INDIAN STATISTICAL INSTITUTE

Bachelor in Mathematics (Hons.) First Year

PROJECT REPORT

Segregation of Students with Special Educational Needs in Brazilian Municipalities and Its Relationship with Student Performance

Introduction to Statistics and Computation with Data

Anuvab De
Arkaprovo Das
Arnav Kanaujia
Siddharto Das

Supervisor: Rituparna Sen

April 17, 2025

Abstract

This is an attempt to extend the results obtained in the data paper **'A dataset on the segregation of students with disabilitites in Brazil' (Rafael Verão Françozo 2024)**. The aim of this project is to verify whether the goal set out in Salamanca declaration and the Sustainable Development Goals of Quality Education can indeed be effected due to reduced segregation. We have used methods such as Multiple Regression and Maximum Likelihood Estimate (MLE) to analyse the given data and draw linear models to make conclusions on this topic. This report details these methods and the conclusions.

Acknowledgements

We would like to convey our deep gratitude to Prof. Rituparna Sen for her insightful feedback and encouragement, without which this project would not have been successful. It was a great learning experience.
We would also like to thank our friends, Souparna Pal and Priyankar Biswas, for their helpful suggestions.

Anuvab De
Arkaprovo Das
Arnav Kanaujia
Siddharto Das

Contents

Chapter 1

Introduction

In 1994, UNESCO organized the Salamanca Conference in Spain, where representatives from 92 countries came together to address the rights of children with special educational needs. The conference resulted in the Salamanca Statement and a Framework for Action, advocating for inclusive education. This landmark document emphasized that all children, regardless of their abilities or disabilities, should learn together in the same educational environments. The Statement calls for schools and education systems to be redesigned to accommodate the diverse learning needs and characteristics of every child. It promotes equality and acceptance, ensuring that all children have access to quality education in inclusive settings. By fostering cooperation and respect among students, inclusive education helps to build more equitable and harmonious societies.

1.1 History of Students with SEN in Brazil

Historically, students with special education needs were segregated from regular schools and sent to specialised schools or specialised classes for assistance. This approach limited their interaction with peers in mainstream education and often reinforced social barriers. In recent years, a series of public policies like the Salamanca Statement (UNESCO 1994) and the National Policy on Special Education (Brazil 2008), aimed at the inclusion of students with SEN have been implemented in Brazil. The Brazilian Inclusion Law (Lei Brasileira de Inclusão - LBI) (Brazil 2015), established comprehensive rights for individuals with disabilities, including access to inclusive education in regular schools. Schools are also increasingly adopting technologies and tools to accommodate the needs of SEN students, such as braille materials, digital resources, and assistive devices.

1.2 Aim of the Project

Our aim is to test whether desegregation benefits students with SEN in terms of their academic performance.

1.3 Solution Approach

The given data (results of students in the ENEM exam) follows a truncated normal distribution. To estimate the mean and variance of the original normal distribution, we use Maximum Likelihood Estimation (MLE). We also apply multiple linear regression along with the Fisher information matrix for our model and use techniques to handle heteroskedastic data.

Chapter 2

Nature of Data

2.1 Index of Dissimilarity (IoD)

The index of dissimilarity (IoD) is a metric used to quantify the segregation between two populations in n spatial units, referred to as tracts. The IoD varies in the closed range [0,1], where 0 represents no segregation between the populations in the tracts being analysed and 1 represents complete segregation.
A sample of the IoD calculation is as follows:

$$D_g = \frac{1}{2} \sum_{i=1}^{n} |X_i - Y_i| \tag{2.1}$$

With X and Y representing the proportion of the two populations being analysed. The value of D_g varies between 0 and 1 and represents the proportion of a group (1 or 2) that would need to move in order to create a uniform distribution of the population. (O. D. Duncan and B. Duncan 1955) The authors used IoD as a measure of inequality between students with SEN and without SEN in Brazil.

2.2 Given Data

We were given a dataset containing the raw segregation rates for each municipality in Brazil from 2008 to 2023. (Rafael Verão Françozo 2024).

In some cases the values are recorded as 'NA'. As shown in Table 1, this ocurred in 2009 for the city of Uiramutã and in 2010 for the city of São João da Baliza. An 'NA' value is recorded in two situations: 1 - when there are no records of students with disabilities in the schools in the city analysed and; 2 - when the city does not exist (there are some cities that were established between 2008 and 2023).

Table 1
Educational segregation between students with and without disabilities in the Brazilian state of Roraima.

