Power

STAT:5201

Week 8: Lecture 1
We have already described Type I and II errors.

<table>
<thead>
<tr>
<th>Reality/True state</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accept $H_0$</td>
</tr>
<tr>
<td>$H_0$ is true</td>
<td>good</td>
</tr>
<tr>
<td>$H_0$ is false</td>
<td>Type II error</td>
</tr>
</tbody>
</table>

**Power** is the probability of rejecting the null when the null is false (something we really hope to do). We want high power.

We commonly use $\alpha$ to represent the Type I error rate, and $\beta$ to represent the Type II error rate. Then, power is $1 - \beta$.

Power (or $1 - \beta$) is a bit more difficult to deal with than the type I error rate or $\alpha$. 

There is a trade-off between Type I and Type II errors.

Decreasing our likelihood of a Type I error means we are less likely to make rejections.

In designing experiments and choosing sample sizes, we should consider power...
* If the power is too low, then you’re wasting your time and resources running experiments with very little chance of detecting something interesting.
* If the power is too high, then you are spending resources in this experiment that might be better spent somewhere else. Like hitting a fly with nuclear weapon, too much...

In practice, we tend to be content with a power of 0.80 or 0.90. This provides high likelihood to detect something interesting when it’s really there.
Refresher: sample size calculation for one-sample test with $\sigma^2$ known, find $n$

- $H_0: \mu = \mu_o$ vs. $H_A: \mu > \mu_o$ (one-sided)

- Choose $\alpha \equiv$ type I error rate
- Choose $\beta \equiv$ type II error rate (i.e. set power $= 1 - \beta$)

- Get cut-off to control type I error rate at $\alpha$ level:

$$P(\bar{Y} > c|H_o\text{true}) = P\left(\frac{\bar{Y} - \mu_o}{\sigma/\sqrt{n}} > \frac{c - \mu_o}{\sigma/\sqrt{n}}|H_o\text{true}\right) = P\left(z > \frac{c - \mu_o}{\sigma/\sqrt{n}}\right) = \alpha$$

$$\Rightarrow c = \mu_o + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$
Refresher: sample size calculation for one-sample test with $\sigma^2$ known, find $n$

- Setting our cut-off for significance at $c = \mu_o + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$ controls the Type I error rate at the $\alpha$ level.

- Now, for a chosen $\beta$ and a given $\mu_A$ under $H_A$ true, we have

$$P(\bar{Y} > c|H_A \text{ true}) = P\left(\frac{\bar{Y} - \mu_A}{\sigma/\sqrt{n}} > \frac{c - \mu_A}{\sigma/\sqrt{n}}|H_A \text{ true}\right)$$

$$= P\left(z > \frac{(\mu_o + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}) - \mu_A}{\sigma/\sqrt{n}}\right)$$

$$= P\left(z > z_\alpha - \frac{d}{\sigma/\sqrt{n}}\right) = 1 - \beta \quad \text{(where} \ d = \mu_A - \mu_o\text{)}$$

(to control Type II error)

So, $z_\alpha - \frac{d}{\sigma/\sqrt{n}} = z_{1-\beta}$

$$\Rightarrow \quad n = \left(\frac{\sigma(z_\alpha - z_{1-\beta})}{d}\right)^2$$

sample size
Refresher: sample size calculation for one-sample test with $\sigma^2$ known, find $n$

- Here, we can see the standardized threshold of $z_\alpha$ controlling the Type I error rate at the $\alpha$ level for sample size $n$.
- Then, given the true mean of $\mu_A$, we see that if we reject for standardized values greater than $z_\alpha - \frac{d}{\sigma/\sqrt{n}}$, then we have a Type II error rate of $\beta$, or power of $1 - \beta$.

\[
\text{Standardized}
\]

\[
\text{Original scale}
\]
Refresher: sample size calculation for one-sample test with $\sigma^2$ known, find $n$

- If using the ‘unstandardized’ values,

$$P(\bar{Y} > \mu_o + z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \mid H_o \text{ true}) = \alpha$$

$$P(\bar{Y} > \mu_o + z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \mid H_A \text{ true}) = 1 - \beta$$

