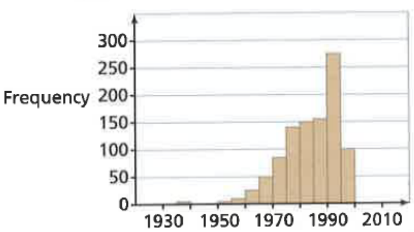
Section 3.1 Worksheet – Shapes of Distributions

MDM4U
Jensen

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

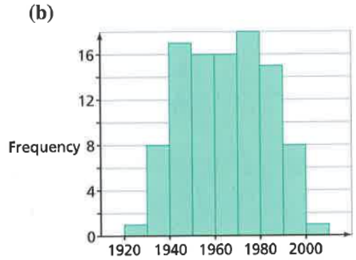
i) cost of the "cheap seats" at 30 baseball stadiums
ii) bowling scores
iii) the gestation period in days of various animals
iv) the year shown on a penny
v) the production year of the American Film Institute's top 100 films
vi) amounts shown on an electric bill

(a)



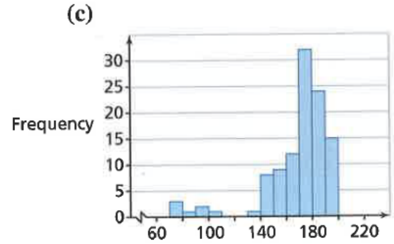
iv) year shown on a penny
skewed left

(b)



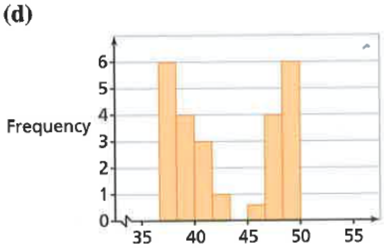
v) year of top 100 films
mound shaped

(c)



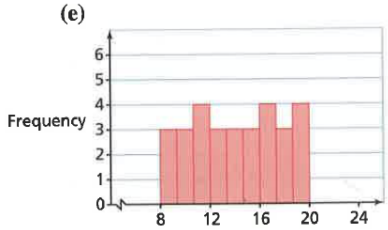
ii) bowling scores
skewed left

(d)



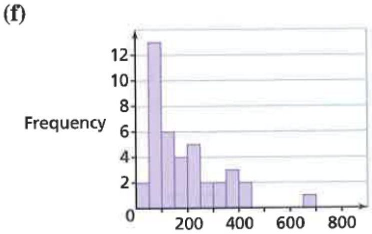
vi) amount of electric bill
bimodal (u-shaped)

(e)



i) the cost of cheap seats
uniform

(f)

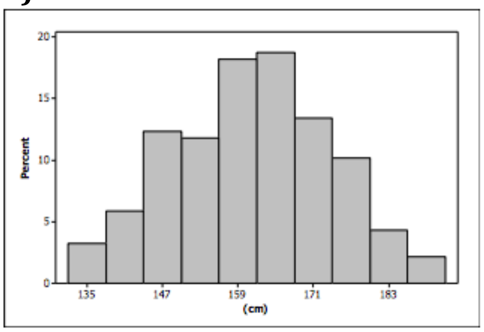


iii) gestation period
skewed right

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

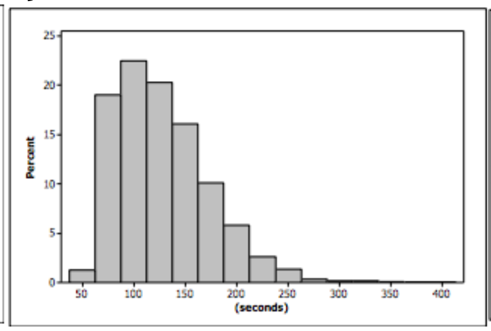
i) Population by age for England (2011 census)
ii) Average rainfall per month for Bermuda
iii) Time taken by students to complete an online quiz
iv) Weight of new-born babies
v) Children's heights
vi) Student scores for a 12 question quiz (5 marks for each correct answer)

a)



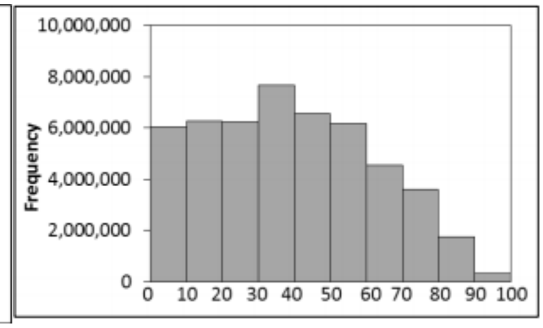
v) children's heights
mound shaped

b)



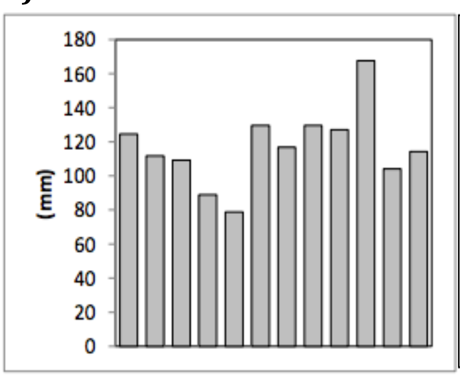
iii) time taken for quiz
rights skewed

c)



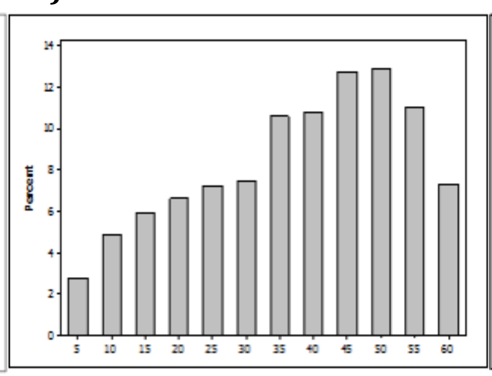
i) population by age
right skewed

d)



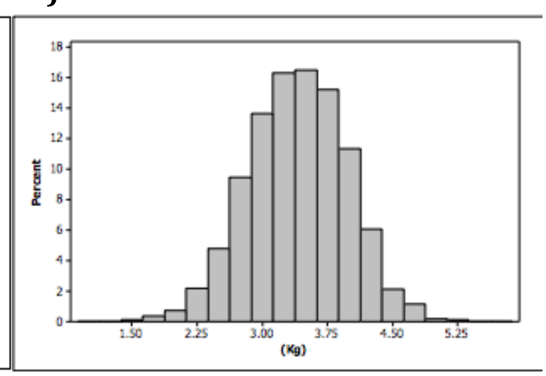
ii) average rainfall
approximately uniform

e)



vi) quiz scores
left skewed

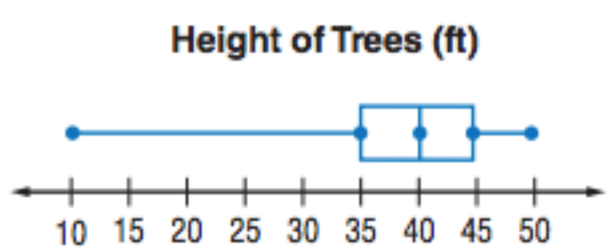
f)



iv) weight of babies
mound shaped

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

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) 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