CITY	COD_CITY	STATE	2023	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008
Alto Alegre	1400050	RR	0,37	0,39	0,27	0,36	0,39	0,38	0,34	0,38	0,31	0,28	0,33	0,32	0,31	0,44	0,51	0,44
Mucajaí	1400308	RR	0,19	0,21	0,24	0,30	0,26	0,22	0,25	0,30	0,26	0,29	0,37	0,39	0,42	0,31	0,67	0,76
Pacaraima	1400456	RR	0,32	0,33	0,33	0,33	0,30	0,29	0,35	0,54	0,48	0,46	0,45	0,41	0,46	0,38	0,54	0,62
Boa Vista	1400100	RR	0,18	0,18	0,18	0,21	0,22	0,23	0,24	0,24	0,25	0,24	0,24	0,24	0,30	0,37	0,44	0,51
Amajari	1400027	RR	0,54	0,45	0,39	0,41	0,43	0,48	0,60	0,58	0,55	0,44	0,55	0,64	0,64	0,57	0,83	0,88
Uiramutã	1400704	RR	0,55	0,51	0,53	0,55	0,60	0,67	0,73	0,63	0,71	0,71	0,78	0,84	0,96	0,99	NA	0,91
Bonfim	1400159	RR	0,33	0,33	0,39	0,38	0,36	0,38	0,39	0,44	0,40	0,45	0,45	0,48	0,71	0,73	0,87	0,73
Cantá	1400175	RR	0,20	0,25	0,24	0,26	0,32	0,28	0,29	0,33	0,27	0,39	0,36	0,38	0,45	0,50	0,66	0,83
Normandia	1400407	RR	0,48	0,50	0,50	0,51	0,52	0,47	0,47	0,50	0,50	0,48	0,60	0,58	0,65	0,71	0,75	0,90
Caracaraí	1400209	RR	0,27	0,29	0,33	0,28	0,27	0,25	0,28	0,30	0,32	0,24	0,29	0,37	0,40	0,37	0,47	0,50
Iracema	1400282	RR	0,24	0,27	0,24	0,26	0,27	0,44	0,39	0,26	0,62	0,50	0,55	0,67	0,57	0,80	0,50	0,57
São João da Baliza	1400506	RR	0,34	0,37	0,36	0,29	0,42	0,37	0,40	0,31	0,42	0,33	0,40	0,55	0,65	NA	0,71	0,95
Caroebe	1400233	RR	0,25	0,20	0,36	0,29	0,35	0,43	0,32	0,45	0,54	0,42	0,60	0,70	0,75	0,78	0,86	0,64
São Luiz	1400605	RR	0,23	0,25	0,32	0,25	0,28	0,30	0,32	0,27	0,37	0,37	0,35	0,25	0,25	0,45	0,40	0,55
Rorainópolis	1400472	RR	0,17	0,17	0,21	0,24	0,21	0,20	0,23	0,23	0,26	0,35	0,34	0,31	0,37	0,52	0,72	0,62

Figure 2.1: A portion of the data from the given data paper

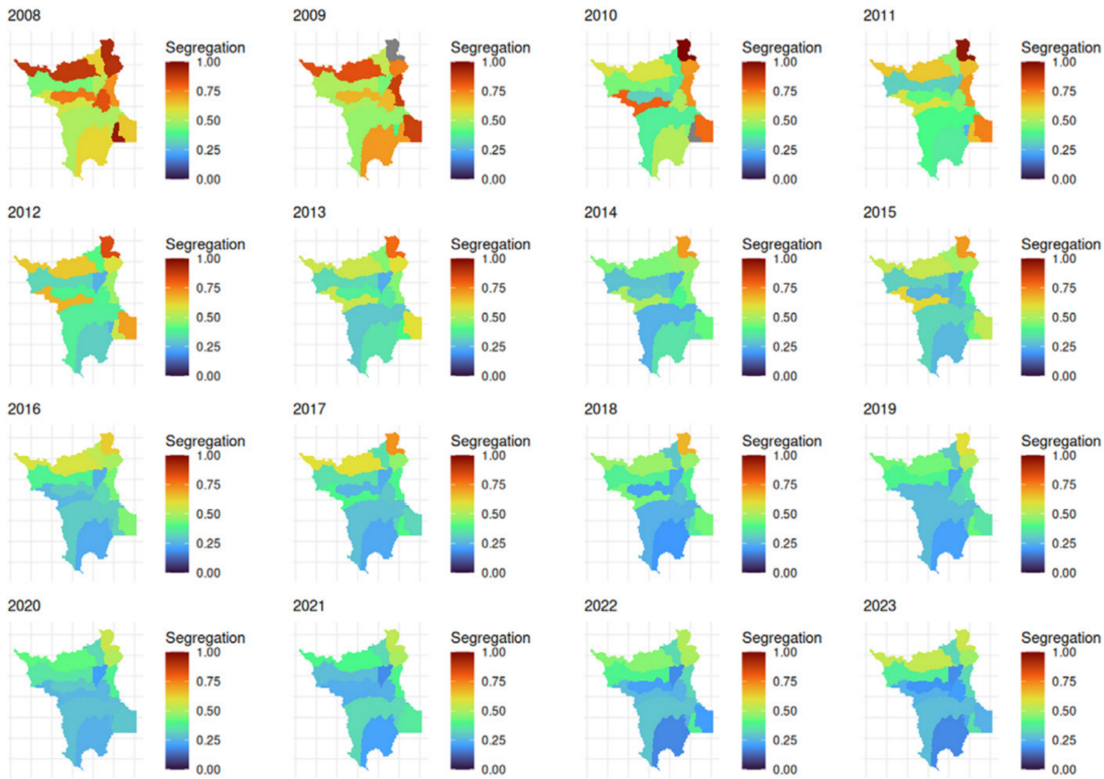
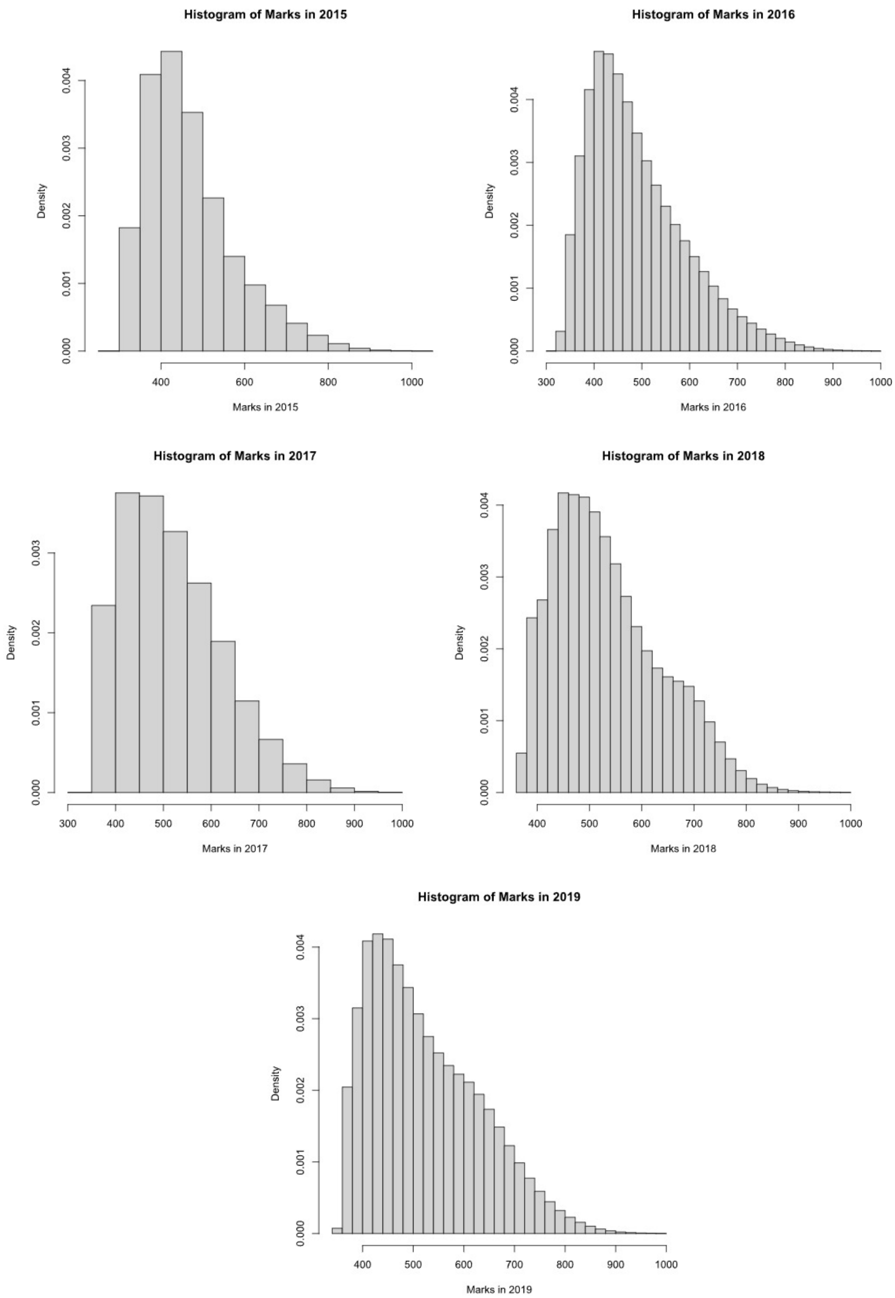


