STAT461 HW5

Xiangyu Ren

10/2/2020

#### Problem 1. Show that by using an argument similar to that used in the lectures to find distributions for and .

#### Problem 2. A test statistic that could be used to test for a significant pairwise differences between thei-th andj-thtreatment is

##### The null hypothesis is . Find the distribution of under the null hypothesis.

First test for the difference. The null hypothesis is:

The least squares estimate of is: In the ANOVA model:

Under , so

therefore:

Thus, we get

#### Problem 3. Recall the soap experiment from Homework 1. Look back at Homework 1 for an explanation of the experiment. The data are the weight lost over 24 hours by different types of soap.

##### (a) Construct the ANOVA table for this experiment by hand. Show the calculations needed to construct the quantities in the ANOVA table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | df | Sum of Square | Mean Square | Ratio |
| Treatment | 3-1=2 | (SST)161.22 | 8.061 | 106.065 |
| Error | 12-3=9 | (SSE)0.687 | 0.076 |  |
| Total | 12-1=11 | (SSTot)16.8 |  |  |

##### (b) Test the null hypothesis that there is no difference in mean weight lost between different soap types. Report the test statistic, the p-value of the statistic under the null hypothesis, and interpret the result of the test. You may use R or an online p-value calculator (like <http://graphpad.com/quickcalcs/PValue1.cfm>) to compute the p-value. Provide the R code used, with output, or clearly describe the online p-value calculator you used.

type=c(rep("Regular",4), rep("Deodorant",4), rep("Moisturizing",4))  
lost=c(-0.3,-0.1,-0.14,0.4,2.63,2.61,2.41,3.15,1.86,2.03,2.26,1.82)  
soap=data.frame(type=as.factor(type),lost=lost)  
library(knitr)  
model=aov(lost~type,data=soap)  
anova(model)

## Analysis of Variance Table  
##   
## Response: lost  
## Df Sum Sq Mean Sq F value Pr(>F)   
## type 2 16.1220 8.0610 104.45 5.914e-07 \*\*\*  
## Residuals 9 0.6946 0.0772   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

test statistic here is 104.45 and the p-value for the two sided test is 5.914e-7, which is much smaller than .05, also means that we will observe test statistics greater than 104.45 only about 5.914e-5% of the time. Thus we can reject the null hypothesis and conclude that the shape of soaps do have significant impact on the dissolution of soaps.

#### Problem 4. Pedestrian light experiment (Larry Lesher, 1985) Recall the pedestrian light experiment from Homework 4. This experiment questions whether pushing a certain pedestrian light button had an effect on the wait time before the pedestrian light showed walk. The treatment factor of interest was the number of pushes of the button, and 32 observations were taken with a mix of 0, 1, 2, and 3 pushes of the button. The waiting times for the walk sign are shown in the following table, with (where the levels of the treatment factor are coded as 0, 1, 2, 3 for simplicity).

##### Answer the question: “Does pushing the button make the light change sooner?”. Clearly state the null and alternative hypotheses, the model used, and all assumptions in the model. Obtain a test statistic (show all code and R output), and interpret the results of the test.

p0 <- c(38.14,38.2,38.31,38.14,38.29,38.17,38.2)  
p1 <- c(38.28,38.17,38.08,38.25,38.18,38.03,37.95,38.26,38.3,38.21)  
p2 <- c(38.17,38.13,38.16,38.3,38.34,38.34,38.17,38.18,38.09,38.06)  
p3 <- c(38.14,38.3,38.21,38.04,38.37)  
push=c(rep(0,7), rep(1,10), rep(2,10), rep(3,5))  
time=c(p0,p1,p2,p3)  
data=data.frame(push=as.factor(push),time=time)  
head(data)

## push time  
## 1 0 38.14  
## 2 0 38.20  
## 3 0 38.31  
## 4 0 38.14  
## 5 0 38.29  
## 6 0 38.17

model\_4=aov(time~push,data=data)  
anova(model\_4)

