Homework 10

Kwasi Mensah

2022-12-16

Author: Kwasi Mensah Homwork_Number: 10 Output: pdf_document Attribution statement:

I did the homework by myself, with help from the book and the professor and the following sources: https://www.r-tutor.com/elementary-statistics/multiple-linear-regression/estimated-multiple-regression-equation https://www.displayr.com/variance-inflation-factors-vifs/

#R Markdown #Run these three functions to get a clean test of homework code

```
dev.off() #Clear the graph window

## null device
##           1

cat('\014') #Clear the console
```

```
rm(list = ls()) #Clear user objects from the environment
```

#R Markdown

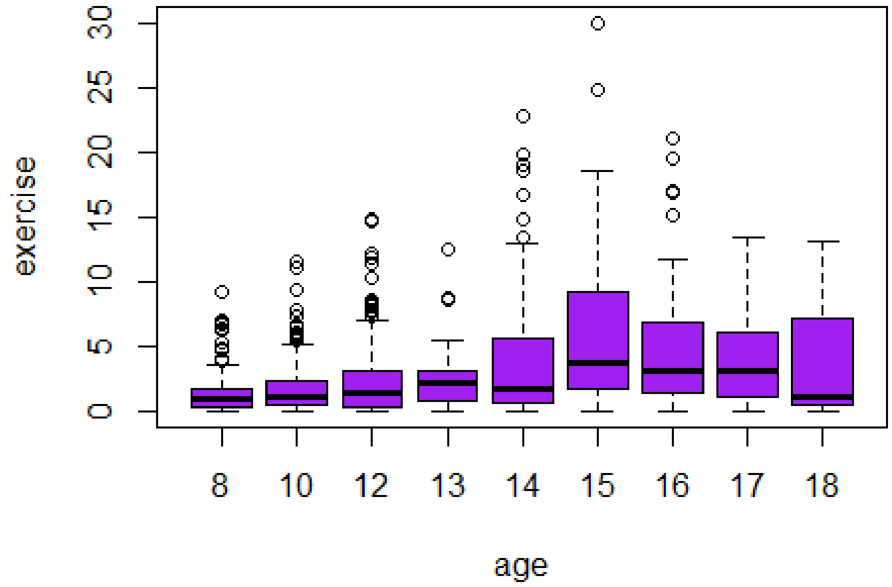
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
##

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.

```
library(car)

## Loading required package: carData
```

```
library(nlme)
library(tidyverse)

## ── Attaching packages
## ───────────────────────────────────────
## tidyverse 1.3.2 ──

## ✓ ggplot2 3.3.6      ✓ purrr   0.3.5
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.4.1
## ✓ readr   2.1.3      ✓ forcats 0.5.2
## ── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
## ✕ dplyr::collapse() masks nlme::collapse()
## ✕ dplyr::filter()   masks stats::filter()
## ✕ dplyr::lag()      masks stats::lag()
## ✕ dplyr::recode()   masks car::recode()
## ✕ purrr::some()     masks car::some()

data(Blackmore)
?Blackmore

## starting httpd help server ... done

Blackmore$age = round(Blackmore$age)
glimpse(Blackmore)

## Rows: 945
## Columns: 4
## $ subject  <fct> 100, 100, 100, 100, 100, 101, 101, 101, 101, 101, 102,
102, 1…
## $ age      <dbl> 8, 10, 12, 14, 16, 8, 10, 12, 14, 17, 8, 10, 12, 15, 8,
10, 1…
## $ exercise <dbl> 2.71, 1.94, 2.36, 1.54, 8.63, 0.14, 0.14, 0.00, 0.00,
5.08, 0…
## $ group    <fct> patient, patient, patient, patient, patient, patient,
patient…

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   :11.43   Mean   : 2.531
##  107    :  5   3rd Qu.:14.00   3rd Qu.: 3.040
##  108    :  5   Max.   :18.00   Max.   :29.960
##  (Other):915

boxplot( exercise ~ age, data=Blackmore, col= 'purple')
```



```
BM1 <- Blackmore[Blackmore$age <= 12,]
head(table(BM1$subject,BM1$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

BM1$age <- as.factor(BM1$age)


list <- rowSums(table(BM1$subject, BM1$age))==3
list <- list[list==TRUE]
list <- as.numeric(names(list))

## Warning: NAs introduced by coercion

BM1 <- BM1[BM1$subject %in% list,]
head(table(BM1$subject, BM1$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

summary(BM1$age)

##   8  10  12
## 177 177 177

summary(aov(exercise ~ age + Error(subject), data=BM1))

##
## 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)
## age         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

#     In the box plot created we can see that the values 8, 10, and 12 all
have almost identical medians. We use the "%in%" operator which lets us only
keep the rows where the value of age for that row is somewhere in the list
created. In our Repeated Anova analysis we are testing the hypothesis that
exercise does not vary over age. When we look at the Error section we see
that there is a df of 530. In the Error: within section the effect of Time is
expressed as the F-ratio or f-value is 27.82, this tests the null hypothesis
that changes are consistently 0 across all time intervals. The p-value is
6.09e-12 which is significantly lower than of threshold of .05 meaning we can
reject the null hypothesis in support of the alternative hypothesis that
exercise does vary across ages.
```

