

Contents

1	Bland-Altman Methodology	2
1.1	Bland-Altman Plots	3
1.2	Limits of Agreement	4
1.2.1	Prevalence of the Bland-Altman plot	5
1.3	Bartko's Ellipse	6
1.4	Coefficient of Repeatability	8
1.4.1	Repeatability	8

Chapter 1

Bland-Altman Methodology

1.1 Bland-Altman Plots

1.2 Limits of Agreement

1.2.1 Prevalence of the Bland-Altman plot

?, which further develops the Bland-Altman methodology, was found to be the sixth most cited paper of all time by the ?. ? describes the rate at which prevalence of the Bland-Altman plot has developed in scientific literature. ? reviewed the use of Bland-Altman plots by examining all articles in the journal ‘Clinical Chemistry’ between 1995 and 2001. This study concluded that use of the Bland-Altman plot increased over the years, from 8% in 1995 to 14% in 1996, and 31-36% in 2002.

The Bland-Altman Plot has since become expected, and often obligatory, approach for presenting method comparison studies in many scientific journals (?). Furthermore ? recommend its use in papers pertaining to method comparison studies for the journal of the British Hypertension Society.

1.3 Bartko's Ellipse

As a complement to the Bland-Altman plot, ? proposes the use of a bivariate confidence ellipse, constructed for a predetermined level. ? provides the relevant calculations for the ellipse. This ellipse is intended as a visual guidelines for the scatter plot, for detecting outliers and to assess the within- and between-subject variances.

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. Consequently Bartko's ellipse provides a visual aid to determining the relationship between variances. If $\text{var}(a)$ is greater than $\text{var}(d)$, the orientation of the ellipse is horizontal. Conversely if $\text{var}(a)$ is less than $\text{var}(d)$, the orientation of the ellipse is vertical.

The Bland-Altman plot for the Grubbs data, complemented by Bartko's ellipse, is depicted in Figure 1.7. The fourth observation is shown to be outside the bounds of the ellipse, indicating that it is a potential outlier.

The limitations of using bivariate approaches to outlier detection in the Bland-Altman plot can demonstrated using Bartko's ellipse. A covariate 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 manipulated data set. By inspection of the confidence interval, a conclusion would be reached that this extra covariate is an outlier, in spite of the fact that this observation is wholly consistent with the conclusion of the Bland-Altman plot.

Importantly, outlier classification must be informed by the logic of the data's formulation. 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 covariate. 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.

In classifying whether a observation from a univariate data set is an outlier, many formal tests are available, such as the Grubbs test for outliers. In assessing whether a covariate in a Bland-Altman plot is an outlier, this test is useful when applied to the case-wise difference values treated as a univariate data set. The null hypothesis of the Grubbs test procedure is the absence of any outliers in the data set. Conversely, the alternative hypotheses is that there is at least one outlier present.

The test statistic for the Grubbs test (G) is the largest absolute deviation from the sample mean divided by the standard deviation of the differences,

$$G = \max_{i=1, \dots, n} \frac{|d_i - \bar{d}|}{S_d}.$$

For the 'F vs C' comparison it is the fourth observation gives rise to the test

statistic, $G = 3.64$. The critical value is calculated using Student's t distribution and the sample size,

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

For this test $U = 0.75$. The conclusion of this test is that the fourth observation in the 'F vs C' comparison is an outlier, with p -value = 0.003, according with the previous result using Bartko's ellipse.

1.4 Coefficient of Repeatability

1.4.1 Repeatability

Barnhart emphasizes the importance of repeatability as part of an overall method comparison study. Before there can be good agreement between two methods, a method must have good agreement with itself. The coefficient of repeatability, as proposed by ? is an important feature of both Carstensen's and Roy's methodologies. The coefficient is calculated from the residual standard deviation (i.e. $1.96 \times \sqrt{2} \times \sigma_m = 2.83\sigma_m$).

The coefficient of repeatability is a measure of how well a measurement method agrees with itself over replicate measurements (?). Once the within-item variability is known, the computation of the coefficients of repeatability for both methods is straightforward.