Advanced Statistics
DS2003 (BDS-4A)
Lecture 07

Instructor: Dr. Syed Mohammad Irteza

Assistant Professor, Department of Computer Science, FAST

08 March, 2022

Previous Lecture

- SD and SE
- Hypothesis Testing
- Quick Review of Conditional Probability
- Testing hypotheses using confidence intervals
- Decision errors

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

- This is why we prefer small values of α -- increasing α increases the Type 1 error rate.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	Type 1 Error
	H_A true	Type 2 Error	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

We (almost) never know if H_0 or H_A is true, but we need to consider all possibilities.

Facebook Interest Categories

- The same survey asked the *850 respondents* how comfortable they are with Facebook creating a list of categories for them.
- *41% of the respondents* said they are comfortable.
- Do these data provide convincing evidence that the proportion of American Facebook users that are comfortable with Facebook creating a list of interest categories for them is *different than 50%*?

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Facebook Interest Categories

- The same survey asked the *850 respondents* how comfortable they are with Facebook creating a list of categories for them.
- *41% of the respondents* said they are comfortable.
- Do these data provide convincing evidence that the proportion of American Facebook users that are comfortable with Facebook creating a list of interest categories for them is *different than 50%*?

- **Setting the hypotheses**
  The *parameter of interest* is the proportion of <u>all</u> American Facebook users who are comfortable with Facebook creating categories of interests for them.

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Facebook Interest Categories

- The same survey asked the *850 respondents* how comfortable they are with Facebook creating a list of categories for them.
- *41% of the respondents* said they are comfortable.
- Do these data provide convincing evidence that the proportion of American Facebook users that are comfortable with Facebook creating a list of interest categories for them is *different than 50%*?

- **Setting the hypotheses**
  The *parameter of interest* is the proportion of <u>all</u> American Facebook users who are comfortable with Facebook creating categories of interests for them.

→ There may be two explanations why our sample proportion is lower than 0.50 (minority).

- *The true population proportion is different than 0.50*.
- *The true population mean is 0.50, and the difference between the true population proportion and the sample proportion is simply due to natural sampling variability*.

Facebook Interest Categories

- The same survey asked the *850 respondents* how comfortable they are with Facebook creating a list of categories for them.
- *41% of the respondents* said they are comfortable.
- Do these data provide convincing evidence that the proportion of American Facebook users that are comfortable with Facebook creating a list of interest categories for them is *different than 50%*?

Setting the hypotheses

We start with the assumption that *50% of American Facebook users are comfortable* with Facebook creating categories of interests for them

H_0: $p = 0.50$

We test the claim that the proportion of American Facebook users who are comfortable with Facebook creating categories of interests for them *is different than 50%*.

H_A: $p \neq 0.50$

Test Statistic

- In order to evaluate if the *observed sample mean is unusual for the hypothesized sampling distribution*, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.50, SE = \sqrt{\frac{0.50 \times 0.50}{850}}\right)$$

$$Z = \frac{0.41 - 0.50}{0.0171} = -5.26$$

- The sample proportion is 5.26 standard errors away from the hypothesized value. Is this considered unusually low? That is, is the result ***statistically significant***?
- *Yes, and we can quantify how unusual it is using a p-value.*

p-Values

- We then use this test statistic to calculate the *p-value*, the *probability of observing data at least as favorable to the alternative hypothesis* as our current data set, *if the null hypothesis were true*.

- If *the p-value is low* (lower than the *significance level, α*, which is usually *5%*) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject* H_0.

- If *the p-value is high* (*higher than α*) we say that it is likely to observe the data even if the null hypothesis were true, and hence *do not reject* H_0.

