

MATH327-HW-01

January 26, 2017

Tim Hung | Probability with Stat Methods | Binghamton University

1 Chapter 1

1.1 1. An election will be held next week and, by polling a sample of the voting population, we are trying to predict whether the Republican or Democratic candidate will prevail. Which of the following methods of selection is likely to yield a representative sample?

(a) Poll all people of voting age attending a college basketball game.

(b) Poll all people of voting age leaving a fancy midtown restaurant.

(c) Obtain a copy of the voter registration list, randomly choose 100 names, and question them.

(d) Use the results of a television call-in poll, in which the station asked its listeners to call in and name their choice.

(e) Choose names from the telephone directory and call these people

1.2 2. The approach used in Problem 1(e) led to a disastrous prediction in the 1936 presidential election, in which Franklin Roosevelt defeated Alfred Landon by a landslide. A Landon victory had been predicted by the Literary Digest. The magazine based its prediction on the preferences of a sample of voters chosen from lists of automobile and telephone owners.

(a) Why do you think the Literary Digest's prediction was so far off ?

In 1936, telephone and automobile owners were probably limited to wealthy and well off voters. Thus, the sample would be more representative of richer voters.

(b) Has anything changed between 1936 and now that would make you believe that the approach used by the Literary Digest would work better today?

More people can afford cars, and most people have cellphones so it would be a better sample today.

1.3 5. A university plans on conducting a survey of its recent graduates to determine information on their yearly salaries. It randomly selected 200 recent graduates and sent them questionnaires dealing with their present jobs. Of these 200, however, only 86 were returned. Suppose that the average of the yearly salaries reported was \$75,000.

- (a) Would the university be correct in thinking that \$75,000 was a good approximation to the average salary level of all of its graduates? Explain the reasoning behind your answer.

No, alumni who were reluctant to respond are more likely to be unhappy with their jobs/salaries. This will result in right-skewed data.

- (b) If your answer to part (a) is no, can you think of any set of conditions relating to the group that returned questionnaires for which it would be a good approximation?

Alumni who have a good relationship with the university.

1.4 10. Based on Graunt's mortality table:

- (a) What proportion of people survived to age 6?

64%

- (b) What proportion survived to age 46?

10%

- (c) What proportion died between the ages of 6 and 36?

48%

2 Chapter 2

2.1 1. The following is a sample of prices, rounded to the nearest cent, charged per gallon of standard unleaded gasoline in the San Francisco Bay area in June 1997.

3.88, 3.90, 3.93, 3.90, 3.93, 3.96, 3.88, 3.94, 3.96, 3.88, 3.94, 3.99, 3.98

2.2 Represent these data in

- (a) a frequency table

- (b) a relative frequency line graph

```
In [42]: prices <- c(3.88, 3.90, 3.93, 3.90, 3.93, 3.96, 3.88,
                    3.94, 3.96, 3.88, 3.94, 3.99, 3.98)
freq <- table(prices)
cat("Frequency Table of Gasoline Prices\n")
print(freq)

cat("\n\nRelative Frequency Table of Gasoline Prices\n")
relFreq <- freq/length(prices)
print(relFreq)
```

```
plot(relFreq, ylab="Relative Frequency", xlab="Prices")
```

