Lecture 15: Chapter 9, Section 2 Inference for Categorical Variable: Hypothesis Tests

- 4 steps in Hypothesis Test; Posing Hypotheses
- Details of 4 Steps, Definitions and Notation
- 3 Forms of Alternative Hypothesis
- *P*-Value
- Example with “Greater Than” Alternative
- 3 Forms of Alternative; Effect on Conclusion
- Small P-values and Statistical Significance

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-3)
  - Displaying and Summarizing (Lectures 3-8)
  - Probability (discussed in Lectures 9-14)
  - Statistical Inference
    - 1 categorical: conf. ints. (L14), hypothesis tests
    - 1 quantitative
    - categorical and quantitative
    - 2 categorical
    - 2 quantitative

Three Types of Inference Problem *(Review)*

In a sample of 446 students, 0.55 ate breakfast.

1. What is our best guess for the proportion of all students who eat breakfast?

   Point Estimate

2. What interval should contain the proportion of all students who eat breakfast?

   Confidence Interval

3. Do more than half (50%) of all students eat breakfast?

   Hypothesis Test

4 Steps in Hypothesis Test About p

(First pose question as choice between 2 opposing views about p.)

1. Check data production for bias, etc.
2. We summarize with $\hat{p}$, standardize to z.
3. Find probability of $\hat{p}$ this extreme.
4. Perform inference, drawing conclusions about population proportion p.

These correspond to 4 Processes of Statistics.

Example: *Posing Hypothesis Test Question*

- **Background**: In a sample of 446 students, 0.55 ate breakfast. Do more than half of all students at that university eat breakfast?
- **Question:** How can we pose above question as two opposing points of view about p?
- **Response:**

Practice: 9.36a p.439

4 Steps in Hypothesis Test About p

(First pose question as choice between 2 opposing views about p.)

1. Check data production for bias.
2. We summarize with $\hat{p}$, standardize to z.
3. Find probability of $\hat{p}$ this extreme.
4. Perform inference, drawing conclusions about population proportion p.

Example: *Considering Data Production*

- **Background**: In a sample of 446 college students, 0.55 ate breakfast. We want to draw conclusions about breakfast habits of all students at that university.
- **Question:** What data production issues should be considered?
- **Response:** (discussed with confidence intervals)
  - Sampling: ______________________ } *(for claims*
  - Study design: ____________________ } *about ________)*

Also, *(for claims about________)* is population ≥ 10n?

And *(for claims about ________)* is n large enough?

Practice: 9.36b-d p.439

4 Steps in Hypothesis Test About p

(First pose question as choice between 2 opposing views about *p*.)

1. Check data production for bias.
2. We summarize with $\hat{p}$, standardize to z.
3. Find probability of $\hat{p}$ this extreme.
4. Perform inference, drawing conclusions about population proportion *p*.

Behavior of Sample Proportion *(Review)*

For random sample of size *n* from population with *p* in category of interest, sample proportion $\hat{p}$ has

- mean *p*
- standard deviation $\sqrt{\frac{p(1-p)}{n}}$

Hypothesis test: assume pop. proportion *p* is proposed value (=0.50 for breakfast example).

Looking Back: *For confidence intervals, we had to substitute sample proportion for unknown p.*

Example: *Summarizing and Standardizing*

- **Background**: In a sample of 446 students, 0.55 ate breakfast. Do more than half (0.50) of all students at that university eat breakfast?
- **Question:** How do we summarize the data?
- **Response:** Summarize with ______________.
  Standardize to

So 0.55 is ______ standard deviations above 0.50: pretty unusual.

Practice: 9.36e p.439

4 Steps in Hypothesis Test About p

(First pose question as choice between 2 opposing views about p.)

1. Check data production for bias.
2. We summarize with $\hat{p}$, standardize to z.
3. Find probability of $\hat{p}$ this extreme.
4. Perform inference, drawing conclusions about population proportion p.