Facebook Interest Categories - p-value

- *p-value*: probability of observing data at least as favorable to H_A as our current data set (*a sample proportion lower than 0.41*), if in fact *H_0 were true* (*the true population proportion was 0.50*).

$$P(\hat{p} < 0.41 \text{ or } \hat{p} > 0.59 \mid p = 0.50) = P(|Z| > 5.26) < 0.0001$$

Facebook Interest Categories – Making a decision

p-value < 0.0001

- If *50%* of all American Facebook users are comfortable with Facebook creating these interest categories, there is *less than a 0.01% chance* of *observing a random sample of 850 American Facebook users where 41% or fewer or 59% of higher feel comfortable with it*.
- This is a *pretty low probability* for us to think that the observed sample proportion, or something more extreme, is *likely to happen simply by chance*.
- Since p-value is low (lower than 5%) we reject H_0.
- *The data provide convincing evidence that the proportion of American Facebook users who are comfortable with Facebook creating a list of interest categories for them is different than 50%.*
- The difference between the null value of 0.50 and observed sample proportion of 0.41 is *not due to chance or sampling variability*.

Number of college applications

A similar survey asked how many colleges students applied to, and 206 students responded to this question. This sample yielded an average of 9.7 college applications with a standard deviation of 7. College Board website states that counselors recommend students apply to roughly 8 colleges. Do these data provide convincing evidence that the average number of colleges all Duke students apply to is higher than recommended?

http://www.collegeboard.com/student/apply/the-application/151680.html

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by <u>all</u> Duke students.

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by all Duke students.
- There may be two explanations why our sample mean is higher than the recommended 8 schools.
  - The true population mean is different.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by all Duke students.
- There may be two explanations why our sample mean is higher than the recommended 8 schools.
  - The true population mean is different.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability
- We start with the assumption the average number of colleges Duke students apply to is 8 (as recommended)

$$H_0 : \boldsymbol{\mu} = 8$$

Setting the hypotheses

- The *parameter of interest* is the average number of schools applied to by all Duke students.
- There may be two explanations why our sample mean is higher than the recommended 8 schools.
    - The true population mean is different.
    - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability
- We start with the assumption the average number of colleges Duke students apply to is 8 (as recommended)

$$H_0 : \boldsymbol{\mu} = 8$$

- We test the claim that the average number of colleges Duke students apply to is greater than 8

$$H_A : \boldsymbol{\mu} > 8$$

Number of college applications - conditions

Which of the following is *not* a condition that needs to be met to proceed with this hypothesis test?

a) Students in the sample should be independent of each other with respect to how many colleges they applied to.
b) Sampling should have been done randomly.
c) The sample size should be less than 10% of the population of all Duke students.
d) There should be at least 10 successes and 10 failures in the sample.
e) The distribution of the number of colleges students apply to should not be extremely skewed.

Number of college applications - conditions

Which of the following is <u>*not*</u> a condition that needs to be met to proceed with this hypothesis test?

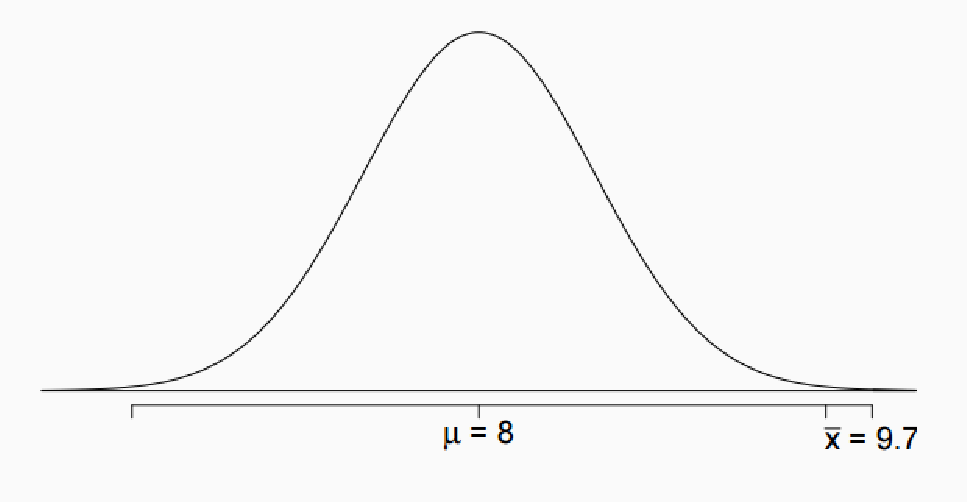
a) Students in the sample should be independent of each other with respect to how many colleges they applied to.
b) Sampling should have been done randomly.
c) The sample size should be less than 10% of the population of all Duke students.
d) *There should be at least 10 successes and 10 failures in the sample.*
e) The distribution of the number of colleges students apply to should not be extremely skewed.