$Range = 14 - 5 = 9$ $\quad$ $Bin\ Width = \frac{10}{5} = 2$

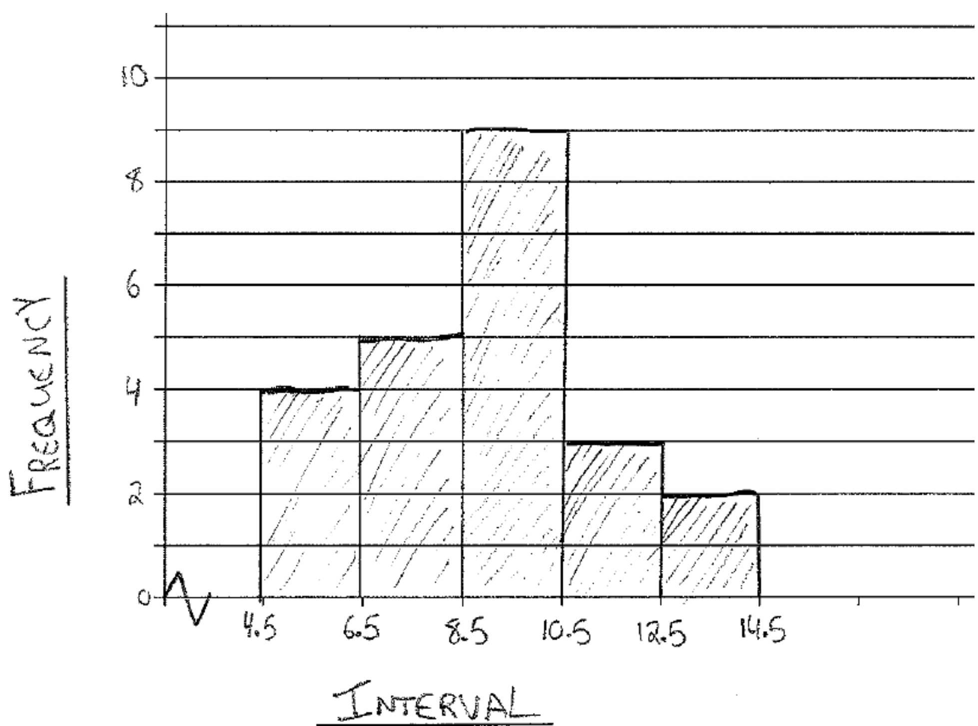
b) Calculate the starting and end point for each of the five intervals. Then complete the frequency distribution.

Rounded range up by 1 from 9 to 10. Subtract $\frac{1}{2} = 0.5$ *from lowest value to determine starting point.*

Starting point = 5 – 0.5 = 4.5

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

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

$Range = 70 - 24 = 46$ $\quad\quad$ $Bin\ Width = \frac{50}{5} = 10$

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

Rounded range up by 4 from 46 to 50. Subtract $\frac{4}{2} = 2$ from lowest value to determine starting point.

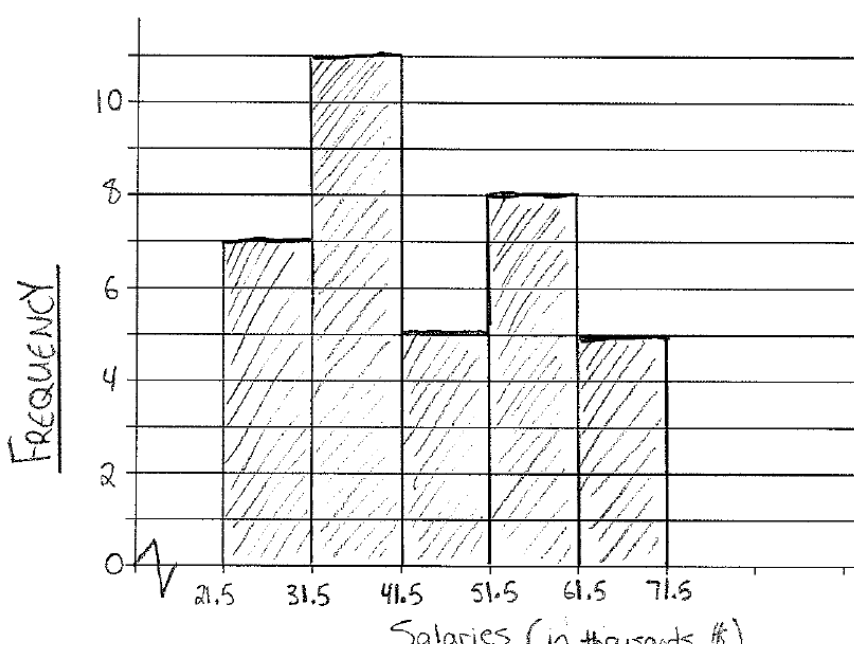
Starting point = 24 – 2 = 22

Also subtract 0.5 to make sure no data falls on a boundary.

Staring point =22 – 0.5 = 21.5

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

c) Create an appropriate histogram.

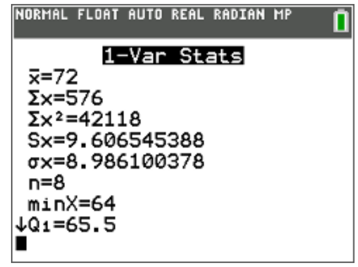


Section 3.2 Worksheet – Measures of Central Tendency

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}



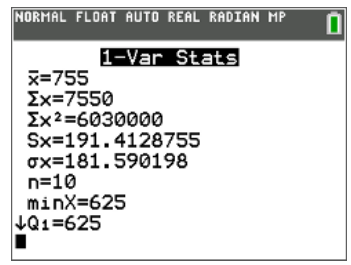
$\bar{x} = 72$

$median = 69$

$mode = none$

$mean > median; skewed\ right$

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



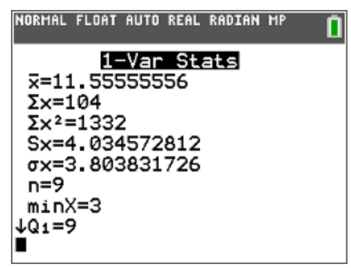
$\bar{x} = 755$

$median = 687.5$

$mode = 625$

$mean > median > mode; skewed\ right$

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



$\bar{x} = 11.56$

$median = 12$

$mode = 15$

$mean < median < mode; skewed\ left$

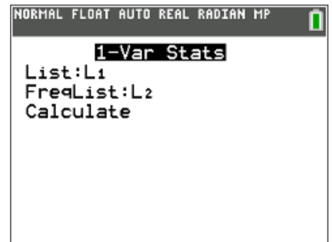
2) Here is a sample of Hakim's Shoes reported sales results:

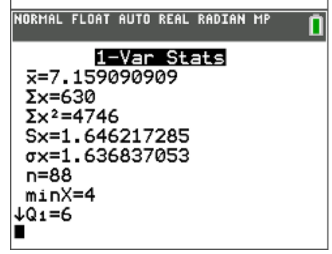
Size	4	5	6	7	8	9	10
Frequency	5	11	15	18	19	13	7

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

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{4(5) + 5(11) + 6(15) + 7(18) + 8(19) + 9(13) + 10(7)}{5 + 11 + 15 + 18 + 19 + 13 + 7} = \frac{630}{88} = 7.16$$

$median: \frac{88+1}{2} = 44.5$ the median data value is between the 44th and 45th piece of data. If we look at cumulative frequencies, this falls within the size 7 category.





$median = 7$

$mode = 8$

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

Mode tells you the most popular size

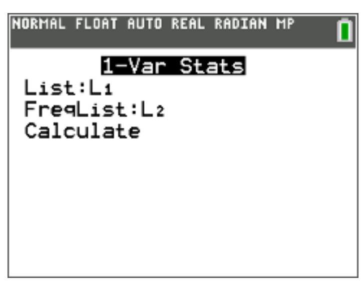
3) A pair of dice is rolled numerous times. The sum of the dice, as well as the frequency, is recorded. Calculate the mean, median, and mode for the results using technology.

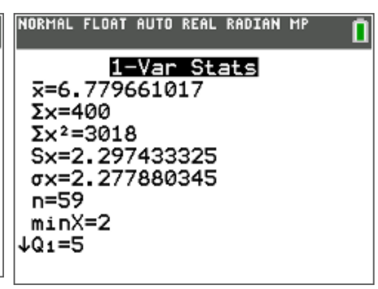
Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	2	3	5	7	9	11	8	7	4	2	1

$\bar{x} = 6.78$

$median = 7$

$mode = 7$





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

Solution using a table:

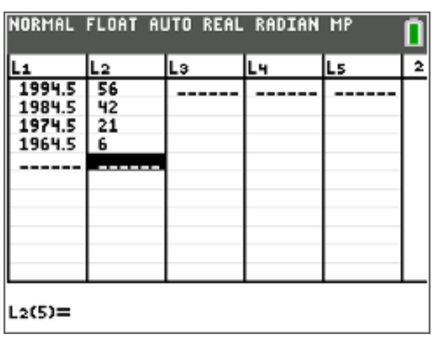
Date	**Frequency, f**	**Midpoint of date, m**	**$f \times m$**
1990 – 1999	56	1994.5	111 692
1980 – 1989	42	1984.5	83 349
1970 – 1979	21	1974.5	41 464.5
1960 – 1969	6	1964.5	11 787
	$\sum f = 125$		$\sum m \times f = 248\ 292.5$

