Portugal fertility survey 1979

- More information
- data source
- data dictionary

File `portugal.RData` on the course web site. Code in `Assignment1.Rmd`

```
head(portugal)
```

```
##   age ageMarried monthsSinceM pregnancies children sons region literacy
## 1  43     22to25          242           3        3    2  lt10k      yes
## 2  32     22to25          124           1        1    0  lt10k      yes
## 3  22     15to18           59           1        1    1  lt10k      yes
## 4  28     22to25           63           1        1    0  lt10k      yes
## 5  30     15to18          169           2        2    2  lt10k      yes
## 6  37     18to20          226           2        2    1  lt10k      yes
```

```
table(portugal$region)
```

```
##
##  lt10k lisbon  porto   20k+ 10-20k
##   3502    470    160    583    433
```

Region is `lt10k` rural areas (less than 10,000 people), towns of size 10-20k, 20k+, and the two largest cities (Lisbon and Porto).

The Question:

- How do literacy and age of marriage affect family size?
- After we account for known explanatory variables, how much variation (if any) is there in birth rates?

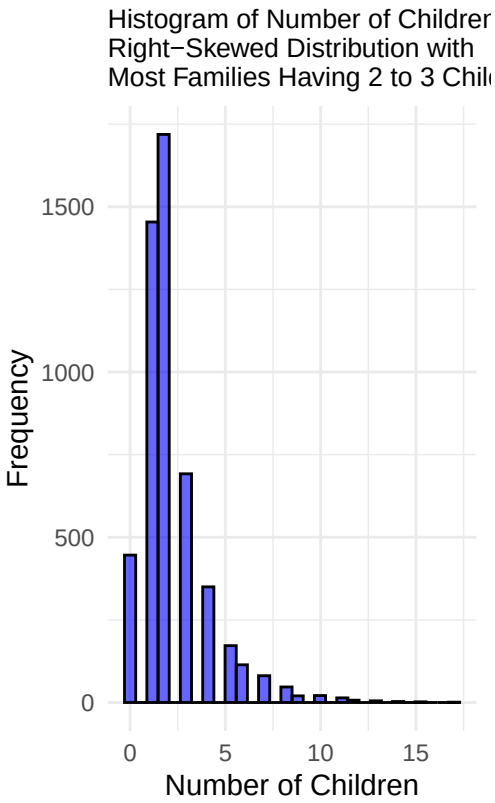
Notes:

- By European standards Portugal is a poor country, and in 1980 it had the same GDP per capita as Mexico.
- It's well known that families are larger in rural areas (i.e. `region` is a confounder).
- You should proceed as if an expert in this area has told you that you do not need to consider zero-inflation. Zero-inflated models do not fit well to this data, probably because birth rates are lower than Fiji, many zeros are expected, and the likelihood is flat.

Figure 1: Statistical Summary and Histogram of response variable 'Children'

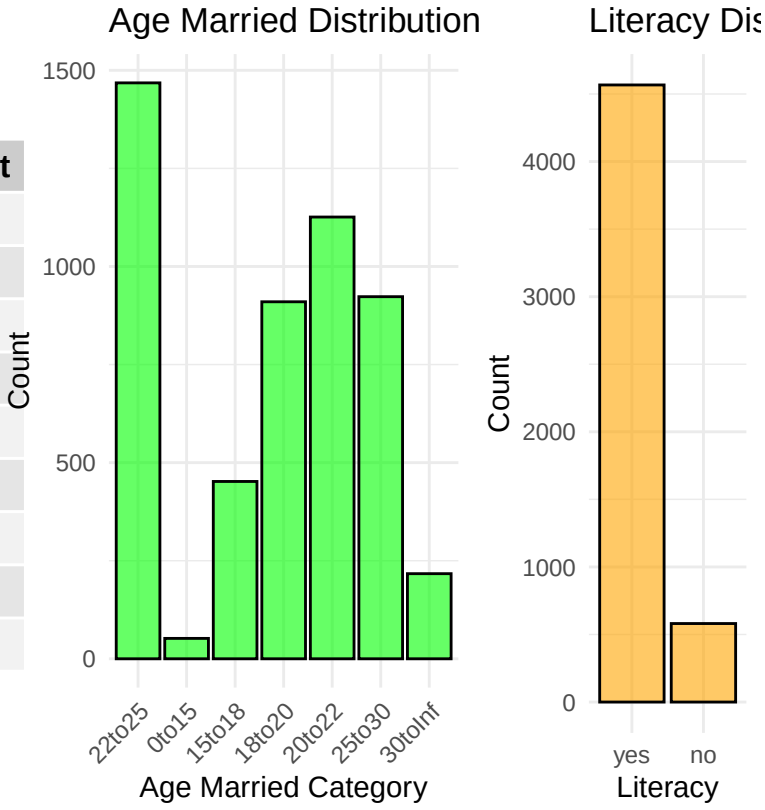
Statistical Summary of Children Showing
Mean, Median, Standard Deviation,
Minimum and Maximum

