Lecture 9: Comparing Groups

What we will do today

- Equivalence between confidence intervals and hypothesis testing
- Chapter 7 (comparing two groups)
  - Comparing means for two groups (e.g., do men get paid more than women)
    - Significance testing
    - Confidence intervals
    - Stata
  - Comparing proportions for two groups (e.g., are women more likely to believe in same-sex marriage than men)
    - Significance testing
    - Confidence intervals
    - Stata

Equivalence between confidence interval and two-sided significance test

Equivalence between confidence interval and two-sided significance test

- 95% confidence interval tells us much of the same information as a **two-sided** significance test with a .05 α (alpha level)
- 99% confidence interval tells us much of the same information as a **two-sided** significance test with a .01 α (alpha level)
- A $(1-\alpha)$% confidence interval tells us much of the same information as a two-sided significance test with alpha level= α

Relationship between confidence interval and two-sided significance test

- Show picture
- If p-value<=.05 (i.e., reject H_0)
  - If p-value<=.05 in a two-sided test, a 95% CI for μ does not contain μ_0
  - Equivalently, if 95% CI for μ does not contains μ_0 then we reject H_0
- If p-value>.05 (i.e., do not reject H_0)
  - When p-value>.05 in a two-sided test, the 95% CI for μ contains μ_0 (associated with null hypothesis, H_0)
  - Equivalently, if 95% CI for μ contains μ_0 then we do not reject H_0

CI and significance test

- Application of equivalence between confidence interval and significance testing
- Credit score example
  - Imagine that 95% CI for population mean credit score is 600 to 700
  - Imagine a two-sided hypothesis, alpha level =.05:
    - Would we reject H_0: $\mu = \mu_0 = 610$?
    - Would we reject H_0: $\mu = \mu_0 = 720$?

CIs vs. significance tests

- Confidence intervals better than significance tests
  - "Most statisticians believe [significance tests] have been overemphasized in social science research….A test merely indicates whether the particular value in H_0 is plausible. It does not tell us which other potential values are plausible. The confidence interval, by contrast, displays all plausible potential values. It shows the extent to which H_0 may be false by showing whether the values in the interval are far from the H_0 value." (Agresti, p. 164)

Chapter 7: Comparing Groups

What we have done; where we are going

- Where we have been: Chapters 5 and 6
  - What is the value of a population mean (quantitative)
    - Significance testing; confidence intervals
  - What is the value of a population proportion (0/1 variable)
    - Significance testing; confidence intervals
- Where we are going: Chap 7 (comparing two groups)
  - Comparing means for two groups (e.g., do men get paid more than women)
    - Significance testing; Confidence intervals; Stata
  - Comparing proportions for two groups (e.g., are women more likely to believe in same-sex marriage than men)
    - Significance testing; Confidence intervals; Stata

Comparing groups: independent and dependent variable

- We want to know whether the population mean for one group is bigger than the population mean for another group
- Research question:
  - Is the mean number of hours worked by men different than mean number of hours worked by women?
    - What is the dependent variable ("outcome variable")?
    - What is the independent variable (variable that affects the outcome variable?
- Today we learn methods to compare values of dependent variable, when independent variable only has *two* groups
  - E.g., we can't answer whether mean number of hours worked differs across, White, Black, Hispanic, Asian because more than two categories

Independent and dependent variable

- What is independent and dependent variable? Is dependent variable quantitative or 0/1 categorical?
  - Do people who take an SAT prep course have higher SAT scores than people who don't?
  - Are women more likely to vote for Obama than men?
  - Do out-of-state students have higher GPA than in-state students?
  - Does using Old Spice deodorant increase swagger?
  - Does belief in abortion rights differ between democrats and republicans?

