Lecture 23: more Chapter 9, Section 2 Inference for Categorical Variable: More About Hypothesis Tests

- Examples of Tests with 3 Forms of Alternative
- How Form of Alternative Affects Test
- When *P*-Value is “Small”: Statistical Significance
- Hypothesis Tests in Long-Run
- Relating Test Results to Confidence Interval

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-4)
  - Displaying and Summarizing (Lectures 5-12)
  - Probability (discussed in Lectures 13-20)
  - Statistical Inference
    - 1 categorical: confidence intervals, hypothesis tests
    - 1 quantitative
    - categorical and quantitative
    - 2 categorical
    - 2 quantitative

Hypothesis Test About *p (Review)*

State null and alternative hypotheses H_0 and H_a :

Null is "status quo", alternative "rocks the boat".

$$H_0 : p = p_0 \quad \text{vs.} \quad H_a : \left\{ \begin{array}{c} p > p_0 \\ p < p_0 \\ p \neq p_0 \end{array} \right\}$$

1. Consider sampling and study design.
2. Summarize with $\hat{p}$, standardize to z, assuming that $H_o : p = p_o$ is true; consider if z is "large".
3. Find P-value=prob.of z this far above/below/away from 0; consider if it is "small".
4. Based on size of P-value, choose H_0 or H_a.

Checking Sample Size: C.I. vs. Test

- Confidence Interval: Require observed counts in and out of category of interest to be at least 10.

$$n\hat{p} = X \geq 10$$
$$n(1-\hat{p}) = n - X \geq 10$$

- Hypothesis Test: Require expected counts in and out of category of interest to be at least 10 (assume $p = p_0$).

$$np_0 \geq 10$$
$$n(1-p_0) \geq 10$$

Example: *Checking Sample Size in Test*

- **Background**: 30/400=0.075 students picked #7 "at random" from 1 to 20. Want to test $H_0: p=0.05$ vs. $H_a: p>0.05$.
- **Question:** Is n large enough to justify finding P-value based on normal probabilities?
- **Response:**

$n p_0 =$

$n(1-p_0) =$

Looking Back: *For confidence interval, checked 30 and 370 both at least 10.*

 Practice: 9.36d p.439

Example: *Test with ">" Alternative (Review)*

- **Note**: Step 1 requires 3 checks:
  - Is sample unbiased? (Sample proportion has mean 0.05?)
  - Is population ≥10*n*? (Formula for s.d. correct?)
  - Are np_0 and $n(1-p_0)$ both at least 10? (Find or estimate *P*-value based on normal probabilities?)

1. Students are "typical" humans; bias is issue at hand.
2. If p=0.05, sd of $\hat{p}$ is $\sqrt{\frac{0.05(1-0.05)}{400}}$ and
$$z=\frac{0.075-0.05}{\sqrt{\frac{0.05(1-0.05)}{400}}}=+2.29$$
3. *P*-value = $P(Z\geq 2.29)$ is small: just over 0.01
4. Reject H_0, conclude Ha: picks were biased for #7.

Example: *Test with "Less Than" Alternative*

- **Background**: 111/230 of surveyed commuters at a university walked to school.

```
Test and CI for One Proportion
Test of p = 0.5 vs p < 0.5
Sample      X      N  Sample p  95.0% Upper Bound  Z-Value  P-Value
1         111    230  0.482609           0.536805    -0.53    0.299
```

- **Question:** Do fewer than half of the university's commuters walk to school?
- **Response:** First write H_0: ____ vs. H_a: ____

1. Students need to be rep. in terms of year. 115≥10
2. Output→ $\hat{p}$ = ____ z = ____ Large? ____.
3. P-value = ____. Small? ____
4. Reject H_0? ____ Conclude? ____

