Análisis de la asociación espacial

Correlación

Gerardo Martín

2022-06-29

Intro

Asociación estadística:

Probar la hipótesis de que dos variables se predicen mutuamente

Asociación espacial:

Variables con estructura espacial que se predicen mutuamente

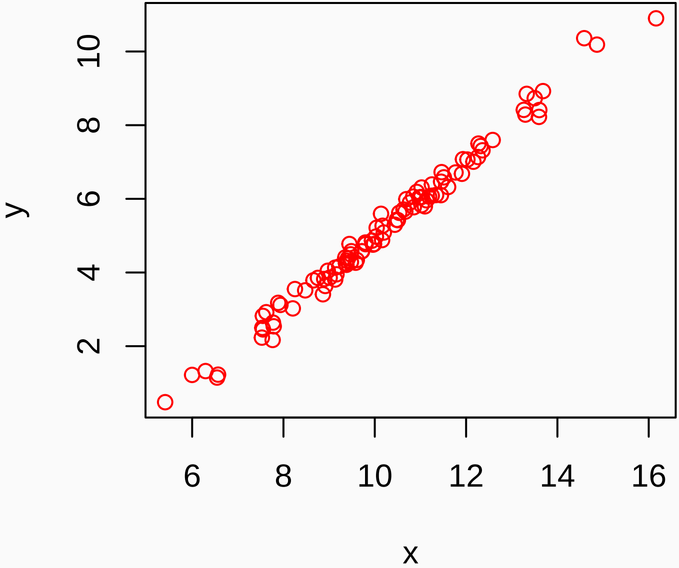


Figure 1: Gráfico de dispersión de dos variables que se predicen mutuamente.

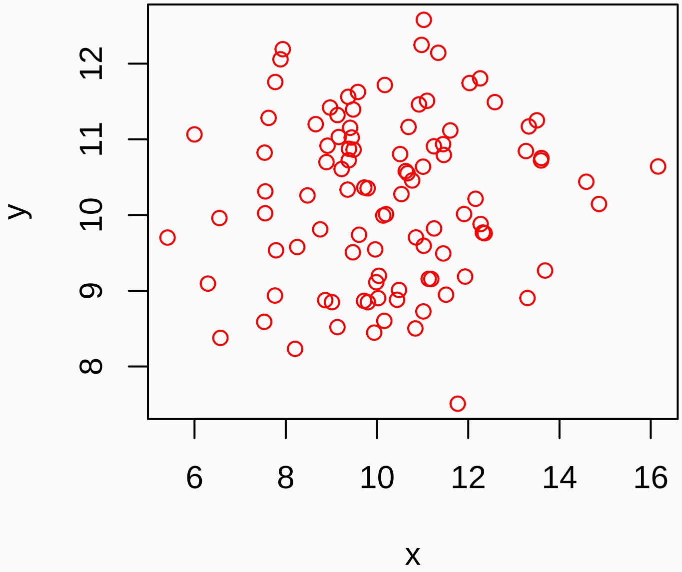


Figure 2: Gráfico de dispersión de dos variables que no se predicen mutuamente.

- Prueba estadística por defecto: Coeficiente de correlación de Pearson
- Estima cociente de covarianza y producto de desviación estándar:

$$r_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y} \tag{1}$$

$$\mathrm{cov}(X,Y) = \mathbb{E}[(X-\mu_x)(Y-\mu_y)] \tag{2}$$

Coeficiente de correlación paso a paso

El caso no espacial

Comenzamos con dos variables X y Y:

Table 1: Primeras seis filas de tabla que contiene *X* y *Y*.

x	y
6.567783	8.377792
9.385544	10.875437
10.025123	8.902984
9.477216	11.394629
10.691021	11.163742
11.768682	7.508639