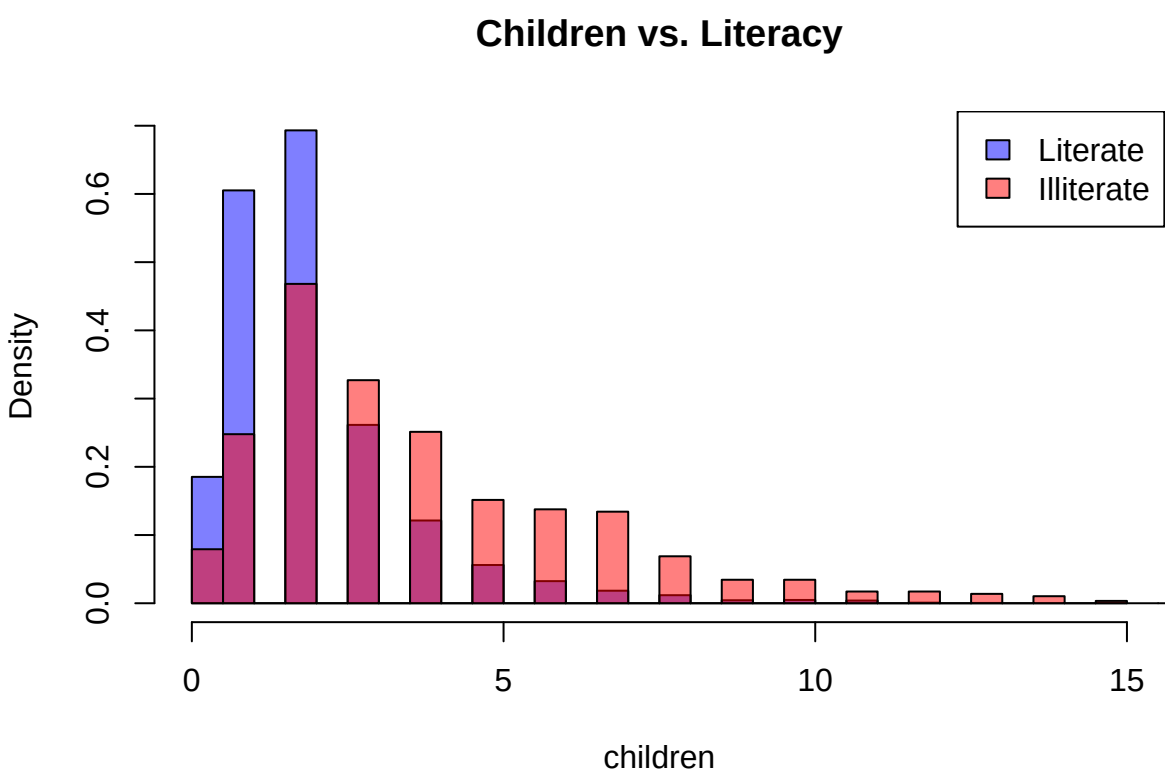
	Mean	Median	SD	Min	Max
1	2.26	2	1.86	0	17



net variables 'Age Married' and 'Literacy'. Out of 5148 samples, most samples married between 20 t

	Variable	Category	Count
1	Age Married	22to25	1468
2	Age Married	0to15	52
3	Age Married	15to18	452
4	Age Married	18to20	910
5	Age Married	20to22	1126
6	Age Married	25to30	923
7	Age Married	30toInf	217
8	Literacy	yes	4567
9	Literacy	no	581





	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.789	0.023	-77.858	0.000
literacyno	0.159	0.024	6.770	0.000
ageMarried22to25	-0.013	0.029	-0.468	0.640
ageMarried0to15	0.036	0.081	0.448	0.654
ageMarried15to18	0.062	0.037	1.702	0.089
ageMarried18to20	0.048	0.031	1.557	0.120
ageMarried20to22	0.016	0.030	0.528	0.598
ageMarried30toInf	0.008	0.060	0.136	0.891

```
##
## Call:
## glm(formula = children ~ offset(logYearsMarried) + literacy +
##     ageMarried, family = poisson, data = portugal)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.788769   0.022975 -77.858  < 2e-16 ***
## literacyno         0.159422   0.023547   6.770 1.29e-11 ***
## ageMarried22to25  -0.013443   0.028703  -0.468   0.6395
## ageMarried0to15    0.036150   0.080721   0.448   0.6543
## ageMarried15to18   0.062495   0.036710   1.702   0.0887 .
## ageMarried18to20   0.048288   0.031022   1.557   0.1196
## ageMarried20to22   0.015770   0.029892   0.528   0.5978
## ageMarried30toInf  0.008235   0.060372   0.136   0.8915
```

