

Advanced Sample Size and Power Estimation

In this section, power and sample size issues are addressed for more sophisticated experimental designs and analyses. The examples below will deal with situations involving analysis of variance, although the similar examples can typically be adapted to cases for linear and multiple regressions.

Evaluating Existing Data

In the first example, the randomized complete block (RCB) experimental design is considered. Specifically, the case for a fixed linear model with normally, identically, and independently distributed data (Niid) is outlined. Such models are often estimated in SAS with the procedure PROC GLM.

The data used in this example will be taken from a previous workshop on GLM and ANOVA (http://webpages.uidaho.edu/cals-statprog/sas/workshops/glm/handout1_glm.pdf).

The data describes a variety (2 levels) and fertilizer (3 levels) study with 4 blocks measuring yield. As a reference, the results of analysis of variance using PROC GLM on these data are provided below:

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|------------------------|----|----------------|-------------|---------|--------|
| Model | 8 | 35.79769477 | 4.47471185 | 7.36 | 0.0005 |
| Error | 15 | 9.11624767 | 0.60774984 | | |
| Corrected Total | 23 | 44.91394244 | | | |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|-----------------|----|-------------|-------------|---------|--------|
| BLOCK | 3 | 11.64506598 | 3.88168866 | 6.39 | 0.0053 |
| VAR | 1 | 2.48868029 | 2.48868029 | 4.09 | 0.0612 |
| FERT | 2 | 20.96013195 | 10.48006597 | 17.24 | 0.0001 |
| VAR*FERT | 2 | 0.70381655 | 0.35190828 | 0.58 | 0.5725 |

| VAR | FERT | YIELD LSMEAN |
|-------------|----------|--------------|
| VAR1 | 1 | 1.85871031 |
| VAR1 | 2 | 3.12521718 |
| VAR1 | 3 | 3.92756253 |
| VAR2 | 1 | 0.86258303 |
| VAR2 | 2 | 2.94528678 |
| VAR2 | 3 | 3.17151761 |

We would like to estimate the statistical power for this analysis. Because of the factorial treatment design, however, there will be multiple hypotheses to compute power for: VAR, FERT, and the interaction VAR*FERT. Contrasts can also be evaluated to compare specific levels of the significant effect, FERT. The SAS procedure for these power calculations is PROC GLMPOWER. Prior to running power assessments, a data set with the analysis results must be created:

Example 4.

```
data yield_means;
  input var$ fert yield;
  cards;
  VAR1 1 1.8587103
  VAR1 2 3.1252172
  VAR1 3 3.9275625
  VAR2 1 0.8625830
  VAR2 2 2.9452868
  VAR2 3 3.1715176
  ;
run;
```

In this data step, the estimated means from the previous ANOVA analysis are entered. The GLMPOWER procedure is then called:

```
proc glmpower data=yield_means;
  class var fert;
  model yield = var fert var*fert;
  contrast 'Fert 1 vs 3' fert 1 0 -1;
  contrast 'Fert 2 vs 3' fert 0 1 -1;

  power
  stddev = 0.779583
  ntotal = 24
  power = .;
  plot x=n min=24 max=275;
  title1 'GLMPOWER';
run;
```

