SAS sample size and power computation

What to input

The SAS procedure PROC POWER will compute the sample size for a desired power. The following information is needed.

(1) The population means for which we would want to have statistically significant F-statistic.

(2) The population standard deviation.

(3) The level of significance of the F-test, typically .05.

(4) The desired power of the F-statistic, typically between .5 and .9.

Example. In Example 3 of the slides, we had population means 15, 16, 17, and 18, population standard deviation 2.20, level of significance 0.05 for the F-test, and we obtained sample sizes required for each of the powers .5, .6, .7, .8, .9.

SAS code

The SAS PROC POWER code for doing this is shown below.

```
proc power;
  onewayanova
    test=overall
    groupmeans = 15 | 16 | 17 | 18
    stddev = 2.20
    alpha = 0.05
    npergroup = .
    power = 0.5 0.6 0.7 0.8 0.9
    ;
run;
```

The specified group means are separated by vertical lines. The `stddev` option defines the presumed population standard deviation, `alpha` sets the desired level of significance, and "`npergroup = .`" asks SAS to compute the number of observations per group to achieve the power levels specified in the `power` option. (The dot indicates a missing value, and Proc Power will calculate these.)

A note about the syntax:

"onewayanova" is ONE statement in the power procedure. The additional lines for "test", "groupmeans", "stddev", etc. are options for the "onewayanova" statement. The semicolon after the line for "power" ends the "onewayanova" statement.

Output

```
        The POWER Procedure
  Overall F Test for One-Way ANOVA

      Fixed Scenario Elements

Method                      Exact
Alpha                        0.05
Group Means           15 16 17 18
Standard Deviation            2.2


       Computed N Per Group

         Nominal    Actual    N Per
 Index     Power     Power    Group

     1       0.5     0.529        7
     2       0.6     0.603        8
     3       0.7     0.725       10
     4       0.8     0.817       12
     5       0.9     0.906       15
```

The last column gives the number of observations needed in each group to achieve the desired power. The column for Nominal Power shows the values for Power that we specified. Achieving this power exactly would require a fractional number of observations in each group. The column for Actual Power is the power when the sample size is increased to an integer.

Determining power for given sample sizes

In addition to determining the sample size that produces a given power, it is also possible to determine power for a given sample size. The code below has the same means and standard deviation as the example above, but it determines the power for sample sizes of n = 3, 4, 5, 6, and 7 per group.

```
proc power;
 onewayanova
   groupmeans = 15| 16 | 17 | 18
   stddev = 2.20
   alpha = 0.05
   npergroup = 3 4 5 6 7
   power = .
   ;
run;
```

```
               The POWER Procedure
        Overall F Test for One-Way ANOVA

             Fixed Scenario Elements

      Method                           Exact
      Alpha                             0.05
      Group Means                15 16 17 18
      Standard Deviation                 2.2


                  Computed Power

                         N Per
               Index     Group      Power

                   1         3      0.190
                   2         4      0.277
                   3         5      0.364
                   4         6      0.450
                   5         7      0.529
```

Note that the last line indicates that the power will be 0.529 when there are 7 observtions in each group. This coincides with the results on page 2.

