Quinto Taller

miércoles, 5 de febrero de 2025 2:02 p. m.

1. Determine el valor de verdad de las siguientes afirmaciones.

- (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. Falsa
- (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. Verdadera
- (c) En el modelo de regresión lineal simple de la clase $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \ldots + \beta_{ic} I_{i,c-1} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ ~~las rectas generadas cambian su pendiente~~. Falsa
- (d) En el modelo de regresión lineal con interacción $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \ldots + \beta_{ic} I_{i,c-1} + \beta_{11} X_{i1} I_{i1} + \beta_{12} X_{i1} I_{i2} + \ldots + \beta_{1,c-1} X_{i1} I_{i,c-1} + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ las rectas generadas cambian su intercepto y pendiente. ✓ Verdadero

(1). Modelo más simple:

$$I_A \begin{cases} 1, & \text{si pertenece a } A \\ 0, & \text{en otro caso} \end{cases}$$

W: Categórica: Categorías A, B, C

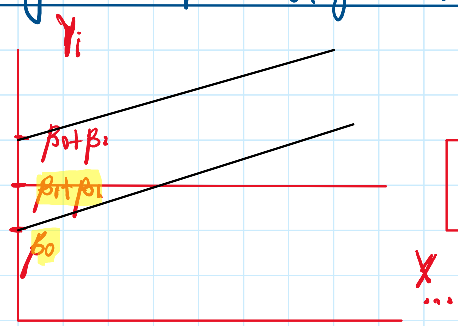
$$Y_i = \beta_0 + \beta_1 I_{i1} + \ldots + \beta_{c-1} I_{i,c-1} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

c: # Categorías

Categoría de referencia I_c

- Determinar la línea generada para categoría 1 de W: Modelo reducido

$$\boxed{Y_i = (\beta_0 + \beta_1) + \varepsilon_i}$$



(2). Modelo con variable continua y categórica sin Int.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \ldots \beta_c I_{i,c-1} + \varepsilon_i$$

- Categoría c (ref)

- Definir recta para cat. 1: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 + \varepsilon_i$

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

- Rectas paralelas (con pendiente)

(3). Modelo con V. continua y W con interacciones

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \ldots + \beta_c I_{i,c-1} + \beta_{1,1} X_{i1} I_{i1} + \beta_{1,2} X_{i1} I_{i1} + \ldots + \beta_{1,c-1} X_{i1} I_{i,c-1} + \varepsilon_i$$

- Definir recta para cat 1: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 + \beta_{1,1} X_{i1} + \varepsilon_i$: $Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_{1,1}) X_{i1} + \varepsilon_i$

Parte práctica:

X_1: Bore W: BodyStyle Y_i: Precio automóvil

factor: Sedan, Otros: wagon, Hatchbaen (ref)

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         -15274.7     8370.8  -1.825  0.06972 .
bore                  7752.2     2581.6   3.003  0.00306 **
bodyStylesedan      -34195.3    11716.2  -2.919  0.00397 **
bodyStylewagon       -1757.3    17855.0  -0.098  0.92171
bore:bodyStylesedan  11359.2     3555.0   3.195  0.00165 **
bore:bodyStylewagon    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
```

β_0, β_1, β_2, β_3, $\beta_{1,1}$, $\beta_{1,2}$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \beta_3 I_{i2} + \beta_{1,1} X_{i1} I_{i1} + \beta_{1,2} X_{i1} I_{i2} + \varepsilon_i$$

$$\hat{Y}_i = -15274.7 + 7752.2 X_{i1} - 34195.3 I_{i1} + \ldots +$$

- Categoría Sedan: $\hat{Y}_i = (-15274.7 - 34195.3) + 7752.2 X_{i1}$ Recta Sedan

$$\hat{Y}_i = (-15274.7 - 34195.3) + 7752.2 \text{ BORE}_i$$

(b) Interpretación β_{11} : $(11359.2) = \hat{\beta}_{11}$: Hatchbach

$Y_i = (-13244{,}4 - 34495{,}5) + 4152{,}2$ Bore

(b). Interpretación β_{11}

$11359{,}2 = \hat{\beta}_{11}$: Hatchback

Por un cambio unitario en la covariable BORE, para los autos con BodyStyle Sedan, el cambio promedio en el precio del automovil es de 11359,2 unidades, en relación a la categoría hatchback (categoría de referencia).

(c). Verificar Intercepto Sedan y Hatchback:

$P = 3$

Sedan: $Y_i = (\beta_0 + \beta_2) + \beta_1 X_{i1} + \varepsilon_i$ ($\beta_2 \to 0$)

Hatchback: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

$p = 2$

$$\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$$

: Suma de cuadrados

$$F = \frac{[SSE(MR) - SSE(MF)]/r}{MSE(MF)}$$

r: gl (SSE(MR)) − gl (SSE(MF))

$P(F_{r,n-p} > F_{calc}) < \alpha$

• $(n-2) - (n-3) = 1 = r$