Power is the area to the right of the critical value under the shifted-right normal curve.
Refresher: sample size calculation for one-sample test with $\sigma^2$ unknown, find $n$

- $H_0 : \mu = \mu_0$ vs. $H_A : \mu > \mu_0$ (one-sided)
- Test utilizes a “$t$-test”, more complex because $\sigma^2$ not known.
- Get cut-off to control type I error rate at $\alpha$ level:

$$P(\bar{Y} > c|H_0 \text{true}) = P(\frac{\bar{Y} - \mu_0}{s/\sqrt{n}} > \frac{c - \mu_0}{s/\sqrt{n}}|H_0 \text{true}) = P(t > \frac{c - \mu_0}{s/\sqrt{n}}) = \alpha$$

$$\Rightarrow c = \mu_0 + t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}$$
Refresher: sample size calculation for one-sample test with $\sigma^2$ unknown, find $n$

- Setting our cut-off for significance at $c = \mu_0 + t_{\alpha,n-1} \cdot \frac{\sigma}{\sqrt{n}}$ controls the Type I error rate at the $\alpha$ level.

- Now, for a chosen $\beta$ and a given $\mu_A$ under $H_A$ true, we have

\[
P(\bar{Y} > c|H_A \text{ true}) = P\left(\frac{\bar{Y} - \mu_A}{s/\sqrt{n}} > \frac{c - \mu_A}{s/\sqrt{n}}|H_A \text{ true}\right) = 1 - \beta
\]

\[
\Rightarrow P(t > \frac{(\mu_0 + t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}) - \mu_A}{s/\sqrt{n}}) = 1 - \beta
\]

\[
\Rightarrow P(t > t_{\alpha,n-1} - \frac{(\mu_A - \mu_0)}{s/\sqrt{n}}) = 1 - \beta
\]

\[
\Rightarrow t_{1-\beta,n-1} = t_{\alpha,n-1} - \frac{d}{s/\sqrt{n}}
\]

So, $t_{1-\beta,n-1} = t_{\alpha,n-1} - \frac{d}{s/\sqrt{n}} \Rightarrow n = \left(\frac{s(t_{\alpha,n-1} - t_{1-\beta,n-1})}{d}\right)^2$

Notice that ‘$n$’ is on both sides of the equation.
Refresher: sample size calculation for one-sample test with $\sigma^2$ unknown, find $n$

- One option is an iterative procedure . . .

0) We need a value for $s$, often this comes from a pilot study.

1) Assume we know $\sigma$ and start by using $z$-values.

$$n^{(0)} = \left( \frac{s(z_{\alpha} - z_{1-\beta})}{d} \right)^2$$

2) Plug into unknown $\sigma^2$ formula.

$$n^{(1)} = \left( \frac{s(t_{\alpha,n^{(0)}-1} - t_{1-\beta,n^{(0)}-1})}{d} \right)^2$$

3) Continuing to use the formula in 2), iterate until convergence.
Example (Sample size calculation in R)

###### one-sample hypothesis test #######
## Use iterative process for a one-sided test:
alpha <- 0.05
beta <- 0.10 # --> gives power of 90%
m.0 <- 4 # --> null mean
m.A <- 6 # --> alternative mean
s <- 3 # --> sample standard deviation

## Get initial n based on a known sigma, so use z-dist instead of t-distribution:
n.0 <- ((s*(qnorm(alpha,lower.tail=FALSE)-qnorm(1-beta,lower.tail=FALSE)))/(m.A-m.0))^2
n.0
[1] 19.26866
Refresher: sample size calculation for one-sample test with $\sigma^2$ unknown, using R

Example (Sample size calculation in R)

```r
############### one-sample hypothesis test #################
## Utilize sample size equation for unknown sigma
## and t-distribution to get next n.

n.1 <- ((s*(qt(alpha,n.0-1,lower.tail=FALSE)-qt(1-beta,n.0-1,lower.tail=FALSE)))/(m.A-m.0))^2
n.1
[1] 21.10018

## Iterate until convergence:
n.old <- n.0
n.new <- n.1
iter <- 1
```

---

12 / 48
Refresher: sample size calculation for one-sample test with $\sigma^2$ unknown, using R

