Pruebas de hipótesis para el modelo de regresión lineal múltiple

En esta sección suponemos que $\boldsymbol{y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$ con $\boldsymbol{X}$ $n \times (k+1)$ dimensional de rango $k+1 < n$.

Prueba de regresión general

Empezamos intereándonos en la hipótesis de que ninguno de los covariables x considerados predicen a la variable de interés y en el modelo de regresión lineal múltiple con errores normales. En términos matemáticos esto significa que $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \ldots, \beta_k)' = \boldsymbol{0}$. Obtenemos la prueba de hipótesis:

$$H_0 \ : \ \boldsymbol{\beta}_1 = \boldsymbol{0} \qquad \text{v.s.} \qquad H_1 \ : \ \boldsymbol{\beta}_1 \neq \boldsymbol{0}.$$

En las notas correspondientes a las distribuciones no centrales se demostró que para $\lambda_1 = \frac{\boldsymbol{\beta}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \boldsymbol{\beta}_1}{2\sigma^2}$

$$F = \frac{SSR/\left(\sigma^2 k\right)}{SSE/\left(\sigma^2(n-k-1)\right)} = \frac{SSR/k}{SSE/(n-k-1)} \sim F(k, n-k-1, \lambda_1)$$

Teorema

Si H_0 es cierta, es decir $\boldsymbol{\beta}_1 = \boldsymbol{0}$, entonces $\lambda_1 = 0$ y $F \sim F(k, n-k-1)$; y si H_0 es falsa, es decir $\boldsymbol{\beta}_1 \neq \boldsymbol{0}$, entonces $\lambda_1 = \frac{\boldsymbol{\beta}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \boldsymbol{\beta}_1}{2\sigma^2}$ y $F \sim F(k, n-k-1, \lambda_1)$.

Observe que $\lambda_1 = 0$ si y sólo si $\boldsymbol{\beta}_1 = \boldsymbol{0}$ dado que $\boldsymbol{X}_c' \boldsymbol{X}_c$ es positio definida. Podemos usar F como cantida pivvotal para realizar la prueba de hipótesis como sigue:

Si $F > F_{\alpha,k,n-k-1}$, con $F_{\alpha,k,n-k-1}$ tal que $\mathbb{P}\left[F > F_{\alpha,k,n-k-1} | \boldsymbol{\beta}_1 = \boldsymbol{0}\right] = \alpha$, entonces rechazamos H_0.

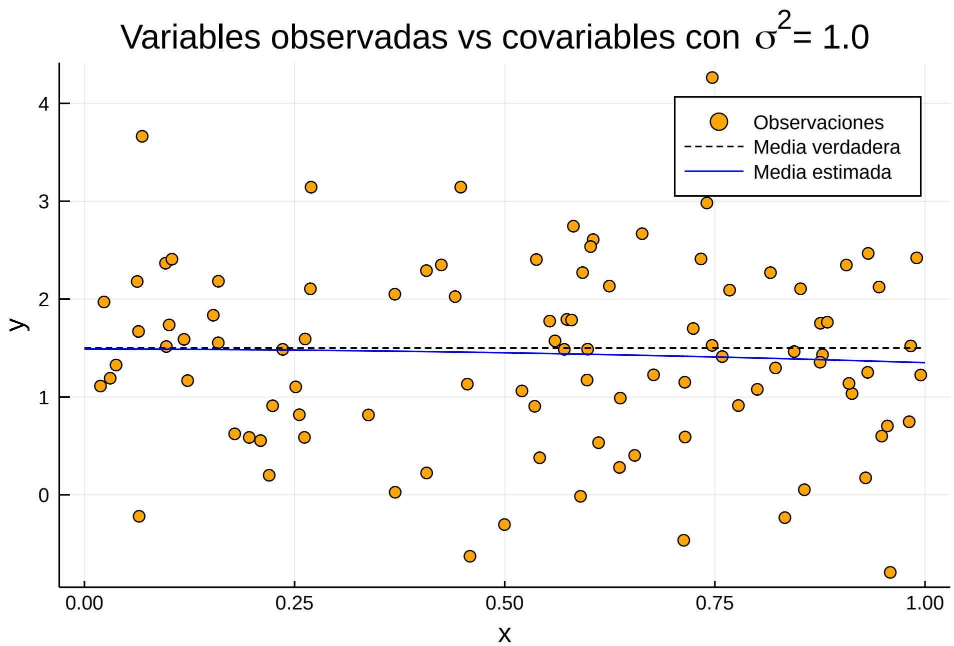
A continuación consideramos la prueba anterior para el caso en que $\boldsymbol{\beta}_1 = \boldsymbol{0}$ y ajustamos un modelo con $\boldsymbol{\beta}_1 \neq \boldsymbol{0}$ y para el caso en que $\boldsymbol{\beta}_1 \neq \boldsymbol{0}$ y ajustamos un modelo con $\boldsymbol{\beta}_1 \neq \boldsymbol{0}$.