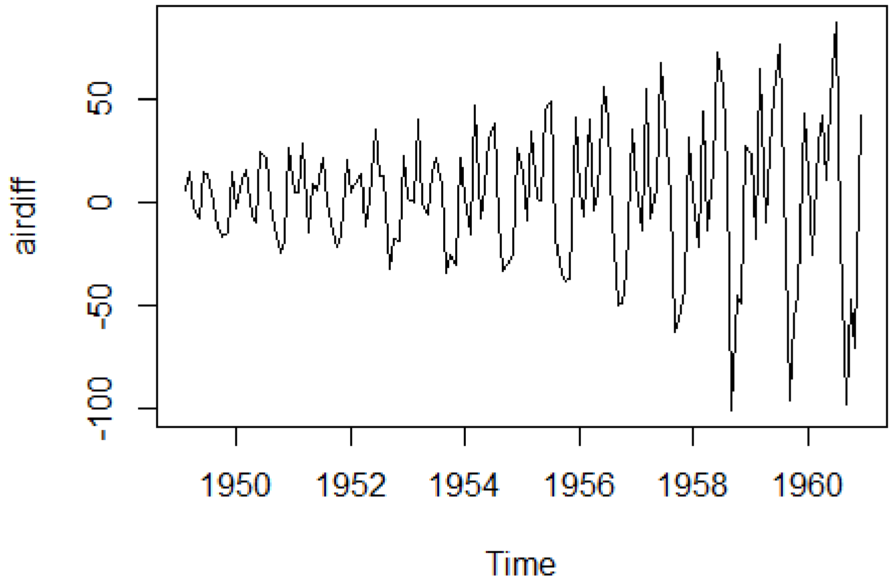
5. Given that the AirPassengers data set has a substantial growth trend, use diff() to create a differenced data set. Use plot() to examine and interpret the results of differencing. Use cpt.var() to find the change point in the variability of the differenced time series. Plot the result and describe in your own words what the change point signifies.

```
library(changepoint)

## Warning: package 'changepoint' was built under R version 4.2.2

## 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 2.2.4
##  See NEWS for details of changes.
```

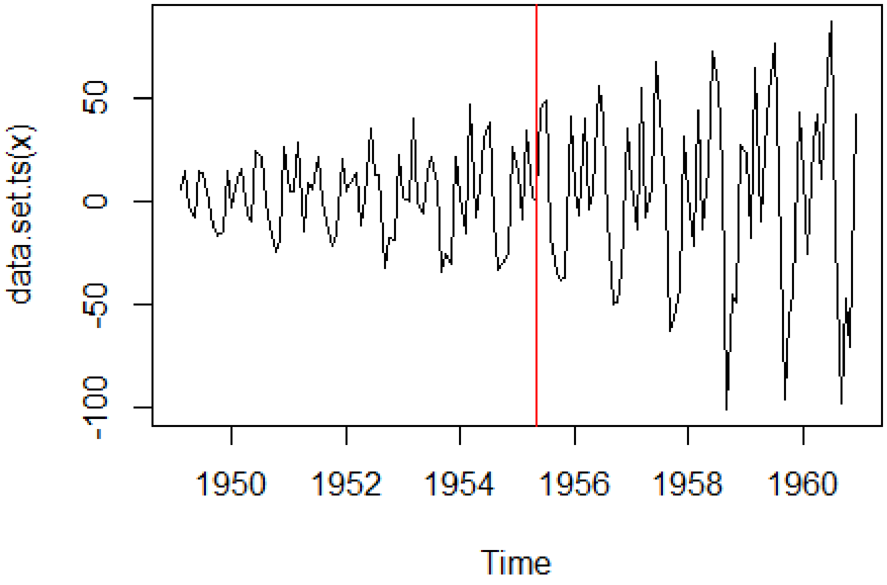
```
data(AirPassengers)
airdiff <- diff(AirPassengers)
airdiff
```

```
##         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(airdiff)
```



```
aircp <- cpt.var(airdiff)
aircp

## 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  : Thu Dec 15 03:47:50 2022
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.4
## 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(aircp)
```