Example (Sample size calculation in R)

```r
############## one-sample hypothesis test ##############
## Iterate until convergence:
while(abs(n.old-n.new)>=1){
  cat("iteration = ",iter,"\n")
  n.old <- n.new
  n.new <- ((s*(qt(alpha,n.old-1,lower.tail=FALSE)-
qt(1-beta,n.old-1,lower.tail=FALSE)))/(m.A-m.0))^2
  print(n.new)
  iter <- iter+1
}
# [1] 20.92262
Use n=21 as your sample size.
```

This was an iterative process when $\sigma^2$ unknown.
Sample size calculation for one-sample test with $\sigma^2$ unknown using Lenth applet

- U of Iowa Professor Russ Lenth has a widely used applet available for power and sample size calculations: [www.stat.uiowa.edu/~rlenth/Power/](http://www.stat.uiowa.edu/~rlenth/Power/)

Example (Lenth sample size website)

We can use these same parameters from the previous example in the applet to see that $n = 21$. 

![Applet screenshot](image.png)
Sample size calculation for one-sample test with $\sigma^2$ unknown using SAS

- Same sample size calculation done in SAS.

Example (SAS sample size)

```sas
options linesize = 79 nocenter nodate formchar = "|----|+|----+=|--|_/<>*" ;

proc power;
   onesamplemeans test=T sides=U /*Upper one-sided as H0:mu>mu.0 */
       alpha=0.05
       nullmean=4
       mean=6
       stddev=3
       power=0.9
       ntotal= . /* The period tells SAS what to solve for.*/
;
run;
```
Sample size calculation for one-sample test with $\sigma^2$ unknown using SAS

Example (SAS sample size)

```
The SAS System

The POWER Procedure
One-Sample t Test for Mean

Fixed Scenario Elements

Distribution     Normal
Method           Exact
Number of Sides  U
Null Mean        4
Alpha            0.05
Mean             6
Standard Deviation 3
Nominal Power   0.9

Computed N Total

Actual Power   N Total
0.904        21
```
Sample size calculation for Confidence Interval (CI) for one-sample scenario with $\sigma^2$ unknown

- $(1 - \alpha)100\%$ CI for $\mu$: $\bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$

  \[
  \text{half-width of CI or MOE}
  \]

- For a given MOE,

  \[
  n = \left( \frac{t_{\alpha/2, n-1} \cdot s}{\text{MOE}} \right)^2
  \]

  We can again use an iterative process to get $n$.

  *round-up to get $n$ at the end.
Sample size calculation for Confidence Interval (CI) for one-sample scenario with $\sigma^2$ unknown, using R

### Example (Sample size calculation in R for CI)

#### 95% confidence interval

```r
alpha=0.05
s=0.5
w=0.1  # half-width

n.0 <- (qnorm(alpha/2,lower.tail=FALSE)*s/w)^2
n.0
[1] 96.03647

n.1 <- (qt(alpha/2,n.0-1,lower.tail=FALSE)*s/w)^2
n.1
# [1] 98.52956
```
Sample size calculation for Confidence Interval (CI) for one-sample scenario with $\sigma^2$ unknown, using R

Example (Sample size calculation in R for CI)

####### 95% confidence interval #######

```r
n.old <- n.0
n.new <- n.1
while (abs(n.old-n.new)>=1){
  n.old <- n.new
  n.new <- (qt(alpha/2,n.old-1,lower.tail=FALSE)*s/w)^2
  print(n.new)
}
[1] 98.46465
```

Use n=99 as your sample size.
Sample size calculation for Confidence Interval (CI) for one-sample scenario with $\sigma^2$ unknown, using Lenth website

Example (Sample size calculation for CI using Lenth website)
Refresher: sample size calculation for two-sample $t$-test (equal sample sizes, constant variance)

- $H_o : \mu_1 - \mu_2 = 0$ vs. $H_A : \mu_1 - \mu_2 > 0$ (one-sided)

- Get cut-off to control type I error rate at $\alpha$ level:

\[ P(\bar{Y}_1 - \bar{Y}_2 > c|H_o \text{true}) = \alpha \]

\[
\Rightarrow P(t > \frac{c}{S_p \sqrt{2/n}}) = \alpha \\
\Rightarrow c = t_{\alpha,2n-2} \cdot S_p \sqrt{2/n}
\]

