The Relationship Between Modern Literacy Levels and Income Per Capita Based on Proximity to Jesuit Mission Sites

Arshnoor Kaur Gill

09/12/2020

Abstract

In this paper we attempt to reproduce 'Human Capital Transmission, Economic Persistance and Culture in South America' by Felipe Valencia Caicedo. We found that there is a 10% increase in modern literacy levels and a 0.2 log increase in the income per capita if one lives in a municipality 100 km closer to a Jesuit mission site in Brazil, Paraguay and Argentina. This suggests that investment in a community may lead to a longterm development of the prosperity of a community, and that the impact of missionaries on communities is a complex relation that deserves further research. A weakness is that through using Jesuit mission sites as proxies for sources of investment in this community, a false linkage could be created between the presence of missionaries and economic and social development in a municipality.

Keywords: Modern literacy, linear regression, South America, income per capita

1 Introduction

The activities of missionaries in developing countries is a hotly contested issue in our modern context. Though all religions have delegates who attempt to spread their faith, this is most associated with the Christian religion, especially in times of colonial expansion. Some believe that the work of building physical infastructure such as schools and hospitals that missionaries associate themselves with justifies the presence of these religious representatives in developing nations. However, there is the other perspective that missionary work is simply a modern by-product of imperialism, which ultimately results in hurting Indigenous populations through unintended consequences of contact, such as contracted disease. Some countries like Nepal have in fact labelled this practice illegal (Luckhurst). What this reproduction of this paper intends to do is examine the effect of Jesuit missions in Argentina, Brazil and Paraguay, in a set of missions called the "reductions" before the Jesuits were then ordered out of South America by the then king of Spain, in order to examine the longlasting effects on the local populations.

The way these findings were derived was through the use of a standard linear regression, the distance from Jesuit mission sites in kiometers used as the explanatory variable to attempt to predict (a) the 2000-2002 literacy rate as a percentage of the municipality (15 or 25 and older), and (b) the ln of the income per capita in 2000 of a municipality. The data is from the initial report published by Valencia Caicedo, which from him was collected from several sourses including: the national censes from Brazil, Paraguay and Argintina, as well as Jesuit Archives and firsthand correspondances from Jesuit officials in Rome (Valencia Caicedo). **The data seems to indicate that being 100 kilometers farther from a Jesuit mission site increases illiteracy in the population 15 or over by approximately 10% in a municipality. Moreover, this reproduction of the paper validates the initial finding that being 100 kilometers farther from a mission site decreases income per capita by 0.2 log in 2000.**

The outline of the paper is as follows. In the **Data** section, I further with reference to Valencia Caicedo's paper discuss the methodology behind the data collection utilized in this report. Within this section, I deliberate on the strengths and purposes of the data for the purpose of this regression analysis, as well as touch on summary statistics for variables of interests. In the following section, **Model**, I lay out the

regression equations that I will be utilize for the OLS analysis. In the **Results** section, there will be graphical and tabulated representation of the three relationships being investigated, and in the **Discussion** section the implications and legitimacy of the results in terms of significance will be deliberated, along with strengths and weaknesses of this reproduction. The most inherent of the weaknesses of this reproduction is that there may be a missing cofounder that creates a false relationship between the presence of Jesuit mission sites and literacy levels or economic prosperity of a location, as well as the lack of data regarding income effects in Argentina. For instance, it may not be anything in particular regarding the Jesuit missions but rather just any investment regardless of source in a community in terms of external aide or money that may lead to eventual relative prosperity. However, on the other hand, a key strength of the data is innovatively simplifying complex geographical calculations into a specific numerical variable, the difference from a mission site, alowing for straight-forward and easily interpretable analysis.

2 Data

This data, initially collected by Caicedo, comes from a variety of different sources. Covering Misiones and Corrientes within Argentina, Misiones and Itapúa within Paraguay and Rio Grande du Sol within Brazil, where the Jesuit mission sites were (based on historical analysis), the 578 observations are all regarding the community on a municipal level.

