Tema 3. El modelo de regresión múltiple

1. Hipótesis básicas

1.1. El modelo

El modelo de regresión lineal múltiple pretende medir **el efecto de las variables más importantes** ($x_1, x_2, ..., x_k$) **sobre la variable respuesta** y, y representa el efecto de las variables restantes mediante una variable aleatoria que llamaremos perturbación del modelo:

$$y = f(x_1, x_2, ..., x_k) + \underbrace{g(x_{k+1}, ..., x_p)}_{u}$$

Supondremos que en el rango de valores de interés, la función f admite una **aproximación lineal**, con lo que resulta **el modelo de regresión lineal múltiple**:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$$

(Véase el Ejercicio 13 "Beneficio de los bancos")

1.2. Hipótesis básicas

Supondremos que se quiere explicar los valores de una **variable aleatoria** y a partir de un conjunto de k variables matemáticas $x_1, x_2, ..., x_k$, que sobre los individuos de de la muestra toman valores predeterminados y conocidos.

La v.a. y recibe el nombre de variable **explicada**, **endógena**, **respuesta** o **dependiente**.

Las variables $x_1, x_2, ..., x_k$ reciben el nombre de variables **explicativas**, **exógenas**, **regresores** o **independientes**.

Admitiremos que en cada individuo de la muestra la relación entre las variables es:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i \text{ , para } i=1,2,...,n,$$

donde

y_i es el valor de la respuesta en el individuo i-ésimo,

β_j mide el **efecto marginal** que produce un aumento unitario en x_j cuando el resto de variables permanecen constantes ($j=1,2,...,k$),

u_i mide el efecto que producen sobre la respuesta las variables que no están incluidas en el modelo. Es el **término de error** o **perturbación**.

Supondremos que **la perturbación aleatoria u_i verifica las hipótesis habituales**:

(1) Su esperanza es cero: $E(u_i) = 0$

(2) Su varianza es constante: $\text{var}(u_i) = \sigma^2$

(3) Son independientes dos a dos: $\text{cov}(u_i, u_j) = 0, \ \forall i \neq j \Leftrightarrow E(u_i u_j) = 0, \ \forall i \neq j$

(4) La distribución de u_i es normal.

Si llamamos $\mathbf{U} = (u_1, u_2, ..., u_n)'$ al vector columna de dimensión n, estas condiciones se resumen en:

$$\mathbf{U} \sim NM_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

es decir, el vector $\mathbf{U}$ tiene una ley normal multivariante de dimensión n, con vector de medias

$$E(\mathbf{U}) = (E(u_1), E(u_2), ..., E(u_n))' = (0, 0, ..., 0)' = \mathbf{0}$$

y matriz de covarianzas

$$\text{var}(\mathbf{U}) = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

Además de estas condiciones sobre la perturbación aleatoria, es necesario **añadir dos hipótesis adicionales para poder estimar el modelo sin ambigüedad**:

(5) El tamaño muestral debe ser mayor que el número de parámetros a estimar, $n>k+1$,

(6) Las variables explicativas $x_1, x_2, ..., x_k$ deben ser linealmente independientes, es decir, el rango de la matriz

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

debe ser $rang(\mathbf{X}) = k+1$.

Las hipótesis respecto a la perturbación aleatoria pueden escribirse en términos de la variable respuesta de la forma siguiente:

(1') Para cada conjunto fijo de valores de las x, la distribución de y tiene media:

$$E(y) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

(2') La varianza de y es constante y no depende de los valores de las x:

$$\mathrm{var}(y) = \sigma^2$$

(3') Las y_i son independientes dos a dos.

(4') La variable respuesta tiene ley normal.

2. Estimación de los parámetros

Puesto que, por hipótesis, la distribución de la variable respuesta es normal, el método de estimación por máxima verosimilitud coincide con el método de estimación por mínimos cuadrados. Así, la función a minimizar es:

$$M = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

donde

$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki}$ son los valores ajustados, para $i=1,2,...,n$,

e_i son los residuos del modelo, para $i=1,2,...,n$.

