

Ejemplo:

Marca de Celular

	Rangos	1	2	3	
		Redmi Honor	Motorol Samsung	Apple	
1	<1500	Q <sub>11</sub> 35 22.195	Q <sub>12</sub> 07 14.268	Q <sub>13</sub> 10 15.537	O <sub>1.</sub> = 52
2	[1500-3500]	30 24.75	18 15.91	10 17.32	O <sub>2.</sub> = 58
3	[3500; +∞)	05 23.05	20 14.81	Q <sub>33</sub> 29 16.13	O <sub>3.</sub> = 54
		O <sub>.1</sub> = 70	O <sub>.2</sub> = 45	O <sub>.3</sub> = 49	164

id	X <sub>1</sub>	X <sub>2</sub>
1	3	3
2	2	2
3	1	1
4	2	1
5	3	2
6	3	2
7	2	1
8	3	2
9	2	1
10	1	1
11	3	3
12	1	1
...	...	...

H<sub>0</sub>: No existe relación entre el rango de los ingresos económicos y la marca de celular  
H<sub>a</sub>: Si existe relación...

α = 0.05

$$\chi^2_c = \sum \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

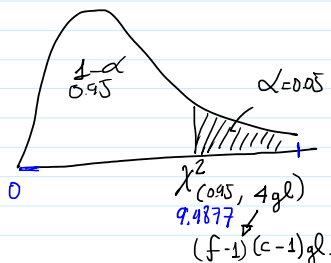
$$\chi^2_c = \frac{(35 - 22.195)^2}{22.195} + \frac{(07 - 14.268)^2}{14.268} + \dots + \frac{(29 - 16.13)^2}{16.13}$$

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{O_{..}}$$

$$E_{11} = \frac{52 \times 70}{164} = 22.195$$

$$\chi^2_c = 43.758$$

Región Crítica



P-value

$$P(\chi^2_{(4 gl)} \geq 43.758) = pchisq(43.758, df=4, lower.tail=F)$$

$$= 7.2 e^{-9} = 0.0000000072$$

$$P\text{-value} = 0.0000$$

Como  $\chi^2_c > \chi^2_0$ , se rechaza H<sub>0</sub>

Existe relación significativa entre los ingresos y marcas de celular

PERFILES FILA

Marca de Celular

	1	2	3	
	Redmi Honor	Motorol Samsung	Apple	
1	Q <sub>11</sub> 35 67.31%	Q <sub>12</sub> 07 13.46%	Q <sub>13</sub> 10 19.23%	O <sub>1.</sub> = 52
2	30 51.72%	18 31.03%	10 17.24%	O <sub>2.</sub> = 58
3	05 9.26%	20 32.04%	Q <sub>33</sub> 29 53.70%	O <sub>3.</sub> = 54
	O <sub>.1</sub> = 70	O <sub>.2</sub> = 45	O <sub>.3</sub> = 49	164

PERFILES COLUMNA

Marca de Celular

	1	2	3	
	Redmi Honor	Motorol Samsung	Apple	
1	Q <sub>11</sub> 35 50%	Q <sub>12</sub> 07	Q <sub>13</sub> 10	O <sub>1.</sub> = 52
2	30 42.85%	18	10	O <sub>2.</sub> = 58
3	05 7.14%	20	Q <sub>33</sub> 29	O <sub>3.</sub> = 54
	O <sub>.1</sub> = 70	O <sub>.2</sub> = 45	O <sub>.3</sub> = 49	164

③ [3500; 4000]

05	20	29	0 <sub>3</sub> = 54
9.26%	32.04%	53.70%	
O <sub>1</sub> = 70	O <sub>2</sub> = 45	O <sub>3</sub> = 49	164

③ [3500; 4000]

05	20	29	0 <sub>3</sub> = 54
7.14%			
O <sub>1</sub> = 70	O <sub>2</sub> = 45	O <sub>3</sub> = 49	164

X<sub>1</sub>: Rango Salarios

- 1 < 1000
- 2 [1000 - 2000]
- 3 [2000 - 3500]
- 4 [3500 - 5000]
- 5 [de 5000 a +]

