Andrew Moore, 09/27/2021

MATH-471, Homework 3.2

Question 1.

Read Ch.4.10 in the textbook and do the following questions from the textbook: pp. 222-223: 4.65, 4.66, 4.70, 4.72.

4.65

Let y be a random variable having a normal distribution with a mean equal to 250 and a standard deviation equal to 50. Find the following probabilities:

a. P(y > 250)

b. P(y > 150)

c. P(150 < y < 350)

d. Find k such that P(250 - k < y < 250 + k) = .60

a.

```
pnorm(250, mean = 250, sd = 50, lower.tail = FALSE)
```

```
## [1] 0.5
```

b.

```
pnorm(150, mean = 250, sd = 50, lower.tail = FALSE)
```

```
## [1] 0.9772499
```

c.

```
pnorm(350, mean = 250, sd = 50) - pnorm(150, mean = 250, sd = 50)
```

```
## [1] 0.9544997
```

d.

```
k <- 42.089

pnorm(250 + k, mean = 250, sd = 50) - pnorm(250 - k, mean = 250, sd = 50)
```

```
## [1] 0.6000889
```

4.66

Suppose that y is a random variable having a normal distribution with a mean equal to 250 and a standard deviation equal to 10.

a. Show that the event $y < 260$ has the same probability as $z < 1$.
b. Convert the event $y > 230$ to the z-score equivalent.
c. Find $P(y < 260)$ and $P(y > 230)$.
d. Find $P(y > 265)$, $P(y < 242)$, and $P(242 < y < 265)$.

a.

```
pnorm(260, mean = 250, sd = 10)
```

```
## [1] 0.8413447
```

```
pnorm(1, mean = 0, sd = 1)
```

```
## [1] 0.8413447
```

b.

```
(z <- (230 - 250) / 10)
```

```
## [1] -2
```

```
pnorm(z, mean = 0, sd = 1)
```

```
## [1] 0.02275013
```

c.

```
pnorm(260, mean = 250, sd = 10)
```

```
## [1] 0.8413447
```

```
pnorm(230, mean = 250, sd = 10, lower.tail = FALSE)
```

```
## [1] 0.9772499
```

d.

```
pnorm(265, mean = 250, sd = 10, lower.tail = FALSE)
```

```
## [1] 0.0668072
```

```
pnorm(242, mean = 250, sd = 10)
```

```
## [1] 0.2118554
```

```
pnorm(265, mean = 250, sd = 10) - pnorm(242, mean = 250, sd = 10)
```

```
## [1] 0.7213374
```

4.70

The College Boards, which are administered each year to many thousands of high school students, are scored so as to yield a mean of 513 and a standard deviation of 130. These scores are close to being normally distributed. What percentage of the scores can be expected to satisfy each of the following conditions?

a. Greater than 600

b. Greater than 700

c. Less than 450

d. Between 450 and 600

Taking the normal distribution as a model for the test scores Y, we can compute $P(Y > 600)$, $P(Y > 700)$, $P(Y < 450)$, and $P(450 < Y < 600)$.

a.

```
pnorm(600, mean = 513, sd = 130, lower.tail = FALSE)
```

```
## [1] 0.2516741
```

b.

```
pnorm(700, mean = 513, sd = 130, lower.tail = FALSE)
```

```
## [1] 0.07515157
```

c.

```
pnorm(450, mean = 513, sd = 130)
```

```
## [1] 0.3139746
```

d.

```
pnorm(600, mean = 513, sd = 130) - pnorm(450, mean = 513, sd = 130)
```

```
## [1] 0.4343513
```

4.72

Refer to Exercise 4.70. An honor society wishes to invite those scoring in the top 5% on the College Boards to join their society.

a. What score is required to be invited to join the society?

b. What score separates the top 75% of the population from the bottom 25%? What do we call this value?

a. We want the 95th percentile for test-takers. Again, using the normal distribution for test scores Y, we can compute this value using `qnorm()`.

```
qnorm(0.95, mean = 513, sd = 130)
```

```
## [1] 726.831
```

b. The difference between the 75th and 25th percentile is known as the inter-quartile range (IQR).

```
qnorm(0.75, mean = 513, sd = 130) - qnorm(0.25, mean = 513, sd = 130)
```

```
## [1] 175.3673
```

Question 2.

