Lecture 8: significance tests for means and proportions (Agresti chapter 6)

Administrative issues

- Midterm
  - Take home mid-term
  - (Most likely) distributed 10/31, due 11/7
  - 10/31 class will be devoted to midterm review

What we will do today

- Significance tests by hand
  - Significance tests for means (quantitative variables)
  - Significance tests for proportions (0/1 categorical variables)
  - Equivalence between significance tests and confidence intervals
- Significance tests in Stata
  - Significance tests for means
  - Significance tests for proportions

Significance tests for means (quantitative variables)

Significance tests

- Significance tests:
  - a significance test uses data to summarize the evidence about a hypothesis.
  - It does this by comparing point estimates (e.g., sample mean) to the parameter values (e.g., population mean) predicted by the hypothesis.
- There are 5 parts to a significance test
  - (1) assumptions
  - (2) hypotheses
  - (3) test statistic
  - (4) p-value (means probability value)
  - (5) conclusion

Strategy for Significance Testing

- Imagine that we are making a hypothesis test about the value of a population mean, μ
  - For example, mean SAT score
- Assuming that the null hypothesis, H_0, is true (i.e., $\mu = \mu_0$), how unlikely would it be to observe the sample mean, $\bar{y}$, that we observed?
- Question:
  - what would the sampling distribution look like if the null hypothesis is true?

Strategy for Significance Testing

- Draw a sampling distribution, assuming that H_0 is true (i.e., $\mu = \mu_0$)
  - We don't know if this is the right sampling distribution; this is the sampling distribution that would be true IF the null hypothesis is true
- Label the sample mean
  - If the sample mean, $\bar{y}$, we got in our actual data is very unlikely to occur with under sampling distribution associated with the null hypothesis, then the null hypothesis is probably wrong
- Essentially, we are going to calculate how many standard deviations our sample mean, $\bar{y}$, is from the population mean associated with the null hypothesis, μ_0.
  - If our sample mean, $\bar{y}$, is many standard deviations away from the population mean associated with null hypothesis, μ_0, then null hypothesis is probably wrong.

Example 1

- Research question: Is the population mean number of hours worked (for those who work) equal to 40?
  - (Test a two sided alternative hypothesis)
- summarize hrs1
  - Sample size=n=2,820
  - Sample mean=40.483
  - Sample std deviation=14.850

(1) Assumptions

- Assumptions
  - The variable is a “quantitative” variable
  - Our data is a random sample from the population
  - The population distribution of this variable has a normal distribution
    - Note: This assumption is robust to violations if sample size is sufficiently large (say n>=30)
  - Do you think the variable “number of hours worked” fulfills these criteria?

(2) Hypotheses

- Generate hypotheses from research question:
  - Research question: Is the population mean number of hours worked (for those who work) equal to 40?
- Null hypothesis (H_o)
  - H_o: $\mu = \mu_o = 40$
- Two sided alternative hypothesis (H_a)
  - Two-sided: H_a: $\mu \neq 40$
    - This means that μ could be greater than 40 (μ >40) or μ could be less than 40 (μ<40)
    - So the two-sided hypothesis encompasses both one-sided hypotheses

(3) Test Statistic

- We conduct a test to see whether we should reject the null hypothesis
- What is the sampling distribution under the assumption that H_0 is true?
- Draw picture

– Test-statistic:
  - If null hypothesis is true, how unlikely would it be to randomly draw a sample mean equal to the observed sample mean

(3) Test Statistic

- Test statistic is based on measuring the distance between μ_0 (associated with the null hypothesis) and $\bar{y}$ (the sample mean we actually observed)
- Test statistic: t-score
  - We conduct a test to see whether we should reject the null hypothesis
  - $t = \frac{\bar{y}-u_0}{se}$, where $se = sample\ std\ err = \frac{sample\ std\ dev}{\sqrt{n}}$
- Hours worked example
  - n=2,820; $\bar{y}$=40.483; $\mu_0 = 40$; s=sample std. dev=14.850
  - $se = \frac{sample\ std\ dev}{\sqrt{n}} = \frac{14.850}{\sqrt{2820}} = .2796$
- $t = \frac{\bar{y}-u_0}{se} = \frac{40.483-40}{.2796} = \frac{.483}{.2796} = 1.73$
  - Question: how many standard dev away from μ_0 is our sample?

