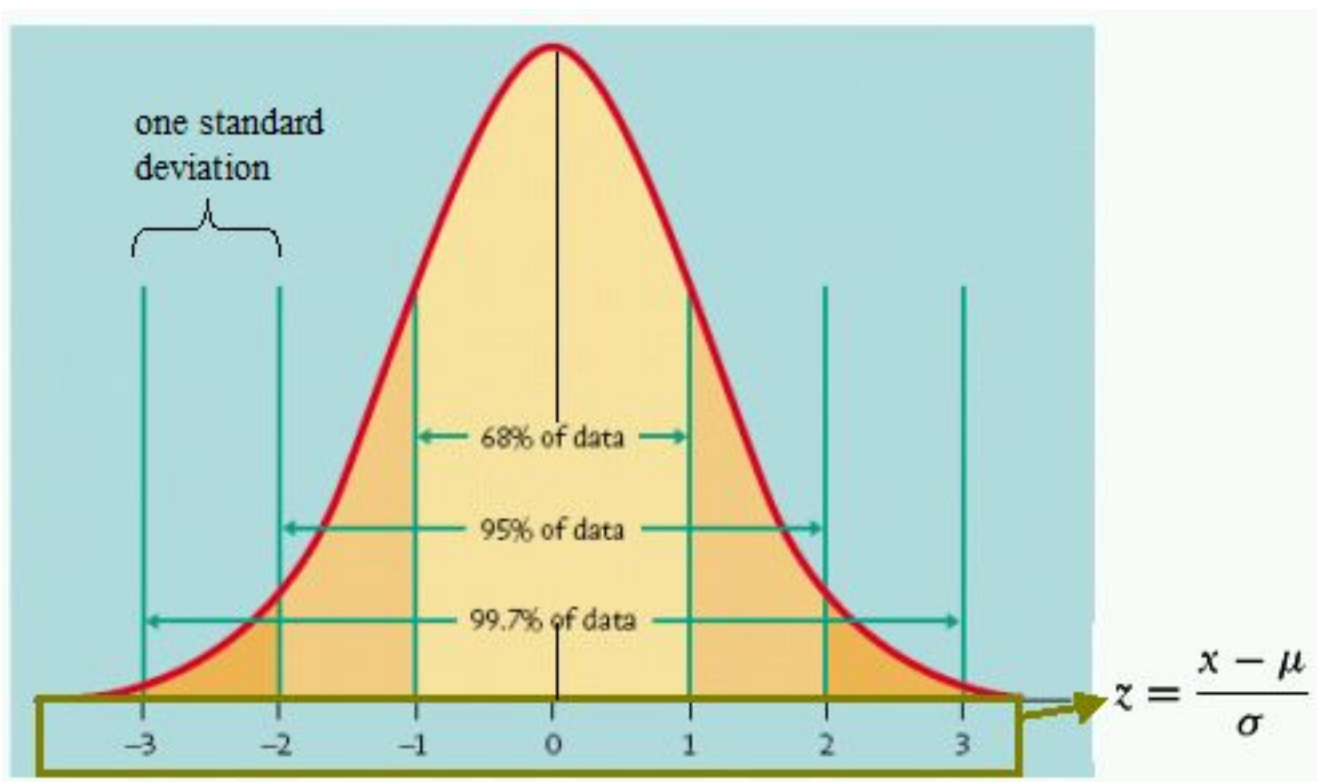
Chapter 3 - Lessons

Analyzing Univariate Data

MDM4U



Unit Outline

Section	Subject	Homework Notes	Lesson and Homework Complete (initial)
3.1	Shapes of Distributions		
3.2	Measures of Central Tendency		
3.3	Measures of Spread		
3.4	Normal Distribution		
3.5	Applications of the normal distribution		
3.6	Confidence Intervals for Population Means		
3.7	Confidence Intervals for Population Proportions		

Unit Performance

Homework Completion: None Some Most All

Days absent:______

Test Review Complete? None Some All

Assignment Mark (%):______

Test Mark (%):______

Notes to yourself to help with exam preparation:

Section 3.1 – Shapes of Distributions

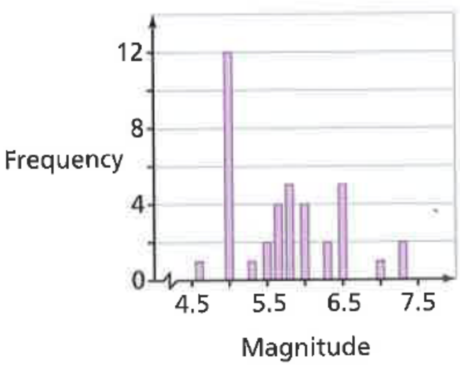
MDM4U
Jensen

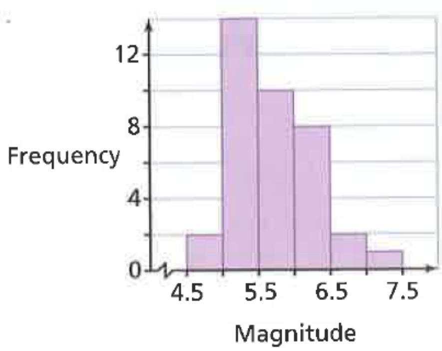
Part 1: Histogram Review

Example 1: Earthquakes are measured on a scale known as the Richter Scale. There data are a sample of earthquake magnitudes in Canada between 1960 and 1965.

5.0	5.0	6.4	5.0	6.0	5.6	6.5	6.5	5.0	5.5
6.4	7.2	5.0	5.7	5.6	5.0	5.0	5.0	5.0	5.7
5.0	7.0	5.5	5.2	4.6	6.3	7.2	6.0	5.4	5.8
6.0	5.7	6.5	5.0	5.7	5.0	5.6	6.0	5.6	6.2

What is wrong with how each of the following histograms display the above data?





Lets Make an Effective Histogram for the Data:

a) Determine the range of the data

b) Determine an appropriate bin (interval) width that will divide the data into 6 intervals.

> Note:
>
> Round your range UP to a value that can be divided easily.

c) Determine the first value of your first interval

We added _____ to 2.6 when we rounded our range, therefore we should subtract _____ from our smallest value _____; which makes our starting point _____.

However, some data will still fall on the border of the intervals, so we should add a decimal place by subtracting .05 from our starting point.

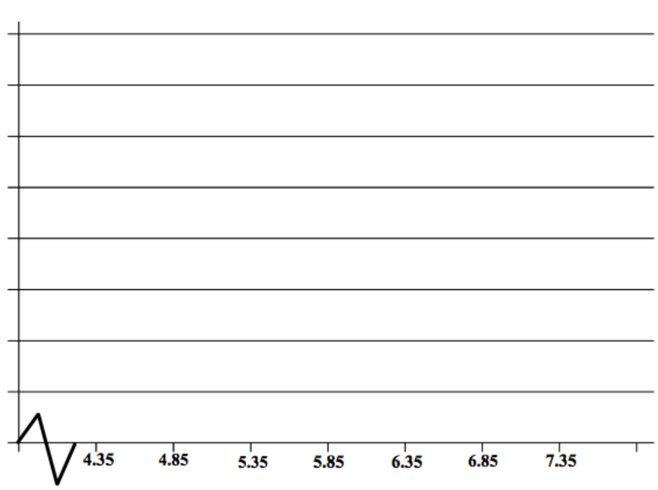
> **Note:**
>
> 1. If you have rounded your range up you should subtract half of the amount you rounded from the smallest value to evenly distribute the 'excess of your range'.
>
> **2. Make sure no data points lie on the border of two intervals. (Do this by subtracting .5 from a whole number, .05 from data with one decimal point, .005 from data with two decimal points and so on)**

d) Create a frequency table using your intervals

> Notice that the number one interval ends with, the next interval starts with the same number. This is because the data for a histogram is continuous!!!

Class Interval	Frequency

e) Create a histogram of the data



Part 2: The Shape of a Distribution

Step back from a histogram. What can you say about the distribution? When you describe a distribution, you should always comment about three things: its shape, center, and spread. In this lesson we will focus on shape.

