Accidents Team Workshop Presentation

Anna, Tomoyoshi, Will, and Yuhan

Data

- ➔ The data being used for this research project is from the Occupational Safety and Health Administration (OSHA).
- ➔ OSHA collects data across different firms each year in hopes to measure workplace safety.
- ➔ This data is cross sectional that has more than 20 variables across approximately 340,000 establishments.
- ➔ Our data was sourced from Professor Singleton’s work on workplace safety and violations.

Hypothesis and Theory

- ➔ Larger establishments will experience less injuries per worker than smaller establishments.
- ➔ Larger establishments will be more subject to economies of scale in terms of OSHA inspections and the costs of implementing/improving safety practices, than when compared to smaller establishments, which would nudge them towards practices that encourage worker safety.

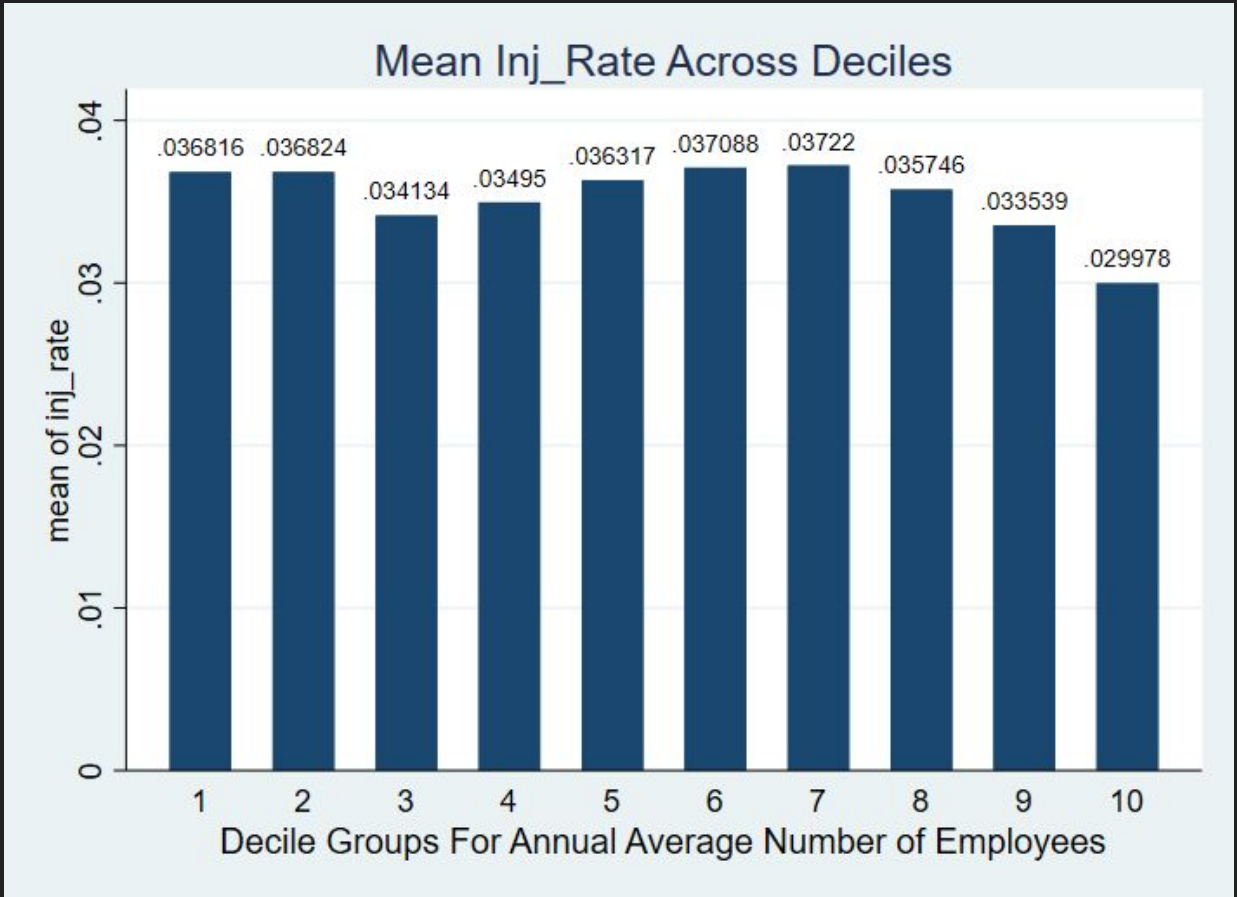
Mechanism

- ➔ After completing the data fix, we generated variables that allowed us to view our key variables at a per capita level. We generated a new variable entitled **`Inj_Rate`** which calculated the quotient of **`total_injuries`** over the **`annual_average_employees`**.

