Abstract. Power analysis calculations for fixed-effects ANOVA designs can now be done in SAS®. Step 1 uses PROC GLM to calculate noncentrality values. Step 2 involves entering those values into a separate SAS program that converts noncentrality values into tables of power probabilities for various scenarios (O'Brien, 1986a). With just a modest effort, the calculations done in Step 2 could be incorporated into PROC GLM itself, making GLM an effective tool for statistical planning as well as for statistical analysis. An example illustrates how easy it should be to perform comprehensive power analyses using an enhanced PROC GLM. A proposed syntax for a POWER statement to set parameters and options is described, and the output that could be produced by it is shown. Both retrospective and prospective power analyses are demonstrated. How the POWER statement would operate in an interactive GLM session is also discussed.

1. Introductory Remarks

This paper illustrates how power analysis capabilities could be incorporated into PROC GLM. Space limitations prevent us from detailing how the calculations are done; for this, see O'Brien (1987, 1986a). We do review those formulas that are necessary to understand what would appear in GLM's output. In addition, we treat some issues not covered in O'Brien (1987, 1986a), in particular, confidence intervals for noncentrality and power.

It should be stressed that the GLM output shown below is imaginary. The capabilities to produce these results may never exist in GLM. If power analysis is something you think should be added to GLM, let SAS Institute know, make sure it gets on the next SASware ballot, and vote for it. Some of the capabilities that we are proposing for GLM have already been implemented elsewhere, notably in the MANOVA procedure in SPSS®.

2. An Overview of Power Analysis

What is statistical power? In short, the power of a statistical test is the probability that a false null hypothesis, \( H_0 \), will be rejected. This is what most investigators want to do, so they try to design their studies to be as powerful as possible. Our focus here is on the normal-theory, fixed-effects general linear model. In this case,

\[
(1) \quad \text{Power} = \text{Prob}[F(\text{sample}) \geq F_{\text{crit}}] 
\]

where \( F(\text{sample}) \) is the usual sample \( F \) statistic:

\[
(2) \quad F(\text{sample}) = \left( \frac{\text{SSH}(\text{sample})}{\text{df}_H} \right) / \sigma^2. 
\]

\( \text{SSH}(\text{sample}) \) is the hypothesis sum of squares having \( \text{df}_H \) degrees of freedom. \( \sigma^2 \) is the mean square error (MSE) with \( \text{df}_E \) degrees of freedom. Calculation of \( \text{SSH}(\text{sample}) \) and \( \sigma^2 \) for various hypotheses is covered in many texts, e.g., Freund, Littell and Spector (1986). \( F_{\text{crit}} \) is the critical value of the \( F \) statistic. That is, it is the value of a random variable, \( F \), having a central \( F \) distribution with \( \text{df}_H \) and \( \text{df}_E \) degrees of freedom such that \( \text{Prob}[F \geq F_{\text{crit}}] = \alpha \) where \( \alpha \) is the chosen significance level for the test.

Here we treat two kinds of power analyses: prospective and retrospective. Prospective power analysis is used in the planning phase of a study. Typically, it addresses the question: "How large must my sample size be?" In practice, however, the question is usually: "The largest sample I can afford is \( N \). Is that large enough to detect an effect of the size I anticipate? Can I get by with a smaller sample?"

Retrospective power analysis is used during and after the data analysis phase of a study. It addresses the question: "How powerful was the test I just conducted?" The retrospective computation of power is one way to indicate the magnitude of an effect and to show whether an effect has any practical significance (as opposed to statistical significance). In a preliminary analysis or a pilot study, one might go on to ask: "What sample size would I need to achieve acceptable power?"

