Effect of Literacy and Age at Marriage on Family Size in Portugal

Heidi Wang AKA Wu Zetian John Zhang

February 09, 2025

Contents

1 Introduction **2**

2 Method **2**

3 Result **2**

3.1 Statistical Summary 2

3.2 Model Selection 4

4 Conclusion **5**

1 Introduction

Family size is a critical determinant of economic stability, social structures, and gender equality, making it essential to understand the factors that influence fertility trends. Among these factors, literacy levels and the age at marriage have been consistently identified as key drivers of family size. Research indicates that higher literacy rates, particularly among women, are associated with smaller families, as education enhances access to family planning resources and expands economic opportunities (Götmark & Andersson, 2020). Similarly, marrying at a younger age tends to extend reproductive years, often resulting in larger families (Dessen & Torres, 2019). Since Portugal, Mexico, and Brazil had similar GDP per capita in 1980, existing studies on fertility trends in these countries provide valuable insights into how literacy and marriage timing may have influenced family size in Portugal.

Studies in Portugal have demonstrated a strong connection between education, marriage age, and fertility rates. Arqueiro et al. (2016) found that rising female literacy, economic development, and improved access to contraception have contributed to declining fertility rates. Their research showed that the average age at first childbirth in Portugal increased from 25 years in 1960 to 29.5 years in 2012, reinforcing the role of delayed marriage in reducing family size. These findings align with broader global patterns. In Brazil, Dessen and Torres (2019) examined how traditional family structures historically encouraged early marriage and larger families, but economic transformations and increased female workforce participation have contributed to a steady decline in birth rates. Similarly, Götmark and Andersson (2020) found that higher education levels are consistently associated with delayed marriage and lower fertility rates across multiple regions, emphasizing education's crucial role in shaping fertility outcomes. Given the economic similarities between Portugal, Brazil, and Mexico in 1980, these studies offer meaningful context for evaluating how literacy and marriage age influenced fertility patterns in Portugal.

This study seeks to investigate the relationship between literacy, age at marriage, and family size in Portugal, employing generalized linear models (GLMs) to quantify these effects. By analyzing birth rate variations while accounting for explanatory variables—including regional differences based on population size—this research aims to provide robust statistical evidence on how literacy and marriage timing influence fertility rates. The findings will contribute to ongoing discussions on family planning, gender equality, and demographic change in Portugal, offering insights that can inform policy development in education, economic growth, and social welfare. Through this analysis, the study will deepen our understanding of the interplay between education, marriage dynamics, and fertility, helping to identify key factors shaping family structures in Portugal.