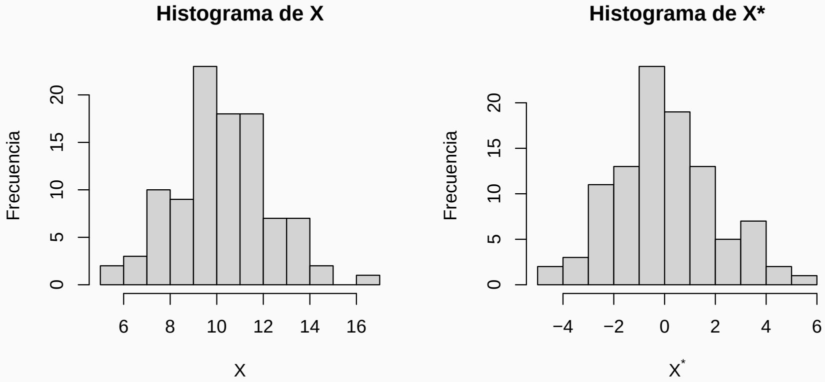
Covarianza entre X y Y

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mu_x)(Y - \mu_y)] \tag{3}$$

De adentro de paréntesis: - $\mu_x =$ Promedio de X; $\mu_y =$ Promedio de Y

- $\mu_x = 10.23; \mu_y = 10.22$
- Restamos μ de todos los valores de X y Y
- $X^* = X - \mu_x; Y^* = Y - \mu_y$

Covarianza entre X y Y



Entonces las variables centradas quedan así:

Table 2: Variables *X* y *Y* centradas (con media de 0).

x	y
-3.6683642	-1.8473745
-0.8506034	0.6502711
-0.2110244	-1.3221819
-0.7589320	1.1694628
0.4548737	0.9385757
1.5325346	-2.7165272
1.3698804	0.8929517
0.6172374	-0.5199712
0.3935418	0.3560402
3.3667466	0.5284583
2.0316619	1.9923603

- Multiplicamos cada valor X_i^* por su correspondiente Y_i^*

$$
\begin{aligned}
X_1^* \times Y_1^* &= -3.668 \times -1.847 = & 6.776 \\
X_2^* \times Y_2^* &= -0.85 \times -0.65 = & -0.553 \\
X_3^* \times Y_3^* &= -0.211 \times -1.322 = & 0.279 \\
&\vdots
\end{aligned}
$$

- Terminamos haciendo: $\frac{1}{n}\sum X_i^* Y_i^*$, es decir el promedio de los productos

Producto de las desviaciones estándar

El denominador de la fórmula para la correlación es:

$$\sigma_x \sigma_y$$

donde σ indica la desviación estándar de la variable en el subíndice.
Recordemos:

$$\sigma_x = \sqrt{\sum \frac{(X_i - \mu_x)^2}{n-1}} \tag{4}$$

Producto de las desviaciones estándar

Dado que ya contamos con $X_i - \mu_x$ y $Y_i - \mu_x$, sólo tenemos que hacer ${X^*}^2$:

$$
\begin{aligned}
{X_1^*}^2 = 6.776^2 = & \quad 45.92 \\
{X_2^*}^2 = -0.553^2 = & \quad 0.305 \\
{X_3^*}^2 = 0.279^2 = & \quad 0.077 \\
\vdots &
\end{aligned}
$$

Producto de las desviaciones estándar

Una vez, obtenidos X^{*2} y Y^{*2}, las sumamos y dividimos entre $n-1=99$:

$$\sum X_i^{*2}/99 = 391.1758/99 = 3.951$$
$$\sum Y_i^{*2}/99 = 119.9985/99 = 1.191$$

Y terminamos sacando las raíces cuadradas:

$$\sqrt{9.951} = 1.987$$
$$\sqrt{1.191} = 1.0911.987 \times 1.091 \qquad = 2.17$$

Una vez obtenidos:

- $\mathrm{cov}(X, Y) = 0.166$
- $\sigma_x \sigma_y = 2.17$

Tenemos:

$$\frac{\mathrm{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{0.166}{2.17} = 0.091$$

El resultado no es preciso por varias operaciones que obviaron decimales

Prueba de correlación en R

- Función para hacer prueba `cor.test`
- Uso:

```
cor.test(x, y)
```

- `x` y `y` son las variables X y X
  - Deben existir en el espacio de trabajo de R

Resultado de la prueba de correlación en R

```r
cor.test(x, y)
```

```
##
##  Pearson's product-moment correlation
##
## data:  x and y
## t = 0.76948, df = 98, p-value = 0.4435
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1207608  0.2698067
## sample estimates:
##       cor
## 0.0774955
```

Interpretación

- `t = 0.76948`, valor del estadístico T-Student
- `df = 98`, grados de libertad
- `p-value = 0.4435`, probabilidad de que $r = 0$
    - Probabilidad de que la correlación no exista
- `cor 0.0774955`, valor del coeficiente de correlación estimado