A consumer testing agency wants to evaluate the claim made by a manufacturer of discount tires. The manufacturer claims that their tires can be driven at least 35,000 miles before wearing out. To determine the average number of miles that can be obtained from the manufacturer's tires, the agency randomly selects 60 tires from the manufacturer's warehouse and places the tires on 15 cars driven by test drivers on a 2-mile oval track. The number of miles driven (in thousands of miles) until the tires are determined to be worn out is given in the following table.

Car	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Miles driven	25	27	35	42	28	37	40	31	29	33	30	26	31	28	30

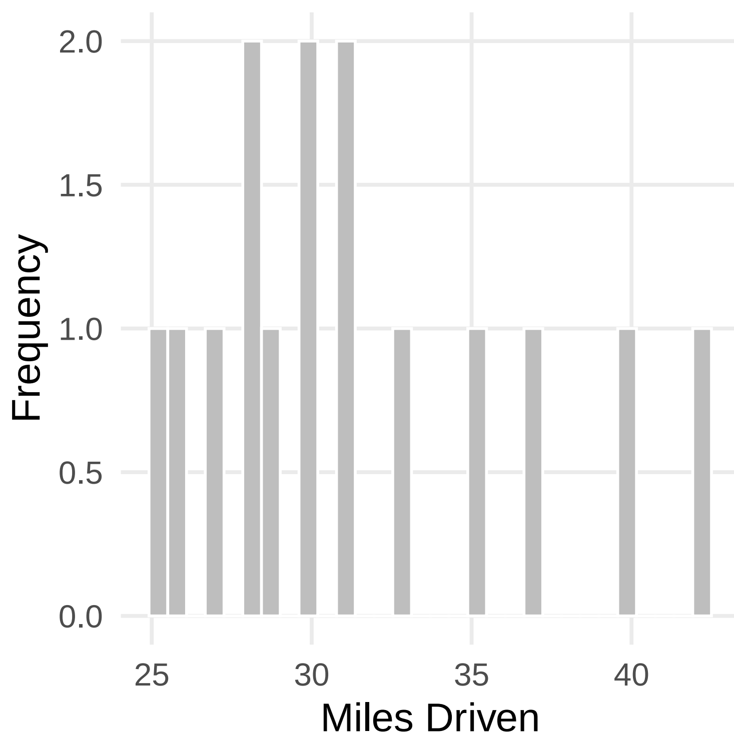
a. Describe the descriptive statistics about the variable 'Miles driven'. Also describe the shape of its distribution.

b. Estimate a population distribution of 'Miles driven' using Q-Q plot.

a. Below is a table of descriptive statistics for the sample of cars, and a histogram.

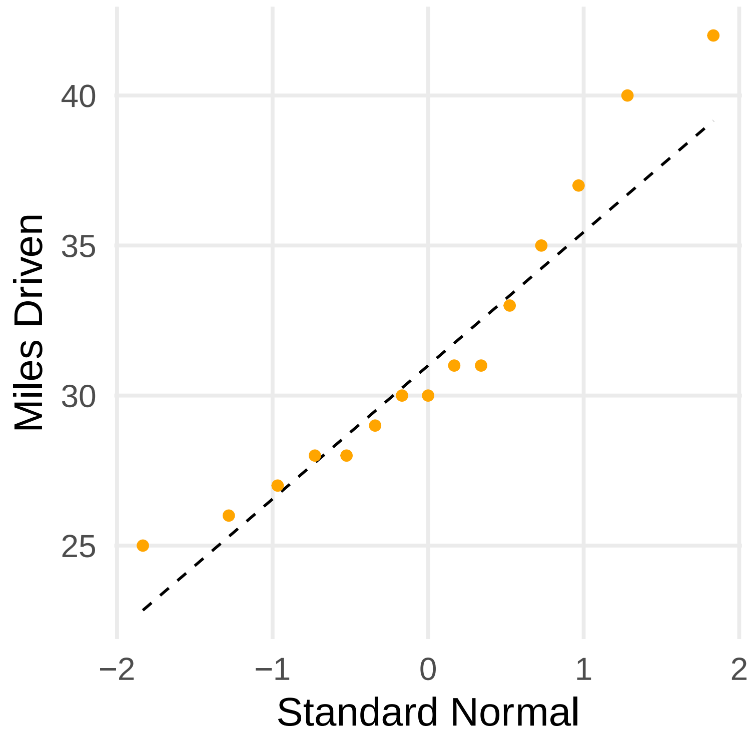
Table 2: Descriptive statistics for the sample of cars.

