Chapter 3 Review – Summarizing Univariate Data

MDM4U

David Chen

Section 3.1: Shapes of Distributions

1) Using the following data:
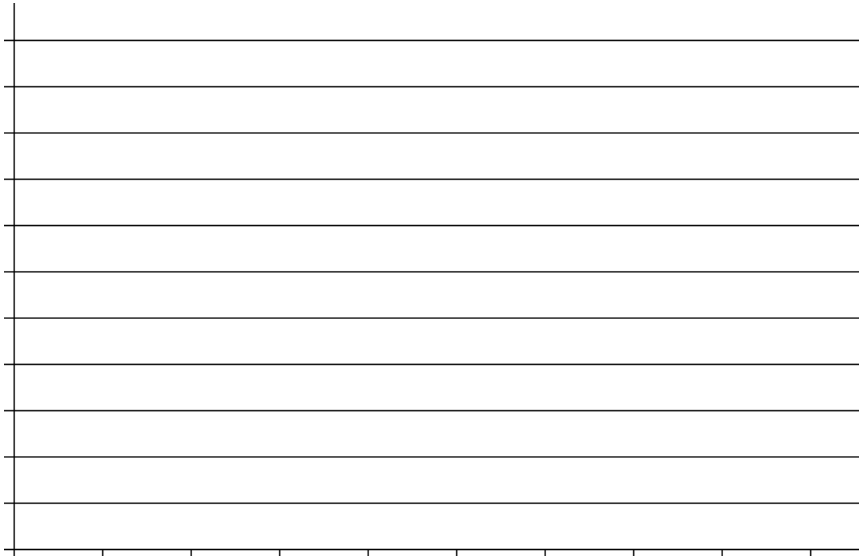
$\{5, 7, 9, 10, 11, 11, 12, 12, 12, 13, 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 15, 15, 15, 15, 16, 16, 17, 18, 20\}$

a) Calculate a bin width that would form six uniform intervals

b) Calculate the starting and end point for each of the five intervals. Then complete the frequency distribution.

Interval	Frequency

c) Create an appropriate histogram.

2) What shape of distribution would you expect for each scenario listed below?

    a) Mr. Hammond's class did very well on the test

    b) Mr. Fusilli's class did very poorly on the test

    c) Dave and his friends have a similar number of speeding tickets

    d) On a range of A to F, most students in Mr. Lee's class had a C+

    e) Billy's scores on a video game are usually either very low or very high

Section 3.2: Measures of Central Tendency

3) Determine the mean, median, and mode of the following sets of sample data. You may use technology or calculate by hand.

a) waiting at a red light, in sec. {20, 28, 33, 33, 34, 35, 36, 37, 39, 40, 40, 45, 47, 53, 54}

b) marks on a sample of Jill's math tests {68, 72, 86, 63, 75, 80}

c) number of hamburgers sold at "Bela's Burgers" daily {140, 113, 97, 136, 85, 93, 105}

4) The scores and their percent of the final grade for a statistics student are given. What is the student's mean score? Calculate by hand and then use technology to check your answer.

	Score (x)	Percent of Final Grade (w)
Homework	85	15%
Quiz	80	10%
Quiz	92	10%
Quiz	76	10%
Project	100	15%
Speech	90	15%
Final Exam	93	25%

5) Jordan's term mark is 86. The term mark counts for 65% of the final mark. What mark must Jordan achieve on the final exam to earn a final mark of:

a) 80   **b)** 90?

6) Mr. Jensen collected data from his entire class on the hours spent studying the night before a statistics test. Calculate, by hand, the mean number of hours spent studying for this population data. Use technology to check your answer.

Hours Spent Studying	Number of Students
0	4
1	5
2	9
3	6
4	3
5	2
6	1

7) The personnel manager for Statsville Marketing Limited considers five criteria when interviewing a job applicant. The manager gives each applicant a score between 1 and 5 in each category, with 5 as the highest score. Each category has a weighting between 1 and 3. The following table lists a recent applicant's scores and the company's weighting factors.

Criterion	Score (x)	Weighting Factor (w)
Education	4	2
Job Experience	2	2
Interpersonal Skills	5	3
Communication Skills	5	3
References	4	1

Determine the weighted mean score for this job applicant by hand. Then check your answer using technology:

8) A group of children were asked how many hours a day they spend watching television. The results were then reported in a frequency distribution. Calculate the mean number of hours for this distribution by hand and then check your answer using technology.

