Statistical Analysis of Method Comparison studies

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
http://staff.pubhealth.ku.dk/~bxc/

Dept. Biostatistics, Univ. of Copenhagen
8–10 February 2011

www.biostat.ku.dk/~bxc/MethComp

Comparing two methods with one measurement on each

Tuesday 8 February, morning

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
www.biostat.ku.dk/~bxc/MethComp

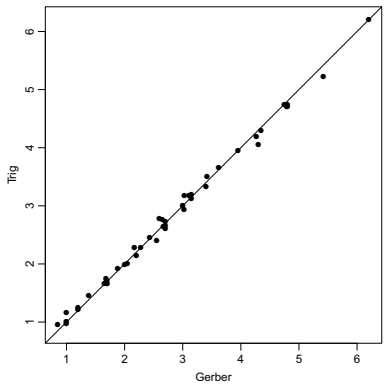
(Comp-simple)

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?

$$D_i = y_{2i} - y_{1i}, \qquad \bar{D}, \qquad \text{s.d.}(D)$$

"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, **?**]

Limits of agreement: Interpretation

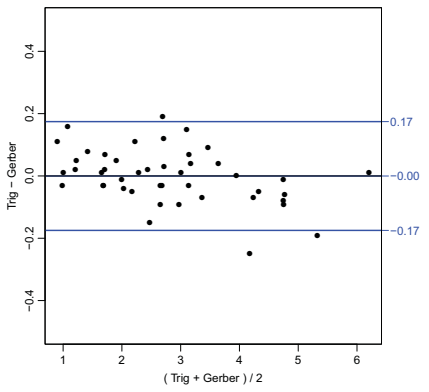
- If a new patient is measured **once** with each of the two methods, the difference between the two values will with 95% probability be within the limits of agreement.
- This is a **prediction** interval for a (future) difference.
- Requires a **clinical** input:
  Are the limits of agreement sufficiently 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.
- It is an **equivalence** problem:
  1: Testing is irrelevant.
  2: Clinical input is required.

Limits of agreement:



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:

$$\begin{aligned} y_{mi} &= \alpha_m + \mu_i + e_{mi} \\ e_{mi} &\sim \mathcal{N}(0, \sigma_m^2) \quad \text{measurement error} \end{aligned}$$

- 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:

$$\widehat{y_2 - y_1} = \hat{D} = \alpha_2 - \alpha_1 \pm 1.96\,\text{s.d.}(D)$$

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 σ/I.

So we should use $t_{0.95} \times \sqrt{(I+1)/I}$ instead. This id 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:

$$\hat{y}_{2|1} = \hat{y}_2|y_1 = \alpha_2 - \alpha_1 + y_1 \pm 2\,\text{s.d.}(D)$$

Correlation
Tuesday 8 February, morning

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
www.biostat.ku.dk/~bxc/MethComp

(Correlation)

Spurious correlation?

Unequal variances induce correlation between D_i and A_i; if variances of y_{1i} and y_{2i} are ζ_1^2 and ζ_2^2 respctively:

$$\text{cov}(D_i, A_i) = \frac{1}{2}(\zeta_2^2 - \zeta_1^2) \neq 0 \quad \text{if } \zeta_1 \neq \zeta_2$$

In correlation terms:

$$\rho(D, A) = \frac{1}{2}\left(\frac{\zeta_2^2 - \zeta_1^2}{\zeta_1^2 + \zeta_2^2}\right)$$

i.e. the correlation depends on whether the difference between the variances is large relative to the sizes of the two.

— not really

The variances we were using were the *marginal* variances of y_1 and y_2:

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

so we have that the marginal variances are:

$$\text{var}(y_m) = \text{var}(\mu_i) + \sigma_m^2$$

and hence the correlation expression is:

$$\rho(D, A) = \frac{1}{2}\left(\frac{\zeta_2^2 - \zeta_1^2}{\zeta_1^2 + \zeta_2^2}\right) = \frac{1}{2}\left(\frac{\sigma_2^2 - \sigma_1^2}{2\text{var}(\mu_i) + \sigma_1^2 + \sigma_2^2}\right)$$

Hence only relevant if $\text{var}(\mu_i)$ is small relative to σ_1^2 and σ_2^2. **Not** likely in practise.

Introduction to computing
Tuesday 8 February, morning

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
www.biostat.ku.dk/~bxc/MethComp

(Intro-comp)

Course structure

The course 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.

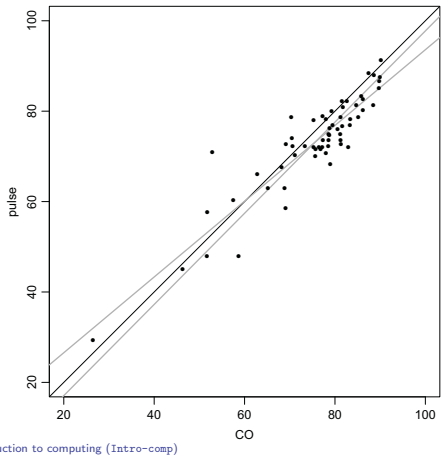
- **R** for data manipulation and graphics.
- Occasionally BUGS for estimation in non-linear variance component models.

Software considerations

- **R**, SAS and Stata all have interfaces to WinBUGS.
- But **R** have more flexible graphical facilities.
- The `MethComp` package is written for **R**.

