

Chapter 313

Passing-Bablok Regression for Method Comparison

Introduction

Passing-Bablok regression for method comparison is a robust, nonparametric method for fitting a straight line to two-dimensional data where both variables, X and Y , are measured with error. It is useful when you have two devices that should give the same measurements and you want to compare them. This is accomplished by estimating a linear regression line and testing whether the intercept is zero and the slope is one. The Passing-Bablok regression procedure fits the intercept (β_0) and the slope (β_1) of the linear equation

$$Y = \beta_0 + \beta_1 X.$$

The estimate of the slope ($B1$) is calculated as the median of all slopes that can be formed from all possible pairs of data points, except those pairs that result in a slope of 0/0 or -1. To correct for estimation bias caused by the lack of independence of these slopes, the median is *shifted* by a factor K which is the number of slopes that are less than -1. This creates an approximately unbiased estimator. The estimate of the intercept ($B0$) is the median of $\{Y_i - B1 X_i\}$.

The intercept is interpreted as the systematic bias (difference) between the two methods. The slope measures the amount of proportional bias (difference) between the two methods.

Experimental Design

Typical designs suitable for Passing-Bablok regression include up to 3000 paired measurements, (x_i, y_i) , $i = 1, \dots, n$, similar to the common input for simple linear regression. Typical data of this type are shown in the table below.

Typical Data for Passing-Bablok regression

Subject	X	Y
1	7	7.9
2	8.3	8.2
3	10.5	9.6
4	9	9
5	5.1	6.5
6	8.2	7.3
7	10.2	10.2
8	10.3	10.6

Technical Details

The methods and results in this chapter are based on the formulas given in Passing and Bablok (1983, 1984) and Bablok, Passing, Bender, and Schneider (1988).

Assumptions

Passing-Bablok regression requires the following assumptions:

1. The variables X and Y are highly, positively correlated (for method comparison only).
2. The relationship between X and Y is linear (straight-line).
3. No special assumptions are made about the distributions (including the variances) of X and Y .

Passing-Bablok Estimation

Define X_i and Y_i , $i = 1, \dots, n$, as the values for two variables, each sampled with error to give the observed values x_i and y_i , respectively. For each of the $N = \binom{n}{2}$ possible pairs of points, define the slope by

$$S_{ij} = \frac{y_i - y_j}{x_i - x_j}$$

If $x_i = x_j$ and $y_i = y_j$, the result is $0/0$ which is undefined. Omit these pairs from the calculation of the slope.

If $x_i = x_j$ and $y_i > y_j$, the result is ∞ . Enter a large positive number. Since we are using the median of the S_{ij} , the value won't matter.

If $x_i = x_j$ and $y_i < y_j$, the result is $-\infty$. Enter a large negative number. Since we are using the median of the S_{ij} , the value won't matter.

If $S_{ij} = -1$, omit the pair from the calculation of the slope. This corrects for the bias caused by the lack of independence in the S_{ij} .

If $S_{ij} < -1$, use the pair and add one to a counter K . This value will be used to shift the median.

The slope B_1 is the shifted median of S_{ij} , where the median is shifted to the right K steps.

Using this slope, calculate intercept B_0 as the median of all n of the quantities $y_i - B_1 x_i$.

Confidence Bounds

Calculate the confidence bounds for β_1 as follows. Let $z_{\alpha/2}$ be the $1 - \alpha/2$ quantile of the standard normal distribution. Let

$$C_{\alpha/2} = z_{\alpha/2} \sqrt{\frac{n(n-1)(2n+5)}{18}}$$

Now calculate

$$M_1 = \left[\frac{N - C_{\alpha/2}}{2} \right]$$

Here, $[U]$ rounds U to the nearest integer.

Also calculate

$$M_2 = N - M_1 + 1$$

Passing-Bablok Regression for Method Comparison

Finally, a confidence interval for β_1 is given by

$$S_{(M_1+K)} \leq \beta_1 \leq S_{(M_2+K)}$$

Where the S_{ij} 's are sorted.

