Module 3: Power

What is Power? + Calculating Power.

Motivation

We haven't yet answered... How much data should we collect?

Outcomes of hypothesis testing

	H_0 **is True**	H_0 **is False**
Reject H_0	Type I (__________)	No Error (__________)
Fail to Reject H_0	No Error (__________)	Type II (__________)

What power means

Power $(1 - \beta)$

The **power** of a test is the probability of correctly rejecting the null hypothesis if a particular alternative scenario is true.

$$\text{Power} = 1 - \beta = P(\text{reject } H_0 \mid \text{a meaningful difference does in fact e}$$

Why β is harder to find than α

Power depends on:

- Variability (σ^2)
- Sample size (replications)
- Design choices

A key idea: You cannot compute power for “something differs”

Example 3.1: Shoe Type (extended)

Suppose, we are preparing to conduct a study for:

- $t = 4$ shoe types
- $r = 3$ runners per shoe
- CRD randomization

Meaningful difference

δ = difference worth detecting

- Example: Fastest vs slowest shoe differs by **10 seconds**

What we must specify

Information Needed to Calculate the Power of a Test

1. a specific alternative (e.g., δ)
2. replications per treatment (r)
3. variability estimate (σ^2)
4. significance level (α)

Reminder: the ANOVA test

$$F = \frac{\mathrm{MSTrt}}{\mathrm{MSE}}$$

- **If** H_0 is true, $F \sim F(\mathrm{df}_1, \mathrm{df}_2)$ (Central F)
- **If a specific** H_A is true $F \sim F(\mathrm{df}_1, \mathrm{df}_2, \lambda)$ (Noncentral F)

Degrees of freedom (CRD)

If we have t treatments and r reps each:

- $\mathrm{df}_1 = t - 1$
- $\mathrm{df}_2 = t(r - 1)$

Example 3.1: Skeleton ANOVA

SV	DF: 12 runners - 1 = 11 total
Shoe Type	(4 - 1) = 3
Runner(Shoe Type) $\rightarrow$ error	(3 - 1)(4) = 8

The noncentrality parameter

$$\lambda = \frac{\sum_{i=1}^{t} (\mu_i - \bar{\mu})^2}{\sigma^2 / r}$$

- _____ = the mean of the i^{th} treatment group
- _____ = the overall mean
- _____ = the experimental error variance
- _____ = the number of replications of the i^{th} treatment group

Sanity check

If all means are equal, then $\lambda = 0$.

Example 3.1: Shoe Type

Recall, the things we need to calculate power are:

1. A specific alternative (e.g., δ)

An ultra-marathon running expert tells us that if there is a difference of 10 seconds between the runners lap times, then that would be of practical importance. Thus, $\delta = 10$.

Example 3.1: Shoe Type

Recall, the things we need to calculate power are:

2. Replications per treatment (r)

We are conducting a study with $t = 4$ and $r = 3$.

Example 3.1: Shoe Type

Recall, the things we need to calculate power are:

3. variability estimate (σ^2)

From an analysis of a pilot study (from Mod 2: CRD Notes), we found $\hat{\sigma}^2 = 17.46$

> **Pilot Studies**
>
> Pilot Studies can be super helpful for estimating experimental error variance $(\hat{\sigma}^2)$ and differences in group means – key pieces of a power analysis. Running a small version of your experiment gives you data to calculate the mean square error and get a sense of the effect size. This makes it easier to plan the full study and ensure your design has enough power to detect meaningful differences.

Example 3.1: Shoe Type

Recall, the things we need to calculate power are:

