

# Analysis on the relation between quantity of apprehended falsified bills and population in Brazil

By: Rafael Neves

## Summary

The goal of this paper is to verify if there is any relation between the population and the quantity of apprehended fake bills for Brazilian states in the time period between the years of 2001 and 2021. The data set for the population of the states is from Brazilian Institute of Geography and Statistics (IBGE) e the data set on the apprehended falsified bills is from Central Bank of Brazil (BCB).

## Explaining the data sets

The data set for the fake bills is composed of 5 variables. Of those, two are qualitative, State and Type. The State variable is the Brazilian Federative Unit (UF) where the bills were found. The Type variable describes which family of bills the apprehended bills came from. This variable can only obtain three values: Cédulas - 1a. família, Cédulas - 2a. família and Cédulas - em polímero. These three classifications just describe the making of the bills. The 1a. família bills are the original real bills from the Plano Real in 1994. They stopped being produced in 2010 and were substituted by the 2a. família bills, but are still in circulation, although it's rare to see them in the day to day. Finally the Cédulas - em polímero was a "celebratory" 10 reais bill for the 500 years of the arrival of the Portuguese on Brazilian soil that were only produced once in the year 2000. In recent years most bills in circulation are from the 2a. família. An image of the different bills can be seen below. The ones to the left are from the 1a família and the ones to the right are from the 2a família. The one on the bottom left with a red dot is the cédula em polímero.



In this paper the differences between the types will be ignored since the emphasis is on population. However it must be acknowledged that the different bills have different mechanisms of identification making some harder to falsify than others. Which may influence the results. For this it will be assumed that the distribution of the types is

uniform among all bills, minimizing the effects from the different types of bills.

As for the 3 quantitative variables. The first one is years, which will be limited to time period between 2001 and 2021. The second quantitative variable is value. Value is the monetary value of each bill apprehended, it can have the following values 1, 2, 5, 10, 20, 50, 100, 200. However, the 1 real bill stopped being produced in 2005, the 20 reais bill started production in 2002 and the 200 reais bill only was produced in 2020. These changes in the production of certain bills could lead to some noise in the data. However it will also be assumed that these changes will uniformly distributed among all bills. The last variable for this data set is Quantity. This represents the quantity of apprehended bills, these values range from 0 to 188792.

