Concepts of Statistical Power

For Variance Known and Large Sample Approximations, we can use the Central Normal (Unit Normal) Distribution as the Referent Distribution for a single degree-of freedom (*df*) hypothesis test, the Sampling Distribution under the Null Hypothesis.

The Non-Central Normal Distribution will be assumed to be the Distribution of the Test Statistic, the Sampling Distribution under the Alternative Hypothesis.

The Type I Error Rate is commonly set at $\alpha = 0.05$.
For $\alpha = 0.05$, the one-tailed Critical Value from the Central Normal Distribution is ±1.645, depending on the direction of the hypothesis. That is, the critical value is either the 100α percentile, 5^{th} percentile, or the $100(1-\alpha)$ percentile, 95^{th} percentile of the Central Normal Distribution.

For $\alpha = 0.05$, the two-tailed Critical Value from the Central Normal Distribution is ±1.96. That is, the critical values are the $100(\alpha/2)$ percentile, 2.5^{th} percentile, and the $100(1-\alpha/2)$ percentile, 97.5^{th} percentile of the Central Normal Distribution.

If a test statistic from the sampled data exceeds the critical value then the null hypothesis is rejected.
The probability that the null is rejected when in actuality it is true (Type I error) is α.

$$\text{Prob[Test} > \text{CV} \mid H_0 \text{ is True]} = \alpha.$$

Hypothesis Testing is based on controlling the Type I Error Rate because the Null Hypothesis is set up so that it can be rejected (i.e., the Principles of Falsification and Testability).

Decison	H_0 is True	H_0 is False
Reject H_0	Type I (α)	Power = $1-\beta$
Do Not Reject H_0	$1-\alpha$	Type II (β)

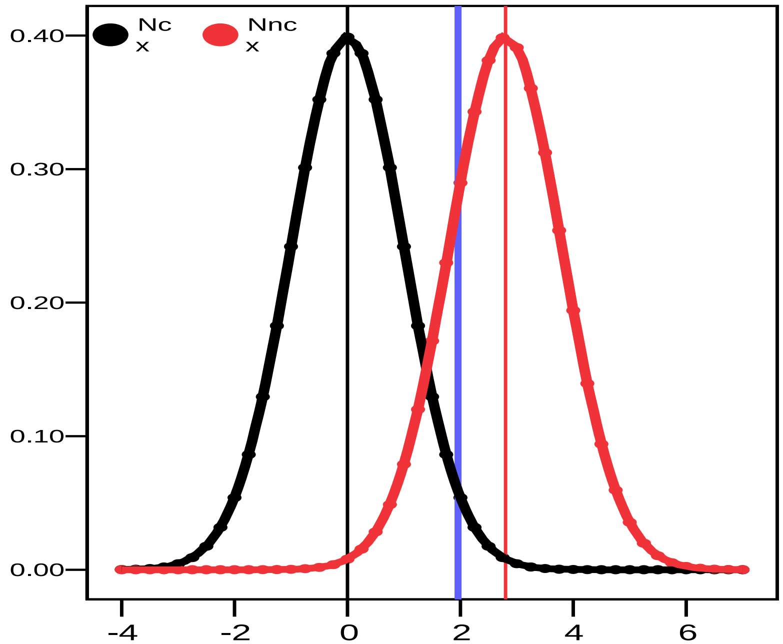
NO *a priori* Type II Error must be set to Test a Hypothesis. However, researchers want to have adequate Statistical Power to reject a False Null.

