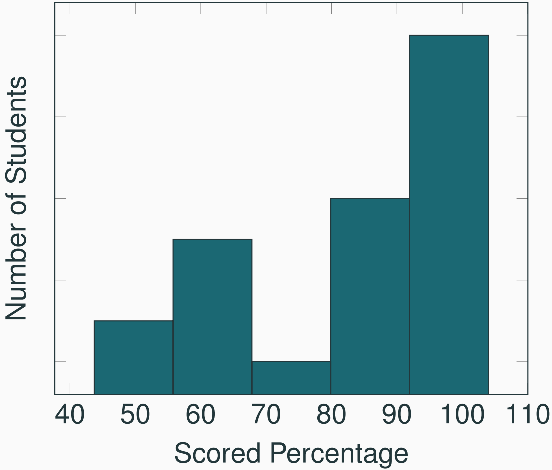
Announcements

- Confidence Intervals Lab due Monday after break
- I’m not assigning other HW over break
- Working on a possible EC opportunity relating to March Madness
  - Hopefully up tonight?
- Read Ch 3.2 and 3.3 for Friday
- Polling: `rembold-class.ddns.net`

Test Results



- High: 104%
- Mean: 79.8%
- Std: 18.5%
- Median: 83.4%

Quick Test Coverage

- I'm going to be quickly going over the test
- Idea is to make sure you understand what might have gone wrong
- If you think a mistake has been made or I misadded points, please swing by the office and let me know

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

A) All 1000 get the drug

B) 500 get the drug, 500 don't

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

A) All 1000 get the drug

B) 500 get the drug, 500 don't

Results from the GSS

The GSS (General Social Survey) asks the same question, below is the distribution of responses from the 2010 survey:

All 1000 get the drug	99
500 get the drug, 500 don’t	571
Total	670

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don’t”? What are the parameter of interest and the point estimate?

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don’t”? What are the parameter of interest and the point estimate?

- ***Parameter of interest:*** Proportion of all Americans who have good intuition about experimental design.

$$p : \text{a population proportion}$$

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don’t”? What are the parameter of interest and the point estimate?

- ***Parameter of interest:*** Proportion of all Americans who have good intuition about experimental design.

$$p \text{ : a population proportion}$$

- ***Point estimate:*** Proportion of sampled Americans who have good intuition about experimental design.

$$\hat{p} \text{ : a sample proportion}$$

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer “500 get the drug 500 don’t”?

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer “500 get the drug 500 don’t”?

- We can answer this research question using a confidence interval, which we know is always of the form

$$\textit{point estimate} \pm z^* \times SE$$

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer "500 get the drug 500 don't"?

- We can answer this research question using a confidence interval, which we know is always of the form

$$\textit{point estimate} \pm z^* \times SE$$

- But what is the standard error of our point estimate?

$$SE_{\hat{p}} = ?$$

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer "500 get the drug 500 don't"?

- We can answer this research question using a confidence interval, which we know is always of the form

$$point\ estimate \pm z^{\star} \times SE$$

- But what is the standard error of our point estimate?

$$SE_{\hat{p}} = ?$$

Standard error of a sample proportion

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

Sample proportions are also nearly normally distributed

Central limit theorem for proportions

Sample proportions will be nearly normally distributed with mean equal to the population mean, p, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$.

$$\hat{p} \sim N\left(mean = p, \quad SE = \sqrt{\frac{p\,(1-p)}{n}} \right)$$

- But of course this is true only under certain conditions...
  - independent observations
  - at least 10 successes and 10 failures

Note: If p is unknown (most cases), we use $\hat{p}$ in the calculation of the standard error.

Back to experimental design...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design?

Back to experimental design...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design?

Given: $n = 670, \hat{p} = 0.85$. First check conditions.

Back to experimental design...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design?

Given: $n = 670, \hat{p} = 0.85$. First check conditions.

1. ***Independence***: The sample is random, and $670 < 10\%$ of all Americans, therefore we can assume that one respondent's response is independent of another.

Back to experimental design...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design?

Given: $n = 670, \hat{p} = 0.85$. First check conditions.

1. ***Independence***: The sample is random, and $670 < 10\%$ of all Americans, therefore we can assume that one respondent's response is independent of another.
2. ***Success-failure***: 571 people answered correctly (successes) and 99 answered incorrectly (failures), both are greater than 10.

Understanding Check

We are given that $n = 670$, $\hat{p} = 0.85$, we also just learned that the standard error of the sample proportion is $SE = \sqrt{\frac{p(1-p)}{n}}$. Which of the below is the correct calculation of the 95% confidence interval?

A) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

B) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

C) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

D) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

Understanding Check

We are given that $n = 670, \hat{p} = 0.85$, we also just learned that the standard error of the sample proportion is $SE = \sqrt{\frac{p(1-p)}{n}}$. Which of the below is the correct calculation of the 95% confidence interval?

A) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}} \rightarrow (0.82, 0.88)$

B) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

C) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

D) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$