GRADE 12 ESSENTIAL
UNIT C – STATISTICS

CLASS NOTES

"*There are three kinds of lies:*
lies, damned lies and statistics"; Mark Twain.

INTRODUCTION

1. If you are adept at Statistics it is possible to credibly and convincingly fool many people about anything most of the time!

2. Statistics is one of the many branches of Mathematics. So far you have worked with **Arithmetic** and **Measurement** and **Geometry** and **Algebra** and **Trigonometry**; special branches of mathematics. The branch of **Probability** is very closely related to **Statistics** as well.

Definition. One simple definition of statistics is the '*collection, processing, and display of information*'.

3. Objectives. The objective of this unit is to learn about statistical measures of **Central Tendency** and how data is distributed.

Prior Knowledge. You would have done much of these same concepts in previous grades so much of this should be a refresher. In Grade 11 Essential you studied how to graph Statistical data and in Grade 9 the ways to collect information.

CENTRAL TENDENCY

4. Many students will be familiar with the ideas of central tendency from previous Grades. This part constitutes a comprehensive review.

5. We often want a single number to represent many numbers in a set. For example if the students if the students in the class room have ages: 23, 25, 26, 23,22, 23, 20, 34, 40 we might ask : 'what **one** age is sort of the average or typical, or the normal, or the middle, or the central age that I can think about when I think about this class?"

6. A single number to represent a bunch of numbers in a set is called a **Central Tendency**. And there are several different ways to find a Central Tendency. And as you might expect the different ways do not necessarily give the same result.

MEAN

7. The **mean** (also called *average*) of data is a statistic that is a measure of the **central value** of the data. It is represented by the symbol: $\bar{x}$, 'x bar' or sometimes the Greek letter: **μ**; 'mu'. A single number that somehow represents a bunch of numbers is called a '***statistic***'. There are many 'statistics' in common use. The mean is a very commonly used statistic, sometimes over-used.

8. There are several types of **mean**, however the one most common and that we will use exclusively (the '*arithmetic mean*') is the mean that is calculated by adding up all the data values and then dividing by the number of values.

9. For example; given the set of data values: **{1, 2, 4, 6, 7, 9}** the mean would be:

$$mean = \bar{\boldsymbol{x}} = \frac{1+2+4+6+7+9}{6} = 4.83$$

Make sure you understand the order of operations! The entire top is summed up then the sum is divided by the size of the set of numbers.

10. You may want to remind yourself, and your calculator, that the top is calculated first by using **parentheses** ( ) around the entire numerator:

$mean = \bar{x} = \frac{(1+2+4+6+7+9)}{6} = 4.83$ (especially if you plan on entering the entire expression into a calculator in one shot)

11. Calculate the mean, $\bar{x}$, for the following data sets to two decimal places:

a. **{1, 2, 2, 3, 4, 6, 9}** Mean, $\bar{x}$ = **Ans: 3.86**	b. {**1.2, 3.5, 3.3, 3.3, 3.4, 0.9, 6.1}** Mean, $\bar{x}$ = **Ans: 3.10**
c. {**70, 80, 90, 100, 110, 120, 130}** Mean, $\bar{x}$ = **Ans: 100.00**	d. {**1, 2, 4, 8, 16, 7, 3, 2, 1}** Mean, $\bar{x}$ = **Ans: 4.89**
e. {**⁻2°, ⁻4°, +7°, ⁻3°, +8°, +8°, 0°}** Mean, $\bar{x}$ = Daily Temperatures in April	f. **{4.2 cm, 5.6 cm, 3.4 cm, 3.1 cm}** Mean, $\bar{x}$ = Thickness of some steaks

12. **General Formula for Mean**. The general formula to calculate a mean is to **sum** up all the numbers in the set of numbers and then divide that **sum** by the size (the number of numbers) of the set of numbers.

The fancy way that statisticians say that rather simple idea is to abbreviate the idea as:

$$mean = \bar{x} = \frac{\Sigma x}{n};$$

where n is the size (how many) numbers in the set of data (the sample size) and **Σx** means the **sum** of all the x's.

* The symbol Σ is the capital Greek letter **SIGMA**; which is their letter 'S'**. 'S'** for **Sum.** There now you know!*
There are different ways to monkey around with the **mean** as we will discover later.

MEAN UNITS

13. Means are made up of numbers that represent some sort of unit. So naturally a mean value of some data should have units attached to it.

Eg: Jeff runs the following daily amount of **km**'s: {5, 6, 7, 8, 9} The mean is 7 **km**. So, don't forget to include units.

MEAN AS AN EQUAL SHARING

14. One way to think about a mean is what everyone would get if everything was equally shared. If Rob has \$4 and Janice has \$6 and Brandon has \$8 and they threw all their money in a pot ($\sum all\,\$$) and added it up it would be \$18, taking away \$3 at a time to divide up the money; each person would get \$6.

$$\frac{\sum all\ \$}{3} = 6$$

MEDIAN, $\tilde{x}$

15. The Median is *another* measure of **central tendency**. It is the **middle** datum in a set of data; like a police line-up, the central datum, …the 'median' on a highway. It is calculated as follows:

a. Arrange the data in numerical order by value;

b. Find the middle data value; and

c. *If* there are two middle data values, use their mean.

Eg: {**9, 2, 4, 6, 6, 7, 1**}. Putting in order: {**1, 2, 4, 6, 6, 7, 9**}. The median, $\tilde{x}$, is 6.

Eg: {**1, 2, 4, 6, 7, 9**}. The median, $\tilde{x}$ is $\frac{4+6}{2}$ = 5. Compare this to the mean. Notice that **means and medians are not necessarily** the same value; sometimes they are close, sometimes very different.

