

Notes on Using the SPSS *manova* Procedure for Power Calculations

This method draws on the idea given by D'Amico, Neilands & Zambarano (2001), but gives a simpler way of setting up the matrix version of the dataset.

Summary

1. Set up a 'dummy' dataset in SPSS which is similar to the one for which you want to obtain the power. The means and standard deviations will be drawn from previous research or may be guesswork. It may be easiest to use standardised measures, so that effects (e.g., differences between group means) can be specified in terms of standard deviations.
2. Run the *manova* procedure in order to save the data in matrix form.
3. Alter the values (the number of cases, differences between means, correlations) and run the *manova* procedure to obtain power calculations. Keep altering the relevant values (usually the number of cases) and rerunning the *manova* analysis to see what values are necessary to obtain an acceptable level of power.

Example

Our example is a pre/post design with two groups, treatment and control. The dependent variable is anxiety, measured on a 10-point scale. We're interested in seeing whether anxiety decreases more for the treatment group than for the control group. In other words whether there's a significant interaction between group and time. From previous research or from our guesswork we think that the mean score for each group (which are randomly assigned from a waiting list of people who have come to an anxiety clinic) at pretest will be around 6, and that the control group score will decline a bit without treatment, to say 5.5, but that the effect of treatment will be quite strong, so the post-test mean for the treatment group will be about 4. The standard deviation of this measure of anxiety is known to be about 1.5. With such changes – 2 versus .5 – we'd certainly like to have good chance of calling the interaction significant at alpha .05. We create a dummy dataset like this:

	group	pre	post
1	.00	5.00	4.50
2	.00	7.00	6.50
3	1.00	5.00	3.00
4	1.00	7.00	5.00
5			

Note that we've put in only two observations per case for this dummy dataset (the minimum we can have at this stage) and that we get the means we want by having one case in each group one unit lower than the mean and the other case one unit higher than the mean.

We now run the *manova* commands:

```
manova pre post by group(0,1)/
  wsfactor=time(2)/
  matrix=out(*)/
  design.
```

The matrix subcommand asks for the data to be saved in matrix format:

	rowtype_	group	varname_	pre	post
1	N	.		4.0000000	4.0000000
2	MEAN	.00		6.0000000	5.5000000
3	N	.00		2.0000000	2.0000000
4	MEAN	1.00		6.0000000	4.0000000
5	N	1.00		2.0000000	2.0000000
6	STDDEV	.		1.4142136	1.4142136
7	CORR	.	PRE	1.0000000	1.0000000
8	CORR	.	POST	1.0000000	1.0000000
9					

We now have the data in a form in which we can alter the various values to simulate the data we think we might get. In the present case, we want to see whether having 10 cases per group would provide enough power, so we'll change the numbers in the top row (Total N) to 20, and the Ns for each group to 10. We also need to change the standard deviation to 1.5 at each time point and reduce the correlation between pre and post scores to a more realistic value, such as .5. The altered dataset looks like this:

	rowtype_	group	varname_	pre	post
1	N	.		20.0000000	20.0000000
2	MEAN	.00		6.0000000	5.5000000
3	N	.00		10.0000000	10.0000000
4	MEAN	1.00		6.0000000	4.0000000
5	N	1.00		10.0000000	10.0000000
6	STDDEV	.		1.5000000	1.5000000
7	CORR	.	PRE	1.0000000	.5000000
8	CORR	.	POST	.5000000	1.0000000

We now run a new set of *manova* commands to obtain the power values:

```
manova pre post by group(0,1)/
  wsfactor=time(2)/
  matrix=in(*)/
  power=f(.05) exact/
  design.
```

The relevant section of the *manova* output shows that the power for the interaction is a bit on the low side:

```
Observed Power at the .0500 Level
                Noncen-
Source of Variation   trality   Power
TIME                  13.889    .941
GROUP BY TIME         5.000    .562
```

so we would increase the number of subjects and repeat the analysis until we obtained an acceptable value. In this case it turns out that if we double the group size, the power for detecting the interaction is .869, which is much more acceptable. In fact, we could probably get away with 15 per group, which gives a power of .753.

Note that, in repeated measures analyses, changes in the correlation between measures may have a dramatic effect on power. In fact, if the correlation between the pre- and post-test measures in the present example was a still-realistic .7 rather than .5, the power with only 10 cases per group is .779 rather than .562.

