Chapter 7 Power & Sample Size Calculation

Yibi Huang

Chapter 7	Power & Sample Size Calculation for CRDs
Section 10.3	Power & Sample Size for Factorial Designs

Power & Sample Size Calculation

When proposing an experiment (applying for funding etc), nowadays one needs to show that the proposed sample size (i.e. the number of experiment units) is

- neither so small that scientifically interesting effects will be swamped by random variation (i.e., has enough power to reject a false H_0)
- nor larger than necessary, wasting resources (time & money)

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, i.e., the chance to reject a H_0 when it is true.
- The *power* of the test is the the chance of rejecting H_0 when H_a is true:

$$\begin{aligned}\text{power } &= 1 - \mathrm{P}(\text{making type II error}|H_0 \text{ is false}) = 1 - \beta \\ &= \mathrm{P}(\text{correctly reject } H_0|H_0 \text{ is false})\end{aligned}$$

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

	H_a is rejected (H_0 is accepted)	H_0 is rejected (H_a is accepted)
H_0 is true	$\surd$	$\alpha = \mathrm{P}(\text{Type } I \text{ error})$
H_0 is false	$\beta = \mathrm{P}(\text{Type } II \text{ error})$	$\surd$

Power Calculation for ANOVA

For a CRD design with g treatments, recall

$$\text{the means model: } y_{ij} = \mu_i + \varepsilon_{ij}$$
$$\text{and the effects model: } y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

for $i = 1, \ldots, g$, and $j = 1, \ldots, n_i$.

The two models are related via the relationship

$$\mu_i = \mu + \alpha_i.$$

For CRDs, we use the means model more often. However, for power and sample size calculation, we need to use the **effects model** and μ must 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.

Recall the H_0 and H_a for the ANOVA F test are

$$H_0: \ \mu_1 = \cdots = \mu_g \quad \text{v.s.} \quad H_a: \ \mu_i\text{'s not all equal,}$$

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

$$H_0: \ \alpha_1 = \cdots = \alpha_g = 0 \quad \text{v.s.} \quad H_a: \ \text{Not all } \alpha_i\text{'s are 0.}$$

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

$$\text{Power} = \mathrm{P}(\text{Reject } \mathrm{H}_0 | H_a \text{ is true}) = \mathrm{P}(F > F_{\alpha,\, g-1,\, N-g}).$$

To find $\mathrm{P}(F > F_{\alpha,\, g-1,\, N-g})$, 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?

For $F = \frac{MS_{Trt}}{\text{MSE}} = \frac{SS_{Trt}/(g-1)}{\text{SSE}/(N-g)}$, recall in Ch3 we've shown that

- no matter under H_0 or H_a, it is always true that

$$\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{N-g}, \quad \text{and} \quad \mathbb{E}(\text{MSE}) = \sigma^2$$

- For the numerator, SS_{Trt}/σ^2 has a χ^2_{g-1} distribution under H_0. But under H_a, it has a **non-central Chi-square distribution $\chi^2_{g-1,\delta}$ with non-central parameter**

$$\delta = \frac{\sum_{i=1}^{g} n_i \alpha_i^2}{\sigma^2}.$$

Moreover,

$$\mathbb{E}(SS_{Trt}) = (g-1)\sigma^2 + \underbrace{\sum_{i=1}^{g} n_i \alpha_i^2}_{\delta\sigma^2}$$

$$= \begin{cases} (g-1)\sigma^2 & \text{under H}_0 \\ (g-1+\delta)\sigma^2 & \text{under H}_a \end{cases}$$

Non-Central F-Distribution

Under H_a, it can be shown that

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

has a **non-central F-distribution** on degrees of freedom $g-1$ and $N-g$, with **non-centrality parameter δ**, 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

$$\alpha_i = \mu_i - \mu, \quad \text{and} \quad \mu = \frac{1}{N}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\mu_i = \frac{1}{N}\sum_{i=1}^{g} n_i\mu_i.$$

Non-central F-distribution is also right skewed, and the greater the non-centrality parameter δ, the further away the peak of the distribution from 0.

