

Análisis de la asociación espacial

Correlación

Gerardo Martín

2022-06-29

Asociación estadística:

Probar la hipótesis de que dos variables se predicen mutuamente

Asociación espacial:

Variables con estructura espacial que se predicen mutuamente

Representación gráfica

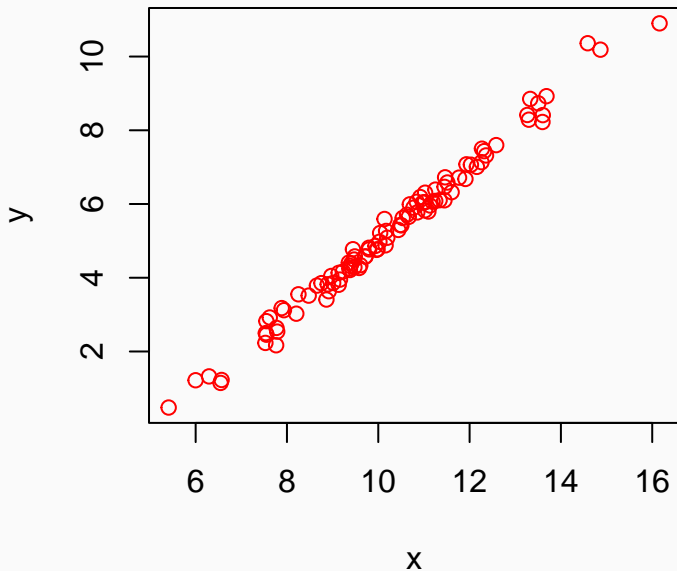


Figure 1: Gráfico de dispersión de dos variables que se predicen mutuamente.

Representación gráfica

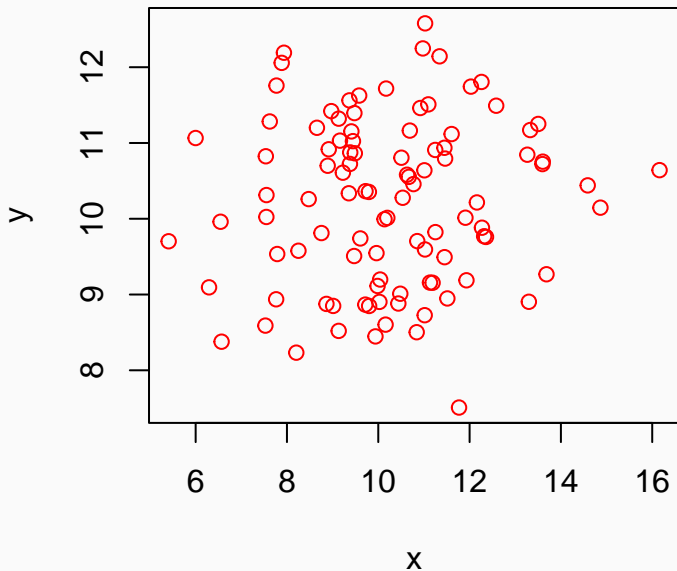


Figure 2: Gráfico de dispersión de dos variables que no se predicen mutuamente.

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

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

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

Coeficiente de correlación paso a paso

Comenzamos con dos variables X y Y :

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

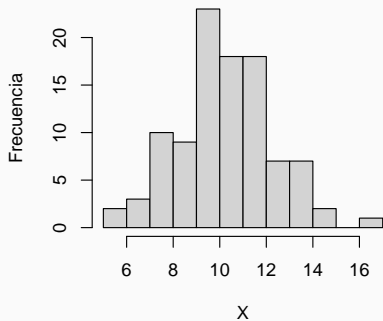
x	y
6.567783	8.377792
9.385544	10.875437
10.025123	8.902984
9.477216	11.394629
10.691021	11.163742
11.768682	7.508639

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

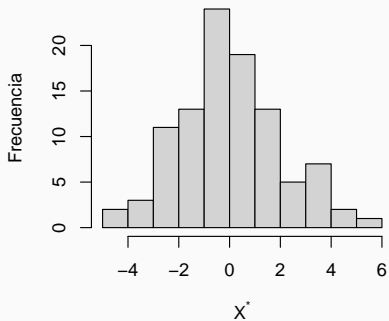
De adentro de paréntesis: - μ_x = Promedio de X ; μ_y = Promedio de Y

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

Histograma de X



Histograma de X^*



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

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

$$X_1^* \times Y_1^* = -3.668 \times -1.847 = 6.776$$

$$X_2^* \times Y_2^* = -0.85 \times -0.65 = -0.553$$

$$X_3^* \times Y_3^* = -0.211 \times -1.322 = 0.279$$

⋮

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

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:

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

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

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

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:

$$\sqrt{3.951} = 1.987$$

$$\sqrt{1.191} = 1.091 \times 1.987 = 2.17$$

Una vez obtenidos:

$$\cdot \text{cov}(X, Y) = 0.166$$

$$\cdot \sigma_x \sigma_y = 2.17$$

Tenemos:

