

# Categorical Data Analysis

## Assignment #1

---

1. Identify each variable as nominal or ordinal.
  - (a) UK political party preference (Labour, Conservative, Social Democrat)
  - (b) Anxiety rating (none, mild, moderate, severe, very severe)
  - (c) Clinic location (London, Boston, Madison, Rochester, Montreal)
  - (d) Response of tumor to chemotherapy (complete elimination, partial reduction, stable, growth progression)
  - (e) Favorite beverage (water, juice, milk, soft drink, beer, wine)
2. Consider the statement, “Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children.” For the 1996 General Social Survey, conducted by the National Opinion Research Center (NORC), 842 replied “yes” and 982 replied “no”. Let  $p$  denote the population proportion who would reply “yes”. Construct a 95% confidence interval for  $p$ . Interpret the result.
3. Suppose that, to collect data in an introductory statistics course, recently I gave the students a questionnaire. One question asked each student whether he or she was a vegetarian. Of  $n = 25$  students, 5 answered “yes”. For testing  $H_0 : p = 0.5$  versus  $H_a : p \neq 0.5$ ,
  - (a) Provide the Pearson’s chi-squared statistic,  $X^2$
  - (b) Provide the likelihood ratio statistic,  $G^2$
  - (c) Conduct hypothesis test for  $H_0$  with both of  $X^2$  and  $G^2$ . (This means you should report p-values for them)
4. The below table shows the results of a study comparing radiation therapy with surgery in treating cancer of the larynx. The response indicates whether the cancer was controlled for at least two years following treatment.

	Cancer Controlled	Cancer not controlled
Surgery	63	6
Radiation therapy	45	9

Researchers are interested in comparing the proportions of cancer control between two methods. Answer the following questions

- (a) Let us denote the proportion of cancer control with surgery by  $p_1$  and that with radiation therapy by  $p_2$ . What are  $\hat{p}_1$  and  $\hat{p}_2$ ?
- (b) Give a 95% confidence interval for  $p_1 - p_2$ .
- (c) Compute the relative risk of cancer control for group “Surgery” relative to group “Radiation therapy” and interpret it.
- (d) Compute the odds of cancer control for group “Surgery” and the odds of cancer control for group “Radiation therapy”.
- (e) Provide the odds ratio of cancer control for group “Surgery” relative to group “Radiation therapy”. And interpret it.
- (f) Give a 95% confidence interval for the odds ratio.
- (g) Now we are interested in the independence between method and response. Based on the result of odds ratio, are they independent?
- (h) We are still interested in the independence. Conduct  $X^2$  and  $G^2$  tests. Interpret the results.

5. A study from the University of Texas Southwestern Medical Center examined whether the risk of hepatitis C was related to whether people had tattoos and to where they got their tattoos. Hepatitis C causes about 10,000 deaths each year in the United States, but often lies undetected for years after infection. The data from this study can be summarized in a two-way table, as follows:

	Hepatitis C	No Hepatitis C	Total
Tatto, parlor	17	35	52
Tatto, elsewhere	8	53	61
None	22	491	513
Total	47	579	626

- (a) Using  $X^2$ , test the hypothesis of independence between tattoo status and Hepatitis C infection status. Report the p-value and interpret.
- (b) Using  $G^2$ , test the hypothesis of independence between tattoo status and Hepatitis C infection status. Report the p-value and interpret.

6. **(SAS problem)** The table below lists results from a simple random sample of front-seat occupants involved in car crashes (based on data from “Who Wants Airbags?” by Meyer and Finney, *Chance*, Vol. 18 No. 2). Use a 0.05 significance level to test the claim that the fatality rate of occupants is different for those in cars equipped with airbags.

	Occupant Fatalities	Occupant Survivals	Total
Airbag Available	41	11,500	11,541
No Airbag Available	52	9,801	9,853

Using SAS system, answer the followings:

- (a) Provide SAS DATA step to create a data set.
- (b) Provide relative risk of fatality for group “Airbag Available” relative to group “No Airbag Available” and its 95% confidence interval. Interpret relative risk and its confidence interval.
- (c) Provide odds of fatality for each group and interpret them.
- (d) Provide odds ratio of fatality for group “Airbag Available” relative to group “No Airbag Available” and 95% confidence interval. Interpret them.
- (e) Provide  $X^2$  and  $G^2$  test results for testing whether airbag status and fatality are associated and interpret them.

Append necessary SAS code should be attached in the end of the answer of each question.