Por tanto, la función a minimizar es:

$$M = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - ... - \beta_k x_{ki})^2$$

Buscamos $\beta_0, \beta_1, ..., \beta_k$ tales que M sea mínimo. Para ello debemos derivar parcialmente la función M respecto de los parámetros $\beta_0, \beta_1, ..., \beta_k$.

Derivando parcialmente M respecto de $\beta_0, \beta_1, ..., \beta_k$ obtenemos $k+1$ **ecuaciones normales**:

$$\frac{\partial M}{\partial \beta_0} = 2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - ... - \beta_k x_{ki})(-1) = 0 \quad \Rightarrow \sum_{i=1}^{n} e_i = 0$$

para $j = 1,...,k$:

$$\frac{\partial M}{\partial \beta_j} = 2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - ... - \beta_k x_{ki})(-x_{ji}) = 0 \quad \Rightarrow \sum_{i=1}^{n} e_i x_{ji} = 0$$

Las ecuaciones normales dan lugar a un sistema de $k+1$ ecuaciones con $k+1$ incógnitas.

Implicaciones de las ecuaciones normales:

a) los residuos deben tener media cero,
b) los residuos no deben estar relacionados con ninguna de las variables explicativas.

Esto garantiza que en la descomposición básica $y_i=\hat{y}_i+e_i$ toda la información de la relación entre la variable respuesta y las variables explicativas está recogida en $\hat{y}_i$, mientras que el residuo e_i contiene la parte de la respuesta que no está relacionada con las x_j.

El sistema de $k+1$ ecuaciones con $k+1$ incógnitas puede escribirse como:

$$\begin{cases}
\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{1i} + \ldots + \beta_k \sum_{i=1}^{n} x_{ki} \\
\sum_{i=1}^{n} y_i x_{1i} = \beta_0 \sum_{i=1}^{n} x_{1i} + \beta_1 \sum_{i=1}^{n} x_{1i}^2 + \ldots + \beta_k \sum_{i=1}^{n} x_{ki} x_{1i} \\
\sum_{i=1}^{n} y_i x_{2i} = \beta_0 \sum_{i=1}^{n} x_{2i} + \beta_1 \sum_{i=1}^{n} x_{1i} x_{2i} + \ldots + \beta_k \sum_{i=1}^{n} x_{ki} x_{2i} \\
\vdots \\
\sum_{i=1}^{n} y_i x_{ki} = \beta_0 \sum_{i=1}^{n} x_{ki} + \beta_1 \sum_{i=1}^{n} x_{1i} x_{ki} + \ldots + \beta_k \sum_{i=1}^{n} x_{ki}^2
\end{cases}$$

y matricialmente:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{pmatrix} \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{Y}} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{pmatrix} \underbrace{\begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}}_{\boldsymbol{\beta}}$$

Es decir: $\mathbf{X'Y=X'X\beta}$

Así, la estimación del vector de parámetros es: $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$

La condición (6), $rang(\mathbf{X})=k+1$, implica que $rang(\mathbf{X'X})=k+1$ y, por tanto, que existe la matriz inversa de $\mathbf{X'X}$.

Planteamiento del modelo:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}, \qquad \mathbf{U} \sim NM_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

Hiperplano de regresión:

$$\boxed{\begin{aligned} &\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}, \qquad \hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y} \\ &\hat{\mathbf{Y}} = \underbrace{\mathbf{X(X'X)}^{-1}\mathbf{X'}}_{\mathbf{H}}\mathbf{Y} = \mathbf{HY} \end{aligned}}$$

La matriz $\mathbf{H}$ recibe el nombre de **proyector ortogonal** (o *hat matrix*) y es simétrica ($\mathbf{H'=H}$) e idempotente ($\mathbf{H^2=H}$).

Residuos:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

La matriz $(\mathbf{I\text{-}H})$ se denomina **proyector ortogonal complementario**.

Observación:

El proyector ortogonal $\mathbf{I\text{-}H}$ define un espacio ortogonal al definido por el proyector $\mathbf{H}$.