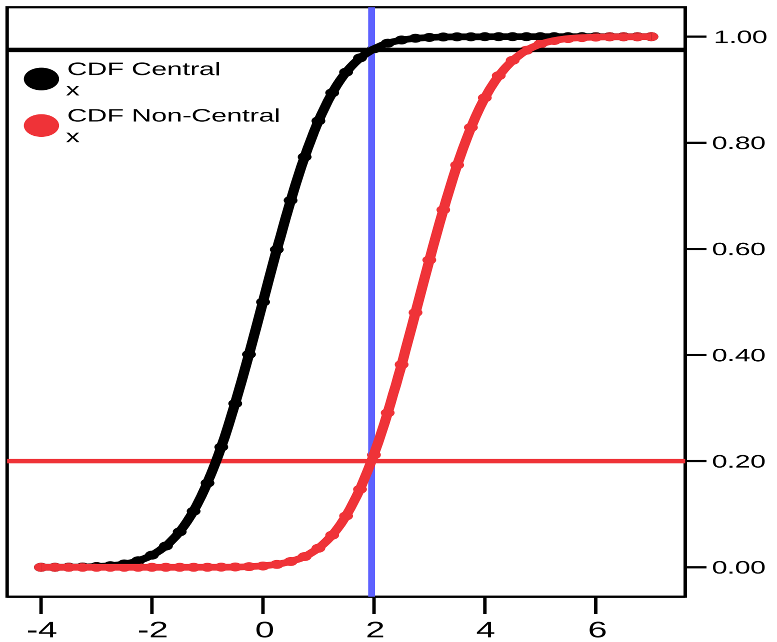
The Type II error rate is commonly set at β = 0.20. Thus Power is set at (1-β) = 0.80.
To calculate power in this context, we need to find the 100(1-β) percentile of the Normal Distribution.
The 100(1-β) percentile, 80th percentile of the Central Normal Distribution is 0.84.

```
data nn;
cv=1.96;
do x = -4 to 7 by 0.25;
PNc=PDF('NORMAL',x,0,1);
PNnc=PDF('NORMAL',x,2.8,1);
CNc=CDF('NORMAL',x,0,1);
CNnc=CDF('NORMAL',x,2.8,1);
output;end;run;
proc plot data=nn;
plot PNc*x='*'
 PNnc*x='o'
 PNnc*cv='I' / overlay box;
plot CNc*x='*'
 CNnc*x='o'
CNnc*cv='I' / overlay box;run;
```

If we add this value to the two-tailed Central Normal Critical Value, we obtain 1.96+ 0.84 = 2.8.

The Non-Central Normal distribution differs from the Central Normal distribution only by a shift in location (mean); it has the same variance and shape. Thus, if a Non-Central Normal distribution is centered at 2.8, then its 20th percentile is 1.96. If we repeatedly sample from Non-Central Normal is distribution is centered at 2.8, 80% of the time the values will be greater than the critical value 1.96, and hence, 80% Power. 80% of the Non-Central Normal PDF (Red) is greater than 1.96.





1.96 (Blue Line) is the 97.5th percentile of the Central Normal
1.96 (Blue Line) is the 20th percentile of the Non-Central Normal

The Central Normal (Black) CDF intersects 1.96 at 0.975.
The Non-Central Normal (Red) CDF intersects 1.96 at 0.20.

Large Sample Approximate Sample Size Calculation for a two-group comparison of Means

The Null Hypothesis is H_0: $(\mu_1 = \mu_0)$ or H_0: $(\mu_1 - \mu_0) = \mu_D = 0$.

Assume Homogeneity of Variance and that each group has a population Standard Deviation of:

$$\sigma_1 = \sigma_0 = \sigma_{POOL} = 2.$$

Assume we are looking for a very small effect of $(\mu_1 - \mu_0) = \mu_D = 0.2$. Cohen's $d = \frac{\bar{Y}_1 - \bar{Y}_0}{s_{pool}} = 0.2/2 = 0.1$.

So we know we will need a large sample size, and therefore, the Normal Approximation should work.

The Student's (pooled) *t*-statistic: $t_{(df=[n_1+n_0-2])} = \dfrac{\bar{Y}_1 - \bar{Y}_0}{s_{pool}\sqrt{1/n_1 + 1/n_0}}$, where $S_{pool} = \dfrac{(n_1-1)s_1^2 + (n_0-1)s_0^2}{(n_1+n_0-2)}$

If the *t*-test exceeds the critical value we will reject the null hypothesis; however, we need to solve for the Distribution Centered at the Critical Value plus the 80th percentile, in this case 2.8.