N (# cars)	Mean	Median	SD	25th Pctile.	75th Pctile.	Range
15	31.5	30	5	28	34	[25, 42]



There are only 15 data points, and there appears to be a fair amount of variability within the sample. The data do not appear to be symmetric; they exhibit positive skew.

b. Below is a QQ-plot of the sample of cars. The departure from linearity in the upper-right is visible, reflecting the long rightward tail in the histogram. A Chi-square distribution or F-distribution seem like possible candidates.


40
35
30
25
Miles Driven
−2
−1
0
1
2
Standard Normal

Question 3.

Read Ch.4.14 in the textbook and submit the following questions from the textbook: pp. 227-229: 4.94 and 4.114.

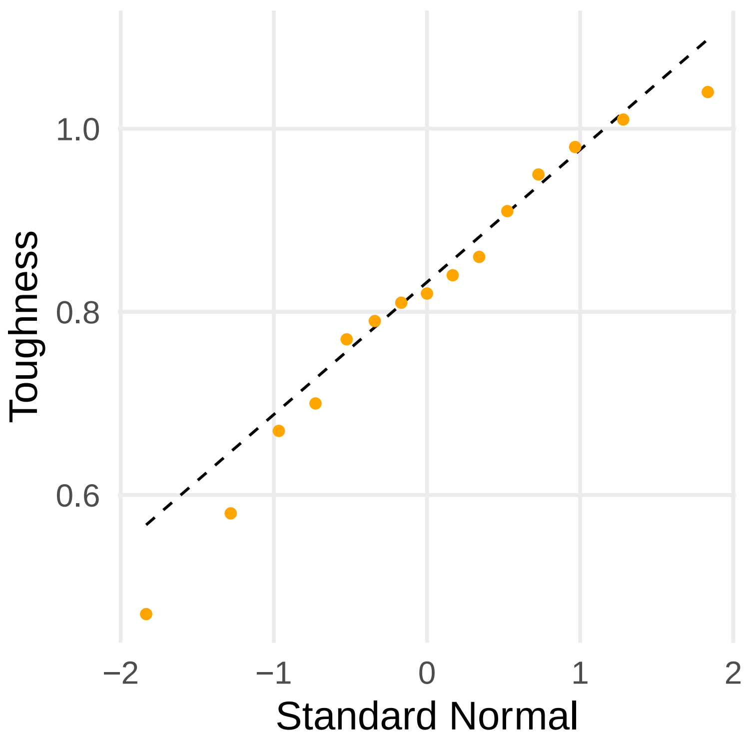
4.94

The fracture toughness in concrete specimens is a measure of how likely it is that blocks used in new home construction may fail. A construction investigator obtains a random sample of 15 concrete blocks and determines the following toughness values:

```
blocks <- c(.47, .58, .67, .70, .77, .79, .81, .82, .84, .86, .91, .95, .98, 1.01, 1.04)
```

a. Use a normal quantile plot to assess whether the data appear to fit a normal distribution.
b. Compute the correlation coefficient and p-value for the normal quantile plot. Comment on the degree of fit of the data to a normal distribution.

a. Visually, the relationship looks highly linear, although the total # of data points is small.



