January 27, 2015

Abstract

Krouwer and Monti (29) presented a graphical method for evaluation of laboratory assays (a mountain plot). They computed the percentile for each ranked difference between the two methods, and by turning at the 50th percentile produced a histogram-like function (the mountain). This method is relevant for detecting large infrequent errors (differences) but lacks the aspect of concentration relationship. These investigators, therefore, recommend use of their plot together with difference plots. Introduction of analytical quality specifications in the mountain plots may be useful in method evaluations.

Davis (1989) proposes the TAM model, which suggests an hypothesis as to why users may adopt particular technologies, and not others. According to this theory, when users are presented with a new technology, two important factors will influence their decision about how and when they will adopt it.

Perceived usefulness (PU) - This was defined by Fred Davis as "the degree to which a person believes that using a particular system would enhance his or her job performance".

Perceived ease-of-use (PEOU) - Davis defined this as "the degree to which a person believes that using a particular system would be free from effort"

Davis's explanations of these term can be rephrased for application to statistical analysis. Perceived Use could refer to the degree to which an user would deem a particular statistical method would properly establish the results of an analaysis. In the case of method comparison studies, proper indication of agreement, or lack thereof.

Perceived ease-of-use requires only applying the context of a

A very modest statistical skill set is the only prerequistive for constructing a Bland-Altman plot, and computing limits of agreement. The main building blocks are simple descriptive, statistics and a knowledge of the normal distribution. These are topics that feature in almost every undergraduate statistics courses.

In short, the user perceives the Bland-Altman methodology to be an easy-to-implement technique, that will properly address the question of agreement.

Conversely the Survival plot is a derivative of the Kaplan-Meier Curve, a non-parametric graphical technique that features in Survival Analysis. This

subject area is a well known domain of statistics, but would be encountered on curriculums of specialist courses. The Mountain Plot is formally called the empirical folder cumulative distribution plot. Currently there is only one software implementation , medcalc.be toolkot (FIX)

The ROC curve is a plot that is commonly used in the appraisal of a statistical analytics systems. Interpretation of the plot, the nearer the curve is to the top left corner of the plot, the better the statistical method is at making predicting outcomes.

The addition of an extra factor

Interaction terms are featured in ANOVA designs.

My search just now found no mention of Cook's distance or influence measures.

The closest I found was an unanswered question on this from April 2003 (http://finzi.psych.upenn.edu/R/Rhelp02a/archive/4797.html).

Beyond that, there is an excellent discussion of "Examining a Fitted Model" in Sec. 4.3 (pp. 174-197) of Pinheiro and Bates (2000) Mixed-Effects Models in S and S-Plus (Springer).

Pinheiro and Bates decided NOT to include plots of Cook's distance among the many diagnostics they did provide. However, 'plot(fit.lme)' plots 'standardized residuals' vs. predicted or 'fitted values'. Wouldn't points with large influence stand apart from the crowd in terms of 'fitted value'?

Of course, there are many things other one could do to get at related information, including reading the code for 'influence' and 'lme', and figure out from that how to write an 'influence' method for an 'lme' object.

Lai et Shiao is interesting in that it extends the usual method comparison study question. It correctly identifies LME models as a methodoloy that can used to make such questions tractable. The Data Set used in their examples is unavailable for independent use. Therefore, for the sake of consistency, a data set will be simulated based on the Blood Data that will allow for extra variables.

A Study of the Bland-Altman Plot and its Associated Methodology
Joseph G. Voelkel Bruce E. Siskowski

0.1 Limits of agreement for Carstensen's data

bxc2008 describes the calculation of the limits of agreement (with the inter-method bias implicit) for both data sets, based on his formulation;

$$\hat{\alpha}_1 - \hat{\alpha}_2 \pm 2\sqrt{2\hat{\tau}^2 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2}.$$

For the 'Fat' data set, the inter-method bias is shown to be 0.045. The limits of agreement are $(-0.23, 0.32)$

Carstensen demonstrates the use of the interaction term when computing the limits of agreement for the 'Oximetry' data set. When the interaction term

is omitted, the limits of agreement are $(-9.97, 14.81)$. Carstensen advises the inclusion of the interaction term for linked replicates, and hence the limits of agreement are recomputed as $(-12.18, 17.12)$.

0.2 Limits of Agreement in LME models

bxc2008 uses LME models to determine the limits of agreement. Between-subject variation for method m is given by d_m^2 and within-subject variation is given by λ_m^2. BXC2008 remarks that for two methods A and B, separate values of d_A^2 and d_B^2 cannot be estimated, only their average. Hence the assumption that $d_x = d_y = d$ is necessary. The between-subject variability D and within-subject variability Λ can be presented in matrix form,

$$D = \begin{pmatrix} d_A^2 & 0 \\ 0 & d_B^2 \end{pmatrix} = \begin{pmatrix} d^2 & 0 \\ 0 & d^2 \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \lambda_A^2 & 0 \\ 0 & \lambda_B^2 \end{pmatrix}.$$

The variance for method m is $d_m^2 + \lambda_m^2$. Limits of agreement are determined using the standard deviation of the case-wise differences between the sets of measurements by two methods A and B, given by

$$\text{var}(y_A - y_B) = 2d^2 + \lambda_A^2 + \lambda_B^2. \tag{1}$$

Importantly the covariance terms in both variability matrices are zero, and no covariance component is present.

