Alpha, Statistical vs. Economic (Business) Significance, Rejection Region (Critical Value) Approach, and Comparing Proportions

Lecture 14

Reading: Sections 12.4 – 12.8

1

Lecture 13: Quick Review Example

- Research question: Do a majority of students favor splitting ECO220Y into two half courses?
  - What's implied hypothesis test (formal notation)?
  - If $\hat{P} = \frac{43}{100}$, do we have any support for H_1?
  - How about $\hat{P} = \frac{53}{100}$? $\hat{P} = \frac{63}{100}$?
  - How would P-values compare?

2

Overview of Hypothesis Testing

- Before collecting any data you choose:
  - Null hypothesis (H_0)
  - Research (alternative) hypothesis (H_1)
- After collecting the data you:
  - Compute the test statistics, P-value
  - Interpret your results and make conclusions
- Often compare with conventional significance levels: $\alpha = 0.01, \alpha = 0.05$, and $\alpha = 0.10$

Lower significance levels correspond to a *higher* burden of proof

3

Statistical Significance

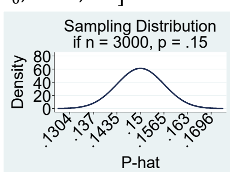
- Statistically significant: Result not likely to be zero; not likely due to chance (sampling error)
  - Significance level (α): The *maximum* probability you are willing to tolerate that sampling error caused observed results: if probability is lower, results are *statistically significant* (at the level α)
    - Often α = 0.05, but arbitrary cut-off for surprising
    - Usually report *best* conventional significance level met
  - You compare the strength of your evidence against the significance level (the burden of proof)

4

Recall Coupon Ex. (p. 396)

- H_0: $p = 0.15$
- H_1: $p > 0.15$
- $\hat{P} = \frac{483}{3{,}000} = 0.161$; P-value = P($\hat{P} \geq 0.161$ | H_0) = 0.0458
  - Statistically signif. at 5% level
  - Say the *best* significance level reached of 1%, 5%, and 10%

$E[\hat{P} \mid H_0, n = 3{,}000] = 0.15$
$SD[\hat{P} \mid H_0, n = 3{,}000] = 0.00652$



Personally, how big of a sample proportion would you want to see to feel comfortable rejecting H_0 in favor of H_1? At least ___.

How does your answer relate to the α you would choose?

5

Two Methods of Hypothesis Testing

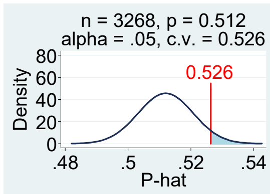
- P-value Approach: Discussed in Lecture 13
  - Small P-value (e.g. 0.009) means strong evidence against H_0 (null) and in favor of H_1 (research)
  - Big P-value (e.g. 0.207) means weak evidence against H_0 and in favor of H_1
- Rejection (Critical) Region Approach: Given significance level (α), find range such that if the test statistic falls in that range, reject H_0
- Connection: if P-value < α then reject H_0

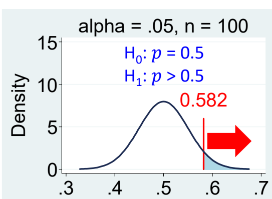
P-value measures strength of evidence, not just yes/no answer 6

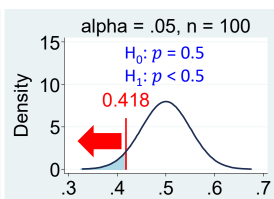
Rejection (Critical) Region

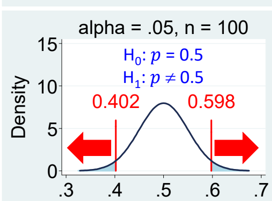
- Rejection Region: Range of values so that if the test statistic falls into it, reject H_0 in favor of H_1
  - Find sampling distribution of test statistic *if* H_0 were true
  - Critical value (c.v.): Edge of rejection region that depends on selected significance level (α)

Recall proportion male for parity=2 Indian-born moms; H_0: $p = 0.512$; H_1: $p > 0.512$; n = 3,268, $\hat{P} = 0.576$ (Lec. 13)









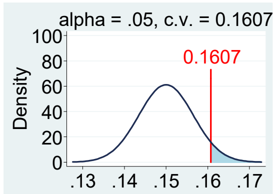
- 100 tosses of a coin to make inferences about its fairness
- Rejection regions shown:
  - (0.582, ∞)
  - (-∞, 0.418)
  - (-∞, 0.402) and (0.598, ∞)

In 400 tosses of a coin, what proportion heads would convince you it is unfair? Show work and illustrate with a fully-labeled graph.

In 400 tosses of a coin, 55% or more heads or 45% or fewer heads would convince me that the coin is unfair at a 5% significance level.

