Statistical Power Analysis Using SAS and R

A Senior Project
Presented to
The Faculty of the Statistics Department
California Polytechnic State University, San Luis Obispo

By
Peter Osmena
March 2010

Table of Contents

This paper will go over power and how power is calculated using SAS or R. Chapter 1 will give an introduction to power, what it is, and what is needed for the calculation of power. Chapter 2 goes in depth of the power calculations for a general ANOVA test and a chi squared test. Chapter 3 contains examples and syntax for calculating power using SAS and R. It will also go through the plotting capabilities of power curves in SAS.

1   Introduction to Power

This chapter will introduce the concept of power and what things are needed to calculate it, why it's important and what affects it. It also goes through a simple example of calculating power and sample size.

1.1 What is Power?

Power is one of the most important things in experimental design. When a hypothesis test does not reject the null hypothesis when it's false a type II error has been made. The power of the test is the probability of rejecting the null hypothesis when it is false, in other words that the test will not make a type II error. Thus power is defined as,

$$\text{Power} = 1 - P(\text{Type II Error}) = P(\text{Rej } H_0 \text{ when } H_0 \text{ is false}) \quad \textbf{(1.1.1)}$$

Since power is the probability of correctly rejecting the null hypothesis when it is false it makes sense that we would like this as large as possible. To do this, as shown in the formula above, we would like the probability of a type II error to be as small as possible. As well as having a small probability of a type II error, for a test to be useful it also has to have a small probability of a type I error, rejecting the null hypothesis when it is true.

1.2 Why is Power Important?

The usefulness of a test will be determined with power. A useful test would be one that, with a high probability, correctly rejects the null hypothesis. One that does this has high power.

For example suppose one has a test designed to detect if a person has cancer. If the test has low power, the probability of correctly determining the patient has cancer is low. Thus, there is a good possibility doctors will tell the patient they don't have cancer, when in fact they do. This would not be an optimal situation. Most people would want to know if they indeed have cancer.

A power analysis is generally used in one of two ways, determining the power of the test or determining the needed sample size to achieve a certain power. It is generally done before the data is collected. Suppose that a company can only afford 10 runs of an

experiment, they run a power analysis and find the power of their experiment is.50. In this situation it would not be beneficial for them to run the experiment, since in reality they can correctly reject the null hypothesis with the same probability as flipping tails on a coin.

A power analysis is useful when determining sample size. One of the issues in running experiments is the cost. In general, the bigger the sample size the more expensive the experiment is to run. Thus, when running experiments most people would like to know the fewest number of runs or samples that are needed to achieve a high pre-specified power.

 The power of a test will determine if the test is useful and worth while to do, or if the analysis was done afterwards, to see if the results of an experiment can be trusted. The higher the power, the higher probability of correctly rejecting the null hypothesis. Sample size calculations determine how big a sample needs to be and how much the experiment will cost. One can see a power analysis is an important and beneficial thing to do before running an experiment or conducting a study.

1.3 What is Needed to Calculate Power and Sample Size?

There are several parameters that are needed in a power or sample size calculation. Each of which will affect power or sample size, either increasing them or decreasing them. All of the parameters of the calculation and their effects are discussed below. The explanation assumes all the other parameters are being held constant.

Type I Error Rate

Since power is defined as the probability of correctly rejecting the null hypothesis it makes sense that the probability of rejecting the null hypothesis when it's true will affect power. This probability, incorrectly rejecting the null hypothesis, is the probability of a Type I error, α. The smaller the alpha, the smaller the probability that the null hypothesis will be rejected. A type II error, β, is the probability that the one fails to reject the null hypothesis when it's false.   As alpha increases, beta decreases, and power increases. As alpha decreases, beta increases, and power decreases.

If one is looking for sample size, as alpha increases the sample size decreases and as alpha decreases sample size increases.

Standard Deviation

Standard deviation is the variation or spread of the data. The less variable the data is, the easier it is to approximate the population value that is being measured. The more variable the data, the harder it is to approximate the population value that is being measured.   Therefore, it is optimal to have the standard deviation be small.

However, the standard deviation of the population is hardly ever known. Thus, one must approximate one. There are several ways to do this. One of the ways this can be done is by taking the range of the data (the maximum value – the minimum value in normally

distributed data) and dividing it by four or six. Since a majority of data falls two deviations of either side of the population value being measured and most of it falls three deviations of either side of the population value being measured dividing the range of the data by 4 or 6 is a good approximation of the standard deviation.

A pilot study could also be used to find an approximation of the standard deviation. One can run a small version of the actual experiment and use the standard deviation found from it, as an approximation for the standard deviation of the regular experiment.

In some cases one can find a previous study that looks at the same population value of interest and use the standard deviation of it. Or one can take their best guess. Regardless of which method is used, a standard deviation must be chosen.

If the standard deviation is small, this will result in greater power than if the standard deviation is large. If the standard deviation is large, then the power will be low.

Thus, a small standard deviation will need fewer samples to achieve the same power as a large standard deviation.

Sample Size/Power
To determine n, the power needs to be specified. To determine power, n needs to be specified. The affect that sample size has on power is fairly intuitive. With a greater sample size the population value of interest can be easier seen against the background of variability. A larger sample size will in turn have a greater power than a small sample size. By the same token gaining higher power requires the test to have more samples than if one were to settle for less power.

Effect Size
The effect size is a very important component in a power analysis. It is the part of the analysis that is the most misunderstood. Effect size, Δ, is the size of the effect that one expects to see in the test. In other words it's how small or big a difference from the null hypothesis that one wants to detect.

For example, if the null hypothesis states the mean IQ of children is 100 and it is found to be 96, the effect size would be 4 IQ units ($\Delta = 4$). For another example, say that from a population of male patients at a psych ward 52% have bipolar disorder. Tested against the assumption that 50% have bipolar disorder, would give a 2% effect size ($\Delta = 2$).

All in all the effect size is an index of the degree of departure that one wants to find from the null hypothesis. It is not hard to see that on average cherries are smaller than oranges, since there is a 'big' effect size. However, it is much harder to see that on average cherries are bigger than grapes, because of the 'small' effect size.

The relationship between power and effect size is fairly intuitive. Big effect sizes are easier to detect and therefore have greater power, all else being equal. In contrast,

detecting a very small effect is more difficult. If there is a small effect size, the power will be low compared to that of a large effect size.

In the case of sample size, one would need a lot more samples to detect a small effect size.

Finding and selecting effect sizes will be discussed in later chapters.

1.4 Simple Example Using a z-test

Example 1.4.1 Calculating Power
Suppose that we are testing H_0: $\mu \leq 80$ versus H_1: $\mu > 80$. We know that n = 50, σ = 5, Δ= 1, and α = .05. Since Δ= 1 we can express the alternative hypothesis as H_1: $\mu > 81$
Calculate power.

To calculate power, first find the standard deviation of the sample mean which is
$$\frac{5}{\sqrt{50}} = .707.$$

Then find the z score of the rejection region, $z_{1-\alpha}$. This P($z* \geq z_{1-\alpha}$), where z^* is the test statistic. Here, $z_{1-\alpha}$ = 1.645 since α = 0.05. Using this we then find this point under the alternate distribution. The alternate distribution is normal with a mean of 81 and a standard deviation of 0.707. Thus 80+(1.645)(0.707) = 81.16. Converting this to a z score we get   z = (81.16 – 81)/0.707 = 0.23. The probability that we will get a z score greater than 0.23 is the power of the test. Thus, the probability that the sample mean is greater than 81.16 is 0.4090. Figure 1.4.1, on the next page, illustrates this. The solid black curve is the null distribution and the dotted red curve is the alternative distribution. The gray vertical line is the critical value and the area to the right represents the power.

Figure 1.4.1

MeanValues

$\boxed{\text{Mean} = 81.16}$
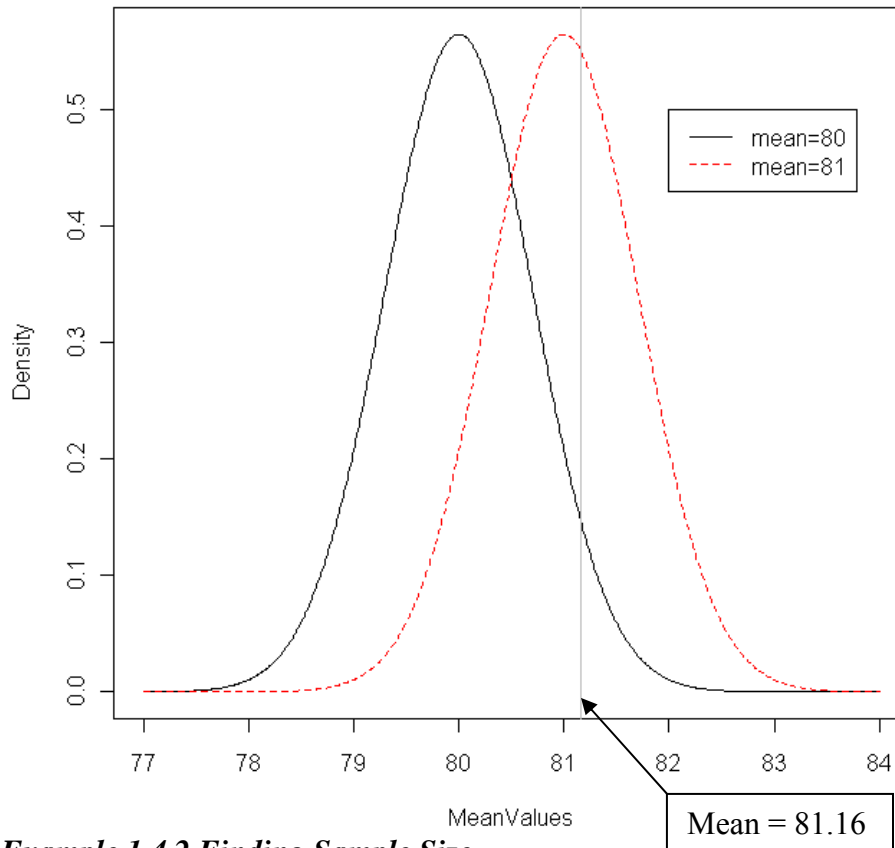
Example 1.4.2 Finding Sample Size

Using the information from example 1.4.1, to find the sample size needed to achieve a power of 0.90, we first must use the null distribution to express the critical point for the test in terms of n. The critical point is $80 + 1.645\left(\dfrac{5}{\sqrt{n}}\right)$. The power of the test is the area of the rejection region under the alternative hypothesis which is 0.90. This gives $z_{1\text{-power}} = -1.28$ so we get the expression $81 - 1.28\left(\dfrac{5}{\sqrt{n}}\right)$. Since there is only one critical point, the two expressions are equal. We then can set them equal to each other and solve for n.

$$80 + 1.645\left(\frac{5}{\sqrt{n}}\right) = 81 - 1.28\left(\frac{5}{\sqrt{n}}\right).$$ Solving gives us $n \approx 214$.

2 Theory of Power

This chapter will go into the theory and computations of calculating power for both a one way ANOVA test and a chi square test. It will also show you some examples of using SAS and R to do the analysis. Details of syntax are discussed in chapter 3.

2.1 Theory of Power for a One WayANOVA

To calculate power for a global F test in a completely randomized design with one treatment at k levels we first need to understand and define the hypothesis. The hypotheses are defined as,

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_1: $\mu_i \neq \mu_j$ for some i,j where i≠j

μ_i = mean of group i
k = number of groups

This test is the F test for equality of means in a one way ANOVA. It assumes that the data is normal with common group variances. Also $N \geq k+1$ and $n_i \geq 1$, where N is the total sample size and n_i is the sample size of group i.

The distribution of the F statistic under the null hypothesis follows a central F distribution, whereas the distribution of the F statistic under the alternative hypothesis follows a noncentral F distribution with the noncentrality parameter, λ. Thus, when the null hypothesis is true it follows a central F distribution and when it's false it follows a noncentral F distribution. Therefore power can be defined as the probability that the F statistic follows a noncentral distribution.

The exact power is given as,

$$\text{Power} = P(F(k-1, N-k, \lambda) \geq F_{1-\alpha}(k-1, N-k)) \qquad \textbf{(2.1.1)}$$

Figure 2.2.1 shows you what the distributions of both the hypotheses look like and will visually show you the power.

