las variables tienen las mismas unidades, ambas alternativas son posibles. Si las diferencias entre las varianzas de las variables son informativas y queremos tenerlas en cuenta en el análisis, no debemos estandarizar las variables: por ejemplo, supongamos dos índices con la misma base, pero uno fluctúa mucho y el otro es casi constante. Este hecho es informativo, y para tenerlo en cuenta en el análisis no se deben

estandarizar las variables, de manera que el índice de mayor variabilidad tenga más peso. Por el contrario, si las diferencias de variabilidad no son relevantes podemos eliminarlas con el análisis normado. En caso de duda, conviene realizar ambos análisis y seleccionar aquel que conduzca a conclusiones más informativas.

Ejemplo 3

La matriz de correlación de los nueve indicadores del ejemplo dos se puede obtener a partir de SPSS con los siguientes pasos.

- Trabajar con el anexo epf9091_v2.sav y en el menú Analizar seleccionar Correlaciones bivariadas, tal como se muestra en la imagen siguiente:



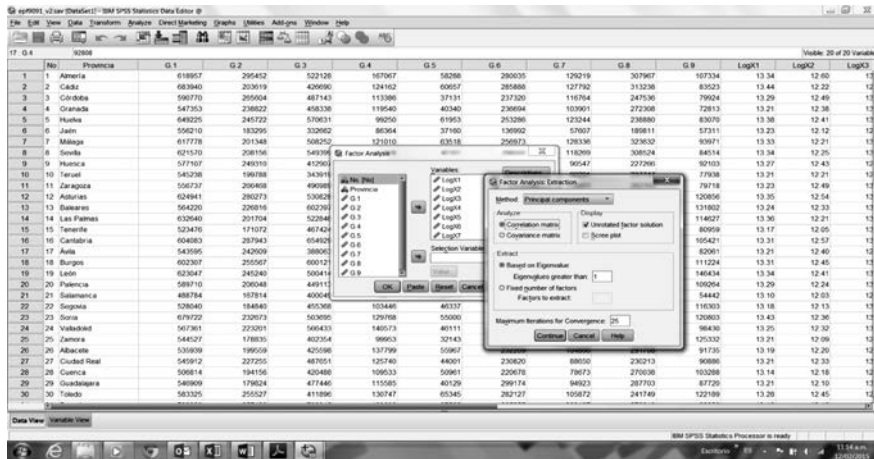
- Seleccionar las variables con sus logaritmos des de la logX1 hasta las LogX9, y el resultado es el siguiente:

Correlation Matrix

	LogX1	LogX2	LogX3	LogX4	LogX5	LogX6	LogX7	LogX8	LogX9
Correlation LogX1	1.000	.538	.503	.531	.458	.500	.614	.458	.508
LogX2	.538	1.000	.566	.631	.541	.577	.572	.465	.524
LogX3	.503	.566	1.000	.724	.668	.705	.832	.808	.586
LogX4	.531	.631	.724	1.000	.727	.758	.807	.783	.684
LogX5	.458	.541	.668	.727	1.000	.718	.779	.721	.612
LogX6	.500	.577	.705	.758	.718	1.000	.759	.727	.731
LogX7	.614	.572	.832	.807	.779	.759	1.000	.848	.607
LogX8	.458	.465	.808	.783	.721	.727	.848	1.000	.619
LogX9	.508	.524	.586	.684	.612	.731	.607	.619	1.000

Puede observarse alta correlación entre x4 y x3; x5 y x4; x6 y x3; x7, x8 y x3.

- Ahora realizamos el cálculo de los componentes principales utilizando como método la matriz de correlaciones y realizando la máxima varianza, tal como se observa en la siguiente imagen:



- Teniendo como resultado los vectores propios para los dos componentes:

Component Matrix^a

	Component	
	1	2
LogX1	.669	.580
LogX2	.714	.464
LogX3	.864	-.163
LogX4	.897	-.053
LogX5	.841	-.171
LogX6	.873	-.075
logX7	.921	-.114
LogX8	.874	-.321
LogX9	.785	.081

Extraction Method: Principal component
Análisis.

a. 2 components extracted

3.8. Interpretación de los componentes

Componentes de tamaño y forma

Cuando existe una alta correlación positiva entre todas las variables, el primer componente principal tiene todas sus coordenadas del mismo signo y puede interpretarse como un promedio ponderado de todas las variables. Se interpreta entonces como un factor global de “tamaño”. Los restantes componentes se interpretan como factores “de forma” y típicamente tienen coordenadas positivas y negativas, lo que implica que contraponen unos grupos de variables frente a otros. Estos factores de forma pueden frecuentemente escribirse como medias

ponderadas de dos grupos de variables con distinto signo y contraponen las variables de un signo a las del otro.

Selección del número de componentes

Se han surgido distintas reglas para seleccionar el número de componentes a mantener, así:

- a) Realizar un gráfico de λ_i frente a i . Comenzar seleccionando componentes hasta que los restantes tengan aproximadamente el mismo valor de λ_i . La idea es buscar un “codo” en el gráfico de sedimentación, es decir, un punto a partir del cual los valores propios son aproximadamente iguales. El criterio es quedarse con un número de componentes que excluya los asociados a valores pequeños y aproximadamente del mismo tamaño.
- b) Seleccionar componentes hasta cubrir una proporción determinada de varianza, como el 80% o el 90%. Esta regla es arbitraria y debe aplicarse con cierto cuidado. Por ejemplo, es posible que un único componente de “tamaño 2” recoja el 90% de la variabilidad, sin embargo, pueden existir otros componentes que sean muy adecuados para explicar la “forma” de las variables.
- c) Desechar aquellos componentes asociados a valores propios inferiores a una cota, que suele fijarse como la varianza media, $\Sigma \lambda_i / p$. En particular, cuando se trabaja con la matriz de correlación, el valor medio de los componentes es uno, y esta regla lleva a seleccionar los valores propios mayores que la unidad. De nuevo esta regla es arbitraria: una variable que sea independiente del resto suele llevar un componente principal y puede tener un valor propio mayor que la unidad. Sin embargo, si está incorrelada con el resto puede ser una variable poco relevante para el análisis y no aportar mucho a la comprensión del fenómeno global.

3.9. Aplicación en SPSS

Ejercicio 1

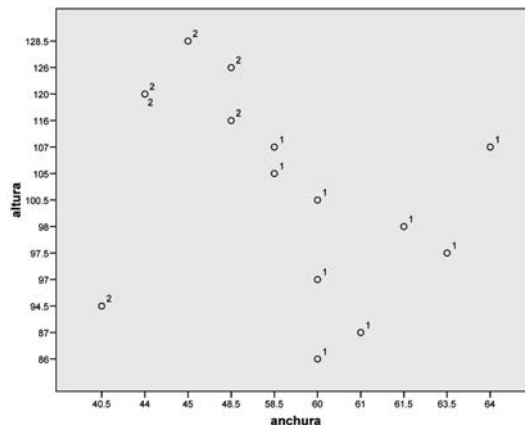
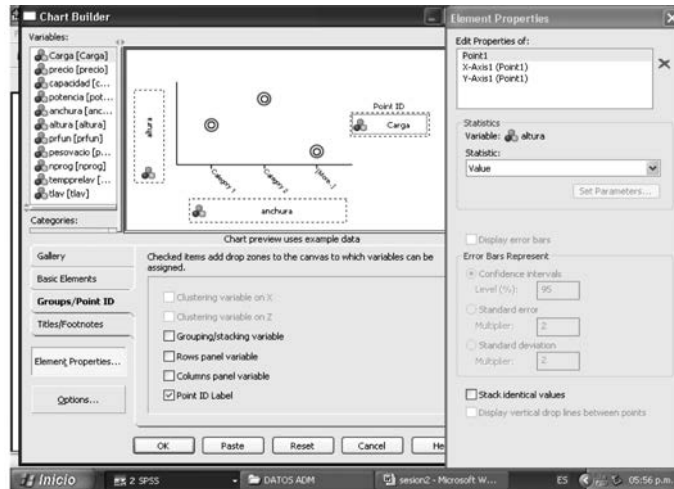
Realizaremos un análisis de componentes principales para el anexo lavadoras.sav, utilizando: a) método de la matriz de covarianzas; b) matriz de correlaciones. Además, considerar los criterios de autovalores, gráfico de sedimentación y máxima varianza para decidir con cuántos componentes se debe quedar en el análisis. También, ver si hay forma de diferenciar las lavadoras de carga superior con carga frontal.

Lavadoras.sav es una base de datos que consta de 15 observaciones y tiene las siguientes características de artículos de lavadoras, las variables son estas:

Precio, capacidad, potencia, anchura, altura, profundidad, peso vacío, programable, temperatura prelavado, tiempo lavado, solución en SPSS.

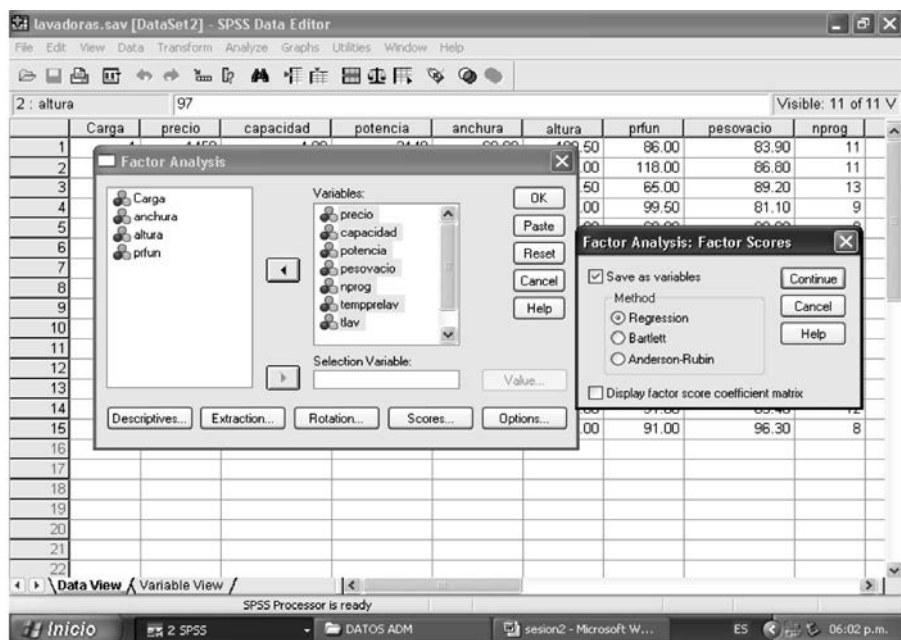
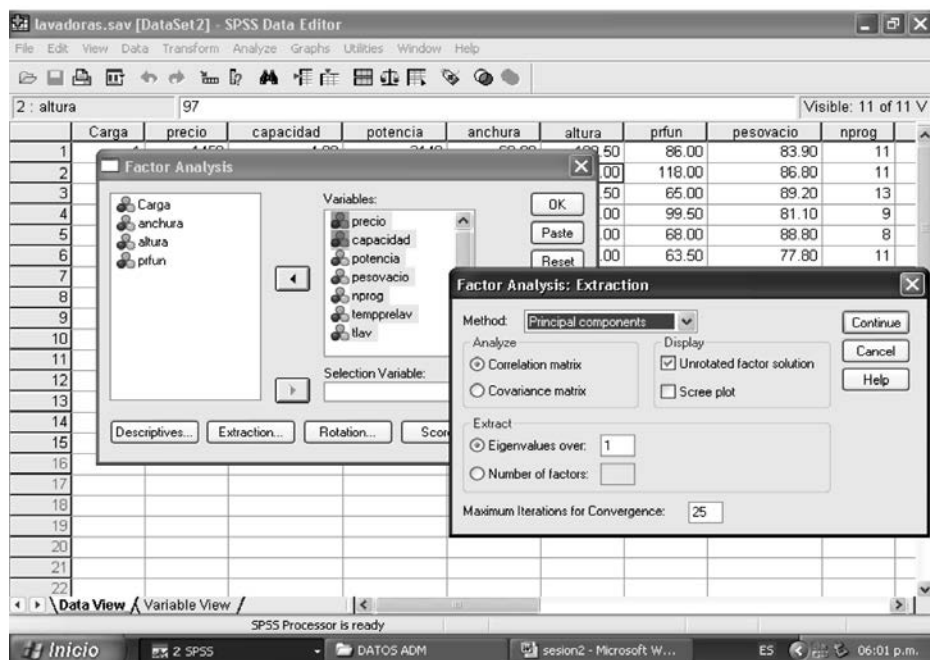
Pasos:

Abrir el anexo lavadoras.sav en SPSS; luego, realicemos un gráfico simple de dispersión (utilizando el generador de gráficos), siguiendo las sentencias tal como se muestra en la imagen siguiente:



Para diferenciar podemos trazar una línea recta y dividir los puntos en cargas superiores en cargas frontales.

- Ahora, realicemos el Análisis de Componentes Principales, utilizando la matriz de correlaciones (b). Veamos la imagen siguiente y sigamos las sentencias en SPSS.



- Teniendo como resultado lo siguiente:

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.155	30.780	30.780	2.155	30.780	30.780
2	1.405	20.072	50.852	1.405	20.072	50.852
3	1.223	17.473	68.325	1.223	17.473	68.325
4	.953	13.615	81.940			
5	.606	8.660	90.599			
6	.511	7.298	97.898			
7	.147	2.102	100.000			

Extraction Method: Principal Component Analysis.

Component Matrix ^a			
	Component		
	1	2	3
precio	.897	-.033	.201
capacidad	-.201	.807	-.020
potencia	.797	.327	-.121
pesovacio	-.023	.767	.353
nprog	.364	-.113	-.033
tempprelav	.668	.076	-.510
tlav	.309	-.196	.884

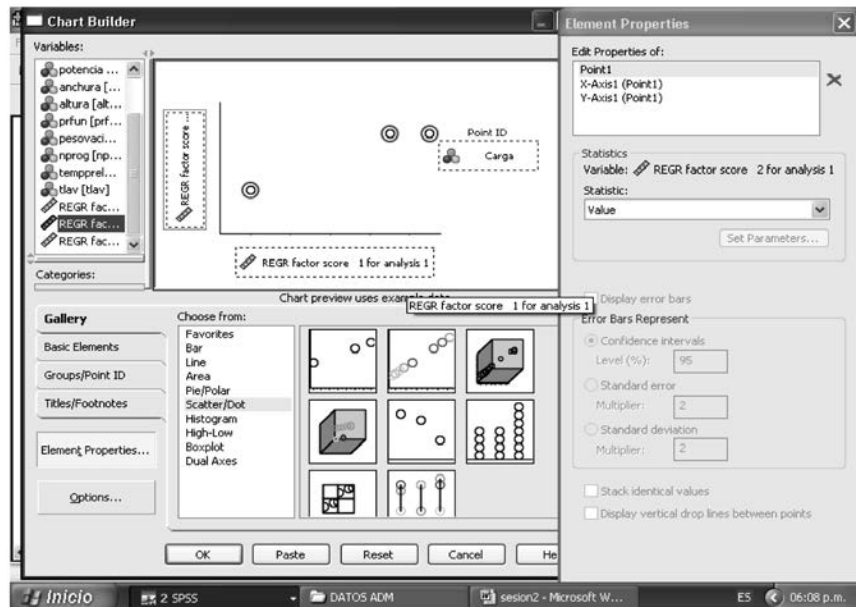
Extraction Method: Principal Component Analysis.

a. 3 components extracted.

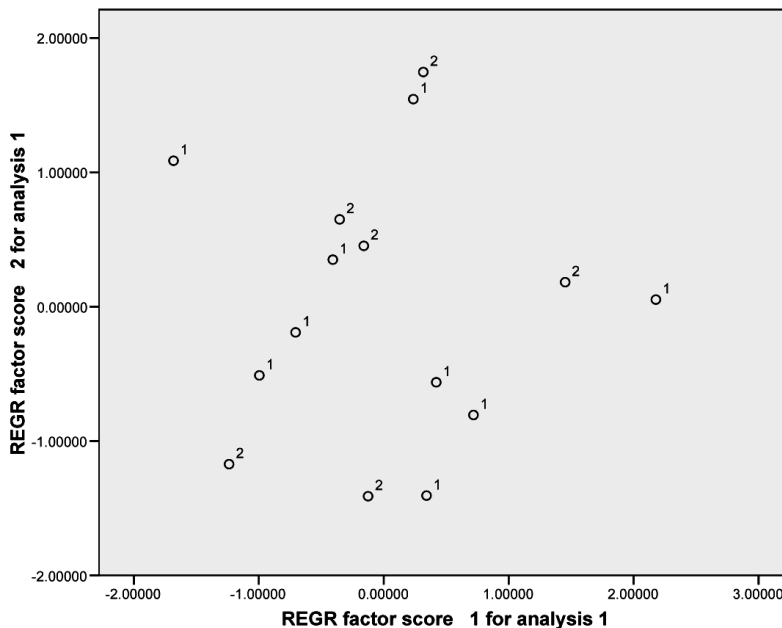
Finalmente, nos quedamos con tres componentes por criterios de autovalores mayores que 1. La interpretación de la matriz de componentes es el precio; la potencia y la temperatura prelavado son las que pesan más (mayores valores).

- Hacemos el gráfico de los componentes. Utilizamos las puntuaciones de factor (Factor scores) generadas automáticamente con el método regresión¹³.

13. Se generan puntuaciones y se almacenan como variables los valores de los factores (método de regresión).



- Tenemos como resultado:



- Ahora, incorporamos las variables altura, anchura y profundidad y ejecutamos nuevamente el método de Componentes principales.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.361	33.614	33.614	3.361	33.614	33.614
2	2.225	22.251	55.865	2.225	22.251	55.865
3	1.414	14.138	70.002	1.414	14.138	70.002
4	1.138	11.378	81.381	1.138	11.378	81.381
5	.613	6.129	87.510			
6	.535	5.354	92.864			
7	.458	4.582	97.445			
8	.145	1.447	98.892			
9	.089	.889	99.781			
10	.022	.219	100.000			

Extraction Method: Principal Component Analysis.

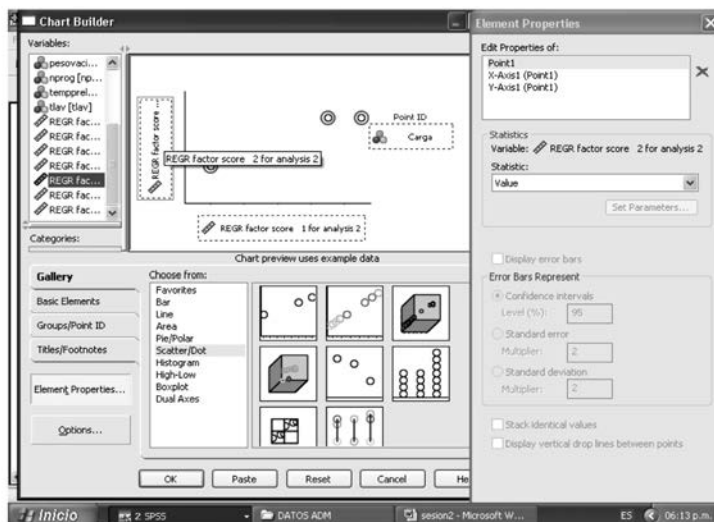
Component Matrix^a

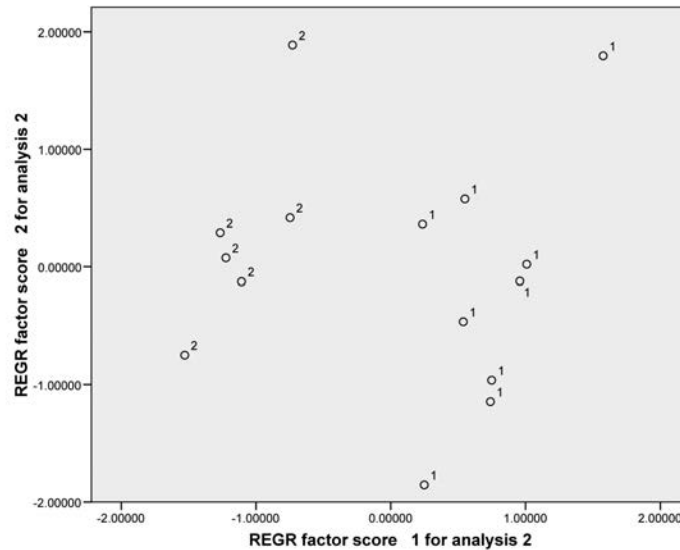
	Component			
	1	2	3	4
precio	.380	.808	-.003	.317
capacidad	-.093	-.216	.798	-.004
potencia	.186	.749	.369	.226
pesovacio	.339	-.153	.747	-.187
nprog	.094	.391	-.105	-.828
tempprelav	-.278	.789	.123	-.113
tlav	.835	.069	-.221	.073
anchura	.875	-.398	.063	.104
altura	-.754	-.024	.015	.463
prfun	.970	.042	-.063	.147

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

- Y graficamos las variables generadas en la corrida anterior (FAC1_2, FAC2_2).





Ejercicio 2

Realizar el mismo análisis de componentes principales para la base medcorp.sav tomando como base el ejercicio anterior.

MEDCORP: Datos de 18 individuos, donde presentan mediciones y observaciones de las siguientes características: cuello, pecho, cintura, caderas, muslo, pantorrilla, tobillo, cabeza, hombros, cintura, rodillas, calzado, peso, género, calle, ojos y nombre de individuo.

Solución en SPSS

A continuación se presenta la matriz de correlaciones. En el cuadro de diálogo de componentes principales, no olvidar generar puntuaciones de factor y almacenar los valores de los factores (método de regresión).

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.259	43.823	43.823	5.259	43.823	43.823
2	2.056	17.137	60.960	2.056	17.137	60.960
3	1.314	10.954	71.913	1.314	10.954	71.913
4	1.154	9.613	81.526	1.154	9.613	81.526
5	.822	6.853	88.380			
6	.531	4.423	92.803			
7	.358	2.986	95.789			
8	.242	2.021	97.810			
9	.145	1.207	99.016			
10	.053	.443	99.460			
11	.039	.322	99.782			
12	.026	.218	100.000			

Extraction Method: Principal Component Analysis.

Component Matrix^a

	Component			
	1	2	3	4
Contorno	.800	-.132	-.166	.152
Pecho	.539	.641	-.391	-.083
Cintura	.765	.280	-.029	.346
Caderas	.726	.394	.239	.291
Muslo	.336	.687	.440	-.276
Pantorrill	.697	.149	.568	-.085
tobillo	.735	.129	-.341	.222
acabeza	.844	-.361	-.157	-.202
ahombro	.745	-.581	.153	-.164
acintura	.555	-.041	-.546	-.338
arodillas	.701	-.531	.304	-.230
calle	.104	-.362	.078	.743

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Análisis

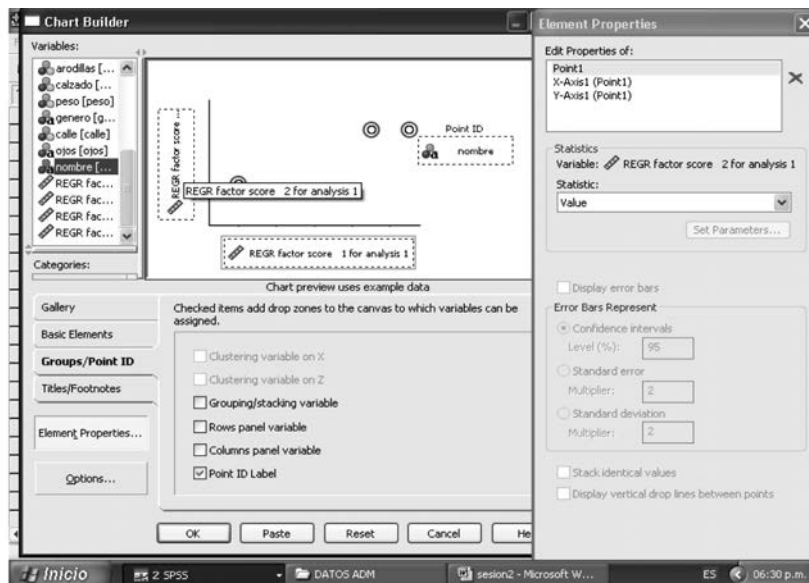
La primera componente representa las variables siguientes: contorno, cintura, cadera, pantorrilla, tobillo, cabeza, hombros, rodillas (tamaño de la persona).

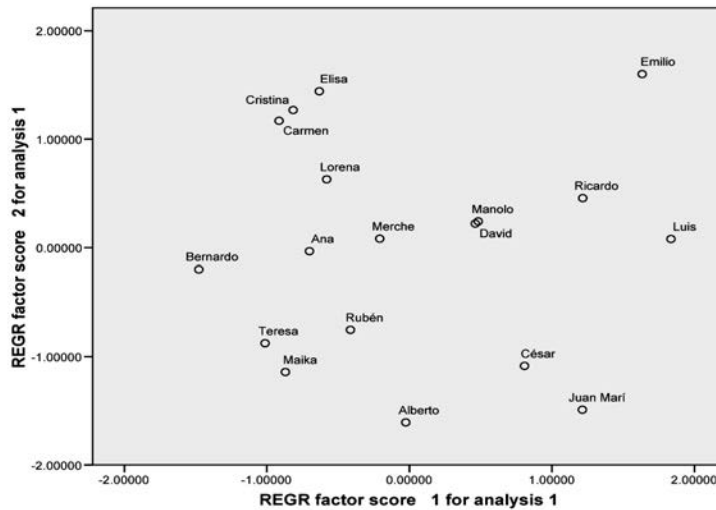
La segunda representa: pecho, muslo, hombros, rodillas (forma o constitución de la persona).

La tercera representa: pantorrilla, cintura.

La cuarta representa: calle.

Ahora, generar un gráfico de dispersión con el factor uno y factor dos. Generador de gráfico.





Si trazamos una línea para dividir en cuadrantes, observamos que *Bernardo* es el dato atípico, por estar más a la izquierda. Las personas de sexo femenino son más obesas (están en el cuadrante II) que los varones (cuadrante III y IV); es decir se puede diferenciar hombres de mujeres. La componente nos dice el tamaño y los ordena. El más pequeño es Bernardo; Luis es alto y delgado y no robusto (gordo). Emilio es alto y gordo. Elisa y Maika tienen el mismo tamaño, pero Maika es menos gorda.

Este gráfico sirve para identificar datos atípicos. Emilio es un individuo atípico, debido a que está alejado del resto. Puede ser porque el dato es tomado erróneamente.

Veamos el efecto de tener un mal dato en la base de datos. Modifiquemos la variable Caderas de Luis (tiene 104 y cambiar por 61).

- Corremos nuevamente el ACP.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.798	43.619	43.619	4.798	43.619	43.619
2	1.905	17.321	60.940	1.905	17.321	60.940
3	1.666	15.146	76.087	1.666	15.146	76.087
4	1.091	9.918	86.005	1.091	9.918	86.005
5	.568	5.163	91.168			
6	.399	3.631	94.799			
7	.221	2.008	96.807			
8	.162	1.477	98.284			
9	.105	.957	99.241			
10	.049	.447	99.688			
11	.034	.312	100.000			

Extraction Method: Principal Component Analysis.

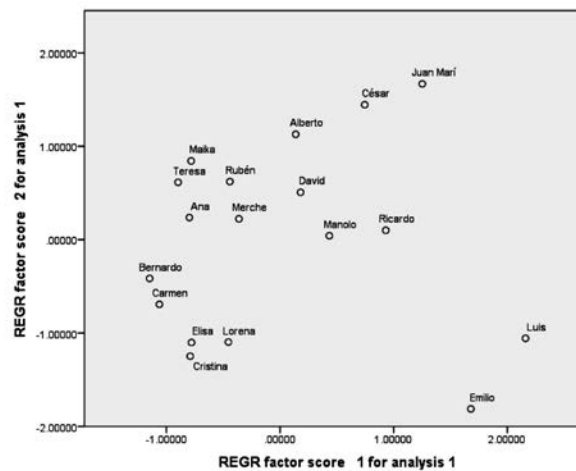
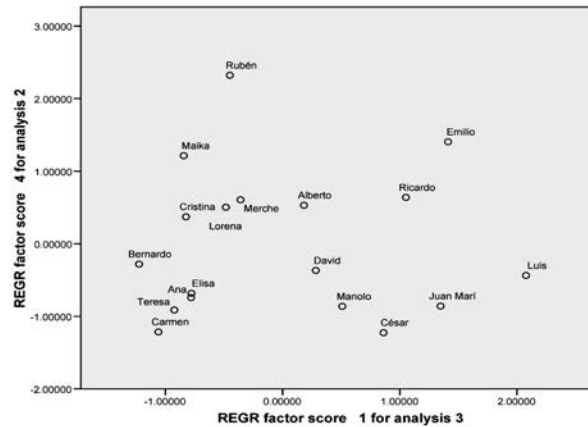
Component Matrix^a

	Component			
	1	2	3	4
Contorno	.801	.009	.100	-.414
Pecho	.518	.483	-.642	.027
Cintura	.710	.441	.023	-.332
Caderas	-.117	.589	.642	-.255
Muslo	.260	.706	-.023	.593
Pantorrilla	.665	.365	.367	.334
tobillo	.736	.155	-.266	-.342
acabeza	.878	-.332	.049	-.046
ahombros	.798	-.386	.382	.085
acintura	.601	-.259	-.584	.210
arodillas	.742	-.387	.368	.331

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

- Y hacemos el gráfico de este segundo análisis:



3.10. Ejercicios

Ejercicio 1

Realizar un análisis de componentes principales para el anexo elecciones.sav. Se solicita:

ELECCIONES: Base de datos que presenta información de votantes por ciudades según el partido político en contienda. Esta base ya se encuentra transformada de acuerdo a la ciudad, es decir, se encuentra totalizado el número de votos por cada partido político.

- a) Realizar un gráfico de dispersión múltiple para los datos.
- b) Realizar un análisis de componentes principales utilizando la matriz de correlaciones.
- c) Nombrar el primer componente, como el tamaño (número de habitantes por municipio).
- d) En la segunda componente, ¿cuáles son los partidos de derecha que agrupa? (Verificar que son PSOE, PCE, EXI). ¿Cómo nombraría dicho componente?
- e) Qué municipio es el que más predomina en el partido de derecha.
- f) Cuáles son los municipios que se concentran en la parte baja del gráfico.
- g) Del (d) al (f) se contrastará a partir del gráfico del primer componente vrs. el segundo componente.
- h) Verificar y decir cuáles municipios tienen menor peso poblacional (esto a partir de aquellos municipios que se encuentran a la izquierda del eje vertical).

Ejercicio 2

Con los datos de índice de bienestar (anexo índice_bienestar.sav). Se solicita:

ÍNDICE_BIENESTAR: Una base de datos que presenta información de 50 ciudades en relación a las siguientes características: provincia, renta, salud, servicios sanitarios, educación cultural, trabajo, empleo, equipos de hogar, acceso a economía, convivencia, seguridad ciudadana y entorno natural.

- a) Decir cuál de todas las provincias vive mejor.
- b) Una de las características primordiales para utilizar la matriz de covarianza es el hecho de que todas las variables están medidas de la misma métrica. Entonces realizar una ACP para que nos ayude a analizar el índice global.
- c) Verificar que la primera componente representa la convivencia y seguridad.
- d) ¿Qué provincias presentan un bienestar con calidad baja?
- e) ¿Son las provincias más inseguras las que se agrupan al lado izquierdo del eje horizontal? (Realizar el gráfico de la componente 1 vrs. componente 2).
- f) ¿Cómo concluiría usted después del ACP?

4. Análisis factorial (AF)

Introducción

El análisis factorial tiene como objeto explicar un conjunto de variables observadas por un pequeño número de variables latentes, o no observadas, que llamaremos factores. Por ejemplo, supongamos que hemos tomado veinte medidas físicas del cuerpo de una persona: estatura, longitud del tronco y de las extremidades, anchura de hombros, peso, etc. Es intuitivo que todas estas medidas no son independientes entre sí, y que conocidas algunas de ellas podemos prever con poco error las restantes. Una explicación de este hecho es que las dimensiones del cuerpo humano dependen de ciertos factores, y si estos fuesen conocidos podríamos prever con pequeño error los valores de las variables observadas.

Como segundo ejemplo, supongamos que estamos interesados en estudiar el desarrollo humano en los países del mundo, y que disponemos de muchas variables económicas, sociales y demográficas, en general dependientes entre sí, que están relacionadas con el desarrollo. Podemos preguntarnos si el desarrollo de un país depende de un pequeño número de factores tales que, conocidos sus valores, podríamos prever el conjunto de las variables de cada país.

Como tercer ejemplo, supongamos que medimos con distintas pruebas la capacidad mental de un individuo para procesar información y resolver problemas. Podemos preguntarnos si existen unos factores, no directamente observables, que explican el conjunto de resultados observados. El conjunto de estos factores será lo que llamaremos inteligencia y es importante conocer cuántas dimensiones distintas tiene este concepto y cómo caracterizarlas y medirlas. El análisis factorial surge impulsado por el interés de Karl Pearson y Charles Sperman en comprender las dimensiones de la inteligencia humana en los años 30, y muchos de sus avances se han producido en el área de psicometría.

El análisis factorial está relacionado con los componentes principales, pero existen ciertas diferencias. En primer lugar, los componentes principales se construyen para explicar las varianzas (varianza total), mientras que los factores se construyen para explicar las covarianzas o correlaciones entre las variables (varianza común).

En segundo lugar, componentes principales es una herramienta descriptiva, mientras que el análisis factorial presupone un modelo estadístico formal de generación de la muestra dada.

La ventaja que tiene el Análisis factorial es que en este se tratan todas las variables simultáneamente, cada una relacionada con las demás y no existe una variable dependiente y otras dependientes.

El objetivo fundamental será el encontrar una manera de resumir la información contenida en una serie de variables originales, mediante una serie de dimensiones compuestas (factores) o valores teóricos con una mínima pérdida de información.

4.1. El modelo factorial

Consideramos las variables observables $X_1, X_2, \dots, X_p$ como variables tipificadas (con media cero y varianza unidad) y vamos a formalizar la relación entre las variables observables y factores, definiendo el modelo factorial de la siguiente forma:

$$\begin{aligned} X_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1k}F_k + e_1 \\ X_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2k}F_k + e_2 \\ &\vdots \\ X_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pk}F_k + e_k \end{aligned}$$

En este modelo, $F_1, F_2, \dots, F_k$ son los factores comunes; $e_1, e_2, \dots, e_k$ son los factores únicos o factores específicos, y l_{jk} es el peso del factor h en la variable j , denominado también *carga factorial* o *saturación de la variable j en el factor h* . Así pues, todas las variables originales están influenciadas por todos los factores comunes, mientras que para cada variable existe un factor único que es específico para esa variable. Tanto los factores comunes como los específicos son variables no observables. El modelo factorial en forma matricial se expresa como sigue:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} l_{11} + l_{12} + \dots + l_{1k} \\ l_{21} + l_{22} + \dots + l_{2k} \\ \vdots \\ l_{p1} + l_{p2} + \dots + l_{pk} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

Equivalentemente: $X = LF + e$

4.2. Hipótesis previas

Dentro de las hipótesis previas la primera de ellas sería la *normalidad*, aunque esta sólo será necesaria si se realizan pruebas estadísticas de significación. También es deseable que exista un cierto grado de *multicolinealidad*, ya que el objetivo es identificar variables relacionadas.

El análisis factorial debe tener suficientes correlaciones como para poder aplicarse. Si no hay un número sustancial de correlaciones mayores de 0.3, entonces es probablemente inadecuado. Las correlaciones entre variables pueden ser analizadas mediante el cálculo de las correlaciones parciales, de tal manera que, si las correlaciones parciales son bajas, entonces no existen factores subyacentes verdaderos y el análisis factorial es inapropiado.

Para poder aplicar la teoría de la inferencia estadística en el modelo factorial es necesario formular hipótesis estadísticas sobre los factores comunes y sobre los factores únicos. Consideraremos los factores comunes $F_1, F_2, \dots, F_k$ como variables tipificadas de media cero y varianza unitaria, y que además no están correlacionadas entre sí. Según esta condición la matriz de covarianzas de los factores comunes es la matriz identidad ($E[FF'] = 1$) y la esperanza del vector de factores comunes es el vector cero ($E[F] = 0$).

Por otra parte, se supone que la matriz de covarianzas de los factores específicos (únicos) es una matriz diagonal, lo que implica que las varianzas de los factores únicos pueden ser distintas y que dichos factores únicos están incorrelacionados entre sí, es decir: ($E[ee'] = \Omega$) con Ω matriz diagonal. Por otro lado, la esperanza del vector de factores comunes se supone que es el vector cero ($E[e] = 0$).

Por último, se tendrá en cuenta que para poder realizar inferencias que permitan distinguir, para cada variable, entre los factores comunes y el factor único, es necesario suponer que los factores comunes están incorrelacionados con el factor único, es decir, que la matriz de covarianzas entre los factores comunes y los factores únicos es la matriz cero ($E[Fe'] = 0$).

Resumiendo las hipótesis previamente citadas tenemos:

$$\text{Modelo} \rightarrow X = LF + e$$

$$\text{Hipótesis} \rightarrow E[FF'] = 1, E[F] = 0, E[ee'] = \Omega, E[e] = 0, E[Fe'] = 0$$

Dado que las variables X son variables tipificadas, su matriz de covarianzas es igual a la matriz de correlación poblacional R_p , matriz que puede descomponerse de la forma siguiente:

$$R_p = E[XX'] = E(LF + e)(LF + e)' = LL' + \Omega$$

La relación anterior puede expresarse en forma matricial como sigue:

$$\begin{bmatrix} 1 + \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} + 1 & \dots & \rho_{2p} \\ \vdots & & \vdots \\ \rho_{p1} + \rho_{p2} & \dots & 1 \end{bmatrix} = \begin{bmatrix} l_{11} + l_{12} & \dots & l_{1k} \\ l_{21} + l_{22} & \dots & l_{2k} \\ \vdots & & \vdots \\ l_{p1} + l_{p2} & \dots & l_{pk} \end{bmatrix} \begin{bmatrix} l_{11} + l_{21} & \dots & l_{p1} \\ l_{12} + l_{22} & \dots & l_{p2} \\ \vdots & & \vdots \\ l_{1k} + l_{2k} & \dots & l_{pk} \end{bmatrix} + \begin{bmatrix} \omega_1^2 + 0 & \dots & 0 \\ 0 + \omega_2^2 & \dots & 0 \\ \vdots & & \vdots \\ 0 + 0 & \dots & \omega_p^2 \end{bmatrix}$$

En esta descomposición LL' es la parte correspondiente a los factores comunes y Ω es la matriz de covarianzas de los factores únicos. Además, en la descomposición se observa que la varianza de la variable tipificada X_j se puede expresar como:

$$V_j = 1 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jp}^2 + \omega_j^2$$

Y si denominamos:

$$h_j^2 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jp}^2$$

Tenemos la descomposición de la varianza poblacional de la variable X_j como:

$$V_j = 1 = h_j^2 + \omega_j^2, j = 1, \dots, p$$

Se observa que h_j^2 es la parte de la varianza de la variable X_j debida a los factores comunes, y se denomina *comunalidad*.

También se observa que ω_j^2 es la parte de la varianza de la variable X_j debida a los factores únicos (o específicos), y se denomina *especificidad*.

De la relación matricial anterior también se deduce que la correlación entre cada par de variables originales X_h y X_j viene dada en función de los coeficientes de los factores comunes como sigue:

$$\rho_{hj} = l_{h1}l_{j1} + l_{h2}l_{j2} + \dots + l_{hp}l_{jp} = \sum_{s=1}^p l_{hs}l_{js}$$

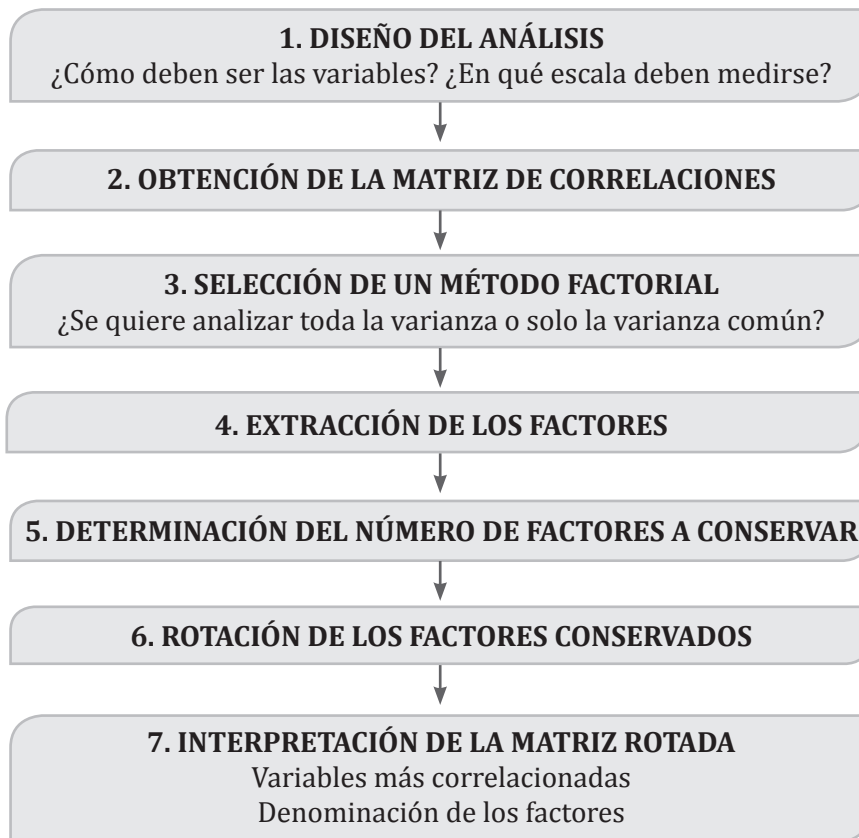
4.3. Importancia de las aplicaciones de Análisis factorial (AF)

1. Utilización de resultados del AF de componentes principales como “variables-releva”. Los resultados de un AF pueden utilizarse como fase previa de cálculo antes de la aplicación de otros métodos; puesto que, los factores obtenidos son independientes, se pueden utilizar como “nuevas variables” evitando dificultades en el caso de que exista una correlación fuerte entre las variables

iniciales, que es uno de los problemas más comunes en el RLM, análisis tipológico o análisis discriminante.

2. Obtención e interpretación de factores. El ACP permitirá, por una parte, estructurar las variables iniciales con el fin de determinar si existe una estructura latente, que no es más que un resumen de la información suministrada por las variables iniciales, y por otra parte, interpretarlas en el caso de que existan dichas variables latentes.
3. Selección de variables. Este tipo de análisis permite seleccionar, de entre un conjunto importante de variables, cuáles son las que más intervienen en la descripción del fenómeno estudiado, haciendo posible que conservemos para análisis posteriores, únicamente aquellas variables iniciales que estén fuertemente correlacionadas con los factores que hemos considerado más importantes.

4.4. Metodología del Análisis factorial



4.5. Aplicación del método en SPSS

Supongamos que tenemos una matriz de orden 1000x20, una muestra de 1000 encuestados (estudiantes de educación media) a quienes se les solicita una valoración, en una escala de 1 a 6, donde 1 significa “nada deseable” y 6 “totalmente deseable”, respecto a 20 características referentes al comportamiento de sus profesores. ¿Podemos analizar los 20,000 valores?, o bien, ¿Podemos sintetizar o resumir toda la información en una, dos o tres variables? ¿No existirá cierta relación entre las variables? ¿Podemos eliminar algunas que no aportan más que una información marginal y poco interesante?

Si existe una relación de interdependencia sistemática en el conjunto de X_i , ¿no puede ser debida originalmente a algunos factores, latentes, más fundamentales?

Como hemos señalado, el análisis factorial tiene como objetivo identificar la estructura de las relaciones entre variables intercorrelacionadas o entre individuos. Cuando el objetivo es identificar la estructura de las relaciones entre variables, el análisis recibe el nombre de componentes principales (ACP).¹⁴

Tomaremos como ejemplo la pregunta 17 de un cuestionario cuyos datos recolecta el archivo trabajo.sav y está referida a la evaluación, por parte de los encuestados, de la importancia que según su opinión, puede tener cada una de las causas del alto índice de desempleo en España:

- b13 La crisis económica.
- b14 La política de empleo del Gobierno.
- b15 La mala gestión de los empresarios.
- b16 La comodidad de la gente, que solo quiere buenos trabajos.
- b17 La falta de preparación del trabajador.
- b18 Las pocas ganas de trabajar de la gente.
- b19 El no saber buscar trabajo.
- b20 Que hay mucho pluriempleo.
- b21 El trabajo que hay no se reparte bien socialmente.

El modelo matemático que subyace a esta técnica es similar al de la RLM y al análisis discriminante, en el que cada variable aparece como una combinación lineal de una serie de factores que no son en este momento observables. Por ejemplo, la variable b13 (las crisis económicas) puede aparecer expresada como:

$$b_{13} = \alpha (\text{sujeto}) + \beta (\text{externas al sujeto}) + \Delta (\text{entorno}) + e_{13}$$

14. El ACP y el AF se utilizan indistintamente y dan resultados similares. En el caso del ACP el objetivo consiste en encontrar una serie de componentes que expliquen el máximo de varianza total de las variables originales; mientras que en el AF el objeto es encontrar una serie de factores que expliquen el máximo de varianza común de las variables originales.

Las descripciones dentro de los paréntesis son grupos de variables desconocidas a priori, correspondientes a los factores subyacentes. El factor “sujeto” puede englobar una serie de variables referidas a las causas del desempleo inherentes al propio sujeto; el factor “externas al sujeto” estaría señalando como causante del desempleo al Gobierno y a los empresarios, y el “entorno” donde podrían estar la crisis económica y el reparto del trabajo.

Observar que el término e_{13} es el factor único referido a la parte de la variable $b13$ que no puede ser explicada por los factores comunes.¹⁵

Ahora aplicaremos la metodología anterior a este caso en particular.

1. Diseño del Análisis factorial

Consideraciones:

- Las variables que se utilizan deben ser comparables; si no lo son, los factores que se obtengan no tendrán sentido.
- Si las escalas de medida de las variables son diferentes, será necesario normalizarlas, ya que las varianzas de cada variable intervienen en el análisis, y las variables que presentan mayor varianza tendrían ciertas ventajas.
- Las variables deben ser medias en escalas cuantitativas (métricas o de intervalo).

Abrir el archivo trabajo.sav e identificar la pregunta 17 de un cuestionario cuyos datos están referidos a la evaluación, por parte de los encuestados, de la importancia que puede tener cada una de las causas del alto desempleo en España.

2. Obtención de la matriz de correlaciones

Acceder al menú de Reducción de dimensiones/factor y seleccionar los cuadros adecuados de dialogo.

A partir de la matriz inicial de datos obtenemos la matriz de correlaciones entre variables. Interesa correlación de Pearson, altamente significativa y un determinante de la matriz relativamente bajo.

15. Los factores únicos se asumen que están incorrelacionados con el resto de factores únicos y con los factores comunes.

Correlation Matrix^a

		Crisis	Política de empleo	Empresarios	Comodidad	Preparación	Ganas de trabajar	Búsqueda	Pluriempleo	Reparto
Correlation	Crisis	1.000	.397	.185	-.120	-.003	-.157	-.101	.019	.084
	Política de empleo	.397	1.000	.202	-.077	-.050	-.104	-.078	.054	.103
	Empresarios	.185	.202	1.000	.028	-.010	-.024	.044	.101	.161
	Comodidad	-.120	-.077	.028	1.000	.336	.559	.387	.214	.043
	Preparación	-.003	-.050	-.010	.336	1.000	.425	.345	.115	.045
	Ganas de trabajar	-.157	-.104	-.024	.559	.425	1.000	.451	.195	.071
	Búsqueda	-.101	-.078	.044	.387	.345	.451	1.000	.231	.134
	Pluriempleo	.019	.054	.101	.214	.115	.195	.231	1.000	.376
	Reparto	.084	.103	.161	.043	.045	.071	.134	.376	1.000
Sig. (1-tailed)	Crisis		.000	.000	.000	.457	.000	.001	.273	.004
	Política de empleo			.000	.007	.057	.000	.006	.044	.001
	Empresarios				.185	.373	.221	.082	.001	.000
	Comodidad					.000	.000	.000	.000	.087
	Preparación						.000	.000	.000	.077
	Ganas de trabajar							.000	.000	.012
	Búsqueda								.000	.000
	Pluriempleo									.000
	Reparto									.000

a. Determinant = .240

El test de Bartlett's se utiliza para verificar si la matriz de correlaciones es una matriz de identidad y se obtiene a través de la transformación X^2 del determinante de la matriz de correlaciones, y cuanto mayor sea y tanto menor sea el grado de significancia, es más improbable que la matriz sea una matriz de identidad; en nuestro caso, resulta evidente que no se trata de una matriz de identidad. Si fuera el caso es que no se pudiese rechazar esta hipótesis, no fuera aconsejable proceder a realizar un AF con los datos.

Inverse of Correlation Matrix

	Crisis	Política de empleo	Empresarios	Comodidad	Preparación	Ganas de trabajar	Búsqueda	Pluriempleo	Reparto
Crisis	1.239	-.441	-.134	.061	-.115	.138	.059	-.008	-.049
Política de empleo	-.441	1.223	-.153	.014	.034	.024	.048	-.043	-.059
Empresarios	-.134	-.153	1.086	-.063	.032	.054	-.058	-.038	-.128
Comodidad	.061	.014	-.063	1.548	-.148	-.674	-.207	-.160	.080
Preparación	-.115	.034	.032	-.148	1.290	-.376	-.231	.005	.005
Ganas de trabajar	.138	.024	.054	-.674	-.376	1.740	-.361	-.067	-.027
Búsqueda	.059	.048	-.058	-.207	-.231	-.361	1.380	-.144	-.086
Pluriempleo	-.008	-.043	-.038	-.160	.005	-.067	-.144	1.247	-.427
Reparto	-.049	-.059	-.128	.080	.005	-.027	-.086	-.427	1.201

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.712
Bartlett's Test of Sphericity	Approx. Chi-Square
	1434.418
	df
	36
	Sig.
	.000

Respecto al índice KMO (Kaiser-Meyer-Olkin), este compara los coeficientes de correlación de Pearson obtenidos en la matriz de correlación con los coeficientes de correlación parcial entre variables. Los valores de este índice pueden ser ≤ 1 , valores entre 0.7 y ≤ 1.0 son considerados aceptables.

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2}$$

Donde r_{ij}^2 es el coeficiente de correlación de Pearson entre las variables i y j; a_{ij}^2 es el coeficiente de correlación Parcial entre las variables i y j. Con el valor obtenido de .712 puede considerarse aceptable y proseguir con el AF.

El procedimiento genera una matriz anti-imagen de covarianzas y correlaciones entre todas las variables de la base de datos; interesan coeficientes pequeños.¹⁶ Además, en la diagonal se presentan los coeficientes MSA (*Measures of sampling adequacy*), que equivalen a los KMO, pero en este caso para cada variable por separado; interesa que estos coeficientes sean bastantes altos.

Anti-image Matrices										
		Crisis	Política de empleo	Empresarios	Comodidad	Preparación	Ganas de trabajar	Búsqueda	Pluriempleo	Reparto
Anti-image Covariance	Crisis	.807	-.291	-.100	.032	-.072	.064	.034	-.005	-.033
	Política de empleo	-.291	.818	-.115	.007	.022	.011	.028	-.028	-.040
	Empresarios	-.100	-.115	.921	-.038	.023	.029	-.039	-.028	-.098
	Comodidad	.032	.007	-.038	.646	-.074	-.250	-.097	-.083	.043
	Preparación	-.072	.022	.023	-.074	.775	-.168	-.130	.003	.003
	Ganas de trabajar	.064	.011	.029	-.250	-.168	.575	-.151	-.031	-.013
	Búsqueda	.034	.028	-.039	-.097	-.130	-.151	.725	-.083	-.052
	Pluriempleo	-.005	-.028	-.028	-.083	.003	-.031	-.083	.802	-.285
	Reparto	-.033	-.040	-.098	.043	.003	-.013	-.052	-.285	.833
Anti-image Correlation	Crisis	.601*	-.358	-.116	.044	-.091	.094	.045	-.006	-.040
	Política de empleo	-.358	.609*	-.132	.010	.027	.016	.037	-.034	-.048
	Empresarios	-.116	-.132	.690*	-.049	.027	.039	-.047	-.033	-.112
	Comodidad	.044	.010	-.049	.744*	-.104	-.411	-.142	-.115	.058
	Preparación	-.091	.027	.027	-.104	.791*	-.251	-.173	.004	.004
	Ganas de trabajar	.094	.016	.039	-.411	-.251	.721*	-.233	-.046	-.019
	Búsqueda	.045	.037	-.047	-.142	-.173	-.233	.816*	-.109	-.067
	Pluriempleo	-.006	-.034	-.033	-.115	.004	-.046	-.109	.669*	-.349
	Reparto	-.040	-.048	-.112	.058	.004	-.019	-.067	-.349	.591*

a. Measures of Sampling Adequacy(MSA)

Todo el análisis previo nos lleva a concluir que el AF resulta pertinente y puede proporcionar conclusiones satisfactorias.

3. Selección de un método factorial

Luego de obtener la matriz de correlaciones, necesitamos reducir el número de variables originales a un número menor de variables. La selección del método depende de los intereses del investigador (Análisis de factores comunes y el Análisis de componentes principales-ACP). El ACP se utiliza cuando el objetivo es resumir la mayoría de información inicial (varianza) en un número mínimo de factores; mientras que el análisis de factores comunes se utiliza para identificar factores o dimensiones subyacentes que reflejen la cantidad de varianzas que las variables originales tienen en común.

16. Los negativos de los coeficientes de correlación parcial entre cada par de variables neutralizan el efecto de todas las variables restantes.

El ACP es el método más utilizado que busca extraer los factores, combinaciones lineales de las variables iniciales, como método para restituir la mayor cantidad de información posible,¹⁷ sin hacer *a priori* ningún tipo de hipótesis sobre el significado de los factores que se van a obtener. Lo que se pretende es descubrir cuáles son las dimensiones principales del espacio que se está estudiando, e interpretar los factores en función de sus relaciones con las variables iniciales.

El método de Factores comunes es la representación clásica desarrollada inicialmente por sociólogos y psicólogos. En lugar de suponer que los factores que se buscan son combinaciones lineales de las variables iniciales, se considera como hipótesis que son las variables iniciales las que son combinaciones lineales de un cierto número de factores subyacentes y no observables.¹⁸

4. Extracción de los factores

A partir de la matriz de correlaciones se obtienen los factores que identifican la estructura subyacente de las relaciones entre las variables iniciales. El sistema o procedimiento estadístico selecciona por defecto el método de componentes principales, el cual como ya hemos comentado, es un método donde se transforman las variables originales en unas nuevas variables, que son combinación lineal de las variables iniciales y que están incorreladas. Estas nuevas variables se denominan componentes principales.

Los componentes principales se obtienen según la varianza restituida por cada uno de ellos. Así, el componente principal que se obtiene en primer lugar es aquel que restituye una mayor proporción de la varianza contenida en la matriz de correlaciones; el segundo componente es aquel que, de la varianza restante, restituye el mayor porcentaje, y así sucesivamente hasta llegar al último factor que restituye la menor cantidad de varianza inicial.

Los componentes extraídos deben cumplir tres condiciones:

1. *Linealidad*. Cada componente principal es una combinación lineal de las variables iniciales:

$$Z_k = a_{k_1} X_1 + a_{k_2} X_2 + \dots + a_{k_p} X_p$$

Donde:

Z_k es el k-ésimo componente;

-
17. La medida de la cantidad de información restituida por cada componente principal es la varianza.
 18. Por ejemplo, en Psicología, en el caso de que los resultados de varios test deben relacionarse con los factores que se busca poner de manifiesto: inteligencia, memoria, etc.

a_{kp} es el coeficiente de la variable X_p en el factor F_k

X_p los valores de la variable p tomados por los n individuos.

- Independencia. Los componentes no están correlacionados entre ellos.

$$\text{Corr}(F_k, F_{k'}) = 0 \quad \forall k \neq k'$$

- Varianza restituida por cada componente.* El primer componente restituye la mayor proporción de varianza contenida en la matriz de correlaciones; el segundo componente la segunda proporción, el tercero la tercera proporción, y así sucesivamente.

$$\text{Var } F_1 \geq \text{Var } F_2 \geq \dots \geq \text{Var } F_k$$

Así, los estadísticos iniciales se presentan a continuación. El valor propio λ_k indica la cantidad de varianza restituida por el factor k . A mayor valor λ_k mayor es el nivel explicativo del componente asociado. La tabla recolecta en porcentajes individuales y acumulados, la proporción de varianza total explicada por cada factor, para la solución no rotada.

Total Variance Explained							Communalities		
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings				Initial	Extraction
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %			
1	2.449	27.211	27.211	2.449	27.211	27.211	Crisis	1.000	.644
2	1.684	18.714	45.925	1.684	18.714	45.925	Política de empleo	1.000	.620
3	1.116	12.395	58.320	1.116	12.395	58.320	Empresarios	1.000	.329
4	.848	9.426	67.747				Comodidad	1.000	.592
5	.705	7.834	75.580				Preparación	1.000	.523
6	.616	6.842	82.422				Ganas de trabajar	1.000	.684
7	.597	6.629	89.051				Búsqueda	1.000	.519
8	.568	6.314	95.365				Pluriempleo	1.000	.646
9	.417	4.635	100.000				Reparto	1.000	.693

Extraction Method: Principal Component Analysis.

Extraction Method: Principal Component Analysis.

A partir de estos valores se puede calcular el porcentaje de varianza restituida por cada componente aplicando la fórmula siguiente:

$$\% V(F_k) = \frac{\lambda_k}{\sum_{k=1}^k \lambda_k}$$

En el ejemplo, el primer componente resume el 27.2% de la varianza total: el segundo componente, el 18.7%; el tercero, el 12.4% y el noveno, 4.6%. Observar que, con los dos primeros componentes, el análisis restituye el 45.9% de la información contenida en la matriz de correlaciones.

La *comunalidad* indica la proporción de varianza de cada variable explicada por los factores seleccionados. Al trabajar inicialmente con todos los factores, la

comunalidad de cada variable es máxima, es decir, igual a uno (cuando utilizamos tantos componentes principales como variables, cada variable puede ser explicada por ella misma y por tanto toda la variabilidad de cada variable, que expresada en unidades de desviación estandarizadas, es igual a la unidad, explicada a su vez por los factores comunes).

5. Determinación del número de factores a conservar

Existe una gran cantidad de reglas y criterios para determinar cuál es el número de ejes factoriales a conservar. La mayoría de los programas más conocidos suele aplicar el denominado *criterio de Kaiser*, según el cual solo se conservan los factores cuyos valores propios son mayores que la unidad. En nuestro ejemplo seleccionaríamos los primeros 3 factores.

Existen otros criterios:

1. Reglas basadas en la restitución mínima.
2. Reglas basadas en la información restituida por cada factor.

Restitución mínima. En este caso se fija *a priori* un nivel correspondiente al porcentaje mínimo de varianza que se quiere restituir y conservar el número de ejes necesarios para alcanzar este nivel.

Sabemos que el porcentaje de varianza explicada por los dos primeros factores, es este:

$$\% V(F_1, F_2) = \frac{\lambda_1 + \lambda_2}{\sum_{k=1}^k \lambda_k}$$

En el caso de que esta cantidad alcance el nivel fijado, conservaremos únicamente estos dos primeros componentes; caso contrario, introduciremos en el análisis el tercer factor, y así sucesivamente, hasta alcanzar el nivel fijado.

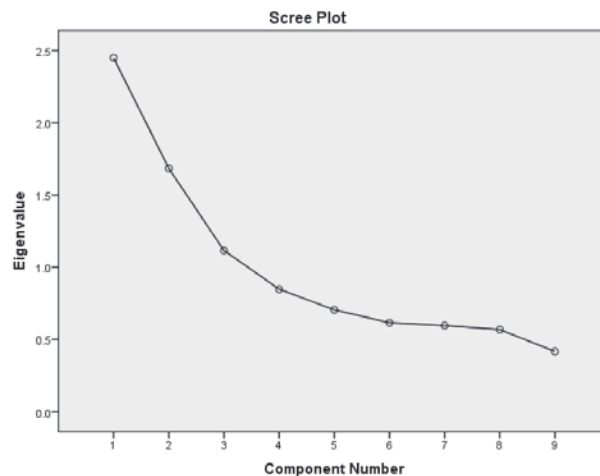
A modo de ejemplo, supongamos que el nivel fijado es el 80%, entonces conservaremos para el análisis aquel número de componentes que cumpla esta fórmula:

$$\% V(F_1, F_2 \dots F_K) = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\sum_{k=1}^k \lambda_k} \geq 0.80$$

En nuestro ejemplo, seleccionaríamos los primeros seis factores.

Información restituida por cada factor. Se puede construir una curva en la cual los puntos en las abscisas sean los números de factores; y en el eje de las ordenadas, el porcentaje de varianza que restituye cada uno de los factores o el valor propio asociado a cada factor.

Se determina el primer punto de inflexión de la curva y se conservan aquellos factores o componentes cuyo número de orden esté situado antes del punto de inflexión. En nuestro ejemplo el cambio de concavidad se produce a partir del tercer factor, lo que indica que deberíamos conservar los primeros tres factores.



En lo referente a los ejes, para hacer las representaciones gráficas de los factores o componentes, estos deben ser tomados dos a dos y, por lo tanto, cada plano factorial explicará un porcentaje de varianza igual a la suma de la que explica cada factor de forma individual. En nuestro ejemplo, el factor uno restituye un 27.2% de la información inicial, y el segundo factor un 18.7%, esto permite afirmar que el primer plano factorial (F1, F2) restituye el 45.9% de la información inicial.

En el ejemplo, el análisis se ha efectuado sobre nueve variables, lo cual permite afirmar que la restitución de los primeros tres factores es considerable. Se decide conservar los tres primeros factores, con los que conservamos el 58.3% de la información inicial.

Un primer resultado que indica la calidad de la representación de la información en los dos factores seleccionados es la *comunalidad*. La comunalidad expresa la proporción de varianza de una variable explicada por los factores seleccionados. Concretamente indica el porcentaje de información de cada variable que estamos perdiendo al trabajar en un espacio determinado. Una comunalidad elevada (cercana a uno) implicará una correlación elevada con *al menos uno* de los factores seleccionados, en cambio una comunalidad baja implicará una correlación baja con

todos los factores seleccionados; es decir, estas variables estarán correlacionadas con otros factores.

6. Rotación de factores conservados

El trabajo en el análisis factorial persigue que los factores comunes tengan una interpretación clara, porque de esa forma se analizan mejor las interrelaciones existentes entre las variables originales. Sin embargo, en muy pocas ocasiones resulta fácil encontrar una interpretación adecuada de los factores iniciales, con independencia del método que se haya utilizado para su extrapolación. Precisamente los procedimientos de rotación de factores se han ideado para obtener, a partir de la solución inicial, unos factores que sean fácilmente interpretables. En la solución inicial cada uno de los factores comunes está correlacionado en mayor o menor medida con cada una de las variables originales. Pues bien, con los factores rotados se trata de que cada una de las variables originales tenga una correlación lo más próxima a una distancia que sea posible con uno de los factores y correlaciones próximas a cero con el resto de los factores. De esta forma, y dado que hay más variables que factores comunes, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de variables.

Examinando las características de las variables de un grupo asociado a un determinado factor se pueden encontrar rasgos comunes que permitan identificar el factor y darle una denominación que responda a esos rasgos comunes. Si se consigue identificar claramente estos rasgos, se habrá dado un paso importante, ya que con los factores comunes no sólo se reducirá la dimensionalidad del problema, sino que también se conseguirá desvelar la naturaleza de las interrelaciones existentes entre las variables originales.

Los ejes factoriales pueden considerarse como las “dimensiones latentes” del problema y el describirlas (interpretarlas) nos conducirá a poder comprender las dimensiones fundamentales del fenómeno objeto de estudio.

Para poder interpretar correctamente estas dimensiones, precisamos saber cuáles son las variables que más contribuyen a la formación de cada factor.¹⁹ Para ello utilizaremos como indicador los coeficientes de correlación entre las variables iniciales y los factores que, en la fase anterior, hayamos decidido conservar para proseguir el análisis. Las variables con los coeficientes de correlación más altos con un factor son las que más contribuyen a la formación de ese factor.

19. Corresponden a los coeficientes utilizados para expresar cada variable estandarizada en términos de los tres factores del modelo. Estos coeficientes son conocidos también como pesos factoriales, cargas factoriales, ponderaciones factoriales o saturaciones factoriales, ya que indican la carga de cada variable en cada factor, de manera que los factores con unos pesos factoriales más elevados (en términos de valor absoluto) indican una relación estrecha con las variables

La matriz que contiene los coeficientes de correlación entre las variables iniciales y los factores se suele llamar matriz factorial inicial o matriz factorial no rotada. Lo ideal es encontrar un modelo en el que todas las variables saturan en algún factor, es decir, pesos factoriales altos en uno y bajos en el resto. Por ejemplo, la variable b18 es una variable con una elevada carga factorial en el primero de los factores y mucho más pequeña en los dos restantes, pudiéndose expresar como sigue:

$$b_{18} = 0.80408 F_1 - 0.08519 F_2 + 0.17407 F_3$$

Tabla x: Matriz de pesos factoriales no rotada

Component Matrix^a

	Component		
	1	2	3
Ganas de trabajar	.804	-.085	.174
Comodidad	.748	-.036	.177
Búsqueda	.718	.045	.040
Preparación	.629	.002	.357
Política de empleo	-.185	.668	.373
Crisis	-.229	.640	.426
Empresarios	.020	.569	.070
Reparto	.236	.551	-.577
Pluriempleo	.436	.437	-.515

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Si los factores están correlacionados (ortogonales), los valores de estos coeficientes no dependen unos de otros y representa la contribución única de cada factor en la variable y en ese sentido no dejan de ser sino los coeficientes de correlación de cada factor con la variable. Para determinar en qué medida los tres factores son capaces de explicar las variables originales, podemos sumar la proporción de varianza de la variable explicada por cada uno de ellos (es decir, los coeficientes al cuadrado) y de este modo obtener las comunales que aparecen en la diagonal de la matriz de correlaciones reproducidas y residuales. Seleccionemos nuevamente la variable b18 y calculemos la comunalidad:

$$b_{18} = 0.80408^2 + 0.08519^2 + 0.17407^2 = 0.68410,$$

lo que indica que aproximadamente el 70% de la variabilidad de b18 es explicada por los tres factores del modelo. Observar que la variable b15 (empresarios) únicamente explica aproximadamente el 33%.²⁰

20. Se reitera que esta proporción de la variabilidad de cada variable explicada por los factores del modelo es lo que se conoce con el nombre de comunalidad de la variable. El valor oscila entre 0 y 1, y la parte de la varianza no explicada por el modelo factorial (1-comunalidad) se conoce con el nombre de factor único o unicidad.

Tabla X: Matriz de correlaciones reproducidas y residuales

Reproduced Correlations

		Crisis	Política de empleo	Empresarios	Preparación	Ganas de trabajar	Comodidad
	Crisis	.644 ^a	.629	.389	.009 ^a	-.164	-.119
	Política de empleo	.629	.620 ^a	.402	.018	-.141 ^a	-.096
	Empresarios	.389	.402	.329 ^a	.039	-.020	.007
	Comodidad	-.119	-.096	.007	.534	.635	.592 ^a
	Preparación	.009	.018	.039	.523	.568	.534
	Ganas de trabajar	-.164	-.141	-.020	.568	.684	.635
	Búsqueda	-.118	-.088	.043	.466	.580	.542
	Pluriempleo	-.039	.019	.221	.091	.223	.219
	Reparto	.053	.109	.277	-.057	.042	.054
Residual ^b	Crisis		-.232	-.204	-.013	.008	-.001
	Política de empleo	-.232		-.201	-.068	.036	.019
	Empresarios	-.204	-.201		-.049	-.004	.021
	Comodidad	-.001	.019	.021	-.197	-.076	
	Preparación	-.013	-.068	-.049		-.143	-.197
	Ganas de trabajar	.008	.036	-.004	-.143		-.076
	Búsqueda	.017	.009	.001	-.121	-.129	-.156
	Pluriempleo	.058	.035	-.120	.024	-.029	-.005
	Reparto	.031	-.006	-.116	.102	.029	-.012

Extraction Method: Principal Component Analysis.

a. Reproduced communalities.

b. Residuals are computed between observed and reproduced correlations. There are 16 (44.0%) nonredundant residuals with absolute values greater than 0.05.

La correlación obtenida entre los distintos factores y las variables puede ser utilizada para estimar la correlación entre las variables. Por ejemplo, de la matriz de coeficientes podemos obtener, para el caso, que la correlación estimada entre las variables b18 (ganas de trabajar) y b16 (comodidad) será la siguiente:

$$r_{b18b16} = (0.80408) (0.74803) + (-0.08519) (-0.03568) + (0.17407) (0.17722) = 0.6353642$$

De hecho, si los factores son ortogonales, el coeficiente de correlación entre las variables i y j es este:

$$r_{ij} = \sum_{f=1}^k r_{fi} r_{fj} = r_{1i} r_{1j} + r_{2i} r_{2j} + \dots + r_{ki} r_{kj}.$$

Donde k es el número de factores comunes, y r_{fi} es el coeficiente de correlación entre el factor f y la variable i.

Estos coeficientes de correlación estimados son los que se presentan en la primera parte de la tabla anterior. La diferencia entre estos coeficientes de correlación así estimados y los coeficientes de correlación observados (encontrados inicialmente) es lo que se conoce como *residual* y son los valores que se presentan en la segunda parte de la tabla anterior. Por ejemplo, para el caso de las dos variables b18 y b16, este residual será:

$$Residual_{b18b16} = 0.55914 - 0.6353642 = -0.0762242$$

A pesar de que la matriz de pesos factoriales no rotada indica las relaciones entre las variables iniciales y los factores, raramente estos factores pueden interpretarse con facilidad, puesto que suele suceder que algunas variables iniciales están altamente correlacionadas con varios factores.²¹ En nuestro ejemplo, la variable *pluriempleo* tiene correlaciones similares en ambos factores.²²

Para solucionar este problema, se suele efectuar lo que se denomina *rotación de los factores*, que consiste en transformar la matriz factorial inicial en una *matriz factorial rotada*, de más fácil interpretación. Se trata de que cada factor tenga coeficientes de correlación significativos con solo algunas de las variables iniciales, y que cada variable inicial tenga coeficientes de correlación significativos con solo algunos factores, a ser posible con solo uno.

