

# Method Comparison Studies with R

**Kevin O'Brien (kevin.obrien@ul.ie, University of Limerick)**

# Medical statistics



- Applications to medicine and the health sciences, including epidemiology, public health, forensic medicine, and clinical research.
- "Biostatistics" more commonly connotes all applications of statistics to biology.
- Clinical Research is main focus for this talk - Method Comparison Studies



## About CLSI

VISION, MISSION, AND VALUES

STRATEGIC PLAN

OUR BOARD OF DIRECTORS

CLSI RESOURCES

FAQS

CLSI CALENDAR OF EVENTS

CAREERS

CONTACT US

COPYRIGHT PERMISSIONS

PRIVACY POLICY

# Committed to Continually Advancing Laboratory Practices

Discover how we are striving to improve clinical laboratory testing quality.

A not-for-profit membership organization, the Clinical and Laboratory Standards Institute (CLSI) brings together the global laboratory community for a common cause: fostering excellence in laboratory medicine. We do so by facilitating a unique process of developing clinical laboratory testing standards based on input from and consensus among industry, government, and health care professionals.

For over 40 years, our members, volunteers, and customers have made CLSI a respected, transformational leader in the development and implementation of clinical laboratory testing standards. Through our unified efforts, we will continue to set and uphold the standards that drive quality test results, enhance



CLINICAL AND  
LABORATORY  
STANDARDS  
INSTITUTE®

August 2013

# EPo9-A3

**Measurement Procedure Comparison and  
Bias Estimation Using Patient Samples;  
Approved Guideline—Third Edition**

AMP



# Medical Measurement





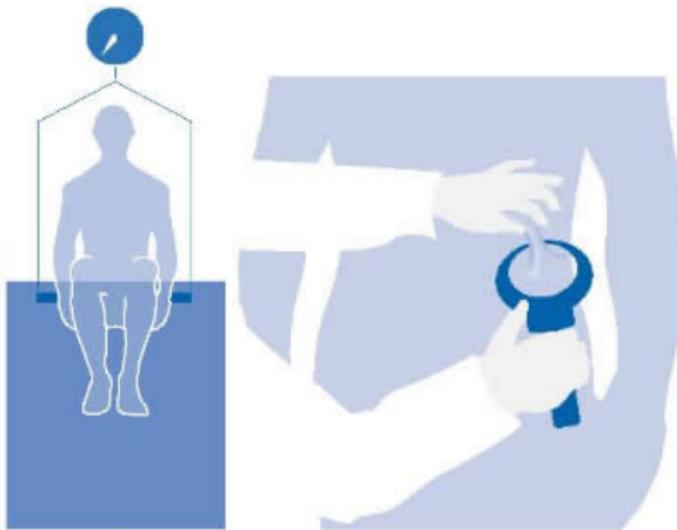
## Common Scenario for comparing two methods of measurements

**Reference Method** Very accurate, but some cost involved in getting measurement.

**Test Method** Not as accurate as reference method, but less cost involved

# Methods of Measurement

## Comparing against a Gold Standard



# Gold Standards

## Gold Standard Methods of Measurement

- Gold standard test usually refers to a diagnostic test or benchmark that is the best available under reasonable conditions.
- Other times, gold standard is used to refer to the most accurate test possible without restrictions.
- Summary: may yield value close to “True Value”, then again it may not.

*For instance, for the diagnosis of aortic dissection, the "gold standard" test used to be the aortogram, which had a sensitivity as low as 83% and a specificity as low as 87%.*

*Since the advancements of magnetic resonance imaging, the magnetic resonance angiogram (MRA) has become the new "gold standard" test for aortic dissection, with a sensitivity of 95% and a specificity of 92%.*

*Before widespread acceptance of any new test, the former test retains its status as the "gold standard."*

December 18, 2012 | By Ioana Patringenaru

# Small, Portable Sensors Allow Users to Monitor Exposure to Pollution on Their Smart Phones

Computer scientists at the University of California, San Diego have built a small fleet of portable pollution sensors that allow users to monitor air quality in real time on their smart phones. The sensors could be particularly useful to people suffering from chronic conditions, such as asthma, who need to avoid exposure to pollutants.

CitiSense is the only air-quality monitoring system capable of delivering real-time data to users' cell



*The CitiSense sensors transmit their air quality readings to smart phones. More pictures of the sensor and its smart phone interface can be found [here](#).*

# Method Comparison Studies

- Commonly encountered issue in medical statistics
  - "Do two methods of measurement agree statistically?".
  - "Can the two methods be used interchangeably?"
  - Sources of disagreement can arise from differing population means (i.e. **inter-method bias**), differing **between-subject variances** and **within subject variances** [1].

# CRAN Clinical Trials Taskview

## CRAN Task View: Clinical Trial Design, Monitoring, and Analysis

**Maintainer:** Ed Zhang and Harry G. Zhang

**Contact:** Ed.Zhang.jr at gmail.com

**Version:** 2014-12-07

This task view gathers information on specific R packages for design, monitoring and analysis of clinical trials. It focuses on including packages for clinical trial design and monitoring in general, plus data analysis packages for a specific type of design. Also, it gives a brief introduction to important packages for analyzing clinical trial data. Please refer to task views [ExperimentalDesign](#), [Survival](#), [Pharmacokinetics](#) for more details on these topics. Please feel free to e-mail me regarding new packages or major package updates.

### Design and Monitoring

## MethComp: Functions for Analysis of Agreement in Method Comparison Studies

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

Version:	1.22.2
Depends:	R ( $\geq$ 3.0.0), <a href="#">nlme</a>
Suggests:	<a href="#">R2WinBUGS</a> , <a href="#">BRugs</a> , <a href="#">rjags</a> , <a href="#">coda</a> , <a href="#">lattice</a> , <a href="#">lme4</a>
Published:	2015-03-31
Author:	Bendix Carstensen, Lyle Gurrin, Claus Ekstrom, Michal Figurski
Maintainer:	Bendix Carstensen <bcx at steno.dk>
License:	<a href="#">GPL-2</a>   <a href="#">GPL-3</a> [expanded from: GPL ( $\geq$ 2)]
URL:	<a href="http://BendixCarstensen.com/MethComp/">http://BendixCarstensen.com/MethComp/</a>
NeedsCompilation:	no
CRAN checks:	<a href="#">MethComp results</a>

## mcr: Method Comparison Regression

This package provides regression methods to quantify the relation between two measurement methods. In particular it addresses regression problems with errors in both variables and without repeated measurements. The package provides implementations of Deming regression, weighted Deming regression, and Passing-Bablok regression following the CLSI EP09-A3 recommendations for analytical method comparison and bias estimation using patient samples.

Version:	1.2.1
Depends:	R ( $\geq 3.0.0$ ), methods
Suggests:	<a href="#">RUnit</a> , <a href="#">XML</a>
Published:	2014-02-12
Author:	Ekaterina Manuilova Andre Schuetzenmeister Fabian Model
Maintainer:	Fabian Model <fabian.model at roche.com>
License:	<a href="#">GPL (&gt; 3)</a>
NeedsCompilation:	yes

## agRee: Various Methods for Measuring Agreement

Bland-Altman plot and scatter plot with identity line for visualization and point and interval estimates for different metrics related to reproducibility/repeatability/agreement including the concordance correlation coefficient, intraclass correlation coefficient, within-subject coefficient of variation, smallest detectable difference, and mean normalized smallest detectable difference.

