

Statistical Analysis Report

Minh Ngo, Sri Sai Sushan Gumadavalli, Will Luo

GitHub Repository Link

https://github.com/UVA-STAT3080/project-team_msw

Introduction and Research Question

Recidivism is the tendency of a convicted criminal to reoffend. Using 2008–2020 New York State data, Part I showed that younger and male offenders have higher recidivism rates. In this analysis, we ask: Is there a significant difference in recidivism rates between male and female offenders in New York State since 1980?

$$\begin{array}{l} H_0 : p_m = p_f \\ H_a : p_m \neq p_f \end{array}$$

Methods

A possible method to answer this question is to use a Two-Proportion Z-Test. This is an appropriate test because we are comparing two independent proportions. In order to use this test, we must make sure all of its conditions are satisfied.

1. Independence

The male and female groups are independent from each other as the recidivism of a male does not meaningfully affect the recidivism of a female. Each of the observations within the male and female groups are also independent from each other, as one person's recidivism does not significantly affect another's.

2. Representative Sample

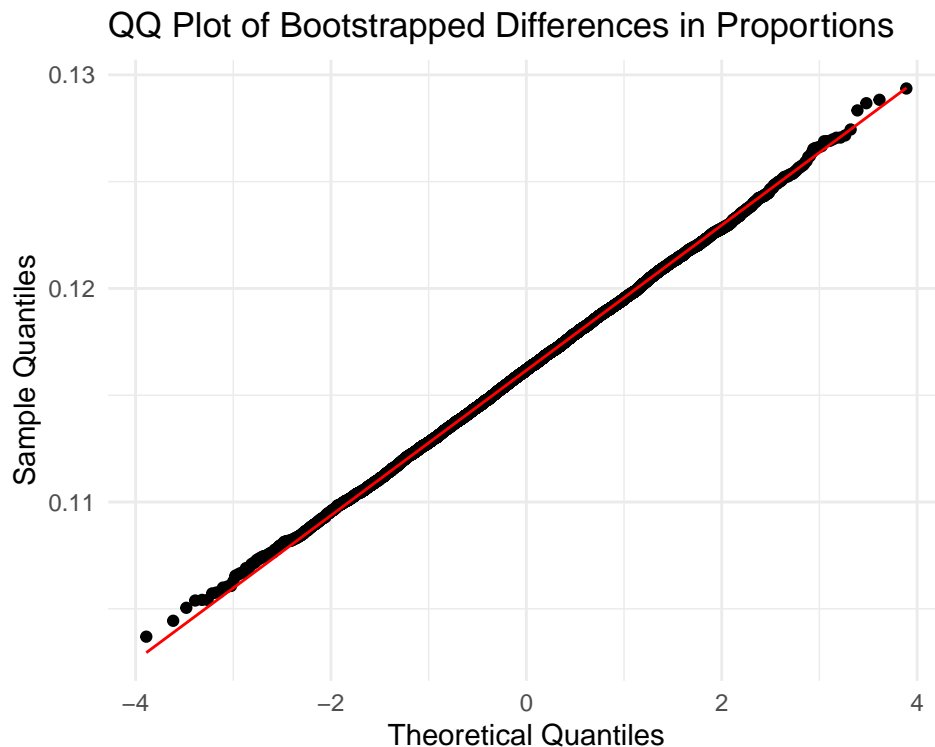
Our data set contains every single releasee in a New York State prison from 2008–2025

and whether or not they returned within three years. Although this is technically a convenience sample, it is still representative of the population of interest—namely, all releasees in New York State since 1960—because the population of New York did not change much since 1980 and our sample covers many years of releases in the time period between 1980 and now. Additionally, most of the population that was alive in 1980 is also still alive today, so the data set actually contains some extrapolated data points from this time.

