# MATH 213 – Homework 5

1. **Reading:** Please read Chapter 5 of our textbook as outlined below. We have been focusing on this chapter for the past week or so. Follow the guidance below and submit your responses to the reading questions labeled (a) – (m). You can type your answers below. It’s helpful for me if you can either bold your answers or make them a different color.

* You can skip 5.1 if you like – we will spend time talking about scatterplots later on.
* Read 5.2.

1. What symbol is used in Statistics to denote a population mean?

The symbol is used in Statistics to denote a population mean is:



1. What symbol is used in Statistics to denote a sample mean?

The symbol is used in Statistics to denote a sample mean is:



* Read 5.3.

1. According to our textbook, for what kind of datasets are dotplots useful? (And what are they implying about when to use histograms?)

Dot plots are useful for small datasets only. Using it for large samples becomes hard to read.

The histogram resembles a more heavily binned version of the stacked dot plot in which is more readable compared to dot plots in term of large datasets.

1. Find a sentence that the textbook authors use to describe one of the distributions that you think is worded nicely and captures an aspect of the distribution well – a sentence that you could see yourself adapting for use in one of your own descriptions. Copy the sentence here.

For instance, there are many more loans with rates between 5% and 10% than loans with rates between 20% and 25% in the dataset.  (My comment: well-constructed)

* Notice that our textbook authors discuss stacked dotplots, histograms, boxplots and so-called “density plots”. We have studied other kinds of graphs also: frequency polygons and jittered dotplots.
* Make sure to read about unimodal, bimodal, multimodal distributions. Try the Guided Practice question for yourself.
* Read 5.4.

1. Our textbook authors say, as we did in class, that the standard deviation “roughly describes how far away a typical observation is from the mean”. Find a similar sentence from the text that describes the intuitive meaning of the *variance* of a dataset. Copy it here.

The variance is the average squared distance from the mean.

1. How are the standard deviation and the variance related to one another?

The standard deviation is the square root of the variance.

* Read 5.5 on box plots, quartiles, and the median.

1. List the statistics and other features of a dataset that are captured by a boxplot.

The statistics and other features of a dataset that are captured by a boxplot are:

the first Quartile, the Median, the third Quartile, the whiskers and the outliers.

1. What is the formal definition of interquartile range (IQR)?

The IQR interquartile range is the length of the box in a box plot.

1. Copy here the textbook’s definition of an “outlier”. Find a sentence that intuitively describes what an outlier is as well as a formula that is commonly used to calculate outliers.

An **outlier** is an observation that appears extreme relative to the rest of the data.

* Read 5.6 on robust statistics – we discussed this on 2/14 class and created some slides together.

1. Which measure(s) of center are considered robust? The measure (s) of center that are considered robust is the Median.
2. Which measure(s) of spread are considered robust?

The measure (s) of center that are considered robust is the two Quartiles (1st and 3rd).

1. Which measure of center and spread are good when describing symmetrical or near-symmetrical distributions? (May have to see our class slides.)

The measure of center and spread that are good when describing symmetrical or near-symmetrical distributions are the Median and the two Quartiles (1st and 3rd).

1. Which measures of center and spread are better for describing heavily skewed distributions? (May have to see our class slides.)

The measures of center and spread that are better for describing heavily skewed distributions are: the Standard Deviation and the Mean.

1. Turn in work for Question 16, parts a, b, c, from the [Chapter 5 exercises.](https://openintro-ims.netlify.app/explore-numerical.html#chp5-exercises)

**Distributions and appropriate statistics.** For each of the following, state whether you expect the distribution to be symmetric, right skewed, or left skewed. Also specify whether the mean or median would best represent a typical observation in the data, and whether the variability of observations would be best represented using the standard deviation or IQR. Explain your reasoning.

* 1. Housing prices in a country where 25% of the houses cost below $350,000, 50% of the houses cost below $450,000, 75% of the houses cost below $1,000,000, and there are a meaningful number of houses that cost more than $6,000,000.

I expect the distribution to be **Right skewed** because of “there are a meaningful number of houses that cost more than $6,000,000.” The Median would best represent this observation. The variability of observations would be best represented using the standard deviation due to outliers. The graph will be **unimodal.**

* 1. Housing prices in a country where 25% of the houses cost below $300,000, 50% of the houses cost below $600,000, 75% of the houses cost below $900,000, and very few houses that cost more than $1,200,000.

I expect the distribution to be **Symmetrical** because of the almost equal sharing of each quartile. The Median would best represent this observation. The variability of observations would be best represented using the IQR. The graph will be **unimodal.**

* 1. Number of alcoholic drinks consumed by college students in a given week. Assume that most of these students do not drink since they are under 21 years old, and only a few drink excessively.

I expect the distribution to be **Right skewed** because of that left outlier caused by the fact that drinking is only allow at 21 years old; in which result to very few people drinking before the age of 21years old. The Mean would best represent this observation. The variability of observations would be best represented using the standard deviation due to outliers. The graph will be **unimodal.**

1. Finish the R Lab activity “JFK and JFKC”. As well as following the directions to make the graphs and compute the statistics, pay close attention to the last question where you are asked to describe what is going on. Remember to speak in context, to talk about center, shape, spread, outliers, any other interesting features, and to compare the graphs on these items. Back up what you say with specific numbers you calculated!