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.

* ANSWERS:
  1. Proportion of Greeks who are suffering. Point estimate is 0.25.
  2. Conditions are met. Don’t evaluate this question, as we did not emphasize checking conditions in class.
  3. CI: (0.223, 0.277). Lower bound might be 0.222 if empirical rule is used instead of 1.96.
  4. It would be wider because a larger critical value would be used.
  5. Assuming the point estimate is similar, it would be narrower because the standard error and therefore the margin of error is inversely proportional to the sample size.

1. **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%.
2. [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.

* ANSWERS:
  1. : taking prenatal vitamins and autism are independent. : taking prenatal vitamins and autism are not independent. Hypotheses could also be stated as association/nonassociation.
  2. The data provide sufficient evidence against the null hypothesis of no association between taking prenatal vitamins and autism ( = 8.865 on 1 degree of freedom, = 0.002907). Students need only provide the -value.
  3. With 95% confidence, the odds of autism are estimated to be 18.5% to 61.1% lower when prenatal vitamins are taken during the periconceptional period, with a point estimate of 43.2%. (The CI for the OR is (0.389, 0.825); the point estimate is 0.568.)
  4. Evidence suggests that taking prenatal vitamins during the periconceptional period may be associated with a lower likelihood of child autism. Data on prenatal vitamin use among mothers of a random sample of autistic children aged 24-60 months and mothers of a random sample of non-autistic children aged 24-60 months provide sufficient evidence to reject the null hypothesis of no association between vitamin use and the odds of autism ( = 8.865 on 1 degree of freedom, = 0.002907). With 95% confidence, the odds of autism are estimated to be 18.5% to 61.1% lower when prenatal vitamins are taken during the periconceptional period, with a point estimate of 43.2%.
  5. The title is misleading because it implies a causal relationship. A better title would be something causally neutral such as “Evidence suggests an association between prenatal vitamin use and likelihood of autism”.

1. **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.
2. [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).

* ANSWERS:
  1. The data do not provide sufficient evidence to reject the null hypothesis of no association between green tea consumption and esophageal carcinoma ( *= 0.065626* on 1 degree of freedom, *p = 0.7978155*).
  2. With 95% confidence, the odds of esophageal carcinoma among green tea drinkers are estimated to be between 49.8% lower and 73% higher than the odds among non-tea-drinkers, with a point estimate that the odds are 8.1% lower. (CI for OR is (0.5028, 1.7364), point estimate is 0.9196.)
  3. There is not sufficient evidence to suggest an association between green tea consumption and the likelihood of esophageal cancer. Data on 47 tea drinkers and 824 non-tea-drinkers do not provide sufficient evidence to reject the null hypothesis of no association ( *= 0.065626* on 1 degree of freedom, *p = 0.7978155*). Although it is estimated that the odds of esophageal carcinoma are 8.1% lower among tea drinkers, at the 95% confidence level, the odds are estimated to be anywhere between 49.8% lower and 73% higher.

1. [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 .)

* ANSWERS:
  1. The data provide sufficient evidence to reject the null hypothesis of no association between vaccination and infection after exposure to malaria (Fisher’s exact test, *p = 0.01408669*).
  2. With 95% confidence, the probability of infection among vaccinated individuals is estimated to be between 0.3919 and 0.8939 less than the probability of infection among unvaccinated individuals, with point estimate of 0.6429.
  3. With 95% confidence, the risk of infection among vaccinated individuals is estimated to be between 27.1% and 82.3% lower than among unvaccinated individuals, with a point estimate of 64.3%. Alternatively: with 95% confidence, the relative risk of infection after exposure to malaria among vaccinated individuals compared with unvaccinated individuals is estimated to be between 0.177 and 0.721 with a point estimate of 0.357. If students give a relative risk comparing unvaccinated to vaccinated, this is acceptable, but a comment should be left.
  4. The malaria vaccine is estimated to be 64.3% effective at preventing infection.
  5. With 95% confidence, the malaria vaccine is estimated to be between 27.1% and 82.3% effective at preventing infection.