Lecture 7: inference and hypothesis testing

What we will do today

- Chapter 5
  - Confidence intervals
    - For 0/1 categorical variables (e.g., did you vote for Romney, did you graduate from college)
- Chapter 6
  - Hypothesis testing for quantitative variables

Review: sampling distribution

- Imagine we want to estimate mean enrollment the population of colleges and universities
  - Draw the three pictures
- Questions
  - In each picture, what does each observation represent?
  - In each picture, what is the unit of analysis?
  - In each picture, what does the X-axis represent?
  - What does standard deviation represent?
  - What does standard error represent?

Review: Bias and efficiency

- Variable=total enrollments at college
- Bias
  - Show unbiased sample; show biased sample
  - Questions
    - How would picture of sampling distribution change if we exclude “less than 2yr for-profit” colleges?
- Efficiency
  - What is efficiency?
  - What would happen to the sampling distribution if we increased the sample size?

Real world application

- Monthly employment figures released on Friday
- What is the national unemployment rate in the U.S. population?
  - Data: Current Population Survey (CPS), a random sample of households
  - Calculate sample mean unemployment rate

$$= \frac{number\ of\ unemployed\ working\ age\ adults}{total\ \#\ of\ working\ age\ adults\ "in\ the\ labor\ marekt"}$$

  - Excludes people not looking for work. Question: Is that problematic?
- Net job creation in the last month
  - Data: Current Employment Statistics (CES) survey
  - Surveys 141,000 business and govt agencies out of an approximate population of 486,000 worksites

Confidence intervals for proportions

Means vs. Proportion (for this book)

- Mean
  - Refers to a quantitative variable (income, number of siblings, number of years married, etc.)
- Proportion
  - Refers to a categorical variable with two categories (will you vote for Obama?; did you graduate from college? Are you male? Are you white?)
  - These are often called “0/1 variables” where, for example, voting for Obama=1 and not voting for Obama=0; graduating from college=1 and not graduating from college=0.
- Statistical methods for means differ from statistical methods for proportions

Notation for proportions

- Population proportion (we usually don't know)
  - $\pi = population\ proportion$ ("pi")
- Sample proportion (we know)
  - $\hat{\pi} = sample\ proportion$ ("pi hat")
- Confidence interval
  - we use the sample proportion, $\hat{\pi}$, to make a confidence interval for the population proportion, π
  - e.g., we are 95% sure that the population proportion, π, of people who prefer Obama is between .49 and .53

Show Proportion in Stata

- IPEDS dataset of colleges and universities
- Variable called "public": is the institution private or public
  - 0= private; 1=public
- Show histogram of population distribution
- Show frequency distribution (tabulate command)
- Show mean (summarize command)
- Important fact:
  - If you code the variable 0="not public" and 1="public", then relative frequency of observations that are public (tabulate command) is equal to the mean of the variable public (summarize command)
  - This is why we use the numeric values 0/1 for "proportion" variables like "public"

CIs for proportions: Conceptual Understanding

- Variable called “Obama”: Do you plan to vote for Obama (0= No; 1= Yes)
- Imagine we asked 200 registered voters whether they planned to vote for Obama
  - 110 said “yes” out of 200
  - Sample proportion= $\hat{\pi} = \frac{number\ that\ said\ "yes"}{sample\ size} = \frac{110}{200} = .55$
- Goal
  - We want to create a 95% confidence interval (CI) for the value of the population proportion, π
  - 95% CI = $\hat{\pi} \pm margin\ of\ error$

CIs for proportions: Conceptual Understanding

- Variable called “Obama”: Do you plan to vote for Obama
  - 0= No; 1= Yes
- Show three pictures
  - Population distribution (unknown)
  - Sample distribution (known for one sample)
  - Sampling distribution (unknown)
- Question for students
  - Is the population distribution of the variable normally distributed?
  - Is the sample distribution of the variable normally distributed?
  - What does each observation of the sampling distribution represent?
  - Is the sampling distribution normally distributed (assume sample size is large)?
    - Why is this the case?

CIs for proportions: Conceptual Understanding

- Variable called "Obama": Do you plan to vote for Obama; 0= No; 1= Yes
- Ask about sampling distribution (assume no bias):
  - What percentage of observations will be within one standard deviation of the population proportion?
  - What percentage of sample proportions will be within one standard error of the population proportion?
  - What percentage of sample proportions will be within 1.96 standard errors of the population proportion?