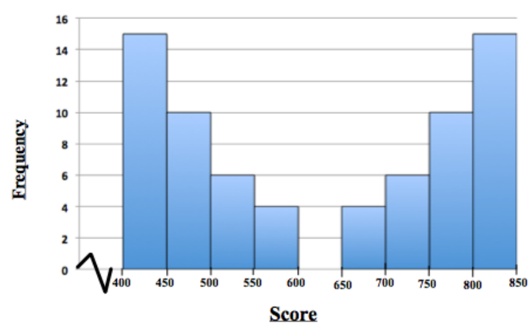
The mode of a histogram is an important characteristic that is often used in describing its shape. The mode of a histogram is the interval with the highest frequency. Does the histogram have a single peak, central peak, or several separated peaks? These peaks are called modes.

The shape of a distribution is generally described in one of four ways:

Note: *A graph is roughly symmetric if the right and left sides of the graph are approximately* ________________________ *of each other.*

1. U-Shaped Distribution

The scores from the game of spider solitaire form this type of distribution.



Score	400–449	450–499	500–549	550–599	600–649	650–699	700–749	750–799	800–849
Frequency	15	10	6	4	0	4	6	10	15

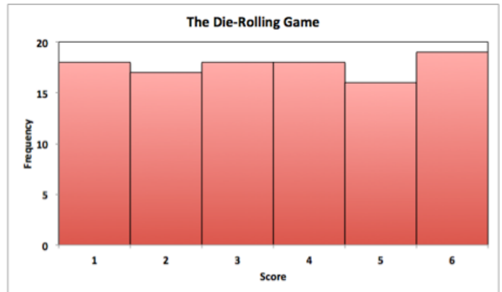
- A U-shaped distribution occurs when there are _______________ at either end of the range

- Because it has two peaks, it can also be described as a __________________ distribution

Can you think of another example of a frequency distribution that would be U-shaped (bimodal)?

2. Uniform Distribution

This is the distribution you would expect from an experiment such as rolling a single die.



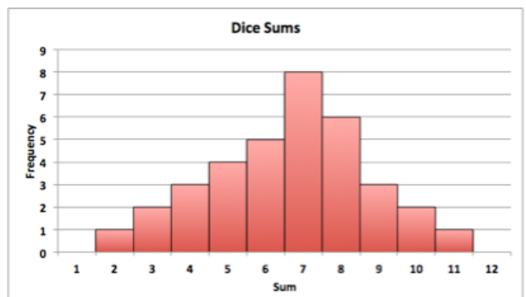
Score	1	2	3	4	5	6
Frequency	18	17	18	18	16	19

- When each outcome has a ________________ frequency, it is called a uniform distribution. The height of each bar is roughly _______________.

- Notice, there doesn't appear to be any one single mode.

3. Mound Shaped Distribution

Rolling a pair of dice and recording the sum results in this type of distribution.



Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	1	2	3	4	5	8	6	3	2	1	0

- In this distribution, there is an interval with the greatest frequency ______________________, and the frequencies of all other intervals __________________ on either side of that

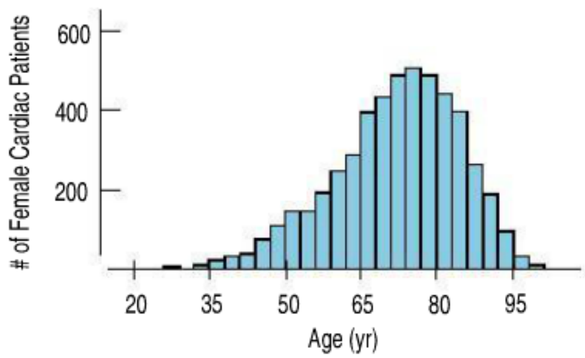
- The frequency distribution takes on a mound (or bell) shape. It can also be described as ______________________ since it has one clear peak (mode).

Do you notice any similarities between the first 3 shapes of distributions?

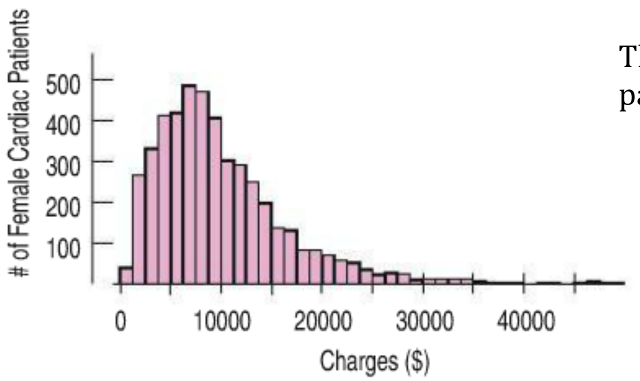
4. Skewed

The thinner ends of a distribution are called the tails. If one tail stretches out farther than the other, the histogram is said to be skewed to the side of the _________________ tail.

Another way to say it is that the interval or group of intervals with the highest frequencies are near one end of the histogram. As a result, the distribution seems to tail off to the left or right.



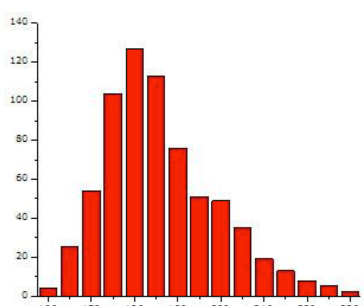
This distribution of ages of female heart attack patients is _____________ skewed.

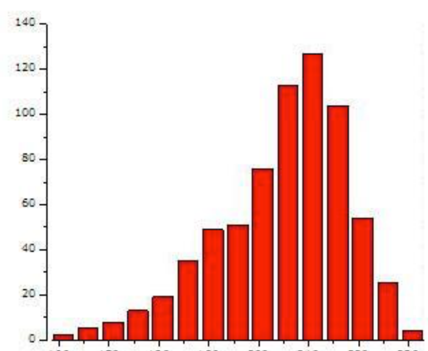


This distribution of cost of treatment for heart attack patients is _______________ skewed.

Tip: *If you get mixed up between left and right skewed……….look at your toes!*

Why call it left or right skewed?





Section 3.2 – Measures of Central Tendency

MDM4U
Jensen

Part 1: Video

This video will review shape of distributions and introduce measures of central tendency. Answer the following questions while watching.

http://www.learner.org/courses/againstallodds/unitpages/unit04.html

1. What variable is examined in comparing men and women workers at the beginning of the video?

2. Would you describe the shape of the distribution of men's weekly wages as symmetric, skewed to the left or skewed to the right?

3. What is the most important difference between the distributions of weekly wages for men and for women?

4. Would a few very large incomes pull the mean of a group of incomes up, down, or leave the mean unaffected?

5. What happens to the median of the set of data of salaries when the president decides to double his salary?

- In this section, you will learn how to describe a set of numeric data using a single value
- The value you calculate will describe the _______________ of the set of data
- The 3 measures of central tendency are:

Part 2: The Mean

The mean: a measure of central tendency found by dividing the _________________________by the ____. In statistics, it is important to distinguish between the mean of a population and the mean of a sample of that population. The sample mean will approximate the actual mean of the population, but the two means could have different values. Different symbols are used to distinguish the two kinds of means.

Population Mean	Sample Mean
$\mu = \frac{\sum x}{N}$	$\bar{x} = \frac{\sum x}{n}$
μ – mu; population mean	x – x-bar; sample mean
$\sum$ - sigma; the sum of	$\sum$ - sigma; the sum of
N - number of values in the population	n - number of values in the sample

Example 1: A group of elementary school children were asked how many pets they have. Here are their responses, arranged from lowest to highest:

1 3 4 4 4 5 7 8 9

What is the mean number of pets for this group of children?

What does this number tell us? One way to think of it would be that if every child in the group had the same number of pets, each would have _____ pets. This is the 'fair share' value.

Part 2: Weighted Mean

Sometimes, certain data within a set are more ________________ than others. For example, the mark on a final exam is often considered to be more important than the mark on a term test for determining an overall grade for a course. A ___________________________ gives a measure of central tendency that reflects the relative importance of the data.

