Power & Sample Size Calculation
Power & Sample Size Calculation

For long or expensive experiments one is nowadays expected to demonstrate that the proposed experiment will not be a waste of time and money. In particular, this means showing that your sample size (i.e. number of experiment units) is

- neither so small that scientifically interesting effects will be swamped by random variation
- nor larger than necessary, wasting resources

Priorities:
It is more important to do the right experiment than to do the right sample size calculation!
Errors and Power in Hypothesis Testing

- A **Type I error** occurs when $H_0$ is true but is rejected.
- A **Type II error** occurs when $H_0$ is false but is accepted.
- The (significance) level of a test is the chance of making a Type I error.

$$\text{test-level} = P(\text{making type I error}|H_0 \text{ is true}) = \tau$$

- The **power** of the test is the chance of rejecting $H_0$ when $H_a$ is true:

$$\text{power} = 1 - P(\text{making type II error}|H_0 \text{ is false}) = 1 - \beta$$

$$= P(\text{correctly reject } H_0|H_0 \text{ is false})$$

- A good test should have small test-level and large power.

<table>
<thead>
<tr>
<th></th>
<th>$H_a$ is rejected (H$_0$ is accepted)</th>
<th>$H_0$ is rejected (H$_a$ is accepted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ is true</td>
<td>$\sqrt{}$</td>
<td>$\tau = P(\text{Type I error})$</td>
</tr>
<tr>
<td>$H_0$ is false</td>
<td>$\beta = (\text{Type II error})$</td>
<td>$\sqrt{}$</td>
</tr>
</tbody>
</table>
Power Calculation for ANOVA

For a CRD design with $g$ treatments, recall

the means model: $y_{ij} = \mu_i + \varepsilon_{ij}$

and the effects model: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$.

in both models we assume $\varepsilon_{ij}$'s are i.i.d. $\mathcal{N}(0, \sigma^2)$, for $i = 1, \ldots, g$, and $j = 1, \ldots, n_i$. The two models are related via the relationship

$\mu_i = \mu + \alpha_i$.

So far, we use the means model more often. But to find the power of the ANOVA test for treatment effects, we have to use the effects model and $\mu$ has to be defined as

$$\mu = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \mu_i = \frac{1}{N} \sum_{i=1}^{g} n_i \mu_i.$$ 

which implies that $\sum_{i=1}^{g} n_i \alpha_i = 0$. Here $N = \sum_{i=1}^{g} n_i$ is the total number of experimental units.

Chapter 7 - 4
The null and alternative hypotheses of the ANOVA $F$ test are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_g \quad \text{v.s.}$$

$$H_a : \mu_i \text{'s not all equal},$$

in terms of the means model, or equivalently in terms of the effects model are

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_g = 0 \quad \text{v.s.}$$

$$H_a : \alpha_i \text{'s not all 0}.$$

Recall we reject $H_0$ at level $\alpha$ if the test statistic $F = \frac{MS_{Trt}}{MSE}$ exceeds the critical value $F_{\alpha, g-1, N-g}$. So

$$\text{Power} = P(\text{Reject } H_0 \mid H_a \text{ is true}) = P\left(F > F_{\alpha, g-1, N-g}\right).$$

To find $P\left(F > F_{\alpha, g-1, N-g}\right)$, we must know the distribution of $F$.

- What is the distribution of $F$ under $H_0$? $F_{g-1, N-g}$.
- And under $H_a$?

Chapter 7 - 5
For $F = \frac{MS_{Trt}}{MSE} = \frac{SS_{Trt}/(g - 1)}{MSE/(N - g)}$, recall in Chapter 3 we’ve shown that

- No matter under $H_0$ or $H_a$, it is always true that
  \[ \frac{SSE}{\sigma^2} \sim \chi^2_{N-g}, \quad \text{and} \quad \mathbb{E}(MSE) = \sigma^2 \]

- For the numerator, $SS_{Trt}/\sigma^2$ has a $\chi^2_{g-1}$ distribution under $H_0$. But under $H_a$, it has a \textbf{non-central Chi-square distribution} $\chi^2_{g-1,\delta}$ with non-central parameter
  \[ \delta = \sum_{i=1}^{g} n_i \alpha_i^2 \]

Moreover,

\[ \mathbb{E}(SS_{Trt}) = (g - 1)\sigma^2 + \sum_{i=1}^{g} n_i (\mu_i - \mu)^2 \delta \sigma^2 \]

\[ = \begin{cases} 
(g - 1)\sigma^2 & \text{under } H_0 \\
(g - 1 + \delta)\sigma^2 & \text{under } H_a 
\end{cases} \]
Non-Central $F$-Distribution