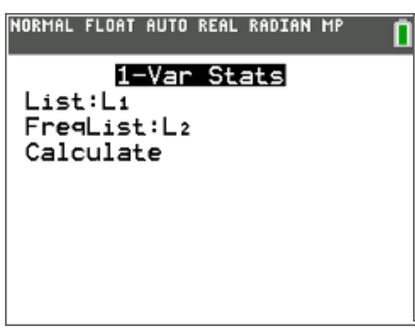
$$\bar{x} = \frac{\sum mf}{\sum f} = \frac{248\ 292.5}{125} = 1986.34$$

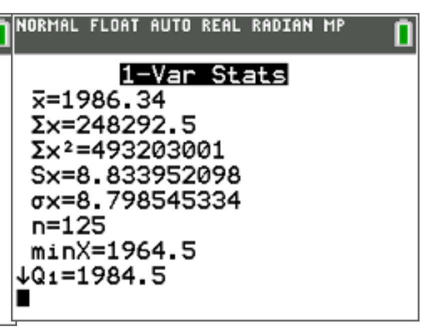
Solution without table:

$$\bar{x} = \frac{\sum mf}{\sum f} = \frac{1994.5(56) + 1984.5(42) + 1974.5(21) + 1964.5(6)}{56 + 42 + 21 + 6} = \frac{248\ 292.5}{125} = 1986.34$$

Check answer with calculator:







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

$\bar{x} = \frac{\sum xw}{\sum w}$

$70 = \frac{75(0.7)+x(0.3)}{0.7+0.3}$

$70 = 52.5 + x(0.3)$

$17.5 = x(0.3)$

$\frac{17.5}{0.3} = x$

$x = 58.33$

The student must get a 58.33% on the final exam to earn a final mark of 70%.

b) 70

$\bar{x} = \frac{\sum xw}{\sum w}$

$80 = \frac{75(0.7)+x(0.3)}{0.7+0.3}$

$80 = 52.5 + x(0.3)$

$27.5 = x(0.3)$

$\frac{27.5}{0.3} = x$

$x = 91.67$

The student must get a 91.67% on the final exam to earn a final mark of 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.

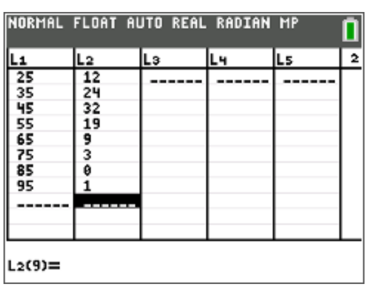
Salary Range ($000)	**Number of Employees, *f***	**Midpoint of Salary, *m***	$f \times m$
20 – 30	12	25	300
30 – 40	24	35	840
40 – 50	32	45	1440
50 – 60	19	55	1045
60 – 70	9	65	585
70 – 80	3	75	225
80 – 90	0	85	0
90 – 100	1	95	95
	$\sum f = 100$		$\sum f \times m = 4530$

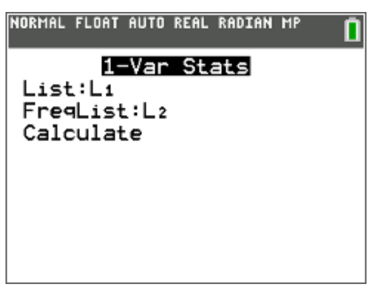
$$\bar{x} = \frac{\sum mf}{\sum f} = \frac{4530}{100} = 45.3$$

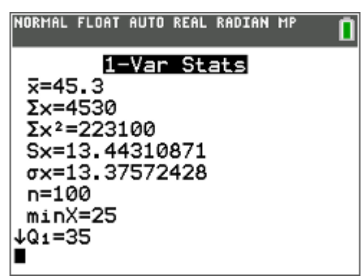
OR without table:

$$\bar{x} = \frac{\sum mf}{\sum f} = \frac{25(12) + 35(24) + 45(32) + 55(19) + 65(9) + 75(3) + 85(0) + 95(1)}{12 + 24 + 32 + 19 + 9 + 3 + 0 + 1} = \frac{4530}{100} = 45.3$$

Check answer with calculator:







The mean salary is about $45 300.

Section 3.3 Worksheet – Measures of Spread

MDM4U
Jensen

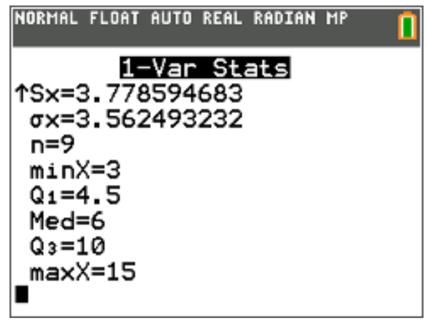
Refer to part 2 of the lesson for help with the following questions

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

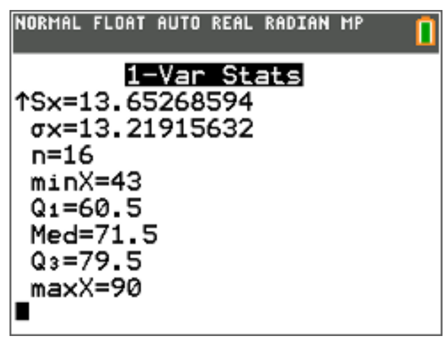
$range = 15 - 3 = 12$
$Q_1 = 4.5$
$median = 6$
$Q_3 = 10$
$IQR = 10 - 4.5 = 5.5$



b) Final grades in a geography class:

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

$range = 90 - 43 = 47$
$Q_1 = 60.5$
$median = 71.5$
$Q_3 = 79.5$
$IQR = 79.5 - 60.5 = 19$



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 *first quartile*
b) 81 *second quartile*
c) 103 *third quartile*

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.

$$\bar{x} = \frac{\sum x}{n} = \frac{95 + 152 + 188 + 205 + 131}{5} = \frac{771}{5} = 154.2$$

Solution using a table:

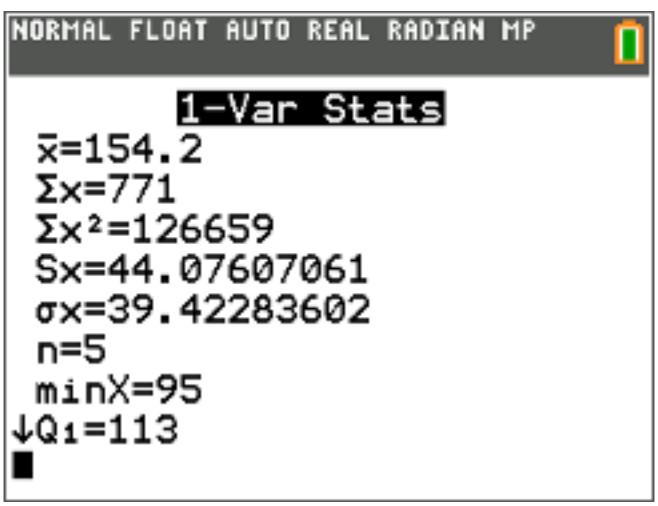
Calories, x	**Deviation, $x - \bar{x}$**	**$(x - \bar{x})^2$**
95	-59.2	3504.64
152	-2.2	4.84
188	33.8	1142.44
205	50.8	2580.64
131	-23.2	538.24
		$\sum(x - \bar{x})^2 = 7770.8$

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{7770.8}{5-1}} = 44.076$$

Solution without a table:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{(95-154.2)^2+(152-154.2)^2+(188-154.2)^2+(205-154.2)^2+(131-154.2)^2}{5-1}} = \sqrt{\frac{7770.8}{5-1}} = 44.076$$

Check with Ti-84:



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

a) Calculate the sample standard deviation by hand.

$$\bar{x} = \frac{\sum x}{n} = \frac{23+17+15+30+25}{5} = \frac{110}{5} = 22$$

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{(23-22)^2+(17-22)^2+(15-22)^2+(30-22)^2+(25-22)^2}{5-1}} = \sqrt{\frac{148}{4}} = 6.08$$

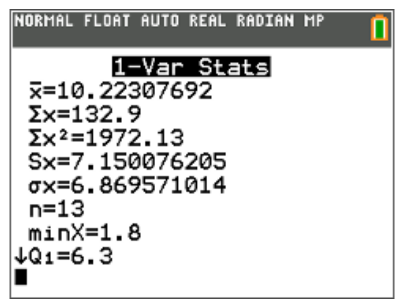
b) Suppose the given data comprise the entire population of all x values. Compute the population standard deviation by hand.

