

Bioestadística Aplicada con SPSS

Santiago Angulo Díaz-Parreño

José Miguel Cárdenas Rebollo

Anselmo Romero Limón

Alfredo Sánchez Alberca



CEU | *Ediciones*

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

Bioestadística Aplicada con SPSS

©2014, Santiago Angulo Díaz-Parreño (sangulo@ceu.es), José Miguel Cárdenas Rebollo (cardenas@ceu.es), Anselmo Romero Limón (arlimon@ceu.es) y Alfredo Sánchez Alberca (asalber@ceu.es).

©2014, Fundación Universitaria San Pablo CEU

CEU *Ediciones*

Julián Romea 18, 28003 Madrid

Correo electrónico: ceuediciones@ceu.es

www.ceuediciones.es

ISBN: 978-84-15949-22-0

Depósito Legal: M-15873-2014

Maquetación: Alfredo Sánchez Alberca

Impresión: Gráficas Vergara, S.A.

Impreso en España - Printed in Spain

Índice general

1.	Introducción a SPSS	1
1.1.	Ejercicios resueltos	1
2.	Distribuciones de Frecuencias y Gráficas	7
2.1.	Ejercicios resueltos	7
2.2.	Ejercicios propuestos	10
3.	Estadísticos Muestrales	11
3.1.	Ejercicios resueltos	11
3.2.	Ejercicios propuestos	17
4.	Regresión Lineal Simple y Correlación	21
4.1.	Ejercicios resueltos	21
4.2.	Ejercicios propuestos	27
5.	Regresión No Lineal	29
5.1.	Ejercicios resueltos	29
5.2.	Ejercicios propuestos	32
6.	Variables Aleatorias Discretas	35
6.1.	Ejercicios resueltos	35
6.2.	Ejercicios propuestos	43

7.	Variables Aleatorias Continuas	45
7.1.	Ejercicios resueltos	45
7.2.	Ejercicios propuestos	55
8.	Intervalos de Confianza para Medias y Proporciones	57
8.1.	Ejercicios resueltos	57
8.2.	Ejercicios propuestos	61
9.	Intervalos de Confianza para Comparar Poblaciones	63
9.1.	Ejercicios resueltos	63
9.2.	Ejercicios propuestos	67
10.	Contraste de Hipótesis	69
10.1.	Ejercicios resueltos	69
10.2.	Ejercicios propuestos	73
11.	Análisis de la Varianza de un Factor	75
11.1.	Ejercicios resueltos	75
11.2.	Ejercicios propuestos	79
12.	ANOVA de Dos o Más Factores, ANOVA de Medidas Re- petidas y ANCOVA	81
12.1.	Ejercicios resueltos	81
12.2.	Ejercicios propuestos	100
13.	Contrastes de Hipótesis No Paramétricos	105
13.1.	Ejercicios resueltos	105
13.2.	Ejercicios propuestos	113
14.	Contrastes Basados en el Estadístico Chi-cuadrado	117
14.1.	Ejercicios Resueltos	117
14.2.	Ejercicios propuestos	122
15.	Análisis de Concordancia	125
15.1.	Ejercicios resueltos	125

Introducción a SPSS

1.1. Ejercicios resueltos

1. Introducir en la matriz de datos los datos de la siguiente muestra y guardarlos en un fichero con el nombre `datos_colesterol.sav`.

Nombre	Sexo	Peso	Altura	Colesterol
José Luis Martínez Izquierdo	H	85	179	182
Rosa Díaz Díaz	M	65	173	232
Javier García Sánchez	H	71	181	191
Carmen López Pinzón	M	65	170	200
Marisa López Collado	M	51	158	148
Antonio Ruiz Cruz	H	66	174	249
Antonio Fernández Ocaña	H	62	172	276
Pilar Martín González	M	60	166	213
Pedro Gálvez Tenorio	H	90	194	241
Santiago Reillo Manzano	H	75	185	280
Macarena Álvarez Luna	M	55	162	262
José María de la Guía Sanz	H	78	187	198
Miguel Angel Cuadrado Gutiérrez	H	109	198	210
Carolina Rubio Moreno	M	61	177	194

i

- a) En la ventana de Vista de variables, crear las variables Nombre, Sexo, Peso, Altura y Colesterol e introducir los datos anteriores, siguiendo las indicaciones del apartado ??.

b) Una vez introducidos los datos, se guardan en un fichero de nombre `datos_colesterol`. siguiendo lo indicado en el apartado ??.

2. Sobre la matriz de datos del ejercicio anterior realizar las siguientes operaciones:

a) Insertar detrás de la variable Nombre una nueva variable Edad con las edades de todos los individuos de la muestra.

Nombre	Edad
José Luis Martínez Izquierdo	18
Rosa Díaz Díaz	32
Javier García Sánchez	24
Carmen López Pinzón	35
Marisa López Collado	46
Antonio Ruiz Cruz	68
Antonio Fernández Ocaña	51
Pilar Martín González	22
Pedro Gálvez Tenorio	35
Santiago Reillo Manzano	46
Macarena Álvarez Luna	53
José María de la Guía Sanz	58
Miguel Angel Cuadrado Gutiérrez	27
Carolina Rubio Moreno	20

i

- 1) En la Vista de variables seleccionar la fila correspondiente a la variable Sexo haciendo click con el ratón sobre la cabecera de la misma y a continuación seleccionar el menú Edición ▶ Insertar variable, con lo que aparece una nueva fila entre las variables Nombre y Sexo.
- 2) En la nueva fila definir la variable Edad.
- 3) En la Vista de datos rellenar los datos de la columna correspondiente a la Edad.

b) Insertar entre los individuos 4º y 5º los datos correspondientes al siguiente individuo

Nombre: Cristóbal Campos Ruiz.

Edad: 44 años.

Sexo: Hombre.
Peso: 70 Kg.
Altura: 178 cm.
Colesterol: 220 mg/dl.

- i*
- 1) Seleccionar la fila correspondiente al 5º individuo, haciendo click con el ratón sobre la cabecera de la misma y a continuación seleccionar el menú Edición ▶ Insertar caso, con lo que aparece una nueva fila entre las correspondientes a los individuos 4º y 5º.
 - 2) Introducir en la nueva fila los datos que se indican.

- c) Cambiar el valor de la variable Peso de Macarena Álvarez Luna por 58.

- i*
- Hacer click con el ratón en la casilla cuyo contenido se desea modificar, escribir 58 y pulsar Intro.

- d) Transformar la variable Altura para que aparezca expresada en metros.

- i*
- 1) Seleccionar el menú Transformar ▶ Calcular variable.
 - 2) En la ventana de transformación de datos introducir el nombre `Altura_metros` en el cuadro Variable de destino.
 - 3) Introducir la expresión `Altura/100` en el cuadro Expresión numérica.
 - 4) Hacer click sobre el botón Aceptar.

- e) Recodificar la variable peso en las siguientes cuatro categorías, teniendo en cuenta el sexo:

Categoría	Hombres	Mujeres
Bajo	≤ 70	≤ 60
Medio	(70,85]	(60,70]
Alto	(85,100]	(70,80]
Muy Alto	> 100	> 80

i

- 1) Seleccionar el menú Transformar ▶ Recodificar en distintas variables.
- 2) En la ventana de recodificación de datos seleccionar la variable Peso y hacer click sobre el botón con una flecha que aparece al lado.
- 3) Escribir el nombre de la variable recodificada Categoría_Peso en el cuadro Nombre de la Variable de resultado y hacer click sobre el botón Cambiar.
- 4) Hacer click en el botón Valores antiguos y nuevos para abrir la ventana de definición de reglas de recodificación.
- 5) Para definir las reglas de recodificación de los hombres,
 - I. Seleccionar la opción Rango INFERIOR hasta valor del cuadro Valor antiguo e introducir 70 en el cuadro correspondiente. Introducir 1 en el cuadro de la opción Valor del cuadro Valor nuevo y hacer click en el botón Añadir.
 - II. Seleccionar la opción Rango del cuadro Valor antiguo e introducir 70 en el cuadro correspondiente y 85 en el cuadro hasta. Introducir 2 en el cuadro de la opción Valor del cuadro Valor nuevo y hacer click en el botón Añadir.
 - III. Seleccionar la opción Rango del cuadro Valor antiguo e introducir 85 en el cuadro correspondiente y 100 en el cuadro hasta. Introducir 3 en el cuadro de la opción Valor del cuadro Valor nuevo y hacer click en el botón Añadir.
 - IV. Seleccionar la opción Rango valor hasta SUPERIOR del cuadro Valor antiguo e introducir 100 en el cuadro correspondiente. Introducir 4 en el cuadro de la opción Valor del cuadro Valor nuevo y hacer click en el botón Añadir.
- 6) Hacer click en el botón Continuar para cerrar la ventana.
- 7) Hacer click en el botón Si la opción... para abrir la ventana de definición de condiciones.
- 8) Seleccionar la opción Incluir si el caso satisface la condición e introducir la condición Sexo=' 'H' ' en el cuadro correspondiente.
- 9) Hacer click en el botón Continuar para cerrar la ventana.

- 10) Hacer click en el botón Aceptar.
- 11) Repetir los mismos pasos para establecer las reglas de recodificación de las mujeres.
- 12) En Vista de variables hacer click en Valores de Categoría_Peso, y en Etiquetas de valor ir asignando a los valores 1, 2, 3 y 4 las etiquetas Bajo, Medio, Alto y Muy alto respectivamente, haciendo click en el botón Añadir después de cada asignación, y una vez terminado hacer click en el Aceptar.

f) Volver a guardar los cambios en el fichero anterior y salir del programa.

i

- 1) Seleccionar el menú Archivo ▶ Guardar.
- 2) Seleccionar el menú Archivo ▶ Salir.

2

Distribuciones de Frecuencias y Gráficas

2.1. Ejercicios resueltos

1. Se realizó una encuesta a 40 personas de más de 70 años sobre el número de medicamentos distintos que tomaban habitualmente. El resultado de dicha encuesta fue el siguiente:

3 - 1 - 2 - 2 - 0 - 1 - 4 - 2 - 3 - 5 - 1 - 3 - 2 - 3 - 1 - 4 - 2 - 4 - 3 - 2
3 - 5 - 0 - 1 - 2 - 0 - 2 - 3 - 0 - 1 - 1 - 5 - 3 - 4 - 2 - 3 - 0 - 1 - 2 - 3

Se pide:

- a) Crear la variable medicamentos e introducir los datos.
- b) Construir la tabla de frecuencias.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Frecuencias.
- 2) Seleccionar la variable medicamentos en el campo Variables del cuadro de diálogo.
- 3) Activar la opción Mostrar tabla de frecuencias y hacer click en el botón Aceptar.

- c) Dibujar el diagrama de barras de las frecuencias absolutas.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Barras ▶ Definir.

- 2) Seleccionar la variable medicamentos en el campo Eje de categorías del cuadro de diálogo, seleccionar la opción N° de casos y hacer click en el botón Aceptar.

d) Dibujar el polígono de frecuencias absolutas.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Líneas ▶ Definir.
- 2) Seleccionar la variable medicamentos en el campo Eje de categorías del cuadro de diálogo, seleccionar la opción N° de casos y hacer click en el botón Aceptar.

e) Dibujar el diagrama de barras de las frecuencias relativas acumuladas.

i

Repetir los mismos pasos del apartado c) pero seleccionando esta vez la opción % acum.

f) Dibujar el diagrama de sectores.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Sectores ▶ Definir.
- 2) Seleccionar la variable medicamentos en el campo Definir sectores por del cuadro de diálogo y hacer click en el botón Aceptar.

2. En un hospital se realizó un estudio sobre el número de personas que ingresaron en urgencias en el mes de noviembre. Los datos observados fueron:

15 – 23 – 12 – 10 – 28 – 7 – 12 – 17 – 20 – 21 – 18 – 13 – 11 – 12 – 26
29 – 6 – 16 – 39 – 22 – 14 – 17 – 21 – 28 – 9 – 16 – 13 – 11 – 16 – 20

Se pide:

- a) Crear la variable urgencias e introducir los datos.
- b) Dibujar el histograma de las frecuencias absolutas agrupando en 5 clases desde el 0 hasta el 40.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Histograma.
- 2) Seleccionar la variable urgencias en el campo Variable del cuadro de diálogo y hacer click en el botón Aceptar.
- 3) Editar el histograma haciendo doble click sobre él.
- 4) En el editor de gráficos hacer click con el botón derecho del ratón en la zona del histograma, el cual quedará rodeado de una línea amarilla y hacer click en la ventana emergente sobre Ventana Propiedades.
- 5) Seleccionar la opción Clases de punto, en Eje X elegir Personalizado y en Número de Intervalos poner 5.
- 6) Hacer click sobre el botón Aplicar, después sobre el botón Cerrar de la ventana de propiedades y cerrar el editor de gráficos.

c) Dibujar el diagrama de caja. ¿Existe algún dato atípico?

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Diagramas de caja.
- 2) Seleccionar la opción Resúmenes para distintas variables y hacer click en el botón Definir.
- 3) Seleccionar la variable urgencias en el campo Las cajas representan del cuadro de diálogo y hacer click sobre el botón Aceptar.

d) En el caso de que exista algún dato atípico, eliminarlo y dibujar el histograma de frecuencias absolutas, de forma que aparezcan clases de amplitud 5, comenzando en el 5 y terminando en el 30.

i

- 1) Identificar el caso que corresponde al dato atípico y eliminarlo en el editor de datos.
- 2) Repetir los pasos del apartado b) para dibujar el histograma.

2.2. Ejercicios propuestos

1. El número de lesiones padecidas durante una temporada por cada jugador de un equipo de fútbol fue el siguiente:

0 – 1 – 2 – 1 – 3 – 0 – 1 – 0 – 1 – 2 – 0 – 1
1 – 1 – 2 – 0 – 1 – 3 – 2 – 1 – 2 – 1 – 0 – 1

Se pide:

- a) Crear la variable lesiones e introducir los datos.
 - b) Construir la tabla de frecuencias.
 - c) Dibujar el diagrama de barras de las frecuencias relativas acumuladas.
 - d) Dibujar el polígono de frecuencias de las frecuencias absolutas acumuladas.
 - e) Dibujar el diagrama de sectores.
2. Para realizar un estudio sobre la estatura de los estudiantes universitarios, seleccionamos, mediante un proceso de muestreo aleatorio, una muestra de 30 estudiantes, obteniendo los siguientes resultados (medidos en centímetros):

179, 173, 181, 170, 158, 174, 172, 166, 194, 185,
162, 187, 198, 177, 178, 165, 154, 188, 166, 171,
175, 182, 167, 169, 172, 186, 172, 176, 168, 187.

Se pide:

- a) Crear la variable estatura e introducir los datos.
- b) Dibujar el histograma de las frecuencias absolutas agrupando desde 150 a 200 en clases de amplitud 10.
- c) Dibujar el diagrama de caja. ¿Existe algún dato atípico?.

Estadísticos Muestrales

3.1. Ejercicios resueltos

1. Se realizó una encuesta a 40 personas de más de 70 años sobre el número de medicamentos distintos que tomaban habitualmente. El resultado de dicha encuesta fue el siguiente:

3 - 1 - 2 - 2 - 0 - 1 - 4 - 2 - 3 - 5 - 1 - 3 - 2 - 3 - 1 - 4 - 2 - 4 - 3 - 2
3 - 5 - 0 - 1 - 2 - 0 - 2 - 3 - 0 - 1 - 1 - 5 - 3 - 4 - 2 - 3 - 0 - 1 - 2 - 3

Se pide:

- a) Crear la variable medicamentos e introducir los datos. Si ya se tienen los datos, simplemente recuperarlos.
- b) Calcular la media aritmética, mediana, moda, varianza y desviación típica de dicha variable. Interpretar los estadísticos.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Frecuencias.
- 2) Seleccionar la variable medicamentos en el campo Variables del cuadro de diálogo.
- 3) Hacer click sobre el botón Estadísticos. Para seleccionar únicamente los estadísticos que nos piden, marcar las casillas correspondientes a dichos estadísticos y hacer click sobre los botones Continuar y Aceptar.

- c) Calcular el coeficiente de asimetría y el de curtosis e interpretar los resultados

i Seguir los mismos pasos del apartado anterior, seleccionando ahora los estadísticos que se piden.

d) Calcular los cuartiles.

i Seguir los mismos pasos de los apartados anteriores, activando la opción Cuartiles.

2. En un grupo de 20 alumnos, las calificaciones obtenidas en Matemáticas fueron:

SS - AP - SS - AP - AP - NT - NT - AP - SB - SS
SB - SS - AP - AP - NT - AP - SS - NT - SS - NT

Se pide:

- a) Crear la variable calificaciones e introducir los datos.
- b) Recodificar esta variable, asignando 2,5 al SS, 5,5 al AP, 7,5 al NT y 9,5 al SB.

i

- 1) Seleccionar el menú Transformar ▶ Recodificar en distintas variables.
- 2) Seleccionar la variable calificaciones y hacer click sobre el botón con la flecha del cuadro de diálogo para llevarla a Variable de entrada.
- 3) Introducir el nombre de la nueva variable en el campo Nombre del cuadro de diálogo y hacer click en el botón Cambiar.
- 4) Hacer click en el botón Valores antiguos y nuevos e introducir las reglas de recodificación y hacer click sobre los botones Continuar y Aceptar.

c) Calcular la moda y la mediana.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Frecuencias.
- 2) Seleccionar la variable recodificada en el campo Variables del cuadro de diálogo.

- 3) Hacer click sobre el botón Estadísticos, seleccionar los estadísticos que se piden y hacer click sobre los botones Continuar y Aceptar.

3. En un estudio sobre la estatura de los estudiantes universitarios, se ha seleccionado una muestra de 30 estudiantes, obteniendo los siguientes resultados (medidos en centímetros):

179, 173, 181, 170, 158, 174, 172, 166, 194, 185,
162, 187, 198, 177, 178, 165, 154, 188, 166, 171,
175, 182, 167, 169, 172, 186, 172, 176, 168, 187.

Se pide:

- a) Crear la variable estatura e introducir los datos.
- b) Obtener un resumen de estadísticos en el que se muestren la media aritmética, mediana, moda, varianza, desviación típica y cuartiles. Interpretar los estadísticos.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Frecuencias.
- 2) Seleccionar la variable estatura en el campo Variables del cuadro de diálogo.
- 3) Hacer click sobre el botón Estadísticos, seleccionar los estadísticos que se piden y hacer click sobre los botones Continuar y Aceptar.

- c) Calcular el tercer decil e interpretarlo.

i

Seguir los mismos pasos de los apartados anteriores, activando la opción Percentiles e introduciendo el percentil deseado en el correspondiente cuadro de texto.

- d) Con los datos obtenidos en apartados anteriores, calcular el coeficiente de variación de Pearson y el rango intercuartílico, e interpretar los resultados.

4. En un estudio sobre la estatura de los estudiantes universitarios, se ha seleccionado una muestra de 30 estudiantes, obteniendo los siguientes resultados (medidos en centímetros):

x_i	Marca	n_i	f_i	N_i	F_i
[150,160)	155	2	0,07	2	0,07
[160,170)	165	7	0,23	9	0,3
[170,180)	175	12	0,4	21	0,7
[180,190)	185	7	0,23	28	0,93
[190,200)	195	2	0,07	30	1

Se pide:

- Crear la variable estatura, en la que vamos a introducir las marcas de clase y crear la variable frecuencias, en la que se introducirán las frecuencias absolutas.
- Ponderar los casos de la variable estatura con las frecuencias de la variable frecuencias

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.
- 2) Activar la opción Ponderar casos mediante, seleccionar la variable frecuencias y hacer click sobre el botón Aceptar.

- Obtener un resumen de estadísticos en el que se muestren la media aritmética, mediana, moda, varianza, desviación típica y cuartiles.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Frecuencias.
- 2) Seleccionar la variable estatura en el campo Variables del cuadro de diálogo.
- 3) Hacer click sobre el botón Estadísticos, seleccionar los estadísticos que se piden y hacer click sobre los botones Continuar y Aceptar.

¿Existen diferencias entre estos estadísticos y los del ejercicio anterior? ¿A qué se deben?

- Calcular el tercer decil.

i Seguir los mismos pasos de los apartados anteriores, activando la opción Percentiles e introduciendo el percentil correspondiente en el cuadro de texto.

e) Calcular el percentil 62.

i Seguir los pasos de los apartados anteriores seleccionando el estadístico deseado.

f) Con los datos obtenidos en apartados anteriores, calcular el coeficiente de variación de Pearson y el rango intercuartílico, e interpretar los resultados.

5. En un hospital se ha tomado nota de la concentración de anticuerpos de inmunoglobulina M en el suero sanguíneo de personas sanas, y han resultado los siguientes datos por litro. Entre paréntesis figura el sexo de la persona (H para hombre y M para mujer).

(H) 1,071	(H) 0,955	(H) 0,73	(M) 0,908	(M) 0,859
(H) 0,927	(M) 0,962	(M) 1,543	(H) 1,094	(M) 0,847
(H) 1,214	(M) 1,456	(M) 1,516	(M) 1,002	(M) 0,799
(M) 0,881	(M) 1,096	(M) 0,964	(H) 0,973	(H) 1,222
(H) 0,887	(H) 1,022	(M) 0,881	(M) 1,42	(M) 1,205

Se pide:

- Crear las variables sexo e inmunoglobulina e introducir los datos.
- Dividir el archivo, usando como variable de segmentación la variable sexo

i

- 1) Seleccionar el menú Datos ▶ Dividir archivo...
- 2) Seleccionar la opción Comparar los grupos u Organizar los resultados por grupos (Se diferencian en la forma de presentar los resultados).
- 3) Seleccionar la variable sexo en el campo Grupos basados en del cuadro de diálogo y hacer clic sobre el botón Aceptar.

c) Calcular la media aritmética, la moda y la mediana de la inmunoglobulina, tanto en hombres como en mujeres.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Frecuencias.
- 2) Seleccionar la variable inmunoglobulina en el campo Variables del cuadro de diálogo.
- 3) Hacer click sobre el botón Estadísticos, seleccionar los estadísticos que se piden, hacer click sobre el botón Continuar y finalmente hacer click sobre el botón Aceptar.

d) Calcular la varianza y la desviación típica tanto en hombres como en mujeres.

i

Seguir los mismos pasos del apartado anterior.

e) ¿En qué población es más representativa la media, en la de hombres o en la de mujeres?

i

Para responder a la pregunta será necesario calcular el coeficiente de variación.

