

HW10_Tully

IST 772 HW10

12/11/21

Assignment:

The homework for this week are exercises:

- 2 (Stanton,p.272)
- 5, 6, 7, and 8 (Stanton,p.273)

Exercise 2 (Stanton,p.272):

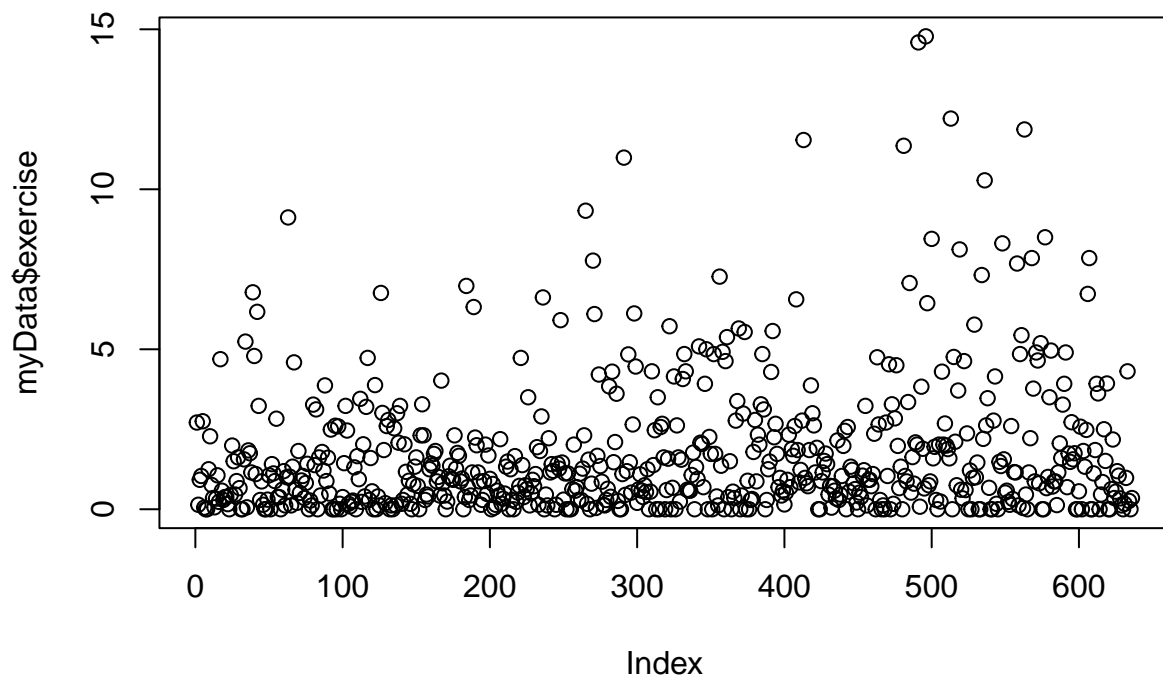
Download and library the nlme package and use data (“Blackmore”) to activate the Blackmore data set. Inspect the data and create a box plot showing the exercise level at different ages. Run a repeated measures ANOVA to compare exercise levels at ages 8, 10, and 12 using aov(). You can use a command like, `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.

```
# display the summary of the data set.
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
```

```
# subset the data
myData <- Blackmore[Blackmore$age<=12,]
# Data cleaning : remove outliers 11.58 & 11.83
myData12 <- Blackmore[Blackmore$age==12,]
myData10 <- Blackmore[Blackmore$age==10,]
myData08 <- Blackmore[Blackmore$age==8,]
#subset/append the three desired ages to myData
myData <- rbind(myData08, myData10, myData12)
#plot the exercise for subset of ages
plot(myData$exercise, main="plot exercise - before balance")
```

