Exploring Data with Excel

File: `TL1_StudentsPerformance.xlsx`

Save As: `TechLab1_YYYMMDD_LastnameFirstName.xlsx`

Purpose:

This lab is designed to increase your proficiency with Excel for data analysis. You will:

- Practice cleaning and exploring real-world data.
- Construct and customize histograms.
- Compute summary statistics.
- Use descriptive analytics to interpret and compare results.

This lab also introduces responsible use of generative AI to help troubleshoot formulas and Excel issues, without outsourcing the analysis itself.

The data dictionary for the data in the Excel file is below for your use and referenced throughout this lab:

Variable	Description
ID #	Unique identifier for each test taker (nominal categorical)
test preparation course	Lists whether a test taker completed a test preparation course (binary categorical); either 'none' or 'complete'
math score	The number of points received on the math test out of 100 points (discrete numerical); range is 0 to 100.
reading score	The number of points received on the reading test out of 100 points (discrete numerical); range is 0 to 100.
writing score	The number of points received on the writing test out of 100 points (discrete numerical); range is 0 to 100.

Part 1. Instructor Walkthrough

The question we want to answer is: Does completing the test preparation course appear to impact math scores?

Before we begin, take two minutes to brainstorm how you might answer this question using the tools we've learned so far in this course. Write down your ideas.

We will walk through how to conduct this analysis together in Excel. Open your Excel file in the application on your computer (don't use the online application) and follow along as we build a model to explore and summarize the data.

1. **Clean the Data**. Use `Sort & Filter` ( ) to:

   - Find unusual entries or missing values.
   - Look for things like text in number fields, out of range values, or inconsistent categories.

   If you make edits or changes to the data: add a new worksheet and rename it to "Edits". Document what you changed and why it was reasonable and necessary. This worksheet should have the following column titles: `ID #`, `Variable`, `Edit`, and `Justification.`

2. **Separate the Data**. In a new worksheet, create two tables, one for the math scores of those who completed a test prep course and one for the math scores of those who did not.

3. **Construct a Histogram**.

   (a) Highlight the data for the students the completed the test prep course.

   (b) Then follow `Insert` → `Chart` → `Histogram` ( ) to insert a histogram into the worksheet.

   (c) Click on the chart that was created and then click on the button that appears on the upper right side of the chart. Ensure the `Chart Title` and `Axes` are checked, then check the box next to `Axes Titles` and `Legend`.

   (d) Change the title and axes titles in your chart to the appropriate titles:
   Chart title: Histogram of the Score Distribution for the Math Test
   x-axis title: Score out of 100
   y-axis title: Number of Students

   (e) **Format Legend**. First, let's adjust the location of the legend. Right click on the legend and then click `Format Legend`. In the side bar that opens, select `Right`. Then let's title the data to make it more clear. Right click on the legend and click `Select Data`.

   A new dialog box should open. Click on the text `Series1` in the `Legend Entries` box. Once Series1 is highlighted, click `Edit` (Edit). In the box that opens under `Series Name`, type in `Completed Test Prep`. Select `OK`.

   (f) **Change Bin Size**. Now let's adjust the bin size of the histogram to make the bins start and end on an integer. Right click on the x-axis labels. Select `Format Axis`. Click on `Bin Width`, then in the box next to the words type in 10. You should see that the first bin now starts at 23 and includes scores up to 33.

   (g) **Add Data**. Histograms in Excel cannot have two sets of data plotted on the same chart. However, you can follow the steps above to create another histogram for the data of those who did not complete test prep. Do this now.

4. **Calculate Summary Statistics.** In a new section of your worksheet, create a table that looks like the following:

	Completed	Not Complete
Mean		
Median		
Standard Deviation		

Within this table use the following formulas to calculate the statistics:

=`AVERAGE(RANGE)`

=`MEDIAN(RANGE)`

=`STDEV.S(RANGE)`

where `RANGE` is the cells that contain the data you want to calculate the statistics of, i.e. `A1:A5`.

5. What does this information tell you about the data for the math scores?

Part 2. Try for Yourself

Now conduct similar analysis, but answer the question: Does completing the test preparation course appear to impact writing scores?

Write your conclusion here:

Responsible Use of Generative AI

Generative AI tools can help you troubleshoot when Excel isn't working the way you expect.

Appropriate questions to ask when troubleshooting may be:

- Why is my STDEV.S formula returning a divide-by-zero error?
- What does _____ error mean?
- How can I change the bin size in my histogram?

Inappropriate questions or prompts would be:

- I've uploaded the data. Make a histogram of this data and write a summary
- Here is the prompt, do the lab for me.