6. Se reciben dos lotes de un determinado fármaco, fabricados con dos modelos de maquinaria diferentes, además uno proviene de Madrid y el otro de Valencia. Se toma una muestra de diez cajas, cinco de cada lote y se mide la concentración del principio activo, obteniendo los siguientes resultados:

Modelo Maquinaria	A	B	A	B	A
Procedencia	Mad	Mad	Val	Mad	Val
Concentración (mg/mm ³)	17,6	19,2	21,3	15,1	17,6

Modelo Maquinaria	B	A	B	B	A
Procedencia	Val	Mad	Val	Mad	Val
Concentración (mg/mm ³)	18,9	16,2	18,3	19	16,4

Se pide:

a) Crear las variables procedencia, maquinaria y concentración e introducir los datos.

- b) Calcular la media aritmética, desviación típica, coeficiente de asimetría y curtosis de la concentración según el lugar de procedencia.

i

- 1) Seleccionar el menú Datos ▶ Dividir archivo...
- 2) Seleccionar la opción Comparar los grupos u Organizar los resultados por grupos.
- 3) Seleccionar la variable procedencia en el campo Grupos basados en del cuadro de diálogo y hacer clic sobre el botón Aceptar.
- 4) Seguir los mismos pasos del ejercicio anterior para seleccionar los estadísticos.

- c) Dibujar el diagrama de cajas de la concentración de principio activo, según la maquinaria de fabricación.

i

- 1) Seleccionar el menú Datos ▶ Dividir archivo...
- 2) Seleccionar la opción Comparar los grupos u Organizar los resultados por grupos.
- 3) Seleccionar la variable maquinaria en el campo Grupos basados en del cuadro de diálogo y hacer clic sobre el botón Aceptar.
- 4) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Diagramas de caja....
- 5) Seleccionar la opción Resúmenes para distintas variables y hacer click en el botón Definir.
- 6) Seleccionar la variable concentracion en el campo Las cajas representan del cuadro de diálogo y hacer click sobre el botón Aceptar.

3.2. Ejercicios propuestos

1. El número de lesiones padecidas durante una temporada por cada jugador de un equipo de fútbol fue el siguiente:

0	1	2	1	3	0	1	0	1	2	0	1
1	1	2	0	1	3	2	1	2	1	0	1

Se pide:

- a) Crear la variable lesiones e introducir los datos. Si ya se tienen los datos, simplemente recuperarlos.
 - b) Calcular: media aritmética, mediana, moda, varianza y desviación típica.
 - c) Calcular los coeficientes de asimetría y curtosis e interpretar los resultados.
 - d) Calcular el cuarto y el octavo decil.
2. En una encuesta sobre la intención de voto en unas elecciones en las que se presentaban tres partidos *A*, *B* y *C*, se preguntó a 30 personas y se obtuvieron las siguientes respuestas:

A - B - VB - A - C - A - VB - C - A - A - B - B - A - B - B
B - A - A - C - B - B - B - A - VB - A - B - VB - A - B - B

Se pide:

- a) Crear la variable voto e introducir los datos.
 - b) Calcular aquellos estadísticos que sea posible para este atributo
3. La siguiente tabla expresa la distribución de las puntuaciones obtenidas por un grupo de alumnos.

0-20	20-40	40-60	60-80	80-100
15	19	13	11	8

Se pide:

- a) Calcular la media aritmética, la mediana y la moda.
 - b) Calcular el percentil 92.
 - c) Calcular la desviación típica.
 - d) Calcular el coeficiente de asimetría.
 - e) Calcular del coeficiente de curtosis.
4. En un estudio de población se tomó una muestra de 27 personas, y se les preguntó por su edad y estado civil (Soltero S, Casado C, Divorciado D y Viudo V) obteniendo los siguientes resultados:

Estado Civil	C	S	S	V	C	C	D	S	S
Edad	62	31	45	80	39	62	31	45	75

Estado Civil	S	S	C	S	D	V	D	S	V
Edad	21	38	59	62	65	68	59	62	65

Estado Civil	C	V	C	D	D	V	V	S	V
Edad	21	40	62	59	65	78	69	31	75

Se pide:

- a) Crear las variables adecuadas e introducir los datos.
 - b) Calcular la media y la desviación típica de la edad según el estado civil.
 - c) Dibujar el diagrama de barras para las frecuencias absolutas de la edad según el estado civil.
5. En un estudio se ha medido la presión arterial diastólica (en mmHg) de 24 individuos. Además se les ha preguntado si fuman y beben:

Fumador	si	no	si	si	si	no	no	si	no	si	no	si
Bebedor	no	no	si	si	no	no	si	si	no	si	no	si
Presión	80	92	75	56	89	93	101	67	89	63	98	58

Fumador	si	no	no	si	no	no	no	si	no	si	no	si
Bebedor	si	no	si	si	no	no	si	si	si	no	si	no
Presión	71	52	98	104	57	89	70	93	69	82	70	49

Se pide :

- a) Crear las variables correspondientes e introducir los datos.
- b) Calcular la media aritmética, desviación típica, coeficiente de asimetría y curtosis de la tensión arterial por grupos dependiendo de si beben y/o fuman.
- c) Dibujar el histograma para las frecuencias absolutas de la tensión arterial según lo grupos hechos anteriormente.

4

Regresión Lineal Simple y Correlación

4.1. Ejercicios resueltos

1. Se han medido dos variables A y B en 10 individuos obteniendo los siguientes resultados:

A	0	1	2	3	4	5	6	7	8	9
B	2	5	8	11	14	17	20	23	26	29

Se pide:

- a) Crear las variables A y B e introducir estos datos.
- b) Dibujar el diagrama de dispersión correspondiente.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos..., elegir la opción Dispersión simple y hacer click sobre el botón Definir.
- 2) Seleccionar la variable B en el campo Eje Y del cuadro de diálogo.
- 3) Seleccionar la variable A en el campo Eje X del cuadro de diálogo y hacer click sobre el botón Aceptar.

En vista del diagrama, ¿qué tipo de modelo crees que explicará mejor la relación entre B y A ?

- c) Calcular la recta de regresión de B sobre A .

i

- 1) Seleccionar el menú *Analizar* ▶ *Regresión* ▶ *Lineales...*
- 2) Seleccionar la variable *B* en el campo *Dependientes* del cuadro de diálogo.
- 3) Seleccionar la variable *A* en el campo *Independientes* del cuadro de diálogo y hacer click sobre el botón *Aceptar*.
- 4) Para escribir la ecuación de la recta, observaremos en la ventana de resultados obtenida, la tabla denominada *Coeficientes*, y en la columna *B* de los *Coeficientes no estandarizados*, encontramos en la primera fila la constante de la recta y en la segunda la pendiente.

d) Dibujar dicha recta sobre el diagrama de dispersión.

i

- 1) Editar el gráfico realizado anteriormente haciendo un doble click sobre él.
- 2) Seleccionar los puntos haciendo click sobre alguno de ellos.
- 3) Seleccionar el menú *Elementos* ▶ *Linea de ajuste total* (También se podría usar en lugar del menu, la barra de herramientas)
- 4) Cerrar la ventana *Propiedades*.
- 5) Cerrar el editor de gráficos, cerrando la ventana.

e) Calcular la recta de regresión de *A* sobre *B* y dibujarla sobre el correspondiente diagrama de dispersión.

i

Repetir los pasos de los apartados anteriores pero escogiendo como variable *Dependiente* la variable *A*, y como variable *Independiente* la variable *B*.

f) ¿Son grandes los residuos? Comentar los resultados.

2. En una licenciatura se quiere estudiar la relación entre el número medio de horas de estudio diarias y el número de asignaturas suspensas. Para

ello se obtuvo la siguiente muestra:

Horas	Suspensos	Horas	Suspensos	Horas	Suspensos
3,5	1	2,2	2	1,3	4
0,6	5	3,3	0	3,1	0
2,8	1	1,7	3	2,3	2
2,5	3	1,1	3	3,2	2
2,6	1	2,0	3	0,9	4
3,9	0	3,5	0	1,7	2
1,5	3	2,1	2	0,2	5
0,7	3	1,8	2	2,9	1
3,6	1	1,1	4	1,0	3
3,7	1	0,7	4	2,3	2

Se pide:

- Crear las variables horas y suspensos e introducir estos datos.
- Calcular la recta de regresión de suspensos sobre horas y dibujarla.

i

- Seleccionar el menú Analizar ▶ Regresión ▶ Lineales...
- Seleccionar la variable suspensos en el campo Dependientes del cuadro de diálogo.
- Seleccionar la variable horas en el campo Independientes del cuadro de diálogo y hacer click sobre el botón Aceptar.
- Para escribir la ecuación de la recta, observaremos en la ventana de resultados obtenida, la tabla denominada Coeficientes, y en la columna B de los Coeficientes no estandarizados, encontramos en la primera fila la constante de la recta y en la segunda la pendiente.
- Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos..., elegir la opción Dispersión simple y hacer click sobre el botón Definir.
- Seleccionar la variable suspensos en el campo Eje Y del cuadro de diálogo.
- Seleccionar la variable horas en el campo Eje X del cuadro de diálogo y hacer click sobre el botón Aceptar.

- 8) Editar el gráfico realizado haciendo un doble click sobre él.
- 9) Seleccionar los puntos haciendo click sobre alguno de ellos.
- 10) Seleccionar el menú Elementos ▶ Línea de ajuste total (También se podría usar en lugar del menú, la barra de herramientas)
- 11) Cerrar la ventana Propiedades.
- 12) Cerrar el editor de gráficos, cerrando la ventana.

c) Indicar el coeficiente de regresión de suspensos sobre horas. ¿Cómo lo interpretarías?

i El coeficiente de regresión es la pendiente de la recta de regresión, que este caso vale $-1,23$ e indica que por cada hora de estudio adicional se obtienen $1,23$ suspensos menos.

d) La relación lineal entre estas dos variables, ¿es mejor o peor que la del ejercicio anterior? Comentar los resultados a partir las gráficas de las rectas de regresión y sus residuos.

i La relación lineal entre estas dos variables es peor que la del ejercicio anterior, pues en este caso hay residuos.

e) Calcular los coeficientes de correlación y de determinación lineal. ¿Es un buen modelo la recta de regresión? ¿Qué porcentaje de la variabilidad del número de suspensos está explicada por el modelo?

i Observaremos en la ventana de resultados obtenida la tabla denominada Resumen del modelo, y en ella encontramos los valores del coeficiente de correlación lineal R y del coeficiente de determinación lineal R^2 .

f) Utilizar la recta de regresión para predecir el número de suspensos correspondiente a 3 horas de estudio diarias. ¿Es fiable esta predicción?

i

- 1) Crear una nueva variable valores e introducir los valores de las horas de estudio para los que queremos predecir.
- 2) Seleccionar el menú Transformar ▶ Calcular variable...
- 3) Introducir el nombre de la nueva variable prediccion en el campo Variable de destino del cuadro de diálogo.
- 4) Introducir la ecuación de la recta en el campo Expresión numérica, utilizando los coeficientes calculados anteriormente y la variable valores y hacer click sobre el botón Aceptar.

- g) Según el modelo lineal, ¿cuántas horas diarias tendrá que estudiar como mínimo un alumno si quiere aprobarlo todo?

i

Seguir los mismos pasos de los apartados anteriores, pero escogiendo como variable dependiente horas, y como independiente suspensos.

3. Después de tomar un litro de vino se ha medido la concentración de alcohol en la sangre en distintos instantes, obteniendo:

Tiempo después (minutos)	30	60	90	120	150	180	210
Concentración (gramos/litro)	1,6	1,7	1,5	1,1	0,7	0,2	2,1

Se pide:

- a) Crear las variables tiempo y alcohol e introducir estos datos.
- b) Calcular el coeficiente de correlación lineal e interpretarlo.

i

- 1) Seleccionar el menú Analizar ▶ Correlaciones ▶ Bivariadas...
- 2) Seleccionar ambas variables en el campo Variables del cuadro de diálogo y hacer click sobre el botón Aceptar.

- c) Dibujar el diagrama de dispersión junto con la recta ajustada correspondiente a alcohol sobre tiempo. ¿Existe algún individuo con un residuo demasiado grande? Si es así, eliminar dicho individuo de la muestra y volver a calcular el coeficiente de correlación. ¿Ha mejorado el modelo?

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos..., elegir la opción Dispersión simple y hacer click sobre el botón Definir.
- 2) Seleccionar la variable alcohol en el campo Eje Y del cuadro de diálogo.
- 3) Seleccionar la variable tiempo en el campo Eje X del cuadro de diálogo y hacer click sobre el botón Aceptar.
- 4) Editar el gráfico realizado anteriormente haciendo un doble click sobre él.
- 5) Seleccionar los puntos haciendo click sobre alguno de ellos.
- 6) Seleccionar el menú Elementos ▶ Línea de ajuste total (También se podría usar en lugar del menú, la barra de herramientas)
- 7) Cerrar la ventana Propiedades.
- 8) Cerrar el editor de gráficos, cerrando la ventana.
- 9) Si existe algún individuo con un residuo demasiado grande, ir a la ventana del Editor de datos, y eliminarlo.
- 10) Repetir los pasos del apartado anterior.

d) Si la concentración máxima de alcohol en la sangre que permite la ley para poder conducir es 0,5 g/l, ¿cuánto tiempo habrá que esperar después de tomarse un litro de vino para poder conducir sin infringir la ley? ¿Es fiable esta predicción?

i

- 1) Seleccionar el menú Analizar ▶ Regresión ▶ Lineales....
- 2) Seleccionar la variable tiempo en el campo Dependientes del cuadro de diálogo.
- 3) Seleccionar la variable alcohol en el campo Independientes del cuadro de diálogo y hacer click sobre el botón Aceptar.
- 4) Para escribir la ecuación de la recta, observaremos en la ventana de resultados obtenida, la tabla denominada Coeficientes, y en la columna B de los Coeficientes no estandarizados, encontramos en la primera fila la constante de la recta y en la segunda la pendiente.

- 5) Crear una nueva variable valores e introducir los valores que queremos estudiar.
- 6) Seleccionar el menú Transformar ▶ Calcular variable...
- 7) Introducir el nombre de la nueva variable prediccion en el campo Variable de destino del cuadro de diálogo.
- 8) Introducir la ecuación de la recta en el campo Expresión numérica, utilizando los coeficientes citados anteriormente y la variable valores y hacer click sobre el botón Aceptar.

4.2. Ejercicios propuestos

1. Se determina la pérdida de actividad que experimenta un medicamento desde el momento de su fabricación a lo largo del tiempo, obteniéndose el siguiente resultado:

Tiempo (en años)	1	2	3	4	5
Actividad restante (%)	96	84	70	58	52

Se desea calcular:

- a) La relación fundamental (recta de regresión) entre actividad restante y tiempo transcurrido.
 - b) ¿En qué porcentaje disminuye la actividad cada año que pasa?
 - c) ¿Cuándo tiempo debe pasar para que el fármaco tenga una actividad del 80%? ¿Cuándo será nula la actividad? ¿Son igualmente fiables estas predicciones?
2. Al realizar un estudio sobre la dosificación de un cierto medicamento, se trataron 6 pacientes con dosis diarias de 2 mg, 7 pacientes con 3 mg y otros 7 pacientes con 4 mg. De los pacientes tratados con 2 mg, 2 curaron al cabo de 5 días, y 4 al cabo de 6 días. De los pacientes tratados con 3 mg diarios, 2 curaron al cabo de 3 días, 4 al cabo de 5 días y 1 al cabo de 6 días. Y de los pacientes tratados con 4 mg diarios, 5 curaron al cabo de 3 días y 2 al cabo de 5 días. Se pide:
 - a) Calcular la recta de regresión del tiempo de curación con respecto a la dosis suministrada.

- b)* Calcular los coeficientes de regresión. Interpretar los resultados.
- c)* Determinar el tiempo esperado de curación para una dosis de 5 mg diarios. ¿Es fiable esta predicción?
- d)* ¿Qué dosis debe aplicarse si queremos que el paciente tarde 4 días en curarse? ¿Es fiable la predicción?

5

Regresión No Lineal

5.1. Ejercicios resueltos

1. En un experimento se ha medido el número de bacterias por unidad de volumen en un cultivo, cada hora transcurrida, obteniendo los siguientes resultados:

Horas	0	1	2	3	4	5	6	7	8
Nº Bacterias	25	32	47	65	92	132	190	275	362

Se pide:

- a) Crear las variables horas y bacterias e introducir estos datos.
- b) Dibujar el diagrama de dispersión correspondiente. En vista del diagrama, ¿qué tipo de modelo crees que explicará mejor la relación entre el número de bacterias y el tiempo transcurrido?

i

- 1) Seleccionar el menú Gráficos > Cuadros de diálogo antiguos > Dispersión/Puntos..., elegir la opción Dispersión simple y hacer click sobre el botón Definir.
- 2) Seleccionar la variable bacterias en el campo Eje Y del cuadro de diálogo.
- 3) Seleccionar la variable horas en el campo Eje X del cuadro de diálogo y hacer click sobre el botón Aceptar.

- c) Hacer una comparativa de los distintos modelos de regresión en función del coeficiente de determinación. ¿Qué tipo de modelo es mejor?

i

- 1) Seleccionar el menú *Analizar > Regresión > Estimación curvilínea...*
- 2) Seleccionar la variable bacterias en el campo Dependientes del cuadro de diálogo.
- 3) Seleccionar la variable horas en el campo Independiente/variable del cuadro de diálogo.
- 4) Desmarcar la opción Representar los modelos.
- 5) Marcar las opciones lineal, cuadrático, cúbico, exponencial y logarítmico, y hacer click sobre el botón Aceptar.

d) En vista de lo anterior, calcular el modelo de regresión que mejor explique la relación entre bacterias y horas.

i

Utilizar los coeficientes que aparecen en el punto anterior y la tabla de la parte de fundamentos teóricos.

e) Según el modelo anterior, ¿cuántas bacterias habrá al cabo de 3 horas y media del inicio del cultivo? ¿Y al cabo de 10 horas? ¿Son fiables estas predicciones?

i

- 1) Crear una nueva variable valores e introducir los valores de las horas para los que queremos predecir las bacterias.
- 2) Seleccionar el menú *Transformar > Calcular variable...*
- 3) Introducir el nombre de la nueva variable prediccion en el campo Variable de destino del cuadro de diálogo.
- 4) Introducir la ecuación del mejor modelo en el campo Expresión numérica, utilizando los coeficientes obtenidos anteriormente y la variable valores y hacer click sobre el botón Aceptar.

f) Dar una predicción lo más fiable posible del tiempo que tendría que transcurrir para que en el cultivo hubiese 100 bacterias.

i

Repetir los pasos del apartado anterior introduciendo la variable horas en el campo Dependientes y la variable bacterias en el campo Independiente/Variabes.

2. Se han medido dos variables S y T en 10 individuos, obteniéndose los siguientes resultados:

$(-1,5, 2,25)$, $(0,8, 0,64)$, $(-0,2, 0,04)$, $(-0,8, 0,64)$, $(0,4, 0,16)$,
 $(0,2, 0,04)$, $(-2,1, 4,41)$, $(-0,4, 0,16)$, $(1,5, 2,25)$, $(2,1, 4,41)$.

Se pide:

- a) Crear las variables S y T e introducir estos datos.
- b) Calcular la recta de regresión de T sobre S . Dibujar dicha recta sobre el diagrama de dispersión. ¿Podemos afirmar que S y T son independientes?

i

- 1) Seleccionar el menú Analizar > Regresión > Lineales....
- 2) Seleccionar la variable T en el campo Dependientes del cuadro de diálogo.
- 3) Seleccionar la variable S en el campo Independientes del cuadro de diálogo y hacer click sobre el botón Aceptar.
- 4) Para escribir la ecuación de la recta, observaremos en la ventana de resultados obtenida, la tabla denominada Coeficientes, y en la columna B de los Coeficientes no estandarizados, encontramos en la primera fila la constante de la recta y en la segunda la pendiente.
- 5) Seleccionar el menú Gráficos > Cuadros de diálogo antiguos > Dispersión/Puntos..., elegir la opción Dispersión simple y hacer click sobre el botón Definir.
- 6) Seleccionar la variable T en el campo Eje Y del cuadro de diálogo.
- 7) Seleccionar la variable S en el campo Eje X del cuadro de diálogo y hacer click sobre el botón Aceptar.
- 8) Editar el gráfico realizado anteriormente haciendo un doble click sobre él.
- 9) Seleccionar los puntos haciendo click sobre alguno de ellos.

- 10) Seleccionar el menú **Elementos > Línea de ajuste total** (También se podría usar en lugar del menú, la barra de herramientas)
- 11) Cerrar la ventana **Propiedades**.
- 12) Cerrar el editor de gráficos, cerrando la ventana.

c) Hacer una comparativa de los distintos modelos de regresión en función del coeficiente de determinación. ¿Qué tipo de relación existe entre T y S ?

i

- 1) Seleccionar el menú **Analizar > Regresión > Estimación curvilínea...**
- 2) Seleccionar la variable T en el campo **Dependientes** del cuadro de diálogo.
- 3) Seleccionar la variable S en el campo **Independiente/variable** del cuadro de diálogo.
- 4) Desmarcar la opción **Representar los modelos**.
- 5) Marcar las opciones **lineal, cuadrático, cúbico y exponencial** y hacer click sobre el botón **Aceptar**.

d) En vista de lo anterior, ajustar el modelo de regresión más apropiado.

i

Utilizar los coeficientes que aparecen en el punto anterior y la tabla de la parte de fundamentos teóricos.

5.2. Ejercicios propuestos

1. En un centro dietético se está probando una nueva dieta de adelgazamiento en una muestra de 12 individuos. Para cada uno de ellos se ha medido el número de días que lleva con la dieta y el número de kilos perdidos desde entonces, obteniéndose los siguientes resultados:

(33, 3,9), (51, 5,9), (30, 3,2), (55, 6,0), (38, 4,9), (62, 6,2),
(35, 4,5), (60, 6,1), (44, 5,6), (69, 6,2), (47, 5,8), (40, 5,3)

Se pide:

Regresión No Lineal

- a) Dibujar el diagrama de dispersión. Según la nube de puntos, ¿qué tipo de modelo explicaría mejor la relación entre los kilos perdidos y los días de dieta?
 - b) Construir el modelo de regresión que mejor explique la relación entre los kilos perdidos y los días de dieta.
 - c) Utilizar el modelo construido para predecir el número de kilos perdidos tras 40 días de dieta y tras 100 días. ¿Son fiables estas predicciones?
2. La concentración de un fármaco en sangre, C en mg/dl, es función del tiempo, t en horas, y viene dada por la siguiente tabla:

t	2	3	4	5	6	7	8
C	25	36	48	64	86	114	168

Se pide:

- a) Según el modelo exponencial, ¿qué concentración de fármaco habría a las 4,8 horas? ¿Es fiable la predicción? Justificar adecuadamente la respuesta.
- b) Según el modelo logarítmico, ¿qué tiempo debe pasar para que la concentración sea de 100 mg/dl?

