PO11Q & PO12Q

Formula Collection and Notation

Dr Florian Linke
Florian.Linke@warwick.ac.uk

1 | PO11Q

Statistic	Formula
Confidence Interval	$Pr(\bar{y} - t_{\alpha/2} \cdot se \leq \mu \leq \bar{y} + t_{\alpha/2} \cdot se) = 1 - \alpha$
Deviation	$d = y_i - \bar{y}$
Mean	$\bar{y} = \frac{\Sigma y_i}{n}$
	$\mu = \Sigma y P(y) = E[y]$
Position of pth percentile	$P = (n + 1) \cdot \frac{p}{100}$
Range	$y_{range} = y_{max} - y_{min}$
Standard Deviation	$s = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}}$
Standard Error	$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$
	$se = \frac{s}{\sqrt{n}}$
t-test	$t = \frac{\bar{y} - \mu_0}{se}$
Variance	$s^2 = \frac{\Sigma(y_i - \bar{y})^2}{n-1}$
z-score	$z = \frac{y - \mu}{\sigma}$

Table 1: Formulae for PO11Q

Statistic	Formula
Crosstabulations	
χ^2-Test Statistic	$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
Two Sample Tests	
Standard Error of Difference	$se = \sqrt{(se_1)^2 + (se_2)^2}$
Means	
Standard Error of Difference	$se = \sqrt{\frac{se_1^2}{n_1} + \frac{se_2^2}{n_2}}$
Standard Error (pooled)	$se_0 = \sqrt{s_0^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
Variance (pooled)	$s_0^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
Proportions	
Standard Error	$se = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$
Standard Error of Difference	$se = \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}$
Standard Error (pooled)	$se_0 = \sqrt{\bar{\pi}(1 - \bar{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
Regression	
Adjusted R^2	$\bar{R}^2 = 1 - \frac{\sum \hat{\epsilon}_i^2 / (n - k - 1)}{\sum (y_i - \bar{y})^2 / (n - 1)}$
Coefficient of Determination	$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{\epsilon}_i^2}{\sum (y_i - \bar{y})^2}$
Estimated Variance	$\hat{\sigma}^2 = \frac{\sum \hat{\epsilon}_i^2}{n - p}$
Estimator of β_0	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
Estimator of β_1	$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$
	$\left.\right\}$ $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$
Explained Sum of Squares	$\sum (\hat{y}_i - \bar{y})^2$
Residual Sum of Squares	$\sum \hat{\epsilon}_i^2 = \sum (y_i - \hat{y}_i)^2$
Estimated Standard Error of $\hat{\beta}_0$	$\hat{se}(\hat{\beta}_0) = \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}\ \hat{\sigma}$
Estimated Standard Error of $\hat{\beta}_1$	$\hat{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}}$
	$\left.\right\}$ $\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$
Total Sum of Squares	$\sum (y_i - \bar{y})^2$

Table 2: Formulae for PO12Q

2.1 Matrices

$$\boldsymbol{X'X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \times \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$(\boldsymbol{X'X})^{-1} = \frac{1}{n \sum x_i^2 - \sum x_i \sum x_i} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

$$\boldsymbol{X'y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

3 | Notation

Symbol	Explanation
PO11Q	
d	Deviation
n	Sample Size
s	Standard Deviation
s^2	Variance
$\bar{y}$	Mean
y_i	Observation i
f	(Absolute) Frequency
cf	Cumulative (Absolute) Frequency
rf	Relative Frequency
crf	Cumulative Relative Frequency
$E[x]$	The expected value of x
μ	Mean of the Population
se	Standard Error (with s of sample)
σ	Standard Deviation of the Population
$\sigma_{\bar{y}}$	Standard Error (with σ of population)
t	t-value
z	z-value
PO12Q	
χ^2	Chi-Squared for test of independence
π	Population Proportion
$\hat{\pi}$	Sample Proportion
$\bar{\pi}$	Pooled Proportion
se_0	Standard Error under the Null Hypothesis
β	Regression Coefficient
$\hat{\beta}$	Estimated Regression Coefficient
ϵ	Error Term
$\hat{\epsilon}$	Estimated Error Term / Residual
A^{-1}	Inverse of Matrix A
A'	Transpose of Matrix A
I	Identity Matrix
R^2	R-Squared / Model Fit
σ^2	Mean Squared Error
k	Number of Slope Coefficients
$\bar{\mathrm{R}}^2$	Adjusted R-Squared
log(x)	Logarithm of variable x
α_i	Regression coefficients of secondary regression models
p	Total Number of Regression Coefficients, including the Intercept
VcV	Variance-Covariance Matrix
Ω	Is equal to $\sigma^2 I$, where σ^2 represents the mean squared error
t	Time period in time-series data

Table 3: Notation