Matthew L. Pergolski
IST 772
Dr. Block
12/14/2021

Homework 10

Beginning Statement

"I produced the material below with no assistance [direct quote from IST 772 class syllabus]." Note: Homework questions from the book may have been copied/pasted into the document for both the student and viewer's convenience.

As per instructions, a website was referenced to find historical quotes for one of the below problems.

The homework for week 10 is exercises 2, 5, 6, 7, and 8 on pages 272 and 273.

Homework Question 2

Question:

2. Download and library the nlme package and use data ("Blackmore") to activate the Blackmore data set. Inspect the data and create a box plot showing the exercise level at different ages. Run a repeated measures ANOVA to compare exercise levels at ages 8, 10, and 12 using aov(). You can use a command like, myData <-Blackmore[Blackmore$age <=12,], to subset the data. Keeping in mind that the data will need to be balanced before you can conduct this analysis, try running a command like this, table(myData$subject,myData$age)), as the starting point for cleaning up the data set.

Answer/Student Response:

The following was generated and observed:

```
library(car)
library(nlme)

myData <- Blackmore[Blackmore$age <= 12,]
table(myData$subject,myData$age)
myData

boxplot(exercise~age, data = myData)

data <- (myData[myData$age <= 12,])
str(data)
data$ageFact <- as.factor(data$age)
list <- rowSums(table(data$subject,data$ageFact))==3
list <- list[list == TRUE]
list <- as.numeric(names(list))

summary(data[data$ageFact == 8,])
summary(data[data$ageFact == 10,])
summary(data[data$ageFact == 12,])

data <- data[data$subject %in% list,]
summary(aov(exercise~ageFact+ Error(subject), data = data))
```

```
>     summary(aov(exercise~ageFact+ Error(subject), data = data))

Error: subject
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 165   1892   11.47

Error: Within
           Df Sum Sq Mean Sq F value   Pr(>F)
ageFact     2  102.7   51.35   27.04 1.33e-11 ***
Residuals 330  626.6    1.90
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The p-value is statistically significant showing a value under the threshold of 0.05 (i.e., 1.33e-11). This signifies that there is a significant difference in terms of the age attributes and exercise performed. Based on this, we can reject the null hypothesis that the difference between the attributes is zero (or essentially zero).

Homework Question 5

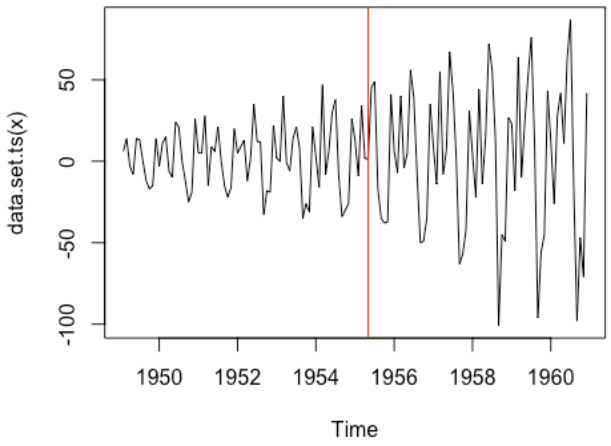
Question:
5. Given that the AirPassengers data set has a substantial growth trend, use diff() to create a differenced data set. Use plot() to examine and interpret the results of differencing. Use cpt.var() to find the change point in the variability of the differenced time series. Plot the result and describe in your own words what the change point signifies.

Answer/Student Response:
The following was generated and observed:

```
install.packages('changepoint')
library(changepoint)
data('AirPassengers')
air.pass <- diff(AirPassengers)
plot(air.pass)

cpt.var(air.pass)

