IST772 Problem Set 11

Abhijith Anil Vamadev

The homework for week 12 is based on exercises 2, 5, 6, 7, and 8 on pages 272 and 273 but with changes as noted in this notebook (i.e., follow the problems as given in this document and not the textbook).

Attribution statement: (choose only one) 1. I did this homework by myself, with help from the book and the professor

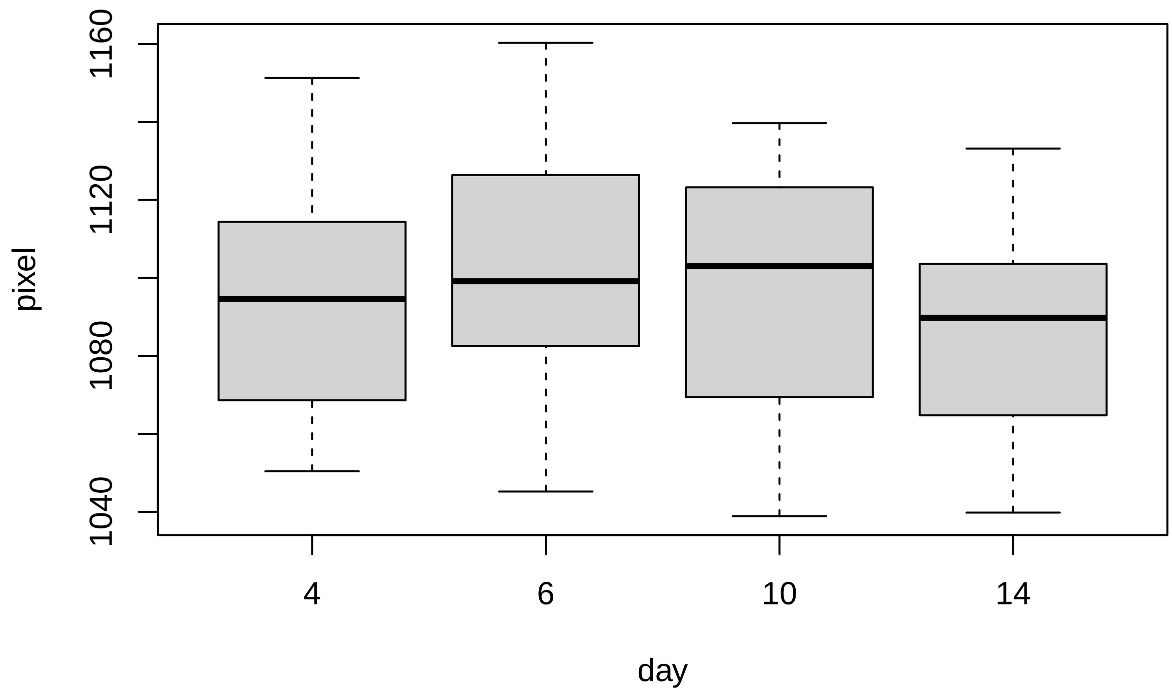
Chapter 11, Exercise 1

Download and library the nlme package and use data("Pixel") to activate the Pixel data set and "? Pixel" for documentation. Inspect the data and create a box plot showing the Pixel intensities on different days. (1 pt) Run a repeated measures ANOVA to compare pixel intensities on days 4, 6, 10 and 14 using aov() (1 pt) and report the results. (1 pt) You can use a command like:

```
#install.packages('nlme')
library(nlme)
```

```
## Warning: package 'nlme' was built under R version 4.1.2
```

```
myData <-subset(Pixel, day %in% c(4,6,10,14))
boxplot(pixel~day, myData)
```