Weighted mean formula:

Example 2: The personnel manager for a marketing company 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. Determine the weighted mean score for this job applicant.

Criterion	Score, x_i	Weighting Factor, w_i
Education	4	2
Job experience	2	2
Interpersonal skills	5	3
Communication skills	5	3
References	4	1

Example 3: The table below shows a student's performance in an MDM4U class. What would their final mark be based on the weightings shown?

Category	Mark, x	Weighting, w
Assignments	92	15%
Tests	86	40%
ISU	85	15%
Exam	88	30%

Part 3: Mean of Grouped Data

Supposed your data have already been organized into a frequency table with ______________________.

You no longer have actual data values, so you must then use the ________________________ of each class to estimate a mean weighted by the frequency.

Finding the average (mean) of grouped data is the same as finding a weighted average; except that you have to use the _________________________________as the data value.

Formula for mean of grouped data:

Example 4: A sample of car owners was asked how old they were when they got their first car. The results were then reported in a frequency distribution. Calculate the mean.

Age	16 - 20	21 - 25	26 - 30	31 - 35	36 - 40
Frequency	10	18	12	8	2

Solution:

Age	**Frequency, f**	**Midpoint of age, m**	**$f \times m$**
16 – 20	10		
21 – 25	18		
26 – 30	12		
31 – 35	8		
36 – 40	2		

Part 4: Median

The median value is the ________________ data point in an ______________________ set, dividing the set into two sets of equal size.

To find the median of a distribution:

1. Arrange all observations in order of size, from smallest to largest
2. If the number of observations n is odd, the median is the middle most observation in the ordered list
3. If the number of observations n is even, the median is the average of the two middle most observations in the ordered list

Tip: *The middle most piece of data is in the* $\frac{n+1}{2}$ *position*

Example 5: Monthly rents downtown and in the suburbs are collected from the classified section of a newspaper. Calculate the median rent in each district

Downtown: 850, 750, 1225, 1000, 800, 1100, 3200
Suburbs: 750, 550, 900, 585, 220, 625, 500, 800

Start by ordering the sets of data…

Downtown:

Suburbs:

Downtown:

There are _____ elements in the set, so the median is the _______________ element. The median is __________/month.

Suburbs:

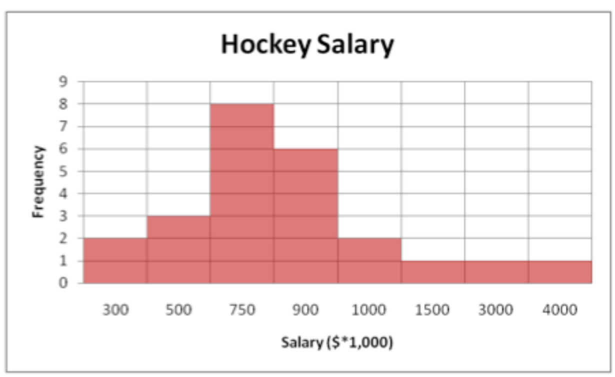
There are _____ elements in the set, so the median is halfway between the _____ and _____ element. Therefore the median is _______________/month.

Part 5: The mode

The mode is simply the _________________________value or range of values in a data set. It is easy to determine the mode from a histogram as it is the ____________________column.

Example 6: The table and histogram show the current salaries for a hockey team.

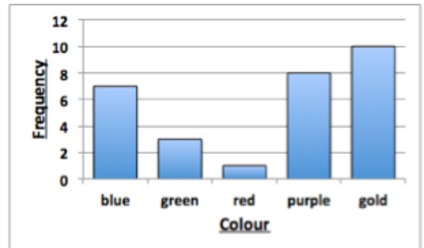
Salary ($)	Number of Players
300 000	2
500 000	3
750 000	8
900 000	6
1 000 000	2
1 500 000	1
3 000 000	1
4 000 000	1



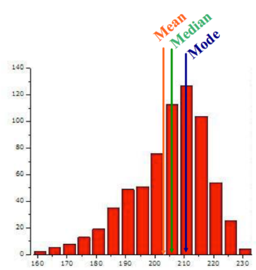
What is the mode of the salaries?

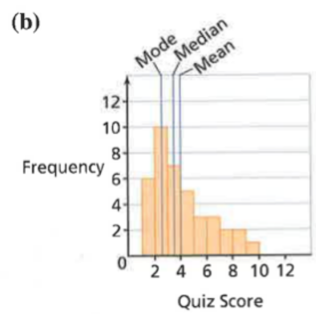
Note: If no measurement is repeated, the data set has no mode. If it has two measurements that occur most often, it is called bimodal and has two modes.

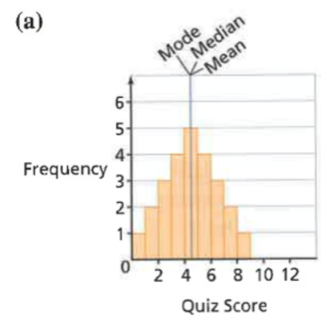
Note: If you have categorical data, the mode is the only appropriate measure of central tendency to use.



Skewed left: **Skewed right:** **Symmetrical:**







The mean is _____ the median

The mean is ________ the median

The mean, median, and mode are _____

Part 6: Application of Weighted Mean

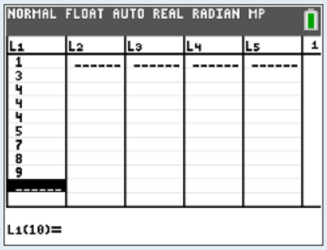
Before you start your homework, use the table below to determine what mark the student needs on the final exam to earn a final grade of 85%.

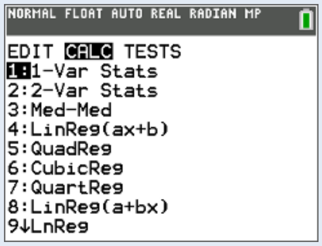
Category	Mark, x	Weighting, w
Assignments	92	15%
Tests	86	45%
ISU	85	10%
Exam	?	30%

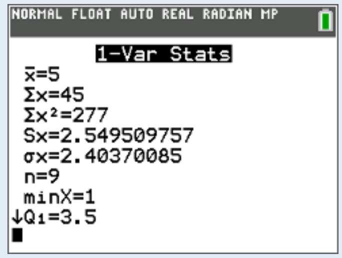
Using your Ti-84 to find the mean of a set of data

Example 1 done with Ti-84:

- Input data in to L1: STAT → EDIT
- Calculate mean: STAT → CALC → 1-VARSTATS → List: L1 → CALCULATE

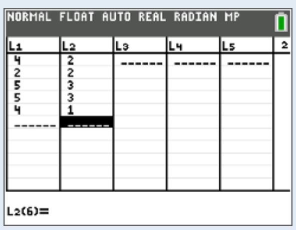


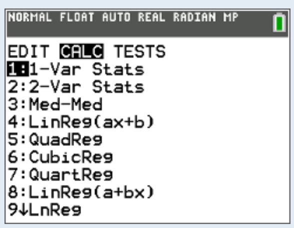


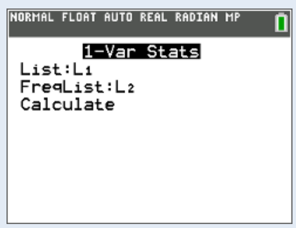


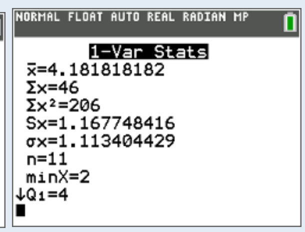
Example 2 done with Ti-84

- Input scores in to L1 and frequencies in L2: STAT → EDIT
- Calculate mean: STAT → CALC → 1-VARSTATS → List: L1, Frequencies: L2 → CALCULATE







