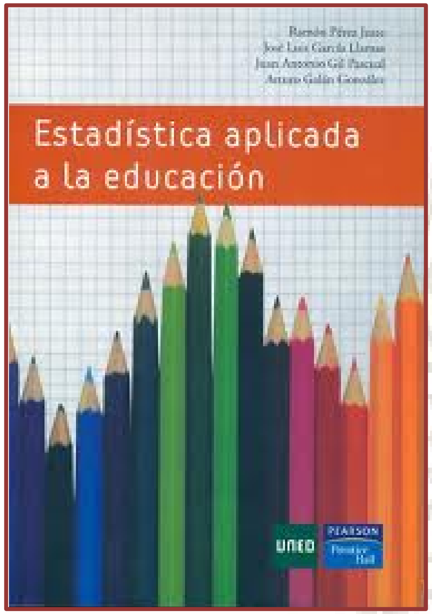
# Estadística Aplicada a la Educación

# Tema 7

Tutor.
**UNED Madrid-Sur (A.U. Parla)**
*Miguel Ángel Daza*
migdaza@madridsur.uned.es

# Estadística Aplicada a la Educación

Índice






# 7. RELACIÓN ENTRE VARIABLES. LAS CORRELACIONES. LA REGRESIÓN.


VIDEOCLASE: Estadística bivariada: Las Medidas de relación.

https://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=56932&ID_Sala=60549&hashData=e7ea4832ba6abb091775a0a59785f696

## 7.1 Introducción.

Al hacer investigación en educación nos interesará en muchas ocasiones **conocer la posible relación entre 2 o más variables.** (por ejemplo X e Y)

En este capítulo estudiaremos:
Las relaciones entre las variables que intervienen en el proceso educativo: **correlación**
Las posibilidades y limitaciones de la predicción de puntuaciones en una variable, conociendo los valores de otra: **regresión**

Estudiaremos los coeficientes de correlación más importantes en el campo educativo:

El coeficiente de **correlación de Pearson** ***(r)***

El coeficiente de **correlación ordinal de Spearman** $(r_s)$

El **coeficiente de contingencia** ***(C)***

El coeficiente de **correlación biserial-puntual** $(r_{bp})$

## 7.2 El concepto de correlación.

La **correlación** nos indica la **tendencia de dos o más conjuntos de datos a variar de forma conjunta**. Para cuantificar la **intensidad** de la correlación contamos con el **coeficiente de correlación** que nos mide el **índice de covariación o variación conjunta** de dos o más series de datos

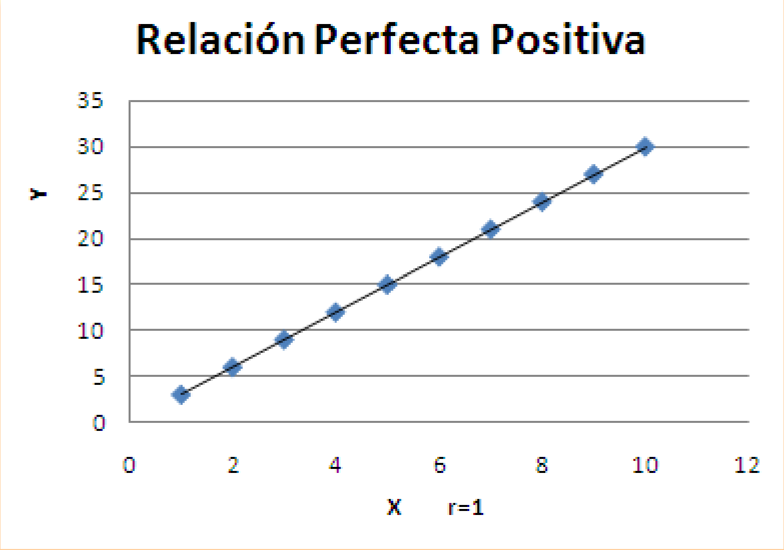
Por ejemplo, podemos estar interesados en conocer qué variables están correlacionadas con el rendimiento académico: tiempo de estudio, motivación, nivel de uso de los recursos tecnológicos, etc... De esta manera, podemos decidir sobre qué variables intervenir para mejorar dicho rendimiento.

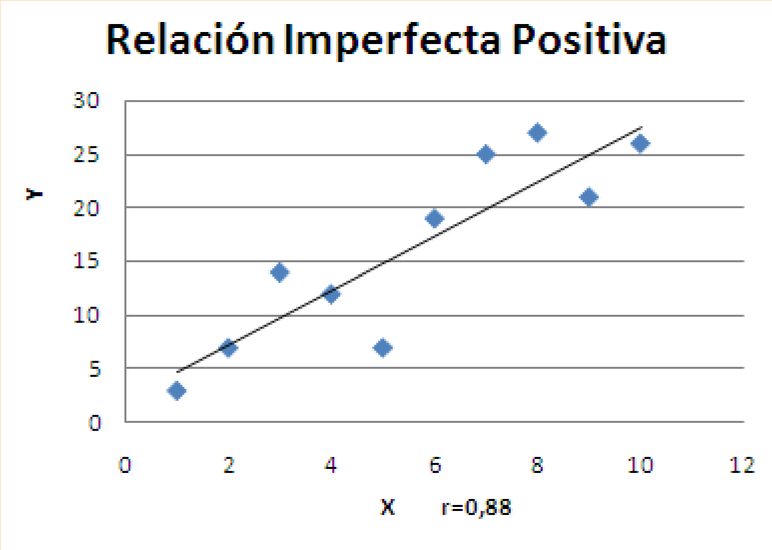
Las situaciones susceptibles de ser analizadas de manera correlacional son esencialmente:

- Estudiar la relación entre 2 o más variables medidas en un mismo grupo de sujetos (lo más habitual)
- Estudiar la relación entre 2 o más grupos de individuos en una sola variable.
- Estudiar el caso de una misma variable medida en dos momentos distintos en una misma muestra (p.ej.: fiabilidad como estabilidad)

## 7.2 El concepto de correlación.

En el análisis de **la correlación entre 2 variables**, teniendo en cuenta la **intensidad** y el **sentido de la relación**, se pueden dar las siguientes situaciones ilustradas mediante un **diagrama de dispersión**.



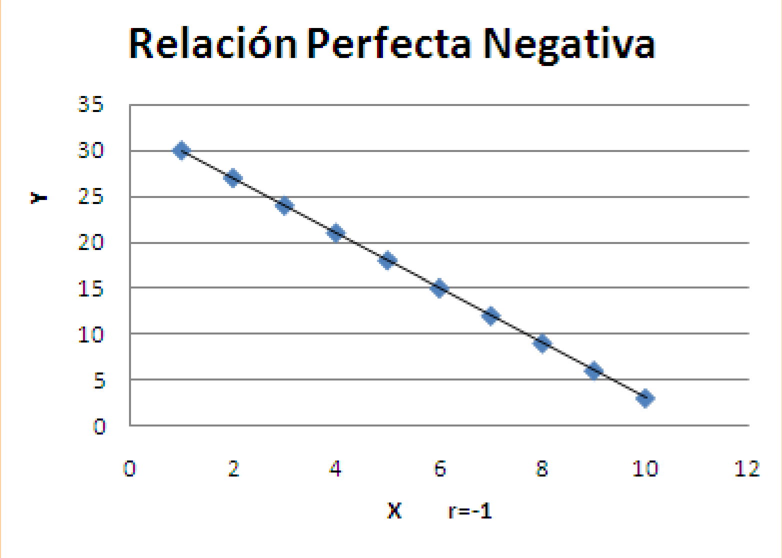



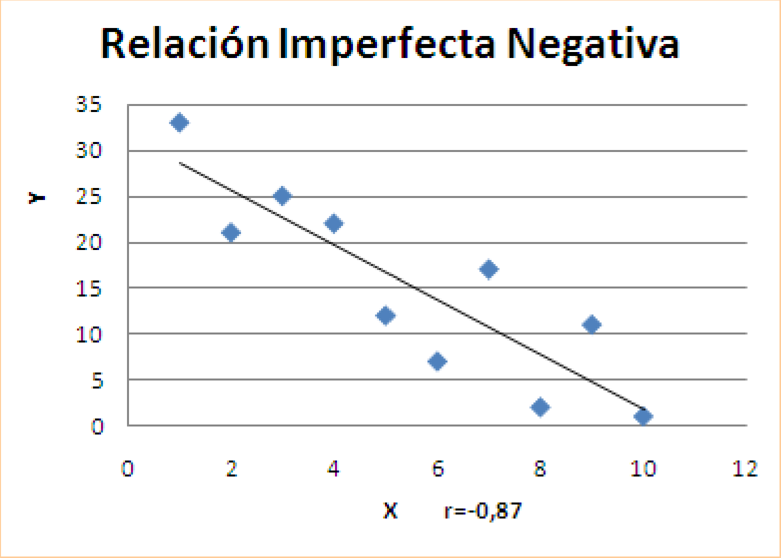

**RELACIÓN PERFECTA POSITIVA** ("función") Valor cuantitativo = **+1**

**RELACIÓN IMPERFECTA POSITIVA** (habitual en el campo educativo) Valor cuantitativo = **entre 0 y +1**

# 7.2 El concepto de correlación.

En el análisis de **la correlación entre 2 variables**, teniendo en cuenta la **intensidad** y el **sentido de la relación**, se pueden dar las siguientes situaciones ilustradas mediante un **diagrama de dispersión**.



