Transition to College Mathematics and Statistics

Christian R. Hirsch • Eric W. Hart • Ann E. Watkins • Beth E. Ritsema

with

James T. Fey • Brin A. Keller • Rebecca K. Walker • James K. Laser

Unit 1: *Interpreting Categorical Data*

Transition to College Mathematics and Statistics (TCMS) consists of eight coherent and focused units with deliberate connections among topics across units. Each unit is comprised of two to four problem-based, inquiry-oriented, and technology-rich multi-day lessons. Each lesson consists of two to four related investigations emphasizing mathematical modeling and important mathematical practices and habits of mind.

Units culminate with a "Looking Back" lesson intended for students to review and synthesize their understanding of key ideas developed in the unit. As such, the following "Looking Back" lessons for each TCMS unit provide potential users an overview of our approach to important mathematical ideas and the expectations and nature of collaborative student work. Preceding each "Looking Back" lesson is the table of contents for the unit.

Table of Contents

Lesson **4** Looking Back

In this unit, you learned the meaning of terms often used in the media to compare the proportion of people in different groups who have some characteristic. These terms include absolute risk, absolute risk reduction, relative risk, and statistically significant difference. You also learned why it is important not to rely on anecdotal evidence but to look for evidence from well-designed experiments. You used the chi-square statistic to decide whether the difference in the proportions that fall into each category in two random samples is statistically significant. Finally, you used a chi-square test to decide whether it is plausible that two categorical variables are independent in the population from which the random sample was taken.

The tasks in this final lesson will help you review and solidify your understanding of key ideas and methods for making sense of categorical data.

1. About 62% of Americans have a pet, with women more likely than men to have one. A Harris Poll of 600 male adult pet owners and 934 female pet owners asked if they considered their pet to be a member of the family. Eighty-five percent of the males said yes, 12% said no, and 3% were not sure. Of the female pet owners, 95% said yes, 3% said no, and 2% were unsure. You may consider these independent random samples of male and female pet owners. (Source: "Pets Really Are Members of the Family," The Harris Poll, June 10, 2011.)

   a. Summarize the information in a two-way table of observed frequencies. Let the columns be men and women pet owners.

   b. Make a bar graph that best compares the proportion of men and women pet owners who gave the different responses. Does the difference in the responses appear to be significant or do the two groups appear almost homogeneous?

   c. To show what homogeneous samples of male and female pet owners would have looked like, make a table of expected frequencies. Round to the nearest tenth.

   d. Compute χ^2 to summarize the difference between male and female responses.

   e. How many categories are there for each sample? What critical value should be used for comparison?

   f. Is the difference in the proportions of male and female pet owners who gave the different responses statistically significant? Explain how you know, and what it means in this context.

② Tonsils, at the back of the throat, are organs of the lymphatic system. They may help the immune system fight disease. Sometimes, when they are chronically inflamed, doctors remove them, called a tonsillectomy. Sweden keeps a register of surgery done on people under the age of 20. During a fifteen-year period, 27,284 young people had a tonsillectomy. For each of them, five young people the same age, sex, and county of residence were randomly selected to serve as controls. The control group ended up with 136,401 young people. After about 23 years, 47 of the tonsillectomy group and 169 of the control group had had a premature heart attack. (Source: Imre Janszky, *et al*. "Childhood Appendectomy, Tonsillectomy, and Risk for Premature Acute Myocardial Infarction—A Nationwide Population-Based Cohort Study," *European Heart Journal*, online access June 1, 2011.)

   a. What is the explanatory variable in this study? What is the response variable? Was this study an experiment? Explain why or why not.

   b. Make a table of observed frequencies.

   c. What is the absolute risk for each group? Why is the incidence of heart attack so low in both groups?

   d. Compute the reduction in absolute risk of a heart attack. Use this in a sentence for parents.

   e. Compute the relative risk of a heart attack. Use this in a sentence for parents.

   f. Assume that you can consider the two groups equivalent to randomly selected samples from all young people. Compute χ^2.

   g. Is the difference in the proportions of people in the two groups who have heart attacks statistically significant? Explain how you know.

   h. Write a short article for your school newspaper about this study. Use and explain the terms absolute risk, relative risk, and statistical significance.

③ Post-traumatic stress disorder (PTSD) is a highly anxious state that can develop after exposure to psychological trauma. Researchers at the U.S. Department of Veterans Affairs (VA) collected data on 852 veterans who screened positive for PTSD. The table on the following page shows some of their results. You may consider these a random sample of all veterans diagnosed with PTSD.

	Received Minimally Adequate Treatment	Did Not Receive Minimally Adequate Treatment	Total
Served In Iraq or Afghanistan	115	280	395
Served Elsewhere	165	292	457
Total	280	572	852

Source: Mary W. Lu, *et al.* "Correlates of Utilization of PTSD Specialty Treatment Among Recently Diagnosed Veterans at the VA," *Psychiatric Services*, Vol. 62, 2011, pp. 943–949. Frequencies estimated from percentages.

a. If you select one of these 852 veterans at random, what is the probability that he or she served in Iraq or Afghanistan? What is the probability that he or she served in Iraq or Afghanistan given that he or she received minimally adequate treatment?

b. According to the mathematical definition of independent events, are the events *served in Iraq or Afghanistan* and *received minimally adequate treatment* independent in this sample? Explain what your conclusion means in practical terms.

c. Using the marginal totals in the table above, complete a table of expected frequencies for a chi-square test of independence. Round to the nearest whole number.

d. Compare the expected frequencies and the observed frequencies. Do the differences seem relatively large or small?

e. Compute the chi-square statistic χ^2.

f. To what critical value should your value of χ^2 be compared? Is the value of χ^2 statistically significant?

g. Write a conclusion that can be drawn from your analysis.

Summarize the Mathematics

Reports in the media often contrast two groups or discuss the association between two conditions.

a) What is the difference between absolute risk reduction and relative risk?

b) What are the characteristics of a well-designed experiment? Why is it imperative to do an experiment rather than rely on anecdotal evidence when deciding how well a treatment, medical or otherwise, works?

c) Describe the importance of each of the following in an experiment: control group, placebo, single blind, double blind.

d) What does it mean for two groups to be homogeneous? What does the stacked bar graph (percent on the vertical axis) look like if the groups are homogeneous?

e) In a test of homogeneity, what does χ^2 measure? How do you use proportional reasoning to compute expected frequencies?

f) Suppose you have two random samples, each classified into the same categories. How can you tell whether the difference in the proportions that fall into each category is statistically significant? What does it mean when the difference is statistically significant?

g) What statistics can be used to evaluate the effectiveness of a screening test? Which is the one you would want to know if you tested positive for some condition? Which is the one you would want to know if you tested negative?

h) What does it mean if two categorical variables are independent? How can you use the definition of independence to compute expected frequencies for a chi-square test of independence?

i) Describe the similarities and differences between a chi-square test of homogeneity and a chi-square test of independence.

Be prepared to share your ideas and reasoning with the class.

✓ Check Your Understanding

Write, in outline form, a summary of the important statistical concepts and methods developed in this unit. Organize your summary so that it can be used as a quick reference in your future work.