International Journal of Pure and Applied Mathematics
Volume 43 No. 3 2008, 309-316

A NOTE ON POWER CALCULATIONS FOR ANOVA DESIGNS

A.J. Watkins[1 §], R.J. Shields[2]
[1]School of Business and Economics
University of Wales Swansea
Singleton Park, Swansea, SA2 8PP, U.K.
e-mail: a.watkins@swansea.ac.uk
[2]School of Environment and Society
University of Wales Swansea
Singleton Park, Swansea, SA2 8PP, U.K.
e-mail: r.j.shields@swansea.ac.uk

Abstract: We present details of statistical power calculations for various experimental designs widely used in comparative studies, with particular emphasis on replacing printed tables and nonograms with readily available mathematical software. We also provide a short discussion on the specification of various experimental scenarios used in the design of a study.

AMS Subject Classification: 62F03, 62K99, 62Q05
Key Words: ANOVA table, experimental design, non-central F-distribution, statistical power

1. Introduction

Ling and Cotter [1] present a summary of statistical power calculations for various experimental designs, with particular reference to the role played by such designs in comparative aquaculture studies. This note continues that discussion, with the twin objectives of presenting further details of the calculations, in which the use of printed tables and nonograms is superseded by the use of readily available mathematical software, and in outlining a range of experimental

Received: September 25, 2007

[§]Correspondence author

features used as part of designing one such study. The approach outlined here is implemented in *Mathematica* (see Wolfram, [4]), and this note is complemented by an electronic notebook file available from the first author.

2. One Way and Nested ANOVA Designs and Analyses

These designs are widely used; for instance, they are employed in comparative aquaculture studies, in which the observed data captures some aspect of interest of fish (such as weight, length, or changes in these variables), and the aim is to compare the effect of, say, diets on this single variable of interest. For completeness, we outline the models and ANOVA tables associated with one way and nested ANOVA designs; see, for instance, Montgomery [2] for further details of and discussion on these designs.

2.1. One Way ANOVA

In the case of a straightforward ANOVA structure, the observed data are assumed to arise from the model

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

for $i = 1, \ldots, a$, and $j = 1, \ldots, n$, where the μ_i are fixed and the ε_{ij} are independently distributed as $N\left(0, \sigma^2\right)$. Thus, μ_i is the population mean of the variable for fish on the i-th diet, and ANOVA tests the null hypothesis that these means are equal; that is, $\mu_i = \mu$ for $i = 1, \ldots, a$. We note the alternative, but equivalent, parameterisation

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

again with $i = 1, \ldots, a$, and $j = 1, \ldots, n$. This version of the model is over-parameterised, so we assume $\sum_{i=1}^{a} \tau_i = 0$, when the null hypothesis above is phrased as $\tau_i = 0$ for $i = 1, \ldots, a$. The basic ANOVA table is outlined in Montgomery [2]; in the usual notation, the total sum of squares is

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} \left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2,$$

which decomposes as

$$SS_T = SS_B + SS_E;$$

that is, the sum of the between groups sum of squares

$$SS_B = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 = n\sum_{i=1}^{a}\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2$$

and the within groups sum of squares

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{i.}\right)^2 .$$

The degrees of freedom associated with these sums of squares complete the ANOVA table, which looks like

Source	Sum of squares	Degrees of freedom	Expected Mean square	F ratio	p-value
Diet	SS_B	$a-1$	$\sigma^2 + \frac{n}{a-1}\sum_{i=1}^{a}\tau_i^2$	$\frac{SS_B/(a-1)}{SS_E/\{a(n-1)\}}$	from tables
Error	SS_E	$a\,(n-1)$	σ^2		
Total	SS_T	$an-1$			

Thus, the test statistic for the null hypothesis is the ratio of the mean squares, which follows the non-central F-distribution with degrees of freedom $(a-1,\, a(n-1))$ and non-centrality parameter

$$\lambda = \frac{n\sum_{i=1}^{a}\tau_i^2}{\sigma^2};$$

see, for example, Pearson and Hartley [3], in which the charts employ the related non-centrality parameter

$$\phi = \sqrt{\frac{\lambda}{a}}.$$

The p-value is computed under the assumption that the null hypothesis is true (so $\lambda = 0$); as per Ling and Cotter [1], we are concerned with the power of the procedure when the null hypothesis is false. A simple example serves to illustrate the ease with which these charts can be superseded by the use of readily available mathematical software. We take $a = 2$, $\mu = 11, \tau_1 = -1, \tau_2 = 1$ (so that $\mu_1 = 10, \mu_2 = 12$), and $\sigma^2 = 1$. We then have $\lambda = 2n$, while the p-value is taken from the F-distribution with degrees of freedom $(1, 2\,(n-1))$. In this case, and working with $n = 5$, we can use the *Mathematica* commands

```
n=5;F0=FRatioDistribution[1,2*(n-1)];critical=Quantile[F0,0.95]
```