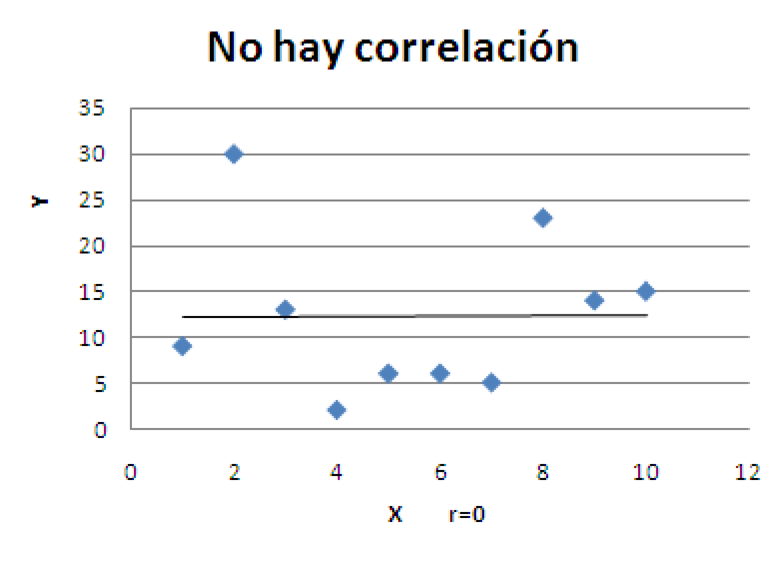




**RELACIÓN PERFECTA NEGATIVA** ("función") Valor cuantitativo = **-1**

**RELACIÓN IMPERFECTA NEGATIVA** (habitual en el campo educativo) Valor cuantitativo = **entre 0 y -1**

# 7.2 El concepto de correlación.

En el análisis de **la correlación entre 2 variables**, teniendo en cuenta la **intensidad** y el **sentido de la relación**, se pueden dar las siguientes situaciones ilustradas mediante un **diagrama de dispersión**.




**RELACIÓN NULA O AUSENCIA DE RELACIÓN** ("Esta ausencia de relación se da cuando dos variables son independientes una de la otra") Valor cuantitativo = **0**

# 7.3 El coeficiente de correlación simple y su interpretación.

**El coeficiente de correlación** nos mide el valor de la covariación o variación conjunta de dos series de datos. El valor de ese coeficiente nos marca la existencia de una relación directa de variables (valores positivos) o inversa (valores negativos).
Es decir, **los valores cuantitativos del coeficiente se sitúan entre +1 y -1**

**¿Cómo interpretar los valores de esos coeficientes?**

1. **El tipo de variables que se relacionan:** ¿hay estudios anteriores que indiquen correlaciones similares?
2. **La variabilidad del grupo:** la correlación tiende a aumentar con la variabilidad del grupo. Ante 2 coeficientes de correlación iguales, el alcanzado en el grupo más homogéneo es de más intensidad.
3. **La finalidad a la que se destina el coeficiente:** fiabilidad (>0,85) o validez (>0,60)

# 7.3 El coeficiente de correlación simple y su interpretación.

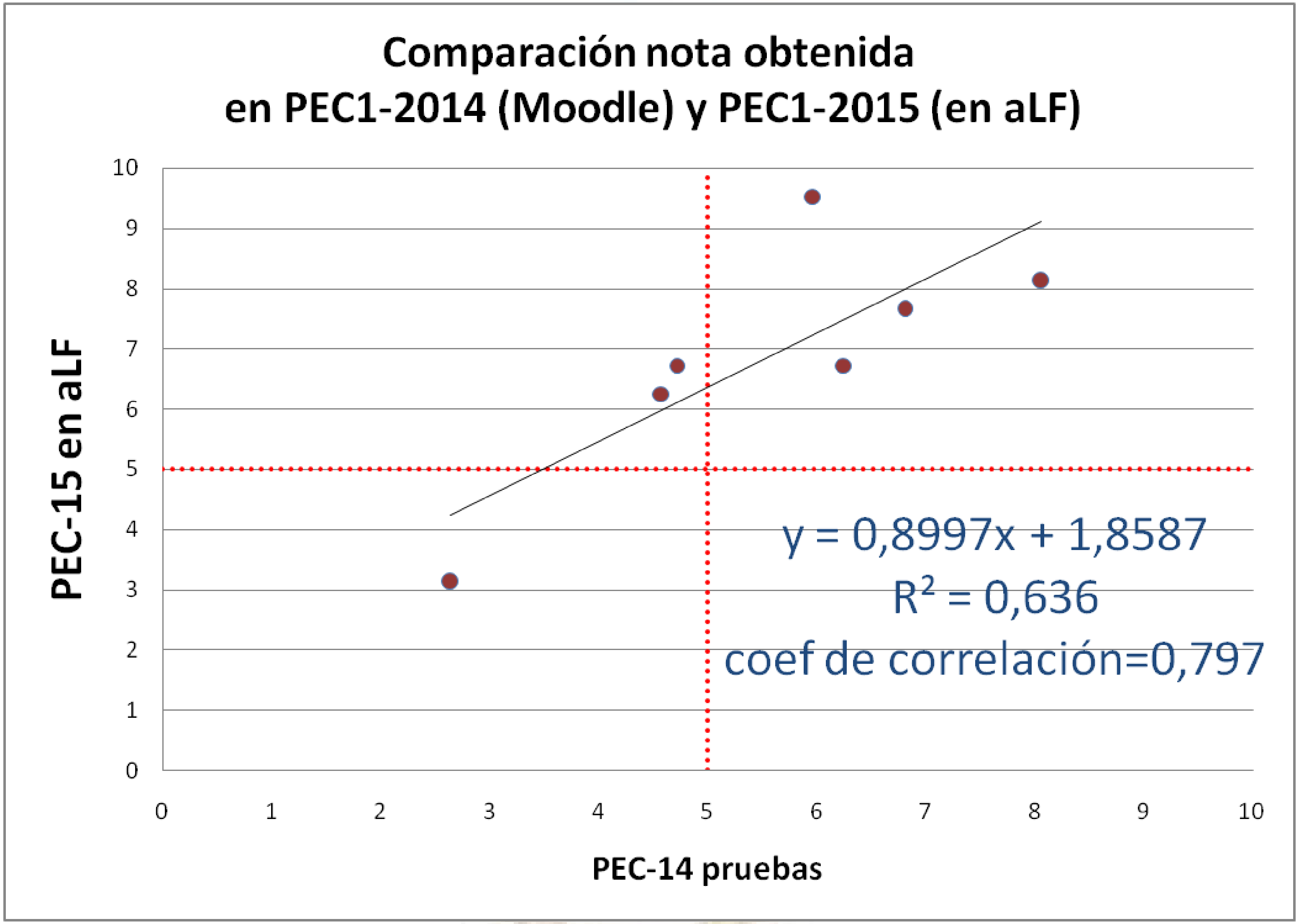
| Valor del coeficiente | Interpretación |
|---|---|
| Entre 0,00 y + o - 0,20 | Correlación muy baja, indifefrente, despreciable. |
| Entre 0,21 y + o - 0,40 | Correlación baja. |
| Entre 0,41 y + o - 0,70 | Correlación media, marcada, notable. |
| Entre 0,71 y + o - 0,90 | Correlación alta, elevada, fuerte |
| Entre 0,91 y + o - 1 | Correlación muy alta, muy elevada |

**Coeficiente de Determinación (d)** = Coeficiente de correlación (r) al cuadrado por 100

$$d = r^2 * 100$$

(Se interpreta como el porcentaje de la varianza de una variable explicada por la otra)

La interpretación de los coeficientes de correlación se suele completar con su **significación estadística**. Se trata de poder afirmar que la correlación entre 2 variables es real y no se puede explicar por efecto del azar.


Comparación nota obtenida
en PEC1-2014 (Moodle) y PEC1-2015 (en aLF)
PEC-15 en aLF
PEC-14 pruebas
0
1
2
3
4
5
6
7
8
9
10
y = 0,8997x + 1,8587
R² = 0,636
coef de correlación=0,797

## 7.3 El coeficiente de correlación simple y su interpretación.

| | | NOMINAL | | | ORDINAL | INTERVALO | RAZÓN |
|---|---|---|---|---|---|---|---|
| | | Categórica | Dicotómica | Dicotomizada | | | |
| **N** | Categórica | **Coef. Contingencia (C)** | | | | | |
| | Dicotómica | | | | | | |
| | Dicotomizada | | | | | | |
| **0** | | | | | **Spearman ($r_s$)** | | |
| **I** | | | **Coef. Biserial-Puntual $r_{bp}$** | | | **Pearson (r)** | |
| **R** | | | | | | | |

La elección de los diferentes coeficientes de correlación depende del nivel de medida y de la categoría de las variables

## 7.4 El coeficiente de correlación de Pearson (r).

Las dos variables son **cuantitativas** y **medidas a nivel de intervalo** (y además se **distribuyen normalmente)**

$$r_{xy} = \frac{n\sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right] \cdot \left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

**Puntuaciones Directas**

$$r_{xy} = \frac{n\sum x \cdot y}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

