

Manual de estadística para Biometría con aplicaciones farmacéuticas en R



José Antonio Sáez Muñoz

Dpto. de Estadística e Investigación Operativa, Universidad de Granada,
Fuente Nueva s/n, 18071 Granada, España (e-mail: joseasaezm@ugr.es).

Biometría
Grado en Farmacia
© 2023 · v1.0

Créditos y derechos

Título: Manual de estadística para Biometría con aplicaciones farmacéuticas en R

Autor: José Antonio Sáez Muñoz (joseasaezm@ugr.es)

Dpto. Estadística e Investigación Operativa, Universidad de Granada, Fuente Nueva s/n, 18071 Granada, España.

Año: © 2023 José Antonio Sáez Muñoz

Versión: v1.0

Licencia: Este material docente se publica bajo la licencia *Creative Commons Attribution 4.0 International* (CC BY 4.0), que permite su uso, copia, distribución y adaptación siempre que se cite correctamente al autor.

Más información en <https://creativecommons.org/licenses/by/4.0/>

Finalidad educativa: Este material docente ha sido elaborado con fines exclusivamente educativos para la formación práctica de la asignatura de Biometría del Grado en Farmacia de la Universidad de Granada.

Citación recomendada: Sáez, J. A. (2023). *Manual de estadística para Biometría con aplicaciones farmacéuticas en R*. Universidad de Granada, Granada, España.

Parte I: Introducción a R	4
1. El entorno R	5
2. Elementos básicos de R	7
3. Ejercicios	13
Parte II: Estadística descriptiva	16
1. Conceptos principales	17
2. Variables cualitativas	20
3. Variables cuantitativas discretas	26
4. Variables cuantitativas continuas	29
5. Medidas descriptivas	34
6. Diagrama de cajas	53
7. Ejercicios	56
Parte III: Regresión	64
1. Análisis de dos variables	65
2. Nube de puntos	66
3. Regresión lineal	68
4. Medidas de relación y ajuste lineal	79
5. Otros modelos de regresión	82
6. Ejercicios	89
Bibliografía	98

Parte I: Introducción a R

1. El entorno R

a) ¿Qué es R?

Lenguaje de programación: Conjunto de instrucciones y reglas que permiten indicar a un ordenador qué operaciones debe realizar.

R: Lenguaje de programación orientado al análisis estadístico y a la visualización de datos.

Ventajas de R:

- Gratuito y de código abierto.
- Gran variedad de *paquetes* (código que amplía las capacidades de R).
- Muy utilizado, con abundante documentación y recursos de apoyo.

Aplicaciones en Farmacia:

- Análisis de datos de ensayos clínicos.
- Interpretación de resultados de pruebas analíticas.
- Apoyo en la selección de tratamientos farmacéuticos.

1. El entorno R

b) Ejecución de código

Entornos de trabajo en R:

1. **Consola**: núcleo básico de R; aquí se **escribe** y se ven **resultados**.
2. **RGui**: interfaz gráfica **sencilla** incluida con R.
3. **RStudio**: entorno **completo y cómodo** con múltiples funcionalidades.

Formas básicas de ejecutar código:

1. **Consola**:
 - Escribir el código y pulsar **Enter**.
 - Útil para pruebas rápidas, pero lo escrito se pierde al cerrar R.
2. **Scripts en RGui / RStudio (.R)**:
 - Seleccionar el bloque de código y pulsar el botón **Run**.
 - Permiten guardar el código para reutilizarlo.

Salida:

- El resultado siempre aparece en la **consola**.
- **Errores y advertencias** se muestran en otro color.

2. Elementos básicos de R

a) Variables

Una **variable** es un contenedor utilizado para almacenar datos.

Tipos de datos más comunes:

- Numéricos: 30, 4.5.
- Caracteres: "Juan", "María".
- Lógicos: TRUE, FALSE.

Asignación de valores a una variable:

Se hace siguiendo la estructura: `variable <- valor`

```
edad <- 30  
nombre <- "Juan"  
activo <- TRUE
```

2. Elementos básicos de R

a) Variables

Reglas para nombrar variables:

1. Comenzar con letra (a-z, A-Z). Luego: letras, números, punto o guion (-).
 2. No utilizar espacios; emplear guion (-), punto (.) o notación camelCase.
 3. Sensibilidad a mayúsculas y minúsculas: edad, Edad y EDAD son distintas.
-

Ejemplos de nombres válidos:

- `mi_variable`
- `edadUsuario`
- `total1`

Ejemplos de nombres no válidos:

- `1raVariable`
- `mi variable`
- `TRUE`

2. Elementos básicos de R

b) Vectores

Un **vector** es una estructura que almacena una secuencia de elementos del mismo tipo: numéricos, de caracteres o lógicos.

Creación de vectores:

Se utiliza la función `c()` (*combine* o *concatenate*).

```
edades <- c(21, 17, 23, 58, 43)
nombres <- c("Juan", "María", "Carlos", "Laura")
```

Acceso y modificación de elementos:

- Se usan corchetes `[]` indicando la posición (los índices comienzan en 1).
- Se puede asignar un nuevo valor a una posición usando `<-`.

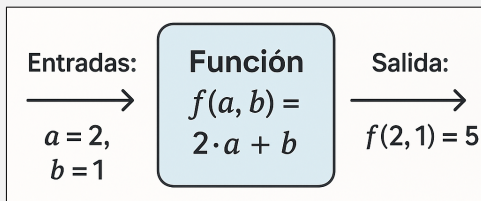
```
notas <- c(7.5, 8.2, 6.9, 9.1)
segunda_nota <- notas[2]
notas[3] <- 7.8
```

2. Elementos básicos de R

c) Funciones

Una **función** es un bloque de código diseñado para realizar una tarea específica.

- Recibe uno o varios **parámetros de entrada**.
- Realiza cálculos o procesos **internos** usando esos parámetros.
- Devuelve un **valor de salida** al usuario.



Ejemplo: Función $f()$ que recibe dos parámetros, a y b , realiza el cálculo $2a + b$ y devuelve el resultado.

2. Elementos básicos de R

c) Funciones

Uso de funciones en R:

Sintaxis básica: `nombre_funcion(parámetros)`

Ejemplo: La función `mean()` recibe un vector numérico y devuelve su media.

```
notas <- c(7.5, 8.2, 6.9, 9.1)
mean(notas)
```

Documentación de funciones:

Anteponiendo `?` al nombre se abre la ayuda, con explicación y ejemplos de uso.

```
?mean
```

2. Elementos básicos de R

d) Comentarios

Los **comentarios** sirven para añadir explicaciones o notas en el código.

- Se escriben con el símbolo #.
- No forman parte de la ejecución del programa.

Código de ejemplo:

```
# Crear el vector de notas
notas <- c(7.5, 8.2, 6.9, 9.1)

# Calcular la media del vector
mean(notas)
```


3. Ejercicios

Ejercicio 1

Determine cuáles de las siguientes asignaciones de variables en R son correctas e incorrectas, justificando cada caso.

- a) `volumenSuero <- "30"`
- b) `via_administracion <- "oral"`
- c) `2fasePacientes <- 30`
- d) `mg diarios <- 2`
- e) `tipo.medicamento <- "Paracetamol"`
- f) `numComprimidos <- 25`
- g) `cantidad$ml <- 500`

3. Ejercicios

Ejercicio 2

Se desea registrar información de pacientes en un estudio farmacéutico. Realizar las siguientes tareas en R:

- Crear un vector llamado `dosis` con las dosis administradas a 5 pacientes: 50, 75, 100, 50, 75.
- Crear un vector llamado `tipo_sangre` con los tipos de sangre de los pacientes: A, B, A, AB, A.
- Acceder a la dosis del tercer paciente y guardarla en la variable `dosis_tercero`.
- Cambiar el tipo de sangre del segundo paciente al valor AB.

