2.1.18. Predicción en el modelo de dos variables

1. Predicción puntual.

La predicción puntual viene dada por el valor de la regresión correspondiente a X_0, es decir;

$$\hat{Y}_0 = a + b \cdot X_0$$

2. Predicción por intervalos.

2.1. Predicción de un valor concreto:

$$\hat{Y}_0 \pm t_{n-2}(\alpha/2) \cdot S_u \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_{ci}^2}}$$

2.2. Predicción para un valor esperado:

$$\hat{Y}_0 \pm t_{n-2}(\alpha/2) \cdot S_u \cdot \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_{ci}^2}}$$

2.2. MODELO DE REGRESIÓN DE DOS VARIABLES

Ejercicio 1

El Ayuntamiento de Madrid quiere construir un modelo de regresión lineal que explique el número de personas que acuden a las piscinas municipales (Y) en función de la diferencia de temperatura con respecto al día anterior (X). Se recogió información durante 46 días y se estimó, por mínimos cuadrados, la siguiente expresión:

$$\hat{Y} = 5.729{,}952 + 117{,}0778 \cdot X$$

A partir de la siguiente información adicional:

$$\bar{Y} = 6.244 \qquad \Sigma\, Y_i^2 = 825.872.111{,}5$$

$$\bar{X} = 4{,}39 \qquad \Sigma\, X_i^2 = 943{,}27812$$

Se pide:

1. ¿Es significativo el incremento de la temperatura en el número de usuarios de las piscinas? Utilizar un nivel de significación del 1 %.
2. Construir un intervalo de confianza para la ordenada en el origen al 90 % de confianza.
3. ¿Podría aceptarse que si la temperatura aumenta en 1 grado acuden 200 personas más a las piscinas municipales? Utilizar una confianza del 95 %.
4. Obtener una medida de la bondad del ajuste.

SOLUCIÓN

1. Para contrastar las hipótesis de significación de la variable explicativa se plantea la siguiente hipótesis nula:

$$H_0\colon \beta = 0 \qquad \text{frente a} \qquad H_1\colon \beta \neq 0$$

Si esta hipótesis es cierta, el estadístico:

$$t_{\text{exp}} = \frac{b}{S_b} \sim t_{n-2}$$

Para calcular el estadístico experimental, se obtienen los datos desconocidos, es decir, la varianza estimada del estimador:

$$S_b^2 = \frac{S_u^2}{\Sigma\, x_{ci}^2}$$

pero S_u^2 es la cuasivarianza estimada de las perturbaciones, que tiene como expresión:

$$S_u^2 = \frac{SCR}{n-2}$$

Es necesario obtener la descomposición de la varianza para poder obtener la SCR.

$$SCT = SCE + SCR$$

donde:

$$SCT = \sum_i y_{ci}^2 = \sum_i Y_i^2 - n\bar{Y}^2$$

$$SCT = 1.794.844.215 - 46 \cdot 6.244^2 = 1.417.559$$

$$SCE = b^2 \cdot \sum_i x_{ci}^2 = b^2(\Sigma X_i^2 - n\bar{X}^2)$$

$$SCE = 117{,}0778^2(943{,}27812 - 46 \cdot 4{,}39^2) = 778.042{,}1455$$

La suma de cuadrados residuales se obtiene por diferencia:

$$SCR = SCT - SCE$$

$$SCR = 1.417.559 - 778.042{,}1455 = 639.516{,}8545$$

Por tanto, la estimación de la varianza de la perturbación toma el valor:

$$S^2 = \frac{SCR}{n-2} = \frac{639.516{,}8545}{46-2} = 14.534{,}47397$$

y la varianza estimada del estimador:

$$S_b^2 = \frac{S_u^2}{\sum_i x_{ci}^2} = \frac{14.534{,}47397}{56{,}76152} = 256{,}0620993$$

El estadístico experimental queda:

$$t_{\exp} = \frac{117{,}0778}{\sqrt{256{,}0620993}} = 7{,}3164$$

La distribución teórica es una t de Student con 44 grados de libertad.

$$t_{44}(0{,}01/2) = t_{44}(0{,}005) = 2{,}6923$$

La comparación de los estadísticos teóricos y experimental resuelve el contraste:

$$|t_{\exp}| > |t_{\text{tco}}|$$

Se rechaza la hipótesis nula. La diferencia en la temperatura sí es significativa para explicar el número de usuarios de las piscinas municipales.

