

TOO MANY VARIABLES, OR TOO FEW SUBJECTS?

K.G. RUSSELL

1. INTRODUCTION

I have been presented with a number of longitudinal studies in which the number of variables, p , has been quite large relative to the number of subjects (cases), n , in the study. I have been concerned that, as a result, many of the statistical tests have had little power. It has also worried me that I have not been able to find any published recommendations of how to determine an appropriate number of subjects for each treatment group. To motivate discussion, an example is given of one particular study which came to my attention.

2. EXAMPLE

The study was conducted by an M.Sc.(Hons) student in the Department of Human Movement Science at The University of Wollongong. His aim was to compare the effects of four treatments on acute stress. These treatments consisted of a Control, a Placebo (listening to music one lunchtime per week), Yoga, and an Exercise programme. Twelve female students under 25 years of age were allocated at random to each treatment group. Four students eventually dropped out of the Yoga group. As a result, the Yoga group had eight subjects, and the other three groups had twelve subjects each.

To create stress, students were required to follow a light around a rotor, and they were scored according to the length of time that they kept a pointer in contact with the light. This was done ten times during a testing session. Observations were taken on a number of variables at one testing session early in the exercise programme, and then at a second session at the end of the programme. The variables which were thought to measure stress were recorded at the beginning, middle and end of the testing session on each of these occasions. This may be represented schematically by the diagram in Table 1.

In all, 97 readings were taken on different variables. Many of these were in the nature of repeated measurements, but there were 18 or 19 separate variables whose values were recorded. They are set out in Table 2, together with the number of measurements made on each one.

Many of the analyses requested were repeated-measures ANOVAs in

which the effects of the four treatments were compared. The response variables were the six repeated measurements on the particular characteristic being measured. For example, one might test whether the patterns of systolic blood pressure readings over the 2 sessions and 3 readings per session were the same for the four treatments.

Testing Session 1
(early in the programme)

Testing Session 2
(at the end of the programme)

Reading taken here →

•

← Reading taken here

Test 1

Test 2

Test 3

Test 4

Test 5

Reading taken here →

•

← Reading taken here

Test 6

Test 7

Test 8

Test 9

Test 10

Reading taken here →

•

← Reading taken here

Table 1: *Schematic representation of the repeated measurements made on a variable*

In other cases, the researcher was interested in the changes between the pre-stress score and post-stress score at the first and second observation sessions. As these readings were often thought to be affected by the initial score on that particular characteristic (for example, tension), this required an Analysis of Covariance using repeated measures.

In addition, it was suggested that we might examine in the one analysis four repeated measurements on each of nine different

variables, using the first reading on each variable as a covariate.

Such requests caused me considerable disquiet, in part because of the difficulties of interpreting the results of some analyses, but chiefly because of the small numbers of subjects in each group.

	<i>Name of Variable</i>	<i>Number of Measurements</i>
	Fitness	2
	Flexibility	2
	Body Fat	2
	JAS (Standard)	2
	JAS (Percentile)	2
	Competition Orientation	1
	Win orientation	1
	Goal orientation	1
	Systolic BP	2 × 3
	Diastolic BP	2 × 3
	Motor performance	2 × 10
Profile of Mood States (POMS)	(Tension	2 × 3
	Depression	2 × 3
	Anger	2 × 3
	Vigour	2 × 3
	Fatigue	2 × 3
	Confusion	2 × 3
	Total Mood Disturbance	2 × 3
	Heart Rate	2 × 5

TABLE 2: *Measurements Taken on Each Individual in the Study*

3. DISCUSSION

The problem of how many subjects one should have per group has arisen on a number of occasions, and I have found little in the books I have read to assist me in advising clients on the number of subjects which they should have in each group. In one of the few references to sample size I have found, FLURY & RIEDWYL [2; page 9] state *If n is not considerably larger than p, say at least three or four times as large, multivariate methods are often not very powerful and depend too*

strongly on certain assumptions. For descriptive methods like principal component analysis it is desirable, as a rule of thumb, for n to be at least ten times as large as p . In discussion following the present paper, I understood Professor D.J. Hand to suggest that the number of subjects should be about nine or ten times as large as the number of variables. In the Study being discussed, the number of subjects in a group was often less than the number of variables being considered in the one analysis.