Example: *Estimating Relevant Probability*

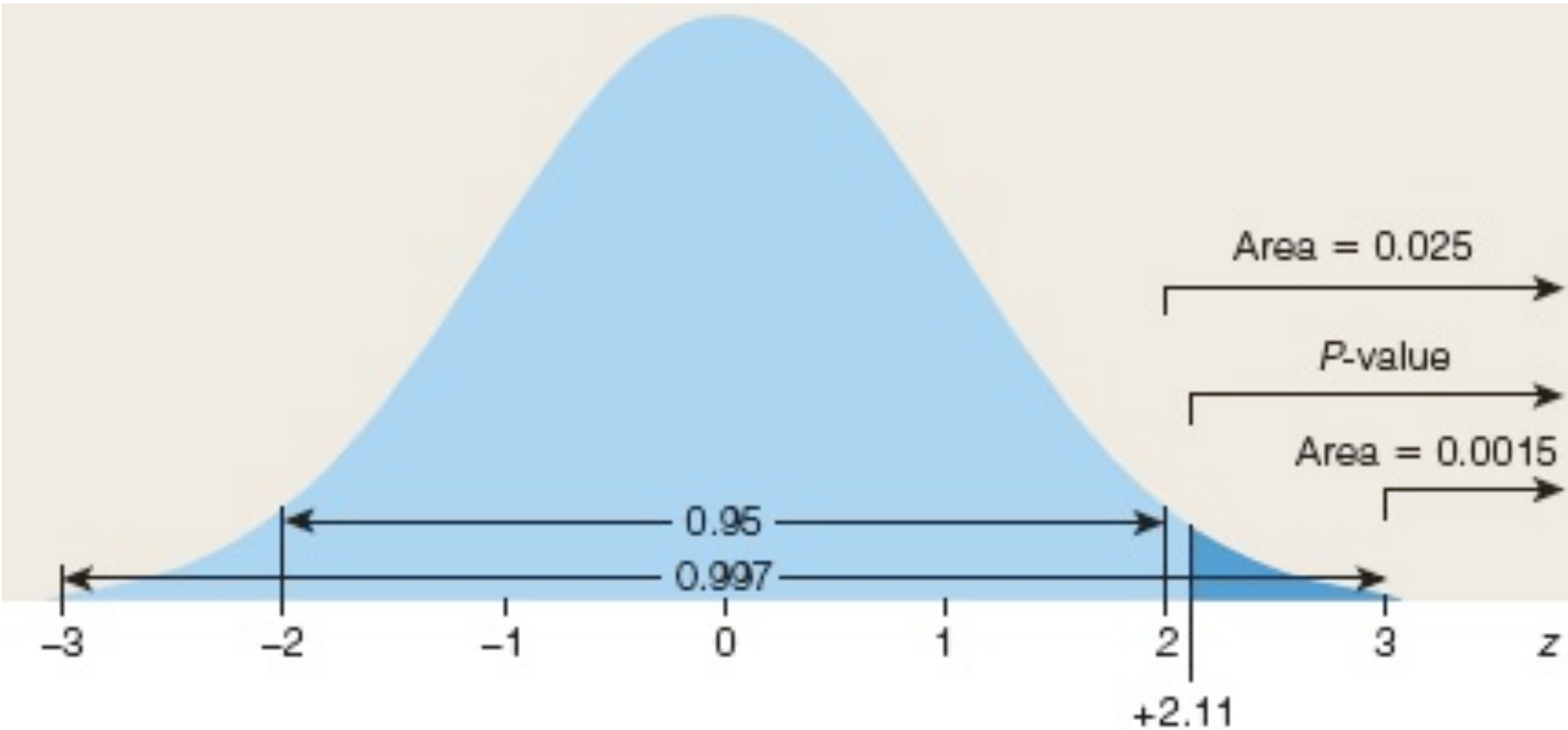
- **Background**: In a sample of 446 students, 0.55 ate breakfast. Do more than half of all students eat breakfast? We summarized with $\hat{p} = 0.55$ and $z = \frac{0.55 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{446}}} = +2.11$
- **Question:** If p=0.50, how unlikely is it to get $\hat{p}$ as high as 0.55 (that is, for z to be ≥ +2.11)?
- **Response:** 68-95-99.7 Rule→ since 2.11 > 2,

P(Z ≥ +2.11) is ______________________________

Such a probability can be considered to be ______________.

Practice: 9.36f p.439

Illustration of Relevant Probability



Looking Ahead: *The relevant probability for testing a hypothesis will be defined as the* ***P-value****.*

4 Steps in Hypothesis Test About p

(First pose question as choice between 2 opposing views about p.)

1. Check data production for bias.
2. We summarize with $\hat{p}$, standardize to z.
3. Find probability of $\hat{p}$ this extreme.
4. Perform inference, drawing conclusions about population proportion p.