3. Ejercicios

Ejercicio 3

En R existe la función `length()` que recibe un vector como parámetro y devuelve su tamaño. Realizar las siguientes tareas:

- Crear un vector llamado `meses` con la duración del tratamiento de 4 pacientes: 3, 6, 12, 9.
- Utilizar la función `length()` para calcular cuántos elementos tiene el vector `meses`.
- Asignar a la variable `num_pacientes` el resultado de aplicar la función `length()` al vector `meses`.
- Mostrar el valor almacenado en `num_pacientes`.

Parte II: Estadística descriptiva

1. Conceptos principales

Estadística:

Disciplina que se encarga de recolectar, organizar, analizar e interpretar **datos** para ayudar a tomar **decisiones informadas** en diversos contextos.

- Las **decisiones basadas en datos** suelen ser más acertadas que las basadas en intuición o suposiciones.
-

Estadística descriptiva:

Rama de la estadística centrada en **organizar, representar y describir** los datos de manera clara y comprensible. Es el primer paso para entender los datos y facilitar análisis posteriores.

1. Conceptos principales

a) Elementos fundamentales para el análisis de datos

1. **Población:** Conjunto de **todos** los elementos que comparten una característica.
2. **Muestra:** **Subconjunto** representativo de la población para el análisis.
3. **Individuo:** Cada unidad dentro de la población o muestra.
4. **Variable:** Característica que se mide o analiza en cada individuo.



1. Conceptos principales

b) Tipos de variables

1. **Variables cualitativas:** Representan categorías o cualidades. Sus valores se expresan generalmente con palabras.
2. **Variables cuantitativas:** Representan cantidades o medidas y se expresan numéricamente. Pueden ser:
 - **Discretas:** Toman valores específicos, normalmente expresados con números enteros, los cuales no tienen valores intermedios.
 - **Continuas:** Sus valores pueden ser cualquier número dentro de un intervalo, incluyendo decimales. Normalmente se expresan con números reales.

2. Variables cualitativas

a) Tabla de frecuencias

Definición: Tabla que organiza los datos y muestra, para **cada valor**, cuántas veces aparece (**frecuencia**).

Ejemplo:

Para la variable $X = \text{"tipo de medicamento vendido"}$

se han recogido los siguientes datos: A, B, A, H, B, A, B, B, B, H, B, B, A, H,
donde: A = Analgésico, B = Antibiótico, H = Antihistamínico.

Tabla de frecuencias:

Medicamento	Frecuencia
A	4
B	7
H	3

Notación:

- n es el **número total de datos**.
- x_i representa el **valor del dato i -ésimo** ($x_1, x_2, x_3, \dots, x_n$).

2. Variables cualitativas

a) Tabla de frecuencias

Cálculo en R:

- Se emplea la función `table()`.
- Recibe como parámetro un **vector**.

Código de ejemplo:

```
# Datos: A = Analgésico, B = Antibiótico, H = Antihistamínico
medicamento <- c("A", "B", "A", "H", "B", "A", "B", "B", "B", "H",
                  "B", "B", "A", "H")

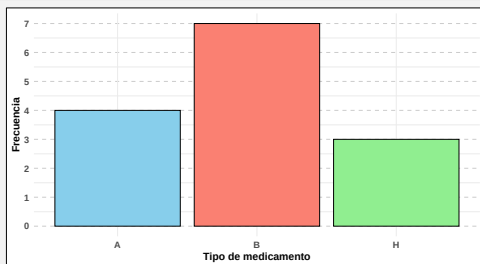
# Crear tabla de frecuencias
table(medicamento)
```

2. Variables cualitativas

b) Gráficos: Gráfico de barras

Gráficos: Representaciones visuales que muestran los datos de forma clara y facilitan comparar frecuencias e identificar valores atípicos.

Gráfico de barras: Cada barra representa un valor y su altura indica la frecuencia. Útil para comparar frecuencias o cuando hay muchos valores.



Ejemplo: Gráfico de barras mostrando la frecuencia de cada tipo de medicamento vendido.

2. Variables cualitativas

b) Gráficos: Gráfico de barras

Cálculo en R:

- Se emplea la función `barplot()`.
- Recibe como parámetro una **tabla de frecuencias**.

Código de ejemplo:

```
# Datos: A = Analgésico, B = Antibiótico, H = Antihistamínico
medicamento <- c("A", "B", "A", "H", "B", "A", "B", "B", "B", "H",
                  "B", "B", "A", "H")

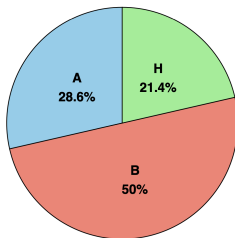
# Crear tabla de frecuencias
freq <- table(medicamento) # freq almacena la tabla

# Crear gráfico de barras
barplot(freq)
```

2. Variables cualitativas

b) Gráficos: Gráfico de sectores

Gráfico de sectores: Cada sector representa un valor y su tamaño es proporcional a su frecuencia. Útil para mostrar proporciones, especialmente cuando hay pocos valores.



Ejemplo: Gráfico de sectores mostrando la proporción de cada tipo de medicamento vendido.

2. Variables cualitativas

b) Gráficos: Gráfico de sectores

Cálculo en R:

- Se emplea la función `pie()`.
- Recibe como parámetro una **tabla de frecuencias**.

Código de ejemplo:

```
# Datos: A = Analgésico, B = Antibiótico, H = Antihistamínico
medicamento <- c("A", "B", "A", "H", "B", "A", "B", "B", "B", "H",
                  "B", "B", "A", "H")

# Crear tabla de frecuencias
freq <- table(medicamento) # freq almacena la tabla

# Crear gráfico de sectores
pie(freq)
```

3. Variables cuantitativas discretas

a) Tabla de frecuencias

Al igual que en las variables cualitativas, los datos pueden organizarse en una **tabla de frecuencias**, que indica cuántas veces aparece cada valor.

Ejemplo:

Para la variable $X = \text{"número de comprimidos diarios por paciente"}$

se han recogido los datos: 0, 1, 2, 2, 1, 2, 3, 3, 2, 1, 3, 2, 3, 0, 1, 2, 3, 2, 3, 1.

Tabla de frecuencias:

Comprimidos	Frecuencia
0	2
1	5
2	7
3	6

3. Variables cuantitativas discretas

a) Tabla de frecuencias

Cálculo en R:

- Se emplea la función `table()`.
- Recibe como parámetro un **vector**.

Código de ejemplo:

```
# Datos: número de comprimidos tomados al día por paciente
comprimidos <- c(0, 1, 2, 2, 1, 2, 3, 3, 2, 1, 3, 2, 3, 0, 1, 2, 3,
2, 3, 1)

# Crear tabla de frecuencias
table(comprimidos)
```

3. Variables cuantitativas discretas

b) Gráficos: Barras y sectores

Con variables discretas pueden emplearse los mismos gráficos que en variables cualitativas: **barras y sectores**.

Cálculo en R:

- **Gráfico de barras:** función `barplot()`.
- **Gráfico de sectores:** función `pie()`.
- Reciben como parámetro una **tabla de frecuencias**.

