# Distribuciones bidimensionales

Dos variables x e y están relacionadas funcionalmente cuando conocida la primera se puede saber con exactitud el valor de la segunda.

***Ejemplo***

Si se deja caer una piedra, existe una fórmula que nos permite calcular exactamente, la altura a la que se encuentra en función del tiempo transcurrido.

$$h = \frac{1}{2} g t^2.$$

## Relación estadística

Dos variables x e y están relacionadas estadísticamente cuando conocida la primera se puede estimar aproximadamente el valor de la segunda.

***Ejemplos***

Ingresos y gastos de una familia.

Producción y ventas de una fábrica.

Gastos en publicidad y beneficios de una empresa.

## Variable estadística bidimensional

Una **variable bidimensional** es una variable en la que cada individuo está definido por un par de caracteres, **(X, Y)**.

Estos dos caracteres son a su vez **variables estadísticas** entre las que existe relación, una de las dos variables es la variable independiente y la otra variable dependiente.

## Distribuciones bidimensionales

Son aquellas en las que a cada individuo le corresponden los valores de dos variables, las representamos por el par $(x_i, y_i)$.

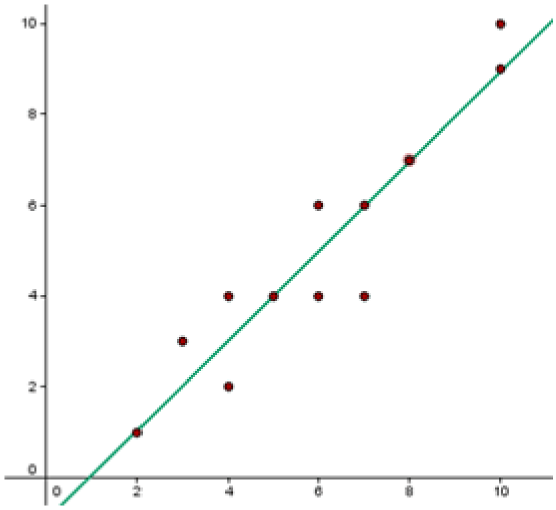
Si representamos cada par de valores como las coordenadas de un punto, el conjunto de todos ellos se llama **nube de puntos** o **diagrama de dispersión**.

Sobre la nube de puntos puede trazarse una recta que se ajuste a ellos lo mejor posible, llamada **recta de regresión**.

***Ejemplo***

Las notas de 12 alumnos de una clase en Matemáticas y Física son las siguientes:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Matemáticas** | **2** | **3** | **4** | **4** | **5** | **6** | **6** | **7** | **7** | **8** | **10** | **10** |
| **Física** | **1** | **3** | **2** | **4** | **4** | **4** | **6** | **4** | **6** | **7** | **9** | **10** |




# Covarianza

La **covarianza** de una variable bidimensional es la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas.

La **covarianza** se representa por $\mathbf{s_{xy}}$ o $\boldsymbol{\sigma_{xy}}$.

$$\sigma_{xy} = \frac{\sum f_i(x_i - \bar{x})(y_i - \bar{y})}{N}$$

$$\sigma_{xy} = \frac{\sum f_i x_i y_i}{N} - \bar{x}\bar{y}$$

La **covarianza** indica el sentido de la correlación entre las variables

Si $\boldsymbol{\sigma_{xy} > 0}$ la correlación es directa.

Si $\boldsymbol{\sigma_{xy} < 0}$ la correlación es inversa.

La **covarianza** presenta como inconveniente, el hecho de que su valor depende de la escala elegida para los ejes.

Es decir, la **covarianza** variará si expresamos la altura en metros o en centímetros. También variará si el dinero lo expresamos en euros o en dólares.

