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 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.

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.

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.

However this is not sufficient; both methods need to have the same level of precision also (i.e. large differences with opposite signs can cancel each other out)

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.

<u>Why is Simple Linear Regression unsuitable?</u>

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

The Bland-Altman plot (Bland & Altman, 1983, 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

Case-wise means: $A_i = (X_i+Y_i)/2$

Case-wise differences: $D_i = X_i-Y_i$

Additional to the Bland-Altman plot is the ***Limits of Agreement***, which are defined as

where is the mean of the case-wise differences and is the standard deviation of the case-wise differences.

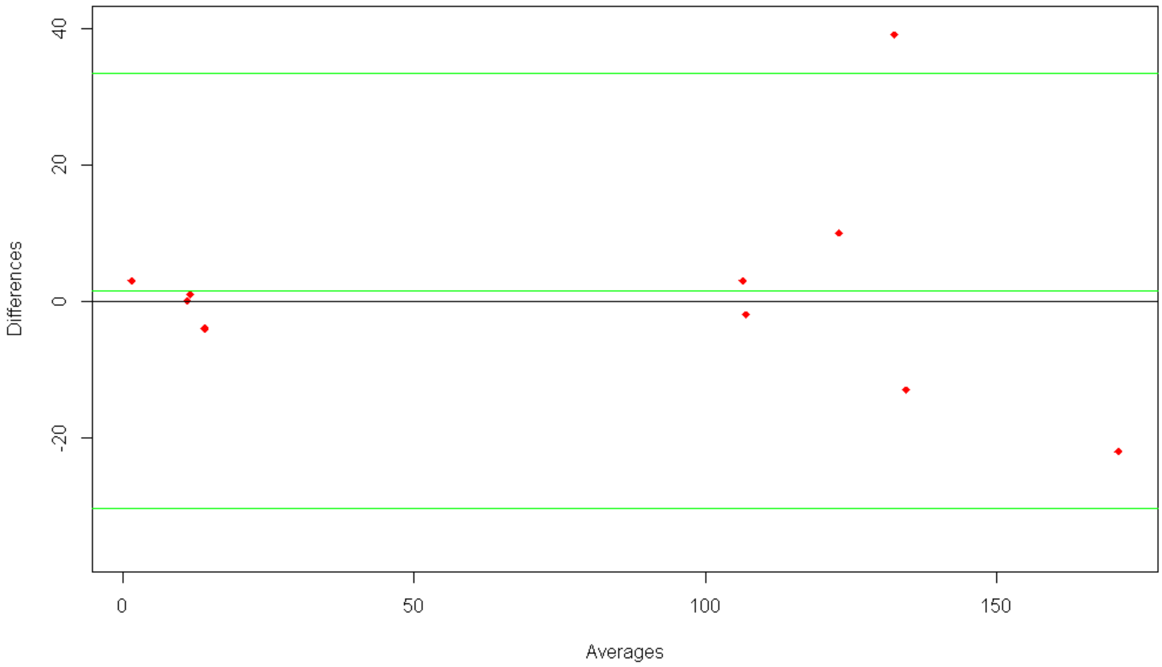
```
ISE =c(108,12,152,3,106,11,128,12,160,128)
Grav=c(105,16,113,0,108,11,141,11,182,118)

D.bar=mean(ISE-Grav)
Sdiff=sd(ISE-Grav)

# Computing the limits of agreement

LOA=c(D.bar-2*Sdiff,D.bar+2*Sdiff)

Averages = (ISE + Grav)/2
Differences = ISE - Grav

plot(Averages,Differences,pch=18,col="red",
ylim=c(1.2*LOA[1],1.2*LOA[2]))

# Put in the horizontal lines

abline(h=0)
abline(h=D.bar,col="green")
abline(h=LOA[1],col="green")
abline(h=LOA[2],col="green")
```

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.

```
> D.bar
[1] 1.5
> Sdiff
[1] 15.95306
> LOA
[1] -30.40611  33.40611
```

Two methods are said to be in agreement if the limits of agreement are acceptably narrow. However interpretation is hindered by the fact that it is unclear as to what constitutes an acceptable threshold for agreement.

The Bland-Altman approach was an improvement on what preceded it, but is now a source of controversy itself. Hence other approaches have been adopted for method comparison.



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 an extension of Orthornomal regression which allows the variance ratio of measurement errors to specified by the user.

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.

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)

```
> 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.