Código de ejemplo:

```
freq <- table(comprimidos) # freq almacena la tabla  
  
barplot(freq) # Crear gráfico de barras  
  
pie(freq) # Crear gráfico de sectores
```


4. Variables cuantitativas continuas

a) Tabla de frecuencias

Para crear una **tabla de frecuencias**, se dividen los datos en **intervalos** y se asigna cada valor a su intervalo correspondiente.

Luego se cuenta cuántos datos caen en cada intervalo.

Ejemplo:

Para la variable $X = \text{"duración de acción de un medicamento (en horas)"}$

se han recogido los datos: 4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4.

Se agrupan en intervalos de 2 horas, de 2 a 8.

Tabla de frecuencias:

Intervalo	Frecuencia
2 – 4	2
4 – 6	6
6 – 8	4

4. Variables cuantitativas continuas

a) Tabla de frecuencias

Cálculo en R:

- Primero se crean los **intervalos** con la función `cut()`.
 - Parámetros: el vector de **datos** y **breaks** (extremos de intervalos).
- Luego se calcula la tabla de frecuencias con `table()`.
 - Parámetros: el vector de **intervalos**.

Código de ejemplo:

```
# Datos: duración de acción de un medicamento (horas)
duracion <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1,
3.8, 5.4)

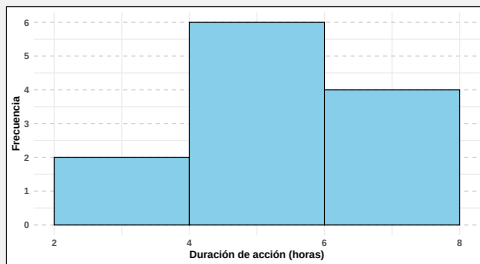
# Crear intervalos de 2 en 2 horas desde 2 hasta 8
intervalos <- cut(duracion, breaks = c(2, 4, 6, 8))

# Crear tabla de frecuencias
table(intervalos)
```

4. Variables cuantitativas continuas

b) Gráficos: Histograma

Histograma: Gráfico donde en el eje X se muestran los valores de la variable divididos en intervalos y en el eje Y la altura de cada barra indica la frecuencia de datos en ese intervalo.



Ejemplo: Histograma mostrando la frecuencia de pacientes según la duración de acción del medicamento, con los datos agrupados en intervalos de 2 horas.

4. Variables cuantitativas continuas

b) Gráficos: Histograma

Cálculo en R:

- Se emplea la función `hist()`.
- Parámetros: el vector de **datos** y **breaks** (extremos de intervalos).

Código de ejemplo:

```
# Datos: duración de acción de un medicamento (horas)
duracion <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1,
3.8, 5.4)

# Crear histograma
hist(duracion, breaks = c(2, 4, 6, 8))
```

4. Variables cuantitativas continuas

c) Tablas de frecuencias y gráficos asociados

Tablas de frecuencias: Normalmente, se usan **tablas sin intervalos** para variables **discretas** y **tablas con intervalos** para variables **continuas**.

Esta elección puede modificarse según factores como la **frecuencia de los valores** o el **nivel de detalle** deseado:

- Pocos valores distintos y repetidos $\Rightarrow$ tabla **sin intervalos**.
- Muchos valores o muy dispersos $\Rightarrow$ tabla **con intervalos**.

Gráficos asociados: Datos sin agrupar (diagramas de barras o de sectores); Datos agrupados en intervalos (histograma).

Ejemplo: Variable discreta agrupada

Para la variable $X =$ “*número de comprimidos anuales tomados por paciente*”, se puede agrupar en intervalos para simplificar la presentación de los datos:

Intervalo	0 – 100	100 – 300	300 – 600
Frecuencia	1025	5024	2620

5. Medidas descriptivas

Definición: Valores numéricos que muestran las **características principales** de un conjunto de datos.

Permiten realizar **comparaciones directas** entre conjuntos de datos, algo que tablas o gráficos no siempre muestran con claridad.

Clasificación:

1. **Medidas de posición:** Muestran qué valor ocupa una posición concreta dentro del conjunto de datos ordenados.
 - **Posición central:** Indican el centro de los datos, es decir, el valor que mejor representa al conjunto.
 - **Posición no central:** Indican valores en distintas posiciones del conjunto de datos, como los valores bajos, medios o altos.
2. **Medidas de dispersión:** Indican qué tan lejos están los datos entre sí.

5. Medidas descriptivas: Posición central

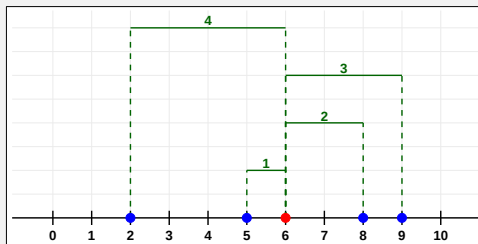
a) Media aritmética ($\bar{x}$)

Fórmula:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Interpretación:

La media puede verse como un *punto de equilibrio*, ya que estima el centro como el valor que iguala la suma de distancias a los datos menores y a los mayores.



Ejemplo: Consideremos las notas de un estudiante en cuatro exámenes: 2, 5, 8 y 9. La media es: $\bar{x} = (2 + 5 + 8 + 9)/4 = 6$. El 6 actúa como punto de equilibrio, ya que $4 + 1 = 2 + 3$.

5. Medidas descriptivas: Posición central

a) Media aritmética ($\bar{x}$)

Cálculo en R:

- Se emplea la función `mean()`.
- Recibe como parámetro un **vector numérico**.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular la media aritmética
mean(x)
```

Resultado: 5.333

5. Medidas descriptivas: Posición central

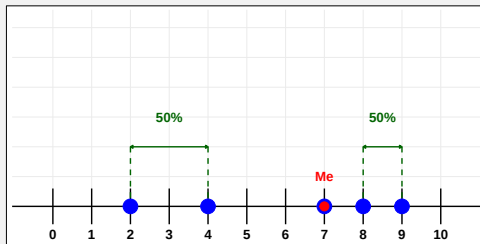
b) Mediana (Me)

Fórmula (datos ordenados):

$$Me = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Interpretación:

La mediana es el valor central cuando los datos están ordenados. Divide el conjunto en dos partes iguales: la mitad de los datos queda por debajo y la otra mitad por encima.



Ejemplo: Consideremos las notas de un estudiante en cinco exámenes: 2, 4, 7, 8 y 9. La mediana es el valor que ocupa la posición central en el conjunto ordenado, por tanto, $Me = 7$.

5. Medidas descriptivas: Posición central

b) Mediana (Me)

Cálculo en R:

- Se emplea la función `median()`.
- Recibe como parámetro un **vector numérico**.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular la mediana
median(x)
```

Resultado: 5.45

5. Medidas descriptivas: Posición central

c) Moda (Mo)

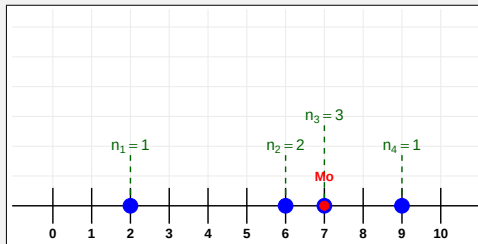
Fórmula:

Si tenemos k valores distintos $x_1, \dots, x_k$ con frecuencias $n_1, \dots, n_k$:

$$Mo = \{x_j : n_j = \max(n_j), \quad j = 1, 2, \dots, k\}$$

Interpretación:

La moda representa el valor del conjunto de datos que ocurre con mayor frecuencia.



Ejemplo: Consideremos las notas de un estudiante en siete exámenes: 2, 6, 6, 7, 7, 7 y 9. La moda es $Mo = 7$, ya que es el valor que más se repite.