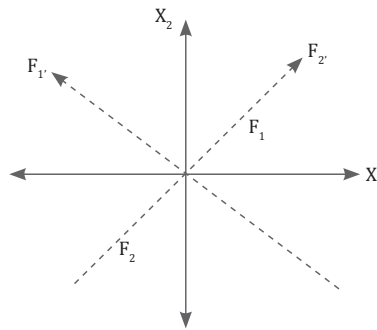
La idea de base de estas rotaciones es muy simple: supongamos que las variables X_1 y X_2 están correlacionadas las dos a la vez con los factores 1 y 2.

Si permaneciendo en el plano definido por los factores F_1 y F_2 , efectuamos una rotación de 45°, las variables X_1 y X_2 aparecen totalmente correlacionadas con el factor F_1 . (eje 1') y nada correlacionadas con el factor F_2 . (eje 2'); lo cual convierte en mucho más fácil la interpretación del factor F_1 . En este caso podríamos decir que el factor F_1 pone de manifiesto la oposición entre las variables X_1 y X_2 .

21. La finalidad de la rotación es ayudarnos a interpretar, en el supuesto de que no quede claro el sentido y significado de los factores de la matriz de pesos factoriales no rotada.

22. En algunos casos pueden presentar altas correlaciones de una variable en todos los factores seleccionados. No olvidar que los valores deben considerarse en valor absoluto.

Figura: Rotación ortogonal



Existen dos tipos de rotaciones: las rotaciones oblicuas y las ortogonales. Las rotaciones oblicuas eliminan la propiedad de independencia de los factores, mientras que las ortogonales las mantienen, siendo estas últimas las más utilizadas.

Dentro de las rotaciones ortogonales,²³ uno de los algoritmos más utilizados es el algoritmo VARIMAX. Este algoritmo trata de minimizar el *número de variables* que hay con pesos o saturaciones elevados en cada factor (las más cercanas a -1 o +1).²⁴

En nuestro ejemplo, si aplicamos una rotación VARIMAX a la matriz factorial no rotada obtenemos el resultado expuesto en la figura siguiente. Efectivamente, observamos que las variables con coeficiente de correlación cercana a uno en el factor uno presentan coeficientes de correlación cercanos a cero en el factor dos, y las variables con coeficientes de correlación cercanos a uno en el factor dos presentan coeficientes de correlación cercanos a cero en el factor uno.

Rotated Component Matrix^a

	Component		
	1	2	3
Ganas de trabajar	.815	-.118	.076
Comodidad	.762	-.065	.083
Preparación	.715	.080	-.074
Búsqueda	.682	.064	.224
Crisis	-.081	.796	-.053
Política de empleo	-.061	.785	.017
Empresarios	.024	.509	.263
Reparto	-.008	.120	.823
Pluriempleo	.205	.023	.777

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a. Rotation Method converged in 5 iterations.

23. Los factores se mantienen incorrelacionados y los ejes forman ángulos rectos.
24. El QUARTIMAX trata de minimizar el número de factores necesarios para explicar un conjunto de variables; el EQUAMAX es una combinación del VARIMAX y el QUARTIMAX. Dentro de los no ortogonales se tienen los de rotación oblicua (OBLIMIN y PROMAX) o no ortogonales, y por tanto existen correlaciones entre factores (no necesariamente iguales a cero).

Al efectuar una rotación hay que tomar en cuenta que el total de la información restituida (en nuestro ejemplo, el 58.3%) permanece constante; pero varía la información restituida por cada uno de los factores, por lo que, en el caso de que sea preciso conocerla, debe recalcularse.

7. Interpretación de la matriz factorial rotada y representación de los resultados

Representación de los resultados

Una vez determinado el número de factores a conservar para el análisis (por medio de cualquiera de las reglas expuestas) también pueden interpretarse los factores obteniendo la representación gráfica de las variables iniciales proyectadas en los planos factoriales (factores tomados dos a dos). En algunos casos también puede ser interesante representar la solución en tres dimensiones.

Las variables iniciales suelen ser representadas en los diferentes planos factoriales, por puntos o vectores que tienen por coordenadas sus coeficientes de correlación con los factores que se estén considerando.

El método utilizado coincidirá a grandes rasgos en la siguiente asignación:

Factor 1

Variables:

- b18 Las pocas ganas de trabajar de la gente.
- b16 La comodidad de la gente, buenos trabajos.
- b17 La falta de preparación del trabajador.
- b19 El no saber buscar trabajo.

Factor 2

Variables:

- b13 La crisis económica.
- b14 La política de empleo del Gobierno.
- b15 La mala gestión de los empresarios.

Factor 3

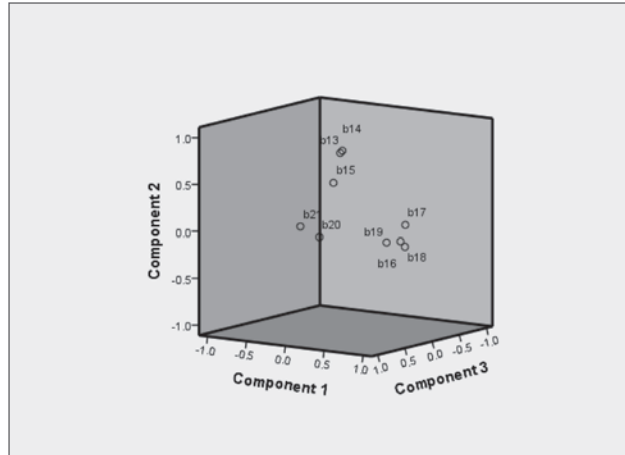
Variables:

- b21 El trabajo que hay no se reparte bien socialmente.
- b20 Que hay mucho pluriempleo.

Analizando estos resultados se podría tratar de tres factores claramente diferenciados y referidos al propio trabajador (el primer factor): el Gobierno y los empresarios (segundo factor) y el reparto o redistribución del trabajo (tercer factor).

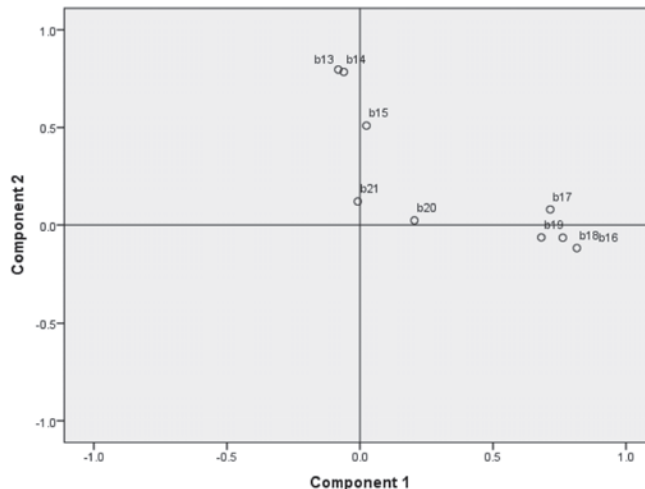
En nuestro caso, la representación de las variables iniciales se presenta en la siguiente figura, que corresponde al gráfico tridimensional de la solución rotada VARIMAX y componentes principales. Este gráfico resulta de seleccionar “Loading PLots” en el subcuadro de diálogo “Rotation”, generándose el gráfico en el que los pesos factoriales actúan como valores en las distintas coordenadas.²⁵

Component Plot in Rotate Space



En el editor del gráfico de tres dimensiones, podemos hacer “clic” derecho sobre el área del gráfico; luego, seleccionar propiedades, haciendo “clic” a la ficha “Variables” y luego excluir el componente Z para lograr una representación de los dos primeros factores.

Component Plot in Rotate Space



25. Si son dos factores extraídos en la solución, el gráfico será bidimensional; pero si son tres o más, el gráfico será tridimensional de los tres primeros factores de la solución.

En cuanto a la calidad de la representación de estas variables, tenemos que ésta es tanto mejor cuanto más lejos estén del centro del eje de coordenadas. En nuestro ejemplo, podemos decir que las variables b18, b16, b19, b20, b15 y b14 están mejor representadas en el plano respecto a b13, b14, b15 y que b20, b21. En el gráfico se observa la formación de 3 grupos de variables. Puede observarse que b16, b17, b18 y b19 están altamente correlacionadas entre sí y a su vez correlacionadas positivamente con el factor 1 (están situadas cercanas al extremo positivo del eje); b13 y b14 lo mismo, pero para el factor 2 (en cambio es negativa la relación con el factor 1).

Las variables mejor representadas son las que se utilizaran para interpretar los ejes. Vemos en nuestro ejemplo que las variables b17, b18, b16, b19 están vinculadas a la noción de “aspectos del propio trabajador” (factor 1). De la misma manera, las variables b13, b14 y b15 están vinculadas a “Gobierno y empresarios” (factor 2); y mientras que las variables b20 y b21 están vinculadas a la noción de “reparto o redistribución del trabajo” (factor 3).

Al igual como hemos representado las variables sobre un plano formado por los factores 1 y 2, se pueden interpretar los factores 1 y 3, 2 y 3 o cualquier combinación binaria de los mismos, si son más de tres en la solución. Los valores de cada variable en las coordenadas corresponden a los pesos factoriales de las mismas en los ejes de cada factor y pueden ser valores comprendidos entre -1 y 1, y obviamente cuanto mayor sea esta coordenada, más contribuyen a la formación del eje. De hecho, la posición de las variables respecto a los ejes es de donde podremos deducir el sentido y el significado de los factores.²⁶

Puntuaciones factoriales

Es aconsejable estimar las puntuaciones factoriales de cada sujeto en la base de datos. Dado que un factor es una combinación lineal de las variables originales:

$$F_j = \sum_{i=1}^p W_{ji} X_i = W_{j1} X_1 + W_{j2} X_2 + \dots + W_{jp} X_p,$$

donde W_i son los coeficientes de las puntuaciones factoriales y p es el número de variables.

Esta misma expresión se puede aplicar a un individuo k determinado así:

$$\hat{F}_{jk} = \sum_{i=1}^p W_{ji} X_{ik},$$

26 En realidad, los planos factoriales están situados en el interior de un círculo de radio unitario, y en ese sentido lo ideal es que los puntos “variables del estudio” no estén concentrados en torno al origen del espacio bidimensional (0,0), sino próximos al borde del círculo o de los ejes factoriales

donde X_{ik} es el valor estandarizado de la variable i para el sujeto k .

W_{ji} es el coeficiente de la puntuación factorial del factor j variable i .

La estimación de las puntuaciones factoriales se generan en el botón "Factor Score". Existen varias formas de generar las puntuaciones factoriales: con el método "Regression" y la opción "Save As Variables", y "Display factor score coefficient matrix", con componentes principales y con rotación VARIMAX se obtiene la tabla siguiente:

Tabla x: Matriz de coeficientes de puntuaciones factoriales

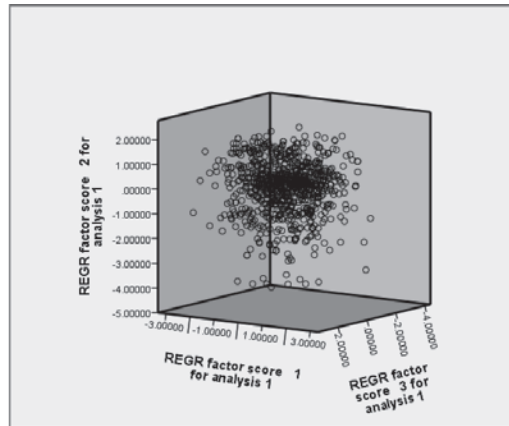
Component Score Coefficient Matrix			
	Component		
	1	2	3
Crisis	.038	.533	-.118
Política de empleo	.037	.519	-.067
Empresarios	.019	.314	.138
Comodidad	.343	.013	-.165
Preparación	.355	.124	-.156
Ganas de trabajar	.364	-.017	-.043
Búsqueda	.285	-.009	.081
Pluriempleo	-.010	-.053	.556
Reparto	-.109	-.012	.610

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Component Scores.

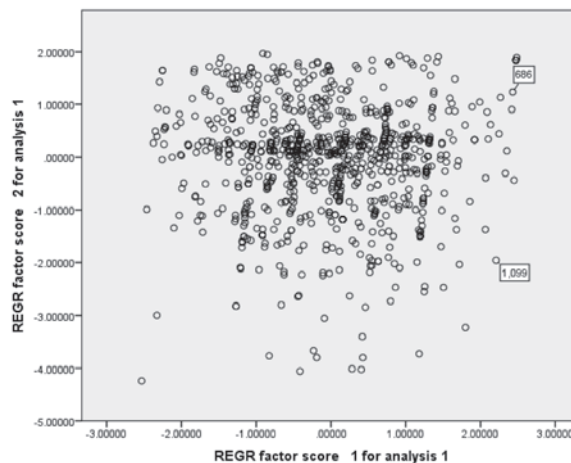
A la vez, en la base de datos se habrán generado tres nuevas columnas correspondientes a las puntuaciones factoriales de cada uno de los tres factores del modelo (ver nuevas variables generadas en la matriz de datos). Estas puntuaciones factoriales tendrán medio igual a cero y desviación estándar igual a uno en todos los casos.

Una vez obtenidas las puntuaciones factoriales de cada sujeto en los tres factores (FAC1_1, FAC2_1, FAC3_1), el análisis factorial termina con la representación gráfica de la posición de cada sujeto en el hiperplano.

Entrar a generador de gráficos y crear el gráfico de dispersión (3-D) con las tres variables generadas.



Una aproximación gráfica que puede permitirnos ubicar a los individuos dentro del espacio formado por los factores aconseja reducir estos a únicamente dos dimensiones. Estando en el gráfico en 3-D, se excluye el factor tres (Eje Z) y se obtiene el gráfico siguiente:



La lectura de este gráfico debe hacerse en relación al gráfico en 2-D, donde se presentaban las variables respecto a los dos factores comunes, donde también se presentaban las variables situadas en el espacio bidimensional. Respecto al primer factor, los sujetos situados más a la derecha han puntuado muy alto aquellas variables que podríamos denominar inherentes al propio “sujeto”; y al revés, con los que están situados muy a la izquierda de este eje, podemos hacer idéntico razonamiento para el segundo de los ejes.

En este sentido, la contribución de un individuo a la formación de un eje o factor es tanto mayor cuanto más alta es la proyección sobre el mismo. Individuos próximos en la dirección del factor uno significa que han dado puntuaciones similares en las variables con cargas factoriales elevadas en el citado factor, y lo mismo para el resto de factores, sin olvidar el peso relativo de cada eje sobre el total.

Interpretación de la matriz factorial rotada

Como hemos señalado, el objetivo de la interpretación de la matriz factorial rotada consiste en identificar cada una de las dimensiones latentes extraídas. Se efectúa escogiendo para cada factor, las variables iniciales cuyas correlaciones con el factor sean las más elevadas (cercanas a +1 o a -1).

En nuestro caso, para cada factor las variables más correlacionadas son estas:

Factor 1

b18	Las pocas ganas de trabajar de la gente	0.815
b16	La comodidad de la gente, buenos trabajos	0.762
b17	La falta de preparación del trabajador	0.715
b19	El no saber buscar trabajo	0.682

Factor 2

b13	La crisis económica	0.796
b14	La política de empleo del Gobierno	0.785
b15	La mala gestión de los empresarios	0.509

Factor 3

b21	El trabajo que hay no se reparte bien socialmente	0.823
b20	Que hay mucho pluriempleo	0.778

A la vista de las variables que constituyen cada uno de los factores, vemos que el factor uno tiene relación con aspectos referentes del propio trabajador; el factor dos, con aspectos referentes a la gestión de los empresarios; y el factor tres, con problemas del entorno económico/Gobierno.

Otra aplicación mediante SPSS

Ya hemos visto que el análisis factorial intenta identificar variables subyacentes o factores que expliquen la configuración de las correlaciones dentro de un conjunto de variables observadas. El análisis factorial se suele utilizar en la reducción de los datos para identificar un pequeño número de factores que expliquen la mayoría de la varianza observada en un número mayor de variables manifiestas. También puede utilizarse para generar hipótesis relacionadas con los mecanismos causales

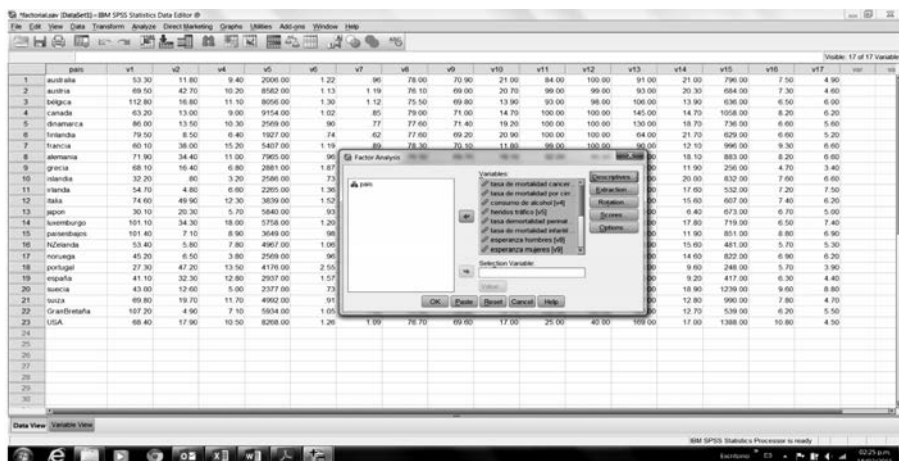
o para inspeccionar las variables para análisis subsiguientes (por ejemplo, para identificar la colinealidad antes de realizar un análisis de regresión lineal).

Con el ejemplo de Análisis factorial vamos a utilizar el anexo *factorial.sav*, el cual contiene los valores de 17 indicadores de tipo sanitario para 23 países. El objetivo es realizar un análisis factorial para encontrar una estructura subyacente en estos datos. Las variables que presenta dicha base son los siguientes:

- V1: tasa de mortalidad cáncer de pulmón.
- V2: tasa de mortalidad por cirrosis hepática.
- V4: consumo de alcohol.
- V5: heridos tráfico.
- V6: tasa de mortalidad perinatal.
- V7: tasa de mortalidad infantil.
- V8: esperanza de vida hombres.
- V9: esperanza de vida mujeres.
- V10: días cama.
- V11: porcentaje de población protegida ambulatorio.
- V12: porcentaje de población protegida hospital.
- V13: población por cama hospitalaria.
- V14: tasa de admisión hospitalaria.
- V15: gasto per cápita en médicos.
- V16: PIB en salud.
- V17: PIB gasto público en salud.

Para realizar un análisis factorial tenemos que realizar los siguientes pasos:

- 1) **Seleccionar Analizar - Reducción de Datos - Factor** y arrastrar todas las variables para darle tratamiento estadístico, así como se muestra en la imagen siguiente:



2) En el botón Descriptivos se podrán obtener:

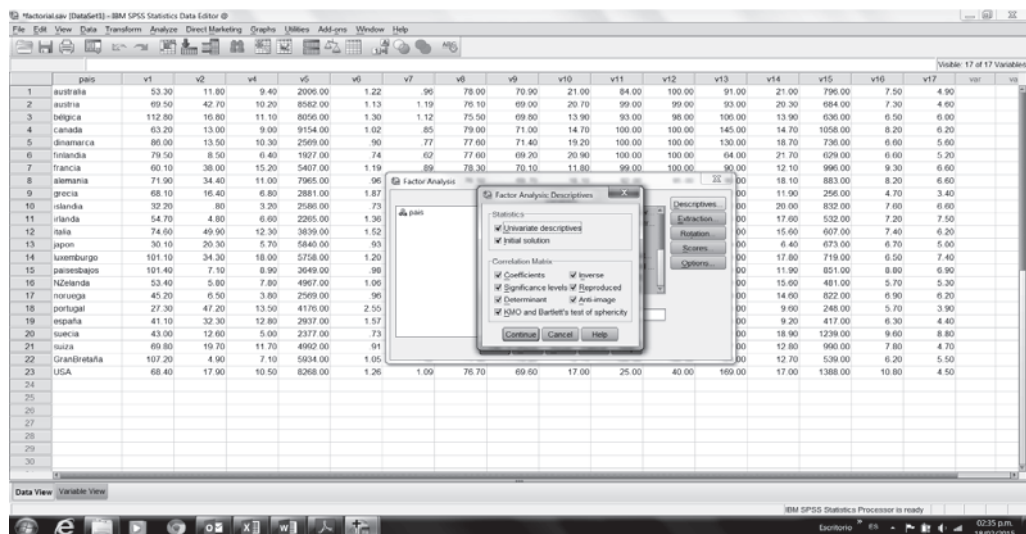
a) **Estadísticos:**

- Descriptivos univariados: la media, la desviación típica y el número de casos válidos para cada variable.
- Solución inicial muestra las communalidades iniciales, los autovalores y el porcentaje de varianza explicada.

b) **Matriz de correlaciones:**

- Coeficientes, niveles de significación y determinante.
- Inversa, reproducida, anti-imagen.
- KMO (media de Kaiser-Meyer-Olkin) y prueba de esfericidad de Barlett.

Así como se muestra en la siguiente imagen:



Al seleccionar todas las opciones, obtendremos los siguientes resultados:

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
tasa de mortalidad cancer de pulmon	65.8217	24.40285	23
tasa de mortalidad por cirrosis hepatica	19.9652	14.75178	23
consumo de alcohol	9.4043	3.64623	23
heridos tráfico	4726.2609	2359.75105	23
tasa demortalidad perinatal	1.1800	.40925	23
tasa de mortalidad infantil	.9917	.29742	23
esperanza hombres	77.4696	1.57637	23
esperanza mujeres	70.7609	1.77117	23
dias cama	15.3522	4.08221	23
porcentaje de poblacion protegida ambulatorio	90.6957	20.06679	23
porcentaje de poblacion protegida hospital	95.7391	12.72155	23
poblacion por cama hospitalaria	113.6957	37.50204	23
tasa de admision hospitalaria	15.3087	4.07653	23
gasto per cápita en médicos	739.6522	280.37793	23
pib en salud	7.3087	1.39281	23
pib gasto publico salud	5.7391	1.27698	23

Con la **matriz de correlaciones**, su significación y el determinante. Es conveniente, como ocurre en este caso, que los coeficientes de la matriz de correlaciones sean significativos en su mayor parte. También es interesante que el determinante de correlaciones sea bajo, 1.35×10^{-10} en este ejemplo:

Correlation Matrix^a

	tasa de mortalidad cancer de pulmon	tasa de mortalidad por cirrosis hepatica	consumo de alcohol	heridos tráfico	tasa demortalidad perinatal	tasa de mortalidad infantil	esperanza hombres	esperanza mujeres	dias cama	porcentaje de poblacion protegida ambulatorio	porcentaje de poblacion protegida hospital	poblacion por cama hospitalaria	tasa de admision hospitalaria	gasto per cápita en médicos	pib en salud	pib gasto publico salud
Correlation	1.000	.004	.319	.311	-.181	.009	-.357	-.246	.152	-.019	-.005	.242	.163	.034	-.016	.141
tasa de mortalidad cancer de pulmon																
tasa de mortalidad por cirrosis hepatica	-.004	1.000	.738	.315	.538	.535	-.338	-.454	-.213	.196	.006	.225	-.211	-.166	-.005	-.201
consumo de alcohol	.319	.738	1.000	.347	.420	.440	-.456	-.585	-.167	.020	-.116	.131	-.161	-.045	.046	-.096
heridos tráfico	.311	.315	.347	1.000	-.022	.190	-.290	-.330	-.114	-.064	-.300	.047	-.122	.297	.259	-.056
tasa demortalidad perinatal	-.181	.538	.420	-.022	1.000	.888	-.495	-.417	-.443	-.100	-.068	.704	-.437	-.595	-.401	-.498
tasa de mortalidad infantil	.009	.535	.440	.190	.888	1.000	-.064	-.577	-.241	-.028	-.058	.559	-.241	-.541	-.398	-.440
esperanza hombres	-.357	-.338	-.456	-.290	-.495	-.064	1.000	.852	-.087	.252	.005	-.169	-.089	.412	.312	.899
esperanza mujeres	-.246	-.454	-.585	-.330	-.417	-.577	.852	1.000	-.203	.156	.079	-.118	-.219	.204	.093	.867
dias cama	.152	-.213	-.167	-.114	-.443	-.241	-.087	-.203	1.000	-.106	.023	-.426	.898	.324	.237	.308
porcentaje de poblacion protegida ambulatorio	-.019	.196	.020	-.064	-.100	-.028	.252	.156	-.106	1.000	.734	-.182	-.113	-.284	-.456	-.031
porcentaje de poblacion protegida hospital	-.005	.006	-.118	-.300	-.068	-.058	.055	.079	.023	.734	1.000	-.366	.016	-.455	-.543	.230
poblacion por cama hospitalaria	-.242	.225	.131	.047	.704	.559	-.169	-.118	-.426	-.182	-.366	1.000	-.432	-.302	-.231	-.550
tasa de admision hospitalaria	.163	-.211	-.161	-.122	-.437	-.241	-.089	-.219	.998	-.113	.018	-.432	1.000	.319	.241	.321
gasto per cápita en médicos	.034	-.166	-.045	.297	-.595	-.541	.412	.204	.324	-.284	-.455	-.302	.319	1.000	.916	.440
pib en salud	-.016	-.005	.046	.258	-.401	-.396	.312	.093	.237	-.456	-.543	-.231	.241	.916	1.000	.443
pib gasto publico salud	.141	-.201	-.096	-.098	-.498	-.446	.099	.067	.338	-.031	.230	-.550	.321	.440	.443	1.000
Big (1-tailed)																
tasa de mortalidad cancer de pulmon		.386	.069	.075	.204	.483	.047	.129	.245	.488	.401	.133	.229	.438	.471	.280
tasa de mortalidad por cirrosis hepatica	.386		.000	.072	.004	.004	.058	.015	.165	.185	.489	.151	.166	.224	.491	.179
consumo de alcohol	.069	.000		.052	.023	.018	.014	.002	.224	.464	.298	.276	.232	.419	.418	.331
heridos tráfico	.075	.072	.052		.460	.192	.089	.062	.302	.386	.082	.416	.289	.085	.117	.329
tasa demortalidad perinatal	.204	.004	.023	.460		.000	.006	.024	.017	.325	.379	.000	.019	.001	.029	.008
tasa de mortalidad infantil	.483	.004	.018	.192	.000		.000	.002	.134	.449	.396	.003	.134	.004	.031	.016
esperanza hombres	.047	.058	.014	.089	.006	.000		.000	.346	.123	.402	.220	.343	.025	.074	.307
esperanza mujeres	.129	.015	.002	.062	.004	.002	.000		.176	.239	.360	.296	.158	.175	.336	.361
dias cama	.245	.165	.224	.302	.017	.134	.346	.176		.315	.459	.021	.000	.006	.138	.057
porcentaje de poblacion protegida ambulatorio	.466	.185	.464	.386	.325	.449	.123	.239	.315		.000	.202	.304	.095	.014	.444
porcentaje de poblacion protegida hospital	.401	.489	.298	.082	.379	.396	.402	.360	.459	.000		.043	.467	.015	.004	.145
poblacion por cama hospitalaria	.133	.151	.276	.416	.000	.003	.220	.296	.021	.202	.043		.020	.061	.144	.003
tasa de admision hospitalaria	.229	.166	.232	.289	.019	.134	.343	.158	.000	.304	.457	.020		.069	.134	.061
gasto per cápita en médicos	.438	.224	.419	.085	.001	.004	.025	.175	.066	.095	.015	.081	.069		.000	.018
pib en salud	.471	.491	.418	.117	.029	.031	.074	.336	.138	.014	.004	.144	.134	.000		.017
pib gasto publico salud	.280	.179	.331	.329	.008	.016	.327	.381	.057	.444	.145	.003	.061	.018	.017	

a. Determinant = 1.35E-010

KMO y prueba de esfericidad de Barlett. Con la prueba KMO contrastamos el hecho de que las correlaciones parciales sean pequeñas. Si el estadístico KMO es grande el procedimiento es adecuado; si es pequeño, será inadecuado. Este estadístico oscila entre cero y uno.

La prueba de Barlett contrasta el hecho de que si la matriz de correlaciones es una matriz de identidad como hipótesis nula, es decir, que las variables están incorreladas; en caso de no rechazarse esta hipótesis, indicaría que el modelo factorial es inadecuado.

Matriz anti-imagen. En un buen modelo factorial la mayoría de los elementos no diagonales deben ser pequeños; y los de la diagonal, grandes. En nuestro ejemplo ocurre esta circunstancia, por lo que el modelo factorial será bueno.

En la imagen siguiente se observa que las pruebas KMO y Bartlett y el estudio de la matriz anti-imagen y la matriz de correlaciones y su significación resultan satisfactorias, por lo que el análisis puede proporcionarnos buenos resultados.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.480
Bartlett's Test of Sphericity	Approx. Chi-Square	359.787
	df	120
	Sig.	.000

3) Extracción

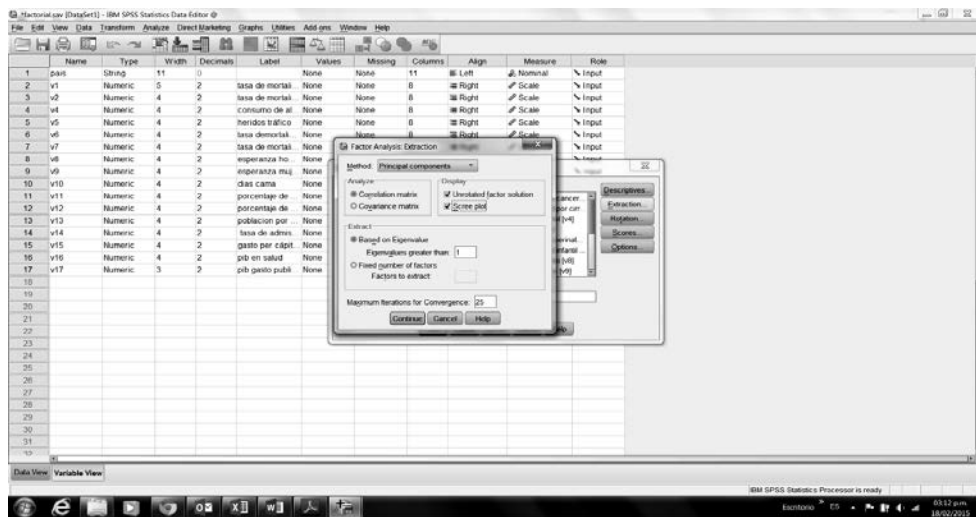
Con este botón el análisis factorial propiamente dicho. Las opciones más importantes son estas:

- a) **Método:** con esta opción seleccionamos el método para extraer los factores. Las opciones disponibles son estas:
- Componentes principales.
 - Mínimos cuadrados no ponderados.
 - Mínimos cuadrados generalizados.
 - Máxima verosimilitud.
 - Factorización de ejes principales.

Nosotros usaremos el método de Componentes principales, que consiste en realizar una combinación lineal de las variables, de tal manera que el primer componente principal sea la combinación que más varianza explique; el segundo, la segunda mayor, y que está incorrelado con el primero, y así sucesivamente.

- a) **Analizar.** Permite especificar o una matriz de correlaciones o una matriz de covarianzas. Utilizaremos la matriz de correlaciones si las variables están en diferente escala y la covarianza en caso contrario.
- b) **Extraer.** Podemos elegir entre utilizar tantos factores como autovalores mayores de un valor. Se conoce como regla de Kaiser la opción de quedarse con tantos factores como autovalores mayores que uno existan. También podemos indicar un número concreto de factores a extraer.
- c) **Mostrar:** ofrece la solución sin rotar y el gráfico de sedimentación.

Así como se muestra en la imagen siguiente:



En este ejemplo, por lo tanto, usaremos la matriz de correlaciones, el método de componentes principales y usaremos la ya comentada regla de Kaiser, es decir, nos quedaremos con tantos factores como autovalores mayores que uno. Los resultados son estos:

Las comunidades: son la proporción de varianza explicada por los factores de cada una de las variables. Vemos que todas las variables van a ser bien explicadas por la estructura factorial, salvo quizás la variable “Heridos de Tráfico”. Ver la imagen siguiente:

Communalities

	Initial	Extraction
tasa de mortalidad cancer de pulmon	1.000	.826
tasa de mortalidad por cirrosis hepatica	1.000	.899
consumo de alcohol	1.000	.782
heridos tráfico	1.000	.680
tasa de mortalidad perinatal	1.000	.945
tasa de mortalidad infantil	1.000	.861
esperanza hombres	1.000	.911
esperanza mujeres	1.000	.854
días cama	1.000	.959
porcentaje de poblacion protegida ambulatorio	1.000	.839
porcentaje de poblacion protegida hospital	1.000	.904
poblacion por cama hospitalaria	1.000	.744
tasa de admision hospitalaria	1.000	.964
gasto per cápita en médicos	1.000	.906

Extraction Method: Principal Component Analysis.

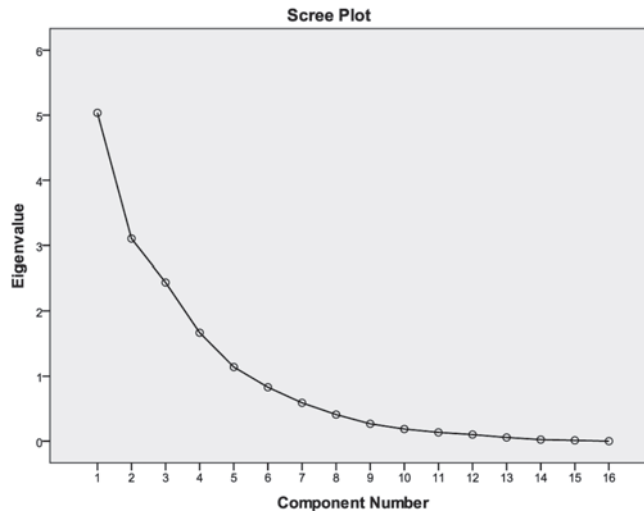
Con la tabla de Varianza Explicada obtenemos una tabla con todos los autovalores y el porcentaje de varianza explicada por cada uno de ellos. En principio tendremos tantos autovalores como variables haya. Mediante la regla de Kaiser tomamos tantos factores como autovalores mayores que uno, en este caso tomaríamos cinco factores. Estos cinco factores explican un 86.23%, que nos parece suficiente como para quedarnos con esos cinco factores, y si nos parece poco aumentaríamos algún factor, y si queremos un modelo más sencillo eliminaríamos algún factor. Como ya vimos, en problemas de tipo socioeconómico es suficiente con una explicación del 60-70%, por lo que sería interesante para simplificar el problema quedarnos con una estructura de cinco factores; para ello, en el botón extraer indicariamos el número concreto que nos interese. Ver la siguiente imagen:

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.563	32.594	32.594	4.563	32.594	32.594	3.476	24.829	24.829
2	2.780	19.854	52.448	2.780	19.854	52.448	2.539	18.133	42.962
3	2.089	14.921	67.369	2.089	14.921	67.369	2.356	16.827	59.789
4	1.595	11.390	78.758	1.595	11.390	78.758	2.151	15.362	75.151
5	1.045	7.467	86.225	1.045	7.467	86.225	1.550	11.074	86.225
6	.625	4.462	90.687						
7	.514	3.670	94.358						
8	.261	1.864	96.222						
9	.192	1.373	97.595						
10	.143	1.019	98.614						
11	.106	.754	99.368						
12	.065	.467	99.835						
13	.022	.159	99.994						
14	.001	.006	100.000						

Extraction Method: Principal Component Analysis.

El gráfico de sedimentación es la representación gráfica de los autovalores, de tal manera que se suele usar para decidir el número de factores a utilizar; tomando factores hasta que la pendiente de la recta sea paralela al Eje X. Veamos el gráfico siguiente:



Por lo tanto, a partir del gráfico de sedimentación y principalmente del estudio de la varianza, nos quedaremos con una estructura factorial con cuatro o cinco factores. Dependiendo de esta decisión las comunales serán estas:

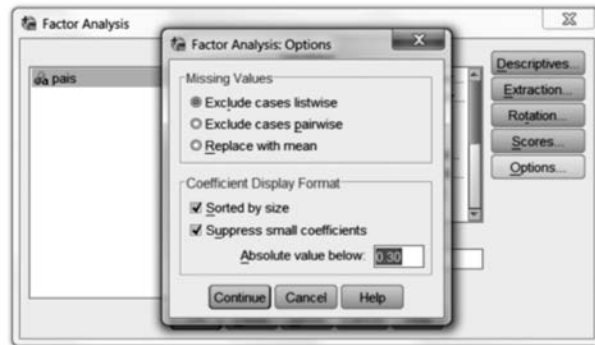
Communalities			Communalities		
	Initial	Extraction		Initial	Extraction
tasa de mortalidad cancer de pulmon	1.000	.411	tasa de mortalidad cancer de pulmon	1.000	.826
tasa de mortalidad por cirrosis hepatica	1.000	.633	tasa de mortalidad por cirrosis hepatica	1.000	.899
consumo de alcohol	1.000	.746	consumo de alcohol	1.000	.782
heridos tráfico	1.000	.673	heridos tráfico	1.000	.680
tasa de mortalidad perinatal	1.000	.938	tasa de mortalidad perinatal	1.000	.945
tasa de mortalidad infantil	1.000	.860	tasa de mortalidad infantil	1.000	.861
esperanza hombres	1.000	.836	esperanza hombres	1.000	.911
esperanza mujeres	1.000	.852	esperanza mujeres	1.000	.854
dias cama	1.000	.896	dias cama	1.000	.959
porcentaje de poblacion protegida ambulatorio	1.000	.819	porcentaje de poblacion protegida ambulatorio	1.000	.839
porcentaje de poblacion protegida hospital	1.000	.904	porcentaje de poblacion protegida hospital	1.000	.904
poblacion por cama hospitalaria	1.000	.740	poblacion por cama hospitalaria	1.000	.744
tasa de admision hospitalaria	1.000	.905	tasa de admision hospitalaria	1.000	.964
gasto per cápita en médicos	1.000	.812	gasto per cápita en médicos	1.000	.906

Extraction Method: Principal Component Analysis.

Extraction Method: Principal Component Analysis.

Evidentemente, con cinco factores las comunales son más altas, el problema está en que este modelo es más complejo que con cuatro factores, por lo que habrá que decidir entre estas dos estructuras, que en nuestro ejemplo vamos a optar por una estructura más simple con cuatro factores.

Una vez decidido el número de factores obtendremos la solución final que es la **matriz de componentes**; pero con una pequeña restricción. La matriz de componentes (también conocida como Factor Scores) comprende las cargas de cada variable en cada uno de los factores originados, de modo que las variables con cargas más altas en un factor, nos indica una estrecha relación entre las variables y ese factor; en definitiva, es la correlación entre el factor y la variable. Por ello, sólo nos interesarán aquellos valores que tengan altas cargas factoriales, y para obtener estos valores seleccionaremos el botón de opciones y seleccionaremos las pestañas de ordenar valores y suprimir aquellas con un valor menor de 0.30, así como se muestra en la imagen siguiente:



El problema está en que, para la interpretación, sería interesante que una variable tuviera una importante carga factorial para un solo factor y no dependiese de varios. Para intentar que se dé esta situación se recurre a las rotaciones factoriales.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.563	32.594	32.594	4.563	32.594	32.594
2	2.780	19.854	52.448	2.780	19.854	52.448
3	2.089	14.921	67.369	2.089	14.921	67.369
4	1.595	11.390	78.758	1.595	11.390	78.758
5	1.045	7.467	86.225			
6	.625	4.462	90.687			
7	.514	3.670	94.358			
8	.261	1.864	96.222			
9	.192	1.373	97.595			
10	.143	1.019	98.614			
11	.106	.754	99.368			
12	.065	.467	99.835			
13	.022	.159	99.994			
14	.001	.006	100.000			

Extraction Method: Principal Component Analysis.

4) Rotación

El fin de la rotación es ayudar a interpretar en los casos en los que las cargas factoriales no estén claras; las rotaciones simplemente son movimientos de los ejes, de tal manera que se mantenga la comunalidad y el porcentaje de varianza explicada por el modelo, aunque no por cada eje. Lo que buscaremos será que cada factor tenga unos pocos pesos altos y el resto, cercanos a cero, y que cada variable esté saturada en solo un factor.

Para realizar una rotación elegiremos el botón rotación, el cual nos permite escoger entre:

- a) **Método:** para elegir el método de rotación entre: Varimax, Equamax, Quartimax, Oblimin directo y Promax.
- b) **Mostrar:**
 - Solución rotada: se selecciona un método de rotación para obtener la solución rotada.
 - Gráficos de saturaciones: muestra las representaciones tridimensionales de las saturaciones para los tres primeros factores. Para una solución de dos factores, se representa un diagrama bidimensional. No se muestra el gráfico si sólo se extrae un factor.

En este caso vamos a elegir una de las rotaciones que SPSS nos ofrece, y concretamente la VARIMAX. Además, marcaremos la solución rotada y el gráfico de saturaciones. Si vemos la tabla de la varianza explicada, observamos que no ha cambiado la varianza total explicada (78.75%), aunque sí lo que explica cada factor, 32.59% en el primer factor y sin rotar, frente a 32.59% rotado; 19.85 frente a 19.85 del segundo factor y 14.92 frente a 14.92 del tercer factor. Ver imagen siguiente:

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.563	32.594	32.594	4.563	32.594	32.594	3.780	27.002	27.002
2	2.780	19.854	52.448	2.780	19.854	52.448	2.755	19.682	46.683
3	2.089	14.921	67.369	2.089	14.921	67.369	2.362	16.871	63.554
4	1.595	11.390	78.758	1.595	11.390	78.758	2.129	15.204	78.758
5	1.045	7.467	86.225						
6	.625	4.462	90.687						
7	.514	3.670	94.358						
8	.261	1.864	96.222						
9	.192	1.373	97.595						
10	.143	1.019	98.614						
11	.106	.754	99.368						
12	.065	.467	99.835						
13	.022	.159	99.994						
14	.001	.006	100.000						

Extraction Method: Principal Component Analysis.

También vemos que las soluciones rotadas se han simplificado:

Component Matrix^a

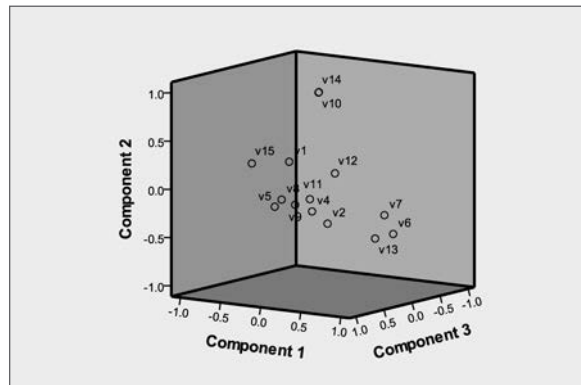
	Component			
	1	2	3	4
tasa de mortalidad cancer de pulmon		.534		.322
tasa de mortalidad por cirrosis hepatica	.692			.360
consumo de alcohol	.662	.336		.437
heridos tráfico		.404		.587
tasa demortalidad perinatal	.893			
tasa de mortalidad infantil	.896			
esperanza hombres	-.701	-.511		
esperanza mujeres	-.650	-.615		
días cama	-.437	.706		-.369
porcentaje de poblacion protegida ambulatorio		-.362	.718	.411
porcentaje de poblacion protegida hospital		-.306	.889	
poblacion por cama hospitalaria	.604	-.342	-.422	
tasa de admision hospitalaria	-.433	.712		-.372
gasto per cápita en médicos	-.529	.402	-.511	.330

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

El primer factor viene determinado por las variables que están más correlacionadas y son tasa de mortalidad por cirrosis hepática, consumo de alcohol, tasa de mortalidad perinatal, tasa de mortalidad infantil, esperanza hombres, mujeres, día camas, población por cama hospitalaria, tasa de admisión hospitalaria y gasto *per cápita* en médicos. De igual manera, el segundo factor viene dado por tasa de mortalidad cáncer de pulmón, heridos tráfico, porcentaje de población protegida ambulatorio, porcentaje de población protegida hospital. Y así sucesivamente, hasta conformar todos los factores. Es importante tener presente que el nombre de cada factor es proporcionado por el investigador, según las correlaciones de las variables que lo explican. No obstante, existirán variables que estarán presentes en todos los factores y en otros no. Para decidir qué variable debería de estar presente en un factor, es necesario graficar los factores en 3D según dónde aparecen las variables. Por ejemplo, vamos a graficar el factor uno en el eje X, factor dos en el eje Y, y factor tres en el eje Z. Ver imagen siguiente:

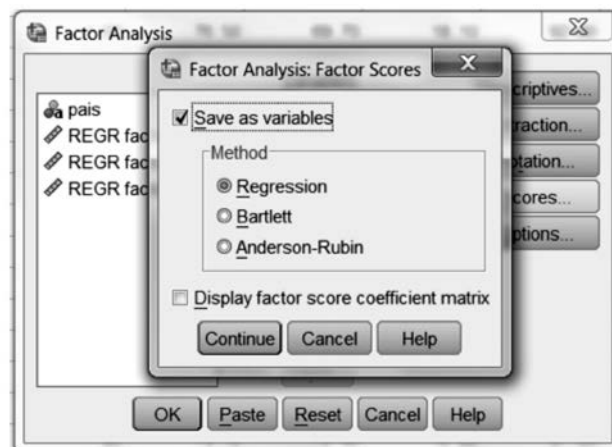
Component Plot in Rotated Space



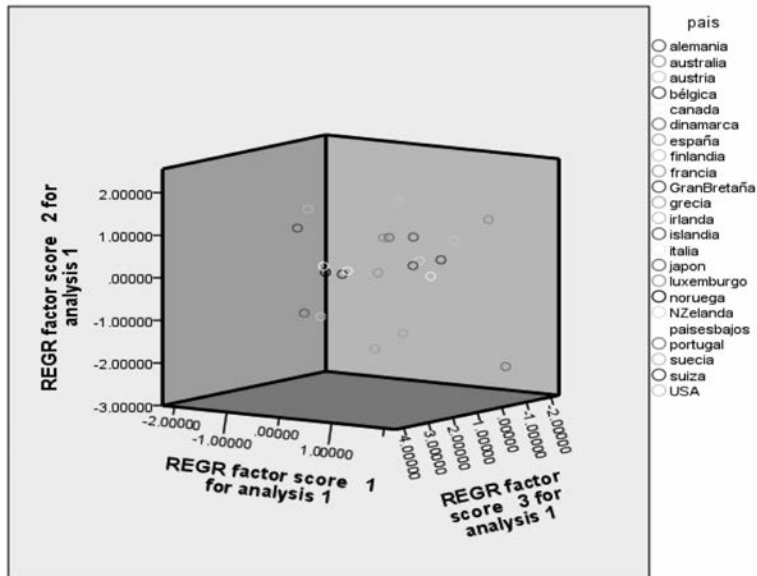
Por medio de los gráficos podemos deducir el sentido y el significado de los factores. A continuación vemos los gráficos bidimensionales y tridimensionales. Estos gráficos son las coordenadas de cada variable con respecto a los factores, que como hemos visto son valores entre -1 y 1, siendo con respecto a los ejes es donde se puede deducir el sentido y el significado de los factores. Lo ideal es que no estén centrados en torno al punto (0,0), sino próximos al borde del gráfico, ya que la idea es identificar las variables con mayor peso factorial. Por ejemplo, con respecto a las componentes uno y dos, vemos que las variables Esperanza de vida masculina y femenina están muy juntas, por lo que están muy correladas entre sí y a su vez con el factor uno negativamente, y Consumo de alcohol y cirrosis hepática lo están entre sí y con el factor uno positivamente, por lo que podríamos concluir que el primer factor contrapone la esperanza de vida con el consumo de alcohol.

5) Puntuaciones factoriales

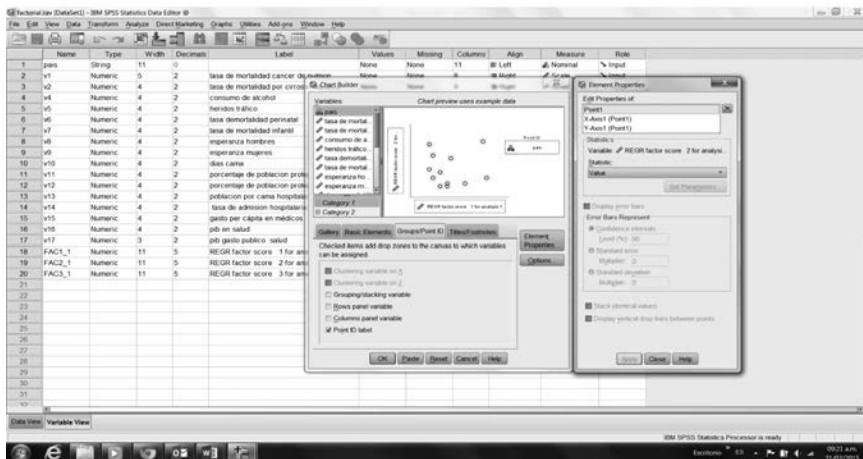
Finalmente, si escogemos el botón Puntuaciones y seleccionamos Guardar como variables, ver imagen siguiente:

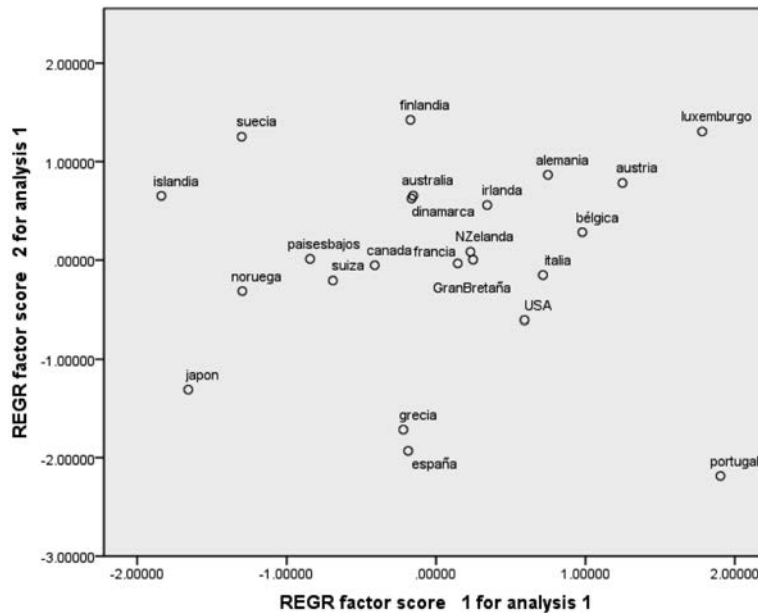


Obtenemos las puntuaciones factoriales, que consisten en reducir el conjunto de variables de cada individuo al conjunto de factores o, dicho en modo matemático, las puntuaciones de cada individuo sobre el conjunto de factores extraído. Estas puntuaciones se añaden en el fichero de datos, con los nombres FAC1_1, FAC1_2, FAC1_3. De tal manera que solo tendremos que representar estas tres variables (Menú Gráficos – dispersión- dispersión 3D-, luego introducir las tres variables nuevas en los ejes y establecer marcas por país). Los resultados se presentan en el siguiente gráfico:



Para obtener gráficos de saturaciones, en la siguiente imagen se muestra la selección para generar dicho gráfico:





4.6. Ejercicios

Ejercicio 1

Una empresa especializada en el diseño de automóviles de turismo desea estudiar cuáles son los deseos del público que compra automóviles. Para ello diseña una encuesta con 10 preguntas, donde se les solicita a cada uno de los 20 encuestados que valore de uno a cinco si una característica es o no muy importante. Los encuestados deberán contestar con un cinco si la característica es muy importante, un cuatro si es importante, un tres si tiene regular importancia, un dos si es poco importante, y un uno si no es nada importante. Las 10 características (v1 a v10) a valorar son: precio, financiación, consumo, combustible, seguridad, comodidad, capacidad, prestaciones, modernidad y aerodinámica. El anexo factorial2.sav, recoge los datos. Realizar un análisis factorial que permita extraer unos factores adecuados a los datos que resuman correctamente la información que contienen.

Ejercicio 2

Para estudiar las ventas de las empresas españolas se consideran siete variables ratio R1 y R7. Concretamente los ratios son beneficios/recursos propios (R1), cash-flow/ventas (R2), inmovilizado/activos totales (R3), ventas/activos totales (R4), ventas/planilla (R5), beneficios/capital social (R6) y beneficios/ventas (R7), que caracterizan a las empresas españolas con mayores ventas. Se trata de

resumir estos ratios por un número menor de factores con mínima pérdida de información que tengan la suficiente calidad como para seguir agrupando a las empresas según sus ventas. ¿Sería coherente identificar un factor financiero, un factor estructural y un factor de rentabilidad? El anexo factorial3.sav contiene la información de las variables.

Ejercicio 3

Con los datos de índice de bienestar (anexo indice bienestar.sav). Se solicita:

- a) Realizar un análisis factorial para reducir la dimensión.
- b) Determinar cuáles serían los factores y que variables son las que explican dichos factores.
- c) Realizar el gráfico de saturaciones y luego analizar para dar sus conclusiones.

Ejercicio 4

Con los datos de las elecciones (elecciones.sav). Se solicita

- a) Realizar un análisis factorial y nombrar según las variables que explican dicho factor.
- b) Analizar el gráfico de saturaciones para el factor uno y el factor dos.

Ejercicio 5

El Departamento de Calidad de Servicio de la empresa INTCOMEX de El Salvador, una empresa que vende material informático, desea diseñar una herramienta permanente de medición de la satisfacción de sus clientes con la calidad del servicio que ofrece. El objetivo es efectuar semanalmente un número importante de encuestas telefónicas, mediante un cuestionario que tuviese en cuenta las dos premisas siguientes:

- i) Debía cumplir los principales aspectos que definen la calidad de servicio desde la perspectiva de los clientes.
- ii) Tenía que ser corto para no cansar a los clientes y para poder cumplir con los objetivos diarios de número de encuestas a realizar.

Consciente de que la única forma de descubrir los aspectos fundamentales del servicio era hablando directamente con los clientes, el departamento de calidad

llevó a cabo una investigación que le permitió descubrir cuál era la estructura latente que, desde la perspectiva del cliente, definía la calidad del servicio de la empresa.

En la primera fase de la investigación de carácter exploratorio se elaboró, a partir de grupos focales con los clientes de la empresa, una lista de 22 aspectos que, desde su punto de vista, definían la calidad del servicio.

Los aspectos que se identificaron como determinantes de la calidad del servicio fueron estos:

- A1 Trato recibido del personal
- A2 Capacidad del personal para resolver problemas
- A3 Aspecto del personal
- A4 Profesionalidad/eficacia del personal
- A5 Aspecto/apariencia interior del punto de venta
- A6 Aspecto exterior del punto de venta
- A7 Claridad de la información
- A8 Novedad/actualidad de la información recibida
- A9 Calidad de la información
- A10 Interés de los folletos informativos en el punto de venta
- A11 Tiempo de espera para ser atendido
- A12 Tiempo de espera para resolver una consulta
- A13 Información sobre ofertas/ promociones especiales
- A14 Precio del *hardware*
- A15 Precio del *software*
- A16 Precio de los consumibles
- A17 Descuentos/bonos ofrecidos
- A18 Rapidez reparación de averías del servicio posventa
- A19 Garantía del servicio posventa
- A20 Asesoramiento recibido sobre nuevo *hardware*
- A21 Asesoramiento recibido sobre nuevo *software*
- A22 Asesoramiento recibido sobre formación y cursillos

En la segunda fase de tipo descriptivo se realizó una encuesta telefónica a una muestra representativa de los clientes, en la que éstos manifestaron su nivel de satisfacción en una escala de cero a 10, donde cero significó “Nada satisfecho”; y 10, “Totalmente satisfecho”, en cada uno de los 22 aspectos.

Efectuar un análisis multivariado descriptivo de esta encuesta (quality_serv_Intx.sav) y aplicar un análisis factorial de componentes principales a los datos obtenidos; es decir, determinar la existencia de una estructura latente dentro de la información analizada, y en el caso de que existiera dicha estructura, identificar el número de variables latentes e interpretarlas.

5. Análisis simple de correspondencias

5.1. Métodos de Reducción: Análisis de correspondencias

El método de Análisis de correspondencias,²⁷ que en adelante denominaremos ACOR, es uno de los más recientes. Sus orígenes se sitúan en 1963 y se debe a J.P. Benzecri, de la Universidad de París.

Existen varias formas de presentar el ACOR. Quizás la más correcta y comprensible sea decir que este tipo de análisis no es más que un caso particular del análisis factorial de componentes principales (ACP), diferenciándose básicamente en los siguientes aspectos:

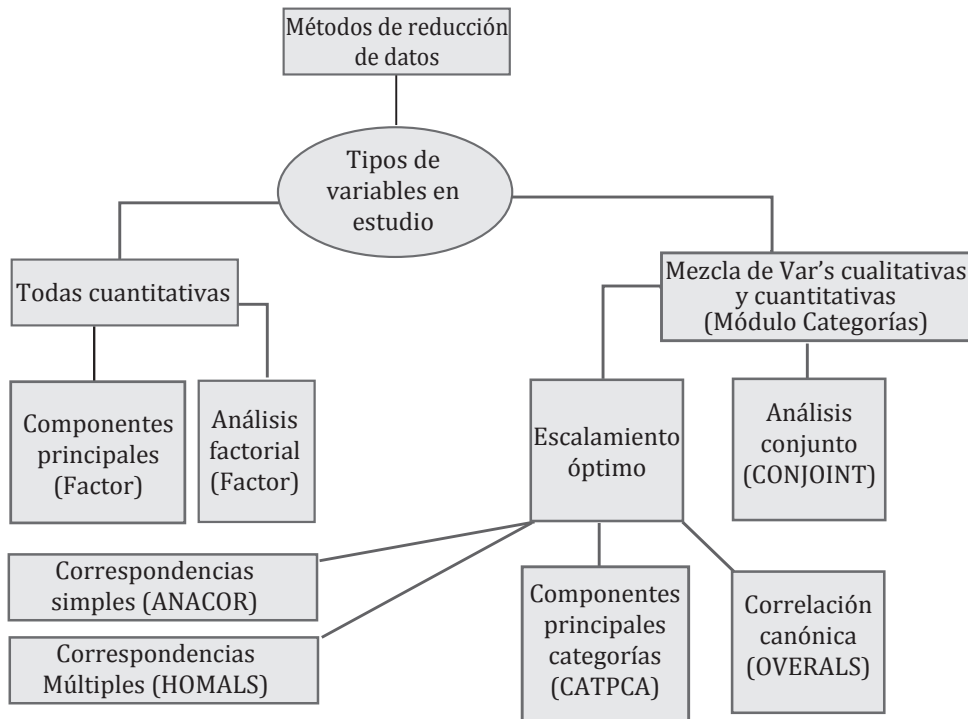
- Mientras que el ACP trata de resumir el número de variables que intervienen en un análisis a través de la construcción de nuevas variables (factores) más sintéticas, el ACOR, *trata de analizar las formas que adoptan las relaciones entre las variables*; puesto que su objetivo es descubrir las dimensiones estructurales que revelan un conjunto de datos. De manera general se dice que el ACOR está orientado especialmente hacia el análisis geométrico de la dependencia entre variables.
- El ACOR permite analizar cualquier matriz de números no negativos, es decir, positivos o nulos (matrices disyuntivas, matrices de datos ordinales, etc.), independientemente de su escala de medida. Por lo tanto, es apropiado para el tratamiento de variables cualitativas, esencialmente nominales.
- Su objetivo básico es analizar las relaciones entre dos conjuntos de variables (habitualmente a partir de matrices de valores absolutos), uno de los cuales suele ser, en investigación de mercados, individuos productos, empresas, marcas) y el otro suele ser atributos/características.²⁸

27. Algunas veces denominado Análisis factorial de correspondencias.

28. Los sujetos de la muestra suelen estar constituidos por segmentos de mercado, individuos agrupados según criterios diversos, partidos políticos, sectores económicos, especies biológicas, personas físicas individuales, etc.; en tanto que las variables pueden ser dos o más y medidas habitualmente en

- Permite la presentación simultánea de los dos conjuntos de variables en un mismo espacio. Algunos autores han definido el ACOR como “Un modo de representación gráfica de las matrices de valores absolutos, que trata de evidenciar en uno o varios gráficos la mayor cantidad posible de información contenida en una matriz”. Para ello no se fija en los valores absolutos, sino en las correspondencias entre las características, es decir en los valores relativos.

A continuación se presenta una clasificación de los métodos de reducción de la dimensión:



Fuente: Elaboración propia.

El análisis de correspondencias es una técnica de reducción de dimensión y elaboración de mapas perceptuales. *Los mapas perceptuales se basan en la asociación entre objetos y un conjunto de características descriptivas o atributos especificados por el investigador.* Su aplicación es muy directa y su beneficio es la capacidad para representar filas y columnas en un mismo espacio. En este sentido

escalas nominales u ordinales. Los gráficos de alta resolución de una, dos o tres dimensiones permitirán situar los objetos/sujetos de la muestra y las variables con sus distintas categorías de modo conjunto o por separado

la finalidad será determinar la posición de una serie de objetos según una serie de características a través de un espacio vectorial en dos, tres o más dimensiones.

En su forma básica, el Análisis de correspondencias examina las relaciones entre categorías de datos nominales mediante la medida de asociación de la chi-cuadrado. Este análisis realiza una reducción de la dimensión del problema, en donde la proximidad, en sentido matemático, entre los individuos, indicará el nivel de asociación.

El Análisis de correspondencias tiene dos objetivos básicos:

- Asociación entre categorías de columnas o filas: se usa para medir la asociación de solo una fila o columna, para ver, por ejemplo, si las modalidades de una variable pueden ser combinadas.
- Asociación entre categorías de filas y columnas: se usa para estudiar si existe relación entre categorías en las filas y columnas.

El Análisis de correspondencias sólo requiere que los datos representen las respuestas a una serie de preguntas y que estén organizadas en categorías. Dependiendo de que si existen dos o más variables, el análisis será simple o múltiple.

Uno de los objetivos del Análisis de correspondencias es describir las relaciones existentes entre dos variables nominales, recogidas en una tabla de correspondencias, sobre un espacio de pocas dimensiones; mientras que al mismo tiempo se describen las relaciones entre las categorías de cada variable. Para cada variable, las distancias sobre un gráfico, entre los puntos de categorías, reflejan las relaciones entre las categorías, con las categorías similares representadas próximas unas a otras.

El análisis de las Tablas de contingencia a menudo incluye examinar los perfiles de fila y de columna, así como contrastar la independencia a través del estadístico de chi-cuadrado. Sin embargo, el número de perfiles puede ser bastante grande y la prueba de chi-cuadrado no revelará la estructura de la dependencia. El procedimiento Tablas de contingencia ofrece varias medidas y pruebas de asociación, pero *no puede representar gráficamente ninguna relación entre las variables*.

Supongamos que la información de un estudio sobre el mercado de computadoras personales, donde se entrevistó a una muestra de clientes de nueve empresas de distribución, con el fin de determinar el posicionamiento de dichas empresas respecto a algunos atributos que habían sido definidos previamente y donde cada cliente tuvo que asociar cada uno de los atributos a una o más empresas.

Atributo	E1	E2		E9	Totales fila
Precios/descuentos	16	17		18	183
Variedad de marcas	8	15		53	174
Rapidez entrega	20	20		25	195
Trato de personal	11	13		44	158
.....					
Totales Columna	125	132		211	1318

Una matriz de este tipo se hace difícil de interpretar (sobre todo en el caso de que contenga un número amplio de filas; (en nuestro ejemplo, atributos) o de las columnas (empresas), por ejemplo:

- Si nos encontramos con un valor muy alto en relación con los demás valores de la misma columna, en nuestra matriz, por ejemplo, el valor 53 en la intersección “empresa 9” y “variedad de marcas”, podemos pensar que esta empresa se caracteriza por una buena “asociación” con este atributo en relación a los demás atributos. Pero ello solamente será cierto si el total de la fila no es a su vez muy alto en relación con los demás valores totales de cada fila (totales fila) del resto de los atributos; en caso contrario, no podemos hacer ninguna conclusión sobre la empresa 9, puesto que será el atributo, que de manera general, se beneficia de las fuertes asociaciones. En nuestro caso, el total de asociaciones del atributo “variedad de marcas” con las distintas empresas es de 174, siendo el segundo atributo, con un total de asociaciones a empresas, más bajo. El análisis de los “totales fila” permite afirmar que la “empresa 9” se asocia fuertemente al atributo “variedad de marcas”.
- Analizando el mismo valor 53, esta vez respecto a los demás valores de la misma fila, observamos que es la asociación más alta. Podemos pensar que la “empresa 9” se diferencia de las demás por una “mejor asociación” en el atributo “variedad de marcas”; pero al igual que en el caso anterior, ello solamente sería cierto en el caso de que el total 211 (total columna) no fuera muy elevado en relación con los otros totales de las otras columnas; ya que, en caso contrario, será la empresa la que tendrá, en general, fuertes asociaciones, en el conjunto de los atributos. En resumen, es la “empresa 9” la que obtiene, en general, asociaciones elevadas en el conjunto de los atributos.

Un análisis visual no es lo adecuado, especialmente cuando tenemos una matriz con suficientes datos; en ese contexto, y para facilitar la interpretación que hemos señalado, se hace indispensable efectuar un tratamiento del conjunto de datos de la matriz de datos que tenga en consideración los siguientes aspectos:

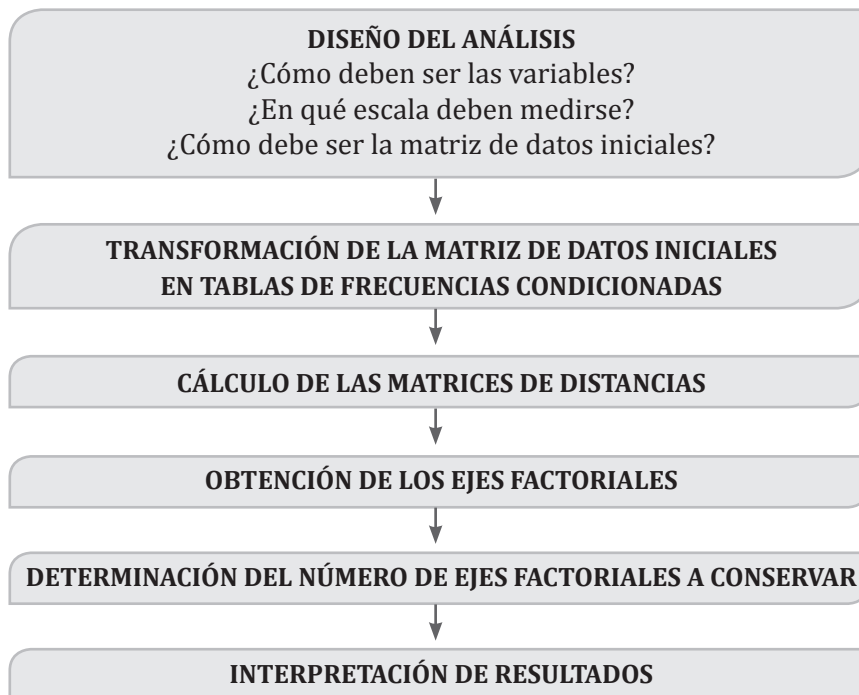
- Los valores calculados en porcentajes según las filas.
- Los valores calculados en porcentajes según las columnas.
- El peso de cada fila y de cada columna.
- Que además reduzca el número de atributos intercorrelacionados a un número más pequeño de “dimensiones independientes”.
- Que establezca las semejanzas entre filas y columnas.

Un tratamiento de este tipo es el que nos permite efectuar el *análisis de correspondencias*.

5.2. Metodología del análisis de correspondencias

La metodología del análisis de correspondencia se puede esquematizar tal como se muestra en la figura siguiente:

Esquema metodológico del análisis de correspondencias



Etapla 1. Diseño del análisis de correspondencias

La principal característica diferenciadora del ACOR con respecto a las demás técnicas de interdependencia es su capacidad para trabajar con escalas cualitativas, es decir, nominales u ordinales.

El ACOR habitualmente trabaja a partir de matrices de valores absolutos. Los datos de base corresponden a una matriz de valores absolutos entre dos variables: I (atributos) y J (productos o marcas, empresas, etc.), en las cuales el conjunto de las filas es de I modalidades y el de las columnas es de J modalidades; es decir, los datos de partida se presentan, pues, bajo la forma de una matriz $I \times J$.

Variable J Variable I	1		j	J
1				
.....				
i			n_{ij}	
.....				
I				

- El análisis es simétrico puesto que las dos variables I y J juegan papeles que son intercambiables. El ACOR pretende estudiar la relación entre los dos conjuntos, encontrando grupos de filas (atributos) similares entre sí por estar asociados o no a las mismas columnas (empresas) y, análogamente, desde el punto de vista de las columnas (empresas) buscará grupos de ellas que se parezcan por estar asociadas o no a determinadas filas (atributos).

Etapla 2. Transformación de la matriz de datos iniciales en tablas de frecuencias condicionadas

Dado que lo que queremos hacer es comparar los perfiles y no las cifras en bruto de la matriz inicial, deberemos efectuar algunas modificaciones antes de realzar el análisis.

Tenemos que los datos iniciales son los valores n_{ij} de la matriz de partida,

$$i \in I \quad i= 1,2,3,...,I \text{ (atributos)}$$

$$j \in J \quad j= 1,2,3,...,J \text{ (empresas)}$$

En donde:

$$n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$$

Es el número de personas entrevistadas en el caso de que cada individuo únicamente contestase a una sola modalidad para las variables I y J , es decir, asociarse un solo atributo a una sola empresa.