Number of Hours	Frequency, f		
0---1	1		
1---2	4		
2---3	7		
3---4	3		
4---5	2		
5---6	1		

Section 3.3: Measures of Spread

9) Listed below are the weights (in pounds) of 25 babies born on Sept. 13 at "Mothers' Hospital":

5.7, 5.9, 6.2, 6.3, 6.9, 7.1, 7.2, 7.3, 7.4, 7.4, 8.0, 8.2, 8.4, 9.1
9.3, 9.4, 9.5, 9.9, 10.4, 10.4, 10.6, 11.7, 12.3, 12.6, 13.8

Use technology to help you calculate the following:

a) Find the quartile values

b) What is the interquartile range?

c) What percent of baby weights are represented by the upper quartile?

d) What is the standard deviation of the sample of birth weights?

10) The number of vacation days used by a sample of 20 employees in a recent year are listed below:

0, 0, 1, 2, 2, 2, 3, 3, 3, 4, 5, 5, 5, 6, 6, 7, 7, 8, 9, 10

Use technology to help you calculate the following:

a) Find the quartile values

b) What is the interquartile range?

c) What percent of data is represented by the interquartile range?

d) What is the standard deviation of the sample of vacation days used?

11) Calculate the sample mean and standard deviation for the following data set by hand:

11, 10, 8, 4, 6, 7, 11, 6, 11, 7

12) The 2011 roster of the Dallas Cowboys football team included 7 defensive line---men. Their weights (in pounds) were 321, 285, 300, 285, 286, 293, and 298. Calculate the population mean and standard deviation for this data by hand. Use your calculator to check your answer.

13) You collect a random sample of the number of children per household in a region. The results are shown at the left. Find the standard deviation of the data set using your graphing calculator.

Number of Children (x)	Frequency (f)
0	10
1	19
2	7
3	7
4	2
5	1
6	4

14) For one week, Matthew and Jordi record how many calls they successfully answer each day while doing tech support. They find that Matthew answers 14, 18, 22, 16, and 8 calls, while Jordi answers 16, 17, 16, 15, and 14 calls. Use this sample data to answer the following questions:

a) Who is the more effective employee? (calculate mean number of calls)

b) Who is the more consistent employee? (calculate the standard deviation)

Section 3.4: The Normal Distribution

15) Fill in the blanks:

Normal distributions are symmetrical and approach _____ at the extremes. Of the data, _____% is within one standard deviation of the mean, _____% is within two standard deviations, and _____% is within three standard deviations of the mean. The area under any normal curve is _____.

16) The thickness of an automobile's windshield can be represented by the normal distribution $X \sim N(14.6, 0.02^2)$
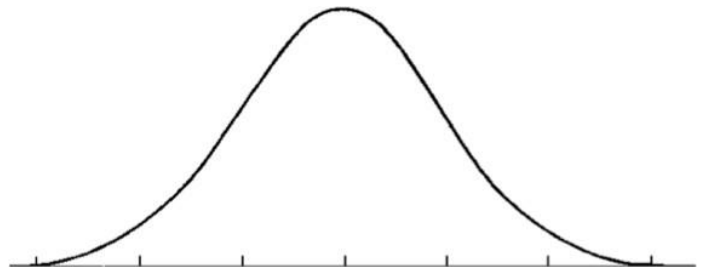
a) Sketch a normal curve for this distribution of vocabulary scores. Label the points that are 1, 2, and 3 standard deviations from the mean.
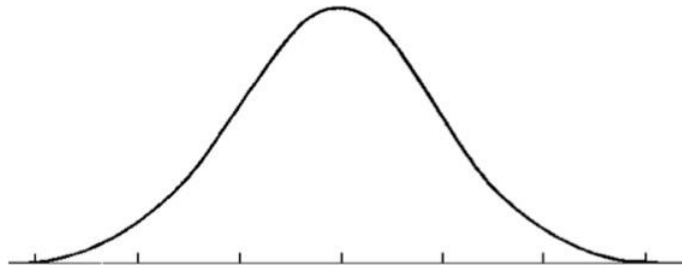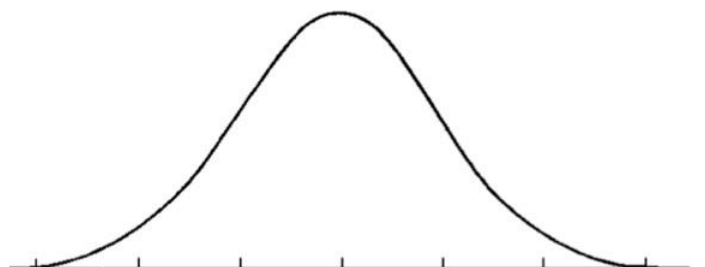
b) What range of thickness does 68% of car windshields fall between?

