Parin Patel
IST 772
HW10

Attribution Statement:

Homework 10 by Parin Patel: I did this homework by myself, with help from the book and the professor. I used these websites to gather historical information related to aviation:

- https://airandspace.si.edu/exhibitions/america-by-air/online/heyday/heyday11.cfm
- https://en.wikipedia.org/wiki/1955_in_aviation
- https://www.fastcompany.com/3022215/what-it-was-really-like-to-fly-during-the-golden-age-of-travel

Exercises:

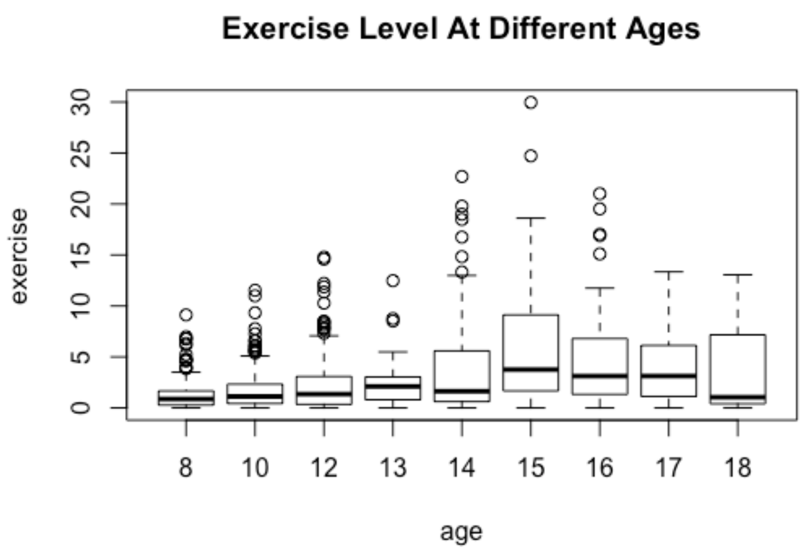
2. Download and library the nlme package and use data ("Blackmore") to activate the Blackmore data set. Inspect the data and create a box plot showing the exercise level at different ages. Run a repeated measures ANOVA to compare exercise levels at ages 8, 10, and 12 using aov(). You can use a command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

The findings of our repeated measure ANOVA for exercise levels at ages 8,10,and 12 showed that, on average, people this age got about 1 hour of exercise a week. This is based on the determined median value of 1. Additionally, as they get older, the more varied the amount of exercise they get becomes. This is because we see the 3rd and 4th quantiles increase as they get older. Same goes for the max. Finally, when we ran our aov function to analyze variance, we continued to see exercise increase with age. The results showed a p-value of 1.33e-11, which means there is a significance between age and the amount of exercise a person does. Also, based on our boxplot, we see that 15 is the age with the highest variance in exercise times. In addition, it has the highest median exercise amount. We start to see a overall dip in the median at the age of 18.

```
> 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 ...
```

Parin Patel
IST 772
HW10

```
> summary(Blackmore)
    subject           age            exercise
 100     :  5   Min.   : 8.00   Min.   : 0.000
 101     :  5   1st Qu.:10.00   1st Qu.: 0.400
 105     :  5   Median :12.00   Median : 1.330
 106     :  5   Mean   :11.43   Mean   : 2.531
 107     :  5   3rd Qu.:14.00   3rd Qu.: 3.040
 108     :  5   Max.   :18.00   Max.   :29.960
 (Other):915
      group
 control:359
 patient:586
```




```
> #Create ANOVA to compare exercise levels at ages 8, 10, and 12
> myData <- Blackmore[Blackmore$age <= 12,]
> table(myData$subject,myData$age)

       8 10 11.58 11.83 12
```

