

STAT 511 Homework 10
Due Date: 11:00 A.M., Friday, April 20

1. Consider question 2 from Homework 9. Use the profile likelihood method to find an approximate 95% confidence interval for μ .
2. The following parts each ask you to come up with expressions for the Poisson case (see slide 86 of the slides on generalized linear models for a precise statement of the model) that are analogous to expressions in your notes on generalized linear models for the binomial response case.
 - (a) Write down the general formula for the likelihood ratio statistic that compares the fit of the Poisson model to a saturated model with one parameter for each count. This is known as the residual deviance statistic for the Poisson case.
 - (b) Write down the general expression for the i th deviance residual in the Poisson case.
 - (c) Write down the general expression for the i th Pearson residual in the Poisson case.
 - (d) Suppose there is overdispersion in the Poisson case. What would we assume about the variance of y_i if we decide to take a quasi-Poisson approach?
3. Do the following counts seem like they might be an independent and identically distributed sample from one Poisson distribution? Explain why or why not.

15, 9, 15, 23, 14, 18, 5, 7, 12, 11

4. Consider an experiment designed to compare the resistance of three plant genotypes (A , B , and C) to a fungal pathogen. Eight plants of each genotype were infected with the pathogen. After 24 hours, a leaf from each plant was sampled and examined under a microscope. The number of infected plant cells was recorded for each leaf. Data are provided below. Is there evidence of a difference in resistance? Analyze this data and explain your conclusions to the researchers.

Genotype	Number of Infected Cells for Each Plant
A	39 31 43 31 34 36 34 24
B	23 28 24 19 16 20 25 12
C	36 38 33 22 23 17 29 16

5. For the case of logistic regression, we discussed in class how to interpret estimated parameters in terms of multiplicative effects on odds. Describe how to interpret parameters in the Poisson case (slide 86 of the slides on generalized linear models).
6. Examine slide 88 of the slides on generalized linear models. Determine the functions a , b , and c and the parameters θ_i and ϕ when y_i has the following distributions.
 - (a) $N(\mu_i, \sigma^2)$
 - (b) Binomial with m trials (known constant) and success probability π_i
 - (c) $Poisson(\mu_i)$
7. Consider an experiment with three treatments (A , B , and C). Suppose there are 10 experimental units for each of treatments A and B . Suppose there are 50 experiment units for treatment C . Imagine that the response for each experimental unit has a binomial distribution with m trials (same for all experiment units) and a success probability that depends on treatment. Suppose that (unknown

to the researcher) the success probabilities for treatments A , B , and C are 0.5, 0.5, and 0.95, respectively. Rather than using logistic regression for analysis, a researcher decides to use a standard three-treatment ANOVA assuming a normal response [$\text{lm}(y \sim \text{trt})$]. The researcher is primarily interested in a comparison of treatments A and B , so he examines the R output for the coefficients to see if the “trtB” (the name R would use) coefficient is significant because he knows that provides a test for the difference in treatment A and B means due to the set first to zero constraints. Explain very carefully and specifically why this might not be a safe analysis strategy.

8. This is essentially computational exercise 16 from *The Statistical Sleuth* by Ramsey and Schafer. Some sociologists suspect that highly publicized suicides may trigger additional suicides. In one investigation of this hypothesis, a researcher collected information about 17 airplane crashes that were known (because of notes left behind) to be murder-suicides. (That means that the pilot intentionally crashed the plane to kill him or herself and the passenger(s).) For each of these crashes, the researcher reported an index of the news coverage and the number of multiple-fatality plane crashes during the week following the publicized crash. The data are available at

<http://www.public.iastate.edu/~dnett/S511/PlaneCrashes.txt>

Is there evidence of an association between the news coverage index and the number of crashes in the following week? Conduct an analysis to address this question.