En el caso que nos ocupa, cada individuo tiene la posibilidad de asociar más de un atributo a cada empresa, por lo que n se interpreta como el número total de asociaciones de atributos a empresas, y el número de asociaciones del atributo i a la empresa j .

Si denominamos:

$$n_i = \sum_{j=1}^J n_{ij} \text{ número total de asociaciones al atributo } i$$

$$n_j = \sum_{i=1}^I n_{ij} \text{ número total de asociaciones a la empresa } j$$

El perfil de asociación del atributo i del I con el conjunto J de empresas es un vector X_i en el cual sus componentes son las divididos por n_i

$$X_i = \left[\frac{n_{i1}}{n_i}, \frac{n_{i2}}{n_i}, \dots, \frac{n_{iJ}}{n_i} \right]$$

Siendo X_i un elemento de R^J

De igual forma se define el perfil de asociación de la empresa j de J con el conjunto I de atributos:

$$Y_j = \left[\begin{array}{c} \frac{n_{1j}}{n_j} \\ \vdots \\ \frac{n_{Ij}}{n_j} \end{array} \right]$$

Siendo Y_j un elemento de R^I .

Se obtienen así dos nubes de puntos: una del conjunto I de atributos representados por R^J y la otra del conjunto J de empresas representadas en R^I .

En el análisis de correspondencias cada elemento estará afectado por su probabilidad de aparición en el proceso. Ello dará lugar a dos tablas: una que representará al conjunto I condicionada por J y otra que representará al conjunto J condicionada por I .

Las frecuencias condicionadas de la variable I por la variable J (X_i) se obtienen dividiendo cada uno de los valores por el total fila. Las frecuencias condicionadas de la variable J por la variable I (Y_j) se obtienen dividiendo cada uno de los valores por el total columna.

Se obtienen así dos nubes de puntos; en el ACOR cada categoría estará afectada por un peso determinado: X_i estará afectado por el peso del total fila; mientras que Y_j estará afectado por el peso del total columna.

Etapa 3. Cálculo de las matrices de distancia

A partir de las dos tablas de frecuencias condicionadas, el ACOR calcula dos matrices de distancias: una de distancias entre categorías de la variable I (en nuestro caso los atributos) y otra de distancias entre categorías de la variable J (en nuestro caso de las empresas).

El objetivo es determinar las similitudes entre categorías de una misma variable en función de su perfil sobre las categorías de la otra variable. En nuestro ejemplo, la matriz de distancias entre empresas indicará el grado de similitud entre ellas en función de su perfil sobre la serie de atributos en que han sido valoradas. De la misma manera, la matriz de distancias entre atributos, valorará la semejanza entre ellos en función de las empresas a las que han sido asociados.²⁹

El ACOR busca respetar el “principio de equivalencia distribucional”: dos elementos con perfiles iguales son considerados como equivalentes, es decir, representan al mismo elemento, por lo que debe ser posible reemplazar estos dos puntos por uno solo, que tenga las mismas coordenadas y esté afectado por la suma de los correspondientes pesos, y todo ello sin modificar las distancias mutuas de los puntos de la nube.

Dado que la distancia euclidiana utilizada en el análisis de componentes principales no respeta este principio, en el caso del ACOR, la distancia que utiliza es la de x^2 (ji cuadrado), cuya expresión viene dada por:

En el caso de i e i' :

$$d^2(i, i') = d^2(X_i, X_{i'}) = \sum_{j=1}^J \frac{1}{P_j} \left[\frac{n_{ij}}{n_i} - \frac{n_{i'j}}{n_{i'}} \right]^2$$

Respecto j y j' por:

$$d^2(j, j') = d^2(X_j, X_{j'}) = \sum_{i=1}^I \frac{1}{P_i} \left[\frac{n_{ij}}{n_j} - \frac{n_{ij'}}{n_{j'}} \right]^2$$

Siendo:

$$P_i = \sum_{j=1}^J \frac{n_{ij}}{n}$$

$$P_j = \sum_{i=1}^I \frac{n_{ij}}{n}$$

29. En cierta forma, las dos matrices de distancias corresponden cada una a una matriz de correlaciones entre variables iniciales en una ACP.

Etapas 4. Obtención de los ejes factoriales

El proceso de extracción de los ejes factoriales no es más que un caso particular del proceso utilizado en el ACP:

- Obtendremos dos conjuntos de ejes factoriales: los que se extraerán de la matriz de distancias asociada a la variable I (atributos), y los que se extraerán de la matriz de distancias asociada a la variable J (empresas).
- Cada eje factorial restituye una cantidad determinada de información contenida en la matriz inicial de datos. La medida de la cantidad de información restituida por cada factor es *la varianza*.³⁰ Así, el factor que restituye una mayor cantidad de información se obtendrá en primer lugar; el segundo que aporte una mayor cantidad de información, en segundo lugar; y el que aporte menor información en último lugar. Además, los factores extraídos son una combinación lineal de las variables iniciales y no están correlacionados.
- La propiedad fundamental es la relación de simetría que existe entre los ejes factoriales obtenidos de estas dos extracciones efectuadas sobre las nubes $N(I)$ y $N(J)$, respectivamente. Dicha simetría permite situar en un mismo plano a los elementos de I y J , es decir, permite la representación conjunta, en un mismo gráfico, de atributos y empresas (marcas, partidos políticos, sectores económicos, etc).

Etapas 5. Determinación del número de ejes factoriales a conservar

La relación simétrica entre los ejes factoriales extraídos de las dos nubes de puntos hacen que no puedan existir más de K ejes factoriales, siendo $K = \min(I, J) - 1$.

Existe una gran cantidad de reglas y criterios para determinar cuál es el número de ejes factoriales que debemos conservar para el análisis final de la información obtenida. Los más habituales coinciden con los utilizados en la selección de factores a retener en el ACP.

Etapas 6. Interpretación de resultados

En el ACOR los elementos están afectados por pesos distintos y, por lo tanto, pueden ejercer mayor o menor influencia en las relaciones³¹ (*correspondencias*). Para medir el grado de influencia de los distintos elementos en la formación de los ejes factoriales seleccionados, se suele utilizar un índice llamado de *contribución*, pudiendo distinguirse entre contribución absoluta y contribución relativa, así:

30. Es por este motivo que los factores se obtienen según la varianza restituida por cada uno de ellos.

31. Se debe poner atención en cuánto contribuye cada atributo en la formación del eje, lo que se denomina *contribución*.

- Se llama *contribución absoluta* de un elemento (atributo o empresa) a la varianza explicada por un eje, al grado en que dicho elemento ha influido en la formación del eje factorial en cuestión.
- Se llama *contribución relativa* de un elemento a la varianza explicada por un eje a la relación entre su contribución absoluta y la varianza explicada por el eje.

Otros términos relacionados en el ACOR son los siguientes:

- *Masa*: Porcentaje de asociaciones a cada empresa o atributo sobre el total de asociaciones efectuadas.
- *Distancia*: Indica cuán lejos está un atributo o empresa del centro de gravedad. A mayor distancia del centro de gravedad, mayor poder diferenciador tendrá un atributo y más se diferenciará una empresa de las demás.
- *Coordenadas*: Indica las coordenadas de cada atributo o empresa en cada uno de los factores extraídos.
- *Contribuciones absolutas*: Grado en el que un atributo o empresa ha influido en la formación de un eje.
- *Correlaciones cuadradas*: Miden la calidad de la representación de un elemento en un eje determinado. Dentro de una misma representación gráfica con una calidad determinada de representación de la información a nivel total, puede haber elementos (atributos y empresas) que estén representados con niveles muy desiguales de calidad. Cuanto más próximo a 1 sea el valor, mejor calidad de representación habrá.³²

La *simetría* entre los ejes extraídos de las dos nubes de puntos (atributos y empresas) nos permite utilizar para la interpretación a las “proximidades”. La proximidad entre un punto i de I y un punto j de J permite llegar a la conclusión de que existe una fuerte relación entre estos puntos.

Las “reglas” de interpretación de las posiciones de atributos y empresas en el mapa son las siguientes:

- *Proximidad entre dos atributos*: cuando dos atributos estén próximos podemos concluir que tienen perfiles parecidos, es decir las variables o modalidades J asociadas a dos atributos próximos son similares, y por lo tanto, estarán

32. Generalmente, una fuerte contribución a la formación de un eje conlleva a una buena representación sobre este eje; pero lo recíproco es a menudo falso: el hecho de que exista una correlación alta no implica que la contribución también sea alta.

fuertemente intercorrelacionadas; caso contrario, se dice que tienen perfiles muy distintos.

- *Proximidad entre dos modalidades J*: cuando dos modalidades J (empresas, etc.) estén próximas podemos decir que tienen los mismos puntos fuertes y los mismos puntos débiles; les han sido asociados los mismos atributos; caso contrario serán muy distintas.
- *Proximidad de atributos o modalidades J al origen*: Cuando una modalidad J se encuentra proyectada cerca del origen, significa que tiene una distribución de valores sobre su columna que, en porcentaje, tiene el mismo perfil que la distribución de valores de la columna “total” y se comporta, pues, como la “media de las modalidades J”, por lo que no se diferencia del resto de las modalidades J en ningún atributo; igual aplica para el caso de las modalidades I (atributos).
- Así concluimos que, a mayor alejamiento del origen de un elemento, habrá mayor poder descriptivo.
- Proximidad entre atributos y modalidades J: Significa que existe una fuerte asociación entre cada atributo y la modalidad J correspondiente.

Como ayuda a la interpretación de los resultados obtenidos a través de un análisis de correspondencias, podría aplicarse el análisis tipológico jerárquico sobre la matriz de coordenadas de las empresas y los atributos en los dos primeros factores.

5.3. Aplicaciones del análisis de correspondencias

Una vez hemos visto las principales características del ACOR, es oportuno concluir los siguientes aspectos:

- El método se aplica principalmente al análisis de matrices de valores absolutos, es decir, matrices obtenidas por simple conteo (o cualquier matriz de números no negativos). Por lo tanto, se hace imprescindible tener siempre en cuenta las siguientes precauciones:
 - Puesto que los datos iniciales serán divididos por los márgenes P_i y P_j es necesario que dichos márgenes tengan un sentido; es decir, la suma de cada fila y de cada columna debe poderse interpretar.
 - Dado que en el cálculo de las distancias entre dos elementos interviene la suma de las columnas, es preciso que esta suma tenga algún sentido.

Basándonos en el hecho de que para la aplicación del ACOR es suficiente con disponer de una matriz de números no negativos, es evidente que sus posibilidades de aplicación en investigación de mercados sean mayores. Su principal campo de aplicación se centra en analizar la percepción del consumidor sobre distintas marcas/productos que compiten en un mercado determinado, valorando así la imagen y el posicionamiento de cada marca/producto a través de conceptos cualitativos como pueden ser atributos, cualidades, etc.

Entre las principales aplicaciones del ACOR se encuentran las siguientes:

- *Visualización de los fenómenos de competencia en un mercado.* Por ejemplo, supongamos que un fabricante de computadoras personales se plantea la necesidad de conocer los puntos fuertes y débiles de las distintas marcas que concurren en este mercado, será suficiente con preparar una lista de atributos (servicios postventa, precio, etc.) y aplicarla a una serie de usuarios (una encuesta) con el fin de que indiquen los atributos que más se asocian a cada marca. El análisis de la matriz de valores absolutos resultante, marca de ordenador por atributo, permitirá visualizar y, por lo tanto, determinar los principales “fenómenos” de competencia en este mercado, lo que nos debería permitir tomar decisiones sobre el posicionamiento o reposicionamiento de la marca, según diferentes alternativas que se desprenden del propio análisis: mantener/reforzar el posicionamiento, reposicionarla cerca de algún competidor, ya sea el líder u otro cuyo posicionamiento sea atractivo, buscar un hueco en el mercado, reposicionando la marca en una zona atractiva en la que estén actuando los competidores.
- *Visualización del posicionamiento de ideal del mercado, y por comparación con el posicionamiento de las marcas competidoras, determinar las marcas más cercanas a dicho ideal.* Se puede introducir en el análisis, como *elemento suplementario*, la denominada *empresa ideal*.³³ Esta empresa no intervendría en el análisis, pero el *software* utilizado permitiría estudiar su posicionamiento (se trata de una variable “pasiva” en el análisis”).
- *Ingeniería de imagen.* Por ejemplo, supongamos que los directivos de la “empresa 5” se plantearan la necesidad de actuar principalmente sobre el criterio de “mejores precios/descuentos” y “mayor variedad de marcas”, pero antes de iniciar su actuación mediante una fuerte campaña publicitaria, quisieran establecer cuál podría ser el nuevo posicionamiento que se obtendría.

Con el fin de intentar satisfacer su “curiosidad” se podría realizar una serie de nuevos análisis de correspondencia, basándonos en la matriz de datos inicial

33. El vector ideal se puede configurar con las frecuencias máximas de cada atributo.

y sustituyéndose, en la misma, los distintos valores de la “empresa 5” por los esperados después de las posibles actuaciones.

Supongamos que, en dicha simulación se pretendiera analizar cuál sería el posicionamiento de la “empresa 5” si dicha empresa disminuyera a la mitad la diferencia con las empresas que reciben máxima puntuación en los atributos “mejores precios/descuentos” y “mayor variedad de marcas”.

Los resultados que se obtengan a través de dicha simulación permitirían constatar que si la “empresa 5” consiguiera aumentar sus asociaciones en los atributos “mejores precios/descuentos” y “mayor variedad de marcas”, esto le permitiría emprender un camino de acercamiento al posicionamiento “ideal”.

5.4. Análisis de correspondencias simple (ANACOR)

Ya hemos mencionado que el análisis factorial de correspondencia simple está particularmente adaptado para tratar tablas de contingencia, representando los efectos existentes en las múltiples modalidades (categorías) combinadas de dos caracteres (variables cualitativas). Si cruzamos en una tabla de contingencia el carácter 1 con modalidades desde $i=1, \dots, n$ (en filas), con el carácter J con modalidades desde $j=1, \dots, p$ (en columnas), podemos representar el número de unidades estadísticas que pertenecen simultáneamente a la modalidad i del carácter I y la modalidad j del carácter J mediante n_{ij} .

En este caso la distinción entre observaciones y variables en el cuadro de doble entrada es artificial; pero, por similitud con componentes principales, suele hablarse a veces de individuos u observaciones cuando nos referimos al conjunto de las modalidades del carácter I (filas), y las variables cuando nos referimos al conjunto de las modalidades del carácter J (columnas).

La finalidad del análisis de correspondencias es determinar la posición de una serie de objetos/sujetos (segmentos de mercado, grupos de individuos o personas físicas, sectores, productos, etc.) en una serie de atributos, características, escalas de valoración, etc., y ello a través de un espacio vectorial de dos, tres o más dimensiones.

5.5. Aplicación mediante SPSS (ANACOR)

Ejemplo 1

En su planteamiento más sencillo, dos variables medidas en una escala nominal (pueden estarlo una ordinal o numérica) y que dan lugar a una tabla de contingencia, es el requerimiento para la aplicación del ANACOR. Para poder realizar un análisis de correspondencia simple tendremos que seguir los siguientes pasos:

- Analizar – Reducción de datos – análisis de correspondencias.

En este ejemplo vamos a trabajar con el anexo “correspondencias.sav” en el que se miden a 1,200 individuos dos variables, el *estado civil* (variable numérica), clasificado en tres grupos: solteros, casados y otros; y sus *preocupaciones más importantes* (variable numérica), en cuatro categorías: la vida afectiva, el dinero, la armonía familiar y la salud.

Podemos analizar previamente los datos a través de una tabla de continencia. Los valores reflejan un $\chi^2 = 55.464$ con seis grados de libertad (g.l.) y un nivel de significación $p = 0.000$, que dan a entender la existencia de una relación significativa entre el estado civil de los individuos y los aspectos de la vida (cuatro en este caso) que más preocupan a los mismos.

Ahora aplicaremos el análisis de correspondencias para profundizar en la estructura de relaciones entre las dos variables en estudio.

Preocupaciones: aspecto que más le preocupa * Estado civil Crosstabulation

Count		Estado civil			Total
		Soltero	Casado	Otro	
Preocupaciones: aspecto que más le preocupa	Su vida afectiva	26	32	5	63
	El dinero	111	166	13	290
	Armonía familiar	59	197	24	280
	Su salud	105	348	72	525
Total		301	743	114	1158

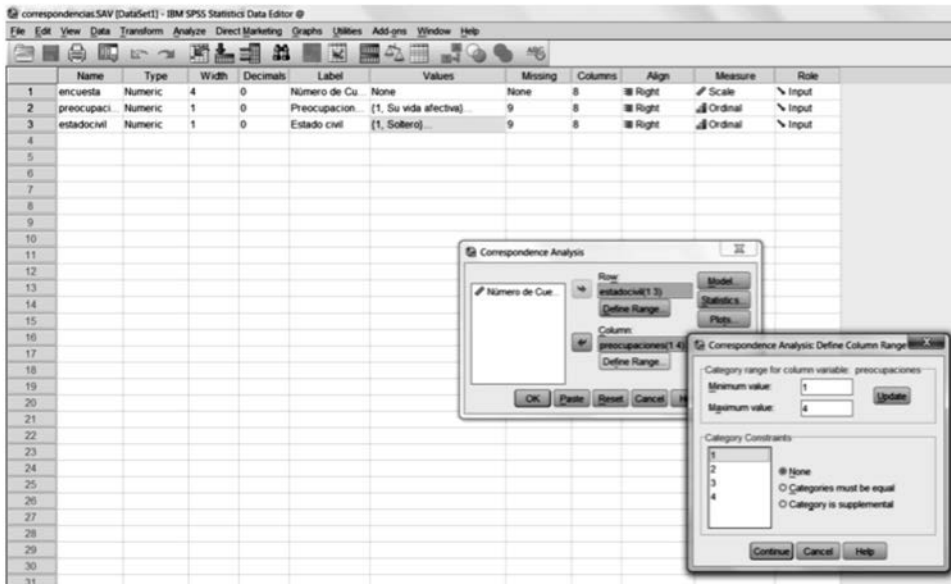
Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	55.464 ^a	6	.000
Likelihood Ratio	54.751	6	.000
Linear-by-Linear Association	44.175	1	.000
N of Valid Cases	1158		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.20.

- Una vez seleccionado el análisis obtendremos el siguiente menú, en donde debemos seleccionar las variables respectivas y las categorías que tiene

cada una. En este ejemplo la variable Fila será *estado civil*, la cual tiene tres categorías (ordenadas de la uno a la tres), por lo que el rango tomará como valor máximo el tres y el mínimo el uno. La variable Columna será *preocupaciones*, y el rango será de uno a cuatro.



La sintaxis generada por el procedimiento en SPSS es la siguiente:

```

DATASET ACTIVATE DataSet1.
CORRESPONDENCE TABLE=preocupaciones(1 4) BY estadocivil(1 3)
/DIMENSIONS=2
/MEASURE=CHISQ
/STANDARDIZE=RCMEAN
/NORMALIZATION=SYMMETRICAL
/PRINT=TABLE RPROFILES CPROFILES
/PLOT=NDIM(1,MAX) BIPLLOT(20).
    
```

Al momento de indicar las categorías de cada variable tenemos una serie de opciones sobre ellas, y son estas:

1. Las categorías deben ser iguales. Utilizaremos esta opción si existen dos categorías que son iguales o las queremos agrupar.
2. La categoría es suplementaria si no influye en el análisis, pero se representa.

Botón de acción “Modelo”

Con este botón especificaremos el número de dimensiones, la medida de distancia, el método de estandarización y el método de normalización.

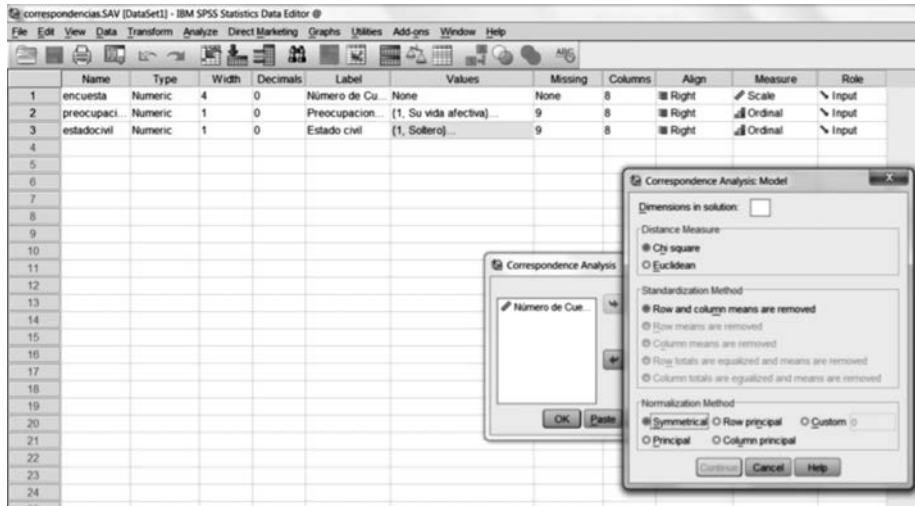
- Dimensiones en la solución: tendremos que tomar en cuenta que el número óptimo de dimensiones será el menor número de dimensiones que necesitaremos para explicar la mayor parte de la variación. Evidentemente, intentaremos que este valor sea el mínimo posible, de tal manera que el número máximo de dimensiones será el mínimo del número de filas menos uno (en este caso $3-1=2$) y el de columnas menos uno ($4-1=3$), es decir, en nuestro ejemplo usaremos dos dimensiones.
- Medida de distancia: por defecto usaremos la distancia chi-cuadrado.
- Método de normalización: esta es una de las decisiones más importantes, ya que dependiendo del método se producirán soluciones que, aunque equivalentes, pueden ser diferentes. La finalidad es asignar o distribuir la inercia de la tabla o por filas, o por columnas o por ambas a la vez.

Las diferentes opciones son estas:

1. Simétrico: en este caso la inercia se reparte igualmente entre filas y columnas. Usaremos este método para examinar las diferencias entre las categorías de las dos variables conjuntamente; posibles diferencias o similitudes entre las dos variables.
2. Principal: usaremos este método si deseamos examinar las diferencias entre las categorías, tanto de la variable fila como de la columna, pero por separado (no conjuntamente, como en el método simétrico).
3. Principal por filas: presenta la inercia de la tabla por fila, es decir, si estamos interesados en indagar cómo difieren entre sí las distintas categorías de la variable fila, de modo que maximice las distancias existentes entre las diversas categorías fila.³⁴
4. Principal por columna: similar al anterior, pero por columna. Se podría estar interesado en diferencias en columnas y no tanto en filas.

34. Si en las filas tenemos, por ejemplo, sectores industriales, productos diferentes, tipologías, etc. y en columnas, aspectos evaluadores de los mismos, lo más probable es que se pueda estar interesado en averiguar posibles diferencias entre filas y no tanto entre columnas.

5. Personalizado. Reparte la inercia entre filas y columnas en diferentes grados. Se puede especificar³⁵ entre -1 y 1.



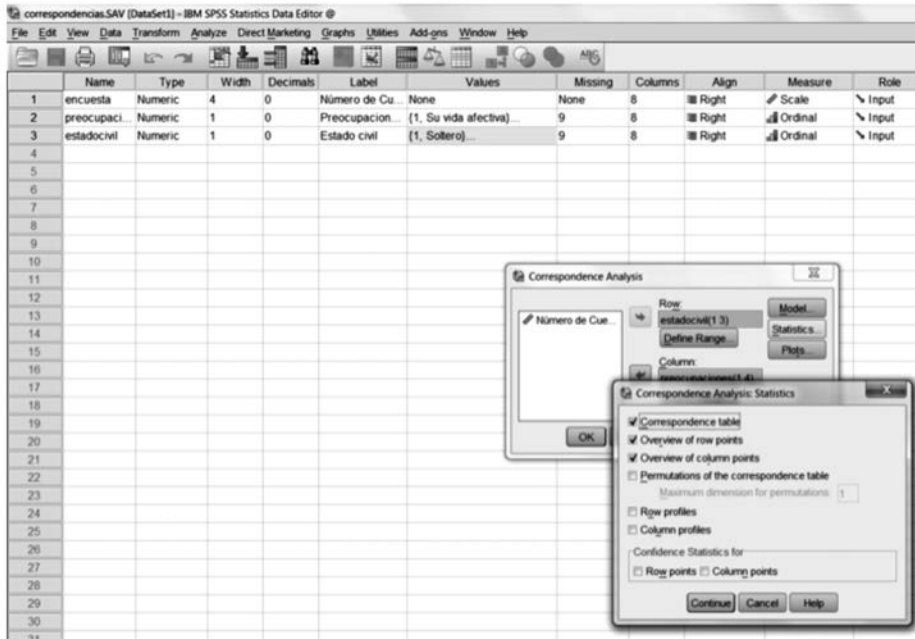
En nuestro ejemplo elegiremos la medida de la chi-cuadrado y como método de normalización el denominado simétrico. Como ya hemos mencionado el número de ejes o dimensiones apropiado será de dos.

Botón de acción “Estadístico”

Con este botón podremos obtener las siguientes salidas:

- Tabla de correspondencias: tabla de contingencia de las variables con los totales marginales de filas y columnas.
- Inspección de los puntos de fila: resumen de cada categoría de filas, las puntuaciones, la masa, la inercia, la contribución a la inercia de la dimensión y la contribución de la dimensión a la inercia del punto.
- Inspección de los puntos de columna: igual que para las filas, pero para cada columna.
- Permutaciones de la tabla de correspondencias.
- Perfiles de fila: marginales por filas.
- Perfiles de columna: marginales por columnas.
- Estadísticos de confianza para puntos de fila.
- Estadísticos de confianza para puntos de columna.

35. Si escogemos el valor uno el procedimiento equivale a “Row principal”, si escogemos -1 equivale a “Column principal” y si escogemos el cero estaremos utilizando el procedimiento “Symmetrical”. Cualquier otro valor reparte la inercia entre filas y columnas en diferentes grados.



Seleccionemos todas las opciones disponibles en este botón de acción.

Análisis de resultados

1. **Tabla de correspondencias:** simplemente es la tabla de doble entrada de resumen de los datos; en la figura siguiente tenemos la tabla a analizar, con sus correspondientes valores marginales. Observamos que en general el aspecto que más preocupa a los individuos de la muestra es la salud (525/1158) y el que menos, su vida afectiva, con 63/1158; además, en ella podemos ver que, por ejemplo, *existen 26 solteros cuya preocupación más importante es la vida efectiva* o 111 solteros cuya mayor preocupación es el dinero.

Correspondence Table					
Estado civil	Preocupaciones: aspecto que más le preocupa				
	Su vida afectiva	El dinero	Armonía familiar	Su salud	Active Margin
Soltero	26	111	59	105	301
Casado	32	166	197	348	743
Otro	5	13	24	72	114
Active Margin	63	290	280	525	1158

2. **Perfiles por fila y columna:** son las distribuciones marginales. Los perfiles por fila son el resultado de dividir cada valor por el total de su fila mientras que los perfiles por columna es el valor de cada casilla entre el total su columna.

En el primer caso la tabla refleja las proporciones que el número de individuos de cada celda representan sobre el total de la fila; y en el segundo, esta misma proporción, pero sobre el total de la columna. El grado de similitud o disimilitud entre estos perfiles tanto por filas como por columnas, quedará reflejado en los gráficos que veremos a continuación en términos de proximidad o lejanía entre las categorías de las variables.

Row Profiles

Preocupaciones: aspecto que más le preocupa	Estado civil			
	Soltero	Casado	Otro	Active Margin
Su vida afectiva	.413	.508	.079	1.000
El dinero	.383	.572	.045	1.000
Armonía familiar	.211	.704	.086	1.000
Su salud	.200	.663	.137	1.000
Mass	.260	.642	.098	

Column Profiles

Preocupaciones: aspecto que más le preocupa	Estado civil			
	Soltero	Casado	Otro	Mass
Su vida afectiva	.086	.043	.044	.054
El dinero	.369	.223	.114	.250
Armonía familiar	.196	.265	.211	.242
Su salud	.349	.468	.632	.453
Active Margin	1.000	1.000	1.000	

3. **Tabla resumen:** en esta tabla destacaremos varios aspectos así:

- Valor propio (eigenvalues): Se interpretan como las correlaciones entre las puntuaciones de las filas y las columnas, y son análogos a las correlaciones de Pearson. Para cada dimensión el cuadrado del valor singular es la inercia y, por lo tanto, una medida de la importancia de la citada dimensión a la hora de explicar los datos de la tabla.
- Inercia: Es el cuadrado del valor propio. *Mide la importancia de un eje;* 0.043 el primero y 0.005 el segundo.
- Inercia total: Es la explicación total de los ejes. Los ejes explican un total de 0.048. La inercia total es definida como la suma pondera de todas las distancias al centroide y dividida entre la suma de todas las celdas en la tabla correspondiente. Las filas con una “masa” pequeña (marginales

pequeños) influyen sobre la inercia solamente cuando están lejos del coentroide, y en cambio las filas con una “masa” muy alta influyen en la inercia total, incluso si están próximas al centroide. Lo mismo podría aplicarse a las columnas.

- d) **Proporción de inercia:** De las dos dimensiones establecidas por defecto, la primera refleja la mayor proporción de inercia de los datos, la segunda dimensión la siguiente, y así sucesivamente hasta completar el 100%. Recordar que la inercia total es la suma a lo largo de las diversas dimensiones de todas las distancias respecto al origen, y que es posible dividir esa inercia en componentes atribuibles a cada dimensión y comparar este parcial con el total de inercia.

Aunque el valor de la inercia podría parecer pequeño, 0.043 y 0.005, entre ambos ejes explican el 100% de la varianza, y más concretamente el primero, el 90.4%; y el segundo, el 9.60%. Son justamente estos valores los que de alguna manera pueden ayudarnos a decidir el número de dimensiones definitivas que seleccionaremos para los datos y que combinen la capacidad explicativa con un porcentaje elevado.

- e) **Chi-cuadrado:** Con este valor obtenemos el contraste de independencia entre ambas variables, *el contraste de la chi-cuadrado en las tablas de contingencia contrasta la independencia como hipótesis nula frente a la dependencia*. En este caso al ser el p-valor *aceptaríamos la hipótesis de dependencia entre variables*. Destacamos que en el caso de que dos variables fuesen independientes, no tendría sentido el análisis de correspondencias.
- f) **Valores singulares de confianza:** Valores bajos indican que el ACOR reflejaría la misma solución con otras muestras de igual tamaño, obtenida de la misma población.
- g) **Correlación:** Valores altos dificultarían a una categoría en el espacio bidimensional.

Summary								
Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation
1	.208	.043			.904	.904	.029	.046
2	.068	.005			.096	1.000	.029	
Total		.048	55.464	.000 ^a	1.000	1.000		

a. 6 degrees of freedom

- h) **Examen de los puntos fila:** Las puntuaciones en la dimensión son los valores que toma cada punto en la representación que hayamos calculado. En la columna “Inercia” vemos lo que aporta cada categoría al total de inercia, por ejemplo, del total de inercia (0.048) los *solteros* aportan 0.028.

La columna “Contribución de los puntos a la inercia de la dimensión” mide la importancia de un punto a la hora de explicar un eje, por ejemplo, la categoría *soltero* es el valor más importante para la orientación del primer eje y la categoría *otro* para el segundo eje.

En la columna “Contribución de la dimensión a la inercia del punto” vemos el porcentaje de cada modalidad explicado por cada eje. El primer eje explica el 98.30% de la modalidad *soltero* frente a solo el 1.7% del segundo eje.

- i) **Examen de los puntos columna:** Igual que el caso anterior, además de los valores para representar en los ejes, podemos ver que *el primer eje viene muy determinado por el dinero* frente al *segundo que lo hace por la armonía familiar*. También podemos ver que el primer eje explica el 99.9% de la varianza de la modalidad de Dinero; y el 93%, de Salud.

En las tablas siguientes tenemos las contribuciones de cada categoría a la inercia de cada dimensión (*Of points to Inertia of Dimensions*) y a su vez la contribución de cada dimensión a la inercia de cada categoría (*Of Dimension to Inertia of Points*), por filas y por columnas. En el primer caso debemos interpretarlas como la proporción de varianza de cada dimensión, que es debida a cada categoría (la suma por columnas igual a la unidad); y en el segundo caso, la proporción de varianza de cada categoría explicada por cada una de las dimensiones o factores (la suma por filas igual a la unidad). En concreto, y por lo que hace referencia a la varianza de la primera dimensión o factor, el *dinero* con un valor 0.547, y la *salud* con 0.289 son las dos categorías con mayor contribución; en tanto que para la segunda dimensión son la *armonía familiar* y la *vida afectiva*, con contribuciones 0.567 y 0.220, respectivamente.

La interpretación de las contribuciones de cada factor o dimensión a la explicación de la variabilidad de cada categoría sigue el mismo criterio, de modo que el primer factor o dimensión explica prácticamente el 100% de la categoría fila *dinero* (99.9 exactamente), en tanto que, en esta primera dimensión, la *armonía familiar* únicamente explica el 36.7%.

Idéntica explicación cabe dar a las contribuciones de las categorías *columna* de la tabla, es decir, de la variable *estado civil*.

Overview Row Points ^a										
		Score in Dimension		Inertia	Contribution					
		1	2		Of Point to Inertia of Dimension		Of Dimension to Inertia of Point			
					1	2	1	2	Total	
⇒	Estado civil	Mass								
	Soltero	.260	-.712	-.166	.028	.634	.106	.983	.017	1.000
	Casado	.642	.171	.168	.005	.091	.268	.762	.238	1.000
	Otro	.098	.764	-.656	.015	.276	.626	.807	.193	1.000
	Active Total	1.000			.048	1.000	1.000			
a. Symmetrical normalization										

Overview Column Points ^a									
Preocupaciones: aspecto que más le preocupa	Mass	Score in Dimension		Inertia	Contribution				
		1	2		Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
					1	2	1	2	Total
Su vida afectiva	.054	-.703	-.523	.007	.129	.220	.848	.152	1.000
El dinero	.250	-.674	.046	.024	.547	.008	.999	.001	1.000
Armonía familiar	.242	.173	.398	.004	.035	.567	.367	.633	1.000
Su salud	.453	.365	-.175	.013	.289	.205	.930	.070	1.000
Active Total	1.000			.048	1.000	1.000			

a. Symmetrical normalization

Permutaciones de la tabla de entrada

Si escogemos las puntuaciones por filas y columnas de la primera dimensión en las tablas anteriores, observaremos que *soltero-casado y otros*, con puntuaciones -0.712 0.71 0.764, están ordenadas de menor a mayor puntuación en la primera dimensión, en tanto que las categorías de la variable columna lo están con el siguiente orden: *su vida afectiva-el dinero-armonía familiar y su salud*, con puntuaciones -0.703, -0.674, -0.173 y 0.365, respectivamente.

La tabla siguiente recoge esta ordenación y reasigna las frecuencias correspondientes a la intersección de cada categoría.

Permuted Correspondence Table According to Dimension 1

Preocupaciones: aspecto que más le preocupa	Estado civil			
	Soltero	Casado	Otro	Active Margin
Su vida afectiva	26	32	5	63
El dinero	111	166	13	290
Armonía familiar	59	197	24	280
Su salud	105	348	72	525
Active Margin	301	743	114	1158

Finalmente, tenemos las desviaciones típicas y las correlaciones entre valores singulares que aparecen en las dos últimas columnas de la tabla resumen (summary)-en la columna "Confident Singular Value", y las desviaciones típicas y las correlaciones de las puntuaciones por filas y columnas que aparecen en las tablas siguientes: Confident Row Points y Confidence Column Points. Estos estadísticos tienen el nombre de *estadísticos de confianza*.

Recordar que la tabla de este ejercicio ilustrativo del ACOR recoge el *estado civil* y las *preocupaciones* de una muestra de los 1158 individuos que constituyen una muestra representativa de una población mayor de 18 años. Las frecuencias de esta tabla siguen una distribución multinomial, que origina la desviación y las correlaciones.

Valores singulares (desviaciones típicas): Si se tienen valores bajos de las desviaciones, esto indica que el análisis de correspondencias dará prácticamente la misma solución con otras muestras de igual tamaño y extraídas con idéntico procedimiento de la misma población.

Valores singulares (correlaciones): Si se tienen valores altos, estos dificultarán situar una categoría en el espacio bidimensional.

Respecto a las desviaciones y las correlaciones para cada una de las categorías *fila* y *columna* de la tabla del ejemplo, el hecho de tener en general valores muy próximos a cero, en el caso de las desviaciones, da a entender que el análisis de correspondencias ha dado una solución global estable.³⁶

Confidence Row Points

Preocupaciones: aspecto que más le preocupa	Standard Deviation in Dimension		Correlation
	1	2	1-2
Su vida afectiva	.162	.129	-.469
El dinero	.051	.063	.168
Armonía familiar	.117	.087	-.184
Su salud	.055	.050	.548

Confidence Column Points

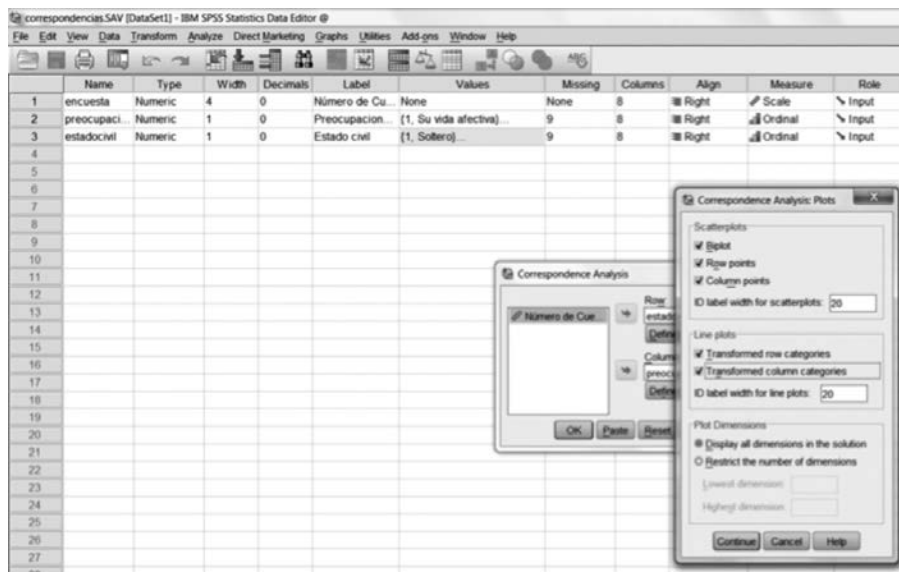
Estado civil	Standard Deviation in Dimension		Correlation
	1	2	1-2
Soltero	.074	.072	-.583
Casado	.054	.039	-.356
Otro	.176	.153	.381

36. Las correlaciones tienen en algunos casos valores más altos, tanto en filas como en columnas. En estos coeficientes influirá sobremanera la normalización utilizada (tenderán obviamente a ser menores por filas o por columnas si la normalización utilizada es por filas o por columnas), al igual que los valores singulares de cada una de estas dimensiones.

Botón de acción “Gráficos”

Dentro del botón de gráficos podremos especificar las siguientes opciones:

- Diagramas de dispersión: Produce los gráficos por parejas de las dimensiones. Entre ellos:
 1. Diagrama de dispersión biespacial: Diagramas conjuntos de los puntos de filas y de columnas.
 2. Puntos de fila.
 3. Puntos de columna.
- Gráfico de líneas. Para obtener los siguientes gráficos unidimensionales:
 1. Categorías de fila transformadas: produce un gráfico de los valores originales para las categorías de fila frente a las puntuaciones de fila.
 2. Categorías de columna transformadas: produce un gráfico de los valores originales para las categorías de columna frente a las puntuaciones de columna.
- Dimensiones del gráfico.
 1. Muestra todas las dimensiones de la solución.
 2. Restringe el número de dimensiones.

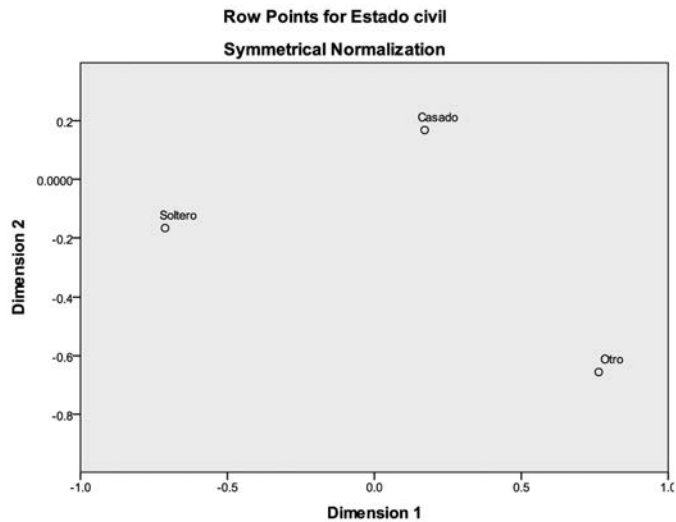


- Seleccionemos todos los gráficos posibles.

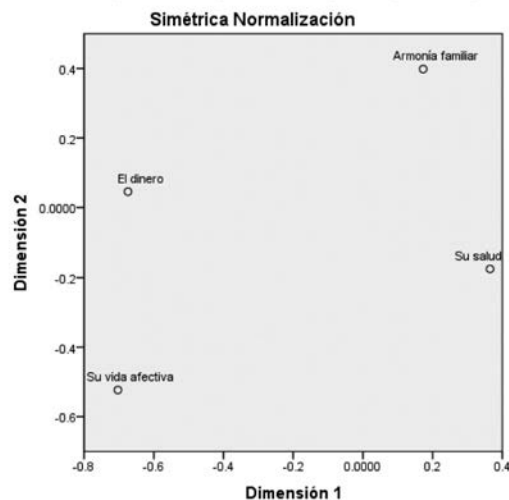
Análisis

Si hubiera más dimensiones obtendríamos las combinaciones de las dimensiones para filas y columnas a la vez y por separado.

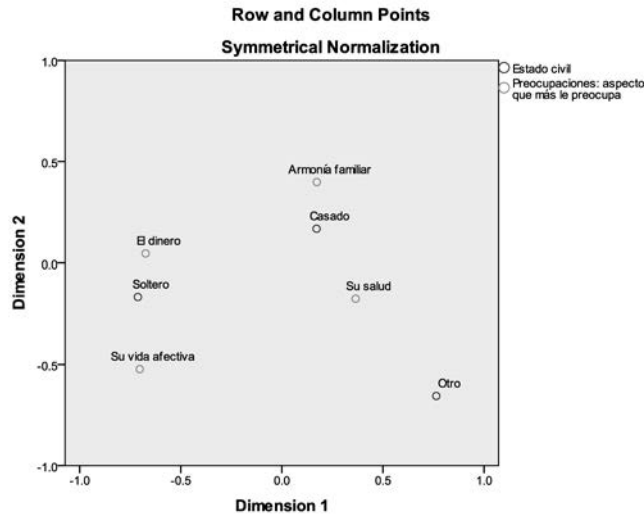
Si analizamos los perfiles fila, observamos que los tres grupos: *solteros*, *casados* y *otros* son bastante diferentes (lejanos). Si observamos los puntos columna, en estos parece que existe semejanza entre dinero y vida afectiva, y entre salud y armonía familiar.



Puntos de columna para Preocupaciones: aspecto que más le preocupa



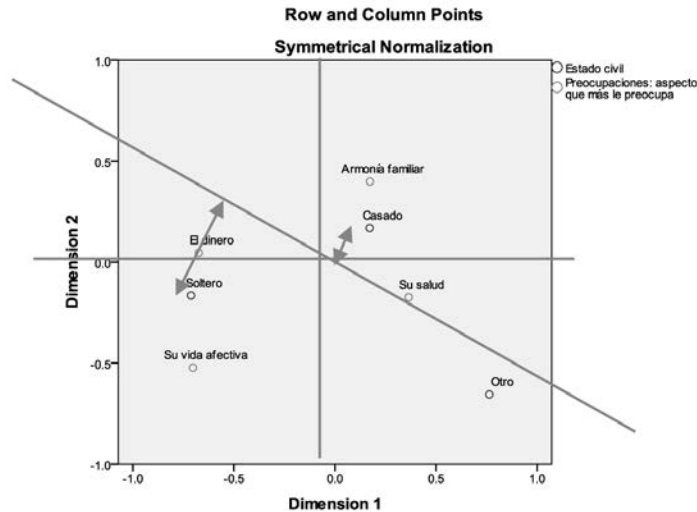
- Puede resultar de interés en muchos casos llevar a cabo una primera aproximación al sentido y significado de cada uno de los ejes, que pueden contraponer aspectos, atributos, cualidades, etc., opuestos y llevarnos a una interpretación más o menos difícil. La primera dimensión contrapone en alguna medida la *salud* y el *dinero*, en tanto que la interpretación de la segunda va más en la línea de *armonía familiar* y *vida afectiva*.



- Conviene tomar en cuenta que ambas dimensiones no explican el mismo porcentaje de variabilidad de los datos de la tabla (90.4% y 9.6% respectivamente), y en ese sentido la interpretación de la estructura de relaciones entre categorías de ambas variables en el plano bidimensional debe tener en cuenta esta circunstancia.
- Próximas al origen del espacio bidimensional o pluridimensional estarán las categorías de ambas variables que menos discriminan cada una de las dimensiones de la solución y, al revés, estas categorías estarán tanto más alejadas del origen, cuanto mayor sea la discriminación.
- Se puede llevar a cabo una lectura superficial del gráfico por cuadrantes, que en el ejemplo asimilaría *casados* a *armonía familiar*, *solteros* a *dinero* y *vida afectiva* y el resto a la *salud*.
- En definitiva, mayor o menor proximidad entre las categorías en el plano equivale a un mayor o menor grado de relación o interdependencia entre las mismas.

A objeto de llevar a cabo una más rigurosa aproximación al grado de relación entre categorías, en el gráfico de la figura siguiente se ha llevado a cabo el trazado de una

recta desde el origen del espacio bidimensional a una de las categorías fila de la tabla (la salud). Posteriormente puede llevarse a cabo una proyección ortogonal desde cada categoría columna hasta la misma (solteros y casados en el ejemplo), y la distancia desde esta intersección entre ambas líneas hasta la categoría columna, puede darnos una idea de cómo las categorías de ambas variables están relacionadas.



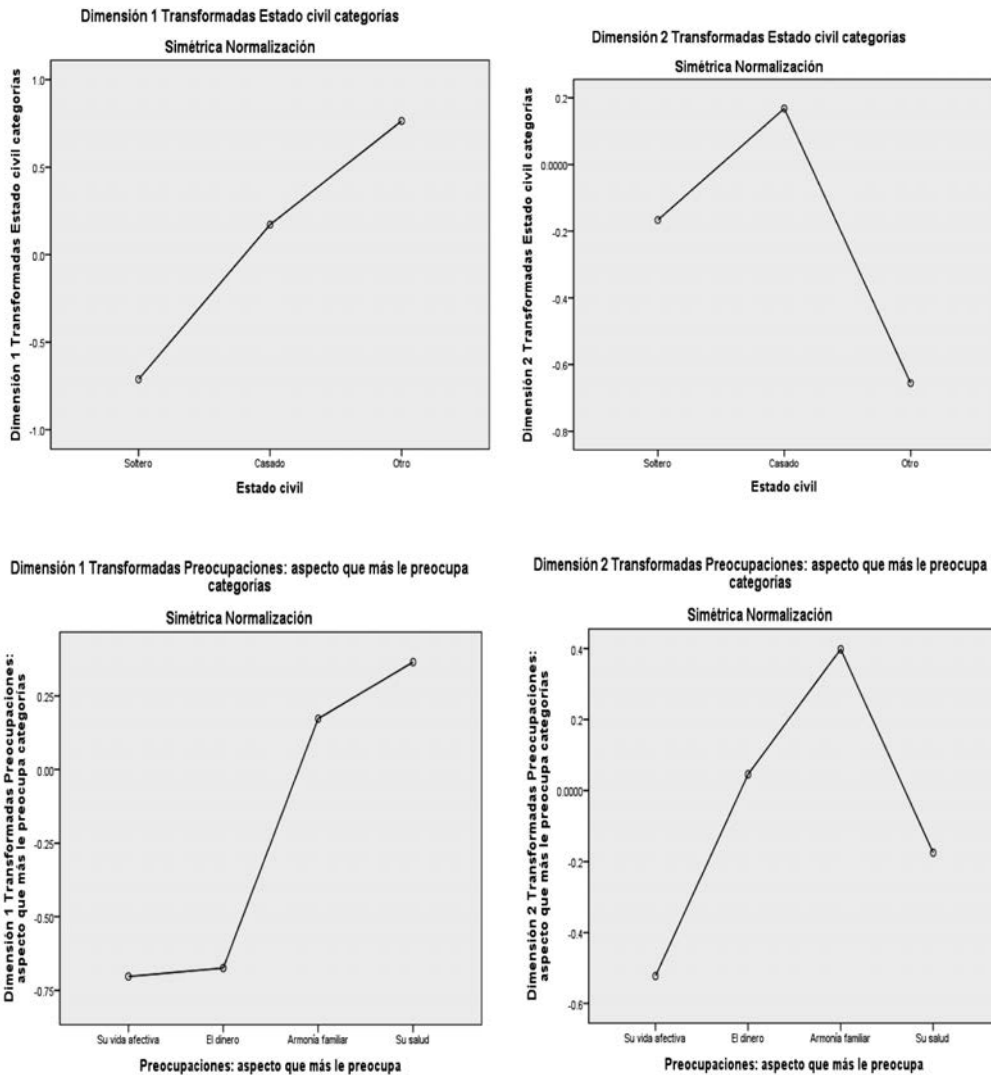
En la interpretación de este y el resto de gráfico del ACOR conviene tener en cuenta, la normalización utilizada, que en nuestro ejemplo hemos utilizado “symmetrical”, que es la opción por defecto.

En resumen, la finalidad del análisis de correspondencias no es otra sino encontrar una configuración en la cual las distancias entre las categorías fila o columna, en un espacio unidimensional, bidimensional o pluridimensional, sea un fiel reflejo de esas mismas distancias entre los perfiles (filas, columnas o ambas) de la tabla. Ahora bien,³⁷ estas distancias son unas distancias ponderadas y basadas en el concepto de masa. La masa afecta al centroide, que es la media ponderada del perfil *fila* o *columna* de modo que el centroide *fila* es la media del perfil *fila*, y el centroide *columna* es la media del perfil *columna*. Según esto, categorías con una gran masa tendrán mayor influencia en el centroide, y al revés, cuando la masa sea pequeña. No tienen influencia en la posición del centroide los 743 individuos “Casados” de la muestra que los 114 de la categoría “Otros”.

Finalmente, los gráficos de líneas muestran los valores de las componentes o puntuaciones por filas y por columnas en cada una de las dos columnas

37. La influencia que ejerce una categoría de acuerdo con su frecuencia marginal.

(unidimensionalmente). Estos gráficos corresponden a las columnas “Score in Dimension” de las tablas *Overview Row/Columns Points*, mencionadas en el literal i.



5.6. Ejercicios

Ejercicio 1

El anexo *correspondencias3.sav*, contiene los datos para 1,529 individuos, agrupados en categorías profesionales (numeradas del uno al cinco), y su consumo de tabaco (categorías uno a cuatro). Se solicita lo siguiente:

1. ¿Cuál sería el número máximo posible de dimensiones?
2. ¿Se podría realizar un análisis de dos dimensiones?, ¿por qué?
3. Realizar un análisis de correspondencias con dos dimensiones. ¿Sería admisible agrupar alguna modalidad de la variable consumo de tabaco? ¿Y de categoría profesional?
4. Comentar el gráfico bidimensional.

Ejercicio 2

Se quiere analizar la influencia de la raza en cada una de las regiones de América del Norte. Para ello se selecciona una muestra de tal manera que a esos individuos se les analiza la región en donde viven y la raza a la que pertenecen. Los resultados aparecen en el anexo *correspondencias4.sav*. Se desea realizar un análisis de correspondencia completo (incluido el análisis de independencia y el estudio de las tablas de frecuencias marginales) en el que se analice esta hipótesis.

Ejercicio 3

Mediante análisis de correspondencia simple analizar la relación entre las categorías de las variables, características recogidas en el anexo *cochessimples.sav*. Este fichero contiene información sobre 12 individuos con las siguientes características: estado civil, procedencia, peso y país de origen.

6. Análisis múltiple de correspondencias (HOMALS)

Introducción

Hemos visto que el análisis de correspondencias simple es de aplicación con dos caracteres o variables cualitativas (análisis de correspondencia simple o sencillamente análisis factorial de correspondencias), cada una de las cuales puede presentar varias modalidades o categorías. Pero el método es generalizable al caso de un número de variables o caracteres cualitativos mayor de dos (análisis de correspondencias múltiples). Cuando el número de variables cualitativas es mayor que dos (en vez de tener sólo los caracteres I, J tenemos los caracteres $J_1, J_2, \dots, J_Q$) *ya no se puede hablar de tabla de contingencia y la representación tabulada de los datos se complica*. No obstante, el análisis en correspondencias múltiples permite estudiar las relaciones entre las modalidades de todas las características cualitativas consideradas.

El Análisis múltiple de correspondencia añade la posibilidad de trabajar con tres o más variables y buscar los patrones y la estructura de relaciones entre las mismas, de un modo gráfico en plots bidimensionales o tridimensionales. HOMALS es un acrónimo para referirse al mismo *Homogeneity Analysis by means of Alternating Least Squares*, que también se conoce con el nombre de “análisis múltiple de correspondencias”.

Los datos de entrada del procedimiento HOMALS lo constituye una matriz de datos, en la que las filas están formadas por sujetos/objetos de diversa índole (personas físicas, marcas, productos, empresas, etc.); y las columnas, por variables que en principio son tratadas como medidas en una escala nominal y, por tanto, con categorías meramente diferentes unas de otras.

El procedimiento HOMALS utiliza en la estimación de sus parámetros el procedimiento conocido como “*Alternating Least Squares*” o alternancia de mínimos cuadrados. La estimación mínimos cuadrados de los valores de los parámetros se realiza en dos fases: una de estimación del modelo propiamente dicha y otra de escalamiento óptimo. Estas dos fases se van alternando iterativamente hasta conseguir una convergencia determinada. En la fase de escalamiento óptimo,

las variables son analizadas de acuerdo con una escala de medidas especificada previamente.

La finalidad del análisis de homogeneidad (HOMALS) es conseguir cuantificaciones de los objetos/sujetos (Object/Scores) y, por tanto, de las categorías de las variables (Category Quantifications) que sean óptimas (Optimal Scores), en el sentido de que las categorías estén separadas unas de otras en la dimensión o dimensiones estudiadas tanto como sea posible y, a su vez, dentro de cada categoría los sujetos estén lo más próximos unos a otros, es decir, con puntuaciones cuanto más homogéneas entre sí (de ahí el nombre de análisis de homogeneidad).

Los términos relacionados que calcularemos más adelante son estos:

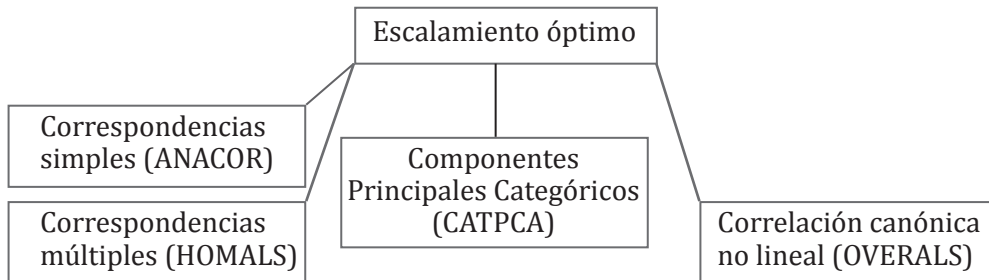
- **Objetc Scores.** A través de un proceso iterativo y mediante el “*Alternating Least Squares*”, el sistema genera para cada objeto/sujeto de la muestra un número de dimensiones de la solución. Estas puntuaciones están estandarizadas (tienen media cero y desviación estándar uno).
- **Optimal Scores (Category Quantifications).** Las puntuaciones individuales en cada dimensión, promediadas para los individuos de cada categoría en las respectivas variables, son las que nos proporcionan “*Category Quantifications*”, que no son otra cosa sino los centroides de cada categoría en cada dimensión. Si trabajamos con una única dimensión, el análisis de homogeneidad asigna puntuaciones óptimas (y por tanto, cuantificaciones por categoría) de modo que la distancia entre estas categorías en la citada dimensión sea la máxima posible. Si son dos las dimensiones, trata de maximizar esta distancia en ambas dimensiones, y así sucesivamente para tres o más dimensiones.

Puesto que las distintas categorías de cada variable tienen, por tanto, tantas puntuaciones como dimensiones solicitadas en la solución, las citadas variables reciben el nombre de “Múltiple Nominal” a nivel de medida.

Si el análisis de homogeneidad (HOMALS) se realiza exclusivamente con dos variables, esto equivale a un análisis de correspondencias simple y aunque sus resultados no sean exactamente los mismos la interpretación con toda seguridad no va a diferir en uno y otro caso.

6.1. Escalamiento óptimo

El esquema de los procedimientos de escalamiento óptimo es el siguiente:



Fuente: Elaboración propia.

El escalamiento óptimo puede detectar relaciones no lineales y producir correlaciones máximas entre variables. Los cuatro procedimientos de escalamiento óptimo antes definidos amplían el ámbito de aplicación de las técnicas estadísticas clásicas de Análisis de componentes principales (ACP) y de Análisis de correlación canónica (ACC), para acomodar variables de niveles mixtos de medida. Si todas las variables del análisis fuesen numéricas y las relaciones entre las variables fuesen lineales, entonces deberían emplearse los procedimientos estadísticos estándares, basados en la correlación, y no habría necesidad de utilizar los procedimientos de escalamiento óptimo. Sin embargo, si las variables de análisis tienen niveles mixtos de medida, o si se sospecha que existen relaciones no lineales entre algunos pares de variables, entonces debería utilizarse el procedimiento de escalamiento óptimo.

En el escalamiento óptimo, el usuario especifica el tipo de medida de cada variable, diferenciando el nivel de medida de cada una de las variables del análisis, permitiendo así la búsqueda de soluciones con el fin de que las variables elegidas por el modelo se ajusten bien a los datos. El escalamiento óptimo también revelará relaciones no lineales.

El escalamiento óptimo proporciona un conjunto de puntuaciones óptimas (o cuantificaciones de categorías) para las categorías de cada variable. Las puntuaciones óptimas se asignan a las categorías de cada variable, basadas en el criterio de optimización del procedimiento en uso. A diferencia de los valores originales de las variables nominales u ordinales del análisis, estas se consideran frecuentemente como una forma de cuantificación de datos cualitativos. Las cuantificaciones de las categorías de cada variable pueden representarse sobre un plano bidimensional o, incluso, en un plano tridimensional, siendo la yuxtaposición en el mismo gráfico, útil para revelar patrones de asociación entre variables.

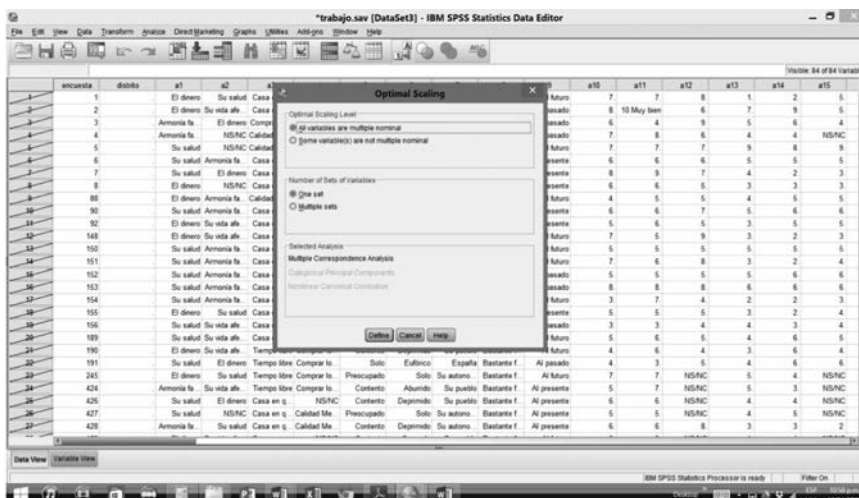
En SPSS suelen utilizarse cuatro procedimientos relacionados con la ejecución del Escalamiento Óptimo, que son los siguientes:

- **Análisis simples de correspondencias (ANACOR):** Ya visto anteriormente, el cual analiza datos de tablas de contingencia de dos dimensiones relativas a dos variables cualitativas.
- **Análisis múltiples de correspondencias u Homogeneidades (HOMALS):** Analiza datos de tablas de contingencia de múltiples dimensiones, donde todas las variables utilizadas son de nivel nominal y donde pueden ignorarse las interacciones de más dimensiones.
- **Análisis de componentes principales categóricos (CATPCA):** Contabiliza los patrones de variación en un solo conjunto de variables de niveles de medición mixtos.
- **Análisis no lineal de correlación canónica (OVERALS):** Contabiliza los patrones de variación cuando se correlacionan dos o más conjuntos de variables de niveles de medición mixta.

6.2. Aplicación mediante SPSS (HOMALS)

Para ilustrar el procedimiento con un ejemplo sencillo, seleccionar en el anexo *trabajo.sav* únicamente los individuos que residen en la Comunidad Valenciana ($c_{15}=17$) y tres variables del archivo (C_1 , C_3 y C_4).

Una vez cargado el archivo *trabajo.sav* y seleccionados los individuos de la Comunidad Valenciana ($c_{15} = 17$) entramos en Analyze/Data Reduction/Optimal Scaling y obtenemos el cuadro de diálogo siguiente:

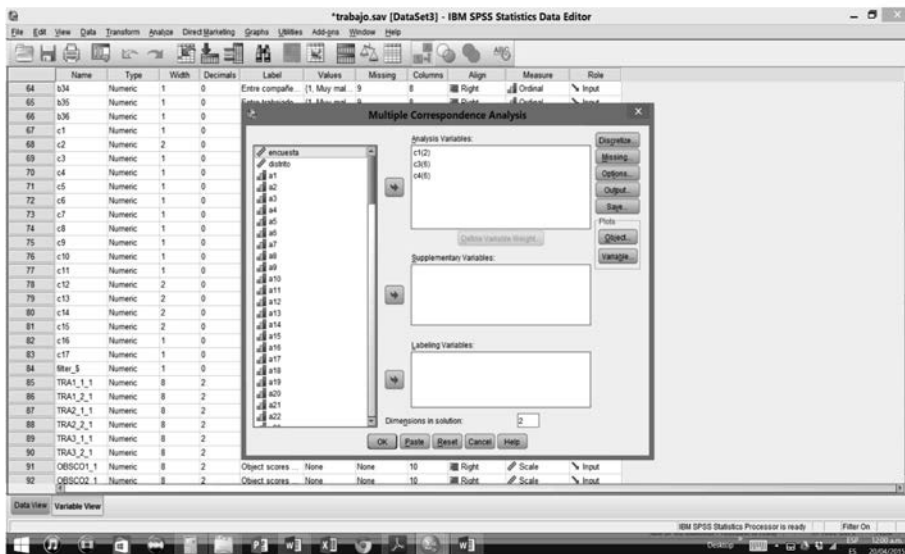


La combinación de los dos distintos niveles de medida y del número de series de variables que intervienen en el análisis es lo que determina el tipo de procedimiento de escalamiento óptimo a utilizar.

Dejamos seleccionadas las dos opciones por defecto, es decir, “All variables múltiple nominal” y “One set”, y pulsamos Define.

Hemos obtenido el análisis de homogeneidad porque en “Measurement Level” de la figura seleccionamos la opción “Al variables múltiples nominal” y en “Number of Sets of Variables” la opción “One Set”. En el primer caso asumimos que todas las variables a analizar puedan tener sus categorías cuantificadas y de modo distinto para cada dimensión, en tanto que en el segundo especificamos que trabajamos exclusivamente con un grupo de variables.

Luego accedemos al cuadro de diálogo principal del análisis de homogeneidad (HOMALS), que aparece en la imagen siguiente:



En este cuadro de diálogo ingresamos como variables a analizar:

c_1 (Sexo) Define Range Maximun: 2 Continue

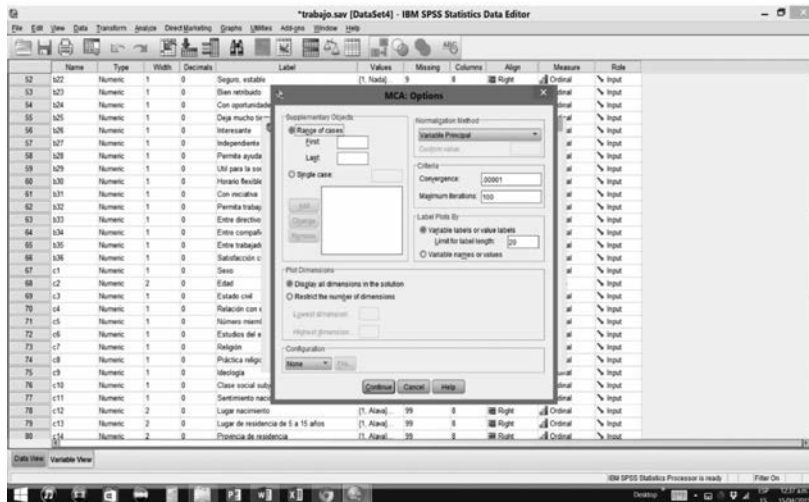
c_3 (Estado civil) Define Range Maximun: 6 Continue

c_4 (Relación con la cabeza de familia) Define Range Maximun: 6 Continue

Dejamos en blanco “Supplementary Variables”; dejamos dos dimensiones en la solución, tal y como aparece en la imagen. Luego pulsamos el botón “Options”.

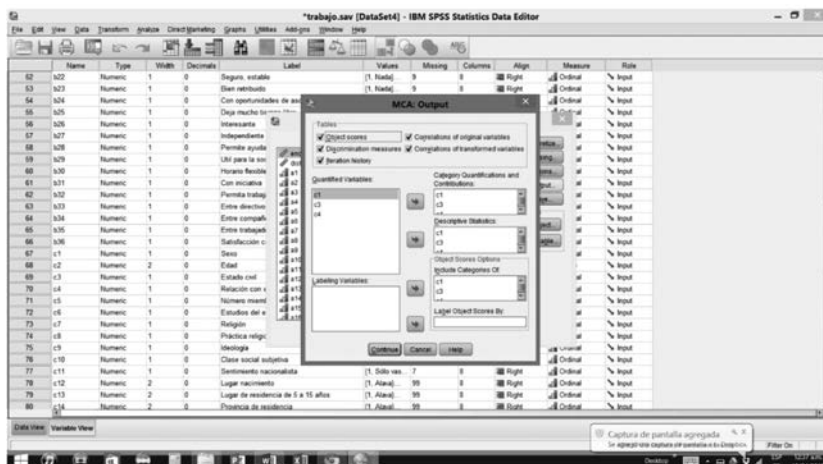
En el cuadro de diálogo de la figura anterior y al entrar cada variable para su análisis hemos debido definir su rango máximo; puesto que el mínimo, siempre por defecto, es igual a la unidad. Únicamente señalar que si en algún caso trabajamos con una variable cuyos valores no comienzan en la unidad o no son consecutivos, se puede proceder a recodificar la misma.

Ahora, entrar en el subcuadro de diálogo “Options” del HOMALS, que aparece en la imagen siguiente:

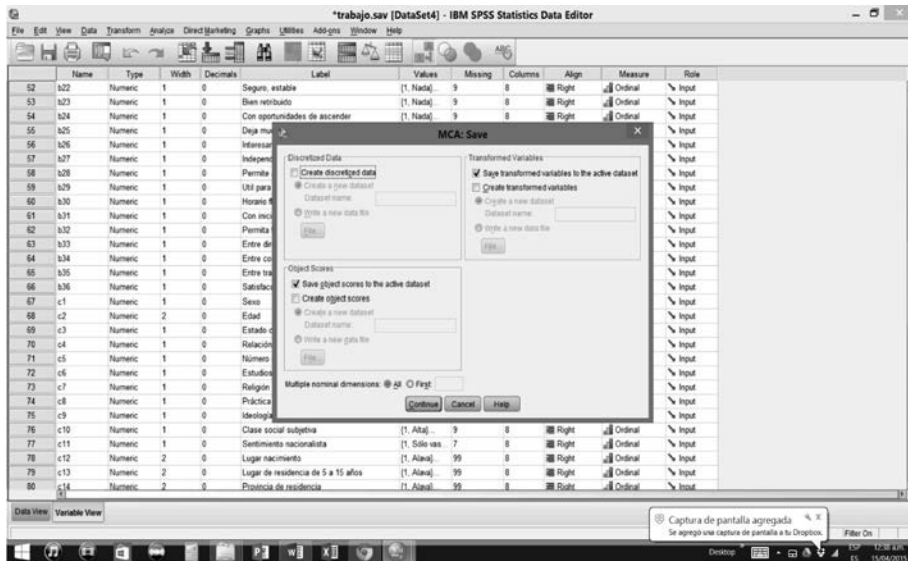


Seleccionamos todas las opciones de este subcuadro de diálogo y dejamos los valores por defecto en “Criterio”. Pulsamos Continúe.

