Beckham Carver

3/11/22

STAT 4025

Dr. Robinson

Design and Analysis of Experiments: Midterm

Question 1:

a. True. In a one-way analysis of variance of this data, the degrees of freedom residual is 25.

b. False. If you wanted to determine if there is a difference in mean change in %cover for Herbicides C and E your null hypothesis would be: H0: μC – μE = 0.

c. The ɛij term in the model represents the fact that each observation has an error calculation respective to it, specifically the random error term respective to the jth wetlands of the ith diet.

d. False. Randomly selecting the ten wetlands from a population of wetlands AND randomly assignment to groups allows us to generalize our data analysis conclusions to the population of wetlands.

e. False. If we were to use all possible pairwise t-tests to determine if there are differences among the herbicides, our experiment wise Type I error rate would be 40.13%

f. False. After conducting a parametric test, it is important to check whether the model residuals follow a normal distribution.

g. True. Antipsychotic medications are commonly prescribed to older adults. These drugs are classified into two major groups: “typical” agents and newer “atypical” agents that are heavily marketed. One study evaluated the association between the type of antipsychotic medication and all-cause mortality. The researchers studied a group of adults aged 65 years and older who were participating in a statewide pharmaceutical assistance program. Researchers identified the type of antipsychotic medication at the start of the study using electronic pharmacy records. This is an example of a cross-sectional study.

h. True. One of the advantages of bootstrapping is that one can easily obtain confidence intervals for a wide array of population parameters (ex. the median, the first quartile, 90th percentile, etc.).

Question 2:

1. H0: μ1 = μ2 = μ3 = μ4

Ha: at least one of the means is significantly different

Where μn represent the true means of the respective temperatures, denoted as groups 1,2,3,4 as seen below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Temp | N | Mean | Variance | SD |
| 1. 100 | 7 | 21.714 | 0.018 | 0.1341641 |
| 1. 125 | 4 | 21.225 | 0.016 | 0.1264911 |
| 1. 150 | 5 | 21.720 | 0.027 | 0.1643168 |
| 1. 175 | 6 | 21.750 | 0.011 | 0.1048809 |

Using the information in the above table, our ANOVA table is:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Sum of Squares | d.F. | MS | F | p |
| Between | 0.8315 | 3 | 0.2772 | 15.6400 | 2.96e-05 |
| Within | 0.3190 | 18 | 0.0177 | -- | -- |
| Total | 1.1505 | 21 | -- | -- | -- |

With a p-value < 0.05 we can indeed reject the null hypothesis and conclude that there is a difference in means between the groups.

1. We are comparing between group variance vs. within group variance, to determine whether there is a true mean difference between the groups. The F-value (MSB/MSW) is the summary of this comparison, by using the f-distribution we can either reject or accept the null hypothesis. The specific models that are competing are MSB and MSW.
2. When utilizing the Bonferroni method for making all pairwise comparisons, our experiment wise error rate would be 0.0493. Our confidence interval regarding firing temperatures 150 and 175 will be [-0.1978, 0.2578] which gives us a resultant p-value of 0.98185. With a CI containing 0 and a p-value > 0.05, we cannot reject the null hypothesis that there is no true difference between 150 and 175.

Question 3:

1. Model, and hypothesis:

clot = β0 + β1x + e

H0: β0 = β1 = 0

Ha: β1 ≠ 0 or β0 = 0

Where β0 represents the intercept, β1 the gradient of the line, and x is the drug administered.

Our regression model produced the following output:

Residuals:

Min 1Q Median 3Q Max

-1.25714 -0.20714 0.03333 0.39286 1.08333

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.6667 0.2688 32.236 3.05e-12 \*\*\*

DrugB 1.1905 0.3664 3.249 0.00774 \*\*

Residual standard error: 0.6585 on 11 degrees of freedom

Multiple R-squared: 0.4897, Adjusted R-squared: 0.4434

F-statistic: 10.56 on 1 and 11 DF, p-value: 0.007745

Checking our assumption of normality, our qq-plot shows two points below the line. However, our Shapiro-wilks test is accepts the null hypothesis that the data is normal, with a p-value > 0.05. The homogeneity of variance assumption is also met, with a largest to smallest sample variance ratio of 1.17816 being close to 1.

Our regression model supports the conclusion that there is a true difference in mean change between the drugs with a p-value < 0.05. Our assumptions are also met, meaning we can reject the null hypothesis that the mean changes are the same.