(4) P-value

- P-value (probability value)
  - Under the presumption that H_0 is true, the p-value is the probability the test statistic equals the observed value or a value even more extreme in the direction predicted by H_a
  - Small p-value means that it would be unusual to find the observed data if H_0 were true.
- Two-sided hypothesis (H_a: $\mu \neq 40$)
  - In two-sided hypothesis "direction predicted by H_a" is both directions; this means that μ could be greater than 40 (μ >40) or μ could be less than 40 (μ<40)
- P-value for two sided hypothesis
  - Pr(obs>t) + pr(obs<-t)= Pr(obs>1.73)+pr(obs<-1.73)
- Draw picture; use z-score table to find probabilities

(4) P-value

- P-value=.0418+.0418=.0813
- Interpretation of p-value
  - Under the assumption that H_0 is true, the probability of observing a test statistic even more extreme than 1.73 (i.e., greater than 1.73 or less than -1.73) is equal to .0813

(4) Rejection Region

- Rejection region
  - $\alpha\ level$ (alpha level) is a number such that we reject H_0 if the observed p-value is less than or equal to the alpha level.
    - We should choose the alpha level prior to conducting analyses
  - We reject the null hypothesis if the observed p-value is less than or equal to the rejection region
  - In practice, most common alpha levels are .05 or .01
    - So if we choose $\alpha\ level$ of .05 and find a p-value of .02, we reject H_0

(4) P-value: rejection region

- Hours worked example
- Assume we choose a rejection region of .05
- We find a p-value of .0836
- Should we reject the null hypothesis?

(5) Conclusion

- H_0: $\mu = \mu_0 = 40$
- H_a: $\mu \neq \mu_0$
- Alpha level= rejection region=.05
- P-value=.0836
- Conclusion:
  - do not reject H_0.
  - We do not have sufficient evidence to reject the null hypothesis that population mean hours worked is equal to 40 hours per week.

Conclusion (continued)

- How to write your conclusion in terms of null and alternative hypotheses

	Conclusion	
P-value	H_0	H_a
P<=.05	Reject H_0	Accept H_a
P>.05	Do not reject H_0	Do not accept H_a

- Note that we never say "Accept H_0" or "Reject H_a"
  - Why? Even if we do not have enough evidence to reject H_0, we might have been able to reject H_0 if we had a bigger sample size

Example 2

- Research question: Is the population mean number of hours worked (for those who work) equal to 40?
  - This time test the one-sided alternative hypothesis that mean hours worked is greater than 40
  - Choose α level (i.e., rejection region) of .05
    - So reject if observed p-value <=.05
- summarize hrs1
  - Sample size=n=2,820
  - Sample mean=40.483
  - Sample std deviation=14.850

(1) Assumptions

- Assumptions
  - The variable is a “quantitative” variable
  - Our data is a random sample from the population
  - The population distribution of this variable has a normal distribution
    - Note: This assumption is robust to violations if sample size is sufficiently large (say n>=30)
  - Do you think the variable “number of hours worked” fulfills these criteria?

(2) Hypotheses

- Generate hypotheses from research question:
  - Research question: Is the population mean number of hours worked (for those who work) equal to 40?
- Null hypothesis (H_o) [same as before]
  - H_o: $\mu = \mu_o = 40$
- Two sided alternative hypothesis (H_a) [different]
  - Two-sided: H_a: $\mu > 40$

(3) Test Statistic

- Test statistic: t-score
  - $t = \frac{\bar{y}-u_0}{se}$, where $se = sample\ std\ err = \frac{sample\ std\ dev}{\sqrt{n}}$
- Question: Does calculation of test statistic change now that we are testing a one sided alternative hypothesis?
- Answer: No!
- Hours worked example
  - n=2,820; $\bar{y}$=40.483; $\mu_0 = 40$; s=sample std. dev=14.850
  - $se = \frac{sample\ std\ dev}{\sqrt{n}} = \frac{14.850}{\sqrt{2820}} = .2796$
- $t = \frac{\bar{y}-u_0}{se} = \frac{40.483-40}{.2796} = \frac{.483}{.2796} = 1.73$

