Lab 3 (80 pts.) - Assessing the Normality of Data
Objectives: Creating and Interpreting Normal Quantile Plots

Note: The data sets are not included in the solutions; they do need to be included as an appendix which is clearly labeled.

A. (10 points) Normal random numbers (no data file required) Use software to generate 10 observations from a normal distribution with $\mu = 2$ and $\sigma = 10$.

1. (5 pts.) Make an appropriate histogram of these observations. How does the shape of the histogram compare with a normal density curve?
2. (5 pts.) Make a normal quantile plot of the data. Does the plot suggest any important deviations from normality? Please provide specifics to explain your answer.

(You must submit your data for this question. No credit will be given without data.)

Solution:

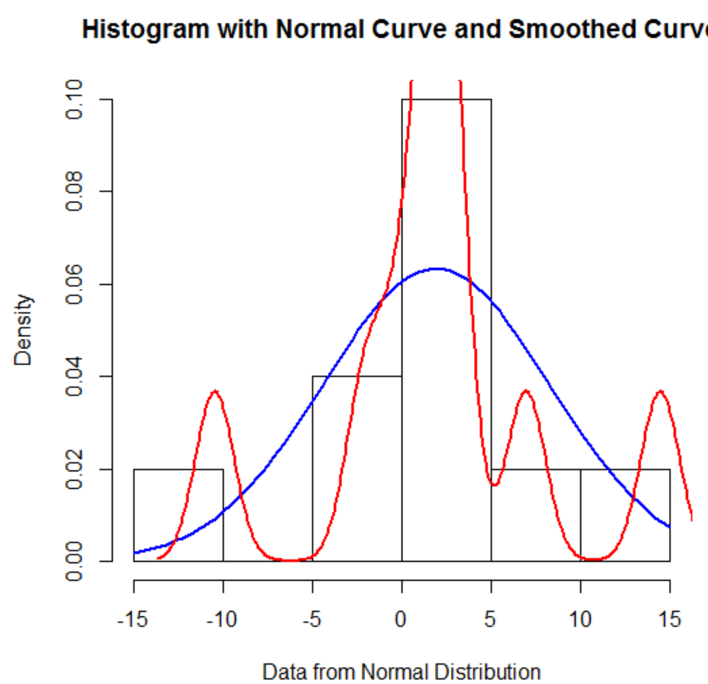
```
> RandomData <- rnorm(10,mean=2,sd=10)
> RandomData
```

1.(5 pts.) Make an appropriate histogram of these observations. How does the shape of the histogram compare with a normal density curve?

Solution:

```
> std<-sd(RandomData)
> m <- mean(RandomData)

> hist(RandomData, xlab="Data from Normal Distribution", freq = FALSE,
+ main="Histogram with Normal Curve and Smoothed Curve")
> curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
> lines(density(RandomData),col = "red", lwd=2)
```

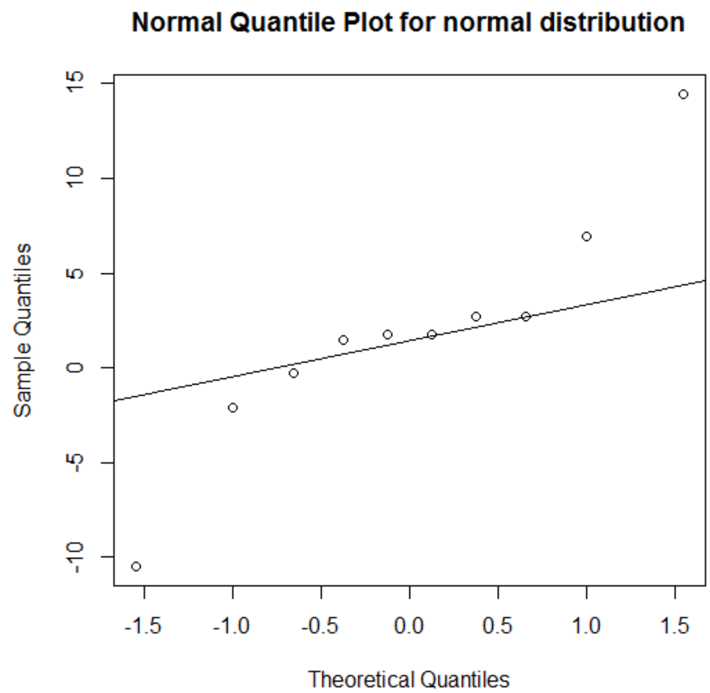


In this case the fitted curve makes it seem that this is a multimodal distribution therefore it is hard to say whether the histogram is normal or not. Different people might get different shapes of the histogram including some that are close to normal and some that are further away.

