

STAT 510 Homework 13

No Due Date: Ungraded

1. In 1846, a group of pioneers traveling west became stranded in the eastern Sierra Nevada mountains. By the time the last survivor was rescued in the spring of 1847, 40 of 87 members in the original group had died from starvation and exposure to extreme cold. The group became known as the Donner Party. The dataset

<http://dnett.github.io/S510/Donner.txt>

contains the age, sex, and status (survived or died) of the members of the group that were 15 years of age or older. Conduct an analysis of this data set to determine how age and sex are associated with the probability of survival. Support your answer with appropriate tests and confidence intervals. State your conclusions in ways that will be easily interpretable by nonstatisticians.

2. 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

3. 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. The smaller the number of infected cells the more resistant a plant tends to be to the fungal pathogen. Data are provided below. Is there evidence of a difference in resistance among the genotypes? Analyze these 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

4. 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 $m = 20$ 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 why this might not be a safe analysis strategy.
5. This is essentially Computational Exercise 16 from Chapter 22 of *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://dnett.github.io/S510/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.