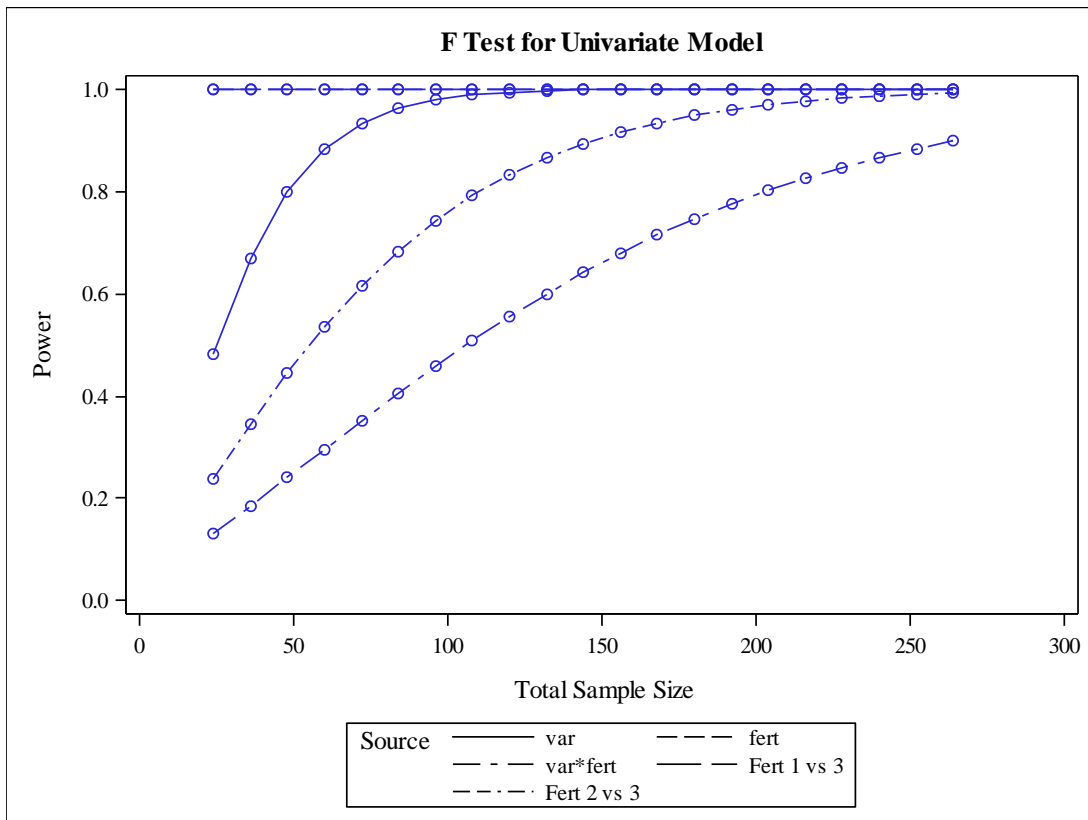
The Class, Model and Contrast statements are identical to those that would be used in PROC GLM. The **power** statement specifies necessary information for power calculations. For example, the **stddev** is given as the root MSE or $\sqrt{0.60774984} = 0.779583$. There were 6 treatment combinations at 4 blocks each giving a total sample size of **ntotal** = 24. The **power** is set to missing in order to solve for that component. The probability of Type I error rate, α , is assumed to be 0.05 and the precision, δ , is determined from the means in the data set. The plot statement produces a power plot of all the specified hypotheses. The statistical power is

SAS Workshop
Power and Sample Size
Hand Out 3

displayed on the Y axis and sample size, n, on the X axis. The X axis has been specified to vary from n = 24 to n = 275.

The output is shown below:

| Computed Power | | | | |
|----------------|----------|-------------|---------|-------|
| Index | Type | Source | Test DF | Power |
| 1 | Effect | var | 1 | 0.482 |
| 2 | Effect | fert | 2 | 0.999 |
| 3 | Effect | var*fert | 2 | 0.131 |
| 4 | Contrast | Fert 1 vs 3 | 1 | >.999 |
| 5 | Contrast | Fert 2 vs 3 | 1 | 0.239 |



The table indicates that there are limited powers for the VAR and VAR*FERT interaction hypotheses (0.482 and 0.131, respectively). The FERT effect, however,

has very high power (0.999). Comparison of FERT means shows that there is high power for FERT 1 vs FERT 3 (0.999) and limited power for FERT 2 vs FERT 3 (0.239).

Examining these hypotheses across sample sizes in the corresponding plot shows flat lines for the overall FERT effect and the contrast of FERT 1 vs FERT 3 (top of figure). These hypotheses have very high power (ability to detect true differences) at all sample sizes. The VAR effect (solid line) has low power at $n = 24$ (0.482) and does not obtain reasonable power until $n \geq 50$, that is, at least 8 blocks. The last two hypotheses, the VAR*FERT interaction and the FERT 2 vs FERT 3 contrast, do not have good power with less than $n = 150$ or 26 blocks. Hence, we can conclude that this experiment only had good power for large differences in the FERT “effect” and that very large sample sizes would have been required in order to detect other effects.