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

El resultado no es preciso por varias operaciones que obviaron decimales

- 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

```
cor.test(x, y)

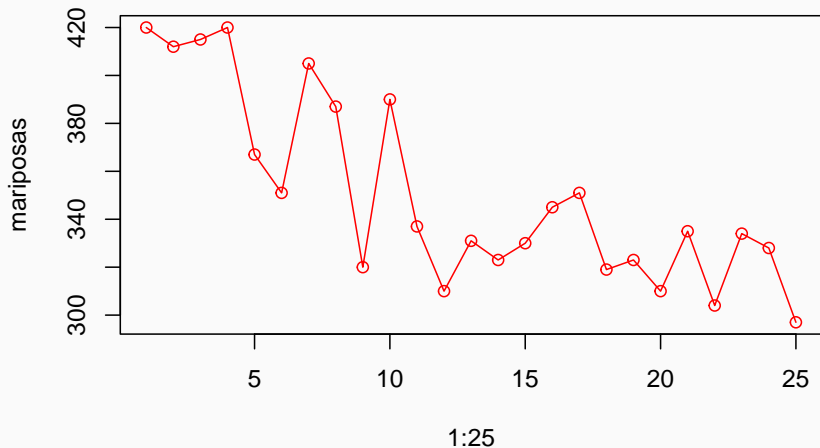
##
## Pearson's product-moment correlation
##
## data:  x and y
## t = 0.76948, df = 98, p-value = 0.4435
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1207608  0.2698067
## sample estimates:
##          cor
## 0.0774955
```

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

Relación entre poblaciones de mariposas, toneladas de pesticida y televisiones (Reino Unido)

- Toneladas de pesticida en aumento
- Densidad de mariposas disminuye
- Cantidad de licencias de televisión en aumento

Densidad de mariposas



Uso de pesticidas

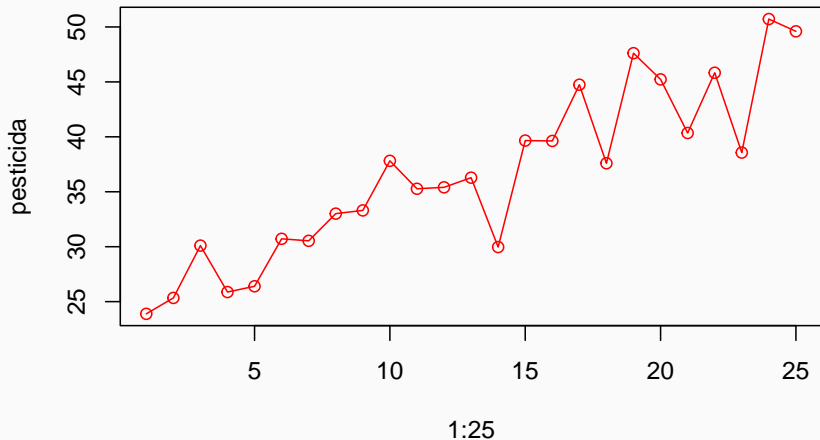


Figure 3: Porcentaje de área cubierta con pesticida.

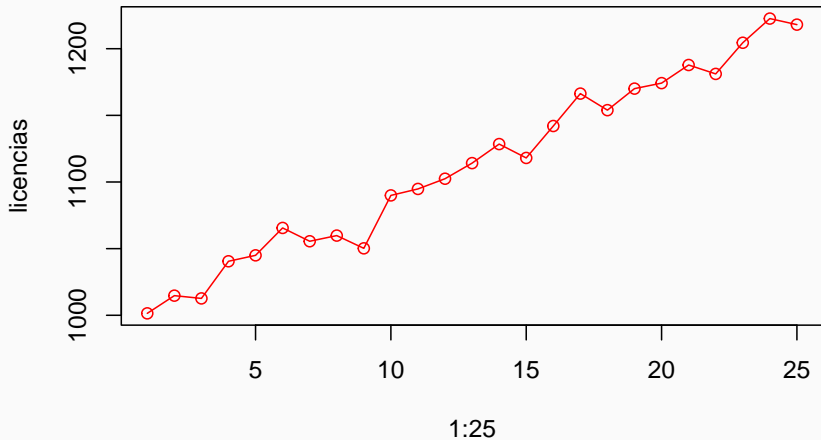


Figure 4: Número mensual promedio de licencias de televisión vendidas.

Limitaciones - Correlaciones entre variables

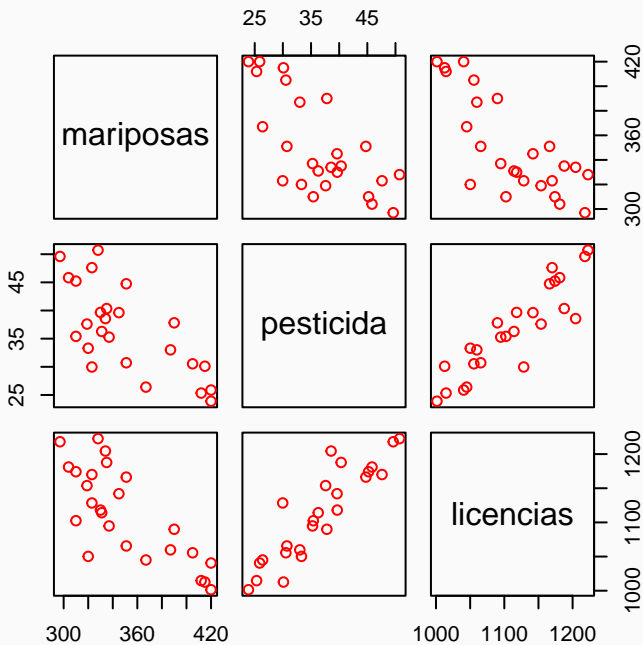


Table 3: Matriz de correlación entre todas las variables.

	mariposas	pesticida	licencias
mariposas	1.0000000	-0.7143064	-0.7764724
pesticida	-0.7143064	1.0000000	0.8986094
licencias	-0.7764724	0.8986094	1.0000000

La venta de licencias de televisión mata a las mariposas

- La correlación no mide causa y efecto
- Correlación puede ocurrir al azar
- Sólo sirve para medir asociación
- Interpretación de asociación → Conocimiento del fenómeno estudiado
 - Es más probable que, aunque la correlación sea menor, la disminución sea producto del uso de pesticidas