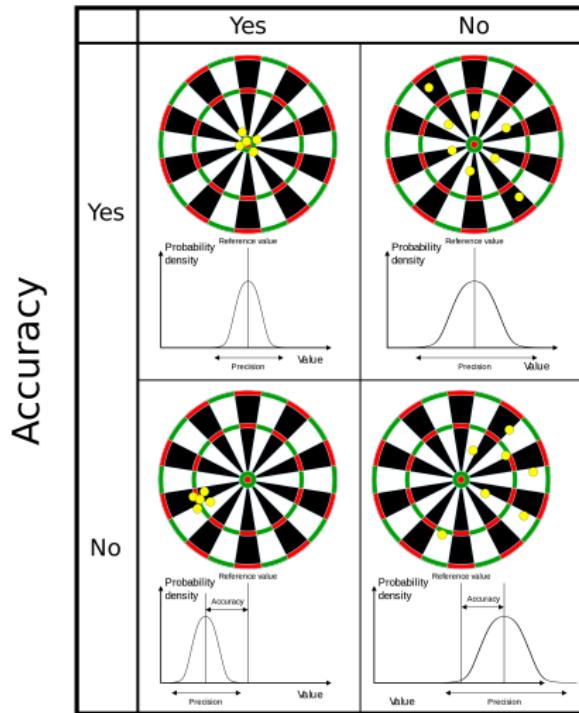
Version:	0.4-0
Depends:	R ( $\geq$ 3.0.2), <a href="#">miscF</a> ( $\geq$ 0.1-2), <a href="#">lme4</a> ( $\geq$ 1.0-4)
Imports:	<a href="#">R2jags</a> ( $\geq$ 0.03-11), <a href="#">coda</a> ( $\geq$ 0.16-1)
Published:	2015-07-10
Author:	Dai Feng
Maintainer:	Dai Feng < <a href="mailto:dai_feng@merck.com">dai_feng@merck.com</a> >
License:	<a href="#">GPL-2</a>   <a href="#">GPL-3</a> [expanded from: GPL]
NeedsCompilation:	no
Materials:	<a href="#">ChangeLog</a>
CRAN checks:	<a href="#">agRee results</a>

## 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.

Version:	0.8-1
Depends:	R ( $\geq$ 2.1.0), <a href="#">R2HTML</a>
Published:	2012-10-29
Author:	Yue Yu AND Lawrence Lin
Maintainer:	Yue Yu <yyu at imyy.net>
License:	<a href="#">GPL-2</a>
URL:	<a href="http://imyy.net">http://imyy.net</a>
NeedsCompilation:	no
CRAN checks:	<a href="#">Agreement results</a>

## Precision



(Wikipedia.org : Accuracy and Precision.svg)

## Three Conditions for Agreement

For two methods of measurement to be considered interchangeable, the following conditions must apply [1]:

- No significant inter-method bias (*accuracy*)
- No difference in the between-subject variabilities of the two methods (*precision*)
- No difference in the within-subject variabilities of the two methods (*repeatability*)

## Repeatability

- **Repeatability** is the variation in measurements taken by a single person or instrument on the same item, under the same conditions, and in a short period of time.
- Methods of measurement should have good repeatability

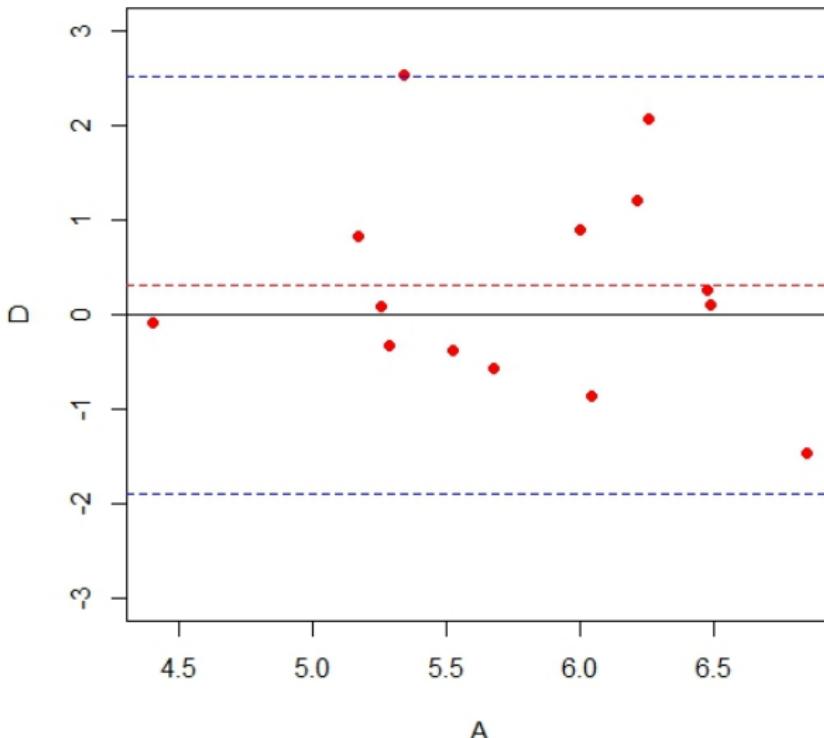
# The Bland-Altman Plot

- The Bland-Altman plot [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.
- 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.