The first step in the calculation of power is obtaining the noncentrality parameter, \( \lambda \), for the distribution of \( F(\text{sample}) \) which has a noncentral \( F \) distribution when \( H_0 \) is false. For the fixed-effects linear model,

\[
(3) \quad \lambda = \frac{\text{SSH}(\text{population})}{\sigma^2} 
\]
where SSH(population) is calculated just like SSH(sample) except that (conjectured) population parameters are used instead of sample parameter estimates. Defining $\delta^2 = SSH(population)/N$,

$$\lambda = N \frac{\delta^2}{\sigma^2}.$$  

This expression explicitly shows the three determinants of the noncentrality parameter and, consequently, of power. $N$ is the sample size: bigger $N$ -- bigger $\lambda$ -- higher power. $\sigma$ is the error (residual, within-cell) standard deviation: smaller $\sigma$ -- bigger $\lambda$ -- higher power. $\delta^2$ can be thought of as the square of the raw effect size: bigger $\delta^2$ -- bigger $\lambda$ -- higher power. Its value depends on the population parameters. For example, in a balanced one-way ANOVA with $J$ groups, $\delta^2 = \frac{\sum l_{ij}}{J}$ where $l = \sum \mu_{ij} / J$. The "effect size" (ES) of Cohen (1977) is a standardized effect size, namely $ES = \delta / \sigma$. From equation (4), it is clear that $\lambda$ is directly proportional to $N$ and inversely proportional to $\sigma^2$, making it easy to calculate changes in $\lambda$ when $N$ or $\sigma^2$ is changed. Writing $\lambda$ as a function of $N$ and $\sigma$,

$$\lambda(aN, b\sigma) = a \lambda(N, \sigma) / b^2.$$  

Finally, power is easily calculated using functions already available in SAS.

$$F_{crit} = \text{FINV}(1-\alpha, df_H, df_E);$$
$$\text{Power} = 1 - \text{PROBF}(F_{crit}, df_H, df_E, \lambda);$$

A retrospective power analysis uses data to estimate the noncentrality parameter and the resulting power. The calculations consist mainly of substituting sample values for population values (e.g., cell means for population means, the mean square error for $\sigma^2$) in the formulas given above. Specifically,

$$\hat{\lambda} = \frac{SSH(sample)}{MSE} = df_H F(sample).$$

$\hat{\lambda}$ is the estimated noncentrality parameter. This estimate is positively biased as is seen from the fact that the expected value of a noncentral $F$ random variable is

$$E[F(sample)] = [df_E/(df_E - 2)] \left[ 1 + (\lambda/df_E) \right]$$

so that

$$E[\hat{\lambda}] = [df_E/(df_E - 2)] [df_H + \lambda] = \lambda + df_H.$$

An unbiased estimate of $\lambda$ is easily found, but this estimate may take on negative values. In practice, one would set negative estimates to zero (even though this reintroduces some bias into the estimate), resulting in an adjusted noncentrality estimate:

$$\lambda_{adj} = \text{Max}(0, [\lambda (df_E - 2)/df_E] - df_H)$$

Dwass (1955) showed how to construct a confidence interval for $\delta$, the raw effect size (see also Miller, 1981, p. 52-53). An approximate confidence interval for $\lambda$ is obtained from using $\lambda = N \frac{\delta^2}{\sigma^2}$ with the sample MSE substituted for $\sigma^2$. The resulting formula is

$$\text{Lower limit} = df_H \text{Max}(0, (\sqrt{F(sample)} - \sqrt{F_{crit}})^2)$$
$$\text{Upper limit} = df_H (\sqrt{F(sample)} + \sqrt{F_{crit}})^2$$

Corresponding estimates and confidence limits for power are obtained from equation (6). Likewise, point and interval estimates for Cohen's effect size follow from $\text{ES} = (\lambda/N)^{1/2}$.

These confidence intervals are generally quite wide as can be seen from the results below. This is because they are based on the conservative Scheffe projection method for simultaneous inference. Only in the case of a one-degree-of-freedom test is the actual coverage of the interval nearly the same as the desired confidence level. Venables (1975) gives another method for constructing a confidence interval for $\lambda$ which gives shorter intervals in many cases. However, Venables' method has the undesirable characteristic that the $1-\alpha$ confidence interval may include the value $\lambda = 0$ when the null hypothesis has been rejected at the $\alpha$ level of significance.