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

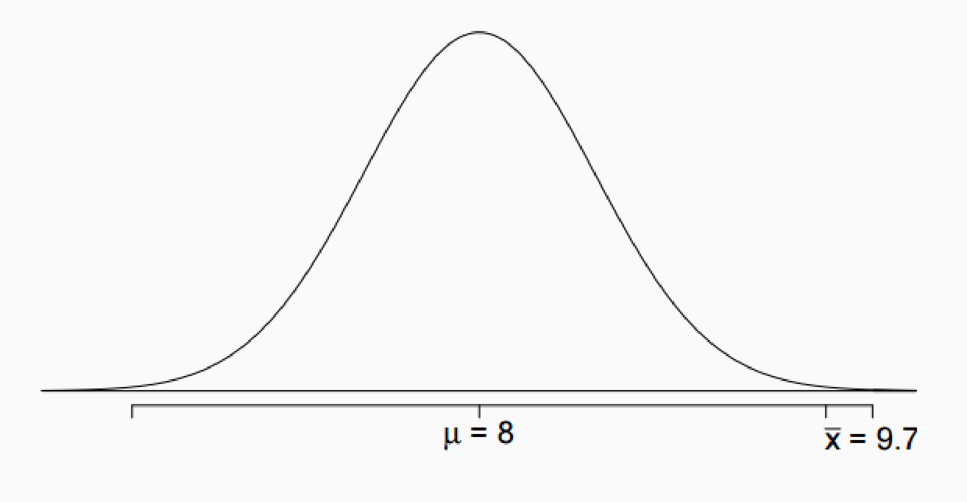
Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



Test Statistic

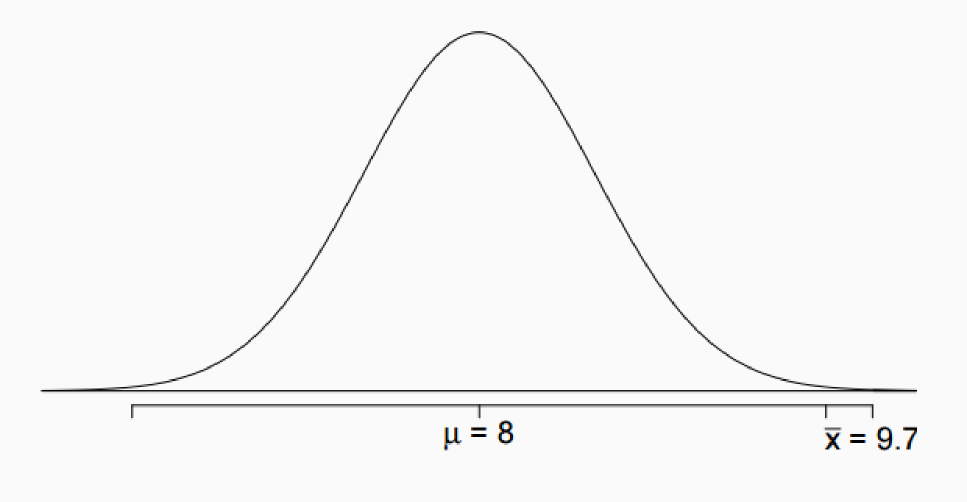
In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

