Method Comparison Studies with R

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

Medical statistics

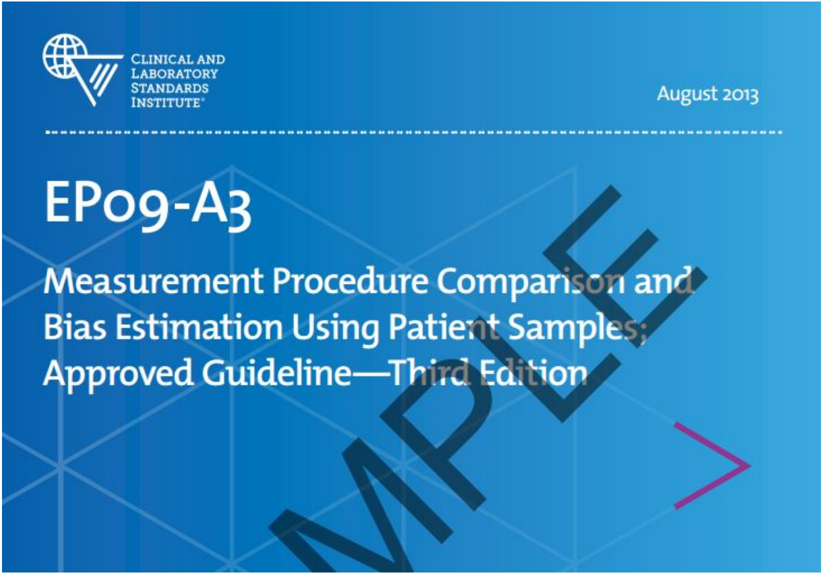
- Applications to medicine and the health sciences, including epidemiology, public health, forensic medicine, and clinical research.
- "Biostatistics" more commonly connotes all applications of statistics to biology.
- Clinical Research is main focus for this talk - Method Comparison Studies

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, transformativ 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, enhanc



CLINICAL AND
LABORATORY
STANDARDS
INSTITUTE®

August 2013

EP09-A3

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

Medical Measurement

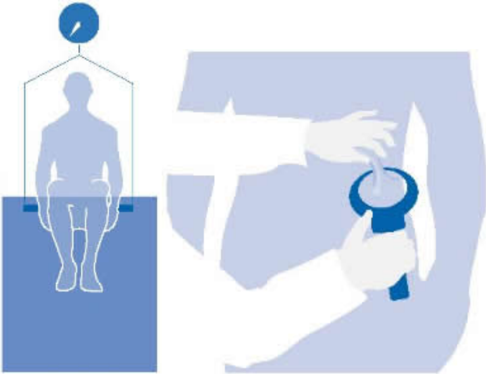
Common Scenario for comparing two methods of measurements

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.

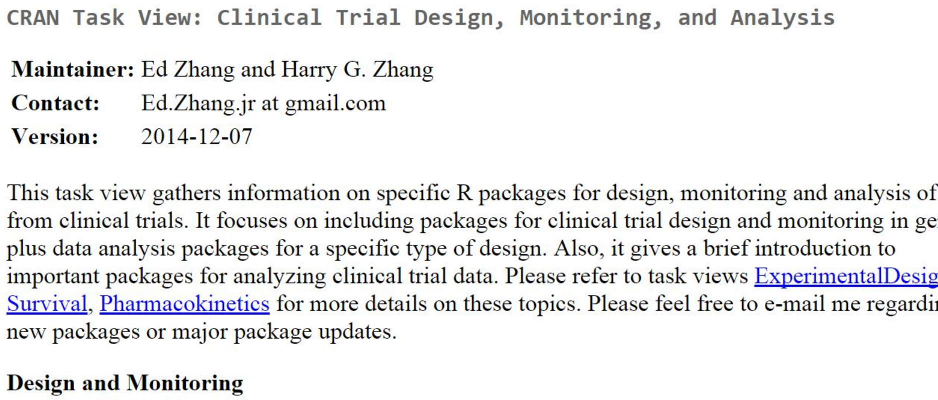
CitiSense is the only air-quality monitoring system capable of delivering real-time data to users' cell

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

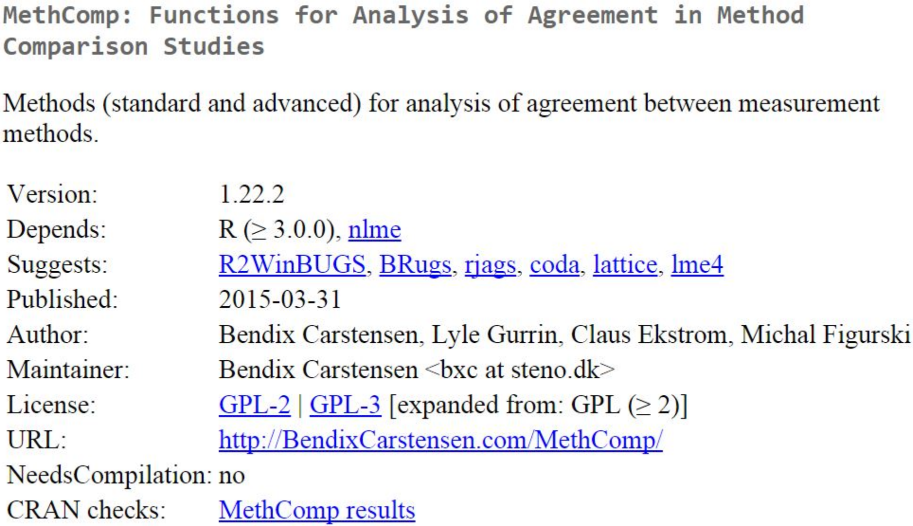


CRAN Task View: Clinical Trial Design, Monitoring, and Analysis

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
from clinical trials. It focuses on including packages for clinical trial design and monitoring in ge
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 ExperimentalDesig
Survival, Pharmacokinetics for more details on these topics. Please feel free to e-mail me regardi
new packages or major package updates.

