

## STAT 231 Assignment 1

The purpose of this assignment is to have you run R code and produce the numerical and graphical summaries discussed in Chapter 1 of the Course Notes for randomly generated data.

Follow the steps in *the Introduction to R and RStudio* posted on Learn to install the software needed for this course (see Section 1 - Introduction). To learn how to run R code see Section 2 – Getting Started. The moments package and the MASS package can be installed using RStudio or by the commands given in the code below. (See Section 4 – Summary Statistics.)

**The code for this assignment is posted both as a text file called RCodeAssignment1.txt and an R file called RCodeAssignment1R.R which are posted in the Assignment 1 folder in the Assignments folder under Content on Learn.**

**Problem 1:** Run the following R code.

```
#####  
# Run this code only once  
install.packages("moments")  
library(moments)  
library(MASS) # truehist is in the library MASS  
#####  
  
#####  
# Problem 1: R code for Gaussian data  
id<-20456458  
mu<-id-10*trunc(id/10) # mu = last digit of ID  
sig<-max(1,trunc(id/10)-10*trunc(id/100)) # sig = second last digit of ID unless last digit is zero  
cat("mu = ", mu, ", sigma = ", sig) # display values of mu and sigma  
set.seed(id)  
yn<-sort(round(rnorm(200,mu,sig),digits=2)) # 200 observations from G(mu,sig)  
yn[1:5] # display first 5 numbers in the data set  
# display sample mean and standard deviation  
cat("sample mean = ", mean(yn), ", sample standard deviation = ", sd(yn))  
cat("five number summary: ",fivenum(yn)) # five number summary  
cat("sample skewness = ", skewness(yn)) # sample skewness  
cat("sample kurtosis = ", kurtosis(yn)) # sample kurtosis  
# plot relative frequency histogram and superimpose Gaussian pdf  
truehist(yn,main="Relative Frequency Histogram of Data")  
curve(dnorm(x,mean(yn),sd(yn)),col="red",add=TRUE,lwd=2)  
# plot Empirical and Gaussian cdf's  
plot(ecdf(yn),verticals=T,do.points=F,xlab="y",ylab="ecdf",main="")
```

```

title(main="Empirical and Gaussian C.D.F.'s")
curve(pnorm(x,mean(yn),sd(yn)),add=TRUE,col="red",lwd=2) # superimpose Gaussian cdf
#####

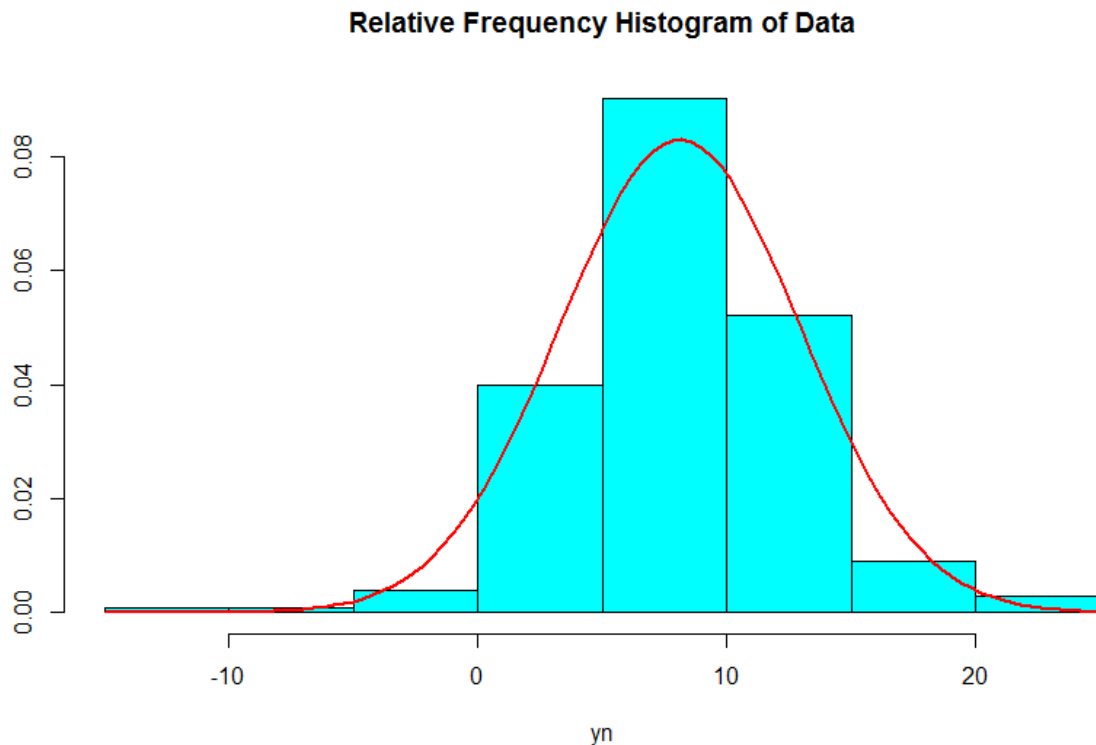
```

