

ÍNDICE

<i>Capítulo 1. PROGRAMAS INFORMÁTICO-ESTADÍSTICOS: SPSS Y R</i>	13
1. El análisis de datos con ordenador	15
2. Introducción al SPSS	16
2.1. Ventanas de SPSS	17
2.2. Menús	18
2.3. Barra de herramientas y barra de estado	20
2.4. Entrada y ejecución del SPSS	20
2.5. Pasos básicos para realizar un análisis de datos con SPSS.....	21
2.6. Matriz de datos	22
3. El programa R y R-Commander.....	23
3.1. Lectura y escritura de datos	25
3.1.1. Lectura de datos	25
3.1.2. Exportar datos	27
3.1.3. Guardar y leer datos	27
4. Análisis descriptivo de datos	28
4.1. Con SPSS.....	28
4.2. Con R	29
4.2.1. La exploración de los datos.....	29
5. La relación entre variable	31
5.1. Con SPSS.....	31
5.2. Con R	31
6. Gráficos	32
6.1. Con SPSS.....	32
6.2. Con R	33
7. La decisión estadística	40
7.1. Pruebas paramétricas	41
7.1.1. Con SPSS	44
7.2. Pruebas no paramétricas	44
7.2.1. Con SPSS	45
7.2.2. Con R.....	46

8. El modo lineal	47
8.1. Regresión lineal	47
8.1.1. Con SPSS	48
8.2. Análisis de la varianza	48
8.2.1. Con SPSS	48
9. Un ejemplo de investigación con SPSS y R	49
Referencias bibliográficas	61

Capítulo 2. INTRODUCCIÓN A LA SELECCIÓN DE MUESTRAS EN EL ÁMBITO EDUCATIVO

1. Introducción	63
1.1. Diseños muestrales	67
1.2. Selección de la muestra en diseños probabilísticos	68
2. Muestreo aleatorio simple	68
2.1. Estimación de media y proporción	69
2.2. Selección del tamaño de la muestra para la estimación de la media y proporción	73
3. Muestreo estratificado	76
3.1. Estimadores de la media y la proporción poblacional	78
3.2. Cálculo del tamaño muestral para la estimación de la media y proporción poblacional	81
3.2.1. Afijación	81
3.2.2. Cálculo tamaño de la muestra en la estimación de la media	82
3.2.3. Cálculo del tamaño de la muestra en la estimación de la proporción	84
4. Muestreo por conglomerados	85
4.1. Estimación de la media y proporción poblacional	86
4.2. Tamaño de la muestra para la estimación de la media y proporción poblacional	90
4.3. Muestreo por conglomerados en dos etapas	91
5. Muestreo sistemático	93
6. Muestreo por cuotas	94
7. Muestreo por contraste de hipótesis estadística	95
7.1 Comparación de dos medias	96
7.2. Comparación de dos proporciones	97
8. Referencias bibliográficas	98

<i>Capítulo 3. TÉCNICAS DESCRIPTIVAS MULTIVARIANTES: ANÁLISIS FACTORIAL</i>	
EXPLORATORIO Y ANÁLISIS DE CLUSTER O CONGLOMERADOS	99
1. Análisis factorial	101
1.1. Introducción	101
1.2. Diseño y método	101
1.2.1. Diseño	101
1.2.2. Modelo matemático	102
1.2.3. Propiedades	103
1.2.4. Soluciones múltiples del modelo	104
1.3. Supuestos y limitaciones	105
1.3.1. Evaluación de la matriz de correlaciones	105
1.3.2. Limitaciones	107
1.4. Procedimiento	108
1.4.1. Pasos para la realización del análisis factorial exploratorio	108
1.4.2. Extracción de factores	109
1.4.3. Rotación de los factores	112
1.4.4. Cálculo de las puntuaciones factoriales	114
1.4.5 Interpretación de resultados	115
2. Análisis de conglomerados	118
2.1. Definición y objetivos	118
2.1.1. Definición	118
2.1.2. Objetivos	118
2.2. Diseño y método	119
2.2.1. Diseño	119
2.2.1.1. Selección de las variables	119
2.2.1.2. Elección de la medida de proximidad	120
2.2.1.3. Elección del método para la formación de conglomerados	137
2.2.1.4. Elección del número de conglomerados	137
2.3. Supuestos y limitaciones	138
2.4. Procedimiento	139
2.4.1 Métodos jerárquicos	139
2.4.1.1. Métodos aglomerativos o ascendentes	139
2.4.1.2. Métodos disociativos o descendentes	140