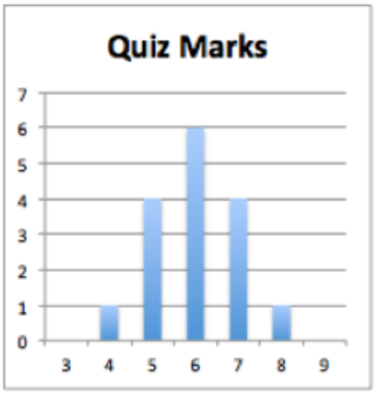


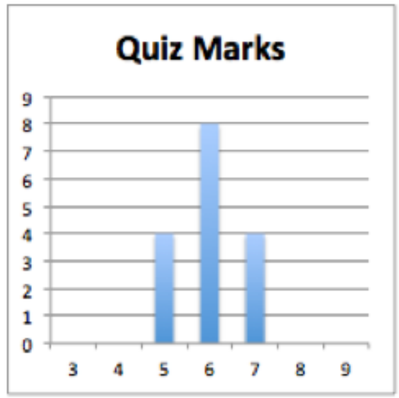
Section 3.3 – Measures of Spread

MDM4U
Jensen

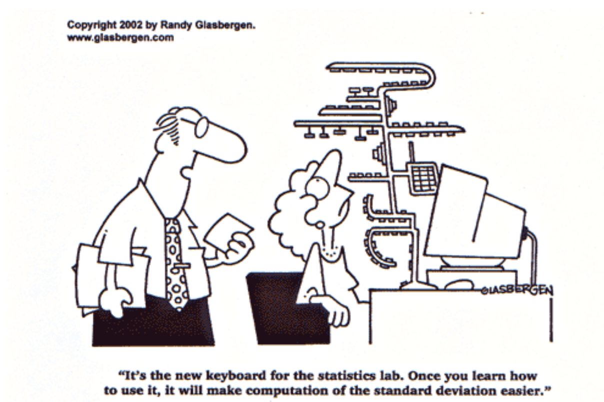
Part 1: Minds On

Describe the similarities and differences between the following two graphs:



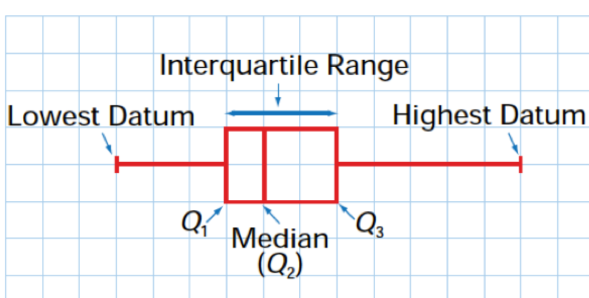


In the previous section we learned how to describe a set of data using measures of central tendency. Just as there are several measures of central tendency, there are also different ________________________________ for a set of data:



Part 2: Interquartile Range (IQR)

A ______________________ is a visual representation of data divided into four groups (quartiles) with equal numbers of values in each quartile.



The three dividing points are:

Q_1 –

Q_2 –

Q_3 –

The ______________________________is the range of the middle half of the data (_______________). The larger the interquartile range, the _______________ the spread of the central half of the data. This is a good measure of spread (dispersion) but can be tedious to calculate.

Remember: The median is the _______________ value in an ordered data set. If the set has an even number of data points, then the median is halfway between the two middle-most values.

Example 1: Here are the final grades in a grade 9 physical education class:

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

a) First put the data into numerical order

This is always the first step when determining the quartiles

b) Calculate the following statistics:

Range =

Median = Q2 =

Q1 =

Q3 =

IQR =

Note: although a quartile is, strictly speaking, a single value, people sometimes speak of datum being within a quartile. What they really mean is that the datum is in the quarter whose upper boundary is the quartile. Example: if a value *x* is within the second quartile, then $Q_1 < x \le Q_2$

c) If your final mark was 75%, what quartile were you within?

Part 3: Standard Deviation

The deviation of a piece of data is the ________________ it is from the ___________ of the set of data. If you were to take the all the deviations for an entire set of data, square each one of them, and then find the average, you would have what is called the variance. The square root of the variance is called the ________________________________.

Population Standard Deviation	Sample Standard Deviation
$\boldsymbol{\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}}$	$\boldsymbol{s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}}$
σ – lower case sigma; population standard deviation	$\boldsymbol{s}$ **–** sample standard deviation
$\sum$ - capital sigma; sum of	$\sum$ - capital sigma; sum of
μ - mu; population mean	$\bar{x}$ – x-bar; sample mean
N - number of items in population	n - number of items in sample

Example 2: The heights of all of the players on a basketball team are shown in the table below. Calculate the standard deviation of the population.

Start by calculating the mean:

Player	Height	Deviation, $\boldsymbol{x-\mu}$	$\boldsymbol{(x-\mu)^2}$
Laura	183		
Jamie	165		
Deepa	148		
Colleen	146		
Ingrid	181		
Justiss	178		
Sheila	154		

Part 4: Comparing Standard Deviations

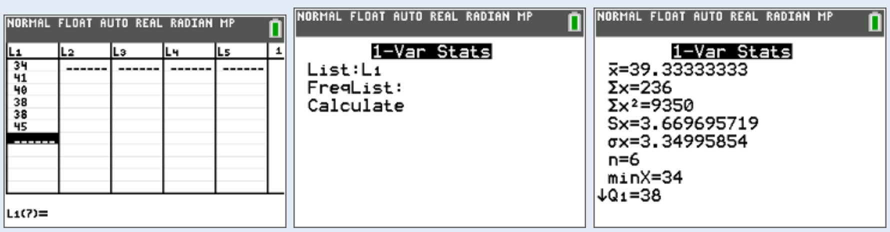
Example 3: Felix and Melanie have a job laying patio stones. Their boss in interested in who the better worker is so randomly throughout the week he chooses a few hours to record how many stones each of the workers lays. The data is recorded in the table below:

Felix	34	41	40	38	38	45
Melanie	51	28	36	44	41	46

Calculate the mean and standard deviation of each sample and compare use them to compare the two workers.

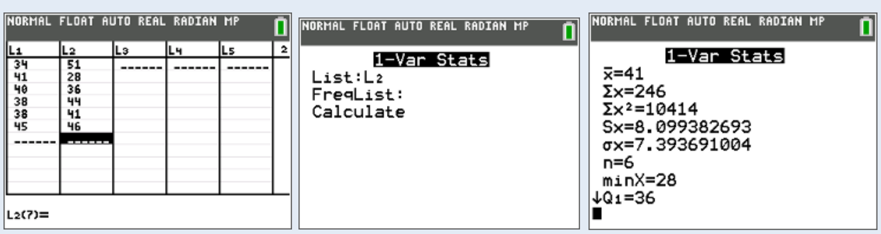
Using the Ti-84 to calculate these statistics:

- Input Felix's data in to L1: STAT → EDIT
- Calculate the statistics for Felix: STAT → CALC → 1-VARSTATS → List: L1 → Calculate



Statistics for Felix:

- Input Melanie's data in to L2: STAT → EDIT
- Calculate the statistics for Melanie: STAT → CALC → 1-VARSTATS → List: L2 → Calculate



Statistics for Melanie:

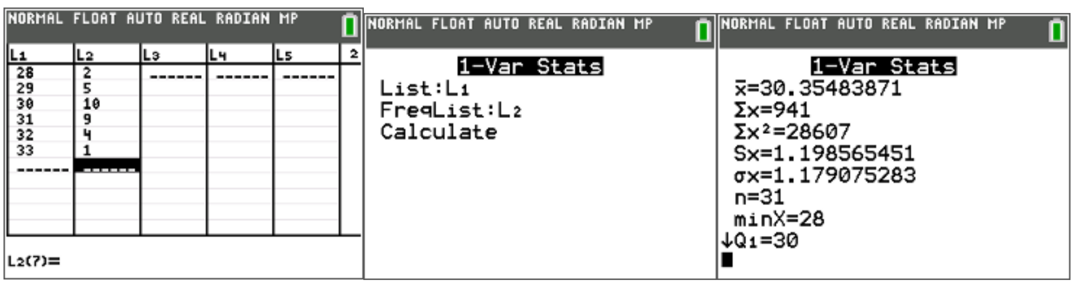
Part 5: Standard Deviation for Data with Weights/Frequencies

A railway line gives out small bags of peanuts to its travellers, and each bag doesn't always contain the same number of peanuts. The following table represents a sample of 31 bags showing the number of peanuts per bag. Calculate the sample standard deviation for the number of peanuts in each bag.

