Political Science 100a/200a

Fall 2001

Hypothesis testing, Part II[1]

- FINALLY we are at the point where we can start asking social science questions.
- Consider the data set **lifeexp** in Stata. Here are a few we might address with this data:
  1. Did former British colonies grow significantly faster (in per capita GDP) between 1960 and 1980 than former French colonies?
  2. Same as 1, but for sub-Saharan African ex-colonies only.
  3. Were former British colonies significantly better or worse off than former French colonies in terms of per capita GDP in 1960? In 1980? In subSaharan Africa only?
  4. Are countries with federal political arrangments 1980 (of which **fedind80** is a measure) significantly more ethnically diverse than are non-federal countries?
  5. Among former colonies in Africa and Asia, do more ethnically diverse states have significantly lower income growth rates for 1960-80?
- All of these questions ask us to test a hypotheses about a *difference* between sample means.
  - e.g.: Let $\bar{x}_{BR}$ be the mean annual growth rate in per capita income for former British colonies, and $\bar{x}_{FR}$ likewise for former French colonies.
  - Question 1 may be interpreted as asking us to see if we can reject the null hypothesis that $\bar{x}_{BR} - \bar{x}_{FR} = 0$.
- To do this, all we need to do is adapt our techniques to ask about the distribution of the *difference* between two sample means, which is not very difficult.
- HOWEVER, first a more conceptual point. We can and probably should distinguish between two sorts of enterprises here: Causal inference versus descriptive inference.
  1. Descriptive inference: Imagine that all we have is a 25 country sample of states, and this data (in **lifeexp**) for each of them.
     - In this case, question 1 can or may be interpreted like this: Can we reject the hypothesis that *in the population of 157 countries*, the difference between the average growth rate for ex-British and ex-French colonies is zero?

[1]Notes by James D. Fearon, Dept. of Political Science, Stanford University, November 2001.

- The problem in this case is to make a *descriptive inference* about a difference between the means of two populations (ex-British and ex-French colonies) from a smaller random sample of the two populations.
- No claim need be advanced here about colonial status *causing* any difference in growth rates.
- To draw an inference about the difference between the unobserved population means for ex-Brit and ex-French colonies, we will need to use the techniques discussed in the last few lectures to measure the uncertainty around our estimate of the difference between the sample means.

2. Causal inference: Imagine that in fact we have data on 157 countries (as we do). In the sense of the preceding, we have "the population."
   - In this case, we can answer the question about descriptive inference just by using the **summarize** or **table** commands in Stata: Illustrate with **gen britfr = 1 if colbrit == 1**, **replace britfr = 0 if colfra == 1**, **table britfr ,contents(mean grw6080 sd grw6080 n grw6080)**.
   - So on average former British colonies grew almost one percent faster per year than former French colonies. This is just describing the (population) data.
   - But suppose that we suspect that there may be some *causal effect* or effects related to colonial policies. We have a theory or suspicion that British and French colonial policies with respect to economy and government, decolonization, or post-colonial relations have differed in ways that have had consequences for the economic performance. (Ideally, you would go in with a theory already in hand that says something like, "If this theory is correct, we should expect ex-British colonies to have performed better economically than ex-French colonies, because ...")
   - In this case, we might be thinking of question 1 as follows: *Can we reject the null hypothesis that British/French colonial status had no systematic impact on post-colonial growth rates?*
   - Here we imagine that each country i's 1960-80 growth rate was produced like the sum of a "colonial effect" and other causes that were random with respect to the colonial effect. e.g.,

     $$grw_{i,BR} = ce_{BR} + \epsilon_{i,BR}$$

     where $grw_{i,BR}$ is annual growth, 1960-80 for a former British colony, ce_{BR} is the growth rate caused by the impact of British colonial policies, and $\epsilon_{i,BR}$ represents the impact of random (with respect to colonial policies) other causes for former British colony i.
   - So in this case of causal inference we are asking about the difference

     $$mean(grw_{i,BR}) - mean(grw_{j,FR}) = ce_{BR} - ce_{FR} + mean(\epsilon_{i,BR}) - mean(e_{j,FR})$$

and we want to draw an inference about the difference in the colonial effects. Can we reject the hypothesis that $ce_{BR} - ce_{FR} = 0$?