Independent vs. Dependent Samples

- Methods for comparing the means or proportions of two groups differ depending on whether you have an "independent" or "dependent" sample
  - an entirely different concept than independent or dependent variable; today we will focus on "independent" samples
- Independent samples (focus of today's class)
  - Observations in one sample are independent of those in other sample; no matching between one sample and the other sample
  - Examples: (A) are Americans more likely to be married than French?; (B) do private colleges charge higher tuition than public
- Dependent samples (next class)
  - Natural matching occurs between each subject in one sample and a subject in another sample, often because each sample has same subjects
  - Examples: (A) does cancer treatment work (before and after); (B) does a college have higher tuition in 2010 than 2000;

Comparing two means

(1) Significance tests
(2) Confidence intervals

Comparing two means (hypothesis test)

- Same steps as before
  - (1) Assumptions
    - Dependent variable is quantitative; random sampling; sufficient sample size
  - (2) Hypotheses
  - (3) Test statistic
  - (4) P-value
  - (5) conclusion

Comparing two means (hypothesis test)

- Specific research question:
  - is the population mean number of hours worked for women, μ_1, different than population mean number of hours worked for men, μ_2?
- Null Hypothesis
  - H_0: $\mu_2 = \mu_1$
- Alternative hypothesis (two-sided)
  - H_a: $\mu_2 \neq \mu_1$
- Show pictures

Comparing two means (hypothesis test)

- What is our general strategy for comparing two means?
  - Instead of thinking of two separate parameters, $\mu_1\ and\ \mu_2$, we think of $(\mu_2 - \mu_1)$ as a single parameter.
  - If the parameter $(\mu_2 - \mu_1) \neq 0$, then we know that the parameter μ_2 is not equal to the parameter μ_1
- Null Hypothesis
  - H_0: $\mu_2 = \mu_1$
  - Same as this: H_0: $\mu_2 - \mu_1 = 0$
- Alternative
  - H_a: $\mu_2 \neq \mu_1$
  - Same as this: H_a: $\mu_2 - \mu_1 \neq 0$

Comparing two means (hypothesis test)

- The parameter μ_1 has a sampling distribution
  - Each observation in the sampling distribution is some sample mean $\bar{y}_1$
- The parameter μ_2 has a sampling distribution
  - Each observation in the sampling distribution is some sample mean $\bar{y}_2$
- The parameter $(\mu_2 - \mu_1)$ has a sampling distribution
  - What is each observation in this sampling distribution?
- Show picture

Hypothesis testing

- (2) Hypotheses
  - H_0: $\mu_2 - \mu_1 = 0$
  - H_a: $\mu_2 - \mu_1 \neq 0$
- (3) Test statistic
  - We test the hypotheses under the assumption that the null hypothesis is true
  - What does sampling distribution of $(\mu_2 - \mu_1)$ look like assuming H_0 is true?
  - Test:
    - How many standard errors is the sample mean we observed $(\bar{y}_2 - \bar{y}_1)$ away from the sample mean associated with the null hypothesis $(\mu_2 - \mu_1 = 0)$?
    - What does standard error of $(\bar{y}_2 - \bar{y}_1)$ represent?
- Show picture

Comparing two means (test statistic)

- Test statistic for value of pop mean (chap 6)
  - $t = \frac{\bar{y} - \mu_0}{se}$; (i.e., how many standard errors $\bar{y}$ is from μ_0)
- General test statistic formula (means, proportions)
  - $t = \frac{point\ estimate - null\ hypothesis\ value}{standard\ error}$
- Test statistic for comparing two means (chap 7)
  - $t = \frac{(\bar{y}_2 - \bar{y}_1) - 0}{se}; \quad where, se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
    - s_1= sample std dev for group 1; s_2= sample std dev for group 2
    - n_1=group 1 sample size; n_2= group 2 sample size
  - What does se represent?

