Some Key Concepts

- In clinical measurement comparison of a new measurement technique with an established one is often needed to see whether they ***agree*** sufficiently for the new to replace the old. [2]
- A method comparison study is a study where two methods of quantitative measurement are compared by measuring the same set of items with both methods.

Bland-Altman Plots

- Historically comparison of two methods of measurement was carried out by use of paired sample t-test, correlation coefficients or simple linear regression.
- Statisticians Martin Bland and Douglas Altman recognized the inadequacies of these analyses and articulated quite thoroughly the basis on which of which they are unsuitable for comparing two methods of measurement [1].

Using Bland-Altman Plots

Bland-Altman plots are a powerful graphical methodology for making a visual assessment of the data. Bland and Altman [1] express the motivation for this plot thusly:

> *“From this type of plot it is much easier to assess the magnitude of disagreement (both error and bias), spot outliers, and see whether there is any trend, for example an increase in (difference) for high values. This way of plotting the data is a very powerful way of displaying the results of a method comparison study."*

Bland-Altman Plots

- Furthermore they proposed their simple methodology specifically constructed for method comparison studies.
- They acknowledge the opportunity to apply other valid, but complex, methodologies, but argue that a simple approach is preferable, especially when the results must be '*explained to non-statisticians*'.

The Bland-Altman Plot

- The Bland-Altman plot [1,2,3] is a very simple graphical method to compare two measurements techniques.
- In this approach the case-wise differences between the two methods are plotted against the corresponding case-wise averages of the two methods.
- As the objective of the Bland-Altman plot is to advise on the agreement of two methods, it is the case-wise differences that are particularly important.
- A horizontal lines is drawn at the mean difference(the inter-method bias) , and at the ***limits of agreement***, which are defined as the inter-method bias plus and minus 2 times the standard deviation of the differences.

- For creating plots, the case wise-averages fulfil several functions, such as expressing the range over which the values were taken, and assessing whether the assumptions of constant variance holds.
- Case-wise averages also allow the case-wise differences to be presented on a two-dimensional plot, with better data visualization qualities than a one dimensional plot.
- *Bland & Altman [2] cautions that it would be the difference against either measurement value instead of their average , as the difference relates to both value.*

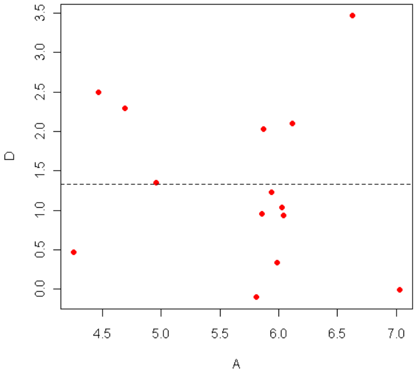
- The magnitude of the ***inter-method bias*** between the two methods is simply the average of the differences $\bar{d}$.
- The variances around this bias is estimated by the standard deviation of the differences $S(d)$.
- This inter-method bias is represented with a line on the Bland-Altman plot.
- These estimates are only meaningful if there is uniform inter-bias and variability throughout the range of measurements, which can be checked by visual inspection of the plot.

Bland-Altman Plot

```
>X = rnorm(14,6,1);Y = rnorm(14,5.3,1.1)
>
>A=(X+Y)/2 #case-wise averages
>D=X-Y #case-wise differences
>
>Dbar=mean(D) #inter-method bias
>SdD=sd(D) #standard deviation of the differences
>
>plot(A,D,pch=16,col="red", ylim=c(-3,3))
>
>abline(h=Dbar,lty=2)
>abline(h=(Dbar-2*SdD),lty=2)
>abline(h=(Dbar+2*SdD),lty=2)
```

Simple Bland-Altman Plot

Inter-method Bias : 0.45 | Limits of Agreement: [-1.32, 2.23]



Round	Fotobalk [F]	Counter [C]	Differences [F-C]	Averages [(F+C)/2]
1	793.8	794.6	-0.8	794.2
2	793.1	793.9	-0.8	793.5
3	792.4	793.2	-0.8	792.8
4	794.0	794.0	0.0	794.0
5	791.4	792.2	-0.8	791.8
6	792.4	793.1	-0.7	792.8
7	791.7	792.4	-0.7	792.0
8	792.3	792.8	-0.5	792.5
9	789.6	790.2	-0.6	789.9
10	794.4	795.0	-0.6	794.7
11	790.9	791.6	-0.7	791.2
12	793.5	793.8	-0.3	793.6

Table: Fotobalk and Counter methods: differences and averages.