16. Try these; find the median and compare with the mean previously calculated:

a. **{4, 8, 1, 1, 3, 5, 7}** Median, $\tilde{x}$ = Mean, $\bar{x}$= **Ans: 4 and 4.14**	b. **{1.2, 3.5, 3.3, 3.3, 3.4, 0.9, 6.1}** Median, $\tilde{x}$ = Mean, $\bar{x}$= **Ans: 3.3 and 3.1**
c. **{70, 80, 90, 100, 110, 120, 130}** $\tilde{x}$ = $\bar{x}$ = **Ans: 100 and 100**	d. **{1, 2, 4, 8, 16, 7, 3, 2, 1}** $\tilde{x}$ = $\bar{x}$ = **Ans: 3 and 4.88**

DIFFERENCE BETWEEN MEAN AND MEDIAN

17. You likely noticed that the mean and the median do not necessarily give the same '*statistic*' of central tendency! If you are sneaky you can use that to your advantage.

20. **Example - The Pay Difference**. John is on a crew of seven workers who shingle roofs for a company. The company is owned by a family; two brothers (Tom and Jerry) and an aunt Brenda. John and his crew claim they do not get paid enough! John does some research and gets the annual income of the ten people in the company.

21. Here is the data that John found: John \$24K, Kyle \$27K, Jeff \$26K, Bev \$21K, Terry \$24K, Mike \$24K, Kevin \$27K, and owners: Tom \$120K, Jerry \$110K, Brenda \$93K.

Calculate the **mean income** of the company ______________________

Calculate the **median income** of the company ________________

What is the most common pay? (ie: **MODE**) __________

Advanced concept*. Notice that when John researched company pay he was just rounding to the nearest 'K**' or thousand. How do you think using rounded numbers might change the 'statistics'?*

22. Explain the big difference in the central tendency statistics:

Do you think that some companies (or politicians) might '*cherry pick*' and use the statistic that makes them look better? (**Y / N**)

MODE

23. The **mode** is yet *another* measure of a 'central' tendency. It is the number value or category that happens the most often; the most popular.

24. Eg: {**2, 2, 4, 5, 6, 6, 7, 7, 7, 8, 9, 3}**. The data value '**7**' occurs the most often, so the mode is **7.** It helps to put the data in numerical order to better notice the most frequent value. Sometimes there may be two values that are equally common, a **bi-modal** distribution of data. If there are two or more most frequent values, list them all as the mode. The mode of {3, 4, 3, 1, 2, 4} is 3 **and** 4.

The mode is not an overly useful statistic for the central tendency. Although it may be useful in some context as below.

25. Carla sells bannock at her bannock shop. Here is a list of how many bannock each of her regular 12 customers ordered today:

{1, 2, 5, 5, 5, 2, 5, 3, 5, 1, 1, 2}.

Carla is thinking that 5 bannock seems to be a popular size quantity to order and maybe she should pre-package some of her bannock into packages of 5. The mode of her order quantities is that order size that occurs the most often. Determine the mean, median, mode, and range of her normal daily orders?

Mean: _____; **Median**: _______; **Mode**: _______; **Range:**____

RANGE (SPREAD OF DATA)

26. The range of your data is another statistic. It tells how **spread out** your data are, a rather simple calculation. To calculate the range of your data find the difference between the highest value in the set of data and lowest value in the set of data.

Range = $x_{max} - x_{min}$

where x_{max} is the highest (maximum) value of the data and x_{min} is the lowest (minimum) value of the data. Range will have units of that which is being measured.

30. Determine the Mean, Median, and Range of the following student course averages in a Math class:
{60%; 45%, 88%, 23%, 76%}

(explain if the teacher like these marks)

SPREAD OF DATA

31. The range of course is not a measure of the Central Tendency; it is a statistic that tells you how ***spread out*** your data is. It is known as a measure of ***Variability***, or sometimes called **Dispersion**, which is not covered on this course. If you see resources that talk about Variance or Deviation or Inter-quartile Range of data then they will be referencing different ways of measuring how ***spread out*** the data is.

32. The importance of the **Range** to Grade 12 studies is that the **Mean** value, the **Median** value and the **Mode** of the data must fall within the x_{max} and x_{min} values of data. In other words; a Mean (average) Math mark cannot be more or less than the highest or lowest of any mark. You cannot have all your marks in the range from 35 to 72 and have a Central Tendency that says you have an 84!

FREQUENCY TABLES

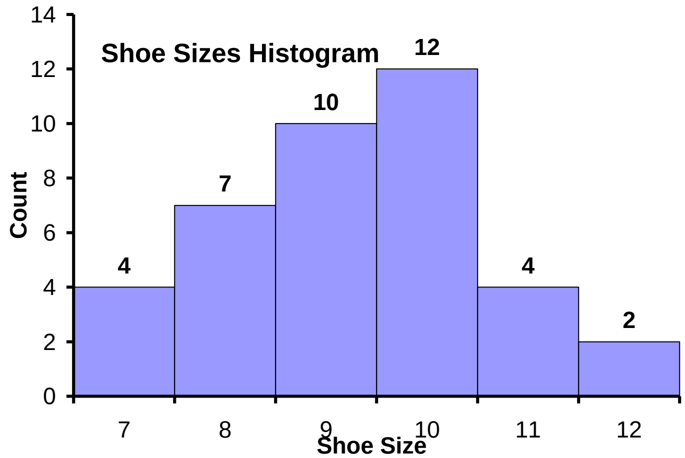
35. Frequency tables just record how frequently different data values occur. You would normally make a Frequency Table when conducting a survey. It is usually easiest to make a stroke tally as you count values and then just count the tally for each value.
Eg: a frequency table from a survey of shoe sizes at our school:

Shoe Size	Tally	Count
7	////	4
8	~~////~~ //	7
9	~~////~~ ~~////~~	10
10	~~////~~ ~~////~~ //	12
11	////	4
12	//	2
	Total sample size, n:	39

HISTOGRAMS

36. You will recall from Grade 11 that Histograms are just a picture of the data. The human eye usually sees patterns better in such a graph or picture. A picture is worth a thousand numbers.

