The Influence of Marriage Patterns and Women's Literacy on Family Size in late 1970s Portugal

STA303 - Winter 2025 - Assignment 1

Betty Li & Thomas Li

Introduction

In the late 1970s, Portugal transitioned from a rural agricultural culture to a modernized family-planning society owing to family size, a crucial demographic indicator that profoundly impacted social structures. A critical examination of family size in late 1970s Portugal would provide tremendous insights into the causal implications behind the substantial societal shift.

Adhikari's (2010) research article records how early marriage and low literacy rates result in high fertility rates in Nepal, where the cultural norms encourage young motherhood and limit access to family planning knowledge. However, Martin (1995) rejects the linear assumption of the relationship between education and fertility, arguing that early education initially increases the birth rate but that higher education ultimately reduces it. Beyond literacy and age of marriage, Duncan et al. (1962) emphasize the role of marriage duration, concluding that longer marriages naturally bring more time for childbearing and, therefore, a greater cumulative number of births. These former studies have outlined how literacy, marriage age, and marriage duration shape fertility patterns in a complex way, offering valuable inferences that align with the goal of this report: provide a clear understanding of the factors that influence family size in Portugal in the late 1970 and generate insightful results into this historical context. The findings will not merely conclude on historical fertility patterns but also spark broader discussions on how these factors will transform societal structure across cultures and times.

This report builds on data from the World Fertility Survey: Portugal, conducted during 1979 and 1980. The dataset contains information from multiple aspects of the families and the parents, but this report will focus on analyzing the influence of the mother's literacy, the mother's age at marriage, and the duration of the marriage on family size using a generalized linear model approach.

Methods

To model the effect of **literacy** and **age at marriage** on **family size** in Portugal, we will apply a Generalized Linear Model (GLM) in this study. The response variable family size is a count variable, which implies that a Poisson regression model is the most suitable model to be applied. The mean and variance should be equal under the assumption of a Poisson regression model. If the variance is much greater than the mean, over-dispersion exists. We will switch to the Negative Binomial model, which includes a dispersion parameter to adjust the variation.

The predictor variables of interests are **age at marriage** (ageMarried) and **literacy** (literacy). **Age at marriage** is a categorical variable with 7 levels and could influence the number of children since marrying younger provides more time for having children. **Literacy** is a binary variable showing the level of education, which can significantly impact reproductive choices and access to family planning. In addition, we include another control variable: months since marriage (monthsSinceM). **Months since marriage** demonstrate the period to have childbirth opportunities, which could significantly impact family size, according to the literature we found. An offset in covariate with log of years since marriage (months are converted into years)

will be implemented into the model, which allows us to compare fertility rates per unit of time rather than just the counts of children.

The number of children per person is modelled as:

$$Y_i \sim \text{Poisson}(\mu_i O_i)$$

Then, we will build the model:

$$\log(\mu_i) = \beta_0 + \beta_1 \text{literacy} + \beta_2 I_1 + \beta_3 I_2 + \beta_4 I_3 + \beta_5 I_4 + \beta_6 I_5 + \beta_7 I_6 + \log(O_i)$$

where μ_i is the expected number of children per unit exposure time per person, and O_i is the offset term representing the total exposure time. Indicators I_1 to I_6 represent 7 categories (including the reference category) of age married. The parameters β_0, β_1, β_2, β_3 represent the estimated effects of each predictor variable. Including the offset term ensures that people who have been married for different periods are appropriately accounted for.

If over-dispersion exists, the number of children per person will be modeled as:

$$Y_i \sim \text{Negetive Binomial}(\mu_i O_i, \tau)$$

We will perform an ANOVA test on the model to determine the significant predictors. A 5% significance level will be used to determine statistical significance, and the 95% confidence intervals will be calculated.

Results

The following data table, with numerical summaries and six graphic summary plots, demonstrates all relevant variables' basic distributions and relationships. For a more straightforward interpretation, **months since married** is mutated to **years since married**.

Variables	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Variance
Age	15.00	28.75	36.00	35.30	43.00	49.00	70.385
Years Since Married	0	6.08	12.33	12.93	19.17	35.92	62.280
Pregnancies	0	1	2	2.282	3	16	3.317
Children	0	1	2	2.26	3	17	3.464
Sons	0	0	1	1.187	3	10	1.474
Age Married	**0 ~ 15**	**15 ~ 18**	**18 ~ 20**	**20 ~ 22**	**22 ~ 25**	**25 ~ 30**	**30+**
	52	452	910	1126	1468	923	217
Region	**Lisbon**	**Porto**	**<10k**	**10k~20k**	**20k+**		
	470	160	3502	433	563		
Literacy	**Literate**	**Illiterate**					
	4567	581					