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 significance 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 μ 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 α_i's are thus $\alpha_i = \mu_i - \mu = \mu_i - 1$:

$$\alpha_1 = 0.6,\ \alpha_2 = -0.4,\ \alpha_3 = 1,\ \alpha_4 = -1,\ \alpha_5 = 0.$$

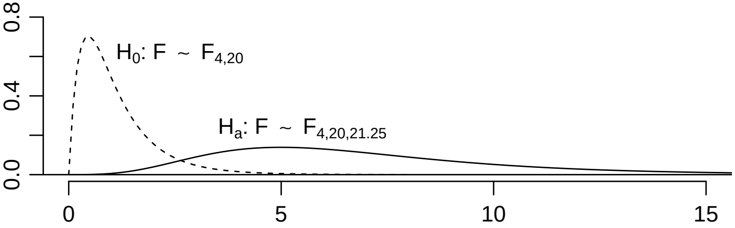
Example 1: Power Calculation (2)

The non-centrality parameter is

$$\delta = \frac{\sum_{i=1}^{g} n_i \alpha_i^2}{\sigma^2} = \frac{5 \times 0.6^2 + 5 \times (-0.4)^2 + 5 \times 1^2 + 6 \times (-1)^2 + 4 \times 0^2}{0.8^2}$$
$$= \frac{13.6}{0.64} = 21.25$$

So

$$F = \frac{MS_{Trt}}{\mathrm{MSE}} \sim \begin{cases} F_{g-1,\, N-g} = F_{4,\, 25-5} & \text{under } \mathrm{H}_0 \\ F_{g-1,\, N-g,\delta} = F_{4,\, 25-5,\, 21.25} & \text{under } \mathrm{H}_a \end{cases}$$

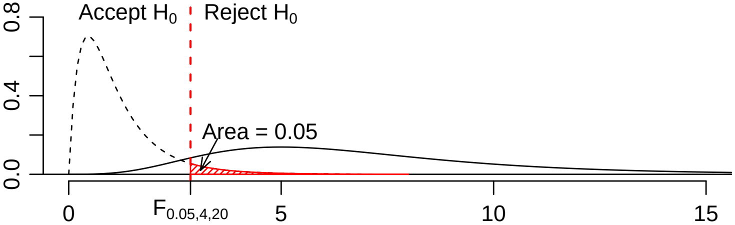


Example 1: Power Calculation (3)

The critical value to reject H_0 keeping the significance level at $\alpha = 0.05$ is $F_{0.05,4,20} \approx 2.866$.

```
> qf(0.05,4,20, lower.tail=F)
[1] 2.866081
```

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



Reminder: Don't confuse the following two:

- in most STAT books, $F_{\alpha,df1,df2}$ means the area of the *upper* tail is α.
- in R, `qf(alpha, df1, df2)` means the the area of the *lower* tail is α.

Example 1: Power Calculation (4)

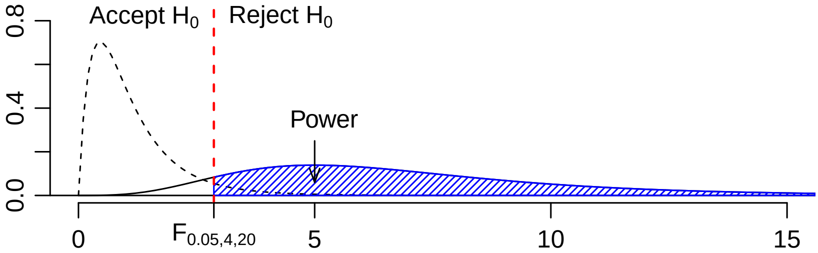
If H_a is true and

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

we know then $F \sim F_{4,20,\delta=21.25}$. The power to reject H_0 is the area under the density of $F_{4,20,\delta=21.25}$ beyond the critical value (the blue shaded area,) which is 0.9249.

```
> pf(qf(.95,4,20),4,20, ncp=21.25, lower.tail=F)
[1] 0.9249342
```

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



Example 1: Power Calculation (5)

Remark

The **power** of a test is not a single value, but a function of the parameters in H_a. If parameter μ_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

Power Of A Test Is Affected By ...