5. Medidas descriptivas: Posición central

c) Moda (*Mo*)

Cálculo en R:

- Se emplea la función `table()` para obtener la frecuencia de cada valor.
- La moda es el dato que aparece con **mayor frecuencia**.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(0, 1, 2, 2, 1, 2, 3, 3, 2, 1, 3, 2, 3, 0, 1, 2, 3, 2, 3, 1)

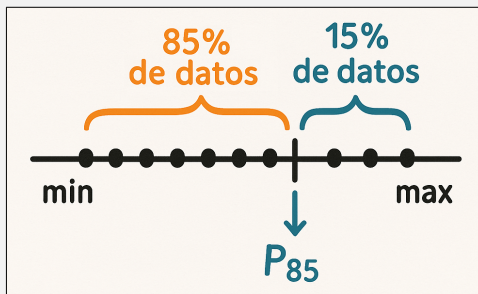
# Calcular la moda: Buscar el valor con mayor frecuencia
table(x)
```

Resultado: 2

5. Medidas descriptivas: Posición no central

a) Percentil (P_k)

Definición: El percentil k (P_k) es el valor que deja por debajo el $k\%$ de los datos ordenados. Existen 99 percentiles: $P_1, P_2, \dots, P_{99}$.



Ejemplo: El percentil P_{85} es el valor que deja por debajo al 85% de los datos cuando están ordenados.

5. Medidas descriptivas: Posición no central

a) Percentil (P_k)

Cálculo en R:

- Se emplea la función `quantile()`.
- Recibe **dos parámetros**:
 - Un **vector numérico** con los datos.
 - Un **número entre 0 y 1** con el percentil (p.e., 0.85 para P_{85}).

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular el percentil 85
quantile(x, 0.85)
```

Resultado: 6.77

5. Medidas descriptivas: Posición no central

b) Cuartiles (Q_i)

Definición: Son tres valores que dividen los datos ordenados en **4 partes iguales**. Se relacionan con los percentiles de la siguiente manera:

$$Q_1 = P_{25}, \quad Q_2 = P_{50}, \quad Q_3 = P_{75}$$

Cálculo en R:

- Se emplea la función `quantile()`, igual que con los percentiles.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular el cuartil 1
quantile(x, 0.25)
```

Resultado: 4.175

5. Medidas descriptivas: Posición no central

c) Deciles (D_j)

Definición: Son nueve valores que dividen los datos ordenados en **10 partes iguales**. Se relacionan con los percentiles de la siguiente manera:

$$D_1 = P_{10}, D_2 = P_{20}, \dots, D_9 = P_{90}$$

Cálculo en R:

- Se emplea la función `quantile()`, igual que con los percentiles.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular el decil 4
quantile(x, 0.4)
```

Resultado: 5.04

5. Medidas descriptivas: Dispersión

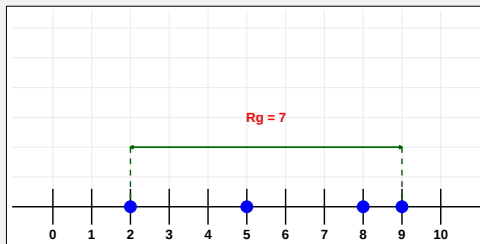
a) Rango (R_g)

Fórmula:

$$R_g = x_{\max} - x_{\min}$$

Interpretación:

El rango mide la amplitud total que abarcan los datos, es decir, la distancia entre el valor máximo y el valor mínimo.



Ejemplo: Consideremos las notas de un estudiante en 4 exámenes: 2, 5, 8 y 9. El rango es $R_g = 9 - 2 = 7$.

5. Medidas descriptivas: Dispersión

a) Rango (R_g)

Cálculo en R:

- Se emplean las funciones `max()` y `min()`.
- Cada una recibe un **vector numérico** con los datos.
- El rango se obtiene como `max(x) - min(x)`.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular el rango
max(x)-min(x)
```

Resultado: 4

5. Medidas descriptivas: Dispersión

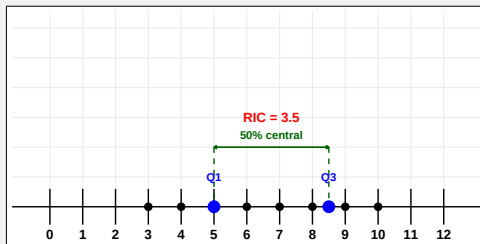
b) Rango intercuartílico (RIC)

Fórmula:

$$RIC = Q_3 - Q_1$$

Interpretación:

El RIC mide la amplitud del 50% central de los datos, es decir, la distancia entre Q_3 y Q_1 . Es una medida de dispersión que evita la influencia de valores extremos.



Ejemplo: Consideremos las notas de un estudiante en siete exámenes: 3, 4, 6, 7, 8, 9 y 10. El rango intercuartílico es $RIC = Q_3 - Q_1 = 8.5 - 5 = 3.5$, que representa la amplitud del 50% central de los datos.

5. Medidas descriptivas: Dispersión

b) Rango intercuartílico (*RIC*)

Cálculo en R:

- Se emplean dos llamadas a la función `quantile()`.
- El RIC se obtiene como `quantile(x, 0.75) - quantile(x, 0.25)`.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular el rango intercuartílico
quantile(x, 0.75) - quantile(x, 0.25)
```

Resultado: 2.075

5. Medidas descriptivas: Dispersión

c) Varianza (s^2)

Fórmula:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Interpretación:

Mide cuánto se alejan, en promedio, los datos respecto a la media. Como está en unidades al cuadrado, su interpretación no es intuitiva.

Corrección de Bessel (división entre $n - 1$ en lugar de n):

- Cuando calculamos la varianza con una muestra dividiendo entre n , el resultado suele ser un poco menor que la varianza real de toda la población.
- Para corregirlo, dividimos entre $n - 1$ en vez de n , acercándonos así más al valor real de la varianza poblacional.

5. Medidas descriptivas: Dispersión

c) Varianza (s^2)

Cálculo en R:

- Se emplea la función `var()`.
- Recibe un **vector numérico** con los datos.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)

# Calcular la varianza
var(x)
```

Resultado: 1.701

5. Medidas descriptivas: Dispersión

d) Desviación típica (s)

Fórmula:

$$s = \sqrt{s^2}$$

Interpretación: Mide cuánto se alejan, en promedio, los datos respecto a la media, pero manteniendo las mismas unidades que los datos originales.

Cálculo en R:

- Se emplea la función `sd()`.
- Recibe un **vector numérico** con los datos.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)
# Calcular la desviación típica
sd(x)
```

Resultado: 1.304

5. Medidas descriptivas: Dispersión

e) Coeficiente de variación (CV)

Fórmula:

$$CV = \frac{s}{\bar{x}}$$

Interpretación: Expresa s como proporción de $\bar{x}$. Permite comparar la dispersión entre conjuntos de datos, ya que es adimensional.

Cálculo en R:

- Se emplean las funciones `sd()` y `mean()`.
- Ambas reciben un **vector numérico** con los datos.

Código de ejemplo:

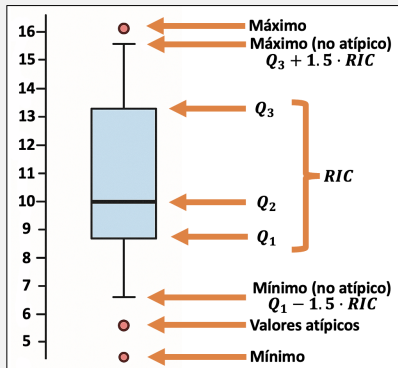
```
# Crear vector con los datos
x <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 3.3, 6.7, 4.1, 3.8, 5.4)
# Calcular el coeficiente de variación
sd(x)/mean(x)
```

Resultado: 0.245

6. Diagrama de cajas

Definición: Gráfico que resume los datos mostrando los valores mínimo y máximo, los cuartiles, la dispersión y los posibles valores atípicos.