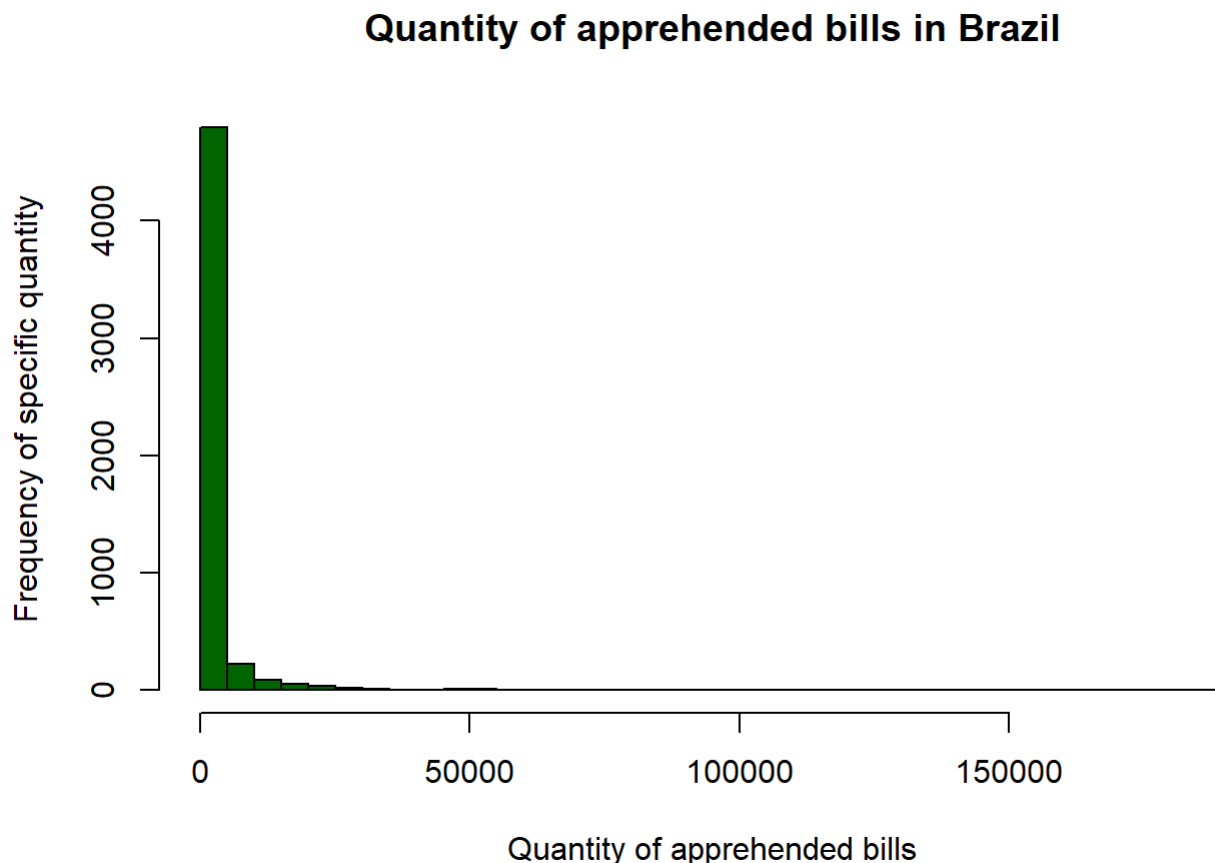
The data set for the population of the states has 29 quantitative variables, one of which represents the years between 2001 and 2021. Another represents the total population for each year and the rest represent the populations for each UF for each year.

## Analysis

The main variable that will be analysed is the Quantity variable from the fake bills data set. The values for this variable have an average of 1933.1 with a standard deviation of 7253.51 for the 5248 observations reported between years of 2001 and 2021.

