

STAT 220 Lecture 16
Hypothesis Testing About One Population Proportion
One-Sided & Two-Sided Tests

Yibi Huang
Department of Statistics
University of Chicago

Example: Children's Play Preference

An observational study conducted at **Chicago Children's Museum** tried to determine the age when children start preferring a **same-sex** peer, instead of a **gender-neutral** playmate.

- ▶ Among 5-year-olds, 78 out of 162 chose a same-sex peer (48%)
- ▶ Among 6-year-olds, 59 out of 97 chose a same-sex peer (61%)

Under the *null hypothesis* of *no preference*, the probability a child chooses a same-sex peer is $p = 0.5$.

Do 6-year-olds (6YO) prefer to play with a same-sex peer more than a different-sex peer?

Example: Children's Play Preference – Population, Parameter, Hypotheses

- ▶ *Population* of interest:
- ▶ *Parameter* of interest:
- ▶ Why is the sample percent (61%) higher than expected (50%)?

Example: Children's Play Preference – Population, Parameter, Hypotheses

- ▶ *Population* of interest: all 6 year-old children (6YO)
- ▶ *Parameter* of interest:
- ▶ Why is the sample percent (61%) higher than expected (50%)?

Example: Children's Play Preference – Population, Parameter, Hypotheses

- ▶ *Population* of interest: all 6 year-old children (6YO)
- ▶ *Parameter* of interest: proportion p of 6YO preferring a same-sex peer
- ▶ Why is the sample percent (61%) higher than expected (50%)?

Example: Children's Play Preference – Population, Parameter, Hypotheses

- ▶ *Population* of interest: all 6 year-old children (6YO)
- ▶ *Parameter* of interest: proportion p of 6YO preferring a same-sex peer
- ▶ Why is the sample percent (61%) higher than expected (50%)?
 - ▶ Explanation 1: The true population proportion p is above 50%.
 - ▶ Explanation 2: The true population proportion is 50%, and this difference is just due to sampling variation.

Example: Children's Play Preference – Population, Parameter, Hypotheses

- ▶ *Population* of interest: all 6 year-old children (6YO)
- ▶ *Parameter* of interest: proportion p of 6YO preferring a same-sex peer
- ▶ Why is the sample percent (61%) higher than expected (50%)?
 - ▶ Explanation 1: The true population proportion p is above 50%.
 - ▶ Explanation 2: The true population proportion is 50%, and this difference is just due to sampling variation.
- ▶ *Null Hypothesis* H_0 : $p = 0.5$ (6YO are gender-neutral)
- ▶ *Alternative Hypothesis* H_A : $p > 0.5$ (6YO are more likely to choose a same-sex peer)

Incorrect Statements of H_0 and H_A

- ▶ H_0 and H_A are always about **population parameters** (not sample statistics).
- ▶ Incorrect hypothesis statements:

$$H_0 : \hat{p} = 0.5, \quad H_A : \hat{p} > 0.5$$

where \hat{p} is the sample proportion, not the population proportion p .

- ▶ Also, statements about the sample (e.g., “50% of the 97 six-year-olds...”) are incorrect for hypotheses.
- ▶ The correct form:

$$H_0 : p = 0.5, \quad H_A : p > 0.5$$

Always specify what p is (here, p = proportion of 6YO choosing a same-sex peer).

Example: Children's Play Preference — Sampling Distribution

What is the sampling distribution for the sample proportion \hat{p} under $H_0: p = 0.5$?

Example: Children's Play Preference — Sampling Distribution

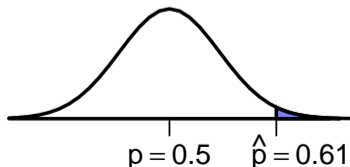
What is the sampling distribution for the sample proportion \hat{p} under $H_0: p = 0.5$?