Confidence limits for the intercept are calculated as follows. Let $B1_L$ and $B1_U$ be the lower and upper limits for the slope from the last calculation. Now calculate the limits for the intercept as

$$B0_L = \text{median}\{y_i - B1_U x_i\}$$

and

$$B0_U = \text{median}\{y_i - B1_L x_i\}$$

Cusum Test of the Linearity Assumption

An important assumption is that the relationship between X and Y is linear. Passing and Bablok (1983) proposed testing this assumption using a modified cusum test. This test proceeds as follows.

Let n_{pos} represent the number of observations with $y_i > B0 + B1 x_i$. Similarly, let n_{neg} represent the number of observations with $y_i < B0 + B1 x_i$.

Step 1 Assign score r_i to each point as follows.

$$\text{If } y_i > B0 + B1 x_i \text{ then } r_i = \sqrt{n_{neg}/n_{pos}}$$

$$\text{If } y_i < B0 + B1 x_i \text{ then } r_i = -\sqrt{n_{pos}/n_{neg}}$$

$$\text{If } y_i = B0 + B1 x_i \text{ then } r_i = 0$$

Step 2 Assign the distance scores D_i to each observation as follows.

$$D_i = \frac{y_i + \frac{x_i}{B1} - B0}{\sqrt{1 + 1/B1^2}}$$

Step 3 Sort the r_i in the order of the D_i . Label these sorted values $r_{(i)}$.

Step 4 Create the $cusum_{(i)}$ statistics using the following.

$$cusum_{(i)} = \sum_{k=1}^i r_{(k)}$$

Step 5 Find the maximum absolute value of $cusum_{(i)}$ across all values of i . Call this $MaxCusum$.

Step 6 The null hypothesis of random arrangement of residuals and therefore a linear relationship between X and Y is tested using the test statistic

$$H = \frac{MaxCusum}{\sqrt{n_{neg} + 1}}$$

This statistic is compared to the Kolmogorov-Smirnov distribution. Details are given in Passing and Bablok (1983). For example, the critical value for a 5% test is 1.36 and for a 1% test is 1.63. Note that we created an approximating polynomial that allows us to calculate p-values for any value of H.

Kendall's Tau Test of the High Correlation Assumption

Passing and Bablok (1983) recommended that a preliminary two-sided test be conducted to determine if Kendall's tau correlation between X and Y is significantly different from zero. They also indicate that this correlation must be positive. Kendall's tau correlation is well documented in the *Correlation* procedure.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

This panel specifies the variables and estimation parameters used in the analysis.

Variables

Y Variable

Specify a single data column to be used for the Y variable. This column contains the measurements from one of the two measurement systems being compared. The X variable contains corresponding measurements from the other (X) measurement system.

It is arbitrary which method is selected for X and which is selected for Y.

The estimated equation will be of the form

$$Y = B0 + B1 \times X$$

The analysis will test whether $B0 = 0$ and $B1 = 1$. If they do, then it follows that $Y = X$.

You can enter the column name or number directly, or click the button on the right to display a Column Selection window that will let you select the column from a list.

X Variable

Specify a single data column to be used for the X variable. This column contains the measurements from one of the two measurement systems being compared. The Y variable contains corresponding measurements from the other (Y) measurement system.

It is arbitrary which method is selected for X and which is selected for Y.

The estimated equation will be of the form

$$Y = B0 + B1 \times X$$

The analysis will test whether $B0 = 0$ and $B1 = 1$. If they do, then it follows that $Y = X$.

You can enter the column name or number directly, or click the button on the right to display a Column Selection window that will let you select the column from a list.

Grouping Variable (Optional)

Grouping Variable

Enter a single categorical grouping variable. The values of this variable indicate which category each subject belongs in. Values may be text or numeric. The grouping variable is optional.