Notes

1. If you really don't know what values to expect, you can set the standard deviation to 1 (as for a standard score), then give differences between means in terms of the standard deviation. Take a simple example in which you wanted to compare two groups. With a standard deviation of 1, if you assigned one group a mean of zero and the other a mean of .5 in the dummy dataset this would be considered a moderate effect size.
2. Another strategy when you don't really know what to expect is to run a series of analyses with various combinations of N and effect sizes (and correlations in a repeated measures design). If the power is reasonable over various more-or-less realistic combinations of values, then the research has a good chance of obtaining a significant result. If not, you may consider that the research is not worth doing with the numbers of subjects you have available.

Further Examples

Oneway ANOVA with Contrasts and Unknown Means and Standard Deviation

In this example, there are three groups of subjects, a control group (*group* = 1) and two treatment groups (2 and 3). In this case we don't know what means and standard deviations to expect, so we'll set the standard deviation equal to one, the mean of the control group to zero, and the means of groups 2 and 3 to .5 and .8 respectively. The difference between the means of groups 1 and 2 is therefore $(.5 - 0) = .5$, and Cohen's $d = .5/1 = .5$, which Cohen (1992) calls a medium effect size. The difference between the control group and group 3 is .8, and $d = .8$, a large effect size.

As before, we type in a dummy dataset with the structure we need, using, in this case, any old numbers for the dependent variable:

	group	score
1	1.00	1.00
2	1.00	2.00
3	2.00	3.00
4	2.00	4.00
5	3.00	5.00
6	3.00	6.00
7		

We then use these manova commands

```
manova score by group(1,3)/
matrix=out(*)/
design.
```

to create the matrix version of the dataset:

	rowtype_	group	varname_	score
1	N	.		6.0000000
2	MEAN	1.00		1.5000000
3	N	1.00		2.0000000
4	MEAN	2.00		3.5000000
5	N	2.00		2.0000000
6	MEAN	3.00		5.5000000
7	N	3.00		2.0000000
8	STDDEV	.		.7071068
9	CORR	.	SCORE	1.0000000
10				

The standard deviation of .7071 is replaced with '1', and the means are given the values mentioned above. We'll assume that we can recruit 60 subjects, 20 in each group. The revised matrix dataset looks like this:

	rowtype_	group	varname_	score
1	N	.		60.0000000
2	MEAN	1.00		.0000000
3	N	1.00		20.0000000
4	MEAN	2.00		.5000000
5	N	2.00		20.0000000
6	MEAN	3.00		.8000000
7	N	3.00		20.0000000
8	STDDEV	.		1.0000000
9	CORR	.	SCORE	1.0000000
10				

We'll also assume that we're not interested in the overall ANOVA result, but rather in two *a priori* contrasts, group 2 versus group 1 and group 3 versus group 1. We can ask *manova* to give the power for each of the contrasts with these commands:

```
manova score by group(1,3)/
  contrast(group)=simple(1)/
  matrix=in(*)/
  power=f(.025) exact/
  design=group(1) group(2).
```

The *simple(1)* option asks for the contrasts we want, and in the *design* statement *group(1)* stands for the group 2 versus group 1 contrast and *group(2)* stands for the group 3 versus group 1 contrast. To hold the overall Type I error at .05, we set *alpha* at .025 for the two contrasts. The relevant part of the output shows that the power for the first contrast is very low, at .24, while that for the second is .59. At this stage we could add more cases to

Observed Power at the .0250 Level		
Source of Variation	Noncen- trality	Power
GROUP(1)	2.500	.244
GROUP(2)	6.400	.592

see how many more cases we would need to achieve an acceptable level of power.

Correlation

Imagine that we want to assess the correlation between measures of anger and narcissism, which we expect to be quite low. We want to have a good chance (power at least .80) of obtaining a significant result if the correlation in the population is .30 or higher. We create the following dummy dataset:

	anger	narciss
1	1.00	2.00
2	2.00	1.00
3	3.00	4.00
4	4.00	3.00
5	5.00	5.00
6		

and use the following commands

```
manova anger with narciss/
  matrix=out(*)/
  design.
```

to produce a matrix version of the data:

type_	rowtype_	varname_	anger	narciss
1	N		5.0000000	5.0000000
2	MEAN		3.0000000	3.0000000
3	STDDEV		1.5811388	1.5811388
4	CORR	ANGER	1.0000000	.8000000
5	CORR	NARCISS	.8000000	1.0000000
6				

The correlation is replaced with .30, and we start the ball rolling with a modest 30 subjects (the values of the means and standard deviations don't matter for our purposes).

The revised matrix dataset is:

rowtype_	varname_	anger	narciss
1	N	30.0000000	30.0000000
2	MEAN	3.0000000	3.0000000
3	STDDEV	1.5811388	1.5811388
4	CORR	ANGER	.3000000
5	CORR	NARCISS	1.0000000
6			

and the relevant output is:

```
-----  
Observed Power at the .0500 Level  
Source of Variation      Noncen-      Power  
                        trality  
Regression                2.769      .362  
-----
```

which is a bit on the low side. Further experimentation shows that 85 subjects are needed to achieve a power of .80.

Using an Existing Dataset

It's often not very sensible to calculate the power for an existing dataset (if the effects are significant, the power will be seen to be adequate; if the effects are not significant, the power was too low), but it can be sensible to ask, for an effect which wasn't significant, "how many more cases would be needed to have a good chance of finding the effect to be significant if the population characteristics are the same as those of my sample?".

For this example, we'll use the *gln_demo.sav* dataset referred to in *Using the GLM Procedure in SPSS for Windows*, and suppose that we have carried out a multivariate analysis with the variables *test1* to *test3* as the dependent variables and *group* (with four categories) as the grouping variable. The results for the 99 cases are as follows:

```
EFFECT .. GROUP  
Multivariate Tests of Significance (S = 3, M = -1/2, N = 45 1/2)  
  
Test Name      Value   Approx. F Hypoth. DF   Error DF   Sig. of F  
Pillais        .14202   1.57358     9.00     285.00    .123  
Hotellings     .15739   1.60308     9.00     275.00    .114  
Wilks          .86122   1.59333     9.00     226.49    .118  
Roys           .11361  
  
-----  
EFFECT .. GROUP (Cont.)  
Univariate F-tests with (3,95) D. F.  
  
Variable      Hypoth. SS   Error SS Hypoth. MS   Error MS      F   Sig. of F  
TEST1         2.04393    59.70197   .68131     .62844     1.08412   .360  
TEST2         5.62424    57.69468   1.87475     .60731     3.08695   .031  
TEST3         2.08582    69.51679   .69527     .73176     .95014   .420  
  
-----
```

The *manova* commands to produce the matrix version of the data are:

```
manova test1 to test3 by group(1,4)/  
matrix=out(*)
```

The dataset is as follows:

	rowtype_	group	varname_	test1	test2	test3
1	N	.		99.0000000	99.0000000	99.0000000
2	MEAN	1		4.4244974	4.5246307	4.2067255
3	N	1		29.0000000	29.0000000	29.0000000
4	MEAN	2		4.6060099	4.8869344	4.2621851
5	N	2		24.0000000	24.0000000	24.0000000
6	MEAN	3		4.7934063	5.1225354	4.2714072
7	N	3		31.0000000	31.0000000	31.0000000
8	MEAN	4		4.6287784	4.9799732	4.6443188
9	N	4		15.0000000	15.0000000	15.0000000
10	STDDEV	.		.7927432	.7793025	.8554272
11	CORR	.	TEST1	1.0000000	.0415152	.1233156
12	CORR	.	TEST2	.0415152	1.0000000	.2428773
13	CORR	.	TEST3	.1233156	.2428773	1.0000000
14						

Before going any further, we can use these commands to calculate the observed power:

```
manova test1 to test3 by group(1,4)/
  matrix=in(*)/
  power=f(.05) exact.
```

The power for the Wilks' Lambda statistic is .62. Let's see what improvement we would get if we added five cases to each group, an addition of 20 more subjects overall. We could add these subjects manually, but the following commands will do it for us, and will make it easier to add more in the future:

```
do if (rowtype_ eq "N" and sysmis(group)).
  compute test1=test1 + 20.
  compute test2=test2 + 20.
  compute test3=test3 + 20.
else if (rowtype_ eq "N" and ~sysmis(group)).
  compute test1=test1 + 5.
  compute test2=test2 + 5.
  compute test3=test3 + 5.
end if.
execute.
```

When the *manova* commands are run again, we find that the power is now .72. If we rerun the above syntax and add another five subjects to each group, giving 139 subjects overall, we obtain a power of exactly .80.

