Quinn Knudsen

Homework 10

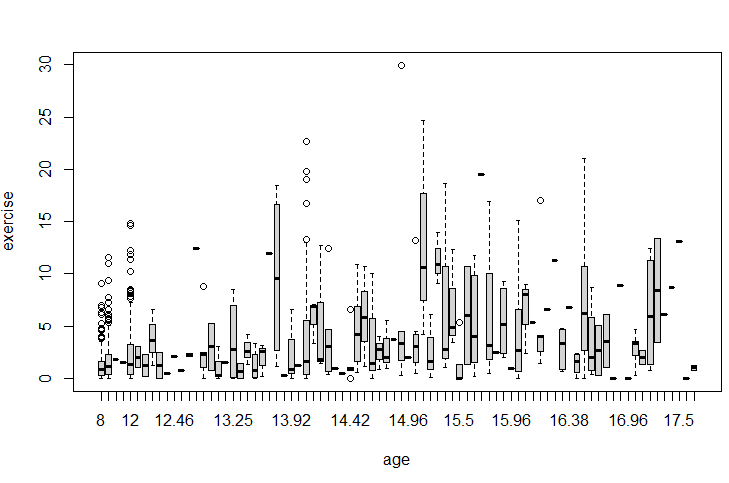
Topic: **Time Series Analysis**

2)

> blm <-data ("Blackmore")

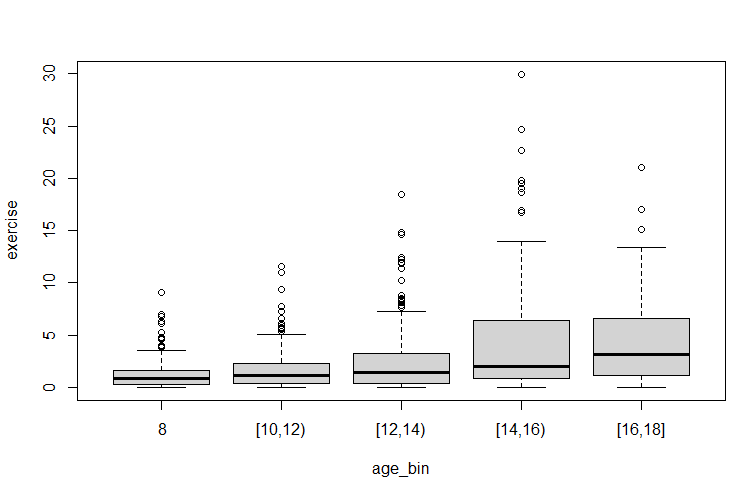
> blm <- Blackmore

> boxplot(exercise~age, data=blm)



> blm$age\_bin <- cut2(blm$age,c(8,10,12,14,16,18))

> boxplot(exercise~age\_bin, data=blm)



blm <- Blackmore[Blackmore$age <=12,]

table(blm$subject,blm$age)

summary(blm)

list <- rowSums(table(myData$subject,myData$age))==3

list <- list[list==TRUE]

list <- as.numeric(names(list))

blm2 <- blm[blm$subject%in% list,]

blm2$age <- as.factor(blm2$age)

> summary(aov(exercise~age,data=blm2))

Df Sum Sq Mean Sq F value Pr(>F)

age 2 102.7 51.35 10.09 5.06e-05 \*\*\*

Residuals 495 2518.4 5.09

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

out<-aov(exercise~age,data=blm2)

library(multcomp)

> TukeyHSD(out, which = "age")

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = exercise ~ age, data = blm2)

$age

diff lwr upr p adj

10-8 0.5397590 -0.042259608 1.121778 0.0756015

12-8 1.1121687 0.530150030 1.694187 0.0000261

12-10 0.5724096 -0.009609006 1.154428 0.0550704

> summary(glht(out, linfct = mcp(age = "Tukey")))

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = exercise ~ age, data = blm2)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

10 - 8 == 0 0.5398 0.2476 2.180 0.0756 .

12 - 8 == 0 1.1122 0.2476 4.492 <1e-04 \*\*\*

12 - 10 == 0 0.5724 0.2476 2.312 0.0551 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- single-step method)

library(ez)

> ezANOVA(data=blm2,dv=.(exercise),within=.(age),wid=(subject),detailed=TRUE)

Warning: You have removed one or more Ss from the analysis. Refactoring "subject" for ANOVA.

$ANOVA

Effect DFn DFd SSn SSd F p p<.05 ges

1 (Intercept) 1 165 1688.7531 1891.8688 147.28519 1.243957e-24 \* 0.40139644

2 age 2 330 102.6938 626.5731 27.04309 1.330658e-11 \* 0.03917912

$`Mauchly's Test for Sphericity`

Effect W p p<.05

2 age 0.7490297 5.115871e-11 \*

$`Sphericity Corrections`

Effect GGe p[GG] p[GG]<.05 HFe p[HF] p[HF]<.05

2 age 0.7993795 9.573171e-10 \* 0.805973 8.316244e-10 \*

The first boxplot as described looked a bit chaotic, so I used the approach of binning the variables to make them a little more meaningful. It appears that there are some outliers on the high end (potentially log-transformation?). Continuing onward, the dataset was subsetted in age less than or equal to 12. Some data prep was then used to take ‘complete’ observations with 3 time measurements to ensure balance. An ANOVA yielded significance for age differences in observed exercise. 12-year-old participants exercise more than 8 year-old participants with near significant values between 12 and 10 and 10 and 8. The general pattern appears linear in what exercise is higher with higher age groups. We see here that we are violating the assumption of homogeneity of variance by the significant p value for Mauchly’s Test for Sphericity, but are spared from shame by the significant GGe and HFe allowing us to confidently reject the null.

