

# An Analysis of effect of literacy and age of a marriage on family size using Portual data

Kyungrok Park

Yushen Chen

Feb 4, 2025

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methods</b>	<b>2</b>
<b>3</b>	<b>Results</b>	<b>2</b>
<b>4</b>	<b>Conclusion</b>	<b>3</b>
<b>5</b>	<b>List of Figures and Tables</b>	<b>4</b>
<b>6</b>	<b>References</b>	<b>8</b>

## 1 Introduction

Fertility trends in Portugal have been closely linked to demographic transitions (Pimentel & Sousa Gomes, 2022). With an aging population, the Portuguese government has become increasingly concerned about low fertility rates (Pimentel & Sousa Gomes, 2022). Since the 1970s, birth rates in Portugal have sharply declined, and the COVID-19 crisis further accelerated this decline, leading to a historic low in fertility levels. Given these concerns, understanding family size and the factors influencing it is crucial for addressing population aging. This study analyzes data from a 1979 fertility survey in Portugal, covering the national population (excluding Azores and Madeira) and focusing on ever-married women aged 15 to 49.

Several studies highlight the role of education and marriage timing in fertility rates. Kebede et al. (2019) found that in sub-Saharan Africa, fertility rates declined until 2000, after which disruptions in women's education slowed the decline. Similarly, Ekoriano et al. (2023) observed that in Indonesia, women who married before 25 had longer reproductive periods and higher fertility rates than those who married later. Likewise, Ariho & Kabagenyi (2020) found that raising the age of first marriage significantly reduces fertility rates in Uganda.

The fertility data from Indonesia and Uganda suggest that delaying marriage reduces family size, reinforcing the need to examine literacy levels and marriage age in Portugal. Higher illiteracy rates are often associated with higher fertility, and early marriage tends to result in larger families. Therefore, this study focuses on how literacy and marriage timing influence family size to provide insights into Portugal's fertility trends.

Therefore, this research aims to examine how literacy and age at marriage influence family size in Portugal by using the Poisson regression and negative binomial regression. Additionally, it will explore the other variations, such as the impact of the number of sons and regional differences on family size.

## 2 Methods

This study examines the effects of literacy and age at marriage on family size using Poisson and Negative Binomial regression models. Family size, as a count variable accumulating over a woman’s reproductive lifetime, makes Poisson regression a natural choice.

The key explanatory variables include literacy (binary) and age at marriage (categorical), as they are the primary focus. Additionally, given that fertility tends to be lower in urban areas (Riederer & Beaujouan, 2023), regional population size is included to differentiate between urban and rural areas. Furthermore, while son preference is observed in some societies (Park, 1983), its presence in Portugal remains uncertain. To investigate this, the proportion of sons among children is incorporated to examine its potential effect on family size.

To refine the model, Poisson regression with an offset (months since marriage) is introduced, allowing birth rates to be estimated instead of total children. However, since overdispersion is likely, a Negative Binomial regression with an offset is also considered. Consequently, three models are analyzed: Poisson regression without an offset, Poisson regression with an offset, and Negative Binomial regression with an offset.

The significance of predictors will be assessed through confidence intervals and p-values, while the precision parameter ( $\tau$ ) will be used to measure overdispersion and unexplained variation in the Poisson model. These analyses will help determine the most appropriate model for understanding how literacy and marriage timing influence family size in Portugal.

## 3 Results

To justify the appropriate regression model for this analysis, the nature of the variables is examined. Figure 1 illustrates that the distribution of family size (number of children) across literacy and age at marriage categories is right-skewed, supporting the use of a Poisson model. The differing means between literacy groups and the greater density of family sizes beyond 3-4 children for illiterate females suggest a potential effect of literacy on fertility. Similarly, the variation in means across age at marriage categories within the same literacy status implies a possible impact of age at marriage on family size.

Given that family size accumulates over time, years since marriage is included as an offset to account for differences in exposure time, ensuring that the model reflects birth rates rather than total number of children. Additionally, Table 1 confirms overdispersion, as the variance-mean ratio (VMR) exceeds 1 for most groups, except for literate and illiterate females who married at ages 0-15. Since the Poisson model assumes equal mean and variance, overdispersion suggests that a Negative Binomial model is a better fit.

