Educational Attainment and Crime in the U.S Population

Kyle Cheng

Is there a relationship between educational attainment and violent crime in the U.S population?

The above question will be explored through statistical testing on the crime and education dataset. The contents of this 20x28 dataset go into great detail. It was created by binding together two separate government data tables that were then modified to improve readability and relevance. The original tables were titled and located as listed below:

- Table 1. "*Crime in the United States, by Volume and Rate per 100,000 Inhabitants, 1997–2016*" from the https://ucr.fbi.gov [1] website

- Table 2. "*Years of School Completed by People 25 years and Over, by Age and Sex: Selected years 1940 to 2019*" from the https://census.gov [2] website

For each year between 1997 and 2016, volumes and rates for 9 categories of crime, along with 8 classifications of educational attainment are described. This level of detail, along with the ample sample size and careful data collection methodology, allows for robust comparison of education and violent crime rates. We will define educational attainment as the percentage of the above 25 American population having some form of college education. All variables are numerical and continuous.

Hypothesis testing on the statistical significance of Pearson's correlation coefficient

In order to determine the existence of a relationship between educational attainment and crime rates, we will be conducting hypothesis testing on the significance of Pearson's correlation coefficient. Our null hypothesis will be that the correlation coefficient is not significantly

different from 0. The alternate hypothesis will state the opposite, that the correlation co-efficient is significantly different from zero. The level of significance (alpha) will be set at .05.

In order to correctly test the statistical significance of the correlation coefficient, several assumptions must hold true [3]

1. Level of measurement: For a Pearson correlation, both variables should be continuous
   - Crime and college education rates are continuous variables on the 0.0-1.0 scale

2. Related pairs: Each observation should have a pair of values
   - Both variables are tied together through year

3. Absence of outliers: An outlier is "defined as a value that is 3.29 standard deviations" away from the mean. [3]
   - All values are within this definition

```
summary(crime$Violent.rate)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   361.6   387.6   466.1   456.3   496.9   611.0
```

```
sd(crime$Violent.rate)*3.29
```

```
## [1] 225.78
```

```
summary(college)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4835  0.5197  0.5390  0.5430  0.5696  0.6026
```

```
sd(college)*3.29
```

```
## [1] 0.1124519
```

Generalization

The results of our statistical significance test on the Pearson correlation coefficient will generalize well to the population our dataset samples: True violent crime rates and educational attainment in the American population 25 years and older. The data was collected for government purposes, with specific measures to produce a sample representative of the overall U.S. population.

Testing

```
cor.test(college, crime$Violent.rate, method = "pearson")$estimate
```

```
##        cor
## -0.9611239
```

```
cor.test(college, crime$Violent.rate, method = "pearson")$p.value
```

```
## [1] 1.672664e-11
```

Conclusion

Our Pearson correlation coefficient has a value of -.9611, with a p-value of $1.672*10^{-11}$. As our p-value is less than the significance level of .05, we have sufficient evidence to reject the null hypothesis. Our correlation coefficient is significantly different from zero.
We conclude that from 1997-2016, there is a negative linear relationship between the sampled rates of violent crime per 100,000 Americans and college education amongst Americans older than 25. This can be generalized to the population of all violent crime and educational attainment in the U.S during this period.

References

[1]: https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/tables/table-1/table-1.xls#overview

[2]: https://www.census.gov/data/tables/time-series/demo/educational-attainment/cps-historical-time-series.html

3: https://www.researchgate.net/post/What-is-the-distribution-assumption-for-Pearson-correlation-coefficient#:~:text=The%20assumptions%20for%20Pearson%20correlation, each%20variable%20should%20be%20continuous