For $X \sim \text{Bin}(n, p)$, when $np \geq 10$ and $n(1 - p) \geq 10$,

$$\hat{p} = \frac{X}{n} \text{ is approx. } \sim N \left(p, \sqrt{\frac{p(1-p)}{n}} \right)$$

Under $H_0: p = 0.5$, as $np = 97(0.5)$ and $n(1 - p) = 97(1 - 0.5)$ are both over 10, we know

$$\hat{p} \sim N \left(p = 0.5, \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(1-0.5)}{97}} \approx 0.051 \right)$$



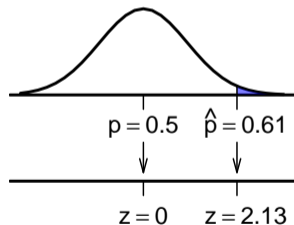
Example: Children's Play Preference — Test Statistic

Is the observed $\hat{p} = 59/97 \approx 0.61$ unusual, given $H_0: p = 0.5$?

$$\hat{p} \sim N \left(p = 0.5, \text{SE} = \sqrt{\frac{0.5(1-0.5)}{97}} \approx 0.051 \right)$$

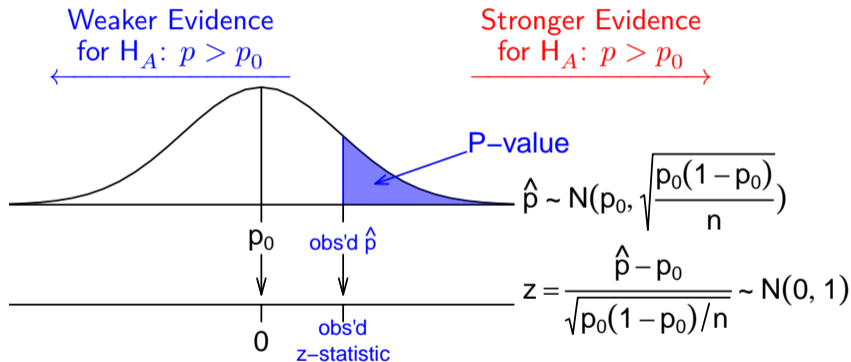
We can use the **z-score** of the sample proportion relative to the distribution above, as the test-statistic, called the **z-statistic**,

$$\begin{aligned} \text{z-statistic} &= \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.61 - 0.5}{\sqrt{0.5(1-0.5)/97}} \approx 2.13 \\ &\sim N(0, 1) \quad \text{under } H_0 : p = p_0 = 0.5 \end{aligned}$$



P-value for Upper One-Sided Test

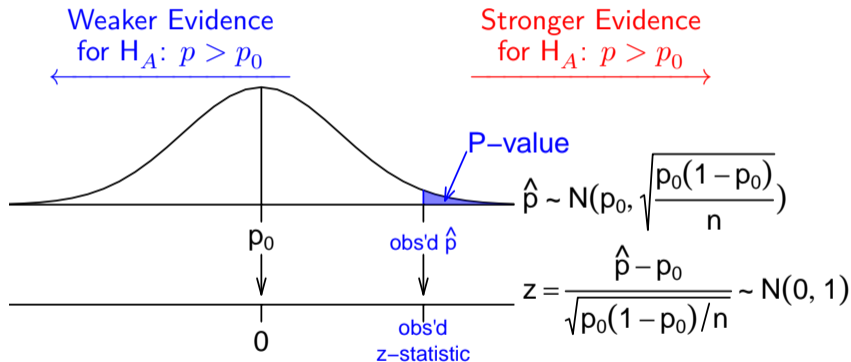
To test $H_0: p = p_0$ against the **upper one-sided** $H_A: p > p_0$, the P-value is the **upper-tail** probability (blue region below).



```
pnorm(z, lower.tail=FALSE) # Finding P-value in R  
1-pnorm(z)                # or
```

P-value for Upper One-Sided Test

To test $H_0: p = p_0$ against the **upper one-sided** $H_A: p > p_0$, the P-value is the **upper-tail** probability (blue region below).



```
pnorm(z, lower.tail=FALSE) # Finding P-value in R  
1-pnorm(z)                # or
```

Example (Children's Play Preference)

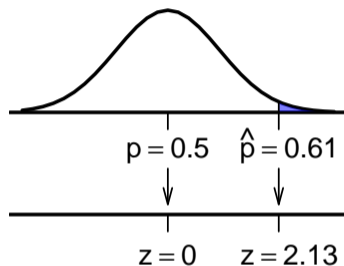
For testing $H_0: p = 0.5$ v.s. $H_A: p > 0.5$,

$$P\text{-value} = P(\hat{p} > 0.5 \mid p = 0.5)$$

$$= P\left(Z > \frac{0.61 - 0.5}{\sqrt{0.5(1 - 0.5)/97}} \approx 2.13\right)$$

$$= 1 - 0.9834$$

$$\approx 0.0166$$



Two-Sided Hypothesis

When children become aware of gender difference, they may tend to:

- ▶ choose a same-sex peer, or
- ▶ choose a different-sex peer.