to find the 95-th percentile of the $F_{1,2(n-1)}$ distribution; this is $5.3176\cdots$. The power of the ANOVA design is then the probability that an observation from

the non-central F distribution with degrees of freedom $(1, 2\,(n-1))$ and non-centrality parameter λ exceeds this critical value; this probability is found from

```
F1=NoncentralFRatioDistribution[1,2*(n-1),2*n];
                                    power=1-CDF[F1,critical]
```

and is $0.79054\cdots$. Clearly, it is straightforward to vary n, to plot power against n, and to investigate further the power of experiments with different a, μ_i (or τ_i) and σ^2. We may further remark that *Mathematica* uses the non-centrality parameter λ, rather than ϕ.

2.2. Nested ANOVA

We next consider the case of a model with fixed main effects and nested random effects, so that the observed data arise from the model

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}$$

for $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n$. The normality assumptions are

$$\beta_{j(i)} \tilde{} N\left(0, \sigma_\beta^2\right) \text{ and } \varepsilon_{(ij)k} \tilde{} N\left(0, \sigma^2\right)$$

with independence everywhere. Again, this model is overparameterised, so we assume $\sum_{i=1}^{a} \tau_i = 0$; as before, we wish to test the hypothesis that $\tau_i = 0$ for $i = 1, \ldots, a$. In this model, there are a levels of the main effect (which, in aquaculture studies, would usually be diet), with $b\,(> 1)$ levels of a second factor (in aquaculture studies, to continue the illustration, this would be tank, and this factor would aspects such as the thickness of individual tanks, and their location and exposure to light, noise and so on) nested under each level of the main factor, and n individual items for each combination of the two factors (to complete the example, this would be the number of fish per tank in an aquaculture study). As Montgomery [2, p. 507] notes, this is a *balanced nested design*, since we assume an equal number of tanks per diet, and equal number of fish in each tank. Further, as the tanks are different from diet to diet, there can be no interaction between tank and diet. In this case, the total sum of squares is

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(Y_{ijk} - \bar{Y}_{...}\right)^2,$$

which decomposes as

$$SS_T = SS_A + SS_{B(A)} + SS_E,$$

in which

$$SS_A = nb\sum_{i=1}^{a}\left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2, \quad SS_{B(A)} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{ij.} - \bar{Y}_{i..}\right)^2$$

and

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}_{ij.}\right)^2.$$

This decomposition, together with the corresponding degrees of freedom, lead to the ANOVA table

Source	Sum of squares	Degrees of freedom	Expected Mean square	F ratio	p-value
Diet	SS_A	$a-1$	$\sigma^2 + n\sigma_\beta^2$ $+\frac{nb}{a-1}\sum_{i=1}^{a}\tau_i^2$	$\frac{SS_A/(a-1)}{SS_{B(A)}/\{a(b-1)\}}$	from tables
Tanks (within diet)	$SS_{B(A)}$	$a\,(b-1)$	$\sigma^2 + n\sigma_\beta^2$		
Error	SS_E	$ab\,(n-1)$	σ^2		
Total	SS_T	$abn-1$			

Thus, the hypothesis of equal means (and power calculations) for this nested design uses a test statistic which follows the non-central F distribution with degrees of freedom $(a-1, a\,(b-1))$ and non-centrality parameter

$$\lambda = \frac{nb}{\left(n\sigma_\beta^2 + \sigma^2\right)}\sum_{i=1}^{a}\tau_i^2;$$

again, we are concerned with the power of the procedure when the null hypothesis is false. Calculation of the power proceeds along the lines indicated above, although there are now more quantities to be specified. For example, using the values in Ling and Cotter [1], we have, with $a = 2, b = 3, n = 100$, $\tau_1 = -0.6, \tau_2 = 0.6$ (to give a minimum detectable difference of 1.2), and $\sigma_\beta^2 = 0.36, \sigma^2 = 17.6$; thus, we have

$$\lambda = \frac{300 \times 0.72}{(36 + 17.6)} = \frac{216}{53.6} = 4.02985\cdots.$$

In *Mathematica*, the critical value for the test statistic is obtained using

```
a=2;b=3;F0=FRatioDistribution[a-1,a*(b-1)];
                                critical=Quantile[F0,0.95]
```

and here is $7.7086\cdots$, while the power of the test is found by first calculating

```
lambda=n*b*0.72/(n*s2b+s2)
```

and then using this in

```
F1=NoncentralFRatioDistribution[a-1,a*(b-1),lambda];
                                        power=1-CDF[F1,critical]
```

and this gives the power as $0.3379\cdots$. Using graphical capabilities within *Mathematica*, we can now easily reproduce Figure 3 in Ling and Cotter [1]; as they observe, the specified minimum detectable difference here is relatively small to the variation within individual fish, so that large samples are required to achieve reasonable power levels. They also note that the power calculations are highly sensitive to changes in one or more of the parameters of the model.

