STA305/1004-Class19

Today's Class

- Sample size for ANOVA
- Factorial designs at two levels
- Cube plots
- Calculation of factorial effects

Sample size for ANOVA - Designing a study to compare more than two treatments

Compare more than 2 means

e.g. $H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_i \neq \mu_j$

- Consider the hypothesis that k means are equal vs. the alternative that at least two differ.
- What is the probability that the test rejects if at least two means differ?
- Power $= 1 - P(\text{Type II error})$ is this probability.

Sample size for ANOVA - Designing a study to compare more than two treatments

The null and alternative hypotheses are:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \text{ vs. } H_1 : \mu_i \neq \mu_j.$$

The test rejects at level α if

$$MS_{Treat}/MS_E \geq F_{k-1,N-K,\alpha}.$$

The power of the test is

$$1 - \beta = P\left(MS_{Treat}/MS_E \geq F_{k-1,N-K,\alpha}\right),$$

when H_0 is false.

Sample size for ANOVA - Designing a study to compare more than two treatments

When H_0 is false it can be shown that:

- MS_{Treat}/σ^2 has a non-central Chi-square distribution with $k-1$ degrees of freedom and non-centrality parameter δ.
- $\boxed{MS_{Treat}/MS_E}$ has a non-central F distribution with the numerator and denominator degrees of freedom $k-1$ and $N-k$ respectively, and non-centrality parameter

between variance

within variance

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

where n_i is the number of observations in group i, $\bar{\mu} = \sum_{i=1}^{k} \mu_i/k$, and σ^2 is the within group error variance .

This is dentoted by $F_{k-1,N-k}(\delta)$.

Direct calculation of Power

- The power of the test is

$$P\left(F_{k-1,N-k}(\delta) > F_{k-1,N-K,\alpha}\right).$$

as σ gets larger then power $\uparrow$

- The power is an increasing function δ
- The power depends on the true values of the treatment means μ_i, the error variance σ^2, and sample size n_i.
- If the experimentor has some prior idea about the treament means and error variance the sample size (number of replications) that will guaruntee a pre-assigned power of the test.

Blood coagulation example - sample size

Suppose that an investigator would like to replcate the blood coagulation study with only 3 animals per diet. In this case $k = 4, n_i = 3$. The treatment means from the initial study are:

Diet	A	B	C	D
Average	61	66	68	61
	μ_1	μ_2	μ_3	μ_4

```
anova(lm.diets)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## diets      3    228    76.0  13.571 4.658e-05 ***
## Residuals 20    112     5.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\rightarrow MS_E \quad \sqrt{5.6} \approx \sigma$

Blood coagulation example - sample size

- $\overbrace{\mu_1 = 61,\ \mu_2 = 66,\ \mu_3 = 68,\ \mu_4 = 61}^{\text{from previous study on coagulation}}$.
- The error variance σ^2 was estimated as $MS_E = 5.6$.
- Assuming that the estimated values are the true values of the parameters, the non-centrality parameter of the F distribution is

$$\delta = 3 \times \left((61-64)^2 + (66-64)^2 + (68-64)^2 + (61-64)^2\right)/5.6 = 20.35714$$

Blood coagulation example - sample size

If we choose $\alpha = 0.05$ as the significance level then $F_{3,20,0.05} = 3.0983912$. The power of the test is then

$$P\,(\overbrace{F_{3,20}(20.36) > 3.10}) = 0.94.$$

→ event test rejects

This was calculated using the CDF for the F distribution in R `pf()`.

```
1-pf(q = 3.10,df1 = 3,df2 = 20,ncp = 20.36)
```

```
## [1] 0.9435208
```

critic value

power

∴ study has 94% power to detect that at least one pair of means differ

Calculating power and sample size using the pwr library

- There are several libraries in R which can calculate power and sample size for statistical tests. The library `pwr()` has a function
- `pwr.anova.test(k = NULL, n = NULL, f = NULL, sig.level = 0.05, power = NULL)`

for computing power and sample size.

- k Number of groups
- n Number of observations (per group)
- f Effect size
- The effect size is the square root of the non-centrality parameter of the non-central F distribution.

$$f = \sqrt{\frac{\sum_{i=1}^{k} n_i (\mu_i - \bar{\mu})^2}{\sigma^2}}.$$

where n_i is the number of observations in group i, $\bar{\mu} = \sum_{i=1}^{k} \mu_i / k$, and σ^2 is the within group error variance.

Calculating power and sample size using the pwr library

as $\alpha \downarrow$ power $\downarrow$

all the properties of power are still true!

In the previous example $\delta = 20.35714$ so $f = \sqrt{20.35714} = 4.5118887$.

