MTH1W – Unit 9: Statistics

Name: Ms. Jacob

Lesson #9.1: Intro to Statistics

Date: May 11, 2026.

Learning Goal: We are learning to understand the different types of data and how to calculate basic statistical summaries.

What is data? Data is information and its RAW form is just a bunch of stuff which has no CONTEXT attached to it.

STATISTICS → It is a branch of Mathematics that deals with COLLECTING data, ORGANIZING data, ANALYZING data, and INTERPRETING data.

Why Study Data?

EDUCATION is based on statistics!

- How effective is this teaching technique for improving student's problem-solving abilities?
- How does improving critical thinking in math class impact a student's ability to think critically about world issues?
- How can we reduce discrimination in education?

PSYCHOLOGY is based on statistics!

- How effective is drug X at reducing anxiety?
- What community programs are most likely to promote healthy lifestyle choices?
- What neural patterns are correlated with a sense of well-being?
- What contexts exacerbate racism? What contexts reduce it?

ADVERTISING and MARKETING is based on statistics!

- Who is the best target audience for this advertising campaign?
- What advertisements will our audience respond to?
- How do you create a 'viral' advertisement?
- What is the probable payout of an advertising campaign?

BUSINESS, FINANCES, and INSURANCE is based on statistics!

- How much risk does the insurance company take on to insure this person, house, car, etc?
- How likely is this stock to grow or shrink?

MEDICINE is based on statistics!

- How effective is a drug for reducing inflammation?
- What environmental and genetic factors contribute to a person's likelihood of getting cancer?
- What are a person's chances of survival given different treatment options?

Statistics is a vast and growing field of mathematics and every day we are exposed to many stats and probabilities. What is the chance that it will rain tomorrow? What is the average in my classes? Am I a normal height for my age? Companies are using statistics and data every minute to make decisions. Netflix uses what you watch to curate what is on your main screen and helps them to decide to make more similar content. Garbage collection needs to be done in an optimal way to save on gas and trips to and from the dump. Sports teams are hiring statisticians as coaches to help makes player decisions. The list is unending. In this unit, we will explore some of the basics ways to summarize and analyze data.

But we also need to recognize that "**Not all Statistics is True!!"**

It is pretty clear that statistics is everywhere. They are in Psychology, health, law, sports, business, entertainment, policy making, and every other discipline. However, Statistics are often added in reports to add credibility which is so evident in the world we live in today. People see numbers and often accept them without questioning. Such numbers and stats, thus, mislead people especially the ones who are not informed about research and statistics. However, not all stats are misleading and it is important to study statistics also to have a greater appreciation for it. You should be able to recognize where the statistical evidence strongly supports conclusions made.

So, it is really important to learn statistics because it will help us to recognize which findings to believe and make you an informed consumer of Statistics!!

The first step to being an informed consumer of statistics is to question the statistics. Check closely the numbers, the sources, and most importantly the procedures used to obtain the findings.

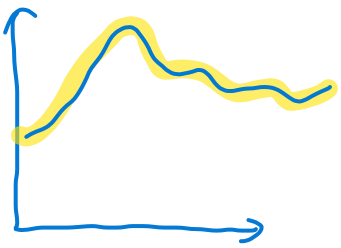
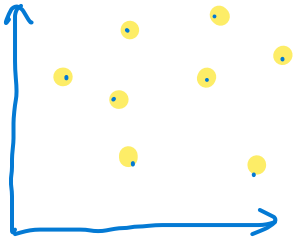
First, we need to understand different types of data. The first two types are called **qualitative** and **quantitative**.

Qualitative	Quantitative
• It is hard to express using numbers. So, they are descriptive in nature. - words, imagination, use of senses, etc. • These need not be facts bcz there may be an opinion attached.	• It can be expressed using numbers. easily • They are facts - How much? - How many? - How often?

Example: Given the following picture, describe what you see using both qualitative and quantitative data.

QUALITATIVE data	QUANTITATIVE data.
smooth dark brown book shelf rectangular prism	3 sections $w = 29\,cm$, $h = 80\,cm$, $l = 40cm$ h for each section $= 25cm$

Quantitative data has two types of data, **continuous** and **discrete**.

Continuous	Discrete
	
• It is measured like height, weight, etc. • It can be decimals, fractions, whole #s etc.	• It is counted # of people in a room, # of chips in a packet. • Always uses counting #s

A small introduction to **Inferential Statistics**

With inferential statistics, we generally take a **SAMPLE,** or a small subset of a larger set of data. We then use this sample to draw inferences about the **POPULATION** as a whole.

