Modelo de regresión lineal múltiple con errores normales

Sea $n \in \mathbb{N}$, $\boldsymbol{X}$ 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

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

equivalentemente

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

con

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

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

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

Por ejemplo, consideramos

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

y

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

In [37]:
```
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 renglo
end
X
```

Out[37]:
```
100×5 Array{Float64,2}:
 1.0  0.227922   0.0519485   0.0118402    0.00269865
 1.0  0.30956    0.0958271   0.0296642    0.00918284
 1.0  0.225516   0.0508576   0.0114692    0.0025865
 1.0  0.886644   0.786138    0.697025     0.618013
 1.0  0.778182   0.605567    0.471241     0.366711
 1.0  0.426053   0.181521    0.0773377    0.0329499
 1.0  0.583341   0.340287    0.198503     0.115795
 1.0  0.659724   0.435236    0.287135     0.18943
 1.0  0.380665   0.144906    0.0551606    0.0209977
 1.0  0.594962   0.35398     0.210605     0.125302
 1.0  0.349628   0.12224     0.0427385    0.0149426
 1.0  0.505719   0.255752    0.129339     0.0654091
 1.0  0.972006   0.944795    0.918346     0.892638
 ⋮
 1.0  0.991558   0.983188    0.974888     0.966659
 1.0  0.690162   0.476324    0.328741     0.226884
 1.0  0.32656    0.106642    0.0348249    0.0113724
 1.0  0.751791   0.565189    0.424904     0.319439
 1.0  0.350597   0.122918    0.0430946    0.0151088
 1.0  0.861125   0.741536    0.638555     0.549875
 1.0  0.204977   0.0420154   0.00861218   0.00176529
 1.0  0.632591   0.400171    0.253144     0.160137
 1.0  0.421625   0.177768    0.0749513    0.0316014
 1.0  0.0975471  0.00951544  0.000928204  9.05436e-5
 1.0  0.793839   0.63018     0.500262     0.397127
 1.0  0.558028   0.311395    0.173767     0.0969671
```

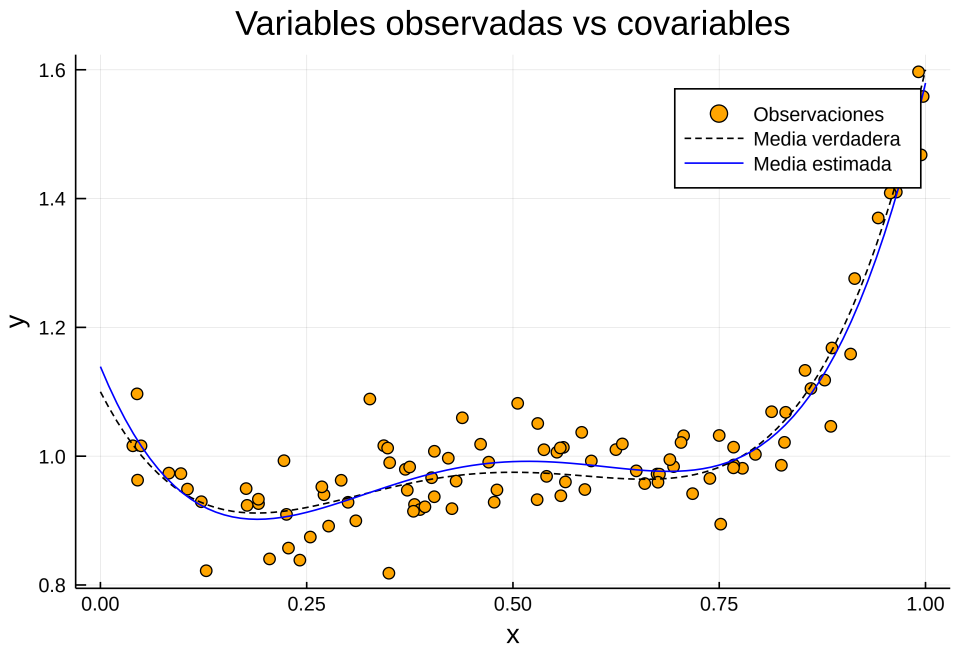
In [38]:
```
ϵ = rand(Normal(0,0.05),n) # Vector de errores normales con media μ=0 y varianza 
y = X * β + ϵ # Observaciones provenientes del modelo
β_ml = ( X' * X)^(-1) * X' * y
```

Out[38]:
```
5-element Array{Float64,1}:
   1.139061647617389
  -3.105966702828301
  13.414753308858511
 -21.30257677825514
  11.434139684381392
```

In [39]:

```
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 verdader
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[39]:



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

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

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

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

dem.

i) $\hat{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{y}$ con $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, entonces, por teorema anteriormente visto en clase, $\hat{\boldsymbol{\beta}} \sim N_{k+1}(\boldsymbol{A}\mathbb{E}[y]\,, \boldsymbol{A}\,\mathrm{cov}(y)\boldsymbol{A}')$. $\boldsymbol{A}\mathbb{E}[\boldsymbol{y}] = \boldsymbol{A}X\boldsymbol{\beta} = \boldsymbol{\beta}$ por otro lado $\boldsymbol{A}\,\mathrm{cov}(y)\boldsymbol{A}' = \sigma^2\boldsymbol{A}\boldsymbol{A}' = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ entonces $\hat{\boldsymbol{\beta}} \sim N_{k+1}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$.

ii) Sabemos que si $\boldsymbol{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \sim \chi^2\left(r, \frac{\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}}{2\sigma^2}\right)$ si sólo si $\boldsymbol{A}$ es idempotente de rango r. Tenemos que $n\hat{\sigma}^2 = \boldsymbol{y}'\left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)\boldsymbol{y}$; al ser $\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)^2$ idempotente se sigue que $\left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)$ es idempotente, de lo cual también se tiene que $\text{rango}\left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right) = \text{tr}\left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right) = n - \text{tr}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\right) = n - (k+1)$. Esto concluye ii).

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

Coeficiente de determinación

Recordemos que

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = (\boldsymbol{y} - \bar{y}\boldsymbol{j})'(\boldsymbol{y} - \bar{y}\boldsymbol{j}) = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2$$

(desarrollando el cuadrado)

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}) = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'X\boldsymbol{\beta} = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'y$$

(usando las ecuaciones normales)

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = (\hat{\boldsymbol{y}} - \bar{y}\boldsymbol{j})'(\hat{\boldsymbol{y}} - \bar{y}\boldsymbol{j})$$

son tales que

$$SST = SSE + SSR$$

por lo que

$$SSR = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'y - n\bar{y}^2$$

El coeficiente de determinación R^2 se define por

$$R^2 = \frac{SSR}{SST} = \frac{\hat{\boldsymbol{\beta}}'\boldsymbol{X}'y - n\bar{y}^2}{\boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2}$$

Habíamos visto que SSR puede ser expresado en términos de la matrix centradora

$$\boldsymbol{X}_c = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix}$$

como $SSR = \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y}$. Vimos que $\boldsymbol{X}_c$ y $\hat{\boldsymbol{\beta}}_1$ satisfacen las ecuaciones normales $\boldsymbol{X}_c' \boldsymbol{y} = \boldsymbol{X}_c' \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_1$ por lo que

$$SSR = \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{y} = \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_1$$

Se sigue que el coeficiente de determinación puede ser escrito en términos de la matriz centradora que es computacionalmente más estable

$$R^2 = \frac{SSR}{SST} = \frac{\hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_1}{\boldsymbol{y}' \boldsymbol{y} - n\bar{y}^2}$$

In [ ]:

In [ ]: