Basic Sample Size and Power Estimation

The SAS procedure PROC POWER is designed to handle many different types of sample size and power computations. While this procedure can deal with more advanced designs such as ANOVA, linear regression and logistic regression, these techniques are better handled through other procedures (see Handout #3). In this workshop, PROC POWER will be used to address two types of problems: Estimation related to a single mean or proportion, and estimation/comparison of two independent means or proportions.

Single means/proportions

The POWER procedure is called using standard SAS format with the PROC POWER statement. For the case of a single mean, the statement ONESAMPLEMEANS is used with several options specifying the components outlined in Handout #1. The first is the **mean=** option, giving the value anticipated for the mean estimate. The **nullmean=** option provides information for δ, the precision component. Several precision values can be specified at once in order to generate several corresponding sample size or power estimates. In the example below, the option is called with the mean value set to 20, which will be evaluated with several precision values: 5, 10, 15 and 20. Note, however, the precisions are specified as the mean + δ, e.g. 25, 30, 35, and 40. The variability component, σ, is given with the **std=** option and the desired statistical power, $\mathbf{1 - \beta}$, is fixed to 0.95. In this example, we are computing the estimated sample size, hence the option **ntotal=** is set to a missing value (a single period). This tells SAS that **n** is to be estimated. If we were computing estimated power, the **ntotal=** statement should be set to a predetermined number, while the **power=** option would be given a missing value, thereby informing SAS we want to estimate power. By default, α is set to 0.05, but can be specified with an **alpha=** option.

Example 1.

```
proc power;
        ONESAMPLEMEANS
                mean=20
                nullmean=25, 30, 35, 40
                std=5
                ntotal=.
                power=.95;
run;
```

Tests for proportions can be run in a similar manner using the ONESAMPLEFREQ statement and **proportion =** and **nullproportion=** options to specify the expected

proportion and precision(s), respectively. The expected proportion in this example is 0.31 or 12/39.

Example 2.

```
proc power;
        onesamplefreq
        proportion=0.31
        nullproportion = 0.01 to 0.3 by 0.01
        ntotal = 20, 30, 50
        sides=1
        power=.;
        ods output Output=power;
run;

proc sort data=power;
        by nullproportion NTotal ;
run;

proc transpose data=power out=power;
        var power;
        id NTotal;
        by nullproportion;
run;

data power;
        set power;
        delta = 12/39 - nullproportion;
run;

proc sgplot;
        series x=delta y = _20/name='twenty' legendlabel="N=20";
        series x=delta y = _30/ legendlabel="N=30";
        series x=delta y = _50/ legendlabel="N=50";
        refline 0.8 /axis=Y lineattrs=(pattern=shortdash color=black);
        xaxis label='Detectable difference at 95% confidence' 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;
```

Here, **n**, is set to three values and power is being estimated. Additionally, several precisions in the form of the **nullproportion=** statement are evaluated from 0.01 to 0.30 in increments of 0.01. The power computation is also designed to test the one sided hypothesis that the observed proportion is greater that the null values using the **side=1** option. The ODS statement outputs the results into data set 'power' and the remainder of the code plots the results in a Power Plot for visual assessment.

Comparison of Two means/proportions

As in the single estimate case, there are different statements for means or proportions: TWOSAMPLEMEANS and TWOSAMPLEFREQ, respectively.

Options in these statements now provide information on pairs or groups. For example, the **groupmeans=** option specifies the means for two samples. They are separated by the vertical line character " **|** ". In the example below, one mean is set to 20, while the second mean is evaluated at four values: 25, 30, 35, and 40. While this is similar to the one sample case above, we are now estimating the sample size for the difference of two means, rather than the sample size for estimating one mean at varying precisions. That is, the sample size will be evaluated for four tests: 20 vs 25, 20 vs 30, 20 vs 35, and 20 vs 40. The precision component, δ, in the two sample problem is now represented as the difference between the two estimated quantities. Similar to the means, the variability must also now be specified for two groups using the **groupstds=** option. In this case, only one standard deviation per group is given and they are both set to 5. The **test=diff_satt** tells SAS that the hypothesis to be tested is the difference of the two means and that the standard deviations of each group *could* be different. The sample size can be specified using either the **ntotal=** or **npergroup=** option. The **power=** option is specified as before.

Example 3.

```
proc power;
        TWOSAMPLEMEANS
                groupmeans=20|25 30 35 40
                groupstds=5|5
                Test=DIFF_SATT
                npergroup=.
                power=.95;
run;
```