Quinto taller

jueves, 29 de mayo de 2025 2:09 p. m.

- (b) En el modelo de regresión lineal simple de la clase $Y_i = \beta_0 + \beta_1 I_{i1} + \beta_2 I_{i2} + \ldots + \beta_{i,c-1} I_{i,c-1} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ las rectas generadas son horizontales.
- (c) En el modelo de regresión lineal simple de la clase $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i2} + \ldots + \beta_{ic} I_{i,c} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ las rectas generadas cambian su pendiente.
- (d) En el modelo de regresión lineal con interacción $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i2} + \ldots + \beta_{ic} I_{i,c} + \beta_{c+1} X_{i1} I_{i2} + \beta_{c+2} X_{i1} I_{i3} + \ldots + \beta_{2c} X_{i1} I_{i,c} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ las rectas generadas cambian su intercepto y pendiente.

2. Considere el modelo de regresión lineal dado por la interacción $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \beta_3 I_{i2} + \beta_4 X_{i1} I_{i1} + \beta_5 X_{i1} I_{i2} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, empleado para evaluar el cambio de precio en automóviles en función de *'Bore' (numérica)* y *'BodyStyle' (categórica)*, esta última con tres categorías *'Sedan'*, *'Wagon'* y *'Hatchback'*. **Se toma como referencia la categoría 'Hatchback'. Responda:**

Modelos con variables indicadoras

Variable indicadora: $\mathbb{I}_{(A)}(i) = \begin{cases} 1, & i \in A \\ 0, & i \notin A \end{cases}$

Variable categórica: W = BodyStyle

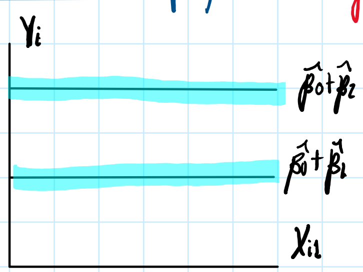
$W = \begin{cases} \text{Sedan} & \mathbb{I}_{i1} \\ \text{Wagon} & \mathbb{I}_{i2} \\ \text{Hatchback} & \mathbb{I}_{i3} \end{cases}$; Y_i = Precio

- Construcción de modelos:

(1). Modelo sencillo: $Y_i = \beta_0 + \beta_1 \mathbb{I}_{i1} + \beta_2 \mathbb{I}_{i2} + \ldots + \beta_c \mathbb{I}_{ic} + \varepsilon_i \; ; \; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

- $Y_i = \beta_0 + \beta_1 \mathbb{I}_{i2} + \beta_2 \mathbb{I}_{i3} + \varepsilon_i$; Asumir una categoría referencia: $\mathbb{I}_{i1}$: Categoría referencia (Sedan)
- Hallar recta específica: Wagon: $Y_i = \beta_0 + \beta_1 + \varepsilon_i$; Hatchback: $Y_i = \beta_0 + \beta_2 + \varepsilon_i$

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1$ $\qquad$ $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_2$



(b). Es verdadero ; C = # Categorías ; C = 3

(2). Modelo con variable continua sin interacciones: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i2} + \beta_3 \mathbb{I}_{i3} + \ldots + \beta_c \mathbb{I}_{ic} + \varepsilon_i \; ; \; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

(Tomando a la primera categoría $\mathbb{I}_{i1}$ como referencia)

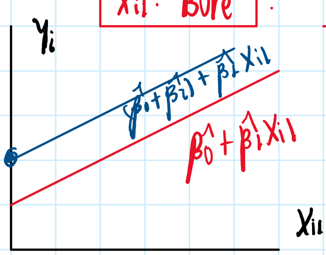
$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i1} + \beta_3 \mathbb{I}_{i2} + \ldots + \beta_c \mathbb{I}_{i,c-1} + \varepsilon_i$

Tomando a la última categoría $\mathbb{I}_{ic}$ como ref.

