Chapter 8 Autocorrelation Assignment

Petra Guy

October 23, 2017

1 Introduction - about the code

I have three scripts (so far!). Initially I wrote a python-ish functions script, but I couldnt work out how to run this from Rscript. I therefore rewrote this as a long ugly code, thinking that if I ran that from Rscript it would simple run through line by line. But the load command would not work. I also checked the results of my calculations against the R acf() function, and they were the same, so I wrote another script using that and dplyr pipes.

```
#!/usr/bin/env Rscript
#Chapter 8.8 Final Practial


prepare_workspace = function(){
  rm(list = ls())
  library(stats) # for plot.ts
  library(ggplot2)
  library(pracma) # for movavg
  graphics.off()
}

# get the data and make two vectors coz easier for loops

GetData = function(){
  load("../Data/KeyWestAnnualMeanTemperature.RData")
  Years = ats[[1]]
  Temps = ats[[2]]
  MyData = as.data.frame(cbind(Years,Temps))
  return(MyData)
}

Examine = function(Data){
  # plot simple time series
  plot.ts(Data)
  #plot terms with lag of 1 to 4 years against each other M Crawley p787
  par(mfrow = c(2,2))
  sapply(1:4, function(x) plot(Data[-c(100:(100-x+1))], Data[-c(1:x)]))
}

#autocorrelation coef is Sum(Y[i+1]-AveY)(Y[i] -AveY)/sum(sqr(Y[i]-AveY))


# calculate sum[(Yi+1 - ave)(Yi - ave)] -- numerator of autocorr coef
Calc_numerator = function(avector){
num = vector("numeric",99)
for (i in seq_along((avector))) {
  if (i <100) {
  num[i] = as.vector( avector[i+1] - mean(avector) ) * ( avector[i] -
mean(avector) )
  }
 else
 totalnum = sum(num)
}
return(totalnum)}


Calc_denom = function(avector){
#  calcuate sum[(Yi - ave)^2] -- denom of auto corr coef
denom = vector("numeric",99)
for (i in seq_along((avector))) {
  if (i <100) {
    denom[i] = as.vector(( avector[i] - mean(avector) )^2)
  }
  else
    totaldenom = sum(denom)
}
return(totaldenom)}


Calc_acf = function(avector){
  num = Calc_numerator(avector)
  denom = Calc_denom(avector)
  autocorrcoef = num/denom
```

Figure 1: Code written in functions

```
#!/usr/bin/env Rscript
#Chapter 8.8 Final Practical

rm(list = ls())
library(stats) # for plot.ts
library(ggplot2)
library(pracma) # for movavg

# Collect the data
MyData = as.data.frame(read.csv("../Data/KeyWest.CSV"))


#load("../Data/KeyWestAnnualMeanTemperature.RData")
#Years = ats[[1]]
Temps = MyData[[3]]
#MyData = as.data.frame(cbind(Years,Temps))


#Examine = function(Data){
  # plot simple time seriesgraphics.off()

pdf("../Results/TAutocorrtimeseries1.pdf")

ggplot(MyData, aes(x = Year, y = Temp))+
 labs(title = "Time series of temperature data for KeyWest 1900 - 2000", x = "date", y =
"Temperature")+
  geom_line()

dev.off()

pdf("../Results/TAutocorrtimeseries2.pdf")
par(mfrow = c(2,2))
sapply(1:4, function(x) plot(MyData$Temp[-c(100:(100-x+1))], MyData$Temp[-c(1:x)], ylab =
"Year", xlab = "year"))
dev.off()

#autocorrelation coef is Sum(Y[i+1]-AveY)(Y[i] -AveY)/sum(sqr(Y[i]-AveY))


# calculate sum[(Yi+1 - ave)(Yi - ave)] -- numerator of autocorr coef
#Calc_numerator = function(avector){
num = vector("numeric",99)
for (i in seq_along((Temps))) {
 if (i <100) {
  num[i] = as.vector( Temps[i+1] - mean(Temps) ) * ( Temps[i] - mean(Temps) )
  }
 else
 totalnum = sum(num)}



#Calc_denom = function(avector){
```

Figure 2: Long ugly script

```
#!/usr/bin/env Rscript
#Chapter 8.8 Final Practial


prepare_workspace = function(){
  rm(list = ls())
   graphics.off()
   library(dplyr)
}

# get the data and make two vectors coz easier for loops


  load("../Data/KeyWestAnnualMeanTemperature.RData")
  Years = ats[[1]]
  Temps = ats[[2]]
  MyData = as.data.frame(cbind(Years,Temps))


  # plot simple time series

  autocorr = acf(Temps,1)[[1]][2]

  print("autoccorrelation coefficient for lag of 1 is ")
  print(autocorr)

  # repeating with the 1000 samples in a for loop
  acfs = vector("numeric",1000)
  for (i in 1:1000){
    acfs[i] = acf(sample(Temps,100))[[1]][2]
  }


  # reapeting using piping - but since acf gives a list, the answer is messh=y,
hnce unlist and select out alternate values
  acfs2 =vector("numeric",1000)
  for (i in 1:1000){
    acfs2[i] <- Temps %>% sample(.,100) %>% acf(.,1)
  }
acfs2 = unlist(acfs2[c(FALSE,TRUE)])
```