Example: *Test with "Not Equal" Alternative*

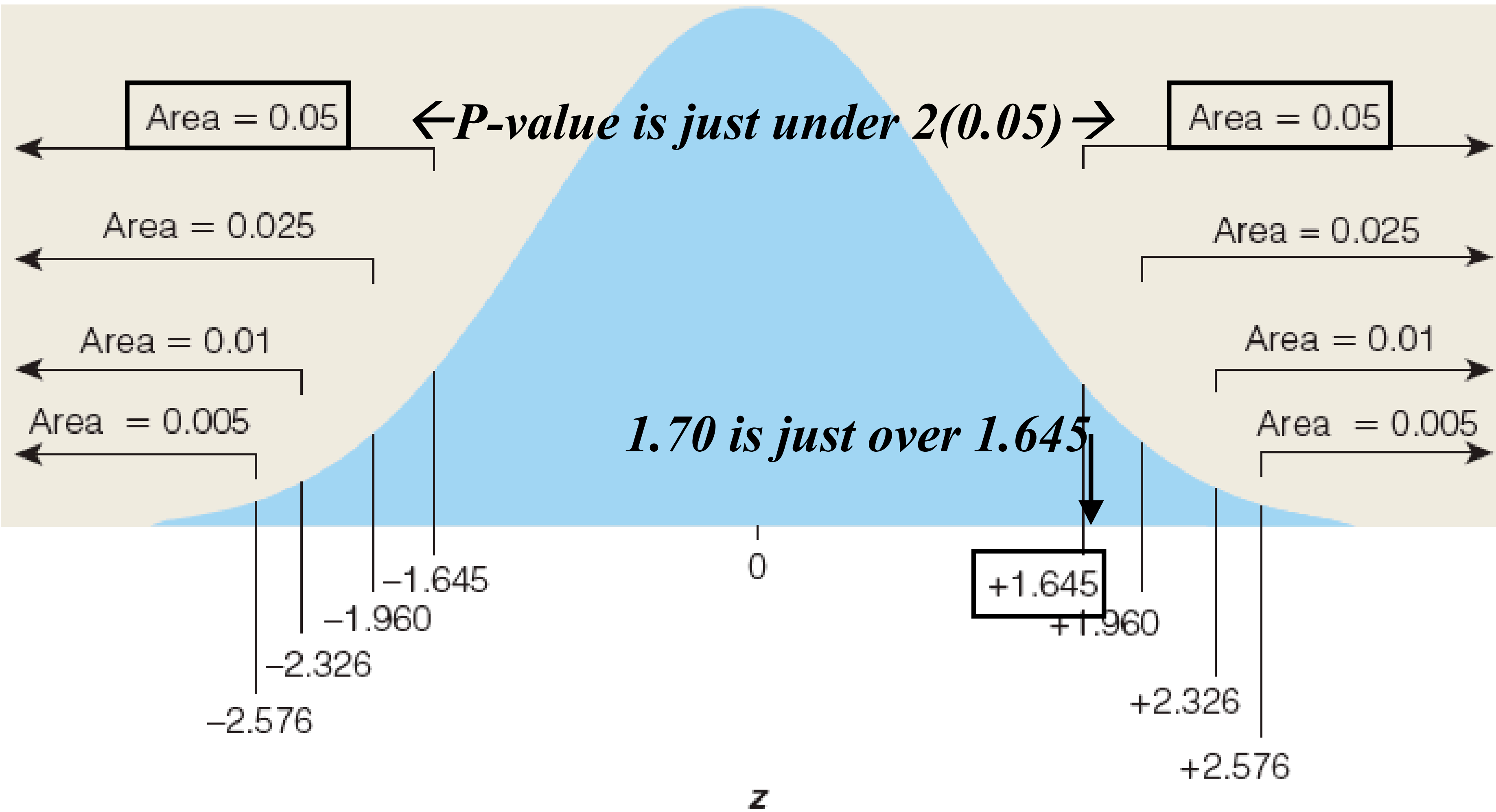
- **Background**: 43% of Florida's community college students are disadvantaged.
- **Question:** Is % disadvantaged at Florida Keys Community College (169/356=47.5%) unusual?

```
Test and CI for One Proportion
Test of p = 0.43 vs p not = 0.43
Sample      X     N  Sample p         95.0% CI          Z-Value  P-Value
1         169   356  0.474719  (0.422847, 0.526592)       1.70    0.088
```

- **Response:** First write $H_0:$ vs. $H_a:$

1. 356(0.43), 356(1-0.43) both ≥10; pop. ≥10(356)
2. $\hat{p} =$, $z =$
3. P-value = small? ______
4. Reject H_0? Is 47.5% unusual?