(4) P-value (probability value)

- P-value
  - Small p-value means that it would be unusual to find the observed data if H_0 were true.
- Question: Does calculation of P-value change in a one-sided vs. a two-sided test?
- Answer: Yes!
- Draw picture

(4) two-sided vs. one-sided p-value

- P-value
  - Small p-value means that it would be unusual to find the observed data if H_0 were true.
- Two-sided. H_a: $\mu \neq \mu_0$ [Example 1]
  - Under the assumption that H_0 is true, the p-value is the probability of finding sample mean at least as far away from μ_0 as $\bar{y}$ (in either direction).
  - P-value=pr(t>1.73)+pr(t<-1.73)
- One sided. H_a: $\mu > \mu_0$ [Example 2]
  - Under the assumption that H_0 is true, the p-value is the probability of finding sample mean at least large $\bar{y}$
  - P-value=pr(t>1.73)

(5) Conclusion

- H_0: $\mu = \mu_0 = 40$
- H_a: $\mu > \mu_0$
- Alpha level= rejection region=.05
- P-value=.0418
- Conclusion: reject H_0; accept H_a
  - We reject the null hypothesis that population mean hours worked is 40
  - We accept the alternative hypothesis that population mean hours worked is greater than 40

One-sided or two-sided hypotheses?

- The data in example 1 and example 2 were exactly the same
  - Example 1 was a two-sided hypothesis
    - We did not reject H_0
  - Example 2 was a one-sided hypothesis
    - We rejected H_0
- You need stronger evidence (i.e., larger t-score) to reject H_0 in a two-sided hypothesis
- Generally, researchers prefer two-sided hypotheses because this is seen as a more conservative approach to hypothesis testing (i.e., only reject H_0 when you have strong evidence)

one-sided vs. two-sided rejection region

- Rejection region
  - $\alpha\ level$ (alpha level) is a number such that we reject if the observed p-value is less than or equal to the alpha level. This is something we decide before we calculate test statistic
  - Show picture of two-sided .05 alpha level
  - Show picture of one-sided .05 alpha level
- Two-sided hypotheses require a more extreme t-score (larger absolute value) in order to reject H_0 than one sided hypotheses. E.g., $\alpha\ level$ =.05:
  - One sided hypothesis: reject H_0 if t>1.645
  - Two sided hypothesis: reject H_0 if t>1.96 or if t<-1.96

In class exercise (answer on next page)

- A random sample of 400 students take the SAT; sample mean is 1030; sample std dev is 300
- Research question: is the population mean SAT score equal to 1,000
- (1) test the research question using a two sided alternative hypothesis
  - Use alpha level=.05; Show all five parts of the significance test; draw & label sampling distribution under assumption H_0 is true
- (2) test the research question using the one-sided alternative hypothesis that the population mean SAT score is greater than 1,000
  - Use alpha level=.05; Show all five parts of the significance test; draw & label sampling distribution under assumption H_0 is true

In class exercise (answers): Question 1

- (1) Assumptions
  - (a) quantitative variable; (b) random sampling (c) normal population distribution (robust to this assumption because large sample size)
- (2) hypothesis
  - H_0: $\mu = \mu_0 = 1{,}000$; H_a: $\mu \neq 1{,}000$
- (3) test statistic
  - $se = \frac{sample\ std\ dev}{\sqrt{n}} = \frac{300}{\sqrt{400}} = \frac{300}{20} = 15$
  - $t = \frac{\bar{y} - u_0}{se} = \frac{1030-1000}{15} = \frac{30}{15} = 2$
- (4) p-value
  - Two-sided p-value is Pr(t>2)+Pr(t<-2)
  - On z-score table, probability of finding z-score>2 is .0228
  - P-value for two-sided hypothesis = Pr(z>2)*2= .0228*2=.0456
- (5) conclusion
  - P-value of .0456 is less than alpha level of .05
  - Reject H_0; Accept H_a; we accept the alternative hypothesis that the population mean SAT score is not equal to 1,000