Parin Patel
IST 772
HW10

```
> is.factor(myData$age) #check if age is factor
[1] FALSE
> myData$ageFact<-as.factor(myData$age)  #change age to factor
> list <- rowSums(table(myData$subject,myData$ageFact))==3
> list <- list[list == TRUE]
> list <- as.numeric(names(list))
Warning message:
NAs introduced by coercion
> summary(myData[myData$ageFact == 8,])
    subject          age         exercise          group       ageFact  
 100    :  1   Min.   :8   Min.   :0.000   control: 93   8    :231  
 101    :  1   1st Qu.:8   1st Qu.:0.280   patient:138   10   :  0  
 102    :  1   Median :8   Median :0.870                 11.58:  0  
 103    :  1   Mean   :8   Mean   :1.259                 11.83:  0  
 104    :  1   3rd Qu.:8   3rd Qu.:1.665                 12   :  0  
 105    :  1   Max.   :8   Max.   :9.120                            
 (Other):225                                                        
```

```
> summary(myData[myData$ageFact == 10,])
    subject          age          exercise           group       ageFact  
 100    :  1   Min.   :10   Min.   : 0.000   control: 92   8    :  0  
 101    :  1   1st Qu.:10   1st Qu.: 0.430   patient:137   10   :229  
 102    :  1   Median :10   Median : 1.120                 11.58:  0  
 103    :  1   Mean   :10   Mean   : 1.746                 11.83:  0  
 104    :  1   3rd Qu.:10   3rd Qu.: 2.330                 12   :  0  
 105    :  1   Max.   :10   Max.   :11.540                            
 (Other):223                                                          
> summary(myData[myData$ageFact == 12,])
    subject          age          exercise           group       ageFact  
 100    :  1   Min.   :12   Min.   : 0.000   control: 59   8    :  0  
 101    :  1   1st Qu.:12   1st Qu.: 0.340   patient:117   10   :  0  
 102    :  1   Median :12   Median : 1.330                 11.58:  0  
 103    :  1   Mean   :12   Mean   : 2.307                 11.83:  0  
 104    :  1   3rd Qu.:12   3rd Qu.: 3.272                 12   :176  
 105    :  1   Max.   :12   Max.   :14.780                            
 (Other):170                                                          
```

Parin Patel
IST 772
HW10

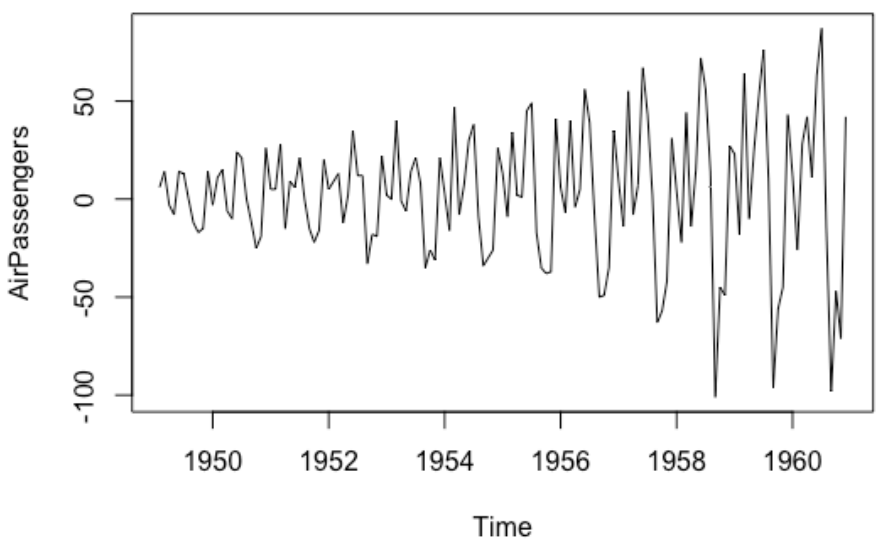
```
> myData <- myData[myData$subject %in% list,]
> summary(aov(exercise~ageFact+ Error(subject), data = myData))

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
>
```

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.

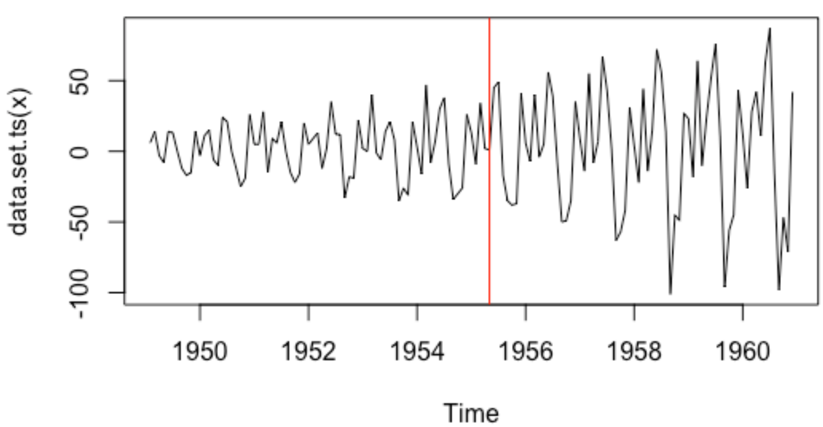
We used the diff(AirPassengers) function to determine the difference between the number of airline passengers, by month. This means that we compared the number of passengers at a certain time (month) against the number of passengers at a previous month to create a time series plot that plots the differences visually to show a trend over time. The red line we see in the second plot shows the changepoint in our dataset. This shows that the major shift in the data occurred in 1955, when there was an increase about 45,000 passengers from the numeric value of 0. This is likely because, based on a review of major events in aviation that year, the mid-1950s were considered the “Golden Age of Flying” due to the rise in the number of more affordable airline tickets and jumbo-jets.