Example: *Drawing Conclusions About p*

- **Background**: In a sample of 446 students, 0.55 ate breakfast. Do more than half of all students eat breakfast? We summarized with $\hat{p} = 0.55$ and $z = \frac{0.55 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{446}}} = +2.11$

The probability of z being +2.11 or higher is less than (1-0.95)÷2= 0.025 (fairly unlikely).

- **Question:** What do we conclude about *p*?
- **Response:**

Hypothesis Test About *p* *(More Details)*

First state 2 opposing views about *p*, called null and alternative hypotheses H_o and H_a.

1. Consider sampling and study design as for C.I.
2. Summarize with $\hat{p}$; does it tend in the suspected direction? Standardize to *z*, assuming $p = p_o$ (p_o is proposed value); consider if *z* is "large".
3. Find prob. of $\hat{p}$ this high/low/different, called '*P*-value' of the test; consider if it is "small".
4. Draw conclusions about *p*: choose between null and alternative hypotheses. (Statistical Inference)

Definitions

- **Null hypothesis** H_o: claim that parameter equals proposed value.
- **Alternative hypothesis** H_a : claim that parameter differs in some way from proposed value.
- ***P*-value:** probability, assuming H_o is true, of obtaining sample data at least as extreme as what has been observed.

Looking Back: *We considered the **probability**, assuming p=0.5 cards are red, of getting as few as 0 red cards in 4 or 5 picks.*

Notation

Proposed value of population proportion: p_o

Null and alternative hypotheses in test about unknown population proportion:

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

Looking Ahead: *The form of the alternative hypothesis will affect Steps 2, 3, 4 of the test.*

Example: *What Are We Testing About?*

- **Background**: Consider 3 problems:
  - 30/400=0.075 students picked #7 "at random" from 1 to 20. Is this evidence of bias for #7?
  - Do fewer than half of commuters walk? 111/230 of surveyed commuters at a university walked.
  - % disadvantaged in Florida community colleges is 43%. Is Florida Keys College unusual with 47.5% disadvantaged?
- **Question:** In each case, are we trying to draw conclusions about a sample proportion $\hat{p}$ or a population proportion p?
- **Response:** ____________________

Looking Ahead: *We'll refer to sample proportion later, to decide which of two claims to believe about the unknown population proportion.*

Example: *Three Forms of Alternative*

- **Background**: Consider 3 problems:
  - 30/400=0.075 students picked #7 "at random" from 1 to 20. Is this evidence of bias for #7?
  - Do fewer than half of commuters walk? 111/230 of surveyed commuters at a university walked.
  - % disadvantaged in Florida community colleges is 43%. Is Florida Keys College unusual with 47.5% disadvantaged?
- **Question:** How do we write the hypotheses in each case?
- **Response:**
  - ______________________________
  - ______________________________
  - ______________________________

Practice: 9.68a p.449

Definitions

- **One-sided alternative hypothesis** refutes equality with > or < sign
- **Two-sided alternative hypothesis** features a not-equal sign

Note: For a one-sided alternative, sometimes the accompanying null hypothesis is written as a (not strict) inequality. Either way, the same conclusions will be reached.

Assessing Merit of Data in One-Sided Test

If sample proportion does not tend in the direction claimed by alternative hypothesis in a 1-sided test, there is no need to proceed further.

Example: *When Test Can Be Cut Short*

- **Background**: The moon has four phases: new moon, first quarter, full moon, and last quarter, each in effect for 25% of the time. A neurologist whose patients claimed their seizures tended to be triggered by a full moon found 20% of 470 seizures were at full moon.
- **Question:** Do we need to carry out all 4 steps in the test?
- **Response:**