Then for given $\beta$:

\[ P(\bar{Y}_1 - \bar{Y}_2 > c|H_A \text{true}) = 1 - \beta \]

\[
\Rightarrow P(t > \frac{c - d}{S_p \sqrt{2/n}}) = 1 - \beta \\
(\text{where } d = \mu_1 - \mu_2)
\]
Refresher: sample size calculation for two-sample $t$-test
(equal sample sizes, constant variance)

$\Rightarrow \frac{t_{\alpha,2n-2} \cdot S_p \sqrt{\frac{2}{n}} - d}{S_p \sqrt{\frac{2}{n}}} = t_{1-\beta,2n-2}$

$\Rightarrow n = 2 \left( \frac{S_p(t_{\alpha,2n-2} - t_{1-\beta,2n-2})}{d} \right)^2$

- And again, an iterative procedure can be used.
Sample size calculation for two-sample $t$-test in R
(equal sample sizes, constant variance)

Example (Sample size calculation in R for 2-sample $t$-test)

```
############################################################################
### two-sample hypothesis test
############################################################################
## For a one-sided test:

alpha <- 0.05
beta <- 0.10  # (gives power of 90%)
d <- 2    ## difference between means.
s <- 1    ## same for both groups (pooled)

## Get initial n based on a known sigma:
n.0 <- 2*(s*(qnorm(alpha,lower.tail=FALSE)-
    qnorm((1-beta),lower.tail=FALSE))/d)^2

n.0
[1] 4.281924
```
Sample size calculation for two-sample \( t \)-test in R
(equal sample sizes, constant variance)

Example (Sample size calculation in R for 2-sample t-test)

```r
### two-sample hypothesis test ###
# Utilize equation for unknown sigma to get next n.
n.1 <- 2*(s*(qt(alpha,2*n.0-2,lower.tail=FALSE)-
         qt((1-beta),2*n.0-2,lower.tail=FALSE))/d)^2
n.1
[1] 5.572524

# Iterate until convergence:
n.old <- n.0
n.new <- n.1
iter <- 1
```
Sample size calculation for two-sample $t$-test in R
(equal sample sizes, constant variance)

Example (Sample size calculation in R for 2-sample t-test)

```r
#################################################
two-sample hypothesis test #################################################
while (abs(n.old-n.new)>=1){
    cat("iteration = ",iter,"\n")
    n.old <- n.new
    n.new <- 2*(s*(qt(alpha,2*n.old-2,lower.tail=FALSE)-
                    qt((1-beta),2*n.old-2,lower.tail=FALSE))/d)^2
    print(n.new)
    iter=iter+1
}
[1] 5.155601
```

Use n=6 as your sample size (6 in each group).
Sample size calculation for two-sample $t$-test, Lenth website (equal sample sizes, constant variance)

Example (Sample size calculation for 2-sample $t$-test, Lenth website)

For $n = 6$ power is 0.9420, for $n = 5$ power is 0.8916.
SAS allows you to consider a variety of differences in means and sample sizes at once to calculate power for each.

Example (Sample size calculation for 2-sample t-test in SAS)

```
proc power;
   twosamplemeans test=diff sides=2
      meandiff=1 2 3
      stddev=1
      npergroup=5 6 7
      power=.;
run;
```

```
The POWER Procedure
Two-Sample t Test for Mean Difference

Distribution          Normal
Method                Exact
Number of Sides       2
Standard Deviation    1
Null Difference       0
Alpha                 0.05

<table>
<thead>
<tr>
<th>Index</th>
<th>Diff</th>
<th>N per Group</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0.286</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>6</td>
<td>0.347</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7</td>
<td>0.406</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5</td>
<td>0.791</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>6</td>
<td>0.876</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>7</td>
<td>0.929</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>5</td>
<td>0.985</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>6</td>
<td>0.996</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>7</td>
<td>&gt;.999</td>
</tr>
</tbody>
</table>
```
Sample size for Confidence Interval (CI) on a contrast

- Recall for our contrast $\gamma = \sum_{i=1}^{a} c_i \mu_i$ we have the estimate:
  \[ \hat{\gamma} = \sum_{i=1}^{a} c_i \bar{Y}_i. \]

- 95% CI for $\gamma$: $\hat{\gamma} \pm t_{\alpha/2,N-a} \sqrt{\frac{MSE \sum_{i=1}^{a} \frac{c_i^2}{n_i}}{MOE}}$

- For a given half-width $MOE$, what sample size do we need?
  