Evaluating Published Data

Researchers are often called upon to evaluate published results, either as part of their own work or as part of a journal review or proposal process. In the following example, the data are taken from a controversial study on the potential histological effects due to genetic modification of potatoes. The paper, *Effect of diets containing genetically modified potatoes expressing Galanthus nivalis lectin on rat small intestine*, by Ewan & Pusztai, (The Lancet, Vol. 354, October 16, 1999), tested three diets (Line effect) consisting of the parent potato line (Parent), the modified parent line (GM: gene for expressing Lectin, a poison for insect control added), and the parent line with the lectin compound added directly (Parent+). Each diet was also assessed with the potatoes served in either raw or boiled form (Cook effect). Hence, the experiment was a 2 x 3 factorial treatment design. Six rats were assigned to each treatment giving a total sample size of $n = 36$. The researchers measured several responses from various parts of the rat intestines. Here, we will concentrate on lymphocyte counts taken in stomach tissue.

As with the previous example, we need to create a data set with the study results. From Table 1, page 1353, we can find the treatment means and their respective standard deviations. These are entered into SAS in the following data step:

```
data Potato;
  input cook$ Part$ line$ mean sd;
  cards;
  Boiled Stomach Parent 294 46
  Raw Stomach Parent 261 32
  Boiled Stomach Parent+ 347 42
  Raw Stomach Parent+ 312 32
  Boiled Stomach GM 339 36
  Raw Stomach GM 323 54
  ;
```

SAS Workshop

Power and Sample Size

Hand Out 3

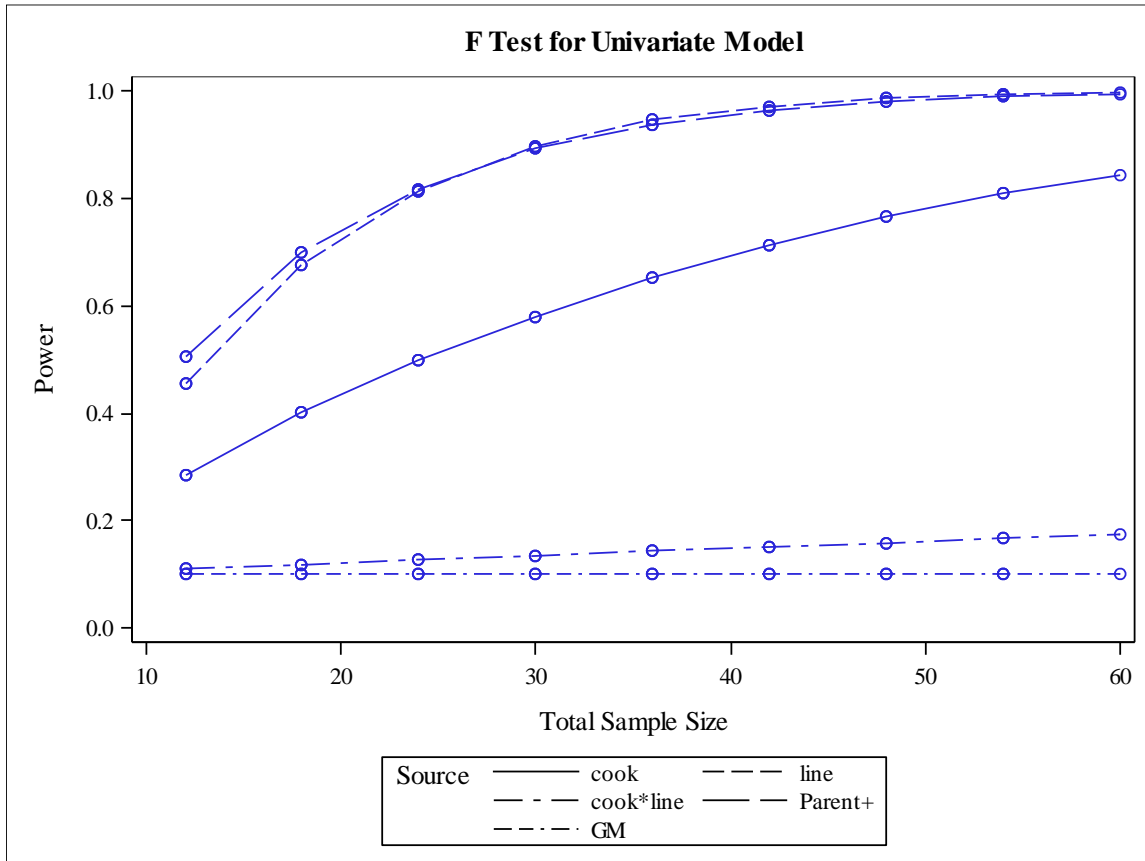
The PROC GLMPOWER procedure can then be called as before:

```
proc glmpower data=Potato;
  class cook line;
  model mean = cook line cook*line;
  contrast 'Parent+' line 1 -1 0;
  contrast 'GM' line 1 0 -1;

  power
    stddev    = 40.3333
    alpha     = 0.10
    ntotal    = 36
    power     = .;
  plot x=n min=12 max=60;
run;
```