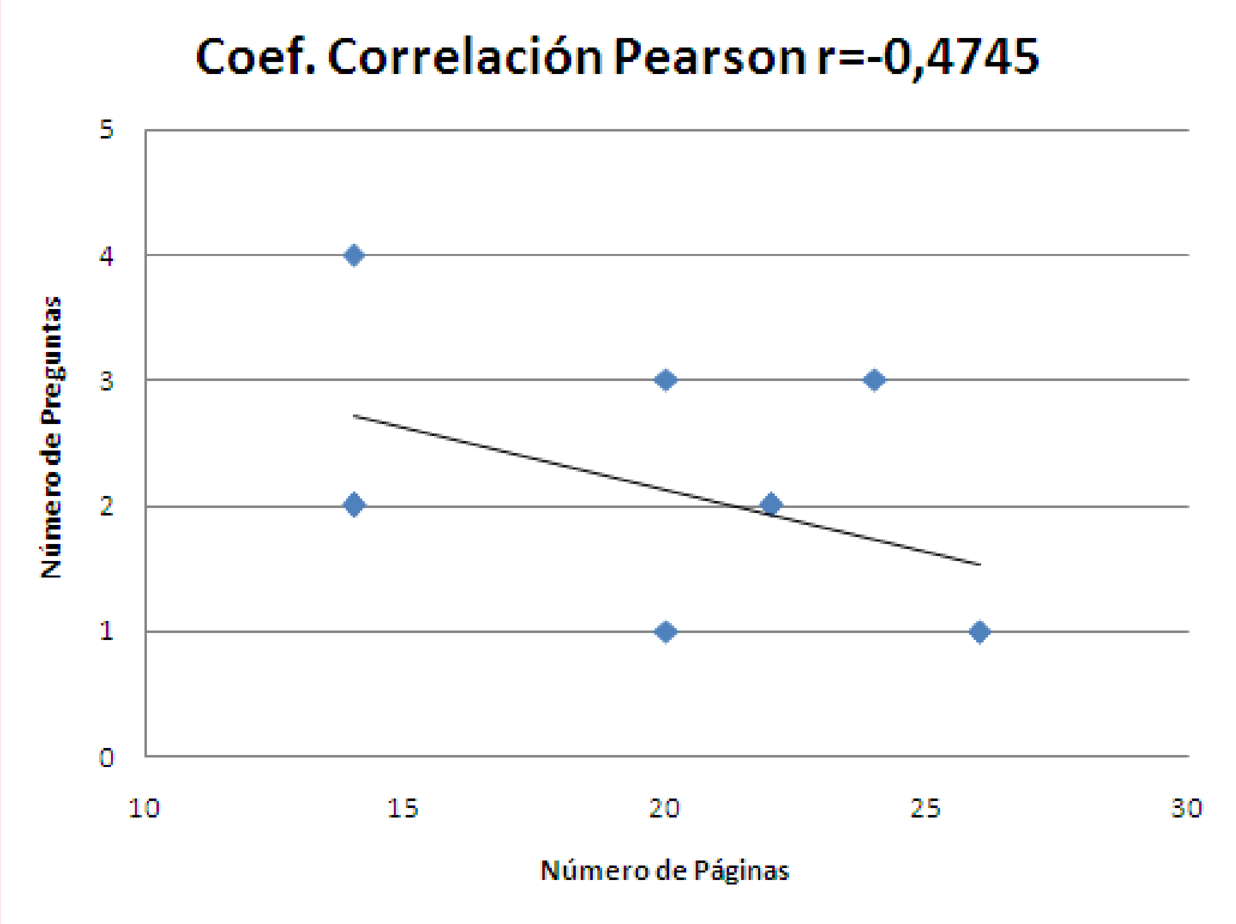
**Puntuaciones Diferenciales**

# 7.4 El coeficiente de correlación de Pearson (r).

**X→ nº de páginas Tema**

**Y→ nº preguntas en examen**

| | X | Y |
|---|---|---|
| T1 | 14 | 4 |
| T2 | 26 | 1 |
| T4 | 14 | 2 |
| T5 | 20 | 3 |
| T6 | 14 | 2 |
| T7 | 22 | 2 |
| T8 | 26 | 1 |
| T9 | 24 | 3 |
| T10 | 22 | 2 |
| T11 | 20 | 1 |
| **Suma** | **202** | **21** |



# 7.4 El coeficiente de correlación de Pearson (r).

$$r_{xy} = \frac{n\sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right] \cdot \left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

**Puntuaciones Directas**

| | X | Y | X2 | Y2 | XY |
|---|---|---|---|---|---|
| T1 | 14 | 4 | 196 | 16 | 56 |
| T2 | 26 | 1 | 676 | 1 | 26 |
| T4 | 14 | 2 | 196 | 4 | 28 |
| T5 | 20 | 3 | 400 | 9 | 60 |
| T6 | 14 | 2 | 196 | 4 | 28 |
| T7 | 22 | 2 | 484 | 4 | 44 |
| T8 | 26 | 1 | 676 | 1 | 26 |
| T9 | 24 | 3 | 576 | 9 | 72 |
| T10 | 22 | 2 | 484 | 4 | 44 |
| T11 | 20 | 1 | 400 | 1 | 20 |
| **Suma** | **202** | **21** | **4284** | **53** | **404** |

$$r_{xy} = \frac{10(404) - 202 \cdot 21}{\sqrt{\left[10(4284) - (202)^2\right] \cdot \left[10(53) - (21)^2\right]}} =$$

$$= \frac{-201}{425{,}68} = -0{,}4722$$

X→ nº de páginas Tema  Y→ nº preguntas en examen

## 7.4 El coeficiente de correlación de Pearson (r).

$$r_{xy} = \frac{\sum x \cdot y}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

**Puntuaciones Diferenciales**

| | X | Y | x=Xi-Xm | y=Yi-Ym | x2 | y2 | xy |
|---|---|---|---|---|---|---|---|
| **T1** | 14 | 4 | -6,2 | 1,9 | 38,44 | 3,61 | -11,78 |
| **T2** | 26 | 1 | 5,8 | -1,1 | 33,64 | 1,21 | -6,38 |
| **T4** | 14 | 2 | -6,2 | -0,1 | 38,44 | 0,01 | 0,62 |
| **T5** | 20 | 3 | -0,2 | 0,9 | 0,04 | 0,81 | -0,18 |
| **T6** | 14 | 2 | -6,2 | -0,1 | 38,44 | 0,01 | 0,62 |
| **T7** | 22 | 2 | 1,8 | -0,1 | 3,24 | 0,01 | -0,18 |
| **T8** | 26 | 1 | 5,8 | -1,1 | 33,64 | 1,21 | -6,38 |
| **T9** | 24 | 3 | 3,8 | 0,9 | 14,44 | 0,81 | 3,42 |
| **T10** | 22 | 2 | 1,8 | -0,1 | 3,24 | 0,01 | -0,18 |
| **T11** | 20 | 1 | -0,2 | -1,1 | 0,04 | 1,21 | 0,22 |
| **Suma** | **202** | **21** | **0** | **0** | **203,6** | **8,9** | **-20,2** |
| **Media** | 20,2 | 2,1 | | | | | |

$$r_{xy} = \frac{(-20,2)}{\sqrt{203,6 \cdot 8,9}} = \frac{-20,2}{42,56} = -0,4745$$

X→ nº de páginas Tema Y→ nº preguntas en examen

## 7.4 El coeficiente de correlación de Pearson (r).

# Propiedades

$$r_{xy} = \frac{n\sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right] \cdot \left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

a) El coeficiente de correlación de Pearson se encuentra comprendido entre los valores -1 y 1.

b) En el caso de que $r_{xy}$ valga 1, tendrá que cumplirse que para toda pareja de valores, sus puntuaciones típicas son iguales: $z_x=z_y$. En el polo opuesto, es decir, si $r_{xy}$ vale -1, entonces se cumple que para todo par de valores, las puntuaciones típicas son iguales pero de distinto signo: $z_x=-z_y$.
Por tanto si $z_x=z_y$, entonces $r_{xy}=1$
y podremos escribir

$$r_{xy} = \frac{\sum Z_x \cdot Z_y}{n} = \frac{\sum Z_x^2}{n} = \frac{\sum \left(Z_x - \bar{Z}_x\right)^2}{n}$$

(ya que, de acuerdo con las propiedades de las puntuaciones típicas, la media de las puntuaciones típicas vale 0, y que la varianza de la puntuaciones z para una variable vale 1).

c) La transformación lineal de las variables no modifica el valor del coeficiente de correlación, aunque sí podría cambiar su signo. Es decir, si calculamos la correlación entre las variables X e Y, el valor de ésta será, en valor absoluto, el mismo que obtengamos entre la variable aX+b, donde a y b son constantes.

## 7.5 El coeficiente de correlación ordinal de Spearman ($r_s$)

Las variables son cuantitativas pero sólo puede garantizarse un **nivel de medida ordinal**

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)}$$

Para calcular este coeficiente, debemos transformar las puntuaciones directas obtenidas con el instrumento en **rangos.**

Se suele comenzar asignando el rango o posición 1 a la puntuación más alta, la siguiente tendrá el rango 2, la siguiente el 3… de tal manera que el último rango que se asigne debe coincidir con el número de sujetos de la muestra.

***¡Ojo!:***

En aquellos casos en que existan varias puntuaciones directas coincidentes, la asignación del rango se realiza calculando el promedio de las posiciones que ocupan.

El criterio de asignación de rangos debe ser el mismo para ambas variables

## 7.5 El coeficiente de correlación ordinal de Spearman ($r_s$)

**Medir la relación entre la talla en cm y el número de hermanos**

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)}$$

| | altura | hermanos |
|---|---|---|
| | X | Y |
| 1 | 155 | 2 |
| 2 | 158 | 3 |
| 3 | 165 | 2 |
| 4 | 154 | 0 |
| 5 | 163 | 2 |
| 6 | 152 | 5 |
| 7 | 170 | 1 |
| 8 | 167 | 5 |
| 9 | 160 | 1 |
| 10 | 172 | 4 |
| 11 | 173 | 2 |
| 12 | 158 | 1 |
| 13 | 168 | 3 |

| | altura | hermanos | | | | |
|---|---|---|---|---|---|---|
| | X | Y | Rango (X) | Rango (Y) | D | D2 |
| 6 | 152 | 5 | 1 | 12,5 | -11,5 | 132,25 |
| 4 | 154 | 0 | 2 | 1 | 1 | 1 |
| 1 | 155 | 2 | 3 | 6,5 | -3,5 | 12,25 |
| 2 | 158 | 3 | 4 | 9,5 | -5,5 | 30,25 |
| 12 | 158 | 1 | 5 | 3 | 2 | 4 |
| 9 | 160 | 1 | 6 | 3 | 3 | 9 |
| 5 | 163 | 2 | 7 | 6,5 | 0,5 | 0,25 |
| 3 | 165 | 2 | 8 | 6,5 | 1,5 | 2,25 |
| 8 | 167 | 5 | 9 | 12,5 | -3,5 | 12,25 |
| 13 | 168 | 3 | 10 | 9,5 | 0,5 | 0,25 |
| 7 | 170 | 1 | 11 | 3 | 8 | 64 |
| 10 | 172 | 4 | 12 | 11 | 1 | 1 |
| 11 | 173 | 2 | 13 | 6,5 | 6,5 | 42,25 |
| | | | | | | **311** |

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)} \rightarrow r_s = 1 - \frac{6 \cdot 311}{13(13^2 - 1)} = 1 - \frac{1866}{2184} = 1 - 0{,}8544 = 0{,}1456$$

## 7.5 El coeficiente de correlación ordinal de Spearman ($r_s$)

# Propiedades

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)}$$

a) El coeficiente de correlación de Spearman se encuentra siempre comprendido entre los valores -1 y 1. Es decir, $-1 < r_s < 1$.

b) Cuando todos los sujetos se sitúan en el mismo puesto para la variable X y para la variable Y, el valor de $r_s$ es 1. Si ocupan valores opuestos, es decir, al primer sujeto en X le corresponde el último lugar en Y, al segundo en X le corresponde el penúltimo en Y, etc., entonces el valor de $r_s$ es -1.

c) El coeficiente $r_s$ es un caso particular de $r_{xy}$, puesto que se calcula a partir de éste, por aplicación del coeficiente de Pearson a valores ordinales considerados como puntuaciones. Por ello, al aplicar la fórmula de $r_{xy}$ a los valores de dos series de rangos, obtendríamos el mismo resultado que con la fórmula de $r_s$.

d) Si calculamos el coeficiente de correlación de Pearson entre dos variables X e Y, y el coeficiente de correlación de Spearman para las mismas puntuaciones pero transformadas en rangos, ambos coeficientes se aproximan en valor según aumenta el número de sujetos n.

