Segundo Taller

jueves, 24 de abril de 2025 2:00 p. m.

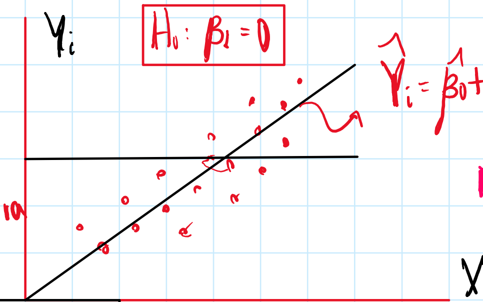
1. Determine el valor de verdad de las siguientes afirmaciones.

- (a) Una suma de cuadrados extra mide la reducción marginal en el SSE cuando una o varias variables predictoras son agregadas al modelo de regresión, dado que las otras predictoras ya fueron agregadas o están en el modelo.

Verdadera : P.H : Significancia

$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i;\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$: R^2 :

$H_0: \beta_1 = 0$



$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}$

No sig.

- (b) El estadístico T correspondiente al procedimiento de prueba empleado para probar la significancia marginal del j- ésimo parámetro es:

$$T_{j,0} = \frac{\hat{\beta}_j - 0}{\sqrt{\sigma^2 c_{jj}}} \sim t_{n-p}$$

Falsa.

Con una región de rechazo asociada de $R_c = \{|T_0| > t_{\alpha/2, n-p}\}$ y p- valor $P(|t_{n-p}| > |T_{j,0}|)$.

- (c) Valores grandes de R^2 implican que la superficie ajustada de respuesta es útil; sin embargo, es menos preferido que R^2_{adj} como medida de bondad de ajuste.

Falso

Verdadera

$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i;\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$j = 0, \dots, k$; σ^2 desconocida ; $\boxed{\hat{\sigma}^2 = MSE \longrightarrow \sigma^2}$

- $T_{j,0} = \dfrac{\hat{\beta}_j - 0}{\sqrt{\hat{\sigma}^2 c_{jj}}} = \dfrac{\hat{\beta}_j - 0}{\sqrt{MSE\, c_{jj}}} \sim t_{n-p}$; $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$; $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$; $\hat{\underline{\beta}} \sim N(\underline{\beta}, (X^TX)^{-1}\sigma^2)$

$c_{jj} = \mathrm{diag}(X^TX)^{-1}$

R^2: proporción $0 \leq R^2 \leq 1$: la proporción de variabilidad explicada por el modelo dada la relación lineal entre la variable respuesta y las demás covariables

- $R^2_{adj} \gg R^2$ → Principio de parsimonia : Afirmación falsa.

No

- (d) Rechazar la hipótesis nula $H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$ permite determinar la significancia global de un modelo de regresión $Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ a través de un procedimiento de sumas de cuadrados extra o procedimiento lineal general. Falso

: Significancia global: Probar el ef. covariables

$H_0: \beta_1 = \dots = \beta_k = 0$

H_1: Algún $\beta_j \neq 0$, $j = 1, \dots, k$

- (e) El estadístico F correspondiente al procedimiento de prueba empleado para probar la significancia global del modelo de regresión lineal múltiple es:

$$F_0 = \frac{SSR/k}{SSE/(n-k)} \sim f_{k,n-k}$$

Falso

Con una región de rechazo asociada de $R_c = \{F_{calc} > f_{\alpha,k,n-k}\}$ y p- valor $P(f_{k,n-k} > F_{calc})$.

$F_0 = \dfrac{SSR/k}{MSE} = \dfrac{SSR/k}{SSE/(n-p)} \sim f_{k,n-p}$

$p = k+1$ → # parámetros

Número observ.

R_c: $\{F_0 > f_{\alpha,k,n-p}\}$; $P(f_{k,n-p} > F_0)$

- (f) Se puede demostrar que $F_{j,0} = T^2_{j,0}$ únicamente si se pretende determinar la significancia marginal de un coeficiente de regresión, lo que implica que $P(f_{1,n-p} > F_{j,0}) \equiv P(|t_{n-p}| > |T_{j,0}|)$.

- (g) Los grados de libertad del cuadrado medio debido a la hipótesis son iguales al rango de la matriz $\mathbf{L}$, asociada la prueba lineal general ($H_0 : \mathbf{L}\beta = 0$ vs $H_1 : \mathbf{L}\beta \neq 0$).

Verdadero.