CIs for proportions: Conceptual Understanding

- Problem:
  - We don’t know the population distribution or the sampling distribution
  - We have one sample and the sample proportion for that sample could be far away from population proportion
- Solution: We think of our sample as being randomly chosen from the sampling distribution
  - 95% of sample proportions (from the sampling distribution) will be within 1.96 standard deviations of the population proportion (show picture)
  - Equivalently, if we select a random sample and calculate the sample proportion, there is a 95% chance that the population proportion will be within 1.96 standard deviations of the sample proportion (show picture)

Calculating confidence intervals

Calculating CI for proportions

- 95% Confidence interval (CI)
  - 95% CI = $\hat{\pi}$ $\pm$ some margin of error
  - $\hat{\pi} \ \pm 1.96 * se$
  - Where $\hat{\pi}$ = sample proportion
    - $\hat{\pi} = \frac{number\ that\ said\ "yes"}{sample\ size} = \frac{110}{200} = .55$
  - se= sample standard error
- General Confidence interval (CI)
  - $\hat{\pi} \ \pm z * se$
  - Where z=z=score associated with desired confidence level
    - Question: where can we find the z-scores associated with each CI?

Calculating sample std. err. For proportions

- Confidence interval (CI) is $\hat{\pi} \pm z * \boldsymbol{se}$
- Population parameters
  - Standard deviation, σ, of the probability distribution
    - $\sigma = \sqrt{\pi(1-\pi)}$
  - Standard error of sample proportion, $\sigma_{\hat{\pi}}$
    - $\sigma_{\hat{\pi}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\pi(1-\pi)}{n}}$
    - but $\sigma_{\hat{\pi}}$ uses π, which is an unknown population parameter
- Sample Statistic
  - Sample standard error of the sample proportion, se
    - $se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$
  - In words: sample standard error, se, is our estimate for how much a random sample proportion differs from the true population proportion

Calculating CI for proportions

- Confidence interval (CI)
  - $\hat{\pi} \pm z * se, \textit{where:}$
  - $se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$
  - z=z-score of desired confidence level
    - Z=1.645 for 90% CI; Z=1.96 for 95% CI; Z=2.58 for 99% CI
- Recommended steps when calculating CI for proportions
  - First, calculate $\hat{\pi}$ = (# of "successes")/n
  - Second, calculate se
  - Third, calculate confidence interval

Calculating CI for proportions, Example

- 200 people sampled; 110 say they will vote for Obama; find 95% CI
  - *Sample size=n=200*
  - $\hat{\pi} = \frac{110}{200} = .55;$
  - $se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = \sqrt{\frac{.55(1-.55)}{200}} = \sqrt{\frac{.2475}{200}} = \sqrt{.0012375} = .03518$
- Confidence interval (CI)
  - $\hat{\pi} \ \pm z * se = .55 \pm 1.96 * .03518 = .55 \pm .069$
    - Lower bound of 95% CI=.55-.069=.481
    - Upper bound of 95% CI=.55+.069=.619
  - We are 95% sure that the pop proportion of people who will vote for Obama lies somewhere between .481 and .619

Confidence Interval Mechanics

- Do we think a confidence interval of .481 to .619 is good enough when trying to predict the proportion of people who will vote for Obama?
- What are two ways we get “more narrow” confidence intervals?

Calculating CI for proportions, Example

- 2,000 people sampled; 1,110 say they will vote for Obama; find 95% CI
  - *Sample size=n=2000;* $\hat{\pi} = \frac{1,100}{2,000} = .55;$
  - $se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = \sqrt{\frac{.55(1-.55)}{2000}} = \sqrt{\frac{.2475}{2000}} = \sqrt{.00012375} = .01112$
- Confidence interval (CI)
  - $\hat{\pi} \pm z * se = .55 \pm 1.96 * .01112 = .55 \pm .022$
    - Lower bound of 95% CI=.55-.022=.528
    - Upper bound of 95% CI=.55+.022=.572
  - We are 95% sure that the pop proportion of people who will vote for Obama lies somewhere between .528 and .572

Properties of Confidence Intervals

- Width of confidence interval decreases as sample size increases
- Width of confidence interval increases as desired confidence level increases
- Sample size considerations
  - Z-distribution is for “large” sample sizes
  - To use z-distribution to calculate CI of proportions, you sample should have at least 15 observations in each category
    - e.g., proportion vegetarian; sample must have at least 15 vegetarians and 15 non-vegetarians to use z-score table

In Class Exercise (answer on next pg)

- CI: $\hat{\pi} \pm z * se = \hat{\pi} \pm z * \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$
- Proportion on Facebook; sample=193; 90 people on facebook
  - What is 95% CI? What is 99% CI?
- Proportion on Facebook; sample=976; 423 people on Facebook
  - What is 95% CI? What is 99% CI?

