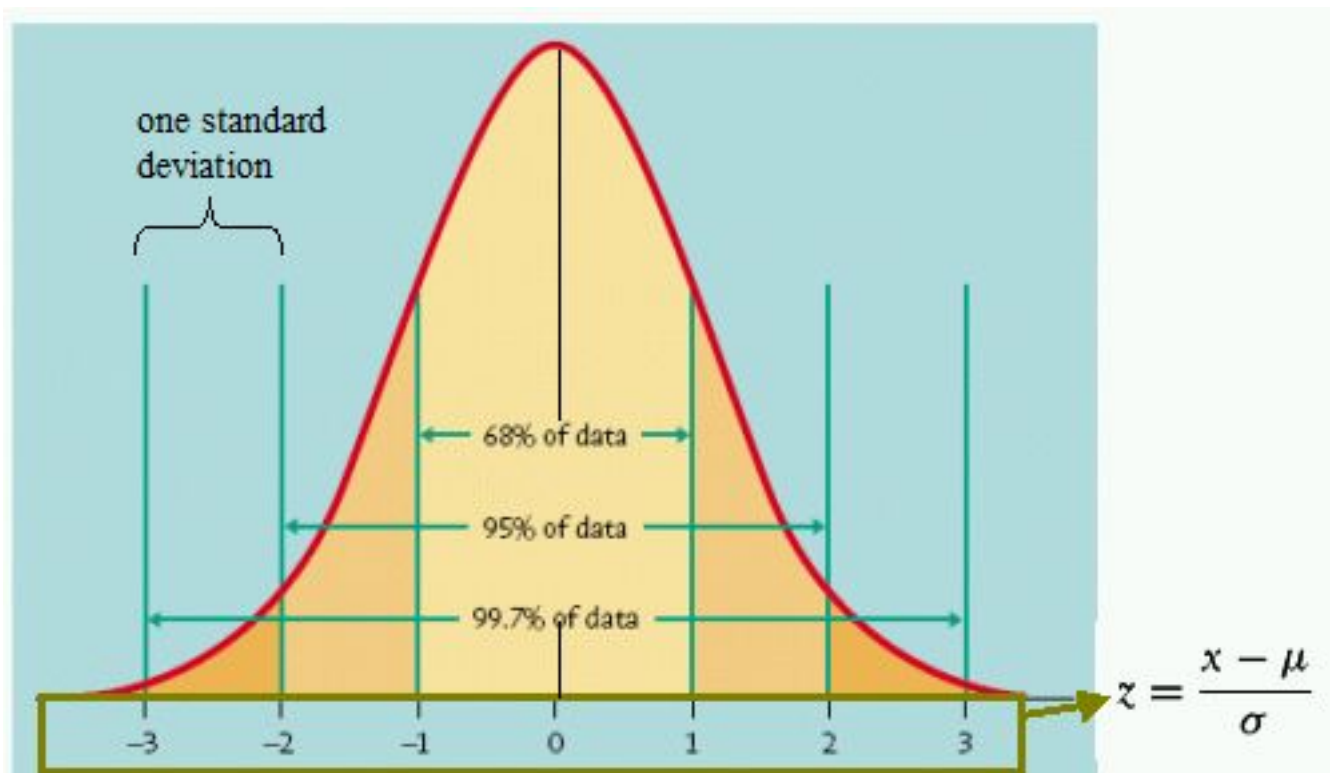


Chapter 3 - Workbook

Analyzing Univariate Data

MDM4U



Section 3.1 Worksheet – Shapes of Distributions

MDM4U

Jensen

1) Match the following distribution curves to the random variables listed below. Also, describe the shape of the distribution.

i) cost of the “cheap seats” at 30 baseball stadiums

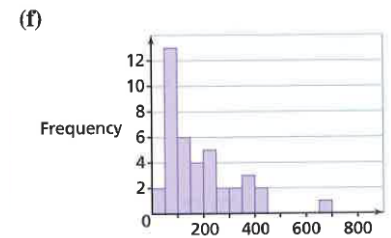
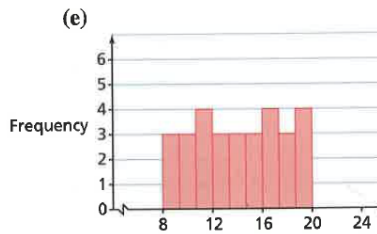
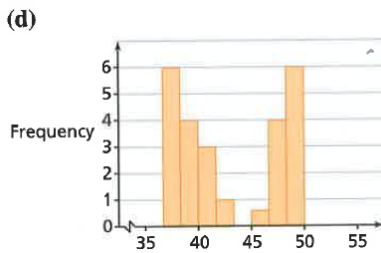
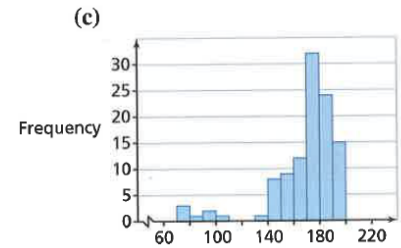
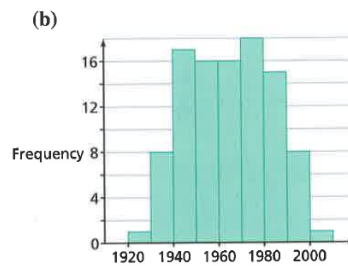
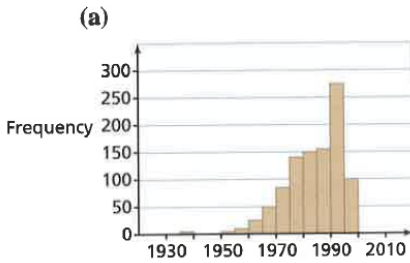
ii) bowling scores