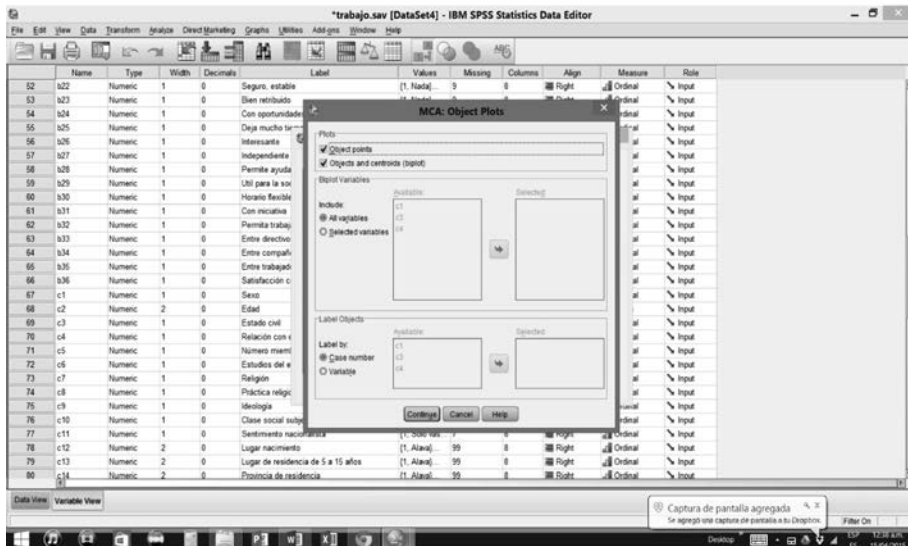
Luego, en el cuadro de diálogo Output, seleccionar todas las tablas y agregar variables a “CAtegory Quantifications” y en la lista de estadísticas descriptivas.

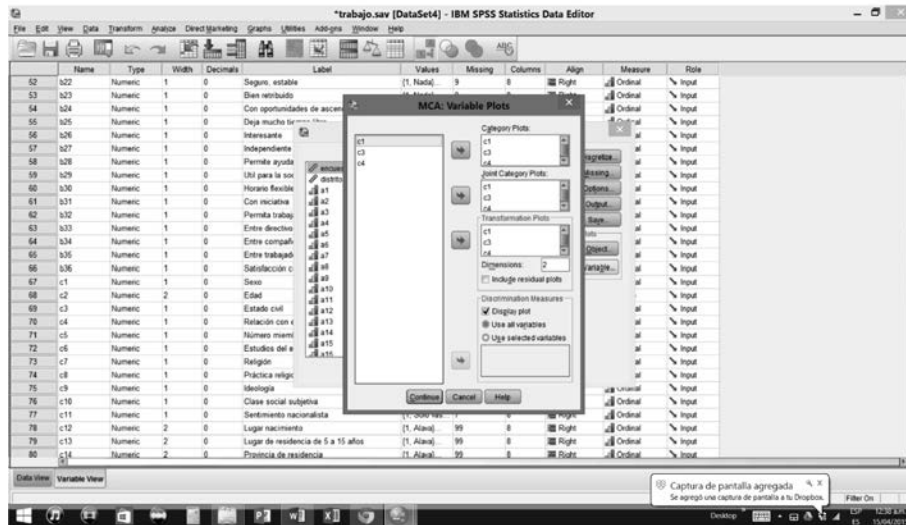


Asimismo, entrar al cuadro de diálogo Save y seleccionar los “Object Scores” y “Variables Transformadas”.



Finalmente, entrar al cuadro de diálogo Plots (Object y Variable) y seleccionar según las gráficas siguientes:





La sintaxis generada por el procedimiento HOMALS es el siguiente:

```
DATASET ACTIVATE DataSet1.
MULTIPLE CORRES VARIABLES=c1 c3 c4
  /ANALYSIS=c1(WEIGHT=2) c3(WEIGHT=6) c4(WEIGHT=6)
  /MISSING=c1(PASSIVE,MODEIMPU) c3(PASSIVE,MODEIMPU)
  c4(PASSIVE,MODEIMPU)
  /DIMENSION=2
  /NORMALIZATION=VPRINCIPAL
  /MAXITER=100
  /CRITITER=.00001
  /PRINT=CORR DESCRIPT(c1 c3 c4) HISTORY DISCRIM OBJECT(c1 c3 c4)
  OCORR QUANT(c1 c3 c4)
  /PLOT=BIPLOT(20) OBJECT(20) CATEGORY(c1 c3 c4) (20) JOINTCAT(c1 c3 c4)
  (20) TRANS(c1 c3 c4(2))
  (20) RESID(c1 c3 c4) (20) DISCRIM (20)
  /SAVE=OBJECT TRDATA.
```

Comentarios:

- Trabajamos con 121 individuos, tres variables y unas frecuencias marginales por categoría, que son las que aparecen al inicio de la tabla (57 hombres, 64 mujeres, 27 solteros, 75 casados, etc.)

Descriptive Statistics

Sexo

		Frequency
Valid	Hombre	57
	Mujer ^a	64
	Total	121

a. Mode.

Estado civil

		Frequency
Valid	Soltero	27
	Casado ^a	75
	Vive pareja	2
	Separado	6
	Divorciado	1
	Viudo	10
	Total	121

a. Mode.

Relación con el cabeza de familia

		Frequency
Valid	Cabeza ^a	58
	Esposa	39
	Hijo	23
	Otro	1
	Total	121

a. Mode.

Iteration History

Iteration Number	Variance Accounted For	
	Total	Increase
1	.052178	.052178
2	1.564983	1.512805
3	1.711181	.146198
4	1.770971	.059790
5	1.792994	.022023
6	1.800691	.007698
7	1.803361	.002669
8	1.804296	.000935
9	1.804629	.000333
10	1.804751	.000121
11	1.804796	.000045
12	1.804813	.000017
13 ^a	1.804820	.000007

a. The iteration process stopped because the convergence test value was reached.

- Si hemos seleccionado la opción “Iteration history” de la figura, nos aparecerá un resumen de las mismas en el output.
- Dimensiones. El número máximo de dimensiones en la solución será igual al número de categorías de las variables menos el número de variables.

Model Summary

Dimension	Cronbach's Alpha	Variance Accounted For		
		Total (Eigenvalue)	Inertia	% of Variance
1	.709	1.896	.632	63.209
2	.625	1.713	.571	57.112
Total		3.610	1.203	
Mean	.669 ^a	1.805	.602	60.161

a. Mean Cronbach's Alpha is based on the mean Eigenvalue.

- A pesar de ello, suelen utilizarse una, dos o tres dimensiones como máximo, dada la dificultad de interpretar un número superior. Por defecto, y tal y como hemos visto en la imagen, el sistema trabaja con dos dimensiones.
- Eigenvalue. Su significado es idéntico al ya comentado en el análisis factorial y en el análisis de correspondencias simple, y sus valores representan la proporción de información categórica explicada por cada una de las dimensiones. En el ejemplo, estos valores son 0.632 y 0.571 para las dimensiones uno y dos, respectivamente. El valor máximo para cada dimensión es igual a la unidad.

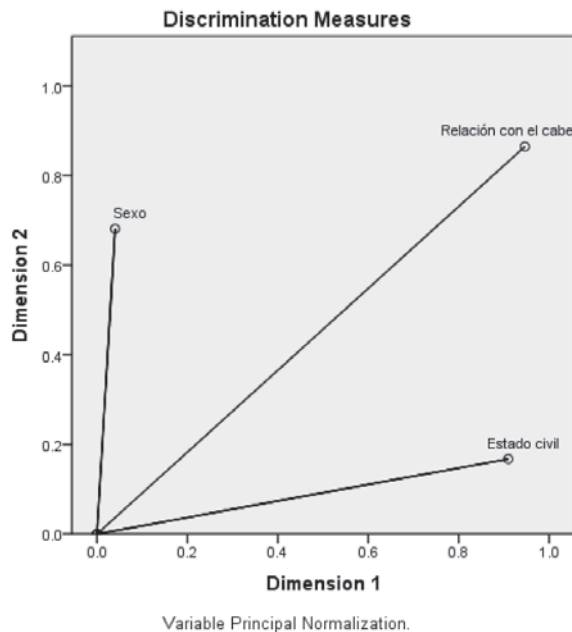
En la figura siguiente tenemos, en primer lugar, los valores de las medidas de discriminación para cada variable en cada dimensión y que equivalen a la varianza de las variables cuantificadas. Su valor máximo es igual a la unidad, que se consigue si todas las puntuaciones de los sujetos caen en grupos mutuamente excluyentes, y a su vez dentro de cada grupo estas puntuaciones son idénticas.

Discrimination Measures

Discrimination Measures

	Dimension		Mean
	1	2	
Sexo	.040	.681	.361
Estado civil	.910	.168	.539
Relación con el cabeza de familia	.946	.865	.906
Active Total	1.896	1.713	1.805
% of Variance	63.209	57.112	60.161

El gráfico siguiente responde a las medidas de discriminación previas, y confirma que en efecto la variable C_3 (Estado civil) está más relacionada con la primera dimensión, en tanto que la variable C_1 (Sexo) lo está con la segunda dimensión, C_4 (relación con el cabeza de familia) discrimina en ambas dimensiones. Si alguna variable estuviese próxima al origen, no discriminaría en ninguna de ambas dimensiones. Esta información se obtiene en el subcuadro de diálogo "Output" que hemos seleccionado, "Discrimination Measures", y en el cuadro de diálogo "Plot", respectivamente.



En la imagen siguiente podemos observar para cada variable, un resumen de frecuencias por categoría y la medida de las puntuaciones en cada una de las dos dimensiones para los individuos de cada categoría. Recordar que SPSS genera en la base de datos, tantas columnas de puntuaciones como dimensiones solicitadas en la solución. Así, 0.212 en la variable C_1 (Sexo), categoría uno (Hombres), dimensión uno, es la media de las puntuaciones en esta dimensión, de los 57 hombres de la muestra. Idéntica interpretación debemos dar al resto de promedios por categorías y dimensiones.

Output del HOMALS: Quantifications

Table

Sexo

Points: Coordinates

Category	Frequency	Centroid Coordinates	
		Dimension	
		1	2
Hombre	57	-.212	.875
Mujer	64	.189	-.779

Variable Principal Normalization.

Estado civil

Points: Coordinates

Category	Frequency	Centroid Coordinates	
		Dimension	
		1	2
Soltero	27	-1.752	-.423
Casado	75	.575	-.028
Vive pareja	2	.573	-.030
Separado	6	-.166	1.053
Divorciado	1	.000	2.557
Viudo	10	.404	.469

Variable Principal Normalization.

Relación con el cabeza de familia

Points: Coordinates

Category	Frequency	Centroid Coordinates	
		Dimension	
		1	2
Cabeza	58	.211	.943
Esposa	39	.840	-1.072
Hijo	23	-1.881	-.487
Otro	1	-1.744	-1.664

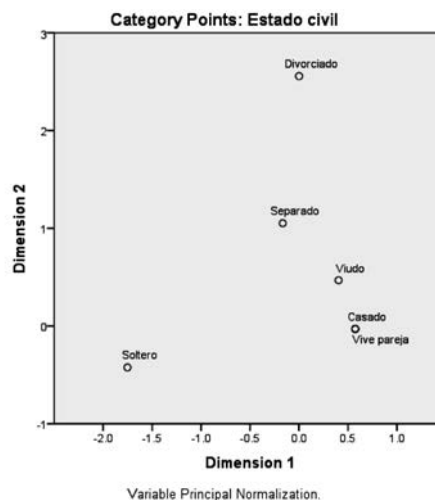
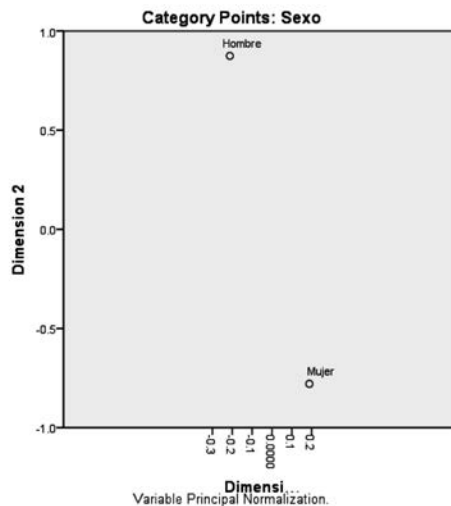
Variable Principal Normalization.

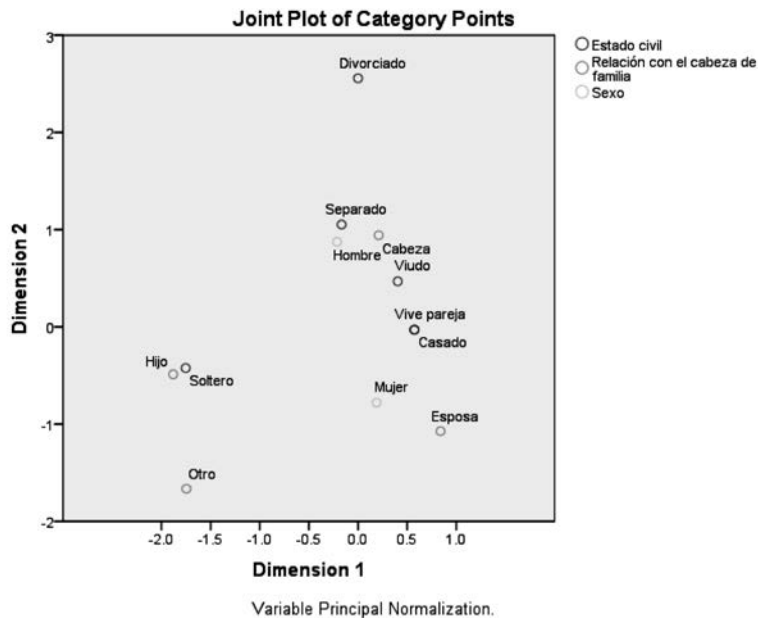
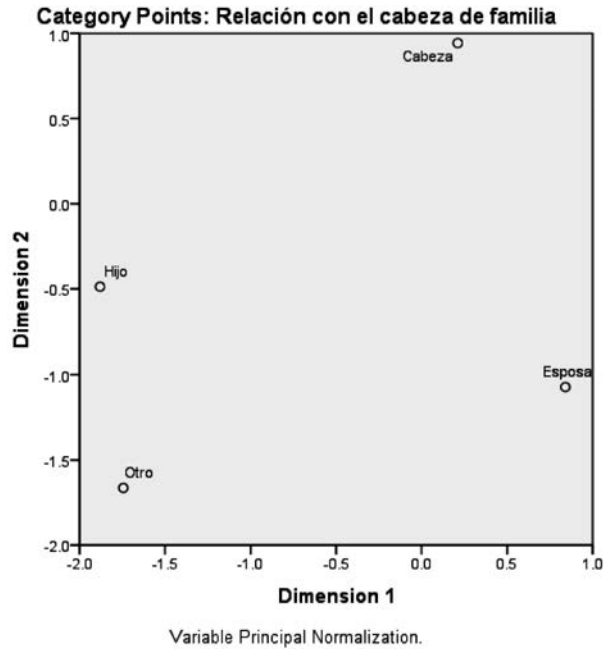
Recordar que la finalidad del HOMALS es encontrar cuantificaciones que sean óptimas en el sentido de que los centroides (medias) de cada categoría estén tan separados unos de otros como sea posible. Esto implica conseguir la máxima homogeneidad entre los sujetos dentro de cada categoría, y en cambio, que éstas

sean lo más heterogéneas posible, lo que se traducirá en una mayor separación en el gráfico. El nombre de Análisis de Homogeneidad responde justamente al primero de los argumentos.

Una aproximación distinta a estos resultados puede llevarse a cabo a través de los gráficos de las imágenes siguientes, los cuales se obtienen directamente en el output del HOMALS, si en el subcuadro de diálogo “Output” y “Plot” hemos seleccionado “Category Quantifications”. Los ejes están dimensionados a partir de las puntuaciones promedio de todas las categorías en todas las variables y para las dos dimensiones, reflejando los puntos del gráfico cada una de estas categorías.

Category Points





La lectura puede resultar difícil por el hecho de superponerse unas categorías a otras, por lo que se puede recurrir a representar en un Scatterplot cada variable por separado, con sus categorías y sus promedios respectivos en cada dimensión.

En otro de los gráficos podemos observar cómo en efecto la segunda dimensión discrimina bastante bien las dos categorías de la variable C_1 (Sexo); en otro gráfico se observa cómo la primera dimensión discrimina mejor las primeras seis categorías de la variable C_3 (Estado civil); finalmente, en el último gráfico, se observa que ambas dimensiones discriminan con bastante claridad las seis categorías de la variable C_4 (Relación con la cabeza de familia).

En las figuras siguientes podemos observar en primer lugar las puntuaciones de los objetos/sujetos en cada una de las dos dimensiones (se presenta únicamente a los primeros sujetos). Si en el subcuadro de diálogo "Save" hemos seleccionado la opción "Objects Scores", en la base de datos tendremos generadas dos nuevas columnas (OBSC01_1 y OBSC02_1) con estas puntuaciones para cada objeto/sujeto de la muestra y en las dimensiones uno y dos, respectivamente.

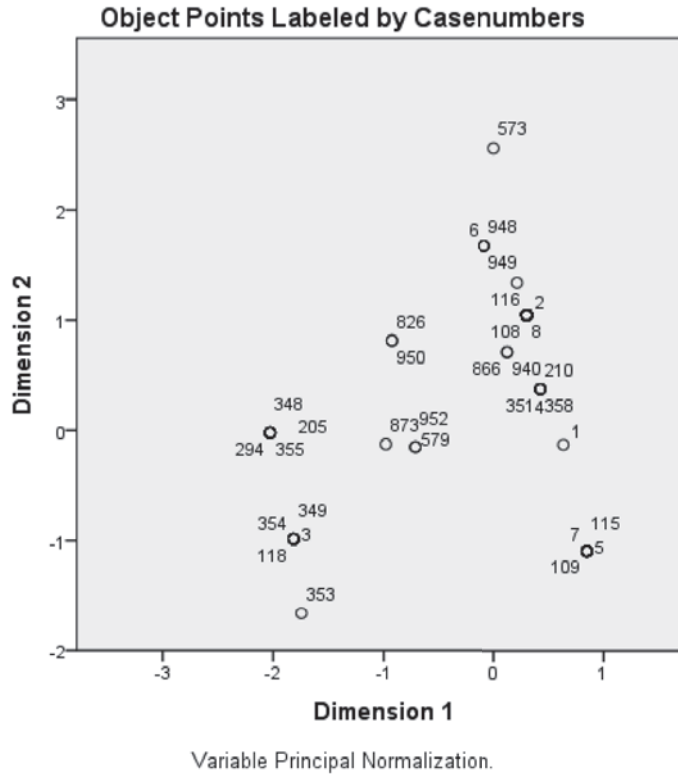
Objects

Object Scores

Case Number	Dimension		Sexo	Estado civil	Relación con el cabeza de familia
	1	2			
1	.634	-.132	Hombre	Casado	Esposa
2	.302	1.044	Hombre	Casado	Cabeza
3	-1.816	-.986	Mujer	Soltero	Hijo
4	.425	.372	Mujer	Viudo	Cabeza
5	.846	-1.097	Mujer	Casado	Esposa
6	-.088	1.674	Hombre	Separado	Cabeza
7	.846	-1.097	Mujer	Casado	Esposa
8	.302	1.044	Hombre	Casado	Cabeza
108	.302	1.044	Hombre	Casado	Cabeza
109	.846	-1.097	Mujer	Casado	Esposa
115	.846	-1.097	Mujer	Casado	Esposa
116	.302	1.044	Hombre	Casado	Cabeza
117	.302	1.044	Hombre	Casado	Cabeza
118	-1.816	-.986	Mujer	Soltero	Hijo
119	.302	1.044	Hombre	Casado	Cabeza
120	.846	-1.097	Mujer	Casado	Esposa
204	.846	-1.097	Mujer	Casado	Esposa
205	-2.028	-.021	Hombre	Soltero	Hijo
993	.008	.170	.001	.009	.003
Active Total	1.000	3.000	1.000	1.000	

Variable Principal Normalization.

Object Points Labeled by



El gráfico de la figura anterior agrupa los individuos de la muestra, de acuerdo con sus puntuaciones en ambas dimensiones. Si la muestra es grande, este gráfico puede resultar lo suficientemente ilustrativo como para situar las mayores concentraciones de individuos en el eje bidimensional. Si, en cambio, la muestra es pequeña, puede recurrirse a la identificación de cada objeto/sujeto de acuerdo con una variable.

6.3. Ejercicios

Ejercicio 1

Mediante análisis de correspondencia múltiple analizar la relación entre las categorías de las variables, características recogidas en el anexo *cochesmul.sav*. Este fichero contiene información sobre 300 individuos con las siguientes características: País de origen, casa (propia, alquilada), ingresos, con/sin hijos y género.

Ejercicio 2

Utilizar el anexo *correspondencias2.sav*, que contiene los 50 primeros individuos del ejemplo anterior, al cual se le ha incluido la variable sexo, y aplicar el HOMALS.

Ejercicio 3

Utilizar el análisis de correspondencias múltiple para observar la relación entre la categoría laboral (catlab), la clasificación étnica (minoría) y el género (sexo) de los empleados de una empresa. Los datos se recogen en el anexo *empleado.sav*.

7. Análisis de correlación canónica no lineal

Introducción

El análisis de correlación canónica no lineal es un procedimiento muy general con numerosas aplicaciones diferentes. El objetivo del análisis de correlación canónica no lineal consiste en analizar las relaciones existentes entre dos o más conjuntos de variables (sets o series de conjunto de variables) en vez de entre las propias variables, como ocurre en el análisis de componentes principales. Por ejemplo, imaginemos que se tienen dos conjuntos de variables, donde un conjunto de variables puede contener elementos de información demográfica general en un conjunto de encuestados, mientras que un segundo conjunto puede contener las respuestas a un conjunto de elementos o ítems actitudinales, a una serie de preguntas de opinión, o indicadores económicos.

Los niveles de escalamiento del análisis pueden ser cualquier combinación de nominal, ordinal y numérico. El análisis de correlación canónica mediante escalamiento óptimo, determina la similitud existente entre los conjuntos mediante la comparación, de manera simultánea de las variables canónicas de cada conjunto, con un conjunto de puntuaciones asignadas a los objetos.

El análisis de correlación canónica lineal consiste en estudiar la relación entre dos series de variables, en la cual una variable es más importante dentro de la serie cuanto mayor información independiente del resto de variables proporcione al modelo. Para ello, el análisis de correlación canónica lineal lleva a cabo una combinación lineal de las variables de una primera serie y otra combinación lineal de variables de la segunda serie de modo que la relación entre ambas sea la máxima posible.

El OVERALS parte de esta misma filosofía, pues trata de explicar tanta variabilidad como sea posible, de las series de variables, a través de la correlación existente de las mismas. Lo que **diferencia** a OVERALS del procedimiento lineal es lo siguiente:

- OVERALS utiliza mínimos cuadrados alternativos, como criterio para la estimación de los parámetros (similar a HOMALS).³⁸
- Las variables no deben estar forzosamente medidas en una escala de intervalo o de razón, sino que pueden estarlo en escalas nominales, ordinales, de intervalo o de razón, y su relación no tiene que ser forzosamente lineal.
- OVERALS establece la similitud entre las series de variables, comparando simultáneamente una combinación lineal de las variables en cada serie, con una serie desconocida *a priori* (la formada por las puntuaciones de los objetos/sujetos generada por el sistema).

Así también, es importante comentar algunas **similitudes** entre este y otros procedimientos analizados hasta el momento.

- Si las variables son todas ellas nominales múltiple, los resultados del OVERALS serán equivalentes a los obtenidos con HOMALS.³⁹
- Con dos series de variables únicamente, si en una de ellas tenemos una variable nominal simple y en las otras, varias, OVERALS puede ser interpretado como un modelo de regresión múltiple con una variable dependiente y varias independientes.
- Si la variable de una de las series en lugar de ser nominal simple es nominal múltiple, OVERALS puede ser considerada de una técnica alternativa al análisis discriminante.
- Finalmente, si tenemos solo dos series de variables y todas ellas son numéricas, el OVERALS será equivalente al análisis de correlación canónica lineal. Aunque SPSS no tiene incorporado este procedimiento, la mayoría de los estadísticos más relevantes no pueden obtenerse a través del mismo y coinciden con los del análisis multivalente de la varianza.

38. Análisis múltiple de correspondencia.

39. Si hay dos o más conjuntos de variables con únicamente una variable por conjunto, el análisis de correlación canónica mediante escalamiento óptimo es equivalente al análisis de componentes principales mediante escalamiento óptimo. Si todas las variables, en un análisis de una variable por conjunto, son nominales múltiples, el análisis de correlación canónica mediante escalamiento óptimo es equivalente al análisis de correspondencias múltiple.

7.1. Aplicación en SPSS

- Cargar el anexo trabajo.sav en el cual trabajaremos con seis variables (a1, a3, a8, a9, c1 y c3). Inicialmente, recodificar la variable C3 en C3_2 (Estado civil recodificado, con tres categorías: solteros, casado y otros). Posteriormente, seleccionamos exclusivamente aquellos individuos de la muestra que han declarado tener ingresos mensuales en su hogar, valores superiores a \$2,700 (Select Cases: If C17 $\geq$ 9). Nos quedaremos justamente con nueve individuos para el análisis.
- Finalmente, entramos en Analyze/Data Reduction/Optimal Scaling, y de nuevo accedemos al cuadro de diálogo principal del escalamiento óptimo. Elijiendo indistintamente cualquiera de los dos niveles de medida,⁴⁰ luego seleccionar “Multiple sets” en “Number of Sets of Variables”, y luego hacemos clic en botón “Define” para entrar en el cuadro de diálogo principal del OVERALS.

Ahora, definir tres series de variables. En la primera, introducir las variables *preo* (aspecto que más le preocupa) y *satis* (aspecto con el que se siente más satisfecho). Antes de ejecutar el procedimiento debemos asignar un rango máximo y un nivel de escalamiento para cada variable en “Define Range and Scale”, el programa nos ofrece las siguientes opciones:

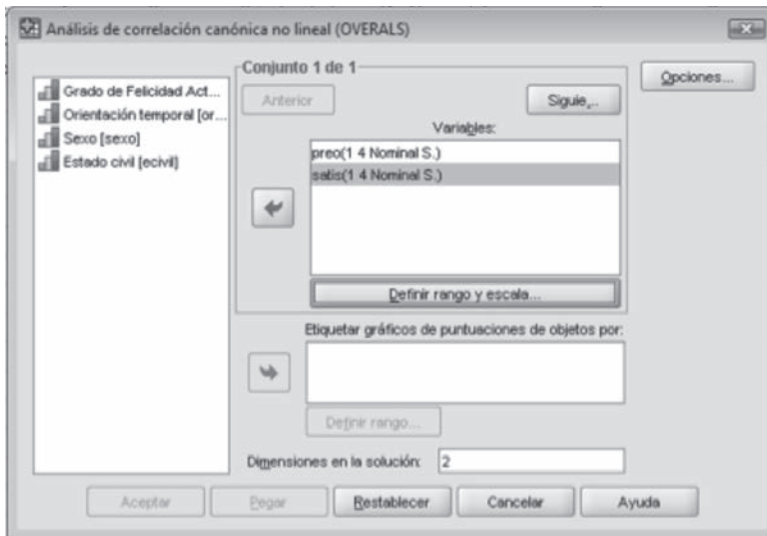
- **Ordinal.** El orden de las categorías de la variable observada se conserva en la variable cuantificada.
- **Nominal simple.** Los objetos que se encuentran en la misma categoría reciben la misma puntuación. Cuando todas las variables son nominales simples, la primera dimensión de esta solución es la misma que la primera del análisis de homogeneidad.
- **Nominal múltiple.** Las cuantificaciones pueden ser distintas para cada dimensión. Cuando todas las variables son nominales múltiples y hay solo una variable en cada conjunto, el análisis de correlación canónica genera los mismos resultados que el análisis de homogeneidad.
- **Numeración discreta.** Las categorías se tratan como si estuvieran ordenadas y espaciadas uniformemente. Las diferencias entre los números de las categorías y el orden de las categorías de la variable observada se conservan en la variable cuantificada. Cuando todas las variables están a nivel numérico y hay dos conjuntos, el análisis de correlación canónica es clásico.

40. En nuestro ejemplo, que incluye seis variables, todas son nominales a excepción del grado de felicidad actual, que será ordinal, por lo que elegiremos la opción “Alguna variable no es nominal múltiple”.

En nuestro ejemplo, asignamos el rango máximo y la escala de medida que se presenta en la tabla siguiente. Una vez introducidas y definidos su rango y escala de medida, pulsamos “Next” y podremos definir la segunda serie. Realizamos la misma operación con la tercera serie, y pulsando los botones “Previous” o “Next” podremos ir visualizando los contenidos de cada serie.

Serie	Variables	Etiqueta	Rango máximo	Escala de medida
1ª	(a1) preo	Aspecto que más le preocupa	4	Single Nominal
	(a3) satis	Aspecto con el que se siente más satisfecho	4	Single Nominal
2ª	(a8) feli	Grado de felicidad actual	4	Ordinal
	(a9) orie	Orientación temporal	3	Single Nominal
3ª	(c1) sexo	Sexo	2	Single Nominal
	(c3_2) ecivil	Estado civil recodificada	3	Single Nominal

En el campo “Label Objects Scores Plot(s) by” del cuadro de diálogo principal del OVERALS podemos introducir una variable que sirva posteriormente para identificar a los individuos en los gráficos. En el ejemplo hemos introducido la variable “encuesta”, que hace referencia al número de encuestado de cada sujeto en la base de datos. En total son 1200 individuos y, por tanto, esta variable tiene valores comprendidos entre 1 y 1200. En cuanto al número de dimensiones en la solución dejamos el valor por defecto (dos).

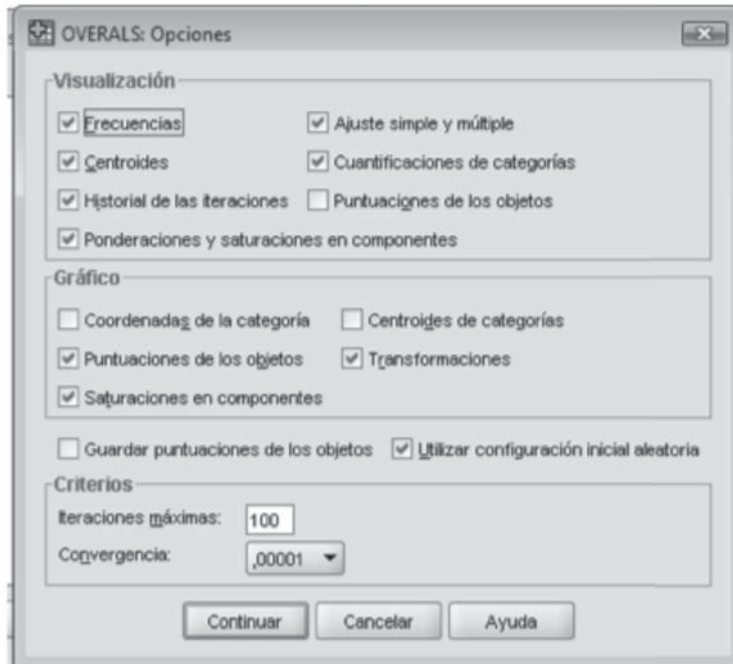


Habiendo seleccionado las variables de las tres series, las etiquetas y el número de dimensiones en el cuadro de diálogo principal del OVERALS, pulsamos el botón “Options” donde se puede elegir lo siguiente:

- **Mostrar:**
 1. Frecuencias marginales (los recuentos).
 2. Centroides.
 3. Historial de iteraciones.
 4. Ponderaciones y las saturaciones en las componentes.
 5. Las cuantificaciones de las categorías.
 6. Las puntuaciones de objeto: puntuación óptima que se asigna a un objeto (caso) en una dimensión particular.
 7. Los estadísticos de ajuste simple y múltiple.
- **Gráfico:** Para generar gráficos de las coordenadas de las categorías, las puntuaciones de objeto, las saturaciones en las componentes, los centroides de la categorías y las transformaciones.
- **Guardar puntuaciones de los objetos:** Guarda las puntuaciones de los objetos como nuevas variables en el archivo de datos. Las puntuaciones de objeto se guardan para el número de dimensiones especificadas en el cuadro de diálogo principal.
- **Utilizar configuración inicial aleatoria:** Se utiliza una configuración inicial aleatoria en el caso de que todas o algunas de las variables sean nominales simples. Si esta opción no se selecciona, se utiliza una configuración inicial anidada.
- **Criterios:** Especificamos el número máximo de iteraciones que el análisis de correlación canónica no lineal puede realizar durante los cálculos.

Para nuestro análisis indicaremos todas las opciones de las secciones de Mostrar y Gráfico y usar configuración inicial aleatoria.

- Seleccionamos todas las opciones y dejamos los valores por defecto, tal y como aparecen en la imagen. Pulsamos “Continue” y “OK” para ejecutar el procedimiento.



Sintaxis generada por el programa SPSS

DATASET ACTIVATE DataSet1.

OVERALS

/VARIABLES=preo(4) satis(4) feli(4) orie(3) sexo(2) ecivil(3) encuesta(99)

/ANALYSIS=preo(SNOM) satis(SNOM) feli(ORDI) orie(SNOM) sexo(SNOM)
ecivil(SNOM)

/SETS 3 (2, 2, 2)

/DIMENSION=2

/PRINT FREQ CENTROID HISTORY WEIGHTS FIT QUANT OBJECT

/PLOT QUANT OBJECT(encuesta) LOADINGS TRANS NDIM(1,MAX)

/MAXITER=100

/CONVERGENCE=.00001

/INITIAL=RANDOM.

El visor de resultados presenta una relación de las variables que intervienen en el procedimiento con sus respectivos niveles de medida y el número de categorías. Luego, una tabla con las frecuencias por categorías y los valores missing de cada variable y dentro de cada serie de variables. De los nueve individuos de la muestra hay dos con valores missing en la variable *a1* y uno en la variable *a3*, por lo que en total son tres en la serie. Por otro lado, tenemos los 0 de la tabla de frecuencias o categorías vacías. Tanto unos como otros (valores missing y categorías vacías) pueden influir enormemente en la solución del OVERALS, por lo que es sumamente importante constatar su existencia y tenerlos presentes a la hora de interpretar los resultados.

List of Variables			
Set		Number of Categories	Optimal Scaling Level
1	Preocupaciones: aspecto que más le preocupa	4	Single Nominal
	Satisfacción: aspecto con el que se siente más satisfecho	4	Single Nominal
2	Grado de Felicidad Actual	4	Ordinal
	Orientación temporal	3	Single Nominal
3	Sexo	2	Single Nominal
	Estado civil	3	Single Nominal

Marginal Frequencies

Set 1

Preocupaciones: aspecto que más le preocupa

	Marginal Frequency
Su vida afectiva	1
El dinero	0
Armonía familiar	1
Su salud	4
Missing	2
Missing within the set	3

Satisfacción: aspecto con el que se siente más satisfecho

	Marginal Frequency
Tiempo libre	5
Casa en que vive	0
Calidad Med.Ambiente	1
Comprar lo que quiere	0
Missing	1
Missing within the set	3

Set 2

Grado de felicidad actual

	Marginal Frequency
Nada feliz	0
Poco feliz	3
Bastante feliz	6
Muy feliz	0
Missing	0
Missing within the set	0

Orientación temporal

	Marginal Frequency
Al pasado	1
Al presente	5
Al futuro	3
Missing	0
Missing within the set	0

Set 3

Sexo

	Marginal Frequency
Hombre	5
Mujer	4
Missing	0
Missing within the set	0

Estado civil

	Marginal Frequency
Soltero	3
Casado	5
otro	1
Missing	0
Missing within the set	0

En la tabla siguiente se tienen los valores de ajuste del modelo, es decir, en qué medida el análisis de correlación canónica ajusta los datos en relación a la asociación entre las diferentes series de variables. El valor máximo de ajuste y equivalente a una correlación perfecta, será siempre igual al número de dimensiones de la solución, **dos** en este caso, por lo que el ajuste es de **1.771**, muy cercano a dos (hay un valor de $0.229=2-1.771$ de desajuste) por lo tanto es un **buen ajuste**.

Los autovalores por dimensión expresan lo mismo, solo que por ejes, y representan en qué medida cada dimensión refleja la relación entre las series y equivalen al valor total del ajuste del modelo, por lo que el **primer eje** ajustará 0.990 sobre uno (desajuste de 0.010) y **el segundo**, de 0.781 sobre 1 (desajuste de 0.219). Ver imagen siguiente:

Summary of Analysis

1	Dimension		Sum
	2		
Set 1	.009	.175	.184
Set 2	.008	.216	.224
Set 3	.014	.265	.278
Mean	.010	.219	.229
Eigenvalue	.990	.781	
Fit			1.771

7.2. Ponderaciones, cuantificaciones y saturaciones

Los valores propios por dimensión tienen la misma interpretación que el procedimiento HOMALS. Representan en qué medida cada dimensión refleja esa relación entre series de variables y equivalen al valor total de ajuste del modelo. Sumados a la medida de desajuste por dimensión deben darnos la unidad.

$$1.^a \text{ dimensión: } 0.990 + 0.010 = 1$$

$$2.^a \text{ dimensión: } 0.781 + 0.219 = 1$$

En la figura siguiente, tenemos en primer lugar los coeficientes de regresión (*weights*) para cada dimensión entre cada una de las variables cuantificadas de la serie y las puntuaciones de los objetos/sujetos generadas por el sistema. Estos pesos nos dan una idea de la contribución de cada variable a cada dimensión y siempre dentro de la serie a que pertenece.

Weights

Set		Dimension	
		1	2
1	Preocupaciones: aspecto que más le preocupa	.166	-.730
	Satisfacción: aspecto con el que se siente más satisfecho	.118	.748
2	Grado de felicidad actual	.195	-.995
	Orientación temporal	1.242	-.084
3	Sexo	.126	-1.037
	Estado civil	-1.250	.343

La cuantificación de cada una de las categorías de las variables se presenta en la siguiente tabla, y estas cuantificaciones son asignadas a cada sujeto, de acuerdo con la categoría a la que pertenece el mismo dentro de la variable. En la regresión múltiple de las puntuaciones de los objetos/sujetos actúan, como variable dependiente, y las cuantificaciones de las variables de la serie, como independientes, todo ello por separado para cada dimensión. De ese modo, los coeficientes de regresión o pesos pueden darnos una idea de la contribución de cada variable de la serie a la dimensión correspondiente.

Quantifications.

Set 1

Preocupaciones: aspecto que más le preocupa^a

	Marginal Frequency	Quantification	Single Category Coordinates		Multiple Category Coordinates	
			Dimension		Dimension	
			1	2	1	2
Su vida afectiva	1	-.778	-.129	.568	-.109	.572
El dinero	0	.000				
Armonía familiar	1	-.579	-.096	.423	-.114	.419
Su salud	4	1.419	.236	-1.037	.237	-1.036
Missing	3					

a. Optimal Scaling Level: Single Nominal

Set 2

Satisfacción: aspecto con el que se siente más satisfecho^a

	Marginal Frequency	Quantification	Single Category Coordinates		Multiple Category Coordinates	
			Dimension		Dimension	
			1	2	1	2
Tiempo libre	5	1.186	.140	.887	.145	.887
Casa en que vive	0	.000				
Calidad Med. Ambiente	1	1.401	.166	1.048	.147	1.051
Comprar lo que quiere	0	.000				
Missing	3					
a. Optimal Scaling Level: Single Nominal						

Set 3

Grado de Felicidad Actual^a

	Marginal Frequency	Quantification	Single Category Coordinates		Multiple Category Coordinates	
			Dimension		Dimension	
			1	2	1	2
Nada feliz	0	.000				
Poco feliz	3	-1.263	-.246	1.256	-.246	1.256
Bastante feliz	6	.838	.164	-.834	.164	-.834
Muy feliz	0	.000				
Missing	0					
a. Optimal Scaling Level: Ordinal						

En la tabla siguiente tenemos las proyecciones de las variables cuantificadas en el espacio de las puntuaciones de los objetos/sujetos de la muestra. En el supuesto de no existir valores missing en los datos (cosa que no ocurre en el ejemplo), estas ponderaciones de cada componente son equivalentes al coeficiente de correlación de Pearson entre las variables cuantificadas y las puntuaciones de los objetos/sujetos en cada dimensión.

Component Loadings

Set		Dimension	
		1	2
1	Preocupaciones: aspecto que más le preocupa ^{a,b}	.232	-.314
	Satisfacción: aspecto con el que se siente más satisfecho ^{a,b}	.211	.342
2	Grado de felicidad actual ^{b,c}	-.128	-.973
	Orientación temporal ^{a,b}	1.192	.175
3	Sexo ^{a,b}	-.382	-.898
	Estado civil ^{a,b}	-1.199	-.078

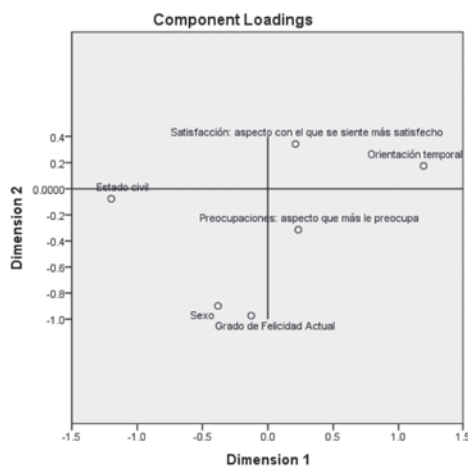
a. Optimal Scaling Level: Single Nominal

b. Projections of the Single Quantified Variables in the Object Space

c. Optimal Scaling Level: Ordinal

La imagen siguiente presenta estas ponderaciones de cada componente reflejadas en un gráfico con las dos dimensiones. Hemos trazado una línea desde el origen del espacio bidimensional hasta cada variable, de modo que la longitud de la misma puede darnos una idea de la importancia de las variables en la solución OVERALS. La relación entre las seis variables del ejemplo aparece con bastante claridad.

- Las variables C1 (Sexo) y a8 (grado de felicidad actual) opuestas a la variable a3 (Aspecto con el que se siente más satisfecho).
- Por otro lado, c3_2 (Estado civil recodificada) opuesta a a9 (Orientación temporal).
- a1 (Aspecto que más le preocupa) es la variable más alejada de cualquiera de estas tendencias.



Ajuste

En la tabla siguiente se observa el ajuste múltiple, el ajuste simple y el “desajuste” simple global y separado por dimensiones para cada variable. Con los resultados de ajuste podremos deducir cuáles son las variables que mejor discriminan y en qué dirección, cuando la contribución de las otras variables de la serie ha sido neutralizada. En el ejemplo anterior, las variables que más discrimina es C3_2 (estado civil recodificada) con un valor total 1.853; y la que menos discrimina es a1 (aspecto que más le preocupa), con un valor 0.561. Pero, así como C3_2 discrimina en gran medida en la primera dimensión (1.574), a1 lo hace en la segunda dimensión (0.533). El desajuste simple (Single Loss) es la diferencia entre los ajustes múltiple y simple, de modo que, si sus valores son cero o próximos, como ocurre en el ejemplo, excepto para las variables C3_2, significa que el nivel de medida asignado a las variables ha sido el correcto, siendo recomendable tratarlas como “múltiple nominal” si estos valores son muy elevados.

Fit

Set		Multiple Fit			Single Fit			Single Loss		
		Dimension		Sum	Dimension		Sum	Dimension		Sum
		1	2		1	2		1	2	
1	Preocupaciones: aspecto que más le preocupa ^a	.028	.533	.561	.028	.533	.561	.000	.000	.000
	Satisfacción: aspecto con el que se siente más satisfecho ^a	.014	.560	.574	.014	.560	.574	.000	.000	.000
2	Grado de Felicidad Actual ^b	.038	.990	1.028	.038	.990	1.028	.000	.000	.000
	Orientación temporal ^a	1.543	.007	1.551	1.543	.007	1.550	.000	.000	.000
3	Sexo ^a	.016	1.076	1.092	.016	1.076	1.092	.000	.000	.000
	Estado civil ^a	1.574	.279	1.853	1.562	.118	1.680	.012	.161	.174

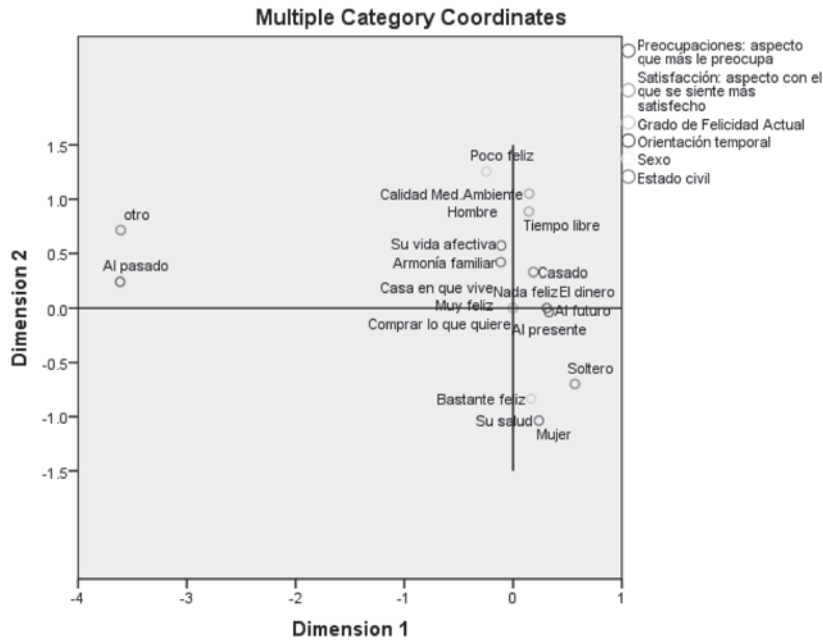
a. Optimal Scaling Level: Single Nominal

b. Optimal Scaling Level: Ordinal

7.3. Cuantificaciones

Con cuantificaciones para cada variable obtenemos las puntuaciones de los individuos frente a cada variable, que no son más que la media de las puntuaciones en esa dimensión por cada variable. También se obtiene el gráfico de todas las modalidades de donde también se podrán obtener conclusiones, como que el Eje dos contrapone a las mujeres, felices, solteras y que se preocupan por su salud frente a los que se sienten satisfechos con calidad ambiental y disfrutan del tiempo libre. Por otra parte, el Eje uno contrapone a los que piensan en el futuro y en el presente frente a los que piensan en el pasado.

Los resultados, tanto gráficos como numéricos, relativos a los centroides, tanto globales como individuales, se pueden interpretar en el mismo sentido, es decir, en qué grado cada variable discrimina a los grupos de individuos.



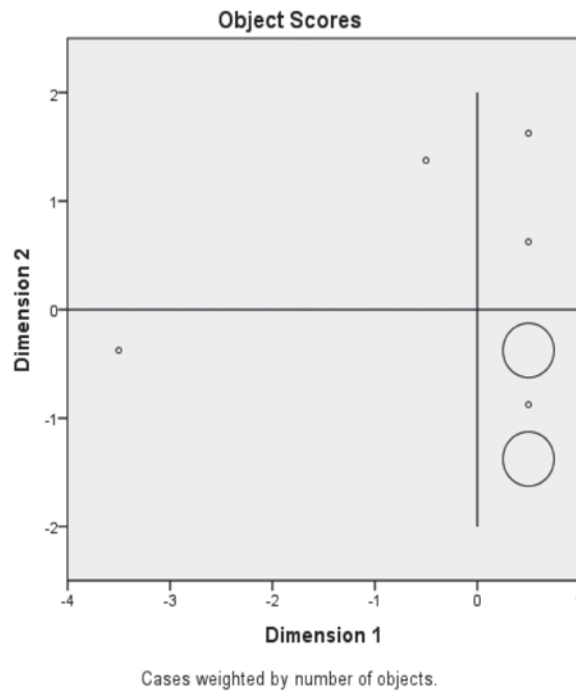
En la tabla que viene se presentan las puntuaciones de los objetos/sujetos en cada una de las dos dimensiones y su representación gráfica. Si, como en el ejemplo, encontramos algún valor extremo deberemos tomar en cuenta que el mismo puede “dominar” la solución y falsear los resultados globales del análisis. En estos casos se puede adoptar una doble solución:

- Repetir el proceso eliminando dichos sujetos de la muestra y contrastar los resultados.
- Procediendo a recodificar (uniendo o combinando categorías) aquellas variables en las que observamos que el citado sujeto puede tener valores o respuestas extremas.

En concreto, el valor extremo del ejemplo corresponde al individuo número 998 de la muestra, una mujer (C1) viuda (C3_2) preocupada por su salud (a1), valor missing en la variable aspecto con el que se siente más satisfecha (a3), que se declara bastante feliz (a8) y que tiene una orientación al pasado (a9).

Object Scores

	Dimension	
	1	2
1	.369	-.996
2	.147	.776
3	.052	1.467
4	.687	-1.507
5	.493	-.188
6	.375	-1.399
7	-3.450	-.591
8	.501	-.188
9	.031	1.460



En los demás resultados del Visor del OVERALS cabe destacar los gráficos correspondientes a los centroides o centroides proyectados, tanto globales para todas las variables como individuales para cada una de ellas por separado. Estos gráficos son equivalentes a los de “Quantifications” del HOMALS (vistos anteriormente) y pueden ser interpretados en el mismo sentido, es decir, en qué grado cada variable discrimina los grupos de individuos, tomando en cuenta que los centroides no son otra cosa sino los valores medios de las puntuaciones de los objetos/sujetos en cada dimensión y para cada una de las categorías de las variables.

Finalmente, aunque no está incluido en el Visor del OVERALS, puesto que sus valores están directamente relacionados con los valores propios del procedimiento, conviene señalar como estadístico de interés el coeficiente de correlación canónico entre series de variables.

Para dos series de variables este coeficiente será:

$$\rho_d = 2 * E_d - 1$$

Donde:

d : es el número de dimensión, y

E : es el valor propio de dicha dimensión.

Si son tres o más las series de variables, como es este ejemplo:

$$\rho_d = ((K * E_d) - 1) / (K - 1)$$

Donde:

d : es el número de dimensión,

K : es el número de series del ejercicio, y

E : es el valor propio de dicha dimensión.

En el ejemplo anterior:

$$\rho_1 = ((3 * 0.9990) - 1) / 3 - 1 = 0,9850 \quad \text{Primera dimensión}$$

$$\rho_2 = ((3 * 0.781) - 1) / 3 - 1 = 0,6715 \quad \text{Segunda dimensión}$$

7. 4. Ejercicios

Ejercicio 1

El anexo Canonica2.sav contiene los datos para 32 individuos de seis variables, las cuales son agrupables en tres conjuntos: el primero, por Ánimo y felicidad; el segundo, por Ideología y residencia; y uno tercero, por Edad y sexo. Realizar sobre estos datos un análisis de correlación canónico.

Ejercicio 2

El anexo *correlacionca2.sav* contiene los resultados de una encuesta en la que se medía el grado de acuerdo o desacuerdo (cada una con cinco niveles) con seis afirmaciones. Las respuestas se codifican en las variables p1 a p6. Realizar un análisis de correlación canónica no lineal tomando como primer conjunto las preguntas uno, dos y tres, y como segundo grupo las preguntas cuatro, cinco y seis.

Ejercicio 3

Utilizar el anexo *trabajo.sav* y seleccionar, de los 1,200 sujetos de la muestra, únicamente aquellos que han manifestado residir en la comunidad de Murcia (muestra de 32 sujetos).

- a) Proceder a recodificar las variables C2 (edad) y C3 (estado civil) en C2_1 (edad codificada) y C3_1 (estado civil recodificada), de acuerdo con los siguientes criterios:

	Categorías	Condición	Frecuencias
C2_1	1	Menos de 25 años	6
	2	Entre 26 y 40	6
	3	Entre 41 y 65	17
	4	Más de 65	3
C3_1	1	Soltero	9
	2	Casado	19
	3	Otros	4

- b) De la muestra de 32, seleccionar exclusivamente aquellos individuos con valores no missing en ninguna de las variables (quedando una muestra de 17 como sigue).

Sujeto	Encuesta	a ₅	a ₈	c ₉	c ₁₆	c _{2_1}	c _{3_1}
1	391	4	3	6	6	3	2
2	392	1	4	2	6	4	3
3	394	7	4	6	6	1	1
4	484	1	2	2	7	1	1
5	485	1	3	6	7	2	2
6	486	1	3	6	7	3	3
7	487	4	3	1	7	1	1
8	488	1	3	2	7	2	2
9	490	6	3	6	7	2	2
10	497	3	4	4	7	2	2

Sujeto	Encuesta	a_5	a_8	c_9	c_{16}	$c_{2,1}$	$c_{3,1}$
11	834	1	3	2	4	2	2
12	835	7	4	6	4	2	2
13	836	3	2	6	4	2	2
14	839	7	2	2	4	1	1
15	848	2	2	4	4	1	1
16	1036	4	3	4	4	1	1
17	1168	4	3	2	7	3	3

c) Proceder a realizar un OVERALS con las siguientes series de variables.

Serie	Variables	Etiquetas	Rango máximo	Escala de media
1 ^a	a_5	Estado de ánimo más frecuente	7	Single nominal
	a_8	Grado de felicidad	4	Single nominal
2 ^a	c_9	Ideología	7	Single nominal
	c_{16}	Hábitat	7	Single nominal
	9. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 10. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 11. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 12. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 13. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 14. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 15. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 16. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables. 17. Se le pide al sujeto que realice un OVERALS con las siguientes series de variables.			
3 ^a	$c_{2,1}$	Edad codificada	4	Single nominal
	$c_{3,1}$	Estado civil recodificado	3	Single nominal

Trabaje con dos dimensiones y determine el ajuste del modelo y los valores propios por dimensión. Interprete los resultados.

8. Análisis de correlación canónica lineal

Introducción

El análisis de correlación canónica es debido a Hotelling. Este investigador estudió en 1936 la relación entre resultados de test de capacidad intelectual y medidas físicas de un grupo de personas. Hotelling pretendía investigar las relaciones entre ambos conjuntos de variables y conocer cuántas dimensiones independientes tenía la relación existente entre ellas. En general, la correlación canónica se utiliza cuando un conjunto de variables multivariantes puede dividirse en dos grupos homogéneos (por criterios económicos, demográficos, sociales, etc.) y se desea estudiar la relación entre ambos conjuntos de variables. En particular, los dos grupos pueden corresponder a las mismas variables medidas en dos momentos distintos en el tiempo y el espacio.

Supongamos que disponemos de un conjunto de datos de n individuos y k variables que pueden subdividirse en dos grupos: el primero incluye p variables y el segundo q , donde $p+q=k$. Por ejemplo, las primeras p variables representan las inversiones realizadas por una empresa y las restantes q variables representan distintas medidas de beneficios. Es posible que $p=q$. Por ejemplo, cuando medimos las p variables en el instante t y en el instante $t+1$. Llamaremos X a la matriz $n \times p$ que contiene los valores de las p primeras variables de n elementos (individuos, países, empresas, etc.) y llamaremos Y a la matriz $n \times q$ que contiene los valores de las q segundas variables medidas sobre esos mismos n elementos. Para medir la relación entre ambos conjuntos se busca una combinación lineal de las p primeras variables, que tenga la máxima correlación con una combinación lineal de las segundas variables. Es decir, definimos a:

$$x^* = Xa = \sum_{i=1}^p a_i x_i$$

como una combinación lineal de las variables del primer grupo, y:

$$y^* = X\beta = \sum_{i=1}^q \beta_i x_i$$

como una combinación lineal de las variables del segundo; se desea encontrar los vectores y tales que las nuevas variables y tengan máxima correlación.

Es posible que una vez encontrada la primera relación entre estas dos variables indicadoras no exista más relación entre ambos conjuntos de variables, y entonces decimos que toda la relación entre ambos conjuntos se resume en una dimensión. Para comprobarlo, podemos buscar una segunda variable indicadora del primer conjunto, que esté incorrelada con la primera, y que tenga correlación máxima con otra variable indicadora del segundo conjunto.

Procediendo de esta manera, podemos obtener $r=\min(p,q)$ relaciones entre variables indicadoras de ambos grupos, que pueden ordenarse por orden de importancia. Determinar el número de relaciones entre las variables (variables indicadoras incorreladas de un conjunto relacionadas con el otro) permite juzgar cuántas dimensiones distintas tiene la relación. En el análisis de correlación canónica ambos conjuntos de variables se consideran simétricamente; pero es posible que querramos explicar el conjunto de variables dependientes Y mediante las independientes X . El procedimiento habitual (regresión múltiple) es construir un indicador de las variables dependientes y relacionarlo con las independientes mediante una regresión múltiple. Este procedimiento es razonable cuando existe sólo una dimensión en la relación, pero puede ser engañoso si existen varias. El análisis de correlación canónica permite determinar cuántas dimensiones relevantes tiene la relación entre ambos conjuntos de variables.

En la mayoría de las aplicaciones, los parámetros son desconocidos y deben estimarse a partir de los datos. Si la distribución conjunta de las observaciones es normal multivariante, utilizar las distancias Mahalanobis estimadas suele dar buenos resultados y será óptimo con muestras grandes. Sin embargo, es frecuente que los datos disponibles para la clasificación no sean normales.

8.1. Fundamentación matemática del Análisis de correlación canónica

Tanto el Análisis de correlación canónica lineal (ACC) como el no lineal parten de conjuntos de variables que pueden agruparse según criterios teóricos: variables sociodemográficas, indicadores económicos, conjunto de ítems, preguntas de opinión. Siendo otra técnica de reducción de dimensión del problema, es decir el objetivo, es la búsqueda de estructuras de relación entre variables.

El Análisis de correlación canónica lineal estudia la relación entre dos series de variables, siendo una variable más importante dentro de la serie cuanto más información independiente del resto de variables aporte al modelo. Para ello el análisis de correlación canónica lineal lleva a cabo una combinación lineal de las variables de la primera serie y otra de la segunda serie, de manera que la relación entre ambas sea al máximo posible.

La idea es determinar el par de combinaciones lineales que tiene la correlación más alta, luego el segundo par cuya correlación es menor o igual a la primera, y así sucesivamente. A estos pares de combinaciones lineales se los denomina *variables canónicas*; por consiguiente, sus correlaciones son llamadas *correlaciones canónicas*, las mismas que miden la fuerza de asociación entre los dos grupos de variables.

El primer grupo de variables está representado por el vector aleatorio p variado $X^{(1)}$, el segundo grupo de q variables es representado por $X^{(2)}$, donde el primer conjunto de variables tiene menos componentes que el segundo, es decir $p \leq q$.

Para los vectores aleatorios $X^{(1)}$ y $X^{(2)}$ se cumplen las siguientes propiedades:

$$\begin{aligned} E(\mathbf{X}^{(1)}) &= \boldsymbol{\mu}^{(1)} & Cov(\mathbf{X}^{(1)}) &= \boldsymbol{\Sigma}_{11} \\ E(\mathbf{X}^{(2)}) &= \boldsymbol{\mu}^{(2)} & Cov(\mathbf{X}^{(2)}) &= \boldsymbol{\Sigma}_{22} \\ Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) &= \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' \end{aligned}$$

Cada par de variables canónicas es representado como una combinación lineal de las variables tanto del grupo $X^{(1)}$ y $X^{(2)}$.

El vector aleatorio y el vector de medias:

$$\mathbf{X}_{((p+q) \times 1)} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} \quad \boldsymbol{\mu} = E[\mathbf{X}] = \begin{bmatrix} E(\mathbf{X}^{(1)}) \\ \dots\dots\dots \\ E(\mathbf{X}^{(2)}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \dots\dots\dots \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}$$

Donde la matriz de covarianzas es esta:

$$\begin{aligned} \Sigma_{(p+q) \times (p+q)} &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\ \Sigma &= \begin{bmatrix} E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & \vdots & E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \\ \dots & \dots & \dots \\ E(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & \vdots & E(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \end{bmatrix} \\ \Sigma &= \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \dots & \dots & \dots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix} \end{aligned}$$

$\begin{matrix} (pxp) & & (pxq) \\ & & \\ (qxp) & & (qxq) \end{matrix}$

Las covarianzas entre dos vectores de diferentes conjuntos, es decir de un vector $X^{(1)}$ y de un vector $X^{(2)}$, están contenidas en Σ_{12} o su transpuesta Σ_{21} . Cuando se tiene gran cantidad de variables, los cálculos de las varianzas se vuelven rutinarios; por ese motivo la técnica de correlación canónica tiene como objetivo resumir las asociaciones entre los conjuntos de vectores de $X^{(1)}$ y $X^{(2)}$ en términos de unas pocas covarianzas debidamente escogidas, en lugar de las pq covarianzas contenidas dentro de la matriz de covarianzas Σ_{12} .

Consideremos las siguientes combinaciones lineales:

$$U = \mathbf{a}'\mathbf{X}^{(1)}$$

$$V = \mathbf{b}'\mathbf{X}^{(2)}$$

Para un par de vectores coeficientes $\mathbf{a}$ y $\mathbf{b}$, se obtiene:

$$Var(U) = \mathbf{a}'Cov(\mathbf{X}^{(1)})\mathbf{a} = \mathbf{a}'\Sigma_{11}\mathbf{a}$$

$$Var(V) = \mathbf{b}'Cov(\mathbf{X}^{(2)})\mathbf{b} = \mathbf{b}'\Sigma_{22}\mathbf{b}$$

$$Cov(U, V) = \mathbf{a}'Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\mathbf{b} = \mathbf{a}'\Sigma_{12}\mathbf{b}$$

Se buscarán vectores de coeficientes tales que:

$$Corr(U, V) = \frac{\mathbf{a}'\Sigma_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\Sigma_{11}\mathbf{a}}\sqrt{\mathbf{b}'\Sigma_{22}\mathbf{b}}}$$

Por lo que se define así:

El primer par de variables canónicas, son las combinaciones lineales correspondientes a U_1, V_1 , que tienen varianzas unitarias y que maximizan la correlación entre ambas.

El segundo par de variables canónicas son las combinaciones lineales correspondientes a U_2, V_2 , que poseen varianzas unitarias y que maximizan la correlación previa entre todas las opciones no correlacionadas con el primer par de variables canónicas.

El i -ésimo par de variables canónicas, son las combinaciones lineales correspondientes al par U_i, V_i , el cual contiene varianzas unitarias y maximizan la correlación previa entre todas las opciones no correlacionadas con las $i-1$ variables canónicas.

Para los cálculos de los vectores a, b se toman en cuenta los siguientes resultados.

Supongamos que $p \leq q$ y que los vectores $X^{(1)}$ y $X^{(2)}$ tienen:

$$\begin{aligned} \text{Cov}(\mathbf{X}^{(1)}) &= \Sigma_{11} \\ \text{Cov}(\mathbf{X}^{(2)}) &= \Sigma_{22} \\ \text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) &= \Sigma_{12} = \Sigma_{21}' \end{aligned}$$

Para los vectores coeficientes a y b , que forman las combinaciones lineales:

$$\begin{aligned} U &= \mathbf{a}^t \mathbf{X}^{(1)} \\ V &= \mathbf{b}^t \mathbf{X}^{(2)} \end{aligned}$$

Entonces, para obtener U_1, V_1 es necesario:

$$\underset{a,b}{\text{MaxCorr}}(U, V) = \rho_1^*$$

Logrando el i -ésimo par de variables canónicas:

$$\begin{aligned} U_i &= \mathbf{e}_i' \Sigma_{11}^{-1/2} \mathbf{X}^{(1)} \\ V_i &= \mathbf{f}_i' \Sigma_{22}^{-1/2} \mathbf{X}^{(2)} \\ \text{con} \quad \text{Cov}(U_i, V_i) &= \rho_i^* \end{aligned}$$

Donde:

$$\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \geq \rho_p^{*2}$$

son los valores propios de la matriz, resultado de la multiplicación de:

$$\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2}$$

Además, $e_1, e_2, \dots, e_p$ son los vectores propios de $PX1$ asociados a ésta. Así mismo, se obtienen los p valores propios más grandes de la matriz:

$$\sum_{22}^{-1/2} \sum_{21} \sum_{11}^{-1} \sum_{12} \sum_{22}^{-1/2}$$

Donde:

$$\Sigma = E[(X - \mu)(X - \mu)^t] \in M_{(p+q)(p+q)}$$

con los correspondientes vectores propios $f_1, f_2, \dots, f_p$.

8.2. El análisis de relaciones mediante la correlación canónica

El análisis de correlación canónica es el método más generalizado de la familia de las técnicas estadísticas multivariantes. Se relaciona directamente con varios métodos de dependencia. Al igual que en la regresión, el objetivo de la correlación canónica es cuantificar la validez de la relación, en este caso entre los dos conjuntos de variables (dependientes e independientes).

Se asemeja al análisis factorial en la creación de compuestos de variables. También se parece al análisis discriminante en su capacidad para determinar las dimensiones independientes (igual que las funciones discriminantes) para cada conjunto de variables que produce la correlación máxima entre las dimensiones. De esta manera la correlación canónica identifica la estructura óptima o la dimensionalidad de cada conjunto de variables, que maximiza la relación entre los conjuntos de variables dependientes e independientes.

El ACC desarrolla varias funciones canónicas que maximizan la correlación entre las combinaciones lineales, también conocidas como valores teóricos canónicos, que son conjuntos de variables dependientes e independientes. Cada función canónica se basa en la correlación entre dos valores teóricos canónicos: un valor teórico para las variables dependientes y otro para las variables independientes. Otra característica única de la correlación canónica es que se obtienen los valores teóricos de forma que se maximice su correlación. Además, la correlación canónica no termina con la obtención de una relación simple entre los conjuntos de variables. En su lugar se pueden conseguir varias funciones canónicas (pares de valores teóricos canónicos).

8.3. Metodología del Análisis de correlación canónica

A continuación se presenta la metodología.

Paso 1: Objetivos del análisis de correlación canónica

Los datos apropiados para el análisis de correlación canónica son dos conjuntos de variables. Suponemos que a cada conjunto se le puede dar un significado teórico, al menos hasta el punto en que un conjunto pueda ser definido como las variables independientes y el otro como las variables dependientes. Una vez que se ha realizado esta distinción, la correlación canónica puede llevarse a cabo con un amplio rango de objetivos, que pueden ser algunos de los siguientes:

- a) Determinar si dos conjuntos de variables (medidas realizadas sobre los mismos objetivos) son independientes uno de otro o, inversamente, determinar la magnitud de las relaciones que pueden existir entre los dos conjuntos.
- b) Obtener un conjunto de ponderaciones para cada conjunto de variables criterio y variables predictoras, para que las combinaciones lineales de cada conjunto estén correlacionadas de forma máxima. Las funciones lineales adicionales que maximizan la restante correlación son independientes del conjunto o de los conjuntos anteriores de combinaciones lineales.
- c) Explicar la naturaleza de cualquiera de las relaciones existentes entre los conjuntos de variables criterio y variables predictoras, generalmente midiendo la contribución relativa de cada variable a las funciones canónicas (relaciones) que son extraídas.

Paso 2: Diseño de un análisis de correlación canónica

Al ser la manera más general del análisis multivariante, el ACC comparte las herramientas básicas comunes a todas las técnicas multivariantes.

Frecuentemente con la correlación canónica se deben resolver cuestiones acerca del impacto del tamaño muestral (tanto pequeño como grande) y la necesidad de una cantidad suficiente de observaciones por variable. Los investigadores pueden tener la tentación de incluir muchas variables tanto en el conjunto de variables independientes como el de dependientes, ignorando sus implicaciones en el tamaño muestral. Una sugerencia es mantener por lo *menos diez observaciones* por variable para evitar el “sobreajuste” de los datos.

La clasificación de las variables como dependientes o independientes tiene poca importancia en la estimación estadística de las funciones canónicas, ya que el análisis de correlación canónica pondera ambos valores teóricos para maximizar

la correlación y no establece ningún énfasis particular en alguno de los valores teóricos para maximizar esta correlación. La composición de cada valor teórico, ya sea dependiente o independiente, llega a ser muy importante. El investigador, antes de aplicar el análisis de correlación canónica, debe relacionar conceptualmente los dos conjuntos de variables. De esta forma, la especificación de los valores teóricos dependientes frente a los independientes es esencial para establecer una base conceptual fuerte para las variables.

Paso 3: Supuestos básicos de la correlación canónica

La generalidad del análisis de correlación canónica también se extiende a sus supuestos estadísticos básicos. El supuesto de *linealidad* afecta a dos aspectos de los resultados de la correlación canónica. Primero, el coeficiente de correlación entre cualesquiera de dos variables está basado en una relación lineal. Segundo, la correlación canónica es la relación lineal entre los valores teóricos. Si los valores teóricos se relacionan de una manera no lineal, la relación no será reflejada por la correlación canónica. De esta manera, aunque el análisis de correlación canónica es el método multivariante más extendido, está restringido a la identificación de relaciones lineales.

El análisis de correlación canónica puede emplear cualquier variable métrica sin que cumpla el estricto supuesto de normalidad. La *normalidad* es deseable porque estandariza una distribución que permite una mayor correlación entre las variables. Pero en un estricto sentido, el análisis de correlación canónica puede utilizar incluso variables no-normales si la forma de la distribución (ejemplo, altamente asimétrica) no disminuye la correlación con otras variables. Con esto se permite la utilización de datos no métricos transformados (en la forma de variables ficticias) también. Sin embargo, se requiere de normalidad multivariante para los contrastes de significación de inferencia estadística de cada función canónica. Dado que los contrastes de normalidad multivariante no están disponibles fácilmente, la línea a seguir que prevalece es asegurar que cada variable presenta normalidad univariante. De este modo, aunque estrictamente no se requiere normalidad, es altamente recomendable que se compruebe la normalidad de todas las variables y que se transformen si fuese necesario.

La *homocedasticidad* también debe ser estudiada, puesto que disminuye la correlación entre las variables. Finalmente, la *multicolinealidad* entre algún conjunto de variables distorsionará la capacidad de la técnica para aislar el impacto de cualquier variable única, haciendo que la interpretación sea menos fiable.