Esto implica que $\hat{\mathbf{Y}}'\mathbf{e} = 0$, puesto que $\hat{\mathbf{Y}}$ y $\mathbf{e}$ pertenecen a dos **espacios distintos y ortogonales**.

Ejercicio 14. (Consumo de 8 provincias españolas) El siguiente cuadro presenta tres indicadores de consumo en 8 provincias españolas: un indicador global de consumo (***y***), un indicador del número de automóviles (***x_1***) y un indicador del número de teléfonos (***x_2***).

Provincia	***Indicador de Consumo (y)***	***Automóviles cada 1000 hab. (x_1)***	***Teléfonos cada 1000 hab.(x_2)***
Ávila	64	58	111
Palencia	78	84	131
Segovia	83	78	158
Burgos	88	81	147
Soria	89	82	121
Valladolid	99	102	165
Logroño	101	85	174
Santander	102	102	169

a) Construir un modelo explicativo que relacione el indicador global con estos dos indicadores de consumo.

b) ¿Qué efecto produce sobre el consumo un aumento de una unidad en el número de automóviles por cada 1000 habitantes?

c) ¿Qué efecto produce sobre el consumo un aumento de una unidad en el número de teléfonos por cada 1000 habitantes?

Las matrices de datos son:

$$\mathbf{X}=\begin{pmatrix}1 & 58 & 111\\ 1 & 84 & 131\\ 1 & 78 & 158\\ 1 & 81 & 147\\ 1 & 82 & 121\\ 1 & 102 & 165\\ 1 & 85 & 174\\ 1 & 102 & 169\end{pmatrix}\qquad \mathbf{Y}=\begin{pmatrix}64\\ 78\\ 83\\ 88\\ 89\\ 99\\ 101\\ 102\end{pmatrix}\qquad \mathbf{X'X}=\begin{pmatrix}8 & 672 & 1176\\ 672 & 57822 & 100453\\ 1176 & 100453 & 176758\end{pmatrix}\qquad \mathbf{X'Y}=\begin{pmatrix}704\\ 60251\\ 105288\end{pmatrix}$$

n $\quad n\bar{x}_1$ $\quad n\bar{x}_2$

$$\mathbf{X'X}=\begin{pmatrix}8 & 672 & 1176\\ 672 & \boxed{57822} & \boxed{100453}\\ 1176 & 100453 & \boxed{176758}\end{pmatrix}\qquad \mathbf{X'Y}=\begin{pmatrix}704\\ \boxed{60251}\\ \boxed{105288}\end{pmatrix}$$

$n\bar{y}$

$n\overline{x_1 y}$

$n\overline{x_1 x_2}$

$n\overline{x_1^2}$ $\quad n\overline{x_2^2}$ $\quad n\overline{x_2 y}$

La inversa de **X'X** es:

$$(\mathbf{X'X})^{-1}=\begin{pmatrix}6.3482 & -0.0317 & -0.0242\\ -0.0317 & 0.0015 & -0.0007\\ -0.0242 & -0.0007 & 0.0005\end{pmatrix}$$

y la estimación de los parámetros es:

$$\hat{\boldsymbol\beta}=\begin{pmatrix}\hat\beta_0\\ \hat\beta_1\\ \hat\beta_2\end{pmatrix}=(\mathbf{X'X})^{-1}\mathbf{X'Y}=\begin{pmatrix}6.3482 & -0.0317 & -0.0242\\ -0.0317 & 0.0015 & -0.0007\\ -0.0242 & -0.0007 & 0.0005\end{pmatrix}\begin{pmatrix}704\\ 60251\\ 105288\end{pmatrix}=\begin{pmatrix}9.0538\\ 0.5203\\ 0.2397\end{pmatrix}$$

El hiperplano de regresión estimado es: $\hat{y}=9.05+0.52x_1+0.24x_2$

3. Propiedades de los estimadores de los parámetros β_j

a) $\hat\beta_j$ tiene ley normal al ser función lineal de v.a. normales.

