# Method Comparison Studies with R

Kevin O'Brien

**Section 3: Additional Remarks on Bland-Altman Analysis**

- Popularity of the Bland-Altman Method
- Treatment of Outliers
- Regression Based Calculation of LoA
- Bartko's Ellipse
- Coefficient of Repeatability

# The Bland-Altman Plot: Prevalence

- Limits of Agreement are used extensively in medical literature for assessing agreement between two methods.
- According to Google Scholar, Bland and Altman's 1986 paper has 22,456 citations.
  ("The Pricing of Options and Corporate Liabilities" by Black and Scholes has 19,019 citations.)

# Popularity of the Bland-Altman Method

- This methodology, now commonly known as the ‘Bland-Altman Plot’, has proved very successful. BA86, which further develops the methodology, was found to be the sixth most cited paper of all time by the BAcite.
- Dewitte also commented on the rate at which prevalence of the Bland-Altman plot has developed in scientific literature.
- The Bland-Altman Plot has since become expected, and often obligatory, approach for presenting method comparison studies in many scientific journals hollis. Furthermore
- BritHypSoc recommend its use in papers pertaining to method comparison studies for the journal of the British Hypertension Society.

- Application of regression techniques to the Bland-Altman plot, and subsequent formal testing for the constant variability of differences is informative.
- The data set may be divided into two subsets, containing the observations wherein the difference values are less than and greater than the inter-method bias respectively.
- For both of these fits, hypothesis tests for the respective slopes can be performed. While both tests can be considered separately, multiple comparison procedures, such as the Benjamini-Hochberg BH test, should be also be used.

# Outliers with Bland-Altman Plots

- The Bland-Altman plot also can be used to identify outliers. An outlier is an observation that is conspicuously different from the rest of the data that it arouses suspicion that it occurs due to a mechanism, or
conditions, different to that of the rest of the observations.
- Classification of outliers can be determined with numerous established approaches, such as the Grubb's test, but always classification must be informed by the logic of the data's formulation. Figure 1.6 is a Bland-Altman plot with two potential outliers.

# Outliers with Bland-Altman Plots

BA99 do not recommend excluding outliers from analyses, but remark that recalculation of the inter-method bias estimate, and further calculations based upon that estimate, are useful for assessing the influence of outliers. The authors remark that ‘we usually find that this method of analysis is not too sensitive to one or two large outlying differences’.

# Outliers with Bland-Altman Plots

In classifying whether a observation from a univariate data set is an outlier, Grubbs' outlier test is widely used. In assessing whether a co-variate in a Bland-Altman plot is an outlier, this test is useful when applied to the difference values treated as a univariate data set. For Grubbs' data, this outlier test is carried out on the differences, yielding the following results.

# Outliers with Bland-Altman Plots

The null and alternative hypotheses is the absence and presence of at least one outlier respectively. Grubbs' outlier test statistic $G$ is the largest absolute deviation from the sample mean divided by the standard deviation of the differences. For the 'F vs C' comparison, $G = 3.6403$. The critical value is calculated using Student's $t$ distribution and the sample size,

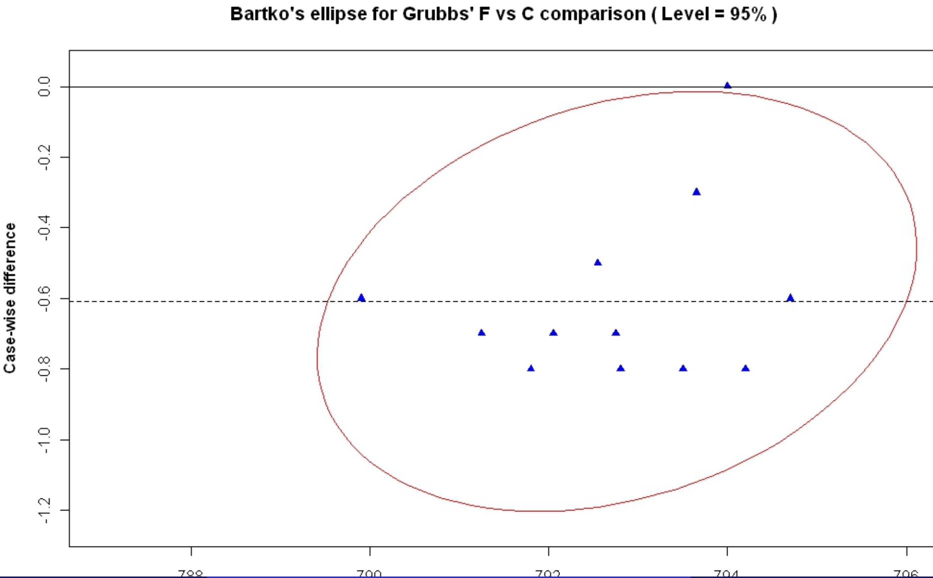
$$U = \frac{n-1}{\sqrt{n}} \sqrt{\frac{t^2_{\alpha/(2n),n-2}}{n-2+t^2_{\alpha/(2n),n-2}}}. \quad (1)$$

