PBIO 294 HW 4 Oct 2017

Reinforcing Likelihood concepts, introducing linear regression, and developing R skills.

Problem 1: Lavine Question 2.8 (slightly modified)

The book *Data by* Andrews and Herzberg [1985] contains lots of data sets that have been used for various purposes in statistics. One famous data set records the annual number of deaths by horse kicks in the Prussian Army from 1875-1894 for each of 14 corps. The data are included below. (The data come from Table 4.1 in the Andrews and Herzberg [1985] book.)

4	1	1	1875	0	0	0	0	0	0	1	1	0	0	0	1	0	3
4	1	2	1876	2	0	0	0	1	0	0	0	0	0	0	1	1	5
4	1	3	1877	2	0	0	0	0	0	1	1	0	0	1	0	2	0	7
4	1	4	1878	1	2	2	1	1	0	0	0	0	0	1	0	1	0	9
4	1	5	1879	0	0	0	1	1	2	2	0	1	0	0	2	1	0	10
4	1	6	1880	0	3	2	1	1	1	0	0	0	2	1	4	3	0	18
4	1	7	1881	1	0	0	2	1	0	0	1	0	1	0	0	0	0	6
4	1	8	1882	1	2	0	0	0	0	1	0	1	1	2	1	4	1	14
4	1	9	1883	0	0	1	2	0	1	2	1	0	1	0	3	0	0	11
4	1	10	1884	3	0	1	0	0	0	0	1	0	0	2	0	1	1	9
4	1	11	1885	0	0	0	0	0	0	1	0	0	2	0	1	0	1	5
4	1	12	1886	2	1	0	0	1	1	1	0	0	1	0	1	3	0	11
4	1	13	1887	1	1	2	1	0	0	3	2	1	1	0	1	2	0	15
4	1	14	1888	0	1	1	0	0	1	1	0	0	0	0	1	1	0	6
4	1	15	1889	0	0	1	1	0	1	1	0	0	1	2	2	0	2	11
4	1	16	1890	1	2	0	2	0	1	1	2	0	2	1	1	2	2	17
4	1	17	1891	0	0	0	1	1	1	0	1	1	0	3	3	1	0	12
4	1	18	1892	1	3	2	0	1	1	3	0	1	1	0	1	1	0	15
4	1	19	1893	0	1	0	0	0	1	0	2	0	0	1	3	0	0	8
4	1	20	1894	1	0	0	0	0	0	0	0	1	0	1	1	0	0	4

Let *Yi j* be the number of deaths in year *i*, corps *j*, for *i* = 1875,...,1894 and *j* = 1,...,14. The *Yij*s are in columns 5–18 of the table.

(a) What are the first four columns of the table?

(b) What is the last column of the table?

(c) What is a good model for the data? Explain.

(d) Suppose you model the data as i.i.d. Poi(λ).

i. Plot the likelihood function for λ.

ii. Find λ hat (i.e., the mle estimate of λ) by corps and by year.

iii. What can you say about the rate of death by horse kick in the Prussian cavalry at the end of the 19th century?

(e) Is there any evidence that different corps had different death rates? Explain.

Problem 2: This question uses the height vs. weight data from 6 Sep. The data can be found on the class website from that week.

a) Fit a linear regression to these data using the 'lm' command within R. The model should predict the observed weight from height and the model should include an intercept and an effect of height. Plot the estimated linear fit superimposed on the data and report the parameter estimates for the intercept, effect of height, and sigma.

b) Repeat a) except fitting your model using your own likelihood function rather than R's built in 'lm'. Plot the estimated linear fit superimposed on the data and report the parameter estimates for the intercept, effect of height, and sigma. Are they the same as from 'lm'?

c) Repeat a) and b) with the addition of an effect of gender in the linear model. Plot the model fit (i.e., with separate lines for males and females) along with the data. Report parameter estimates for the intercept, effect of height, gender and sigma.

d) Report AIC values for the linear model with and without the gender term based on your own likelihood functions and the 'lm' fits. Which model is better supported based on AIC values?