

## Data Analytics Fundamentals

---

There are two parts to this Assignment: Part A and Part B. Please complete both parts.

### Part A

**Watch:** <https://youtu.be/X6uS8FiNTIU>

**To Do 1:** Download the Excel file "MGMT3540\_Assignment2\_Data", go to sheet "Part A".

Q1. Calculate Descriptives (mean, median, and mode of SALES) in Excel using Excel formulas.

*\*\*Hint: There is more than one mode in the data\*\**

### Part B

**Watch:** <https://youtu.be/jgjoyLADz3Y>

**To Do 2:** Download the Excel file "Manufacturing Plant Salary Data"

Q1. Calculate the average age, length of service, and salary for all employees in the plant.

Q2. Calculate the standard deviation for age, length of service, and salary.

Q3. Calculate average salary for Males and Females. Could the other variables (age, length of service) potentially explain any differences in the average salary for Males and Females. Explain.

Report your answers for both Part A and Part B in the Excel document (in the green cells) and upload it to Elearning.

Only EXCEL files are accepted.

Keep all of the formulas.

## Study Guide for Part A

---

### Overview

To get you started, I'm providing you with a step-by-step guide for Part A of Assignment #2. Try to complete the assignment on your own first before you look at my guide below. When you complete Part A on your own, come back and look at my guide below. Then move on to Part B.

### Part A Guide

Part A asks you to calculate summary statistics for fourth quarter sales in this data set. We're going to look at three particular summary statistics, the mean, the mode, and the median, and we're going to use Excel's built in functions to calculate them.

#### Mean

Let's start with the mean. You can calculate this in cell B30.

To do that, we'll click in cell B30 and we'll use the built-in formula. You start with the equal sign, and the formula for the mean is “=average()”. Then you need to point Excel to the data. You're going to do that by opening the parentheses, and then highlighting the cells between C7 and C21. You close the parentheses, and you hit enter, and you see the mean of this data set is 13.783.

#### Median

Next, we're going to calculate the median in cell B31.

To do that, we're going to use the built-in median function. We start with an equal sign, we use the median formula, which is “=median()”. We open parentheses, highlight C7 to C21, then we close the parentheses and hit enter. The median is 13.75.

#### Mode

The mode is the most frequent observation in the data set. We'll calculate the mode in B32. We'll use the equal sign and we'll use the built-in mode formula, mode. And then we'll open the parentheses and we'll highlight between C7 and C21. We'll close the parentheses and we'll hit enter. We'll see the mode is eight. Now if you look closely at the data, you'll notice that the value eight appears twice in our Q4 Sales column.

But it also looks like there are other values that also appear twice. This would be a circumstance where there are potentially multiple modes. In other words, multiple values that appear equally frequently in the data. So, Excel can help us with that as well, but we have to treat it a little bit differently. We use a different formula. I'm going to show you that now. The first thing I would do is give yourself some extra room by inserting several rows so that if there are multiple modes, you have plenty of room for Excel to tell you that.

We're going to delete our prior formula. Now we're going to highlight these four cells between B32 and B37 to give us room to see how many modes there are. Now for the formula we're going to use equal and this is a different formula. It's “=mode.mult()”, standing for multiple modes. We'll open the parentheses and we'll highlight between C7 and C21. Then we'll close parentheses.

Now because we're looking for multiple modes and we're using what's called an array, we need to not just hit enter, but on a PC, you have to do **shift + control + enter** at the same time. For Mac users, you need to do **shift + control + enter** at the same time as well (do not replace "control" with "command" as you usually do). And now, you get the output of a formula. What this tells us is there are actually three modes in this data set, 8, 11, and 15. The last value, N/A, tells you there are no additional modes.

So, looking at these summary statistics, what can we learn? Well first, noticing that the mean and the median are fairly close to each other means we likely have some type of normal distribution, but we have a unique data set in that we have three modes, eight, 11, and 15.

Now, watch the video and move on to Part B.