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

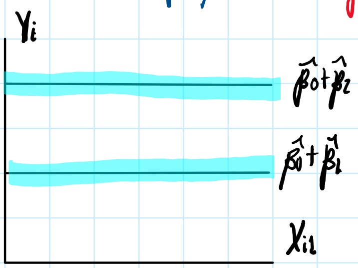
$W = \begin{cases} \text{Sedan} & \mathbb{I}_{i1} \\ \text{Wagon} & \mathbb{I}_{i2} \\ \text{Hatchback} & \mathbb{I}_{i3} \end{cases}$; Y_i = Precio

- 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)

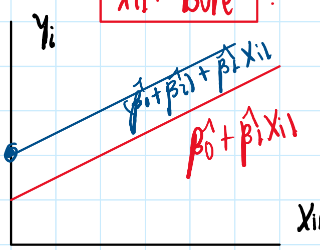
$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i1} + \beta_3 \mathbb{I}_{i2} + \ldots + \beta_c \mathbb{I}_{i,c-1} + \varepsilon_i$

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:

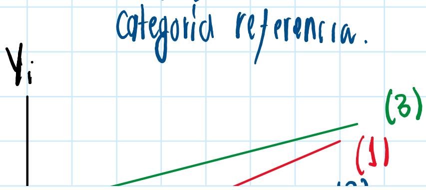
$Y_i = \underbrace{[\beta_0 + \beta_1 X_{i1} + \beta_2 \mathbb{I}_{i2} + \ldots + \beta_c \mathbb{I}_{ic}]}_{\text{Parte 1: Modelo (2)}} + \underbrace{[\beta_{c+1} X_{i1} \mathbb{I}_{i2} + \ldots + \beta_{2c} X_{i1} \mathbb{I}_{ic}]}_{\text{Parte 2: Interacción}} + \varepsilon_i$

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

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

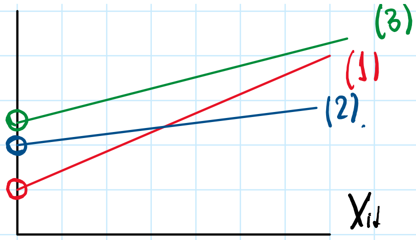
$Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_4) X_{i1} + \varepsilon_i$ (2)

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

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

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

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i2} + \beta_3 I_{i3} + \beta_4 X_{i1} I_{i2} + \beta_5 X_{i1} I_{i3} + \varepsilon_i$: 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:

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

- 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.

$\hat{\beta}_4$: BodStyle sedan:bore : Interacción entre X_{i1} y 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: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}$ (1)

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

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

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

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

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

```
> 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)

- (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:

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

$H_1: \beta_4 \neq \beta_5$

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

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

$H_1: \beta_4 \neq \beta_5$

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

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

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

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

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**.

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

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

Eliminar excelencia (X_2)

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

```
> 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
_______
```

```
------------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
_______
```

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