Statistical Analysis of Method Comparison studies

Bendix Carstensen  Steno Diabetes Center, Gentofte, Denmark & Dept. Biostatistics, Medical Faculty, University of Copenhagen
http://BendixCarstensen.com

Claus Thorn Ekstrøm  Statistics, Faculty of Life Sciences, University of Copenhagen
www.statistics.life.ku.dk/~ekstrom/

Tutorial, SISMEC, Ancona, Italy
28 September 2011

http://BendixCarstensen.com/MethComp/Ancona.2011
Comparing two methods with one measurement on each Morning

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
Comparing measurement methods

General questions:

- Are results systematically different?
- Can one method safely be replaced by another?
- What is the size of measurement errors?
- Different centres use different methods of measurement: How can we convert from one method to another?
- How precise is the conversion?
Two methods for measuring fat content in human milk:

The relationship looks like:

$$y_1 = a + by_2$$
Two methods — one measurement by each

How large is the difference between a measurement with method 1 and one with method 2 on a (randomly chosen) person?

\[ D_i = y_{2i} - y_{1i}, \quad \bar{D}, \quad \text{s.d.}(D) \]

“Limits of agreement:”

\[ \bar{D} \pm 2 \times \text{s.d.}(D) \]

95% prediction interval for the difference between a measurement by method 1 and one by method 2. [1, 2]
Limits of agreement: Interpretation

▸ If a new patient is measured once with each of the two methods, the difference between the two values will with 95% probability be within the limits of agreement.

▸ This is a prediction interval for a (future) difference.

▸ Requires a clinical input: Are the limits of agreement sufficiently narrow to make the use of either of the methods clinically acceptable?

▸ Is it relevant to test if the mean is 0?
Limits of agreement: Test?

Testing whether the difference is 0 is a bad idea:

- If the study is sufficiently small this will be accepted even if the difference is important.
- If the study is sufficiently large this will be rejected even if the difference is clinically irrelevant.
Limits of agreement:

Comparing two methods with one measurement on each

Plot differences ($D_i$) versus averages ($A_i$).
Model in “Limits of agreement”

Methods $m = 1, \ldots, M$, applied to $i = 1, \ldots, I$ individuals:

$$y_{mi} = \alpha_m + \mu_i + e_{mi}$$

$$e_{mi} \sim \mathcal{N}(0, \sigma_{m}^2) \quad \text{measurement error}$$

- Two-way analysis of variance model, with unequal variances in columns.
- Different variances are not identifiable without replicate measurements for $M = 2$ because the variances cannot be separated.
Limits of agreement:

Usually interpreted as the likely difference between two future measurements, one with each method:

\[ y_2 - y_1 = \hat{D} = \alpha_2 - \alpha_1 \pm 1.96 \text{s.d.}(D) \]

Normally we use 2 instead of 1.96. Neither are formally correct if we take the model seriously:

- Use a t-quantile with \( I - 1 \) d.f.
- Estimation s.d. of \( \alpha_2 - \alpha_1 \) is \( \sigma / \sqrt{I} \).

So we should use \( t_{0.95} \times \sqrt{(I + 1)/I} \) instead. This is 2.08 for \( I = 30 \) and less than 2 if \( I > 85 \).
Limits of agreement:

Limits of agreement can be converted to a prediction interval for $y_2$ given $y_1$, by solving for $y_2$:

$$y_2 - y_1 = \alpha_2 - \alpha_1 \pm 2 \text{s.d.}(D)$$

which gives:

$$\hat{y}_2|_1 = \hat{y}_2|y_1 = \alpha_2 - \alpha_1 + y_1 \pm 2 \text{s.d.}(D)$$
Introduction to computing

Morning

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
Structure of practicals

This tutorial is both theoretical and practical, i.e. the aim is to convey a basic understanding of the problems in method comparison studies, but also to convey practical skills in handling the statistical analysis.

- \textbf{\texttt{R}} for data manipulation and graphics.
- So we assume familiarity with \texttt{R}.
- Occasionally \texttt{BUGS} for estimation in non-linear variance component models.
- \texttt{BUGS} is hidden inside an \texttt{R}-function.
How it works

Example data sets are included in the MethComp package.

Functions in MethComp are based on a data frame with a particular structure; a Meth object:

meth — method (factor)
item — item, person, individual, sample (factor)
repl — replicate (if present) (factor)
y — the actual measurement (numerical)

Once converted to Meth, just use summary, plot etc.
> subset(ox, as.integer(item) < 3)
Getting your own data into R

Take a look in “The R Primer” by Claus Ekstrøm, or:

If your data are not too large, the simplest is to edit your data in Excel or some other spreadsheet to look like this:

```
item repl  id  CO  pulse
1   1  1.1  78.0  71
1   2  1.2  76.4  72
1   3  1.3  77.2  73
2   1  2.1  68.7  68
2   2  2.2  67.6  67
2   3  2.3  68.3  68
```

The first line is variable names; the following lines are data.
Analysis options in this course

- Scatter plots.
- Bland-Altman plots \(((y_2 - y_1) \text{ vs. } (y_1 + y_2)/2)\)
- Limits of Agreement (LoA).
- Models with constant bias.
- Models with linear bias.
- Conversion formulae between methods (single replicates)
- Transformation of measurements.
- Plots of conversion equations.
- Reporting of variance components.
Requirements

- **R** for data manipulation and graphics.
- **Keep a script of what you did:**
  - Use the built-in editor in **R**
  - the nerds can use ESS
  - or you can download **R-Studio**.
- **You need the packages:**
  - MethComp
  - R2WinBUGS
  - coda
  - BRugs
  - **Epi - Version 1.10 !!!**
Functions in the MethComp package

5 broad categories of functions in MethComp:

- Graphical — exploring data.
- Data manipulation — reshaping and changing.
- Simulation — generating datasets or replacing variables.
- Analysis functions — fitting models to data.
- Reporting functions — displaying results from analyses.

