Especificación erronea del modelo

Particionemos el modelo de regresión lineal múltiple $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ en la siguiente manera

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= (\boldsymbol{X}_1,\ \boldsymbol{X}_2) \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\epsilon} \\ &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \end{aligned}$$

Si realizamos inferencia sin incluir $\boldsymbol{X}_2\boldsymbol{\beta}_2$ cuando los datos provienen del modelo con $\boldsymbol{\beta}_2 \neq \boldsymbol{0}$ entonces decimos que estamos subajustando (underfitting). Si incluimos $\boldsymbol{X}_2\boldsymbol{\beta}_2$ cuando los datos provienen del modelo con $\boldsymbol{\beta}_2 = \boldsymbol{0}$ decimos que estamos sobreajustando (overfitting).

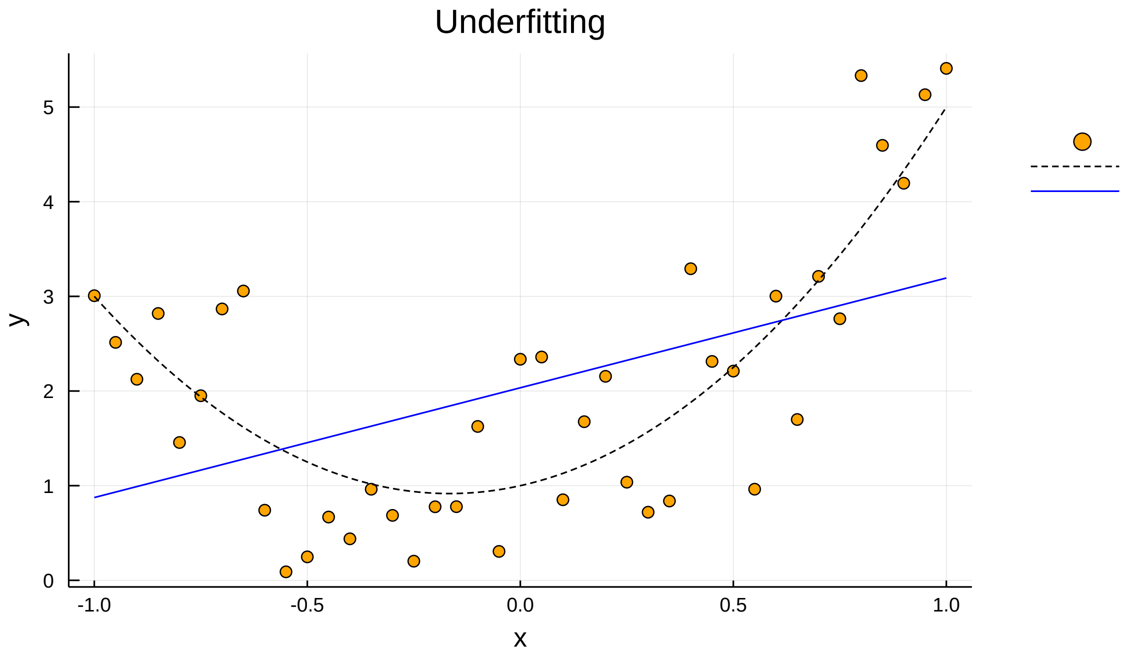
Esto puede ser visualizado en el ejemplo de regresión polinomial.

In [1]:
```
using Distributions
y = [ 1.0 + x + 3.0*x^2 for x in -1:0.05:1 ] .+ rand(Normal(),41)
x = collect(-1:0.05:1)
X = zeros(41,7) # Matriz de ceros para construir X
for i in 1:41
    X[i,:] = [ 1, x[i], x[i]^2, x[i]^3, x[i]^4, x[i]^5, x[i]^6 ] # iteramos para 
end
X_1 = X[:,1:2];
```

In [2]:
```
using Plots, Measures #, Plots.PlotMeasures # Paquete para producir imágenes
default(size = (900, 400))
f(x) = 1.0 + x + 3.0*x^2 # Media cuadrática verdadera
β_u = ( X_1' * X_1)^(-1) * X_1' * y
β_o = ( X' * X)^(-1) * X' * y;
```