Parin Patel
IST 772
HW10

Plot of Differenced Data Set



Plot of Change Point in Variability



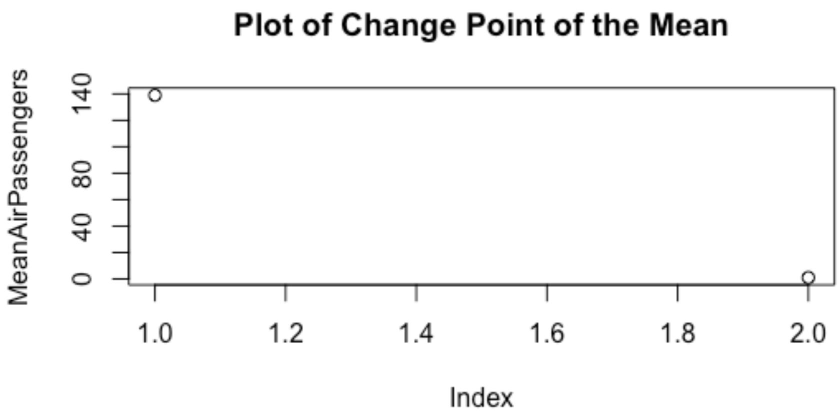
Parin Patel
IST 772
HW10

```
> #####Question 5:
> library("changepoint")
> data("AirPassengers")
> AirPassengers <- diff(AirPassengers) #create differenced data set.
> plot(AirPassengers,main="Plot of Differenced Data Set") #plot
>
> AP_Plot<-cpt.var(AirPassengers) #find change point in variability
> plot(AP_Plot, main="Plot of Change Point in Variability")
> AP_Plot
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  : Fri Apr 26 15:58:35 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
>
```

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?

By plotting the cpt.mean of the airline passenger data, we get a confidence value of exactly 1. Our change point analysis from the previous exercise (#5) shows that there has been an invent in 1955 that caused a surge in passengers in the time series data. Since our function used was cpt.mean, our change point analysis was focused in #6 on the mean. Our analysis in #5 was focused on the variance. From number 5's second change point line in the time series graph, we can see that the numeric value of the passengers (x) is 0 before it increases and then drops towards the end of 1995. Based on this, we can expect that our change point around the mean to be 0, which is the result of our analysis for #6. Therefore, the results of our overall findings suggests that there has been a change in the mean number of passengers over time.

Parin Patel
IST 772
HW10




```
> ####Question 6:
> # retrieve the confidence value
> MeanAirPassengers <- cpt.mean(AirPassengers, class = FALSE)
> MeanAirPassengers
          cpt conf.value
          139          1
> plot(MeanAirPassengers,main="Plot of Change Point of the Mean")
> MeanAirPassengers["conf.value"] #confidence value
conf.value
         1
>
```