Frequency Table of Gasoline Prices

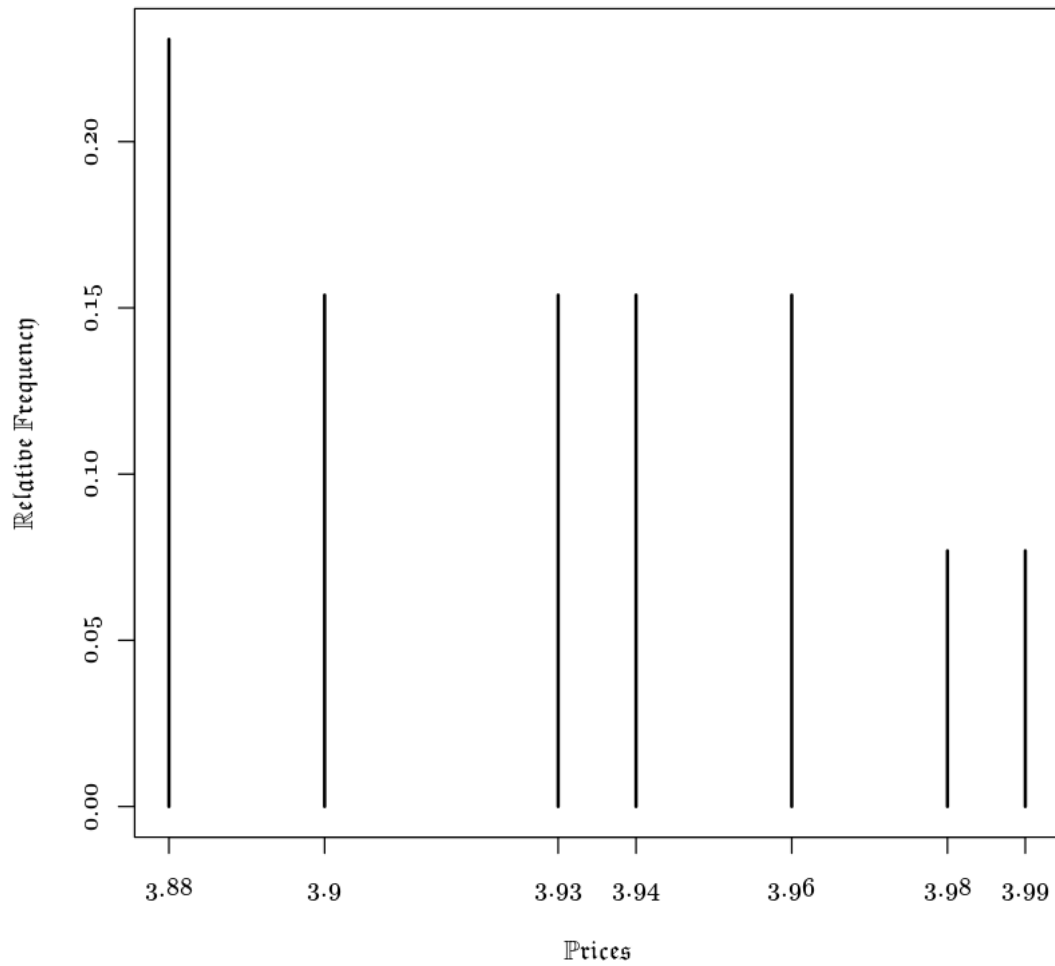
prices

3.88	3.9	3.93	3.94	3.96	3.98	3.99
3	2	2	2	2	1	1

Relative Frequency Table of Gasoline Prices

prices

3.88	3.9	3.93	3.94	3.96	3.98	3.99
0.23076923	0.15384615	0.15384615	0.15384615	0.15384615	0.07692308	0.07692308