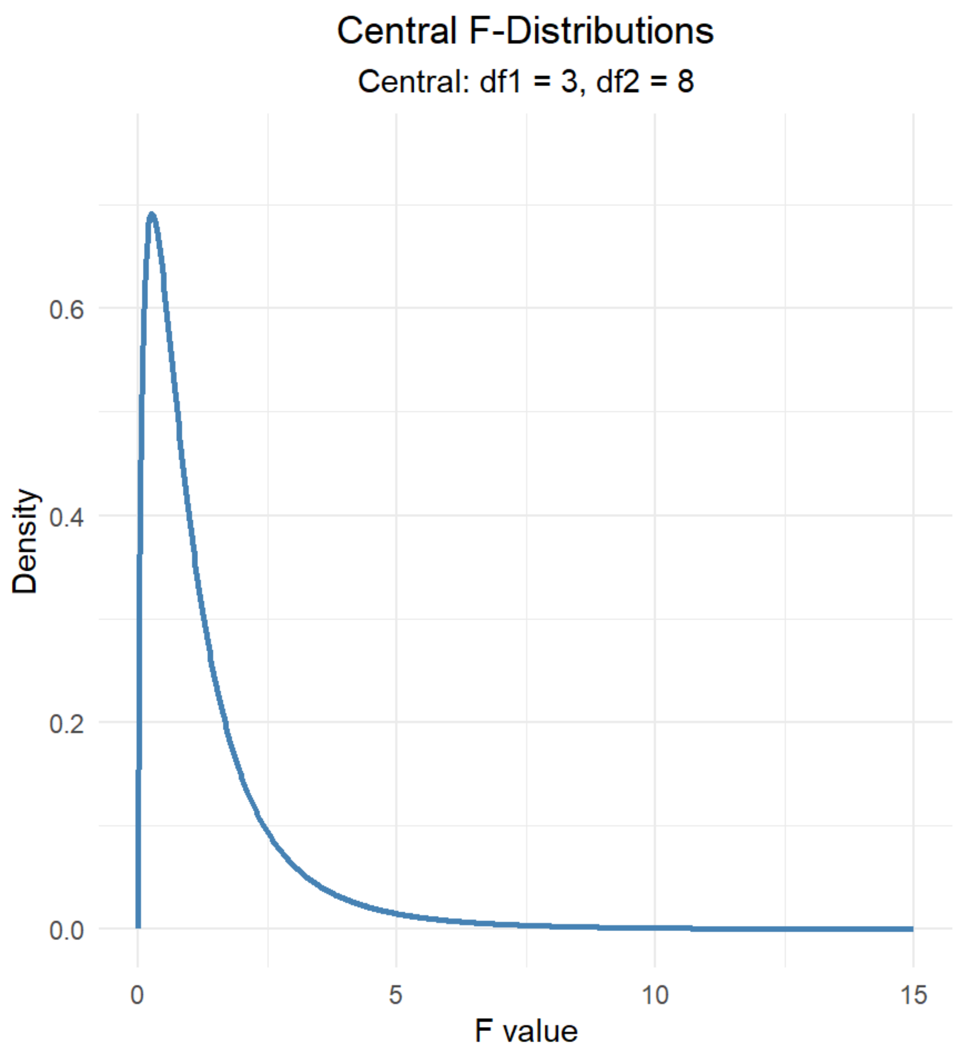
4. Significance level (α)

We get to set this at $\alpha = 0.05$.