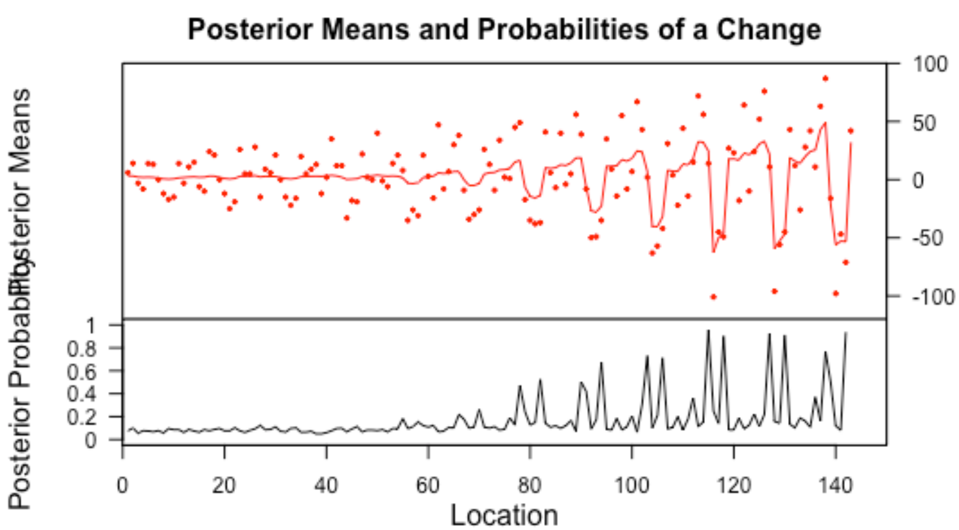
7. Find historical information about air travel on the Internet and/or in reference materials that sheds light on the results from Exercises 5 and 6. Write a mini-article (less than 250 words) that interprets your statistical findings from Exercises 5 and 6 in the context of the historical information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

I briefly touched on this earlier in my answers, but the 1950's were considered the Golden Age of plane travel. According to the Smithsonian National Air and Space Museum, the 1950s was the age that saw the air travel become more common. The service became more widely spread and commonplace. For passengers, while tickets were still expensive compared to today's values, more passengers were able to buy them and travel to a increasing number of destinations. Airlines were becoming more comfortable and faster for public travel as well. They offered low-fare night coach services and had "red eye" specials. The piston-engine airplanes from the 1940's was replaced with modern day jet aircrafts. This also made flying more efficient. It was the age where airlines were increasing in the number of passengers they could fly and the distance they were flying them. This would explain the rise and change in passengers flying in our dataset. It was a change in an era that made air travel an important necessity.

Parin Patel
IST 772
HW10

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

Our Bayesian approach to time series yielded similar results as our frequentist model in number 6. Based on the graph below, we can see that both the posterior probability (bottom line) and the posterior mean (top line) plots show a change in both probability and mean difference around the 76th location. This point, 76th, is the same change point we calculated in #6, when using the frequentist approach. We can then say with some certainty that these findings show that shift has indeed occurred in the dataset in and after the 76th location.



Appendix I: Source Code

```
##Question 2:
library("nlme")
library("carData")


data("Blackmore")
Blackmore <- data.frame(Blackmore)
str(Blackmore) #decimal place noticed. Lets remove by rounding.
Blackmore$age <- round(Blackmore$age)
summary(Blackmore)

boxplot(exercise~age, data = Blackmore, main="Exercise Level At Different Ages")#get boxplot
#Create ANOVA to compare exercise levels at ages 8, 10, and 12
```

Parin Patel
IST 772
HW10

```
myData <- Blackmore[Blackmore$age <= 12,]
table(myData$subject,myData$age)
is.factor(myData$age) #check if age is factor
myData$ageFact<-as.factor(myData$age)  #change age to factor
list <- rowSums(table(myData$subject,myData$ageFact))==3
list <- list[list == TRUE]
list <- as.numeric(names(list))

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

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

#####Question 5:
library("changepoint")
data("AirPassengers")
AirPassengers <- diff(AirPassengers) #create differenced data set.
plot(AirPassengers,main="Plot of Differenced Data Set") #plot

AP_Plot<-cpt.var(AirPassengers) #find change point in variability
plot(AP_Plot, main="Plot of Change Point in Variability")



####Question 6:
# retrieve the confidence value
MeanAirPassengers <- cpt.mean(AirPassengers, class = FALSE)
plot(MeanAirPassengers,main="Plot of Change Point of the Mean")
MeanAirPassengers["conf.value"] #confidence value


###Question 8:
library("bcp")
bcpAirPassengers <- bcp(as.vector(AirPassengers))
plot(bcpAirPassengers)
```

Appendix II: Output

```
> ##Question 2:
> library("nlme")
> library("carData")
```

