Homework 10

Kyle Walter

6/19/2021

——– 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(myDatasubject,myDataage)), as the starting point for cleaning up the data set.

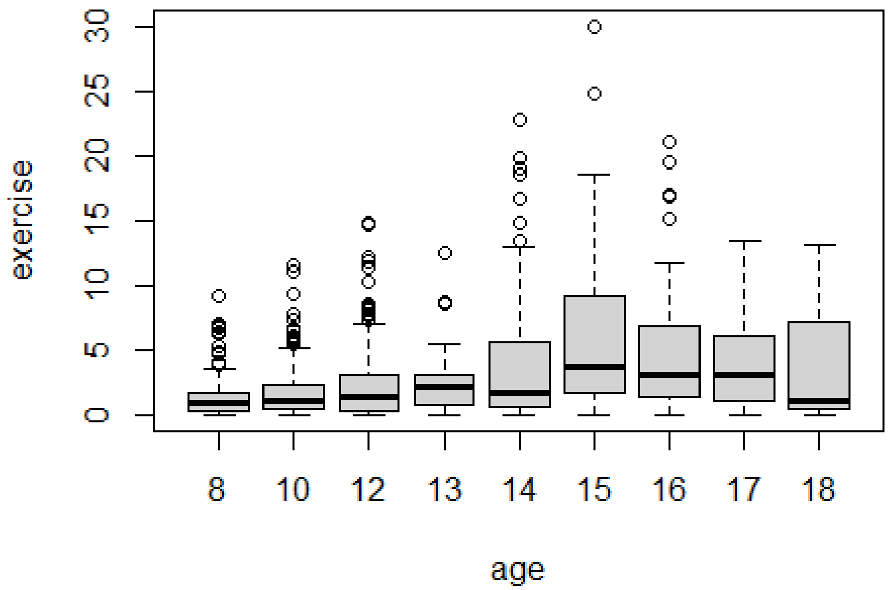
```
require(nlme)

## Loading required package: nlme

require(carData)

## Loading required package: carData

data(Blackmore)
summary(Blackmore)

##     subject          age           exercise          group
##  100    :  5   Min.   : 8.00   Min.   : 0.000   control:359
##  101    :  5   1st Qu.:10.00   1st Qu.: 0.400   patient:586
##  105    :  5   Median :12.00   Median : 1.330
##  106    :  5   Mean   :11.44   Mean   : 2.531
##  107    :  5   3rd Qu.:14.00   3rd Qu.: 3.040
##  108    :  5   Max.   :17.92   Max.   :29.960
##  (Other):915
```

Started by calling the data and looking at the summary. We see th ages rang from 8 years old to max of 17.9 years. In order to answer our question lets round the dates to the closest year.

```
Blackmore$age<- round(Blackmore$age)
```

Create a boxplot of the data

```
boxplot(exercise ~ age, data = Blackmore)
```



Now that we see all the ages rang from 8 through 18 over a 10 year period. We want to grab just those years that we need for the study group

```
FilteredBlackMore <- Blackmore[Blackmore$age<=12,]
```

Next we want to grab where the measurments are take for each group in all three years. This helps balace the data set

```
List <- rowSums(table(FilteredBlackMore$subject, FilteredBlackMore$age))==3
List <- List[List==T]
List <- factor(names(List))
FilteredBlackMore <- FilteredBlackMore[FilteredBlackMore$subject %in% List,]
```

Check that the set has balanced

```
table(FilteredBlackMore$age)

##
##   8  10  12
## 187 187 187
```

Next we will run the frequentist expirement. Our null hypthoesis is that exercise does not vary over age, while

```
summary(aov(exercise ~ age + Error(subject), data = FilteredBlackMore))

##
## Error: subject
```