  - In effect, for this causal inference version we imagine that the population of 157 countries and growth rates that we see is really a particular *sample* from a hypothetical set of outcomes that "might have been." I.e., we imagine that countries growth rates for 1960-80 were like draws from a box of tickets, and we are testing to see if we can reject the null hypothesis that ex-British and ex-French colonies are drawn from two boxes that have the same mean.
  - (Note: FPP might be unhappy with this, because the "box model" is too conjectural or hypothetical – see p. 558.)

- Ok, back to answering question 1. To test the null hypothesis, whether in the case of descriptive or causal inference, we need to estimate the probable error around the difference $\bar{x}_{BR} - \bar{x}_{FR}$. To the extent that we can imagine that these means are produced from samples that are like independent draws from two different boxes (a British box and a French box), we have

$$\begin{aligned} var(\bar{x}_{BR} - \bar{x}_{FR}) &= var(\bar{x}_{BR}) + var(\bar{x}_{FR}) \\ &= \frac{\sigma^2_{BR}}{n_{BR}} + \frac{\sigma^2_{FR}}{n_{FR}} \end{aligned}$$

  which we can estimate with

$$var(\bar{x}_{BR} - \bar{x}_{FR}) = \frac{s^2_{BR}}{n_{BR}} + \frac{s^2_{FR}}{n_{FR}}$$

- Plugging in the relevant numbers and computing the sd gives:

$$sd(\bar{x}_{BR} - \bar{x}_{FR}) = \sqrt{\frac{(2.23)^2}{40} + \frac{(1.91)^2}{22}} = .5386$$

- To do the t test we ask what is the probability of getting a difference this large if the truth were that the difference were zero:

$$t = \frac{\text{observed diff.} - \text{expected diff}|H_0}{\text{se of diff.}} = \frac{(2.62 - 1.71) - 0}{.5386} = 1.69.$$

- What is the probability of getting a number this large with a t distribution with $40 + 22 - 2$ degrees of freedom?

- How do I know that this is the number of "degrees of freedom"? In general, *degrees of freedom equals the number of data points you have minus the number of parameters you are estimating.* Here there are 40 + 22 data points and two parameters (the two means). So in general, when estimating a difference of means where the sample sizes are n and m, the df's are $m + n - 2$.

- **di ttail(60, 1.69)** = .048. So what this is referring to with a picture. .... If we have a two-tailed test, then p = .096, which is barely significant at the .10 level. (Note that this was a two tailed test. Means what? What if, by contrast, we had "gone in" with a theory that said that we should expect British ex-colonies to grow faster than ex-French?)
- So we can't really reject the hypothesis that there may be no systematic effect of British versus French colonial status on post-colonial growth rates.
- Check our results with **ttest grw6080 ,by(britfr) unequal** ...
- IMPORTANT: Even if we HAD found a statistically significant difference, this would not mean that we had proven that colonial policies CAUSED the average difference in growth rates, since there might be omitted variables that just happen to be correlated with colonial status but which are in fact driving growth rates. In terms of our "model,"

  $$mean(grw_{i,BR}) - mean(grw_{j,FR}) = ce_{BR} - ce_{FR} + mean(\epsilon_{i,BR}) - mean(e_{j,FR})$$

  the causal inference is only justified if we can plausibly argue that the difference between the average effect of the "other causes" (the ϵ's) for British and French colonies is zero. In words, it has to be that there is no other factor that significantly affected growth rates that happened to be correlated with colonial status.
- So at best, the result would give us a license to look more closely.

Question: In Africa, Asia, and the Middle East, have ethnically more diverse states been significantly more likely to experience violent civil conflict in the period since 1960?

