

Range Charts for Agreement in Measurement Comparison Studies, With Application to Replicate Mass Spectrometry Experiments

James A. Koziol¹*, Anne C. Feng¹, Jingyi Yu², Noelle M. Griffin², Jan E. Schnitzer²

¹Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037

²Sidney Kimmel Cancer Center, 10905 Road to the Cure, San Diego, CA 92122

*Corresponding author: James A. Koziol, Ph.D, Department of Molecular and Experimental Medicine, MEM216 The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, E-mail: koziol@scripps.edu; Tel: 858-784-2703; Fax: 858-784-2664

Received August 02, 2008; Accepted September 02, 2008; Published September 13, 2008

Citation: James AK, Anne CF, Jingyi Y, Noelle MG, Jan ES (2008) Range Charts for Agreement in Measurement Comparison Studies, With Application to Replicate Mass Spectrometry Experiments. *J Proteomics Bioinform* 1: 287-292. doi:10.4172/jpb.1000036

Copyright: © 2008 James AK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

It is important to investigate the reproducibility of raw mass spectrometry (MS) features of abundance, such as spectral count, peptide number and ion intensity values, when conducting replicate mass spectrometry measurements. Reproducibility can be inferred from these replicate data either formally with analyses of variance techniques or informally with graphical procedures, particularly, Bland-Altman plots on paired runs. In this note, we suggest range plots to provide a suitable generalization of Bland-Altman plots to experiments with more than two replicate runs. We describe range charts and their interpretation, and illustrate their use with data from a recent proteomic study relating to label-free analysis.

Introduction

Bland and Altman (1986) introduced a useful graphical procedure for comparing two methods of measurement of a continuous biological variable. The Bland-Altman plot of differences (ordinate) versus means (abscissa) of paired measurements is simple and elegant, and has become a standard for agreement studies, especially when combined with limits of agreement. Typically, 95% limits of agreement lines are superimposed on the Bland-Altman plots; these provide an interval within which 95% of differences between future measurements by the two methods would be expected to lie.

If one wishes to compare more than two matched measurements, pairwise Bland-Altman plots can be shown. This might prove cumbersome, however, with increasing numbers of measurements being compared. Bland and Altman (1999; 2007) have proposed rigorous statistical methods for such studies, devolving from formal analyses of variance of the method comparison studies. Similarly, an informal graphical procedure is available from the quality control literature, namely, Shewhart control charts for the range (Shewhart, 1939; Montgomery, 2001). Conceptually, one plots the ranges (ordinate) versus the means (abscissa) of the matched samples for the range chart.

In this note, we describe range charts for multiply-matched comparison studies, and illustrate some techniques for imposing limits of agreement on the charts. We illustrate this method with data arising from a recent study we had undertaken, to investigate the reproducibility of raw mass spectrometry features of abundance, including spectral count, peptide number and ion intensity values, across replicate 2DLC mass spectrometry measurements.

Reference intervals

In comparison studies of two methods, the limits of agreement advocated by Bland and Altman can be interpreted as reference intervals, and delineate the range within which most differences between measurements might be expected to lie. The reference interval is generally defined by the range between two centile values of a population, centered about the median value. For example, the standard limits of a 95% reference interval would be the 2.5th and 97.5th centiles. [See Altman and Bland (1994) for some discussion of terminology regarding centiles, quantiles, and related quantities].

There are many ways of constructing reference intervals, as elegantly summarized by Wright and Royston (1999)

and Bland and Altman (2007). We here describe two simple nonparametric techniques, which make no assumptions about the underlying distributions of the data comprising the range charts, followed by a semi-parametric technique, related to underlying normality.

Harrell-Davis

Harrell and Davis (1982) introduced an elegant distribution-free quantile estimator, which may be invoked to estimate the centiles of an underlying distribution.

This approach is formulaically straightforward, but does require computational capabilities involving mathematical functions. For reference, we state Harrell and Davis's formula in the Appendix.

Bootstrap

A second nonparametric approach utilizes the bootstrap (Efron and Tibshirani, 1993), as an extension of the simple approach of estimating empirical centiles directly from the ordered observations. Draw, say, 1000 bootstrap samples of the same cardinality from the underlying observations. From each bootstrap sample, compute the desired empirical centiles, using interpolation if necessary between the nearest two order statistics. Then, average the corresponding empirical centiles over the bootstrap samples to establish the reference interval limits. This approach is computationally involved, but formulaically simple.

