Youth, Unemployment and Crime in US

Huda Feroz, Syed Bilal Hoda, Daniyal Aziz

Math 107 - Lie Detector: Introduction to the Practice of Statistics.

17th April 2020

**Research question**

At the March, 1995, International Meeting on Population and Social Development in Copenhagen, during the session on unemployment, underemployment, and population it was discussed that one cause of unemployment is that there is a difference in supply of employment opportunities and availability or supply of labour [1]. During the meeting, the age structure of the population was discussed as a demographic variable [1]. It is because the rising youth population is accompanied by a rise in the labor force of the community which is said to lead the substantial chunk of population to unemployment. Since economic growth of a region is directly dependent upon unemployment [2], it is important to investigate the relationships which impact youth unemployment rate, such as youth population. The youth unemployment rate is the number of unemployed 15-24-year-olds expressed as a percentage of the youth labour force. Unemployed people are those who report that they are without work, that they are available for work and that they have taken active steps to find work in the last four weeks [3].

The existence of relationship between crime and youth unemployment is an age-old question in the social science literature. There have been many studies of the relationships from a range of perspectives, but the research effort has failed to produce clear evidence of having a link. In order to establish a clear link, it is crucial to investigate that is this relationship present in all states of the United States.

At the times when unemployment rate of United States are worrisome or criminal activities in the region are increasing, people increasingly demand for policies that are focused on increasing count of youth in labor market. People demand for more programs and policies which can provide basic skills and access to secondary and tertiary education, better information, access to credit, and creating job roles which are more suitable for the younger population. Similarly, in order to stabilize economy and youth unemployment rates, there is a demand to introduce stricter population control policies and more population control programs. It is why, we need to investigate relationship between population of youth and youth unemployment, and relationship between youth unemployment in each state and crime activities [4].

The purpose of this research is to study if there is a relationship between crime and youth unemployment in all states of the United States, and if there is a relationship between youth population and unemployment, by using statistical inferences.

## Hypotheses

The continued high unemployment, specifically youth unemployment in many countries, and particularly in the United States, has given rise to a discussion of the reasons for this persistence of the unemployment. One reason people propose for the persistence of unemployment is that there is a youth population bulge in United Stated. The analysis is based on the null hypothesis that there is no relationship between proportion of youth in the population and proportion of youth unemployment across the US states. However, our alternative hypothesis suggests that there exists a relationship between proportion of youth in the population and proportion of youth unemployment across the US states.

Moreover, due to increase in crime rates in the United Sates, there is a growing opinion that criminal offences are linked to high youth unemployment. Our analysis is based on the null hypothesis that the average crime rate for all groups of proportions of unemployed youth is same across all states in the United States. However, our alternative hypothesis suggests that the average crime rate for all groups of proportions of unemployed youth is not the same across all states in the United States.

## Statistical methods

Our data consists of 48 observations and 3 variables:
1. Youth: categorical variable which states the proportion of the youth in the population is similar to national average (average), lower than the national average (low) or higher than the national average (high)
2. Youth unemployment: categorical variable which states the proportion of unemployed youth is similar to the national average (average), lower than the national average (low) or higher than the national average (high).

3.  Crime rate: continuous variable which states the number of offences per million people.

To summarize the data, we have used mean and median as measures of center for our crime rate as the variable is continuous and have used standard deviation and range as measures of spread of data. However, mean is the most suitable choice for center of data and standard deviation for spread of data as the distribution of data is reasonably symmetric and there are no outliers in our crime rate data. Since youth and youth unemployment are categorical variables, mode is the best choice of measure of center. We have also calculated the 95% confidence interval for the mean. For the confidence interval of the mean we calculated the standard error for the sample and we multiplied it with 1.96. The product was added to the mean to get the upper confidence interval and the product was subtracted from the mean to get the lower confidence interval. We multiplied by 1.96 because in a 95% confidence interval, 95% of the area of a normal distribution is within 1.96 standard deviations of the mean.

