

# Bioestadística Aplicada con R y rk.Teaching

Santiago Angulo Díaz-Parreño ([sangulo@ceu.es](mailto:sangulo@ceu.es))

José Enrique Blanco Durango ([jose.blancodurango@ceu.es](mailto:jose.blancodurango@ceu.es))

Juan Carlos Garro Garro ([garro.eps@ceu.es](mailto:garro.eps@ceu.es))

Eduardo López Ramírez ([elopez@ceu.es](mailto:elopez@ceu.es))

Rubén Oncala Mesa ([ruben.oncalamesa@ceu.es](mailto:ruben.oncalamesa@ceu.es))

Alfredo Sánchez Alberca ([asalber@ceu.es](mailto:asalber@ceu.es))

Alberto Zaragoza de Lorite ([alberto.zaragozalorite@ceu.es](mailto:alberto.zaragozalorite@ceu.es))

Departamento de Matemáticas y Ciencia de Datos  
CEU San Pablo

Septiembre 2024



CEU

*Universidad  
San Pablo*

## Términos de la licencia

Esta obra está bajo una licencia Atribución–No comercial–Compartir igual 4.0 Internacional de Creative Commons. Para ver una copia de esta licencia, visite <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>.

Con esta licencia eres libre de:

- Compartir – Copiar y redistribuir el material en cualquier medio o formato.
- Adaptar – Remezclar, transformar y construir a partir del material.

Bajo los siguientes términos:



**Atribución.** Usted debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



**No comercial.** Usted no puede hacer uso del material con propósitos comerciales.



**Compartir igual.** Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original.

No hay restricciones adicionales — No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

# Índice general

<b>1</b>	<b>Introducción a R y RKWard</b>	<b>1</b>
1.1	Introducción	1
1.2	Instalación	2
1.2.1	Windows	2
1.2.2	MacOs	3
1.2.3	Linux	3
1.3	Arranque	3
1.4	Tipos de datos y operadores aritméticos y lógicos	4
1.5	Introducción y manipulación de datos	5
1.5.1	Introducción de datos en línea de comandos	5
1.5.2	Introducción de datos en RKWard	6
1.5.3	Ponderación de datos	7
1.5.4	Guardar datos	9
1.5.5	Abrir datos	9
1.5.6	Eliminación de datos	9
1.6	Transformación de datos	10
1.6.1	Filtrado de datos	10
1.6.2	Cálculo de variables	10
1.6.3	Recodificación de variables	11
1.7	Manipulación de ficheros de resultados	11
1.7.1	Guardar los resultados	11
1.7.2	Limpiar la ventana de resultados	12
1.8	Manipulación de guiones de comandos	12
1.8.1	Creación de un guión de comandos	12
1.8.2	Guardar un guión de comandos	13
1.8.3	Abrir un guión de comandos	13
1.9	Ayuda	13
1.10	Ejercicios resueltos	15
1.11	Ejercicios propuestos	17
<b>2</b>	<b>Distribuciones de Frecuencias y Representaciones Gráficas</b>	<b>19</b>
2.1	Ejercicios resueltos	19
2.2	Ejercicios propuestos	21
<b>3</b>	<b>Estadísticos Muestrales</b>	<b>23</b>
3.1	Ejercicios resueltos	23
3.2	Ejercicios propuestos	25
<b>4</b>	<b>Regresión Lineal Simple y Correlación</b>	<b>27</b>
4.1	Ejercicios resueltos	27
4.2	Ejercicios propuestos	31

<b>5</b>	<b>Regresión no lineal</b>	<b>33</b>
5.1	Ejercicios resueltos	33
5.2	Ejercicios propuestos	36
<b>6</b>	<b>Probabilidad</b>	<b>39</b>
6.1	Ejercicios resueltos	39
6.2	Ejercicios propuestos	43
<b>7</b>	<b>Variables Aleatorias Discretas</b>	<b>45</b>
7.1	Ejercicios resueltos	45
7.2	Ejercicios propuestos	47
<b>8</b>	<b>Variables Aleatorias Continuas</b>	<b>49</b>
8.1	Ejercicios resueltos	49
8.2	Ejercicios propuestos	54
<b>9</b>	<b>Intervalos de Confianza para Medias y Proporciones</b>	<b>55</b>
9.1	Ejercicios resueltos	55
9.2	Ejercicios propuestos	57
<b>10</b>	<b>Intervalos de Confianza para la Comparación de 2 Poblaciones</b>	<b>59</b>
10.1	Ejercicios resueltos	59
10.2	Ejercicios propuestos	61
<b>11</b>	<b>Contraste de Hipótesis</b>	<b>63</b>
11.1	Ejercicios resueltos	63
11.2	Ejercicios propuestos	67
<b>12</b>	<b>Análisis de la Varianza de 1 Factor</b>	<b>69</b>
12.1	Ejercicios resueltos	69
12.2	Ejercicios propuestos	71

# Introducción a R y RKWard

## 1 Introducción

La gran potencia de cálculo alcanzada por los ordenadores ha convertido a los mismos en poderosas herramientas al servicio de todas aquellas disciplinas que, como la estadística, requieren manejar un gran volumen de datos. Actualmente, prácticamente nadie se plantea hacer un estudio estadístico serio sin la ayuda de un buen programa de análisis estadístico.

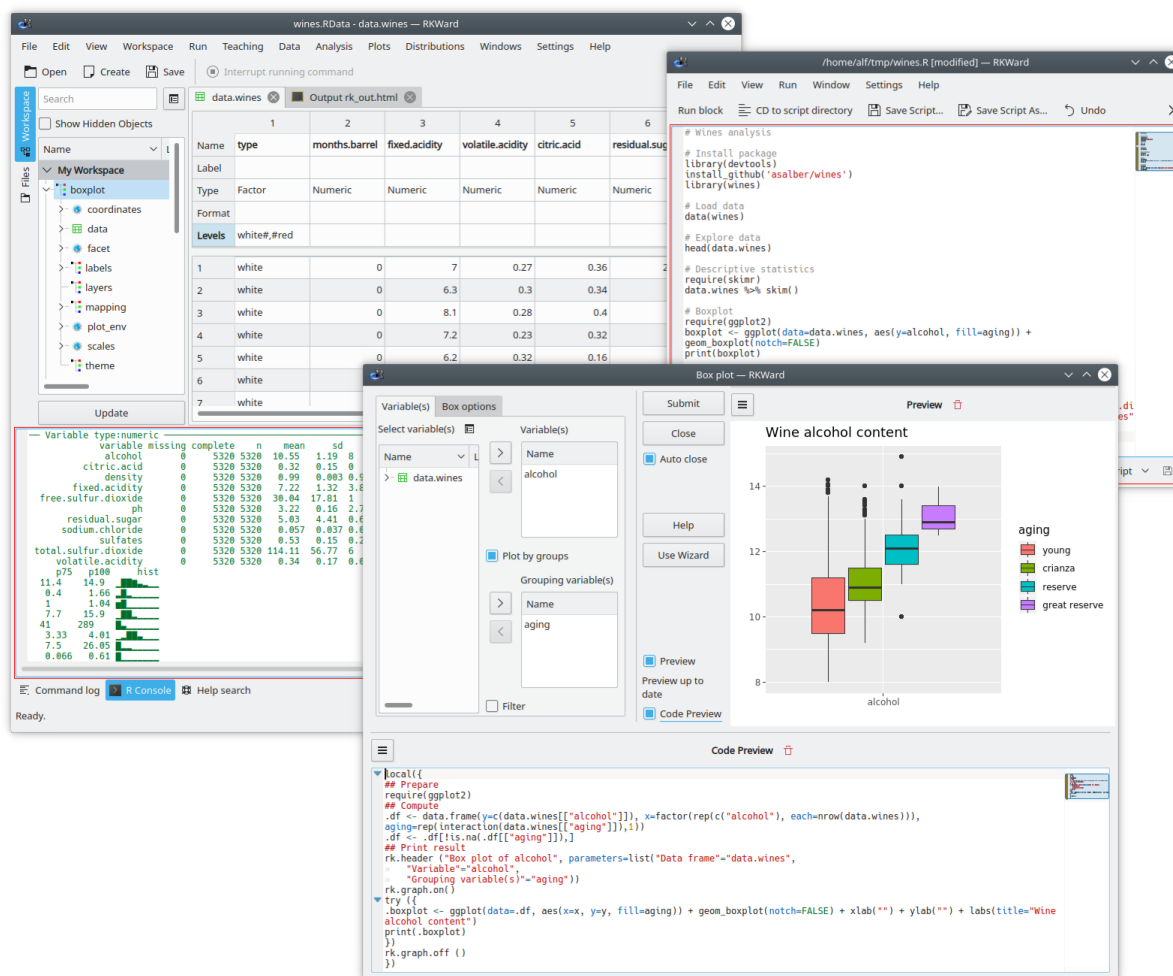
R es un potente lenguaje de programación que incluye multitud de funciones para la representación el análisis de datos. Fue desarrollado por Robert Gentleman y Ross Ihaka en la Universidad de Auckland en Nueva Zelanda, aunque actualmente es mantenido por una enorme comunidad científica en todo el mundo.



Las ventajas de R frente a otros programas habituales de análisis de datos, como pueden ser SPSS, SAS, SPlus, Matlab o Minitab, son múltiples:

- Es software libre y por tanto gratuito. Puede descargarse desde la web <http://www.r-project.org/>.
- Es multiplataforma. Existen versiones para Windows, Macintosh, Linux y otras plataformas.
- Está avalado y en constante desarrollo por una amplia comunidad científica que lo utiliza como estándar para el análisis de datos.
- Cuenta con multitud de paquetes para todo tipo de análisis estadísticos y representaciones gráficas, desde los más habituales, hasta los más novedosos y sofisticados que no incluyen otros programas. Los paquetes están organizados y documentados en un repositorio CRAN (Comprehensive R Archive Network) desde donde pueden descargarse libremente. En España hay una copia de este repositorio en la web <http://cran.es.r-project.org/>.
- Es programable, lo que permite que el usuario pueda crear fácilmente sus propias funciones o paquetes para análisis de datos específicos.
- Existen multitud de libros, manuales y tutoriales libres que permiten su aprendizaje e ilustran el análisis estadístico de datos en distintas disciplinas científicas como las matemáticas, la física, la biología, la psicología, la medicina, etc.

Por defecto el entorno de trabajo de R es en línea de comandos, lo que significa que los cálculos y los análisis se realizan mediante comandos o instrucciones que el usuario teclea en una ventana de texto. No obstante, existen distintas interfaces gráficas de usuario que facilitan su uso, sobre todo para usuarios noveles. La interfaz gráfica que se utilizará para realizar estas prácticas será *RKWard*, desarrollada por Thomas Friedrichsmeier, junto al paquete *rkTeaching* especialmente desarrollado por el departamento de Matemáticas de la Universidad San Pablo CEU para la docencia de estadística.



El objetivo de esta práctica es introducir al alumno en la utilización de este programa, enseñándole a realizar las operaciones básicas más habituales de carga y manipulación de datos.

## 2 Instalación

La instalación del software necesario para realizar estas prácticas depende del sistema operativo.

### 2.1 Windows

Para usuarios de plataformas Windows existe un programa de instalación que incluye R, RKward y rkTeaching. El programa puede descargarse desde la dirección <https://aprendeconalf.es/proyecto/rkteaching/#instalaci%C3%B3n-en-windows>.

### 2.2 MacOs

Para usuarios de MacOs primero es necesario instalar R, después RKward y finalmente el paquete rkTeaching. Los pasos para realizar la instalación están descritos en la dirección <https://aprendeconalf.es/proyecto/rkteaching/#instalaci%C3%B3n-en-mac-os>.

## 2.3 Linux

Para usuarios de Linux primero es necesario instalar R, después RKWard y finalmente el paquete rkTeaching. Los pasos para realizar la instalación están descritos en la dirección <https://aprendeconalf.es/proyecto/rkteaching/#instalaci%C3%B3n-en-linux>.

## 3 Arranque

Como cualquier otra aplicación de Windows, para arrancar el programa hay que hacer clic sobre la opción correspondiente del menú Inicio » Programas » RKWard, o bien sobre el icono de escritorio



Al arrancar, aparece la ventana de bienvenida de RKWard (figura 1.1).

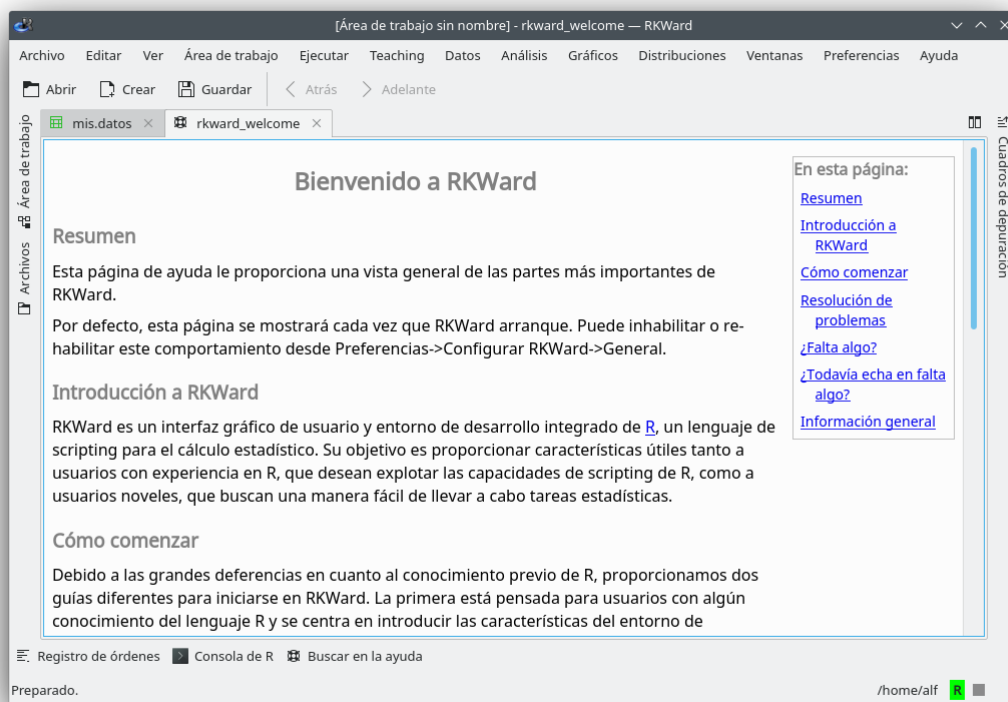


Figura 1.1 – Interfaz gráfica de usuario de RKWard.

La interfaz gráfica de usuario RKWard consta de los siguientes elementos:

- **Barra de menús.** Contiene distintos menús con operaciones que pueden realizarse con R. Si se ha instalado el paquete rkTeaching debe de aparecer el menú Teaching.
- **Barra de botones.** Contiene botones para abrir, crear y guardar conjuntos de datos, espacios de trabajo y guiones de comandos.

- **Ventana principal.** Es la ventana central donde aparecerán la ventana de introducción de datos, los resultados de los comandos ejecutados o de las búsquedas realizadas.
- **Área de trabajo.** Es una ventana desplegable al hacer clic sobre la solapa situada en el lado izquierdo que contiene todos los elementos del área de trabajo de R. Entre estos elementos aparecen los paquetes cargados, los conjuntos de datos y las variables que contienen los datos de la sesión actual.
- **Registro de órdenes** Es una solapa desplegable situada en la parte inferior donde aparece un registro de todas las acciones realizadas o comandos ejecutados en la sesión de trabajo actual. Cada vez que se seleccione un menú que lleve asociado la ejecución de algún comando, dicho comando aparecerá en esta ventana. Esto permite modificar fácilmente los parámetros del comando y volver a ejecutarlo rápidamente sin necesidad de volver al menú.
- **Consola de R** Es una solapa desplegable situada también en la parte inferior que da acceso al intérprete de comandos de R. En esta ventana pueden teclearse y ejecutarse directamente los comandos de R.
- **Buscar en la ayuda** Es una solapa desplegable situada en la parte inferior que permite hacer búsquedas sobre comandos de R o de algún paquete.
- **Mensajes.** Es la línea de texto que aparece en la parte inferior, donde se muestra información adicional sobre errores, advertencias u otra información auxiliar al ejecutar un comando, así como la ruta del área de trabajo activo.

## 4 Tipos de datos y operadores aritméticos y lógicos

En R existen distintos tipos de datos. Los más básicos son:

**Numeric** : Es cualquier número decimal. Se utiliza el punto como separador de decimales. Por defecto, cualquier número que se teclee tomará este tipo.

**Integer** : Es cualquier número entero. Para convertir un número de tipo Numeric en un entero se utiliza el comando `as.integer()`

**Logical** : Puede tomar cualquiera de los dos valores lógicos TRUE (verdadero) o FALSE (falso).

**Character** : Es cualquier cadena de caracteres alfanuméricos. Deben introducirse entre comillas. Para convertir cualquier número en una cadena de caracteres se utiliza el comando `as.character()`.

Los valores de estos tipos de datos pueden operarse utilizando distintos operadores o funciones predefinidas para cada tipo de datos. Los más habituales son:

**Operadores aritméticos** : + (suma), - (resta), \* (producto), / (cociente), ^ (potencia).

**Operadores de comparación** : > (mayor), < (menor), >= (mayor o igual), <= (menor o igual), == (igual), != (distinto).

**Operadores lógicos** : & (conjunción y), | (disyunción o), ! (negación no).

**Funciones predefinidas** : `sqrt()` (raíz cuadrada), `abs()` (valor absoluto), `log()` (logaritmo neperiano), `exp()` (exponencial), `sin()` (seno), `cos()` (coseno), `tan()` (tangente).

Al evaluar las expresiones aritméticas existe un orden de prioridad entre los operadores de manera que primero se evalúan las funciones predefinidas, luego las potencias, luego los productos y cocientes, luego las sumas y restas, luego los operadores de comparación, luego las negaciones, luego las conjunciones y finalmente las disyunciones. Para forzar un orden de evaluación distinto del predefinido se pueden usar paréntesis. Por ejemplo

```
> 2^2+4/2
[1] 6
> (2^2+4)/2
[1] 4
> 2^(2+4/2)
```



```
[1] 16
> 2^(2+4)/2
[1] 32
> 2^((2+4)/2)
[1] 8
```

También es posible asignar valores a variables mediante el operador de asignación `=`. Una vez definidas, las variables pueden usarse en cualquier expresión aritmética o lógica. Por ejemplo,

```
> x=2
> y=x+2
> y
[1] 4
> y>x
[1] TRUE
> x>=y
[1] FALSE
> x==y-2
[1] TRUE
> x!=0 & !y<x
[1] TRUE
```

## 5 Introducción y manipulación de datos

Antes de realizar cualquier análisis de datos hay que introducir los datos que se quieren analizar.

