Repeated measurement and analysis units

Review of basic principles

Jonas Ranstam

The NEPI Foundation, Malmö University Hospital, SE-205 02 Malmö, Sweden. Tel +46 40-332664. Fax -336215
Submitted 98-05-19. Accepted 98-06-02

Much research is performed using repeated measurements, i.e., every subject contributes more than one observation. This arises naturally in studies on knees, elbows, fingers and other multiple body parts. In some experiments, repeated measurements are also used to reduce the number of subjects needed for the study by getting more accurate information from each subject than would be the case with a single measurement.

An observed difference between two groups of subjects is usually considered statistically significant if it is less than 5% likely that it was caused only by normal variations between subjects. This is often assessed with a statistical test based on the assumption that observations are independent. Violations of this assumption easily lead to false conclusions.

Assume that 6 subjects were divided into 2 groups and had their systolic blood pressure measured in each arm, leaving the data as presented in Table 1. If measurements performed on a subject's left and right arms measure the same blood pressure, any variation between these is attributable to measurement errors only. Hence, when a subject's blood pressure is measured in one of the arms, all of the information on his or her blood pressure has been acquired (assume that blood pressure is constant during the examination and that no physiological anomaly exists). Measurements of blood pressure in the other arm contribute only information on the accuracy of the measurement.

We could thus test the differences in systolic blood pressure between the subjects in groups 1 and 2, using a t-test of the left arm recordings (n 6, $p = 0.4$) or of the right arm recordings (n 6, $p = 0.07$). However, since we have two measurements from each subject, it would be more reliable to use the mean of the two measurements (n 6, $p = 0.1$) and we would also avoid the subjectivity in choosing one of the arms. Hitherto, all the results have been based on correct units of analysis, i.e., the patient. None of them indicates statistically significant differences between the two groups.

An incorrect analysis could be performed using the same dataset, re-arranged as in Table 2. A t-test of the mean difference in systolic blood pressure between

Table 1. Data from a hypothetical experiment on systolic blood pressure measured using both arms

	Systolic pressure (mmHg)		
Subject	Left arm	Right arm	Group
1	125	130	1
2	140	135	1
3	125	125	1
4	130	140	2
5	135	145	2
6	145	135	2

Left arm: Group mean difference (standard deviation in groups 1 and 2, respectively) = 6.7 mmHg (8.7, 7.6).
Right arm: Group mean difference (standard deviation in groups 1 and 2, respectively) = 10.0 mmHg (5.0, 5.0).
Mean arm: Group mean difference (standard deviation in groups 1 and 2, respectively) = 8.3 mmHg (6.6, 2.9)

Table 2. Data set-up when ignoring the link between the two measurements

Measurement	Systolic pressure (mmHg)	Group
Subject 1, left arm	125	1
Subject 1, right arm	130	1
Subject 2, left arm	140	1
Subject 2, right arm	135	1
Subject 3, left arm	125	1
Subject 3, right arm	125	1
Subject 4, left arm	130	2
Subject 4, right arm	140	2
Subject 5, left arm	135	2
Subject 5, right arm	145	2
Subject 6, left arm	145	2
Subject 6, right arm	135	2

Group mean difference (standard deviation in groups 1 and 2, respectively) = 8.3 mmHg (6.3, 6.0).

Acta Orthop Scand 1998; 69 (4): 345–346

measurements from groups 1 and 2 would then indicate statistical significance (n 12, p = 0.04). The difference observed in mean systolic blood pressure between the groups will be exactly the same as in the analysis based on the arm mean values above. Since mean values vary less than their components, the variations between the units are, on average, slightly greater. This is an often-used argument against single measurements. More importantly, however, due to the doubled number of analysis units, the p-value is too low; remember that the second reading of a subject's blood pressure gave no new information. When using the recording as unit of analysis, we mean that all recordings provide an equal amount of information. As a consequence of this fault, the difference in systolic blood pressure between the two groups appears to be statistically significant.

The importance of using correct units of analysis has been reported previously and simple yet correct ways to analyze data from repeated measurement experiments have been suggested (Matthews et al. 1990, Morris 1993). Other, more complex, statistical methods exist, but these are, in general, not easily accessible.

Matthews J N S, Altman D G, Campbell M J, Royston P. Analysis of serial measurements in medical research. Br Med J 1990; 300: 230-5.

Morris R W. Bilateral procedures in randomised controlled trials. JBJS (Br) 1993; 75: 675-6.