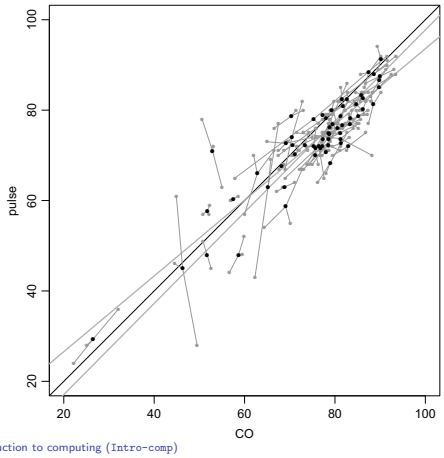
Therefore we use **R** in this course.

Oximetry data



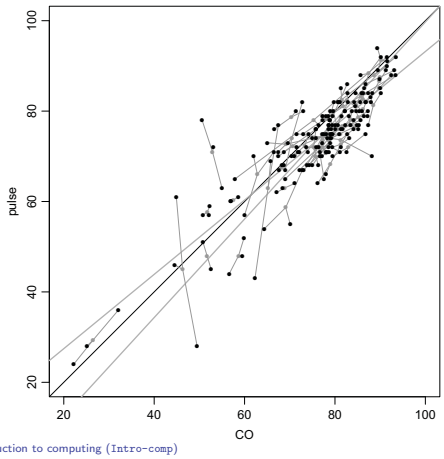
Means over replicates.

Oximetry data



Linked replicates.

Oximetry data



Linked replicates.

How it works

Example data sets are included in the `MethComp` package.

The function 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.

How it looks:

```
> subset(ox,as.integer(item)<3)
     meth item repl    y
1      CO    1    1 78.0
2      CO    1    2 76.4
3      CO    1    3 77.2
4      CO    2    1 68.7
5      CO    2    2 67.6
6      CO    2    3 68.3
184 pulse    1    1 71.0
185 pulse    1    2 72.0
186 pulse    1    3 73.0
187 pulse    2    1 68.0
188 pulse    2    2 67.0
189 pulse    2    3 68.0
```

```
> subset(to.wide(ox),as.integer
Note:
 Replicate measurements are t
  item repl  id   CO pulse
1    1    1 1.1 78.0    71
2    1    2 1.2 76.4    72
3    1    3 1.3 77.2    73
4    2    1 2.1 68.7    68
5    2    2 2.2 67.6    67
6    2    3 2.3 68.3    68
```


Analysis options in this course

- Scatter plots.
- Bland-Altman plots ($(y_2 - y_1)$ vs. $(y_1 + y_2)/2$)
- Limits of Agreement (LoA).
- Models with constant bias.
- Models with linear bias.
- Conversion formulae between methods (single replicates)
- Tansformation of measurements.
- Plots of converison equations.
- Reporting of variance components.

Requirements

- **R** for data manipulation and graphics:
- `Tinn-R` text editor with syntax highlighting for **R**. Alternatively you can use the bulit-in editor in **R**, or the nerds can use `ESS`.
- `nlme`-package for variance component models — constant bias.
- `BUGS` for fitting models with linear bias (non-linear variance component models, over-parametrized).

All of it works from within **R**.

About R

- **R** uses *objects* — this can ee a data-frame, a single number, a table or a vector (set of numbers)
- and *functions* that take one or more objects and produces:
  - printed output
  - graph
  - another object

```
oxim <- Meth( ox )
plot( oxim )
```


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.

Graphical functions (basic)

- BA.plot Makes a Bland-Altman plot of two methods from a data frame with method comparison data, and computes limits of agreement. The plotting etc is really done by a call to
- BlandAltman Draws a Bland-Altman plot and computes limits of agreement.
- plot.Meth Plots all methods against all other, both as a scatter plot and as a Bland-Altman plot.
- bothlines Adds regression lines of y on x and vice versa to a scatter plot.

Data manipulation functions

- make.repl Generates a repl column in a data frame with columns meth, item and y.
- perm.repl Randomly permutes replicates within (method,item) and assigns new replicate numbers.
- to.wide/to.long Transforms a data frame in the long form to the wide form and vice versa.
- Meth.sim Simulates a dataset (a Meth object) from a method comparison experiment.

Analysis functions (simple)

- DA.reg, regresses the differences on the averages. Also regresses the absolute residuals on the averages to check whether the variance is constant.
- Deming Performs Deming regression, i.e. regression with errors in both variables.
- BA.est Estimates in the variance components models underlying the concept of limits of agreement, and returns the bias and the variance components. Assumes constant bias between methods.
- VC.est The workhorse behind BA.est.

Analysis functions (general)

- AltReg Estimates via ad-hoc procedure (alternating regressions) in a model with linear bias between methods. Returns a matrix of estimates with the conversion parameters as well as the variance components.
- MCmcmc Estimates via BUGS in the general model with non-constant bias. Produces an MCmcmc object.

Reporting functions

- summary.Meth Tabulates replicates by methods and items.
- print.MCmcmc Prints a table of conversion equation between methods analyzed, with prediction standard deviations. Also gives summaries of the posteriors for the parameters that constitute the conversion algorithms.
- plot.MCmcmc Plots the conversion lines between methods with prediction limits.
- post.MCmcmc Plots smoothed posterior densities for the estimates.
- trace.MCmcmc Plots the simulation traces from an MCmcmc object.

Does it work?

