2.1.3 Formas funcionales cuadráticas

La regresión múltiple permite establecer relaciones funcionales de una variable que no se pueden tratar o modelizar mediante la regresión simple.

Supongamos una relación cuadrática del siguiente tipo:

$$Y = \beta_0 - \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon \quad \longrightarrow \quad \frac{\partial Y}{\partial X_1} = -\beta_1 + 2\beta_2 X_1$$

Para estudiar la variación esperada en Y tras un cambio unitario en X_1 derivamos:

$$dY = -\beta_1 dX_1 + 2\beta_2 X_1 dX_1 = dX_1\left(-\beta_1 + 2\beta_2 X_1\right)$$

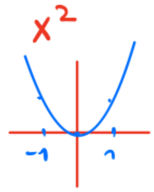
y sustituyendo los diferenciales por incrementos

$$\Delta Y = \left(-\beta_1 + 2\beta_2 X_1\right)\Delta X_1$$

de manera que el incremento de la variables explicada depende del incremento de la variable independiente pero también de su nivel.

Igualando a cero obtenemos su máximo o mínimo:

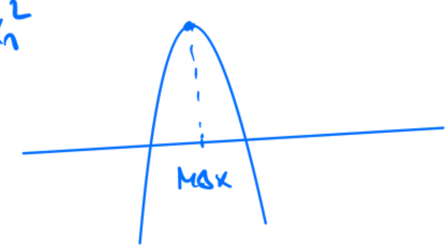
$$-\beta_1 + 2\beta_2 X_1 = 0;.$$



Para ese nivel inicial de X_1 existe:

- Un Mínimo: si $\beta_2 > 0$ $\longrightarrow$ $\oplus \beta_2 X_1^2$
- Un Máximo: si $\beta_2 < 0$

$\longrightarrow \ominus \beta_2 X_1^2$



EJERCICIO 1: Salarios en el sector turístico español

Estimamos el modelo en que el salario hora en el sector turístico español depende, con una relación del nivel de estudios acabados, y también de la antigüedad en la empresa.

Modelo poblacional:

$$salario = \beta_0 + \beta_1 estudios + \beta_2 estudios^2 + \beta_3 antigüedad + \beta_4 antigüedad^2 + \varepsilon,$$

Y su estimación (FRM):

$\hat{\beta}_2$

$$\widehat{salario}_i = 8,04 - 0,385 \cdot estudios_i + 0,189 \cdot estudios_i^2$$
$$+ 0,299 \cdot antigüedad_i - 0,0017 \cdot antigüedad_i^2,$$

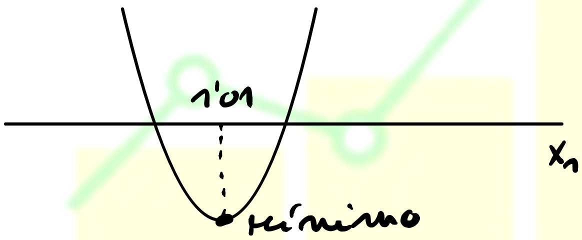
$$n = 5286, R^2 = 0,2165, \bar{R}^2 = 0,2159.$$

a) ¿En qué nivel de estudios se alcanza el salario mínimo fijada la antigüedad?

$\hookrightarrow \hat{\beta}_2 > 0$

$$\frac{\partial \widehat{salario}}{\partial estudios} = -0'385 + 2 \cdot 0'189 \cdot estudios = 0$$

$$2 \cdot 0'189 \cdot estudios = 0'385 \rightarrow estudios = \frac{0'385}{2 \cdot 0'189} = 1'01$$

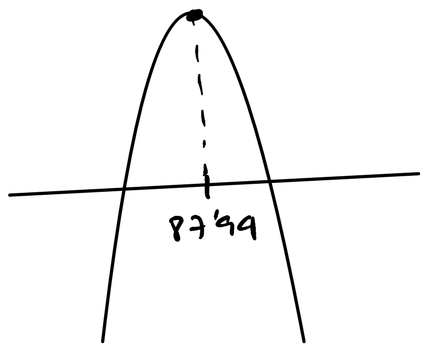


b) ¿Para cuántos años de antigüedad alcanza el salario su máximo fijados los estudios?