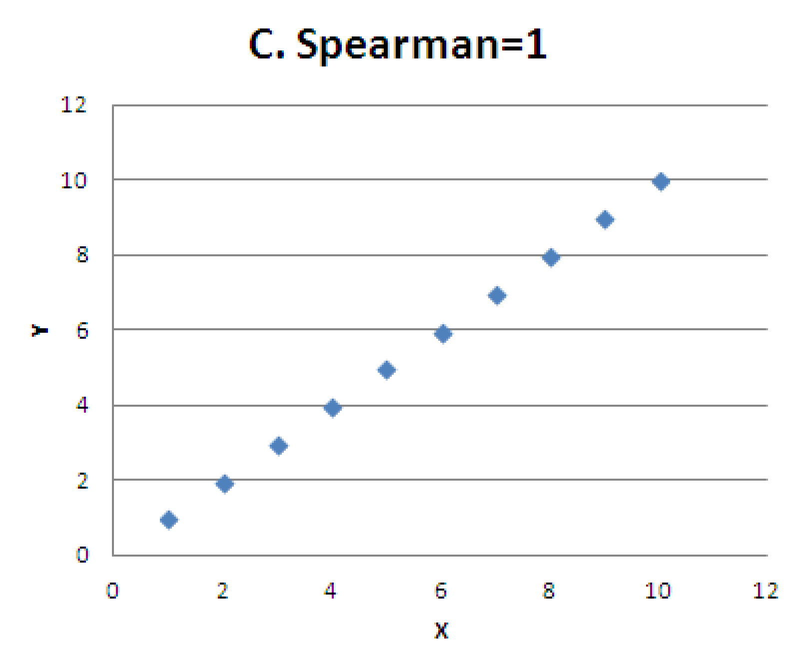
## 7.5 El coeficiente de correlación ordinal de Spearman ($r_s$)

**Si ambas variables tienen el mismo orden**

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)}$$

| X | Y | Rango(x) | Rango(y) | di | di2 |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 0 |
| 2 | 2 | 2 | 2 | 0 | 0 |
| 3 | 3 | 3 | 3 | 0 | 0 |
| 4 | 4 | 4 | 4 | 0 | 0 |
| 5 | 5 | 5 | 5 | 0 | 0 |
| 6 | 6 | 6 | 6 | 0 | 0 |
| 7 | 7 | 7 | 7 | 0 | 0 |
| 8 | 8 | 8 | 8 | 0 | 0 |
| 9 | 9 | 9 | 9 | 0 | 0 |
| 10 | 10 | 10 | 10 | 0 | 0 |
| | | | | | **0** |

**n= 10**



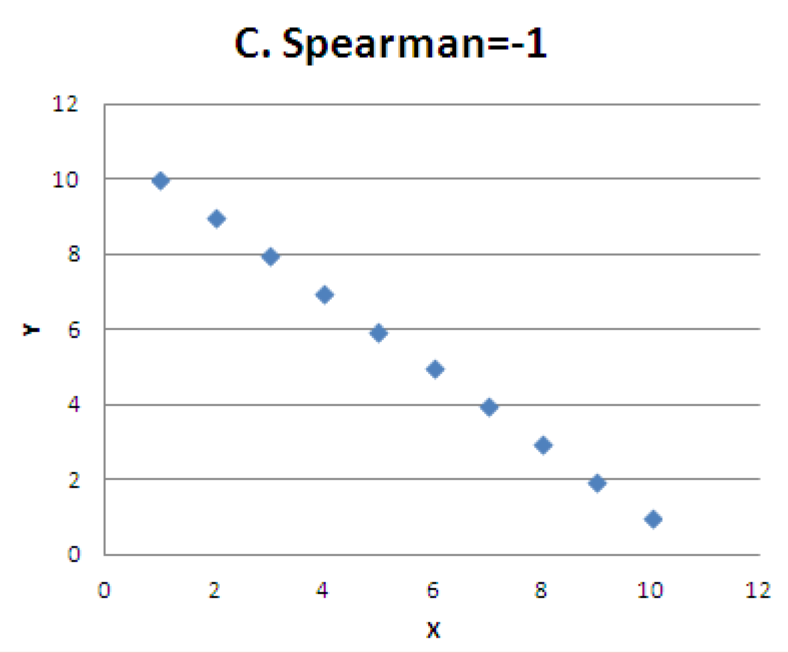

**C. Spearman= 1**

## 7.5 El coeficiente de correlación ordinal de Spearman ($r_s$)

**Si ambas variables tienen el orden invertido**

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)}$$

| X | Y | Rango(x) | Rango(y) | di | di2 |
|---|---|---|---|---|---|
| 1 | 10 | 1 | 10 | -9 | 81 |
| 2 | 9 | 2 | 9 | -7 | 49 |
| 3 | 8 | 3 | 8 | -5 | 25 |
| 4 | 7 | 4 | 7 | -3 | 9 |
| 5 | 6 | 5 | 6 | -1 | 1 |
| 6 | 5 | 6 | 5 | 1 | 1 |
| 7 | 4 | 7 | 4 | 3 | 9 |
| 8 | 3 | 8 | 3 | 5 | 25 |
| 9 | 2 | 9 | 2 | 7 | 49 |
| 10 | 1 | 10 | 1 | 9 | 81 |
| | | | | | **330** |




**n= 10**

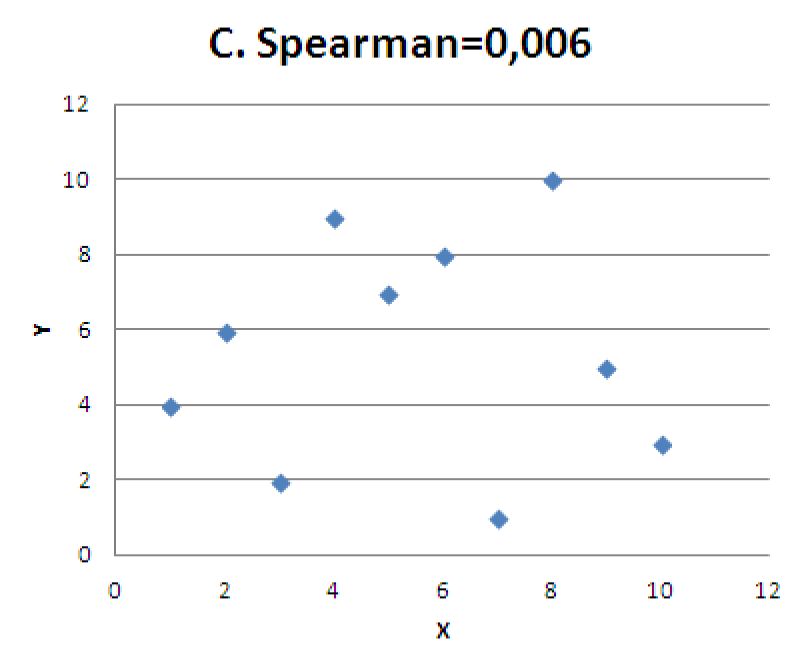
**C. Spearman= -1**

## 7.5 El coeficiente de correlación ordinal de Spearman ($r_s$)

**Si ambas variables no tienen relación**

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)}$$

| X | Y | Rango(x) | Rango(y) | di | di2 |
|---|---|---|---|---|---|
| 1 | 4 | 1 | 4 | -3 | 9 |
| 2 | 6 | 2 | 6 | -4 | 16 |
| 3 | 2 | 3 | 2 | 1 | 1 |
| 4 | 9 | 4 | 9 | -5 | 25 |
| 5 | 7 | 5 | 7 | -2 | 4 |
| 6 | 8 | 6 | 8 | -2 | 4 |
| 7 | 1 | 7 | 1 | 6 | 36 |
| 8 | 10 | 8 | 10 | -2 | 4 |
| 9 | 5 | 9 | 5 | 4 | 16 |
| 10 | 3 | 10 | 3 | 7 | 49 |
| | | | | | **164** |




**n= 10**

**C. Spearman= 0,006**

## 7.6 El coeficiente de contingencia (C).

Para hallar el **grado de asociación entre 2 variables nominales** (categóricas)

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

$$\chi^2 = \sum_{g=1}^{G} \sum_{c=1}^{C} \frac{(f_o - f_e)^2}{f_e}$$

$$f_e = \frac{f_f \cdot f_c}{f_t}$$

En el caso de **variables nominales o atributos,** se suele hablar de *“**grado de asociación**” (no de “grado de correlación”).*

*Se utiliza en aquellos supuestos en que se recogen datos de **variables clasificadas en categorías.***

*Así, las **tablas de contingencia** reflejan una asignación de sujetos a grupos y categorías en cada una de las variables.*

