IST 777

Joe Hernandez

9/12/2019

Homework #10

2. Install and load package

```
> library("carData")
```

Use Blackmore data set and inspect 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
> str(Blackmore)
'data.frame':	945 obs. of  4 variables:
 $ subject : Factor w/ 231 levels "100","101","102",..: 1 1 1 1 1 2 2 2 
2 2 ...
 $ age     : num  8 10 12 14 15.9 ...
 $ exercise: num  2.71 1.94 2.36 1.54 8.63 0.14 0.14 0 0 5.08 ...
 $ group   : Factor w/ 2 levels "control","patient": 2 2 2 2 2 2 2 2 2 
2 ...
> View(Blackmore)
```

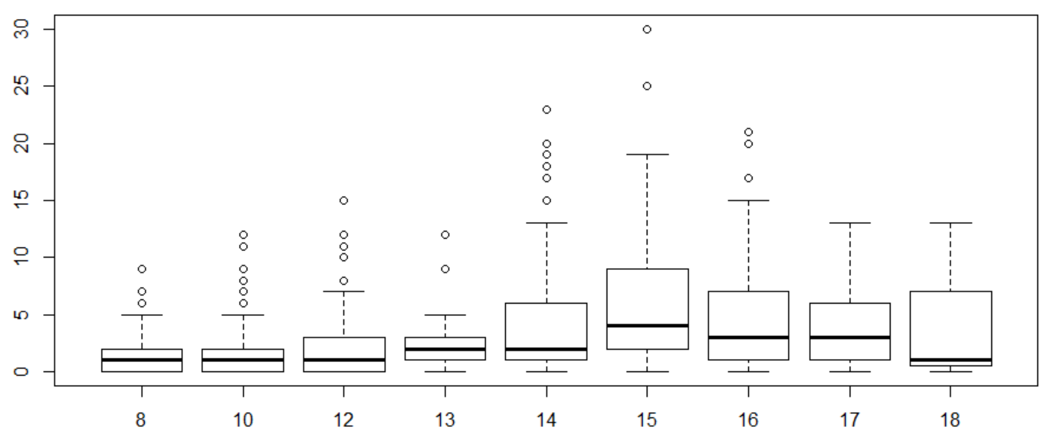
Creating rounded data set for age and exercise for easier processing

```
> Blkmr <- Blackmore
> Blkmr$exercise <- round(Blkmr$exercise, digits = 0)
> Blkmr$age <- round(Blkmr$age, digits = 0)
> ## Detail and justify new data set
> # Inspect new data set
> summary(Blkmr)
    subject        age            exercise          group
 100    :  5   Min.   : 8.00   Min.   : 0.000   control:359
 101    :  5   1st Qu.:10.00   1st Qu.: 0.000   patient:586
 105    :  5   Median :12.00   Median : 1.000
 106    :  5   Mean   :11.43   Mean   : 2.521
 107    :  5   3rd Qu.:14.00   3rd Qu.: 3.000
 108    :  5   Max.   :18.00   Max.   :30.000
 (Other):915
> str(Blkmr)
'data.frame':	945 obs. of  4 variables:
 $ subject : Factor w/ 231 levels "100","101","102",..: 1 1 1 1 1 2 2 2
```

```
2 2 ...
 $ age     : num  8 10 12 14 16 8 10 12 14 17 ...
 $ exercise: num  3 2 2 2 9 0 0 0 0 5 ...
 $ group   : Factor w/ 2 levels "control","patient": 2 2 2 2 2 2 2 2 2
2 ...
```

Create box plot showing exercise level at different ages

```
> boxplot(exercise ~ age, data = Blkmr)
```



The boxplot shows an improving trend in the exercise level of kids from ages 8 – 15. Their peak exercise is achieved at 15 where the three subsequent years show a decline in exercise levels.

Run repeated measured ANOVA comparing exercise levels at ages 8, 10, and 12 (I struggled with this but this was the best I could produce)