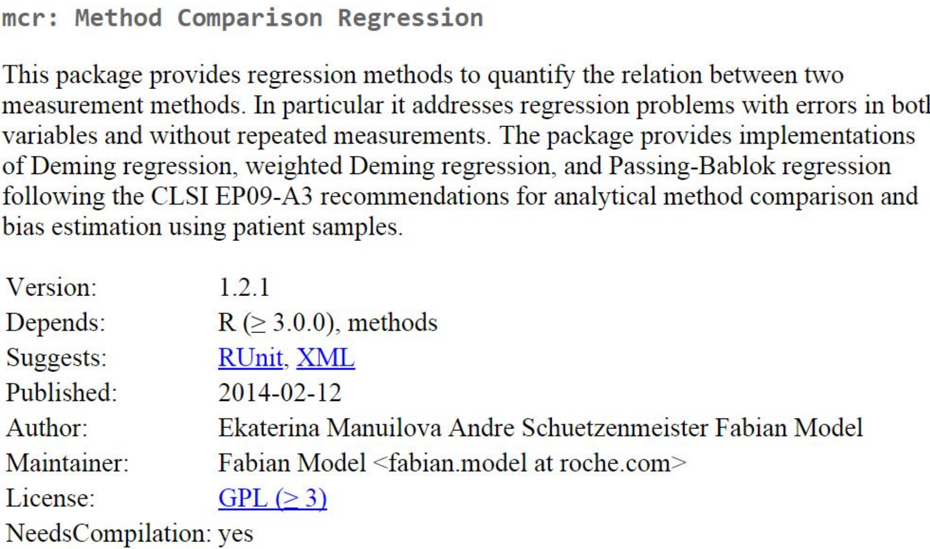
Design and Monitoring



MethComp: Functions for Analysis of Agreement in Method Comparison Studies

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

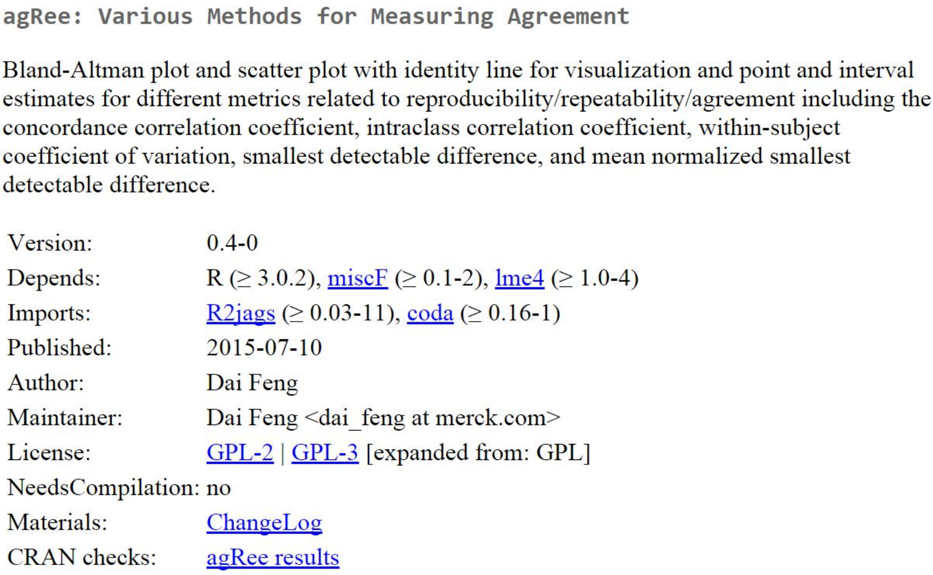
Version:	1.22.2
Depends:	R (≥ 3.0.0), nlme
Suggests:	R2WinBUGS, BRugs, rjags, coda, lattice, lme4
Published:	2015-03-31
Author:	Bendix Carstensen, Lyle Gurrin, Claus Ekstrom, Michal Figurski
Maintainer:	Bendix Carstensen <bxc at steno.dk>
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]
URL:	http://BendixCarstensen.com/MethComp/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



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 bot 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.

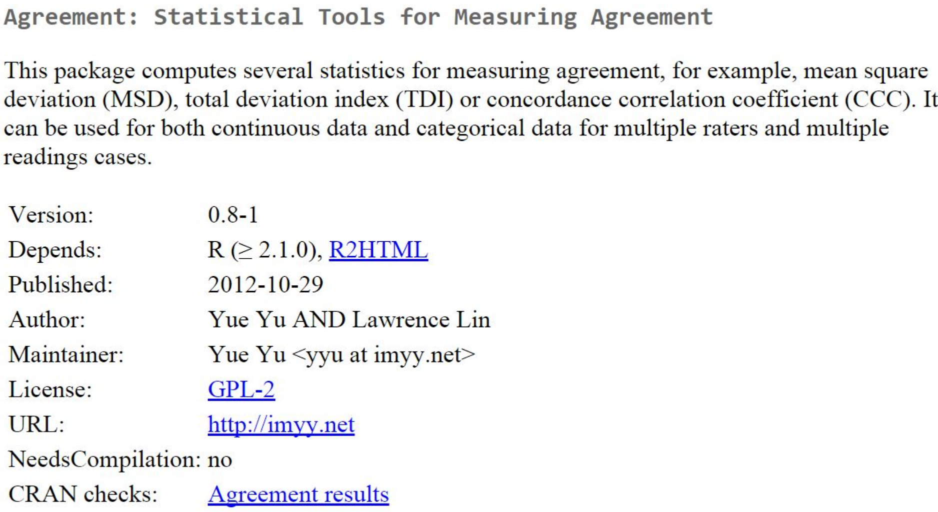
Version:	1.2.1
Depends:	R (≥ 3.0.0), methods
Suggests:	RUnit, XML
Published:	2014-02-12
Author:	Ekaterina Manuilova Andre Schuetzenmeister Fabian Model
Maintainer:	Fabian Model <fabian.model at roche.com>
License:	GPL (≥ 3)
NeedsCompilation:	yes