Overview of these in the Practicals.
Does it work?

library(MethComp)
library(help=MethComp)  # Do you have version 1.10??
data(ox)
ox <- Meth(ox)
summary(ox)
plot(ox)
BA.plot(ox)
BA.est(ox)
( AR.ox <- AltReg(ox,linked=TRUE,trace=TRUE) )
MCmcmc(ox,code.only=TRUE)
MC.ox <- MCmcmc(ox,n.iter=100)
MethComp(MC.ox)
plot(MC.ox)
plot(MC.ox)
trace.MCmcmc(MC.ox)
post.MCmcmc(MC.ox)
Non-constant difference
Morning

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
Limits of agreement — assumptions

- The difference between methods is constant
- The variances of the methods (and hence of the difference) is constant.

Check this by:

- Regress differences on averages.
- Regress absolute residuals from this on the averages.
Glucose measurements

Non-constant difference
Glucose measurements

Non-constant difference
Regress difference on average

\[ D_i = a + bA_i + e_i, \quad \text{var}(e_i) = \sigma_D^2 \]

If \( b \) is different from 0, we could use this equation to derive LoA:

\[ a + bA_i \pm 2\sigma_D \]

or convert to prediction as for LoA:

\[ y_{2|1} = y_1 + a + bA_i \approx y_1 + a + by_1 = a + (1 + b)y_1 \]

Exchanging methods would give:

\[ y_{1|2} = -a + (1 - b)y_1 \]

instead of:

\[ y_{1|2} = \frac{-a}{1 + b} + \frac{1}{1 + b}y_1 \]
Variable limits of agreement

\[ \text{Non-constant difference} \]
Improving the regression of $D$ on $A$

$$y_{2i} - y_{1i} = a + b(y_{1i} + y_{2i})/2 + e_i$$

$$y_{2i}(1 - b/2) = a + (1 + b/2)y_{1i} + e_i$$

$$y_{2i} = \frac{a}{1 - b/2} + \frac{1 + b/2}{1 - b/2}y_{1i} + \frac{1}{1 - b/2}e_i$$

$$y_{1i} = \frac{-a}{1 + b/2} + \frac{1 - b/2}{1 + b/2}y_{2i} + \frac{1}{1 + b/2}e_i$$

This is what comes out of the functions DA.reg and BA.plot
Variable limits of agreement

plasma = $-2.49 + 1.38 \text{capil}$ (95% p.i.: $+/−2.57$)
capil = $1.80 + 0.73 \text{plasma}$ (95% p.i.: $+/−1.87$)
capil−plasma = $2.09 − 0.32 (\text{capil+plasma})/2$ (95% p.i.: $+/−2.16$)

Non-constant difference
Conversion equation with prediction limits

![Graph showing the relationship between Venous plasma and Capillary blood. The graph includes a scatter plot with a regression line and prediction limits. The x-axis represents Venous plasma, and the y-axis represents Capillary blood. The data points are scattered across the graph, with some lying close to the regression line, indicating a strong correlation. The non-constant difference is highlighted, showing a variation in the response across different levels of Venous plasma.](image-url)
Prediction intervals

- Prediction s.e. for $y_{1|2}$ is $\sigma/(1 - b/2)$
- Prediction s.e. for $y_{1|2}$ is $\sigma/(1 + b/2)$
- The slope of the prediction line is the ratio of the prediction s.e.s.
- Hence prediction limits can be used both ways:
Conversion equation with prediction limits

Non-constant difference 27/90
Why does this work?

The general model for the data is:

\[ y_{1i} = \alpha_1 + \beta_1 \mu_i + e_{1i}, \quad e_{1i} \sim \mathcal{N}(0, \sigma^2_1) \]
\[ y_{2i} = \alpha_2 + \beta_2 \mu_i + e_{2i}, \quad e_{2i} \sim \mathcal{N}(0, \sigma^2_2) \]

- Work out the prediction of \( y_1 \) given an observation of \( y_2 \) in terms of these parameters.
- Work out how differences relate to averages in terms of these parameters.
- Then the prediction is as we just derived it.
Why is it wrong anyway?

- Introducing linear bias, \( y_{mi} = \alpha_m + \beta_m \mu_i + e_{mi} \) puts measurements by different methods on different scales. Hence it has formally no meaning to form the differences.

- In the induced model for \( D_i \sim a + bA_i + e_i \), \( e_i \) and \( A_i \) are not independent.

- But if \( \beta \) is not too far from 1 it not a big problem, though.
Comparing two methods with replicate measurements

Morning

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
**Replicate measurements**

**Fat data; exchangeable replicates:**

<table>
<thead>
<tr>
<th>item</th>
<th>repl</th>
<th>KL</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4.5</td>
<td>4.9</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4.4</td>
<td>5.0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4.7</td>
<td>4.8</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6.4</td>
<td>6.5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6.2</td>
<td>6.4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6.5</td>
<td>6.1</td>
</tr>
</tbody>
</table>

**Oximetry data; linked replicates:**

<table>
<thead>
<tr>
<th>item</th>
<th>repl</th>
<th>CO</th>
<th>pulse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>78.0</td>
<td>71</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>76.4</td>
<td>72</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>77.2</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>68.7</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>67.6</td>
<td>67</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>68.3</td>
<td>68</td>
</tr>
</tbody>
</table>

**Linked or exchangeable replicates!**
Extension of the model: exchangeable replicates

\[ y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir} \]

\[ \text{s.d.}(c_{mi}) = \tau_m \quad \text{— “matrix”-effect} \]

\[ \text{s.d.}(e_{mir}) = \sigma_m \quad \text{— measurement error} \]