The model statement assumes that the experimental design was completely random (CRD). The contrast statements in this case, test both the Parent + lectin treatment and the GM line against the original parent line. Because the paper did not specify a MSE value, we must approximate one from the standard deviations supplied. While we could compute a pooled variance estimate based on sample sizes and standard deviations, it is often sufficient to simply average the standard deviations for sample size/power computations. The small difference in the two methods will have little impact on the final estimates. In this case, the average standard deviation is 40.3333. Additionally, to provide a liberal estimate of power, alpha has been set to 0.10 or 90% confidence coefficient. A plot is requested showing the power over the range of n from n = 12 to n = 60. The results follow:

| Computed Power | | | | |
|----------------|----------|-----------|---------|-------|
| Index | Type | Source | Test DF | Power |
| 1 | Effect | cook | 1 | 0.652 |
| 2 | Effect | line | 2 | 0.945 |
| 3 | Effect | cook*line | 2 | 0.143 |
| 4 | Contrast | Parent+ | 1 | 0.937 |
| 5 | Contrast | GM | 1 | 0.101 |



The table tells us we have moderate power on the cooking method and good power on the potato lines treatments. Within the later effect, the contrast of adding lectin to potatoes (Parent +) with the original parent line (Parent) has good power (0.937), but the same comparison for the GM line is very low (0.101). The plot visually shows this as well as the evidence that there is little chance of improving the power at any sample size.

Mixed Models and Non-normal Responses

In most experimental settings, it is often more appropriate to specify a mixed model or non-normal response in order to fully account for the underlying data structure. For example, while the normality and independence assumptions used in the RCB demonstration shown earlier were appropriate, the effect of blocks would be better represented as a random. In such cases, another means of assessing statistical

SAS Workshop

Power and Sample Size

Hand Out 3

power is required. The RCB example can be reanalyzed in a mixed model format using the following PROC GLIMMIX code:

```
proc glimmix data=yield1;  
  class block var fert;  
  model yield = var fert var*fert/dist=normal;  
  random block;  
run;
```

where, block is now assumed to be a random effect (for more information on generalized linear mixed models, please refer to the GLMMIX workshop handouts: <http://webpages.uidaho.edu/cals-statprog/sas/workshops/glimmix/outline.html>). The resulting output is:

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------------|
| Cov Parm | Estimate | Standard Error |
| BLOCK | 0.5457 | 0.5295 |
| Residual | 0.6077 | 0.2219 |

| Type III Tests of Fixed Effects | | | | |
|---------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| VAR | 1 | 15 | 4.09 | 0.0612 |
| FERT | 2 | 15 | 17.24 | 0.0001 |
| VAR*FERT | 2 | 15 | 0.58 | 0.5725 |

Note that the tests are identical to those given earlier, as are the residual error (root MSE). The block effect now has a variance component estimate of 0.5457. The similarity in the two analyses occurs because of the relatively simple design, normality assumption, and balance nature of the data (no missing values).

