1. Responda las siguientes preguntas.

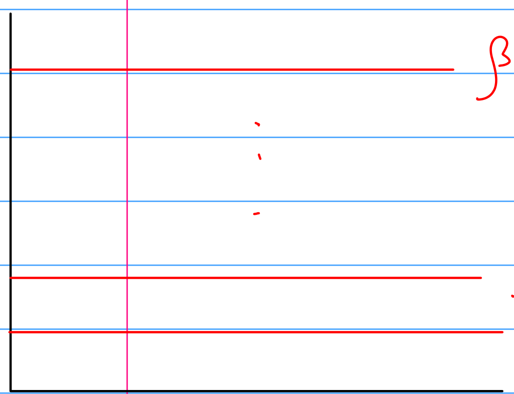
a) Suponga que se ajusta un modelo de regresión con una variable categórica, sin interacción, ¿dicho modelo genera rectas secantes?

b) En un modelo de regresión lineal simple ajustado solo con factores, las rectas generadas son horizontales.

c) El parámetro β_j es la media de Y en la categoría j en el modelo de regresión $Y = \beta_0 + \sum_{k=1}^{c-1} \beta_k I_k$, en caso de que no, ¿cuál es la media?

d) La interacción entre variables numéricas y categóricas hace variar la tasa de cambio de la respuesta en cada categoría de la variable categórica.

a) $Y=\beta_0+\beta_1 I_1+\cdots+\beta_{c-1} I_{c-1}+\varepsilon,\ \varepsilon\sim N(\underline{0},\sigma^2 I_n)$



$\beta_0+\beta_{c-1}$ $I_{c-1}=1$ y las otras son 0

$\beta_0+\beta_1,\ I_1=1$ y $I_2=\cdots=I_c=0$

β_0 $I_1=\cdots I_{c-1}=0,\ I_c=1$

b) Son horizontales, mirar la figura de a)

c) No

$Y=\beta_0+\beta_1 I_1+\cdots+\beta_{c-1} I_{c-1}+\varepsilon,\ \varepsilon\sim N(\underline{0},\sigma^2 I_n)$

i) Suponga $I_J=1$

$Y=\beta_0+\beta_J+\varepsilon \Rightarrow E[Y]=E[\beta_0+\beta_J+\varepsilon]=\beta_0+\beta_J$

ii) $\beta_0=\beta_0\times 1=\beta_0(I_1+I_2+\cdots+I_{c-1}+I_c)$

Veamos que la suma de todas las indicadoras es igual a 1.

$$\sum_{k=1}^{c} I_k = 1 \quad \text{Suponga que } I_J = 1 \Rightarrow I_K = 0 \text{ si } K \neq J$$

$$\| $$

$$\sum_{K=1}^{J-1} \cancel{I_K}^{\,0} + \cancel{I_J}^{\,1} + \sum_{K=J+1}^{c} \cancel{I_K}^{\,0} = 0 + 1 + 0 = 1$$

Reemplazando β_0 por $\beta_0(I_1 + \dots + I_c)$ se tiene que

$$Y = \beta_0(I_1 + \dots + I_c) + \beta_1 I_1 + \dots + \beta_{c-1} I_{c-1} + \varepsilon$$
$$= I_1(\beta_0 + \beta_1) + I_2(\beta_2 + \beta_0) + \dots + I_{c-1}(\beta_0 + \beta_{c-1}) + I_c \beta_0 + \varepsilon$$

Ahora si $I_J = 1$ queda que $Y = \beta_0 + \beta_J + \varepsilon$ y $E[Y] = \beta_0 + \beta_J$

d) $Y = \beta_0 + \beta_1 X_1 + \beta_{1,1} X_1 I_1 + \beta_{1,2} X_1 I_2 + \dots + \beta_{1,c-1} X_1 I_{c-1} + \varepsilon$

Si $I_J = 1$ el modelo se reduce a

$$Y = \beta_0 + \beta_1 X_1 + \beta_{1,J} X_1 \cancel{I_J}^{\,1} + \varepsilon = \beta_0 + \underbrace{(\beta_1 + \beta_{1,J})}_{} X_1 + \varepsilon$$

La Tasa de cambio varió respecto a la original

2) El modelo a ajustar es $mpg = \beta_0 + \beta_1 wt + \beta_2 \, cyl + \beta_3 \, cyl * wt + \varepsilon$