In order to understand summary statistics for different groups, we use barplots for youth and youth unemployment [figure.1]. To visualize the distribution of crime rate and to see if our distribution is normal, we used histograms [figure.2]. In order to confirm normality of the data, we performed Shapiro-Wilk Test for Normality. To visualize the spread of data, and to check for presence of outliers, we used boxplots [figure.3].
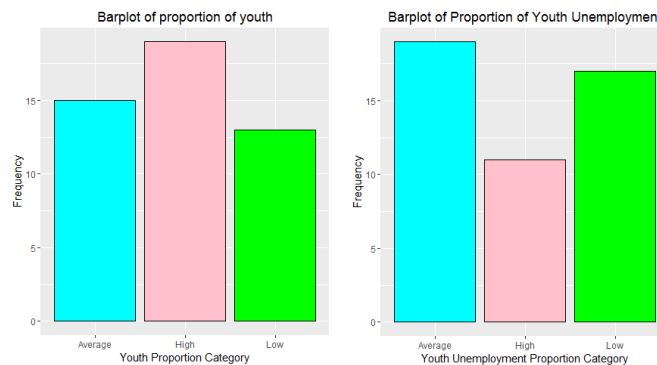


*Figure 1 Barplots for: i) proportions of youth ii) proportions of youth unemployment*
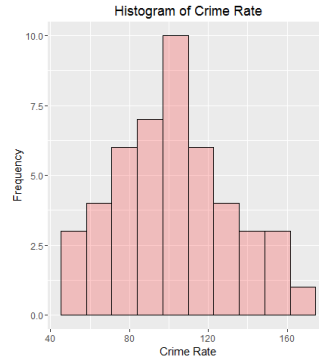
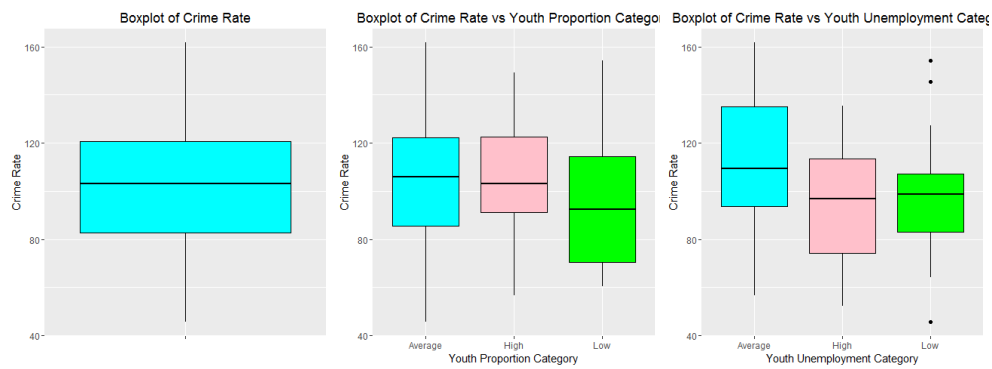*Figure 2 Histogram/distribution of crime rate*



*Figure 3 boxplot of: i) crime rate, ii) crime rate vs youth proportion, iii) crime rate vs youth unemployment*
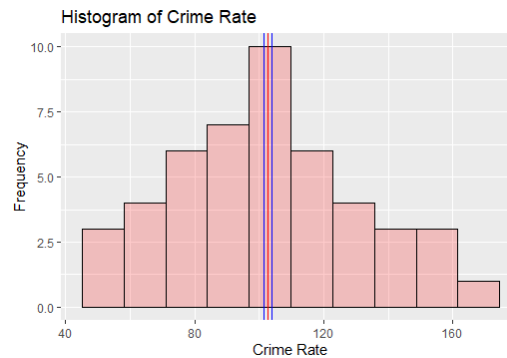


*Figure 4 histogram of: crime rate with mean and 95% confidence interval of mean*