3. Prospective Power Analysis: An Example

Noah D. Kay, DDS, is studying two treatments that he thinks will reduce the formation of dental calculus (plaque). He plans to use a one-way ANOVA with two treatment groups and a control group, and the control group is to have twice as many subjects as each treatment group. He expects the second treatment to be more effective than the first. Specifically, in terms of the units by which he measures plaque formation, the design is:

<table>
<thead>
<tr>
<th>Group</th>
<th>Control</th>
<th>Treatment 1</th>
<th>Treatment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size:</td>
<td>2n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Conjectured mean:</td>
<td>32</td>
<td>26</td>
<td>24</td>
</tr>
</tbody>
</table>

The common within-cell standard deviation is conjectured to be between 6.0 and 9.0.
The proposed SAS statements for the analysis are:

```
TITLE 'Prospective Power Analysis';
DATA plaque;
  INPUT grp $ grp_n grp_mean;
  CARDS;
  CO 2 32
  T1 1 26
  T2 1 24
PROC GLM; CLASS grp; FREQ grp_n;
  POWER N = 40 60 S = 9.0 6.0;
  MODEL grp_mean=grp/SS3;
  CONTRAST 'Ctrl-Trt· grp 2 -1
                  -1;
```

The only new statement is the POWER statement. In that statement, "N = 40 60" specifies that results are to be calculated for a total sample size of 40 (20 controls + 10 in each treatment group) and 60 (30+15+15). "S = 9.0 6.0" specifies two different error standard deviations to be used to calculate power. Note that in a prospective analysis, the error sum of squares reported by GLM will ordinarily be zero; thus, the error variance to be used in the power analysis must be specified by the investigator in the POWER statement.

The proposed GLM output for the analysis follows. Abbreviations used in the output are: 

- N = total sample size; DF = denominator degrees of freedom; Sigma = the error standard deviation (σ); NC = the noncentrality parameter (λ); Alpha = the significance level of the test (α); Power = the power of the test. Unless told otherwise (see the "Proposed Syntax" section below), GLM would produce a power analysis for the overall model F test, for each effect in the model, and for each contrast. In a one-way ANOVA like this one, of course, the test of the GRP effect is the same as the model F test, so the overall ANOVA table and the power analysis for the model F test are not shown.

```
Prospective Power Analysis

<table>
<thead>
<tr>
<th>Type III Mean</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>DF</td>
<td>SS</td>
<td>Square</td>
<td>F Value</td>
</tr>
<tr>
<td>GRP</td>
<td>2</td>
<td>51.0000</td>
<td>25.5000</td>
<td>9999.99</td>
</tr>
</tbody>
</table>
```

4. Retrospective Power Analysis: An Example

The data are from the dental calculus study reported in Finn (1974). Two treatment groups and one control group were measured during two consecutive years. (For simplicity, it is assumed that the same two treatments were used each year though, in fact, this was not the case. Also two control groups were used the first year; these have been combined. The dependent variable is total calculus formation, the sum of the six measurements reported by Finn. (For comparability with Finn's analysis, untransformed measurements are used though, in fact, the dependent variable is not normally distributed.) The analysis, then, is an unbalanced two-way ANOVA. The within-cells standard deviation ("Root MSE") was 8.374. The cell means and cell counts (in parentheses) were:

```
<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment 1</th>
<th>Treatment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>13.82 (17)</td>
<td>5.57 (7)</td>
<td>5.60 (5)</td>
</tr>
<tr>
<td>Year 2</td>
<td>10.00 (28)</td>
<td>6.75 (24)</td>
<td>3.58 (26)</td>
</tr>
</tbody>
</table>
```

The SAS statements to run the analysis are:
TITLE 'Finn (1974) Dental Calculus Data';
DATA teeth; INFILE 'finn.datl';
INPUT group $ year dentcalc;
PROC GLM; CLASS group year;
POWER ALPHA = .05 .01
   NFACTOR = 1.0 2.0 CL = .95;
MODEL dentcalc=group year group*year/SS3;
CONTRAST 'Ctrl-Trt' group 2 -1 -1;

In this POWER statement, "ALPHA = .05 .01" specifies that two different Type I error rates be used to estimate power; "NFACTOR = 1.0 2.0" specifies that power be calculated for the sample size actually used and for a sample twice as large; and "CL = .95" specifies that 95% confidence intervals be calculated for noncentrality parameters and for powers. Selected portions of our proposed GLM output are shown below.