Coupon Ex: Unstandardized Rejection Region Approach

- H$_0$: p = 0.15
- H$_1$: p > 0.15
- Critical value, rejection region
  - $P(\hat{P} \geq c.v.) = \alpha$
  - Already verified Normal approx. (L13)



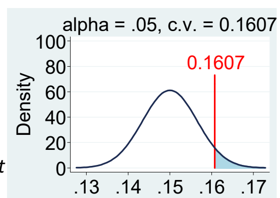
$$P\left(\hat{P} \geq 0.15 + 1.6449 * \sqrt{\frac{0.15 * 0.85}{3000}}\right) = 0.05$$

$$P(\hat{P} \geq .1607) = 0.05$$

10

H$_0$: p = 0.15; H$_1$: p > 0.15; $\hat{P}$ = 0.1610

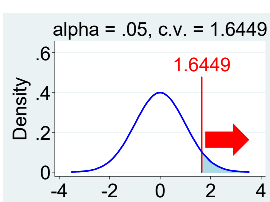
- Significance level = 0.05
- Critical value = 0.1607
- Rejection region = (0.1607, ∞)
- Test statistic = $\hat{P}$ = 0.1610
- Reject H$_0$ and infer H$_1$: the result is *statistically significant* at the 5% level, we've (sufficiently) proven that the redemption rate is above 15% (at least for a 5% burden of proof)



11

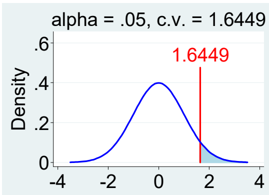
Standardized Rejection Region Approach

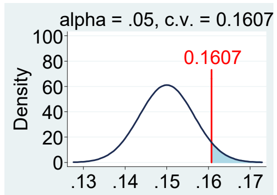
- H$_0$: p = 0.15
- H$_1$: p > 0.15
- Standardized critical value, rejection region
  - $P(Z \geq c.v.) = \alpha$
  - $P(Z \geq 1.6449) = 0.05$
- Standardized test statistic
  - $z = \frac{0.1610-0.15}{0.0065192} = 1.687$



12

Standardized vs. Unstandardized Approach





Standardized approach: Find the standardized rejection region (above) and then compare $z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ with it

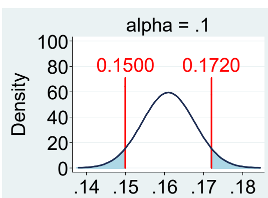
What is p_0?

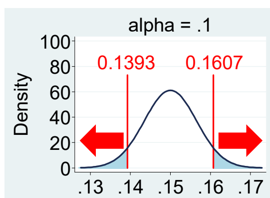
Unstandardized approach: Find the unstandardized rejection region (above) and then compare $\hat{P}$ with it

Where did 0.1607 come from?

13

Difference is Very Important





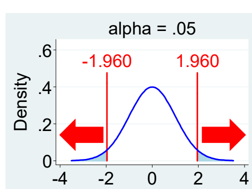
Confidence interval estimator centered at point estimate (best guess) ($\hat{P} = 0.1610$ and $n = 3000$)

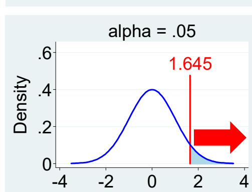
Rejection region for hypothesis test based on presumption that H_0 is true ($H_0: p = 0.15$ and $n = 3000$)

14

Two Tail Standard

- Economists often use two tailed tests
  - Even when a directional test seems more obvious
- One justification: two-tailed test is more conservative
  - Fewer statistically significant results





15

Recall Karlan and List (2007)

ratio	gave 0	1	Total
0	16,389	298	16,687
1	10,902	231	11,133
2	10,882	252	11,134
3	10,876	253	11,129
Total	49,049	1,034	50,083

(1) How big of an effect does offering a match have on the response rate? [Answering requires CI estimation: Lecture 12]

(2) Does offering a match affect the response rate? [Answering requires hypothesis testing]

“Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment” (See Lecture 12)

16

When testing proportions, null says no difference: $H_0: (p_1 - p_2) = 0$

- Comparing proportions: $H_0: (p_1 - p_2) = 0$
  - E.g. response rate w/ match is same as w/o match
  - Hence, under the presumption that the null is true, we pool the two groups together: $\bar{P} = \frac{X_1+X_2}{n_1+n_2}$ to get $SE[\hat{P}_1 - \hat{P}_2] = \sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}$
  - In *contrast*, for CI estimation (Lecture 12): $SE[\hat{P}_1 - \hat{P}_2] = \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$

17

(2) Hypothesis Testing: Does Matching Have an Effect?