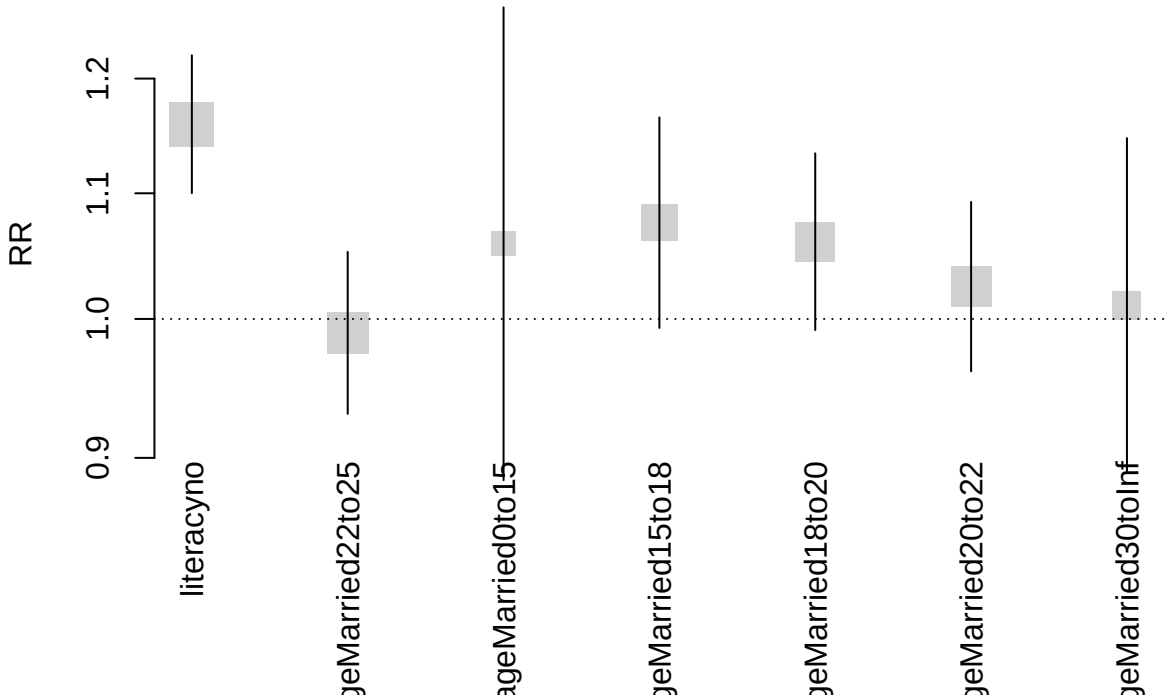
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5654.4  on 5147  degrees of freedom
## Residual deviance: 5600.1  on 5140  degrees of freedom
## AIC: 18047
##
## Number of Fisher Scoring iterations: 5
```

	Estimate	2.5 %	97.5 %
(Intercept)	-1.7724898	-1.8216692	-1.7233104
literacyno	0.1476897	0.0953899	0.1999895
ageMarried22to25	-0.0105444	-0.0719655	0.0508767
ageMarried0to15	0.0570068	-0.1223318	0.2363453
ageMarried15to18	0.0730426	-0.0067813	0.1528666
ageMarried18to20	0.0586100	-0.0084053	0.1256252
ageMarried20to22	0.0245153	-0.0397481	0.0887786
sd	0.2650000	0.2650000	0.2650000

```
##                 2.5 %    97.5 %  Estimate
## (Intercept) 0.1617555 0.1784744 0.1699094
```

```
##           2.5 %    97.5 %  Estimate
## sigma 0.2977129 0.2350557 0.2645356
```

```
##                      2.5 %   97.5 % Estimate            level variable x
## literacyno       1.1000877 1.221390 1.159153       literacyno          1
## ageMarried22to25 0.9305630 1.052193 0.989511 ageMarried22to25          2
## ageMarried0to15  0.8848547 1.266612 1.058663  ageMarried0to15          3
##                       cex
## literacyno       3.091968
## ageMarried22to25 2.853160
## ageMarried0to15  1.669737
```




```
## [1] 0.2645351
```

```
## [1] 0.004897076
```

literacy	agemarried	Mean	Variance
yes	25to30	1.97	1.98
no	25to30	3.27	5.49
yes	22to25	1.97	1.98
no	22to25	3.92	7.05
yes	0to15	2.79	2.64
no	0to15	4.46	2.27
yes	15to18	2.40	4.10
no	15to18	4.31	6.22
yes	18to20	2.15	2.98
no	18to20	4.87	12.98
yes	20to22	2.12	2.68
no	20to22	3.98	7.91
yes	30toInf	1.42	1.79
no	30toInf	1.68	3.41