A separate analysis will be performed for each group.

Reports Tab

This tab controls which statistical reports are displayed in the output.

Significance Levels and Confidence Levels

Confidence Level

This confidence level, entered as a percentage, is used in the calculation of confidence intervals for regression coefficients.

Typical confidence levels are 90%, 95%, and 99%, with 95% being the most common.

Range

$50 < \text{Confidence Level} < 100$

Assumptions Alpha

Specify the alpha value (significance level) used for tests of assumptions. Alpha is the probability of rejecting the null hypothesis when it is actually true.

Recommended

Often, an alpha of 0.05 is used. However, many statisticians recommend a higher value of alpha for tests of assumptions such as 0.10 or even 0.20.

Range

Typical choices for alpha range between 0.001 and 0.200.

Select Reports

Run Summary to Residuals

Each of these options specifies whether the indicated report is displayed.

Report Options

The following options control the format of the reports.

Variable Names

Specify whether to use variable names, variable labels, or both to label output reports. In this discussion, the variables are the columns of the data table.

- **Names**

Variable names are the column headings that appear on the data table. They may be modified by clicking the Column Info button on the Data window or by clicking the right mouse button while the mouse is pointing to the column heading.

- **Labels**

This refers to the optional labels that may be specified for each column. Clicking the Column Info button on the Data window allows you to enter them.

- **Both**

Both the variable names and labels are displayed.

Passing-Bablok Regression for Method Comparison

Comments

1. Most reports are formatted to receive about 12 characters for variable names.
2. Variable Names cannot contain blanks or math symbols (like + - * / . ,), but variable labels can.

Value Labels

Value Labels are used to make reports more legible by assigning meaningful labels to numbers and codes.

The options are

- **Data Values**

All data are displayed in their original format, regardless of whether a value label has been set or not.

- **Value Labels**

All values of variables that have a value label variable designated are converted to their corresponding value label when they are output. This does not modify their value during computation.

- **Both**

Both data value and value label are displayed.

Example

A variable named GENDER (used as a grouping variable) contains 1's and 2's. By specifying a value label for GENDER, the report can display "Male" instead of 1 and "Female" instead of 2. This option specifies whether (and how) to use the value labels.

Report Options – Decimal Places

Item Decimal Places

These decimal options allow the user to specify the number of decimal places for items in the output. Your choice here will not affect calculations; it will only affect the format of the output.

- **Auto**

If one of the "Auto" options is selected, the ending zero digits are not shown. For example, if "Auto (0 to 7)" is chosen,

0.0500 is displayed as 0.05

1.314583689 is displayed as 1.314584

The output formatting system is not designed to accommodate "Auto (0 to 13)", and if chosen, this will likely lead to lines that run on to a second line. This option is included, however, for the rare case when a very large number of decimals is needed.

Plots Tab

These options specify which plots are produced as well as the plot format.

Passing-Bablok Regression Plots

Passing-Bablok Regression Scatter Plot

Indicate whether to display this plot. Click the plot format button to change the plot settings.

Show Combined Plot (If Grouping Variable Present)

If you have a grouping variable present, this option allows you to plot all groups on one plot for comparison. Passing-Bablok regression is still performed on each group separately.

Residual Plot

Residual Plot

Indicate whether to display this plot. Click the plot format button to change the plot settings.

Example 1 – Passing-Bablok Regression for Method Comparison

This section presents an example of how to run a Passing-Bablok regression method comparison of the data in the *PassBablok 1* dataset. This dataset contains measurements from two measurement methods on each of 30 items. The goal is to determine if the two measurement systems obtain equivalent measurements. 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 Passing-Bablok Regression for Method Comparison window.

1 Open the PassBablok 1 example dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **PassBablok 1.NCSS**.
- Click **Open**.