Box-Cox transformation procedure

Box and Cox (1964) introduced a power function that has been widely adopted to transform data to approximate normality. In the present context, one would invoke the Box-Cox procedure on the sample ranges, determine limits of agreement on the Box-Cox transformed scale, and then back-transform to the original range scale. There is some computational complexity involved in finding the Box-Cox transformation parameter [see Appendix], but determination of limits of agreement on a putatively normal scale is straightforward.

An application

We illustrate range charts and reference intervals with the following example. Yu and colleagues (2008) have undertaken an extensive study of computational and statistical considerations relating to proteomic analyses of label-free mass spectroscopy (MS) data. We utilize one dataset from their study, consisting of normalized spectral indices of 174 peptides found in normal liver tissue from three replicate MS runs. Yu et al. (2008) investigate fundamental issues of reproducibility and effects of normalization in this and other

datasets, and we make no attempt to recapitulate their analyses here. Rather, our use of range charts is meant to be complementary to their more comprehensive study.

In Figure 1, we give the range chart for the 174 replicates. Clearly, variability increases with sample mean, an important diagnostic from these plots. We may take a simple log transformation (Bland and Altman, 1986; 1996) of the spectral indices, with the intent to render the variance independent of the mean. As shown in Figure 2, the log transform goes far toward ameliorating the increasing variability with mean in the original spectral data.

We turn next to the issue of tolerance intervals. In Table 1 we give the limits of the 90% and 95% tolerance intervals for the original data as well as the log-transformed data, for the three methods described earlier. That the Harrell-Davis and bootstrap estimates are nearly coincident is unsurprising: Sheather and Marron (1990) have noted the interconnection of the two methods. The Box-Cox method is somewhat deviant from the other methods, particularly in the upper limits for the original data. Here, there are 14 values (8.0%) exceeding 211, 7 (4.5%) exceeding 386 or 396, 5 (2.9%) exceeding 537, and 4 (2.3%) exceeding 887 or 889; the nonparametric approaches appear to have coverages nearer the desired levels of 5% or 2.5% than Box-Cox. We comment that with the original data, Box-Cox was not altogether successful in achieving normality: sample skewness was reduced from 5.27 to 0.004 (target 0), but sample kurtosis was reduced from 33.32 to 3.59 (target 3). Further correction for kurtosis might be desirable. In Figures 1 and 2 we have also superimposed the nonparametric two-sided 90% tolerance intervals; again, the log transformation seems quite suitable.

Discussion

Our motivation for investigating range charts was a comprehensive study we had undertaken, to investigate the reproducibility of raw mass spectrometry (MS) features of abundance, including spectral count, peptide number and ion intensity values, across replicate 2DLC mass spectrometry measurements. We found that typically raw features were not very reproducible across replicates. We could verify this formally via analyses of variance our search for an informal graphical procedure led us first to Bland-Altman plots, and then to range plots as an appropriate extension to greater than two replicate runs. The range plots gave us immediate indication that log transformation would ameliorate the increasing variability problem. We went on to develop a more involved normalization technique which also tended to work well with our experimental data.

The reproducibility of a proteomics experiment can be viewed in two ways, 1) reproducibility in terms of the pro-