We can express the Large Sample Approximate Power as:

Power = Prob[$Z(\mu=2.8) \geq 1.96$]   or

Power = 1- Prob[$Z(\mu=2.8) < 1.96$]

In the case of the Large Sample Approximate will use the *Z*-statistic instead of the *t*-statistic and assume the *t* approximates the Normal.

Solving for Sample Size in the Large Sample Approximate

Assume the sample sizes are equal, $n_1 = n_0 = n$ $t_{(df=[n_1+n_0-2])} = \dfrac{\bar{Y}_1 - \bar{Y}_0}{s_{pool}\sqrt{2/n}} \geq \sim 2.8$

To solve for *n*, it will help to square both sides of the equation.

$$\frac{(\bar{Y}_1 - \bar{Y}_0)^2}{s^2_{Pool}\frac{2}{n}} \geq \sim 7.84 \;\rightarrow\; \frac{n(\bar{Y}_1 - \bar{Y}_0)^2}{2s^2_{Pool}} \geq \sim 7.84 \;\rightarrow\; n \geq \sim \frac{2s^2_{Pool}7.84}{(\bar{Y}_1 - \bar{Y}_0)^2} \;\rightarrow n \geq \sim \frac{15.68s^2_{Pool}}{(\bar{Y}_1 - \bar{Y}_0)^2}$$

In this case with $\sigma_{POOL} = 2$ and the small effect of $(\mu_1 - \mu_0) = \mu_D = 0.2$, $n \geq \sim \dfrac{(15.68)2^2}{(.2)^2}$

$$n \geq \sim 62.72/0.04 \rightarrow n \geq \sim 1{,}568 \text{ per group or } 3{,}136 \text{ Total}$$

For dummy-coded regression, the *F*-statistic for the correlation from a single predictor model is:

$t^2 = F_{[1,(N-2)]} = \dfrac{(N-2)r^2}{(1-r^2)} \geq$ **CV. Thus,** $N = \dfrac{CV(1-r^2)}{r^2} + 2$**. The the Effect Size Conversion Table**

$r^2 = \dfrac{d^2}{4+d^2} = .01/4.01 = 0.0025$**, Thus** $N = \dfrac{7.84(0.9975)}{0.0025} + 2 = 3{,}138$

If SAS® PROC POWER is used to solve for the Sample Size

```
proc power;
   onewayanova test=overall
   alpha = 0.05
   groupmeans = 0 | 0.2
   stddev = 2
   npergroup = .
   power = .80;  run;
```

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

      Fixed Scenario Elements
  Method                    Exact
  Alpha                      0.05
  Group Means                0 0.2
  Standard Deviation           2
  Nominal Power               0.8

    Computed N Per Group
      Actual     N Per
       Power     Group
       0.800      1571
```

```
proc power;
  onewayanova test=overall
  alpha = 0.05
  groupmeans = 0 | 0.2
  stddev = 2
  ntotal = .
  power = .80;  run;
```

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

         Fixed Scenario Elements
Method                    Exact
Alpha                      0.05
Group Means                0 0.2
Standard Deviation           2
Nominal Power               0.8

      Computed N Total
        Actual     N
         Power     Total
         0.800      3142
```

$n \geq \sim$ 1,568 per group or 3,136 Total Is Good Approximation

Small Sample Size Calculation for a two-group comparison of Means

With smaller sample sizes this may not be a very good approximate. Technically, the Central *t*-distribution, as a Referent Distribution for the *t*-statistic, has a different critical value for each value for the *df*s.

The Expected Value of the Central *t*-distribution is 0. The Variance of the Central *t*-distribution is *df*/(*df*-2).

As the *df*s become "large," the Central *t*-distribution approximates the Central Normal with a Variance of one.

