1. Responda las siguientes preguntas.

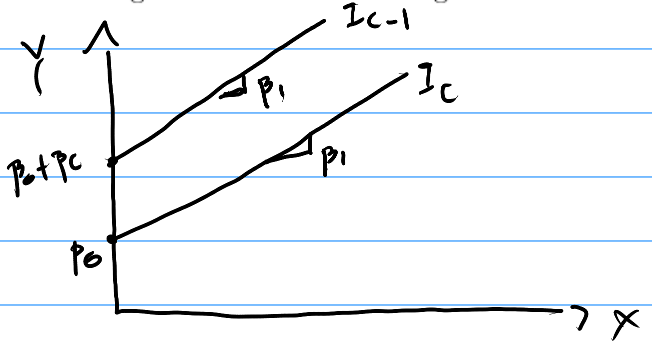
a) Suponga que se ajusta un modelo de regresión con una variable categórica, sin interacción, ¿dicho modelo genera c rectas secantes?

b) En un modelo de regresión lineal simple ajustado solo con variables cualitativas, 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? F

d) ¿La interacción entre variable numérica y las indicadoras asociadas a una variable categórica hace variar la tasa de cambio de la respuesta dada por la predictora numérica en cada categoría de la variable categórica?

a)



$$Y = \beta_0 + \beta_1 X + \beta_2 I_1 + \beta_3 I_2 + \dots + \beta_c I_{c-1} + \varepsilon$$

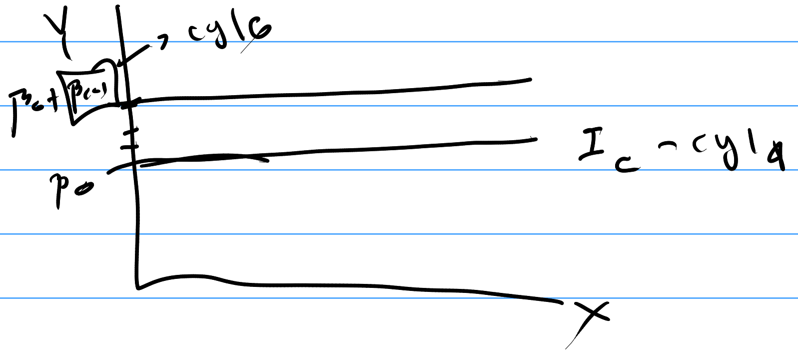
$$Y = \beta_0 + \beta_1 X$$

$I_{c-1} = 1$?

$$Y = \beta_0 + \beta_1 X + \beta_c + \varepsilon$$
$$= (\beta_0 + \beta_c) + \beta_1 X + \varepsilon$$

b)

$$Y = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \beta_3 I_3 + \dots + \beta_{c-1} I_{c-1} + \varepsilon$$



¿$E[Y] = \beta_j$?

$Y = \beta_0 \leftarrow$

$I_{c-1} = 1$

$Y = \beta_0 + \beta_{c-1}$

β_0

$\beta_0 + \beta_1$

$\beta_0 + \beta_2$

$\boxed{\beta_0 + \beta_j}$

$I_j \quad j = 1, \dots, K$

Y vs $(X \wedge S)$

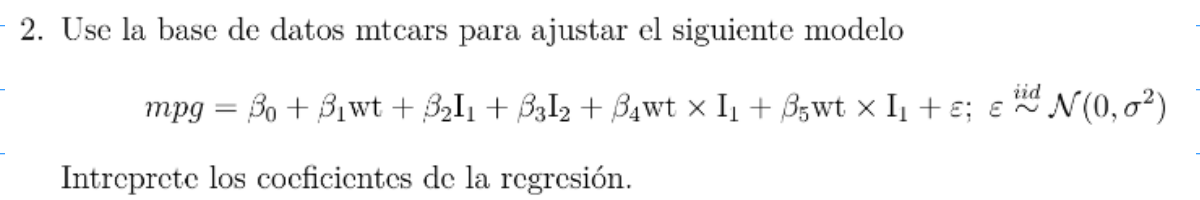
d) $Y = \beta_0 + \beta_1 X + \beta_2 I_1 + \beta_3 I_2 + \beta_{1,1} I_1 X + \beta_{1,2} I_2 X$