## Analysis of Variance Table  
##   
## Response: time  
## Df Sum Sq Mean Sq F value Pr(>F)  
## push 3 0.008047 0.0026824 0.2455 0.8638  
## Residuals 28 0.305953 0.0109269

kable(anova(model\_4), format = "markdown")

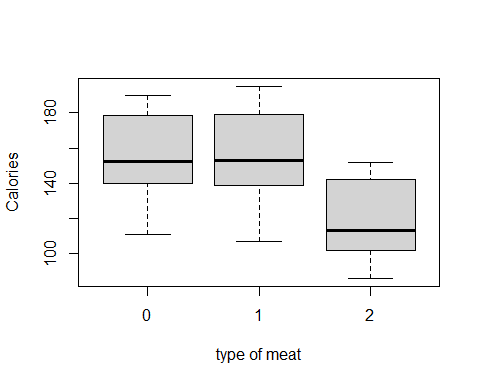
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| push | 3 | 0.0080471 | 0.0026824 | 0.2454844 | 0.8638292 |
| Residuals | 28 | 0.3059529 | 0.0109269 | NA | NA |

test statistic here is 0.2455, and the p-value is 0.8638, means that we will observe test statistics greater than 0.2455 about 86.38% of the time, which is greater than .05. Thus we accept the null hypothesis and conclude that pushing the button does not have significant impact on making the light change sooner.

#### Problem 5. Hot Dogs A study was conducted to compare the calories and sodium in hot dogs made with different types of meat.

##### (a) Read the data into R and plot calories as a response variable with the type of meat on the x-axis. Your plot could be either a boxplot or a plot with one dot for each hot dog.

Cal=c(186,181,176,149,184,190,158,139,175,148,152,111,141,153,190,157,131,149,135,132,  
173,191,182,190,172,147,146,139,175,136,179,153,107,195,135,140,138,  
129,132,102,106,94,102,87,99,107,113,135,142,86,143,152,146,144)  
Sod=c(495,477,425,322,482,587,370,322,479,375,330,300,386,401,645,440,317,319,298,253,  
458,506,473,545,496,360,387,386,507,393,405,372,144,511,405,428,339,  
430,375,396,383,387,542,359,357,528,513,426,513,358,581,588,522,545)  
type=c(rep(0,20),rep(1,17),rep(2,17))  
t=as.factor(type)  
boxplot(Cal~t,xlab="type of meat",ylab="Calories")



##### (b) Answer the following question: “Are there differences in the average calories of hot dogs made with different kinds of meat?”. To answer this question, write down a statistical model (clearly state the response variable, treatment levels, number of replicates, …), express the above question as a testable null hypothesis, and report the p-value of the test statistic under the null hypothesis. Your answer should include all R code used, and the important R output.

data=data.frame(type=as.factor(type),Cal=Cal)  
head(data)

## type Cal  
## 1 0 186  
## 2 0 181  
## 3 0 176  
## 4 0 149  
## 5 0 184  
## 6 0 190

model\_5=aov(Cal~type,data=data)  
anova(model\_5)

## Analysis of Variance Table  
##   
## Response: Cal  
## Df Sum Sq Mean Sq F value Pr(>F)   
## type 2 17692 8846.1 16.074 3.862e-06 \*\*\*  
## Residuals 51 28067 550.3   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

kable(anova(model\_5), format = "markdown")

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| type | 2 | 17692.20 | 8846.098 | 16.07399 | 3.9e-06 |
| Residuals | 51 | 28067.14 | 550.336 | NA | NA |

test statistic here is 16.074 and the p-value is 3.862e-6, which is smaller than .05, means that we will observe test statistics greater than 16.074 about 3.862e-4% of the time. Thus we can say we reject the null hypothesis, and conclude that hot dogs made with different kinds of meat does have significant impact on average calories of hot dogs.