

## Clase práctica 25 (16-11-2017)

Alejandro Nasif Salum

Para las variables

$Y$ : tiempo de envío en milisegundos

$X$ : cantidad de paquetes adicionales al inicial (0 indica un único paquete de datos)

se quiere estudiar la correlación lineal y teniendo en cuenta que es esperable teóricamente la causalidad de  $X$  sobre  $Y$ , se busca estudiar el modelo de regresión lineal

$$Y = \beta_0 + \beta_1 X + u.$$

Se tienen los datos:

x	0	0	0	1	1	3	3	4	4	4	6	6	8	8	8
y	1.19	2.27	1.77	1.99	2.34	3.79	3.50	2.94	3.45	3.78	4.96	4.29	5.55	5.20	5.26

Se pide:

1. Graficar los puntos  $(x_1, y_1), \dots, (x_n, y_n)$ .
2. Calcular el coeficiente de correlación muestral  $r_{xy}$ .
3. Hallar la recta de mínimos cuadrados de la regresión de  $Y$  sobre la covariable  $X$ . Graficar sobre el diagrama del punto 1) dicha recta.
4. Calcular las sumas de cuadrados total (SST), de la regresión (SSR) y del error (SSE) y hallar el  $R^2 = 1 - \frac{SSE}{SST}$ . Verificar que  $R^2 = (r_{xy})^2$ .
5. Hallar intervalos de confianza de nivel 0,95 para  $\beta_0$  y  $\beta_1$ .
6. Contrastar las hipótesis  $H_0: \beta_1 = 0$  vs,  $H_1: \beta_1 \neq 0$ . Hallar el valor  $p$ .

Tener en cuenta las siguientes funciones:

- `plot`, en su forma `plot(x,y)`, con `x` e `y` vectores de igual longitud.
- `abline`.
- `pt`, `qt` (para la distribución *t-Student*).

### Para aprender un poco más...

— Se sugiere correr el siguiente código, donde se entiende que  $y$  y  $x$  son los nombres de las correspondientes variables, como vectores de  $\mathbb{R}$ .

```
modelo<-lm(y~x)
modelo
summary(modelo)
```

— Intentar identificar las respuestas de los puntos anteriores (no están los intervalos, pero sí toda la información para el cálculo) para así interpretar el resultado.

— La especificación

$y \sim x$

determina un modelo como el que se planteó. Sin embargo, pueden intentarse modelos más complejos. Consideremos la variable  $Z$  con valores

z	0	1	0	2	0	3	2	3	2	2	4	4	5	2	5
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Repetir lo anterior para los modelos

$y \sim x+z$ ,

$y \sim x+x^2$ ,

$y \sim x*z$ .

Entre otras posibilidades. Y analizar cuál es la expresión completa del modelo (por ejemplo, en el último caso no se considera solo una variable  $x \cdot z$ , sino que se trata de un modelo más complejo).