Setting up a Test for a Population Proportion

Reed Coots
Statistics Course Developer

Hypothesis Testing

Why do we do Hypothesis Tests?

Hypothesis Testing

Why do we do Hypothesis Tests?

Could the value of the parameter be ____?

Hypothesis Testing

Why do we do Hypothesis Tests?

Could the value of the parameter be _____?

Use data to help support that claim

C.S. Mott Children's Hospital Poll

C.S. Mott Children's Hospital conducted a national poll on an issue in children's health, sleep habits. We will be looking at an example about lack of sleep in teens.

Research Question

In previous years 52% of parents believed that electronics and social media was the cause of their teenager's lack of sleep. Do more parents today believe that their teenager's lack of sleep is caused due to electronics and social media?

Research Question

In previous years 52% of parents believed that electronics and social media was the cause of their teenager's lack of sleep. Do more parents today believe that their teenager's lack of sleep is caused due to electronics and social media?

Population - Parents with a teenager (age 13-18)

Parameter of Interest - p

Research Question

In previous years 52% of parents believed that electronics and social media was the cause of their teenager's lack of sleep. Do more parents today believe that their teenager's lack of sleep is caused due to electronics and social media?

Population - Parents with a teenager (age 13-18)

Parameter of Interest - p Population Proportion

Test for a significant increase in the proportion of parents with a teenager who believe that electronics and social media is the cause for lack of sleep.

Hypotheses

Hypotheses

$H_0 : p = 0.52$

Hypotheses

$H_0 : p = 0.52$

$H_a : p \ ? \ 0.52$

IVQ

What symbol should replace the ?

Hypotheses

$H_0 : p = 0.52$

$H_a : p > 0.52$

"Significant Increase"

Hypotheses

$H_0 : p = 0.52$

$H_a : p > 0.52$

Where p is the population proportion of parents with a teenager who believe that electronics and social media is the cause of their teenager's lack of sleep

Hypotheses

$H_0 : p = 0.52$

$H_a : p > 0.52$

Where p is the population proportion of parents with a teenager who believe that electronics and social media is the cause of their teenager's lack of sleep

significance level

$\alpha = 0.05$

Survey Results

A random sample of **1018** parents with a teenager was taken and **56%** said they believe electronics and social media was the cause of their teenager's lack of sleep.

Assumptions

We need a **random sample** of parents

Assumptions

We need a **random sample** of parents

We also need a **large enough sample size** to ensure our distribution of sample proportions is normal

Assumptions

We need a **random sample** of parents

We also need a **large enough sample size** to ensure our distribution of sample proportions is normal

That is: **n · p** be at least 10

n · (1-p) be at least 10

Assumptions

We need a **random sample** of parents

We also need a **large enough sample size** to ensure our distribution of sample proportions is normal

That is: **n · p** be at least 10 → **n · p_o**

n · (1-p) be at least 10 → **n · (1-p_o)**

Checking Assumptions

Random Sample

Checking Assumptions

Random Sample ✔

In background information

Checking Assumptions

Random Sample ✅

$n \cdot p_o$

$n \cdot (1-p_o)$

$p_o = 0.52$

Checking Assumptions

Random Sample ✓

$n \cdot p_o = 1018 \cdot (0.52) = 529$ ✓

$n \cdot (1-p_o) = 1018 \cdot (1-0.52) = 489$ ✓

Segment Title Here!

Testing a One Population Proportion

Reed Coots
Statistics Course Developer

Hypotheses

$H_0 : p = 0.52$

$H_a : p > 0.52$

Best Estimate of p is $\hat{p} = 0.56$

Where p is the population proportion of parents with a teenager who believe that electronics and social media is the cause of their teenager's lack of sleep

$\alpha = 0.05$

Test Statistic

Best estimate - Hypothesized estimate

Standard error of estimate

Test Statistic

Best estimate - Hypothesized estimate

Standard error of estimate

$$\frac{\hat{p} - p_o}{s.e.}$$

Test Statistic

Best estimate - Hypothesized estimate

Standard error of estimate

$$\frac{\hat{p} - p_o}{s.e.}$$

$$s.e.(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$$

Test Statistic

Best estimate - Hypothesized estimate

Standard error of estimate

$$\frac{\hat{p} - p_o}{s.e.}$$

$$s.e.(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}} \quad \rightarrow \quad s.e.(\hat{p}) = \sqrt{\frac{p_0 \cdot (1-p_o)}{n}} \checkmark$$