Table 1. *Statistical summaries of the data, the relevant variables, and their categories and distributions.*
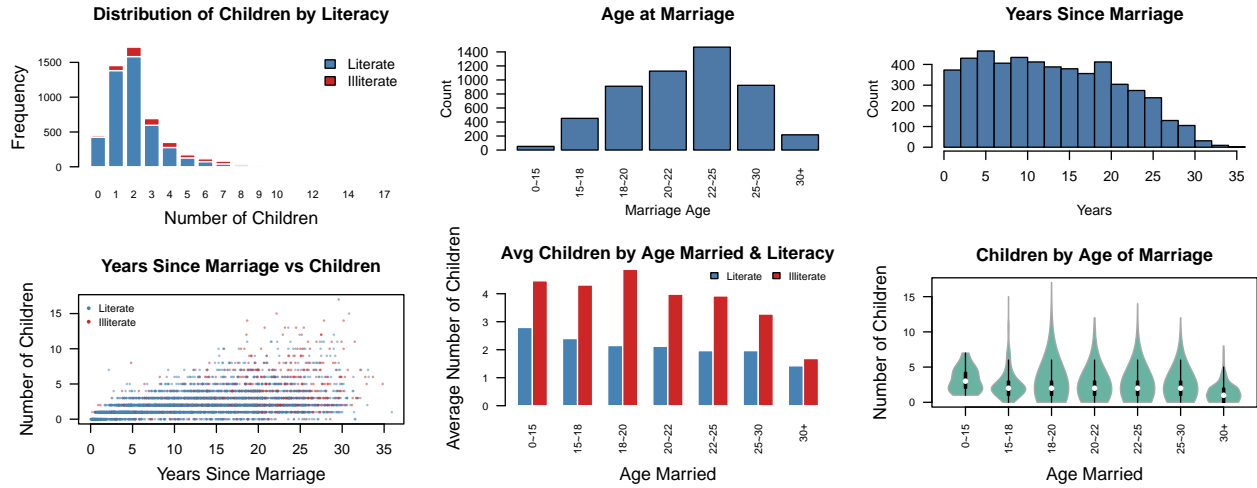
Figure 1. *Six summary plots on distributions and relationships of age married, years since married, literacy, and children.*

The distribution plots show that **age married** is unimodal with a slight left skew. Meanwhile, the frequency counts for **years since marriage** exhibit a decreasing pattern as the year increases, which holds on the entire spread of data, from Year 0 to roughly Year 36. According to the table summaries, literate individuals are 7 times more than illiterate ones from this survey. **Children**, the response variable, is right-skewed with an unimodal peak at 2. It spreads from 0 to 17 children and centers at about 2 children.

Regarding multivariate relationships, **age married**, **years since married**, and **literacy** are all likely to affect the number of **children**. Illiterate women seem to give birth more often, and those who entered marriages early saw higher fertility rates. Longer duration of marriage may also lead to larger family sizes. These patterns are a preliminary result of the pivotal contributors to more children in families, consistent with existing fertility studies.

The data summaries show noticeable over-dispersion in the response variable **children**, with a much greater variance than the mean. Therefore, the Negative Binomial model will also be taken into consideration to fit the data.

	Estimate	Std. Error	z value	$\Pr(>|z|)$
(Intercept)	-1.740	0.022	-79.980	0.000
literacyno	0.159	0.024	6.770	0.000
ageMarried0to15	-0.012	0.080	-0.151	0.880
ageMarried15to18	0.014	0.036	0.395	0.693
ageMarried20to22	-0.033	0.029	-1.124	0.261
ageMarried22to25	-0.062	0.028	-2.228	0.026
ageMarried25to30	-0.048	0.031	-1.557	0.120
ageMarried30toInf	-0.040	0.060	-0.669	0.504

```
## Waiting for profiling to be done...

##                   2.5 %    97.5 %  level   variable x      cex
## literacyno        1.1196890 1.2279667    no    literacy 1 3.291384
## ageMarried0to15   0.8409047 1.1524807  0to15 ageMarried 2 1.781194
## ageMarried15to18  0.9451070 1.0880403 15to18 ageMarried 3 2.664674
## ageMarried20to22  0.9146822 1.0245271 20to22 ageMarried 4 2.969442
## ageMarried22to25  0.8905246 0.9926858 22to25 ageMarried 5 3.034438
## ageMarried25to30  0.8966122 1.0125613 25to30 ageMarried 6 2.867521
```

Table 2. *Summary table of the Poisson Regression model and the confidence intervals of the coefficients of the predictors.*

