

Modelo de regresión lineal múltiple con errores normales

Sea $n \in \mathbb{N}$, \mathbf{X} matriz de constantes $n \times (k + 1)$ dimensional con rango $k + 1 < n$,

$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)' \in \mathbb{R}^{k+1}$ y $\sigma^2 \in \mathbb{R}$. Consideramos el modelo

$$\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}),$$

equivalentemente

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

con

$$\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}),$$

En clase vimos que los estimadores máximo verosimil para $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ son $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$ con

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \hat{\sigma}^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\end{aligned}$$

Por ejemplo, consideramos

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{pmatrix}$$

y

$$\boldsymbol{\beta} = \begin{pmatrix} 1.1 \\ -2.5 \\ 11 \\ -18 \\ 10 \end{pmatrix}$$

```
In [5]: using Distributions # Paquete con distribuciones de probabilidad
n = 100 # Consideramos 100 observaciones
β = [ 1.1, -2.5, 11.0, -18.0, 10.0] # Vector β
x = rand(Uniform(),n) # n puntos aleatorios uniformes en (0,1), esta sería la pri
X = zeros(n,5) # Matriz de ceros para construir X
for i in 1:n
    X[i,:] = [ 1, x[i], x[i]^2, x[i]^3, x[i]^4 ] # iteramos para construir renglon
end
X
```

```
Out[5]: 100x5 Array{Float64,2}:
 1.0  0.0751632  0.00564951  0.000424635  3.19169e-5
 1.0  0.970433  0.94174      0.913895   0.886874
 1.0  0.516609  0.266885   0.137875   0.0712277
 1.0  0.95873   0.919163   0.881228   0.84486
 1.0  0.725705  0.526648   0.382191   0.277358
 1.0  0.260591  0.0679079  0.0176962  0.00461148
 1.0  0.94333   0.889872   0.839443   0.791873
 1.0  0.318422  0.101393   0.0322856  0.0102805
 1.0  0.928333  0.861803   0.80004     0.742704
 1.0  0.736149  0.541915   0.39893     0.293672
 1.0  0.951602  0.905545   0.861718   0.820013
 1.0  0.131648  0.0173311  0.0022816  0.000300367
 1.0  0.424466  0.180171   0.0764766  0.0324617
 ⋮
 1.0  0.357086  0.12751    0.0455321  0.0162589
 1.0  0.168822  0.0285007  0.00481154 0.000812293
 1.0  0.598327  0.357995   0.214198   0.12816
 1.0  0.829282  0.687709   0.570305   0.472943
 1.0  0.9105    0.82901    0.754814   0.687258
 1.0  0.93165   0.867972   0.808646   0.753375
 1.0  0.58738   0.345015   0.202655   0.119035
 1.0  0.407024  0.165669   0.0674311  0.0274461
 1.0  0.478355  0.228824   0.109459   0.0523602
 1.0  0.690414  0.476672   0.329101   0.227216
 1.0  0.278767  0.0777111  0.0216633  0.00603901
 1.0  0.53553   0.286792   0.153586   0.0822499
```