Figure 2.2: Segregation rate over the years in the Brazilian state of Roraima

2.3 Description of the data taken from ENEM exams

We further took data from the ENEM(Exame Nacional do Ensino Médio, shortened as ENEM, is a non-mandatory, standardized Brazilian national exam, which evaluates high school students in Brazil) exams (Wikipedia ENEM) for the years 2015 to 2019 and the histograms of student scores seems like a truncated normal distribution.(INEP 2020)



The dataset from the ENEM exams includes columns such as Student ID, type of disability, subject-wise scores, and exam centre location. However, for our analysis, we focus only on the columns related to disability type, Student ID, Mathematics scores, and exam centre location.

Chapter 3

Model

$A_i = M(t) + \epsilon_i$
$B_i = M(t) - (C\nu(t) + D) + \epsilon'_i$

- A_i is the marks obtained by the i^{th} student without Special Education Needs categorized by district and year of data collection.
- B_i is the marks obtained by i^{th} student with Special Education Needs categorized by district and year of data collection.
- M(t) is the expected score of a student without SEN in the ENEM exam at year t.
- ϵ_i represents the variability in students without SEN and ϵ'_i represents the variability in students with SEN. We assume that ϵ_i are i.i.d. normal and ϵ'_i are i.i.d. normal. We further assume ϵ_i and ϵ'_j are all mutually independent.
- We assume the disadvantage is a linear function of $\nu(t)$. $C\nu(t) + D$ represents the disadvantage faced by people with disabilities due to Segregation, where $\nu(t)$ denotes segregation rate with respect to year t and, C and D are constants.
- We assume disadvantage due to segregation is uncorrelated between different years.

We aim to do multiple linear regression on $C\nu(t) + D$ to find estimates for C and D. Let $f(\nu(t))$ represent the disadvantage caused by segregation. We want to fit $f(\nu(t_i)) = C\nu(t) + D$.