```
> myData <- Blkmr[Blkmr$age <= 12, ]
> # Balance data before analysis (get all 1 across table)
> Blkmrage <- table(myData$subject, myData$age)
> Blkmrage <- Blkmrage[which(Blkmrage[, 1] == 1 & Blkmrage[, 2] == 1 &
Blkmrage[, 3] == 1),]
>
> # Subset age 8
> Blkmr8 <- Blkmr[which(Blkmr$age == 8),]
> aov(exercise ~ age, data = Blkmr8)
Call:
   aov(formula = exercise ~ age, data = Blkmr8)

Terms:
                Residuals
Sum of Squares   543.4372
Deg. of Freedom       230

Residual standard error: 1.537131
```

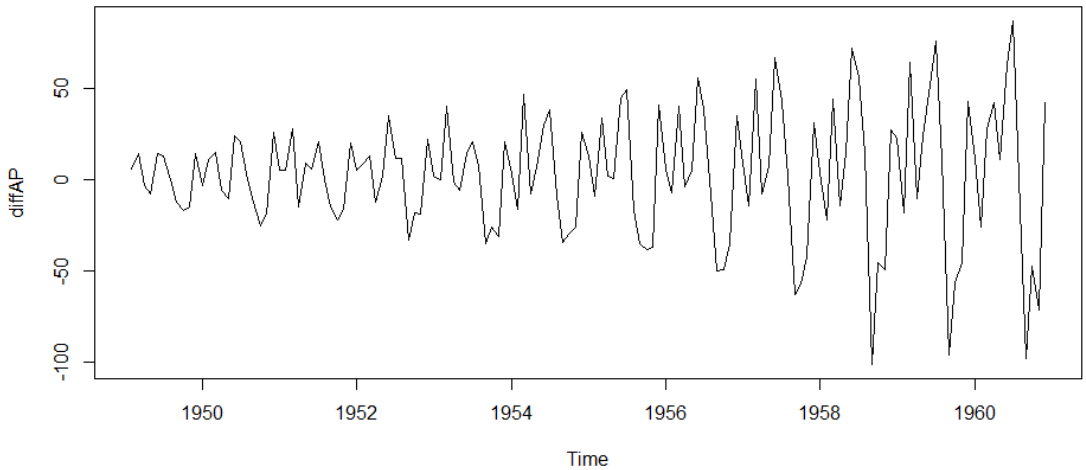
```
> # Subset age 10
> Blkmr10 <- Blkmr[which(Blkmr$age == 10),]
> aov(exercise ~ age, data = Blkmr10)
Call:
   aov(formula = exercise ~ age, data = Blkmr10)

Terms:
                Residuals
Sum of Squares   918.3057
Deg. of Freedom       228

Residual standard error: 2.006902
> # Subset age 12
> Blkmr12 <- Blkmr[which(Blkmr$age == 12),]
> aov(exercise ~ age, data = Blkmr12)
Call:
   aov(formula = exercise ~ age, data = Blkmr12)

Terms:
                Residuals
Sum of Squares   1466.201
Deg. of Freedom       188

Residual standard error: 2.792659
```

5. Create a differenced data set of AirPassengers data set

```
> diffAP <- diff(AirPassengers)
> diffAP
      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 to examine and interpret results