2 Open the Passing-Bablok Regression for Method Comparison window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Passing-Bablok Regression for Method Comparison** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- Select the **Variables** tab.
- For **Y Variable**, enter **Method2**.
- For **X Variable**, enter **Method1**.
- Leave the **Grouping Variable** blank.

4 Specify the reports.

- Select the **Reports** tab.
- In addition to the reports already selected, check **Residuals**.

5 Specify the plots.

- Select the **Plots** tab.
- Click on the **Passing-Bablok Regression Scatter Plot** button.
- Click on the **Residuals Plot** button.

6 Run the procedure.

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

Run Summary

Item	Value	Item	Value
Y Variable	Method2	Rows Used	30
X Variable	Method1	B0: Intercept	-0.0092
		B1: Slope	0.9986

This report gives a summary of the input and various descriptive measures about the Passing-Bablok regression.

Passing-Bablok Regression for Method Comparison

Descriptive Statistics

Item	Y	X	Y - X
Variable	Method2	Method1	Difference
N	30	30	30
Mean	62.99333	63.14	-0.1466667
Std Dev	23.5886	23.6153	2.2578
Std Error	4.3067	4.3115	0.4122
COV	0.3745	0.3740	-15.3942
Minimum	12.7	14.2	-3.8
First Quartile	47.425	48.8	-0.5
Median	64.15	61.4	-0.1
Third Quartile	83.075	84.1	0.3
Maximum	98.6	99.1	10.2

This report gives descriptive statistics about the variables used in the regression.

Kendall's Tau Correlation Confidence Interval and Hypothesis Test

Kendall's Tau Correlation	Lower 95%: Conf. Limit of Tau	Upper 95%: Conf. Limit of Tau	Z-Value for Testing H0: $\rho = 0$	P-Value	Reject H0 that $\rho = 0$ at $\alpha = 0.05$?
0.9724	0.9546	0.9833	7.5289	0.0000	Yes

This procedure assumes that Kendall's Tau is significantly greater than zero.
This assumption cannot be rejected.

The section reports an analysis of Kendall's tau correlation. It provides both a confidence interval and a significance test. The main thing to look for here is that the correlation (0.9724 in this example) is large and positive. This can be surmised from both the confidence interval and the p-value.

Cusum Test of Linearity

Maximum Cusum	Linearity Z Value	Linearity P-Value	Reject Linearity at $\alpha = 0.05$?
3.0000	0.7500	> 0.3	No

This procedure assumes that only a positive linear relationship exists between the two measurements.
This assumption cannot be rejected.

The section reports the results of the cusum test of linearity. When linearity is rejected, the assumption that the data supports this assumption is rejected and the analysis should not be used.

Passing-Bablok Regression for Method Comparison

Regression Coefficients

Regression Coefficients

Item	Intercept B0	Slope B1
Regression Coefficient	-0.0092	0.9986
Lower 95% Conf. Limit of $\beta(i)$	-0.7486	0.9856
Upper 95% Conf. Limit of $\beta(i)$	0.7987	1.0099

Method Comparison Hypothesis Test

Hypothesis being tested, $H_0: \beta_0.LCL < 0 < \beta_0.UCL$ and $\beta_1.LCL < 1 < \beta_1.UCL$.
 Cannot reject H_0 . There is not enough evidence to conclude that the methods are not equal.

This section reports the regression coefficients, along with their analytic confidence limits. The conclusion is reported on the last line. **This is the main conclusion desired of this report.**