A histogram of the above shoe size frequency table would look like this.



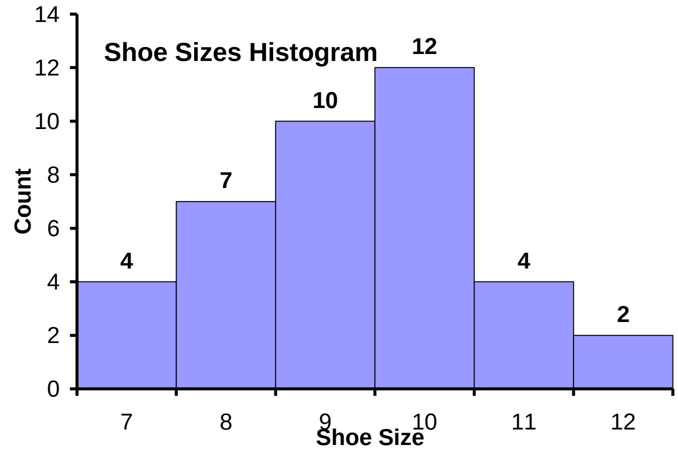
37. Notice that in a histogram that the area of each bar represents a percentage of the entire sample. Each category is the same width (one shoe size in this case). In the above case there were **39 samples**. So the category for shoe size **9** takes up **10 out of 39** parts of the graph, or **25.64%** of the data. In a histogram, there should not normally be any gaps between the bars in the categories (unlike a '**bar graph**').

38. A histogram can also show percentages or probabilities (instead of a raw count) of categories along the vertical axis, simply by taking the count and dividing by the total number of the sample. We could have said, for example, that 4/39 of shoes are size 7, so 10.26%, and graphed the percentage along the vertical axis instead.

In the above survey on shoe size the histogram tells us that there is a 46% probability that next random person to walk in the door will have a size 10 or bigger shoe. Histograms are useful to study probability (another unit in Grade 12)

FINDING CENTRAL TENDENCIES FROM A HISTOGRAM

39. Given the shoe size histogram it is not difficult to find the important statistics of **mean**, **median**, and **mode** (almost just by looking) directly from the histogram!



40. **Find the mean from the histogram**. Add up all the data values, divide by the total number of values. There are 39 data values, their total is (4*7) + (7*8) + (10*9) + (12*10) + (4*11) + (2*12) = 362.

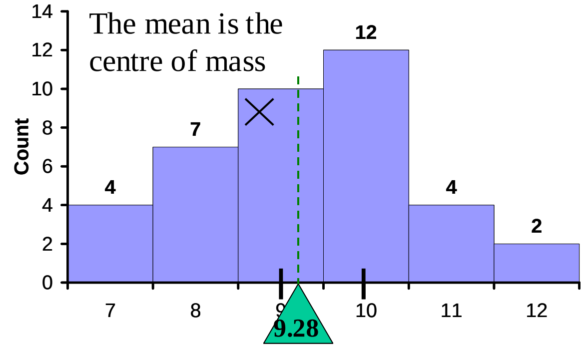
41. So the mean, when we count the frequency of data values, is given by:

$$\bar{x} = \frac{\sum(f_i x_i)}{N} = \frac{(4*7)+(7*8)+(10*9)+(12*10)+(4*11)+(2*12)}{39}$$ which is the mathematical way of saying "*take the frequency count, f_i, of each data value category, x_i, multiply them together, add them all up, and divide by the total number*".

The Greek letter Σ (capital 'sigma'), is used almost exclusively to mean ' add up all this crap'.
(But I don't think crap is a Greek word.)

42. So we have **362 ÷ 39 = 9.28**. The **mean** shoe size is **9.28**.
If you went to the shoe store and asked for a size **9.28**, then you would have a 'mean' or 'average' foot size. Obviously, stores don't actually sell that size, you have a choice between discrete shoe size values only: either size **9** or **10**.

43. One rather important interpretation of the mean is that it is the '*centre of balance*' of the histogram, the histogram would balance at the mean if we stuck a knife edge there along the bottom.



You may know a bit about 'Centre of Mass' from science.

45. Find the median from the Histogram. The median is the exact half-way piece of data when you line the data up in order of value. There are **39** data values. If we knock off **19** values from the *left*, and **19** values from the *right*, the ***20th*** value from either end would be the central one. The first 4 values from the left were sevens, the next 7 were eights, so the 20th would occur in the size nine category. The **median** is 9. I put an X in the median category where the 20th person was. A median will generally be one of the categories, unless it is exactly half way between two categories with an even number of samples.

Find the Mode from the Histogram

46. Find the mode from the histogram. This is the easiest statistic of the histogram! It is readily obvious that the most frequent shoe size is size **10**. So the **mode** is **10**. Occasionally you might notice two peaks (or more) that are the same frequency. It is possible to be *multi-modal* and have more than one most common value.

Notice this histogram is not a nice perfect (normal) symmetrical shape. We say it is '**skewed**'. The data favours the lower data by a bit. There is a statistic to measure 'skewedness' too which we do not cover.

47. *The Three Different Measures of Central Tendency Generally Give Different Answers*. The wonder of statistics! So, as you can see…, the **three different measures** of central tendency give similar but slightly different answers. For our shoe sizes the **mean,** $\bar{x}$, was 9.28, the **median,** $\tilde{x}$, was 9 and the **mode** was 10. You could say the average was 9.28 (if we all had our feet surgically adjusted so we were all the same shoe size), or a median: where half the people have a size 9 or more and half have a size 9 or less. The mode, the most common size, is 10.

If you can master statistics, you would be a great politician; you can give three different answers for the same question!!! Pick the one that best makes you look good!

A good example of a misleading measure of central tendency.