Power as area

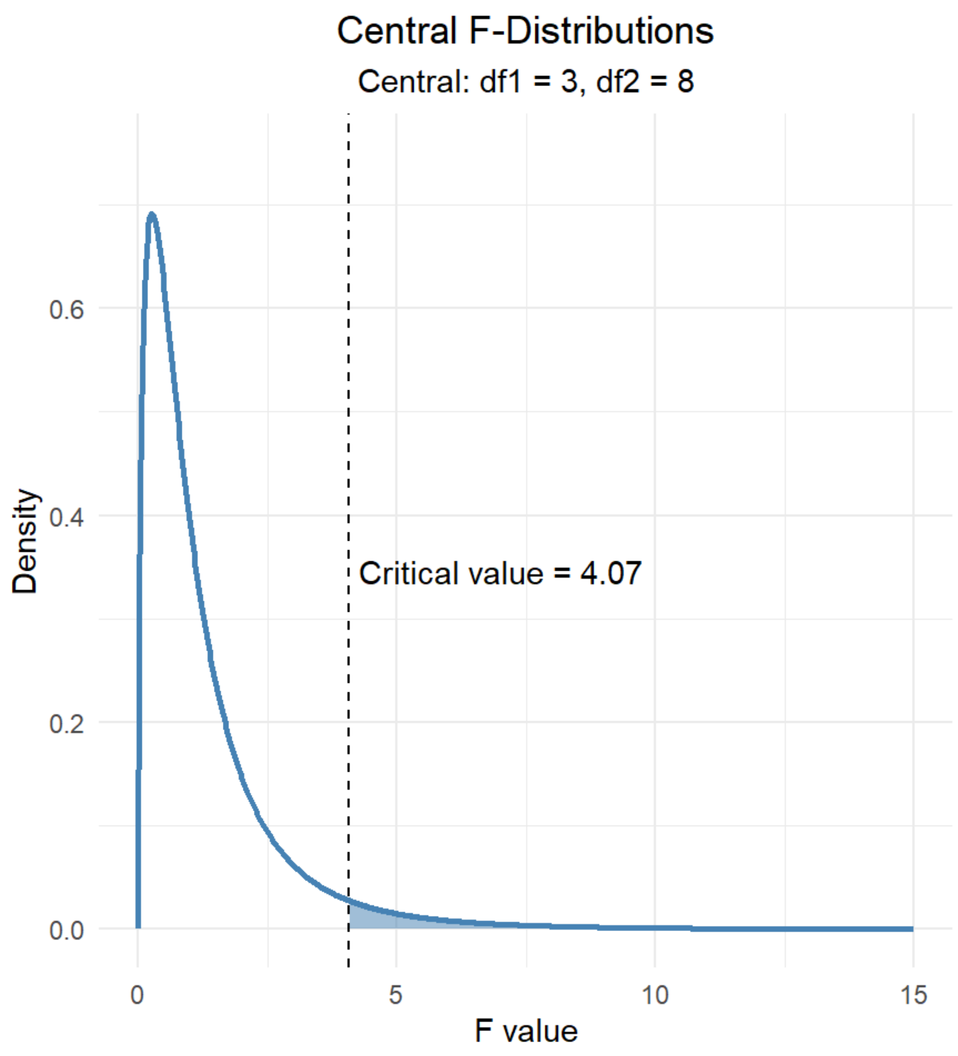
Power as an area

Consider the central F-distribution with these parameters:



