Method Comparison Studies

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

Medical Measurement

Common Scenario for comparing two methods of measurements

Reference Method

Test Method

Methods of Measurement

Comparing against a Gold Standard
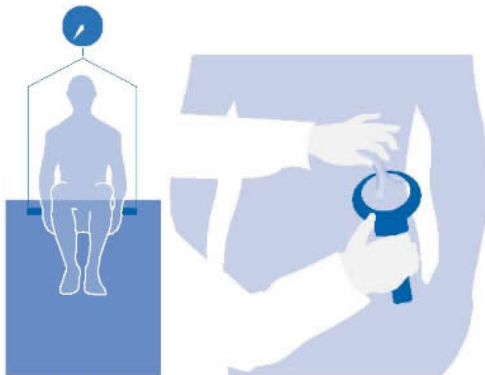
Gold Standards

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 ($\geq$ 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 ($\geq$ 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 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:	RUnit, XML
Published:	2014-02-12
Author:	Ekaterina Manuilova Andre Schuetzenmeister Fabian Model
Maintainer:	Fabian Model <fabian.model at roche.com>
License:	GPL ($\geq$ 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

Precision
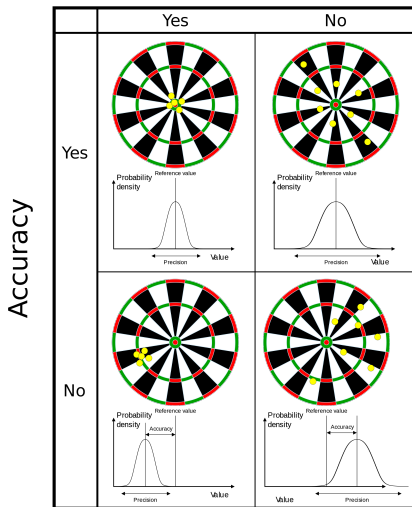
(Wikipedia.org : Accuracy and Precision.svg)

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)*
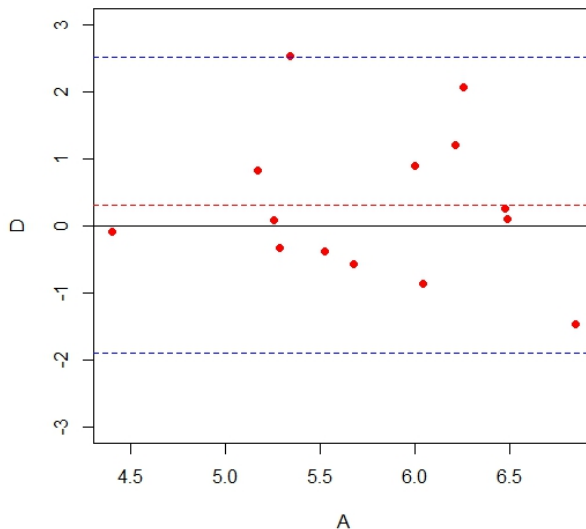
The Bland-Altman Plot

- 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

Google   Bland Altman 1986

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

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

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).
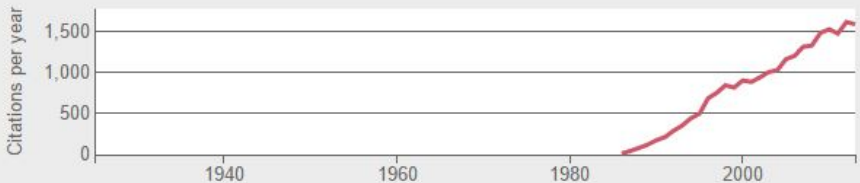
The top 100 papers

< ◆ >

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).
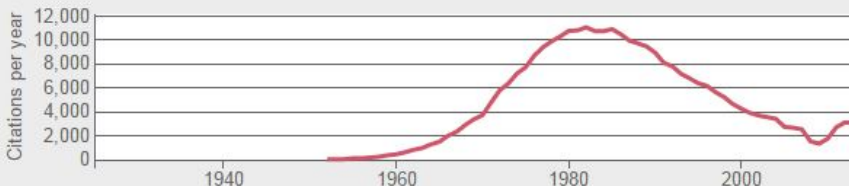
Nature.com