plot(cpt.var(air.pass))
```




```
Created on  : Tue May 25 17:44:21 2021

summary(.)  :
----------
Created Using changepoint version 2.2.2
Changepoint type      : Change in variance
Method of analysis    : AMOC
Test Statistic  : Normal
Type of penalty       : MBIC with value, 14.88853
Minimum Segment Length : 2
Maximum no. of cpts   : 1
Changepoint Locations : 76
```

The graph indicates a timeline in which data has been recorded in terms of passengers over several months/years the red-line indicates an 'inflection point' in which the number of passengers began to substantially increase versus before. In looking at the graph from a visual perspective, we can see that the change happened at some point between 1954 and 1956, presumably in 1955. When looking at the function, we can see the 'Changepoint Locations' value equals 76.

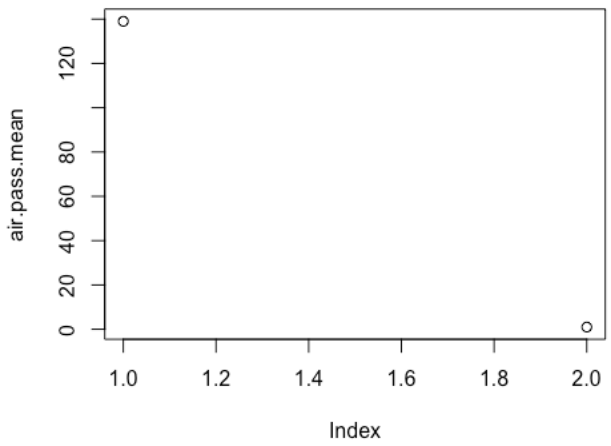
Homework Question 6

Question:
6. Use cpt.mean() on the AirPassengers time series. Plot and interpret the results. Compare the change point of the mean that you uncovered in this case to the change point in the variance that you uncovered in Exercise 5. What do these change points suggest about the history of air travel?

Answer/Student Response:
The following was developed and observed:

```
air.pass.mean <- cpt.mean(air.pass, class = FALSE)
plot(air.pass.mean)

air.pass.mean['conf.value']
```



From the graph generated through plot(air.pass.mean), we see that there's a change from the first data point when compared to the data point that follows, we detect a clear difference in value, which suggests a change has indeed been witnessed by those paying attention as time moved on.

Homework Question 7

Question:
7. Find historical information about air travel on the Internet and/or in reference materials that sheds light on the results from Exercises 5 and 6. Write a mini-article (less than 250 words) that interprets your statistical findings from Exercises 5 and 6 in the context of the historical information you found.

Answer/Student Response:
According to an article found on https://airandspace.si.edu/exhibitions/america-by-air/online/heyday/heyday11.cfm, it indicates, "By the end of the 1950s...[flying] was becoming a necessity." The article mentions that air traffic began to increase in the 1950s, which seems to match our historical data found in R with the Airpassengers dataset. When looking at the graph generated through R for the plot(air.pass.mean), we can confirm that the article's information suggests that our initial takeaways from our inflection/change point are fairly accurate -- if not more tailored to the exact year when air travel become more 'popular.'

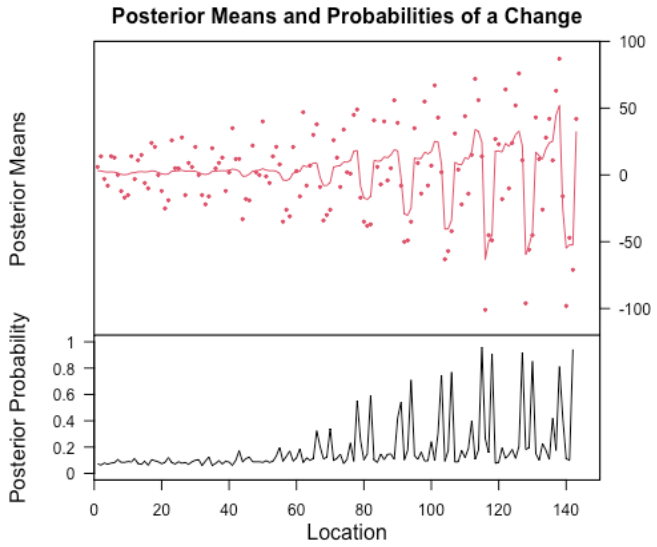
Homework Question 8

Question:
8. Use bcp() on the AirPassengers time series. Plot and interpret the results. Make sure to contrast these results with those from Exercise 6.

Answer/Student Response:
The following was generated and observed:

```
install.packages('bcp')
library(bcp)

b.c.p <- bcp(as.vector(air.pass))
b.c.p
plot(b.c.p)
```



The findings from the Bayesian approach seem to draw similar conclusions. We can see, from the posterior probability distribution, we see that, from visually looking at the graph, that our change point seems to correspond to the 60 through 80 interval period.