Our alpha-value (significance-level) is 0.05, this is because we have kept our confidence interval as 95% [5]. To check if our null hypothesis if true, that there is no relationship between proportion of youth in the population and proportion of youth unemployment across the US states, we perform the chi-squared test on our categorical variables for youth and youth employment. However, if a chi squared test is conducted on a sample with a smaller size, then

the chi squared test yields an inaccurate inference. An adequate expected cell counts is required for chi-squared test. A common rule is 5 or more in all cells of a 2-by-2 table, and 5 or more in 80% of cells in larger tables, but no cells with zero expected count. Therefore, to perform chi-squared test, we allow the computation of p-values by Monte Carlo simulation. Since our assumptions for chi-squared are not completely met, we also perform Fisher's exact test.



*Figure 5 Table visualization of youth vs youth unemployment proportion*

To check if our null hypothesis is true, that the average crime rate for all groups of proportions of unemployed youth is same across all states in the United States, we use one-way analysis of variance test. We used ANOVA because we have a categorical independent variable, youth unemployment, and a normally distributed interval dependent variable, crime rate, and we needed to test for differences in the means of the dependent variable broken down by the levels of the independent variable [5]. ANOVA requires that the variance of data in groups should be the same; to check for it we used Bartlett test. ANOVA model assumes independence of observations. Since, each of the observations are from different states therefore they are assumed to be independent.

**Results**

The statistical analysis shows that the measures of center for our data is as under:
1. Crime rate: mean = 102.8, median = 103.0
2. Youth: mode = high
3. Youth unemployment: mode = average

The statistical analysis shows that the measures of spread for crime rate is that standard deviation is 28.89327 and range is (45.5, - 161.8). The 95% confidence intervals of the mean are:

Lower Interval ~ 101.6036

Upper Interval ~ 104.0134

Figure.1(i), bar plots for proportions of youth, shows that out of 48 states, 15 states have average youth proportions, 19 have high youth proportions, and 13 have low youth proportions. Figure 1(ii), bar plots for proportions of youth unemployment, shows that out of 48 states, 19 states have average youth unemployment proportions, 11 have high youth unemployment proportions, and 17 have low youth unemployment proportions. Figure.2, histogram of crime rate, shows that the crime rate is normally distributed with a very slight skew on the right. Shapiro-Wilk Test for Normality gives p-value as 0.8508, confirming that crime rate data follows a normal distribution. Figure.3(i), boxplot of crime rate, shows that there are no outliers. Figure.3(ii), boxplot of crime rate vs youth proportion, shows that the spread of crime rate compared to the proportion of youth in the states is more or less the same, where the mean of crime rates in states with low average of youths is slightly lower than rest of the areas, but the overall crime rate has more spread in this category. Figure.3(iii), boxplot of crime rate vs youth unemployment, shows that the mean of crime rates for states with average proportion of youth unemployment is higher than the rest. The figure shows three outliers, where two outliers are at the upper end and one outlier is at the lower end of the plot. The mean crime rate for states with high proportion of youth unemployment and the mean crime rate for states with low proportion of youth unemployment is same.

The applied chi-squared test on youth proportion and youth unemployment proportion yields 0.951 as p-value. Fisher's exact test on youth proportion and youth unemployment proportion yields 0.9467 as p-value. For chi-squared test and for the Fisher test as well, since our p-value is greater than our significance level (0.05), we do not reject the null hypothesis.

Bartlett test on crime rate and youth unemployment proportion yields 0.9491 as p-value. The null hypothesis of Bartlett test is that all population variances are equal. The alternative hypothesis is that at least two are different. Since the p-value is 0.9491, hence we do not reject the null hypothesis and we say variances are equal. Moreover, Figure.3(ii), boxplot of crime rate vs youth

proportion, shows that the overall spread is more or less the same. ANOVA test yields 0.139 as our p-value so we do not reject the null hypothesis at a significance level of 0.05.