Number of Peanuts	28	29	30	31	32	33
Frequency	2	5	10	9	4	1

Solution:

- Input number of peanuts in to L1: STAT → EDIT
- Input frequencies in to L2
- Calculate statitsics: STAT → CALC → 1-VARSTATS → List: L1 → Frequencies: L2 → CALCULATE



Section 3.4 – Normal Distribution

MDM4U
Jensen

Part 1: Dice Rolling Activity

a) Roll two 6-sided number cubes 18 times. Record a tally mark next to the appropriate number after each roll. After rolling the cube 18 times, determine the frequency for each number by counting the tally marks.

Individual Data		
Sum of the Numbers Rolled	Tally Marks	Frequency
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		

b) Create a histogram or bar graph for your Individual Data for Activity 2.

c) Determine the value of each of the following for your data set.

Mean =
Median =
Mode =

Range =
Interquartile Range =
Standard Deviation =

Record the combined class data for Activity 2 in the table below.

Sum of the Numbers Rolled	Frequency
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

d) Create a histogram or bar graph for the Class Data for Activity 2.

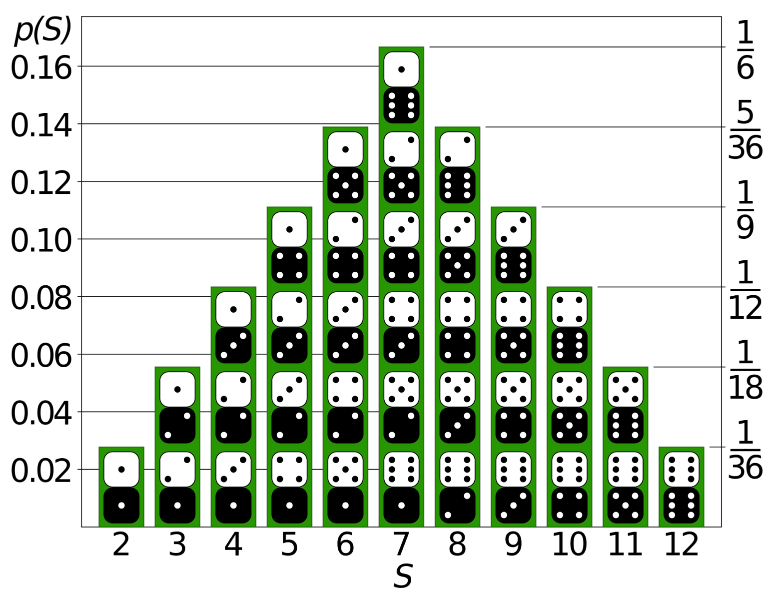
e) Determine the value of each of the following for the class data set.

Mean =
Median =
Mode =

Range =
Interquartile Range =
Standard Deviation =

f) Describe the frequency distribution of class set of data. What do you notice about shape, measures of central tendency, etc.?

Summary of activity: If you were to draw a smooth curve through the tops of the rectangles on the distribution of the class data, it would approximately form what is called a __________________________. Shown to the right is the theoretical probability distribution for rolling two dice. The more trials we complete, the closer our distribution would look to the actual probability distribution shown:



Part 2: Properties of a Normal Distribution

If you are to measure many similar things which have differences caused by a random variation, those results will typically be distributed ___________________________ and ___________________________about the mean. Statisticians observe this mound shaped curve so often that its mathematical model is known as the normal distribution. Distributions that are close to normal include: scores on tests taken by many people, repeated measurements of the same quantity (heights, weights), characteristics of biological populations (yields of corn), and chance outcomes (dice rolling example).

The normal curve was first used in the 1700's by French mathematicians and early 1800's by German mathematician and physicist Karl Gauss. The curve is known as the Gaussian distribution and is also sometimes called a bell curve.

General Properties of all Normal Distributions:

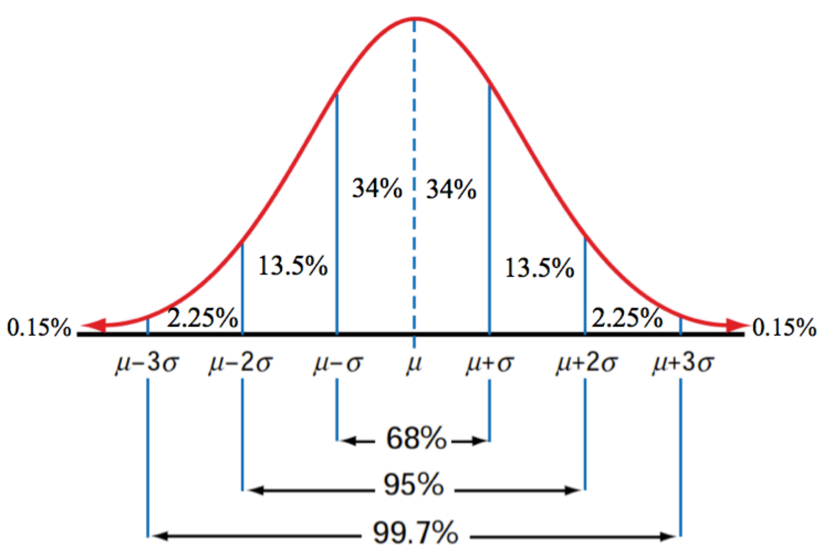
- It is ______________________ and ______________________; the mean, median and mode are ___________ and fall at the line of symmetry
- It is shaped like a ___________, peaking in the middle and sloping down toward the sides. It approaches ___________ at the extremes

The Empirical Rule:

Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean. We'll soon show how to find these numbers precisely; but one simple rule (the Empirical Rule) is usually all we need.

In any normal distribution…

- Approximately _____ of the observations fall within 1 standard deviation of the mean
- Approximately _____ of the observations fall within 2 standard deviations of the mean
- Approximately _________ of the observations fall within 3 standard deviations of the mean



Notation Used:

Any particular normal distribution is completely described by two numbers:

$$X \sim N(\bar{\mu},\ \sigma^2)$$

X -

N –

μ –

σ –

Example:

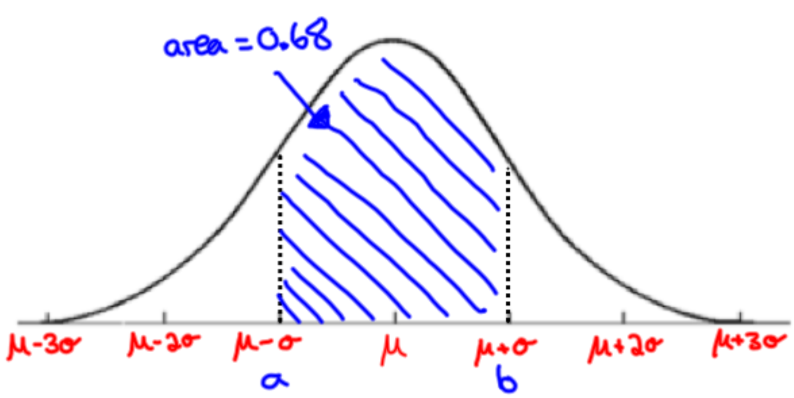
$X \sim N(50, 5^2)$

This normal distribution would have a mean of ______, a standard deviation of ______ and a variance of ______.