# Outliers with Bland-Altman Plots

For this test $U = 0.7501$. The conclusion of this test is that the fourth observation in the 'F vs C' comparison is an outlier, with $p - value = 0.002799$.

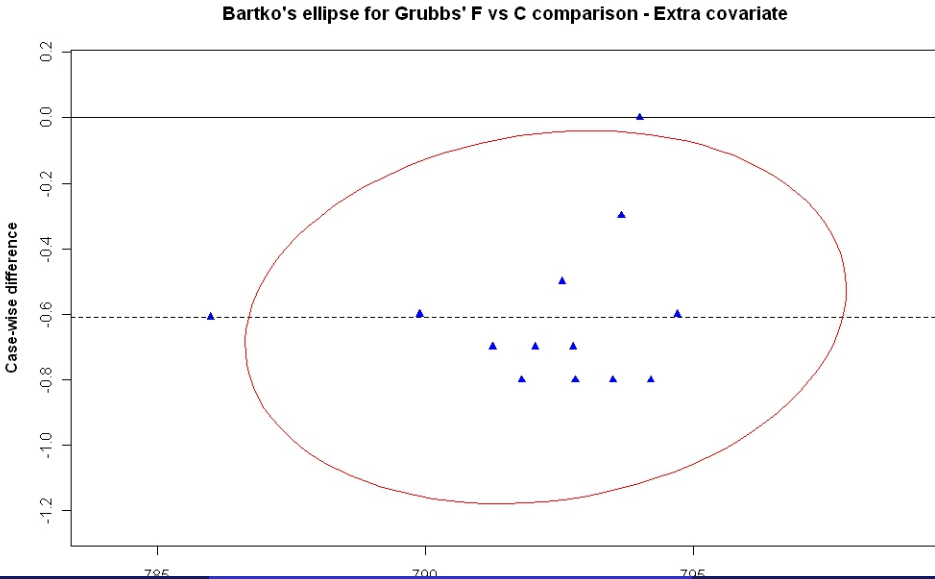
As a complement to the Bland-Altman plot, Bartko proposes the use of a bivariate confidence ellipse, constructed for a predetermined level. The minor axis relates to the between subject variability, whereas the major axis relates to the error mean square, with the ellipse depicting the size of both relative to each other.

AltmanEllipse provides the relevant calculations for the ellipse. Bartko states that the ellipse can, inter alia, be used to detect the presence of outliers (furthermore Bartko proposes formal testing procedures, that shall be discussed in due course). Inspection of Figure 1.7 shows that the fourth observation is outside the bounds of the ellipse, concurring with the conclusion that it is an outlier.


Bartko's ellipse for Grubbs' F vs C comparison ( Level = 95% )
Case-wise difference
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2

The limitations of using bivariate approaches to outlier detection in the Bland-Altman plot can demonstrated using Bartko's ellipse. A co-variate is added to the 'F vs C' comparison that has a difference value equal to the inter-method bias, and an average value that markedly deviates from the rest of the average values in the comparison, i.e. 786.

Table 1.8 depicts a 95% confidence ellipse for this enhanced data set. By inspection of the confidence interval, a conclusion would be reached that this extra co-variate is an outlier, in spite of the fact that this observation is consistent with the intended conclusion of the Bland-Altman plot.


Bartko's ellipse for Grubbs' F vs C comparison - Extra covariate
Case-wise difference
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2

In the Bland-Altman plot, the horizontal displacement of any observation is supported by two independent measurements. Any observation should not be considered an outlier on the basis of a noticeable horizontal displacement from the main cluster, as in the case with the extra co-variate. Conversely, the fourth observation, from the original data set, should be considered an outlier, as it has a noticeable vertical displacement from the rest of the observations.