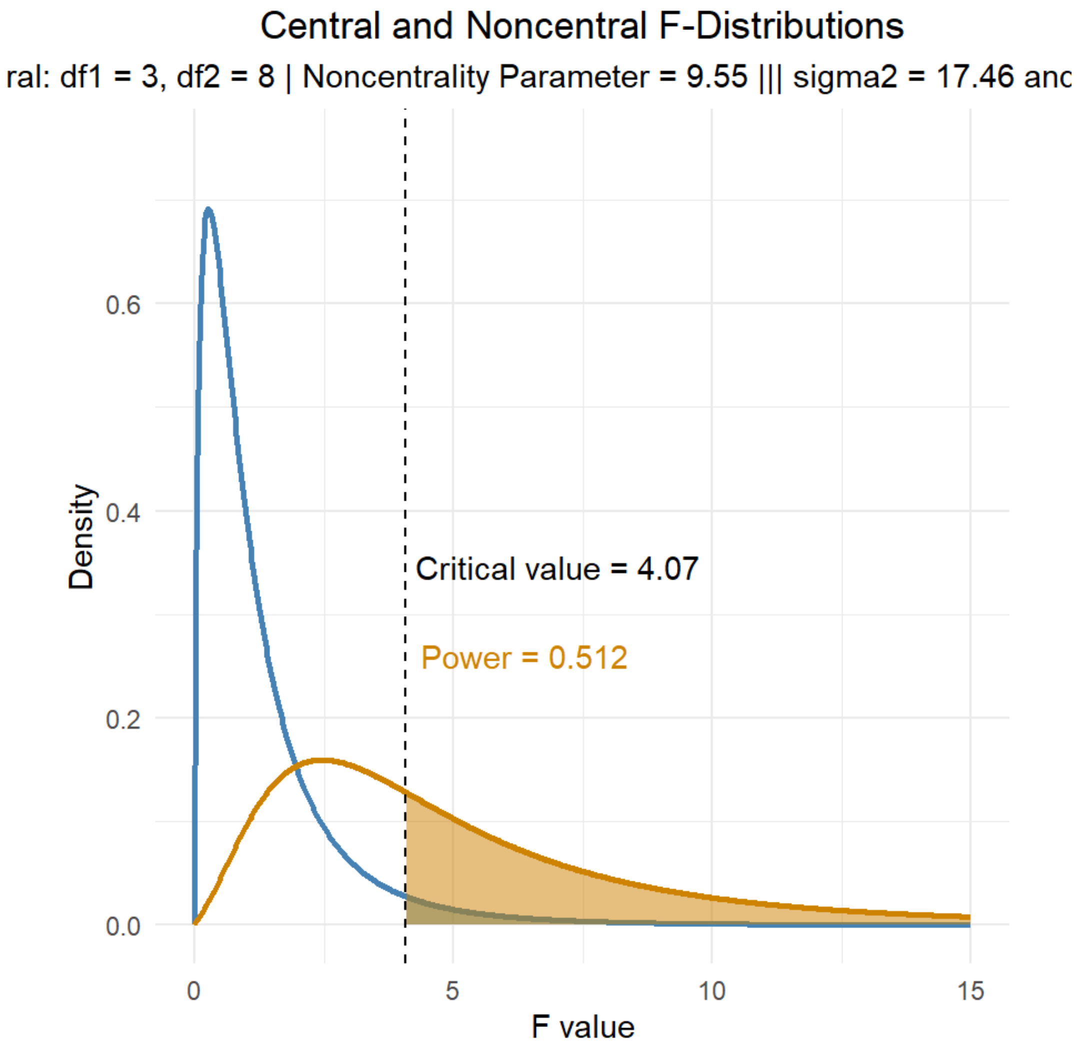
Power as area

Recall, that the power represents the probability of **rejecting the null** when a specific alternative is true. So let’s focus on the region where we’ll reject the null hypothesis (the Rejection Region)



Power as area

Now, we must incorporate the part about the **specific alternative being true**. If our particular alternative is true, then