Multiple Regression

The proposed analysis will regress the dependent variable *y* on *x1* and *x2*, which are numeric variables, and *x3*, which is dichotomous. As usual, we set up a dummy dataset, using any old numbers, with one exception. With the dichotomous variable, *x3*, it's probably a good idea to put in zeroes and ones (e.g., female = 0, male = 1) in the same proportion as you would expect them to occur in your sample. The reason is that the mean and standard deviation of a proportion are linked, so we don't want to arbitrarily change one or both once the matrix dataset has been created. In this example, we'll assume a 50/50 distribution.

The initial dataset is:

	y	x1	x2	x3
1	1.00	2.00	4.00	.00
2	2.00	1.00	1.00	.00
3	3.00	4.00	2.00	.00
4	4.00	3.00	5.00	1.00
5	5.00	6.00	3.00	1.00
6	6.00	5.00	6.00	1.00
7				

The manova commands to produce the matrix version of the data are:

```
manova y with x1 x2 x3/
matrix=out(*).
```

The initial matrix version of the dataset is:

rowtype_	varname_	y	x1	x2	x3
1	N	6.0000000	6.0000000	6.0000000	6.0000000
2	MEAN	3.5000000	3.5000000	3.5000000	.5000000
3	STDDEV	1.8708287	1.8708287	1.8708287	.5477226
4	CORR Y	1.0000000	.8285714	.5428571	.8783101
5	CORR X1	.8285714	1.0000000	.3714286	.6831301
6	CORR X2	.5428571	.3714286	1.0000000	.6831301
7	CORR X3	.8783101	.6831301	.6831301	1.0000000
8					

We're going to assume (in a real example, this could be on the basis of past research, or theory) that *x1* is moderately (.5) correlated with *y*, but *x2* and *x3* are only weakly correlated with *y* (both .3). Furthermore we're going to assume that *x1* and *x2* are quite highly correlated (.6), but neither is correlated with *x3* (.1). We think we can obtain 50 subjects. The amended dataset is:

rowtype_	varname_	y	x1	x2	x3
1	N	50.0000000	50.0000000	50.0000000	50.0000000
2	MEAN	3.5000000	3.5000000	3.5000000	.5000000
3	STDDEV	1.8708287	1.8708287	1.8708287	.5477226
4	CORR Y	1.0000000	.5000000	.3000000	.3000000
5	CORR X1	.5000000	1.0000000	.6000000	.1000000
6	CORR X2	.3000000	.6000000	1.0000000	.1000000
7	CORR X3	.3000000	.1000000	.1000000	1.0000000
8					

The manova commands

```
manova y with x1 x2 x3/
matrix=in(*)/
power=f(.05).
```

give the following output:

```

-----
Observed Power at the .0500 Level
Source of Variation      Noncen-      Power
                        trality
Regression                20.986      .970

```

```

-----
Regression analysis for WITHIN CELLS error term
--- Individual Univariate .9500 confidence intervals
--- two-tailed observed power taken at .0500 level
Dependent variable .. Y

```

COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t
X1	.48418	.00000	.153	3.166	.003
X2	-.01582	-.11076	.153	-.103	.918
X3	.86472	6.05305	.420	2.059	.045

COVARIATE	Lower -95%	CL- Upper	Noncent.	Power
X1	.176	.792	10.025	.873
X2	-.324	.292	.011	.051
X3	.019	1.710	4.240	.522

There is ample power (.97) for the overall regression, but we are also interested in the power for each independent variable. The above results show the power for an alpha of .05; we might want to use a different *alpha*, e.g., $.05/3 = .0167$, for the predictors. We can run the *manova* commands again, with *power=f(.0167)*, and obtain (only part of the output is shown):

```

-----
--- two-tailed observed power taken at .0167 level

```

COVARIATE	Lower -95%	CL- Upper	Noncent.	Power
X1	.176	.792	10.025	.750
X2	-.324	.292	.011	.017
X3	.019	1.710	4.240	.345

The power for *x1* is still adequate, but a bigger sample would be needed to achieve acceptable values for *x2* and *x3*. (A sample of 100 gives a value of .70 for *x3*; *x2* is a lost cause because of its correlation with *x1*.)

Alan Taylor
 Department of Psychology
 20th October 2004

References

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.

D'Amico, E., Neilands, T., & Zambarano, R. (2001). Power designs for multivariate and repeated measures designs: A flexible approach using the SPSS MANOVA procedure. *Behavior Research Methods, Instruments, & Computers*, 33, 479-484. [BF180.B4, R]