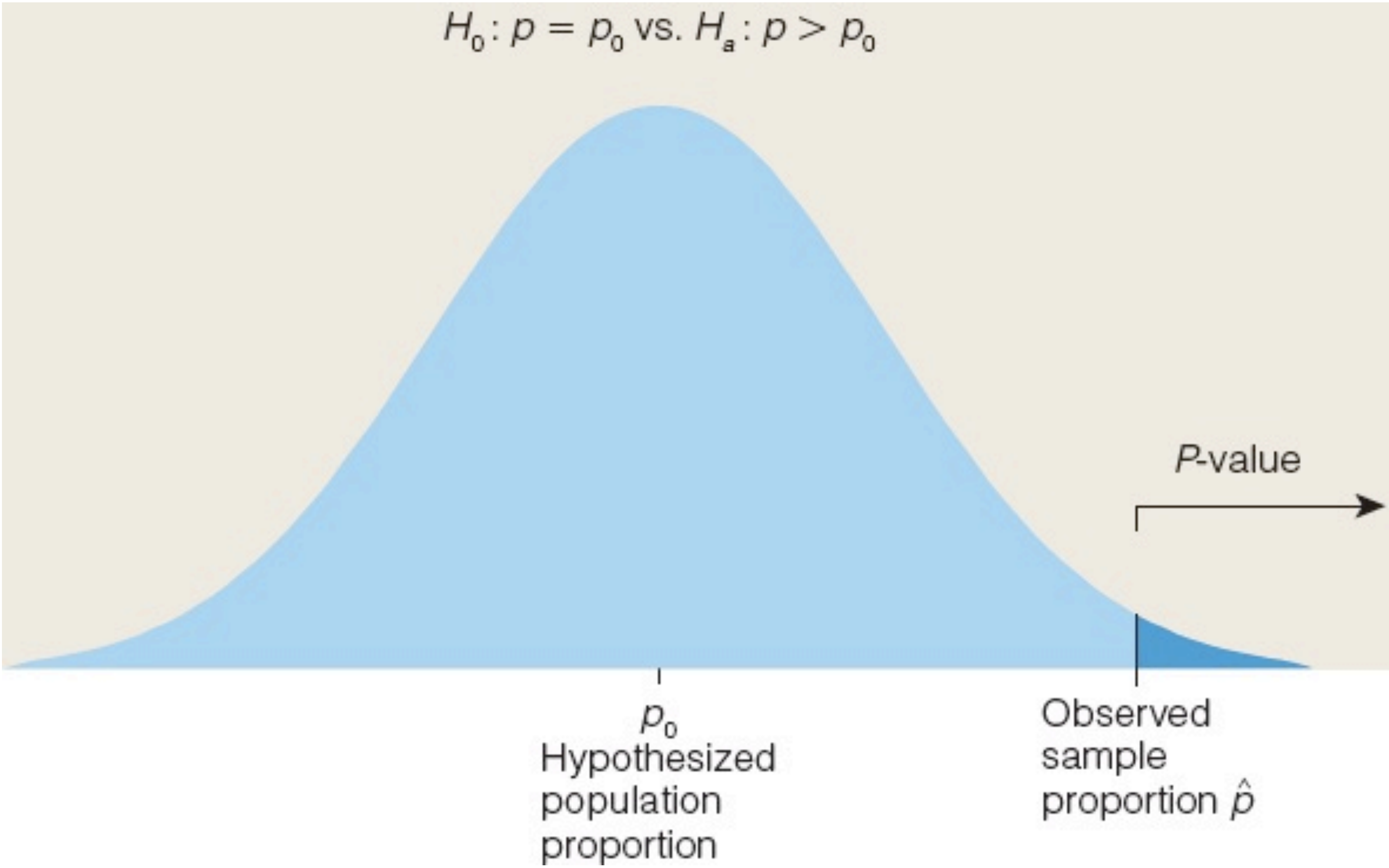
Practice: 9.40a p.441

How to Assess *P*-Value

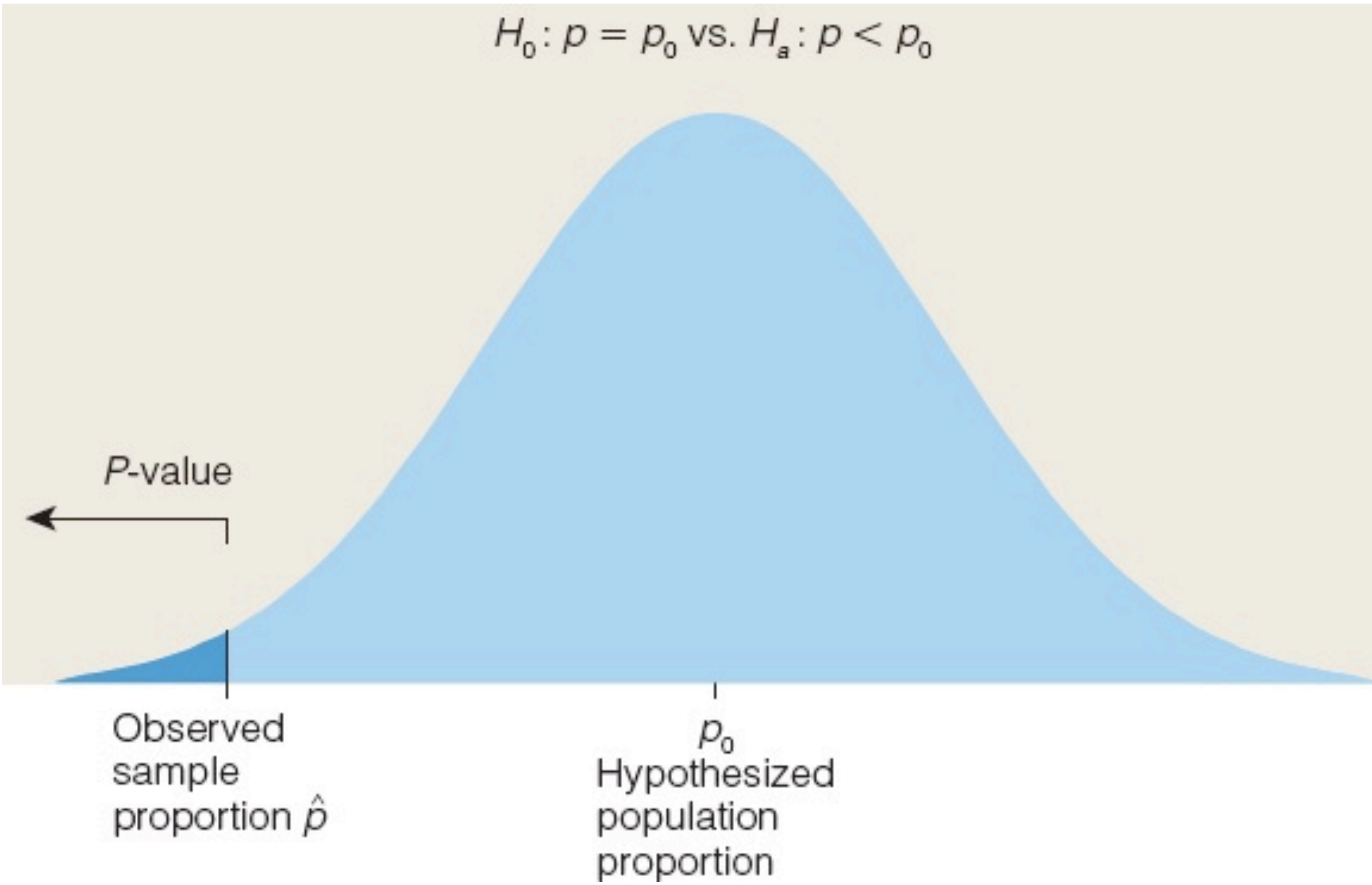
***P*-value:** probability, assuming H_o is true, of obtaining sample data at least as extreme as what has been observed. How to find *P*-value depends on form of alternative hypothesis:

- Right-tailed probability for $H_a : p > p_o$
- Left-tailed probability for $H_a : p < p_o$
- Two-tailed probability for $H_a : p \neq p_o$

