Log Transformation

Lucky

Log Transformation

You've picked up a bunch of rocks from a rocky beach and want to estimate the weight of all the rocks at the beach with a Confidence level of 93.47%.

a. Plot the qqline and boxplot of the data. Also get the skewness. What is your conclusion about the distribution being normal?
b. Do a log transformation and perform the steps in a. What's your conclusion?

Use Log transformed data for the following questions.

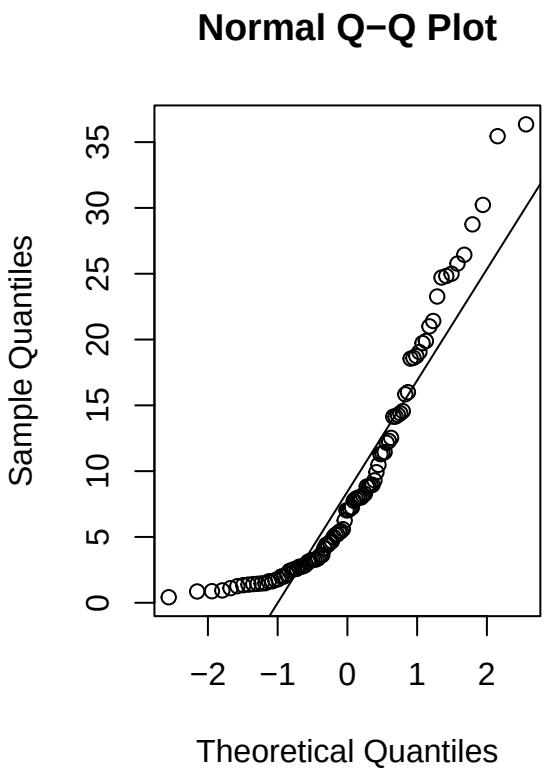
c. What is the Mean, Std dev, and the Sample size.

d. Find Standard Error.

e. Find the t-score for the 93.47% confidence interval.

f. Use this t-score, sample mean, std error to get the upper and lower limit of the confidence interval.
g. Do reverse transformation to get the confidence interval in Ounces.

```
# Clear the environment
rm(list = ls())
# Load moments library for function "skewness"
library(moments)
# Load readxl library
library(readxl)
# Read this excel file
prob4 <- read_excel("Example3.xlsx", sheet = "Prob4")
# Plot  QQ Line and Box Plot
# Plot the QQ Line
# Set graphical parameters
par(mfrow = c(1,2))
qqnorm(prob4$Weight)
qqline(prob4$Weight)
# Plot box plot
# Set graphical parameters
par(mfrow = c(1,3))
```




```
boxplot(prob4$Weight, main = "Box Plot")
# Get the skewness
skew <- skewness(prob4$Weight); skew
```

```
## [1] 1.239558
```

```
# Conclusion about the distribution being normal?
# The distribution is not normal as most of dots in the QQ Plot are not on a straight line
# Since median is closer to the bottom of the box, and whisker is shorter on the lower end then the dis
cat("The distribution is not normal as most of dots in the QQ Line are not on a straight line")
```

```
## The distribution is not normal as most of dots in the QQ Line are not on a straight line
```

```
cat("Median is closer to bottom, whisker shorter lower. The distribution is positively skewed")
```

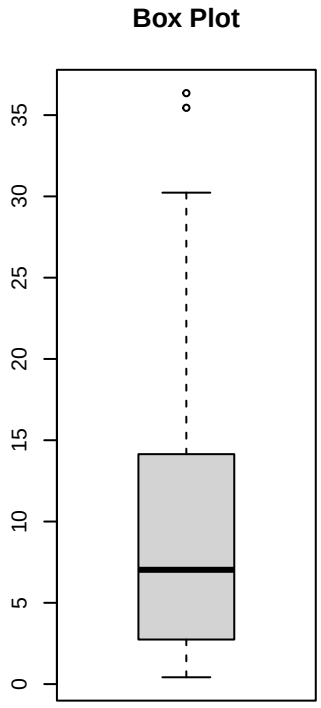
```
## Median is closer to bottom, whisker shorter lower. The distribution is positively skewed
```

```
# Log Transformation
log1 <- log(prob4$Weight); log1
```