The data Caicedo uses regarding literacy rates (the percentage of the population of a municipality which is literate) and income (in terms of the Brazilian currency) comes from the national censes of these countries. For Brazil, this is the Brazilian Institute of Geography and Statistics (IBEG), the agency responsible for undertaking the census in Brazil. It appears that there is a discrepency for the literacy rate, in that for some municipalities it may have been collected for people fifteen or older, and other municipalities collected for people twenty-five or older. This sampling technique is described as "probabilistic", with the frame being permanent residents who live in private homes and the population being every citizen in Brazil. The technique appears to be systemic in intent, with the chosen housholds rotated on a yearly basis for whom will be picked for the sample ("Continuous National Household Sample Survey - Continuous PNAD"). For Paraguay, the literacy data comes from the Dirección General de Estadìstica, Encuestas y Censos (DGEEC), the national statistical agency, for the population per municipality 10 years or older in 2002. This appears to be probabilistic as well, but I was unable to recover the exact sampling technique undergone for the collection of this data. In Argentina, the literacy rate is from the National Institute of Statistics and Census of Argentina, for the population (10 years or older), in 2001. This data was collected through sampling individuals, households and dwellings ("OECD Assessment of the Statistical System of Argentina and Key Statistics of Argentina") as the frame, with the total population of Argentina as the population. There are obvious weaknesses in accumulating this data in this way—not only do they come from different agencies from different years, the cut-offs for the literacy are also in fact different. That being said, the data is from a time relatively close together, and the analysis later may be able to account for state effects.

In terms of income, direct income data could only be derived for Brazil, from the Institute of Applied Economic Research for 2000, in terms of annual earnings in the domestic currency for 2000, with the same sampling technique as described earlier. For Paraguay, data was collected through the World Bank for 2008 in the same technique and scope such that it could be compared to the Brazil observations on a muncipal level. That being said, there was no data in terms of income available for Argentinian municipalities in this time period, so the data regarding income in terms of distance from Jesuit sites will ignore Argentina. Income as a variable was transformed to "ln" of income, the natural logarithm, both because it makes the relationship with income further linear, and also because it makes much more intuitive sense to consider how the distance from a Jesuit site would change income in terms of percentage rather than dollars. There are evident weaknesses to this—through a lack of information regarding income for Argentinian municipalities, we cannot account for the direct impact that the sites had in an area where they were highly prominent. However, a strength is the conversion to the ln of income—not only does it follow the accepted standard when it comes to economic analysis, using the natural logarithm also just makes sense. Someone who is making 50,000 dollars a year would be much more impacted by a wage increase of 1,000 dollars, versus someone making 500,000 dollars a year. The quality of life is not impacted uniformly by a unit increase of

income regardless of where you are on the income scale, so using percentages universalizes the experience somewhat. Regardless of how much money you make, it makes intuitive sense that a 10% increase would be of same relative importance to you as someone who makes half as much or double your yearly income.

The variables of important for the purpose of this analysis are: the distance from a Jesuit site each municipality is from (in kilometres), the literacy rate of each municipality, from the year between 2000 and 2002 that it was collected, and the ln of the yearly income in the Brazilian currency. For literacy, there were a wealth of options of how this could've been measured—for instance, one variable from Caicedo was centered around "CFE"s, or country-fixed effects. These effects essentially take into account dummy variables for the country that the municipality is a part of, and uses it to affect the literacy variable. I've decided that I will use the literacy rates without the CFEs, such that it is less transformed and more "raw" data, with the intention to use the country each municipality is from as a categorical variable in the linear regressions that will be undergone. Moreover, other variables that will be considered in the analysis are variables which represent geographic information, including: the country, the distance from a river, the altitude, the precipitation levels and the distance from the coast. This geographical data was derived from the work of the organization BIOCLIM. The intent is hopefully to use these variables to stabalize any effect terrain may have had to access to information.

Below is distribution of the income of the observations from Brazil and Paraguay:

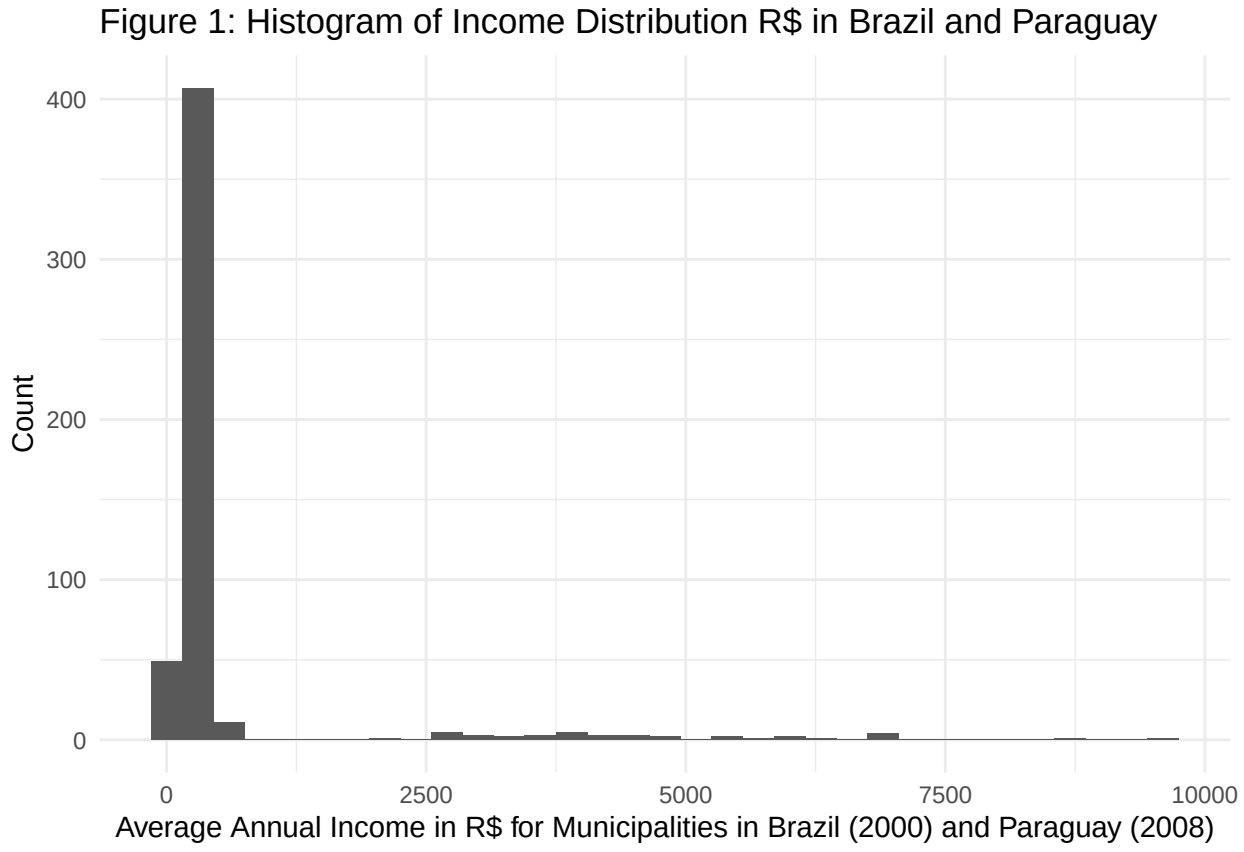


Figure 1: Histogram of Income

```
## Warning: Removed 29 rows containing non-finite values (stat_bin).
```

