

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:

Lin's Concordance Correlation Coefficient

$$\begin{aligned}\rho_C &= 1 - \frac{E[(Y - X)^2]}{E[(Y - X)^2 | \rho = 0]} \\ &= 1 - \frac{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2 - 2\rho\sigma_Y\sigma_X}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2} \\ &= \frac{2\rho\sigma_Y\sigma_X}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2}\end{aligned}$$

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

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

where

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

Lin's Concordance Correlation Coefficient

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.

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

Other confidence intervals can be obtained similarly.

Missing Values

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

Procedure Options

This section describes the options available in this procedure.

Variables Tab

This panel specifies the variables used in the analysis.

Variables

Y: First Measurement Variable

This variable contains the values of Y, the candidate measurement.

X: Second Measurement Variable

This variable contains the values of X, the gold standard measurement. This is the variance to which Y is compared.

In some situations, there is no gold standard. You just want to study the agreement of two measurement variables. In that case, this is the second measurement variable.

Frequency Variable

Specify an optional frequency (count) variable here. This variable contains integers that represent the number of observations (frequency) associated with each row of the dataset.

If this option is left blank, each dataset row has a frequency of one. This variable lets you modify that frequency. This may be useful when your data were previously tabulated and you want to enter counts.

Alpha Levels

Alpha for C.I.'s

This is the value of alpha for the asymptotic confidence intervals of ρ_c . A 100 (1 – alpha)% confidence interval is calculated. Usually, this number will range from 0.1 to 0.001. A common choice for alpha is 0.05, which results in a 95% confidence interval.

Reports Tab

The following options control which reports and plots are displayed.

Select Reports

Run Summary ... Linear Regression

These options specify which reports are displayed.

Report Options

Precision

Specify the precision of numbers in the report. Single precision will display seven-place accuracy, while the double precision will display thirteen-place accuracy. Note that all reports are formatted for single precision only.

Variable Names

Specify whether to use variable names or (the longer) variable labels in report headings.

Decimal Places

Specify the number of digits after the decimal point to display on the output of values of this type. Note that this option in no way influences the accuracy with which the calculations are done.

Enter *All* to display all digits available. The number of digits displayed by this option is controlled by whether the *Precision* option is *Single* or *Double*.

Select Plots

Y vs X

These options control whether the scatter plot is displayed, its size, and its format.

Click the large plot format button to change the plot settings.

Example 1 – Comparing Two Measurements

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

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Lin's Concordance Correlation Coefficient window.

1 Open the Lins CCC dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Lins CCC.NCSS**.
- Click **Open**.

2 Open the Lin's Concordance Correlation Coefficient window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Lin's Concordance Correlation Coefficient** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- On the procedure window, select the **Variables tab**.
- Double-click in the **Y: First Measurement Variable** box. This will bring up the variable selection window.
- Select **Quick** from the list of variables and then click **Ok**.
- Double-click in the **X: Second Measurement Variable** box. This will bring up the variable selection window.
- Select **GoldStd** from the list of variables and then click **Ok**.

4 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

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		
$\sqrt{\text{MSE}}$	2.20682		

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

Lin's Concordance Correlation Coefficient Section

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

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

Lin's Concordance Correlation Coefficient**Concordance Correlation Coefficient, ρ_c**

The estimated value of Lin's concordance correlation coefficient.

Lower, One-Sided 95% C.L. of ρ_c

This is the limit of a one-sided, 95% lower confidence interval for ρ_c . 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 ρ_c

This is the limit of a one-sided, 95% upper confidence interval for ρ_c . 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 ρ_c

These are the limits of a two-sided, 95% confidence interval for ρ_c . 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

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

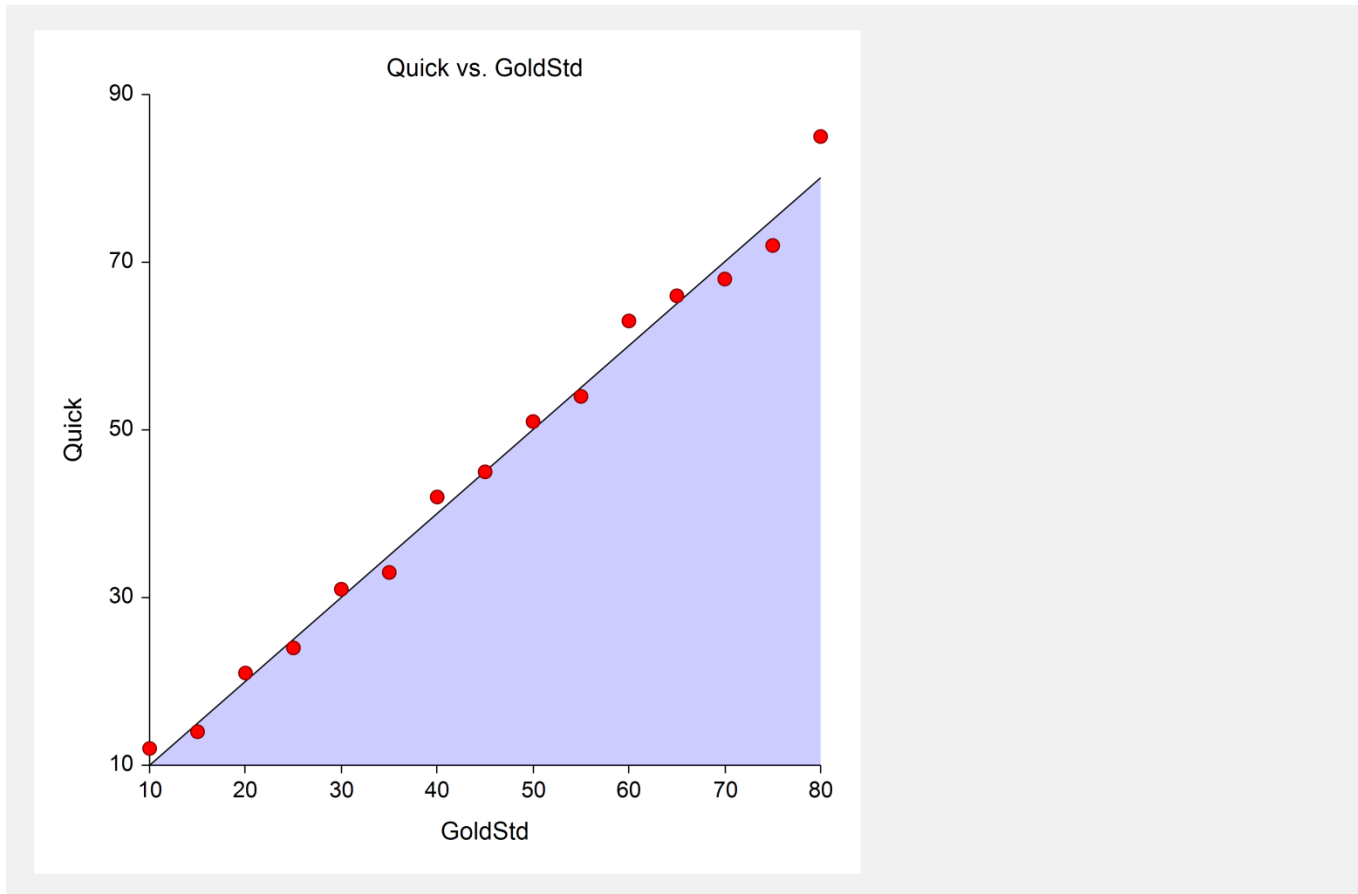
This report presents the usual descriptive statistics.

Linear Regression Section

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



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.