Part 3: Area Under a Normal Curve

The area under every normal curve equals ________. In any normal distribution, the percent of the data that lies between two specific values, *a* and *b*, is the area under the normal curve between endpoints *a* and *b*

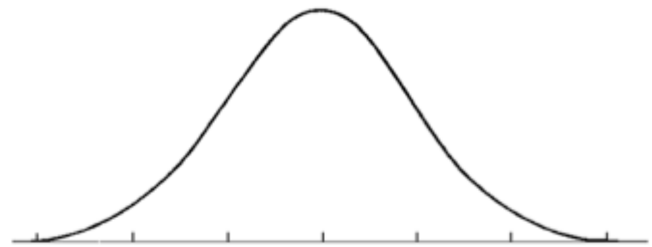
Example:



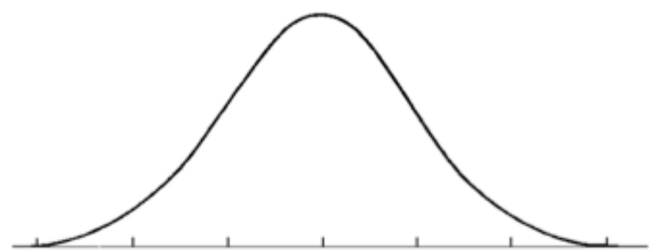
For ALL normal distribution problems you should sketch, label, shade the region of the curve, and answer the question.

Example 1: If $X \sim N(50, 5^2)$, shade in the area of the given interval on a normal curve and find the area of the shaded region using the Empirical Rule.

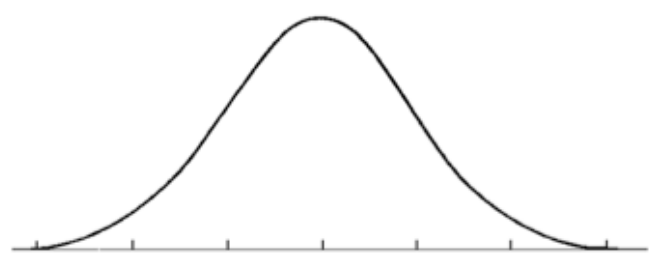
a) $x > 55$



b) $40 < x < 60$

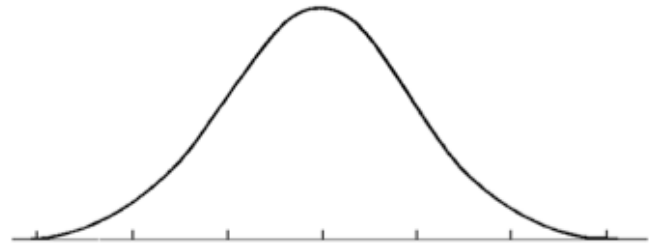


c) $x < 40$



Note: different methods get slightly different answers because we are using the Empricial Rule which has rounded values.

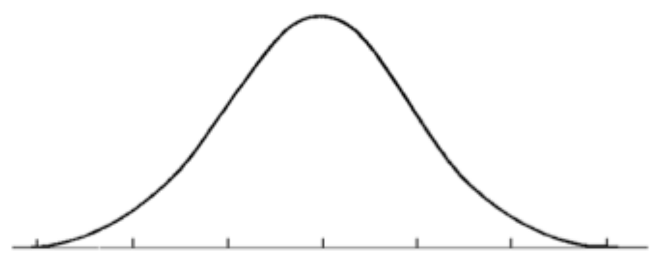
Example 2: A real estate agent, working entirely on commission, makes an average of $850 with a standard deviation of $260 weekly selling property in the city. If we assume this distribution is roughly normal, what is the probability that a real estate agent will make between $1110 and $1,370 selling property in the city?

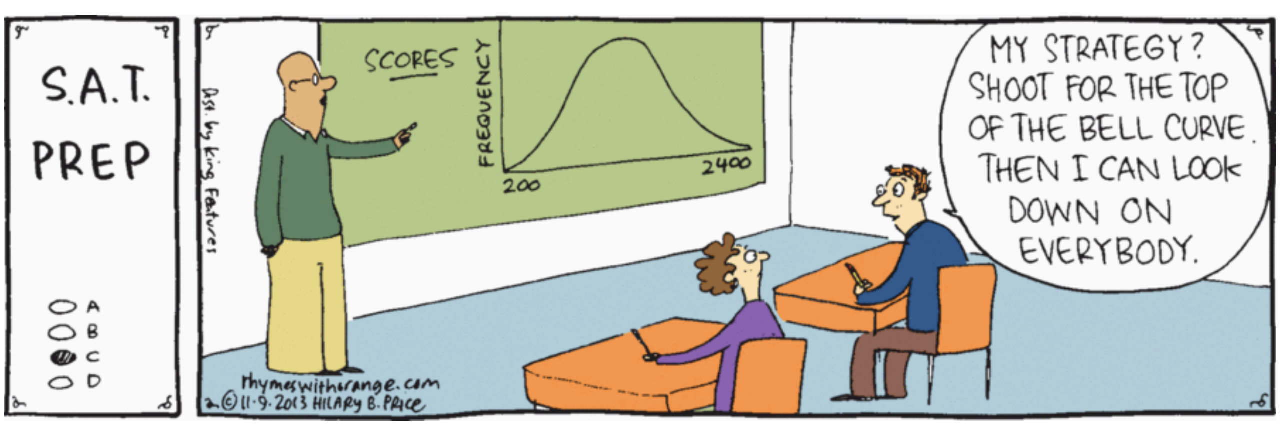


Example 3: Julie is an engineer who designs roller coasters. She wants to develop a ride that 95% of the population can ride. The average adult in North America has a mass of 71.8 kg, with a standard deviation of 13.6 kg.

a) Describe this information in the normal distribution notation

b) If she wanted to provide for 95% of the general population, what range of masses should she anticipate?





Section 3.5a – Applying the Normal Distribution

MDM4U
Jensen

Part 1: Normal Distribution Video

While watching the video, answer the following questions

1. What is another name for the Empirical rule?

2. How tall must a woman be to join the Boston Beanstalks Club?

3. How do you calculate a z-score?

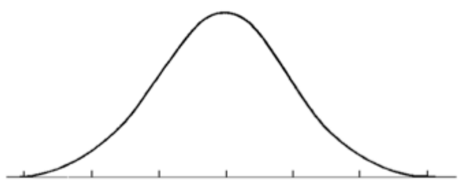
4. Based on z-scores, are eligibility requirements to join the Boston Beanstalks more difficult to meet for men or for women?

Part 2: The Standard Normal Distribution

All normal distributions are the same when we measure how many ________________________ an observation x lies away from the _____________, which we calculate as follows:

The standardized value for x, calculated by the formula above, is called its _____________. Observations from different normal distributions are best compared by comparing their standardized values, or z-scores. The z-score states how many _______________________________the original observation falls away from the ___________ and in which direction. Observations larger than the mean have __________________ z-scores, while observations smaller than the mean have _______________ z-scores. Converting to standardized values allows us to find proportions that we can't get from the Empirical Rule.

Example 1: Suppose we want to know the percentage of data from a normal distribution with mean $\mu = 8$ and standard deviation $\sigma = 2$ that falls below $x = 9$. We can use the Empirical Rule to learn that the percentage is between 50% and 84%, but that is not a very accurate estimate. Instead, we convert $x = 9$ into a z-score:

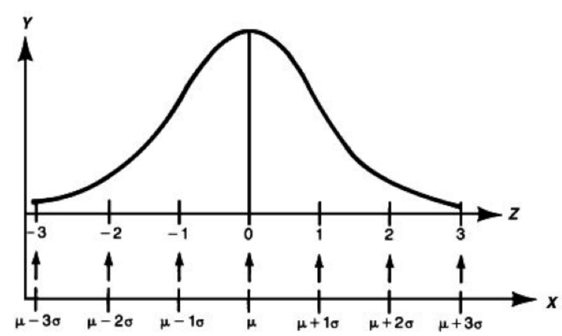


The standard normal graph shows how the value $x = 9$ is ______ standard deviations ____________ the mean. This value is in the same position on this graph as it was for the normal distribution graph.

	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

Now, we find the proportion of standard normal data that falls below 0.5. To find this proportion, we use a standard normal table, which gives the proportion of data that falls ____________ any value for z. Using the portion of a z-table, we look down the z-column to locate 0.5 and then move to the right under the .00 column. Our answer is ________, or around ____________.

Standardizing a data value by finding the z-score places it on the ____. This is a special normal distribution with a mean of ____ and a standard deviation of ____.