- Replicates within \((m, i)\) are needed to separate \(\tau\) and \(\sigma\).
- Even with replicates, the separate \(\tau\)s are only estimable if \(M > 2\).
- Still assumes that the difference between methods is constant.
- Assumes exchangeability of replicates.
Extension of the model: linked replicates

\[ y_{mir} = \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \]

- \( s.d.(a_{ir}) = \omega \) — between replicates
- \( s.d.(c_{mi}) = \tau_m \) — “matrix”-effect
- \( s.d.(e_{mir}) = \sigma_m \) — measurement error

- Still assumes that the difference between methods is constant.
- Replicates are \textit{linked} between methods:
  \( a_{ir} \) is common across methods, i.e. the first replicate on a person is made under similar conditions for all methods (i.e. at a specific day or the like).
Replicate measurements

Three approaches to limits of agreement with replicate measurements:

1. Take means over replicates within each method by item stratum.
2. Replicates within item are taken as items.
3. Fit the correct variance components model and use this as basis for the LoA.
   The model is fitted using:
   > BA.est( data, linked=TRUE ).
Oximetry data

Comparing two methods with replicate measurements
Replicate measurements

- The limits of agreement should still be for difference between future **single** measurements.
- Analysis based on the **means** of replicates is therefore **wrong**:
- Model:
  \[ y_{mirt} = \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \]
  \[
  \text{var}(y_{1jr} - y_{2jr}) = \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2
  \]
  — note that the term \( a_{ir} - a_{ir} \) cancels because we are referring to the **same** replicate.
Wrong or almost right

In the model the correct limits of agreement would be:

\[ \alpha_1 - \alpha_2 \pm 1.96 \sqrt{\tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2} \]

But if we use means of replicates to form the differences we have:

\[ \bar{d}_i = \bar{y}_{1i} - \bar{y}_{2i} = \alpha_1 - \alpha_2 + \frac{\sum_r a_{ir}}{R_{1i}} - \frac{\sum_r a_{ir}}{R_{2i}} + c_{1i} - c_{2i} + \frac{\sum_r e_{1ir}}{R_{1i}} - \frac{\sum_r e_{2ir}}{R_{2i}} \]
The terms with \( a_{ir} \) are only relevant for linked replicates in which case \( R_{1i} = R_{2i} \) and therefore the term vanishes. Thus:

\[
\text{var}(\bar{d}_i) = \frac{\tau_1^2 + \tau_2^2 + \sigma_1^2}{R_{1i}} + \frac{\sigma_2^2}{R_{2i}} < \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2
\]

so the limits of agreement calculated based on the means are much too narrow as prediction limits for differences between future *single* measurements.
(Linked) replicates as items

If replicates are taken as items, then the calculated differences are:

\[ d_{ir} = y_{1ir} - y_{2ir} = \alpha_1 - \alpha_2 + c_{1i} - c_{2i} + e_{1ir} - e_{2ir} \]

which has variance \( \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2 \), and so gives the correct limits of agreement. However, the differences are not independent:

\[ \text{cov}(d_{ir}, d_{is}) = \tau_1^2 + \tau_2^2 \]

Negligible if the residual variances are very large compared to the interaction, variance likely to be only slightly downwards biased.
Recommendations

- Fit the correct model, and get the estimates from that, e.g. by using `BA.est`.
- If you must use over-simplified methods:
  - Use linked replicates as item.
  - If replicates are not linked; make a random linking.

Note: If this give a substantially different picture than using the original replicate numbering as linking key, there might be something fishy about the data.

Further details, see [3].
Oximetry data

Linked replicates used as items

Mean over replicates as items

Limits based on model — dashed line assuming exchangeable replicates

Comparing two methods with replicate measurements
Repeatability and reproducibility
Wednesday 9 February, morning

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
Accuracy of a measurement method

- **Repeatability:**
The accuracy of the method under exactly similar circumstances; i.e. the same lab, the same technician, and the same day. *(Repeatability conditions)*

- **Reproducibility:**
The accuracy of the method under comparable circumstances, i.e. the same machinery, the same kit, but possibly different days or laboratories or technicians. *(Reproducibility conditions)*
Quantification of accuracy

- Upper limit of a 95% confidence interval for the difference between two measurements.
- Suppose the variance of the measurement is $\sigma^2$:
  \[
  \text{var}(y_{m1} - y_{m2}) = 2\sigma^2
  \]
  i.e the standard error is $\sqrt{2}\sigma$, and a confidence interval for the difference:
  \[
  0 \pm 1.96 \times \sqrt{2}\sigma = 0 \pm 2.772\sigma \approx 2.8\sigma
  \]
- This is called the reproducibility coefficient or simply the reproducibility. (The number 2.8 is used as a convenient approximation).
Quantification of accuracy

- Where do we get the $\sigma$?
- Repeat measurements on the same item (or even better) several items.
- The conditions under which the repeat (replicate) measurements are taken determines whether we are estimating repeatability or reproducibility.
- In larger experiments we must consider the exchangeability of the replicates — i.e. which replicates are done under (exactly) similar conditions and which are not.
Extension with non-constant bias

\[ y_{mir} = \alpha_m + \beta_m \mu_i + \text{random effects} \]

There is now a scaling between the methods. Methods do not measure on the same scale — the relative scaling is estimated, between method 1 and 2 the scale is \( \beta_2 / \beta_1 \).

Consequence: Multiplication of all measurements on one method by a fixed number does not change results of analysis:

The corresponding \( \beta \) is multiplied by the same factor as is the variance components for this method.
Variance components

Two-way interactions:

\[ y_{mir} = \alpha_m + \beta_m (\mu_i + a_{ir} + c_{mi}) + e_{mir} \]

The random effects \( c_{mi} \) and \( e_{mir} \) have variances specific for each method.