# 7.6 El coeficiente de contingencia (C).

## Medir la relación entre el Color de Pelo y el Color de Ojos

| Color Pelo | | Color Ojos: Azul | Gris/Verde | Negro/Pardo |
|---|---|---|---|---|
|  | Rubio | 260 | 140 | 17 |
| | Castaño | 118 | 204 | 64 |
| | Negro | 28 | 110 | 42 |
| | Pelirrojo | 7 | 8 | 2 |

$$52{,}13 = \frac{417 \cdot 125}{1000} = \frac{52125}{1000}$$

$$f_e = \frac{f_f \cdot f_c}{f_t}$$

| Color Pelo | | Color Ojos: Azul | Gris/Verde | Negro/Pardo | |
|---|---|---|---|---|---|
| | Rubio | 260 | 140 | 17 | **417** |
| | | 172,22 | 192,65 | 52,13 | |
| | Castaño | 118 | 204 | 64 | **386** |
| | | 159,42 | 178,33 | 48,25 | |
| | Negro | 28 | 110 | 42 | **180** |
| | | 74,34 | 83,16 | 22,50 | |
| | Pelirrojo | 7 | 8 | 2 | **17** |
| | | 7,02 | 7,85 | 2,13 | |
| | | **413** | **462** | **125** | **1.000** |

$$413 = 260 + 118 + 28 + 7$$

# 7.6 El coeficiente de contingencia (C).

| | | Color Ojos | | |
|---|---|---|---|---|
| | | Azul | Gris/Verde | Negro/Pardo |
| Color Pelo | Rubio | 44,74 | 14,39 | 23,67 |
| | Castaño | 10,76 | 3,69 | 5,14 |
| | Negro | 28,89 | 8,66 | 16,90 |
| | Pelirrojo | 0,00 | 0,00 | 0,01 |

$$\chi^2 = \sum_{g=1}^{G}\sum_{c=1}^{C}\frac{(f_o - f_e)^2}{f_e}$$

$$f_e = \frac{f_f \cdot f_c}{f_t}$$

$\chi^2$
**156,86**

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}} \rightarrow C = \sqrt{\frac{156,86}{1000+156,86}} = \sqrt{\frac{156,86}{1156,86}} = 0,3682$$

$$C \in \left[0, \sqrt{\frac{k-1}{k}}\right] \rightarrow k = \min\{Filas, Columnas\} \rightarrow C \in \left[0, \sqrt{\frac{3-1}{3}}\right] \rightarrow C \in [0\,,\,0'816]$$

0.3682/0.816=0.451→ relación de nivel medio entre las dos variables color del PELO y color de los OJOS

## 7.6 El coeficiente de contingencia (C).

# Propiedades

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

a) El coeficiente de contingencia C está comprendido entre 0 y 1. Es decir, 0 < C < 1. En este caso, no tendría sentido hablar de coeficientes negativos o positivos. El signo suele indicar que las variables consideradas varían en una misma dirección o en dirección opuesta. Cuando trabajamos con variables nominales, no es posible hablar de incrementos o decrementos en el valor de las mismas, puesto que entre las modalidades de este tipo de variables no se dan ni siquiera relaciones de orden.
Por otra parte, el coeficiente C presenta el valor 0 cuando la relación entre las variables es nula, pero nunca alcanza el valor 1. El máximo que puede alcanzar C depende del número de filas y columnas.

b) El coeficiente C indica la intensidad de la relación, pero no cuáles son las modalidades de ambas variables que tienden a darse conjuntamente. **La relación se establece entre aquellas modalidades correspondientes a la fila y la columna de celdas con frecuencia esperada superior a la frecuencia observada.**

c) El valor de C depende del número de filas y de columnas de la tabla de contingencia construida para su cálculo. Por este motivo, no son comparables dos valores de C obtenidos para una misma pareja de variables, salvo en el caso en que correspondan a tablas de las mismas dimensiones.

d) El coeficiente de contingencia C no es comparable a otros coeficientes de correlación, tales como $r_{xy}$ de Pearson o $r_s$ de Spearman. Valores similares en C y en cualquiera de estos coeficientes no tendrían que indicar un similar grado de correlación entre las parejas de variables implicadas en cada caso.

## 7.7 El Coeficiente de correlación biserial-puntual ($r_{bp}$).

Entre una variable **cuantitativa** (continua o discreta) y otra **auténticamente dicotómica**

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_t}{S_t} \cdot \sqrt{\frac{p}{q}}$$

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_q}{S_t} \cdot \sqrt{p \cdot q}$$

Cuando buscamos el grado de relación que se manifiesta entre una variable cuantitativa, continua o discreta, y otra auténticamente dicotómica, debemos recurrir al coeficiente biserial-puntual

## 7.7 El Coeficiente de correlación biserial-puntual ($r_{bp}$).

**Comprobar la relación entre el rendimiento académico de los estudiantes de Matemáticas (Secundaria), medido con prueba objetiva (50 ítems) y el sexo.**

| Puntos Intervalo | Femenino | Masculino | Total |
|---|---|---|---|
| 1-7 | 2 | 1 | 3 |
| 8-14 | 6 | 4 | 10 |
| 15-21 | 9 | 5 | 14 |
| 22-28 | 11 | 13 | 24 |
| 29-35 | 8 | 10 | 18 |
| 36-42 | 3 | 6 | 9 |
| 43-49 | 2 | 4 | 6 |
| **TOTAL** | **41** | **43** | **84** |

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_t}{S_t} \cdot \sqrt{\frac{p}{q}}$$

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_q}{S_t} \cdot \sqrt{p \cdot q}$$

# 7.7 El Coeficiente de correlación biserial-puntual ($r_{bp}$).

| Intervalo | fp | fq | ft | Xi | Xifp | Xifq | Xift | Xi2ft |
|---|---|---|---|---|---|---|---|---|
| 1-7 | 2 | 1 | 3 | 4 | 8 | 4 | 12 | 48 |
| 8-14 | 6 | 4 | 10 | 11 | 66 | 44 | 110 | 1.210 |
| 15-21 | 9 | 5 | 14 | 18 | 162 | 90 | 252 | 4.536 |
| 22-28 | 11 | 13 | 24 | 25 | 275 | 325 | 600 | 15.000 |
| 29-35 | 8 | 10 | 18 | 32 | 256 | 320 | 576 | 18.432 |
| 36-42 | 3 | 6 | 9 | 39 | 117 | 234 | 351 | 13.689 |
| 43-49 | 2 | 4 | 6 | 46 | 92 | 184 | 276 | 12.696 |
| **TOTAL** | **41** | **43** | **84** | | **976** | **1.201** | **2.177** | **65.611** |
| Medias | | | | | 23,80 | 27,93 | 25,92 | |

$$\bar{X}_p = \frac{\sum X_i f_p}{f_p} \rightarrow \bar{X}_p = \frac{976}{41} = 23{,}80$$

$$\bar{X}_q = \frac{\sum X_i f_q}{f_q} \rightarrow \bar{X}_q = \frac{1201}{43} = 27{,}93$$

$$\bar{X}_t = \frac{\sum X_i f_t}{f_t} \rightarrow \bar{X}_t = \frac{2177}{84} = 25{,}92$$

$$S_t = \sqrt{\frac{\sum X_i^2 f_t}{f_t} - \left(\frac{\sum X_i f_t}{f_t}\right)^2} \rightarrow$$

$$S_t = \sqrt{\frac{65611}{84} - \left(\frac{2177}{84}\right)^2} = 10{,}45$$

$$p = \frac{n_p}{n_t} = \frac{41}{84} = 0{,}488$$

$$q = \frac{n_q}{n_t} = \frac{43}{84} = 0{,}512$$

# 7.7 El Coeficiente de correlación biserial-puntual ($r_{bp}$).

| Intervalo | fp | fq | ft | Xi | Xifp | Xifq | Xift | Xi2ft |
|---|---|---|---|---|---|---|---|---|
| 1-7 | 2 | 1 | 3 | 4 | 8 | 4 | 12 | 48 |
| 8-14 | 6 | 4 | 10 | 11 | 66 | 44 | 110 | 1.210 |
| 15-21 | 9 | 5 | 14 | 18 | 162 | 90 | 252 | 4.536 |
| 22-28 | 11 | 13 | 24 | 25 | 275 | 325 | 600 | 15.000 |
| 29-35 | 8 | 10 | 18 | 32 | 256 | 320 | 576 | 18.432 |
| 36-42 | 3 | 6 | 9 | 39 | 117 | 234 | 351 | 13.689 |
| 43-49 | 2 | 4 | 6 | 46 | 92 | 184 | 276 | 12.696 |
| **TOTAL** | **41** | **43** | **84** | | **976** | **1.201** | **2.177** | **65.611** |
| Medias | | | | | 23,80 | 27,93 | 25,92 | |

$$S_t = \sqrt{\frac{\sum X_i^2 f_t}{f_t} - \left(\frac{\sum X_i f_t}{f_t}\right)^2} \rightarrow$$

$$S_t = \sqrt{\frac{65611}{84} - \left(\frac{2177}{84}\right)^2} = 10,45$$

$$p = \frac{n_p}{n_t} = \frac{41}{84} = 0,488$$

$$q = \frac{n_q}{n_t} = \frac{43}{84} = 0,512$$

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_t}{S_t} \cdot \sqrt{\frac{p}{q}} \rightarrow r_{bp} = \frac{23,80 - 25,92}{10,45} \cdot \sqrt{\frac{0,488}{0,512}} = 0,20$$

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_q}{S_t} \cdot \sqrt{p \cdot q} \rightarrow r_{bp} = \frac{23,80 - 27,93}{10,45} \cdot \sqrt{0,488 \cdot 0,512} = 0,20$$

## 7.7 El Coeficiente de correlación biserial-puntual ($r_{bp}$).

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_t}{S_t} \cdot \sqrt{\frac{p}{q}}$$

# Propiedades

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_q}{S_t} \cdot \sqrt{p \cdot q}$$

a) Se demuestra que el coeficiente $r_{bp}$ es resultado de aplicar el coeficiente de correlación de Pearson al caso en que una de las variables tiene carácter dicotómico.

b) El valor de $r_{bp}$ no puede ser mayor que 1 ni menor que -1 Es decir, se cumple $-1 < r_{bp} < 1$. Cuanto mayor sea la distancia entre la media de los sujetos que presentan la primera modalidad y la media del total de sujetos, más próximo a 1 ó -1 será el coeficiente de correlación que obtengamos.

c) **Un coeficiente de correlación positivo indicará que a puntuaciones altas de X corresponde pertenecer a la categoría cuya proporción es p**, mientras que a puntuaciones bajas de X corresponde pertenecer a la categoría cuya proporción es q. **Un coeficiente negativo deberá ser interpretado en sentido contrario, es decir, a puntuaciones altas de X correspondería la categoría cuya proporción es q**, y a puntuaciones bajas aquélla cuya proporción es p.