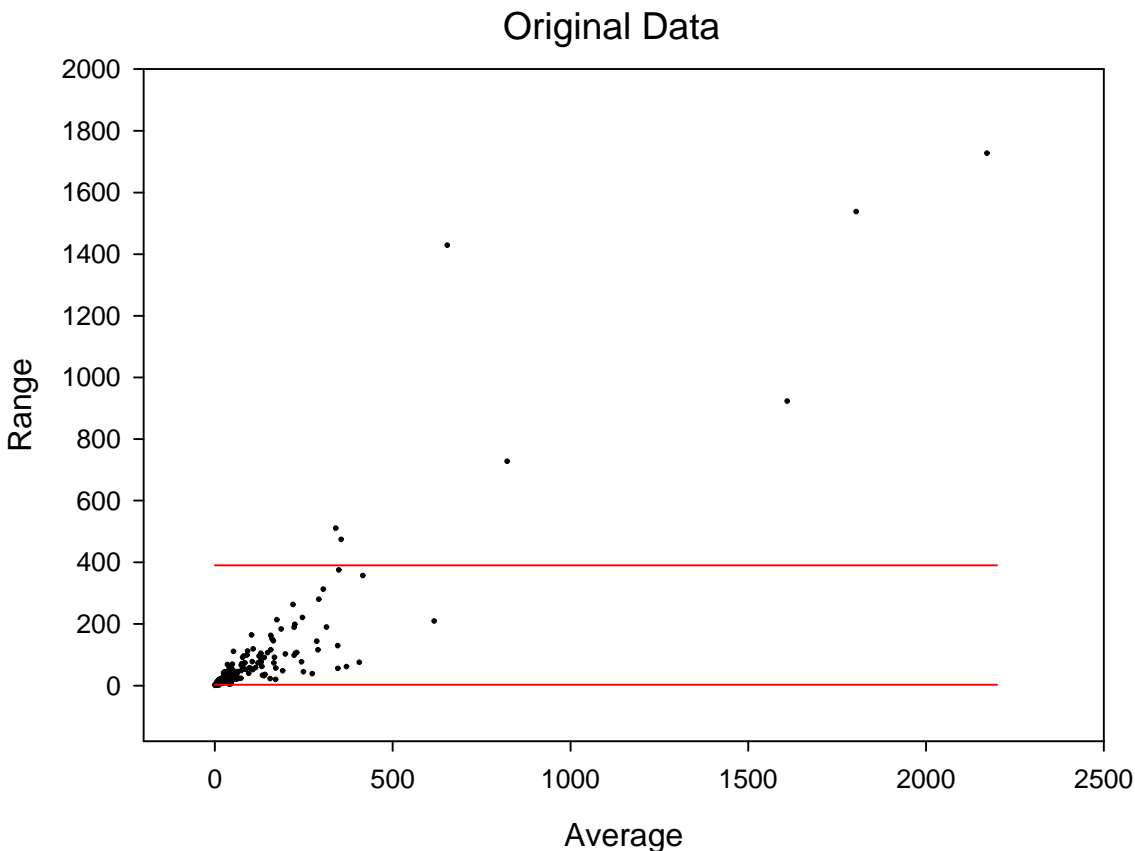


Figure 1: Range chart for spectral index measures from 174 peptides, as determined from three independent mass spectrometry runs. Red lines delineate a nonparametric 90% tolerance interval for future determinations.