Power and Sample Size calculations are based on obtaining a Critical Value from the Central *t*-distribution (based on *df*s) and finding where that Critical Value falls within the Non-Central *t*-distribution (same *df*s) with a Non-Centrality Parameter of δ.

The Central *t*-Distribution is the Referent Distribution for a single *df* hypothesis test, the Sampling Distribution under the Null Hypothesis.

The Non-Central *t*-Distribution is the Distribution of the Test Statistic, the Sampling Distribution under the Alternative Hypothesis.

The Expected Value, Variance, and higher moments of the Non-Central *t*-distribution are rather complicated functions. In fact, the Non-Central *t*-distribution is a bit skewed with smaller *df*s. However, as the *df*s become "large" the Non-Central *t*-distribution approximates the Non-Central Normal with a Variance of one.

In the two-group mean comparison scenario the Non-Centrality Parameter, δ = $\frac{|\mu_1 - \mu_0|}{\sigma}\sqrt{\frac{n_1 n_0}{n_1 + n_0}}$

Suppose, we assume the Homogeneity of Variance and that each group has a population Standard Deviation of: $\sigma_1 = \sigma_0 = \sigma_{POOL} = 1$.

and Assume we are looking for a fairly large effect of $(\mu_1 - \mu_0) = \mu_D = 1.2$.

Based on a prior calculation, to have **80% Power at α = 0.05**, there will need to be **12 subjects per group**.

$$\delta = \frac{|\mu_1 - \mu_0|}{\sigma}\sqrt{\frac{n_1 n_0}{n_1 + n_0}} = \frac{1.2}{1}\sqrt{\frac{12 * 12}{12 + 12}} = \mathbf{2.94}$$

The Variance of the Central *t*-distribution with *df* = (12+12-2) = 22 is 1.1.

The 97.5th percentile of the Central *t*-distribution with *df* = 22 is 2.074

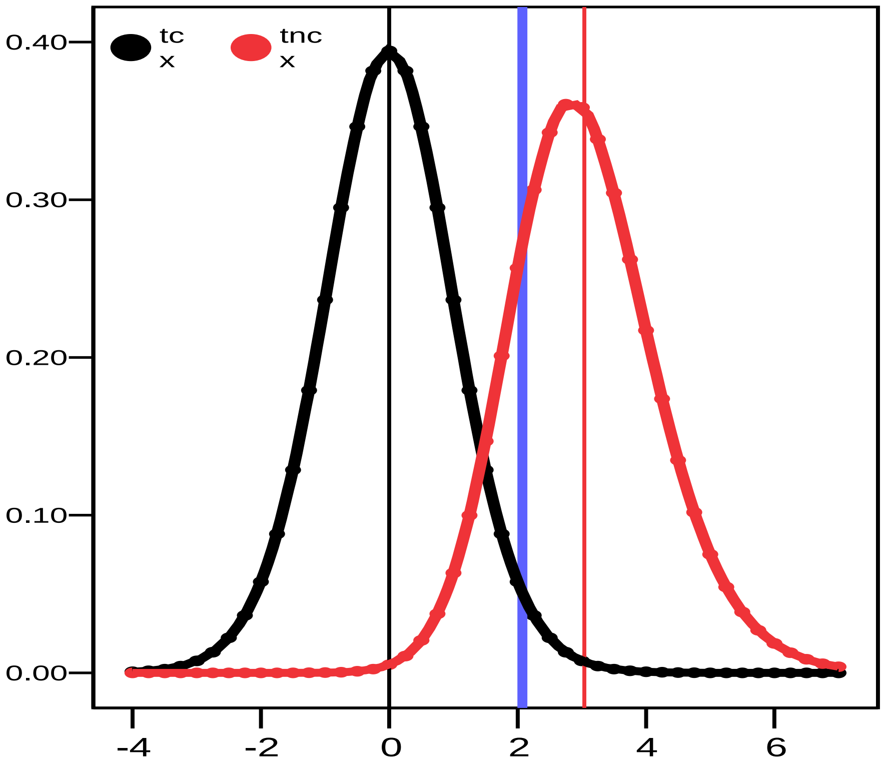
(Critical Value from the Referent Distribution)

