Comparing within-subject variances in a study to compare two methods of measurement

Martin Bland, 22 April, 2001

In the design for comparing two methods of measurement proposed by Bland and Altman (1986), two observations are made by each method on each subject. This design was use to compare a Wright peak flow meter and a min Wright peak flow meter. The following measurements of peak expiratory flow (litres/min) were obtained:

Subject	Wright meter		Mini meter	
	Obs 1	Obs 2	Obs 1	Obs 2
1	494	490	512	525
2	395	397	430	415
3	516	512	520	508
4	434	401	428	444
5	476	470	500	500
6	557	611	600	625
7	413	415	364	460
8	442	431	380	390
9	650	638	658	642
10	433	429	445	432
11	417	420	432	420
12	656	633	626	605
13	267	275	260	227
14	478	492	477	467
15	178	165	259	268
16	423	372	350	370
17	427	421	451	443

We recommended that the repeatability should be calculated for each method separately and compared. I was recently asked how we could carry out a statistical comparison of the two repeatabilities.

The problem is how to compare the within-subject standard deviations in a matched sample.

Denote the pairs of measurements by the same method on subject i by x_i and y_i. The standard deviation for a single subject s_i is given by the following formula for variance, i.e. standard devation squared:

$$s_i^2 = \frac{1}{2-1}\left(x_i^2 + y_i^2 - \frac{(x_i + y_i)^2}{2} \right)$$

$$= \frac{x_i^2}{2} + \frac{y_i^2}{2} - x_i y_i$$

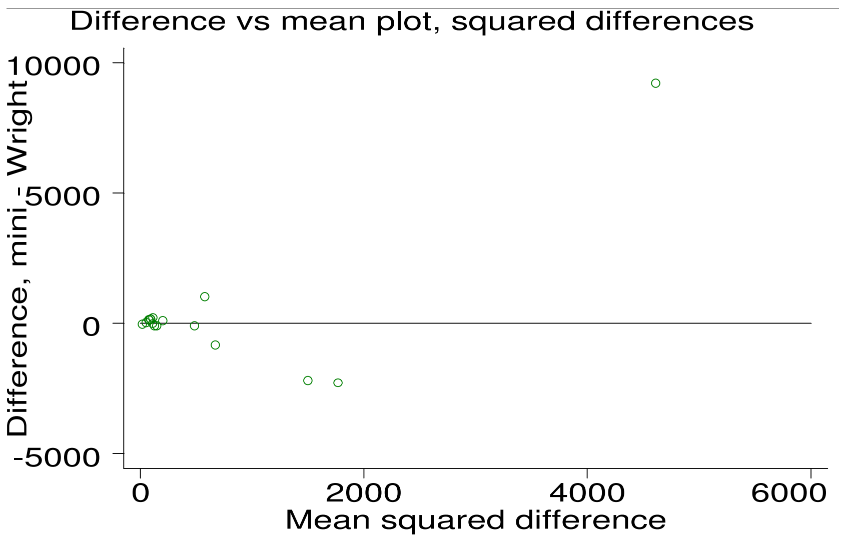
$$= \frac{1}{2}(x_i - y_i)^2$$

Hence for each subject the squared difference $(x_i - y_i)^2$ is an estimate of the within-subject variance for that method of measurement times 2, and the absolute value $|x_i - y_i|$ is an estimate of the within-subject standard deviation for that method of measurement times root 2. We can compare these estimates between the two methods of measurement using the two sample t method. It is usually preferable to compare variances rather than to compare standard deviations directly.

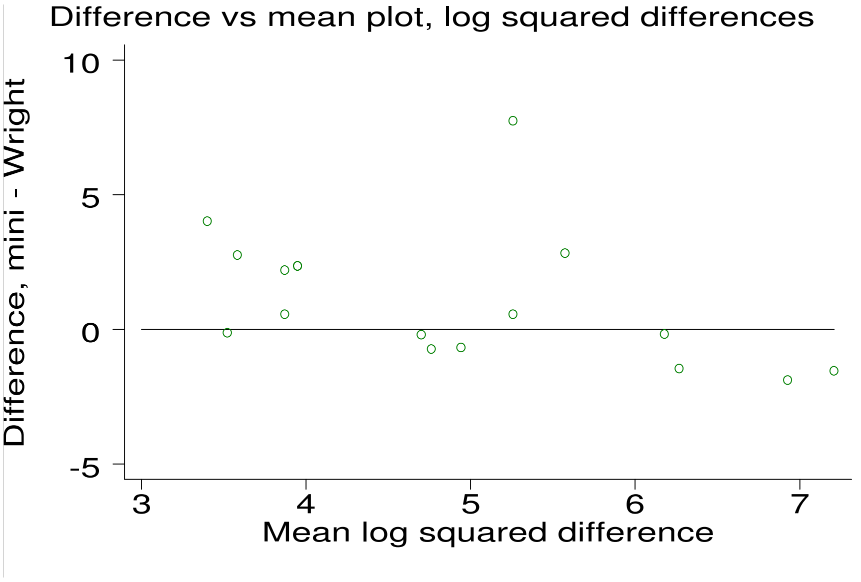
For the PEFR meter data, the squared differences are:

Subject	Wright meter	Mini meter
1	16	169
2	4	225
3	16	144
4	1089	256
5	36	0
6	2916	625
7	4	9216
8	121	100
9	144	256
10	16	169
11	9	144
12	529	441
13	64	1089
14	196	100
15	169	81
16	2601	400
17	36	64

For the paired t method, the differences between the squared differences by the two methods should follow a Normal distribution and be unrelated to the average squared difference for the subject. This is clearly not the case here, as the graph shows:



The assumptions of the paired t method a re clearly not met in this case and I suspect that this will always be so. A log transformation of the squared differences is quite effective:



One of the differences for the Wright meter was zero. It was replaced by half the next smallest value, 64, for this analysis.

Proceeding with the paired t test (Stata output) we get:

```
One-sample t test                                      Number of obs =       17

------------------------------------------------------------------------------
Variable |       Mean    Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   lslws |   1.098972    .5972562   1.84003   0.0844      -.1671547    2.365098
------------------------------------------------------------------------------
Degrees of freedom: 16
```

Thus in the example there is only very weak evidence that there is a difference between the within-subject variances. Antilogging the mean difference we get exp(1.098972) = 3.00, showing that the within-subject variance for the mini meter is estimated to be 3 times that for the Wright meter, but there is a very wide confidence interval for this ratio, from exp(-0.1671547) = 0.85 to exp(2.365098) = 10.65.

The square root of the ratio of within-subject variances will be the ratio of the within-subject standard deviations for the two methods of measurement.

Reference

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307-10.