Comparing two means (test statistic)

- Hours worked data
  - Women (group 1): $\bar{y}_1 = 37.60; n_1 = 1501; s_1 = 13.94$
  - Mean (group 2): $\bar{y}_2 = 43.76; n_2 = 1319; s_2 = 15.18$
- Test statistic for comparing two means (chap 7)
  - $\text{t} = \frac{(\bar{y}_2 - \bar{y}_1) - 0}{se}; \quad where, se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
    - s_1= sample std dev for group 1; s_2= sample std dev for group 2
    - n_1=group 1 sample size; n_2= group 2 sample size
- Calculations
  - $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}} = \sqrt{\frac{13.94^2}{1501} + \frac{15.18^2}{1319}} = .5515$
  - $\text{t} = \frac{(\bar{y}_2 - \bar{y}_1) - 0}{se} = \frac{(43.76 - 37.60) - 0}{.5515} = \frac{6.16}{.5515} = 11.2$

Comparing two means (p-value)

- P-value
  - t=11.2
  - Two-sided alternative hypothesis:
    - P-value=Probability of randomly choosing a sample mean and having t-score greater than 11.2 + probability of randomly choosing a sample mean and having t-score less than -11.2
    - P-value=pr(t>11.2)+pr(t<-11.2)
    - Use z-score table, probability is essentially 0
    - Show picture
- Conclusion
  - Reject H_0 and accept H_a

Note on standard error

- Se for mean, one group (s=sample std. dev)
  - $se = \frac{s}{\sqrt{n}}$; therefore, $se^2 = \left(\frac{s}{\sqrt{n}}\right)^2$
- For two groups
  - $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\left(\frac{s_1}{\sqrt{n_1}}\right)^2 + \left(\frac{s_2}{\sqrt{n_2}}\right)^2} = \sqrt{(se_1)^2 + (se_2)^2}$
    - s_1= sample std dev for group 1; s_2= sample std dev for group 2
    - n_1=group 1 sample size; n_2= group 2 sample size

Confidence interval for comparing two means

- Confidence interval for μ
  - e.g., we are 95% confident that the population mean number of hours worked lies somewhere between 39.93 and 41.03
- Confidence interval for $(\mu_2 - \mu_1)$
  - e.g., We are 95% confident that the weekly number of hours worked is between 5.09 and 7.24 higher for men than women
- Show pictures

CI for comparing two means

- Confidence interval for μ
  - $\bar{y} \pm z * se;$
  - where $se = s/\sqrt{n};$
- Confidence interval for $(\mu_2 - \mu_1)$
  - $(\bar{y}_2 - \bar{y}_1) \pm z * se$
  - $where, se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
    - s_1= sample std dev for group 1; s_2= sample std dev for group 2
    - n_1=group 1 sample size; n_2= group 2 sample size

CI for comparing two means

- Hours worked data
  - Women (group 1): $\bar{y}_1 = 37.60; n_1 = 1501; s_1 = 13.94$
  - Mean (group 2): $\bar{y}_2 = 43.76; n_2 = 1319; s_2 = 15.18$
- Confidence interval for $(\mu_2 - \mu_1)$
  - $(\bar{y}_2 - \bar{y}_1) \pm z * se;\ se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = .55$ (last example)
- 95% CI (z=1.96)
  - $43.76 - 37.60 \pm 1.96 * .55 = 6.16 \pm 1.078$
  - Lower=6.16-1.078=5.082; upper=6.16+1.078=7.238
  - e.g., We are 95% confident that the weekly number of hours worked is between 5.082 and 7.238 higher for men than women

Note on CI and confidence intervals

- If your confidence interval does not include zero
  - Then you would reject H_0: $\mu_2 - \mu_1 = 0$
  - If confidence level=(1 – alpha-level)
    - E.g., 95% and .05 alpha level
- If your confidence interval includes zero
  - Then you would not reject H_0: $\mu_2 - \mu_1 = 0$
  - If confidence level=(1 – alpha-level)
    - E.g., 95% and .05 alpha level

In-class exercises

- Question
  - We sampled SAT score of 400 men ($\bar{y}_1 = 1000; s_1 = 100$) and we sampled SAT scores of 484 women($\bar{y}_2 = 1030; s_2 = 120$)
  - (1) test the alternative hypothesis that mean SAT score is higher for women than men
  - (2) Calculate a 99% CI for difference in SAT scores between men and women
- You can assume that the assumptions were satisfied