plot exercise – before balance

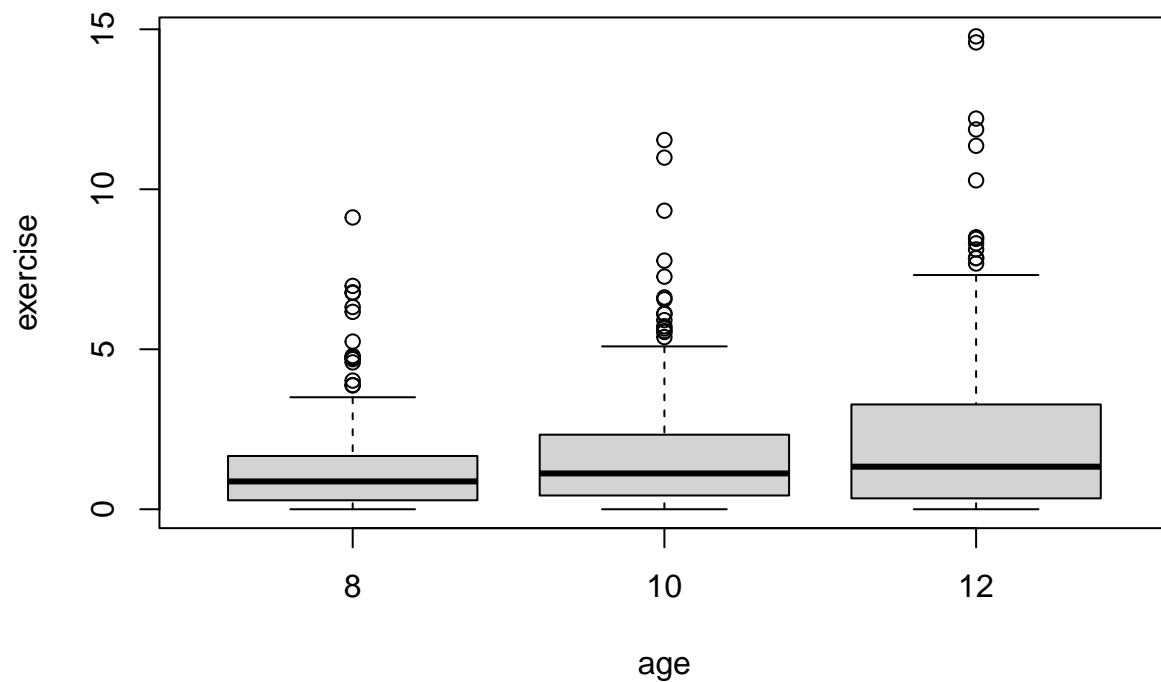


```
#Show unbalanced table  
table(myData$age)
```

```
##  
##      8  10  12  
## 231 229 176
```

```
#boxplot exercise by age - before balance  
boxplot(exercise~age, data=myData, main="boxplot exercise by age - before balance")
```