Both “majority choosing same-sex” and “majority choosing diff-sex” are evidence of awareness.

So the alternative hypothesis is

$$H_A : p \neq 0.5$$

Two-Sided Hypothesis Tests

Testing $H_0: p = p_0$ vs. two-sided alternative $H_A: p \neq p_0$, the test statistic is still:

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

In this case, a sample proportion \hat{p} **far above** or **far below** p_0 are both evidence in favor of H_A .

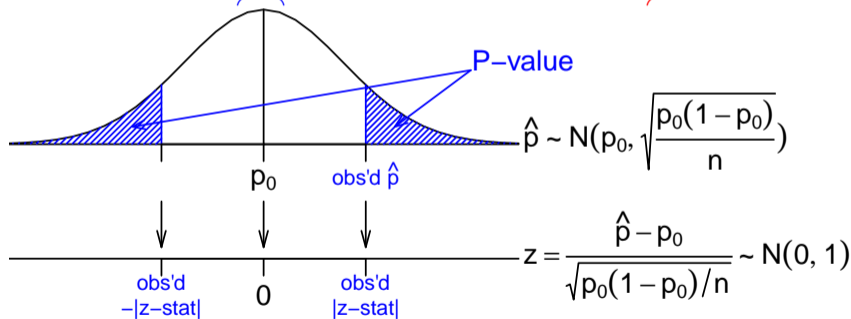
P-values for Two-Sided Hypothesis Tests

Testing $H_0: p = p_0$ vs. two-sided alternative $H_A: p \neq p_0$,
the P -value is the **two-tail probability** (blue regions below).

Stronger evidence
for $H_A: p \neq p_0$

Weak evidence
for $H_A: p \neq p_0$

Stronger evidence
for $H_A: p \neq p_0$



```
2*pnorm(abs(z), lower.tail=FALSE) # Finding P-value in R
```

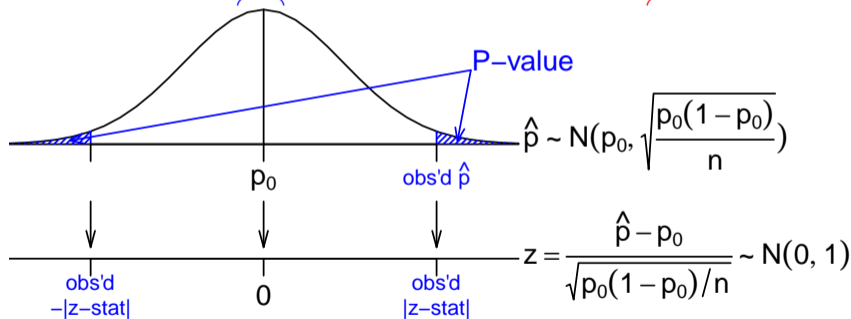
P-values for Two-Sided Hypothesis Tests

Testing $H_0: p = p_0$ vs. two-sided alternative $H_A: p \neq p_0$, the P -value is the **two-tail probability** (blue regions below).

Stronger evidence
for $H_A: p \neq p_0$

Weak evidence
for $H_A: p \neq p_0$

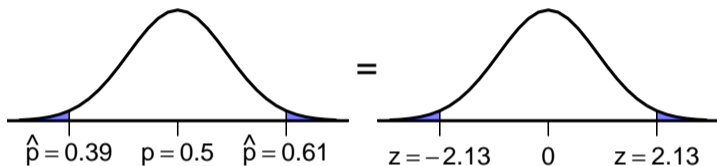
Stronger evidence
for $H_A: p \neq p_0$



```
2*pnorm(abs(z), lower.tail=FALSE) # Finding P-value in R
```

Example (Children's Play Preference)

For testing $H_0: p = 0.5$ v.s. $H_A: p \neq 0.5$,



$$\begin{aligned} P\text{-value} &= 0.0166 \times 2 \\ &= 0.0332 \end{aligned}$$

Lower One-Sided Hypothesis Test

If we want to know whether 6YO prefer **different**-sex peers ($< 50\%$ choose same-sex), then:

$$H_A : p < 0.5$$

Three types of alternative hypotheses:

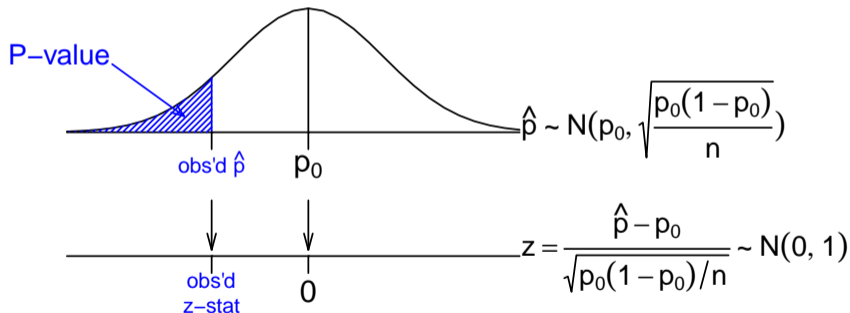
- ▶ Upper one-sided: $H_A : p > 0.5$
- ▶ Lower one-sided: $H_A : p < 0.5$
- ▶ Two-sided: $H_A : p \neq 0.5$

P-values for Lower One-Sided Hypothesis Tests

To test $H_0: p = p_0$ vs. **lower one-sided** alternative $H_A: p < p_0$, the P-value is the **lower-tail** probability (blue region below).

Stronger Evidence
for $H_A: p < p_0$

Weaker Evidence
for $H_A: p < p_0$



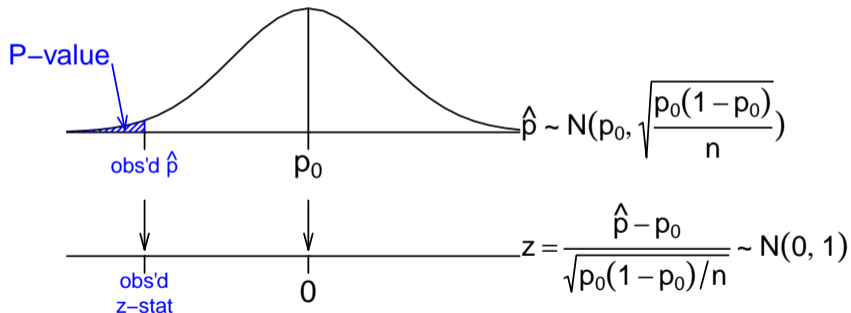
`pnorm(z)` # Finding P-value in R

P-values for Lower One-Sided Hypothesis Tests

To test $H_0: p = p_0$ vs. **lower one-sided** alternative $H_A: p < p_0$, the *P*-value is the **lower-tail** probability (blue region below).

Stronger Evidence
for $H_A: p < p_0$

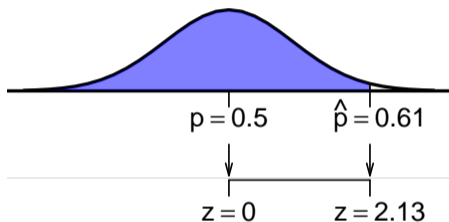
Weaker Evidence
for $H_A: p < p_0$



`pnorm(z)` # Finding P-value in R

Example (Children's Play Preference)

For $H_0: p = 0.5$ vs. $H_A: p < 0.5$, the P -value is lower-tail area $1 - 0.0166 = \boxed{0.9834}$.

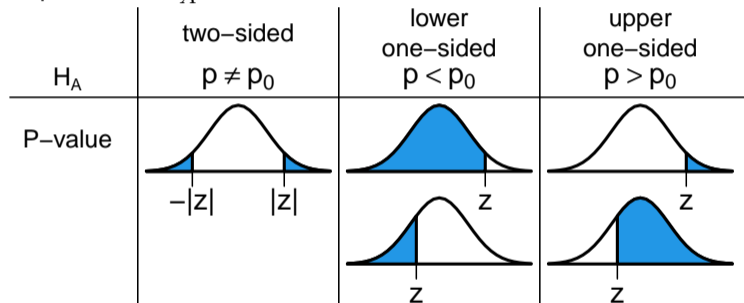


$H_A: p < 0.5$ is less plausible than $H_0: p = 0.5$, since $\hat{p} = 0.61 > 0.5$.
No reason to reject H_0 .

Recap: How to Compute One-Sided & Two Sided P -values

The z -statistic for $H_0 : p = p_0$ is $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0, 1)$.