Figure 2 further supports this conclusion by showing that the Poisson model without an offset produces substantially different estimates compared to the Poisson and Negative Binomial models with offsets. Additionally, confidence intervals (95%) reveal significant differences between the Poisson model without an offset and the two models that include offsets. Notably, most age categories in the Poisson and Negative Binomial models with offsets include zero in their confidence intervals, indicating a lack of statistical significance. Furthermore, the estimates exhibit a consistent trend—they decrease from the Poisson model without an offset to the Poisson model with an offset, then increase under the Negative Binomial model, confirming that the extra variation in birth rates is accounted for. Since the proportion of sons is not significant in all three models, it is excluded from the final regression.

To finalize model selection, dispersion and log-likelihood values are examined. The precision parameter ( $\tau$ ) is estimated at 16.81, confirming high overdispersion, meaning that a Poisson model does not sufficiently explain variation in birth rates. Additionally, the log-likelihood values for the Negative Binomial model with an offset (-8909.044) are higher than those of the Poisson models (-9239.154, -8944.741), reinforcing the decision to adopt the Negative Binomial model with an offset.

The interaction term between age at marriage and literacy is excluded. Figure 3 reveals that the distribution of birth rates remains similar across literacy groups within each age category, with no substantial variation in

median, spread, or outliers. This indicates that the relationship between age at marriage and fertility does not vary significantly by literacy status, making an interaction term unnecessary.

Table 2 presents the final Negative Binomial model results. Literate females who married at 22-25 have an 11.2% lower birth rate than illiterate females in the same age group ( $IRR = 0.888$ ,  $z = -4.49$ ,  $p < 0.001$ ), with a 95% confidence interval (0.843, 0.935) confirming statistical significance. This finding suggests that literacy is linked to lower fertility when controlling for age at marriage.

Regarding age at marriage, results indicate that illiterate females who married at 15-18 and 18-20 have 8.6% and 6.7% higher birth rates than those who married at 22-25 ( $IRR = 1.086$ ,  $p = 0.027$  and  $IRR = 1.067$ ,  $p = 0.031$ , respectively). However, other age categories (0-15, 25-30, 30+) show no statistically significant impact on birth rates, suggesting that fertility differences at these ages may be due to random variation.

In terms of regional differences, illiterate females in Lisbon and regions with 20k+ population have 24.2% and 25.5% lower birth rates than their rural counterparts ( $IRR = 0.758$ ,  $z = -6.98$ ,  $p < 0.001$  and  $IRR = 0.745$ ,  $z = -8.40$ ,  $p < 0.001$ ). Women in regions with 10k-20k populations exhibit a 14.9% lower birth rate ( $IRR = 0.851$ ,  $z = -4.20$ ,  $p < 0.001$ ). However, the birth rate in Porto is 9.2% lower than in rural areas but is not statistically significant ( $IRR = 0.908$ ,  $p = 0.095$ ), indicating that this difference may be due to random variation.

## 4 Conclusion

With the justification of model choice, including the analysis of the Variance-Mean Ratio (VMR), precision parameter ( $\tau$ ) of the Gamma distribution, and log-likelihood comparison, the Negative Binomial model with an offset is selected for this study. The results suggest that literacy status significantly influences birth rates, as illiterate females who married at 22-25 have an 11.2% lower birth rate than illiterate females in the same age group. Furthermore, illiterate females who married at 15-18 and 18-20 exhibit 8.6% and 6.7% higher birth rates, respectively, compared to those who married at 22-25, suggesting that early marriage is associated with higher fertility. However, no significant effects are observed for other age categories (0-15, 25-30, 30+), implying that birth rate variation based on age at marriage may be due to random effects. This challenges previous findings by Kebede et al. (2019) and Ekoriano et al. (2023), which suggested that delaying marriage reduces fertility. In terms of regional differences, illiterate females in Lisbon and regions with 20K+ population exhibit 24.2% and 25.5% lower birth rates, respectively, compared to other areas. However, birth rates in Porto and regions with less than 20K population show no significant differences, indicating that regional birth rate variations may be random rather than systematic. While Riederer & Beaujouan (2023) found lower birth rates in urban areas, our study does not confirm a systematic effect of urbanization or population size. Overall, despite the dataset’s regional (Portugal) and temporal (1980s) context, literacy emerges as a stronger predictor of fertility than age at marriage or urbanization, with illiterate females generally exhibiting higher birth rates.