Variables Aleatorias Discretas

6.1. Ejercicios resueltos

En los ejercicios prácticos vamos a utilizar las siguientes funciones matemáticas, también llamadas operadores, de SPSS:

PDF.BINOM (c, n, p) que calcula el valor de la función de probabilidad en el valor c , de la variable binomial de parámetros n y p .

PDF.POISSON (c, λ) que calcula el valor de la función de probabilidad en el valor c , de la variable de Poisson de parámetro λ .

CDF.BINOM (c, n, p) que calcula el valor de la función de distribución en el valor c , de la variable binomial de parámetros n y p .

CDF.POISSON (c, λ) que calcula el valor de la función de distribución en el valor c , de la variable de Poisson de parámetro λ .

1. Dada la v.a.d. con distribución Binomial $X \sim B(10, 0,4)$ se pide:

- a) Crear la variable X e introducir todos sus posibles valores en el editor de datos.
- b) Crear la variable probabilidad que contenga la probabilidad de cada uno de los valores de la variable X. Interpretar los valores de la variable obtenidos.

i

1) Seleccionar el menú Transformar ▶ Calcular variable.

- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso probabilidad, dentro del campo Variable de destino:.
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función $PDF.BINOM(c, n, p)$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función Pdf .Binom. El primer parámetro de la función $PDF.BINOM(c, n, p)$ puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X. El segundo parámetro es el número de veces que se repite el experimento, que en nuestro caso es 10 y el último parámetro es la probabilidad de éxito, en nuestro caso 0.4. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

c) Dibujar la gráfica de la función de probabilidad de la variable X.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
- 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
- 3) Introducir la variable X en el cuadro de diálogo Eje X.
- 4) Introducir la variable creada probabilidad en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.

d) Crear la variable probacumulada que contenga la función de distribución (probabilidad acumulada) de cada uno de los valores de la variable X. Interpretar los valores de la variable obtenidos.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso probacumulada, dentro del campo Variable de destino:.

3) En el cuadro de diálogo Expresión numérica, escribir la función $CDF.BINOM(c, n, p)$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función $Cdf.Binom$. El primer parámetro de la función $CDF.BINOM(c, n, p)$ puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X . El segundo parámetro es el número de veces que se repite el experimento, que en nuestro caso es 10 y el último parámetro es la probabilidad de éxito, en nuestro caso 0.4. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

e) Dibujar la gráfica de la función de distribución de la variable X .

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
- 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
- 3) Introducir la variable X en el cuadro de diálogo Eje X.
- 4) Introducir la variable creada probacumulada en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.

f) Calcular las siguientes probabilidades:

i. $P(X = 7)$.

i

Calcular $PDF.BINOM(7, 10, 0.4)$.

ii. $P(X \leq 4)$.

i

Calcular $CDF.BINOM(4, 10, 0.4)$.

iii. $P(X > 5)$.

i

Calcular $1-CDF.BINOM(5, 10, 0.4)$.

IV. $P(2 \leq X < 9)$.

i

Calcular $CDF.BINOM(8, 10, 0.4) - CDF.BINOM(1, 10, 0.4)$.

i

Se podrían calcular todos los valores a la vez definiendo la variable X, introduciendo todos los valores que necesitamos, en nuestro caso 7, 4, 5, 8, y 1 y creando las variables probabilidad y probabilidad acumulada utilizando para ello las funciones $PDF.BINOM$ y $CDF.BINOM$ respectivamente.

2. Sea X una variable aleatoria binomial tal que con $X \sim B(40, 0.1)$
- a) Crear la variable X e introducir todos sus posibles valores en el editor de datos.
 - b) Crear la variable probabilidad que contenga la probabilidad de cada uno de los valores de la variable X. Interpretar los valores de la variable obtenidos.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso probabilidad, dentro del campo Variable de destino:
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función $PDF.BINOM(c, n, p)$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función Pdf .Binom. El primer parámetro de la función $PDF.BINOM(c, n, p)$ puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X. El segundo parámetro es el número de veces que se repite el experimento, que en nuestro caso es 40 y el último parámetro es la probabilidad de éxito, en nuestro caso 0.1. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

- c) Dibujar la gráfica de la función de probabilidad de la variable X.

- i*
- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
 - 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
 - 3) Introducir la variable X en el cuadro de diálogo Eje X.
 - 4) Introducir la variable creada probabilidad en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.

d) Calcular:

I. $P(X = 8)$.

i Calcular PDF .BINOM(8,40,0.1).

II. $P(X \geq 7)$.

i Calcular 1-CDF .BINOM(6,40,0.1).

III. $P(X < 4)$.

i Calcular CDF .BINOM(3,40,0.1).

IV. $P(3 \leq X \leq 9)$.

i Calcular CDF .BINOM(9,40,0.1)-CDF .BINOM(2,40,0.1).

i Se podrían calcular todos los valores a la vez definiendo la variable X, introduciendo todos los valores que necesitamos, en nuestro caso 8, 6, 3, 9, y 2 y creando las variables probabilidad y probacumulada utilizando para ello las funciones PDF.BINOM y CDF.BINOM respectivamente.

3. Dada la v.a.d. con distribución de Poisson $X \sim P(4)$, se pide:

- a) Crear la variable X e introducir los valores de dicha variable desde 0 hasta 10 en el editor de datos. ¿Podríamos introducir todos los valores de la variable X?

- b) Crear la variable probabilidad que contenga la probabilidad de cada uno de los valores de la variable X. Interpretar los valores de la variable obtenidos.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso probabilidad, dentro del campo Variable de destino:.
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función $PDF.POISSON(c,media)$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función Pdf .Poisson. El primer parámetro de la función $PDF.POISSON(c,media)$ puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X y el segundo parámetro es la media de la variable, en nuestro caso 4. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

- c) Dibujar la gráfica de la función de probabilidad de la variable X. ¿Se parece esta gráfica a la gráfica de la función de probabilidad de la binomial $B(40, 0,1)$ del ejercicio anterior? ¿Cuál puede ser la causa del parecido?

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
- 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
- 3) Introducir la variable X en el cuadro de diálogo Eje X.
- 4) Introducir la variable creada probabilidad en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.

- d) Crear la variable probacumulada que contenga la función de distribución (probabilidad acumulada) de cada uno de los valores de la variable X. Interpretar los valores de la variable obtenidos.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso probacumulada, dentro del campo Variable de destino:.
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función $CDF.POISSON(c,media)$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función $Cdf.Poisson$. El primer parámetro de la función $CDF.POISSON(c,media)$ puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X y el segundo parámetro es la media de la variable, en nuestro caso 4. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

e) Dibujar la gráfica de la función de distribución de la variable X .

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
- 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
- 3) Introducir la variable X en el cuadro de diálogo Eje X.
- 4) Introducir la variable creada probacumulada en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.

f) Calcular las siguientes probabilidades:

I. $P(X = 9)$.

i

Calcular $PDF.POISSON(9,4)$.

II. $P(X \leq 6)$.

i

Calcular $CDF.POISSON(6,4)$.

III. $P(X \geq 5)$.

i Calcular $1 - \text{CDF} . \text{POISSON}(4, 4)$.

iv. $P(4 \leq X < 50)$.

i Calcular $\text{CDF} . \text{POISSON}(49, 4) - \text{CDF} . \text{POISSON}(3, 4)$.

4. La probabilidad de curación de un paciente al ser sometido a un determinado tratamiento es 0,85. Calcular la probabilidad de que en un grupo de 6 enfermos sometidos a tratamiento:

a) Se curen la mitad.

i El número de enfermos que se curan sigue una distribución Binomial $X \sim B(6, 0,85)$, por lo que la probabilidad de que se curen la mitad, es decir $P(X = 3)$, se calculará mediante $\text{PDF} . \text{BINOM}(3, 6, 0.85)$.

b) Se curen al menos 4.

i La $P(X \geq 4)$, se calculará mediante $1 - \text{CDF} . \text{BINOM}(3, 6, 0.85)$.

5. En un servicio de urgencias de cierto hospital se sabe que, en media, llegan 2 pacientes cada hora. Calcular:

a) ¿Cuál es la probabilidad de que en una hora lleguen 3 pacientes?

i El número de pacientes que llegan en una hora sigue una distribución de Poisson de media 2, $X \sim P(2)$, por lo que la probabilidad de que en una hora lleguen 3 pacientes, es decir $P(X = 3)$, se calculará mediante $\text{PDF} . \text{POISSON}(3, 2)$.

b) Si los turnos de urgencias son de 8 horas, ¿cuál será la probabilidad de que en un turno lleguen más de 5 pacientes?

i El número de pacientes que llegan en un turno de urgencias sigue una distribución de Poisson de media 16, $X \sim P(16)$, por lo que

la probabilidad de que lleguen más de 5, $P(X > 5)$, se calculará mediante $1 - \text{CDF.POISSON}(5, 16)$.

6.2. Ejercicios propuestos

1. Al lanzar 100 veces una moneda, ¿cuál es la probabilidad de obtener entre 40 y 60 caras?
2. La probabilidad de que al administrar una vacuna dé una determinada reacción es 0,001. Si se vacunan 2000 personas ¿cuál es la probabilidad de que aparezca alguna reacción adversa?
3. El número medio de llamadas por minuto que llegan a una centralita telefónica es igual a 120. Hallar las probabilidades de los sucesos siguientes:
 - a) A =durante 2 segundos lleguen a la centralita menos de 4 llamadas.
 - b) B =durante 3 segundos lleguen a la centralita 3 llamadas como mínimo.

VARIABLES ALEATORIAS CONTINUAS

7.1. Ejercicios resueltos

En los ejercicios prácticos vamos a utilizar las siguientes funciones matemáticas, también llamadas operadores, de SPSS:

PDF.CHISQ (c, n) que calcula el valor de la función de densidad en el valor c , de la variable Chi-cuadrado con n grados de libertad.

PDF.F (c, m, n) que calcula el valor de la función de densidad en el valor c , de la variable F de Fisher-Snedecor con m y n grados de libertad.

PDF.NORMAL (c, μ, σ) que calcula el valor de la función de densidad en el valor c , de la variable normal de media μ y desviación típica σ .

PDF.T (c, n) que calcula el valor de la función de densidad en el valor c de la variable T de Student con n grados de libertad.

PDF.UNIFORM (c, a, b) que calcula el valor de la función de densidad en el valor c de la variable uniforme U de parámetros a y b .

CDF.CHISQ (c, n) que calcula el valor de la función de distribución en el valor c , de la variable Chi-cuadrado con n grados de libertad.

CDF.F (c, m, n) que calcula el valor de la función de distribución en el valor c , de la variable F de Fisher-Snedecor con m y n grados de libertad.

CDF.NORMAL (c, μ, σ) que calcula el valor de la función de distribución en el valor c , de la variable normal de media μ y desviación típica σ .

CDF.T (c, n) que calcula el valor de la función de distribución en el valor c de la variable T de Student con n grados de libertad.

CDF.UNIFORM (c, a, b) que calcula el valor de la función de distribución en el valor c de la variable uniforme U con parámetros a y b .

IDF.CHISQ (c, n) que calcula el valor de la variable Chi-cuadrado con n grados de libertad, cuando el valor de la función de distribución es c .

IDF.F (c, m, n) que calcula el valor de la variable F de Fisher-Snedecor con m y n grados de libertad, cuando el valor de la función de distribución es c .

IDF.NORMAL (c, μ, σ) que calcula el valor de la variable normal de media μ y desviación típica σ , cuando el valor de la función de distribución es c .

IDF.T (c, n) que calcula el valor de la variable T de Student con n grados de libertad, cuando el valor de la función de distribución es c .

IDF.UNIFORM (c, a, b) que calcula el valor de la variable uniforme U de parámetros a y b , cuando el valor de la función de distribución es c .

1. Dada la v.a.c. con distribución uniforme $X \sim U(0, 2)$, se pide:

- a) Definir la variable X, dándole valores desde 0 hasta 2, con un incremento de 0,1.
- b) Calcular el valor de la función de densidad en los valores de X introducidos.

Observación: Los valores de la función de densidad no son las probabilidades de cada valor de la variable.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso densidad, dentro del campo Variable de destino:.
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función `PDF.UNIFORM(c, mín, máx)` o seleccionar del cuadro de diálogo Grupo de funciones la opción `Todo` y en el cuadro de diálogo Funciones y variables especiales la función `Pdf.Uniform`. El primer parámetro de la función `PDF.UNIFORM(c, mín, máx)` puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir

la variable X. El segundo parámetro es el valor mínimo de la variable, en nuestro caso 0 y el último parámetro es el valor máximo de la variable, en nuestro caso 2. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

c) Representar la función de densidad.

- i*
- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
 - 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
 - 3) Introducir la variable X en el cuadro de diálogo Eje X.
 - 4) Introducir la variable creada densidad en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.
 - 5) Editar el gráfico haciendo doble click sobre el mismo, y como queremos unir los puntos los seleccionaremos todos haciendo click sobre uno de ellos.
 - 6) Hacer click con el botón derecho del ratón y seleccionar Añadir Línea de interpolación
 - 7) Marcar la opción Recto, hacer click sobre el botón Cerrar y por último cerrar el editor de gráficos.

d) Calcular el valor de la función de distribución de dicha variable en los valores de X introducidos.

- i*
- 1) Seleccionar el menú Transformar ▶ Calcular variable.
 - 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso probacumulada, dentro del campo Variable de destino.
 - 3) En el cuadro de diálogo Expresión numérica:, escribir la función $CDF.UNIFORM(c, \text{mín}, \text{máx})$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función $Cdf.Uniform$. El pri-

mer parámetro de la función CDF .UNIFORM(c , mín , máx) puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X. El segundo parámetro es el valor mínimo de la variable, en nuestro caso 0 y el último parámetro es el valor máximo de la variable, en nuestro caso 2. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

e) Dibujar la gráfica de la función de distribución.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
- 2) Seleccionar la opción Dispersión Simple y hacer click sobre el botón Definir.
- 3) Introducir la variable X en el cuadro de diálogo Eje X.
- 4) Introducir la variable creada probacumulada en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.
- 5) Editar el gráfico haciendo doble click sobre el mismo, y como queremos unir los puntos los seleccionaremos todos haciendo click sobre uno de ellos.
- 6) Hacer click con el botón derecho del ratón y seleccionar Añadir Línea de interpolación
- 7) Marcar la opción Recto, hacer click sobre el botón Cerrar y por último cerrar el editor de gráficos.

2. Dada la v.a.c. con distribución normal $X \sim N(0, 1)$, se pide:

- a) Crear la variable X e introducir en el editor de datos de entre sus infinitos valores (toda la recta real), por ejemplo aquellos que van desde -3 hasta 3, en incrementos de 0.2 (para introducir los 31 valores diferentes resulta aconsejable la utilización de un programa de hoja de cálculo como Excel).
- b) Calcular el valor de la función de densidad en los valores de X introducidos.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso densidad, dentro del campo Variable de destino:.
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función $PDF.NORMAL(c,media,desv_tip)$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función Pdf .Normal. El primer parámetro de la función puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X. El segundo parámetro es la media de la variable, en nuestro caso 0, y el último parámetro es la desviación típica, en nuestro caso 1. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

c) Dibujar la gráfica de la función de densidad de la variable X.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
- 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
- 3) Introducir la variable X en el cuadro de diálogo Eje X.
- 4) Introducir la variable creada densidad en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.
- 5) Editar el gráfico haciendo doble click sobre el mismo, y como queremos unir los puntos los seleccionaremos todos haciendo click sobre uno de ellos.
- 6) Hacer click con el botón derecho del ratón y seleccionar Añadir Línea de interpolación
- 7) Marcar la opción Recto, hacer click sobre el botón Cerrar y por último cerrar el editor de gráficos.

d) Calcular el valor de la función de distribución de dicha variable en los valores de X introducidos.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso probacumulada, dentro del campo Variable de destino.
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función $CDF.NORMAL(c,media,desv_tip)$ o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función $Cdf.Normal$. El primer parámetro de la función puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable X. El segundo parámetro es la media de la variable, en nuestro caso 0, y el último parámetro es la desviación típica, en nuestro caso 1. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

e) Dibujar la gráfica de la función de distribución.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos.
- 2) Seleccionar la opción Dispersión simple y hacer click sobre el botón Definir.
- 3) Introducir la variable X en el cuadro de diálogo Eje X.
- 4) Introducir la variable creada probacumulada en el cuadro de diálogo Eje Y y hacer click sobre el botón Aceptar.
- 5) Editar el gráfico haciendo doble click sobre el mismo, y como queremos unir los puntos los seleccionaremos todos haciendo click sobre uno de ellos.
- 6) Hacer click con el botón derecho del ratón y seleccionar Añadir Línea de interpolación
- 7) Marcar la opción Recto, hacer click sobre el botón Cerrar y por último cerrar el editor de gráficos.

- f) Crear la variable `invprobacumu` que, aplicada al valor de la función de distribución (probabilidad acumulada), nos devuelva el valor de X . Interpretar los valores de la variable obtenidos.

i

- 1) Seleccionar el menú Transformar ▶ Calcular variable.
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso `invprobacumu`, dentro del campo Variable de destino.
- 3) En el cuadro de diálogo Expresión numérica:, escribir la función `IDF.NORMAL(c,media,desv_tip)` o seleccionar del cuadro de diálogo Grupo de funciones la opción Todo y en el cuadro de diálogo Funciones y variables especiales la función `Idf.Normal`. El primer parámetro de la función puede ser un único valor o el nombre de una variable para calcular todos sus valores a la vez, en cuyo caso se debe introducir la variable probacumulada. El segundo parámetro es la media de la variable, en nuestro caso 0, y el último parámetro es la desviación típica, en nuestro caso 1. En este caso al introducir expresiones numéricas, se utilizará el punto como separador de decimales, ya que la coma se usa como separador de parámetros. Por último hacer click sobre el botón Aceptar.

- g) Calcular las siguientes probabilidades:

i. $P(X < -1)$.

i

Calcular `CDF.NORMAL(-1,0,1)`.

ii. $P(X \leq 2)$.

i

Calcular `CDF.NORMAL(2,0,1)`.

iii. $P(X > 1,2)$.

i

Calcular `1-CDF.NORMAL(1.2,0,1)`.

iv. $P(-2 \leq X < 2,5)$.

i Calcular $\text{CDF.NORMAL}(2.5, 0, 1) - \text{CDF.NORMAL}(-2, 0, 1)$.

v. $P(-2,223 \leq X \leq 2,581)$.

i Calcular $\text{CDF.NORMAL}(2.581, 0, 1) - \text{CDF.NORMAL}(-2.223, 0, 1)$.

vi. $P(1,002 < X < 6,234)$.

i Calcular $\text{CDF.NORMAL}(6.234, 0, 1) - \text{CDF.NORMAL}(1.002, 0, 1)$.

i Se podrían calcular todos los valores a la vez definiendo la variable X, introduciendo todos los valores que necesitamos, en nuestro caso $-1, 2, 1, 2, 2, 5, -2, 2, 581, -2, 223, 6, 234$ y $1, 002$ y creando la variable probacumulada utilizando para ello la función CDF.NORMAL .

h) Calcular los valores x_0 que cumplan lo siguiente.

i. $P(X \leq x_0) = 0,995$.

i Calcular $\text{IDF.NORMAL}(0.995, 0, 1)$.

ii. $P(X > x_0) = 0,025$.

i Calcular $\text{IDF.NORMAL}(1 - 0.025, 0, 1)$.

i Se podrían calcular todos los valores a la vez definiendo la variable X, introduciendo todos los valores que necesitamos, en nuestro caso $0,995$ y $1 - 0,025$ y creando la variable invprobacumu utilizando para ello la función IDF.NORMAL .

3. Dada la v.a.c. con distribución normal $X \sim N(2,5, 5)$, se pide:

a) Calcular las siguientes probabilidades:

i. $P(X > 4,5)$.

i Calcular $1 - \text{CDF.NORMAL}(4.5, 2.5, 5)$.

II. $P(2,58 \leq X \leq 3,65)$.

i Calcular $CDF.NORMAL(3.65, 2.5, 5) - CDF.NORMAL(2.58, 2.5, 5)$.

i Se podría calcular todos los valores a la vez definiendo la variable X , introduciendo todos los valores que necesitamos, en nuestro caso 4,5, 3,65 y 2,58 y creando la variable probacumulada utilizando para ello la función $CDF.NORMAL$.

b) Calcular los valores de x_0 que cumplen:

I. $P(X < x_0) = 0,238$.

i Calcular $IDF.NORMAL(0.238, 2.5, 5)$.

II. $P(X > x_0) = 0,621$.

i Calcular $IDF.NORMAL(1-0.621, 2.5, 5)$.

i Se podría calcular todos los valores a la vez definiendo la variable X , introduciendo todos los valores que necesitamos, en nuestro caso 0,238 y $1 - 0,621$ y creando la variable invprobacumu utilizando para ello la función $IDF.NORMAL$.

4. Dada la v.a.c. con distribución T de Student $X \sim T(6)$, calcular las siguientes probabilidades:

a) $P(X \leq 0)$

i Calcular $CDF.T(0, 6)$.

b) $P(X \leq 1,2)$

i Calcular $CDF.T(1.2, 6)$.

c) $P(X > 1,2)$

i Calcular $1 - \text{CDF} . T(1.2, 6)$.

d) $P(0 \leq X \leq 1,2)$

i Calcular $\text{CDF} . T(1.2, 6) - \text{CDF} . T(0, 6)$.

i Se podría calcular todos los valores a la vez definiendo la variable X, introduciendo todos los valores que necesitamos, en nuestro caso 0 y 1,2 y creando la variable probabilidad utilizando para ello la función $\text{CDF} . T$.

5. Dada la v.a.c. con distribución Chi-cuadrado $X \sim \chi(5)$ se pide calcular los valores x_0 que cumplan lo siguiente:

a) $P(X \leq x_0) = 0,90$

i Calcular $\text{IDF} . \text{CHISQ}(0.90, 5)$.

b) $P(X \leq x_0) = 0,99$

i Calcular $\text{IDF} . \text{CHISQ}(0.99, 5)$.

c) $P(X > x_0) = 0,025$

i Calcular $\text{IDF} . \text{CHISQ}(1 - 0.025, 5)$.

i Se podría calcular todos los valores a la vez definiendo la variable X, introduciendo todos los valores que necesitamos, en nuestro caso 0,90, 0,99 y $1 - 0,025$ y creando la variable invprobacumu utilizando para ello la función $\text{IDF} . \text{CHISQ}$.