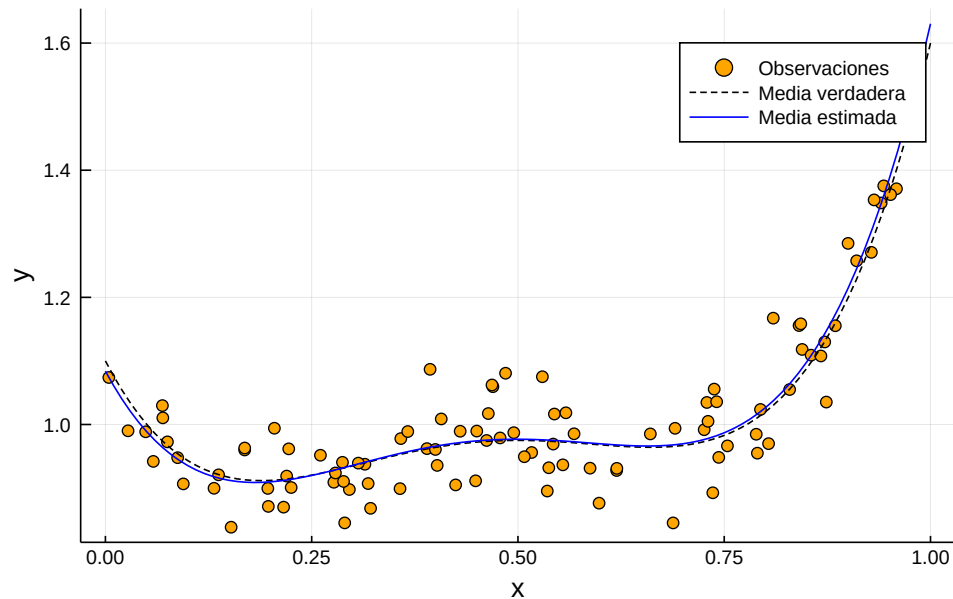
```
In [27]: ε = rand(Normal(0,0.05),n) # Vector de errores normales con media μ=0 y varianza σ²=0.05
y = X * β + ε # Observaciones provenientes del modelo
β ml = ( X' * X )^(-1) * X' * y
```

```
Out[27]: 5-element Array{Float64,1}:
 1.083888379682323
 -2.4028418162868106
 10.860178346395301
 -18.02078565589942
 10.109959797886674
```

```
In [32]: using Plots # Paquete para producir imágenes
f(x) = β[1] + β[2]*x + β[3]*x^2 + β[4]*x^3 + β[5]*x^4 # Media polinomial verdadera
f_ml(x) = β_ml[1] + β_ml[2]*x + β_ml[3]*x^2 + β_ml[4]*x^3 + β_ml[5]*x^4 # Media p
mesh = collect(0.0:1.0/100.0:1.0)
scatter(x,y,color="orange",label="Observaciones")
plot!(mesh,f.(mesh), color = :black, linestyle=:dash ,label="Media verdadera")
plot!(mesh,f_ml.(mesh), color = :blue, label="Media estimada")
ylabel!("y")
xlabel!("x")
title!("Variables observadas vs covariables")
```

Out[32]:

Variables observadas vs covariables



Teorema Sean $\hat{\beta}, \hat{\sigma}^2$ los estimadores máximo verosímil en el modelo de regresión lineal múltiple con errores normales, entonces

i) $\hat{\beta} \sim N_{k+1}(\beta, \sigma^2(X'X)^{-1})$.

ii) $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n - k - 1)$.

iii) $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes.

dem.

i) $\hat{\beta} = Ay$ con $A = (X'X)^{-1}X'$, entonces, por teorema anteriormente visto en clase,

$\hat{\beta} \sim N_{k+1}(AE[y], A \text{cov}(y)A')$. $AE[y] = AX\beta = \beta$ por otro lado

$A \text{cov}(y)A' = \sigma^2 AA' = \sigma^2(X'X)^{-1}$ entonces $\hat{\beta} \sim N_{k+1}(\beta, \sigma^2(X'X)^{-1})$.

ii) Sabemos que si $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces $\mathbf{y}'\mathbf{A}\mathbf{y} \sim \chi^2\left(r, \frac{\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}}{2\sigma^2}\right)$ si sólo si \mathbf{A} es idempotente de rango r .

Tenemos que $n\hat{\sigma}^2 = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$; al ser $(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')^2$ idempotente se sigue que $(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ es idempotente, de lo cual también se tiene que $\text{rango}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{tr}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = n - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = n - (k + 1)$. Esto concluye ii).

iii) Sabemos que si $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces $\mathbf{y}'\mathbf{A}\mathbf{y}$ es independiente de $\mathbf{B}\mathbf{y}$ si sólo si $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$. Tenemos que $\hat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{y}$ con $\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ y $n\hat{\sigma}^2 = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$, por lo que basta ver que $\mathbf{B}\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{0}$.

Coeficiente de determinación

In []: