STAT 222 Lecture 7
Power & Sample Size Calculation
Section 3.6

Yibi Huang

Power & Sample Size Calculation for ANOVA F-tests

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 noise (i.e., unable to reject a false H_0)
- nor larger than necessary, which is a waste of 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 failing to reject a false H_0
- 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:

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

A good test has a small significance level and a large power.

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

Recall the model for multi-sample data

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij}\text{'s are i.i.d. } N(0, \sigma^2)$$

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

The H_0 and H_a for the ANOVA F test are

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

Recall the model for multi-sample data

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij}\text{'s are i.i.d. } N(0, \sigma^2)$$

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

The H_0 and H_a for the ANOVA F test are

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

Recall we reject H_0 if the F-statistic $F = \frac{MS_{trt}}{\mathrm{MSE}}$ exceeds some critical value. The *power* of the ANOVA F-test is hence

$$\text{Power} = \mathrm{P}(\text{Reject } H_0 \,|\, H_a \text{ is true}) = \mathrm{P}(F > \text{critical value} \,|\, H_a \text{ is true}).$$

Need to know the distribution of F to calculate the power.

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

Non-Central F-Distribution

Under H_a: μ_i's not all equal, it can be shown that

$$F = \frac{\mathrm{MS}_{trt}}{\mathrm{MSE}} = \frac{\mathrm{SS}_{trt}/(g-1)}{\mathrm{SSE}/(N-g)}$$

has a ***non-central F-distribution*** on degrees of freedom $g-1$ and $N-g$, with ***non-centrality parameter*** δ^2, denoted as

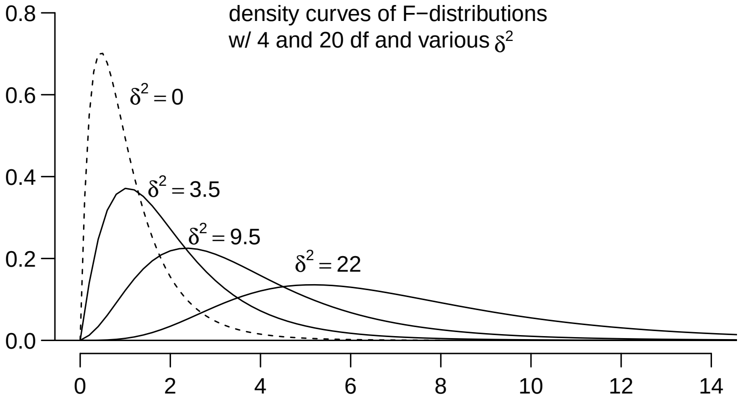
$$F \sim F_{g-1,N-g,\delta^2} \quad \text{where} \quad \delta^2 = \frac{\sum_{i=1}^{g} n_i(\mu_i - \mu)^2}{\sigma^2}.$$

where

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

Non-Central F-Distribution

Non-central F-distribution is also right skewed, and the greater the non-centrality parameter δ^2, the further away the peak of the distribution is 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

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

Sol. The grand mean μ is

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

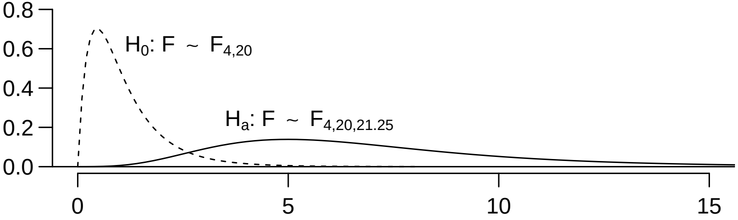
Example 1: Power Calculation (2)

The non-centrality parameter is

$$\begin{aligned}\delta^2 &= \frac{\sum_{i=1}^{g} n_i(\mu_i - \mu)^2}{\sigma^2} \\ &= \frac{5(1.6-1)^2 + 5(0.6-1)^2 + 5(2-1)^2 + 6(0-1)^2 + 4(1-1)^2}{0.8^2} \\ &= \frac{13.6}{0.64} = 21.25\end{aligned}$$

So

$$F = \frac{MS_{trt}}{\text{MSE}} \sim \begin{cases} F_{g-1,\,N-g} = F_{4,\,25-5} & \text{under H}_0 \\ F_{g-1,\,N-g,\delta^2} = F_{4,\,25-5,\,21.25} & \text{under H}_a \end{cases}$$

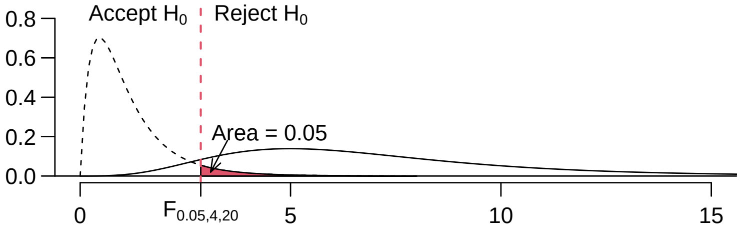


Example 1: Power Calculation (3)

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

```
qf(0.05,4,20, lower.tail=F)
[1] 2.866
```

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



Example 1: Power Calculation (4)