3.1 Multiple Linear Regression

Let $\hat{f}(\nu(t_i))$ be an estimator of $C\nu(t_i) + D$ in the year t_i, and let it be normally distributed. Let $\hat{\sigma}_i^2$ be a point estimate for the variance of the sampling distribution of $\hat{f}(\nu(t_i))$. (Rao 2002)

$$\text{Let } Y = \begin{bmatrix} \hat{f}(\nu(t_1)) \\ \hat{f}(\nu(t_2)) \\ \vdots \\ \hat{f}(\nu(t_k)) \end{bmatrix} \implies \mathbb{E}[Y] = \begin{bmatrix} \nu(t_1) & 1 \\ \nu(t_2) & 1 \\ \vdots & \vdots \\ \nu(t_k) & 1 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix}$$

$\hat{f}(\nu(t_i)) = C\nu(t) + D + \epsilon_i''$

Notation

$$\mathbb{D}(Y) = \begin{bmatrix} {\sigma_1}^2 & & & \\ & {\sigma_2}^2 & & \\ & & \ddots & \\ & & & {\sigma_k}^2 \end{bmatrix} \approx \begin{bmatrix} \hat{\sigma}_1^2 & & & \\ & \hat{\sigma}_2^2 & & \\ & & \ddots & \\ & & & \hat{\sigma}_k^2 \end{bmatrix} = G$$

$$X = \begin{bmatrix} \nu(t_1)) & \nu(t_2)) & \dots & \nu(t_k) \\ 1 & 1 & \dots & 1 \end{bmatrix}^T \qquad \beta = \begin{bmatrix} C & D \end{bmatrix}^T$$

$U = G^{-1/2}X$

$Z = G^{-1/2}Y$

$\mathbb{E}[Z] = G^{-1/2}\mathbb{E}[Y] = G^{-1/2}X\beta = U\beta \implies \mathbb{E}[Z] = U\beta$

$\mathbb{D}(Z) = G^{-1/2}\mathbb{D}(Y)(G^{-1/2})^T = G^{-1/2}G(G^{-1/2})^T = I$
$\implies \mathbb{D}(Z) = I$

If $U^TU\hat{\beta} = U^TZ$,
then $||Z - U\beta_1||^2 \geq ||Z - U\hat{\beta}||^2 \quad \forall\beta_1$

So, let $\hat{\beta} = (U^TU)^{-1}U^TZ$

then, $\mathbb{E}[\hat{\beta}] = (U^TU)^{-1}U^T\mathbb{E}[Z] = (U^TU)^{-1}U^TU\beta = \beta$

$$\begin{aligned} \text{and, } \mathbb{D}(\hat{\beta}) &= (U^TU)^{-1}U^T\mathbb{D}(Z)U((U^TU)^{-1})^T \\ &= (U^TU)^{-1}U^TU((U^TU)^T)^{-1} = (U^TU)^{-1} \end{aligned}$$

so that, $\hat{\beta}$ is a least square estimate for β

Proof

If U^TU is invertible then let $\hat{\beta} = (U^TU)^{-1}U^TZ$
$(Z - U\beta)^T(Z - U\beta)$
$= [Z - U\hat{\beta} + U(\hat{\beta} - \beta)]^T[Z - U\hat{\beta} + U(\hat{\beta} - \beta)]$
$= (Z - U\hat{\beta})^T(Z - U\hat{\beta}) + (\hat{\beta} - \beta)^TU^TU(\hat{\beta} - \beta)$
$\geq (Z - U\hat{\beta})^T(Z - U\hat{\beta})$
This shows the minimum of $(Z-U\beta)^T(Z-U\beta)$ is $(Z-U\hat{\beta})^T(Z-U\hat{\beta})$ and is attained at $\beta = \hat{\beta}$

3.2 Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is a method used to estimate the parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ of a probability distribution by maximizing the likelihood function. Given an independent and identically distributed (i.i.d.) sample $X = \{x_1, x_2, \ldots, x_n\}$, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} P(x_i|\theta_1, \theta_2, \ldots, \theta_k)$$

We normally take the log of the maximum likelihood function to ease computations:

$$\ell(\theta) = \sum_{i=1}^{n} \log P(x_i|\theta_1, \theta_2, \ldots, \theta_k)$$

Now we can use any numerical method to maximize this log likelihood function. Usually we take partial derivatives and setting them to zero. In our data we have a truncated normal so we use a MLE to estimate μ and σ.(Zeng and Gui 2021)

The distribution from which the marks of the students with SEN and those without SEN coming from a particular year and district is assumed to be normal.
However, in this case, we have a truncated sample.
Thus, the mean of the true normal is estimated using the MLE $\hat{\mu}_M$.

Let $\varphi_{\mu,\sigma}(x)$ denote the pdf of $\mathcal{N}(\mu, \sigma)$
And let $\phi_{\mu,\sigma}(x)$ denote the cdf of $\mathcal{N}(\mu, \sigma)$

To find the pdf of a truncated $\mathcal{N}(\mu, \sigma)$ at a we can find the conditional pdf.
$f_Z(z|z \geq a) = \frac{f_Z(z)}{P(z \geq a)} = \frac{\varphi_{\mu,\sigma}(z)}{1-\Phi_{\mu,\sigma}(a)}$

Likelihood Function:

$$L(\mu, \sigma) = \prod_{i=1}^{N} \frac{\varphi_{\mu,\sigma}(x_i)}{1 - \Phi_{\mu,\sigma}(a)}$$

Log Likelihood Function:

$$\ell(\mu, \sigma) = \sum_{i=1}^{N} log(\varphi_{\mu,\sigma}(x_i)) - N log(1 - \Phi_{\mu,\sigma}(a))$$

We have the MLE's $\hat{\mu}_M$,$\hat{\sigma}_M$ such that L(μ,σ) (or equivalently ℓ) is maximized at $\mu = \hat{\mu}$, $\sigma = \hat{\sigma}$
The M.L.E $\hat{\mu}_M$ is approximately normally distributed with expectation μ, the true mean.
The M.L.E $\hat{\sigma}_M$ is also approximately normally distributed with sd σ, the true sd. We can estimate the variance using the Fisher Information Matrix, $I(\mu, \sigma)$, and this is discussed in the next section.

