1 - Practice Problems

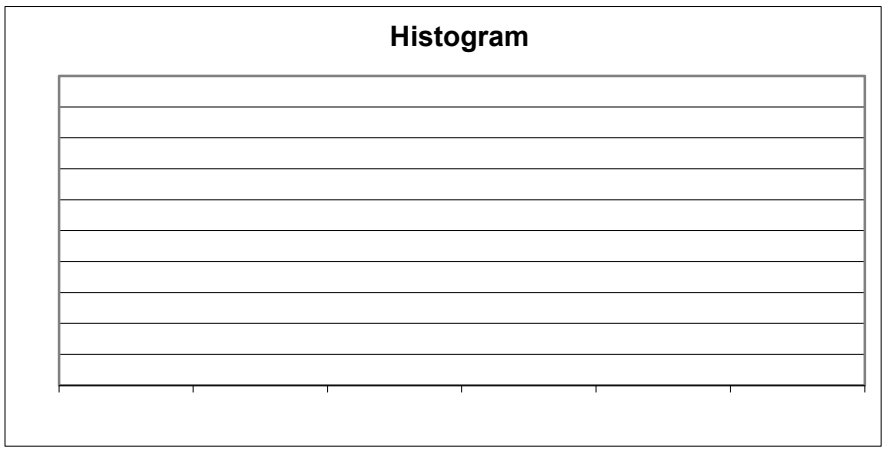
Thickness (")
0.754
0.735
0.754
0.748
0.740
0.752
0.747
0.740
0.751
0.741
0.740
0.742
0.748
0.732
0.750
0.747
0.750
0.752

1. To test new equipment for making ¾" plywood, measurements of thickness were obtained.

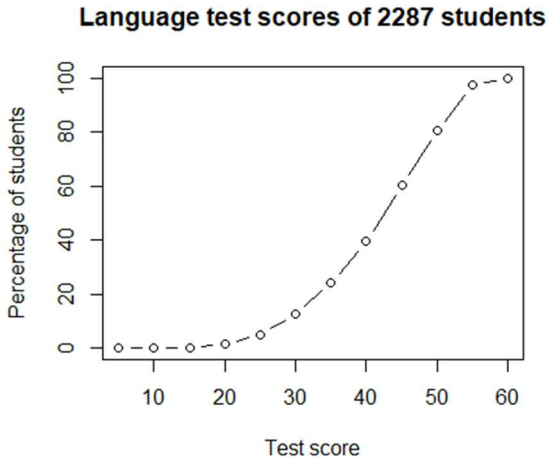
   a. Create a frequency table with five classes:

Class Limits	Frequency	Relative Frequency

   b. Sketch a histogram (or use R):



2. The test scores of 2287 Grade 8 students in the Netherlands are represented by the ogive below.



   a. What proportion of students had scores below 50?
   b. What proportion of students had scores between 30 and 40?
   c. What range of scores represents the top 25% of the class?

3. The times between Old Faithful eruptions, in minutes, are given in the stem plot below.

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

   4 | 3
   4 | 55566666777788899999
   5 | 000001111112222233333333444444444
   5 | 55555566667778888999999999
   6 | 00000022223334444
   6 | 555667899
   7 | 00001111123333333444444
   7 | 55555555666666666777777777777888888888888888999999999
   8 | 000000001111111111111222222222222333333333333334444444444
   8 | 55555566666677888888999
   9 | 00000012334
   9 | 6

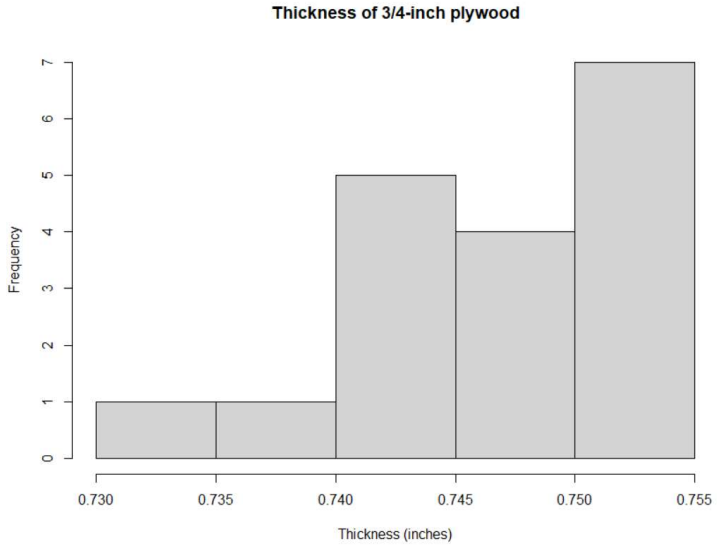
   a. What is the range of the data?
   b. Describe the distribution of the data.
   c. Which 10-minute range contains the largest number of waiting times between eruptions?

