

FORMULARIO ANÁLISIS DE DATOS II

1. Media de una variable

$$\bar{X} = \frac{\sum X}{N}$$

$$\bar{X}_{muestra} = \frac{n_0 \bar{X}_0 + n_1 \bar{X}_1}{N}$$

2. Varianza de una variable

$$S_x^2 = \frac{\sum (X - \bar{X})^2}{N-1} = \frac{1}{N-1} \left(\sum X^2 - \frac{(\sum X)^2}{N} \right) = \frac{\sum x^2}{N-1} = \frac{\sum X^2}{N} - \bar{X}^2$$

$$S_x^2 = pq$$

3. Desviación típica

$$S_x = \sqrt{S_x^2}$$

4. Puntuaciones típicas o estandarizadas

$$Z_x = \frac{X - \bar{X}}{S_x}$$

5. Covarianza entre dos variables

$$S_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N-1} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{N-1} = \frac{\sum xy}{N-1}$$

6. Coeficiente de correlación lineal de Pearson

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\frac{\sum XY - \bar{X}\bar{Y}}{N}}{S_x S_y} = \frac{b S_x}{S_y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}} \sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}}$$

$$r_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{(\sum (X - \bar{X})^2) \sum (Y - \bar{Y})^2}} = \frac{\sum xy}{\sqrt{x^2} \sqrt{y^2}} = \frac{\sum Z_x Z_y}{N-1}$$

7. Recta de regresión

$$\hat{Y} = Y - e$$

$$Y = \hat{Y} + e$$

$$e = Y - \hat{Y}$$

8. Ordenada en el origen y pendiente de la recta de regresión

$$b = \frac{S_{XY}}{S_X^2} = r_{XY} \frac{S_Y}{S_X} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sum X^2 - \frac{(\sum X)^2}{N}} = \frac{\sum xy}{\sum x^2} = \bar{Y}_1 - \bar{Y}_0$$

$$b_0 = \bar{Y} - (b_1 \bar{X}_1 + b_2 \bar{X}_2 + \dots + b_k \bar{X}_k) \text{ (regresión lineal múltiple)}$$

$$a = \bar{Y} - b\bar{X} \rightarrow \bar{Y} = a + b\bar{X} = \bar{Y}_0$$

9. Bondad de ajuste o coeficiente de determinación

$$R^2 = r_{XY}^2 = \frac{SC_{\text{exp}}}{SC_t} = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} = \frac{b^2 \sum (X - \bar{X})^2}{\sum (Y - \bar{Y})^2} = \frac{b^2 S_X^2}{S_Y^2}$$

$$R^2 = \frac{(\bar{Y}_1 - \bar{Y}_0)^2 pq}{S_Y^2}$$

10. Varianza total en el modelo de regresión lineal simple en puntuaciones directas y diferenciales

$$S_t^2 = \frac{SC_t}{N-1} = \frac{\sum (Y - \bar{Y})^2}{N-1} = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N-1} = \frac{\sum y^2}{N-1}$$

11. Varianza total en el modelo de regresión lineal simple en puntuaciones típicas

$$S_t^2 = \frac{\sum Z_Y^2}{N-1} = \frac{N-1}{N-1} = 1$$

12. Varianza explicada en el modelo de regresión lineal simple en puntuaciones directas y diferenciales

$$S_{\text{exp}}^2 = \frac{SC_{\text{exp}}}{k} = \frac{\sum (\hat{Y} - \bar{Y})^2}{k} = \frac{b^2 NS_X^2}{k} = \frac{(\bar{Y}_1 - \bar{Y}_0)^2 Npq}{k}$$

$$S_{\text{exp}}^2 = \frac{b^2 \sum (X - \bar{X})^2}{k} = \frac{b^2 \left[\sum X^2 - \frac{(\sum X)^2}{N} \right]}{k}$$

13. Varianza explicada en el modelo de regresión lineal simple en puntuaciones típicas

$$S_{\text{exp}}^2 = \frac{SC_{\text{exp}}}{k} = \frac{\sum \hat{Z}_Y^2}{k} = \frac{r_{XY}^2 \sum Z_X^2}{k} = \frac{r_{XY}^2 (N-1)}{k}$$

14. Varianza residual o no explicada en el modelo de regresión lineal simple en puntuaciones directas y diferenciales

$$S_{\text{res}}^2 = \frac{SC_{\text{res}}}{N-k-1} = \frac{\sum (Y - \hat{Y})^2}{N-k-1} = \frac{SC_t - SC_{\text{exp}}}{N-k-1} = \frac{NS_Y^2 - b^2 NS_X^2}{N-k-1}$$

$$S_{\text{res}}^2 = N(S_Y^2 - b^2 S_X^2) = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N} - b^2 \left[\sum X^2 - \frac{(\sum X)^2}{N} \right]}{N-k-1}$$

$$S_{\text{res}}^2 = \frac{n_0 S_{Y_0}^2 + n_1 S_{Y_1}^2}{n_0 + n_1 - 2}$$

15. Varianza residual o no explicada en el modelo de regresión lineal simple en puntuaciones típicas

$$S_{\text{res}}^2 = \frac{SC_t - SC_{\text{exp}}}{N-k-1} = \frac{(N-1)(1-r_{XY}^2)}{N-k-1}$$

16. Validación del modelo de regresión

Fuentes de variación	Sumas de cuadrados	Grados de libertad	Varianza	F
Regresión o explicada	$\sum (\hat{Y} - \bar{Y})^2$	k	$S_{\text{exp}}^2 = \frac{SC_{\text{exp}}}{k}$	$\frac{S_{\text{exp}}^2}{S_{\text{res}}^2} = \frac{\frac{R_{XY}^2}{k}}{\frac{1-R_{XY}^2}{N-k-1}}$
Residual o no explicada	$\sum (Y - \hat{Y})^2$	N-k-1	$S_{\text{res}}^2 = \frac{SC_{\text{res}}}{N-k-1}$	
Total	$\sum (Y - \bar{Y})^2$	N-1	$S_t^2 = \frac{SC_t}{N-1}$	

$$SC_{\text{total}} = SC_{\text{exp}} + SC_{\text{res}}$$

17. Significación de parámetros

$$t = \frac{|b|}{S_b} = \frac{|b|}{\sqrt{\frac{S_{res}^2}{\sum (X - \bar{X})^2}}} = \frac{|b|}{\sqrt{\frac{S_{res}^2}{\sum X^2 - \frac{(\sum X)^2}{N}}}} = \frac{|b|}{\sqrt{\frac{S_Y^2(1 - R_{XY}^2)}{N - 2}}}$$

$$t = \frac{|r_{XY}|}{\sqrt{\frac{1 - r_{XY}^2}{N - 2}}} = \frac{|\bar{Y}_0 - \bar{Y}_1|}{\sqrt{S_Y^2 \left(\frac{1}{n_0} + \frac{1}{n_1} \right)}}$$

18. Intervalos de predicción: pronósticos

$$Y_0 : \hat{Y} \pm t_{(\alpha, N-k-1)} \sqrt{S_{res}^2 \left(1 + \frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum (X - \bar{X})^2} \right)}$$