- H_0: $p_2 - p_1 = 0$ (1 is control group; 2 is match treatment)
- H_1: $p_2 - p_1 \neq 0$
- $\hat{P}_1 = \frac{X_1}{n_1} = \frac{298}{16{,}687} = 0.01786$; $\hat{P}_2 = \frac{X_2}{n_2} = \frac{736}{33{,}396} = 0.02204$
- Pooled proportion: $\bar{P} = \frac{X_1+X_2}{n_1+n_2} = \frac{298+736}{16{,}687+33{,}396} = 0.02065$
- $z = \frac{\hat{P}_2-\hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1}+\frac{\bar{P}(1-\bar{P})}{n_2}}} = \frac{0.02204-0.01786}{\sqrt{0.02065(1-0.02065)\left(\frac{1}{16{,}687}+\frac{1}{33{,}396}\right)}} = 3.10$

Using Normal table, compute P-value as 0.002. (With just the Empirical (68-95-99.7) Rule, know it must be less than 0.003.)

Is the result statistically significant? If so, at which level?

Infer that offering a match *causes* a response rate change?

18

Labour Force Survey (LFS)

- Every month Statistics Canada runs the LFS
  - "Recently, the monthly LFS sample size has been approximately 56,000 households, resulting in the collection of labour market information for approximately 100,000 individuals [aged 15 years and over]."
- Does being born outside of Canada increase risk of being unemployed? If so, how much?

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1209254
(Retrieved Jan. 13, 2019)

19

Which populations are we comparing?

- p_1: Of those aged 25 – 54, born in Canada & in labor force, the proportion unemployed
- p_2: Of those aged 25 – 54, *not* born in Canada & in labor force, the proportion unemployed
- Is $\hat{P}_2 = 0.0780$ with $n_2 = 11{,}170$ a statistically higher unemployment rate than $\hat{P}_1 = 0.0541$ with $n_1 = 33{,}370$?
  - To answer, hypothesis testing or CI estimation?

Note: Sample sizes are realistic approximations given expected fraction of 100,000 people in survey to be in each of the two populations

20

Is unemployment rate higher for non-Canadian born people in LF?

- H_0: $p_2 - p_1 = 0$ (1 is Canadian-born and 2 is not)
- H_1: $p_2 - p_1 > 0$
- $\hat{P}_1 = 0.0541, n_1 = 33{,}370$; $\hat{P}_2 = 0.0780, n_2 = 11{,}170$
- Pooled proportion: $\bar{P} = \frac{X_1+X_2}{n_1+n_2} = \frac{1{,}805+871}{33{,}370+11{,}170} = 0.0601$
- $z = \frac{\hat{P}_2-\hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1}+\frac{\bar{P}(1-\bar{P})}{n_2}}} = \frac{0.0780-0.0541}{\sqrt{0.0601(0.9399)\left(\frac{1}{11{,}170}+\frac{1}{33{,}370}\right)}} = 9.20$

Approximate P-value? Conclusion?

21

How Big is the Difference?

- Canadian born unemp.: $\hat{P}_1 = 0.0541, n_1 = 33{,}370$
- Non-Canadian born unemp.: $\hat{P}_2 = 0.0780, n_2 = 11{,}170$

$$(0.0780 - 0.0541) \pm 1.96\sqrt{\frac{0.0541\,(0.9459)}{33{,}370} + \frac{0.0780(0.9220)}{11{,}170}}$$

$$0.0239 \pm 1.96 * 0.0028$$

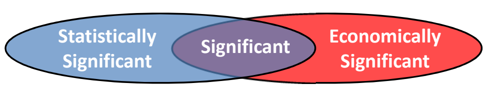
$$0.0239 \pm 0.0055$$

- LCL = 0.0184 and UCL = 0.0294

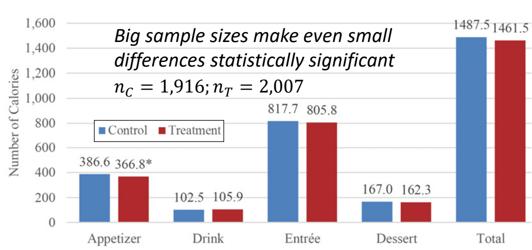
We are 95% confident the 2012 unemployment rate is between 1.84 and 2.94 percentage points higher for those born outside Canada. This is huge, but we cannot say discrimination against immigrants caused higher unemployment (observational data).

Significant: Has meaning; Is important

- Economically significant: Effect big enough so that decision makers would think it important
- Use "significant" for results that are *both* large enough to care about & statistically significant



  – E.g. difference in employment rates by birthplace is *significant*: a tiny P-value *and* a big difference



Excerpt (p. 10): Figure 2 shows average calories ordered by the treatment and control group, both overall and by course. Total calories is slightly lower for the treatment group (1,461.5 versus 1,487.5) but the difference is not statistically significant. The only ~~significant~~ [statistically significant] difference is in calories from appetizers. (http://www.nber.org/papers/w24889)

"Significant" only if __both__ statistically & economically significant

2018 *NBER Working Paper*: "The Impact of Information Disclosure on Consumer Behavior: Evidence from a Randomized Field Experiment of Calorie Labels on Restaurant Menus"