boxplot exercise by age – before balance

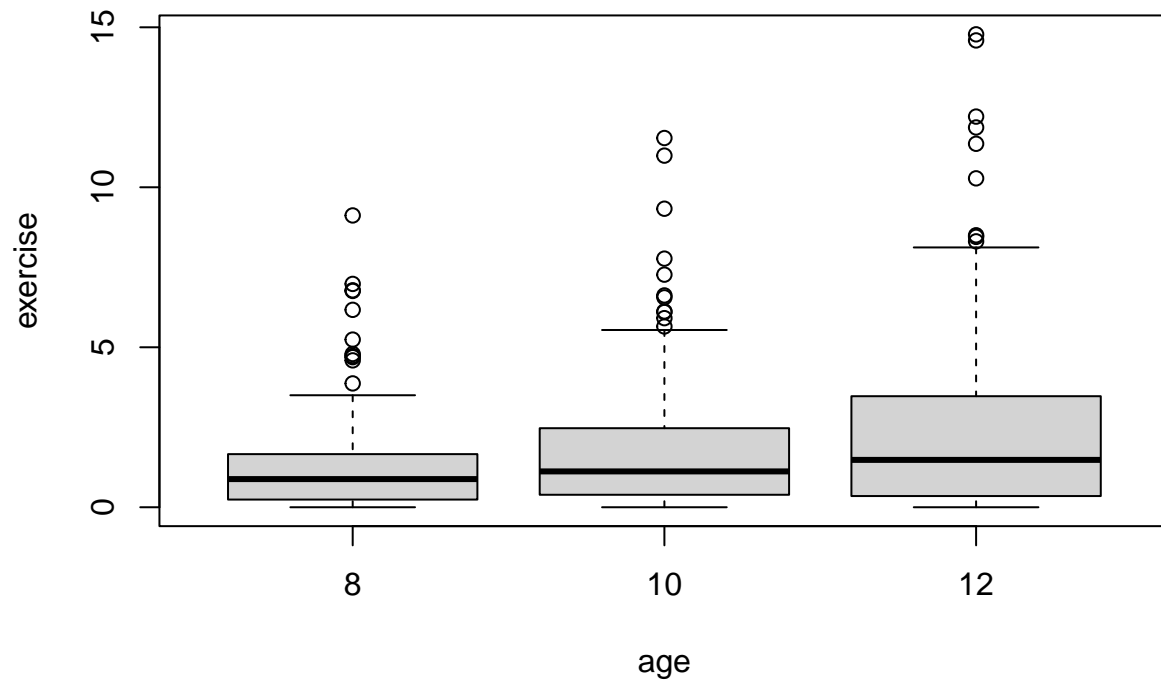


```
#suppress warning start
defaultW <- getOption("warn")
options(warn = -1)
#Make a list of rows - Balance data set
list <- rowSums(table(myData$subject,myData$age))==3
#Keep only those with 3 observations
list <- list[list==TRUE]
#Extract the row indices
list <- as.numeric(names(list))
#suppress warning close
options(warn = defaultW)
#Match against the data
mdBal <- myData[myData$subject %in% list,]
#Show balanced table
table(mdBal$age)
```

```
##
## 8 10 12
## 166 166 166
```

```
#box plot exercise by age - after balance
boxplot(exercise~age, data=mdBal, main="boxplot exercise by age - after balance")
```

boxplot exercise by age – after balance



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

```
##
## 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)
## age         1  102.7   102.66   54.23 1.43e-12 ***
## Residuals 331   626.6     1.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Running a repeated measures ANOVA to compare exercise levels at ages 8, 10, and 12 using `aov()` demonstrates statistical significance in the age variable to determining estimated hours per week dedicated to exercise. The p-value is well below the standard alpha of 0.05. In this case, we would reject the null hypothesis.

Exercise 5 (Stanton,p.273):

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.