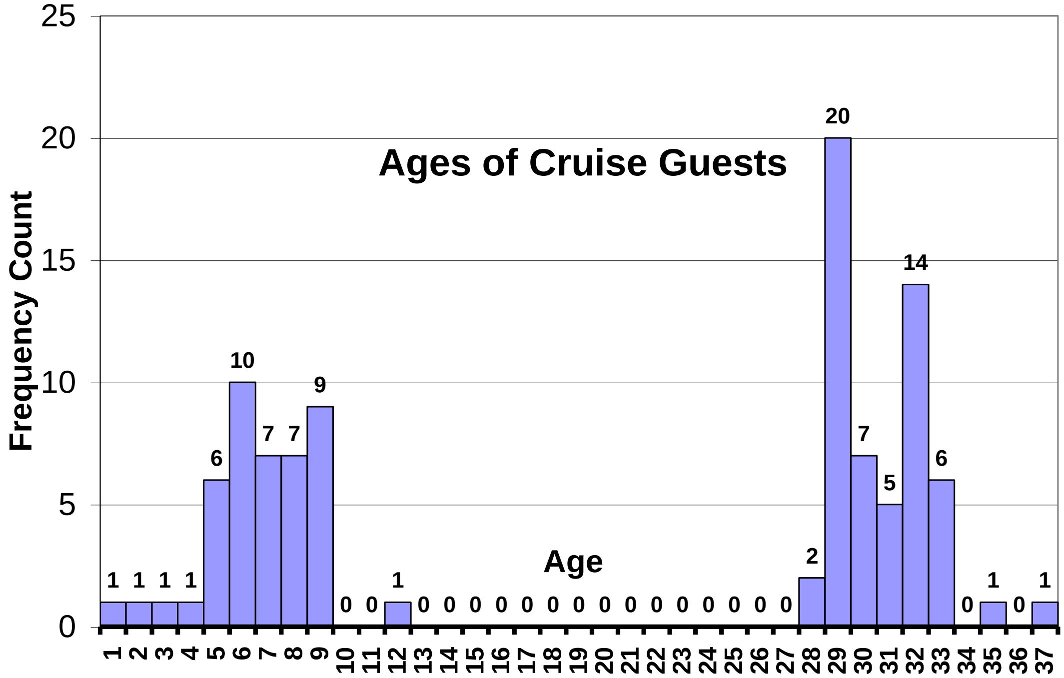
48. We get the sense that measures of central tendency are a good representation of the measure of the 'centre' of data. But be careful!! Statistics can be misleading, especially in certain hands! If you really understand statistics, you can pretty well rule the world! (ie: **BS** baffles brains!)

50. Let's say you decide to go on a holiday after grad! You really deserve a cruise in the Caribbean after all that Math! So, you book a holiday with Funtastic Cruise Lines that ***guarantee*** the average (ie: ***mean***) age is about ***20*** years old!

Check out the graph below and calculate the mean, median, mode, and Range!

Aren't you surprised then when you get on board and find out that there are lots of married **30** year olds, many with their young toddlers!

Oh well, at least you can make some babysitting money while on-board!



51. The 'bimodal' distribution in the histogram above shows the hazards of believing a single statistic. So *be very careful* which statistics you use. You might '*accidentally*' mislead people! Often **median** is a more relevant statistic. Further, any suspicious statistic that isn't supported by a histogram could (should) be readily challenged! In this particular case with the cruise ship I would certainly want to see the histogram of all the data, not just one number that summarizes everything!

52. The important rule to remember is that **using a single statistical number to represent a lot of numbers can be very dangerous** unless you are familiar with the 'distribution' of the data too! Data that is '*skewed*' generally has a large difference between the Mean, Median, and Mode. Data that is not 'skewed' is called 'normal'. The best way to see if the data is distributed in a well-behaved manner is to ask to see the graph of the data.

Be very careful with statistical numbers!! **Ask to see a graph also**.

Finding the middle number of an ordered set of numbers (ordinals).

To find the middle position of any set of objects (including an ordered list of numbers); add one to the size of the set, n, and divide the total by 2. If the set size 'n' is an even number then the middle position will be between two objects (numbers).

Eg: the middle position between 7 objects is (7 + 1)/2 so the 4^{th} position or 4^{th} number is the centre object or number.

The middle position between 12 objects is: (12 + 1)/2 = 6.5; so the position in between the 6^{th} position and the 7^{th} position.

Can you think why this is true?

TRIMMED MEANS

53. Trimmed means are means that have been trimmed! Some of the numbers on the end have been trimmed off. The main reason to do this is to eliminate rogue '**outlier**' values.

54. Take for example the ages of kids in a day care that an assistant collected for you!
{2, 4, 3, 5, 4, 4, 3, 23} years of age. It is difficult to believe you have a 23 your-old kid in daycare. Perhaps your assistant wrote down the age wrong! Or maybe she did not understand the survey question. But you certainly do not want to include that number. So, you can '**trim**' it off. Normally however, we trim from both ends if we trim data 'to be fair'. Regardless, whoever gives you the statistic should be perfectly clear what it was they may have trimmed.

55. Example: do a 10% trim of this set of numbers.
{1, 2, 4, 4, 6, 7, 8, 9,10, 15}. The size of this set, ***n***, is 10. 10% of ten is one. So we trim off one number from each end. The trimmed mean becomes $\frac{\mathbf{2+4+4+6+7+8+9+10}}{\mathbf{8}} = \mathbf{6.25}$

56. You have likely seen the Trimmed Mean often and never realized it, particularly in judging figure skating. Because of a scandal many years ago where a skating judge was bribed to give elevated marks, the Olympic Committee decided that when six judges give their score the low one and the high one will be thrown out. If there was no bribery or cheating then throwing away the top score and the bottom score should not change the mean score by much since neither end score would have been excessively away from the normal mean anyway.

WEIGHTED MEANS

58. Teacher Rick has a class of 20 students with a class average of 74%. Teacher Courtney has an English class of 40 students with a class average of 86%. What is the average mark for both classes combined?