  If we assume $n_i = n$ for all $i$ (i.e. balanced design), then
  
  $n = \left( \frac{t_{\alpha/2,N-a}}{MOE} \right)^2 \cdot MSE \cdot \sum_{i=1}^{a} c_i^2$ \quad where \quad $N = na$.

  Again, you can use an iterative procedure to find $n$. 
Power and Sample Size for 1-way ANOVA

- What about power for a 1-way ANOVA?

- We consider power in the overall F-test...

\[ H_0 : \mu_1 = \mu_2 = \ldots = \mu_a \quad \text{vs.} \quad H_A : \text{not } H_0 \]

Test statistic

\[ F_0 = \frac{\frac{SSE_{\text{red}} - SSE_{\text{full}}}{\Delta df}}{MSE_{\text{full}}} = \frac{\frac{SS_{\text{total}} - SSE_{\text{full}}}{(a-1)}}{MSE_{\text{full}}} \]

\[ = \frac{\frac{SS_{\text{trt}}}{(a-1)}}{MSE_{\text{full}}} = \frac{MS_{\text{trt}}}{MSE_{\text{full}}} \]

Reduced model: only 1 mean
Full model: \( a \) means
Under $H_0$ true, $F_o$ follows a ‘central $F$-distribution’ or 
$F_o \sim F_{(a-1,N-a,\phi=0)}$ where $\phi$ is the non-centrality parameter.

This information sets us up with a cut-off for significance for controlling the Type I error rate at the $\alpha$ level...
What about the Type II error rate? What about power?

As always, to compute power, we must provide the true state of the means under $H_A$ true.

For the two-sample t-test, we only needed to state the difference between the two means.

Now, with ‘a’ means, there are many different configurations of the means where $H_o$ is false.

This suggests we need to consider numerous configurations.
Under $H_A$ true, $F_o$ follows a ‘non-central $F$-distribution’ or $F_o \sim F_{(a-1,N-a,\phi)}$ where $\phi > 0$.

We establish a cut-off for significance $c = F_{(\alpha,a-1,N-a,\phi=0)}$ based on a given Type I error rate. Then, the power is the probability that $F_o$ is greater than or equal to $c$ under $H_A$ true.
What is the value of $\phi$? The non-centrality parameter?

This depends on the particular configuration of $a$ means.

$$\phi = \frac{\sum_{i=1}^{a} n_i \alpha_i^2}{\sigma^2}$$

where $\alpha_i = \mu_i - \mu$ and $\sum_i n_i \alpha_i = 0$

The non-centrality parameter $\phi$ measures how far the treatment means are from being equal relative to the squared-standard error of a single mean or $\frac{\sigma^2}{n_i}$

Recall,

$$E[MS_{TRT}] = \sigma^2 + \frac{\sum_{i=1}^{a} n_i \alpha_i^2}{a-1}$$

With $E[MSE] = \sigma^2$ we can see how $\phi$ works with these values to form our $F$-statistic for Treatment.
$$\phi = \sum_{i=1}^{a} \frac{n_i \alpha_i^2}{\sigma_i^2}$$

- Larger differences in means provides a larger $\phi$ and provides a larger power (all things being equal). In general...
  - $\phi$ increases if you increase sample sizes.
  - $\phi$ increases if the error variance is smaller.
  - $\phi$ increases if the means $\mu_i$ are farther apart.

![Central and non-central F(2,15,ncp=phi)](image-url)
Power and Sample Size for 1-way ANOVA

Non-central $F(2,15,\text{ncp}=\phi)$

$x$

$F(2,15,\text{ncp}=10)$

$F(2,15,\text{ncp}=20)$

$F(2,15,\text{ncp}=30)$

$x$
If there are three means, which configuration is easier to detect?