2. Es necesario recordar que la forma que toma el intervalo de confianza para la ordenada en el origen es:

$$P[a - t_{n-2}(\alpha/2)\cdot S_a < \alpha \leqslant a + t_{n-2}(\alpha/2)\cdot S_a] = 1 - \alpha$$

El estadístico t será:

$$t_{46-2}(0{,}1/2) = t_{44}(0{,}05) = 1{,}6802$$

La varianza estimada de a será:

$$S_a^2 = S_u^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i x_{ci}^2}\right) = 14.534{,}47397\left(\frac{1}{46} + \frac{4{,}39^2}{56{,}76152}\right) = 5.250{,}8212$$

y, por tanto, su desviación típica:

$$S_a = \sqrt{S_a^2} = \sqrt{5.250{,}8212} = 72{,}46255$$

El intervalo será:

$$(5.729{,}952 \pm 1{,}6802\cdot 72{,}46255) = (5.729{,}952 \pm 121{,}75157)$$

es decir, los límites del intervalo de confianza para la ordenada en el origen, al 90 % de confianza, serán:

$$(5.608{,}2; 5.851{,}7)$$

3. En esta cuestión se plantea un contraste de hipótesis con la siguiente forma:

$$H_0\colon \beta = 200 \qquad H_1\colon \beta \neq 200$$

de manera que, si H_0 es cierta:

$$t_{\exp} = \frac{b - 200}{S_b} \sim t_{n-2}$$

El estadístico experimental será:

$$t_{\text{exp}} = \frac{117{,}0778 - 200}{\sqrt{256{,}0620993}} = -5{,}182$$

mientras que el estadístico teórico:

$$t_t\text{: } t_{46-2}(0{,}05/2) = t_{44}(0{,}025) = 2{,}0154$$

como:

$$|t_{\text{exp}}| > |t_{\text{tco}}|$$

Se rechaza la hipótesis nula. Si aumenta 1 grado la temperatura no acudirán 200 personas más a las piscinas.

4. Para obtener la medida de la bondad del ajuste, se calcula el coeficiente de determinación, R^2.

$$R^2 = \frac{SCE}{SCT} = \frac{778.042{,}1455}{1.417.559} = 0{,}5488$$

es decir, el 54,88 % de las variaciones en el número de personas que acuden a las piscinas están explicados por el ajuste mínimo-cuadrático sobre las diferencias de temperatura.

Ejercicio 2

La empresa «La onda veloz» ha especificado un modelo de regresión lineal clásico para explicar la función de demanda de sus aparatos de radio (Y), expresada en miles de unidades, y en función del precio de los aparatos de radio (X), en miles de pesetas.

Para su estimación se recoge una muestra de 24 observaciones, con los siguientes resultados:

$$\sum_{i=1}^{24} Y_i = 2.637{,}252 \qquad \sum_{i=1}^{24} X_i = 85{,}249992$$

$$\sum_{i=1}^{24} y_{ci}^2 = 159{,}8926 \qquad \sum_{i=1}^{24} x_{ci}^2 = 3{,}09161$$

$$\sum_{i=1}^{24} x_{ci} y_{ci} = -16{,}5862$$

Se pide:

1. Estimar los parámetros del modelo.
2. Obtener el coeficiente de determinación.
3. Con esta estimación, ¿sería posible decir, con una confianza del 95 %, que la empresa tendrá una demanda de 80.000 unidades si el precio de los aparatos de radio fuese de 10.000 pesetas?
4. Contrastar la significación global del modelo.

