Exercises on Methods for Count Data

Once again, the real exercises are analysing real data. Use S-PLUS to help you answer the questions if you wish.

1.

(i) Find the mean and variance of $\widehat{\phi}$ and $\log \widehat{\phi}$ for the empirical odds ratio in a 2×2 table.

(ii) [Rather technical] Suppose we modify the empirical odds ratio to

$$\widetilde{\phi} = \frac{(n_{11} + 0.5)(n_{22} + 0.5)}{(n_{12} + 0.5)(n_{21} + 0.5)}$$

Show that for large n

$$\text{var}(\log \widetilde{\phi}) \approx \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}.$$

[A proof can be found in Agresti, A. (2002) *Categorical Data Analysis* Second Edition §3.1.7 and doubtless in the first edition.]

2. Compare the chi-squared and Fisher exact tests for the following table and interpret the results you get.

Row	1	2	3	4	5	6	7	8	9
1	0	7	0	0	0	0	0	1	1
2	1	1	1	1	1	1	1	0	0
3	0	8	0	0	0	0	0	0	0

3. Consider a linear logistic regression on just one variable, the dose x. Derive formulae for the estimator of ED50 and its approximate standard error. [Venables & Ripley (2002, p. 193)]

4. Laird & Oliver (1981) collected data on deaths for patients after a heart valve operation, with the age at the operation and the type of valve. Subjects were followed up for between 3 and 97 months.

Age		Aortic	Mitral
< 55	Deaths	4	1
	months at risk	1259	2082
≥ 55	Deaths	7	9
	months at risk	1417	1647

(a) What type of sampling is this?

(b) Fit a log-linear model and interpret the coefficients you obtain (including giving suitable confidence intervals).

5. Both the gamma distribution and the negative binomial distribution

$$f_Y(y;\theta,\mu) = \frac{\Gamma(\theta+y)}{\Gamma(\theta)\,y!}\frac{\mu^y\,\theta^\theta}{(\mu+\theta)^{\theta+y}}$$

for fixed θ provide examples of GLMs. Identify the various components of the GLM setup, and give the canonical links.

[The gamma case can be found in McCullagh & Nelder (1989) and many other places. The results for the negative binomial case are given in Venables & Ripley (2002, §7.4).]