TPS REPORT 03

Measures of Central Tendency & the Normal Distribution in Excel

1. Download the TPS 03 Excel file from *www.terevaka.net/nau/ant568/tps01.html*

2. In this report, you'll be generating and analyzing data meant to represent the results of a health-related telephone survey in Flagstaff. Column **A** contains 50 different phone numbers that were included in the survey. Your first task will be to format the phone numbers. In Cell **B2**, try the following formula:

   ="("&left(A2,3)&")"

   Your output should be (928). The quotations tell Excel to render a text character, and the **&** allows us to concatenate characters, formulas, numbers, etc.

   Now delete the contents of Cell **B2**.

3. Create a new formula in Cell **B2** that will convert the contents of Cell **A2** to the format (###) ###-####. Make sure you include a space after the end-parentheses.

   - Write your formula in your TPS Report.

   Now Copy your formula from **B2** and Paste to Cells **B3** through **B51**. Add a heading in Cell **B1**.

4. In Column **C**, you'll randomly assign a sex to the telephone numbers. A simple way to create sex data where there is an equal likelihood of each phone number being assigned to "male" or "female" would be to use the formula:

   =if(randbetween(0,1)=0,"male","female")

   Enter that formula in Cell **C2**, then Copy the formula and Paste it to Cells **C3** through **C51**. Add a heading in Cell **C1**.

5. Next, you'll generate height and weight figures randomly generated from a uniform distribution. Since we'd expect certain differences based on sex, you'll need to sort your spreadsheet alphabetically according to the values in Column **C**. But before you do that, Copy all of Column **C** and then right-click on the highlighted area and **Paste Values**. That eliminates the risk of Excel generating new sex assignments during the sorting process.

6. On the HOME tab at the top of your screen use the button to perform a **Custom Sort**. Make sure you have checked the box to indicate that your **Data has headers**. Then sort alphabetically based on your Column **C**.

7. In Cell **C53** write a formula to count how many of your values between Cells **C2** and **C51** are "female".

   - Write your formula in your TPS Report.

8. In Column **D**, you'll generate height data in centimeters. Use the  button on the DATA tab at the top of your screen for this. Using the **Random Number Generation** option, you'll generate height for the females. Make sure your **Distribution** is **Normal**. You'll be generating only one variable (height), but you'll generate the same **Number of Random Numbers** as your value obtained in Cell **C53**. The **Mean** is the average value you want to generate. Suppose the average height of females is 163.5 cm. Use that value as your **Mean**. The **Standard Deviation** (a measure of the dispersion of a set of data from its mean) is 3.5 inches. To convert that to cm, Google "3.5 in to cm"—a handy way to convert almost any units in quantitative research to any other units.

   In the **Output Options**, select the **Output Range** bubble. Click in the  box, then drag your mouse beginning in Cell **D2** down to select all of the cells in Column **D** that correspond to "female" values in Column **C**.

   Hit the OK button to complete the Random Number Generation action.

9. Now repeat the process above for the remaining Cells in Column **D** down to Cell **D51** to generate height data for the males. Use 177.8 as your **Mean** and 10.2 as your **Standard Deviation**.

   Add a heading in Cell **D1**.

10. Next you'll use the same routines to generate weight data in pounds in Column **E**. Use the following information:

	Mean	Standard Deviation
Male	195.5	29
Female	166.2	18

    Add a heading in Cell **E1**.

    Add borders between all Cells **B1** through **E51.**

    Now Copy that table with borders and

    - Paste it into your TPS report so that it fits on a single page.

11. Now you'll turn your attention to the **Summary** sheet. Create appropriate formulas to calculate the corresponding values for Cells **B2** through **E5**.

- In your TPS Report, write your formulas for Cells **B2**, **C2**, **D2**, and **E2**.

12. Copy Cells **A1** through **E5**.

    - Paste that table into your TPS report.

13. Your table likely includes at least one "#N/A" value within the **Mode** column.

    - Explain in your TPS report why Excel is not able to provide a mode in these cases.

14. You may be accustomed to the mean as the most popular measure of central tendency. However, in many cases the median is a better choice.

    - In your TPS Report, explain in your own words when the median is a more appropriate measure of central tendency.

15. In some cases, a single point within data is irrelevant as a measure of central tendency.

    - In your TPS Report, give a hypothetical example of when you would choose to use the range as your measure of central tendency.

16. Now you'll turn your attention to the **Coins** sheet. In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.[1]

    Essentially, this means that in theory we should be able to approximate a normal distribution through repeated coin-flip experiments.

    In Cell **A2**, create a formula such that there is an equal likelihood the cell value will result in "Heads" or "Tails".

    - Write that formula in your TPS report.

17. Now Copy your formula from Cell **A2** and Paste it to Cells **A3** through **A51**. You've just created a single experiment where a coin is flipped 50 times.

18. In Cell **A53**, create a formula that counts the number of "Heads" results in Cells **A2** through **A51**.

    - Write that formula in your TPS report.

[1] https://en.wikipedia.org/wiki/Central_limit_theorem

19. Now you'll repeat the experiment 49 more times. Copy all of Column **A** and Paste to Columns **B** through **AX**.

    Notice in Row **53** that even though we *expect* 25 heads in each experiment, it might turn out that way relatively infrequently.

20. To see if your experiment actually generated data approximating a normal distribution, you'll create a histogram using all of the values from Row **53**.

    On the DATA tab at the top of your screen use the 

button and select **Histogram**.

    Click in the  box next to **Input Range**, then drag your mouse beginning in Cell **A53** all the way across to Cell **AX53**.

    Then click in the  box next to **Bin Range**, then navigate to the **Bins** sheet, and drag your mouse beginning in Cell **A1** all the way down to Cell **A21**.

    For **Output Options**, click the **New Worksheet Ply**, and check the **Chart Output Box**. Hit the OK button to complete the histogram action.

    Now Copy the chart and

    - Paste it into your TPS report.

{ &, left, mid, countif, average, mode, median, min, max }