Comparación variables

$$Y_i^* = \beta_1 X_{i1}^* + \beta_2 X_{i2}^* + \dots + \beta_n X_{in}^* + \varepsilon_i$$
$$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

$$Y_i^* = \frac{Y_i - \bar{Y}}{\sqrt{\frac{1}{n-1}\sum_{i=1}(Y_i-\bar{Y})^2}}$$

$$X_i^* = \frac{X_i - \bar{X}}{\sqrt{\frac{1}{n-1}\sum_{i=1}(X_i-\bar{X})^2}}$$

Difiere un poco de las notas de clase

Selección modelos

(1). Minimizar C_p

$$C_p = \frac{SSE_p}{MSE} - (n-2p)$$

(2). Minimizar $PRESS_{(i)}$

$$PRESS_{(i)} = \sum_{i=1}^{n} e_{(i)}^2$$

(3). Maximizar R_p^2

(4). Maximizar R_{adj}^2

(5). Minimizar SSE_p (MSE)

Bajo el principio parsimonia

Criterios multicolinealidad

$$VIF_j = \frac{1}{1-R_j^2}$$

- $VIF_j \le 5$ no hay mult.
- $5 < VIF_j < 10$ moderada
- $VIF_j \ge 10$ severa

R_j^2: Variabilidad que es explicada por el grado de asociación lineal entre la variable X_j y las demás covariables en general $X_j \sim \dots$

$$X_j = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{j-1} X_{j-1} + \beta_{j+1} \dots + \beta_n X_{in} + \varepsilon_i$$

Índice de condición: Descomposición en valores propios

$$K_j = \frac{\lambda_{max}}{\lambda_j} \quad : \quad j = 1, \dots, n$$

- $10 > \sqrt{K_j}$ No hay multicol.
- $10 \le \sqrt{K_j} < 31{,}6$ moderada
- $\sqrt{K_j} \ge 31{,}6$ severa

Número condición: último índice condición (más grande)

Proporción descomposición varianza $\pi_{ij} > 0{,}5$

- VIF_j, I condición, # cond: Detectar si hay o no hay multicolinealidad y en qué grado
- π_{ij}: Identificar el conjunto de variables que causan la mult.

Métodos selección variables

Forward: $Y_i = \beta_0 + \varepsilon_i \Rightarrow Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in} + \varepsilon_i$

Backward: $Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in} + \varepsilon_i \Rightarrow Y_i = \beta_0 + \varepsilon_i$

Stepwise: Combinación: Bidireccional

Modelos con variables indicadoras

c: # categorías

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

- Solo cambia el intercepto; rectas horizontales

(2). Modelo con variable continua sin interacciones:

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

(Tomando a la primera categoría I_{i1} como referencia)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \beta_3 I_{i2} + \dots + \beta_c I_{i,c-1} + \varepsilon_i$$

Tomando a la última categoría I_{ic} como ref.

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

- Solo cambia el intercepto; rectas con pendiente

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i2} + \beta_3 I_{i3} + \epsilon_i$; $\epsilon_i \sim N(0, \sigma^2)$ — Tomando a la última categoría I_{i1} como ref.

- Solo cambia el intercepto; rectas con pendiente

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

$$Y_i = \underbrace{[\beta_0 + \beta_1 X_{i1} + \beta_2 I_{i2} + \dots + \beta_c I_{ic}]}_{\text{Parte 1: Modelo (2)}} + \underbrace{[\beta_{c1} X_{i1} I_{i2} + \dots + \beta_{2c} X_{i1} I_{ic}]}_{\text{Parte 2: Interacción}} + \epsilon_i$$

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

Tomando I_{i1} como categoría referencia.

- Cambia el intercepto y la pendiente

Ejemplo importante: Método de todas las regresiones posibles

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

tabla todas regresiones posibles

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

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

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

- 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