***Ejemplo 1:***

Las notas de 12 alumnos de una clase en Matemáticas y Física son las siguientes:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Matemáticas** | **2** | **3** | **4** | **4** | **5** | **6** | **6** | **7** | **7** | **8** | **10** | **10** |
| **Física** | **1** | **3** | **2** | **4** | **4** | **4** | **6** | **4** | **6** | **7** | **9** | **10** |

Hallar la **covarianza** de la distribución.

| $x_i$ | $y_i$ | $x_i \cdot y_i$ |
|---|---|---|
| **2** | **1** | **2** |
| **3** | **3** | **9** |
| **4** | **2** | **8** |
| **4** | **4** | **16** |
| **5** | **4** | **20** |
| **6** | **4** | **24** |
| **6** | **6** | **36** |
| **7** | **4** | **28** |
| **7** | **6** | **42** |
| **8** | **7** | **56** |
| **10** | **9** | **90** |
| **10** | **10** | **100** |
| **72** | **60** | **431** |

Después de tabular los datos hallamos las **medias aritméticas**:

$$\bar{x} = \frac{72}{12} = 6 \qquad \bar{y} = \frac{60}{12} = 5$$

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

***Ejemplo 2:***

Los valores de dos variables X e Y se distribuyen según la tabla siguiente:

| **Y/X** | **0** | **2** | **4** |
|---|---|---|---|
| **1** | **2** | **1** | **3** |
| **2** | **1** | **4** | **2** |
| **3** | **2** | **5** | **0** |

Hallar la **covarianza** de la distribución.

En primer lugar convertimos la tabla de doble entrada en tabla simple y calculamos las medias aritméticas.

| $x_i$ | $y_i$ | $f_i$ | $x_i \cdot f_i$ | $y_i \cdot f_i$ | $x_i \cdot y_i \cdot f_i$ |
|---|---|---|---|---|---|
| **0** | **1** | **2** | **0** | **2** | **0** |
| **0** | **2** | **1** | **0** | **2** | **0** |
| **0** | **3** | **2** | **0** | **6** | **0** |
| **2** | **1** | **1** | **2** | **1** | **2** |
| **2** | **2** | **4** | **8** | **8** | **16** |
| **2** | **3** | **5** | **10** | **15** | **30** |
| **4** | **1** | **3** | **12** | **3** | **12** |
| **4** | **2** | **2** | **8** | **4** | **16** |
| | | **20** | **40** | **41** | **76** |

$$\bar{x} = \frac{40}{20} = 2 \qquad \bar{y} = \frac{41}{20} = 2.05$$

$$\sigma_{xy} = \frac{76}{20} - 2 \cdot 2.05 = -0.3$$

# Correlación

La **correlación** trata de establecer la relación o dependencia que existe entre las dos variables que intervienen en una **distribución bidimensional**.

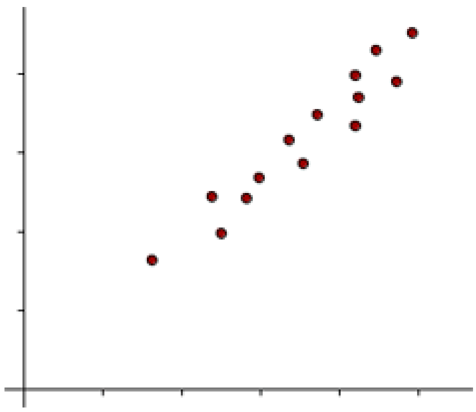
Es decir, determinar si los cambios en una de las variables influyen en los cambios de la otra. En caso de que suceda, diremos que las variables están correlacionadas o que hay **correlación** entre ellas.

## Tipos de correlación

### 1º Correlación directa

La correlación directa se da cuando al aumentar una de las variables la otra aumenta.

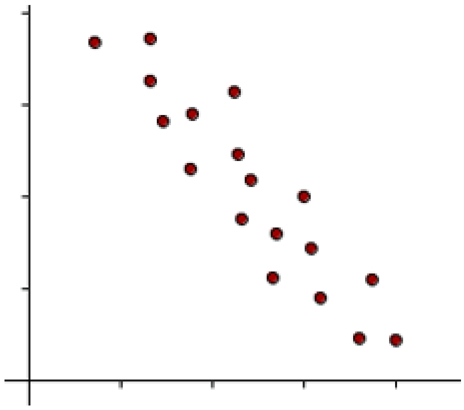
La recta correspondiente a la nube de puntos de la distribución es una recta creciente.



