HW 19 Solutions

For problems 1-3 consider the following table which gives the measurements on chlorpheniramine maleate tablets from a particular manufacturer (as in Figure 12.1 of your textbook). This data represents the reports from seven different labs that analyzed 10 different portions of a composite (each) whose weight was equivalent to that of one chlorpheniramine maleate tablet. The composites being measured were prepared beforehand by grinding and mixing together tablets that had nominal dosage levels of 4mg chlorpheniramine maleate.

The purpose of the study is to evaluate the use of a semiautomated method for measuring the amount of chlorpheniramine maleate in tablets. Your boss wants you (the data scientist) to analyze the data to see if there could be some systematic differences in the measurements among the labs. Your task is to determine if the data indicates that any differences in the means of the measurements from the various labs are signals of a systematic difference or if they might as well be due to chance.

You can import this data set into (base) R using the following code:

```
mydata <- read.csv("https://raw.githubusercontent.com/ProfSuzy/Stat61/main/tablets2.csv",
                   header=T)
mydata
```

```
##      Lab1 Lab2 Lab3 Lab4 Lab5 Lab6 Lab7
## 1   4.15 3.93 4.10 4.16 4.14 4.09 4.10
## 2   4.08 3.92 4.10 4.14 4.14 4.08 4.05
## 3   4.09 4.08 4.05 4.16 4.14 4.08 4.05
## 4   4.08 4.09 4.07 4.02 4.17 4.12 4.06
## 5   4.01 4.06 4.06 4.15 4.13 4.17 4.06
## 6   4.01 4.06 4.03 4.15 4.24 4.15 4.12
## 7   4.00 4.02 4.04 4.18 4.18 4.12 4.07
## 8   4.09 4.00 4.03 4.12 4.14 4.10 4.18
## 9   4.08 4.01 4.03 4.09 4.25 4.12 4.15
## 10  4.00 4.01 4.06 4.04 4.17 4.12 4.18
```

Problem 1 (12.20)

(a) List one possible pro and one possible con about the statistical decision to use an exploratory, graphical technique to answer this research question.

(b) List one possible pro and one possible con about the statistical decision to use a formal F-test to answer this research question.

Problem 2 (12.20) List one possible pro and one possible con about the statistical decision to use both exploratory, graphical techniques and a formal F-test to answer this research question.

Problem 3 Make a decision to use only one of the three different analysis approaches listed above to address this research question and report the results of your analysis. (Make sure you include any relevant graphs or R output.)

```
##Put data table into proper form (each unit on one row)
measurements = as.vector(unlist(mydata))
portion=rep(1:10, 7)
lab = c(rep("Lab1", 10), rep("Lab2", 10), rep("Lab3", 10), rep("Lab4", 10),
        rep("Lab5", 10), rep("Lab6", 10), rep("Lab7", 10))
## Double check that the lengths of these vectors match
length(measurements); length(portion); length(lab)
```
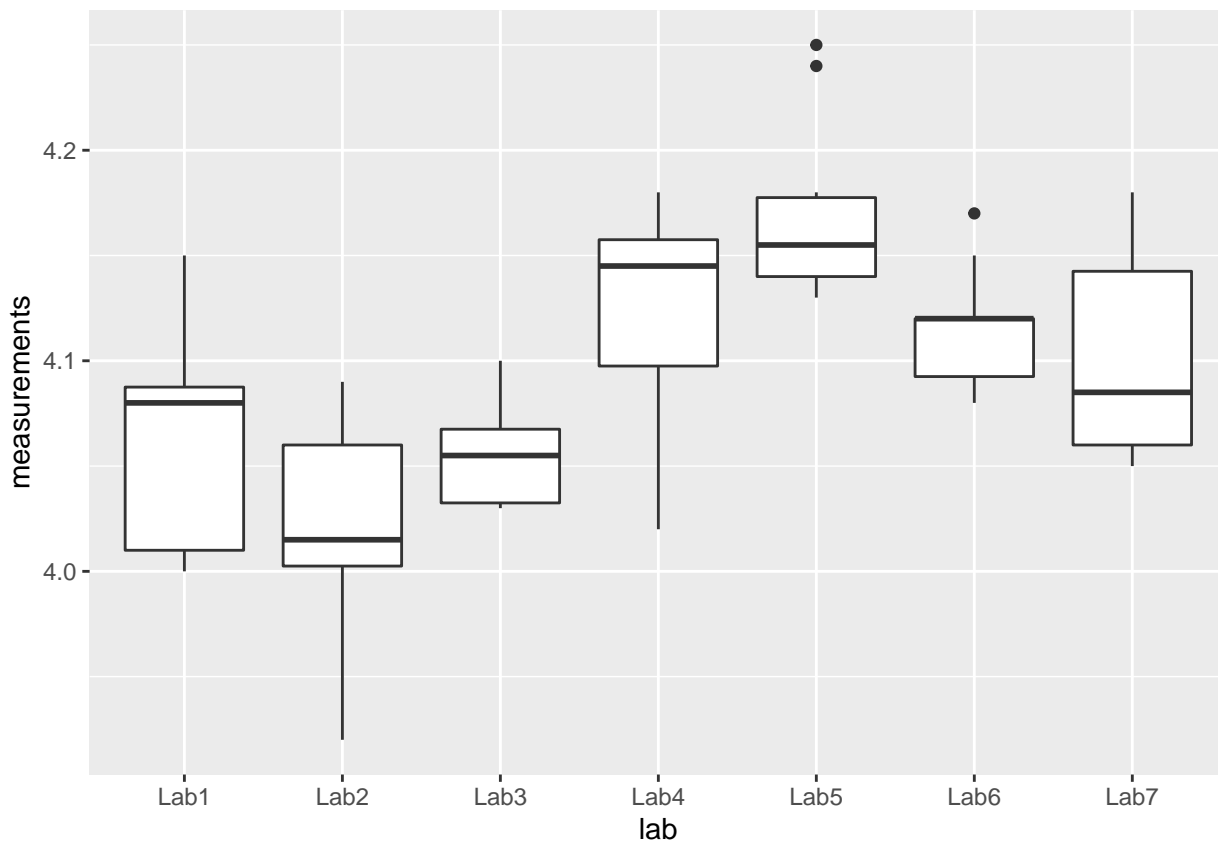
```
## [1] 70
```

```
## [1] 70
```

```
## [1] 70
```

```
## Create the data frame to analyze
unit_data <- data.frame(measurements, portion, lab)
## Create boxplots of the 10 observations for each of the 7 labs
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
ggplot(unit_data)+
  geom_boxplot(aes(y=measurements, x=lab))
```



```
mod <- lm(measurements ~ factor(lab), data = unit_data)
anova(mod)
```

3

```
## Analysis of Variance Table
##
## Response: measurements
##             Df  Sum Sq   Mean Sq F value    Pr(>F)
## factor(lab)  6 0.15343 0.0255724  11.902 6.975e-09 ***
## Residuals   63 0.13536 0.0021486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```