Chart, scatter chart

Description automatically generated

Shapiro-Wilk normality test

data: avg\_elk$resids

W = 0.94909,

p-value = 0.5846

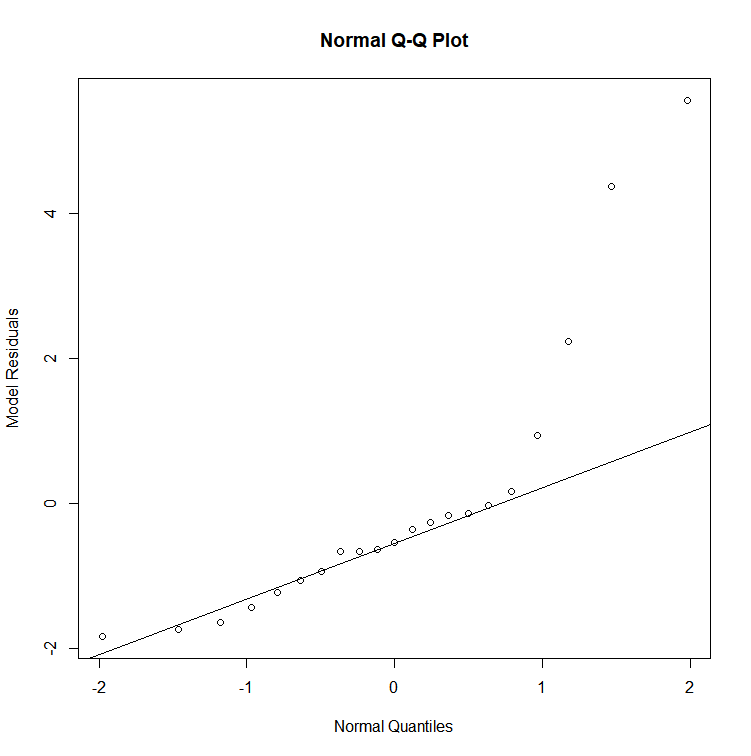
1. Our residual standard error is 0.658, which means our estimated error variance is 0.65852 = 0.4336486. This is the estimated variance around the true regression line, in this case it is very low which coincides to our low p-value. The variance is low in part because each drug has consistent and small residuals.

Question 4:

1. Sampling from readers of a Wyoming paper and testing a Wyoming whiskey implies the intended scope (or at least the effective scope) of the experiment is for Wyoming locals. It would be inaccurate to extrapolate conclusions beyond residents Wyoming, as Wyomingite’s opinions towards Wyoming made whiskey will be skewed. Participants were also allowed to select their own group, which nullifies RATG and in turn nullifies establishing causation. Because participants were allowed to select their own environment, the study informs more about the correlation between **tastes in** atmosphere/whiskey, rather than the **effect of** atmosphere on whiskey.
2. With the goal of determining the impact of atmosphere on Wyoming whiskey drinkers perceived taste, I would implement RATG, and ensure that participants are only informed that the study’s purpose is to rate the taste of the whiskey and not the effect of the atmosphere.

(optionally) After the rating were given, I would then poll the participants on what environment they *would* have preferred, and their age as these are cheap data points to collect that can help inform the study.

1. For assessing normality, we will use a qq-plot and a Shapiro wilks test:

Our qq-plot shows heavy outliers, as well as many residuals directly on the true regression line. This hints towards non-normal data and unequal variance of residuals.

Our Shapiro-Wilks test confirms this, with a p-value << 0.05 we reject the null hypothesis that the data is normal.

Shapiro-Wilk normality test

W = 0.76245,

p-value = 0.0001846

Calculating the largest to smallest sample variance ratio, we get 6.789551 which greatly exceeds the cutoff of 2.5, this means we also do not have equal variance of residuals. We conclude that the data is not normal and does not have equal variance of residuals. This nullifies any results given through ANOVA or regression.

1. To determine if there is a mean difference in taste ratings, a permutation test was performed over 10,000 permutations, where:

Ho: μa = μb = μc

Ha: at least one mean is different than the others

With the permutation test a p-value of 0.0018 was attained. This is less than 0.05 and does not depend on the assumptions rejected previously. Without RATG we cannot conclude causation, but **we can conclude** that there is a true difference in mean taste scores across the groups.

Chart, histogram

Description automatically generated

The following CI were also obtained using bootstrapping:

|  |  |  |
| --- | --- | --- |
| A vs. B | A vs. C | B vs. C |
| 2.5% 97.5%  2.043 5.150 | 2.5% 97.5%  2.242 6.496 | 2.5% 97.5%  -1.683 3.283 |

Code:

library(tidyverse)

# for question 2 I wanted to use rpsychi package as I had seen online however it

# appears to be unsupported now, I did find a website that performed the same

# anova from summary data tests, and double checked by hand

# https://acetabulum.dk/cgi-bin/anova

#### begin question 3

elk <- read.csv("elk.csv", header = T)

# cleaning data, experimental units are elk, not individual blood samples

avg\_elk <- elk %>%

group\_by(elk) %>%

mutate(clot= mean(clot)) %>%

distinct() %>%

group\_by(Drug) %>%

mutate(drugAvg = mean(clot))