But \( a_{ir} \) does not depend on \( m \) — must be scaled to each of the methods by the corresponding \( \beta_m \).

Implies that \( \omega = \text{s.d.}(a_{ir}) \) is irrelevant — the scale is arbitrary. The relevant quantities are \( \beta_m \omega \) — the between replicate variation within item as measured on the \( m \)th scale.
\[ y_{mir} = \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \]
\[ y_{mir} = \alpha_m + \beta_m (\mu_i + a_{ir} + c_{mi}) + e_{mir} \]
Estimation: Alternating random effects regression
Morning

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
Alternating random effects regression

Carstensen [4] proposed a ridiculously complicated approach to fit the model

\[ y_{mir} = \alpha_m + \beta_m \mu_i + c_{mi} + e_{mir} \]

based in the observation:

- For fixed \( \mu \) the model is a linear mixed model.
- For fixed \((\alpha, \beta)\) it is a regression through 0.

This has be improved in [5]
Alternating random effects regression

Now consider instead the correctly formulated version of the slightly more general model:

\[ y_{mir} = \alpha_m + \beta_m (\mu_i + a_{ir} + c_{mi}) + e_{mir} \]

Here we observe

- For fixed \( \zeta_{mir} = \mu_i + a_{ir} + c_{mi} \) the model is a linear model, with residual variances different between methods.
- For fixed \( (\alpha, \beta) \) scaled responses \( y \) are used:

\[
\frac{y_{mir} - \alpha_m}{\beta_m} = \mu_i + a_{ir} + c_{mi} + e_{mir} / \beta_m
\]
Estimation algorithm

\[ y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir} \]

1. Start with \( \zeta_{mir} = \bar{y}_{mi} \).
2. Estimate \((\alpha_m, \beta_m)\).
3. Compute the scaled responses and fit the random effects model.
4. Use the estimated \( \mu_i \)s, and BLUPs of \( a_{ir} \) and \( c_{mi} \) to update \( \zeta_{mir} \).
5. Check convergence in terms of identifiable parameters.
The residual variances

- The variance components are estimated in the model for the scaled response.
- The parameters \((\alpha_m, \beta_m)\) are not taken into account in the calculation of the residual variance.
- Hence the residual variances must be corrected *post hoc*.
- This machinery is implemented in the function AltReg in the MethComp package.
> AR.ox <- AltReg(ox, linked=T, trace=T)
AltReg uses 354 obs. out of 354 in the supplied data.

iteration 1 criterion: 1

\[
\begin{array}{cccccccc}
\text{alpha} & \text{beta} & \text{sigma} & \text{Intercept: CO} & \text{Slope: CO} & \text{Slope: pulse} & \text{IxR} & \text{sd. MxI} & \text{sd. res.sd.} \\
\text{CO} & 0.911 & 0.988 & 1.861 & 74.419 & 74.417 & 1.000 & 0.974 & 3.371 & 3.502 & 2.292 \\
pulse & -1.039 & 1.014 & 1.860 & 74.422 & 74.419 & 1.027 & 1.000 & 3.460 & 3.595 & 3.958 \\
\end{array}
\]

... 

iteration 14 criterion: 0.000986339

\[
\begin{array}{cccccccc}
\text{alpha} & \text{beta} & \text{sigma} & \text{Intercept: CO} & \text{Slope: CO} & \text{Slope: pulse} & \text{IxR} & \text{sd. MxI} & \text{sd. res.sd.} \\
\text{CO} & -20.548 & 1.281 & 1.027 & 74.419 & 76.938 & 1.000 & 1.063 & 3.521 & 2.978 & 2.055 \\
pulse & -17.301 & 1.205 & 3.308 & 72.049 & 74.419 & 0.941 & 1.000 & 3.313 & 2.802 & 4.079 \\
\end{array}
\]

There were 14 warnings (use warnings() to see them)

> round(AR.ox,3)

\[
\begin{array}{cccccccc}
\text{From} & \text{To} & \text{Intercept: CO} & \text{Slope: CO} & \text{Slope: pulse} & \text{IxR} & \text{sd. MxI} & \text{sd. res.sd.} \\
\text{CO} & 0.000 & -2.159 & 1.000 & 1.063 & 3.521 & 2.978 & 2.055 \\
pulse & 2.031 & 0.000 & 0.941 & 1.000 & 3.313 & 2.802 & 4.079 \\
\end{array}
\]
Your turn:

Start on the practical titled:

“Oximetry: Linked replicates with non-constant bias”
Converting between methods
Afternoon

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
Predicting method 2 from method 1

\[
y_{10r} = \alpha_1 + \beta_1 (\mu_0 + a_{0r} + c_{10}) + e_{10r}
\]
\[
y_{20r} = \alpha_2 + \beta_2 (\mu_0 + a_{0r} + c_{20}) + e_{20r}
\]
\[
\downarrow
\]
\[
y_{20r} = \alpha_2 + \frac{\beta_2}{\beta_1} (y_{10r} - \alpha_1 - e_{10r})
\]
\[
\quad + \beta_2 (-c_{10} + c_{20}) + e_{20r}
\]

The random effects have expectation 0, so:

\[
E(y_{20} | y_{10}) = \hat{y}_{20} = \alpha_2 + \frac{\beta_2}{\beta_1} (y_{10} - \alpha_1)
\]
\[ y_{20r} = \alpha_2 + \frac{\beta_2}{\beta_1} (y_{10r} - \alpha_1 - e_{10r}) + \beta_2 (-c_{10} + c_{20}) + e_{20r} \]

\[
\text{var}(\hat{y}_{20}|y_{10}) = \left(\frac{\beta_2}{\beta_1}\right)^2 (\beta_1^2 \tau_1^2 + \sigma_1^2) + (\beta_2^2 \tau_2^2 + \sigma_2^2)
\]

The slope of the prediction line from method 1 to method 2 is \( \beta_2/\beta_1 \).

