TPS REPORT 04

Comparing Data with Descriptive Statistics in Excel

1. Download the TPS 04 Excel file from *www.terevaka.net/nau/ant568/tps01.html*

2. In this report, you'll be analyzing data meant to represent water quality from three different towns. Column **A** contains 150 different measurements of the number of micrograms (*μg*) of lead in each decaliter (*dal*) of water collected at different locations within Town 1. Column **B** contains analogous data for Town 2, and Column **C** contains analogous data for Town 3.

3. In Cell **F2**, write a formula to calculate the average amount of lead in each sample drawn from Town 1.

   - Write your formula in your TPS Report.

   Now Copy your formula and Paste it to Cells **G2** and **H2**.

4. Next you'll calculate the mode for the data set. It's possible that one or more of the town datasets will contain multiple modes, so select Cells **F3** through **F7**, and then use the **mode.mult** primitive in an *array formula* to calculate up to five possible modes for data from Town 1 (Hint: see TPS 02 if you don't remember how to calculate an array formula).

   Repeat the process for Town 2 (Cells **G3** to **G7**) and Town 3 (**H3** to **H7**).

5. Calculate the medians for the data for Towns 1, 2, and 3 in Cells **F8**, **G8**, and **H8** respectively.

6. Calculate the standard deviation for samples from Town 1 in Cell **F9**. Be careful – in Excel there are a number of choices when it comes to primitives related to standard deviation.

   - Write your formula in your TPS Report.
   - In your TPS Report, explain which primitive you chose and why.

   Now calculate standard deviations for data from Towns 2 and 3 in Cells **G9** and **H9**.

7. Excel provides us with hundreds of primitives and an almost limitless supply of potential analyses. However, it's still useful in a pinch to know how to calculate standard deviation (or variance) on your own.

   - In Word, use the π Equation button on the INSERT tab at the top of your window to create a general equation for an unbiased standard deviation in your TPS report. Use σ to represent standard deviation, μ to represent the mean of all samples, X

to represent the individual sample value, and N to represent the number of samples.

8. Write an equation in Cell **F10** to calculate the variance for data from Town 1.

   - Write your formula in your TPS Report.

   Copy your equation from Cell **F10** and Paste to Cells **G10** and **H10**.

   - Copy all cells between **E1** and **H10** and Paste your table in your TPS Report.

9. Before proceeding to more questions, create individual scatter plots for data from each town to help you visualize the distribution of lead measurements.

   - Copy *each* scatter plot and Paste it in your TPS Report.

10. Suppose that you are asked which town has the most lead in its water supply, and to support your conclusion based on a single statistic.

    - Write your answer and justification to this question in your TPS Report.

11. Suppose that you are asked which town has the most variability in lead content in its water sources.

    - Write your answer and justification to this question in your TPS Report.

12. Suppose following sampling of water quality, the three towns implement a treatment procedure to reduce lead content. Town 1 uses a treatment procedure that will randomly reduce lead measurements to anywhere between 60% and 75% of what they were in Column **A**. In Cell **J2**, write a formula to offer a hypothetical recalculation of the value in Cell **A2**, given these parameters.

    - Write your formula in your TPS Report.

    Copy your formula from Cell **J2** and Paste it to Cells **J3** through **J151**.

13. Town 2 uses a treatment procedure that will randomly reduce lead measurements to anywhere between 80% and 85% of what they were in Column **B**. In Cell **K2**, write a formula to offer a hypothetical recalculation of the value in Cell **B2**, given these parameters.

    Copy your formula from Cell **K2** and Paste it to Cells **K3** through **K151**.

14. Town 3 uses a treatment procedure that will randomly reduce lead measurements to anywhere between 50% and 90% of what they were in Column **C**. In Cell **L2**, write a formula to offer a hypothetical recalculation of the value in Cell **C2**, given these parameters.

    Copy your formula from Cell **L2** and Paste it to Cells **L3** through **L151**.

15. Write the appropriate formulas or array formulas to calculate values for Cells **O2** through **Q10**.

    - Copy all cells between **N1** and **Q10** and Paste your table in your TPS Report.

16. If you had to select one treatment procedure, based on the results in Cells **O2** through **Q10**, which town's treatment procedure would you select?

    - In your TPS Report, write a paragraph explaining and justifying your answer. Make sure you refer to specific statistics from your table(s).

17. The general public may often look to us, as scientists, to reduce the answer to a question to a single statistic or number, but this might also be a relatively poor representation of the overall distribution of data. Sometimes a table or even a graph does a better job of telling the story. In TPS Report 03, you created a histogram using the **Data Analysis** feature of Excel. Now you'll create a histogram using your own knowledge of primitives and analysis in Excel.

    In Cells **T2** through **T11**, write formulas that will count up the number of values from Column **J** in each of the bin ranges in Cells **S2** through **S11**.

    - Write your formulas in your TPS Report.

18. Repeat the steps from #17 to Paste the appropriate formulas to Cells **W2** through **W11** and Cells **Z2** through **Z11**, for treatment data from Towns 2 and 3 respectively.

19. Create a graph to represent the histogram data from Columns **T**, **W**, and **Z** all in the same visualization.

    - Copy the visualization and Paste it in your TPS Report.

{ average, mode.mult, median, min, max, randbetween, countif }