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.

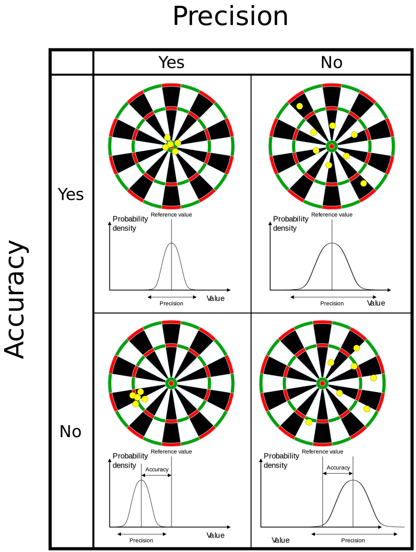
Version:	0.4-0
Depends:	R (≥ 3.0.2), miscF (≥ 0.1-2), lme4 (≥ 1.0-4)
Imports:	R2jags (≥ 0.03-11), coda (≥ 0.16-1)
Published:	2015-07-10
Author:	Dai Feng
Maintainer:	Dai Feng <dai_feng at merck.com>
License:	GPL-2 | GPL-3 [expanded from: GPL]
NeedsCompilation:	no
Materials:	ChangeLog
CRAN checks:	agRee results



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), R2HTML
Published:	2012-10-29
Author:	Yue Yu AND Lawrence Lin
Maintainer:	Yue Yu <yyu at imyy.net>
License:	GPL-2
URL:	http://imyy.net
NeedsCompilation:	no
CRAN checks:	Agreement results

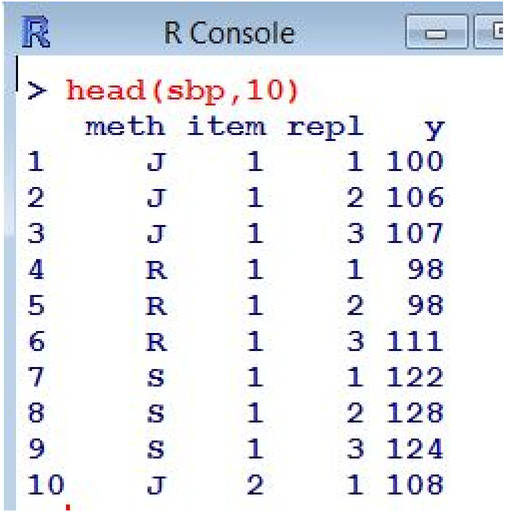


(Wikipedia.org : Accuracy and Precision.svg)

Two Types of Method Comparison Problem

(1) Single Measurement per subject by each method *(straightforward enough problem)*

(2) Multiple Measurement per subject by each method *(Basis of the approaches discussed here)*



```
> head(sbp,10)
   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
```

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 neasurement should have good repeatability

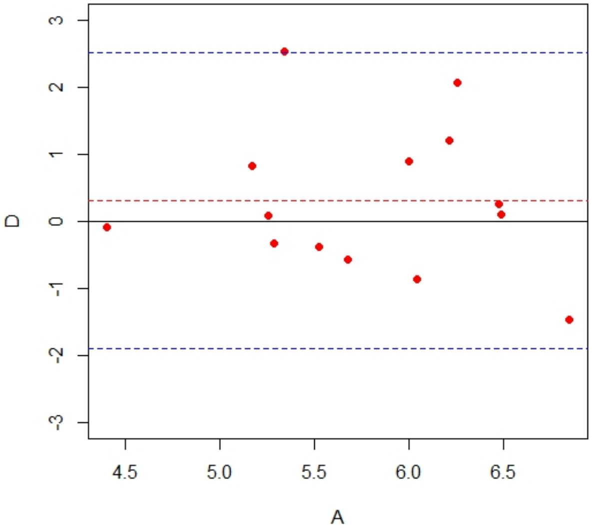
The Bland-Altman Plot

- 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]



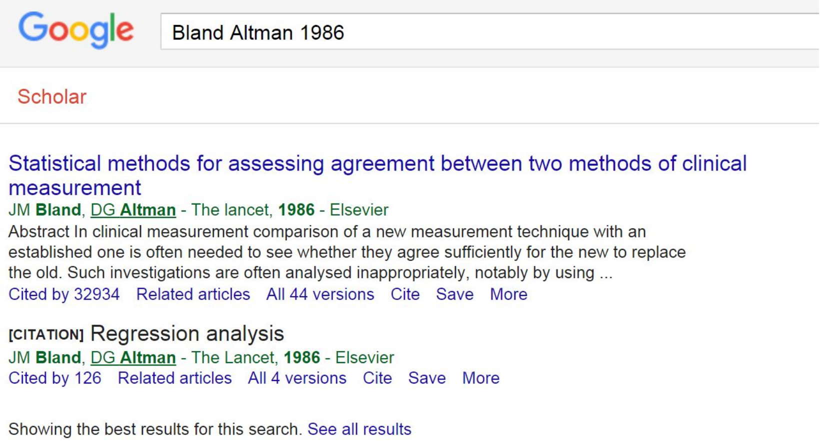
Bland-Altman Plot

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



Statistical methods for assessing agreement between two methods of clinical measurement

