

MA4605 Lecture 5A: Method Comparison Procedures

In Clinical Statistics, researchers often need to compare two methods of measurement to determine are in “*agreement*” i.e. whether these measurement methods give similar measurements consistently, and consequently whether or not these two methods can be used interchangeably.

Such a situation could arise if one method of measurement had high levels of accuracy and precision, but was expensive to use.

A cheaper method may be available, but for that method to be worthwhile this (“test”) method would have to agree sufficiently with the previous (“reference”) method.

Inter-Method Bias

We have encountered this problems in laboratory classes already, but (inappropriately) applying a simple linear regression to the problem.

Part V – Comparing analytical methods An ion-selective electrode (ISE) determination of sulphide from sulphate-reducing bacteria was compared with a gravimetric determination. The result, obtained were expressed in milligrams of sulphide.

Sulphide (ISE method): 108,12,152,3,106,11,128,12,160,128

Sulphide (gravimetry): 105,16,113,0,108,11,141,11,182,118

The mean of the case-wise differences is used as an estimate for the ***inter-method bias***: the tendency of one method to systematically out-measure the other.

$$\bar{d} = \frac{\sum_i^n (x_i - y_i)}{n}$$

Having a negligible mean of case-wise differences indicates that there is no inter-method bias, which is necessary for two methods to be considered in agreement.

Consider the table of measurement data on the next page. Methods A and B can be said to be in agreement. The inter-method bias is -0.763, not significantly different from 0. (N.B. the sign doesn't matter, all we are concerned with is the absolute value.)

Conversely there is a consistent tendency for Method D to give a reading greater than C. The inter-method bias between Methods C and D is -3.9346, which would be significantly different from 0. We would not say Methods D and C are in agreement.

(Remark: It may be possible to re-calibrate Method D to account for this bias, and remove this source of disagreement)

Equivalent Precision

Absence of a significant inter-method bias is not sufficient for two methods to be in agreement; both methods need to have the same level of precision also (i.e. large differences with opposite signs can cancel each other out).

Consider the comparison of Methods E and F. By inspection of the data, we can quickly determine that these methods do not give consistent results. However, if we were to compute the inter-method bias, we would find it to be 0.6458, not significantly different from 0.

	Method A	Method B		Method C	Method D		Method E	Method F
	99.31	101.14		98.645	101.42		106.44	97.48
	99.45	99.81		100.915	105.71		94.85	104.48
	102.05	102.74		100.67	105.52		103.14	97.6
	99.21	99.98		100.63	104.24		93.51	106.41
	100.7	102.82		100.175	103.18		95.86	107.22
	101.99	101.77		101.745	105.28		103.07	94.44
	100.39	101.28		99.63	103.41		94.49	109.75
	99.84	99.76		100.685	104.07		106.31	93.04
	99.1	101.45		98.21	104.62		97.55	103.17
	98.85	99.67		99.57	103.44		102.82	97.12
	100.29	101.11		99.76	103.63		105.29	94.97
	101.24	100.05		103.1	106.43		105.46	95.36
Mean	100.2017	100.9650		100.3113	104.2458		100.7325	100.0867
Variance	1.2272	1.2633		1.7359	1.8752		25.6389	33.1726
Bias	-0.7633			-3.9346			0.6458	

Using the Paired t-test

We will use a paired t-test to determine if there is inter-method bias present between the two methods of measurement.

```
> ISE
[1] 108 12 152 3 106 11 128 12 160 128
> Grav
[1] 105 16 113 0 108 11 141 11 182 118
>
> Diff =ISE-Grav
> Diff
[1] 3 -4 39 3 -2 0 -13 1 -22 10
>
> mean(Diff)
[1] 1.5

> t.test(ISE,Grav,paired=TRUE)
Paired t-test
data: ISE and Grav
t = 0.2973, df = 9, p-value = 0.773

alternative hypothesis: true difference in means
is not equal to 0

95 percent confidence interval:
-9.912129 12.912129

sample estimates:
mean of the differences
1.5
```

Here we are testing the null hypothesis that the mean of the case-wise differences is zero.