Bartko's ellipse provides a visual aid to determining the relationship between variances. If var($a_i$) is greater than var($d_i$), the orientation of the ellipse is horizontal. Conversely if var($a_i$) is less than var($d_i$), the orientation of the ellipse is vertical.

Referring to the assumption that bias and variability are constant across the range of measurements, BA99 address the case where there is an increase in variability as the magnitude increases. They remark that it is possible to ignore the issue altogether, but the limits of agreement would wider apart than necessary when just lower magnitude measurements are considered.

Conversely the limits would be too narrow should only higher magnitude measurements be used. To address the issue, they propose the logarithmic transformation of the data. The plot is then formulated as the difference of paired log values against their mean. Bland and Altman acknowledge that this is not easy to interpret, and may not be suitable in all cases.

BA99 offers two variations of the Bland-Altman plot that are intended to overcome potential problems that the conventional plot would inappropriate for. The first variation is a plot of casewise differences as percentage of averages, and is appropriate when there is an increase in variability of the differences as the magnitude increases.

The second variation is a plot of casewise ratios as percentage of averages. This will remove the need for log transformation. This approach is useful when there is an increase in variability of the differences as the magnitude of the measurement increases. Eksborg proposed such a ratio plot, independently of Bland and Altman. Dewitte commented on the reception of this article by saying ‘Strange to say,this report has been overlooked’.

Assuming that there will be no curvature in the scatter-plot, the methodology regresses the difference of methods ($d$) on the average of those methods ($a$) with a simple intercept slope model; $\hat{d} = b_0 + b_1 a$. Should the slope $b_1$ be found to be negligible, $\hat{d}$ takes the value $\bar{d}$. The next step to take in calculating the limits is also a regression, this time of the residuals as a function of the scale of the measurements, expressed by the averages $a_i$; $\hat{R} = c_0 + c_1 a_i$

With reference to absolute values following a half-normal distribution with mean $\sigma\sqrt{\frac{2}{\pi}}$, BA99 formulate the regression based limits of agreement as follows

$$\hat{d} \pm 1.96\sqrt{\frac{\pi}{2}}\hat{R} = \hat{d} \pm 2.46\hat{R} \tag{2}$$

# Formal Methods

**Section 3 - Formal Methods**

-
-

# Formal Methods

- The approach proposed by BA83 is a formal test on the Pearson correlation coefficient of case-wise differences and means ($\rho_{ad}$).
- According to the authors, this test is equivalent to the 'Pitman Morgan Test'. For the Grubbs data, the correlation coefficient estimate ($r_{ad}$) is 0.2625, with a 95% confidence interval of (-0.366, 0.726) estimated by Fishers 'r to z' transformation Cohen.

# Formal Methods

- The null hypothesis ($\rho_{ad}$ =0)would fail to be rejected. Consequently the null hypothesis of equal variances of each method would also fail to be rejected. There has no been no further mention of this particular test in BA86, although BA99 refers to Spearman's rank correlation coefficient.
- BA99 comments 'we do not see a place for methods of analysis based on hypothesis testing'. BA99 also states that consider structural equation models to be inappropriate.

# Formal Methods

DunnSEME highlights an important issue regarding using models such as these, the identifiability problem. This comes as a result of there being too many parameters to be estimated. Therefore assumptions about some parameters, or estimators used, must be made so that others can be estimated. For example $\alpha$ may take the value of the inter-method bias estimate from Bland-Altman methodology. Another assumption is that the precision ratio $\lambda = \frac{\sigma^2_\epsilon}{\sigma^2_\delta}$ may be known.

# Formal Methods

DunnSEME considers methodologies based on two methods with single measurements on each subject as inadequate for a serious study on the measurement characteristics of the methods. This is because there would not be enough data to allow for a meaningful analysis. There is, however, a contrary argument that is very difficult to get replicate observations when the measurement method requires invasive medical procedure.

# Formal Methods

DunnSEME recommends the following approach for analyzing method comparison data. Firstly he recommends conventional Bland-Altman methodology; plotting the scatterplot and the Bland-Altman plot, complemented by estimate for the limits of agreement and the correlation coefficient between the difference and the mean. Additionally boxplots may be useful in considering the marginal distributions of the observations. The second step is the calculations of summary statistics; the means and variances of each set of measurements, and the covariances.

# Formal Methods

When both methods measure in the same scale (i.e. $\beta = 1$), DunnSEME recommends the use of Grubbs estimators to estimate error variances, and to test for their equality. A test of whether the intercept $\alpha$ may be also be appropriate.