FOM 11 **5.1 Exploring Data**

Statistics is a field of mathematics that deals with the collecting and summarizing of data. There are three measures of central tendency:

1. **Mean** - average
   – Computed by adding a set of values and dividing by the number of values.

2. **Median** – center or middle value
   – Computed by ordering the values from least to greatest, then taking the middle value or the average of the two middle values.

3. **Mode** - most frequent.
   – Computed by taking the value that occurs the most often.

Example 1: For the set of values: 1, 6, 3, 8, 9, 3, 6, 1, 6 determine the:

a. mean

b. median

c. mode

Example 2: Ten numbers have a mean of 37. If one is removed, the mean is 38. What number was removed?

Example 3: The following table gives the frequency distribution of the number of orders received each day during the past 50 days at the office of a publishing company.

# of Orders	# of Days
10-12	7
13-15	12
16-18	17
19-21	14

Calculate the mean, median, and mode.

Example 4: Paulo needs a new battery for his car. He is trying to decide between two different brands. Both brands are the same price. He obtains data for the lifespan, in years, of 30 batteries of each brand, as shown below.

Measured Lifespans of 30 Car Batteries (years)									
Brand X					Brand Y				
5.1	7.3	6.9	4.7	5.0	5.4	6.3	4.8	5.9	5.5
6.2	6.4	5.5	5.7	6.8	4.7	6.0	4.5	6.6	6.0
6.0	4.8	4.1	5.2	8.1	5.0	6.5	5.8	5.4	5.1
6.3	7.5	5.0	5.7	8.2	5.7	6.8	5.6	4.9	6.1
3.3	3.1	4.3	5.9	6.6	4.9	5.7	6.2	7.0	5.8
5.8	6.4	6.1	4.6	5.7	6.8	5.9	5.3	5.6	5.9

a. Describe how the data in each set is distributed. Describe any similarities or differences between the two sets of data.

b. Explain why the mean and median don't fully describe the difference between these two brands of batteries. Why can additional information be learned from the **range** of the data?

c. Is the mode useful in this situation?

Assignment: Pg. 211 #1-3

FOM 11 **5.2 Frequency Tables, Histograms and Frequency Polygons**

A **frequency distribution** is a set of intervals (table or graph) into which raw data is organized; each interval is associated with a frequency that indicates the number of measurements in this interval.

A **histogram** is the graph of a frequency distribution, in which equal intervals of values are marked on the horizontal axis and the frequencies associated with these intervals are indicated by the areas of the rectangles drawn for these intervals.

A **frequency polygon** is the graph of a frequency distribution, produced by joining the midpoints of the intervals using straight lines.

Example 1: The following is a set of test scores out of 100.

45	68	94	76	89	99
53	79	61	61	98	72
61	81	98	59	91	68
72	32	57	69	42	78
48	87	78	74	93	71

a. Create a frequency table.

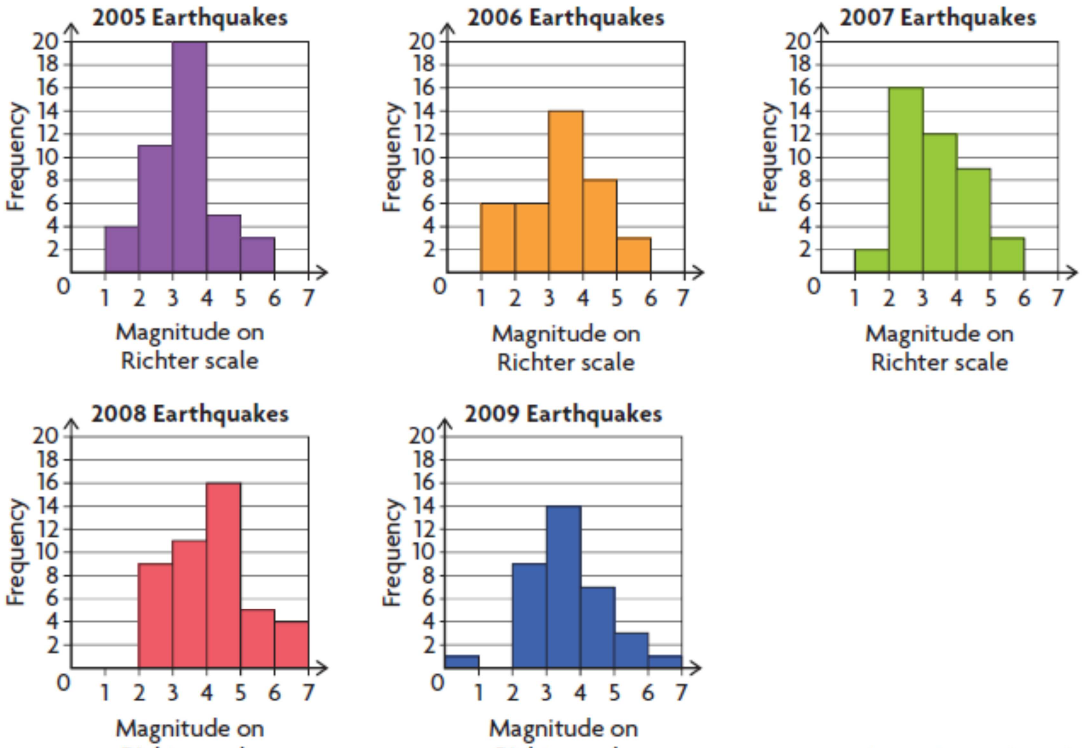
Score	**Tally**	**Frequency**
0-10		
10-20		
20-30		
30-40		
40-50		
50-60		
60-70		
70-80		
80-90		
90-100		