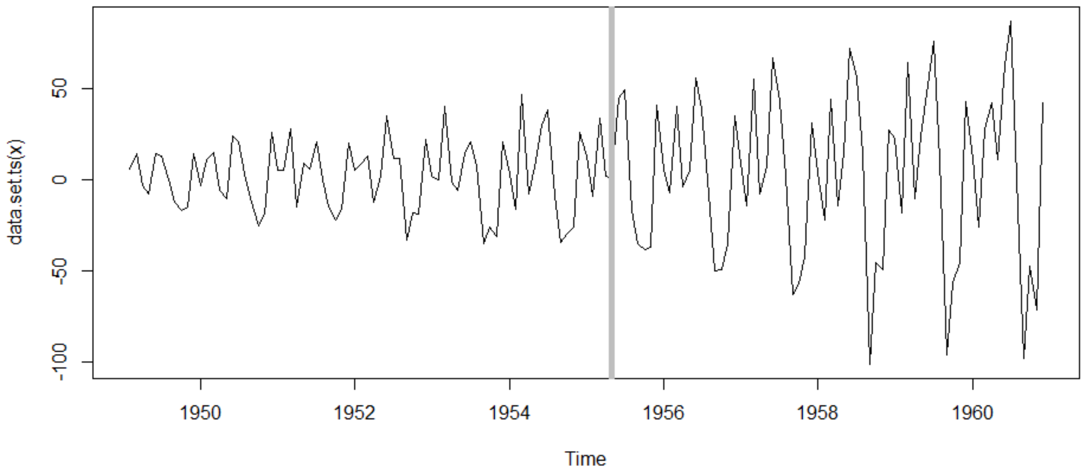
```
> plot(diffAP)
```



The plot is a noise component after decomposition in air passengers but shows signs of seasonality for each year. The Y-scale for air passengers is huge ranging from -100 to 100 in variability. The overall plot has an expanding variability over time where, around 1955 and onward, we notice the variability significant where it could escape the threshold of a normal trend.

Use cpt.var() to find change point in variability

```
> diffAPcp <- cpt.var(diffAP)
> diffAPcp
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  : Mon Sep 02 17:42:55 2019

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

Plot result

> plot(diffAPcp,cpt.col="grey",cpt.width=5)
```



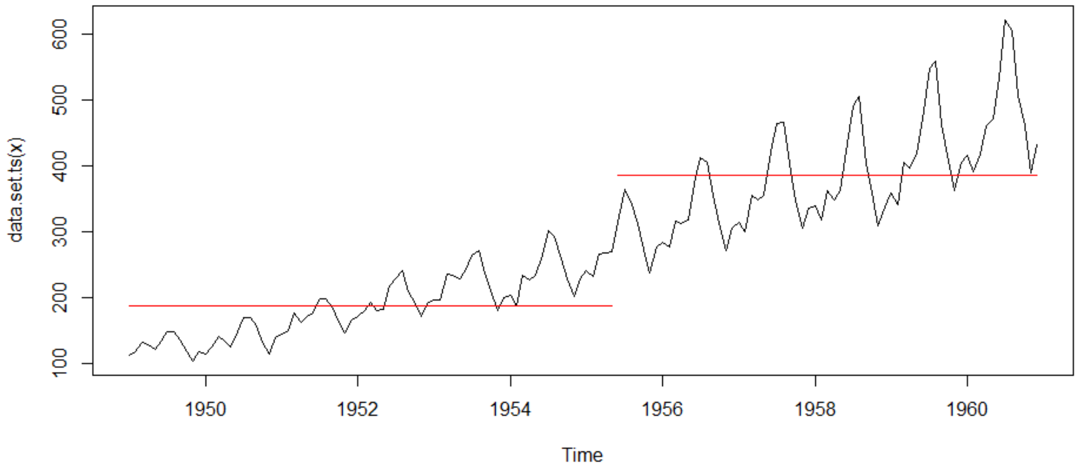
The plot shows the change point in variability at 14.88853 as the most significant where this occurs at location 76 slightly beyond 1955.

6. Use cpt.mean() on AirPassengers time series

```
> APcp <- cpt.mean(AirPassengers)
> APcp
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  : Mon Sep 02 17:42:55 2019

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 result

