1. Considere las siguientes afirmaciones y determine su valor de verdad.

a) La matriz sombrero cumple que $H = (H^{2021})^T$.

b) Suponga que se ajustó el modelo $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$ y se tiene que $gl(SSE) = 98$. El modelo fue ajustado con 102 observaciones.

c) Una suma de cuadrados extra, mide la reducción marginal en la SSE.

d) En la hipótesis lineal general ($H_0 : \mathbf{L}\beta = 0$ vs $H_1 : \mathbf{L}\beta \neq 0$) los grados de libertad del cuadrado medio debido a la hipótesis son iguales al rango de la matriz $\mathbf{L}$

a) H es simétrica e idempotente $\Rightarrow H^n = H$ y $H^T = H$

$(H^{2021})^T = (H)^T = H \Rightarrow$ es verdadero.

$$\underline{\hat{y}} = X\underline{\hat{\beta}} = X\underbrace{(X^TX)^{-1}X^T\underline{y}}_{\underline{\hat{\beta}}} = \underbrace{[X(X^TX)^{-1}X^T]}_{H \text{ (matrizhat)}}\underline{y}$$

b) ¿Quién es p? # de parámetros (cuantos β's hay en el modelo)

¿Quién es K? # de covariables

$$p = \underbrace{K}_{\text{cuantas covariables tengo}} + \overbrace{1}^{\text{el } \beta_0}$$

$gl(SSE) = n - p = 98 \Rightarrow n = 98 + p = 98 + 4 = 102$

Es verdadero

c) Verdadero (teoría), también se puede medir el aumento en la SSR.

c) Es una definición.

Nota: OJO el rango $\neq$ # filas

Ej: $\begin{cases} H_0: L\underline{\beta} = \underline{0} \\ H_1: L\underline{\beta} \neq 0 \end{cases}$ $\quad \beta = [\beta_0\ \beta_1\ \beta_2]^T$

$\begin{cases} H_0: \beta_1 = \beta_2\ ;\ 2\beta_1 = 2\beta_2 \quad (\text{¡!}) \\ H_1: \text{Alguna de los de arriba no se cumple} \end{cases}$

esta es L

¡! matricialmente $\begin{bmatrix} 0 & 1 & -1 \\ 0 & 2 & -2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Downarrow$

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

Anotaciones R

```
12 mod1 <- lm(Y ~ X1 + X2 + X3 + X4, data = datos)
```

Esto significa que se ajusta el modelo

(*)

$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_4 X_{4i} + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2);\ 1 \leq i \leq 54$

ó $\underline{Y} = \mathbb{X}\underline{\beta} + \underline{\varepsilon},\ \underline{\varepsilon} \sim N_{54}(\underline{0}, \sigma^2 I_n)$

a)

$\hat{\beta}_J$ $SE(\hat{\beta}_J)$ Valor de $T_{0,J}$ Valor p

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.148707  14.855006   0.279  0.78121
X1           -3.690499   2.970780  -1.242  0.22005
X2            0.009458   0.046297   0.204  0.83897
X3           47.940199  15.709131   3.052  0.00367 **
X4           11.371019   7.868536   1.445  0.15479
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Si el valor-p es $< \alpha \Rightarrow$ se rechaza Ho en el juego de hipótesis $\begin{cases} H_0: \beta_J = 0 \\ H_1: \beta_J \neq 0 \end{cases}$

Si se rechaza Ho eso significa que β_J es significativo dadas las otras covariables

b)

Vamos a ver por qué es marginal (β_3 por ejemplo)

$$\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases}$$

Bajo Ho el modelo se convierte en

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_4 X_{4i} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

Modelo reducido (M_R)

(*) Modelo completo (M_F)

Acá van los que se incluyeron en Ho

lo pega todo

lo que va después del dado

$$SSR(\beta_3 \mid \beta_0, \beta_1, \beta_2, \beta_4) = \underbrace{SSR(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)}_{M_F} - \underbrace{SSR(\beta_0, \beta_1, \beta_2, \beta_4)}_{M_R}$$

$$= \underbrace{SSE(\beta_0, \beta_1, \beta_2, \beta_4)}_{M_R} - \underbrace{SSE(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)}_{M_F}$$

Vamos a definir $F_0 = \dfrac{MSR(\beta_3 \mid \beta_0, \beta_1, \beta_2, \underbrace{\beta_4)}_{M_R}}{\underbrace{MSE(\beta_0, \ldots, \beta_4)}_{M_F}}$

$$MSR(\beta_3 \mid \beta_0, \beta_1, \beta_2, \beta_4) = \frac{SSR(\beta_3 \mid \beta_0, \beta_1, \beta_2, \beta_4)}{r}$$

¿Cómo calcular r?