iii) the gestation period in days of various animals

iv) the year shown on a penny

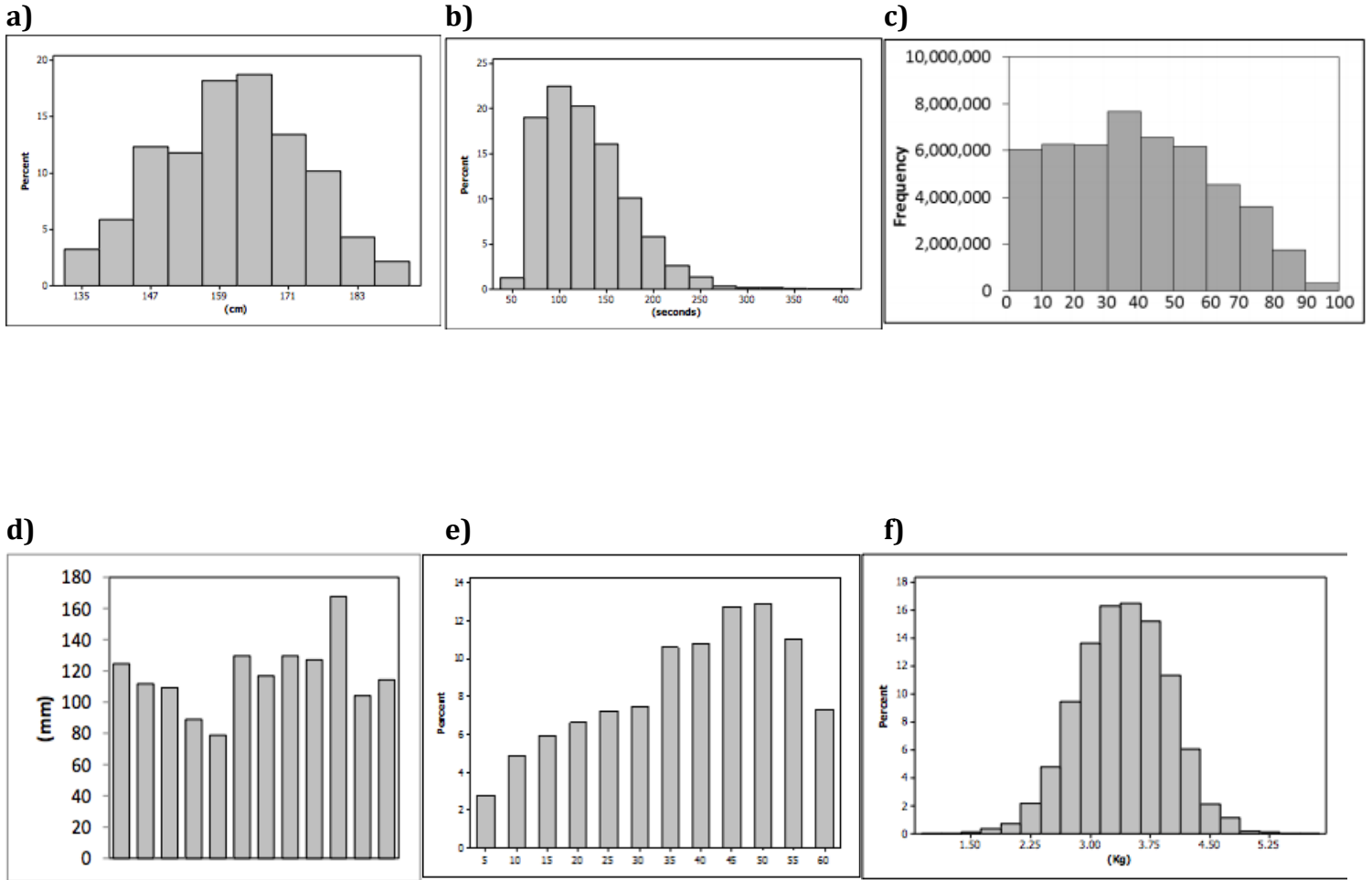
v) the production year of the American Film Institute’s top 100 films

