Measurement error

J Martin Bland, Douglas G Altman

This is the 21st in a series of occasional notes on medical statistics.

Several measurements of the same quantity on the same subject will not in general be the same. This may be because of natural variation in the subject, variation in the measurement process, or both. For example, table 1 shows four measurements of lung function in each of 20 schoolchildren (taken from a larger study[1]). The first child shows typical variation, having peak expiratory flow rates of 190, 220, 200, and 200 l/min.

Let us suppose that the child has a "true" average value over all possible measurements, which is what we really want to know when we make a measurement. Repeated measurements on the same subject will vary around the true value because of measurement error. The standard deviation of repeated measurements on the same subject enables us to measure the size of the measurement error. We shall assume that this standard deviation is the same for all subjects, as otherwise there would be no point in estimating it. The main exception is when the measurement error depends on the size of the measurement, usually with measurements becoming more variable as the magnitude of the measurement increases. We deal with this case in a subsequent statistics note. The common standard deviation of repeated measurements is known as the *within-subject standard deviation*, which we shall denote by ζ_w.

To estimate the within-subject standard deviation, we need several subjects with at least two measurements for each. In addition to the data, table 1 also shows the mean and standard deviation of the four readings for each child. To get the common within-subject standard deviation we actually average the variances, the squares of the standard deviations. The mean within-subject variance is 460.52, so the estimated within-subject standard deviation is $\zeta_w = \sqrt{460.5} = 21.5$ l/min. The calculation is easier using a program that performs one way analysis of variance[2] (table 2). The value called the residual mean square is the within-subject variance. The analysis of variance method is the better approach in practice, as it deals automatically with the case of subjects having different numbers of observations. We should check the assumption that the standard deviation is unrelated to the magnitude of the measurement. This can be done graphically, by plotting the individual subject's standard deviations against their means (see fig 1). Any important relation should be fairly obvious, but we can check analytically by calculating a rank correlation coefficient. For the figure there does not appear to be a relation (Kendall's $\tau = 0.16$, $P = 0.3$).

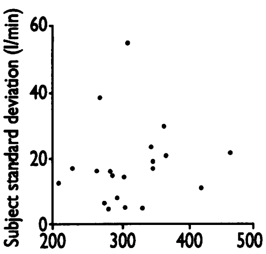

Fig 1—*Individual subjects' standard deviations plotted against their means*

A common design is to take only two measurements per subject. In this case the method can be simplified because the variance of two observations is half the square of their difference. So, if the difference between the two observations for subject I is d_i the within-subject standard deviation ζ_w is given by when n is the number of subjects. We can check for a relation between standard deviation and mean by plotting for each subject the absolute value of the difference—that is, ignoring any sign—against the mean.

Table 1—*Repeated peak expiratory flow rate (PEFR) measurements for 20 schoolchildren*

Child No	PEFR (l/min) 1st	2nd	3rd	4th	Mean	SD
1	190	220	200	200	202.50	12.58
2	220	200	240	230	222.50	17.08
3	260	260	240	280	260.00	16.33
4	210	300	280	265	263.75	38.60
5	270	265	280	270	271.25	6.29
6	280	280	270	275	276.25	4.79
7	260	280	280	300	280.00	16.33
8	275	275	275	305	282.50	15.00
9	280	290	300	290	290.00	8.16
10	320	290	300	290	300.00	14.14
11	300	300	310	300	302.50	5.00
12	270	250	330	370	305.00	55.08
13	320	330	330	330	327.50	5.00
14	335	320	335	375	341.25	23.58
15	350	320	340	365	343.75	18.87
16	360	320	350	345	343.75	17.02
17	330	340	380	390	360.00	29.44
18	335	385	360	370	362.50	21.02
19	400	420	425	420	416.25	11.09
20	430	460	480	470	460.00	21.60

The measurement error can be quoted as ζ_w. The difference between a subject's measurement and the true value would be expected to be less than $1.96\ \zeta_w$ for 95% of observations. Another useful way of presenting measurement error is sometimes called the *repeatability*, which is $\sqrt{2} \times 1.96\ \zeta_w$ or $2.77\ \zeta_w$. The difference between two measurements for the same subject is expected to be less than $2.77\ \zeta_w$ for 95% of pairs of observations. For the data in table 1 the repeatability is $2.77 \times 2.5 = 60$ l/min. The large variability in peak expiratory flow rate is well known, so individual readings of peak expiratory flow are seldom used. The variable used for analysis in the study from which table 1 was taken was the mean of the last three readings.[1]

Other ways of describing the repeatability of measurements will be considered in subsequent statistics notes.

1 Bland JM, Holland WW, Elliott A. The development of respiratory symptoms in a cohort of Kent schoolchildren. *Bull Physio-Path Resp* 1974;10:699-716.

2 Altman DG, Bland JM. Comparing several groups using analysis of variance. *BMJ* 1996;312:1472.

Table 2—*One way analysis of variance for the data of table 1*

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio (F)	Probability (P)
Children	19	285318.44	15016.78	32.6	<0.0001
Residual	16	27631.25	460.52		
Total	79	312949.69			

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
J Martin Bland, *professor of medical statistics*

IRCF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, PO Box 777, Oxford OX3 7LF
Douglas G Altman, *head*

Correspondence to: Professor Bland.

BMJ 1996;312:1654