(Bickel and Doksum 2001)

3.3 Fisher Information Matrix

Let us consider a random variable Y. Assume Y admits a density g with respect to a given measure depending on some parameter θ taking values in an open subset Θ of $\mathbb{R}^d$, such that the log-likelihood function log g is differentiable on Θ and that $\|\partial_\theta \log(g(y_i, \theta))(\partial_\theta \log(g(y_i, \theta)))^t\|$ is integrable, where x^t stands for the transpose of a vector or a matrix x. Then, by definition, the Fisher information matrix is given for all $\theta \in \Theta$ by :

$$I(\theta) = \mathbb{E}_\theta[\partial_\theta \log(g(y_i, \theta))(\partial_\theta \log(g(y_i, \theta)))^t]$$

When this expression can not be analytically evaluated, people are interested in computing an estimate of the Fisher information matrix. Considering this expression, one can derive a first moment estimator of the Fisher information matrix based on a n-sample $(y_1, ..., y_n)$ of observations:

$$\hat{I}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\nabla_\theta \log(g(y_i, \theta))\Big|_{\theta=\hat{\theta}}\right)\left(\nabla_\theta \log(g(y_i, \theta))\Big|_{\theta=\hat{\theta}}\right)^t$$

Where:

$$\nabla_\theta f(x, \theta) = \begin{bmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \vdots \\ \frac{\partial f}{\partial \theta_n} \end{bmatrix}$$

In our case:

$$\nabla_{(\mu,\sigma)} log(p(x, (\mu, \sigma))) = \begin{bmatrix} \frac{\partial log(p(x,(\mu,\sigma))}{\partial \mu} \\ \frac{\partial log(p(x,(\mu,\sigma))}{\partial \sigma} \end{bmatrix}$$

$p(x, (\mu, \sigma)) = \frac{\varphi_{\mu,\sigma}(z)}{1-\Phi_{\mu,\sigma}(a)}$

Therefore,

$$\nabla_{(\mu,\sigma)} log(p(x, (\mu, \sigma))) = \begin{bmatrix} \frac{\partial \varphi_{\mu,\sigma}(x))}{\partial \mu}\frac{1}{\varphi_{\mu,\sigma}(x)} - \frac{\partial \varphi_{\mu,\sigma}(a))}{\partial \mu}\frac{1}{1-\phi_{\mu,\sigma}(a)} \\ \frac{\partial \varphi_{\mu,\sigma}(x))}{\partial \sigma}\frac{1}{\varphi_{\mu,\sigma}(x)} - \frac{\partial \varphi_{\mu,\sigma}(a))}{\partial \sigma}\frac{1}{1-\phi_{\mu,\sigma}(a)} \end{bmatrix}$$

Finally,

$$\hat{I}(\mu, \sigma) = \frac{1}{n}\sum_{i=1}^{n}\left(\nabla_{(\mu,\sigma)} log(p(x, (\mu, \sigma)))\Big|_{\mu,\sigma=\hat{\mu},\hat{\sigma}}\right)\left(\nabla_{(\mu,\sigma)} log(p(x, (\mu, \sigma)))\Big|_{\mu,\sigma=\hat{\mu},\hat{\sigma}}\right)^t$$

Where $\hat{\mu}$ and $\hat{\sigma}$ are the estimates of μ and σ using MLE.
The dispersion of the MLE approaches the Cramér-Rao bound.
So, we can approximate the variance, co-variance matrix by the inverse of estimated fisher matrix. Therefore, in our case: $\hat{\sigma}_M = (\hat{I}^{-1}(\mu, \sigma))_{1,1}$ where $\hat{\sigma}_M$ is the estimate of the variance of the M.L.E. estimators.
From this we get the estimate for $\mathbb{D}(Y)$ that we needed to do the multiple linear regression as we have now got the $\hat{\sigma}_i^2$ for all i. (How we do so is discussed in the next section.)

(Wu 2021)(Delattre and Kuhn 2023)

3.4 Plan

1. We have a truncated normal data but in our model we need the mean of the true normal as an estimate for the mean of $\hat{f}(\nu(t_i))$ so first we do an M.L.E to find out $\hat{\mu}_M$ we also get $\hat{\sigma}_M$.

2. Now we claim that, $\hat{f}(\nu(t_i)) = \hat{\mu}_{M_{disabled}} - \hat{\mu}_{M_{abled}}$ is a good estimator for the disadvantage.

   Proof. As described before $\hat{\mu}_{M_{abled}}$ and $\hat{\mu}_{M_{disabled}}$ are normally distributed. And their distributed as follows:
   $\hat{\mu}_{M_{abled}}(t_i) \sim \mathcal{N}(M(t_i), \sigma_a^2(t_i))$
   $\hat{\mu}_{M_{disabled}}(t_i) \sim \mathcal{N}(M(t_i) - (C\nu(t_i) + D), \sigma_b^2(t_i))$
   Where $\sigma_a^2(t_i)$ is the variance of ϵ_i and $\sigma_b^2(t_i)$ is the variance of ϵ_i' as given in the model. Therefore, $\hat{\mu}_{M_{disabled}} - \hat{\mu}_{M_{abled}} \sim \mathcal{N}(C\nu(t_i) + D, \sigma_a^2(t_i) + \sigma_b^2(t_i))$.
   So we get that $\mathbb{E}(\hat{f}(\nu(t_i))) = C\nu(t_i) + D = f(\nu(t_i))$. □

3. Now we also need the variance of $\hat{f}(\nu(t_i))$ to do our regression so we use the Fisher information matrix within which we need to use the estimates obtained before in the M.L.E. to find $\hat{\sigma}_a^2(t_i)$ from $(1,1)^{th}$ entry of the inverse of the Fisher matrix of the population without SEN and we get $\hat{\sigma}_b^2(t_i)$ from $(1,1)^{th}$ entry of the inverse of the Fisher matrix of the population with SEN. We get: $\hat{\sigma}_i = \hat{\sigma}_a^2(t_i) + \hat{\sigma}_b^2(t_i)$
   And from this we can finally get $\mathbb{D}(Y)$ in our model.