- **Scenario 1**
  - $\mu_1 = -1$
  - $\mu_2 = 0$
  - $\mu_3 = 1$

- **Scenario 2**
  - $\mu_1 = 0$
  - $\mu_2 = 0$
  - $\mu_3 = 2$

$$\phi = \frac{\sum_{i=1}^{a} n_i (\mu_i - \mu)^2}{\sigma^2}$$

Assuming $n_i = n$,

$$\phi_1 = \frac{n \sum_{i=1}^{a} (\mu_i - 0)^2}{\sigma^2} = \frac{n \cdot 2}{\sigma^2}$$

$$\phi_2 = \frac{n \sum_{i=1}^{a} (\mu_i - 2/3)^2}{\sigma^2} = \frac{n \cdot 2^2}{\sigma^2} \uparrow$$

Larger $\phi$ is easier to detect in scenario 2.
Example (Computing power for a given scenario)

\[ a = 3, \ n = 6, \ \alpha = 0.05, \ \sigma^2 = 0.075 \]

\[ \mu_1 = 2.82 \]
\[ \mu_2 = 3.89 \]
\[ \mu_3 = 3.04 \]

\[ \Rightarrow \mu = 3.25 \]

\[ \text{cut-off for significance is } F_{(0.05,2,15,\phi=0)} = 3.68 \]
\[ \alpha_1 = -0.43 \]
\[ \alpha_2 = 0.64 \]
\[ \alpha_3 = -0.21 \]
\[ \phi = \frac{6 \sum_{i=1}^{3} \alpha_i^2}{0.075} = 51.09 \]

Power = \[ P(F_{(2,15,\phi=51.09)} > 3.68) = 0.9999 \]

*Very small MSE relative to \( \alpha_i \)'s gave very large power.*
Power and Sample Size for 1-way ANOVA

Example (1-way ANOVA sample size in SAS)

```sas
/*Power for overall F-test.*/
proc power;
    onewayanova test=overall
      alpha=0.05
      groupmeans=2.82|3.89|3.04
      stddev=0.2739
      npergroup=6
      power=.
    ;
run;
```

The SAS System

The POWER Procedure
Overall F Test for One-Way ANOVA

Fixed Scenario Elements

<table>
<thead>
<tr>
<th>Method</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Group Means</td>
<td>2.82  3.89  3.04</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.2739</td>
</tr>
<tr>
<td>Sample Size per Group</td>
<td>6</td>
</tr>
</tbody>
</table>

Computed Power

<table>
<thead>
<tr>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;.999</td>
</tr>
</tbody>
</table>
Software will compute power for you.

Depending on the software you use, it may ask for something that doesn’t seem like the non-centrality parameter described above, but is actually a slight transformation of it.

R will ask for ‘between group variance’ and ‘within group variance’ instead of the non-centrality parameter.

\[
\phi = \frac{n \sum_{i=1}^{a} \alpha_i^2}{\sigma^2} = n(a - 1) \left( \frac{\sum (\mu_i - \mu)^2}{a-1} \right) = n(a - 1) \frac{\text{between group variance}}{\text{within group variance}}
\]

So, ‘between group variance’ = \(\sum \frac{\alpha_i^2}{a-1}\) and ‘within group variance’ = \(\sigma^2\).
In the previous example, ‘between group variance’ = \( \frac{\sum \alpha_i^2}{a-1} = 0.3193 \), and ‘within group variance’ = 0.075.

### Example (1-way ANOVA sample size in R)

#### power in 1-way ANOVA

```r
> power.anova.test(groups=3, n=6, between.var=0.3193, within.var=0.075, sig.level=0.05)
```

Balanced one-way analysis of variance power calculation

groups = 3
n = 6
between.var = 0.3193
within.var = 0.075
sig.level = 0.05
power = 0.999976

NOTE: n is number in each group
The Lenth website requests $SD[TREATMENT]$ which is $\sqrt{\text{between group variance}}$ and $SD[WITHIN]$ which is $\sqrt{\text{within group variance}}$.