Elementos principales:

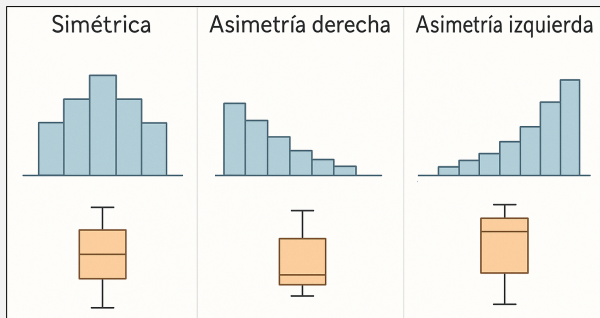


- **Caja:** Intervalo entre Q_1 y Q_3 .
- **Línea en la caja:** Me .
- **Bigotes:** Se extienden hasta el máximo y mínimo no atípico ($Q_3 + 1.5 \cdot RIC$ y $Q_1 - 1.5 \cdot RIC$).
- **Puntos fuera de los bigotes:** Valores atípicos o extremos.

6. Diagrama de cajas

Formas de la distribución:

1. **Distribución simétrica.** Los datos están equilibrados alrededor de la mediana.
2. **Asimetría a la derecha (positiva).** Los datos se concentran en los valores bajos de la variable y la cola se extiende hacia valores altos.
3. **Asimetría a la izquierda (negativa).** Los datos se concentran en los valores altos de la variable y la cola se extiende hacia valores bajos.



Ejemplo: Principales formas de distribución de los datos.

6. Diagrama de cajas

Cálculo en R:

- Se utiliza la función `boxplot()`.
- Recibe un **vector numérico** con los datos.

Código de ejemplo:

```
# Crear vector con los datos
x <- c(16, 23, 15, 20, 20, 15, 16, 17, 23, 9, 23, 17, 18, 17)

# Generar diagrama de cajas
boxplot(x)
```

7. Ejercicios

Ejercicio 4

Utilizar los datos sobre síntomas reportados por pacientes que se muestran a continuación para resolver las siguientes cuestiones:

```
sintomas <- c("DC", "F", "T", "DC", "DM", "T", "F", "DC", "T", "DG",  
"DM", "F", "DC", "T")
```

Abreviaturas: DC = Dolor de cabeza, DM = Dolor muscular, F = Fiebre, T = Tos, DG = Dolor de garganta.

- ¿Qué tipo de variable es `sintomas`?
- Crear en R la tabla de frecuencias.
- Generar el gráfico de barras.
- Generar el gráfico de sectores.
- ¿En qué situaciones es más útil cada tipo de gráfico anterior?
- ¿Qué medidas descriptivas pueden calcularse con la variable? Calcularlas.

7. Ejercicios

Ejercicio 5

Emplear el siguiente conjunto de datos, que representa el número de medicamentos que recibe cada paciente en una visita a la farmacia, para resolver las siguientes cuestiones:

```
medicamentos <- c(0, 1, 2, 1, 3, 2, 0, 1, 4, 2, 1, 3, 0, 2, 1)
```

- a) ¿Qué tipo de variable es medicamentos?
- b) Crear en R la tabla de frecuencias.
- c) Generar el gráfico de barras.
- d) ¿Cuál es el valor más frecuente?
- e) Calcular la media y la desviación típica.

7. Ejercicios

Ejercicio 6

Emplear el siguiente conjunto de datos, que representa la duración en minutos de las consultas atendidas por los farmacéuticos durante un día en la farmacia, para resolver las siguientes cuestiones:

```
duracion <- c(4.2, 6.1, 5.5, 7.3, 5.9, 4.8, 6.9, 5.3, 6.7,  
4.1, 3.8, 6.4)
```

- ¿Qué tipo de variable es duracion?
- Crear en R la tabla de frecuencias con intervalos de 1 minuto.
- Generar el gráfico más adecuado para representar esta variable.

7. Ejercicios

Ejercicio 7

Emplear el siguiente conjunto de datos sobre el tiempo en minutos que tarda en completarse la preparación de diferentes tipos de medicamentos, para responder a las siguientes cuestiones:

```
tiempos <- c(20, 25, 40, 35, 42, 572, 32, 30, 55)
```

- ¿Qué tipo de medidas son la media y la mediana? ¿Cómo se calculan?
- Calcular la media y la mediana en R.
- Comparar los resultados anteriores. ¿A qué se deben las diferencias?

7. Ejercicios

Ejercicio 8

Utilizar el siguiente conjunto de datos, que representa la cantidad de unidades de medicamento dispensadas por día en una farmacia durante una semana, para resolver las siguientes cuestiones:

```
unidades <- c(50, 75, 60, 85, 70, 90, 55)
```

- Definir el rango y el rango intercuartílico y calcularlos en R.
- Definir la varianza y la desviación típica y calcularlas en R.
- ¿Qué ventajas tiene usar la desviación típica en lugar de la varianza?
- Calcular el CV y explicar cuándo puede resultar más útil.

7. Ejercicios

Ejercicio 9

Utilizar el conjunto de datos sobre el tiempo de respuesta en minutos de una serie de ensayos clínicos de un nuevo medicamento para responder las siguientes cuestiones:

```
tiempos <- c(12.3, 14.1, 15.5, 16.2, 17.0, 17.8, 18.4, 18.9, 19.3,  
19.7, 20.1, 21.6, 22.0, 40.5)
```

- Crear un diagrama de cajas en R.
- Analizar la forma de la distribución de los tiempos de respuesta.
- ¿Qué parte del diagrama refleja la dispersión de los datos?
- ¿Se identifican valores atípicos en el conjunto de datos?

7. Ejercicios

Ejercicio 10

Los siguientes datos representan las presiones arteriales sistólicas (en mmHg) de 20 mujeres seleccionadas al azar:

```
presion_sistolica <- c(110, 118, 120, 122, 124, 125, 128, 130, 132,  
134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 155)
```

- Calcular P_{70} , D_1 y Q_1 .
- Determinar la presión máxima correspondiente al 5% de las mujeres con menor presión.
- Calcular la presión que es superada por el 30% de las mujeres.
- Indicar entre qué dos valores de presión se encuentra el 40% central de los datos.

7. Ejercicios

Ejercicio 11

En un estudio farmacéutico sobre la eficacia de un nuevo medicamento, se registraron los resultados de cuatro lotes de pacientes tratados con distintas dosis:

Lote	A	B	C	D
Eficacia media (%)	60	78	52	36
Varianza	9.2	13.9	8.1	6.4

- Sabiendo que en R existe la función `sqrt()` para calcular raíces cuadradas, calcular el coeficiente de variación de cada lote.
- Determinar qué lote muestra mayor homogeneidad en la eficacia del tratamiento.

Parte III: Regresión

1. Análisis de dos variables

En algunos estudios, se miden **dos variables**, X e Y , en cada individuo.

Ejemplo: Datos de **horas de trabajo** en un laboratorio de análisis clínicos (X) y número de **análisis realizados** (Y) por distintos técnicos:

x_i	y_i
5	15
32	18
60	35
75	65

- Cada fila de la tabla representa un individuo.
- Tenemos 4 individuos: $n = 4$.
- El primer individuo trabajó 5 horas y realizó 15 análisis.

