Quinto taller

sábado, 4 de octubre de 2025 11:50 a. 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.
   - (b) El estadístico T correspondiente al procedimiento de prueba empleado para probar la significancia marginal del j- ésimo parámetro es:

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

     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.

Parte teórica

(a). Verdadera

(b). Falso

- Varianza σ^2; Estimación: $\hat{\sigma}^2 = MSE$

$Z = \frac{\hat{\beta}_j - 0}{\sqrt{\sigma^2 c_{jj}}} \sim N(0,1)$ (conocida)

$T = \frac{\hat{\beta}_j - 0}{\sqrt{MSE\, c_{jj}}} \sim t_{n-p}$ T- Centrada

(c). Falsa:

Primera parte ( ): No es correcto porque no explica la utilidad de la superficie respuesta

- $0 < R^2 < 1$: la proporción de variabilidad explicada por el modelo, dada la relación lineal entre la variable respuesta (Y); las demás covariables.
- $R^2 > R^2_{adj}$: El R^2_{adj} penaliza el número de covariables en el modelo
- Siempre se busca el modelo más parsimonioso

- (d) Rechazar la hipótesis nula $H_0 : \beta_0 = \beta_1 = ... = \beta_k = 0$ permite determinar la significancia global de un modelo de regresión $Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ a través de un procedimiento de sumas de cuadrados extra o procedimiento lineal general.
- (e) El estadístico F correspondiente al procedimiento de prueba empleado para probar la significancia global del modelo de regresión lineal múltiple es:

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

  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) 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 $H_1 : \mathbf{L}\beta \neq 0$).

(d). Falso: $\begin{cases} H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1: \text{Algún } \beta_j \neq 0;\ j = 1, \dots, k \end{cases}$

(e). Falso: $F = \frac{SSR/k}{SSE/n-p} = \frac{MSR}{MSE} \sim f_{k,n-p}$

$p = k+1$ $R_c = \{F_0 > f_{\alpha,n-p}\}$; $P(f_{k,n-p} > F_0)$

(f). $F_{j,0} = T^2_{j,0}$ Regresión lineal simple

$F_{j,0} \sim f_{1,n-p}$: $P(f_{1,n-p} > F_{j,0}) \equiv P(|t_{n-p}| > |T_{j,0}|)$

$F = \frac{SSR/1}{MSE} \sim f_{1,n-p}$

(g). Verdadero: $\begin{cases} H_0: \mathbf{L}\beta = 0 \\ H_1: \mathbf{L}\beta \neq 0 \end{cases}$ $\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$

$F = \frac{MSH}{SSE/n-p} = \frac{SSH/r}{MSE} \sim f_{r,n-p}$

$\mathbf{L} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $r = 2$

$R(\mathbf{L})$ = # Filas linealmente independientes no-nulas

$\mathbf{L}$: Grados de libertad del estadístico

Parte práctica

Table 1: Información en análisis

Performance (Y_i)	Strength (X_1)	Skills (X_2)	Speed (X_3)
337.4335	60.22040	37.55007	54.67685
418.0430	45.23407	72.26675	53.60737
413.5214	32.09499	86.78852	44.96376
378.6311	63.44574	44.25407	53.85836
428.8907	49.40831	93.26630	43.28241

Considere a *'Overall Performance'* como la variable respuesta. *Las covariables en análisis se especifican en la tabla mostrada con anterioridad.* **Suponga que los supuestos del modelo se cumplen. De respuesta a los siguientes planteamientos:**

1. Determine cuál es el modelo empleado en esta situación, junto con sus supuestos, además, reporte la recta de regresión ajustada.

(1). $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$
$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$\Rightarrow Y_n = X\beta + \varepsilon$; $\varepsilon \sim N_n(0, \sigma^2 I)$

Modelo ajustado: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3}$

$\widehat{Per}_i = \hat{\beta}_0 + \hat{\beta}_1 S_{ti} + \hat{\beta}_2 S_{ki} + \hat{\beta}_3 S_{pi}$

se especifican en la tabla mostrada con anterioridad. **Suponga que los supuestos del modelo se cumplen. De respuesta a los siguientes planteamientos:**

1. Determine cuál es el modelo empleado en esta situación, junto con sus supuestos, además, reporte la recta de regresión ajustada.