Example (1-way ANOVA sample size Lenth applet)
Example (1-way ANOVA power calculation)

Another example...

What power will I have with the following set-up?

\[ a = 4, \ n = 5, \ \alpha = 0.05, \ \sigma^2 = 1.2 \]

\[ \mu_1 = 9.5 \]
\[ \mu_2 = 10.3 \]
\[ \mu_3 = 11 \]
\[ \mu_4 = 12.4 \]

Get the non-centrality parameter... find power.
Power and Sample Size for 1-way ANOVA

Checking with the Lenth applet: $\hat{\sigma} = 1.096$ and $\sqrt{\frac{\sum \alpha_i^2}{a-1}} = 1.230$

Example (1-way ANOVA power calculation)
Power and Sample Size for 1-way ANOVA

Checking with the R: \( \sigma^2 = 1.2 \) and \( \frac{\sum \alpha_i^2}{a-1} = 1.51 \)

Example (1-way ANOVA power calculation in R)

```r
############# power in 1-way ANOVA #########################
> power.anova.test(groups=4,n=5,between.var=1.51,
                   within.var=1.2,sig.level=0.05)

Balanced one-way analysis of variance power calculation

  groups = 4
  n = 5
  between.var = 1.51
  within.var = 1.2
  sig.level = 0.05
  power = 0.9111074

NOTE: n is number in each group
Power and Sample Size for 1-way ANOVA

\[ \phi = \frac{\sum_{i=1}^{a} n(\mu_i - \mu)^2}{\sigma^2} \]

- In the previous example with 4 groups, we had \( \mu_1 = 9.5, \mu_2 = 10.3, \mu_3 = 11, \) and \( \mu_4 = 12.4 \) which lead to \( \sum \alpha_i^2 = 4.54. \)

- Actually, any configuration of means giving \( \sum \alpha_i^2 = 4.54 \) with all the same other values (\( n, \sigma, \) type I error rate) will have the same power.

- When working with clients, they may be more comfortable hypothesizing what the range of \( \mu_i \)'s is rather than specifically stating what each \( \mu_i \) is.

Let \( D = \mu_{max} - \mu_{min} \) and assume \( n_i = n \) (balanced design).
Given that the range of the means is $D$, what do you think is the ‘most favorable’ configuration of means? (i.e. the easiest to detect, or reject $H_0$).

Given that the range of the means is $D$, what do you think is the ‘least favorable’ configuration of means? (i.e. the hardest to detect, or lowest power).
To make power analysis useful, you must be able to specify some scientifically or practically meaningful set of alternative means, and you must be able to guess as to how large the error variance is $\sigma^2$.

Perhaps the client can think about... “If this were true, I would want to know about it,” and then design for those interesting alternatives.

Examples:
- The doubling of a mutation rate is practically significant, so I want to design for that.
- An increase in MPG of 1 is relevant, so I will design for that.
- A decrease in time-to-pain-relief of 20 minutes is meaningful to the patient, so I will design for that.

Many granting agencies (NSF, NIH, etc.) require a power analysis before funding a proposal.
Rejection of the overall F-test can arise from any departure from the null hypothesis of equal means.

In some situations, you may be interested in one or two particular contrasts (and less interested in others).

You can design an experiment with the power of rejection of a given contrast in mind. $H_0 : \gamma = 0$ vs. $H_A : \gamma \neq 0$ where $\gamma = \sum_{i=1}^{a} c_i \mu_i$ and $\hat{\gamma} = \sum_{i=1}^{a} c_i \bar{Y}_i$.

Under $H_0$ true, $F = \frac{SS_{\text{contrast}}}{MSE} \sim F_{1,N-a}$

The non-centrality parameter under $H_A$ true is

$$\phi = \frac{(\sum_{i=1}^{a} c_i \alpha_i)^2}{\sigma^2 \sum_{i=1}^{a} \left( \frac{c_i^2}{n_i} \right)} = \frac{(\sum_{i=1}^{a} c_i \mu_i)^2}{\sigma^2 \sum_{i=1}^{a} \left( \frac{c_i^2}{n_i} \right)}$$

And you find the power $P(F_{1,N-a,\phi} > F_{\alpha,1,N-a})$.