- I have incorporated some new variables into the data set **lifeexp** that allow us to address this question.
  - **ethfrac** is a commonly used measure of ethnic diversity based on a 1960 Soviet ethnographic Atlas. It ranges from 0 to 1 and represents the probability that two randomly selected individuals from the country will be members of different ethnolinguistic groups (according the Soviet geographers). Thus **ethfrac** = 0 implies total homogeneity and **ethfrac** $\rightarrow$ 1 implies extreme heterogeneity. (Using your probability theory, you should be able to figure out how you would construct this measure if you had data for each country giving the proportion of the population belonging to each group.)
  - I created a dichotomous version of **ethfrac**, called **eth2val**, which is 1 for **ethfrac** $>$ .5 and 0 otherwise.
  - **cwar** is a dichotomous variable that is '1' for countries that had a violent civil conflict involving an organized military opposition that killed more than 1000 people

over its course, according to data from the International Institute for Strategic Studies and Ruth Leger Sivard, ed., *World Military Expenditures*. This is a pretty low threshold (**tab cwar**).

- The indicator variable **colony** is '1' for all countries in subSaharan Africa, North Africa/Middle East, and Asia (excluding Japan). I want to look at the relationship between ethnic diversity and civil war only within this set in order to control a bit for the possible confounding influence of per capita income – can you see what the potential problem would be?

- Let's **tabulate eth2val cwar if colony == 1,row** and show the row percentages. Note that the more diverse countries are more likely to have had a civil war by this definition, but the effect is not enormous. Could it be due to chance rather than any causal influence of ethnic diversity on the propensity of a country to have significant civil violence?

- Let the null hypothesis H_0 = civil wars occur randomly with respect to ethnic diversity, and our alternative H_1 = more ethnically diverse countries are more likely to have civil wars.

- We can formulate this as a test of the null hypothesis that the difference between two sample means is zero:
  - Let $\bar{x}_{ed=1}$ be the proportion of highly ethnically diverse (**eth2val = 1**) countries that had civil wars.
  - Let $\bar{x}_{ed=0}$ be the proportion of low ethnic diversity countries that had civil wars.
  - We can now ask if we can reject the hypothesis that $\bar{x}_{ed=1} - \bar{x}_{ed=0} = 0$, using a t test.
  - To do so we need to figure the standard error of the difference, which is:

$$\sqrt{\frac{s^2_{ed=1}}{n_{ed=1}} + \frac{s^2_{ed=0}}{n_{ed=0}}} = \sqrt{\frac{.62(1-.62)}{47} + \frac{.49(1-.49)}{35}} = .11$$

  - So our test statistic is

$$t = \frac{\text{observed diff.} - \text{expected diff}|H_0}{\text{se of diff.}} = \frac{(.62 - .49) - 0}{.11} = 1.18.$$

  - Degrees of freedom are $n_{ed=1} + n_{ed=0} - 2 = 47 + 35 - 2 = 80$. From Stata the probability of a draw from t distribution with 80 df's that is more than 1.18 away from zero *in the positive direction* is **di ttail(80, 1.18)** = .12.
  - Since our alternative hypothesis H_1 is directional (i.e. we expected greater ethnic diversity to imply a greater risk of civil war), this is our p value for the test – that is, it is the probability that we would see a difference this large or larger if the null hypothesis of no systematic difference were true.

- $p = .12$ is not statistically significant by any standard convention.
- Questions about any step?

The χ^2 test

- There is another way to approach the last problem, using something called a χ^2 *test*, that is important to be familiar with because (1) it is very commonly used in the analysis of cross-tabulations (e.g., joint democracy and war, etc.), and (2) the χ^2 probability distribution appears in many social science statistical applications (e.g., the sample variance s^2 has a χ^2 distribution when the population is roughly normal).
- First, let me introduce the χ^2 probability distribution.
- Intuitively, the χ^2 is the distribution you get if you (1) draw a bunch of numbers from a standard normal distribution, (2) square them, and (3) add them together.
- eg: in Stata,
  - **clear, set obs 5000**
  - **for num 1/10: gen xX = invnorm(uniform())**
  - **gen z = 0**
  - **gen z = z1^2 + z2^2 + ... + z10^**
  - **graph z ,bin(25)**
- This is a χ^2 distribution with m degrees of freedom, where m is the number of terms in the sum of the normals.
- OK, NOW let's reconsider the cross-tab of **eth2val** and **cwar**. **tab eth2val cwar**.
- We are interested in the likelihood that numbers like these would appear if in fact there were NO systematic relationship between ethnic diversity and the probability of civil war in these parts of the war.
- Another way to go at this problem is to think of these two variables, **eth2val** and **cwar** as having a *joint probability distribution* $f(x, y)$, and ask if we can reject the hypothesis that the two variables are distributed independently, i.e., that $f(x, y) = f(x)f(y)$.
- The first step is to figure what numbers we would expect in the cells if in fact the two variables were independent. work through this ...
- A natural measure of how different the actual distribution is from what we would expect if the two variables were independent will thus be some function of how far each observed value in each cell is from the expected value under the null hypothesis.

