IDS 702

Poisson regression with rates

Poisson regression with rates

- The poisson distribution models counts in a specified time interval
- When the time interval is the same for each subject (e.g., awards in one year), we don’t need to account for this in the model
- But if the time interval varies among subjects, we can account for this in the poisson model
    - Example:
- This could also mean modeling incidence rates in different populations
    - Example:

Poisson regression with rates

```
> ## Check data
> nonmel
   cases      n        city age.range
1      1 172675 Minneapolis     15_24
2     16 123065 Minneapolis     25_34
3     30  96216 Minneapolis     35_44
4     71  92051 Minneapolis     45_54
5    102  72159 Minneapolis     55_64
6    130  54722 Minneapolis     65_74
7    133  32185 Minneapolis     75_84
8     40   8328 Minneapolis       85+
9      4 181343      Dallas     15_24
10    38 146207      Dallas     25_34
11   119 121374      Dallas     35_44
12   221 111353      Dallas     45_54
13   259  83004      Dallas     55_64
14   310  55932      Dallas     65_74
15   226  29007      Dallas     75_84
16    65   7583      Dallas       85+
```

https://rpubs.com/kaz_yos/poisson

Poisson regression with rates

- We can write the rate as counts/population, i.e., $\pi_i = \frac{\lambda_i}{n_i}$
- Then we can write the model as $log\left(\pi_i = \frac{\lambda_i}{n_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$
- $\implies log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + log(n_i)$

Implementation in R

```
> poismod_rates <- glm(cases ~ city + age.range, offset = log(n), family = poisson, data =
 nonmel)
> summary(poismod_rates)

Call:
glm(formula = cases ~ city + age.range, family = poisson, data = nonmel,
    offset = log(n))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.50598  -0.48566   0.01639   0.36926   1.24763

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -5.4834     0.1037 -52.890  < 2e-16 ***
cityDallas      0.8039     0.0522  15.399  < 2e-16 ***
age.range15_24 -6.1742     0.4577 -13.488  < 2e-16 ***
age.range25_34 -3.5440     0.1675 -21.160  < 2e-16 ***
age.range35_44 -2.3268     0.1275 -18.254  < 2e-16 ***
age.range45_54 -1.5790     0.1138 -13.871  < 2e-16 ***
age.range55_64 -1.0869     0.1109  -9.800  < 2e-16 ***
age.range65_74 -0.5288     0.1086  -4.868 1.13e-06 ***
age.range75_84 -0.1157     0.1109  -1.042    0.297
```

Interpretation

```
> cbind(exp(coef(poismod_rates)),exp(confint(poismod_rates)),summary(poismod_rates)$coeffic
ient[,4])
Waiting for profiling to be done...
                                  2.5 %      97.5 %
(Intercept)     0.004155093 0.0033714705 0.005063810 0.000000e+00
cityDallas      2.234233723 2.0183157107 2.476754713 1.663497e-53
age.range15_24  0.002082493 0.0007355278 0.004598702 1.830153e-41
age.range25_34  0.028897843 0.0206756999 0.039925425 2.223891e-99
age.range35_44  0.097605351 0.0761403712 0.125574089 1.909191e-74
age.range45_54  0.206182619 0.1655570968 0.258794339 9.487693e-44
age.range55_64  0.337260039 0.2724828938 0.421069706 1.125469e-22
age.range65_74  0.589325065 0.4784318120 0.732709014 1.126463e-06
age.range75_84  0.890782074 0.7196290043 1.112221150 2.972102e-01
```

Controlling for age group, the risk of non-melanoma skin cancer is 2.2 times higher in Dallas than Minneapolis. We are 95% confident that the true risk ratio is between 2.02 and 2.48, which is statistically significant ($p<.001$).