**HOMEWORK**

**Chi-Square Test**

Name: Michael Chambers

**Background**

**Analysis of Categorical Data**: in epidemiological and other types of research, categorical data are often available to test a hypothesis on the association between 2 variables, for example between a disease and a possible risk factor, where both variables fall into discrete categories. The first step is to construct a 2-way classification table consisting of the rows of one variable and the columns of the second variable. This is called a contingency table. If there are 2 rows and 2 columns it is referred to as a “2 by 2” (or 2x2) table, and higher numbers are also possible (e.g. 3 by 2, or 2 by 4, etc… depending on the variables in the study).

**Instructions**

In this exercise, you will calculate and interpret the chi-square test for different types of studies.

***Fill in the attached worksheet and answer to the questions below.***

Problem 1. In a study of infant birth weight and maternal factors, the newborn babies were categorized as being either small size for gestational age (N=201) or normal size (N=2089). The following data were observed in relation to the mother’s educational level and the onset of prenatal care.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Small size** | **Normal size** |
| Maternal education: |  |  |
| < high school | 32 | 111 |
| high school | 94 | 800 |
| college | 75 | 1178 |
| Onset of prenatal care: |  |  |
| < 3rd month of pregnancy | 55 | 145 |
| 3rd month or later | 146 | 1944 |

What proportion of small size infants were born to mothers who did not complete high school?

32/(32+94+75) = 32/201 = .159 = 15.9%

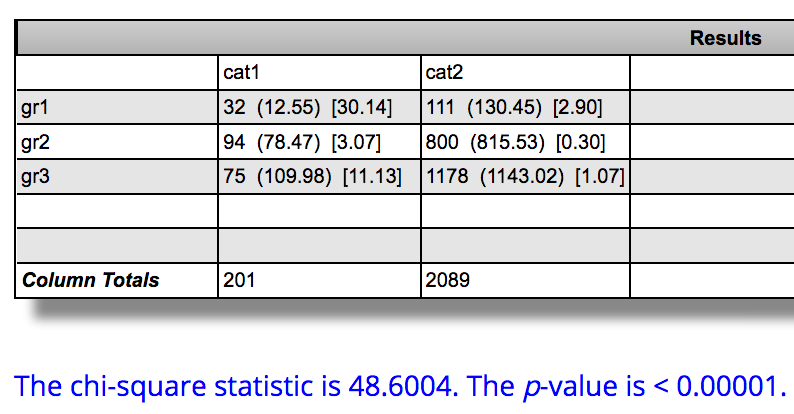
What proportion were born to mothers who completed college, and how does that compare to the proportion of normal size infants born to such mothers?

Proportion of small size infants born to mothers who completed college: 75/201 = .373 = 37.3%

Proportion of normal size infants born to mothers who completed college: 1178/(800+111+1178) = .564 = 56.4%

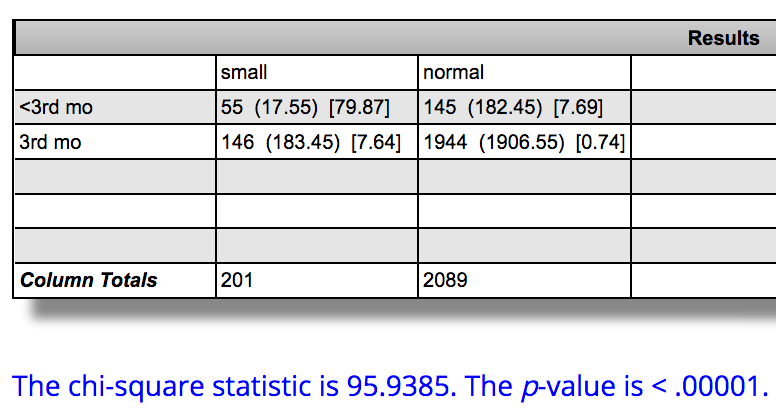
The proportion of normal size children born to college educated mothers is larger than the proportion of small size children born to college educated mothers.

(a) Is there a statistically significant association between maternal educational level and the size of the infant?



Yes, there is a statistically significant association b/w maternal education and the size of the infant.

(b) If the null hypothesis states that the proportion of small size babies is the same for mothers with onset of prenatal care at 3 months or later as it is for mothers who started prenatal care earlier than 3 months, would you accept it or reject it based on the chi-square test? Interpret the result.



I would reject the null hypothesis that the proportion of small size babies is the same for mothers with onset of prenatal care at 3 months or later as it is for mothers who started prenatal care earlier than 3 months. There is a statistically significant association between the onset of prenatal care and the size of the infant.

Problem 2.

A study was conducted to evaluate the relative efficacy of supplementation with calcium versus calcitriol in the treatment of postmenopausal osteoporosis. Calcitriol is an agent that has the ability to increase gastrointestinal absorption of calcium. A number of patients withdrew from this study prematurely due to the adverse effects of treatment, which include thirst, skin problems, and neurologic symptoms. The relevant data appear below.

|  |  |  |  |
| --- | --- | --- | --- |
| Treatment | Withdrawal: yes | Withdrawal: no | Total |
| Calcitriol | 27 | 287 | 314 |
| Calcium | 20 | 288 | 308 |
| Total | 47 | 575 | 622 |

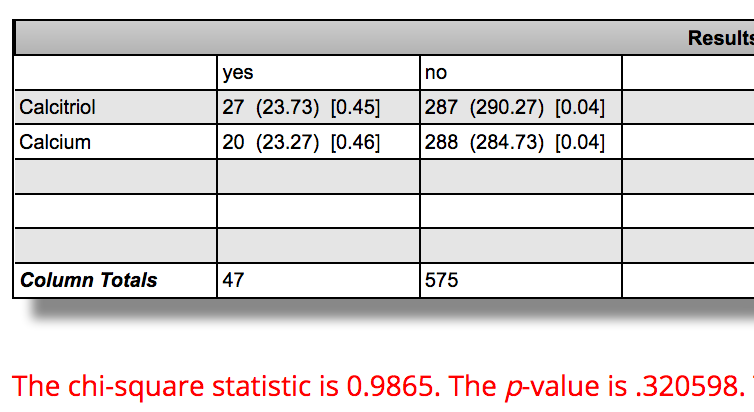
(a) Compute the sample proportion of subjects who withdrew from the study in each treatment group.

Calcitrol: 27/314 = .086 = 8.6%

Calcium: 20/308 = .065 = 6.5%

(b) Test the hypothesis that there is no association between treatment group and withdrawal from the study at the 0.05 level of significance. What do you conclude?

We cannot reject the null hypothesis that the proportion of study participants in the calcitriol treatment group who withdrew is identical to the proportion of participants in the calcium treatment group who withdrew.



Problem 3. In a study of the possible association between age and the preference for a new type of natural sweetener, the manufacturer conducted a survey. They wanted to see if those who preferred the new sweetener to sugar differed from those who liked sugar better, with respect to age. Of the 15 people who liked sugar, 11 were younger than 18 years old, compared to 13 out of 40 sweetener lovers. Based on this study, should the company begin an advertising campaign targeting the young consumers? Why or why not?

Because the sample size is low we will be using a Fisher exact test on this dataset. With a Fisher exact test value of .013 we can conclude that there is a statistically significant difference between the proportions of people who prefer sugar or the sweetener base on age differences and we can reject the null hypothesis. Even though we can conclude that there is a statistically significant difference, it would be nice to get more observations in this sample.