Let's say you wanted to know:

- Which political party will win the next provincial/federal election?
- How do Ontario students compare to students in Quebec in their mathematical problem-solving ability?
- Which sports person is most loved by Canadians?

You can't ask every Canadian (the whole population) these questions. And if you ask these questions to just your friends, you will be facing the issue of a terribly biased answer. So, one thing that is clear is that you definitely need to collect the data from a sample but the sample must be an unbiased **RANDOM** sample.

Inferential Statistics assumes that the sample is random but because it is very easy for a sample to be biased, so there are many strategies that researchers adopt in an attempt to eliminate or **decrease the bias** in the sampling. One of the strategies used is **SIMPLE RANDOM SAMPLING.**

Simple random sampling occurs when every member of the sample has an equal chance of being selected into the sample. In addition, the selection of one member is independent from the selection of another member. Thus, in theory, it is a selection by **PURE CHANCE**.

Consider the following examples:

1. *A substitute teacher wants to know how students in the class did on their last test. She asks only the 10 students sitting on the front row of the class to report how they did on their last test and she concludes from them that the class did extremely well.*

Sample? Population? Problems with the Conclusion?

Now what if the substitute teacher put all the names of the students in the class on a piece of paper and put it in a hat from which she drew out 10 names at random?

2. *A coach is interested in how many cartwheels the average Grade 9 student can do in his high school. The coach chooses 8 'volunteers', all of whom happen to be women, and concludes that Grade 9 students can do an average of 16 cartwheels in a row without stopping.*

Sample? Population? Problems with the Conclusion?

You may argue that sometimes it may simply not be possible or feasible to take a truly random sample. You are right!! We have to accept that estimates derived from using a sample have a chance of being inaccurate. This cannot be avoided unless we measure the entire population. The researcher has to accept that there could be variations in the sample due to chance that lead to changes in the population estimate. The difference between the true parameter and the statistic obtained by sampling is called **sampling error** and a statistician would report: "I am fairly confident that the true number of my result is actually between _____and _____ (interval estimate)''.

Error in sampling is called **bias**. Statisticians go to great lengths to avoid the many potential sources of bias.

Measures of CENTRAL TENDENCY

Lastly, let's look at how we can calculate some statistical summaries: mean, median, mode and range.

The **mean** is the AVERAGE. This is an easy, yet tedious calculation:

$$\text{Mean} = \frac{\text{Sum of all data values}}{\#\text{ of data values.}}$$

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$ $\quad \therefore M = \dfrac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$

The **median** is the Middle number. Calculating the median depends on whether you have an even or odd list of numbers.

But First place the #s in an ORDER (↗ ↘) ascending / descending

ODD data values

$x_1 \quad x_2 \quad \boxed{x_3} \quad x_4 \quad x_5$

Median

EVEN data values

$x_1 \quad x_2 \quad \boxed{x_3 \quad x_4} \quad x_5 \quad x_6$

$\text{Median} = \dfrac{x_3 + x_4}{2}$

The **mode** is the number that shows up the most. If there is no mode, state "no mode". If there is more than one, they state all the numbers that apply.

eg 1, 3, 5, 1, 7, 9, 3, 6, 2 $\quad \therefore$ Mode = 1 and 3

The **range** is a measure of spread, meaning how wide or narrow is the spread of the data is. The range is calculated by:

$$\text{Range} = (\text{Highest data value}) - (\text{Lowest data value})$$

Two examples: Calculate the mean, median, mode, and range for each data set.

Hours Slept

7.5 6 7.25 6.5 7.25 6
6 6 5.75 7 7.5

$$\text{Mean} = \frac{\text{Sum}}{\#} = \frac{72.75}{11} \approx 6.61$$

5.75, 6, 6, 6, 6, 6.5, 7, 7.25, 7.25, 7.5, 7.5

Median = 6.5

Mode = 6

$$\therefore \text{Range} = 7.5 - 5.75 = 1.75$$

Minutes to Run 5km

27.3 17.9 34.3 31.8 36.9 29
34.1 33.4 29 22.3

$$\text{Mean} = \frac{\text{Sum}}{\#} = \frac{296}{10} = 29.6$$

17.9
22.3
27.3
29
29
31.8
33.4
34.1
34.3
36.9

$$\text{Median} = \frac{29 + 31.8}{2} = \frac{60.8}{2} = 30.4$$

Mode = 29

$$\text{Range} = 36.9 - 17.9 = 19$$

Success Criteria:

- I can identify the difference between qualitative vs quantitative data and discrete vs continuous
- I understand the difference between population and a sample
- I recognize that in a simple random sample every member of the population has an equal chance of being chosen
- I can calculate the mean, median, mode, and range from a given data set