b) $\hat\beta_j$ es un estimador **insesgado** de β_j,

c) La varianza de $\hat\beta_j$ es:

$$\mathrm{var}(\hat\beta_j)=\sigma^2 q_{jj},\text{ siendo } q_{jj} \text{ el } (j+1)\text{-ésimo elemento de } diag((\mathbf{X'X})^{-1})$$

Estas propiedades se resumen en:

$$\hat\beta_j \sim N(\beta_j,\sigma^2 q_{jj}),\text{ para } j=0,1,\ldots,k.$$

Observación: a) se demuestra viendo que el estimador mínimo-cuadrático se ha obtenido como función lineal de las y_i (que tenían ley normal). Es decir: $\hat{\boldsymbol\beta}=(\mathbf{X'X})^{-1}\mathbf{X'Y}$.

3.1 Cálculo de esperanzas

Recordemos que el modelo es: $\mathbf{Y} = \mathbf{X\beta} + \mathbf{U}, \qquad \mathbf{U} \sim NM_n(\mathbf{0}, \sigma^2\mathbf{I})$

A partir de la definición del estimador mínimo-cuadrático, tenemos que:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X\beta}+\mathbf{U}) = \underbrace{(\mathbf{X'X})^{-1}\mathbf{X'X}}_{=\mathbf{I}_{k+1}}\boldsymbol{\beta} + \underbrace{(\mathbf{X'X})^{-1}\mathbf{X'}}_{=\mathbf{C}\ \text{constante}}\mathbf{U} = \boldsymbol{\beta} + \mathbf{CU}$$

y tomando esperanzas:

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}+\mathbf{CU}) = \boldsymbol{\beta} + \mathbf{C}\underbrace{E(\mathbf{U})}_{=\mathbf{0}} = \boldsymbol{\beta}$$

3.2 Cálculo de varianzas

La matriz de varianzas-covarianzas de los estimadores $\hat{\boldsymbol{\beta}}$ será una matriz *(k+1)*x*(k+1)*, simétrica, en cuya diagonal estarán las varianzas de cada $\hat{\beta}_j$ *(j=0,1,...,k)* y fuera de la diagonal estarán las covarianzas entre $\hat{\beta}_j$ y $\hat{\beta}_l$, para $l \neq j$.

Esta matriz se define como:

$$\text{var}(\hat{\boldsymbol{\beta}}) = E((\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))').$$

Teniendo en cuenta que

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \text{ y que } \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{CU}, \text{ donde } \mathbf{C} = (\mathbf{X'X})^{-1}\mathbf{X'},$$

sustituyendo en la definición, se tiene que:

$$\text{var}(\hat{\boldsymbol{\beta}}) = E(\mathbf{CUU'C'}) = \mathbf{C}\underbrace{E(\mathbf{UU'})}_{=\sigma^2\mathbf{I}_n}\mathbf{C'} = \sigma^2\mathbf{CC'} = \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1} = \sigma^2(\mathbf{X'X})^{-1}$$

Si llamamos q_{ij} *(i,j=0,1,...,k)* a los elementos de la matriz $(\mathbf{X'X})^{-1}$:

$$(\mathbf{X'X})^{-1} = \begin{pmatrix} q_{00} & q_{01} & \cdots & q_{0k} \\ q_{10} & q_{11} & \cdots & q_{1k} \\ \vdots & \vdots & & \vdots \\ q_{k0} & q_{k1} & \cdots & q_{kk} \end{pmatrix} \quad \text{con} \quad q_{ij} = q_{ji} \quad \forall i \neq j$$

entonces:

$$\text{var}(\hat{\beta}_j) = \sigma^2 q_{jj}, \text{ para } j = 0,1,...,k,$$

$$\text{cov}(\hat{\beta}_j, \hat{\beta}_l) = \sigma^2 q_{jl}, \text{ para } j,l = 0,1,...,k, \text{ con } j \neq l.$$

Puesto que, en general, σ^2 será desconocida, en su lugar utilizaremos la varianza residual. Entonces:

$s_R^2\, q_{jj}$ será un estimador de la $\text{var}(\hat{\beta}_j) = \sigma^2 q_{jj}$

Observación: En general, la matriz $\mathbf{X'X}$ no será diagonal, por lo que su inversa tampoco lo será. Esto implica que los estimadores $\hat{\beta}_j$ **no serán independientes**, pues sus covarianzas no serán nulas. Solamente serán independientes si $\mathbf{X'X}$ es una matriz diagonal, lo que requiere que las variables explicativas tengan media cero y covarianzas nulas.

