Project EDDIE: Environmental Pollution & Public Health

Student Handout

By the end of this module you will be able to answer the question: How does environmental pollution impact human health?

Learning Objectives

By the end of this module students will be able to:

* Calculate a cancer mortality rate
* Describe how environmental pollutants are correlated with cancer mortality rates
* Execute a comparison of means t-test in Google Sheets
* Determine if the means of two data sets are statistically distinguishable

Why This Matters

A healthy community begins with a healthy environment. Other health measures such as diet, exercise, and vaccination can all be undermined by polluted air and water. In conjunction with a strong public health system, an emphasis on environmental quality can reduce disease and increase well-being. Since the beginning of the industrial revolution, certain industries and practices have polluted the environment with carcinogens, endocrine disruptors, airway and skin irritants, and other toxicants. The presence of these pollutants has serious implications for human health, a conclusion supported by analysis of large population data sets, as will be demonstrated in this module.

Outline

Part A: Cancer Mortality Rates in the US

Part B: Louisiana’s Cancer Alley

Part A: Cancer Mortality Rates

The goal of public health practitioners is for every member of the community to live a long and healthy life. To help understand the underlying processes that impede this goal, public health researchers use a variety of statistical metrics to aggregate health data.

One metric that scientists use to understand human health outcomes is **mortality rate**. Since population sizes vary (e.g. Alaska has a much smaller population than California), using total deaths from a specific cause is not a useful way to compare health outcomes. Scientists must normalize the data to account for variations in population size, age, and other demographics to enable comparison. For state-level data, mortality rates are calculated using the equation below, dividing the number of deaths by state population size. This calculation allows deaths to be expressed as deaths per 100,000 people. Find out more about cancer mortality rates at the [National Cancer Institute](https://seer.cancer.gov/statistics/types/mortality.html).

1) Consider the following data. This data set is the number of deaths from heart disease from the states listed and the approximate population of those states in 2019. Calculate the heart disease mortality rate for each of the states listed.

| State | Heart Disease Deaths in 2019 | Population in 2019 | Heart Disease Mortality Rate per 100,000 people |
| --- | --- | --- | --- |
| Alabama | 13,448 | 6,123,000 |  |
| Alaska | 843 | 649,000 |  |
| Arizona | 12,587 | 9,393,000 |  |
| Arkansas | 8,669 | 3,827,000 |  |
| California | 7,762 | 5,669,000 |  |

For the rest of this module we’ll be using cancer mortality rates collected from the [National Center for Health Statistics](https://www.cdc.gov/nchs/pressroom/sosmap/cancer_mortality/cancer.htm), organized by the CDC. You can download the data[**HERE.**](https://docs.google.com/spreadsheets/d/1M-yVvm2cW-Ajx3kzp_ID_KVT9sbda9wM/copy)

2) Open the data in Google Sheets. Look at the headers of the data in the tab titled ‘CancerRates.’ Fill in the table below with what column each of the data are stored in.

| Data | Column Letter |
| --- | --- |
| State |  |
| Year |  |
| Death Rate |  |
| Total Deaths |  |
| Link to Data Source |  |

The goal of this section of the module is to calculate and visualize the average cancer mortality rate per 100,000 for each state across the years 2020, 2019, 2018, 2017, 2016, 2015, and 2014. The sheet also contains data for 2005, but we are going to exclude this set of data since there is such a large time gap between it and when the rest of the data was collected. To calculate the average cancer mortality rate, you will need to reorganize the data.

3) Click on the Insert drop-down menu, and then click on “Pivot table.” Accept the default data range of CancerRates!A1:E401 and Insert to New sheet. Click “Create.” You should have created a new sheet named Pivot Table 1 which you can rename to “CancerRatesPivot.” The Pivot table editor should be on the right hand side of the screen.

Graphical user interface, application, table, Excel

Description automatically generatedGraphical user interface, application, table, Excel

Description automatically generated

4) Set up the Pivot table as shown below by setting options with the Pivot table editor. The left-most column should contain the state names, followed by 6 columns of data, one for each year of data, and a summary column that will contain the average for each state.