## 7.8 Otros coeficientes de correlación.

- 7.8.1 El coeficiente PHI ($\Phi$)
- 7.8.2 El coeficiente de correlación tetracórico ($r_t$)
- 7.8.3 Coeficiente de correlación biserial ($r_b$)

## 7.8.1 El coeficiente PHI (Φ)

Entre variables **dicotómicas** (excepcionalmente se puede usar con variables dicotomizadas)

$$\phi = \frac{\left|B \cdot C - A \cdot D\right|}{\sqrt{(A+B)(A+C)(C+D)(B+D)}}$$

| | Cat. 1 | Cat .2 | Total |
|---|---|---|---|
| **Cat. 1** | (A) | (B) | |
| **Cat. 2** | (C) | (D) | |
| **Total** | | | |

## 7.8.1 El coeficiente PHI (Φ)

**Comprobar la relación entre la pertenencia a un centro de enseñanza (Público o Privado) y la respuesta dada a un ítem (Verdadero/Falso).**

| | Público | Privado | TOTAL |
|---|---|---|---|
| **Falso** | 39(A) | 29(B) | 68 |
| **Verdadero** | 64(C) | 55(D) | 119 |
| **TOTAL** | 103 | 84 | 187 |

$$\phi = \frac{|B \cdot C - A \cdot D|}{\sqrt{(A+B)(A+C)(C+D)(B+D)}}$$

$$\phi = \frac{|B \cdot C - A \cdot D|}{\sqrt{(A+B)(A+C)(C+D)(B+D)}} = \frac{|29 \cdot 64 - 39 \cdot 55|}{\sqrt{68 \cdot 103 \cdot 119 \cdot 84}} = \frac{289}{8367{,}32} = 0{,}034$$

# 7.8.1 El coeficiente PHI (Φ)

## Propiedades

$$\phi = \frac{|B \cdot C - A \cdot D|}{\sqrt{(A+B)(A+C)(C+D)(B+D)}}$$

a) El coeficiente φ es un caso particular de $r_{xy}$, puesto que se calcula a partir de éste, por aplicación del coeficiente de Pearson a una serie de valores de carácter dicotómico.

b) El coeficiente de correlación φ se encuentra comprendido entre los valores -1 y 1. Es decir, -1 < φ < 1.

Este coeficiente de correlación será positivo cuando cb sea mayor que ad. En este caso, existe una relación entre las dos variables en el sentido de que los sujetos que presentan el valor 0 en la variable X tienden a presentar el valor 0 también en Y, y sujetos que presentan el valor 1 en X tienden al valor 1 en la variable Y.

Por el contrario, el coeficiente será negativo cuando cb sea menor que ad. En tal situación, predominan los sujetos situados en las casillas correspondientes a las frecuencias a y d. Es decir, existe relación entre presentar el valor 0 en X y presentar el valor 1 en Y. De forma recíproca, existe relación entre presentar el valor 1 en X y el valor 0 en la variable Y.

c) El valor de φ será 1 cuando todos los sujetos que presentan la modalidad 1 en X presentan la modalidad 1 en Y, y todos los sujetos con 0 en X obtienen 0 en Y. El valor de φ será -1 cuando todos los sujetos que presentan la modalidad 1 en X presentan la modalidad 0 en Y, y todos los sujetos con 0 en X obtienen 1 en Y

## 7.8.2 El coeficiente de correlación tetracórico ($r_t$)

Entre variables **cuantitativas y continuas** (por razones de la investigación se **dicotomizan**)

$$r_t = \frac{B \cdot C}{A \cdot D}$$

$$r_t = \cos\left( \frac{180\sqrt{A \cdot D}}{\sqrt{B \cdot C} + \sqrt{A \cdot D}} \right)$$

| | | **Md(y)** | | |
|---|---|---|---|---|
| | | **-** | **+** | **Total** |
| **Md(x)** | **+** | (A) | (B) | |
| | **-** | (C) | (D) | |
| | **Total** | | | |

## 7.8.2 El coeficiente de correlación tetracórico ($r_t$)

**Comprobar la relación entre las puntuaciones alcanzadas en matemáticas (X) y una prueba de razonamiento matemático (y)**

| Estudiantes | Puntuación en Matemáticas | Prueba de Razonamiento | Estudiantes | Puntuación en Matemáticas | Prueba de Razonamiento |
|---|---|---|---|---|---|
| 1 | 1 | 35 | 11 | **5** | 48 |
| 2 | 2 | 40 | 12 | 5 | 55 |
| 3 | 2 | 28 | 13 | 5 | 43 |
| 4 | 3 | 20 | 14 | 6 | 40 |
| 5 | 3 | 25 | 15 | 6 | 60 |
| 6 | 3 | 42 | 16 | 6 | 80 |
| 7 | 4 | 52 | 17 | 7 | 70 |
| 8 | 4 | 41 | 18 | 8 | 90 |
| 9 | 4 | 55 | 19 | 9 | 75 |
| 10 | **4** | 58 | 20 | 10 | 95 |

2|805
3|5
4|021830
5|2585
6|0
7|05
8|0
9|05

Para calcular la Mediana de la variable Y

2|058
3|5
4|00123**8**
5|**2**558
6|0
7|05
8|0
9|05

$$Md_{(x)} = \frac{4+5}{2} = 4{,}5 \qquad Md_{(y)} = \frac{48+52}{2} = 50$$

**Como hay 20 datos la Media está entre la posición 10 y la posición 11**

# 7.8.2 El coeficiente de correlación tetracórico ($r_t$)

| Estudiantes | Puntuación en Matemáticas | X-Mdx | signo | Prueba de Razonamiento | Y-Mdy | signo | Estudiantes | Puntuación en Matemáticas | X-Mdx | signo | Prueba de Razonamiento | Y-Mdy | signo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | -3,5 | - | 35 | -15 | - | 11 | 5 | 0,5 | + | 48 | -2 | - |
| 2 | 2 | -2,5 | - | 40 | -10 | - | 12 | 5 | 0,5 | + | 55 | 5 | + |
| 3 | 2 | -2,5 | - | 28 | -22 | - | 13 | 5 | 0,5 | + | 43 | -7 | - |
| 4 | 3 | -1,5 | - | 20 | -30 | - | 14 | 6 | 1,5 | + | 40 | -10 | - |
| 5 | 3 | -1,5 | - | 25 | -25 | - | 15 | 6 | 1,5 | + | 60 | 10 | + |
| 6 | 3 | -1,5 | - | 42 | -8 | - | 16 | 6 | 1,5 | + | 80 | 30 | + |
| 7 | 4 | -0,5 | - | 52 | 2 | + | 17 | 7 | 2,5 | + | 70 | 20 | + |
| 8 | 4 | -0,5 | - | 41 | -9 | - | 18 | 8 | 3,5 | + | 90 | 40 | + |
| 9 | 4 | -0,5 | - | 55 | 5 | + | 19 | 9 | 4,5 | + | 75 | 25 | + |
| 10 | 4 | -0,5 | - | 58 | 8 | + | 20 | 10 | 5,5 | + | 95 | 45 | + |

Md(y)

Md(x)

| | - | + | Total |
|---|---|---|---|
| + | 3(A) | 7(B) | 10 |
| - | 7(C) | 3(D) | 10 |
| Total | 10 | 10 | 20 |

$$r_t = \frac{B \cdot C}{A \cdot D} \rightarrow r_t = \frac{7 \cdot 7}{3 \cdot 3} = \frac{49}{9} = 5{,}44$$

$$5{,}44 \text{ está entre } 5{,}389 \; y \; 5{,}595 \rightarrow 0{,}59$$

$$r_t = \cos\left(\frac{180\sqrt{A \cdot D}}{\sqrt{B \cdot C} + \sqrt{A \cdot D}}\right) \rightarrow r_t = \cos\left(\frac{180\sqrt{3 \cdot 3}}{\sqrt{7 \cdot 7} + \sqrt{3 \cdot 3}}\right) = \cos\left(\frac{180 \cdot 3}{7+3}\right) = \cos(54) = 0{,}59$$

## 7.8.2 El coeficiente de correlación tetracórico ($r_t$)

Valores estimados de $r_t$ correspondientes a los valores del cociente BC/AD

*Ejemplo:* Si $BC/AD = 3{,}28$, el $r_t$ correspondiente es 0,44. La interpolación entre valores $BC/AD$ no se recomienda, porque la exactitud de los valores $r_t$ no supera el segundo decimal. Si $AD$ es mayor que $BC$, buscar el cociente $AD/BC$ y dar un signo negativo al $r_t$.