Answer (significance test)

- (2) hypotheses (men=group 1; women=group2)
  - H_0: $\mu_2 - \mu_1 = 0$; H_a: $\mu_2 - \mu_1 > 0$
- (3) test statistic
  - $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{100^2}{400} + \frac{120^2}{484}} = \sqrt{25 + 29.75} = 7.40$
  - $\mathrm{t} = \frac{(\bar{y}_2 - \bar{y}_1) - 0}{se} = \frac{1030 - 1000}{7.4} = \frac{30}{7.4} = 4.05$
- (4) p-value (one-sided hypothesis)
  - p-value=pr(t>4.05)
  - Using z-score table: $pr(z > 4.05) = .0000317$
- (5) conclusion
  - P-value of .0000317 is less than alpha level of .05
  - Reject H_0; accept H_a: mean SAT score is higher for women than men

Answer (99% confidence interval)

- men=group 1; women=group 2
- Confidence interval for $(\mu_2 - \mu_1)$
  - $(\bar{y}_2 - \bar{y}_1) \pm z * se;\ se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- 99% CI
  - z=2.58 (from z-score table); se=7.4 (from significance test example)
  - $(\bar{y}_2 - \bar{y}_1) \pm z * se = 1030 - 1000 \pm 2.58 * 7.4$
  - $= 30 \pm 19.09 = 10.91\ to\ 40.09$
  - We are 99% confident that population mean SAT score is between 10.91 to 40.09 higher for women than men

Comparing two proportions

(1) Significance tests
(2) Confidence intervals

Comparing two proportions (significance tests)

- Significance tests for comparing two proportions
  - (1) assumptions
    - Outcome variable is a 0/1 variable; random assignment; sufficient sample size (at least 10 observations in each category for group 1, at least 10 obs in each category for group 2)
  - (2) hypotheses
  - (3) test statistic
  - (4) p-values
  - (5) conclusion

Comparing two proportions (significance tests)

- Research question:
  - Is the population proportion of men (π_1) who voted for Obama in 2008 equal the population proportion of women (π_2) who voted for Obama in 2008
  - Show sample proportions
- (2) Use two-sided alternative hypothesis, ($\alpha = .05$)
  - $H_0\colon \pi_1 = \pi_2 \; same \; as \; H_0\colon \pi_2 - \pi_1 = 0$
  - $H_a\colon \pi_1 \neq \pi_2 \; same \; as \; H_0\colon \pi_2 - \pi_1 \neq 0$
- (3) Testing strategy
  - Think of $(\pi_2 - \pi_1)$ as a single parameter; draw sampling distribution assuming $(\pi_2 - \pi_1) = 0$; calculate how likely it would be to observe $(\hat{\pi}_2 - \hat{\pi}_1)$ if the H_0 was true
  - show picture of sampling dist

(3) Test statistic

- $z = \frac{estimate - null\ hypothesis\ value}{null\ standard\ error} = \frac{(\hat{\pi}_2 - \hat{\pi}_1) - 0}{se_0}$
- se_0= sample standard error assuming $H_0$is true
  - If H_0 is true, then there is only one proportion (i.e., $\pi_1 = \pi_2 = \pi$)
  - $se_0 = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}} = \sqrt{\hat{\pi}(1 - \hat{\pi})(\frac{1}{n_1} + \frac{1}{n_2})}$
  - $\hat{\pi} = pooled\ proportion$
  - $\hat{\pi} = \frac{\text{total "successes"}\ in\ sample_1 and\ sample_2}{total\ sample\ size\ in\ sample_1 and\ sample_2}$