You should get something reasonable out of this:

```
library(MethComp)
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)
trace.MCmcmc(MC.ox)
post.MCmcmc(MC.ox)
```

Non-constant difference
Tuesday 8 February, afternoon

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
www.biostat.ku.dk/~bxc/MethComp

(Non-const)

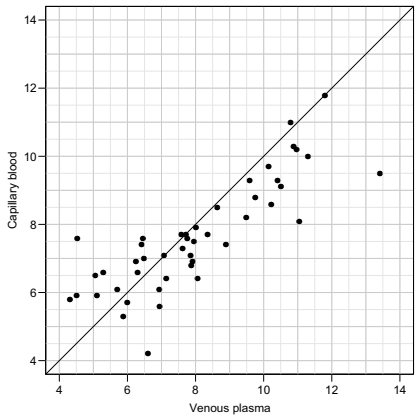
Limits of agreement — assumptions

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

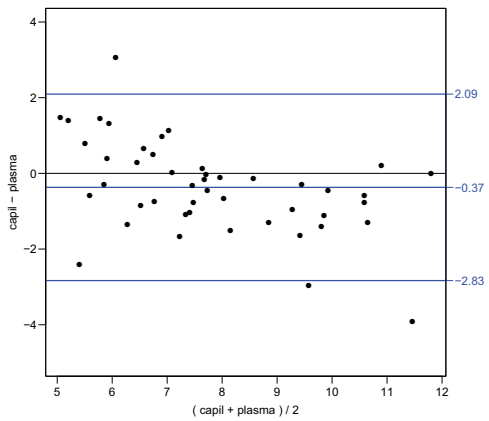
Check this by:

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

Glucose measurements



Glucose measurements



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:

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

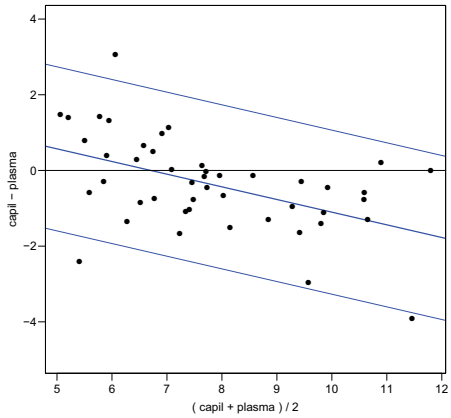
Exchanging methods would give:

$$y_{1|2} = -a + (1-b)y_1$$

instead of:

$$y_{1|2} = \frac{-a}{1+b} + \frac{1}{1+b}y_1$$

Variable limits of agreement



Improving the regression of D on A

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

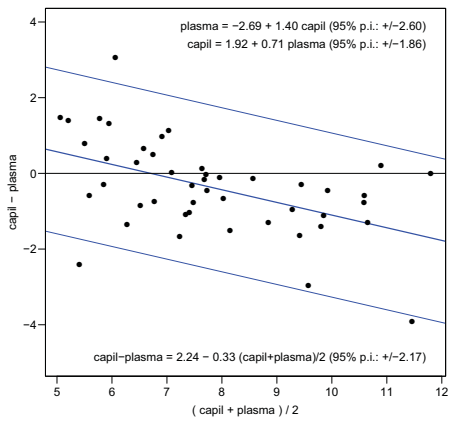
$$y_{2i}(1 - b/2) = a + (1 + b/2)y_{1i} + e_i$$

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

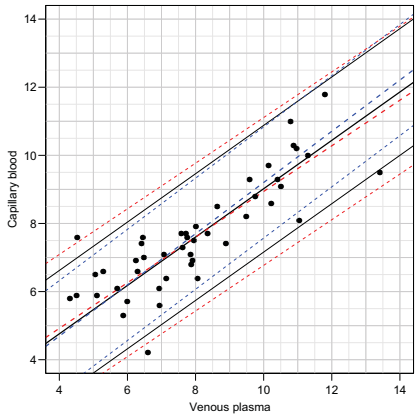
$$y_{1i} = \frac{-a}{1+b/2} + \frac{1-b/2}{1+b/2}y_{2i} + \frac{1}{1+b/2}e_i$$

This is what comes out of the functions `DA.reg` and `BA.plot`

Variable limits of agreement



Conversion equation with prediction limits



Why does this work?

The general model for the data is:

$$\begin{aligned} y_{1i} &= \alpha_1 + \beta_1\mu_i + e_{1i}, & e_{1i} &\sim \mathcal{N}(0, \sigma_1^2) \\ y_{2i} &= \alpha_2 + \beta_2\mu_i + e_{2i}, & e_{2i} &\sim \mathcal{N}(0, \sigma_2^2) \end{aligned}$$

- 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?

Conceptually:
Once the β_m is introduced:

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

measurements by different methods are on different scales.

The scalings, β_m, of the "true" μs are different for the two methods.

Hence it has formally no meaning to form the differences.

So why is it wrong anyway?

Statistically:
Under the specified model for the ya, the induced model for the differences on the averages A_i, these contain the error terms, and so does the residuals:

$$D_i = a + bA_i + e_i,$$

where:

$$\begin{aligned} D_i &= (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)\mu_i + e_{1i} - e_{2i} \\ A_i &= (\alpha_1 + \alpha_2)/2 + (\beta_1 + \beta_2)\mu_i/2 + (e_{1i} + e_{2i})/2 \\ e_i &= e_{1i}\left(1 - \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}\right) - e_{2i}\left(1 + \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}\right) \end{aligned}$$

So why is it wrong anyway?

Statistically:
So the covariate is not independent of the error terms:

$$\mathrm{cov}(A_i, e_i) = \frac{1}{2}\left\{\sigma_1^2 - \sigma_2^2 - \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}(\sigma_1^2 + \sigma_2^2)\right\}$$

Thus the assumptions behind regression are violated.

Then why use it?

- With only one observation per (method,item) there is not much else to do.
- If the slope linking the two methods (β_1/β_2) is not dramatically different from 1, the violatiosn are not that big.
- The transformatiion $(y_{1i}, y_{2i}) \mapsto (D_i, A_i)$ is a transformation to two quantities *approximately* (marginally) independent, and therefore better suited for regreassion.