59. We cannot just add the two marks together and divide by two since Courtney had twice as many students, her average should actually count twice as much. Had we got all the students into one room and made an average of their individual marks their individual averages we get a different answer. The proper average mark combining both classes is:

$\frac{\mathbf{74*1+86*2}}{\mathbf{3}}$ =82%; closer to the English class average since there were more students in the English class. The important point here is that you **cannot** generally ***average averages*** with considering the size of the sets of numbers involved.

60. Teachers often use weighted and trimmed means. Some teachers will knock off the lowest mark on a set of assignments based on the principal every one has a bad day. And most teachers do weighted mean in which tests count twice as much as quizzes and quizzes count twice as much as assignments for example.

61. **Example**: Teacher Mike has three quizzes and one test the first part of the course. Tests count twice as much as a quiz. Bev's quiz scores were 45, 68, and 77 and her test score was 85. What is Bev's mark?

$$\frac{\mathbf{45(1)+68(1)+77(1)+85(2)}}{\mathbf{1+1+1+2}} = 72.$$

Compare this to Bev's mark if the test **did not** have a heavier weight factor than the quiz:

$$\frac{45+68+77+85}{4} =$$

Example:

62. Last year the Jets won 35% of their games in the 80 game season. This year after 40 games they have won *25*% of their games. Combine those two statistics into one average:

PERCENTILES

Finding the percentile rank of a particular score or value

65. Often, we are interested in where a value sits in a range of values.

You might have asked yourself a question like this before: "Is my mark in the bottom 10% of the marks in class? Is my mark in the top quarter of the class marks? Is my baby too chubby?"

Finding where a value sits in an ordered set of data is called a finding a Percentile Rank.

It is like where you fit in in a hundred dog sled team! You prefer to be a highly rated dog up at the front! Dog number one is at the back of the pack! You want to be the 100^{th} dog forward! You want to be a top dog, the high ranking boss dog!

66. To calculate a percentile rank follow this formula:

$\boldsymbol{Percentile\ Rank = \frac{B+0.5E}{n} * 100}$; where **B** is the number of data in the ordered set that are **Below** the given score or value and **E** is the number of scores **E**qual to that score value (in the event there is more than one)

Finally, **percentile rank** is normally rounded up to the next whole number

***Be aware that there is no one single agreement on how to calculate a percentile ranking, so you may see variant explanations in other sources. The concept remains the same, you are to find how high up the list in the numbers you are*.**

67. **Example 1**. Let's say you have these values for something: **{20, 20, 45, 50, 60, 60, 70, 75, 80, 95}**, They are already put in ascending (increasing) order. There are 10 of them, n =10. Two of the marks were lower than 45 and there was only one 45

$Percentile\ Rank = \frac{2+0.5(1)}{10} * 100$= 25.

The score of 45 has a percentile rank of 25. Had there been 100 numbers, the 45 would have been in the 25th place.

68. **Example 2**: Say your class of 52 members wrote a test. There were 37 students that had a mark lower than yours and no one else had your mark. Your percentile rank would be:

$\boldsymbol{Percentile\ Rank = \frac{37+0.5(1)}{52} * 100}$ =72.11... Your Percentile Rank is 73. Percentile rankings are always whole numbers, so if there is a fractional (decimal) part **round up** to the nearest whole value. Notice we **do not** call it a 73**%** percentile ranking. It is the position of your place in a hundred dog sled team! How far up the ladder you are.

Be careful to distinguish between your mark on the test, perhaps it was 93%, and the 73 percentile ranking which indicates that 73% of the students had a lower mark than you. **Marks and rankings are not the same thing!**

69. **Example 3**: John wrote a test, the entire class was thirty students. John got a mark of 48 out of 60. So *his* test mark as a percentage on *his* test was $\frac{48}{60} = \frac{x}{100}$; so his mark on the test was 80%. But how did he do compared to the remainder of the class? If only 6 students had a mark Below his (B = 6) and he and two others had a 48 mark **E**qual to his (**E** = 3), then what was his percentile rank?

$\boldsymbol{Percentile\ Rank} = \frac{6+0.5(3)}{30} * \mathbf{100} = 25.$ The percentile rank is a 25. **Not 25%.** So John had a percentile **rank** of 25, meaning 25% of students had a mark at or below his, which of course means that 75% had a mark that was better.

70. Consequently one would think that this test must have been rather easy if only 25% of the students had less than an 80% test mark and 75% of students had better than an 80% test mark.

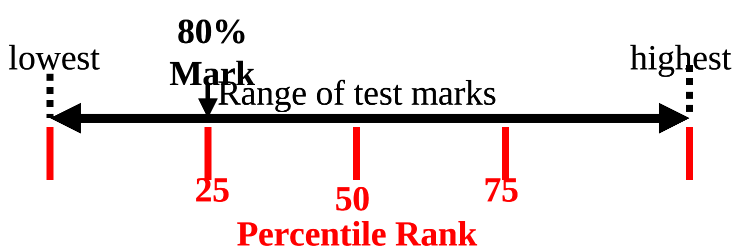


Diagram of John's and the classes inflated (?) test mark(s)

71. In some schools and universities, the staff of the institution would review the marks and determine that obviously the test must have been too easy and they would actually reduce the marks sent in by the teacher! This has been in the news a lot, there are lots of instructional places that are considered '**credit mills'**; they graduate everyone with a good mark.

72. In the same context, some Universities know full well from their own statistics that they keep which schools submit inflated marks and so the Universities make their own adjustments accordingly.

73. It is called '*adjusting marks to the curve*', or '*bell curving*' when marks are adjusted so that they meet the expected statistically '*normal*' mark.