### 5.1 Introducción de datos en línea de comandos

Existen muchas formas de introducir datos en R pero aquí sólo veremos las más habituales. La forma más rápida de introducir datos es usar la consola de R para crear un vector de datos mediante el comando `c()`. Por ejemplo, para introducir las notas de 5 alumnos se debe teclear en la consola de R

```
> nota = c(5.6,7.2,3.5,8.1,6.4)
```

Esto crea el vector `nota` con el que posteriormente se pueden realizar cálculos como por ejemplo la media

```
> mean(nota)
[1] 6.16
```

Otra forma habitual de introducir los datos de una muestra es crear un conjunto de datos mediante el comando `data.frame()`. Por ejemplo, para crear un conjunto de datos a partir de las notas anteriores, hay que teclear

```
> curso = data.frame(nota)
```

Esto crea una matriz de datos en la que cada columna se corresponde con una variable y cada fila con un individuo de la muestra. En el ejemplo la matriz `curso` sólo tendría una columna que se correspondería con las notas y 5 filas, cada una de ellas correspondiente a un alumno de la muestra. Es posible acceder a las variables de un conjunto de datos con el operador dólar `$`. Por ejemplo, para acceder a las notas hay que teclear

```
> curso$nota
[1] 5.6 7.2 3.5 8.1 6.4
```

Es fácil añadir nuevas variables a un conjunto de datos, pero siempre deben tener el mismo tamaño muestral. Por ejemplo, para añadir una nueva variable con el grupo (mañana o tarde) de los alumnos, hay que teclear

```
> curso$grupo = c("m","t","t","m","m")
```

Ahora el conjunto de datos `curso` tendría dos columnas, una para la nota y otra para el grupo de los alumnos. Tecleando el nombre de cualquier objeto, se muestra su información:

```
> curso
  nota grupo
1  5.6     m
2  7.2     t
3  3.5     t
4  8.1     m
5  6.4     m
```

Cuando se introducen datos se puede utilizar el código NA (not available), para indicar la ausencia del dato.

Las variables definidas en cada sesión de trabajo quedan almacenadas en la memoria interna de R en lo que se conoce como *área de trabajo*. Es posible obtener un listado de todos los objetos almacenados en el área de trabajo mediante los comandos `ls()`. Si se desea más información, el comando `ls.str()` además de mostrar los objetos de la memoria indica sus tipos y sus valores.

```
> ls()
[1] "curso" "nota"  "x"      "y"
> ls.str()
curso : 'data.frame': 5 obs. of 2 variables:
 $ nota : num 5.6 7.2 3.5 8.1 6.4
 $ grupo: chr "m" "t" "t" "m" ...
nota : num [1:5] 5.6 7.2 3.5 8.1 6.4
x : num 2
y : num 4
```

Para eliminar un objeto de la memoria se utiliza el comando `rm()`.

```
> ls()
[1] "curso" "nota"  "x"      "y"
> rm(x,y)
> ls()
[1] "curso" "nota"
```

## 5.2 Introducción de datos en RKWard

RKWard dispone de una interfaz gráfica para introducir los datos sin necesidad de saberse los comandos anteriores. Para ello hay que ir al menú **Archivo** » **Nuevo** » **Conjunto de datos**. Con esto aparecerá una ventana donde hay que darle un nombre al conjunto de datos y tras esto aparece la ventana de la figura 1.2 con una tabla en la que se pueden introducir los datos de la muestra. Al igual que antes, cada variable debe introducirse en una columna y cada individuo en una fila.

Haciendo clic en las casillas de la cabecera cada fila es posible cambiar el nombre de la variable, ponerle una etiqueta, su tipo, su formato y los niveles en caso de tratarse de un factor o variable categórica. Los nombres de variables deben comenzar con una letra o un punto y pueden contener cualquier letra, punto, subrayado (`_`) o número. En particular, no se pueden utilizar espacios en blanco. Además, R es distingue entre mayúsculas y minúsculas.

Una vez definida la variable, para introducir los datos basta con teclearlos en las casillas que aparecen más abajo en la misma columna.

R permite definir más de un conjunto de datos en un mismo área de trabajo.

Los objetos definidos en el área de trabajo pueden verse haciendo clic en la solapa **Área de trabajo**. Para editar una variable o un conjunto de datos basta con hacer doble clic sobre él. También puede obtenerse un resumen como el que se muestra en la figura 1.3 haciendo clic en el botón derecho y seleccionando **ver** en el menú contextual que aparece.

## 5.3 Ponderación de datos

Cuando una variable o un conjunto de datos tiene unos pocos valores que se repiten mucho, en lugar de teclearlos es más rápido indicar los valores y ponderarlos por sus frecuencias. Para ello se utiliza el menú **Teaching**

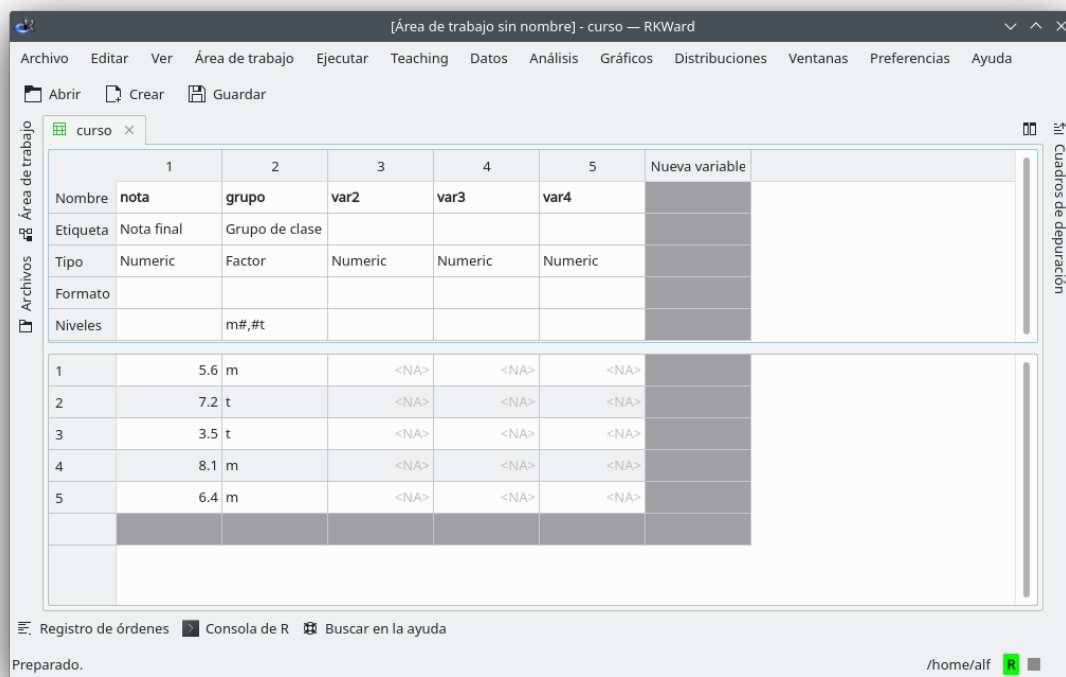


Figura 1.2 – Ventana de introducción de datos

[Datos] » [Ponderar datos]. Al seleccionarlo aparece una ventana donde hay que seleccionar el conjunto de datos a ponderar, la variable numérica de dicho conjunto de datos que contiene las frecuencias de ponderación, e indicar un nombre para el nuevo conjunto de datos. Por ejemplo, si en una clase hay 20 chicas y 30 chicos, se puede crear un conjunto de datos con la variables sexo y frecuencia, tal y como se muestra en la figura 1.4, y después llamar al menú de ponderación con los datos que aparecen la figura 1.5.

## 5.4 Guardar datos

Una vez introducidos los datos, conviene guardarlos en un fichero para no tener que volver a introducirlos en futuras sesiones. Para guardar los conjunto de datos definidos en el área de trabajo, se utiliza el menú [Área de trabajo] » [Guardar área de trabajo]. Con esto aparece una ventana donde hay que darle un nombre al fichero y seleccionar la carpeta donde se guardará. Los conjuntos de datos se guardan siempre en ficheros de R con extensión rda o rData.

También es posible guardar los datos en un fichero de texto plano mediante el menú [Archivo] » [Exportar] » [Exportar tabular data]. Tras esto aparece una ventana donde hay que seleccionar el conjunto de datos a exportar, darle un nombre al fichero de texto y seleccionar la carpeta donde se guardará. Esta ventana contiene también solapas donde se puede indicar entre otras cosas si incluir los nombres de las variables o no, el separador de decimales o el separador de los datos, que puede ser un espacio, tabuladores, comas u otro caracter.

## 5.5 Abrir datos

Si los datos con los que se pretende trabajar ya están guardados en un fichero de R, entonces tendremos que abrir dicho fichero. Para ello se utiliza el [Área de trabajo] » [Abrir área de trabajo] y en la ventana que aparece se selecciona el fichero que se desea abrir. Automáticamente se cargará el conjunto de datos del fichero y pasará a ser el conjunto de datos activo.

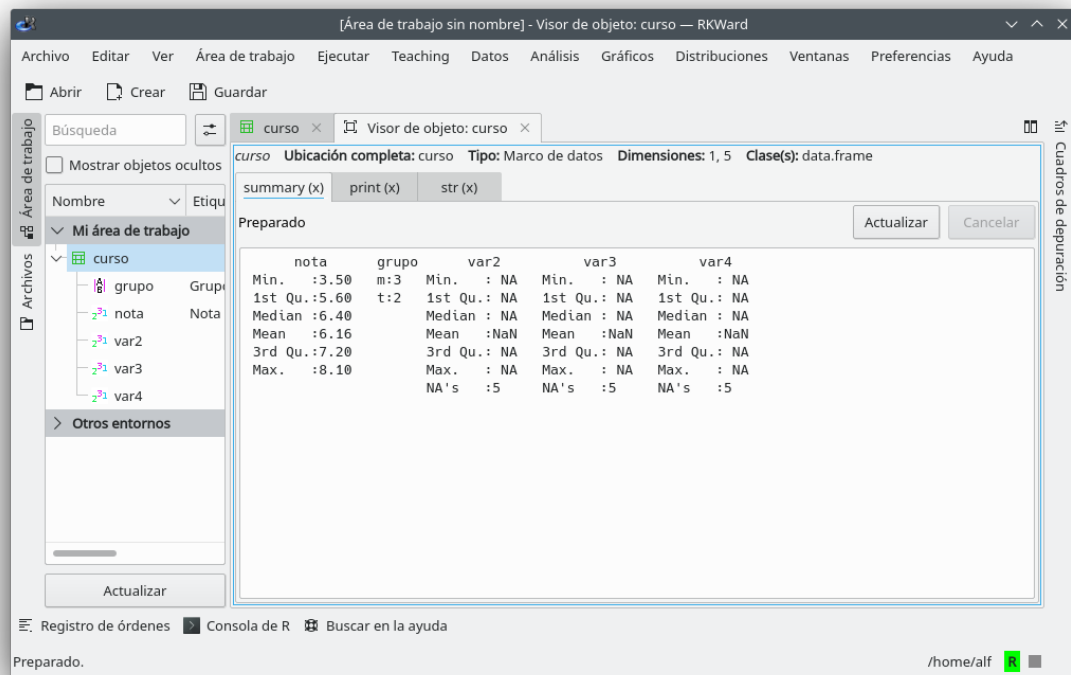


Figura 1.3 – Ventana de resumen descriptivo de un conjunto de datos

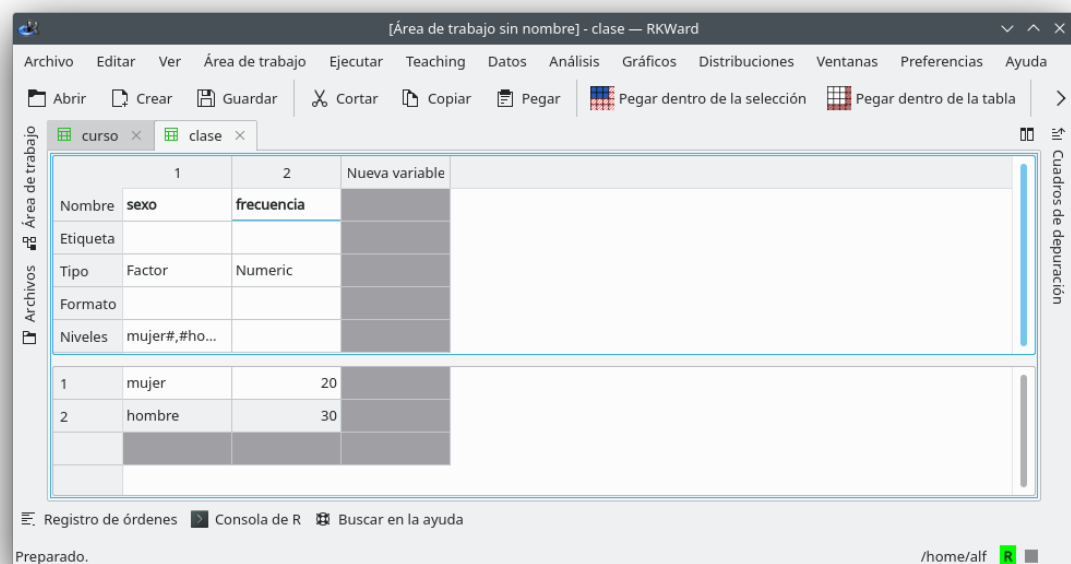


Figura 1.4 – Conjunto de datos preparado para ser ponderado

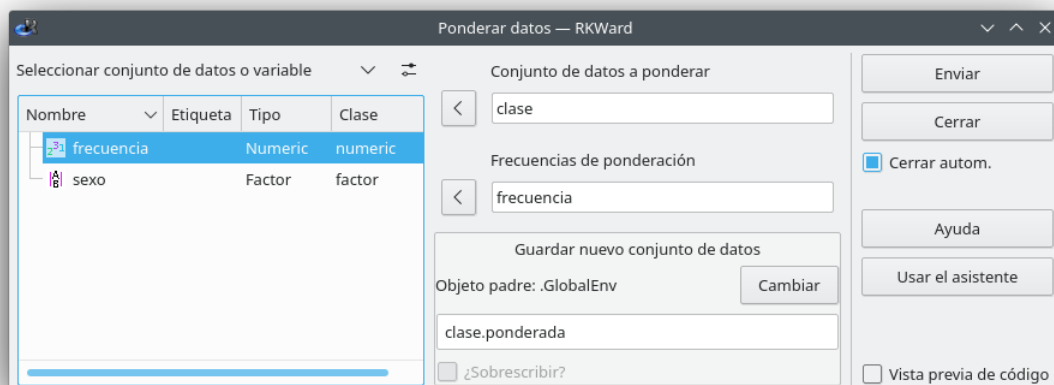


Figura 1.5 – Ventana de ponderación de datos

También es posible cargar datos de ficheros con otros formatos, como por ejemplo un fichero de texto. Para ello se utiliza el menú **Archivo** » **Importar** » **Importar datos** y en la ventana que aparece se selecciona el fichero de texto que se desea abrir y en el cuadro desplegable del formato de archivo se debes seleccionar Text. Después aparecerá una ventana donde habrá que darle un nombre al conjunto de datos y seleccionar el tipo de separador y si los nombres de las variables aparecen en la primera línea del fichero.

## 5.6 Eliminación de datos

Para eliminar una variable del conjunto de datos primero hay que editar el conjunto de datos, y después, en la ventana de edición de datos, hay que hacer clic con el botón derecho del ratón sobre la cabecera de la columna correspondiente y seleccionar en el menú contextual que aparece **Borrar esta variable**.

Para eliminar individuos del conjunto de datos que hacer clic con el botón derecho del ratón sobre la cabecera de la fila correspondiente y seleccionar en el menú contextual que aparece **Borrar esta fila**.

En la ventana del área de trabajo también es posible borrar cualquier objeto del área de trabajo de R haciendo clic con el botón derecho del ratón sobre él y seleccionando el menú **Eliminar**.

## 6 Transformación de datos

A menudo en los análisis hay que realizar transformaciones en los datos originales. A continuación se presentan las transformaciones más habituales.

### 6.1 Filtrado de datos

Cuando se desea realizar un análisis con un subconjunto de individuos del conjunto de datos activo que cumplen una determinada condición es posible filtrar el conjunto de datos para quedarse con esos individuos. Para ello se utiliza el menú **Teaching** » **Datos** » **Filtrar datos**. Con esto aparece un cuadro de diálogo en el que hay que seleccionar el conjunto de datos que se desea filtrar, y en el cuadro de texto Condición de selección indicar la condición lógica que tienen que cumplir los individuos seleccionados. También hay que indicar el nombre del nuevo conjunto de datos. Por ejemplo, para seleccionar los alumnos del grupo de la mañana habría que indicar la condición `grupo=="m"` tal y como se muestra en la figura 1.6.

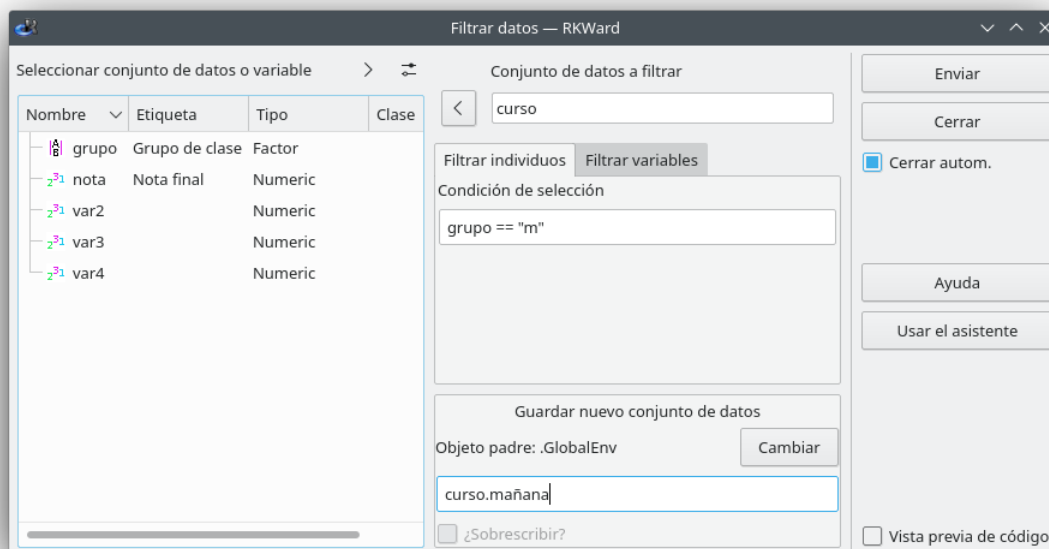


Figura 1.6 – Ventana de filtrado de datos.