In class exercise (answers): Question 2

- (1) Assumptions
  - (a) quantitative variable; (b) random sampling (c) normal population distribution (robust to this assumption because large sample size)
- (2) hypothesis
  - H_0: $\mu = \mu_0 = 1{,}000$; H_a: $\mu > 1{,}000$
- (3) test statistic
  - $se = \frac{sample\ std\ dev}{\sqrt{n}} = \frac{300}{\sqrt{400}} = \frac{300}{20} = 15$
  - $t = \frac{\bar{y} - u_0}{se} = \frac{1030 - 1000}{15} = \frac{30}{15} = 2$
- (4) p-value
  - One-sided p-value is Pr(t>2)
  - On z-score table, probability of finding z-score>2 is .0228
  - P-value=.0228
- (5) conclusion
  - P-value of .0228 is less than alpha level of .05
  - Reject H_0; Accept H_a; We accept the alternative hypothesis that the population mean SAT score is greater than 1,000

Significance Tests for a Proportion

Significance Tests for a Proportion

- Same five parts as before:
  - (1) assumptions
  - (2) hypotheses
    - Note: decide on alpha level (i.e., rejection level) here
  - (3) test statistic
  - (4) p-value
  - (5) conclusion

Example 1

- Does the population proportion of people who believe that same sex couples should be allowed to marry equal .5?
  - Test using a two-sided alternative hypothesis
  - Use α level (i.e., rejection region) of .05
- Created a 0/1 variable called “msamesex” out of the ordinal variable “marhomo”
  - Show in Stata

Significance Tests for a Proportion

- Does the π of people who believe same sex couples should be allowed to marry equal .5?
  - Test with two-sided hypothesis
- (1) assumptions
  - Variable is a 0/1 variable (i.e., has two categories)
  - Random sampling
  - Sufficient sample size (at least 15 in each category)
- (2) hypotheses
  - H_0: $\pi = \pi_0 = .5$
  - H_a: $\pi \neq .5$

(3) Test statistic

- Does the π of people who believe same sex couples should be allowed to marry equal .5?
- (3) Test statistic
  - If H_0 is true (i.e., $\pi = .5$), how unlikely would it be to observe the sample proportion we observed, $\hat{\pi} = .4368$?
  - $z = \frac{\hat{\pi} - \pi_0}{se_0}$;
    - $\hat{\pi} = sample\ proportion$; π_0= proportion associated with null hypothesis; se_0= standard error associated with null hypothesis
- Draw sampling distribution associated with H_0

(3) Test statistic

- $z = \frac{\hat{\pi}-\pi_0}{se_0}$;
  - $n = sample\ size$
  - $\hat{\pi} = sample\ proportion = \frac{\#\ that\ say\ "yes"}{sample\ size}$
  - π_0= proportion associated with null hypothesis;
  - se_0= standard error associated with null hypothesis
    - $se_0 = \sqrt{\pi_0(1-\pi_0)/n}$
- Recommended steps
  - (1) calculate $\hat{\pi}$
  - (2) calculate se_0
  - (3) calculate z

(3) Test statistic

- Why do we use se_0 instead of se in z-test?
  - "null" standard error: $se_0 = \sqrt{\pi_0(1 - \pi_0)/n}$
  - estimated standard error: $se = \sqrt{\hat{\pi}(1 - \hat{\pi})/n}$
- We test H_0 under presumption that H_0 is true
  - therefore, use standard error under assumption that population proportion equals π_0
- Confidence intervals
  - Confidence intervals do not have a hypothesized value. Therefore, for CI use estimated se, not se_0

(3) Test statistic

- calculations
  - n=3,182
  - $\pi_0 = .5$
  - $\hat{\pi} = \frac{\#\ that\ say\ "yes"}{sample\ size} = \frac{1{,}390}{3{,}182} = .4368$

  – $se_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{.5(1-.5)}{3182}} = \sqrt{\frac{.25}{3182}} = .0088638$