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}} = \sqrt{\frac{(23-22)^2+(17-22)^2+(15-22)^2+(30-22)^2+(25-22)^2}{5}} = \sqrt{\frac{148}{5}} = 5.44$$

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.

$\mu = 10.22$

$\sigma = 6.87$



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

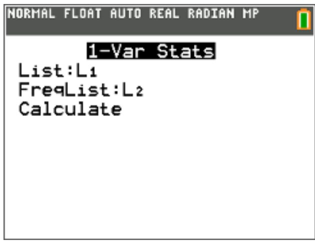
	Mean, $\bar{x}$	***Standard Deviation, s***
Class A	***71.87***	***6.22***
Class B	***71***	***4.12***
Class C	***70.4***	***5.88***
Class D	***76.93***	***1.98***

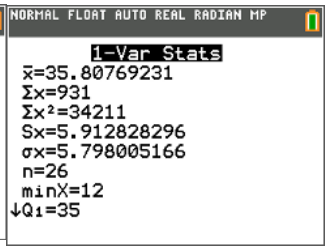
Class C has the lowest pulse rate. Class D has the most consistent pulse rate because it has the lowest standard deviation.

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





$\mu = 35.8$

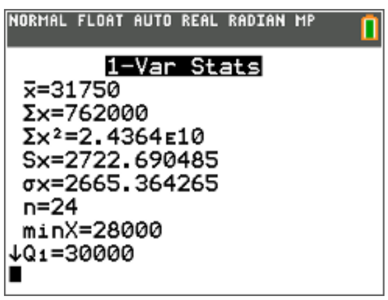
$\sigma = 5.8$

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

$\bar{x} = 31\ 750$

$s = 2\ 722.7$

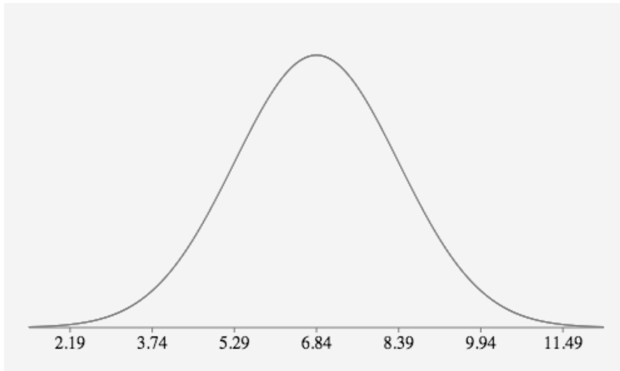


Section 3.4 Worksheet – Normal Distribution

MDM4U
Jensen

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

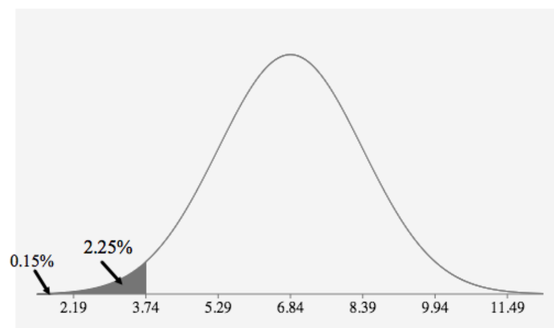
a) Sketch a normal curve for this distribution of vocabulary scores. Label the points that are 1, 2, and 3 standard deviations from the mean.



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

Method 1:

Area under curve = 0.15% + 2.25% = 2.4%



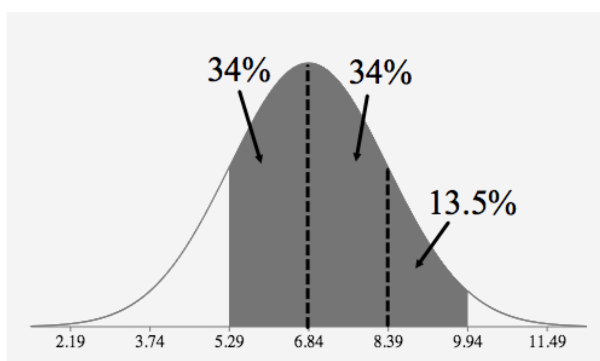
Method 2:

95% are within 2σ of μ (between 3.74 and 9.94). The other 5% are outside of this range. Because the normal distribution is symmetric, 2.5% are lower than 3.74 and 2.5% are bigger than 9.94.

Area under curve = 2.5%

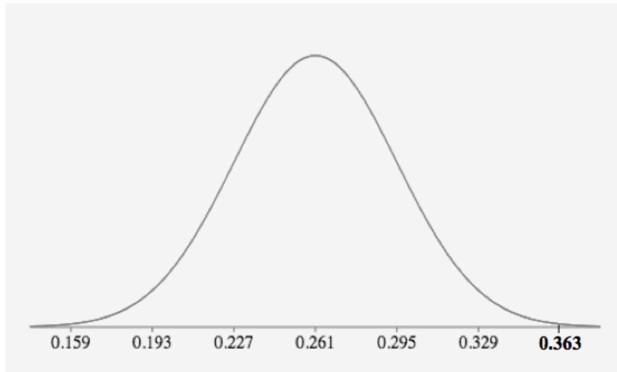
c) What percent of scores are between 5.29 and 9.94?

Area under curve = 34% +34% +13.5% = 81.5%



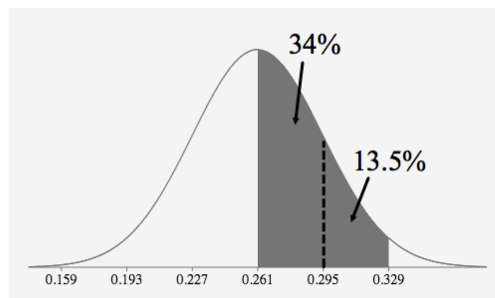
2) For Major League Baseball players, the mean of 432 batting averages is 0.261 with a standard deviation of 0.034. Suppose that the distribution is normally distributed.

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?

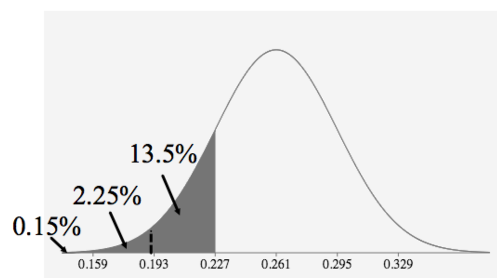
Area under curve = 34% + 13.5% = 47.5%



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

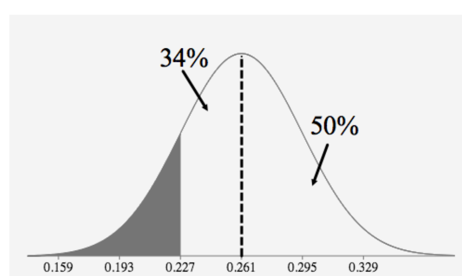
Method 1:

Area under curve = 0.15% + 2.25% + 13.5% = 15.9%



Method 2:

Area under curve = 100% - 50% - 34% = 16%

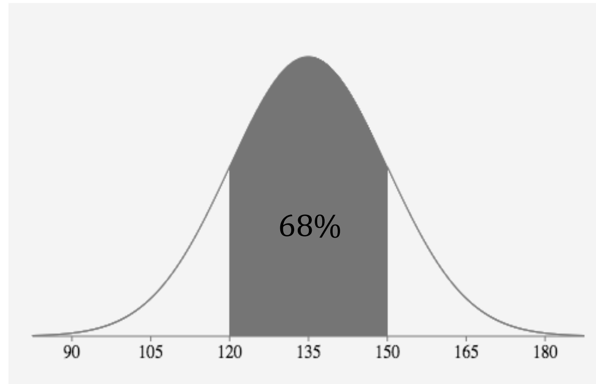


3) Out of 100 packages of jawbreakers, 68 packages contain between 120 and 150. Use your knowledge of normal distribution to estimate the average number of jawbreakers and the standard deviation of the sample.