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
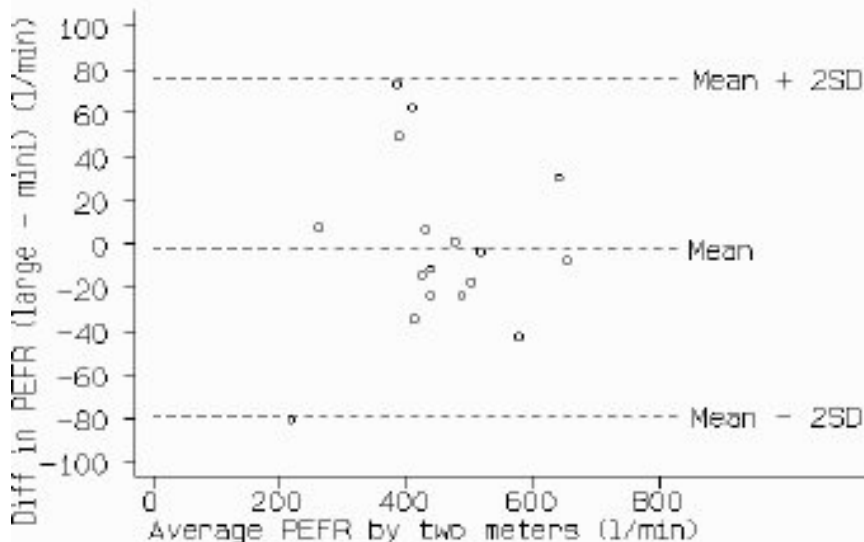
Interpreting the Bland-Altman Plot

Interpreting the Bland-Altman Plot

Bland-Altman plot: indicating potential outliers

Interpreting the Bland-Altman Plot

Bland-Altman plot: lack of constant variance
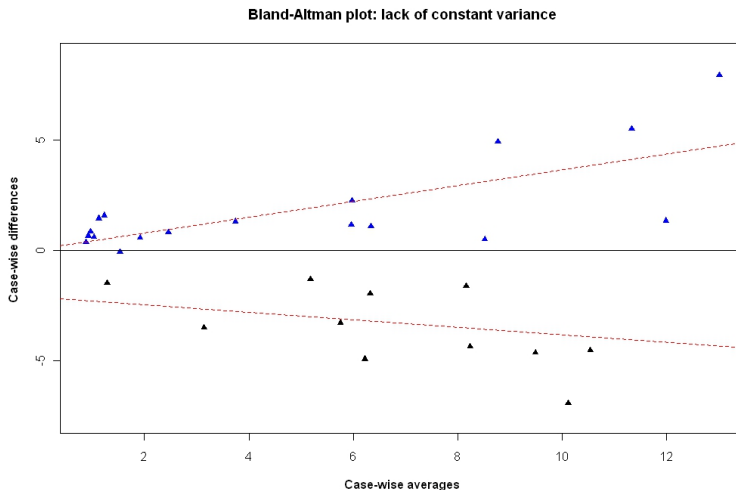
Interpreting the Bland-Altman Plot
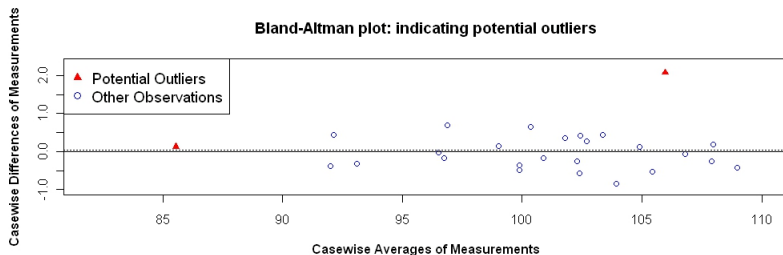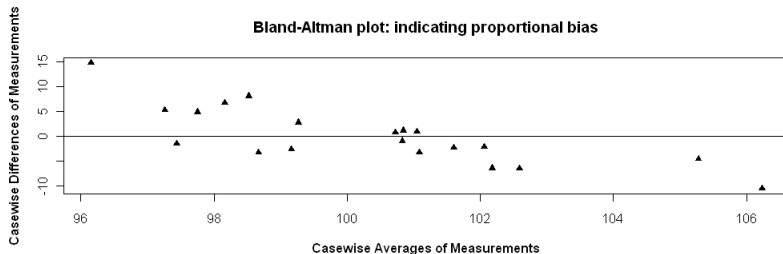
The Bland-Altman Plot: Prevalence

- 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
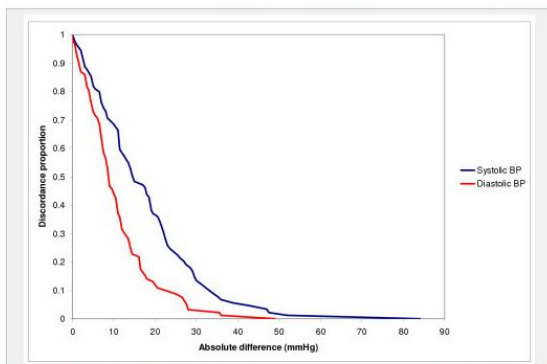
© Copyright Policy - open-access
License