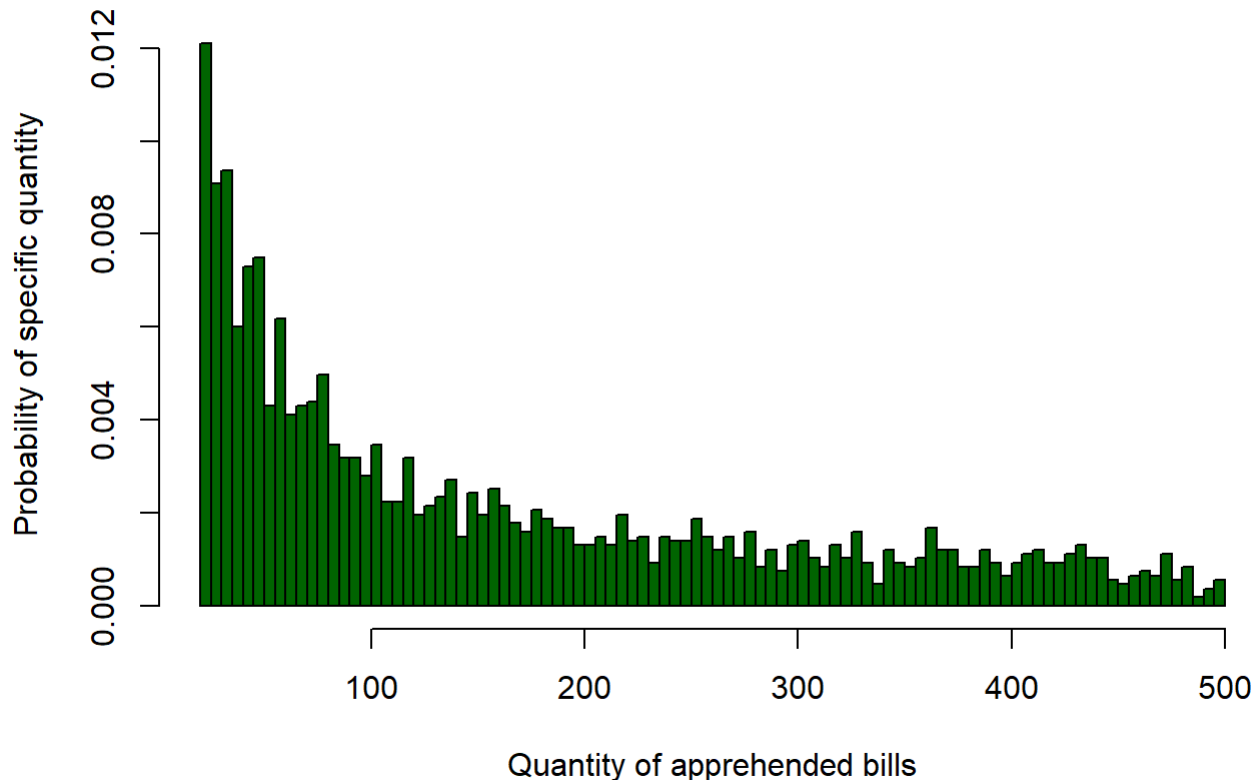
Now looking exclusively at the quantity for the bills apprehended in the state of São Paulo, the average becomes 16028.8, with a standard deviation of 27731.69 for the 220 observations reported between years of 2001 and 2021. This is interesting since on average the population of São Paulo for this time period is 42529551, with a standard deviation of 2852649.86, which represents about 21.78% of the average population of Brazil. This is a first indication of a relation between population and quantity of apprehended bills.

The following histogram shows the quantities of apprehended bills between the years of 2001 and 2021 for the entire country.



However, since 25.59% of the quantity of apprehended bills is for quantities less than 20 bills and 33.75% of the quantity of apprehended bills is for quantities larger than 500. Thus, a histogram bounded by quantities larger than 20 and less than 500 and with the probability of each specific quantity instead of frequency, can show a different perspective on the values of apprehended bills.

### Quantity of apprehended bills in Brazil in the range (20, 500)



For this smaller sample the histogram is still right-skewed. Which indicates that, even if considering that smaller apprehended quantities tend to happen due to just random chance, most apprehensions are more likely to be from individual bills being collected over the year rather than single large scale operations.

Returning now to the relation between the average of apprehended bills and the population of each UF. The comparison will be between the most populated UF, São Paulo with an average population over the years of 2001 and 2021 of 42529551, and the second most populated UF, Minas Gerais with an average population over the same time period of 20075257. Each average has a standard deviation of 2852649.86 and 1060101, respectively. The average quantities of apprehended bills for each UF over the time period is 16028.8 for São Paulo and 4730.6 for Minas Gerais. With a standard deviation of 27731.69 and 7111.26, respectively.

The confidence interval with 90% confidence for the average difference between the average quantities of apprehended bills for the states of São Paulo and Minas Gerais is  $[11277.451, 11318.949]$ . Meaning that the true difference will be in 90% of intervals constructed in this way.

Perhaps such a large difference between the averages of São Paulo and Minas Gerais is because of an outlier in the quantity of apprehended bills for a given year.

The year of 2006 in São Paulo has the largest value for quantity apprehended. With 188792 bills of 50 reais being apprehended. Comparing 2006 with the following year, 2007, to observe if the quantity for 2006 in São Paulo is an outlier or not.