4. TWO RULES OF THUMB

(A) One assumption of the multivariate Analysis of Variance is that the covariance matrices of the p variables being analysed are equal in the various treatment groups. Box's M-test of the equality of the covariance matrices requires that the sample covariance matrices be non-singular, which in turn requires that the number of subjects in a group, n , must satisfy $(n - 1) \geq p$. I feel that one ought not to carry out a test if the number of subjects in a group is so small that one cannot examine whether the underlying assumptions of the test are valid. Therefore, the minimum number of subjects per group ought to be one more than the largest number of variables which one wants to examine simultaneously (here counting four repeated measurements on the one variable as four different variables).

In later discussions, Dr Ari Verbyla warned me against using Box's M-test, because of its lack of robustness to departures from Normality. I agree with this warning, but I do not feel that it warrants a change to this rule. If the number of subjects per group is so small that one cannot even use a test in the idealised circumstances for which it is appropriate, then I do not believe that one should use a smaller number of subjects in a less ideal situation. In the study described here, several of the variables being measured on each subject were ordinal, which suggests to me that the number of subjects per group should be larger than the number suggested by the rule described above.

(B) In discussions with a researcher who consulted me, it seemed to me that he considered b repeated measurements on the one variable from one individual to be carrying the same amount of information in a comparison of group means as one single observation from each of b separate individuals. This prompted me to wonder: *how many independent observations is a repeated measurement worth?*

In unpublished work presented to the 10th Australian Statistical Conference in 1990, I assumed a correlation between consecutive repeated measurements of ρ , and postulated two models:

- the *split-plot model*, in which the correlation of ρ is assumed to hold between any two of the b repeated measurements on the variable in question; and
- the *autoregressive model*, in which the correlation between two repeated measurements b time units apart is $\rho^{|b|}$.

By comparing the variance of the mean of b repeated measurements to the mean of b independent observations (assuming a common value of σ^2 throughout), I calculated the value of b repeated measurement to be:

$\frac{b}{1 + (b-1)\rho}$ ($\rightarrow \frac{1}{\rho}$ as $b \rightarrow \infty$) independent observations in the split-plot model; and

$\frac{b^2(1-\rho)^2}{b(1-\rho)^2 + 2(\rho^{b+1} - b\rho^2 + (b-1)\rho)}$ ($\rightarrow b(\frac{1-\rho}{1+\rho})$ as $b \rightarrow \infty$) independent observations in the autoregressive model.

In my (limited) experience, the correlation between repeated measurements i time units apart is usually less than ρ but greater than $\rho^{|i|}$, so that the values in the previous paragraph probably represent extremes. These could be taken as lower and upper limits for the amount of information in b repeated measurements on a given variable.

Then one could use the well-known formula

$$n \geq 2\left(\frac{\sigma}{\delta}\right)^2 (z_{\alpha/2} + z_{\beta})^2$$

(for example, see COCHRAN & COX [1; Section 2.2]) for the numbers of observations needed in samples from each of two Normal populations with a common standard deviation σ if a z-test at the $100\alpha\%$ level of significance is to have power $1 - \beta$ of detecting a difference between the population means at least as large as δ . With a suitable estimate for ρ , one could then allow for the additional information given by the b repeated measurements to determine a minimum sample size for each treatment group.

For example, to have 90% probability of detecting a significant difference between population means which are 5 units apart if the standard deviation of each group is 3 units, a test at the 5% level would

require at least $2\left(\frac{5}{3}\right)^2(1.96 + 1.2816)^2 = 58.4$ (approx.) observations per group. If there were to be four repeated measurements on the variable in question, and if we were to assume a correlation between consecutive repeated measurements of 0.3, the split-plot model suggests that the four repeated measurements would contribute as much information on the group mean as $\frac{4}{1 + 3 \times 0.3} = 2.1$ independent observations, and we could reduce the number of subjects per group to $58.4/2.1 = 28$ (approx) subjects per group for the purpose of inter-group comparisons. Using the autoregressive model, the four repeated measurements would contribute as much information as $\frac{4^2 \times 0.49}{4 \times 0.49 + 2(0.3^5 - 4 \times 0.09 + 3 \times 0.3)} = 2.57$ independent observations, suggesting that we could reduce the number of subjects per group to $58.4/2.57 = 23$ (approx.) subjects per group. One might then feel reasonably happy in recommending to a client that 28 (or 30, say) subjects per group should provide the required power.

REFERENCES

- [1] COCHRAN, W.G. and COX, G.M., (1957) *Experimental Designs*, 2nd edition New York: Wiley.
- [2] FLURY, B. and RIEDWYL, H., (1988) *Multivariate Statistics: A Practical Approach* London: Chapman and Hall

Department of Mathematics
The University of Wollongong
P.O. Box 1144
Wollongong, N.S.W. 2500
Australia