Answers

1a

Class Limits	Frequency	Rel. Freq
0.730 - 0.735	1	0.055556
0.735 - 0.740	1	0.055556
0.740 - 0.745	5	0.277778
0.745 - 0.750	4	0.222222
0.750 - 0.755	7	0.388889

1b



2. a. 80%
   b. 25%
   c. 48-60

3. a. 43-96 min
   b. bimodal, with a lot of eruptions between 50 and 55 minutes apart, and a lot between 75 and 80 minutes apart
   c. 75-85 min

2 – Practice Problems

1. To test new equipment for making ¾" plywood, measurements of thickness were obtained.

   Calculate the range and standard deviation of the data.

Thickness (")
0.754
0.735
0.754
0.748
0.740
0.752
0.747
0.740
0.751
0.741
0.740
0.742
0.748
0.732
0.750
0.747
0.750
0.752

2. Calculate the mean and standard deviation of a sample of 350 steel rods (use R or your calculator's built in functions):

Diameter (mm)	Frequency
10.00	40
10.01	75
10.02	100
10.03	90
10.04	45

3. The following set of data is the *maximum wind velocity* measured at Vancouver International Airport over a one-year period. The data are not ordered by month as given.

    a. Summarize the data set using at least four common sample statistics.
    b. Imagine you are considering the possible wind effects on a new multi-storey hotel to be built in Richmond. Let's assume there is a very low probability that maximum wind velocity will be greater than 3 standard deviations above the mean.
        i. What wind velocity is 3 sample standard deviations above the sample mean?
        ii. What is the probability that maximum wind velocity will exceed this value, assuming that maximum wind velocities follow a normal distribution (Empirical Rule)?

Wind Velocity(km/h)
39.16
115.22
21.11
43.04
37.01
30.1
19.33
70.54
76.11
24.35
109.58
50.27

4. The two sets of sample data (Site A and Site B) are measurements of the drainage rate of the soil in two different building sites (liters per m^2 per s).

Site A	Site B
2.99	1.02
4.75	3.56
8.79	3.5
5.59	3.45
2.32	4.5
1.9	13.6
	4.5
	2.3
	3.5
	2.6
	3.31
	3.1

   a. Based on relevant sample statistics, which of the sites has the most "relatively-variable" drainage? Conclude which site you would be more likely to recommend constructing a building on.
   b. For Site B data set, assume that the drainage rates are known to be distributed in a bell-shaped curve. Could either (or both) the max and min values could be discounted as being too extreme?
   c. If you were to exclude the minimum data point from the data set for site A, what would be the effect on the mean, median and standard deviation?

5. The following soil density measurements were made on a strip mall development site in Richmond before and after compacting the sand.

   Make a graphical comparison and calculate summary statistics to describe the differences before and after. Imagine you are trying to describe how the compacting of the soil has improved the site for building purposes to the client. Use 3 or 4 sample statistics most suitable for making comparisons.

Before Compacting (g/cm^3)	After Compacting (g/cm^3)
0.1	15.46
1.03	17.54
2.69	18.66
3.54	19.03
4.33	19.31
13.41	20.08
14.04	21.2
14.04	21.29
14.83	21.4
53.26	24.68

6. Hank and Pete work in a fabrication facility. Records indicate that Hank completes an average of 62 items per shift, with a standard deviation of 4.2 items. Pete completes an average of 59 items per shift, with a standard deviation of 4.1 items. Which of the two would you consider to be the more consistent worker? Explain.