| *BC/AD* | $r_t$ | *BC/AD* | $r_t$ | *BC/AD* | $r_t$ |
|---|---|---|---|---|---|
| 1,74-1,78 | 0,22 | 4.83-4.99 | 0,56 | 33,61-37,79 | 0,90 |
| 1,79-1,83 | 0,23 | 5,00-5,18 | 0,57 | 37,80-43,06 | 0,91 |
| 1,84-1,88 | 0,24 | 5,19-5,38 | 0,58 | 43,07-49,83 | 0,92 |
| 1,89-1,92 | 0,25 | 5,39-5,59 | 0,59 | 49,84-58,79 | 0,93 |
| 1,94-1,98 | 0,26 | 5,60-5,80 | 0,60 | 58,80-70,95 | 0,94 |
| 1,99-2,04 | 0,27 | 5,81-6,03 | 0,61 | 70,96-89,01 | 0,95 |

$$r_t = \frac{B \cdot C}{A \cdot D} \rightarrow r_t = \frac{7 \cdot 7}{3 \cdot 3} = \frac{49}{9} = 5{,}44$$

$$5{,}44 \text{ está entre } 5{,}389 \; y \; 5{,}595 \rightarrow 0{,}59$$

**Página 27 del Formulario y Tablas**

## 7.8.2 El coeficiente de correlación tetracórico ($r_t$)

# Propiedades

$$r_t = \frac{B \cdot C}{A \cdot D} \qquad r_t = \cos\left(\frac{180\sqrt{A \cdot D}}{\sqrt{B \cdot C} + \sqrt{A \cdot D}}\right)$$

a) El coeficiente $r_t$ puede valer más que 1 ó menos que -1. Este coeficiente de correlación será positivo cuando cb sea mayor que ad. En tal caso, existe una relación entre las dos variables en el sentido de que los sujetos que presentan el valor 0 en la variable X tienden a presentar el valor 0 también en Y, y sujetos que presentan el valor 1 en X tienden al valor 1 en Y.
Por el contrario, el coeficiente será negativo cuando cb sea menor que ad. En tal situación, predominan los sujetos situados en las casillas correspondientes a las frecuencias a y d. Es decir, existe relación entre presentar el valor 0 en X y presentar el valor 1 en Y. De forma recíproca, existe relación entre presentar el valor 1 en X y cl valor 0 en Y.

b) Si una de las cuatro frecuencias de la tabla de distribución conjunta es nula, el coeficiente de correlación tetracórica tendrá un valor $r_t$ = l ó $r_t$ = -l.
Si a ó d adoptan el valor 0, tendremos que en cb/ad, el denominador es 0, y por tanto el cociente tiende a infinito. La tabla 11 asigna, en este caso, un coeficiente $r_t$ = l. Si b ó c adoptan el valor 0, tendremos que en ad/cb, el denominador se hace 0, y consecuentemente el cociente tiende a infinito. La tabla 11 asigna ahora un coeficiente $r_t$=-l.

c) Para un mismo conjunto de datos, se cumple que $r_t$ vale aproximadamente (3/2)φ. Esta aproximación es tanto mejor cuanto más próximos se encuentren a la mediana los puntos de dicotomización de ambas variables y cuando $r_t$ es menor o igual a 0.50.

## 7.8.3 Coeficiente de correlación biserial ($r_b$)

Entre una variable **cuantitativa, continua o discreta** y otra **dicotomizada.**

$$r_b = \frac{\left|\overline{X}_p - \overline{X}_t\right|}{S_t} \cdot \frac{p}{y}$$

$$r_b = \frac{\left|\overline{X}_p - \overline{X}_q\right|}{S_t} \cdot \frac{p \cdot q}{y}$$

# 7.8.3 Coeficiente de correlación biserial ($r_b$)

**Relación entre las puntuaciones en una prueba de rendimiento (40 ítems) y el nivel de integración de los estudiantes de secundaria**

| Prueba de rendimiento | Integrado (p) | No Integrado (q) | Total |
|---|---|---|---|
| 1-5 | 3 | 3 | 6 |
| 6-10 | 4 | 6 | 10 |
| 11-15 | 8 | 9 | 17 |
| 16-20 | 15 | 14 | 29 |
| 21-25 | 23 | 10 | 33 |
| 26-30 | 12 | 5 | 17 |
| 31-35 | 10 | 4 | 14 |
| 36-40 | 5 | 2 | 7 |
| **TOTAL** | **80** | **53** | **133** |

# 7.8.3 Coeficiente de correlación biserial ($r_b$)

| Intervalo | fp | fq | ft | Xi | Xifp | Xifq | Xift | Xi2ft |
|---|---|---|---|---|---|---|---|---|
| 1-5 | 3 | 3 | 6 | 3 | 9 | 9 | 18 | 54 |
| 6-10 | 4 | 6 | 10 | 8 | 32 | 48 | 80 | 640 |
| 11-15 | 8 | 9 | 17 | 13 | 104 | 117 | 221 | 2.873 |
| 16-20 | 15 | 14 | 29 | 18 | 270 | 252 | 522 | 9.396 |
| 21-25 | 23 | 10 | 33 | 23 | 529 | 230 | 759 | 17.457 |
| 26-30 | 12 | 5 | 17 | 28 | 336 | 140 | 476 | 13.328 |
| 31-35 | 10 | 4 | 14 | 33 | 330 | 132 | 462 | 15.246 |
| 36-40 | 5 | 2 | 7 | 38 | 190 | 76 | 266 | 10.108 |
| **TOTAL** | **80** | **53** | **133** | | **1.800** | **1.004** | **2.804** | **69.102** |
| | | Medias | | | 22,50 | 18,94 | 21,08 | |

**p=** **0,60**

**q=** **0,40**

**S= 8,67**

**P=0,60 → área mayor q=0,40 → área menor → y=0,3857 (ver tabla)**

$$r_b = \frac{\left|\overline{X}_p - \overline{X}_t\right|}{S_t} \cdot \frac{p}{y} \rightarrow r_b = \frac{\left|22{,}50 - 21{,}08\right|}{8{,}67} \cdot \frac{0{,}60}{0{,}3857} = 0{,}255$$

## 7.8.3 Coeficiente de correlación biserial ($r_b$)

| Intervalo | fp | fq | ft | Xi | Xifp | Xifq | Xift | Xi2ft |
|---|---|---|---|---|---|---|---|---|
| 1-5 | 3 | 3 | 6 | 3 | 9 | 9 | 18 | 54 |
| 6-10 | 4 | 6 | 10 | 8 | 32 | 48 | 80 | 640 |
| 11-15 | 8 | 9 | 17 | 13 | 104 | 117 | 221 | 2.873 |
| 16-20 | 15 | 14 | 29 | 18 | 270 | 252 | 522 | 9.396 |
| 21-25 | 23 | 10 | 33 | 23 | 529 | 230 | 759 | 17.457 |
| 26-30 | 12 | 5 | 17 | 28 | 336 | 140 | 476 | 13.328 |
| 31-35 | 10 | 4 | 14 | 33 | 330 | 132 | 462 | 15.246 |
| 36-40 | 5 | 2 | 7 | 38 | 190 | 76 | 266 | 10.108 |
| **TOTAL** | **80** | **53** | **133** | | **1.800** | **1.004** | **2.804** | **69.102** |
| Medias | | | | | 22,50 | 18,94 | 21,08 | |

**p=** **0,60**

**q=** **0,40**

**S= 8,67**

**P=0,60 → área mayor q=0,40 → área menor → y=0,3857 (ver tabla)**

$$r_b = \frac{\left|\overline{X}_p - \overline{X}_q\right|}{S_t} \cdot \frac{p \cdot q}{y} \rightarrow r_b = \frac{\left|22{,}50 - 18{,}94\right|}{8{,}67} \cdot \frac{0{,}60 \cdot 0{,}40}{0{,}3857} = 0{,}255$$

## 7.8.3 Coeficiente de correlación biserial ($r_b$)

Tabla I.—Areas y ordenadas de la curva de distribución normal en función $x/\sigma$

**Página 17 del Formulario y Tablas**

| (1) $z$ Puntuación tipificada $\left(\frac{x}{\sigma}\right)$ | (2) $A$ Area desde la media a $\frac{x}{\sigma}$ | (3) $B$ Area de la parte mayor | (4) $C$ Area de la parte menor | (5) $y$ Ordenada en $\frac{x}{\sigma}$ |
|---|---|---|---|---|
| 0.25 | .0987 | .5987 | .4013 | .3867 |
| 0.26 | .1026 | .6026 | .3974 | .3857 |
| 0.27 | .1064 | .6064 | .3936 | .3847 |
| 0.28 | .1103 | .6103 | .3897 | .3836 |
| 0.29 | .1141 | .6141 | .3859 | .3825 |

**P=0,60 → área mayor q=0,40 → área menor → y=0,3857 (ver tabla)**

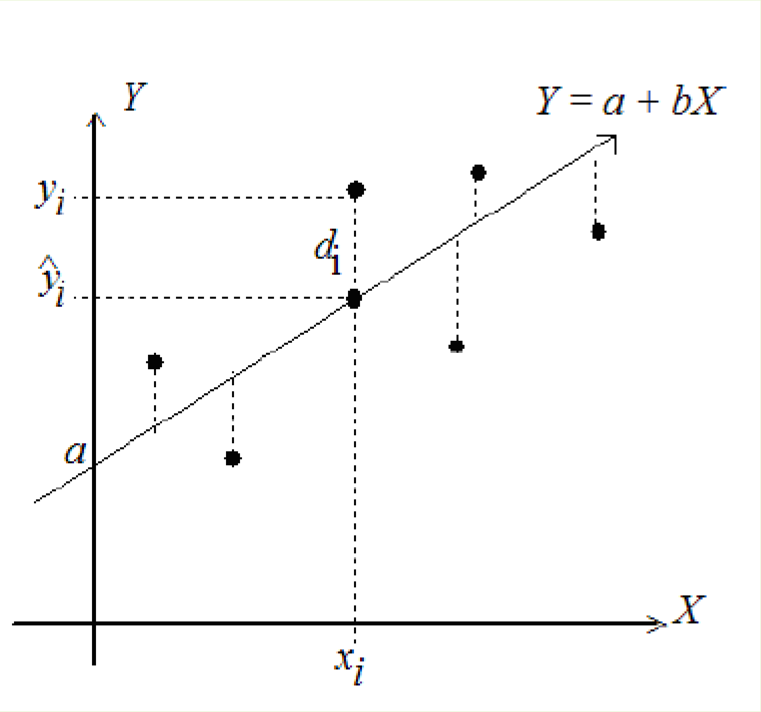
| | | NOMINAL | | | ORDINAL | INTERVALO | RAZÓN |
|---|---|---|---|---|---|---|---|
| | | Categórica | Dicotómica | Dicotomizada | | | |
| **N** | Categórica | **Coef. Contingencia (C)** | | | | | |
| | Dicotómica | | **Coef. PHI (Φ)** | | | | |
| | Dicotomizada | | | **Coef. Tetracórico ($r_t$)** | | | |
| **0** | | | | | **Spearman ($r_s$)** | | |
| **I** | | | **Coef. Biserial-Puntual $r_{bp}$** | **Coef. biserial ($r_b$)** | | **Pearson (r)** | |
| **R** | | | | | | | |

## 7.9 La regresión lineal simple.

Elevando al cuadrado el coeficiente de correlación obtenemos el coeficiente de determinación, que permite conocer el porcentaje de varianza compartida.