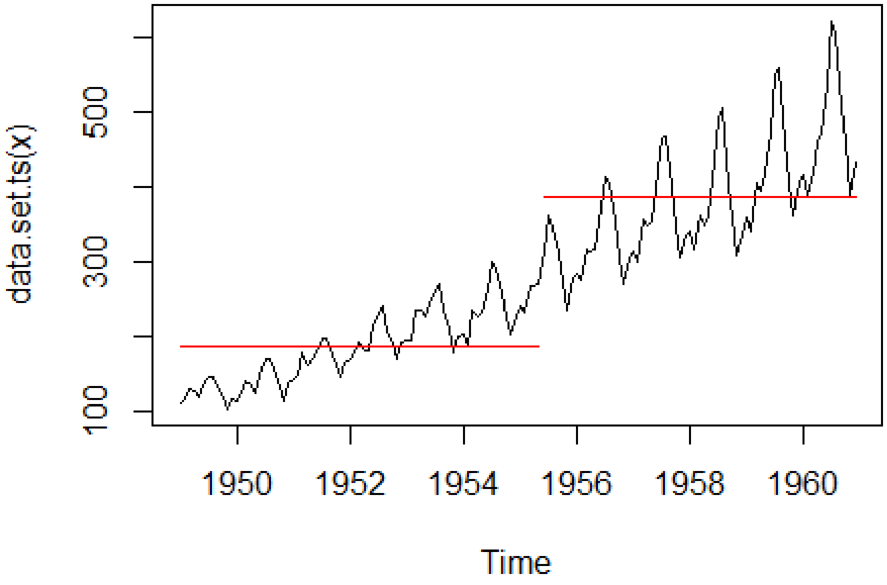


```
#     If we subtract the second element in a time series from the first
element, we will get the difference between two observations. When we plot
the differencing trends, we see these difference over the subjected time
period.  The red line in the graph signifies the change-point location. The
```

```
change-point shows us where there is a significant shift occurs in the mean
level during a certain period of time. We see that there is a large jump
between 1994 and 1996, closer to 1996 so we could assume the year is 1995.
```

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?

```
# change point in mean
library(changepoint)
aircp1=cpt.mean(AirPassengers)
aircp1

## 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  : Thu Dec 15 03:47:50 2022
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.4
## 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(aircp1)
```



```
aircp2 <- cpt.mean(AirPassengers, class=FALSE)
aircp2["conf.value"]

## conf.value
##          1

#     When looking at the graph change point of mean is represented by the
red lines. Each red lines represent the mean level of the index across the
whole period of time that is covered in the graph. We see that there is a
large jump between 1994 and 1996, closer to 1996 so we could assume the year
is 1995. The major change point occurs at 77 which is one point higher than
the change point in Exercise 5, 76. We see that out confidence value is 1
which is the strongest possible value. This signifies that our analysis has
detected a powerful change in the mean of the time series or the history of
air travel. I believe this could be due to the it being the post era to the
two World Wars.
```

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.

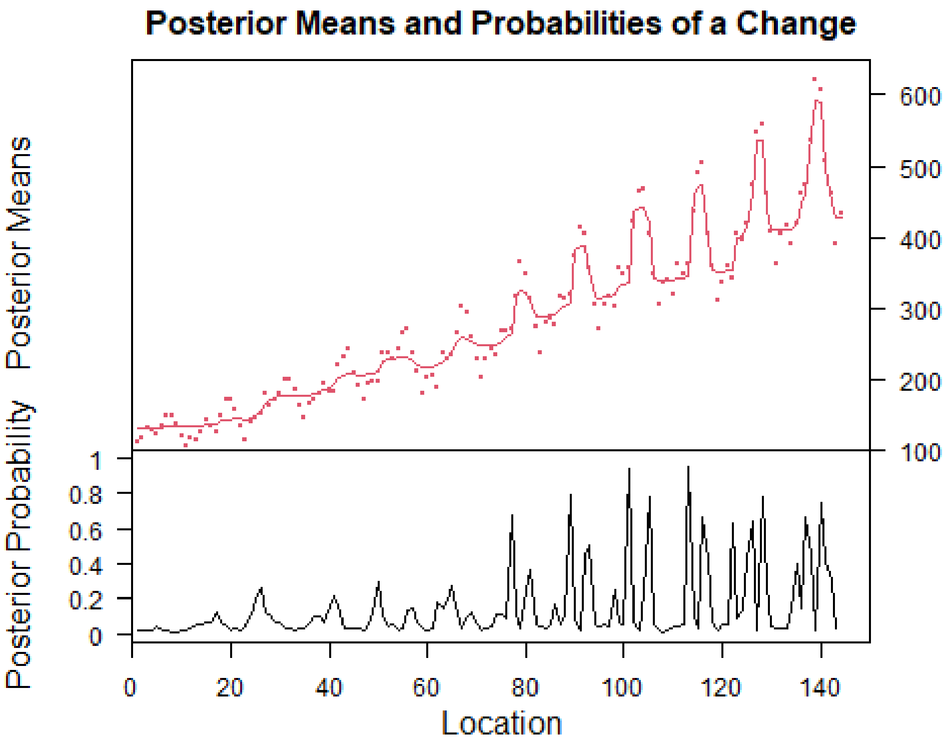
```
#SOURCE: https://metroairportnews.com/travel-by-air-the-golden-years-1920s-
1960s/

#   The Golden Age of Air Travel was between 1920s to 1960s. Aircraft evolved
from wooden material to metal aircraft especially during the two World Wars.
The Jet Age occurred between the 1950s and the 1960s. During this period
```

```
trans-Atlantic travel routes and growing trade resulted in competition
between airlines. In addition, with the introduction of the Boeing 707 for
commercial travel both internationally and domestically can help explain why
there is a big change between 1994 and 1996 in terms of the change-point in
mean and variance.
```