- (3) Test statistic

  – $z = \frac{\hat{\pi}-\pi_0}{se_0} = \frac{.4368-.5}{.0088638} = \frac{-.0632}{.0088638} = -7.13$

(4) P-value

- (4) find p-value
  - Remember that for two tailed hypothesis we use “two sided” p-value
    - Pr(obs>z) + pr(obs<-z)= Pr(obs>z=7.13)+pr(obs<-z=-7.13)
  - P-value ≈.0000+.0000=.0000
  - Show picture
- (5) Conclusion (associated with two-sided H_a)
  - alpha level=.05
  - P-value=.0000
  - We reject the null hypothesis that the population proportion of people who agree with same sex marriage is equal to .5; we accept the alternative hypothesis that the population proportion of people who agree with same sex marriage is not equal to .5

Example 2

- Does the population proportion of people who believe that same sex couples should be allowed to marry equal .5?
  - Test using a one-sided alternative hypothesis that population proportion is less than .5
  - Use α level (i.e., rejection region) of .05
- Created a 0/1 variable called “msamesex” out of the ordinal variable “marhomo”
  - Show in Stata

Significance Tests for a Proportion

- Does the π of people who believe same sex couples should be allowed to marry equal .5?
  - One sided hypothesis
- (1) assumptions
  - Variable is a 0/1 variable (i.e., has two categories)
  - Random sampling
  - Sufficient sample size (at least 15 in each category)
- (2) hypotheses
  - H_0: $\pi = \pi_0 = .5$
  - H_a: $\pi < .5$

(3) Test statistic

- Test statistic (same calculations as before)
  - $z = \frac{\hat{\pi}-\pi_0}{se_0}$
- calculations
  - n=3,182; $\pi_0 = .5$
  - $\hat{\pi} = \frac{\#\ that\ say\ "yes"}{sample\ size} = \frac{1,390}{3,182} = .4368$
  - $se_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{.5(1-.5)}{3182}} = \sqrt{\frac{.25}{3182}} = .0088638$
- (3) Test statistic
  - $z = \frac{\hat{\pi}-\pi_0}{se_0} = \frac{.4368-.5}{.0088638} = \frac{-.0632}{.0088638} = -7.13$

(4) P-value

- (4) find p-value (different from before)
  - One-sided alternative hypothesis, H_a: $\pi < .5$
    - P-value=pr(obs<-z=-7.32)
  - P-value $\approx .0000$
  - Show picture
- (5) Conclusion (associated with two-sided H_a)
  - alpha level=.05
  - P-value=.0000
  - We reject the null hypothesis that the population proportion of people who agree with same sex marriage is equal to .5; we accept the alternative hypothesis that the population proportion of people who agree with same sex marriage is less than .5

In-class exercise (answer on next pg)