Under $H_a$, it can be shown that

$$F = \frac{MS_{Trt}}{MSE}$$

has a non-central $F$-distribution on degrees of freedom $g - 1$ and $N - g$, with non-centrality parameter $\delta$, denoted as

$$F \sim F_{g-1, N-g, \delta} \quad \text{where} \quad \delta = \frac{\sum_{i=1}^{g} n_i \alpha_i^2}{\sigma^2}.$$

Recall that

$$\alpha_i = \mu_i - \mu, \quad \text{and} \quad \mu = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \mu_i = \frac{1}{N} \sum_{i=1}^{g} n_i \mu_i.$$
Example 1: Power Calculation (1)

- \( g = 5 \) treatment groups
  - group sizes: \( n_1 = n_2 = n_3 = 5, n_4 = 6, n_5 = 4 \)
- assume \( \sigma = 0.8 \)
- desired test level \( \alpha = 0.05 \)
- find the power of the test when \( H_a \) is true with

\[
\mu_1 = 1.6, \mu_2 = 0.6, \mu_3 = 2, \mu_4 = 0, \mu_5 = 1.
\]

**Solution.**

The grand mean \( \mu \) is

\[
\mu = \frac{\sum_{i=1}^{g} n_i \mu_i}{N} = \frac{5 \times 1.6 + 5 \times 0.6 + 5 \times 2 + 6 \times 0 + 4 \times 1}{5 + 5 + 5 + 6 + 4} = \frac{25}{25} = 1
\]

The \( \alpha_i \)'s are thus \( \alpha_i = \mu_i - \mu = \mu_i - 1: \)

\[
\alpha_1 = 0.6, \alpha_2 = -0.4, \alpha_3 = 1, \alpha_4 = -1, \alpha_5 = 0.
\]
Example 1: Power Calculation (2)

The non-centrality parameter is

\[ \delta = \sum_{i=1}^{g} \frac{n_i \alpha_i^2}{\sigma^2} = \frac{5 \times 0.6^2 + 5 \times (-0.4)^2 + 5 \times 1^2 + 6 \times (-1)^2 + 4 \times 0^2}{0.8^2} \]

\[ = \frac{13.6}{0.64} = 21.25 \]

So

\[ F = \frac{MS_{Trt}}{MSE} \sim \begin{cases} 
F_{g-1, N-g} = F_{4, 25-5} & \text{under } H_0 \\
F_{g-1, N-g, \delta} = F_{4, 25-5, 21.25} & \text{under } H_a
\end{cases} \]
Example 1: Power Calculation (3)

To keep the test level at $\alpha = 0.05$, the critical value to reject $H_0$ must be $F_{0.05,4,25}$.

```r
> qf(.95, 4, 20)  # not qf(0.05, 4, 20)!
[1] 2.866081
```

Then when $H_0$ is true, the chance that $H_0$ is rejected is only 0.05 (the red shaded area.)

Remark: Don’t confuse the following two

- In most STAT books, $F_{\alpha, df_1, df_2}$ means the upper tail has an area of $\alpha$.
- In R, `qf(alpha, df1, df2)` means the lower tail has an area of $\alpha$.  

Chapter 7 - 10
Example 1: Power Calculation (4)

If $H_a$ is true and

\[ \mu_1 = 1.6, \mu_2 = 0.6, \mu_3 = 2, \mu_4 = 0, \mu_5 = 1, \]

we know then $F \sim F_{4,20,21.5}$. The power to reject $H_0$ is the area under the density of $F_{4,20,21.5}$ beyond the critical value (the blue shaded area,) which is 0.9249.

\[
> 1 - \text{pf(qf(.95,4,20),4,20, ncp=21.25)}
\]

[1] 0.9249342

In the R codes, `ncp` means the “non-centrality parameter.”
Remark

The power of a test is not a single value, but a function of the parameters in $H_a$. If parameter $\mu_i$’s change their values, the power of the test also changes. It makes no sense to talk about the power of a test without specifying the parameters in $H_a$. 
Example 2: Sample Size Calculation (1)

- $g = 5$ treatment groups of equal sample size $n_i = n$ for all $i$
- assume $\sigma = 0.8$
- desired test level $\alpha = 0.05$
- What is the minimal sample size $n$ per treatment to have power $0.95$ when

  $\alpha_1 = 0.5, \alpha_2 = -0.5, \alpha_3 = 1, \alpha_4 = -1, \alpha_5 = 0$?

