CHAPTER 13

The Chi Square Test

Summary _____

Chi square ($\chi 2$) is yet another null hypothesis statistical test (NHST). A chi square test results in a conclusion about the relationship between two variables. Chi square is appropriate when the information on a subject or event is not a qualitative score but the category the subject or event is in. Chi square compares the *observed frequencies* (*O*) in the categories to *expected frequencies (E)*. The expected frequencies come from a hypothesis about the population that the sample data are from. In every case, ΣO must equal ΣE.

Two kinds of $\chi 2$ tests were described. In an *independence test*, the null hypothesis is that the two variables are independent. A *rejected* null hypothesis means that the two variables are related; they are not independent. A *retained* null hypothesis means that the data support the conclusion that the variables are independent (but independence is *not* proved).

In a *goodness-of-fit* $\chi 2$ test, the null hypothesis comes from a theory. If the null hypothesis is rejected, conclude that the theory *fails* to be consistent with the data. If the null hypothesis is retained, conclude that the theory is in *accord* with the data (but not that the theory is *proved* correct).

If one or more expected frequency is *small* (around 5 or less), the probabilities given by the chi square distribution may not be accurate. The best solution is to gather sufficient data to avoid this problem. A second solution is to combine categories, which results in fewer categories and larger expected frequencies.

Each $\chi 2$ problem has an associated *degrees of freedom.* For tests of independence, (contingency tables of rows and columns), $df = (R-1)(C-1)$. For goodness-of-fit tests with one variable and two or more categories, degrees of freedom is equal to the number of categories minus one. If other restrictions are added to goodness-of-fit tests, one *df* is subtracted for each parameter estimated. An example of a restriction is that of arranging the mean and standard deviation of the

expected scores so that they are equal to the mean and standard deviation of the observed scores.

The *effect size index* for a 2 x 2 test of independence is phi φ (pronounced "fee"). Small, medium, and large values of φ are 0.10, 0.30, and 0.50 respectively. Chi square tests are appropriate if:

1. the measurement of a subject or event places it into a category.
2. the measurement of one subject or event does not influence the measurement of another subject or event (the measurements must be independent).
3. samples are representative of the populations that the conclusions apply to.

Multiple-Choice Questions _____

1. The person who developed χ^2 was
   (1) Ronald A. Fisher;
   (2) William S. Gosset;
   (3) Karl Pearson;
   (4) Helen Walker.

2. The shape of the theoretical χ^2 distribution is determined by
   (1) the number of observations;
   (2) the size of the expected frequency of events;
   (3) the number of categories of events;
   (4) all of the above.

3. The sum of the expected frequencies must be equal to
   (1) the sum of the observed frequencies;
   (2) the *df*;
   (3) (R – 1)(C – 1);
   (4) none of the above.

4. A developmental psychologist developed a theory that predicted the proportion of children who would, in a period of stress, cling to the mother, attack the mother, or attack a younger sibling. The stress situation was set up and the responses of 50 children were recorded. The appropriate χ^2 test is a test of
   (1) goodness of fit with 2 *df*;
   (2) goodness of fit with 40 *df*;
   (3) independence with 2 *df*;
   (4) independence with 40 *df*.

5. The null hypothesis for a goodness-of-fit test is that the observed frequencies
   (1) fit the expected frequencies;
   (2) do not fit the expected frequencies;
   (3) either (1) or (2), depending of the size of the χ^2 value;
   (4) either (1) or (2), depending on the size of the χ^2 and the *df*.

6. In a χ^2 test of independence between sex and kinds of phobias, the null hypothesis was rejected. The proper conclusion is that
   (1) sex and phobias are independent of each other;
   (2) sex and phobias are related to each other;
   (3) knowing a person's phobia gives you no clue to his or her sex;
   (4) none of the above.

7. A χ^2 test of goodness of fit was used to evaluate a model. The null hypothesis was rejected. The proper conclusion is that the model is
   (1) adequate;
   (2) inadequate;
   (3) either (1) or (2), depending on the *df*;
   (4) models must be evaluated with a test of independence.

8. To use the χ^2 distribution with confidence you must assume that the observations you make
   (1) are normally distributed;
   (2) have equal variances;
   (3) are independent;
   (4) all of the above.

9. To find expected frequencies in a test of independence,
   (1) begin by assuming that the categories are independent;
   (2) use predictions based on a theory;
   (3) begin by assuming that the categories of events are related in some way;
   (4) none of the above.

10. For a 2×2 χ^2 test of independence, _____ is the effect size index, and _____ is a medium value.
   (1) θ (theta); .50;
   (2) θ (theta); .30;
   (3) φ (phi); .50;
   (4) φ (phi); .30.

11. Suppose you were analyzing data from a 2×2 test of independence and you had one expected frequency that was very small. According to your text you are very likely to make a
   (1) Type I error;
   (2) Type II error;
   (3) both (1) and (2);
   (4) neither (1) nor (2).

12. If you compare chi square curves for skewness, you'll find that as degrees of freedom increase the curves are
   (1) positively skewed but becoming less so;
   (2) positively skewed and becoming more so;
   (3) negatively skewed and becoming less so;
   (4) negatively skewed and becoming more so.

13. After the data are gathered, a statistician may decide to combine logically related categories. The most likely reason for this is to
   (1) hide mistakes;
   (2) increase the size of some expected values;
   (3) increase the degrees of freedom;
   (4) decrease the skew of the chi square distribution.