```
#add dataset AirPassengers
data("AirPassengers")
# diff() to create a differenced data set
diff(AirPassengers)
```

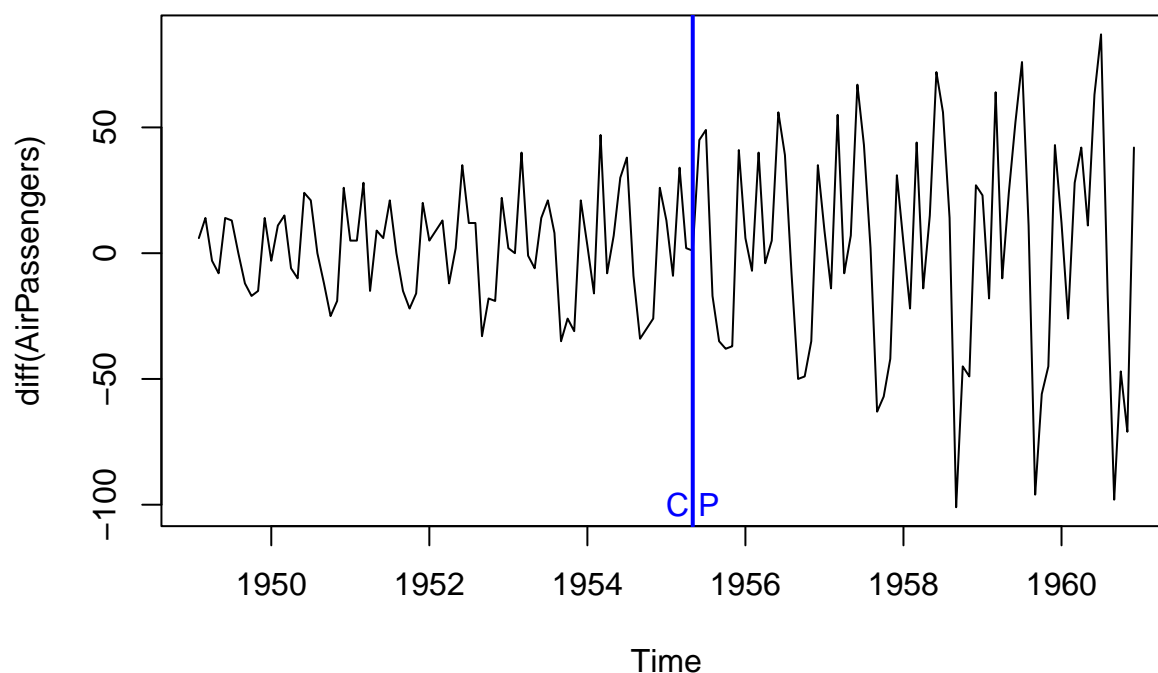
```
##      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1949      6   14   -3   -8   14   13    0  -12  -17  -15   14
## 1950   -3   11   15   -6  -10   24   21    0  -12  -25  -19   26
## 1951    5    5   28  -15    9    6   21    0  -15  -22  -16   20
## 1952    5    9   13  -12    2   35   12   12  -33  -18  -19   22
## 1953    2    0   40   -1   -6   14   21    8  -35  -26  -31   21
## 1954    3  -16   47   -8    7   30   38   -9  -34  -30  -26   26
## 1955   13   -9   34    2    1   45   49  -17  -35  -38  -37   41
## 1956    6   -7   40   -4    5   56   39   -8  -50  -49  -35   35
## 1957    9  -14   55   -8    7   67   43    2  -63  -57  -42   31
## 1958    4  -22   44  -14   15   72   56   14 -101  -45  -49   27
## 1959   23  -18   64  -10   24   52   76   11  -96  -56  -45   43
## 1960   12  -26   28   42   11   63   87  -16  -98  -47  -71   42
```

```
# plot the differenced data set
#plot(diff(AirPassengers))
cvdap <- cpt.var(diff(AirPassengers))
summary(cvdap)
```

