DATA557 - Homework 8

Will Wright

3/6/2019

Question 1.

a) Use Poisson regression to estimate the effect of dose on post-treatment cell count in the "cells" data from HW 7. Treat dose as a quantitative variable, and adjust for sex, age and pre-treatment cell count. Report the coefficient estimate table.

Est	imate Std	. Error z v	alue Pr(	>|z|)
(Intercept)	4.4301	0.0467	94.950	0
dose	0.0277	0.0003	80.506	0
sex	0.1920	0.0213	9.008	0
age	-0.0228	0.0007	-33.533	0
count0	0.0052	0.0001	36.180	0

b) State the interpretation of the exponentiated coefficient estimate for dose.

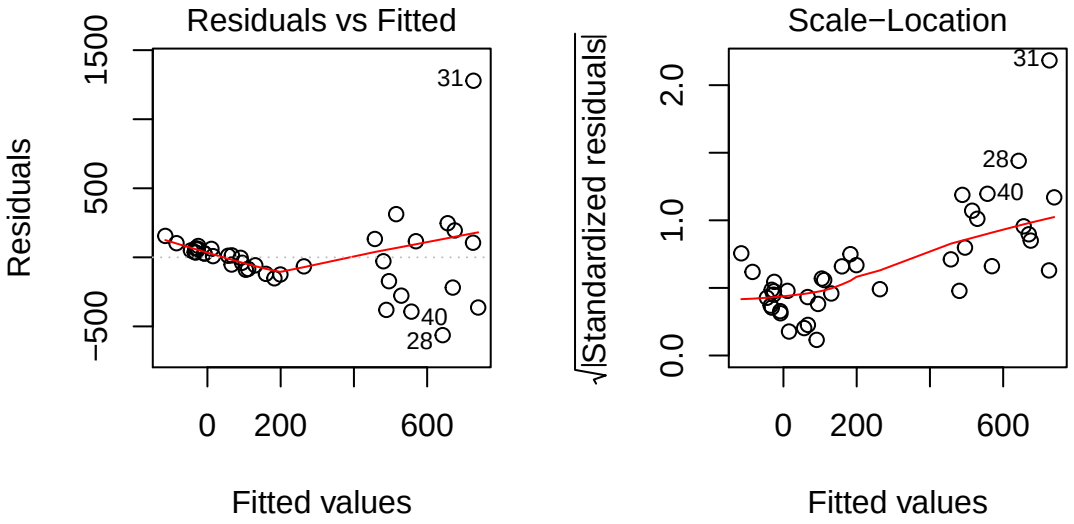
$$e^{B_1} = \frac{\mu(X+1)}{\mu(X)}$$

In our case, $e^{0.0278} = 1.028$ and the interpretation of the exponentiated coefficient is that for increase of 1mg in dose, there is an average of a 2.8% increase in post-treatment cell count, when holding sex, age, and pre-treatment cell count constant.

c) Is there evidence of an effect of dose?

Yes, at $p < 0.05$, there is a significant effect.

d) Are robust SEs necessary?

