STAT 461 Midterm 1

Xiangyu Ren

10/12/2020

#### Question 3. Use R and R-markdown to answer the following questions and show all code used to answer these questions. Assume that a CRD was used for an experiment with three treatments (A, B, and C).The following table records the response variable for each experimental unit:

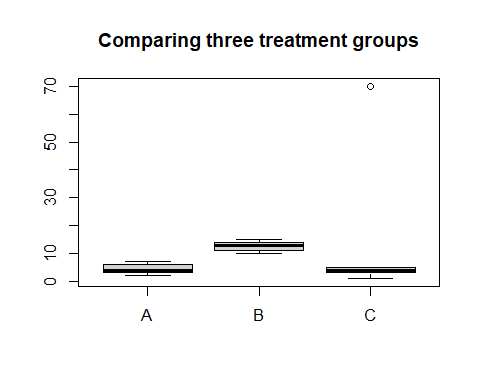


#### (a) Create a box-plot comparing the three treatment groups. Are there any peculiar points in the figure?

data = data.frame("A" = c(6,4,3,7,2), "B" = c(10,11,15,14,13), "C" = c(1,4,3,5,70))  
data

## A B C  
## 1 6 10 1  
## 2 4 11 4  
## 3 3 15 3  
## 4 7 14 5  
## 5 2 13 70

boxplot(data, main = "Comparing three treatment groups")



There is a peculiar point in group c with value 70.

#### (b) The scientist is interested in finding out if there isanyof the treatments differs from another. Write down the appropriate null hypothesis in statistical notation.

Null hypothesis:

#### (c) Run a One-Way ANOVA analysis in R to compute the p-value for the hypothesis test outlined in (b). Do you reject or accept the null hypothesis at an level?

data = data.frame("treatment" = c(rep("A", 5), rep("B", 5), rep("C", 5)), "value" = c(6,4,3,7,2,10,11,15,14,13,1,4,3,5,70))  
model = aov(value ~ treatment, data = data)  
anova(model)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)  
## treatment 2 386.8 193.40 0.6433 0.5428  
## Residuals 12 3607.6 300.63

According to the anova table, we can conclude that we will accept the null hypothesis since the p-value is large than the that we picked.

#### (d) The scientist now wants to test the hypothes is: . Compute the p-value for this hypothesis test. Do you accept the null hypothesis at the level? You must write out the test statistic and the distribution of the test statistic for full credit.

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'.

lsm = lsmeans(model, "treatment")  
contrast1 = c(1/2, 1/2, -1)  
contrastList = list("1/2(tau\_a+tau\_b)-tau\_c" = contrast1)   
contrast(lsm, contrastList)

## contrast estimate SE df t.ratio p.value  
## 1/2(tau\_a+tau\_b)-tau\_c -8.1 9.5 12 -0.853 0.4104

According to the contrast table, we can see that the estimate is -8.1 and the p-value is 0.4104. Therefore, we can conclude that we accept the null hypothesis since 0.4104 is larger than the that we picked. Thus .

#### (e) While working on their analysis, the scientist notes that they made a mistake in entering one of their data points. Instead, . Repeat (c) but with this change. Did the outcome of the hypothesis test change? Why? (Hint: Compute the , , and ).

data2 = data.frame("treatment" = c(rep("A", 5), rep("B", 5), rep("C", 5)), "value" = c(6,4,3,7,2,10,11,15,14,13,1,4,3,5,7))  
model2 = aov(value ~ treatment, data = data2)  
anova(model2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## treatment 2 235.6 117.800 25.985 4.357e-05 \*\*\*  
## Residuals 12 54.4 4.533   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

According to the anova table, the p-value that we get is extremely smaller than the , after we change to 7. Therefore, we can say that we reject the null hypothesis, thus at least one treatment has significant difference from others.

At this point, we say the outcome of the hypothesis test changes from part(c).

# sig\_A^2  
var(c(6,4,3,7,2))

## [1] 4.3

# sig\_B^2  
var(c(10,11,15,14,13))

## [1] 4.3

# old sig\_C^2  
var(c(1,4,3,5,70))

## [1] 893.3

# new sig\_C^2  
var(c(1,4,3,5,7))

## [1] 5

The outcome of the hypothesis test changes, because in part(c), with incorrect data, the variance of treatment group c is extremely larger than other groups, 70 is an outlier, thus we would loose precision on part(c)’s hypothesis test. Therefore after we correct the data, with smaller variance, we get more precise conclusion and p-value.