2.4.2. Métodos no jerárquicos	140
2.4.2.1. Método de k-medias	140
2.5. Interpretación de resultados	141
3. Un ejemplo de análisis factorial y de cluster	143
3.1. Análisis factorial	143
3.1.1. Pertinencia del análisis	144
3.1.2. Extracción de los factores	145
3.1.3. Solución rotada	147
3.1.4. Análisis factorial con R	147
3.2. Análisis de cluster	152
Referencias bibliográficas	157

Capítulo 4 TÉCNICAS EXPLICATIVAS MULTIVARIANTES: ANÁLISIS DISCRIMINANTE, ANÁLISIS DE REGRESIÓN Y ANÁLISIS DE SEGMENTACIÓN

1. Análisis discriminante	161
1.1. Definición y objetos	163
1.2. Diseño y método	164
1.2.1. Modelo matemático	164
1.2.2. Tamaño de la muestra	165
1.2.3. División de la muestra	165
1.3. Supuestos y limitaciones	166
1.3.1. Supuestos del modelo	166
1.3.2. Comprobación de los supuestos paramétricos	166
1.4. Procedimiento	167
1.4.1. Fases del análisis	167
1.4.1.1. Selección de las variables discriminantes	167
1.4.1.2. Extracción de las funciones discriminantes	170
1.4.1.3. Clasificación de los individuos	173
1.5. Interpretación de resultados	174
2. Análisis de regresión	175
2.1. Objetivos	175
2.2. Diseño y método	177
2.2.1. Diseño de investigación	177
2.3. Supuestos y limitaciones	180
2.4. Procedimiento	184
2.5. Interpretación de resultados	191
3. Análisis de segmentación. El modelo Chaid	191

3.1. Introducción	191
3.2. El análisis de segmentación. El modelo Chaid	194
3.2.1. Definiciones	194
3.2.2. Objetivos y aplicaciones del método	196
3.2.3. Relación con otros métodos multivariantes	198
3.2.4. El estadístico ji-cuadrado en la segmentación	199
3.2.5. El procedimiento Chaid	200
3.2.6. Un ejemplo de aplicación del procedimiento Chaid	203
3.2.7. Supuestos y limitaciones	207
3.2.8. Interpretación de resultados	208
4. Ejemplos de análisis discriminante, de regresión y de segmentación ..	208
4.1. Análisis discriminante.	208
4.2. Análisis de regresión.	221
4.3. Análisis de segmentación	233
5. Referencias bibliográficas	238

Capítulo 5. TÉCNICAS RELACIONADAS CON LA MEDIDA. VALIDACIÓN DE INSTRUMENTOS DE MEDIDA. TEORÍA DE RESPUESTA AL ÍTEM. DISEÑOS DE CASO ÚNICO.

1. Técnicas relacionadas con la medida.	243
2. Análisis de los ítems	243
2.1. Teoría clásica de los test (TCT).	243
2.1.1. Dificultad	244
2.1.2. Discriminación.	245
2.1.3. Análisis de distractores	248
2.1.4. Coeficiente de fiabilidad	248
2.1.5. Validez.	255
2.1.6. Dimensionalidad	257
2.2. Teoría de respuesta al ítem.	258
2.2.1. Supuestos	259
2.2.2. Modelos.	259
2.2.3. Curva característica del test	261
2.2.4. Funciones de información.	262
2.2.5. Aplicaciones de la TRI.	263
2.2.5.1. Banco de ítems.	263
2.2.5.2. Equiparación de puntuaciones	263
2.2.5.3. Funcionamiento diferencial de los ítems	264
2.2.5.4. Tests adaptativos	264

3. Diseños de caso único	265
3.1. Características de los diseños experimentales de caso único.. . . .	265
3.2. Fases en la experimentación de caso único.	266
3.3. Tipos de diseño de caso único	267
3.4. Ventajas e inconvenientes de este tipo de diseño	269
4. Un ejemplo de análisis de un instrumento de medida desde la TCT. . .	270
5. Un ejemplo de análisis de un instrumento de medida desde la TRI . . .	273
Referencias bibliográficas	283

1. EL ANÁLISIS DE DATOS CON ORDENADOR

El análisis de datos con la ayuda del ordenador resulta en la actualidad imprescindible en todo proceso de análisis de datos recogidos de investigaciones. El tratamiento de los mismos mediante técnicas estadísticas, o la representación gráfica de estos, se ve ayudado por el uso de paquetes estadísticos.

Hoy en día la proliferación de dichos paquetes, para realizar tareas concretas y programas de propósito general que abarcan un conjunto amplio de las mismas, hacen difícil la elección al investigador. No obstante, en Ciencias Sociales el uso del paquete SPSS, de propósito general, está muy difundido, aunque su alto precio lo hace costoso para el investigador novel o en el mundo académico. Como respuesta a este problema ha tomado fuerza, dentro del software libre, el paquete estadístico R.

Para realizar el análisis estadístico de los datos de una investigación, en la mayoría de los casos, se realizará el siguiente proceso:

- a) **Introducir los datos** en el programa estadístico o grabar los datos en una hoja de cálculo, archivo de texto plano (tipo notepad de Windows) o una base de datos. En todos los casos se tendrá un fichero de datos en forma matricial, donde en filas estarán los entes o casos de los que se han recogido los datos (personas, instituciones, actividades, etc.) y en columnas los atributos o variables que tienen las entidades (altura, peso, edad, nivel académico, etc.)
- b) **Filtrado y depuración de los datos** eliminando los casos anómalos, donde las variables tienen un valor fuera del rango de las mismas, o cuyo valor sea extremo (mayor o menor de 1,5 veces el rango inter-cuartílico).

- c) **Imputación o eliminación de casos ausentes** bien por falta de respuesta o dejados exprofeso en esa condición.
- d) **Seleccionar un procedimiento para calcular algún estadístico o crear un gráfico.** Para lo cual se seleccionará las variables y/o los casos para el análisis.
- e) **Ejecutar el procedimiento y elegir los resultados que se necesiten como salida.** Esta selección llevará al investigador a salvar los resultados en un fichero que permita su posterior agregación a los informes de investigación.

2. INTRODUCCIÓN AL SPSS

La investigación en ciencias sociales, y en especial la investigación cuantitativa, necesita el uso de paquetes informático-estadístico para su desarrollo. En este apartado se fijará los rudimentos mínimos para utilizar uno de los programas estadísticos más populares, el SPSS.

Statistics de SPSS es un paquete estadístico de análisis de datos utilizado en la investigación en ciencias sociales, económicas y de la salud. Contiene programas capaces de realizar desde unos simples análisis descriptivos hasta diferentes tipos de análisis multivariante. Está formado por un conjunto de módulos que conforman el paquete básico llamado SPSS Statistics donde en la pestaña de **Análisis** permite realizar:

- **Informes** (libro de códigos, cubos OLAP, informes por filas o columnas).
- **Estadísticos descriptivos** (Frecuencias, Descriptivos, Explorar, Tablas cruzadas, Gráficos P-P y Q-Q).
- **Tablas** (Personalizadas y Respuestas múltiples).
- **Comparar medias** (Medias, Prueba T para una muestra, Prueba T para muestras independientes, Prueba T para muestras relacionadas, ANOVA de un factor).
- **Modelo lineal general** (Univariante, Multivariante, Medidas repetidas, Componentes de la Varianza).
- **Modelos lineales generalizados** (Modelos lineales generalizados, Ecuaciones de estimación generalizadas).