## 5 List of Figures and Tables

Figure 1: Density of Number of Children by Age at Marriage and Literacy

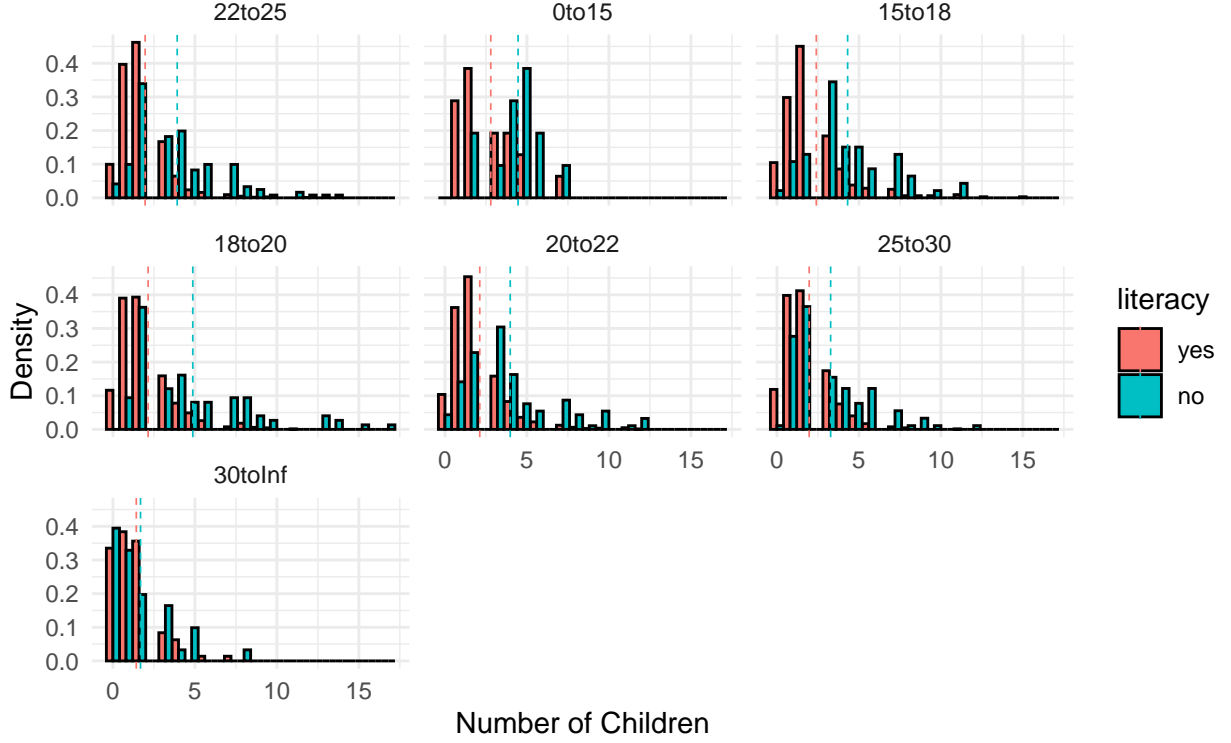


Figure 1: A Density plot which shows the distribution of Family size based on the Literacy Status and Age of Marriage. Notably, the density of family size for each plot has approximately a right-skewed distribution. Furthermore, for each density plot, the density of of number of children of illeracy females become greater than the number of children of literate females after a family size of 3-4, and it suggests a possible effect of literacy status on family size. Additionally, the mean of family sizes for different age at marriage and same literacy status (dotted line with same colour) shows fairly different values, and it also indicates that the age at marrage of females could be a possible factor for different family sizes.