Finn (1974) Dental Calculus Data
General Linear Models Procedure

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>Pr&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP</td>
<td>2</td>
<td>858.049</td>
<td>429.0243</td>
<td>6.12</td>
</tr>
<tr>
<td>YEAR</td>
<td>1</td>
<td>42.106</td>
<td>42.1055</td>
<td>0.60</td>
</tr>
<tr>
<td>GROUP*YEAR</td>
<td>2</td>
<td>89.698</td>
<td>44.8492</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Power Analysis for GROUP*YEAR using Type III SS

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>SS</th>
<th>Square</th>
<th>F Value</th>
<th>Pr&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctrl-Trt</td>
<td>1</td>
<td>853.481</td>
<td>853.481</td>
<td>12.17</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

Power Analysis for Ctrl-Trt Contrast

<table>
<thead>
<tr>
<th>N</th>
<th>DF</th>
<th>Sigma</th>
<th>NC</th>
<th>Adj NC</th>
<th>NC 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>107</td>
<td>101</td>
<td>8.374</td>
<td>12.237</td>
<td>9.994</td>
<td>1.027-35.792</td>
</tr>
<tr>
<td>214</td>
<td>208</td>
<td>8.374</td>
<td>24.473</td>
<td>19.969</td>
<td>2.055-71.583</td>
</tr>
</tbody>
</table>

Power Analysis for GROUP using Type III SS

<table>
<thead>
<tr>
<th>N</th>
<th>DF</th>
<th>Sigma</th>
<th>Adj Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>107</td>
<td>101</td>
<td>8.374</td>
<td>0.05</td>
</tr>
<tr>
<td>107</td>
<td>101</td>
<td>8.374</td>
<td>0.01</td>
</tr>
<tr>
<td>214</td>
<td>208</td>
<td>8.374</td>
<td>0.05</td>
</tr>
<tr>
<td>214</td>
<td>208</td>
<td>8.374</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Power Analysis for YEAR using Type III SS

<Output is similar to that for GROUP>

5. Proposed Syntax for the POWER Statement

POWER H= effects and contrasts
OFF
ALPHA = a1 a2 ...
N = n1 n2 ...
NFACTOr | INF = a1 a2 ...
SIGMA | S = a1 a2 ...
SFACTOr | SF = b1 b2 ...
CILEVEL | CL = v1 v2 ...
ORDER=N SIGMA ALPHA
OUT=SASdataset ;

Any number of POWER statements may appear. Specifications in the POWER statement remain in effect until changed by a subsequent POWER statement. The terms below may be specified on the POWER statement.

H = effects and contrasts
specifies model effects and contrasts for which power analysis is desired. Contrasts are referred to by the label given in the CONTRAST statement, and the label must be enclosed in quotes. If effects are listed, the POWER statement must appear after the MODEL statement. If contrasts are listed, the POWER statement
must appear after the CONTRAST statements in which
the labels appear. Several abbreviations may be used.
_ALL_ requests power analyses for all effects and
contrasts and the overall model F test. This is the
default when a POWER statement is used but H = is not
specified. _OVERALL_ specifies the model F test.
_EFFECTS_ specifies all effects in the MODEL
statement. _CONTRASTS_ specifies all defined
contrasts. _NONE_ specifies that no power analysis is
to be done. This could be used to turn off power
analysis during an interactive GLM session. This is the
default if no POWER statement is used.

OFF is equivalent to specifying H = _NONE_.

**ALPHA** = p1 p2 ...
specifies a list of significance levels to be used in power
calculations. Each specified value must be between
zero and one.

**N** = n1 n2 ...
specifies a list of sample sizes to be used in power
calculations. This is the overall sample size, the total
number of individuals in the sample. **N** and **NF** =
cannot both be in effect at the same time; the one most
recently specified will be in effect.

**NFACTOR** = a1 a2 ...
specifies a list of sample size multipliers. That is,
power analyses will be calculated for samples of size
n1\*N, n2\*N, ... where N is the number of observations
used in the analysis (see Equation 5). Multipliers may
de be decimal fractions. **NF** = and **N** = cannot both be in effect at the same time; the one most
recently specified will be in effect.