Probar significancia individual: $\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}$ $\forall j = 0, \dots, k$

- $T_{j,0} = \dfrac{\hat{\beta}_j - 0}{\sqrt{\hat{\sigma}^2 c_{jj}}} = \dfrac{\hat{\beta}_j - 0}{\sqrt{MSE\, c_{jj}}} \sim t_{n-p}$

$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i;\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$F_0 = \dfrac{SSR/1}{SSE/(n-p)} \sim f_{1,n-p}$

$k = 1$: $T_{j,0}^2 = F_{j,0}$

- $P(f_{1,n-p} > F_{j,0}) = P(|t_{n-p}| > |T_{j,0}|)$

$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

$\mathbf{L}\underline{\beta} = \underline{0}$ Prueba lineal general : $F = \dfrac{MSH}{MSE} = \dfrac{SSH/r}{MSE}$: $L = [0\ 1\ 0\ 1]$ = $r = 2$

$L\beta = 0$ Prueba lineal general: $F = \frac{MSH}{MSE} = \frac{SSH/r}{MSE}$: $L = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ = $r = 2$ gl

Rango matriz: # filas no nulas linealmente independientes

Parte práctica: Y: Performance; X_1: Fuerza; X_2: Habilidades; X_3: Velocidad

- $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$; $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$: $Y = X\beta + \varepsilon$; $\varepsilon_n \sim N_n(0, \sigma^2 I_n)$

$\widehat{Per}_i = \hat\beta_0 + \hat\beta_1 F_i + \hat\beta_2 H_i + \hat\beta_3 V_i$

$\widehat{Per}_i = 70,80 + 0,88 X_{i1} + 1,87 X_{i2} + 3,04 X_{i3}$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 70.80225    2.45665   28.82   <2e-16 ***
X1           0.88463    0.02848   31.06   <2e-16 ***
X2           1.87368    0.02073   90.41   <2e-16 ***
X3           3.04502    0.02789  109.19   <2e-16 ***
---
```

2. Determine la significancia de la regresión global a través del método de sumas de cuadrados extra. ¿Cree usted que puede realizarse esta prueba empleando otro método? De ser así, pruébelo.

$\begin{cases} H_0: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_1: \text{Algún } \beta_j \neq 0;\ j = 1,2,3 \end{cases}$

- Considerar un modelo completo y reducido: MF: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$

MR: $Y_i = \beta_0 + \varepsilon_i$; $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

w $s = 0$ # cov. mod red.

$F = \frac{[SSR(MF) - SSR(MR)]/v}{MSE(MF)} \sim f_{v, n-p} = \frac{[SSE(MR) - SSE(MF)]/v}{MSE(MF)}$

$R_c: \{F_{cal} > f_{\alpha, v, n-p}\}$; $P(f_{v,n-p} > F_{cal})$

- Sumas de cuadrados extra

$F = \frac{SSR(\beta_1, \beta_2, \beta_3 \mid \beta_0)}{MSE(\beta_0, \beta_1, \beta_2, \beta_3)} = \frac{[SSR(\beta_0, \beta_1, \beta_2, \beta_3) - SSR(\beta_0)]/v}{MSE(\beta_0, \beta_1, \beta_2, \beta_3)} = \frac{[SSE(\beta_0) - SSE(\beta_0, \beta_1, \beta_2, \beta_3)]/v}{MSE(\beta_0, \beta_1, \beta_2, \beta_3)}$

v: Grados de libertad : $v = gl(SSR(MF)) - gl(SSR(MR)) = gl(SSE(MR)) - gl(SSE(MF))$ = $k - s$ = $(n-p) - (n-m) = -p + m$

- En resumen: v = # parámetros en la hipótesis nula

$v = 3$

$n = 500$ $p = 4$

```
> anova(modelo_reducido)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value Pr(>F)
Residuals 499 2183751  4376.3
```

$SSE(MR) = 2183751$

```
> MSE <- SSE_completo/(n- p)
> MSE
[1] 103.6835
```

```
> anova(modelo)
Analysis of Variance Table

Response: Y
           Df  Sum Sq Mean Sq  F value    Pr(>F)
X1          1   78308   78308   755.26 < 2.2e-16 ***
X2          1  817773  817773  7887.27 < 2.2e-16 ***
X3          1 1236243 1236243 11923.33 < 2.2e-16 ***
Residuals 496   51427     104
```

$SSE(MF) = 51427$

$v: 499 - 496 = 3$

$F = \frac{[2183751 - 51427]/3}{103,6835}$

```
> anova_comparacion <- anova(modelo_reducido, model
> anova_comparacion
Analysis of Variance Table

Model 1: Y ~ 1
Model 2: Y ~ X1 + X2 + X3
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1    499 2183751
```

→ Utilizando el truco

- α: A un nivel de significancia del 5% se puede establecer que el modelo el