5)

> air <- AirPassengers

> airdiff<-diff(air)

> plot(airdiff)

> library(changepoint)

> aircp<-cpt.var(airdiff)

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

> 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 : Mon Feb 08 03:15:38 2021

summary(.) :

----------

Created Using changepoint version 2.2.2

Changepoint type : Change in variance

Method of analysis : AMOC

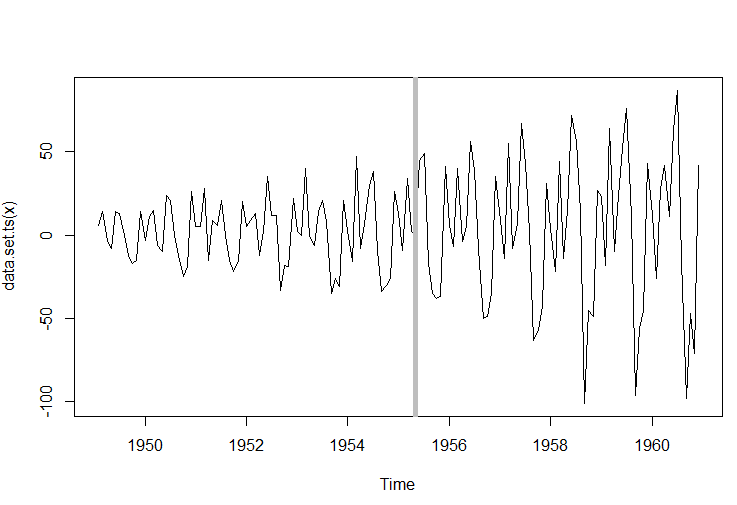
Test Statistic : Normal

Type of penalty : MBIC with value, 14.88853

Minimum Segment Length : 2

Maximum no. of cpts : 1

Changepoint Locations : 76



As indicated by the gray line on the chart above, this signifies a change in variance in the middle of 1955 (changepoint 76) for the difference data. After this changepoint, the variability significantly increased. In this instance, the changepoint is easily observed by the eye.

6)

> aircp<-cpt.mean(airdiff)

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

> 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 : Mon Feb 08 03:15:38 2021

summary(.) :

----------

Created Using changepoint version 2.2.2

Changepoint type : Change in mean

Method of analysis : AMOC

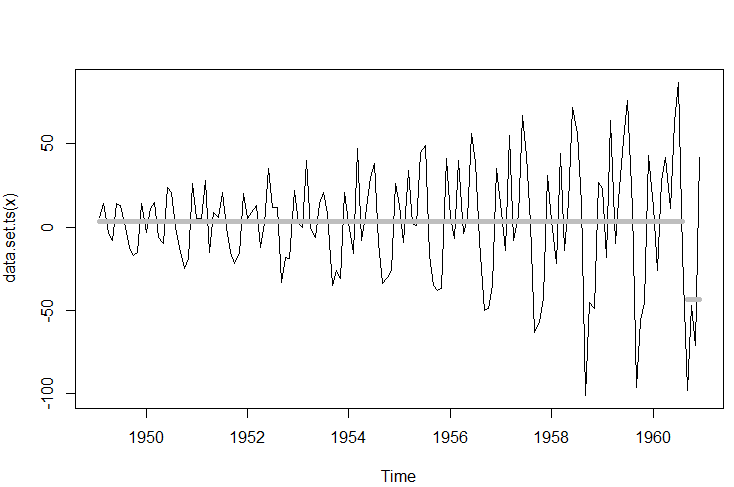
Test Statistic : Normal

Type of penalty : MBIC with value, 14.88853

Minimum Segment Length : 1

Maximum no. of cpts : 1

Changepoint Locations : 139



This indicates mean changes exemplified in the 1960s and cyclical/seasonal effect. Months September through November tend to be negative values.

7)

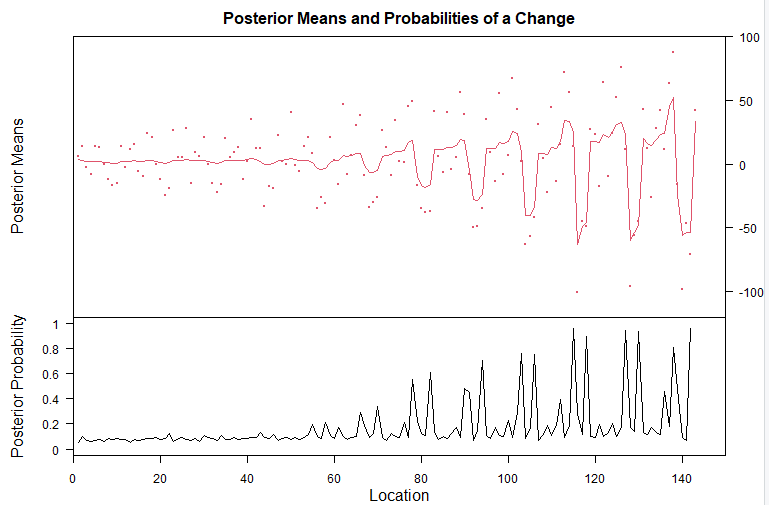
The 1950s were known as the Golden Age for flight travel filled with boozy and smoke filled cabins. As time went on flights became more popular (and eventually cheaper, but not for quite some time). In 1958, the first commercial flight of the Boeing 707 was launched. Coming off the war plagued 1940s the ability and desire for international travel may have gradually increased with time.

8)

> library(bcp)

> bcpair<-bcp(as.vector(airdiff))

> plot(bcpair)



The results here demonstrate more pronounced/noteworthy changepoints based upon the posterior probability. These changepoints occur beyond just location 139 as shown in question 6 or 79 as shown in question 5 and visually depict the changes and variability in a much more optically intuitive manner.