```
## 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
```

```
# identify change point
cpLoc <- AirPassengers[cvdap@cpts[1]]
# Add a blue vertical line to plot at changepoint
plot(diff(AirPassengers), main = "AirPassengers Differencing with Variance Changepoint (CP)")
abline(v = 1955.3333, col="blue", lwd = 2)
text(1955.3333, -100, "C P", col="blue")
```

AirPassengers Differencing with Variance Changepoint (CP)



Using the AMOC (“at most one change”) method of analysis with variance as the changepoint type. The changepoint identified, at observation location # 76, is denoted as the first observation of the new segment. With a data set of 144 observations, this changepoint is just slightly right of center or during the year 1955 as reflected on the chart with the vertical blue line.

Exercise 6 (Stanton,p.273):

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?

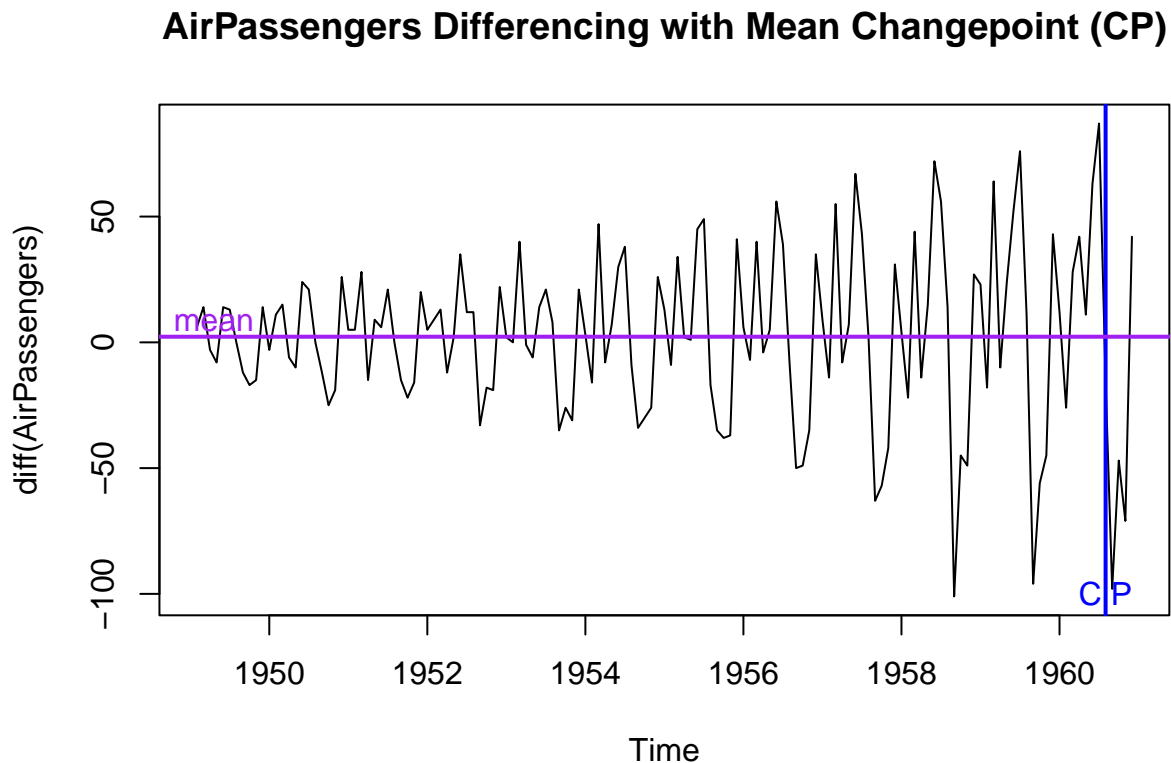
```
#Use cpt.mean() on the AirPassengers time series
cmdap <- cpt.mean(diff(AirPassengers))
summary(cmdap)
```

```
## 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, 14.88853
## Minimum Segment Length : 1
## Maximum no. of cpts   : 1
## Changepoint Locations : 139
```