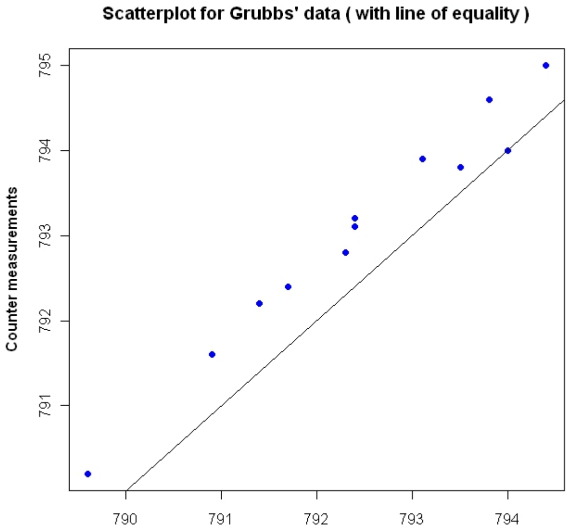
Round	Fotobalk [F]	Terma [T]	Differences [F-T]	Averages [(F+T)/2]
1	793.80	793.20	0.60	793.50
2	793.10	793.30	-0.20	793.20
3	792.40	792.60	-0.20	792.50
4	794.00	793.80	0.20	793.90
5	791.40	791.60	-0.20	791.50
6	792.40	791.60	0.80	792.00
7	791.70	791.60	0.10	791.65
8	792.30	792.40	-0.10	792.35
9	789.60	788.50	1.10	789.05
10	794.40	794.70	-0.30	794.55
11	790.90	791.30	-0.40	791.10
12	793.50	793.50	0.00	793.50

Table: Fotobalk and Terma methods: differences and averages.

Grubbs Data - Identity Plot

- Notwithstanding previous remarks about regression, the first step recommended, which the authors argue should be mandatory, is construction of a simple scatter plot of the data.
- The ***line of equality*** must also be shown, as it is necessary to give the correct interpretation of how both methods compare.
- A scatter plot of the Grubbs data (forthcoming example) is shown in Figure 1.1.
- Visual inspection confirms the previous conclusion that there is an inter-method bias present, i.e. Fotobalk device has a tendency to record a lower velocity.

Grubbs Data - Identity Plot



Grubbs Example

- In the case of Grubbs data the inter-method bias is -0.61 metres per second, and is indicated by the dashed line on Figure 1.2.
- By inspection of the plot, it is also possible to compare the precision of each method.
- Noticeably the differences tend to increase as the averages increase.

Next Slide

- The Bland-Altman plot for comparing the ‘Fotobalk’ and ‘Counter’ methods, which shall henceforth be referred to as the ‘F vs C’ comparison, is depicted in Figure 1.2, using data from Table 1.3.
- The presence and magnitude of the inter-method bias is indicated by the dashed line.