# 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)
```

# Inter-method Bias : 0.27 | Limits of Agreement: [-1.98, 2.52]



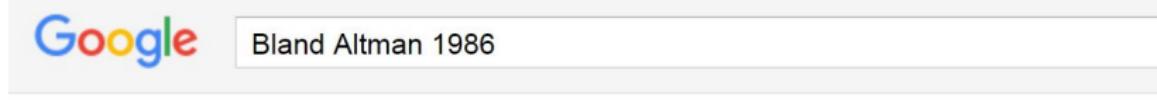
# Bland-Altman Plot

## Building Blocks

- ① Simple Arithmetic Operations
- ② Sample Mean - `mean()`
- ③ Sample Standard deviation - `sd()`
- ④ Scatter plot - `plot()`
- ⑤ Normal Distribution
- ⑥ Enhancing plots - basic R knowledge

**Remark:** Nothing here that is beyond a Stats 101 course in college.

# In excess of 32000 citations



Scholar

## Statistical methods for assessing agreement between two methods of clinical measurement

JM Bland, DG Altman - The lancet, 1986 - Elsevier

Abstract 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. Such investigations are often analysed inappropriately, notably by using ...

Cited by 32934 Related articles All 44 versions Cite Save More

## [CITATION] Regression analysis

JM Bland, DG Altman - The Lancet, 1986 - Elsevier

Cited by 126 Related articles All 4 versions Cite Save More

Showing the best results for this search. See all results

# May 2015 - In excess of 30000 citations



bland altman 1986

[Scholar](#)

About 53,600 results (0.06 sec)

[Articles](#)[Statistical methods for assessing agreement between two methods of clinical measurement](#)JM Bland, DG Altman - *The lancet*, 1986 - Elsevier

Abstract 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. Such investigations are often analysed inappropriately, notably by using ...

Cited by 30519 Related articles All 47 versions Cite Save

[Case law](#)[My library](#)[Any time](#)[Since 2015](#)[Since 2014](#)[Since 2011](#)[Custom range...](#)[Sort by relevance](#)[Sort by date](#) include patents[Agreement between methods of measurement with multiple observations per individual](#)JM Bland, DG Altman - *Journal of biopharmaceutical statistics*, 2007 - Taylor & Francis

... View all references; **Bland and Altman, 1986**. **Bland , JM , Altman , DG (1986)**. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet i:307–310. ... View all references; **Bland and Altman, 1986**. **Bland , JM , Altman , DG (1986)** ...

Cited by 513 Related articles All 7 versions Cite Save

[\[HTML\] Applying the right statistics: analyses of measurement studies](#)JM Bland, DG Altman - *Ultrasound in obstetrics & gynecology*, 2003 - Wiley Online Library

... For each parameter, agreement between MR imaging and arthrography was investigated using the method of **Bland and Altman [1986]**. Arthrography was considered to be the standard and differences between methods were calculated and plotted. ...

Cited by 688 Related articles All 10 versions Cite Save

# nature

International weekly journal of science

[Home](#) | [News & Comment](#) | [Research](#) | [Careers & Jobs](#) | [Current Issue](#) | [Archive](#) | [Audio & Video](#) | [For Authors](#)

[Archive](#) > [Volume 514](#) > [Issue 7524](#) > [News Feature](#) > Article

NATURE | NEWS FEATURE

عربي

## The top 100 papers

*Nature* explores the most-cited research of all time.

**Richard Van Noorden, Brendan Maher & Regina Nuzzo**

29 October 2014



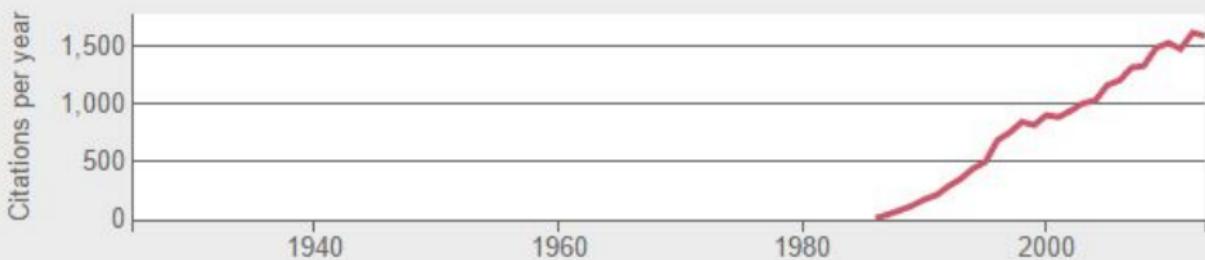
To

Rank: 29 Citations: 23,826

Statistical methods for assessing agreement between two methods of clinical measurement.

Bland, J. M. & Altman, D. G.

*Lancet* 327, 307–310 (1986).



## The top 100 papers



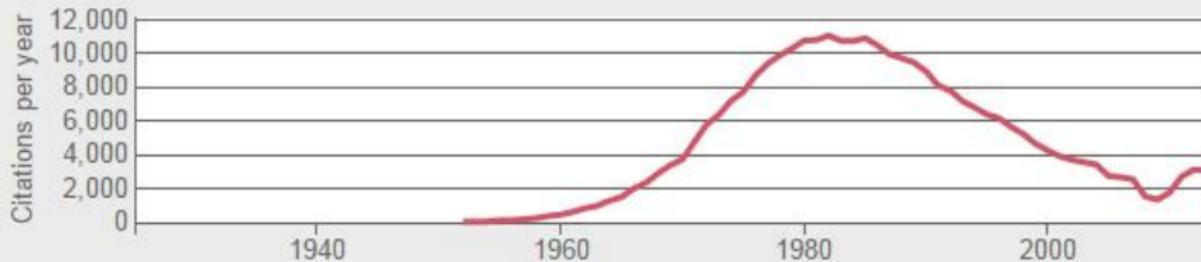
Click through to explore the Web of Science's all-time top-cited papers. (Data provided by Thomson Reuters, extracted on 7 October 2014).

Rank: 1 Citations: 305,148

Protein measurement with the folin phenol reagent.

Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J.

*J. Biol. Chem.* **193**, 265–275 (1951).



*The Kaplan Meier paper was a sleeper hit, receiving almost no citations until computing power boomed in the 1970s, making the methods accessible to non-specialists. Simplicity and ease of use also boosted the popularity of papers in this field.*

*British statisticians Martin Bland and Douglas Altman made the list (number 29) with a technique, now known as the Bland Altman plot, for visualizing how well two measurement methods agree.*

*The same idea had been introduced by another statistician 14 years earlier, but Bland and Altman presented it in an accessible way that has won citations ever since.*

*(Richard Van Noorden, Brendan Maher & Regina Nuzzo)*

# R Packages for Bland-Altman Analysis

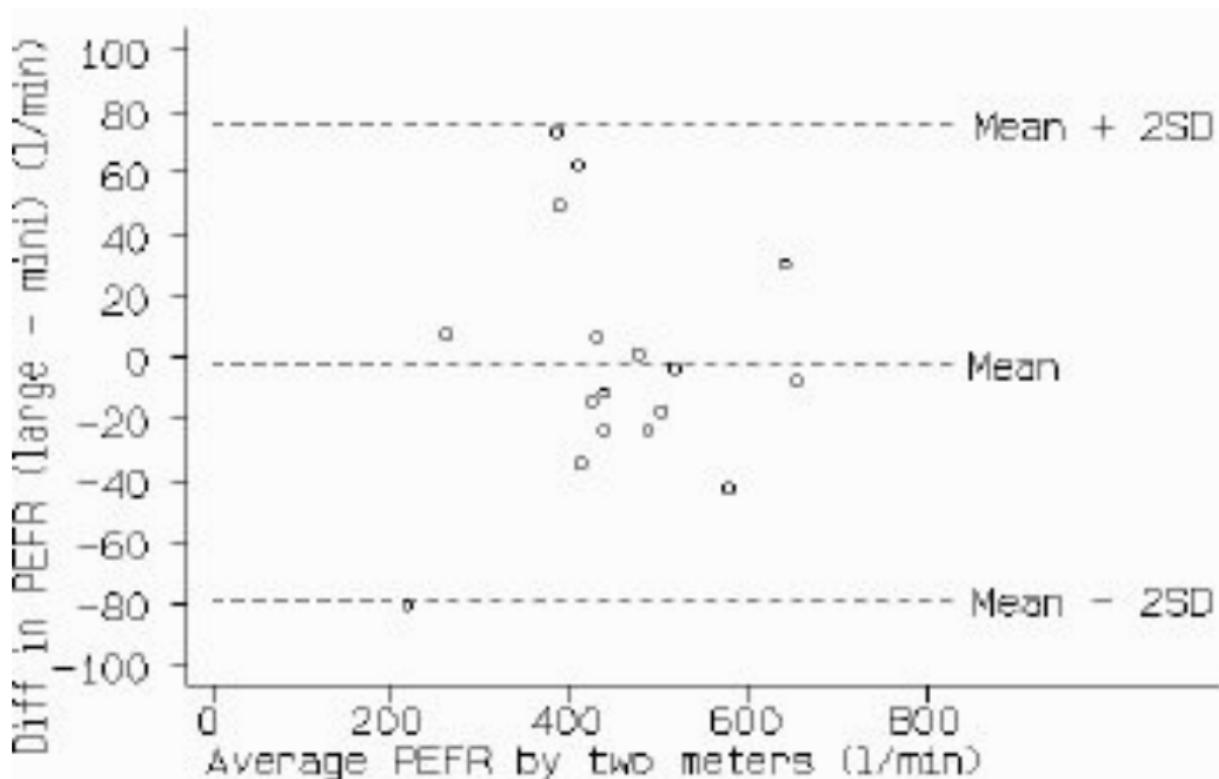
`PairedData` has a function `plotBA` based on `ggplot2` and no stats as return value

`ResearchMethods` has a function `BlandAltman` which focuses on a GUI and has no return values.

`epade` has a function `bland.altman.ade` which appears to have no return values.

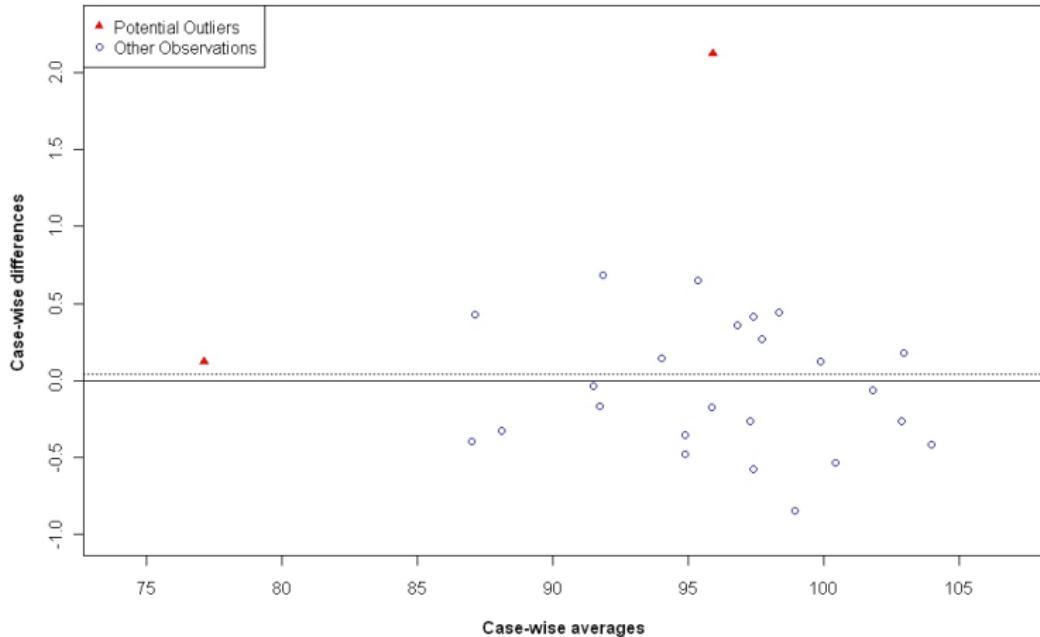
`MethComp` has a function `BlandAltman` that is deprecated and a function `ba.plot` which does a lot, mainly regression methods

# Interpreting the Bland-Altman Plot



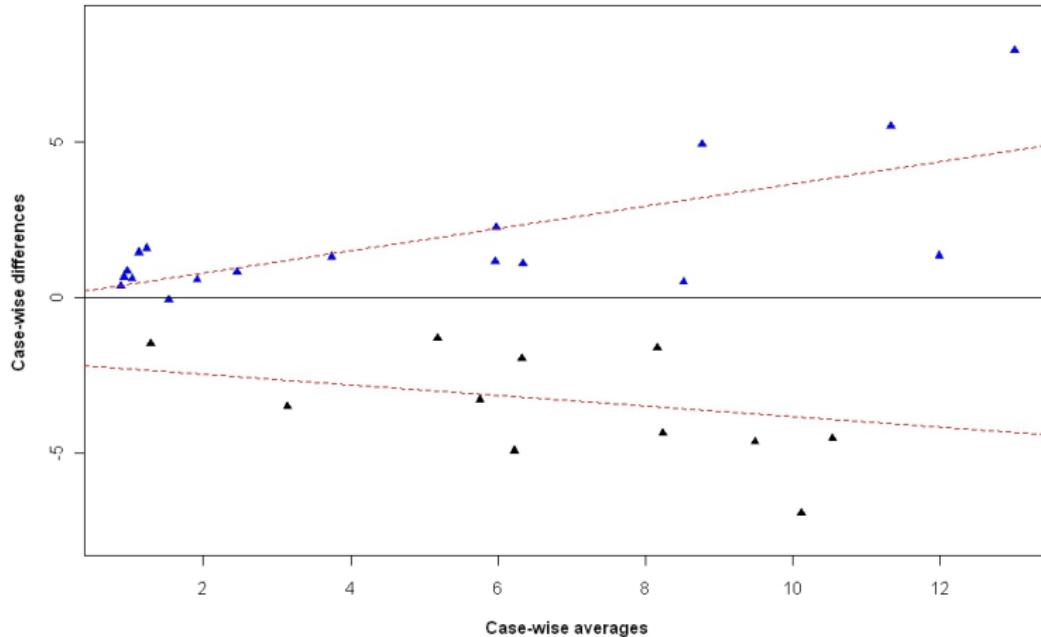
# Interpreting the Bland-Altman Plot

### Bland-Altman plot: indicating potential outliers

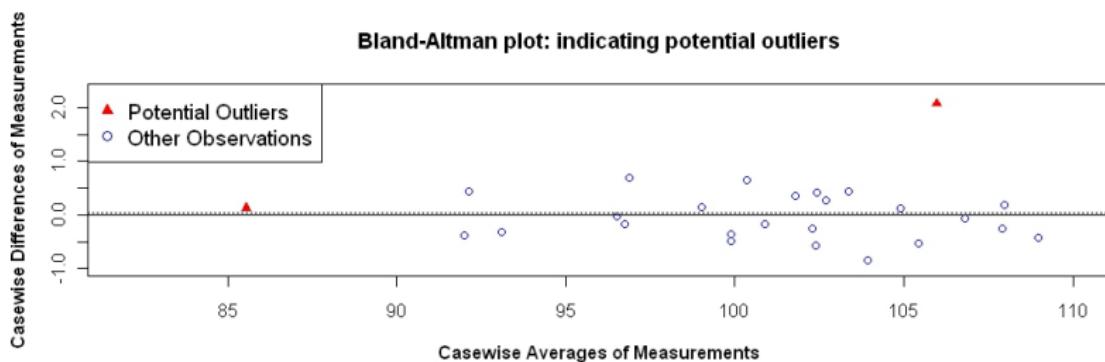
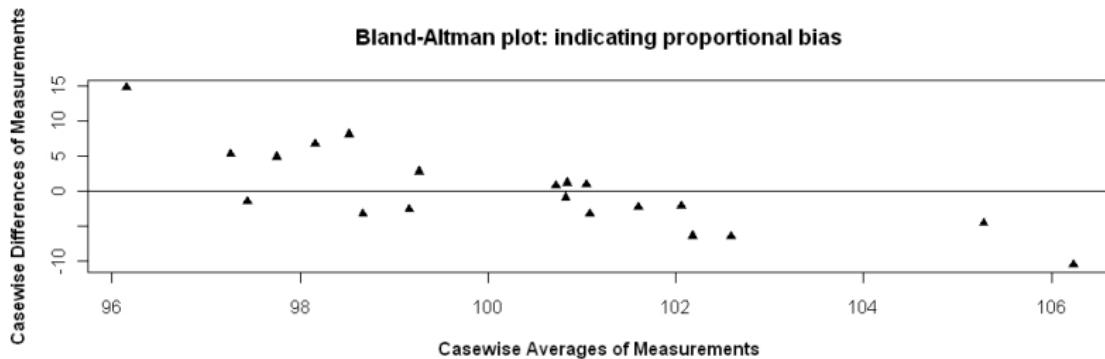


# Interpreting the Bland-Altman Plot

Bland-Altman plot: lack of constant variance



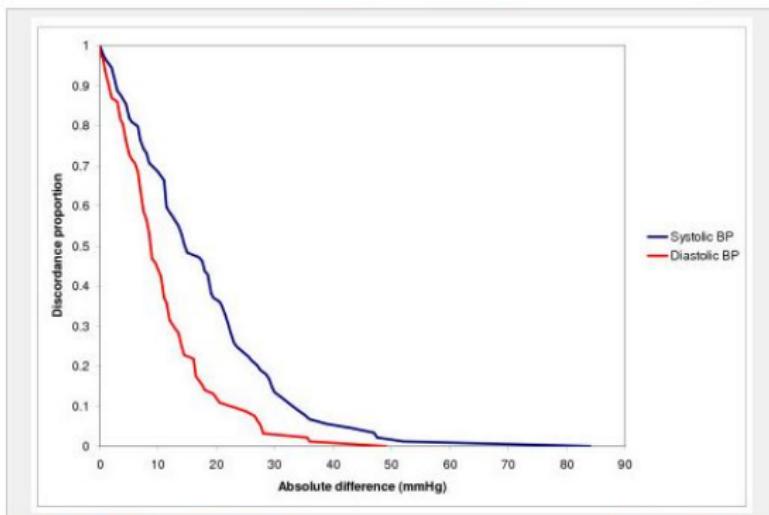
# Interpreting the Bland-Altman Plot



# The Bland-Altman Plot: Prevalence

- Limits of Agreement are used extensively in medical literature for assessing agreement between two methods.
- Building Blocks are featured in almost every undergraduate statistics course (i.e. Mean, Standard Deviation, Scatterplot, Normal Distribution)
- Other graphical techniques, such as *Survival-Agreement Plot* (based on Kaplan-Meier Curve) and *Mountain Plot* have been developed, but are not prevalent at all.