**Discussion**

Evidence on the link between youth population and youth unemployment show that our null hypothesis is true i.e. we can not assume a relationship between proportion of youth in the population and proportion of youth unemployment across the US states, which means youth bulge does not contribute to youth unemployment. Evidence on the link between crime rate and youth unemployment proportion show that our null hypothesis is true i.e. there is no difference in the average crime rate for all groups of proportions of unemployed youth in each state. This means that there is no relationship between the proportion of youth unemployed in each state and the crime rate.

Our analysis therefore shows that, programs and polices focusing on youth development and shaping economy as per demands of youth, may not improve youth unemployment rates or decrease criminal activity. Moreover, we can say that decreasing youth bulge (or population) can not improve youth employment rates.

There are several reasons to suspect that the available data and evidence does not correctly establish the relationship between proportion of youth and unemployment, and unemployment and crime rate. It is because, the data and evidence include several biases which were not accounted for or were difficult to account for and they may have affected the results.

1. In our analysis, the provided data only takes proportion of youth and proportion of youth unemployment into account. Whereas, many important demographic and economic variables have not been accounted for, and this may bias the relationships we seek to measure. Some of these omitted factors (variables) are:
   - Alcohol and drug consumption: alcohol consumption is procyclic to crime rate and tends to have independent effects on criminal behavior [6]. Similar patterns may exist for drug use and gun availability [6].

- Economic growth: GDP is an important contributing factor to unemployment. Okun's law says that "GDP must grow at about a 4% rate for one year to achieve a one percentage point reduction in the rate of unemployment" [7]. Moreover, since GDP per capita can be used as a good proxy for personal wealth, it may have an impact on criminal activities.

- Economic recessions: when economic growth slows, companies generate less revenue so they lay off workers to cut costs, which generates higher levels of unemployment. Moreover, declining incomes during recessions reduce purchases theft-worthy items, thus providing fewer targets for criminal activity.

- Education and skill level: greater education levels and having more technical skills provide opportunities for greater number of jobs, so lower risk of unemployment [6]. It also means that the jobs are comparatively high paying, which may count to economic prosperity and lower crime rates.

- Law enforcement: criminal activities are controlled by justice system and implementation of the law. Therefore, the number police officers per 100,000 inhabitants, and training and education of these police offices impact criminal activities in the area.

Some other omitted variables may be gender, health conditions of young population in the area, childhood upbringing of population, presence of labor unions, and global competition.

2. We only check for presence of relationship between youth and youth unemployment proportions, and crime rate and youth unemployment but do not investigate on the direction of impact. For example, higher crime rates in an area may itself impede employment growth and contribute to regional unemployment levels.

3. Aggregation bias: we estimate the relationships without looking into types of crimes or areas of crime. Whereas, we could investigate for a clear relationship if we investigate the relationship between crime rate and youth unemployment by considering crime level

margins. Similarly, we did not look into unemployment as per their types, for example full-time jobs, part-times jobs, summer employments.

4. United States has 52 states and 48 have been included in the sample. The remaining states may have value which can change the current results. So not including all states may lead to weak outcomes.

**Bibliography:**

1. US National Library of Medicine National Institutes of Health, march 4, 1995, author:N/A , Asian Forum NewsUS National Library of Medicine National Institutes of Health, march 4, 1995, author:N/A , Asian Forum News

2. Levine, L. (2011). Economic growth and the unemployment rate. Congressional Research Service, 7–75700, Washington, D.C.Levine, L. (2011). Economic growth and the unemployment rate. Congressional Research Service, 7–75700, Washington, D.C.

3. OECD (2020), Youth unemployment rate (indicator). doi: 10.1787/c3634df7-en (Accessed on 17 April 2020)

4. Lam, D. (2020). Youth bulges and youth unemployment. Retrieved 10 April 2020

5. Institute for Digital Research and Education Statistical Counseling

6. Ruhm, C. J. (1995). Economic conditions and alcohol problems. Journal of health economics, 14(5), 583-603.

7. Alamro, H., & Al-dalaien, Q. F. (2014). Modeling the Relationship between GDP and Unemployment for Okun's Law Specific to Jordan. Available at SSRN 2440674.