STAT 252

Week 1: Review

2024-04-01

Day Two

Purpose of Statistics: Make inferences about a population from a sample

---

Statistical Question: A question that can be answered by collecting data that varies

For example...

How old am I?

(not statistical)

How old are Cal Poly students on average?

(statistical)

How tall is Jamear?

(not statistical)

How tall is the average 12-year-old in the US?

(statistical)

Let's Practice

Statistical or Not?

1. **How many hours per week do students spend studying for exams?**
2. **How many siblings do you have?**
3. **What was the temperature at 12pm today?**

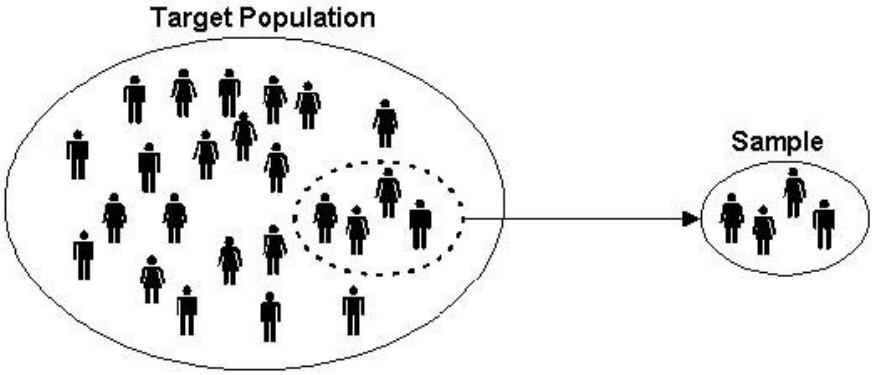
Classifying Units of Study

Population: The entire group of study from which a sample is drawn

Sample: Part of the population from which data is gathered

Observational Unit: A single person, place, or thing from the sample

Sample Size: Total number of observational units in a sample (denoted n)



For example...

Population, Sample, Observational Unit: Cal Poly Students

Population: All students at Cal Poly

Sample: Students currently enrolled in a STAT 252 course

Observational Unit: A Cal Poly student

Population, Sample, Observational Unit: Sodas in college dining halls

Population: All sodas served in college dining halls across the US

Sample: Sodas served at Cal Poly's dining halls

Observational Unit: A soda

Let's Practice

State the population, sample, and observational unit.

Statistical Question: According to current Cal Poly students, what dining hall venue is the best?

Population:

Sample:

Observational Unit:

Statistical Question: What proportion of Cal Poly students hand-write versus type their notes?

Population:

Sample:

Observational Unit:

Statistical Question: How tall are buildings in SLO on average?

Population:

Sample:

Observational Unit:

Variables

Variable: A characteristic that can be measured and assume different values

Categorical: Variable that takes on three or more category designations

For example...

1. **What is the most common hair color at Cal Poly?**

- Variable of Interest: Hair Color
- Variable Type: Categorical

2. **What is the least liked ice cream flavor among Cal Poly students?**

- Variable of Interest: Ice Cream Flavor
- Variable Type: Categorical

Categorical Binary: Variable that takes on two category designations

For example...

1. **What proportion of US citizens are married?**

- Variable of Interest: Marital Status
- Variable Type: Categorical Binary

2. **What percentage of Cal Poly students have a job?**

- Variable of Interest: Employment Status
- Variable Type: Categorical Binary

Quantitative: Variable that takes on a continuous range of numerical values

For example...

1. **What is the temperature on average in a Cal Poly dorm?**

- Variable of Interest: Temperature
- Variable Type: Quantitative

2. **How heavy is a student's backpack on average?**

- Variable of Interest: Weight
- Variable Type: Quantitative

Let's Practice

State the variable of interest and its type (categorical, categorical binary, quantitative).

1. **How many students at Cal Poly participate in sports?**

- Variable of Interest:
- Variable Type:

2. **What mode of transportation (bus, Uber, personal vehicle) is most commonly used by Cal Poly students to go downtown?**

- Variable of Interest:
- Variable Type:

3. **What is the average monthly revenue for a small business in SLO?**

- Variable of Interest:
- Variable Type:

Variables (cont.)

Explanatory Variable: Variable that explains variation in the response variable

Response Variable: Variable of interest that the outcome of a study measures

For example...

1. Does going to office hours affect a student's performance on a test?

- Explanatory Variable: Whether a student went to office hours
- Response Variable: Score on test

2. Does eating before or after a workout allow you to squat more?

- Explanatory Variable: Whether you ate before or after your workout
- Response Variable: Squat weight

Let's Practice

State the explanatory and response variable for each scenario.