The larger the non-centrality parameter

$$\delta = \frac{\sum_{i=1}^{g} n_i \alpha_i^2}{\sigma^2},$$

the greater the power.

The power of a test will increase if

- the number of replicate n_i per treatment increases
- the magnitude of treatment effects α_i's increase (under the constraint $\sum_i n_i \alpha_i = 0$)
- the size of noise σ^2 decreases.

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 significance level $\alpha = 0.05$
- Assuming the design is balanced that all groups have identical sample size n, 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

$$\begin{aligned}\delta = \frac{\sum_{i=1}^{g} n_i \alpha_i^2}{\sigma^2} &= \frac{n}{0.8^2}\ [0.5^2 + (-0.5)^2 + 1^2 + (-1)^2 + 0^2] \\ &= \frac{n}{0.8^2} \times 2.5 = 3.9n\end{aligned}$$

So

$$F = \frac{MS_{Trt}}{\text{MSE}} \sim \begin{cases} F_{g-1,\,N-g} = F_{4,\,5n-5} & \text{under } \mathrm{H}_0 \\ F_{g-1,\,N-g,\delta} = F_{4,\,5n-5,\,3.9n} & \text{under } \mathrm{H}_a \end{cases}$$

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

```
> F.crit = qf(alpha, g-1, N-g, lower.tail=F)   # syntax
> F.crit = qf(0.05, 4, 5*n-5, lower.tail=F)    # sub-in the values
```

By definition,

$$\begin{aligned} \textit{Power} &= \mathrm{P}(\textit{reject } H_0 \,|H_a \textit{ is true}) \\ &= \mathrm{P}(\textit{the non central F statistic } \geq F^*) \\ &= \mathrm{P}(F(g-1, N-g, \delta) \geq F^*) \\ &= \mathrm{P}(F(4, 5n-5, 3.9n) \geq F^*) \end{aligned}$$

which can be found in R via the command

```
> pf(F.crit, g-1, N-g, ncp=delta, lower.tail=F)  # syntax
> pf(F.crit, 4, 5*n-5, ncp=3.9*n, lower.tail=F)  # sub-in values
```