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

X_{i1}: Bore ; Categoría referencia: Sedan ($\mathbb{I}_{i1}$)

(c). Falsa



- Recta wagon: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 + \varepsilon_i = (\beta_0 + \beta_2) + \beta_1 X_{i1} + \varepsilon_i$

$\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 X_{i1}$

- Recta Sedan: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$; $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}$

- En este modelo solo cambia el intercepto

(3). Modelo con variable continua y con interacciones:

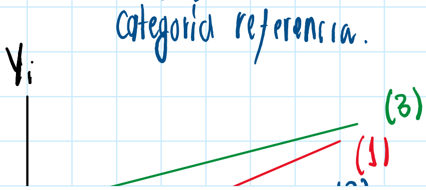
$$Y_i = \underbrace{[\beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i2} + \ldots + \beta_c \mathbb{I}_{ic}]}_{\text{Parte 1: Modelo (2)}} + \underbrace{[\beta_{c+1} X_{i1} \mathbb{I}_{i2} + \ldots + \beta_{2c} X_{i1} \mathbb{I}_{ic}]}_{\text{Parte 2: Interacción}} + \varepsilon_i$$

$\Downarrow$ Tomando $\mathbb{I}_{i1}$ como categoría referencia.

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i2} + \beta_3 \mathbb{I}_{i3} + \beta_4 X_{i1} \mathbb{I}_{i2} + \beta_5 X_{i1} \mathbb{I}_{i3} + \varepsilon_i$

- Hallar las rectas específicas para cada categoría:

Hatchback: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}$ (1)

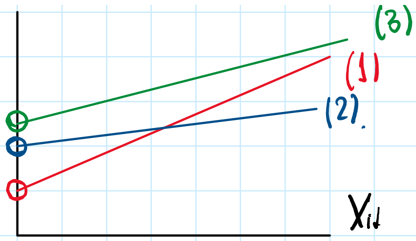


Hatchback: $\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \; : \; \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}}$ (1)

Sedan: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 + \beta_4 X_{i1} + \varepsilon_i$

$\boxed{Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_4) X_{i1} + \varepsilon_i}$ (2)

Recta ajustada: $\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_4) X_{i1}$



Wagon: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_3 + \beta_5 X_{i1} + \varepsilon_i$

$\boxed{Y_i = (\beta_0 + \beta_3) + (\beta_1 + \beta_5) X_{i1} + \varepsilon_i}$ (3)

Recta ajustada: $\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_3) + (\hat{\beta}_1 + \hat{\beta}_5) X_{i1}$

(d). Verdadera

- En este caso, las rectas cambian su intercepto y su pendiente.

- (a) Caracterize la información en análisis, reportando la ecuación de las rectas en función de los diferentes niveles, así como las rectas ajustadas.

Modelo clase (3)

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i2} + \beta_3 \mathbb{I}_{i3} + \beta_4 X_{i1} \mathbb{I}_{i2} + \beta_5 X_{i1} \mathbb{I}_{i3} + \varepsilon_i$: lm(price ~ BodyStyle * Bore) (reflejar interacción)

