DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Research Methods Posc 302

TESTING STATISTICAL HYPOTHESIS
THE CHI SQUARE TEST

I. TODAY'S SESSION:
   A. The basic ideas of hypothesis testing
   B. The chi square test of statistical independence
   C. Writing tips:
      1. **R**epublican, not **r**epublican; **D**emocrat, not **d**emocrat when referring to American parties.
      2. Pay attention to the meaning of words..
         i. "There is a large **difference** in percentages between whites and blacks with respect to turnout." Not, "There is a large **void** in percentages between whites and blacks with respect to voting."
            1) Find help in ***Woe Is I***, especially Chapter 5.

II. SAMPLING DISTRIBUTION:
   A. These comments repeat some of the material presented in Class 15 notes.
   B. The basic ideas and uses of a sampling distribution
      1. A sampling distribution shows the relationship or connection between the possible outcomes of N independent random trials and the probability associated with each outcome.
      2. It's based on sound mathematical theory.
         i. In this course we don't need to worry much about the underlying mathematics.
      3. The use of a sampling distribution
         i. We can use a sampling distribution to make a decision about a null hypothesis based on an observed result because it (the sampling distribution) tells us how likely a result is, if the null hypothesis is true.
         ii. It connects results with probabilities and hence helps us make informed or educated inferences.
         iii. We can in effect observe a result and then conclude or infer with the help of a sampling distribution that this outcome is or is not likely under the null hypothesis.
   C. Example:
      1. Consider one final time the initial problem, judging whether or not a person is fair in a coin tossing contest.
      2. We translated that question into a question about a null hypothesis.
         i. We wondered if the H_0 holds water.

ii. If this hypothesis turns out to be false, we'll entertain an alternative, H_A
iii. Both are stated ahead of time.

3. Example:
   i. H0: P = .5 (i.e., the coin is "fair" or "honest"
   ii. HA: P ≠ .5 (i.e., the coin is not fair but we don't know ahead of time how it might be unfair.
4. We then designed an experiment to test it.
   i. Conduct N independent random trials and observe some result.
      1) Example: 10 tosses of the person's coin.
5. Determine an appropriate sampling distribution.
   i. Example: for N = 10 trials each of which eventuates or ends in just one of two possible outcomes the sampling distribution is, assuming that the probability of a particular outcome is ½ is:

Number of heads in 10 flips	Probability
10.00	0.0010
9.00	0.0098
8.00	0.0439
7.00	0.1172
6.00	0.2051
5.00	0.2461
4.00	0.2051
3.00	0.1172
2.00	0.0439
1.00	0.0098
0.00	0.0010
	1.0000

Box 1: Sampling Distribution P = .5, N = 10

   ii. Box 1 shows possible outcomes–the number of heads in 10 tosses–and the probabilities of those outcomes.

D. Before conducting the experiment and using the sampling distribution to interpret the results, we need to establish a decision rule.

III. DECISION RULE:

A. Again these notes are repeated from Class 15.
B. Implicit in our plan is the necessity of making a decision rule.

1. We have to decide, **ahead of time**, what outcomes are **in our minds** probable and what outcomes are **again in our minds** improbable.
2. One way to do this is to classify outcomes as falling in to a **region of acceptance** or a **region of rejection**.
    i. The terms acceptance and rejection pertain to the null hypothesis.
3. If an outcome falls in the region of acceptance, we say that its likelihood of occurrence is relatively high given the null hypothesis and hence we will accept that hypothesis.
4. If, on the other hand, a result is considered to be unlikely if the null hypothesis is true, then we will reject that hypothesis and tentatively accept the alternative.
5. This approach gives us a way to answer the questions: "If in a series of 10 flips of an honest coin a person gets 9 heads, should we distrust the individual?"