Figure 2.2.1
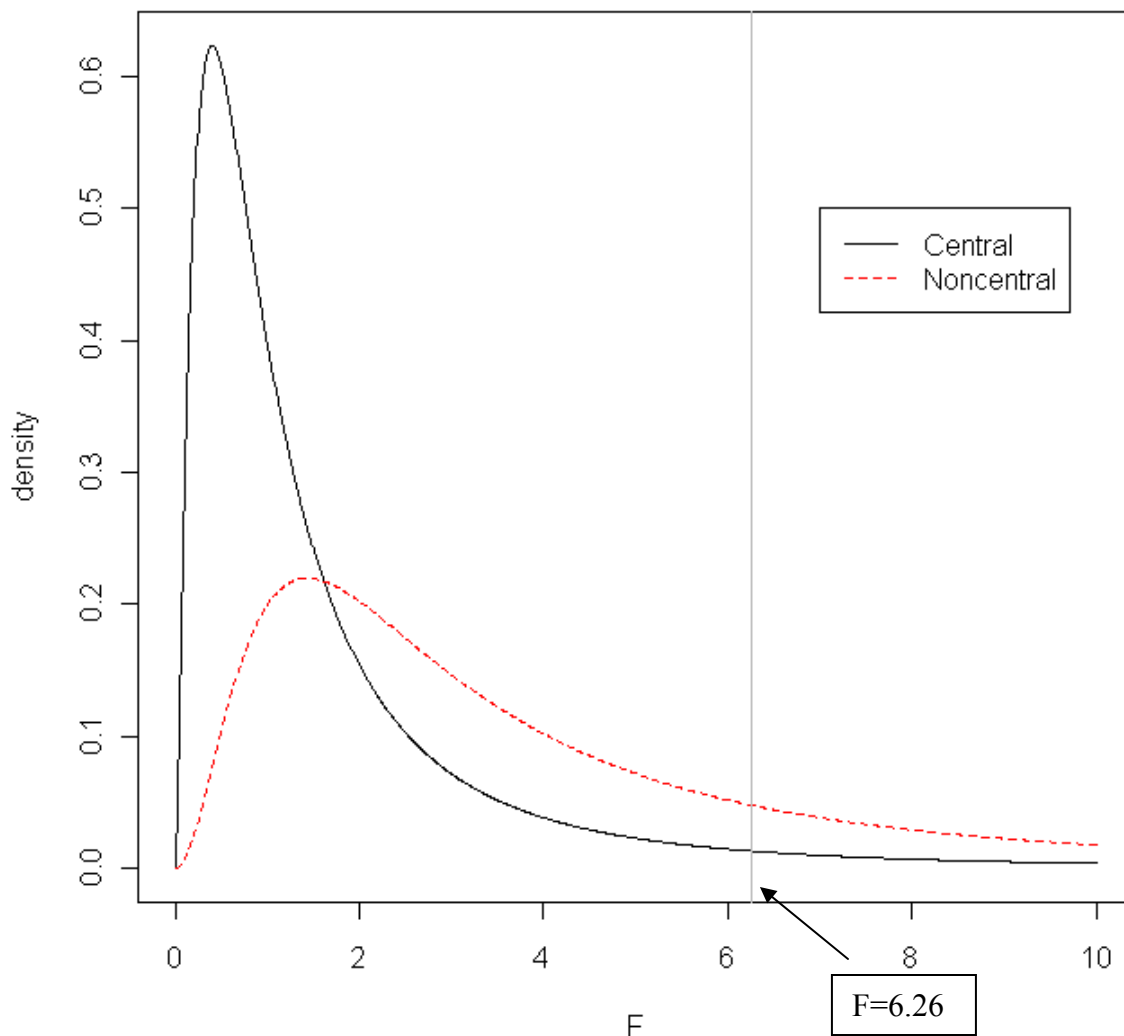
In figure 2.2.1 the solid black curve is the F-distribution under the null hypothesis (this is arbitrarily chosen to be an F distribution with 5 and 4 degrees of freedom). The dotted red curve is the distribution under the alternative hypothesis (same F with a noncentrality parameter of 9). The vertical line represents the F value with a probability of 0.05 under the null hypothesis. If we get an F statistic to the left of the line we conclude that it came from the null distribution and if it is to the right of the line then we conclude it came from the alternative distribution. The area under the alternative distribution to the right of the line represents the power of the test. As you can see this particular test will have relatively low power since the area under the curve is rather small.

SAS and R both use formula 2.1.1 to compute power. However, they differ in the way they define λ, the noncentrality parameter.

SAS defines the noncentrality parameter as,

$$\lambda = N \left(\frac{\sum_{i=1}^{k} \omega_i \, (\mu_i - \mu)^2}{\sigma^2} \right) \qquad (2.1.2)$$

Where μ is the overall mean and ω_i is the weight for the ith group. The weights would be used if one was using an unbalanced design. An unbalanced design is discussed later in the section.

R defines λ in a more complicated way. The computation relies on the effect size of the test. Recall form section 1.3 effect size is how big or small a difference you want to detect from the null hypothesis.

To find the effect size, Δ, we use the formula,

$$\Delta = \frac{\sigma_\mu}{\sigma} \qquad (2.1.3)$$

where,

$$\sigma_\mu = \sqrt{\frac{\sum_{i=1}^{k} (\mu_i - \mu)^2}{k}} \qquad (2.1.4)$$

σ_μ = between "mean" variation
σ = error variaiton

Using this λ is defined as,

$$\lambda = \Delta^2 n(k) \qquad (2.1.5)$$

n = sample size of balanced groups

Either way you do it you will come up with the same answer. This is shown below.

Formula 2.1.4 can be rewritten as

$$\sigma_\mu = \sqrt{\sum_{i=1}^{k} \frac{1}{k} (\mu_i - \mu)^2}$$

Then rewriting formula 2.1.3 we get,

$$\Delta = \frac{\sqrt{\sum_{i=1}^{k} \frac{1}{k}(\mu_i - \mu)^2}}{\sigma}$$

Using formula 2.1.5 we have,

$$\lambda = \frac{\sum_{i=1}^{k} \frac{1}{k}(\mu_i - \mu)^2}{\sigma^2} nk$$

nk = N so,

$$\lambda = N\left(\frac{\sum_{i=1}^{k} \frac{1}{k}(\mu_i - \mu)^2}{\sigma^2}\right)$$

This is equivalent to formula 2.1.2 where $\omega_i = \dfrac{1}{k}$

To find the sample size needed for a specific power you would set formula 2.1.1 equal to the desired power and solve for N. Both SAS and R will go through an algorithm to find an N that will satisfy the equation.

Examples 2.1.1 and 2.1.2 show you syntax and results for finding power and n in both SAS and R.

Example 2.1.1 Finding power of a balanced one way ANOVA
Suppose we have,

$\mu_1 = 59$	$\sigma = 12$	k = 3
$\mu_2 = 66$	$\alpha = 0.05$	
$\mu_3 = 42$	n = 4	

Using this information we have:
f = 0.84

> pwr.anova.test(f=.84,k=3,n=4,sig.level=0.05)

Balanced one-way analysis of variance power calculation

      k = 3
      n = 4
      f = 0.84
  sig.level = 0.05
    power = 0.5848498

 NOTE: n is number in each group

<u>SAS Input/Output</u>
```
proc power;
      onewayanova test=overall
      groupmeans=59|66|42
      std=12
      npergroup=4
      power=.
      ;
run;
```

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

                        Fixed Scenario Elements

             Method                              Exact
             Group Means                      59 66 42
             Standard Deviation                     12
             Sample Size Per Group                   4
             Alpha                                0.05


                          Computed Power

                              Power

                              0.585
```

Example 2.1.2 Finding n of a balanced one way ANOVA
Suppose,

μ_1= 15	σ = 1.5	k = 3
μ_2= 14	α = 0.05	
μ_3= 18	power = .8	

Using this information we have:
f = 1.13

R Input/Output:
> **pwr.anova.test(f=1.13,k=3,power=0.80,sig.level=0.05)**

    Balanced one-way analysis of variance power calculation

        k = 3
        n = 3.718485
        f = 1.13
    sig.level = 0.05
      power = 0.8

 NOTE: n is number in each group

SAS Input/Output:
```
proc power;
      onewayanova test=overall
      groupmeans=15|14|18
      std=1.5
      npergroup=.
      power=.80
      ;
run;
```

The POWER Procedure
Overall F Test for One-Way ANOVA

Fixed Scenario Elements

Method	Exact
Group Means	15 14 18
Standard Deviation	1.5
Nominal Power	0.8
Alpha	0.05

Computed N Per Group

Actual Power	N Per Group
0.846	4

Note that SAS rounds n up to the nearest whole number where R leaves n as a fraction. Thus the actual power may be greater than the power specified in SAS.

Unbalanced ANOVA
If an unbalanced design is used then formula 2.1.2 must be used for finding the noncentrality parameter. This formula takes into account the weights of the sample sizes, whereas formula 2.1.5 does not. For a balanced design these weights are all 1, meaning that all the samples are equal. If you want one sample to be twice as large as the other two samples then you would want the weights to be 2, 1, and 1.

There are many reasons why one would want to use an unbalanced design, one being that maybe one of the treatment combinations of an experiment costs a lot more than the others. So one must decide what they should do; have fewer expensive runs and more cheaper runs or more of the expensive and fewer cheaper runs.

Another reason for using an unbalanced design is that the experiment has a control. Because we most likely know a lot about the control already, we don't need a lot of runs in the experiment. So we decrease the number runs for the control and increase the number of runs for the experimental treatments.

Only SAS will accommodate the unequal sample sizes since it uses formula 2.1.2. Examples 3.2.5 and 3.2.6 show the SAS code and output for finding power and sample for this type of design.

Randomized Complete Block
The purpose of a randomized complete block design is to take out some of the error variance caused by a nuisance factor. If we know what this nuisance factor is then we can block on this factor to decrease variability.

We can find the power of this type of design in a couple of different ways using a one way ANOVA procedure. If one knows, or can approximate, the proportion of error variance the blocks account for, then the standard deviation can be adjusted to reflect this smaller with group variability. This method is easily implemented in either SAS or R.

Or one can take an effect size approach. By reducing the error variation you're increasing the effect size. Having this increased effect size will increase the power. However, the effect size calculation is a little different. The blocked effect size is defined as,

$$\Delta_{block} = \frac{\Delta_{CR}}{\sqrt{1-PV_b}} \qquad \textbf{(2.1.6)}$$

Δ_{block} = Effect size of RCB design
ΔCR = Effect size of the analogous completely randomized design

$$PV_{block} = \frac{\sigma_{block}^2}{\sigma^2} \quad \text{(proportion of between block variation)}$$

$\sigma_{block}^2 = $ between block variance $\quad \sigma^2 = $ design error variance

After finding this effect size using formula 2.2.6 you can then use R to find the power.

You can run this analysis in either way. Doing it the first way is simpler and you can use both SAS and R to find power. The second way using formula 2.2.6 is more involved and you would have to use R. They are essentially the same, but in SAS since you specify the means, you need to adjust the standard deviation yourself. Since R has you specify effect size, you need to adjust it.

2.2.1 Theory of Power for ANOVA Contrasts

Not only does SAS compute power for a global F test, it also computes power for contrasts. Thus we determine power for comparisons. The null and alternative hypotheses for a test of subsets of means to other subsets are,

$$H_0 : c_1\mu_1 + \cdots + c_k\mu_k = c_0$$

$$H_1 : \begin{cases} c_1\mu_1 + \cdots + c_k\mu_k \neq c_0 \\ c_1\mu_1 + \cdots + c_k\mu_k > c_0 \\ c_1\mu_1 + \cdots + c_k\mu_k < c_0 \end{cases}$$

k is the number of groups, $\{c_1, \ldots, c_k\}$ are the contrast coefficients, and c_0 is the null contrast value.

This type of test assumes normal data with common group variances and requires $N \geq k$ and $n_i \geq 1$.

Power is defined as,

$$power = \begin{cases} P(F(1,(N-k,\delta^2 \geq F_{1-\alpha}(1, N-k)) \\ \quad P(t(N-k,\delta) \geq t_{1-\alpha}(N-k) \\ \quad P(t(N-k,\delta) \leq t_{\alpha}(N-k) \end{cases}$$

Where,

$$\delta = N^{\frac{1}{2}} \left(\frac{\sum_{i=1}^{k} c_i \mu_i - c_0}{\sigma \left(\sum_{i=1}^{k} \frac{c_i^2}{\omega_i} \right)^{\frac{1}{2}}} \right)$$

To find the sample size needed for a specific power you would set the power equation equal to the desired power and solve for N. SAS will go through an algorithm to find an N that will satisfy the equation.

The next section will go over the difference in an overall versus a contrast test and show some examples comparing both of them when finding power and n.

2.2.2 Overall Power versus Contrast Power

When doing an ANOVA power analysis, the results may change whether you are using an overall F test or testing for a specific contrast. If sample size is held constant then power may increase or decrease, the same thing is true for sample size if power is held constant.

The reason for this has to do with the effect sizes of the means being tested. Generally, if there is a big effect size there will be greater power. A small effect size yields smaller power.

For example, suppose the group means are 5, 4, 6, 11, 12, and 9 with a standard deviation of 4 and group sample size of 5. A majority of these means fall more than one standard deviation away from the grand mean, so when comparing them all to each other, the effect size would be large (the effect size is 3.02). Since the effect size is large we can expect that that the power of the overall F test will be large. Example 2.2.1 shows this calculation in SAS.

Example 2.2.1
```
proc power;
      onewayanova test=overall
      groupmeans=(5 4 6 11 12 9 )
      std= 4
      npergroup= 5
      power=.;
run;
```

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

                        Fixed Scenario Elements

                 Method                          Exact
                 Group Means             5 4 6 11 12 9
                 Standard Deviation                  4
                 Sample Size Per Group               5
                 Alpha                            0.05


                           Computed Power

                              Power

                              0.828
```

The power that was calculated is 0.828, a high power as expected.