vi) amounts shown on an electric bill

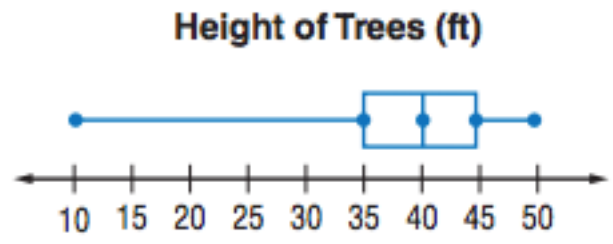


2) Match the following distribution curves to the random variables listed below. Also, describe the shape of the distribution.

- i) Population by age for England (2011 census)
- ii) Average rainfall per month for Bermuda
- iii) Time taken by students to complete an online quiz
- iv) Weight of new-born babies
- v) Children's heights
- vi) Student scores for a 12 question quiz (5 marks for each correct answer)



3) The box and whisker plot shows the heights in feet of several trees. Is the distribution skewed left or right? Explain.



4) Using the following data:

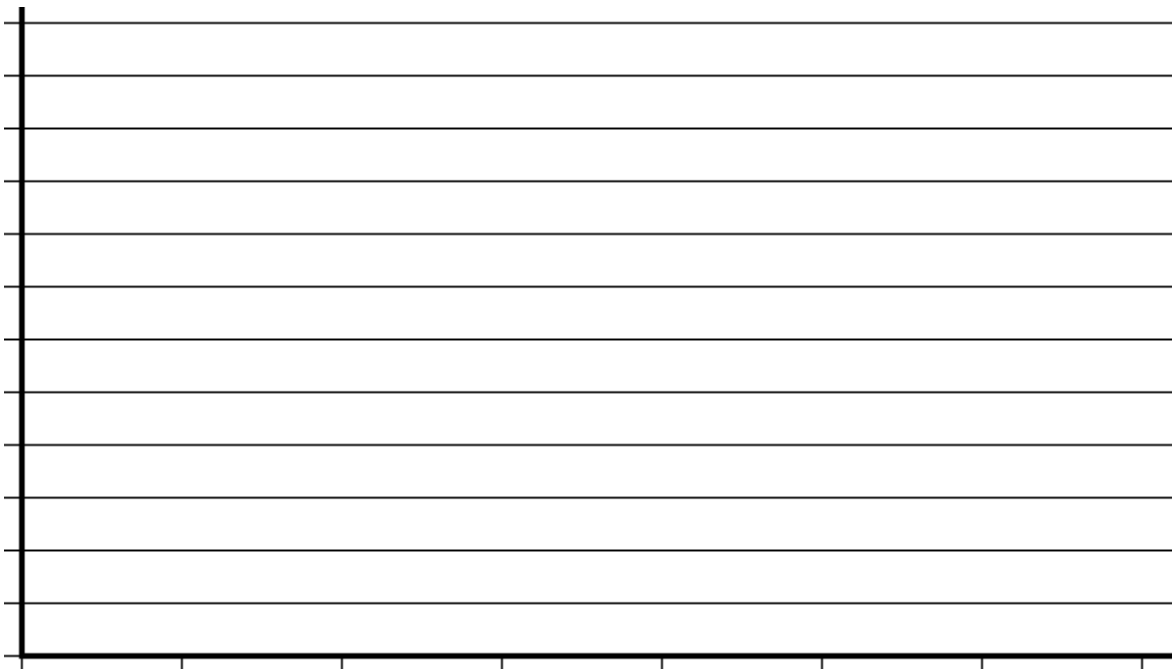
13, 7, 5, 7, 9, 10, 5, 11, 8, 7, 9, 10, 10, 11, 14, 10, 6, 9, 7, 12, 9, 10, 6

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

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

Interval	Frequency

c) Create an appropriate histogram.



5) The following data represent salaries, in thousands of dollars, for employees of a small company. Notice the data have been sorted in increasing order.