Luego de ajustar en R se obtuvo lo siguiente

```
Call:
lm(formula = mpg ~ wt * cyl, data = datos)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.1513 -1.3798 -0.6389  1.4938  5.2523 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   39.571      3.194  12.389 2.06e-12 ***
wt            -5.647      1.359  -4.154 0.000313 ***
cyl6         -11.162      9.355  -1.193 0.243584    
cyl8         -15.703      4.839  -3.245 0.003223 ** 
wt:cyl6        2.867      3.117   0.920 0.366199    
wt:cyl8        3.455      1.627   2.123 0.043440 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.449 on 26 degrees of freedom
Multiple R-squared:  0.8616,	Adjusted R-squared:  0.8349 
F-statistic: 32.36 on 5 and 26 DF,  p-value: 2.258e-10
```

$\hat{mpg} = 39.571 - 5.647 wt - 11.162 \, I_{cyl=6} - 15.703 \, I_{cyl=8} + 2.867 \, wt \, I_{cyl=6} + 3.455 \, I_{cyl=8}$

Si $cyl = 6 \Rightarrow \hat{mpg} = \underbrace{(39.571 - 11.162)}_{28.409} - \underbrace{(5.647 - 2.867)}_{2.78} wt$

Si $cyl = 8$, el análisis es análogo

3) X_1 = peri, X_2 = perm, X_3 = shape

	k	R_sq	adj_R_sq	SSE	Cp	Variables_in_model
1	1	0.677	0.669	109513013	20.830	peri
2	1	0.157	0.139	285283187	124.882	perm
3	1	0.033	0.012	327309336	149.761	shape
4	2	0.774	0.764	76348142	3.197	peri perm
5	2	0.714	0.701	96883762	15.353	peri shape
6	2	0.159	0.122	284550132	126.448	shape perm
7	3	0.780	0.765	74326644	4.000	peri shape perm

Forward: i) Se ajusta el modelo $Y = \beta_0 + \varepsilon$ y se busca el modelo de una covariable tal que el SSE de ese modelo sea mínimo

La candidata a entrar es X_1

Se ajusta el modelo $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ y se evalúa el siguiente contraste de hipótesis $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$

$$F_0 = \frac{SSE(MR) - SSE(MF)}{MSE(MF)} = \frac{338543101 - 109513013}{109513013/46} \simeq 96.2021$$

$$F_{0.95, 1, 48-2} = 4.0517$$

Como $F_0 > F_{0.95,1,48-2}$ X_1 entra al modelo

ii) Lo que sigue es encontrar el modelo con 2 variables cuyo SSE sea mínimo e incluya a X_1

La candidata a entrar es X_2

Ahora $MF: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ y $MR: Y = \beta_0 + \beta_1 X_1 + \varepsilon$

$$F_0 = \frac{SSE(MR) - SSE(MF)}{MSE(MF)} \quad \frac{109513013 - 76348142}{76348142/45} \approx 19.54755$$

$$F_{0.95,1,45} = 4.056612$$

Como $F_0 > F_{0.95,1,45}$ X_2 entra

Backward: i) Se ajusta el modelo con todas las covariables $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

$$F_0 = \frac{SSE(MR) - SSE(MF)}{MSE(MF)}$$

Se busca el modelo con 2 covariables que tenga menor SSE

La candidata a salir es X_3 y el modelo reducido es
MR: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

$$F_0 = \frac{76348142 - 74326644}{74326644/44} = 1.19 \qquad \begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases}$$

$$F_{0.95,1,44} = 4.06$$

Como $F_0 < F_{0.95,1,44}$ Sale X_3 del modelo

ii) Se busca el modelo de una covariable que contenga a X_1 o X_2 (Se elimina una de las dos) que tenga menor SSE
la candidata a salir es X_2

MR: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$, MF: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

$$F_0 = \frac{109513013 - 76348142}{76348142/45} \simeq 19.54 \qquad \begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$$

$F_{0.95,1,45} = 4.056612$ Como $F_0 > F_{0.95,1,45}$ Se acaba el backward

Stepwise: i) Se hace la primera iteración como en forward

$Y = \beta_0 + \beta_1 X_1 + \varepsilon$ (quedamos con este modelo)

ii) Se hace como si fuera forward y quedamos con

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

OJO: Cuando se agrega una covariable al modelo hay que verificar que los demás β's sean marginal/ significativos

En este caso hay que ver $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$

$MR: Y = \beta_2 X_2 + \varepsilon$, $MF: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

$$F_0 = \frac{285283187 - 76348142}{76348142 / 45} \approx 123.15$$

$F_{0.95,1,45} = 4.056612$ $\quad F_0 > F_{0.95,1,45}$ y X_1 se queda

iii) la candidata a entrar es X_3

$MF: Y = \beta_0 + \beta_1 X_1 \cdots + \beta_3 X_3 + \varepsilon$ $\quad MR: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

X_3 no entra y se acaba el stepwise