Paso 4: Obtención de las funciones canónicas y valoración del ajuste global

La obtención de sucesivos valores teóricos *es similar al procedimiento empleado en el análisis factorial sin rotación*. El primer factor extraído explica la máxima

cantidad de varianza en el conjunto de variables. Después se calcula el segundo factor para que explique lo más posible la varianza no explicada por el primer factor, y así sucesivamente, hasta que todos los factores han sido considerados. Por tanto, los posteriores factores se calculan a partir de los residuos o de la varianza restante de los primeros factores. El análisis de correlación canónica sigue un procedimiento similar pero centrándose en la explicación de la cantidad máxima de relación entre los dos conjuntos de variables, en lugar de un solo conjunto. El resultado es que el primer par de valores teóricos se calcula con el fin de obtener la mayor inter-correlación posible entre los dos conjuntos de variables. El segundo par de valores teóricos canónicos es obtenido después para que represente la máxima relación entre los dos conjuntos de variables (valores teóricos) que no se ha explicado por el primer par de valores teóricos. En resumen, los sucesivos pares de valores teóricos canónicos están basados en la varianza residual, y sus respectivas correlaciones canónicas (que reflejan las interrelaciones entre los valores teóricos) disminuyen a medida que se calculan funciones adicionales; es decir, el primer par de valores teóricos canónicos refleja la mayor inter-correlación, el siguiente par la segunda correlación, y así sucesivamente.

Se debe tomar en cuenta un aspecto adicional acerca de la obtención de valores teóricos canónicos. Ya se ha mencionado que los sucesivos pares de valores teóricos canónicos están basados en la varianza residual. Por tanto, cada uno de los pares de valores teóricos es ortogonal e independiente respecto de todos los otros valores teóricos obtenidos a partir del mismo conjunto de datos.

La validez de la relación entre los pares de valores teóricos se refleja en la correlación canónica. Cuando se eleva al cuadrado, la correlación canónica representa la cantidad de varianza de un valor teórico, explicada por el otro valor teórico. A esto también se le puede definir como la cantidad de varianza compartida entre los dos valores teóricos canónicos. *Las correlaciones canónicas al cuadrado son denominadas raíces canónicas o autovalores.*

¿Qué funciones canónicas se deben interpretar? Al igual que cualquier investigación que utiliza otras técnicas estadísticas, la práctica más común es analizar las funciones cuyos coeficientes de correlación son estadísticamente significativas para un nivel, normalmente .005 o mayor. La interpretación de los valores teóricos canónicos en una función significativa está basada en la premisa de que las variables en cada conjunto, que contribuyen fuertemente a las varianzas comparativas por estas funciones, son consideradas como relacionadas una con otras.

Se emplean algunos criterios conjuntamente para decidir qué funciones canónicas se deben interpretar.

- a) **Nivel de significación:** El nivel de significación de una correlación canónica, que generalmente se considera como el mínimo aceptable para

la interpretación, es el nivel 0.005, que junto con el nivel 0.001 se ha llegado a convertir en el nivel más habitualmente aceptado para considerar que un coeficiente de correlación es estadísticamente significativo.

- b) **Magnitud de las relaciones canónicas:** La significación práctica de las funciones canónicas, representada por el tamaño de las correlaciones canónicas, también se debe tener en cuenta para decidir qué funciones interpretar. La decisión se basa habitualmente en la contribución de los resultados para una mejor comprensión del problema de investigación que se está estudiando.

Paso 5: Interpretación del valor teórico canónico

Si la relación canónica resulta estadísticamente significativa y las magnitudes de la raíz canónica son aceptables, el investigador aún necesita realizar interpretaciones de los resultados. La realización de estas interpretaciones comprende el examen de las funciones canónicas para determinar la importancia relativa de cada una de las variables originales en las relaciones canónicas. Se han propuesto tres métodos:

- a) **Ponderaciones canónicas.** El enfoque tradicional para interpretar las funciones canónicas comprende el examen del signo y la magnitud de la ponderación asociada a cada variable en su valor teórico canónico. Las variables con ponderaciones tienen signos contrarios y presentan una relación inversa. Igualmente, las variables cuyas ponderaciones del mismo signo contrario presentan una relación directa. Sin embargo, la interpretación de la importancia o contribución relativa de una variable por su ponderación canónica está sujeta a las mismas críticas asociadas con la interpretación de los coeficientes beta en las técnicas de regresión.

Por ejemplo, una ponderación pequeña puede significar o bien que sus correspondiente variable es irrelevante para explicar la relación o bien que ha sido apartada de la relación debido a un alto grado de multicolinealidad. Otro problema del uso de las ponderaciones canónicas es que están sujetas a una inestabilidad considerable (variabilidad) de una muestra a otra. Esta inestabilidad se da porque el procedimiento de cálculo del análisis canónico genera ponderaciones que maximizan las correlaciones canónicas para una muestra determinada de conjuntos de variables dependientes e independientes observadas. Este problema hace que el uso de las ponderaciones canónicas para interpretar los resultados de un análisis canónico, deba realizarse con precaución.

- b) **Cargas canónicas.** El empleo de las cargas canónicas ha sustituido al uso de ponderaciones canónicas como base de interpretación, debido a las deficiencias inherentes a estas últimas. Las cargas canónicas, también

denominadas *correlaciones de estructura canónica*, miden la correlación simple entre una variable original observada del conjunto dependiente o independiente y el valor teórico canónico del conjunto. *Las cargas canónicas reflejan la varianza que la variable observada comparte con el valor teórico canónico y puede ser interpretada como una carga factorial para valorar la contribución de forma separada*, y se calcula la correlación dentro del conjunto entre variable y valor teórico. *Cuando mayor sea el coeficiente, mayor será la importancia que tiene para calcular el valor teórico canónico. Los criterios para determinar la significación de las correlaciones de estructura canónica también son los mismos que resultan con las cargas factoriales.*

- c) **El enfoque de interpretación a emplear.** Se han abordado diferentes métodos para interpretar la naturaleza de las relaciones canónicas. Sin embargo, la pregunta permanece: ¿Qué método debe emplear el investigador? Dado que la mayoría de los problemas canónicos requiere de un computador, el investigador debe emplear aquel método que esté disponible en los paquetes estadísticos estándar.

Resumen

- El análisis de correlación canónica es una técnica útil y potente para explorar las relaciones entre variables dependientes e independientes múltiples. La técnica es ante todo descriptiva, aunque puede ser empleada con fines predictivos. Los resultados obtenidos a partir de un análisis canónico deben dar respuestas a cuestiones relacionadas con el número de maneras en las que se relacionan dos conjuntos de múltiples variables, la validez de las relaciones y la naturaleza de las relaciones definidas. El análisis canónico posibilita al investigador combinar en una medida compuesta, lo que de otra forma podría ser un gran número difícil de manejar de correlaciones bivariantes entre conjuntos de variables. Es útil para identificar relaciones globales entre múltiples variables dependientes e independientes, especialmente cuando el analista tiene poco conocimiento *a priori* sobre las relaciones entre los conjuntos de variables. Fundamentalmente, el investigador puede aplicar el análisis de correlación canónica a un conjunto de variables, seleccionar aquellas variables (tanto dependientes como independientes) que parecen ser significativamente relacionadas, y llevar a cabo posteriores correlaciones canónicas con las restantes variables más significativas, o realizar regresiones con estas variables.
- Todo lo que ocurre en el ACC está diseñado para maximizar esta simple correlación.

- En la forma más general, el ACC es conducido a responder estas dos preguntas básicas de investigación:
 - a) ¿Hay una relación entre el conjunto de variables latentes predictor y criterio? ¿El modelo ACC suficientemente captura la relación entre el conjunto de variables predictor y criterio para garantizar la interpretación? Después de eso, si la relación es pequeña (y no interpretable) hará no sensitiva/sensible para probar identificar variables que contribuyan a esa relación.
 - b) Si una relación importante se encuentra que existe entre el conjunto de variables latentes predictor y criterios, entonces, ¿Cuál es la contribución relativa de cada variable de las funciones canónicas (relaciones) extraíbles?

Lista de requerimientos

- El tamaño de la muestra tiene un impacto directo en la significancia “práctica” de los resultados del estudio. Tamaños de muestras pequeñas no representarían las correlaciones como buenas y pueden empañar cualquier relación significativa.
- Muestras muy grandes, de otra manera, tienen la tendencia a producir resultados estadísticos significativos aun por pequeños, efectos no importantes que no tienen significancia práctica.

Hair et.al. (1995) y Tabachnick y Fidell (2001) recomiendan al menos 10 casos por variable para evitar sobreajuste de datos.

8.4. Supuestos y términos clave

- Linealidad. Todas las relaciones entre todos los pares de variables son lineales.
- Normalidad multivariable. El ACC no impone de manera estricta que las variables sean normalmente distribuidas; sin embargo, la normalidad es deseable porque la interpretación derivada del test de la estadística inferencial de la significancia de esta función canónica es basada en el supuesto de normalidad multivariada.
- Homocedasticidad. El análisis canónico resulta mejor cuando las relaciones entre pares de variables son homocedásticas; es decir, cuando una variable expone similar cantidad de varianza a través del rango de valores para la otra variable.

Términos clave en el ACC

- **Variables canónicas.** Es la combinación lineal que representa la suma ponderada de variables que comprenden cualquiera de las dos variables latentes dependientes e independientes.
- **Coefficiente de correlación canónica (R_c).** R_c es análogo al estadístico R del análisis de regresión lineal múltiple. Es la correlación bivariada (r de Pearson) ente dos variables canónicas. Es una medida de la fuerza de todas las relaciones entre la composición lineal (variables) para las variables predictoras y variables dependientes.
- **Correlación canónica cuadrada (R_c^2).** Raíz canónica y valor propio, es el cuadrado de la correlación canónica y provee un estimado de la cantidad de varianza compartida entre los dos conjuntos latentes de las variables predictoras y dependientes. Es directamente análogo al estadístico R^2 en la regresión múltiple.
- **Función canónica.** Es la relación entre dos variables canónicas, una variable para la dependiente y una para la variable independiente. La fuerza de la relación es dada por el coeficiente de correlación canónica (R_c).

El ACC extraerá tantas funciones como variables en el conjunto más pequeño de variables.

Ejemplo: Si hay cinco variables independientes y tres variables dependientes, el máximo número de funciones canónicas que pueden ser extraídas son tres. Las funciones extraídas son ortogonales (independientes) una de la otra.

- **Coefficientes de la función canónica estandarizada.** También conocidos como pesos canónicos; esos son coeficientes estandarizados usados en la ecuación lineal las variables predictoras y dependientes observadas dentro de las variables latentes respectivas.

El signo y la magnitud del coeficiente asignado a cada variable en su variable canónica ayudan en la interpretación de las funciones canónicas (relación entre las variantes dependiente e independiente).

Variables con coeficientes (pesos) relativamente grandes que contribuyen más a la varianza y viceversa.

Esos pesos pueden ser interpretados en la misma forma como los pesos beta (B) en la regresión múltiple.

- **Coefficiente de estructura** (r_s). También conocido como *carga canónica*, es una medida de la correlación bivariada simple (r de Pearson) entre una variable observada (variable predictor) y las puntuaciones de la función canónica (varianza canónica) para el conjunto de variables (ejemplo, la variable latente creada desde todas las variables predictoras a través de la ecuación lineal).

Como factor de carga, ayuda en la interpretación de la estructura del cálculo de la variable latente, a través de identificar cuáles variables observadas contribuyen a la creación de la variable latente.

- **Coefficiente de estructura canónica cuadrada** (rs^2) son los cuadrados de los coeficientes de estructura.

Este estadístico indica la cantidad de varianza compartida entre una variable observada y la variable latente, generada desde el conjunto de variables observadas.

8.5. Aplicación en SPSS

Para ilustrar la aplicación de la correlación canónica, emplearemos variables tomadas de la base de datos, consistente en una serie de medidas obtenidas a partir de una muestra de 100 clientes de la empresa DIST_COM_SV. Las variables incluyen las calificaciones de DIST_COM_SV sobre 7 atributos (X_1 a X_7) y dos medidas que reflejan los efectos de los esfuerzos de la empresa (X_9 = uso de los productos de DIST_COM_SV; y X_{10} = satisfacción del cliente con DIST_COM_SV).

Seguir la metodología antes descrita para aplicar el análisis de correlación canónica, en cada uno de estos pasos:

Primer paso: Objetivos del análisis de correlación canónica

Utilizaremos las nueve variables como datos de entrada. Las clasificaciones de DIST_COM_SV (X_1 a X_7) son designadas como el conjunto de múltiples variables independientes o variables predictoras. Las medidas del nivel de uso y nivel de satisfacción (variables X_9 y X_{10}) se especifican como el conjunto de múltiples variables dependientes o variables criterio. El problema estadístico comprende la identificación de cualquier relación latente entre las percepciones de un cliente sobre DIST_COM_SV y el nivel de uso/satisfacción del cliente.

Segundo y tercer pasos: Diseño de un análisis de correlación canónica y contrastación de los supuestos

El diseño de las variables incluye dos variables dependientes métricas y siete variables independientes métricas. La base conceptual de ambos conjuntos

está bien establecida, por lo que no hay necesidad de realizar formulaciones de modelos alternativos para contrastar los diferentes conjuntos de variables. Las siete variables generan un ratio de observaciones frente a variables de 13 a una, porque excede correctamente el supuesto de 10 observaciones para variables. No se considera que el tamaño muestral de 100 afecta las estimaciones del error de muestreo notablemente, y por tanto, no debería tener impacto sobre la significación estadística de los resultados. Como paso previo se deberá verificar que las variables superen todos los test estadísticos (supuestos).

Cuarto paso: Obtención de las funciones canónicas y valoración del ajuste global

El ACC está restringido a obtener dos funciones canónicas, ya que el conjunto de variables dependientes contiene solamente dos variables. Para determinar el número de funciones canónicas a incluir en el caso de interpretación, el análisis se centra en el nivel de significación estadística.

Significación estadística

El primer contraste de significación estadística es para las correlaciones canónicas de cada una de las dos funciones canónicas. En este ejemplo, ambas correlaciones canónicas son estadísticamente significativas (ver Tabla al final del documento). Junto a los test por cada función canónica de forma separada, también se llevan a cabo contrastes multivariantes para ambas funciones simultáneamente. Los contrastes estadísticos empleados son la lambda de Wilks, el criterio de Pillai, la traza de Hotelling y la mayor raíz de Roy. La tabla también presenta estos test multivariantes, los cuales indican que las funciones canónicas, consideradas conjuntamente, son estadísticamente significativas a un nivel de 0.01.

Además de la significación estadística, ambas correlaciones canónicas tienen el tamaño suficiente como para ser consideradas significativas de forma práctica.

Quinto paso: Interpretación de los valores teóricos canónicos

Una vez que se ha considerado estadísticamente significativa la relación canónica y aceptable la magnitud de la raíz canónica, el investigador procede a realizar interpretaciones de los resultados.

Estas interpretaciones comprenden el examen de las funciones canónicas para determinar la importancia relativa de cada una de las variables originales para obtener las relaciones canónicas. Los métodos de interpretación son (1) las ponderaciones canónicas (coeficientes estandarizados) y (2) las cargas canónicas (correlaciones de estructuras).

Ponderaciones canónicas

La Tabla que viene a continuación contiene las ponderaciones canónicas estandarizadas por cada valor teórico canónico, tanto para las variables dependientes como independientes. Como ya se discutió, la magnitud de las ponderaciones representa su contribución relativa al valor teórico. Basado en el tamaño de las ponderaciones, el orden de contribución de las variables independientes al primer valor teórico es X_5 , X_1 , X_3 , X_4 , X_6 , X_2 y X_7 , mientras que el orden de las variables dependientes sobre el primer valor teórico es X_{10} y después X_9 . Clasificaciones similares se pueden encontrar para los valores teóricos de la segunda función canónica. Dado que las ponderaciones canónicas son generalmente inestables, particularmente en casos de multicolinealidad, debido a que se calculan exclusivamente para optimizar la correlación canónica, se consideran más apropiadas las cargas canónicas.

Cargas canónicas

La Tabla que viene a continuación representa las cargas canónicas para los valores teóricos dependiente e independiente para ambas funciones canónicas. El objetivo de maximizar los valores teóricos para la correlación entre ellos tiene como resultados unos valores teóricos “optimizados” no para la interpretación sino para la predicción. Con esto, la identificación de las relaciones es más difícil. En el primer valor teórico dependiente, ambas variables tienen cargas que exceden el 0.90, reflejándose en una alta varianza compartida (0.855). Con lo que se indica un alto grado de inter-correlación entre las dos variables y sugiere que ambas medidas o bien una u otra son representativas de los efectos de los esfuerzos de DIST_COM_SV.

El primer valor teórico independiente sigue un modelo bastante diferente, con cargas que varían desde 0.06189 a 0.76560 con una variable independiente (X_7), que incluso tiene una carga negativa, aunque es bastante pequeña y no de interés sustantivo. Las tres variables son las cargas más altas en cuanto al valor teórico independiente: X_5 (servicio global), X_1 (velocidad de entrega) y X_3 (flexibilidad de precios).

De manera general, la primera función canónica corresponde más estrechamente a los resultados de una regresión múltiple, con el valor teórico independiente, representando el conjunto de variables que realiza la mejor predicción de las dos medidas dependientes. El investigador también debe realizar un análisis de sensibilidad del valor teórico independiente, en este caso para ver si las cargas cambian cuando es eliminada una variable independiente.

Una consideración gerencial de los resultados

Con el estudio de esta relación observamos primero que las dos variables dependientes están estrechamente relacionadas y crean una dimensión claramente definida para representar los resultados de los esfuerzos de DIST_COM_SV. Cuando se interpreta el valor teórico independiente, observamos que tres variables: X_5 (servicio global), X_1 (velocidad de entrega) y X_3 (flexibilidad de precio) proporcionan las contribuciones sustantivas y, por lo tanto, son los predictores claves de la dimensión de resultado. Estos deberían ser los puntos de concentración en el desarrollo de cualquier estrategia que afecte los resultados de DIST_COM_SV.

Tabla de resultados del ACC: Caso empresa DIST_COM_SV

En una ventana de sintaxis del programa SPSS, digitar la siguientes instrucciones; luego ejecutar el procedimiento.

MANOVA x9 x10 WITH x1 x2 x3 x4 x5 x6 x7

/PRINT = SIGNIF (MULT EIGEN DIMENR)

/DISCRIM = (STAN ESTIM COR ALPHA (.999)).

```
MANOVA x9 x10 WITH x1 x2 x3 x4 x5 x6 x7
/PRINT = SIGNIF (MULT EIGEN DIMENR)
/DISCRIM = (STAN ESTIM COR ALPHA (.999)).
```

Manova

[DataSet3] C:\Users\ElnorCrespin\Dropbox\UCA_MEAI\Curso_2015\DistComSV.sav

The default error term in MANOVA has been changed from WITHIN CELLS to WITHIN+RESIDUAL. Note that these are the same for all full factorial designs.

```
***** Analysis of Variance *****
100 cases accepted.
  0 cases rejected because of out-of-range factor values.
  0 cases rejected because of missing data.
  1 non-empty cell.

1 design will be processed.
```

EFFECT .. WITHIN CELLS Regression Multivariate Tests of Significance (S = 2, M = 2, N = 44 1/2)

Test Name	Value	Approx. F	Hypo. DF	Error DF	Sig. of F
Pillais	1.13740	17.32978	14.00	184.00	.000
Hotellings	7.54524	48.50513	14.00	180.00	.000
Wilks	.09037	30.24460	14.00	182.00	.000

Roys
Note.. F statistic for WILKS' Lambda is exact.

La cantidad de varianza compartida entre el conjunto de variables.

Ho: No existe relación entre los dos conjuntos de variables latentes (Rc=0). En nuestro caso la Ho puede ser rechazada.

Eigenvalues and Canonical Correlation				
Root No.	Eigenvalue	Pct.	Cum. Pct.	Sq. Cor
1	7.19494	95.35729	95.35729	.87797
2	.35030	4.64271	100.00000	.25943
Canon Cor.				.93700
				.50934
Dimension Reduction Analysis				
Roots	Wilks L.	F	Hypoth. DF	Error DF
1 TO 2	.09037	30.24460	14.00	182.00
2 TO 2	.74057	5.37132	6.00	92.00
Sig. of F				.000
				.000
EFFECT .. WITHIN CELLS Regression (Cont.)				
Univariate F-tests with (7,92) D. F.				
Variable	Sq. Mul. R	Adj. R-sq.	Hypoth. MS	Error MS
x9	.77495	.75783	885.68589	19.57016
x10	.80118	.78605	8.29404	.15660
F				45.25695
				52.96181
Sig. of F				.000
				.000
Standardized canonical coefficients for DEPENDENT variables				
Function No.				
Variable	1	2		
x9	.50077	-1.33009		

Medidas de ajuste global del modelo.

Explica una cantidad razonable de variación entre los 2 sets var's latentes

La primera función explica el 88% de la varianza.

Coefficientes. La magnitud representa su contribución relativa a la función; sin embargo, son inestables (ver estructura de coeficientes).

x10 .58003 1.29750

Correlations between DEPENDENT and canonical variables
Function No.

Variable 1 2
x9 .91293 -.40811
x10 .93587 .35235

Estructura de coeficientes. Las 2 variables retornan altos coeficientes de estructura

Variance in dependent variables explained by canonical variables

CAN. VAR.	Pct Var DEP	Cum Pct DEP	Pct Var COV	Cum Pct COV
1	85.46470	85.46470	75.03574	75.03574
2	14.53530	100.00000	3.77083	78.80657

Standardized canonical coefficients for COVARIATES

CAN. VAR.

COVARIATE	1	2
x1	.22593	.96049
x2	.10409	.86277
x3	.56800	-.15711
x4	.34802	1.45700
x5	.44429	-1.52682
x6	-.05023	-.73949
x7	.00032	-.47625

Ver el orden de contribución y el signo (es importante).

[Cada columna corresponde a rs. Elevando al cuadrado -una fila en las dos dimensiones- se obtiene la estructura de coeficientes canónicos al cuadrado- h^2 , sumando ambos valores]. Este valor se utiliza para evaluar que tan útil (o no útil) fue la variable para la solución del ACC.

Correlations between COVARIATES and canonical variables									
CAN. VAR.									
4									
Covariate	1	2							
x1	.76423	-.10881							
x2	.06189	-.14420							
x3	.62312	-.12038							
x4	.41496	.62435							
x5	.76560	-.22361							
x6	.34845	.19699							
x7	-.27816	-.21963							
Variance in covariates explained by canonical variables									
CAN. VAR.	Pct Var DEP	Cum Pct DEP	Pct Var COV	Cum Pct COV					
1	24.24807	24.24807	27.61823	27.61823					
2	2.12722	26.37530	8.19974	35.81797					
Regression analysis for WITHIN CELLS error term									
--- Individual Univariate .9500 confidence intervals									
Dependent variable .. x9 X9. Nivel de fidelidad									
COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t	Lower -95%	CL- upper		
x1	-.0435188925	-.0063937600	2.01273	-.02162	.983	-4.04097	3.95394		
x2	-.6789072515	-.0902990875	2.09025	-.32480	.746	-4.83032	3.47251		

x3	3.3619678619	.5185369533	.41125	8.17502	.000	2.54519	4.17874
x4	-.0410055580	-.0051609518	.66683	-.06149	.951	-1.36539	1.28338
x5	8.3453720194	.6974292935	3.91830	2.12985	.036	.56329	16.12745
x6	1.2914708389	.1107446736	.94720	1.36346	.176	-.58976	3.17270
x7	.5629510281	.0992730912	.35544	1.58380	.117	-.14299	1.26889
X10: Nivel de satisfacción							
Dependent variable .. x10							
COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t	Lower -95%	CL- Upper
x1	.2400012475	.3704905244	.18005	1.33298	.186	-.11759	.59759
x2	.1761110801	.2461187511	.18698	.94186	.349	-.19525	.54748
x3	.2899512696	.4698900800	.03679	7.88162	.000	.21689	.36302
x4	.4285017041	.5666628484	.05965	7.18345	.000	.31003	.54697
x5	.1316382477	.1155903780	.35051	.37556	.708	-.56451	.82778
x6	-.1961871365	-.1767640239	.08473	-2.31538	.023	-.36447	-.02790
x7	-.0459750654	-.0851860325	.03180	-1.44593	.152	-.10913	.01718

***** Analysis of Variance *****

EFFECT .. CONSTANT
Multivariate Tests of Significance (S = 1, M = 0, N = 44 1/2)

Test Name	Value	Exact F	Hypoth. DF	Error DF	Sig. of F
Pillais	.05975	2.89158	2.00	91.00	.061
Hotellings	.06355	2.89158	2.00	91.00	.061
Wilks	.94025	2.89158	2.00	91.00	.061
Rois	.05975				

Note.. F statistics are exact.

Eigenvalues and Canonical Correlations						
Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.		
1	.06355	100.00000	100.00000	.24445		

EFFECT .. CONSTANT (Cont.)						
Univariate F-tests with (1,92) D. F.						
Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
x9	81.57914	1800.45515	81.57914	19.57016	4.16855	.044
x10	.25410	14.40759	.25410	.15660	1.62254	.206

EFFECT .. CONSTANT (Cont.)						
Standardized discriminant function coefficients						
Function No.						
Variable	1					
x9	.85004					
x10	.53578					

Estimates of effects for canonical variables						
Canonical Variable						
Parameter	1					

1	-2.72034	
Correlations between DEPENDENT and canonical variables		
Canonical Variable		
Variable	1	
x9	.84438	
x10	.52680	

8.6. Ejercicios

Ejercicio 1

- Robert Ho (2000) investigó la eficacia en la competencia de dos modelos de motivación-protección con la intención de predecir el comportamiento del sexo seguro (uso del condón) seguido de la comunicación del VIH/AIDS.
- Los ítems fueron escritos para representar la motivación a la protección y la intención del uso del condón (ver lista de variables abajo)—anexo HIV-2.sav
- Utilizar ACC para investigar si hay una relación entre la motivación a la protección y la intención para el uso del condón, como multioperacionalizado en los dos conjuntos de variables latentes.

Variables de motivación-protección:

- **Riesgo percibido.** No usar un condón (o no insistir en que puede utilizarse durante las relaciones sexuales) conducirá a un mayor riesgo de contraer SIDA.
- **Gravedad percibida.** Contraer SIDA es equivalente a recibir una sentencia de muerte.
- **Autoeficacia.** Al practicar sexo seguro (por ejemplo, utilizando un condón o insistir en que se use) me siento seguro de poder evitar el contagio del SIDA.
- **Eficacia de respuesta.** Puedo disminuir la probabilidad de contraer SIDA al practicar siempre sexo seguro (por ejemplo, usando condón o insistiendo que se use).
- **Afrontamiento desadaptativo.** No me preocupo acerca del SIDA porque soy capaz de tener parejas libres de SIDA.
- **Temor.** El hecho de que muchas personas hayan muerto de SIDA, realmente me asusta.

Intención sobre el uso del condón:

- **Disposición para su uso.** Estoy dispuesto a usar el condón (o insistir en su uso) cuando tenga relaciones en el futuro.
- **Uso probable.** Es altamente probable que use el condón (o insistir en su uso) cuando tenga relaciones en el futuro.

- **Intención de uso.** Pretendo usar el condón (o insistir en su uso) cuando tenga relaciones en el futuro.
- **Uso certero.** Realmente usaré el condón (o insistir en su uso) cuando tenga relaciones en el futuro.

9. Bibliografía

- Anderson, David Ray; Sweeney, Dennis J. y Williams, Thomas Arthur. (2008). Estadística para administración y economía. Cengage Learning, décima edición, México.
- Bienvenido Visauta. (2003). Análisis Estadístico con SPSS para Windows. Volumen II. Estadística Multivariante.
- Cuadras, Carles, M. (2008). Nuevos métodos de análisis multivariante. CMC Ediciones, Barcelona, España.
- Grim, L. and Yarnold, P.R. (1994). Reading and understanding multivariate statistics. American Psychological Association. Washington, D.C.
- Hair, Anderson; Tathan, Black. (1999). Análisis Multivariante, 5.^a Edición Prentice Hall.
- J.D. Jobson. (1992). Applied Multivariate Data Analysis. Volumen II: Categorical and Multivariate Methods
- Lawrence C. Hamilton. (2013). Statistics with Stata Version 12.
- Malhotra, Naresh K. (2008). Investigación de mercados. Pearson Educación, quinta edición, México.
- McDaniel, Carl y Gates, Roger. (1999). Investigación de mercados contemporánea. International Thomson, cuarta edición, México.
- Pardo, A.; Ruiz, M. (2002). SPSS: Guía para el análisis de datos. Mc Graw Hill-Interamericana de España.
- Peña, D. (2002). Análisis de Datos Multivariantes, Mc Graw Hill - Interamericana de España.
- Ho, Robert. (2014). Handbook of Univariate and Multivariate Data Analysis with IBM SPSS.
- Pedret, Ramon; Sagnier, Laura; Camp, Francesc. (2003). Herramientas para segmentar mercados y posicionar productos, Ediciones Deusto.

10. Anexos

Capítulo 1

Base de Datos mundodes.sav

País	var1	var2	var3	var4	var5	var6	var7
Albania	24.70	5.70	30.80	69.60	75.50	600.00	1.00
Bulgaria	12.50	11.90	14.40	68.30	74.70	2250.00	1.00
Czechoslovaki	13.40	11.70	11.30	71.80	77.70	2980.00	1.00
Hungary	11.60	13.40	14.80	65.40	73.80	2780.00	1.00
Poland	14.30	10.20	16.00	67.20	75.70	1690.00	1.00
Romania	13.60	10.70	26.90	66.50	72.40	1640.00	1.00
USSR	17.70	10.00	23.00	64.60	74.00	2242.00	1.00
Byelorussia	15.20	9.50	13.10	66.40	75.90	1880.00	1.00
Ukrainian	13.40	11.60	13.00	66.40	74.80	1320.00	1.00
Argentina	20.70	8.40	25.70	65.50	72.70	2370.00	2.00
Bolivia	46.60	18.00	111.00	51.00	55.40	630.00	2.00
Brasil	28.60	7.90	63.00	62.30	67.60	2680.00	2.00
Chile	23.40	5.80	17.10	68.10	75.10	1940.00	2.00
Columbia	27.40	6.10	40.00	63.40	69.20	1260.00	2.00
Ecuador	32.90	7.40	63.00	63.40	67.60	980.00	2.00
Guyana	28.30	7.30	56.00	60.40	66.10	330.00	2.00
Paraguay	34.80	6.60	42.00	64.40	68.50	1110.00	2.00
Perú	32.90	8.30	109.90	56.80	66.50	1160.00	2.00
Uruguay	18.00	9.60	21.90	68.40	74.90	2560.00	2.00
Venezuela	27.50	4.40	23.30	66.70	72.80	2560.00	2.00
México	29.00	23.20	43.00	62.10	66.00	2490.00	2.00
Belgium	12.00	10.60	7.90	70.00	76.80	15540.00	3.00
Finland	13.20	10.10	5.80	70.70	78.70	26040.00	3.00
Denmark	12.40	11.90	7.50	71.80	77.70	22080.00	3.00
France	13.60	9.40	7.40	72.30	80.50	19490.00	3.00
Germany	11.40	11.20	7.40	71.80	78.40	22320.00	3.00
Greece	10.10	9.20	11.00	65.40	74.00	5990.00	3.00
Ireland	15.10	9.10	7.50	71.00	76.70	9550.00	3.00
Italy	9.70	9.10	8.80	72.00	78.60	16830.00	3.00
Netherlands	13.20	8.60	7.10	73.30	79.90	17320.00	3.00
Norway	14.30	10.70	7.80	67.20	75.70	23120.00	3.00
Portugal	11.90	9.50	13.10	66.50	72.40	7600.00	3.00
Spain	10.70	8.20	8.10	72.50	78.60	11020.00	3.00
Sweden	14.50	11.10	5.60	74.20	80.00	23660.00	3.00
Switzerland	12.50	9.50	7.10	73.90	80.00	34064.00	3.00
U.K.	13.60	11.50	8.40	72.20	77.90	16100.00	3.00
Austria	14.90	7.40	8.00	73.30	79.60	17000.00	3.00
Japan	9.90	6.70	4.50	75.90	81.80	25430.00	3.00
Canada	14.50	7.30	7.20	73.00	79.80	20470.00	3.00
U.S.A.	16.70	8.10	9.10	71.50	78.30	21790.00	3.00
Afghanistan	40.40	18.70	181.60	41.00	42.00	168.00	5.00
Bahrain	28.40	3.80	16.00	66.80	69.40	6340.00	4.00
Iran	42.50	11.50	108.10	55.80	55.00	2490.00	4.00
Iraq	42.60	7.80	69.00	63.00	64.80	3020.00	4.00
Israel	22.30	6.30	9.70	73.90	77.40	10920.00	4.00
Jordan	38.90	6.40	44.00	64.20	67.80	1240.00	4.00

País	var1	var2	var3	var4	var5	var6	var7
Kuwait	26.80	2.20	15.60	71.20	75.40	16150.00	4.00
Oman	45.60	7.80	40.00	62.20	65.80	5220.00	4.00
Saudi Arabia	42.10	7.60	71.00	61.70	65.20	7050.00	4.00
Turkey	29.20	8.40	76.00	62.50	65.80	1630.00	4.00
United Arab Emir	22.80	3.80	26.00	68.60	72.90	19860.00	4.00
Bangladesh	42.20	15.50	119.00	56.90	56.00	210.00	5.00
China	21.20	6.70	32.00	68.00	70.90	380.00	5.00
Hong Kong	11.70	4.90	6.10	74.30	80.10	14210.00	5.00
India	30.50	10.20	91.00	52.50	52.10	350.00	5.00
Indonesia	28.60	9.40	75.00	58.50	62.00	570.00	5.00
Malaysia	31.60	5.60	24.00	67.50	71.60	2320.00	5.00
Mongolia	36.10	8.80	68.00	60.00	62.50	110.00	5.00
Nepal	39.60	14.80	128.00	50.90	48.10	170.00	5.00
Pakistan	30.30	8.10	107.70	59.00	59.20	380.00	5.00
Philippines	33.20	7.70	45.00	62.50	66.10	730.00	5.00
Singapore	17.80	5.20	7.50	68.70	74.00	11160.00	5.00
Sri Lanka	21.30	6.20	19.40	67.80	71.70	470.00	5.00
Thailand	22.30	7.70	28.00	63.80	68.90	1420.00	5.00
Algeria	35.50	8.30	74.00	61.60	63.30	2060.00	6.00
Angola	47.20	20.20	137.00	42.90	46.10	610.00	6.00
Botswana	48.50	11.60	67.00	52.30	59.70	2040.00	6.00
Congo	46.10	14.60	73.00	50.10	55.30	1010.00	6.00
Egypt	38.80	9.50	49.40	57.80	60.30	600.00	6.00
Ethiopia	48.60	20.70	137.00	42.40	45.60	120.00	6.00
Gabon	39.40	16.80	103.00	49.90	53.20	390.00	6.00
Gambia	47.40	21.40	143.00	41.40	44.60	260.00	6.00
Ghana	44.40	13.10	90.00	52.20	55.80	390.00	6.00
Kenya	47.00	11.30	72.00	56.50	60.50	370.00	6.00
Libya	44.00	9.40	82.00	59.10	62.60	5310.00	6.00
Malawi	48.30	25.00	130.00	38.10	41.20	200.00	6.00
Morocco	35.50	9.80	82.00	59.10	62.50	960.00	6.00
Mozambique	45.00	18.50	141.00	44.90	48.10	80.00	6.00
Namibia	44.00	12.10	135.00	55.00	57.50	1030.00	6.00
Nigeria	48.50	15.60	105.00	48.80	52.20	360.00	6.00
Sierra Leone	48.20	23.40	154.00	39.40	42.60	240.00	6.00
Somalia	50.10	20.20	132.00	43.40	46.60	120.00	6.00
South Africa	32.10	9.90	72.00	57.50	63.50	2530.00	6.00
Sudan	44.60	15.80	108.00	48.60	51.00	480.00	6.00
Swaziland	46.80	12.50	118.00	42.90	49.50	810.00	6.00
Tunisia	31.10	7.30	52.00	64.90	66.40	1440.00	6.00
Uganda	52.20	15.60	103.00	49.90	52.70	220.00	6.00
Tanzania	50.50	14.00	106.00	51.30	54.70	110.00	6.00
Zaire	45.60	14.20	83.00	50.30	53.70	220.00	6.00
Zambia	51.10	13.70	80.00	50.40	52.50	420.00	6.00
Zimbabwe	41.70	10.30	66.00	56.50	60.10	640.00	6.00

Acciones

Obs	X1	X2	X3
1	3.4	89.7	30.2
2	5.1	55.7	9.9
3	4.5	52.3	11.5
4	3.5	47	11.2
5	5.9	42.7	7
6	5.1	30.6	6.9
7	4.6	64.4	11.8
8	5	51	9.6
9	3.2	54.4	14.7
10	3.4	45.7	13.2
11	6.5	39.9	5.2
12	4.4	40.3	13.7
13	5.1	52.4	11
14	5.8	43.9	8
15	4.6	52.8	14.4
16	7.2	65.8	7.8
17	7.2	58.1	7.7
18	4.4	58.5	12.1
19	7.8	84.3	11
20	16	96.5	6
21	16.7	100	6.8
22	15.2	92.3	5.2
23	17.5	99.9	6.8
24	16.2	93.5	6.1
25	14.7	100	6.6
26	15.3	99.9	5.9
27	15.8	100	6.9
28	18.3	96.3	5.7
29	15.9	100	6.9
30	16.1	92.5	6.1
31	9.7	87.6	7.7
32	6.9	53.6	6.6
33	14.4	87.8	5.2
34	14.9	34.5	4.69

Capítulo 2

Base de Datos Bd_empleados.sav

Variable	Etiqueta	Nivel de medida
id	Código de empleado	Escala
sexo	Sexo	Nominal
fechnac	Fecha de nacimiento	Escala
educ	Nivel educativo	Nominal
catlab	Categoría laboral	Nominal
salario	Salario actual	Escala
salini	Salario inicial	Escala
tiempemp	Meses desde el contrato	Escala
expprev	Experiencia previa (meses)	Escala
minoría	Clasificación de minorías	Nominal

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
1	Hombres	03-feb-52	15	Directivo	\$57,000.00	\$27,000.00	98	144	No
2	Hombres	23-may-58	16	Administrativo	\$40,200.00	\$18,750.00	98	36	No
3	Mujeres	26-jul-29	12	Administrativo	\$21,450.00	\$12,000.00	98	381	No
4	Mujeres	15-abr-47	8	Administrativo	\$21,900.00	\$13,200.00	98	190	No
5	Hombres	09-feb-55	15	Administrativo	\$45,000.00	\$21,000.00	98	138	No
6	Hombres	22-ago-58	15	Administrativo	\$32,100.00	\$13,500.00	98	67	No
7	Hombres	26-abr-56	15	Administrativo	\$36,000.00	\$18,750.00	98	114	No
8	Mujeres	06-may-66	12	Administrativo	\$21,900.00	\$9,750.00	98	Ausente	No
9	Mujeres	23-ene-46	15	Administrativo	\$27,900.00	\$12,750.00	98	115	No
10	Mujeres	13-feb-46	12	Administrativo	\$24,000.00	\$13,500.00	98	244	No
11	Mujeres	07-feb-50	16	Administrativo	\$30,300.00	\$16,500.00	98	143	No
12	Hombres	11-ene-66	8	Administrativo	\$28,350.00	\$12,000.00	98	26	Si
13	Hombres	17-jul-60	15	Administrativo	\$27,750.00	\$14,250.00	98	34	Si
14	Mujeres	26-feb-49	15	Administrativo	\$35,100.00	\$16,800.00	98	137	Si
15	Hombres	29-ago-62	12	Administrativo	\$27,300.00	\$13,500.00	97	66	No
16	Hombres	17-nov-64	12	Administrativo	\$40,800.00	\$15,000.00	97	24	No
17	Hombres	18-jul-62	15	Administrativo	\$46,000.00	\$14,250.00	97	48	No
18	Hombres	20-mar-56	16	Directivo	\$103,750.00	\$27,510.00	97	70	No
19	Hombres	19-ago-62	12	Administrativo	\$42,300.00	\$14,250.00	97	103	No
20	Mujeres	23-ene-40	12	Administrativo	\$26,250.00	\$11,550.00	97	48	No
21	Mujeres	19-feb-63	16	Administrativo	\$38,850.00	\$15,000.00	97	17	No
22	Hombres	24-sep-40	12	Administrativo	\$21,750.00	\$12,750.00	97	315	Si
23	Mujeres	15-mar-65	15	Administrativo	\$24,000.00	\$11,100.00	97	75	Si
24	Mujeres	27-mar-33	12	Administrativo	\$16,950.00	\$9,000.00	97	124	Si

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
25	Mujeres	01-jul-42	15	Administrativo	\$21,150.00	\$9,000.00	97	171	Si
26	Hombres	08-nov-66	15	Administrativo	\$31,050.00	\$12,600.00	96	14	No
27	Hombres	19-mar-54	19	Directivo	\$60,375.00	\$27,480.00	96	96	No
28	Hombres	11-abr-63	15	Administrativo	\$32,550.00	\$14,250.00	96	43	No
29	Hombres	28-ene-44	19	Directivo	\$135,000.00	\$79,980.00	96	199	No
30	Hombres	17-sep-61	15	Administrativo	\$31,200.00	\$14,250.00	96	54	No
31	Hombres	24-feb-64	12	Administrativo	\$36,150.00	\$14,250.00	96	83	No
32	Hombres	28-ene-54	19	Directivo	\$110,625.00	\$45,000.00	96	120	No
33	Hombres	18-mar-61	15	Administrativo	\$42,000.00	\$15,000.00	96	68	No
34	Hombres	02-feb-49	19	Directivo	\$92,000.00	\$39,990.00	96	175	No
35	Hombres	22-ago-61	17	Directivo	\$81,250.00	\$30,000.00	96	18	No
36	Mujeres	07-ago-63	8	Administrativo	\$31,350.00	\$11,250.00	96	52	No
37	Hombres	09-oct-54	12	Administrativo	\$29,100.00	\$13,500.00	96	113	Si
38	Hombres	27-abr-62	15	Administrativo	\$31,350.00	\$15,000.00	96	49	Si
39	Hombres	22-jun-60	16	Administrativo	\$36,000.00	\$15,000.00	96	46	Si
40	Mujeres	28-ago-33	15	Administrativo	\$19,200.00	\$9,000.00	96	23	Si
41	Mujeres	18-mar-61	12	Administrativo	\$23,550.00	\$11,550.00	96	52	Si
42	Hombres	23-sep-60	15	Administrativo	\$35,100.00	\$16,500.00	95	90	No
43	Hombres	18-ene-64	12	Administrativo	\$23,250.00	\$14,250.00	95	46	No
44	Hombres	15-jun-63	8	Administrativo	\$29,250.00	\$14,250.00	95	50	No
45	Hombres	02-ago-38	12	Seguridad	\$30,750.00	\$13,500.00	95	307	No
46	Mujeres	18-nov-40	15	Administrativo	\$22,350.00	\$12,750.00	95	165	No
47	Mujeres	28-abr-38	12	Administrativo	\$30,000.00	\$16,500.00	95	228	No
48	Hombres	07-jun-47	12	Seguridad	\$30,750.00	\$14,100.00	94	240	No
49	Hombres	16-sep-58	15	Administrativo	\$34,800.00	\$16,500.00	94	93	No
50	Hombres	09-feb-60	16	Directivo	\$60,000.00	\$23,730.00	94	59	No
51	Hombres	08-jul-62	12	Administrativo	\$35,550.00	\$15,000.00	94	48	No
52	Hombres	12-nov-63	15	Administrativo	\$45,150.00	\$15,000.00	94	40	No
53	Hombres	21-abr-54	18	Directivo	\$73,750.00	\$26,250.00	94	56	No
54	Hombres	04-jun-31	12	Administrativo	\$25,050.00	\$13,500.00	94	444	No
55	Hombres	25-jun-60	12	Administrativo	\$27,000.00	\$15,000.00	94	120	No
56	Hombres	16-abr-62	15	Administrativo	\$26,850.00	\$13,500.00	94	5	No
57	Hombres	15-abr-63	15	Administrativo	\$33,900.00	\$15,750.00	94	78	No
58	Mujeres	14-nov-64	15	Administrativo	\$26,400.00	\$13,500.00	94	3	No
59	Hombres	07-may-61	15	Administrativo	\$28,050.00	\$14,250.00	94	36	Si
60	Hombres	16-feb-59	12	Administrativo	\$30,900.00	\$15,000.00	94	102	Si
61	Hombres	28-abr-64	8	Administrativo	\$22,500.00	\$9,750.00	94	36	Si
62	Hombres	18-jul-62	16	Directivo	\$48,000.00	\$21,750.00	93	22	No
63	Hombres	20-ago-61	17	Directivo	\$55,000.00	\$26,250.00	93	32	No
64	Hombres	28-sep-63	16	Directivo	\$53,125.00	\$21,000.00	93	48	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
65	Hombres	28-mar-64	8	Administrativo	\$21,900.00	\$14,550.00	93	41	No
66	Hombres	16-feb-62	19	Directivo	\$78,125.00	\$30,000.00	93	7	No
67	Hombres	28-may-64	16	Directivo	\$46,000.00	\$21,240.00	93	35	No
68	Hombres	05-may-63	16	Directivo	\$45,250.00	\$21,480.00	93	36	No
69	Hombres	23-jun-60	16	Directivo	\$56,550.00	\$25,000.00	93	34	No
70	Hombres	08-feb-62	15	Administrativo	\$41,100.00	\$20,250.00	93	27	No
71	Hombres	26-ago-48	17	Directivo	\$82,500.00	\$34,980.00	93	207	No
72	Mujeres	07-ene-64	16	Administrativo	\$54,000.00	\$18,000.00	93	11	No
73	Mujeres	09-feb-68	12	Administrativo	\$26,400.00	\$10,500.00	93	Ausente	No
74	Mujeres	28-abr-33	15	Administrativo	\$33,900.00	\$19,500.00	93	192	No
75	Mujeres	12-ago-65	15	Administrativo	\$24,150.00	\$11,550.00	93	Ausente	No
76	Mujeres	03-sep-67	15	Administrativo	\$29,250.00	\$11,550.00	93	11	No
77	Mujeres	09-sep-68	12	Administrativo	\$27,600.00	\$11,400.00	93	6	No
78	Mujeres	20-ago-68	12	Administrativo	\$22,950.00	\$10,500.00	93	10	No
79	Mujeres	23-ene-62	16	Administrativo	\$34,800.00	\$14,550.00	93	8	No
80	Mujeres	25-may-61	16	Administrativo	\$51,000.00	\$18,000.00	93	22	No
81	Mujeres	12-mar-68	12	Administrativo	\$24,300.00	\$10,950.00	93	5	No
82	Mujeres	28-ago-47	12	Administrativo	\$24,750.00	\$14,250.00	93	193	Si
83	Mujeres	12-oct-67	12	Administrativo	\$22,950.00	\$11,250.00	93	Ausente	Si
84	Mujeres	12-mar-67	8	Administrativo	\$25,050.00	\$10,950.00	93	8	Si
85	Hombres	09-abr-62	15	Administrativo	\$25,950.00	\$17,100.00	92	42	No
86	Hombres	25-ago-61	15	Administrativo	\$31,650.00	\$15,750.00	92	64	No
87	Hombres	20-oct-59	12	Administrativo	\$24,150.00	\$14,100.00	92	130	No
88	Hombres	10-feb-62	19	Directivo	\$72,500.00	\$28,740.00	92	10	No
89	Hombres	24-jun-61	19	Directivo	\$68,750.00	\$27,480.00	92	8	No
90	Mujeres	27-feb-38	8	Administrativo	\$16,200.00	\$9,750.00	92	Ausente	No
91	Mujeres	04-nov-67	12	Administrativo	\$20,100.00	\$11,250.00	92	24	No
92	Mujeres	25-jun-68	8	Administrativo	\$24,000.00	\$10,950.00	92	6	No
93	Mujeres	05-mar-68	12	Administrativo	\$25,950.00	\$10,950.00	92	Ausente	No
94	Mujeres	04-ago-50	12	Administrativo	\$24,600.00	\$10,050.00	92	44	No
95	Mujeres	08-ago-68	12	Administrativo	\$28,500.00	\$10,500.00	92	6	No
96	Hombres	02-oct-33	8	Seguridad	\$30,750.00	\$15,000.00	92	432	Si
97	Hombres	18-ene-53	17	Administrativo	\$40,200.00	\$19,500.00	92	168	Si
98	Hombres	17-may-56	8	Seguridad	\$30,000.00	\$15,000.00	92	144	Si
99	Mujeres	07-jul-68	12	Administrativo	\$22,050.00	\$10,950.00	92	5	Si
100	Hombres	25-oct-63	18	Directivo	\$78,250.00	\$27,480.00	91	47	No
101	Hombres	14-mar-60	16	Directivo	\$60,625.00	\$22,500.00	91	44	No
102	Hombres	28-mar-63	14	Administrativo	\$39,900.00	\$15,750.00	91	59	No
103	Hombres	17-mar-59	19	Directivo	\$97,000.00	\$35,010.00	91	68	No
104	Hombres	05-nov-62	15	Administrativo	\$27,450.00	\$15,750.00	91	48	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
105	Hombres	07-mar-66	15	Administrativo	\$31,650.00	\$13,500.00	91	18	No
106	Hombres	04-ago-62	19	Directivo	\$91,250.00	\$29,490.00	91	23	No
107	Mujeres	16-ago-60	12	Administrativo	\$25,200.00	\$14,400.00	91	83	No
108	Mujeres	16-jul-30	12	Administrativo	\$21,000.00	\$11,550.00	91	108	No
109	Hombres	10-nov-63	12	Administrativo	\$30,450.00	\$15,000.00	91	49	Si
110	Hombres	29-oct-52	15	Administrativo	\$28,350.00	\$18,000.00	91	151	Si
111	Hombres	27-nov-40	12	Seguridad	\$30,750.00	\$9,000.00	91	314	Si
112	Hombres	21-jun-48	12	Seguridad	\$30,750.00	\$15,000.00	91	240	Si
113	Hombres	06-oct-59	16	Directivo	\$54,875.00	\$27,480.00	90	68	No
114	Hombres	25-ago-61	14	Administrativo	\$37,800.00	\$16,500.00	90	60	No
115	Hombres	12-may-61	15	Administrativo	\$33,450.00	\$14,100.00	90	85	No
116	Hombres	09-jun-62	15	Administrativo	\$30,300.00	\$16,500.00	90	16	No
117	Mujeres	14-ene-32	12	Administrativo	\$31,500.00	\$18,750.00	90	205	No
118	Mujeres	04-mar-64	12	Administrativo	\$31,650.00	\$14,250.00	90	48	No
119	Mujeres	23-jul-63	12	Administrativo	\$25,200.00	\$14,100.00	90	55	No
120	Mujeres	12-nov-64	16	Directivo	\$37,800.00	\$15,750.00	90	7	No
121	Mujeres	06-ago-36	15	Administrativo	\$18,750.00	\$10,500.00	90	54	No
122	Mujeres	26-sep-65	15	Administrativo	\$32,550.00	\$13,500.00	90	22	No
123	Mujeres	24-abr-49	12	Administrativo	\$33,300.00	\$15,000.00	90	3	No
124	Mujeres	29-may-63	16	Administrativo	\$38,550.00	\$16,500.00	90	Ausente	No
125	Hombres	06-ago-56	12	Administrativo	\$27,450.00	\$15,000.00	90	173	Si
126	Hombres	21-ene-51	15	Seguridad	\$24,300.00	\$15,000.00	90	191	Si
127	Hombres	01-sep-50	12	Seguridad	\$30,750.00	\$15,000.00	90	209	Si
128	Mujeres	25-jul-46	12	Administrativo	\$19,650.00	\$9,750.00	90	229	Si
129	Hombres	18-jul-59	17	Directivo	\$68,750.00	\$27,510.00	89	38	No
130	Hombres	06-sep-58	20	Directivo	\$59,375.00	\$30,000.00	89	6	No
131	Hombres	08-feb-62	15	Administrativo	\$31,500.00	\$15,750.00	89	22	No
132	Hombres	17-may-53	12	Administrativo	\$27,300.00	\$17,250.00	89	175	No
133	Hombres	12-sep-59	15	Administrativo	\$27,000.00	\$15,750.00	89	87	No
134	Mujeres	10-nov-41	16	Directivo	\$41,550.00	\$24,990.00	89	285	No
135	Mujeres	22-may-68	12	Administrativo	\$26,250.00	\$10,950.00	89	Ausente	No
136	Hombres	25-ago-39	12	Administrativo	\$22,200.00	\$15,000.00	88	324	No
137	Hombres	12-jul-37	21	Directivo	\$65,000.00	\$37,500.00	88	264	No
138	Hombres	09-ene-47	12	Administrativo	\$30,900.00	\$15,000.00	88	252	No
139	Mujeres	18-jun-31	8	Administrativo	\$20,100.00	\$13,200.00	88	90	No
140	Mujeres	05-abr-65	12	Administrativo	\$22,350.00	\$13,500.00	88	26	No
141	Mujeres	14-jun-66	15	Administrativo	\$35,550.00	\$13,350.00	88	32	No
142	Mujeres	13-nov-66	12	Administrativo	\$28,500.00	\$13,950.00	88	34	No
143	Mujeres	24-ago-39	12	Administrativo	\$24,450.00	\$13,200.00	88	107	No
144	Mujeres	28-ago-31	8	Administrativo	\$16,650.00	\$9,750.00	88	412	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
145	Hombres	15-ene-64	12	Administrativo	\$26,700.00	\$13,500.00	88	38	Si
146	Hombres	02-ene-51	18	Administrativo	\$43,950.00	\$23,250.00	88	182	Si
147	Hombres	11-sep-36	15	Administrativo	\$23,700.00	\$13,500.00	88	359	Si
148	Mujeres	05-oct-59	15	Administrativo	\$26,550.00	\$14,250.00	88	61	Si
149	Hombres	28-jun-62	12	Administrativo	\$27,600.00	\$15,000.00	87	75	No
150	Hombres	05-mar-54	12	Administrativo	\$25,800.00	\$15,000.00	87	143	No
151	Hombres	16-ago-54	16	Directivo	\$42,300.00	\$26,250.00	87	126	No
152	Hombres	03-may-29	8	Seguridad	\$30,750.00	\$15,000.00	87	451	No
153	Mujeres	13-may-67	12	Administrativo	\$26,700.00	\$12,900.00	87	18	No
154	Mujeres	07-jun-40	12	Administrativo	\$20,850.00	\$12,000.00	87	163	No
155	Hombres	06-mar-63	15	Administrativo	\$35,250.00	\$15,000.00	87	54	Si
156	Hombres	12-ene-63	15	Administrativo	\$26,700.00	\$15,000.00	87	56	Si
157	Mujeres	02-jul-67	12	Administrativo	\$26,550.00	\$13,050.00	87	11	Si
158	Mujeres	20-sep-64	12	Administrativo	\$27,750.00	\$12,000.00	87	11	Si
159	Mujeres	13-ene-51	16	Administrativo	\$25,050.00	\$12,750.00	87	123	Si
160	Hombres	27-ago-51	16	Directivo	\$66,000.00	\$47,490.00	86	150	No
161	Hombres	11-feb-63	16	Administrativo	\$52,650.00	\$19,500.00	86	20	No
162	Hombres	25-may-61	16	Directivo	\$45,625.00	\$23,250.00	86	60	No
163	Hombres	10-nov-65	15	Administrativo	\$30,900.00	\$15,000.00	86	25	No
164	Mujeres	13-ago-65	15	Administrativo	\$29,400.00	\$16,500.00	86	24	No
165	Mujeres	29-abr-66	15	Administrativo	\$33,300.00	\$13,500.00	86	24	No
166	Mujeres	12-jul-42	12	Administrativo	\$21,900.00	\$9,750.00	86	156	No
167	Mujeres	26-ago-32	12	Administrativo	\$18,150.00	\$9,750.00	86	72	No
168	Mujeres	23-abr-62	16	Directivo	\$46,875.00	\$17,250.00	86	19	No
169	Hombres	23-may-65	12	Administrativo	\$25,500.00	\$14,400.00	86	37	Si
170	Hombres	13-jun-64	12	Administrativo	\$26,550.00	\$15,000.00	86	38	Si
171	Hombres	21-ene-30	12	Administrativo	\$26,700.00	\$13,500.00	86	367	Si
172	Mujeres	13-jun-53	15	Administrativo	\$29,850.00	\$15,000.00	86	79	Si
173	Hombres	15-ene-50	20	Directivo	\$69,250.00	\$42,480.00	85	134	No
174	Hombres	07-ene-35	8	Seguridad	\$31,950.00	\$15,000.00	85	438	No
175	Hombres	08-ene-38	8	Administrativo	\$26,250.00	\$15,600.00	85	171	No
176	Mujeres	05-oct-62	16	Administrativo	\$35,700.00	\$17,250.00	85	19	No
177	Mujeres	11-ago-62	12	Administrativo	\$28,500.00	\$16,500.00	85	69	No
178	Mujeres	17-ene-38	12	Administrativo	\$17,100.00	\$10,200.00	85	72	No
179	Hombres	19-ene-65	12	Administrativo	\$25,200.00	\$13,050.00	85	29	Si
180	Mujeres	23-ene-60	12	Administrativo	\$24,000.00	\$12,750.00	85	59	Si
181	Mujeres	04-ene-39	12	Administrativo	\$27,450.00	\$10,200.00	85	101	Si
182	Mujeres	08-mar-43	12	Administrativo	\$18,450.00	\$10,200.00	85	228	Si
183	Hombres	01-mar-60	15	Administrativo	\$39,300.00	\$15,750.00	84	72	No
184	Hombres	05-nov-64	15	Administrativo	\$38,850.00	\$15,000.00	84	53	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
185	Hombres	28-ago-30	8	Seguridad	\$30,750.00	\$15,000.00	84	380	No
186	Hombres	06-jul-64	15	Administrativo	\$37,500.00	\$20,400.00	84	33	No
187	Hombres	29-jul-63	16	Directivo	\$58,750.00	\$21,750.00	84	13	No
188	Mujeres	04-feb-34	12	Administrativo	\$34,500.00	\$18,750.00	84	208	No
189	Mujeres	28-mar-50	12	Administrativo	\$36,000.00	\$19,980.00	84	240	No
190	Mujeres	15-nov-32	8	Administrativo	\$29,100.00	\$16,500.00	84	35	No
191	Mujeres	23-oct-31	12	Administrativo	\$16,500.00	\$10,200.00	84	288	No
192	Hombres	09-oct-49	12	Administrativo	\$19,650.00	\$12,750.00	84	180	Si
193	Mujeres	22-jul-66	12	Administrativo	\$24,750.00	\$12,000.00	84	41	Si
194	Mujeres	10-abr-38	15	Administrativo	\$27,150.00	\$15,750.00	84	231	Si
195	Mujeres	06-nov-55	12	Administrativo	\$26,400.00	\$12,750.00	84	36	Si
196	Mujeres	03-mar-46	16	Administrativo	\$23,100.00	\$12,000.00	84	214	Si
197	Hombres	01-oct-62	15	Directivo	\$54,900.00	\$25,500.00	83	49	No
198	Hombres	17-mar-52	19	Directivo	\$70,875.00	\$43,500.00	83	156	No
199	Hombres	02-feb-58	16	Directivo	\$51,250.00	\$27,480.00	83	69	No
200	Hombres	13-feb-63	17	Directivo	\$67,500.00	\$34,980.00	83	9	No
201	Hombres	08-may-55	12	Administrativo	\$29,340.00	\$19,500.00	83	150	No
202	Hombres	17-mar-63	15	Administrativo	\$39,600.00	\$16,500.00	83	47	No
203	Hombres	17-mar-64	12	Administrativo	\$29,100.00	\$15,000.00	83	50	No
204	Hombres	21-oct-60	15	Administrativo	\$33,150.00	\$16,500.00	83	69	No
205	Hombres	22-jun-44	16	Directivo	\$66,750.00	\$52,500.00	83	258	No
206	Hombres	22-may-43	12	Seguridad	\$33,750.00	\$15,000.00	83	284	No
207	Hombres	15-feb-59	15	Administrativo	\$27,300.00	\$17,250.00	83	91	No
208	Mujeres	28-nov-68	12	Administrativo	\$24,000.00	\$11,250.00	83	16	No
209	Mujeres	14-ene-34	8	Administrativo	\$19,800.00	\$10,200.00	83	75	No
210	Hombres	04-may-49	15	Administrativo	\$30,600.00	\$16,500.00	83	216	Si
211	Hombres	08-nov-50	15	Administrativo	\$28,950.00	\$15,000.00	83	108	Si
212	Hombres	14-may-65	15	Administrativo	\$38,400.00	\$16,500.00	83	64	Si
213	Hombres	20-oct-45	8	Seguridad	\$30,750.00	\$15,000.00	83	302	Si
214	Mujeres	13-sep-68	12	Administrativo	\$20,400.00	\$10,950.00	83	9	Si
215	Mujeres	15-sep-68	12	Administrativo	\$19,200.00	\$11,100.00	83	7	Si
216	Hombres	16-may-64	15	Administrativo	\$30,150.00	\$15,750.00	82	72	No
217	Hombres	15-mar-49	16	Administrativo	\$34,620.00	\$27,750.00	82	149	No
218	Hombres	21-mar-64	15	Administrativo	\$80,000.00	\$15,750.00	82	34	No
219	Hombres	02-feb-63	12	Administrativo	\$25,350.00	\$15,000.00	82	32	No
220	Hombres	17-ago-63	12	Administrativo	\$29,850.00	\$15,750.00	82	85	No
221	Mujeres	22-may-47	12	Administrativo	\$24,000.00	\$13,800.00	82	97	No
222	Mujeres	15-sep-35	15	Administrativo	\$27,750.00	\$19,500.00	82	265	No
223	Mujeres	14-mar-42	8	Administrativo	\$22,350.00	\$10,200.00	82	48	No
224	Mujeres	20-nov-34	12	Administrativo	\$16,200.00	\$10,200.00	82	Ausente	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
225	Mujeres	13-oct-67	15	Administrativo	\$21,900.00	\$12,750.00	82	Ausente	No
226	Mujeres	21-jun-64	16	Administrativo	\$23,250.00	\$15,750.00	82	4	No
227	Mujeres	17-jul-65	12	Administrativo	\$33,900.00	\$12,000.00	82	11	No
228	Mujeres	21-sep-63	15	Administrativo	\$25,650.00	\$14,250.00	82	51	No
229	Mujeres	21-ago-40	12	Administrativo	\$17,250.00	\$10,200.00	82	358	No
230	Hombres	04-feb-34	15	Administrativo	\$22,500.00	\$15,000.00	82	371	Si
231	Hombres	09-ene-65	16	Directivo	\$40,200.00	\$21,000.00	81	3	No
232	Hombres	13-sep-63	19	Directivo	\$55,500.00	\$33,750.00	81	62	No
233	Hombres	21-jul-63	15	Administrativo	\$26,550.00	\$15,000.00	81	52	No
234	Hombres	12-nov-60	15	Administrativo	\$50,550.00	\$19,500.00	81	44	No
235	Hombres	11-sep-62	19	Directivo	\$75,000.00	\$31,500.00	81	13	No
236	Mujeres	02-may-69	8	Administrativo	\$27,450.00	\$12,000.00	81	9	No
237	Mujeres	04-jun-69	12	Administrativo	\$22,650.00	\$11,250.00	81	Ausente	No
238	Mujeres	15-jun-69	12	Administrativo	\$27,300.00	\$11,250.00	81	5	No
239	Mujeres	14-sep-69	12	Administrativo	\$27,750.00	\$11,225.00	81	5	No
240	Mujeres	15-jun-61	16	Directivo	\$54,375.00	\$18,750.00	81	81	No
241	Mujeres	27-ago-36	8	Administrativo	\$17,400.00	\$10,200.00	81	390	No
242	Mujeres	03-nov-67	12	Administrativo	\$40,800.00	\$18,000.00	81	4	No
243	Mujeres	10-feb-69	12	Administrativo	\$23,100.00	\$10,950.00	81	Ausente	No
244	Mujeres	15-sep-69	8	Administrativo	\$22,500.00	\$10,950.00	81	5	No
245	Mujeres	16-mar-69	12	Administrativo	\$26,700.00	\$11,550.00	81	18	No
246	Mujeres	14-ene-68	12	Administrativo	\$24,900.00	\$11,250.00	81	Ausente	No
247	Mujeres	13-may-68	12	Administrativo	\$19,650.00	\$10,950.00	81	11	No
248	Mujeres	09-abr-69	12	Administrativo	\$22,050.00	\$10,950.00	81	9	No
249	Mujeres	09-may-69	12	Administrativo	\$25,500.00	\$12,000.00	81	11	No
250	Mujeres	10-oct-66	15	Administrativo	\$28,200.00	\$12,750.00	81	19	No
251	Mujeres	19-ene-69	12	Administrativo	\$23,100.00	\$11,250.00	81	13	No
252	Hombres	18-sep-69	12	Administrativo	\$25,500.00	\$11,400.00	81	9	Si
253	Mujeres	21-feb-42	8	Administrativo	\$17,100.00	\$10,200.00	81	Ausente	Si
254	Hombres	08-feb-64	18	Directivo	\$68,125.00	\$32,490.00	80	29	No
255	Hombres	15-ago-32	12	Seguridad	\$30,600.00	\$15,750.00	80	460	No
256	Hombres	03-ene-48	19	Directivo	\$52,125.00	\$27,480.00	80	221	No
257	Hombres	24-sep-51	19	Directivo	\$61,875.00	\$36,750.00	80	199	No
258	Hombres	09-mar-69	8	Administrativo	\$21,300.00	\$11,550.00	80	24	No
259	Mujeres	15-mar-69	12	Administrativo	\$19,650.00	\$11,250.00	80	5	No
260	Mujeres	20-feb-69	12	Administrativo	\$22,350.00	\$11,250.00	80	5	No
261	Mujeres	27-mar-69	12	Administrativo	\$23,400.00	\$11,250.00	80	18	No
262	Mujeres	20-jun-68	12	Administrativo	\$24,300.00	\$10,950.00	80	8	No
263	Mujeres	13-oct-68	12	Administrativo	\$28,500.00	\$11,250.00	80	4	No
264	Mujeres	16-ene-69	12	Administrativo	\$19,950.00	\$11,250.00	80	8	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
265	Mujeres	24-sep-69	12	Administrativo	\$23,400.00	\$11,250.00	80	Ausente	No
266	Mujeres	07-oct-63	16	Administrativo	\$34,500.00	\$17,250.00	80	3	No
267	Mujeres	24-ago-69	12	Administrativo	\$18,150.00	\$10,950.00	80	Ausente	No
268	Hombres	16-may-36	12	Administrativo	\$22,350.00	\$15,000.00	80	272	Si
269	Hombres	17-mar-66	15	Administrativo	\$40,200.00	\$17,250.00	79	38	No
270	Hombres	29-oct-47	15	Administrativo	\$28,650.00	\$18,000.00	79	261	No
271	Hombres	24-ago-63	15	Administrativo	\$27,750.00	\$16,500.00	79	55	No
272	Hombres	17-jun-64	18	Administrativo	\$66,875.00	\$31,980.00	79	30	No
273	Hombres	12-nov-39	12	Seguridad	\$30,000.00	\$15,750.00	79	308	No
274	Hombres	04-ago-64	16	Directivo	\$83,750.00	\$21,750.00	79	12	No
275	Hombres	14-ene-63	12	Administrativo	\$33,900.00	\$16,500.00	79	94	No
276	Hombres	11-may-65	16	Directivo	\$56,500.00	\$21,000.00	79	12	No
277	Mujeres	20-may-65	16	Directivo	\$43,000.00	\$17,490.00	79	20	No
278	Mujeres	12-jun-43	8	Administrativo	\$20,850.00	\$12,000.00	79	70	No
279	Mujeres	16-abr-69	12	Administrativo	\$24,450.00	\$12,000.00	79	8	No
280	Mujeres	20-oct-69	12	Administrativo	\$24,750.00	\$10,950.00	79	5	No
281	Hombres	18-feb-45	8	Seguridad	\$34,500.00	\$15,750.00	79	246	Si
282	Hombres	15-sep-63	14	Administrativo	\$27,900.00	\$15,000.00	79	47	Si
283	Hombres	25-feb-63	19	Directivo	\$68,125.00	\$32,010.00	78	35	No
284	Hombres	10-oct-61	19	Directivo	\$73,500.00	\$33,000.00	78	45	No
285	Hombres	28-may-30	8	Seguridad	\$30,750.00	\$15,750.00	78	429	No
286	Hombres	07-mar-58	15	Directivo	\$40,050.00	\$25,500.00	78	133	No
287	Hombres	18-ene-65	16	Administrativo	\$40,350.00	\$19,500.00	78	20	No
288	Hombres	22-ene-55	15	Directivo	\$38,700.00	\$23,730.00	78	176	No
289	Hombres	27-oct-64	17	Directivo	\$65,000.00	\$30,750.00	78	26	No
290	Hombres	05-abr-54	18	Directivo	\$51,450.00	\$36,240.00	78	149	No
291	Hombres	23-sep-34	12	Seguridad	\$35,250.00	\$15,750.00	78	387	No
292	Hombres	07-may-63	14	Administrativo	\$25,950.00	\$15,000.00	78	53	No
293	Hombres	19-sep-65	15	Administrativo	\$25,050.00	\$14,250.00	78	24	No
294	Hombres	10-oct-68	12	Administrativo	\$26,700.00	\$12,750.00	78	25	No
295	Hombres	20-ago-32	8	Administrativo	\$24,000.00	\$15,750.00	78	476	No
296	Hombres	12-feb-64	12	Administrativo	\$26,850.00	\$15,000.00	78	48	No
297	Mujeres	16-abr-42	12	Administrativo	\$23,400.00	\$15,300.00	78	209	No
298	Mujeres	24-ago-66	12	Administrativo	\$24,600.00	\$13,500.00	78	47	No
299	Mujeres	11-may-65	15	Administrativo	\$32,550.00	\$18,000.00	78	6	No
300	Hombres	26-may-60	16	Administrativo	\$26,550.00	\$15,000.00	78	105	Si
301	Hombres	04-ago-68	12	Administrativo	\$31,500.00	\$13,500.00	78	7	Si
302	Hombres	28-sep-39	8	Administrativo	\$22,350.00	\$15,000.00	78	320	Si
303	Hombres	10-feb-38	12	Seguridad	\$35,250.00	\$15,750.00	78	281	Si
304	Mujeres	28-sep-45	15	Administrativo	\$25,800.00	\$13,500.00	78	51	Si