6. Un autobús pasa por una parada cada 15 minutos. Si una persona llega a la parada en un momento cualquiera:

a) Calcular la probabilidad de que tenga que esperar más de diez minutos al autobús.

i El tiempo de espera sigue una distribución Uniforme continua de parámetros 0 y 15, es decir, $X \sim U(0, 15)$, y la probabilidad de que tenga que esperar más de 10 minutos vendrá dada por $1 - \text{CDF} . \text{UNIFORM}(10, 0, 15)$.

b) Calcular la probabilidad de que tenga que esperar entre 7 y 10 minutos

i La probabilidad de que tenga que esperar entre 7 y 10 minutos vendrá dada por $\text{CDF} . \text{UNIFORM}(10, 0, 15) - \text{CDF} . \text{UNIFORM}(7, 0, 15)$.

7. Se sabe que en una población la estatura de sus habitantes varones de entre 20 y 40 años sigue una distribución aproximadamente Normal de media 176 cm. y desviación típica 5 cm.

a) Si se elige aleatoriamente un varón de entre 20 y 40 años, ¿cuál es la probabilidad de que mida menos de 165 cm.?

i La estatura sigue una distribución Normal de media 176 cm. y desviación típica 5 cm., es decir, $X \sim N(176, 5)$, y la probabilidad de que mida menos de 165 cm. vendrá dada por $\text{CDF} . \text{NORMAL}(165, 176, 5)$.

b) ¿Cuál será la estatura por encima de la cual estará el 10 % de los varones?

i Si por encima de la estatura buscada está el 10 % de los varones, por debajo de dicha estatura estará el 90 %, por lo que para calcular dicho valor habrá que hacer $\text{IDF} . \text{NORMAL}(1 - 0.10, 176, 5)$.

7.2. Ejercicios propuestos

1. Hallar las siguientes probabilidades:

a) $P(-2,4 \leq Z \leq -1,2)$ si Z es $N(0, 1)$

b) $P(|Z| > 1,2)$ si Z es $N(0, 1)$

c) $P(1,3 \leq X \leq 3,3)$ si X es $N(2, 1)$

- d) $P(|X - 3| > 2)$ si X es $N(3, 4)$
2. Entre los diabéticos, el nivel de glucosa en la sangre en ayunas X , puede suponerse de distribución aproximadamente normal, con media 106mg/100ml y desviación típica 8mg/100ml.
- a) Hallar $P(X \leq 120\text{mg}/100\text{ml})$
- b) ¿Qué porcentaje de diabéticos tendrá niveles entre 90 y 120mg/100ml?
- c) Encontrar un valor que tenga la propiedad de que el 25 % de los diabéticos tenga un nivel de glucosa por debajo de dicho valor.
3. Se sabe que el nivel de colesterol en varones de más de 30 años sigue una distribución normal, de media 220 y desviación típica 30. Realizado un estudio sobre 20000 varones mayores de 30 años,
- a) ¿Cuántos se espera que tengan su nivel de colesterol entre 210 y 240?
- b) ¿Cuántos se espera que tengan su nivel de colesterol por encima de 250?
- c) ¿Cuál será el nivel de colesterol, por encima del cual se espera que esté el 20 % de la población?

8

Intervalos de Confianza para Medias y Proporciones

8.1. Ejercicios resueltos

1. Se analiza la concentración de principio activo en una muestra de 10 envases tomados de un lote de un fármaco, obteniendo los siguientes resultados en mg/mm^3 :

17,6 – 19,2 – 21,3 – 15,1 – 17,6 – 18,9 – 16,2 – 18,3 – 19,0 – 16,4

Se pide:

- a) Crear la variable concentracion, e introducir los datos de la muestra.
- b) Calcular el intervalo de confianza para la media de la concentración del lote con nivel de confianza del 95 % (nivel de significación $\alpha = 0,05$).

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar.
- 2) En el cuadro de diálogo que aparece seleccionar la variable concentracion en el campo Lista de dependientes y hacer click en el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece habilitar la opción Descriptivos, introducir el nivel de confianza deseado en el cuadro Intervalo de confianza para la media, y hacer click en el botón Continuar y Aceptar.

- c) Calcular los intervalos de confianza para la media con niveles del 90 % y del 99 % (niveles de significación $\alpha = 0,1$ y $\alpha = 0,01$).

i Repetir los mismos pasos del apartado anterior, cambiando el nivel de confianza para cada intervalo.

- d) Si definimos la precisión del intervalo como la inversa de su amplitud, ¿cómo afecta a la precisión del intervalo de confianza el tomar niveles de significación cada vez más altos? ¿Cuál puede ser la explicación?
 - e) Si, para que sea efectivo, el fármaco debe tener una concentración mínima de 16 mg/mm³ de principio activo, ¿se puede aceptar el lote como bueno? Justificar la respuesta.
2. Una central de productos lácteos recibe diariamente la leche de dos granjas X e Y. Para analizar la calidad de la leche, durante una temporada, se controla el contenido de materia grasa de la leche que proviene de ambas granjas, con los siguientes resultados:

X		Y	
0,34	0,34	0,28	0,29
0,32	0,35	0,30	0,32
0,33	0,33	0,32	0,31
0,32	0,32	0,29	0,29
0,33	0,30	0,31	0,32
0,31	0,32	0,29	0,31
		0,33	0,32
		0,32	0,33

- a) Crear las variables grasa y granja, e introducir los datos de la muestra.
- b) Calcular el intervalo de confianza con un 95 % de confianza para el contenido medio de materia grasa de la leche sin tener en cuenta si la misma procede de una u otra granja.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar.
- 2) En el cuadro de diálogo que aparece seleccionar la variable grasa en el campo Lista de dependientes y hacer click en el botón Estadísticos.

- 3) En el cuadro de diálogo que aparece habilitar la opción Descriptivos, introducir el nivel de confianza deseado en el cuadro Intervalo de confianza para la media, y hacer click en el botón Continuar y Aceptar.

- c) Calcular los intervalos de confianza con un 95 % de confianza para el contenido medio de materia grasa de la leche dividiendo los datos según la granja de procedencia de la leche.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar.
- 2) En el cuadro de diálogo que aparece seleccionar la variable grasa en el campo Lista de dependientes, seleccionar la variable granja en el campo Lista de factores y hacer click en el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece habilitar la opción Descriptivos, introducir el nivel de confianza deseado en el cuadro Intervalo de confianza para la media, y hacer click en el botón Continuar y Aceptar.

- d) A la vista de los intervalos obtenidos en el punto anterior, ¿se puede concluir que existen diferencias significativas en el contenido medio de grasa según la procedencia de la leche? Justificar la respuesta.

3. En una encuesta realizada en una facultad, sobre si los alumnos utilizan habitualmente (al menos una vez a la semana) la biblioteca de la misma, se han obtenido los siguientes resultados, en los que se ha anotado 1 si la respuesta ha sido positiva y 0 si ha sido negativa:

Alumno	1	2	3	4	5	6	7	8	9	10	11	12
Respuesta	0	1	0	0	0	1	0	1	1	1	1	0

Alumno	13	14	15	16	17	18	19	20	21	22	23
Respuesta	1	0	1	0	0	0	1	1	1	0	0

Alumno	24	25	26	27	28	29	30	31	32	33	34
Respuesta	1	0	0	1	1	0	0	1	0	1	0

- a) Crear la variable respuesta e introducir los datos de la muestra.
- b) Calcular el intervalo de confianza con $\alpha = 0,01$ para la media de la variable respuesta.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar.
- 2) En el cuadro de diálogo que aparece seleccionar la variable respuesta en el campo Lista de dependientes y hacer click en el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece habilitar la opción Descriptivos, introducir el nivel de confianza deseado en el cuadro Intervalo de confianza para la media, y hacer click en el botón Continuar y Aceptar.

- c) ¿Qué interpretación tiene dicho intervalo en términos de proporción de alumnos que habitualmente utilizan la biblioteca?
4. El Ministerio de Sanidad está interesado en la elaboración de un intervalo de confianza para la proporción de personas mayores de 65 años con problemas respiratorios que han sido vacunadas en una determinada ciudad. Para ello, después de preguntar a 200 pacientes mayores de 65 años con problemas respiratorios en los hospitales de dicha ciudad, 154 responden afirmativamente.
- a) Crear la variable respuesta cuyos dos únicos valores serán 0 para las respuestas negativas, y 1 para las positivas, y una segunda variable que podemos denominar frecuencia cuyos valores son las frecuencias absolutas de cada una de las respuestas (46 para la respuesta 0 y 154 para la respuesta 1).
 - b) Ponderar los valores de la variable respuesta mediante los pesos introducidos en la variable frecuencia.

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.
- 2) En el cuadro de diálogo que aparece activar la opción Ponderar casos mediante, seleccionar la variable frecuencia y hacer click en el botón Aceptar.

- c) Calcular el intervalo de confianza al 95 % para la proporción de pacientes vacunados.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar.
- 2) En el cuadro de diálogo que aparece seleccionar la variable respuesta en el campo Lista de dependientes y hacer click en el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece habilitar la opción Descriptivos, introducir el nivel de confianza deseado en el cuadro Intervalo de confianza para la media, y hacer click en el botón Continuar y Aceptar.

- d) Si entre los objetivos del Ministerio se encontraba alcanzar una proporción del al menos un 70 % de vacunados en dicho colectivo, ¿se puede concluir que se han cumplido los objetivos? Justificar la respuesta.

8.2. Ejercicios propuestos

1. Para determinar el nivel medio de colesterol en la sangre de una población, se realizaron análisis sobre una muestra de 8 personas, obteniéndose los siguientes resultados:

196 – 212 – 188 – 206 – 203 – 210 – 201 – 198

Hallar los intervalos de confianza para la media del nivel de colesterol con niveles de significación 0,1, 0,05 y 0,01.

2. Utilizar el fichero Hipertensos Datos Claves para estimar mediante un intervalo de confianza del 95 % la presión diastólica inicial media en todos los individuos. Calcular también los intervalos del 95 % de confianza correspondientes a los grupos tratados con placebo, IECA y Ca Antagonista + Diurético. ¿Existen diferencias significativas en la presión diastólica inicial media de esos grupos?
3. Para tratar un determinado síndrome neurológico se utilizan dos técnicas *A* y *B*. En un estudio se tomó una muestra de 60 pacientes con dicho

síndrome y se le aplicó la técnica *A* a 25 de ellos y la técnica *B* a los 35 restantes. De los pacientes tratados con la técnica *A*, 18 se curaron, mientras que de los tratados con la técnica *B*, se curaron 21. Calcular un intervalo de confianza del 95 % para la proporción de curaciones con cada técnica.

4. A las siguientes elecciones locales en una ciudad se presentan tres partidos: *A*, *B* y *C*. Con el objetivo de hacer una estimación sobre la proporción de voto que cada uno de ellos obtendrá, se realiza una encuesta a la que responden 300 personas, de las cuales 60 piensan votar a *A*, 80 a *B*, 90 a *C*, 15 en blanco y 55 abstenciones. Calcular un intervalo de confianza para la proporción de votos, sobre el total del censo, de cada uno de los partidos que se presentan.

Intervalos de Confianza para Comparar Poblaciones

9.1. Ejercicios resueltos

1. Para ver si una campaña de publicidad sobre un fármaco ha influido en sus ventas, se tomó una muestra de 8 farmacias y se midió el número de unidades de dicho fármaco vendidas durante un mes, antes y después de la campaña, obteniéndose los siguientes resultados:

Antes	147	163	121	205	132	190	176	147
Después	150	171	132	208	141	184	182	145

- a) Crear las variables antes y despues e introducir los datos de la muestra.
- b) Obtener un resumen estadístico en el que aparezcan la media y la desviación típica de ambas variables. A la vista de los resultados: ¿son las medias diferentes?, ¿ha aumentado la campaña el nivel de ventas?, ¿crees que los resultados son estadísticamente significativos?

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Frecuencias.
- 2) En el cuadro de diálogo que aparece seleccionar las dos variables en el campo Variables y hacer click en el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece activar los estadísticos deseados y hacer click en el botón Continuar y Aceptar.

- c) Obtener los intervalos de confianza para la media de la diferencia entre ambas variables con niveles de significación 0,05 y 0,01.

i

- 1) Seleccionar el menú *Analizar* ▶ *Comparar medias* ▶ *Prueba T para muestras relacionadas*.
- 2) En el cuadro de diálogo que aparece seleccionar ambas variables en el campo *Variables emparejadas* y hacer click en el botón *Opciones*.
- 3) En el cuadro de diálogo que aparece introducir el nivel de confianza deseado en el campo *Porcentaje del intervalo de confianza* y hacer click en el botón *Continuar* y *Aceptar*.

- d) Crear la variable diferencia, que se obtiene como antes-despues.

i

- 1) Seleccionar el menú *Transformar* ▶ *Calcular variable*.
- 2) En el cuadro de diálogo que aparece introducir el nombre de la nueva variable diferencia en el campo *Variable de destino*, en el campo *Expresión numérica* introducir *antes-despues* y hacer click en el botón *Aceptar*.

- e) Calcular el intervalo de confianza para la media de la variable diferencia con un nivel de significación del 0,05 y comparar el intervalo obtenido con el del apartado anterior.

i

- 1) Seleccionar el menú *Analizar* ▶ *Comparar medias* ▶ *Prueba T para una muestra*.
- 2) En el cuadro de diálogo que aparece seleccionar la variable diferencia en el campo *Variables para contrastar* y hacer click en el botón *Opciones*.
- 3) En el cuadro de diálogo que aparece introducir el nivel de confianza deseado en el campo *Porcentaje del intervalo de confianza* y hacer click en el botón *Continuar* y *Aceptar*.

- f) ¿Existen pruebas suficientes para afirmar con un 95 % de confianza que la campaña de publicidad ha aumentado las ventas? ¿Y si cambiamos los dos últimos datos de la variable *despues* y ponemos 190

en lugar de 182 y 165 en lugar de 145? Observar qué le ha sucedido al intervalo para la diferencia de medias y darle una explicación.

2. Una central de productos lácteos recibe diariamente la leche de dos granjas X e Y. Para analizar la calidad de la leche, durante una temporada, se controla el contenido de materia grasa de la leche que proviene de ambas granjas, con los siguientes resultados:

X		Y	
0,34	0,34	0,28	0,29
0,32	0,35	0,30	0,32
0,33	0,33	0,32	0,31
0,32	0,32	0,29	0,29
0,33	0,30	0,31	0,32
0,31	0,32	0,29	0,31
		0,33	0,32
		0,32	0,33

- a) Crear las variables grasa y granja, e introducir los datos de la muestra.
b) Calcular el intervalo de confianza con un 95 % de confianza para la diferencia en el contenido medio de materia grasa de la leche procedente de ambas granjas.

i

- 1) Seleccionar el menú Analizar ▶ Comparar medias ▶ Prueba T para muestras independientes.
- 2) En el cuadro de diálogo que aparece seleccionar la variable grasa en el campo Variables para contrastar, seleccionar la variable granja en el campo Variable de agrupación y hacer click sobre el botón Definir grupos.
- 3) En el cuadro de diálogo que aparece introducir en el campo Grupo 1 el valor de la variable granja correspondiente a la granja X y en el campo Grupo 2 el correspondiente a la granja Y, y hacer click sobre el botón Continuar.
- 4) En el cuadro de diálogo inicial hacer click sobre el botón Opciones.

5) En el cuadro de diálogo que aparece introducir el nivel de confianza deseado en el campo Porcentaje del intervalo de confianza y hacer click en el botón Continuar y Aceptar.

- c) A la vista de los intervalos obtenidos en el punto anterior, ¿se puede concluir que existen diferencias significativas en el contenido medio de grasa según la procedencia de la leche?
- d) De los dos intervalos de confianza que se obtienen, uno da como su- puesta la igualdad de varianzas y el otro no. En este ejemplo concre- to, ¿sería metodológicamente correcto considerar adecuado el que supone igualdad de varianzas?

i Entre los resultados aparece la prueba de Levene para contrastar la igualdad de varianzas. Se trata de un contraste de hipótesis cu- yo resultado final es un p -valor, denominado Sig. por el progra- ma, que aprenderemos a interpretar en prácticas sucesivas. Por ahora es suficiente tener en cuenta que cuando el p -valor es ma- yor que el nivel de significación elegido no podemos rechazar la hipótesis de igualdad de varianzas.

3. Un profesor universitario ha tenido dos grupos de clase a lo largo del año: uno con horario de mañana y otro de tarde. En el de mañana, sobre un total de 80 alumnos, han aprobado 55; y en el de tarde, sobre un total de 90 alumnos, han aprobado 32.

- a) Crear las variables: grupo, cuyos valores serán 0 (mañana) y 1 (tarde); aprobado, cuyos valores serán 1 (aprobado) y 0 (suspense); y frecuen- cia, cuyos valores serán el número de aprobados y suspensos en cada grupo.
- b) Ponderar los valores de la variable aprobado mediante los pesos de la variable frecuencia.

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.
- 2) En el cuadro de diálogo resultante activar la opción Ponderar casos mediante, seleccionar la variable frecuencia en el campo Variable de frecuencia y hacer click en el botón Aceptar.

- c) Calcular el intervalo para la diferencia de proporciones de alumnos aprobados en cada grupo.

i

- 1) Seleccionar el menú Analizar ▶ Comparar medias ▶ Prueba T para muestras independientes.
- 2) Seleccionar la variable aprobado en el campo Variables para contrastar, la variable grupo en el campo Variable de agrupación y hacer click en el botón Definir grupos.
- 3) En el cuadro de diálogo que aparece introducir en el campo Grupo 1 el valor de la variable grupo correspondiente al grupo de mañana y en el campo Grupo 2 el correspondiente al grupo de tarde, y hacer click sobre el botón Continuar.
- 4) En el cuadro de diálogo inicial hacer click sobre el botón Opciones.
- 5) En el cuadro de diálogo que aparece introducir el nivel de confianza deseado en el campo Porcentaje del intervalo de confianza y hacer click en el botón Continuar y Aceptar.

- d) Suponiendo que el resto de los factores (temario, complejidad de examen, nivel previo de conocimientos, expediente académico previo de los alumnos,...) no han influido en el aprobado o suspenso en la asignatura, ¿se puede concluir que el factor horario ha sido determinante en la proporción de suspensos? Justificar la respuesta.

9.2. Ejercicios propuestos

1. Se ha realizado un estudio para investigar el efecto del ejercicio físico en el nivel de colesterol en la sangre. En el estudio participaron once personas, a las que se les midió el nivel de colesterol antes y después de desarrollar un programa de ejercicios. Los resultados obtenidos fueron los siguientes:

Antes	182	232	191	200	148	249	276	213	241	280	262
Después	198	210	194	220	138	220	219	161	210	213	226

- a) Hallar el intervalo de confianza del 90 % para la diferencia del nivel medio de colesterol antes y después del programa de ejercicios.

b) A la vista de dicho intervalo, ¿se concluye que el ejercicio físico disminuye el nivel de colesterol con una confianza del 90 %?

2. Dos químicos *A* y *B* realizan respectivamente 14 y 16 determinaciones de la actividad radiactiva de una muestra de material. Sus resultados en Curios fueron:

<i>A</i>		<i>B</i>	
263,36	254,68	286,53	254,54
248,64	276,32	284,55	286,30
243,64	256,42	272,52	282,90
272,68	261,10	283,85	253,75
287,33	268,41	252,01	245,26
287,26	282,65	275,08	266,08
250,97	284,27	267,53	252,05
		253,82	269,81

a) Calcular el intervalo de confianza para la diferencia de las medias de actividad detectada por cada uno de los químicos con un 95 % de confianza.

b) ¿Se puede decir que existen diferencias significativas en la media de actividad detectada por cada químico?

3. Utilizar el fichero Hipertensos Datos Claves para calcular el intervalo de confianza del 95 % para la comparación de la presión sistólica media inicial y final, en cada uno de los tratamientos. ¿Qué tratamiento ha sido más efectivo para reducir la presión?

4. En una encuesta realizada por la Junta de Comunidades de Castilla la Mancha, en los dos hospitales de una ciudad, se pregunta a los pacientes hospitalizados cuando salen del hospital por si consideran que el trato recibido ha sido correcto. En el primero de ellos se pregunta a 200 pacientes y 140 responden que sí, mientras que en el segundo, se pregunta a 300 pacientes y 160 responden que sí.

a) Calcular el intervalo de confianza para la diferencia de proporciones de pacientes satisfechos con el trato recibido.

b) ¿Hay pruebas significativas de que el trato recibido en un hospital es mejor que en el otro?

10

Contraste de Hipótesis

10.1. Ejercicios resueltos

1. Se analiza la concentración de principio activo en una muestra de 10 envases tomados de un lote de un fármaco, obteniendo los siguientes resultados en mg/mm^3 :

17,6 – 19,2 – 21,3 – 15,1 – 17,6 – 18,9 – 16,2 – 18,3 – 19,0 – 16,4

Se pide:

- a) Crear la variable concentración, e introducir los datos de la muestra.
- b) Realizar el contraste de hipótesis bilateral $H_0 : \mu = 18$ y $H_1 : \mu \neq 18$ con un nivel de significación 0,05.

i

- 1) Seleccionar el menú **Analizar** ▶ **Comparar medias** ▶ **Prueba T para una muestra**.
- 2) En el cuadro de diálogo que aparece, seleccionar la variable concentración en el campo **Variables** para contrastar, introducir el valor de la media en la hipótesis nula (en este caso 18) en el campo **Valor de prueba** y hacer click en el botón **Opciones**.

3) En el cuadro de diálogo que aparece, introducir el nivel de confianza del contraste en el campo Porcentaje del Intervalo de confianza y hacer click sobre el botón Continuar y Aceptar. Aunque el nivel de confianza no afecta al p -valor del contraste, que el programa denomina sig. (bilateral), si afecta al intervalo de confianza que se muestra junto con el p -valor.

- c) De igual manera realizar los contrastes bilaterales $H_0 : \mu = 19,5$ y $H_1 : \mu \neq 19,5$ con niveles de significación 0,05 y 0,01. ¿Cómo afecta la disminución en el nivel de significación en la facilidad para rechazar H_0 ?

i Seguir los mismos pasos del apartado anterior.