- ➔ We replicated this process across all the different injury types: skin disorders, respiratory issues, and poisonings. We also created a new variable called **`decile`** which separates the variable **`annual_average_employees`** ten groups**.**

- ➔ Professor Singleton used a similar method in his paper “The Effect of Workplace Inspections on Worker Safety”, and due to the origin of our data and guidance, we proceeded with a similar route.

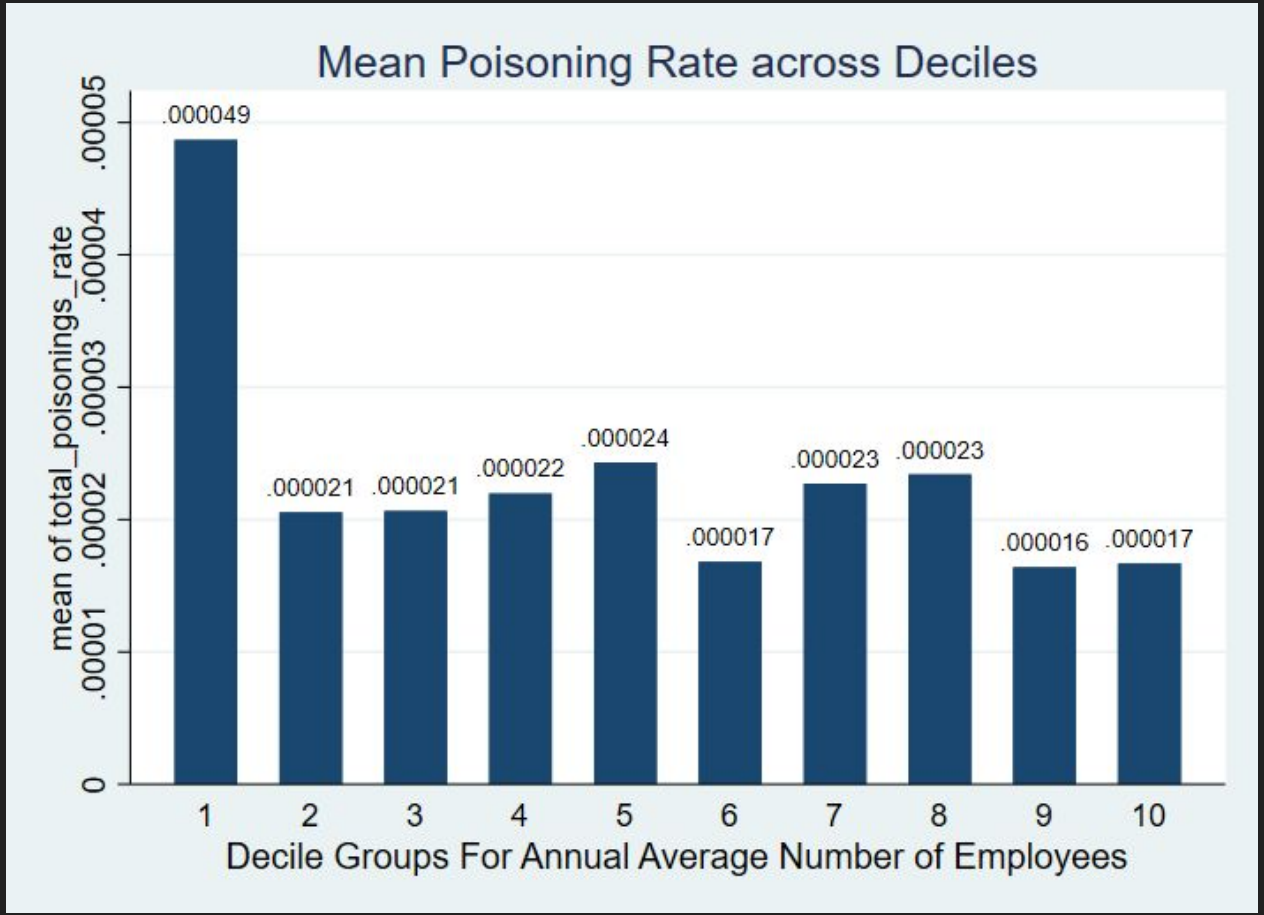
Figure 1:


Mean Inj_Rate Across Deciles
mean of inj_rate
0
.01
.02
.03
.04
.036816
.036824
.034134
.03495
.036317
.037088
.03722
.035746
.033539
.029978
1
2
3
4
5
6
7
8
9
10
Decile Groups For Annual Average Number of Employees

Analysis for Figure 1

- **Correlation Testing:**
  - We found a correlation coefficient of -0.0136 which suggests an extremely weak negative correlation.
  - The small magnitude implies that the relationship is not practically significant. In other words, the variables have little linear association.
- **Summary Statistics:**
  - Mean Injury Rate of .0352, Standard Deviation of .0882, Median of .0171
  - Based on this information a substantial number of the observations is between 0 and 0.3.
- **Outliers:**
  - This variable had substantial outliers, 7.5, 9.37, 10.76, and 28.
  - These data points also imply that a firm with 1 worker is getting injured 7 to 28 times.
  - The median is considerably lower than the mean which could indicate some skewness in the distribution of injury rates, i.e pointing to issues with outliers.

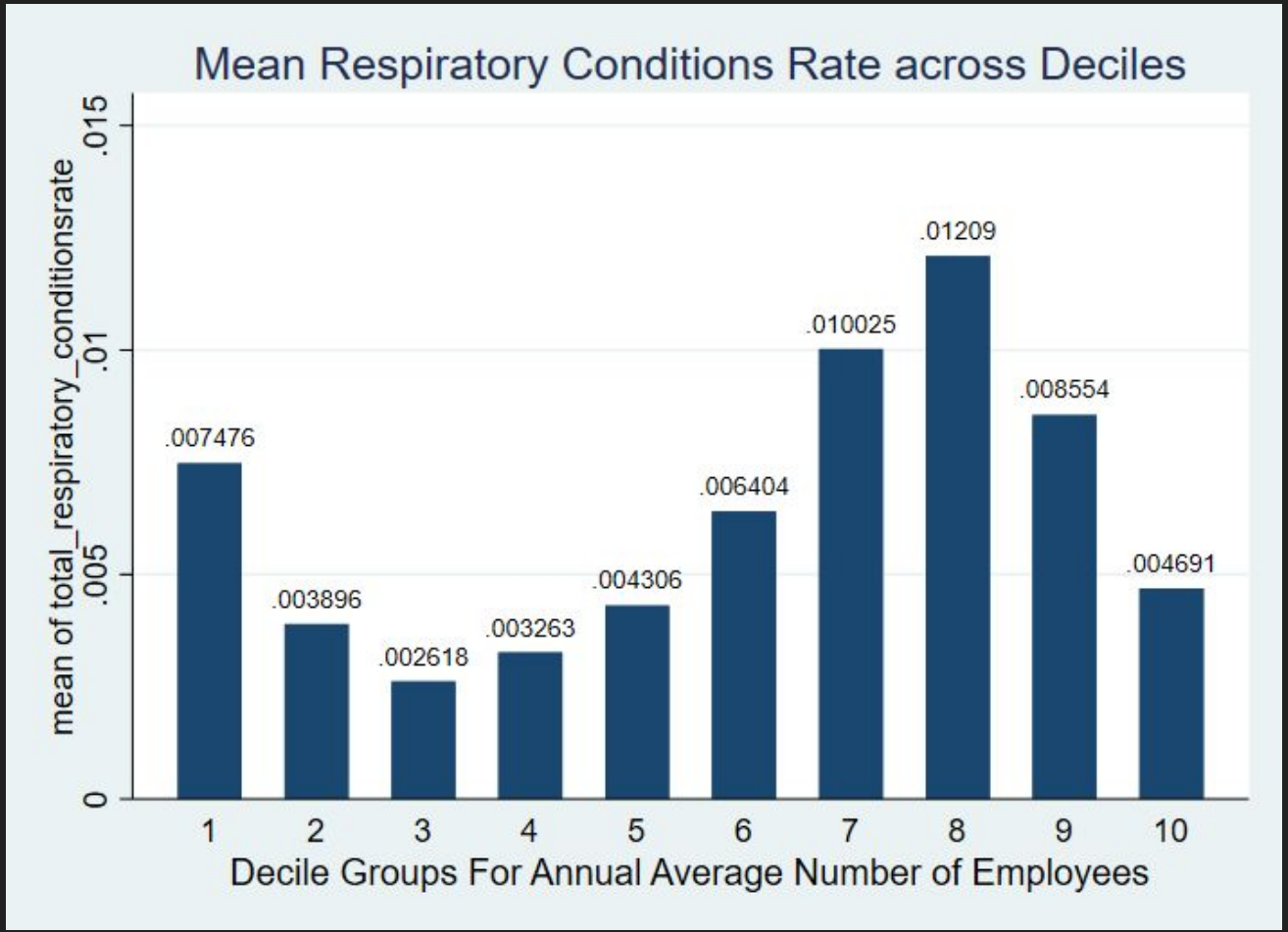
Figure 2:


Mean Poisoning Rate across Deciles
mean of total_poisonings_rate
0
.00001
.00002
.00003
.00004
.00005
.000049
.000021
.000021
.000022
.000024
.000017
.000023
.000023
.000016
.000017
1
2
3
4
5
6
7
8
9
10
Decile Groups For Annual Average Number of Employees

Analysis for Figure 2

- **Correlation Testing:**
  - We found a correlation coefficient of -.0025 which does not support our hypothesis however it is not a strong correlation and could have been caused by outliers in the data.
- **Summary Statistics:**
  - The mean poisoning rate was .0000231, the standard deviation was .0021069, and the median was 0.
  - We then used tabulate and codebook to look through the data for possible outliers causing the smallest decile group to have the highest rate.
- **Outliers:**
  - We found that one observation had a Skin Disorder rate of exactly 1, implying that every person at that firm had a skin disorder.
  - This and the difference between the mean and median explain the issues with outliers in our data and explain the difference between decile 1 and the remaining deciles.

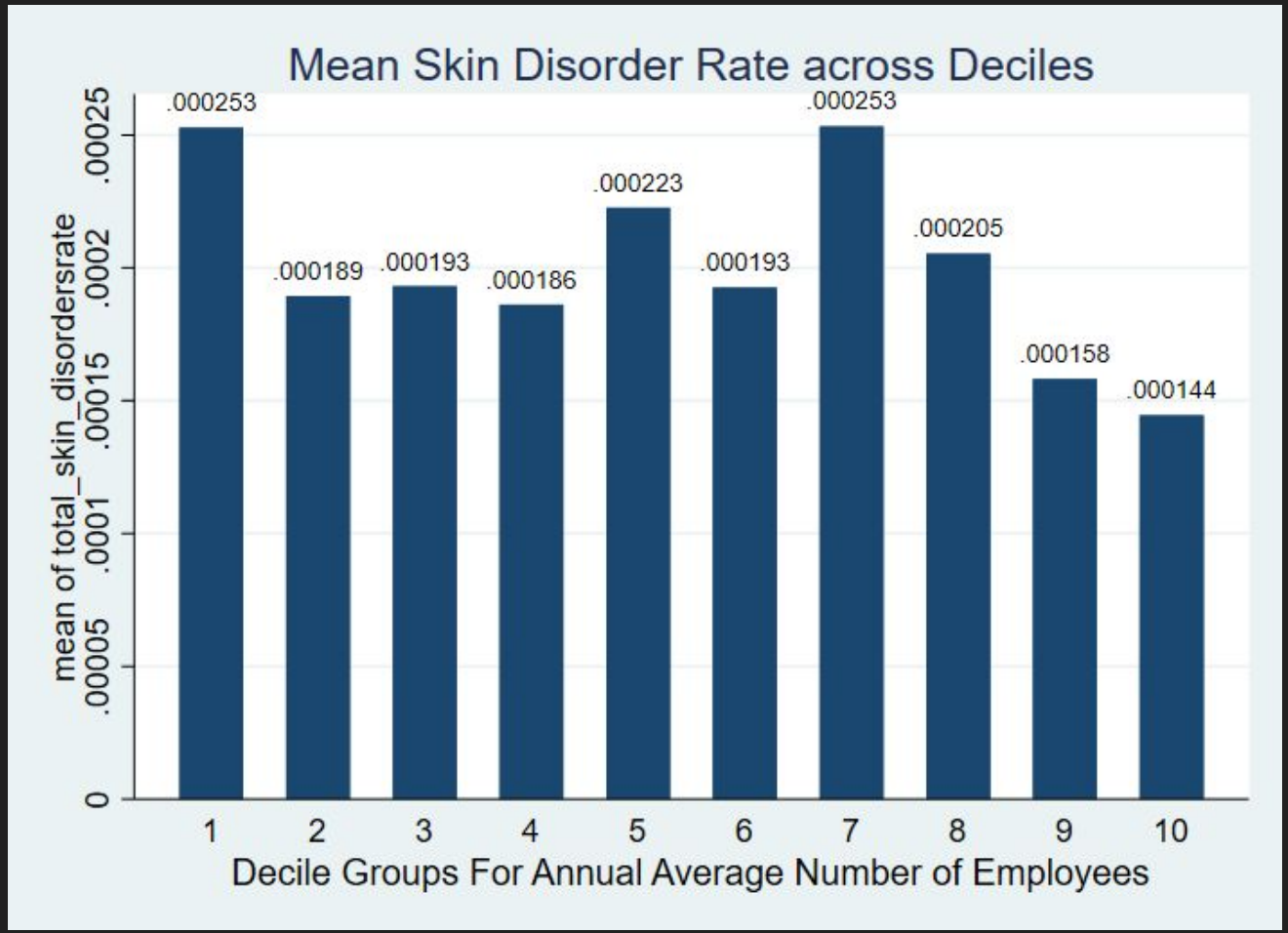
Figure 3:


Mean Respiratory Conditions Rate across Deciles
mean of total_respiratory_conditionsrate
0
.005
.01
.015
.007476
.003896
.002618
.003263
.004306
.006404
.010025
.01209
.008554
.004691
1
2
3
4
5
6
7
8
9
10
Decile Groups For Annual Average Number of Employees

Analysis for Figure 3

- **Correlation Testing:**
  - The bar chart shows a clear negative correlation between respiratory conditions and decile groups.
  - To confirm this, we ran a correlation test and found a correlation coefficient of .0286, indicating a weak positive correlation.
- **Summary Statistics:**
  - The mean for this variable is .00633, the standard deviation is .04745, and the median is 0.
  - We found a distribution with a range from 0 to 4.625, implying that this variable may have more outliers impacting its analysis.
- **Outliers:**
  - The information about the range and the higher standard deviation implies that there are outliers impacting the distribution and with a maximum of 4.625, one data point where each worker is being impacted by respiratory conditions 4 plus times per year.

Figure 4:


Mean Skin Disorder Rate across Deciles
mean of total_skin_disordersrate
0
.00005
.0001
.00015
.0002
.00025
.000253
.000189
.000193
.000186
.000223
.000193
.000253
.000205
.000158
.000144
1
2
3
4
5
6
7
8
9
10
Decile Groups For Annual Average Number of Employees

Analysis for Figure 4

- **Correlation Testing:**
  - It is hard to see if there is a correlation between the two variables, so we ran a correlation test and found a correlation coefficient of .0031.
  - A correlation coefficient of 0.0031 suggests a very weak positive correlation between the two variables and it's important to note that correlation does not imply causation.
- **Summary Statistics:**
  - We found a mean skin disorder rate of .00019, a standard deviation of .005423, a median of 0, and a range of 0 to 1.
  - Compared to the overall injury rate and the other types of injuries within our data this variable has a much smaller distribution and is very uncommon.
- **Outliers:**
  - Outliers were not an issue for this variable.

Conclusion/Discussion

- **Data Collection Considerations:**
  - Recognition of potential issues within the data, such as incorrect reporting or replication of data points.
- **Future Research Direction:**
  - Interest in determining if correlations between firm size and injury rates are unique to certain sectors.
- **Reflections on the Research Process:**
  - Specific focus on gaining more insights into data collection processes and addressing data-related issues for a more comprehensive analysis.
  - New data sources?