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
305	Hombres	25-oct-41	12	Seguridad	\$30,750.00	\$15,750.00	77	317	No
306	Hombres	05-abr-66	15	Administrativo	\$30,750.00	\$16,500.00	77	41	No
307	Hombres	24-jul-45	16	Directivo	\$50,000.00	\$32,490.00	77	264	No
308	Hombres	01-feb-63	15	Administrativo	\$34,500.00	\$18,000.00	77	63	No
309	Hombres	19-jul-64	15	Administrativo	\$26,250.00	\$15,750.00	77	38	No
310	Hombres	23-mar-64	16	Administrativo	\$44,875.00	\$21,240.00	77	22	No
311	Mujeres	15-feb-51	12	Administrativo	\$22,500.00	\$12,000.00	77	63	No
312	Mujeres	12-mar-63	12	Administrativo	\$25,650.00	\$14,250.00	77	64	No
313	Mujeres	25-oct-69	12	Administrativo	\$21,300.00	\$11,250.00	77	Ausente	No
314	Mujeres	24-nov-66	12	Administrativo	\$29,850.00	\$13,500.00	77	38	No
315	Mujeres	19-sep-69	12	Administrativo	\$34,500.00	\$12,150.00	77	4	No
316	Hombres	14-ene-60	15	Administrativo	\$27,750.00	\$15,000.00	77	52	Si
317	Hombres	27-feb-69	12	Administrativo	\$27,750.00	\$11,550.00	77	12	Si
318	Hombres	19-feb-61	16	Directivo	\$48,750.00	\$21,990.00	76	61	No
319	Hombres	02-mar-66	15	Administrativo	\$43,410.00	\$15,750.00	76	12	No
320	Hombres	10-nov-36	12	Administrativo	\$22,050.00	\$15,000.00	76	385	No
321	Mujeres	26-oct-53	12	Administrativo	\$22,050.00	\$12,000.00	76	6	No
322	Mujeres	02-mar-38	12	Administrativo	\$22,500.00	\$14,250.00	76	90	No
323	Mujeres	13-may-67	15	Administrativo	\$25,500.00	\$12,000.00	76	7	No
324	Mujeres	14-oct-66	12	Administrativo	\$29,160.00	\$15,000.00	76	22	No
325	Mujeres	04-nov-34	8	Administrativo	\$16,800.00	\$10,200.00	76	76	No
326	Hombres	24-jul-58	8	Seguridad	\$29,550.00	\$15,750.00	76	144	Si
327	Hombres	02-mar-65	12	Administrativo	\$26,700.00	\$15,750.00	76	18	Si
328	Hombres	25-oct-52	18	Directivo	\$55,000.00	\$32,490.00	75	125	No
329	Hombres	24-ago-58	18	Directivo	\$62,500.00	\$34,980.00	75	74	No
330	Hombres	09-may-59	15	Administrativo	\$27,300.00	\$17,250.00	75	132	No
331	Mujeres	04-feb-42	12	Administrativo	\$24,450.00	\$12,000.00	75	144	No
332	Mujeres	10-may-64	16	Administrativo	\$33,000.00	\$18,000.00	75	26	No
333	Mujeres	26-ene-65	15	Administrativo	\$37,050.00	\$18,000.00	75	5	No
334	Mujeres	09-may-66	12	Administrativo	\$24,450.00	\$10,950.00	75	32	Si
335	Hombres	26-feb-30	8	Seguridad	\$31,950.00	\$15,750.00	74	408	No
336	Hombres	10-ene-64	16	Directivo	\$47,250.00	\$21,240.00	74	45	No
337	Mujeres	27-ene-70	12	Administrativo	\$26,100.00	\$11,550.00	74	2	No
338	Mujeres	12-ago-38	8	Administrativo	\$15,900.00	\$10,200.00	74	43	No
339	Mujeres	07-nov-42	8	Administrativo	\$23,700.00	\$10,650.00	74	281	No
340	Mujeres	06-may-34	8	Administrativo	\$21,750.00	\$12,450.00	74	318	No
341	Hombres	20-ene-45	12	Directivo	\$59,400.00	\$33,750.00	74	272	Si
342	Mujeres	01-jun-48	12	Administrativo	\$24,450.00	\$14,250.00	74	117	Si
343	Hombres	09-jun-53	16	Directivo	\$103,500.00	\$60,000.00	73	150	No
344	Hombres	13-oct-63	12	Administrativo	\$35,700.00	\$16,500.00	73	72	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
345	Mujeres	05-may-69	12	Administrativo	\$22,200.00	\$16,500.00	73	7	No
346	Mujeres	16-ago-68	15	Administrativo	\$22,950.00	\$13,950.00	73	22	No
347	Mujeres	10-ago-44	12	Administrativo	\$23,100.00	\$12,000.00	73	228	No
348	Mujeres	11-jun-62	16	Directivo	\$56,750.00	\$30,000.00	73	15	No
349	Mujeres	22-mar-38	17	Administrativo	\$29,100.00	\$12,750.00	73	375	No
350	Mujeres	21-abr-39	12	Administrativo	\$37,650.00	\$15,750.00	73	132	No
351	Mujeres	27-ene-66	12	Administrativo	\$27,900.00	\$13,500.00	73	32	No
352	Mujeres	26-nov-33	8	Administrativo	\$21,150.00	\$12,000.00	73	159	No
353	Hombres	10-mar-55	12	Seguridad	\$31,200.00	\$15,750.00	73	155	Si
354	Mujeres	08-may-46	12	Administrativo	\$20,550.00	\$11,250.00	73	154	Si
355	Hombres	19-may-61	15	Administrativo	\$25,950.00	\$17,250.00	72	83	No
356	Hombres	10-nov-65	15	Administrativo	\$28,350.00	\$15,000.00	72	48	No
357	Mujeres	18-ene-32	8	Administrativo	\$17,700.00	\$10,200.00	72	184	No
358	Mujeres	11-abr-44	15	Administrativo	\$23,550.00	\$13,500.00	72	49	No
359	Mujeres	12-jun-41	12	Administrativo	\$19,950.00	\$10,200.00	72	56	No
360	Mujeres	11-ago-68	12	Administrativo	\$29,400.00	\$15,300.00	72	30	No
361	Mujeres	09-jun-66	12	Administrativo	\$28,800.00	\$13,950.00	72	18	No
362	Mujeres	08-abr-37	8	Administrativo	\$16,950.00	\$10,200.00	72	319	No
363	Hombres	03-jun-54	16	Administrativo	\$35,700.00	\$18,000.00	72	138	Si
364	Mujeres	26-mar-47	12	Administrativo	\$17,400.00	\$10,200.00	72	116	Si
365	Mujeres	16-oct-48	8	Administrativo	\$21,450.00	\$10,200.00	72	194	Si
366	Mujeres	21-mar-61	12	Administrativo	\$24,750.00	\$12,000.00	72	68	Si
367	Mujeres	22-may-42	12	Administrativo	\$16,950.00	\$10,200.00	72	271	Si
368	Mujeres	20-jul-43	12	Administrativo	\$26,100.00	\$13,500.00	72	169	Si
369	Hombres	14-nov-67	14	Administrativo	\$28,050.00	\$15,000.00	71	15	No
370	Mujeres	20-ene-63	16	Administrativo	\$36,600.00	\$18,000.00	71	12	No
371	Mujeres	05-abr-66	16	Directivo	\$58,125.00	\$18,000.00	71	11	No
372	Hombres	22-ago-35	15	Administrativo	\$21,300.00	\$15,750.00	70	372	No
373	Hombres	29-oct-49	12	Administrativo	\$22,500.00	\$16,500.00	70	216	No
374	Hombres	26-jul-67	15	Administrativo	\$29,400.00	\$15,750.00	70	15	No
375	Hombres	08-oct-66	12	Administrativo	\$27,450.00	\$14,700.00	70	41	No
376	Hombres	09-oct-64	15	Administrativo	\$29,850.00	\$15,750.00	70	48	No
377	Hombres	29-nov-65	15	Administrativo	\$25,350.00	\$15,750.00	70	56	No
378	Mujeres	21-sep-30	8	Administrativo	\$15,750.00	\$10,200.00	70	275	No
379	Mujeres	12-may-38	8	Administrativo	\$19,650.00	\$13,050.00	70	102	No
380	Mujeres	22-feb-41	12	Administrativo	\$21,000.00	\$13,500.00	70	82	No
381	Hombres	15-jul-46	17	Administrativo	\$27,000.00	\$18,000.00	70	192	Si
382	Hombres	20-oct-59	12	Administrativo	\$24,000.00	\$15,750.00	70	120	Si
383	Hombres	03-jun-61	17	Directivo	\$78,500.00	\$28,740.00	70	67	Si
384	Mujeres	11-nov-55	12	Administrativo	\$20,850.00	\$13,050.00	70	127	Si

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
385	Hombres	01-oct-30	12	Seguridad	\$30,000.00	\$15,750.00	69	348	No
386	Hombres	18-ago-34	8	Seguridad	\$28,500.00	\$15,750.00	69	174	No
387	Hombres	03-feb-65	19	Directivo	\$65,000.00	\$31,980.00	69	74	No
388	Hombres	02-ene-59	14	Administrativo	\$30,150.00	\$16,500.00	69	110	No
389	Hombres	15-abr-59	19	Directivo	\$66,875.00	\$32,490.00	69	81	No
390	Mujeres	09-nov-68	15	Administrativo	\$24,150.00	\$13,500.00	69	7	No
391	Mujeres	12-ene-69	12	Administrativo	\$24,450.00	\$12,450.00	69	12	No
392	Mujeres	12-may-70	12	Administrativo	\$21,600.00	\$12,000.00	69	Ausente	No
393	Mujeres	24-jun-69	12	Administrativo	\$27,900.00	\$12,450.00	69	Ausente	No
394	Mujeres	04-feb-70	8	Administrativo	\$29,100.00	\$12,450.00	69	17	No
395	Mujeres	09-mar-70	12	Administrativo	\$22,650.00	\$11,250.00	69	2	No
396	Mujeres	17-ago-70	12	Administrativo	\$20,850.00	\$11,250.00	69	Ausente	No
397	Mujeres	17-ene-70	12	Administrativo	\$22,950.00	\$12,300.00	69	5	No
398	Mujeres	21-nov-70	12	Administrativo	\$30,600.00	\$12,450.00	69	5	No
399	Mujeres	06-feb-70	12	Administrativo	\$20,400.00	\$11,250.00	69	Ausente	No
400	Mujeres	06-ago-69	12	Administrativo	\$23,850.00	\$12,750.00	69	20	No
401	Mujeres	14-mar-70	12	Administrativo	\$22,800.00	\$11,250.00	69	Ausente	No
402	Mujeres	07-feb-70	12	Administrativo	\$20,700.00	\$11,250.00	69	2	No
403	Mujeres	28-abr-70	12	Administrativo	\$21,300.00	\$11,250.00	69	3	Si
404	Mujeres	01-may-53	12	Administrativo	\$24,300.00	\$15,000.00	69	121	Si
405	Mujeres	12-jul-44	12	Administrativo	\$19,650.00	\$13,950.00	69	133	Si
406	Hombres	05-oct-65	17	Directivo	\$60,000.00	\$32,490.00	68	17	No
407	Hombres	06-sep-65	15	Administrativo	\$30,300.00	\$15,750.00	68	55	No
408	Hombres	12-feb-64	19	Directivo	\$61,250.00	\$33,000.00	68	9	No
409	Hombres	21-nov-62	19	Administrativo	\$36,000.00	\$19,500.00	68	21	No
410	Mujeres	09-ene-42	8	Administrativo	\$25,200.00	\$18,750.00	68	344	No
411	Mujeres	21-ago-31	12	Administrativo	\$16,200.00	\$10,200.00	68	180	No
412	Mujeres	16-jun-70	12	Administrativo	\$22,800.00	\$11,250.00	68	2	No
413	Mujeres	13-mar-66	16	Directivo	\$43,500.00	\$19,500.00	68	11	No
414	Hombres	08-ene-61	8	Seguridad	\$30,300.00	\$15,750.00	68	155	Si
415	Hombres	02-may-63	15	Administrativo	\$31,950.00	\$15,750.00	68	70	Si
416	Hombres	16-ene-65	15	Administrativo	\$35,250.00	\$13,500.00	67	6	No
417	Hombres	22-jul-67	15	Administrativo	\$37,800.00	\$15,000.00	67	36	No
418	Hombres	05-jun-65	15	Administrativo	\$31,200.00	\$15,750.00	67	46	No
419	Hombres	20-oct-64	15	Administrativo	\$29,400.00	\$16,500.00	67	68	No
420	Hombres	16-jul-57	19	Directivo	\$70,000.00	\$35,040.00	67	75	No
421	Hombres	01-nov-61	15	Administrativo	\$33,900.00	\$15,750.00	67	96	No
422	Hombres	20-jul-63	15	Administrativo	\$27,150.00	\$16,500.00	67	78	No
423	Mujeres	26-abr-36	12	Administrativo	\$22,200.00	\$13,800.00	67	196	No
424	Mujeres	06-jul-66	15	Administrativo	\$31,350.00	\$11,100.00	67	47	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
425	Mujeres	26-ene-42	12	Administrativo	\$20,850.00	\$13,500.00	67	181	No
426	Hombres	27-oct-52	16	Administrativo	\$33,300.00	\$17,490.00	67	120	Si
427	Hombres	21-abr-42	8	Administrativo	\$26,250.00	\$16,050.00	67	97	Si
428	Hombres	28-feb-64	15	Administrativo	\$31,950.00	\$15,750.00	67	58	Si
429	Hombres	13-ago-46	8	Seguridad	\$30,000.00	\$15,750.00	67	305	Si
430	Hombres	29-abr-56	19	Directivo	\$66,250.00	\$34,980.00	67	99	Si
431	Hombres	15-ene-59	18	Directivo	\$86,250.00	\$45,000.00	66	50	No
432	Hombres	12-mar-64	12	Administrativo	\$30,750.00	\$15,000.00	66	56	No
433	Hombres	15-nov-66	12	Administrativo	\$33,540.00	\$15,750.00	66	47	No
434	Hombres	#NULL!	16	Administrativo	\$34,950.00	\$20,250.00	66	55	No
435	Hombres	24-may-64	15	Administrativo	\$40,350.00	\$16,500.00	66	80	No
436	Hombres	27-may-65	12	Administrativo	\$30,270.00	\$15,750.00	66	80	No
437	Hombres	04-sep-33	8	Administrativo	\$26,250.00	\$16,050.00	66	264	No
438	Hombres	11-oct-64	15	Administrativo	\$32,400.00	\$15,000.00	66	64	No
439	Mujeres	25-jun-70	12	Administrativo	\$20,400.00	\$11,250.00	66	Ausente	No
440	Mujeres	10-nov-47	8	Administrativo	\$24,150.00	\$12,750.00	66	96	No
441	Mujeres	02-ago-49	15	Administrativo	\$23,850.00	\$13,500.00	66	122	No
442	Mujeres	18-sep-68	12	Administrativo	\$29,700.00	\$13,500.00	66	26	No
443	Mujeres	10-feb-29	8	Administrativo	\$21,600.00	\$13,500.00	66	228	No
444	Mujeres	16-sep-61	12	Administrativo	\$24,450.00	\$15,750.00	66	87	No
445	Hombres	04-ago-63	15	Administrativo	\$28,050.00	\$16,500.00	66	84	Si
446	Hombres	03-ago-58	16	Directivo	\$100,000.00	\$44,100.00	66	128	Si
447	Hombres	02-jul-61	15	Administrativo	\$49,000.00	\$20,550.00	66	86	Si
448	Mujeres	05-jun-33	12	Administrativo	\$16,350.00	\$10,200.00	66	163	Si
449	Hombres	02-ene-66	16	Directivo	\$70,000.00	\$21,750.00	65	19	No
450	Hombres	21-jul-54	19	Directivo	\$55,000.00	\$34,980.00	65	129	No
451	Hombres	19-jul-69	15	Administrativo	\$28,500.00	\$14,250.00	65	20	No
452	Hombres	13-ago-53	12	Administrativo	\$28,800.00	\$18,000.00	65	210	No
453	Hombres	07-ago-30	15	Administrativo	\$24,450.00	\$15,750.00	65	338	No
454	Hombres	28-jul-65	19	Directivo	\$90,625.00	\$31,250.00	65	18	No
455	Hombres	17-ene-64	16	Directivo	\$43,650.00	\$19,500.00	65	19	No
456	Hombres	17-oct-59	19	Directivo	\$75,000.00	\$42,510.00	65	54	No
457	Hombres	27-may-68	15	Administrativo	\$31,650.00	\$14,250.00	65	10	No
458	Hombres	06-jul-65	19	Directivo	\$61,875.00	\$28,740.00	65	26	No
459	Mujeres	10-feb-71	12	Administrativo	\$21,750.00	\$11,250.00	65	Ausente	No
460	Mujeres	12-ago-69	12	Administrativo	\$22,500.00	\$12,750.00	65	24	No
461	Mujeres	08-nov-43	8	Administrativo	\$21,600.00	\$13,500.00	65	173	No
462	Mujeres	18-oct-63	16	Directivo	\$34,410.00	\$19,500.00	65	79	No
463	Mujeres	15-oct-34	15	Administrativo	\$20,700.00	\$14,250.00	65	241	No
464	Hombres	20-mar-62	19	Directivo	\$47,550.00	\$33,000.00	64	27	No

id	sexo	fechnac	educ	catlab	salario	salini	tiempemp	expprev	minoría
465	Hombres	20-jul-62	12	Administrativo	\$33,900.00	\$16,500.00	64	106	No
466	Mujeres	15-jun-48	12	Administrativo	\$23,400.00	\$13,500.00	64	198	No
467	Mujeres	18-ago-67	16	Administrativo	\$32,850.00	\$19,500.00	64	20	No
468	Mujeres	28-nov-65	16	Directivo	\$55,750.00	\$19,980.00	64	36	No
469	Mujeres	01-jun-64	15	Administrativo	\$25,200.00	\$13,950.00	64	57	No
470	Hombres	22-ene-64	12	Administrativo	\$26,250.00	\$15,750.00	64	69	Si
471	Hombres	03-ago-66	15	Administrativo	\$26,400.00	\$15,750.00	64	32	Si
472	Hombres	21-feb-66	15	Administrativo	\$39,150.00	\$15,750.00	63	46	No
473	Mujeres	25-nov-37	12	Administrativo	\$21,450.00	\$12,750.00	63	139	No
474	Mujeres	05-nov-68	12	Administrativo	\$29,400.00	\$14,250.00	63	9	No

Base de Datos Deserción_d_extorsión.sav

Departamento	Municipio	cambio domic ilio	traslado otra c	emigra a Ed ucame	aband ona P ais	Total_d esercio n	año	Delito	Total_d elitos1	Rural	Urbana	Total_ delitos 2
AHUACHAPÁN	AHUACHAPÁN	676	209	6	42	933	2013	EXTORSIONES	37		37	37
AHUACHAPÁN	APANCA	84	38	1	1	124	2013	EXTORSIONES	2		2	2
AHUACHAPÁN	ATQUIZAYA	166	48	3	2	219	2013	EXTORSIONES	12		12	12
AHUACHAPÁN	ATACO	94	8		4	106	2013	EXTORSIONES	1		1	1
AHUACHAPÁN	EL REFUGIO	93	28	3	1	125	2013	EXTORSIONES	2		2	2
AHUACHAPÁN	GUAYMANGO	109	32		11	152	2013	EXTORSIONES	3		3	3
AHUACHAPÁN	JUJUTLA	136	74	3	59	272	2013	EXTORSIONES	0			
AHUACHAPÁN	SAN FRANCISCO MENENDEZ	158	118	20	114	410	2013	EXTORSIONES	2		2	2
AHUACHAPÁN	SAN LORENZO	46	5	4	2	57	2013	EXTORSIONES	1		1	1
AHUACHAPÁN	SAN PEDRO PUXTLA	17	5	4	0	26	2013	EXTORSIONES	0			
AHUACHAPÁN	TACUBA	148	13	1	7	169	2013	EXTORSIONES	5		5	5
AHUACHAPÁN	TURÍN	115	8	2	2	127	2013	EXTORSIONES	2		2	2
SANTA ANA	CANDELARIA DE LA FRONTERA	240	29	5	13	287	2013	EXTORSIONES	4		4	4
SANTA ANA	COATEPEQUE	226	44	3	29	302	2013	EXTORSIONES	3	1	2	3
SANTA ANA	CHALCHUAPA	528	226	12	38	804	2013	EXTORSIONES	28	3	25	28
SANTA ANA	EL CONGO	140	66	2	14	222	2013	EXTORSIONES	6		6	6
SANTA ANA	EL PORVENIR	92	24		2	118	2013	EXTORSIONES	2		2	2
SANTA ANA	MASAHUAT	22			7	29	2013	EXTORSIONES	16	1	15	16
SANTA ANA	METAPAN	374	62	1	114	551	2013	EXTORSIONES	0			
SANTA ANA	SAN ANTONIO PAJONAL	12	6		14	32	2013	EXTORSIONES	0			
SANTA ANA	SAN SEBASTIAN SALITRILLO	115	23	1	7	146	2013	EXTORSIONES	6		6	6
SANTA ANA	SANTA ANA	1718	512	6	120	2356	2013	EXTORSIONES	266	5	261	266
SANTA ANA	SANTA ROSA GUACHIPILÍN	49	21		12	82	2013	EXTORSIONES	1		1	1
SANTA ANA	SANTIAGO DE LA FRONTERA	12	10		11	33	2013	EXTORSIONES	0			
SANTA ANA	TEXISTEPEQUE	77	44	3	33	157	2013	EXTORSIONES	1		1	1
SONSONATE	ACAJUTLA	276	104	8	57	445	2013	EXTORSIONES	9	1	8	9
SONSONATE	ARMENIA	187	42	3	10	242	2013	EXTORSIONES	17	2	15	17
SONSONATE	CALUCO	57	9		6	72	2013	EXTORSIONES	0			

Departamento	Municipio	cambio domic- ilio	traslado otra es- cuela	emigra- ción	aband- ona	Total d esercio n	año	Delito	Total d delitos	Rural	Urbana	Total delitos 2
SONSONATE	CUISNAHUAT	21	24		1	46	2013	EXTORSIONES	1		1	1
SONSONATE	ISHUATAN	61	20	12	4	97	2013	EXTORSIONES	1		1	1
SONSONATE	IZALCO	325	115	6	10	456	2013	EXTORSIONES	32		32	32
SONSONATE	JUAYÚA	162	27		2	191	2013	EXTORSIONES	17	3	14	17
SONSONATE	NAHUZALCO	220	51		3	274	2013	EXTORSIONES	24	3	21	24
SONSONATE	NAHULINGO	24	24		0	48	2013	EXTORSIONES	1		1	1
SONSONATE	SALCOATITÁN	41	8		0	49	2013	EXTORSIONES	3		3	3
SONSONATE	SAN ANTONIO DEL MONTE	108	20		3	131	2013	EXTORSIONES	8	1	7	8
SONSONATE	SAN JULIÁN	100	23	4	26	153	2013	EXTORSIONES	11	1	10	11
SONSONATE	SANTA CATARINA											
SONSONATE	MASAHUAT	14	3	1	1	19	2013	EXTORSIONES	4		4	4
SONSONATE	SANTO DOMINGO DE GUZMAN	28	4		0	32	2013	EXTORSIONES	4	1	3	4
SONSONATE	SONSONATE	669	193	26	137	1025	2013	EXTORSIONES	42	4	38	42
SONSONATE	SONZACATE	171	66	9	24	270	2013	EXTORSIONES	10		10	10
CHALATENANGO	AGUA CALIENTE	32	12		41	85	2013	EXTORSIONES	3		3	3
CHALATENANGO	ARCATAO	7	4	1	8	20	2013	EXTORSIONES	0			
CHALATENANGO	AZACUALPA	7		1	2	10	2013	EXTORSIONES	0			
CHALATENANGO	CITALA	6	1		3	10	2013	EXTORSIONES	1		1	1
CHALATENANGO	COMALAPA	22	13		10	45	2013	EXTORSIONES	0			
CHALATENANGO	CONCEPCIÓN QUEZALTEPEQUE	13	17		13	43	2013	EXTORSIONES	0			
CHALATENANGO	CHALATENANGO	180	54	4	101	339	2013	EXTORSIONES	15		15	15
CHALATENANGO	DULCE NOMBRE DE MARÍA	20	27		35	82	2013	EXTORSIONES	0			
CHALATENANGO	EL CARRIZAL	17	22		5	44	2013	EXTORSIONES	1		1	1
CHALATENANGO	EL PARAISO	90	7	6	57	160	2013	EXTORSIONES	0			
CHALATENANGO	LA LAGUNA	4	16		15	35	2013	EXTORSIONES	0			
CHALATENANGO	LA PALMA	52	20		22	94	2013	EXTORSIONES	2		2	2
CHALATENANGO	LA REINA	22	17		31	70	2013	EXTORSIONES	0			
CHALATENANGO	LAS VUELTAS	2			5	7	2013	EXTORSIONES	0			
CHALATENANGO	NOMBRE DE JESÚS	10	12	2	34	58	2013	EXTORSIONES	0			
CHALATENANGO	NUEVA CONCEPCIÓN	185	57	2	100	344	2013	EXTORSIONES	8		8	8

Departamento	Municipio	cambio _domicilio	traslado _otra_esc	emigra _a_Educame	aband ona_Pais	Total_d esercio n	año	Delito	Total_d delitos1	Rural	Urbana	Total_ delitos 2
CHALATENANGO	NUEVA TRINIDAD	3	1		1	5	2013	EXTORSIONES	0			
CHALATENANGO	OJOS DE AGUA	12			7	19	2013	EXTORSIONES	0			
CHALATENANGO	POTONICO	10	2		2	14	2013	EXTORSIONES	0			
CHALATENANGO	SAN ANTONIO DE LA CRUZ	13	17		5	35	2013	EXTORSIONES	0			
CHALATENANGO	SAN ANTONIO DE LOS RANCHOS				7	7	2013	EXTORSIONES	0			
CHALATENANGO	SAN FERNANDO	18	3		4	25	2013	EXTORSIONES	0			
CHALATENANGO	SAN FRANCISCO LEMPA	2	2		4	8	2013	EXTORSIONES	0			
CHALATENANGO	SAN FRANCISCO MORAZÁN	7	19		1	27	2013	EXTORSIONES	0			
CHALATENANGO	SAN IGNACIO	24	26		13	63	2013	EXTORSIONES	1	1	1	1
CHALATENANGO	SAN ISIDRO LABRADOR		1		4	5	2013	EXTORSIONES	0			
CHALATENANGO	SAN JOSE CANCASQUE	7	1	7	7	22	2013	EXTORSIONES	0			
CHALATENANGO	SAN JOSE LAS FLORES	6		2	7	15	2013	EXTORSIONES	0			
CHALATENANGO	SAN LUIS DEL CARMEN	1			0	1	2013	EXTORSIONES	0			
CHALATENANGO	SAN MIGUEL DE MERCEDES	9	2		3	14	2013	EXTORSIONES	0			
CHALATENANGO	SAN RAFAEL	25	9		12	46	2013	EXTORSIONES	1	1	1	1
CHALATENANGO	SANTA RITA	39			2	41	2013	EXTORSIONES	1	1	1	1
CHALATENANGO	TEJUTLA	138	48	1	37	224	2013	EXTORSIONES	0			
LA LIBERTAD	ANTIGUO CUSCATLÁN	32	17	4	9	62	2013	EXTORSIONES	7	7	7	7
LA LIBERTAD	CIUDAD ARCE	444	174	4	53	675	2013	EXTORSIONES	8	8	8	8
LA LIBERTAD	COLÓN	811	188	11	37	1047	2013	EXTORSIONES	39	39	39	39
LA LIBERTAD	COMASAGUA	99	16	3	3	121	2013	EXTORSIONES	1	1	1	1
LA LIBERTAD	CHILTIUPAN	49	9		10	68	2013	EXTORSIONES	0			
LA LIBERTAD	HUIZÚCAR	54	5		0	59	2013	EXTORSIONES	2	2	2	2
LA LIBERTAD	JAYAQUE	136	6		3	145	2013	EXTORSIONES	1	1	1	1
LA LIBERTAD	JICALAPA	51	17		1	69	2013	EXTORSIONES	0			
LA LIBERTAD	LA LIBERTAD	549	95	9	39	692	2013	EXTORSIONES	45	45	45	45
LA LIBERTAD	NUEVO CUSCATLÁN	31			3	34	2013	EXTORSIONES	0			
LA LIBERTAD	SANTA TECLA	347	214	30	156	747	2013	EXTORSIONES	43		43	43
LA LIBERTAD	QUEZALTEPEQUE	345	81	1	42	469	2013	EXTORSIONES	19	1	18	19
LA LIBERTAD	SACACOYO	173	34	8	19	234	2013	EXTORSIONES	2		2	2

Departamento	Municipio	cambio domic ilio	traslado otra es c	emigra a Ed ucame	aband ona_P ais	Total d esercio n	año	Delito	Total d elitos1	Rural	Urbana	Total delitos 2
LA LIBERTAD	SAN JOSÉ VILLANUEVA	31	12		5	48	2013	EXTORSIONES	2		2	2
LA LIBERTAD	SAN JUAN OPICO	513	149	12	78	752	2013	EXTORSIONES	12		12	12
LA LIBERTAD	SAN MATÍAS	58	11	1	21	91	2013	EXTORSIONES	1	1		1
LA LIBERTAD	SAN PABLO TACACHICO	146	83	2	53	284	2013	EXTORSIONES	3		3	3
LA LIBERTAD	TAMANIQUE	102	43	1	33	179	2013	EXTORSIONES	4		4	4
LA LIBERTAD	TALNIQUE	28	12	1	1	42	2013	EXTORSIONES	0			
LA LIBERTAD	TEOTEPEQUE	63	4		21	88	2013	EXTORSIONES	0			
LA LIBERTAD	TEPECOYO	68	20		3	91	2013	EXTORSIONES	3		3	3
LA LIBERTAD	ZARAGOZA	164	49		8	221	2013	EXTORSIONES	2		2	2
SAN SALVADOR	AGUILARES	162	103	2	66	333	2013	EXTORSIONES	0			
SAN SALVADOR	APOPA	989	239	15	76	1319	2013	EXTORSIONES	30		30	30
SAN SALVADOR	AYUTUXTEPEQUE	163	61	5	16	245	2013	EXTORSIONES	8		8	8
SAN SALVADOR	CUSCATANCINGO	375	57	5	14	451	2013	EXTORSIONES	10		10	10
SAN SALVADOR	EL PAISNAL	131	35	5	25	196	2013	EXTORSIONES	8	1	7	8
SAN SALVADOR	GUAZAPA	135	66		6	207	2013	EXTORSIONES	9	1	8	9
SAN SALVADOR	ILOPANGO	818	142	1	46	1007	2013	EXTORSIONES	20	1	19	20
SAN SALVADOR	MEJICANOS	714	194	6	48	962	2013	EXTORSIONES	70	1	69	70
SAN SALVADOR	NEJAPA	93	57		3	153	2013	EXTORSIONES	11		11	11
SAN SALVADOR	PANCHIMALCO	185	54	1	4	244	2013	EXTORSIONES	23	2	21	23
SAN SALVADOR	ROSARIO DE MORA	57	22		3	82	2013	EXTORSIONES	1		1	1
SAN SALVADOR	SAN MARCOS	456	53	20	36	565	2013	EXTORSIONES	43		43	43
SAN SALVADOR	SAN MARTÍN	566	198	10	53	827	2013	EXTORSIONES	26		26	26
SAN SALVADOR	SAN SALVADOR	2160	785	54	298	3297	2013	EXTORSIONES	352	2	350	352
SAN SALVADOR	SANTIAGO TEXACUANGOS	95	26	1	6	128	2013	EXTORSIONES	6		6	6
SAN SALVADOR	SANTO TOMÁS	136	51		5	192	2013	EXTORSIONES	5		5	5
SAN SALVADOR	SOYAPANGO	1373	377	2	128	1880	2013	EXTORSIONES	63	1	62	63
SAN SALVADOR	TONACATEPEQUE	474	117	9	29	629	2013	EXTORSIONES	12		12	12
SAN SALVADOR	CIUDAD DELGADO	485	166	12	30	693	2013	EXTORSIONES	36		36	36
CUSCATLÁN	CANDELARIA	98	20		3	121	2013	EXTORSIONES	0			
CUSCATLÁN	COJUTEPEQUE	295	135	2	59	491	2013	EXTORSIONES	35	1	34	35
CUSCATLÁN	EL CARMEN	93	9		0	102	2013	EXTORSIONES	2		2	2

Departamento	Municipio	cambio _domicilio	traslado _otra_escuela	emigración _a_Educación	abandonos _por	Total_desercion	año	Delito	Total_delitos1	Rural	Urbana	Total_delitos2
CUSCATLÁN	EL ROSARIO	26	3		2	31	2013	EXTORSIONES	3	1	2	3
CUSCATLÁN	MONTE SAN JUAN	90	14	2	5	111	2013	EXTORSIONES	0			
CUSCATLÁN	ORATORIO DE CONCEPCIÓN	36	15	6	0	57	2013	EXTORSIONES	0			
CUSCATLÁN	SAN BARTOLOME PERULAPIA	77	12	2	1	92	2013	EXTORSIONES	0			
CUSCATLÁN	SAN CRISTOBAL	20	2		3	25	2013	EXTORSIONES	0			
CUSCATLÁN	SAN JOSE GUAYABAL	49	16	3	2	70	2013	EXTORSIONES	5		5	5
CUSCATLÁN	SAN PEDRO PERULAPAN	220	74		10	304	2013	EXTORSIONES	9	1	8	9
CUSCATLÁN	SAN RAFAEL CEDROS	95	39		49	183	2013	EXTORSIONES	10	1	9	10
CUSCATLÁN	SAN RAMÓN	44	11		6	61	2013	EXTORSIONES	1	1	1	1
CUSCATLÁN	SANTA CRUZ ANALQUITO	15	2		1	18	2013	EXTORSIONES	1	1	1	1
CUSCATLÁN	SANTA CRUZ MICHAPA	93	19		24	136	2013	EXTORSIONES	2	2	2	2
CUSCATLÁN	SUCHITOTO	140	42		41	223	2013	EXTORSIONES	6	2	4	6
CUSCATLÁN	TENANCINGO	54	13	2	3	72	2013	EXTORSIONES	0			
LA PAZ	CUYULTITAN	48	3		2	53	2013	EXTORSIONES	0			
LA PAZ	EL ROSARIO	84	13	1	34	132	2013	EXTORSIONES	1	1	1	1
LA PAZ	JERUSALEN	9			0	9	2013	EXTORSIONES	0			
LA PAZ	MERCEDES LA CEIBA	1	12		0	13	2013	EXTORSIONES	0			
LA PAZ	OLOCUILTA	193	35		14	242	2013	EXTORSIONES	6		6	6
LA PAZ	PARAISO DE OSORIO	18			1	19	2013	EXTORSIONES	0			
LA PAZ	SAN ANTONIO MASAHUAT	11	22		7	40	2013	EXTORSIONES	0			
LA PAZ	SAN EMIGDIO	16	38	4	1	59	2013	EXTORSIONES	0			
LA PAZ	SAN FRANCISCO CHINAMECA	71	23		5	99	2013	EXTORSIONES	0			
LA PAZ	SAN JUAN NONUALCO	154	40		17	211	2013	EXTORSIONES	5		5	5
LA PAZ	SAN JUAN TALPA	73	44		7	124	2013	EXTORSIONES	1	1	1	1
LA PAZ	SAN JUAN TEPEZONTES	32	6		8	46	2013	EXTORSIONES	0			
LA PAZ	SAN LUIS TALPA	239	50		30	319	2013	EXTORSIONES	5	1	4	5
LA PAZ	SAN MIGUEL TEPEZONTES	35	3	3	0	41	2013	EXTORSIONES	0			
LA PAZ	SAN PEDRO MASAHUAT	211	70	2	26	309	2013	EXTORSIONES	10		10	10
LA PAZ	SAN PEDRO NONUALCO	55	16	2	0	73	2013	EXTORSIONES	5		5	5
LA PAZ	SAN RAFAEL OBRAJUELO	64	22	1	14	101	2013	EXTORSIONES	11		11	11

Departamento	Municipio	cambio domic- ilio	traslado otra es- c	emigra- a Ed- ucame	aband- ona_P- ais	Total d esercio n	año	Delito	Total d elitos1	Rural	Urbana	Total delitos 2
LA PAZ	SANTA MARIA OSTUMA	38	15		1	54	2013	EXTORSIONES	1		1	1
LA PAZ	SANTIAGO NONUALCO	321	75	4	74	474	2013	EXTORSIONES	17	1	16	17
LA PAZ	TAPALHUACA	34	4	1	4	43	2013	EXTORSIONES	1		1	1
LA PAZ	ZACATECOLUCA	416	228	50	99	793	2013	EXTORSIONES	42		42	42
LA PAZ	SAN LUIS LA HERRADURA	167	80	2	45	294	2013	EXTORSIONES	15		15	15
CABAÑAS	CINQUERA	15	6	3	3	27	2013	EXTORSIONES	0			
CABAÑAS	GUACOTECTI	35	4	1	3	43	2013	EXTORSIONES	0			
CABAÑAS	ILOBASCO	451	84	6	252	793	2013	EXTORSIONES	22	1	21	22
CABAÑAS	JUTIAPA	73	2		35	110	2013	EXTORSIONES	0			
CABAÑAS	SAN ISIDRO	36	9	2	22	69	2013	EXTORSIONES	3		3	3
CABAÑAS	SENSUNTEPEQUE	242	86	55	122	505	2013	EXTORSIONES	29	1	28	29
CABAÑAS	TEJUTEPEQUE	38	22		10	70	2013	EXTORSIONES	2		2	2
CABAÑAS	VICTORIA	50	15		36	101	2013	EXTORSIONES	2		2	2
CABAÑAS	DOLORES	59	8		33	100	2013	EXTORSIONES	0			
SAN VICENTE	APASTEPEQUE	96	24	4	67	191	2013	EXTORSIONES	7		7	7
SAN VICENTE	GUADALUPE	54	5	1	4	64	2013	EXTORSIONES	0			
SAN VICENTE	SAN CAYETANO ISTEPEQUE	13	10		6	29	2013	EXTORSIONES	1		1	1
SAN VICENTE	SANTA CLARA	20	6		20	46	2013	EXTORSIONES	1		1	1
SAN VICENTE	SANTO DOMINGO	43	8	1	11	63	2013	EXTORSIONES	2		2	2
SAN VICENTE	SAN ESTEBAN CATARINA	25	14		16	55	2013	EXTORSIONES	3		3	3
SAN VICENTE	SAN ILDEFONSO	30	24	4	23	81	2013	EXTORSIONES	0			
SAN VICENTE	SAN LORENZO	36	7	4	17	64	2013	EXTORSIONES	3		3	3
SAN VICENTE	SAN SEBASTIAN	32	11		36	79	2013	EXTORSIONES	2		2	2
SAN VICENTE	SAN VICENTE	297	145	27	128	597	2013	EXTORSIONES	101	2	99	101
SAN VICENTE	TECOLUCA	213	68	1	82	364	2013	EXTORSIONES	9		9	9
SAN VICENTE	TEPETITAN				2	2	2013	EXTORSIONES	1		1	1
SAN VICENTE	VERAPAZ	33	9		4	46	2013	EXTORSIONES	0			
USULUTÁN	ALEGRIA	37	11		5	53	2013	EXTORSIONES	5		5	5
USULUTÁN	BERLIN	130	45		33	208	2013	EXTORSIONES	12	1	11	12
USULUTÁN	CALIFORNIA	36			0	36	2013	EXTORSIONES	1		1	1
USULUTÁN	CONCEPCION BATRES	73	21		29	123	2013	EXTORSIONES	5	1	4	5

Departamento	Municipio	cambio _domicilio	traslado _otra_escuela	emigra _a_Educación	abandona _aulas	Total_d _esercion	año	Delito	Total_d _delitos1	Rural	Urbana	Total_ _delitos2
USulután	VILLA EL TRIUNFO	48	20		17	85	2013	EXTORSIONES	4		4	4
USulután	EREGUAYQUIN	49	11	9	9	78	2013	EXTORSIONES	3		3	3
USulután	ESTANZUELAS	71	19		51	141	2013	EXTORSIONES	6		6	6
USulután	JIQUILISCO	329	132	2	197	660	2013	EXTORSIONES	22		22	22
USulután	JUCUAPA	115	49		19	183	2013	EXTORSIONES	21	2	19	21
USulután	JUCUAPAN	116	15		48	179	2013	EXTORSIONES	4		4	4
USulután	MERCEDES UMANA	100	15		27	142	2013	EXTORSIONES	8		8	8
USulután	NUEVA GRANADA	53	3		7	63	2013	EXTORSIONES	5	1	4	5
USulután	OZATLAN	62	19	3	30	114	2013	EXTORSIONES	3		3	3
USulután	PUERTO EL TRIUNFO	94	35	2	39	170	2013	EXTORSIONES	1		1	1
USulután	SAN AGUSTIN	28	7		4	39	2013	EXTORSIONES	0			
USulután	SAN BUENAVENTURA	9	2		1	12	2013	EXTORSIONES	0			
USulután	SAN DIONISIO	42	1		8	51	2013	EXTORSIONES	2		2	2
USulután	SANTA ELENA	143	32	3	95	273	2013	EXTORSIONES	5	1	4	5
USulután	SAN FRANCISCO JAVIER	33	4		14	51	2013	EXTORSIONES	1		1	1
USulután	SANTA MARIA	39	16	1	19	75	2013	EXTORSIONES	6		6	6
USulután	SANTIAGO DE MARIA	139	49	2	31	221	2013	EXTORSIONES	15		15	15
USulután	TECAPAN	37	3		10	50	2013	EXTORSIONES	4		4	4
USulután	USULUTÁN	673	245	27	205	1150	2013	EXTORSIONES	60	1	59	60
SAN MIGUEL	CAROLINA	45	12		20	77	2013	EXTORSIONES	1		1	1
SAN MIGUEL	CIUDAD BARRIOS	121	69	5	68	263	2013	EXTORSIONES	15		15	15
SAN MIGUEL	COMACARAN	25	4		26	55	2013	EXTORSIONES	1		1	1
SAN MIGUEL	CHAPELTIQUE	71	15		44	130	2013	EXTORSIONES	2	1	1	2
SAN MIGUEL	CHINAMECA	79	40	8	10	137	2013	EXTORSIONES	4		4	4
SAN MIGUEL	CHIRILAGUA	148	19		56	223	2013	EXTORSIONES	5	1	4	5
SAN MIGUEL	EL TRANSITO	125	90	2	49	266	2013	EXTORSIONES	13	2	11	13
SAN MIGUEL	LOLOTIQUE	52	22		19	93	2013	EXTORSIONES	3		3	3
SAN MIGUEL	MONCAGUA	114	81		46	241	2013	EXTORSIONES	7		7	7
SAN MIGUEL	NUEVA GUADALUPE	26	8		11	45	2013	EXTORSIONES	3		3	3
SAN MIGUEL	NUEVO EDEN DE SAN JUAN	16	4		8	28	2013	EXTORSIONES	0			
SAN MIGUEL	QUELEPA	20	7		3	30	2013	EXTORSIONES	2		2	2

Departamento	Municipio	cambio _domicilio	traslado _otra_es _c	emigra _a_Ed _ucame	aband ona_P _alis	Total d esercio _n	año	Delito	Total d elitos1	Rural	Urbana	Total delitos 2
SAN MIGUEL	SAN ANTONIO	17	6		8	31	2013	EXTORSIONES	0			
SAN MIGUEL	SAN GERARDO	24	1		16	41	2013	EXTORSIONES	1		1	1
SAN MIGUEL	SAN JORGE	11	6		16	33	2013	EXTORSIONES	1		1	1
SAN MIGUEL	SAN LUIS DE LA REINA	26	11		28	65	2013	EXTORSIONES	1		1	1
SAN MIGUEL	SAN MIGUEL	1881	468	11	407	2767	2013	EXTORSIONES	355	7	348	355
SAN MIGUEL	SAN RAFAEL ORIENTE	67	5	2	27	101	2013	EXTORSIONES	1		1	1
SAN MIGUEL	SESORI	45	14	2	36	97	2013	EXTORSIONES	2		2	2
SAN MIGUEL	ULUAZAPA	22	1	2	11	36	2013	EXTORSIONES	1	1		1
MORAZÁN	ARAMBALA	4	7		0	11	2013	EXTORSIONES	3		3	3
MORAZÁN	CACAOPERA	46	11	1	22	80	2013	EXTORSIONES	4		4	4
MORAZÁN	CORINTO	85	55	3	77	220	2013	EXTORSIONES	2		2	2
MORAZÁN	CHILANGA	9	13		14	36	2013	EXTORSIONES	5		5	5
MORAZÁN	DELICIAS DE CONCEPCION	13	10		0	23	2013	EXTORSIONES	1		1	1
MORAZÁN	EL DIVISADERO	31	15	3	18	67	2013	EXTORSIONES	5		5	5
MORAZÁN	EL ROSARIO	5	6		1	12	2013	EXTORSIONES	0			
MORAZÁN	GUALOCOCTI	4	3		1	8	2013	EXTORSIONES	1		1	1
MORAZÁN	GUATAJAGUA	61	37		17	115	2013	EXTORSIONES	3		3	3
MORAZÁN	JOATECA	4	8		14	26	2013	EXTORSIONES	1		1	1
MORAZÁN	JOCOAITIQUE	30	9		6	45	2013	EXTORSIONES	3		3	3
MORAZÁN	JOCORO	57	46		15	118	2013	EXTORSIONES	6	1	5	6
MORAZÁN	LOLOTQUILLO	8	8	1	10	27	2013	EXTORSIONES	1		1	1
MORAZÁN	MEANGUERA	26	18		33	77	2013	EXTORSIONES	4		4	4
MORAZÁN	OSICALA	45	19		10	74	2013	EXTORSIONES	9		9	9
MORAZÁN	PERQUIN	14	5		3	22	2013	EXTORSIONES	0			
MORAZÁN	SAN CARLOS	7	27		8	42	2013	EXTORSIONES	7		7	7
MORAZÁN	SAN FERNANDO	17	13	2	1	33	2013	EXTORSIONES	0			
MORAZÁN	SAN FRANCISCO GOTERA	333	43	7	68	451	2013	EXTORSIONES	101	4	97	101
MORAZÁN	SAN ISIDRO	11			1	12	2013	EXTORSIONES	0			
MORAZÁN	SAN SIMON	24	17		4	45	2013	EXTORSIONES	4		4	4
MORAZÁN	SENSEMBRA	20	4		8	32	2013	EXTORSIONES	0			
MORAZÁN	SOCIEDAD	72	21	1	32	126	2013	EXTORSIONES	7	1	6	7

Departamento	Municipio	cambio domic ilio	traslado otra_es c	emigra a_Ed ucame	aband ona_P a/s	Total_d esercio n	año	Delito	Total_d elitos1	Rural	Urbana	Total_ delitos 2
MORAZÁN	TOROLA	2	6	1	5	14	2013	EXTORSIONES	1		1	1
MORAZÁN	YAMABAL	11	17		2	30	2013	EXTORSIONES	4		4	4
MORAZÁN	YOLOAIQUIN	10	7		4	21	2013	EXTORSIONES	3		3	3
LA UNIÓN	ANAMOROS	55	28	1	40	124	2013	EXTORSIONES	6	1	5	6
LA UNIÓN	BOLIVAR	17	3		10	30	2013	EXTORSIONES	1		1	1
LA UNIÓN	CONCEPCION DE ORIENTE	33	10		12	55	2013	EXTORSIONES	1		1	1
LA UNIÓN	CONCHAGUA	205	87		106	398	2013	EXTORSIONES	6	1	5	6
LA UNIÓN	EL CARMEN	66	27		47	140	2013	EXTORSIONES	2		2	2
LA UNIÓN	EL SAUCE	12	8		11	31	2013	EXTORSIONES	1		1	1
LA UNIÓN	INTIPUCA	34	16		46	96	2013	EXTORSIONES	2	1	1	2
LA UNIÓN	LA UNION	197	108	6	200	511	2013	EXTORSIONES	31	1	30	31
LA UNIÓN	LISLIQUE	39	25	1	28	93	2013	EXTORSIONES	1		1	1
LA UNIÓN	MEANGUERA DEL GOLFO	9	1		26	36	2013	EXTORSIONES	0			
LA UNIÓN	NUEVA ESPARTA	26	31		29	86	2013	EXTORSIONES	0			
LA UNIÓN	PASAQUINA	85	46		48	179	2013	EXTORSIONES	1		1	1
LA UNIÓN	POLOROS	47	27		39	113	2013	EXTORSIONES	3	1	2	3
LA UNIÓN	SAN ALEJO	80	24	5	49	158	2013	EXTORSIONES	1		1	1
LA UNIÓN	SAN JOSE	6	4		3	13	2013	EXTORSIONES	0			
LA UNIÓN	SANTA ROSA DE LIMA	193	44	2	37	276	2013	EXTORSIONES	16		16	16
LA UNIÓN	YAYANTIQUE	47	8		7	62	2013	EXTORSIONES	1		1	1
LA UNIÓN	YUCUAIQUIN	30	15		6	51	2013	EXTORSIONES	0			

Base de Datos Deserción_d_homicidio.sav

DEPTO	Municipio	cambio _domicilio	traslado _otr_esc	emigra _a_Educame	abando no_Pais	Total_dese rccion	año	Delito	Total_d elitos_1	Mujer	Hom bre	ND	Total_d elitos_2
AHUACHAPÁN	AHUACHAPÁN	676	209	6	42	933	2013	HOMICIDIOS	27	4	23	0	27
AHUACHAPÁN	APANECA	84	38	1	1	124	2013	HOMICIDIOS	2	0	2	0	2
AHUACHAPÁN	ATIZAPAZ	166	48	3	2	219	2013	HOMICIDIOS	17	1	16	0	17
AHUACHAPÁN	ATACO	94	8		4	106	2013	HOMICIDIOS	2	0	2	0	2
AHUACHAPÁN	EL REFUGIO	93	28	3	1	125	2013	HOMICIDIOS	5	0	5	0	5
AHUACHAPÁN	GUAYMANGO	109	32		11	152	2013	HOMICIDIOS	2	0	2	0	2
AHUACHAPÁN	JUJUTLA	136	74	3	59	272	2013	HOMICIDIOS	12	2	10	0	12
AHUACHAPÁN	SAN FRANCISCO MENENDEZ	158	118	20	114	410	2013	HOMICIDIOS	13	1	11	1	13
AHUACHAPÁN	SAN LORENZO	46	5	4	2	57	2013	HOMICIDIOS	0				
AHUACHAPÁN	SAN PEDRO PUXTLA	17	5	4	0	26	2013	HOMICIDIOS	2	0	2	0	2
AHUACHAPÁN	TACUBA	148	13	1	7	169	2013	HOMICIDIOS	9	4	5	0	9
AHUACHAPÁN	TURIN	115	8	2	2	127	2013	HOMICIDIOS	4	0	4	0	4
SANTA ANA	CANDELARIA DE LA FRONTERA	240	29	5	13	287	2013	HOMICIDIOS	9	1	8	0	9
SANTA ANA	COATEPEQUE	226	44	3	29	302	2013	HOMICIDIOS	16	3	13	0	16
SANTA ANA	CHALCHUAPA	528	226	12	38	804	2013	HOMICIDIOS	39	2	37	0	39
SANTA ANA	EL CONGO	140	66	2	14	222	2013	HOMICIDIOS	9	1	8	0	9
SANTA ANA	EL PORVENIR	92	24		2	118	2013	HOMICIDIOS	2	0	2	0	2
SANTA ANA	MASAHUAT	22			7	29	2013	HOMICIDIOS	1	1	0	0	1
SANTA ANA	METAPAN	374	62	1	114	551	2013	HOMICIDIOS	14	2	12	0	14
SANTA ANA	SAN ANTONIO PAJONAL	12	6		14	32	2013	HOMICIDIOS	1	0	1	0	1
SANTA ANA	SAN SEBASTIAN SALITRILLO	115	23	1	7	146	2013	HOMICIDIOS	12	1	11	0	12
SANTA ANA	SANTA ANA	1718	512	6	120	2356	2013	HOMICIDIOS	73	3	70	0	73
SANTA ANA	SANTA ROSA GUACHIPILIN	49	21		12	82	2013	HOMICIDIOS	1	0	1	0	1
SANTA ANA	SANTIAGO DE LA FRONTERA	12	10		11	33	2013	HOMICIDIOS	0				
SANTA ANA	TEXISTEPEQUE	77	44	3	33	157	2013	HOMICIDIOS	3	0	3	0	3
SONSONATE	ACAJUTLA	276	104	8	57	445	2013	HOMICIDIOS	4	0	4	0	4

DEPTO	Municipio	cambio _domicilio	traslado _otr_esc	emigra _a_Educame	abando no_Pais	Total_dese rccion	año	Delito	Total d elitos_1	Mujer	Hom bre	ND	Total d elitos_2
SONSONATE	ARMENIA	187	42	3	10	242	2013	HOMICIDIOS	39	7	32	0	39
SONSONATE	CALUCO	57	9		6	72	2013	HOMICIDIOS	14	0	14	0	14
SONSONATE	CUISNAHUAT	21	24		1	46	2013	HOMICIDIOS	3	0	3	0	3
SONSONATE	ISHUATAN	61	20	12	4	97	2013	HOMICIDIOS	3	0	3	0	3
SONSONATE	IZALCO	325	115	6	10	456	2013	HOMICIDIOS	35	2	33	0	35
SONSONATE	JUAYUA	162	27		2	191	2013	HOMICIDIOS	12	1	11	0	12
SONSONATE	NAHUZALCO	220	51		3	274	2013	HOMICIDIOS	35	2	33	0	35
SONSONATE	NAHULINGO	24	24		0	48	2013	HOMICIDIOS	3	0	3	0	3
SONSONATE	SALCOATITAN	41	8		0	49	2013	HOMICIDIOS	1	0	1	0	1
SONSONATE	SAN ANTONIO DEL MONTE	108	20		3	131	2013	HOMICIDIOS	4	1	3	0	4
SONSONATE	SAN JULIAN	100	23	4	26	153	2013	HOMICIDIOS	3	1	2	0	3
SONSONATE	SANTA CATARINA MASAHUAT	14	3	1	1	19	2013	HOMICIDIOS	1	0	1	0	1
SONSONATE	SANTO DOMINGO DE GUZMAN	28	4		0	32	2013	HOMICIDIOS	0				
SONSONATE	SONSONATE	669	193	26	137	1025	2013	HOMICIDIOS	16	1	15	0	16
SONSONATE	SONZACATE	171	66	9	24	270	2013	HOMICIDIOS	2	1	1	0	2
CHALATENANGO	AGUA CALIENTE	32	12		41	85	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	ARCATAO	7	4	1	8	20	2013	HOMICIDIOS	2	0	2	0	2
CHALATENANGO	AZACUALPA	7		1	2	10	2013	HOMICIDIOS	0				
CHALATENANGO	CITALA	6	1		3	10	2013	HOMICIDIOS	2	0	2	0	2
CHALATENANGO	COMALAPA	22	13		10	45	2013	HOMICIDIOS	2	0	2	0	2
CHALATENANGO	CONCEPCION												
CHALATENANGO	QUEZALTEPEQUE	13	17		13	43	2013	HOMICIDIOS	5	3	2	0	5
CHALATENANGO	CHALATENANGO	180	54	4	101	339	2013	HOMICIDIOS	2	0	2	0	2
CHALATENANGO	DULCE NOMBRE DE MARIA	20	27		35	82	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	EL CARRIZAL	17	22		5	44	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	EL PARAISO	90	7	6	57	160	2013	HOMICIDIOS	5	1	4	0	5
CHALATENANGO	LA LAGUNA	4	16		15	35	2013	HOMICIDIOS	4	2	2	0	4
CHALATENANGO	LA PALMA	52	20		22	94	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	LA REINA	22	17		31	70	2013	HOMICIDIOS	0				
CHALATENANGO	LAS VUELTAS	2			5	7	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	NOMBRE DE JESUS	10	12	2	34	58	2013	HOMICIDIOS	0				
CHALATENANGO	QUEZALTEPEQUE	185	57	2	100	344	2013	HOMICIDIOS	7	0	7	0	7
CHALATENANGO	NUEVA TRINIDAD	3	1		1	5	2013	HOMICIDIOS	0				

DEPTO	Municipio	cambio _domicilio	traslado _otr_esc	emigra _a_Ed ucame	abando no_Pai s	Total_dese rcion	año	Delito	Total_d elitos_1	Mujer	Hom bre	ND	Total_d elitos_2
CHALATENANGO	OJOS DE AGUA	12			7	19	2013	HOMICIDIOS	0				
CHALATENANGO	POTONICO	10	2		2	14	2013	HOMICIDIOS	0				
CHALATENANGO	SAN ANTONIO DE LA CRUZ	13	17		5	35	2013	HOMICIDIOS	0				
CHALATENANGO	SAN ANTONIO DE LOS RANCHOS				7	7	2013	HOMICIDIOS	0				
CHALATENANGO	SAN FERNANDO	18	3		4	25	2013	HOMICIDIOS	0				
CHALATENANGO	SAN FRANCISCO LEMPA	2	2		4	8	2013	HOMICIDIOS	0				
CHALATENANGO	SAN FRANCISCO MORAZAN	7	19		1	27	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	SAN IGNACIO	24	26		13	63	2013	HOMICIDIOS	0				
CHALATENANGO	SAN ISIDRO LABRADOR		1		4	5	2013	HOMICIDIOS	0				
CHALATENANGO	SAN JOSE CANCASQUE	7	1	7	7	22	2013	HOMICIDIOS	0				
CHALATENANGO	SAN JOSE LAS FLORES	6		2	7	15	2013	HOMICIDIOS	0				
CHALATENANGO	SAN LUIS DEL CARMEN	1			0	1	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	SAN MIGUEL DE MERCEDES	9	2		3	14	2013	HOMICIDIOS	1	0	1	0	1
CHALATENANGO	SAN RAFAEL	25	9		12	46	2013	HOMICIDIOS	0				
CHALATENANGO	SANTA RITA	39			2	41	2013	HOMICIDIOS	4	2	2	0	4
CHALATENANGO	TEJUTLA	138	48	1	37	224	2013	HOMICIDIOS	11	1	10	0	11
LA LIBERTAD	ANTIGUO CUSCATLAN	32	17	4	9	62	2013	HOMICIDIOS	3	0	3	0	3
LA LIBERTAD	CIUDAD ARCE	444	174	4	53	675	2013	HOMICIDIOS	14	1	13	0	14
LA LIBERTAD	COLON	811	188	11	37	1047	2013	HOMICIDIOS	62	6	49	7	62
LA LIBERTAD	COMASAGUA	99	16	3	3	121	2013	HOMICIDIOS	5	1	4	0	5
LA LIBERTAD	CHILTIUPAN	49	9		10	68	2013	HOMICIDIOS	1	0	1	0	1
LA LIBERTAD	HUIZUCAR	54	5		0	59	2013	HOMICIDIOS	12	0	12	0	12
LA LIBERTAD	JAYAQUE	136	6		3	145	2013	HOMICIDIOS	7	0	7	0	7
LA LIBERTAD	JICALAPA	51	17		1	69	2013	HOMICIDIOS	1	0	1	0	1
LA LIBERTAD	LA LIBERTAD	549	95	9	39	692	2013	HOMICIDIOS	18	2	16	0	18
LA LIBERTAD	NUOVO CUSCATLAN	31			3	34	2013	HOMICIDIOS	0				
LA LIBERTAD	SANTA TECLA	347	214	30	156	747	2013	HOMICIDIOS	17	6	11	0	17
LA LIBERTAD	QUEZALTEPEQUE	345	81	1	42	469	2013	HOMICIDIOS	36	4	32	0	36
LA LIBERTAD	SACACOYO	173	34	8	19	234	2013	HOMICIDIOS	3	0	3	0	3
LA LIBERTAD	SAN JOSE VILLANUEVA	31	12		5	48	2013	HOMICIDIOS	7	1	6	0	7
LA LIBERTAD	SAN JUAN OPICO	513	149	12	78	752	2013	HOMICIDIOS	19	1	18	0	19
LA LIBERTAD	SAN MATIAS	58	11	1	21	91	2013	HOMICIDIOS	5	1	4	0	5

DEPTO	Municipio	cambio domicilio	traslado otr_esc	emigra a_Educame	abando no_Pais	Total_dese rccion	año	Delito	Total_d elitos_1	Mujer	Hom bre	ND	Total_d elitos_2
LA LIBERTAD	SAN PABLO TACACHICO	146	83	2	53	284	2013	HOMICIDIOS	2	0	2	0	2
LA LIBERTAD	TAMANIQUE	102	43	1	33	179	2013	HOMICIDIOS	1	1	0	0	1
LA LIBERTAD	TALNIQUE	28	12	1	1	42	2013	HOMICIDIOS	2	0	2	0	2
LA LIBERTAD	TEOTEPEQUE	63	4		21	88	2013	HOMICIDIOS	2	0	2	0	2
LA LIBERTAD	TEPECOYO	68	20		3	91	2013	HOMICIDIOS	4	0	4	0	4
LA LIBERTAD	ZARAGOZA	164	49		8	221	2013	HOMICIDIOS	12	1	11	0	12
SAN SALVADOR	AGUILARES	162	103	2	66	333	2013	HOMICIDIOS	11	2	9	0	11
SAN SALVADOR	APOPA	989	239	15	76	1319	2013	HOMICIDIOS	72	10	62	0	72
SAN SALVADOR	AYUTUXTEPEQUE	163	61	5	16	245	2013	HOMICIDIOS	13	1	12	0	13
SAN SALVADOR	CUSCATANCINGO	375	57	5	14	451	2013	HOMICIDIOS	27	2	25	0	27
SAN SALVADOR	EL PAISNAL	131	35	5	25	196	2013	HOMICIDIOS	9	0	9	0	9
SAN SALVADOR	GUAZAPA	135	66		6	207	2013	HOMICIDIOS	19	1	18	0	19
SAN SALVADOR	ILOPANGO	818	142	1	46	1007	2013	HOMICIDIOS	46	3	43	0	46
SAN SALVADOR	MEJICANOS	714	194	6	48	962	2013	HOMICIDIOS	91	6	85	0	91
SAN SALVADOR	NEJAPA	93	57		3	153	2013	HOMICIDIOS	13	2	11	0	13
SAN SALVADOR	PANCHIMALCO	185	54	1	4	244	2013	HOMICIDIOS	31	2	29	0	31
SAN SALVADOR	ROSARIO DE MORA	57	22		3	82	2013	HOMICIDIOS	4	0	4	0	4
SAN SALVADOR	SAN MARCOS	456	53	20	36	565	2013	HOMICIDIOS	16	3	13	0	16
SAN SALVADOR	SAN MARTIN	566	198	10	53	827	2013	HOMICIDIOS	25	5	20	0	25
SAN SALVADOR	SAN SALVADOR	2160	785	54	298	3297	2013	HOMICIDIOS	214	18	196	0	214
SAN SALVADOR	SANTIAGO TEXACUANGOS	95	26	1	6	128	2013	HOMICIDIOS	15	2	13	0	15
SAN SALVADOR	SANTO TOMAS	136	51		5	192	2013	HOMICIDIOS	13	1	12	0	13
SAN SALVADOR	SOYAPANGO	1373	377	2	128	1880	2013	HOMICIDIOS	96	9	87	0	96
SAN SALVADOR	TONACATEPEQUE	474	117	9	29	629	2013	HOMICIDIOS	21	5	16	0	21
SAN SALVADOR	CIUDAD DELGADO	485	166	12	30	693	2013	HOMICIDIOS	45	1	43	1	45
CUSCATLÁN	CANDELARIA	98	20		3	121	2013	HOMICIDIOS	0				
CUSCATLÁN	COJUTEPEQUE	295	135	2	59	491	2013	HOMICIDIOS	32	2	30	0	32
CUSCATLÁN	EL CARMEN	93	9		0	102	2013	HOMICIDIOS	8	1	7	0	8
CUSCATLÁN	EL ROSARIO	26	3		2	31	2013	HOMICIDIOS	5	0	5	0	5
CUSCATLÁN	MONTE SAN JUAN	90	14	2	5	111	2013	HOMICIDIOS	2	0	2	0	2
CUSCATLÁN	ORATORIO DE CONCEPCION	36	15	6	0	57	2013	HOMICIDIOS	2	0	2	0	2
CUSCATLÁN	SAN BARTOLOME PERULAPIA	77	12	2	1	92	2013	HOMICIDIOS	5	0	5	0	5
CUSCATLÁN	SAN CRISTOBAL	20	2		3	25	2013	HOMICIDIOS	2	0	2	0	2

DEPTO	Municipio	cambio domic ilio	traslado otr_esc	emigra a Ed ucame	abando no_Pai s	Total_dese rccion	año	Delito	Total d elitos_1	Mujer	Hom bre	ND	Total d elitos_2
CUSCATLÁN	SAN JOSE GUAYABAL	49	16	3	2	70	2013	HOMICIDIOS	4	0	4	0	4
CUSCATLÁN	SAN PEDRO PERULAPAN	220	74		10	304	2013	HOMICIDIOS	35	1	34	0	35
CUSCATLÁN	SAN RAFAEL CEDROS	95	39		49	183	2013	HOMICIDIOS	6	0	6	0	6
CUSCATLÁN	SAN RAMON	44	11		6	61	2013	HOMICIDIOS	3	1	2	0	3
CUSCATLÁN	SANTA CRUZ ANALQUITO	15	2		1	18	2013	HOMICIDIOS	2	0	2	0	2
CUSCATLÁN	SANTA CRUZ MICHAPA	93	19		24	136	2013	HOMICIDIOS	16	0	16	0	16
CUSCATLÁN	SUCHITOTO	140	42		41	223	2013	HOMICIDIOS	18	0	18	0	18
CUSCATLÁN	TENANCINGO	54	13	2	3	72	2013	HOMICIDIOS	5	1	4	0	5
LA PAZ	CUYULTITAN	48	3		2	53	2013	HOMICIDIOS	2	0	2	0	2
LA PAZ	EL ROSARIO	84	13	1	34	132	2013	HOMICIDIOS	12	0	12	0	12
LA PAZ	JERUSALEN	9			0	9	2013	HOMICIDIOS	1	0	1	0	1
LA PAZ	MERCEDES LA CEIBA	1	12		0	13	2013	HOMICIDIOS	0				
LA PAZ	OLOCUILTA	193	35		14	242	2013	HOMICIDIOS	19	2	17	0	19
LA PAZ	PARAISO DE OSORIO	18			1	19	2013	HOMICIDIOS	0				
LA PAZ	SAN ANTONIO MASAHUAT	11	22		7	40	2013	HOMICIDIOS	0				
LA PAZ	SAN EMIGDIO	16	38	4	1	59	2013	HOMICIDIOS	0				
LA PAZ	SAN FRANCISCO CHINAMECA	71	23		5	99	2013	HOMICIDIOS	6	0	6	0	6
LA PAZ	SAN JUAN NONUALCO	154	40		17	211	2013	HOMICIDIOS	0				
LA PAZ	SAN JUAN TALPA	73	44		7	124	2013	HOMICIDIOS	5	0	5	0	5
LA PAZ	SAN JUAN TEPEZONTES	32	6		8	46	2013	HOMICIDIOS	0				
LA PAZ	SAN LUIS TALPA	239	50		30	319	2013	HOMICIDIOS	19	0	19	0	19
LA PAZ	SAN MIGUEL TEPEZONTES	35	3	3	0	41	2013	HOMICIDIOS	0				
LA PAZ	SAN PEDRO MASAHUAT	211	70	2	26	309	2013	HOMICIDIOS	22	1	21	0	22
LA PAZ	SAN PEDRO NONUALCO	55	16	2	0	73	2013	HOMICIDIOS	3	0	3	0	3
LA PAZ	SAN RAFAEL OBRAJUELO	64	22	1	14	101	2013	HOMICIDIOS	8	0	8	0	8
LA PAZ	SANTA MARIA OSTUMA	38	15		1	54	2013	HOMICIDIOS	0				
LA PAZ	SANTIAGO NONUALCO	321	75	4	74	474	2013	HOMICIDIOS	39	0	38	1	39
LA PAZ	TAPALHUACA	34	4	1	4	43	2013	HOMICIDIOS	2	1	1	0	2
LA PAZ	ZACATECOLUCA	416	228	50	99	793	2013	HOMICIDIOS	33	4	27	2	33
LA PAZ	SAN LUIS LA HERRADURA	167	80	2	45	294	2013	HOMICIDIOS	21	0	21	0	21
CABAÑAS	CINQUERA	15	6	3	3	27	2013	HOMICIDIOS	0				
CABAÑAS	GUACOTECTI	35	4	1	3	43	2013	HOMICIDIOS	2	2	0	0	2
CABAÑAS	ILOBASCO	451	84	6	252	793	2013	HOMICIDIOS	52	6	46	0	52

