**Epidemiology Assignment 5**

Read the following sections of the CDC Online Epidemiology Manual:

Lesson 3: Measures of Risk: <https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson3/index.html>

[(Links to an external site.)](https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson3/index.html)

[Links to an external site.](https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson3/index.html)

Section 1: Frequency Measures:

Section 5: Measures of Association

Read Lesson 4 in the CDC Online Epidemiology Manual:

<https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson4/index.html>

[(Links to an external site.)](https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson4/index.html)

[Links to an external site.](https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson4/index.html)

        Section 2: Tables

Section 3: Graphs

Section 4: Other Data Displays

The main goal of this week's Epidemiology project is to help you become familiar with how Public Health data is analyzed, displayed, and interpreted. Chapter 6 of the textbook is filled with graphs and charts to express the findings of many epidemiology studies. This worksheet is designed to help you become comfortable with the many different concepts that epidemiologists deal with on a daily basis. If you want to play around with this a little more, I've included a couple of optional extra credit graphs for you to make yourself on Excel. Have fun!

**Worksheet: Epidemiology Module 10: Reading and Interpreting Graphs and Charts**

The main goal of the Epidemiology studies for the past few weeks has been to help you become familiar with how Public Health data is analyzed and displayed. Most readers only have time to size up the data quickly. They are looking for a snapshot of the data that will allow them to make a quick assessment of what is going on. It is for this reason that data is displayed in the form of graphs, tables, figures, etc. Actually, no one, not even the professional, really knows the answers to the critical questions when first looking at the numerical measurements (raw data). It’s usually not possible to know much, without first condensing it into a snapshot.

There is no single formula for displaying data. When the data is first being collected, it will be entered into a preliminary database, known as a line listing. This could be in the form of an Excel file for smaller data sets, or a larger type of data file that would be analyzed by more powerful software. This is reviewed in the CDC Manual in Lesson 4, Section 1.

“To analyze data effectively, an epidemiologist must become familiar with the data before applying analytic techniques. The epidemiologist may begin by examining individual records such as those contained in a line listing. This review will be followed by production of a table to summarize the data. Sometimes, the resulting tables are the only analysis that is needed, particularly when the amount of data is small and relationships are straightforward.”

Usually epidemiologists will take a preliminary look at the numbers to see what trends are standing out. By trends, we mean associations, correlations, etc. Which of these are important? Which are not.? Is a correlation an indication of a causal relationship? Is it merely a coincidence? Or, does it suggest the involvement of a third factor that links the first two? Where does the professional begin?

“When the data are more complex, graphs and charts can help the epidemiologist visualize broader patterns and trends and identify variations from those trends. Variations in data may represent important new findings or only errors in typing or coding which need to be corrected. Thus, tables and graphs can be helpful tools to aid in verifying and analyzing the data.”

If you take a look at an Excel program, you will notice that under the heading of charts you will see more than a dozen types of charts available. If you have an excel program available, and you know how to plot data, you can type some data into to a worksheet page to recreate one of the data sets shown in the CDC Manual. For example, type the data in from **Table 4.1a Reported Cases of Primary and Secondary Syphilis by Age — United States, 2002,** just as it is displayed.You will have two columns. You can make a bar graph. Designate the age brackets to be along the x-axis, and the frequency, or number of cases, along the as the y-axis. Practice adding labels to the data.

If you want to try something more complex, you can add some variables to your table. Type the gender data from **Table 4.2 Reported Cases of Primary and Secondary Syphilis by Age and Sex — United States, 2002** into your worksheet, and add those to your bar graph. You will get a bar graph similar to the graphs in Figure 6-1 on p.108 of your textbook.

**NOTE: This is not a required exercise, but you will receive extra credit if you do it (20 points for each graph). You should email this to me as a separate file, and explain what you did, so I’ll be sure to give you credit for it.**

**Assignment Worksheet:**

For your assignment, review the following sections: (please change the text in your answers to a different color or highlight with yellow).

Lesson 3: Measures of Risk: https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson3/index.html

 Section 1: Frequency Measures:

 Section 5: Measures of Association

Lesson 4: Displaying Public Health Data

 Section 2: Tables

 Section 3: Graphs

 Section 4: Other Data Displays

**Question 1:** Now, as you look through Chapter 6 in the textbook, you will recognize many of these types of charts, graphs or tables that were described in the CDC Manual. Let’s look at a few of them in more detail.

**In Figure 6-1. These two graphs are similar in that these are both both bar graphs with two variables being compared. They are both comparing frequencies during the individual years between 1978-2010.**

**1. Explain how are these two graphs are different.**

**Questions 2-7:** In Figure 6-4 these figures are comparing Risk Groups of AIDS patients from two different years. Answer these questions;

Look in Lesson 3: Measures of Risk

 Section 1: Frequency Measures

2.Which of the following parameters is being compared in this graph? Highlight your answer.

 a. Ratio

 b. Proportion

 c. Incidence Rate

3. Explain what the term “Cumulative” means in this analysis.

4. In comparing the data from 1986 to 2010, name 4 major changes that occurred in frequency among the different Risk groups.

5. Explain why the Blood Transfusion and Hemophiliac patient group appear to disappear?

6. Why does the frequency in the Homosexual group go down so dramatically? Are there fewer homosexual patients in 2010 than there were in 1986?

7. Why did the percentage of Heterosexual patients increase between 1986 and 2010?

**Questions 8-9:** When we see tables and graphs, we are often too quick to jump to conclusions about what they are trying to tell us. Read the Section in the textbook on 110-111, Epidemiology and Modes of HIV Transmission and Anal Sex - A High-Rsik Mode. Now look at Table 6-2.

Let’s see if you can figure out what this Table is telling you. The frequencies (percent HIV Seropositive) do not add up to 100%. They would have, if they had been comparing the Proportion of patients who had converted to Seropositive as the result of the different types of sexual activities. By reading the text carefully, you can figure out what these percentages are actually referring to.

8. What are the percentages in Table 6-2 referring to? Percent of what?

9. Explain the main finding of Table 6-2.

**Question 10:** Sometimes we see charts that we are not familiar with. We don’t know what to think! What do most people do when confronted with a chart they don’t understand? They skip over it! Figure 6-7 is an example of this. Actually, the finding of of this chart is dramatic. It would be a shame to miss out on it because you didn’t know how to read it.

You will find an explanation of this type of chart, also known as a pyramid chart in Lesson 4, Section 3 (scroll down to Figure 4.10 Population Distribution of Zambia by Age and Sex, 2000). Read that section, then use it to interpret the pyramid chart in Figure 6-7, on p. 119 of the textbook.

The goal of this graph is to examine how the age distribution of the population of Lesotho, a country in southern Africa, has changed over time as a result of the AIDS epidemic. For the graphs in this figure, they are looking at a the Proportion of the population in each age group.

Each bar represents what percentage of the population was in each age group during the year that they are plotting (age 0-5; 6-10; etc). The total should add up to 100%. Note: This is a Proportion that they are looking at (see Lesson 3, section 1).

10. Explain why the overall shape of the pyramid changed between 1950 and 2007. What is the main finding of this pair of charts?