A REVIEW OF SOME METHODS OF ANALYSIS OF REPEATED MEASUREMENTS EXPERIMENTS

Margaret A. Cecce
Biometrics Unit, Cornell University

BU-772-M*

May 1982

Abstract

The necessary and sufficient conditions for F-tests to be appropriate in univariate analysis are discussed. Alternative approaches to the statistical analysis are also discussed for the case where these conditions are not met, through a review of the pertinent literature.

---

*BU-772-M in the Biometrics Unit Series, Cornell University, Ithaca, NY 14853.

A REVIEW OF SOME METHODS OF ANALYSIS OF REPEATED MEASUREMENTS EXPERIMENTS

Margaret A. Cecce
Biometrics Unit, Cornell University

BU-772-M May 1982

A repeated measure design or experiment is one in which treatments are randomly assigned to a set of experimental units, with the response measured on each experimental unit repeatedly over the duration of the experiment. The rationale behind taking multiple measurements on an individual experimental unit is that the error variability within an experimental unit should be considerably smaller than that among different experimental units.

A pretest-posttest experiment is an example of a repeated measure experiment. A possible pretest-posttest experiment would be one designed to determine the effectiveness of different teaching methods. A pretest is given to groups of students, then is followed by a treatment period, followed by the posttest. Each student serves as his own control for the experiment.

Another repeated measures example is a nutrition experiment in which a set of experimental animals is randomly assigned to three groups. Each group receives each treatment for a limited amount of time, and two or more measurements are taken through time on each experimental unit for each treatment. The sequence of treatments is randomly assigned to a group. The treatments may be different food additives with the measurement being weight gain.

A repeated measures design may have a built-in analytic problem associated with the variance-covariance structure of the design. The assumption of independence of observations, necessary for standard statistical procedures, is violated in a repeated measures design since several observations are taken on the same experimental units (usually through time). It is assumed that some degree of 1st order correlation exists with the possible existence of higher order correlations.

Several approaches are possible for the analysis of a repeated measures investigation. These include: a univariate analysis of variance with standard F-test, a univariate analysis of variance using F-tests with the degrees of freedom adjusted for failure to meet the assumptions, and a multivariate analysis of variance.

In order to use the univariate analysis of variance with standard F-tests, one must consider the conditions necessary for the existence of an exact F distribution. The literature, from various disciplines, seems to confuse the issue by variously stating the need for uniformity, sphericity, or circularity as conditions for valid F-tests. Are these conditions necessary and sufficient for valid F-tests; if so, are they equivalent; if not, what constitutes the necessary conditions for valid F-tests?

Definition: Uniformity — A variance-covariance matrix, Σ, is said to be uniform if it is of the form

$$\sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} .$$

This is referred to as a type S matrix by Huynh and Feldt (1970), or a matrix exhibiting the property of compound symmetry by Geisser (1963). With a uniform variance-covariance structure all F-tests are valid with the ratios having exact F distributions. This uniformity condition is a familiar condition for valid F-tests, used by Winer (1971) and other basic texts.

Definition: Sphericity — Mauchly (1940) defined a population as spherical if it exhibited a symmetric variance-covariance matrix of the form $\sigma^2 \underset{\sim}{I}$. In other words, if the observations are independent, the Σ matrix is spherical. He developed a likelihood ratio test statistic for sphericity, W . W is defined as

$$W = V/\{(\text{tr } V)/p\}^p \quad ,$$

where V is the sample covariance matrix and p is the number of component vectors, Consul (1967). One additional property of spherical populations, described by Mauchly, is that spherical populations remain spherical under linear orthogonal transformations.

Sphericity ties in as a condition for valid F-tests in repeated measure designs through type H matrices. Huynh and Feldt (1970) define a matrix $\underset{\sim}{Y}$ which has a covariance matrix $\Sigma^{*} = \lambda I$ as type H; where $Y = CX$ and C is an orthogonal contrast matrix (such as a Helmert matrix). Type H matrices have the property of sphericity; therefore, F-tests of these contrasts are valid.