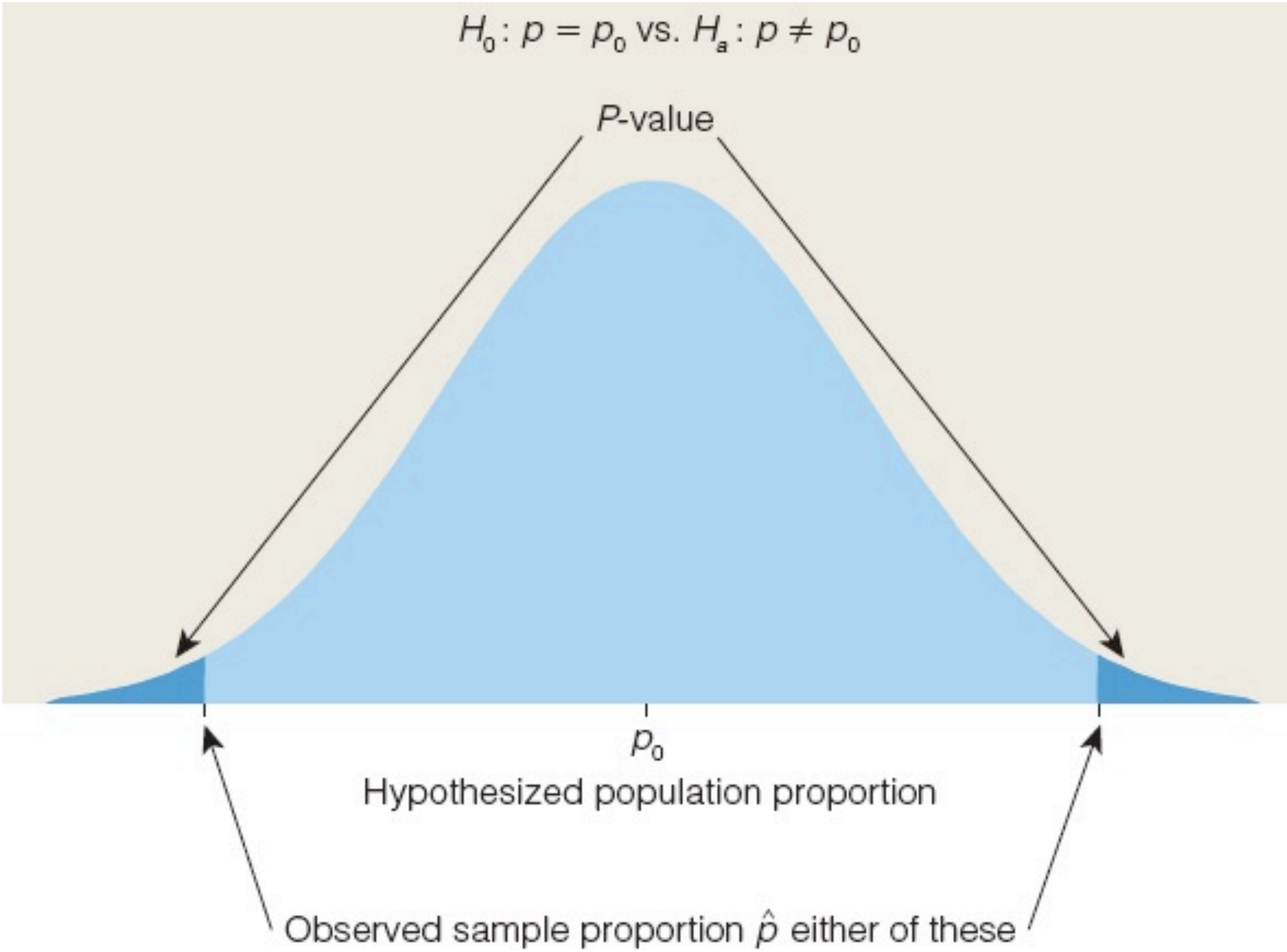
P-Value for $H_a : p > p_o$ is Right-tailed Probability



P-Value for $H_a : p < p_o$ is Left-tailed Probability



P-Value for $H_a : p \neq p_o$ is Two-tailed Probability



Drawing Correct Conclusions

Two possible conclusions:

- *P*-value small→reject H_o→conclude H_a. State we have evidence in favor of H_a.

  (not same as **proving** H_a true and H_o false).

- *P*-value not small→don't reject H_o →conclude H_o may be true.

  (not same as **proving** H_o true and H_a false)