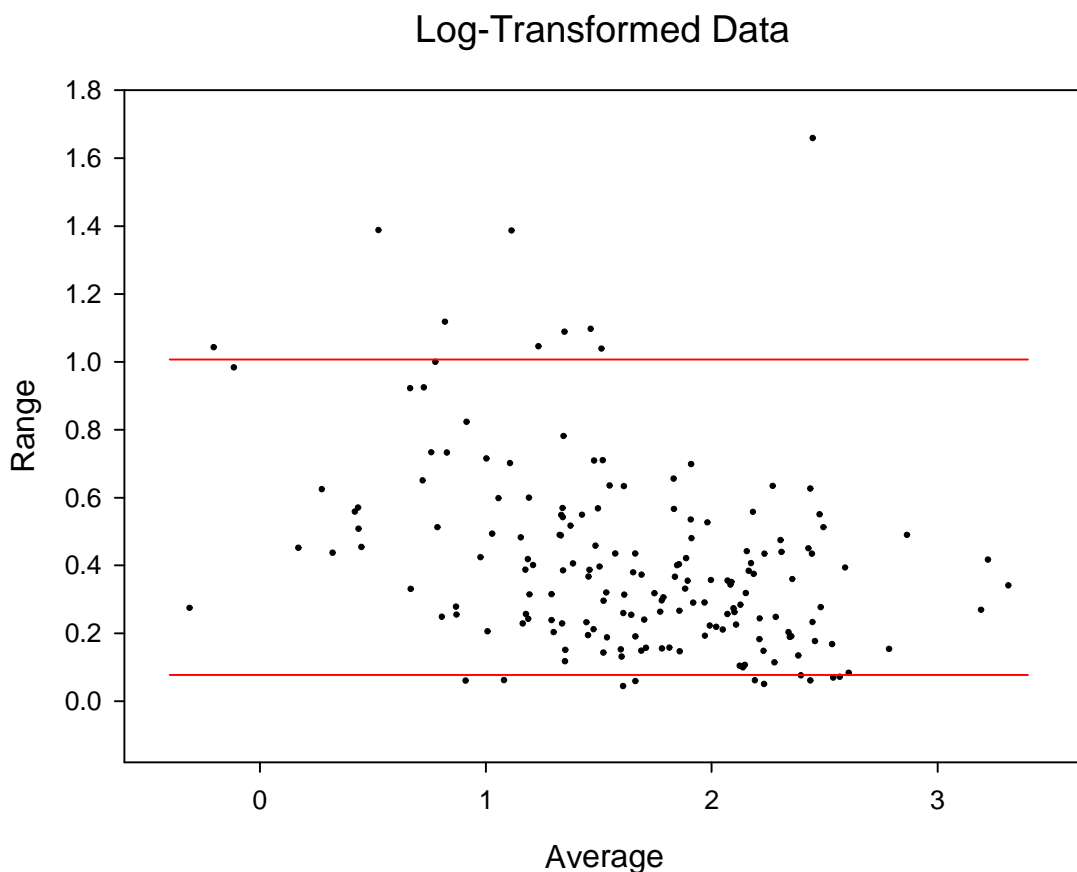


Figure 2: Range chart derived from the log-transformed data described in Figure 1, along with nonparametric 90% tolerance intervals.

Original Data

	90% Tolerance Interval		95% Tolerance Interval	
	Lower Limit	Upper Limit	Lower Limit	Upper Limit
Harrell-Davis	3.081	386.60	1.927	887.37
Bootstrap	3.104	396.42	1.945	889.80
Box-Cox	3.265	211.90	2.050	537.72

Log-Transformed Data

Harrell-Davis	0.0769	1.0070	0.0608	1.1723
Bootstrap	0.0772	1.0085	0.0608	1.1754
Box-Cox	0.0713	1.0399	0.0604	1.1099

Table 1: Tolerance Intervals for Replicate MS Data.

teins identified, i.e., are we identifying the same proteins in replicate 2 as we did in replicate 1, and 2) reproducibility in the relative quantity of these proteins detected across the replicates, i.e., is the relative abundance of protein A the same in replicate 1 as in replicate 2. In terms of protein identification, Durr et al. (2004) previously showed that any second replicate "shotgun" MS measurement will identify 30-40% of proteins not found in a single MS measurement of an identical sample, inferring an approximate 60% overlap in protein identifications between replicate measurements. This also demonstrates the requirement for multiple replicate measurements in order to achieve significant coverage of the sample of interest; we refer the interested reader to Koziol et al. (2006), where we give some practical guidelines for experimental design. Nevertheless, the Yu et al. (2008) study clearly demonstrated [both formally and with range charts] that the abundance features for these common proteins are not very reproducible across replicate MS measurements. As a result, normalization of these features (by such method as developed by Yu et al., 2008) is required to control for the variation, which in turn enhances the reproducibility of the replicates thus allowing their direct comparison.

Range charts are conceptually straightforward, and are an economical way of depicting levels of agreement among multiple measurements. Compared to pairwise Bland-Altman plots, one range chart summarizes the entire agreement study; however, signed differences revelatory of systematic large or small values in one method compared to a second method may be obscured.

There are various ways of establishing reference intervals on range charts. We have focused on nonparametric and semiparametric methods that are relatively easy to implement, and dispense with sometimes restrictive distributional assumptions attending ranges. We mention in passing that we also explored quantile estimation based on Edgeworth and Cornish-Fisher expansions (Cornish and Fisher, 1937; Stuart and Ord, 1987), but results were less satisfactory than with Box-Cox. The distribution of the range with normally distributed random variables has been extensively studied, leading to parametric approaches to establishing reference intervals; see Montgomery (2001) for details. And, one could plausibly impose one-sided limits of agreement with range charts, in settings where upper limits of agreement are of primary interest.

