STAT461 HW6

Xiangyu Ren

10/9/2020

#### Problem 1. Define the following terms:

#### a. treatment

Combinations of factor levels are called treatments, where factor, also called an independent variable, is an explanatory variable manipulated by the experimenter.

#### b. contrast

A contrast is a linear combination of variables such that all the coefficients add up to zero. One way to think of it is as a set of weighted variables.

#### c. replicate

Replication is the repetition of an experimental condition so that the variability associated with the phenomenon can be estimated.

#### d. confounder

A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association.

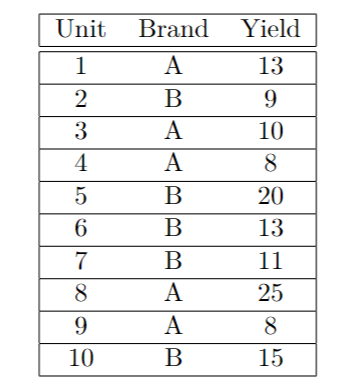
#### e. response

A response variable, also known as a dependent variable, is a concept, idea, or quantity that someone wants to measure.

#### f. generalizability

Generalizability is a measure of how well a researcher thinks their experimental results from a sample can be extended to the population as a whole.

#### Problem 2.



#### a. Enter the above data into R.

Unit = 1:10  
Brand = c("A", "B", "A", "A", "B", "B", "B", "A", "A", "B")   
Yield = c(13, 9, 10, 8, 20, 13, 11, 25, 8, 15)  
data1 = data.frame(Unit, Brand, Yield)  
data1

## Unit Brand Yield  
## 1 1 A 13  
## 2 2 B 9  
## 3 3 A 10  
## 4 4 A 8  
## 5 5 B 20  
## 6 6 B 13  
## 7 7 B 11  
## 8 8 A 25  
## 9 9 A 8  
## 10 10 B 15

#### b. Run the one-way ANOVA model in R.

library(lsmeans)

## Loading required package: emmeans

## The 'lsmeans' package is now basically a front end for 'emmeans'.  
## Users are encouraged to switch the rest of the way.  
## See help('transition') for more information, including how to  
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.

model = aov(Yield ~ Brand, data = data1)  
anova(model)

## Analysis of Variance Table  
##   
## Response: Yield  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Brand 1 1.6 1.60 0.0467 0.8343  
## Residuals 8 274.0 34.25

lsm.data = lsmeans(model, ~Brand)  
lsm.data

## Brand lsmean SE df lower.CL upper.CL  
## A 12.8 2.62 8 6.76 18.8  
## B 13.6 2.62 8 7.56 19.6  
##   
## Confidence level used: 0.95

#### c. What is the least squares estimate for the contrast ? Compute the least square estimate in R.

contrast1 = c(1, -1)  
contrastlist = list("tau\_A-tau\_B" = contrast1)  
contrast(lsm.data, contrastlist)

## contrast estimate SE df t.ratio p.value  
## tau\_A-tau\_B -0.8 3.7 8 -0.216 0.8343

#### Problem 3. Consider the one-way ANOVA model where and . Are the following contrasts estimable and why? If so, what are the least-squares estimate?

#### a.

, we get

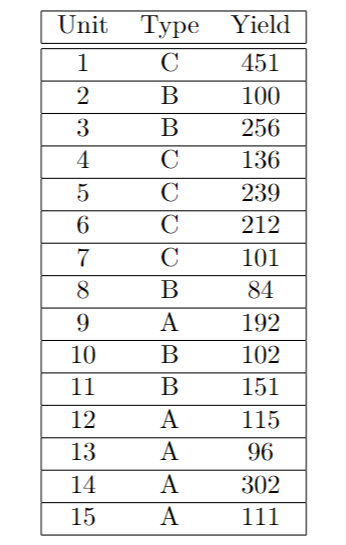
#### b.

This cannot be estimated because we cannot find any in to estimate it.