the F-statistic actually follows a non-central F-distribution with the aforementioned parameters. This is plotted on the same plot as the central F-distribution below:



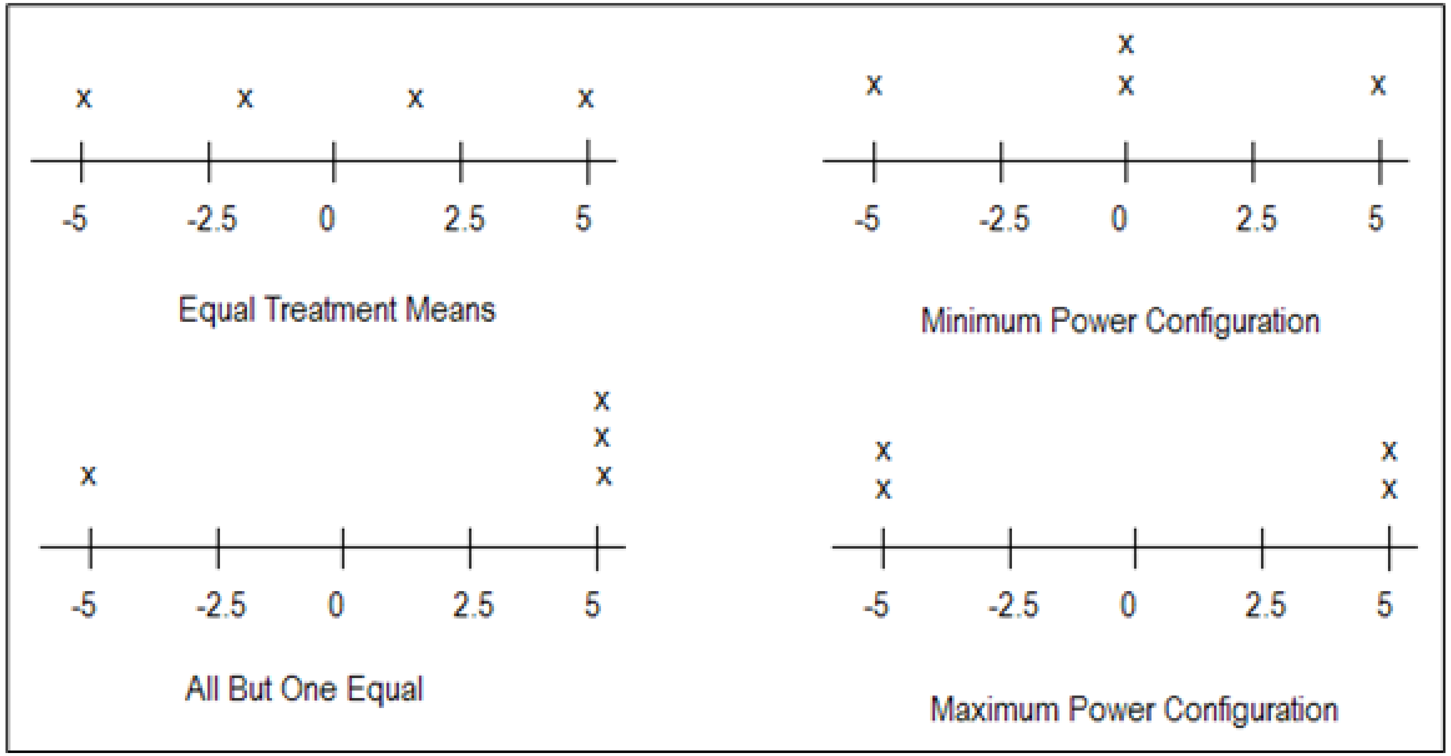
The power for this test is about ________. That is, there is only a 51% chance of correctly rejecting the null hypothesis under these conditions. Do you think this is a good test?