90-95-98-99 Rule to Estimate *P*-value



One-sided or Two-sided Alternative

- Form of alternative hypothesis impacts *P*-value
- *P*-value is *the* deciding factor in test
- Alternative should be based on what researchers hope/fear/suspect is true *before* "snooping" at the data
- If < or > is not obvious, use two-sided alternative (more conservative)

Example: *How Form of Alternative Affects Test*

- **Background**: 43% of Florida's community college students are disadvantaged.
- **Question:** Is % disadvantaged at Florida Keys Community College (47.5%) unusually **high**?

```
Test of p = 0.43 vs p > 0.43
Sample       X      N  Sample p  95.0% Lower Bound  Z-Value  P-Value
1          169    356  0.474719           0.431186     1.70    0.044
```

- **Response:** Now write H_0: $p = 0.43$ vs. H_a:

1. Same checks of data production as before.
2. Same $\hat{p}$ =0.475 *(Note: 0.475>0.43)*, same z=+1.70.
3. Now P-value = . Small?
4. Is 47.5% significantly higher than 43%?

Practice: 9.46d p.444

P-value for One- or Two-Sided Alternative

- *P*-value for one-sided alternative is **half** *P*-value for two-sided alternative.
- *P*-value for two-sided alternative is **twice** *P*-value for one-sided alternative.

For this reason, two-sided alternative is more conservative (larger *P*-value, harder to reject Ho).

Example: *Thinking About Data at Hand*

- **Background**: 43% of Florida's community college students are disadvantaged. At Florida Keys, the rate is 47.5%.
- **Question:** Is the rate at Florida Keys significantly **lower**?
- **Response:**

Definition; How Small is a "Small" *P*-value?

alpha (α): cut-off level which signifies a *P*-value is small enough to reject H_0

- Avoid blind adherence to cut-off α =0.05
- Take into account…
  - Past considerations: is H_0 "written in stone" or easily subject to debate?
  - Future considerations: What would be the consequences of either type of error?
    - Rejecting H_0 even though it's true
    - Failing to reject H_0 even though it's false

Example: *Reviewing P-values and Conclusions*

- **Background**: Consider our prototypical examples:
  - Are random number selections biased? *P*-value=0.011
  - Do fewer than half of commuters walk? *P*-value=0.299
  - Is % disadvantaged significantly different? *P*-value=0.088
  - Is % disadvantaged significantly higher? *P*-value=0.044
- **Question:** What did we conclude, based on *P*-values?
- **Response:** (Consistent with 0.05 as cut-off α)
  - *P*-value=0.011→ Reject H_0?
  - *P*-value=0.299→ Reject H_0?
  - *P*-value=0.088→ Reject H_0?
  - *P*-value=0.044→ Reject H_0?

Example: *Cut-Offs for "Small" P-Value*

- **Background**: Bookstore chain will open new store in a city if there's evidence that its proportion of college grads is higher than 0.26, the national rate.
- **Question:** Choose cut-off (0.10, 0.05, 0.01):
  - if no other info is provided
  - if chain is enjoying considerable profits; owners are eager to pursue new ventures
  - if chain is in financial difficulties, can't afford losses if unsuccessful due to too few grads
- **Response:**
  -
  -
  -

Practice: 9.56 p.445

Definition

Statistically significant data: produce P-value small enough to reject H_0. z plays a role:

$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{(\hat{p}-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$$

Reject H_0 if P-value small; if $|z|$ large; if…

- Sample proportion $\hat{p}$ far from p_0
- Sample size n large
- Standard deviation small (if p_0 is close to 0 or 1)

Role of Sample Size *n*

- **Large *n***: may reject H_0 even though observed proportion isn't very far from p_0, from a practical standpoint.

Very small *P*-value→strong evidence against Ho but *p* not necessarily very far from *po*.

- **Small *n***: may fail to reject H_0 even though it is false.