68% implies within 1 standard deviation of the mean. In other words, a total of 2 standard deviations. Therefore:

$$\sigma = \frac{150-120}{2} = 15$$

$$\mu = 120 + 15 = 135$$



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

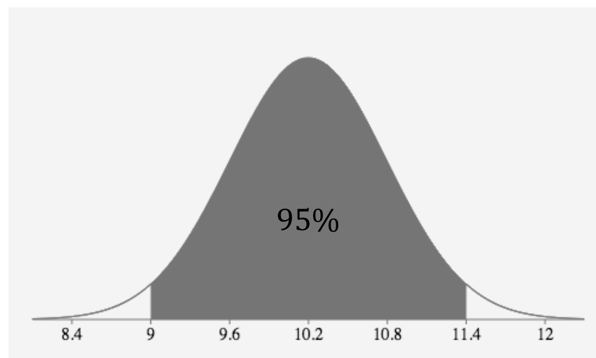
a) What range does 95% of the quantity of coffee dispensed lie between?

68% of the data is within 1 standard deviation of the mean.

$$\mu + 2\sigma = 10.2 + 2(0.6) = 9$$

$$\mu - 2\sigma = 10.2 - 2(0.6) = 11.4$$

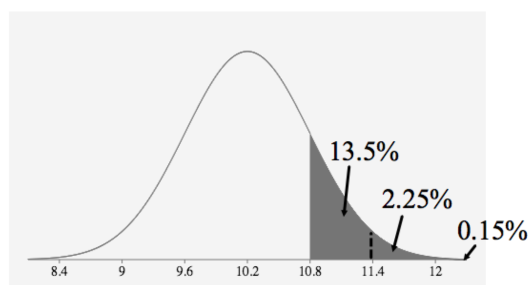
Therefore, 95% of the quantity of coffee dispensed should be between 9 and 11.4 ounces.



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

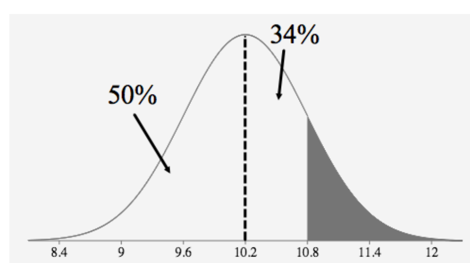
Method 1:

Area under curve = 13.5% + 2.25% + 0.15% = 15.9%



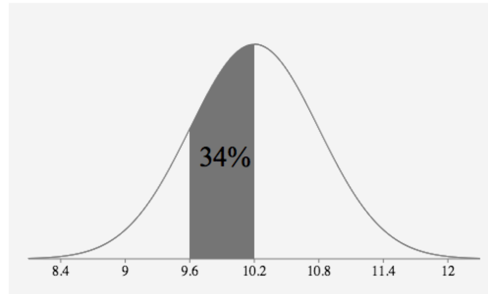
Method 2:

Area under curve = 100% - 50% - 34% = 16%



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

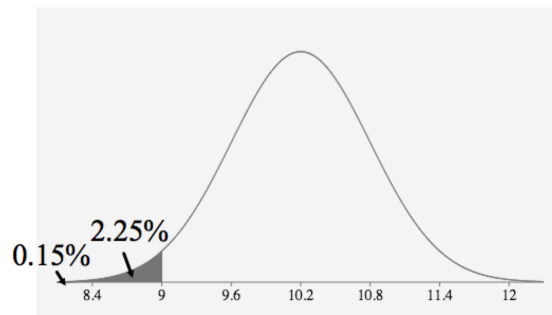
Area under curve = 34%



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

Method 1:

Area under curve = 0.15% + 2.25% = 2.4%



Method 2:

95% are within 2σ of μ (between 9 and 11.4). The other 5% are outside of this range. Because the normal distribution is symmetric, 2.5% are lower than 9 and 2.5% are bigger than 11.4.

Area under curve = 2.5%

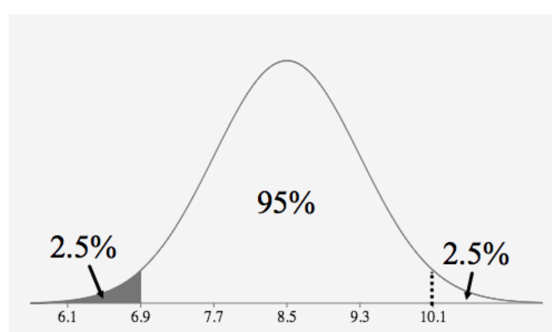
5) Burns Appliance Co. offers a replacement warranty on their toaster ovens, which have a mean lifespan of 8.5 years, with a standard deviation of 0.8 years. How long a warranty would they establish if they could only afford to repair no more than 2.5% of the toaster ovens they make?

2.5% at the end of the lower half means 5% at both ends combined. Since 95% of data in a normal distribution is within 2 standard deviations of the mean, a possible warranty they could establish is…

$warranty = \mu - 2\sigma = 8.5 - 2(0.8) = 6.9.$

Area under curve less than 6.9 years is 2.5%.

Therefore they should offer a 6.9 year warranty.



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) $x = 5.3$

$$z_{5.3} = \frac{x-\mu}{\sigma} = \frac{5.3-6}{2} = -0.35$$

b) $x = 7.2$

$$z_{7.2} = \frac{x-\mu}{\sigma} = \frac{7.2-6}{2} = 0.6$$

c) $x = 9.9$

$$z_{9.9} = \frac{x-\mu}{\sigma} = \frac{9.9-6}{2} = 1.95$$

d) $x = 0.8$

$$z_{0.8} = \frac{x-\mu}{\sigma} = \frac{0.8-6}{2} = -2.6$$

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

a) $z = 2.33$

$$area\ to\ the\ left\ of\ 2.33 = normalcdf(lower = -E99, upper = 2.33, \mu = 0, \sigma = 1) = 0.99$$

A z-score of 2.33 is in the 99th percentile.

b) $z = -0.83$

$$area\ to\ the\ left\ of -0.83 = normalcdf(lower = -E99, upper = -0.83, \mu = 0, \sigma = 1) = 0.203$$

A z-score of -0.83 is in the 20th percentile.

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?

$$z = \frac{x - \mu}{\sigma}$$

$$2.5 = \frac{x - 50}{10}$$

$$25 = x - 50$$

$$x = 75$$

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?

	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704

Method 1: z-score

$$z_{200} = \frac{x-\mu}{\sigma} = \frac{200-174}{35} = 0.74$$

From table:

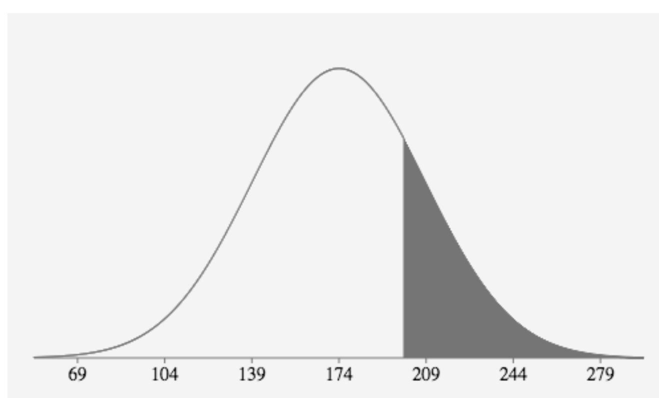
Area to right of 0.74 = 1 – area to left of 0.74 = 1 – 0.7704 = 0.2296

About 22.96% of her games she has a score more than 200 points.

Method 2: Calculator

Area to the right of 200 $= normalcdf(lower = 200, upper = E99, \mu = 174, \sigma = 35) = 0.2288$

About 22.88% of her games she has a 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?

Method 1: z-score

We are looking for the x value in the 90th percentile. From the table, the z-score with area to the left closest to 0.90 is 1.28.

$$z = \frac{x - \mu}{\sigma}$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997

$$1.28 = \frac{x - 170}{11}$$

$$14.08 = x - 170$$

$x = 184.08$ *Adrian needs a score of about 184 to get a spot in the all-star game.*

Method 2: Calculator

x-value in 90th percentile = $invnorm(area = 0.9, \mu = 170, \sigma = 11) = 184.1$

Adrian needs a score of about 184 to get 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?