4. Now using all the previous results we can do the regression to find $\hat{C}$ and $\hat{D}$ for specific places over different years.

Chapter 4

Implementation

We considered only the following columns from the raw datasets for the ENEM examination. The mathematics score is used as a measure of the academic performance of students. The place of the examination is used to approximate the place of schooling since this information is not available for many records.

Name of column	**Description**
NO_MUNICIPIO_PROVA	Exam municipality name for candidate
SG_UF_PROVA	Acronym of the federative unit where the candidate took the exam
NU_NOTA_MT	Marks obtained in Mathematics
IN_BAIXA_VISAO	Does the candidate have low vision?
IN_CEGUEIRA	Is the candidate blind?
IN_SURDEZ	Is the candidate deaf?
IN_DEFICIENCIA_AUDITIVA	Does the candidate have hearing impairment?
IN_SURDO_CEGUEIRA	Does the candidate have deafblindness?
IN_DEFICIENCIA_FISICA	Does the candidate have physical disability?
IN_DEFICIENCIA_MENTAL	Does the candidate have mental disability?
IN_DEFICIT_ATENCAO	Does the candidate have attention deficit?
IN_DISLEXIA	Is the candidate dyslexic?
IN_DISCALCULIA	Does the candidate have dyscalculia?
IN_AUTISMO	Is the candidate autistic?
IN_VISAO_MONOCULAR	Does the candidate have monocular vision?
IN_OUTRA_DEF	Does the candidate have any other disability or special condition?

We constructed new columns from the initial set. 'PLACE' is constructed by joining the columns 'NO_MUNICIPIO_PROVA' and 'SG_UF_PROVA' as strings. 'D_UNION' is 1 if any of the columns for disability is 1, and is 0 otherwise. The columns 'PLACE', 'D_UNION' and 'NU_NOTA_MT' are collected in new files. These reduced files are named red15.csv, red16.csv, red17.csv, red18.csv and red19.csv.

Name of column	**Description**
PLACE	NO_MUNICIPIO_PROVA@SG_UF_PROVA
D_UNION	Union of IN_*
NU_NOTA_MT	Same as before.

These reduced files are processed to find MLE estimates for the mean and standard deviation of the true normal underlying the distribution of the scores of the students without SEN and with SEN.

means.R

```
# data in reduced
M <- mean(reduced$NU_NOTA_MT)
S <- sd(reduced$NU_NOTA_MT)
ui <- c(0, 1)
dim(ui) <- c(1, 2)
ci <- c(0)
llike = function (dat, k) {
  function (ms) {
    m <- ms[1]
    s <- ms[2]
    -sum(log(dnorm(dat, m, s)))+length(dat)*log(1-pnorm(k, m, s))
  }
}
abled <- by(reduced, reduced$PLACE, function (dat) {
  constrOptim(c(M, S), llike(dat[dat$D_UNION == 0, 'NU_NOTA_MT'], a),
              NULL, ui, ci)
})
abled <- sapply(abled, `[[`, 'par')
disabled <- by(reduced, reduced$PLACE, function (dat) {
  constrOptim(c(M, S), llike(dat[dat$D_UNION == 1, 'NU_NOTA_MT'], a),
              NULL, ui, ci)
})
disabled <- sapply(disabled, `[[`, 'par')
tot <- as.data.frame(cbind(t(abled), t(disabled)))
```

The left cut-off values are found by inspection:

Year	***a* (Left cut-off point)**
2015	283.5
2016	319
2017	310
2018	355.5
2019	358.5

For some municipalities, insufficient data, particularly for students with SEN, prevented us from carrying out the computations, so these municipalities were dropped from the set. The variable tot is written into a csv file which gives us a table with the following columns:

Column	**Description**
1st column	PLACE
2nd column	Mean of students without SEN in that PLACE.
3rd column	SD of students without SEN in that PLACE.
4th column	Mean of students with SEN in that PLACE.
5th column	SD of students with SEN in that PLACE.

These files are named 15.csv, 16.csv, 17.csv, 18.csv and 19.csv.
To estimate the Fisher information matrix we need functions to compute some partial derivatives and we also need a function to calculate the estimated variance. These are listed in the file 'derivs.R'.

derivs.R

```
library(pracma)

phim <- function (x, m, s) {
  ((x-m)*exp(-(x-m)^2/(2*s*s)))/(sqrt(2*pi)*s^3)
}

phis <- function (x, m, s) {
  ((((x-m)^2)*exp(-(x-m)^2/(2*s^2)))/(sqrt(2*pi)*s^4))-(exp(-(x-m)^2/(2*s^2))/(sqrt(2*pi)*s^2))
}

E<-function (a, m, s) {exp(-((a-m)**2)/(2*s**2))}
PHIm<-function (a, m, s) {-E(a,m,s)/(sqrt(2*pi)*s)}
PHIs<-function (a, m, s) {-((a-m)*E(a,m,s))/(sqrt(2*pi)*s**2)}

esigma <- function(x, a, m, s) {
 n <- length(x)
 a11 <- sum(((phim(x,m,s)/dnorm(x,m,s))+(PHIm(a,m,s)/(1-pnorm(a,m,s))))**2)/n
 a12 <- sum(((phis(x,m,s)/dnorm(x,m,s))+(PHIs(a,m,s)/(1-pnorm(a,m,s))))*((phim(x,m,s)/dnorm(x,m,s))+(PHIm(a,m,s)/(1-pnorm(a,m,s)))))/n
 a21 <- a12
 a22 <- sum(((phis(x,m,s)/dnorm(x,m,s))+(PHIs(a,m,s)/(1-pnorm(a,m,s))))**2)/n
 mx <- c(a11, a12, a21, a22)
 dim(mx) <- c(2, 2)
 inv(mx)[1,1]
}
```

The function esigma gives the estimated variance.