The width of the prediction interval is:

\[
2 \times 2 \times \sqrt{\left(\frac{\beta_2}{\beta_1}\right)^2 (\beta_1^2 \tau_1^2 + \sigma_1^2) + (\beta_2^2 \tau_2^2 + \sigma_2^2)}
\]
If we do the prediction the other way round \((y_1 | y_2)\) we get the same relationship i.e. a line with the inverse slope, \(\beta_1 / \beta_2\).

The width of the prediction interval in this direction is (by permutation of indices):

\[
2 \times 2 \times \sqrt{\left( \beta_1^2 \tau_1^2 + \sigma_1^2 \right) + \left( \frac{\beta_1}{\beta_2} \right)^2 \left( \beta_2^2 \tau_2^2 + \sigma_2^2 \right)}
\]

\[
= 2 \times 2 \times \frac{\beta_1}{\beta_2} \sqrt{\left( \frac{\beta_2}{\beta_1} \right)^2 \left( \beta_1^2 \tau_1^2 + \sigma_1^2 \right) + \left( \beta_2^2 \tau_2^2 + \sigma_2^2 \right)}
\]

i.e. if we draw the prediction limits as straight lines they can be used both ways.
pulse = 2.11 + 0.94 CO
( 6.00 )

CO = −2.25 + 1.06 pulse
( 6.39 )

Converting between methods
What happened to the curvature?

Usually the prediction limits are curved:

\[
\hat{y} | x \pm t_{0.975} \times \hat{\sigma} \sqrt{1 + x'x}
\]

In our prediction we have ignored the last term \((x'x)\), i.e. effectively assuming that there is no estimation error on \(\alpha_{2|1}\) and \(\beta_{2|1}\).
Transformation of data
Afternoon

Bendix Carstensen

MethComp
28 September 2011
Tutorial, SISMEC, Ancona, Italy

http://BendixCarstensen.com/MethComp/Ancona.2011
If variances are not constant

A transformation might help:

```r
> round( ftable( DA.reg(ox) ), 3 )

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>alpha</th>
<th>beta</th>
<th>sd.pred</th>
<th>beta=1</th>
<th>s.d.=K</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>CO</td>
<td>0.000</td>
<td>1.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>pulse</td>
<td></td>
<td>1.864</td>
<td>0.943</td>
<td>5.979</td>
<td>0.142</td>
<td>0.000</td>
</tr>
<tr>
<td>pulse</td>
<td>CO</td>
<td>-1.977</td>
<td>1.061</td>
<td>6.342</td>
<td>0.142</td>
<td>0.000</td>
</tr>
<tr>
<td>pulse</td>
<td></td>
<td>0.000</td>
<td>1.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

> oxt <- transform( ox, y=log(y/(100-y)) )

> round( ftable( DA.reg(oxt) ), 3 )

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>alpha</th>
<th>beta</th>
<th>sd.pred</th>
<th>beta=1</th>
<th>s.d.=K</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>CO</td>
<td>0.000</td>
<td>1.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>pulse</td>
<td></td>
<td>-0.034</td>
<td>0.900</td>
<td>0.306</td>
<td>0.009</td>
<td>0.246</td>
</tr>
<tr>
<td>pulse</td>
<td>CO</td>
<td>0.038</td>
<td>1.111</td>
<td>0.340</td>
<td>0.009</td>
<td>0.246</td>
</tr>
<tr>
<td>pulse</td>
<td></td>
<td>0.000</td>
<td>1.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
```

Transformation of data
pulse =
2.11 + 0.94 CO 
( 6.00 )

CO =
−2.25 + 1.06 pulse 
( 6.39 )
Analysis on the transformed scale

> ARoxt <- AltReg( ox, linked=T, trace=T, Transform="pctlogit" )

iteration 1 criterion: 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Intercept: CO</th>
<th>pulse Slope: CO</th>
<th>pulse Slope: I</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>0.003</td>
<td>0.998</td>
<td>0.098</td>
<td>1.151</td>
<td>1.151</td>
</tr>
<tr>
<td>pulse</td>
<td>-0.003</td>
<td>1.003</td>
<td>0.098</td>
<td>1.151</td>
<td>1.151</td>
</tr>
</tbody>
</table>

iteration 2 criterion: 0.08547255

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Intercept: CO</th>
<th>pulse Slope: CO</th>
<th>pulse Slope: I</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>-0.024</td>
<td>1.032</td>
<td>0.100</td>
<td>1.151</td>
<td>1.181</td>
</tr>
<tr>
<td>pulse</td>
<td>-0.039</td>
<td>1.019</td>
<td>0.121</td>
<td>1.121</td>
<td>1.151</td>
</tr>
</tbody>
</table>

...
Analysis on the transformed scale

> ARoxt <- AltReg( ox, linked=T, trace=T, Transform="pctlogit" )

AltReg converged after 15 iterations
Last convergence criterion was 0.0008526646

> ARoxt
Note: Response transformed by: log p/(100 - p)

Conversion between methods:

<table>
<thead>
<tr>
<th></th>
<th>alpha</th>
<th>beta</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>To:</td>
<td>From:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>CO</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>pulse</td>
<td>0.042</td>
<td>1.105</td>
</tr>
<tr>
<td>pulse</td>
<td>CO</td>
<td>-0.038</td>
<td>0.905</td>
</tr>
<tr>
<td></td>
<td>pulse</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Variance components (sd):