- Karl Pearson established the following results:

  1. For each cell in the cross-tab, square the difference between the observed value and the expected value under the null hypothesis of independence, then divide this number by the expected value. Formally, for each cell, calculate

$$\frac{(\text{observed \# - expected \#})^2}{\text{expected \#}}$$

  2. Then add these numbers up for all the cells. The result is the χ^2 *test statistic.*
  3. Pearson showed that this statistic has a particular probability distribution, called a χ^2 *distribution with k degrees of freedom.* In the case of a cross-tab, the degrees of freedom $k = (\#\text{ rows} - 1) * (\#\text{ columns} - 1)$. (The intuition is that the degrees of freedom are the number of terms in the sum that makes up the χ^2 statistic that are independently determined. With a two-by-two cross tab, if you know the marginal distribution, the total number n, and any two cells, you can exactly determine the other two.)
  4. Draw rough graphs of the prob. distribution.
  5. Technically, this is the distribution you get when you take one standard normal distribution for each degree of freedom, square each of them, and add them together. (Intuitively, with a large sample, the central limit theorem implies that the variation of the observed value around the expected value in each cell under the null hypothesis should follow a normal distribution. Squaring the difference gets you a squared normal distribution, which is then added to those for the other cells.)

- The bigger the divergence of the observed numbers from the expected numbers in each cell, the bigger the χ^2 test statistic. The χ^2 distribution tells you "What is the probability you would get a test statistic this large if the truth were that the two variables are independent?" Draw graph ...

- Work it out for this example ...

- Stata is much quicker: **tab eth2val cwar, row chi**.

- Compare results to difference of means test: Same answer, though χ^2 is by nature a "two tailed" test. (see discussion in FPP, p. A-30, note 3 on the equivalence of different tests in this situation)

- Notice that you can use the χ^2 statistic to test for independence in cases where a difference of means test would be hard to apply.

  – e.g.: In the literature on ethnic politics, one sometimes encounters theoretical arguments that imply the relation between ethnic diversity and the probability of violent conflict should NOT be monotonically increasing (i.e., increasing steadily

with greater diversity). Instead, some arguments imply that violent conflict should be most likely in societies that are neither highly homogenous nor highly heterogenous. e.g., when there is one big majority group and one relatively large minority (such as in Rwanda, Burundi, Sri Lanka, Israel/Palestine, etc.)

- We can create a 3-valued indicator variable that divides countries into Low, Medium, and High levels of ethnic fractionalization as follows:
  - **xtile eth3val = ethfrac ,nq(3)** (see help xtile). **tab eth3val**
  - **tab eth3val cwar, row chi**. Interpret.
- You can also use the χ^2 test when you have a null hypothesis that predicts a probability distribution across $n > 2$ events.
  - eg: The last US Senate had 15 Democratic-Democratic delegations, 19 Republican-Republican delegation, and 16 mixed (one Dem, one Rep) delegations.
  - What distribution would you expect to observe under a (naive) null hypothesis that suppose that voters vote without any reference to party ID, and thus elect Democrats and Republicans to each Senate seat with equal probabilities?
  - Then the probability of getting two Dem's is what? 2 Reps? A mixed delegation?
  - Try to figure out how you could use a χ^2 test to evaluate the null hypothesis. (And be sure to read the relevant FPP chapter, first section, if you are confused by any of this.)