### 2º Correlación inversa

La correlación inversa se da cuando al aumentar una de las variables la otra disminuye.

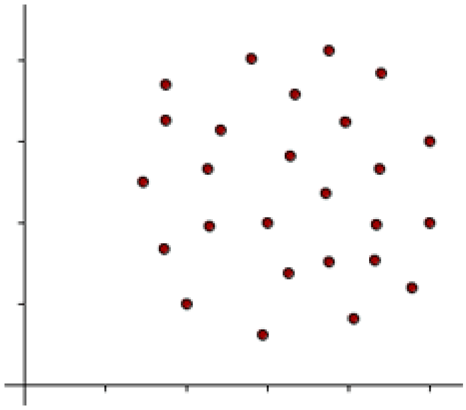
La recta correspondiente a la nube de puntos de la distribución es una recta decreciente.



**3º Correlación nula**

La correlación nula se da cuando no hay dependencia de ningún tipo entre las variables.

En este caso se dice que las variables son incorreladas y la nube de puntos tiene una forma redondeada.

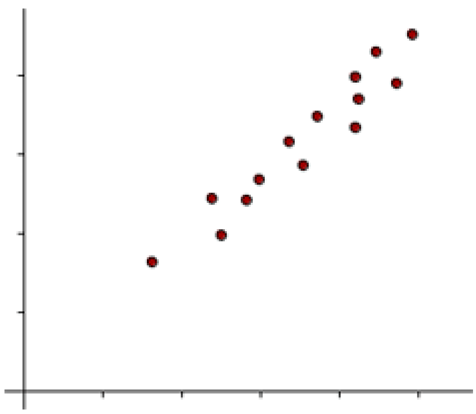


## Grado de correlación

El **grado de correlación** indica la proximidad que hay entre los puntos de la nube de puntos. Se pueden dar tres tipos:

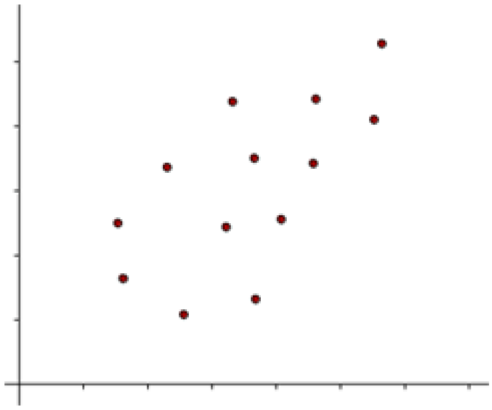
**1. Correlación fuerte**

La correlación será fuerte cuanto más cerca estén los puntos de la recta.



**2. Correlación débil**

La correlación será débil cuanto más separados estén los puntos de la recta.



**3. Correlación nula**

## Coeficiente de correlación lineal

El **coeficiente de correlación lineal** es el cociente entre la **covarianza** y el producto de las **desviaciones típicas** de ambas variables.

El **coeficiente de correlación lineal** se expresa mediante la letra **r**.

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

## Propiedades del coeficiente de correlación

**1.** El **coeficiente de correlación** no varía al hacerlo la escala de medición.

Es decir, si expresamos la altura en metros o en centímetros el coeficiente de correlación no varía.

**2.** El signo del **coeficiente de correlación** es el mismo que el de la **covarianza**.

Si la covarianza es positiva, la correlación es directa.

Si la covarianza es negativa, la correlación es inversa.

Si la covarianza es nula, no existe correlación.