Verify that you obtain the following output and plots:

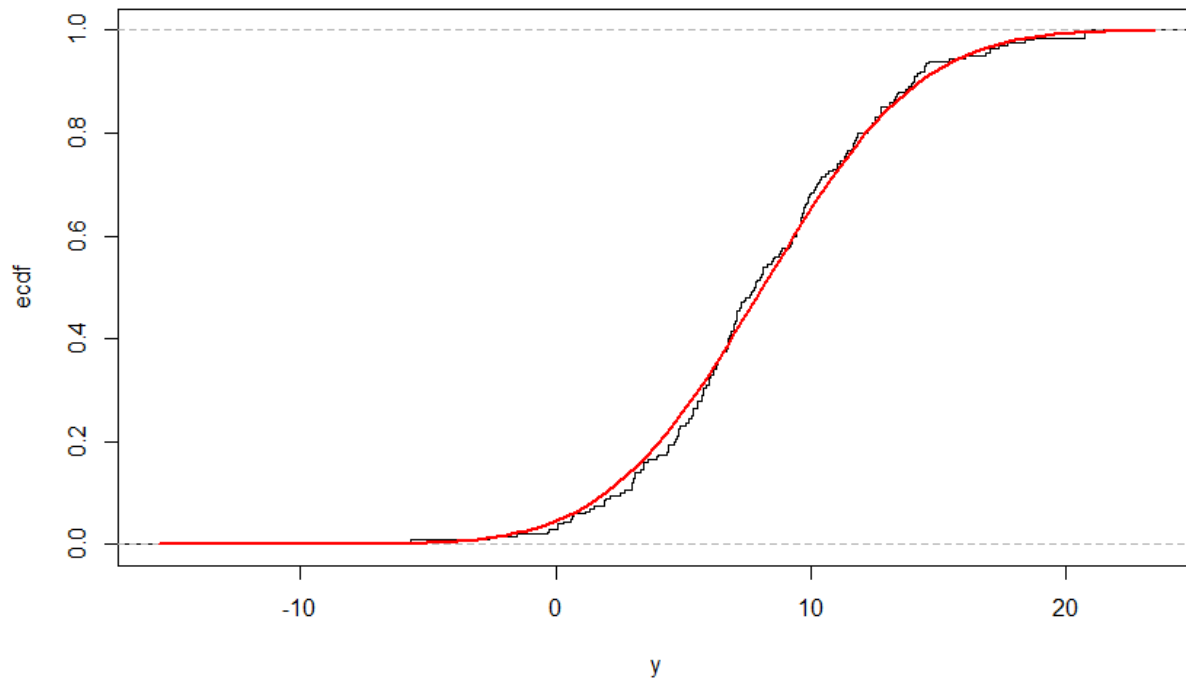
```

> yn[1:5] # display first 5 numbers in the data set
[1] -12.89 -5.67 -2.60 -1.54 -0.31
> # display sample mean and standard deviation
> cat("sample mean = ", mean(yn), ", sample standard deviation = ", sd(yn))
sample mean = 8.11465 , sample standard deviation = 4.812293
> cat("five number summary: ", fivenum(yn)) # five number summary
five number summary: -12.89 5.36 7.815 11.32 20.77
> cat("sample skewness = ", skewness(yn)) # sample skewness
sample skewness = -0.2029152
> cat("sample kurtosis = ", kurtosis(yn)) # sample kurtosis
sample kurtosis = 4.486426