The same principle of effect size is taken into consideration when we are testing a contrast. If the contrast we are testing has bigger effect size than the one computed for the overall F test then the power will increase. If the contrast has a smaller effect size, then the power will decrease. If it is about the same, then the power will also be approximately the same. Example 2.2.2 illustrates this.

Example 2.2.2
The hypotheses below are the null and alternative hypotheses for the contrasts defined below. They are in the order as specified in the syntax.

$H_0: \mu_1 = \mu_2$ $\quad H_1: \mu_1 \neq \mu_2$
$\quad \mu_2 = \mu_5$ $\qquad \mu_2 \neq \mu_5$
$\quad \mu_3 = \mu_4$ $\qquad \mu_3 \neq \mu_4$

```
proc power;
      onewayanova test=contrast
      contrast = (1 -1 0 0 0 0) (0 1 0 0 -1 0) (0 0 1 -1 0 0)
      groupmeans=(5 4 6 11 12 9)
      std= 4
      npergroup= 5
      power=.;
run;
```

```
                    The POWER Procedure
              Single DF Contrast in One-Way ANOVA

                    Fixed Scenario Elements

        Method                             Exact
        Group Means                5 4 6 11 12 9
        Standard Deviation                     4
        Sample Size Per Group                  5
        Number of Sides                        2
        Null Contrast Value                    0
        Alpha                               0.05


                       Computed Power

        Index     -----------Contrast-----------    Power

          1     1    -1    0    0    0    0    0.067
          2     0     1    0    0   -1    0    0.858
          3     0     0    1   -1    0    0    0.475
```

The first contrast (1 -1 0 0 0 0) is comparing the first two means. We can see that these are very close to each other especially considering that the standard deviation is 4. Thus, this would translate into a very small effect size. Since the effect size is much smaller than with the overall test (the effect size is 0.5), the power should reflect this and be much lower. As we can see the power is 0.067, significantly lower than 0.828.

The second contrast (0 1 0 0 -1 0) is comparing the second and fifth means. These two means are very different, more than a standard deviation apart. Thus this effect size will be large (the effect size is 4), comparable to the effect size of the overall test of 3.02. Thus the two powers should be very similar, and they are 0.858 for the contrast and 0.828 for the overall.

The last contrast (0 0 1 -1 0 0) is comparing the third and fourth means. These means are not that close to each other, a tiny bit over a standard deviation away. Considering this, the effect size should be larger than the first contrast but smaller than the second contrast (the effect size is 2.5), thus the power will be below the overall test but above the power of the first contrast. As you can see it is with a power of 0.475.

This can also be used for finding sample sizes while holding power constant. The same principles hold. If the effect size for the overall test is larger than the effect size for the contrast then the sample size for the overall test will be greater than for that of the test of the contrast. If the effect size for the overall test is smaller than that of the test for the contrast, then the sample size of the contrast will be larger. This is shown in example 2.2.3. Example 2.2.3 uses the same group means and standard deviation as Example 2.2.1 and holds power at 0.8.

Example 2.2.3
```
proc power;
      onewayanova test=overall
      groupmeans=(5 4 6 11 12 9 )
      std= 4
      npergroup= .
      power=.8;
run;

proc power;
      onewayanova test=contrast
      contrast = (1 -1 0 0 0 0) (0 1 0 0 -1 0) (0 0 1 -1 0 0)
      groupmeans=(5 4 6 11 12 9)
      std= 4
      npergroup=.
      power=.8;
run;
```

```
                        Fixed Scenario Elements

                Method                      Exact
                Group Means        5 4 6 11 12 9
                Standard Deviation              4
                Nominal Power                 0.8
                Alpha                        0.05


                        Computed N Per Group

                    Actual     N Per
                    Power      Group

                    0.828        5

                        Fixed Scenario Elements

                Method                      Exact
                Group Means        5 4 6 11 12 9
                Standard Deviation              4
                Nominal Power                 0.8
                Number of Sides                 2
                Null Contrast Value             0
                Alpha                        0.05


                        Computed N Per Group
```

			Contrast				Actual Power	N Per Group
Index								
1	1	-1	0	0	0	0	0.801	252
2	0	1	0	0	-1	0	0.858	5
3	0	0	1	-1	0	0	0.822	11

The different tests will have different or approximately the same power or sample size (depending on what you're interested in) depending on how the effect sizes compare to each other while holding everything constant. Large effect sizes will have larger powers or sample sizes, and smaller effect sizes will have smaller powers or sample sizes.

So when you are powering your study you should keep in mind what you want to investigate. Are you more interested in comparing all the means or just some of them? The answer to the question may dramatically change the answer you get from the analysis.

2.3 Theory of Power for a Chi Squared Test

Chi Squared Power in R
To calculate power for a chi squared test we first need to understand and define the hypotheses. The hypotheses are defined as,

H_0: p_{01}, p_{02}, …, p_{0m} $(\sum_{i=1}^{m} p_{0i} = 1)$

H_1: p_{10}, p_{12}, …, p_{1m} where the proportions are different than the null but also sum to 1.

The chi squared test statistic is defined as,

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

where,

> O_i = an observed frequency
> E_i = an expected (theoretical) frequency, asserted by the null hypothesis
> n = the number of possible outcomes of each event

The distribution of the χ^2-statistic when the null hypothesis is true, follows a central chi square distribution. When it's false it follows a noncentral chi squared distribution with the noncentrality parameter, λ. So essentially power is the probability that the data comes from a noncentral chi squared distribution.

R finds power using this expression,
Power = P(χ^2 (df, λ) $\geq \chi^2_{1-\alpha}$ (df))   **(2.3.1)**

Figure 2.2.2 shows you what the distributions of both the hypotheses look like and will visually show you the power.

Figure 2.2.2
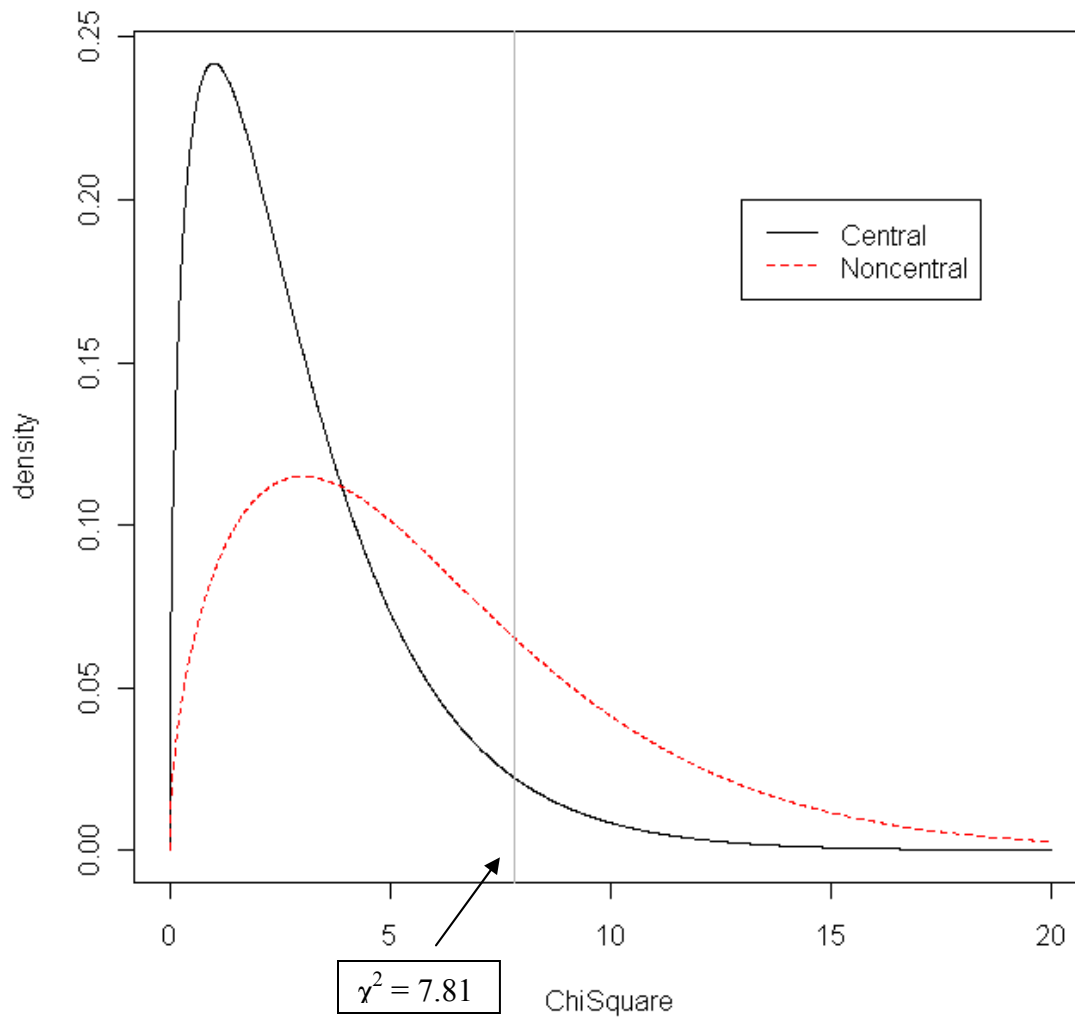
In figure 2.2.2 the solid black curve is the distribution under the null hypothesis (this is arbitrarily chosen to be a chi square distribution with 3 degrees of freedom). The dotted red curve is the distribution under the alternative hypothesis (same chi square with a noncentrality parameter of 3). The vertical line represents the chi square value with a probability of 0.05 under the null hypothesis. If we get a chi square statistic to the left of the line we conclude that it came from the null distribution and if it is on the right of the line we conclude it came from the alternative distribution. The area under the alternative hypothesis curve to the right of the line represents the power of the test. This particular test will have moderate power since there is a fair amount of area under the non central curve.

The noncentrality parameter, λ, is defined as,

$$\lambda = \Delta^2 N \qquad \textbf{(2.3.2)}$$

Δ = effect size
N = total sample size

where,

$$\Delta = \sqrt{\sum_{i=1}^{m} \frac{(p_{1i} - p_{0i})^2}{p_{0i}}}$$

To find the sample size required for a specified power, R finds an N that will satisfy equation 2.3.1.

Chi Square Power in SAS
SAS computes power in a different way. For a chi square test it only will compute the power for two proportions. So the hypotheses for this test are,

$$H_0 \;:\; p_2 - p_1 = p_0$$

$$H_1 \;:\; \begin{cases} p_2 - p_1 \neq p_0, & \text{2-sided} \\ p_2 - p_1 > p_0, & \text{upper 1-sided} \\ p_2 - p_1 < p_0, & \text{lower 1-sided} \end{cases}$$

The chi square test statistic is defined as,

$$z_P = \frac{\hat{p}_2 - \hat{p}_1 - p_0}{\left[\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]^{\frac{1}{2}}} = [N w_1 w_2]^{\frac{1}{2}} \frac{\hat{p}_2 - \hat{p}_1 - p_0}{\hat{p}(1-\hat{p})} \qquad \textbf{(2.3.3)}$$

Using this power is defined as,

$$power - \begin{cases} \Phi\left(\dfrac{(p_2-p_1-p_0)(N w_1 w_2)^{\frac{1}{2}} - z_{1-\alpha}[(w_1 p_1 + w_2 p_2)(1 - w_1 p_1 - w_2 p_2)]^{\frac{1}{2}}}{[w_2 p_1(1-p_1) + w_1 p_2(1-p_2)]^{\frac{1}{2}}}\right), & \text{upper 1-sided} \\[4mm] \Phi\left(\dfrac{-(p_2-p_1-p_0)(N w_1 w_2)^{\frac{1}{2}} - z_{1-\alpha}[(w_1 p_1 + w_2 p_2)(1 - w_1 p_1 - w_2 p_2)]^{\frac{1}{2}}}{[w_2 p_1(1-p_1) + w_1 p_2(1-p_2)]^{\frac{1}{2}}}\right), & \text{lower 1-sided} \\[4mm] \Phi\left(\dfrac{(p_2-p_1-p_0)(N w_1 w_2)^{\frac{1}{2}} - z_{1-\frac{\alpha}{2}}[(w_1 p_1 + w_2 p_2)(1 - w_1 p_1 - w_2 p_2)]^{\frac{1}{2}}}{[w_2 p_1(1-p_1) + w_1 p_2(1-p_2)]^{\frac{1}{2}}}\right) + \\[4mm] \Phi\left(\dfrac{-(p_2-p_1-p_0)(N w_1 w_2)^{\frac{1}{2}} - z_{1-\frac{\alpha}{2}}[(w_1 p_1 + w_2 p_2)(1 - w_1 p_1 - w_2 p_2)]^{\frac{1}{2}}}{[w_2 p_1(1-p_1) + w_1 p_2(1-p_2)]^{\frac{1}{2}}}\right), & \text{2-sided} \end{cases}$$

For the 1-sided cases, a closed-form inversion of the power equation yield an approximate total sample size

$$N = \frac{\left[z_{1-\alpha} \left\{ (w_1 p_1 + w_2 p_2)(1 - w_1 p_1 - w_2 p_2) \right\}^{\frac{1}{2}} + z_{power} \left\{ w_2 p_1 (1 - p_1) + w_1 p_2 (1 - p_2) \right\}^{\frac{1}{2}} \right]^2}{w_1 w_2 (p_2 - p_1 - p_0)^2}$$

For the 2-sided case, the solution for N is obtained by numerically inverting the power equation.