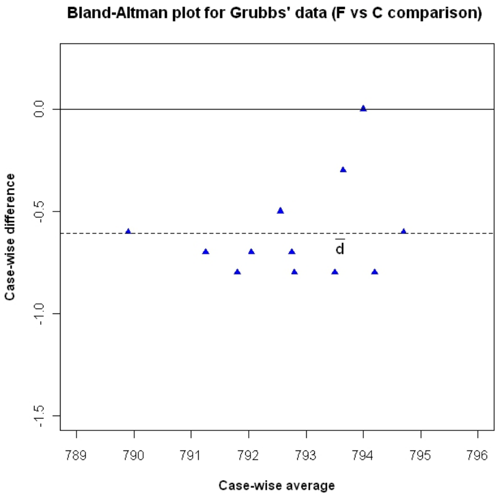
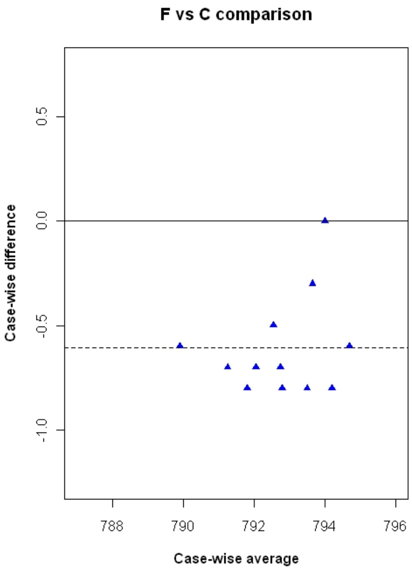
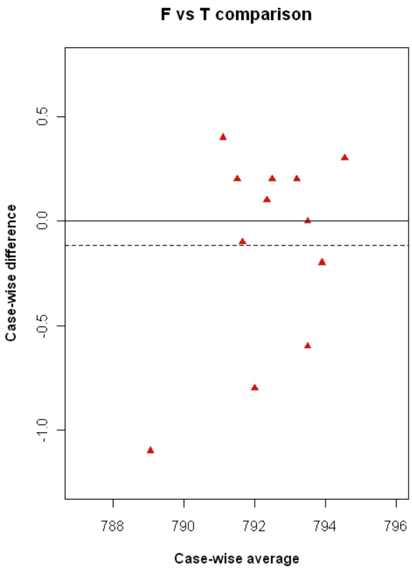


Figure: Bland-Altman plot For Fotobalk and Counter methods.

Next Slide

In the next slide Bland-Altman plots for the ‘F vs C’ and ‘F vs T’ comparisons are shown, where ‘F vs T’ refers to the comparison of the ‘Fotobalk’ and ‘Terma’ methods. Usage of the Bland-Altman plot can be demonstrate in the contrast between these comparisons.


F vs C comparison
Case-wise difference
Case-wise average
0.5
0.0
-0.5
-1.0
788
790
792
794
796


F vs T comparison
Case-wise difference
Case-wise average
0.5
0.0
-0.5
-1.0
788
790
792
794
796

- By inspection, there exists a larger inter-method bias in the ‘F vs C’ comparison than in the ‘F vs T’ comparison.
- Conversely there appears to be less precision in‘F vs T’ comparison, as indicated by the greater dispersion of co-variates.

Interpreting Bland-Altman Plots

- In the following slides are some prototype Bland-Altman plots derived from simulated data, each for the purpose of demonstrating how the plot would inform an analyst of features that would adversely affect use of the recommended methodology.

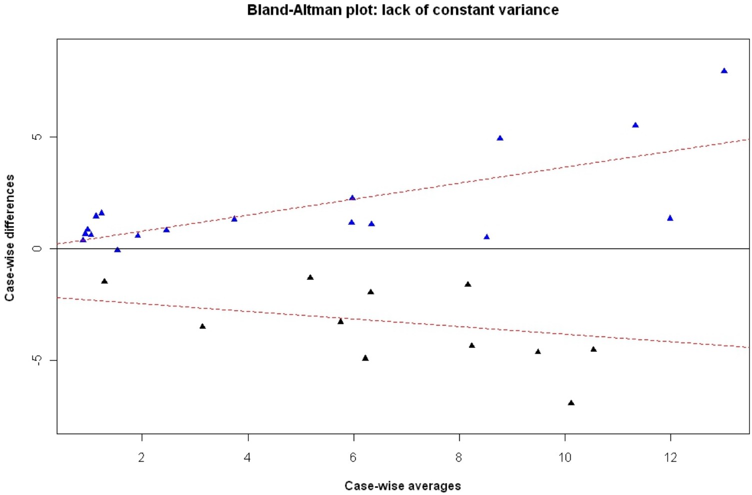


Figure: Bland-Altman plot demonstrating the increase of variance over the range.