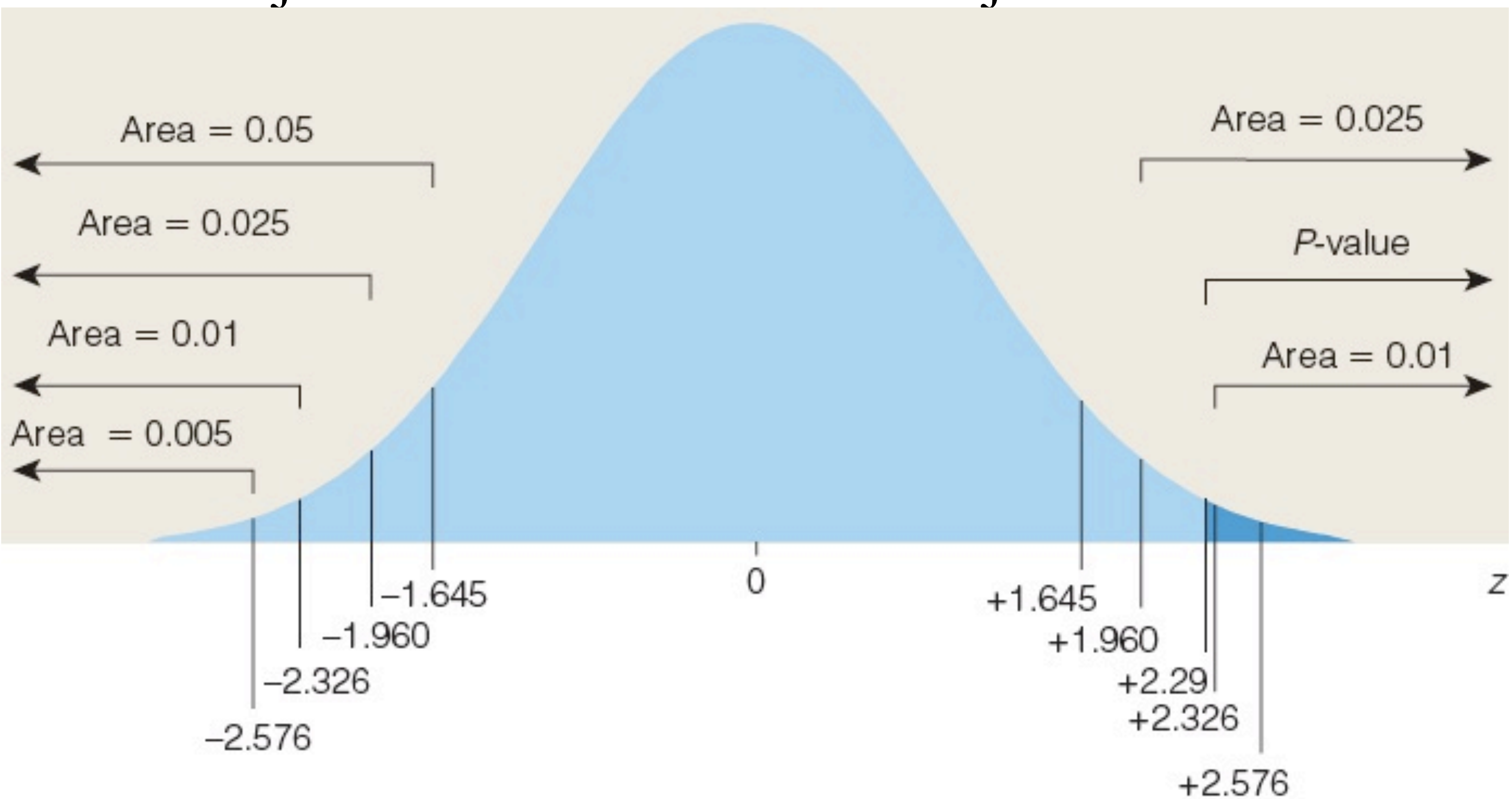
Example: *Test with "Greater Than" Alternative*

- **Background**: 30/400=0.075 students picked #7 "at random" from 1 to 20.
- **Question:** In general, is p>0.05? (evidence of bias?)
- **Response:** First write H_o:__________ vs. H_a: __________

1. Students are "typical" humans; bias is issue at hand.
2. 0.075 > 0.05 so the *sample* did favor #7. If p = 0.05, $\hat{p}$ standardizes to z =
3. P-value = ______________________________
4. Reject H_o? _____ Conclude? ________________________

Assessing a *P*-value with 90-95-98-99 Rule

2.29 just under 2.326 ➔ *P*-value just over 0.01



Three Types of Inference Problem *(Review)*

In a sample of 446 students, 0.55 ate breakfast.