Given these results, we can manipulate the GLIMMIX procedure itself to compute values necessary for power assessment. The required codes are given below:

SAS Workshop

Power and Sample Size

Hand Out 3

Statistical Programs
College of Agricultural and Life Sciences
<http://webpages.uidaho.edu/cals-statprog>

```
data rcb_means;
  input var$ fert yield;
  do block = 1 to 4;
    output;
  end;
cards;
VAR1 1 1.8587103
VAR1 2 3.1252172
VAR1 3 3.9275625
VAR2 1 0.8625830
VAR2 2 2.9452868
VAR2 3 3.1715176
run;

proc glimmix data=rcb_means noprofile;
  class block var fert;
  model yield= var fert var*fert/dist=normal;
  random block;
  parms (0.5457) (0.6077)/hold=1,2;
  ods output tests3=results;
run;

data power;
  set results;
  alpha = 0.05;
  nc = numdf*fvalue;
  fcrit = finv(1 - alpha, numdf, dendif, 0);
  power = 1 - probf(fcrit, numdf, dendif, nc);
run;

proc print;
run;
```

(Adapted from Walter W. Stroup. 1999. *Mixed Model Procedures To Assess Power, Precision, and Sample Size In The Design Of Experiments*. Available online at: <http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms/Power/stroup-1999-power.pdf>)

The analysis means are put into a dataset as before with the addition of a block variable. This data step essentially just repeats the means once for each block.

PROC GLIMMIX is called as before with the addition of the **noprofile** option which prevents the estimation of the random effects. Two additional statements are also present. The **parms** statement lists the values for the random effects (Blocks and the residual error, in that order). The **hold=1,2** option requests that parameters 1 and 2 be held constant, again telling SAS to just use these values and not to optimize them for the data. The ODS statement outputs the test statistics to a data set “results” for further computations. Those computations are listed in the data step that follows. This data step utilizes the theoretical relationship between the F statistic, degrees of

SAS Workshop
Power and Sample Size
Hand Out 3

freedom and a value called the non-centrality parameter (NC) to calculate estimated power. For the user, all that needs to be specified here is the value for alpha. The resulting data set produces:

| Obs | Effect | NumDF | DenDF | FValue | ProbF | alpha | nc | fcrit | power |
|-----|----------|-------|-------|--------|--------|-------|---------|---------|---------|
| 1 | var | 1 | 15 | 4.10 | 0.0612 | 0.05 | 4.0952 | 4.54308 | 0.47365 |
| 2 | fert | 2 | 15 | 17.25 | 0.0001 | 0.05 | 34.4909 | 3.68232 | 0.99844 |
| 3 | var*fert | 2 | 15 | 0.58 | 0.5725 | 0.05 | 1.1582 | 3.68232 | 0.12847 |

The power estimates are very similar to those seen previously, but differ slightly due to differences in the estimation procedure and the addition of a random effect for blocks.

The real value of a procedure such as PROC GLIMMIX, however, is its ability to handle non-normal responses. The following example is, again, a RCB design with two treatments (0 and 1) and 8 blocks. The response in this case, however, is binomial, measuring a positive or negative result. Each treatment was recorded 100 times per block. An analysis of variance assuming a generalized linear mixed model for a binomial response and a logit link function was carried out for these data and the following results were obtained (for more information on generalized linear mixed models and analysis of these data, see the GLIMMIX workshop handout at <http://webpages.uidaho.edu/cals-statprog/sas/workshops/glimmix/outline.html>):

