Portugal Fertility Rate based on literacy rate and age

Shreya Sakura Noskor and Yuming Huang

2025-02-06

Abstract

Introduction

Global fertility rates have been declining, often linked to factors like education and access to healthcare. In particular, educated women tend to have fewer children and delay childbirth. In countries like Portugal, literacy and age at marriage can play significant roles in family size decisions. This report will explore how these factors influence family size in Portugal.

Of course, it is not exactly the number of children that directly affects the literacy rates of countries but usually it is the other way around: where the literacy rates affect the number of children a family has. This is a vast area of research so there's been a considerable amount of research that has already been done. For example, one group of researchers decided to explore the effects of literacy rates (among other things) on HRFB which is high-risk fertility behavior in women (such as having children too soon or too late and/or having too many children too close together) (Woldeamanuel et al., 2023). They found that "the women who had received no education or only primary education are more likely to experience high-risk fertility behavior than women with a higher level of education" meaning there is a direct correlation between literacy rates and the health/behavior of the mother (Woldeamanuel et al., 2023). Another group of researchers in 2013 took a look at the correlation between literacy rates and infant health and population stabilization. The main results from that report were that education allows women to spend more money on childcare, and encourages healthy decision-making roles about their and their child's nutrition and lifestyle. They also saw that female literacy was the deciding factor as opposed to male literacy rates (Saurabh et al., 2013). The last group of researchers took a look at the relationship between literacy rates and institutional delivery services, where they see if mothers safely give birth to children. Their key finding was also that there was a positive correlation where the more educated one was, the more well-versed they were in their delivery options, lowering the maternal and infant mortality rate (Khan et al., 2022). These studies show that literacy impacts maternal and child mortality rates, which directly correlates to family size.

With all of this information in mind, this report will further investigate this phenomenon by asking "How do literacy and age of a marriage affect family size?". We will create models to see which factors, like literacy, age, and region, affect family size in Portugal. Our report is structured as such: Methods section outlines the statistical procedure used to model the data to gather more insights into our research question, the Results section where we discuss the properties of data and the model to highlight the key areas, and lastly our conclusion section which will interpret the key areas that were talked about in the Results section. This will allow us to answer our research question as well as connect it to the real world.

Methods

The data that we use was extracted from data source and the code was cleaned by instructors in STA303, using R and its various libraries. From the data gathered there were 4 variables that we use to make a model that will answer our research questions.

Poisson Model

Because the value we are investigating is the size of the family and so the columns 'children' becomes our outcome variable. Because the value of that coloumn is a a continuous count variable, we chose a Poisson Model.

$$Y_i \sim \text{Poisson}(\mu_i)$$
$$\mu_i = \exp(\beta_0 + \beta_1 \cdot \text{logAgeMarried}_i + \beta_2 \cdot \text{ageMarried}_i + \beta_3 \cdot \text{literacy}_i)$$

Now we consider the fact that some people who have been married longer may have a bigger family size because they simply had more time then others (independent from other variables we are trying to investigate). Hence we are trying to account for that offset by accounting for the logarithm of the age married variable.

$$Y_i \sim \text{Poisson}(p_i)$$
$$log(p_i) = \beta_0 + \beta_1 \cdot \text{logAgeMarried}_i + \beta_2 \cdot \text{ageMarried}_i + \beta_3 \cdot \text{literacy}_i + log(O_i)$$
$$p_i = \mu_i O_i$$

Next we look at the mean and standard deviation of the data to see if there is an overdispersion. We account for this due to the fact that real world data may have more variance then we have expected and since of the rules of poisson distribution is that the mean and the standard deviation must be similar we investigate the data to see if this is the case.

We see from figure in Results that over dispersion is definitely an issue because of the difference in mean and variance. And hence we decide to use a Negative Binomial distribution instead to account for the overdispersion.

$$Y_i \sim \text{NegBin}(O_i\mu_i, \tau)$$
$$log(\mu_i) = \beta_0 + \beta_1 \cdot \text{logAgeMarried}_i + \beta_2 \cdot \text{ageMarried}_i + \beta_3 \cdot \text{literacy}_i$$

Results

(Intercept)	0.676
ageMarried0to15	0.352
ageMarried15to18	0.199
ageMarried18to20	0.089
ageMarried20to22	0.077
ageMarried25to30	0.000
ageMarried30toInf	-0.322
literacyno	0.690
ageMarried0to15:literacyno	-0.223
ageMarried15to18:literacyno	-0.104
ageMarried18to20:literacyno	0.128
ageMarried20to22:literacyno	-0.061
ageMarried25to30:literacyno	-0.181
ageMarried30toInf:literacyno	-0.523

