Chapter 301

Lin's Concordance Correlation Coefficient

Introduction

This procedure calculates Lin's concordance correlation coefficient (ρ_C) from a set of bivariate data.

The statistic, ρ_c , is an index of how well a new test or measurement (Y) reproduces a gold standard test or measurement (X). It quantifies the agreement between these two measures of the same variable (e.g., chemical concentration). Like a correlation, ρ_c ranges from -1 to 1, with perfect agreement at 1. It cannot exceed the absolute value of ρ, Pearson's correlation coefficient between Y and X. It can be legitimately calculated on as few as ten observations.

The formulas used are found in an appendix of McBride (2005). The development of this statistic is presented Lin (1989, 1992, 2000), Lin, Hedayat, Sinha, and Yang (2002), and Lin, Hedayat, and Wu (2012).

Technical Details

Following Lin et al. (2002), assume that n observations Y_k and X_k are selected from a bivariate population with means μ_Y and μ_X, variances σ_Y^2 and σ_X^2, and correlation ρ (the Pearson correlation coefficient). Here, Y represents a measure from a candidate test or method and X represents the corresponding measure of the gold standard test or method.

The degree of concordance (agreement) between the two measures can be characterized by the expected value of their squared difference

$$E[(Y-X)^2] = (\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2 - 2\rho\sigma_Y\sigma_X \,.$$

This is the expected squared perpendicular deviation from a 45° line through the origin.

If every pair from the bivariate population is in exact agreement, the above expectation would be 0. In order to create an index of concordance scaled to between -1 and 1, Lin used:

$$\rho_C = 1 - \frac{E[(Y-X)^2]}{[(Y-X)^2 \mid \rho = 0]}$$

$$= 1 - \frac{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2 - 2\rho\sigma_Y\sigma_X}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2}$$

$$= \frac{2\rho\sigma_Y\sigma_X}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2}$$

The value of ρ_C is estimated from a sample by $\hat{\rho}_C$ where the usual sample counterparts are substituted into the above formula to obtain

$$\hat{\rho}_C = \frac{2S_{YX}}{(\bar{Y} - \bar{X})^2 + S_Y^2 + S_X^2}$$

where

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

$$\bar{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k$$

$$S_X^2 = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{X})^2$$

$$S_Y^2 = \frac{1}{n}\sum_{k=1}^{n} (Y_k - \bar{Y})^2$$

$$S_{XY} = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{X})(Y_k - \bar{Y})$$

In order to achieve a better approximation with the normal distribution, Lin (1989) transforms $\hat{\rho}_C$ using Fisher's z transformation to obtain

$$\hat{\lambda} = \tanh^{-1}(\hat{\rho}_C) = \frac{1}{2}\ln\left(\frac{1 + \hat{\rho}_C}{1 - \hat{\rho}_C}\right).$$

This quantity has an asymptotically normal distribution with mean

$$\lambda = \tanh^{-1}(\rho_C) = \frac{1}{2}\ln\left(\frac{1 + \rho_C}{1 - \rho_C}\right)$$

and variance

$$\sigma(\rho, \upsilon, n)^2 = \frac{1}{n-2}\left\{\frac{(1 - \rho^2)\rho_C^2}{(1 - \rho_C^2)\rho^2} + \frac{2\rho_C^3(1 - \rho_C)\upsilon^2}{\rho(1 - \rho_C^2)^2} - \frac{\rho_C^4\upsilon^4}{2\rho^2(1 - \rho_C^2)^2}\right\}.$$

where

$$\upsilon = \frac{|\mu_Y - \mu_X|}{\sqrt{\sigma_Y \sigma_X}}$$

This variance is estimated using

$$S_{\hat{\lambda}}^2 = \frac{1}{n-2}\left\{\frac{(1-r^2)\hat{\rho}_C^2}{(1-\hat{\rho}_C^2)r^2} + \frac{2\hat{\rho}_C^3(1-\hat{\rho}_C)u^2}{r(1-\hat{\rho}_C^2)^2} - \frac{\hat{\rho}_C^4 u^4}{2r^2(1-\hat{\rho}_C^2)^2}\right\}$$

where

$$u = \frac{|\bar{Y}-\bar{X}|}{\sqrt{S_Y S_X}}$$

Note that we follow the advice of McBride and do not multiple u by (n – 1)/n as he originally suggested.

A lower *100(1 – α)%* confidence limit can be calculated using the normal distribution as follows.

$$\hat{\rho}_{C,lower} = \tanh\left(\hat{\lambda} - z_\alpha S_{\hat{\lambda}}\right)$$

Other confidence intervals can be obtained similarly.

Missing Values

Rows with missing values in either of the variables are ignored.

Example 1 – Comparing Two Measurements

In this example, we will compare a new quick measurement to an expensive, gold standard measure.