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

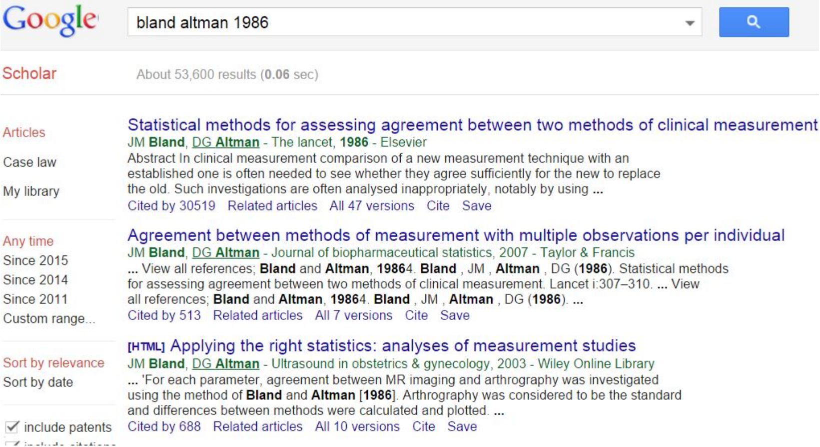
Abstract In clinical measurement comparison of a new measurement technique with an established one is often needed to see whether they agree sufficiently for the new to replace the old. Such investigations are often analysed inappropriately, notably by using ...

[CITATION] Regression analysis

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

Showing the best results for this search. See all results

May 2015 - In excess of 30000 citations



About 53,600 results (**0.06** sec)

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

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**4. **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**4. **Bland** , JM , **Altman** , DG (**1986**). **...**
Cited by 513 Related articles All 7 versions Cite Save

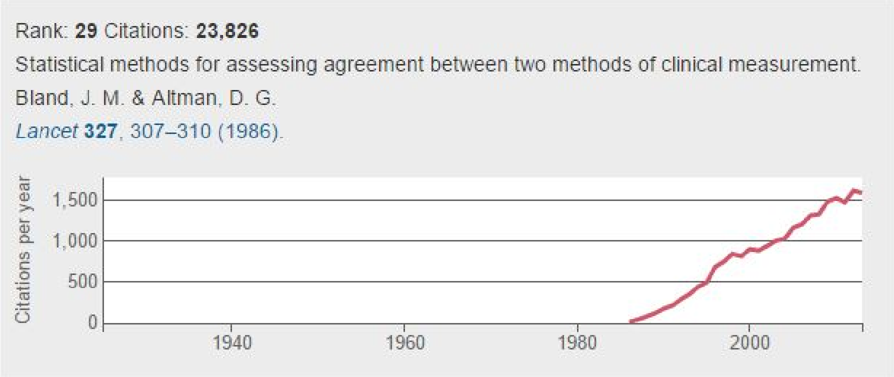
[HTML] Applying the right statistics: analyses of measurement studies
JM **Bland**, DG **Altman** - Ultrasound in obstetrics & gynecology, 2003 - Wiley Online Library
... 'For each parameter, agreement between MR imaging and arthrography was investigated using the method of **Bland** and **Altman** [**1986**]. 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

The top 100 papers

Nature explores the most-cited research of all time.

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

29 October 2014

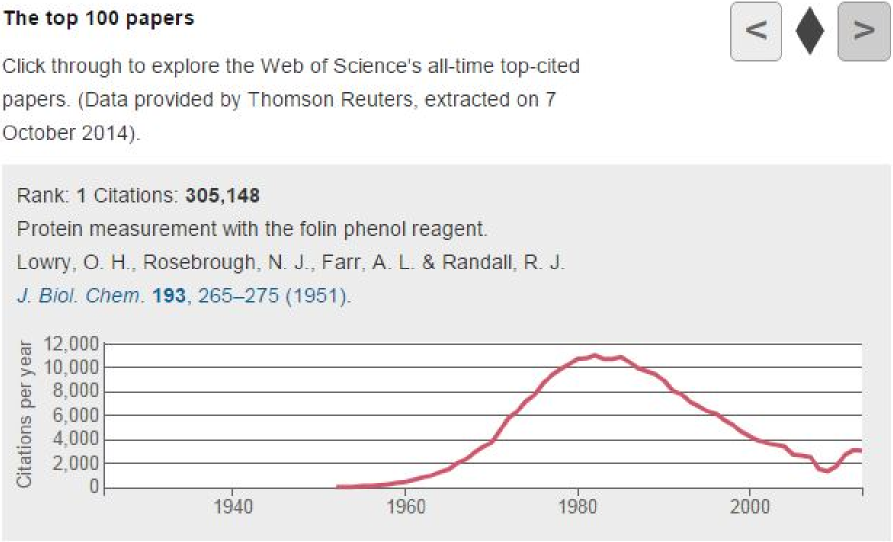


Rank: **29** Citations: **23,826**

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).