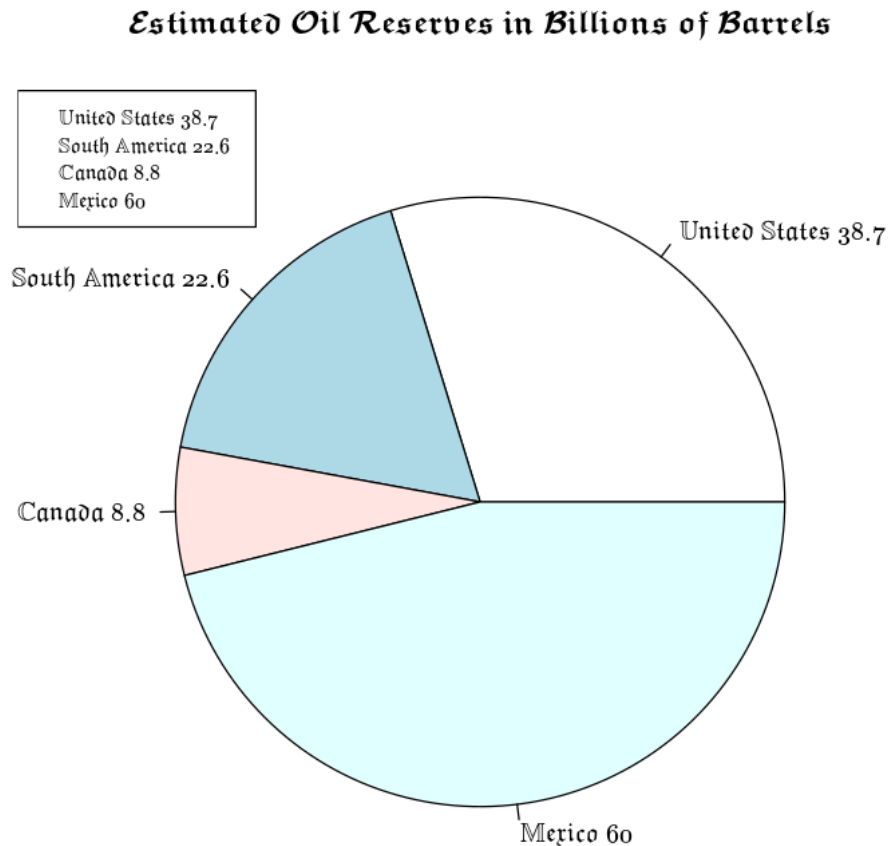


2.3 3. The following are the estimated oil reserves, in billions of barrels, for four regions in the Western Hemisphere.

United States 38.7
South America 22.6
Canada 8.8
Mexico 60.0

2.4 Represent these data in a pie chart

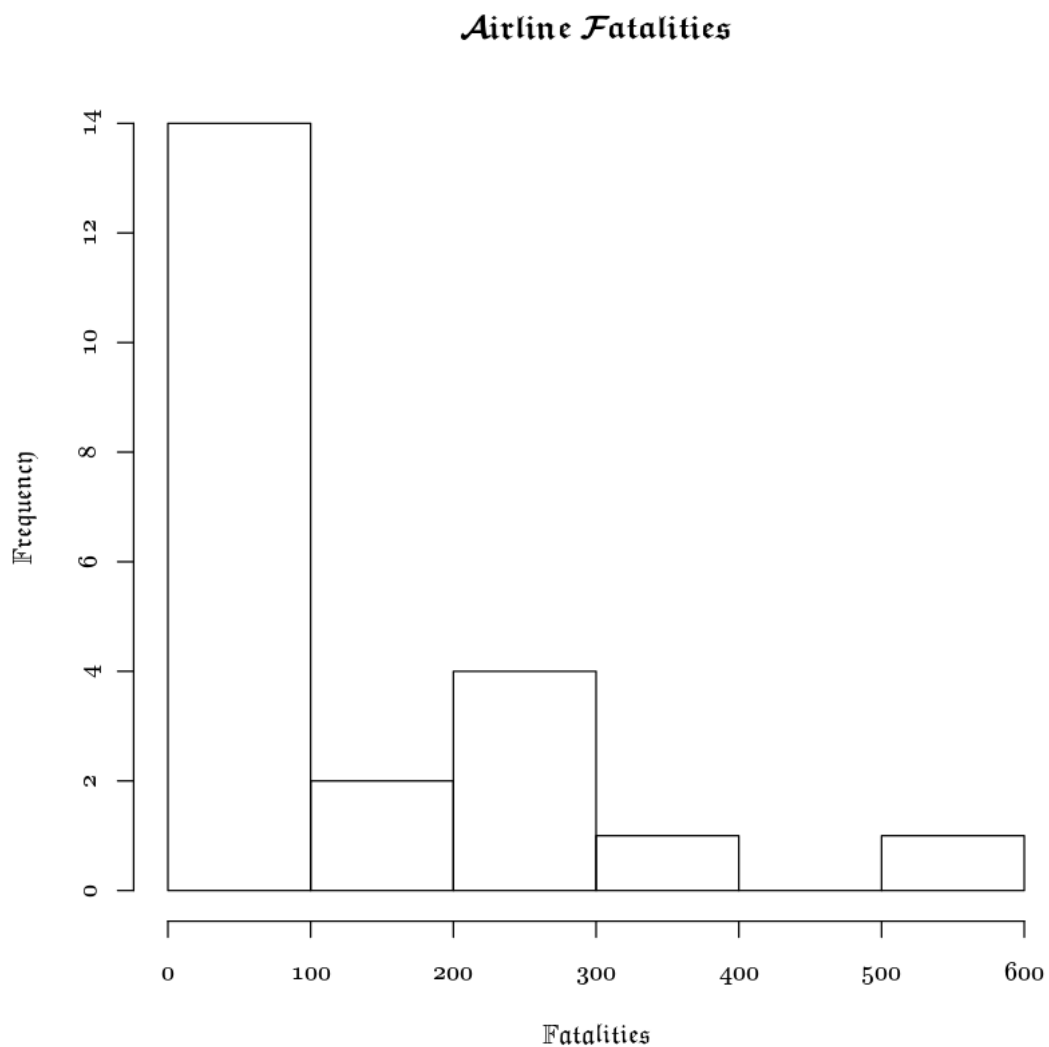
```
In [4]: regions <- c("United States", "South America", "Canada", "Mexico");  
data <- c(38.7, 22.6, 8.8, 60.0);  
regions <- paste(regions, data);  
pie(data, regions, main="Estimated Oil Reserves in Billions of Barrels");  
legend("topleft",regions, cex=0.8);
```