$$\hat{Per_i} = \hat\beta_0 + \hat\beta_1 S_{ti} + \hat\beta_2 S_{ki} + \hat\beta_3 S_{pi}$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 70.80225    2.45665   28.82   <2e-16 ***
Strength     0.88463    0.02848   31.06   <2e-16 ***
Skills       1.87368    0.02073   90.41   <2e-16 ***
Speed        3.04502    0.02789  109.19   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.18 on 496 degrees of freedom
Multiple R-squared:  0.9765,    Adjusted R-squared:  0.9763
F-statistic:  6855 on 3 and 496 DF,  p-value: < 2.2e-16
```

$$\hat{Per_i} = 70,80 + 0,88 S_{ti} + 1,87 S_{ki} + 3,04 S_{pi}$$

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.

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1:$ Algún $\beta_j \neq 0$; $j = 1,2,3$

$$F = \frac{SSR/h}{SSE/n-p} \sim f_{h,n-p}$$

$F = 6855$

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

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

$\begin{cases} MF: Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \\ MR: Y_i = \beta_0 + \varepsilon_i \;; \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \end{cases}$

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

v = # Grados de libertad:

$v = gl(SSR(MF)) - gl(SSR(MR))$
$gl(SSE(MR)) - gl(SSE(MF))$

Truco: v = # parámetros en la hipótesis nula

```
> myAllRegTable(modelo)
  k  R_sq adj_R_sq       SSE        Cp    Variables_in_model
1 1 0.545    0.544  994638.7  9097.102                 Speed
2 1 0.373    0.372 1369507.8 12712.643                Skills
3 1 0.036    0.034 2105443.3 19810.602              Strength
4 2 0.931    0.930  151433.3   966.546          Skills Speed
5 2 0.588    0.587  898833.1  8175.076        Strength Speed
6 2 0.410    0.408 1287670.0 11925.333       Strength Skills
7 3 0.976    0.976   51426.6     4.000 Strength Skills Speed
>
```

$n = 500$

$$F = \frac{[2183751 - 51426,6]/3}{51426,6 / 500-4}$$

- $SSR(MR) = SSR(\beta_0) = SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$
$= 2183751$

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

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

$$F = \frac{[2183751 - 51426,6]/3}{51426,6/496}$$

$P_{val} < \alpha$

- Si, se puede usar otro método

$\begin{cases} H_0: L\beta = 0 \\ H_1: L\beta \neq 0 \end{cases}$: $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1:$ Algún $\beta_j \neq 0$; $j = 1,2,3$

$H_0: \begin{matrix} \beta_1 = 0 \\ \beta_2 = 0 \\ \beta_3 = 0 \end{matrix}$: $\begin{matrix} \beta_1 + 0\beta_2 + 0\beta_3 = 0 \\ 0\beta_1 + \beta_2 + 0\beta_3 = 0 \\ 0\beta_1 + 0\beta_2 + \beta_3 = 0 \end{matrix}$

$L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$; $v = 3$

$$F = \frac{MSH}{MSE}$$

$F = 6855$

$P_{val} < \alpha = 0,05$

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

$H_0: \beta_j = 0$; $j = 0, \dots, h$
$H_1: \beta_j \neq 0$

$$T = \frac{\hat\beta_j - 0}{\sqrt{MSE\,C_{jj}}} \sim t_{n-p}$$

$$T_{0,0} = \frac{70,802}{2,456} = 28,82$$

$P_{val} < \alpha$: A un nivel de sign del 5% ...

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

$$\hat\beta_j \pm t_{\alpha/2, n-p} \cdot \sqrt{Var[\hat\beta_j]}$$

$0,884 \pm t_{\alpha/2, n-p} \cdot 0,02848$

```
(Intercept) 70.80225   2.45665   28.82  <2e-16 ***
Strength     0.88463   0.02848   31.06  <2e-16 ***
Skills       1.87368   0.02073   90.41  <2e-16 ***
Speed        3.04502   0.02789  109.19  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.18 on 496 degrees of freedom
Multiple R-squared:  0.9765,	Adjusted R-squared:  0.9763
F-statistic:  6855 on 3 and 496 DF,  p-value: < 2.2e-16
```

$\boxed{\hat\beta_j \pm t_{\alpha/2,n-p} \cdot \sqrt{\widehat{var}[\hat\beta_j]}}$

$\boxed{0{,}884 \pm t_{\alpha/2,n-p} \cdot 0{,}02848}$

$\boxed{0{,}884 \pm 1{,}64 \cdot 0{,}02848}$ $\boxed{LI \; ; \; LS}$

Con una confianza del 95%

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.**

$H_0: \beta_1 = \beta_3 \; : \; \beta_1 - \beta_3 = 0$

$\quad\;\; \beta_1 = \beta_2 \; : \; \beta_1 - \beta_2 = 0$

$H_1: \beta_1 \neq \beta_3$

$\quad\;\; \beta_1 \neq \beta_2$

$$L\beta = 0: \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\boxed{r = 2}$

$F = \dfrac{SSH/2}{MSE} \sim f_{2,n-p}$

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

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

$\quad = \boxed{\beta_0 + \beta_1 (X_{i1} + X_{i2} + X_{i3}) + \varepsilon_i}$