Implemented in `BA.plot` and in `DA.reg`, which also checks the residuals.

For further details, see [2].

Comparing two methods with replicate measurements

Tuesday 8 February, afternoon

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
`www.biostat.ku.dk/~bxc/MethComp`

`(comp-repl)`

Replicate measurements

Fat data; exchangeable replicates:

```
item repl  KL  SL
   1    1 4.5 4.9
   1    2 4.4 5.0
   1    3 4.7 4.8
   3    1 6.4 6.5
   3    2 6.2 6.4
   3    3 6.5 6.1
```

Oximetry data; linked replicates:

```
item repl   CO pulse
   1    1 78.0    71
   1    2 76.4    72
   1    3 77.2    73
   2    1 68.7    68
   2    2 67.6    67
   2    3 68.3    68
```

Linked or exchangeable replicates!

Extension of the model: exchangeable replicates

$$\begin{aligned} y_{mir} &= \alpha_m + \mu_i + c_{mi} + e_{mir} \\ \text{s.d.}(c_{mi}) &= \tau_m \quad \text{— "matrix"-effect} \\ \text{s.d.}(e_{mir}) &= \sigma_m \quad \text{— measurement error} \end{aligned}$$

- Replicates within (m, i) are needed to separate τ and σ.
- Even with replicates, the separate τs 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

$$\begin{aligned} y_{mir} &= \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \\ \text{s.d.}(a_{ir}) &= \omega \quad \text{— between replicates} \\ \text{s.d.}(c_{mi}) &= \tau_m \quad \text{— "matrix"-effect} \\ \text{s.d.}(e_{mir}) &= \sigma_m \quad \text{— measurement error} \end{aligned}$$

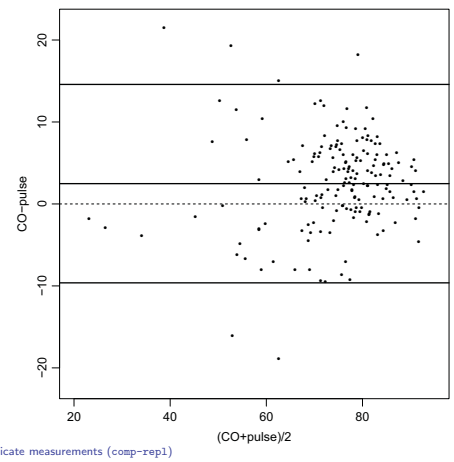
- Still assumes that the difference between methods is constant.
- Replicates are *linked* between methods: a_{ir} is common across methods, i.e. the first replicate on a person is made under similar conditions for all methods (i.e. at a specific day or the like).

Replicate measurements

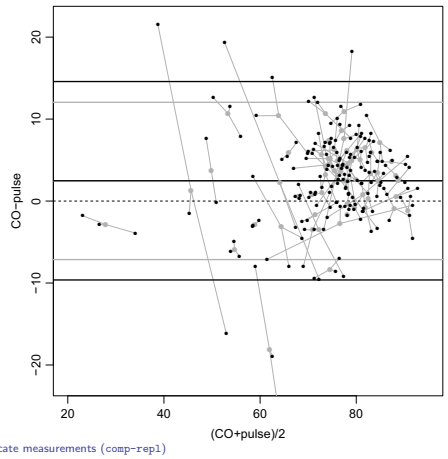
Three approaches to limits of agreement with replicate measurements:

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 )`.

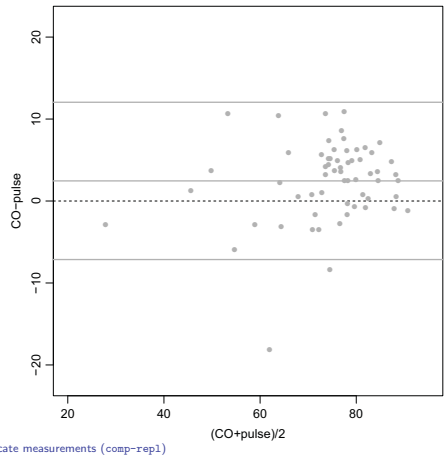
Oximetry data



Oximetry data



Oximetry data



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:

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

- $\text{var}(y_{1jr} - y_{2jr}) = \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2$
  — 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:

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

If we are using means of replicates to form the differences we have:

$$\begin{aligned} \bar{d}_i = \bar{y}_{1i\cdot} - \bar{y}_{2i\cdot} &= \alpha_1 - \alpha_2 + \frac{\sum_r a_{ir}}{R_{1i}} - \frac{\sum_r a_{ir}}{R_{2i}} \\ &\quad + c_{1i} - c_{2i} + \frac{\sum_r e_{1ir}}{R_{1i}} - \frac{\sum_r e_{2ir}}{R_{2i}} \end{aligned}$$

The terms with a_{ir} are only relevant for linked replicates in which case $R_{1i} = R_{2i}$ and therefore the term vanishes. Thus:

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

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.

Exchangeable replicates as items?

If replicates are exchangeable it is not clear how to produce the differences using replicates as items.

If replicates are paired at random (se the function `perm.repl`), the variance will still be correct using the model without the $i \times r$ interaction term (a_{ir}):

$$\text{var}(y_{1ir} - y_{2is}) = \tau_1^2 + \sigma_1^2 + \tau_2^2 + \sigma_2^2$$

Differences will be positively correlated within item:

$$\text{cov}(y_{1ir} - y_{2is}, y_{1it} - y_{2iu}) = \tau_1^2 + \tau_2^2$$

— slight underestimate of the true variance.