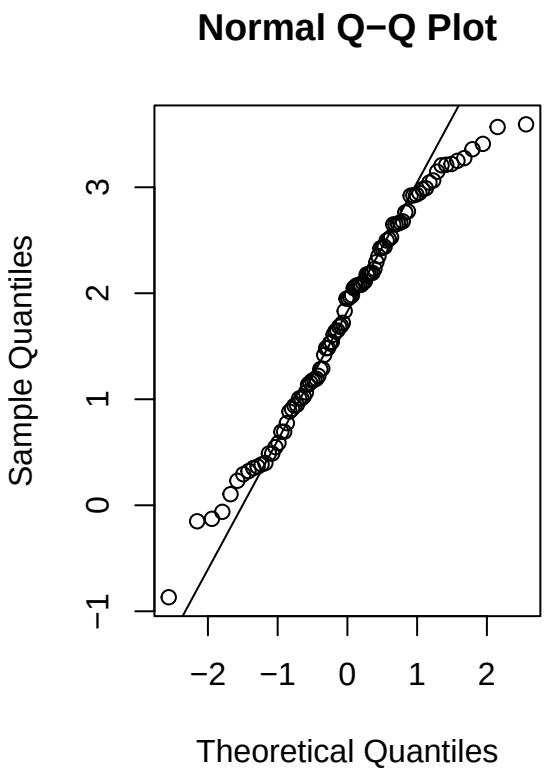
```
##  [1] -0.8675006 -0.1508229 -0.1278334  0.1043600  0.2926696  0.3220835
##  [7]  0.3506569  0.3852624  0.4885800  0.5481214  0.7747272  0.8837675
## [13]  0.9082586  1.0079579  1.0079579  1.0260416  1.0647107  2.3494687
```

```
## [19]  2.4239174  2.5095993  1.6409366  2.0992442  1.6937791  2.2300144
## [25]  2.1792869  2.0794415  2.0476928  1.9810015  1.9671124  1.9516082
## [31]  1.7209793  1.6486586  1.6094379  1.4816045  1.4182774  1.2892326
## [37]  1.2837078  1.1939225  1.1847900  1.1568812  1.1378330  2.5281258
## [43]  3.2120526  2.6490077  2.6497146  2.6581594  2.6782780  2.7625384
## [49]  2.4379897  2.9821403  2.9204698  1.8325815  1.6845454  1.5411591
## [55]  1.5260563  1.4838747  3.2748780  3.2492110  2.9902171  1.2208299
## [61]  1.1755733  0.9477894  0.9400073  0.6931472  0.6931472  0.5877867
## [67]  0.4885800  0.3987761  0.3646431  0.2311117 -0.0618754  1.9487632
## [73]  2.0655961  2.0756845  2.0794415  2.1138430  2.1781550  2.7725887
## [79]  2.6672282  2.4973292  2.4300984  2.2945529  2.1927702  2.1871742
## [85]  2.9236991  2.9311938  2.9470671  3.0445224  3.0638581  3.1471650
## [91]  3.2072080  3.2188758  3.3586378  3.4088348  3.5681233  3.5931942
```

```
# Normal Q-Q Plot and Box Plot of Log Transformation
# Plot the QQ Line
# Set graphical parameters
par(mfrow = c(1,2))
```




```
qqnorm(log1)
qqline(log1)
# Plot box plot of the data
# Set graphical parameters
par(mfrow = c(1,3))
```




```
boxplot(log1, main = "Log Transformed Box Plot")
# Get the skewness of log transformed
skew1 <- skewness(log1); skew1
```

```
## [1] -0.2234782
```

```
# What's your conclusion?
# The distribution is normal as most of dots in the QQ Plot are on a straight line
# Since median is closer to the top of the box, and whisker is shorter on the upper end then the distri
cat("The distribution is normal as most of dots in the QQ Line are on a straight line")
```

```
## The distribution is normal as most of dots in the QQ Line are on a straight line
```

```
cat("Median is closer to top, whisker shorter at upper, the distribution is negatively skewed")
```

```
## Median is closer to top, whisker shorter at upper, the distribution is negatively skewed
```

```
# Log Transformed data
# What is the Mean, Std Dev, and the Sample Size?
# N
n <- length(log1); n
```

```
## [1] 96
```

```r
# Mean
m <- mean(log1); m
```

```
## [1] 1.780766
```

```r
# Std Dev
std <- sd(log1); std
```

```
## [1] 1.024418
```

```r
# Find Standard Error
sx_bar <- std / sqrt(n); sx_bar
```

```
## [1] 0.1045542
```

```r
# Find t-score for the 93.47% confidence interval
t_score <- m / sx_bar; t_score
```

```
## [1] 17.03199
```

```r
# Get the Upper and Lower limit of the confidence interval
# Find critical
p1 <- 1 - (93.47/100)
critical <- qt(p1/2, n-1, lower.tail = FALSE)
# Upper and Lower limit.
upperlimit <- m + critical * sx_bar; upperlimit
```

```
## [1] 1.975736
```

```r
lowerlimit <- m - critical * sx_bar; lowerlimit
```

```
## [1] 1.585796
```

```r
# Reverse transformation to get the confidence interval in Ounces
reverseupperlimit <- exp(`upperlimit`); reverseupperlimit
```

```
## [1] 7.211927
```

```r
reverselowerlimit <- exp(`lowerlimit`); reverselowerlimit
```

```
## [1] 4.883177
```

Log Transformed Box Plot