Test statistic, P-value, conclusion

- *(3) Test statistic*
  - $pooled\ proportion = \hat{\pi} = \frac{1972}{3378} = .5838$
  - $se_0 = \sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{n_1} + \frac{1}{n_2})} = \sqrt{.5838\ (1 - .5838)(\frac{1}{1{,}440} + \frac{1}{1{,}938})}$
  - $se_0 = \sqrt{(.2430) * (.00121)} = .01715$
  - $z = \frac{(\hat{\pi}_2 - \hat{\pi}_1) - 0}{se_0} = \frac{.6161 - .5403}{.01715} = \frac{.0758}{.01715} = 4.42$
- (4) P-value (two-sided H_a)
  - P-value=pr(z>4.42)+pr(z<-4.42)=.0000034+.0000034
- (5) conclusion
  - P-value is less than .05; reject H_0; accept H_a

Confidence intervals, comparing proportions

- Confidence interval for π (one sample)
  - $\hat{\pi} \pm z(se)$;
  - where $se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$
- Confidence interval for $(\pi_2 - \pi_1)$ (two samples)
  - $(\hat{\pi}_2 - \hat{\pi}_1) \pm z(se)$
  - $se = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} = \sqrt{(se_1)^2 + (se_2)^2}$
- Note we do not use se_0 because we are not assuming that $\pi_2 - \pi_1 = 0$
  - Where, $se_0 = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}}$

Example: CI comparing proportions

- Data
  - Group 1=men: $n_1 = 1{,}440;\ \hat{\pi}_1 = \frac{778}{1440} = .5403$
  - Group 2=women: $n_2 = 1{,}938;\ \hat{\pi}_2 = \frac{1{,}194}{1{,}938} = .6161$
  - $se = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} = \sqrt{\frac{.5403\ (1-.5403)}{1{,}440} + \frac{.6161\ (1-.6161)}{1{,}938}}$
  - $se = \sqrt{.000172 + .000122} = .017162$
- 95% Confidence interval for $(\pi_2 - \pi_1)$ (two samples)
  - $(\hat{\pi}_2 - \hat{\pi}_1) \pm z(se) = (.6161 - .5403) \pm 1.96 * .01716$
  - $= .0758 \pm .0336 = .0422\ to\ .1094$
  - We are 95% confident that proportion voting for Obama in 2008 is between .0422 to .1094 higher for women than men

(Fictitious) in-class exercise

- Research question
  - 93 out of 186 men (group 1) play video games; 90 out of 200 women (group 2) play video games
  - Is the population proportion of men who play video games, π_1, different than proportion of women who play video games, π_2?
  - Test using two-sided alternative hypothesis with alpha level=.05

Answer

- (2) Hypotheses
  - H_0: $\pi_2 - \pi_1 = 0$; H_a: $\pi_2 - \pi_1 \neq 0$
- (3) Test statistic
  - $\hat{\pi}_1 = \frac{93}{186} = .5; \hat{\pi}_2 = \frac{90}{200} = .45$
  - $\hat{\pi} = \frac{\text{total "successes" in } s_1 and\ s_2}{total\ sample\ size\ in\ s_1 and\ s_2} = \frac{93+90}{186+200} = .4741$
  - $se_0 = \sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{n_1} + \frac{1}{n_2})} = \sqrt{.4741\ (1 - .4741)(\frac{1}{186} + \frac{1}{200})} = .0509$
  - $z = \frac{(\hat{\pi}_2 - \hat{\pi}_1) - 0}{se_0} = \frac{(.45 - .50) - 0}{.0509} = \frac{-.05}{.0509} = -.98$
- (4) p-value
  - One-sided p-value with z=-.98 is .1635
  - Two-sided p-value= .1635*2=.327

– (5) Conclusion
  - P-value of .327 is not less than .05; do not reject H_0

Comparing (Independent) groups in Stata

(1) Means
(2) proportions

Comparing two Groups in Stata

- Means (quantitative variables)
  - Significance tests and CIs
- Proportions (0/1 categorical variables)
  - Significance tests and Cis
- Open the dataset in this link:
  - https://www.dropbox.com/s/nxz8zlpf7dj48w8/gss-2010-big-comparing-groups.dta
  - Note: this dataset is different than last week; it will not work if you use the dataset from last week