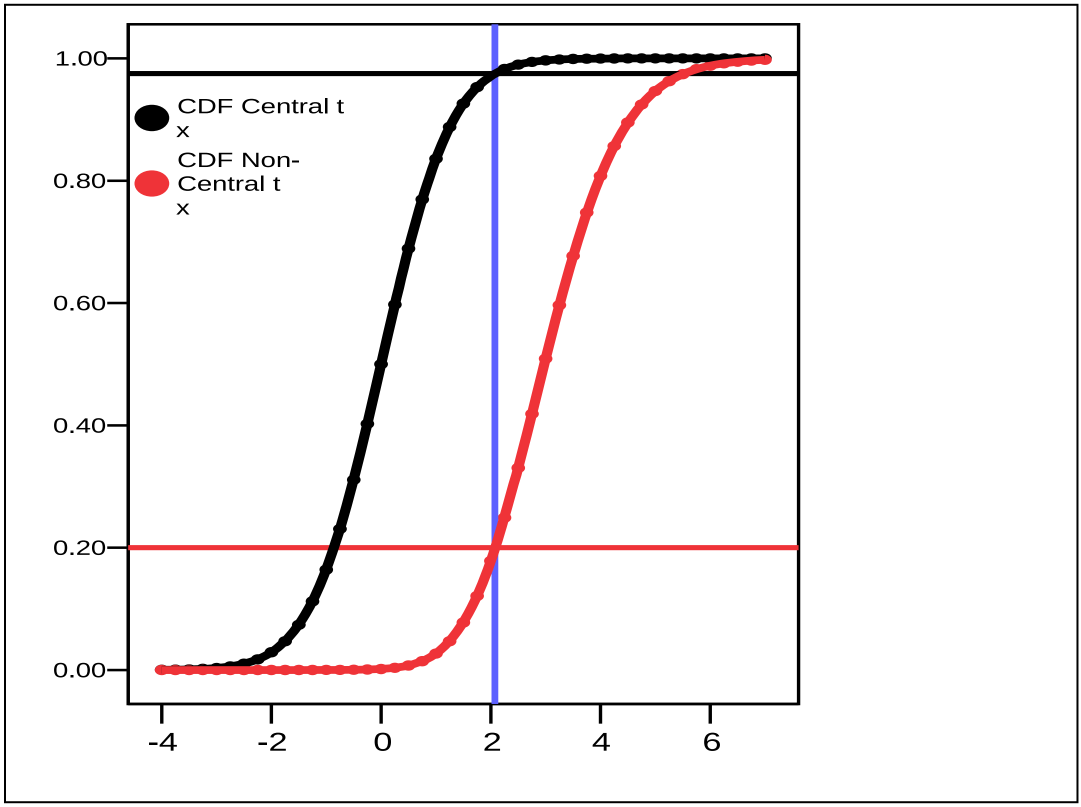
The Expected Value of the Non-Central *t*-distribution with δ = 2.94 and *df* = 22 is 3.04.

The Variance of the Non-Central *t*-distribution with δ = 2.94 and *df* = 22 is 1.33.

The 20th percentile of the Non-Central *t*-distribution with δ = 2.94 and *df* = 22 is 2.081

Power = Prob[$t(df, \delta) \geq t([1-\alpha/2, df, 0)]$





2.074 (Blue Line) is the 97.5th percentile of the Central *t*
2.074 (Blue Line) is approximately the 20th percentile of the Non-Central *t*. Approximately 80% of the Non-Central t PDF (Red) is greater than 2.074.