- d) Realizar los contrastes de hipótesis unilaterales $H_0 : \mu = 17$ y $H_1 : \mu > 17$, y $H_0 : \mu = 17$ y $H_1 : \mu < 17$ con un nivel de significación de 0,1.

i Repetir los mismos pasos de apartados anteriores pero teniendo en cuenta que lo que el programa denomina Sig. (bilateral) es el p -valor del contraste bilateral; por lo tanto, para el contraste unilateral de mayor el p -valor será Sig. (bilateral)/2, y para el contraste unilateral de menor el p -valor será 1-(Sig. (bilateral)/2).

- e) Si el fabricante del lote asegura haber aumentado la concentración de principio activo con respecto a anteriores lotes, en los que la media era de 17.5 mg/mm³, con un nivel de confianza del 95%; ¿aceptamos o rechazamos lo dicho por el fabricante?

2. Varios investigadores desean saber si es posible concluir que dos poblaciones de niños difieren respecto a la edad promedio en la cual pueden caminar por sí solos. Los investigadores obtuvieron los siguientes datos para la edad al comenzar a andar (expresada en meses):

Muestra en la población A : 9,5 – 10,5 – 9,0 – 9,8 – 10,0 – 13,0 – 10,0 – 13,5
10,0 – 9,8

Muestra en la población B : 12,5 – 9,5 – 13,5 – 13,8 – 12,0 – 13,8 – 12,5
9,5 – 12,0 – 13,5 – 12,0 – 12,0

Contraste de Hipótesis

- a) Crear las variables población y edad.

- b) Realizar un contraste de hipótesis, con un nivel de significación de 0,05, para dar respuesta a la conclusión que buscan los investigadores.

i

Se trata de un contraste bilateral de comparación de medias en poblaciones independientes.

- 1) Seleccionar el menú *Analizar* ▶ *Comparar medias* ▶ *Prueba T para muestras independientes*.
- 2) En el cuadro de diálogo que aparece, seleccionar la variable *edad* en el campo *Variables para contrastar*, la variable *población* en el campo *Variable de agrupación* y hacer click sobre el botón *Definir grupos*.
- 3) En el cuadro de diálogo que aparece, escribir el valor de la variable *población* correspondiente a la población *A* en el campo *Grupo 1* y el correspondiente a la población *B* en el campo *Grupo 2*, y hacer click sobre el botón *Continuar y Aceptar*.

3. Algunos investigadores han observado una mayor resistencia de las vías respiratorias en fumadores que en no fumadores. Para confirmar dicha hipótesis, se realizó un estudio para comparar el porcentaje de retención traqueobronquial en las mismas personas cuando aún eran fumadoras y transcurrido un año después de dejarlo. Los resultados se indican en la tabla siguiente:

Porcentaje de retención	
Cuando fumaba	Transcurrido un año sin fumar
60,6	47,5
12,0	13,3
56,0	33,0
75,2	55,2
12,5	21,9
29,7	27,9
57,2	54,3
62,7	13,9
28,7	8,9
66,0	46,1
25,2	29,8
40,1	36,2

- a) Crear las variables antes y después e introducir los datos.
- b) Plantear el contraste de hipótesis adecuado para confirmar o rechazar la hipótesis de los investigadores.

i

Se trata de un contraste unilateral ($H_0 : \mu_1 = \mu_2$ y $H_1 : \mu_1 > \mu_2$) de igualdad de medias en datos pareados.

- 1) Seleccionar el menú *Analizar* ▶ *Comparar medias* ▶ *Prueba T para muestras relacionadas*.
- 2) En el cuadro de diálogo que aparece, seleccionar ambas variables y pasarlas al campo *Variables emparejadas*.

Al tratarse de un contraste unilateral de mayor el p -valor será $\text{Sig. (bilateral)}/2$.

4. Un profesor universitario ha tenido dos grupos de clase a lo largo del año: uno con horario de mañana y otro de tarde. En el de mañana, sobre un total de 80 alumnos, han aprobado 55; y en el de tarde, sobre un total de 90 alumnos, han aprobado 32.
 - a) Crear las variables: grupo, cuyos valores serán mañana y tarde; calificación, cuyos valores serán 1 (aprobado) y 0 (suspense); y frecuencia, cuyos diferentes valores son el número de aprobados y suspensos en cada grupo.

b) Ponderar los datos mediante la variable frecuencia.

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.
- 2) En el cuadro de diálogo resultante activar la opción Ponderar casos mediante, seleccionar la variable frecuencia en el campo Variable de frecuencia y hacer click en el botón Aceptar.

c) Realizar un contraste de hipótesis para determinar si el factor horario ha sido o no determinante en la proporción de suspensos con un nivel de significación 0,05

i

Se trata de un contraste bilateral de comparación de medias en poblaciones independientes.

- 1) Seleccionar el menú Analizar ▶ Comparar medias ▶ Prueba T para muestras independientes.
- 2) Seleccionar la variable calificación en el campo Variables para contrastar, la variable grupo en el campo Variable de agrupación y hacer click en el botón Definir grupos.
- 3) En el cuadro de diálogo que aparece introducir en el campo Grupo 1 el valor de la variable grupo correspondiente al grupo de mañana y en el campo Grupo 2 el correspondiente al grupo de tarde, y hacer click sobre el botón Continuar.

10.2. Ejercicios propuestos

1. Un grupo de médicos intenta probar que un programa de ejercicios moderadamente activos puede beneficiar a los pacientes que han sufrido previamente un infarto de miocardio. Para ello escogieron once individuos para participar en el estudio y midieron su capacidad de trabajo, entendida como el tiempo que se tarda en alcanzar una tasa de 160 latidos por minuto mientras se camina con una cierta velocidad sobre una cinta de andar, al comienzo del estudio y después de 25 semanas de ejercicio controlado. Los resultados, expresados en minutos, fueron los siguientes:

Sujeto	Antes	Después
1	7,6	14,7
2	9,9	14,1
3	8,6	11,8
4	9,5	16,1
5	8,4	14,7
6	9,2	14,1
7	6,4	13,2
8	9,9	14,9
9	8,7	12,2
10	10,3	13,4
11	8,3	14,0

¿Sostienen estos datos el argumento de los investigadores?

2. Se acepta generalmente que existen diferencias ligadas al sexo relacionadas con la respuesta a la tensión producida por el calor. Para comprobarlo, se sometió a un grupo de 10 hombres y 8 mujeres a un programa de ejercicios duros, en un medio con temperatura alta (40 °C) y sin posibilidad de beber. La variable de interés medida fue el porcentaje de peso corporal perdido. Se obtuvieron los datos siguientes:

Varones		Mujeres	
2,9	3,7	3,0	3,8
3,5	3,8	2,5	4,1
3,9	4,0	3,7	3,6
3,8	3,6	3,3	4,0
3,6	3,7		

Según los datos recogidos, ¿hay diferencias en el porcentaje medio de peso corporal perdido como respuesta al ejercicio físico desarrollado en alta temperatura entre hombres y mujeres?

Análisis de la Varianza de un Factor

11.1. Ejercicios resueltos

1. Se realiza un estudio para comparar la eficacia de tres programas terapéuticos para el tratamiento del acné. Se emplean tres métodos:
 - a) Lavado, dos veces al día, con cepillo de polietileno y un jabón abrasivo, junto con el uso diario de 250 mg de tetraciclina.
 - b) Aplicación de crema de tretinoína, evitar el sol, lavado dos veces al día con un jabón emulsionante y agua, y utilización dos veces al día de 250 mg de tetraciclina.
 - c) Evitar el agua, lavado dos veces al día con un limpiador sin lípidos y uso de crema de tretinoína y de peróxido benzoílico.

En el estudio participan 35 pacientes. Se separó aleatoriamente a estos pacientes en tres grupos de tamaños 10, 12 y 13, a los que se asignó respectivamente los tratamientos I, II, y III. Después de 16 semanas se anotó para cada paciente el porcentaje de mejoría en el número de lesiones.

Tratamiento					
I		II		III	
48,6	50,8	68,0	71,9	67,5	61,4
49,4	47,1	67,0	71,5	62,5	67,4
50,1	52,5	70,1	69,9	64,2	65,4
49,8	49,0	64,5	68,9	62,5	63,2
50,6	46,7	68,0	67,8	63,9	61,2
		68,3	68,9	64,8	60,5
				62,3	

- a) Crear las variables tratamiento y mejora e introducir los datos de la muestra.

i

Para cualquier ANOVA, a pesar de que la variable tratamiento es cualitativa, conviene que se introduzca como cuantitativa, ya que SPSS sólo admite como factor de clasificación variables cuantitativas. Si hubiese que mostrar los diferentes niveles de la variable factor como cualidades, bastará con asignarles etiquetas).

- b) Dibujar el diagrama de dispersión. ¿Qué conclusiones sacas de la nube de puntos?

i

- 1) Seleccionar el menú Gráficos > Cuadros de diálogo antiguos > Dispersión/Puntos.
- 2) En el cuadro de diálogo que aparece, seleccionar la opción Dispersión simple y hacer click en el botón Definir.
- 3) En el siguiente cuadro de diálogo que aparece, seleccionar la variable mejora al campo Eje Y y la variable tratamiento al campo Eje X, y hacer click sobre el botón Aceptar.

- c) Obtener la tabla de ANOVA correspondiente al problema. ¿Se puede concluir que los tres tratamientos tienen el mismo efecto medio con un nivel de significación de 0,05?

i

- 1) Seleccionar el menú Analizar > Comparar medias > ANOVA de un factor.

2) En el cuadro de diálogo que aparece, seleccionar la variable mejora al campo Lista de dependientes y la variable tratamiento al campo factor, y hacer click sobre el botón Aceptar.

d) Obtener la tabla de ANOVA correspondiente al problema pero que además muestre los intervalos de confianza de los 3 diferentes tratamientos, con una significación de 0,05, y diversos estadísticos de cada uno de ellos.

i Repetir los mismos pasos del apartado anterior, haciendo click en el botón Opciones del último cuadro de diálogo y activar la opción Descriptivos.

e) Si se puede concluir que los tratamientos no han tenido el mismo efecto medio, ¿entre qué parejas de tratamientos hay diferencias estadísticamente significativas?

i Repetir los mismos pasos del apartado anterior, haciendo click en el botón Post hoc del último cuadro de diálogo y activar la opción de Bonferroni con un nivel de significación de 0,05.

f) Si se puede concluir que los tratamientos no han tenido el mismo efecto medio, ¿cuáles son los grupos homogéneos (grupos con un comportamiento similar en cuanto a mejora se refiere) de tratamientos que se pueden establecer?

i Repetir los mismos pasos del apartado anterior, activando la opción de Duncan.

2. Se sospecha que hay diferencias en la preparación del examen de selectividad entre los diferentes centros de bachillerato de una ciudad. Con el fin de comprobarlo, de cada uno de los 5 centros, se eligieron 8 alumnos al azar, con la condición de que hubieran cursado las mismas asignaturas, y se anotaron las notas que obtuvieron en el examen de selectividad.

Los resultados fueron:

Centros				
1	2	3	4	5
5,5	6,1	4,9	3,2	6,7
5,2	7,2	5,5	3,3	5,8
5,9	5,5	6,1	5,5	5,4
7,1	6,7	6,1	5,7	5,5
6,2	7,6	6,2	6,0	4,9
5,9	5,9	6,4	6,1	6,2
5,3	8,1	6,9	4,7	6,1
6,2	8,3	4,5	5,1	7,0

- Crear las variables nota y centro e introducir los datos de la muestra.
- Dibujar el diagrama de dispersión. ¿Qué conclusiones sacas sobre la nota media de selectividad en los distintos centros?

i

- 1) Seleccionar el menú Gráficos > Cuadros de diálogo antiguos > Dispersión/Puntos.
- 2) En el cuadro de diálogo que aparece, seleccionar la opción Dispersión simple y hacer click en el botón Definir.
- 3) En el siguiente cuadro de diálogo que aparece, seleccionar la variable nota al campo Eje Y y la variable centro al campo Eje X, y hacer click sobre el botón Aceptar.

- Realizar el contraste de ANOVA. ¿Se puede confirmar la sospecha de que hay diferencias entre las notas medias de los centros?

i

- 1) Seleccionar el menú Analizar > Comparar medias > ANOVA de un factor.
- 2) En el cuadro de diálogo que aparece, seleccionar la variable nota al campo Lista de dependientes y la variable centro al campo factor, y hacer click sobre el botón Aceptar.

- ¿Qué centros son los mejores en la preparación de la selectividad?

i

Repetir los mismos pasos del apartado anterior, haciendo click en el botón Post hoc del último cuadro de diálogo y activar las opciones de Bonferroni, para ver los intervalos de diferencias entre centros, y de Duncan para establecer grupos de comportamiento homogéneo.

11.2. Ejercicios propuestos

1. Se midió la frecuencia cardíaca (latidos por minuto) en cuatro grupos de adultos; controles normales (A), pacientes con angina (B), individuos con arritmias cardíacas (C) y pacientes recuperados del infarto de miocardio (D). Los resultados son los siguientes:

A	B	C	D
83	81	75	61
61	65	68	75
80	77	80	78
63	87	80	80
67	95	74	68
89	89	78	65
71	103	69	68
73	89	72	69
70	78	76	70
66	83	75	79
57	91	69	61

¿Proporcionan estos datos la suficiente evidencia para indicar una diferencia en la frecuencia cardiaca media entre esos cuatro tipos de pacientes? Considerar $\alpha = 0,05$.

2. Se midió la frecuencia respiratoria (inspiraciones por minuto) en ocho animales de laboratorio y con tres niveles diferentes de exposición al monóxido de carbono. Los resultados son los siguientes:

Nivel de exposición		
Bajo	Moderado	Alto
36	43	45
33	38	39
35	41	33
39	34	39
41	28	33
41	44	26
44	30	39
45	31	29

Con base en estos datos, ¿es posible concluir que los tres niveles de exposición, en promedio, tienen un efecto diferente sobre la frecuencia respiratoria? Tomar $\alpha = 0,05$.

ANOVA de Dos o Más Factores, ANOVA de Medidas Repetidas y ANCOVA

12.1. Ejercicios resueltos

1. En un estudio diseñado para analizar la influencia de un tipo de dieta y de un fármaco en el peso corporal perdido, expresado en Kg, se ha anotado el número de Kg perdidos en un grupo de personas al cabo de 3 meses de dieta y de tomar el fármaco, obteniendo los siguientes resultados (si algún individuo presenta un dato negativo significa que en lugar de perder Kg de peso los ha ganado):

	Fármaco No	Fármaco Sí
Dieta No	1,5; 0,5; 0,0; -1,0; -1,0	6,5; 5,0; 7,0; 3,0; 4,5; 5,0
Dieta Sí	3,5; 3,0; 4,0; 2,5; 2,0	9,5; 8,0; 7,5; 7,0; 8,5; 7,5

- a) Crear la variables Kilosperdidos, Dieta y Fármaco e introducir los datos en PASW.

i

Al introducir los datos, tener en cuenta que tanto la Dieta como el Fármaco son variables cualitativas pero hay que introducirlas como numéricas. Posteriormente se puede asignar una etiqueta a los valores numéricos introducidos.

- b) Realizar un ANOVA de dos vías con los datos e interpretar la tabla de ANOVA obtenida.

i

Para realizar el ANOVA propuesto, utilizar el menú Analizar > Modelo lineal general > Univariante y utilizar como Variable dependiente la variable Kilos perdidos y como Factores fijos las variables Dieta y Fármaco. Por último pinchar en el botón Aceptar.

Para la interpretación de la tabla de ANOVA, prestar especial atención a las siguientes líneas de la tabla:

- 1) Modelo Corregido: muestra si el modelo planteado para explicar el peso perdido en función de la dieta, el fármaco y la interacción de ambas es o no significativo globalmente, es decir, si es o no capaz de explicar un porcentaje considerable de la variabilidad de la variable dependiente. De hecho, en la parte baja de la tabla aparece el R cuadrado del modelo (porcentaje de variabilidad explicado).
- 2) Dieta: muestra si la dieta resulta o no significativa para explicar la variabilidad del peso perdido.
- 3) Fármaco: muestra si el fármaco resulta o no significativo.
- 4) Dieta x Fármaco: muestra si la interacción de dieta y fármaco resulta o no significativa.

Una conclusión muy importante a la luz de los resultados es que no hay una interacción significativa entre dieta y fármaco, es decir que el efecto del fármaco no dependerá de si una persona toma o no dieta, y a la inversa, que el efecto de la dieta no dependerá de si se toma o no fármaco.

- c) Mostrar el gráfico de medias correspondiente al anterior ANOVA. ¿Qué conclusiones cualitativas pueden sacarse del gráfico obtenido?

i

Para mostrar el gráfico de medias, que PASW llama gráfico de perfiles, repetir los pasos del anterior punto para realizar el ANOVA de dos vías pero antes de Aceptar pinchar en el botón Gráficos y pasar al campo Eje horizontal, por ejemplo, la variable Dieta, y al campo Líneas separadas la variable Fármaco. Por último pinchar en el botón Añadir y Continuar. Igualmente se

podría obtener un gráfico de medias con la variable Fármaco en el Eje horizontal y la Dieta en Líneas separadas.

Las conclusiones cualitativas: se observa claramente que no hay interacción (líneas paralelas), que los dos puntos del grupo de los que no hacen dieta están por debajo de los que sí que la hacen, lo cual hace sospechar que el factor dieta será significativo, e igualmente los dos puntos de los que no toman fármaco están por debajo de los que sí que lo toman, lo cual hace sospechar que el factor fármaco también será significativo.

- d) Calcular las medias y desviaciones típicas de los Kg perdidos en todos los grupos, tanto las reales como las estimadas por el modelo.

i

El procedimiento estadístico Modelo Lineal General Univariante de PASW, que es el que se está aplicando en este ejercicio, da la posibilidad de mostrar dos tipos diferentes de medias junto con sus desviaciones típicas:

- Las medias *reales*, observadas o empíricas de cada uno de los grupos, teniendo en cuenta que en el problema se pueden calcular: la media global, las dos de los grupos que se forman con la variable dieta, las dos de los que se forman con la variable fármaco y las 4 que se forman con el cruce de dieta y fármaco.
- Las medias que el programa llama *estimadas*, que no son las reales sino las estimadas a partir de los parámetros del modelo que el programa genera cada vez que le pedimos que realice un ANOVA de varias vías. Internamente, para PASW un ANOVA de varias vías es un modelo de regresión multivariante en el que la variable dependiente se expresa como una función lineal de las diferentes variables independientes. Dicho modelo genera unos coeficientes de la combinación lineal que permiten estimar cuáles serían las medias de los diferentes grupos, con la ventaja de que estas medias llevarían eliminado el efecto que las demás variables, interacciones, o incluso covariables, podrían estar

introduciendo. Por ejemplo, si se plantea calcular la media del grupo de los que siguen dieta y en dicho grupo da la casualidad de que prácticamente todos los individuos toman fármaco, su valor será más grande del que de verdad le correspondería si el grupo hubiese estado equilibrado (tantos que tomen fármaco como que no lo tomen). Esa media calculada eliminando la influencia del fármaco sería la estimada para el grupo de los que hacen dieta, y así cualquier otra. En realidad, siempre que aparezcan interacciones significativas, covariables significativas o los grupos que surjan de cruzar las variables no estén equilibrados (en nuestro problema hay dos grupos de 5 y dos de 6), se recomienda trabajar con las medias estimadas.

Para obtener un resumen estadístico con las medias y las desviaciones típicas reales: *Analizar > Modelo lineal general > Univariante*, en el botón *Opciones*, en la parte de *Visualización* activar *Estadísticos descriptivos* y pinchar en el botón *Continuar*. Las medias y desviaciones típicas se podrían obtener igualmente segmentando el archivo y aplicando los procedimientos oportunos de estadística descriptiva.

Para obtener un resumen estadístico con las medias estimadas: *Analizar > Modelo lineal general > Univariante*, y en el botón *Opciones*, en la parte de *Medias marginales estimadas*, dentro de *Factores e interacciones de los factores* pasar toda la lista a *Mostrar las medias para* y pinchar en el botón *Continuar*.

- e) Teniendo en cuenta que no hay interacción significativa, calcular los intervalos de confianza para las medias estimadas de los dos grupos que se forman con la variable dieta y el intervalo de confianza para la diferencia, todos con un 95 % de confianza. Igualmente con la variable fármaco.

i

Como la interacción no ha sido significativa, sólo cabe plantearse si hay diferencias significativas entre los que hacen dieta y los que no, y entre los que toman fármaco y los que no,

pero en ningún caso habría que cruzar las variables (diferencias entre los que hacen dieta y los que no dentro del grupo de los que toman fármaco, diferencias entre los que hacen dieta y los que no dentro del grupo que no toma fármaco, diferencias entre los que toman fármaco y los que no dentro de los que hacen dieta o diferencias entre los que toman fármaco y los que no dentro de los que no hacen dieta). Además, ya sabemos que como tanto la dieta como el fármaco han resultado variables significativas, debe haber diferencias entre dieta sí y dieta no, y también entre fármaco sí y fármaco no, pero ahora se pide que se cuantifique el intervalo de confianza para la diferencia. Por otra parte, como los grupos no están equilibrados y además han salido significativas las dos variables, conviene trabajar con las medias estimadas en lugar de las reales. Afortunadamente, PASW junto con las medias estimadas también puede dar sus correspondientes intervalos de confianza tanto de forma individual como para las diferencias. Para ello: Analizar > Modelo lineal general > Univariante, y en el botón Opciones, en la parte de Medias marginales estimadas, dentro de Factores e interacciones de los factores pasar Dieta y Fármaco a Mostrar las medias para, activar la casilla Comparar los efectos principales y como método de Ajuste del intervalo de confianza podemos escoger Bonferroni (en realidad, en este ejemplo no hay necesidad de ajustar los intervalos porque sólo hay dos categorías en cada variable, pero si hubiese tres o más, en lugar de una comparación se podrían realizar bastantes más y habría que controlar la posibilidad de un error de tipo I, por ejemplo, con el método de Bonferroni). Por último, pinchar en el botón Continuar.

2. En un estudio diseñado para analizar la influencia de un tipo de dieta y de un fármaco en el peso corporal perdido, expresado en Kg, se ha anotado el número de Kg perdidos en un grupo de personas al cabo de 3 meses de dieta y de tomar el fármaco, obteniendo los siguientes resultados (si algún individuo presenta un dato negativo significa que en lugar de perder Kg de peso los ha ganado):

	Fármaco No	Fármaco Sí
Dieta No	1,5; 0,5; 0,0; -1,0; -1,0	6,5; 5,0; 7,0; 3,0; 4,5; 5,0
Dieta Sí	3,5; 3,0; 4,0; 2,5; 2,0	12,5; 12,0; 11,5; 13,5; 12,5; 10,0

- a) Crear la variables Kilosperdidos, Dieta y Fármaco e introducir los datos en PASW.

i Seguir el mismo proceso del apartado a del ejercicio anterior.