Comparing means

- Research question:
  - Does mean number of hours worked by women differ than mean number of hours worked by mn?
- Get to know your data
  - describe sex hrs1
  - tab sex
  - sum hrs1
  - sum hrs1 if sex==1
  - sum hrs1 if sex==2
- Side note on "operators"
  - help operators
  - sum hrs1 if age<=25
  - sum hrs1 if age>25 & age<=65

Comparing means

- Is hours worked different for women and men?
- ttest command syntax (same command as before)
  - Compare one variable for two different groups
    - ttest varname, by(group-varname) [options]
  - Important options
    - unequal: assumes that the two groups have unequal variances
    - level(#): what confidence level on confidence interval (e.g., level(99))
- Conduct significance test
  - ttest hrs1, by(sex) unequal
  - Show output in OneNote
- Conduct significance test (99% CI)
  - ttest hrs1, by(sex) level(99) unequal

Comparing proportions

- Research question
  - Is the proportion of men who voted for Obama in 2008 different than proportion of women who voted for Obama in 2008?
- Get to know your data
  - describe sex male obama08
  - tab sex
  - tab male
  - tab obama08
  - tab obama08 male
  - tab obama08 male, col

Comparing proportions

- Is the proportion of men who voted for Obama in 2008 different than proportion of women who voted for Obama in 2008?
- Compare one variable for two different groups
  - prtest varname, by(group-var-name) [options]
  - Important options:
    - level(#): specify confidence level for confidence interval
- Conduct prtest
  - prtest obama08, by(male)
    - Show output in OneNote
  - prtest obama08, by(male) level(99)
  - prtest obama08, by(sex)

Decisions and Decision Errors

Decisions and Errors

- Type 1 error
  - Probability of rejecting H_0 when H_0 is true
  - Example:
    - Null hypothesis: Amanda Knox is innocent
    - Truth: Amanda Knox is innocent
    - Type 1 error: jury finds Amanda Knox guilty, when in fact she is innocent
- Type 2 error
  - Probability of not rejecting H_0 when H_0 is false
  - Example
    - Null hypothesis: Amanda Knox is innocent
    - Truth: Amanda Knox is guilty
    - Type 1 error: jury finds Amanda Knox innocent, when in fact she is guilty

Decisions and Errors

- Type 1 error
  - Probability of rejecting H_0 when H_0 is true
- Type 2 error
  - Probability of not rejecting H_0 when H_0 is false

Truth (usually unknown)	Decision	
	Reject $\boldsymbol{H_0}$	**Do not reject $\boldsymbol{H_0}$**
H_0 is true	Type I error	Correct decision
H_0 is false	Correct Decision	Type II error

Type 1 error (optional slide)

- Type 1 error
  - Probability of rejecting H_0 when H_0 is true
- Type 2 error
  - Probability of not rejecting H_0 when H_0 is false

- Prior to conducting test, decide your tolerance for Type 1 error
  - probability of Type 1 error is alpha-level (i.e., rejection region) for test
  - Example: H_0: proportion of public institutions=.4
    - Alpha=.05, willing to accept 5% chance that we reject H_0 when H_0 is true.

Statistical vs. Practical Significance

- t-test
  - $t = \frac{\bar{y} - \mu_0}{se}; se = \frac{sample\ std.dav}{\sqrt{n}}$
  - $\uparrow n \rightarrow \downarrow se \rightarrow \uparrow t$
- Example in Stata
  - H_0:Proportion public=.29; (a) population (b) sample
- When you have a big enough sample, every relationship is significant
  - Example of research on English FE Colleges
- Funny business:
  - When sample sizes big, look for “strong” relationships

Statistical vs. Practical Significance

- The too small sample size problem
  - Cannot detect significant relationships even if those relationships are extremely strong in the population
- The too big sample size problem
  - Even the most trivial relationship is significant
  - Growing problem with more "administrative" data
- This is another reason to prefer confidence intervals over significance tests
  - For sample size too small: CI shows population relationship could be quite large
  - For sample size too big: CI shows that population relationship is very small.