```
Hypothesis:
bodyStylesedan = 0

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

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

$\alpha = 0{,}05$: Rechazar H_0 — Los interceptos son diferentes.

(d). Analizar precio para Sedan y otros (Wagon)

• Modelos clase (2)

Sedan: $Y_i = (\beta_0 + \beta_2) + \beta_1 X_{i1} + \varepsilon_i$;

$P = 3$

Wagon: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_3 + \beta_{1,2} X_{i1} + \varepsilon_i$

$Y_i = (\beta_0 + \beta_3) + (\beta_1 + \beta_{1,2}) X_{i1} + \varepsilon_i$ ($\beta_{1,2} \to 0$)

$p = 4$

$$\begin{cases} H_0: \beta_{1,2} = 0 \\ H_1: \beta_{1,2} \neq 0 \end{cases}$$

: Suma de cuadrados

$$F = \frac{[SSE(MR) - SSE(MF)]/r}{MSE(MF)}$$

r: gl (SSE(MR)) − gl (SSE(MF))

$P(F_{r,n-p} > F_{calc}) < \alpha$

```
> linearHypothesis(modelo, "bore:bodyStylewagon=0")
Linear hypothesis test

Hypothesis:
bore:bodyStylewagon = 0

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

  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1    178 5993128379
2    177 5992191967  1    936412 0.0277 0.8681
```

α • No rechazo H_0

• Los efectos entre la interacción por categoría con X_1 no es significativo

Backward

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

X_1: Estudio; X_2: Excelencia; X_3: dueño; Y_i = Calificación

X_1: Estudio; X_2: Excelencia; X_3: Sueño; Y_i = Calificación

Backward: $Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in} + \varepsilon_i$ → Modelo reducido : Tabla[..., ...]

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$: -5

```
> 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 32 .4
Estudio     1      12.6  2312.3 3   8   0.5251 0.4705
Excelencia  1      15.3  2315.1 32  0   0.6391 0.4260
Sueno       1   12529.9 14829.6 505.92 523.0416 <2e-16 ***
---
```

Suma parcial: No es sig. No es sig.

```
Term dropped in step 1 : Estudio
```

Suma de cuadrados

$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$

$F = \dfrac{[SSE(MR) - SSE(MF)] / r}{MSE(MF)}$; r = gl(SSE(MR)) - gl(SSE(MF))

$P(f_{r,n-p} > F_{calc}) < \alpha$

- Modelo sin estudio: $Y_i = \beta_0 + \beta_2 X_{i2} + \beta_2 X_{i3} + \varepsilon_i$

```
> myAnova(modelo2)
      Sum_of_Squares DF Mean_Square F_Value     P_value
Model       18443.56  3   6147.8537 256.634 1.052e-45
Error        2299.75 96     23.9557
```

Model completo

```
> myAnova(modelo_reducido)
      Sum_of_Squares DF Mean_Square F_Value      P_value
Model       18430.98  2   9215.4915 386.581 6.13516e-47
Error        2312.33 97     23.8385
```

Modelo reducido

```
Call:
lm(formula = Calificacion ~ Sueno)

Coefficients:
(Intercept)        Sueno
     48.100        5.333
```

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

- Con el método stepwise se obtiene el mismo modelo
- Con el mét. forward se obtiene el mismo.