Failing to reject false Ho is 2nd type of error

Definition

- **Type I Error:** reject null hypothesis even though it is true (false positive)
  - Probability is cut-off α
- **Type II Error:** fail to reject null hypothesis even though it's false (false negative)

Hypothesis Test and Long-Run Behavior

20 coin flips	test Ho: p=.50 vs. Ha: p not equal .50 (reject if p-value<.05)	
TTTTTHTHTTHHTHTHTTHH proportion of heads 9/20=.45	z=-.45, p-value=.655	→ do not reject Ho
HTTHHTHHTTTHTHTTTHHT proportion of heads 8/20=.40	z=-.89, p-value=.371	→ do not reject Ho
TTTHHHHHHHHTHTHHTHTT proportion of heads 12/20=.60	z=+.89, p-value=.371	→ do not reject Ho
THHHHTHHHTHTHHHTHHHH proportion of heads 15/20=.75	z=+2.24, p-value=.025	→ reject Ho
• • • repeated flips of 20 coins	• • •	in the long run 95% of tests do not reject Ho 5% of tests reject Ho
TTHTTTTHTTTHTTHHTHHH proportion of heads 8/20=.40	z=-.89, p-value=.371	→ do not reject Ho

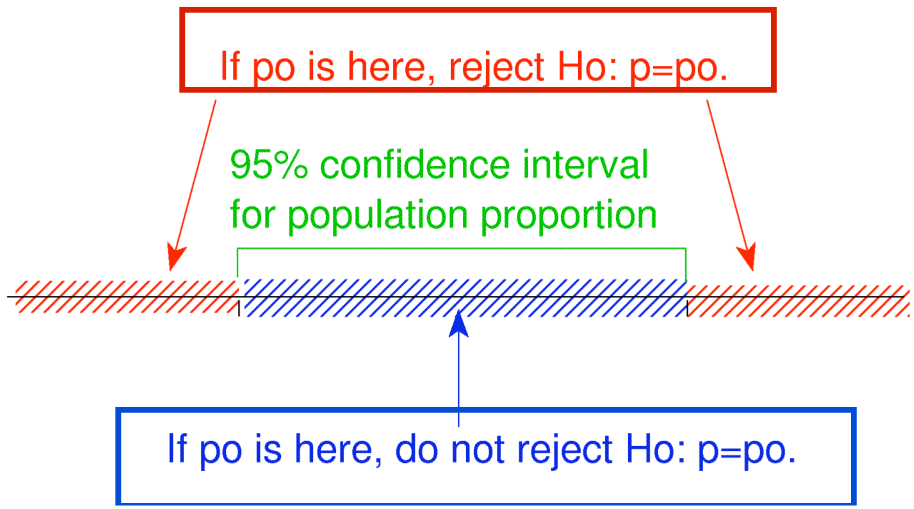
Confidence Interval and Hypothesis Test Results

- Confidence Interval: range of plausible values
- Hypothesis Test: decides if a value is plausible

Informally,

- If *po* is in confidence interval, don't reject Ho: *p*=*po*
- If *po* is outside confidence interval, reject Ho: *p*=*po*

Relationship between 95% confidence interval
and two-sided test with .05 as cut-off for p-value



Example: *Test Results, Based on C.I.*

- **Background**: A 95% confidence interval for proportion of all students choosing #7 "at random" from numbers 1 to 20 is (0.055, 0.095).
- **Question:** Would we expect a hypothesis test to reject the claim p=0.05 in favor of the claim p>0.05?
- **Response:**

Example: *C.I. Results, Based on Test*

- **Background**: A hypothesis test did not reject H_0: p=0.5 in favor of the alternative H_a: p<0.5.
- **Question:** Do we expect 0.5 to be contained in a confidence interval for p?
- **Response:**

Lecture Summary
(More Hypothesis Tests for Proportions)

- Examples with 3 forms of alternative hypothesis
- Form of alternative hypothesis
  - Effect on test results
  - When data render formal test unnecessary
  - *P*-value for 1-sided vs. 2-sided alternative
- Cut-off for "small" *P*-value
- Statistical significance; role of n, Type I or II Error
- Hypothesis tests in long-run
- Relating tests and confidence intervals