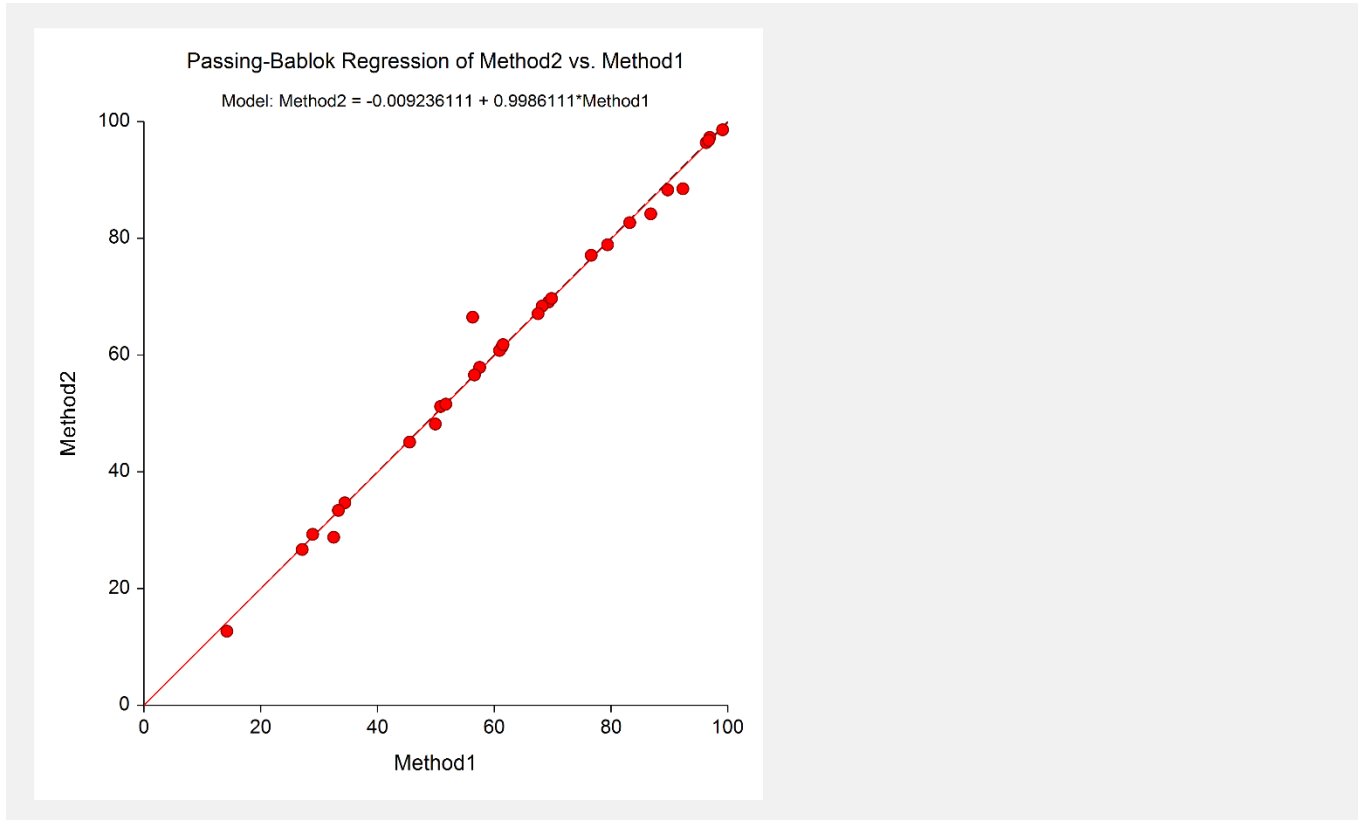
Residuals

Residuals

Row	X	Y	Difference Y - X	Predicted Yhat	Residual (Y-Yhat)
1	69.3	69.1	-0.2000	69.1945	0.0945
2	27.1	26.7	-0.4000	27.0531	0.3531
3	61.3	61.4	0.1000	61.2056	-0.1944
4	50.8	51.2	0.4000	50.7202	-0.4798
5	34.4	34.7	0.3000	34.3430	-0.3570
6	92.3	88.5	-3.8000	92.1626	3.6626
7	57.5	57.9	0.4000	57.4109	-0.4891
8	45.5	45.1	-0.4000	45.4276	0.3276
.
.
.