X<sub>2</sub>: marca celular

- 1: Xiaomi
- 2: Honor
- 3: Huawei-Motorola
- 4: Samsung
- 5: Iphone

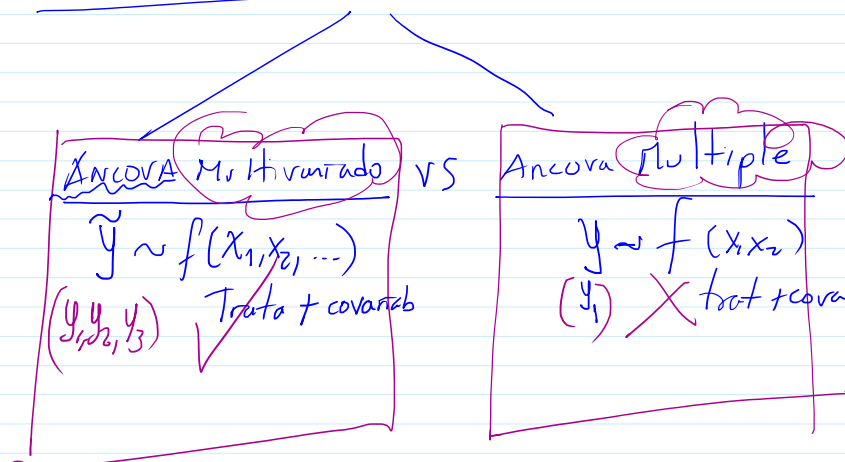
Multivariado

Analisis discriminante; Técnica de dependencia de clasificación  
 $y$   $\nearrow$  categorica (y)

Machine learning

Analisis discriminante Técnica Supervisada de clasificación  
 $y$   $\nearrow$  categorica

Analisis de Covarianza multivariado



Regresión logística binaria Simple

$$y \sim f(\vec{x}) \quad \begin{matrix} y \text{ categorica (2)} \\ x \text{ tiene q ser solo 1} \end{matrix}$$

Regresión logística binaria Multiple

$$y \sim f(\vec{x}) \quad \begin{matrix} y \text{ categorica (2)} \\ x_1, x_2, x_3, \dots > 1 \end{matrix}$$

Regresión logística Multinomial Simple

$$y \sim f(x) \quad \begin{matrix} y \text{ categorica (>2 cate no tienen jerarquía)} \\ x (1) \end{matrix}$$

Regresión logística Multinomial Multiple

$$y \sim f(x) \quad \begin{matrix} y \text{ categorica (>2 cate no tienen jerarquía)} \\ x (x_1, x_2, x_3, \dots) \end{matrix}$$

Regresión logística Ordinal Simple

$$y \text{ categorica (>2 cate tienen jerarquía)} \\ x (1)$$

Regresión logística Ordinal Multiple

Analisis Discriminante

ADLS  $\rightarrow$  Simple ( $y \begin{matrix} \leftarrow K_1 \\ \leftarrow K_2 \end{matrix} \begin{matrix} E \\ F \end{matrix}$ )  
 $\hookrightarrow$  función Discriminante una linea

ADLM  $\rightarrow y \begin{matrix} \leftarrow K_1 \\ \leftarrow K_2 \end{matrix} > 2 \text{ categorias}$   
 $\hookrightarrow$  función (LINEA)

$$\underset{\text{dependiente}}{y} = f(\underbrace{x_1, x_2, x_3, x_4, x_5}_{\text{Independientes}})$$

~~multicolinealidad~~  $\left. \begin{matrix} \nearrow \\ \searrow \end{matrix} \right\} \text{VIF}$

Selección de Variables

Backward  $\bigcirc \rightarrow$

Forward  $\bigcirc \leftarrow$

Stepwise

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_5 x_5$$

$$R^2_{\text{just}} = 0.58$$

$$\text{mod} = \left\{ \begin{matrix} \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots - x_5 \\ \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots - x_5 \end{matrix} \right. \quad \text{Reg} = 0.87$$

$$\text{mod 1 } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3$$

$$= \beta_1 x_1 + \beta_4 x_4$$

$$\neq \beta_1 x_1 + \beta_5 x_5 \quad \text{mod 2}$$

mixtura de  
Backward y  
Forward

$$\text{mod} = \begin{cases} \hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 & -X_5 \quad R^2_{\text{aj}} = 0.67 \\ \hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 & -X_4 \quad R^2_{\text{aj}} = 0.70 \\ \hat{y} & -X_3 \quad R^2_{\text{aj}} = 0.56 \\ \hat{y} & -X_2 \quad R^2_{\text{aj}} = 0.58 \\ \hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 & -X_1 \quad R^2_{\text{aj}} = 0.59 \end{cases}$$

$$\begin{aligned} &= \beta_1 X_1 + \beta_4 X_4 \\ &= \beta_1 X_1 + \beta_5 X_5 \quad R^2 = 0.70 \quad \text{For ward} \\ &X_1 + X_5 \begin{cases} X_2 \Rightarrow R^2 = 0.68 \\ X_3 \Rightarrow R^2 = 0.72 \\ X_4 \Rightarrow R^2 = 0.60 \end{cases} \end{aligned}$$

elimino  $X_4$

$$y = f(X_1, X_5, X_3)$$

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_5 \quad R^2 = 0.70 \quad \checkmark$$

$$\begin{cases} y = & -X_1 \quad R^2 = 0.68 \\ y = & -X_2 \quad R^2 = 0.71 \\ y = & -X_3 \quad R^2 = 0.69 \\ y = & -X_5 \quad R^2 = 0.65 \end{cases}$$

elimino  $X_2$

$$y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_5 X_5 \quad R^2 = 0.71$$