- b) Realizar un ANOVA de dos vías con los datos e interpretar la tabla de ANOVA obtenida. ¿Hay interacción significativa? ¿Cómo se interpretaría?

i Para obtener la tabla de ANOVA, seguir el mismo proceso del apartado b del ejercicio anterior.

Sí que hay interacción significativa y eso implica que no hay la misma diferencia en Kg perdidos entre los que hacen dieta y los que no si consideramos el grupo de los que no toman fármaco que si consideramos el grupo de los que sí que lo toman.

- c) Mostrar el gráfico de medias correspondiente al anterior ANOVA. ¿Qué conclusiones cualitativas pueden sacarse del gráfico obtenido?

i Para obtener el gráfico de medias, seguir el mismo proceso del apartado c del ejercicio anterior.

La interpretación del gráfico es que ahora se ve claramente que además de ser la dieta y el fármaco significativos, también lo es la interacción (las líneas que unen las medias tienen pendientes muy diferentes).

- d) ¿Entre qué medias sería adecuado buscar diferencias significativas en este problema? Calcular las medias, junto con sus intervalos de confianza con un 95 %, tanto de las medias como de sus diferencias.

i

Por tratarse de un problema en el que hay interacción entre dieta y fármaco, se debería ver si hay diferencias entre dieta sí y dieta no en el grupo de los que toman fármaco, y entre dieta sí y dieta no en el grupo de los que no lo toman. Igualmente se podría analizar la diferencia entre fármaco sí y fármaco no en el grupo de los que hacen dieta, y entre fármaco sí y fármaco no en el grupo de los que no la hacen. Trabajando con las medias estimadas: Analizar > Modelo lineal general > Univariante, y en el botón Opciones, en la parte de Medias marginales estimadas, dentro de Factores e interacciones de los factores pasar toda la lista a Mostrar las medias para, activar la casilla Comparar los efectos principales y escoger como Método de ajuste del intervalo de confianza Bonferroni.

Pero se puede observar en el resultado que aparece que PASW, por defecto, no realiza nada más que las comparaciones en los factores principales, que por otra parte no son las que interesan en este problema ya que hay interacción. Para que realice las comparaciones debidas a la interacción no queda más remedio que obligar al programa a realizarlas mediante comandos. Para ello, hay que generar el *Archivo de Sintaxis* (conjunto de órdenes) del procedimiento estadístico con el que se está trabajando y añadirle los comandos adecuados.

Para generar el archivo de sintaxis, en lugar de pinchar en el botón Aceptar al final de todo el proceso, se pincha en el botón Pegar. Eso hace que se abra el Editor de Sintaxis, con un conjunto de comandos que son los que dan las órdenes adecuadas al programa para que ejecute los procesos pedidos. Para que realice la comparación dentro de los grupos generados mediante el cruce de variables, hay que retocar la línea en la que muestra sus medias, /EMMEANS=TABLES(Dieta*Fármaco), que debe incluir: COMPARE(Fármaco) ADJ(BONFERRONI). Su aspecto final debe ser:

```
/EMMEANS=TABLES(Dieta*Fármaco) COMPARE(Fármaco) ADJ(BONFERRONI)
```

Y también hay que añadir una línea más justo a continuación de la anterior:

```
/EMMEANS=TABLES(Fármaco*Dieta) COMPARE(Dieta) ADJ(BONFERRONI)
```

Una vez hechos los cambios, se marca todo el texto arrastrando el ratón y se ejecutan los comandos seleccionados con el menú *Ejecutar > Selección*. Con ello, se obtienen todas las comparaciones junto con sus intervalos de confianza.

3. Se ha realizado un experimento que consiste en que se ha anotado el tiempo, en días, que han tardado en contestar correctamente a un cuestionario 30 personas, 15 hombres y 15 mujeres, distribuidos en grupos que han seguido tres métodos diferentes de aprendizaje de la materia del cuestionario. Los resultados aparecen en la siguiente tabla:

	Método a	Método b	Método c
Hombre	15, 16, 18, 19, 14	25, 27, 28, 23, 29	21, 22, 18, 17, 20
Mujer	24, 27, 29, 25, 23	17, 15, 13, 16, 18	20, 19, 22, 17, 23

- Crear las variables Sexo, Método y Días, e introducir los datos.
- Generar la tabla de ANOVA de dos vías con interacción e interpretar los resultados.

i

Utilizar el menú *Analizar > Modelo lineal general > Univariante* y utilizar como Variable dependiente la variable Días y como Factores fijos las variables Sexo y Método. Por último pinchar en el botón *Aceptar*.

Interpretación: no hay diferencias significativas asociadas al sexo ni al método. Sin embargo sí que hay interacción, es decir la diferencia en el tiempo de respuesta entre hombres y mujeres depende del método seguido, e igualmente las diferencias entre los tiempos de respuesta según los diferentes métodos dependen del sexo.

- Generar dos gráficos de medias: un primero en el que en el eje horizontal aparezca el Método y con líneas distintas según Sexo, y un segundo en el que en el eje horizontal aparezca el Sexo y líneas distintas para el Método. ¿Qué se puede decir de la interacción de las variables?

i

Repetir los pasos del anterior punto para realizar el ANOVA de dos vías pero antes de Aceptar pinchar en el botón Gráficos y pasar al campo Eje horizontal la variable Método, y al campo Líneas separadas la variable Sexo y se pincha en el botón Añadir. Para el siguiente gráfico, sin necesidad de repetir todo el proceso, se pasa al Eje horizontal la variable Sexo y al campo Líneas separadas la variable Método y se pincha en el botón Añadir. Posteriormente se pincha en Continuar y Aceptar.

Interpretación: es evidente que las líneas se cruzan, lo cual indica que hay interacción.

- d) Calcular las medias estimadas y sus correspondientes intervalos de confianza, además de analizar entre qué medias hay diferencias y sus correspondientes intervalos de confianza para la diferencia. Trabajar con un 95 % de confianza.

i

De nuevo, las medias que de verdad interesan son las que surgen del cruce de los dos factores. Para poder ver la diferencia entre hombre y mujeres, tanto en el método a como el b y el c; y para poder ver la diferencia entre los métodos a y b, a y c y b y c, tanto en el grupo de hombres como en el de mujeres, acudimos de nuevo a las medias estimadas y hacemos las comparativas: Analizar > Modelo lineal general > Univariante, y en el botón Opciones, en la parte de Medias marginales estimadas, dentro de Factores e interacciones de los factores pasar toda la lista a Mostrar las medias para, activar la casilla Comparar los efectos principales y escoger como Método de ajuste del intervalo de confianza Bonferroni. Posteriormente se genera el archivo de sintaxis, pinchando en el botón Pegar en lugar de Aceptar. Eso hace que se abra el Editor de Sintaxis, donde, para que realice la comparación dentro de los grupos generados mediante el cruce de variables, hay que retocar la línea en la que muestra sus medias, /EMMEANS=TABLES(Sexo*Método), que debe incluir: COMPARE(Método) ADJ(BONFERRONI). Su aspecto final debe ser:

```
/EMMEANS=TABLES(Sexo*Método) COMPARE(Método)  
ADJ(BONFERRONI)
```

Y también hay que añadir una línea más justo a continuación de la anterior:

```
/EMMEANS=TABLES(Método*Sexo) COMPARE(Sexo) ADJ(BONFERRONI)
```

Una vez hechos los cambios, se marca todo el texto arrastrando el ratón y se ejecutan los comandos seleccionados con el menú *Ejecutar > Selección*. Con ello, se obtienen todas las comparaciones junto con sus intervalos de confianza.

4. Se ha realizado un experimento para comprobar si los Kg perdidos por un grupo de personas dependen o no de la dieta utilizada durante 3 meses. Para ello se ha trabajado con dos dietas diferentes y se ha anotado el número de Kg perdidos y el índice de masa corporal (IMC) de los individuos. Los resultados aparecen en la siguiente tabla:

Kg perdidos	Dieta	IMC
4,3	1	24
1,8	1	20
6,1	1	25
2,3	1	22
1,1	1	21
0,5	1	21
7,2	1	27
3,2	1	24
9,1	2	30
5,2	2	28
7,3	2	29
10,3	2	32
5,1	2	24
7,6	2	28
6,2	2	27
4,1	2	25

- a) Crear las variables Kgperdidos, Dieta y IMC e introducir los datos de la tabla.

- b) Sin tener en cuenta la variable IMC, ¿se puede considerar que la variable dieta influye significativamente en la cantidad de Kg perdidos?

i

Como en este caso sólo hay dos categorías de la variable dieta, se puede hacer la comparación mediante una T de Student de datos independientes: Analizar > Comparar Media > Prueba T para muestras independientes, la Variable para contrastar son los Kg perdidos y la Variable de agrupación la Dieta, en la que habrá que definir adecuadamente los grupos. A la vista de los resultados, con un 95 % de confianza se puede afirmar que la dieta influye en la cantidad de Kg perdidos.

La T de Student de datos independientes tiene la ventaja de que da directamente el intervalo para la diferencia. No obstante, como la pregunta es si la dieta influye significativamente en la cantidad de Kg perdidos y no se piden los intervalos de confianza, para realizar el contraste de igualdad de las dos medias también se puede hacer un ANOVA de una vía mediante: Analizar > Comparar medias > ANOVA de un factor, o incluso a través del procedimiento Analizar > Modelo lineal general > Univariante, que en principio está diseñado para realizar ANOVAS de varias vías pero también se puede utilizar para un único factor.

- c) Representar la nube de puntos de la cantidad de Kg perdidos en función del IMC coloreando los puntos según la Dieta seguida. A la vista del gráfico, ¿la cantidad de Kg perdidos depende de la dieta o más bien del IMC del individuo?

i

Como cualquier otro diagrama de dispersión, se puede generar mediante: Gráficos > Cuadros de diálogo antiguos > Dispersión/Puntos, escoger Dispersión simple, pinchar en Definir, escoger como Eje Y la variable Kg perdidos, como Eje X el IMC y Establecer marcas por la variable Dieta.

Interpretación: se ve claramente que todos los puntos, ya sean los de la dieta 1 o la 2, aparecen en una nube que se puede

ajustar con una única recta de pendiente no nula que marca la influencia de la covariable IMC, pero eliminado el efecto del IMC, es decir, si arrastramos los puntos hacia abajo para dejar la recta con pendiente 0, todos los puntos quedarían a una altura similar, por lo que no habría influencia de la dieta.

- d) Realizar un ANCOVA con la variable dependiente Kgperdidos, el factor dieta y la covariable IMC. Interpretar los resultados obtenidos.

i

Para generar la tabla del ANOVA con la covariable añadida, Analizar > Modelo lineal general > Univariantes, seleccionar como Variable dependiente Kgperdidos, como Factores Fijos la Dieta y como Covariables el IMC.

Interpretación: El número de Kg perdidos no depende significativamente de la dieta y es simplemente la covariable la que hace que parezca que los individuos que siguen la dieta 1 pierdan más Kg que los que siguen la dieta 2.

- e) Mostrar las medias reales y las estimadas. ¿Hay diferencias estadísticamente significativas entre las estimadas?

i

Para mostrar las medias reales, se pueden seguir múltiples caminos dentro del programa; entre otros, repetir los pasos dados en el punto anterior, pinchar en el botón Opciones y en la parte Visualización escoger Estadísticos descriptivos.

Para mostrar las medias estimadas (en las que se ha eliminado el efecto de la covariable) seguir los mismos pasos pero, después de pinchar en el botón Opciones, ir a la parte de Medias marginales estimadas, escoger en Factores e interacciones de los factores tanto OVERALL como Dieta y pasar al campo Mostrar las medias para. Por último, activar Comparar los efectos principales y como método de Ajuste del intervalo de confianza vale cualquiera porque en este caso sólo hay dos niveles (habitualmente se puede escoger Bonferroni).

Interpretación: Las medias muestrales reales de los dos grupos y sus correspondiente estimadas son claramente diferentes. Al eliminar el efecto de la covariable incluso la media muestral estimada de los que toman la dieta 2 es menor que la del grupo 1 cuando en las reales era justo al revés. No obstante, entre las estimadas no hay diferencias estadísticamente significativas (el intervalo para la diferencia contiene al 0) con un 95 % de confianza.

- f) Repetir el ANCOVA con la comparación de medias, tanto reales como estimadas, pero suponiendo que los Kg perdidos por los que han seguido la dieta 1 son: 6,3; 3,8; 8,1; 4,3; 3,1; 4,3; 8,6; 5,2. Obtener también el diagrama de dispersión con el código de colores según dieta.

i Seguir exactamente los mismos pasos que en los puntos anteriores del ejercicio.

5. Se desea comparar la rapidez con la que aparece el efecto de tres nuevos agentes repigmentadores: A, B y C. Con esta intención, se aplican de manera tópica dosis equivalentes de los tres repigmentadores en zonas de la piel con pérdida total de pigmentación en los mismos ocho pacientes con vitíligo. A continuación, se recoge el tiempo, en días, que tardan en aparecer los primeros signos de repigmentación:

A	B	C
19	3	31
11	2	9
7	4	16
4	1	6
3	2	8
5	7	18
7	1	5
4	3	9

- a) Crear las variable A, B y C e introducir los datos en PASW conforme se dan en la tabla anterior (cada individuo en una fila y las 3 medidas diferentes en 3 variables).

- b) Realizar un ANOVA de medidas repetidas e interpretar el resultado obtenido.

i

PASW dispone de un procedimiento estadístico específico para la realización de ANOVAS de medidas repetidas cuando los datos se introducen en el formato clásico de una fila por individuo y una variable por cada una de las medidas repetidas. Para acceder al procedimiento: Analizar > Modelo lineal general > Medidas repetidas. En el campo Nombre del factor intra-sujetos se da nombre al factor que contendrá las medidas repetidas, que en este caso son tiempos expresados en días, por lo que se puede llamar Tiempo. En el campo Número de niveles hay que poner el número de medidas repetidas, es decir, el número de niveles del factor Tiempo que son 3. Posteriormente se pincha en el botón Añadir para añadir el factor a la lista de factores intra-sujetos (permite trabajar con varios) y en el botón Definir para dar las variables que forman los niveles del factor y en el campo Variables intra-sujetos introducimos A, B y C. Antes de pinchar en el botón Aceptar para obtener el ANOVA de medidas repetidas, se puede observar que el procedimiento estadístico permite incluir en el análisis factores inter-sujetos (que más bien son factores inter-grupos), por ejemplo el sexo, y covariables, por ejemplo la edad; por lo tanto, este procedimiento permite realizar el ANOVA más completo posible: varios factores intra-sujetos, varios factores inter-sujetos y varias covariables. Por último, se pincha en Aceptar y se obtiene el resultado del ANOVA.

Interpretación:

El resultado obtenido contiene varias tablas pero, esencialmente, las más importantes son:

- La tabla que el programa llama Prueba de los efectos intra-sujetos, que en su primera línea (Esfericidad Asumida) da el p -valor del contraste de igualdad de medias de todos los niveles del factor intra-sujetos. Esta tabla presentaría más líneas como la comentada si se hubiesen añadido más factores intra-sujeto al estudio, y también la inter-

acción entre los factores intra-sujeto con los factores inter-sujetos y las covariables, si es que los hubiese en el estudio considerado.

- La tabla que el programa llama Prueba de los efectos inter-sujetos, que en este caso no tiene ninguna línea importante ya que no se han añadido al estudio factores inter-sujeto ni covariables. Mostraría el p-valor de todos los factores inter-sujetos, de las covariables y de la interacción entre de los factores inter-sujetos entre sí.
- También muestra una tabla de Prueba de esfericidad de Mauchly que es la prueba con la que se contrastan los supuestos del modelo de ANOVA de medidas repetidas. Si el p-valor de la prueba sale mayor que el nivel de significación entonces no se puede rechazar la hipótesis nula de esfericidad de los datos (distribuciones normales en todos los grupos con homogeneidad de varianzas). Si se rechaza la esfericidad mejor que en el ANOVA de medidas repetidas convendría fijarse en el resultado de los Contrastes Multivariados que aparecen en otra tabla y que no necesitan esfericidad para poder ser calculados.

En el caso concreto del problema, el factor intra-sujetos definido, el tiempo, sí que resulta significativo, por lo que al menos una de las medidas repetidas tiene una media diferente de las demás.

- c) ¿Entre qué medidas del factor intra-sujetos se dan diferencias estadísticamente significativas?

i

Después de Analizar > Modelo lineal general > Medidas repetidas y definir el factor intra-sujetos, Tiempo, con 3 niveles, dentro del cuadro de diálogo que aparece (Medidas Repetidas), pinchando en el botón Opciones, en la parte de Medias marginales estimadas ir a la lista de Factores e interacciones de los factores y escoger OVERALL y Tiempo y pasarlas a Mostrar las medias para. Después, activar

la casilla Comparar los efectos principales y como Ajuste del intervalo de confianza utilizar Bonferroni.

Interpretación: en la tabla de Comparaciones por pares que aparece se ve que hay diferencias significativas entre los tiempos 2 y 3, es decir entre el tiempo en que comienzan a aparecer signos de repigmentación con el agente B y el C.

- d) Realizar el mismo ANOVA de medidas repetidas pero introduciendo los datos como se darían en un ANOVA de dos vías con la variable cuantitativa dependiente Tiempo y los factores cualitativos Repigmentador e Individuo.

i

Si se pretenden hacer un ANOVA de medidas repetidas como si fuese un ANOVA de dos vías sin interacción, lo primero es introducir los datos de tal forma que cada individuo ocupa tantas filas como medidas repetidas se han dado. Para ello se genera una variable Tiempo que contiene los 24 tiempos, una variable Repigmentador numérica con valores 1, 2 y 3 (a los que se pueden asignar las etiquetas de valor A, B y C), y una variable Individuo con valores desde el 1 al 8 repetidos 3 veces:

Tiempo	Repigmentador	Individuo
19	1	1
11	1	2
...
4	1	8
3	2	1
2	2	2
...
3	2	8
31	3	1
9	3	2
...
9	3	8

Una vez introducidos los datos en el formato adecuado, el ANOVA de dos vías sin interacción se hace con: Analizar

> Modelo lineal general > Univariante; escoger como Variable dependiente el Tiempo y como Factores fijos Repigmentador e Individuo, y pinchando en el botón Modelo escoger en Especificar modelo el tipo Personalizado, en Construir términos escoger como Tipo la opción Efectos principales (no debe haber interacción) y pasar Repigmentador e Individuo al campo Modelo. Por último pinchar en los botones Continuar y Aceptar. Con ello se obtiene la tabla del ANOVA en la que se puede ver que el p -valor obtenido para el factor inter-sujetos Repigmentador es el mismo que el obtenido para el factor intra-sujetos del ANOVA de medidas repetidas.

6. En un estudio sobre la memoria realizado con 12 personas se ha registrado el número de aciertos en una secuencia de números en tres periodos temporales distintos: tras una hora de entrenamiento, tras un día y tras una semana. De las 12 personas, 6 han seguido un método de entrenamiento y las otras 6 otro diferente. Además, se piensa que la edad de los individuos puede influir en el resultado y temiendo que los dos grupos de 6 individuos no tengan una distribución homogénea de edades, la misma se incluye como covariable. Los datos obtenidos han sido:

	Hora	Día	Semana	Edad
Método 1	4	5	7	43
	6	8	10	48
	1	6	5	56
	2	10	12	36
	5	10	10	35
	1	7	8	51
Método 2	1	2	4	59
	3	6	6	46
	3	5	4	53
	1	4	7	45
	5	6	5	49
	2	8	7	39

- a) Crear las variables Método (se introduce como numérica pero es cualitativa, aunque se le pueden asignar etiquetas de valor), Hora, Día, Semana y Edad e introducir los datos en PASW.

- b) Realizar un ANOVA de medidas repetidas pero teniendo en cuenta el factor inter-sujetos Método y también la covariable Edad. Interpretar el resultado obtenido.

i

Para obtener las tablas ANOVA de medidas repetidas: Analizar > Modelo lineal general > Medidas repetidas. En el campo Nombre del factor intra-sujetos se puede poner cualquiera, por ejemplo aprovechando las iniciales de hora, día y semana que son los momentos en los que se realizan las diferentes medidas se puede llamar Hds; su Número de niveles es 3; se pincha en el botón Añadir y luego en Definir para escoger Variables intra-sujetos que son Hora, Día y Semana; en Factores inter-sujetos la variable Método y en Covariables la Edad. Por último, se pincha en Aceptar.

Interpretación de las tablas:

Mirando en la tabla de Pruebas de efectos intra-sujetos se ve que el factor Hds, con esfericidad asumida, ha resultado significativo. Es decir, hay una evolución en el número de aciertos según el tiempo; en definitiva, hay alguna media (ya sea la de Hora, Día o Semana) cuya diferencia con alguna de las demás es significativa. En la misma tabla se observa que no hay interacción significativa entre Hds y Edad ni entre Hds y Método. Por lo tanto esa evolución significativa, esa diferencia en el número de aciertos entre los niveles del factor intra no depende ni del Método ni de la Edad.

Mirando en la tabla de Pruebas de efectos inter-sujetos se ve que tanto la Edad como el Método han resultado significativos. Es decir, el número de aciertos tiene una correlación no nula con la Edad, y además también hay diferencias entre el número de aciertos de los que han seguido un método de aprendizaje y los que han seguido otro.

- c) ¿Entre qué medias estimadas hay diferencias significativas?

i

Para ver entre qué medias estimadas hay diferencias significativas, después de definir los factores intra-sujetos, inter-sujetos

y las covariables, se pincha en el botón Opciones, en Medias marginales estimadas ir a Factores e interacciones de los factores y escoger Método y Hds (también se puede escoger la global, OVERALL, y las que surgen de la interacción, pero la interacción no ha resultado significativa) y pasarlas a Mostrar las medias para. Posteriormente activar Comparar los efectos principales y como método de Ajuste del intervalo de confianza escoger Bonferroni.

Interpretación: Al analizar la comparativa de medias estimadas se observa que el Método 1 logra una cantidad significativamente mayor que el 2. Además también hay diferencias entre los aciertos que se obtienen entre una hora de entrenamiento y al cabo de un día y una semana, pero no hay diferencia significativa entre los resultados al cabo de un día y al cabo de una semana.