SOLUCIÓN

1. Los parámetros estimados por MCO serán:

$$b = \frac{\sum_i x_{ci} y_{ci}}{\sum_i x_{ci}^2} = \frac{-16{,}5862}{3{,}09161} = -5{,}3649$$

$$a = \overline{Y} - b\overline{X}$$

donde

$$\bar{y} = \frac{\sum_i Y_i}{n} = \frac{2.637{,}252}{24} = 109{,}8855$$

$$\bar{x} = \frac{\sum_i X_i}{n} = \frac{85{,}249992}{24} = 3{,}552083$$

$$a = 109{,}8855 - (-5{,}3649) \cdot 3{,}552083 = 128{,}942$$

El modelo estimado será:

$$\hat{Y}_i = 128{,}942 - 5{,}3649 \cdot X_i$$

2. El coeficiente de determinación, R^2, será:

$$R^2 = \frac{SCE}{SCT}$$

donde:

$$SCT = \sum_i y_{ci}^2 = 159{,}8926$$

$$SCE = b \cdot \sum_i x_{ci} y_{ci} = (-5{,}3649)(-16{,}5862) = 88{,}9833$$

por diferencia:

$$SCR = SCT - SCE = 159{,}8926 - 88{,}9833 = 70{,}9093$$

El coeficiente de determinación:

$$R^2 = \frac{88{,}983}{159{,}8926} = 0{,}5565$$

El 55,65 % de las variaciones en la demanda de los aparatos de radio está explicada por el ajuste mínimo-cuadrático sobre su precio.

3. Hay que construir un intervalo de predicción para un valor concreto, al 95 %, y comprobar si la demanda igual a 80.000 unidades pertenece a él.

Predicción puntual:

$$X_f = 10$$

$$\hat{Y}_f = a + bX_f = 128{,}942 - 5{,}3649 \cdot 10 = 75{,}193$$

El intervalo será:

$$P\left[\hat{Y}_f - t_{n-2}(\alpha/2) \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_i x_{ci}^2}} < Y_f \Big/ X = X_f \leqslant \right.$$

$$\left. \leqslant \hat{Y}_f + t_{n-2}(\alpha/2) \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_i x_{ci}^2}} \right]$$

donde:

$$t_{\text{tco}}\colon t_{24-2}(0{,}05/2) = t_{22}(0{,}025) = 2{,}0739$$

La varianza residual:

$$S^2 = \frac{SCR}{n-2} = \frac{70{,}9093}{24-2} = 3{,}22315$$

y su desviación típica:

$$S = 1{,}7953$$

El intervalo tomará los valores:

$$I_c\left(Y_f \Big/ X = X_f\right) \equiv \left[\hat{Y}_f \pm t_{n-2}(\alpha/2) \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_i x_{ci}^2}} \right] \equiv$$

$$\equiv \left(75{,}193 \pm 2{,}0739 \cdot 1{,}7953 \cdot \sqrt{1 + \frac{1}{24} + \frac{(10 - 3{,}552083)^2}{3{,}09161}} \right) \equiv$$

$$\equiv (75{,}193 \pm 14{,}17269) \equiv (61{,}02031; 89{,}36569)$$

Como el valor de 80.000 pesetas pertenece al intervalo, se puede afirmar, con un 95 %, que si el precio de los aparatos de radio fuera de 10.000 pesetas, la demanda se cifrará en 80.000 pesetas.

4. La hipótesis de significación global, en el modelo simple de dos variables, se formula igual de la de significación individual:

$$H_0\colon \beta = 0 \qquad H_1\colon \beta \neq 0$$

si la hipótesis es cierta:

$$F_{\exp} = \frac{\dfrac{SCE}{1}}{\dfrac{SCR}{n-2}} \sim F^1_{n-2}$$

los estadísticos serán:

$$F_{\exp} = \frac{\dfrac{88,9833}{1}}{\dfrac{70,9093}{24-2}} = 27,6075$$

$$F_{\text{tco}}: F^{1}_{22}(0,05) = 4,3009$$

Como $F_{\exp} > F_t$, se rechaza H_0, es decir, el modelo sí es globalmente significativo.

Ejercicio 3

Un analista financiero está interesado en conocer la relación existente entre el Índice General de la Bolsa de Madrid (Y) y el tipo de cambio a largo plazo fijado por el Banco de España (X). Para ello construye un modelo de regresión del que obtiene los siguientes resultados:

$$\hat{Y}_i = 750 - 1,2 \cdot X_i \qquad i = 1, 2, ..., 36$$
$$R^2 = 0,85$$
$$S_x = 2,34$$

A partir de esta información usted debe responder a las siguientes preguntas:

1. ¿Es sensible al tipo de interés el nivel del Índice General de la Bolsa de Madrid?
2. ¿Cuál es el incremento del Índice General de la Bolsa de Madrid si el tipo de interés aumenta en 7 unidades?