4. Estimación de la varianza

El estimador máximo-verosímil de la varianza de la perturbación aleatoria σ^2 es la varianza residual:

$$s_R^2 = \frac{1}{n\text{-}k\text{-}1}\sum_{i=1}^{n} e_i^2$$

Matricialmente, si $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ es el vector columna de dimensión n que contiene los residuos, entonces:

$$s_R^2 = \frac{1}{n\text{-}k\text{-}1}\mathbf{e'e} = \frac{1}{n\text{-}k\text{-}1}(\mathbf{Y}-\hat{\mathbf{Y}})'(\mathbf{Y}-\hat{\mathbf{Y}}) = \frac{1}{n\text{-}k\text{-}1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})$$

$$= \frac{1}{n\text{-}k\text{-}1}(\mathbf{Y'Y} - \mathbf{Y'X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X'Y} + \underbrace{\hat{\boldsymbol{\beta}}'}\mathbf{X'X}\hat{\boldsymbol{\beta}})$$

$$= \frac{1}{n\text{-}k\text{-}1}(\mathbf{Y'Y} - \mathbf{Y'X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X'Y} + \mathbf{Y'X}(\mathbf{X'X})^{-1}\mathbf{X'X}\hat{\boldsymbol{\beta}})$$

$$= \frac{1}{n\text{-}k\text{-}1}(\mathbf{Y'Y} - \hat{\boldsymbol{\beta}}'\mathbf{X'Y})$$

Observación: como en modelos anteriores, se puede demostrar que

$$\frac{n\text{-}k\text{-}1}{\sigma^2}s_R^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} e_i^2 \sim \chi^2_{n-k-1}$$

La **esperanza** y **varianza** del estimador s_R^2 son:

$$E(s_R^2) = \sigma^2, \ \ \mathrm{var}(s_R^2) = \frac{2\sigma^4}{n-k-1}.$$

Éstas se deducen a partir de la esperanza y varianza de la ley Gamma, teniendo en cuenta que:

$$\chi_k^2 = Gamma\left(\frac{k}{2}, 2\right) \Rightarrow E(\chi_k^2) = k, \ \ \mathrm{var}(\chi_k^2) = 2k.$$

Ejercicio 15. Con los datos del **Ejercicio 14** (consumo de 8 provincias españolas) calcular la varianza residual y una estimación de las varianzas y covarianzas entre los coeficientes.

Del Ejercicio 14 sabemos que:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 9.0538 \\ 0.5203 \\ 0.2397 \end{pmatrix} \qquad \mathbf{X'Y} = \begin{pmatrix} 704 \\ 60251 \\ 105288 \end{pmatrix} \qquad \mathbf{Y} = (64,78,83,88,89,99,101,102)'$$

Calculamos:

$$\mathbf{Y'Y} = (64,78,\ldots,102)\begin{pmatrix} 64 \\ 78 \\ \vdots \\ 102 \end{pmatrix} = 63140, \quad \hat{\boldsymbol{\beta}}'\mathbf{X'Y} = (9.0538 \quad 0.5203 \quad 0.2397)\begin{pmatrix} 704 \\ 60251 \\ 105288 \end{pmatrix} = 62970.84$$

Y la varianza residual es:

$$s_R^2 = \frac{1}{n\text{-}k\text{-}1}\mathbf{e'e} = \frac{1}{n\text{-}k\text{-}1}(\mathbf{Y'Y} - \hat{\boldsymbol{\beta}}'\mathbf{X'Y}) = \frac{6314 - 62970.84}{8-2-1} = 33.83$$