Let the null hypothesis be that there is no difference between the average apprehended quantity in 2006 and 2007. Let the alternative hypothesis is that the difference is greater than zero.

$$H_0 : \mu_{2006} - \mu_{2007} = 0$$

$$H_1 : \mu_{2006} - \mu_{2007} > 0$$

$$P \left( \frac{\overline{X}_{2006} - \overline{X}_{2007} - (\mu_{2006} - \mu_{2007})}{\sqrt{\frac{\sigma_{2006}^2}{n_{2006}} + \frac{\sigma_{2007}^2}{n_{2007}}}} \leq \frac{33231.25 - 33252.75 - 0}{\sqrt{\frac{65229.9550749^2}{8} + \frac{61753.6470033^2}{8}}} \right)$$

Calculating the hypothesis, results in a  $p - value$  of 0.4997394, indicating that the year of 2006 is not likely to be an abnormality for São Paulo.

It could also be the case that both years are outliers for the values São Paulo. To verify if this is the case a hypothesis test with a level of significance of  $\alpha = 0.05$  will be performed. Let the null hypothesis be that the average quantity of bills apprehended between the years of 2001 and 2021 is the same as the average quantity of bills apprehended in the year 2006. Let the alternative hypothesis be that it is actually less than the average quantity apprehended in 2006.

$$H_0 : \mu_{avg} = \mu_{2006}$$

$$H_1 : \mu_{avg} < \mu_{2006}$$

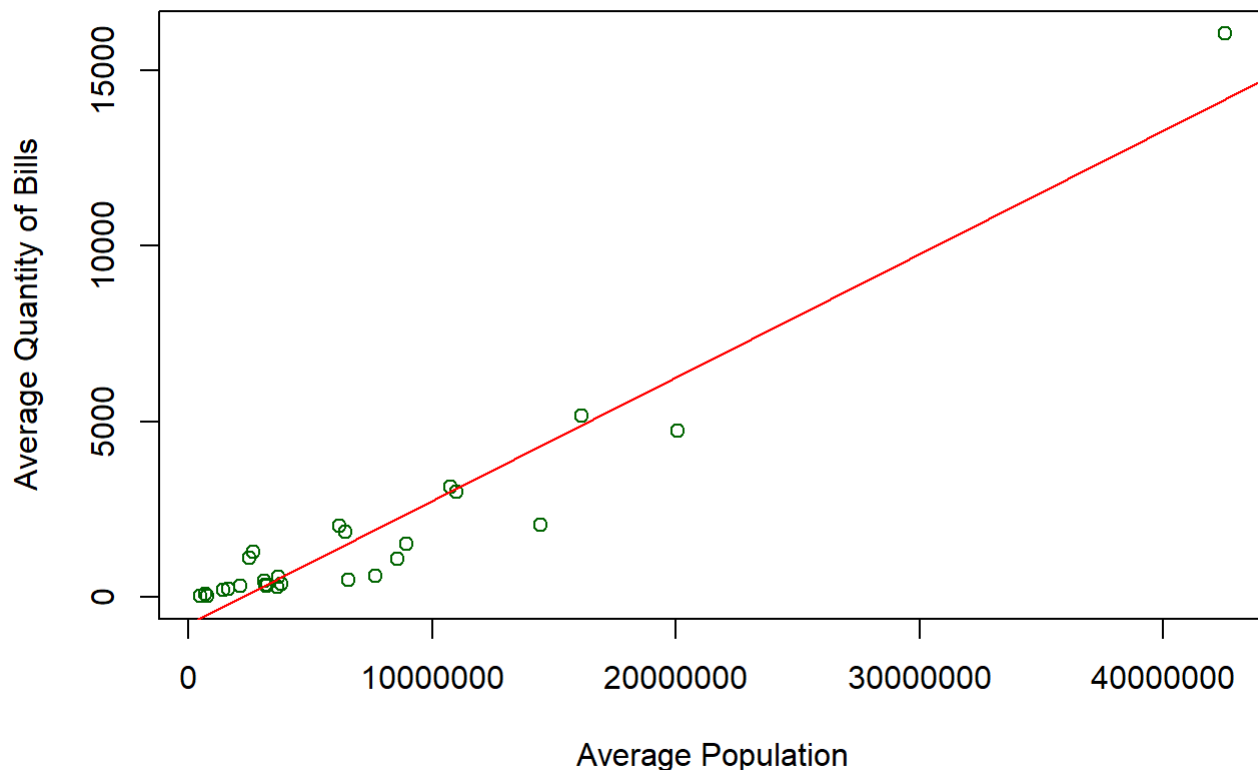
$$P \left( \frac{\overline{X}_{avg} - \overline{X}_{2006}}{\sqrt{\frac{\sigma_{avg}^2}{n_{avg}}}} \leq \frac{16028.8 - 33231.25}{\sqrt{\frac{27731.6916486^2}{220}}} \right)$$