## 6.2 Cálculo de variables

Para calcular una nueva variable a partir de otras ya existentes en el área de trabajo de R se utiliza el menú **Teaching > Datos > Calcular variable**. Con esto aparece un cuadro de diálogo en el que hay que introducir la expresión a partir de la que se calculará la nueva variable en el cuadro de texto **Expresión de cálculo**, e indicar el nombre de la nueva variable. La expresión de cálculo puede ser cualquier expresión aritmética o lógica de R, en las que pueden utilizarse cualquiera de las variables del área de trabajo de R. Por ejemplo, para eliminar los decimales de la variable **nota** podría crearse una nueva variable **puntuacion** multiplicando por 10 las notas, tal y como se muestra en la figura 1.7.

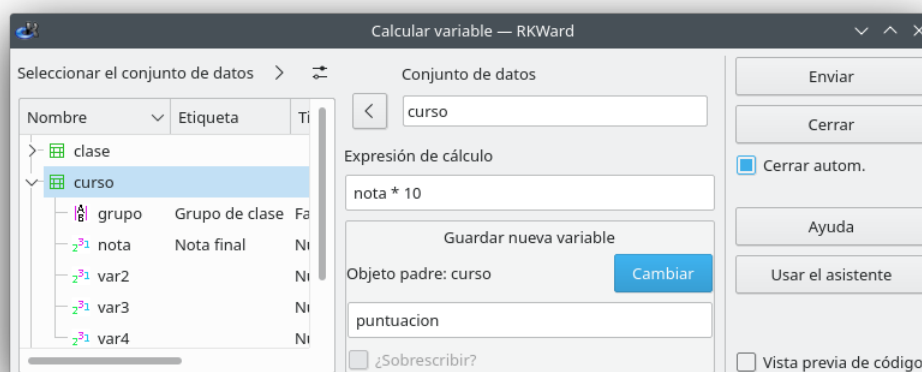


Figura 1.7 – Ventana de cálculo de nuevas variables.

## 6.3 Recodificación de variables

Otra transformación habitual es la recodificación de variables que permite transformar los valores de una variable de acuerdo a un conjunto de reglas de reescritura. Normalmente se utiliza para convertir una variable numérica en una variable categórica que pueda usarse como un factor.

Para recodificar una variable se utiliza el menú **Teaching > Datos > Recodificar variable**. Con esto aparece una ventana en la que hay que seleccionar la variable que se desea recodificar, indicar el nombre de la nueva variable recodificada e introducir las reglas de recodificación en el cuadro de texto Reglas de recodificación. Las reglas de recodificación siempre siguen la sintaxis **valor o rango de valores = nuevo valor** y pueden introducirse tantas reglas como se desee, cada una en una línea. Al lado izquierdo de la igualdad puede introducirse un único valor, varios valores separados por comas, o un rango de valores indicando el límite inferior y el límite superior del intervalo separados por el operador **:**. A la hora de definir el límite inferior puede utilizarse la palabra clave **lo** para referirse al menor de los valores de la muestra y **hi** para referirse al mayor de los valores. Por ejemplo, para recodificar la variable **nota** en categorías correspondientes a las calificaciones ([0,5) Suspenso, [5,7) Aprobado, [7,9) Notable y [9,10] Sobresaliente), habría que introducir las reglas que se muestran en la figura 1.8. Después, en la ventana de introducción de datos, se pueden renombrar los niveles del factor introduciendo el valor suspenso para la categoría 1, aprobado para la categoría 2, notable para la categoría 3 y sobresaliente para la categoría 4.

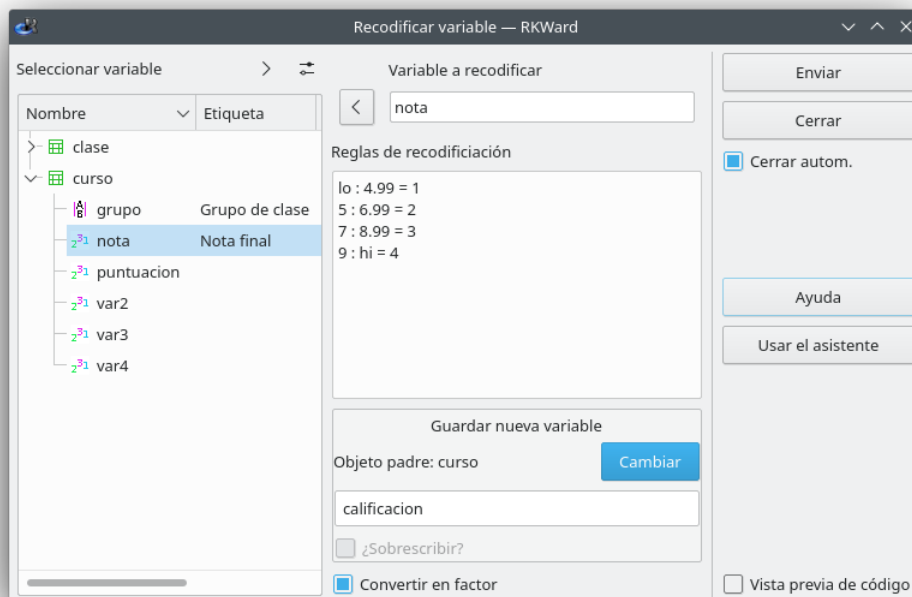


Figura 1.8 – Ventana de recodificación de variables

## 7 Manipulación de ficheros de resultados

### 7.1 Guardar los resultados

Cada vez que se ejecuta un comando de R, bien en la consola de comandos o a través de un menú, el comando ejecutado y su salida quedan registrados en la bitácora de comandos. Sin embargo, esta salida es en texto plano

sin formato por lo que muchos de los procedimientos recogidos en los menús producen además una salida mucho más comprensible en formato HTML en la ventana de resultados.

Para guardar el contenido de la ventana de resultados en un fichero se utiliza el menú **Archivo** » **Exportar página como HTML**. Con esto aparece un cuadro de diálogo en el que hay que indicar el nombre del fichero y la carpeta donde se desea guardar. El fichero resultante está en formato HTML por lo que se podrá visualizar con cualquier navegador web.

## 7.2 Limpiar la ventana de resultados

La ventana de resultados va acumulando todas las salidas de los análisis realizados en cada sesión de trabajo. Para no mezclar los resultados de estudios distintos, conviene limpiar la ventana de resultados cada vez que se empiece un estudio nuevo. Para ello hay que seleccionar el menú **Edición** » **Limpiar salida**.

# 8 Manipulación de guiones de comandos

## 8.1 Creación de un guión de comandos

RKward también incorpora un entorno de desarrollo para programadores de R que permite crear guiones de comandos que pueden ejecutarse todos seguidos. Esta opción es muy interesante para repetir análisis o automatizar tareas repetitivas. Para crear un guión de comandos hay que seleccionar el menú **Archivo** » **Nuevo** » **Archivo de guiones**. Con esto aparecerá una venta como la que aparece en la figura 1.9 donde se podrán teclear los comandos de R para después ejecutarlos uno a uno o en bloque.

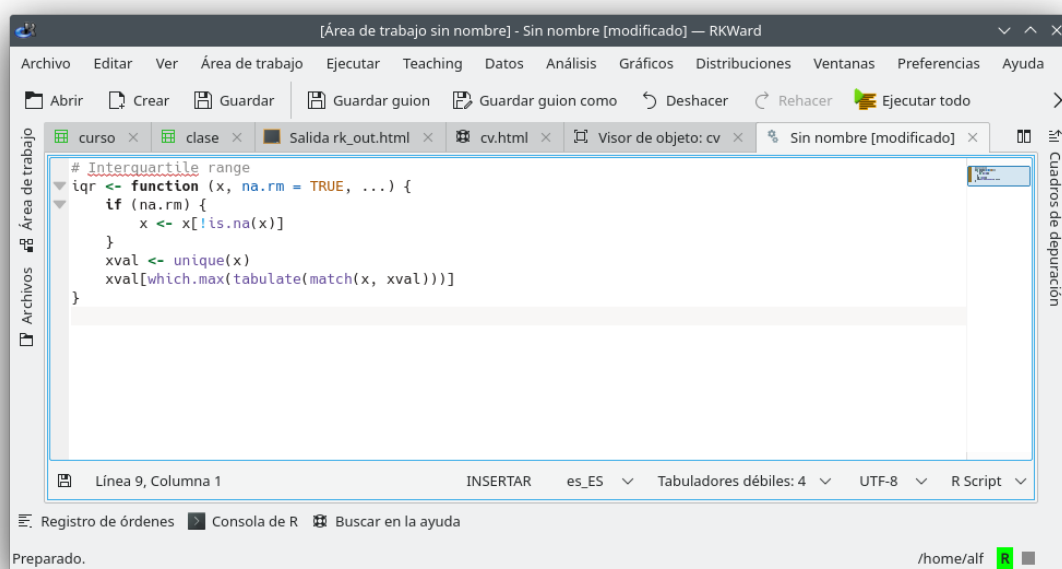


Figura 1.9 – Ventana de edición de guiones de comandos



## 8.2 Guardar un guión de comandos

Los guiones de comandos también pueden guardarse en un fichero de texto plano mediante el menú **Archivo** **Guardar guión** e indicando el nombre del fichero y la carpeta donde se guardará en el cuadro de diálogo que aparece.

## 8.3 Abrir un guión de comandos

Para abrir un fichero con un guión de comandos se utiliza el menú **Archivo** **Abrir archivo de guiones de R** y después seleccionar el fichero que se desea abrir en el cuadro de diálogo que aparece.

# 9 Ayuda

Otra de las ventajas de R es que tiene un sistema de ayuda muy documentado. Es posible conseguir ayuda sobre cualquier función, procedimiento o paquete simplemente tecleando el comando `help()`. Por ejemplo, para obtener ayuda sobre el comando `mean` se teclearía

```
> help("mean")
```

y con esto aparecerá una ventana de ayuda donde se describe la función y también aparecen ejemplos que ilustran su uso. Si no se conoce exactamente el nombre de la función o comando, se puede hacer una búsqueda aproximada con el comando `help.search()`. Por ejemplo, si no se recuerda el nombre de la función logarítmica, se podría teclear

```
> help("logarithm")
```

y con esto aparecerá una ventana con todos los ficheros de ayuda que contienen la palabra `logarithm`.

Finalmente, también es posible invocar la ayuda general de R en RKWard con el menú **Ayuda** **Ayuda de R** con lo que aparecerá una página web desde donde podremos navegar a la información deseada. También es posible buscar ayuda sobre un comando concreto en el menú **Ayuda** **Buscar en la ayuda de R**.

Para más información sobre R se recomienda visitar la página <http://www.r-project.org/>, y para más información sobre RKWard se recomienda visitar la página <https://rkward.kde.org/>.

## 10 Ejercicios resueltos

1. Crear un conjunto de datos con los datos de la siguiente muestra y guardarlo con el nombre `colesterol.rda`

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



Para crear el conjunto de datos:

- (a) Seleccionar el menú **Archivo** » **Nuevo** » **Conjunto de datos**.
- (b) En el cuadro de diálogo que aparece introducir el nombre del conjunto de datos `colesterol` y hacer clic en el botón **Aceptar**.
- (c) En la ventana del editor de datos hay que definir una variable en cada columna introduciendo su nombre y tipo en las casillas de la cabecera de cada columna.
- (d) Una vez definidas las variables hay que introducir los datos de cada variable en la columna correspondiente.

Para guardar los datos:

- (a) Seleccionar el menú **Área de trabajo** » **Guardar área de trabajo**.
- (b) En el cuadro de diálogo que aparece hay que darle un nombre al fichero, seleccionar la carpeta donde guardarlo y hacer clic en el botón **Aceptar**.

2. Abrir el fichero creado en el ejercicio anterior y realizar las siguientes operaciones:

- (a) Insertar 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



Para abrir el conjunto de datos del ejercicio anterior:

1. Seleccionar el menú **Área de trabajo** » **Abrir área de trabajo**.
2. En el cuadro de diálogo que aparece seleccionar la carpeta donde se encuentra el fichero con los datos del ejercicio anterior, seleccionar el fichero y hacer clic en el botón **Aceptar**.

Para insertar la variable `Edad`:

1. Hacer clic en la solapa **Área de trabajo**.
2. En la ventana del área de trabajo doble clic sobre el conjunto de datos `colesterol`.
3. En la ventana del editor de datos introducir el nombre de la variable `edad` y su tipo en las casillas de la cabecera de una nueva columna vacía, e introducir los datos de las edades en las celdas de más abajo.

- (b) Insertar un nuevo individuo con siguientes datos

Nombre: Cristóbal Campos Ruiz.

Edad: 44 años.

Sexo: Hombre.

Peso: 70 Kg.

Altura: 178 cm.

Colesterol: 220 mg/dl.



1. En la ventana del editor de datos introducir los datos de del nuevo individuo en la primera fila vacía.

- (c) Crear una nueva variable donde se calcule el índice de masa corporal de cada paciente mediante la formula:

$$\text{imc} = \frac{\text{Peso (en Kg)}}{\text{Altura (en mt)}^2}$$



1. Seleccionar el menú **Teaching >> Datos >> Calcular variable**.
2. En el cuadro de diálogo que aparece introducir la fórmula para calcular el índice de masa corporal en el campo Expresión de cálculo.
3. En el cuadro Guardar nueva variable hacer clic sobre el botón Cambiar.
4. En el cuadro de diálogo que aparece seleccionar como objeto padre la el conjunto de datos colesterol y hacer clic sobre el botón Aceptar.
5. Introducir el nombre de la nueva variable imc y hacer clic sobre el botón Aceptar.

- (d) Recodificar el índice de masa corporal en una nueva variable de acuerdo a las siguientes categorías:

Menor de 18.5	Bajo peso
De 18.5 a 24.5	Saludable
De 24.5 a 30	Sobrepeso
Mayor de 30	Obeso



1. Seleccionar el menú **Teaching >> Datos >> Recodificar variable**.
2. En el cuadro de diálogo que aparece seleccionar como variable a recodificar la variable imc.
3. Introducir las reglas de recodificación en el campo Reglas de recodificación:
  - 10:18.5 = 1
  - 18.5:24.5 = 2
  - 24.5:30 = 3
  - 30:hi = 4
4. En el cuadro Guardar nueva variable hacer clic sobre el botón Cambiar.
5. En el cuadro de diálogo que aparece seleccionar como objeto padre la el conjunto de datos colesterol y hacer clic sobre el botón Aceptar.
6. Introducir el nombre de la nueva variable obesidad y hacer clic sobre el botón Aceptar.
7. En la ventana de edición de datos introducir los niveles del factor, asignando Bajo peso a la categoría 1, Saludable a la categoría 2, Sobrepeso a la categoría 3 y Obeso a la categoría 4.

- (e) Filtrar el conjunto de datos para obtener un nuevo conjunto de datos con los datos de los hombres



1. Seleccionar el menú **Teaching >> Datos >> Filtrar**.
2. En el cuadro de diálogo que aparece seleccionar como conjunto de datos colesterol.
3. En el campo Condición de selección introducir la condición `sexo=="H"`.
4. Introducir el nombre del nuevo conjunto de datos colesterol.hombres y hacer clic sobre el botón Aceptar.

## 11 Ejercicios propuestos

1. El conjunto de datos neonatos del paquete `rk.Teaching`, contiene información sobre una muestra de 320 recién nacidos en un hospital durante un año que cumplieron el tiempo normal de gestación. Se pide:

(a) Cargar el conjunto de datos.



1. Hacer clic en la solapa Área de trabajo para desplegarla y ver los paquetes del área de trabajo.
2. Hacer doble clic sobre el paquete `rk.Teaching` para ver todos los conjuntos de datos que contiene.
3. Hacer clic con el botón derecho sobre el conjunto de datos `neonatos` y en el menú contextual que aparece seleccionar `Copiar a .GlobalEnv` para hacer una copia del conjunto de datos en nuestro entorno de trabajo.

- (b) Calcular la variable `apgar.medio` como la media de las variables `apgar1` y `apgar5`.
- (c) Recodificar la variable `peso` en el factor `categoria.peso` con dos categorías que se correspondan con los pesos menores y mayores de 2.5 Kg.
- (d) Recodificar la variable `apgar1` en el factor `estado.apgar1` con tres categorías: deprimido ( $\text{Apgar} \leq 3$ ), moderadamente deprimido ( $3 < \text{Apgar} \leq 6$ ) y normal ( $\text{Apgar} > 6$ ).
- (e) Filtrar el conjunto de datos para quedarse con los hijos de las madres no fumadoras con una puntuación Apgar al minuto de nacer menor o igual que 3. ¿Cuántos niños hay?

# Distribuciones de Frecuencias y Representaciones Gráficas

## 1 Ejercicios resueltos

1. En una encuesta a 25 matrimonios sobre el número de hijos que tenían se obtuvieron los siguientes datos:

1, 2, 4, 2, 2, 2, 3, 2, 1, 1, 0, 2, 2, 0, 2, 2, 1, 2, 2, 3, 1, 2, 2, 1, 2

Se pide:

- Crear un conjunto de datos con la variable hijos e introducir los datos.
- Construir la tabla de frecuencias.



- Seleccionar el menú **Teaching** » **Distribución de frecuencias** » **Tabla de frecuencias**.
- En el cuadro de diálogo que aparece, seleccionar la variable hijos en el campo Variable a tabular y hacer clic en el botón Enviar.

- Dibujar el diagrama de barras de las frecuencias absolutas.



- Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de barras**.
- En el cuadro de diálogo que aparece, seleccionar la variable hijos en el campo Variable y hacer clic en el botón Enviar.

- Para la misma tabla de frecuencias anterior, dibujar también el diagrama de barras de las frecuencias relativas, el de absolutas acumuladas y el de relativas acumuladas, además de sus correspondientes polígonos.



Repetir los pasos del apartado anterior activando, en la solapa de Opciones de las barras, la opción Frecuencias relativas si se desea el diagrama de barras de frecuencias relativas, activando la opción Frecuencias acumuladas si se desea el diagrama de barras de frecuencias acumuladas y activando la opción Polígono para obtener el polígono asociado.

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

15, 23, 12, 10, 28, 50, 12, 17, 20, 21, 18, 13, 11, 12, 26  
30, 6, 16, 19, 22, 14, 17, 21, 28, 9, 16, 13, 11, 16, 20

Se pide:

- Crear un conjunto de datos con la variable urgencias e introducir los datos.
- Dibujar el diagrama de cajas. ¿Existe algún dato atípico? En el caso de que exista, eliminarlo y proceder con los siguientes apartados.



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de cajas**.
2. En el cuadro de diálogo que aparece, seleccionar la variable urgencias en el campo Variables y hacer clic en el botón Enviar.
3. En la ventana que aparece con el diagrama de cajas identificar el dato atípico.
4. Ir a la ventana de edición de datos y eliminar la fila del dato atípico haciendo clic con el botón derecho del ratón en la cabecera de la fila y seleccionando **Borrar esta fila**.