C. What is probable? What is improbable?
1. At some point we are going to have to say that a result is or is not probable under the null hypothesis.
2. The judgment or evaluation of probability is subjective. What I think is a highly unlikely eventuality you may regard as reasonably probable.
3. That is, we both may agree that something is possible but I maintain that its chances of occurring are very remote whereas you may think that probability is not too low.
4. Knowing the exact probability really won't help us decide. Why?
    i. Consider: if an outcome occurs with probability .1 or one time out of ten, is it improbable or probable?
    ii. I might say "That's a pretty unlikely result."
    iii. By contrast, you might argue "Sure the probability is only 1 out of 10, but that's not an exceptionally low probability."
    iv. Or take the previous sampling distribution and a hypothetical result, say, the person we are dealing with gets 7 heads on 10 tosses.
        1) Is he or she cheating–is the coin biased?
        2) Well, the chances of getting 7 heads on 10 flips of an honest coin is .1172. (See Box 1 above.)
        3) Is that a low probability? Is it so low that you would think something is fishy with the coin or the game or the person?
        4) Or is it a low probability but not so low that you want to shout "Foul!" and reach for a revolver?

D. The long and short is that
1. We have to divide the sampling distribution into an areas of acceptance and rejection;
2. But the choice of these areas is somewhat arbitrary and subjective.

E. Critical region of a sampling distribution.

1. **Definition of critical region**: a set of outcomes considered so unlikely under the null hypothesis that should any one of them occur we will reject that null hypothesis.
2. Let's say that any result that occurs with probability of .1 or lower will lead us to reject the null hypothesis.
3. With this guideline in mind we divide the sampling distribution as follows in the next box. (Box 2).
4. It shows that we will reject H_0 that P = .5 if we obtain any of the following results:
    i. 8 or more heads in 10 tosses, because too many heads for a fair coin.
    ii. 2 or fewer heads in 10 tosses, because not enough heads for a fair coin.

Number of heads in 10 flips	Probability	
10.00	0.0010	
9.00	0.0098	Reject H_0
8.00	0.0439	
7.00	0.1172	
6.00	0.2051	
5.00	0.2461	Accept H_0
4.00	0.2051	
3.00	0.1172	
2.00	0.0439	
1.00	0.0098	Reject H_0
0.00	0.0010	
	1.0000	

Box 2: My Critical Region in Sampling Distribution

5. This may not seem like a fair decision rule because we reject the null hypothesis if events that could occur but are moderately unlikely do in fact occur.
    i. So we can modify the decision rule to say that if we observe an outcome that occurs with probability less than .01 or 1 time out of 100, we'll reject H_0; otherwise, we'll accept it.
    ii. The critical region then becomes (see Box 35):
    iii. 9 or more heads in 10 tosses, because too many heads for a fair

coin.

iv. 1 or fewer heads in 10 tosses, because not enough heads for a fair coin.

6. Which of these decision rules is best?

i. The answer depends to some degree on subjective factors.

Number of heads in 10 flips	Probability	
10.00	0.0010	
9.00	0.0098	Reject H_0
8.00	0.0439	
7.00	0.1172	
6.00	0.2051	
5.00	0.2461	Accept H_0
4.00	0.2051	
3.00	0.1172	
2.00	0.0439	
1.00	0.0098	Reject H_0
0.00	0.0010	
	1.0000	

Box 3: Smaller Critical Region

F. NOTE: CREATE THE DECISION RULE BEFORE SEEING DATA OR RUNNING THE EXPERIMENT.

1. It's not fair looking at the data and then creating a decision rule.
2. That procedure is acceptable in some circumstances but that type of statistics is not part of this course.

IV. SUMMARY:

A. Study the substantive problem and specify research hypotheses.

B. Translate these propositions to statistical hypotheses.

1. H_0: parameter of interest equals a particular number.
2. H_A: parameter of interest does not equal or is greater than or is less than the number specified in the null hypothesis.

C. Design an "experiment"

1. Determine number of trials.

D. Find an appropriate sampling distribution.

E. Make a decision rule.
   1. Choose areas of acceptance and rejection.
F. Collect data
   1. Determine observed value
G. Make a decision about the hypothesis, H_0.
H. Derive a substantive conclusion.