2.(5 pts.) Make a normal quantile plot of the data. Does the plot suggest any important deviations from normality? Please provide specifics to explain your answer.

Solution:

```
> qqnorm(RandomData,main="Normal Quantile Plot for normal distribution")
> qqline(RandomData)
```



This does not look like a straight line but curving away in a regular pattern. This suggest important deviations from normality

B. (10 points) Normal random numbers (no data file required) Use software to generate 100 observations from a normal distribution with μ = 2 and σ = 10 .

1. (3 pts.) Make an appropriate histogram of these observations. How does the shape of the histogram compare with a normal density curve?
2. (3 pts.) Make a normal quantile plot of the data. Does the plot suggest any important deviations from normality? Please provide specifics to explain your answer.
3. (4 pts.) Are the plots from part A and part B the same or different? Please explain your answer. Remember both of these parts are from the same normal distribution.

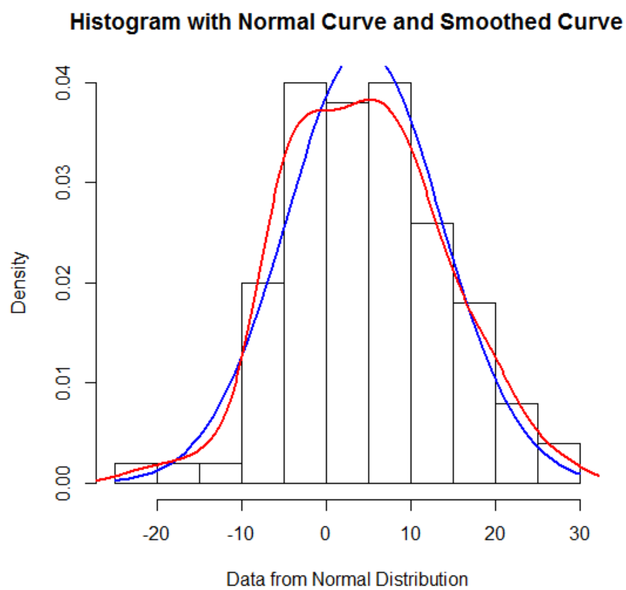
Solution:

```
> RandomData <- rnorm(100,mean=2,sd=10)
> RandomData
```

1. (3 pts.) Make an appropriate histogram of these observations. How does the shape of the histogram compare with a normal density curve?

Solution:

```
> std<-sd(RandomData)
> m <- mean(RandomData)
> hist(RandomData, xlab="Data from Normal Distribution", freq = FALSE,
+ main="Histogram with Normal Curve and Smoothed Curve")
> curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
```

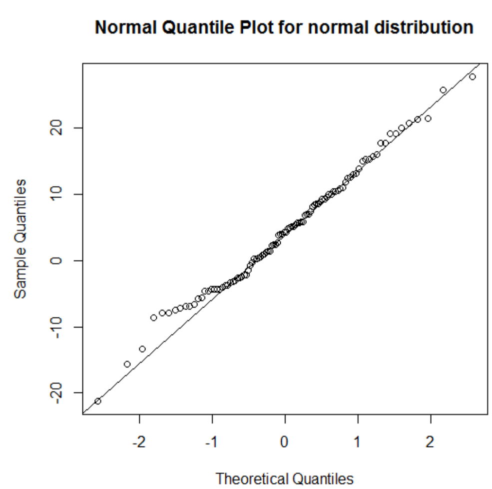


It seems that the shape of the histogram fits the normal shape very well.

2.(3 pts.) Make a normal quantile plot of the data. Does the plot suggest any important deviations from normality? Please provide specifics to explain your answer.