to subset the data. Keeping in mind that the data will need to be balanced before you can conduct this analysis. (1 pt) Try running a command like:

```
new <- myData
table(myData$Dog,myData$day)
```

```
##
##      4 6 10 14
##   1  2 2  2  2
##   10 2 2  0  0
##   2  2 2  2  2
##   3  2 2  2  2
##   4  2 2  2  2
##   5  2 2  2  2
##   6  2 2  2  2
##   7  2 2  2  2
##   8  2 2  2  2
##   9  2 0  0  0
```

```
new$dayFact <- as.factor(new$day) # Convert Time to a factor
list <- rowSums(table(new$Dog,new$dayFact))==8# Make a list of rows
list <- list[list==TRUE] # Keep only those with 8 observations
list <- as.numeric(names(list)) # Extract the row indices
new <- new[new$Dog %in% list,] # Match against the data
table(new$Dog,new$dayFact) #result
```

```
##
##     4 6 10 14
##   1 2 2  2  2
##   2 2 2  2  2
##   3 2 2  2  2
##   4 2 2  2  2
##   5 2 2  2  2
##   6 2 2  2  2
##   7 2 2  2  2
##   8 2 2  2  2
```

```
aovOut<-aov(pixel ~ day+Error(Dog), data = new)
summary(aovOut)
```

```
##
## Error: Dog
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  7  48506    6929
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## day        1    529   528.5   3.635 0.0618 .
## Residuals 55   7997   145.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

as the starting point for cleaning up the data set.

- In the first section of the output, marked "Error: Dog," the "Residuals" refers to the variance attributable to individual differences among chicks. The df = 7 signifies that we have 8 Dogs in the data set. The sum of squares of 48,506 represents variation in the weight variable that is directly attributable to individual differences among dogs.
- In the second section, marked "Error: Within," the effect of Time is expressed as an F-ratio, $F(11,484) = 231.6, p < .001$. The df = 1 for the numerator reflects the 1 points in time where we measured the weights for each of the 7 chicks. The df = 5 for the denominator is the remaining error variance that is not attributable to individual differences. This F-ratio tests the null hypothesis that changes in Dog pixels are consistently 0 across all time intervals. With a p-value of 0.0618 that is not below the conventional threshold of significance we fail to reject this null hypothesis that changes in Dog pixels are consistently 0 across all time intervals.

Chapter 11, Exercise 5

Given that the built-in AirPassengers data set (see "? AirPassengers" for documentation) has a substantial growth trend, use diff() to create a differenced data set. (1 pt) 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. (1 pt) Plot the result and describe in your own words what the change point signifies. (1 pt)

```
AirPassengers
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
```

```
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
```

```r
library(changepoint)
```

```
## Warning: package 'changepoint' was built under R version 4.1.2
```

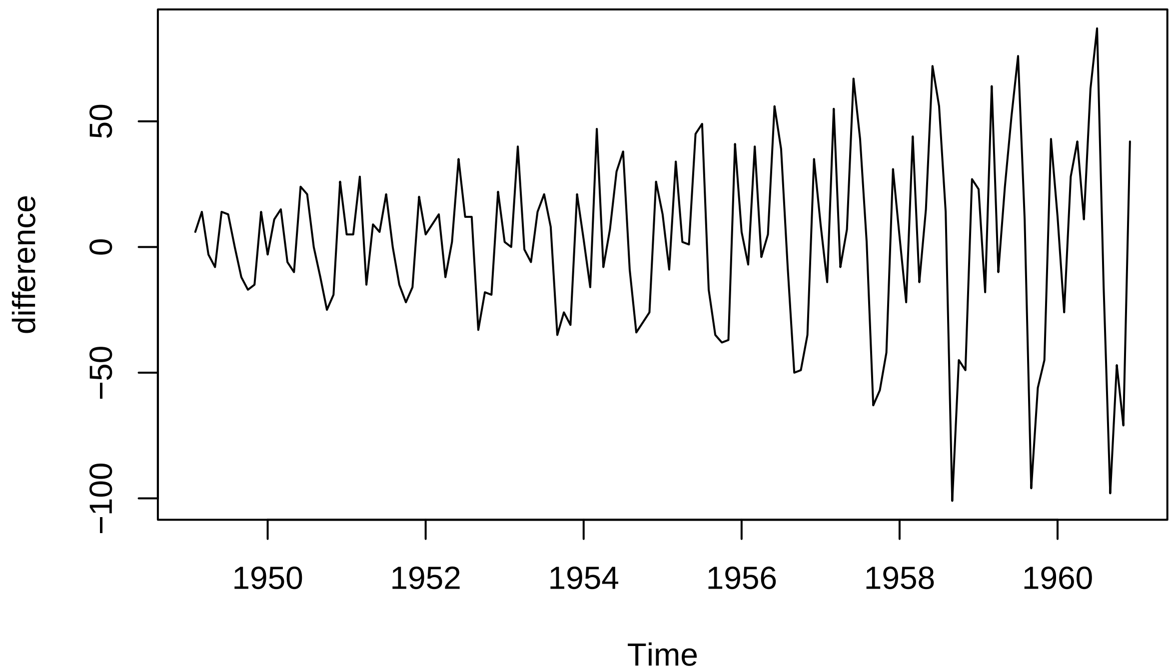
```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