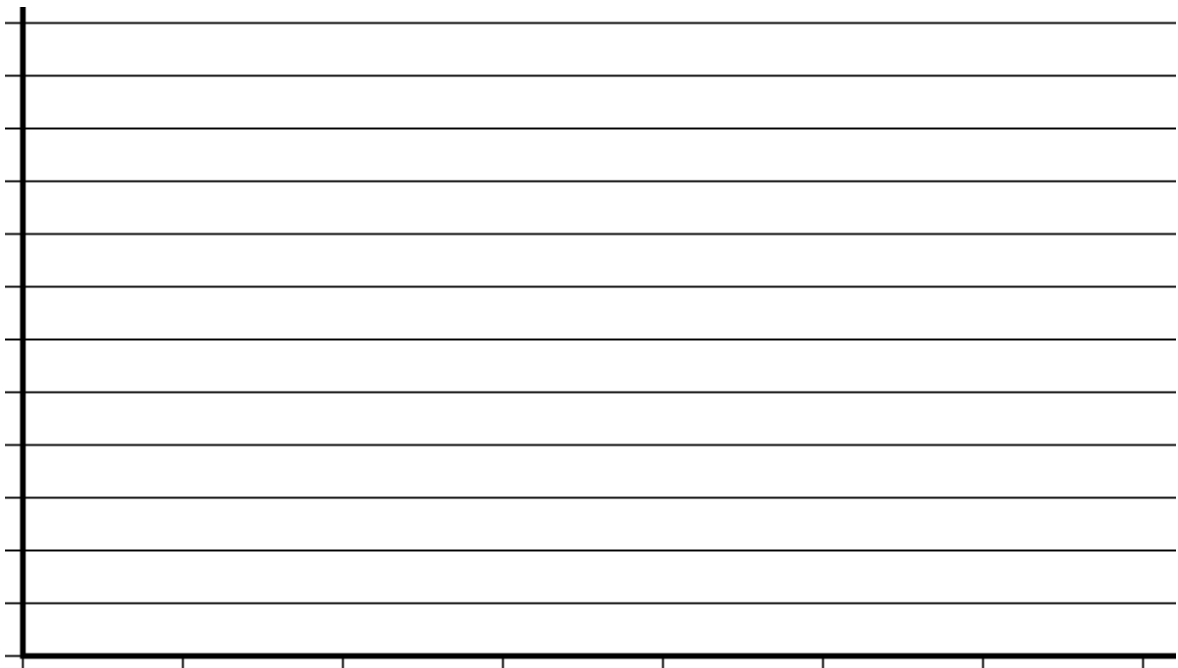
24, 25, 25, 27, 27, 29, 30, 35, 35, 35, 36, 38, 38
39, 39, 40, 40, 40, 45, 45, 45, 45, 47, 52, 52, 52
58, 59, 59, 61, 61, 67, 68, 68, 68, 70

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

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

Salary Interval	Frequency

c) Create an appropriate histogram.



Section 3.2 Worksheet – Measures of Central Tendency

MDM4U

Jensen

1) Use technology to calculate the mean, median, and mode for the following samples. Then use the relative location of the mean, median, and mode to describe the sets as symmetric, skewed left, or skewed right.

a) Marks on a set of tests {66, 65, 72, 78, 93, 70, 68, 64}

b) Monthly rent (\$) {625, 750, 800, 650, 725, 850, 625, 650, 625, 1250}

c) Points scored by a basketball player {12, 15, 8, 12, 15, 10, 3, 14, 15}

2) Here is a sample of Hakim's Shoes reported sales results:

Size	4	5	6	7	8	9	10
Frequency	5	11	15	18	19	13	7

a) Calculate the mean, median, and mode shoe size by hand.

b) Which measure of central tendency is most appropriate? Why?

3) A pair of dice is rolled numerous times. The sum of the dice, as well as the frequency, is recorded. Calculate the mean, median, and mode for the results using technology.

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	2	3	5	7	9	11	8	7	4	2	1

4) Jasmine records the dates on 125 pennies. Find the mean date of the sample by hand. Check your answer using technology.

Date	1990 - 1999	1980 -1989	1970 - 1979	1960 - 1969
Frequency	56	42	21	6

5) A student's term mark is 75. The term mark counts for 70% of the final mark. What mark must the student achieve on the exam to earn a final mark of..

a) 70

b) 80

6) The following table shows the salary structure of Statsville Plush Toys, Inc. Assume that salaries exactly on an interval boundary have been placed in the higher interval. Calculate the mean salary by hand. Check your answer using technology.

Salary Range (\$000)	Number of Employees, f
20 - 30	12
30 - 40	24
40 - 50	32
50 - 60	19
60 - 70	9
70 - 80	3
80 - 90	0
90 - 100	1

Section 3.3 Worksheet – Measures of Spread

MDM4U

Jensen

Refer to part 2 of the lesson for help with the following questions

1) Determine the range, median, Q_1 , Q_3 , and the interquartile range for the following sets of data. You may use technology.

a) Number of home runs hit by players on the Statsville little league team:

6 4 3 8 9 11 6 5 15

b) Final grades in a geography class:

88 56 72 67 59 48 81 62
90 75 75 43 71 64 78 84

2) For a recent standardized test, the median was 88, Q_1 was 67, Q_3 was 105. Describe the following scores in terms of quartiles

a) 8

b) 81

c) 103

Refer to part 3 of the lesson for help with the following questions

3) The number of calories in 12-ounce servings of five popular beers are {95, 152, 188, 205, 131}. Calculate the sample mean and standard deviation of calories by hand. Check your answer using your calculator.