<table>
<thead>
<tr>
<th></th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>IxR</td>
</tr>
<tr>
<td>CO</td>
<td>0.232</td>
</tr>
<tr>
<td>pulse</td>
<td>0.210</td>
</tr>
</tbody>
</table>

This is an analysis for the transformed data.
Backtransformation for plotting

```r
prpulse <- seq(20,100,1)
lprpulse <- log( prpulse / (100-prpulse) )
lprCO <- ARoxt["CO",2] + ARoxt["CO",4]*lprpulse
lprCOlo <- ARoxt["CO",2] + ARoxt["CO",4]*lprpulse - 2*sd.CO.pred
lprCOhi <- ARoxt["CO",2] + ARoxt["CO",4]*lprpulse + 2*sd.CO.pred
prCO <- 100/(1+exp(-cbind( lprCO, lprCOlo, lprCOhi )))
prCO[nrow(prCO),] <- 100
```

But this is not necessary; it is implemented in `plot.MethComp`:

```r
plot( ARoxt, pl.type="conv" )
```
pulse = 2.03 + 0.94 CO (6.01)

CO = −2.16 + 1.06 pulse (6.38)
Transformation to a Bland-Altman plot

Just convert to the differences versus the averages:

```r
prpulse <- cbind( prpulse, prpulse, prpulse )
with( to.wide(ox),
    plot( (CO+pulse)/2, CO-pulse, pch=16,
         ylim=c(-40,40), xlim=c(20,100),
         xaxs="i", yaxs="i" )
    abline( h=-4:4*10, v=2:10*10, col=gray(0.8) )
matlines( (prCO+prpulse)/2, prCO-prpulse, lwd=c(3,1,1),
          col="blue", lty=1 )

But this is not necessary; it is implemented in plot.MethComp:

plot( ARoxt, pl.type="BA" )
```
pulse =
2.03 + 0.94 CO
( 6.01 )

CO =
−2.16 + 1.06 pulse
( 6.38 )
Implementation in **BUGS**

\[ y_{mir} = \alpha_m + \beta_m (\mu_i + a_{ir} + c_{mi}) + e_{mir} \]

Non-linear hierarchical model: Implement in **BUGS**.

- The model is *symmetrical* in methods.
- Mean is overparametrized.
- Choose a prior (and hence posterior!) for the \( \mu s \) with finite support.
- Keeps the chains nicely in place.

This is the philosophy in the function MCmcmc.
Results from fitting the model

The posterior dist’n of \((\alpha_m, \beta_m, \mu_i)\) is singular.

But the relevant translation quantities are identifiable:

\[
\alpha_{2|1} = \alpha_2 - \alpha_1 \beta_2 / \beta_1 \\
\beta_{2|1} = \beta_2 / \beta_1
\]

So are the variance components.

Posterior medians used to devise prediction equations with limits.
The MethComp package for R

Implemented model:

\[ y_{mir} = \alpha_m + \beta_m (\mu_i + a_{ir} + c_{mi}) + e_{mir} \]

- Replicates required.
- R2WinBUGS, BRugs or JAGS is required.
- Dataframe with variables meth, item, repl and y (a Meth object)
- The function MCmcmc writes a BUGS-program, initial values and data to files.
- Runs BUGS and sucks results back in to R, and gives a nice overview of the conversion equations.
Example output: Oximetry

```r
> summary( ox )
#Replicates
Method  1  2  3 #Items #Obs: 354 Values: min  med   max
  CO   1  4  56   61  177   22.2  78.6  93.5
  pulse 1  4  56  61  177   24.0  75.0  94.0
>
> Mcox <- MCmcmc( ox, linked=TRUE, n.iter=2000 )
Loading required package: coda
Loading required package: lattice
Loading required package: R2WinBUGS
Loading required package: BRugs
Welcome to BRugs running on OpenBUGS version 3.0.3

Comparison of 2 methods, using 354 measurements on 61 items, with up to 3 replicate measurements, (replicate values are in the set: 1 2 3 )
( 2 * 61 * 3 = 366 ):

No. items with measurements on each method:
#Replicates
Method  1  2  3 #Items #Obs: 354 Values: min  med   max
  CO  1  4  56   61  177   22.2  78.6  93.5
```
Simulation run of a model with
- method by item and item by replicate interaction:
- using 4 chains run for 2000 iterations
  (of which 1000 are burn-in),
- monitoring all values of the chain:
- giving a posterior sample of 4000 observations.

model is syntactically correct
data loaded
model compiled
Initializing chain 1: initial values loaded but this or another
Initializing chain 2: initial values loaded but this or another
Initializing chain 3: initial values loaded but this or another
Initializing chain 4: initial values loaded but this or another
initial values generated, model initialized
Sampling has been started ...
1000 updates took 38 s
deviance set
monitor set for variable 'alpha'
monitor set for variable 'beta'
monitor set for variable 'sigma.mi'
monitor set for variable 'sigma.ir'
monitor set for variable 'sigma.res'
monitor set for variable 'deviance'
> MCox

## Conversion between methods:

<table>
<thead>
<tr>
<th>To</th>
<th>From</th>
<th>alpha</th>
<th>beta</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>CO</td>
<td>0.000</td>
<td>1.000</td>
<td>1.740</td>
</tr>
<tr>
<td>pulse</td>
<td>CO</td>
<td>-9.342</td>
<td>1.159</td>
<td>5.328</td>
</tr>
<tr>
<td>pulse</td>
<td>pulse</td>
<td>8.061</td>
<td>0.863</td>
<td>4.508</td>
</tr>
<tr>
<td>pulse</td>
<td>pulse</td>
<td>0.000</td>
<td>1.000</td>
<td>6.115</td>
</tr>
</tbody>
</table>

Implementation in BUGS
Variance components (sd):