## Survival-Agreement Curve



© Copyright Policy - open-access  
License

Related In: Results - Collection  
Show All Figures

**Figure 1:** Survival-agreement plot, as proposed by Luiz et al.[10] The x-axis shows the absolute difference between self-reported and measured blood pressure (BP), and the y-axis shows the proportion of observations with differences that are at least the observed difference. Separate lines for systolic and diastolic BP.

## Technology Acceptance Model

Davis (1989) proposes the TAM model, which suggests an hypothesis as to why users may adopt particular technologies, and not others.

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"

# Technology Acceptance Model

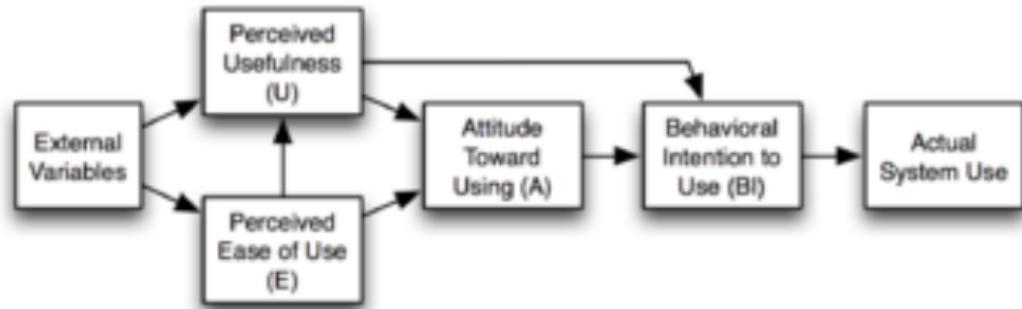


Figure: Technology Acceptance Model Flowchart (Davis, 1989)

- Bland-Altman method not very good on it's own.
- Does not account for Replicate Measurements.
- Useful as a diagnostic method subsequent to other methods.
- Develop a proper methodology for MCS and Get people to use it!

# Shiny

by RStudio

# A web application framework for R

# Shiny Web Applications with R

## Useful Shiny Resources

- [shiny.rstudio.com](http://shiny.rstudio.com)
- [showmeshirey.com](http://showmeshirey.com)
- [shiny.snap.uaf.edu/](http://shiny.snap.uaf.edu/)

# Shiny-phyloseq

[Shiny-phyloseq](#) is an interactive web application that provides a graphical user interface to the microbiome analysis package for R, called [phyloseq](#). For details about using the [phyloseq](#) package directly, see [The phyloseq Homepage](#).

## Citation

Shiny-phyloseq is provided under a free-of-charge, open-source license (A-GPL3). All we require is that you cite/attribute the following in any work that benefits from this code or application.

## Citing the Web Application

McMurdie and Holmes (2014) Shiny-phyloseq: Web Application for Interactive Microbiome Analysis with Provenance Tracking. *Bioinformatics* *in press*.

# 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".

## Three Conditions

For two methods of measurement to be considered interchangeable, the following conditions must apply [1]:

- No significant inter-method bias
- No difference in the between-subject variabilities of the two methods
- No difference in the within-subject variabilities of the two methods (repeatability)

# Part 2 : Using LME Models

# WIKIPEDIA : Linear Models (fixed effects only)

## Introduction to linear regression [ edit]

Given a data set  $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$  of  $n$  statistical units, a linear regression model assumes that relationship between the dependent variable  $y_i$  and the  $p$ -vector of regressors  $x_i$  is linear. This relationship is modeled through a disturbance term or error variable  $\varepsilon_i$  — an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n,$$

where  $T$  denotes the transpose, so that  $\mathbf{x}_i^T \boldsymbol{\beta}$  is the inner product between vectors  $\mathbf{x}_i$  and  $\boldsymbol{\beta}$ .

Often these  $n$  equations are stacked together and written in vector form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

# WIKIPEDIA : Linear Mixed Effects Models

## Definition [\[edit\]](#)

---

In matrix notation a mixed model can be represented as

$$\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{u} + \boldsymbol{\epsilon}$$

where

- $\mathbf{y}$  is a known vector of observations, with mean  $E(\mathbf{y}) = X\boldsymbol{\beta}$ ;
- $\boldsymbol{\beta}$  is an unknown vector of fixed effects;
- $\mathbf{u}$  is an unknown vector of random effects, with mean  $E(\mathbf{u}) = \mathbf{0}$  and variance-covariance matrix  $\text{var}(\mathbf{u}) = G$ ;
- $\boldsymbol{\epsilon}$  is an unknown vector of random errors, with mean  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  and variance  $\text{var}(\boldsymbol{\epsilon}) = R$ ;
- $X$  and  $Z$  are known design matrices relating the observations  $\mathbf{y}$  to  $\boldsymbol{\beta}$  and  $\mathbf{u}$ , respectively.

[Figure: Wikipedia Entry on LME Models](#)

# The nlme R package

## nlme: Linear and Nonlinear Mixed Effects Models

Fit and compare Gaussian linear and nonlinear mixed-effects models.

Version: 3.1-120

Priority: recommended

Depends: graphics, stats, R ( $\geq 3.0.0$ )

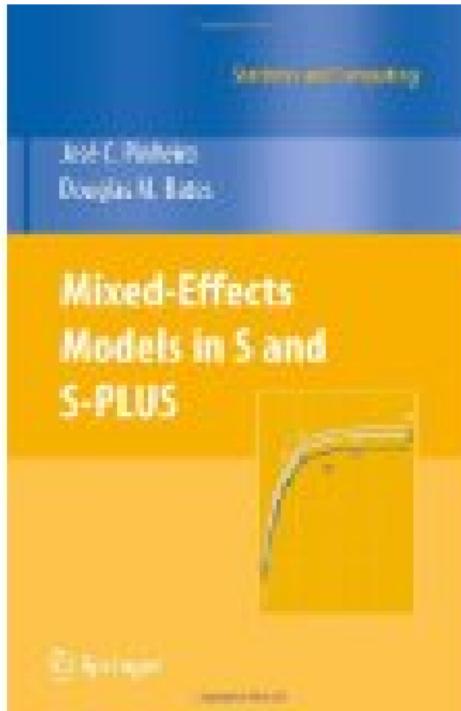
Imports: [lattice](#)

Suggests: [Hmisc](#), [MASS](#)

Published: 2015-02-20

Author: José Pinheiro [aut] (S version), Douglas Bates [aut] (up to 2007), Saikat [ctb] (up to 2002), Deepayan Sarkar [ctb] (up to 2005), EISPACK author (src/rs.f), R-core [aut, cre]

Maintainer: R-core <R-core at R-project.org>



# Chapter 5 : Extending the Basic LME Model

<b>5 Extending the Basic Linear Mixed-Effects Model</b>	<b>201</b>
5.1 General Formulation of the Extended Model . . . . .	202
5.1.1 Estimation and Computational Methods . . . . .	202
5.1.2 The GLS model . . . . .	203
5.1.3 Decomposing the Within-Group Variance–Covariance Structure . . . . .	205
5.2 Variance Functions for Modeling Heteroscedasticity . . . . .	206
5.2.1 <code>varFunc</code> classes in <code>nlme</code> . . . . .	208
5.2.2 Using <code>varFunc</code> classes with <code>lme</code> . . . . .	214
5.3 Correlation Structures for Modeling Dependence . . . . .	226
5.3.1 Serial Correlation Structures . . . . .	226
5.3.2 Spatial Correlation Structures . . . . .	230
5.3.3 <code>corStruct</code> classes in <code>nlme</code> . . . . .	232
5.3.4 Using <code>corStruct</code> Classes with <code>lme</code> . . . . .	239
5.4 Fitting Extended Linear Models with <code>gls</code> . . . . .	249
5.5 Chapter Summary . . . . .	266
Exercises . . . . .	267

