# probability lab
## Dr. Sarah's 1010: Introduction to Mathematics

Consumer Statistics and Probability: To recognize misrepresentations of studies and statistical data in the real world by applying statistical techniques and understanding the role of chance and probability

goals:
- critically analyze the role of probability and chance in real world situations
- utilize technology to adapt and use mathematical formulas that include cell referencing to answer real-world questions and interpret results
- explore quantitative connections within geographical regions
- communicate statistics and probability information in written documents

**Birthday Simulations**: In a simulation, we can have a computer mimic real-life scenarios so that we can understand the likelihood of an event happening.

1. Go to the **Birthday Simulation** from the assignment in ASULearn. First enter 34 for the **Number** in room, and then repeatedly **Simulate** until you have conducted 100 Trials or so. What is the probability of a shared birthday among at least two of the people, according to the simulation?

2. Based on this simulation only, is it likely that two people in a class with 34 people have the same birthday?

   Circle one:         no          yes

3. Next enter 15 for the **Number** in room, and then repeatedly **Simulate** until you have conducted 100 Trials or so. What is the probability of a shared birthday among at least two of the people, according to the simulation?

4. Based on this simulation only, is it likely that two people in a class with 15 people have the same birthday?

   Circle one:          no          yes

**Country Populations**: The world is awash with data we can easily access and analyze to come to our own conclusions. Appalachian's strategic plan has a statement about sustainability: "Sustainability at Appalachian State University is not a trend, it is a tradition. We are active stewards of our... interconnected financial, cultural and natural resources... We balance critical, creative and global thinking in a living laboratory..." Population sizes have a direct impact on sustainability such as resource distribution. We have already seen that country populations are important in financial measurements via the debt per citizen. We'll continue to see connections in statistics and probability, in comparison of countries, in deciding about policies on HIV testing, and more.

5. The World Bank notes: "Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates." What are some real-life considerations in the problems of trying to count everyone in a country?

6. Open the Excel file **WorldBank_Population2023.xlsx** from the assignment (I've collated the 2023 data and added some items to help you). Country populations are known to typically satisfy Benford's Law, which looks at the frequency of the first digit and finds a logarithmic pattern in the distribution. The first digit in many real-life data approximates a logarithmic trend (where 1 occurs about 30% of the time while 9 less than 5% of the time) such as country populations, elections, and stock prices. A common usage of Benford's law is in fraud detection.

7. In G2 type
   =LEFT(E2,1)
   and hit return. You'll see the first digit of Afghanistan's population in G2. This is Excel's formula for extracting from the very left side of an entry, i.e. the first digit. You'll also see increase by 1 a summary count of the number of countries whose populations have that first digit in the J column.

8. Back on G2, use Excel's fill down feature to fill the formula all the way to the last country. Here is a review of how to fill down from earlier labs: Move your cursor to the lower-right corner so that it turns into a black plus sign. Click so that you keep the black + sign, hold down, and fill down the Excel formula by scrolling down the column vertically. If you have problems, watch the fill down helper video in the assignment link on ASULearn.

9. Fill down G through all the countries. In column J you'll see the the numbers fill in. Fill in the following table from J:

| first digit | number of countries whose populations have that first digit |
|---|---|
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |

10. To create a geometric representation of the data—DO NOT fill down from the bottom right—but DO leave the plus sign thicker and both white and black ✚ as you click in the middle of J2 and scroll down to J10 so that only those numbers are highlighted. If you have problems, watch the highlighter helper video in the assignment link on ASULearn.

11. Next, **Insert** a **Clustered Column**, which might be under Recommended Charts, or under the 2-D Column chart options for column graphs, or might just say something like Column, depending on your system. You should see a graph that looks something like the one below except your graph won't have the percentages and it won't perfectly satisfy the idealized column chart showing Benford's law. Sketch your graph to the right (a rough sketch is fine):
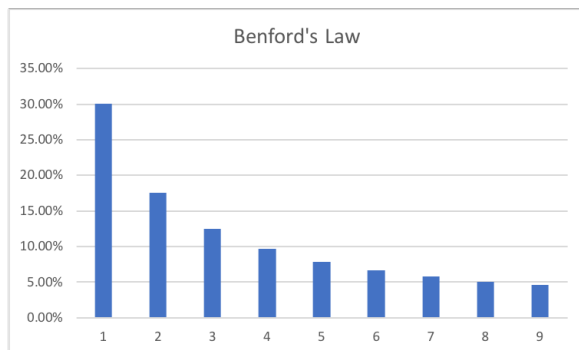
Benford's Law

12. Next compare your Excel chart to the idealized one and critically analyze how well the 2023 country population data fits Benford's law. Is there anything higher than we would expect from Benford's? Lower? If so, which ones?

13. The first country's population is in E2 (row 2 column E). Scroll down to the end of the data. What box contains the last country's population?

   E____

14. Excel's command to add entries in column E from row 2 to row 1000 would be =sum(E2:E1000) What command will add all the populations in the E column from row 2 to the row you found in your last response?

   =sum(E2:      )

15. Enter this entire command in Excel in F2 (don't forget the =). What is the total world population according to the World Bank file?

16. Search the internet for the world population in 2023 and write it here.

17. Examine the World Bank data. Name a country whose population data is missing?

18. Excel's command to find the smallest entry in column E from row 2 to row 1000 would be =small(E2:E1000,1). Use your answer in #14 to replace the 1000—type this this modified formula in F3 and find the smallest country population and then match it to the country by looking back at the E and A columns.

   What is the population?                    What country has this population?

19. Adapt the last command to find the largest country population by using the large command in F4, similar to the small command.

   What is the population?                    What country has this population?

20. Take a look at the A and E columns. Write down something you found interesting or surprising.

21. Go to
    `https://databank.worldbank.org/source/population-estimates-and-projections`
    which is where I obtained the data from, and open the **Database** items. Next, select a different **Database** that you are interested in, different than Population estimates and projections. Which did you select?

22. After selecting yours and choosing OK, next open **Series** and write down how many Series are Available for your Database. In the image below, it shows how many Series were available for the populations (189). Yours will be different. Today we only looked at one column, but this would tell us the number of higher dimensions of data we could work in!
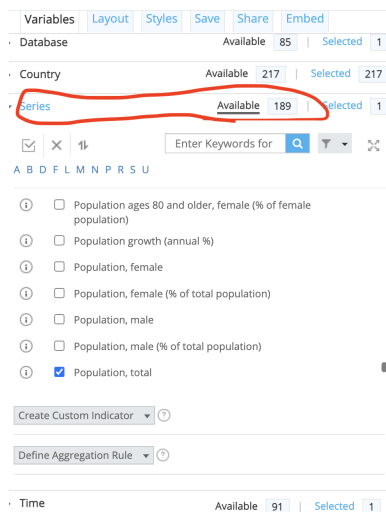


23. Write down a Series for your Database that you found interesting.

24. Collate your handwritten responses, preferably on this handout, into one full size multipage PDF for submission in the ASULearn assignment.