Table 1: Mean and Variance of Family Size by Literacy Status and Age at Marriage

literacy	ageMarried	Mean Family Size	Variance of Family Size	Variance-to-Mean Ratio (VMR)
no	22to25	3.920530	7.046976	1.7974550
no	0to15	4.461538	2.269231	0.5086207
no	15to18	4.310345	6.217786	1.4425263
no	18to20	4.870968	12.983170	2.6654189
no	20to22	3.982609	7.911976	1.9866314
no	25to30	3.274336	5.486568	1.6756274
no	30toInf	1.684211	3.411095	2.0253378
yes	22to25	1.965831	1.976795	1.0055771
yes	0to15	2.794872	2.641026	0.9449541
yes	15to18	2.398477	4.102898	1.7106264
yes	18to20	2.148103	2.984166	1.3892101
yes	20to22	2.122651	2.676031	1.2607024

yes	25to30	1.966667	1.975402	1.0044416
yes	30toInf	1.424581	1.785010	1.2530073

*Note:* The table illustrates the mean, variance and variance-mean ratio(VMR) of family size for corresponding category of literacy status and age at marriage. Besides two categories (literate and illiterate females with age at marriage of 0 to 15), the VMR of family exceeds 1, and it suggests a potential overdispersion of family size when we account for two main explanatory variables, literacy status and age at marriage.

Table 2: Exponentiated Model Coefficients (Incident Rate Ratios)

Predictor	IRR (Exp Coeff)	Std. Error	z-Statistic	p-Value	2.5% CI (IRR)	97.5% CI (IRR)
<b>(Intercept)</b>	0.2032976	0.0293841	-54.2158740	0.0000000	0.1919201	0.2153496
<b>literacyyes</b>	0.8880462	0.0264200	-4.4939994	0.0000070	0.8432315	0.9352427
<b>ageMarried0to15</b>	1.0576321	0.0888727	0.6304801	0.5283805	0.8885584	1.2588768
<b>ageMarried15to18</b>	1.0862765	0.0373454	2.2159555	0.0266946	1.0096060	1.1687694
<b>ageMarried18to20</b>	1.0674562	0.0303022	2.1542475	0.0312208	1.0059046	1.1327741
<b>ageMarried20to22</b>	1.0299062	0.0287770	1.0240022	0.3058342	0.9734253	1.0896642
<b>ageMarried25to30</b>	1.0164169	0.0310176	0.5249807	0.5995966	0.9564663	1.0801252
<b>ageMarried30toInf</b>	1.0286729	0.0623127	0.4536722	0.6500648	0.9104093	1.1622992
<b>regionlisbon</b>	0.7584510	0.0396006	-6.9816328	0.0000000	0.7018097	0.8196636
<b>regionporto</b>	0.9077341	0.0580395	-1.6678953	0.0953365	0.8101313	1.0170960
<b>region20k+</b>	0.7449635	0.0350394	-8.4025461	0.0000000	0.6955197	0.7979223
<b>region10-20k</b>	0.8509630	0.0383938	-4.2034577	0.0000263	0.7892777	0.9174693

*Note:* The table demonstrates coefficient estimate, standard errors, z-statistics, p-values and 95% confidence interval of our final model (negative binomial with an offset) for each categorical variable on birth rate with a reference variable as illiterate females who married between 22 to 25. Since the coefficient estimates are exponentiated, they can be interpreted as the birth rate increases/decreases by Beta if it belongs to a certain category compared to illiterate females with age of marriage of 22-25. For example, for the coefficient of age at marriage: 18-20, it can be interpreted as illiterate female with age at marriage between 18-20 have higher birth rate than illiterate females with age at marriage between 22-25. Also, the Z-statistic is used for corresponding p-values and CIs since the dataset contains more than 5000 datas