# creating linear model

m1 <- lm(clot~Drug,data=avg\_elk)

summary(m1)

# adding residuals

avg\_elk$resids <- residuals(m1)

# constructing qq-plot, has outliers

{qqnorm(avg\_elk$resids,xlab="Normal Quantiles",

ylab="Model Residuals")

qqline(avg\_elk$resids)}

# conducting shapiro-test, is normal

shapiro.test(avg\_elk$resids)

boxplot(clot~Drug,data=avg\_elk)

# checking for equal variance

avg\_elk$logthresh <- log(avg\_elk$clot)

elk\_summary <- avg\_elk %>%

group\_by(Drug) %>%

summarise(vargroup = sd(clot)^2,

mngroup = mean(clot),

varlogs = var(logthresh))

varcheck <- max(elk\_summary$vargroup)/min(elk\_summary$vargroup)

varcheck

# var is close to 1, which supports homogeneity of variance.

###### begin question 4

# transferred data to a csv using Google sheets

whiskey <- read.csv("part4.csv")

# creating linear model

m2 <- lm(Rating~Atmosphere,data=whiskey)

summary(m2)

# adding residuals

whiskey$resids <- residuals(m2)

# creating anova

anovam1 <- summary(aov(m2))

# constructing qq-plot, whole lot of outliers

{qqnorm(whiskey$resids,xlab="Normal Quantiles",

ylab="Model Residuals")

qqline(whiskey$resids)}

# conducting shapiro-test, is not normal

shapiro.test(whiskey$resids)

boxplot(Rating~Atmosphere,data=whiskey)

# checking for equal variance

whiskey$logthresh <- log(whiskey$Rating)

whisk\_summary <- whiskey %>%

group\_by(Atmosphere) %>%

summarise(vargroup = sd(Rating)^2,

mngroup = mean(Rating),

varlogs = var(logthresh))

varcheck <- max(whisk\_summary$vargroup)/min(whisk\_summary$vargroup)

varcheck

# variance exceeds 2.5, fails

# bootstrapping time, data is not normal

Fobs <- anovam1[[1]]$"F value"[1]

Fobs

permutation.test <- function(obsdat,respvar,treatvar,numsims){

Fperm=c()

result=0

for(i in 1:numsims){

newtreat <- sample(treatvar,nrow(obsdat),replace=FALSE)

mod <- lm(respvar ~ newtreat,data=obsdat)

anovamod <- summary(aov(mod))

Fperm[i] <- anovamod[[1]]$"F value"[1]

}

result=sum(Fperm >= Fobs)

return(list(result, Fperm))

}

test1 <- permutation.test(obsdat=whiskey,

respvar=whiskey$Rating,

treatvar=whiskey$Atmosphere,

numsims=10000)

hist(test1[[2]], breaks=50, col='grey', main="Permutation Distribution of Variance Ratio",

xlab='variance ratio',xlim=c(1,30))

abline(v=Fobs, lwd=3, col="red")

perm\_pvalue <- test1[[1]]/10000

perm\_pvalue

bootf <- function(dat,sampsize,B){

meanvec <- 1:B

for (i in 1:B){

samp <- sample(dat,sampsize, replace=TRUE)

mnsamp <- mean(samp)

meanvec[i] <- mnsamp

} # end of i loop

return (list(meanvec))

} #end of function bootf

options(digits=4)

whiskey\_a <- whiskey[whiskey$Atmosphere=="Classical",]

whiskey\_a <- whiskey\_a$Rating

whiskey\_b <- whiskey[whiskey$Atmosphere=="Normal",]

whiskey\_b <- whiskey\_b$Rating

whiskey\_c <- whiskey[whiskey$Atmosphere=="Cowboy",]

whiskey\_c <- whiskey\_c$Rating

aboot <- bootf(whiskey\_a,8,10000)

bboot <- bootf(whiskey\_b,7,10000)

cboot <- bootf(whiskey\_c,6,10000)

fullboot <- data.frame(Atmosphere = rep(c("Classical"),each = 10000),

ratings = c(aboot))

# I cant get all of the memebrs into the dataframe :((

bootdiff\_ab <- aboot[[1]]-bboot[[1]]

conf\_int\_diff\_sw\_cy <- quantile(bootdiff\_ab,c(0.025,0.975))

conf\_int\_diff\_sw\_cy

bootdiff\_ac <- aboot[[1]]-cboot[[1]]

conf\_int\_diff\_sw\_cy <- quantile(bootdiff\_ac,c(0.025,0.975))

conf\_int\_diff\_sw\_cy

bootdiff\_bc <- bboot[[1]]-cboot[[1]]

conf\_int\_diff\_sw\_cy <- quantile(bootdiff\_bc,c(0.025,0.975))

conf\_int\_diff\_sw\_cy