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------------|
| Cov Parm | Estimate | Standard Error |
| block | 0.5201 | 0.6037 |
| block*Treatment | 0.8335 | 0.4926 |

| Type III Tests of Fixed Effects | | | | |
|---------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| Treatment | 1 | 7 | 6.75 | 0.0355 |

SAS Workshop

Power and Sample Size

Hand Out 3

Statistical Programs
College of Agricultural and Life Sciences
<http://webpages.uidaho.edu/cals-statprog>

| Treatment | Mean | Standard Error Mean | Lower Mean | Upper Mean |
|-----------|--------|---------------------|------------|------------|
| 0 | 0.9276 | 0.02958 | 0.8189 | 0.9732 |
| 1 | 0.7807 | 0.07245 | 0.5669 | 0.9064 |

As demonstrated in the previous examples, power computations begin by defining a data set with the estimated means:

```
data intro_binomial;
  input Treatment P;
  N = 200;
  Y = N*P;
  do B = 2 to 16;
  do block = 1 to B;
    output;
  end;
end;
cards;
0 .9276
1 .7807
;
```

```
run;
```

In this data step, the means are defined as well as a block variable, B. The do loop structure creates multiple sets of means having from 2 to 16 blocks, which allows us to assess the change in statistical power with changes in the number of blocks.

Once the data set is created, the following codes could be used to compute and plot the estimated power for the treatment effect:

```
proc sort;
  by B;

proc glimmix data=intro_binomial noprofile;
  class block Treatment;
  model Y/N= Treatment;
  random block block*Treatment;
  parms (0.5201) (0.8335)/hold=1,2;
  ods output tests3=results;
  by B;

run;

data power;
  set results;
  alpha = 0.05;
  nc = numdf*fvalue;
  fcrit = finv(1 - alpha, numdf, dendif, 0);
  power = 1 - probf(fcrit, numdf, dendif, nc);

run;
```

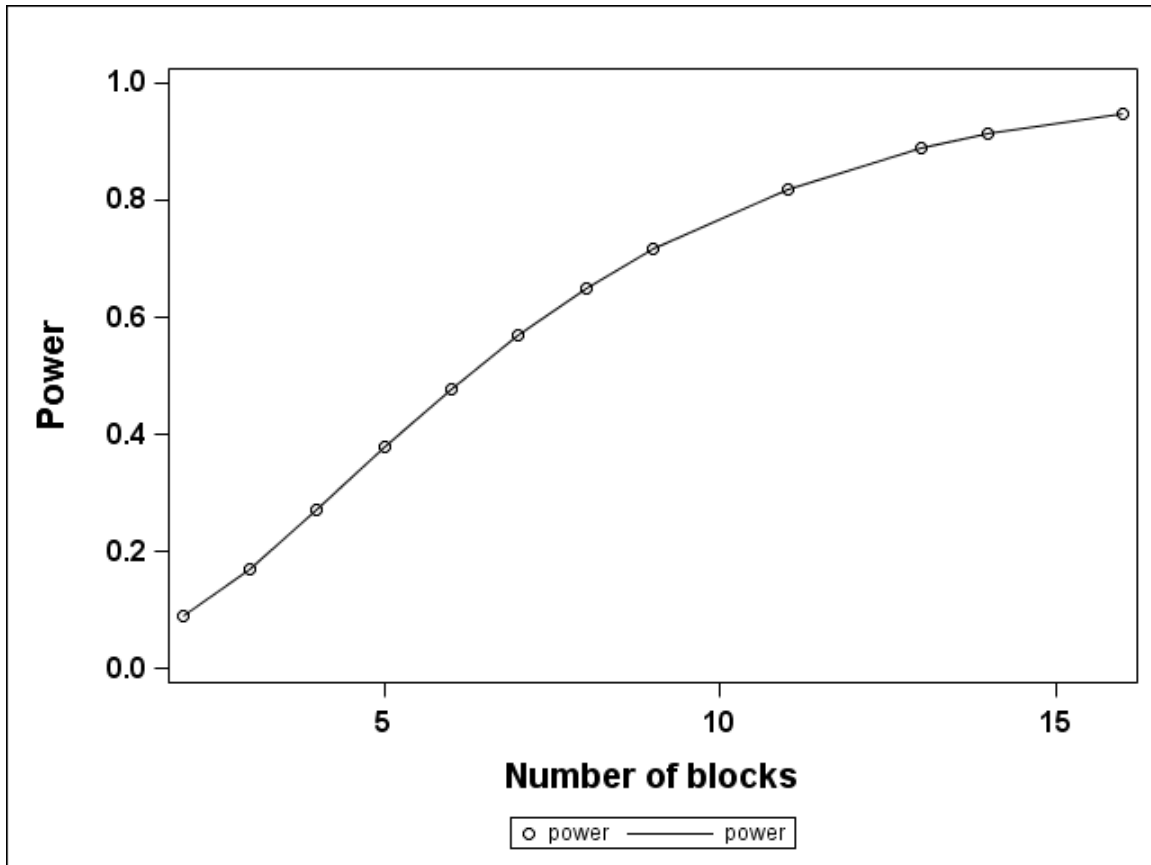
SAS Workshop
Power and Sample Size
Hand Out 3