Nature.com

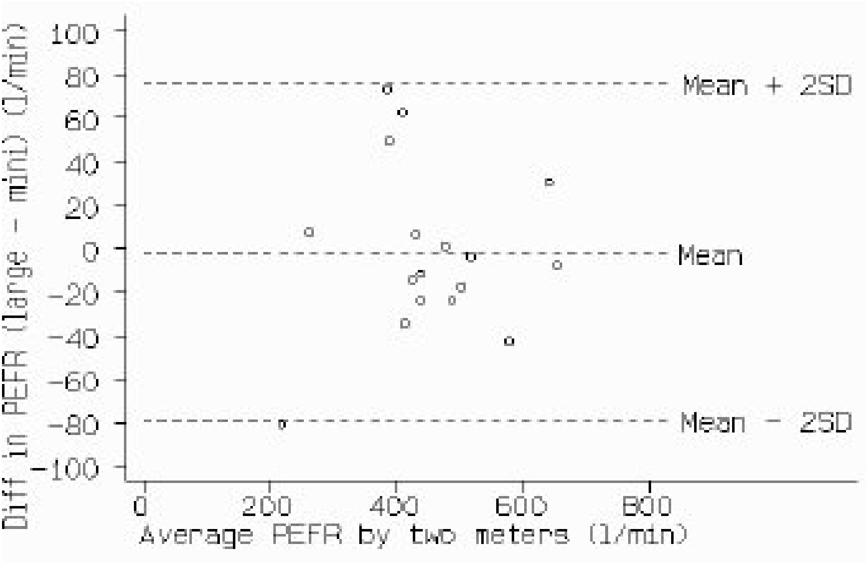
> *The Kaplan Meier paper was a sleeper hit, receiving almost no citations until computing power boomed in the 1970s, making the methods accessible to non-specialists. Simplicity and ease of use also boosted the popularity of papers in this field.*
>
> *British statisticians Martin Bland and Douglas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