In Class Exercise: Answers

- $\hat{\pi} \ \pm z * se;$
- sample=193; 90 people on facebook
  - $\hat{\pi}$=0.466321244; se= 0.035909049
  - 95% CI: 0.466321244 +- 0.070381736
  - 99% CI: 0.466321244 +- 0.092645346
- sample=976; 423 people on Facebook
  - $\hat{\pi}$= 0.433401639; se= 0.015862003
  - 95% CI: 0.433401639 +- 0.031089526
  - 99% CI: 0.433401639 +- 0.040923967

Assumptions and Assumption Violations

Assumptions and Assumption Violations

- Assumptions for confidence interval of a mean
  - (1) sample is a random sample from population
  - (2) population distribution of variable is normal
- “Robust”
  - A statistical method is robust with respect to a particular assumption, when it performs adequately even when that assumption is violated
  - Statisticians have shown that CI for a mean is robust against violations of normal population assumption, especially when sample size > 30

Assumptions and Assumption Violations

- Why is CI for mean robust to normal population assumption?
- **Central limit theorem**:
  - when sample size is large, the sampling distribution of the sample mean, $\bar{y}$, is approximately normal, even if the population distribution of the variable is not normal
- Why?
  - Because confidence interval about a statistic (e.g., the mean) is based on the shape of the sampling distribution, not the shape of population distribution of the variable; the shape of the population distribution will be normal as long as sample size is sufficiently large
  - If sample size is small (and population distribution is not normal), then sampling distribution is not normal, and we probably shouldn't trust the z-table to create confidence interval
- How large is large enough?
  - If population distribution is normal then sampling distribution is normal for any sample size
  - If population distribution is not normal, sample size of about 30 is sufficient

Assumptions and Assumption Violations

- Assumptions for confidence interval of a mean:
  - Assumption (1): The sample is a random sample from the population
- Do you think the CI for the mean is "robust" to violations of assumption (1)?
  - Why or why not?

Chapter 5: Significance tests

Hypotheses

- Science and social science often proceeds by testing hypotheses
- What is a hypothesis?
  - In statistics, a hypothesis is a declarative statement about a population.
  - It is usually a prediction that a parameter (e.g., population mean) takes a particular numerical value or falls in a certain range
  - Example hypotheses:
    - Average household income in the U.S. is $50,000
    - Men get paid more than women
- Remember:
  - We make hypotheses about the population, but we use sample data to test hypotheses
- What are some other hypotheses you can think of?

Hypotheses

- Example hypotheses:
  - Men get paid more than women
  - Holding job title constant, men get paid more than women
  - Public universities increase out of state enrollments when state appropriations decline
  - The mean score on a test will be 80%
  - Holding other factors constant, prestigious institutions will produce more master’s degrees than non prestigious institutions

Why are hypotheses useful?

- Force the researcher to make their predictions explicit
- Helps the researcher organize their analysis
  - I organize my literature review, conceptual framework, methods section, and results around the hypotheses I test
  - Example of current research project
- Helps the research community identify what relationships are significant
  - If prior research does not test hypotheses, hard to accumulate knowledge about a field

Null and alternative hypotheses

- Imagine we want to want to know whether the mean number of hours worked (by people who work) is 40
- Null hypothesis (H_o)
  - A statement that the population parameter has a specific value
  - (in words) H_o: the population mean number of hours worked is 40
  - (using symbols) H_o: $\mu = \mu_0 = 40$)
    - μ_0 is the parameter value associated with the null hypothesis (null pop mean)
- Alternative hypothesis (H_a)
  - A statement that the parameter falls in some alternative range of values
  - Two-sided alternative hypothesis
    - (in words) H_a: the population mean number of hours worked is not equal to 40
    - (using symbols) H_a: $\mu \neq 40$
  - One-sided alternative hypothesis (mean is greater than 40)
    - (in words) H_a: the population mean number of hours worked is greater than 40
    - (using symbols) H_a: $\mu > 40$
  - One-sided alternative hypothesis (mean is less than 40)

Null and alternative hypotheses

- Is the proportion of people who believe in global warming = .5?
- Null hypothesis (H_o)
  - (in words) H_o: the population proportion of people who believe in global warming equals .5
  - (using symbols) H_o: $\pi = \pi_0 = .5$
    - π_0 = parameter value associated with the null hypothesis (null pop proportion)
- Alternative hypothesis (H_a)
  - Two-sided alternative hypothesis
    - H_a: the population proportion of people who believe in global warming is not equal to .5
    - (using symbols) H_a: $\pi \neq .5$
  - One-sided alternative hypothesis (mean is greater than 40)
    - H_a:the population proportion of people who believe in global warming is greater than .5
    - (using symbols) H_a: $\pi > .5$
  - One-sided alternative hypothesis (mean is less than 40)
    - H_a:the population proportion of people who believe in global warming is less than .5
    - (using symbols) H_a: $\pi < .5$

In Class exercise (didn't have time to write answers)

- For each research question, write:
  - The null hypothesis
    - Write in words and write in symbols
  - All three alternative hypotheses
    - Write in words and write in symbols
- Question (1):
  - Is the population mean average SAT score equal to 1000?
- Question (2):
  - Is the population proportion of people who think that same-sex couples should have the right to marry equal to .5?