Solution:

```
> qqnorm(RandomData,main="Normal Quantile Plot for normal distribution")
> qqline(RandomData)
```



There are a few points that are deviant from the straight line. In general, I think the QQ-plot doesn't suggest important deviations from normality.

3. (4 pts.) Are the plots from part A and part B the same or different? Please explain your answer. Remember both of these parts are from the same normal distribution.

Solution:

The plots from A and B are different in that the they are two independent datasets of different sizes, albeit, produced from the same parent distribution. The large size of the second dataset endows it with a better sense of normality as visualized with the histogram and the probability plot. In fact, the plots from A don't necessarily look like they come from a normal distribution at all.

C. (40 points) Random numbers from other distributions (no data file required.) Use software to generate 100 observations from the distributions called (I) right skewed, (II) left skewed, (III) short tailed and (IV) long tailed in the tutorial. For each of the distributions answer the following questions (the answers for each distribution should be right after each other):

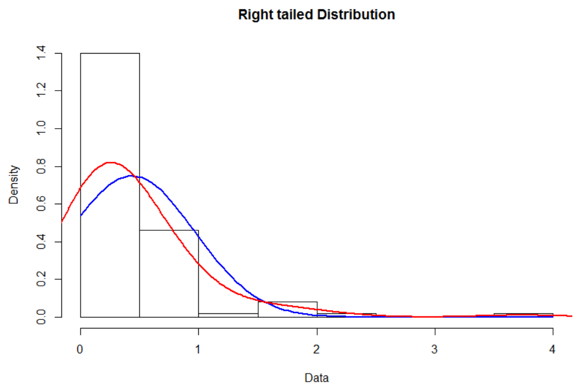
1. (5 pts. for each distribution) Make an appropriate histogram of these observations. Please describe the shape of the distribution. How does the histogram compare with a normal density curve?
2. (5 pts. for each distribution) Make a normal quantile plot of your data. Please describe the shape of the plot. Does the plot suggest any important deviations from normality?

Solution:
Note: the solutions of two questions above are organized by the type of the distribution.

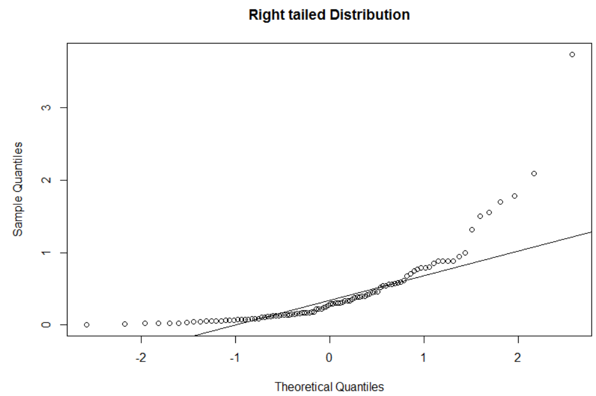
I. Right skewed distribution:
Solution:

```
> n = 100
> right <- rexp(n,rate=2)
> left <- rbeta(n,2,0.5,ncp=2)
> long <- rcauchy(n,location=0,scale=1)
> short <- runif(n,min=0,max=2)
> RandomData <- right
> #Data
> RandomData
> title <- "Right tailed Distribution"
> std<-sd(RandomData)
> m <- mean(RandomData)
> hist(RandomData, xlab="Data", freq = FALSE, main=title)
> curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
> lines(density(RandomData,adjust=3),col = "red", lwd=2)
> dev.new()
> qqnorm(RandomData,main=title)
> qqline(RandomData)
```

1.



2.



The shape of the histogram is right skewed. The histogram deviates from the normal.	The plot is curved upward (concave). The QQ-plot suggests important deviations from normality.