This table shows the coefficients of the poisson model where the intercept is when they are literate, got married at age 20 to 22. The value for that is 0.753 so that means exp(0.753) = 2.12 children for the baseline family. Next we look at women that are not literate showing 0.629, which means women that are not literate will have exp(0.629)=1.87 more children then ones who are literate without taking age married into consideration. Taking a look at people who are married between 22 to 25 we see they have -0.077, which means they will have exp(-0.077) = 0.925 less child then the baseline group. And lastly looking at people

married from 25 to 30 with no literacy we see that their value is -0.119 which means they have exp(-0.119) = 0.887 less child then people who married at 20-22 years old that are literate.

```
##
##  22to25   0to15  15to18  18to20  20to22  25to30 30toInf   never
##    1468      52     452     910    1126     923     217       0
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.751	0.022	-81.100	0.000
literacyno	0.039	0.051	0.751	0.452
ageMarried22to25	-0.068	0.029	-2.329	0.020
ageMarried0to15	0.024	0.098	0.247	0.805
ageMarried15to18	0.053	0.039	1.366	0.172
ageMarried18to20	-0.005	0.032	-0.153	0.878
ageMarried25to30	-0.046	0.033	-1.384	0.166
ageMarried30toInf	-0.011	0.066	-0.159	0.873
literacyno:ageMarried22to25	0.213	0.069	3.092	0.002
literacyno:ageMarried0to15	0.047	0.170	0.277	0.782
literacyno:ageMarried15to18	-0.011	0.088	-0.120	0.904
literacyno:ageMarried18to20	0.197	0.074	2.671	0.008
literacyno:ageMarried25to30	0.162	0.077	2.098	0.036
literacyno:ageMarried30toInf	0.028	0.149	0.188	0.851

Now in this model above we take a look at the overdispersion case where we may have more variance in our data due to measurement and other circumstance outside of our data's predictors. Here we take log of marriage into consideration.The intercept again is a literate women who got married at 20 to 22 years old. They have around 0.17 children per year of marriage. Now looking at the same variables as last time non-literate people's value is 0.039 or exp(0.039) = 1.039. This means that non-literate 20-22 year olds will have 1.039 more child then literate people per year. Taking a look at people who are married between 22 to 25 we see they have -0.068, which means they will have exp(-0.068) = 0.934 less child per year than the baseline group. And lastly looking at people married from 25 to 30 with no literacy we see that their value is 0.162 which means they have exp(0.162) = 1.175 less child per year then people who married at 20-22 years old that are literate.

Literacy	Age Married	Region Type	Mean	Variance
yes	20to22	Rural	2.16	2.81
no	20to22	Rural	4.04	8.15
yes	22to25	Rural	1.99	2.05
no	22to25	Rural	4.03	7.28
yes	0to15	Rural	2.83	2.79
no	0to15	Rural	4.42	2.45
yes	15to18	Rural	2.41	4.21
no	15to18	Rural	4.14	6.57
yes	18to20	Rural	2.22	3.12
no	18to20	Rural	4.94	13.22
yes	25to30	Rural	2.02	2.05
no	25to30	Rural	3.26	5.59
yes	30toInf	Rural	1.49	1.89
no	30toInf	Rural	1.85	3.70
yes	20to22	Urban	1.84	1.66
no	20to22	Urban	3.14	4.14
yes	22to25	Urban	1.77	1.46
no	22to25	Urban	3.00	4.40

Literacy	Age Married	Region Type	Mean	Variance
yes	0to15	Urban	2.50	1.67
no	0to15	Urban	5.00	NA
yes	15to18	Urban	2.33	3.25
no	15to18	Urban	5.38	3.12
yes	18to20	Urban	1.59	1.61
no	18to20	Urban	2.67	1.33
yes	25to30	Urban	1.68	1.47
no	25to30	Urban	3.43	4.62
yes	30toInf	Urban	0.96	0.86
no	30toInf	Urban	0.60	0.30

Here we see that the mean and the variance for certain rows are off by a large amount, anywhere between 2 to 9. This shows evidence of overdispersion so we must make a negative binomial function like stated in Methods section.

	Estimate	2.5 %	97.5 %
(Intercept)	-1.727	-1.774	-1.680
literacyno	0.022	-0.092	0.136
ageMarried22to25	-0.071	-0.134	-0.009
ageMarried0to15	0.035	-0.181	0.250
ageMarried15to18	0.054	-0.030	0.139
ageMarried18to20	-0.001	-0.070	0.069
ageMarried25to30	-0.054	-0.125	0.016
ageMarried30toInf	-0.020	-0.159	0.118
literacyno:ageMarried22to25	0.215	0.063	0.367
literacyno:ageMarried0to15	0.055	-0.328	0.438
sd	0.295	0.232	0.262