Finalmente, la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\beta}}$ es:

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = s_R^2(\mathbf{X'X})^{-1} = 33.83\begin{pmatrix} 6.3482 & -0.0317 & -0.0242 \\ -0.0317 & 0.0015 & -0.0007 \\ -0.0242 & -0.0007 & 0.0005 \end{pmatrix} = \begin{pmatrix} 214.76 & -1.0724 & -0.8187 \\ -1.0724 & 0.0507 & -0.0237 \\ -0.8187 & -0.0237 & 0.0169 \end{pmatrix}$$

(la inversa de $\mathbf{X'X}$ se ha obtenido en el Ejercicio 14).

Las desviaciones típicas de los coeficientes son las raíces cuadradas de los términos diagonales.

5. Predicción

Un modelo de regresión múltiple permite:

1) **Estimar las medias condicionadas**: Estimar las medias de las distribuciones de y para cada valor de x,
2) **Realizar pronósticos**: Prever futuros valores de la variable respuesta.

Tanto la estimación de la media como la predicción de y se obtienen sustituyendo en la ecuación del hiperplano de regresión. Por tanto, **sus valores numéricos son idénticos**.

Pero la precisión de estas dos estimaciones es distinta y, por tanto, **los intervalos de confianza serán distintos**.

5.1. Estimación de las medias condicionadas

Modelo: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}, \qquad \mathbf{U} \sim NM_n(\mathbf{0}, \sigma^2 \mathbf{I})$

$E(y/\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$,

El nuevo parámetro que se quiere estimar es la media de y condicionada a que las variables explicattivas x_j tomen un valor concreto $\mathbf{x}_h' = (1, x_{1h}, x_{2h}, \ldots, x_{kh})$, es decir:

$$E(y/\mathbf{x}=\mathbf{x}_h) = \mathbf{x}_h' \boldsymbol{\beta} = m_h$$

El **estimador puntual** de m_h es:

$$\hat{y}_h = \mathbf{x}_h' \hat{\boldsymbol{\beta}},$$

que tiene ley normal univariante, al ser combinación lineal de las componentes de un vector normal multivariante.

Proposición: $\hat{y}_h = \mathbf{x}_h' \hat{\boldsymbol{\beta}}$ es un estimador **insesgado** de m_h, cuya varianza es

$$\mathrm{var}(\hat{y}_h) = \sigma^2 v_{hh}, \quad \text{donde} \quad v_{hh} = \mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h$$

(Demostración)

Observaciones:

- $v_{hh} = \mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h$ es el elemento h-ésimo de la diagonal del proyector $\mathbf{H}$.
- Puede demostrarse que (apéndice 12, Peña II):

$$v_{hh} = \mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h = \frac{1}{n}\left(1 + \underbrace{(\tilde{\mathbf{x}}_h - \bar{\mathbf{x}})'\mathbf{S}_{xx}^{-1}(\tilde{\mathbf{x}}_h - \bar{\mathbf{x}})}_{dist.Mahalanobis}\right),$$

  donde $\tilde{\mathbf{x}}_h' = (x_{1h}, x_{2h}, \ldots, x_{kh})$, de manera que $\mathbf{x}_h' = (1, \ \tilde{\mathbf{x}}_h')$, $\bar{\mathbf{x}}$ es el vector de medias de las k variables explicativas y $\mathbf{S}_{xx}$ es su matriz de covarianzas.
- v_{hh} es una medida de distancia del punto $\tilde{\mathbf{x}}_h$ al centro de gravedad $\bar{\mathbf{x}}$. Además,

$$\text{Si } \tilde{\mathbf{x}}_h = \bar{\mathbf{x}} \Rightarrow v_{hh} = \frac{1}{n} \Rightarrow \mathrm{var}(\hat{y}_h) = \frac{\sigma^2}{n}$$

$$\text{Si } \tilde{\mathbf{x}}_h \neq \bar{\mathbf{x}} \Rightarrow v_{hh} > \frac{1}{n} \Rightarrow \mathrm{var}(\hat{y}_h) > \frac{\sigma^2}{n}$$