I_3

↳ $Y = \beta_0 + \underbrace{\beta_1}_{\beta_1} X$

$I_1 = 1$

$Y = (\beta_0 + \beta_2) + \underbrace{(\beta_1 + \beta_{1,1})} X$

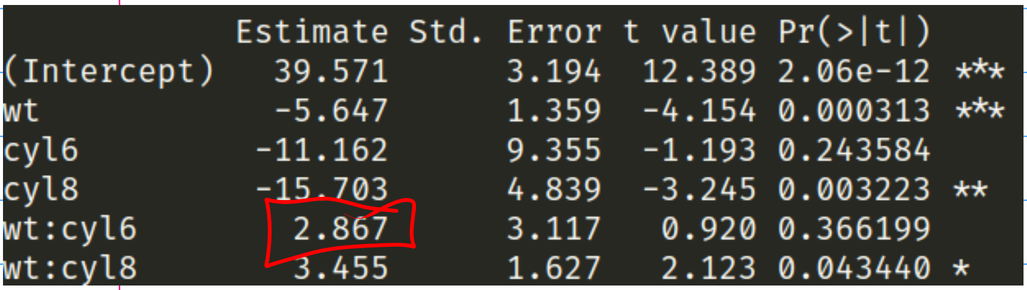


2. Use la base de datos mtcars para ajustar el siguiente modelo

$$mpg = \beta_0 + \beta_1 \mathrm{wt} + \beta_2 \mathrm{I}_1 + \beta_3 \mathrm{I}_2 + \beta_4 \mathrm{wt} \times \mathrm{I}_1 + \beta_5 \mathrm{wt} \times \mathrm{I}_1 + \varepsilon;\ \varepsilon \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

Intreprete los coeficientes de la regresión.

mpg vs $(wt \wedge cyl)$



```
            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 *
```

$Y = 39{,}571 - 5{,}647wt - 11{,} \ldots$

$cyl6 = 1$

$Y = 39{,}571 - 5{,}647wt - 11{,}162 + 2{,}867 \cdot wt$

$mpg = (39{,}571 - 11{,}162) + (2{,}867 - 5{,}647)wt$

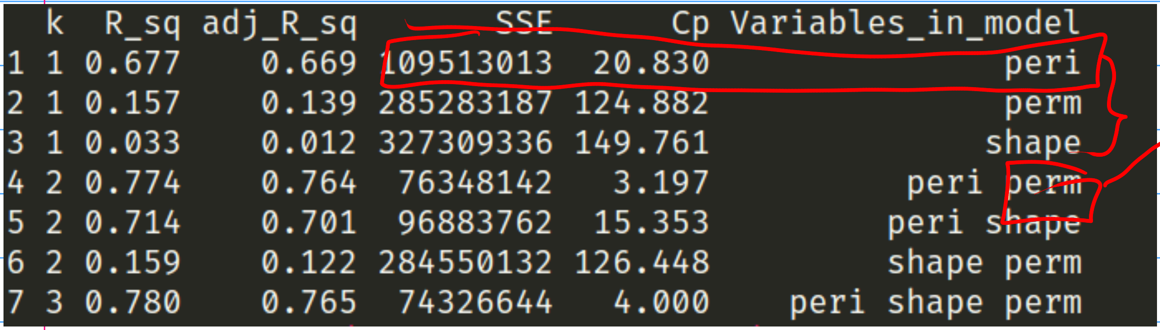
$wt = 0$

$cyl4 = 1$

$mpg = 39{,}571 - 5{,}647wt$

3. Usando la base de datos rock lleve a cabo los métodos de selección forward, backward y stepwise. Concluya cual de los tres modelos obtenidos es el mejor. Para dicha tarea se puede apoyar en la tabla de todas las regresiones posibles.

Método Forward. } Se busca rechazar para seguir



```
  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
```

$> X_2$

$x_1 = \text{peri}$

$x_2 = \text{perm}$

$x_3 = \text{shape}$

1 var

Partimos de $MR: Y = \beta_0 + \varepsilon$

Candidata X_1 $MF: Y = \beta_0 + \beta_1 X_1 + \varepsilon$

$$\begin{cases} H_0: \beta_1 = 0 \\ H_a: \beta_1 \neq 0 \end{cases}$$

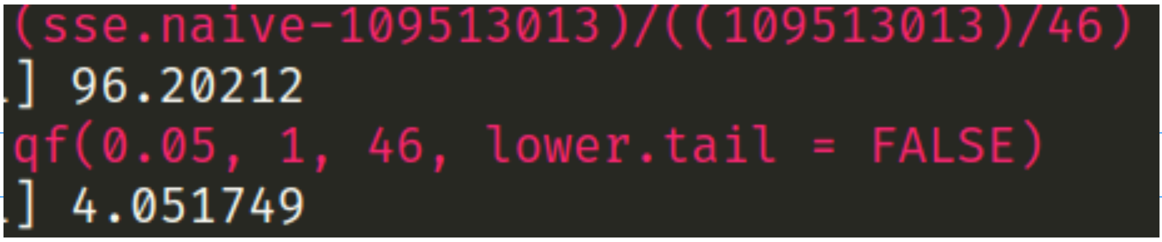
$$F_0 = \frac{\frac{SSE(MR) - SSE(MF)}{1}}{\frac{SSE(MF)}{n-p}} = \frac{\frac{SSE(MR) - SSE(MF)}{1}}{\frac{SSE(MF)}{46}}$$

$$F_0 = \frac{338543101 - 109513013}{\frac{109513013}{46}} =$$

Si $F_0 > f_{\alpha,1,46}$ entonces rechazo H_0

$\therefore$ β_1 es signif.