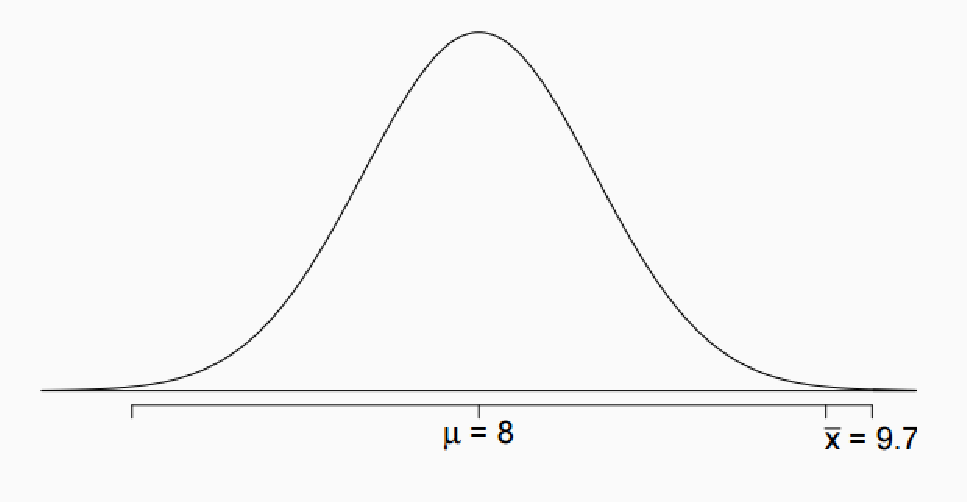


$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



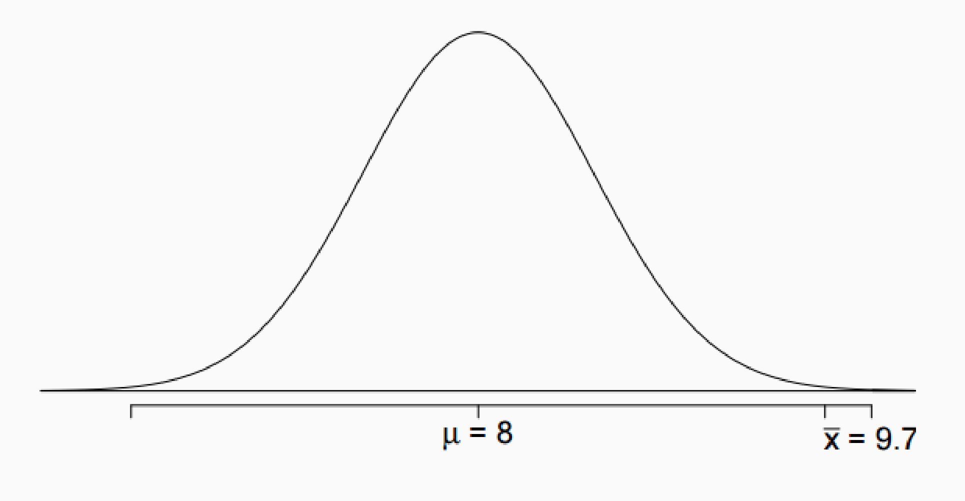
$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

The sample mean is 3.4 standard errors away from the hypothesized value. Is this considered unusually high? That is, is the result *statistically significant*?

Test Statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

The sample mean is 3.4 standard errors away from the hypothesized value. Is this considered unusually high? That is, is the result *statistically significant*?

Yes, and we can quantify how unusual it is using a p-value.

p-values

- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.

p-values

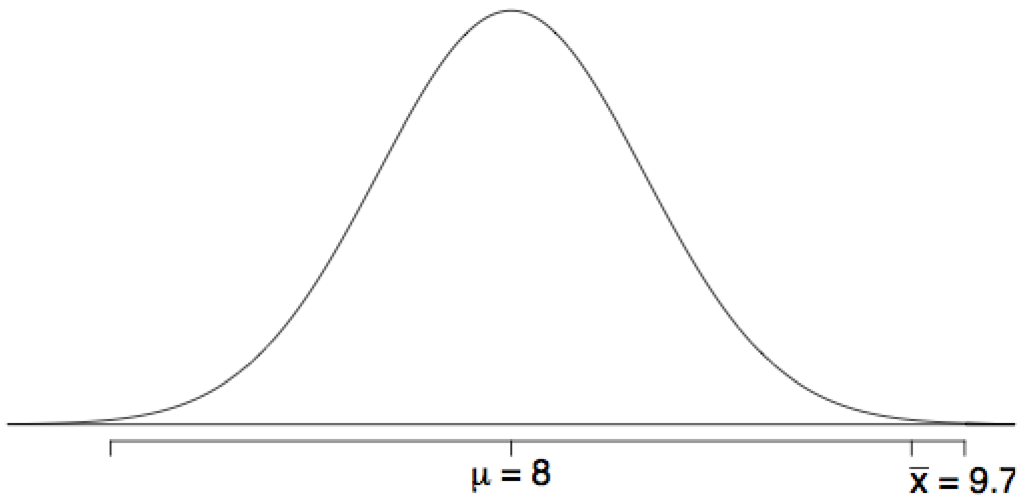
- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.
- If the p-value is *low* (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject* H_0.

