Project Part 5

Research Question

The research question that will be answered in this report is:

Has the percentage of severe injuries that have lead to amputations decreased in recent years?

Data

The data that helps to answer this comes from the Bureau of Labor Statistics (BLS, ref. 3) and the United States Department of Labor-Occupational Safety and Healthy administration (OSHA) (ref. 1,2). The two data sources provided data on workplace injuries- each with data on 2019, which is the year I have chosen to focus on. Each has data on specific industry, which will allow me to compare each industry individually. After subsetting the OSHA data to this year, I was able to merge the data based on a NAICS (North American Industry Classification) code (ref.2), which classifies each industry under a code (or summary of injuries, based on the the BLS).

Each row of the data set contains a specific instance of a severe injury report which comes from the OSHA dataset, as well as summary data on that industry from the BLS data. The columns of the dataset include the date the specific severe injury occurred, the employer, the NAICS code, the corresponding industry name, and ID code (correspond the injury to filed paperwork), if a hospitalization or Amputation was necessary, the nature of the injury (amputation, fractures, crushing, etc.), the part of the body that was injured, an event title which explains how the injury occured, the source (and a secondary source if neccessary) of the injury, as well as the industry name and total fatal injuries (a death at the result of an injury, ref.3) in the year 2019. The data is appropriate for answering the given research question because it includes data from 2019, which can be used to compare to years in the past.

Choosing a test

The test I have chosen to use to answer this question is a one-tailed one sample z-test for proportions. According to CNN Money, OSHA reported that in 2015 around 25.45% of severe injuries reported to them lead to an amputation (ref. 4). This can be used as a population proportion to compare to the proportion of severe injuries from a sample of injuries from the year 2019 that lead to amputations (my dataset).

Assumptions for the one-sample z-test for proportions:

- The data are random instances from the population.
- The population matches a binomial distribution.
- When both the mean (np) and the variance(n(1-p)) values are greater than 10, the binomial distribution can be approximated by the normal distribution.

The above assumptions can be identified in our data by taking a random sample of values from my given dataset, which would fulfill the first assumption. To fulfill the second two assumptions, we can make sure that our values for the mean and variance are greater 10 so tahat we can use a normal distribution. For this test, because $p = 0.2545$, we would need to make sure our sample includes at least 40 different instances of injuries in order to be valid. I will be using a sample of 100 to make sure this assumption is valid.

Conducting the test

Test hypotheses:
Null hypothesis: The population proportion is equal to the sample proportion of severe injuries that lead to amputations.
Alternative hypothesis: The population proportion is greater than the sample proportion of severe injuries that lead to amputations.

```
set.seed(07312001)
sample <- sample_n(data_together, 100)
amps <- sum(sample$Amputation, na.rm= T)
amps
```

```
## [1] 21
```

```
z.prop <- prop.test(amps, 100, p=0.2545, alternative="less", correct=FALSE)
z.prop

##
##  1-sample proportions test without continuity correction
##
## data:  amps out of 100, null probability 0.2545
## X-squared = 1.0437, df = 1, p-value = 0.1535
## alternative hypothesis: true p is less than 0.2545
## 95 percent confidence interval:
##  0.0000000 0.2841872
## sample estimates:
##    p
## 0.21
```

Results

From the test, we have a p-value of 0.1535, which means we fail to reject the null hypothesis. This means that there is no statistically significant difference among the proportion of severe injuries that lead to amputations from the population and the sample chosen. In less statistical terms, the test showed no significant difference between severe injuries causing amputations in 2015 and 2019. This makes sense just by looking at the number of amputations within our sample, which is 21.

Conclusions and Generalizing Beyond the Data

From this test, we can see that there is no difference in the amount of severe injuries that lead to amputations from 2015 to 2019. While one test with one sample is not enough to make any sweeping generalization, we can most likely say that there has been no decrease in reported severe injuries that lead to amputations between 2015 and 2019. This could mean that no steps have been taking to help reduce these incidents in recent years, which would lead to the proportions in the above test being the same like we see here.

References

1. https://www.osha.gov/severeinjury/
2. https://www.osha.gov/Establishment-Specific-Injury-and-Illness-Data
3. https://www.bls.gov/iif/oshcfoi1.htm
4. https://money.cnn.com/2016/03/18/news/workplace-injuries/