2.5 7. (Use the table from Problem 6.)

- (a) Represent the number of yearly airline fatalities in a histogram.
- (b) Represent the number of yearly airline fatalities in a stem and leaf plot.
- (c) Find the sample mean of the number of yearly airline fatalities.
- (d) Find the sample median of the number of yearly airline fatalities.
- (e) Find the sample standard deviation of the number of yearly airline fatalities

```
In [12]: fatalities <- c(197, 5, 231, 285, 278, 39, 62, 33, 1, 239, 166,  
                        342, 3, 1, 12, 89, 531, 0, 22, 13, 22, 50)  
hist(fatalities, main="Airline Fatalities", xlab="Fatalities")
```



```
In [14]: fatalities <- c(197, 5, 231, 285, 278, 39, 62, 33, 1, 239, 166,
                        342, 3, 1, 12, 89, 531, 0, 22, 13, 22, 50)
      stem(fatalities, scale=2)
```

The decimal point is 2 digit(s) to the right of the |

```
0 | 00001112234569
1 | 7
2 | 03489
3 | 4
4 |
5 | 3
```

```
In [23]: fatalities <- c(197, 5, 231, 285, 278, 39, 62, 33, 1, 239, 166, 342, 3, 1, 12, 89, 531,
                        cat("Mean: ", mean(fatalities))
                        cat("\n\nMedian: ", median(fatalities))
                        cat("\n\nStandard Deviation: ", sd(fatalities))
```

Mean: 119.1364

Median: 44.5

Standard Deviation: 144.785

2.6 8. The sample mean of the weights of the adult women of town A is larger than the sample mean of the weights of the adult women of town B. Moreover, the sample mean of the weights of the adult men of town A is larger than the sample mean of the weights of the adult men of town B. Can we conclude that the sample mean of the weights of the adults of town A is larger than the sample mean of the weights of the adults of town B? Explain your answer.

No, because we have no data on how many men and women there are for each town.