Parin Patel
IST 772
HW10

```
>
>
> data("Blackmore")
> Blackmore <- data.frame(Blackmore)
> str(Blackmore) #decimal place noticed. Lets remove by rounding.
'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 ...
> Blackmore$age <- round(Blackmore$age)
> summary(Blackmore)
    subject         age           exercise
 100    :  5   Min.   : 8.00   Min.   : 0.000
 101    :  5   1st Qu.:10.00   1st Qu.: 0.400
 105    :  5   Median :12.00   Median : 1.330
 106    :  5   Mean   :11.43   Mean   : 2.531
 107    :  5   3rd Qu.:14.00   3rd Qu.: 3.040
 108    :  5   Max.   :18.00   Max.   :29.960
 (Other):915
     group
 control:359
 patient:586
```

```
>
> boxplot(exercise~age, data = Blackmore, main="Exercise Level At Different Ages")#get
boxplot
> #Create ANOVA to compare exercise levels at ages 8, 10, and 12
> myData <- Blackmore[Blackmore$age <= 12,]
> table(myData$subject,myData$age)

      8 10 12
  100  1  1  1
  101  1  1  1
  102  1  1  1
  103  1  1  1
  104  1  1  1
  105  1  1  1
  106  1  1  1
  107  1  1  1
  108  1  1  1
  109  1  1  0
```

Parin Patel
IST 772
HW10

110	1	1	1
111	1	1	1
112	1	1	1
113	1	1	1
114	1	1	1
115	1	1	1
116	1	1	1
117	1	1	1
118	1	1	1
119	1	1	1
120	1	1	1
121	1	1	1
122	1	1	1
123	1	1	1
124	1	1	1
125	1	1	1
126	1	1	1
127	1	1	1
128	1	1	0
129	1	1	1
130	1	1	1
132	1	1	0
133	1	1	1
134	1	1	1
135	1	1	1
136	1	1	1
137	1	1	1
138	1	1	1
139	1	1	1
140	1	1	1
141	1	1	1
142	1	1	1
143	1	1	1
144	1	1	1
145	1	1	1
146	1	1	1
147	1	1	0
148	1	1	1
149	1	1	1
150	1	1	1
151	1	1	1
152	1	1	1
153	1	1	1
154	1	1	1
155	1	1	0
156	1	1	1

Parin Patel
IST 772
HW10

157	1	1	1
158	1	1	1
159	1	1	1
160	1	1	1
161	1	1	1
162	1	1	0
163	1	1	1
164	1	1	1
165	1	1	1
166	1	1	1
167	1	1	1
168	1	0	1
169	1	1	0
170	1	1	1
171	1	1	1
172	1	1	1
173	1	1	1
174	1	1	0
175	1	1	1
176	1	1	1
177	1	1	1
178	1	1	1
179	1	1	1
180	1	1	1
181	1	1	1
182	1	1	1
183	1	1	1
184	1	1	0
185	1	1	1
186	1	1	1
187	1	1	1
188	1	1	1
189	1	1	1
190	1	1	1
192	1	1	1
193	1	1	0
194	1	1	1
195	1	1	1
196	1	1	0
198	1	1	1
199	1	1	1
200	1	1	1
201	1	1	0
202	1	1	1
203	1	1	1
204	1	1	1

Parin Patel
IST 772
HW10

205	1	1	1
206	1	1	1
207a	1	1	1
207b	1	1	1
208	1	1	1
209	1	1	1
210	1	1	1
211	1	1	0
212	1	1	1
213	1	1	1
214	1	1	1
215	1	1	1
216	1	1	0
217	1	1	0
218	1	1	1
219	1	1	1
220	1	1	1
221	1	1	1
222	1	0	1
223	1	1	0
224	1	1	1
225	1	1	1
226	1	1	0
227	1	1	0
228	1	1	1
229a	1	1	1
229b	1	1	1
230	1	1	1
231	1	1	1
232	1	1	1
233	1	1	1
234	1	1	0
235	1	1	1
236	1	1	1
237	1	1	1
238	1	1	1
239	1	1	0
240	1	1	1
241	1	1	0
242	1	1	1
243	1	1	1
244	1	1	1
245	1	1	1
246	1	1	1
247	1	1	1
248	1	1	0

Parin Patel
IST 772
HW10

249	1	1	1
250	1	1	1
251	1	1	1
252	1	1	1
253	1	1	1
254	1	1	1
255	1	1	1
255b	1	1	1
256	1	1	1
257	1	1	1
258	1	1	0
259	1	1	1
260	1	1	0
261	1	1	0
262	1	1	0
263	1	1	0
264	1	1	0
265	1	1	0
266	1	1	0
267	1	1	1
268	1	1	1
269	1	1	0
270	1	1	1
271	1	1	1
272	1	1	0
273a	1	1	1
273b	1	1	1
274	1	1	0
275	1	1	1
276	1	1	1
277	1	1	1
278	1	1	1
279a	1	1	0
279b	1	1	1
280a	1	1	1
280b	1	1	1
281	1	1	1
282	1	1	1
283	1	1	0
284	1	1	1
285	1	1	0
286	1	1	1
300	1	1	1
301	1	1	1
302	1	1	1
303	1	1	1