(c) Construir la tabla de frecuencias agrupando en 5 clases.



1. Seleccionar el menú **Teaching** » **Distribución de frecuencias** » **Tabla de frecuencias**.
2. En el cuadro de diálogo que aparece seleccionar la variable urgencias.
3. En la solapa de Clases activar la casilla Agrupar en intervalos, marcar la opción Número de intervalos e introducir el número deseado de intervalos en el campo Intervalos sugeridos y hacer clic sobre el botón Enviar.

(d) Dibujar el histograma de frecuencias absolutas correspondiente a la tabla anterior.



1. Seleccionar el menú **Teaching** » **Gráficos** » **Histograma**.
2. En el cuadro de diálogo que aparece seleccionar la variable urgencias en el campo Variable.
3. En la solapa de Clases activar la casilla Agrupar en intervalos, marcar la opción Número de intervalos e introducir el número deseado de intervalos en el campo Intervalos sugeridos y hacer clic sobre el botón Enviar.

(e) Para la misma tabla de frecuencias anterior, dibujar también el histograma de las frecuencias relativas, el de absolutas acumuladas y el de relativas acumuladas, además de sus correspondientes polígonos.



Repetir los pasos del apartado anterior activando, en la solapa de Opciones del histograma, la opción Frecuencias relativas si se desea el histograma de frecuencias relativas, activando la opción Frecuencias acumuladas si se desea el histograma de frecuencias acumuladas y activando la opción Polígono para obtener el polígono asociado.

3. Los grupos sanguíneos de una muestra de 30 personas son:

A, B, B, A, AB, o, o, A, B, B, A, A, A, A, AB,  
A, A, A, B, o, B, B, B, A, A, A, o, A, AB, o.

Se pide:

- (a) Crear un conjunto de datos con la variable grupo.sanguineo e introducir los datos.
- (b) Construir la tabla de frecuencias.



1. Seleccionar el menú **Teaching** » **Distribución de frecuencias** » **Tabla de frecuencias**.
2. En el cuadro de diálogo que aparece, seleccionar la variable grupo.sanguineo en el campo Variable a tabular y hacer clic en el botón Enviar.

(c) Dibujar el diagrama de sectores.



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de sectores**.
2. En el cuadro de diálogo que aparece, seleccionar la variable grupo.sanguineo en el campo Variables y hacer clic sobre el botón Enviar.

4. En un estudio de población se tomó una muestra de 27 personas, y se les preguntó por su edad y estado civil, obteniendo los siguientes resultados:

## 2. Distribuciones de Frecuencias y Representaciones Gráficas

Estado civil	Edad									
Soltero	31	45	35	65	21	38	62	22	31	
Casado	62	39	62	59	21	62				
Viudo	80	68	65	40	78	69	75			
Divorciado	31	65	59	49	65					

Se pide:

- Crear un conjunto de datos con la variables estado.civil y edad e introducir los datos.
- Construir la tabla de frecuencias de la variable edad para cada categoría de la variable estado.civil.



- Seleccionar el menú **Teaching** » **Distribución de frecuencias** » **Tabla de frecuencias**.
- En el cuadro de diálogo que aparece, seleccionar la variable edad en el campo Variable a tabular, activar la casilla Tabular por grupos, seleccionar la variable estado.civil en el campo Variable de agrupación y hacer clic en el botón Enviar.

- Dibujar los diagramas de cajas de la edad según el estado civil. ¿Existen datos atípicos? ¿En qué grupo hay mayor dispersión?



- Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de cajas**.
- En el cuadro de diálogo que aparece, seleccionar la variable edad en el campo Variables, activar la casilla Dibujar por grupos, seleccionar la variable estado.civil en el campo Variable de agrupación y hacer clic en el botón Enviar.

## 2 Ejercicios propuestos

- 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:

- Construir la tabla de frecuencias.
  - Dibujar el diagrama de barras de las frecuencias relativas y de frecuencias relativas acumuladas.
  - Dibujar el diagrama de sectores.
- 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:

- Dibujar el histograma de las frecuencias absolutas agrupando desde 150 a 200 en clases de amplitud 10.
  - Dibujar el diagrama de cajas. ¿Existe algún dato atípico?
- El conjunto de datos neonatos del paquete rk.Teaching, contiene información sobre una muestra de 320 recién nacidos en un hospital durante un año que cumplieron el tiempo normal de gestación. Se pide:

- (a) Construir la tabla de frecuencias de la puntuación Apgar al minuto de nacer. Si se considera que una puntuación Apgar de 3 o menos indica que el neonato está deprimido, ¿qué porcentaje de niños está deprimido en la muestra?
- (b) Comparar las distribuciones de frecuencias de las puntuaciones Apgar al minuto de nacer según si la madre es mayor o menor de 20 años. ¿En qué grupo hay más neonatos deprimidos?
- (c) Construir la tabla de frecuencias para el peso de los neonatos, agrupando en clases de amplitud 0.5 desde el 2 hasta el 4.5. ¿En qué intervalo de peso hay más niños?
- (d) Comparar la distribución de frecuencias relativas del peso de los neonatos según si la madre fuma o no. Si se considera como peso bajo un peso menor de 2.5 kg, ¿En qué grupo hay un mayor porcentaje de niños con peso bajo?
- (e) Si en los recién nacidos se considera como peso bajo un peso menor de 2.5 kg, calcular la prevalencia del bajo peso de recién nacidos en el grupo de madres fumadoras y en el de no fumadoras.
- (f) Calcular el riesgo relativo de que un recién nacido tenga bajo peso cuando la madre fuma, frente a cuando la madre no fuma.
- (g) Construir el diagrama de barras de la puntuación Apgar al minuto. ¿Qué puntuación Apgar es la más frecuente?
- (h) Construir el diagrama de frecuencias relativas acumuladas de la puntuación Apgar al minuto. ¿Por debajo de qué puntuación estarán la mitad de los niños?
- (i) Comparar mediante diagramas de barras de frecuencias relativas las distribuciones de las puntuaciones Apgar al minuto según si la madre ha fumado o no durante el embarazo. ¿Qué se puede concluir?
- (j) Construir el histograma de pesos, agrupando en clases de amplitud 0.5 desde el 2 hasta el 4.5. ¿En qué intervalo de peso hay más niños?
- (k) Comparar la distribución de frecuencias relativas del peso de los neonatos según si la madre fuma o no. ¿En qué grupo se aprecia menor peso de los niños de la muestra?
- (l) Comparar la distribución de frecuencias relativas del peso de los neonatos según si la madre fumaba o no antes del embarazo. ¿Qué se puede concluir?
- (m) Construir el diagrama de caja y bigotes del peso. ¿Entre qué valores se considera que el peso de un neonato es normal? ¿Existen datos atípicos?
- (n) Comparar el diagrama de cajas y bigotes del peso, según si la madre fumó o no durante el embarazo y si era mayor o no de 20 años. ¿En qué grupo el peso tiene más dispersión central? ¿En qué grupo pesan menos los niños de la muestra?
- (o) Comparar el diagrama de cajas de la puntuación Apgar al minuto y a los cinco minutos. ¿En qué variable hay más dispersión central?



# Estadísticos Muestrales

## 1 Ejercicios resueltos

1. En una encuesta a 25 matrimonios sobre el número de hijos que tenían se obtuvieron los siguientes datos:

1, 2, 4, 2, 2, 2, 3, 2, 1, 1, 0, 2, 2, 0, 2, 2, 1, 2, 2, 3, 1, 2, 2, 1, 2

Se pide:

- Crear un conjunto de datos con la variable hijos e introducir los datos. Si ya se tienen los datos, simplemente recuperarlos.
- Calcular la media aritmética, varianza y desviación típica de dicha variable. Interpretar los estadísticos.



- Seleccionar el menú **Teaching** » **Estadística descriptiva** » **Estadísticos**.
- En el cuadro de diálogo que aparece seleccionar la variable hijos en el campo Variables.
- En la solapa **Estadísticos básicos** seleccionar Media y Desviación típica, y hacer click sobre el botón Enviar.

- Calcular los cuartiles, el recorrido, el rango intercuartílico, el tercer decil y el percentil 68.



- Seleccionar el menú **Teaching** » **Estadística descriptiva** » **Estadísticos**.
- En el cuadro de diálogo que aparece seleccionar la variable hijos en el campo Variables.
- En la solapa **Estadísticos básicos** seleccionar Cuartiles, Rango, Rango intercuartílico, introducir los valores 0.3 y 0.68 en el campo Percentiles, y hacer click sobre el botón Enviar.

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

15, 23, 12, 10, 28, 50, 12, 17, 20, 21, 18, 13, 11, 12, 26  
30, 6, 16, 19, 22, 14, 17, 21, 28, 9, 16, 13, 11, 16, 20

Se pide:

- Crear un conjunto de datos con la variable urgencias e introducir los datos.
- Calcular la media aritmética, varianza, desviación típica y coeficiente de variación de dicha variable. Interpretar los estadísticos.



- Seleccionar el menú **Teaching** » **Estadística descriptiva** » **Estadísticos**.
- En el cuadro de diálogo que aparece seleccionar la variable urgencias en el campo Variables.
- En la solapa **Estadísticos básicos** seleccionar Media, Varianza, Desviación típica y Coeficiente de variación, y hacer click sobre el botón Enviar.

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



Seguir los mismos pasos del apartado anterior, seleccionando Coeficiente de asimetría y Coeficiente de Curtosis en la solapa **Estadísticos básicos**.

3. 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:

- Crear un conjunto de datos curso con la variable calificaciones e introducir los datos.
- Recodificar esta variable, asignando 2.5 al SS, 6 al AP, 8 al NT y 9.5 al SB.



- Seleccionar el menú **Teaching > Datos > Recodificar variable**.
- En el cuadro de diálogo que aparece seleccionar como variable a recodificar la variable calificaciones.
- Introducir las reglas de recodificación en el campo Reglas de recodificación:
 

```
"SS" = 2.5
"AP" = 6
"NT" = 8
"SB" = 9.5
```
- En el cuadro Guardar nueva variable hacer click sobre el botón Cambiar.
- En el cuadro de diálogo que aparece seleccionar como objeto padre la el conjunto de datos curso y hacer click sobre el botón Enviar.
- Introducir el nombre de la nueva variable nota, desmarcar la casilla Convertir en factor y hacer click sobre el botón Enviar.

- La mediana y el rango intercuartílico.



- Seleccionar el menú **Teaching > Estadística descriptiva > Estadísticos**.
- En el cuadro de diálogo que aparece seleccionar la variable nota en el campo Variables.
- En la solapa **Estadísticos básicos** seleccionar Mediana y Rango intercuartílico, y hacer click sobre el botón Enviar.

4. Para realizar un estudio sobre la estatura de los estudiantes universitarios se ha seleccionado mediante un proceso de muestreo aleatorio, una muestra de 30 estudiantes, obteniendo los siguientes resultados (medidos en centímetros):

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

Se pide:

- Crear un conjunto de datos con las variables estatura y sexo e introducir los datos.
- Obtener un resumen de estadísticos en el que se muestren la media aritmética, mediana, varianza, desviación típica y cuartiles según el sexo. Interpretar los estadísticos.



- Seleccionar el menú **Teaching > Estadística descriptiva > Estadísticos**.
- En el cuadro de diálogo que aparece seleccionar la variable estatura en el campo Variables, marcar la casilla Estadística por grupos y seleccionar la variable sexo en el campo Variables de agrupación.

3. En la solapa **Estadísticos básicos** seleccionar Media, Mediana, Varianza, Desviación típica y Cuartiles, y hacer click sobre el botón Enviar.

## 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:

- Calcular la media aritmética, mediana, varianza y desviación típica de las lesiones e interpretarlas.
  - Calcular los coeficientes de asimetría y curtosis e interpretarlos.
  - Calcular el cuarto y el octavo decil e interpretarlos.
2. En un estudio de población se tomó una muestra de 27 personas, y se les preguntó por su edad y estado civil, obteniendo los siguientes resultados:

Estado civil	Edad									
Soltero	31	45	35	65	21	38	62	22	31	
Casado	62	39	62	59	21	62				
Viudo	80	68	65	40	78	69	75			
Divorciado	31	65	59	49	65					

Se pide:

- Calcular la media y la desviación típica de la edad según el estado civil e interpretarlas.
  - ¿En qué grupo es más representativa la media?
3. En un estudio se ha medido la tensión arterial de 25 individuos. Además se les ha preguntado si fuman y beben:

Fumador	si	no	si	si	si	no	no	si	no	si	no	si	no
Bebedor	no	no	si	si	no	no	si	si	no	si	no	si	si
Tensión arterial	80	92	75	56	89	93	101	67	89	63	98	58	91

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	
Tensión arterial	71	52	98	104	57	89	70	93	69	82	70	49	

Calcular la media aritmética, desviación típica, coeficiente de asimetría y curtosis de la tensión arterial por grupos dependiendo de si beben o fuman e interpretarlos.

4. El conjunto de datos neonatos del paquete `rk. Teaching`, contiene información sobre una muestra de 320 recién nacidos en un hospital durante un año que cumplieron el tiempo normal de gestación. Se pide:
- Calcular la media y la mediana muestral del peso de los nacidos e interpretarlos.
  - Calcular el peso medio de los recién nacidos de la muestra según si la madre ha fumado o no durante el embarazo. Calcular también el peso medio de los recién nacidos de madres que no han fumado durante el embarazo, según si la madre fumaba o no antes del embarazo. ¿Qué conclusiones se pueden sacar?
  - ¿Cuál es la puntuación Apgar al minuto de nacer más frecuente?
  - Calcular la media de la diferencia entre las puntuaciones Apgar a los 5 minutos y al minuto de nacer. ¿Cómo evolucionan los recién nacidos?
  - Calcular los cuartiles muestrales del peso de los recién nacidos e interpretarlos.

- (f) Comparar los cuartiles muestrales del peso de los recién nacidos según el sexo.
  - (g) ¿Por encima de qué peso estarán el 10% de los niños con mayor peso?
  - (h) Si se considera que un niño es atípico por bajo peso si se encuentra entre el 5% de los pesos más bajos, ¿por debajo de qué peso tiene que estar?
  - (i) Calcular el recorrido y el rango intercuartílico muestrales del peso de los recién nacidos e interpretarlos.
  - (j) Calcular la varianza y la desviación típica del peso de los recién nacidos e interpretarlos.
  - (k) ¿En qué grupo hay más variabilidad del peso de los recién nacidos, en las madres fumadoras o en las madres no fumadoras durante el embarazo? ¿En qué grupo será más representativo el peso medio?
  - (l) ¿Qué variable presenta más variabilidad relativa, el peso de los recién nacidos o el Apgar al minuto de nacer?
  - (m) Calcular el coeficiente de asimetría y de apuntamiento muestrales del peso de los recién nacidos e interpretarlos.
  - (n) ¿Qué distribución es más asimétrica, la de los pesos de recién nacidos en madres mayores de 20 años o en madres menores de 20 años?
  - (o) ¿Qué distribución es más apuntada, la del peso de los recién nacidos en hombres o en mujeres?
  - (p) De acuerdo a la forma de la distribución, ¿puede considerarse la puntuación Apgar al minuto de nacer como una variable normal? ¿Y el número de cigarros fumados al día durante el embarazo?
5. Se quiere comparar la precisión de dos tensiómetros, uno de brazo y otro de muñeca, y para ello se han realizado 8 medidas repetidas de la tensión arterial de una misma persona con cada uno de ellos, obteniendo los siguientes valores en mmHg:
- tens.brazo: 111, 109, 112, 111, 113, 113, 114, 111.
  - tens.muñeca: 115, 113, 117, 116, 112, 112, 117, 112.
- ¿Qué tensiómetro es más preciso?

# Regresión Lineal Simple y Correlación

## 1 Ejercicios resueltos

1. Se han medido dos variables  $X$  e  $Y$  en 10 individuos obteniendo los siguientes resultados:

$X$	0	1	2	3	4	5	6	7	8	9
$Y$	2	5	8	11	14	17	20	23	26	29

Se pide:

- Crear un conjunto de datos con las variables  $X$  y  $Y$  e introducir estos datos.
- Dibujar el diagrama de dispersión correspondiente.



- Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de Dispersión**.
- En el cuadro de diálogo que aparece, seleccionar la variable  $Y$  en el campo Variable  $Y$ , la variable  $X$  en el campo Variable  $X$ , y hacer clic en el botón Enviar.

En vista del diagrama, ¿qué tipo de modelo crees que explicará mejor la relación entre  $X$  e  $Y$ ?

- Calcular la recta de regresión de  $Y$  sobre  $X$ .



- Seleccionar el menú **Teaching** » **Regresión** » **Regresión lineal**.
- En el cuadro de diálogo que aparece, seleccionar la variable  $Y$  en el campo Variable dependiente y la variable  $X$  en el campo Variable independiente, y hacer clic sobre el botón Enviar.

- Dibujar dicha recta sobre el diagrama de dispersión.



- Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de Dispersión**.
- En el cuadro de diálogo que aparece, seleccionar la variable  $Y$  en el campo Variable  $Y$ , la variable  $X$  en el campo Variable  $X$ , y hacer clic en el botón Enviar.
- En la solapa **Línea de ajuste**, seleccionar Dibujar recta de regresión y hacer clic en el botón Enviar.

- Calcular la recta de regresión de  $X$  sobre  $Y$  y dibujarla sobre el correspondiente diagrama de dispersión.



Repetir los pasos de los apartados anteriores pero escogiendo como Variable dependiente la variable  $X$ , y como Variable independiente la variable  $Y$

- ¿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 un conjunto de datos con las variables horas.estudio y suspensos e introducir estos datos.
- Construir la tabla de frecuencias bidimensional de las variables horas.estudio y suspensos.



- Seleccionar el menú [Teaching](#) » [Distribución de frecuencias](#) » [Tabla de frecuencias bidimensional](#).
- En el cuadro de diálogo que aparece, seleccionar la variable horas.estudio en el campo Variable a tabular en filas, la variable suspensos en el campo Variable a tabular en columnas, y hacer clic sobre el botón Enviar.

- Calcular la recta de regresión de suspensos sobre horas.estudio y dibujarla.



Para calcular la recta de regresión:

- Seleccionar el menú [Teaching](#) » [Regresión](#) » [Regresión lineal](#).
- En el cuadro de diálogo que aparece, seleccionar la variable suspensos en el campo Variable dependiente y la variable horas.estudio en el campo Variable independiente, seleccionar Guardar el modelo, introducir un nombre para el modelo y hacer clic sobre el botón Enviar.

