# Apprentice Migration to Edinburgh (1775-1799)

## Abstract

This report investigates the migration patterns of apprentices to Edinburgh using a dataset that incorporates variables such as distance, population, degree of urbanization, and direction from Edinburgh, we aim to discern the influential factors affecting this migration. A Poisson Regression model was proposed to understand the relationships between these variables and the number of migrating apprentices. Preliminary visualizations suggest various relationships, with distance likely having a negative impact, while population and urbanization might boost migration. The report provides a comprehensive analysis, from data exploration to model conclusions, to shed light on historical migration patterns.

## Introduction

The dataset pertains to the migration of apprentices to Edinburgh from 33 regions in Scotland between the years 1775 and 1799. The goal is to discern how many apprentices migrated from each region and how various factors such as distance, population, degree of urbanization, and direction from Edinburgh influence these numbers. The principal question to answer is: "How do these factors impact the number of apprentices migrating to Edinburgh from different regions?"

## Data

This data is pertinent as it provides insights into apprentice migration to Edinburgh during the mentioned period. By understanding these patterns, we can glean insights into the motivations behind apprentice migrations, which might be influenced by socio-economic, geographic, and other factors. The exact origins of how these data were collected are not mentioned. However, given the source and the context, it's probable that they were derived from historical records or censuses.

As observed, there aren't missing values in the dataset, making it suitable for analysis. In the dataset there are seven “County Name” with 0 “Apprentices”, in this case, replace with 0.0001 and transformed in logarithmic.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

## Model

The Poisson Regression model is apt for this kind of data since we are dealing with a count (number of apprentices) which follows a discrete distribution. Given counts cannot be negative and are whole numbers, Poisson regression is a natural choice. With Poisson regression, we can model how the different factors (distance, population, degree of urbanization, and direction from Edinburgh) influence the number of apprentices migrating to Edinburgh.

Distance is expected to have a negative effect (greater distance, less migration); population and degree of urbanization might have a positive effect (larger values indicate more potential migrants); direction might have a mixed effect depending on geographical and socio-economic factors.

Where

: This represents the observed count (e.g., number of apprentices) for the i-th observation comes from a Poisson distribution with mean .

: The logarithm of the expected count is being modeled as a linear function of the predictors.

: This is the intercept for the -th group, where gi can be 1, 2, or 3. This allows for a different baseline count of apprentices for each group.

: These are the regression coefficients associated with the predictors "Distance", "Population (1000s)", and "Degree of Urbanization", respectively. They quantify the effect of each predictor on the logarithm of the expected count of apprentices.

The prior distributions would be non-informative to let the data inform the model. Flat or slightly regularizing priors could be used.

### Results

For all parameters ([1], [2], [3], 1, 2, 3), the point estimates of the PSRFs are very close to 1, which is a good sign. A value of 1.01, as shown in your output, is a positive indication that the chains have converged when considering all parameters simultaneously. In the autocorrelation table the values are close to zero for all parameters, indicating that samples separated by 50 steps are nearly independent. The effective sample sizes provide insights into the efficiency of the MCMC sampling process. The number of iterations for the model was 7 000 for will be run.

#### Interpretation

Intercepts For a direction from Edinburgh 1 (North), 2 (West) and 3 (South), with all other predictors being zero, we expect 69.74, 87.65, 211.89 apprentices to migrate to Edinburgh, respectively.

Distance : This suggests a decrease of about 3.35% in the expected count of migrating apprentices for each unit increase in distance, keeping other factors constant.

Population: This suggests an increase of about 2.16% in the expected count of migrating apprentices for each unit increase in population (1000s).

Population: This suggests a decrease of about 3.53% in the expected count of migrating apprentices for each unit increase in urbanization.

## Conclusions

The model suggests that the expected number of migrating apprentices decreases as the distance from Edinburgh increases and as the degree of urbanization increases. On the other hand, it increases with the population of the region.

The group to which a region belongs (as indicated by the intercepts) also plays a significant role in influencing the expected count of migrating apprentices, with group 3 regions having the highest base rate of migration when other factors are zero.