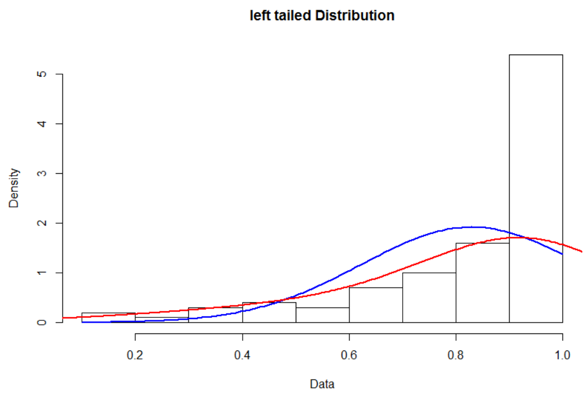
II. Left skewed distribution:

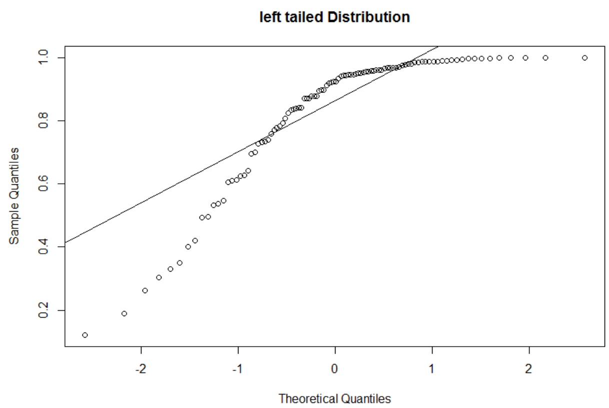
Solution:

```
> RandomData <- left
> #Data
> RandomData
> title <- "left tailed Distribution"
> std<-sd(RandomData)
> m <- mean(RandomData)
> hist(RandomData, xlab="Data", freq = FALSE, main=title)
> curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
> lines(density(RandomData,adjust=3),col = "red", lwd=2)
> lines(density(RandomData,adjust=3),col = "red", lwd=2)
> dev.new()
> qqnorm(RandomData,main=title)
> qqline(RandomData)
```

1.

2.





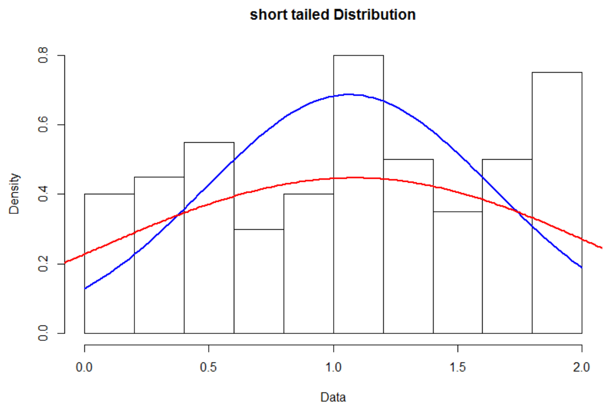
It is left skewed. The histogram deviates from normal curve.	The curve is open downward (convex). The QQ-plot suggests important deviations from normality.

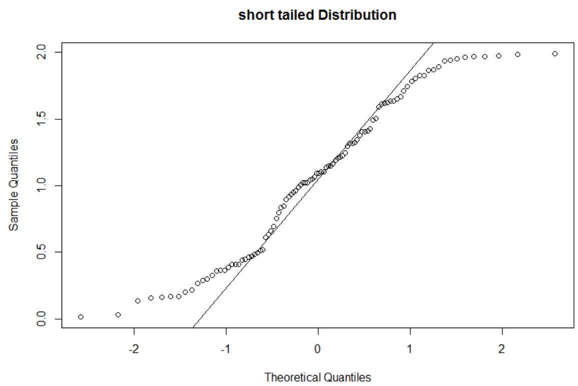
III. Short tailed Distribution:

Solution:

```
> RandomData <- short
> #Data
> RandomData
> title <- "short tailed Distribution"
> std<-sd(RandomData)
> m <- mean(RandomData)
> hist(RandomData, xlab="Data", freq = FALSE, main=title)
> curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
> lines(density(RandomData,adjust=3),col = "red", lwd=2)
> lines(density(RandomData,adjust=3),col = "red", lwd=2)
> dev.new()
> qqnorm(RandomData,main=title)
> qqline(RandomData)
```

1. 2.