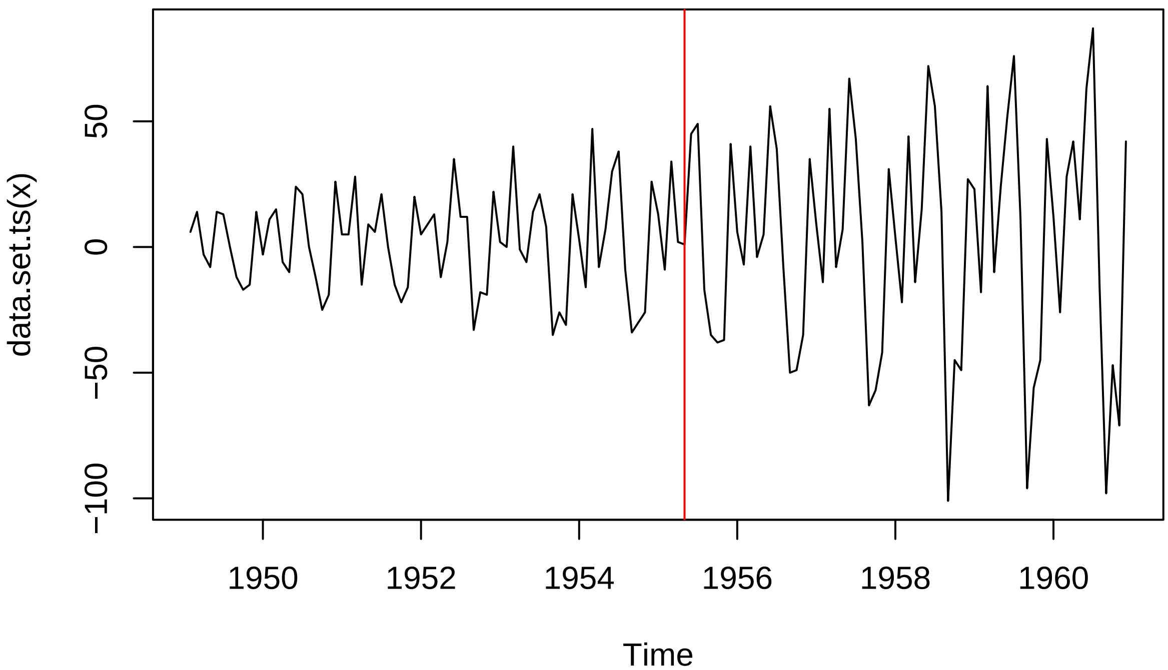
```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## 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
```

```r
difference <- diff(AirPassengers)
plot(difference)
```




```
plot(cpt.var(difference))
```