- d) Obtener todos los gráficos de medias estimadas posibles en el problema.

i

En el problema se pueden obtener hasta 4 gráficos de medias:

- El correspondiente a las medias estimadas en ambos métodos de entrenamiento. Con este gráfico se puede comprobar cualitativamente que hay diferencias entre los dos métodos.
- El correspondiente a las medias estimadas de Hds (Horas, Días, Semanas). Con este gráfico se puede comprobar cualitativamente que hay diferencias entre las medias de las 3 medidas relacionadas diferentes.
- Las medias que surgen del cruce de Método con Hds, con el Método en el Eje X y líneas diferentes para Hds. Con este gráfico se puede comprobar cualitativamente que no hay interacción, es decir que las diferencias dentro del método 1 son parecidas a las que se dan en el método 2.
- Las medias que surgen del cruce de Método con Hds, con el Hds en el Eje X y línea diferentes para el Método. Con este gráfico se puede comprobar cualitativamente que no

hay interacción, es decir que las diferencias entre los dos métodos son parecidas al cabo de una hora, al cabo de un día y al cabo de una semana.

Para obtener los gráficos, después de definir los factores intra-sujetos, inter-sujetos y las covariables, se pincha en el botón Gráficos, se escoge el factor que va al Eje horizontal y el que genera las Líneas separadas (si es que las hay), y se pincha en el botón Añadir para ir añadiendo consecutivamente los gráficos a la lista. Cuando se han añadido los 4, se pincha en Continuar.

12.2. Ejercicios propuestos

1. En un estudio se quiere analizar la influencia sobre la ansiedad social, cuantificada mediante una escala numérica que va de 0 a 10, de la edad, dividida en tres categorías, y si se fuma o no. Los datos obtenidos fueron:

	Fumar No	Fumar Sí
Edad 1	3,91; 5,01; 4,47; 3,33; 4,71	4,83; 3,95; 4,04; 3,66; 9,44
Edad 2	5,65; 6,49; 5,50; 5,72; 5,44	9,66; 7,68; 9,57; 7,98; 7,39
Edad 3	4,94; 7,13; 5,54; 5,94; 6,16	5,92; 5,48; 5,19; 6,12; 4,45

- a) Considerando la posibilidad de interacción entre las variables independientes, ¿se puede considerar que la edad, expresada en forma de categorías, influye en la ansiedad? ¿Y el fumar? ¿Se puede considerar que el fumar o no influye de forma diferente en la ansiedad dependiendo de la categoría de edad analizada?
 - b) Dependiendo de los resultados del apartado anterior, ¿entre qué medias habría diferencias estadísticamente significativas? Calcular los intervalos de confianza para las diferencias.
2. Se investiga la influencia de tres técnicas terapéuticas sobre los días de hospitalización de una enfermedad infantil. Para ello, se han tomado tres muestras de 8 pacientes, a cada una de las cuales se ha aplicado uno de los tratamientos, y se anotan los correspondientes días que han permanecido ingresados. Se pretende comparar las medias producidas según las tres técnicas, pero se piensa que la edad del niño puede

influir en la estancia hospitalaria. Por ello, también se anota la edad del paciente. Los datos recogidos se presentan en la siguiente tabla:

Días Hospitalizado	Tratamiento	Edad
15	1	10
1	1	6
4	1	5
6	1	8
10	1	9
0	1	4
7	1	9
13	1	12
6	2	4
13	2	8
5	2	8
18	2	8
9	2	6
7	2	11
15	2	10
15	2	9
14	3	7
9	3	8
16	3	7
7	3	3
13	3	6
18	3	8
13	3	6
6	3	8

- a) Sin tener en cuenta la edad del paciente, ¿se podría concluir que el tiempo de hospitalización depende del tratamiento?
 - b) Considerando el efecto, si es que lo hay, de la edad del paciente, ¿se podría concluir que el tiempo de hospitalización depende del tratamiento?
3. En un estudio se quiere analizar la eficacia de dos tipos de entrenamiento (A1: entrenamiento sólo físico, A2: entrenamiento físico + entrenamiento psicológico) para mejorar el rendimiento físico. Para ello,

se dispone de una muestra de 8 individuos con los que se generan dos grupos de 4 asignados aleatoriamente, y se mide su rendimiento físico mediante un test de rendimiento numérico que va de 0 a 15 puntos. Los 8 individuos son sometidos al test en 4 momentos diferentes (B1: al cabo de una semana de entrenamiento, B2: al cabo de dos, B3: al cabo de tres y B4: al cabo de 4). Los datos obtenidos fueron:

	B1	B2	B3	B4
A1	3	4	7	7
	6	5	8	8
	3	4	7	9
	3	3	6	8
A2	1	2	5	10
	2	3	6	10
	2	4	5	9
	2	3	6	11

- a) ¿Influye significativamente la semana en la que se realiza el test en el resultado? ¿Y el tipo de entrenamiento? ¿Es significativa la interacción entre tipo de entrenamiento y la semana en la que se realiza el test?
- b) Entre qué medias hay diferencias estadísticamente significativas? Calcular los intervalos de confianza para las medias y para las diferencias.
4. Se ha aplicado un dispositivo electrónico que mide la frecuencia cardíaca a 10 estudiantes. Se realizó una primera medición un minuto antes de que comenzasen a hacer un examen, la segunda medición se hizo cuando llevaban 15 minutos realizando el examen, la tercera un minuto después de entregarlo y la cuarta 15 minutos después de terminar. Los resultados fueron:

ANOVA de Dos o Más Factores, ANOVA de Medidas Repetidas y ANCOVA

Estudiante	Medida1	Medida2	Medida3	Medida4
1	57	61	77	70
2	73	87	88	83
3	75	89	89	65
4	75	60	67	68
5	77	87	67	67
6	88	96	84	55
7	89	65	89	60
8	101	80	77	60
9	103	85	76	66
10	107	73	69	60

¿Son las mediciones significativamente distintas entre sí? Si hay diferencia, ¿entre qué mediciones se dan?

Contrastes de Hipótesis No Paramétricos

13.1. Ejercicios resueltos

1. Procedemos a lanzar una moneda 30 veces y obtenemos la siguiente secuencia C (cara) y X (Cruz):

X-C-X-C-X-C-X-X-C-C-X-C-C-X-C-X-C-X-X-C-X-C-X-C-X-C-C-X-X-X

Se pide:

- a) Crear las variables Secuencia e introducir los datos de la muestra.

i

Para comprobar, como se pedirá en el siguiente apartado, si los datos de una secuencia siguen o no una distribución aleatoria, es necesario definir la variable como numérica e introducir un número diferente para cada categoría. Por ejemplo, podemos introducir el valor 1 para la cruz X, y el valor 2 para la cara C. Posteriormente, se le puede asignar una etiqueta de valor a cada número.

- b) Comprobar si sería correcto suponer esa secuencia como aleatoria.

i

- 1) Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadro de diálogos antiguo ▶ Rachas.
- 2) En el cuadro de diálogo que aparece, seleccionar la variable Secuencia y pasarla al campo Contrastar Variables. En el campo Punto de corte seleccionar la opción Personalizado, y esco-

ger el valor 1,5 si se ha introducido la X como 1 y la C como 2 (si se han introducido otros valores numéricos, escoger como Punto de corte personalizado la media de los dos valores introducidos), y por último hacer click sobre el botón Aceptar.

c) ¿Qué interpretación darías al p valor obtenido?

2. Una central de productos lácteos recibe diariamente la leche de dos granjas X e Y. Para analizar la calidad de la leche, durante una temporada, se controla el contenido de materia grasa de la leche que proviene de ambas granjas, con los siguientes resultados:

X		Y	
0,34	0,34	0,30	0,29
0,32	0,33	0,30	0,32
0,33	0,30	0,30	0,32
0,31	0,32	0,29	0,31
		0,32	0,31
		0,32	0,33

- a) Crear las variables Grasa y Granja, e introducir los datos de la muestra.
 b) Comprobar la hipótesis de normalidad de los datos en cada grupo.

i

Para realizar el contraste de normalidad vamos a utilizar el contraste de Kolmogorov-Smirnov-Lilliefors. Para ello:

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar . . .
- 2) Seleccionar en Lista de dependientes la variable Grasa, y en Lista de factores la variable Granja. En Visualización debe estar activa la opción Gráficos, y en el botón Gráficos debe estar activa la opción Gráficos con pruebas de normalidad.

- c) Contrastar la hipótesis de igualdad de varianzas (homocedasticidad) entre los datos de los dos grupos.

i

Para realizar el contraste de igualdad de varianzas entre los dos grupos, vamos a utilizar el con traste de Levene. Para ello:

- 1) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar . . .
- 2) Seleccionar en Lista de dependientes la variable Grasa, y en Lista de factores la variable Granja. En Visualización debe estar activa la opción Gráficos, y pinchando en el botón Gráficos, en Dispersión por nivel con prueba de Levene se debe escoger la opción No transformados.

- d) Utilizando el contraste más adecuado, ¿se puede concluir que existen diferencias significativas en el contenido medio de grasa según la procedencia de la leche?

i

Aunque según el análisis anterior, no podemos rechazar la hipótesis de normalidad en los grupos que queremos comparar y tampoco la igualdad de varianzas en los datos de los dos grupos, pero como la muestra es muy pequeña e incluso uno de los grupos tiene menos de 10 observaciones, lo más correcto sería aplicar el contraste de la U de Mann Whitney. Para ello:

- 1) Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadro de diálogos antiguo ▶ 2 muestras independientes.
- 2) En el cuadro de dialogo que aparece seleccionar la variable Grasa al campo Lista Contrastar variables, seleccionar la variable Granja al campo Variable de agrupación y hacer click sobre el botón Definir grupos.
- 3) En el cuadro de diálogo que aparece introducir en el campo Grupo 1 el valor de la variable Granja correspondiente a la granja X y en el campo Grupo 2 el correspondiente a la granja Y, hacer click sobre el botón Continuar y hacer click en el botón Aceptar.

3. Para ver si una campaña de publicidad sobre un fármaco ha influido en sus ventas, se tomó una muestra de 8 farmacias y se midió el número de unidades de dicho fármaco vendidas durante un mes, antes y después de

la campaña, obteniéndose los siguientes resultados:

Antes	147	163	121	205	132	190	176	147
Después	150	171	132	208	141	184	182	149

- a) Crear las variables Antes y Despues e introducir los datos de la muestra.
- b) Comprobar la hipótesis de normalidad de la variable diferencia.

i

- 1) Seleccionar el menú Transformar ▶ Calcular.
- 2) En el cuadro de diálogo que aparece introducir el nombre de la nueva variable Diferencia en el campo Variable de destino, en el campo Expresión numérica introducir Antes-Despues y hacer click en el botón Aceptar.
- 3) Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar...
- 4) Seleccionar en Lista de dependientes la variable Diferencia. En Visualización debe estar activa la opción Gráficos, y en el botón Gráficos debe estar activa la opción Gráficos con pruebas de normalidad.

- c) Utilizando el contraste más adecuado, ¿se puede concluir que la campaña de publicidad ha aumentado las ventas?

i

- 1) Aunque, según el análisis anterior, no podemos rechazar la hipótesis de normalidad en la variable diferencia, como la muestra es muy pequeña, lo más correcto sería aplicar el contraste de Wilcoxon.
- 2) Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadro de diálogos antiguo ▶ 2 muestras relacionadas.
- 3) En el cuadro de diálogo que aparece seleccionar las variables Antes y Despues al campo Contrastar Pares y hacer click sobre el botón Aceptar.

4. Se han valorado los cambios en la presión arterial sistólica (antes del tratamiento con el fármaco menos después del tratamiento, en mm Hg) en quince pacientes, que se dividieron en tres grupos, aplicando a cada gru-

po un fármaco diferente. Los resultados obtenidos tras un año de tratamiento fueron:

Cambio de presión		
Fármaco 1	Fármaco 2	Fármaco 3
12	-3	1
15	5	5
16	-8	19
6	-2	45
8	4	3

- Crear las variables Farmaco y Cambio e introducir los datos de la muestra.
- Generar un diagrama de dispersión con los datos de los 15 pacientes, con el fármaco en el eje X y el cambio en la presión arterial sistólica en el eje Y. A la vista del diagrama, ¿crees que los datos presentan homogeneidad de varianzas? ¿crees que hay algún grupo con un cambio en la presión arterial diferente del resto?

i

- Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos...
- Escoger la opción Dispersión simple, y pinchar en botón Definir.
- Seleccionar como Eje X la variable Farmaco, y como Eje Y la variable Cambio.

- Analizar la normalidad de los datos de los 3 grupos y la homogeneidad de varianzas.

i

Para analizar la normalidad de los datos de los 3 grupos utilizamos el con traste de Kolmogorov-Smirnov-Lilliefors, y para la homogeneidad de varianzas el con traste de Levene. Para ello:

- Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar...
- Seleccionar en Lista de dependientes la variable Cambio, y en Lista de factores la variable Farmaco.

- 3) Para la prueba de normalidad, en Visualización debe estar activa la opción Gráficos, y en el botón Gráficos debe estar activa la opción Gráficos con pruebas de normalidad.
- 4) Para la prueba de homogeneidad de varianzas, en Visualización debe estar activa la opción Gráficos, y pinchando en el botón Gráficos, en Dispersión por nivel con prueba de Levene se debe escoger la opción No transformados.

d) Utilizando el contraste más adecuado, ¿se puede concluir que existen diferencias en los cambios de la presión sistólica en función del fármaco recibido?

i

En este caso no se cumple la homogeneidad de varianzas, por lo que no se puede aplicar una ANOVA y tendremos que recurrir al contraste de Kruskal-Wallis.

- 1) Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadro de diálogos antiguo ▶ k muestras independientes.
- 2) En el cuadro de dialogo que aparece seleccionar la variable Cambio al campo Lista Contrastar variables, seleccionar la variable Farmaco al campo Variable de agrupación y hacer click sobre el botón Definir rango.
- 3) En el cuadro de diálogo que aparece introducir en el campo Mínimo el valor 1, y en el campo Máximo el valor 3, hacer click sobre el botón Continuar y hacer click en el botón Aceptar.

e) ¿Entre qué grupos se dan las diferencias significativas?

i

Para analizar entre qué grupos se dan las diferencias significativas mediante un contraste no paramétrico, hay que utilizar la U de Mann Whitney, y ver si hay diferencias significativas entre todas las categorías de la variable Farmaco tomadas dos a dos; es decir, hay que ver si hay diferencias entre 1-2, 1-3 y 2-3. Posteriormente, el p valor obtenido en cada uno de los análisis lo multiplicamos por 3 que es el número de categorías de la variable Farmaco. Por ejemplo, para la comparación 1-3:

- 1) Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadro de diálogos antiguo ▶ 2 muestras independientes...
- 2) En el cuadro de dialogo que aparece seleccionar la variable Cambio al campo Lista Contrastar variables, seleccionar la variable Farmaco al campo Variable de agrupación y hacer click sobre el botón Definir grupos.
- 3) En el cuadro de diálogo que aparece introducir en el campo Grupo 1 el valor 1 (valor de la variable Farmaco correspondiente al primero de los grupos que consideramos), y en el campo Grupo 2 el valor 3 (valor de la variable Farmaco correspondiente al segundo de los grupos que consideramos). Posteriormente, hacer click sobre el botón Continuar y hacer click en el botón Aceptar.
- 4) Una vez obtenido el p -valor, no olvidar multiplicar por 3.

5. Se quiere contrastar la dificultad de cuatro modelos de examen que se van a poner en la convocatoria ordinaria de la asignatura de bioestadística. Para ello se pide a cinco profesores diferentes que valoren cada uno de los modelos de 0 a 10, y los resultados fueron:

	Modelo1	Modelo2	Modelo3	Modelo4
Profesor1	6	8	5	8
Profesor2	5	4	7	9
Profesor3	5	4	5	6
Profesor4	7	4	6	7
Profesor5	6	3	7	8

- a) Crear las variables Modelo1, Modelo2, Modelo3 y Modelo4 e introducir los datos de la muestra.
- b) ¿Podemos afirmar que el grado de dificultad de los modelos es diferente? Utilizar un contraste de Friedman para dar respuesta a la pregunta.

i

- 1) Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadro de diálogos antiguo ▶ K muestras relacionadas.

2) En el cuadro de diálogo que aparece seleccionar las variables modelo1, modelo2, modelo3 y modelo4 al campo Variables de contraste y hacer click sobre el botón Aceptar.

6. El test de Apgar es un examen clínico de neonatología en donde el médico realiza una prueba medida en 3 estándares sobre el recién nacido para obtener una primera valoración simple (macroscópica) y clínica sobre el estado general del neonato después del parto. El recién nacido es evaluado de acuerdo a cinco parámetros fisioanatómicos simples, que son: color de la piel, frecuencia cardíaca, reflejos, tono muscular y respiración. A cada parámetro se le asigna una puntuación entre 0 y 2, y sumando las cinco puntuaciones se obtiene el resultado del test. El test se realiza al minuto, a los cinco minutos y, en ocasiones, a los diez minutos de nacer. La puntuación al primer minuto evalúa el nivel de tolerancia del recién nacido al proceso del nacimiento y su posible sufrimiento, mientras que la puntuación obtenida a los 5 minutos evalúa el nivel de adaptabilidad del recién nacido al medio ambiente y su capacidad de recuperación.

En la siguiente tabla se refleja la puntuación obtenida por 22 recién nacidos en el test de Apgar al minuto y a los cinco minutos de haber nacido:

Apgar1	10	3	8	9	8	9	8	8	8	8	7
Apgar5	10	6	9	10	9	10	9	9	9	9	9
Apgar1	8	6	8	9	9	9	9	8	9	3	9
Apgar5	9	6	8	9	9	9	9	8	9	3	9

Con los datos anteriores se pretende realizar un contraste de hipótesis para analizar si existe o no relación entre las dos puntuaciones. Para ello:

- Crear las variables Apgar1 y Apgar5 e introducir los datos de la muestra.
- Comprobar si las variables siguen distribuciones normales.

i

- Seleccionar el menú Analizar ▶ Estadísticos descriptivos ▶ Explorar...
- Seleccionar en Lista de dependientes la variables Apgar1 y Apgar2, y dejar vacía la Lista de factores. En Visualización debe es-

tar activa la opción Gráficos, y en el botón Gráficos debe estar activa la opción Gráficos con pruebas de normalidad.

- c) Realizar el contraste de hipótesis bilateral: H_0 : No hay relación entre las variables, H_1 : Sí que hay relación.

i

Como las variables analizadas no siguen distribuciones normales, para realizar un contraste de relación entre ambas hay que obtener el coeficiente de correlación de Spearman y ver si es o no significativamente diferente de 0. Para ello:

- 1) Seleccionar el menú Analizar ▶ Correlaciones ▶ Bivariadas.
- 2) En el cuadro de diálogo que aparece, pasar las variable Apgar1 y Apgar2 al campo Variables. Seleccionar en la lista de Coeficientes de correlación la opción Spearman y hacer click sobre el botón Aceptar.

13.2. Ejercicios propuestos

1. El departamento de calidad de un laboratorio farmacéutico, recoge una muestra de una de las maquinas que producen ampollas de suero fisiológico y comprueba los siguientes resultados, C (correcto) y D (Defectuoso):

C-C-C-C-D-C-C-C-C-D-D-D-C-C-C-C-C-C-C-C-C-C-D-D-D-C-C-C-C-D-D-D-C-C-C-C-C-C-C

¿Se puede considerar que el resultado obtenido sigue una secuencia aleatoria?

2. Se ha realizado un estudio para investigar el efecto del ejercicio físico en el nivel de colesterol en la sangre. En el estudio participaron once personas, a las que se les midió el nivel de colesterol antes y después de desarrollar un programa de ejercicios. Los resultados obtenidos fueron los siguientes:

Antes	223	212	221	210	225	202	198	200	185	220
Después	226	211	222	212	225	201	196	217	130	220

Utilizando el contraste más adecuado, ¿se puede concluir que el ejercicio físico disminuye el nivel de colesterol?

3. Dos químicos *A* y *B* realizan respectivamente 14 y 16 determinaciones de la actividad radiactiva de una muestra de material. Sus resultados en Curios:

<i>A</i>		<i>B</i>	
263,10	262,60	286,53	254,54
262,10	259,60	284,55	286,30
257,60	262,20	272,52	282,90
261,70	261,20	283,85	253,75
260,70	259,20	252,01	245,26
269,13	268,63	275,08	266,08
268,13	217,00	267,53	252,05
		253,82	269,81

Utilizando el contraste más adecuado, ¿se puede concluir que existen diferencias significativas en la actividad detectada por cada químico?

4. En un hospital se están evaluando dos tratamientos diferentes para ver si existen diferencias entre ellos, para lo cual se seleccionaron dos grupos de 32 pacientes cada uno y se aplicó un tratamiento a cada grupo. Los resultados fueron:

	Empeoraron	Igual	Mejoraron	Curaron
A	9	12	5	6
B	5	6	11	10

Utilizando el contraste más adecuado, ¿se puede concluir que existen diferencias significativas entre ambos tratamientos?

5. Queremos comparar las calificaciones iniciales de un grupo de 20 alumnos, con las obtenidas al final del curso, para ver si existen diferencias, las calificaciones fueron (SS suspenso, A aprobado, N notable y SB sobresaliente):

Contrastes de Hipótesis No Paramétricos

Alumno	1	2	3	4	5	6	7	8	9	10
Calificación inicial	SS	A	A	SS	N	SS	SS	SB	A	A
Calificación final	A	A	SS	A	SB	N	A	N	SS	A

Alumno	11	12	13	14	15	16	17	18	19	20
Calificación inicial	SS	N	A	SS	SB	A	N	SS	A	SB
Calificación final	SB	A	N	A	SB	A	N	A	N	SB

Utilizando el contraste más adecuado, ¿se puede concluir que existen diferencias significativas entre las calificaciones al comienzo y al final del curso?

6. Disponemos de la evaluación que han obtenido tres grupos de prácticas de la asignatura de bioestadística (MM muy mal, M mal, R regular, B bien y MB muy bien):

Grupo 01	R	B	R	M	MM	B	MB	R
Grupo 02	B	R	M	B	R	M	B	MB
Grupo 03	MB	B	M	R	B	MB	B	R

Grupo 01	M	B	M	R	R	MM	M
Grupo 02	M	R	M	R			
Grupo 03	B	MB	B	R	MB		

Utilizando el contraste más adecuado, ¿se puede concluir que existen diferencias significativas en la evaluación de los diferentes grupos?