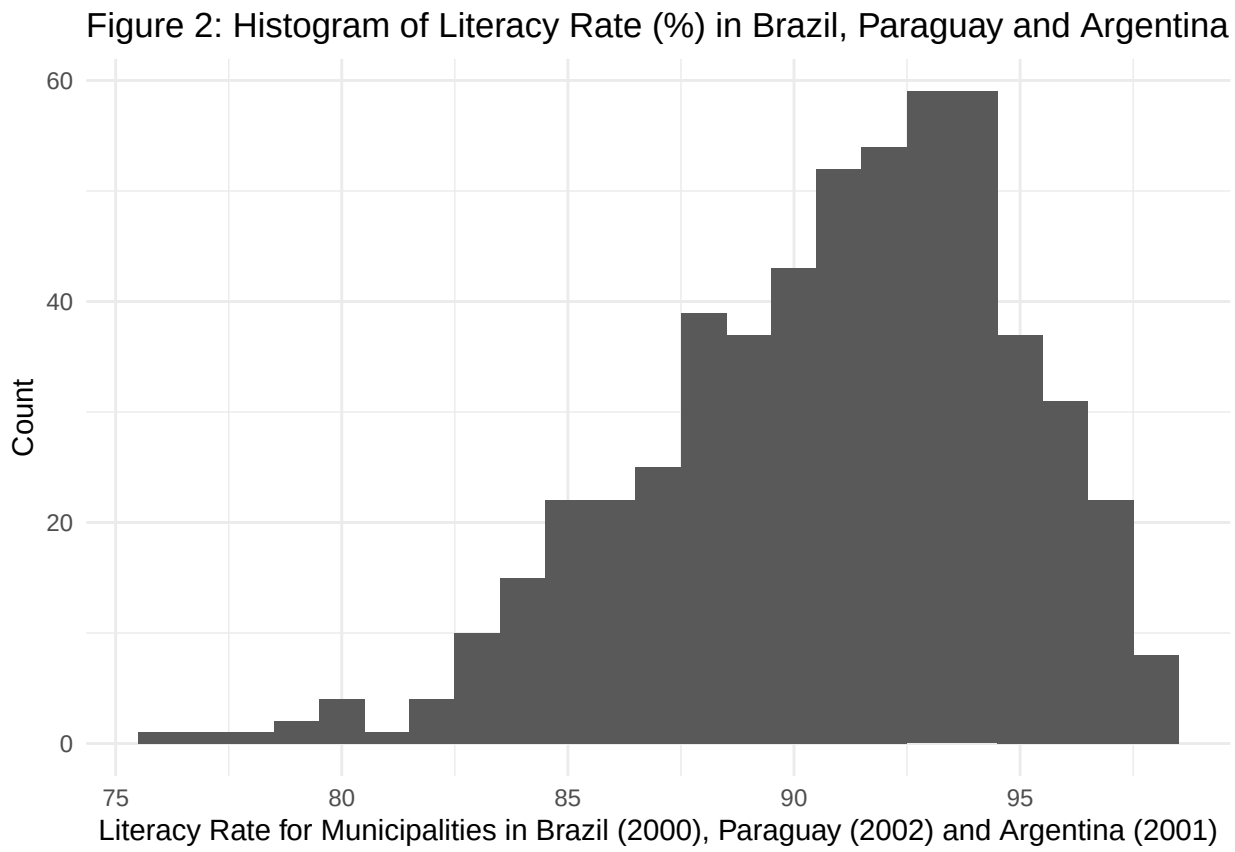


Figure 2: Histogram of Modern Literacy Rates

```
new_inc %>%
  summarize(mean = mean(income), median = median(income), min = min(income), max = max(income), q1 = qua
```

```
##       mean   median      min      max       q1       q3     var
## 1 578.9704 244.1659 84.79019 9637.997 188.0417 311.4556 1566601
```

```
literacy %>%
  drop_na %>%
  summarize(mean = mean(literacy), median = median(literacy), min = min(literacy), max = max(literacy),
```

