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

Heidi Wang John Zhang

February 10, 2025

Contents

1 Introduction

Family size plays a crucial role in economic stability, social structures, and gender equality, making it important to understand the factors influencing fertility trends. Among these, literacy levels and age at marriage have been consistently identified as key drivers. Research indicates that higher literacy rates, particularly among women, are linked to smaller families, as education improves access to family planning and economic opportunities (Götmark & Andersson, 2020). Conversely, early marriage extends reproductive years, often resulting in larger families (Dessen & Torres, 2019). Given that Portugal, Mexico, and Brazil had similar GDP per capita in 1980, studies on fertility trends in these countries provide valuable context for understanding how literacy and marriage timing influenced family size in Portugal.

Research has shown that education, marriage age, and fertility rates are interconnected. In Portugal, increased female literacy, economic growth, and better access to contraception have led to declining fertility rates, with the average age at first childbirth rising from 25 years in 1960 to 29.5 years in 2012 (Arqueiro et al., 2016). Similar trends have been observed in Brazil, where early marriage and large families were once common due to traditional family structures, but economic shifts and higher female workforce participation have contributed to a steady decline in birth rates (Dessen & Torres, 2019). Additionally, whether individuals live in rural or urban areas also plays a role, as fertility rates tend to be higher in less populated regions with stronger traditional family norms and lower access to education and healthcare (Dessen & Torres, 2019). These patterns align with global findings, as higher education levels are consistently associated with delayed marriage and lower fertility rates across multiple regions in Mexico(Götmark & Andersson, 2020). Given the economic similarities between Portugal, Brazil, and Mexico in 1980, these studies provide important insights into how literacy and marriage timing influenced Portugal's fertility trends.

This study examines how literacy, age at marriage, and regional differences influence family size in Portugal using generalized linear models (GLMs). By analyzing variations in family size while accounting for these key explanatory variables, the research aims to provide robust statistical evidence on their impact. The findings contribute to discussions on family planning, gender equality, and demographic trends in Portugal, offering valuable insights to inform policy development in education, economic growth, and social welfare.

2 Method

This study employs a generalized linear model (GLM) with a Poisson distribution to examine the relationship between literacy, age at marriage, and family size in Portugal. Since family size is a count variable representing the number of children per family, the Poisson model is well-suited for non-negative integer-valued data. The predictor variables in the model include age at marriage (categorical), literacy (categorical), and region (categorical). The inclusion of region as a predictor is supported by peer-reviewed literature, which suggests that geographic differences may influence fertility rates. A fundamental assumption of the Poisson model is that the mean and variance of the response variable are equal, which will be evaluated through exploratory data analysis (EDA).

To assess the suitability of the Poisson model, the empirical distribution of family size will be plotted and mean-variance comparisons will be performed to check for over-dispersion, which occurs when the variance exceeds the mean, violating Poisson assumptions. Additionally, a likelihood ratio test (LRT) will be conducted to determine whether region is a necessary predictor. This will be achieved by comparing a full Poisson model (which includes region) with a reduced model that excludes it. A statistically significant result will indicate that region significantly contributes to explaining variation in family size, justifying its inclusion in the final model.

If over-dispersion is detected, negative binomial (NB) GLM will be considered, as it introduces an additional dispersion parameter to account for excess variability. Once the most appropriate model is selected, the estimated coefficients will be interpreted directly to assess the effects of literacy, marriage age, and region on family size. This approach ensures that the final model provides a robust and meaningful statistical analysis of fertility patterns in Portugal.

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(the category with most proportion is placed leftmost), 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%).