**3.** El **coeficiente de correlación lineal** es un número real comprendido entre −1 y 1.

$$-1 \le r \le 1$$

**4.** Si el **coeficiente de correlación lineal** toma valores cercanos a −1 la correlación es **fuerte e inversa**, y será tanto más fuerte cuanto más se aproxime r a −1.

**5.** Si el **coeficiente de correlación lineal** toma valores cercanos a 1 la correlación es **fuerte y directa**, y será tanto más fuerte cuanto más se aproxime r a 1.

**6.** Si el **coeficiente de correlación lineal** toma valores cercanos a 0, la correlación es **débil**.

**7.** Si r = 1 ó −1, los puntos de la nube están sobre la recta creciente o decreciente. Entre ambas variables hay **dependencia funcional**.

***Ejemplo 1:*** Las notas de 12 alumnos de una clase en Matemáticas y Física son las siguientes:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Matemáticas** | **2** | **3** | **4** | **4** | **5** | **6** | **6** | **7** | **7** | **8** | **10** | **10** |
| **Física** | **1** | **3** | **2** | **4** | **4** | **4** | **6** | **4** | **6** | **7** | **9** | **10** |

Hallar el **coeficiente de correlación** de la distribución e interpretarlo.

| $x_i$ | $y_i$ | $x_i \cdot y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|
| 2 | 1 | 2 | 4 | 1 |
| 3 | 3 | 9 | 9 | 9 |
| 4 | 2 | 8 | 16 | 4 |
| 4 | 4 | 16 | 16 | 16 |
| 5 | 4 | 20 | 25 | 16 |
| 6 | 4 | 24 | 36 | 16 |
| 6 | 6 | 36 | 36 | 36 |
| 7 | 4 | 28 | 49 | 16 |
| 7 | 6 | 42 | 49 | 36 |
| 8 | 7 | 56 | 64 | 49 |
| 10 | 9 | 90 | 100 | 81 |
| 10 | 10 | 100 | 100 | 100 |
| 72 | 60 | 431 | 504 | 380 |

**1º** Hallamos las **medias aritméticas**.

$$\bar{x} = \frac{72}{12} = 6 \qquad \bar{y} = \frac{60}{12} = 5$$

**2º** Calculamos la **covarianza**.

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

**3º** Calculamos las **desviaciones típicas**.

$$\sigma_x = \sqrt{\frac{504}{12} - 6^2} = 2.45 \qquad \sigma_y = \sqrt{\frac{380}{12} - 25} = 2.58$$

**4º** Aplicamos la fórmula del **coeficiente de correlación lineal**.

$$r = \frac{5.92}{2.45 \cdot 2.58} = 0.94$$

Al ser el **coeficiente de correlación** positivo, la correlación es directa.

Como **coeficiente de correlación** está muy próximo a 1 la correlación es muy fuerte.

Los valores de dos variables X e Y se distribuyen según la tabla siguiente:

| **Y/X** | **0** | **2** | **4** |
|---|---|---|---|
| **1** | **2** | **1** | **3** |
| **2** | **1** | **4** | **2** |
| **3** | **2** | **5** | **0** |

Determinar el **coeficiente de correlación**.

Convertimos la tabla de doble entrada en tabla simple.

| $x_i$ | $y_i$ | $f_i$ | $x_i \cdot f_i$ | $x_i^2 \cdot f_i$ | $y_i \cdot f_i$ | $y_i^2 \cdot f_i$ | $x_i \cdot y_i \cdot f_i$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 0 | 0 | 2 | 2 | 0 |
| 0 | 2 | 1 | 0 | 0 | 2 | 4 | 0 |
| 0 | 3 | 2 | 0 | 0 | 6 | 18 | 0 |
| 2 | 1 | 1 | 2 | 4 | 1 | 1 | 2 |
| 2 | 2 | 4 | 8 | 16 | 8 | 16 | 16 |
| 2 | 3 | 5 | 10 | 20 | 15 | 45 | 30 |
| 4 | 1 | 3 | 12 | 48 | 3 | 3 | 12 |
| 4 | 2 | 2 | 8 | 32 | 4 | 8 | 16 |
| | | 20 | 40 | 120 | 41 | 97 | 76 |

