019 - Poisson Regression

EPIB 607 - FALL 2020

Sahir Rai Bhatnagar
Department of Epidemiology, Biostatistics, and Occupational Health
McGill University

sahir.bhatnagar@mcgill.ca

slides compiled on November 6, 2020

1. Breastfeeding and respiratory infection I

2. Breastfeeding and respiratory infection II

3. Malaria control with bednets

4. Population mortality rates in Denmark

A total of 189,612 person-years of follow up were accumulated over the course of the study: 151,690 among infants who were being breastfed and 37,922 among infants not being breastfed. Over the course of follow up the investigators identified 514,230 incident cases of respiratory infection among breastfeeding infants and 140,312 among non-breastfeeding infants. Calculate the crude incidence rate difference and 95% CI comparing infants who were not breastfed with those who were.

```
fit <- glm(cases ~ -1 + PT + PT:not_breastfed, family = poisson(link = identity))
print(summary(fit), signif.stars = F)
```

```
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## PT                 3.39001    0.00473   717.1   <2e-16
## PT:not_breastfed  0.31001    0.01095    28.3   <2e-16
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance:        Inf  on 2  degrees of freedom
## Residual deviance: 1.1195e-10  on 0  degrees of freedom
## AIC: 32.68
##
## Number of Fisher Scoring iterations: 2
```

1. Breastfeeding and respiratory infection I

2. Breastfeeding and respiratory infection II

3. Malaria control with bednets

4. Population mortality rates in Denmark

Calculate the crude incidence rate ratio and 95% CI comparing infants who were not breastfed with those who were.

```
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.22083    0.00139   875.5   <2e-16
## not_breastfed  0.08751    0.00301    29.1   <2e-16
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 8.3002e+02  on 1  degrees of freedom
## Residual deviance: 1.1533e-10  on 0  degrees of freedom
## AIC: 32.68
##
## Number of Fisher Scoring iterations: 2
```

See the 2018 Lancet article *Efficacy of Olyset Duo, a bednet containing pyriproxyfen and permethrin, versus a permethrin-only net against clinical malaria in an area with highly pyrethroid-resistant vectors in rural Burkina Faso: a cluster-randomised controlled trial* by Tiono et. al. Reproduce the Rate ratio (95% CI) in Table 2. Calculate the rate difference and 95% CI comparing PPF-treated to Standard long-lasting insecticidal nets. Check the goodness of fit.

```
## Call:
## glm(formula = cases ~ exposure + offset(log(years)), family = poisson(link = log),
##     data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.6831     0.0243   28.09  < 2e-16
## exposure     -0.2669     0.0329   -8.12  4.6e-16
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1381.2  on 23  degrees of freedom
## Residual deviance: 1316.0  on 22  degrees of freedom
## AIC: 1477
##
## Number of Fisher Scoring iterations: 5
```

We can fit the following simple (multiplicative) rate ratio model to the patterns of mortality rates for 1980-1984 and 2000-2004. The reference cell is females 70-74, 1980-84. R = rate. M = multiplier.

Year	Age	Female (F)			Male (M)			
	70-74	R_F			R_F		$\times M_M$	
1980-	75-79	R_F	$\times M_{75}$		R_F	$\times M_{75}$	$\times M_M$	
1984	80-84	R_F	$\times M_{80}$		R_F	$\times M_{80}$	$\times M_M$	
	85-89	R_F	$\times M_{85}$		R_F	$\times M_{85}$	$\times M_M$	
	70-74	R_F		$\times M_{20y}$	R_F		$\times M_M$	$\times M_{20y}$
2000-	75-79	R_F	$\times M_{75}$	$\times M_{20y}$	R_F	$\times M_{75}$	$\times M_M$	$\times M_{20y}$
2004	80-84	R_F	$\times M_{80}$	$\times M_{20y}$	R_F	$\times M_{80}$	$\times M_M$	$\times M_{20y}$
	85-89	R_F	$\times M_{85}$	$\times M_{20y}$	R_F	$\times M_{85}$	$\times M_M$	$\times M_{20y}$

Year	Age	Female_deaths	Female_PT	Female_rate	Male_deaths	Male_PT	Male_rate
1980-1984	70-74	15989	586883	0.02724	23810	456908	0.05211
1980-1984	75-79	20838	454143	0.04588	24707	300319	0.08227
1980-1984	80-84	24073	297679	0.08087	20319	167304	0.12145
1980-1984	85-89	20216	147772	0.13681	13524	74296	0.18203
2000-2004	70-74	13912	521562	0.02667	17360	436995	0.03973
2000-2004	75-79	19731	471946	0.04181	22477	341363	0.06584
2000-2004	80-84	25541	369990	0.06903	22992	217930	0.10550
2000-2004	85-89	27135	226798	0.11964	17444	104010	0.16772
2005-2009	70-74	12179	540569	0.02253	15782	472013	0.03344
2005-2009	75-79	17273	444474	0.03886	19547	344351	0.05676
2005-2009	80-84	23513	363534	0.06468	21781	230530	0.09448
2005-2009	85-89	26842	237877	0.11284	17811	114485	0.15557

Rate =	________	$\times$________	$\times$________	$\times$________	$\times$________	$\times$________
		if 75-79	if 80-84	if 85-89	if male	if 2000-04
$\log[Rate] =$	________	$+$________	$+$________	$+$________	$+$________	$+$________
		if 75-79	if 80-84	if 85-89	if male	if 2000-04
$\log[Rate] =$	________	$+$________	$+$________	$+$________	$+$________	$+$________
		$\times$ I_{75-79}	$\times$ I_{80-84}	$\times$ I_{85-89}	$\times$ I_{male}	$\times$ $I_{2000-04}$

where each 'I'' is a (0/1) indicator of the category in question. By using both the 0 and 1 values of each I, this 6-parameter equation produces a fitted value for each of the $4 \times 2 \times 2 = 16$ cells.