Preguntas clave:

- ¿Existe alguna relación entre las variables X e Y ?
- Si la hay, ¿podemos usar una variable para estimar la otra?

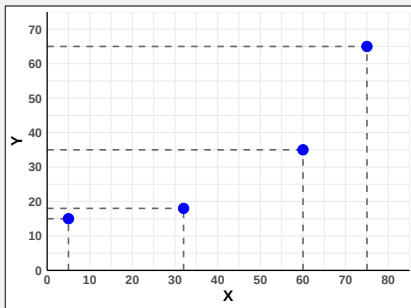
2. Nube de puntos

Definición: Gráfico que representa cada individuo mediante un punto (x_i, y_i) , mostrando los valores de las variables X e Y .

Interpretación: Si los puntos siguen un patrón (una línea recta, una curva, etc.), esto indica que las variables X e Y podrían estar relacionadas.

Ejemplo: Datos de horas de trabajo (X) y análisis realizados (Y).

x_i	y_i
5	15
32	18
60	35
75	65



2. Nube de puntos

Cálculo en R:

- Se utiliza la función `plot()`.
- Recibe **dos parámetros**:
 - Dos **vectores numéricos**: uno para X y otro para Y .
 - El primero se representa en el eje **horizontal** y el segundo en el **vertical**.

Código de ejemplo:

```
# Crear vectores con los datos
x <- c(5, 32, 60, 75)
y <- c(15, 18, 35, 65)

# Generar nube de puntos
plot(x, y)
```

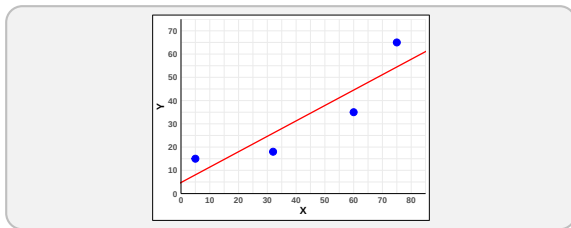
3. Regresión lineal

Modelo: Representación simplificada de la **realidad**. En **matemáticas**, los modelos se expresan generalmente mediante **ecuaciones**.

Regresión lineal: Permite encontrar un modelo (ecuación) que describa la relación entre las variables X e Y , reflejando el patrón observado en los datos. Se expresa mediante la ecuación de una recta:

$$y = ax + b, \quad \text{con } a, b \in \mathbb{R}$$

donde y es el valor predicho de Y a partir de un valor dado x de X .

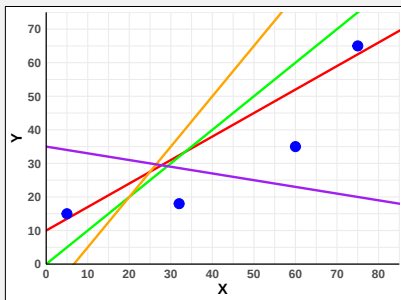


Objetivo: Predecir el valor de la variable **dependiente** Y a partir de un valor conocido de la variable **independiente** X .

3. Regresión lineal

a) Cálculo de la mejor recta de regresión

En la ecuación $y = ax + b$, los parámetros a y b pueden tomar infinitos valores, lo que genera infinitas rectas posibles.



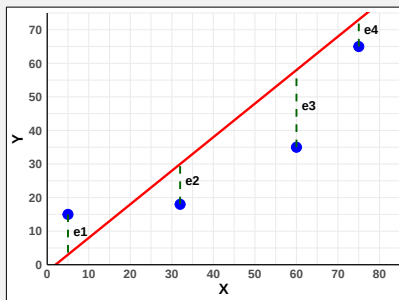
El objetivo es encontrar la **mejor recta**, es decir, aquella que **se acerque lo más posible** a todos los puntos simultáneamente.

3. Regresión lineal

a) Cálculo de la mejor recta de regresión

Para medir el ajuste en cada punto, definimos el **residuo** e_i como la diferencia entre el valor observado y_i y el valor estimado por la recta, $\hat{y}_i = ax_i + b$:

$$e_i = y_i - \hat{y}_i$$



Buscamos la recta (los valores de a y b) que minimice los residuos en conjunto.

3. Regresión lineal

a) Cálculo de la mejor recta de regresión

Una forma inicial de hallar la recta es minimizar la suma de los residuos:

$$\min \sum e_i = \min \sum (y_i - \hat{y}_i)$$

Sin embargo, los residuos positivos y negativos pueden cancelarse entre sí.

Para evitar esto, se utiliza la **suma de los cuadrados** de los residuos, lo que asegura que todos los términos sean positivos:

$$\min \sum e_i^2 = \min \sum (y_i - \hat{y}_i)^2 = \min \sum (y_i - (ax_i + b))^2$$

Este procedimiento para determinar los valores de a y b se conoce como el **Método de los Mínimos Cuadrados**, y es la técnica estándar para obtener la recta de regresión.

3. Regresión lineal

a) Cálculo de la mejor recta de regresión

Obtención de los parámetros a y b :

- Se aplica derivación parcial respecto a a y b y se iguala a cero.
- Esto genera un sistema de ecuaciones con incógnitas a y b .
- Al resolverlo, se obtienen valores únicos de a y b que ajustan la recta de forma óptima a los datos. Las soluciones son:

$$a = \frac{s_{xy}}{s_x^2}$$

$$b = \bar{y} - a\bar{x}$$

donde

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

es la covarianza entre X e Y , y s_x^2 es la varianza de X .

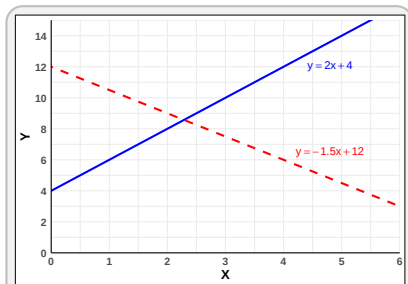
3. Regresión lineal

b) Coeficientes de la recta de regresión

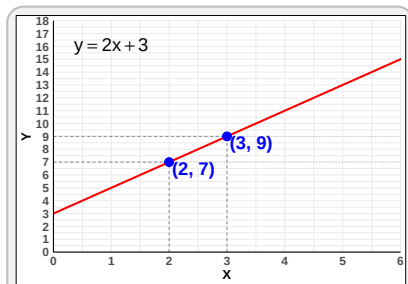
En la recta $y = ax + b$, tenemos dos coeficientes:

Pendiente (a):

- Indica si la recta es **creciente** ($a > 0$) o **decreciente** ($a < 0$).
- Representa el cambio en y cuando x aumenta en una unidad.



Ejemplo: Recta creciente ($y = 2x + 4$) y recta decreciente ($y = -1.5x + 12$).



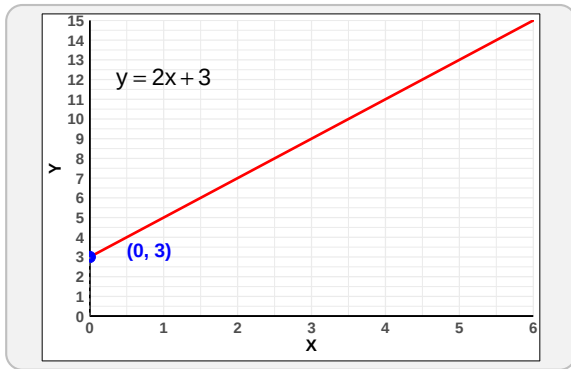
Ejemplo: Incremento en y cuando x aumenta en 1 en la recta $y = 2x + 3$.