As we have a high p-value, we fail to reject the null hypothesis. As we conclude that the mean of the case-wise differences is zero; we conclude that there is no inter-method bias between the two methods of measurement.

Why is Simple Linear Regression unsuitable?

Recall the nature of the independent and dependent variables. The independent variable X is said to "cause" Y . In the case of two methods of measurement, both variables are in fact response ("Y") variables.

Another thing to consider is that in SLR models, the measurement error is associated with the Y variable only. This is not consistent with the case of comparing two measurements, where some level of measurement error is associated with each individual value.

Bland and Altman Plots

The Bland-Altman plot (Bland & Altman, 1986 and 1999), or difference plot, is a graphical method to compare two measurements techniques.

In this graphical method the differences (or sometimes the ratios) between the two techniques are plotted against the averages of the two techniques.

Where X and Y are the underlying sets of measurements

$$\textbf{Case-wise means:} \quad \mathbf{A_i = (X_i + Y_i) / 2}$$

$$\textbf{Case-wise differences:} \quad \mathbf{D_i = X_i - Y_i}$$

Further to the calculations on the previous piece of R code, we can determine the mean and standard deviation of the case-wise differences and Limits of Agreement.

Limits of agreement are an estimate of the interval within which a proportion of the differences between measurements lie.

The limits of agreement includes both systematic (bias) and random error (precision), and is a useful measure for comparing the likely differences between individual results measured by two methods.

Method E	Method F	Casewise Difference
106.44	97.48	8.96
94.85	104.48	-9.63
103.14	97.6	5.54
93.51	106.41	-12.9
95.86	107.22	-11.36
103.07	94.44	8.63
94.49	109.75	-15.26
106.31	93.04	13.27
97.55	103.17	-5.62
102.82	97.12	5.7
105.29	94.97	10.32
105.46	95.36	10.1
d.bar		0.6458
S.d		10.6608

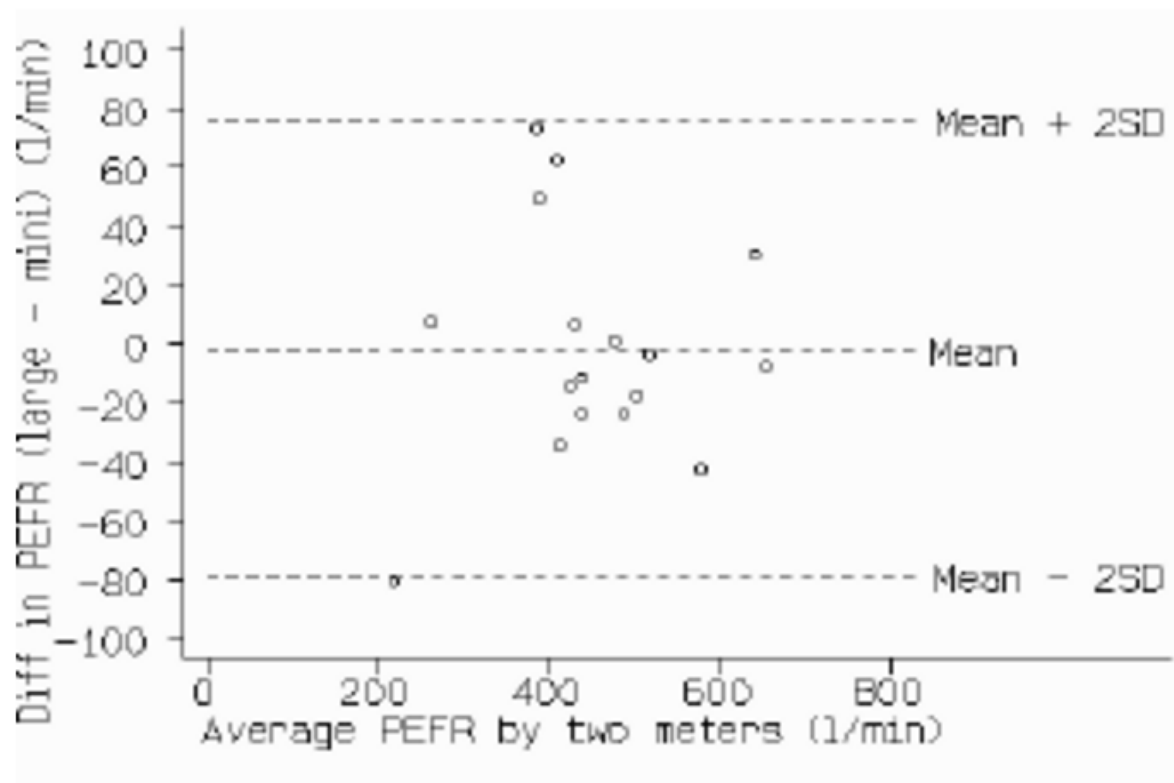
$$LOA = \bar{d} \pm (1.96 \times S_d)$$

```

> D.bar
[1] 0.6458
>
> Sdiff
[1] 10.6608
>
> LOA
[1] -20.2493 21.5409