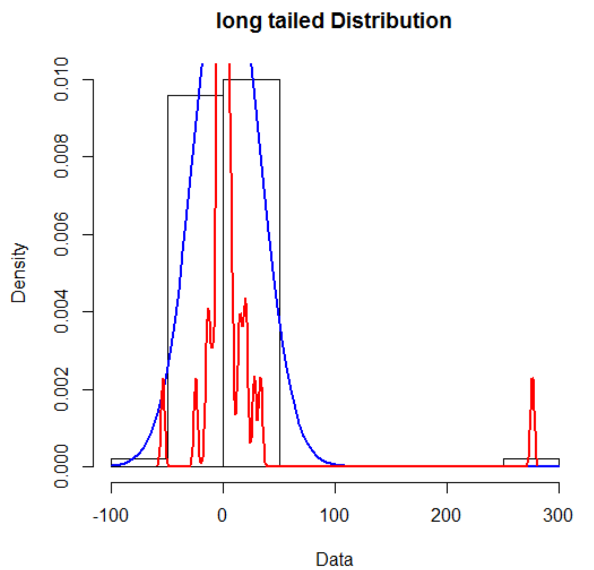
The histogram does not "peak" well like the normal density curve. Also, it fails to produce the tails of the normal density curve. It deviates from normal curve.	The points generally have high slope near the center but low slope near the end (S-shaped) The QQ-plot suggests important deviations from normality.

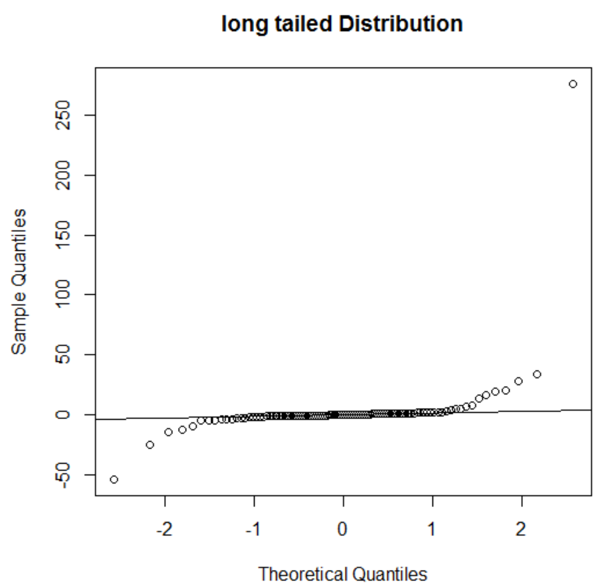
IV. Long tailed distribution:

Solution:

```
> RandomData <- long
> #Data
> RandomData
> title <- "long tailed Distribution"
> std<-sd(RandomData)
> m <- mean(RandomData)
> hist(RandomData, xlab="Data", freq = FALSE, main=title)
> curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
> lines(density(RandomData,adjust=3),col = "red", lwd=2)
> dev.new()
> qqnorm(RandomData,main=title)
> qqline(RandomData)
```

1. 2.





This histogram has a sharp peak with outliers. It deviates from normal curve.	The points generally have low slope near the center but high slope near the end (S-shaped). The QQ-plot suggests important deviations from normality.

D. (10 points) How long are customer service center calls? (data file: eg01-15calls80.txt) We have data on the lengths of all 31,492 calls made to the customer service center of a small bank in a month. The table below displays the lengths of the first 80 calls, in seconds.

TABLE 1.2 Service Times (Seconds) for Calls to a Customer Service Center

77	289	128	59	19	148	157	203
126	118	104	141	290	48	3	2
372	140	438	56	44	274	479	211
179	1	68	386	2631	90	30	57
89	116	225	700	40	73	75	51
148	9	115	19	76	138	178	76
67	102	35	80	143	951	106	55
4	54	137	367	277	201	52	9
700	182	73	199	325	75	103	64
121	11	9	88	1148	2	465	25

1. (5 pts.) Make an appropriate histogram of these observations. Which distribution do you think this data is? (normal, right skewed, left skewed, short tailed or long tailed). Please explain your answer.
2. (5 pts.) Make a normal quantile plot of the data. Which distribution do you think this data is? (normal, right skewed, left skewed, short tailed or long tailed). Please explain your answer.

