Modeling Fertility Data

Junru Zhang

06/07/2020

Loading and Cleaning the Data

```
fijiFile = 'fijiDownload.RData'
load(fijiFile)
```

```
# get rid of newly married women and those with missing literacy status
fijiSub = fiji[fiji$monthsSinceM > 0 & !is.na(fiji$literacy),]
fijiSub$logYears = log(fijiSub$monthsSinceM/12)
fijiSub$ageMarried = relevel(fijiSub$ageMarried, '15to18')
fijiSub$urban = relevel(fijiSub$residence, 'rural')
```

Fitting Generalized Linear Models

```
fijiRes = glm(
children ~ offset(logYears) + ageMarried + ethnicity + literacy + urban,
family=poisson(link=log), data=fijiSub)
logRateMat = cbind(est=fijiRes$coef, confint(fijiRes, level=0.99))
```

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

```
knitr::kable(cbind(
  summary(fijiRes)$coef,
  exp(logRateMat)),
digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)	est	0.5 %	99.5 %
(Intercept)	-1.181	0.017	-69.196	0.000	0.307	0.294	0.321
ageMarried0to15	-0.119	0.021	-5.740	0.000	0.888	0.841	0.936
ageMarried18to20	0.036	0.021	1.754	0.079	1.037	0.983	1.093
ageMarried20to22	0.018	0.024	0.747	0.455	1.018	0.956	1.084
ageMarried22to25	0.006	0.030	0.193	0.847	1.006	0.930	1.086
ageMarried25to30	0.056	0.048	1.159	0.246	1.057	0.932	1.195
ageMarried30toInf	0.138	0.098	1.405	0.160	1.147	0.882	1.462
ethnicityindian	0.012	0.019	0.624	0.533	1.012	0.964	1.061
ethnicityeuropean	-0.193	0.170	-1.133	0.257	0.824	0.514	1.242
ethnicitypartEuropean	-0.014	0.069	-0.206	0.837	0.986	0.822	1.171
ethnicitypacificIslander	0.104	0.055	1.884	0.060	1.110	0.959	1.276
ethnicityroutman	-0.033	0.132	-0.248	0.804	0.968	0.675	1.336
ethnicitychinese	-0.380	0.121	-3.138	0.002	0.684	0.492	0.920
ethnicityother	0.668	0.268	2.494	0.013	1.950	0.895	3.622
literacyno	-0.017	0.019	-0.857	0.391	0.984	0.936	1.034
urbansuva	-0.159	0.022	-7.234	0.000	0.853	0.806	0.902
urbanotherUrban	-0.068	0.019	-3.513	0.000	0.934	0.888	0.982

```
fijiSub$marriedEarly = fijiSub$ageMarried == '0to15'
fijiRes2 = glm(
children ~ offset(logYears) + marriedEarly + ethnicity + urban,
 family=poisson(link=log), data=fijiSub)
logRateMat2 = cbind(est=fijiRes2$coef, confint(fijiRes2, level=0.99))
```

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

```
knitr::kable(cbind(
     summary(fijiRes2)$coef,
     exp(logRateMat2)),
  digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)	est	0.5 %	99.5 %
(Intercept)	-1.163	0.012	-93.674	0.000	0.313	0.303	0.323
marriedEarlyTRUE	-0.136	0.019	-7.189	0.000	0.873	0.832	0.916
ethnicityindian	-0.002	0.016	-0.154	0.877	0.998	0.958	1.039
ethnicityeuropean	-0.175	0.170	-1.034	0.301	0.839	0.524	1.262
ethnicitypartEuropean	-0.014	0.068	-0.202	0.840	0.986	0.823	1.171
ethnicitypacificIslander	0.102	0.055	1.842	0.065	1.107	0.957	1.273
ethnicityroutman	-0.038	0.132	-0.285	0.775	0.963	0.672	1.330
ethnicitychinese	-0.379	0.121	-3.130	0.002	0.684	0.493	0.921
ethnicityother	0.681	0.268	2.545	0.011	1.976	0.907	3.667
urbansuva	-0.157	0.022	-7.162	0.000	0.855	0.808	0.904
urbanotherUrban	-0.066	0.019	-3.414	0.001	0.936	0.891	0.984

