Homework 10

Luke Daniels

3/28/2018

```
FileBuilder <- function(fileN=10,
                        fileFolder = "RandomFiles/",
                        fileSize = c(15,100),
                        fileNA = 3){
  for(i in seq_len(fileN)) {
    fileLength <- sample(fileSize[1]:fileSize[2], size=1)
    varX <- runif(fileLength) #random x values
    varY <- runif(fileLength) #random y values
    dF <- data.frame(varX, varY) #bind to data frame
    badvals <- rpois(n=1, lambda = fileNA) # number of NA values
    dF[sample(nrow(dF), size = badvals),1] <- NA # NA is X
    dF[sample(nrow(dF), size = badvals),2] <- NA # NA in Y

    # create a file name for this data frame
    fileLabel <- paste(fileFolder,
                       "ranFile",
                       formatC(i,
                               width = 3,
                               format="d",
                               flag = "0"),
                       ".csv", sep = "")
    # Set up data file and incorporate time stamp
    # and minimal metadata

    write.table(cat("# Simulated random data file",  #writing the metadata
                    "for batch processing", "\n",
                    "# timesstamp: ",as.character(Sys.time()), "\n",
                    "# NJG", "\n",
                    "#------------", "\n",
                    "\n",
                    file = fileLabel,
                    row.names = "",
                    col.names = "",
                    sep = ""))

    #### Files are now built
    #### Now we need to add data frame now

    write.table(x = dF,
                file = fileLabel,
                sep = ",",
                row.names = FALSE,
                append = TRUE)

  } # Close the for loop
} #close the function
```

```
################################################

# FUNCTION: regStats
# fit linear model, get regression stats
# input: 2-column data frame
# output: slope, p-value, and r2
#
# ------------------------------------------
regStats <- function(d=NULL) {
  if(is.null(d)) {
    xVar <- runif(10)
    yVar <- runif(10)
    d <- data.frame(xVar, yVar)
  }
  . <- lm(data = d, d[,2]~d[,1]) # . as temporary storage
  . <- summary(.)
  statsList <- list(Slope = .$coefficients[2,1],
                    pVal = .$coefficients[2,4],
                    r2 = .$r.squared)
  return(statsList)
} # close function

#-------------------------------------------
```

Body for Batch Proccessing

```
# Start of body of program
library(TeachingDemos)
char2seed("Freezing March")

#-------------------------------------------

#Global Variables
fileFolder <- "RandomFiles/"
nFiles <- 100
fileOut <- "StatsSummary.csv"

# Create 100 random data
FileBuilder(fileN = nFiles)
```

```
## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""
```

```
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
```

```
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
```

```
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
```

```
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
```

```
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
```

```
## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""

## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file

## ""
```

```
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
## ""
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
```

```
## ""
```

```
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
```

```
## ""
```

```
## Warning in write.table(x = dF, file = fileLabel, sep = ",", row.names =
## FALSE, : appending column names to file
```

```
# Create data frame to hold file summary statistics

fileNames <- list.files(path = fileFolder)
ID <- seq_along(fileNames)
fileName <- fileNames
slope <- rep(NA, nFiles) # Vector to hold output
pVal <- rep(NA, nFiles)  # Vector to hold output
r2 <- rep(NA, nFiles)    # Vector to hold output

statsOut <- data.frame(ID, fileName, slope, pVal, r2) #Organize Vectors

# batch process by looping through individual files

for (i in seq_along(fileNames)){
  data <- read.table(file = paste(fileFolder, fileNames[i], sep=""),
                     sep=",",
                     header = TRUE) #read in next data file
  dClean <- data[complete.cases(data),] #get clean cases

  . <- regStats(dClean) # pull regression stats from clean file
  statsOut[i, 3:5] <- unlist(.) #unlust, copy into last 3 columns
}

# set up output file and incorporate time stamp and minimal metadata

write.table(cat("# Summary stats for ",
                "batch processing of regression models", "\n",
                "# timestamp: ",as.character(Sys.time()), "n",
                "# LD", "\n",
                "# ---------------", "\n",
                "\n",
                file = fileOut,
                row.names = "",
                col.names = "",
                sep = ""))
```

```
## ""
```

```
# now add the data frame

write.table(x=statsOut,
            file = fileOut,
            row.names = FALSE,
            col.names = TRUE,
            sep = ",",
            append = TRUE)
```

```
## Warning in write.table(x = statsOut, file = fileOut, row.names = FALSE, :
## appending column names to file
```

Question 4: "Breaking the Program"

To break the code, I adjusted the value of lambda (NA). If we do not change the number of rows, it is impossible to assign 16 NA values to the matrix. R cannot assign a sample that is larger than the population. Altering lambda does not affect the files created. However, it makes so that a Stats Summary is not produced. If the value of lamda is inbetween the range of number of rows then the model will break. If lambda is greater than the minimum value of rows, then eventually in the for loop, more NA's will be assigned than the minimum number of rows.

Data that cannot be fit with a linear model

The data cannot be fit with a linear model when lambda is change to 10 and the minimum number of rows is 15. All of the files are created, however, the summary statistics are not.

```
#Produces the following error:
#Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
  #0 (non-NA) cases
```

The error seems to indicate that there are zero non NA cases. It could be because some files had data sets completely filled with NA values and therefore a regression could not be performed.

5.)Question 5 Add two columns to Statsmmary.csv

To add new columns we modified the code in the following way.

```
for (i in seq_along(fileNames)){
  data <- read.table(file = paste(fileFolder, fileNames[i], sep=""),
                     sep=",",
                     header = TRUE) #read in next data file
  dClean <- data[complete.cases(data),] #get clean cases

  . <- regStats(dClean) # pull regression stats from clean file
  statsOut[i, 3:5] <- unlist(.) #unlust, copy into last 3 columns
  statsOut[i,6] <- nrow(data)
  statsOut[i,7] <- nrow(dClean)
}

write.table(cat("# Summary stats for ",
                "batch processing of regression models", "\n",
                "# timestamp: ",as.character(Sys.time()), "n",
                "# LD", "\n",
                "# --------------", "\n",
                "\n",
                file = fileOut,
                row.names = "",
                col.names = "",
                sep = ""))
```

```
## ""
```

```
# Adding the data frame

write.table(x=statsOut,
            file = fileOut,
            row.names = FALSE,
            col.names = TRUE,
```

```
            sep = ",",
            append = TRUE)
```

```
## Warning in write.table(x = statsOut, file = fileOut, row.names = FALSE, :
## appending column names to file
```

```
head(statsOut)
```

```
##   ID       fileName       slope      pVal           r2 V6 V7
## 1  1 ranFile001.csv -0.01712517 0.8529012 0.0003756598 98 94
## 2  2 ranFile002.csv  0.07113830 0.6304443 0.0053063845 50 46
## 3  3 ranFile003.csv -0.06279787 0.6088401 0.0047931624 65 57
## 4  4 ranFile004.csv  0.20033061 0.6277713 0.0272295482 17 11
## 5  5 ranFile005.csv -0.11200916 0.2567368 0.0156581241 90 84
## 6  6 ranFile006.csv -0.18345108 0.2774899 0.0405274549 37 31
```

As excepted V7 (the cleaned rows have lower values!