Para dibujar la recta de regresión:

- Seleccionar el menú [Teaching](#) » [Gráficos](#) » [Diagrama de Dispersión](#).
- En el cuadro de diálogo que aparece, seleccionar la variable suspensos en el campo Variable Y y la variable horas.estudio en el campo Variable X.
- En la solapa [Línea de ajuste](#), seleccionar Lineal y hacer clic en el botón Enviar.

- Indicar el coeficiente de regresión de suspensos sobre horas.estudio. ¿Cómo lo interpretarías?



El coeficiente de regresión es la pendiente de la recta de regresión.

- 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.
- 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?



El coeficiente de determinación aparece en la ventana de resultados como  $R^2$ , y el coeficiente de correlación es su raíz cuadrada.

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



1. Seleccionar el menú **Teaching** » **Regresión** » **Predicciones**.
2. En el cuadro de diálogo que aparece seleccionar como modelo de regresión la recta calculada en el segundo apartado, introducir los valores para los que se desea la predicción en el campo Predicciones para y hacer clic sobre el botón Enviar.

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



Seguir los mismos pasos de los apartados anteriores, pero escogiendo como variable dependiente horas.estudio, y como independiente suspensos, y haciendo la predicción para o 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 entre el alcohol y el tiempo e interpretarlo. ¿Es bueno el modelo lineal?



1. Seleccionar el menú **Teaching** » **Regresión** » **Regresión lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable alcohol en el campo Variable dependiente y la variable tiempo en el campo Variable independiente, y hacer clic sobre el botón Enviar.

- (c) Dibujar la recta de regresión del alcohol sobre el 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?



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de Dispersión**.
2. En el cuadro de diálogo que aparece, seleccionar la variable alcohol en el campo Variable Y y la variable tiempo en el campo Variable X.
3. En la solapa **Línea de ajuste**, seleccionar Lineal y hacer clic en el botón Enviar.

Se observa que hay un residuo atípico para el punto que corresponde a los 210 minutos. Para eliminarlo: En la ventana de edición del conjunto de datos hacer clic con el botón derecho del ratón sobre la fila correspondiente al dato con el residuo atípico y seleccionar Borrar esta fila.

- (d) Si la concentración máxima de alcohol en la sangre que permite la ley para poder conducir es 0.3 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?



Para construir la recta de regresión:

1. Seleccionar el menú **Teaching** » **Regresión** » **Regresión lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable tiempo en el campo Variable dependiente y la variable alcohol en el campo Variable independiente.
3. Seleccionar Guardar el modelo, introducir un nombre para el modelo y hacer clic sobre el botón Enviar.

Para hacer la predicción:

1. Seleccionar el menú **Teaching** » **Regresión** » **Predicciones**.

2. En el cuadro de diálogo que aparece seleccionar como modelo de regresión la recta calculada e introducir los valores para los que se desea la predicción en el campo Predicciones para y hacer clic sobre el botón Enviar.

4. El conjunto de datos edad.estatura del paquete rk.Teaching contine la edad y la estatura de 30 personas. Se pide:

- (a) Cargar datos del conjunto de datos edad.estatura desde el paquete rk.Teaching.
- (b) Calcular la recta de regresión de la estatura sobre la edad. ¿Es un buen modelo la recta de regresión?



1. Seleccionar el menú **Teaching >> Regresión >> Regresión lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable estatura en el campo Variable dependiente y la variable edad en el campo Variable independiente, y hacer clic en el botón Enviar.

- (c) Dibujar el diagrama de dispersión de la estatura sobre la edad. ¿Alrededor de qué edad se observa un cambio en la tendencia?



1. Seleccionar el menú **Teaching >> Gráficos >> Diagrama de Dispersión**.
2. En el cuadro de diálogo que aparece, seleccionar la variable estatura en el campo Variable Y, la variable edad en el campo Variable X, y hacer clic en el botón Enviar.

- (d) Recodificar la variable edad en dos grupos para mayores y menores de 20 años.



1. Seleccionar el menú **Teaching >> Datos >> Recodificar variable**.
2. En el cuadro de diálogo que aparece seleccionar en el campo Variable a recodificar la variable edad.
3. En el campo Reglas de recodificación introducir
 

```
10:20 = "menores"
20:hi = "mayores"
```
4. En el cuadro Guardar nueva variable hacer clic sobre el botón Cambiar.
5. En el cuadro de diálogo que aparece seleccionar como objeto padre la el conjunto de datos edad\_estatura y hacer clic sobre el botón Aceptar.
6. Introducir el nombre de la nueva variable grupo.edad y hacer clic sobre el botón Enviar.

- (e) Calcular la recta de regresión de la estatura sobre la edad para cada grupo de edad. ¿En qué grupo explica mejor la recta de regresión la relación entre la estatura y la edad? Justificar la respuesta.



1. Seleccionar el menú **Teaching >> Regresión >> Regresión lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable estatura en el campo Variable dependiente y la variable edad como Variable independiente.
3. Seleccionar la opción Ajuste por grupos, introducir la variable grupo.edad en el campo Variable de agrupación, y hacer clic en el Enviar.

- (f) Dibujar las rectas de regresión anteriores.



1. Seleccionar el menú **Teaching >> Gráficos >> Diagrama de Dispersión**.
2. En el cuadro de diálogo que aparece, seleccionar la variable estatura en el campo Variable Y y la variable edad en el campo Variable X.
3. Seleccionar la opción Dibujar por grupos e introducir la variable grupo.edad en el campo Variable de agrupación.
4. En la solapa **Línea de ajuste**, seleccionar Lineal y hacer clic en el botón Enviar.



(g) ¿Qué estatura se espera que tenga una persona de 14 años? ¿Y una de 38?



Para predecir la estatura de la persona de 14 años:

1. Seleccionar el menú **Teaching** » **Regresión** » **Predicciones**.
2. En el cuadro de diálogo que aparece seleccionar como modelo de regresión la recta calculada para los menores e introducir 14 en el campo Predicciones para y hacer clic sobre el botón Enviar.

para predecir la estatura de la persona de 38 años, repetir lo mismo pero seleccionando la recta de regresión para los mayores e introducir 38 en el campo Predicciones para.

## 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 4 días. Se pide:
    - (a) Calcular la recta de regresión del tiempo de curación con respecto a la dosis suministrada.
    - (b) Calcular el coeficiente de regresión del tiempo de curación con respecto a la dosis e interpretarlo.
    - (c) Calcular el coeficiente de correlación lineal e interpretarlo.
    - (d) Determinar el tiempo esperado de curación para una dosis de 5 mg diarios. ¿Es fiable esta predicción?
    - (e) ¿Qué dosis debe aplicarse si queremos que el paciente tarde 4 días en curarse? ¿Es fiable la predicción?
  3. El fichero estaturas.pesos.alumnos del paquete rk.Teaching, contiene la estatura, el peso y el sexo de una muestra de alumnos universitarios. Se pide:
    - (a) Cargar el conjunto de datos estaturas.pesos.alumnos desde el paquete rk.Teaching.
    - (b) Calcular la recta de regresión del peso sobre la estatura y dibujarla.
    - (c) Calcular las rectas de regresión del peso sobre la estatura para cada sexo y dibujarlas.
    - (d) Calcular los coeficientes de determinación de ambas rectas. ¿Qué recta es mejor modelo? Justificar la respuesta.
    - (e) ¿Qué peso tendrá un hombre que mida 170 cm? ¿Y una mujer de la misma estatura?
  4. El conjunto de datos neonatos del paquete rk.Teaching, contiene información sobre una muestra de 320 recién nacidos en un hospital durante un año que cumplieron el tiempo normal de gestación. Se pide:
    - (a) Construir la tabla de frecuencias bidimensional del Agpar al minuto de nacer frente a si la madre ha fumado o no durante el embarazo. ¿Qué conclusiones se pueden sacar?

- (b) Construir la tabla de frecuencias bidimensional del peso de los recién nacidos frente a la edad de la madre. ¿Qué conclusiones se pueden sacar?
- (c) Construir la recta de regresión del peso de los recién nacidos sobre el número de cigarros fumados al día por las madres. ¿Existe una relación lineal fuerte entre el peso y el número de cigarros?
- (d) Dibujar la recta de regresión calculada en el apartado anterior. ¿Por qué la recta no se ajusta bien a la nube de puntos?
- (e) Calcular y dibujar la recta de regresión del peso de los recién nacidos sobre el número de cigarros fumados al día por las madres en el grupo de las madres que si fumaron durante el embarazo. ¿Es este modelo mejor o peor que la recta de los apartados anteriores?  
Según este modelo, ¿cuánto disminuirá el peso del recién nacido por cada cigarro más diario que fume la madre?
- (f) Según el modelo anterior, ¿qué peso tendrá un recién nacido de una madre que ha fumado 5 cigarros diarios durante el embarazo? ¿Y si la madre ha fumado 30 cigarros diarios durante el embarazo? ¿Son fiables estas predicciones?
- (g) ¿Existe la misma relación lineal entre el peso de los recién nacidos y el número de cigarros fumados al día por las madres que fumaron durante el embarazo en el grupo de las madres menores de 20 y en el grupo de las madres mayores de 20? ¿Qué se puede concluir?

# Regresión no lineal

## 1 Ejercicios resueltos

El procedimiento más sencillo para construir un modelo no lineal, siempre que sea posible, es transformar las variables para convertirlo en un modelo lineal. En el caso de los modelos de regresión simple más comunes las transformaciones que convierten cada modelo en un modelo lineal aparecen en la tabla siguiente:

Modelo	Modelo no lineal	Modelo lineal	Transformación
Potencial	$y = ax^b$	$\log(y) = \log(a) + b \log(x)$	Se toma el logaritmo de ambas variables
Exponencial	$y = e^{a+bx}$	$\log(y) = a + bx$	Se toma el logaritmo de la variable dependiente
Logarítmico	$y = a + b \log x$	$y = a + b \log x$	Se toma el logaritmo de la variable independiente
Inverso	$y = a + b/x$	$y = a + b \frac{1}{x}$	Se toma el inverso de la variable independiente
Curva S	$y = e^{a+b/x}$	$\log(y) = a + b \frac{1}{x}$	Se toma el logaritmo de la variable dependiente y el inverso de la independiente

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	1	2	3	4	5	6	7	8	9
Nº Bacterias	25	28	47	65	86	121	190	290	362

Se pide:

- (a) Crear un conjunto de datos con 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?



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de dispersión**.
2. En el cuadro de diálogo que aparece, seleccionar la variable bacterias en el campo Variable Y y la variable horas en el campo Variable X, y hacer clic en el botón Enviar.

- (c) Calcular los modelos exponencial y cuadrático de las bacterias sobre las horas. ¿Qué tipo de modelo es el mejor?



Para el modelo exponencial:

1. Seleccionar el menú **Teaching** » **Regresión** » **Regresión no lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable bacterias en el campo Variable dependiente y la variable horas en el campo Variable independiente.

3. En la solapa de Modelo de regresión seleccionar el modelo Exponencial.
4. Seleccionar Guardar modelo e introducir un nombre para el modelo y hacer clic sobre el botón Enviar.

Para el modelo cuadrático repetir los pasos pero seleccionando como modelo el Cuadrático. El modelo mejor será aquel que tenga un coeficiente de determinación mayor.

(d) Dibujar la curva del mejor de los modelos anteriores.



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de dispersión**.
2. En el cuadro de diálogo que aparece, seleccionar la variable bacterias en el campo Variable Y y la variable horas en el campo Variable X.
3. En la solapa **Línea de ajuste** seleccionar la opción Exponencial y hacer clic sobre el botón Enviar.

(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?



1. Seleccionar el menú **Teaching** » **Regresión** » **Predicciones**.
2. En el cuadro de diálogo que aparece seleccionar el modelo de regresión exponencial construido antes.
3. Introducir los valores 3.5, 10 en el campo Predicciones para y hacer clic sobre el botón Enviar.
4. Como se trata de un modelo exponencial, las predicciones obtenidas corresponden al logaritmo de bacterias. Para obtener la predicción de bacterias basta con aplicar la función exponencial a los valores obtenidos.

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



Para construir el modelo logarítmico:

1. Seleccionar el menú **Teaching** » **Regresión** » **Regresión no lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable horas en el campo Variable dependiente y la variable bacterias en el campo Variable independiente.
3. Seleccionar como modelo el Logarítmico.
4. Seleccionar Guardar modelo e introducir un nombre para el modelo y hacer clic sobre el botón Enviar.

Para hacer la predicción:

1. Seleccionar el menú **Teaching** » **Regresión** » **Predicciones**.
2. En el cuadro de diálogo que aparece seleccionar el modelo de regresión logarítmico construido antes.
3. Introducir el valor 100 en el campo Predicciones para y hacer clic sobre el botón Enviar.

2. El conjunto de datos dieta del paquete rk.Teaching contiene los datos de un estudio llevado a cabo por un centro dietético para probar una nueva dieta de adelgazamiento. Para cada individuo se ha medido el número de días que lleva con la dieta, el número de kilos perdidos desde entonces y si realizó o no un programa de ejercicios. Se pide:

- (a) Cargar el conjunto de datos dieta desde el paquete rk.Teaching.
- (b) 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?



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de dispersión**.

2. En el cuadro de diálogo que aparece, seleccionar la variable kilos en el campo Variable Y, la variable días en el campo Variable X, y hacer clic en el botón Enviar.

(c) Construir el modelo de regresión que mejor explique la relación entre los kilos perdidos y los días de dieta.



1. Seleccionar el menú **Teaching > Regresión > Comparación de modelos**.
2. En el cuadro de diálogo que aparece, seleccionar la variable kilos en el campo Variable dependiente y la variable días en el campo Variable independiente.
3. En la solapa **Modelos de regresión** seleccionar todos los modelos y hacer clic sobre el botón Enviar.
4. El mejor modelo aparece en primer lugar y es el que tenga el coeficiente de determinación mayor.

(d) Dibujar el modelo del apartado anterior.



1. Seleccionar el menú **Teaching > Gráficos > Diagrama de Dispersión**.
2. En el cuadro de diálogo que aparece, seleccionar la variable kilos en el campo Variable Y y la variable días en el campo Variable X.
3. En la solapa **Línea de ajuste** seleccionar la opción correspondiente al mejor modelo y hacer clic sobre el botón Enviar.

(e) Construir el modelo de regresión que mejor explique la relación entre los kilos perdidos y los días de dieta para los que no hacen ejercicio.



Para ver qué modelo es mejor:

1. Seleccionar el menú **Teaching > Regresión > Comparación de modelos**.
2. En el cuadro de diálogo que aparece, seleccionar la variable kilos en el campo Variable dependiente y la variable días en el campo Variable independiente.
3. Seleccionar la opción Filtro e introducir la condición `ejercicio=="no"` en el campo Condición de selección.
4. En la solapa **Modelos de regresión** seleccionar todos los modelos y hacer clic sobre el botón Enviar.
5. El mejor modelo aparece en primer lugar y es el que tenga el coeficiente de determinación mayor.

Para construir el modelo:

1. Seleccionar el menú **Teaching > Regresión > Regresión no lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable kilos en el campo Variable dependiente y la variable días en el campo Variable independiente.
3. Seleccionar la opción Filtro e introducir la condición `ejercicio=="no"` en el campo Condición de selección.
4. Seleccionar Guardar modelo e introducir un nombre para el modelo y hacer clic sobre el botón Enviar.

(f) Construir el modelo de regresión que mejor explique la relación entre los kilos perdidos y los días de dieta para los que si hacen ejercicio.



Para ver qué modelo es mejor:

1. Seleccionar el menú **Teaching > Regresión > Comparación de modelos**.
2. En el cuadro de diálogo que aparece, seleccionar la variable kilos en el campo Variable dependiente y la variable días en el campo Variable independiente.
3. Seleccionar la opción Filtro e introducir la condición `ejercicio=="si"` en el campo Condición de selección.
4. En la solapa **Modelos de regresión** seleccionar todos los modelos y hacer clic sobre el botón Enviar.

5. El mejor modelo aparece en primer lugar y es el que tenga el coeficiente de determinación mayor.

Para construir el modelo:

1. Seleccionar el menú **Teaching >> Regresión >> Regresión no lineal**.
2. En el cuadro de diálogo que aparece, seleccionar la variable kilos en el campo Variable dependiente y la variable días en el campo Variable independiente.
3. Seleccionar la opción Filtro e introducir la condición `ejercicio=="si"` en el campo Condición de selección.
4. Seleccionar Guardar modelo e introducir un nombre para el modelo y hacer clic sobre el botón Enviar.

- (g) Utilizar el modelo construido para predecir el número de kilos perdidos tras 40 y 500 días de dieta, tanto para los que hacen ejercicio como para los que no. ¿Son fiables estas predicciones?



1. Seleccionar el menú **Teaching >> Regresión >> Predicciones**.
2. En el cuadro de diálogo que aparece seleccionar el modelo de regresión construido antes para los que no hacen ejercicio.
3. Introducir los valores 40, 500 en el campo Predicciones para y hacer clic sobre el botón Enviar. Repetir los pasos anteriores seleccionando el modelo de regresión construido antes para los que si hacen ejercicio.

## 2 Ejercicios propuestos

1. 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?
2. El fichero `naciones.txt` contiene información sobre el desarrollo de distintos países (tasa de fertilidad, tasa de uso de anticonceptivos, tasa de mortalidad infantil, producto interior bruto per cápita y continente). Se pide:
  - (a) Importar el fichero `naciones.txt` en un conjunto de datos.
  - (b) Construir el mejor modelo de regresión de la tasa de fertilidad sobre el producto interior bruto. ¿Cómo explicarías esta relación?
  - (c) Dibujar el modelo del apartado anterior.
  - (d) ¿Qué tasa de fertilidad le corresponde a una mujer que viva en un país con un producto interior bruto per cápita de 10000 \$? ¿Y si la mujer vive en Europa?

# Probabilidad

## 1 Ejercicios resueltos

1. Construir los espacios probabilísticos correspondientes a los siguientes experimentos aleatorios:

(a) Sacar una carta de una baraja española.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Juegos de azar](#) » [Naipes](#) » [Espacio probabilístico](#).
2. En el cuadro de diálogo que aparece y hacer clic en el botón Enviar.

(b) Lanzar dos monedas.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Juegos de azar](#) » [Monedas](#) » [Espacio probabilístico](#).
2. En el cuadro de diálogo que aparece, introducir 2 en el campo Número de monedas y hacer clic en el botón Enviar.

(c) Lanzar dos dados.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Juegos de azar](#) » [Dados](#) » [Espacio probabilístico](#).
2. En el cuadro de diálogo que aparece, introducir 2 en el campo Número de dados y hacer clic en el botón Enviar.

(d) Lanzar dos dados y dos monedas.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Combinar espacios probabilísticos independientes](#).
2. En el cuadro de diálogo que aparece, seleccionar los conjuntos de datos correspondientes a los espacios muestrales del lanzamiento de dos monedas y del lanzamiento de dos dados generados en los apartados anteriores, y hacer clic en el botón Enviar.