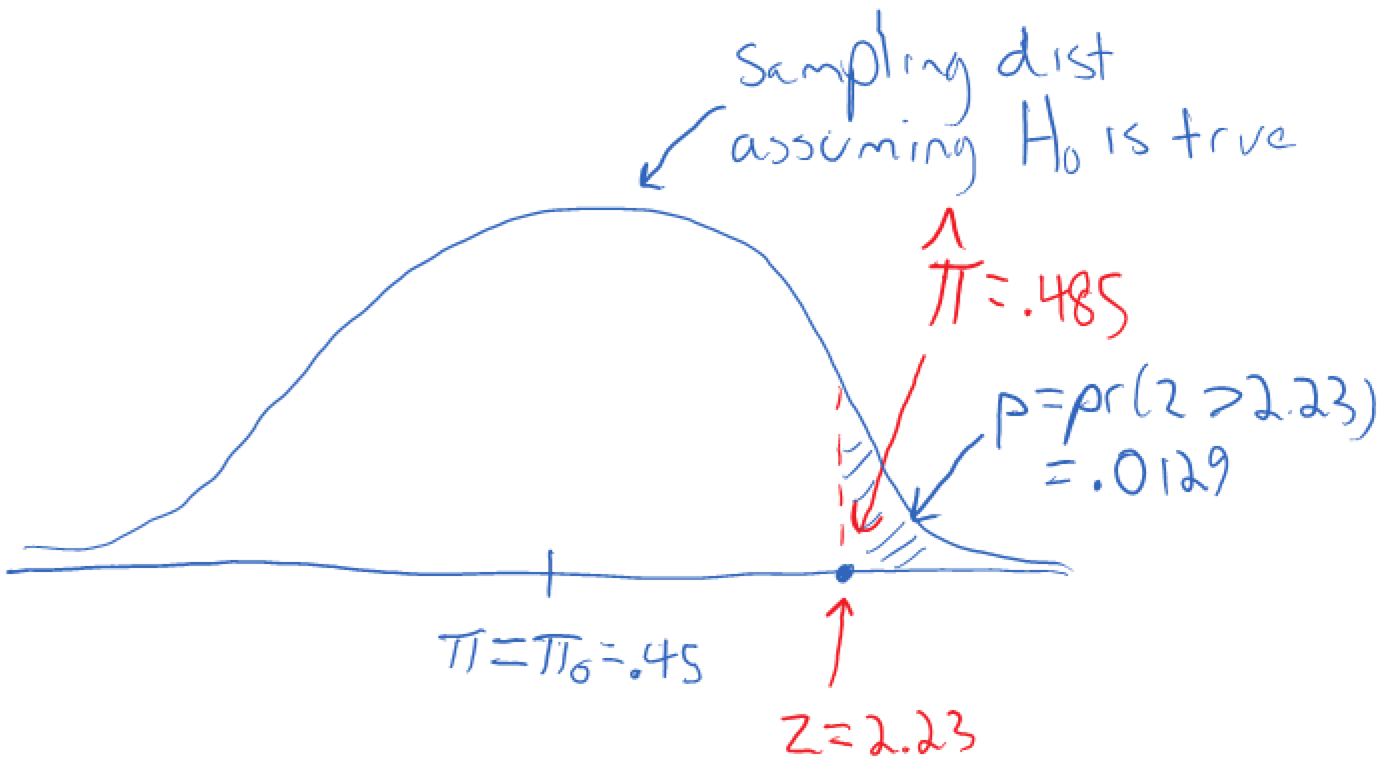
- Question: Nationally, 45% of community college students transfer to a 4-year institution within five years (I made this up). Researchers collected data on a random sample of 1,000 first-year Arizona community college student. Five years later, 485 had transferred to a 4-year institution. Does the population proportion of Arizona community college students who transfer equal the national population proportion of students who transfer?
  - Test using the one-sided hypothesis that the population proportion of Arizona is greater than .45
  - Use alpha level of .05; show all five steps (but can skip the "assumptions" step; draw and label sampling distribution associated with H_0

In-class exercises (answer)

- (2) Hypotheses
  - H_0: $\pi = \pi_0 = .45$; $\mathrm{H_a}$: $\pi > .45$
- (3) Test-statistic
  - $\hat{\pi} = \frac{485}{1000} = .485$; $se_0 = \sqrt{\frac{.45*(1-.45)}{1000}} = \sqrt{\frac{.2475}{1000}} = .01573$
  - $z = \frac{\hat{\pi}-\pi_0}{se_0} = \frac{.485-.45}{.01573} =\approx 2.23$
  - See sampling distribution on next page
- (4) P-value (use z-score table)
  - One sided hypothesis; p=pr(z>2.23)=.0129
- (5) Conclusion
  - P-value (=.0129) is less than alpha level (=.05)
  - Reject H_0; Accept H_a: $\pi > .45$

In-class exercises (answer)

- Sampling distribution (assuming H_0 is true)



Significance test for proportions

- Question:
  - Would we have rejected if we used a two-sided hypothesis and alpha-level=.05?
    - Yes! P-value=pr(z>2.23) + pr(z<-2.23)=.0129+.0129=.0258, which is less than .05
  - Would we have rejected if we used a one-sided hypothesis and an alpha-level of .01
    - No! P-value=.0129 is greater than alpha level of .01