In Class exercise answers

- Note: I didn't have time to write answers
- Question (1): Is the population mean average SAT score equal to 1000?
  - Null hypothesis
  - Alternative hypotheses
    - Two-sided
    - One-sided (greater than)
    - One-sided (less than)
- Question (2): Is the population proportion of people who think that same-sex couples should have the right to marry equal to .5?
  - Null hypothesis
  - Alternative hypotheses
    - Two-sided
    - One-sided (greater than)
    - One-sided (less than)

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

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

Significance tests

- 5 parts of a significance test
  - (1) assumptions
  - (2) specify null and alternative hypotheses
  - (3) test statistic
  - (4) p-value
  - (5) conclusion

(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?
  - Remember that hypotheses are about population mean, μ
- Null hypothesis (H_o)
  - H_o: $\mu = \mu_o = 40$
- Two sided alternative hypothesis (H_a)
  - Two-sided: H_a: $\mu \neq 40$

(3) Test Statistic

- We conduct a test to see whether we should reject the null hypothesis
- Very important to remember:
    - We conduct our test under the assumption that the null hypothesis is true
- – Test-statistic:
    - If null hypothesis is true, how unlikely would it be to randomly draw a sample mean equal to the observed sample mean
- Draw picture

(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
  – $se = \frac{sample\ std\ dev}{\sqrt{n}} = \frac{14.850}{\sqrt{2820}} = .2796$
- $t = \frac{\bar{y}-u_0}{se} = \frac{40.483-40}{.2796} = \frac{.483}{.2796} = 1.73$

(4) P-value

- P-value
  - Under the assumption that $H_0$is true, the p-value is the probability that 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_0were true.
  - t= value of your t-test
- Two-sided hypothesis ($H_a: \mu \neq \mu_0$)
  - Pr(obs>t) + pr(obs<-t)
- Use z-score table to find probabilties
- Draw picture

(4) P-value

- P-value
  - Under the assumption that $H_0$is true, the p-value is the probability that the test statistic equals the observed value or a value even more extreme in the direction predicted by H_a
- 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 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"

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?
  - Remember that hypotheses are about population mean, μ
- Null hypothesis (H_o)
  - H_o: $\mu = \mu_o = 40$
- Two sided alternative hypothesis (H_a)
  - Two-sided: H_a: $\mu > 40$

(3) Test Statistic

- Question:
  - Does calculation of test statistic change now that we are testing a one sided alternative hypothesis?
- Test statistic: t-score
  - $t = \frac{\bar{y}-u_0}{se}$, 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
  - $se = \frac{sample\ std\ dev}{\sqrt{n}} = \frac{14.850}{\sqrt{2820}} = .2796$
- $t = \frac{\bar{y}-u_0}{se} = \frac{40.483-40}{.2796} = \frac{.483}{.2796} = 1.73$

(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_0were 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).
- 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 $\bar{y}$
- One sided. H_a: $\mu < \mu_0$
  - Under the assumption that H_0 is true, the p-value is the probability of finding sample mean as small or smaller than $\bar{y}$
- Draw picture

(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
  - Show picture of p-values
- 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
  - Assume alpha level=.05; Show all five parts of the significance test
- (2) test the research question using the one-sided alternative hypothesis that the population mean SAT score is greater than 1,000
  - Assume alpha level=.05; Show all five parts of the significance test

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

Equivalence between confidence interval and two-sided significance test

- Confidence interval for value of population parameter, μ
  - Goal of CI: some range of values between which we believe the population parameter, μ, lies.
  - Confidence interval: $\bar{y} \pm t * se$
  - 95% CI: $\bar{y} \pm 1.96 * se$
    - Assume sample size is greater than 100
- (1) Construct 95 % CI; (2) Construct a two-sided significance test with alpha=.05

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

- Example: Hours per week on internet
  - H_0: $\mu = \mu_0 = 10; H_a$: $\mu \neq 10$
  - Imagine that we reject H_0 using an alpha level of .05
    - Question: does 95% CI include the value 10?
  - Imagine we fail to reject H_0 using an alpha level of .05
    - Question: does 95% CI include the value 10?
- Example: Credit score
  - Imagine that 95% CI for population mean credit score is 600 to 700
  - Question (imagine 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)