```

Example of a Bland-Altman plot from a Scientific Publication



Orthonormal Regression

Orthonormal regression (more commonly called Deming Regression) is the term used in laboratory medicine to refer to linear regression analysis in which the random error of both variables is taken into account.

Orthonormal Regression assumes that the variances of the measurement errors for both methods are equal.

Deming regression is actually an extension of Orthonormal regression (rather than an alternative name) which allows the variance ratio of measurement errors to be specified by the user, and there is no requirement for measurement error variances to be equal.

The two terms are now used interchangeably. However many studies refer to Deming regression, when in fact what they are using is Orthonormal regression. It is important to know the difference.

MethComp R Package

To implement Deming regression, we require the use of the **R** package, known as **MethComp**. This package was written by Bendix Carstensen, (Steno Diabetes Centre, Copenhagen)

```
> D.bar
[1] 1.5
>
> Sdiff
>
> library(MethComp)
> Deming(trig,gerb)

Intercept Slope sigma.trig sigma.gerb
-0.08025171 1.02870424 0.05679647 0.05679647
>
> Deming(gerb,trig)

Intercept Slope sigma.gerb sigma.trig
0.07801243 0.97209670 0.05679647 0.05679647
```

We have two Deming regression equations, depending on which order we specify the variables.

- $trig = -0.0802 + 1.028\ gerb$
- $gerb = 0.0780 + 0.9720\ trig$

However it can quickly be algebraically determined that these two equations are merely re-expressions of each other, notwithstanding some rounding error.

- $gerb = 0.0780 + 0.9720\ trig$
- $gerb - 0.0780 = 0.9720\ trig$
- $(gerb/0.9720) - (0.0780/0.9720) = trig$
- $1.028\ gerb - 0.0802 = trig$

Variance Ratios

Orthonormal Regression assumes that the variances of the measurement errors for both methods are equal. If one method has a higher level of precision than the other, this is not a valid assumption.

The Deming regression approach allows a variance ratio of the measurement errors to be specified, but there is no accepted way for determining an estimate.

While these factors undermine the Deming/Orthonormal regression approach, they are still preferable to simple linear regression.

Other Approaches

As computational power has been improved significantly since the time that Orthonormal /Deming Regression and Bland-Altman Methods were devised, and as such, more appropriate techniques have been devised to assess the agreement of two methods.

One key approach is ***Linear Mixed Effects Models***, which can properly account for repeated measured by both methods on each subject.

LME models are quite complex, and also outside the scope of your course. However should you be carrying out Method Comparison Procedures in your careers, consider LME software to test for agreement.