DEPTO	Municipio	cambio _domic _ilio	traslado _otr_esc	emigra _a_Ed _ucame	abando no_Pai _s	Total_dese _rcion	año	Delito	Total_d elitos_ 1	Mujer	Hom bre	ND	Total_d elitos_ 2
CABAÑAS	JUTIAPA	73	2		35	110	2013	HOMICIDIOS	2	0	2	0	2
CABAÑAS	SAN ISIDRO	36	9	2	22	69	2013	HOMICIDIOS	2	0	2	0	2
CABAÑAS	SENSUNTEPEQUE	242	86	55	122	505	2013	HOMICIDIOS	11	0	11	0	11
CABAÑAS	TEJUTEPEQUE	38	22		10	70	2013	HOMICIDIOS	10	1	9	0	10
CABAÑAS	VICTORIA	50	15		36	101	2013	HOMICIDIOS	3	1	2	0	3
CABAÑAS	DOLORES	59	8		33	100	2013	HOMICIDIOS	3	0	0	0	3
SAN VICENTE	APASTEPEQUE	96	24	4	67	191	2013	HOMICIDIOS	4	0	4	0	4
SAN VICENTE	GUADALUPE	54	5	1	4	64	2013	HOMICIDIOS	1	0	1	0	1
SAN VICENTE	SAN CAYETANO ISTEPEQUE	13	10		6	29	2013	HOMICIDIOS	0				
SAN VICENTE	SANTA CLARA	20	6		20	46	2013	HOMICIDIOS	3	0	3	0	3
SAN VICENTE	SANTO DOMINGO	43	8	1	11	63	2013	HOMICIDIOS	1	0	1	0	1
SAN VICENTE	SAN ESTEBAN CATARINA	25	14		16	55	2013	HOMICIDIOS	0				
SAN VICENTE	SAN ILDEFONSO	30	24	4	23	81	2013	HOMICIDIOS	0				
SAN VICENTE	SAN LORENZO	36	7	4	17	64	2013	HOMICIDIOS	2	0	2	0	2
SAN VICENTE	SAN SEBASTIAN	32	11		36	79	2013	HOMICIDIOS	7	0	7	0	7
SAN VICENTE	SAN VICENTE	297	145	27	128	597	2013	HOMICIDIOS	23	2	21	0	23
SAN VICENTE	TECOLUCA	213	68	1	82	364	2013	HOMICIDIOS	33	2	31	0	33
SAN VICENTE	TEPETITAN				2	2	2013	HOMICIDIOS	2	0	2	0	2
SAN VICENTE	VERAPAZ	33	9		4	46	2013	HOMICIDIOS	1	0	1	0	1
USULUTÁN	ALEGRIA	37	11		5	53	2013	HOMICIDIOS	3	0	3	0	3
USULUTÁN	BERLIN	130	45		33	208	2013	HOMICIDIOS	3	0	3	0	3
USULUTÁN	CALIFORNIA	36			0	36	2013	HOMICIDIOS	1	0	1	0	1
USULUTÁN	CONCEPCION BATRES	73	21		29	123	2013	HOMICIDIOS	8	1	7	0	8
USULUTÁN	VILLA EL TRIUNFO	48	20		17	85	2013	HOMICIDIOS	3	0	3	0	3
USULUTÁN	EREGUAYQUIN	49	11	9	9	78	2013	HOMICIDIOS	6	0	6	0	6
USULUTÁN	ESTANZUELAS	71	19		51	141	2013	HOMICIDIOS	4	0	4	0	4
USULUTÁN	JIQUILISCO	329	132	2	197	660	2013	HOMICIDIOS	39	4	35	0	39
USULUTÁN	JUCUAPA	115	49		19	183	2013	HOMICIDIOS	16	2	14	0	16
USULUTÁN	JUCUARAN	116	15		48	179	2013	HOMICIDIOS	9	0	9	0	9
USULUTÁN	MERCEDES UMANA	100	15		27	142	2013	HOMICIDIOS	2	1	1	0	2
USULUTÁN	NUOVA GRANADA	53	3		7	63	2013	HOMICIDIOS	1	1	0	0	1
USULUTÁN	OZATLAN	62	19	3	30	114	2013	HOMICIDIOS	13	0	13	0	13
USULUTÁN	PUERTO EL TRIUNFO	94	35	2	39	170	2013	HOMICIDIOS	9	1	8	0	9

DEPTO	Municipio	cambio _domicilio	traslado _otr_esc	emigra _a_Educame	abando no_Pais	Total_dese rccion	año	Delito	Total_d elitos_1	Mujer	Hom bre	ND	Total_d elitos_2
USulután	SAN AGUSTIN	28	7		4	39	2013	HOMICIDIOS	0				
USulután	SAN BUENAVENTURA	9	2		1	12	2013	HOMICIDIOS	0				
USulután	SAN DIONISIO	42	1		8	51	2013	HOMICIDIOS	3	0	3	0	3
USulután	SANTA ELENA	143	32	3	95	273	2013	HOMICIDIOS	4	0	4	0	4
USulután	SAN FRANCISCO JAVIER	33	4		14	51	2013	HOMICIDIOS	2	0	2	0	2
USulután	SANTA MARIA	39	16	1	19	75	2013	HOMICIDIOS	4	1	3	0	4
USulután	SANTIAGO DE MARIA	139	49	2	31	221	2013	HOMICIDIOS	8	1	7	0	8
USulután	TECAPAN	37	3		10	50	2013	HOMICIDIOS	3	1	2	0	3
USulután	USULUTAN	673	245	27	205	1150	2013	HOMICIDIOS	36	4	32	0	36
SAN MIGUEL	CAROLINA	45	12		20	77	2013	HOMICIDIOS	3	0	3	0	3
SAN MIGUEL	CIUDAD BARRIOS	121	69	5	68	263	2013	HOMICIDIOS	10	1	9	0	10
SAN MIGUEL	COMACARAN	25	4		26	55	2013	HOMICIDIOS	3	0	3	0	3
SAN MIGUEL	CHAPELTIQUE	71	15		44	130	2013	HOMICIDIOS	4	0	4	0	4
SAN MIGUEL	CHINAMECA	79	40	8	10	137	2013	HOMICIDIOS	13	0	13	0	13
SAN MIGUEL	CHIRILAGUA	148	19		56	223	2013	HOMICIDIOS	12	0	12	0	12
SAN MIGUEL	EL TRANSITO	125	90	2	49	266	2013	HOMICIDIOS	3	0	3	0	3
SAN MIGUEL	LOLOTIQUE	52	22		19	93	2013	HOMICIDIOS	3	0	3	0	3
SAN MIGUEL	MONCAGUA	114	81		46	241	2013	HOMICIDIOS	5	1	4	0	5
SAN MIGUEL	NUOVA GUADALUPE	26	8		11	45	2013	HOMICIDIOS	1	0	1	0	1
SAN MIGUEL	NUOVO EDEN DE SAN JUAN	16	4		8	28	2013	HOMICIDIOS	0				
SAN MIGUEL	QUELEPA	20	7		3	30	2013	HOMICIDIOS	1	0	1	0	1
SAN MIGUEL	SAN ANTONIO	17	6		8	31	2013	HOMICIDIOS	1	0	1	0	1
SAN MIGUEL	SAN GERARDO	24	1		16	41	2013	HOMICIDIOS	1	0	1	0	1
SAN MIGUEL	SAN JORGE	11	6		16	33	2013	HOMICIDIOS	2	0	2	0	2
SAN MIGUEL	SAN LUIS DE LA REINA	26	11		28	65	2013	HOMICIDIOS	0				
SAN MIGUEL	SAN MIGUEL	1881	468	11	407	2767	2013	HOMICIDIOS	78	7	71	0	78
SAN MIGUEL	SAN RAFAEL ORIENTE	67	5	2	27	101	2013	HOMICIDIOS	3	0	3	0	3
SAN MIGUEL	SESORI	45	14	2	36	97	2013	HOMICIDIOS	1	0	1	0	1
SAN MIGUEL	ULUAZAPA	22	1	2	11	36	2013	HOMICIDIOS	2	0	2	0	2
MORAZÁN	ARAMBALA	4	7		0	11	2013	HOMICIDIOS	0				
MORAZÁN	CACAOPERA	46	11	1	22	80	2013	HOMICIDIOS	2	1	1	0	2
MORAZÁN	CORINTO	85	55	3	77	220	2013	HOMICIDIOS	10	1	9	0	10
MORAZÁN	CHILANGA	9	13		14	36	2013	HOMICIDIOS	1	0	1	0	1

DEPTO	Municipio	cambio _domicilio	traslado _otr_esc	emigra _a_Educame	abando no_Pais	Total_dese rccion	año	Delito	Total_d elitos_1	Mujer	Hom bre	ND	Total_d elitos_2
MORAZÁN	DELICIAS DE CONCEPCION	13	10		0	23	2013	HOMICIDIOS	0				
MORAZÁN	EL DIVISADERO	31	15	3	18	67	2013	HOMICIDIOS	1	0	1	0	1
MORAZÁN	EL ROSARIO	5	6		1	12	2013	HOMICIDIOS	0				
MORAZÁN	GUALOCOCTI	4	3		1	8	2013	HOMICIDIOS	0				
MORAZÁN	GUATAJAGUA	61	37		17	115	2013	HOMICIDIOS	2	0	2	0	2
MORAZÁN	JOATECA	4	8		14	26	2013	HOMICIDIOS	0				
MORAZÁN	JOEOATIQUE	30	9		6	45	2013	HOMICIDIOS	0				
MORAZÁN	JOCORO	57	46		15	118	2013	HOMICIDIOS	4	0	4	0	4
MORAZÁN	LOLOTIQUILLO	8	8	1	10	27	2013	HOMICIDIOS	2	0	2	0	2
MORAZÁN	MEANGUERA	26	18		33	77	2013	HOMICIDIOS	0				
MORAZÁN	OSICALA	45	19		10	74	2013	HOMICIDIOS	3	0	3	0	3
MORAZÁN	PERQUIN	14	5		3	22	2013	HOMICIDIOS	0				
MORAZÁN	SAN CARLOS	7	27		8	42	2013	HOMICIDIOS	1	0	1	0	1
MORAZÁN	SAN FERNANDO	17	13	2	1	33	2013	HOMICIDIOS	0				
MORAZÁN	SAN FRANCISCO GOTERA	333	43	7	68	451	2013	HOMICIDIOS	6	1	5	0	6
MORAZÁN	SAN ISIDRO	11			1	12	2013	HOMICIDIOS	1	1	0	0	1
MORAZÁN	SAN SIMON	24	17		4	45	2013	HOMICIDIOS	0				
MORAZÁN	SENSEMBRA	20	4		8	32	2013	HOMICIDIOS	1	0	1	0	1
MORAZÁN	SOCIEDAD	72	21	1	32	126	2013	HOMICIDIOS	4	0	4	0	4
MORAZÁN	TOROLA	2	6	1	5	14	2013	HOMICIDIOS	0				
MORAZÁN	YAMABAL	11	17		2	30	2013	HOMICIDIOS	1	0	1	0	1
MORAZÁN	YOLOAIQUIN	10	7		4	21	2013	HOMICIDIOS	2	0	2	0	2
LA UNIÓN	ANAMOROS	55	28	1	40	124	2013	HOMICIDIOS	3	0	3	0	3
LA UNIÓN	BOLIVAR	17	3		10	30	2013	HOMICIDIOS	3	1	2	0	3
LA UNIÓN	CONCEPCION DE ORIENTE	33	10		12	55	2013	HOMICIDIOS	2	0	2	0	2
LA UNIÓN	CONCHAGUA	205	87		106	398	2013	HOMICIDIOS	29	1	28	0	29
LA UNIÓN	EL CARMEN	66	27		47	140	2013	HOMICIDIOS	8	1	7	0	8
LA UNIÓN	EL SAUCE	12	8		11	31	2013	HOMICIDIOS	2	0	2	0	2
LA UNIÓN	INTIPIUCA	34	16		46	96	2013	HOMICIDIOS	3	0	3	0	3
LA UNIÓN	LA UNION	197	108	6	200	511	2013	HOMICIDIOS	32	4	28	0	32
LA UNIÓN	LISIQUE	39	25	1	28	93	2013	HOMICIDIOS	5	2	3	0	5
LA UNIÓN	MEANGUERA DEL GOLFO	9	1		26	36	2013	HOMICIDIOS	1	0	1	0	1
LA UNIÓN	NUOVA ESPARTA	26	31		29	86	2013	HOMICIDIOS	1	0	1	0	1

DEPTO	Municipio	cambio _domic _ilio	traslado _otr_esc	emigra _a_Ed _ucame	abando no_Pai _s	Total_dese _rcion	anho	Delito	Total_d elitos_ _1	Mujer	Hom bre	ND	Total_d elitos_ _2
LA UNIÓN	PASAJUNA	85	46		48	179	2013	HOMICIDIOS	6	0	6	0	6
LA UNIÓN	POLOROS	47	27		39	113	2013	HOMICIDIOS	3	0	3	0	3
LA UNIÓN	SAN ALEJO	80	24	5	49	158	2013	HOMICIDIOS	12	0	12	0	12
LA UNIÓN	SAN JOSE	6	4		3	13	2013	HOMICIDIOS	0				
LA UNIÓN	SANTA ROSA DE LIMA	193	44	2	37	276	2013	HOMICIDIOS	9	0	9	0	9
LA UNIÓN	YAYANTIQUE	47	8		7	62	2013	HOMICIDIOS	1	0	1	0	1
LA UNIÓN	YUCUAIQUIN	30	15		6	51	2013	HOMICIDIOS	0				

Base de Datos Mortalidad_Infantil_RLM.sav

Información sobre las variables

Variable	Etiqueta	Nivel de medida
Obs	Observación	Escala
Mort_Inf	Mortalidad Infantil	Escala
Tasa AlfM	Tasa de alfabetismo de Mujeres-en porcentaje	Escala
PIBP	PIB per cápita	Escala
Tas_FerT	Tasa de fertilidad total	Escala
Tasa_AnalfM	Tasa de analfabetismo de Mujeres-en porcentaje	Escala
Recip_pib	<ninguno>	Escala

Obs	Mort_Inf	Tasa AlfM	PIBP	Tas_FerT	Tasa_AnalfM	Recip_pib
1	128	37	1870	6.66	63	0.000535
2	204	22	130	6.15	78	0.007692
3	202	16	310	7	84	0.003226
4	197	65	570	6.25	35	0.001754
5	96	76	2050	3.81	24	0.000488
6	209	26	200	6.44	74	0.005
7	170	45	670	6.19	55	0.001493
8	240	29	300	5.89	71	0.003333
9	241	11	120	5.89	89	0.008333
10	55	55	290	2.36	45	0.003448
11	75	87	1180	3.93	13	0.000847
12	129	55	900	5.99	45	0.001111
13	24	93	1730	3.5	7	0.000578
14	165	31	1150	7.41	69	0.00087
15	94	77	1160	4.21	23	0.000862
16	96	80	1270	5	20	0.000787
17	148	30	580	5.27	70	0.001724
18	98	69	660	5.21	31	0.001515
19	161	43	420	6.5	57	0.002381
20	118	47	1080	6.12	53	0.000926
21	269	17	290	6.19	83	0.003448
22	189	35	270	5.05	65	0.003704
23	126	58	560	6.16	42	0.001786
24	12	81	4240	1.8	19	0.000236
25	167	29	240	4.75	71	0.004167
26	135	65	460	4.1	35	0.002174
27	107	87	3020	6.66	13	0.000331
28	72	63	1420	7.28	37	0.000704
29	128	49	420	8.12	51	0.002381
30	27	63	19830	5.23	37	0.00005
31	152	84	420	5.79	16	0.002381
32	224	23	530	6.5	77	0.001887

Obs	Mort_Inf	Tasa AlfM	PIBP	Tas_FerT	Tasa_AnalfM	Recip_pib
33	142	50	8640	7.17	50	0.000116
34	104	62	350	6.6	38	0.002857
35	287	31	230	7	69	0.004348
36	41	66	1620	3.91	34	0.000617
37	312	11	190	6.7	89	0.005263
38	77	88	2090	4.2	12	0.000478
39	142	22	900	5.43	78	0.001111
40	262	22	230	6.5	78	0.004348
41	215	12	140	6.25	88	0.007143
42	246	9	330	7.1	91	0.00303
43	191	31	1010	7.1	69	0.00099
44	182	19	300	7	81	0.003333
45	37	88	1730	3.46	12	0.000578
46	103	25	780	5.66	75	0.001282
47	67	85	1300	4.82	15	0.000769
48	143	78	930	5	22	0.001075
49	83	85	690	4.7	15	0.001449
50	223	33	200	8.49	67	0.005
51	240	19	450	6.5	81	0.002222
52	312	21	280	6.5	79	0.003571
53	12	79	4430	1.69	21	0.000226
54	52	83	270	3.25	17	0.003704
55	79	43	1340	7.17	57	0.000746
56	61	88	670	3.52	12	0.001493
57	168	28	410	6.09	72	0.002439
58	28	95	4370	2.86	5	0.000229
59	121	41	1310	4.88	59	0.000763
60	115	62	1470	3.89	38	0.00068
61	186	45	300	6.9	55	0.003333
62	47	85	3630	4.1	15	0.000275
63	178	45	320	6.09	55	0.003125
64	142	67	560	7.2	33	0.001786

Capítulo 3

Base de Datos Elecciones.sav

id	distrito	Partido	NumVotos
Centro	CENTRO	ucd	42243
Centro	CENTRO	psoe	28714
Centro	CENTRO	pce	9863
Centro	CENTRO	cd	13153
Centro	CENTRO	exde	8754
Centro	CENTRO	exi	3945
Centro	CENTRO	vari	1752
Centro	CENTRO	abst	49423
Arganzuela	ARGANZUE	ucd	26569
Arganzuela	ARGANZUE	psoe	20117
Arganzuela	ARGANZUE	pce	6143
Arganzuela	ARGANZUE	cd	7659
Arganzuela	ARGANZUE	exde	4111
Arganzuela	ARGANZUE	exi	2507
Arganzuela	ARGANZUE	vari	893
Arganzuela	ARGANZUE	abst	25430
Retiro	RETIRO	ucd	28351
Retiro	RETIRO	psoe	14921
Retiro	RETIRO	pce	4650
Retiro	RETIRO	cd	11837
Retiro	RETIRO	exde	6408
Retiro	RETIRO	exi	1707
Retiro	RETIRO	vari	898
Retiro	RETIRO	abst	21677
Salamanca	SALAMANC	ucd	45733
Salamanca	SALAMANC	psoe	20567
Salamanca	SALAMANC	pce	5992
Salamanca	SALAMANC	cd	23161
Salamanca	SALAMANC	exde	13032
Salamanca	SALAMANC	exi	2287
Salamanca	SALAMANC	vari	1294
Salamanca	SALAMANC	abst	39187
Chamartin	CHAMARTI	ucd	33518
Chamartin	CHAMARTI	psoe	16535
Chamartin	CHAMARTI	pce	5196
Chamartin	CHAMARTI	cd	15985
Chamartin	CHAMARTI	exde	9106
Chamartin	CHAMARTI	exi	2023
Chamartin	CHAMARTI	vari	853
Chamartin	CHAMARTI	abst	28562
Tetuan	TETUAN	ucd	34663
Tetuan	TETUAN	psoe	29514
Tetuan	TETUAN	pce	10092
Tetuan	TETUAN	cd	10120
Tetuan	TETUAN	exde	5883
Tetuan	TETUAN	exi	3986
Tetuan	TETUAN	vari	1794
Tetuan	TETUAN	abst	33857
Chamberi	CHAMBERI	ucd	47752

id	distrito	Partido	NumVotos
Latina	LATINA	ucd	55208
Latina	LATINA	psoe	50969
Latina	LATINA	pce	18999
Latina	LATINA	cd	11485
Latina	LATINA	exde	4615
Latina	LATINA	exi	7146
Latina	LATINA	vari	1981
Latina	LATINA	abst	54406
Carabanchel	CARABANC	ucd	46959
Carabanchel	CARABANC	psoe	45814
Carabanchel	CARABANC	pce	19609
Carabanchel	CARABANC	cd	8901
Carabanchel	CARABANC	exde	4714
Carabanchel	CARABANC	exi	6773
Carabanchel	CARABANC	vari	2263
Carabanchel	CARABANC	abst	49264
Villaverde	VILLAYER	ucd	28511
Villaverde	VILLAYER	psoe	39656
Villaverde	VILLAYER	pce	21256
Villaverde	VILLAYER	cd	4684
Villaverde	VILLAYER	exde	2585
Villaverde	VILLAYER	exi	6239
Villaverde	VILLAYER	vari	1572
Villaverde	VILLAYER	abst	38446
Mediodia	MEDIODIA	ucd	16831
Mediodia	MEDIODIA	psoe	36408
Mediodia	MEDIODIA	pce	21177
Mediodia	MEDIODIA	cd	1896
Mediodia	MEDIODIA	exde	1047
Mediodia	MEDIODIA	exi	5554
Mediodia	MEDIODIA	vari	1339
Mediodia	MEDIODIA	abst	30845
Vallecas	VALLECAS	ucd	24939
Vallecas	VALLECAS	psoe	44920
Vallecas	VALLECAS	pce	21424
Vallecas	VALLECAS	cd	3030
Vallecas	VALLECAS	exde	1891
Vallecas	VALLECAS	exi	6996
Vallecas	VALLECAS	vari	1629
Vallecas	VALLECAS	abst	42856
Moratalaz	MORATALA	ucd	25330
Moratalaz	MORATALA	psoe	27715
Moratalaz	MORATALA	pce	10433
Moratalaz	MORATALA	cd	4721
Moratalaz	MORATALA	exde	2015
Moratalaz	MORATALA	exi	3847
Moratalaz	MORATALA	vari	1039
Moratalaz	MORATALA	abst	39753
Ciudad Lineal	CIUDAD	ucd	47966

id	distrito	Partido	NumVotos
Chamberí	CHAMBERI	psoe	22896
Chamberí	CHAMBERI	pce	6860
Chamberí	CHAMBERI	cd	22336
Chamberí	CHAMBERI	exde	13960
Chamberí	CHAMBERI	exi	2547
Chamberí	CHAMBERI	vari	1300
Chamberí	CHAMBERI	abst	43886
Fuencarral	FUENCARR	ucd	29483
Fuencarral	FUENCARR	psoe	27039
Fuencarral	FUENCARR	pce	8502
Fuencarral	FUENCARR	cd	6528
Fuencarral	FUENCARR	exde	3397
Fuencarral	FUENCARR	exi	3497
Fuencarral	FUENCARR	vari	1194
Fuencarral	FUENCARR	abst	28556
Moncloa	MONCLOA	ucd	24114
Moncloa	MONCLOA	psoe	14720
Moncloa	MONCLOA	pce	5049
Moncloa	MONCLOA	cd	9566
Moncloa	MONCLOA	exde	5737
Moncloa	MONCLOA	exi	1840
Moncloa	MONCLOA	vari	858
Moncloa	MONCLOA	abst	21645

id	distrito	Partido	NumVotos
Ciudad Lineal	CIUDAD	psoe	40421
Ciudad Lineal	CIUDAD	pce	14606
Ciudad Lineal	CIUDAD	cd	11050
Ciudad Lineal	CIUDAD	exde	5829
Ciudad Lineal	CIUDAD	exi	5022
Ciudad Lineal	CIUDAD	vari	1596
Ciudad Lineal	CIUDAD	abst	47772
San Blas	SANBLAS	ucd	18356
San Blas	SANBLAS	psoe	28881
San Blas	SANBLAS	pce	14143
San Blas	SANBLAS	cd	2791
San Blas	SANBLAS	exde	2020
San Blas	SANBLAS	exi	4321
San Blas	SANBLAS	vari	1284
San Blas	SANBLAS	abst	28045
Hortaleza	HORTALE	ucd	24523
Hortaleza	HORTALE	psoe	26618
Hortaleza	HORTALE	pce	10318
Hortaleza	HORTALE	cd	4659
Hortaleza	HORTALE	exde	2182
Hortaleza	HORTALE	exi	3332
Hortaleza	HORTALE	vari	969
Hortaleza	HORTALE	abst	26485

Base de Datos epf9091_v2.sav

No	Provincia	G.1	G.2	G.3	G.4	G.5	G.6	G.7	G.8	G.9
1	Almería	618957	295452	522128	167067	58288	280035	129219	307967	107334
2	Cádiz	683940	203619	426690	124162	60657	285888	127792	313238	83523
3	Córdoba	590770	265604	487143	113386	37131	237320	116764	247536	79924
4	Granada	547353	238822	458338	119540	40340	236694	103901	272308	72813
5	Huelva	649225	245722	570631	99250	61953	253286	123244	238880	83070
6	Jaén	556210	183295	332662	86364	37160	136992	57607	189811	57311
7	Málaga	617778	201348	508252	121010	63518	256973	128336	323632	93971
8	Sevilla	621570	208156	549399	137408	45101	298000	118269	308524	84514
9	Huesca	577107	249310	412907	107976	39602	335334	90547	227266	92103
10	Teruel	545238	199788	343919	122154	42281	224286	90291	237747	77938
11	Zaragoza	556737	266468	496989	132517	54106	235188	118931	282369	79718
12	Asturias	624941	280273	530828	132066	57679	340013	149265	315478	120856
13	Baleares	564220	226816	602397	144005	86803	358290	150551	351555	131802
14	Las Palmas	632640	201704	522846	153775	84148	327988	173031	305628	114627
15	Tenerife	523476	171072	467424	118857	65247	303598	142620	283563	80959
16	Cantabria	604083	287943	654929	119269	63320	302277	116752	276663	105421
17	Ávila	543595	242609	388063	92808	47035	254563	74522	250853	82061
18	Burgos	602307	255567	600121	162166	51308	280023	132161	301813	111224
19	León	623047	245240	500414	136030	41667	333066	119657	267506	146434

No	Provincia	G.1	G.2	G.3	G.4	G.5	G.6	G.7	G.8	G.9
20	Palencia	589710	206048	449113	113714	34787	248620	115825	294068	109264
21	Salamanca	488784	167814	400049	78217	24476	195065	69846	193056	54442
22	Segovia	528040	184840	455368	103446	46337	217156	91436	259705	116303
23	Soria	679722	232673	503695	129768	55000	272249	117587	300014	120803
24	Valladolid	567361	223201	566433	140573	46111	254216	149041	327774	98430
25	Zamora	544527	178835	402354	99953	32143	227163	70283	231577	125332
26	Albacete	535939	199559	425598	137799	55967	232209	104866	291708	91735
27	Ciudad Real	545912	227255	487651	125740	44001	230820	88650	230213	90886
28	Cuenca	506814	194156	420488	109533	50961	220678	78673	270038	103288
29	Guadalajara	546909	179824	477446	115585	40129	299174	94923	287703	87720
30	Toledo	583325	255527	411896	130747	65345	282127	105872	241749	122189
31	Barcelona	702920	257429	702315	168696	97595	365255	239187	379319	99929
32	Gerona	684186	285047	566149	149308	77553	259839	191400	329089	134786
33	Lérida	696542	283134	508906	146773	90828	402073	180652	353124	152924
34	Tarragona	586122	283112	557653	150464	62853	331848	185713	381485	114876
35	Alicante	579355	205685	490235	134254	68141	297939	117710	316675	111756
36	Castellón	496559	201606	411972	107739	42939	212051	84610	241795	77370
37	Valencia	539570	228072	464127	138419	62471	285948	134751	384939	96564
38	Badajoz	430442	204529	332948	91831	41112	187500	77481	203808	61478
39	Cáceres	569808	222756	403547	119078	47904	248571	100282	285880	89736
40	La Coruña	692445	249121	506616	141805	56114	277401	142246	289111	108489
41	Lugo	719078	286277	414893	142629	62779	301867	101889	216693	119398
42	Orense	598669	182378	370866	106873	31779	221028	114728	205921	90184
43	Pontevedra	736441	263479	468984	136204	50815	344289	129685	309349	100255
44	Madrid	670577	253928	864553	148014	86601	393664	232387	440275	130290
45	Murcia	610718	210169	470859	128627	46866	318508	102978	311262	114457
46	Navarra	669082	324877	704572	221954	81180	415313	185493	411027	156493
47	Alava	664450	234132	631137	189169	58406	313033	164730	355280	98942
48	Guipuzcua	643141	254653	668435	151454	61985	302491	169527	405259	109995
49	Vizcaya	635929	283160	677817	156612	67899	337253	176222	423122	132572
50	La Rioja	634839	209753	542656	127615	54684	269843	126717	322845	121844
51	Ceuta y Melilla	678733	192344	362317	81673	27191	138705	81979	226279	65135

Base de Datos Índice_bienestar.sav

ID	Provincia	Renta	Salud	Sanitarios	educativ oycultural	Cultural	Trabajo	Empleo	EquipHogar	AccesEconom	Convivencia	Seguridad	Entornonatural
1	Alava	5	2	9	6	5	4	4	5	4	4	8	9
2	Albacete	5	7	2	2	8	3	6	2	2	8	7	5
3	Alicante	9	9	10	10	8	9	4	9	7	2	1	6
4	Almería	5	7	1	4	4	7	8	5	6	6	5	5
5	Ávila	4	5	3	4	4	5	6	6	10	5	4	6
6	Badajoz	1	8	4	3	3	4	7	5	4	7	6	4
7	Balears	8	3	7	5	10	8	5	6	6	3	4	10
8	Barcelona	8	3	10	10	5	7	6	9	9	3	5	7
9	Burgos	7	6	7	7	6	7	5	5	2	7	6	4
10	Cáceres	2	6	5	3	6	2	10	4	4	7	7	6
11	Cádiz	1	6	3	5	2	2	1	6	6	4	7	5
12	Castellón	7	4	5	4	5	7	7	6	6	6	4	5
13	Ciudad Real	3	5	4	2	3	3	2	4	7	7	7	3
14	Córdoba	2	5	3	3	2	1	5	5	6	7	7	4
15	Cuenca	4	9	1	1	7	4	5	2	1	10	7	6
16	Girona	10	5	6	8	10	9	5	9	5	4	5	6
17	Granada	2	5	5	6	5	3	7	5	5	4	4	2
18	Guadalajara	4	10	5	6	6	7	5	6	6	7	9	3
19	Guipúzcoa	10	6	9	10	7	10	4	10	7	1	1	10
20	Huelva	2	5	3	2	2	3	4	5	5	6	6	7
21	Huesca	7	4	7	7	10	9	4	7	3	7	6	4
22	Jaén	2	7	3	2	1	1	8	5	5	7	7	7
23	Coruña	3	2	5	7	4	4	4	3	7	4	7	9
24	Palmas	5	5	6	8	6	5	8	5	10	4	6	8
25	León	5	4	6	5	4	5	5	4	4	4	7	6
26	Lleida	10	6	4	6	10	9	7	7	4	5	6	5
27	Rioja	8	6	4	7	6	8	5	6	5	6	6	5
28	Lugo	3	1	5	1	3	6	10	1	3	6	9	7
29	Madrid	9	7	9	9	5	7	10	9	10	4	4	3
30	Málaga	3	5	5	4	4	4	3	5	9	4	4	6
31	Murcia	3	3	3	4	3	7	9	7	9	6	5	3
32	Navarra	10	8	10	10	6	10	6	10	7	4	2	5
33	Ourense	4	2	5	3	2	6	5	1	5	5	8	6
34	Asturias	3	8	2	4	4	6	3	6	6	8	7	3
35	Palencia	6	4	3	6	9	5	3	5	3	8	8	4
36	Pontevedra	3	4	5	6	2	4	4	3	6	4	6	10
37	Salamanca	4	7	7	7	10	3	7	3	4	7	6	5
38	Tenerife	5	6	5	6	5	5	5	4	8	3	5	7
39	Cantabria	6	7	8	9	6	6	3	6	5	5	8	9
40	Zaragoza	7	5	10	7	7	7	3	7	6	6	4	3
41	Segovia	7	8	3	5	8	6	5	6	4	9	7	5

ID	Provincia	Renta	Salud	SSanitarios	educativoycultural	Cultural	Trabajo	Empleo	EquipHogar	AccesEconómico	Convivencia	Seguridad	Entorno natural
42	Sevilla	2	6	4	5	1	3	5	7	6	6	5	3
43	Soria	7	5	7	6	8	8	5	2	2	10	10	6
44	Tarragona	7	5	6	7	8	8	5	7	7	4	3	6
45	Teruel	7	7	6	3	7	7	5	4	1	8	9	1
46	Toledo	4	8	3	2	4	4	3	5	4	8	7	2
47	Valencia	6	3	9	6	4	7	6	7	7	4	4	5
48	Valladolid	6	8	7	9	7	6	4	8	6	6	5	3
49	Vizcaya	9	4	9	9	3	6	5	8	8	3	1	8
50	Zamora	4	6	1	4	5	2	6	4	4	7	8	3

Base de Datos Lavadoras.sav

Carga	precio	capacidad	potencia	anchura	altura	prfun	pesovacio	nprog	tempprelav	tlav
1	1450	4	3140	60	100.5	86	83.9	11	43	85
1	2150	4.5	3300	60	97	118	86.8	11	54	93
2	1390	5	2700	45	128.5	65	89.2	13	35	84
1	1537	4	3200	58.5	107	99.5	81.1	9	47	90
2	1535	5.5	3230	48.5	116	68	88.8	8	52	85
2	1380	4	2400	44	120	63.5	77.8	11	55	85
1	1640	4	2700	58.5	105	99.5	81.2	10	32	96
2	1399	5	2900	48.5	126	64	81.3	8	52	84
1	1119	4.5	2200	61	87	89	88	12	36	92
2	1590	4.5	3000	40.5	94.5	68	86.5	13	80	85
1	1476	4.5	3050	61.5	98	96	105.8	9	44	89
1	1476	5	2290	63.5	97.5	95	82	8	37	92
2	1280	4	2400	44	120	64	77.6	7	36	80
1	1350	5	2950	60	86	91	83.4	12	27	85
1	1090	5	2300	64	107	91	96.3	8	32	86

Base de Datos Medcorp.sav

Cuello	Pecho	Cintura	Caderas	Muñeca	Pantorrilla	Tobillo	Acabeza	Alombr	Acintura	Arrodillas	Calzado	Peso	Genero	Calle	Ojos	Nombre
31	95	72	90	52	32	24	169	141	110	47	39	59	M	6	C	Lorena
32	95	70	96	57	34	22	162	136	103	49	39	62	M	16	O	Elisa
38	102	82	61	53	37	26	192	155	120	57	43	78	H	5	O	Luis
33	96	77	91	51	36	23	158	137	97	45	38	60	M	2	O	Cristina
32	88	70	82	50	33	21	160	140	102	47	40	50	H	5	C	Bernardo
38	102	94	104	56	39	27	175	148	106	51	41	82	H	89	O	Emilio
36	86	80	94	46	32	25	171	140	99	49	42	65	H	89	O	Rubén
36	85	80	96	53	38	24	178	162	102	58	42	67	H	29	O	César
36	90	80	94	52	40	24	181	166	105	60	42	75	H	98	C	Juan Mari
29	85	64	89	52	33	21	169	146	107	52	38	53	M	60	O	Teresa
32	88	68	90	48	33	22	167	144	104	50	37	53	M	153	O	Maika
35	92	72	97	53	35	22	167	143	105	50	39	58	M	123	O	Merche
39	94	92	103	54	36	23	183	151	113	51	44	74	H	50	O	Ricardo
30	88	72	95	56	36	22	162	133	98	49	38	54	M	2	C	Carmen
36	90	86	101	54	36	22	179	150	100	52	42	74	H	32	O	David
39	89	69	90	46	33	23	183	154	104	52	44	74	H	18	O	Alberto
31	89	65	95	50	37	22	170	140	100	50	39	60	M	26	O	Ana
39	94	70	95	55	38	23	180	150	102	51	43	79	H	17	O	Manolo

Capítulo 4

Base de Datos Factorial.sav

Información sobre las variables

Variable	Etiqueta	Nivel de medida
país	<ninguno>	Nominal
v1	tasa de mortalidad cáncer de pulmón	Escala
v2	tasa de mortalidad por cirrosis hepática	Escala
v4	consumo de alcohol	Escala
v5	heridos tráfico	Escala
v6	tasa de mortalidad perinatal	Escala
v7	tasa de mortalidad infantil	Escala
v8	esperanza hombres	Escala
v9	esperanza mujeres	Escala
v10	días cama	Escala
v11	porcentaje de población protegida ambulatorio	Escala
v12	porcentaje de población protegida hospital	Escala
v13	población por cama hospitalaria	Escala
v14	tasa de admisión hospitalaria	Escala
v15	gasto per cápita en médicos	Escala
v16	pib en salud	Escala
v17	pib gasto publico salud	Escala

país	v1	v2	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17
Australia	53.3	11.8	9.4	2006	1.22	0.96	78	70.9	21	84	100	91	21	796	7.5	4.9
Austria	69.5	42.7	10.2	8582	1.13	1.19	76.1	69	20.7	99	99	93	20.3	684	7.3	4.6
Bélgica	112.8	16.8	11.1	8056	1.3	1.12	75.5	69.8	13.9	93	98	106	13.9	636	6.5	6
Canadá	63.2	13	9	9154	1.02	0.85	79	71	14.7	100	100	145	14.7	1058	8.2	6.2
Dinamarca	86	13.5	10.3	2569	0.9	0.77	77.6	71.4	19.2	100	100	130	18.7	736	6.6	5.6
Finlandia	79.5	8.5	6.4	1927	0.74	0.62	77.6	69.2	20.9	100	100	64	21.7	629	6.6	5.2
Francia	60.1	38	15.2	5407	1.19	0.89	78.3	70.1	11.8	99	100	90	12.1	996	9.3	6.6
Alemania	71.9	34.4	11	7965	0.96	1.03	76.5	69.7	18.1	92	95	90	18.1	883	8.2	6.6
Grecia	68.1	16.4	6.8	2881	1.87	1.46	77.8	73.2	11.9	98	98	162	11.9	256	4.7	3.4
Islandia	32.2	0.8	3.2	2586	0.73	0.61	80.5	73.6	20.2	100	100	92	20	832	7.6	6.6
Irlanda	54.7	4.8	6.6	2265	1.36	0.98	75	69.5	17.6	36	100	103	17.6	532	7.2	7.5
Italia	74.6	49.9	12.3	3839	1.52	1.24	77.4	70.7	15.4	100	100	127	15.6	607	7.4	6.2
Japón	30.1	20.3	5.7	5840	0.93	0.62	79.2	73.7	6.7	100	100	84	6.4	673	6.7	5
Luxemburgo	101.1	34.3	18	5758	1.2	1.12	75.1	68	18.1	100	100	76	17.8	719	6.5	7.4
Países bajos	101.4	7.1	8.9	3649	0.98	0.84	79.2	72.5	11.8	76	88	83	11.9	851	8.8	6.9
NZelandia	53.4	5.8	7.8	4967	1.06	1.25	76.4	69.7	15.7	100	100	100	15.6	481	5.7	5.3
noruega	45.2	6.5	3.8	2569	0.96	0.79	79	72.2	14.9	100	100	148	14.6	822	6.9	6.2
Portugal	27.3	47.2	13.5	4176	2.55	1.9	75	67	9.6	100	100	195	9.6	248	5.7	3.9
España	41.1	32.3	12.8	2937	1.57	0.96	78	71.5	9.2	87	86	185	9.2	417	6.3	4.4
Suecia	43	12.6	5	2377	0.73	0.7	78.9	72.6	19.2	100	100	71	18.9	1239	9.6	8.8
suiza	69.8	19.7	11.7	4992	0.91	0.8	79.1	72.4	12.8	97	98	87	12.8	990	7.8	4.7
Gran Bretaña	107.2	4.9	7.1	5934	1.05	1.02	75.9	70.2	12.7	100	100	124	12.7	539	6.2	5.5
USA	68.4	17.9	10.5	8268	1.26	1.09	76.7	69.6	17	25	40	169	17	1388	10.8	4.5

Base de Datos Factorial2.sav

encuesta	precio	financia cion	consu mo	gasolin a	segurida d	confor	capacidad	presta ciones	juvenil	aerodina mica
1	4	1	4	3	3	2	4	4	4	4
2	5	5	4	4	3	3	4	1	1	3
3	2	1	3	1	4	2	1	5	4	5
4	1	1	1	1	4	4	2	5	5	4
5	1	1	2	1	5	5	4	3	3	2
6	5	5	5	5	3	3	4	2	2	1
7	4	5	4	4	2	2	5	1	1	1
8	3	2	3	1	4	4	2	5	5	5
9	4	4	4	3	4	4	3	1	1	1
10	5	5	5	5	2	2	3	2	2	2
11	2	2	2	1	5	4	4	3	4	3
12	4	4	5	5	4	5	5	2	1	2
13	3	2	2	1	4	5	4	4	3	3
14	5	5	4	4	5	4	4	1	2	2
15	4	3	3	1	4	4	5	3	4	4
16	5	5	4	4	4	5	4	2	1	1
17	4	4	5	2	4	5	5	4	4	2
18	5	5	4	4	2	2	1	2	2	3
19	3	3	2	2	4	4	5	4	5	4
20	5	5	4	4	4	5	4	3	2	1

Base de Datos Factorial3.sav

V1	V2	V3	V4	V5	V6	V7
4.1	0.6	6.9	4.7	2.4	2.3	5.2
1.8	3	6.3	6.6	2.5	4	8.4
3.4	5.2	5.7	6	4.3	2.7	8.2
2.7	1	7.1	5.9	1.8	2.3	7.8
6	0.9	9.6	7.8	3.4	4.6	4.5
1.9	3.3	7.9	4.8	2.6	1.9	9.7
4.6	2.4	9.5	6.6	3.5	4.5	7.6
1.3	4.2	6.2	5.1	2.8	2.2	6.9
5.5	1.6	9.4	4.7	3.5	3	7.6
4	3.5	6.5	6	3.7	3.2	8.7
2.4	1.6	8.8	4.8	2	2.8	5.8
3.9	2.2	9.1	4.6	3	2.5	8.3
2.8	1.4	8.1	3.8	2.1	1.4	6.6
3.7	1.5	8.6	5.7	2.7	3.7	6.7
4.7	1.3	9.9	6.7	3	2.6	6.8
3.4	2	9.7	4.7	2.7	1.7	4.8
3.2	4.1	5.7	5.1	3.6	2.9	6.2
4.9	1.8	7.7	4.3	3.4	1.5	5.9
5.3	1.4	9.7	6.1	3.3	3.9	6.8
4.7	1.3	9.9	6.7	3	2.6	6.8
3.3	0.9	8.6	4	2.1	1.8	6.3
3.4	0.4	8.3	2.5	1.2	1.7	5.2
3	4	9.1	7.1	3.5	3.4	8.4
2.4	1.5	6.7	4.8	1.9	2.5	7.2
5.1	1.4	8.7	4.8	3.3	2.6	3.8
4.6	2.1	7.9	5.8	3.4	2.8	4.7
2.4	1.5	6.6	4.8	1.9	2.5	7.2
5.2	1.3	9.7	6.1	3.2	3.9	6.7
3.5	2.8	9.9	3.5	3.1	1.7	5.4
4.1	3.7	5.9	5.5	3.9	3	8.4
3	3.2	6	5.3	3.1	3	8
2.8	3.8	8.9	6.9	3.3	3.2	8.2
5.2	2	9.3	5.9	3.7	2.4	4.6
3.4	3.7	6.4	5.7	3.5	3.4	8.4
2.4	1	7.7	3.4	1.7	1.1	6.2
1.8	3.3	7.5	4.5	2.5	2.4	7.6
3.6	4	5.8	5.8	3.7	2.5	9.3
4	0.9	9.1	5.4	2.4	2.6	7.3
0	2.1	6.9	5.4	1.1	2.6	8.9
2.4	2	6.4	4.5	2.1	2.2	8.8
1.9	3.4	7.6	4.6	2.6	2.5	7.7
5.9	0.9	9.6	7.8	3.4	4.6	4.5
4.9	2.3	9.3	4.5	3.6	1.3	6.2
5	1.3	8.6	4.7	3.1	2.5	3.7
2	2.6	6.5	3.7	2.4	1.7	8.5
5	2.5	9.4	4.6	3.7	1.4	6.3
3.1	1.9	10	4.5	2.6	3.2	3.8
3.4	3.9	5.6	5.6	3.6	2.3	9.1
5.8	0.2	8.8	4.5	3	2.4	6.7
5.4	2.1	8	3	3.8	1.4	5.2

V1	V2	V3	V4	V5	V6	V7
3.7	0.7	8.2	6	2.1	2.5	5.2
2.6	4.8	8.2	5	3.6	2.5	9
4.5	4.1	6.3	5.9	4.3	3.4	8.8
2.8	2.4	6.7	4.9	2.5	2.6	9.2
3.8	0.8	8.7	2.9	1.6	2.1	5.6
2.9	2.6	7.7	7	2.8	3.6	7.7
4.9	4.4	7.4	6.9	4.6	4	9.6
5.4	2.5	9.6	5.5	4	3	7.7
4.3	1.8	7.6	5.4	3.1	2.5	4.4
2.3	4.5	8	4.7	3.3	2.2	8.7
3.1	1.9	9.9	4.5	2.6	3.1	3.8
5.1	1.9	9.2	5.8	3.6	2.3	4.5
4.4	1.1	9.3	5.5	2.5	2.7	7.4
3	3.8	5.5	4.9	3.4	2.6	6
1.1	2	7.2	4.7	1.6	3.2	10
3.7	1.4	9	4.5	2.6	2.3	6.8
4.2	2.5	9.2	6.2	3.3	3.9	7.3
1.6	4.5	6.4	5.3	3	2.5	7.1
5.3	1.7	8.5	3.7	3.5	1.9	4.8
2.3	3.7	8.3	5.2	3	2.3	9.1
3.6	5.4	5.9	6.2	4.5	2.9	8.4
5.6	2.2	8.2	3.1	4	1.6	5.3
3.6	2.2	9.9	4.8	2.9	1.9	4.9
5.2	1.3	9.1	4.5	3.3	2.7	7.3
3	2	6.6	6.6	2.4	2.7	8.2
4.2	2.4	9.4	4.9	3.2	2.7	8.5
3.8	0.8	8.3	6.1	2.2	2.6	5.3
3.3	2.6	9.7	3.3	2.9	1.5	5.2
1	1.9	7.1	4.5	1.5	3.1	9.9
4.5	1.6	8.7	4.6	3.1	2.1	6.8
5.5	1.8	8.7	3.8	3.6	2.1	4.9
3.4	4.6	5.5	8.2	4	4.4	6.3
1.6	2.8	6.1	6.4	2.3	3.8	8.2
2.3	3.7	7.6	5	3	2.5	7.4
2.6	3	8.5	6	2.8	2.8	6.8
2.5	3.1	7	4.2	2.8	2.2	9
2.4	2.9	8.4	5.9	2.7	2.7	6.7
2.1	3.5	7.4	4.8	2.8	2.3	7.2
2.9	1.2	7.3	6.1	2	2.5	8
4.3	2.5	9.3	6.3	3.4	4	7.4
3	2.8	7.8	7.1	3	3.8	7.9
4.8	1.7	7.6	4.2	3.3	1.4	5.8
3.1	4.2	5.1	7.8	3.6	4	5.9
1.9	2.7	5	4.9	2.2	2.5	8.2
4	0.5	6.7	4.5	2.2	2.1	5
0.6	1.6	6.4	5	0.7	2.1	8.4
6.1	0.5	9.2	4.8	3.3	2.8	7.1
2	2.8	5.2	5	2.4	2.7	8.4
3.1	2.2	6.7	6.8	2.6	2.9	8.4
2.5	1.8	9	5	2.2	3	6

Base de Datos Quality_serv_Intx.sav

Información sobre las variables

Variable	Etiqueta	Nivel de medida
A1	Trato recibido del personal	Nominal
A2	Capacidad del personal para resolver problemas	Nominal
A3	Aspecto del personal	Nominal
A4	Profesionalidad/eficacia del personal	Nominal
A5	Aspecto/apariencia interior del punto de venta	Nominal
A6	Aspecto exterior del punto de venta	Nominal
A7	Claridad de la información	Nominal
A8	Novedad/actualidad de la información recibida	Nominal
A9	Calidad de la información	Nominal
A10	Interés de los folletos informativos en el punto de venta	Nominal
A11	Tiempo de espera para ser atendido	Nominal
A12	Tiempo de espera para resolver una consulta	Nominal
A13	Información sobre ofertas/ promociones especiales	Nominal
A14	Precio del hardware	Nominal
A15	Precio del software	Nominal
A16	Precio de los consumibles	Nominal
A17	Descuentos/bonos ofrecidos	Nominal
A18	Rapidez reparación de averías del servicio postventa	Nominal
A19	Garantía del servicio postventa	Nominal
A20	Asesoramiento recibido sobre nuevo hardware	Nominal
A21	Asesoramiento recibido sobre nuevo software	Nominal
A22	Asesoramiento recibido sobre formación y cursos	Nominal

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22
1	4	4	4	9	2	7	5	5	3	3	5	6	1	1	1	9	5	4	8	10	4
2	5	5	5	8	4	6	8	6	4	4	3	7	2	2	4	0	7	5	9	6	5
3	6	6	4	9	2	8	6	5	5	5	6	8	3	3	2	6	8	6	6	8	6
4	4	5	5	8	5	9	7	6	6	4	4	9	4	4	4	8	9	5	7	7	9
5	5	6	4	7	2	8	5	5	4	5	5	7	5	3	3	7	7	7	8	9	7
4	7	5	7	6	4	7	6	4	5	6	6	6	6	4	5	9	8	6	5	4	8
5	2	7	8	8	2	6	4	5	6	4	7	8	7	5	3	8	6	7	6	5	6
6	6	6	7	7	4	7	6	6	7	5	3	7	8	4	6	7	7	8	8	6	7
5	6	7	6	8	2	8	8	7	6	6	4	6	9	5	4	8	5	9	5	5	5
4	4	8	7	6	3	9	7	6	5	7	5	5	9	4	5	9	6	7	6	8	6

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22
5	5	9	8	7	5	8	6	5	6	3	4	4	8	5	4	6	7	6	7	7	5
6	7	8	7	8	1	7	8	4	7	4	6	5	7	6	5	5	6	7	3	9	4
7	3	7	9	7	5	8	7	5	8	5	4	6	6	5	2	4	7	8	4	6	6
6	4	8	4	8	2	9	6	6	9	6	6	7	5	4	5	3	8	6	5	7	4
5	2	6	5	9	5	8	5	7	8	4	5	8	4	5	2	2	7	5	4	10	7
6	6	7	4	8	3	7	6	6	7	5	6	6	3	6	4	1	6	4	6	4	3
7	7	5	5	9	4	6	7	5	6	6	7	7	2	7	3	2	7	3	5	6	7
6	8	4	4	8	2	7	4	6	5	5	5	5	1	6	5	3	8	4	6	5	8
5	6	5	2	7	5	8	5	5	6	6	4	6	2	5	2	4	7	5	7	7	4
6	7	6	4	6	2	9	6	6	7	7	5	3	3	7	5	5	6	6	5	8	7
7	8	7	3	7	4	8	7	8	8	8	6	4	4	4	2	6	7	5	6	9	5
6	5	5	4	6	2	7	6	7	9	7	7	3	5	5	6	7	6	6	7	10	6
5	6	6	7	7	5	6	7	8	4	6	5	4	6	7	1	8	5	5	8	3	8
4	7	8	6	8	4	7	6	7	5	5	6	5	7	3	5	9	6	6	9	5	4
3	5	7	8	9	5	8	7	4	6	4	7	4	8	5	2	8	5	5	5	7	5
4	6	9	9	8	4	9	4	5	7	5	8	5	9	3	4	7	6	7	6	6	6
5	7	8	7	7	6	8	5	4	8	6	6	6	0	6	1	6	7	6	7	8	7
6	4	3	8	8	4	7	4	3	9	7	5	5	9	7	3	5	8	7	6	7	8
7	5	5	5	9	2	6	5	4	8	8	5	6	8	4	2	4	6	8	7	10	9
8	6	4	7	7	7	7	4	3	7	9	6	7	7	5	3	3	7	9	8	8	3
9	4	6	6	8	2	8	9	3	6	7	7	6	6	8	1	3	8	8	6	7	2
8	5	5	7	9	8	9	8	6	7	6	8	7	5	5	4	4	7	7	7	8	5
7	7	6	4	5	2	0	6	7	8	5	9	8	4	6	2	5	6	6	5	9	4
8	8	7	7	6	3	9	7	3	9	5	0	9	3	4	5	6	7	7	5	6	3
9	9	8	5	5	4	8	5	4	7	4	7	2	2	5	3	7	8	4	6	7	5
8	10	6	6	7	3	9	6	3	6	3	8	3	3	9	4	8	7	3	7	8	4
7	2	7	8	6	2	8	5	4	5	4	7	2	4	3	3	9	6	5	8	10	6
8	3	8	9	8	3	9	6	5	4	5	6	3	5	4	4	9	8	4	6	8	5
9	4	6	7	6	2	7	7	6	3	6	8	4	6	6	3	8	9	6	7	9	7
7	5	7	8	9	3	8	8	4	5	3	9	3	7	5	4	7	7	5	8	7	8
8	4	8	6	7	4	7	9	5	6	8	7	4	8	6	2	6	5	7	9	8	9
9	6	7	7	9	4	6	8	6	7	6	6	3	9	8	4	5	6	8	10	6	5
7	5	9	6	6	4	7	7	5	8	7	5	4	0	9	2	4	7	4	9	7	6
8	7	8	5	8	3	8	6	4	9	5	6	5	9	5	4	3	6	1	3	5	8
7	5	7	6	6	2	6	5	5	7	6	7	6	8	4	3	2	8	4	4	8	5
6	6	4	7	8	2	7	4	7	8	8	8	7	7	3	5	3	7	3	5	7	6
5	8	5	6	9	3	8	6	6	9	6	5	5	6	4	4	4	8	5	4	8	7
6	7	4	7	7	4	4	5	5	6	4	4	6	5	5	2	5	9	4	5	9	3
6	8	6	8	5	3	5	7	8	5	3	5	7	4	6	3	6	7	5	6	6	4
6	5	5	4	6	2	4	6	4	6	4	4	6	3	7	4	7	9	5	5	7	6
5	6	6	8	8	5	6	7	4	7	5	5	5	2	6	5	8	6	7	7	8	5
5	7	7	9	7	7	3	4	5	8	4	6	6	1	5	6	9	7	4	6	6	7

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22
9	6	8	6	4	4	4	6	4	6	5	7	8	3	4	7	7	5	8	4	7	6
7	7	7	6	3	3	5	3	5	9	4	8	9	4	5	2	4	4	4	5	8	8
8	8	8	7	5	2	4	4	6	4	5	9	0	5	6	6	6	6	7	3	9	8
9	7	9	8	5	2	5	3	5	7	5	8	4	1	5	7	7	3	5	4	6	7
5	8	6	7	4	3	3	4	6	6	4	9	5	1	6	4	5	4	6	6	7	9

Capítulo 5

Base de Datos Correspondencias.sav

Información sobre las variables

Variable	Etiqueta	Nivel de medida
encuesta	Número de Cuestionario	Escala
preocupaciones	Preocupaciones: aspecto que más le preocupa	Ordinal
estadocivil	Estado civil	Ordinal

encuesta	preocupaciones	estado civil
1	El dinero	Casado
2	El dinero	Casado
3	Armonía familiar	Soltero
4	Armonía familiar	Otro
5	Su salud	Casado
6	Su salud	Otro
7	Su salud	Casado
8	El dinero	Casado
9	El dinero	Otro
10	Su salud	Casado
11	Armonía familiar	Casado
12	Su salud	Casado
13	El dinero	Soltero
14	Su salud	Otro
15	El dinero	Casado
16	Su vida afectiva	Otro
17	Armonía familiar	Casado
18	Su salud	Soltero
19	Su salud	Casado
20	Su vida afectiva	Casado
21	El dinero	Casado
22	Su salud	Casado
23	Su salud	Otro
24	El dinero	Casado
25	Su salud	Casado
26	Armonía familiar	Soltero
27	Armonía familiar	Soltero
28	Armonía familiar	Casado
29	Su vida afectiva	Soltero
30	Su salud	Casado
31	Su salud	Otro
32	Armonía familiar	Soltero
33	Su salud	Casado
34	Su salud	Otro
35	El dinero	Casado
36	El dinero	Casado
37	El dinero	Soltero
38	Su salud	Casado
39	Su salud	Casado
40	El dinero	Casado

encuesta	preocupaciones	estado civil
401	Armonía familiar	Casado
402	Su salud	Otro
403	Armonía familiar	Casado
404	Su salud	Otro
405	Armonía familiar	Soltero
406	Su salud	Casado
407	Su vida afectiva	Casado
408	Su salud	Otro
409	Armonía familiar	Casado
410	El dinero	Soltero
411	El dinero	Soltero
412	Su salud	Soltero
413	Armonía familiar	Soltero
414	El dinero	Casado
415	Su salud	Casado
416	Su salud	Casado
417	Su salud	Casado
418	Su salud	Casado
419	Armonía familiar	Casado
420	Su salud	Casado
421	Su salud	Casado
422	El dinero	Soltero
423	El dinero	Casado
424	Armonía familiar	Casado
425	El dinero	Casado
426	Su salud	Casado
427	Su salud	Soltero
428	Armonía familiar	Soltero
429	El dinero	Casado
430	Armonía familiar	Casado
431	Su salud	Otro
432	Armonía familiar	Casado
433	Su salud	Casado
434	Armonía familiar	Casado
435	El dinero	Casado
436	Armonía familiar	Casado
437	Su salud	Casado
438	El dinero	Soltero
439	Armonía familiar	Casado
440	Armonía familiar	Soltero

encuesta	preocupaciones	estado civil
801	El dinero	Casado
802	El dinero	Casado
803	El dinero	Casado
804	Su salud	Casado
805	Armonía familiar	Otro
806	Su salud	Soltero
807	El dinero	Soltero
808	Su salud	Casado
809	Armonía familiar	Casado
810	El dinero	Casado
811	El dinero	Casado
812	Su salud	Casado
813	Su salud	Soltero
814	Su salud	Soltero
815	Su salud	Soltero
816	El dinero	Casado
817	El dinero	Casado
818	El dinero	Casado
819	Su salud	Soltero
820	Su salud	Soltero
821	Su vida afectiva	Casado
822	Su salud	Casado
823	El dinero	Soltero
824	Su salud	Casado
825	Su salud	Otro
826	Su salud	Casado
827	Armonía familiar	Casado
828	Armonía familiar	Soltero
829	Su salud	Casado
830	Su salud	Casado
831	Armonía familiar	Casado
832	El dinero	Casado
833	Su salud	Casado
834	El dinero	Casado
835	El dinero	Casado
836	El dinero	Casado
837	NS/NC	Casado
838	Su vida afectiva	Casado
839	El dinero	Soltero
840	Su salud	Casado

encuesta	preocupaciones	estado civil
41	Su salud	Casado
42	Armonía familiar	Casado
43	El dinero	Casado
44	El dinero	Soltero
45	Su salud	Otro
46	Su salud	Casado
47	Su vida afectiva	Casado
48	NS/NC	Soltero
49	Su salud	Casado
50	Su vida afectiva	Soltero
51	Su salud	Casado
52	NS/NC	Casado
53	El dinero	Casado
54	Armonía familiar	Otro
55	Su salud	Otro
56	Armonía familiar	Casado
57	Armonía familiar	Casado
58	Su salud	Otro
59	NS/NC	Soltero
60	Armonía familiar	Soltero
61	NS/NC	Soltero
62	Armonía familiar	Casado
63	Su salud	Soltero
64	Su salud	Casado
65	Su vida afectiva	Soltero
66	El dinero	Casado
67	Su salud	Soltero
68	El dinero	Casado
69	Armonía familiar	Casado
70	El dinero	Soltero
71	NS/NC	Casado
72	El dinero	Casado
73	Su salud	Soltero
74	Armonía familiar	Casado
75	El dinero	Soltero
76	Su salud	Casado
77	Armonía familiar	Casado
78	Su salud	Otro
79	El dinero	Casado
80	Su salud	Casado
81	Su vida afectiva	Casado
82	Armonía familiar	Casado
83	Armonía familiar	Soltero
84	El dinero	Soltero
85	Armonía familiar	Soltero
86	Su salud	Casado
87	Su salud	Soltero
88	El dinero	Casado
89	El dinero	Soltero
90	Su salud	Casado
91	NS/NC	Casado
92	El dinero	Soltero

encuesta	preocupaciones	estado civil
441	El dinero	Soltero
442	Armonía familiar	Casado
443	Su salud	Casado
444	Su salud	Casado
445	Armonía familiar	Casado
446	Armonía familiar	Casado
447	Su salud	Soltero
448	Su vida afectiva	Soltero
449	El dinero	Soltero
450	Su salud	Casado
451	Su salud	Casado
452	NS/NC	Soltero
453	Armonía familiar	Casado
454	El dinero	Casado
455	Su salud	Casado
456	Su salud	Soltero
457	El dinero	Soltero
458	Su salud	Casado
459	Armonía familiar	Casado
460	Su salud	Soltero
461	Su salud	Otro
462	Su salud	Soltero
463	Su salud	Casado
464	El dinero	Soltero
465	Su salud	Casado
466	El dinero	Otro
467	El dinero	Casado
468	Su salud	Casado
469	Su salud	Soltero
470	Armonía familiar	Casado
471	El dinero	Casado
472	NS/NC	Soltero
473	Su vida afectiva	Soltero
474	Su salud	Soltero
475	Su vida afectiva	Casado
476	Su salud	Casado
477	Su salud	Casado
478	Su salud	Casado
479	Armonía familiar	Casado
480	El dinero	Casado
481	El dinero	Soltero
482	Su salud	Casado
483	Su salud	Casado
484	Armonía familiar	Soltero
485	El dinero	Casado
486	Su salud	Otro
487	Su salud	Soltero
488	Armonía familiar	Casado
489	Su salud	Casado
490	Armonía familiar	Casado
491	Su salud	Soltero
492	Armonía familiar	Otro

encuesta	preocupaciones	estado civil
841	El dinero	Casado
842	Su salud	Casado
843	Su salud	Otro
844	El dinero	Casado
845	Su vida afectiva	Casado
846	El dinero	Casado
847	El dinero	Soltero
848	El dinero	Soltero
849	Su salud	Casado
850	El dinero	Soltero
851	El dinero	Casado
852	Su vida afectiva	Soltero
853	Su salud	Casado
854	El dinero	Casado
855	El dinero	Otro
856	NS/NC	Otro
857	Su salud	Otro
858	El dinero	Soltero
859	El dinero	Casado
860	El dinero	Casado
861	El dinero	Casado
862	Su salud	Casado
863	Su salud	Otro
864	Armonía familiar	Otro
865	Armonía familiar	Casado
866	Su salud	Soltero
867	Armonía familiar	Casado
868	Su salud	Otro
869	El dinero	Casado
870	Armonía familiar	Casado
871	Armonía familiar	Casado
872	Su salud	Casado
873	Su salud	Casado
874	Su salud	Otro
875	Su salud	Casado
876	Su salud	Casado
877	El dinero	Soltero
878	Su salud	Casado
879	Armonía familiar	Casado
880	Su salud	Otro
881	Su salud	Casado
882	Armonía familiar	Soltero
883	Su salud	Soltero
884	Su salud	Soltero
885	Su salud	Casado
886	Su salud	Casado
887	NS/NC	Casado
888	Armonía familiar	Soltero
889	Su salud	Casado
890	Su salud	Casado
891	Su salud	Casado
892	Armonía familiar	Casado

encuesta	preocupaciones	estado civil
93	Armonía familiar	Casado
94	Armonía familiar	Casado
95	Armonía familiar	Otro
96	Armonía familiar	Casado
97	Armonía familiar	Casado
98	Su salud	Casado
99	Armonía familiar	Soltero
100	Armonía familiar	Casado
101	Su salud	Casado
102	Su salud	Casado
103	Su vida afectiva	Casado
104	Su salud	Soltero
105	Su salud	Otro
106	NS/NC	Casado
107	El dinero	Otro
108	NS/NC	Casado
109	NS/NC	Casado
110	El dinero	Casado
111	El dinero	Soltero
112	Armonía familiar	Casado
113	El dinero	Casado
114	Su salud	Soltero
115	El dinero	Casado
116	Su salud	Casado
117	Su salud	Otro
118	Su salud	Casado
119	Armonía familiar	Casado
120	Armonía familiar	Casado
121	El dinero	Casado
122	Su vida afectiva	Casado
123	Su salud	Casado
124	El dinero	Soltero
125	El dinero	Soltero
126	Su salud	Casado
127	El dinero	Casado
128	Su salud	Casado
129	Armonía familiar	Casado
130	El dinero	Casado
131	Armonía familiar	Casado
132	Su salud	Casado
133	Armonía familiar	Casado
134	Su salud	Casado
135	Su vida afectiva	Soltero
136	Su salud	Casado
137	Su salud	Casado
138	Armonía familiar	Soltero
139	Su salud	Casado
140	Su salud	Casado
141	Su salud	Otro
142	Su salud	Casado
143	El dinero	Casado
144	El dinero	Soltero

encuesta	preocupaciones	estado civil
493	Su salud	Casado
494	Su vida afectiva	Casado
495	El dinero	Casado
496	Su vida afectiva	Otro
497	Armonía familiar	Casado
498	Armonía familiar	Casado
499	Armonía familiar	Casado
500	Su salud	Otro
501	El dinero	Casado
502	NS/NC	Casado
503	Armonía familiar	Casado
504	Su salud	Otro
505	El dinero	Soltero
506	Armonía familiar	Casado
507	Su salud	Casado
508	Su salud	Casado
509	El dinero	Otro
510	Armonía familiar	Casado
511	Su salud	Soltero
512	Su salud	Casado
513	Su salud	Casado
514	NS/NC	Soltero
515	Armonía familiar	Otro
516	Armonía familiar	Soltero
517	Armonía familiar	Soltero
518	El dinero	Casado
519	El dinero	Soltero
520	Armonía familiar	Casado
521	Su salud	Otro
522	El dinero	Otro
523	NS/NC	Casado
524	Su salud	Casado
525	Su salud	Soltero
526	Armonía familiar	Soltero
527	El dinero	Casado
528	NS/NC	Casado
529	Su salud	Soltero
530	Su salud	Soltero
531	Armonía familiar	Soltero
532	Armonía familiar	Soltero
533	El dinero	Soltero
534	El dinero	Casado
535	Armonía familiar	Soltero
536	Su salud	Casado
537	Su salud	Otro
538	Su vida afectiva	Soltero
539	Su salud	Casado
540	Su salud	Soltero
541	Su salud	Casado
542	El dinero	Casado
543	Su salud	Casado
544	Su salud	Casado

encuesta	preocupaciones	estado civil
893	Su salud	Soltero
894	Su salud	Casado
895	Armonía familiar	Casado
896	El dinero	Casado
897	El dinero	Soltero
898	Su salud	Casado
899	Su salud	Casado
900	Armonía familiar	Casado
901	Su salud	Casado
902	Armonía familiar	Casado
903	El dinero	Casado
904	Su vida afectiva	Casado
905	El dinero	Casado
906	Su salud	Casado
907	El dinero	Soltero
908	El dinero	Soltero
909	Su salud	Casado
910	Su salud	Casado
911	Su salud	Otro
912	El dinero	Soltero
913	Armonía familiar	Casado
914	Armonía familiar	Otro
915	Armonía familiar	Casado
916	Armonía familiar	Casado
917	El dinero	Casado
918	Su salud	Casado
919	El dinero	Soltero
920	Su salud	Otro
921	Armonía familiar	Casado
922	Su salud	Casado
923	Su salud	Otro
924	Armonía familiar	Casado
925	El dinero	Soltero
926	Su salud	Casado
927	NS/NC	Casado
928	El dinero	Casado
929	Su salud	Soltero
930	Su salud	Soltero
931	Armonía familiar	Otro
932	Armonía familiar	Soltero
933	Armonía familiar	Casado
934	Su vida afectiva	Casado
935	Su salud	Soltero
936	El dinero	Casado
937	Su salud	Soltero
938	Su salud	Soltero
939	Armonía familiar	Casado
940	El dinero	Casado
941	Su salud	Casado
942	El dinero	Casado
943	Su salud	Casado
944	NS/NC	Soltero

encuesta	preocupaciones	estado civil
145	Su salud	Soltero
146	Armonía familiar	Casado
147	NS/NC	Otro
148	El dinero	Casado
149	Armonía familiar	Casado
150	Su salud	Casado
151	Su salud	Soltero
152	Su salud	Casado
153	Su salud	Casado
154	Su salud	Soltero
155	El dinero	Soltero
156	Su salud	Casado
157	Su salud	Soltero
158	Su salud	Casado
159	El dinero	Casado
160	Su salud	Casado
161	Su salud	Casado
162	El dinero	Casado
163	El dinero	Soltero
164	El dinero	Soltero
165	Su salud	Casado
166	Su salud	Soltero
167	Su salud	Soltero
168	Su vida afectiva	Otro
169	El dinero	Soltero
170	Su salud	Casado
171	El dinero	Casado
172	Su salud	Casado
173	Su vida afectiva	Casado
174	Armonía familiar	Casado
175	Su salud	Otro
176	El dinero	Otro
177	Su vida afectiva	Otro
178	Su salud	Casado
179	Su salud	Casado
180	Su salud	Casado
181	Su salud	Casado
182	El dinero	Casado
183	Armonía familiar	Casado
184	El dinero	Soltero
185	Su salud	Casado
186	El dinero	Soltero
187	Su salud	Casado
188	Armonía familiar	Casado
189	Su salud	Soltero
190	El dinero	Soltero
191	Su salud	Soltero
192	Armonía familiar	Casado
193	Su salud	Soltero
194	Armonía familiar	Casado
195	Armonía familiar	Casado
196	Armonía familiar	Casado

encuesta	preocupaciones	estado civil
545	Su salud	Casado
546	Su salud	Casado
547	Armonía familiar	Casado
548	Armonía familiar	Soltero
549	Su salud	Casado
550	Armonía familiar	Soltero
551	El dinero	Casado
552	Armonía familiar	Casado
553	Su salud	Casado
554	Armonía familiar	Soltero
555	El dinero	Casado
556	El dinero	Casado
557	Su salud	Casado
558	El dinero	Casado
559	Armonía familiar	Soltero
560	Su salud	Otro
561	Su salud	Casado
562	Armonía familiar	Casado
563	Armonía familiar	Casado
564	Su salud	Soltero
565	El dinero	Soltero
566	Su salud	Otro
567	NS/NC	Casado
568	Su salud	Casado
569	El dinero	Casado
570	Armonía familiar	Soltero
571	Su vida afectiva	Casado
572	El dinero	Soltero
573	Su salud	Casado
574	Armonía familiar	Otro
575	Su salud	Casado
576	Su salud	Casado
577	Su salud	Casado
578	El dinero	Soltero
579	Su salud	Casado
580	Armonía familiar	Casado
581	El dinero	Soltero
582	El dinero	Casado
583	Armonía familiar	Casado
584	Su salud	Casado
585	El dinero	Soltero
586	Armonía familiar	Casado
587	Su salud	Soltero
588	Su salud	Casado
589	Su salud	Casado
590	Su salud	Casado
591	Su salud	Soltero
592	Armonía familiar	Otro
593	El dinero	Soltero
594	Armonía familiar	Casado
595	Armonía familiar	Casado
596	Armonía familiar	Otro