  Cuánto más alejado esté $\tilde{\mathbf{x}}_h$ de $\bar{\mathbf{x}}$, mayor será la varianza $\mathrm{var}(\hat{y}_h)$ y, por tanto, **menor será la precisión** de la estimación.
- Se define $\hat{n}_h = 1/v_{hh}$ como el **nº equivalente de observaciones** para estimar m_h. Utilizando esta cantidad, $\mathrm{var}(\hat{y}_h) = \sigma^2 / \hat{n}_h$.

Intervalo de confianza para m_h

Utilizaremos la ley de probabilidad de su estimador puntual $\hat{y}_h = \mathbf{x}_h'\hat{\boldsymbol{\beta}}$.

Caso 1: σ^2 conocida,

$$\hat{y}_h \sim N\left(m_h, \frac{\sigma^2}{\hat{n}_h}\right) \Rightarrow \frac{\hat{y}_h - m_h}{\sigma / \sqrt{\hat{n}_h}} \sim N(0,1)$$

Caso 2: σ^2 desconocida,

$$\frac{\hat{y}_h - m_h}{s_R / \sqrt{\hat{n}_h}} \sim t_{n-k-1}$$

En general, se estará en la hipótesis del Caso 2, con lo que el I.C. para m_h al $(1\text{-}\alpha)100\%$ será:

$$I.C.(m_h) = \left[\hat{y}_h \mp t_{1-\frac{\alpha}{2}} \, s_R \sqrt{\frac{1}{\hat{n}_h}}\right]$$

donde $t_{1-\alpha/2}$ es el percentil $(1\text{-}\alpha/2)100\%$ de la ley t de Student con $n\text{-}k\text{-}1$ gr. libertad y

$$\hat{n}_h = \frac{1}{v_{hh}} = \frac{1}{\mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h}$$

5.2 Predicción de una nueva observación (pronóstico)

Se quiere prever el **valor futuro** de la variable respuesta cuando $\mathbf{x}=\mathbf{x}_h$. Por tanto, se tiene una nueva v.a.

$y_h = \mathbf{x}_h'\boldsymbol{\beta} + u_h$, donde $u_h \sim N(0, \sigma^2)$ y es independiente de $u_1, u_2, \ldots, u_n$.

La predicción o pronóstico de la variable respuesta se obtiene sustituyendo en la ecuación del hiperplano de regresión:

$$\hat{y}_h = \mathbf{x}_h'\hat{\boldsymbol{\beta}}$$

pero, ¿y si se quiere un **intervalo de predicción**?

El **intervalo de predicción o de pronóstico** es un intervalo de valores probables para la variable aleatoria y_h. y al $(1\text{-}\alpha)100\%$ es

$$y_h \in \left[\hat{y}_h \mp t_{1-\frac{\alpha}{2}} \, s_R \sqrt{1 + \frac{1}{\hat{n}_h}}\right]$$

donde $t_{1-\alpha/2}$ es el percentil $(1\text{-}\alpha/2)100\%$ de la ley t de Student con $n\text{-}k\text{-}1$ gr. libertad y

$$\hat{n}_h = \frac{1}{v_{hh}} = \frac{1}{\mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h}$$

¿Cómo se deduce este intervalo?

Se considera la variable diferencia, es decir, el residuo $e_h = y_h - \hat{y}_h$, que tiene ley normal al ser combinación lineal de normales.