```
Model 1: Y ~ 1
Model 2: Y ~ X1 + X2 + X3
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1    499 2183751
2    496   51427  3   2132324 6855.3 < 2.2e-16 ***
---
```

¿ α: A un nivel de significancia del 5% se puede establecer que el modelo es significativo: Al menos un β_j es diferente de cero / Al menos un β_j es significativo

- Esto también se puede con la prueba lineal general

$$\begin{cases} H_0: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_1: \text{Algún } \beta_j \neq 0 \ ; \ j = 1,2,3 \end{cases} \quad : \quad L\beta = 0$$

H_0: $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$ vs H_1: Algún $\beta_j \neq 0$; $j = 1,2,3$

- $L\beta = 0$:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\boxed{r = 3}$ $\quad F = \dfrac{SSH/r}{MSE} = \boxed{F = \dfrac{(2183751 - 51427)/3}{103{,}6835}}$

3. Determine la significancia de los parámetros individuales β_j, junto con intervalo de confianza. Brinde una interpretación apropiada.

$$\begin{cases} H_0: \beta_j = 0 \ ; \ j = 0,1,2,3 \\ H_1: \beta_j \neq 0 \end{cases}$$

$$T_{0} = \frac{\hat{\beta_0} - 0}{\sqrt{Var[\hat{\beta_0}]}} = \frac{\hat{\beta_0} - 0}{\sqrt{MSE\,C_{00}}} \sim t_{n-p} \quad : \quad P(|t_{n-p}| \geq |T_0|) \ ; \ R_c = \{|T_0| > t_{\alpha/2,\, n-p}\}$$

$\boxed{\alpha = 0{,}05}$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 70.80225    2.45665   28.82   <2e-16 ***
X1           0.88463    0.02848   31.06   <2e-16 ***
X2           1.87368    0.02073   90.41   <2e-16 ***
X3           3.04502    0.02789  109.19   <2e-16 ***
---
```

- A un nivel de significancia del 5%, se puede establecer que el intercepto es significativo en presencia de las demás cov.

- A un nivel de sig. del 5%, se puede establecer que el efecto de la fuerza es significativo ($\hat{\beta_1}$ es significativamente distinto de cero) en presencia de las demás cov.

- Intervalo de confianza: $\boxed{\hat{\beta_j} \pm t_{\alpha/2,\, n-p} \cdot \sqrt{Var[\hat{\beta_j}]}}$: $\boxed{\hat{\beta_1} \pm t_{0{,}025,\, 500-4} \cdot \sqrt{Var[\hat{\beta_1}]}}$

```
> confint(modelo)
                 2.5 %     97.5 %
(Intercept) 65.9755369 75.6289713
X1           0.8286681  0.9405966
X2           1.8329626  1.9144035
X3           2.9902333  3.0998132
>
```

- Con una confianza del 95% se puede establecer que el efecto promedio de la fuerza está comprendido entre 0,82 y 0,94 unidades; es decir, que por unidad de cambio en la fuerza, el rendimiento promedio varía entre 0,82 y 0,94 unidades

En presencia demás covariables.

De esto depende la cobertura de un intervalo

Verificar la significancia de parámetros : si el IC contiene al cero, entonces, el parámetro no es signif.

4. Determine si el efecto de la primera covariable es el mismo que el efecto de la tercera covariable; al mismo tiempo, verifique si el correspondiente efecto de la primera co-

$H_0: \beta_1 = \beta_3$ vs $H_1: \beta_1 \neq \beta_3$

no es signif.

4. Determine si el efecto de la primera covariable es el mismo que el efecto de la tercera covariable; al mismo tiempo, verifique si el correspondiente efecto de la primera covariable es el mismo que el de la segunda covariable. Plantee una prueba de hipótesis para ello y realice el procedimiento adecuado. **Reporte el modelo completo y el modelo reducido.**

$H_0: \beta_1 = \beta_3$ vs $H_1: \beta_1 \neq \beta_3$ v
$\beta_1 = \beta_2$ $\qquad \beta_1 \neq \beta_2$

- $\boxed{H_0: \begin{array}{l}\beta_1 - \beta_3 = 0 \\ \beta_1 - \beta_2 = 0\end{array} \text{ vs } H_1: \begin{array}{l}\beta_1 \neq \beta_3 \text{ v} \\ \beta_1 \neq \beta_2\end{array}}$

$$L\beta = 0 \rightarrow \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \boxed{r = 2}$$

MF: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$

MR: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_1 X_{i2} + \beta_1 X_{i3} + \varepsilon_i$

$Y_i = \beta_0 + \beta_1 \underbrace{(X_{i1} + X_{i2} + X_{i3})}_{X_{123}} + \varepsilon_i$

$$\boxed{F = \frac{SSH/2}{MSE}}$$

```
> anova_LG <- anova(modelo_LG, modelo)
> anova_LG
Analysis of Variance Table

Model 1: Y ~ X123
Model 2: Y ~ X1 + X2 + X3
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    498 366319
2    496  51427  2    314893 1518.5 < 2.2e-16 ***
---
```

$$F = \frac{[366319 - 51427]/2}{MSE} = 1518,5$$

$< \alpha$

A un nivel de significancia del 5%, se puede rechazar la hipótesis nula, así, se puede establecer que el efecto de la fuerza es distinto al efecto de la Velocidad, o bien, el efecto de la primera covariable es distinto al efecto de las habilidades.

$\boxed{\text{Se cumple una o ambas}}$