Ideal Power

For most experiments, we would like the power of the test to be at least ________.

Power Configurations

Even with the same max difference δ...



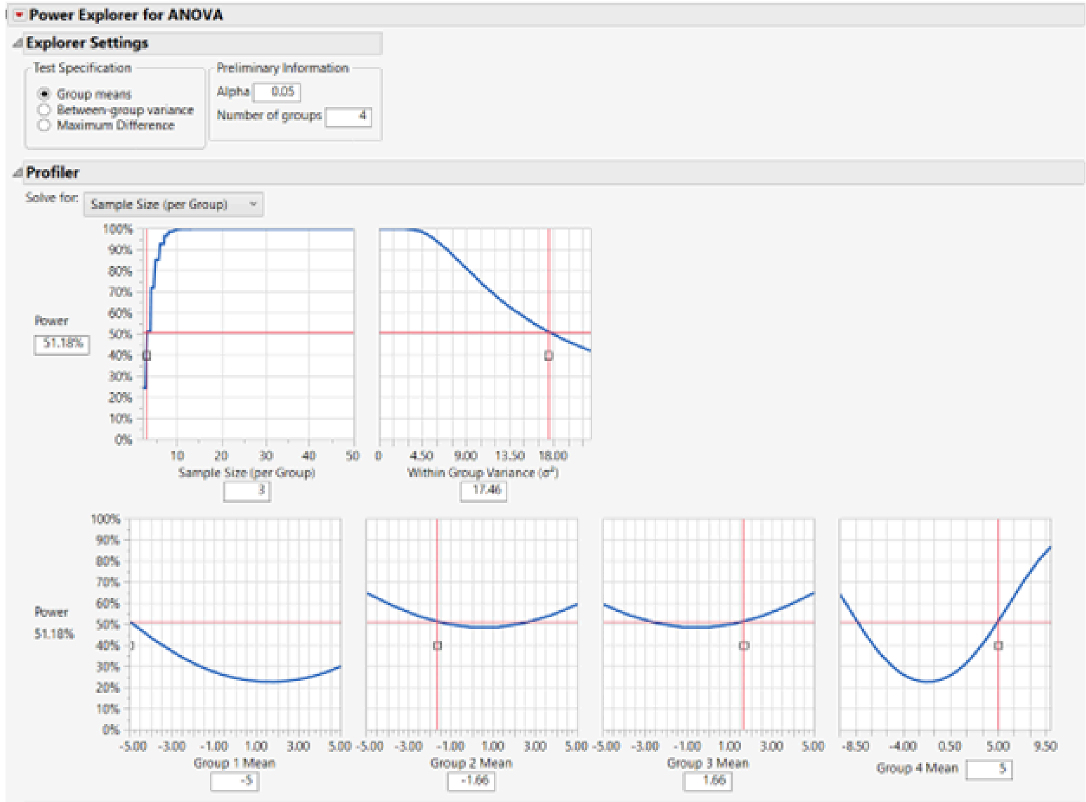
Power depends on the configuration

Mean pattern	Power (approx.)
Maximum power	0.78
Minimum power	**0.47**
Equally spaced	0.51
All but one equal	0.65