SAS goes through more complicated calculations to find power than R, and it is limited in what it is capable of. SAS only calculates power for two samples while R calculates power for as many as we want.

Examples of finding power and sample size are shown in section 3.2.4.

2.4 More on Effect Size

ANOVA Effect Size
The ANOVA effect size is the degree of departure from having no effect. Thus if there is no effect then all the populations means would be equal. In a case that there is no effect f would equal 0. Thus this effect size, f, can take on values from 0 to an upper limit defined by the nature of the problem.

As mentioned early,

$$\Delta = \frac{\sigma_\mu}{\sigma}$$

and

$$\sigma_\mu = \sqrt{\frac{\sum_{i=1}^{k}(\mu_i - \mu)^2}{k}}$$

The sum of the squared differences in the above equation are the departures of the populations means from the mean of the combined populations or the mean of the means for equal sample sizes. The σ is the standard deviation within the populations. Thus f is the ratio of standard deviation of population means to the standard deviation of the populations. However this assumes that we have a balanced design.

In R an effect size is needed for the calculation, but in SAS that is not one of the options. In SAS the equivalent to an effect size is asking for the group means and the standard deviations. This is all you need to calculate an effect size as seen in the equations above. Essentially SAS does the effect size calculation itself making things a little easier.

Chi Square Effect Size

For a chi square test, the effect size is a value which increases with the degree of discrepancy between the two distributions given by the null and the alternative hypotheses. The effect size, w, measures the discrepancy between the paired proportions of the null and the alternative hypotheses over the cells. As mentioned earlier, w is defined as,

$$\Delta = \sqrt{\sum_{i=1}^{m} \frac{(p_{1i} - p_{0i})^2}{p_{0i}}}$$

As defined the value of w can be 0 when all the paired proportions, in all cells are equal, thus meaning there is no effect and the null hypothesis is true. And just like the ANOVA effect size it can have an upper limit defined by the problem.

3 Using SAS and R for a Power Analysis

Finding power by hand can be quite difficult, as seen in the previous chapter. This chapter will explain how to use SAS or R for a power or sample size analysis.

3.1 Finding Power Using R

To conduct a power or sample size analysis using R the pwr package must be installed and loaded. In this package there are different functions to find power for different tests. For all the power calculations exactly one of the arguments (the one you want to find, most likely power or sample size) has to be left NULL for the calculation to be completed.

Two Proportions With Equal Sample Sizes
Function: pwr.2p.test
Arguments: h: Effect size
               n: Number of observations in sample
               sig.level: Significance level
               power: Power of test
               alternative: Character string specifying the alternative hypothesis (
                                 'two.sided', 'greater', 'less')

Two Proportions With Different Sample Sizes
Function: pwr.2p2n.test
Arguments: h: Effect size
               n1: Number of observations in first sample
               n2: Number of observations in second sample
               sig.level: Significance level
               power: Power of test
               alternative: Character string specifying the alternative hypothesis (
                                 'two.sided', 'greater', 'less')

ANOVA
Function: pwr.anova.test
Arguments: k: Number of groups
               n: Number of observations per group
               f: Effect size
               sig.level: Significance level
               power: Power of test

Chi Square
Function: pwr.chisq.test
Arguments: w: Effect size
               N: Total number of observations

df: Degrees of freedom
sig.level: Significance level
power: Power of test

General Linear Model
Function: pwr.f2.test
Arguments: u: Numerator degrees of freedom
              v: Denominator degrees of freedom
              f2: Effect Size
              sig.level: Significance level
              power: Power of test

Mean of a Normal Distribution
Function: pwr.norm.test
Arguments: d: Effect size ($d = \mu - \mu_0$)
              n: number of observations
              sig.level: Significance level
              power: Power of test
              alternative: Character string specifying the alternative hypothesis (
                     'two.sided', 'greater', 'less')

One Sample Proportion Tests
Function: pwr.p.test
Argumetns: h: Effect size
              n: Number of observations
              sig.level: Significance level
              power: Power of test
              alternative: Character string specifying the alternative hypothesis (
                     'two.sided', 'greater', 'less')

Correlation Test
Function: pwr.r.test
Arguments: n: Number of observations
              r: Correlation coefficient
              sig.level: Significance level
              power: Power of test
              alternative: Character string specifying the alternative hypothesis (
                     'two.sided', 'greater', 'less')

One Sample, Two Sample, or Paired t-test
Function: pwr.t.test
Arguments: n: Sample size
              d: Effect size
              sig.level: Significance level
              power: Power of test

type: Type of t-test ('one.sample', 'two.sample', 'paired.sample')
alternative: Character string specifying the alternative hypothesis (
'two.sided', 'greater', 'less')

Two Sample of Different Sizes t-test
Function: pwr.t2n.test
Arguments: n1: Number of observations in first sample
n2: Number of observations in second sample
d: Effect size
sig.level: Significance level
power: Power of test
alternative: Character string specifying the alternative hypothesis (
'two.sided', 'greater', 'less')

3.2.1 Finding Power Using SAS

In SAS there is a power procedure that can handle power and sample size calculations for many different tests. This procedure is contained in proc power. To find power or sample use one of the options below with a proc power statement.

PROC POWER < options > **;**

MULTREG < options > **;**

ONECORR < options > **;**

ONESAMPLEFREQ < options > **;**

ONESAMPLEMEANS < options > **;**

ONEWAYANOVA < options > **;**

PAIREDFREQ < options > **;**

PAIREDMEANS < options > **;**

TWOSAMPLEFREQ < options > **;**

TWOSAMPLEMEANS < options > **;**

TWOSAMPLESURVIVAL < options > **;**

One or More Coefficients in Multiple Linear Regression
Statement: MULTREG

Fisher's z Test or t Test of Correlation
Statement: ONECORR

Single Binomial Proportion
Statement: ONESAMPLEFREQ

One-Sample t Test, Confidence Interval Precision, or Equivalence Test
Statement: ONESAMPLEMEANS

One-Way ANOVA Including Single-Degree-of-Freedom Contrasts
Statement: ONEWAYANOVA
McNemar's test for Paired Proportions
Statement: PAIREDFREQ

Paired t Test, Confidence Interval Precision, or Equivalence Test
Statement: PAIREDMEANS

Chi-Square, Likelihood Ratio, and Fisher's Exact Tests for Two Independent Proportions
Statement: TWOSAMPLEFREQ

Two Sample t Test, Confidence Interval Precision, or Equivalence Test
Statement: TWOSAMPLEMEANS

Log-Rank, Gehan, and Tarone-Ware Tests for Comparing Two Survival Curves
Statement: TWOSAMPLESURVIVAL
Table 3.2.1 shows a more detailed summary of the possible analyses and the options that can be used using proc power.

Table 3.2.1

Statement	Options	
Multiple linear regression: Type III F test	MULTREG	
Correlation: Fisher's z test	ONECORR	DIST=FISHERZ
Correlation: t test	ONECORR	DIST=T
Binomial proportion: Exact test	ONESAMPLEFREQ	TEST=EXACT
Binomial proportion: z test	ONESAMPLEFREQ	TEST=Z
Binomial proportion: z test with continuity adjustment	ONESAMPLEFREQ	TEST=ADJZ
One-sample t test	ONESAMPLEMEANS	TEST=T
One-sample t test with lognormal data	ONESAMPLEMEANS	TEST=T DIST=LOGNORMAL
One-sample equivalence test for mean of normal data	ONESAMPLEMEANS	TEST=EQUIV

One-sample equivalence test for mean of lognormal data	ONESAMPLEMEANS	TEST=EQUIV DIST=LOGNORMAL
Confidence interval for a mean	ONESAMPLEMEANS	CI=T
One-way ANOVA: One-degree-of-freedom contrast	ONEWAYANOVA	TEST=CONTRAST
One-way ANOVA: Overall F test	ONEWAYANOVA	TEST=OVERALL
McNemar exact conditional test	PAIREDFREQ	
McNemar normal approximation test	PAIREDFREQ	DIST=NORMAL
Paired t test	PAIREDMEANS	TEST=DIFF
Paired t test of mean ratio with lognormal data	PAIREDMEANS	TEST=RATIO
Paired additive equivalence of mean difference with normal data	PAIREDMEANS	TEST=EQUIV_DIFF
Paired multiplicative equivalence of mean ratio with lognormal data	PAIREDMEANS	TEST=EQUIV_RATIO
Confidence interval for mean of paired differences	PAIREDMEANS	CI=DIFF
Pearson chi-square test for two independent proportions	TWOSAMPLEFREQ	TEST=PCHI
Fisher's exact test for two independent proportions	TWOSAMPLEFREQ	TEST=FISHER
Likelihood ratio chi-square test for two independent proportions	TWOSAMPLEFREQ	TEST=LRCHI
Two-sample t test assuming equal variances	TWOSAMPLEMEANS	TEST=DIFF

Two-sample Satterthwaite *t* test assuming unequal variances	TWOSAMPLEMEANS	TEST=DIFF_SATT
Two-sample pooled *t* test of mean ratio with lognormal data	TWOSAMPLEMEANS	TEST=RATIO
Two-sample additive equivalence of mean difference with normal data	TWOSAMPLEMEANS	TEST=EQUIV_DIFF
Two-sample multiplicative equivalence of mean ratio with lognormal data	TWOSAMPLEMEANS	TEST=EQUIV_RATIO
Two-sample confidence interval for mean difference	TWOSAMPLEMEANS	CI=DIFF
Log-rank test for comparing two survival curves	TWOSAMPLESURVIVAL	TEST=LOGRANK
Gehan rank test for comparing two survival curves	TWOSAMPLESURVIVAL	TEST=GEHAN
Tarone-Ware rank test for comparing two survival curves	TWOSAMPLESURVIVAL	TEST=TARONEWARE

3.2.2 Details of ANOVA in SAS

The ONEWAYANOVA statement is used for power and sample size analysis. It can run analysis for one degree of freedom contrasts and for the overall F-test in a one-way ANOVA. It can also handle a balanced and an unbalanced design.

To specify an unbalanced design the GROUPWEIGHTS or GROUPNS can be specified.

To specify a balanced design you can leave the NFRACTIONAL options null then the default of this will be 1 resulting in a balanced design. Also using the NPERGROUP implicitly specifies a balanced design.

ALPHA=
Specifies a significance level for the statistical test (default is .05).

CONTRAST=
Specifies coefficients for the different contrasts you want to test. You must have a coefficient for every mean in the GROUPMEANS= option. You can specify multiple with additional set of coefficients or with another contrast statement.

GROUPMEANS= or GMEANS=
Specifies group sample means.

GROUPNS= or GNS=
Specifies group sample sizes.

GROUPWEIGHTS= or GWEIGHTS=
Specifies group weights for sample size allocation. This controls how the sample size is divided into each of the groups. If the NFRACTIONAL option is not used, the total sample size has to be equal to a multiple of the sum of the group weights and has to be integer values. Also the number of groups must be the same as with the GROUPMEANS= option.

NFRACTIONAL or NFRAC
Allows for fractional input and output for sample sizes.

NPERGROUP= or NPERG=
Specifies a common sample size per group or requests a solution for a common sample per group with a missing value (NPERG=.).

NTOTAL=
Specifies the sample size or requests a solution for the sample size with a missing value (NTOTAL=.).

NULLCONTRAST= or NULLC
Specifies the null value of the contrast, default being 0. This can only be used with the TEST=CONTRAST analysis.

OUTPUTORDER=
Controls how input and default analysis are ordered in output.

POWER=
Specifies power or requests a solution for power (POWER=.).

SIDES=
Specifies direction of test, one tailed, two tailed, upper, lower.
STDEV= or STD=
Specifies the standard deviation.
TEST=CONTRAST or TEST=OVERALL
Specifies type of test.