Method 1: z-score

$$z_{140} = \frac{140 - 100}{15} = 2.67$$

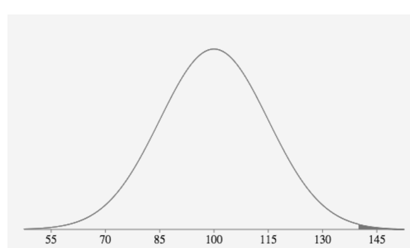
From the z-score table, the area to the right of 2.67 is = 1 – area to the left = 1 – 0.9962 = 0.0038

About 0.38% of the population is considered genius.

Method 2: calculator

% IQ > 140 = $normalcdf(lower = 140, upper = E99, \mu = 100, \sigma = 15) = 0.0038$

About 0.38% of the population is considered genius.

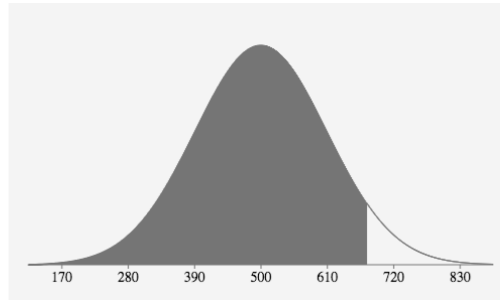


7) A standardized test is known to be normally distributed with a mean of 500 and a standard deviation of 110.

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

Method 1: z-score

$$z_{675} = \frac{675 - 500}{110} = 1.59$$



From the z-score table, the area to the left of 1.59 is 0.9441

She is in about the 94th percentile.

Method 2: Calculator

% scores < 675 = $normalcdf(lower = -E99, upper = 675, \mu = 500, \sigma = 110) = 0.9442$

She is in about the 94th percentile.

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

Method 1: z-score

We are looking for the x value in the 80th percentile. From the table, the z-score with area to the left closest to 0.80 is 0.84.

	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995

$$z = \frac{x - \mu}{\sigma}$$

$$0.84 = \frac{x - 500}{110}$$

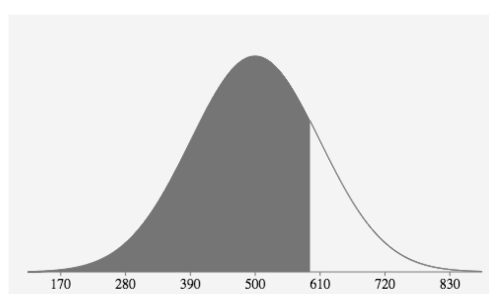
$$92.4 = x - 500$$

$x = 592.4$ *He must get a score of about 592 to be in the 80th percentile.*

Method 2: Calculator

x-value in 80th percentile = $invnorm(area = 0.8, \mu = 500, \sigma = 110) = 592.6$

He must get a score of about 593 to be in the 80th percentile.



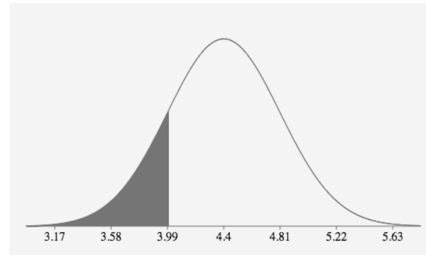
8) The weights of 75 model planes at a local convention are normally distributed. The average weight is 4.4 kg, with a standard deviation of 0.41 kg.

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

% planes < 4 = $normalcdf(lower = -E99, upper = 4, \mu = 4.4, \sigma = 0.41) = 0.1646$

planes < 4 = $0.1646 \times 75 = 12.345$

About 12 planes have a mass less than 4 kg.

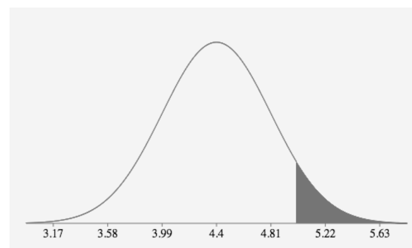


b) How many planes would be disqualified if it were against the rules to have a plane with a mass of more than 5 kg?

% planes > 5 = $normalcdf(lower = 5, upper = E99, \mu = 4.4, \sigma = 0.41) = 0.0717$

planes > 5 = 0.0717×75 = 5.38

Approximately 5 planes would be disqualified.

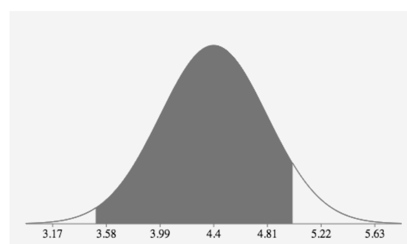


c) How many planes have a mass between 3.5 kg and 5 kg?

% 3.5 < planes < 5 = $normalcdf(lower = 3.5, upper = 5, \mu = 4.4, \sigma = 0.41) = 0.9142$

3.5 < planes < 5 = 0.9142×75 = 68.57

Approximately 69 planes are between 3.5 and 5 kg.

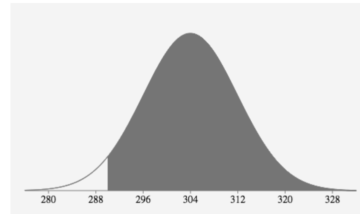


9) On the driving range, Tiger Woods practices his swing with a particular club by hitting many, many balls. Suppose that wen Tiger hits his driver, the distance the ball travels follows a normal distribution with mean 304 yards and standard deviation 8 yards.

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

% drives > 290 = $normalcdf(lower = 290, upper = E99, \mu = 304, \sigma = 8) = 0.9599$

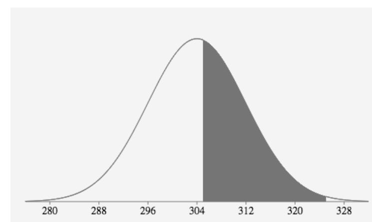
Approximately 95.99% of his drives travel at least 290 yards.



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

% 305 < drives < 325 = $normalcdf(lower = 305, upper = 325, \mu = 304, \sigma = 8) = 0.4459$

Approximately 44.59% of his drives travel between 305 and 325 yards.



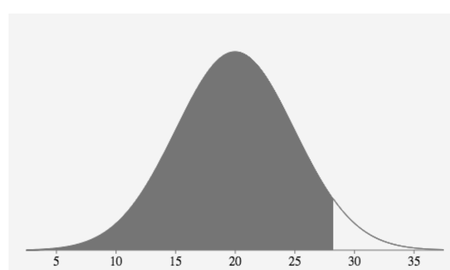
10) For the distribution $X \sim N(3, 0.74^2)$, determine the percent of the data that is within the given interval

a) $X > 2.44 = normalcdf(lower = 2.44, upper = E99, \mu = 3, \sigma = 0.74) = 0.7754 = 77.54\%$
b) $1.8 < X < 2.3 = normalcdf(lower = 1.8, upper = 2.3, \mu = 3, \sigma = 0.74) = 0.1196 = 11.96\%$
c) $X < 1.91 = normalcdf(lower = -E99, upper = 1.91, \mu = 3, \sigma = 0.74) = 0.0704 = 7.04\%$

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?

Length of fish in the 95th percentile $= invnorm(area = 0.95, \mu = 20, \sigma = 5) = 28.2$

The 95th percentile length of fish is about 28.2 cm.



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	$P(X < 9) = normalcdf(lower = -E99, upper = 9, \mu = 12, \sigma = 3) = 0.1587$
30	5	$P(X < 25) = normalcdf(lower = -E99, upper = 25, \mu = 30, \sigma = 5) = 0.1587$
5	2.2	$P(X > 6) = normalcdf(lower = 6, upper = E99, \mu = 5, \sigma = 2.2) = 0.3247$
245	18	$P(233 < X < 242) = normalcdf(lower = 233, upper = 242, \mu = 245, \sigma = 18) = 0.1813$

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

Ty Cobb: $z_{0.42} = \frac{x-\mu}{\sigma} = \frac{0.42-0.266}{0.0371} = 4.15$

Ted Williams: $z_{0.406} = \frac{x-\mu}{\sigma} = \frac{0.406-0.267}{0.0326} = 4.26$

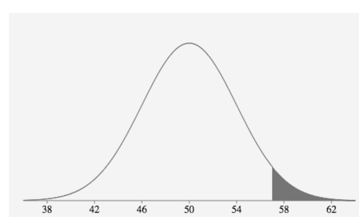
George Brett: $z_{0.39} = \frac{x-\mu}{\sigma} = \frac{0.39-0.261}{0.0317} = 4.07$

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

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.