The GLIMMIX procedure is run as before, separately for each setting of B, the block variable. The subsequent data step computes the power separately for each block.

Finally, the estimated power values from the data set are printed and plotted:

```
proc print;
proc sgplot;
    scatter x=b y=power;
    series x=b y=power;
    axis label='Number of blocks' LABELATTRS=(Family=Arial
Size=13 Weight=Bold) VALUEATTRS=(Family=Arial
Size=12 Weight=Bold);
    yaxis label='Power' min=0 LABELATTRS=(Family=Arial
Size=15 Weight=Bold) VALUEATTRS=(Family=Arial
Size=12 Weight=Bold);
run;
```

| Obs | B | Effect | Num DF | DenDF | FValue | ProbF | alpha | nc | fcrit | power |
|-----|----|-----------|-----------|-------|--------|--------|-------|---------|---------|---------|
| 1 | 2 | Treatment | 1 | 1 | 1.85 | 0.4034 | 0.05 | 1.8525 | 161.448 | 0.09005 |
| 2 | 3 | Treatment | 1 | 2 | 2.78 | 0.2375 | 0.05 | 2.7787 | 18.513 | 0.17035 |
| 3 | 4 | Treatment | 1 | 3 | 3.70 | 0.1499 | 0.05 | 3.7050 | 10.128 | 0.27210 |
| 4 | 5 | Treatment | 1 | 4 | 4.63 | 0.0978 | 0.05 | 4.6312 | 7.709 | 0.37772 |
| 5 | 6 | Treatment | 1 | 5 | 5.56 | 0.0650 | 0.05 | 5.5574 | 6.608 | 0.47804 |
| 6 | 7 | Treatment | 1 | 6 | 6.48 | 0.0437 | 0.05 | 6.4837 | 5.987 | 0.56874 |
| 7 | 8 | Treatment | 1 | 7 | 7.41 | 0.0297 | 0.05 | 7.4099 | 5.591 | 0.64811 |
| 8 | 9 | Treatment | 1 | 8 | 8.34 | 0.0203 | 0.05 | 8.3361 | 5.318 | 0.71594 |
| 9 | 11 | Treatment | 1 | 10 | 10.19 | 0.0096 | 0.05 | 10.1886 | 4.965 | 0.81984 |
| 10 | 13 | Treatment | 1 | 12 | 12.04 | 0.0046 | 0.05 | 12.0411 | 4.747 | 0.88911 |
| 11 | 14 | Treatment | 1 | 13 | 12.97 | 0.0032 | 0.05 | 12.9673 | 4.667 | 0.91383 |
| 12 | 16 | Treatment | 1 | 15 | 14.82 | 0.0016 | 0.05 | 14.8198 | 4.543 | 0.94881 |



This demonstrates that the number of blocks needs to be 11 or more to reach a statistical power for the treatment effect of at least 0.80.