```



## Empirical and Gaussian C.D.F.'s



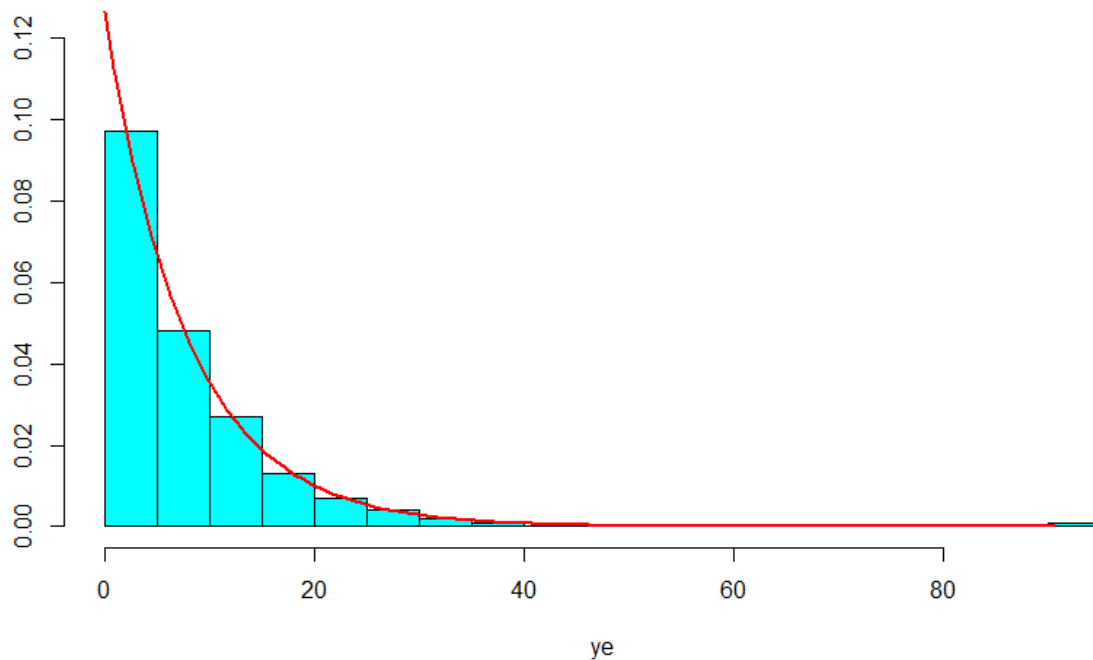
**Problem 2:** Run the following R code.

```
#####
# Problem 2: R code for Exponential data
set.seed(id)
mu<-max(1,id-10*trunc(id/10))          # mu = last digit of ID unless it is zero
ye<-sort(round(rexp(200,1/mu),digits=2)) # 200 observations from Exponential(1/mu)
ye[1:5]          # display first 5 numbers in the data set
# display sample mean and standard deviation
cat("sample mean = ", mean(ye), ", sample standard deviation = ", sd(ye))
cat("five number summary: ",fivenum(ye))      # five number summary
cat("sample skewness = ", skewness(ye))      # sample skewness
cat("sample kurtosis = ", kurtosis(ye))      # sample kurtosis
# plot relative frequency histogram and superimpose Exponential pdf
truehist(ye,ymax=1/mean(ye),main="Relative Frequency Histogram of Data")
curve(dexp(x,1/mean(ye)),from=0.001,to=max(ye),col="red",add=TRUE,lwd=2)
# plot Empirical and Exponential cdf's
plot(ecdf(ye),verticals=T,do.points=F,xlab="y",ylab="ecdf",main="")
title(main="Empirical and Exponential C.D.F.'s")
curve(pexp(x,1/mean(ye)),col="red",add=TRUE,lwd=2)
#Plot side by side boxplots
boxplot(yn,ye,col="cyan",names=c("Gaussian Data","Exponential Data"))
#####
```