Table 3.2.2 Summary of Options in ONEWAYANOVA

Task	Options
Define analysis	TEST=
Specify analysis information	ALPHA=
	CONTRAST=
	SIDES=
	NULLCONTRAST=
Specify effects	GROUPMEANS=
Specify variability	STDDEV=
Specify sample size and allocation	GROUPNS=
	GROUPWEIGHTS=
	NPERGROUP=
	NTOTAL=
Specify power	POWER=
Control sample size rounding	NFRACTIONAL
Control ordering in output	OUTPUTORDER=

3.2.3 Details of Chi Squared in SAS

The TWOSAMPLEFREQ statement performs power and sample size analyses for tests of two independent proportions. Pearson's chi-square, Fisher's exact, and likelihood ratio chi-squared tests can be calculated.

ALPHA=
Specifies the level of significance of the statistical test. The default is 0.05.

GROUPPROPORTIONS= or GPROPORTIONS= or GROUPPS= or GPS=
Specifies the two independent proportions, p_1 and p_2.

GROUPNS= or GNS=
Specifies the two group sample sizes or requests a solution for one group sample size given the other.

GROUPWEIGHTS= or GWEIGHTS=
Specifies the sample size allocation weights for the two groups, or requests a solution for one group weight given the other. This option controls how the total sample size is divided between the two groups. Each pair of values for the two groups represents relative allocation weights.

NFRACTIONAL or NFRAC
Enables fractional input and output for sample sizes.

NPERGROUP= or NPERG=
Specifies the common sample size per group or requests a solution for the common sample size per group with a missing value (NPERGROUP=.).

NTOTAL=
Specifies the sample size or requests a solution for the sample size with a missing value (NTOTAL=.).

NULLODDSRATIO= or NULLOR=
Specifies the null odds ratio. The default value is 1. This option can only be used along with the ODDSRATIO= option in the TEST=PCHI analysis.

NULLPROPORTIONDIFF= or NULLPDIFF=
Specifies the null proportion difference. The default value is 0. This option can only be used along with the GROUPPROPORTIONS= or PROPORTIONDIFF= option in the TEST=PCHI analysis.

NULLRELATIVERISK= or NULLRR=
Specifies the null relative risk. The default value is 1. This option can only be used along with the RELATIVERISK= option in the TEST=PCHI analysis

ODDSRATIO= or OR=
Specifies the odds ratio.

OUTPUTORDER= or OUTPUTORDER= or OUTPUTORDER=
Controls how the input and default analysis parameters are ordered in the output. OUTPUTORDER=INTERNAL (the default) produces output sorted respectively by

The OUTPUTORDER=SYNTAX option arranges the parameters in the output in the same order that their corresponding options are specified in the TWOSAMPLEFREQ statement. The OUTPUTORDER=REVERSE option arranges the parameters in the output in the reverse of the order that their corresponding options are specified in the TWOSAMPLEFREQ statement.

POWER=
Specifies the desired power of the test or requests a solution for the power with a missing value (POWER=.).

PROPORTIONDIFF=number-list or PDIFF=number-list
Specifies the proportion difference $p_2 - p_1$.

REFPROPORTION= or REFP=
Specifies the reference proportion p_1.

RELATIVERISK= or RR=

Specifies the relative risk p_2 / p_1

SIDES=

specifies the number of sides (or tails) and direction of the statistical test or confidence interval.

> 1
> 1-sided with alternative hypothesis in same direction as effect
>
> 2
> 2-sided
>
> U
> upper 1-sided with alternative greater than null value
>
> L
> lower 1-sided with alternative less than null value
>
> The default value is 2.

TEST=FISHER or TEST=LRCHI or TEST=PCHI

Specifies the statistical analysis. TEST=FISHER specifies Fisher's exact test. TEST=LRCHI specifies the likelihood ratio chi-square test. TEST=PCHI (the default) specifies Pearson's chi-square test.

Table 3.2.3 Summary of Options in TWOSAMPLEFREQ

Task	Options
Define analysis	TEST=
Specify analysis information	ALPHA=
	NULLPROPORTIONDIFF=
	NULLODDSRATIO=
	NULLRELATIVERISK=
	SIDES=
Specify effects	GROUPPROPORTIONS=
	ODDSRATIO=
	PROPORTIONDIFF=
	REFPROPORTION=
	RELATIVERISK=
Specify sample size and allocation	GROUPNS=
	GROUPWEIGHTS=

	NPERGROUP=
	NTOTAL=
Specify power	POWER=
Control sample size rounding	NFRACTIONAL
Control ordering in output	OUTPUTORDER=

3.2.4 Examples of Power Analyses for ANOVA and Chi Squared

This section will go over examples for different types of ANOVA and Chi Square power analysis in both R and SAS.

Example 3.2.1 Finding power of a balanced one way ANOVA

Suppose we have,

$\mu_1 = 59$ $\sigma = 12$ $k = 3$
$\mu_2 = 66$ $\alpha = .05$
$\mu_3 = 42$ $n = 4$

Using this information we have:
$f = .84$

R Input/Output:
> pwr.anova.test(f=.84,k=3,n=4,sig.level=0.05)

Balanced one-way analysis of variance power calculation

k = 3
n = 4
f = 0.84
sig.level = 0.05
power = 0.5848498

NOTE: n is number in each group

SAS Input/Output

```
proc power;
      onewayanova test=overall
      groupmeans=59|66|42
      std=12
      npergroup=4
      power=.
      ;
run;
```

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

                      Fixed Scenario Elements

               Method                        Exact
               Group Means               59 66 42
               Standard Deviation              12
               Sample Size Per Group            4
               Alpha                         0.05


                         Computed Power

                            Power

                            0.585
```

Example 3.2.2 Finding power of a balanced one way ANOVA

Suppose we have,

$\mu_1 = 2.24$ $\sigma = .3$
$\mu_2 = 2.20$ $\alpha = .05$
$\mu_3 = 2.29$ $n = 10$
$\mu_4 = 2.34$ $k = 5$
$\mu_5 = 2.19$

Using the above information we have:
$f = .187$

R Input/Output
> pwr.anova.test(f=.187,k=5,n=10,sig.level=0.05)

    Balanced one-way analysis of variance power calculation

       k = 5
       n = 10
       f = 0.187
   sig.level = 0.05
     power = 0.1425926

 NOTE: n is number in each group

SAS Input/Output
proc power;
```
      onewayanova test=overall
      groupmeans=2.24|2.20|2.29|2.34|2.19
      std=.3
      npergroup=10
      power=.
      ;
```
run;

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

                  Fixed Scenario Elements

          Method                                Exact
          Group Means          2.24 2.2 2.29 2.34 2.19
          Standard Deviation                      0.3
          Sample Size Per Group                    10
          Alpha                                  0.05


                      Computed Power

                          Power

                          0.144
```

Example 3.2.3 Finding n of a balanced one way ANOVA
Suppose,
$\mu_1 = 15$ $\sigma = 1.5$ $k = 3$
$\mu_2 = 14$ $\alpha = .05$
$\mu_3 = 18$ power = .8

Using this information we have:
f = 1.13

R Input/Output:
> pwr.anova.test(f=1.13,k=3,power=0.80,sig.level=0.05)

   Balanced one-way analysis of variance power calculation

      k = 3
      n = 3.718485
      f = 1.13
   sig.level = 0.05
     power = 0.8

 NOTE: n is number in each group

SAS Input/Output:
```
proc power;
     onewayanova test=overall
     groupmeans=15|14|18
     std=1.5
     npergroup=.
     power=.80
     ;
run;
```

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

                   Fixed Scenario Elements

         Method                          Exact
         Group Means                  15 14 18
         Standard Deviation                1.5
         Nominal Power                     0.8
         Alpha                            0.05


                   Computed N Per Group

                  Actual    N Per
                  Power     Group
                  0.846       4
```

Example 3.2.4 Finding n of a balanced one way ANOVA

$\mu_1 = 23$ $\sigma = 5$
$\mu_2 = 32$ $\alpha = .05$
$\mu_3 = 25$ power = .85
$\mu_4 = 29$ $k = 5$
$\mu_5 = 26$

Using the above information we have:
f = .632

R Input/Output:
> pwr.anova.test(f=.632,k=5,power=0.85,sig.level=0.05)

Balanced one-way analysis of variance power calculation

      k = 5
      n = 7.73407
      f = 0.632
  sig.level = 0.05
    power = 0.85

 NOTE: n is number in each group

SAS Input/Output
```
proc power;
     onewayanova test=overall
     groupmeans=23|32|25|29|26
     std=5
     npergroup=.
     power=.85
     ;
run;
```

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


                 Fixed Scenario Elements


    Method                               Exact
    Group Means              23 32 25 29 26
    Standard Deviation                       5
    Nominal Power                         0.85
    Alpha                                 0.05



                 Computed N Per Group

                  Actual     N Per
                  Power      Group

                  0.866         8
```

As you can see in the previous four examples both SAS and R came up with the same results.

Example 3.2.5 Finding n of an unbalanced one way ANOVA

<u>SAS Input/Output</u>

```
proc power;
   onewayanova test=overall
      groupmeans = 3 | 7 | 8
      stddev = 4
      groupweights = (1 2 2)
      ntotal = .
      power = 0.8;
run;
```

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


                 Fixed Scenario Elements


    Method                               Exact
    Group Means                          3 7 8
    Standard Deviation                       4
    Group Weights                        1 2 2
    Nominal Power                          0.8
    Alpha                                 0.05