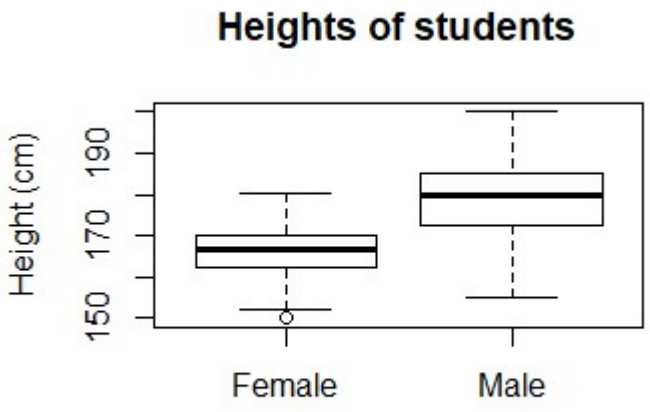
7. Suppose a lumber mill ships 4x4's with a mean moisture content of $\mu = 18\%$ and $\sigma = 0.5\%$. A histogram reveals that the data is bell-shaped.

   a. Use Chebyshev's Rule to determine the minimum percentage of scores that lie between 17% and 19%.
   b. Repeat question a. using the Empirical Rule.
   c. Use the Empirical Rule to estimate the percentage of values greater than 19.5%.
   d. Use the Empirical Rule to estimate the percentage of values between 18% and 18.5%.
   e. Building specifications require that the moisture content be under 19% to be closed in. Use the Empirical Rule to estimate the percentage of 4x4's that are suitable to be closed in.

8. Human body temperatures have a mean of 98.20°F and a standard deviation of 0.62°F. Determine whether each of the following temperatures is usual or unusual.
   a. 101.00°F
   b. 96.90°F
   c. 96.98°F

9. The heights of 237 statistics students are represented in the boxplot below. Complete the sentence below to make one observation about the relative heights of male and female students.

   "x% of female students are shorter than…".



Answers

1.

Range	0.022
Standard Dev	0.006560179

2.

Stdev	0.0120078	mm

3a.
The mean maximum monthly wind velocity is 52.99 km/h with a standard deviation of 32.98 km/h. There is considerable variability seen in the standard deviation. Also, the range of values is 95.89 km/h. As a result, it is difficult to predict the maximum wind velocity during a month.

3b.
mean + 3s = 52.98 + 3(32.98) = 151.92 km/h

3c.
It is not reasonable to apply the Empirical Rule since our data does not reflect a normal distribution. If we assume that the population is normal anyway, the probability of being more than 3 standard deviation above the mean is:

$$(100 - 99.7)/2 = 0.15\%$$

4a.
Site B has a larger coefficient of variation, larger standard deviation, and larger range so it is relatively more variable than Site A. I would recommend Site A as its mean drainage is higher, and it is more consistent.

4b.
xbar - 3sigma = 4.07 - 3(3.14) = -5.3
xbar - 2sigma = 4.07 - 2(3.14) = - 2.2
xbar + 2sigma = 4.07 + 2(3.14) = 10.3
xbar + 3sigma = 4.07 + 3(3.14) = 13.5 $L/m^2/s$
The maximum value is more than 3 standard deviations above the mean and thus is a very rare value.

4c.
If we remove the minimum from Site A the mean and median should increase and the standard deviation should decrease.

5.

Before Compaction (g/cm3)		After Compaction (g/cm3)	
Mean	12.127	Mean	19.865
Median	8.87	Median	19.695
Mode	14.04	Mode	#N/A
Standard Deviation	15.63321	Standard Deviation	2.50923738
Skewness	2.347438	Skewness	0.17569909
Range	53.16	Range	9.22
Minimum	0.1	Minimum	15.46
Maximum	53.26	Maximum	24.68
Sum	121.27	Sum	198.65
Count	10	Count	10

The mean and median soil density have increased after compaction, so overall compacting the soil increases the density. More importantly, by comparing the Ranges and the Coefficients of Variation, we can see that there is much less variability in the soil density after compaction.

6.

	Mean	St. Dev.	CV	
Hank	62	4.2	6.774194	%
Pete	59	4.1	6.949153	%

Comparing the CV's we see that Hank is a more consistent worker even though his standard deviation is slightly higher than Pete's.

7a.
17% is 2σ below the mean and 19 is 2σ above the mean. Applying Chebyshev with $k = 2$, we have at least 75% of values between 17% and 19%

7b.
95%

7c.
19.5% is 3σ above the mean. 99.7% are within 3σ of the mean when the data is normally distributed. 0.3% of the data is outside the 3σ range. We are only interested in the upper tail of the distribution so 0.3%/2 = 0.15%

7d.
We know that 68% of the values are between -1σ and 1σ. 34% of the values should then be located between the mean and 1σ.

7e.
19% is 2σ above the mean. We want the entire bell except the part that is more than 2σ.
5%/2 = 2.5%

8a.
$Z = 4.52$; unusual

8b.
$Z = -2.10$; unusual

8c.
$Z = -1.97$; usual

9.
50% of the male students are taller than the tallest female student, who is 180 cm tall.
(Other answers are possible.)