Statistical Methods

Analysis 1

For my first analysis, I compared the amount of deaths with an underlying cause of lung cancer and the amount of deaths with an underlying cause of death other than lung cancer. To do this, I performed a chi-squared goodness of fit test using the number of lung cancer deaths, the number of non-lung cancer deaths, and probabilities of dying by lung cancer and not dying by lung cancer.

Analysis 2

For my second analysis, I wanted to see what the top five underlying causes of death were. To do this, I created a frequency table of all the causes of death. After I created this frequency table, I sorted it so that the events appeared from the least to most frequent. I then noted which codes were reported to have occurred most frequently and referred to the codebook to identify the causes of death and confirm their status as one of the top five most frequent causes of death as the number of cases for each code is also listed in the codebook.

Analysis 3

For my third analysis, I wanted to see what other multiple causes of death and diseases were comorbid with lung cancer. To do this, I created a subset of data consisting of only the cases in which lung cancer was listed as the underlying cause of death. I then opened the dataset and observed which other causes of death were additionally listed for each individual. Because additional malignant neoplasms (cancer) did not specify type or location, I excluded this from the comorbidities as I was unable to differentiate if the neoplasm was in an additional lung

location or if it was in an additional body location and thus constituting as a different disease. After this, the top three comorbid causes of death were residual disease, chronic lower respiratory disease and diabetes. To test whether these additional causes of death were beyond what is expected, I performed three chi-squared goodness of fit tests. The first chi-squared test examined unspecified residual disease, and there were different probabilities of this co-occurring with lung cancer depending upon the stage of cancer. Because the stage of cancer is not specified in the dataset, I averaged probability across the three stages. The second chi-squared test examined chronic lower respiratory disease and the third chi-squared test examined diabetes.

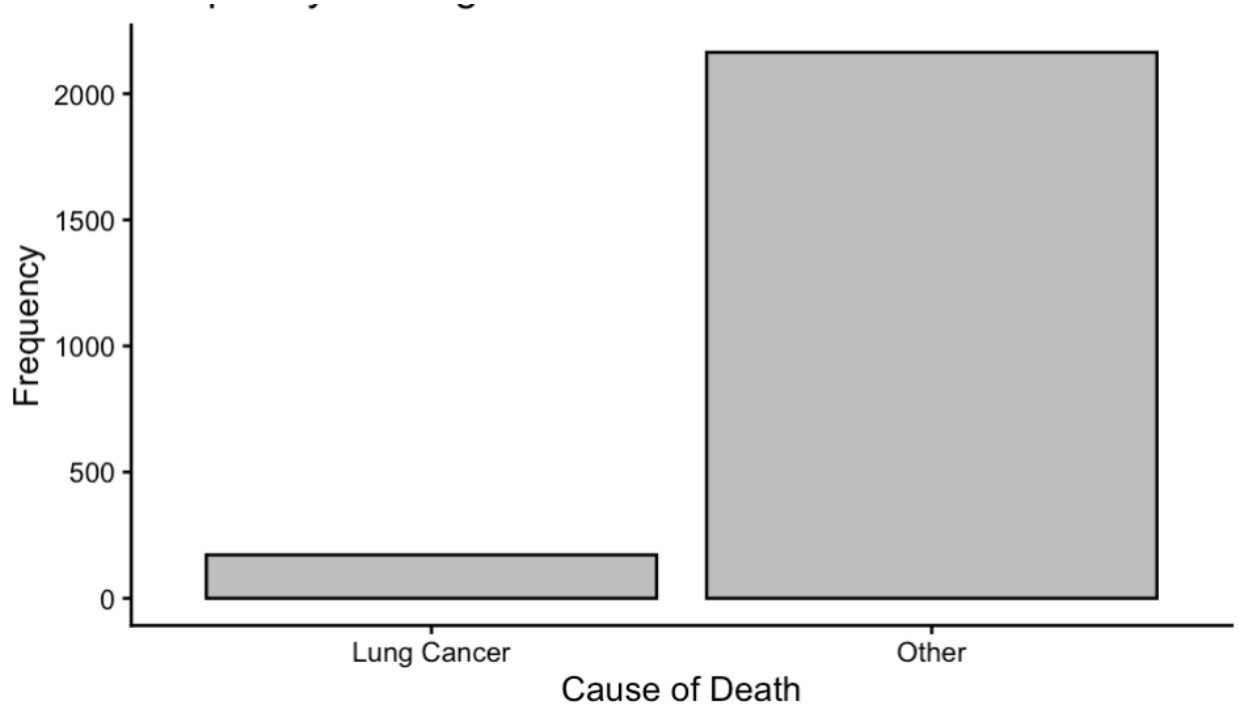
Results

Analysis 1

The results of the chi-squared goodness of fit test was that $p = 9.91 \times 10^{-5}$ ($df = 1$). These results indicate that the observed frequency of deaths from lung cancer do not match the expected frequency. Given that the observed frequency of lung cancer deaths does not meet the expected frequency, it is likely that the observed frequency is higher. While the probability of dying of lung cancer is only .05, you would expect about 5 percent of the population to have an underlying cause of death of lung cancer, however, the number of observed deaths with an underlying cause of lung cancer was 157 out of 2321 people, as observed in Figure 1, amounting to 6.7% of the population identifying lung cancer as the underlying cause of death.