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          -15274.7     8370.8  -1.825  0.06972 .
bodyStylesedan       -34195.3    11716.2  -2.919  0.00397 **
bodyStylewagon        -1757.3    17855.0  -0.098  0.92171
bore                   7752.2     2581.6   3.003  0.00306 **
bodyStylesedan:bore   11359.2     3555.0   3.195  0.00165 **
bodyStylewagon:bore     879.8     5290.0   0.166  0.86810
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5818 on 177 degrees of freedom
Multiple R-squared:  0.3574,	Adjusted R-squared:  0.3393
F-statistic: 19.69 on 5 and 177 DF,  p-value: 1.411e-15
```

$\hat{\beta}_0$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_1$, $\hat{\beta}_4$, $\hat{\beta}_5$

- Categoría referencia Hatchback:

$$W = \begin{cases} \text{Sedan} & \mathbb{I}_{i2} \\ \text{Wagon} & \mathbb{I}_{i3} \\ \text{Hatchback} & \mathbb{I}_{i1} \end{cases};$$

- Recta ajustada Wagon: $\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_4) X_{i1}$

$\hat{Y}_i = (-15274{,}7 - 34195{,}3) + (7752{,}2 + 11359{,}2) X_{i1}$

- (b) Interprete la estimación del parámetro β_4 en términos del problema.

$\hat{\beta}_4$: BodyStyle sedan:bore : Interacción entre X_{i1} y $\mathbb{I}_{i2}$ (Sedan) :

- Por un cambio unitario en Bore (variable continua), para los automóviles de la clase sedan, el precio promedio cambia en 11359,2 unidades en relación a la categoría de referencia (Hatchback)

- (c) Determine si existe diferencia entre las ordenadas en el origen de las rectas correspondientes a los carros con *'BodyStyle'* de la clase *'Hatchback' y 'Sedan'*. Escriba la hipótesis correspondiente y el criterio de decisión.

Hatchback: $\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \; : \; \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}}$ (1)

Sedan: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 + \beta_4 X_{i1} + \varepsilon_i$

$\boxed{Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_4) X_{i1} + \varepsilon_i}$ (2)

Recta ajustada: $\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_4) X_{i1}$

$$\begin{cases} H_0: \beta_0 = \beta_0 + \beta_2 \rightarrow \beta_0 - \beta_0 = \beta_2 \rightarrow \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$$

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

```
> linearHypothesis(modelo, "bodyStylesedan=0")