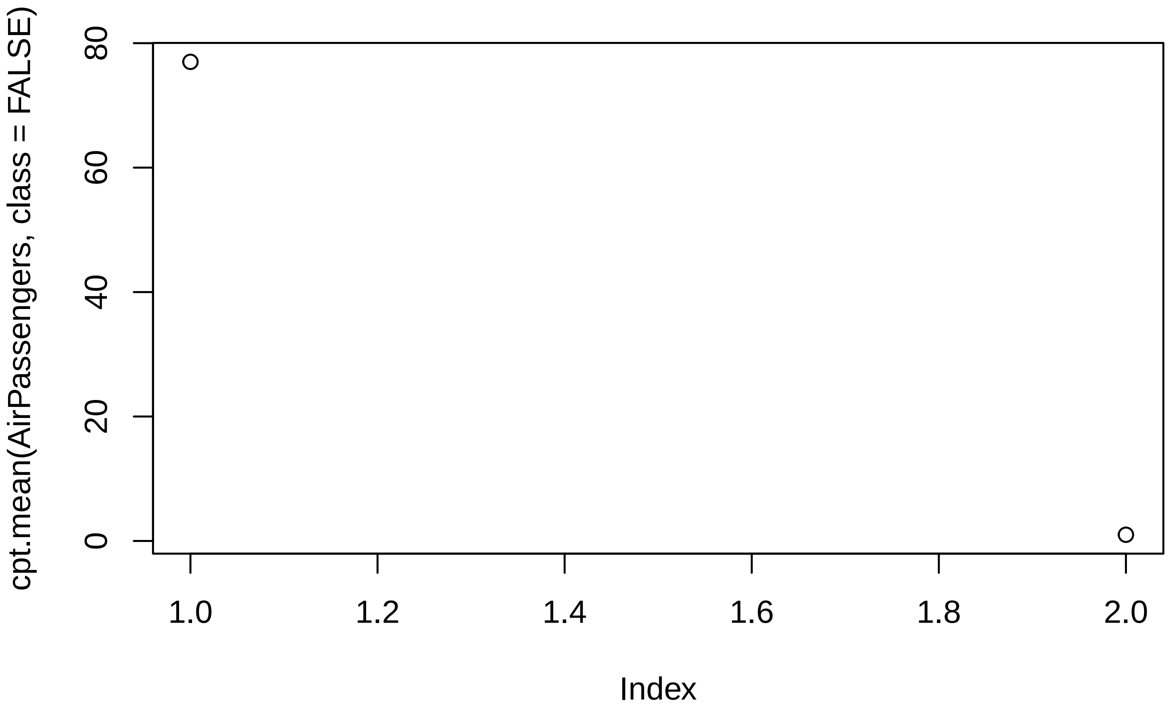


* As the years grew, there is a definite difference increase among air passengers. More monthly totals of international airline passengers are visible through the years. * The redline in the plot indicates the changepoint of the dataset. The changepoint scans the data and find a major shift occured in the mean level of the data. In this plot, the major shift occured in point 76, which is around mid 1955 where the amount of international passengers increased by about 45,000 in one month.

Chapter 11, Exercise 6

Use cpt.mean() on the undifferenced AirPassengers time series. (1 pt) 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? (1 pt)

```
plot(cpt.mean(AirPassengers, class = FALSE))
```



* These change points indicates that there has been a shift in the mean over time.

Chapter 11, 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. (1 pt)

```
#Source:https://www.loveexploring.com/gallerylist/86315/how-air-travel-
#has-changed-in-every-decade-from-the-1920s-to-today
```

- In 1920s, air travel was mostly viewed as a luxury item which was only affordable by the very upper class in socity, and travelers were all rich members from the society. Passengers were treated similarly; they normally find food and drink. However, during 1950s, this trend began to shift. Air travel became more like a necessity. Big airplane started to take a train of passengers. This might explain the results in question 5 and 6. There was a significant passenger increase in 1955.