We remark that, with two matched measurements, the range chart displays the absolute values of the signed differences from the Bland-Altman plot. This relationship generalizes to our setting of three MS determinations, a consequence of the mathematical identity: $\text{range}\{x_1, x_2, x_3\} = \frac{1}{2} (|x_1 - x_2| + |x_2 - x_3| + |x_3 - x_1|)$, for any real numbers x_1, x_2, x_3 . That is, the range of 3 observations [as in a range chart] can be recovered from the pairwise differences [as in the Bland-Altman plots], though not conversely.

In summary, reproducibility is an important criterion for assessing proteomic experiments, and is an essential property of validation. In this regard, range charts provide immediate graphical representation of comparison studies, from which unacceptable or untoward levels of variability can be discerned. Range charts constitute straightforward assessments of method comparison and agreement, and are complementary to more formal assessments of agreement, as described by Bland and Altman (1999; 2007).

Acknowledgments

We thank the reviewers for their insightful comments. This research was supported in part by grant PO1CA164898 from the National Institutes of Health.

References

1. Altman DG, Bland JM (1994) Quartiles, quintiles, centiles, and other centiles. *BMJ* 309: 996. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
2. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1: 307-310. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
3. Bland JM, Altman DG (1996) Transforming data. *BMJ* 312: 770. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
4. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Statist Meth Med Res* 8: 135-160. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
5. Bland JM, Altman DG (2007) Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 17: 571-582. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
6. Box GEP, Cox DR (1964) An analysis of transformations (with discussion). *J Roy Statist Soc B* 26: 211-252.
7. Cornish EA, Fisher RA (1938) Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute* 5: 307-320. » [CrossRef](#) » [Google Scholar](#)
8. Durr E, Yu J, Krasinska KM, Carver LA, Yates JR, et al. (2004) Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat Biotechnol* 22: 985-992. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)

9. Efron B, Tibshirani RJ (1993) *An Introduction to the Bootstrap*. Chapman Hall, London. » [CrossRef](#) » [Google Scholar](#)
10. Horn PS, Pesce AJ, Copeland BE (1998) A robust approach to reference interval estimation and evaluation. *Clin Chem* 44: 622-631. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
11. Koziol JA, Feng AC, Schnitzer JE (2006) Application of capture-recapture models to estimation of protein count in MudPIT experiments. *Anal Chem* 78: 3203-3207. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
12. Montgomery DC (2001) *Introduction to Statistical Quality Control* 4th Ed John Wiley New York.
13. Shewhart WA (1939) (with the editorial assistance of Deming W). *Statistical Method from the Viewpoint of Quality Control* Graduate School US Department of Agriculture Washington DC.
14. Stuart A, Ord JK (1987) *Kendall's Advanced Theory of Statistics Volume I* 5th Ed Oxford University Press New York.
15. Wright EM, Royston P (1999) Calculating reference intervals for laboratory measurements. *Statist Meth Med Res* 8: 93-112.
16. Yu J, Griffin NM, Oh P, Shore S, et al. (2008) Quantification and normalization of complex label-free mass spectrometry data for proteomics analysis. submitted.

Appendix

The Harrell-Davis quantile estimator

Harrell and Davis (1982) introduced an elegant approach to distribution-free quantile estimation from a sample of independent, identically distributed random variables $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. The Harrell-Davis estimator of the p^{th} quantile is

$$Q_p = \sum_{i=1}^n W_{n,i} X_{(i)}$$

where

$$W_{n,i} = I_{i/n}[p(n+1), (1-p)(n+1)] - I_{(i-1)/n}[p(n+1), (1-p)(n+1)];$$

here, $I_q(\alpha, \beta)$ is the incomplete beta function and

$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ are the observed order statistics of the sample.

The Box-Cox transformation procedure

Box and Cox (1964) proposed the power function

$$\begin{aligned} f(x,\lambda) &= (x^\lambda - 1)/\lambda, & \lambda \neq 0, \\ &= \log(x), & \lambda = 0. \end{aligned}$$

Box and Cox (1964) obtained a likelihood function for estimating the transformation parameter λ by assuming that the transformed variable would be normally distributed for suitable λ . In practice, the maximum likelihood estimate $\hat{\lambda}$ of λ , calculated from the Box-Cox likelihood function, is used for transformation. The function f is continuous in λ at 0, so numerical maximization is not at all problematic.