```

# identify change point
cmLoc <- AirPassengers[cmdap@cpts[1]]
# get the mean of the differencing of AirPassengers (2.2378)
mdap <- mean(diff(AirPassengers))
# Add a blue vertical line to plot at changepoint
plot(diff(AirPassengers), main = "AirPassengers Differencing with Mean Changepoint (CP)")
abline(v = 1960.5833, col="blue",lwd = 2)
text(1960.5833,-100,"C P", col="blue")
abline(h = mdap, col="purple",lwd = 2)
text(1949.3,8,"mean", col="purple")

```



Using the AMOC (“at most one change”) method of analysis with mean as the changepoint type. The changepoint identified, at observation location #139, is denoted points furthest from the differencing mean. With a data set of 144 observations, this changepoint is in the year 1960 as reflected on the chart with the vertical blue line. Comparing this change point of the mean to the change point in the variance uncovered in Exercise 5, we see this one is marked in the 1960s. The air travel advancements made in 1955 had continued volatility on AirPassenger travel well into the 1960.

Exercise 7 (Stanton,p.273):

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.

Air travel changed in the 1950s-1960s: (1) The speed of Jet engines replaced piston-engine airliners, shortening flight times. (2) low-fare night coach service made it more affordable for individuals to fly out of necessity. (3) The arrival of nonstop transcontinental service allowed people to get around the world easier. (The Era of Mass Air Travel Begins | America by Air, 2007)

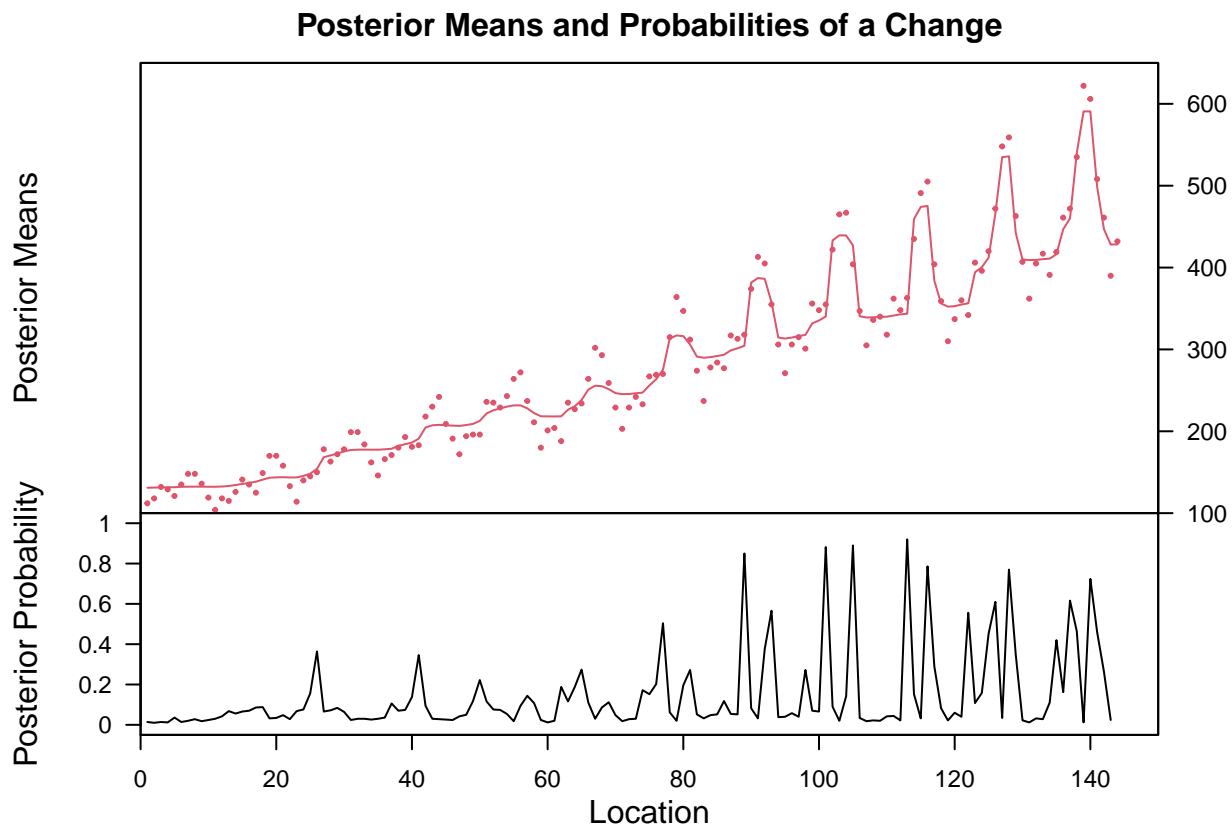
Exercise 8 (Stanton,p.273):

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)
```

```
## Loading required package: grid
```

```
bcpap <- bcp(as.vector(AirPassengers))  
#plot the bcp  
plot(bcpap)
```



```
#check to see if posterior prob > .95  
table(bcpap$posterior.prob > .95)
```



```
##  
## FALSE  
## 143
```

```
#summary on posterior prob  
summary(bcpap$posterior.prob)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
## 0.0100  0.0300  0.0640  0.1453  0.1480  0.9200         1
```

```
#examine the posterior.prob at location 139 (CP in ex. 6)  
bcpap$posterior.prob[139]
```

```
## [1] 0.012
```

This posterior means and probabilities of a change chart denotes that the largest spike is around 101 location (Max Probability 0.928). Even at that max probability point, it still falls short of the 0.95 we would like to observe for statistical significance. In contrast with what we observed from exercise 6, where the change point is shown at location 139 here in the probabilities of change chart only show a 0.032 probability of change at that location.

Reference(s):

Stanton, Jeffrey M.. Reasoning with Data. Guilford Publications. Kindle Edition.

The Era of Mass Air Travel Begins | America by Air. (2007). America By Air. Retrieved December 11, 2021, from <https://airandspace.si.edu/exhibitions/america-by-air/online/heyday/heyday11.cfm>