**Figure 2: Comparison of Regression Models**

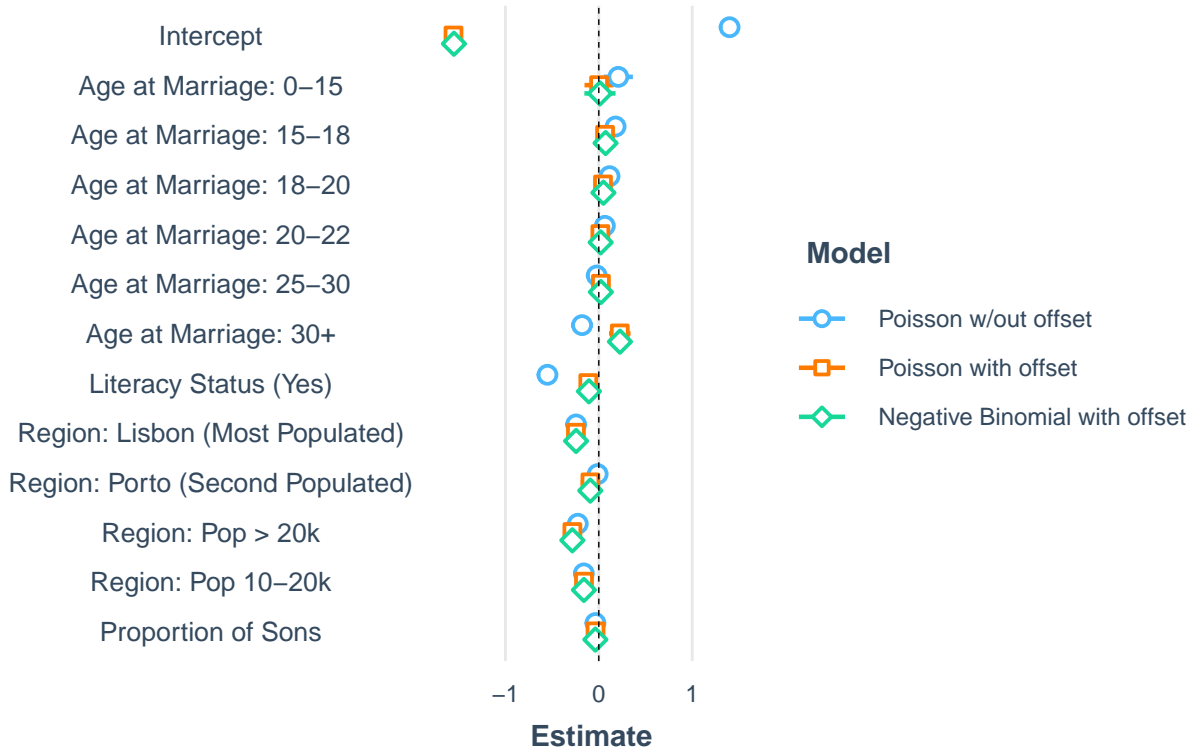


Figure 2: The figure displays the estimates with 95% confidence interval for each corresponding variable of three different models (Poisson with the offset, Poisson without the offset and Negative Binomial with the offset). Notably, most of coefficients estimates for age of marriage become statistically insignificant as the negative binomial model is applied, and the coefficient estimates for the proportion of sons in family are not statistically significant in all three models. Also, after excluding the proportion of sons in family from each model and examining the log-likelihood estimates as -9239.154, -8944.741 and -8909.044 for Poisson without the offset, Poisson with the offset and Negative Binomial with the offset respectively, it highlights that the negative binomial with the offset is the best fit model for our response variable (birth rate) and explanatory variables (literacy status, age of marriage and region)