$\hookrightarrow \hat{\beta}_4 < 0$

$$\frac{\partial \widehat{salario}}{\partial antigüedad} = 0'299 - 2 \cdot 0'0017 \cdot antigüedad = 0$$

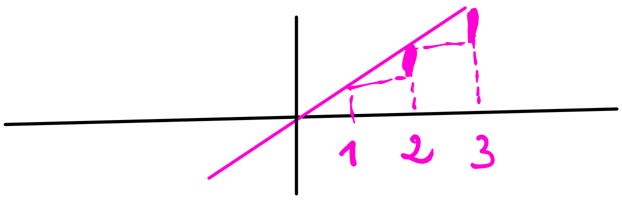
$$antigüedad = \frac{0'299}{2 \cdot 0'0017} = 87'94$$

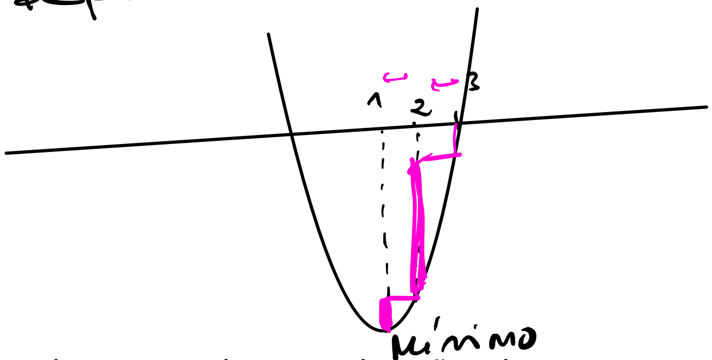


c) ¿En cuánto se incrementa el salario por cada nivel de estudios superado, ceteris paribus la antigüedad?

$$\frac{\partial \widehat{salario}}{\partial\, estudios} = -0'385 + 2 \cdot 0'189 \cdot estudios$$

No es constante. Depende del nivel de estudios.





d) ¿En cuánto se incrementa el salario medio cuando se pasa de uno a dos años de antigüedad? ¿y cuando se pasa de 29 a 30 años? ¿Por qué se produce distinto incremento?

$$\frac{\partial \widehat{salario}}{\partial\, antigüedad} = 0'299 - 2 \cdot 0'0017 \cdot antigüedad = 0$$

1 → 2 años → $0'299 - 2 \cdot 0'0017 \cdot 1 = 0'2956$

29 - 30 años → $0'299 - 2 \cdot 0'0017 \cdot 29 = 0'2004$

2.1.4 Términos de interacción

En ocasiones es adecuado para dotar de mayor realismo o afinación al modelo previsto hacer que una variable explicativa dependa de la magnitud o nivel que alcanza otra variable independiente. Es como si ambas variables explicativas tuvieran un efecto parcial no solo aisladamente, sino también conjuntamente. Este tipo de interacción se puede considerar introduciendo en el modelo un término nuevo que actúe como término de interacción:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon \longrightarrow \frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3 X_2$$

Diferenciando en en ambas partes respecto a Y y X_1 tenemos:

$$dY = \beta_1 dX_1 + \beta_3 X_2 dX_1 = dX_1\left(\beta_1 + \beta_3 X_2\right)$$

y sustituyendo los diferenciales por incrementos:

$$\Delta Y = \left(\beta_1 + \beta_3 X_2\right)\Delta X_1$$

de manera que el incremento de la variables explicada depende del incremento de la variable independiente pero también de su nivel de la variable con la que interacciona.