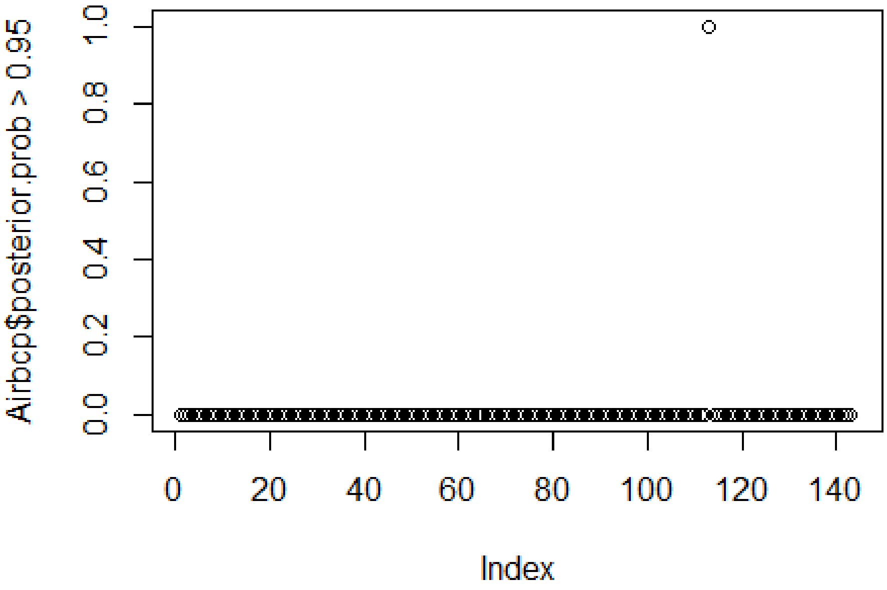
8.Use bcp() on the AirPassengers time series. Plot and interpret the results. Make sure to contrast these results with those from Exercise 6.

```
# install.packages("bcp")
library(bcp)

## Warning: package 'bcp' was built under R version 4.2.2

## Loading required package: grid

Airbcp <- bcp(as.vector(AirPassengers))
summary(Airbcp)

##
## Bayesian Change Point (bcp) summary:
##
##
## Probability of a change in mean and posterior means:
##
##     Probability    X1
## 1         0.018 130.7
## 2         0.020 131.0
## 3         0.020 131.4
## 4         0.016 131.7
## 5         0.038 131.8
## 6         0.022 132.3
## 7         0.018 132.5
## 8         0.006 132.6
## 9         0.010 132.6
## 10        0.020 132.5
## 11        0.024 132.5
## 12        0.034 132.8
## 13        0.048 133.3
## 14        0.058 134.3
## 15        0.060 135.4
## 16        0.066 136.6
## 17        0.124 138.0
## 18        0.060 141.4
## 19        0.042 142.8
## 20        0.024 143.7
## 21        0.028 143.8
## 22        0.016 143.8
## 23        0.044 143.6
## 24        0.110 145.0
## 25        0.192 148.7
```

```
plot(Airbcp)
```



```
plot(Airbcp$posterior.prob > .95)
```



```
#     Our Bayesian change-point analysis summary and graph shows, we can see
that there is a large jump, in mean and probability at the 77th/144 location.
When compared to Exercise 6 we see that we get the same result for were
change-point of the mean, or a large shift occurs at the point of 77. Thus,
we can assume that this is definitely where the change-point of the mean
occurs, and a real shift in air travel occurred at this point in time.
```