**Figure 3: Distribution of Birth Rate by Literacy and Age at Marriage**

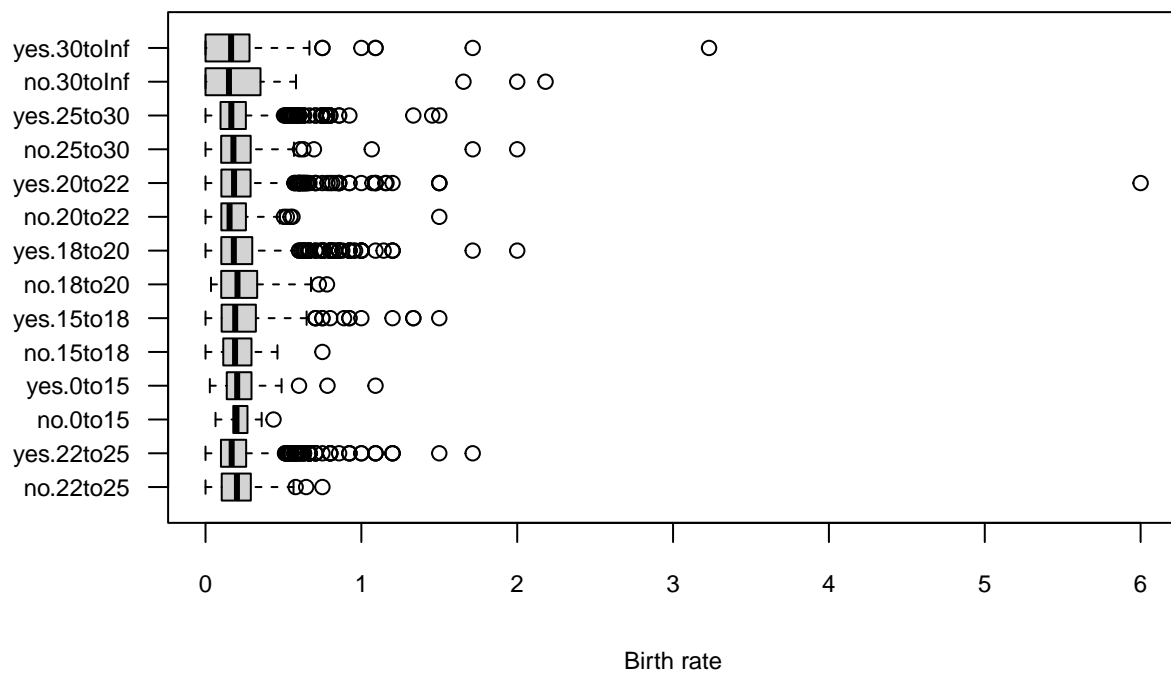


Figure 3: The figure illustrates the distribution of birth rates based on literacy status and age at marriage. As observed, the median, spread and outlier for different literacy status of same age at marriage are approximately same, and it indicates that there is no potential interaction or effect of literacy status on birth rates for corresponding age at marriage.

## 6 References

- Ariho, P., & Kabagenyi, A. (2020). Age at first marriage, age at first sex, family size preferences, contraception and change in fertility among women in Uganda: Analysis of the 2006–2016 period. *BMC Women's Health*, 20(1). <https://doi.org/10.1186/s12905-020-0881-4> Ekoriano, M., Muthmainnah, M., Titisari, A., Devi, Y. P., Widodo, T., & Purwoko, E. (2023). The average age of first marriage for Indonesian women in their reproductive period who give birth to an average of two children: National survey (2017 – 2019). *F1000Research*, 12, 35. <https://doi.org/10.12688/f1000research.126816.1> Kebede, E., Goujon, A., & Lutz, W. (2019). Stalls in Africa's fertility decline partly result from disruptions in female education. *Proceedings of the National Academy of Sciences*, 116(8), 2891–2896. <https://doi.org/10.1073/pnas.1717288116>
- Pimentel, D., & Sousa Gomes, C. (2022). Beyond the crisis: Fertility variations and the family policies in the Portuguese municipalities. *Espace Populations Sociétés*, 2022/2–3. <https://doi.org/10.4000/eps.12990>
- Park, C. B. (1983). Preference for Sons, Family Size, and Sex Ratio: An Empirical Study in Korea. *Demography*, 20(3), 333–352. <https://doi.org/10.2307/2061246>
- Riederer, B., & Beaujouan, É. (2024). Explaining the urban–rural gradient in later fertility in Europe. *Population, Space and Place*, 30, e2720. <https://doi.org/10.1002/psp.2720>