Chapter 11, Exercise 8

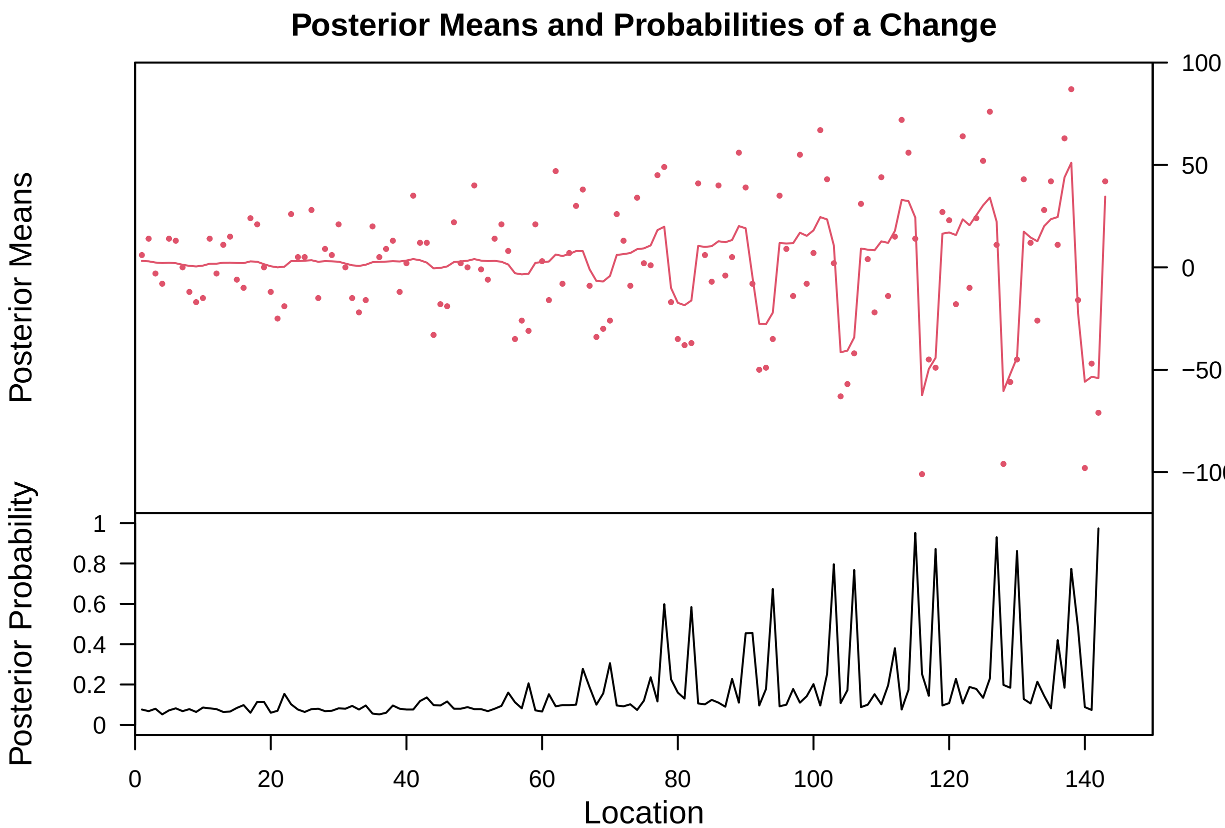
Use bcp() on the AirPassengers time series. (1 pt) Plot and interpret the results. Make sure to contrast these results with those from Exercise 6. (1 pt)

```r
library(bcp)
```

```
## Warning: package 'bcp' was built under R version 4.1.2

## Loading required package: grid
```

```r
bcpAD <- bcp(as.vector(difference))
plot(bcpAD)
```



- The upper pane shows the original time series and the lower pane shows the probabilities of a mean change at each point in time. You will note that there are isolated spikes that show probabilities near 1 at many points across the timeline. Yet, somewhere after the data point of 100, more spikes can be observed, we can see there is substantial density of probability values near 1
- The posterior means and probability plots showed a shift in probability and mean difference around the same location in question 6, 76 was the change point in the previous result, so it can be explained that there was a big shift occurred in the data at this point and after.