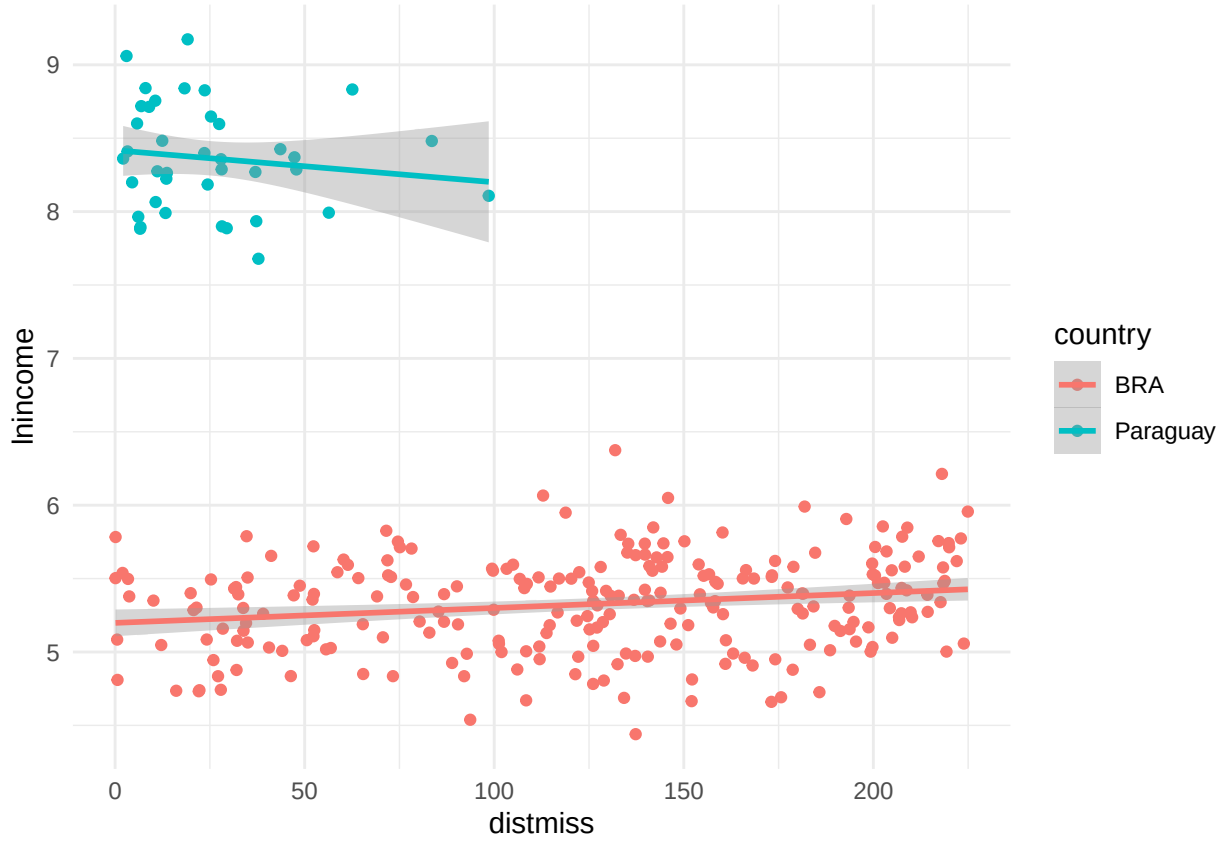
```
## # A tibble: 1 x 7
##    mean median   min   max    q1    q3   var
##   <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  90.8   91.5  75.7  98.4  88.0  94.0  17.7
```

```
new_income <- income %>%
  filter(distmiss < 225)

scatter1 <- ggplot(data = new_income, aes(x = distmiss, y = lnincome, color = country)) +
  geom_point() +
  geom_smooth(method = lm) +
  theme_minimal()

scatter1
```

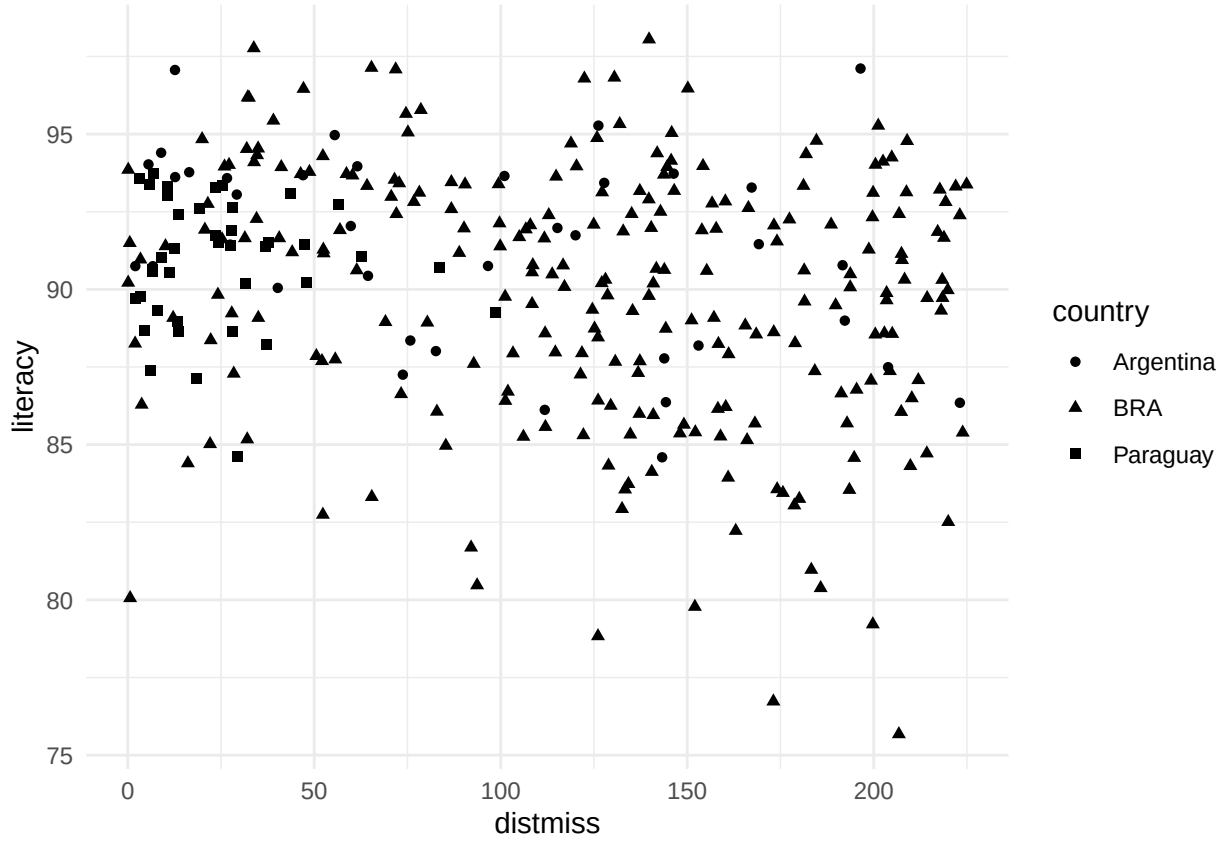
```
## `geom_smooth()` using formula 'y ~ x'
```




```
new_literacy <- literacy %>%
  filter(distmiss < 225)

scatter2 <- ggplot(data = new_literacy, aes(x = distmiss, y = literacy, shape = country)) +
  geom_point() +
  theme_minimal()

scatter2
```




```
model <- lm(literacy ~ distmiss + country, data = figureii)
summary(model)
```

```
##
## Call:
## lm(formula = literacy ~ distmiss + country, data = figureii)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.0371  -2.3183   0.6035   2.6718   8.4135
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     92.621048   0.698918 132.521  < 2e-16 ***
## distmiss        -0.013786   0.003567  -3.865 0.000135 ***
## countryBRA      -1.057539   0.668316  -1.582 0.114560
## countryParaguay -1.408368   0.890434  -1.582 0.114726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.777 on 316 degrees of freedom
## Multiple R-squared:  0.06427,    Adjusted R-squared:  0.05539
## F-statistic: 7.235 on 3 and 316 DF,  p-value: 0.0001041
```

```r
model2 <- lm(lnincome ~ distmiss, data = figureiiii_iv)
summary(model2)
```