In [2]:
```
using Distributions # Paquete con distribuciones de probabilidad
using Plots # Paquete para producir imágenes
using LaTeXStrings # Paquete para usar latex en strings
```

In [4]:
```
n = 100 # Consideramos 100 observaciones
x = rand(Uniform(),n) # n puntos aleatorios uniformes en (0,1), esta sería la pri
X = zeros(n,3) # Matriz de ceros para construir X
for i in 1:n
    X[i,:] = [ 1, x[i], x[i]^2 ] # iteramos para construir renglones de X
end
```

In [6]:
```
ϵ = rand(Normal(0,1.0),n) # Vector de errores normales con media μ=0 y varianza σ
y = 1.5 .+ ϵ # Observaciones provenientes del modelo con varianza 0.1
β_ml = ( X' * X)^(-1) * X' * y
f(x) = 1.5
f_ml(x) = β_ml[1] + β_ml[2]*x + β_ml[3]*x^2.0
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 con \\sigma^{2}  = 1.0")
```

Out[6]:



Si calculamos el estadístico de prueba F

In [11]:
```
SSR = β_ml' * X' * y - n*mean(y)^2.0
SSE = y'*y - β_ml' * X' * y
F = (n-2-1)*SSR/( 2* SSE )
Fα = cquantile( FDist( 2, n-2-1), 0.05 );
```

Out[11]:
```
3.0901866751548672
```

Vemos que

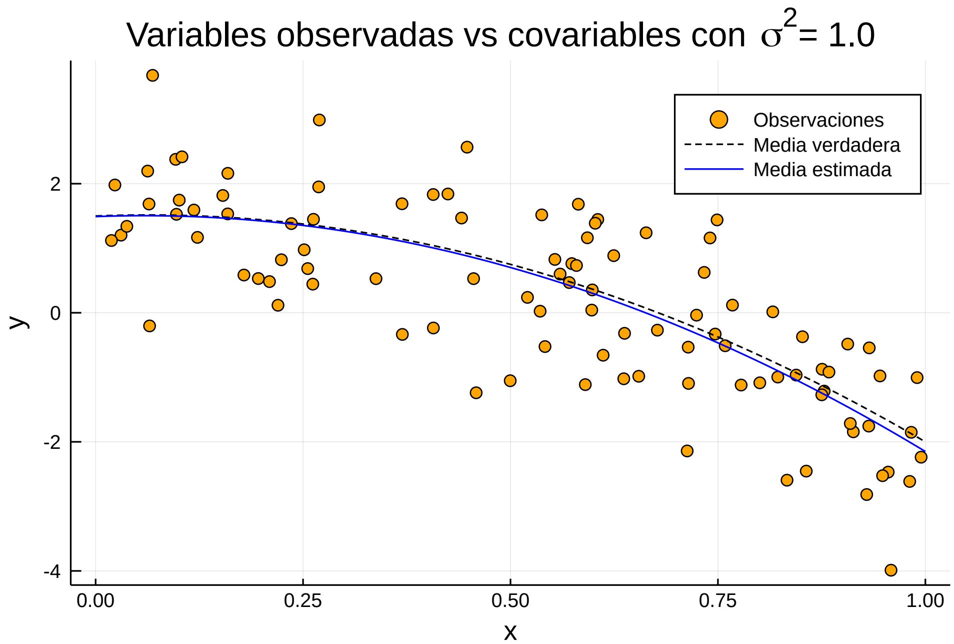
In [14]:
```
F >= Fα
```

Out[14]:
```
false
```

```
Por lo que no rechazamos $H 0 : \pmb{\beta} 1=\pmb{0}$.
```

In [18]:
```
y = 1.5 .+ 0.5.*x .- 4.0.*x.^2.0 .+ ϵ # Observaciones provenientes del modelo con
β_ml = ( X' * X)^(-1) * X' * y
f(x) = 1.5 + 0.5*x - 4.0.*x^2.0
f_ml(x) = β_ml[1] + β_ml[2]*x + β_ml[3]*x^2.0
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 con \\sigma^{2}  = 1.0")
```

Out[18]:



In [19]:
```
SSR = β_ml' * X' * y - n*mean(y)^2.0
SSE = y'*y - β_ml' * X' * y
F = (n-2-1)*SSR/( 2* SSE )
Fα = cquantile( FDist( 2, n-2-1), 0.05 );
```

Vemos que

In [20]:
```
F >= Fα
```

Out[20]:
```
true
```

Por lo que rechazamos $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$.