- Modelos mixtos (Lineales, Lineales generalizados).
- Correlaciones (Bivariadas, Parciales, Distancias).
- Regresión (Modelación lineal automatizada, Lineales, Estimación curvilínea, Mínimos cuadrados parciales, Logística binaria, Logística multinomial, Ordinal, Probit, No lineal, Estimación ponderada, Mínimos cuadrados en dos fases, Escalamiento óptimo (CATREG).
- Loglineal (General, Logit, Selección de modelo).
- Clasificar (Cluster bietápico, Cluster de k-medias, Cluster jerárquico, Discriminante, Vecino más cercano).
- Reducción de dimensiones (Factor, Análisis de correspondencias, Escalamiento óptimo).
- Escala (Análisis de fiabilidad, PREFSCAL, PROXSCAL, ALSCAL).
- Pruebas no paramétricas (Una muestra, muestras independientes, muestras relacionadas).
- Previsiones — Modelos para series temporales.
- Supervivencia (Tablas de mortalidad, Kaplan-Meier, Regresión de Cox, Cox con covariable dep. del tiempo).
- Respuesta múltiple
- Muestras complejas.
- Simulación.
- Control de Calidad (Gráficos de control, Gráficos de Pareto).
- Curva COR.

2.1. Ventanas de SPSS

SPSS utiliza diferentes ventanas:

Datos	Contiene el fichero de datos. Con este editor se puede crear un fichero de datos o modificar uno ya existente. Esta ventana se abre automáticamente cuando se ejecuta el SPSS. El fichero de datos tiene extensión.sav
-------	--

Resultados	Contiene los resultados de la ejecución de los comandos del SPSS, tanto de lectura de datos, transformación de datos, etc., como de análisis de datos. Esta ventana se abre automáticamente cuando se ejecuta el SPSS. Se pueden abrir varias ventanas de salida. Los ficheros de resultados aparecen en el visor del mismo nombre y tienen extensión.spv
Sintaxis	Esta ventana se utiliza para ejecutar comandos del SPSS. Estos comandos pueden pertenecer a un fichero que ha sido creado con el SPSS, o pueden escribirse directamente, o pueden proceder de ejecutar la opción Pegar de alguna de las ventanas de diálogo. Los mencionados comandos se pueden grabar en un fichero para utilizarlos en otro momento de la sesión de SPSS. Se pueden abrir varias ventanas de sintaxis. Para ejecutar los comandos SPSS de una ventana de sintaxis, se marca con el <i>ratón</i> los comandos que se desean ejecutar y se <i>pincha</i> en el icono  (ejecutar). Los ficheros de sintaxis tienen extensión.sps
Script	Ejecuta programas en Sax Basic sobre un objeto de SPSS, por ejemplo una tabla pivote. Hay programas que contienen procesos standard almacenados en el subdirectorio SPSS y con extensión sbs. En Windows necesita tener instalado WinHlp32.exe. También puede ejecutar script en Python, siempre que esté instalado IBM® SPSS Statistics - Essentials for Python.

En todas las ventanas del SPSS están disponibles unos *menús*, una *barra de herramientas*, una *barra de estado*, y unos *botones*, algunos de los cuales son comunes a todas las ventanas y otros son específicos de cada una de ellas.

2.2. Menús

Los menús más importantes del SPSS son los siguientes:

Archivo	Se utiliza para nuevo (datos, sintaxis, resultados, script), abrir, abrir base de datos, leer datos de texto, cerrar, guardar, guardar como, mostrar información de datos, hacer caché de datos, imprimir, presentación preliminar, cambiar servidor, detener procesador, datos usados recientemente, archivos usados recientemente, salir de SPSS.
Edición	Se utiliza para <i>deshacer</i> , <i>rehacer</i> , <i>cortar</i> , <i>copiar</i> , <i>pegar</i> , <i>pegar variables</i> , <i>eliminar</i> en los ficheros de datos y procedimientos y <i>buscar</i> en el fichero de datos y <i>opciones</i> para definir las condiciones de entorno para tablas y datos.

