Comparing Japanese Interregional Populations

Greg Pollock

9/26/2020

Introduction

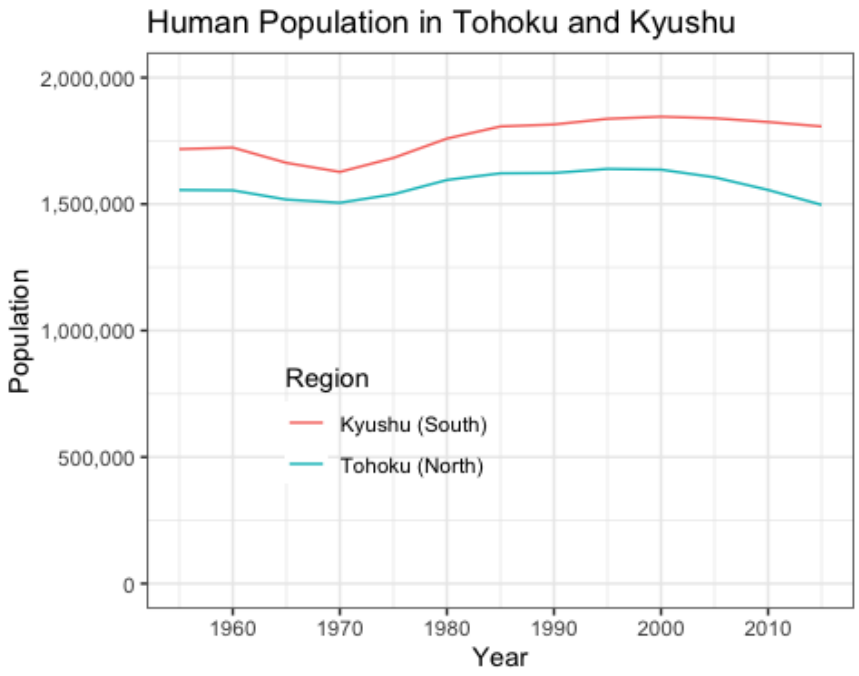
The human population in Japan is a subject of interest to many sociologists and economists, and of particular interest are the population differences between geographical regions. The question of interest in this analysis is whether there is a population difference between prefectures (a prefecture is like a country) in northern Japan and southern Japan. The Null Hypothesis is that there is no difference in population between the two regions' prefectures, and the Alternative Hypothesis is that there exists a difference in population between the two regions' prefectures. The alpha level for statistical significance will be $\alpha = 0.05$.

Methodology

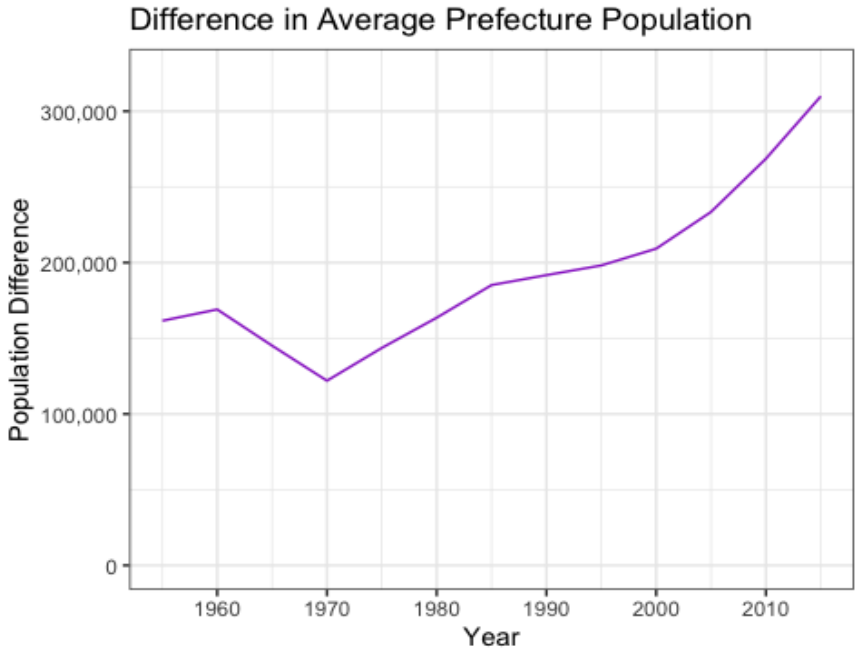
The dataset used in this analysis contains population counts from all of Japan's prefectures in five-year increments from 1955-2015. Two regions: the northern region of Tohoku and the southern region of Kyushu were selected for the analysis, and population counts from the six prefectures of Tohoku and the seven prefectures of Kyushu were aggregated, after which a mean for each year was used to produce a measure of average prefecture population. This produced two datapoints per year - one for each region. A Wilcoxon signed rank test was used to determine if the regions' average prefecture populations are different from one another. This test was appropriate because the data is paired by year and assumptions of normality may not be appropriate for the samples used. The dataset used in the analysis is publicly accessible: https://www.kaggle.com/jd1325/japan-population-data

Statistics

The following graph shows the average prefecture populations of Tohoku and Kyushu from 1955-2015. It can easily be seen that the population of Kyushu's prefectures are always greater than the population of Tohoku's even in times of decreasing population.



Overall, both populations increase and decrease during the same time periods although potentially at different rates. To investigate this idea, a plot of the difference in average population was created, and it was determined that there is a consistent average population difference between the two regions' prefectures. Further study can better quantify whether or not the difference in population is changing significantly.



Results

The Wilcoxon signed rank test confirms what was seen in the plots that there is a difference in average population between the two regions' prefectures.

```
#
#                Wilcoxon signed rank test
#
# data:      Tohoku$Population  and  Kyushu$Population
# V = 0,     p-value = 0.0002441
# alternative hypothesis: true location shift is not equal to 0
```

The p-value is well below $\alpha = 0.05$ meaning that the Null Hypothesis is rejected in favor of the Alternative Hypothesis. Thus, it is concluded that there exists a significant difference between prefectures in Japan's northern region of Tohoku and its southern region of Kyushu. Due to clear graphical evidence, it is also reasonable to conclude that prefectures in the southern region of Kyushu have a greater average population than the northern region of Tohoku.

Conclusion

This analysis found that there is certainly a difference in average population between prefectures from the two regions in question. In this preliminary exploration of the data it was reasonable to use nonparametric methods since little exploration has been done to verify if normality assumptions are met. For further research, broader analysis of all regions of Japan should be done to better describe the human population of Japan over time as well as in depth analysis among prefectures. Plots of the data also have shown that time series analysis methods will be useful in futher research to better describe the differences in populations. In addition to quantitative research, historical and demographic research will also help to explain the causes of population change in Japan across its several regions.