4) Given the sample data x : {23, 17, 15, 30, 25}

a) Calculate the sample standard deviation by hand.

b) Suppose the given data comprise the entire population of all x values. Compute the population standard deviation by hand.

5) The Points Per Game during the 2012-2013 season of all the New York Knicks players were {14.2, 28.7, 10.4, 1.8, 6.6, 13.9, 6.0, 18.1, 6.8, 7.0, 8.7, 3.5, 7.2}. Calculate the population mean and standard deviation using technology.

Refer to part 4 of the lesson for help with the following question

6) Four classes recorded a sample of their pulse rates. Use the sample mean and standard deviation to determine which class has the lowest pulse rate and which class has the most consistent pulse rate using technology.

Class A	63	78	79	75	73	72	62	75	63	77	77	65	70	69	80
Class B	72	66	73	80	74	75	64	68	67	70	70	69	69	74	74
Class C	68	75	78	73	75	68	71	78	65	67	63	69	59	68	79
Class D	78	75	76	76	79	78	78	76	74	81	78	76	79	74	76

Refer to part 5 of the lesson for help with the following question

7) The nurses' union collects data on the hours worked by all of the operating-room nurses at the Statsville General Hospital. Calculate the population mean and standard deviation using technology.

Hours Per Week	Number of Employees
12	1
32	5
35	7
38	8
42	5

8) The following table shows a sample of salaries for employees at a company. Calculate the sample mean and standard deviation using technology.

Salary (\$)	Frequency
28 000	4
30 000	6
32 000	7
34 000	4
36 000	2
38 000	1

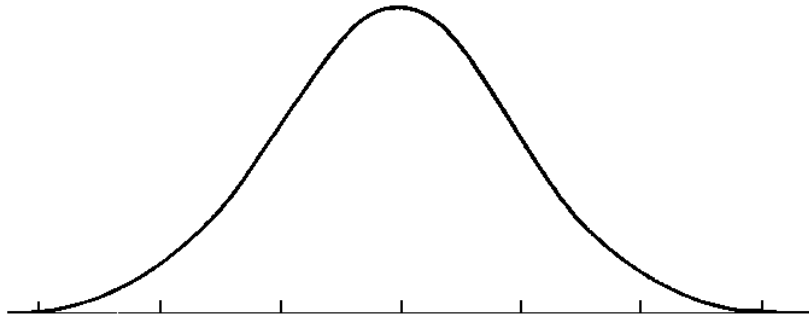
Section 3.4 Worksheet – Normal Distribution

MDM4U

Jensen

1) The distribution of vocabulary scores for seventh-graders in Indiana is $N(6.84, 1.55^2)$.

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.

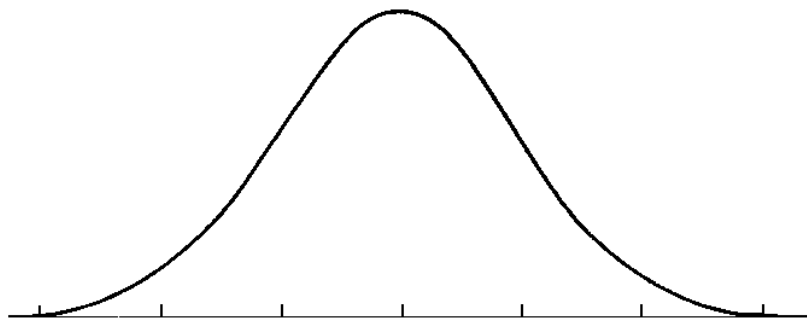


b) What percent of the vocabulary scores are less than 3.74?

c) What percent of scores are between 5.29 and 9.94?

2) For Major League Baseball players, the mean of 432 batting averages is 0.261 with a standard deviation of 0.034. Suppose that the distribution is normally distributed.

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

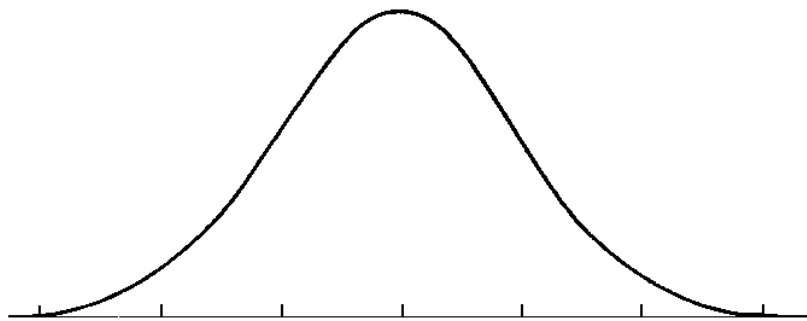


b) What percent of batting averages are between 0.261 and 0.329?

c) What percent of batting averages are less than 0.227?