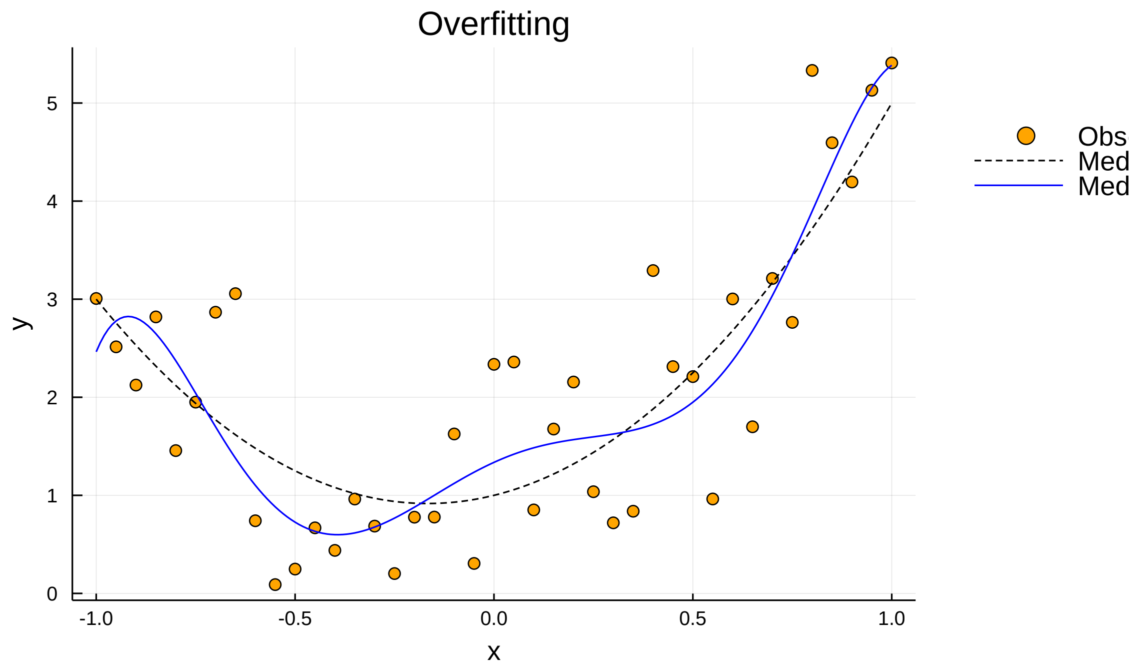
In [3]:
```
# Underfitting
f_u(x) = β_u[1] + β_u[2]*x # Media lineal dada por máxima verosimilitud al subest
mesh = collect(-1.0:1.0/100.0:1.0)
scatter(x,y,color="orange",label="Observaciones")
plot!(mesh,f.(mesh), color = :black, linestyle=:dash ,label="Media cuadrática ver
plot!(mesh,f_u.(mesh), color = :blue, label="Media lineal estimada", legend=:oute
ylabel!("y")
xlabel!("x")
title!("Underfitting")
```

Out[3]:



In [4]:
```
# Overfitting
f_o(x) = β_o[1] + β_o[2]*x + β_o[3]*x^2 + β_o[4]*x^3 + β_o[5]*x^4 + β_o[6]*x^5 + β
mesh = collect(-1.0:1.0/100.0:1.0)
scatter(x,y,color="orange",label="Observaciones")
plot!(mesh,f.(mesh), color = :black, linestyle=:dash ,label="Media cuadrática verd
plot!(mesh,f_o.(mesh), color = :blue, label="Media polinomial grado 6 estimada", l
ylabel!("y")
xlabel!("x")
title!("Overfitting")
```

Out[4]:



Ortogonalización

Recordemos que al considerar dos modelos:

$$\boldsymbol{y} = \boldsymbol{X}_1 \boldsymbol{\beta}_1^\star + \boldsymbol{\epsilon}^\star$$

y

$$\boldsymbol{y} = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

no se tiene necesariamente que los estimadores OLS satsifacen $\hat{\boldsymbol{\beta}}_1^\star = \hat{\boldsymbol{\beta}}_1$. Observemos que si usamos el modelo con $\boldsymbol{\beta}_1^\star$ cuando el modelo verdadero es el que contiene $\boldsymbol{\beta}_1$ y $\boldsymbol{\beta}_2$ en las ecuaciones anteriores, se tiene que

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\beta}_1^\star\right] &= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\mathbb{E}[y] \\
&= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2) \\
&= \boldsymbol{\beta}_1 + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2
\end{aligned}$$

por lo que la estimación de $\boldsymbol{\beta}_1$ se ve afectada por $\boldsymbol{X}_2$ a menos que $\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{0}$. En el siguiente teorema vemos que si $\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{0}$ entonces $\hat{\boldsymbol{\beta}}_1^\star$ y $\hat{\boldsymbol{\beta}}_1$ más que tener el mismo valor esperado, son iguales.

Teorema Si $\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{0}$ entonces $\boldsymbol{\beta}_1$, en el modelo completo $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, es igual a $\boldsymbol{\beta}_1^\star$, en el modelo reducido $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1^\star + \boldsymbol{\epsilon}^\star$.

Antes de proceder a la demostración enunciamos el siguiente Lemma para trabajar con la inversión de matrices expresadas en bloques

Lemma Si

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{1,2} & \boldsymbol{A}_{1,2} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} \end{pmatrix}$$

es no singular, simétrica y tal que $\boldsymbol{A}_{1,1}^{-1}$ existe, y $\boldsymbol{B} = \boldsymbol{A}_{2,2} - \boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}$ es invertible, entonces

$$\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{A}_{1,1}^{-1} + \boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\boldsymbol{B}^{-1}\boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1} & -\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\boldsymbol{B}^{-1} \\ -\boldsymbol{B}^{-1}\boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1} & \boldsymbol{B}^{-1} \end{pmatrix}$$

Prueba del teorema

De $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X}) - 1\boldsymbol{X}'\boldsymbol{y}$ se sigue que

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{y} \end{pmatrix}$$

Usando el lemma anterior tenemos que

$$\hat{\boldsymbol{\beta}}_1 = \left(\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{B}^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \right)\boldsymbol{X}_1'\boldsymbol{y} - \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{B}^{-1}\boldsymbol{X}_2'\boldsymbol{y}$$

Por lo que $\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{0}$ implica

$$\hat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{y} = \hat{\boldsymbol{\beta}}_1^\star$$

El resultado anterior será de ultilidad para realizar pruebas de hipótesis.