Can you see a possible problem with this '*marking to the curve?*'

You Try:

75. Bev got a mark of 37 out of 50 on a test. Her class was 20 students. 18 students had a mark below hers and she had the only 37. What was her percentage mark on her test? What was her percentile rank amongst the class for that test? What can you say about the difficulty of the test?

QUARTILES

76. Quartiles involve breaking an ordered set of data into four equal size portions and seeing where any particular value of interest falls in which quarter of the data.

The 25th percentile, P_{25}, is the cut off for the lower quartile (Q_1, the first quartile, in some computer programs), it is the lower **median of the median**!

the 50th percentile, P_{50}, is the cut off for the lower half of the ordered data (Q_2, the second quartile, in some programs),

the 75th percentile, P_{75}, is the cut off for the lower three quarters (or upper one quarter) of the ordered data; it is the **upper median of the top half of data** (Q_3, in some programs).

Notice how P_{50}, also known as the 2nd quartile, Q_2, is really just the Median because half of the ordered data values are below and half are above the Median.

77. **Example**: A report by a certain interest group does a survey of 1,000 families and reports that the income information for a local community as follows:

P_{25} or $\mathbf{Q_1}$ = \$30,000; P_{50} or $\mathbf{Q_2}$= \$50,000; P_{75} or $\mathbf{Q_3}$ = \$75,000 ; and $\mathbf{P_{90}}$ = \$100,000. Draw it out on a number line!

a. what percentage of the families earn more than \$75,000?

b. what percentage of the families earn between \$75,000 and \$100,000.

c. what is the median income of the community?

d. from the information given what is the **range** of the incomes in the community?

e. How many families make more than $75K? How many families make less than $50K?

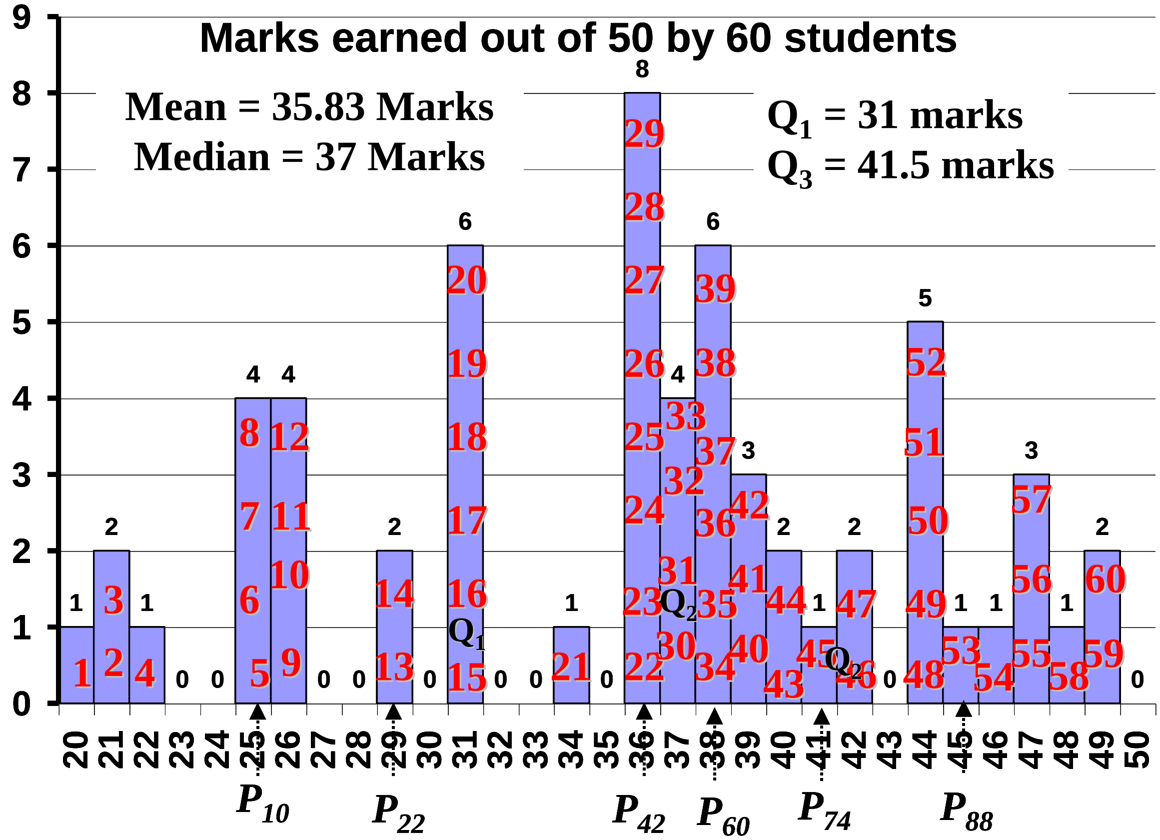
f. although you cannot find the full range of the variable income values, you can find the '*inter-quartile range*'! The range of incomes between Q_1 and Q_3, or P_{25} and P_{75}. Find the difference to calculate the Inter-quartile range. The interquartile range could be called 'the middle class'! The inter-quartile range is:

g. How many of these families are middle class?

____

78. You may be starting to notice that statistics are not overly descriptive and reliable for small sample sizes (?) and that larger sample sizes give more useful statistics. You might also notice that there is a variety of ways to manipulate data, something that may be good or bad depending on your intent!

80. Perhaps this comprehensive picture will be better. This is a histogram showing the test scores for 60 students. I think this explains lots of statistics so much more easily!



Deciles

81. Be aware there are also ‘decile’ rankings where instead of being broken into 100 whole numbered chunks the ordered data is broken into 10 chunks. If the doctor told you your child’s weight was in the lower decile of the weight scale he meant the child had a P_{10} ranking and below, your child was well below normal weight, below 90% of all babies in weight.