b. Create a histogram for this data.

c. Create a frequency polygon of this data.

Example 2: pg. 218

The magnitude of an earthquake is measured using the Richter scale. Examine the histograms for the frequency of earthquake magnitudes in Canada from 2005 to 2009. Which of these years could have had the most damage from earthquakes?



National Research Council Canada

Understanding the Richter Scale*	
Magnitude	**Effects**
less than 3.0	recorded by seismographs; not felt
3.0–3.9	feels like a passing truck; no damage
4.0–4.9	felt by nearly everyone; movement of unstable objects
5.0–5.9	felt by all; considerable damage to weak buildings
6.0–6.9	difficult to stand; partial collapse of ordinary buildings
7.0–7.9	loss of life; destruction of ordinary buildings
more than 7.9	widespread loss of life and destruction

*Every unit increase on the Richter scale represents an earthquake 10 times more powerful. For example, an earthquake measuring 5.6 is 10 times more powerful than an earthquake measuring 4.6.

Assignment: Pg. 222 #3-6, 9

FOM 11 **5.3 Standard Deviation**

To describe data numerically, we often use two numbers:

1. **Mean**: the average

   Let $x_1, x_2, x_3, ..., x_n$ represent any set of values.

   Mean: $\bar{x} = \mu = \frac{\sum x_i}{n}$

2. **Standard Deviation:** a measure of the extent to which data cluster around the mean.

   Let $x_1, x_2, x_3, ..., x_n$ represent any set of values.

   Standard Deviation: $\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$

   The smaller the standard deviation, the more consistent the results and the closer the data to the mean.

Example 1: Calculate the standard deviation from the following sets of values:

a. 7, 8, 9, 10, 11

b. 7, 9, 11, 13, 15

Example 2: Calculate the standard deviation for the following sets of data:

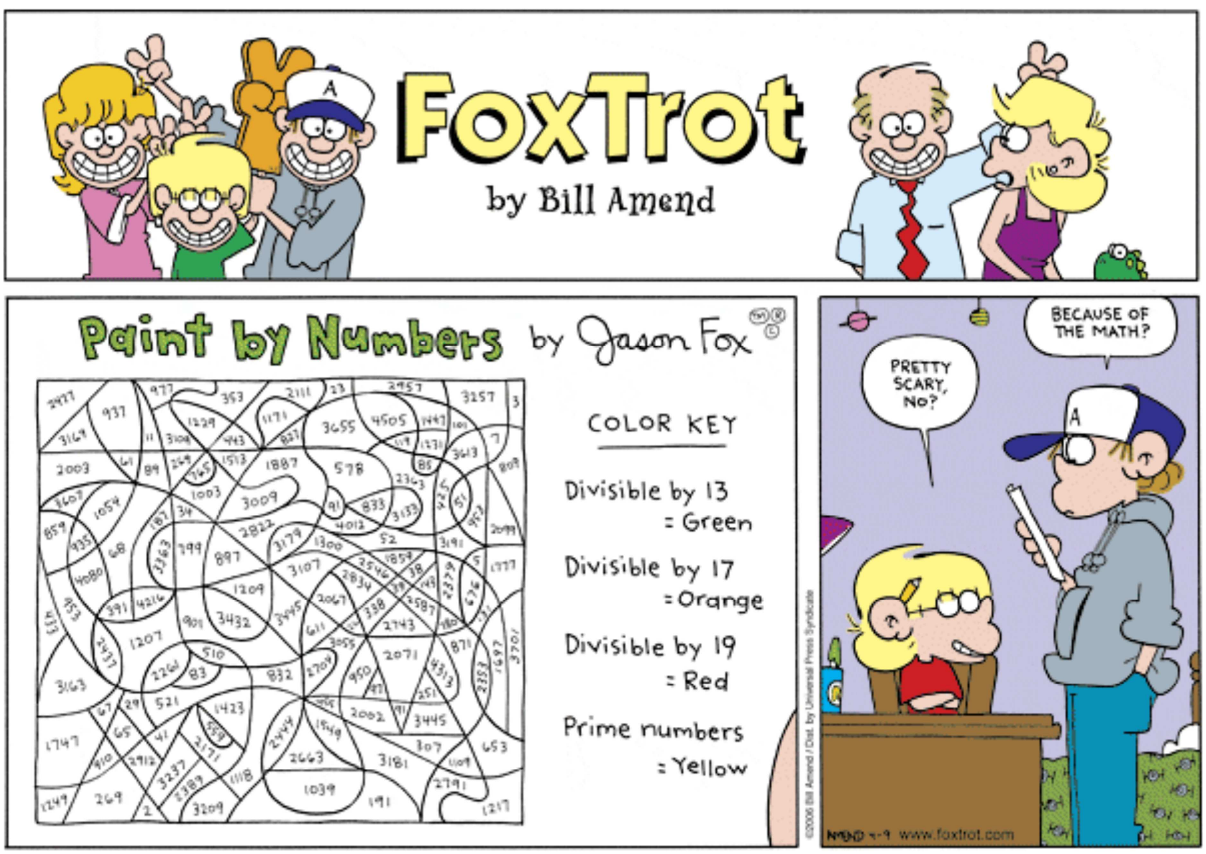
a.