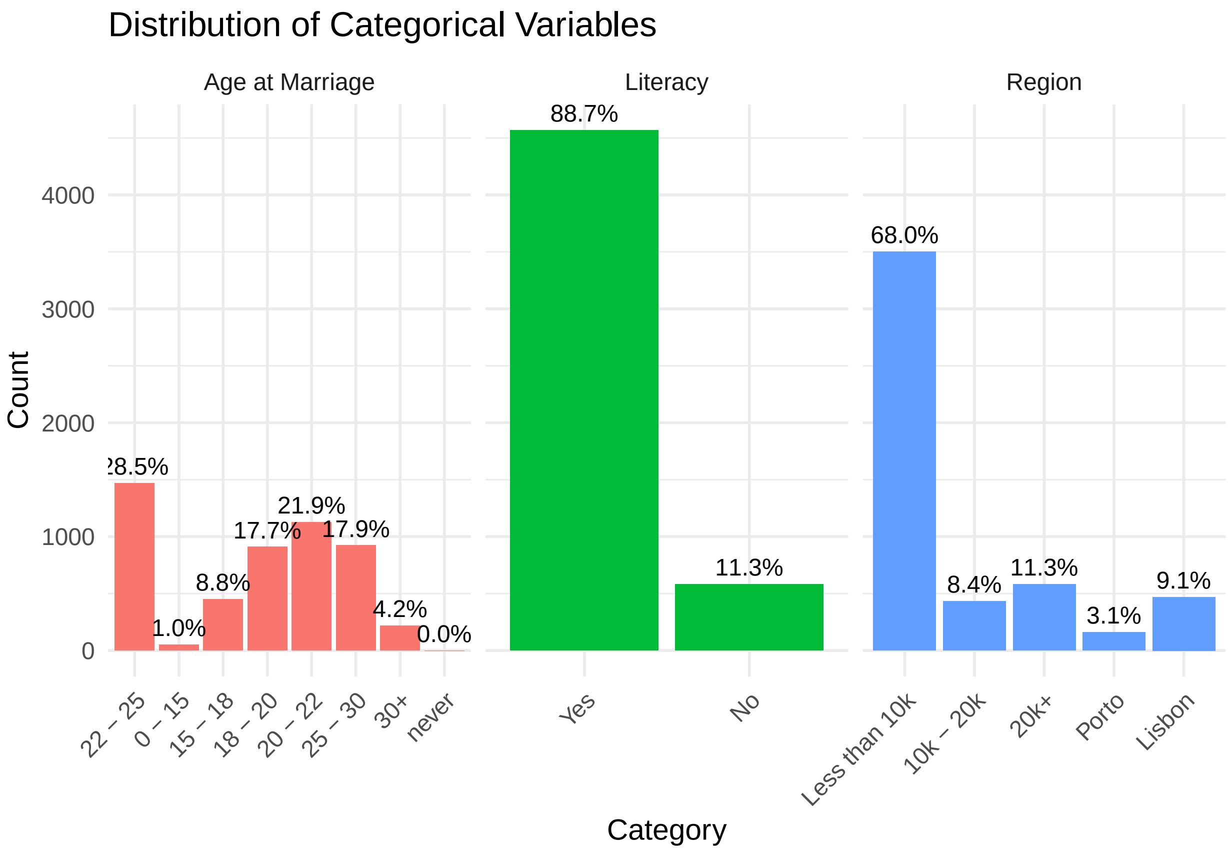
2 Method

We use a Generalized Linear Model (GLM) with a Poisson distribution to model the number of children per family, accounting for potential overdispersion using a quasi-Poisson or negative binomial approach.