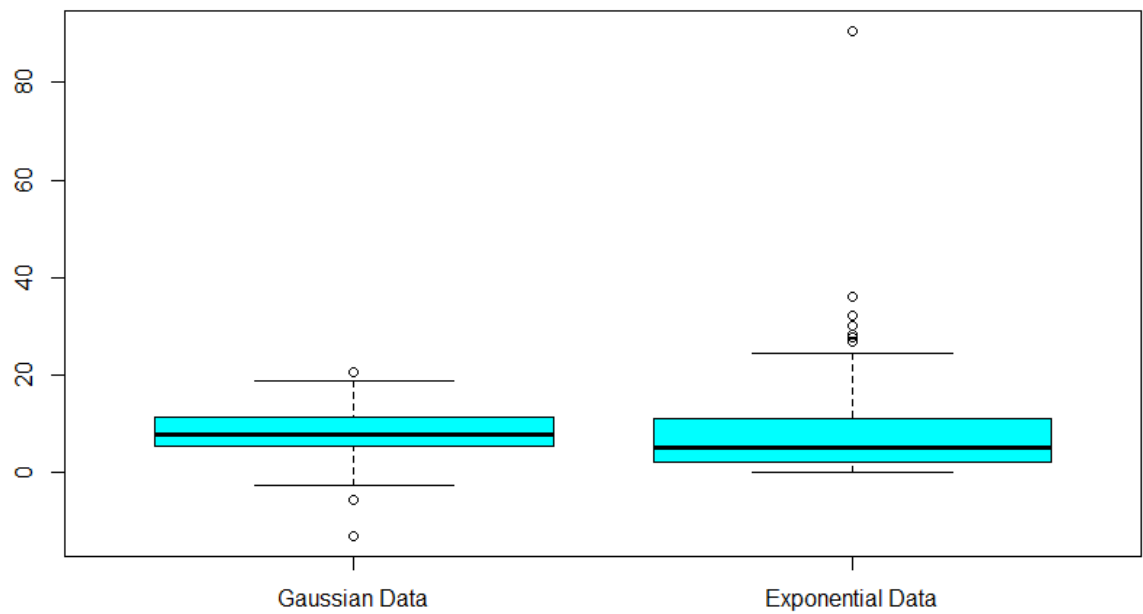
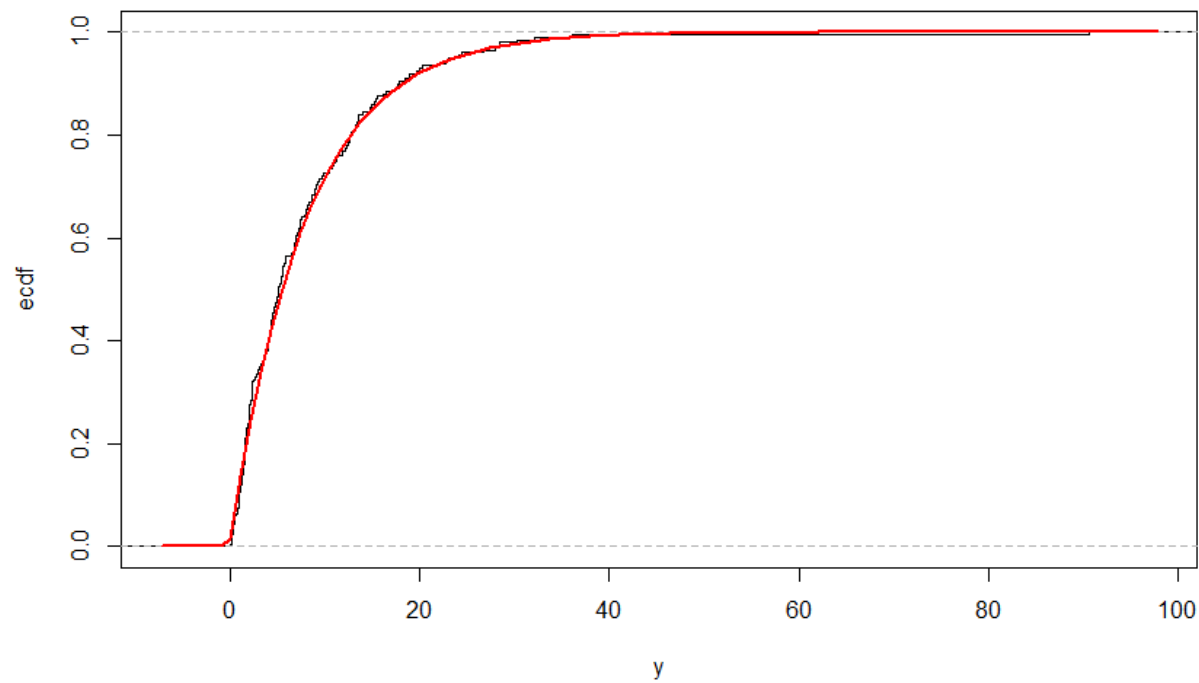
Verify that you obtain the following output and plots.

```
> ye[1:5]                                # display first 5 numbers in the data set
[1] 0.01 0.13 0.18 0.24 0.26
> # display sample mean and standard deviation
> cat("sample mean = ", mean(ye), ", sample standard deviation = ", sd(ye))
sample mean = 7.9169 , sample standard deviation = 9.249768
> cat("five number summary: ", fivenum(ye))          # five number summary
five number summary: 0.01 2.07 5.095 11.12 90.52
> cat("sample skewness = ", skewness(ye))            # sample skewness
sample skewness = 4.198336
> cat("sample kurtosis = ", kurtosis(ye))            # sample kurtosis
sample kurtosis = 33.82573
```