The programme to estimate the variance for the estimator is in fishers.R.

fishers.R

```
# with cities in cities, data in reduced, info in qest, a set.

abled <- sapply(cities, function (city) {
  m <- qest[city, 1]
  s <- qest[city, 2]
  x <- reduced[reduced$D_UNION == 0 &
                 reduced$PLACE == city,
               "NU_NOTA_MT"]
  esigma(x, a, m, s)
})

disabled <- sapply(cities, function (city) {
  m <- qest[city, 3]
  s <- qest[city, 4]
  x <- reduced[reduced$D_UNION == 1 &
                 reduced$PLACE == city,
               "NU_NOTA_MT"]
  esigma(x, a, m, s)
})
```

In this case, due to insufficient data points within some municipalities, particularly for students with SEN, the estimated Fisher information matrix turned out to be nearly singular so these municipalities were dropped from the set. We are left with 925 municipalities. These are listed in the file fsselected.txt.

After all the required quantities were found, multiple linear regression, as described in section 3.1, can be performed. The following listing shows the programme to perform this.

linreg.R

```
library(stringi)

fs15$v15 <- fs15$x + fs15$y
fs16$v16 <- fs16$x + fs16$y
fs17$v17 <- fs17$x + fs17$y
fs18$v18 <- fs18$x + fs18$y
fs19$v19 <- fs19$x + fs19$y
selection <- Reduce(
  intersect,
  lapply(list(fs15, fs16, fs17, fs18, fs19), function (x) {x$Row.names})
)

vars <- as.data.frame(cbind(
  selection,
  fs15[fs15$Row.names %in% selection, "v15"],
  fs16[fs16$Row.names %in% selection, "v16"],
  fs17[fs17$Row.names %in% selection, "v17"],
  fs18[fs18$Row.names %in% selection, "v18"],
  fs19[fs19$Row.names %in% selection, "v19"]
))

names(vars) <- c('PLACE', 'v15', 'v16', 'v17', 'v18', 'v19')

dissimilarity <- raw.dissimilarity[,c(1, 3, 12:8)]
dissimilarity$PLACE <- stri_join(dissimilarity$NO_MUNICIPIO,
                                 dissimilarity$SG_UF,
                                 sep='@')

`15`$m15 <- `15`[1] - `15`[3]
`16`$m16 <- `16`[1] - `16`[3]
`17`$m17 <- `17`[1] - `17`[3]
`18`$m18 <- `18`[1] - `18`[3]
`19`$m19 <- `19`[1] - `19`[3]

xs <- as.data.frame(cbind(
  selection,
  `15`[selection, "m15"],
  `16`[selection, "m16"],
  `17`[selection, "m17"],
  `18`[selection, "m18"],
  `19`[selection, "m19"]
))
names(xs) <- c('PLACE', 'm15', 'm16', 'm17', 'm18', 'm19')
```