                 Computed N Total

                  Actual        N
                  Power       Total

                  0.819         50
```

3.3 Overview of Plotting Power Curves in SAS

Power curves are visual tools that compare power for different parameters. For example a power curve can visually show the power for a variety of different sample sizes. This makes it easy to see approximately how big the sample size has to be in order to obtain a specified power. It can also do the same thing using standard deviations and so on.

SAS has the capability of producing these curves and can make them several different ways. In order to produce these curves using SAS, all that needs to be done is an addition of a plot statement at the end of the code where the power analysis is ran. This can be done for all the different tests SAS offers. Details and examples of plotting power for ANOVA and chi square are shown in the next sections.

There are many different options you can choose when using a plot statement. These options are discussed in section 3.4.

ANOVA Power Curves

To produce a power curve for an ANOVA test comparing sample size versus power all that is needed is the code specified in section 3.2.2 and a plot statement. However the only difference is that in the npergroup statement one needs to specify more than one number for it to be possible to create a curve. The more sample sizes that are specified the better the fit of the curve.

Example 3.3.1 shows you a power curve of sample size versus power. If you want to switch the axis all you would need to do is use the y= option. Example 3.3.2 shows you this.

SAS can also plot power and sample size for different standard deviations. All you would need to do is in addition to specifying multiple sample sizes is to specify different standard deviations. This is shown in Example 3.3.3.

One thing SAS does not plot for the ANOVA power curves is the effect size. However it will produce curves for different group means. So the imputed group means must be used to compute the effect sizes. Example 3.3.4 shows this in addition to the different standard deviations.

It is also possible to produce these curves for different contrasts by specifying the contrasts you want to test. Like the overall test, SAS can also do this using different standard deviations and different means. Example 3.3.5 is a simple example of power using contrasts with just one standard deviation and one set of group means.

Not only does SAS produce these curves, in the output it produces a table of all the powers for the different combinations of the parameters specified. Table 3.3.1 shows a partial table for the combinations of standard deviations, group means, and sample sizes using the specifications of example 3.3.4.

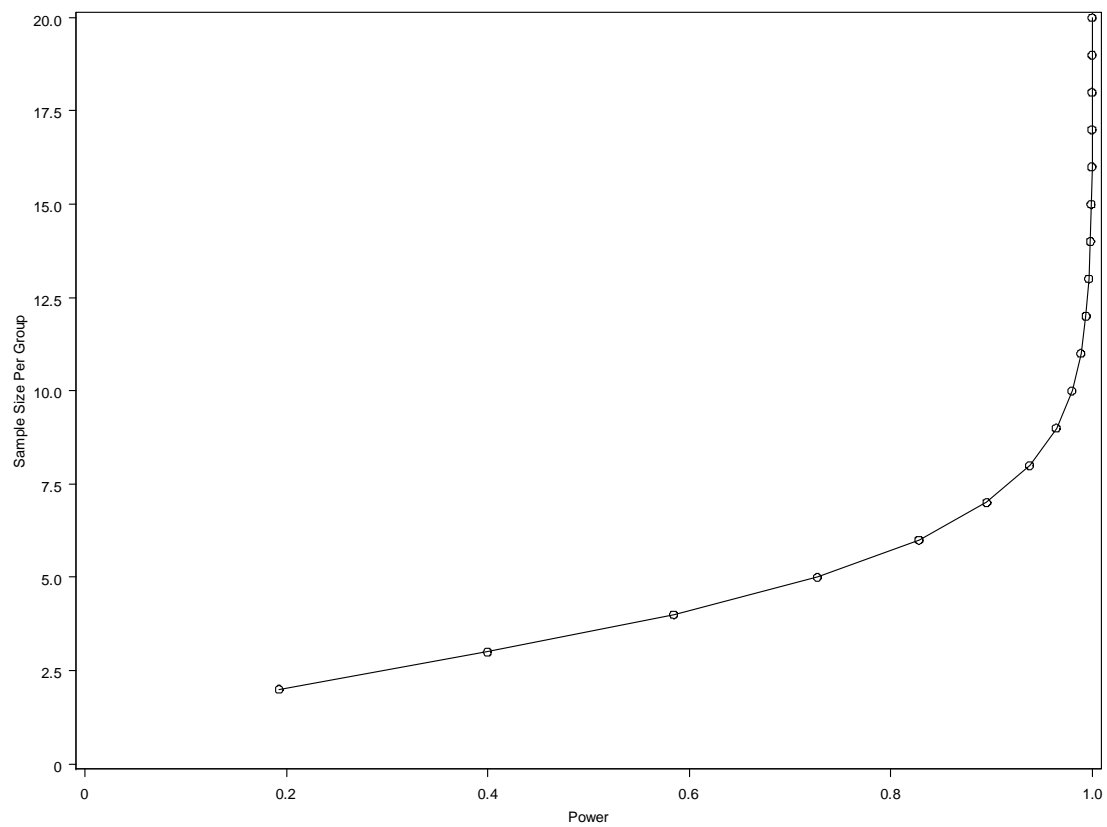*Example 3.3.1*
```
proc power;
      onewayanova test=overall
      groupmeans=59|66|42
      std=12
      npergroup=2 5 7 10 15 20
      power=.;
      plot;
run;
```
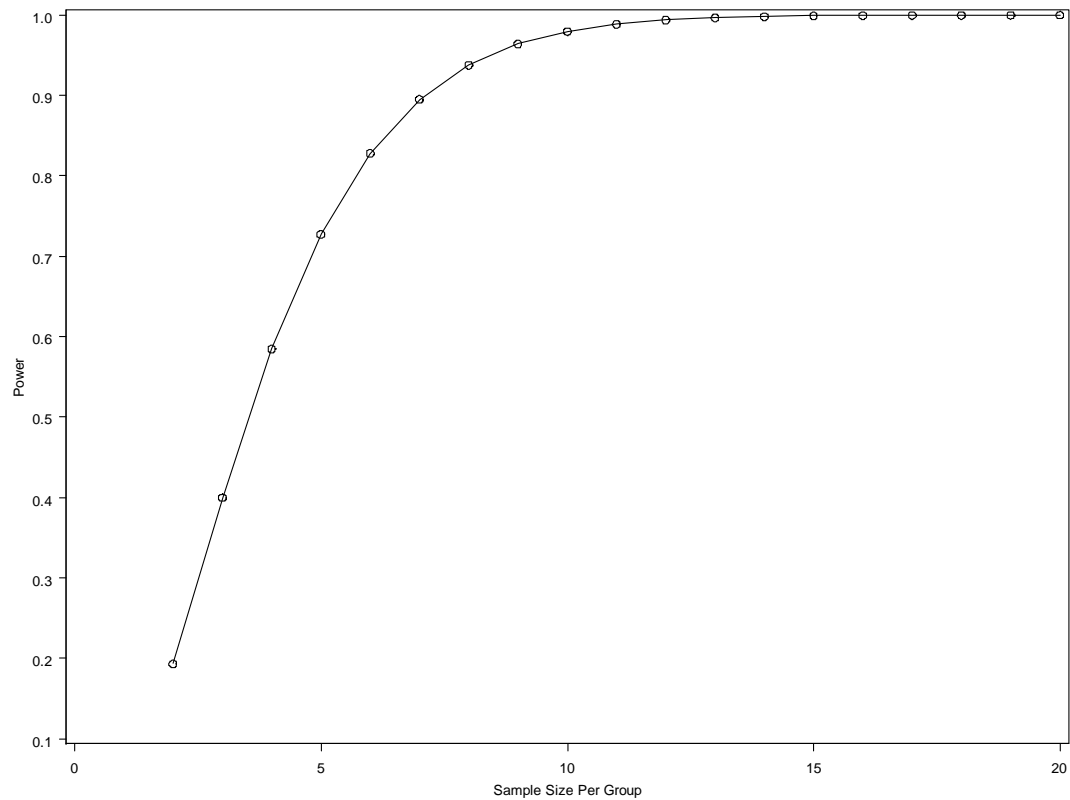
Example 3.3.2
```
proc power;
      onewayanova test=overall
      groupmeans=59|66|42
      std=12
      npergroup= 2 5 7 10 15 20
      power=.;
      plot y=n;
run;
```
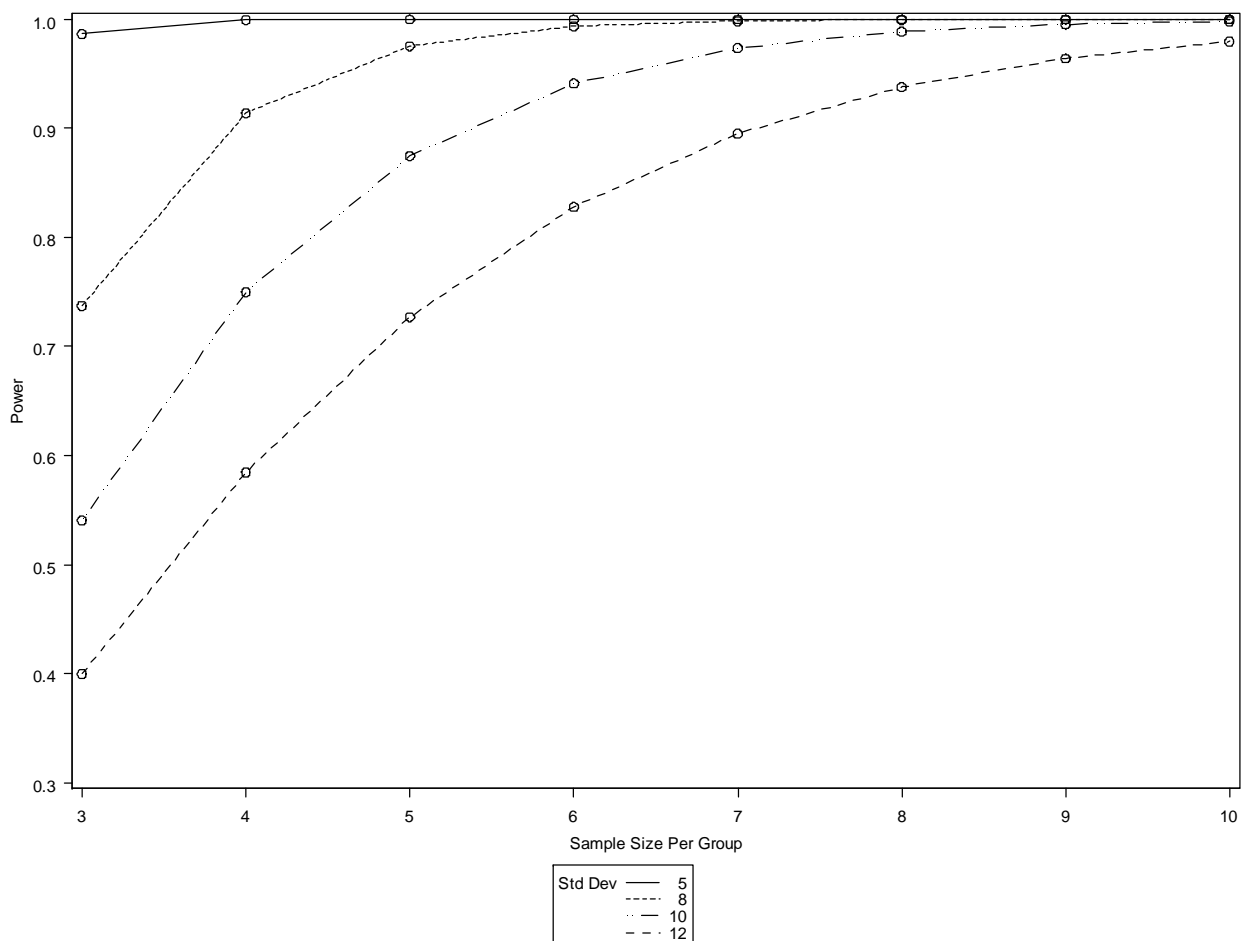
Example 3.3.3
```
proc power;
      onewayanova test=overall
      groupmeans=59|66|42
      std= 5 8 10 12
      npergroup= 3 4 5 6 7 8 9 10
      power=.;
      plot y=n;
run;
```

Example 3.3.4
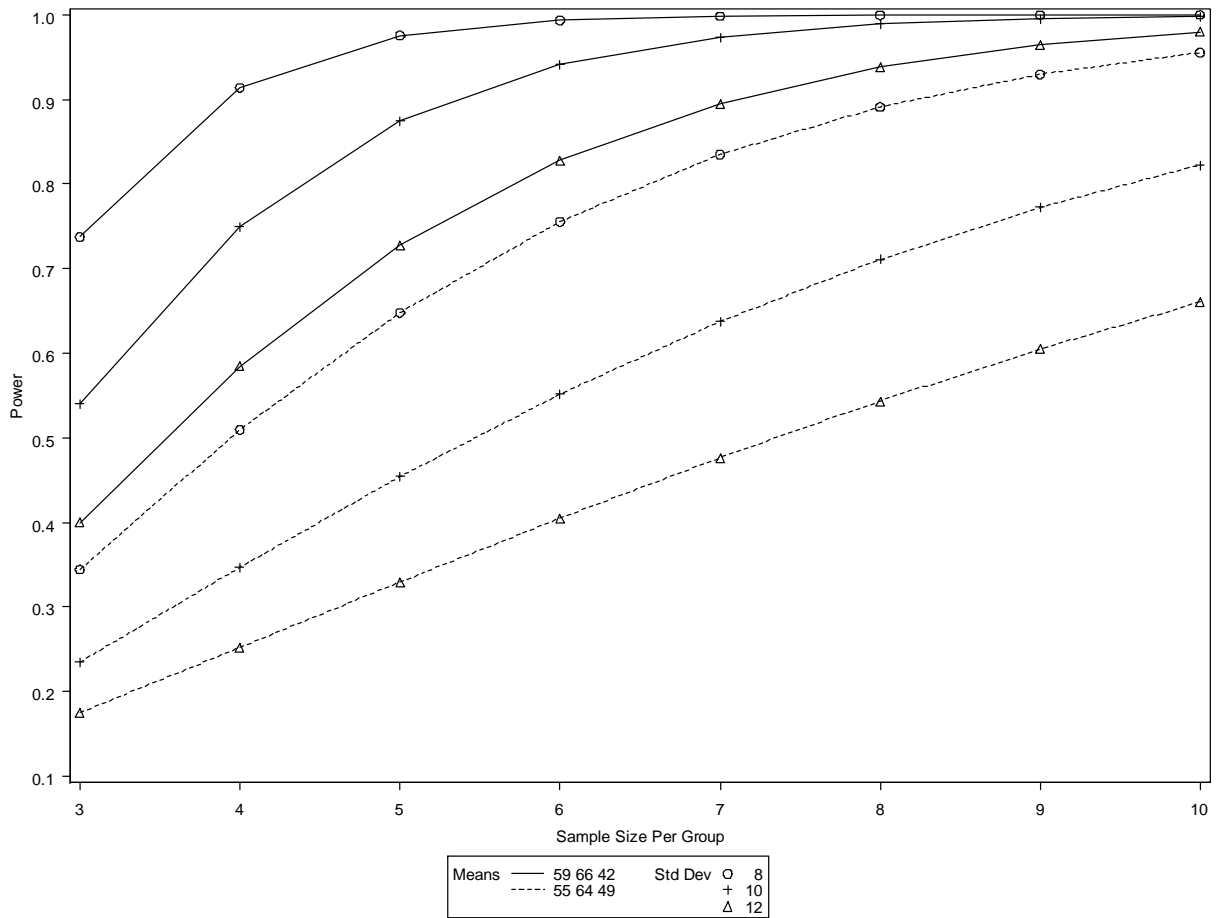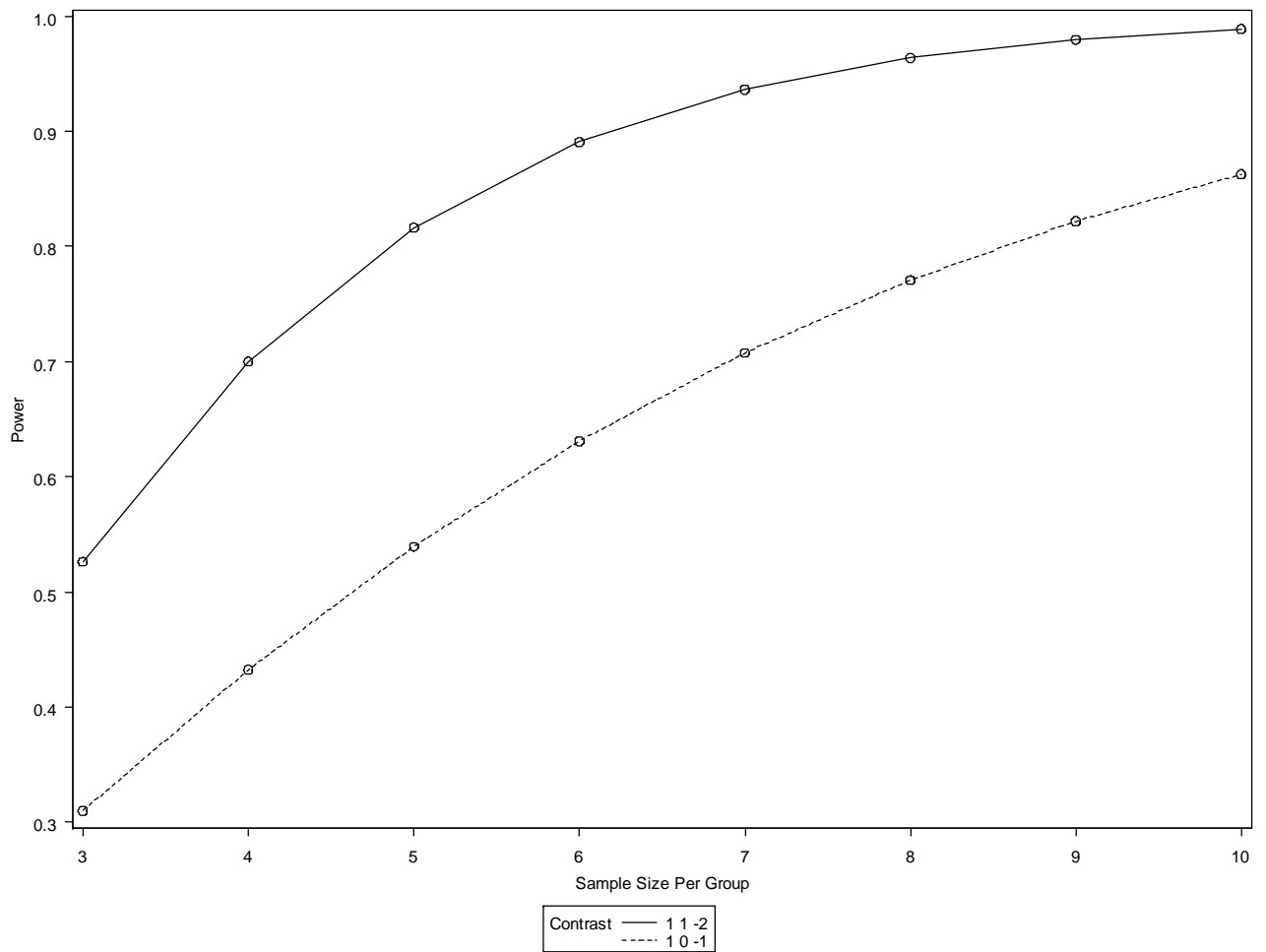```
proc power;
      onewayanova test=overall
      groupmeans=(59 66 42) (55 64 49)
      std= 8 10 12
      npergroup= 3 4 5 6 7 8 9 10
      power=.;
      plot y=n;
run;
```