#### c.

we get

#### Problem 4.



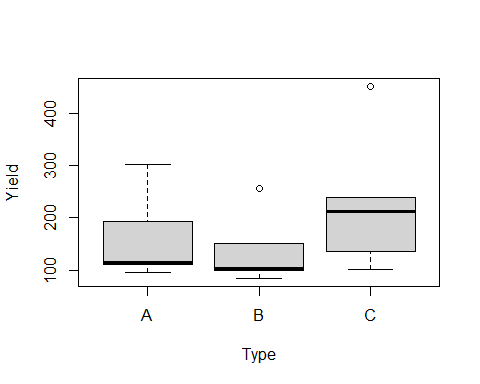
#### a. Input the data into R.

Unit = 1:15  
Type = c("C","B","B","C","C","C","C","B","A","B","B","A","A","A","A")   
Yield = c(451,100,256,136,239,212,101,84,192,102,151,115,96,302,111)   
data2 = data.frame(Unit, Type, Yield)  
data2

## Unit Type Yield  
## 1 1 C 451  
## 2 2 B 100  
## 3 3 B 256  
## 4 4 C 136  
## 5 5 C 239  
## 6 6 C 212  
## 7 7 C 101  
## 8 8 B 84  
## 9 9 A 192  
## 10 10 B 102  
## 11 11 B 151  
## 12 12 A 115  
## 13 13 A 96  
## 14 14 A 302  
## 15 15 A 111

#### b. Create boxplots for each Type.

boxplot(Yield ~ Type, data = data2, main = "")



#### c.Say we fit a one-way ANOVA model to this data. What are the values of the following and in words, describe what the term is.

model2=aov(Yield ~ Type, data = data2)   
anova(model2)

## Analysis of Variance Table  
##   
## Response: Yield  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Type 2 21225 10612 1.0261 0.3878  
## Residuals 12 124105 10342

lsm2=lsmeans(model2,"Type")  
lsm2

## Type lsmean SE df lower.CL upper.CL  
## A 163 45.5 12 64.1 262  
## B 139 45.5 12 39.5 238  
## C 228 45.5 12 128.7 327  
##   
## Confidence level used: 0.95

sum(163, 139, 228) / 3

## [1] 176.6667

is the grand mean for the anova model,

is the ls estimate for type A,

is the ls estimate for type B,

is the ls estimate for type C,

#### d. Use the lsmeans package to compute .

contrast2 = c(1, -1/2, -1/2)  
contrastList2 = list("tau\_A-1/2(tau\_B+tau\_C)" = contrast2)   
contrast(lsm2, contrastList2)

## contrast estimate SE df t.ratio p.value  
## tau\_A-1/2(tau\_B+tau\_C) -20 55.7 12 -0.359 0.7258

#### e.Say we want to test the hypothesis that all treatments (A, B, and C) have the same treatment mean. Write down the null hypothesis.

The null hypothesis is given by the following,

#### f. What is the reduced model under ?

The reduced model is

#### g. Using R, test this hypothesis at an level.

library(knitr)  
summary(contrast(lsm2, method = "pairwise", adjust = "tukey"), infer = c(T, T), level = 0.95, side = "two-sided")

## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## A - B 24.6 64.3 12 -147 196.2 0.382 0.9230   
## A - C -64.6 64.3 12 -236 107.0 -1.004 0.5882   
## B - C -89.2 64.3 12 -261 82.4 -1.387 0.3780   
##   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 3 estimates   
## P value adjustment: tukey method for comparing a family of 3 estimates

Based on the p-value, at 5% significant level, we don’t reject the null hypothesis.

#### Problem 5. Let and . Let for . All random variables are independent to one another.

#### a. Let , what is the distribution of F?

#### b. Let , what is the distribution of G?

#### c.Let , what is the distribution of R?

#### d. Let , what is the distribution of K?

#### e. Find and such that:

#### follows an distribution. What are the numerator degrees of freedom and denominator degrees of freedom for this distribution?

Therefore, we can get . The numerator degrees of freedom is 1 and denominator degrees of freedom is 20.