```
## [1] 1 3
```

Since we find overdispersion we will interpret the values of the negative binomial as well. The intercept again is literate people that got married at 20-22 years old. The intercept shows a value of -1.727, which is the log rate of children per year for the baseline group. Which means exp(-1.727) = 0.177 children per year for literate people that got married at 20-22 years old. Then non-literate people have a value of 0.022 and exp(0.022) = 1.022, which means that non-illiterate people that got married at 20-22 years old have 1.022 more children per year then the baseline literate group (which means 0.177 x 1.022 = 0.18 children per year for non-literate people). Then we take a look at people married when they were 22-25 years old who are also literate and we see their value is -0.071 or exp(-0.071) = 0.932.This means that their rate of children per year decreased by 0.932 compared to our baseline group. And lastly the non-literate group of people that married when they were 25-30 years old have a value of 0.173, or exp(0.173) = 1.189. which means that their rate of children per year increased by 1.189 compared to the baseline (0.177 x 1.189 = 0.21 children per year).

	Estimate	2.5 %	97.5 %
(Intercept)	-1.709	-1.756	-1.662
literacyno	0.014	-0.099	0.128
ageMarried22to25	-0.068	-0.130	-0.005
ageMarried0to15	0.039	-0.176	0.253
ageMarried15to18	0.052	-0.032	0.137
ageMarried18to20	-0.001	-0.070	0.068
ageMarried25to30	-0.048	-0.119	0.022

	Estimate	2.5 %	97.5 %
ageMarried30toInf	-0.018	-0.157	0.120
region_typeUrban	-0.171	-0.237	-0.105
literacyno:ageMarried22to25	0.216	0.065	0.368
literacyno:ageMarried0to15	0.057	-0.324	0.438
literacyno:ageMarried15to18	0.008	-0.187	0.204
literacyno:ageMarried18to20	0.192	0.028	0.357
literacyno:ageMarried25to30	0.167	-0.002	0.336
literacyno:ageMarried30toInf	0.059	-0.255	0.373
sd	0.291	0.228	0.258

```
## [1] 1 3
```

Then just for the sake of looking we also take a look at how region affects our data, but much to our surprise it doesn't change the values too much. The only difference is that now the non-literate people that were married at 15 to 18 years old changed from -0.008 to 0.008 but even then it is not a significant change. We also see a new value for people who live in urban areas now (as that is how we decided to split our regions; rural and urban areas). Now the baseline is literate people who got married at 20-25 years old and they live in rural towns. All our interpretation stays the same with respect to the baseline group and the urban city value of -0.171 or exp(-0.171) = 0.843 tells us that compared to the baseline there is a decrease of 0.843 children per year in each household.

Conclusion

In conlcusion we see through out the results section that the literacy rate and the age you were married does affect the size of your family. We see that the sooner you get married the more children they have (as seen with the positive coeffecients). We also see the same outcome when we look at their literacy rate where if they are non-literate their coeffecients are positive which in turn means that there is a postive correlation between literacy and family size as well. Now since both literacy and age married affect the size of the family, the interaction variable, where we take both into account, also shows the same results where if they are both married early and have no literacy then their values are a lot higher.

Some outliners in these conclusions are obviously the catagory of married between 0 to 13 years olds as it is not always biologically possible for a child that young to conceive as well as the fact that it is frowned upon in most societies.

Limitations

We are unable to factor other outside factors that might come into play such as the fact that one does not need to be married to have children.

Because we categorized the region into 2 factors instead of the original 5, we may have limited some of the data we could have gathered from there. It seemed redundant for the sake of the report as our results do not show any significant change and our pre-research (in the introduction) did not seem to focus on how regions may affect family size, or even our other predictor variables, as well.

Reference

Saurabh, S., Sarkar, S., & Pandey, D. K. (2013). Female Literacy Rate is a Better Predictor of Birth Rate and Infant Mortality Rate in India. Journal of family medicine and primary care, 2(4), 349–353. https://doi.org/10.4103/2249-4863.123889

Woldeamanuel, B. T., Gessese, G. T., Demie, T. G., Handebo, S., & Biratu, T. D. (2023). Women's education, contraception use, and high-risk fertility behavior: A cross-sectional analysis of the demographic and health survey in Ethiopia. Frontiers in global women's health, 4, 1071461. https://doi.org/10.3389/fgwh.2023.1071461

Khan, J. R., Ara, T., Rahman, Md. M., Hossain, Md. B., & Muurlink, O. (2022). A multilevel assessment of the influence of education on women's uptake of institutional delivery services during childbirth in Bangladesh. Midwifery, 113, 103425–103425. https://doi.org/10.1016/j.midw.2022.103425