# LME4 R Package

## lme4: Linear mixed-effects models using Eigen and S4

Fit linear and generalized linear mixed-effects models. The models and their components are represented using S4 classes and methods. The core computational algorithms are implemented using the Eigen C++ library for numerical linear algebra and RcppEigen "glue".

Version:

1.1-7

Depends:

R ( $\geq$  2.15.1), [Matrix](#) ( $\geq$  1.1.1), methods, stats, [Rcpp](#) ( $\geq$  0.10.5)

Imports:

graphics, grid, splines, parallel, [MASS](#), [nlme](#), [lattice](#), [minqa](#) ( $\geq$  1.1.15), [nlopt](#)

LinkingTo:

[Rcpp](#), [RcppEigen](#)

Suggests:

[knitr](#), [boot](#), [PKPDmodels](#), [MEMSS](#), [testthat](#) ( $\geq$  0.8.1), [ggplot2](#), [mlmRev](#), [opt](#)  
( $\geq$  2013.8.6), [gamm4](#), [pbkrtest](#)

Published:

2014-07-19

Author:

Douglas Bates [aut], Martin Maechler [aut], Ben Bolker [aut, cre], Steven W  
[aut], Rune Haubo Bojesen Christensen [ctb], Henrik Singmann [ctb], Bin D  
[ctb]





# Douglas Bates

dmbates

University of Wisconsin  
 Madison, WI, U.S.A.  
 Joined on Aug 20, 2010

Contributions

Repositories

Public activity

## Popular repositories

### MixedModels.jl

A Julia package for fitting (statistical) mixed-e...

51

### JuliaWorkshop

Materials for a workshop on Julia programmi...

13

### RePsychLing

Data sets from subject/item type studies in Ps...

9

### stat692

Materials for Statistics 692 at UW-Madison, F...

7

### ParallelGLM.jl

Parallel fitting of GLMs using SharedArrays

5

## Public contributions

## Repos

### Stat

Rev...

### Julia

Emb...

### Julia

Meta...

### Julia

Julia...

### bria

Mark...

# The nlme Package

(For review)

- LME models can be implemented in R using the nlme package, one of the core packages.
- Authors: Jose Pinheiro, Douglas Bates (up to 2007), Saikat DebRoy (up to 2002), Deepayan Sarkar (up to 2005), the R Core team  
(source: nlme package manual)
- “Mixed-Effects Models in S and S-PLUS” by JC Pinheiro and DM Bates (Springer,2000)

# LME models

- In a linear mixed-effects (LME) model, responses from a subject are due to both fixed and random effects.
- A random effect is an effect associated with a sampling procedure. Replicate measurements would require use of random effect terms in model.
- **(Essentially : Use of random effects allows sets of observations to be grouped together )**
- Can have differing number of replicate measurements for different subjects.

The image shows a screenshot of an R console window. The title bar says "R Console". The command "head(sbp, 10)" is entered in red at the top. Below it is a table with columns labeled "meth", "item", "repl", and "y". The data consists of 10 rows, numbered 1 to 10. The "meth" column has values J, J, J, R, R, R, S, S, S, J. The "item" column has values 1, 1, 1, 1, 1, 1, 1, 1, 1, 2. The "repl" column has values 1, 2, 3, 1, 2, 3, 1, 2, 3, 1. The "y" column has values 100, 106, 107, 98, 98, 111, 122, 128, 124, 108.

	meth	item	repl	y
1	J	1	1	100
2	J	1	2	106
3	J	1	3	107
4	R	1	1	98
5	R	1	2	98
6	R	1	3	111
7	S	1	1	122
8	S	1	2	128
9	S	1	3	124
10	J	2	1	108

Figure: Systolic Blood Pressure Data (MethComp Package, Carstensen et al)

## Example: Blood Data

- Used in Bland and Altman's 1999 paper [3]. Data was supplied by Dr E O'Brien.
- Simultaneous measurements of systolic blood pressure each made by two experienced observers, J and R, using a sphygmometer.
- Measurements also made by a semi-automatic blood pressure monitor, denoted S.
- On 85 patients, 3 measurement made in quick succession by each of the three observers (765 measurements in total)

# Roy's Approach

- Roy proposes an LME model with Kronecker product covariance structure in a doubly multivariate setup.
- Response for  $i$ th subject can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + b_{1i} z_{i1} + b_{2i} z_{i2} + \epsilon_i$$

- $\beta_1$  and  $\beta_2$  are fixed effects corresponding to both methods.  
( $\beta_0$  is the intercept.)
- $b_{1i}$  and  $b_{2i}$  are random effects corresponding to both methods.

## Roy's LME model

- Let  $\mathbf{y}_i$  be the set of responses for subject  $i$  (in matrix form).
  - $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$
  - $\mathbf{b}_i \sim N_m(0, \mathbf{D})$  (m: number of methods)
  - $\boldsymbol{\epsilon}_i \sim N_{n_i}(0, \mathbf{R})$  ( $n_i$ : number of measurements on subject  $i$ )

**(Remark:** Using Roy's own notation, which is different from Wikipedia)

# Variance-covariance matrix

- Overall variance covariance matrix for response vector  $\mathbf{y}_i$

$$\text{Var}(\mathbf{y}_i) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i + \mathbf{R}_i$$

- can be re-expressed as follows:

$$\mathbf{Z}_i \begin{bmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{bmatrix} \mathbf{Z}'_i + \left( V \otimes \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right)$$

- Overall variability between the two methods is sum of between-subject and within-subject variability,

$$\text{Block } \Omega_i = \begin{bmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

# Variance-Covariance Structures

Further to Chapter 5 of Pinheiro Bates

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

- Symmetric structure specifies that  $\sigma_1^2$  may differ from  $\sigma_2^2$ .
  - Compound symmetric (CS) structure specifies that  $\sigma_1^2 = \sigma_2^2$ .
  - In both cases,  $\sigma_{12}$  may take value other than 0.
  - (*What is stated here is applicable to  $D$  also*)

- Roy's uses an LME model approach to provide a set of formal tests for method comparison studies.
- Four candidate models are fitted to the data. One is a **reference model**, and three are **nested models**.
- All of these models are similar to one another, but for the imposition of equality constraints in the nested models (i.e. Using CS structure)

# The Reference Model