EJERCICIO 2: Usuarios de internet

Nos plantemos si los ingresos per cápita y los años de estudio influyen en la proporción de la población usuaria de internet. Consideramos además que el efecto sobre los usuarios de internet de una variación porcentual en los ingresos depende de los años de educación.

Su estimación (FRM):

$$\widehat{internet} = 52{,}608 - 6,26\ln(ingresos) - 19,08 \cdot estudios + 2,511\,[\ln(ingresos) \cdot estudios]$$
$$n = 169,\ R^2 = 0,8024,\ \bar{R}^2 = 0,7988.$$

a) ¿Cuál es el efecto parcial sobre internet de los ingresos, si consideramos que el valor del nivel de estudios es 7,59?

$$\frac{\partial\, internet}{\partial \ln ingresos} = -6'26 + 2'511 \cdot estudios$$
$$-6'26 + 2'511 \cdot 7'59 = 12'79$$

lin – log → Δ 1% de los ingresos ⇒ Δ $\frac{12'79}{100}$ = 0'1279 la proporción de usuarios

b) ¿Cuál será el efecto parcial sobre internet de los estudios, fijando el valor de los ingresos en el ingreso medio per cápita en términos de PPA en logaritmos de la muestra que es 8,8.

$$\frac{\partial\, \widehat{internet}}{\partial\, estudios} = -19'08 + 2'511 \cdot \ln ingresos$$
$$-19'08 + 2'511 \cdot 8'8 = 3'01$$

lin – lin : Δ de 1 año de educacion ⇒ Δ 3'01 la proporción de usuarios

2.1.5 Regresiones con variables estandarizadas

Cuando alguna de las variables tiene una escala de valores de difícil interpretación puede ser interesante medirla en términos tipificados o estandarizados. Tipificar no es más que restar la media a todos los valores de la variable y dividirla por su desviación típica o error estándar.

$$Z=(X-\bar{X})/S_X \quad \rightarrow \quad z = \frac{x_i - \bar{x}}{S_x} \left\} \begin{array}{l} \mu_x = 0 \\ \sigma_x = 1 \end{array} \right.$$

Puede resultar adecuado expresar todo el modelo estandarizado, en este caso se llama modelo de coeficientes beta.

Tipificando las variables del modelo tenemos que la estimación MCO es:

$$\underbrace{\frac{y_i}{S_Y}}_{z_y} = \left(\frac{S_{X_1}}{S_Y}\right)\hat{\beta}_1\left(\frac{x_{1i}}{S_{X_1}}\right) + \left(\frac{S_{X_2}}{S_Y}\right)\hat{\beta}_2\left(\frac{x_{2i}}{S_{X_2}}\right) + \dots + \left(\frac{S_{X_k}}{S_Y}\right)\hat{\beta}_k\left(\frac{x_{ki}}{S_{X_k}}\right) + \frac{\hat{\varepsilon}_i}{S_Y}$$

$$\frac{y_i - \bar{y}}{S_y} = z_y$$

$$\frac{x_i - \bar{x}_1}{S_{x_1}} = z_1$$

Y podemos expresar el modelo en función de la variable tipificada z:

$$\boxed{Z_y = \tilde{\beta}_1 Z_1 + \tilde{\beta}_2 Z_2 + \dots \tilde{\beta}_k Z_k + \tilde{\varepsilon}. \\ =}$$

Donde:

$$\tilde{\beta}_2 = (S_{X_2}/S_Y)\hat{\beta}_2, \dots, \tilde{\beta}_k = (S_{X_k}/S_Y)\hat{\beta}_1 \quad \rightarrow \quad \tilde{\beta}_j = \hat{\beta}_j \frac{S_{x_j}}{S}$$

Una de las ventajas de los coeficientes beta es que no dependen de las unidades de medida utilizadas y permiten determinar la influencia de las variables explicativas sobre la explicada a partir de la magnitud del coeficiente, lo que normalmente no ocurre en los otros casos en que los coeficientes pueden modificarse cambiando las unidades de medida de las variables.