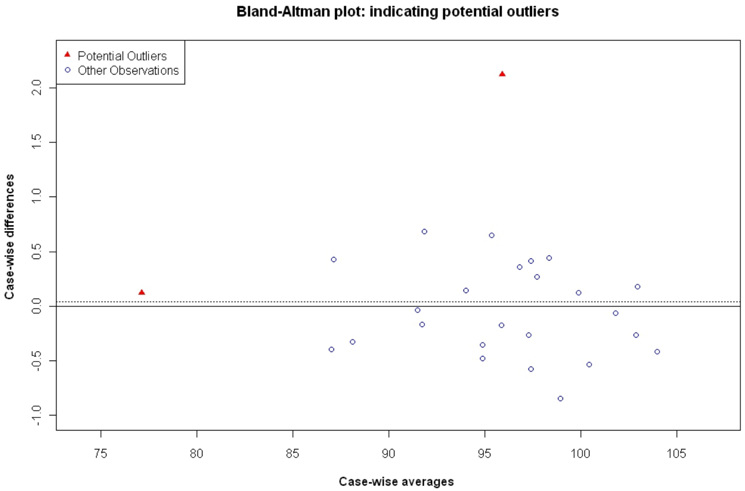
R Packages for Bland-Altman Analysis

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

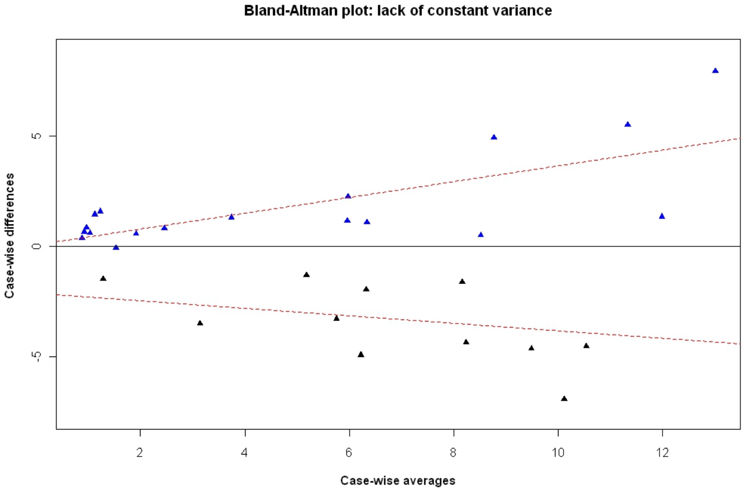
Interpreting the Bland-Altman Plot



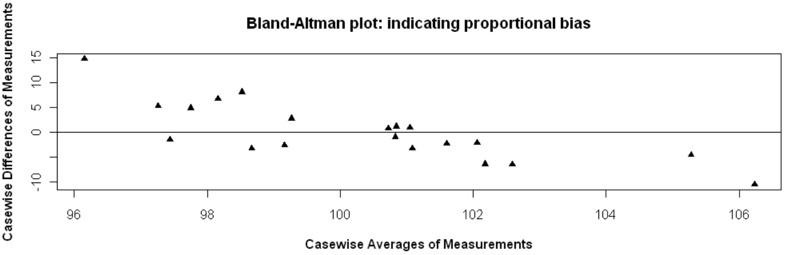
Interpreting the Bland-Altman Plot



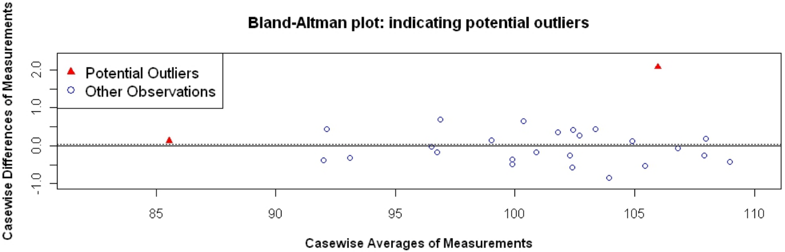
Interpreting the Bland-Altman Plot



Interpreting the Bland-Altman Plot



Bland-Altman plot: indicating proportional bias



Bland-Altman plot: indicating potential outliers

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

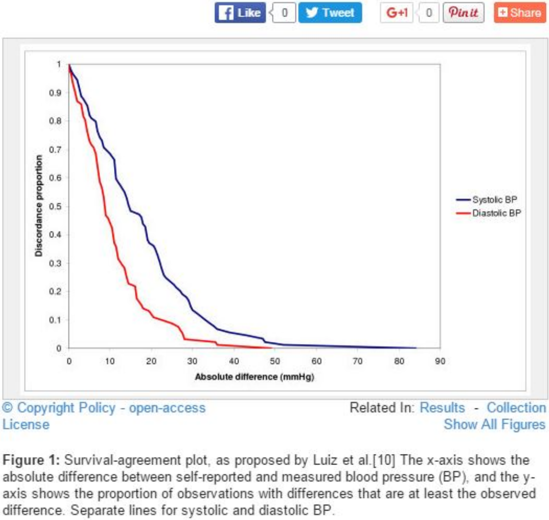


Figure 1: Survival-agreement plot, as proposed by Luiz et al.[10] The x-axis shows the absolute difference between self-reported and measured blood pressure (BP), and the y-axis shows the proportion of observations with differences that are at least the observed difference. Separate lines for systolic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

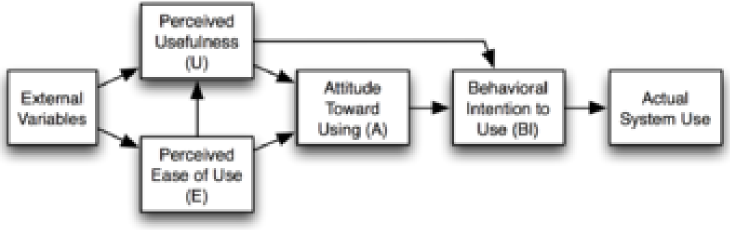


Figure: Technology Acceptance Model Flowchart (Davis,1989)

- 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

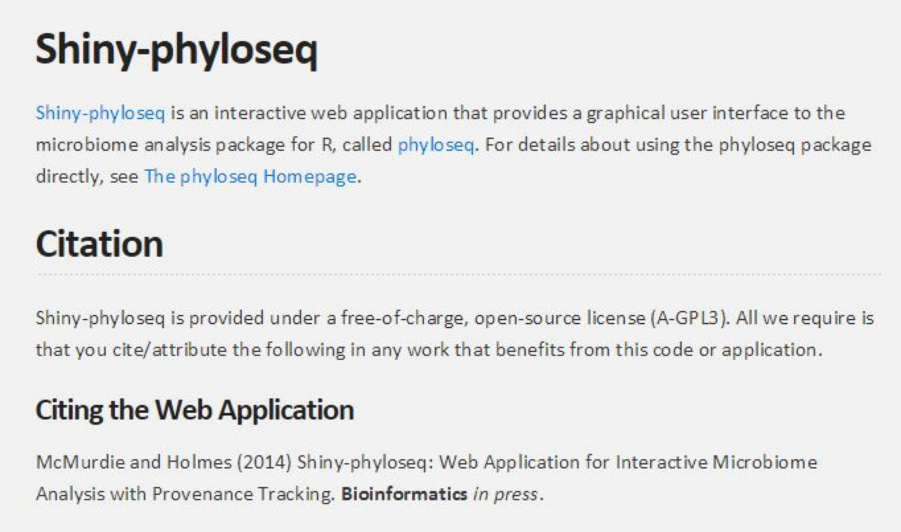
by RStudio

A web application framework for R

Shiny Web Applications with R

Useful Shiny Resources

- shiny.rstudio.com
- showmeshiny.com
- 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".

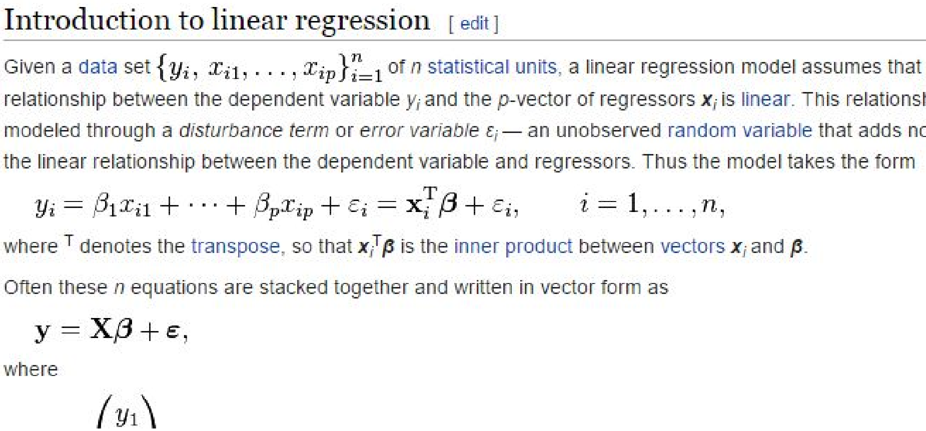
Three Conditions

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 $\{y_i,\, x_{i1}, \ldots, x_{ip}\}_{i=1}^{n}$ of *n* statistical units, a linear regression model assumes that relationship between the dependent variable y_i and the *p*-vector of regressors $\boldsymbol{x}_i$ is linear. This relations modeled through a *disturbance term* or *error variable* ε_i — an unobserved random variable that adds n the linear relationship between the dependent variable and regressors. Thus the model takes the form