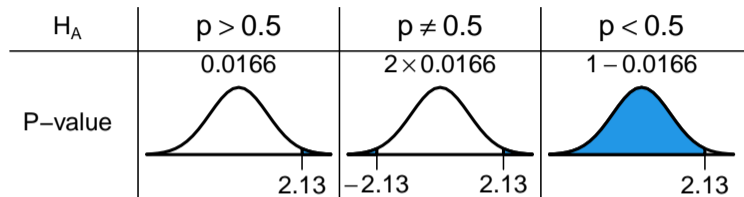
The P -value depends on H_A .



The bell-shape curve is the standard normal curve. In R:

- ▶ two-sided P -value = `2*pnorm(abs(z), lower.tail=FALSE)`
- ▶ lower one-sided P -value = `pnorm(z)`
- ▶ upper one-sided P -value = `pnorm(z, lower.tail=FALSE)`

Back to the Children's Play Preference Example



For $H_A: p > 0.5$ or $H_A: p \neq 0.5$, we **reject H_0** since P -value is **low** (under 5%)

- ▶ Significant evidence: 6YO are more likely to choose same-sex peers.
- ▶ The difference between null value (50%) and observed (61%) is beyond random variation.

For $H_A: p < 0.5$, no reason to reject $H_0: p = 0.5$ since $H_A: p < 0.5$ is less plausible than $H_0: p = 0.5$ given the $\hat{p} = 0.61 > 0.5$.

Conclusion when the P -value is Low

When the P -value is below the significance level, we say

- ▶ H_0 is rejected
- ▶ Strong evidence that more than 50% of 6YO prefer same-sex peers
- ▶ The percentage of 6YO choosing same-sex peers is **significantly** above 50%

We do NOT say:

- ▶ H_A is “accepted”
- ▶ We “fail to reject” H_A

Conclusion when the P -value is Not Low

When the P -value exceeds the significance level, we say

- ▶ We fail to reject H_0
- ▶ No strong evidence that 6YO favor same-sex peer over diff-sex peer
- ▶ The percentage of 6YO choosing same-sex peers is **NOT significantly** above 50%

We do NOT say:

- ▶ H_0 is “accepted”
- ▶ We “fail to accept” H_A
- ▶ There is strong evidence H_0 is true—which might be a Type 2 error
The chance of making a Type 2 error is not controlled, which can be quite big

More Incorrect Statements of Hypotheses

Use **significant(ly)** and **reject** ONLY when making the conclusion of a hypothesis test.
Do NOT use them in stating the hypotheses.

Incorrect hypotheses statements:

- ▶ H_0 : Percentage of 6YO choosing same-sex peers is **not significantly** over 50%
- ▶ H_A : Percentage of 6YO choosing same-sex peers is **significantly** over 50%

or

- ▶ H_0 : We **don't reject** that 6YO have no gender preference
- ▶ H_A : We **reject** that 6YO have no gender preference

Recap: Inference for a Population Proportion

Hypothesis Tests

1. Set hypotheses: $H_0: p = p_0$, vs. $H_A: p < p_0, p > p_0$, or $p \neq p_0$
2. Calculate *test statistic* and *P-value*

$$z\text{-stat} = \frac{\hat{p} - p_0}{SE}, \quad \text{where } SE = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

3. (Optional) Choose a significance level α and make a decision
 - ▶ If $P\text{-value} < \alpha$, reject H_0
 - ▶ If $P\text{-value} > \alpha$, do not reject H_0

.....

100(1 - α)% Confidence Interval for p :

$$\hat{p} \pm z_{\alpha/2} \underbrace{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}}_{SE} \quad \text{where}$$

Confidence level	90%	95%	99%
$z_{\alpha/2}$	1.645	1.960	2.576

SE

For confidence intervals, use

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

For hypothesis testing, use

$$SE = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

Why?

- ▶ By Normal approx to Binomial, $\hat{p} \sim N(p, \sqrt{p(1-p)/n})$ for large n .
- ▶ For CI, p is unknown, so estimate with \hat{p} .
- ▶ For hypothesis testing, $p = p_0$ under H_0 , so use p_0 .

Conditions for Using the Large-Sample CI and Tests for Proportions

- ▶ The observations are (nearly) i.i.d. from the population studied.
 - ▶ If SRS, the sample size is at most 10% of population.
- ▶ Sample size n should be large. Rule of thumb:
 - ▶ to use the large-sample CI:
 $n\hat{p}$ and $n(1 - \hat{p})$ need to be both ≥ 10
 - ▶ to use the large-sample test for the H_0 of $p = p_0$:
 np_0 and $n(1 - p_0)$ need to be both ≥ 10