**Solution.** The non-centrality parameter is

$$\delta = \frac{1}{\sigma^2} \sum_{i=1}^{g} n_i \alpha_i^2 = \frac{n}{0.8^2} \left[ 0.5^2 + (-0.5)^2 + 1^2 + (-1)^2 + 0^2 \right]$$

$$ = \frac{n}{0.8^2} \times 2.5 = 3.9n$$

So

$$F = \frac{MS_{Trt}}{MSE} \sim \begin{cases} F_{g-1, N-g} &= F_{4, 5n-5} \quad \text{under } H_0 \\ F_{g-1, N-g, \delta} &= F_{4, 5n-5, 3.9n} \quad \text{under } H_a \end{cases}$$
Recall the critical value $F^*$ for rejecting $H_0$ at level $\alpha = 0.05$ is

$$F^* = F_{\alpha, g-1, N-g} = F_{0.05, 4, 5n-5}$$

which can be find in R via the command

```r
> qf(1-alpha, g-1, N-g)  # syntax
> qf(1-0.05, 4, 5*n-5)    # sub-in the values
```

By definition,

$$\text{Power} = P(\text{reject } H_0 \mid H_a \text{ is true}) = P(\text{the non central } F \text{ statistic } \geq F^*) = P(F(g-1, N-g, \delta) \geq F^*) = P(F(4, 5n-5, 3.9n) \geq F^*)$$

which can be found in R via the command

```r
> 1 - pf(qf(1-alpha, g-1, N-g), g-1, N-g, ncp=delta)  # syntax
> 1 - pf(qf(1-0.05, 4, 5*n-5), 4, 5*n-5, ncp=3.9*n)  # sub-in values
```

In the R codes, `ncp` means the “non-centrality parameter.”

Chapter 7 - 14
Now we find the R code to find the power of the ANOVA $F$-test when $n$ is known. Let’s plug in different values of $n$ and see what is the smallest $n$ to make power $\geq 0.95$.

\[
> 1 - \text{pf}\left(\text{qf}(1-\text{alpha}, g-1, N-g), g-1, N-g, \text{ncp}=\delta)\right)
\]

\[
> n = 5
> 1 - \text{pf}\left(\text{qf}(1-0.05, 4, 5*n-5), 4, 5*n-5, \text{ncp}=3.9*n\right)
\]

\[
[1] 0.8994675 \quad \# \text{less than 0.95, not high enough}
\]

\[
> n = 6
> 1 - \text{pf}\left(\text{qf}(1-0.05, 4, 5*n-5), 4, 5*n-5, \text{ncp}=3.9*n\right)
\]

\[
[1] 0.9578791 \quad \# \text{greater than 0.95, bingo!}
\]

So we need 6 replicates in each of the 5 treatment groups to ensure a power of 0.95 when

\[
\alpha_1 = 0.5, \; \alpha_2 = -0.5, \; \alpha_3 = 1, \; \alpha_4 = -1, \; \alpha_5 = 0.
\]

**Remark:** Again, we also have to specify the alternative hypothesis $H_a$ fully to to find the appropriate sample size.
But $\sigma^2$ is Unknown . . .

As $\sigma^2$ is usually unknown, here are a few ways to make a guess.

▶ Make a small-sample pilot study to get an estimate of $\sigma$.

▶ Based on prior studies or knowledge about the experimental units, can you think of a range of plausible values for $\sigma$? If so, choose the biggest one.

▶ You could repeat the sample size calculations for various levels of $\sigma$ to see how it affects the needed sample size.
How to Specify the $H_a$?

As the power of a test depends on the alternative hypothesis $H_a$, that is, the $\alpha_i$’s, one might has to try several sets of $\alpha_i$’s to find the appropriate sample size. But how many $H_a$’s we have to try?

Here is a useful trick.

1. Suppose we would be interested if any two means differed by $D$ or more.
2. The smallest value of $\delta$ in this case is when two means differ by exactly, $D$, and the other $g - 2$ means are halfway between.
3. So try $\alpha_1 = D/2$, $\alpha_2 = -D/2$, and $\alpha_i = 0$ for all other $\tau$.

Assuming equal sample sizes, the non-centrality parameter is

$$\delta = \sum_i \frac{n\alpha_i^2}{\sigma^2} = \frac{n(D^2/4 + D^2/4)}{\sigma^2} = \frac{nD^2}{2\sigma^2}.$$