Explaining the fijiRes Model

$$Y_i \sim Poisson(O_i\mu_i) \quad \text{with link function } log(\mu_i) = X_i\beta$$

where the response Y_i is the number of children born of the individual i after marriage, and covariates X_i include: $X_{i1}, ..., X_{i6}$: the range of age when married of the individual i, $=1$ if married in the range 0to15, 18to20, 20to22, 22to25, 25to30, 30+ **respectively**, $=0$ otherwise,

$X_{i7}, ..., X_{i13}$: the ethnic group that individual i belongs to, $=1$ if individual i is Indian, European, part European, pacific Islander, Rotuman, Chinese, other, **respectively**, otherwise $=0$,

X_{i14} : whether the individual i is literate or not, $=1$ if not literate, and $=0$ if literate,

X_{i15} : the residence area of the individual i, $=1$ if living in the capital city of Suva and $=0$ otherwise,

X_{i16} : the residence area of the individual i, $=1$ if living in urban area and $=0$ otherwise.

O_i is the number of years since married. $(Offset\ terms)$

μ_i is the rate of children born per year.

Note: since the response variable is the number of children of each individual after marriage, this is a counting process. Thus it makes sense that the response variable follows a **poisson distribution**. And we think that age when married, ethnic group, being literiate or not, area of residence and number of months since married are all factors that affect the outcome, which is the number of children born.

Likelihood Ratio Test

```
lmtest::lrtest(fijiRes2, fijiRes)
```

```
## Likelihood ratio test
##
## Model 1: children ~ offset(logYears) + marriedEarly + ethnicity + urban
## Model 2: children ~ offset(logYears) + ageMarried + ethnicity + literacy +
##     urban
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1  11 -9604.3
## 2  17 -9601.1  6 6.3669     0.3834
```

Likelyhood ratio test can only compare nested models. Although it might look like fijiRes2 does not have the MarriedEarly covaraite, which fijiRes has, in fact it does. The MarriedEarly factor in fijiRes evaluates to TRUE if age married is in the range of 0to15, and it evaluates to FLASE if age married is not in the range of 0to15. On the other hand, the ageMarried factor in fijiRes2 has a covariate that evaluates to 1 if age married is in the range of 0to15, and 0 otherwise. So essentially, fijiRes2 contain all covariates included in fijiRes and hence they are nested models.

The constraints on the regression coefficients are that **the coefficents ageMarried18to20, ageMarried20to22, ageMarried22to25, ageMarried25to30, ageMarried30toInf, and literacy are all at the same level** in the restricted model, fijiRes.

Using Data to Verify Hypothesises

Hypothesis 1

Improving girls' education and delaying marriage will result in women choosing to have fewer children and increase the age gaps between their children

This hypothesis is **not acceptable**.

First of all, the LRT suggests adding the literacy term isn't helping explain the data significantly better (p-value= 0.3834 >0.05). Therefore we can conclude that there is no evidence that improving education would affect how many children women choose to have. This is then inconsistent with the hypothesis.

Considering the estimates from fijiRes for the levels of age married, none of the later ages married have a significantly different rate of having children from the 15to18 group (1 in all the CIs equivalent here to p-value>0.05), which is not consistent with a "delayed marriage effect". We can also see this from the LR test as the simpler version of the variable appears to work just as well as the version with more levels, also supporting this idea of no meaningful differences across these groups.

Also, in fijiRes2, we observe that the MarriedEarly coefficient has an odds ratio ("est") of 0.873 and a 95% confidence of (0.832, 0.916). We interpret this as, holding all other covariates constant, getting married before age 15 results in having fewer children (between 8 and 17% lower) than getting married after the age of 15. This is consistent with women married before 15 not being fertile in the early years of their marriage.

Hypothesis 2

Contraception was not widely available in Fiji in 1974 and as a result there was no way for married women to influence their birth intervals.

This hypothesis is **acceptable**.

Our results seem consistent with the idea that contraception wasn't widely available in Fiji, as neither being better educated (literate) nor marrying later seemed to explain the rate at which women were having children. Education and delayed marriage won't readily influence the rate of having children if the mechanisms to control getting pregnant, e.g. contraception, aren't available to anyone.