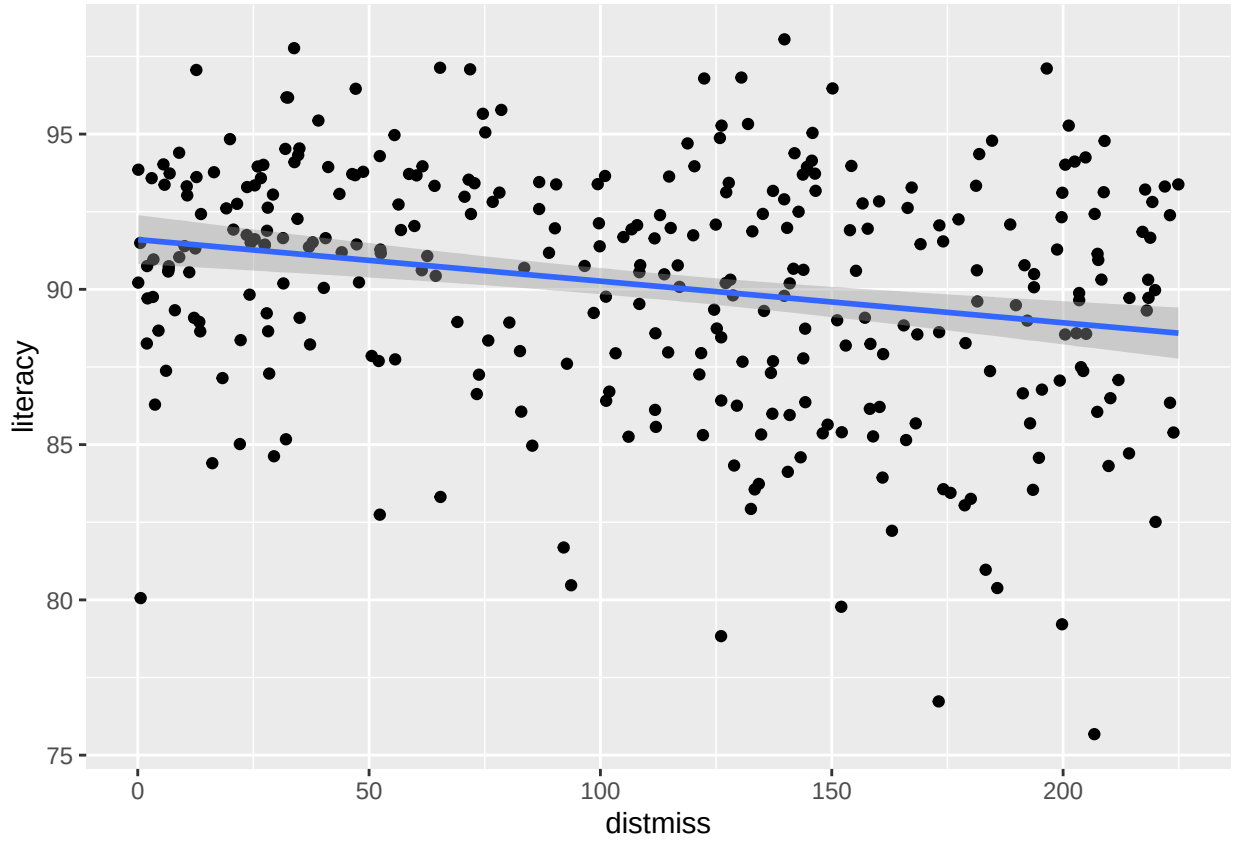
```
##
## Call:
## lm(formula = lnincome ~ distmiss, data = figureiiii_iv)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0436 -0.7781 -0.1616  0.6309  4.2950
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.7912950  0.1265976  61.544  < 2e-16 ***
## distmiss    -0.0037111  0.0009702  -3.825 0.000161 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.106 on 279 degrees of freedom
## Multiple R-squared:  0.04983,    Adjusted R-squared:  0.04643
## F-statistic: 14.63 on 1 and 279 DF,  p-value: 0.0001612
```