$P(rainfall > 57) = normalcdf(lower = 57, upper = E99, \mu = 50, \sigma = 4) = 0.0401$

There is about a 4.01% chance of floods happening in a randomly selected year.

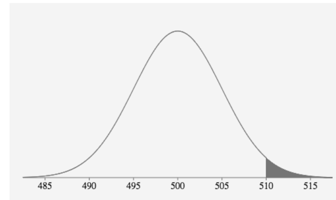


4) The weight of food packed in certain containers is a normally distributed random variable with a mean weight of 500 pounds and a standard deviation of 5 pounds. If a container is picked at random, find the probability that it contains:

a) more than 510 pounds

$P(pounds > 510) = normalcdf(lower = 510, upper = E99, \mu = 500, \sigma = 5) = 0.0228$

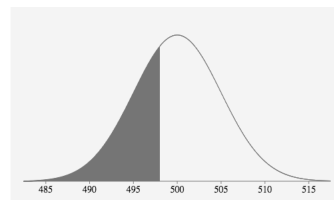
There is about a 2.285% chance a randomly picked container has more than 510 pounds of food.



b) less than 498 pounds

$P(pounds < 498) = normalcdf(lower = -E99, upper = 498, \mu = 500, \sigma = 5) = 0.3446$

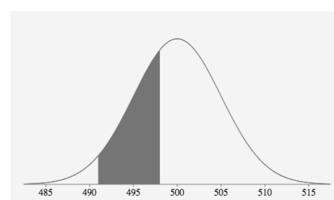
There is about a 34.46% chance a randomly picked container has less than 498 pounds of food.



c) between 491 and 498 pounds

$P(491 < pounds < 498) = normalcdf(lower = 491, upper = 498, \mu = 500, \sigma = 5) = 0.3086$

There is about a 30.86% chance a randomly picked container has between 491 and 498 pounds of food.

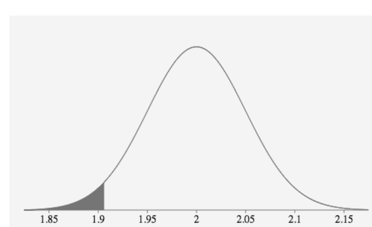


5) The diameter of a lead shot has a normal distribution with a mean diameter equal to 2 inches and a standard deviation equal to 0.05 inches. Find what diameter a circular hole should be so that only 3 percent of the lead shots can pass through it.

We are looking for the size of a lead shot in the 3rd percentile. A lead shot in the 3rd percentile is bigger than 3% of all lead shots. This means 3% of lead shots will be smaller and therefore be able to fit through a hole of that size.

3rd percentile = $invnorm(area = 0.03, \mu = 2, \sigma = 0.05) = 1.906$ inches

The hole should be about 1.906 inches in diameter so that only 3% of the lead shots can pass through it.

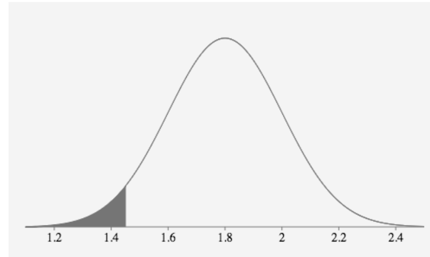


6) The nicotine content in a brand of king-size cigarettes has a normal distribution with a mean content of 1.8 mg and a standard deviation of 0.2 mg. Find the probability that the nicotine content of a randomly selected cigarette of this brand will be:

a) less than 1.45 mg

$P(nicotine < 1.45) = normalcdf(lower = -E99, upper = 1.45, \mu = 1.8, \sigma = 0.2) = 0.0401$

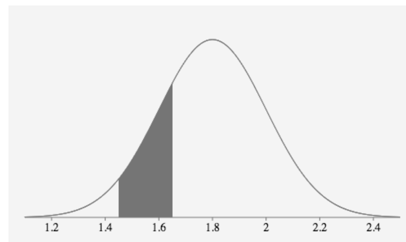
There is about a 4.01% chance a randomly selected cigarette has less than 1.45 mg of nicotine.



b) between 1.45 and 1.65 mg

$P(1.45 < nicotine < 1.65) = normalcdf(lower = 1.45, upper = 1.65, \mu = 1.8, \sigma = 0.2) = 0.1866$

There is about an 18.66% chance a randomly selected cigarette has between 1.45 and 1.65 mg of nicotine.

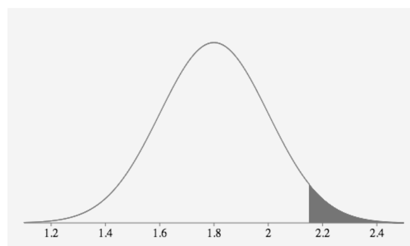


c) more than 2.15 mg.

$P(nicotine > 2.15) = normalcdf(lower = 2.15, upper = E99, \mu = 1.8, \sigma = 0.2) = 0.0401$

There is about a 4.01% chance a randomly selected cigarette has more than 2.15 mg of nicotine.

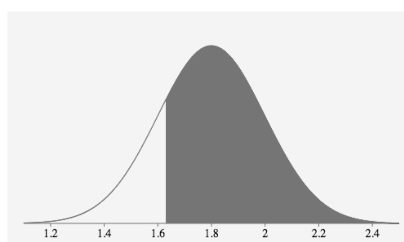
Note: We could have used the property that normal distributions are symmetrical and the answer from part a) to determine this probability.



d) What value is needed so that 80 percent of the cigarettes will exceed it in their nicotine content?

20^{th} percentile of nicotine content = $invnorm(area = 0.2, \mu = 1.8, \sigma = 0.2) = 1.63$

80% of cigarettes have more than about 1.63 mg of nicotine.

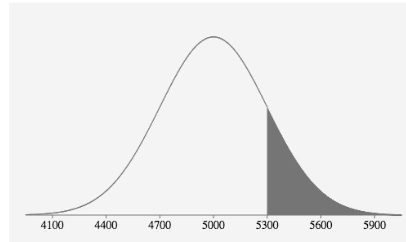


7) The demand for meat at a grocery store during any week is approximately normally distributed with a mean demand of 5000 pounds and a standard deviation of 300 pounds.

a) If the store has 5300 pounds of meat, what is the probability that they will run out during a random week?

$P(meat > 5300) = normalcdf(lower = 5300, upper = E99, \mu = 5000, \sigma = 300) = 0.1587$

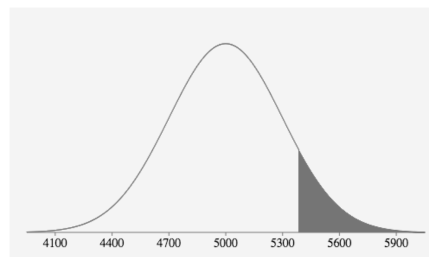
There is about a 15.87% chance the grocery store will run out of meat.



b) How much meat should the store have in stock per week so as to only run short 10 percent of the time?

90^{th} percentile for amount of meat sold = $invnorm(area = 0.9, \mu = 5000, \sigma = 300) = 5384.5$

They should stock about 5384.5 pounds of meat in order to only run short 10% 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

a)

Method 1: Formula

$C.I.$ for $\mu = 215 \pm 1.96\left(\frac{8}{\sqrt{40}}\right)$

$C.I. = 215 \pm 2.479$

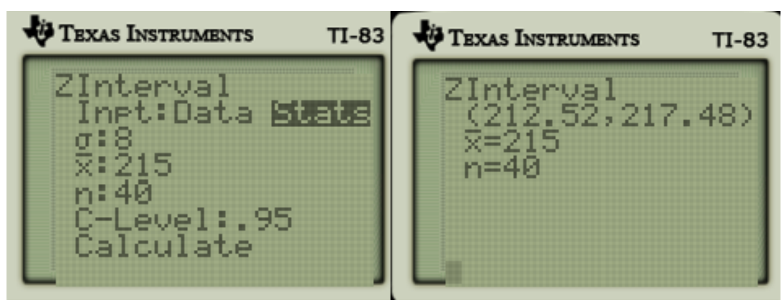
$C.I. = (212.5, 217.5)$

We can be 95% confident that the interval 212.5 to 217.5 captures the true population mean.