```
> plot(APcp)
```



The changepoint analysis using the mean shows a similar result where significant change to the average number of air passengers increased beyond it's normal threshold of 14.90944 just beyond 1955 at location 77.

Compared to the changepoint in variability, the changepoint in mean has closely similar results in both penalty and changepoint location where it is almost negligible to switch between the two without being too far off.

This suggests there must be a significant event happening in the mid to late 1950's that would allow for a noticeable change in air passengers.

7. Find historical information on air travel shedding light to results from Exercises 5 & 6

   In the early 1950's, jet aircraft was still in its infancy and many airliners maintained stock of propeller piston-engine airliners. By 1955, the jet aircraft was more economical in range and fuel consumption where it became more mainstream mode of air transportation. This would result in direct flights to farther destinations and less fuel expended from improved aerodynamics. Reports show the jet engine achieving 60% less fuel consumption than its predecessor leading to nearly double the flying range favorable for international flights. Ideally the propeller engine was favored for shorter distance travelling but the jet engine offered more convenience and safety as well. Propeller and piston-engine aircraft were also considered more dangerous means of transportation compared to the jet engines. Eventually as time progressed, airliners would adopt jet engine aircraft to transport travelers once their aircraft life cycle for propeller planes offered no additional benefit to using them. Additionally, flying in that time period was exclusive to the rich with personal services which made flying a unique experience such as cooking services and to increase flying throughput, the focus was on convenience rather than the novelty of flying. Commercial flying in the early 50's were exclusive to wealthy celebrities, politicians, businessmen, and sports teams. With the introduction of a reduced fare for air service this attracted tourists which turned air travel to class systems such as tourist and economy classes. As a result this would increase the capacity of aircraft while sacrificing earlier amenities. Business decisions helped in driving the increased variability during the "Golden Age of Flight" consistent with the data we observed.
   https://airandspace.si.edu/exhibitions/america-by-air/online/heyday/heyday11.cfm

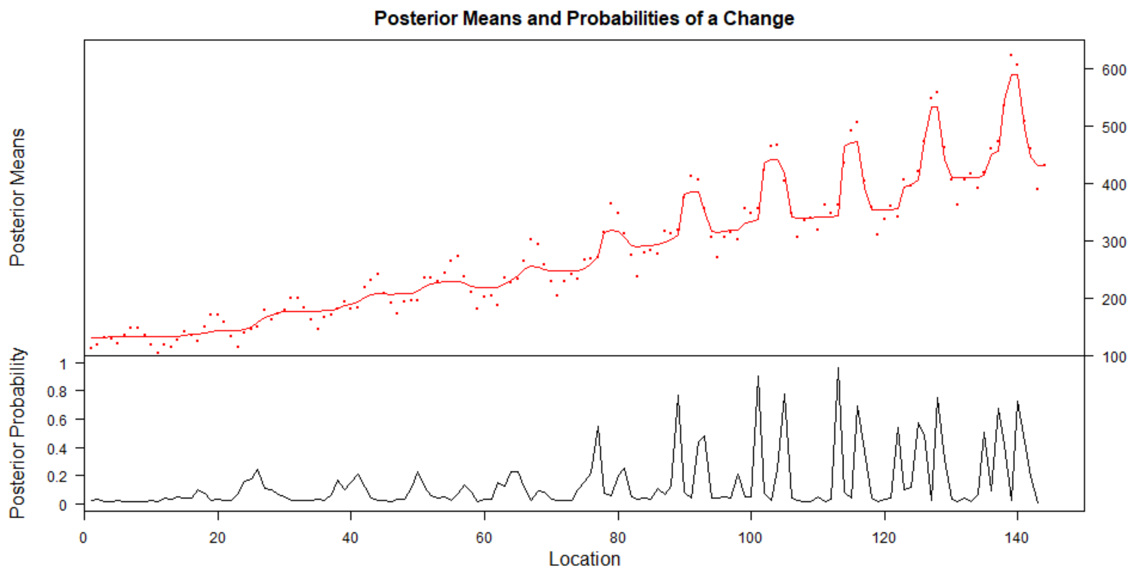
https://www.thevintagenews.com/2019/04/13/air-travel-was-wonderful/
https://www.retrowow.co.uk/transport/50s/Air_travel_50s.html

8. Use bcp() on AirPassengers time series

```
> APbcp <- bcp(AirPassengers)
```

Plot results

```
> plot(APbcp)
```



The posterior means shows a gradual increasing trend in air passengers with some cyclical patterns; however, around location 80, there is a noticable difference in the trend with larger peaks and troughs in its cyclicality. When observing the posterior probability, around location 76-77 (similar to change point analyses for means and variance) there is a drastic change in the probabilities. It adds value where after 1955 there was an event that would change the data outside of its normal trend.