Ver	Se utiliza para visualizar la <i>barra de estado</i> , visualizar los iconos de la <i>barra de herramientas</i> , cambiar las <i>fuentes</i> , poner o quitar la <i>cuadrícula</i> en los datos y para visualizar las <i>etiquetas de valor</i> , y finalmente <i>variables</i> (si se está en la vista de datos).
Datos	Se utiliza para realizar cambios en el fichero de datos, tales como: definir propiedades de variables, copiar propiedades de datos, definir fechas, insertar variable, insertar caso, ir a caso, ordenar casos, transponer, reestructurar, fundir archivos (añadir casos, añadir variables), agregar- es decir, crear grupos de casos para un análisis-, diseño ortogonal, segmentar archivos, seleccionar casos y ponderar casos. Estos cambios, excepto el caso de agregar que crea un archivo independiente, son temporales a no ser que se graben antes de salir del programa.
Transformar	Se utiliza para realizar cambios en las variables seleccionadas del fichero de datos y crear nuevas variables basadas en valores de otras variables que ya existen. Estos cambios son temporales y se guardan al salvar el fichero de datos. Las opciones del menú transformar son las siguientes: <i>calcular</i> , <i>semilla de aleatorización</i> , <i>contar apariciones</i> , <i>recodificar (en las mismas variables, en distintas variables)</i> , <i>categorizar variables</i> , <i>asignar rango a casos</i> , <i>recodificación automática</i> , <i>crear serie temporal</i> , <i>reemplazar valores perdidos</i> . Generar valores aleatorios.
Analizar	Se utiliza para realizar el análisis estadístico deseado. Las opciones son: Informes, Estadísticos descriptivos, Tablas, Comparar medias, Modelo lineal general, Modelos lineales generalizados, Modelos mixtos, Correlaciones, Regresión, Loglineal, Clasificar, Reducción de dimensiones, Escala, Pruebas no paramétricas, Previsiones, Supervivencia, Respuesta múltiple, Muestras complejas, Simulación, Control de Calidad, Curva COR.
Gráficos	Se utiliza para crear gráficos: Generar gráficos, de Barras, Barras 3D, Líneas, Áreas, Circular, Máximos y mínimos, Diagramas de caja y Barras de error, Pirámide de población, Dispersión, Histograma.
Utilidades	Se utiliza para visualizar información sobre el contenido del fichero de datos, del fichero de parámetros, o definir grupos de variables. Las distintas opciones disponibles son: <i>variables</i> , <i>información del archivo</i> , <i>definir conjuntos</i> , <i>usar conjuntos</i> , <i>ejecutar proceso y editor de menús</i> .
Ventana	Para minimizar las ventanas y donde van apareciendo los distintos ficheros que se ejecutan.
?	Este menú abre una ventana de ayuda que contiene información sobre el uso de cualquiera de las características del SPSS.

2.3. Barra de herramientas y barra de estado

Statistics tiene una barra de herramientas donde se reflejan mediante iconos las distintas opciones de la ventana activa. Consiste, de alguna forma, un método abreviado de acceder a los menús. La barra de estado, situada en la parte inferior de la pantalla, tiene varias secciones:

- Área del procesador: nos indica el estado del procesador
- Área de recuento: iteraciones realizadas en el fichero de datos
- Área de filtrado: si existen datos filtrados
- Área de ponderación: si los datos están ponderados
- Área de segmentación de archivos

Solo se reflejará información cuando el procesador esté realizando su actividad y cuando se ejecute la misma.

2.4. Entrada y ejecución del SPSS

Como cualquier programa Windows, SPSS tiene dos procedimientos de ejecución: 1) abriendo cualquier fichero de sintaxis o procedimientos (.sps), de datos (.sav) o de resultados (.spo o.spv); 2) al dar en el icono correspondiente al programa. Cualquiera de las dos opciones anteriores dará acceso a una ventana principal «**SPSS para Windows**», y a una ventana con un fichero de datos en blanco.

