Fiji_Fertility

Qizheng Wang

5/7/2020

Fiji Fertility Survey Analysis (Generalised linear models - Poisson)

```
setwd("C:/Users/roder/OneDrive/Documents/GitHub/For2020Job/Fiji_Fertility_Analysis")

fijiFile = load("fiji.RData")
```

```
#Eliminate newly married women and women who don't have 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')

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

	Estimate	Std. Error	z value	Pr(>|z|)	est	0.5 %	99.5 %
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

Assumption made: a woman's fertility rate is zero before marriage and constant thereafter until menopause.

fijiRes1 Model:

$$\log(\text{number of children}_i) = \boldsymbol{X_i\beta} + \log(\text{years married}_i)$$

It's a Generalised linear model using poisson regression. Response is number of children had per year. Link function is a log link. Offset is log(year married). Covariate are age married, ethinicity, literate, and urban.

By oberserving the summary table, there are some really big p-value, I want to further analyse these covariate, thus I created another model fijiRes2:

```
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

```
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
```

Then use likelihood ratio test to test which model is preferable.

```
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
```

Since fijiRes2 is nested with fijiRes 2, the comparison is valid. The constraints on the vector of regression coefficients, β, would be that literacy would have $\beta = 0$ as it is not included in the model, and the levels of age married, other than 0to15 would be constrained to all have the same β as marriedEarly collapses all of these into one level. I.e., $\beta_{15to18} = \beta_{18to22} = \ldots = \beta_{30toInf}$
The p-value is bigger than 0.05, thus the literay isn't helping for explaining the data significantly. Thus we can assume that for high possibility, the increase in education is not significantly related to children had per year.
From the summary of fijiRes, we can see a series of large p-value for levels of age married, thus a delated marriage is not significant with child birth either.
But howev by examing married early varialbe in fijiRes2.