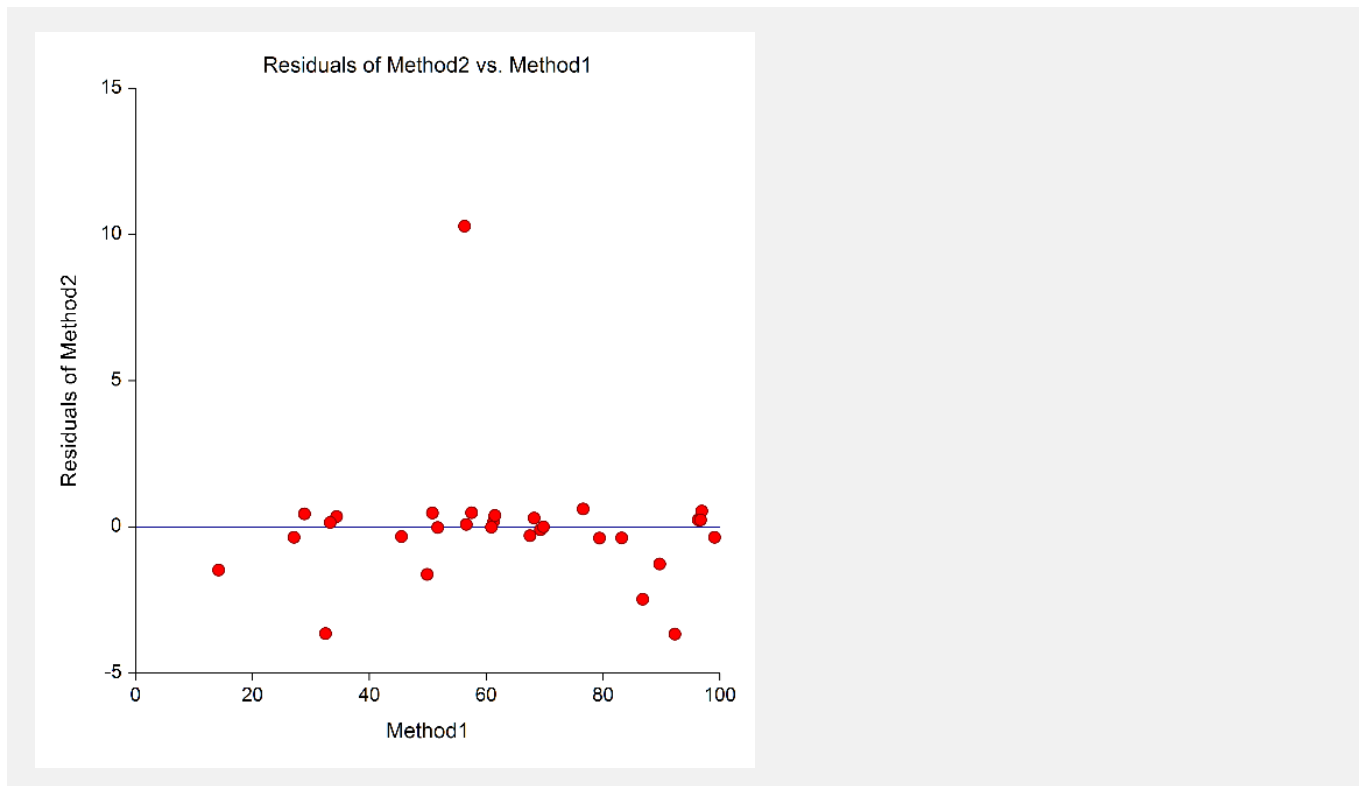
This section reports the residuals and the difference for each of the input data points.

Passing-Bablok Regression for Method Comparison

Passing-Bablok Regression Scatter Plot

This report shows both the 45-degree line ($Y = X$) and the fitted Passing-Bablok regression line. In this example, the regression line hides the 45-degree line.

Residual Plot



This plot emphasizes the deviation of the points from the regression line.