Parin Patel
IST 772
HW10

```
  304  1  1  1
  305  1  1  1
  306  1  1  1
  307  1  1  1
  308  1  1  1
  309  1  1  1
  310  1  1  1
  311  1  1  1
  312  1  1  1
  313  1  1  1
  314  1  1  1
  315  1  1  0
  316  1  1  0
  317  1  1  1
  318  1  1  1
  319  1  1  1
  320  1  1  1
  321  1  1  1
  322  1  1  0
  323  1  1  1
  324  1  1  0
  325  1  1  1
  326  1  1  1
  327  1  1  1
  328  1  1  1
  329  1  1  1
  330  1  1  1
  331  1  1  1
  332  1  1  1
  333  1  1  1
  334  1  1  1
  335  1  1  1
  336  1  1  1
  337  1  1  0
  338  1  1  1
  340  1  1  1
  341  1  1  0
> is.factor(myData$age) #check if age is factor
[1] FALSE
> myData$ageFact<-as.factor(myData$age)  #change age to factor
> list <- rowSums(table(myData$subject,myData$ageFact))==3
> list <- list[list == TRUE]
> list <- as.numeric(names(list))
Warning message:
NAs introduced by coercion
>
```

Parin Patel
IST 772
HW10

```
> summary(myData[myData$ageFact == 8,])
   subject        age       exercise         group
 100    :  1   Min.   :8   Min.   :0.000   control: 93
 101    :  1   1st Qu.:8   1st Qu.:0.280   patient:138
 102    :  1   Median :8   Median :0.870
 103    :  1   Mean   :8   Mean   :1.259
 104    :  1   3rd Qu.:8   3rd Qu.:1.665
 105    :  1   Max.   :8   Max.   :9.120
 (Other):225
 ageFact
 8 :231
 10:  0
 12:  0
```

```
> summary(myData[myData$ageFact == 10,])
   subject        age        exercise          group
 100    :  1   Min.   :10   Min.   : 0.000   control: 92
 101    :  1   1st Qu.:10   1st Qu.: 0.430   patient:137
 102    :  1   Median :10   Median : 1.120
 103    :  1   Mean   :10   Mean   : 1.746
 104    :  1   3rd Qu.:10   3rd Qu.: 2.330
 105    :  1   Max.   :10   Max.   :11.540
 (Other):223
 ageFact
 8 :  0
 10:229
 12:  0
```

```
> summary(myData[myData$ageFact == 12,])
   subject        age        exercise          group
 100    :  1   Min.   :12   Min.   : 0.000   control: 68
 101    :  1   1st Qu.:12   1st Qu.: 0.350   patient:121
 102    :  1   Median :12   Median : 1.350
 103    :  1   Mean   :12   Mean   : 2.289
 104    :  1   3rd Qu.:12   3rd Qu.: 3.080
 105    :  1   Max.   :12   Max.   :14.780
 (Other):183
 ageFact
 8 :  0
 10:  0
```

Parin Patel
IST 772
HW10
12:189

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

Error: subject
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 176   1931   10.97

Error: Within
          Df Sum Sq Mean Sq F value   Pr(>F)
ageFact     2  105.2   52.60   27.82 6.09e-12 ***
Residuals 352  665.7    1.89
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> #####Question 5:
> library("changepoint")
> data("AirPassengers")
> AirPassengers <- diff(AirPassengers) #create differenced data set.
> plot(AirPassengers,main="Plot of Differenced Data Set") #plot
>
> AP_Plot<-cpt.var(AirPassengers) #find change point in variability
> plot(AP_Plot, main="Plot of Change Point in Variability")
>
>
>
> ####Question 6:
> # retrieve the confidence value
> MeanAirPassengers <- cpt.mean(AirPassengers, class = FALSE)
> plot(MeanAirPassengers,main="Plot of Change Point of the Mean")
> MeanAirPassengers["conf.value"] #confidence value
conf.value
         1
>
>
> ###Question 8:
> library("bcp")
> bcpAirPassengers <- bcp(as.vector(AirPassengers))
> plot(bcpAirPassengers)
```