```
REF = lme(y ~ meth,  
           data = dat,  
           random = list(item=pdSymm(~ meth-1)),  
           weights=varIdent(form=~1|meth),  
           correlation = corSymm(form=~1 | item/repl),  
           method="ML")
```

- LME model that specifies a symmetric matrix structure for both between-subject and within-subject variances.
- No Equality Constraints

# The Nested Model 1 (Between-Subject Variances)

```
NMB = lme(y ~ meth,  
           data = dat,  
           random = list(item=pdCompSymm(~ meth-1)),  
           weights=varIdent(form=~1|meth),  
           correlation = corSymm(form=~1 | item/repl),  
           method="ML")
```

- LME model that specifies a compound symmetric matrix structure for between-subject (**i.e. Equality Constraint imposed**) and symmetric structure within-subject variances.

# The Nested Model 2 (Within-Subject Variances)

```
NMW = lme(y ~ meth,  
           data = dat,  
           random = list(item=pdSymm(~ meth-1)),  
           #weights=varIdent(form=~1|meth),  
           correlation = corCompSymm(form=~1|item/repl),  
           method="ML")
```

- LME model that specifies a symmetric matrix structure for between-subject and compound symmetric structure within-subject variances (*i.e. Equality Constraint imposed*).

# The Nested Model 3 (Overall Variances)

```
NMO = lme(y ~ meth,  
           data = dat,  
           random = list(item=pdCompSymm(~ meth-1)),  
           #weights=varIdent(form=~1|meth),  
           correlation = corCompSymm(form=~1|item /repl),  
           method="ML")
```

- LME model that specifies a compound symmetric matrix structure for both between-subject and within-subject variances.

# Example: Blood Data

Inter-method Bias between J and S: 15.62 mmHg

```
>summary(REF)
```

.....

Fixed effects: y ~ meth

	Value	Std.Error	DF	t-value	p-value
(Intercept)	127.41	3.3257	424	38.310	0
methS	15.62	2.0456	424	7.636	0

.....

# Between-subject variance covariance matrix

..

Random effects:

Formula: ~method - 1 | subject

Structure: General positive-definite

StdDev Corr

methodJ 30.396975 methdJ

methodS 31.165565 0.829

Residual 6.116251

..

$$\hat{\mathbf{D}} = \begin{pmatrix} 923.97 & 785.34 \\ 785.34 & 971.29 \end{pmatrix}$$

# Within-subject variance covariance matrix

Correlation Structure: General

Formula: ~1 | subject/obs

Parameter estimate(s):

Correlation:

1

2 0.288

Variance function:

Structure: Different standard deviations per stratum

Formula: ~1 | method

Parameter estimates:

J S

1.000000 1.490806

$$\hat{\Sigma} = \begin{pmatrix} 37.40 & 16.06 \\ 16.06 & 83.14 \end{pmatrix}$$

## Overall variance covariance matrix

- Overall variance

$$\text{Block } \hat{\Omega} = \hat{D} + \hat{\Sigma} = \begin{pmatrix} 961.38 & 801.40 \\ 801.40 & 1054.43 \end{pmatrix}$$

- Standard deviation of the differences can be computed accordingly : 20.32 mmHg.
- Furthermore, limits of agreement can be computed:  
 $[15.62 \pm (2 \times 20.32)]$  (mmHg).

## Some useful R commands

- intervals :

This command obtains the estimate and confidence intervals on the parameters associated with the model. This is particularly useful in writing some code to extract estimates for inter-method bias and variances, and hence estimates for the limits of agreement.

- anova :

When a reference model and nested model are specified as arguments, this command performs a **likelihood ratio test**.

## Formal Tests: Between-subject Variances

- Test the hypothesis that both methods have equal between-subject variances.
- Constructed an alternative model “Nested Model B” using ***compound symmetric*** form for between-subject variance (hence specifying equality of between-subject variances).
- Use a likelihood ratio test to compare models.

...