```
library(pwr)
pwr.anova.test(k = 4,n = 3,f = 4.5)
```

H_1 is true $f=4.5$

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 4
##               n = 3
##               f = 4.5
##       sig.level = 0.05
##           power = 1
##
## NOTE: n is number in each group
```

3 units per group

4 groups

$\therefore$ total sample size is $4 \times 3 = 12$

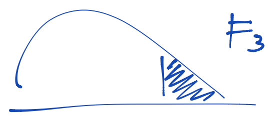
when H_0 is true $f=0$

then power $= 0.05$

When H_0 is true

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ is true

$$\delta = \frac{\sum_{i=1}^{4} n_i (\mu_i - \bar{\mu})^2}{\sigma^2}$$



$F_{3,20}$

$= 0$

$P(F_{3,20}(0) > F_{3,20}) = 0.05$

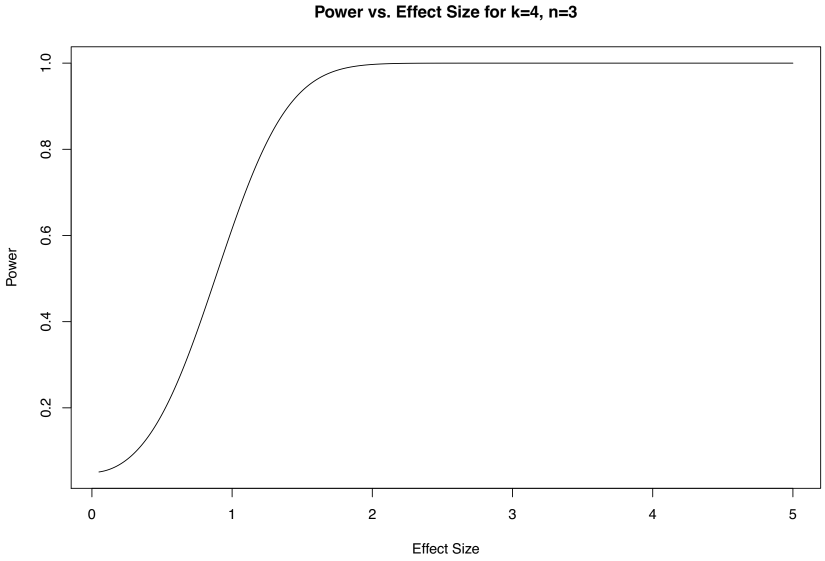
$=$ type I error rate

- Which ANOVA study should have a greater type II error rate (assume both studies with use the same significance level)?
  - A study that is designed to detect an effect size of 0.5
  - — — — — — 0.75

Study 1: $1-\beta_1$ $\qquad 1-\beta_2 > 1-\beta_1$

Study 2: $1-\beta_2$ $\qquad \Rightarrow \beta_1 > \beta_2$

Calculating power and sample size using the pwr library



Calculating power using simulation

The general procedure for simulating power is:

1. Use the underlying model to generate random data with (a) specified sample sizes, (b) parameter values that one is trying to detect with the hypothesis test, and (c) nuisance parameters such as variances.
2. Run the estimation program (e.g., `t.test()`,`lm()`) on these randomly generated data.
3. Calculate the test statistic and p-value.
4. Do Steps 1–3 many times, say, N, and save the p-values. The estimated power for a level alpha test is the proportion of observations (out of N) for which the p-value is less than alpha.

