Association of Categorical Variables

MATH 360 Homework for Pre-Service Teachers

1) The chi-square test statistic is defined as the sum of for all cells.

A. Define what O and E represent in this formula & how to find each.

B. Explain why the chi-square test statistic is defined this way. Include reasoning for every aspect of this formula, including the difference of *O-E*, the squaring of this difference, the division by *E,* and why these values are summed.

C. Look at a chi-square distribution when the degrees of freedom is 6. What is its shape? Why?

D. Describe the shape of the chi-square distribution. Include graphs of the distribution for various degrees of freedom. How does its shape change as the degrees of freedom increases? Why?

E. What is the smallest possible value chi-square can be in any situation? Why?

2)Adapted from Burrill, G., Franklin, C., Godbold, L., & Young, L. (2003). *Navigating through data*

*analysis in grades 9-12.* Reston, VA: NCTM.

Scenario: A bank had 48 candidates for promotion, with equal numbers of males and females. A total of 35 people were recommended for promotion.

A. Suppose that the recommendations of male & female candidates showed no evidence of gender discrimination. Fill in the table below with a possible scenario that shows this.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Recommended for  Promotion | Not Recommended for  Promotion | Total |
| Male |  |  | 24 |
| Female |  |  | 24 |
| Total | 35 | 13 | 48 |

B. Now suppose that the recommendations for promotion showed strong evidence of discrimination against the female candidates for promotion. Complete the table to show a possible example of this case.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Recommended for  Promotion | Not Recommended for  Promotion | Total |
| Male |  |  | 24 |
| Female |  |  | 24 |
| Total | 35 | 13 | 48 |

C. Suppose the evidence of discrimination against the women fell into a “gray” area, making any discrimination against the women not clearly obvious without further study. Complete the table below to show such a case.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Recommended for  Promotion | Not Recommended for  Promotion | Total |
| Male |  |  | 24 |
| Female |  |  | 24 |
| Total | 35 | 13 | 48 |

D. It turned out that at the actual bank, of the 24 females, 14 candidates were recommended for promotion, and 21 of the 24 male candidates were recommended for promotion. Complete the table below with this actual data.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Recommended for  Promotion | Not Recommended for  Promotion | Total |
| Male |  |  | 24 |
| Female |  |  | 24 |
| Total | 35 | 13 | 48 |

E. Some of the female employees felt like there was gender discrimination in this decision. Set up and carry out a simulation to investigate this claim. Note: Do not investigate the claim through a hypothesis test.

F. Based on your simulation results, is there an association between gender and promotion recommendation at this bank? Back this up with numerical justification based on your simulation.

G. Suppose a similar study at another company showed a clearly strong, statistically significant association between gender and promotion recommendation. Pretend you are the company’s lawyer. Argue that this is not sufficient evidence that gender discrimination took place.

3) Adapted from College Board (2009). *AP Statistics Free-Response Questions.* New York, NY: Author.

A simple random sample of 100 high school seniors was selected from a large school district. The gender of each student was recorded, and each student was asked the following questions:

1. Have you ever had a part-time job?

2. If you answered yes to the previous question, was your part-time job in the summer only?

The responses are summarized in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| Job Experience | Male | Female | Total |
| Never had a part-time job | 21 | 31 | 52 |
| Had a part-time job during summer only | 15 | 13 | 28 |
| Had a part-time job but not only during summer | 12 | 8 | 20 |
| Total | 48 | 52 | 100 |

(a) Construct a graphical display that represents the association between gender and job experience for the students in the sample. You may use the grid below or paste a graphic in from a program of your choice.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

(b) Write a few sentences summarizing what the display in part (a) reveals about the association between gender and job experience for the students in the sample.

(c) Give 3 different numerical comparisons you can make using the data in the table to determine if there is an association between gender & job experience for the students in the sample.

(d) Which test of significance should be used to test if there is an association between gender and job experience for the population of high school seniors in the district?

State the null and alternative hypotheses for the test, but do not perform the test.

4) A student named Martha comes to you for help. She says she knows you use chi-square tests with categorical data, but she doesn’t know when to use which one (goodness of fit, independence, & homogeneity). Write below your response to Martha as her teacher, perhaps including examples to illustrate your points. Be specific in your description of your response, writing exactly what you would say and/or drawing anything you would use in your response.

5) The Common Core State Standards for Mathematics includes the following standard: “Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies).Recognize possible associations and trends in the data.”

Research has found that students commonly have three incorrect conceptions about association of categorical variables.

* Determinist: students believed that an association meant all cases must show an association with no exceptions. These students believed that the cells in the two-way table that did not agree with the association should have zero frequency.
* Unidirectional: students believed dependence occurred only when it was direct. This could be explained by the tendency of students to give more relevance to positive cases than negative cases that confirm a given hypothesis.
* Localist: students looked at part of the data to determine if an association existed, often only looking at the cell with the highest frequency or at only one conditional distribution.

Design a classroom lesson focused on the given standard which addresses these misconceptions. You do not need to create homework or exam questions; only the classroom lesson.

6) You are leading your students in an activity in which they complete all 4 steps of the GAISE framework (formulate a research question, collect their own data, analyze the data, and interpret the results) in a situation relevant to chi-square tests. Students have brainstormed the following research questions. For each question, determine if it is relevant to chi-square tests and provide your response to the student. Consider data collection issues as well. For those that are not relevant and/or have data collection issues, suggest some modifications.

LAURA: I want to study if girls are smarter than boys, so I am going to compare the GPAs of boys and girls at school.

BILL: I think people with higher GPAs are less likely to have had a car crash while driving. I am going to ask a senior with drivers licenses their GPAs and if they’ve crashed while driving.

ANNE MARIE: I want to see whether boys are more likely to own smartphones than girls. So I’m going to count during passing periods the number of boys I see with smartphones and the number of girls I see with no phones.

NICK: I think girls are more likely to bring their lunch to school than boys. So I’m going to stand at the entrance to the lunchroom and count the number of boys and girls that bring their lunches.

JEFF: I want to study the relationships between gender, race, grade level, political party and whether students make honor roll or not at our school.

KRISTIN: I read that 77% of high school students go on to college. I want to see if that’s true of the students here.