Homework 5: Categorical data analysis

STAT218

2024-03-14

1. [L6] **Vu and Harrington 8.6**. Greece has faced a severe economic crisis since the end of 2009. A Gallup poll surveyed 1,000 randomly sampled Greeks in 2011 and found that 25% of them said they would rate their lives poorly enough to be considered “suffering”.
   1. Describe the population parameter of interest. What is the value of the point estimate of this parameter?
   2. Check if the conditions required for constructing a confidence interval based on these data are met.
   3. Construct a 95% confidence interval for the proportion of Greeks who are “suffering”.
   4. Without doing any calculations, describe what would happen to the confidence interval if we decided to use a higher confidence level.
   5. Without doing any calculations, describe what would happen to the confidence interval if we used a larger sample.
2. **Vu and Harrington 8.14**. On June 28, 2012 the U.S. Supreme Court upheld the much debated 2010 healthcare law, declaring it constitutional. A Gallup poll released the day after this decision indicates that 46% of 1,012 Americans agree with this decision. At a 95% confidence level, this sample has a 3% margin of error. Based on this information, determine if the following statements are true or false, and explain your reasoning.
   1. We are 95% confident that between 43% and 49% of Americans in this sample support the decision of the U.S. Supreme Court on the 2010 healthcare law.
   2. We are 95% confident that between 43% and 49% of Americans support the decision of the U.S.Supreme Court on the 2010 healthcare law.
   3. If we considered many random samples of 1,012 Americans, and we calculated the sample proportions of those who support the decision of the U.S. Supreme Court, 95% of those sample proportions will be between 43% and 49%.
   4. The margin of error at a 90% confidence level would be higher than 3%.
3. [L7, L8] **Vu and Harrington 8.24**. Researchers studying the link between prenatal vitamin use and autism surveyed the mothers of a random sample of children aged 24 - 60 months with autism and conducted another separate random sample for children with typical development. The table vitamin shows the number of mothers in each group who did and did not use prenatal vitamins during the three months before pregnancy (periconceptional period).
   1. State appropriate hypotheses to test for association between use of prenatal vitamins during the three months before pregnancy and autism.
   2. Complete the hypothesis test and state an appropriate conclusion.
   3. Estimate the association between use of prenatal vitamins and autism using an appropriate measure (relative risk or odds ratio). Produce a point estimate and a confidence interval.
   4. Write a narrative summary of your results from (a) – (c).
   5. A New York Times article reporting on this study was titled “Prenatal Vitamins May Ward Off Autism”. Do you find the title of this article to be appropriate? Explain your answer. Additionally, propose an alternative title.
4. **Vu and Harrington 8.27**. Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.
   1. The chi-square model, just like the normal model, has two parameters, mean and standard deviation.
   2. The chi-square model is always right skewed, regardless of the value of the degrees of freedom parameter.
   3. The chi-square statistic is always positive.
   4. As the degrees of freedom increases, the shape of the chi-square model becomes more skewed.
5. [L7, L8] **Vu and Harrington 8.45**. In a study examining the association between green tea consumption and esophageal carcinoma, researchers recruited 300 patients with carcinoma and 571 without carcinoma and administered a questionnaire about tea drinking habits. Out of the 47 individuals who reported that they regularly drink green tea, 17 had carcinoma. Out of the 824 individuals who reported they never drink green tea, 283 had carcinoma. These data are stored in the table tea.
   1. Analyze the data to assess evidence for an association between green tea consumption and esophageal carcinoma from these data.
   2. Report and interpret a point estimate and conficence interval for an appropriate measure of association.
   3. Write a narrative summary of your results in (a)-(b).
6. [L8] Volunteer patients were randomized into one of two experiment groups where they would receive an experimental vaccine or a placebo. They were subsequently exposed to a drug-sensitive strain of malaria and observed to see whether they came down with an infection. The raw data from the trial are stored as the dataset openintro::malaria.
   1. Test for association between the vaccination and the probability of infection. (*Hint*: take note of the small counts.)
   2. Estimate the difference in probability of infection between the vaccine group and the placebo group. Provide a 95% confidence interval and interpret the interval in context.
   3. Estimate the relative risk of infection in the vaccine group compared with the placebo group. Provide a 95% confidence interval and interpret the interval in context.
   4. Vaccine efficacy is often measured by the relative *reduction* in risk, estimated by . Provide a point estimate of vaccine efficacy from the trial data.
   5. Provide a 95% confidence interval for vaccine efficacy based on your answers in (c)-(d). (*Hint*: if are a confidence interval for parameter , then an interval for is given by .)