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.

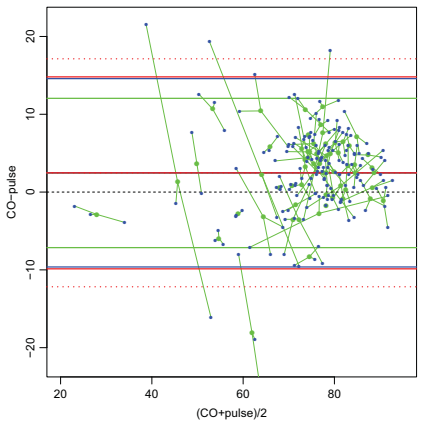
Further details, see [3].

Oximetry data

Linked replicates used as items

Mean over replicates as items

Limits based on model — dashed line assuming exchangeable replicates



A general model

Wednesday 9 February, morning

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
`www.biostat.ku.dk/~bxc/MethComp`

`(General)`

Extension of the model:

$$\begin{aligned} y_{mir} &= \alpha_m + \mu_i + a_{ir} + c_{mi} + d_{mr} + e_{mir} \\ &\text{s.d.}(a_{ir}) = \omega \quad \text{— between replicates} \\ &\text{s.d.}(c_{mi}) = \tau_m \quad \text{— "matrix"-effect} \\ &\text{s.d.}(d_{mr}) = \nu_m \quad \text{— replicate structre} \\ &\text{s.d.}(e_{mir}) = \sigma_m \quad \text{— measurement error} \end{aligned}$$

Method, Item, Replicate

- 1 3-way interaction
- 3 2-way interactions

What part of the interactions should be systematic (fixed) and what part should be random?

(m, r) - between replicates within method

This effect has $M \times R$ levels, usually a rather small number.

This effect will therefore normally be modelled as a fixed effect, but not necessarily with $M \times R$ parameters, presumably fewer.

If replicates are times of sampling or analysis, we may consider different time trends for each method, e.g.

$$d_{mr} = \gamma_m t_r$$

A random $m \times r$-effect would be hard to interpret. Omitted in the following.

(i, r) - between replicates within individual

Observations with same (i, r) — but different by different methods — will be correlated.

Use if all methods are applied to each item at

- different times
- at different locations
- at different conditions

This means there is a some common structure to replicates with the same number — they are **linked**.

(m, i) - between methods within individual

This is what is often called a "matrix" effect.

Matrix in the chemical sense: The surrounding matter ("matrix") in which the substance of interest is dissolved.

Represents random effects of items reacting differently on each measurement method.

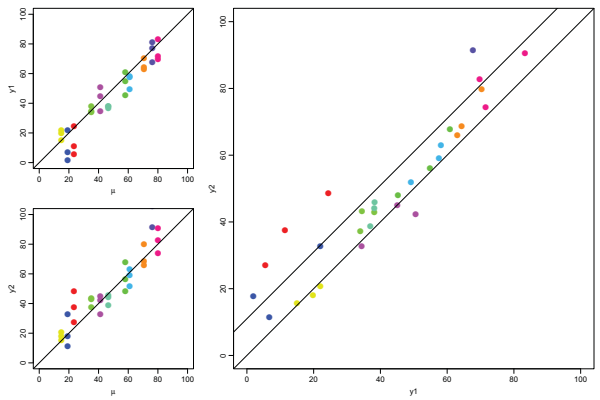
Logical to require that the variance of these effects is allowed to differ between methods.

Variance component model!

$$
\begin{aligned}
y_{mir} &= \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \\
&\text{s.d.}(a_{ir}) = \omega \quad \text{— between replicates} \\
&\text{s.d.}(c_{mi}) = \tau_m \quad \text{— "matrix"-effect} \\
&\text{s.d.}(e_{mir}) = \sigma_m \quad \text{— measurement error}
\end{aligned}
$$

Note we do not consider the method by replicate interaction any more.

The model is a (standard) variance component model, where two of the variance components depend on method.



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

Fitting the variance component model

Complicated and counter-intuitive in **R**:

```
> library( nlme )
> lme( y ~ meth + item,
       random = list( item = pdIdent(~meth - 1),
                      repl = ~1),
       weights = varIdent(form = ~1 | meth),
       data = ox)
```

Teasing out the estimates of the variance components is quite an ordeal, hence it is packaged in the `BA.est` function.


```
Random effects:
 Formula: ~meth - 1 | item
 Structure: Multiple of an Identity
          methCO methpulse
StdDev: 2.928042  2.928042

 Formula: ~1 | repl %in% item
        (Intercept) Residual
StdDev:    3.415692 2.224868

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | meth
 Parameter estimates:
      CO    pulse
1.000000 1.795365
Number of Observations: 354
Number of Groups:
          item repl %in% item
            61             177
```


Packed solution

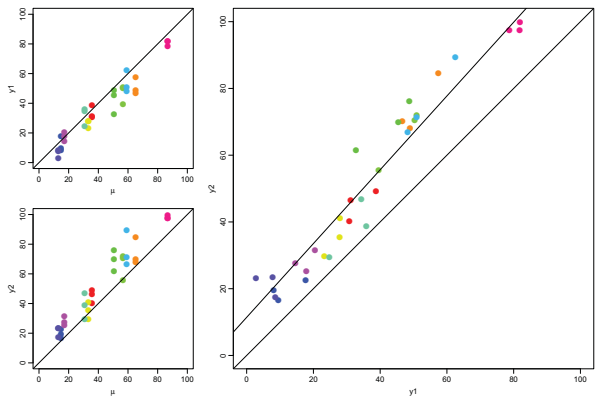
This model has been packaged in a function that calls `lme` and then tease out the relevant parameters.

```
> BA.est(ox,linked=TRUE)
$Bias
       CO     pulse
 0.000000 -2.470446

$VarComp
             IxR      MxI      res
CO    3.415692 2.928042 2.224868
pulse 3.415692 2.928042 3.994451

$LoA
                 Mean     Lower    Upper       SD
pulse - CO   -2.470446 -14.80779 9.866901 6.168674

$RepCoef
             SD    Coef.
CO    5.764892 11.52978
pulse 7.432710 14.86542
```