3 Data

```r
figureii <- figureii %>%
  mutate(mission_char = ifelse(mission == 1, "Mission Site", "Not Mission Site"))

ggplot(figureii, aes(x = distmiss, y = literacy)) + geom_point() + geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



4 Model

In order to graph the relations between the predictor variable, distance from a Jesuit mission, and the two response variables described thus far, (a) the literacy rate and (b) natural logarithm of the income of the sampled municipalities. Both relations will be derived through the use of ordinary least squares regression analysis, using the *lm()* function. This is a frequentist approach and relies on the concept of fixed parameters in the following equations:

$$\text{literacy rate} = \beta_0 + \beta_1 \cdot \text{distance from Jesuit mission} + \beta_2 \cdot \text{Brazil} + \beta_3 \cdot \text{Paraguay} + \beta_4 \cdot \text{distance from river}$$

$$+\beta_5 \cdot \text{distance from coast} + \beta_6 \cdot \text{precipitation level} + \beta_7 \cdot \text{altitude} + \text{error}$$

In which: -

$$\beta_0$$

is the intercept parameter, in which the distance from a Jesuit mission is 0 km. -

$$\beta_1$$

is the slope parameter, which represents how the percentage of the total population's literacy would change with a one kilometer difference in km. —

$$\beta_2$$

is the difference between the expected value of the literacy at any level —

Brazil

is a dummy variable which equals "0" if the municipality being observed is in not in Brazil, and "1" if the municipality is in Brazil. —

Paraguay

is a dummy variable which equals "0" if the municipality being observed is in not in Paraguay, and "1" if the municipality is in Paraguay. -

5 Results

6 Discussion

7 References

- JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2020). rmarkdown: Dynamic Documents for R. R package version 2.3. URL https://rmarkdown.rstudio.com.
- Luckhurst, Toby (2018). John Allen Chau: Do missionaries help or harm? BBC News, https://www.bbc.com/news/world-46336355.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- Valencia Caicedo, Felipe, 2018, "Replication Data for: 'The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America' ", https://doi.org/10.7910/DVN/ML1155, Harvard Dataverse, V1
- Yihui Xie (2019). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.26.
- https://www.ibge.gov.br/en/statistics/social/population/16833-monthly-dissemination-pnadc1.html
- http://www.oecd.org/statistics/good-practice-toolkit/countryassessments/OECD-Assessment-of-the-Statistical-System-and-Key-Statistics-of-Argentina.pdf