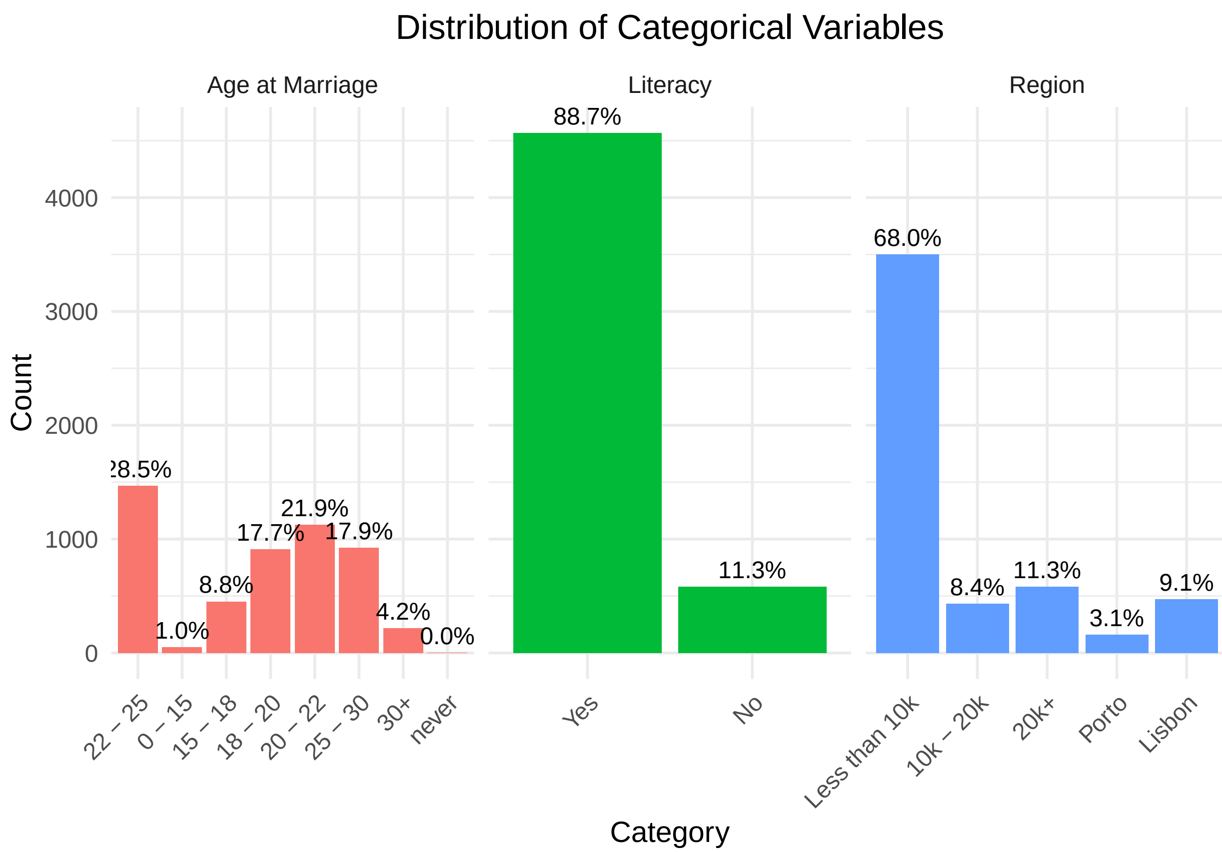

Figure 1: Most women married at 22-25 years (28.5%), were literate (88.7%), and lived in rural areas (68.0%)

