Lab Assignment 6

STA 100 | A. Farris | Spring 2021

A pdf copy of your assignment is due at 5:00pm PT) Wednesday, June 2. Submission of the pdf will be through gradescope. Please put in the effort to make it look reasonably professional – you're encouraged to use R Markdown. You are free to work in groups, but you must submit your own writeup, and run your own code. Work for you is <mark>highlighted</mark>.

Part 1: Testing for differences between group means

As a case study, let's investigate differences in the luminosity between several different kinds of Egyptian cotton. We have at our disposal `egyptianCotton.csv`, containing 20 measurements of luminosity from samples of cotton. Represented among these are five different kinds of cotton identified as `Giza67`, `Giza68`, `Giza69`, `Giza70`, and `Menoufi`.

Let's read this into memory:

```
ec <- read.csv("egyptianCotton.csv")
```

We can inspect the first six rows of the data to assess it's format:

```
head(ec)
```

```
  Luminance Variety
1     89.94  Giza69
2     89.59  Giza69
3     89.22  Giza69
4     88.98  Giza69
5     89.00  Giza67
6     88.52  Giza67
```

Preliminary to testing for equality between the means of these groups, we can fit an ANOVA model:

```
fittedModel <- aov(Luminance ~ Variety,
                   data = ec
                   )
```

We can inspect boxplots and a histogram of residuals in order to check whether there are substantial violations of the equal variances and normality assumptions of the F-test that we will apply:

```
par(mfrow=c(1,2)) ## two side-by-side plots
boxplot(Luminance ~ Variety,
        data = ec,
        xlab="",
        las = 2 ## vertical labels
        )
hist(fittedModel$residuals,
     ylab="Frequency",
     xlab="Residual",
     main=""
     )
```
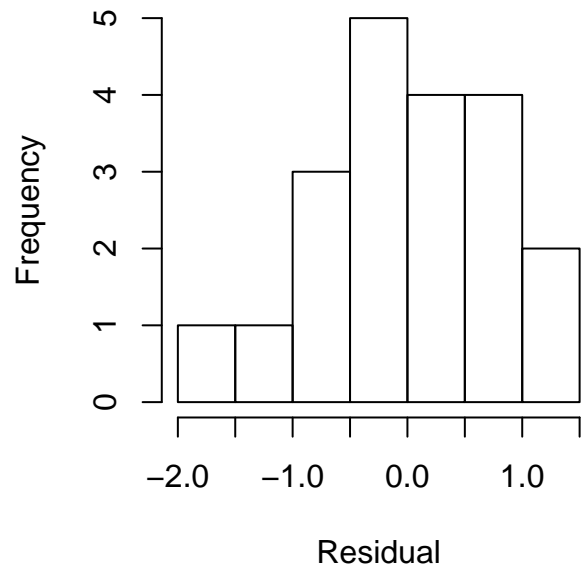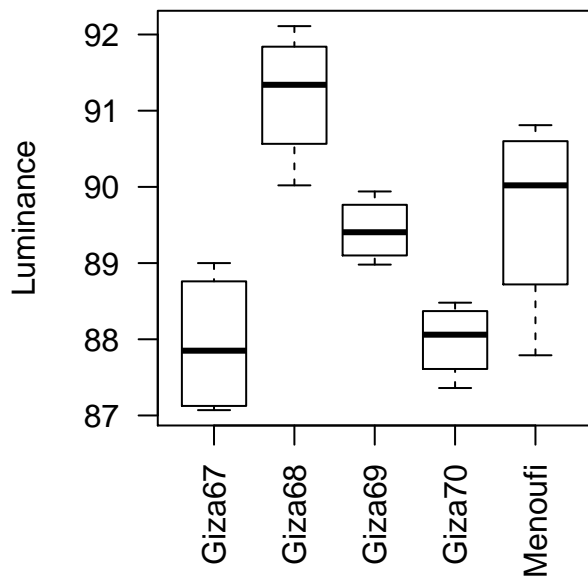
These depart slightly from equal variances and from normality, but by no more than we would expect them to by chance.