*Example 3.3.5*
```
proc power;
      onewayanova test=contrast
      contrast = (1 1 -2) (1 0 -1)
      groupmeans=(59 66 42)
      std= 12
      npergroup= 3 4 5 6 7 8 9 10
      power=.;
      plot y=n;
run;
```

## Table 3.3.1 Power For Different Combinations of Arguments For ANOVA

```
          Computed Power


                                   Std    N Per
         Index    ------Means------ Dev   Group   Power

           1      59    66    42     8      3     0.737
           2      59    66    42     8      4     0.914
           3      59    66    42     8      5     0.976
           4      59    66    42     8      6     0.994
           5      59    66    42     8      7     0.998
           6      59    66    42     8      8     >.999
           7      59    66    42     8      9     >.999
           8      59    66    42     8     10     >.999
           9      59    66    42    10      3     0.541
          10      59    66    42    10      4     0.750
          11      59    66    42    10      5     0.875
          12      59    66    42    10      6     0.941
          13      59    66    42    10      7     0.974
          14      59    66    42    10      8     0.989
          15      59    66    42    10      9     0.995
          16      59    66    42    10     10     0.998
          17      59    66    42    12      3     0.400
          18      59    66    42    12      4     0.585
          19      59    66    42    12      5     0.727
          20      59    66    42    12      6     0.828
          21      59    66    42    12      7     0.895
          22      59    66    42    12      8     0.938
          23      59    66    42    12      9     0.964
          24      59    66    42    12     10     0.980
          25      55    64    49     8      3     0.345
          26      55    64    49     8      4     0.510
          27      55    64    49     8      5     0.648
          28      55    64    49     8      6     0.756
          29      55    64    49     8      7     0.835
          30      55    64    49     8      8     0.891
          31      55    64    49     8      9     0.930
          32      55    64    49     8     10     0.955
          33      55    64    49    10      3     0.235
          34      55    64    49    10      4     0.347
          35      55    64    49    10      5     0.454
          36      55    64    49    10      6     0.552
          37      55    64    49    10      7     0.638
          38      55    64    49    10      8     0.711
          39      55    64    49    10      9     0.772
```

Chi Square Power Curves

Just like plotting power curves using ANOVA, chi squared power curves can be plotted in the same way, by using the plot statement.

Example 3.3.6 plots group sample size versus power. Again one needs to specify more than one group sample size to produce a curve, however the more sample sizes that are used the better the curve will fit.

For a Pearson's chi square test SAS is not able to produce a power curve using effect size, to get around this you can specify different group proportions and calculate the effect size by hand. Example 3.3.7 shows this.

One can also specify different null hypothesis proportions in addition to different group means proportions. This is shown in Example 3.3.8.

Just like the ANOVA test, tables for power of the different combinations of the parameters will be produced in the output. An example table is shown in Table 3.3.2 for example 3.3.8.
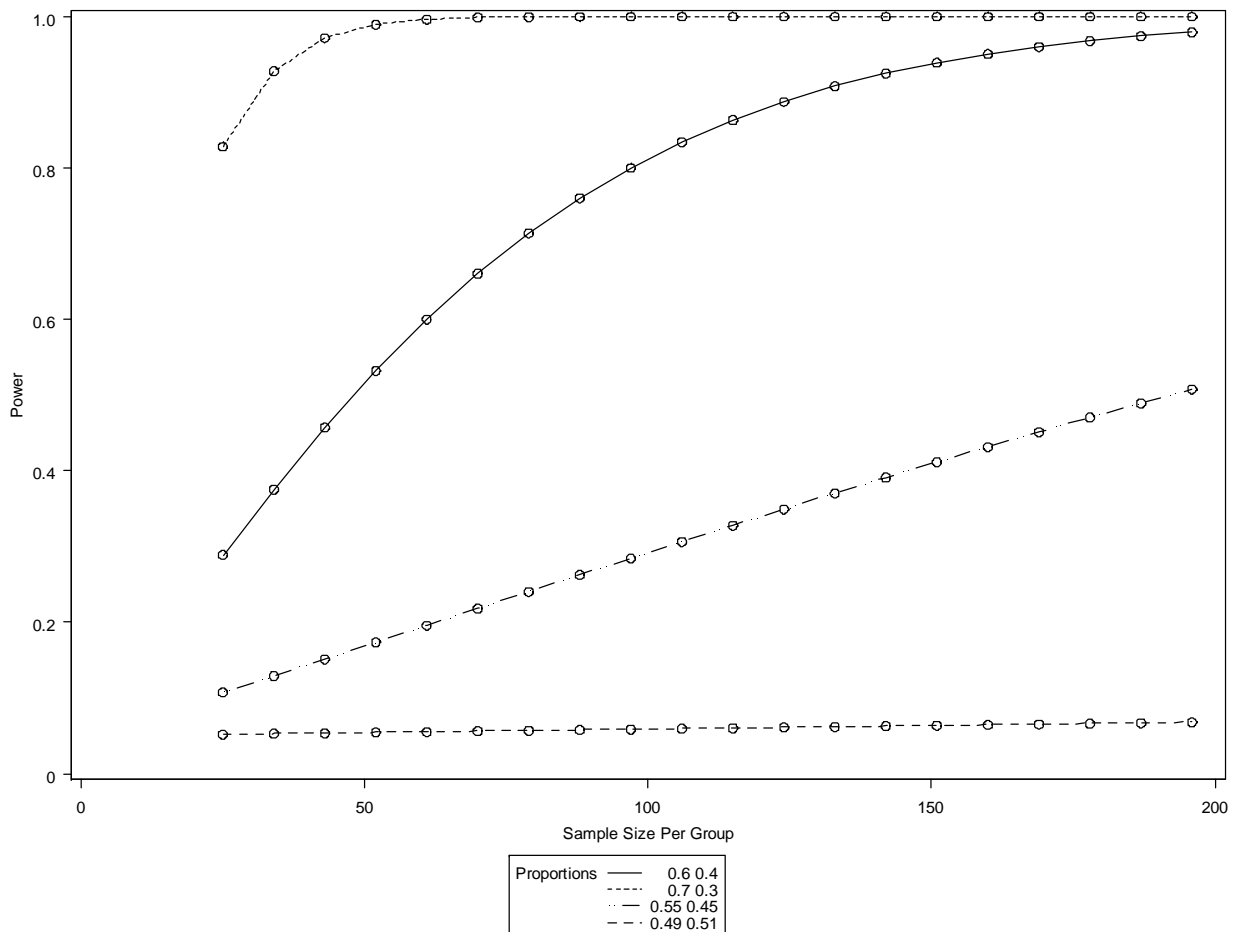
Example 3.3.6
```
proc power;
      twosamplefreq test=pchi
      groupproportions = (.6 .4)
      nullproportiondiff = 0
      npergroup = 25 50 75 100 200
      power = .;
        plot;
 run;
```
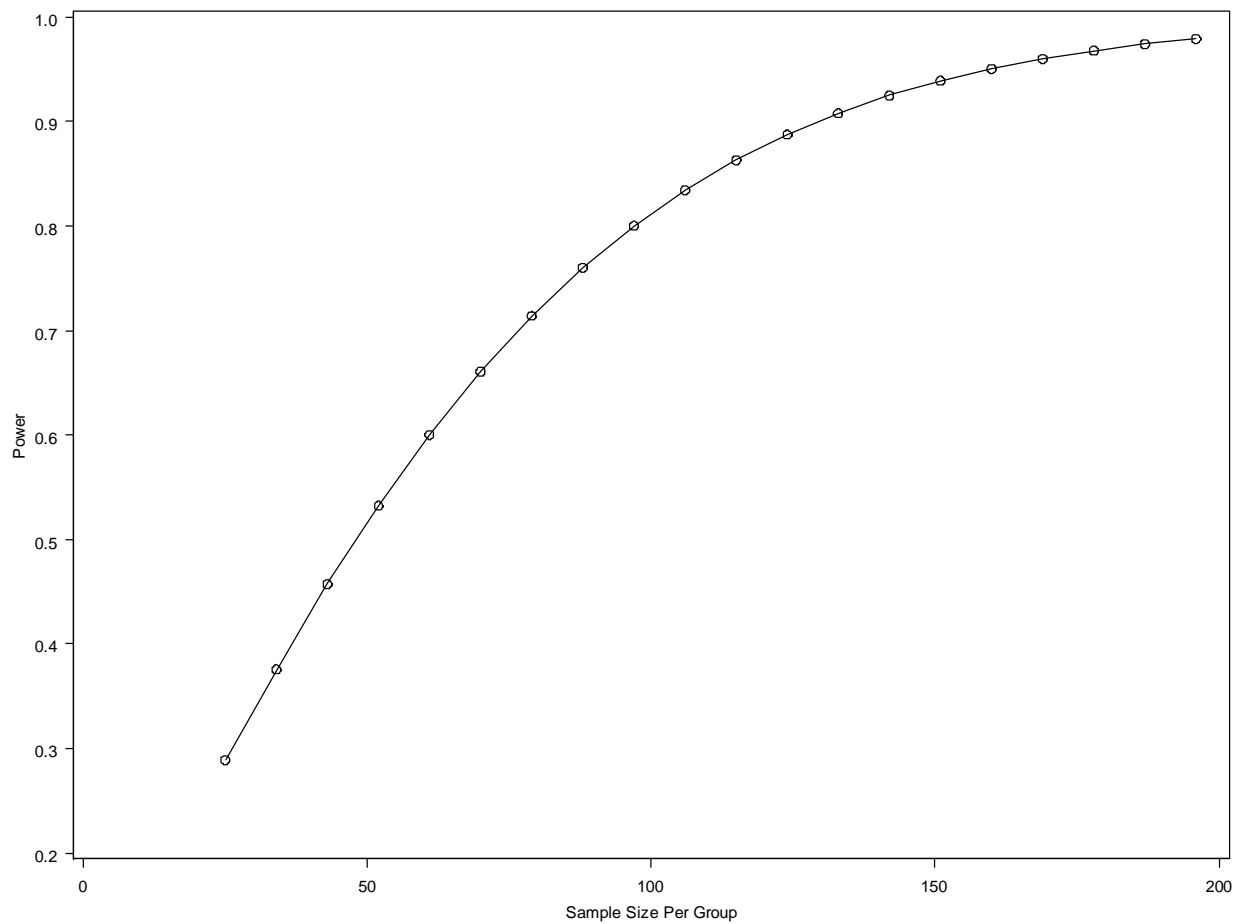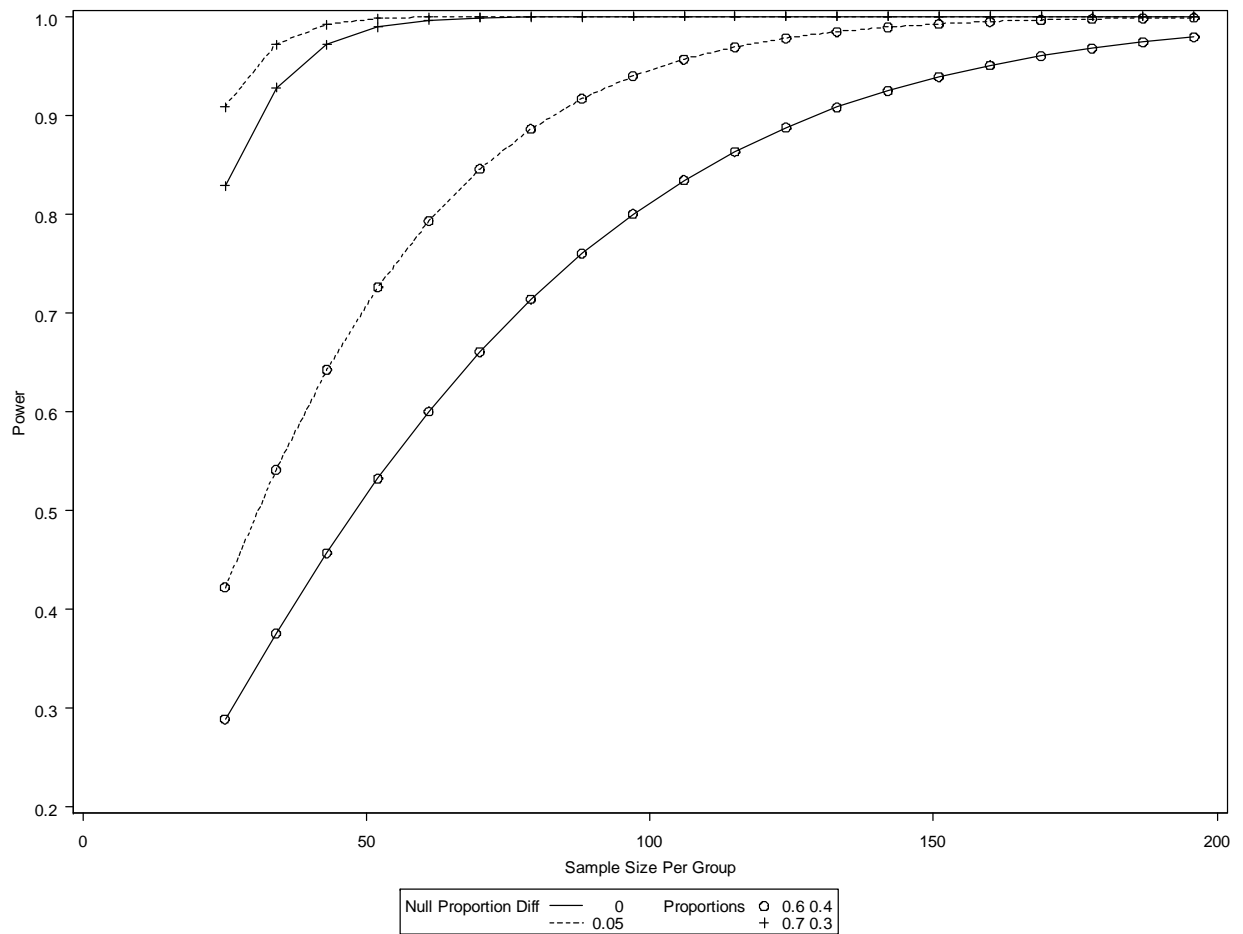
Example 3.3.7
```
proc power;
      twosamplefreq test=pchi
      groupproportions = (.6 .4) (.7 .3) (.55 .45) (.49 .51)
      nullproportiondiff = 0
      npergroup = 25 50 75 100 200
      power = .;
        plot;
  run;
```