SOLUCIÓN

1. Hay que realizar un contraste de significación individual:

$$H_0: \beta = 0 \qquad H_1: \beta \neq 0$$

si la hipótesis nula es cierta:

$$t_{\exp} = \frac{b}{S_b} \sim t_{n-2}$$

o bien:

$$F_{exp} = \frac{\dfrac{R^2}{k-1}}{\dfrac{1-R^2}{n-k}} \sim F_{n-k}^{k-1}$$

los estadísticos experimental y teórico serán:

$$F_{exp} = \frac{\dfrac{0,85}{2-1}}{\dfrac{1-0,85}{36-2}} = 192,6667$$

F_{tco}: $F_{34}^{1}(0,05) = 4,13$ (al 95 % de confianza)

Como $F_{exp} > F_t$, se rechaza H_0. El tipo de cambio a largo plazo influye sensiblemente en el Índice General de la Bolsa de Madrid.

2. Como

$$b = \frac{\Delta y}{\Delta x} \quad ; \quad \Delta y = b \cdot \Delta X$$

$$\Delta y = -1,2 \cdot 7 = -8,4$$

Si el tipo de interés incrementa en 7 unidades su valor, el IGB de Madrid disminuirá en 8,4 unidades.

Ejercicio 4

Una empresa de telefonía está interesada en operar en el mercado español de llamadas internacionales. Antes de iniciar este proyecto de inversión desea cuantificar los ingresos del sector en función del precio de dichas llamadas en España. Para ello plantea el siguiente modelo:

$$Y_i = \alpha + \beta \cdot X_i + u_i$$

donde:

Y: Ingresos anuales del sector de telefonía internacional en miles de millones de pesetas.

X: Precio medio de las llamadas internacionales en cientos de pesetas por paso.

Suponga que usted es consultado por esta empresa, que le facilita la siguiente información:

$$n = 35 \qquad \Sigma X = 17{,}22 \qquad \Sigma Y = 84{,}945$$
$$SCT = 12{,}144 \qquad R^2 = 0{,}7378 \qquad S_{XY} = -0{,}134$$

Las cuestiones de interés para el operador de telefonía que usted debe responder son:

1. Estime el modelo teórico planteado.
2. Construya un intervalo de confianza para los ingresos autónomos del sector. Considere un nivel de confianza del 95 %.
3. ¿Considera usted que el precio medio es significativo en la explicación de los ingresos? Considere un nivel de significación del 5 %.
4. ¿Qué nivel de ingresos se espera para el año próximo año si el precio medio previsto es de 0,52 unidades?

SOLUCIÓN

1. Los estimadores de MCO para los parámetros α y β del modelo son:

$$b = \frac{\sum_{i=1}^{n} x_{ci} y_{ci}}{\sum_{i=1}^{n} x_{ci}^2}$$
$$a = \overline{Y} - b \cdot \overline{X}$$

Para obtener los valores es necesario recurrir a:

$$R^2 = \frac{SCE}{SCT}$$

es decir:

$$0{,}7378 = \frac{SCE}{12{,}144}$$

La SCE será:

$$SCE = 0{,}7378 \cdot 12{,}144 = 8{,}9598$$

pero, a su vez,

$$SCE = b \cdot \sum_i x_{ci} y_{ci} = b \cdot n \cdot S_{x,y}$$

por tanto:

$$b = \frac{SCE}{n \cdot S_{x,y}} = \frac{8{,}9598}{35 \cdot (-0{,}134)} = \frac{8{,}9598}{-4{,}69} = -1{,}9104$$