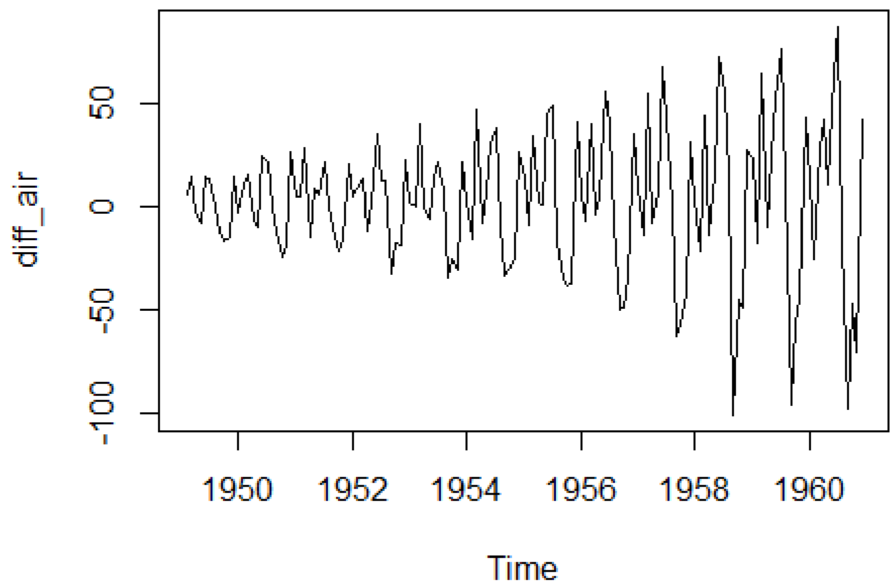
```
##            Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 186   1979   10.64
##
## Error: Within
##            Df Sum Sq Mean Sq F value   Pr(>F)
## age         1  101.9   101.9   56.51 4.19e-13 ***
## Residuals 373  672.8     1.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model shows the effect of age expresed in the F value. We have a quite high F value of 56.51 and p value of 4.19e-13. Since the p-value is virtually zero, we reject the null hypthosis that age has no effect on exercise and accept the alternative hypthosis that exercise varies across age. of children.

——– Exercise 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.

First we'lll bring in the dataset for air passengers and then lag the actuals to view variance over time.

```
require(changepoint)

## Loading required package: changepoint

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Successfully loaded changepoint package version 2.2.2
##  NOTE: Predefined penalty values changed in version 2.2.  Previous penalty
values with a postfix 1 i.e. SIC1 are now without i.e. SIC and previous
penalties without a postfix i.e. SIC are now with a postfix 0 i.e. SIC0. See
NEWS and help files for further details.

data("AirPassengers")
diff_air <- diff(AirPassengers)
plot(diff_air)
```

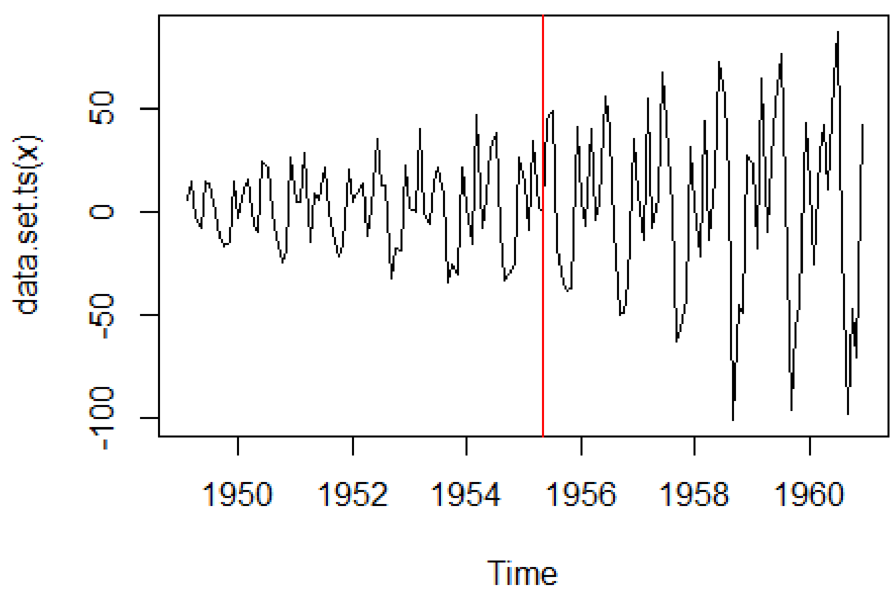


THe variances show a patterns of seasonality of how many passengers fly in each month of the year. There are periods of higher traffic such as the summer and winter holidays while other months.As the graph hits about halfway we see a larger variance in the time of the year when peope fly.

```
varAir <- cpt.var(diff_air)
varAir

## Class 'cpt' : Changepoint Object
##        ~~   : S4 class containing 12 slots with names
##               cpttype date version data.set method test.stat pen.type
pen.value minseglen cpts ncpts.max param.est
##
## Created on  : Sat Jun 12 12:48:47 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
```

I next ran the change point variance to determine where there is a significant change as the 76th value. Our significant change point is 14.88853. In order to better detect this we plot the line at point to be able to see where the shift happens in the data. As we can the jumps in size more than double to the right.