Relative Frequency Histogram of Data



Empirical and Exponential C.D.F.'s



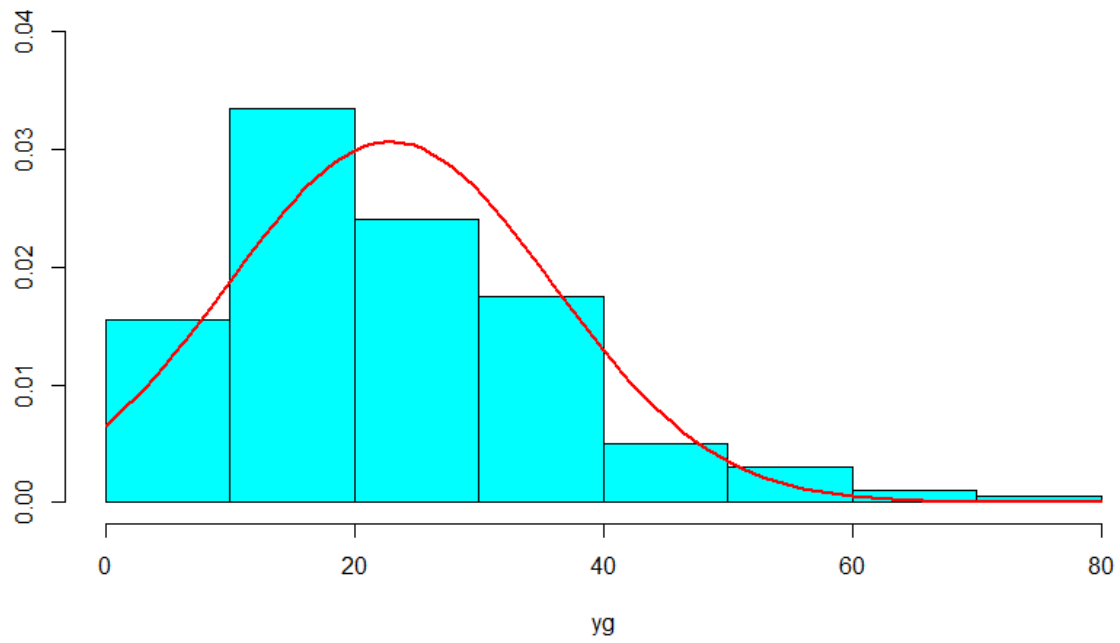
**Problem 3:** Run the following R code.

```
#####  
# Problem 3: R code for Gamma data  
set.seed(id)  
yg<-sort(round(rgamma(200,3,1/mu),digits=2)) # 200 observations from Gamma(3,1/mu)  
yg[1:5]          # display first 5 numbers in the data set  
cat("sample mean = ", mean(yg), ", sample standard deviation = ", sd(yg))  
cat("five number summary: ", fivenum(yg))      # five number summary  
cat("sample skewness = ", skewness(yg))        # sample skewness  
cat("sample kurtosis = ", kurtosis(yg))        # sample kurtosis  
# plot relative frequency histogram and superimpose Gaussian pdf  
truehist(yg,ymax=1/mean(yg),main="Relative Frequency Histogram of Data")  
curve(dnorm(x,mean(yg),sd(yg)),col="red",add=TRUE,lwd=2)  
# plot Empirical and Gaussian cdf's  
plot(ecdf(yg),verticals=T,do.points=F,xlab="y",ylab="ecdf",main="")  
title(main="Empirical and Gaussian C.D.F.'s")  
curve(pnorm(x,mean(yg),sd(yg)),add=TRUE,col="red",lwd=2) # superimpose Gaussian cdf  
#####
```