Linear hypothesis test:
bodyStylesedan = 0

Model 1: restricted model
Model 2: price ~ bodyStyle * bore

  Res.Df        RSS Df Sum of Sq      F   Pr(>F)
1    178 6280574850
2    177 5992191967  1 288382883 8.5184 0.003972 **
---
```

Pval < α = 0,05

- A un nivel de significancia del 5%, se puede establecer que el efecto de la categoría Sedan es significativo (sobre la ordenada)

Ejemplo importante: Método de todas las regresiones posibles

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i2} + \beta_3 \mathbb{I}_{i3} + \beta_4 X_{i1} \mathbb{I}_{i2} + \beta_5 X_{i1} \mathbb{I}_{i3} + \varepsilon_i; \; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$: Modelo completo

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i2} + \beta_3 I_{i3} + \beta_4 X_{i1} I_{i2} + \beta_5 X_{i1} I_{i3} + \varepsilon_i;\quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$: Modelo completo (MF)

tabla todas regresiones posibles

	k	R_sq	adj_R_sq	SSE	Cp
1	1	0.270	0.266	6808439489	22.111
2	1	0.106	0.101	8339332596	67.331
3	1	0.077	0.072	8607903816	75.264
4	1	0.000	-0.005	9324673834	96.436
5	1	0.000	-0.006	9325424921	96.459
6	2	0.323	0.315	6313303524	9.485
7	2	0.316	0.309	6375515576	11.323
8	2	0.299	0.291	6537854051	16.118
9	2	0.278	0.270	6737360745	22.011
10	2	0.277	0.269	6746166218	22.271
11	2	0.127	0.117	8145559383	63.607
12	2	0.124	0.114	8169682089	64.320
13	2	0.092	0.082	8465035217	73.044
14	2	0.090	0.080	8485820856	73.658
15	2	0.013	0.002	9206352559	94.941
16	3	0.355	0.344	6017124167	2.736
17	3	0.324	0.312	6308226252	11.335
18	3	0.323	0.312	6310049522	11.389
19	3	0.316	0.305	6373969292	13.277
20	3	0.316	0.305	6374972350	13.307
21	3	0.314	0.303	6394985452	13.898
22	3	0.312	0.300	6415771091	14.512
23	3	0.283	0.271	6690547232	22.628
24	3	0.137	0.122	8051360814	62.825
25	3	0.103	0.088	8367499581	72.163
26	4	0.357	0.343	5992519891	4.010
27	4	0.357	0.343	5993128379	4.028
28	4	0.327	0.311	6280574850	12.518
29	4	0.325	0.310	6297449816	13.017
30	4	0.320	0.305	6337833534	14.210
31	5	0.357	0.339	5992191967	6.000

	Variables_in_model
1	bore
2	bore:bodyStylesedan
3	bodyStylesedan
4	bodyStylewagon
5	bore:bodyStylewagon
6	bore bore:bodyStylesedan
7	bore bodyStylesedan
8	bodyStylesedan bore:bodyStylesedan
9	bore bore:bodyStylewagon
10	bore bodyStylewagon
11	bore:bodyStylesedan bore:bodyStylewagon
12	bodyStylewagon bore:bodyStylesedan
13	bodyStylesedan bore:bodyStylewagon
14	bodyStylesedan bodyStylewagon
15	bodyStylewagon bore:bodyStylewagon
16	bore bodyStylesedan bore:bodyStylesedan
17	bore bodyStylewagon bore:bodyStylesedan
18	bore bore:bodyStylesedan bore:bodyStylewagon
19	bore bodyStylesedan bodyStylewagon
20	bore bodyStylesedan bore:bodyStylewagon
21	bodyStylesedan bore:bodyStylesedan bore:bodyStylewagon
22	bodyStylesedan bodyStylewagon bore:bodyStylesedan
23	bore bodyStylewagon bore:bodyStylewagon
24	bodyStylewagon bore:bodyStylesedan bore:bodyStylewagon
25	bodyStylesedan bodyStylewagon bore:bodyStylewagon
26	bore bodyStylesedan bore:bodyStylesedan bore:bodyStylewagon
27	bore bodyStylesedan bodyStylewagon bore:bodyStylesedan
28	bore bodyStylewagon bore:bodyStylesedan bore:bodyStylewagon
29	bodyStylesedan bodyStylewagon bore:bodyStylesedan bore:bodyStylewagon
30	bore bodyStylesedan bodyStylewagon bore:bodyStylewagon
31	bore bodyStylesedan bodyStylewagon bore:bodyStylesedan bore:bodyStylewagon

- $H_0: \beta_2 = 0$: $H_1: \beta_2 \neq 0$

- Modelo reducido: Reemplazar H_0 en MF: MR.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_3 I_{i3} + \beta_4 X_{i1} I_{i2} + \beta_5 X_{i1} I_{i3} + \varepsilon_i$$

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

- Bore (X_{i1}); Sedan (I_{i2}); Wagon (I_{i3}); Bore: Sedan ($X_{i1} \cdot I_{i2}$); Bore: Wagon ($X_{i1} \cdot I_{i3}$)
- Bore (X_{i1}); Wagon (I_{i3}); Bore: Sedan ($X_{i1} \cdot I_{i2}$); Bore: Wagon ($X_{i1} \cdot I_{i3}$)

$$F = \frac{[6280574850 - 5992191967]/1}{5992191967/176}$$

: (1). 177 gl
(2). 176 gl

- (d) Determine si el cambio promedio *'Price'* por unidad de cambio en *'Bore'* es igual para los niveles de *'BodyStyle'* de la clase *'Sedan'* y *'Wagon'*. Escriba la hipótesis correspondiente y el criterio de decisión.

- Por una pendiente:

Sedan: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 + \beta_4 X_{i1} + \varepsilon_i$

$\boxed{Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_4) X_{i1} + \varepsilon_i}$ (2)

Recta ajustada: $\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_4) X_{i1}$

Wagon: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_3 + \beta_5 X_{i1} + \varepsilon_i$

$\boxed{Y_i = (\beta_0 + \beta_3) + (\beta_1 + \beta_5) X_{i1} + \varepsilon_i}$ (3)

Recta ajustada: $\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_3) + (\hat{\beta}_1 + \hat{\beta}_5) X_{i1}$

$H_0: \beta_1 + \beta_4 = \beta_1 + \beta_5 \rightarrow \beta_4 = \beta_5 : \beta_4 - \beta_5 = 0$

$H_1: \beta_4 \neq \beta_5$

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

- Criterio decisión: $P(f_{v,n-p} > F_{cal})$