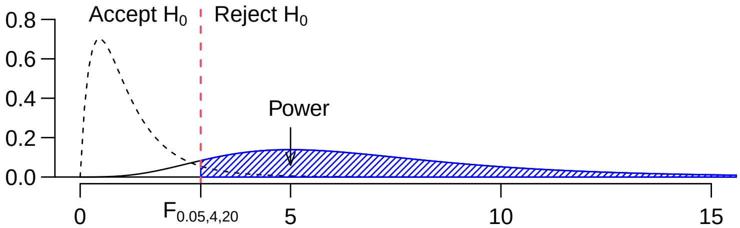
If H_a is true and

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

we know then $F \sim F_{4,20,\delta^2=21.25}$. The power to reject H_0 is the area under the density of $F_{4,20,\delta^2=21.25}$ 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] 0.9249
```

In R, `ncp` stands for the "non-centrality parameter."



Power Depends On the Parameters in H_a

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, the power of the test also changes. Cannot talk about the power of a test without specifying the parameters in H_a

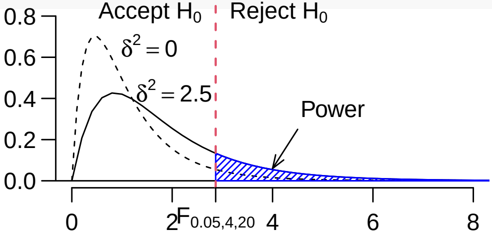
Ex. For H_a: $\mu_1 = 1.4,\ \mu_2 = 0.6,\ \mu_3 = \mu_4 = \mu_5 = 1$, then

$$\mu = \frac{5 \cdot 1.4 + 5 \cdot 1 + 5 \cdot 0.6 + 6 \cdot 1 + 4 \cdot 1}{5+5+5+6+4} = 1$$

$$\delta^2 = \frac{5(1.4-1)^2 + 5(0.6-1)^2 + 5(1-1)^2 + 6(1-1)^2 + 4(1-1)^2}{0.8^2} = 2.5$$

```
pf(qf(.95,4,20),4,20, ncp=2.5, lower.tail=F)
[1] 0.1713
```

Power is 0.1713.



Power of a Test Is Affected By . . .

The larger the non-centrality parameter

$$\delta^2 = \frac{\sum_{i=1}^{g} n_i(\mu_i - \mu)^2}{\sigma^2},$$

the greater the power.

The power of a test will increase if

- the number of replicate n_i per treatment increases
- the difference of treatment means $\mu_i - \mu$'s increase
- 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 equal sample size n in all groups, what is the minimal sample size n per treatment to have power 0.95 when

$$\mu_1 = 0.5,\ \mu_2 = -0.5,\ \mu_3 = 1,\ \mu_4 = -1,\ \mu_5 = 0?$$

Sol. Can calculate that $\mu = \frac{1}{N}\sum_{i=1}^{g} n\mu_i = 0$. The non-centrality parameter is

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

So
$$F = \frac{MS_{trt}}{\text{MSE}} \sim \begin{cases} F_{g-1,\,N-g} = F_{4,\,5n-5} & \text{under } \mathrm{H}_0 \\ F_{g-1,\,N-g,\delta^2} = F_{4,\,5n-5,\,3.9n} & \text{under } \mathrm{H}_a \end{cases}$$

Recall the critical value F^* for rejecting H_0 at level $\alpha = 0.05$ is

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

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,

$$\begin{aligned}
Power &= \mathrm{P}(\textit{reject } H_0 \,|H_a \textit{ is true}) \\
&= \mathrm{P}(\textit{the non central F statistic } \geq F^*) \\
&= \mathrm{P}(F(g-1, N-g, \delta^2) \geq F^*) \\
&= \mathrm{P}(F(4, 5n-5, 3.9n) \geq F^*)
\end{aligned}$$

which can be found in R via the command

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

In 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, 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] 0.8995
```

Power is 0.8995, less than 0.95. $n = 5$ is 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] 0.9579
```

Greater than 0.95, bingo!

So we need 6 replicates in each of the 5 groups to ensure a power of 0.95 when $\mu_1 = 0.5,\ \mu_2 = -0.5,\ \mu_3 = 1,\ \mu_4 = -1,\ \mu_5 = 0.$
Remark: Again, must fully specify H_a to calculate 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^2.
- Based on prior studies or knowledge about the experimental units, can you think of a range of plausible values for σ^2? If so, choose the biggest one.
- You could repeat the sample size calculations for various levels of σ^2 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 μ_i's, one might has to try several sets of μ_i's to find the appropriate sample size. But how many H_a's we have to try?

- μ_i's only affects power through $\delta^2 = \sum_{i=1}^{g} \frac{n_i(\mu_i - \mu)^2}{\sigma^2}$. Identical power if two sets of μ_i's have identical δ^2 values.

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^2 in this case is when two means differ by exactly, D, and the other $g-2$ means are halfway between.

So try $\mu_1 = D/2$, $\mu_2 = -D/2$, and $\mu_i = 0$ for all other groups. Assuming equal sample sizes, μ would be 0 and the non-centrality parameter is

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