Para obtener a:

$$\bar{Y} = \frac{\sum_i Y_i}{n} = \frac{84{,}945}{35} = 2{,}427$$

$$\bar{X} = \frac{\sum_i X_i}{n} = \frac{17{,}22}{35} = 0{,}492$$

$$a = 2{,}427 - (-1{,}9104) \cdot 0{,}492 = 3{,}3669$$

$$\hat{Y}_i = 3{,}3669 - 1{,}9104 \cdot X_i$$

2. El intervalo de confianza para los ingresos autónomos, la ordenada en el origen, será:

$$I_c(\alpha) \equiv [a \pm t_{n-2}(\alpha/2) \cdot S_a]$$

Se van a ir obteniendo los datos necesarios:

$$S_a^2 = S_u^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i x_{ci}^2} \right)$$

donde S^2 es la varianza estimada de la perturbación:

$$S^2 = \frac{SCR}{n-2}$$

La SCR se obtiene de la descomposición de la varianza, por diferencia:

$$SCR = SCT - SCE$$

$$SCR = 12{,}144 - 8{,}9598 = 3{,}1842$$

por tanto:

$$S^2 = \frac{3{,}1842}{35 - 2} = 0{,}09649$$

Para obtener la suma del regresor centrado al cuadrado, se puede utilizar la fórmula de b:

$$b = \frac{\sum_i x_{ci} y_{ci}}{\sum_i x_{ci}^2}$$

$$\sum_i x_{ci}^2 = \frac{\sum_i x_{ci} y_{ci}}{b} = \frac{-4{,}69}{-1{,}9104} = 2{,}4549$$

La varianza estimada de a:

$$S_a^2 = 0{,}09649\left(\frac{1}{35} + \frac{0{,}492^2}{2{,}4549}\right) = 0{,}01227$$

y su desviación típica:

$$S_a = \sqrt{0{,}01227} = 0{,}11077$$

Los límites del intervalo al 95 % de confianza serán:

$$I_c(\alpha) = (3{,}3669 \pm 2{,}0345 \cdot 0{,}11077) = (3{,}3669 \pm 0{,}22536) = (3{,}1415; 3{,}5922)$$

donde

$$t_{\text{tco}} = t_{35-2}(0{,}05/2) = t_{33}(0{,}025) = 2{,}0345$$

3. La hipótesis de significación de la variable explicativa se refleja en:

$$H_0\colon \beta = 0 \qquad H_1\colon \beta \neq 0$$

Si la hipótesis nula es cierta:

$$t_{\exp} = \frac{b}{S_b} \sim t_{n-2}$$

Se obtiene, en primer lugar, la varianza estimada de b:

$$S_b^2 = \frac{S_u^2}{\sum_i x_{ci}^2} = 0{,}09649 \cdot \frac{1}{2{,}4549} = 0{,}0393$$

y su desviación típica:

$$S_b = \sqrt{S_b^2} = \sqrt{0{,}0393} = 0{,}1982$$

El estadístico experimental será:

$$t_{\exp} = \frac{-1{,}9104}{0{,}1982} = -9{,}636$$

El teórico será:

$$t_{35-2}\left(\frac{0{,}05}{2}\right) \equiv t_{33}(0{,}025) = 2{,}0345$$

Como $|t_{\exp}| > |t_{\text{tco}}|$, rechazamos H_0, es decir, el precio medio de las llamadas internacionales es significativa.

4. Se busca la predicción puntual, cuando el regresor sea 0,52 unidades, es decir;

$$X_f = 0{,}52$$
$$\hat{Y}_f = a + b \cdot X_f = 3{,}3669 - 1{,}9140 \cdot 0{,}52 = 2{,}37349$$

Si el precio de las llamadas es de 5,2 pesetas por paso, los ingresos anuales serán 2.373 millones de pesetas.

Ejercicio 5

En el estudio de la función de consumo (*CONS*) de un país se especificó un modelo lineal a partir de la renta disponible (*YD*) y se estimó por MCO, a través del programa Eviews, obteniéndose la siguiente salida de regresión incompleta:

```
==========================================================================
LS // Dependent Variable is CONS
Sample: 1 15
Included observations: 15
==========================================================================
        Variable        Coefficien   Std. Error   t-Statistic       Prob.
==========================================================================
           C             2.126752     0.427095
           YD            0.803865                   46.25190
==========================================================================
R-squared                            Mean dependent var           20.13333
Adjusted R-squared                   S.D. dependent var           8.433493
S.E. of regression       0.680184    Akaike info criter          -0.647218
Sum squared resid        6.014451    Schwarz criterion           -0.552812
Log likelihood                       F-statistic
Durbin-Watson stat       1.998103    Prob(F-statistic)            0.000000
==========================================================================
```