Using JMP

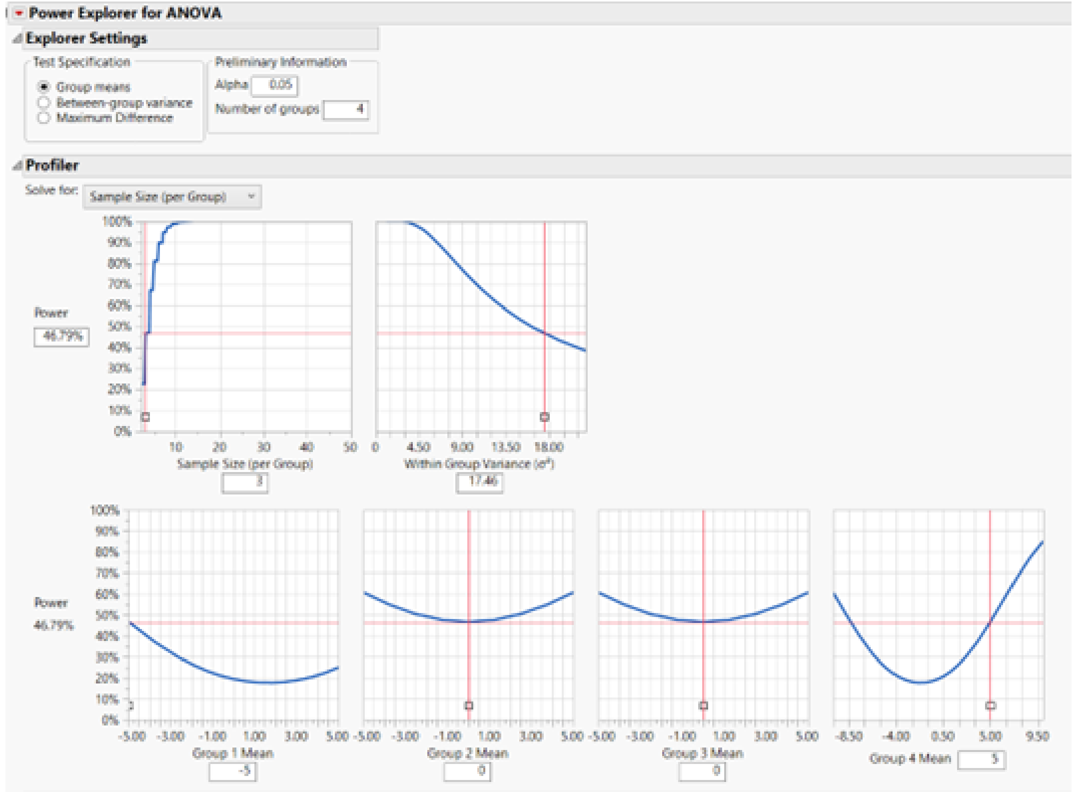
JMP: Power Explorer

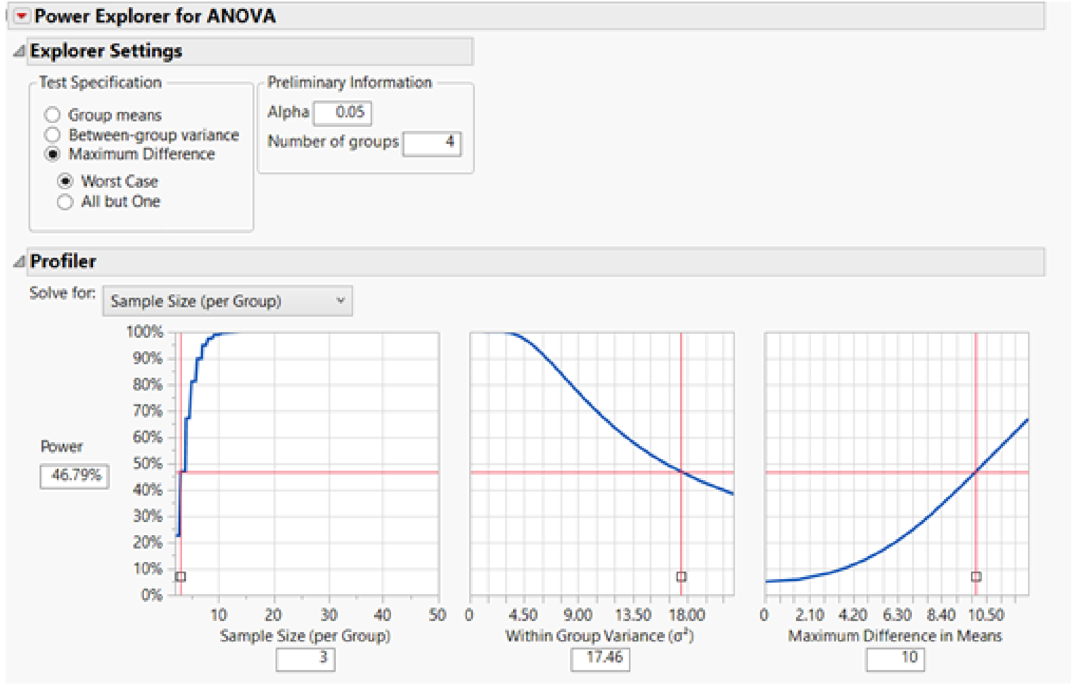
DOE > Sample Size Explorers > Power > Power for ANOVA



