DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 5 | Issue 1 Article 5

5-1-2006

Limitations Of The Analysis Of Variance

Phillip I. Good
Information Research, Huntington Beach, CA., pigood@verizon.net

Clifford E. Lunneborg
University of Washington

Journal of Modern Applied Statistical Methods
May, 2006, Vol. 5, No. 1, 41-43

1538 – 9472/06/$95.00

Limitations Of The Analysis Of Variance

Phillip I. Good
Information Research
Huntington Beach, C.A.

Cliff Lunneborg
Department of Statistics
University of Washington

Conditions under which the analysis of variance will yield inexact p-values or would be inferior in power to a permutation test are investigated. The findings for the one-way design are consistent with and extend those of Miller (1980).

Key words: Analysis of variance, permutation tests, exact tests, robust tests, one-way designs, k-sample designs.

Introduction

The analysis of variance has three major limitations:

1. It is designed to test against any and all alternatives to the null hypothesis and thus may be suboptimal for testing against a specific hypothesis.

2. It is optimal when losses are proportional to the square of the differences among the unknown population means, but may not be optimal otherwise. For example, when losses are proportional to the absolute values of the differences among the unknown population means, expected losses would be minimized via a test that makes use of the absolute values of the differences among the sample means; see, for example, Good (2005).

3. It is designed for use when the observations are drawn from a normal distribution and though it is remarkably robust, it may not yield exact p-values when the observations come from distributions that are heavier in the tails than the normal. Even in cases when the analysis of variance yields almost exact p-values, it may be less powerful than the corresponding permutation test when the observations are drawn from non-normal distributions under the alternative.

The use of the F-distribution for deriving p-values for the analysis of variance is based upon the assumption of normality; see, for example, the derivation in Lehmann (1986). Nevertheless, Jagers (1980) shows that the F-ratio is almost exact in many non-normal situations.

The purpose of the present note is to explore the conditions under which a distribution would be sufficiently non-normal that the analysis of variance applied to observations from that distribution would be either inexact or less powerful than a permutation test.

Findings: General Hypotheses

When the form of the distribution is known explicitly, one often can transform the observations to normally-distributed ones and then apply the analysis of variance; see, Lehman (1986) for a list of citations. Consequently, the

Philip Good is a statistical consultant. He authored numerous books that include, *Introduction to Statistics via Resampling Methods and R/S-PLUS* and *Common Errors in Statistics and How to Avoid Them*. Email: pigood@verizon.net. The late Cliff Lunneborg was Professor Emeritus, Statistics & Psychology and author of *Modeling Experimental and Observational Data* and *Data Analysis by Resampling: Concepts and Applications.*

present investigation is limited to the study of observations drawn from contaminated normal distributions, both because such distributions are common in practice and because they cannot be readily transformed.

In R, examples of samples such distributions would include the following:

```
rnorm(n,2*rbinom(n,1,0.3))
ifelse(rbinom(n,1,0.3),rnorm(n,0.5),
  rnorm(n,1.5,1.5))
```

for both of which the analysis of variance was exact in 1000 simulations of an unbalanced 1x3 design with 3, 4, and 5 observations per cell.

Regardless of the underlying distribution, providing the observations are exchangeable under the null hypothesis, one can always make use of the permutation distribution of a test statistic to obtain an exact test. Let X_{ij} denote the jth observation in the ith cell of a one-way design. Eliminating factors from the F-ratio that are invariant under rearrangement of the observations between cells, such as the within sum of squares that forms its denominator, a permutation test based on the F-ratio reduces to a test based on the sum $\sum_i (\sum_j X_{ij})^2$. It was this test that was used in head-to-head comparisons with the one-way analysis of variance.

When a 1x3 design was formed using the following code

```
s1=rnorm(size[1],rbinom(size[1],1,0.3))
s2=ifelse(rbinom(size[2],1,0.3),
  rnorm(size[2],0.5),rnorm(size[2],1.5,1.5))
s3=ifelse(rbinom(size[2],1,0.3),
  rnorm(size[3],1),rnorm(size[3],2,2))
```

the power of the analysis of variance and the permutation test based upon 1000 simulations were comparable for a balanced design with as few as three observations per cell (α=10%, β=22%). But for an unbalanced design with 3, 4, and 5 observations per cell, the permutation test was more powerful at the 10% level with β=30%, compared to 18% for the analysis of variance.

When a 1x4 design was formed using the following code:

```
s0=rnorm(size[1],rbinom(size[1],1,0.5))
s1=rnorm(size[2],rbinom(size[2],1,0.5))
s2=rnorm(size[3],rbinom(size[3],1,0.5))
s3=rnorm(size[4],2 + rbinom(size[4],1,0.5))
```

the power of the analysis of variance and the permutation test were comparable for a balanced design with as few as three observations per cell (α=10%, β=57%). However, for an unbalanced design with 2, 3, 3, and 4 observations per cell, the permutation test was more powerful at the 10% level with β=86%, compared with 65% for the analysis of variance.

If the designs are balanced, the simulations support Jagers (1980) result, that the analysis of variance is both exact and powerful, whether observations are drawn from a contaminated normal distribution, a distorted normal distribution (z=2*z if z>0), a censored normal distribution (z = -0.5 if z< -0.5), or a discrete distribution such as would arise from a survey on a five-point Likert scale. When the design is unbalanced, Jagers' result does not apply, and the permutation test has superior power. The results confirm and extend the findings of Miller (1986).

Findings: Specific Hypotheses

When testing for an ordered dose response, the Pearson's product moment correlation coefficient is usually employed as a test statistic with p-values obtained from a t distribution. Alternatively, the exact permutation procedure due to Pitman (1937) could be employed. In the simulations with contaminated normal distributions, it was found that the parametric procedure for testing for an ordered dose response was both exact (to within the simulation error) and as powerful as the permutation method.

For testing other specific hypotheses, the permutation method may be preferable, simply because no well-tabulated parametric distribution exists. An example would be the alternative that exactly one of the k-populations from which the samples are drawn is different from the others for which an exact test based on the distribution of $\max_k |\bar{X}_{.} - \bar{X}_k|$ is readily obtained by permutation means.

To further explore the possibilities, a copy of the code along with a complete listing of the simulation results is provided at mysite.verizon.net/res7sf1o/AnovPower.txt. (A manuscript assessing the robustness of the two-way analysis of variance is in preparation.)

References

Good, P. (2005). *Permutation, parametric, and bootstrap tests of hypotheses* (3rd Ed.). New York: Springer.

Jagers, P. (1980). Invariance in the linear model: An argument for chi-square and F in nonnormal situations. *Mathematische Operationsforschung und Statistik, 11*, 455-464.

Lehmann, E. L. (1986). *Testing statistical hypotheses*. (2nd ed.). New York: Wiley.

Miller, R. G. (1986). *Beyond ANOVA, basics of applied statistics*. New York: Wiley.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any population. *Journal of the Royal Statistical Society, 4*, 225-232.