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
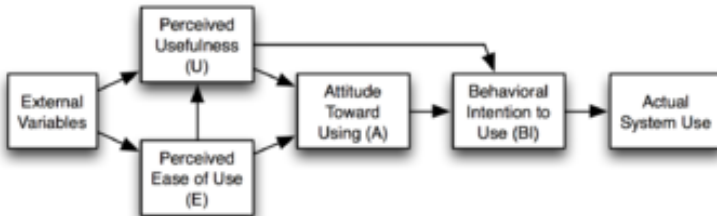
Figure: Technology Acceptance Model Flowchart (Davis,1989)

- 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

- 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)

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 Deb [ctb] (up to 2002), Deepayan Sarkar [ctb] (up to 2005), EISPACK authors [c (src/rs.f), R-core [aut, cre]
Maintainer:	R-core <R-core at R-project.org>
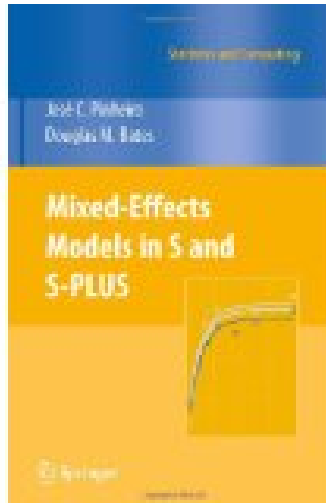
Chapter 5 : Extending the Basic LME Model

LME models

- In a linear mixed-effects 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.

- Can have differing number of replicate measurements for different subjects.

Roy's Approach

- Roy proposes an LME model with Kronecker product covariance structure in a doubly multivariate setup.

- Response for ith 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).

- $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$

- $\boldsymbol{b}_i \sim N_m(0, \boldsymbol{D})$ (m: number of methods)

- $\boldsymbol{\epsilon}_i \sim N_{n_i}(0, \boldsymbol{R})$ (n_i: number of measurements on subject i)

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

$$\left(\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right)$$

- Symmetric structure specifies that σ_1^2 may differ from σ_2^2.
- Compound symmetric structure specifies that $\sigma_1^2 = \sigma_2^2$.
- In both cases, σ_{12} may take value other than 0.

The `nlme` Package

- 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)

- Recall: 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.
- These models are similar to one another, but for the imposition of equality constraints in the nested models

Variance Covariance Matrices

- Under Roy's model, random effects are defined using a bivariate normal distribution.

- Consequently, the variance-covariance structures can be described using 2×2 matrices.

- A discussion of the various structures a variance-covariance matrix can be specified under is required before progressing.

- The following structures are relevant:
  1. the identity structure,
  2. the compound symmetry structure
  3. the symmetric structure.

Variance Covariance Matrices

- The **identity** structure is simply an abstraction of the identity matrix.
- The **compound symmetric** structure and **symmetric** structure can be described with reference to the following matrix (here in the context of the overall covariance Block-Ω_i, but equally applicable to the component variabilities G and Σ);

$$\left(\begin{array}{cc} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{array} \right)$$

```
R R Console

>
>
> summary(sbp)
 meth         item           repl           y
 J:255   Min.   : 1    Min.   :1    Min.   : 74.0
 R:255   1st Qu.:22    1st Qu.:1    1st Qu.:110.0
 S:255   Median :43    Median :2    Median :126.0
         Mean   :43    Mean   :2    Mean   :132.6
         3rd Qu.:64    3rd Qu.:3    3rd Qu.:150.0
         Max.   :85    Max.   :3    Max.   :228.0
>
```

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.

The Nested Model 1

```
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 and symmetric structure within-subject variances.

The Nested Model 2

```
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.

The Nested Model 3

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

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

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)

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
..
```

$$\hat{\boldsymbol{D}} = \left(\begin{array}{cc} 923.97 & 785.34 \\ 785.34 & 971.29 \end{array} \right)$$

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
```

$$\hat{\boldsymbol{\Sigma}} = \left(\begin{array}{cc} 37.40 & 16.06 \\ 16.06 & 83.14 \end{array} \right)$$

Overall variance covariance matrix

- Overall variance

$$\text{Block } \hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{D}} + \hat{\boldsymbol{\Sigma}} = \left(\begin{array}{cc} 961.38 & 801.40 \\ 801.40 & 1054.43 \end{array} \right)$$

- 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 symmetry 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