$r = gl(SSE_{reducido}) - gl(SSE_{Full})$ $\quad gl(SSR_{reducido})$

$= (n-4) - (n-5) = \underbrace{5}_{gl(SSR_{Full})} - \overbrace{4}^{gl(SSR_{reducido})} = 1$

Simplemente # parámetros del full - # de parámetros del reducido

Finalmente $F_0 = \dfrac{SSR(\beta_3 \mid \beta_0, \beta_1, \beta_2, \beta_4)/1}{\underbrace{MSE(\beta_0, \ldots, \beta_4)}_{\frac{SSE(\beta_0, \ldots, \beta_4)}{n-p}}} \overset{\text{bajo } H_0}{\sim} F_{1, n-p}$

La única significativa individual (marginal) / es X_3.

Tabla ANOVA

c)

	Sum_of_Squares	DF	Mean_Square	F_Value	P_value
Model	409.934	4	102.4834	3.50058	0.0136397
Error	1434.532	49	29.2762		

SSR, K, $\frac{SSR}{K}=MSR$, $P(F_{K,n-p} > F_0)$

SSE, $n-p$, $\frac{SSE}{n-p}=MSE$, $\frac{MSR}{MSE}=F_0$

Usando $\alpha=0.05$ se rechaza Ho porque p-valor $<\alpha$

¿Quién es Ho en la ANOVA?

$$\begin{cases} H_0: \beta_1 = \cdots = \beta_4 = 0 \ \{\beta_1=0, \beta_2=0, \ldots, \beta_4=0 \\ H_1: \text{Al menos uno es distinto de } 0 \end{cases}$$

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

$$F_0 = \frac{SSR(\beta_1,\ldots,\beta_4|\beta_0)/r}{MSE(\beta_0,\ldots,\beta_4)} = \frac{[SSE(M_R)-SSE(M_F)]/r}{\underbrace{MSE(M_F)}_{\frac{SSE(M_F)}{n-p}}} \sim F_{4,\underbrace{49}_{n-p}}^{r}$$

$r = 5-1 = 4$

También se puede ver como $\begin{cases} H_0: L\underline{\beta} = \underline{0} \\ H_1: L\underline{\beta} \neq 0 \end{cases}$

$$\text{Con } L = \begin{array}{c} \begin{matrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

d)

d1) $\begin{cases} H_0 : \beta_1 = \beta_2 = \beta_4 \\ H_1 : \text{A determinar} \end{cases}$

d2) $\begin{cases} H_0 : \beta_1 + \beta_2 = \beta_4;\ \beta_3 = -\beta_4 \\ H_1 : \text{A determinar} \end{cases}$

d1) $\beta_2 = \beta_1$ y $\beta_4 = \beta_1$ reemplazando en el modelo original quedaría

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_1 X_{2i} + \beta_3 X_{3i} + \beta_1 X_{4i} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$
$$= \beta_0 + \beta_1 X^*_{1i} + \beta_3 X_{3i} + \varepsilon_i$$

Donde $X^*_{1i} = X_{1i} + X_{2i} + X_{4i}$

Matricialmente

$\beta_1 = \beta_2 = \beta_4$ (resto β_1 de todas partes)

$0 = \beta_2 - \beta_1 = \beta_4 - \beta_1$

$$\Rightarrow \begin{cases} \beta_2 - \beta_1 = 0 \\ \beta_4 - \beta_1 = 0 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix}}_{L} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$F_0 = \frac{[SSE(M_R) - SSE(M_F)]/r}{MSE(M_F)} \overset{\text{bajo } H_0}{\sim} F_{r,\, n-p}$$

rango de L

$d_2)$ $\begin{cases} H_0: \beta_1+\beta_2=\beta_4 \text{ ; } \beta_3=-\beta_4 \\ H_1: \text{Al menos una de las igualdades no se cumple} \end{cases}$

Matricial / $H_0: \begin{bmatrix} 0 & \underline{1} & 1 & 0 & -1 \\ 0 & 0 & 0 & \underline{1} & 1 \end{bmatrix} \underline{\beta} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\left.\begin{array}{l} \beta_1 = \beta_4 - \beta_2 \\ \beta_3 = -\beta_4 \end{array}\right\}$ Pivotes en función de las libres

$\left.\begin{array}{l} \beta_0 = \beta_0 \\ \beta_2 = \beta_2 \\ \beta_4 = \beta_4 \end{array}\right\}$ variables libres

$$Y_i = \beta_0 + (\beta_4 - \beta_2)X_{1i} + \beta_2 X_{2i} - \beta_4 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$
$$= \beta_0 + \beta_2 \underbrace{(X_{2i} - X_{1i})}_{X_{2i}^*} + \beta_4 \underbrace{(X_{1i} - X_{3i} + X_{4i})}_{X_{4i}^*}$$

Modelo reducido

$$F_0 = \frac{[SSR(M_F) - SSR(M_R)]/\overset{r=2}{r}}{MSE} \overset{\text{bajo } H_0}{\sim} F_{2,49}$$