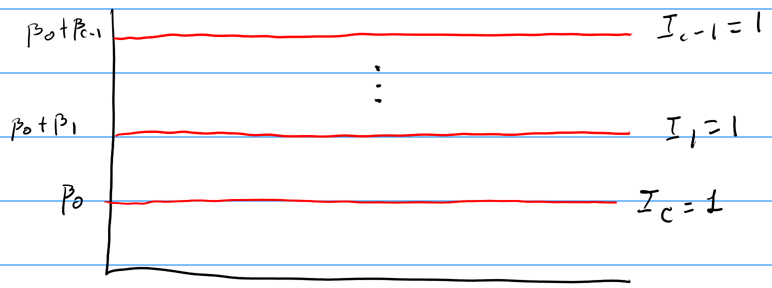
1. Responda las siguientes preguntas.

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 $Y = \beta_0 + \sum_{k=1}^{c-1} \beta_k I_k + \varepsilon$; $\varepsilon \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, en caso de que no, ¿cuál es la media?
d) La interacción entre variables numéricas y categóricas hace variar la tasa decambio de la respuesta en cada categoría de la variable categórica.

a) No, considere el modelo

$$Y = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \dots + \beta_{c-1} I_{c-1} + \varepsilon; \ \varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$



Al no haber interacción, sólo se traslada la recta de regresión pues se ve modificado su intercepto. Como no se modifica su pendiente, se generan rectas paralelas, no secantes.

b) Sí, es el caso particular del literal anterior.

c) No. Suponga $I_j = 1$, luego:

$$Y = \beta_0 + \beta_j + \varepsilon; \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$

Similarmante para cualquier $j = 1, 2, \dots, k$.
Tomando la media en ambos lados queda:

$$E[Y] = E[\beta_0 + \beta_j + \varepsilon] = \beta_0 + \beta_j$$

Por tanto, la media de Y en la categoría j es $\beta_0 + \beta_j \neq \beta_j$

d) Considere el modelo Y vs X con indicadores e interacciones:

$$Y = \beta_0 + \beta_1 X + \beta_2 I_1 + \beta_3 I_2 + \dots + \beta_c I_{c-1} + \beta_{1,1} X I_1 + \beta_{1,2} X I_2$$
$$+ \beta_{1,c-1} X I_{c-1} + \varepsilon; \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$

Si $I_j = 1$

$$\therefore Y = \beta_0 + \beta_1 X + \beta_{j+1} + \beta_{1,j} X + \varepsilon; \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$
$$= (\beta_0 + \beta_{j+1}) + \underbrace{(\beta_1 + \beta_{1,j})}_{\text{Tasa de cambio}} X + \varepsilon$$

Como la tasa de cambio $\beta_1 + \beta_{1,j}$ depende de la categoría j, la afirmación del enunciado es correcta.

2) El modelo a ajustar es:

$$mpg = \beta_0 + \beta_1 wt + \beta_2 cyl + \beta_3 cyl * wt + \varepsilon; \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$

Donde cyl es una var. categórica con los niveles 4, 6, 8. De modo tal que el modelo al considerar interacciones se convierte en:

$$mpg = \beta_0 + \beta_1 wt + \beta_2 cyl_6 + \beta_3 cyl_8 + \beta_{1,6} wt * cyl_6$$
$$+ \beta_{1,8} wt * cyl_8 + \varepsilon; \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2)$$

Nivel de referencia cyl_4.

```
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
```

$\therefore \hat{Y} = 39,571 - 5,697wt - 11,162\,cyl_6 - 15,703\,cyl_8 + 2,867\,wt*cyl_6 + 3,955\,wt*cyl_8$

Si $cyl = 6$, luego $\hat{Y} = (39,571 - 11,162) - (5,697 - 2,867)\,wt$

$= 28,909 - 2,78wt$

la media de $mpg = 28,909$ cuando $wt = 0$ en la categoría $cyl = 6$ y su cambio por unidad de aumento en wt decrece en $2,78$ en la misma categoría.

Analogamente en el taller se hace para $cyl = 4$ y $cyl = 8$.

3) Para mayor facilidad, se definen:

$X_1 = peri,\; X_2 = perm,\; X_3 = shape$

```
> myAllRegTable(lm(area ~ ., data = rock))
  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

1) Ajustar el modelo $Y = \beta_0 + \varepsilon;\; \varepsilon \overset{iid}{\sim} N(0, \sigma^2) \rightarrow MR$.

Identificar modelo de una covariable cuyo SSE sea mínimo, así la candidata a entrar es X_1

Por lo tanto se ajusta $Y = \beta_0 + \beta_1 X_1 + \varepsilon;\; \varepsilon \overset{iid}{\sim} N(0, \sigma^2) \rightarrow MF$

y se evalúa su significancia.

$$\begin{cases} H_0: \beta_1 = 0 \\ H_a: \beta_1 \neq 0 \end{cases} \qquad F_0 = \frac{SSE(MR) - SSE(MF)}{MSE(MF)} \overset{H_0}{\sim} F_{1,48-2}$$

$$F_0 = \frac{338543101 - 109513013}{\frac{109513013}{48-2}} = 96{,}2021$$

$f_{0,05,1,46} = 4{,}0517$

como $F_0 > F_{\alpha,1,46}$ X_1 entra al modelo.

Pasos siguientes analogamente se hacen en el taller.

Backward

i) Se parte del ajuste con todas las covariables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon; \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2) \rightarrow MF$$

Se ajusta el modelo con 2 predictoras con mínimo SSE el cual es el ajustado con X_1 y X_2, por lo que la candidata a salir es X_3. El modelo es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon; \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2) \rightarrow MR$$

Se quiere probar

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

$$F_0 = \frac{SSE(MR) - SSE(MF)}{MSE(MF)} \overset{H_0}{\sim} F_{1,48-4}$$

$$F_0 = \frac{7639842 - 7432644}{\frac{7432644}{48-4}} = 1{,}19$$

$f_{0,05,1,44} = 4,06$

Como $F_0 < f_{\alpha,1,44}$, x_3 sale del modelo y el algoritmo continúa.

Su continuación en el taller.

Stepwise

i) Primer paso es el mismo que en forward, por lo que:

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

$$MF: Y = \beta_0 + \beta_1 X_1 + \varepsilon \, ;$$

Entró X_1. Como no hay más variables se continúa a la siguiente iteración.

ii) Variable a entrar X_2. En taller, en la segunda iteración de forward efectivamente X_2 entra, por lo que lo siguiente en stepwise es verificar significancia individual de las predictoras que ya estaban en el modelo, en este caso X_1:

se quiere probar

$$\begin{cases} H_0: \beta_1 = 0 \\ H_a: \beta_1 \neq 0 \end{cases} \qquad F_0 = \frac{SSE(MR) - SSE(MF)}{MSE(MF)} \overset{H_0}{\sim} f_{1,48-3}$$

luego:

$$MR: Y = \beta_0 + \beta_2 X_2 + \varepsilon$$

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

Así, $F_0 = \dfrac{285283187 - 76398192}{\dfrac{76398192}{95}} = 123.15$

$f_{0,05,1,95} = 4,056612$

Como $F_0 > f_{\alpha,1,95}$, X_1 se queda en el modelo.

continuación taller.