Power curves

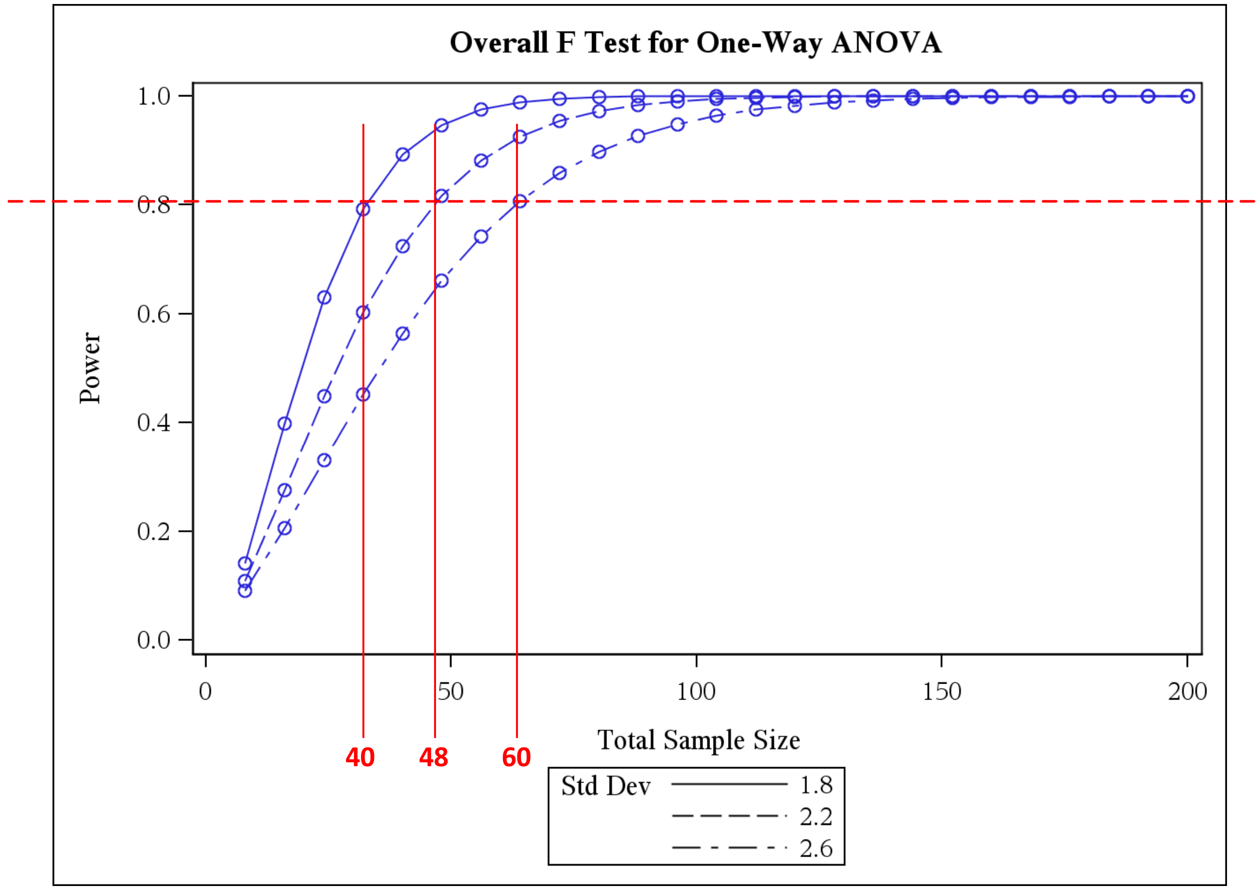
```
ods graphics on;
proc power;
 onewayanova
   test=overall
   groupmeans=15 | 16 | 17 | 18
   stddev= 1.8 to 2.6 by 0.4
   alpha = 0.05
   ntotal = 8 to 200 by 4
   power=.;
 plot  x=n min=8 max=200;
 run;
ods graphics off;
```

The 'ods graphics on' and 'ods graphics off' statements may be omitted if you are using SAS on a local server. They are required if you are using SAS OnDemand (in the cloud).

We are keeping the group (treatment) means the same as before, but we are providing a range of values for the population standard deviation and the sample sizes. Note that the sample sizes are defined in the "ntotal" option, and we are providing the total sample size (across all four groups). To get the sample size per group, you will need to divide by 4. To avoid issues with fractional sample sizes in each group, we increment the total sample size "by 4", so the increment in each group is 1.

The plot statement is a new statement in proc power. In other words, 'plot' is NOT an option for the 'onewayanova' statement.

The resulting power curve, shown on the next page, is very useful to gauge how sensitive the sample size is to changes in the presumed population standard deviation. The standard deviations is rarely known, and must be estimated from prior studies. It is possible for minor changes in the standard deviation to greatly affect the sample size, so it is important to know the sample sizes for a range of plausible standard deviations. This is precisely what we have done in this example -- we originally assumed the standard deviation was 2.2 and now we are seeing the changes in sample size if the standard deviation is "off" by 0.4 (i.e., from a low of 1.8 to a high of 2.6). It is also possible to vary the values of alpha, and generate curves similar to those shown on the next page.



This is the same graph shown in the slides for this lesson, but we have added some additional detail in the interpretation.

The three curves correspond to the three values for the standard deviation, as shown in the legend at the bottom. The curve corresponding to the smallest standard deviation (the solid line) will have the largest power because less variation in the population makes it easier to detect differences in the means.

The horizontal red dashed line indicates 80% power. The three vertical red lines indicate the point on each curve that corresponds to 80% power.

- If the population standard deviation is 1.8, we need a total sample size of approximately 40, or 10 per group.
- If the population standard deviation is 2.2, we need a total sample size of approximately 48, or 12 per group.
- If the population standard deviation is 2.6, we need a total sample size of approximately 60, or 15 per group.