The Central *t* (Black) CDF intersects 2.074 at 0.975.
The Non-Central Normal (Red) CDF intersects 2.074 at approximately 0.20.

```
data tt;
dfe=22;alpha=0.05; ncp=2.9393877;
tcv=TINV((1-(alpha/2)), dfe,0);
do x = -4 to 7 by 0.25;
Ptc=PDF('T',x,dfe,0);Ptnc=PDF('T',x,dfe,ncp);
Ctc=CDF('T',x,dfe,0);Ctnc=CDF('T',x,dfe,ncp);
output;end;run;
```

```
proc plot data=tt;
   plot Ptc*x='*'
        Ptnc*x='o'
        Ptnc*tcv='I' / overlay box;
   plot Ctc*x='*'
        Ctnc*x='o'
        Ctnc*tcv='I' / overlay box;run;
```

If SAS® PROC POWER is used to solve for Power and Sample Size

```
proc power;
   onewayanova test=overall
   alpha = 0.05
   groupmeans = 0 | 1.2
   stddev = 1
   npergroup = 12
   power = .;    run;
```

```
proc power;
   onewayanova test=overall
   alpha = 0.05
   groupmeans = 0 | 1.2
   stddev = 1
   npergroup = .
   power = .80;    run;
```

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

       Fixed Scenario Elements
Method                      Exact
Alpha                       0.05
Group Means                 0 1.2
Standard Deviation            1
Sample Size Per Group       12

           Computed Power
              Power
              0.802
```

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

       Fixed Scenario Elements
Method                      Exact
Alpha                       0.05
Group Means                 0 1.2
Standard Deviation            1
Nominal Power               0.8

        Computed N Per Group
            Actual     N Per
            Power      Group
            0.802        12
```

The power exceeds 0.80 because the 20th percentile of the Non-Central *t*-distribution with δ = 2.94 and *df* = 22 is 2.081 and the 97.5th percentile of the Central *t*-distribution with *df* = 22 is 2.074. That is, 2.074 is actually at the 19.8th percentile of the Non-Central *t*-distribution with δ = 2.94 and *df* = 22.

However, Large Sample Approximate $n \geq \sim \dfrac{15.68 s^2_{Pool}}{(\bar{Y}_1 - \bar{Y}_0)^2}$ **$n \geq \sim 15.68/(1.2)^2 = 10.88$; $n \geq \sim 11$ per group**

OR $= r^2 = \dfrac{d^2}{4+d^2} = 1.44/5.44 = 0.2647$, **Thus** $N = \dfrac{7.84(0.7353)}{0.2647} + 2 = 23.78 \approx 24 \approx 12$ **per group**

Is NOT that BAD in this circumstance, but with smaller effects and smaller α the approximation gets worse.

```
proc glm data=cta2x3;class group;
model score = group;means group;means group /tukey ;run;
***** Means are from the Output *****;
***** The Root MSE from Output is used as the common stddev **;
***** Observed Power at alpha = 0.05 ****************************************;
proc power;
      onewayanova test=overall
         alpha = 0.05
         groupmeans = 73.67 | 79.78  | 75.94
         stddev = 12.27
         groupns = (55 45 50)
         power = .;    run;
***** Future Total Sample Size for study to have 80%
 *** at alpha=.01  Assuming balanced design  *****************;
proc power;
      onewayanova test=overall
         alpha = 0.01
         groupmeans = 73.67 | 79.78  | 75.94
         stddev = 12.27
         ntotal = .
         power = .80;    run;

***** Future Total Sample Size for Pairwise Contrast to have 80%
 *** at alpha=.01  Assuming balanced design  *****************;
proc power;
         onewayanova test=contrast
         contrast = (1 -1 0)
         alpha = 0.01
         groupmeans = 73.67 | 79.78  | 75.94
         stddev = 12.27
         ntotal = .
         power = .80;    run;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
group	3	0 1 2

Number of Observations Read	150
Number of Observations Used	150

Dependent Variable: score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	930.15313	465.07657	3.09	0.0485
Error	147	22128.70687	150.53542		
Corrected Total	149	23058.86000			

R-Square	Coeff Var	Root MSE	score Mean
0.040338	16.08876	**12.26929**	76.26000

Level of group	N	score Mean	score Std Dev
0	**55**	**73.6727273**	14.0185783
1	**45**	**79.7777778**	10.7044784
2	**50**	**75.9400000**	11.4951809

Tukey's Studentized Range (HSD) Test for score

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	147
Error Mean Square	150.5354
Critical Value of Studentized Range	3.34848

Comparisons significant at the 0.05 level are indicated by ***.

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 2	3.838	-2.132	9.807	
1 - 0	6.105	0.266	11.944	***
2 - 1	-3.838	-9.807	2.132	

