FORMULARIO ANÁLISIS DE DATOS II

1. **Media de una variable**

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

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

2. **Varianza de una variable**

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

$$S_x^2 = pq$$

$$\sum Y^2 = \sum Y_0^2 + \sum Y_1^2$$

$$\sum Y_0^2 = \left(S_{Y_0}^2 + \overline{Y}_0^2\right)n$$

$$\sum Y_1^2 = \left(S_{Y_1}^2 + \overline{Y}_1^2\right)n$$

3. Desviación típica

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

4. **Puntuaciones típicas o estandarizadas**

$$Z_X = \frac{X - \overline{X}}{S_X}$$

5. Covarianza entre dos variables

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

6. Coeficiente de correlación lineal de Pearson

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\frac{\sum XY}{N} - \overline{X}\,\overline{Y}}{S_X S_Y} = \frac{bS_X}{S_Y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}\sqrt{\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}}}$$

$$r_{XY} = \frac{\sum \left(X - \overline{X}\right)\left(Y - \overline{Y}\right)}{\sqrt{\left(\sum X - \overline{X}\right)^2}\sqrt{\left(\sum Y - \overline{Y}\right)^2}} = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = \frac{\sum Z_X Z_Y}{N-1}$$

7. Recta de regresión

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

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

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

$$\overline{Y} = \overline{Y}_{estimada}$$

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

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

$$b_0 = \overline{Y} - \left(b_1 \overline{X}_1 + b_2 \overline{X}_2 + \ldots + b_k \overline{X}_k\right) \text{ (regresión lineal múltiple)}$$

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

9. Bondad de ajuste o coeficiente de determinación

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

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

$$R^2 = \frac{b^2\left(\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right)}{\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}}$$

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

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

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

$$S^2_{\exp} = \frac{SC_{\exp}}{k} = \frac{\sum\left(\hat{Y} - \overline{Y}\right)^2}{k} = \frac{b^2 N S^2_X}{k} = \frac{\left(\overline{Y}_1 - \overline{Y}_0\right)^2 Npq}{k}$$

$$S^2_{\exp} = \frac{b^2 \sum\left(X - \overline{X}\right)^2}{k} = \frac{b^2\left[\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right]}{k}$$

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

$$S^2_{\exp} = \frac{SC_{\exp}}{k} = \frac{\sum \hat{Z}_Y^2}{k} = \frac{r^2_{XY} \sum Z^2_X}{k} = \frac{r^2_{XY}(N-1)}{k}$$

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

$$S^2_{res} = \frac{SC_{res}}{N-k-1} = \frac{\sum\left(Y - \hat{Y}\right)^2}{N-k-1} = \frac{SC_t - SC_{\exp}}{N-k-1} = \frac{NS^2_Y - b^2 NS^2_X}{N-k-1}$$

$$S^2_{res} = \frac{N\left(S^2_Y - b^2 S^2_X\right)}{N-k-1} = \frac{\sum Y^2 - \frac{\left(\sum Y\right)^2}{N} - b^2\left[\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right]}{N-k-1}$$

$$S^2_{res} = \frac{n_0 S^2_{Y_0} + n_1 S^2_{Y_1}}{n_0 + n_1 - 2}$$

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

$$S^2_{res} = \frac{SC_t - SC_{\exp}}{N-k-1} = \frac{(N-1)\left(1 - r^2_{XY}\right)}{N-k-1}$$

16. Validación del modelo de regresión

Fuentes de variación	Sumas de cuadrados	Grados de libertad	Varianza	F
Regresión o explicada	$\sum\left(\hat{Y} - \overline{Y}\right)^2$	k	$S^2_{\exp} = \frac{SC_{\exp}}{k}$	$\frac{S^2_{\exp}}{S^2_{res}} = \frac{\frac{R^2_{XY}}{k}}{\frac{1-R^2_{XY}}{N-k-1}}$
Residual o no explicada	$\sum\left(Y - \hat{Y}\right)^2$	N-k-1	$S^2_{res} = \frac{SC_{res}}{N-k-1}$	
Total	$\sum\left(Y - \overline{Y}\right)^2$	N-1	$S^2_t = S^2_Y = \frac{SC_t}{N-1}$	

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

17. Significación de parámetros

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

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

18. Intervalos de predicción: pronósticos

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