

Unidad 3 - Modulo 1

Pablo Anicich

5 de julio de 2016

Ejercicio 1.2.9

Se tienen los datos de los siguientes diez países:

Pais	PBI	Población	PBI.x.Cápita
Estados Unidos	16799700	321234000	0.052
China	9181377	1370793000	0.007
Japón	4901532	126806000	0.039
Alemania	3635959	81276000	0.045
Francia	2737361	64297000	0.043
Reino Unido	2535761	64915000	0.039
Brasil	2242854	204519000	0.011
Rusia	2118006	144031000	0.015
Italia	2071955	60963000	0.034
India	1870651	1299499000	0.001

Dada una muestra de las variables aleatorias $\{x_i\}$ y $\{y_i\}$ la correlación entre ambas se estima usando:

$$C = \frac{\sum_{k=1}^n (x_i - x_{medio})(y_i - y_{medio})}{\sqrt{\sum_{k=1}^n (x_i - x_{medio})^2} \sqrt{\sum_{k=1}^n (y_i - y_{medio})^2}}$$

Entonces, la matriz de correlaciones para la tabla de arriba es:

$$\begin{pmatrix} 1 & 0.2084859 & 0.3361903 \\ 0.2084859 & 1 & -0.6997094 \\ 0.3361903 & -0.6997094 & 1 \end{pmatrix}$$

Ejercicios adicionales

Para cada conjunto de datos:

1. Decidir el tipo de correlación que se va a intentar y/o si hay que linealizar transformando Y.
2. Calcular los σ_{ij} correspondientes.
3. Plantear las fórmulas necesarias resolviendo si hace falta el sistema de ecuaciones dejado como ejercicio (o buscando por internet).
4. Calcular los valores de los parámetros.
5. Verificar haciendo la regresión con Excel (o con el paquete de elección).

Ej01

1. Correlación lineal.
2. $\sigma_{jl} = \sum_{i=1}^n x_i^j y_i^l$.
 - $\sigma_{10} = 5050$.

- $\sigma_{01} = 1.7190432 \times 10^4$.
- $\sigma_{11} = 1.1348696 \times 10^6$
- $\sigma_{20} = 338350$.

3. Cálculo de la recta de mínimos cuadrados $y = ax + b$. La recta de cuadrados mínimos se estima a partir de: $\chi^2 = \sum_{i=1}^n (y_i - ax_i - b)^2$.

Tomando las derivadas parciales e igualando a cero se tiene:

$$\begin{pmatrix} \sigma_{20} & \sigma_{10} \\ \sigma_{10} & N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sigma_{11} \\ \sigma_{01} \end{pmatrix}$$

resolviendo las ecuaciones anteriores se llega a:

$$a = \frac{\sigma_{10}\sigma_{01} - N\sigma_{11}}{\sigma_{10}^2 - N\sigma_{20}}$$

$$b = \frac{\sigma_{01} - a\sigma_{10}}{N}$$

4. Resolviendo las ecuaciones anteriores, para este caso se obtienen:

$$a = 3.2013538$$

$$b = 10.235951$$

5. Resolución usando R:

```
##
## Call:
## lm(formula = Y ~ X, data = Ej01)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.2142  -3.5365  -0.1745   3.2362  13.5826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.23595    1.07333   9.537 1.23e-15 ***
## X           3.20135    0.01845 173.493 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.326 on 98 degrees of freedom
## Multiple R-squared:  0.9968, Adjusted R-squared:  0.9967
## F-statistic: 3.01e+04 on 1 and 98 DF,  p-value: < 2.2e-16
```

Datos y regresión lineal