```
> # realizar prueba sobre pendiente
> linearHypothesis(modelo, "bodyStylesedan:bore=bodyStylewagon:bo
re")

Linear hypothesis test:
bodyStylesedan:bore - bodyStylewagon:bore = 0

Model 1: restricted model
Model 2: price ~ bodyStyle * bore

  Res.Df        RSS Df Sum of Sq      F  Pr(>F)
1    178 6128414009
2    177 5992191967  1 136222042 4.0238 0.04639 *
---
```

$< \alpha$

- A un nivel de significancia del 5%, el efecto de la interacción entre Bore y Sedan es significativamente distinto que el efecto de la interacción entre Bore y Wagon.

: el efecto sobre Bore depende de la referencia del automóvil.

- (a) El método *'Forward'* parte del modelo con todas las variables, eliminando secuencialmente de a una variable con el propósito de reducir la suma de cuadrados del error asociado. El método *'Backward'* sigue la lógica contraria.

Falso: al revés

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

1. Use el método de selección *'Backward'* para determinar el mejor modelo (paso por paso) y compárelo con el resultado dado por el método *'Forward'* y *'Stepwise'*. **Use el conjunto de datos del taller anterior.**

```
> myBackward(datos2[, -5])
─────────────STEP  1 ─────────────
 The drop statistics :
```

1. Use el método de selección *'Backward'* para determinar el mejor modelo (paso por paso) y compárelo con el resultado dado por el método *'Forward'* y *'Stepwise'*. **Use el conjunto de datos del taller anterior.**

Calificacion	Estudio	Excelencia	Sueno	Tipo estudiante
75.78204	5.284089	2.642755	5.013464	Relajado
79.08557	4.867878	2.427432	4.551853	Relajado
74.94530	5.660917	2.821258	5.353560	Intermedio
93.44042	5.955115	2.970340	9.269884	Intermedio

Y_i X_{i1} X_{i2} X_{i3}

Variable menos significativa: Valor P más alto

```
> myBackward(datos2[, -5])
-------------STEP  1 ------------
 The drop statistics :
Single term deletions

Model:
Calificacion ~ Estudio + Excelencia + Sueno
           Df Sum of Sq     RSS    AIC  F value Pr(>F)
<none>                   2299.8 321.54
Estudio     1      12.6  2312.3 320.08   0.5251 0.4705
Excelencia  1      15.3  2315.1 320.20   0.6391 0.4260
Sueno       1   12529.9 14829.6 505.92 523.0416 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

_______
Term dropped in step 1 : Estudio
```

- Paso (1): $Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$

Eliminar excelencia (X_{i2})

```
-------------STEP  2 ------------
 The drop statistics :
Single term deletions

Model:
Calificacion ~ Excelencia + Sueno
           Df Sum of Sq     RSS    AIC  F value Pr(>F)
<none>                   2312.3 320.08
Excelencia  1       4.2  2316.5 318.26   0.1748 0.6768
Sueno       1   12564.2 14876.5 504.24 527.0561 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

_______
Term dropped in step 2 : Excelencia
```

- Paso (2): $Y_i = \beta_0 + \beta_3 X_{i3} + \varepsilon_i$

```
-------------STEP  3 ------------
 The drop statistics :
Single term deletions

Model:
Calificacion ~ Sueno
       Df Sum of Sq     RSS    AIC F value    Pr(>F)
<none>               2316.5 318.26
Sueno   1     18427 20743.3 535.48  779.55 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:
lm(formula = Calificacion ~ Sueno)

Coefficients:
(Intercept)        Sueno
     48.100        5.333
```

Significativo

: $Y_i = \beta_0 + \beta_3 X_{i3} + \varepsilon_i$