

Project 10 - Categorical Data Analysis

Statistics 401: Fall 2006

Due Wednesday, November 29

Use R to complete this project. Attach all R commands used to complete this project in an appendix. Annotate with the problem number.

1. Read the *Bozeman Daily Chronicle Article* from October 27, 2006, “Attorneys argue over unsealing homicide documents” available at the STAT401 web site. The following table summarizes the claims given by the defendant’s attorney’s:

Had not Heard of Case	Heard of Case and Guilty	Heard of Case and No Opinion	Total
20	148	233	401

Use a Goodness of Fit Test to determine whether the percentage of people who have heard of the case is 90%, and that these people are evenly split on whether or not the defendants are guilty.

- (a) Construct a bar chart to display the data. Include the plot in your report.
 - (b) State the appropriate hypotheses to be tested.
 - (c) Check the assumptions.
 - (d) Display a table of the expected counts. Label it and refer to it from the text of your report.
 - (e) Display a table of the χ^2 contributions. Label it and refer to it from the text of your report.
 - (f) Give the test statistic value.
 - (g) Give the distribution of the test statistic assuming H_0 is true.
 - (h) Give the p -value
 - (i) Make a decision at $\alpha = .05$.
 - (j) Draw a conclusion in the context of this problem.
 - (k) Is a follow-up necessary? If not, explain why. If so, do the follow-up. Explain what output you are using and specify which proportion(s) appears to be different than its hypothesized value.
2. In November, 1990, the *Student Right To Know and Campus Security Act* was signed into law. This law has been expanded in subsequent years and is widely referred to as the *Clery Act*. The Act mandates that institutions of higher education report and make available to students and employees the reported occurrences of specific types of crime which occur on campus. The following table, excerpted from the report *Montana State University - Bozeman Calendar Year 2002 Annual Crime Report*, gives the number of reported Burglaries and Forcible Sex Offences for three different years:

OFFENSE	2000	2001	2002
Burglary	12	10	10
Sex-Offense-Forcible	5	9	11

Is there an association between the type of offense and the year?

- (a) Construct a bar chart to display the data. Include the plot in your report.
- (b) State the appropriate hypotheses to be tested.

- (c) Check the assumptions.
 - (d) Display a table of the expected counts. Label it and refer to it from the text of your report.
 - (e) Display a table of the χ^2 contributions. Label it and refer to it from the text of your report.
 - (f) Give the test statistic value.
 - (g) Give the distribution of the test statistic assuming H_0 is true.
 - (h) Give the p -value
 - (i) Make a decision at $\alpha = .05$.
 - (j) Draw a conclusion in the context of this problem.
 - (k) Is a follow-up necessary? If not, explain why. If so, do the follow-up. Explain what output you are using and specify which categories exhibit the association.
3. In March 25, 2005, *Animal Behavior* published Creel and Winnie's "Responses of elk herd size to fine scale spatial and temporal variation in the risk of predation by wolves." In the paper, Creel and Winnie are studying how far from the timber elk herds aggregate when there are wolves present. The following table shows the numbers of bulls, cows and calves at three different ranges of distance to the timber when there are wolves present (from Table 1 in the paper):

	0-30	31-100	101-300	> 300	Totals
Bulls	479	493	158	18	1148
Cows	454	517	192	53	1216
Calves	86	108	25	9	228

Are the distributions of the distances from the timber the same for all three age populations of elk?

- (a) Construct a bar chart to display the data. Include the plot in your report.
- (b) State the appropriate hypotheses to be tested.
- (c) Check the assumptions. Assume that Creel and Winnie's data is from SRS's of calves, cows and bulls.
- (d) Display a table of the expected counts. Label it and refer to it from the text of your report.
- (e) Display a table of the χ^2 contributions. Label it and refer to it from the text of your report.
- (f) Give the test statistic value.
- (g) Give the distribution of the test statistic assuming H_0 is true.
- (h) Give the p -value
- (i) Make a decision at $\alpha = .05$.
- (j) Draw a conclusion in the context of this problem.
- (k) Is a follow-up necessary? If not, explain why. If so, do the follow-up using the two highest χ^2 contributions. Explain what output you are using and specify which population(s) and which proportion(s) are different than the others.