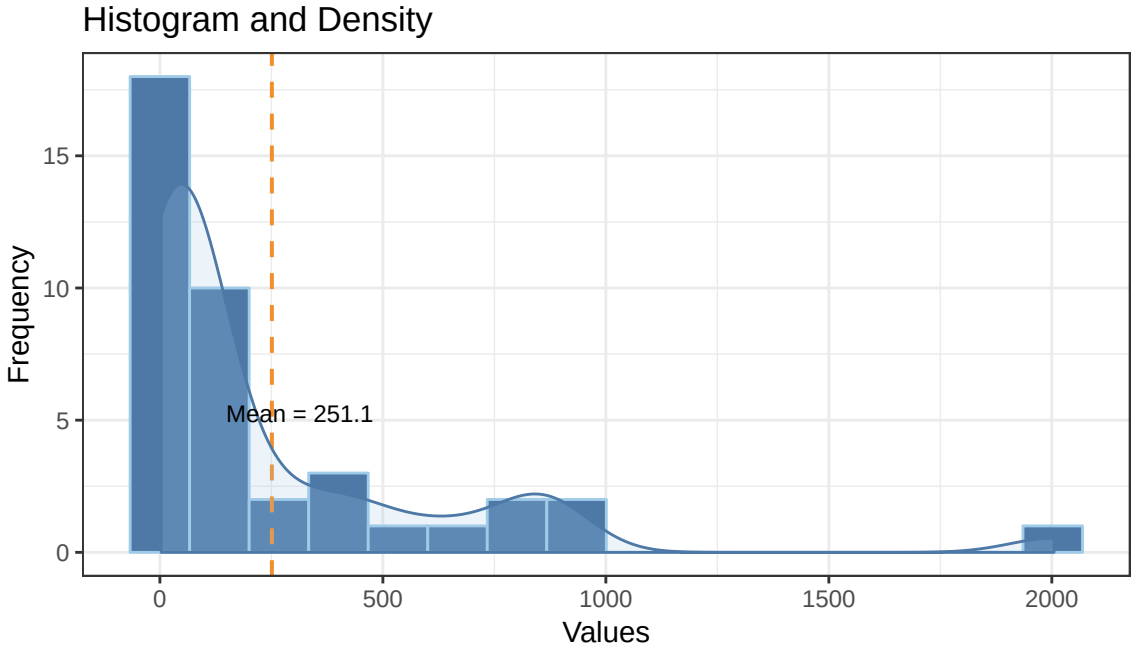


The above uses a regular linear regression and we can see that the data is heteroscedastic so robust SEs are necessary.

e) Assess the assumptions of the Poisson regression model.

Assumptions:

1. The responses are independent - Assumed to be true, but unknown
2. Each y_i is Poisson-distributed with mean $e^{B_0+B_1X_i}$



It does not appear to follow the poisson distribution very closely, but it has the general shape.

f) Compare results with results from the linear regression model for HW 7, Q1.

In HW7, Q1, the linear model has a dose coefficient of 5.81, which is difficult to directly compare to our 2.8% for this question. Both are significant, but clearly, the assumptions matter.

Question 2.

a) Use Poisson regression with dose as a categorical (i.e., factor) variable. Report the coefficient estimate table.

Est	imate Std	. Error z v	alue Pr(	>|z|)
(Intercept)	4.5181	0.0545	82.975	0.0000
factor(dose)10	0.1000	0.0598	1.672	0.0946
factor(dose)100	2.6789	0.0457	58.628	0.0000
sex	0.1935	0.0213	9.076	0.0000
age	-0.0228	0.0007	-33.507	0.0000
count0	0.0054	0.0002	33.572	0.0000

b) Interpret the estimated effects of dose in this model using exponentiated coefficient estimates.

The 0.100 coefficient for 'factor(dose)10' is exponentiated to a 10.5% increase in post-treatment cell count for every 1mg increase in dose, all else held constant between the 0 dose level and 100 dose level. The 2.68 coefficient for 'factor(dose)100' is exponentiated to a 1457% increase in post-treatment cell count for every 1mg increase in dose, all else held constant for the 0 dose level and the 100 dose level.

Question 3.

a) For this question you are to use the "Teeth" data to estimate the association between having at least one tooth extracted (i.e., EXTR > 0) and disease severity (PDALL), with adjustment for age and gender. Estimate this association using logistic regression. Report the coefficient estimate table.

Est	imate Std	. Error z v	alue Pr(	>|z|)
(Intercept)	-8.3603	1.0396	-8.042	0.0000
PDALL	1.4517	0.1671	8.687	0.0000
AGE	0.0304	0.0146	2.079	0.0376
GENDERM	-0.3309	0.2059	-1.607	0.1080

b) Provide an interpretation of the association using an exponentiated coefficient estimate from the logistic regression model.

The exponentialted coefficient for PDALL is 4.27. For every one unit change in PDALL, the exponentiated-odds of at least 1 EXTR increases by 327%, all else held constant.

c) Estimate the association using Poisson regression. Report the coefficient estimate table.

Est	imate Std	. Error z v	alue Pr(	>|z|)
(Intercept)	-6.1123	0.7718	-7.920	0.0000
PDALL	0.9124	0.1035	8.816	0.0000
AGE	0.0199	0.0123	1.612	0.1069
GENDERM	-0.2336	0.1746	-1.337	0.1811

d) Provide an interpretation of the association using an exponentiated coefficient estimate from the Poisson regression model.

The exponentiated coefficient for PDALL is 2.49. For every one unit change in PDALL, EXTR increases by 141%, all else held constant).

e) Compare results from the two models.

Both are significant, but the logistic regression is coming out at 327% compared to 141% for the poisson. Given that the data is binary and not count data, it seems like the logistic makes more sense.