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

Repeatability and reproducibility

Wednesday 9 February, morning

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
www.biostat.ku.dk/~bxc/MethComp

(Repro)

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.
  (**Repeata**bility 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.
  (**Reproduci**bility conditions)

Quantification of accuracy

- Upper limit of a 95% confidence interval for the difference between two measurments.
- Suppose the variance of the measurement is σ^2:

$$\mathrm{var}(y_{mi1} - y_{mi2}) = 2\sigma^2$$

  i.e the standard error is $\sqrt{2}\sigma$, and a confidnece 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 σ?
- 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.

Linear bias between methods

Wednesday 9 February, afternoon

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
www.biostat.ku.dk/~bxc/MethComp

(Lin-bias)

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 β_2/β_1.

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

The corresponding β 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 β_m.

Implies that $\omega = \mathrm{s.d.}(a_{ir})$ is irrelevant — the scale is arbitrary. The relevant quantities are $\beta_m\omega$ — the between replicate variation within item *as measured on the mth scale*.

Variance components

Method, Item, Replicate.

$$\begin{aligned} y_{mir} &= \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir} \\ &\mathrm{s.d.}(c_{mi}) = \tau_m \end{aligned}$$

Matrix-effect: Each item reacts differently to each method.

If only two methods compared:
τ_1 and τ_2 cannot be separated. Variances must be reported on the scale of each method, as $\beta_m\tau_m$.

Variance components

Method, Item, Replicate.

$$\begin{aligned} y_{mir} &= \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir} \\ &\text{s.d.}(a_{ir}) = \omega \end{aligned}$$

Common across methods — must be scaled relative to the methods.

Included if replicates are linked across methods, e.g. if there is a sequence in the replicates.

The relevant quantities to reports are $\beta_m\omega$ — the s.d. on the scale of the mth method.

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 μ the model is a linear mixed model.
- For fixed (α, β) 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 (α, β) 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 $\zeta_{mir} = \bar{y}_{mi\cdot}$
2. Estimate (α_m, β_m).
3. Compute the scaled responses and fit the random effects model.
4. Use the estimated μ_is, and BLUPs of a_{ir} and c_{mi} to update ζ_{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 (α_m, β_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
       alpha  beta sigma Intercept: CO  pulse Slope: CO pulse Ix
CO     0.911 0.988 1.861         74.419 74.417      1.000 0.974
pulse -1.039 1.014 1.860         74.422 74.419      1.027 1.000
...

iteration 14 criterion: 0.000986339
        alpha  beta sigma Intercept: CO  pulse Slope: CO pulse I
CO    -20.548 1.281 1.027         74.419 76.938      1.000 1.063
pulse -17.301 1.205 3.308         72.049 74.419      0.941 1.000
There were 14 warnings (use warnings() to see them)