3. Normality

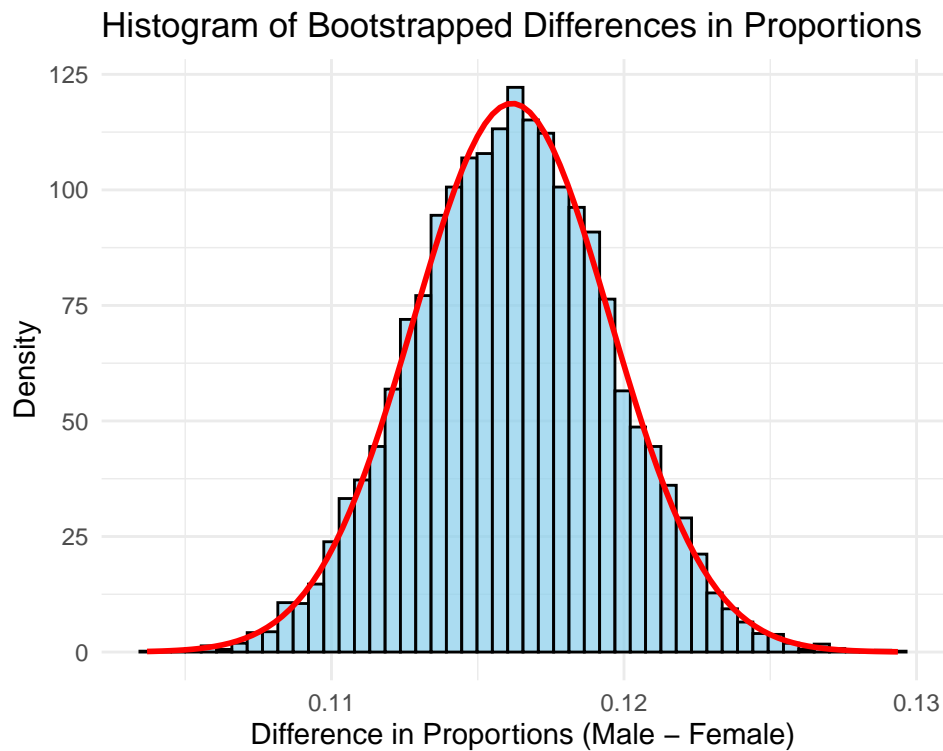
We can use a normal quantile-quantile plot to see if the distribution of the differences in proportions are normally distributed. We will do this by making 10,000 bootstrap samples of the two groups and calculating their difference proportions. These bootstrap samples were drawn from two theoretical binomial distributions with parameters matching each of the two groups, because this is mathematically the same as drawing from the actual data and less computationally intensive. With an ordered list of these values, we can calculate the quantile value for each of them and compare them to what their quantile value would have been, given that they were normally distributed. The QQ-plot generated is shown below.

Normal Quantile-Quantile Plot



As we can see, the quantile values of the bootstrapped samples closely follow the line of normality. This gives strong evidence that the true distribution differences is approximately normal. We can also visually see the distribution using a histogram, as shown below.

Histogram



The distribution of the data seems to very closely follow the normal distribution curve (shown in red). From these two graphics, we can be confident that the distribution of the differences in proportions are approximately normal. With all the conditions of the Two-Proportion Z-Test satisfied, we can calculate the Z-statistic and p-value for our data.

Results

Two-Proportion Z-Test

The extremely small p-value providing strong evidence against the null hypothesis that recidivism rates are equal between genders, indicating the test is statistically significant. The observed difference of around 11.6 percentage points presents that males are more likely to re-offend than females.

Table 1: Two-Proportion Z-Test Results

Statistic	Value
Male Proportion	0.38967
Female Proportion	0.27355
Difference (p1 - p2)	0.11612
Z-Statistic	31.62078
p-value	0.00000

Conclusions

Our results indicate that there is a significant difference in recidivism rates between male and female offenders in New York State since 1980. To be specific, males have a rate of ~39% while females have a rate of ~27%, which means gender can be a determining factor in predicting re-offense probability, as males are more likely to recidivize.

However, we cannot necessarily say that gender directly influences recidivism rates because our analysis does not include potential confounding variables such as age, race, offense type, and other relevant factors. We also assume the recidivism patterns have remained stable over time since 1980, but they could be dynamic instead.

These findings are most suitable to generalize for New York State prison releases since 2008. We should take precautions when applying these results to releases from other states or from different time periods, since criminal behavior and justice system practices vary across time and location. However, the significant gender disparity observed aligns with broader national trends (Spjeldnes & Goodkind, 2009).

References

Portions of the code, data visualization formatting, and written text in this report were developed with the assistance of OpenAI's ChatGPT (Version 5).

New York State Department of Corrections and Community Supervision. (2025, March 7). Recidivism: Beginning 2008 [Data set]. State of New York: data.ny.gov. <https://catalog.data.gov/dataset/recidivism-beginning-2008>

OpenAI. (2025). ChatGPT (Version 5) [Large language model]. OpenAI. <https://chat.openai.com/>

Spjeldnes, S., & Goodkind, S. (2009). Gender differences and offender reentry: A review of the literature. *Journal of Offender Rehabilitation*, 48(4), 314–335. <https://doi.org/10.1080/10509670902850812>