2.7 13. The following are the percentages of ash content in 12 samples of coal found in close proximity:

9.2, 14.1, 9.8, 12.4, 16.0, 12.6, 22.7, 18.9, 21.0, 14.5, 20.4, 16.9

2.8 Find the

- (a) sample mean
- (b) sample standard deviation of these percentages

```
In [13]: ashContent = c(9.2, 14.1, 9.8, 12.4, 16.0, 12.6, 22.7,
                        18.9, 21.0, 14.5, 20.4, 16.9)
cat("\nMean: ", mean(ashContent))
cat("\n\nStandard Deviation: ", sd(ashContent))
```

Mean: 15.70833

Standard Deviation: 4.39534

2.9 16. The following data represent the lifetimes (in hours) of a sample of 40 transistors:

112, 121, 126, 108, 141, 104, 136, 134
 121, 118, 143, 116, 108, 122, 127, 140
 113, 117, 126, 130, 134, 120, 131, 133
 118, 125, 151, 147, 137, 140, 132, 119
 110, 124, 132, 152, 135, 130, 136, 128

(a) Determine the sample mean, median, and mode.

(b) Give a cumulative relative frequency plot of these data.

```
In [40]: lifetimes <- c(112, 121, 126, 108, 141, 104, 136, 134,
                        121, 118, 143, 116, 108, 122, 127, 140,
                        113, 117, 126, 130, 134, 120, 131, 133,
                        118, 125, 151, 147, 137, 140, 132, 119,
                        110, 124, 132, 152, 135, 130, 136, 128)
cat("Mean: ", mean(lifetimes))
cat("\n\nMedian: ", median(lifetimes))
cat("\n\nFrequency Table")
sort(table(lifetimes))
cat("\nMode: 108, 118, 121, 126, 130, 132, 134, 136, 140 all occur twice")
```

Mean: 127.425

Median: 127.5

Frequency Table

```
lifetimes
104 110 112 113 116 117 119 120 122 124 125 127 128 131 133 135 137 141 143 147
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
151 152 108 118 121 126 130 132 134 136 140
  1   1   2   2   2   2   2   2   2   2   2
```

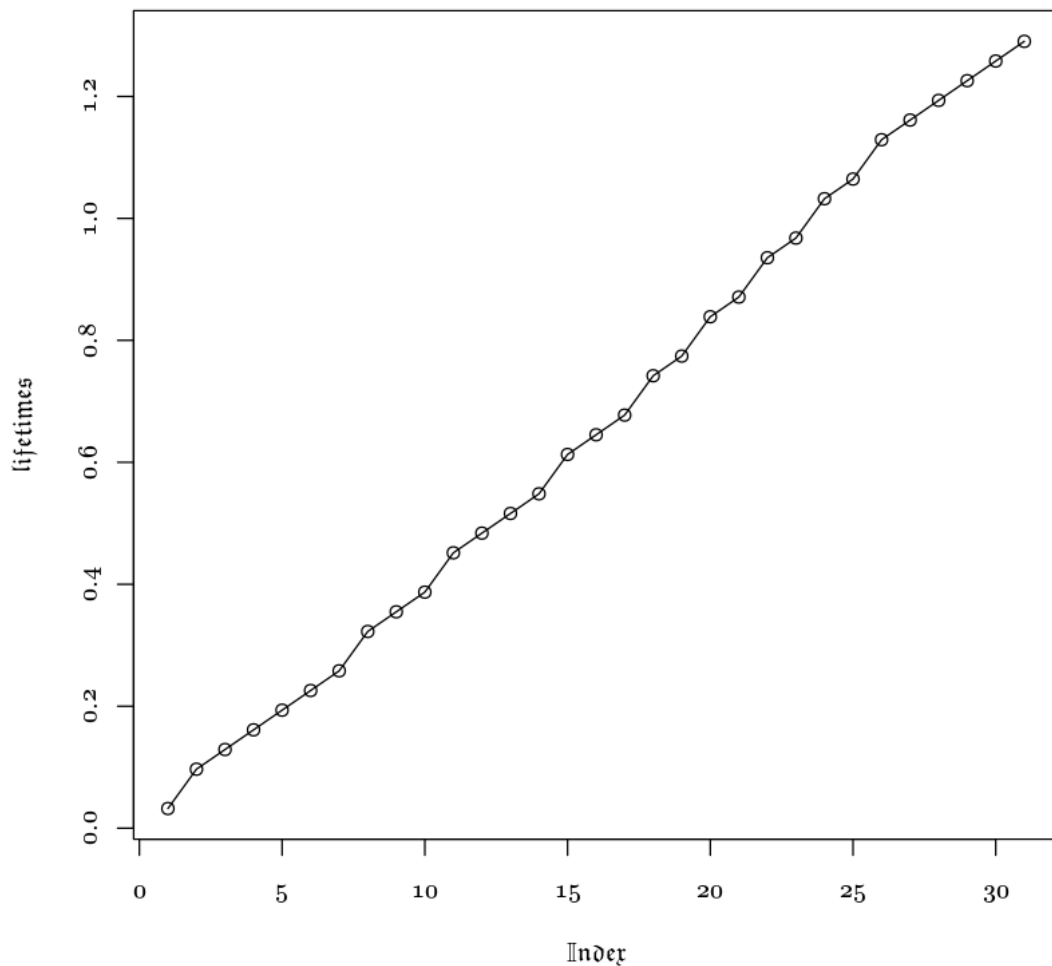
Mode: 108, 118, 121, 126, 130, 132, 134, 136, 140 all occur twice

```

In [50]: lifetimes <- c(112, 121, 126, 108, 141, 104, 136, 134,
                        121, 118, 143, 116, 108, 122, 127, 140,
                        113, 117, 126, 130, 134, 120, 131, 133,
                        118, 125, 151, 147, 137, 140, 132, 119,
                        110, 124, 132, 152, 135, 130, 136, 128)

lifetimes <- sort(lifetimes)
lifetimes <- table(lifetimes)
lifetimes <- cumsum(lifetimes)/length(lifetimes)
plot(lifetimes, type="o")