3. Experimental Design

We briefly outline some ideas on experimental design. In many applications (and in discussions such as Ling and Cotter [1]), the focus of power calculations is n, the number of fish per tank, with other quantities being partially or completely specified by experimental conditions. For instance, the experimenter may wish to investigate a specific number of diets (thus specifying a), with physical constraints then determining b, while estimates of variances σ^2, σ_β^2 may also be available from other or pilot studies. Thus, it remains to consider the range in the μ_i (equivalently, the τ_i); in some cases, it may be sensible to consider the range $[\mu_l, \mu_h]$ for the μ_i, and then to specify how the μ_i are spread out between the lowest value μ_l and the highest value μ_h. At least three possibilities readily suggest themselves:

— an extreme scenario, in which $a-1$ diets have mean μ_l, but one diet has mean μ_h. In this case,

$$\sum_{i=1}^{a} \tau_i^2 = \left[\frac{a-1}{a}\right] (\mu_h - \mu_l)^2 .$$

— a uniform scenario, in which the mean of the i-th diet is

$$\mu_i = \frac{(a-i)\,\mu_l + (i-1)\,\mu_h}{(a-1)};$$

here, we find

$$\sum_{i=1}^{a} \tau_i^2 = \left[\frac{a\,(a+1)}{12\,(a-1)}\right] (\mu_h - \mu_l)^2 .$$

— an intermediate scenario, with diets divided into three sub-groups, in which a_l of the diets have mean μ_l, a_h of the diets have mean μ_h, and the

remaining $a - a_l - a_h$ diets have mean

$$\frac{\mu_l + \mu_h}{2};$$

here, we have

$$\sum_{i=1}^{a} \tau_i^2 = \left[\frac{a\left(a_l + a_h\right) - \left(a_l - a_h\right)^2}{4a} \right] \left(\mu_h - \mu_l\right)^2 .$$

Algebraic inspection, or direct numerical evaluation, confirm that, even with a, μ_l, μ_h fixed, the three expressions for $\sum_{i=1}^{a} \tau_i^2$ can vary considerably, and since these expressions directly influence the non-centrality parameter λ, further affect power calculations. For instance, with $a = 5, b = 3$, $\mu_l = 10, \mu_h = 12$, and with $\sigma_\beta^2 = \frac{1}{4}, \sigma^2 = 9$, we obtain the following power calculations for $n = 10\,(10)\,100$:

n	Extreme $\sum_{i=1}^{a} \tau_i^2 = \frac{16}{5}$	Uniform $\sum_{i=1}^{a} \tau_i^2 = \frac{5}{2}$	Intermediate $(a_l = a_h = 2)$ $\sum_{i=1}^{a} \tau_i^2 = 4$
10	0.42281	0.33578	0.51673
20	0.64587	0.52869	0.75298
30	0.76131	0.64309	0.85584
40	0.82555	0.71413	0.90584
50	0.86441	0.76097	0.93304
60	0.88964	0.79356	0.94928
70	0.90697	0.81725	0.95970
80	0.91943	0.83510	0.96679
90	0.92872	0.84895	0.97183
100	0.93587	0.85996	0.97555

This shows that the power calculations are highly sensitive to changes in the experimental framework, including those specified by the experimenter.

4. Discussion and Conclusions

We have shown the use of printed tables and nonograms in statistical power calculations can be replaced by readily available mathematical software; we have used *Mathematica*, but similar functionality may be found in other packages. Furthermore, the graphical capabilities of such software packages facilitate systematic consideration of any feature (such as n, the number of fish per tank) of a proposed experimental study. Since the underlying power calculations

are generally also dependent on remaining aspects of the design (such as a, the number of diets, and $\sum_{i=1}^{a} \tau_i^2$, a measure of variation of average response within the a diets) specified by the experimenter, we see that careful specification of these values is required if the experimenter is to gain a genuine insight into the statistical power of a proposed experiment.

References

[1] E.N. Ling, D. Cotter, Statistical power in comparative aquaculture studies, *Aquaculture*, **224** (2003), 159-168.

[2] D.C. Montgomery, *Design and Analysis of Experiments*, Fourth Edition, John Wiley and Sons, New York (1997).

[3] E.S. Pearson, H.O. Hartley, Charts of the power function for analysis of variance tests, derived from the non-central F-distribution, *Biometrika*, **38** (1951), 112-130.

[4] S. Wolfram, *The Mathematica Book*, Fourth Edition, Wolfram Media, Cambridge University Press, Cambridge (1999).