<table>
<thead>
<tr>
<th>Method</th>
<th>IxR</th>
<th>MxI</th>
<th>res</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>3.878</td>
<td>3.122</td>
<td>1.230</td>
</tr>
<tr>
<td>pulse</td>
<td>3.222</td>
<td>2.757</td>
<td>4.324</td>
</tr>
</tbody>
</table>

Variance components with 95 % cred.int.:

<table>
<thead>
<tr>
<th>Method</th>
<th>CO</th>
<th>pulse</th>
</tr>
</thead>
<tbody>
<tr>
<td>qnt</td>
<td>50%</td>
<td>2.5%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>2.5%</td>
</tr>
<tr>
<td>SD IxR</td>
<td>3.878</td>
<td>3.053</td>
</tr>
<tr>
<td></td>
<td>3.222</td>
<td>2.426</td>
</tr>
<tr>
<td></td>
<td>3.122</td>
<td>2.193</td>
</tr>
<tr>
<td></td>
<td>2.757</td>
<td>1.915</td>
</tr>
<tr>
<td>res</td>
<td>1.230</td>
<td>0.143</td>
</tr>
<tr>
<td></td>
<td>4.324</td>
<td>3.709</td>
</tr>
<tr>
<td>tot</td>
<td>5.220</td>
<td>4.507</td>
</tr>
<tr>
<td></td>
<td>6.135</td>
<td>5.457</td>
</tr>
</tbody>
</table>

Implementation in BUGS 73/90
Mean parameters with 95 % cred.int.:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>50%</th>
<th>2.5%</th>
<th>97.5%</th>
<th>P(&gt;0/1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha[pulse.C0]</td>
<td>8.057</td>
<td>-2.457</td>
<td>29.884</td>
<td>0.969</td>
</tr>
<tr>
<td>alpha[CO.pulse]</td>
<td>-9.346</td>
<td>-49.949</td>
<td>2.476</td>
<td>0.031</td>
</tr>
<tr>
<td>beta[pulse.C0]</td>
<td>0.863</td>
<td>0.604</td>
<td>0.997</td>
<td>0.024</td>
</tr>
<tr>
<td>beta[CO.pulse]</td>
<td>1.159</td>
<td>1.003</td>
<td>1.657</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Note that intercepts in conversion formulae are adjusted to get conversion formulae that represent the same line both ways, and hence the median intercepts in the posterior do not agree exactly with those given in the conversion formulae.
Program

- Example
- Random rater vs. fixed methods
- Statistical modelling
Example: depression ratings

<table>
<thead>
<tr>
<th>Patient</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

"Doctor doctor! Am I depressed?"

Getting second opinions ... and third ... and fourth
Example: depression ratings

Research question

How well will two doctors agree on the diagnosis?

In this example we use humans as “measurement methods” or raters.

However, we are not interested in making statements about specific raters.
Fixed versus random effects

Definition: Factors can either be fixed or random.

- A factor is fixed when the levels (e.g. raters) under study are the only levels of interest.
- A factor is random when the levels under study are a random sample from a larger population of raters and the goal of the study is to make a statement regarding the larger population.

Raters can be defined as fixed or random factors:

- If the raters themselves are of interest (you want to use them again) then use fixed model.
- If raters are randomly chosen of possible pool of raters (you do not have specific raters in mind) then use the random model.
Fixed versus random effects

Rater: either fixed or random

Fixed raters: The raters themselves are of interest (you want to use the exact same raters again).

Random raters: raters are randomly chosen from a (large) pool of possible raters.

\[ \text{random.raters=} \text{FALSE} \] (default)

\[ \text{random.raters=} \text{TRUE} \]
Modelling: exchangeable replicates

The model for fixed methods is:

\[ y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir} \]

s.d.\((c_{mi}) = \tau_m \quad \text{— "matrix"-effect}\)

s.d.\((e_{mir}) = \sigma_m \quad \text{— measurement error}\)

- Replicates within \((m, i)\) are needed to separate \(\tau\) and \(\sigma\).
- Even with replicates, the separate \(\tau\)s are only estimable if \(M > 2\).
- Assumes that the difference between methods is constant.
- Assumes exchangeability of replicates.

If no replicates then disregard the \(c_{mi}'s\).
Modelling: exchangeable replicates

The model for random methods/raters is:

\[ y_{mir} = b_m + \mu_i + c_{mi} + e_{mir} \]

- \( \text{s.d.}(b_m) = \xi \) — variation among raters
- \( \text{s.d.}(c_{mi}) = \tau_m \) — “matrix”-effect
- \( \text{s.d.}(e_{mir}) = \sigma_m \) — measurement error

- Replicates within \((m, i)\) are needed to separate \(\tau\) and \(\sigma\).
- Even with replicates, the separate \(\tau\)s are only estimable if \(M > 2\).
- Note: average difference is 0!
- Assumes exchangeability of replicates.

If no replicates then disregard the \(c_{mi}\)’s.
Model for replicate measurements

Same approach as before: Fit the correct variance components model and use this as the basis for LoA.

- Extremely flexible.
- Can even be used to analyze the situation where every rater not necessarily has scored every item.

Exchangeable replicates are not uncommon, e.g.,
- Experts scoring/extracting information from images
- Measurements taken on couples/twins.

Linked replicates do not make sense, when it is arbitrary which person is partner 1 or partner 2.
Replicate measurements

The limits of agreement / prediction interval for two random raters scoring a new future observation is

\[ 0 \pm 1.96 \sqrt{\frac{2\xi^2 + \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2}{\text{Extra variation}}} \]

However, since we are considering the prediction interval for two random raters we use the average variance components in the formula

\[ 0 \pm 1.96 \sqrt{2(\xi^2 + \bar{\tau}^2 + \bar{\sigma}^2)} \]

Note that the expected difference is zero since we have no fixed order of the raters.
Example: Stress scoring of dogs

10 judges scoring stress indicators from 10 dogs.