```



2.10 20. Find the quartiles of the following ages at death as given in obituaries of the New York Times in the 2 weeks preceding 1 August 2013.

92, 90, 92, 74, 69, 80, 94, 98, 65, 96, 84, 69, 86,

91, 88, 74, 97, 85, 88, 68, 77, 94, 88, 65, 76, 75,
60, 69, 97, 92, 85, 70, 80, 93, 91, 68, 82, 78, 89

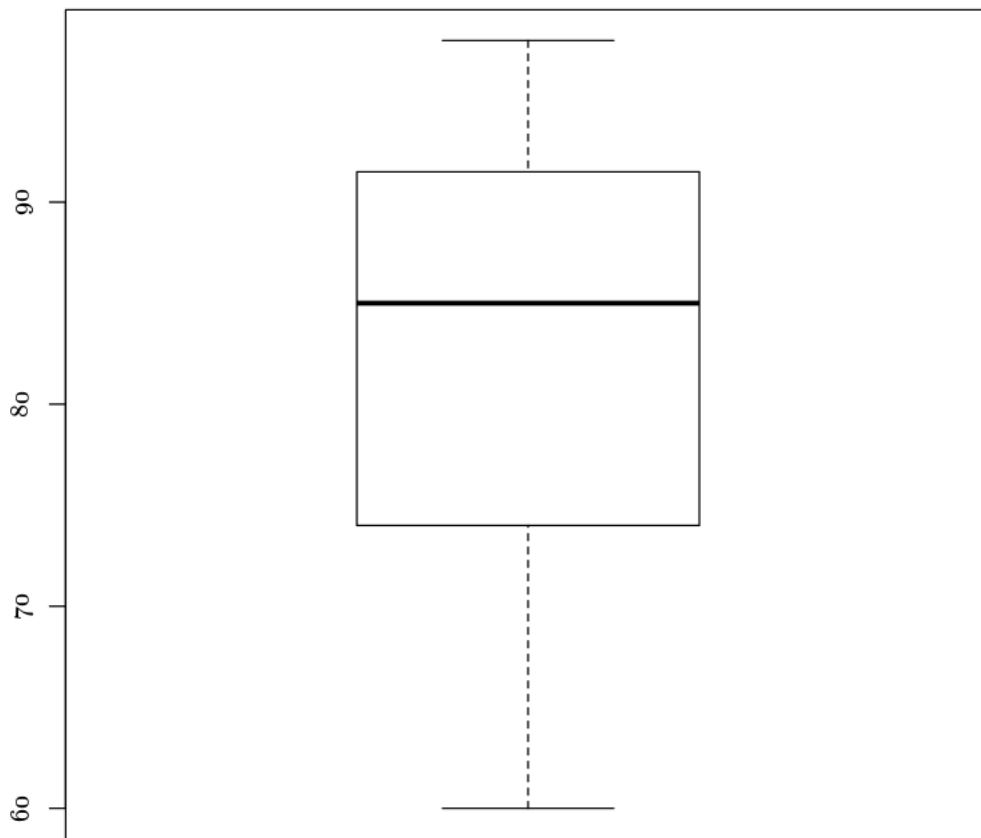
2.11 23. Represent the data of Problem 20 in a box plot.

```
In [20]: death <- c(92, 90, 92, 74, 69, 80, 94, 98, 65, 96, 84, 69, 86,  
                    91, 88, 74, 97, 85, 88, 68, 77, 94, 88, 65, 76, 75,  
                    60, 69, 97, 92, 85, 70, 80, 93, 91, 68, 82, 78, 89)  
cat("\nQuartiles of ages of death from given data:")  
quantile(death)  
cat("\nBoxplot of ages of death from given data:")  
boxplot(death)
```