Daily Commute Time (min)	**Number of Employees**
0-10	4
10-20	9
20-30	6
30-40	4
40-50	2

b.

# of Orders	# of Days
10-12	4
13-15	12
16-18	20
19-21	14

Assignment: Pg. 233 #1-3, 5, 6, 8, 11, 14



FOM 11 **5.4 The Normal Distribution**

Normal Distribution: Data that, when graphed as a histogram or frequency polygon, results in a unimodal symmetric distribution about the mean.

The **normal curve** is a symmetrical curve that represents the normal distribution. It is also called a **bell curve**.

Properties of a normal distribution:

- Has a mean μ and a standard deviation σ.
- Symmetrical about the mean.
- Almost all the population lies within 3 standard deviations of the mean.
- The horizontal axis is an asymptote.
- The total area under the curve is 1.
- It follows the 68-95-99.7 Rule.

Example 1: A company has determined that the lifetime of the car battery it manufacturers is normally distributed with a mean of 48 months and a standard deviation of 8 months.

a. Sketch it.

b. What percent have life spans less than or equal to 64 months.

c. What percent have life spans between 40 and 72 months?

d. Between which life spans do 95% of the batteries lie?

Example 2: Two baseball teams flew to the North American Indigenous Games. The members of each team had carry-on luggage for their sports equipment. The masses of the carry-on luggage were normally distributed, with the characteristics shown in the table.

Team	**μ (kg)**	**σ (kg)**
Men	6.35	1.04
Women	6.35	0.59

a. Sketch a graph to show the distribution of the masses of the luggage for each team.

b. The women's team won the championship. Each member received a medal and a souvenir baseball, with a combined mass of 1.18 kg, which they packed in their carry-on luggage. Sketch a graph that shows the distribution of the masses of their carry-on luggage change for the flight home.

Example 3: Shirley wants to buy a new cellphone. She researches the cellphone she is considering and finds the following data on its longevity, in years:

2.0	2.4	3.3	1.7	2.5	3.7	2.0	2.3	2.9	2.2
2.3	2.7	2.5	2.7	1.9	2.4	2.6	2.7	2.8	2.5
1.7	1.1	3.1	3.2	3.1	2.9	2.9	3.0	2.1	2.6
2.6	2.2	2.7	1.8	2.4	2.5	2.4	2.3	2.5	2.6
3.2	2.1	3.4	2.2	2.7	1.9	2.9	2.6	2.7	2.8

a. Does the data approximate a normal distribution?

b. If Shirley purchases the cellphone, what is the likelihood that it will last for more than 3 years?

Assignment: Pg. 251 #1-4, 6, 10, 11, 13, 14

FOM 11 **5.5 Z – Scores**

Since there are many different possible curves with different values of μ and σ, we can standardize the curve by transforming **each** score into a **z-score** (a measure of how many standard deviations a value is from the mean).

Standard Normal Distributions can be used in **every** problem for **any** data values.