> dogdata <- Meth(item=1, y=2:11, data=dogs)
> BA.est(dogdata, random.raters=TRUE, linked=FALSE)

Variance components (sd):

<table>
<thead>
<tr>
<th></th>
<th>IxR</th>
<th>MxI</th>
<th>M</th>
<th>res</th>
</tr>
</thead>
<tbody>
<tr>
<td>j1</td>
<td>0</td>
<td>18.145</td>
<td>14.11</td>
<td>20.948</td>
</tr>
<tr>
<td>j10</td>
<td>0</td>
<td>8.122</td>
<td>14.11</td>
<td>12.736</td>
</tr>
<tr>
<td>j2</td>
<td>0</td>
<td>0.009</td>
<td>14.11</td>
<td>11.350</td>
</tr>
<tr>
<td>j3</td>
<td>0</td>
<td>0.004</td>
<td>14.11</td>
<td>9.524</td>
</tr>
<tr>
<td>j4</td>
<td>0</td>
<td>0.004</td>
<td>14.11</td>
<td>9.614</td>
</tr>
<tr>
<td>j5</td>
<td>0</td>
<td>8.924</td>
<td>14.11</td>
<td>12.588</td>
</tr>
<tr>
<td>j6</td>
<td>0</td>
<td>18.534</td>
<td>14.11</td>
<td>21.135</td>
</tr>
<tr>
<td>j7</td>
<td>0</td>
<td>0.023</td>
<td>14.11</td>
<td>11.991</td>
</tr>
<tr>
<td>j8</td>
<td>0</td>
<td>0.004</td>
<td>14.11</td>
<td>9.384</td>
</tr>
<tr>
<td>j9</td>
<td>0</td>
<td>0.003</td>
<td>14.11</td>
<td>9.789</td>
</tr>
</tbody>
</table>

Inter-rater agreement 84/90
Example: Stress scoring of dogs

10 judges scoring stress indicators from 10 dogs.

```r
> res <- BA.est(dogdata, random.raters=TRUE,
+            linked=FALSE)
> res$LoA

Mean       Lower       Upper       SD
Rand. rater - rater  0 -61.02451  61.02451  30.51225
```
Linked replicates

For linked replicates, extend the model as before:

\[ y_{mir} = b_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \]

\( \text{s.d.}(b_m) = \xi \) — variation among raters
\( \text{s.d.}(a_{ir}) = \omega \) — between replicates
\( \text{s.d.}(c_{mi}) = \tau_m \) — "matrix"-effect
\( \text{s.d.}(e_{mir}) = \sigma_m \) — measurement error

The variation between replicates, \( \omega \), does not enter the limits-of-agreement since the LoA’s are for a single new future observation (ie., the same replicate from one item/individual for both raters).

\[
0 \pm 1.96 \sqrt{2(\xi^2 + \tau^2 + \sigma^2)}
\]
Linked replicates

```r
> dogdata <- Meth(item=1, y=2:11, data=dogs)
> BA.est(dogdata, random.raters=TRUE)

Variance components (sd):

<table>
<thead>
<tr>
<th></th>
<th>IxR</th>
<th>MxI</th>
<th>M</th>
<th>res</th>
</tr>
</thead>
<tbody>
<tr>
<td>j1</td>
<td>6.466</td>
<td>18.754</td>
<td>13.994</td>
<td>21.672</td>
</tr>
<tr>
<td>j10</td>
<td>6.466</td>
<td>8.163</td>
<td>13.994</td>
<td>10.454</td>
</tr>
<tr>
<td>j2</td>
<td>6.466</td>
<td>3.213</td>
<td>13.994</td>
<td>10.439</td>
</tr>
<tr>
<td>j3</td>
<td>6.466</td>
<td>0.011</td>
<td>13.994</td>
<td>5.718</td>
</tr>
<tr>
<td>j4</td>
<td>6.466</td>
<td>0.033</td>
<td>13.994</td>
<td>11.716</td>
</tr>
<tr>
<td>j5</td>
<td>6.466</td>
<td>8.817</td>
<td>13.994</td>
<td>10.449</td>
</tr>
<tr>
<td>j7</td>
<td>6.466</td>
<td>4.732</td>
<td>13.994</td>
<td>10.523</td>
</tr>
<tr>
<td>j8</td>
<td>6.466</td>
<td>0.009</td>
<td>13.994</td>
<td>5.701</td>
</tr>
<tr>
<td>j9</td>
<td>6.466</td>
<td>0.026</td>
<td>13.994</td>
<td>11.441</td>
</tr>
</tbody>
</table>
```

Inter-rater agreement

87/90
## Linked replicates

```r
> res2 <- BA.est(dogdata, random.raters=TRUE)
> res2$LoA

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Lower</th>
<th>Upper</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rand. rater - rater</td>
<td>0</td>
<td>-60.47317</td>
<td>60.47317</td>
<td>30.23658</td>
</tr>
</tbody>
</table>
```

**Inter-rater agreement**

88/ 90
Random raters

- Fit the correct variance component model where variation among raters is considered a random effect
- Since each rater can have his/her individual variance we need to average the individual variance components
- Extract the relevant variance components and compute the limits-of-agreement
DG Altman and JM Bland.
Measurement in medicine: The analysis of method comparison studies.

JM Bland and DG Altman.
Statistical methods for assessing agreement between two methods of clinical measurement.

B Carstensen, J Simpson, and LC Gurrin.
Statistical models for assessing agreement in method comparison studies with replicate measurements.

B Carstensen.
Comparing and predicting between several methods of measurement.

B. Carstensen.
*Comparing Clinical Measurement Methods: A practical guide.*