Mendoza, Toothaker and Crain (1976), in discussing the necessary and sufficient conditions for valid F-ratios in repeated measures design, state that the variance-covariance matrix must meet a circularity assumption. This circularity assumption is that $C\Sigma C' = \sigma^2 I$, which is the condition for sphericity defined by Huynh and Feldt (1970). Therefore, circularity and sphericity can be considered equivalent conditions. With a type H matrix we have $C\Sigma C' = \lambda I$, which also is a spherical form. Therefore, if the matrix is type H we also have the necessary and sufficient condition for valid F-tests in the analysis of variance. Huynh and Feldt also state that a type S matrix (uniformity) provides a sufficient condition for valid F-tests. If C is a Helmert matrix and Σ is type S, then it can be demonstrated that $C\Sigma C' = \lambda I$.

All of these criteria for valid F-ratios are based on either having independent observations or looking at contrasts in which we can obtain independence.

Evaluating the equality of variance assumptions of a repeated measures design is achieved by using the sample covariance matrix. Huynh and Mandeville (1979) discuss preliminary testing of the sample covariance matrix using Mauchly's sphericity statistic, W, to test for sphericity, followed by the Box criterion for testing equality of the covariance matrices. The authors state that simulation studies indicate that with departures from normality, Mauchly's W statistic has type I errors which vary with the tails of the distribution. It was also suggested that Box's criterion may also be sensitive to departures from normality.

Rogan, Keselman and Mendoza (1979) described four possible strategies for deciding a course of action following a test for sphericity or circularity as discussed above. The test for circularity does not predetermine which of the strategies should be employed. In each strategy an adjustment is made in the F-statistic degrees of freedom; the adjustment, e, is a multiplicative adjustment to both the numerator and denominator degrees of freedom. The strategies are as follows:

1. Using the conservative F-test (Geisser and Greenhouse, 1958) following a significant test for circularity or using the conventional F-test following a non-significant test for circularity. The conservative F-test is one which adjusts the F-test degrees of freedom by the lower bound of Box's (1954b) e, where

$$e = k^2(\sigma_{tt}-\bar{\sigma}_{..})^2/(k-1)\left\{\sum_{t=1}^{k}\sum_{s=1}^{k}\sigma_{ts}^2 - 2k\sum_{t=1}^{k}\bar{\sigma}_{t.}^2 - k^2\bar{\sigma}_{..}^2\right\} \quad ,$$

with k = number of columns and σ_{ts}'s are elements of the population covariance matrix. This lower bound is $(k-1)^{-1}$, where k is the number of occasions (columns) measured. This is the most conservative estimate of adjustments to the degrees of freedom of the F-tests.

2. Using the $\hat{e}$-adjusted F-test (Collier, et al., 1967) following a significant test for circularity or using the conventional F-test following a non-significant test for circularity. The $\hat{e}$ used here is a sample-based estimate of e . Each of the σ_{ts}'s are replaced by variance or covariance estimates based on the sample in hand. This estimate is biased, being negatively skewed for high values of $\hat{e}$ and positively skewed for small values of $\hat{e}$. With a sample of 15 observations per group Collier, et al. showed that the bias is reduced.

3. Using the $\tilde{e}$-adjusted F-test (Huynh and Feldt, 1970) following a significant test for circularity or using the conventional F-test following a non-significant test for circularity, where $\tilde{e}$ is defined as

$$\tilde{e} = \frac{n(k-1)\hat{e}-2}{[k-1][n-1-(k-1)\hat{e}]}$$

and n = number of rows (treatments). $\tilde{e} \geq \hat{e}$ with equality occurring at the lower bound. It is possible for $\tilde{e}$ to exceed 1; if this occurs the value of $\tilde{e}$ is taken to be 1 . Huynh and Feldt have shown in a Monte Carlo study that $\hat{e}$ is a better estimate in the neighborhood of .5, but when $\hat{e} > .75$, $\tilde{e}$ is less biased than $\hat{e}$ and, therefore, a better estimate for e .

4. Using a multivariate test following a significant test for circularity or using the conventional F-test following a non-signficant test for circularity. The multivariate analysis does not depend on the particular

form of the variance-covariance matrix, whereas the univariate analysis does. In the case where the sample covariance matrix varies considerably from the conditions of circularity, that is when e is small, .5 or less, then the multivariate procedures are appropriate.

In considering these different strategies it is necessary to compare the performance of the various estimates with respect to bias and precision of the estimates. The Monte Carlo study of Huynh and Feldt and simulation studies of Rogan, et al. indicate that the $\tilde{e}$-adjustment is most powerful and less biased than $\hat{e}$ for values of $.75 \leq e \leq 1$, and $\hat{e}$ is more powerful and less biased than $\tilde{e}$ for values of e in the range of $.5 < e < .75$. With values of $e \leq .5$, the multivariate analysis is most appropriate, although for small sample sizes the multivariate approach is less sensitive.