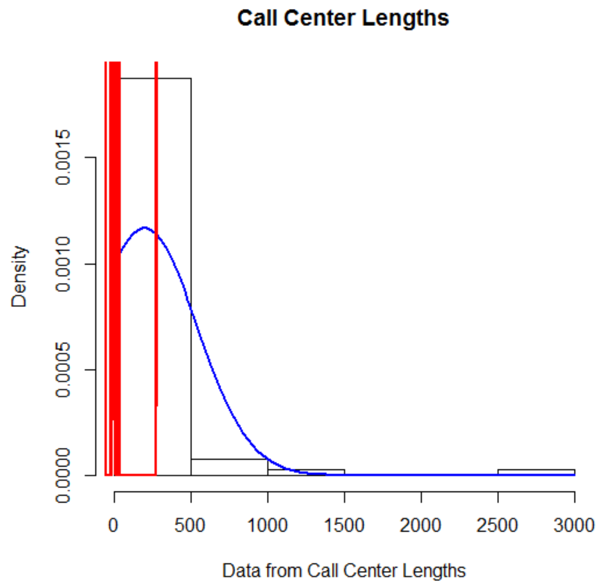
Solution:

```
> call<- read.table("eg01-15calls80.txt", header=TRUE)
>call
>attach (call)
```

1.(5 pts.) Make an appropriate histogram of these observations. Which distribution do you think this data is? (normal, right skewed, left skewed, short tailed or long tailed). Please explain your answer.

Solution:

```
> std<-sd(length)
> m <- mean(length)
> hist(length, xlab="Data from Call Center Lengths", freq = FALSE,
main="Call Center Lengths")
> curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
> lines(density(length,adjust=3),col = "red", lwd=2)
> dev.new()
```

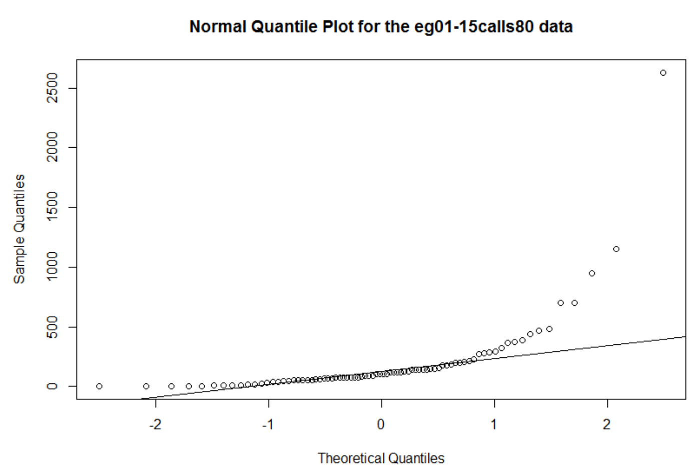


Based on the histogram, I would say this distribution is skewed-right and resembles the simulated skewed-right dataset.

2.(5 pts.) Make a normal quantile plot of the data. Which distribution do you think this data is? (normal, right skewed, left skewed, short tailed or long tailed). Please explain your answer.

Solution:

```
> qqnorm(length, main=Normal Quantile Plot for the eg01-15calls80 data")
> qqline(length)
```



This QQ-plot is very similar to the one produced for the skewed-right dataset. This confirms my assessment of the histogram.

E. (10 points) Comparison of data. This is a group assignment and is due on Blackboard at Midnight on Monday, Feb. 9. Each group must consist of 3-4 people (in any of my sections) and will submit a combined report. Be sure that the names and sections of each person are at the top of the page.

1. (3 pts.) For Parts A, B and C, present all of the graphs for each part from each student ordered by part. Therefore, there will be 3-4 histograms and 3-4 normal quantiles plots for each type of random number simulation (6 in total).
2. (7 pts.) After each of the six sets of plots, please answer the following question: “Are each of these plots the same or different? If they are different, please propose a possible explanation.”

Solution:

Response for 2: I would expect all of them to look similar except for A with only 10 data points and maybe the Cauchy in C.

The problem with A is that the sample size is too small. The problem with Cauchy is that a change in the location and number of outliers will change the shape of the plot.