Lecture 9: Comparing Groups

PUAD 601 with Nathan Favero

April 1, 2020

AMERICAN UNIVERSITY
WASHINGTON, DC

Review: Hypothesis testing

Q: My research hypothesis is that men are taller than women. What should my null and alternative hypotheses be?

A:

Null: $H_0 : \mu_{height,men} = \mu_{height,women}$

Alternative: $H_A : \mu_{height,men} > \mu_{height,women}$ (one-tailed)

Review: Hypothesis testing

```
. ttest price=6000

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% C
---------+--------------------------------------------------------------------
   price |      74    6165.257    342.8719    2949.496    5481.9
------------------------------------------------------------------------------
    mean = mean(price)                            t =   0.4820
Ho: mean = 6000                 degrees of freedom =       73

 Ha: mean < 6000      Ha: mean != 6000       Ha: mean > 6000
Pr(T < t) = 0.684   Pr(|T| > |t|) = 0.631   Pr(T > t) = 0.315
```

Q: What is the null hypothesis? A: $H_0 : \mu_{price} = 6000$

Q: What can I conclude? A: Nothing

Review: Hypothesis testing

Q: Identify whether the following alternative hypotheses imply a one-tailed or two-tailed hypothesis test:

1. $Corr(x, y) \neq 0.2$
2. $\mu_x < 52$
3. $Corr(x, y) > 0$

A:

1. Two-tailed
2. One-tailed
3. One-tailed

Chi-squared tests

- When we have two categorical variables, talking about the two variables being related or unrelated is a bit more complicated than when at least one of our variables is continuous
- Example:
  - Let's look at some more survey data from the American National Election Survey in 2016
  - First, let's look at the breakdown of who people voted for in the 2012 presidential election (among those who voted)

Chi-squared tests

- Example:

```
. tab vote
```

Vote Choice	Freq.	Percent
1. Barack Obama	1,728	56.58
2. Mitt Romney	1,268	41.52
5. Other SPECIFY	58	1.90
Total	3,054	100.00

Chi-squared tests

- Example:
  - In the total sample (restricted to those who voted), 56.58% of the respondents voted for Obama
  - Q: Is race associated with vote choice?
  - If race is unrelated to vote choice, then the proportion of voters who support Obama should be the same for all racial groups (in the population)
  - If race is related to vote choice, then different racial groups will have a different percentage who supported Obama (in the population)
  - We want to know, based on our sample, can we conclude that the proportion of people who support Obama (vs. Romney vs. third party) is different depending on the racial group?

Chi-squared tests

- Example:
  - In Stata, we run:

    ```
    tab race vote, row
    ```

  - We can see that the percentage supporting Obama is not the same across all racial groups
  - For example, 48% of whites voted for Obama versus 98% of blacks
  - But remember, these are just differences in our sample
  - A chi-squared test will let us determine whether these sample differences are big enough to conclude that there must also be differences among racial groups (in terms of vote choice) in the population (or after taking into account random noise)

Chi-squared tests

- Example:
  - Null hypothesis: the two categorical variables are unrelated to one another in the population; in other words, the breakdown percentages among the three vote choices are equal for all racial groups
  - Alternative hypothesis: at least one racial group has a different vote breakdown (by percentage) than the other groups
  - In Stata, we run:

    ```
    tab race vote, row chi2
    ```

  - Now, we get a p-value: p=0.000
  - So we reject the null hypothesis and conclude that race is related to vote choice

Chi-squared tests

- Any time there's a hypothesis test, the most important thing to figure out is what the null hypothesis is
- If $p < \alpha$, we reject the null hypothesis (for any hypothesis test)
- For this chi-squared test, the null hypothesis is that the variables are unrelated (as the null hypothesis usually is)
- How do you know if two variables related?
- If knowing the value of one variable will help you predict the other variable (even a little bit), the two variables are related/associated
- In our example, knowing the race of a citizen helps us predict the likelihood that they will support Obama, so we'd say race and vote choice are related