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

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

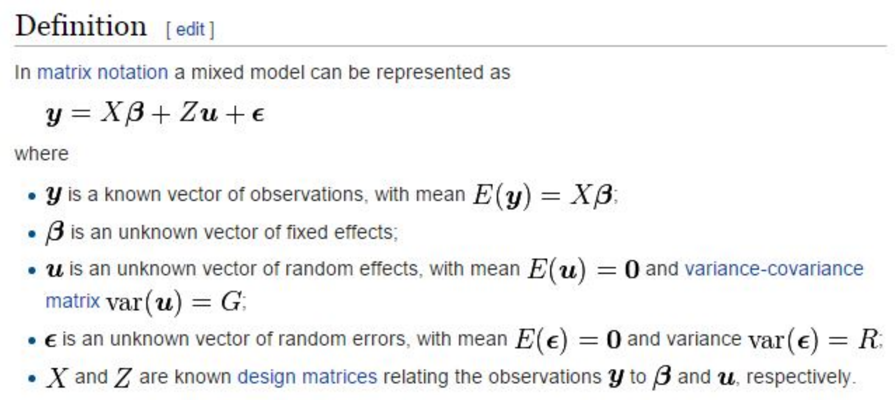
Often these *n* equations are stacked together and written in vector form as

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

where

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

WIKIPEDIA : Linear Mixed Effects Models



Definition [edit]

In matrix notation a mixed model can be represented as

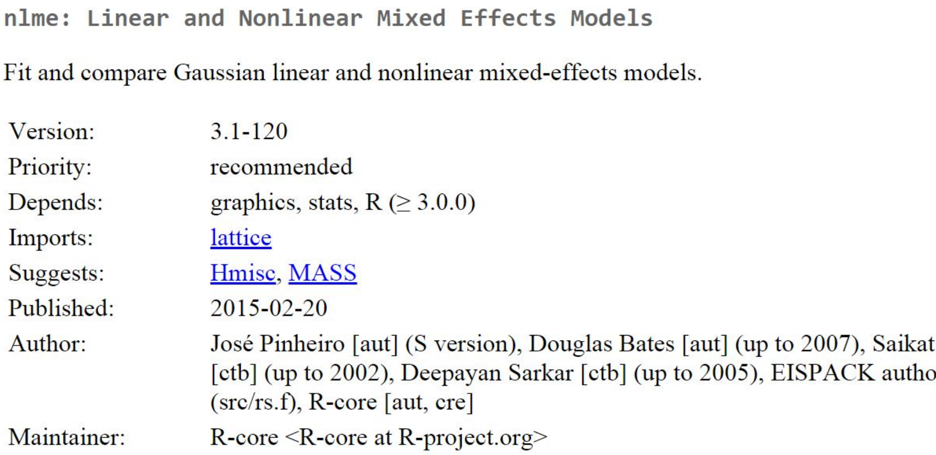
$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{u} + \boldsymbol{\epsilon}$$

where

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

Figure: Wikipedia Entry on LME Models

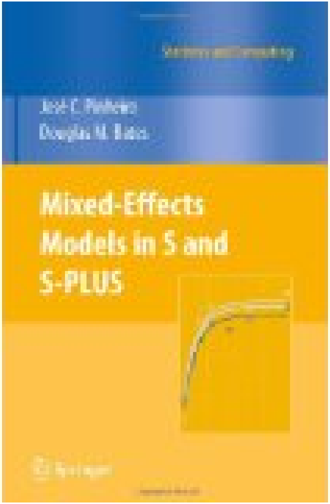
The nlme R package



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 (≥ 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 autho (src/rs.f), R-core [aut, cre]
Maintainer:	R-core <R-core at R-project.org>