```
> anova (REF, NMB)
```

	Model	df	...	logLik	Test	L.Ratio	p-value
REF	1	8	...	-2030.736			
NMB	2	7	...	-2030.812	1 vs 2	0.1529142	0.6958

...

- Fail to reject hypothesis of equality.

## Formal Tests: Within-subject Variances

- Test the hypothesis that both methods have equal within-subject variances.
- Constructed an alternative model "Nested Model W" using compound symmetric form for within-subject variance (hence specifying equality of within-subject variances).
- Again, use a likelihood ratio test to compare models.

...

```
> anova (REF, NMW)
```

	Model	df	...	logLik	Test	L.Ratio	p-value
REF		1	8	...	-2030.736		
NMW		2	7	...	-2045.044	1 vs 2	28.61679 <.0001

- Reject hypothesis of equality.

## Formal Tests : Outcomes

- Inter-method bias: Significant difference in mean values detected.
- Between-subject variance: No significant difference in between-subject variances between the two methods detected.
- Within-subject variance: A significant difference in within-subject variances is detected.
- Can not recommend switching between the two methods.

## Remarks

- Can perform a test for equality of overall variances.
- This can be done by specifying a compound symmetry structure for both between-subject and within-subject variances when constructing a nested model.
- Roy controls the family-wise error rate in paper, using Bonferroni correction procedure.

# Part 3 : Influence and Case Deletion Diagnostics

# influence.ME: Tools for Detecting Influential Data in Mixed Effects Models

by Rense Nieuwenhuis, Manfred te Grotenhuis, and Ben Pelzer

**Abstract** **influence.ME** provides tools for detecting influential data in mixed effects models. The application of these models has become common practice, but the development of diagnostic tools has lagged behind. **influence.ME** calculates standardized measures of influential data for the point estimates of generalized mixed effects models, such as DFBETAS, Cook's distance, as well as percentile change and a test for changing levels of significance. **influence.ME** calculates these measures of influence while accounting for the nesting structure of the data. The package and measures of influential data are introduced, a practical example is given, and strategies for dealing with influential data are suggested.

The application of mixed effects regression models has become common practice in the field of social sci-

large numbers of observations at the individual level while the number of higher level groups is relatively small. For instance, Van der Meer, te Grotenhuis, and Pelzer (2010) were unable to find any country-level comparative studies involving more than 54 countries. With such a relatively low number of countries, a single country can easily be overly influential on the parameter estimates of one or more of the country-level variables.

## Detecting Influential Data

All cases used to estimate a regression model exert some level of influence on the regression parameters. However, if a single case has extremely high or low scores on the dependent variable relative to its expected value — given other variables in the model, one or more of the independent variables, or both — this case may overly influence the regression parameters by 'pulling' the estimated regression line towards itself. The simple inclusion or exclusion of

## Abstract for `influence.ME` Package

*`influence.ME` provides tools for detecting influential data in mixed effects models. The application of these models has become common practice, but the development of diagnostic tools has lagged behind.*

*`influence.ME` calculates standardized measures of influential data for the point estimates of generalized mixed effects models, such as **DFBETAS**, **Cook's distance**, as well as percentile change and a test for changing levels of significance.*

- `influence.ME` only works for models fitted with LME4.
- Can't use with previous approach so lets try new model with LME4

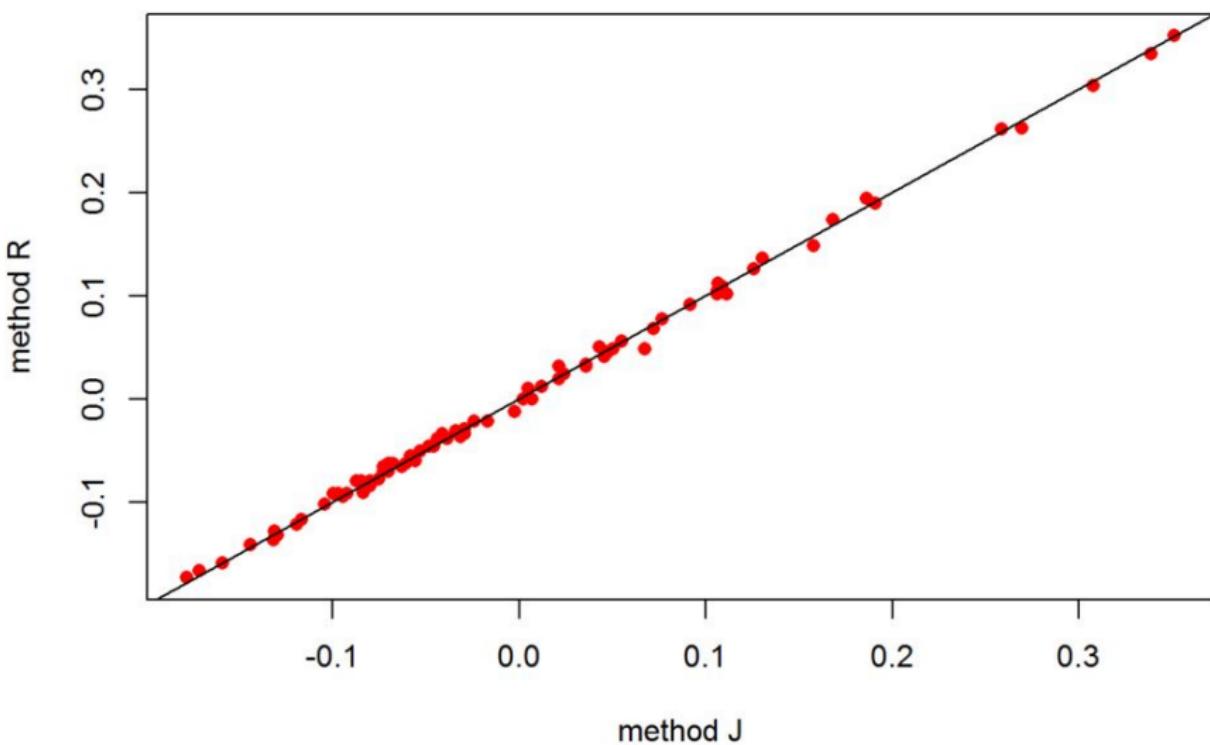
```
sbp.LME4 = lmer(y ~ method-1 + (1 | subject ),  
                  data = sbp,  
                  REML = FALSE)
```

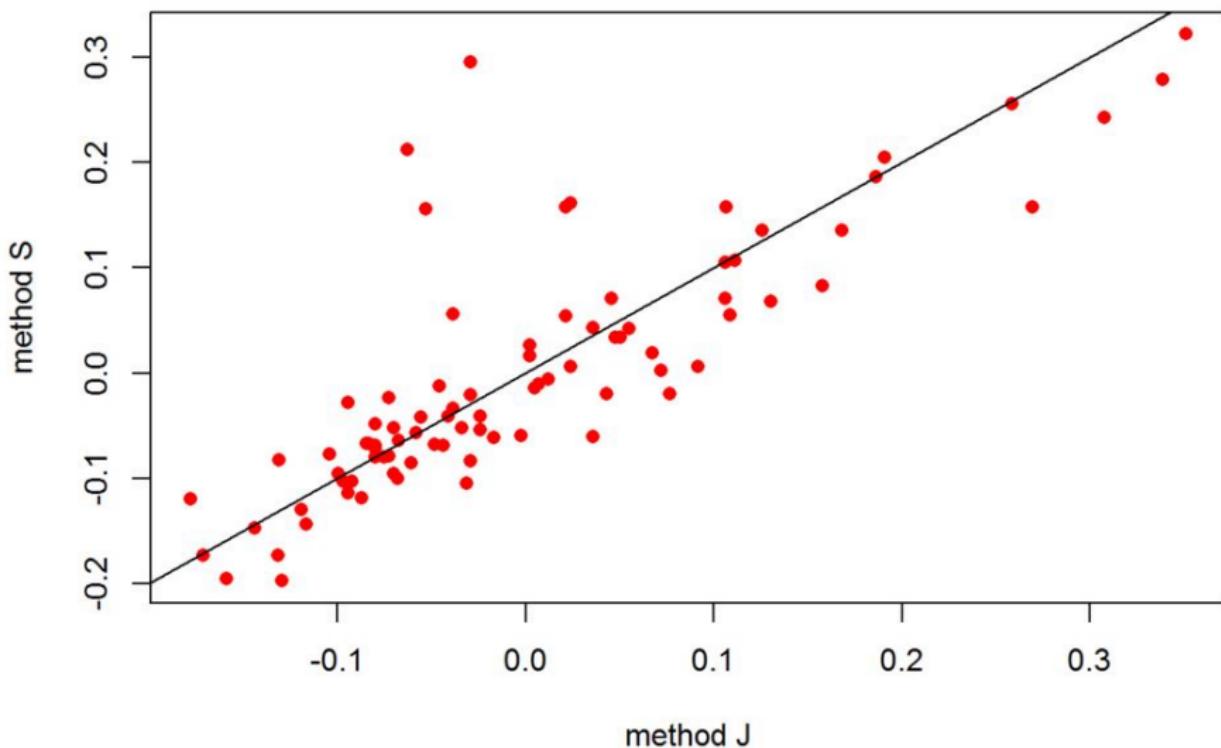
```
## Fixed effects:  
  