3) Out of 100 packages of jawbreakers, 68 packages contain between 120 and 150. Use your knowledge of normal distribution to estimate the average number of jawbreakers and the standard deviation of the sample.

4) The amount of coffee an automatic machine dispenses (in ounces) can be represented by the normal distribution $X \sim N(10.2, 0.6^2)$.



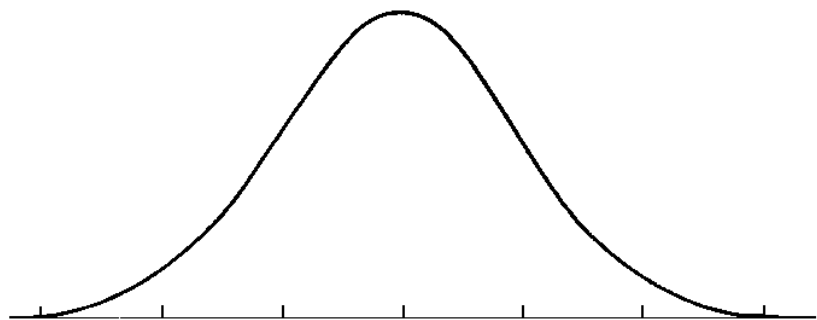
a) What range does 95% of the quantity of coffee dispensed lie between?

b) What percent of cups dispensed contain greater than 10.8 ounces?

c) What percent of cups dispensed contain between 9.6 and 10.2 ounces?

d) What percent of cups dispensed contain less than 9 ounces?

5) Burns Appliance Co. offers a replacement warranty on their toaster ovens, which have a mean lifespan of 8.5 years, with a standard deviation of 0.8 years. How long a warranty would they establish if they could only afford to repair no more than 2.5% of the toaster ovens they make?



Section 3.5a Worksheet – Applying the Normal Distribution

MDM4U

Jensen

1) Calculate a z-score for each x-value given $\mu = 6$ and $\sigma = 2$.

a) $x = 5.3$

b) $x = 7.2$

c) $x = 9.9$

d) $x = 0.8$

2) Using the z-score table (or your calculator), find the percentile that corresponds to each of the following z-scores.

a) $z = 2.33$

b) $z = -0.83$

3) Given a normally distributed data set whose mean is 50 and whose standard deviation is 10, what value of x would a z-score of 2.5 have?

4) Adrian's average bowling score is 174, and is normally distributed with a standard deviation of 35. In what percent of games does Adrian score more than 200 points?

5) The top 10% of bowlers in Adrian's league get to play in an all-star game. If the league average is 170, with a standard deviation of 11 points, and is normally distributed what average score does Adrian need to have to obtain a spot in the all-star game?

6) IQ score of people around the world are normally distributed, with a mean of 100 and a standard deviation of 15. A genius is someone with an IQ greater than or equal to 140. What percent of the population is considered genius?

7) A standardized test is known to be normally distributed with a mean of 500 and a standard deviation of 110.

a) A student's score is 675, what percentile is she in?

b) Another student taking the same test wants to score in the 80th percentile. What score must he get?

8) The weights of 75 model planes at a local convention are normally distributed. The average weight is 4.4 kg, with a standard deviation of 0.41 kg.

a) How many planes have a mass less than 4 kg?

b) How many planes would be disqualified if it were against the rules to have a plane with a mass of more than 5 kg?

c) How many planes have a mass between 3.5 kg and 5 kg?

9) On the driving range, Tiger Woods practices his swing with a particular club by hitting many, many balls. Suppose that when Tiger hits his driver, the distance the ball travels follows a normal distribution with mean 304 yards and standard deviation 8 yards.

a) What percent of Tiger's drives travel at least 290 yards?

b) What percent of Tiger's drives travel between 305 and 325 yards?

10) For the distribution $X \sim N(3, 0.74^2)$, determine the percent of the data that is within the given interval

a) $X > 2.44$

b) $1.8 < X < 2.3$

c) $X < 1.91$

11) Perch in a lake have a mean length of 20 cm and a standard deviation of 5 cm. What would be the length of a fish in the 95th percentile?

Section 3.5b Worksheet – Applying the Normal Distribution

MDM4U

Jensen

1) Copy and complete the chart below, assuming a normal distribution for each situation.