3. Regresión lineal

b) Coeficientes de la recta de regresión

Término independiente (b):

- Valor de y en la recta $y = ax + b$ cuando $x = 0$ (corte con el eje Y).



3. Regresión lineal

b) Coeficientes de la recta de regresión

Cálculo en R:

Para crear el modelo, se emplea la función `lm()`.

- Recibe como parámetro una **fórmula** de la forma $y \sim x$, donde:
 - y es la variable dependiente y x la independiente.
- Devuelve como salida los **coeficientes de la recta de regresión**:
 - El valor (Intercept) corresponde al **término independiente** (b).
 - El coeficiente de x corresponde a la **pendiente** (a).

Código de ejemplo:

```
# Crear vectores con los datos
x <- c(5, 32, 60, 75)
y <- c(15, 18, 35, 65)
# Ajustar modelo y obtener coeficientes
lm(y~x)
```

Resultado: pendiente $a = 0.6626$, término independiente $b = 4.7576$

3. Regresión lineal

b) Coeficientes de la recta de regresión

Para representar los datos y la recta de regresión en R se siguen estos pasos:

1. Para graficar los datos, se emplea la función `plot()`, colocando la variable independiente en el eje horizontal y la dependiente en el eje vertical.
2. Para agregar la recta de regresión, se emplea la función `abline()`, que recibe como parámetros el término independiente (**b**) y la pendiente (**a**).

Código de ejemplo:

```
# Graficar los datos
plot(x, y)

# Dibujar la recta (b = 4.7576, a = 0.6626)
abline(4.7576, 0.6626)
```


3. Regresión lineal

c) Predicción de valores de una variable conocida la otra

El modelo de regresión lineal permite **predecir** el valor de una variable dependiente a partir de una variable independiente.

Dadas dos variables X e Y , se pueden calcular dos rectas de regresión: una para predecir Y a partir de X y otra para predecir X a partir de Y .

Característica	Recta de Y sobre X	Recta de X sobre Y
Ecuación	$y = ax + b$	$x = a'y + b'$
Variable dependiente	Y	X
Variable independiente	X	Y
Predicción	Predice Y dado X	Predice X dado Y
Pendiente	$a = \frac{s_{xy}}{s_x^2}$	$a' = \frac{s_{xy}}{s_y^2}$
Término independiente	$b = \bar{y} - a\bar{x}$	$b' = \bar{x} - a'\bar{y}$

3. Regresión lineal

c) Predicción de valores de una variable conocida la otra

Cálculo en R:

- Tras extraer los coeficientes, se puede calcular el valor de **y** para un valor dado de **x** aplicando la ecuación de la recta $y = a \cdot x + b$

Código de ejemplo (predecir el valor de y cuando $x = 10$):

```
# Recta de regresión obtenida con lm()
# y = 0.6626 * x + 4.7576

# Aplicar la ecuación de la recta
0.6626 * 10 + 4.7576
```

Resultado: 11.384

4. Medidas de relación y ajuste lineal

a) Coeficiente de correlación lineal de Pearson (r)

Utilidad: Permite conocer las características (**sentido** y **fuerza**) de la relación lineal entre dos variables cuantitativas.

Fórmula:

$$r = \frac{s_{xy}}{s_x \cdot s_y} \in [-1, 1]$$

Interpretación:

1. **Sentido de la relación lineal:**

- **Directo** ($r > 0$): Ambas variables aumentan o disminuyen juntas.
- **Inverso** ($r < 0$): Al aumentar una variable, la otra disminuye.

2. **Fuerza de la relación lineal:**

- $|r| \geq 0.8$: Relación lineal **muy fuerte**.
- $|r| \in [0.5, 0.8)$: Relación lineal **moderada**.
- $|r| < 0.5$: Relación lineal **débil**.
- $r = 0$: **No hay relación lineal**.

4. Medidas de relación y ajuste lineal

b) Coeficiente de determinación (R^2)

Utilidad: Evalúa qué tan bien el modelo de regresión se ajusta a los datos y permite hacer predicciones fiables.

Fórmula (en regresión lineal):

$$R^2 = r^2 \in [0, 1]$$

Interpretación (proporción de variabilidad explicada):

R^2 indica la proporción de la variabilidad de Y explicada por su relación lineal con X según el modelo de regresión lineal. **Ejemplo:** Si $R^2 = 0.755$, entonces el 75.5% de la variabilidad de Y se atribuye a los cambios en X .

- Si R^2 está cerca de 1, el modelo explica gran parte de la variabilidad de los datos, lo que implica un buen ajuste y predicciones fiables.
- Si R^2 está cerca de 0, el modelo explica poca variabilidad, indicando un mal ajuste y predicciones poco fiables.

4. Medidas de relación y ajuste lineal

Cálculo en R:

- Para calcular el **coeficiente de correlación**, se emplea la función `cor()`, que recibe como parámetros los vectores de datos de las variables X e Y .
- Para calcular el **coeficiente de determinación**, se eleva al cuadrado el coeficiente de correlación.

Código de ejemplo:

```
# Calcular el coeficiente de correlación
cor(x, y)

# Calcular el coeficiente de determinación
cor(x, y)^2
```

5. Otros modelos de regresión

Limitaciones del ajuste lineal: En algunos casos, los datos no se ajustan bien a la recta de regresión estándar: $y = ax + b$

Modelos alternativos:

Existen otros modelos de regresión que permiten capturar distintos patrones de relación entre variables. Entre los más comunes se encuentran:

- **Regresión forzada al origen:** Cuando se espera que la variable dependiente sea cero si la independiente es cero.
- **Regresión parabólica:** Para relaciones curvilíneas con máximos o mínimos dentro del rango de los datos.
- **Regresión hiperbólica:** Cuando la influencia de la variable independiente sobre la dependiente disminuye a medida que la independiente aumenta.

5. Otros modelos de regresión

a) Regresión forzada al origen

En la **regresión forzada al origen**, el modelo se expresa como:

$$y = ax$$

con $b = 0$, por lo que la recta pasa por el origen.

Casos de uso:

- Cuando se espera que $y = 0$ si $x = 0$.
- Cuando b no tiene sentido o se busca simplificar el modelo.

Utilidad: Reduce parámetros y facilita la interpretación en contextos concretos.

5. Otros modelos de regresión

a) Regresión forzada al origen

Cálculo en R:

Para crear el modelo, se emplea la función `lm()`.

- Recibe como parámetro una **fórmula** de la forma $y \sim x + 0$.
- Devuelve la **pendiente de la recta de regresión** (a).

Código de ejemplo:

```
# Datos de ejemplo
x <- c(0, 1, 2, 3, 4)
y <- c(0.2, 2.0, 4.1, 5.2, 8.1)

# Regresión forzada al origen
lm(y~x+0)
```

Resultado: pendiente $a = 1.94$

5. Otros modelos de regresión

b) Regresión parabólica

Si los datos siguen una relación curvilínea, pueden ajustarse con una parábola.

Para una **parábola de segundo grado**, el modelo se expresa como:

$$y = ax^2 + bx + c$$

Casos de uso:

- Cuando la relación entre X e Y no es lineal y se observa curvatura.
- Para fenómenos con máximo o mínimo dentro del rango de datos.

Utilidad: Permite capturar la curvatura en los datos, ofreciendo un ajuste más preciso que una recta simple.

5. Otros modelos de regresión

b) Regresión parabólica

Cálculo en R:

Para crear el modelo, se emplea la función `lm()`.

- Recibe como parámetro una **fórmula** de la forma $y \sim x + I(x^2)$.
- Devuelve los coeficientes a, b, c de la parábola.