##           Estimate Std. Error t value  
## methodJ   127.408     3.257  39.12  
## methodR   127.322     3.257  39.10  
## methods  143.027     3.257  43.92  
##
```

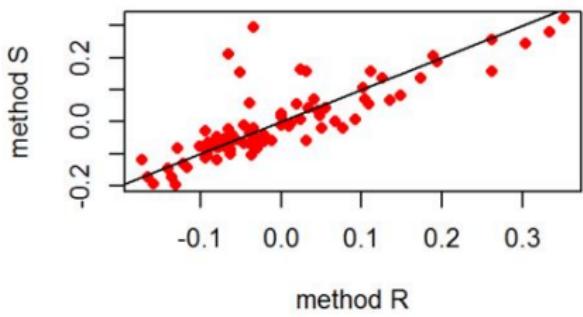
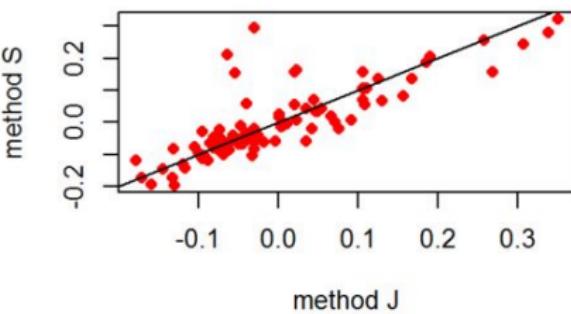
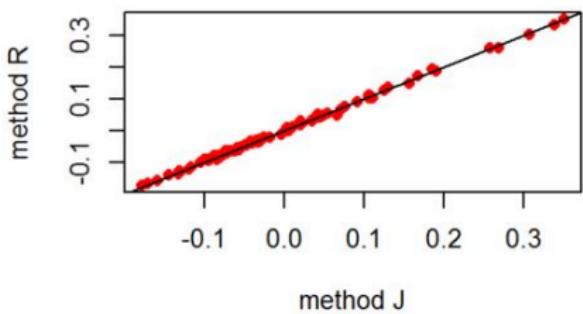
```
MCS.LME4.inf <- influence(sbp.LME4, "subject")
CDs<- cooks.distance(MCS.LME4.inf)
DFBs <- dfbetas(MCS.LME4.inf)
```

## DFBetas ranked by Cook's Distance

```
##      case      CDs    methodJ    methodR   methods
## 78  78 0.61557407 -0.02934556 -0.03387780 0.2954937
## 80  80 0.41590973 -0.06305026 -0.06515241 0.2123881
## 68  68 0.22536651 -0.05334867 -0.05062375 0.1555187
## 72  72 0.09348500  0.02388626  0.02419887 0.1617474
## 48  48 0.08706988  0.02147541  0.03145273 0.1581591
## 30  30 0.07118415  0.26925807  0.26215970 0.1581569
```







## Case Deletion Diagnostics for nlme Model

- Going back to `nlme` model
- We can compute Between Subject Variance and Within-Subject Variance for both methods.
- Express these as ratios (BWVR and WSVR).
- Use Case Deletion diagnostics to study effect of each case.
- *Remove one subject each time and refit model (Not Computationally Efficient, but still very useful)*

# Between-subject variance covariance matrix

..

Random effects:

Formula: ~method - 1 | subject

Structure: General positive-definite

StdDev Corr

methodJ 30.396975 methdJ

methodS 31.165565 0.829

Residual 6.116251

..

$$\hat{\mathbf{D}} = \begin{pmatrix} 923.97 & 785.34 \\ 785.34 & 971.29 \end{pmatrix}$$

# Within-subject variance covariance matrix

Correlation Structure: General

Formula: ~1 | subject/obs

Parameter estimate(s):

Correlation:

1

2 0.288

Variance function:

Structure: Different standard deviations per stratum

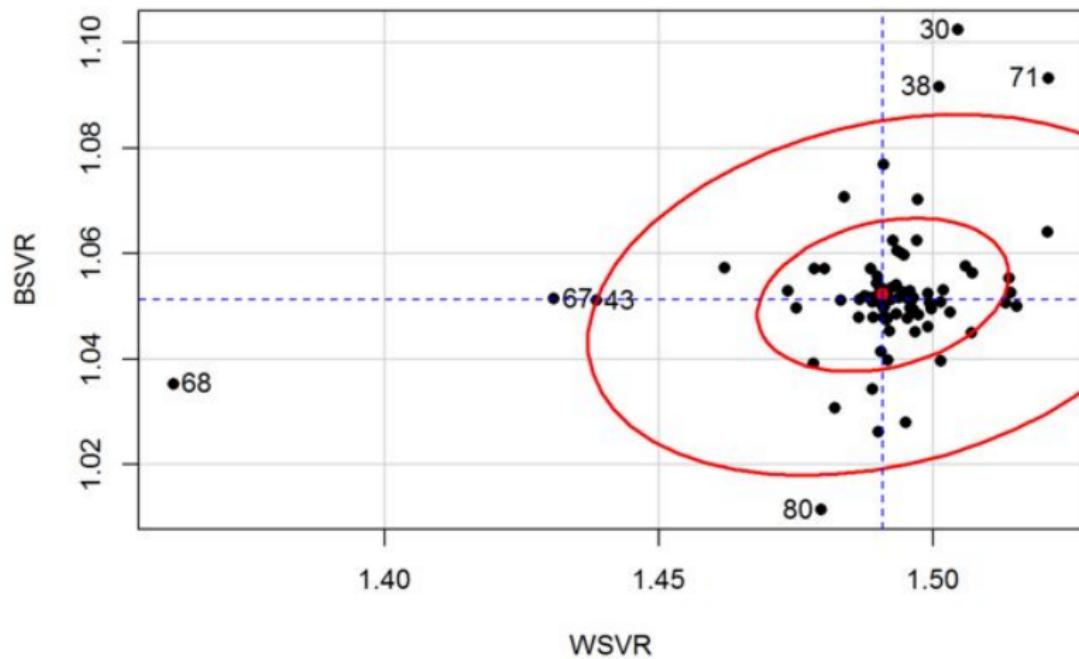
Formula: ~1 | method

Parameter estimates:

J S

1.000000 1.490806

$$\hat{\Sigma} = \begin{pmatrix} 37.40 & 16.06 \\ 16.06 & 83.14 \end{pmatrix}$$



# References

-  A Roy (2009): *An application of linear mixed effects model to assess the agreement between two methods with replicated observations* Journal of Biopharmaceutical Statistics
-  Bland JM, Altman DG (1986) *Statistical method for assessing agreement between two methods of clinical measurement.*
-  Bland JM, Altman DG (1999) *Measuring agreement in method comparison studies.* Statistical Methods in Medical Research
-  Pinheiro JC, Bates DM (2000): *Mixed-effects models in S and S-PLUS*, Springer.

# Thanks

- Dr Kevin Hayes, University of Limerick
- Dr Kevin Burke, University of Limerick
- Peter Fennell