Calculating power using simulation

One of the advantages of calculating power via simulation is that we can investigate what happens to power if, say, some of the assumptions behind one-way ANOVA are violated.

pwr.anova.test
assumes that data is normally distributed

Calculating power using simulation - R program

```
#Simulate power of ANOVA for three groups

NSIM <- 1000 # number of simulations
res <- numeric(NSIM) # store p-values in res

mu1 <- 2; mu2 <- 2.5;mu3 <- 2  # true mean values of treatment groups
sigma1 <- 1; sigma2 <- 1; sigma3 <- 1 #variances in each group
n1 <- 40; n2 <- 40; n3 <- 40 #sample size in each group

for (i in 1:NSIM) # do the calculations below N times
  {
# generate sample of size n1 from N(mu1,sigma1^2)
y1 <- rnorm(n = n1,mean = mu1,sd = sigma1)
# generate sample of size n2 from N(mu2,sigma2^2)
y2 <- rnorm(n = n2,mean = mu2,sd = sigma2)
# generate sample of size n3 from N(mu3,sigma3^2)
y3 <- rnorm(n = n3,mean = mu3,sd = sigma3)
y <- c(y1,y2,y3) # store all the values from the groups
# generate the treatment assignment for each group
trt <- as.factor(c(rep(1,n1),rep(2,n2),rep(3,n3)))
m <- lm(y~trt) # calculate the ANOVA
res[i] <- anova(m)[1,5] # p-value of F test
}
sum(res<=0.05)/NSIM # calculate p-value
```

```
## [1] 0.613
```

Example of a factorial design

Suppose that an investigator is interested in examining three components of a weight loss intervention. The three components are:

1. Keeping a food diary (yes/no)
2. Increasing activity (yes/no)
3. Home visit (yes/no)

Factorial designs

The investigator plans to investigate all $2x2x2 = 2^3 = 8$ combinations of experimental conditions.

The experimental conditions will be.

Expt condition	Keep food diary	Increase physical activity	Home visit	weight loss
1	No	No	No	y_1
2	No	No	Yes	y_2
3	No	Yes	No	y_3
4	No	Yes	Yes	y_4
5	Yes	No	No	y_5
6	Yes	No	Yes	y_6
7	Yes	Yes	No	y_7
8	Yes	Yes	Yes	y_8

Factorial designs at two levels

- To perform a factorial design, you select a fixed number of levels of each of a number of factors (variables) and then run experiments in all possible combinations.

Factorial designs at two levels

- The factors can be quantitative or qualitative.
- Two levels of a quantitative variable could be two different temperatures or two different concentrations.
- Qualitative factors might be two types of catalysts or the presence and absence of some entity.

Factorial designs at two levels

Is a study where patients randomly assigned to one of four groups:

1. Aspirin only (beta-carotene placebo)
2. Beta carotene only (aspirin placebo)
3. Aspirin and beta carotene
4. Neither (both aspirin & beta carotene placebos) a factorial design.

YES

Factor 1 Aspirin	Factor 2 Beta-carotene	
N	N	(4)
Y	N	(1)
N	Y	(2)
Y	Y	(3)

\# factors

2^2 design = 4 possible expt. conditions

\# levels

∴ the study involves all factor-level combinations

it's a factorial design

Factorial design

The notation 2^3 identifies: - the number of factors (3) - the number of levels of each factor (2) - how many experimental conditions are in the design ($2^3 = 8$)

Factorial experiments can involve factors with different numbers of levels.

Factorial design

$2+2+1=5$

Consider a $4^2 \times 3^2 \times 2$ design.

1. How many factors?
2. How many levels of each factor?
3. How many experimental conditions (runs)?

2 factors at 4 levels
2 factors at 3 levels
1 factor at 2 levels

$16 \times 9 \times 2 = 288$ expt. conditions

END

Difference between ANOVA and Factorial Designs

In ANOVA the objective is to compare the individual experimental conditions with each other. In a factorial experiment the objective is generally to compare combinations of experimental conditions.

Let's consider the food diary study above. What is the effect of keeping a food diary?

Expt condition	Keep food diary	Increase physical activity	Home visit	weight loss
1	No	No	No	y_1
2	No	No	Yes	y_2
3	No	Yes	No	y_3
4	No	Yes	Yes	y_4
5	Yes	No	No	y_5
6	Yes	No	Yes	y_6
7	Yes	Yes	No	y_7
8	Yes	Yes	Yes	y_8

We can estimate the effect of food diary by comparing the mean of all conditions where food diary is set to NO (conditions 1-4) and mean of all conditions where food diary set to YES (conditions 5-8). This is also called the **main effect** of food diary, the adjective *main* being a reminder that this average is taken over the levels of the other factors.