Además de estas formas de entrar en SPSS, existe dos formas de ejecutar SPSS: a) pulsando el botón **ACEPTAR** en el menú correspondiente del procedimiento estadístico, b) marcando con el ratón en el fichero de procedimientos (.sps) los comandos a ejecutar y pulsando el icono realizar.

Recordemos que cuando en el menú de SPSS damos la opción **PEGAR**, aparecerán en el fichero de sintaxis los comandos que realizan las opciones seleccionadas en el menú.

	p1	p2	p3	p4	p5	p6	p7	p8
1	5,00	2,00	28,00	3,00	19,00	29,00	28,00	30,0
2	8,00	2,00	23,00	5,00	25,00	37,00	19,00	11,0
3	14,00	3,00	30,00	2,00	39,00	38,00	30,00	20,0
4	8,00	3,00	24,00	3,00	36,00	32,00	28,00	15,0
5	3,00	4,00	38,00	2,00	30,00	42,00	30,00	22,0
6	8,00	3,00	30,00	3,00	32,00	33,00	16,00	25,0
7	12,00	2,00	27,00	5,00	26,00	37,00	36,00	26,0
8	21,00	3,00	26,00	3,00	37,00	46,00	17,00	30,0
9	13,00	2,00	22,00	5,00	27,00	35,00	18,00	25,0
10	5,00	2,00	18,00	3,00	18,00	30,00	32,00	24,0
11	12,00	4,00	30,00	4,00	22,00	34,00	24,00	19,0
12	2,00	4,00	26,00	1,00	33,00	41,00	24,00	26,0
13	7,00	4,00	24,00	2,00	25,00	40,00	20,00	40,0

Figura 1.1 Pantalla principal de SPSS.

2.5. Pasos básicos para realizar un análisis de datos con SPSS

Introducir los datos en SPSS. Se puede abrir un archivo de datos previamente grabado en SPSS, importar de una hoja de cálculo o una base de datos, leer de un archivo de texto o introducir datos directamente en el editor de datos del SPSS.

Seleccionar un procedimiento para calcular algún estadístico o crear un gráfico.

Seleccionar las variables para el análisis. Cuando elijamos un procedimiento estadístico hay que seleccionar las variables a las que vamos a aplicar el mismo.

Ejecutar el procedimiento y seleccionar los resultados que se necesite como salida. Esta selección se realizará en el *navegador de resultados*.

2.6. Matriz de datos

Para realizar un análisis estadístico es necesario disponer de una matriz de datos. Dicha matriz se puede estructurar de la siguiente forma:

	VARIABLE 1	VARIABLE 2		VARIABLE m
Caso 1				
Caso 2				
Caso 3				
•				
•				
•				
Caso n				

Es decir, es una matriz rectangular de dimensión $n \times m$, donde n (número de filas) corresponde a las unidades o elementos de análisis, y m (número de columnas) corresponde a las variables.

El proceso de construcción de esta matriz de datos se compone de:

1. **Introducir o importar los datos.** Para realizar esta labor se puede utilizar el editor de datos de SPSS o importarlos de una base de datos, hoja de cálculo, de otros programas estadísticos, o bien con un procesador de textos crear un fichero ASCII contenedor de dicha matriz de datos.

Como se ha comentado, los datos se pueden introducir directamente en el editor del SPSS. Es aconsejable definir primero las variables y después introducir los datos. Ambas operaciones se realizan desde la ventana «**Nuevo Datos**».

La ventana de datos tiene dos visiones: a) **vista de datos**, donde se muestran los valores reales o las etiquetas de valor definidas y b) **vista de variables**, con la información relativa a las mismas.

2. **Definir variables.** Para crear un archivo de datos de SPSS (.sav), se necesita además «*definir variables*». Esta labor incluye: nombrar las variables, el tipo, el formato de lectura, la etiqueta de la variable, la etiqueta de las categorías (cuando la variable es discreta) y la definición de los valores «*missing*» (ausentes).