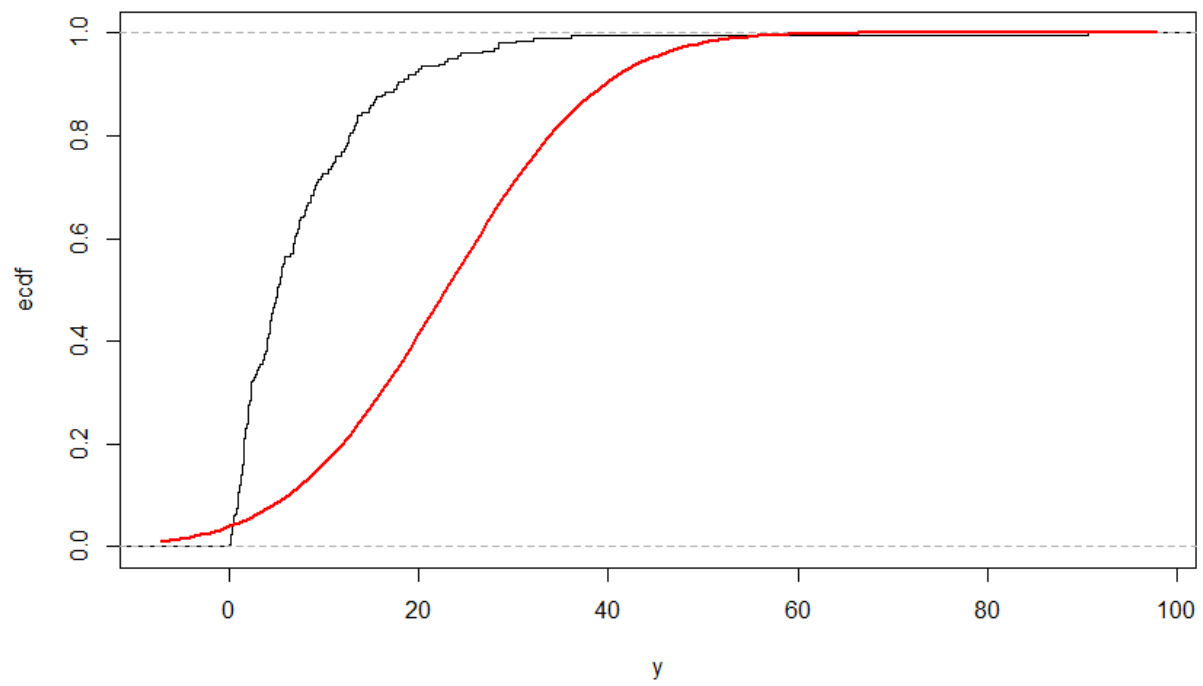
Verify that you obtain the following output and plots:

```
> yg[1:5]          # display first 5 numbers in the data set  
[1] 1.32 2.27 3.62 4.18 4.66  
> cat("sample mean = ", mean(yg), ", sample standard deviation = ", sd(yg))  
sample mean = 22.89415 , sample standard deviation = 4.812293  
> cat("five number summary: ", fivenum(yg))      # five number summary  
five number summary: 1.32 13.515 20.775 30.555 74.39  
> cat("sample skewness = ", skewness(yg))        # sample skewness  
sample skewness = 0.9927479  
> cat("sample kurtosis = ", kurtosis(yg))        # sample kurtosis  
sample kurtosis = 4.116735
```