c) Draw a diagram and state the percent of data that represents each of the following:
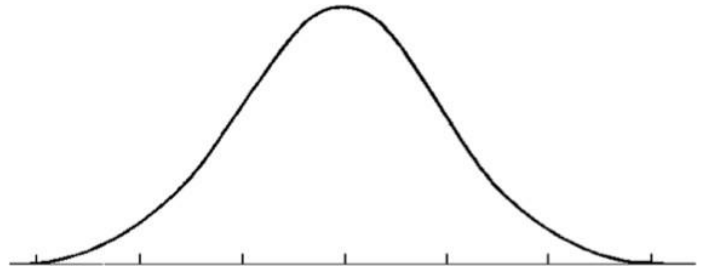
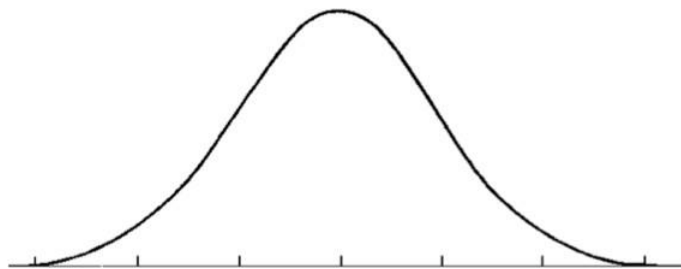**i)** less than 14.58 mm

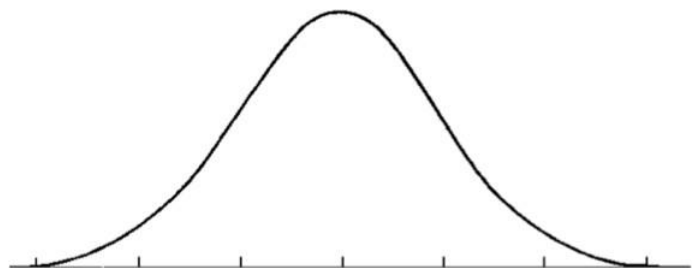**ii)** between 14.58 mm and 14.6 mm

iii) more than 14.64 mm

17) If the mass of 35 dogs in your neighbourhood were represented by the normal distribution
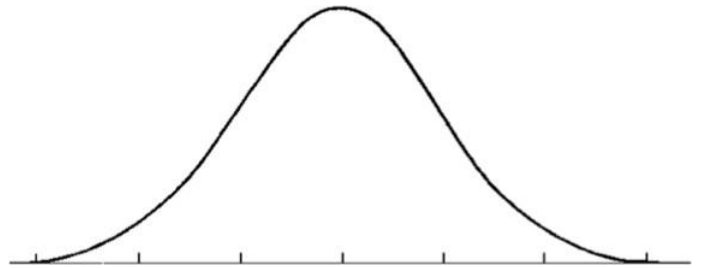$X\sim N(11.2, 2.8^2)$

a) Sketch a normal curve for this distribution of vocabulary scores. Label the points that are 1, 2, and 3 standard deviations from the mean.

b) How many dogs would you expect to have a mass 11.2 kg or less?

c) How many dogs would you expect to have a mass between 8.4 kg and 14 kg?

Section 3.5: Applying the Normal Distribution

18) Calculate the z---score for each x---value, correct to one decimal place, given the distribution $X \sim N(8, 1.5^2)$

a) $x = 4.8$ **b)** $x = 9.3$ **c)** $x = 2.9$

19) Given a normally distributed data set whose mean is 30 and whose standard deviation is 5, what values of x would each of the following z---scores have?

a) z = 1.00   **b)** z = 0.48

20) The grades on a test follow a normal distribution with a mean mark of 60 and a standard deviation of 18.

a) Jean---Paul's mark is 77; what percentile is he in?

b) If you want to be in the 90th percentile, what mark must you get on the test?

20) Bubba Watson led the PGA tour in driving distance during the 2015 season. The length of his drives followed a normal distribution with a mean of 317 yards and a standard deviation of 10 yards. Find the percent of his drives that you would expect to have been:

a) greater than 300 yards

b) less than than 305 yards

c) between 320 and 330 yards

21) For the distribution $N(3, 0.74^2)$, determine the percent of the data that is within the given interval.

a) $X > 2.44$

b) $1.8 < X < 2.3$

c) $X < 1.91$

22) Researchers have observed that regular smokers have an average lifespan that is normally distributed and is 68 years with a standard deviation of 10 years. What percent of smokers will live beyond age 76?

Section 3.6: Confidence Intervals

23) The most popular name for leisure boats is "Serenity". A Survey of 200,000 boat owners found that 12% of their boats were names "Serenity". Determine a 90% confidence interval for the proportion of leisure boats that have this name.

24) A study of 55 patients with back pain reported that the mean duration of pain was 17.6 months. It is known that the standard deviation of duration of back pain is 5.1 months. Assuming that the duration of this problem is normally distributed in the population; determine a 99% confidence interval for the mean duration of back pain in the population.