2. Repetir el experimento de lanzar dos monedas 10 veces, 100 veces, 1000 veces y 1000000 de veces y calcular las frecuencias relativas de cada resultado. ¿Hacia dónde tienden las frecuencias? Construir el espacio probabilístico de este experimento y comprobar que se cumple la ley de los grandes números, es decir, que las frecuencias anteriores se aproximan a las probabilidades de cada suceso elemental.



Para la realización del experimento:

- (a) Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Juegos de azar](#) » [Monedas](#) » [Lanzamiento de monedas](#).
- (b) En el cuadro de diálogo que aparece, introducir 2 en el campo Número de monedas, introducir 10 en el campo Número de lanzamientos, activar la casilla Distribución de frecuencias y hacer clic en el botón Enviar.

Repetir los mismos pasos pero introduciendo 100, 1000 y 1000000 respectivamente en el campo Número de lanzamientos.

Para construir el espacio probabilístico correspondiente:

- Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Juegos de azar](#) » [Monedas](#) » [Espacio probabilístico](#).
- En el cuadro de diálogo que aparece, introducir 2 en el campo Número de monedas y hacer clic en el botón Enviar.

3. En una estantería hay tres cajas de un medicamento A, dos de un medicamento B y una de un medicamento C. Construir los espacios probabilísticos asociados a los siguientes experimentos aleatorios:

- Elegir tres medicamentos al azar sin reemplazamiento.



- Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Juegos de azar](#) » [Urna](#) » [Espacio probabilístico](#).
- En el cuadro de diálogo que aparece, seleccionar la opción Lista de objetos, introducir los objetos A,A,A,B,B,C en el campo Lista de objetos, introducir 3 en el campo Número de extracciones, y hacer clic en el botón Enviar.

- Elegir tres medicamentos al azar con reemplazamiento.



Repetir los mismos pasos del apartado anterior pero además activando la casilla Con reemplazamiento.

4. En una población se ha hecho un estudio epidemiológico sobre tres enfermedades asociadas habitualmente a la infancia, como son la varicela, el sarampión y la rubeola. Las frecuencias observadas aparecen en la siguiente tabla:

Varicela	Sarampión	Rubeola	Frecuencia
No	No	No	2654
No	No	Si	1436
No	Si	No	1682
No	Si	Si	668
Si	No	No	1747
Si	No	Si	476
Si	Si	No	876
Si	Si	Si	265

Se pide:

- Crear el conjunto de datos enfermedades.infantiles con las variables varicela, sarampión, rubeola y frecuencia e introducir datos de la población.
- Crear el espacio probabilístico asociado a la población.



- Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Construir espacio probabilístico](#).
- En el cuadro de diálogo que aparece seleccionar el conjunto de datos enfermedades.infantiles, activar la casilla Definir frecuencias, seleccionar la variable frecuencia en el campo Frecuencia, darle el nombre enfermedades.infantiles.ep al espacio probabilístico y hacer clic en el botón Enviar.

- Calcular la probabilidad de que una persona de esta población haya tenido varicela.



- Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
- En el cuadro de diálogo que aparece seleccionar el espacio probabilístico enfermedades.infantiles.ep, introducir `varicela == "Si"` en el campo Suceso y hacer clic en el botón Enviar.



(d) Calcular la probabilidad de que una persona de esta población haya tenido varicela o sarampión.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico enfermedades.infantiles.ep, introducir `varicela == "Si" | sarampion=="Si"` en el campo Suceso y hacer clic en el botón Enviar.

(e) Calcular la probabilidad de que una persona de esta población haya tenido sarampión y rubeola.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico enfermedades.infantiles.ep, introducir `sarampion == "Si" & rubeola=="Si"` en el campo Suceso y hacer clic en el botón Enviar.

(f) Calcular la probabilidad de que una persona de esta población haya tenido varicela si no ha tenido sarampion. ¿Son independientes el haber tenido varicela y el haber tenido sarampión?



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico enfermedades.infantiles.ep, introducir `varicela == "Si"` en el campo Suceso, activar la casilla Probabilidad condicionada, introducir `sarampion == "No"` en el campo Condición y hacer clic en el botón Enviar.

(g) Calcular la probabilidad de que una persona de esta población no haya tenido rubeola ni sarampión si ha tenido varicela.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico enfermedades.infantiles.ep, introducir `rubeola == "No" & sarampion=="No"` en el campo Suceso, activar la casilla Probabilidad condicionada, introducir `varicela == "Si"` en el campo Condición y hacer clic en el botón Enviar.

5. Se ha probado un test diagnóstico para detectar el embarazo en un grupo de mujeres en edad de procrear, obteniendo los siguientes resultados

Embarazo	Test	Frecuencia
No	—	3876
No	+	47
Si	—	12
Si	+	131

Se pide:

- (a) Crear el conjunto de datos test.embarazo con las variables embarazo, test, y frecuencia e introducir datos de la muestra.
- (b) Crear el espacio probabilístico asociado a la población.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Construir espacio probabilístico](#).
2. En el cuadro de diálogo que aparece seleccionar el conjunto de datos test.embarazo, activar la casilla Definir frecuencias, seleccionar la variable frecuencia en el campo Frecuencia, darle el nombre test.embarazo.ep al espacio probabilístico y hacer clic en el botón Enviar.

(c) Calcular la prevalencia del embarazo.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico test.embarazo.ep, introducir `embarazo == "Si"` en el campo Suceso y hacer clic en el botón Enviar.

(d) Calcular la probabilidad de que el test de positivo.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico test.embarazo.ep, introducir `test == "+"` en el campo Suceso y hacer clic en el botón Enviar.

(e) Calcular la sensibilidad del test.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico test.embarazo.ep, introducir `test=="+"` en el campo Suceso, activar la casilla Probabilidad condicionada, introducir `embarazo == "Si"` en el campo Condición y hacer clic en el botón Enviar.

(f) Calcular la especificidad del test.



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico test.embarazo.ep, introducir `test=="-"` en el campo Suceso, activar la casilla Probabilidad condicionada, introducir `embarazo == "No"` en el campo Condición y hacer clic en el botón Enviar.

(g) Calcular el valor predictivo positivo del test. ¿Es útil el test para detectar el embarazo?



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico test.embarazo.ep, introducir `embarazo=="Si"` en el campo Suceso, activar la casilla Probabilidad condicionada, introducir `test=="+"` en el campo Condición y hacer clic en el botón Enviar.

(h) Calcular el valor predictivo negativo del test. ¿Es útil el test para descartar el embarazo?



1. Seleccionar el menú [Teaching](#) » [Probabilidad](#) » [Calcular probabilidad](#).
2. En el cuadro de diálogo que aparece seleccionar el espacio probabilístico test.embarazo.ep, introducir `embarazo=="No"` en el campo Suceso, activar la casilla Probabilidad condicionada, introducir `test=="-"` en el campo Condición y hacer clic en el botón Enviar.

## 2 Ejercicios propuestos

1. Construir el espacio muestral correspondiente a tirar una moneda, un dado y sacar una carta de una baraja española.
2. Para comprobar la eficacia de una vacuna contra la gripe se tomó una muestra de 1000 y se observó si fueron vacunadas y si finalmente tuvieron gripe o no. Los resultados obtenidos fueron los siguientes

Vacuna	Gripe	Frecuencia
No	No	418
No	Si	312
Si	No	233
Si	Si	37

Se pide:

- (a) Construir el espacio probabilístico asociado al experimento.
  - (b) Calcular la probabilidad de haberse vacunado contra la gripe.
  - (c) Calcular la prevalencia de la gripe.
  - (d) Calcular la probabilidad de desarrollar la gripe tras haberse vacunado. ¿Es efectiva la vacuna?
3. Para probar la eficacia de un test diagnóstico para detectar el ébola en un país centroafricano, se tomó una muestra de personas a las que se le ha aplicado el test. El test dió positivo en 147 personas con ébola, pero también dió positivo en 28 personas sin ébola. Por otro lado el test dió negativo en 97465 personas sin ébola, pero también dió negativo en 65 personas con ébola. Se pide:
- (a) Construir el espacio probabilístico asociado al test diagnóstico.
  - (b) Calcular la prevalencia del ébola en ese país.
  - (c) Calcular la probabilidad de que el test de negativo.
  - (d) Calcular la sensibilidad y la especificidad del test.
  - (e) ¿Para qué es más efectivo el test, para detectar o para descartar el ébola?



# Variables Aleatorias Discretas

## 1 Ejercicios resueltos

1. Sea  $X$  la variable que mide el número de caras obtenidas al lanzar 10 monedas. Se pide:

(a) Generar la distribución de probabilidad de una variable Binomial  $B(10, 0.5)$ .



1. Seleccionar el menú [teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir 0,1,2,3,4,5,6,7,8,9,10 en el campo Valores de la variable, introducir 10 en el campo Número de repeticiones, 0.5 en el campo Probabilidad de éxito, y hacer clic en el botón Enviar.

(b) Dibujar la gráfica de la función de probabilidad de la Binomial  $X \sim B(10, 0.5)$ .



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir 10 en el campo Número de repeticiones, 0.5 en el campo Probabilidad de éxito y hacer clic en el botón Enviar.

(c) Dibujar la gráfica de la función de distribución.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir 10 en el campo Número de repeticiones, 0.5 en el campo Probabilidad de éxito, seleccionar la opción Función de distribución y hacer clic en el botón Enviar.

(d) Calcular  $P(X = 7)$ .



1. Seleccionar el menú [teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir 7 en el campo Valores de la variable, introducir 10 en el campo Número de repeticiones, 0.5 en el campo Probabilidad de éxito, y hacer clic en el botón Enviar.

(e) Calcular  $P(X \leq 4)$ .



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir 4 en el campo Valores de la variable, 10 en el campo Número de repeticiones, 0.5 en el campo Probabilidad de éxito.
3. Seleccionar la opción Probabilidades acumuladas y hacer clic en el botón Enviar.

(f) Calcular  $P(X > 5)$ .



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir 5 en el campo Valores de la variable, 10 en el campo Número de repeticiones, 0.5 en el campo Probabilidad de éxito.
3. Seleccionar la opción Probabilidades acumuladas, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

(g) Calcular  $P(2 \leq X < 9)$ .



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir los valores 1, 8 en el campo Valores de la variable, 10 en el campo Número de repeticiones, 0.5 en el campo Probabilidad de éxito.
3. Seleccionar la opción Probabilidades acumuladas y hacer clic en el botón Enviar.

La probabilidad del intervalo  $P(2 \leq X < 9)$  es la resta de las probabilidades obtenidas  $P(X < 9) = P(X \leq 8)$  y  $P(X < 2) = P(X \leq 1)$ .

2. El número de nacimientos diarios en una determinada población sigue una distribución de Poisson de media 6 nacimientos al día. Se pide:

(a) Dibujar la gráfica de la función de probabilidad.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Poisson](#) » [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir el valor 6 en el campo Media y hacer clic en el botón Enviar.

(b) Dibujar la gráfica de la función de distribución.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Poisson](#) » [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir el valor 6 en el campo Media, marcar la opción Función de distribución y hacer clic en el botón Enviar.

(c) Calcular la probabilidad de un día haya 1 nacimiento.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Poisson](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir el valor 1 en el campo Valores de la variable, introducir el valor 6 en el campo Media, y hacer clic en el botón Enviar.

(d) Calcular la probabilidad de que un día haya menos de 6 nacimientos.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Poisson](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir 5 en el campo Valores de la variable y 6 en el campo Media.
3. Seleccionar la opción Probabilidades acumuladas y hacer clic en el botón Enviar.

(e) Calcular la probabilidad de que un día haya 4 o más nacimientos.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Poisson](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir 3 en el campo Valores de la variable y 6 en el campo Media.
3. Seleccionar la opción Probabilidades acumuladas, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

(f) Calcular la probabilidad de que un día haya entre 4 y 8 nacimientos, inclusivos.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Poisson](#) » [Probabilidades](#).
  2. En el cuadro de diálogo que aparece, introducir los valores 3, 8 en el campo Valores de la variable y 6 en el campo Media.
  3. Seleccionar la opción Probabilidades acumuladas y hacer clic en el botón Enviar.
- La probabilidad del intervalo  $P(4 \leq X \leq 8)$  es la resta de las probabilidades obtenidas  $P(X \leq 8)$  y  $P(X < 4) = P(X \leq 3)$ .

3. La ley de los casos raros dice que en una distribución Binomial  $B(n, p)$ , cuando  $n \geq 30$  y  $p \leq 0.1$  la distribución se parece mucho a una distribución Poisson  $P(np)$ . Para comprobar hasta qué punto se parecen estas distribuciones, se pide:

- (a) Generar la distribución de probabilidad de una variable Binomial  $B(30, 0.1)$ .



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir los valores 0,1,2,3,4,5,6,7,8,9,10 en el campo Valores de la variable, introducir el valor 30 en el campo Número de repeticiones, 0.1 en el campo Probabilidad de éxito y hacer clic en el botón Enviar.

- (b) Generar la distribución de probabilidad de una variable Poisson  $P(3)$  y compararla con la de la binomial  $B(30, 0.1)$ .



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Poisson](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir los valores 0,1,2,3,4,5,6,7,8,9,10 en el campo Valores de la variable, introducir el valor 3 en el campo Media y hacer clic en el botón Enviar.

- (c) Generar la distribución de probabilidad de una variable Binomial  $B(100, 0.03)$  y compararla con la de la Poisson  $P(3)$ . ¿Se parecen más estas distribuciones que las anteriores?



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir los valores 0,1,2,3,4,5,6,7,8,9,10 en el campo Valores de la variable, introducir el valor 100 en el campo Número de repeticiones, 0.03 en el campo Probabilidad de éxito y hacer clic en el botón Enviar.

- (d) Dibujar las gráficas de las distribuciones anteriores y ver cuáles se parecen más. ¿Se cumple la ley de los casos raros?



1. Seleccionar el menú [Teaching](#) » [Simulaciones](#) » [Ley de los casos raros](#).
2. En el cuadro de diálogo que aparece, cambiar el valor de n hasta 30 y el de p hasta 0.1.
3. Después cambiar el valor de n hasta 100 y el de p hasta 0.03.

## 2 Ejercicios propuestos

1. Al lanzar 100 veces una moneda, ¿cuál es la probabilidad de obtener entre 40 y 60 caras inclusive?
2. 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.
  - (b) Se curen al menos 4.
3. 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?

4. El número medio de llamadas por minuto que llegan a una centralita telefónica es igual a 120. Se pide:
  - (a) Dar la distribución de probabilidad del número de llamadas en 2 segundos y dibujar su gráfica.
  - (b) Calcular la probabilidad de que durante 2 segundos lleguen a la centralita menos de 4 llamadas.
  - (c) Calcular la probabilidad de que durante 3 segundos lleguen a la centralita 3 llamadas como mínimo.
5. Se sabe que la probabilidad de que aparezca una bacteria en un  $\text{mm}^3$  de cierta disolución es de 0.002. Si en cada  $\text{mm}^3$  a lo sumo puede aparecer una bacteria, determinar la probabilidad de que en un  $\text{cm}^3$  haya como máximo 5 bacterias.



# Variables Aleatorias Continuas

## 1 Ejercicios resueltos

1. Supongase que un autobús pasa por una parada cada 15 minutos y que una persona puede llegar a la parada en cualquier instante, entonces la variable que mide el tiempo que la persona espera al autobús es una variable Uniforme continua  $U(0, 15)$ , ya que cualquier valor entre 0 y 15 minutos es equiprobable. Se pide:

(a) Dibujar la gráfica de la función de densidad de la Uniforme  $X \sim U(0, 15)$ .



1. Seleccionar el menú Teaching » Distribuciones » Continuas » Uniforme » Gráfico de probabilidad.
2. En el cuadro de diálogo que aparece, introducir el valor 0 en el campo Mínimo, 15 en el campo Máximo y hacer clic en el botón Enviar.

(b) Dibujar la gráfica de la función de distribución.



1. Seleccionar el menú Teaching » Distribuciones » Continuas » Uniforme » Gráfico de probabilidad.
2. En el cuadro de diálogo que aparece, introducir el valor 0 en el campo Mínimo, 15 en el campo Máximo, marcar la opción Función de distribución y hacer clic en el botón Enviar.

(c) Calcular la probabilidad de esperar al autobús menos de 5 minutos.



1. Seleccionar el menú Teaching » Distribuciones » Continuas » Uniforme » Probabilidades.
2. En el cuadro de diálogo que aparece, introducir el valor 5 en el campo Valores de la variable, 0 en el campo Mínimo, 15 en el campo Máximo y hacer clic en el botón Enviar.

(d) Calcular la probabilidad de esperar al autobús más de 12 minutos.



1. Seleccionar el menú Teaching » Distribuciones » Continuas » Uniforme » Probabilidades.
2. En el cuadro de diálogo que aparece, introducir el valor 12 en el campo Valores de la variable, 0 en el campo Mínimo, 15 en el campo Máximo, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

(e) Calcular la probabilidad de esperar al autobús entre 5 y 10 minutos.



1. Seleccionar el menú Teaching » Distribuciones » Continuas » Uniforme » Probabilidades.
2. En el cuadro de diálogo que aparece, introducir los valores 10, 5 en el campo Valores de la variable, 0 en el campo Mínimo, 15 en el campo Máximo y hacer clic en el botón Enviar.

La probabilidad del intervalo  $P(5 \leq X \leq 10)$  es la resta de las probabilidades obtenidas  $P(X \leq 10) - P(X \leq 5)$ .

(f) ¿Por debajo de qué tiempo esperará al autobús la mitad de las veces?



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Uniforme](#) » [Cuantiles](#).
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.5 en el campo Probabilidades acumuladas, 0 en el campo Mínimo, 15 en el campo Máximo y hacer clic en el botón Enviar.

(g) ¿Por encima de qué tiempo esperará al autobús el 10% de las veces?



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Uniforme](#) » [Cuantiles](#).
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.1 en el campo Probabilidades acumuladas, 0 en el campo Mínimo, 15 en el campo Máximo, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

2. La variable aleatoria normal de media 0 y desviación típica 1,  $Z \sim N(0, 1)$ , se conoce como normal estándar y es la variable normal más importante. Se pide:

(a) Dibujar la gráfica de la función de densidad.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Normal](#) » [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir el valor 0 en el campo Media, 1 en el campo Desviación típica y hacer clic en el botón Enviar.

(b) ¿Cómo afectan los dos parámetros de la normal, su media y su desviación típica, a la forma de la campana de Gauss?



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Normal](#) » [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, seleccionar la opción Previsualizar.
3. Incrementar el valor de la media y ver cómo cambia la forma de la campana.
4. Después disminuir el valor de la desviación típica y ver cómo cambia la forma de la campana.