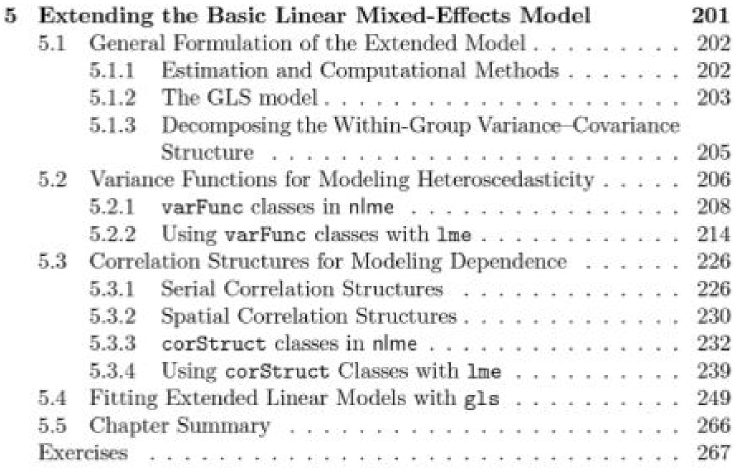


José C. Pinheiro
Douglas M. Bates

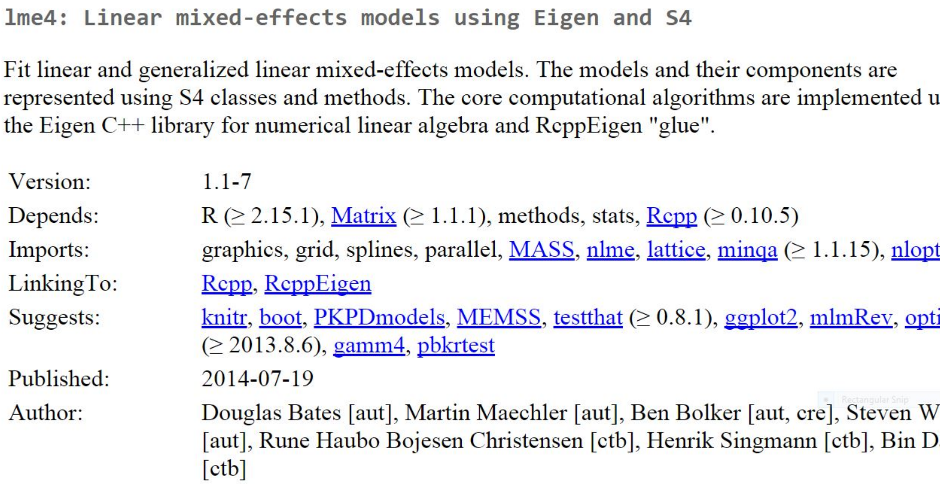
Mixed-Effects Models in S and S-PLUS

Springer

Chapter 5 : Extending the Basic LME Model



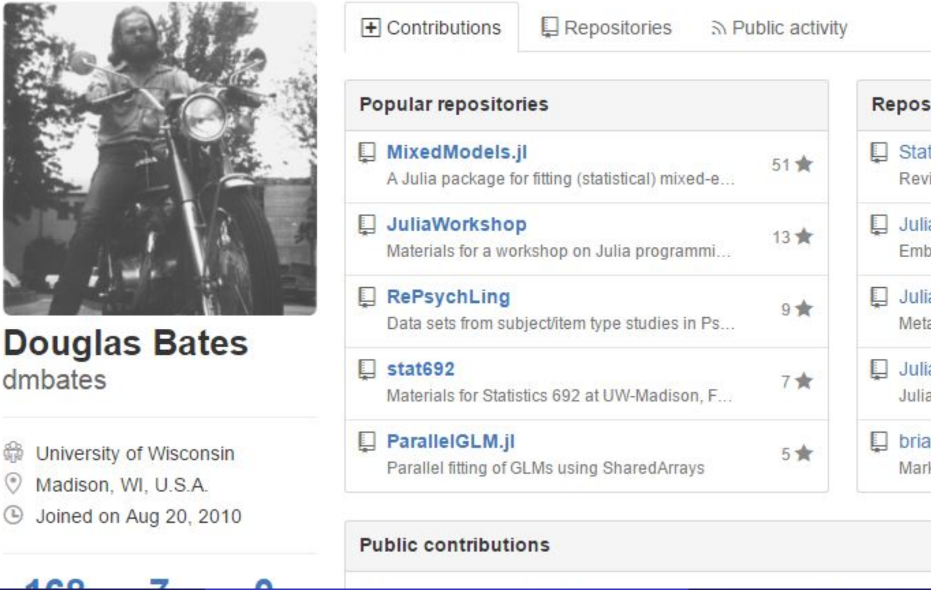
LME4 R Package



`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 u the Eigen C++ library for numerical linear algebra and RcppEigen "glue".

Version:	1.1-7
Depends:	R (≥ 2.15.1), Matrix (≥ 1.1.1), methods, stats, Rcpp (≥ 0.10.5)
Imports:	graphics, grid, splines, parallel, MASS, nlme, lattice, minqa (≥ 1.1.15), nlopt
LinkingTo:	Rcpp, RcppEigen
Suggests:	knitr, boot, PKPDmodels, MEMSS, testthat (≥ 0.8.1), ggplot2, mlmRev, opti (≥ 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

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 ★

Repos

- Stat
  Revi
- Julia
  Emb
- Julia
  Meta
- Julia
  Julia
- bria
  Mark

Public contributions

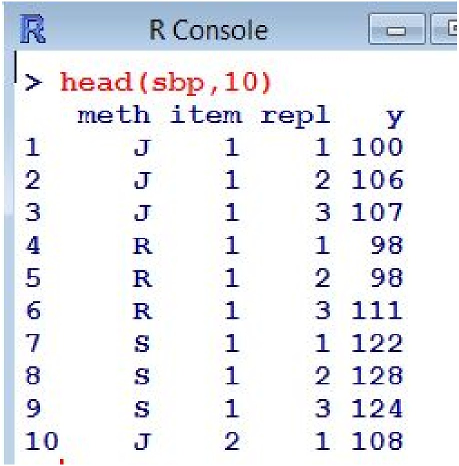
The `nlme` Package

(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.



```
> head(sbp,10)
   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

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

- β_1 and β_2 are fixed effects corresponding to both methods. (β_0 is the intercept.)
- b_{1i} and b_{2i} are random effects corresponding to both methods.

Roy's LME model

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

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

- can be re-expressed as follows:

$$\boldsymbol{Z}_i \left[\begin{array}{cc} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{array} \right] \boldsymbol{Z}_i' + \left(V \otimes \left[\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right] \right)$$

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

$$\text{Block } \boldsymbol{\Omega}_i = \left[\begin{array}{cc} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{array} \right] + \left[\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right].$$

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 σ_1^2 may differ from σ_2^2.
- Compound symmetric (CS) structure specifies that $\sigma_1^2 = \sigma_2^2$.
- In both cases, σ_{12} may take value other than 0.
- *(What is stated here is applicable to* ***D*** *also)*

- Roys uses an LME model approach to provide a set of formal tests for method comparison studies.
- Four candidates 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
..
```

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

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
```

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

Overall variance covariance matrix

- Overall variance

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

- 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.

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