Example 3.3.8
```
proc power;
      twosamplefreq test=pchi
      groupproportions = (.6 .4) (.7 .3)
      nullproportiondiff = 0 .05
      npergroup = 25 50 75 100 200
      power = .;
        plot;
  run;
```

Table 3.3.2 Power For Different Combinations of Arguments For Chi Square

```
                              Computed Power

                 Null
               Proportion                               N Per
     Index       Diff     Proportion1    Proportion2    Group     Power

       1         0.00         0.6            0.4          25       0.289
       2         0.00         0.6            0.4          50       0.516
       3         0.00         0.6            0.4          75       0.691
       4         0.00         0.6            0.4          100      0.812
       5         0.00         0.6            0.4          200      0.981
       6         0.00         0.7            0.3          25       0.828
       7         0.00         0.7            0.3          50       0.987
       8         0.00         0.7            0.3          75       >.999
       9         0.00         0.7            0.3          100      >.999
      10         0.00         0.7            0.3          200      >.999
      11         0.05         0.6            0.4          25       0.422
      12         0.05         0.6            0.4          50       0.709
      13         0.05         0.6            0.4          75       0.870
      14         0.05         0.6            0.4          100      0.946
      15         0.05         0.6            0.4          200      >.999
      16         0.05         0.7            0.3          25       0.909
      17         0.05         0.7            0.3          50       0.997
      18         0.05         0.7            0.3          75       >.999
      19         0.05         0.7            0.3          100      >.999
      20         0.05         0.7            0.3          200      >.999
```

3.4 Plot Options in SAS

You can specify the following plot options in the PLOT statement.

INTERPOL=JOIN or INTERPOL=NONE
Specifies the type of curve to draw through the computed points.

KEY= BYCURVE or KEY= BYFEATURE or KEY= ONCURVES
Specifies the style of key for the plot. The default is KEY=BYFEATURE. Each entry shows the mapping between a value of the feature and the value(s) of the analysis parameter(s) linked to that feature

MARKERS=ANALYSIS or MARKERS=COMPUTED or MARKERS=NICE or MARKERS=NONE
Specifies the locations for plotting symbols.

MAX=
Specifies the maximum of the range of values for the parameter associated with the "argument" axis (the axis that is *not* representing the parameter being solved for).

MIN=
Specifies the minimum of the range of values for the parameter associated with the "argument" axis (the axis that is *not* representing the parameter being solved for).

NPOINTS= or NPTS=
Specifies the number of values for the parameter associated with the "argument" axis. You cannot use the NPOINTS= and STEP= options simultaneously.

STEP=
Specifies the increment between values of the parameter associated with the "argument" axis. You cannot use the STEP= and NPOINTS= options simultaneously.

VARY (*feature < BY parameter-list > ... feature < BY parameter-list >* **)**
Specifies how plot features should be linked to varying analysis parameters. Available plot *features* are COLOR, LINESTYLE, PANEL, and SYMBOL. A "panel" refers to a separate plot with a heading identifying the subset of values represented in the plot.

X=EFFECT or X=N or X=POWER
Specifies a plot with the requested type of parameter on the x-axis and the parameter being solved for on the y-axis. When X=EFFECT, the parameter assigned to the x-axis is the one most representative of "effect size." When X=N, the parameter assigned to the x-axis is the sample size. When X=POWER, the parameter assigned to the x-axis is the one most representative of "power" You cannot use the X= and Y= options simultaneously.

You can only use the X=N option when a scalar sample size parameter is used as input in the analysis. For example, X=N can be used with total sample size or sample size per group, or with two group sample sizes when one is being solved for.

3.5 Advantages and Disadvantages of SAS and R
SAS and R both have their advantages and disadvantages when it comes to computing power. They both have different inputs and what they can do for different tests. Thus it's an issue of the right software for the right problem. What one wants to do and what information one knows will guide the decision of what to use, SAS or R.

One Way ANOVA Power Analysis
R is limited in what it can do when doing this kinds of power analyis. The only type of design that R can find power for is a balanced one way ANOVA design. It cannot handle an unbalanced design which is a big limitation in what it can do. It also only computes power for an overall F test. Another thing that can be good or bad with R is that one of its inputs is the effect size, Δ. This means that in order to calculate power you first must calculate the effect size. This computation can be a difficult to do especially when you have a lot of groups to deal with. An alternative is to use what Cohen defines as a small, medium, or large effect size (.10, .25, .40 respectively), in his book *Statistical Power for the Behavioral Sciences*. This is a good option to use the group means are not known,

but when they are, it does not make sense to not calculate the effect size. One can also go by the convention of effect size defined in the field of the experiment.

When doing this same analysis in SAS, it has a lot more flexibility in what can be done. It can handle both a balanced and unbalanced design. To specify an unbalanced design the options GROUPWEIGHTS or GROUPN can be used. Not only can SAS handle both these designs, it can also test the overall F and it can test different contrasts, whereas R only tests the overall F. This gives SAS more flexibility and more options than R.

So overall using SAS would be a better choice. It can do a power analysis for a balanced and unbalanced design and it can handle both an overall F and different contrasts. The only time R has an advantage is when the group means are not known and the only option is to use a general effect. Otherwise I would use SAS in computing power for a one way ANOVA.

Chi Square Power Analysis
Both R and SAS are capable in doing a power analysis for a chi square test. However SAS has more options in which to choose from to do a variety of different things that R cannot. In SAS one is able to choose what kind of chi square test they want to run, Pearson, Fisher's exact test for two proportions, or a likelihood ratio chi square test. Also it allows for testing for different things like proportions, proportion differences, odds ratios, and relative risks. In R you are limited to a Pearson test and testing for a proportion. But there is a disadvantage in using SAS for these types of calculations; it only handles two proportions which limits its capabilities.

The advantage in using R is that it can handle two, three, five, or as many proportions as needed, whereas SAS only does two. However it does not have as many options as SAS does as mentioned earlier. R only deals with proportions, not odds ratios or relative risks and you can't use Fisher's exact test or a likelihood ratio test. Also effect size must be one of the inputs. But included in the pwr pakage are functions that will calculate this effect size. ES.w1 and ES.w2 are the functions that will calculate the effect size. This adds an extra step into the calculation but it is much easier to use these functions than to calculate them by hand. If one does not want to go through this step but want to use R then Cohen's suggestion of a small, medium, and large effect size (.10, .30, .50 respectively) can be used depending on what one wants to find from the study. One can also use the convention of effect size determined from the field one is in.

Therefore if one wants to test more than two proportions use R. If one is only testing two proportions and wants more flexibility SAS should be used.

Graphing Capabilities
An advantage of using SAS is that it is very easy to create a power curve. All it takes is an addition of a plot statement and the power curve will be created. If one is using R they have to create the graph themselves by using the results of the power analysis that was ran. This is a little more difficult and more labor intensive than SAS.

However there is more flexibility and control in how the graph looks in R. One can control the colors, the plotting symbols, location of the legend, and other things of that nature. In SAS these things are already predetermined by the software.

Works Cited

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New Jersey: Lawrence Erlbaum Associates, Inc.

Greenwood, P. E. (1996). *A Guide to Chi-Squared Testing.* New York: John Wiley & Sons.

Lipsey, M. W. (1990). *Design Sensitivity.* Newbury Park: SAGE Publications Inc.

Montgomery, D. C. (2008). *Design and Analysis of Experiments, 7th ed.* Wiley.

Navidi, W. (2007). *Statistics for Engineers and Scientist.* New York: McGraw-Hill Science/Engineering/Math.

SAS. (n.d.). *The Power Procedure*. Retrieved 2010, from SAS Documentation: http://support.sas.com/onlinedoc/913/docMainpage.jsp

cran.r-project. (2009, October). *R project*. Retrieved 2010, from Package 'pwr': http://cran.r-project.org/web/packages/pwr/pwr.pdf

Baur, D. (2004). *Proc Power in SAS 9.1.* SUGI.