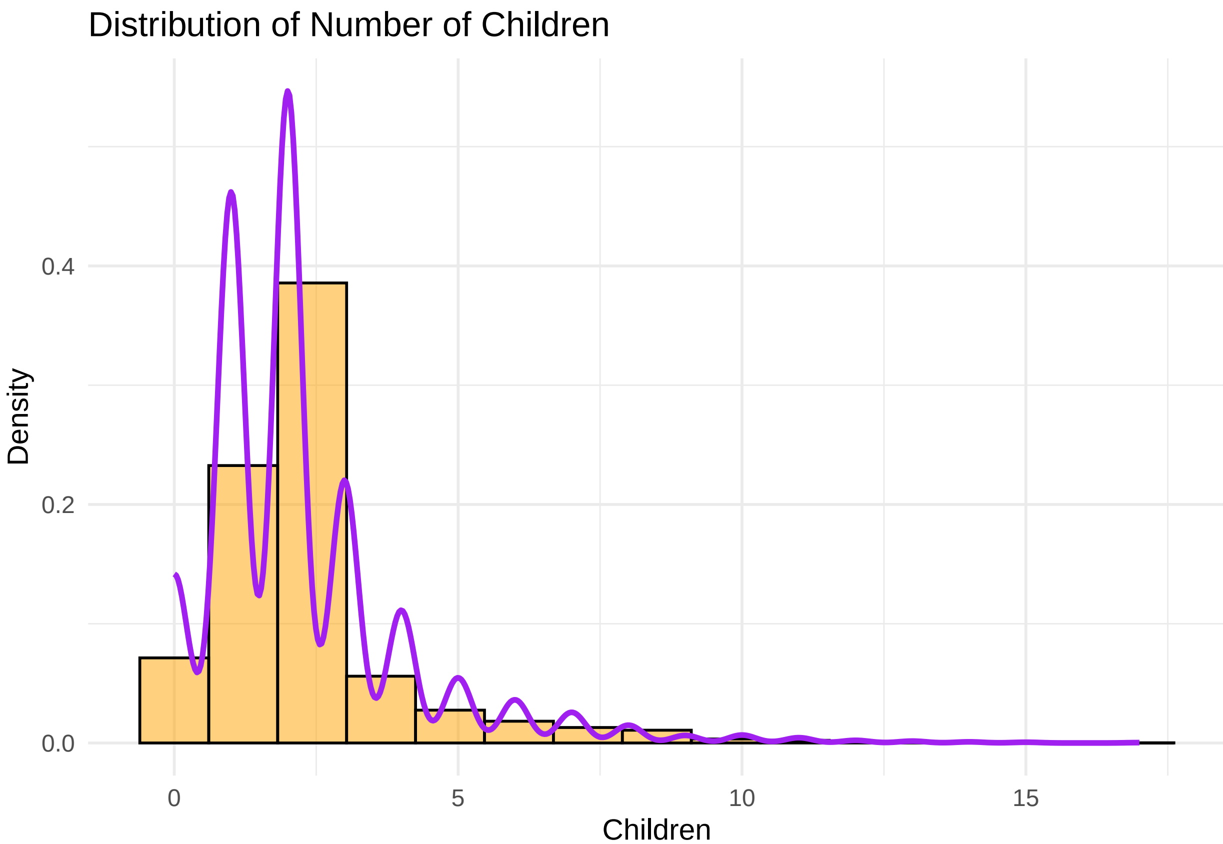
3 Result

3.1 Statistical Summary

The dataset analyzed in this study includes key demographic and fertility-related variables, with the number of children per woman serving as the response variable. The predictor variables considered include literacy, age at marriage, and region. Figure 1 illustrates the distribution of these categorical variables, showing that the majority of women were literate (88.7%), while a smaller proportion (11.3%) were illiterate. Regarding age at marriage, the most common category was 22-25 years (28.5%), followed by 20-22 years (21.9%) and 25-30 years (17.9%), while early marriage before 15 years was rare (1.0%). The regional distribution indicates that most respondents lived in rural areas with populations below 10,000 (68.0%), while smaller proportions resided in Lisbon (9.1%) and Porto (3.1%).



The response variable, the number of children per woman, exhibits a right-skewed distribution (Figure 2). While most women have between 0 and 5 children, a small proportion has significantly larger families. The mean number of children per woman is 2.26, while the variance is 3.46, suggesting greater variability than expected under a Poisson distribution, which assumes an equal mean and variance. This initial observation indicates potential overdispersion, necessitating further examination of model assumptions.



3.2 Model Selection

To investigate the impact of literacy, age at marriage, and region on fertility, a Poisson regression model was first estimated using log-transformed years since marriage as an offset (Table 1). This offset accounts for differences in marital duration, ensuring meaningful comparisons across individuals. The Poisson model results indicate that literacy is associated with a small but statistically significant increase in fertility, with illiterate women having an estimated 12.9% higher fertility rate than literate women ($p < 0.001$). Additionally, women who married between 15-18 years had significantly higher fertility rates than those who married between 18-20 years (reference category), supporting existing literature linking early marriage to increased childbearing. Regional effects were also observed, with women residing in urban areas (Lisbon, Porto, and regions with populations over 20,000) having significantly lower fertility rates compared to those in rural regions (populations below 10,000), aligning with demographic trends.

Table 1: Model Coefficients from Poisson Regression with Offset (Log Years Married)

	Estimate	Std_Error	p_Value
(Intercept)	-1.728	0.019	0.000
literacyNo	0.129	0.024	0.000
ageMarried0 - 15	0.040	0.080	0.619
ageMarried15 - 18	0.075	0.034	0.027
ageMarried18 - 20	0.060	0.028	0.031
ageMarried20 - 22	0.024	0.026	0.367
ageMarried25 - 30	0.019	0.029	0.516
ageMarried30+	0.029	0.059	0.621
region10k - 20k	-0.162	0.036	0.000

	Estimate	Std_Error	p_Value
region20k+	-0.294	0.033	0.000
regionPorto	-0.094	0.053	0.079
regionLisbon	-0.278	0.037	0.000

To assess whether region is a necessary predictor, a likelihood ratio test was conducted by comparing the full Poisson model (which included region) with a reduced model that excluded region. The results, presented in Table 2, indicate a highly significant chi-square test statistic (141.675, $p < 0.001$). Since the p-value is below the 0.05 significance level, we reject the null hypothesis that region has no effect on fertility. This finding supports the inclusion of region as an essential predictor, consistent with prior demographic research emphasizing the role of urbanization in fertility patterns.