SOME PRACTICE PROBLEMS

SAMPLE QUESTIONS

1. Calculate the 10% trimmed mean (10% from each ends) for the values:
{45, 96, 8, 54, 48, 30, 49, 52, 38, 44}

2. Kevin had four tests and a final exam. All the tests had the same 'weight'. The final exam was worth twice as much as a single test.
If his marks were as follows:
Test 1: 67%, **Test 2**: 87%; **Test 3**: 90%; **Test 4**: 35%; and his final exam was a 92%; what was his final course average?

3. Scott presently has a course average of **42%**. If the final exam is worth 25% of the course average, what exam score (%) does Scott need to get an overall course average of **at least** (ie: greater than or equal to; $\geq$) 50%.

4. Colleen has been jogging every day! She wants to jog *at least* an average of 4 km per day. The last six days she has jogged 22.2 km; how far will she need to jog today to get an average of *at least* 4 km / day for the week?

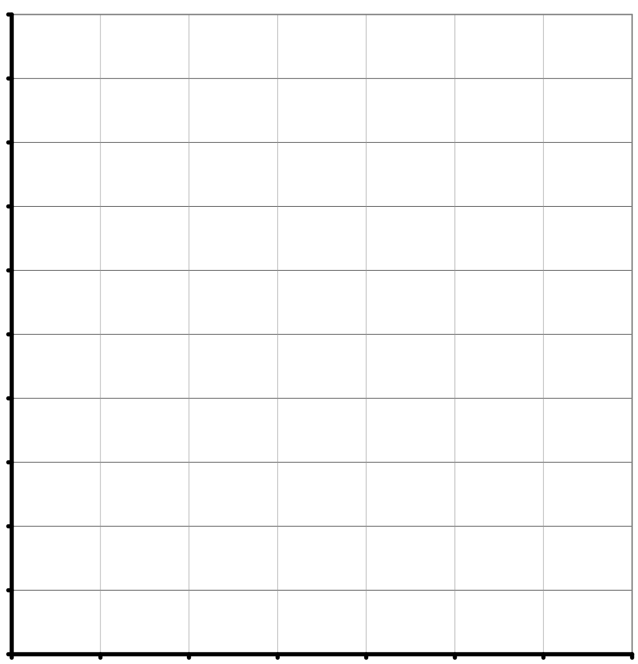
5. A Quality Assurance supervisor from the Canadian Nuts and Bolts company measured the masses of 300 bolts to see how consistently they were being manufactured. Here are the results of his quality testing sample.

Mass [g]	7.40	7.50	7.60	7.70	7.80	7.90	8.00
Frequency	3	5	38	86	91	60	17

a. Determine the mean mass of the bolts. (You may want to do your own frequency data table)

b. Calculate the median mass.

c. do a properly labelled histogram and clearly mark the mean, median, mode, the **P_{50} (Q_2)**, the **Q_3,** and the approximate **P_{80}.**



d. What is the range of the masses of this sample?

e. what is the range of the inner quartile sizes, from Q_1 (P_{25}) to Q_3 (P_{75}). Does this suggest that the masses tend to be clumped more in the centre.

6. Candice is 1.6 metres (160 cm) tall. She is taller than 55 of the students in her grade and no one is exactly the same height category as she is. There are 121 students in her grade. What is Candice's percentile rank? What percent of the students are taller than Candice?

7. Baby Reeve visits his doctor for his 3-month check up. He weighs 14 pounds and 8 ounces. When his weight is compared with 800 other 3-month old babies, Reeve weighs more than 321 of the babies and the same as 16 of the babies. What is his percentile rank for the weight of 3-month old babies?

8. Problem Solving. Mr. F normally has an 80% success rate with his curling shots. Today he only had three out of his first six shots that were good. If he has ten more shots to go in the game (there are only 16 shots total for each player in a game), how many good shots must he make in his last 10 shots to at least maintain his 80% success average.

A frequency data table to record and calculate large samples

	Frequency Data Table (to calculate statistics of large samples)				
x Value of variable being measured	Tally ticks (if doing a survey)	**f** frequency each value happens [count]	Cumulative Frequency (running total)	**f*x** *f times x*	
					Mode; most frequent x: _______
					Mean, μ **or** $\bar{x}$ **=** $\frac{\Sigma(f*x)}{n} =$
					Median, $\tilde{x}$ Halfway into the data; in between two values if n is EVEN. = ________
		sum: **n =** ______		sum: $\Sigma all\ the\ f*x's$ ________	

*A quick way to find the middle place of a string of numbers is to take **(n + 1)/2**. That will tell you where the middle place would fall. If the result is a half value then you then you are in between the two places. So in a string of 83 numbers the middle number would be in the 42nd place. In a string of an even number of numbers however, say 180, the middle place would be in the 181÷ 2 place or the '*ninety and a halfth*' place; so you would need find the mean of the two numbers either side; so the mean of the two numbers in the 90th and the 91st place.

Use the Cumulative Frequency column to find the half way value of the data.

USING TECHNOLOGY

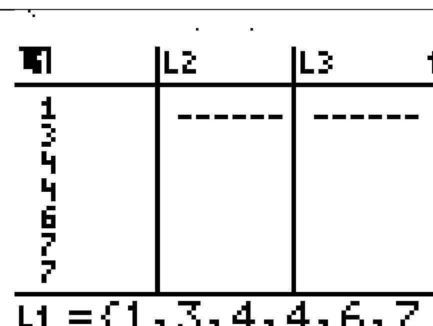
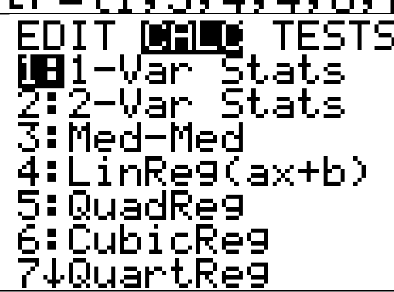
There are multiple device and websites that will perform Statistical Calculations.