Mean, μ	Standard Deviation, σ	Probability
12	3	$P(X < 9) =$
30	5	$P(X < 25) =$
5	2.2	$P(X > 6) =$
245	18	$P(233 < X < 242) =$

2) There have been some outstanding hitters in baseball. In 1911, Ty Cobb's batting average was 0.420. In 1941, Ted Williams batted 0.406. George Brett's 0.390 average in 1980 was one of the highest since Ted Williams. Batting averages have historically been approximately normally distributed with means and standard deviations as shown below. Compute z-scores for each of these three hitters. Can you rank the three hitters? Explain.

Decade	Mean, μ	Standard Deviation, σ
1910's	0.266	0.0371
1940's	0.267	0.0326
1970s-1980s	0.261	0.0317

3) The amount of annual rainfall in a certain region is known to be a normally distributed random variable with a mean of 50 inches and a standard deviation of 4 inches. If the rainfall exceeds 57 inches during the year, it leads to floods. Find the probability that during a randomly selected year there will be floods.

- 4)** The weight of food packed in certain containers is a normally distributed random variable with a mean weight of 500 pounds and a standard deviation of 5 pounds. If a container is picked at random, find the probability that it contains:
- a)** more than 510 pounds
 - b)** less than 498 pounds
 - c)** between 491 and 498 pounds
- 5)** The diameter of a lead shot has a normal distribution with a mean diameter equal to 2 inches and a standard deviation equal to 0.05 inches. Find what diameter a circular hole should be so that only 3 percent of the lead shots can pass through it.
- 6)** The nicotine content in a brand of king-size cigarettes has a normal distribution with a mean content of 1.8 mg and a standard deviation of 0.2 mg. Find the probability that the nicotine content of a randomly selected cigarette of this brand will be:
- a)** less than 1.45 mg
 - b)** between 1.45 and 1.65 mg
 - c)** more than 2.15 mg.
 - d)** What value is needed so that 80 percent of the cigarettes will exceed it in their nicotine content?
- 7)** The demand for meat at a grocery store during any week is approximately normally distributed with a mean demand of 5000 pounds and a standard deviation of 300 pounds.
- a)** If the store has 5300 pounds of meat, what is the probability that they will run out during a random week?
 - b)** How much meat should the store have in stock per week so as to only run short 10 percent of the time?

Section 3.6 Worksheet – Confidence Intervals

MDM4U

Jensen

1) For each set of data, determine the margin of error and confidence interval for a 95% confidence level.

	n	\bar{x}	σ
a)	40	215	8
b)	130	35	3.4
c)	30	9.65	0.56

2) A machine that produces control arms for a vehicle gas pedal generates pedals that have a length with standard deviation of 0.08 cm. Thirty pedals are tested to see if their lengths are acceptable. The sample has a mean of 18.2 cm. What would be the acceptable range of lengths for a 95% confidence level for the mean length?

3) You have a part-time job maintaining a water-jug-refilling machine. The machine rarely fills each jug to the same volume and sometimes needs recalibrating. The manufacturer states that the standard deviation of the machine is 0.3 L. You monitor the next 20 fillings and determine that their mean volume is 18.8 L. Assuming the data are normally distributed, determine the acceptable range of volumes for a confidence level of 95%.

4) An exit poll is done outside a voting location. People who have just voted are asked if they will state who they voted for. In a close election, an exit poll states that Larry Liberal has 48% of the vote, while Constance Conservative has 46% of the vote, with the rest split up among other candidates. The polling firm states that 500 people were polled. Find a 95% confidence interval for the proportion of people who support each candidate.

5) A market-research firm asked 300 people in Toronto who their favourite hockey team is. 55 said the Leafs are their favourite team. Determine a 99% confidence interval for the proportion of people in Toronto that have chosen the Leafs to be their favourite team.

CHAPTER 3 ANSWER KEY

SECTION 3.1

- 1) **i)** e; uniform
ii) c; skewed left
iii) f; skewed right
iv) a; skewed left
v) b; mound
vi) d; bimodal(u-shaped)

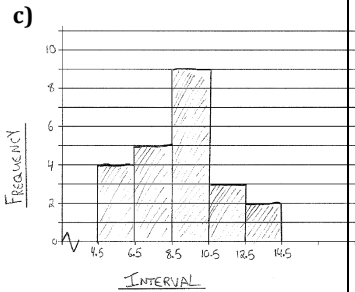
- 2) **i)** c; right skewed
ii) d; approximately uniform
iii) b; right skewed
iv) f; mound
v) a; mound
vi) e; left skewed

3) Skewed left. 75% of the data is clustered on the right side (between 35 and 50). Only 25% of the data is on the left (between 10 and 35)

4) a) 2

b)

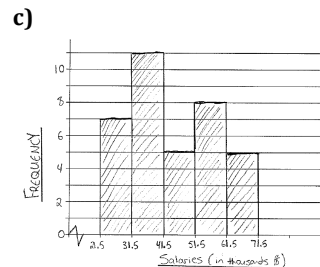
Interval	Frequency
4.5 - 6.5	4
6.5 - 8.5	5
8.5 - 10.5	9
10.5 - 12.5	3
12.5 - 14.5	2