$\therefore$ X_1 entra



```
(sse.naive-109513013)/((109513013)/46)
] 96.20212
qf(0.05, 1, 46, lower.tail = FALSE)
] 4.051749
```

Fo

F

$F_0 > F_{\alpha,1,46}$ rechaza H_0

$\therefore \beta_1$

$\therefore x_1$ entra

Partimos de $MR: Y = \beta_0 + \beta_1 X_1 + \varepsilon$

Candidata x_2 $MF: Y = \beta_0 + \beta_1 X + \beta_2 X_2 + \varepsilon$

$$\begin{cases} H_0: \beta_2 = 0 \\ H_a: \beta_2 \neq 0 \end{cases}$$

$$F_0 = \frac{SSE(MR) - SSE(MF)}{\frac{SSE(MF)}{45}}$$

$F_0 > F$ rechaza H_0

$\therefore \beta_2$ signif.

$\therefore x_2$ entra

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

Candidata x_3 $MF: Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

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

$$F_0 = \frac{SSE(MR) - SSE(MF)}{\frac{SSE(MF)}{44}} \sim F_{1,44}$$

$F_0 < F_{\alpha,1,44} \Rightarrow$ No rechazo

$\therefore \beta_3$ no es sign.

$\therefore x_3$ no entra

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

Método Backward → Busca no rechazar para seguir

partimos de MF: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

candidata X_3 MR: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

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

$$F_0 = \frac{SSE(MR) - SSE(MF)}{\frac{SSE(MF)}{44}} \sim F_{1,\,44}$$

$F_0 < F_{\alpha,1,44}$ ⇒ No rechazo

$\therefore \beta_3$ no es signif.

$\therefore X_3$ sale

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

Candidata X_2 MR: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

$$F_0 = \frac{SSE(MR) - SSE(MF)}{\frac{SSE(MF)}{45}} \sim F_{1,\,45}$$

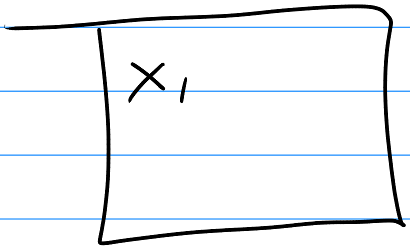
$F_0 > F_{\alpha,1,45}$ ⇒ se rechaza

$\therefore \beta_2$ signif.

$\therefore X_2$ se queda

M. final $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

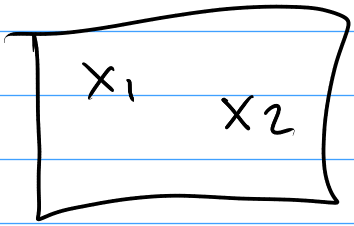
Método Stepwise



Paso 1

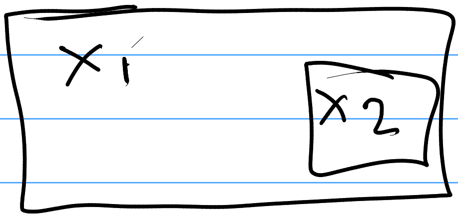
x_1 entró

Paso 2



x_2 entró

Paso 3



MF: $Y=\beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$

MR: $Y=\beta_0+\beta_2x_2+\varepsilon$

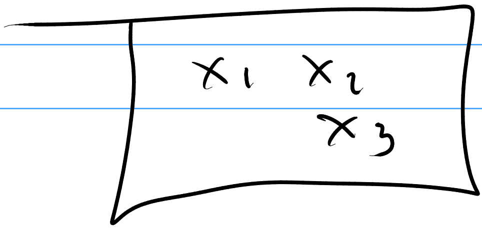
$$\begin{cases} H_0: \beta_1 = 0 \\ H_a: \beta_1 \neq 0 \end{cases}$$

$$F_0 = \frac{SSE(MR) - SSE(MF)}{\frac{SSE(MF)}{45}} \sim F_{1,45}$$

X_1 se quedó

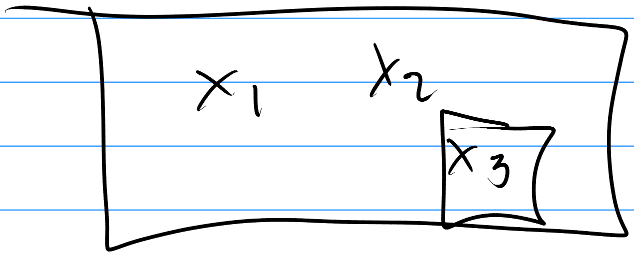
suposición

Paso 4



X_3 entra

Paso 5



¿ x_1 o x_2 sale?