$$\bar{x} = \frac{40}{20} = 2 \qquad \bar{y} = \frac{41}{20} = 2.05$$

$$\sigma_x^2 = \frac{120}{20} - 2^2 = 2 \qquad \sigma_y^2 = \frac{97}{20} - 2.05^2 = 0.65$$

$$\sigma_x = \sqrt{2} = 1.41 \qquad \sigma_y = \sqrt{0.65} = 0.81$$

$$\sigma_{xy} = \frac{76}{20} - 2 \cdot 2.05 = -0.3$$

$$r = \frac{-0.3}{1.41 \cdot 0.81} = -0.26$$

Al ser el **coeficiente de correlación** negativo, la correlación es inversa.

Como **coeficiente de correlación** está muy próximo a 0 la correlación es muy débil.

# Recta de regresión

La **recta de regresión** es la que mejor se ajusta a la **nube de puntos**.

La **recta de regresión** pasa por el punto $(\bar{x}, \bar{y})$ llamado **centro de gravedad**.

### Recta de regresión de Y sobre X

La recta de regresión de Y sobre X se utiliza para estimar los valores de la Y a partir de los de la X.

La **pendiente** de la recta es el cociente entre la covarianza y la varianza de la variable X.

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

### Recta de regresión de X sobre Y

La **recta de regresión** de X sobre Y se utiliza para estimar los valores de la X a partir de los de la Y.

La **pendiente** de la recta es el cociente entre la covarianza y la varianza de la variable Y.

$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

Si la correlación es nula, r = 0, las rectas de regresión son perpendiculares entre sí, y sus ecuaciones son:

$y = \bar{y}$

$x = \bar{x}$

***Ejemplo***

Las notas de 12 alumnos de una clase en Matemáticas y Física son las siguientes:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Matemáticas** | **2** | **3** | **4** | **4** | **5** | **6** | **6** | **7** | **7** | **8** | **10** | **10** |
| **Física** | **1** | **3** | **2** | **4** | **4** | **4** | **6** | **4** | **6** | **7** | **9** | **10** |

Hallar las **rectas de regresión** y representarlas.

| $x_i$ | $y_i$ | $x_i \cdot y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|
| 2 | 1 | 2 | 4 | 1 |
| 3 | 3 | 9 | 9 | 9 |
| 4 | 2 | 8 | 16 | 4 |
| 4 | 4 | 16 | 16 | 16 |
| 5 | 4 | 20 | 25 | 16 |
| 6 | 4 | 24 | 36 | 16 |
| 6 | 6 | 36 | 36 | 36 |
| 7 | 4 | 28 | 49 | 16 |
| 7 | 6 | 42 | 49 | 36 |
| 8 | 7 | 56 | 64 | 49 |
| 10 | 9 | 90 | 100 | 81 |
| 10 | 10 | 100 | 100 | 100 |
| 72 | 60 | 431 | 504 | 380 |

**1º** Hallamos las **medias aritméticas**.

$$\bar{x} = \frac{72}{12} = 6 \qquad \bar{y} = \frac{60}{12} = 5$$

**2º** Calculamos la **covarianza**.

$$\sigma_{xy} = \frac{431}{12} - 6 \cdot 5 = 5.92$$

**3º** Calculamos las **varianzas**.

$$\sigma_x^2 = \frac{504}{12} - 6^2 = 6 \qquad \sigma_y^2 = \frac{380}{12} - 25 = 6.66$$

**4º**Recta de regresión de Y sobre X.

$$y - 5 = \frac{5.92}{6}(x - 6) \qquad y = 0.987x - 0.922$$

**4º**Recta de regresión de X sobre Y.

$$x - 6 = \frac{5.92}{6.66}(y - 5) \qquad x = 0.889y + 1.556$$