p-values

- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.
- If the p-value is *low* (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject H_0*.
- If the p-value is *high* (higher than α) we say that it is likely to observe the data even if the null hypothesis were true, and hence *do not reject H_0*.

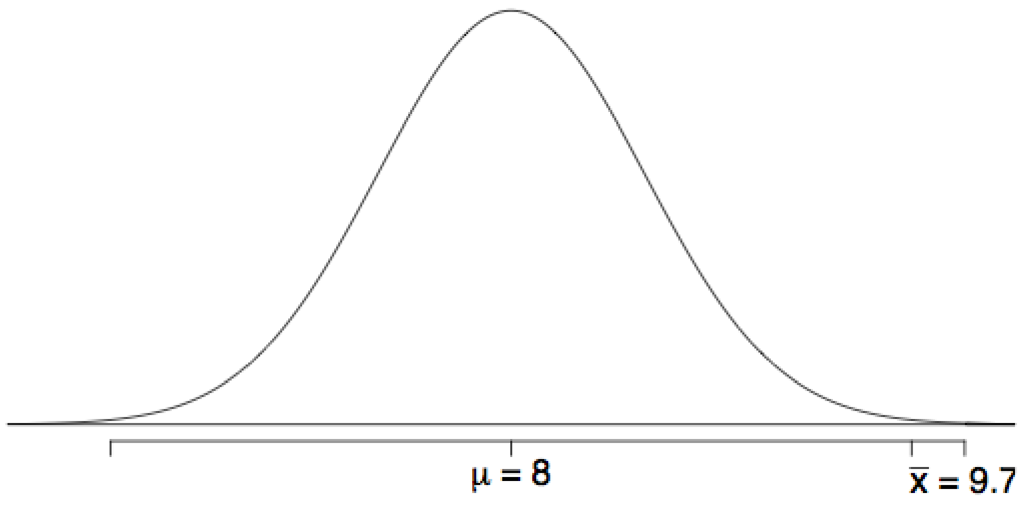
Number of college applications - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).



Number of college applications - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).



$$P(\bar{x} > 9.7 \mid \mu = 8) = P(Z > 3.4) = 0.0003$$

Number of college applications - Making a decision

- p-value = 0.0003

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject* H_0.

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject H_0*.
- The data provide convincing evidence that Duke students apply to more than 8 schools on average.

Number of college applications - Making a decision

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject H_0*.
- The data provide convincing evidence that Duke students apply to more than 8 schools on average.
- The difference between the null value of 8 schools and observed sample mean of 9.7 schools is *not due to chance* or sampling variability.

Recap: Hypothesis testing framework

1. Set the hypotheses.
2. Check assumptions and conditions.
3. Calculate a *test statistic* and a p-value.
4. Make a decision, and interpret it in context of the research question.

Recap: Hypothesis testing for a population mean

1. Set the hypotheses

    - H_0: μ = null value
    - H_A: μ < or > or ≠ null value

2. Calculate the point estimate

3. Check assumptions and conditions

    - Independence: random sample/assignment, 10% condition when sampling without replacement
    - Normality: nearly normal population or $n \geq 30$, no extreme skew -- or use the t distribution (Ch 5)

4. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{x} - \mu}{SE}, \text{ where } SE = \frac{s}{\sqrt{n}}$$

5. Make a decision, and interpret it in context

- If p-value < α, reject H_0, data provide evidence for H_A
- If p-value > α, do not reject H_0, data do not provide evidence for H_A

Sources

- openintro.org/os (Chapter 5)
- Type I and II Errors (utexas.edu)