



# Major Group Project

By: Madison Phillips

# Table of Contents

- Project overview
- Hypotheses
- Data
- One-Way ANOVA
- Analysis
- Conclusion



# Project Overview

For the final project, my group decided to study homelessness and how it is intertwined with things such as age, income, and substance abuse. For my portion of the project I attempted to answer whether or not there is a correlation between income for the homeless and whether or not they are substance abusers. To answer this, we used the dataset `homeless_pep.csv`. This dataset, which contained 242 participants, was obtained from Kaggle and a link can be found below. It is important to note that this dataset solely contains information on homeless veterans, but it is still possible to make inferences on the homeless population as whole from this data.

Data source: [https://www.kaggle.com/code/ikarhumi/homeless-dataset-focus-on-veteran/data?select=homeless\\_prep.csv](https://www.kaggle.com/code/ikarhumi/homeless-dataset-focus-on-veteran/data?select=homeless_prep.csv)



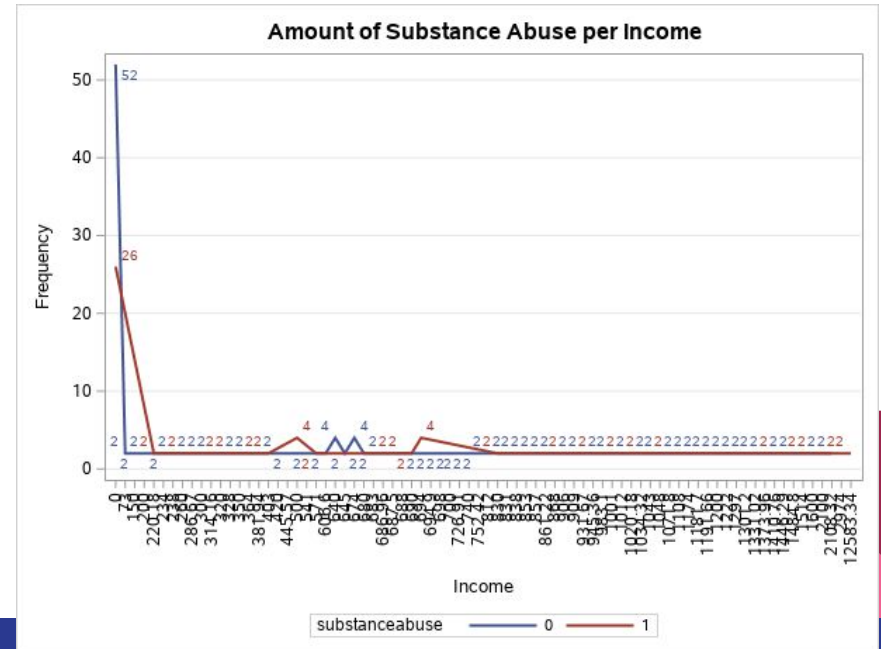
# Hypotheses

- With a research question of 'How does income affect the likelihood of substance abuse?' we can identify two hypotheses for our research.
- Null Hypothesis ( $H_0$ ): Income does not have an affect on the likelihood of substance abuse.
- Alternative Hypothesis ( $H_1$ ): Income does have an affect on the likelihood of substance abuse.