Properties of a Standard Normal Distribution:

- Mean is 0.
- Standard Deviation is 1.
- Area under the curve is equal to 1.
- The graph is symmetrical about the mean.
- We use **z** instead of **x** to represent numbers along the horizontal axis.

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

- A(z) is the **<u>area under the curve to the left of z</u>**.
- We can find the areas by using a graphing calculator or a z-table.

Example 1: If IQ scores are normally distributed with a mean of 100 and standard deviation of 15, determine:

a. the z-score for 120.

b. the probability that a randomly selected person has an IQ less than 120.

Example 2: The GPA at GW Graham Secondary is 2.6, with a standard deviation of 0.5. If the top 10% of all students are eligible to attend UBC, what is the minimum GPA needed to attend UBC?

Example 3: At a high school, the average grade for Science is 66, with a standard deviation of 10. If 20 students with grades between 73 and 85 receive B's, how many students are taking Science at the high school?

Example 4: A manufacturer of cell phones has determined a mean of 26 months before a need of repairs, with a standard deviation of 6 months. What length of time for this warranty should the manufacturer set so that less than 10% of all cell phones will need repairs during the warranty period?

Assignment: Pg. 264 #5-7(ac), 8, 9, 11, 13-15, 17, 20

FOM 11 **5.6 Confidence Intervals**

We often hear statements such as:

"80% of dentists recommend this toothpaste with results accurate within 2 percentage points, 19 times out of 20".

When we interpret this result, we need to consider the **confidence interval**, the **margin of error**, and the **confidence level.**

A confidence interval is the interval in which the true value is estimated to lie, with a stated degree of probability. The confidence interval may be expressed using ± notation, such as 54.0% ± 3.5%, or ranging from 50.5% to 57.5%.

The **margin of error** is the possible difference between the estimate of the value you are trying to determine and the true value for the population. (Usually expressed as a plus or minus percent, such as ± 5%)

The **confidence level** is the likelihood that the result for the "true" population lies within the range of the confidence interval.

Example 1: A survey of 320 users of the skateboard park indicates that 40% of them would like the parks board to extend the evening use of the facility. This survey is considered accurate to within 5.4%, 19 times out of 20.

a. Calculate the range of people that want to extend the evening use of the facility.

b. Determine the accuracy of the results.

Example 2: A poll was conducted to ask voters the following question: If an election were held today, whom would you vote for? The results indicated that 53% would vote for Smith and 47% would vote for Jones. The results were stated as being accurate within 3.8 percent points, 19 times out of 20. Who will win the election?

Example 3: Polling organizations in Canada frequently survey samples of the population to gauge voter preference prior to elections. People are asked:

1. "If an election were held today, which party would you vote for?" If they say they don't know, then they are asked:

2. "Which party are you leaning toward voting for?"

The results of three different polls taken during the first week of November, 2010, are shown below. The results of each poll are considered accurate 19 times out of 20.

Polling Organization & Data	Conservative (%)	Liberal (%)	NDP (%)	Bloc Quebecois (%)	Green Party (%)	Undecided (%)
Ekos	29	29	19	9	11	12.6
sample size, 1815	margin of error, ±2.3%					
Nanos	37	32	15	11	5	19.2
sample size, 844	margin of error, ±3.4%					
Ipsos	35	29	12	11	12	n.a.
sample size, 1000	margin of error, ±3.1%					

source: http://www.sfu.ca/~aheard/elections/polls.html

a. How does the sample size used in the poll affect the margin of error in the reported results?

b. Compare the confidence intervals for the Liberal Party for each of the three polls. How does the sample size used in the poll affect the confidence interval?

Example 4: To meet regulation standards, baseballs must have a mass from 142.0 g to 149.0 g. A manufacturing company has set its production equipment to create baseballs that have a mean mass of 145.0 g. To ensure that the production equipment continues to operate as expected, the quality control engineer takes a random sample of baseballs each day and measures their mass to determine the mean mass. If the mean mass of the random sample is 144.7 g to 145.3 g, then the production equipment is running correctly. If the mean mass of the sample is outside the acceptable level, the production equipment is shut down and adjusted. The quality control engineer refers to the chart shown below when conducting random sampling.

Confidence Level	Sample Size Needed
99%	110
95%	65
90%	45

a. What is the confidence interval and margin of error the engineer is using for quality control tests?

b. What is the relationship between confidence level and sample size?

c. After making adjustments in equipment, the quality control engineer decided that the mean mass of baseballs must lie in the range 144.2 g to 146.4 g.

   i. What is the mean mass and the margin of error?

   ii. Will the new baseballs meet regulation standards?

Assignment: Pg. 274 #3, 4, 6-9