5) a) 10

b)

Salary Interval	Frequency
21.5 - 31.5	7
31.5 - 41.5	11
41.5 - 51.5	5
51.5 - 61.5	8
61.5 - 71.5	5



SECTION 3.2

- 1) **a)** $\bar{x} = 72$; median = 69; mode = none;
mean > median; skewed right
b) $\bar{x} = 755$; median = 687.5; mode = 625;
mean > median > mode; skewed right
c) $\bar{x} = 11.56$; median = 12; mode = 15;
mean < median < mode; skewed left
2) a) $\bar{x} = 7.16$; median = 7; mode = 8
b) Mode tells you the most popular size

3) $\bar{x} = 6.78$; median = 7; mode = 7

4) $\bar{x} = 1986.34$

5) **a)** 58.33% **b)** 91.67%

6) \$45 300

SECTION 3.3

- 1) **a)** range = 12; $Q_1 = 4.5$; median = 6; $Q_3 = 10$; IQR = 5.5
b) range = 47; $Q_1 = 60.5$; median = 71.5; $Q_3 = 79.5$; IQR = 19

2) **a)** first quartile **b)** second quartile **c)** third quartile

3) $\bar{x} = 154.2$; $s = 44.076$

4) **a)** $s = 6.08$ **b)** $\sigma = 5.44$

5) $\mu = 10.22$; $\sigma = 6.87$

6)

	Mean, \bar{x}	Standard Deviation, s
Class A	71.87	6.22
Class B	71	4.12
Class C	70.4	5.88
Class D	76.93	1.98

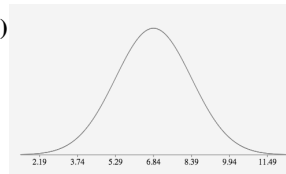
Class C has the lowest pulse rate. Class D has the most consistent pulse rate because it has the lowest standard deviation.

7) $\mu = 35.8$; $\sigma = 5.8$

8) $\bar{x} = 31\ 750$; $s = 2\ 722.7$

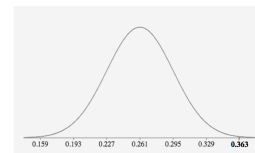
SECTION 3.4

1) a)



- b)** 2.4% or 2.5%
c) 81.5%

2) a)



- b)** 47.5%
c) 15.9% or 16%

3) $\sigma = 15$; $\mu = 135$

- 4) **a)** between 9 and 11.4 ounces
b) 15.9% or 16%
c) 34%
d) 2.4% or 2.5%

5) 6.9 year warranty

SECTION 3.5a

1) **a)** -0.35 **b)** 0.6 **c)** 1.95 **d)** -2.6

2) **a)** 99th **b)** 20th

3) 75

4) 22.96% or 22.88%

5) about 184

6) 0.38%

7) a) 94th percentile b) 592.4 or 592.6

8) a) about 12 b) about 5 c) about 69

9) a) 95.99% b) 44.59%

10) a) 77.54% b) 11.96% c) 7.04%

11) 28.2 cm

SECTION 3.5b

1) a) 0.1587 b) 0.1587 c) 0.3247 d) 0.1813

2) Ty Cobb: $z_{0.42} = 4.15$
Ted Williams: $z_{0.406} = 4.26$
George Brett: $z_{0.39} = 4.07$

Based on z-scores, which tell us how many standard deviations a player's batting average is above the mean, Ted Williams has the best average, then Ty Cobb, then George Brett.

3) 4.01%

4) a) 2.285% b) 34.46% c) 30.86%

5) 1.906 inches

6) a) 4.01% b) 18.66% c) 4.01% d) 1.63 mg

7) a) 15.87% b) 5384.5 pounds

SECTION 3.6

1) a) We can be 95% confident that the interval 212.52 to 217.48 captures the true population mean.

b) We can be 95% confident that the interval from 34.416 to 35.584 captures the true population mean.

c) We can be 95% confident that the interval from 9.45 to 9.85 captures the true population mean.

2) We can be 95% confident that the interval from 18.17 cm to 18.23 cm captures the true mean length of pedals.

3) We can be 95% confident that the interval from 18.67 L to 18.93 L captures the true mean volume.

4) We can be 95% confident that the interval from 0.4362 to 0.5238 captures the true proportion of voters who support Larry.

We can be 95% confident that the interval from 0.4163 to 0.5037 captures the true proportion of voters who support Constance.

5) We can be 99% confident that the interval from 0.1258 to 0.2409 captures the true population proportion of people in Toronto who have chosen the Leafs to be their favourite team.