The Translation of X to Z by the Transformation $Z = (X - \mu)/\sigma$

Figure 3

Part 3: Practice Finding and Interpreting Z-Scores

Example 2: Caley scored 84% in her Data Management course, while Lauren, who attends a different Data Management class, scored 83%. If Caley's class average is 74% with a standard deviation of 8, and Lauren's class average is 70% with a standard deviation of 9.8, use z-scores to determine who has the better mark.

Note: *z-scores are used to standardize the data so that they can be accurately compared.*

Part 4: Area to Left (Percentile)

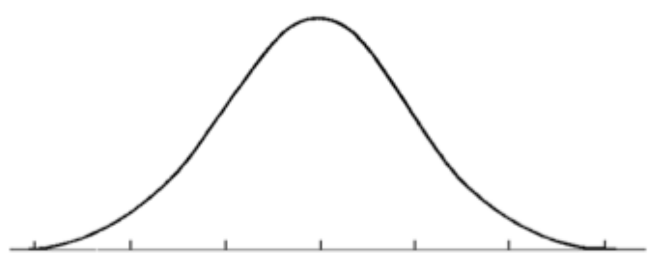
The proportion (percentage) of data with a ____________ z-score is equal to the __________________ the value can be categorized by. The areas given on the z-score table are considered percentiles because the z-score table always gives areas to the left of the given z-score.

Percentile: the kth percentile is the least data value that is greater than k% of the population.

For example; If your test result is in the 85th percentile, that means that 85% of tests had a lower score than yours.

Example 3: Perch in a lake have a mean length of 20cm and a standard deviation of 5cm. Find the percent of the population that is less than 22 cm.

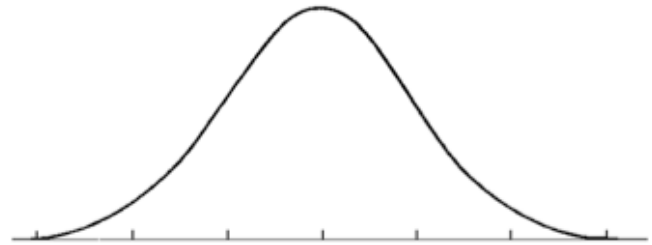
	0.00	0.01	0.0
0.0	0.5000	0.5040	0.50
0.1	0.5398	0.5438	0.54
0.2	0.5793	0.5832	0.58
0.3	0.6179	0.6217	0.62
0.4	0.6554	0.6591	0.6
0.5	0.6915	[illegible]	



Part 5: Area to the Right

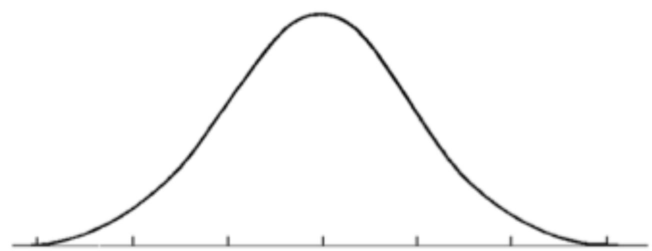
Example 4: Perch in a lake have a mean length of 20cm and a standard deviation of 5cm. Find the percent of the population that is greater than to 17 cm.

−0.8	0.2119	0.2090	0.2061	0.2033
−0.7	0.2420	0.2389	0.2358	0.2327
−0.6	0.2743	0.2709	0.2676	0.2643
−0.5	0.3085	0.3050	0.3015	0.2981
−0.4	0.3446	0.3409	0.3372	0.3336
−0.3	0.3821	0.3783	0.3745	0.3707



Part 6: Area Between Two Values

Example 5: Using the normal distribution $X \sim N(7, 2.2^2)$, find the percent of data that is within the interval $3 < x < 6$



Part 7: Percentile to Scores: z in Reverse

Finding areas from z-scores is the simplest way to work with the normal distribution. But sometimes we start with areas and are asked to work backward to find the corresponding z-score or even the original data value.

Example 6: SAT test scores have a mean of 500 and a standard deviation of 100. Suppose a college says it only admits students who have scores in at least the 90^{th} percentile. How high a score does it take to be eligible?

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

<u>**Using the Ti-84**</u>

Using ti-83 for a Normal Distribution

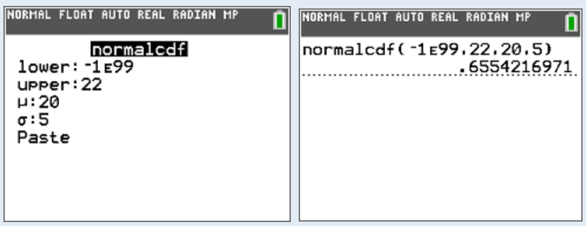
Access normalcdf function: 2nd → VARS (DISTR) → NORMALCDF

For a normal distribution:
normalcdf(left boundary, right boundary, mean, standard deviation)

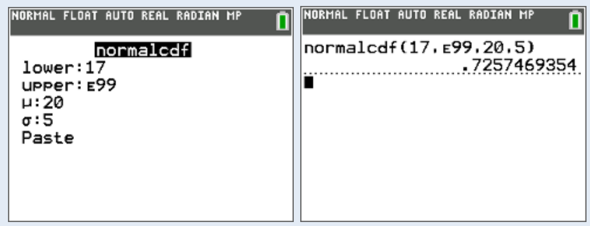
For a standard normal distribution:
normalcdf(left boundary z-score, right boundary z-score)

Note: *Use 1E99 for positive infinity and -1E99 for negative infinity*

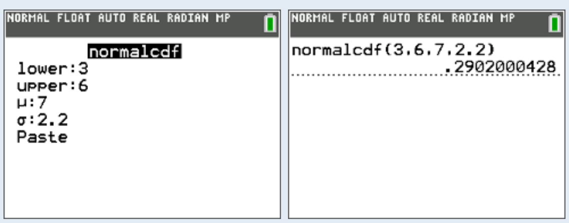
Example 3: $X \sim N(20, 5^2)$. Find percent less than 22.



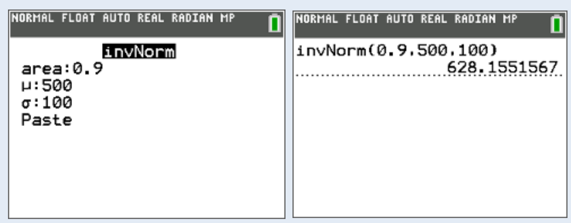
Example 4: $X \sim N(20, 5^2)$. Find percent greater than 17.



Example 5: $X \sim N(7, 2.2^2)$. Find percent between 3 and 6.



Example 6: $X \sim N(500, 100^2)$. Score in 90th percentile.

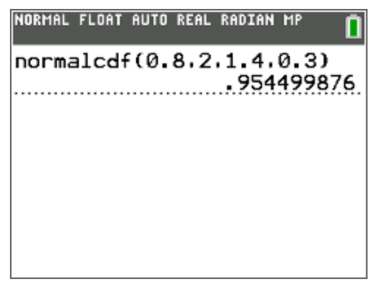


Section 3.5b – Applying the Normal Distribution
MDM4U
Jensen

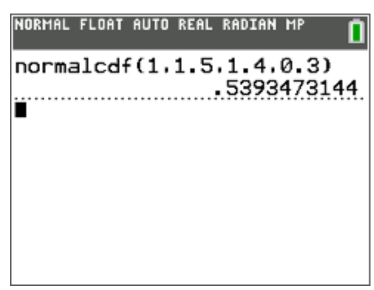
Height of cacti are normally distributed with a mean of 1.4 m and a standard deviation of 0.3 m.

1. 68% of the cacti are between...

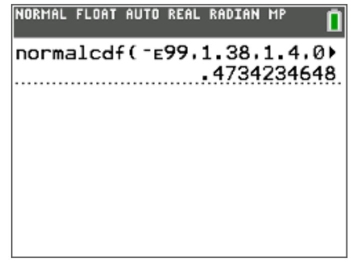
2. What percent of cacti are between 0.8 and 2m tall?