If you're ever unsure of, for example, how close the histogram should look to bell-shaped, run something like

```
normalData <- rnorm(dim(ec)[1])
resids <- aov(normalData~Variety, data=ec)$residuals
hist(resids)
```

a few times. This will generate some random normally distributed data with the same dimension as your data, so the resulting histograms show you what to expect under normality. You'll see by doing this that, with small sample sizes, even seemingly clear violations of normality might not be decisive.

The F-test itself is easy enough to perform:

```
anova(fittedModel)
```

```
Analysis of Variance Table

Response: Luminance
          Df Sum Sq Mean Sq F value    Pr(>F)
Variety    4 29.243  7.3107  9.3158 0.0005455 ***
Residuals 15 11.771  0.7848
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we use, say, significance level 0.01, we do in fact reject the null hypothesis here with p-value 0.00055. Because we have found evidence that the groups are not all the same, let's perform a post-hoc analysis. To do so, we can use Tukey's Honestly Significant Difference procedure, with familywise error rate 0.01 (i.e. confidence 0.99):

```
tuky <- TukeyHSD(fittedModel,
                 conf.level = 0.99
                 )
par(mar=c(5,6,4,1)+1.2) ## so that labels don't get cut off
plot(tuky,
     las=1 ## horizontal labels
     )
```

2

99% family–wise confidence level
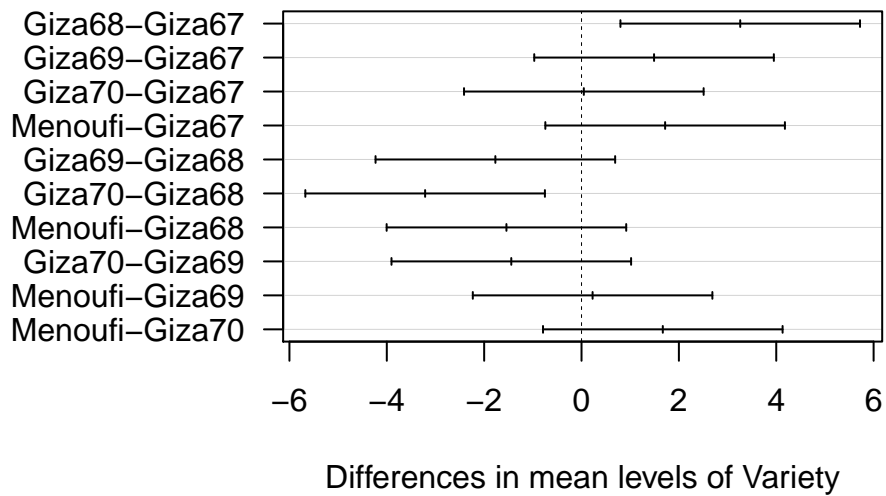
From the confidence intervals above, we can be 99% confident that Giza67 and Giza70 have lower average luminosity than Giza68, but we don't have sufficient evidence to make any other conclusions.

Part 2: Assignment

Perform the analysis requested in the memo on the next page.

Memo

FROM: Maria McYourNewBoss

TO: New hire

SUBJECT: Ventilation analysis

DATE: July 27, 2020

Congratulations on your new position in the McYourNewBoss Lab! I hope you've had a chance to get settled and to meet your new labmates. Before you get started with some other projects, I'm going to need you to wrap one up that had already been started. The data for this one has already been collected, so this should be easy.

The goal of this project is to investigate how three methods of ventilation during anesthesia affect post-op blood folate levels. The three methods are

- 50% nitrous oxide and 50% oxygen, continuously for 24 hours;
- 50% nitrous oxide and 50% oxygen, only during the operation; and
- no nitrous oxide but 35%–50% oxygen continuously for 24 hours.

You can find data on blood folate levels (in μg) for patients treated with these methods here.

I'm concerned about limiting the risk of anemias that are associated with low blood folate levels in patients, so you need to identify for me whether there is evidence for any of these methods being better or worse than others in this sense.

In our field, it is typical to use significance level 0.05.

Remember that, like many similarly distinguished people in my field, I rely on students or employees to do my data analysis for me. So, it is helpful to me for you to explain what you are doing that I may follow along. However, also like many distinguished employers in this field, I have a strong and somewhat mercurial sense of how valuable my time is; so you need to do so in as brief and succinct a way as you can.