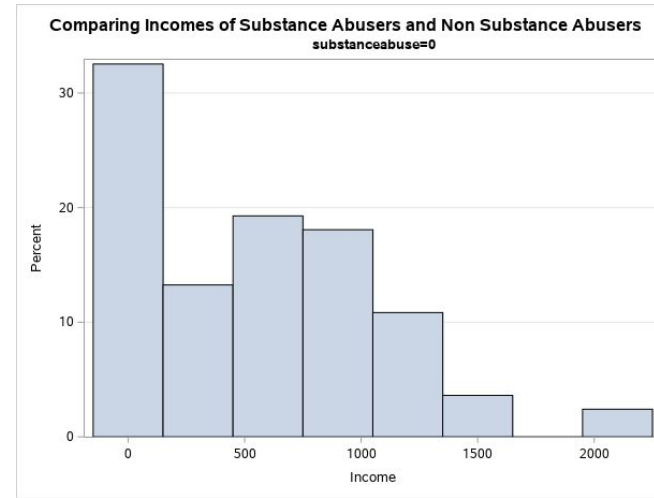
# Data

- A line graph was created to compare the data. The blue line represent those in the shelter who do are not substance abusers and the red line represents the substance abusers within this shelter. It can be seen that there was a frequency of 26 substance abusers with an income of \$0, as well as 52 non-substance abusers with an income of \$0.
- It can also be noted that there were peaks on the graph for substance abusers at incomes of \$500 and \$694. Other peaks were for non substance abusers and they were located at incomes of \$640 and \$675.

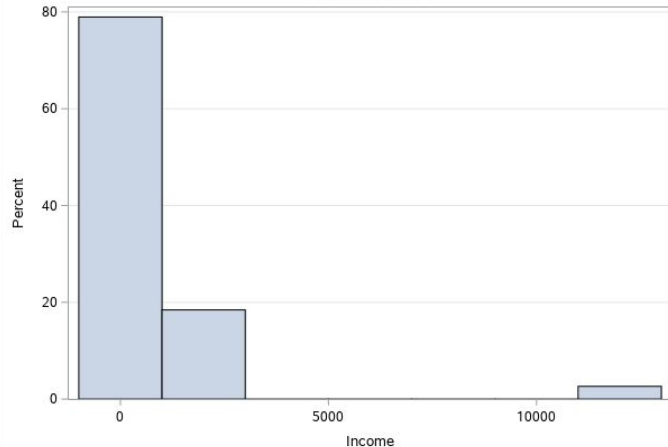


# Data Continued

- To get an additional look at the data and how it compared, I also created separate histograms in SAS.
- The histogram that is located on the right of the slide shows the amount of non substance abusers in the shelter and their respective incomes



Comparing Incomes of Substance Abusers and Non Substance Abusers  
substanceabuse=1



- The histogram to the left shows the amount of substance abusers in the shelter as well as their incomes.
- These histograms show that most substance abusers were found in the lowest income category (\$0)
- There were a few substance abusers not in the lowest income category. Roughly 19 substance abusers had an income above \$0 but below \$5000 and around 1 substance abuser with an income over \$10,000

# One-Way ANOVA

- A One-Way ANOVA test was conducted through SAS. This test resulted in an F-Value of 3.06 which concludes that our data was statistically significant since a critical score of 0.05 was used.
- The One-Way ANOVA also resulted in a P-Value of 0.0816. Our P-Value measures probability and lets us know whether or not any statistical differences are due to chance or if the differences are significant.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4525967.9	4525967.9	3.06	0.0816
Error	240	355222088.6	1480092.0		
Corrected Total	241	359748056.4			

# Analysis

- The One-Way ANOVA provided us with a F-Value of 3.06 and a P-Value of 0.0816.
- With a F-Value that is higher than the P-Value, there is sufficient evidence to reject the Null Hypothesis that income does not have an impact on the likelihood of substance abuse.





# Conclusion

Since the Null Hypothesis was rejected, we can infer that there likely is a connection between income and substance abuse for those who are homeless. This data provides evidence for homeless shelters that rehabilitative measures need to be implemented within the shelters to try and combat the amount of substance abuse that is occurring in the homeless community. These rehabilitative measures can include things such as drug counseling or even providing clean needles to those coming to the shelters when drug counseling is not available.

All graphs and tests were run on SAS. SAS results were collected and put into individual folders on GitHub.

GitHub link: [https://github.com/aronchen1/Stats216\\_Project/tree/main/Madis%20Folder](https://github.com/aronchen1/Stats216_Project/tree/main/Madis%20Folder)