* For Rows, add “STATE”
* For Columns, add “YEAR”
* For Values, add “RATE” and under the “Summarize by” drop down menu, select “AVERAGE”
* For Filters, add “YEAR” and remove the check mark next to “2005” (click on it)Graphical user interface, application, table, Excel

  Description automatically generated

5) To make the data easier to interpret, the next step is to plot it as a bar graph. We want to plot just the state names as labels and the averages for the 2014-2020 period, which will be represented by the length of the bars. Copy and paste the Pivot table *STATE* and *Grand Total* columns to the right of the Pivot table. Rename *Grand Total to Average Cancer Mortality*. Highlight the state names and cancer mortality rate data (STATE and Average Cancer Mortality). Then, under the “Insert” menu, select “Chart.”

Graphical user interface, application, table, Excel

Description automatically generated

7) Click the upside-down triangle next to Average and select “Sort Z-> A” to order the mortality rates from highest to lowest. The chart should re-order states along the x-axis.**You may need to widen the plot to see all of the labels on the x-axis. Be sure each column has a state label lining up with it. If it does not, make the graph wider**

Graphical user interface, application, table, Excel

Description automatically generated

9) Double click on the “Average” bar on the chart. Change the color of that single bar so that it stands out.

10) Looking at the plot, which state has the highest cancer mortality rate over the entire time period for which we have data?

Ans:

11) Name 2 other states that have cancer mortality rates higher than the national average.

Ans:

12) Which state has the lowest cancer mortality rate on average?

Ans:

13) Name 2 other states that have cancer mortality rates lower than the national average?

Ans:

Part B: Louisiana’s Cancer Alley

Now we will look at a case study from a single location to understand how environmental pollution impacts human health, using cancer mortality rates as a proxy for human health impacts.

Along the lower portion of the Mississippi River is an 85-mile stretch of land between Baton Rouge and New Orleans that houses almost 150 chemical factories and refineries. Located on the river, for ease of shipping, these factories and refineries are also located near, and in, some of the most populous areas of Louisiana. Residents of this area are exposed to emissions of hazardous toxicants at much higher concentrations than most of the United States population.

14) Look at the bar graph you made of cancer mortality rates in the United States. How does the cancer mortality rate of Louisiana compare to the national average?

Ans:

This area of Louisiana with all the refineries has not only a higher cancer mortality rate than the rest of the U.S., but also one of the highest rates in Louisiana. Due to the hotspots of cancer in this area, it’s earned the nickname “Cancer Alley.” However, just because the bar graph shows that one state has a higher or lower cancer rate than the national average doesn’t mean that the difference is meaningful or real. The spread of data between the highest and lowest points might be so big that the two sets of data mostly overlap, so much so, that they’re indistinguishable even though the means look different. In other words, the difference of means is not *statistically* *significant*.

In statistics, there are methods to determine if two sets of data are significantly different. One such test is the comparison of means t-test. A t-test creates a *p-value*, which is a probability*. If the p-value is below some threshold, usually 0.05, then the difference between the means of the two sets of data is said to be statistically significant.*

Google Sheets has a formula to calculate a t-test which returns a p-value. The formula is =t.test(). Within the parentheses there are four arguments. The first two arguments are the two sets of data to be tested. The first set of data is highlighted, followed by a comma, then the second set of data is highlighted. The next two arguments tell Sheets what type of t-test to execute. For this module, we will be using a two-tailed t-test (put a 2 in the formula), and two sample unequal variances (put a 3 in the formula). The formula should be typed as follows:

=t.test(first set of data, second set of data, 2,3)

15) Somewhere on your Sheet, create a table. Label the table “t-test results.” In the left-hand column, put a label for the t-test you are executing. The right column will have the p-values. The first test you will execute in this table is to compare the cancer mortality rate of Louisiana to the National Average. Arrange that information in your table as shown below:

| **T-Test Results:** |  |
| --- | --- |
| Comparison | p-value |
| Louisiana vs. National Average |  |
| Hawaii vs National Average |  |

16) In the cell next to the “Louisiana vs. National Average” label and below the “p-value” label, type the formula for a t-test, highlighting the data for the years 2014-2020 for Louisiana and then for the national average, 2014-2020. The p-value will appear in the cell. What is the p-value of the t-test? *Sheets may give the answer in scientific notation.*

Ans:

17) Does this p-value indicate that there is a significant difference between the cancer mortality rate of Louisiana and the national average?

Ans:

18) How do you think that the pollution in Louisiana’s Cancer Alley influenced this result?

Ans:

Note that this finding would need to be followed with more research on how the specific pollutants influence cancer for the link to be confirmed.