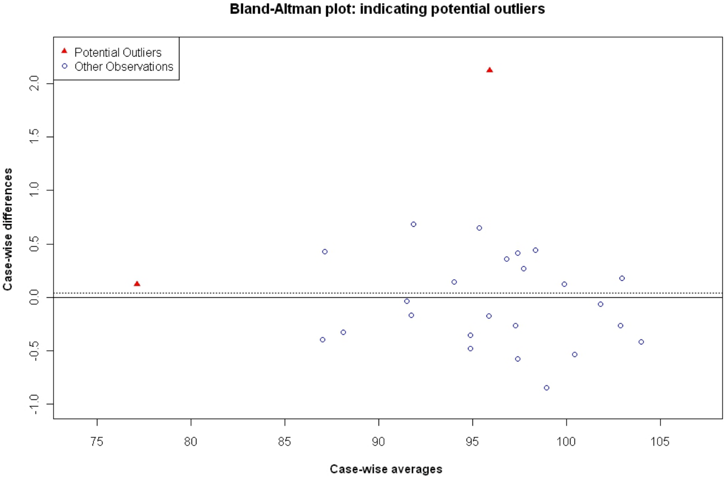


Figure: Bland-Altman plot indicating the presence of potential outliers.

Live Demonstration

Live Demonstration of Method Comparison Analysis on Grubbs Data with

- Agreement
- MethComp

Agreement

Agreement
Statistical Tools for Measuring Agreement

This package computes several statistics for measuring agreement, for example, mean square deviation (MSD), total deviation index (TDI) or concordance correlation coefficient (CCC).

It can be used for both continuous data and categorical data for multiple raters and multiple readings cases.

Author	BYue Yu AND Lawrence Lin
Maintainer	Yue Yu <yyu at imyy.net>
License	GPL-2
URL	http://imyy.net

MethComp

MethComp:
Functions for analysis of agreement in method comparison studies

Methods (standard and advanced) for analysis of agreement between measurement methods.

Author	Bendix Carstensen, Lyle Gurrin, Claus Ekstrom, Michal Figurski
Maintainer	Bendix Carstensen <bxc at steno.dk>
License	GPL-2 | GPL-3
URL	http://BendixCarstensen.com/MethComp/

Replicate Measurements

- Bland and Altman's approach originally devised for a single measurement on each item by each of the methods.

- Their 1999 paper [3] extended their approach to replicate measurements:
  By replicates we mean two or more measurements on the same individual taken in identical conditions.
  In general this requirement means that the measurements are taken in quick succession.

- Emphasis put on "repeatability".

Limits of Agreement with Replicates

- Bland Altman 1999 addresses the issue of computing LoAs in the presence of replicate measurements, suggesting several computationally simple approaches. When repeated measures data are available, it is desirable to use all the data to compare the two methods.
- However, the original Bland-Altman method was developed for two sets of measurements done on one occasion (i.e. independent data), and so this approach is not suitable for replicate measures data.
- Bland and Altman [3] address this problem by offering two different approaches.

- The premise of the first approach is that replicate measurements can be treated as independent measurements.
- The second approach is based upon using the mean of the each group of replicates as a representative value of that group.
- Using either of these approaches will allow an analyst to estimate the inter method bias.
- Carstensen et al [4,5] and Roy [6] computes the limits of agreement to the case with replicate measurements by using LME models.

Limits of Agreement with Replicates

- Carstensen et al [5] takes issue with the limits of agreement based on mean values, in that they can only be interpreted as prediction limits for difference between means of repeated measurements by both methods, as opposed to the difference of all measurements. Incorrect conclusions would be caused by such a misinterpretation.
- Carstensen et al [5] demonstrates how the limits of agreement calculated using the mean of replicates are ‘much too narrow as prediction limits for differences between future single measurements’.

Limits of Agreement with Replicates

- In computing limits of agreement, it is first necessary to have an estimate for the standard deviations of the differences.
- When the agreement of two methods is analyzed using LME models, a clear method of how to compute the standard deviation is required.
- Different approaches in fitting LME models will yield different results.
- As the estimate for inter-method bias and the quantile would be the same under various approaches, the focus is solely on the standard deviation.

Limits of Agreement

- Carstensen et al [5] computes the limits of agreement to the case with replicate measurements by using LME models.
- Roy [6] formulates a very powerful method of assessing whether two methods of measurement, with replicate measurements, also using LME models.
- Roy's approach is based on the construction of variance-covariance matrices.
- Importantly, Roy's approach does not directly address the issue of limits of agreement (although another related analysis , the *Coefficient of Repeatability*, is mentioned).