V. OVERVIEW OF THE CHI SQUARE TEST:
A. Suppose we are interested in the relationship between people's religiosity, as measured by how frequently they attend church, and their political ideology.
   1. Do liberals tend to be atheists, for instance?
B. We can turn to the General Social Survey and obtain a cross-classification between an indicator of church attendance (X) and an indicator of ideology (Y).
   1. An aside: in this instance labeling the variables independent and dependent is arbitrary.
C. This observed table shows a relationship but only for a sample.
   1. What we really want to know is whether or not there is an association in the population as whole.
   2. But, as indicated in previous classes this kind of population characteristic is unknown and has to be estimated.
D. One approach is to state a (statistical) null hypothesis, namely that there is no relationship between X and Y.
   1. For now the null hypothesis might be written as H_0: XY = 0, where "XY" means "the relationship between X and Y" and the whole phrase reads: "The hull hypothesis is that the relationship between X and Y is non-existent or nil."
      i. The alternative hypothesis is that there is a relationship in the population or H_A: XY $\neq$ 0.
   2. Naturally, to make use of these hypotheses and investigate them with the methods outlined in the previous sections we need to refine it.
E. After doing so we can then set up an "experimental" procedure, determine a sampling distribution, and make a decision rule.
   1. Then once the data have been collect–the table generated with the sample–we can test the hypothesis.
      i. That is, we can "accept" it, which means that whatever the sample shows we believe there is no relationship in the population.
      ii. Or we can reject H_0 (and accept H_A), which means that we think there is a relationship in the population.
F. Preview:
   1. We'll obtain an outcome, called an observed **chi square** statistic, and ask the question: if there is no relationship between the religiosity and ideology in the population, how likely is it that we would observe a chi square as

large or larger than the one we actually observed?

2. The computer program we've been using reports a chi square statistic along with probability.
   i. The interpretation of these numbers is: if there were no relationship between the variables in the population, the probability of observing a chi square of ... or larger would be ...
   ii. This question is analogous to asking if a coin is fair how likely is it we would observe a given number of heads in N tosses?

G. Example:

1. Null hypothesis: in the population of Americans there is **no** relationship between X (Church attendance) by Y (Ideology).
2. Here is a sample cross-tabulation:

Cells contain: -Column percent -N of cases		attend				
		1 Seldom	2 Moderate	3 Often	4 A lot	*ROW TOTAL*
polviews	1 Liberal	**18.0** 179	**12.3** 69	**12.4** 50	**8.4** 53	***13.6*** *351*
	2 Lean L	**14.8** 147	**14.6** 82	**11.7** 47	**8.4** 53	***12.7*** *329*
	3 Mod	**40.8** 406	**41.7** 234	**37.6** 151	**37.0** 233	***39.6*** *1,024*
	4 Lean C	**14.5** 144	**18.0** 101	**19.4** 78	**18.4** 116	***17.0*** *439*
	5 Cons	**12.0** 119	**13.4** 75	**18.9** 76	**27.7** 174	***17.2*** *444*
	COL TOTAL	***100.0*** *995*	***100.0*** *561*	***100.0*** *402*	***100.0*** *629*	***100.0*** *2,587*

Figure 1: Association Between Church Attendance and Ideology

3. The observed chi square for this association is 112.17.
4. Knowledge of the sampling distribution of the chi square lets us conclude:
   i. If there were no association between the two variables, the probability of observing a chi square this large or larger is less than .0001.
   ii. That's like getting 490 heads out of 500 tosses.
   iii. Hence, we should perhaps reject the null hypothesis and conclude that there is a relationship between church attendance and political ideology.

VI. STATISTICAL INDEPENDENCE:

A. So far in the semester we've just "verbalized" the meaning of "relationship" and occasionally used figures or examples to get explain the idea.

1. Now let's state the notion of relationship, or more precisely, lack of relationship more formally.

B. Marginal probabilities

1. Consider two categorical variables such as X = number of times a person attends church and Y = self-placement on a liberalism conservatism.
   i. Let's denote that probability that a randomly selected person has a particular value on X as P(X = a).
      1) Read this statement as "the probability that a person's score on X is a." The letter "a" is just a symbol for some arbitrary value of X.
   ii. Similarly, let P(Y = g) be the probability that a randomly drawn person has some value on Y.
      1) Again, the letter "g" is just a symbol that means "some value."
   iii. In the present example, we might refer to the probability that a person says "never" when asked how often he or she attends church during the month as P(X = "never").
      1) Usually, we use the term P(X) to mean "the probability that X equals some value.)
      2) And a similar phrase, P(Y) can be used for the probability that Y equals some value.
   iv. **These are called marginal probabilities.**

C. Joint probabilities

1. A randomly selected and interviewed individual can be classified on both variables.
   i. Example: a person has value X = "never" **and** Y = "very liberal."
   ii. Another person might be jointly classified X = "twice a week" and Y = "moderate."
2. We are interested in the probability that a person has a particular value on X and a particular value on Y.
   i. Let's call this probability P(X = aY = g), which can be read "the probability that X = a **and** Y = g), where of course a and g are simply letters that stand for particular values of the variables.
   ii. It's usually easiest and sufficient to just write **P(XY)** to mean the joint probability that X equals a value and Y (simultaneously) equals some value.