Table 2: Likelihood Ratio Test: Evaluating the Importance of Region

Model	Log Likelihood	Degrees of Freedom	Chi Square	p-value
Full Model (With Region)	-8944.741	NA	NA	NA
Reduced Model (Without Region)	-9015.579	-4	141.675	0.000

Although the Poisson model provides valuable insights, evidence of overdispersion in the response variable suggests that the Poisson assumption of equal mean and variance is violated. This overdispersion was further confirmed by the dispersion parameter estimate (0.245), which indicates greater variability than the Poisson model can accommodate. As a result, a negative binomial model was estimated to account for excess variability (Table 3). The negative binomial model confirmed the key findings of the Poisson model, with the effects of literacy, age at marriage, and region remaining statistically significant. This suggests that our results are robust across model specifications.

Table 3: Model Coefficients from Negative Binomial Regression with Offset

	Estimate	2.5 %	97.5 %
(Intercept)	-1.712	-1.753	-1.671
literacyNo	0.119	0.067	0.171
ageMarried0 - 15	0.056	-0.118	0.230
ageMarried15 - 18	0.083	0.010	0.156
ageMarried18 - 20	0.065	0.006	0.125
ageMarried20 - 22	0.029	-0.027	0.086
ageMarried25 - 30	0.016	-0.045	0.077
ageMarried30+	0.028	-0.094	0.150
region10k - 20k	-0.161	-0.237	-0.086
region20k+	-0.294	-0.363	-0.226
regionPorto	-0.097	-0.211	0.017
regionLisbon	-0.276	-0.354	-0.199
sd	0.245	0.214	0.279

4 Conclusion

The negative binomial regression results (Table 3) reveal that literacy, age at marriage, and regional factors significantly influence fertility rates in Portugal. Exponentiating the coefficients provides a clearer interpretation in terms of relative fertility rates. Women who are illiterate have a **12.6% higher fertility**

rate compared to literate women ($e^{0.119} = 1.126$, 95% CI: 1.069, 1.187). This finding is consistent with previous research linking lower educational attainment to higher fertility, as education enhances family planning awareness and economic opportunities, reducing the number of children born.

Age at marriage also plays a significant role in determining fertility levels. Compared to women who married between 18-20 years (the reference category), those who married between **15-18 years** have **8.7% higher fertility rates** ($e^{0.083} = 1.086$, 95% CI: 1.010, 1.169), supporting the idea that earlier marriage increases the reproductive window. In contrast, women who married at older ages (22-25, 25-30, and 30+) do not show significant differences in fertility, suggesting that the impact of delayed marriage on childbearing diminishes after a certain age threshold.

The importance of region in predicting fertility was confirmed by the likelihood ratio test (Table 2), where excluding regional factors significantly worsened model fit ($\chi^2 = 141.675$, $p < 0.001$). Women living in urbanized areas, particularly in Lisbon and regions with populations over 20,000, have substantially lower fertility rates compared to those in rural areas with fewer than 10,000 people. Specifically, fertility rates are **13.7% lower** for women in mid-sized regions (10k - 20k, $e^{-0.161} = 0.851$, 95% CI: 0.789, 0.918)), **25.5% lower** for women in larger regions (20k+, $e^{-0.294} = 0.745$, 95% CI: 0.696, 0.798)), and **24.1% lower** for women in Lisbon ($e^{-0.276} = 0.759$, 95% CI: 0.701, 0.819)). These results align with demographic transition theory, where fertility declines with increasing urbanization due to enhanced educational, economic, and contraceptive access.

Finally, the estimated dispersion parameter (0.245) confirmed that overdispersion was present in the data, validating the choice of a negative binomial model over a Poisson regression, which assumes equal mean and variance. These findings underscore the importance of addressing educational disparities and regional fertility patterns when designing policy interventions. Expanding access to education and reproductive health services, especially in rural areas, could contribute to moderating fertility rates and improving overall family well-being.