Greenhouse and Geisser (1959) propose a three-step procedure for the univariate approach. This procedure considers the difficulty of calculating $\hat{e}$ or $\tilde{e}$, described above from the sample covariance matrix, and suggests the following steps to avoid this calculation whenever possible.

Step 1. Test the F-ratio with the conventional degrees of freedom. This is the most liberal test, and if it is not significant none of the more conservative tests will be significant. Therefore, if the F-test is not significant here, stop; but if the test is significant go to step 2.

Step 2. Test the F-ratio with the Geisser and Greenhouse lower limit adjustment. This is the most conservative test, since the degrees of freedom are reduced the maximum amount. If the most conservative test is significant, any test between the most conservative and most liberal would also be significant. Therefore, if this test is significant, stop at this step.

If, on the other hand, this test is not significant, while the most liberal test is significant, we need to consider an actual estimate of e as the adjustment to the degrees of freedom. In that case go to step 3.

Step 3. Estimate e from S, the sample estimate of Σ, the covariance matrix. This can be accomplished using either the $\hat{e}$ or $\tilde{e}$ estimates discussed earlier. The F-ratios would then be tested with the degrees of freedom reduced by the $\hat{e}$ or $\tilde{e}$ estimate. This will fall between the conventional degrees of freedom and the lower bound degrees of freedom, and will allow a final decision on significance of the F-tests of interest.

It is evident that, for many sets of data, one would be able to stop the analysis before reaching step three, and therefore eliminate the need to make the calculations of an estimate of e .

In the general case of $Y = X\beta + \epsilon$, $Y \sim N(X\beta, \Sigma)$, the F-tests of the standard univariate analysis of variance holds if the $\text{cov}(y'Ay, y'By) = 0$. That is, the sums of squares of the different factors are independent of each other and of the error sums of squares. This is easily verified for the cases where sphericity or uniformity hold that all of these sums of squares are indeed independent using Theorem 2.3 and 2.4 from Searle's (1971) Linear Models.

References

Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. Annals of Mathematical Statistics 25:290-302.

Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and correlation between errors in the two-way classification. Annals of Mathematical Statistics 25:484-498.

Cole, J. W. L. and Grizzle, J. E. (1966). Applications of multivariate analysis of variance to repeated measurements experiments. Biometrics 22: 810-828.

Collier, R. O., Jr., Baker, F. B., Mandeville, G. K. and Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in repeated measures design. Psychometrika 32:339-353.

Consul, P. C. (1967). On the exact distribution of the criterion W for testing sphericity in a p-variate normal distribution. Annals of Mathematical Statistics 38:1170-1174.

Geisser, S. and Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics 29:855-891.

Greenhouse, S. W. and Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika 24:95-112.

Huynh, H. and Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. JASA 65: 1582-1585.

Huynh, H. and Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. Journal of Educational Statistics 1:69-82.

Huynh, H. and Mandeville, G. K. (1979). Validity conditions in repeated measures designs. Psychological Bulletin 86:964-973.

Kershner, R. P. and Federer, W. T. (1979). Two treatment crossover designs for estimating a variety of effects. BU-675-M. Cornell University.

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. Annals of Mathematical Statistics 11:204-209.

Mendoza, J. L., Toothaker, L. E. and Crain, B. R. (1976). Necessary and sufficient conditions for F ratios in the $L \times J \times K$ factorial design with two repeated factors. JASA 71:992-993.

Morrison, D. F. (1976). Multivariate Statistical Methods, McGraw-Hill, New York.

Rogan, J. C., Keselman, H. J. and Mendoza, J. L. (1979). Analysis of repeated measurements. British Journal of Mathematical and Statistical Psychology 32:269-286.

Searle, S. R. (1971). Linear Models, John Wiley and Sons, New York.

Timm, Neil H. (1975). Multivariate Analysis with Applications in Education and Psychology, Brooks/Cole Publishing Company, Monterey, California.

Winer, B. J. (1971). Statistical Principles in Experimental Designs, 2nd ed. McGraw-Hill, New York.

Woodward, J. A. and Overall, J. E. (1973). Nonorthogonal analysis of variance in repeated measures experimental designs. Educational and Psychological Measurement 36:855-859.