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

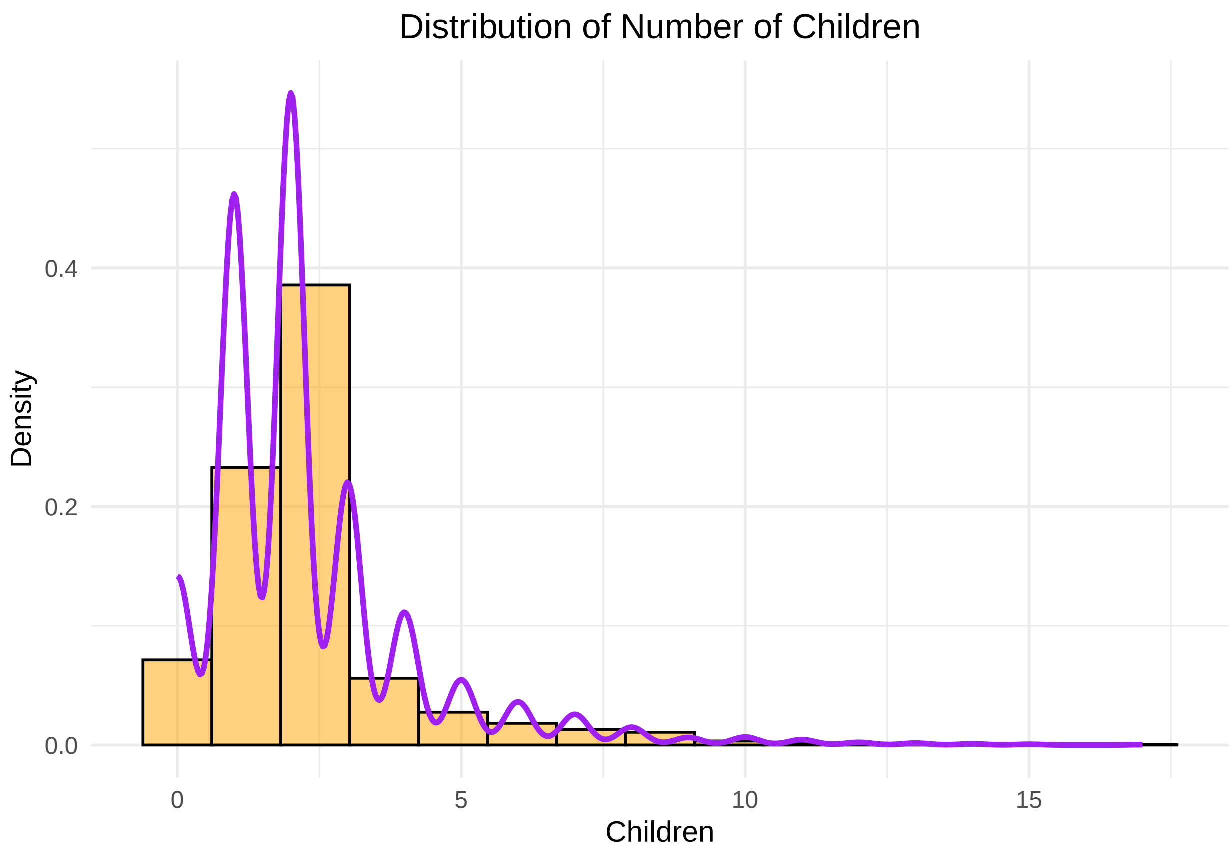

Figure 2: The histogram shows a right-skewed distribution of children per woman, with most having 0 to 5 children. The density plot suggests overdispersion, indicating a potential need for a negative binomial model

women having an 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), women who married at age 15-18 tend to have higher fertility than those who married at age 22-25, while those unliterated tend to have higher fertility than those literated

	Estimate	Standard 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
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 as suggested by the literature, 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, where region is significant in the model

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(Log Years Married), the results are similar with previous poisson model where women who married at age 15-18 tend to have higher fertility than those who married at age 22-25, while those unliterated tend to have higher fertility than those literated.

	Estimate	Standard Error	p-Value	2.5%	97.5%
(Intercept)	-1.712	0.021	0.000	-1.753	-1.671
literacyNo	0.119	0.026	0.000	0.067	0.171
ageMarried0 - 15	0.056	0.089	0.528	-0.118	0.230
ageMarried15 - 18	0.083	0.037	0.027	0.010	0.156
ageMarried18 - 20	0.065	0.030	0.031	0.006	0.125
ageMarried20 - 22	0.029	0.029	0.306	-0.027	0.086
ageMarried25 - 30	0.016	0.031	0.600	-0.045	0.077
ageMarried30+	0.028	0.062	0.650	-0.094	0.150
region10k - 20k	-0.161	0.038	0.000	-0.237	-0.086
region20k+	-0.294	0.035	0.000	-0.363	-0.226
regionPorto	-0.097	0.058	0.095	-0.211	0.017
regionLisbon	-0.276	0.040	0.000	-0.354	-0.199
sd	0.245	NA	NA	0.214	0.279

4 Conclusion

The negative binomial regression results (Table 3) confirm 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 aligns with prior studies indicating that lower educational attainment is associated with higher fertility due to reduced access to family planning and economic opportunities (Götmark & Andersson, 2020). Similarly, age at marriage plays a crucial role in fertility outcomes, with women who married between **15-18 years** exhibiting an **8.6% higher fertility rate** ($e^{0.083} = 1.086$, 95% CI: 1.010, 1.169)). This result supports earlier research showing that early marriage extends the reproductive window and leads to higher fertility (Dessen & Torres, 2019). In contrast, the fertility rates for women who married at older ages (25-30 and 30+) do not differ significantly from the reference category (22-25 years), suggesting that the impact of delayed marriage on childbearing diminishes after a certain threshold.

The impact of regional differences on fertility was confirmed by the likelihood ratio test (Table 2), where excluding regional variables significantly worsened model fit ($\chi^2 = 141.675$, $p < 0.001$). Fertility rates were found to be **14.9% 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 findings align with demographic transition theory, which posits that urbanization leads to lower fertility due to increased access to education, economic opportunities, and contraception (Arqueiro et al., 2016). This is consistent with prior research in Brazil and Portugal, where economic development and increased female workforce participation have contributed to declining fertility rates (Dessen & Torres, 2019).

These results reinforce the importance of education and regional socio-economic factors in shaping fertility patterns in Portugal. The findings highlight that policies aimed at increasing female literacy and delaying marriage could help moderate fertility rates. Furthermore, regional disparities suggest that improving access to reproductive health services in rural areas could contribute to more balanced fertility outcomes. Given Portugal's historical similarities with Brazil and Mexico in terms of economic development, these insights offer valuable context for policymakers addressing demographic changes and family planning strategies.

5 Reference

Arqueiro, A., Cunha, J., Garbe, C., Lafontaine, D., Dionisio, M. D. L., Shiel, G., ... & Géron, S. (2016). Literacy in Portugal. Country Report. Children and adolescents.

Dessen, M. A., & Torres, C. V. (2019). Family and socialization factors in Brazil: An overview. Online readings in psychology and culture, 6(3), 2.

Götmark, F., & Andersson, M. (2020). Human fertility in relation to education, economy, religion, contraception, and family planning programs. BMC Public Health, 20, 1-17.