Figure 3: Using R acf() and dplyr

2 Graphs

The time series does not immediately show a clear trend. The next four scatter plots are for years plotted against years with lags of 1 to 4 which appear to show a weak upward trend. The histogram show the distribution of autocorrelation coefficients from a random sample of mean temperatures. The values are all below my calcualted value of 0.309, implying the data are correlated (p = 0). Since they're apppears to be a correlation I calculated 2 point moving averages and plotted these with a linear fit which shows an increase in mean annual temperatures.

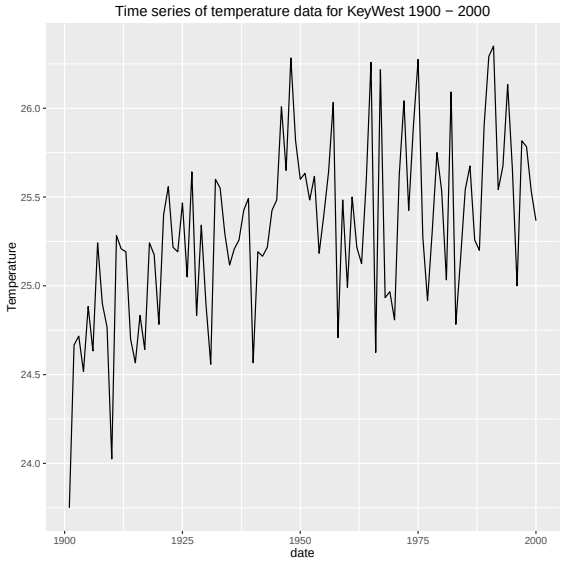

Figure 4: Time series plot.

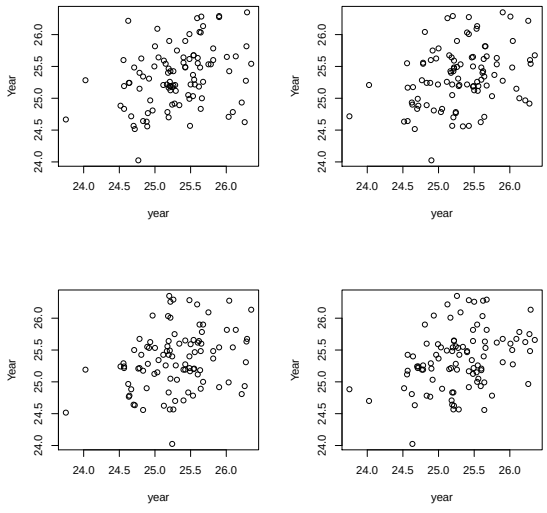

Figure 5: Scatter plots of lag 1 to 4.

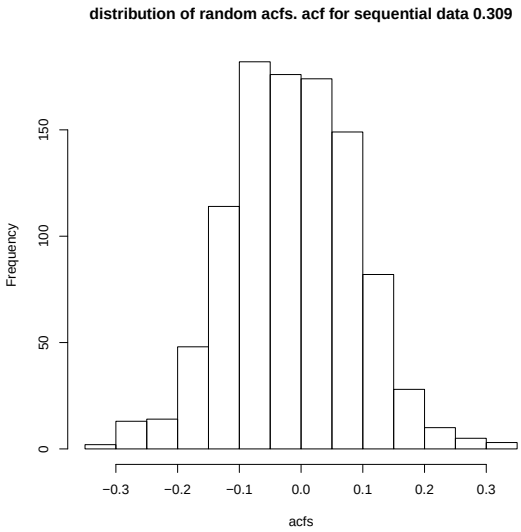

Figure 6: Frequency distribution for randomly sampled acf

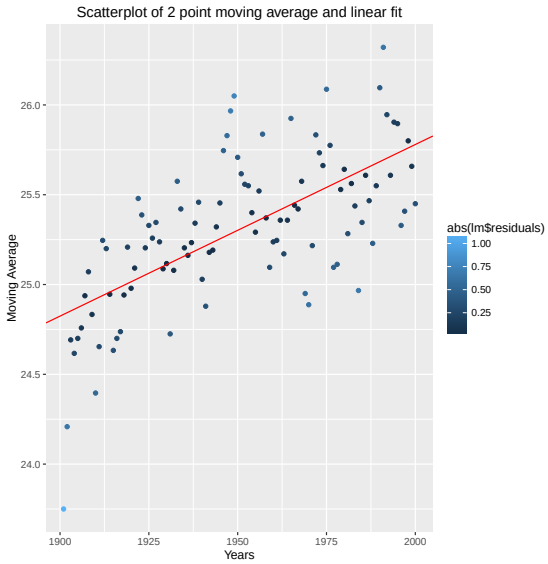

Figure 7: Moving Averages and straight line fit.