Test Statistic

$$\frac{\hat{p} - p_o}{s.e.}$$

$$\text{Null } s.e.(\hat{p}) = \sqrt{\frac{p_0 \cdot (1 - p_o)}{n}}$$

Test Statistic

$$\frac{\hat{p} - p_o}{s.e.} \qquad \text{Null } s.e.(\hat{p}) = \sqrt{\frac{p_0 \cdot (1 - p_o)}{n}}$$

$$Z = \frac{0.56 - 0.52}{0.0157}$$

Z = 2.555

Test Statistic Interpretation

$$Z = 2.555$$

That means that our observed sample proportion is 2.555 null standard errors above our hypothesized population proportion

Test Statistic Distribution

- A Z test statistic is another random variable! It has a distribution.

Test Statistic Distribution

- A Z test statistic is another random variable! It has a distribution.
- The Z test statistic will always follow a N(0,1)

Test Statistic Distribution

- A Z test statistic is another random variable! It has a distribution.
- The Z test statistic will always follow a N(0,1)
- This is due to us centering and scaling our original data

Test Statistic Distribution

- A Z test statistic is another random variable! It has a distribution.
- The Z test statistic will always follow a N(0,1)
- This is due to us centering and scaling our original data

$$\frac{\hat{p} - p_o}{s.e.(\hat{p})}$$

Test Statistic Distribution

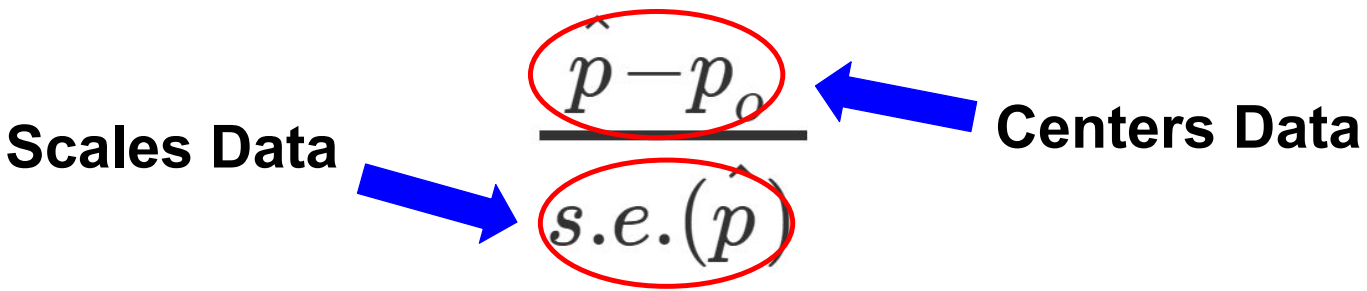
- A Z test statistic is another random variable! It has a distribution.
- The Z test statistic will always follow a N(0,1)
- This is due to us centering and scaling our original data