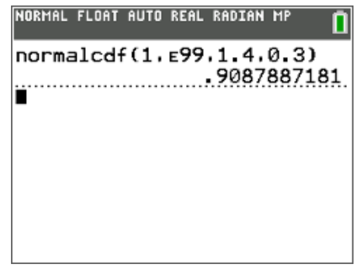
3. What percent of cacti are between 1.0 and 1.5 meters tall?



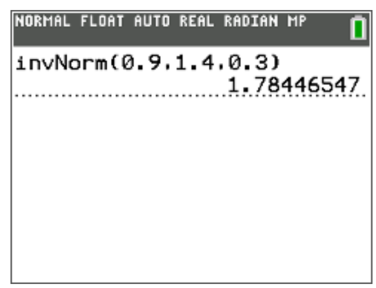
4. What percent of cacti are less than 1.38 meters tall?



5. What percent of cacti are likely more than 1.0 meters tall?

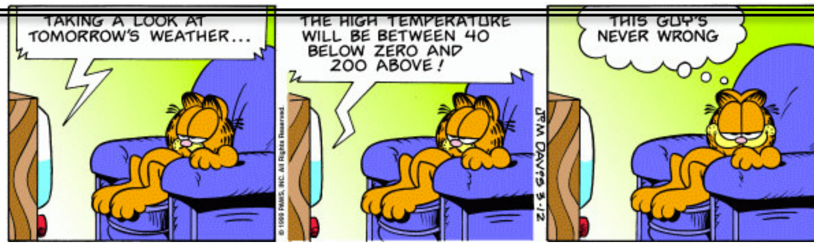


6. Find the 90th percentile of cactus height...



Section 3.6 – Confidence Intervals

MDM4U
Jensen



Part 1: Intro to Confidence Intervals

A confidence interval uses _______________ data to estimate an unknown ____________________ parameter with an indication of how precise the estimate is and of how confident we are that the result is correct.

When we take a sample, we do our best to try and obtain values that accurately represent the true values for the population.

For example, if we took a simple random sample of 500 from a population of a town with 10,000 people and found that in the upcoming election, 285 plan to got for Candidate Y, then our best guess for the proportion of the town that will vote for Candidate Y is 285/500 or 57%.

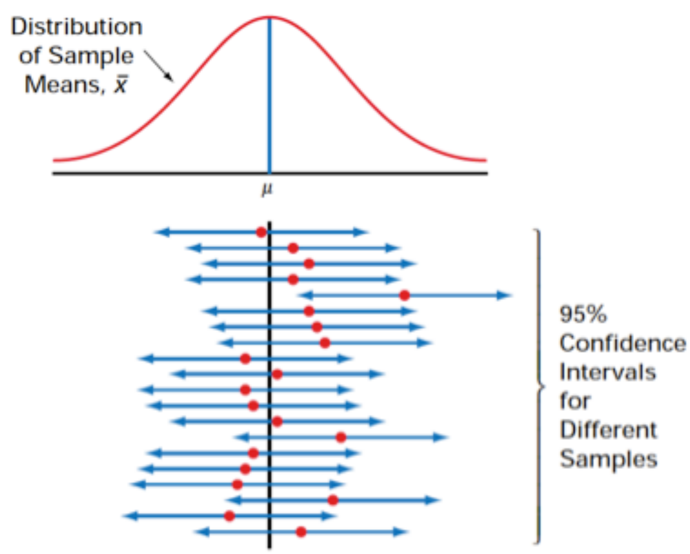
But we don't know if the TRUE proportion of people who will vote for Candidate Y is 57%. For this reason, we report a ________________________. From this information, we can create a confidence interval.

The confidence interval will tell us with a certain percentage of confidence (usually 95% is used) that the population mean will fall between some lower limit and some upper limit.

To interpret a confidence interval we say:

A computed confidence interval shows the proportion of confidence intervals of size *n* that actually contain the population mean.

In this example, the 95% confidence intervals for various samples are shown. Only 1 of 20 does not contain the population mean, meaning that there is a 95% chance the population mean is within the confidence interval given by different sample means.



Part 2: Confidence Intervals for Population Means

Formula:

$$C.I. = sample\ mean \pm margin\ of\ error$$

$$C.I. = \bar{x} \pm z^* \left(\frac{\sigma}{\sqrt{n}}\right)$$

$\bar{x} =$

$\sigma =$

$n =$

$z^* =$

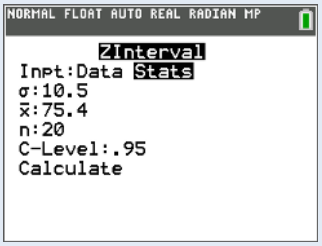
This table gives a list of common confidence levels and their associated critical values:

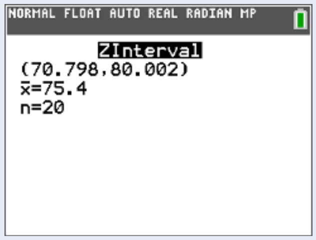
Confidence Level	Tail Size	z^*
90%	0.05	1.645
95%	0.025	1.960
99%	0.005	2.576

Example 1: A paint manufacturer knows from experience that drying times for latex paints have a standard deviations of 10.5 minutes. The manufacturer wants to use the slogan: "Dries in *T* minutes" on its advertising. Twenty test areas of equal size are painted and the mean drying time is found to be 75.4 minutes. Find a 95% confidence interval for the actual mean drying time of the paint.

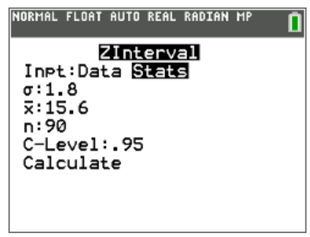
Using the Ti-84 for example 1:

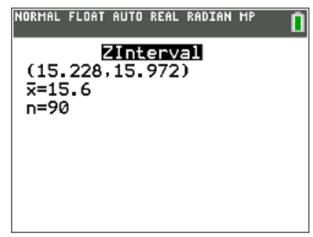
STAT → TESTS → ZINTERVAL





Example 2: Julia enjoys jogging. She has been jogging over a period of several years, during which time her physical condition has remained constantly good. Usually, she jogs 2 miles per day. The standard deviation of her times is 1.8 minutes. During the past year, Julia has recorded her times to run 2 miles. She has a random sample of 90 of these times. For these 90 times, the mean was 15.6 minutes. Find a 95% confidence interval for the mean jogging time for the entire distribution of Julia's 2 -mile running times.





<u>Part 2: Confidence Intervals for Population Proportions</u>

Formula:

$$C.I. = sample\ mean \pm margin\ of\ error$$

$$C.I. = \hat{p} \pm z^* \left(\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

$\hat{p}$ =

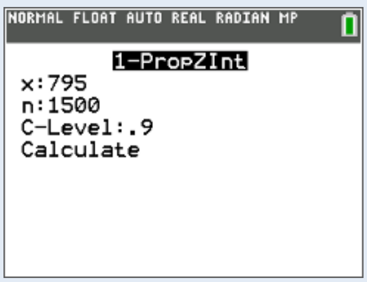
$n =$

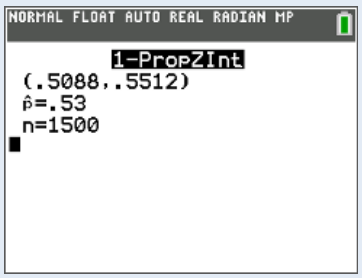
$z^* =$

Example 3: Voter turnout in municipal elections is often very low. In a recent election, the mayor got 53% of the voters, but only about 1500 voters turned out. Construct a 90% confidence interval for the proportion of people who support the mayor.

Using the Ti-84 for example 3:

STAT → TESTS → 1-PropZInt





Note: x was calculated by doing $\hat{p} \times n$

Example 4: A random sample of 188 books purchased at a local bookstore showed that 66 of the books were murder mysteries. Find a 90% confidence interval for the proportion of books sold by this store that are murder mysteries.