Method 2: Calculator

$\sigma = 8$
$\bar{x} = 215$
$n = 40$
C-Level: 0.95

C.I. = (212.52, 217.48)



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

b)

Method 1: Formula:

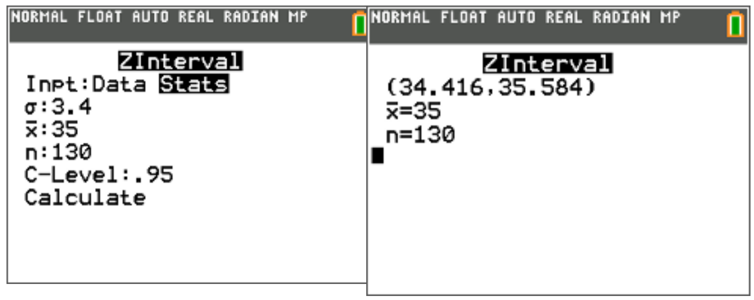
$C.I.\text{ for }\mu = 35 \pm 1.96\left(\frac{3.4}{\sqrt{130}}\right)$

$C.I. = 35 \pm 0.5845$

$C.I. = (34.4, 35.6)$

We can be 95% confident that the interval from 34.4 to 35.6 captures the true population mean.

Method 2: Calculator:



$\sigma = 3.4$
$\bar{x} = 35$
$n = 130$
C-Level: 0.95

$C.I. = (34.416, 35.584)$

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

c)

Method 1: Formula:

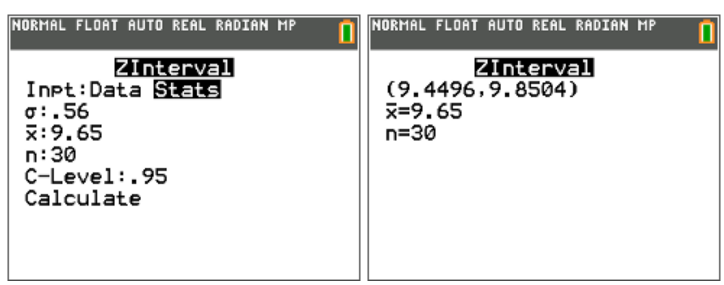
$C.I.\ for\ \mu = 9.65 \pm 1.96\left(\frac{0.56}{\sqrt{30}}\right)$

$C.I. = 9.65 \pm 0.2$

$C.I. = (9.45, 9.85)$

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

Method 2: Calculator:



$\sigma = 0.56$
$\bar{x} = 9.65$
$n = 30$
C-Level: 0.95

$C.I. = (9.45, 9.85)$

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

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?

Method 1: Formula

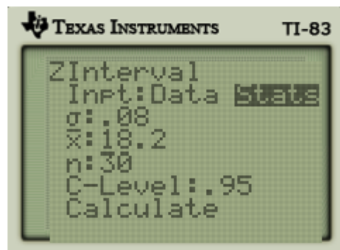
$$C.I.\, for\ \mu = 18.2 \pm 1.96\left(\frac{0.08}{\sqrt{30}}\right)$$

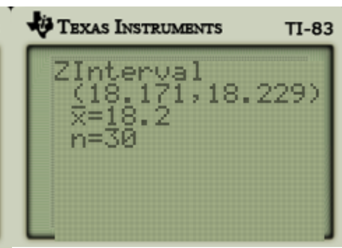
$$C.I. = 18.2 \pm 0.0286$$

$$C.I. = (18.17, 18.23)$$

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

Method 2: Calculator





$\sigma = 0.08$
$\bar{x} = 18.2$
$n = 30$
C-Level: 0.95

$$C.I. = (18.17, 18.23)$$

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

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%.

Method 1: Formula

$$C.I.\, for\ \mu = 18.8 \pm 1.96\left(\frac{0.3}{\sqrt{20}}\right)$$

$$C.I. = 18.8 \pm 0.1315$$

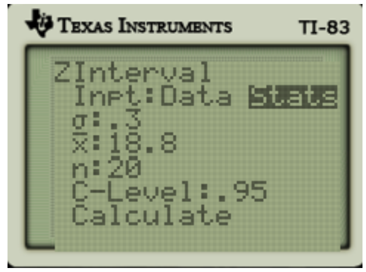
$$C.I. = (18.67, 18.93)$$

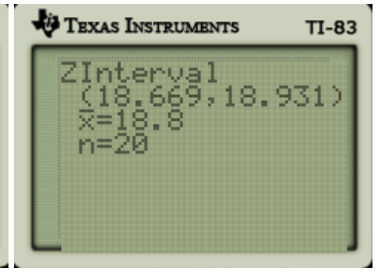
We can be 95% confident that the interval from 18.67 L to 18.93 L captures the true mean volume.

Method 2: Calculator

$\sigma = 0.3$
$\bar{x} = 18.8$
$n = 20$
C-Level: 0.95

$C.I. = (18.67, 18.93)$





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

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.

Larry

Method 1: Formula

$$C.I.\, for\ p = 0.48 \pm 1.96\left(\frac{\sqrt{0.48 \cdot 0.52}}{\sqrt{500}}\right)$$

$C.I. = 0.48 \pm 0.0438$

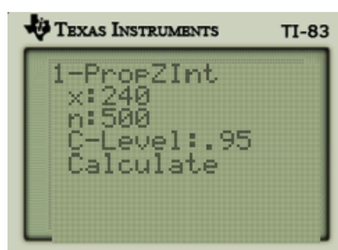
$C.I. (0.4362, 0.5238)$

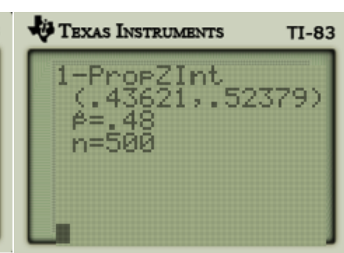
We can be 95% confident that the interval from 0.4362 to 0.5238 captures the true proportion of voters who support Larry.

Method 2: Calculator

$x = 0.48 \times 500 = 240$
$n = 500$
C-Level = 0.95

$C.I. (0.4362, 0.5238)$





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

Constance

Method 1: Formula

$$C.I.\,for\,p = 0.46 \pm 1.96\left(\frac{\sqrt{0.46 \cdot 0.54}}{\sqrt{500}}\right)$$

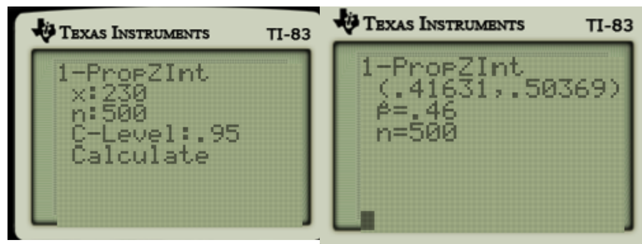
$$C.I. = 0.46 \pm 0.0437$$

$$C.I. = (0.4163, 0.5037)$$

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

Method 2: Calculator

$x = 0.46 \times 500 = 230$
$n = 500$
C-Level = 0.95



$$C.I. = (0.4163, 0.5037)$$

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

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.

Note: $\hat{p} = \frac{55}{300} = 0.183$

Method 1: Formula

$$C.I.\,for\,p = 0.183 \pm 2.576\left(\frac{\sqrt{0.183 \cdot 0.817}}{\sqrt{300}}\right)$$

$$C.I. = 0.183 \pm 0.0575$$

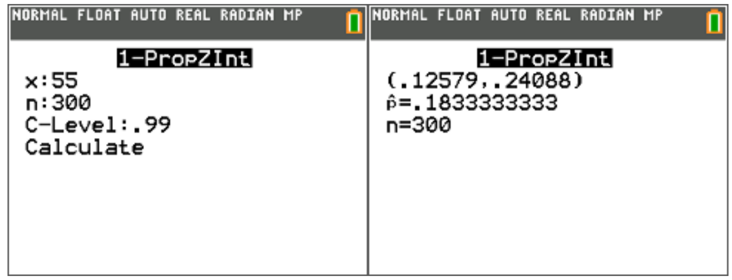
$$C.I. = (0.1255, 0.2405)$$

We can be 99% confident that the interval from 0.1255 to 0.2405 captures the true population proportion of people in Toronto who have chosen the Leafs to be their favourite team.

Method 2: Calculator

$x = 55$
$n = 300$
C-Level = 0.99

$C.I. = (0.1258, 0.2409)$



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.