Quartiles of ages of death from given data:

0\%	60 25\%	74 50\%	85 75\%	91.5 100\%	98
-----	---------	---------	---------	------------	----

Boxplot of ages of death from given data:



2.12 26. The following are the grade point averages of 30 students recently admitted to the graduate program in the Department of Industrial Engineering and Operations Research at the University of California at Berkeley.

3.46, 3.72, 3.95, 3.55, 3.62, 3.80, 3.86, 3.71, 3.56, 3.49, 3.96, 3.90, 3.70, 3.61, 3.72, 3.65,

(a) Represent the preceding data in a stem and leaf plot.

(b) Calculate the sample standard deviation s

```
In [7]: gpas <- c(3.46, 3.72, 3.95, 3.55, 3.62, 3.80, 3.86, 3.71,
                 3.56, 3.49, 3.96, 3.90, 3.70, 3.61, 3.72, 3.65,
                 3.48, 3.87, 3.82, 3.91, 3.69, 3.67, 3.72, 3.66,
                 3.79, 3.75, 3.93, 3.74, 3.50, 3.83)
```

```
stem(gpas)
cat("\nStandard Deviation: ", sd(gpas))
```

The decimal point is 1 digit(s) to the left of the |

```
34 | 689
35 | 056
36 | 125679
37 | 01222459
38 | 02367
39 | 01356
```

Standard Deviation: 0.1456717

2.13 33. A random group of 12 high school juniors were asked to estimate the average number of hours they study each week. The following give these hours along with the student's grade point average.

Hours	GPA
6	2.8
14	3.2
3	3.1
22	3.6
9	3.0
11	3.3
12	3.4
5	2.7
18	3.1
24	3.8
15	3.0
17	3.9

2.14 Find the sample correlation coefficient between hours reported and GPA.

```
In [2]: hours <- c(6, 14, 3, 22, 9, 11, 12, 5, 18, 24, 15, 17)
      gpa <- c(2.8, 3.2, 3.1, 3.6, 3.0, 3.3, 3.4, 2.7, 3.1, 3.8, 3.0, 3.9)
      cat("Correlation Coefficient: ", cor(hours, gpa))
```

Correlation Coefficient: 0.7428882

2.15 36. In a study of children in grades 2 through 4, a researcher gave each student a reading test. When looking at the resulting data the researcher noted a positive correlation between a student's reading test score and height. The researcher concluded that taller children read better because they can more easily see the blackboard. What do you think?

The conclusion is naive because the 4th graders are probably taller than the 2nd and 3rd graders

2.16 37. A recent study yielded a positive correlation between breast-fed babies and scores on a vocabulary test taken at age 6. Discuss the potential difficulties in interpreting the results of this study.

Correlation does not imply causation.