Performing the hypothesis test results in a  $p - value$  of  $1.776581 * 10^{-20} < \alpha = 0.05$ . Therefore, it rejects the null hypothesis and shows that the years 2006 and 2007 are very likely to be outliers.

Analizing the data from the most populous UF, São Paulo, leads to the conclusion that the years of 2006 and 2007 are outliers from the data. However, every UF is likely to have outliers in their quantities. The best way to minimize the influence of such, is to use the maximum amount of data and analyze the entire time period.

The following scatter-plot has the average population on the x-axis and the average quantity on the y-axis. It represents the averages for the years 2001 through 2021.

## Relation between average population and quantity of bills by UF



Linear relation between population and quantity of bills per UF between 2001 and 2021

Notice the positive linear relation between population and quantity of bills. This shows that the initial assumption that population and quantity of apprehended bills are related was correct.

This relation can be explained by the fact the the more people a UF has, the greater the amount of bills in that UF. If the probability of any given bill being fake bill is the same for all bills, it would explain the relation between population and falsified bills.

However, it could also be the case that the quantity of falsified bills is related to crime rates. Which are then related to population. This hypothesis escapes the goals of this analysis, but let it be an indication for future research topics.

In conclusion, there seems to be a relation between population and quantity of apprehended bills. However a broader analysis is needed, since there could be to other factors that influence the quantity of apprehended bills.

## References

- Banco Central do Brasil, . (2015). Cédulas do real. Banco Central do brasil. Retrieved November 23, 2021, from <https://www.bcb.gov.br/cedulasemoedas/cedulas> (<https://www.bcb.gov.br/cedulasemoedas/cedulas>).
- Banco Central do Brasil, . (2015). Cédulas do real. Retrieved 2021, from <https://www.bcb.gov.br/content/cedulasemoedas/PublishingImages/indexcedulas/notassegundafamilia.PNG> (<https://www.bcb.gov.br/content/cedulasemoedas/PublishingImages/indexcedulas/notassegundafamilia.PNG>).
- Banco Central do Brasil, . (2021). Falsificações por ano e Por Estado. Portal Brasileiro de Dados Abertos. Retrieved November 12, 2021, from <https://dados.gov.br/dataset/falsificacoes-por-ano-e-por-estado> (<https://dados.gov.br/dataset/falsificacoes-por-ano-e-por-estado>).
- IBGE, . (2021). População residente estimada. Tabela 6579: População Residente Estimada. Retrieved November 21, 2021, from <https://sidra.ibge.gov.br/tabela/6579> (<https://sidra.ibge.gov.br/tabela/6579>).

Wikipedia, . (2021). Hyperinflation in Brazil. Retrieved 2021, from

<https://i1.wp.com/radames.manosso.nom.br/palavras/files/2019/10/real-1a-familia.jpg?resize=474%2C881>

(<https://i1.wp.com/radames.manosso.nom.br/palavras/files/2019/10/real-1a-familia.jpg?resize=474%2C881>).