Se obtienen su esperanza y varianza:

$$E(y_h - \hat{y}_h) = E(\mathbf{x}_h'\boldsymbol{\beta} + u_h - \mathbf{x}_h'\hat{\boldsymbol{\beta}}) = \underbrace{\mathbf{x}_h'\boldsymbol{\beta}}_{=m_h} + 0 - \underbrace{\mathbf{x}_h'\boldsymbol{\beta}}_{=m_h} = 0$$

$$\mathrm{var}(y_h - \hat{y}_h) = \mathrm{var}(y_h) + \mathrm{var}(\hat{y}_h) = \sigma^2 + \frac{\sigma^2}{\hat{n}_h} = \sigma^2\left(1 + \frac{1}{\hat{n}_h}\right)$$

Observación: La varianza se obtiene como suma de varianzas puesto que $y_h, \hat{y}_h$ son independientes al serlo u_h de $u_1, u_2, \ldots, u_n$.

Por tanto, se tiene que:

$$y_h - \hat{y}_h \sim N\left(0, \sigma^2\left(1 + \frac{1}{\hat{n}_h}\right)\right) \Rightarrow \frac{y_h - \hat{y}_h}{\sigma\sqrt{1 + \frac{1}{\hat{n}_h}}} \sim N(0,1)$$

siempre que σ^2 sea conocida.

Si σ^2 es desconocida, entonces se sustituye por su estimador, y se tiene que:

$$\frac{y_h - \hat{y}_h}{s_R\sqrt{1 + \frac{1}{\hat{n}_h}}} \sim t_{n-k-1}$$

Puesto que esta será la situación habitual, el intervalo de predicción para la respuesta concreta y_h es:

$$y_h \in \left[\hat{y}_h \mp t_{1-\frac{\alpha}{2}}\; s_R\sqrt{1 + \frac{1}{\hat{n}_h}}\right]$$

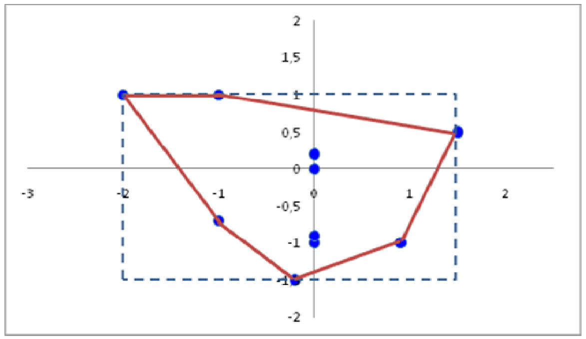
donde

$$\hat{n}_h = \frac{1}{v_{hh}} = \frac{1}{\mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h}$$

Extrapolación oculta y predicción

En la estimación de la respuesta media o la predicción de nuevas respuestas en un punto concreto $\mathbf{x}=\mathbf{x}_h$, debemos ser extremadamente cautos con la **extrapolación**.

- Si sólo se tiene en cuenta el **producto cartesiano**, es fácil considerar la predicción para un punto que puede estar fuera de la nube de puntos con la que se ha calculado el hiperplano de regresión (**Extrapolación oculta!!**)
- Para evitar este problema, debemos ceñirnos al **casco convexo** (*convex hull*) de las variables regresoras: el menor conjunto convexo que contiene a los n puntos originales.



producto cartesiano (-2,1.5)x(-1.5,1)

casco convexo

En general, cuando el número de variables regresoras es $k > 2$, el casco convexo resulta difícil de calcular.

Sin embargo, es posible calcular de forma sencilla un **elipsoide que contiene al casco conexo**, de la forma siguiente:

Sea

$$v_{\max} = \max_{1 \le h \le n}\{v_{hh}\}, \quad \text{donde} \quad v_{hh} = \mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h$$

entonces, el elipsoide de ecuación:

$$\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} \le v_{\max}, \quad \forall \mathbf{x} \in \mathbb{R}^{k+1}$$

contiene al casco convexo. (Atención! Este elipsoide no es el menor de los elipsoides que contienen al casco convexo, pero sí el más sencillo de calcular.)

Para evitar en lo posible la extrapolación, comprobaremos si el punto $\mathbf{x}_h$ para el que se quiere realizar la predicción se encuentra dentro de este elipsoide:

$\mathbf{x}_h$ pertenece al elipsoide si se cumple que

$$\mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h \le v_{\max}$$

(Véase el ejemplo "Predicción")