Código de ejemplo:

```
# Datos de ejemplo
x <- c(0, 1, 2, 3, 4)
y <- c(1.0, 2.9, 6.1, 11.2, 17.8)

# Ajustar modelo cuadrático
lm(y ~ x + I(x^2))
```

Resultado: $a = 0.8071, b = 0.9614, c = 1.0343$

5. Otros modelos de regresión

c) Regresión hiperbólica

Cuando el efecto de x sobre y disminuye a medida que x aumenta, puede emplearse un **modelo hiperbólico**:

$$y = \frac{a}{x} + b$$

Casos de uso:

- Cuando el efecto de x sobre y disminuye a medida que x aumenta.
- En fenómenos de saturación o rendimientos decrecientes.

Utilidad: Permite modelar relaciones inversas entre las variables, donde un incremento de x produce una disminución progresiva en y .

5. Otros modelos de regresión

c) Regresión hiperbólica

Cálculo en R:

Para crear el modelo, se emplea la función `lm()`.

- Recibe como parámetro una **fórmula** de la forma $y \sim I(1/x)$.
- Devuelve los coeficientes a y b del modelo hiperbólico.

Código de ejemplo:

```
# Datos de ejemplo
x <- c(1, 2, 3, 4, 5)
y <- c(5.0, 3.2, 2.4, 2.0, 1.7)

# Ajustar modelo hiperbólico
lm(y~I(1/x))
```

Resultado: $a = 4.062$, $b = 1.005$

6. Ejercicios

Ejercicio 12

Se dispone de los siguientes datos que relacionan la concentración de un fármaco (en mg) con su eficacia terapéutica, medida en una escala de 0 a 30:

```
concentracion <- c(1.2, 2.7, 3.5, 5.8, 6.9, 6.6, 8.6, 8.9, 9.7)
eficacia <- c(5, 7, 10, 12, 15, 18, 20, 25, 27)
```

- Representar en R la nube de puntos.
- Calcular la ecuación de la recta de regresión lineal que permite predecir la eficacia del fármaco en función de su concentración.
- Estimar la eficacia correspondiente a una concentración de 4.5 mg.

6. Ejercicios

Ejercicio 13

Se dispone de los siguientes datos que relacionan el pH de una solución con la estabilidad del medicamento (en horas):

```
ph <- c(2, 3, 5, 6, 7, 8, 9)
estabilidad <- c(20, 25, 30, 35, 40, 38, 30)
```

- Representar en R la nube de puntos.
- Utilizando un modelo de regresión lineal, calcular la estabilidad del medicamento cuando el pH de la solución es de 4.

6. Ejercicios

Ejercicio 14

Se dispone de los siguientes datos que relacionan la dosis de un medicamento (en mg) con la frecuencia cardíaca medida en pulsaciones por minuto:

```
dosis <- c(10, 20, 30, 40, 50, 60, 70, 80, 90, 100)
frec_card <- c(70, 72, 75, 77, 79, 81, 83, 85, 86, 87)
```

- Ajustar un modelo de regresión lineal que permita predecir la frecuencia cardíaca en función de la dosis del medicamento.
- Estimar la frecuencia cardíaca esperada cuando la dosis administrada es de 75 mg.
- Determinar la dosis que correspondería a una frecuencia cardíaca de 74 pulsaciones por minuto.

6. Ejercicios

Ejercicio 15

La siguiente tabla muestra la cantidad de horas de sueño y la eficacia de un tratamiento medida en una escala de 0 a 100:

```
horas <- c(4, 5, 6, 7, 8, 9, 10)
eficacia <- c(50, 55, 60, 65, 70, 75, 78)
```

- Crear la nube de puntos en R.
- Ajustar un modelo de regresión lineal que permita predecir la eficacia en función de las horas de sueño. Representar la recta obtenida.
- Evaluar la bondad del ajuste del modelo obtenido.
- Describir la relación lineal observada entre las variables.

6. Ejercicios

Ejercicio 16

Se dispone de los siguientes datos sobre la dosis administrada de un fármaco (en mg) y la reducción de la presión arterial observada (en mmHg) en cinco pacientes:

Dosis	10	20	30	40	50
Reducción	5	8	12	14	17

- Determinar las dos rectas de regresión.
- Calcular la reducción esperada en la presión arterial para un paciente que recibe una dosis de 35 mg.

6. Ejercicios

Ejercicio 17

Se ha medido la concentración de un fármaco en sangre (en mg/L) y el efecto sobre la disminución de los niveles de glucosa (en mg/dL) en cinco pacientes:

Concentración	2	4	6	8	10
Reducción glucosa	8	15	24	30	38

Se sabe que, sin fármaco administrado, la reducción de glucosa es cero.

- Determinar el modelo de regresión lineal forzado a pasar por el origen para predecir la reducción de glucosa en función de la concentración.
- Predecir la reducción esperada de glucosa para un paciente con concentración de 7 mg/L.

6. Ejercicios

Ejercicio 18

Se ha medido la actividad enzimática (en unidades por litro) tras administrar distintas dosis de un nuevo fármaco (en mg) a 5 pacientes:

Dosis	0	2	4	6	8
Actividad	5	12	20	23	22

- Ajustar un modelo cuadrático para predecir la actividad enzimática en función de la dosis.
- Predecir la actividad para un paciente que reciba 5 mg de fármaco.

6. Ejercicios

Ejercicio 19

Se ha medido el efecto de un fármaco analgésico (en porcentaje de alivio del dolor) tras administrar distintas dosis (en mg) a cinco pacientes:

Dosis	1	2	4	6	8
Alivio	15	28	45	55	60

- Ajustar un modelo hiperbólico para predecir el alivio del dolor a partir de la dosis administrada.
- Predecir el alivio para un paciente que reciba 5 mg de fármaco.

6. Ejercicios

Ejercicio 20

Se ha medido la concentración plasmática de vitamina C (en mg/dL) tras administrar un suplemento (en mg) a cinco voluntarios:

Suplemento	50	100	200	300	400
Concentración	2	4	7	9	10

- Usando un modelo parabólico, predecir la concentración plasmática para una dosis de suplemento de 250 mg.
- Usando un modelo hiperbólico, estimar la dosis de suplemento necesaria para alcanzar una concentración de 8 mg/dL.

Bibliografía

Cobo, E., Muñoz, P., & González, J. A. (2007). *Bioestadística para no estadísticos: Bases para interpretar artículos científicos*. Elsevier-Masson.

Gómez Rubio, V., & López Cano, E. (2017). *Teoría y problemas resueltos de matemática aplicada y estadística para Farmacia*. Paraninfo.

Marín Morales, J., & Carrasco Ribelles, L. A. (2022). *Introducción a los análisis estadísticos en R*. Marcombo.

Martín Andrés, A., & Luna del Castillo, J. de D. (2004). *Bioestadística para las Ciencias de la Salud*. Norma-Capitel.

Rius, F., & Barón, F. J. (2008). *Bioestadística*. International Thomson.

Ross, S. (2007). *Introducción a la Estadística*. Reverté.

Sánchez, M., Frutos, G., & Cuesta, P. L. (1996). *Estadística y matemáticas aplicadas*. Síntesis.

Valderrama Bonnet, M. J., Ocaña Lara, F. A., & Ocaña Peinado, F. M. (2020). *Biometría: Grado en Farmacia* (2.^a ed.). Técnica Avicam.



UNIVERSIDAD
DE GRANADA



FACULTAD DE
FARMACIA



Manual de estadística para Biometría
con aplicaciones farmacéuticas en R