STA461: Two-Factor Studies with Equal Sample Sizes

Lynn Lin

October 26, 2018

Introduction

- So far, we have only considered one treatment factor

- However, often we have two or more treatment factors

- Now we are focusing on the analysis of variance for two-factor studies where

  - the factors are crossed

  - equal sample sizes

Examples

- Weight loss clinic: for a given client

    - Have a choice of several diets

    - A choice of several exercise regimens

    - Different diets may result in different average weight loss

    - Average weight loss may depend on exercise programs

    - Effectiveness of a given diet may depend on the exercise regimen being followed by a client (interaction effect of diet and exercise on weight loss)

Example

- We could analyze one factor at a time, i.e., hold one factor constant at one level, and then alter the levels of the second factor (one-factor-at-a-time approach)

Example

- We could analyze one factor at a time, i.e., hold one factor constant at one level, and then alter the levels of the second factor (one-factor-at-a-time approach)

- But we are not able to assess the extent of an interaction effect of diet and exercise on average weight loss

- Advantages to two-factor studies:

  - we can alter both factor levels at the same time

  - we can assess interaction terms, or interactions between the two factors

Definitions

- In a **two-factor experiment**, we are interested in the effects of two factors, and possible interaction effects of the two factors, on a response variable

- In a **factorial design**, all combinations of levels of the factors are included in the experiments

  - Suppose the first factor, factor A has a levels, and the second factor, factor B has b levels, then we have $a \times b$ treatments altogether

- An experimental design is **balanced** if there is the same number of observations for each combination of factors (treatments)

- An experiment is **completely randomized** if experimental units are randomly assigned to combinations of factors

- **Two-way analysis of variance** is the classical, parametric, approach to analyzing the results of a two-factor experiment that has a balanced, completely randomized, factorial design

Factor effects model

$$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}$$

Partitioning the degrees of freedom

- Total sample size is n

- Total treatments is ab

- $SSTot$: $n - 1$

- SST: $ab - 1$

- SSE: $n - ab$

- Breakdown df for SST:

    - SSA: $a - 1$

    - SSB: $b - 1$

    - $SSAB$: $ab - 1 - (a - 1) - (b - 1) = (a - 1)(b - 1)$

Mean Squares

- Take the sum of squares and divide by their corresponding degrees of freedom

$$SS/df$$

Example

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
ToothGrowth$dose = factor(ToothGrowth$dose, levels=c(0.5,1.0,2.0),
                                   labels=c("low","med","high"))
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "low","med","high": 1 1 1 1 1 1 1 1 1 1 ...
```

Check for balanced design
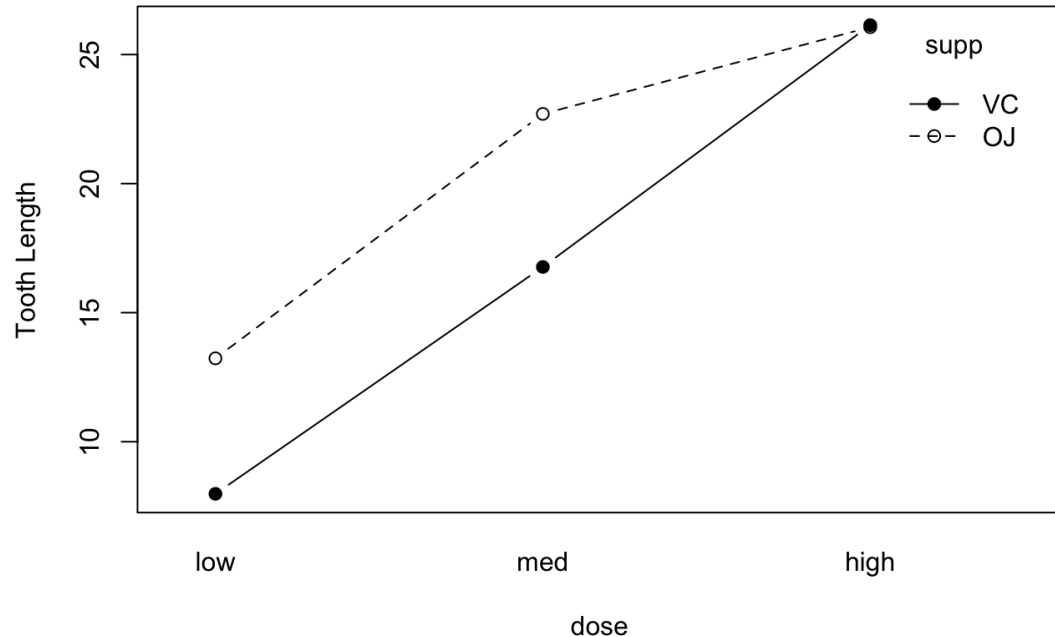
```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      low med high
##   OJ  10  10   10
##   VC  10  10   10
```

Boxplot

```
boxplot(len ~ supp * dose, data=ToothGrowth,
        ylab="Tooth Length", main="Boxplots of Tooth Growth Data")
```

Interaction plot

```
with(ToothGrowth, interaction.plot(x.factor=dose, trace.factor=supp,
                    response=len, fun=mean, type="b", legend=T,
                    ylab="Tooth Length",
                    pch=c(1,19)))
```

Numerical summaries

```
with(ToothGrowth, tapply(len, list(supp,dose), mean))
```

```
##       low    med   high
## OJ 13.23 22.70 26.06
## VC  7.98 16.77 26.14
```

```
with(ToothGrowth, tapply(len, list(supp,dose), var))
```

```
##        low        med       high
## OJ 19.889 15.295556   7.049333
## VC  7.544  6.326778 23.018222
```

Numerical summaries

```
aov.out = aov(len ~ supp * dose, data=ToothGrowth)
model.tables(aov.out, type="means", se=T)
```

```
## Tables of means
## Grand mean
##
## 18.81333
##
##   supp
## supp
##      OJ     VC
## 20.663 16.963
##
##   dose
## dose
##     low    med   high
## 10.605 19.735 26.100
##
##   supp:dose
##      dose
## supp low    med    high
##   OJ 13.23 22.70 26.06
```

ANOVA table

```
summary(aov.out)
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## supp           1  205.4   205.4  15.572 0.000231 ***
## dose           2 2426.4  1213.2  92.000  < 2e-16 ***
## supp:dose      2  108.3    54.2   4.107 0.021860 *
## Residuals     54  712.1    13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Checking assumptions

```
par(mfrow = c(2,2))
plot(aov.out)
```

Example

```
TukeyHSD(aov.out)
```

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = len ~ supp * dose, data = ToothGrowth)
##
## $supp
##       diff       lwr       upr      p adj
## VC-OJ -3.7 -5.579828 -1.820172 0.0002312
##
## $dose
##             diff       lwr       upr    p adj
## med-low    9.130  6.362488 11.897512 0.0e+00
## high-low 15.495 12.727488 18.262512 0.0e+00
## high-med  6.365  3.597488  9.132512 2.7e-06
##
## $`supp:dose`
##                     diff        lwr       upr      p adj
## VC:low-OJ:low    -5.25 -10.048124 -0.4518762 0.0242521
## OJ:med-OJ:low     9.47   4.671876 14.2681238 0.0000046
## VC:med-OJ:low     3.54  -1.258124  8.3381238 0.2640208
```