(c) Dibujar la gráfica de la función de distribución.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Normal](#) » [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir el valor 0 en el campo Media, 1 en el campo Desviación típica, marcar la opción Función de distribución y hacer clic en el botón Enviar.

(d) Calcular la probabilidad de que la normal estándar tome un valor menor que  $-1$ .



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Normal](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir el valor  $-1$  en el campo Valores de la variable, 0 en el campo Media, 1 en el campo Desviación típica, y hacer clic en el botón Enviar.

(e) Calcular la probabilidad de que la normal estándar tome un valor mayor que 1.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Normal](#) » [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir el valor 1 en el campo Valores de la variable, 0 en el campo Media, 1 en el campo Desviación típica, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

(f) Calcular la probabilidad de que la normal estándar tome un valor entre  $-1$  (la media menos la desviación típica) y 1 (la media más la desviación típica).



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **Normal** » **Probabilidades**.
  2. En el cuadro de diálogo que aparece, introducir los valores 1, -1 en el campo Valores de la variable, 0 en el campo Media, 1 en el campo Desviación típica y hacer clic en el botón Enviar.
- La probabilidad del intervalo  $P(-1 \leq Z \leq 1)$  es la resta de las probabilidades obtenidas  $P(Z \leq 1) - P(Z \leq -1)$ .

- (g) Calcular la probabilidad de que la normal estándar tome un valor entre -2 (la media menos dos veces la desviación típica) y 2 (la media más dos veces la desviación típica).



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **Normal** » **Probabilidades**.
  2. En el cuadro de diálogo que aparece, introducir los valores 2, -2 en el campo Valores de la variable, 0 en el campo Media, 1 en el campo Desviación típica y hacer clic en el botón Enviar.
- La probabilidad del intervalo  $P(-2 \leq Z \leq 2)$  es la resta de las probabilidades obtenidas  $P(Z \leq 2) - P(Z \leq -2)$ .

- (h) Calcular la probabilidad de que la normal estándar tome un valor entre -3 (la media menos tres veces la desviación típica) y 3 (la media más tres veces la desviación típica).



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **Normal** » **Probabilidades**.
  2. En el cuadro de diálogo que aparece, introducir los valores 3, -3 en el campo Valores de la variable, 0 en el campo Media, 1 en el campo Desviación típica y hacer clic en el botón Enviar.
- La probabilidad del intervalo  $P(-3 \leq Z \leq 3)$  es la resta de las probabilidades obtenidas  $P(Z \leq 3) - P(Z \leq -3)$ .

- (i) Calcular los cuartiles.



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **Normal** » **Cuantiles**.
2. En el cuadro de diálogo que aparece, introducir las probabilidades 0.25, 0.5, 0.75 en el campo Probabilidades acumuladas, 0 en el campo Media, 1 en el campo Desviación típica y hacer clic en el botón Enviar.

- (j) Calcular el valor que deja acumulada por debajo una probabilidad 0.95.



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **Normal** » **Cuantiles**.
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.95 en el campo Probabilidades acumuladas, 0 en el campo Media, 1 en el campo Desviación típica y hacer clic en el botón Enviar.

- (k) Calcular el valor que deja acumulada por encima una probabilidad 0.025.



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **Normal** » **Cuantiles**.
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.025 en el campo Probabilidades acumuladas, 0 en el campo Media, 1 en el campo Desviación típica, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

3. El teorema central del límite establece que la variable resultante de sumar 30 o más variables independientes sigue una distribución normal de media la suma de las medias de cada una de las variables y de varianza la suma de sus varianzas. Esta es la explicación de que una gran parte de las variables continuas que aparecen en la naturaleza sean variables normales. Para observar de manera experimental el teorema central del límite se realiza un experimento que consiste en lanzar varios dados muchas veces y sumar los valores obtenidos. Se pide:

- (a) Simular el lanzamiento de un dado 100000 veces y dibujar el diagrama de barras asociado. ¿Tiene forma de campana de Gauss?



Para generar los lanzamientos del dado:

1. Seleccionar el menú [Teaching](#) [Simulaciones](#) [Lanzamiento de dados](#).
2. En el cuadro de diálogo que aparece, introducir 1 en el campo Número de dados, introducir 100000 en el campo Número de lanzamientos, seleccionar la opción Incluir suma, introducir un nombre para el conjunto de datos y hacer clic en el botón Enviar.

Para dibujar el diagrama de barras:

1. Seleccionar el menú [Teaching](#) [Gráficos](#) [Diagrama de barras](#).
2. En el cuadro de diálogo que aparece seleccionar la variable sum.
3. En la solapa [Opciones de las barras](#), seleccionar la opción Frecuencias relativas y hacer clic en el botón Enviar.

(b) Repetir el apartado anterior con 2 y 30 dados. ¿Se cumple el teorema central del límite?

4. La suma de  $n$  variables normales estándar independientes elevadas al cuadrado es una variable con distribución Chi-cuadrado con  $n$  grados de libertad  $\chi^2(n)$ . Sea  $X$  una variable Chi-cuadrado con 6 grados de libertad  $\chi^2(6)$ . Se pide:

(a) Dibujar la gráfica de la función de densidad.



1. Seleccionar el menú [Teaching](#) [Distribuciones](#) [Continuas](#) [Chi-cuadrado](#) [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir el valor 6 en el Grados de libertad y hacer clic en el botón Enviar.

(b) Calcular la probabilidad de que la variable tome un valor menor que 6.



1. Seleccionar el menú [Teaching](#) [Distribuciones](#) [Continuas](#) [Chi-cuadrado](#) [Probabilidades](#).
2. En el cuadro de diálogo que aparece, introducir el valor 6 en el campo Valores de la variable, 6 en el campo Grados de libertad y hacer clic en el botón Enviar.

(c) Calcular el valor que deja acumulada por debajo una probabilidad 0.05.



1. Seleccionar el menú [Teaching](#) [Distribuciones](#) [Continuas](#) [Chi-cuadrado](#) [Cuantiles](#).
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.05 en el campo Probabilidades acumuladas, 6 en el campo Grados de libertad y hacer clic en el botón Enviar.

(d) Calcular el valor que deja acumulada por arriba una probabilidad 0.1.



1. Seleccionar el menú [Teaching](#) [Distribuciones](#) [Continuas](#) [Chi-cuadrado](#) [Cuantiles](#).
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.1 en el campo Probabilidades, 6 en el campo Grados de libertad, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

5. La variable que se obtiene al dividir una normal estándar entre la raíz de una variable Chi-cuadrado de  $n$  grados de libertad dividida por  $n$ , sigue una distribución  $t$  de student de  $n$  grados de libertad  $T(n)$ . Sea  $X$  una variable  $t$  de student de 8 grados de libertad  $T(8)$ . Se pide:

(a) Dibujar la gráfica de la función de probabilidad y compararla con la de la normal estándar.



1. Seleccionar el menú [Teaching](#) [Distribuciones](#) [Continuas](#) [T de student](#) [Gráfico de probabilidad](#).
2. En el cuadro de diálogo que aparece, introducir el valor 8 en el campo Grados de libertad y hacer clic en el botón Enviar.

(b) Calcular el percentil octavo.



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **T de student** » **Cuantiles**.
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.08 en el campo Probabilidades acumuladas, 8 en el campo Grados de libertad y hacer clic en el botón Enviar.

(c) Calcular el valor por encima del cual está el 5% de la población.



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **T de student** » **Cuantiles**.
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.05 en el campo Probabilidades, 8 en el campo Grados de libertad, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

6. La variable resultante de dividir una variable Chi-cuadrado de  $n$  grados de libertad dividida por  $n$ , entre una variable Chi-cuadrado de  $m$  grados de libertad dividida por  $m$ , sigue un modelo de distribución  $F$  de Fisher de  $n$  y  $m$  grados de libertad  $F(n, m)$ . Sea  $X$  una variable  $F$  de Fisher de 10 y 20 grados de libertad  $F(10, 20)$ . Se pide:

(a) Dibujar la gráfica de la función de densidad



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **F de Fisher** » **Gráfico de probabilidad**.
2. En el cuadro de diálogo que aparece, introducir 10 en el campo Grados de libertad del numerador, introducir 20 en el campo Grados de libertad del denominador y hacer clic en el botón Enviar.

(b) Calcular la probabilidad acumulada por encima de 1.



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **F de Fisher** » **Probabilidades**.
2. En el cuadro de diálogo que aparece, introducir el valor 1 en el campo Valores de la variable, 10 en el campo Grados de libertad del numerador, 20 en el campo Grados de libertad del denominador, seleccionar la opción derecha en el campo cola de acumulación y hacer clic en el botón Enviar.

(c) Calcular el rango intercuartílico.



1. Seleccionar el menú **Teaching** » **Distribuciones** » **Continuas** » **F de Fisher** » **Cuantiles**.
2. En el cuadro de diálogo que aparece, introducir las probabilidades 0.75, 0.25 en el campo Probabilidades, 10 en el campo Grados de libertad del numerador, 20 en el campo Grados de libertad del denominador y hacer clic en el botón Enviar.

El rango intercuartílico es la resta de los valores obtenidos correspondientes al tercer y primer cuartiles.

## 2 Ejercicios propuestos

1. 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.

2. Se sabe que el nivel de colesterol en varones de más de 30 años de una determinada población sigue una distribución normal, de media 220mg/dl y desviación típica 30mg/dl. Si la población tiene 20000 varones mayores de 30 años,
  - (a) ¿Cuántos se espera que tengan su nivel de colesterol entre 210mg/dl y 240mg/dl?
  - (b) ¿Cuántos se espera que tengan su nivel de colesterol por encima de 250mg/dl?
  - (c) ¿Cuál será el nivel de colesterol por encima del cual se espera que esté el 20% de la población?
3. Calcular la probabilidad de obtener entre 40 y 60 caras, inclusive, al lanzar 100 veces una moneda. Utilizar la aproximación de la distribución binomial mediante una normal.

# Intervalos de Confianza para Medias y Proporciones

## 1 Ejercicios resueltos

- 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  $\text{mg}/\text{mm}^3$ :

$$17.6 - 19.2 - 21.3 - 15.1 - 17.6 - 18.9 - 16.2 - 18.3 - 19.0 - 16.4$$

Se pide:

- Crear un conjunto de datos con la variable concentracion.
- Calcular el intervalo de confianza para la media de la concentración del lote con nivel de confianza del 95% (nivel de significación  $\alpha = 0.05$ ).



- Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para una muestra**.
- En el cuadro de diálogo que aparece seleccionar la variable concentracion en el campo Variable y hacer clic sobre el botón Enviar.

- Calcular los intervalos de confianza para la media con niveles del 90% y del 99% (niveles de significación  $\alpha = 0.1$  y  $\alpha = 0.01$ ).



Repetir los mismos pasos del apartado anterior, cambiando el nivel de confianza para cada intervalo en la solapa Opciones de contraste

- 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?
- ¿Qué tamaño muestral sería necesario para obtener una estimación del contenido medio de principio activo con un margen de error de  $\pm 0.5 \text{ mg}/\text{mm}^3$  y una confianza del 95%?



- Seleccionar el menú **Teaching** » **Estadística descriptiva** » **Estadísticos**.
- En el cuadro de diálogo que aparece seleccionar la variable concentracion en el campo Variable.
- En la solapa **Estadísticos básicos** marcar el estadístico Cuasidesviación típica y hacer clic en el botón Enviar.
- Seleccionar el menú **Teaching** » **Test paramétricos** » **Cálculo del tamaño muestral para la media**.
- En el cuadro de diálogo que aparece introducir la cuasidesviación típica muestral en el campo Desviación típica, el nivel de significación deseado, en este caso 0.05, en el campo Nivel de significación, el margen de error deseado, en este caso 0.5, en el campo Error, y hacer clic en el botón Enviar.

- Si, para que sea efectivo, el fármaco debe tener una concentración mínima de  $16 \text{ mg}/\text{mm}^3$  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 un conjunto de datos con las variables grasa y granja.  
 (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.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para una muestra**.
2. En el cuadro de diálogo que aparece seleccionar la variable grasa en el campo Variable y hacer clic sobre el botón Enviar.

- (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.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para una muestra**.
2. En el cuadro de diálogo que aparece seleccionar la variable grasa en el campo Variable.
3. Seleccionar la casilla de Filtro e introducir la condición `granja==`X`` hacer clic sobre el botón Enviar.
4. Repetir los mismos pasos para el intervalo de confianza de la granja Y, introduciendo la condición `granja==`Y`` en el campo Condición de selección.

- (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 el alumnado utiliza habitualmente (al menos una vez a la semana) la biblioteca de la misma, se han obtenido los siguientes resultados:

Alumno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Respuesta	no	si	no	no	no	si	no	si	si	si	si	no	si	no	si	no	no

Alumno	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Respuesta	no	si	si	si	no	no	si	no	no	si	si	no	no	si	no	si	no

- (a) Crear un conjunto de datos con la variable respuesta.  
 (b) Calcular el intervalo de confianza con  $\alpha = 0.01$  para la proporción del alumnado que utiliza habitualmente la biblioteca.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Proporciones** » **Test de proporciones para una muestra**.
2. En el cuadro de diálogo que aparece seleccionar la variable respuesta en el campo Variable e introducir si en el campo Categoría.
3. En la solapa **Opciones de contraste** introducir 0.99 en el campo Nivel de confianza y hacer clic en el botón Enviar.

- (c) ¿Qué interpretación tiene dicho intervalo? ¿Cómo es su precisión?



- (d) ¿Qué tamaño muestral sería necesario para obtener una estimación del porcentaje de alumnos que utilizan regularmente la biblioteca con un margen de error de un 1% y una confianza del 95%?



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Proporciones** » **Cálculo del tamaño muestral para una proporción**.
2. En el cuadro de diálogo que aparece introducir la proporción muestral en el campo **p**, el nivel de significación deseado, en este caso 0.05, en el campo **Nivel de significación**, el margen de error deseado, en este caso 0.01, en el campo **Error**, y hacer clic en el botón **Enviar**.

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



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Proporciones** » **Test para una proporción**.
2. En el cuadro de diálogo que aparece marcar la opción **Introducción manual de frecuencias**, introducir 154 en el campo **Frecuencia muestral**, introducir 200 en el campo **Tamaño muestral** y hacer clic en el botón **Enviar**.

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

## 2 Ejercicios propuestos

1. Para determinar el nivel medio de colesterol (en mg/dl) 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. ¿Se puede afirmar que el nivel de colesterol medio de la población está por debajo de 210 mg/dl?

2. 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. ¿Qué intervalo es más preciso?
3. 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 en 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.
4. El conjunto de datos neonatos del paquete **rk**. **Teaching**, contiene información sobre una muestra de 320 recién nacidos en un hospital durante un año que cumplieron el tiempo normal de gestación. Se pide:
  - (a) Calcular el intervalo de confianza del 99% para el peso medio de los recién nacidos. ¿Entre qué valores estará el peso medio?
  - (b) Calcular el intervalo de confianza para la puntuación media del Apgar al minuto de nacer y compararlo con el de la puntuación Apgar a los 5 minutos. ¿Existen diferencias significativas entre las medias de ambas puntuaciones?
  - (c) Calcular el intervalo de confianza para el porcentaje de niños con peso menor o igual que 2.5 Kg en el grupo de las madres que han fumado durante el embarazo y en el de las que no.



# Intervalos de Confianza para la Comparación de 2 Poblaciones

## 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 un conjunto de datos con las variables antes y después.  
 (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?



1. Seleccionar el menú **Teaching** » **Estadística descriptiva** » **Estadísticos**.
2. En el cuadro de diálogo que aparece seleccionar las variables antes y después en el campo Variables.
3. En la solapa **Estadísticos básicos** activar la casilla de selección para la Media y la Desviación típica y hacer clic en el botón Enviar.

- (c) Obtener los intervalos de confianza para la media de la diferencia entre ambas variables con niveles de significación 0.05 y 0.01.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para dos muestras pareadas**.
2. En el cuadro de diálogo que aparece seleccionar la variable antes en el campo Comparar, la variable después en el campo Con.
3. En la solapa **Opciones de contraste** introducir 0.95 en el campo Nivel de confianza y hacer clic en el botón Enviar.
4. Repetir los pasos para el intervalo de confianza con nivel de significación 0.01 poniendo 0.99 en el campo Nivel de confianza.

- (d) ¿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 después 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.



1. En la ventana de edición de datos, cambiar los datos de las dos últimas farmacias y cerrar la ventana.
2. Repetir los pasos del apartado anterior.

Existen diferencias entre las medias con el nivel de confianza fijado siempre que el intervalo resultante no contenga el valor 0.

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 un conjunto de datos con las variables grasa y granja.  
 (b) Calcular el intervalo de confianza para el cociente de varianzas del contenido de materia grasa de la leche procedente de ambas granjas.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Varianzas** » **Test F de Fisher**.
2. En el cuadro de dialogo que aparece seleccionar la variable grasa en el campo Comparar y seleccionar la variable granja en el campo Según.
3. En la solapa **Opciones de contraste** introducir 0.95 en el campo Nivel de confianza y hacer clic sobre el botón Enviar.

Se mantiene la hipótesis de igualdad de varianzas con la confianza fijada si el intervalo resultante contiene el valor 1.

- (c) 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.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para muestras independientes**.
2. En el cuadro de dialogo que aparece seleccionar la variable grasa en el campo Comparar y seleccionar la variable granja en el campo Según.
3. En la solapa **Opciones de contraste** introducir 0.95 en el campo Nivel de confianza, marcar la opción Si en el campo Suponer varianzas iguales y hacer clic sobre el botón Enviar.

- (d) A la vista del intervalo obtenido 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.



Existen diferencias entre las medias con el nivel de confianza fijado siempre que el intervalo resultante no contenga el valor 0.

3. En una encuesta realizada en una facultad, sobre si el alumnado utiliza habitualmente (al menos una vez a la semana) la biblioteca de la misma, se han obtenido los siguientes resultados:

Alumno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Respuesta	no	si	no	no	no	si	no	si	si	si	si	no	si	no	si	no	no
Sexo	H	M	M	H	H	H	M	M	M	M	H	H	M	H	M	H	H

Alumno	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Respuesta	no	si	si	si	no	no	si	no	no	si	si	no	no	si	no	si	no
Sexo	M	H	M	M	M	H	M	H	H	M	M	H	H	M	M	M	H

- (a) Crear un conjunto de datos con las variables respuesta y sexo.
- (b) ¿Existen diferencias significativas entre las proporciones de chicos y chicas que usan habitualmente la biblioteca? Justificar la respuesta.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Proporciones** » **Test para la comparación de dos proporciones**.
2. En el cuadro de dialogo que aparece seleccionar la variable respuesta en el campo Comparar, seleccionar la variable sexo en el campo Según, introducir el valor si al campo Categoría y hacer clic sobre el botón Enviar.

