Comparing Relative Standing with Excel

File: `TL2_StudentsPerformance.xlsx`

Save As: `TechLab2_YYYMMDD_LastnameFirstName.xlsx`

Purpose:

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

- Construct and customize scatter plots.
- Use equations and cell references to calculate z-scores.
- Use sort and filter tools to identify potential outliers.
- Compare students across multiple variables using measures of relative standing

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

Variable	Description
ID #	Unique identifier for each test taker (nominal categorical)
test preparation course	Lists whether a test take 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.

The question we want to answer is: Do students who completed the test preparation course perform differently across all three tests compared to those who did not?

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 analyze the data. At this time turn to your neighbors – double check their file is open in the application on their computer and not the online application.

1. **Separate the Data**. In a new worksheet, create two tables, one for those students who took the test prep course and those who didn't.

2. **Visualize the Data**. Now we will create a scatter plot to visualize the data points comparing math scores and reading scores.

   (a) Highlight the `math score` and `reading score` for the students who did not complete a test prep course. Then follow `Insert` → `Charts` → `Scatter Plot` () → `Markers Only` ()

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

   (c) Change the title and axes titles in your chart to the appropriate titles:

   Chart title: Comparison of Math to Reading Scores

   x-axis title: Math Score

   y-axis title: Reading Score

   (d) **Add More Data**. Right click on the center of your chart and click on `Select Data`. In the dialog box that opens click `Add` ().

      i. In the next box that opens, click in the `Series X values:` cell, then click and drag over the math scores of those who completed a test prep course.

      ii. In the `Series Y values:` cell, delete the `=1` and then click and drag over the reading scores of those who completed a test prep course.

      iii. Finally, in the `Series name` cell type: Completed Test Prep Course.

      iv. Click `OK`.

   (e) While you're in the `Select Data Source` dialog box, edit () the original series and change the name to: Did Not Complete Test Prep Course. Then click `OK`.

3. You should now have a scatter plot with two different colored markers. To change the color of a set of markers, right click on one of the points you want to change and select `Format Data Series`. In the panel that opens, click on the paint can () and then select `Marker` (). Use the drop downs for `Fill` and `Border` to change the coloring of the marker.

4. What seems likely about the math and reading scores for students who took the test prep course versus those who didn't take the course by looking at your scatter plot? Start your answer with *It seems likely that...*

5. **Calculate summary statistics.** For each score type, calculate the mean and standard deviation using `=AVERAGE(RANGE)` and `=STDEV.S(RANGE)`, respectively.

6. **Calculate z-scores.**

   (a) Next to the table containing the information for students who did not take a test prep course, add three extra columns titled: `z_math`, `z_reading`, and `z_writing`.

   (b) Now, in the first cell underneath `z_math` type the equation `=(C2-$J$2)/$J$4`, changing the cell numbers to represent where you have the data stored.

> **Use Generative AI**
> Using your preferred generative AI tool, ask the following question:
>
> In Excel, what do dollar signs around a cell mean? For example, if I type C3 in an equation, what does that mean?
>
> Based on what you learned, write down how you would explain to a confused classmate what it is beneficial to write the equation in 6.b with no dollar signs on the reference to C2 and with dollar signs on the references to J2 and J4.

   (c) Auto-populate the equation down the column to calculate the z-scores for all of the math scores by double clicking in the bottom right hand of the cell (your cursor should look like a crosshair).

   (d) Calculate the z-scores for the other two columns (`z_reading` and `z_writing`) by modifying steps b-c. Make sue you use the mean and standard deviation corresponding to the appropriate columns. Before typing anything into your computer, write the exact equation (starting with the $\bar{}$) that you plan to type into the first cell underneath `z_reading` and `z_writing`.

   (e) Which student has the highest z-score in math? Which student has the lowest?

   (f) **On Your Own.** Calculate the z-scores for the scores of the students who took the test prep course. Which student has the highest z-score in math? Which student has the lowest?

7. **Compare Across Variables.**

   (a) Now let's assess overall performance using z-scores. Add another column next to the `z_writing` in the non-test prep dataset and label it `Cumulative`.

   (b) Type the formula `=SUM(RANGE)` and select the cells with the different z-scores as the range. This calculates the cumulative score for each student in relation to the overall average.

Reminder: Z-scores tell us how far above or below average a value is, in standard deviations. A cumulative z-score gives a quick way to see which students perform consistently above average across all subjects.

(c) Which student(s) have the highest cumulative z-score and which have the worst? Write down these scores.

(d) **On Your Own**. Do the same thing, but for the students who took the prep course. Which student(s) have the highest cumulative z-score and which have the worst? Write down these scores.

8. Reviewing your scatter plot and cumulative scores, answer the following:
   - What take-aways can you draw about the overall performance of students who took the prep course versus those who did not?
   - Are there any surprising outliers or patterns worth noting?

9. Reflect. Take 2 minutes – what did you learn today? What questions do you still have? What new questions do you have? How could you see yourself using what you learned.