b. The correlation coefficient is very large; with the associated p value indicating an extremely low probability of observing this data if the true correlation was 0. Overall, both graphical and correlational analysis indicate that the toughness of the blocks fits a normal distribution.

```
##
##  Pearson's product-moment correlation
##
## data:  x and q
## t = 19.41, df = 13, p-value = 5.54e-11
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9487576 0.9945445
## sample estimates:
```

```
##       cor
## 0.9831813
```

4.114

As part of a study to determine factors that may explain differences in animal species relative to their size, the following body masses (in grams) of 50 different bird species were re ported in the paper "Temperature and the Northern Distributions of Wintering Birds," by Richard Repasky (1991).

```
mass <- c(
  7.7, 10.1, 21.6, 8.6, 12.0, 11.4, 16.6, 9.4, 11.5, 9.0, 8.2, 20.2, 48.5, 21.6,
  26.1, 6.2, 19.1, 21.0, 28.1, 10.6, 31.6, 6.7, 5.0, 68.8, 23.9, 19.8, 20.1, 6.0,
  99.6, 19.8, 16.5, 9.0, 448.0, 21.3, 17.4, 36.9, 34.0, 41.0, 15.9, 12.5, 10.2,
  31.0, 21.5, 11.9, 32.5, 9.8, 93.9, 10.9, 19.6, 14.5
)
```

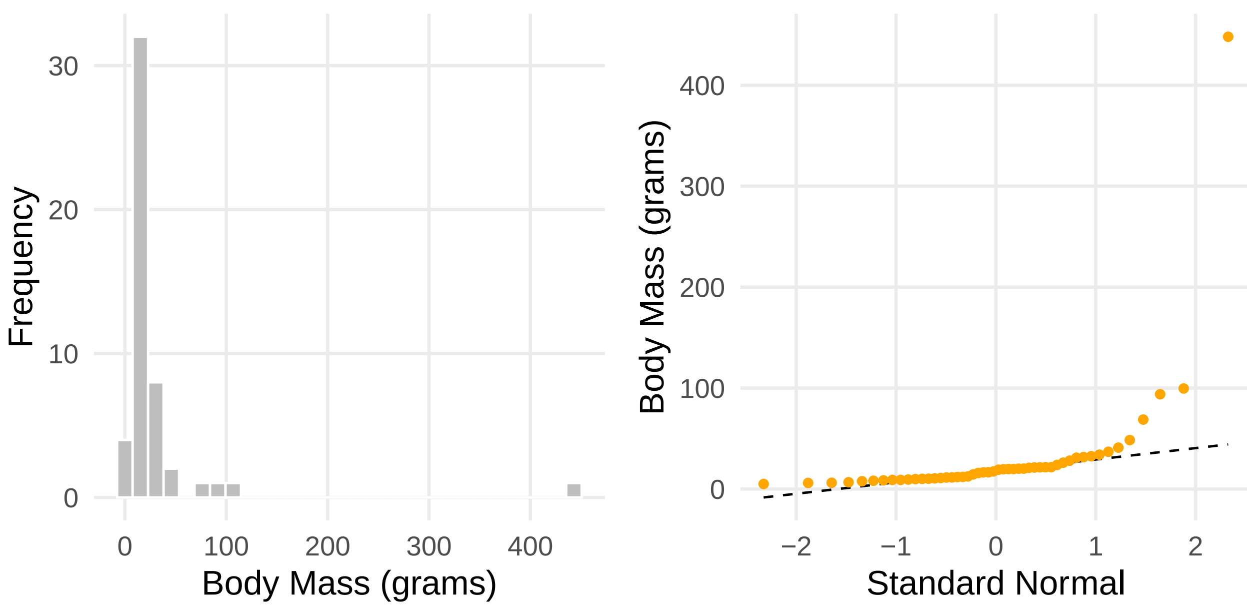
a. Does the distribution of the body masses appear to follow a normal distribution? Provide both a graphical and a quantitative assessment.

b. Repeat part (a), with the outlier 448.0 removed.

c. Determine the sample mean and median with and without the value 448.0 in the data set.

d. Determine the sample standard deviation and MAD with and without the value 448.0 in the data set.

a. Histogram and QQ-plot for the body-mass data (*including* the outlier). Visually, these data do not appear to come from a normal distribution.