**Relative Frequency Histogram of Data**



**Empirical and Gaussian C.D.F.'s**



**Problem 4:** Run the following R code.

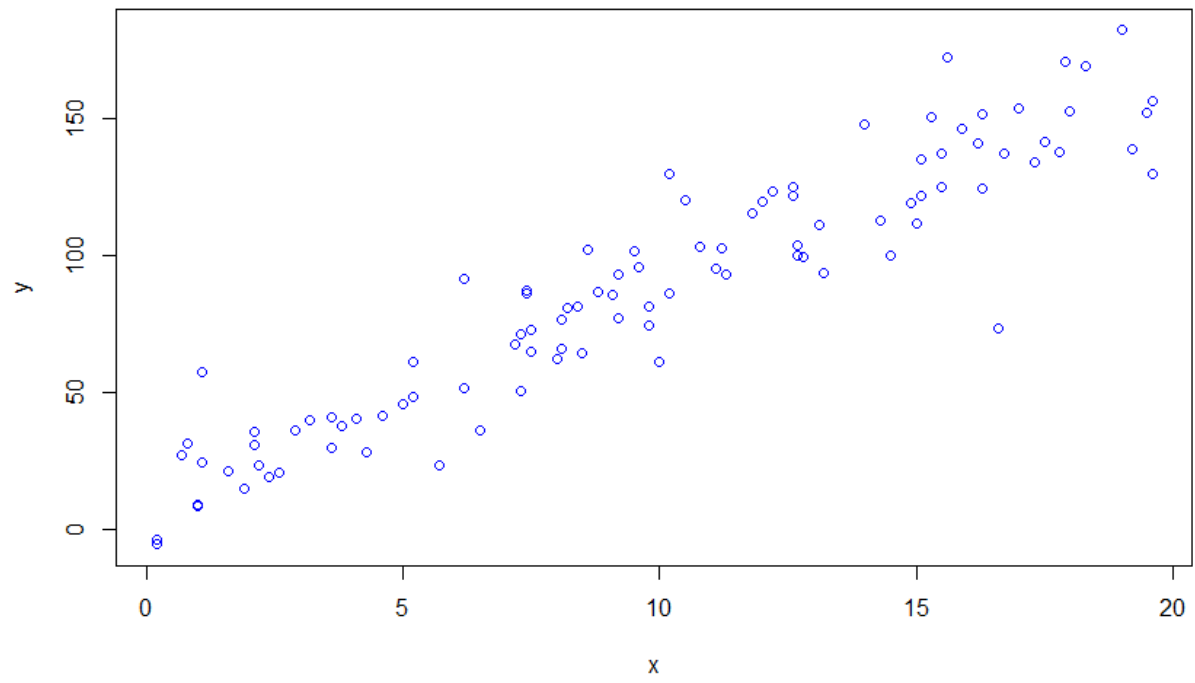
```
#####  
# Problem 4: R code for bivariate data  
set.seed(id)  
x<-round(runif(100,0,20),digits=1)  
alpha<-mean(yn)  
beta<-mean(ye)  
# display values of alpha and beta  
cat("alpha = ", alpha, ", beta = ", beta)  
y<-round(alpha+beta*x+rnorm(100,0,beta*2),digits=1)  
# display first 5 pairs of data  
matrix(c(x[1:5],y[1:5]),nrow=5,ncol=2,byrow=F)  
# display sample correlation  
cat("sample correlation = ", cor(x,y))  
plot(x,y,col="blue",main="Scatterplot of Data")  
#####
```

**Verify that you obtain the following output and plots:**

```
> cat("alpha = ", alpha, ", beta = ", beta)  
alpha = 8.11465 , beta = 7.9169  
> y<-round(alpha+beta*x+rnorm(100, 0, beta*2), digits=1)  
> # display first 5 pairs of data  
> matrix(c(x[1:5], y[1:5]), nrow=5, ncol=2, byrow=F)  
      [, 1] [, 2]  
[1, ]  1.1 24.1  
[2, ]  1.9 14.9  
[3, ]  8.5 64.1  
[4, ] 15.5 136.9  
[5, ] 19.6 156.0  
> # display sample correlation  
> cat("sample correlation = ", cor(x,y))  
sample correlation = 0.9365159
```



Scatterplot of Data



Run the R code for the 4 problems above again except modify the line

```
"id<-20456458"
```

in Problem 1 by replacing the number 20456458 with your UWaterloo ID number.

When you run the R code with your ID number you will generate 8 new plots. Export these 8 plots as .png files using RStudio (See Introduction to R and RStudio Section 6).

Download the Assignment 1 Template which is posted as a Word document on Learn. Fill in the required information and plots based on the output for the data generated using your ID number. Your assignment must follow the template exactly. See Assignment 1 Example posted on Learn.

Create a .pdf file for the answer to EACH problem.

Here are some options for creating pdf files:

Most word processing software will allow you to save your file as a PDF; however, if you require software to create PDFs, some free options are listed below:

- Use a free word processing program that can export directly to PDF, such as [OpenOffice.org](https://www.openoffice.org).
- Download and install a PDF printer driver such as [PrimoPDF](https://www.primopdf.com).
- Other alternatives can be found by searching the Internet using the search words "convert files to PDF."

Upload your assignment to Crowdmark one problem at a time using the link which was emailed to you.

Follow the Crowdmark instructions for completing and submitting at <https://crowdmark.desk.com/customer/portal/articles/1639407-completing-and-submitting-an-assignment>