7. Para comparar las dificultades presentados por un grupo de problemas de lógica, se han seleccionado aleatoriamente a ocho individuos a los que se les ha planteado tres pruebas iguales, a cada uno y se han anotado los tiempos, en minutos, que han tardado en resolverlos. Los resultados obtenidos son:

Prueba 1	Prueba 2	Prueba 3
38	6	35
22	4	9
14	8	8
8	2	4
6	4	8
10	14	10
14	2	5
8	6	3

Utilizando el contraste no paramétrico más adecuado, ¿se puede concluir que existen diferencias significativas en los tiempos de resolución de las tres pruebas?

8. La siguiente tabla muestra los datos de 9 pacientes con anemia aplásica:

Reticulocitos (%)	3,6	2,0	0,3	0,3	0,2
Linfocitos (por mm ²)	1700	3078	1820	2706	2086
Reticulocitos (%)	3,0	0,0	1,0	2,2	
Linfocitos (por mm ²)	2299	676	2088	2013	

Mediante el adecuado contraste de hipótesis basado en el coeficiente de correlación de Spearman, ¿existe relación entre ambas variables?

Contrastes Basados en el Estadístico Chi-cuadrado

14.1. Ejercicios Resueltos

1. Dadas dos parejas de genes Aa y Bb, la descendencia del cruce efectuado según las leyes de Mendel, debe estar compuesto del siguiente modo:

Fenotipo	Frecuencias Relativas
AB	$9/16 = 0,5625$
Ab	$3/16 = 0,1875$
aB	$3/16 = 0,1875$
ab	$1/16 = 0,0625$

Elegidos 300 individuos al azar de cierta población, se observa la siguiente distribución de frecuencias:

Fenotipo	Frecuencias Observadas
AB	165
Ab	47
aB	67
ab	21

Se pide

- a) Crear las variables fenotipo y frecuencia e introducir los datos de la muestra.
- b) Ponderar los datos mediante la variable frecuencia.

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.
- 2) En el cuadro de diálogo resultante activar la opción Ponderar casos mediante, seleccionar la variable frecuencia en el campo Variable frecuencia y hacer click en el botón Aceptar.

c) Comprobar si esta muestra cumple las leyes de Mendel.

i

- 1) Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadro de diálogo antiguos ▶ Chi-cuadrado.
- 2) En el cuadro de diálogo que aparece seleccionar la variable fenotipo al campo Lista Contrastar Variables, y en Valores esperados marcar la opción valores e introducir las proporciones según las leyes de Mendel y siguiendo el orden en el que aparecen los fenotipos, y hacer click sobre el botón Aceptar.

d) A la vista de los resultados del contraste, ¿se puede aceptar que se cumplen las leyes de Mendel en los individuos de dicha población?

2. En un estudio sobre úlceras pépticas se determinó el grupo sanguíneo de 1655 pacientes ulcerosos y 10000 controles, los datos fueron:

	O	A	B	AB
Paciente	911	579	124	41
Controles	4578	4219	890	313

a) Crear las variables participantes, grupo_sanguíneo y frecuencia e introducir los datos.

b) Ponderar los datos mediante la variable frecuencia.

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.
- 2) En el cuadro de diálogo resultante activar la opción Ponderar casos mediante, seleccionar la variable frecuencia en el campo Variable frecuencia y hacer click en el botón Aceptar.

c) Construir la tabla de contingencia y realizar el contraste Chi-cuadrado.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos Descriptivos ▶ Tablas de contingencia.
- 2) En el cuadro de diálogo que aparece, seleccionar la variable participantes al campo Filas y la variable grupo_sanguíneo al campo Columnas, y hacer click sobre el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece, marcar la casilla Chi-cuadrado y hacer click en el botón Continuar.
- 4) En el cuadro de diálogo inicial, hacer click sobre el botón Casillas.
- 5) En el cuadro de diálogo que aparece, marcar las casillas Frecuencias observadas y Esperadas, y hacer click sobre el botón Continuar y Aceptar.

d) A la vista de los resultados del contraste, ¿existe alguna relación entre el grupo sanguíneo y la úlcera péptica?, es decir, ¿se puede concluir que la proporción de pacientes y de controles es diferente dependiendo del grupo sanguíneo?

3. Mitchell et al. (1976, Annals of Human Biology), partiendo de una muestra de 478 individuos, estudiaron la distribución de los grupos sanguíneos en varias regiones del sur-oeste de Escocia, obteniendo los resultados que se muestran:

	Eskdale	Annandale	Nithsdale	
A	33	54	98	185
B	6	14	35	55
O	56	52	115	223
AB	5	5	5	15
	100	125	253	478

- a) Crear las variables grupo_sanguíneo, región y frecuencia e introducir los datos.
- b) Ponderar el estudio, por la variable frecuencia

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.

- 2) En el cuadro de diálogo resultante activar la opción Ponderar casos mediante, seleccionar la variable frecuencia en el campo Variable frecuencia y hacer click en el botón Aceptar.

- c) Construir la tabla de contingencia y realizar el contraste Chi-cuadrado.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos Descriptivos ▶ Tablas de contingencia.
- 2) En el cuadro de diálogo que aparece, seleccionar la variable grupo_sanguíneo al campo Filas y la variable región al campo Columnas, y hacer click sobre el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece, marcar la casilla Chi-cuadrado y hacer click en el botón Continuar.
- 4) En el cuadro de diálogo inicial, hacer click sobre el botón Casillas.
- 5) En el cuadro de diálogo que aparece, marcar las casillas Frecuencias observadas y Esperadas, y hacer click sobre el botón Continuar y Aceptar.

- d) En vista de los resultados del contraste, ¿se distribuyen los grupos sanguíneos de igual manera en las diferentes regiones?

4. En un estudio para saber si el habito de fumar está relacionado con el sexo, se ha preguntado a 26 personas. De los 9 hombres consultados 2 respondieron que fumaban, mientras que de las 17 mujeres consultadas, 6 fumaban. ¿Podemos afirmar que existe relación entre ambas variables?

- a) Crear las variables sexo, fuma y frecuencia e introducir los datos.
- b) Ponderar el estudio, por la variable frecuencia

i

- 1) Seleccionar el menú Datos ▶ Ponderar casos.
- 2) En el cuadro de diálogo resultante activar la opción Ponderar casos mediante, seleccionar la variable frecuencia en el campo Variable frecuencia y hacer click en el botón Aceptar.

- c) Construir la tabla de contingencia y realizar el contraste Chi-cuadrado.

- i*
- 1) Seleccionar el menú Analizar ▶ Estadísticos Descriptivos ▶ Tablas de contingencia.
 - 2) En el cuadro de diálogo que aparece, seleccionar la variable sexo al campo Filas y la variable fuma al campo Columnas, y hacer click sobre el botón Estadísticos.
 - 3) En el cuadro de diálogo que aparece, marcar la casilla Chi-cuadrado y hacer click en el botón Continuar.
 - 4) En el cuadro de diálogo inicial, hacer click sobre el botón Casillas.
 - 5) En el cuadro de diálogo que aparece, marcar las casillas Frecuencias observadas y Esperadas, y hacer click sobre el botón Continuar y Aceptar.

d) En vista de los resultados del contraste, ¿se distribuyen los fumadores de igual manera en ambos sexos?

i En este caso el procedimiento a seguir es igual que para la Chi cuadrado, pero vemos que ahora no se cumplen las condiciones para poder aplicar esta prueba, por eso nos tendremos que fijar en el Estadístico exacto de Fisher, que si podemos aplicar, teniendo en cuenta si estamos realizando un contraste bilateral o unilateral.

5. Para probar la eficacia de dos fármacos diferentes contra las migrañas, se seleccionaron a 20 personas que padecían migrañas habitualmente, y se les dió a tomar a cada uno los fármacos en momentos diferentes. Luego se les preguntó si habían obtenido mejoría o no con el fármaco tomado. Los resultados fueron los siguientes:

Fármaco 1	Sí	Sí	Sí	Sí	Sí	No	Sí	No	Sí	Sí
Fármaco 2	No	No	Sí	No	Sí	Sí	No	No	No	No
Fármaco 1	Sí	No	Sí	No	Sí	Sí	Sí	No	Sí	Sí
Fármaco 2	Sí	No	Sí	No	No	Sí	No	Sí	No	No

a) Crear las variables Mejora_Farmaco1, y Mejora_Farmaco2 e introducir los datos.

b) Construir la tabla de contingencia y realizar el contraste de McNemar.

i

- 1) Seleccionar el menú Analizar ▶ Estadísticos Descriptivos ▶ Tablas de contingencia.
- 2) En el cuadro de diálogo que aparece, seleccionar la variable Mejora_Farmaco1 al campo Filas y la variable Mejora_Farmaco2 al campo Columnas, y hacer click sobre el botón Estadísticos.
- 3) En el cuadro de diálogo que aparece, marcar la casilla McNemar y hacer click en el botón Continuar.
- 4) En el cuadro de diálogo inicial, hacer click sobre el botón Casillas.
- 5) En el cuadro de diálogo que aparece, marcar las casillas Frecuencias observadas y Esperadas, y hacer click sobre el botón Continuar y Aceptar.

c) En vista de los resultados del contraste, ¿podemos afirmar que existen diferencias significativas entre los dos fármacos?

i

Otra forma de realizar este mismo contraste, sería seleccionado el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadros de diálogo antiguos ▶ 2 muestras relacionadas. Luego pasar las dos variables a contrastar al cuadro Contrastar pares, marcar la casilla McNemary hacer click sobre el botón Aceptar.

14.2. Ejercicios propuestos

1. Supongamos que queremos comprobar si un dado está bien equilibrado o no. Lo lanzamos 1200 veces, y obtenemos las frecuencias de la

siguiente tabla:

Número	Frecuencias de aparición
1	120
2	275
3	95
4	310
5	85
6	315

- a) A la vista de los resultados, ¿se puede aceptar que el dado está bien equilibrado?
 - b) Nos dicen que, en este dado, los números pares aparecen con una frecuencia 3 veces superior a la de los impares. Contrastar dicha hipótesis.
2. Se realiza un estudio en una población de pacientes críticos hipotéticos y se observan, entre otras, dos variables, la evolución (si sobreviven SV o no NV) y la presencia o ausencia de coma, al ingreso. Se obtienen los siguientes resultados:

	No coma	Coma	
SV	484	37	521
NV	118	89	207
	602	126	728

Nos preguntamos: ¿es el coma al ingreso un factor de riesgo para la mortalidad?

3. La recuperación producida por dos tratamientos distintos A y B, se clasifican en tres categorías: muy buena, buena y mala. Se administra el tratamiento A a 32 pacientes y el B a otros 28. De las 22 recuperaciones muy buenas, 10 corresponden al tratamiento A; de las 24 recuperaciones buenas, 14 corresponden al tratamiento A y de las 14 que tienen una mala recuperación, 8 corresponden al tratamiento A. ¿Son igualmente efectivos ambos tratamientos para la recuperación de los pacientes?
4. Para contrastar la hipótesis de que las mujeres tienen más éxito en sus estudios que los hombres, se ha tomado una muestra de 10 chicos y

otra de 10 chicas que han sido examinados por un profesor que aprueba siempre al 40 % de los alumnos presentados a examen. Teniendo en cuenta que sólo aprobaron 2 chicos, utiliza el contraste de hipótesis más adecuado para decidir si la citada hipótesis es cierta.

5. Se ha preguntado a los 150 alumnos de un curso, si estaban de acuerdo o no, con la metodología de enseñanza de dos profesores, distintos que les han dado clase en la asignatura de bioestadística. Los resultados se recogen en la siguiente tabla:

Profesor 1 \ Profesor 2	Opinión favorable	Opinión desfavorable
Opinión favorable	37	48
Opinión desfavorable	44	21

¿Podemos afirmar que existe diferente opinión por parte de los alumnos, sobre los dos profesores?

Análisis de Concordancia

15.1. Ejercicios resueltos

1. Se ha medido la presión arterial sistólica en un grupo de 20 pacientes con el tensiómetro habitual y con un tensiómetro de muñeca, obteniéndose los siguientes resultados:

Tensiómetro habitual	Tensiómetro de muñeca	Tensiómetro habitual	Tensiómetro de muñeca
112	124	133	126
124	116	115	121
96	88	104	98
106	110	86	94
138	144	93	102
155	150	144	132
86	82	125	120
126	118	112	104
114	120	108	108
92	97	98	98

Se pide

- a) Crear las variables `tens_habitual` y `tens_muñeca` e introducir los datos de la muestra.
- b) Calcular el coeficiente de correlación intraclase e interpretar el resultado.

i

- 1) Seleccionar el menú **Analizar** ▶ **Escala** ▶ **Análisis de fiabilidad**.
- 2) Seleccionar las variables **tens_habitual** y **tens_muñeca** al campo **Elementos**, y hacer click en el botón **Estadísticos**.
- 3) En el cuadro de diálogo resultante marcar **Coefficiente de correlación intraclase**.
- 4) En el campo **Modelo**: seleccionar la opción **Dos factores, efectos mixtos** y en el campo **Tipo**: seleccionar la opción **Acuerdo absoluto**. Se ha escogido el modelo indicado pues el primer factor de variabilidad serían los pacientes, que entendemos que se han escogido de forma aleatoria y sus efectos serían aleatorios, y el segundo factor es el aparato de medida que en nuestro caso es de efectos fijos, pues el tensiómetro habitual y el de muñeca no se han sacado de una muestra aleatoria de tensiómetros.
- 5) Hacer click sobre el botón **Continuar y Aceptar**. arterial obtenidas con el tensiómetro habitual y con el de muñeca es muy buena.

- c) Dibujar un gráfico de dispersión en el que aparezca la recta de regresión de la presión arterial medida con el tensiómetro de muñeca en función de la medida con el tensiómetro habitual.

i

- 1) Seleccionar el menú **Gráficos** ▶ **Cuadros de diálogo antiguos** ▶ **Dispersión/Puntos**, elegir la opción **Dispersión simple** y hacer click sobre el botón **Definir**.
- 2) Seleccionar la variable **tens_muñeca** en el campo **Eje Y** del cuadro de diálogo.
- 3) Seleccionar la variable **tens_habitual** en el campo **Eje X** del cuadro de diálogo y hacer click sobre el botón **Aceptar**.
- 4) Editar el gráfico realizado haciendo un doble click sobre él.
- 5) Seleccionar los puntos haciendo click sobre alguno de ellos.
- 6) Hacer click con el botón derecho del ratón y seleccionar **Añadir línea de ajuste total**.

- 7) Cerrar la ventana Propiedades y la ventana del editor de gráficos.

d) Añadir al gráfico obtenido en el apartado anterior, la recta de regresión de la presión arterial medida con el tensiómetro de muñeca en función de la medida con el tensiómetro habitual, que se obtendría si ambos procesos de medida dieran exactamente el mismo resultado.

- i*
- 1) Editar el gráfico realizado en el apartado anterior haciendo un doble click sobre él.
 - 2) Hacer click con el botón derecho del ratón y seleccionar Añadir línea de referencia desde la ecuación y en el campo Ecuación personalizada que aparece en Línea de referencia de la ventana Propiedades, escribir x para representar la recta $y = x$.
 - 3) Cerrar la ventana Propiedades y la ventana del editor de gráficos.

e) Crear la variable diferencia obtenida restando `tens_habitual` menos `tens_muñeca`

- i*
- 1) Seleccionar el menú Transformar ▶ Calcular...
 - 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso `diferencia`, dentro del campo Variable de destino:.
 - 3) En el cuadro de diálogo Expresión numérica: escribir `tens_habitual-tens_muñeca` y hacer click sobre el botón Aceptar.

f) Crear la variable `tens_media` cuyos valores serán las medias de `tens_habitual` y `tens_muñeca`

- i*
- 1) Seleccionar el menú Transformar ▶ Calcular...
 - 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso `tens_media`, dentro del campo Variable de destino:.
 - 3) En el cuadro de diálogo Expresión numérica: escribir $(\text{tens_habitual} + \text{tens_muñeca}) / 2$ y hacer click sobre el botón Aceptar.

- g) Representar gráficamente los valores de la diferencia de tensiones en función de su media.

i

- 1) Seleccionar el menú Gráficos ▶ Cuadros de diálogo antiguos ▶ Dispersión/Puntos, elegir la opción Dispersión simple y hacer click sobre el botón Definir.
- 2) Seleccionar la variable diferencia en el campo Eje Y del cuadro de diálogo.
- 3) Seleccionar la variable tens_media en el campo Eje X del cuadro de diálogo y hacer click sobre el botón Aceptar.

- h) Crear la variable dif_abs, cuyos valores serán los valores absolutos de la variable diferencia, y la variable desacuerdo, que tomará el valor 0 si las medidas con ambos métodos son iguales y el valor 1 si son distintas.

i

- 1) Seleccionar el menú Transformar ▶ Calcular...
- 2) Introducir el nombre de la nueva variable que vamos a crear, en este caso dif_abs, dentro del campo Variable de destino:.
- 3) En el cuadro de diálogo Expresión numérica: escribir ABS(diferencia) y hacer click sobre el botón Aceptar.
- 4) Crear una variable denominada desacuerdo e introducir un 0 cuando dif_abs sea 0 y un 1 en caso contrario.

- i) Representar gráficamente el porcentaje de discordancia entre ambos métodos de medición de la tensión arterial.

i

- 1) Seleccionar el menú Analizar ▶ Superviv. ▶ Kaplan-Meier.
- 2) Seleccionar la variable dif_abs al campo Hora: y la variable desacuerdo al campo Estado:.
- 3) Hacer click en el botón Definir evento....
- 4) En el cuadro de diálogo resultante seleccionar la opción Valor único:, poner un 1 en dicho campo y hacer click sobre el botón Continuar.
- 5) Hacer click sobre el botón Opciones y marcar Suprviv. en Gráficos.

6) Hacer click sobre el botón Continuar y Aceptar.

2. Se entregaron a dos radiólogos A y B un conjunto de radiografías de tórax de pacientes oncológicos, para que informaran si presentaban metástasis en los pulmones o no. Ambos radiólogos analizaron todas las radiografías y cada uno de ellos emitió su informe, indicando en cuáles de ellas se apreciaban metástasis y en cuáles no. Como resultado de dichos informes hubo 32 radiografías en que ambos radiólogos apreciaron metástasis, 68 radiografías en que ninguno las apreció, 6 radiografías en que el radiólogo A las apreció y el B no, y 10 en que el B las apreció y el A no.

a) Crear las variables opiniónA, opiniónB y frecuencia e introducir los datos de la muestra.

b) Ponderar los datos mediante la variable frecuencia.

i

1) Seleccionar el menú Datos ▶ Ponderar casos.

2) En el cuadro de diálogo resultante activar la opción Ponderar casos mediante, seleccionar la variable frecuencia en el campo Variable de frecuencia y hacer click en el botón Aceptar.

c) Construir la tabla de contingencia, calcular el índice Kappa de Cohen e interpretar el resultado.

i

1) Seleccionar el menú Analizar ▶ Estadísticos Descriptivos ▶ Tablas de contingencia.

2) En el cuadro de diálogo que aparece, seleccionar la variable opiniónA al campo Filas y la variable opiniónB al campo Columnas, y hacer click sobre el botón Estadísticos.

3) En el cuadro de diálogo que aparece, marcar la casilla Kappa, y hacer click en el botón Continuar y Aceptar.

3. Un tribunal médico tiene que evaluar las aptitudes psicofísicas de un grupo de aspirantes a ingreso en un cuerpo de la Administración. El tribunal está formado por tres médicos, cada uno de los cuales tiene que analizar los datos de todos los aspirantes, mantener una entrevista personal con cada uno de ellos y emitir un informe que se concreta en: Apto, Dudoso

o No Apto. Los resultados de dicho informe fueron los siguientes:

Médico 1	Médico 2	Médico 3
Apto	Apto	Apto
Apto	Dudoso	Apto
Dudoso	Apto	Apto
No Apto	No Apto	Dudoso
No Apto	No Apto	No Apto
Apto	Apto	Apto
Apto	Apto	Apto
Dudoso	Apto	Dudoso
No Apto	Dudoso	No Apto
Apto	Apto	Apto

- Crear las variables médico1, médico2 y médico3 e introducir los datos, asignando el valor 0 a No Apto, el 1 a Dudoso y el 2 a Apto.
- Calcular el coeficiente de concordancia W de Kendall e interpretar el resultado..

i

- Seleccionar el menú Analizar ▶ Pruebas no paramétricas ▶ Cuadros de diálogo antiguos ▶ K muestras relacionadas.
- En el cuadro de diálogo resultante seleccionar las variables médico1, médico2 y médico3 en el campo Variables de contraste, marcar en Tipo de prueba la opción W de Kendall y hacer click en el botón Aceptar.
- El coeficiente de correlación W de Kendall es 0,233, lo que indica que hay poca concordancia.

15.2. Ejercicios propuestos

- Se ha medido la concentración de ácido úrico en sangre en un grupo de diez pacientes con el equipo tradicional y con un equipo nuevo, obteniéndose los siguientes resultados, expresados en mg./dl.:

Tradicional	5,4	6,2	3,7	7,6	4,5	3,8	5,2	4,7	4,9	5,5
Nuevo	5,8	6,9	3,4	6,4	4,5	4,4	5,8	5,6	4,2	6,8

Se desea:

Análisis de Concordancia

- a) Calcular el coeficiente de correlación intraclase e interpretarlo en términos de concordancia de las medidas de la concentración de ácido úrico en sangre obtenidas con ambos equipos.
 - b) Dibujar el diagrama de dispersión, en el que aparezca la recta de regresión de las concentraciones de ácido úrico obtenidas con el equipo nuevo sobre las obtenidas con el equipo tradicional, y la recta de regresión que se obtendría si el equipo nuevo siempre diera un resultado 0,8 mg/dl superior al que proporciona el equipo tradicional.
2. Se plantean a un grupo de personas dos tests A y B para determinar si su régimen alimenticio es adecuado o no. Como resultado de los test hubo 72 personas cuyo régimen alimenticio evaluado con ambos tests resultó adecuado, 34 personas en que con ninguno de los tests resultó adecuado, 12 personas cuyo régimen era adecuado según el test A pero no según el test B y 10 personas cuyo régimen era adecuado según el test B pero no según el test A. Se pide:
 - a) Obtener la tabla de contingencia correspondiente a los resultados obtenidos con ambos tests.
 - b) ¿Hay mucha coincidencia entre los resultados obtenidos con ambos tests? Calcular el índice de Kappa y contestar a partir del resultado obtenido.