$$\frac{\hat{p} - p_o}{s.e.(\hat{p})}$$

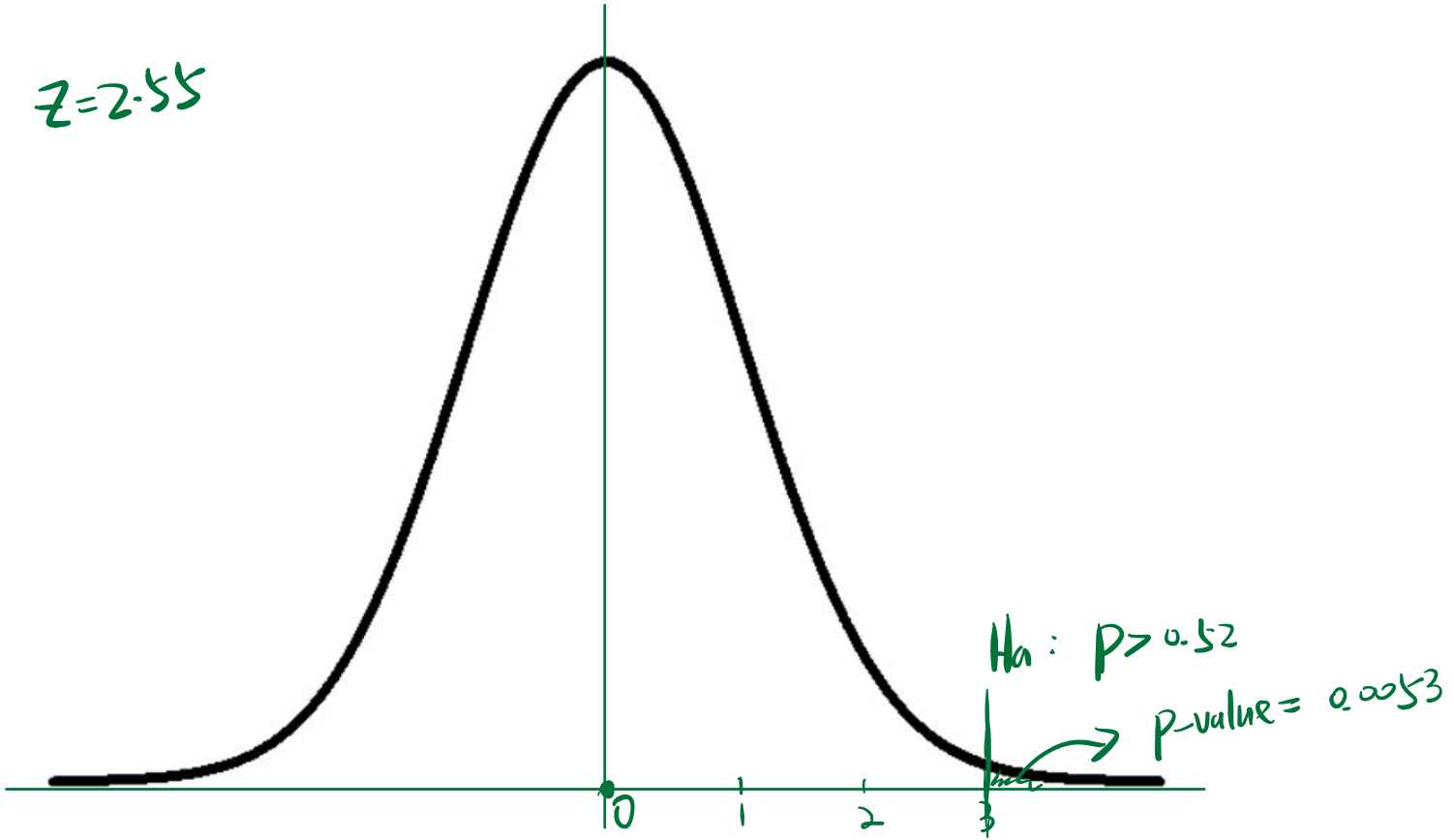
Centers Data

Test Statistic Distribution

- A Z test statistic is another random variable! It has a distribution.
- The Z test statistic will always follow a N(0,1)
- This is due to us centering and scaling our original data



The P-Value



Conclusions

p-value = 0.0053 < α = 0.05

Conclusions

p-value = 0.0053 < α = 0.05

Reject the null hypothesis (H_0: p = 0.52)

Conclusions

p-value = 0.0053 < α = 0.05

Reject the null hypothesis (H_0: p = 0.52)

There is sufficient evidence to conclude that the population proportion of parents with a teenager who believe that electronics and social media is the cause for lack of sleep is greater than 52%.

Summary

- 4 main steps to a hypothesis test
  - Stating hypothesis & select significance level (α)
  - Checking assumptions
  - Calculating a test statistic and getting a p-value from the test statistic
  - Drawing a conclusions from the p-value

Summary

- 4 main steps to a hypothesis test
  - Stating hypothesis & select significance level (α)
  - Checking assumptions
  - Calculating a test statistic and getting a p-value from the test statistic
  - Drawing a conclusions from the p-value
- The Z test statistic distribution is N(0,1)