Setup

To run this example, complete the following steps:

1 Open the Lins CCC example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Lins CCC** and click **OK**.

2 Specify the Lin's Concordance Correlation Coefficient procedure options

- Find and open the **Lin's Concordance Correlation Coefficient** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y: First Measurement Variable ....**Quick**
X: Second Measurement Variable ....**GoldStd**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Run Summary Section

Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	Quick	Rows Processed	15
Independent Variable	GoldStd	Rows Used in Estimation	15
Frequency Variable	None	Rows with X Missing	0
R-Squared	0.9911	Rows with Freq Missing	0
Correlation	0.9956	Sum of Frequencies	15
Coefficient of Variation	0.0486		
Mean Square Error, MSE	4.870055		
√MSE	2.20682		

This report shows information about which variables and rows were processed.

Lin's Concordance Correlation Coefficient Section

Lin's Concordance Correlation Coefficient Section

Parameter	Value
Concordance Correlation Coefficient, ρc	0.9953
Lower, One-Sided 95% C.L. of ρc	0.9885
Upper, One-Sided 95% C.L. of ρc	0.9984
Lower, Two-Sided 95% C.L. of ρc	0.9863
Upper, Two-Sided 95% C.L. of ρc	0.9984

This report shows Lin's coefficient and corresponding confidence intervals.

Concordance Correlation Coefficient, ρc

The estimated value of Lin's concordance correlation coefficient.

Lower, One-Sided 95% C.L. of ρc

This is the limit of a one-sided, 95% lower confidence interval for ρc. The interval is all values higher than this value. Thus, the confidence interval estimates that the true value is at least this value.

Upper, One-Sided 95% C.L. of ρc

This is the limit of a one-sided, 95% upper confidence interval for ρc. The interval is all values lower than this value. Thus, the confidence interval estimates that the true value is less than this value.

Two-Sided 95% C.L. of ρc

These are the limits of a two-sided, 95% confidence interval for ρc. The interval is all values between the upper and lower values. Thus, the confidence interval estimates that the true value is between these values.

Descriptive Statistics Section

Descriptive Statistics Section

Parameter	Measurement 1	Measurement 2
Variable	Quick	GoldStd
Count	15	15
Mean	45.4	45
Standard Deviation	22.60468	22.36068
Minimum	12	10
Maximum	85	80

This report presents the usual descriptive statistics.

Linear Regression Section

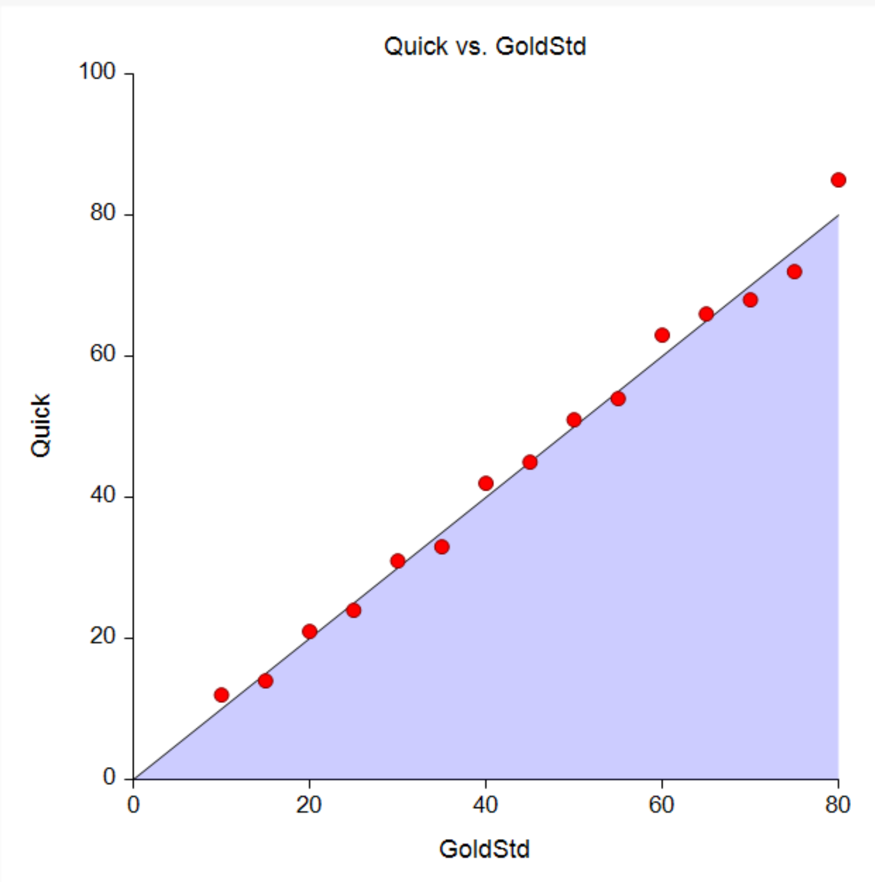
Linear Regression Section

Parameter	Intercept B(0)	Slope B(1)
Regression Coefficients	0.1107	1.0064
Lower 95% C.L.	-2.7337	0.9494
Upper 95% C.L.	2.9551	1.0634
Standard Error	1.3166	0.0264
T Value	0.0841	38.1562
Prob Level (T Test)	0.9343	0.0000

This report displays a brief summary of a linear regression of Y on X.

Scatter Plot

Plots Section



Scatter Plot

This scatter plot helps you to assess the agreement of the two measurements. The plot is shaded below the 45° line to make the assessment easier.