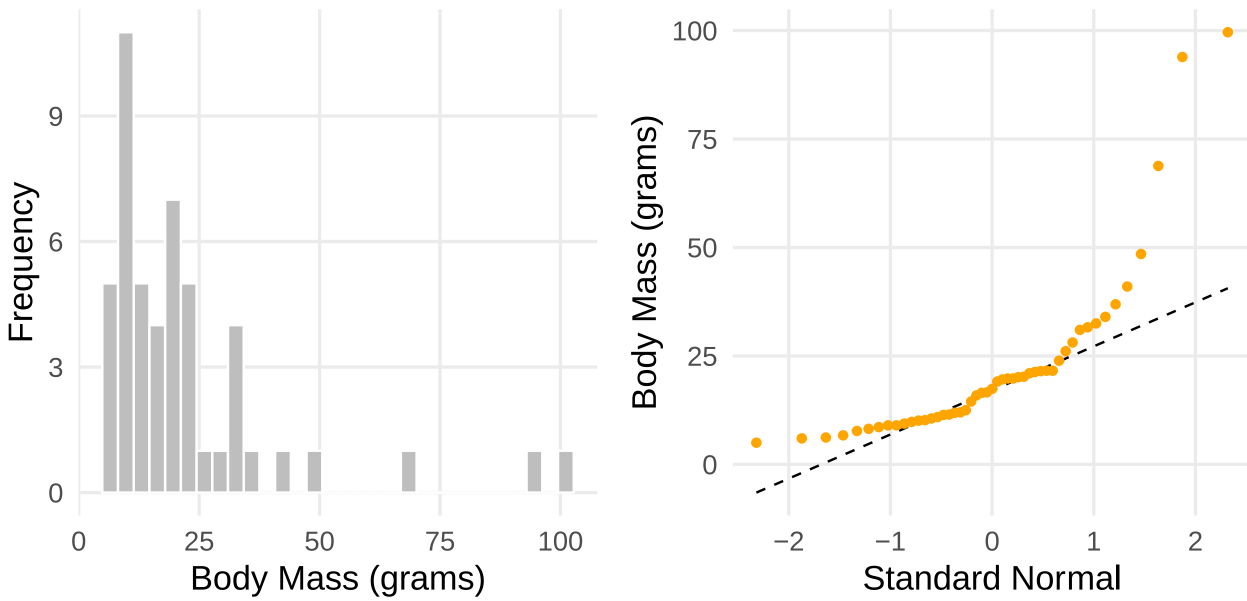


A very large outlier can be observed in both the histogram and QQ-plot. Although the QQ-plot doesn't appear strictly linear, we can compute the correlation between the observed values and normal-quantiles. The r value is positive, but not very strong.

```
##
##  Pearson's product-moment correlation
##
## data:  x and q
## t = 4.6082, df = 48, p-value = 3.018e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3256761 0.7210189
## sample estimates:
```

```
##       cor
## 0.5538194
```

b. Histogram and QQ-plot for the body-mass data (*excluding* the outlier). Even after removing the outlier, the data still seem non-normal.



In the absence of the outlier, the shape of the distribution is clearer. There is moderate positive skew, and the QQ-plot shows a non-linear relationship. It is likely that a correlational analysis is inappropriate, but in the spirit of repeating steps from **a.**, we can compute the r value for the QQ-plot.

The r value is larger than what was seen in **a.**, but it would be misleading to rely on this statistic when considering the graphical results.

```
##
##  Pearson's product-moment correlation
##
## data:  x and q
## t = 9.2975, df = 47, p-value = 3.184e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6768409 0.8856157
## sample estimates:
##       cor
## 0.8048535
```

c. Means and medians, when retaining or excluding the outlier.

	Mean	Median
With Outlier	30.74	18.25
Without Outlier	22.23	17.40

d. Standard deviations and MADs, when retaining or excluding the outlier.

	SD	MAD
With Outlier	63.30	11.79

	SD	MAD
Without Outlier	19.71	10.67