14. Being able to examine each group's contribution to the final value of a chi square value was described as being due to the _____ nature of chi square.
    (1) distributive;
    (2) additive;
    (3) combinatorial;
    (4) summative.

15. The result of Cambridge-Somerville Youth Study described in your text was that
    (1) youth from low socioeconomic status homes were twice as likely to have a police record as youth from higher socioeconomic status homes;
    (2) youth from low socioeconomic status homes were no more likely to have a police record than youth from higher socioeconomic status hones;
    (3) youth with counselors and enrichment opportunities were less likely to have a police record than youth without counselors and opportunities;
    (4) youth with counselors and enrichment opportunities were just as likely to have a police record as youth without counselors and opportunities.

16. A friend brings you data that can be used to establish the independence of two variables. You run an analysis on the data and find a chi square value smaller than the tabled value. Your analysis supports the idea that the two variables
    (1) are independent;
    (2) are not independent;
    (3) either (1) or (2) depending whether the calculated value is significantly smaller than the tabled value.

17. The goodness-of-fit test is useful in helping us determine
    (1) whether or not the data fit our hypothesis;
    (2) whether or not our data fit what is expected by chance;
    (3) both (1) and (2);
    (4) neither (1) nor (2).

18. Assume you are interested in determining if there is a preference in the dining hall for a particular night. You would expect, by chance, that there would be an equal number of students eating in the dining hall each night. You collect the data. Which χ^2 test is MOST appropriate?
    (1) χ^2 test for independence;
    (2) goodness of fit;
    (3) either, depending on degrees of freedom;
    (4) neither; you should use a *t* test.

19. Suppose your data produce an effect size estimate of 0.30. According to the text, this is a
    (1) large effect size;
    (2) medium effect size;
    (3) small effect size.

20. For the χ^2 test described in question 18, how many degrees of freedom are there?
    (1) 8;
    (2) 7;
    (3) 6;
    (4) 12.

Interpretation _____

1. In the study of human origins, anthropologists divide themselves into two camps: the "lumpers" (who argue that the pieces of evidence are not all that different) and the "splitters" (who argue that the differences represent separate species). A budding anthropologist wondered if being a lumper or splitter was related to experience as an anthropologist. She gathered the following data, which produced a χ^2 value of 16.67. Calculate φ and write a conclusion about anthropologists.

	Lumpers	Splitters	Σ
Less than 5 yrs. Experience	10	30	40
More than 5 yrs. Experience	40	20	60
Σ	50	50	

2. Let's say you collect data on skipping class. You examine male students and female students who have low, medium and high GPA's. You obtain a χ^2 value of 9.96. Write a conclusion about skipping class.

	Low	Medium	High	Σ
Males	35	21	14	70
Females	21	25	31	77
Σ	56	46	45	

Problems

1. In an activity-wheel experiment, animals were either isolated or housed together. Both groups were allowed food for one hour each day. After 25 days, the animals were either alive with stabilized weights or they were dead. Test for the independence of these two conditions with a χ^2 test. Calculate φ. Write a conclusion. For the reasoning behind this experiment, see Spatz and Jones (1971).

	Living Conditions	
	Isolated	Together
Alive	10	8
Dead	2	12

2. With tomatoes, the color of the fruit and the height of the plant are genetically determined (with red dominant over yellow, and tall dominant over short). Suppose for a few minutes that you are a geneticist during the 1920s or 1930s when these facts were unknown. If the above facts about dominance are true, a particular set of crossings will result in a 9:3:3:1 ratio, with 9 tall reds to 3 short reds to 3 tall yellows to 1 short yellow. Suppose you carried out the crossings and found 90 tall reds, 39 short reds, 39 tall yellows, and 18 short yellows. Use a χ^2 test to determine if such data fit a 9:3:3:1 model. (Note: 9:3:3:1 is a shortcut way to write 9/16, 3/16, 3/16, 1/16.) Is this a test for independence or goodness of fit?

3. You may remember a textbook problem that used M&M candy. The answer to that problem suggested that a larger sample would make a more realistic problem. I took the textbook's advice and bought a 22.7 ounce bag of plain M&M's. There were 756 pieces of candy in the bag. The Mars company says that the color distribution in plain M&M's is brown—30%; yellow—20%; red—20%; blue—10%; green—10% orange—10% (www.mms.com/us/about/products/index.jsp). The frequency count in my bag was: brown—222; yellow—149; red—132; blue—102; green—44; orange—107. Identify this problem as a test of independence or of goodness-of-fit. Calculate χ^2 and write an interpretation.

4. A researcher asked 173 people with different educational accomplishments tc respond to the statement "Our planet is being observed by intelligent life forms from outer space." Responses were given on a 7-point scale from "Strong Agreement" to "Strong Disagreement." The number of people responding in each educational category follows. Analyze the data, and comment on the relationship between level of education and agreement with the statement.

Level of Agreement	College Educated	High School Dropouts
Strong Agreement	2	4
Mild Agreement	7	10
Slight Agreement	15	24
Neutral	19	21
Slight Disagreement	23	18
Mild Disagreement	12	9
Strong Disagreement	6	3

5. One year just before the end of the 20th century in America, 1,985,596 baby boys and 1,895,298 baby girls were born. Use a chi square test to determine if these data fit the theory that girls and boys are born in equal numbers.