1. What is our best guess for the proportion of all students who eat breakfast?

   Point Estimate

2. What interval should contain the proportion of all students who eat breakfast?

   Confidence Interval

3. Do more than half (50%) of all students eat breakfast?

   Hypothesis Test

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_o and $n(1-p_o)$ 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*

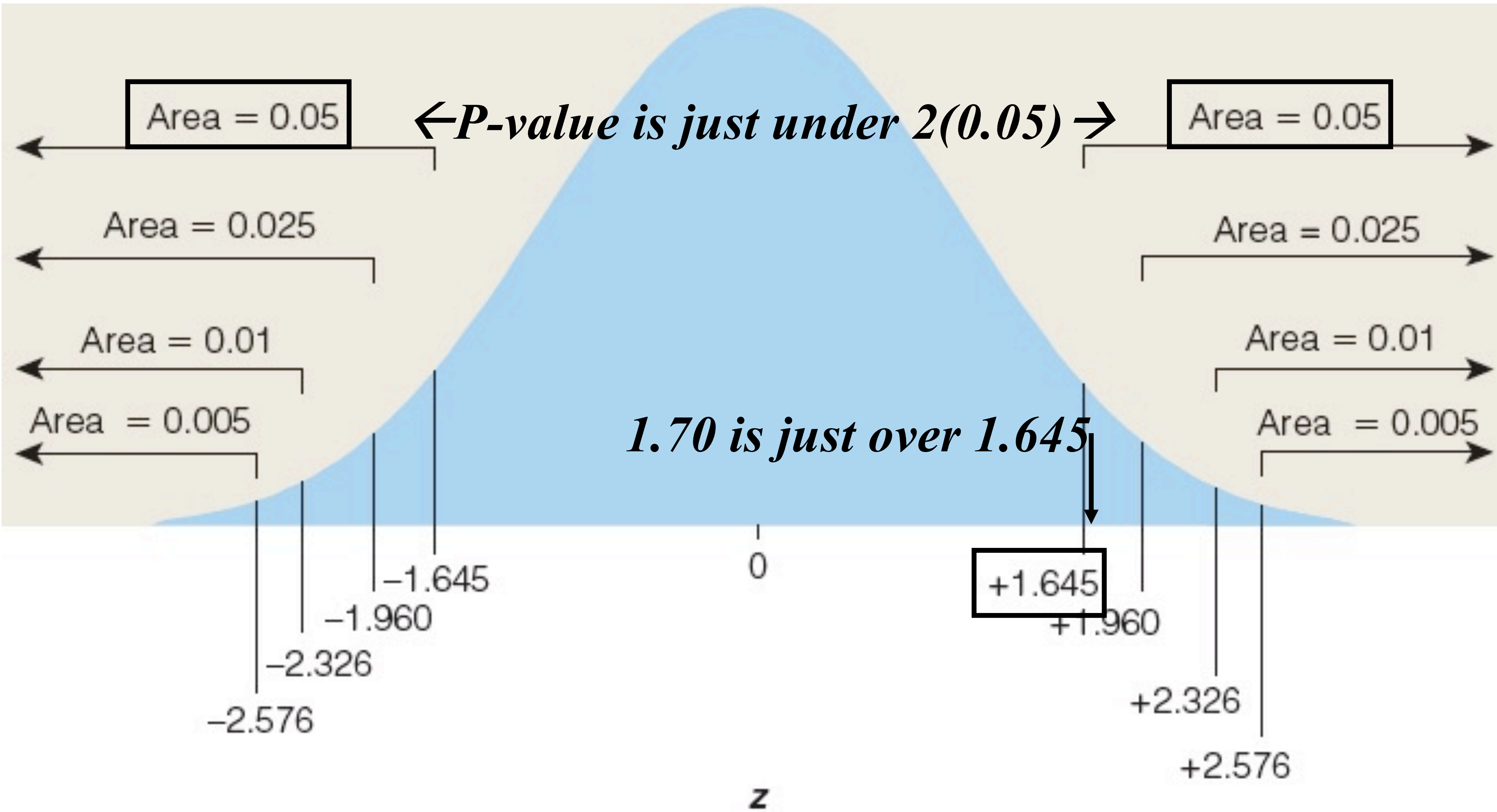
- **Background**: 43% of Florida's community college students are disadvantaged.
- **Question:** Is % disadvantaged at Florida Keys Community College (169/356=47.5%) unusual?

```
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%? ______

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).

Thinking About Data

Before getting caught up in details of test, consider evidence at hand.

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:**

Practice: 9.93a p.459

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 $\alpha = 0.05$
- 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? _____

Lecture Summary
(Inference for Proportions: Hypothesis Test)

- 4 steps in hypothesis test
  - Checking data production
  - Summarizing and standardizing
  - Finding a probability (*P*-value)
  - Conclusions as inference
- Posing null and alternative hypotheses
- Definitions and notation
- 3 forms of alternative hypothesis
- Assessing *P*-value
- Example with "greater than" alternative

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