```
*** Means are form the Output  The Root MSE from Output is used as the common stddev **;
****************************************************************;

***** Observed Power at alpha = 0.05 ****************************************;
```

The POWER Procedure

Overall F Test for One-Way ANOVA

Fixed Scenario Elements

Method	Exact
Alpha	0.05
Group Means	73.67 79.78 75.94
Standard Deviation	12.27
Group Sample Sizes	55 45 50

Computed Power

Power
0.589

```
* Future Total Sample Size for study to have 80% at alpha=.01  Assuming balanced design *;
```

The POWER Procedure

Overall F Test for One-Way ANOVA

Fixed Scenario Elements

Method	Exact
Alpha	0.01
Group Means	73.67 79.78 75.94
Standard Deviation	12.27
Nominal Power	0.8
Group Weights	1 1 1

Computed N Total

Actual Power	N Total
0.804	**336**

```
***** Observed Power at alpha = 0.05 for Pairwise Contrast *********************;
```

***** Uncorrected **************************

The POWER Procedure

Single DF Contrast in One-Way ANOVA

Fixed Scenario Elements

Method	Exact
Contrast Coefficients	1 -1 0
Alpha	**0.05**
Group Means	73.67 79.78 75.94
Standard Deviation	12.27
Group Sample Sizes	55 45 50
Number of Sides	2
Null Contrast Value	0

Computed Power

Power
0.692

This Power is higher than **0.589** for the omnibus test; however, there has been no adjustment for Multiple Comparisions.

Bonferroni Corrected

The POWER Procedure

Single DF Contrast in One-Way ANOVA

Fixed Scenario Elements

Method	Exact
Contrast Coefficients	1 -1 0
Alpha	**0.016667**
Group Means	73.67 79.78 75.94
Standard Deviation	12.27
Group Sample Sizes	55 45 50
Number of Sides	2
Null Contrast Value	0

Computed Power

Power
0.524

Bonferroni Correction applied to the Alpha
$\alpha_{BON} = \alpha/k = 0.05/3 = 0.016667$
Unfortunately PROC POWER does not have alpha adjustment options such as Tukey or Dunnett

```
* Future Total Sample Size for study to have 80% at alpha=.01  Assuming balanced design *;
***** Uncorrected **************************     Bonferroni Corrected
```

Uncorrected

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

         Fixed Scenario Elements
Method                                   Exact
Contrast Coefficients                   1 -1 0
Group Means                  73.67 79.78 75.94
Standard Deviation                       12.27
Nominal Power                              0.8
Number of Sides                              2
Null Contrast Value                          0
Alpha                                     0.01
Group Weights                            1 1 1

       Computed N Total

       Actual         N
        Power     Total
        0.803       288
```

This Sample Size is lowerr than **336** for the omnibus test; however, there has been no adjustment for Multiple Comparisions.

Bonferroni Corrected

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

         Fixed Scenario Elements
Method                                   Exact
Contrast Coefficients                   1 -1 0
Group Means                  73.67 79.78 75.94
Standard Deviation                       12.27
Nominal Power                              0.8
Number of Sides                              2
Null Contrast Value                          0
Alpha                                 0.003333
Group Weights                            1 1 1

       Computed N Total

       Actual         N
        Power     Total
        0.802       351
```

Bonferroni Correction applied to the Alpha
$\alpha_{BON} = \alpha/k = 0.01/3 = 0.003333$
Unfortunately PROC POWER does not have alpha adjustment options such as Tukey or Dunnett