Hay diferencias entre las proporciones con el nivel de confianza fijado si el intervalo resultante no contiene el valor 0.

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. ¿Existen diferencias significativas en el porcentaje de aprobados en ambos grupos? ¿Pueden ser debidas al turno horario? Justificar la respuesta.



- (a) Seleccionar el menú **Teaching** » **Test paramétricos** » **Proporciones** » **Test para la comparación de dos proporciones**.
- (b) En el cuadro de diálogo que aparece seleccionar la opción Introducción manual de frecuencias, introducir 55 en el campo Frecuencia muestral 1, introducir 80 en el campo Tamaño muestral 1, introducir 32 en el campo Frecuencia muestral 2, introducir 90 en el campo Tamaño muestral 2 y hacer clic en el botón Enviar.

## 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 (en mg/dl) antes y después de desarrollar un programa de ejercicios. Los resultados obtenidos fueron los siguientes:

Nivel Previo	182	232	191	200	148	249	276	213	241	280	262
Nivel Posterior	198	210	194	220	138	220	219	161	210	213	226

- (a) Hallar el intervalo de confianza del 95% para la diferencia del nivel medio de colesterol antes y después del ejercicio.
- (b) Hallar el intervalo de confianza del 99% para la diferencia del nivel medio de colesterol antes y después del ejercicio.
- (c) A la vista de los intervalos anteriores, ¿se concluye que el ejercicio físico disminuye el nivel de colesterol?
2. En una encuesta realizada 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 180 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 para un nivel de significación  $\alpha = 0.01$  de que el trato recibido en un hospital es mejor que en el otro?
3. El conjunto de datos neonatos del paquete rk.Teaching, contiene información sobre una muestra de 320 recién nacidos en un hospital durante un año que cumplieron el tiempo normal de gestación. Se pide:
  - (a) Calcular el intervalo de confianza para la comparación de los pesos medios de los recién nacidos en el grupo de las madres fumadoras y en el de las no fumadoras durante el embarazo. ¿Influye el hecho de fumar en el peso del recién nacido?

- (b) Considerando sólo la muestra de niños correspondientes a madres que no han fumado durante el embarazo, calcular el intervalo de confianza para la comparación de los pesos medios de los recién nacidos en el grupo de las madres que no fumaban antes del embarazo y en el de las que si fumaban. ¿Influye el hecho de haber fumado antes del embarazo en el peso del recién nacido?
- (c) Calcular el intervalo de confianza para la media de la diferencia entre las puntuaciones Apgar al minuto de nacer y a los 5 minutos. ¿Cómo es la evolución de los niños en los primeros minutos de vida?
- (d) Si se considera que una puntuación Apgar al minuto menor o igual que 3 es un indicativo de que el niño ha nacido deprimido, calcular el intervalo de confianza del 90% para la diferencia de proporciones de niños deprimidos en el grupo de las madres fumadoras y en el de las no fumadoras.
- (e) ¿Se puede afirmar que la edad de la madre influye en el porcentaje de niños deprimidos al nacer?

# Contraste de Hipótesis

## 1 Ejercicios resueltos

1. Para averiguar si en una determinada población existen menos hombres que mujeres se plantea un contraste de hipótesis sobre la proporción de hombres que hay en la población:  $H_0 : p = 0.5$  frente a  $H_1 : p < 0.5$  y para ello se toma una muestra aleatoria de 10 personas. Se pide:

- (a) Suponiendo cierta la hipótesis nula, ¿qué distribución sigue la variable que mide el número de hombres en la muestra de tamaño 10?
- (b) Suponiendo cierta la hipótesis nula, ¿cuál es la probabilidad de que en la muestra se obtengan 0 hombres? ¿Se aceptaría la hipótesis nula en tal caso? Justificar la respuesta.



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades acumuladas](#).
2. En el cuadro de diálogo que aparece, introducir 0 en el campo Valor(es) de la variable, 10 en el campo Nº de repeticiones, 0.5 en el campo Probabilidad de éxito, marcar la opción Cola izquierda y hacer click en el botón Aceptar.

- (c) Suponiendo cierta la hipótesis nula, si se decide rechazarla cuando en la muestra haya 2 o menos hombres, ¿cuál es el riesgo de equivocarse?



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades acumuladas](#).
2. En el cuadro de diálogo que aparece, introducir 2 en el campo Valor(es) de la variable, 10 en el campo Nº de repeticiones, 0.5 en el campo Probabilidad de éxito, marcar la opción Cola izquierda y hacer click en el botón Aceptar.

- (d) Si el máximo riesgo de error  $\alpha$  que se tolera es 0.05, ¿qué número de hombres en la muestra formarían la región de rechazo de la hipótesis nula?



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades acumuladas](#).
2. En el cuadro de diálogo que aparece, introducir 1 en el campo Valor(es) de la variable, 10 en el campo Nº de repeticiones, 0.5 en el campo Probabilidad de éxito, marcar la opción Cola izquierda y hacer click en el botón Aceptar.

- (e) Suponiendo que la proporción real de hombres en la población fuese de 0.4, ¿cuál es la potencia del contraste para la región de rechazo del apartado anterior?



1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Discretas](#) » [Binomial](#) » [Probabilidades acumuladas](#).
2. En el cuadro de diálogo que aparece, introducir 1 en el campo Valor(es) de la variable, 10 en el campo Nº de repeticiones, 0.4 en el campo Probabilidad de éxito, marcar la opción Cola izquierda y hacer click en el botón Aceptar.

- (f) Si en lugar de una muestra de tamaño 10 se tomase una muestra de tamaño 100, y haciendo uso de la aproximación de una distribución binomial mediante una normal, ¿qué número de hombres en la muestra formarían la región de rechazo para un riesgo  $\alpha = 0.05$ ? ¿Qué potencia tendría ahora el contraste si la proporción real de hombres fuese de 0.4? ¿Es mejor o peor contraste que el anterior? Justificar la respuesta.



Una distribución binomial  $B(100, 0.5)$  puede aproximarse mediante una normal  $N(50, 5)$ .

1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Normal](#) » [Cuantiles](#).
2. En el cuadro de diálogo que aparece, introducir la probabilidad 0.05 en el campo Probabilidades, 50 en el campo media, 5 en el campo desviación típica, marcar la opción Cola izquierda y hacer click en el botón Aceptar.

El valor obtenido es la frontera entre la región de aceptación y la región de rechazo. Si en la muestra se obtienen menos hombres de dicho valor se rechazará la hipótesis nula, mientras que si se obtienen más se aceptará. Para calcular la potencia de contraste:

1. Seleccionar el menú [Teaching](#) » [Distribuciones](#) » [Continuas](#) » [Normal](#) » [Probabilidades acumuladas](#).
2. En el cuadro de diálogo que aparece, introducir el valor de la frontera en el campo Valor(es) de la variable, 40 en el campo media, 4.899 en el campo desviación típica, marcar la opción Cola izquierda y hacer click en el botón Aceptar.

- (g) Si se toma una muestra de tamaño 100 y se observan 41 hombres, ¿cuál es  $p$ -valor del contraste? ¿Podría rechazarse la hipótesis nula para un riesgo  $\alpha = 0.05$ ? ¿y para un riesgo  $\alpha = 0.01$ ?



1. Seleccionar el menú [Teaching](#) » [Test paramétricos](#) » [Proporciones](#) » [Test para una proporción](#).
2. En el cuadro de diálogo que aparece marcar la casilla de Introducción manual de frecuencias, introducir 41 en el campo Frecuencia muestral e introducir 100 en el campo Tamaño muestral.
3. En la solapa [Opciones de contraste](#), introducir 0.5 en el campo Hipótesis nula, seleccionar como hipótesis alternativa Unilateral menor y hacer click en el botón Aceptar.

2. 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  $\text{mg}/\text{mm}^3$ :

$$17.6 - 19.2 - 21.3 - 15.1 - 17.6 - 18.9 - 16.2 - 18.3 - 19.0 - 16.4$$

Se pide:

- (a) Crear un conjunto de datos con la variable concentracion.
- (b) Realizar el contraste de hipótesis bilateral:  $H_0: \mu = 18$  y  $H_1: \mu \neq 18$  con un nivel de significación 0.05.



1. Seleccionar el menú [Teaching](#) » [Test paramétricos](#) » [Medias](#) » [Test t para una muestra](#).
2. En el cuadro de diálogo que aparece seleccionar la variable concentracion.
3. En la solapa [Opciones de contraste](#), introducir 18 en el campo Hipótesis nula, seleccionar como hipótesis alternativa la opción Bilateral y hacer click sobre el botón Aceptar.

- (c) De igual manera realizar los contrastes bilaterales:  $H_0: \mu = 19.5$  y  $H_1: \mu \neq 19.5$  con un niveles de significación 0.05 y 0.01. ¿Cómo afecta la disminución en el nivel de significación en la facilidad para rechazar  $H_0$ ?



Seguir los mismos pasos del apartado anterior introduciendo 19.5 en el campo Hipótesis nula.

- (d) Realizar los contrastes bilaterales y unilaterales para la hipótesis nula  $H_0: \mu = 17$  con un nivel de significación de 0.05. ¿Qué relación hay entre el  $p$ -valor de los contrastes bilateral y unilaterales?





1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para una muestra**.
2. En el cuadro de diálogo que aparece seleccionar la variable concentración.
3. En la solapa **Opciones de contraste**, introducir 17 en el campo Hipótesis nula, seleccionar como hipótesis alternativa Bilateral y hacer click sobre el botón Aceptar.
4. Para el contraste de menor repetir lo mismo seleccionando como hipótesis alternativa Unilateral menor.
5. Para el contraste de mayor repetir lo mismo seleccionando como hipótesis alternativa Unilateral mayor.

- (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  $17 \text{ mg/mm}^3$ , ¿se acepta o se rechaza la afirmación del fabricante?
- (f) ¿Cuál sería el tamaño muestral requerido para poder detectar una diferencia de  $0.5 \text{ mg/mm}^3$  más con un nivel de significación  $\alpha = 0.05$  y una potencia  $1 - \beta = 0.8$ ?



Para calcular el tamaño muestral se necesita saber la desviación típica de la población o una estimación suya. Para ello se calcula previamente la cuasivarianza muestral

1. Seleccionar el menú **Teaching** » **Estadística descriptiva** » **Estadísticos**.
2. En el cuadro de diálogo que aparece seleccionar la variable concentración.
3. En la solapa **Estadísticos básicos** marcar la opción Cuasidesviación típica y hacer click en el botón Aceptar.

Para calcular el tamaño muestral:

1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Cálculo del tamaño muestral para el test T**.
2. En el cuadro de diálogo que aparece introducir en el campo Diferencia en las medias el valor 0.5, introducir en el campo Desviación típica el valor de la cuasidesviación típica obtenido, introducir el valor 0.05 en el campo Nivel de significación, introducir el valor 0.8 en el campo Potencia, marcar la opción Una muestra en el campo Tipo de test, seleccionar como hipótesis alternativa Unilateral y hacer click sobre el botón Aceptar.

3. En una encuesta realizada en una facultad, sobre si el alumnado utiliza habitualmente (al menos una vez a la semana) la biblioteca de la misma, se han obtenido los siguientes resultados:

Alumno	1	2	3	4	5	6	7	8	9	10	11	12
Respuesta	No	Si	No	No	No	Si	No	Si	Si	Si	Si	No

Alumno	13	14	15	16	17	18	19	20	21	22	23
Respuesta	Si	No	Si	No	No	No	Si	Si	Si	No	No

Alumno	24	25	26	27	28	29	30	31	32	33	34
Respuesta	Si	No	No	Si	Si	No	No	Si	No	Si	No

- (a) Crear un conjunto de datos con la variable respuesta como factor.
- (b) Contrastar si el porcentaje de alumnos que utiliza regularmente la biblioteca es superior al 40%.



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Proporciones** » **Test para una proporción**.
2. En el cuadro de diálogo que aparece seleccionar la variable respuesta e introducir Si en el campo Categoría.
3. En la solapa **Opciones de contraste**, introducir 0.4 en el campo Hipótesis nula, seleccionar como hipótesis alternativa Unilateral mayor y hacer click en el botón Aceptar.

4. 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

- Crear un conjunto de datos con las variables población y edad.
- Realizar un contraste de hipótesis con un nivel de significación de 0.05 para dar respuesta a la conclusión que buscan los investigadores.



- Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para dos muestras independientes**.
  - En el cuadro de dialogo que aparece seleccionar la variable edad en el campo Comparar y la variable población en el campo Según.
  - En la solapa **Opciones de contraste** seleccionar como hipótesis alternativa Bilateral, marcar la casilla Suponer varianzas iguales y hacer click sobre el botón Aceptar.
- Hay diferencias entre las poblaciones si el *p*-valor es menor que 0.05.

- 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	8.90
66.0	46.1
25.2	29.8
40.1	36.2

- Crear un conjunto de datos con las variables antes y después e introducir los datos.
- Plantear el contraste de hipótesis adecuado para confirmar o denegar la hipótesis de los investigadores.



- Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **Test t para dos muestras pareadas**.
- En el cuadro de diálogo que aparece seleccionar la variable antes en el campo Comparar y la variable después en el campo Con.
- En la solapa **Opciones de contraste** seleccionar como hipótesis alternativa Unilateral mayor y hacer click en el botón Aceptar.

- 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. ¿Se puede afirmar que hay diferencias significativas entre los porcentajes de aprobados en ambos grupos? Justificar la respuesta.



- (a) Seleccionar el menú **Teaching** » **Test paramétricos** » **Proporciones** » **Test para dos proporciones**, marcar la casilla **Introducción manual de frecuencias**, introducir 55 en el campo **Frecuencia muestral 1**, introducir 80 en el campo **Tamaño muestral 1**, introducir 32 en el campo **Frecuencia muestral 2** e introducir 90 en el campo **Tamaño muestral 2**.
- (b) En la solapa **Opciones de contraste**, seleccionar como hipótesis alternativa **Bilateral** y hacer click en el botón **Aceptar**.

## 2 Ejercicios propuestos

1. El fichero **pulso.txt** contiene información sobre el pulso de un grupo de pacientes que han realizado distintos ejercicios: pulso en reposo (**pulse1**), pulso después de hacer ejercicio (**pulse2**), tipo de ejercicio (**ran**, 1=correr, 2=andar), sexo (**sex**, 1=hombre, 2=mujer) y peso (**weight**). Se pide:
  - (a) Contrastar si el pulso en reposo está por debajo de 75 pulsaciones.
  - (b) ¿Qué tamaño muestral sería necesario para detectar una diferencia de 2 pulsaciones más en la media de las pulsaciones en reposo, con un nivel de significación 0.05 y una potencia de 0.9?
  - (c) Contrastar si el pulso después de correr está por encima de 85 pulsaciones.
  - (d) Contrastar si el porcentaje de personas con taquicardia leve (número de pulsaciones en reposo por encima de 90) supera el 5%.
  - (e) ¿Se puede afirmar que el ejercicio aumenta las pulsaciones con una significación de 0.05? ¿y con una significación 0.01? Justificar la respuesta.
  - (f) ¿Existen diferencias entre las pulsaciones después de andar y después de correr? Justificar la respuesta.
  - (g) ¿Existen diferencias entre las pulsaciones en reposo entre hombres y mujeres? ¿Y entre las pulsaciones después de correr? Justificar la respuesta.



# Análisis de la Varianza de 1 Factor

## 1 Ejercicios resueltos

- Se realiza un estudio para comparar la eficacia de tres programas terapéuticos para el tratamiento del acné. Se emplean tres métodos:

- Lavado, dos veces al día, con cepillo de polietileno y un jabón abrasivo, junto con el uso diario de 250 mg de tetraciclina.
- 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.
- Evitar el agua, lavado dos veces al día con un limpiador sin lípidos y uso de crema de tretinoína y de peróxido benzofílico.

En el estudio participan 35 pacientes. Se separó aleatoriamente a estos pacientes en tres subgrupos 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	

- Crear un conjunto de datos con las variables tratamiento y mejora e introducir los datos de la muestra.
- Dibujar el diagrama de puntos. ¿Se observan diferencias entre los tratamientos en el diagrama?



- Seleccionar el menú **Teaching > Gráficos > Diagrama de puntos**.
- En el cuadro de diálogo que aparece, seleccionar la variable mejora en el campo Variable, seleccionar la variable tratamiento en el campo Grupos, y hacer click sobre el botón Aceptar.

- Realizar el contraste de ANOVA. ¿Se puede concluir que los tres tratamientos tienen el mismo efecto medio con un nivel de significación de 0.05?



- Seleccionar el menú **Teaching > Test paramétricos > Medias > ANOVA**.
- En el cuadro de diálogo que aparece, seleccionar el conjunto de datos acne en el campo Conjunto de datos.
- Seleccionar la variable mejora en el campo Variable dependiente, la variable tratamiento en el campo Factores entre individuos 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 comparación de los tres tratamientos con una significación de 0.05. ¿Entre qué parejas de tratamientos hay diferencias estadísticamente significativas?



Repetir los mismos pasos del apartado anterior activando la opción Comparaciones de medias por pares de la solapa **Comparación por pares**.

- (e) Comprobar que se cumple la hipótesis de igualdad de varianzas entre los tratamientos.



Repetir los mismos pasos del apartado anterior activando la opción Test de Levene de la solapa **Opciones**.

- (f) Dibujar el gráfico de los intervalos de confianza para la media de cada tratamiento.



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de medias**.
2. En el cuadro de diálogo que aparece, seleccionar la variable mejora en el campo Variable y seleccionar la variable tratamiento en el campo Grupos.
3. Seleccionar la opción Intervalos de confianza y hacer click sobre el botón Aceptar.

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	

- (a) Crear un conjunto de datos con las variables nota y centro e introducir los datos de la muestra.  
 (b) Dibujar el diagrama de puntos. ¿Se observan diferencias entre los centros en el diagrama?



1. Seleccionar el menú **Teaching** » **Gráficos** » **Diagrama de puntos**.
2. En el cuadro de diálogo que aparece, seleccionar la variable nota en el campo Variable, seleccionar la variable centro en el campo Grupos y hacer click sobre el botón Aceptar.

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



1. Seleccionar el menú **Teaching** » **Test paramétricos** » **Medias** » **ANOVA**.
2. En el cuadro de diálogo que aparece, seleccionar el conjunto de datos selectividad en el campo Conjunto de datos.
3. Seleccionar la variable notas en el campo Variable dependiente, la variable centro en el campo Factores entre individuos y hacer click sobre el botón Aceptar.

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



Repetir los mismos pasos del apartado anterior activando la opción Comparaciones de medias por pares de la solapa [Comparación por pares](#).

## 2 Ejercicios propuestos

- 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  $\alpha = 0.05$ .

- 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  $\alpha = 0,05$ .