encuesta	preocupaciones	estado civil
945	El dinero	Casado
946	Su salud	Soltero
947	El dinero	Soltero
948	Su salud	Casado
949	El dinero	Casado
950	Armonía familiar	Casado
951	Armonía familiar	Otro
952	Su vida afectiva	Soltero
953	Su salud	Casado
954	Armonía familiar	Casado
955	Armonía familiar	Otro
956	Armonía familiar	Soltero
957	Armonía familiar	Casado
958	Su vida afectiva	Soltero
959	Armonía familiar	Casado
960	Su salud	Otro
961	Su salud	Casado
962	Armonía familiar	Otro
963	Su salud	Otro
964	Su vida afectiva	Soltero
965	El dinero	Casado
966	Su salud	Casado
967	Armonía familiar	Casado
968	Su salud	Casado
969	Su salud	Soltero
970	Armonía familiar	Casado
971	Su salud	Casado
972	El dinero	Soltero
973	Su salud	Casado
974	Su salud	Casado
975	Su salud	Casado
976	Su salud	Soltero
977	Su salud	Otro
978	Su salud	Casado
979	Su salud	Casado
980	El dinero	Soltero
981	Armonía familiar	Casado
982	Armonía familiar	Soltero
983	El dinero	Casado
984	Su salud	Casado
985	Su salud	Casado
986	Armonía familiar	Casado
987	Su salud	Casado
988	Armonía familiar	Casado
989	Armonía familiar	Casado
990	Armonía familiar	Casado
991	Su salud	Soltero
992	Armonía familiar	Casado
993	El dinero	Casado
994	Su vida afectiva	Casado
995	Armonía familiar	Casado
996	Su salud	Casado

encuesta	preocupaciones	estado civil
197	Su salud	Soltero
198	Su vida afectiva	Casado
199	El dinero	Casado
200	Armonía familiar	Casado
201	El dinero	Soltero
202	El dinero	Casado
203	Su salud	Casado
204	Su salud	Soltero
205	Armonía familiar	Casado
206	Su salud	Casado
207	Armonía familiar	Casado
208	Armonía familiar	Soltero
209	El dinero	Casado
210	Su salud	Casado
211	El dinero	Soltero
212	El dinero	Casado
213	Su salud	Casado
214	Su salud	Casado
215	Su salud	Casado
216	Su salud	Casado
217	Armonía familiar	Casado
218	Armonía familiar	Casado
219	Armonía familiar	Casado
220	Su vida afectiva	Casado
221	Armonía familiar	Casado
222	Armonía familiar	Casado
223	Armonía familiar	Soltero
224	Su salud	Soltero
225	El dinero	Casado
226	Su salud	Otro
227	Armonía familiar	Casado
228	Su salud	Casado
229	Armonía familiar	Casado
230	Su salud	Soltero
231	El dinero	Soltero
232	El dinero	Soltero
233	Armonía familiar	Casado
234	Armonía familiar	Casado
235	Armonía familiar	Otro
236	El dinero	Casado
237	El dinero	Casado
238	Su salud	Casado
239	Armonía familiar	Casado
240	Su salud	Casado
241	El dinero	Soltero
242	Su salud	Otro
243	Su salud	Casado
244	El dinero	Soltero
245	El dinero	Soltero
246	Su salud	Casado
247	Su salud	Otro
248	El dinero	Casado

encuesta	preocupaciones	estado civil
597	Armonía familiar	Casado
598	Su salud	Casado
599	NS/NC	Casado
600	Su salud	Casado
601	NS/NC	Casado
602	El dinero	Casado
603	El dinero	Casado
604	El dinero	Soltero
605	Su salud	Casado
606	El dinero	Casado
607	Su salud	Otro
608	El dinero	Soltero
609	Su salud	Casado
610	Su salud	Casado
611	Armonía familiar	Soltero
612	Su salud	Casado
613	Su salud	Casado
614	Su salud	Casado
615	Su salud	Soltero
616	Su salud	Casado
617	Su salud	Soltero
618	Su salud	Otro
619	Su salud	Otro
620	Su salud	Otro
621	El dinero	Casado
622	El dinero	Casado
623	El dinero	Casado
624	El dinero	Soltero
625	Su salud	Casado
626	Armonía familiar	Casado
627	Su salud	Casado
628	Su salud	Casado
629	Su salud	Otro
630	Su salud	Casado
631	Su vida afectiva	Soltero
632	Armonía familiar	Casado
633	Su salud	Otro
634	Su salud	Casado
635	Su salud	Otro
636	Su salud	Casado
637	Armonía familiar	Casado
638	Su salud	Soltero
639	El dinero	Soltero
640	El dinero	Otro
641	El dinero	Soltero
642	Su vida afectiva	Casado
643	Su salud	Soltero
644	Armonía familiar	Soltero
645	Su salud	Casado
646	Su salud	Casado
647	Su salud	Casado
648	Su salud	Casado

encuesta	preocupaciones	estado civil
997	El dinero	Casado
998	Su salud	Otro
999	El dinero	Casado
1000	Armonía familiar	Soltero
1001	El dinero	Casado
1002	Su salud	Otro
1003	NS/NC	Casado
1004	Su salud	Casado
1005	Armonía familiar	Casado
1006	Armonía familiar	Casado
1007	Armonía familiar	Casado
1008	El dinero	Casado
1009	Su salud	Casado
1010	Su salud	Casado
1011	Su salud	Casado
1012	El dinero	Casado
1013	Su salud	Casado
1014	Su salud	Otro
1015	Su salud	Otro
1016	Su salud	Casado
1017	Armonía familiar	Casado
1018	Su salud	Casado
1019	El dinero	Soltero
1020	El dinero	Soltero
1021	Su salud	Casado
1022	NS/NC	Casado
1023	Armonía familiar	Casado
1024	Armonía familiar	Casado
1025	Armonía familiar	Casado
1026	Armonía familiar	Casado
1027	Su salud	Soltero
1028	El dinero	Otro
1029	El dinero	Casado
1030	El dinero	Soltero
1031	Su salud	Casado
1032	NS/NC	Casado
1033	Armonía familiar	Soltero
1034	NS/NC	Soltero
1035	El dinero	Casado
1036	El dinero	Soltero
1037	Armonía familiar	Casado
1038	Su salud	Soltero
1039	El dinero	Casado
1040	Armonía familiar	Casado
1041	Su salud	Otro
1042	Armonía familiar	Casado
1043	Su salud	Soltero
1044	Su salud	Casado
1045	Su salud	Soltero
1046	El dinero	Casado
1047	Su salud	Casado
1048	Su salud	Casado

encuesta	preocupaciones	estado civil
249	Su salud	Soltero
250	Su salud	Casado
251	Su vida afectiva	Soltero
252	El dinero	Soltero
253	Su vida afectiva	Soltero
254	Su salud	Casado
255	Armonía familiar	Casado
256	Armonía familiar	Casado
257	Su salud	Casado
258	Su salud	Casado
259	Su salud	Casado
260	Armonía familiar	Soltero
261	Su vida afectiva	Casado
262	Su salud	Casado
263	Su vida afectiva	Casado
264	Su salud	Casado
265	Su vida afectiva	Casado
266	Armonía familiar	Casado
267	El dinero	Soltero
268	Su salud	Otro
269	Su salud	Casado
270	Su salud	Soltero
271	Su salud	Casado
272	Su vida afectiva	Casado
273	El dinero	Casado
274	Armonía familiar	Casado
275	Su salud	Casado
276	El dinero	Casado
277	El dinero	Soltero
278	Armonía familiar	Casado
279	Su salud	Soltero
280	Su salud	Casado
281	Su salud	Soltero
282	El dinero	Soltero
283	Su salud	Casado
284	Su salud	Casado
285	El dinero	Casado
286	El dinero	Casado
287	Armonía familiar	Casado
288	Su salud	Otro
289	Armonía familiar	Soltero
290	Armonía familiar	Casado
291	Armonía familiar	Casado
292	Armonía familiar	Casado
293	Su salud	Casado
294	Su salud	Casado
295	El dinero	Casado
296	Su vida afectiva	Casado
297	NS/NC	Casado
298	Su salud	Casado
299	Su salud	Casado
300	El dinero	Casado

encuesta	preocupaciones	estado civil
649	El dinero	Soltero
650	Su salud	Soltero
651	Su salud	Otro
652	Su salud	Casado
653	Su salud	Soltero
654	El dinero	Casado
655	Armonía familiar	Casado
656	Su salud	Casado
657	Su vida afectiva	Otro
658	Su salud	Otro
659	Armonía familiar	Soltero
660	Armonía familiar	Casado
661	Su salud	Casado
662	Su salud	Casado
663	Su salud	Casado
664	Su salud	Casado
665	Su salud	Casado
666	Su salud	Casado
667	Su vida afectiva	Casado
668	Su salud	Casado
669	Su salud	Casado
670	Su salud	Soltero
671	El dinero	Otro
672	Su salud	Soltero
673	Su salud	Casado
674	Su salud	Casado
675	El dinero	Soltero
676	Su salud	Casado
677	Su salud	Casado
678	Armonía familiar	Soltero
679	Su salud	Casado
680	Armonía familiar	Soltero
681	Su salud	Casado
682	Su salud	Soltero
683	Su salud	Casado
684	Su salud	Casado
685	Armonía familiar	Otro
686	Su salud	Casado
687	Su salud	Casado
688	El dinero	Soltero
689	El dinero	Otro
690	Su salud	Casado
691	Su salud	Casado
692	Su salud	Casado
693	Su salud	Casado
694	El dinero	Casado
695	NS/NC	Casado
696	NS/NC	Casado
697	Armonía familiar	Casado
698	Su salud	Casado
699	Su salud	Soltero
700	El dinero	Casado

encuesta	preocupaciones	estado civil
1049	NS/NC	Casado
1050	Su salud	Casado
1051	El dinero	Soltero
1052	Su salud	Casado
1053	Su salud	Casado
1054	Su salud	Casado
1055	Armonía familiar	Soltero
1056	El dinero	Otro
1057	Su salud	Casado
1058	Su salud	Casado
1059	Su salud	Otro
1060	Armonía familiar	Casado
1061	El dinero	Soltero
1062	Su salud	Soltero
1063	El dinero	Casado
1064	El dinero	Casado
1065	Su salud	Casado
1066	El dinero	Soltero
1067	Armonía familiar	Otro
1068	El dinero	Casado
1069	Armonía familiar	Soltero
1070	Su salud	Soltero
1071	Su salud	Soltero
1072	Su vida afectiva	Soltero
1073	Su salud	Casado
1074	Su salud	Soltero
1075	Su salud	Casado
1076	Su salud	Casado
1077	El dinero	Soltero
1078	Su salud	Soltero
1079	Su salud	Casado
1080	El dinero	Soltero
1081	Su salud	Casado
1082	NS/NC	Casado
1083	Su salud	Casado
1084	Armonía familiar	Soltero
1085	Su salud	Casado
1086	Armonía familiar	Casado
1087	Su salud	Casado
1088	Su salud	Casado
1089	El dinero	Casado
1090	Su salud	Casado
1091	Su salud	Casado
1092	El dinero	Soltero
1093	Su salud	Casado
1094	Su salud	Soltero
1095	Su salud	Casado
1096	Armonía familiar	Otro
1097	Armonía familiar	Otro
1098	Su salud	Soltero
1099	Armonía familiar	Soltero
1100	Su salud	Casado

encuesta	preocupaciones	estado civil
301	Su salud	Casado
302	El dinero	Casado
303	El dinero	Soltero
304	Su salud	Casado
305	El dinero	Casado
306	Su vida afectiva	Casado
307	Armonía familiar	Casado
308	El dinero	Soltero
309	Armonía familiar	Casado
310	NS/NC	Casado
311	El dinero	Casado
312	Su salud	Casado
313	Armonía familiar	Soltero
314	Armonía familiar	Casado
315	Armonía familiar	Casado
316	Su salud	Casado
317	El dinero	Casado
318	Armonía familiar	Casado
319	Armonía familiar	Casado
320	Su salud	Otro
321	Armonía familiar	Casado
322	Armonía familiar	Casado
323	Su salud	Casado
324	Su salud	Casado
325	Su salud	Casado
326	Su salud	Casado
327	El dinero	Soltero
328	Su salud	Casado
329	Su salud	Otro
330	Armonía familiar	Casado
331	Su salud	Soltero
332	Su salud	Casado
333	Su salud	Casado
334	Su salud	Casado
335	Su salud	Casado
336	El dinero	Casado
337	El dinero	Casado
338	Su salud	Otro
339	El dinero	Casado
340	El dinero	Casado
341	El dinero	Soltero
342	El dinero	Casado
343	El dinero	Casado
344	El dinero	Casado
345	Su salud	Casado
346	Armonía familiar	Casado
347	Su salud	Casado
348	El dinero	Casado
349	Armonía familiar	Soltero
350	NS/NC	Casado
351	Su salud	Soltero
352	Armonía familiar	Casado

encuesta	preocupaciones	estado civil
701	El dinero	Casado
702	Su salud	Casado
703	Su salud	Otro
704	Su salud	Otro
705	Su vida afectiva	Soltero
706	Armonía familiar	Casado
707	El dinero	Casado
708	Su salud	Casado
709	Su vida afectiva	Casado
710	Su salud	Casado
711	Su vida afectiva	Casado
712	Su vida afectiva	Casado
713	Su vida afectiva	Soltero
714	Su vida afectiva	Soltero
715	El dinero	Casado
716	Armonía familiar	Casado
717	Armonía familiar	Casado
718	Su salud	Casado
719	El dinero	Soltero
720	Armonía familiar	Casado
721	Su salud	Soltero
722	Su salud	Casado
723	Armonía familiar	Casado
724	Armonía familiar	Casado
725	Armonía familiar	Casado
726	El dinero	Soltero
727	Su salud	Casado
728	Armonía familiar	Casado
729	Armonía familiar	Casado
730	Armonía familiar	Casado
731	Su salud	Otro
732	Su salud	Soltero
733	Armonía familiar	Casado
734	El dinero	Soltero
735	Armonía familiar	Casado
736	El dinero	Casado
737	Su salud	Casado
738	Su salud	Casado
739	Armonía familiar	Casado
740	El dinero	Casado
741	Su salud	Casado
742	Su vida afectiva	Casado
743	El dinero	Casado
744	Armonía familiar	Casado
745	Su salud	Soltero
746	El dinero	Casado
747	El dinero	Casado
748	Armonía familiar	Soltero
749	Armonía familiar	Otro
750	Su salud	Casado
751	El dinero	Soltero
752	Armonía familiar	Casado

encuesta	preocupaciones	estado civil
1101	El dinero	Soltero
1102	Armonía familiar	Casado
1103	Armonía familiar	Casado
1104	Su salud	Soltero
1105	Armonía familiar	Casado
1106	Su salud	Casado
1107	El dinero	Casado
1108	Armonía familiar	Otro
1109	Armonía familiar	Casado
1110	Su salud	Casado
1111	Su salud	Casado
1112	El dinero	Casado
1113	El dinero	Soltero
1114	Su salud	Casado
1115	Su salud	Casado
1116	Su salud	Casado
1117	Armonía familiar	Soltero
1118	Su salud	Casado
1119	Su salud	Casado
1120	Armonía familiar	Casado
1121	Su salud	Casado
1122	El dinero	Casado
1123	NS/NC	Casado
1124	El dinero	Casado
1125	El dinero	Casado
1126	El dinero	Casado
1127	Armonía familiar	Casado
1128	Armonía familiar	Casado
1129	Su salud	Casado
1130	El dinero	Casado
1131	Su salud	Casado
1132	Su salud	Soltero
1133	Su vida afectiva	Casado
1134	NS/NC	Casado
1135	El dinero	Casado
1136	Su salud	Casado
1137	Su salud	Casado
1138	Armonía familiar	Casado
1139	Armonía familiar	Casado
1140	Su salud	Otro
1141	Su salud	Soltero
1142	Armonía familiar	Casado
1143	Su salud	Otro
1144	Su salud	Soltero
1145	El dinero	Casado
1146	El dinero	Casado
1147	Su salud	Casado
1148	Su vida afectiva	Soltero
1149	Su salud	Soltero
1150	El dinero	Casado
1151	Su salud	Casado
1152	Su salud	Casado

encuesta	preocupaciones	estado civil
353	El dinero	Casado
354	Su salud	Casado
355	El dinero	Otro
356	Su salud	Casado
357	Su salud	Casado
358	Armonía familiar	Casado
359	Su salud	Casado
360	Su salud	Casado
361	El dinero	Casado
362	Su salud	Soltero
363	Armonía familiar	Casado
364	Su salud	Casado
365	Su salud	Otro
366	Armonía familiar	Casado
367	El dinero	Soltero
368	Su salud	Casado
369	Su salud	Soltero
370	Su vida afectiva	Soltero
371	Su salud	Casado
372	Su salud	Casado
373	Su salud	Casado
374	Su salud	Casado
375	El dinero	Soltero
376	Su salud	Casado
377	Su salud	Casado
378	Su salud	Casado
379	El dinero	Casado
380	Su salud	Casado
381	Armonía familiar	Soltero
382	El dinero	Casado
383	Su salud	Soltero
384	El dinero	Soltero
385	Su vida afectiva	Casado
386	Su salud	Casado
387	Su salud	Soltero
388	Su salud	Otro
389	Armonía familiar	Soltero
390	Armonía familiar	Casado
391	Su salud	Casado
392	Su salud	Otro
393	Su salud	Casado
394	El dinero	Soltero
395	NS/NC	Casado
396	Armonía familiar	Casado
397	Armonía familiar	Soltero
398	El dinero	Soltero
399	Su salud	Casado
400	Armonía familiar	Casado

encuesta	preocupaciones	estado civil
753	El dinero	Casado
754	Su salud	Casado
755	Armonía familiar	Casado
756	Su salud	Soltero
757	Su salud	Casado
758	Su salud	Soltero
759	Su salud	Casado
760	El dinero	Soltero
761	El dinero	Casado
762	Su salud	Soltero
763	Su salud	Casado
764	Su salud	Soltero
765	Su salud	Casado
766	Armonía familiar	Soltero
767	Armonía familiar	Casado
768	El dinero	Casado
769	Su salud	Casado
770	Armonía familiar	Soltero
771	Armonía familiar	Casado
772	Armonía familiar	Casado
773	Armonía familiar	Casado
774	Su salud	Casado
775	Su salud	Soltero
776	Armonía familiar	Casado
777	Armonía familiar	Casado
778	Su salud	Otro
779	Su vida afectiva	Soltero
780	NS/NC	Casado
781	Su vida afectiva	Soltero
782	Armonía familiar	Casado
783	El dinero	Casado
784	Armonía familiar	Casado
785	Su salud	Casado
786	El dinero	Soltero
787	Su salud	Casado
788	Armonía familiar	Casado
789	Su salud	Casado
790	Armonía familiar	Soltero
791	Su salud	Casado
792	Su salud	Casado
793	Su salud	Casado
794	El dinero	Soltero
795	El dinero	Soltero
796	El dinero	Soltero
797	Su salud	Casado
798	Su salud	Casado
799	Armonía familiar	Casado
800	El dinero	Soltero

encuesta	preocupaciones	estado civil
1153	Su salud	Casado
1154	Armonía familiar	Casado
1155	El dinero	Soltero
1156	El dinero	Soltero
1157	Su salud	Soltero
1158	Su salud	Otro
1159	El dinero	Soltero
1160	El dinero	Casado
1161	Su salud	Casado
1162	Su salud	Casado
1163	El dinero	Soltero
1164	Su salud	Casado
1165	El dinero	Casado
1166	Armonía familiar	Casado
1167	El dinero	Casado
1168	Armonía familiar	Otro
1169	Su salud	Casado
1170	Su salud	Casado
1171	El dinero	Soltero
1172	Armonía familiar	Casado
1173	El dinero	Casado
1174	Su vida afectiva	Soltero
1175	NS/NC	Soltero
1176	Armonía familiar	Soltero
1177	Armonía familiar	Soltero
1178	Su salud	Casado
1179	Su vida afectiva	Soltero
1180	Su vida afectiva	Soltero
1181	El dinero	Casado
1182	Su salud	Casado
1183	Armonía familiar	Soltero
1184	Su salud	Casado
1185	Su vida afectiva	Soltero
1186	Su salud	Casado
1187	El dinero	Casado
1188	Su salud	Casado
1189	El dinero	Soltero
1190	NS/NC	Casado
1191	NS/NC	Casado
1192	Su salud	Soltero
1193	El dinero	Casado
1194	Armonía familiar	Otro
1195	Su salud	Casado
1196	Su salud	Otro
1197	Su salud	Soltero
1198	El dinero	Soltero
1199	Armonía familiar	Casado
1200	Su salud	Soltero

Base de Datos Correspondencias3.sav

Información sobre las variables

Variable	Etiqueta	Nivel de medida
personal	Grupo de personal	Nominal
fuman	Fumadores	Nominal

personal	fuman
Directores Sr	Nada
Directores Jr	Nada
Empleados Sr	Nada
Empleados Jr	Nada
Secretarias	Nada
Empleados Jr	Nada
Secretarias	Nada
Directores Jr	Nada
Empleados Sr	Nada
Directores Sr	Nada
Directores Sr	Poco
Directores Jr	Poco
Empleados Sr	Poco
Empleados Jr	Poco
Secretarias	Poco
Empleados Sr	Poco
Empleados Jr	Poco
Secretarias	Poco
Directores Jr	Poco
Empleados Sr	Poco
Directores Sr	Medio
Directores Jr	Medio
Empleados Sr	Medio
Empleados Jr	Medio
Secretarias	Medio
Directores Sr	Medio
Directores Jr	Medio
Empleados Sr	Medio
Empleados Jr	Medio
Secretarias	Medio
Directores Jr	Medio
Secretarias	Medio
Directores Sr	Medio
Directores Sr	Medio
Directores Sr	Medio
Directores Sr	Mucho
Directores Jr	Mucho
Empleados Sr	Mucho
Empleados Jr	Mucho
Secretarias	Mucho

[illegible][illegible]

personal	fuman
Empleados Jr	Poco
Empleados Jr	Poco
Empleados Jr	Poco
Empleados Jr	Poco
Empleados Jr	Poco
Empleados Jr	Poco
Empleados Jr	Poco
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Medio
Empleados Jr	Mucho
Empleados Jr	Mucho
Empleados Jr	Mucho
Empleados Jr	Mucho
Empleados Jr	Mucho
Empleados Jr	Mucho
Empleados Jr	Mucho
Secretarias	Nada
Secretarias	Nada
Secretarias	Nada
Secretarias	Nada
Secretarias	Nada
Secretarias	Nada
Secretarias	Poco
Secretarias	Poco
Secretarias	Poco
Secretarias	Poco
Secretarias	Poco
Secretarias	Poco
Secretarias	Medio
Secretarias	Medio

Información sobre las variables

Variable	Etiqueta	Nivel de medida
raza	Raza del encuestado	Nominal
región	Región de los Estados Unidos	Nominal

[illegible][illegible][illegible][illegible]

[illegible]

raza	región
Blanca	Oeste
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Otra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Negra	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Negra	Nor-Este
Negra	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este
Blanca	Nor-Este

[illegible][illegible][illegible]

[illegible]

raza	región
Blanca	Oeste
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Otra	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Otra	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Blanca	Sur-Este
Negra	Sur-Este
Negra	Sur-Este
Negra	Sur-Este
Negra	Sur-Este
Negra	Sur-Este
Negra	Sur-Este
Negra	Sur-Este
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Negra	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Negra	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste
Negra	Oeste
Negra	Oeste
Blanca	Oeste
Otra	Oeste
Otra	Oeste
Blanca	Oeste
Blanca	Oeste
Blanca	Oeste

[illegible][illegible][illegible]

[illegible]

[illegible]

raza	región	raza	región	raza	región	raza	región
Blanca	Sur-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Otra	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste
Blanca	Nor-Este	Blanca	Sur-Este	Blanca	Nor-Este	Blanca	Oeste

Base de Datos Coches_simple.sav

Información sobre las variables

Variable	Etiqueta	Nivel de medida
ECivil	E Civil	Nominal
procedencia	Procedencia	Nominal
peso	<ninguno>	Nominal
escivil	E Civil	Nominal
Origen	Procedencia	Nominal

ECivil	procedencia	peso	escivil	Origen
casado	Europeo	14	casado	Europeo
casado	Americano	37	casado	Americano
casado	Japonés	51	casado	Japonés
cahijos	Europeo	15	cahijos	Europeo
cahijos	Americano	52	cahijos	Americano
cahijos	Japones	44	cahijos	Japones
soltero	Europeo	15	soltero	Europeo
soltero	Americano	33	soltero	Americano
soltero	Japones	63	soltero	Japones
soltchijos	Europeo	1	soltchijos	Europeo
soltchijos	Americano	6	soltchijos	Americano
soltchijos	Japones	8	soltchijos	Japones

Capítulo 6

Base de Datos Cochesmul-9.sav

Origin	Type	Income	Kids	Sex
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Altos	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Foráneo	Propiedad	Altos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	varón
Nacional	Alquiler	Bajos	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Foráneo	Propiedad	Altos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	varón
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	varón
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	varón
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	varón
Nacional	Alquiler	Bajos	con hijos	varón
Foráneo	Alquiler	Medios	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	varón
Foráneo	Propiedad	Bajos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	varón
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Foráneo	Propiedad	Bajos	sin hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer

Origin	Type	Income	Kids	Sex
Nacional	Alquiler	Bajos	sin hijos	varón
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	mujer
Foráneo	Propiedad	Medios	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Altos	con hijos	mujer
Nacional	Alquiler	Altos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	#NULL!	mujer
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	varón
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Foráneo	Alquiler	Medios	sin hijos	varón
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	varón
Foráneo	Alquiler	Bajos	sin hijos	varón
Foráneo	Propiedad	Bajos	sin hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Bajos	sin hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	varón

Origin	Type	Income	Kids	Sex
Nacional	Propiedad	Altos	sin hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Alquiler	Altos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	mujer
Nacional	Alquiler	Altos	con hijos	varón
Foráneo	Alquiler	Bajos	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	sin hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Alquiler	Bajos	sin hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Alquiler	Altos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	sin hijos	varón
Foráneo	Alquiler	Medios	sin hijos	varón
Foráneo	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Altos	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	varón

Origin	Type	Income	Kids	Sex
Nacional	Alquiler	Bajos	sin hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Altos	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	sin hijos	varón
Foráneo	Propiedad	Bajos	sin hijos	varón
Foráneo	Alquiler	Bajos	con hijos	varón
Foráneo	Propiedad	Bajos	sin hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	sin hijos	varón
Nacional	Alquiler	Medios	sin hijos	varón
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Propiedad	Altos	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	varón
Nacional	Propiedad	Altos	con hijos	mujer
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	varón
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Nacional	Alquiler	Altos	sin hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Foráneo	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Altos	sin hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Foráneo	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Foráneo	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Altos	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	varón
Nacional	Alquiler	Medios	con hijos	mujer

Origen	Type	Income	Kids	Sex
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Foráneo	Propiedad	Bajos	sin hijos	varón
Foráneo	Alquiler	Bajos	sin hijos	varón
Foráneo	Alquiler	Medios	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Nacional	Alquiler	Bajos	con hijos	varón
Foráneo	Propiedad	Altos	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Foráneo	Alquiler	Medios	con hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Foráneo	Propiedad	Altos	con hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Altos	con hijos	mujer
Nacional	Alquiler	Medios	sin hijos	varón
Foráneo	Alquiler	Medios	#NULL!	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Foráneo	Alquiler	Altos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Foráneo	Propiedad	Altos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Altos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Medios	sin hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Altos	con hijos	varón
Nacional	Propiedad	Medios	sin hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer

Origen	Type	Income	Kids	Sex
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Propiedad	Altos	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	varón
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	varón
Nacional	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Propiedad	Medios	sin hijos	varón
Foráneo	Propiedad	Bajos	con hijos	varón
Nacional	Propiedad	0	sin hijos	varón
Foráneo	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	varón
Foráneo	Propiedad	Bajos	con hijos	mujer
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Nacional	Alquiler	Medios	con hijos	varón
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Propiedad	Bajos	con hijos	mujer
Foráneo	Propiedad	Medios	sin hijos	mujer
Nacional	Alquiler	Bajos	con hijos	varón
Foráneo	Propiedad	Medios	con hijos	mujer
Nacional	Alquiler	Bajos	con hijos	mujer
Nacional	Alquiler	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	varón
Foráneo	Alquiler	Bajos	sin hijos	varón
Foráneo	Alquiler	Altos	con hijos	mujer
Nacional	Propiedad	Medios	con hijos	varón
Nacional	Propiedad	Bajos	sin hijos	varón
Foráneo	Propiedad	Medios	sin hijos	varón
Foráneo	Alquiler	Medios	sin hijos	mujer
Nacional	Alquiler	Medios	sin hijos	varón

Base de Datos Correspondencias2-3.sav

encuesta	preocupaciones	estadocivil	sexo
1	El dinero	Casado	mujer
2	El dinero	Casado	mujer
3	Armonía familiar	Soltero	hombre
4	Armonía familiar	Otro	mujer
5	Su salud	Casado	mujer
6	Su salud	Otro	hombre
7	Su salud	Casado	hombre
8	El dinero	Casado	mujer
9	El dinero	Otro	mujer
10	Su salud	Casado	mujer
11	Armonía familiar	Casado	hombre
12	Su salud	Casado	mujer
13	El dinero	Soltero	hombre
14	Su salud	Otro	hombre
15	El dinero	Casado	mujer
16	Su vida afectiva	Otro	mujer
17	Armonía familiar	Casado	hombre
18	Su salud	Soltero	hombre
19	Su salud	Casado	mujer
20	Su vida afectiva	Casado	hombre
21	El dinero	Casado	mujer
22	Su salud	Casado	mujer
23	Su salud	Otro	mujer
24	El dinero	Casado	mujer
25	Su salud	Casado	mujer

encuesta	preocupaciones	estadocivil	sexo
26	Armonía familiar	Soltero	hombre
27	Armonía familiar	Soltero	mujer
28	Armonía familiar	Casado	mujer
29	Su vida afectiva	Soltero	hombre
30	Su salud	Casado	hombre
31	Su salud	Otro	mujer
32	Armonía familiar	Soltero	mujer
33	Su salud	Casado	hombre
34	Su salud	Otro	hombre
35	El dinero	Casado	mujer
36	El dinero	Casado	hombre
37	El dinero	Soltero	mujer
38	Su salud	Casado	hombre
39	Su salud	Casado	mujer
40	El dinero	Casado	mujer
41	Su salud	Casado	mujer
42	Armonía familiar	Casado	hombre
43	El dinero	Casado	mujer
44	El dinero	Soltero	hombre
45	Su salud	Otro	mujer
46	Su salud	Casado	hombre
47	Su vida afectiva	Casado	hombre
48	NS/NC	Soltero	mujer
49	Su salud	Casado	hombre
50	Su vida afectiva	Soltero	mujer

Capítulo 7

Base de Datos Canonica2-1.sav

encuesta	eanimo	gfelici	edad	ecivil	ideología	zonaresidencia
389	Preocupado	Poco feliz	menos de 25	Soltero	derecha	más 100000
390	Cansado	Bastante feliz	entre 41 65	Casado	derecha	más 100000
391	Cansado	Bastante feliz	entre 41 65	Casado	derecha	más 100000
392	Contento	Muy feliz	más de 65	Otros	Izquierda	más 100000
393	Cansado	Bastante feliz	entre 41 65	Casado	derecha	más 100000
394	Preocupado	Muy feliz	menos de 25	Soltero	derecha	más 100000
484	Contento	Poco feliz	menos de 25	Soltero	Izquierda	más 100000
485	Contento	Bastante feliz	entre 41 65	Casado	derecha	más 100000
486	Contento	Bastante feliz	más de 65	Otros	derecha	más 100000
487	Cansado	Bastante feliz	entre 41 65	Soltero	Izquierda	más 100000
488	Contento	Bastante feliz	entre 41 65	Casado	Izquierda	más 100000
490	Eufórico	Bastante feliz	menos de 25	Casado	derecha	más 100000
497	Aburrido	Muy feliz	entre 25 y 40	Casado	centro	más 100000
500	Cansado	Poco feliz	entre 41 65	Otros	derecha	más 100000
501	Cansado	Bastante feliz	entre 25 y 40	Casado	derecha	más 100000
503	Contento	Bastante feliz	entre 41 65	Casado	derecha	más 100000
834	Contento	Bastante feliz	entre 41 65	Casado	Izquierda	menos 100000
835	Preocupado	Muy feliz	entre 41 65	Casado	derecha	menos 100000
836	Aburrido	Poco feliz	entre 25 y 40	Casado	derecha	menos 100000
837	Contento	Muy feliz	entre 41 65	Casado	derecha	menos 100000
838	Contento	Muy feliz	entre 25 y 40	Casado	derecha	menos 100000
839	Preocupado	Poco feliz	menos de 25	Soltero	Izquierda	menos 100000
840	Preocupado	Bastante feliz	más de 65	Casado	derecha	menos 100000
841	Contento	Bastante feliz	entre 41 65	Casado	derecha	menos 100000
847	Deprimido	Poco feliz	entre 25 y 40	Soltero	derecha	menos 100000
848	Solo	Poco feliz	entre 41 65	Soltero	centro	menos 100000
1036	Cansado	Bastante feliz	menos de 25	Soltero	centro	menos 100000
1037	Contento	Muy feliz	entre 25 y 40	Casado	derecha	menos 100000
1038	Cansado	Nada feliz	entre 41 65	Soltero	derecha	menos 100000
1039	Preocupado	Poco feliz	entre 41 65	Casado	derecha	menos 100000
1040	Cansado	Bastante feliz	entre 41 65	Casado	derecha	menos 100000
1168	Cansado	Bastante feliz	entre 41 65	Otros	Izquierda	más 100000

Base de Datos Correlacionca2-1.sav

pregunta1	pregunta2	pregunta3	pregunta4	pregunta5	pregunta6
desacuerdo	mucho	poco	totalmente de acuerdo	normal	mucho
normal	normal	normal	normal	mucho	normal
totalmente de acuerdo	poco	poco	poco	poco	poco
mucho	desacuerdo	normal	desacuerdo	totalmente de acuerdo	mucho
desacuerdo	totalmente de acuerdo	mucho	totalmente de acuerdo	mucho	normal
poco	mucho	totalmente de acuerdo	mucho	mucho	poco
mucho	totalmente de acuerdo	normal	normal	normal	mucho
normal	normal	mucho	totalmente de acuerdo	normal	normal
poco	poco	poco	poco	poco	mucho
desacuerdo	mucho	normal	desacuerdo	poco	poco
normal	normal	mucho	poco	mucho	normal
totalmente de acuerdo	mucho	mucho	normal	totalmente de acuerdo	poco

pregunta1	pregunta2	pregunta3	pregunta4	pregunta5	pregunta6
desacuerdo	normal	normal	poco	normal	totalmente de acuerdo
desacuerdo	mucho	normal	desacuerdo	totalmente de acuerdo	mucho
normal	normal	poco	mucho	normal	poco
mucho	mucho	poco	poco	totalmente de acuerdo	normal
totalmente de acuerdo	normal	poco	desacuerdo	poco	normal
poco	mucho	totalmente de acuerdo	totalmente de acuerdo	normal	normal
desacuerdo	normal	normal	normal	normal	mucho
mucho	totalmente de acuerdo	normal	poco	mucho	mucho
normal	normal	poco	mucho	poco	poco
poco	mucho	poco	normal	normal	poco
desacuerdo	poco	normal	poco	mucho	mucho
mucho	poco	mucho	desacuerdo	poco	mucho
totalmente de acuerdo	mucho	normal	desacuerdo	poco	desacuerdo
normal	mucho	normal	desacuerdo	normal	totalmente de acuerdo
poco	normal	normal	mucho	normal	poco
desacuerdo	poco	mucho	totalmente de acuerdo	mucho	poco
poco	mucho	mucho	desacuerdo	poco	mucho
normal	normal	normal	totalmente de acuerdo	poco	normal
desacuerdo	poco	normal	poco	totalmente de acuerdo	mucho
normal	poco	normal	normal	mucho	desacuerdo
desacuerdo	mucho	poco	mucho	desacuerdo	totalmente de acuerdo
normal	normal	poco	poco	desacuerdo	desacuerdo
totalmente de acuerdo	poco	normal	mucho	totalmente de acuerdo	poco
normal	normal	normal	normal	normal	normal
mucho	totalmente de acuerdo	normal	poco	normal	mucho
desacuerdo	normal	normal	desacuerdo	mucho	normal
poco	mucho	desacuerdo	mucho	desacuerdo	mucho
desacuerdo	normal	normal	poco	mucho	desacuerdo
desacuerdo	normal	totalmente de acuerdo	totalmente de acuerdo	poco	poco
poco	desacuerdo	totalmente de acuerdo	mucho	poco	normal
normal	mucho	normal	totalmente de acuerdo	poco	desacuerdo
desacuerdo	normal	desacuerdo	totalmente de acuerdo	mucho	poco
poco	mucho	normal	normal	totalmente de acuerdo	normal
mucho	poco	desacuerdo	normal	totalmente de acuerdo	mucho
normal	normal	normal	mucho	poco	totalmente de acuerdo
mucho	normal	mucho	normal	totalmente de acuerdo	normal
mucho	normal	mucho	normal	mucho	poco
desacuerdo	totalmente de acuerdo	normal	normal	poco	mucho
desacuerdo	mucho	normal	poco	normal	totalmente de acuerdo
mucho	mucho	normal	normal	poco	normal
normal	mucho	normal	totalmente de acuerdo	normal	poco
mucho	totalmente de acuerdo	mucho	normal	poco	normal
mucho	mucho	poco	totalmente de acuerdo	poco	normal
mucho	totalmente de acuerdo	desacuerdo	normal	totalmente de acuerdo	poco
mucho	poco	normal	totalmente de acuerdo	normal	normal
normal	normal	desacuerdo	totalmente de acuerdo	mucho	poco
mucho	poco	normal	poco	mucho	mucho
poco	normal	mucho	normal	totalmente de acuerdo	poco
poco	mucho	poco	normal	poco	desacuerdo
desacuerdo	totalmente de acuerdo	normal	normal	poco	mucho
poco	poco	desacuerdo	normal	mucho	normal
normal	normal	totalmente de acuerdo	mucho	poco	desacuerdo
poco	totalmente de acuerdo	normal	normal	mucho	mucho
poco	poco	mucho	mucho	normal	normal
desacuerdo	totalmente de acuerdo	totalmente de acuerdo	normal	mucho	normal

Capítulo 8

Base de Datos DistComSV-2.sav

id	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
1	4.1	0.6	6.9	4.7	2.4	2.3	5.2	0	32	4.2	1	0	1	1
2	1.8	3	6.3	6.6	2.5	4	8.4	1	43	4.3	0	1	0	1
3	3.4	5.2	5.7	6	4.3	2.7	8.2	1	48	5.2	0	1	1	2
4	2.7	1	7.1	5.9	1.8	2.3	7.8	1	32	3.9	0	1	1	1
5	6	0.9	9.6	7.8	3.4	4.6	4.5	0	58	6.8	1	0	1	3
6	1.9	3.3	7.9	4.8	2.6	1.9	9.7	1	45	4.4	0	1	1	2
7	4.6	2.4	9.5	6.6	3.5	4.5	7.6	0	46	5.8	1	0	1	1
8	1.3	4.2	6.2	5.1	2.8	2.2	6.9	1	44	4.3	0	1	0	2
9	5.5	1.6	9.4	4.7	3.5	3	7.6	0	63	5.4	1	0	1	3
10	4	3.5	6.5	6	3.7	3.2	8.7	1	54	5.4	0	1	0	2
11	2.4	1.6	8.8	4.8	2	2.8	5.8	0	32	4.3	1	0	0	1
12	3.9	2.2	9.1	4.6	3	2.5	8.3	0	47	5	1	0	1	2
13	2.8	1.4	8.1	3.8	2.1	1.4	6.6	1	39	4.4	0	1	0	1
14	3.7	1.5	8.6	5.7	2.7	3.7	6.7	0	38	5	1	0	1	1
15	4.7	1.3	9.9	6.7	3	2.6	6.8	0	54	5.9	1	0	0	3
16	3.4	2	9.7	4.7	2.7	1.7	4.8	0	49	4.7	1	0	0	3
17	3.2	4.1	5.7	5.1	3.6	2.9	6.2	0	38	4.4	1	1	1	2
18	4.9	1.8	7.7	4.3	3.4	1.5	5.9	0	40	5.6	1	0	0	2
19	5.3	1.4	9.7	6.1	3.3	3.9	6.8	0	54	5.9	1	0	1	3
20	4.7	1.3	9.9	6.7	3	2.6	6.8	0	55	6	1	0	0	3
21	3.3	0.9	8.6	4	2.1	1.8	6.3	0	41	4.5	1	0	0	2
22	3.4	0.4	8.3	2.5	1.2	1.7	5.2	0	35	3.3	1	0	0	1
23	3	4	9.1	7.1	3.5	3.4	8.4	0	55	5.2	1	1	0	3
24	2.4	1.5	6.7	4.8	1.9	2.5	7.2	1	36	3.7	0	1	0	1
25	5.1	1.4	8.7	4.8	3.3	2.6	3.8	0	49	4.9	1	0	0	2
26	4.6	2.1	7.9	5.8	3.4	2.8	4.7	0	49	5.9	1	0	1	3
27	2.4	1.5	6.6	4.8	1.9	2.5	7.2	1	36	3.7	0	1	0	1
28	5.2	1.3	9.7	6.1	3.2	3.9	6.7	0	54	5.8	1	0	1	3
29	3.5	2.8	9.9	3.5	3.1	1.7	5.4	0	49	5.4	1	0	1	3
30	4.1	3.7	5.9	5.5	3.9	3	8.4	1	46	5.1	0	1	0	2
31	3	3.2	6	5.3	3.1	3	8	1	43	3.3	0	1	0	1
32	2.8	3.8	8.9	6.9	3.3	3.2	8.2	0	53	5	1	1	0	3
33	5.2	2	9.3	5.9	3.7	2.4	4.6	0	60	6.1	1	0	0	3
34	3.4	3.7	6.4	5.7	3.5	3.4	8.4	1	47.3	3.8	0	1	0	1
35	2.4	1	7.7	3.4	1.7	1.1	6.2	1	35	4.1	0	1	0	1
36	1.8	3.3	7.5	4.5	2.5	2.4	7.6	1	39	3.6	0	1	1	1
37	3.6	4	5.8	5.8	3.7	2.5	9.3	1	44	4.8	0	1	1	2
38	4	0.9	9.1	5.4	2.4	2.6	7.3	0	46	5.1	1	0	1	3
39	0	2.1	6.9	5.4	1.1	2.6	8.9	1	29	3.9	0	1	1	1
40	2.4	2	6.4	4.5	2.1	2.2	8.8	1	28	3.3	0	1	1	1
41	1.9	3.4	7.6	4.6	2.6	2.5	7.7	1	40	3.7	0	1	1	1
42	5.9	0.9	9.6	7.8	3.4	4.6	4.5	0	58	6.7	1	0	1	3
43	4.9	2.3	9.3	4.5	3.6	1.3	6.2	0	53	5.9	1	0	0	3
44	5	1.3	8.6	4.7	3.1	2.5	3.7	0	48	4.8	1	0	0	2
45	2	2.6	6.5	3.7	2.4	1.7	8.5	1	38	3.2	0	1	1	1
46	5	2.5	9.4	4.6	3.7	1.4	6.3	0	54	6	1	0	0	3
47	3.1	1.9	10	4.5	2.6	3.2	3.8	0	55	4.9	1	0	1	3
48	3.4	3.9	5.6	5.6	3.6	2.3	9.1	1	43	4.7	0	1	1	2
49	5.8	0.2	8.8	4.5	3	2.4	6.7	0	57	4.9	1	0	1	3

id	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
50	5.4	2.1	8	3	3.8	1.4	5.2	0	53	3.8	1	0	1	3
51	3.7	0.7	8.2	6	2.1	2.5	5.2	0	41	5	1	0	0	2
52	2.6	4.8	8.2	5	3.6	2.5	9	1	53	5.2	0	1	1	2
53	4.5	4.1	6.3	5.9	4.3	3.4	8.8	1	50	5.5	0	1	0	2
54	2.8	2.4	6.7	4.9	2.5	2.6	9.2	1	32	3.7	0	1	1	1
55	3.8	0.8	8.7	2.9	1.6	2.1	5.6	0	39	3.7	1	0	0	1
56	2.9	2.6	7.7	7	2.8	3.6	7.7	0	47	4.2	1	1	1	2
57	4.9	4.4	7.4	6.9	4.6	4	9.6	1	62	6.2	0	1	0	2
58	5.4	2.5	9.6	5.5	4	3	7.7	0	65	6	1	0	0	3
59	4.3	1.8	7.6	5.4	3.1	2.5	4.4	0	46	5.6	1	0	1	3
60	2.3	4.5	8	4.7	3.3	2.2	8.7	1	50	5	0	1	1	2
61	3.1	1.9	9.9	4.5	2.6	3.1	3.8	0	54	4.8	1	0	1	3
62	5.1	1.9	9.2	5.8	3.6	2.3	4.5	0	60	6.1	1	0	0	3
63	4.1	1.1	9.3	5.5	2.5	2.7	7.4	0	47	5.3	1	0	1	3
64	3	3.8	5.5	4.9	3.4	2.6	6	0	36	4.2	1	1	1	2
65	1.1	2	7.2	4.7	1.6	3.2	10	1	40	3.4	0	1	1	1
66	3.7	1.4	9	4.5	2.6	2.3	6.8	0	45	4.9	1	0	0	2
67	4.2	2.5	9.2	6.2	3.3	3.9	7.3	0	59	6	1	0	0	3
68	1.6	4.5	6.4	5.3	3	2.5	7.1	1	46	4.5	0	1	0	2
69	5.3	1.7	8.5	3.7	3.5	1.9	4.8	0	58	4.3	1	0	0	3
70	2.3	3.7	8.3	5.2	3	2.3	9.1	1	49	4.8	0	1	1	2
71	3.6	5.4	5.9	6.2	4.5	2.9	8.4	1	50	5.4	0	1	1	2
72	5.6	2.2	8.2	3.1	4	1.6	5.3	0	55	3.9	1	0	1	3
73	3.6	2.2	9.9	4.8	2.9	1.9	4.9	0	51	4.9	1	0	0	3
74	5.2	1.3	9.1	4.5	3.3	2.7	7.3	0	60	5.1	1	0	1	3
75	3	2	6.6	6.6	2.4	2.7	8.2	1	41	4.1	0	1	0	1
76	4.2	2.4	9.4	4.9	3.2	2.7	8.5	0	49	5.2	1	0	1	2
77	3.8	0.8	8.3	6.1	2.2	2.6	5.3	0	42	5.1	1	0	0	2
78	3.3	2.6	9.7	3.3	2.9	1.5	5.2	0	47	5.1	1	0	1	3
79	1	1.9	7.1	4.5	1.5	3.1	9.9	1	39	3.3	0	1	1	1
80	4.5	1.6	8.7	4.6	3.1	2.1	6.8	0	56	5.1	1	0	0	3
81	5.5	1.8	8.7	3.8	3.6	2.1	4.9	0	59	4.5	1	0	0	3
82	3.4	4.6	5.5	8.2	4	4.4	6.3	0	47.3	5.6	1	1	1	2
83	1.6	2.8	6.1	6.4	2.3	3.8	8.2	1	41	4.1	0	1	0	1
84	2.3	3.7	7.6	5	3	2.5	7.4	0	37	4.4	1	1	0	1
85	2.6	3	8.5	6	2.8	2.8	6.8	1	53	5.6	0	1	0	2
86	2.5	3.1	7	4.2	2.8	2.2	9	1	43	3.7	0	1	1	1
87	2.4	2.9	8.4	5.9	2.7	2.7	6.7	1	51	5.5	0	1	0	2
88	2.1	3.5	7.4	4.8	2.8	2.3	7.2	0	36	4.3	1	1	0	1
89	2.9	1.2	7.3	6.1	2	2.5	8	1	34	4	0	1	1	1
90	4.3	2.5	9.3	6.3	3.4	4	7.4	0	60	6.1	1	0	0	3
91	3	2.8	7.8	7.1	3	3.8	7.9	0	49	4.4	1	1	1	2
92	4.8	1.7	7.6	4.2	3.3	1.4	5.8	0	39	5.5	1	0	0	2
93	3.1	4.2	5.1	7.8	3.6	4	5.9	0	43	5.2	1	1	1	2
94	1.9	2.7	5	4.9	2.2	2.5	8.2	1	36	3.6	0	1	0	1
95	4	0.5	6.7	4.5	2.2	2.1	5	0	31	4	1	0	1	1
96	0.6	1.6	6.4	5	0.7	2.1	8.4	1	25	3.4	0	1	1	1
97	6.1	0.5	9.2	4.8	3.3	2.8	7.1	0	60	5.2	1	0	1	3
98	2	2.8	5.2	5	2.4	2.7	8.4	1	38	3.7	0	1	0	1
99	3.1	2.2	6.7	6.8	2.6	2.9	8.4	1	42	4.3	0	1	0	1
100	2.5	1.8	9	5	2.2	3	6	0	33	4.4	1	0	0	1

Base de Datos HIV-2.sav

sex	age	involvement	healed	perceived_risk	perceived_severity	self_efficacy	maladaptive_coping	willing_use	response_efficacy	likely_use	fear	intended_use	certain_use
female	27	yes	no	5	5	2	3	4	4	3	2	3	3
male	19	no	no	4	5	2	2	3	4	3	3	3	3
female	39	yes	no	5	4	4	4	2	4	2	2	2	2
male	22	yes	no	5	5	5	2	3	5	3	4	4	2
female	23	yes	no	4	4	4	4	4	4	3	4	3	3
female	20	yes	no	5	4	5	4	5	5	5	3	5	4
female	21	yes	yes	5	4	4	4	2	4	2	4	2	2
female	19	no	no	5	3	4	2	4	4	4	4	4	4
female	26	yes	no	5	4	4	1	5	5	4	3	4	4
female	48	no	no	5	5	4	1	4	5	5	5	5	4
male	20	yes	yes	5	4	2	2	4	4	2	2	4	3
female	21	yes	yes	5	1	5	2	2	4	2	4	2	2
male	19	yes	no	5	4	2	4	4	5	2	4	3	2
male	54	yes	no	5	3	4	5	5	5	5	3	5	5
female	52	yes	no	5	4	4	5	5	5	1	2	1	1
male	19	yes	yes	5	4	4	3	5	4	5	4	5	5
female	31	yes	no	5	4	4	1	5	1	5	5	5	5
male	23	yes	no	4	2	4	4	2	4	2	2	2	2
female	38	yes	yes	5	4	4	4	4	5	2	1	2	2
female	19	yes	no	4	4	3	2	4	4	4	4	4	3
male	21	no	yes	5	4	4	2	5	5	5	1	5	5
female	20	yes	no	5	5	2	2	5	4	5	5	5	5
female	21	yes	no	5	5	2	3	5	5	5	5	4	4
female	20	yes	no	5	4	4	2	4	4	2	4	4	4
male	19	no	yes	4	5	4	1	5	5	5	5	5	5
female	19	yes	yes	4	5	4	3	4	4	4	4	4	4
female	20	1.333	yes	5	4	4	1	5	1	5	4	4	5
male	23	yes	no	5	2	2	2	4	4	3	2	3	2
female	28	yes	no	5	3	4	1	5	4	4	5	5	5
male	20	no	no	4	4	4	1	4	4	4	2	5	5
female	29	yes	yes	5	5	4	1	5	5	5	4	5	5
female	21	yes	yes	5	3	3	3	3	4	3	4	3	3
female	25	yes	no	5	4	4	2	5	4	5	4	5	5
male	21	no	yes	5	4	3	1	4	4	4	3	4	4
male	19	no	yes	5	4	2	2	5	4	4	5	5	5
male	19	no	no	5	5	4	2	4	4	4	4	4	4
male	24	yes	no	4	3	4	4	4	4	2	2	2	2
male	29	yes	no	5	4	5	1	5	5	2	4	4	2
male	26.9	yes	1.761	5	1	2	4	4	4	2	1	2	2
female	17	no	no	4	4	4	2	5	5	5	3	5	5
female	23	no	yes	5	4	3	2	5	4	5	4	5	5
female	51	no	yes	5	5	4	2	4	4	4	2	4	5
female	34	no	no	5	1	5	1	5	5	5	4	5	5
female	22	yes	no	5	5	4	2	5	5	5	2	5	5
female	20	yes	no	5	5	5	1	5	5	5	5	5	5

sex	age	involv e	healedu	perceive d_risk	percei ved_se verity	self_effi cacy	maladaptive _coping	willing_ use	response_ef ficacy	likely_ use	fear	inten d_use	certain_use
female	21	yes	no	5	4	4	3	4	5	3	4	4	3
male	27	yes	no	5	4	4	2	4	5	3	4	4	3
female	21	yes	yes	5	4	4	3	4	5	1	4	1	3
female	32	yes	no	5	4	4	2	4	4	5	1	5	4
female	38	yes	no	5	1	4	5	5	4	5	4	5	5
female	30	yes	no	5	5	4	2	4	4	3	4	3	3
female	28	yes	no	5	4	2	2	4	4	5	5	5	5
female	17	yes	yes	5	5	4	2	5	4	5	5	5	5
male	17	no	no	3	2	5	2	5	5	4	3	5	3
male	25	yes	no	4	3	4	3	3	5	1	5	3	3
female	37	yes	yes	4	2	5	1	5	5	4	5	5	4
female	38	yes	yes	4	2	3	1	5	4	5	4	5	5
male	22	yes	no	5	5	4	1	5	5	5	5	5	5
male	22	yes	no	5	3	3.1	1	5	5	4	5	4	4
male	18	no	no	5	3	3	3	5	5	5	2	5	5
female	18	no	no	5	2	4	2	5	4	5	4	5	5
female	35	yes	no	5	5	5	1	5	5	5	5	5	5
female	19	no	yes	5	4	4	2	5	5	5	2	5	5
male	27	yes	no	4	2	2	1	4	4	2	2	1	1
male	21	no	no	5	2	2	4	5	5	1	2	1	1
female	32	yes	no	5	5	2	2	5	5	5	5	5	5
female	48	yes	no	5	3	5	1	3	5	3	5	3	3
female	32	yes	no	5	5	4	5	5	5	5	5	5	5
male	49	yes	no	5	4	1	1	5	4	4	2	4	5
female	31	yes	no	5	2	4	2	3	5	3	3	3	3
male	24	yes	no	4	4	5	2	4	5	3	2	4	3
male	17	no	yes	1	1	1	5	1	1	1	3	5	1
male	19	yes	yes	4	2	4	1	5	5	4	3	5	4
male	43	no	yes	4	4	4	2	4	4	4	4	4	4
female	19	yes	no	5	4	4	2	4	5	4	3	4	2
female	17	yes	no	5	1	4	4	5	5	5	4	5	5
female	18	no	no	4	4	4	3	5	5	5	5	5	5
male	18	yes	no	5	4	4	4	4	1	5	5	4	1
female	17	no	no	1	3	4	2	5	4	4	3	5	5
female	17	no	no	5	5	2	2	5	5	5	5	5	5
female	17	no	no	4	2	4	2	4	4	4	2	4	4
male	19	yes	no	4	4	3	3	4	4	4	5	4	4
female	20	yes	no	5	5	1	5	1	1	1	5	1	1
male	17	no	no	5	5	5	2	5	5	5	2	5	3
male	17	yes	yes	4	4	4	2	4	4	4	4	4	4
male	17	no	no	5	4	4	4	4	4	4	4	4	4
female	17	no	no	4	2	4	2	5	5	5	3	5	5
female	18	yes	no	5	5	5	1	5	5	5	5	5	5
male	19	yes	no	4	5	3	5	1	5	2	2	2	1
male	19	yes	yes	5	5	5	2	4	5	1	3	5	4
male	18	yes	no	5	5	5	1	5	5	5	3	5	5
male	22	yes	no	5	5	4	2	4	4	4	4	4	4
male	17	yes	yes	5	2	5	4	2	5	2	2	2	1
male	20	yes	yes	4	2	3	3	3	4	3	2	3	3
male	19	yes	no	5	2	3	2	4	4	3	2	3	3
male	17	no	no	4	4	4	4	5	5	4	2	5	4

sex	age	involv e	healedu	perceive d_risk	percei ved_se verity	self_effi cacy	maladaptive _coping	willing_ use	response_ef ficacy	likely_ use	fear	inten d_use	certain_use
male	19	no	no	5	4	5	1	5	5	5	1	5	5
male	17	yes	yes	4	4	4	2	5	5	5	5	5	5
female	17	no	yes	5	4	4	1	5	5	5	4	5	5
female	17	no	yes	5	5	2	2	5	4	5	4	5	5
female	18	yes	no	4	4	3	2	5	4	5	4	5	4
male	17	no	no	5	3	4	5	5	5	5	4	5	5
male	17	no	yes	5	2	4	5	5	5	5	3	5	5
female	18	yes	no	5	4	3	1	5	5	5	3	5	5
female	27	yes	yes	4	5	3	2	5	5	5	5	5	4
male	44	yes	yes	4	2	2	4	2	2	2	4	2	2
female	19	no	no	5	1	1	5	1	5	1	1	1	1
male	19	yes	yes	2	5	4	1	3	4	3	3	3	3
male	25	no	no	3	4	4	3	4	4	4	4	4	4
female	20	yes	no	5	4	3	3	5	4	5	5	5	5
female	23	yes	no	5	2	4	2	4	4	4	2	4	2
female	28	yes	yes	5	4	3	5	3	4	3	5	1	1
male	38	yes	yes	5	4	5	5	5	4	2	2	4	2
female	18	yes	no	5	4	5	2	2	5	2	5	2	2
female	32	yes	no	5	5	5	1	5	5	5	5	5	5
male	27	yes	no	4	4	3	5	1	1	1	1	1	1
female	22	yes	no	5	4	3	2	4	5	4	4	4	4
female	20	yes	no	5	4	4	2	5	5	4	3	4	5
male	55	yes	yes	5	2	2	4	4	2	2	4	2	2
female	27	yes	no	5	4	2	4	4	4	4	4	4	4
female	18	yes	yes	5	4	5	5	4	5	4	5	4	4
female	25	yes	no	5	3	2	2	3	3	3	2	3	3
female	25	yes	yes	5	4	4	3	5	5	5	2	5	5
female	22	yes	yes	5	4	2	2	5	4	5	4	5	4
male	19	yes	no	5	4	5	5	4	5	3	2	3	4
female	21	yes	no	5	4	4	1	4	5	4	4	4	5
female	18	yes	yes	5	2	4	3	4	4	3	2	3	2
male	24	yes	no	5	3	4	1	5	5	3	4	3	2
female	32	yes	no	4	5	4	2	3	4	2	2	2	2
female	21	yes	yes	4	4	4	2	5	4	5	4	5	5
male	20	yes	no	5	4	4	4	4	4	4	4	4	4
male	25	yes	yes	4	2	2	2	4	4	4	2	4	2
female	19	yes	no	4	1	4	1	5	5	5	2	5	5
female	19	no	yes	5	5	2	1	5	4	5	4	5	5
female	23	yes	yes	5	2	4	4	4	5	3	4	4	3
female	47	yes	no	4	5	2	4	2	2	2	4	2	2
male	24	yes	no	5	5	5	1	3	5	2	3	3	3
female	19	no	yes	5	3	4	2	4	5	4	5	5	4
female	23	yes	no	5	4	4	3	4	4	4	4	4	4
male	19	no	yes	4	4	5	2	5	5	5	4	5	5
female	20	yes	yes	5	4	4	1	5	5	5	5	5	5
female	19	yes	no	4	2	4	3	5	5	2	4	2	3
female	19	no	no	5	4	2	2	5	5	5	5	5	5
female	20	no	yes	5	5	4	2	5	5	5	4	5	5
female	21	yes	no	5	5	4	4	5	5	4	4	4	4
female	24	yes	no	5	5	2	1	4	4	4	4	4	2
male	33	yes	yes	5	2	4	1	4	4	4	4	4	4

sex	age	involv e	healedu	perceive d_risk	percei ved_se verity	self_effic acy	maladaptive _coping	willing_ use	response_ef ficacy	likely_ use	fear	inten d_use	certain_use
female	19	yes	yes	5	3	3	2	4	4	3	4	2	3
female	18	no	yes	4	4	3	2	4	4	4	4	4	4
female	18	yes	no	5	4	2	1	5	4	4	4	4	2
female	24	yes	yes	5	4	5	3	4	5	4	3	4	2
female	24	yes	no	5	5	4	2	5	5	4	5	4	4
female	18	yes	yes	4	3	3	3	4	4	4	2	4	4
female	25	yes	no	4	3	2	3	2	4	2	3	2	2
female	33	yes	no	5	3	5	1	5	5	4	4	5	5
female	20	no	yes	5	5	3	2	5	4	5	4	5	5
female	30	yes	no	5	2	4	2	4	4	4	2	4	4
female	18	yes	yes	4	3	4	3	4	4	3	3	3	3
female	38	yes	yes	5	1	5	1	5	5	5	2	5	4
female	18	yes	no	4	3	3	3	4	4	4	2	4	4
female	18	yes	yes	5	5	4	1	5	5	5	5	5	5
male	40	yes	yes	5	4	2	2	3	4	4	2	3	3
female	18	yes	no	5	4	4	2	5	4	4	2	4	5
male	25	yes	no	4	4	4	4	4	4	3	2	3	3
female	20	yes	yes	5	5	2	1	4	5	4	5	4	3
female	20.1	yes	yes	5	5	4	1	5	5	5	4	5	5
male	24	yes	no	5	5	4	1	5	5	5	4	5	5
male	23.8	no	no	5	3	5	2	5	4	4	2	4	4
male	17	yes	yes	2	5	3	4	1	5	1	4	1	3
male	17	yes	yes	4	2	4	2	5	5	5	3	5	5
male	17	yes	yes	5	5	4	2	5	4	5	4	5	5
male	18	yes	no	5	4	4	2	5	4	4	4	4	4
female	37	yes	no	4	3	5	2	4	4	4	4	4	4
female	18	no	no	5	4	3	5	3	4	2	2	2	2
female	17	no	no	5	4	2	2	5	5	5	4	5	5
female	17	yes	yes	4	1	4	3	4	4	3	2	2	3
female	18	yes	yes	4	5	2	2	5	4	5	2	4	4
female	22	yes	yes	5	5	2	1	5	5	5	5	5	5
female	27	yes	no	5	5	5	2	5	5	4	5	4	2
female	20	yes	no	5	4	2	1	2	2	1	5	1	1
male	19	yes	no	2	2	4	4	2	2	2	1	2	2
female	17	yes	no	4	3	2	2	4	1	4	4	4	5
male	33	yes	no	5	2	4	5	5	4	2	2	2	2
female	33	yes	no	5	5	4	1	2	4	4	2	2	2
female	21.4	yes	no	5	4	4	1	5	4	1	4	5	5
female	17	yes	no	5	4	4	3	4	4	4	2	4	4
female	17	yes	yes	5	5	4	2	5	4	4	3	5	5
female	17	yes	no	4	4	3	2	4	5	4	3	4	4
female	20	yes	no	4	5	4	1	4	3	5	5	3	5
female	17	yes	no	5	4	4	2	5	5	5	5	5	5
male	17	yes	yes	5	5	2	2	4	4	4	4	4	2
female	17	no	no	5	5	4	2	5	4	5	4	5	5
male	17	yes	yes	5	4	2	2	5	5	4	4	4	2
female	18	yes	yes	5	2	5	1	4	5	3	2	3	2
male	18	yes	yes	5	5	4	2	5	4	5	4	5	5
female	17	yes	no	5	4	5	2	4	5	3	3	3	3
male	18	yes	no	5	3	3	3	3	4	3	1	3	3
female	18	yes	no	5	2	4	2	4	4	4	1	4	4

sex	age	involv e	healedu	perceive d_risk	percei ved_se verity	self_effi cacy	maladaptive _coping	willing_ use	response_ef ficacy	likely_ use	fear	inten d_use	certain_use
male	18	no	no	4	3	4	3	4	4	4	2	4	5
male	29	yes	no	5	4	4	2	4	5	3	4	3	3
female	26	yes	yes	5	1	5	1	5	4	5	2	5	5
male	18	yes	yes	5	2	4	2	5	5	5	2	5	5
female	18	no	yes	5	5	3	3	5	5	5	4	5	5
male	19	yes	yes	5	4	5	3	5	5	3	3	5	5
male	33	yes	yes	4	2	4	1	4	4	2	2	2	2
male	20	no	yes	5	4	4	1	5	4	5	4	5	5
female	23	yes	yes	5	5	5	4	3	5	3	3	3	3
female	17	yes	no	5	5	4	1	5	5	4	5	5	5
male	17	yes	no	5	2	4	2	3	5	3	3	3	3
female	32	yes	no	4	5	4	2	3	4	2	3	2	2
female	18	no	no	4	2	4	2	5	4	4	2	4	5
female	17	yes	yes	5	5	5	3	5	5	5	3	5	5
male	17	1.393	1.38455	5	2	4	2	5	4	5	5	5	5
male	17	yes	yes	5	5	5	3	5	5	5	3	5	5
male	19	yes	yes	4	4	5	2	3	5	3	2	3	1
male	18	yes	yes	5	5	4	4	1	5	1	4	1	1
female	19	yes	no	5	2	4	2	5	4	4	3	3	2
male	18	yes	yes	4	3	3	2	5	4	5	3	5	5
male	20	yes	no	4	3	2	1	2	4	1	4	1	1
female	21	yes	no	5	5	4	5	2	4	1	3	2	1
male	17	yes	yes	4	4	4	1	5	4	5	5	4	5
female	18	yes	no	3	4	4	2	4	4	3	2	2	3
male	21	yes	no	5	1	2	1	5	5	5	4	5	5
female	17	no	no	5	5	5	1	5	5	5	4	5	5
female	17	no	no	4	2	4	2	5	4	5	2	5	4
female	18	yes	yes	5	4	4	1	2	5	2	4	2	1
female	17	yes	yes	4	3	3	2	4	4	4	3	4	4
female	17	yes	yes	5	4	4	1	4	5	4	5	4	4
female	17	no	no	4	3	4	2	4	4	4	4	4	4
female	37	yes	no	5	4	3	1	5	5	4	4	5	5
female	17	no	no	5	5	1	2	5	4	5	5	5	5
male	20	yes	yes	5	5	2	3	5	5	3	3	3	3
male	17	no	yes	4	3	4	3	5	4	4	4	5	5
male	17	yes	no	2	1	1	5	1	5	2	1	2	5
female	17	yes	yes	1	1	1	5	1	5	1	1	5	1
female	17	yes	no	5	1	2	2	3	4	3	1	1	1
female	21	yes	yes	5	4	4	2	4	4	4	4	4	4
female	21	yes	yes	4	2	5	1	5	5	5	5	5	5
male	19	yes	no	5	5	4	1	1	5	3	1	2	3
female	17	no	no	5	2	2	3	4	5	3	3	4	4
female	34	yes	yes	5	3	1	1	3	1	1	5	1	3
female	45	yes	no	2	4	4	2	3	4	2	2	2	2
female	27	no	yes	1	1	4	2	4	5	4	2	4	4
male	26	yes	yes	4	3	3	2	4	4	4	3	4	4
female	32	yes	no	3	3	2	1	4	3	3	5	3	3



Impreso en
Talleres Gráficos UCA,
en abril del 2016.
San Salvador, El Salvador, C. A.
La edición consta de 500 ejemplares.

Análisis multivariante: Aplicaciones con SPSS

El análisis multivariante tiene múltiples aplicaciones, desde crear perfiles de grupos de individuos, de niveles de satisfacción, realizar tratamiento y análisis de encuestas de satisfacción (políticas, económicas), encuestas de opinión pública; utilizando escalas de percepciones, identificar ciudades que sean comparables para abordar políticas (sociales y económicas) similares. El objetivo del análisis multivariante es propiciar el estudio conjunto de múltiples variables (cualitativas y cuantitativas), procurando obtener información útil y de apoyo a la toma de decisiones; para ello, reduce el número de variables y sintetiza grandes conjuntos de datos, buscando una mejor interpretación y, por lo tanto, una mayor comprensión de los fenómenos que se estudian.



Elnor Crespín Elías (Tonacatepeque, El Salvador, 1968)

Doctor en Ciencias Económicas por la Universidad Autónoma de Baja California-México; Maestría en Dirección de Empresas (UCA), Licenciado en Matemática en la Especialidad de Computación y Estadística (UES); Postgrado en Gestión de la Investigación Científica; Diplomado en Gestión de la Calidad Total en Educación por la Universidad Católica de Honduras, y en Administración de la Calidad Total aplicado a Empresas Salvadoreñas, por la Universidad de Georgetown (EEUU).

Es investigador en el campo de la economía de la educación y teoría de juegos, con experiencia en dirección de proyectos de investigación y consultorías en las áreas de Educación, Economía, Estadística y Sistemas de Calidad. Ha trabajado en el sector de educación superior en los últimos 20 años, y se ha desempeñado como docente de postgrado en las áreas relacionadas a Métodos estadísticos, Teoría de juegos y Metodología de investigación; además forma parte del equipo de Pares Evaluadores del Ministerio de Educación y del Consejo Nacional de Evaluación y Acreditación Universitaria de Panamá (CONEAUPA). A la fecha es Coordinador de Proyectos del Instituto de Ciencia, Tecnología e Innovación, de la Universidad Francisco Gavidia.



Visita nuestro
sitio web

ISBN 978-999-23-47-54-6