JMP: Power Explorer

Minimum Power = `Maximum Difference` > `Worst Case`





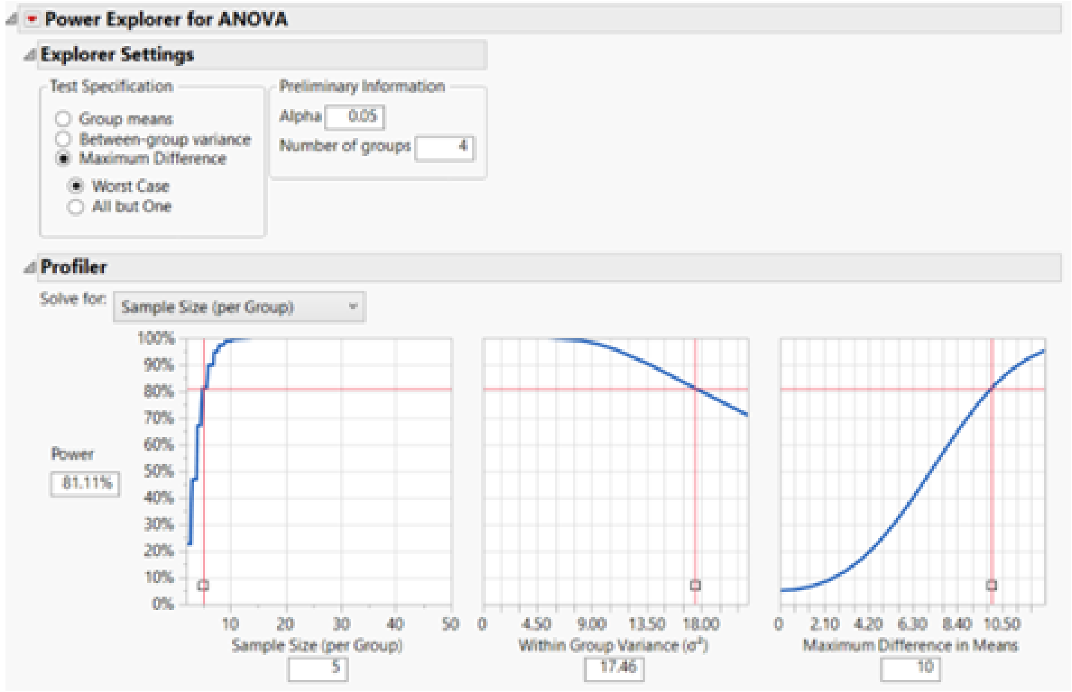
JMP: Power Explorer

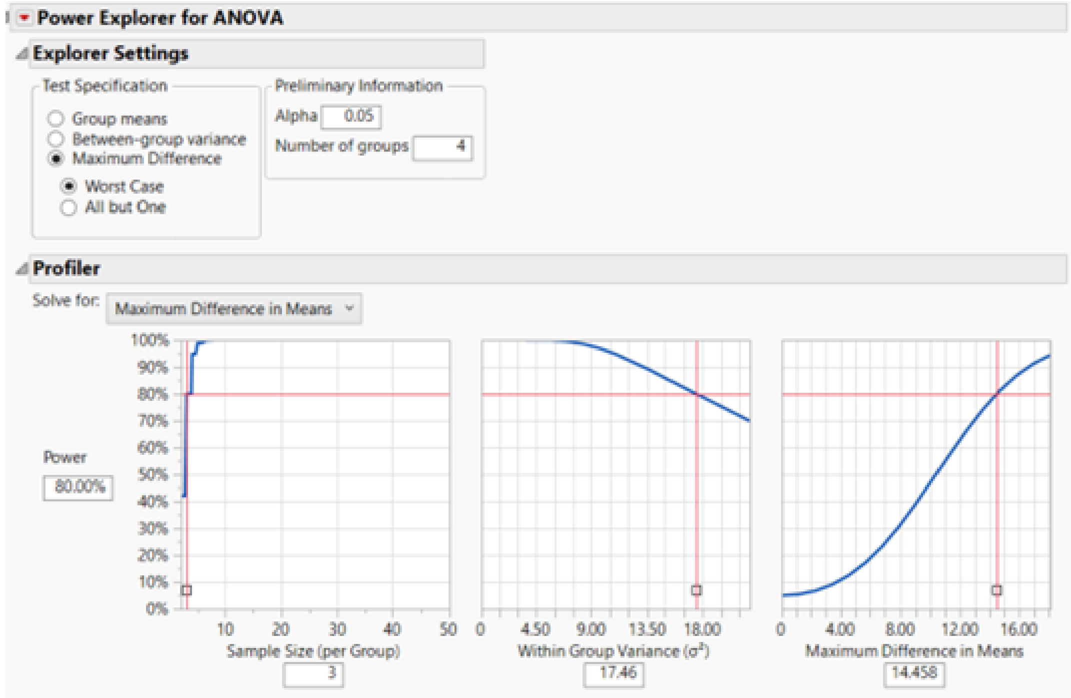
How many reps do we need?

Fix α, σ^2, δ, power
Solve for r

What difference can we detect?

Fix α, σ^2, r, power
Solve for δ





Power in R

What is the the power of my test?

```r
library(pwr2)
pwr.1way(
  k = 4,               # Number of groups
  n = 2,               # Sample size per group
  alpha = 0.05,        # Significance level
  delta = 10,          # Difference in group means
  sigma = sqrt(17.46) # Standard devision of obs
)
```

```
     Balanced one-way analysis of variance power calculation 

              k = 4
              n = 2
          delta = 10
          sigma = 4.178516
    effect.size = 0.8461218
      sig.level = 0.05
          power = 0.2254578

NOTE: n is number in each group, total sample = 8 power = 0.225457781758005
```

What is the sample size I need?

```r
ss.1way(
  k = 4,                    # Number of groups
  alpha = 0.05,             # Significance level
  beta = 1 - 0.80,          # Type II error rate (1 - power)
  delta = 10,               # Difference in group means
  sigma = sqrt(17.46),      # Standard deviation of obs
  B = 100                   # Number of iterations for numerical search
)
```

```
     Balanced one-way analysis of variance sample size adjustment 

              k = 4
      sig.level = 0.05
          power = 0.8
              n = 5

NOTE: n is number in each group, total sample = 20
```