roy has demonstrated a methodology whereby d_A^2 and d_B^2 can be estimated separately. Also covariance terms are present in both D and Λ. Using Roy's methodology, the variance of the differences is

$$\text{var}(y_{iA} - y_{iB}) = d_A^2 + \lambda_B^2 + d_A^2 + \lambda_B^2 - 2(d_{AB} + \lambda_{AB}) \tag{2}$$

All of these terms are given or determinable in computer output. The limits of agreement can therefore be evaluated using

$$\bar{y_A} - \bar{y_B} \pm 1.96 \times \sqrt{\sigma_A^2 + \sigma_B^2 - 2(\sigma_{AB})}. \tag{3}$$

For Carstensen's 'fat' data, the limits of agreement computed using Roy's method are consistent with the estimates given by BXC2008; $0.044884 \pm 1.96 \times 0.1373979 = (-0.224, 0.314)$.

0.3 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 BA99 is an important feature of both Carstensen's and Roy's methodologies. The coefficient is calculated from the residual standard deviation (i.e. $1.96 \times \sqrt{2} \times \sigma_m = 2.83\sigma_m$).

1 Hamlett and Lam

The methodology proposed by Roy2009 is largely based on hamlett, which in turn follows on from lam.

1.1 Roy's variability tests

Variability tests proposed by Roy2009 affords the opportunity to expand upon Carstensen's approach.

The first test allows of the comparison the begin-subject variability of two methods. Similarly, the second test assesses the within-subject variability of two methods. A third test is a test that compares the overall variability of the two methods.

The tests are implemented by fitting a specific LME model, and three variations thereof, to the data. These three variant models introduce equality constraints that act null hypothesis cases.

Other important aspects of the method comparison study are consequent. The limits of agreement are computed using the results of the first model.

The desired outcome of this research is to

- Formulate a methodology that represents Best practice in Method Comparison Studies. Indeed the methodology is envsiaged to advance what is considered best practice, inter alia, by making diagnostics procedures a standard part of MCS.

- Provide for ease of use such that non-statisticians can master and implement the method, with a level of training that one would expect as part of a Professional CPD programe.

Apropos of the matter of ease-of-use, certain assumptions must be made.

The user has a reasonable amount of computer literacy. The user would have a reasonable understanding of statistics, consistent with an undergraduate statistics module. That is to say, that the user is acquainted with the idea of $p-$values.

Easy to follow set of instructions to properly implement the method.

Linear Mixed Effects Models can be implemented by using one of the following R packages. lme4 nlme

The first package to be introducted was nlme, developled by Jose Pinheiro and Douglas Bates (Authors of the the companion textbook, NAME)

As this package has been under ongoing development for quite a long time, it is now allows for a lot of complex LME implementations. Furthermore, nlme is one of the base R packages. That is to say, when one downloads and installs R, nlme is automatically installed also, and can be called immediately.

Having said that, the authors have pointed to several limitations of the overall methodology thrugh R. The original developers have both left the project, but other statisticians have taken over the development, and indeed a new version of nlme was released.

LME4 is a more recent package. at a glance, the syntax is easier, but the development is less advanced. There are several functionalities that can not be implemented with lme4 yet. As an example - CHAP5 in PB - has no equivalent in LME4. Indeed no textbook exists to co-incide with LME4.

The main author, Douglas Bates, has turned his attention to development of LME models in the Julia programming language.

The nlmeU package is described by its authors as an extesntion of the nlme package, and indeed provides for additionally functionality. The package is also useful as it serves as a companion piece to the book by Galecki and Burzwhatski.

The nlme package also allows for the specification of GLS models.

Objects and Classes

The main nlme object is an `nlme` model.

The main lme4 object is called an `lmer` model

The lattice package is used for graphical methods.

Model Diagnostics with `nlme`

2 Method Comparison Stduies with `R`

2.1 Accuracy and Precision

An important consideration in discussing methods of measurement are the issues of accuracy and precision.

2.2 What is Agreement

Agreement between two methods of clinical measurement can be quantified using the differences between observations made using the two methods on the same subjects. (Bland and Altman 1999)

2.3 Inappropriate Techniques for MCS

2.4 Links and Papers

`Westgard Statistics - http://www.westgard.com/lesson23.htm`

2.5 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 Bland & Altman (1999) is an important feature of both Carstensen's and Roy's methodologies. The coefficient is calculated from the residual standard deviation (i.e. $1.96 \times \sqrt{2} \times \sigma_m = 2.83\sigma_m$).

Measurement Systems Analysis

The topic of measurement sensitivity anaylysis (MSA, also known as Gauge R&R) is prevalent in industrial statistics (i.e Six Sigma).

There is extensive literature that covers the area. For the sake of brevity, we will use Cano et al.

For sake of clarity, Cano's definitions of repeatability and reproducibility are listed, with added emphasis.

Reproducibility is rarely, if ever, discussed in the domain of Method Comparison Studies. This may be due to the fact that prevalent methodologies can be used for the problem.However the methodologies proposed by this research can easily be extended.