D. Statistical independence:

1. Variables X and Y are **statistically independent** if and only if P(XY) = P(X)P(Y).
   i. In words, X and Y are independent if and only if the joint probability that X = a and Y = g equals the product of the marginal

probabilities.

E. This is pretty abstract, but fortunately there are zillions of ways to provide an intuitive understanding.

1. Example 1: you toss a coin and simultaneously and independently roll one die with six faces.
   i. The probability of getting “heads” is ½. (Assume the coin is fair.)
   ii. The probability of getting “.”(one dot) is 1/6.
   iii. The probability of getting “head” and “.” is ½ X 1/6 = 1/12.
   iv. Explanation:
      1) The marginal probability of “X = head” is ½.
      2) The marginal probability of “Y = one” is 1/6.
      3) Since the coin toss and roll are done independently, the joint probability that “X = head” and “Y = one” is the product of the marginal probabilities or ½ X 1/6 = 1/12.
2. Example 2: you draw a card at random from a 52-card deck and independently toss an honest coin.
   i. What is the probability of getting the joint occurrence “king of hearts” **and** “tail.”
   ii. Since there is (presumably) only one king of hearts in the deck, the marginal probability of drawing it is 1/52.
   iii. And as we’ve seen countless times before the probability of tossing a tail in one flip of a fair coin is ½.
   iv. Consequently, the probability P(X = king of hearts, Y = tail) equals P(kind of heart) times P(tail) = 1/52 X ½ = 1/104.
3. Example 3: suppose you interview people from a well studied population of Americans.
   i. In this population there are 55 percent females so the marginal probability of drawing a female is P(gender = female) = P(F) = .55.
   ii. Moreover, the proportion of Republicans in the population is .40 and consequently the probability of interviewing a Republican is P(party = Republican) = P(R) = .40.
   iii. Consequently, the joint probability that a randomly selected individual is a female Republican is P(F) X P(R) = .55 X .40 = .22.

F. The chi square statistic lets us test the hypothesis that the values of one variable are independent of the values of another variable.

1. If this situation turns out to be true, we’ll say that there is not relationship between them in the population.

VII. TEST YOURSELF:

A. It’s important that you actively try to understand these ideas. So here are a couple of questions.

1. Consider the people in the United States. The probability of being over 65

years of age is .35. (I'm just making this up.) The probability of being a strong liberal is .1. (Again, I'm making this up.) Moreover, for now let's assume that there is no connection whatever between ideology and age.

   i. What is the probability that someone you chance to meet on the street will be a senior citizen **and** call herself a strong liberal?
   ii. Given what was said above, what is the probability that someone you chance upon will be younger than 65 **and** will be a strong liberal?

2. Do you suppose that X, eye (or hair) color is independent of Y, vote choice in the 1999 Philadelphia election for mayor? Could you express this lack of independence in the probability symbols we used in the preceding pages?

B. Interpret the relationship shown in Figure 1.

1. Which of these statements is true?
   i. 18 percent of the people who "Seldom" attend church are liberal.
   ii. 18 percent of the sample are liberal.
   iii. About 8 percent of the 629 people who say they attend church "A lot" are liberal.
   iv. 27.7 percent of the conservatives attend church "A lot."
2. Hint: two statements are true and two are false.
3. If you have trouble with these, make sure you e-mail me for an appointment or explanation.

VIII. NEXT TIME:

A. The chi square test of the hypothesis that one variable (X) is statistically independent of another variable (Y).

B. Reading:

1. Johnson and Joslyn, ***Research Methods***, discuss the chi square test on pages 343 to 346. I am going into the background in more detail.