Difference between ANOVA and Factorial Designs

Expt condition	Keep food diary	Increase physical activity	Home visit	weight loss
1	No	No	No	y_1
2	No	No	Yes	y_2
3	No	Yes	No	y_3
4	No	Yes	Yes	y_4
5	Yes	No	No	y_5
6	Yes	No	Yes	y_6
7	Yes	Yes	No	y_7
8	Yes	Yes	Yes	y_8

The main effect of food diary is:

$$\frac{y_1 + y_2 + y_3 + y_4}{4} - \frac{y_5 + y_6 + y_7 + y_8}{4}.$$

The main effect of physical activity is:

$$\frac{y_1 + y_2 + y_5 + y_6}{4} - \frac{y_3 + y_4 + y_7 + y_8}{4}.$$

The main effect of home visit is:

$$\frac{y_1 + y_3 + y_5 + y_7}{4} - \frac{y_2 + y_4 + y_6 + y_8}{4}.$$

Factorial designs at two levels

To perform a factorial design:

1. Select a fixed number of levels of each factor.
2. Run experiments in all possible combinations.

Factorial designs at two levels

- We will discuss designs where there are just two levels for each factor.
- Factors can be quantitative or qualitative.
- Two levels of quantitative variable could be two different temperatures or concentrations.
- Two levels of a quantitative variable could be two different types of catalysts or presence/absence of some entity.

Pilot plant investigation - example of factorial design

A pilot plant invsetiagtion employed a 2^3 factorial design (Box, Hunter, and Hunter (2005)) with

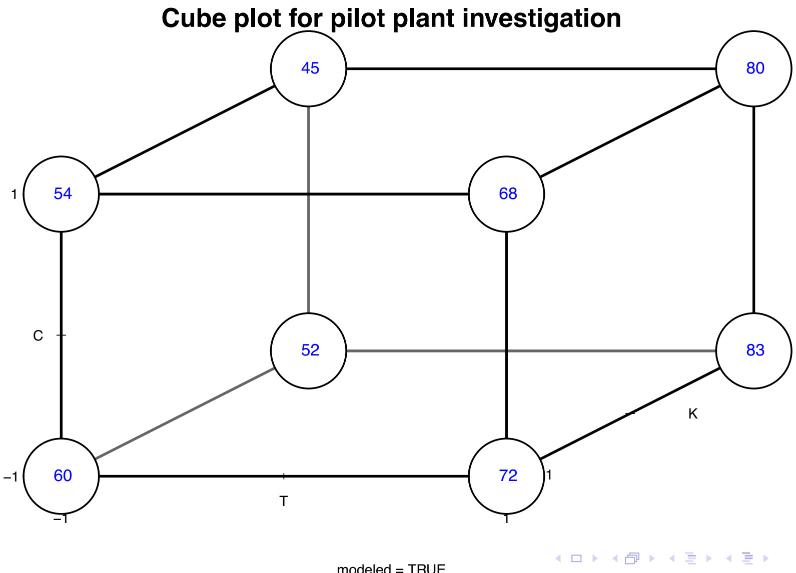
Factors	level 1	level 2
Temperature	$160\mathrm{C}^\circ$(-1)	$180\mathrm{C}^\circ(+1)$
Concentration	20% (-1)	40% $(+1)$
Catalyst	A (-1)	B$(+1)$

run	T	C	K	y
1	-1	-1	-1	60
2	1	-1	-1	72
3	-1	1	-1	54
4	1	1	-1	68
5	-1	-1	1	52
6	1	-1	1	83
7	-1	1	1	45
8	1	1	1	80

- Each data value recorded is for the response yield y averaged over two duplicate runs.

Cube plots

```
library("FrF2")
bhh54 <- lm(y~T*C*K,data=tab0502)
cubePlot(bhh54,"T","K","C",main="Cube plot for pilot plant investigation")
```



Cube plots

- 8 run design produces 12 comparisons
- Each edge of cube only one factor changed while other 2 held constant.
- Therefore experimenter that believes in only changing one factor at a time is satisfied.