Equivalence between confidence interval and two-sided significance test

Using Stata to conduct Significance tests

You conduct tests in Stata as I do

Using Stata to conduct tests

- Download Stata dataset
  - https://www.dropbox.com/s/whpdrryj707w1hw/gss-2010-big-sig-test.dta
- Open Stata dataset called 2010-big-sig-test.dta
- Copy this code into "command line"
  - describe hrs1 msamesex
  - sum hrs1
  - tab msamesex

Using Stata to conduct tests

- Test significance of a mean
  - ttest command:
- Test significance of proportions
  - prtest command
- One or two-tailed test
  - Stata reports both one- and two-tailed significance tests
  - In practice, researchers usually use two-tailed tests; we will focus on interpreting the two-tailed test
  - Converting a two-tailed test to a one-tailed test
    - Cut the two-tailed probability in half
    - If something has a .1 probability using a two-tailed test, its probability would be .05 using a one tailed test

Significance test for mean

- command (for one-sample test of mean)
  - ttest varname== μ_0
- Test null hypothesis that mean number of hours worked is equal to 40
  - H_0: $\mu = \mu_0 = 40$
  - Type this in Stata
    - ttest satp50==40
  - Why use two equal signs "=="?
    - In Stata, use two equal signs whenever we are *comparing* whether one value equals another value
- Copy test output into OneNote and discuss results
  - Stata results are very close to results calculated by hand

Significance test for proportion

- command (for one-sample test of proportion)
  - prtest varname== π_0
- Test null hypothesis that population proportion of people who agree w/ same sex marriage equals .5
  - $H_0\!: \pi = \pi_0 = .5$
  - Type this in Stata
    - prtest msamesex==.5
  - Why use two equal signs "=="?
    - In Stata, use two equal signs whenever we are *comparing* whether one value equals another value
- Copy test output into OneNote and discuss results
  - Stata results are very close to results calculated by hand

Significance of proportion

- command (for one-sample test of proportion)
  - prtest varname== π_0
  - Example:
    - prtest public==.4
    - H_0: proportion of public institutions is equal to .4
- help prtest
  - "prtest performs tests on the equality of proportions using large-sample statistics"
  - Note that prtest is calculated based on z-scores
- Stata examples
  - proportion of: (1) public institutions; (2) MA granting

Comments on testing with Software

- Note that Stata does not tell you whether to reject H_0 or not.
  - Stata merely gives you the p-values associated with different alternative hypotheses.
  - You decide whether to reject H_0 depending on what alpha levels (rejection regions) you decided on before hand.
- For testing proportions (prtest command), the variable must be coded 0/1; coding 1=no and 2=yes will not work

In class exercises (answers on next pg)

- For both questions:
  - Write all significance testing steps, except for assumptions; test a two-sided hypothesis; use alpha level of .05
- Significance test for mean (ttest command)
  - Test whether population mean number of children=2
  - (variable is called “childs”)
- Significance test for proportion (prtest command)
  - Test whether population proportion of people who think marijuana should be legal=.45
  - (variable is called “legalize”)

In class exercise answers (#1)

- Test whether population mean number of children=2
- Hypotheses
  - H_0: $\mu = \mu_0 = 2$; H_a: $\mu \neq 2$
- Calculate test statistic
  - Stata command:
    - ttest childs==2
  - $t \approx 2.64$
- P-value (for two-sided test) $\approx$.0083
- Conclusion
  - P-value of .0083 is less than alpha level of .05; reject null hypothesis

In class exercise answers (#2)

- Test whether population proportion of people who think marijuana should be legal=.45
- Hypotheses
  - H_0: $\pi = \pi_0 = .45; H_a: \pi \neq .45$
- Calculate test statistic
  - Stata command:
    - prtest legalize==.45
  - $z \approx 1.134$
- P-value (for two-sided test) $\approx$.2554
- Conclusion
  - P-value of .2554 is not less than alpha level of .05; do not reject null hypothesis

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 alpha level
- 99% confidence interval tells us much of the same information as a **two-sided** significance test with a .01 alpha level
- Etc.

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)

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.