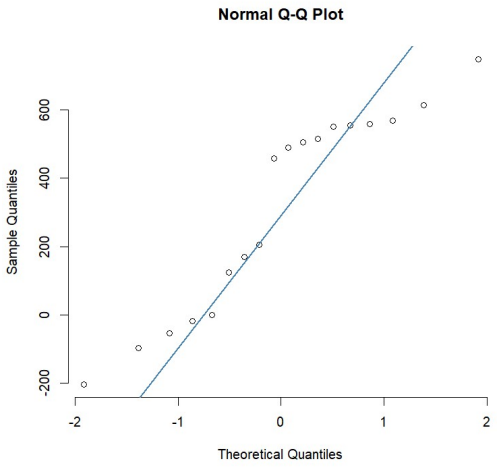
Submission for Homework 8 – James Fogg

For *Concepts in Statistical Data Analysis*, Spring 2026

Note: See the end of the document at the end for the full R script. I just commented out the code except for the specific plot that I needed, one at a time.

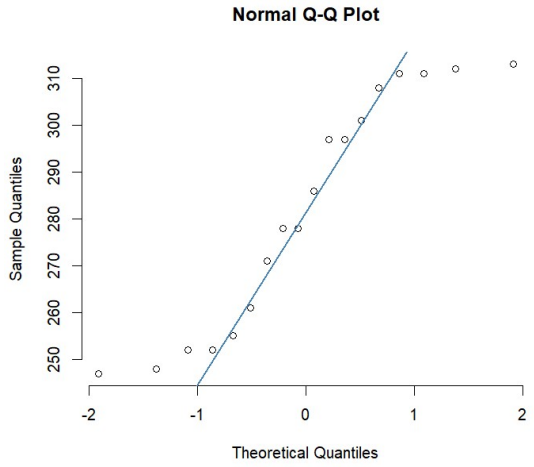
a.)



This data does not come from a normal distribution. Towards either end of the graph, there are significant outliers both above and below the line. Further analysis with the Shapiro-Wilks test as shown in the lecture notes confirms this, as the P value is extremely small. The R code that was used for this:

```
mydata_1 = c(747,-1,-17,555,125,169,-96,-53,614,488,205,514,457,505,568,551,558,-203)

#a

qqnorm(mydata_1, pch = 1, frame = FALSE)

qqline(mydata_1, col = "steelblue" , lwd = 2)

shapiro.test(mydata_1) #optional test? It's included in the notes.
```

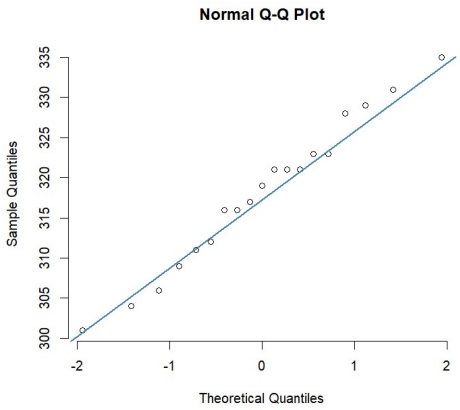
b.) The graph was too big to fit here. See next page.



This data likely did not come from a normal distribution. There are extreme outliers on either end of the graph. This is further confirmed with the Shapiro-Wilks test, which shows a P value of 0.02967, which is small. This is included because the sample code included in the lecture notes includes this. The R code that can be used for this is:

```
mydata_2 =
c(311,312,278,278,271,313,308,297,252,297,311,255,248,301,247,286,261,252)

#b

qqnorm(mydata_2, pch = 1, frame = FALSE)

qqline(mydata_2, col = 'steelblue', lwd = 2)

shapiro.test(mydata_2)
```

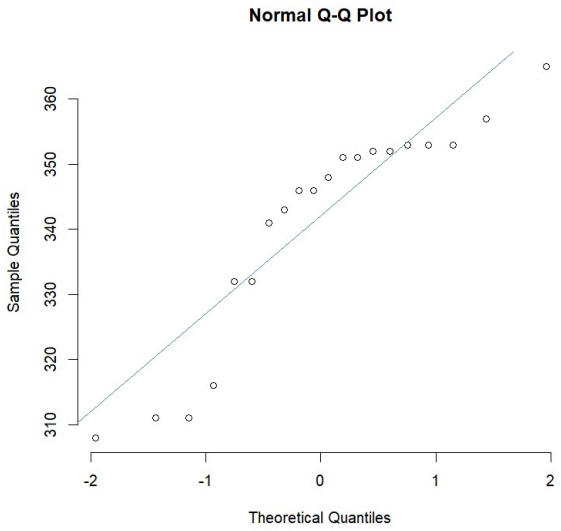
c.) (If the graph was any larger it wouldn't fit on the page)



This graph shows that the data likely DID come form a normal distribution. There are almost no discernible outliers. This is further reinforced by running shapiro.test() as included in the lecture notes, which produces a high p value of 0.954. The R codes that were used for this:

```
mydata_3 =
c(323,304,321,321,321,301,306,323,331,316,319,312,329,309,311,328,335,316,317)

#c

qqnorm(mydata_3, pch = 1, frame = FALSE)

qqline(mydata_3, col = 'steelblue', lwd = 2)

shapiro.test(mydata_3)
```

d.)



This graph likely did not come from a normal distribution. There are extreme outliers at either end of the graph, and almost nothing stays centered around the line that is shown. In addition, the Shapiro-wilks test backs this up bny producing a p-value of 0.005791.

R Code that was used for this:

```
mydata_4 =
c(351,348,352,351,353,353,357,353,341,311,343,365,352,311,316,346,308,346,332,332)

#d

qqnorm(mydata_4, pch = 1, frame = FALSE)
```

qqline(mydata_4, col = 'steelblue', lwd = 2)

shapiro.test(mydata_4)

The full R script that was used for this assignment:

```
# hw08.R -- R code used for homework 8 of Concepts in Statistical Data Analysis, Spring 2026.

# Author: James C. Fogg

# Version: 2026.02.23


#setting up the datasets to use.

mydata_1 = c(747,-1,-17,555,125,169,-96,-53,614,488,205,514,457,505,568,551,558,-203)

mydata_2 = c(311,312,278,278,271,313,308,297,252,297,311,255,248,301,247,286,261,252)

mydata_3 = c(323,304,321,321,321,301,306,323,331,316,319,312,329,309,311,328,335,316,317)

mydata_4 = c(351,348,352,351,353,353,357,353,341,311,343,365,352,311,316,346,308,346,332,332)


#generate the qq plots


#a

qqnorm(mydata_1, pch = 1, frame = FALSE)

qqline(mydata_1, col = "steelblue" , lwd = 2)

shapiro.test(mydata_1) #optional test? It's included in the notes.


#b

qqnorm(mydata_2, pch = 1, frame = FALSE)

qqline(mydata_2, col = 'steelblue', lwd = 2)
```

```
shapiro.test(mydata_2)


#c
qqnorm(mydata_3, pch = 1, frame = FALSE)
qqline(mydata_3, col = 'steelblue', lwd = 2)
shapiro.test(mydata_3)


#d
qqnorm(mydata_4, pch = 1, frame = FALSE)
qqline(mydata_4, col = 'steelblue', lwd = 2)
shapiro.test(mydata_4)
```