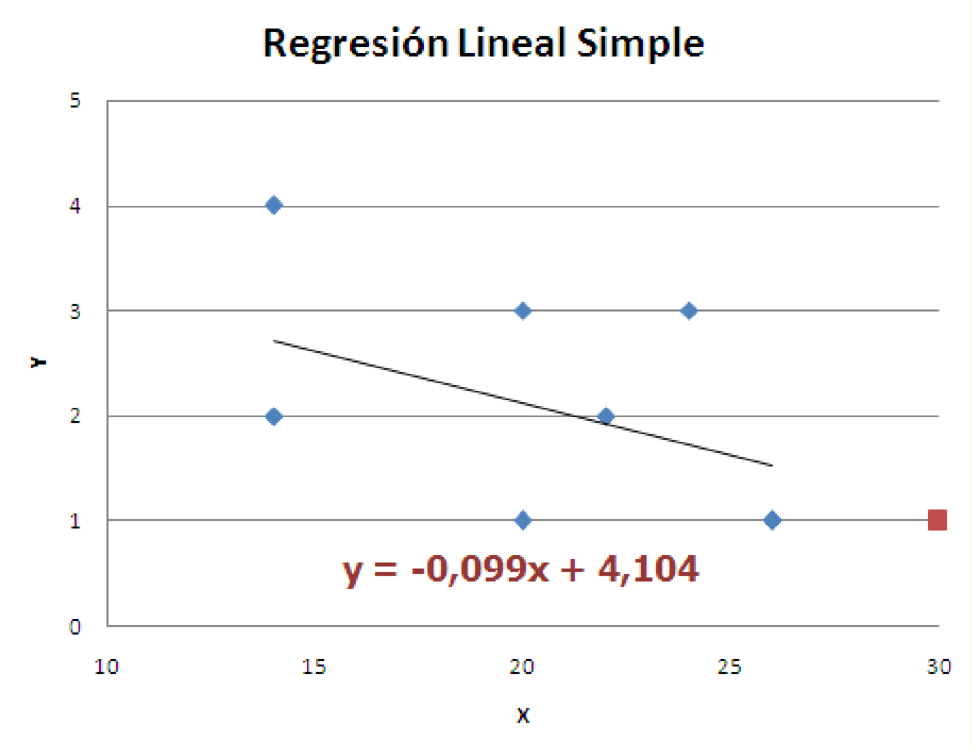
Esto hace posible que se puedan estimar los valores de una variable conociendo los valores en la otra: **regresión lineal simple**

La recta de regresión ***(Y = a + bX)*** *nos permite llevar a cabo la predicción o estimación de los valores en una variable (**variable criterio**) a partir del conocimiento de los valores en la otra variable (**variable predictora**), con la que mantiene una alta correlación.*



# 7.9 La regresión lineal simple.

| | X | Y |
|---|---|---|
| T1 | 14 | 4 |
| T2 | 26 | 1 |
| T4 | 14 | 2 |
| T5 | 20 | 3 |
| T6 | 14 | 2 |
| T7 | 22 | 2 |
| T8 | 26 | 1 |
| T9 | 24 | 3 |
| T10 | 22 | 2 |
| T11 | 20 | 1 |
| Suma | 202 | 21 |




$$y = -0{,}099x + 4{,}104$$

| | | |
|---|---|---|
| T12 | 30 | 1 |

$$y = -0{,}099(30) + 4{,}104 \cong 1$$

## CORRELACIONES

***a) Coeficiente de correlación de Pearson (r)***

$$r_{xy} = \frac{n \cdot \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{\left[n \sum X^2 - \left(\sum X\right)^2\right]\left[n \sum Y^2 - \left(\sum Y\right)^2\right]}}$$ Puntuaciones directas

$$r_{xy} = \frac{\sum x \cdot y}{\sqrt{\sum x^2 \cdot \sum y^2}}$$ Puntuaciones diferenciales

***b) Coeficiente de correlación de Spearman ($r_s$)***

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n\left(n^2 - 1\right)}$$

Dónde n nos indica el número de sujetos o de pares de puntuaciones y D es la diferencia de rangos o posiciones que ocupa un mismo sujeto en dos variables distintas.

***c) Coeficiente de Contingencia (C)***

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$ Dónde $$\chi^2 = \sum_{g=1}^{G} \sum_{c=1}^{c} \frac{\left(f_o - f_e\right)^2}{f_e}$$ A su vez $$f_e = \frac{f_f \cdot f_c}{f_t}$$

*d) Coeficiente de correlación biserial-puntual ($r_{bp}$)*

$$r_{bp} = \frac{\overline{X}_p - \overline{X}_t}{s_t} \cdot \sqrt{\frac{p}{q}} \qquad r_{bp} = \frac{\overline{X}_p - \overline{X}_q}{s_t} \cdot \sqrt{p \cdot q}$$

*e) El coeficiente PHI ($\phi$)*

$$\phi = \frac{|B \cdot C - A \cdot D|}{\sqrt{(A+B)(A+C)(C+D)(B+D)}}$$

*f) El coeficiente de correlación tetracórico ($r_t$)*

$$r_t = \frac{B \cdot C}{A \cdot D}$$

En el numerador figura el producto cruzado de la diagonal donde coinciden los mismos signos, mientras que en el denominador figura el cruce en que no coinciden los valores, es decir, son distintos.

Existe otro procedimiento directo, si bien exige disponer de una calculadora que incorpore las funciones trigonométricas, en este caso el coseno, mediante a siguiente fórmula:

$$r_t = \cos\left(\frac{180\sqrt{A \cdot D}}{\sqrt{B \cdot C} + \sqrt{A \cdot D}}\right)$$ El valor así obtenido es el coeficiente de correlación.

*g) Coeficiente de correlación biserial ($r_b$)*

$$r_b = \frac{|\overline{X}_p - \overline{X}_t|}{s_t} \cdot \frac{p}{y} \qquad r_b = \frac{|\overline{X}_p - \overline{X}_q|}{s_t} \cdot \frac{p \cdot q}{y}$$

## Resumen

Concepto de correlación.

Relación Perfecta Positiva.

Relación Imperfecta Positiva.

Relación Perfecta Negativa.

Relación Imperfecta Negativa.

Covariación.

Coeficiente de Correlación de Pearson.

Coeficiente de Correlación Ordinal de Spearman.

Coeficiente de Contingencia.

Coeficiente de Correlación Biserial-puntual

Coeficiente PHI

Coeficiente de Correlación Tetracórico

Coeficiente de Correlación Biserial.

Regresión Lineal Simple.

# Fe de erratas

# PREGUNTAS

Exámenes anteriores

Si en una muestra de 200 sujetos; 100 mujeres y 100 varones, pretendemos establecer la correlación entre la variable sexo (variable X), y los ingresos de cada persona, medidos en euros (variable Y), debemos utilizar':

Seleccione una:

- a. El $r_{xy}$ de Pearson
- b. El coeficiente de Spearman
- c. El coeficiente biserial puntual ( $r_{bp}$ )

La transformación de las puntuaciones de los sujetos a rangos o posiciones es una operación propia del coeficiente de correlación de:

Seleccione una:

- a. Pearson
- b. Spearman
- c. Contingencia

Si en una muestra de 200 sujetos; 100 mujeres y 100 varones, pretendemos establecer la correlación entre la variable sexo (variable X), y los ingresos de cada persona, medidos en euros (variable Y), debemos utilizar':

Seleccione una:

- a. El $r_{xy}$ de Pearson
- b. El coeficiente de Spearman
- c. El coeficiente biserial puntual ( $r_{bp}$ ) ✓

La respuesta correcta es: El coeficiente biserial puntual ( $r_{bp}$ )

La transformación de las puntuaciones de los sujetos a rangos o posiciones es una operación propia del coeficiente de correlación de:

Seleccione una:

- a. Pearson
- b. Spearman ✓
- c. Contingencia

La respuesta correcta es: Spearman

El Coeficiente de Correlación phi ( $\Phi$ ) busca la existencia de relaciones entre dos variables...

Seleccione una:

- a. Continuas
- b. Una dicotómica y otra continua
- c. Dicotómicas

Si al aumentar los valores de una variable en un grupo de sujetos, los valores de la otra aumentan y, además, lo hacen en la misma proporción, nos encontramos ante

Seleccione una:

- a. Una relación imperfecta positiva
- b. Una relación perfecta negativa
- c. Una relación perfecta positiva

El Coeficiente de Correlación phi ( $\Phi$ ) busca la existencia de relaciones entre dos variables...

Seleccione una:

- a. Continuas
- b. Una dicotómica y otra continua
- c. Dicotómicas

La respuesta correcta es: Dicotómicas

Si al aumentar los valores de una variable en un grupo de sujetos, los valores de la otra aumentan y, además, lo hacen en la misma proporción, nos encontramos ante

Seleccione una:

- a. Una relación imperfecta positiva
- b. Una relación perfecta negativa
- c. Una relación perfecta positiva

La respuesta correcta es: Una relación perfecta positiva