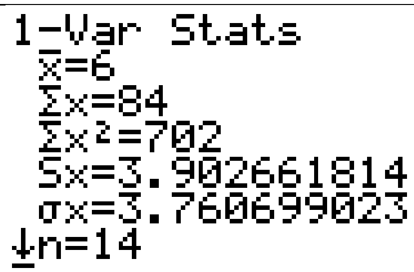
This Appendix will demonstrate the Texas Instruments Device and the more recently popular DESMOS App on your mobile device.
.
Most Spreadsheets of course will also do the calculations.

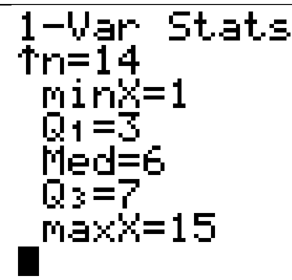
TEXAS INSTRUMENTS GRAPHING CALCULATORS

Given the data set: {1, 3, 4, 4, 6, 7, 7, 7, 8, 2, 2, 6, 15, 12}

a. determine the mean, median, mode from a single list
b. determine the mean, median, mode from a frequency table
b. make a histogram of the data from a frequency table
c. determine the quartiles

Enter the 14 data into the TI-83 into **L_1** in the familiar **STAT** **EDIT** process	
Select **STAT** **CALC** **1-VAR STATS**	
PRESS **ENTER**	





It is an extended screen in two parts, cursor down for all the values.

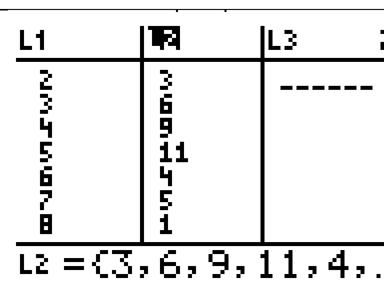
The mean is 6; the sum of all the data is 84, the sum of the each data element squared is 702 (we do not use that directly); the $\mathbf{S_x}$ and $\boldsymbol{\sigma_x}$ are measures of variability (sample deviation and population deviation) which we do not consider here, the **n, sample size,** is the number of data, the **minX** is x_{min}, **Q1** is the first quartile, **Med** is the median (or Q2), **Q3** is the third quartile, and **maxX** is the X_{max} of our data set.

ENTERING DATA FROM A FREQUENCY TABLE

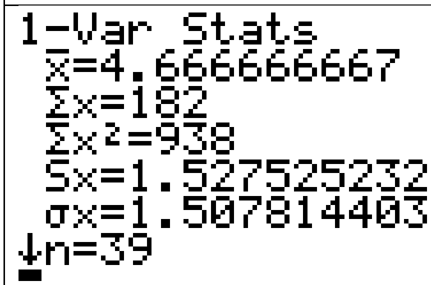
One you get past a couple dozen data values and given that many of them are the same it is very useful to enter data in a frequency table. Often the data has been grouped too into average class values.

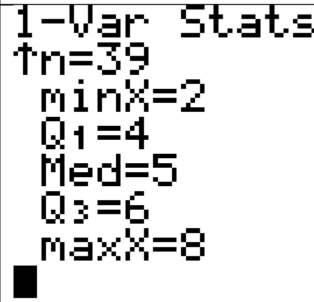
Value	Frequency
2	3
3	6
4	9
5	11
6	4
7	5
8	1

Enter data into List 1 and List 2



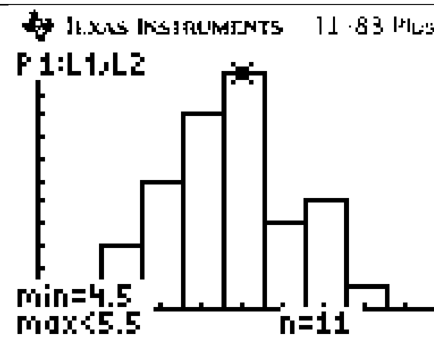
But this time make sure when you do the Stats Calculation that you tell the Calculator that the two lists you are using for data and for Frequency.





The Texas Instruments family of calculators will also graph the data for you if you are familiar with the graphing buttons.

See your teacher if you want to see how!



DESMOS GRAPHING CALCULATOR AND STATISTICS

A more recently popular mathematical suite of tools is the DESMOS suite of Math Tools. The DESMOS Graphing Calculator is especially helpful for Statistical Calculations.

The DESMOS Graphing Calulator is available on line in your browser: https://www.desmos.com/calculator

or as an App for download in Google and Apple devices

Here are some examples of it’s capabilities.

Determine some statistics and a histogram for the data set:

{ 2, 2, 3, 4, 4, 4, 4, 5, 5, 5 ,6, 6, 6, 6, 7, 7, 8, 9 }

1

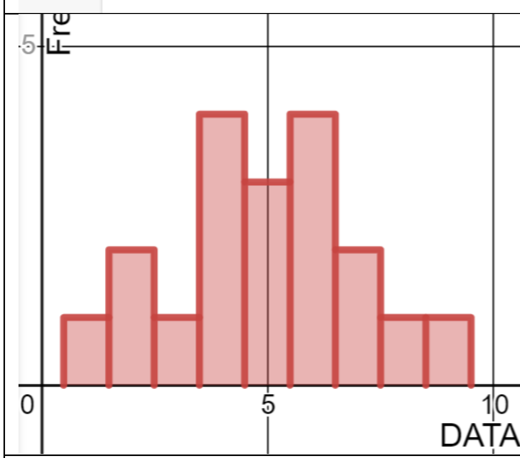
$$\mathrm{mean}([2,2,3,4,4,4,4,5,5,5,6,6,6,6,7,7,8,9])$$

$$= 5.16666666667$$

2

$$\mathrm{stats}([2,2,3,4,4,4,4,5,5,5,6,6,6,6,7,7,8,9])$$

Min	2
Q1	4
Median	5
Q3	6
Max	9



You’ll figure it out!

Have fun!