> round(AR.ox,3)
    From
To   Intercept: CO  pulse Slope: CO pulse IxR sd. MxI sd. res.sd.
  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
```


Converting between methods

Wednesday 9 February, afternoon

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
`www.biostat.ku.dk/~bxc/MethComp`

(Convert)

Predicting method 2 from method 1

$$\begin{aligned} y_{10r} &= \alpha_1 + \beta_1(\mu_0 + a_{0r} + c_{10}) + e_{10r} \\ y_{20r} &= \alpha_2 + \beta_2(\mu_0 + a_{0r} + c_{20}) + e_{20r} \\ &\Downarrow \\ y_{20r} &= \alpha_2 + \frac{\beta_2}{\beta_1}(y_{10r} - \alpha_1 - e_{10r}) \\ &\qquad + \beta_2(-c_{10} + c_{20}) + e_{20r} \end{aligned}$$

The random effects have expectation 0, so:

$$\text{E}(y_{20}|y_{10}) = \hat{y}_{20} = \alpha_2 + \frac{\beta_2}{\beta_1}(y_{10} - \alpha_1)$$

$$
\begin{aligned}
y_{20r} &= \alpha_2 + \frac{\beta_2}{\beta_1}(y_{10r} - \alpha_1 - e_{10r}) \\
&\quad + \beta_2(-c_{10} + c_{20}) + e_{20r} \\
\mathrm{var}(\hat{y}_{20}|y_{10}) &= \left(\frac{\beta_2}{\beta_1}\right)^2(\beta_1^2\tau_1^2 + \sigma_1^2) + (\beta_2^2\tau_2^2 + \sigma_2^2)
\end{aligned}
$$

The slope of the prediction line from method 1 to method 2 is β_2/β_1.

The width of the prediction interval is:

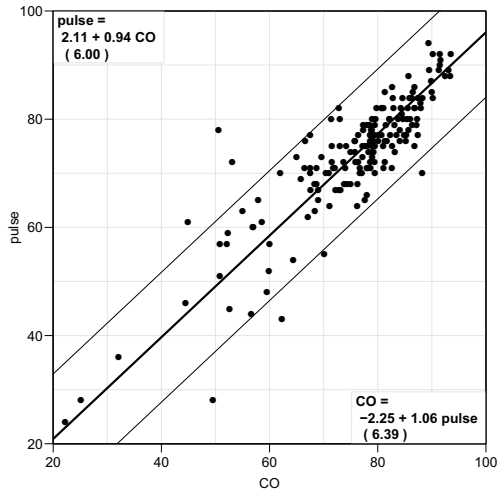
$$
2 \times 2 \times \sqrt{\left(\frac{\beta_2}{\beta_1}\right)^2(\beta_1^2\tau_1^2 + \sigma_1^2) + (\beta_2^2\tau_2^2 + \sigma_2^2)}
$$

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, β_1/β_2.

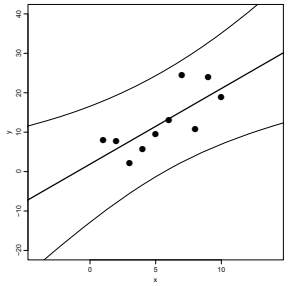
The width of the prediction interval in this direction is (by permutation of indices):

$$
\begin{aligned}
&2 \times 2 \times \sqrt{(\beta_1^2\tau_1^2 + \sigma_1^2) + \left(\frac{\beta_1}{\beta_2}\right)^2(\beta_2^2\tau_2^2 + \sigma_2^2)} \\
&= 2 \times 2 \times \frac{\beta_1}{\beta_2}\sqrt{\left(\frac{\beta_2}{\beta_1}\right)^2(\beta_1^2\tau_1^2 + \sigma_1^2) + (\beta_2^2\tau_2^2 + \sigma_2^2)}
\end{aligned}
$$

i.e. if we draw the prediction limits as straight lines they can be used both ways.



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 $\beta_{2|1}$.

Variance components
Wednesday 9 February, afternoon

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
`www.biostat.ku.dk/~bxc/MethComp`

(`Var-comp`)

Variance components

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

3 variance components / random effects:

- a_{ir}: between replicates within item, ω^2
  $\beta_m\omega$ is the relevant quantity.
- c_{mi}: matrix effect τ_m^2
  $\beta_m\tau_m$ is the relevant quantity.
- e_{mir}: measurement error, residual variation σ_m^2
  σ_m is the relevant quantity.

Variance components

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

The total variance of a measurement is:

$$
\sqrt{\beta_m^2\omega^2 + \beta_m^2\tau_m^2 + \sigma_m^2}
$$

These are the variance components returned by `AltReg` or `MCmcmcm` using `print.MCmcmc` and shown by `post.MCmcmc`.

Repeatability and reproducibility

Repeatability is based on the difference between measurements made under comparable, though not exactly identical conditions.

Reproducibility is based on the difference between measurements made under comparable, though not exactly identical conditions.

This is a different setting from the one underlying the modelling of data from a comparison experiment.

The exchangeability has no meaning, we are discussing future measurements in different circumstances.

Repeatability and reproducibility

Repeatability: $2.8\sigma_m$:
same individual, same replicate, but not considering the variation that constitute differences between replicates *in the experiment*.

Hence *reproducibility* is not estimable from a classical experiment, unless an extra layer of replication is introduced — i.e. different laboratories.

Transformation of data
Wednesday 9 February, afternoon

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
www.biostat.ku.dk/~bxc/MethComp

(Transform)

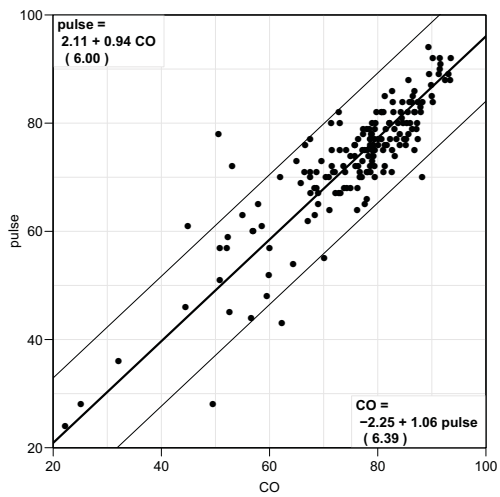
If variances are not constant

A transformation might help:

```
> round( ftable( DA.reg(ox) ), 3 )
              alpha   beta sd.pred beta=1 s.d.=K
From: To:
CO    CO      0.000  1.000      NA     NA     NA
      pulse   1.864  0.943   5.979  0.142  0.000
pulse CO     -1.977  1.061   6.342  0.142  0.000
      pulse   0.000  1.000      NA     NA     NA

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

> round( ftable( DA.reg(oxt) ), 3 )
              alpha   beta sd.pred beta=1 s.d.=K
From: To:
CO    CO      0.000  1.000      NA     NA     NA
      pulse  -0.034  0.900   0.306  0.009  0.246
pulse CO      0.038  1.111   0.340  0.009  0.246
      pulse   0.000  1.000      NA     NA     NA
```

Analysis on the transformed scale

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

iteration 1 criterion: 1
       alpha  beta sigma Intercept: CO pulse Slope: CO pulse   I
CO     0.003 0.998 0.098         1.151 1.151     1.000 0.994 0.2
pulse -0.003 1.003 0.098         1.151 1.151     1.006 1.000 0.2

iteration 2 criterion: 0.08547255
       alpha  beta sigma Intercept: CO pulse Slope: CO pulse   I
CO    -0.024 1.032 0.100         1.151 1.181     1.000 1.013 0.2
pulse -0.039 1.019 0.121         1.121 1.151     0.987 1.000 0.2

...

iteration 15 criterion: 0.0008526646
       alpha  beta sigma Intercept: CO pulse Slope: CO pulse   I
CO    -0.528 1.506 0.082         1.151 1.314     1.000 1.105 0.2
pulse -0.516 1.362 0.144         1.003 1.151     0.905 1.000 0.2
```