```
plot(varAir, cpt.col="red", cpt.width=.05)
```



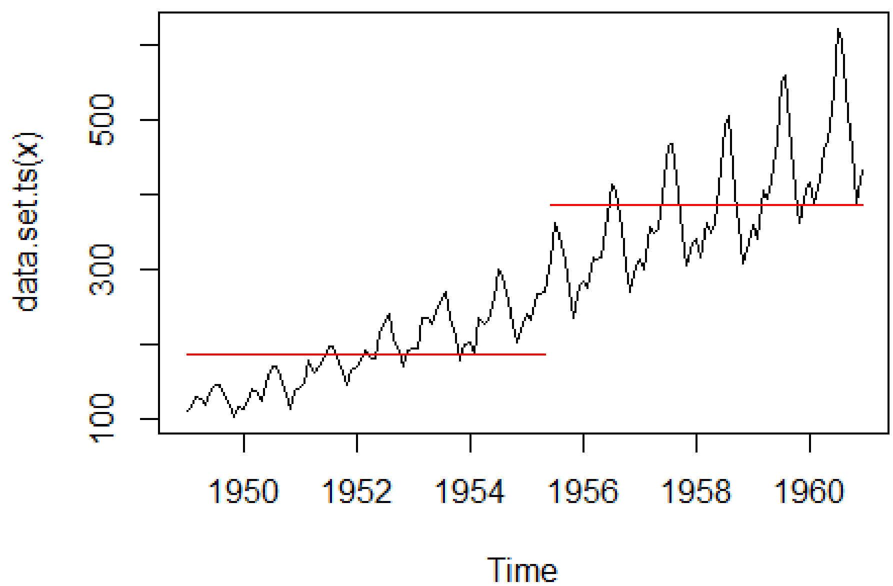
***

Exercise 6 - Changepoint mean

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? change point in mean

```
varMean <- cpt.mean(AirPassengers)
varMean

## Class 'cpt' : Changepoint Object
##        ~~   : S4 class containing 12 slots with names
##              cpttype date version data.set method test.stat pen.type
pen.value minseglen cpts ncpts.max param.est
##
## Created on  : Sat Jun 12 12:48:47 2021
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.2
```

```
## Changepoint type      : Change in mean
## Method of analysis    : AMOC
## Test Statistic  : Normal
## Type of penalty       : MBIC with value, 14.90944
## Minimum Segment Length : 1
## Maximum no. of cpts   : 1
## Changepoint Locations : 77