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

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 ≥ 0.95.

```
> F.crit = qf(alpha, g-1, N-g, lower.tail=F)
> pf(F.crit, g-1, N-g, ncp=delta, lower.tail=F)  # syntax

> n = 5
> F.crit = qf(0.05, 4, 5*n-5, lower.tail=F)
> pf(F.crit, 4, 5*n-5, ncp=3.9*n, lower.tail=F)
[1] 0.8994675        # less than 0.95, not high enough

> n = 6
> F.crit = qf(0.05, 4, 5*n-5, lower.tail=F)
> pf(F.crit, 4, 5*n-5, ncp=3.9*n, lower.tail=F)
[1] 0.9578791        # 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 have to specify the H_a fully to to find the appropriate sample size.

But σ^2 is Unknown ...

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

- Make a small-sample pilot study to get an estimate of σ.
- Based on prior studies or knowledge about the experimental units, can you think of a range of plausible values for σ? If so, choose the biggest one.
- You could repeat the sample size calculations for various levels of σ 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 α_i's, one might has to try several sets of α_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 δ 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 τ. Assuming equal sample sizes, the non-centrality parameter is

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

Power and Sample Size Calculation for Complete Block Designs

For complete block designs

$$
\begin{aligned}
y_{ij} &= \mu + \alpha_i + \beta_j + \varepsilon_{ij} && \text{RCBD} \\
y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} && \text{Latin Square} \\
y_{ijk\ell} &= \mu + \alpha_i + \beta_j + \gamma_k + \delta_\ell + \varepsilon_{ijk\ell} && \text{Graeco-Latin Square}
\end{aligned}
$$

where α_i's are the treatment effect, and $\beta_j, \gamma_k, \delta_\ell$ are effects of blocking factors, when H_a is true, treatments have different effects, the ANOVA F-statistic also has a non-central F-distribution

$$F = \frac{\mathrm{MS}_{trt}}{\mathrm{MSE}} \sim \begin{cases} F_{g-1,\,\text{df of MSE}} & \text{under } \mathrm{H}_0 \\ F_{g-1,\,\text{df of MSE},\delta} & \text{under } \mathrm{H}_a \end{cases}$$

where the non-centrality parameter δ is

$$\delta = \frac{r \sum_i \alpha_i^2}{\sigma^2}, \quad \text{where } r = \#\text{ of replicates per treatment}$$

Note that the df changes from $N - g$ to the df of MSE.

Power and Sample Size Calculation for Balanced Factorial Designs

- Section 10.3 in Oehlert's book
- We may ignore factorial structure and compute power and sample size for the overall null hypothesis of no model effects.
- We will address methods for balanced data only because most factorial experiments are designed to be balanced, and simple formulae Power for balanced data for noncentrality parameters exist only for balanced data.
- For factorial data, we usually test H_0 about main effects or interactions in addition to the overall H_0 of no model effects.

Non-Centrality Parameter for Balanced Factorial Designs

Ex, consider a 3-way factorial design

$$y_{ijk\ell} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \beta\gamma_{jk} + \alpha\gamma_{ij} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijk\ell}$$

to calculate the power of the F-test of a main effect or an interaction effect under H_a, the non-centrality parameter can be calculated as follows:

- Square the factorial effects (using the zero-sum constraint) and sum them,
- Multiply this sum by the total number of data in the design divided by the number of levels in the effect, and
- Divide that product by the error variance.

E.g., consider a $a \times b \times c$ factorial design with r replicates per treatment

$$y_{ijk\ell} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \beta\gamma_{jk} + \alpha\gamma_{ij} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijk\ell}$$

To test about the B main effects

$$\mathrm{H}_0: \ \beta_1 = \cdots = \beta_b = 0 \quad \text{v.s.} \quad \mathrm{H}_a: \ \text{not all } \beta_j \text{ are } 0$$

the ANOVA F-statistic for B main-effect is

$$F = \frac{\mathrm{MS}_B}{\mathrm{MSE}} \sim \begin{cases} F_{b-1, \text{df of MSE}} & \text{under } \mathrm{H}_0 \\ F_{b-1, \text{df of MSE}, \delta} & \text{under } \mathrm{H}_a \end{cases}$$

where the non-centrality parameter δ is

$$\delta = \left(\frac{N}{b} \sum\nolimits_j \beta_j^2\right) \Big/ \sigma^2 = \left(rac \sum\nolimits_j \beta_j^2\right) \Big/ \sigma^2.$$

Here $N = abcr$ is the total number of observations.

E.g., consider a $a \times b \times c$ factorial design with r replicates per treatment

$$y_{ijk\ell} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \beta\gamma_{jk} + \alpha\gamma_{ij} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijk\ell}$$

To test about the AB interactions

$$\mathrm{H}_0: \text{ all } \alpha\beta_{ij} = 0 \quad \text{v.s.} \quad \mathrm{H}_a: \text{ not all } \alpha\beta_{ij} = 0$$

the ANOVA F-statistic for AB interactions

$$F = \frac{\mathrm{MS}_{AB}}{\mathrm{MSE}} \sim \begin{cases} F_{(a-1)(b-1),\, \text{df of MSE}} & \text{under } \mathrm{H}_0 \\ F_{(a-1)(b-1),\, \text{df of MSE}, \delta} & \text{under } \mathrm{H}_a \end{cases}$$

where the non-centrality parameter δ is

$$\delta = \left(\frac{N}{ab} \sum\nolimits_{ij} (\alpha\beta_{ij})^2\right) \Big/ \sigma^2 = \left(rc \sum\nolimits_{ij} (\alpha\beta_{ij})^2\right) / \sigma^2.$$

Here $N = abcr$ is the total number of observations.

Sample Size Should Be The Last Thing To Consider

In practice, sample size calculation is the *last* thing to do in designing an experiment. Consider the followings before calculating the required sample size.

- asking the right question,
- choosing an appropriate response,
- thinking about what factors need to be blocked on,
- choosing a right design,
- setting up the right procedure to recruit subjects, and so on,

All the above are more important than doing the right sample size calculation!

Sometimes, choosing a better design can substantially reduces the required sample size.