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:
                alpha   beta     sd
To:   From:
CO    CO        0.000  1.000  0.202
      pulse     0.042  1.105  0.341
pulse CO       -0.038  0.905  0.309
      pulse     0.000  1.000  0.271

 Variance components (sd):
         s.d.
Method     IxR   MxI   res
  CO     0.232 0.160 0.143
  pulse  0.210 0.145 0.191
```

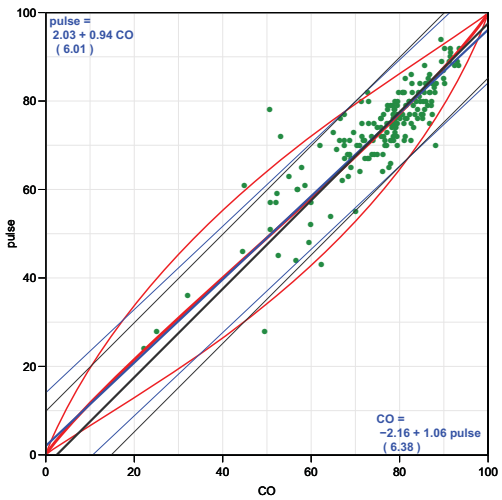
This is an analysis for the *transformed* data.

Backtransformation for plotting

```
 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`:

```
plot( ARoxt, pl.type="conv" )
```

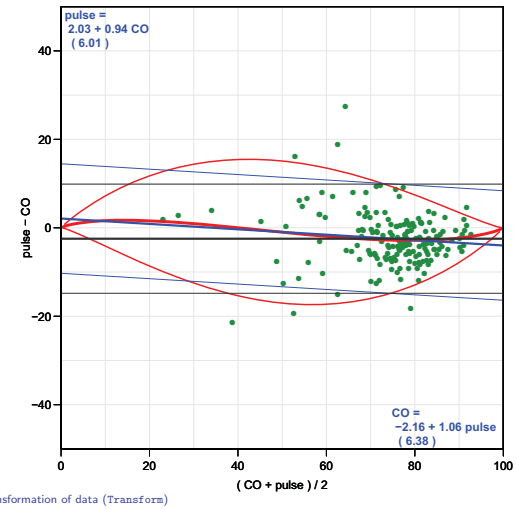
Transformation to a Bland-Altman plot

Just convert to the differences versus the averages:

```
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" )
```

Implementation in BUGS

Thursday 10 February, morning

Bendix Carstensen

MethComp
8–10 February 2011
Dept. Biostatistics, Univ. of Copenhagen
`www.biostat.ku.dk/~bxc/MethComp`

(BUGS-impl)

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 μs 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:

$$\begin{aligned} \alpha_{2|1} &= \alpha_2 - \alpha_1\beta_2/\beta_1 \\ \beta_{2|1} &= \beta_2/\beta_1 \end{aligned}$$

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` or `BRUGS` 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

```
> 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:
              alpha   beta     sd
To:   From:
CO    CO      0.000  1.000  1.740
      pulse  -9.342  1.159  5.328
pulse CO      8.061  0.863  4.508
      pulse   0.000  1.000  6.115
```


DG Altman and JM Bland.
Measurement in medicine: The analysis of method comparison studies.
The Statistician, 32:307–317, 1983.

B. Carstensen.
Comparing methods of measurement: Extending the LoA by regression.
Stat Med, 29:401–410, Feb 2010.

B Carstensen, J Simpson, and LC Gurrin.
Statistical models for assessing agreement in method comparison studies with replicate measurements.
International Journal of Biostatistics, 4(1):Article 16, 2008.

B Carstensen.
Comparing and predicting between several methods of measurement.
Biostatistics, 5(3):399–413, Jul 2004.

B. Carstensen.
Comparing Clinical Measurement Methods: A practical guide.
Wiley, 2010.

JM Bland and DG Altman.
Measuring agreement in method comparison studies.
Statistical Methods in Medical Research, 8:136–160, 1999.


```
 Variance components (sd):
       s.d.
Method    IxR   MxI   res
  CO    3.878 3.122 1.230
  pulse 3.222 2.757 4.324
Variance components with 95 % cred.int.:
    method     CO                  pulse
    qnt       50%   2.5%  97.5%    50%   2.5%  97.5%
SD
IxR         3.878  3.053  4.533  3.222  2.426  3.930
MxI         3.122  2.193  9.764  2.757  1.915  5.902
res         1.230  0.143  2.639  4.324  3.709  5.019
tot         5.220  4.507 10.645  6.135  5.457  7.849
```



```
Mean parameters with 95 % cred.int.:
                     50%    2.5%  97.5% P(>0/1)
alpha[pulse.CO]  8.057  -2.457 29.884   0.969
alpha[CO.pulse] -9.346 -49.949  2.476   0.031
beta[pulse.CO]   0.863   0.604  0.997   0.024
beta[CO.pulse]   1.159   1.003  1.657   0.976

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


The `MethComp` package

Also (currently) contains:

- `BA.plot` — make a Bland-Altman plot and compute limits of agreement.
- `BA.est` — estimates in the variance component model for the constant bias situation.
- `Deming` — regression with errors in both variables.
  A `.pdf` with a detailed derivation of the formulae (by Anders C Jensen) is included in the package too.
- A number of example data sets, amongst them all examples from [6].