plot(varMean, cpt.col="red", cpt.width=.05)
```



The red line in the graph represents where the change in mean is significant. We see the line is showing us around 1955 at this point the number of passengers flying is not only different from based on season, but after 1955 the number of people flying increases significantly. It maybe changes in disposable income or redesign of seating by the airlines to squeeze more passengers in thus increasing supply for the number of people to take to the sky.

——– Exercise 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.

SOURCE: https://www.chicagotribune.com/coronavirus/ct-nw-coronavirus-airline-tsa-travelers-20200409-ylrq2ztctbe4fh35cfhbgrxczy-story.html

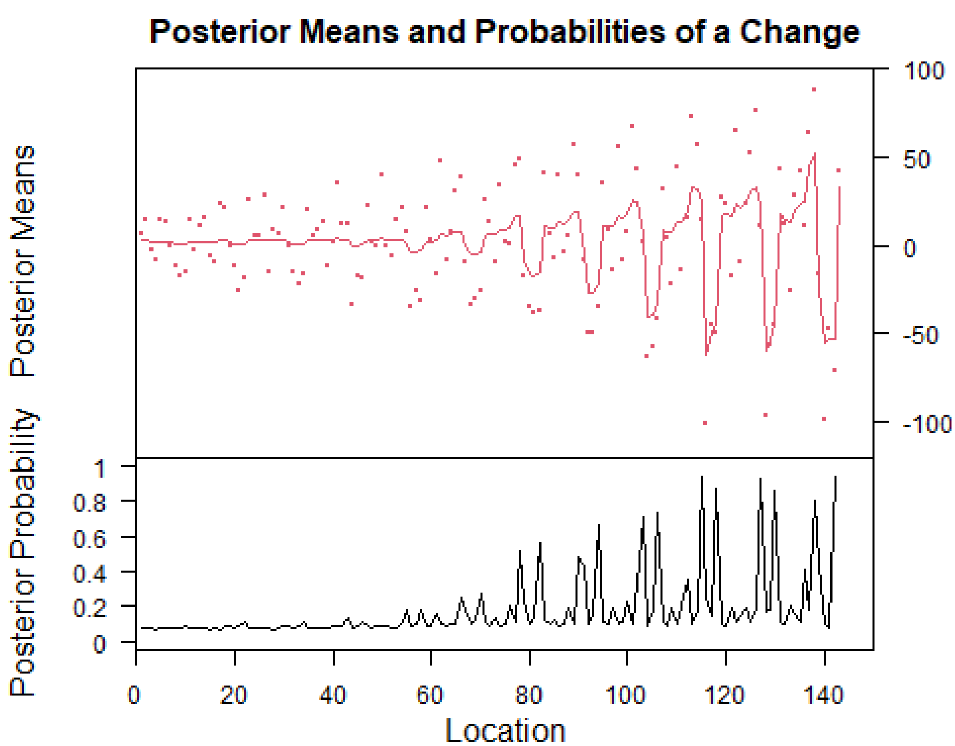
Accoring to the article, during the Coronavirus pandemic we the flight levels dropped to levels not seen since 1954, which supports what the data was showing us. The artilce also mentions that advancements in safety around that time and lower costs lead more people to take to the air for their travel plans. All

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

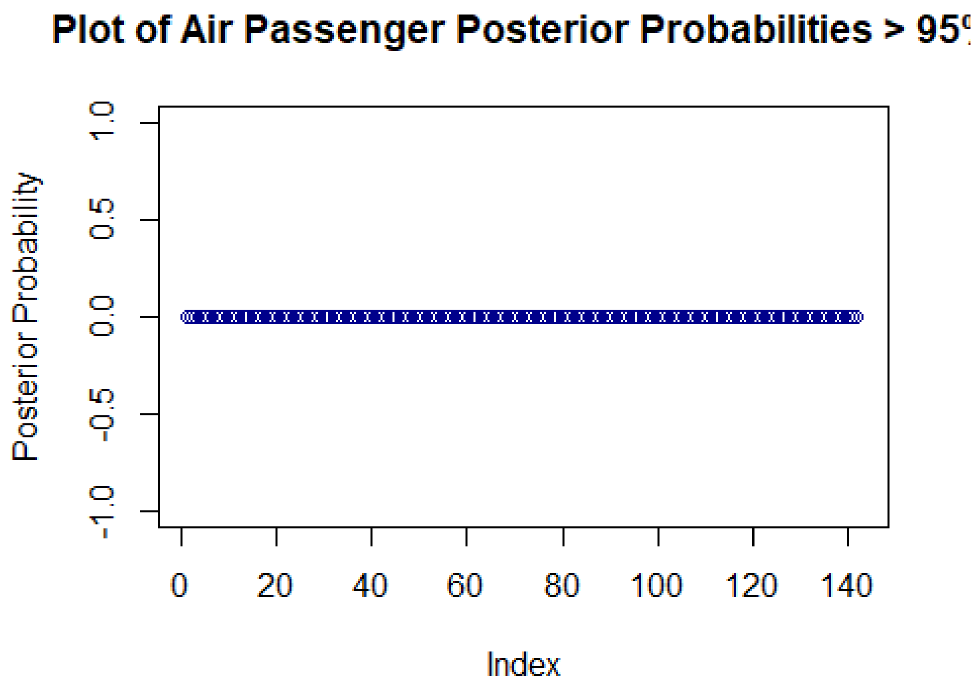
install.packages(“bcp”) library(bcp)

Let’s calculate the bcp and

```
require(bcp)

## Loading required package: bcp

## Loading required package: grid

bcpAir <- bcp(as.vector(diff_air), mcmc = 10000)
plot(bcpAir)
```




```
plot(bcpAir$posterior.prob>.95,
    main="Plot of Air Passenger Posterior Probabilities > 95%",
    ylab="Posterior Probability", col="darkblue")
```



When reviewing the the position probability using a Bayesian change point analysis, we notice a change happens around th 77th point of the graph. This supports the earlier frequentest findings and the change is noticeable on the graph is the change point is