We first built the Poisson regression model to analyze the relationship between literacy, age at marriage, and family size. The estimate for illiterate people is positive, with a p-value less than 0.05. This shows that literacy has a significant relationship with the number of children. The category of people who married between 22 and 25 years old has a statistically significant relationship with fertility, while other groups do not since their p-values are larger than 0.05.
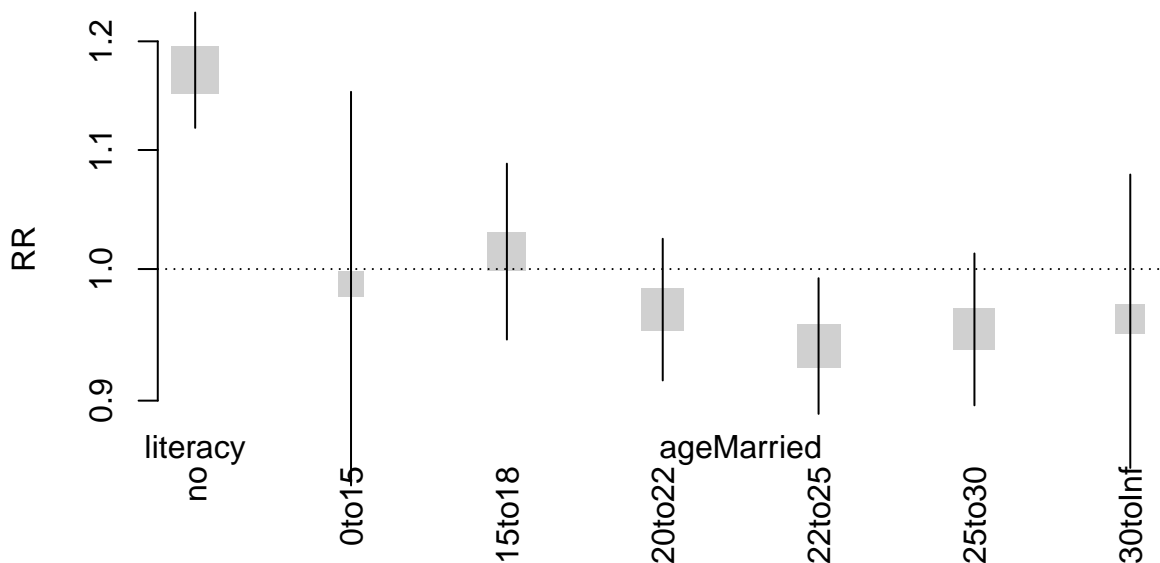
Figure 2. *Graph of the effect of predictors on the number of children.*

The effect plot above visually shows the variance around these estimates. The confidence interval of literacy does not overlap with 1, which shows its significant relationship with the number of children. The confidence intervals for most age groups include 1, meaning their effects are not statistically significant. The variance of age at marriage is inconsistent across different categories. This suggests that the impact of age at marriage varies depending on the specific age range.

	Estimate	2.5 %	97.5 %
(Intercept)	-1.714	-1.762	-1.666
literacyno	0.148	0.095	0.200
ageMarried0to15	-0.002	-0.180	0.177
ageMarried15to18	0.014	-0.064	0.093
ageMarried20to22	-0.034	-0.097	0.029
ageMarried22to25	-0.069	-0.129	-0.009
ageMarried25to30	-0.059	-0.126	0.008
sd	0.265	0.235	0.298

```
##                    2.5 %   97.5 % Estimate               level variable x
## literacyno       1.1000716 1.221372 1.1591364         literacyno        1
## ageMarried0to15  0.8348857 1.193901 0.9983841  ageMarried0to15         2
## ageMarried15to18 0.9376778 1.097703 1.0145402 ageMarried15to18         3
##                       cex
## literacyno       3.091964
## ageMarried0to15  1.672048
## ageMarried15to18 2.519213
```

Table 3. *Summary table of the Negative Binomial model and the confidence intervals of the coefficients of the predictors.*

A Negative Binomial model was built to solve the over-dispersion problem, as it introduces a dispersion parameter to better model the response variable. The results from the Negative Binomial model provide more reliable estimates of the predictors. Literacy shows a higher likelihood of having more children than the original model. The effect of age at marriage on number of children is also more consistent. The dispersion parameter in the Negative Binomial model captures the excess variability, which makes the confidence intervals and hypothesis tests more accurate. Comparing the two models, the Negative Binomial model is clearly the better choice since it solves the problem of over-dispersion.

Conclusion

The results from the Negative Binomial model suggest that illiterate women bear more children on average than literate ones, meaning that literacy significantly impacts family size. Despite Martin (1995) proposing the contrasting influences of lower and higher education on children born, our finding reveals a clear inverse linear relationship between literacy and fertility, contradicting any non-linear effect. Lastly, Adhikari's (2010) finding of the negative correlation between literacy and fertility rates would support our conclusion. However, there is less consistency in the effect of age at marriage on family size since only the 22~25 age group shows statistical significance.

In conclusion, our model posits that fertility patterns in late 1970s Portugal are shaped by the dominant predictor of literacy and the moderate factor of age at marriage. Our finding offers valuable insights into how women's education and marriage decisions could result in demographic changes and social structure shifts. It highlights the need to implement policies targeting education and marriage laws, thereby fostering family planning and further enhancing the nation's development.

Bibliography

1. Adhikari, R. (2010). Demographic, socio-economic, and cultural factors affecting fertility differentials in Nepal. *BMC pregnancy and childbirth, 10*, 1–11.

2. Martin, T. C. (1995). Women's education and fertility: results from 26 Demographic and Health Surveys. *Studies in family planning*, 187–202.

3. Duncan, O. D., Freedman, R., Coble, J. M., & Slesinger, D. P. (1965). Marital fertility and size of family of orientation. *Demography, 2(1)*, 508–515.