**SIGMA** = s1 s2 ...
specifies a list of error standard deviations (square root
of mean square error) to be used in power calculations.
**S** = and **SF** = cannot both be in effect at the same time; the one most
recently specified will be in effect.

**SFACTOR** = b1 b2 ...
specifies a list of error standard deviation multipliers.
That is, power analyses will be calculated for b1*s, b2*s, ...
where s is the Root MSE from the analysis of
variance (see Equation 5). Multipliers may be decimal
fractions. **SF** = and **S** = cannot both be in effect at the same
time; the one most recently specified will be in effect.

**CLEVEL** = CL = v1 v2 ...
specifies a list of confidence levels for interval
estimates. Each "v" should be between zero and one.
The default is **CL** = .95 for 95% confidence intervals.
Confidence intervals are calculated whenever a
retrospective analysis is done, that is, whenever the
error sum of squares in the analysis of variance is non-
zero.

**ORDER** = N SIGMA ALPHA
specifies the order of appearance of these three columns
in the power analysis results. By default, **N** and **DF**
appear first and change most slowly. **SIGMA** appears
next, and **ALPHA** appears last, changing most rapidly.

**OUT** = SASdataset
specifies the name of a SAS data set to contain results
from the power analyses.

6. Interactive Power Analysis: An Example

To show how the POWER statement might be used
in an interactive GLM session, a variation of the prospective
analysis of section 3 is shown below. The output would
be nearly the same as above and so is omitted.

**TITLE** 'Prospective Power Analysis';
**DATA** plaque;
**INPUT** grp $ grp_n grp_mean;
**CARDS**;
c0 2 32
t1 1 26
t2 1 24
**RUN**;
**PRGC** GLM; **CLASS** grp; **FREQ** grp_n;
**MODEL** grp_mean=grp/SS3;
**POWER** H='grp' **N**=40 60 8=6;
**RUN**;

<ANOVA and power analysis for GRP appear here>

**CONTRAST** 'Ctrl vs Trtmnt' grp 2 -1 -1;
**POWER** H='Ctrl vs Trtmnt';
**RUN**;

<Contrast test result and power analysis appear here>

**POWER** H = 'Ctrl vs Trtmnt' S=9.0;
**RUN**;

<Power analyses for GRP and the contrast are repeated>
<with the new error standard deviation>
POWER N=100;
RUN;

<Power analyses for GRP and the contrast are repeated>
<with the new sample size>
QUIT;

7. Additional Issues Concerning Power Analysis

This paper has focused on power analysis in the univariate fixed-effects model. PROC GLM, of course, does much more than that. GLM does multivariate analysis, random and mixed models can be analyzed (with the help of RANDOM and TEST statements), and repeated measures designs can be analyzed by either the mixed-model or multivariate method. Muller and Peterson (1984) demonstrate a simple method for multivariate power analysis. Their method would also work with repeated measures designs (multivariate method). For random effects in the model, the non-null distribution of F(sample) is not a noncentral F; rather, it is $\theta$ times that of a central F variate, where $\theta > 1$ is a function of the variance components (Scheffé, 1959).

Power analysis could also be incorporated into other SAS procedures. An obvious candidate is PROC REG since it also handles normal theory linear models. Another is PROC CATMOD since the methods of this paper transfer to categorical data models with little change (O'Brien, 1986b).

References


---

1 S. Paul Wright, Statistics Department and UT Medical Center, 328 Stokely Management Center, Knoxville, TN, 37996-0532. BITNET: PA81668@UTKVM1.
2 Ralph G. O'Brien, Statistics Department and UT Computing Center, 338 Stokely Management Center, Knoxville, TN, 37996-0532. BITNET: PA87458@UTKVM1.

This work was supported by a Faculty Research Fellowship from the University of Tennessee College of Business Administration.

SAS is a registered trademark of SAS Institute, Inc., Cary, NC, USA.

SPSS® is a trademark of SPSS Inc., Chicago, IL, USA.