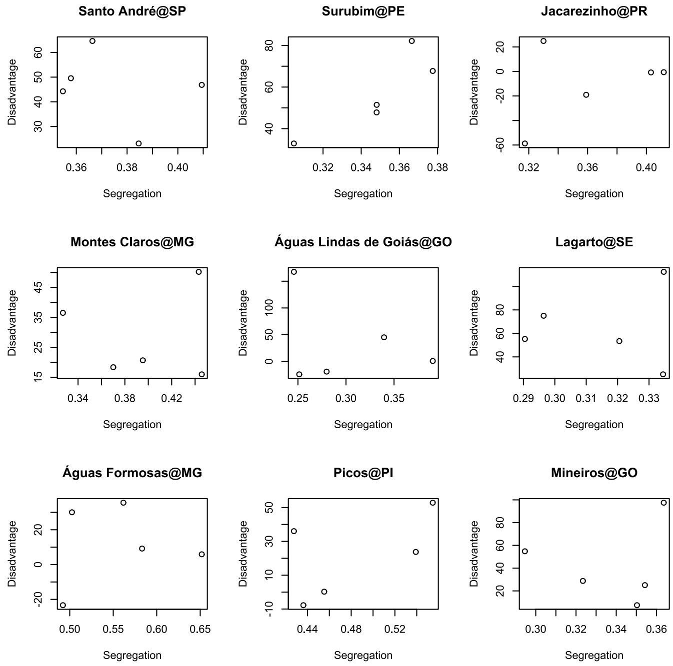
linreg.R

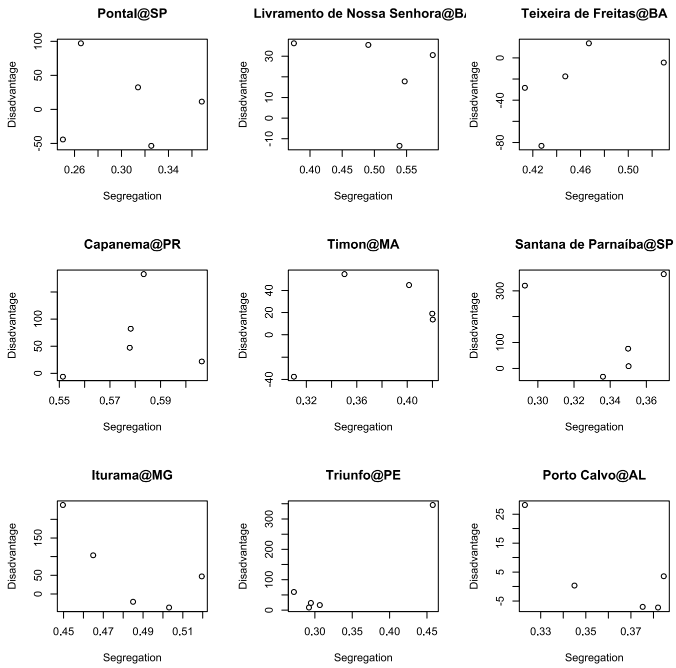
```
coefsolve <- function (place) {
  X <- simplify2array(c(
    dissimilarity[
      dissimilarity$PLACE == place,
      c("X2015", "X2016", "X2017", "X2018", "X2019")
    ],
    rep(1.0, 5)
  ))
  dim(X) <- c(5, 2)
  sG <- diag(
    as.numeric(simplify2array(vars[vars$PLACE == place,
                                   c("v15", "v16", "v17", "v18", "v19")])
  )^(-0.5)
  )
  U <- sG %*% X
  Y <- simplify2array(xs[xs$PLACE == place,
                         c('m15', 'm16', 'm17', 'm18', 'm19')
  ])
  Z <- sG %*% Y
  beta <- solve(t(U) %*% U, t(U) %*% Z)
  as.vector(beta)
}

coeffs <- as.data.frame(t(sapply(selection, coefsolve)))
names(coeffs) <- c('C', 'D')
```

The function coefsolve returns the estimate $\hat{\beta} = [\hat{C}, \hat{D}]$ given the municipality name. The results were compiled in the file coeffs.csv. The following is a ten-line sample from coeffs.csv.

PLACE	**C**	**D**
Alta Floresta@MT	171.077428638927	-29.4144832849468
Xinguara@PA	-321.353558846477	138.992588480574
Jataí@GO	-1703.33349027495	687.360494231541
Chapadinha@MA	-237.46568978277	103.125549496272
Novo Hamburgo@RS	-734.973148173562	328.162752531221
Teixeira de Freitas@BA	603.366922645836	-307.684138180409
Marabá@PA	-108.530869612552	71.2823115777139
Assis Chateaubriand@PR	-146.088995415837	98.7336817550898
Alvinópolis@MG	-1581.9808468883	1023.98859681936
Frutal@MG	-566.450127486294	261.149150292796

The plots given below are a plot of disadvantage vs. segregation over different years in 18 of the 925 municipalities considered.





4.1 Some Results

Some results obtained from the computations are as follows.

- 485 of the 925 municipalities considered showed a decreasing relationship of disadvantage with segregation (using the value of $\hat{C}$).
- 509 of the 925 municipalities considered showed a positive intercept of disadvantage with respect to segregation, i.e., showed a net disadvantage projected at zero segregation (using the value of $\hat{D}$).

Chapter 5

Conclusion

This study examined whether segregation affects academic performance of Special Educational Needs (SEN) students in Brazil using ENEM exam data from 2015 to 2019. Despite policies aligned with the Salamanca Statement and UN Sustainable Development Goals, we found no consistent relationship of inclusive education policies with student scores nationwide. In 52.4% of the municipalities considered, the disadvantage decreased with increase in segregation. This evidence suggests weak implementation or limited translation of policies into measurable academic benefits. Our conclusion indicates that the desegregation policy, as implemented currently in Brazil, does not correlate with improved academic outcomes for SEN students, suggesting the need for deeper evaluation of policy execution.

Bibliography

Bickel, Peter J. and Kjell A. Doksum (2001). “Mathematical Statistics: Basic Ideas and Selected Topics”. In: Prentice Hall. Chap. 2.

Brazil (2008). *Política Nacional de Educação Especial na Perspectiva da Educação Inclusiva [The National Policy of Special Education on the Perspective of Inclusive Education]*.

— (2015). *Brazilian Inclusion Law (Lei Brasileira de Inclusão-LBI)*.

Delattre, Maud and Estelle Kuhn (2023). *Estimating Fisher Information Matrix in Latent Variable Models based on the Score Function*. https://arxiv.org/pdf/1909.06094.

Duncan, Otis Dudley and Beverly Duncan (1955). *A Methodological Analysis of Segregation Indexes*. https://www.jstor.org/stable/2088328?origin=crossref&seq=1.

INEP (2020). *ENEM Microdata*. https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem.

Rafael Verão Françozo, et al. (2024). *A dataset on the segregation of students with disabilities in Brazil*. https://www.sciencedirect.com/science/article/pii/S235234092400934X.

Rao, C. R. (2002). “Linear Statistical Inference and Its Applications”. In: Wiley. Chap. 4.

UNESCO (1994). *The Salamanca Statement*.

Wikipedia (ENEM). *Exame Nacional do Ensino Médio*.

Wu, Xuan (2021). *Enhanced Monte Carlo Estimation of the Fisher Information Matrix with Independent Perturbations for Complex Problems*. https://arxiv.org/pdf/2104.07180.

Zeng, X. and W Gui (2021). “Statistical Inference of Truncated Normal Distribution Based on the Generalized Progressive Hybrid Censoring”. In: *Entropy* 23.2. url: https://doi.org/10.3390/e23020186.