Figure 1

Frequency of lung cancer as the underlying cause of death

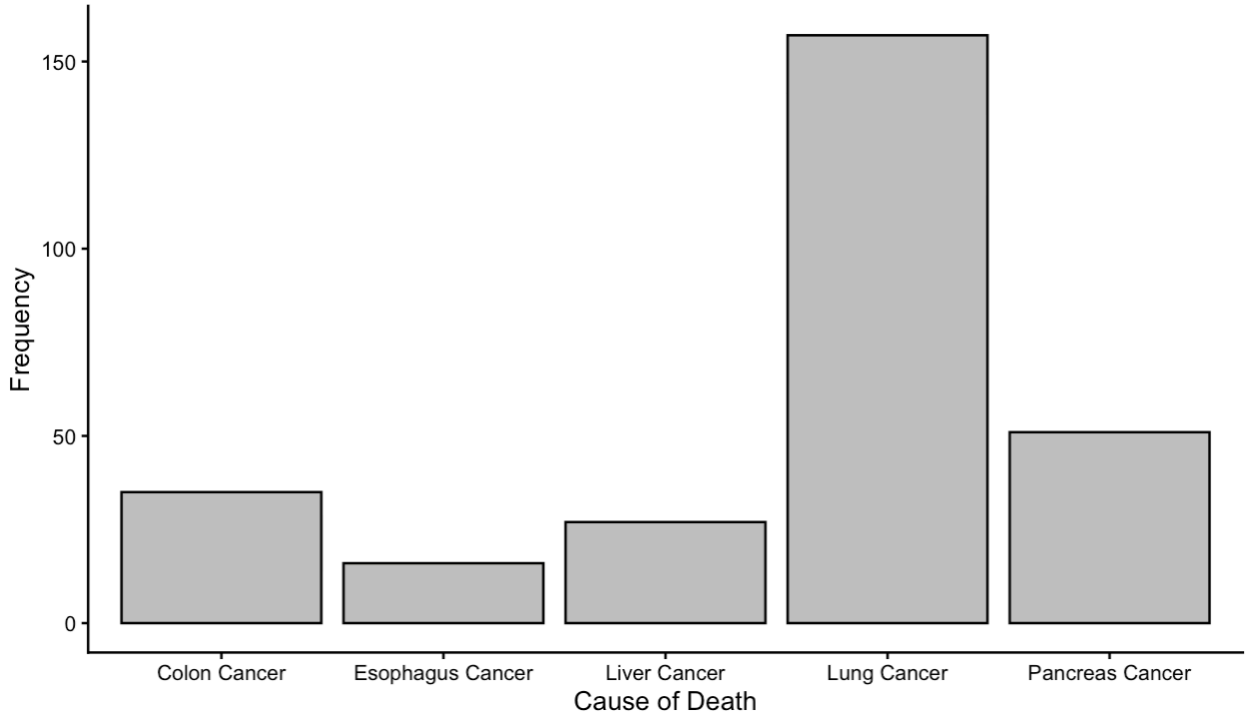


Analysis 2

The results indicated by this sorted frequency table was that the top five most frequent causes of death were Lung Cancer, Pancreas Cancer, Colon Cancer, Liver Cancer and Esophagus cancer. This was confirmed when I referred to the codebook to validate the findings from the frequency table and are displayed in Figure 2.

Figure 2

The five most frequent causes of death



Analysis 3

The results of the first chi-squared, which tested to see if the frequency of cases involving comorbidity with unspecified residual disease and lung cancer, was $p = 2.2 \times 10^{-16}$ ($df = 1$). These results suggest that the observed frequency of cases where unspecified residual disease was comorbid with lung cancer do not happen as expected and likely occurred more frequently in our sample. The results from the second chi-squared test, which tested to see if the frequency of cases involving comorbidity with chronic lower respiratory disease and lung cancer, was $p = 1.307 \times 10^{-5}$ ($df = 1$). These results suggest that the observed frequency of cases where chronic lower respiratory disease was comorbid with lung cancer do not happen as expected and likely occurred more frequently in our sample. The results of the final chi-squared test, which tested to see if the frequency of cases involving comorbidity with diabetes and lung cancer, was $p = .004$ ($df = 1$). These results suggest that the observed frequency of cases where diabetes was comorbid

with lung cancer do not happen as expected and likely occurred more frequently in our sample. The results are summarized in Figure 3.

Figure 3

Causes of death comorbid with lung cancer