1. How does the type of fertilizer used affect plant growth?

- Explanatory Variable:
- Response Variable:

2. Does the type of exercise regimen (aerobic or strength training) impact weight loss?

- Explanatory Variable
- Response Variable:

Day Three

Purpose of Statistics: Make inferences about a population from a sample

---

Parameter v. Statistics

_____: Numbers that summarize data for an entire population

_____: Numbers that summarize data from a sample

Why are samples necessary in statistics?

Can a parameter ever be known?

For the purpose of this class, we will primarily focus on 3 measures, each with a unique symbol for their parameter and statistic. Each measure also has a certain variable type associated with it:

Measure	Parameter	Statistic	Variable Type
Mean			
Proportion			
Standard Deviation			

Let's Practice

Fill in the missing elements.

1. **Measure: Sample Mean**

- Parameter or Statistic? _____
- Symbol? _____

2. **Measure: _____**

- Parameter or Statistic? _____
- Symbol? s

3. **Measure: _____ Proportion**

- Parameter or Statistic? Parameter
- Symbol?

Types of Study Designs

Observational Study: A study in which the researcher collects data through observations without any manipulation of participants

For example...

1. A researcher observes study habits and time management behaviors of college students in a library setting to understand factors influencing academic performance.
2. A psychologist conducts observations in a college cafeteria to analyze eating habits and food choices among students, investigating factors such as social influences and dietary preferences.

Experimental Study: A study in which the researcher actively manipulates participants and splits them into a control group and treatment group(s)

Control Group: Participants who do not receive the experimental treatment

For example...

1. A researcher examines the effects of different study environments on academic performance by randomly assigning college students to study either in a quiet library or a bustling café and comparing their subsequent test scores.
2. An education researcher investigates the impact of personalized feedback on essay writing skills by randomly assigning students to receive either generic or tailored feedback on their writing assignments and assessing improvements in writing quality.

Let's Practice

State whether the situation best describes an observational or experimental study.

1. A researcher examines the relationship between caffeine consumption and productivity levels in employees at a workplace by reviewing employee performance reports and coffee logs.

- Experimental or Observational Study?

2. A group of scientists studies the relationship between temperature variations and crop yields in agricultural regions by exposing different test plants to various conditions in the lab.

- Experimental or Observational Study?

3. A researcher investigates the impact of music genres on mood regulation in listeners by collecting survey data from Spotify users.

- Experimental or Observational Study?

Randomness

Random Selection: Process by which participants are chosen for a study by chance

For example...

1. A professor selects students from a class roster to participate in a research study on sleep patterns by assigning each student a unique identification number and using a random number generator to select participants.
2. A university divides its student body into distinct groups based on academic programs. Then, it randomly picks students from each group for its survey on student satisfaction.

Random Assignment: Process by which the treatment is given to the observational unit by chance

For example...

1. In a clinical trial testing the effectiveness of a new medication for migraine headaches, participants are randomly assigned to either the treatment group, where they receive the medication, or the control group, where they receive a placebo.
2. In an educational study examining the impact of different teaching methods on student learning outcomes, classrooms are randomly assigned to one of several experimental conditions, such as traditional lecture-based instruction or active learning strategies.

What is the purpose of each aspect of randomness? (hint: confounding variables & bias)

Preview of Statistical Inference

Statistical Inference: Statistical technique that draws a conclusion about a population parameter based on a sample statistic

Two Types:

_____: Statistical inference that assesses the plausibility of a particular claim about the parameter

For example...

1. A boss claims that it takes its employees 10 minutes, on average, to complete a particular task. To verify this claim, a random sample of 30 employees is selected, and their task completion times are recorded.

_____: Statistical inference that estimates the value of a parameter with a range of plausible values

For example...

1. A manufacturing company has recently introduced a new product to the market and wants estimate its population mean defect rate. To achieve this, the company randomly selects a sample of 200 units from their production line.

4 Techniques:

1. Hypothesis Test for Population Mean μ
2. Hypothesis Test for Population Proportion p
3. Confidence Interval for Population Mean μ
4. Confidence Interval for Population Proportion p

Video 1: One-Population Hypothesis Test (Mean, Proportion)

Purpose of Statistics: Make inferences about a population from a sample

---

One-Population Mean Hypothesis Test

Assumptions: Data is random; observations are independent

Hypotheses for Two-Tailed Test:

- Null Hypothesis:
  - Words: The population mean of [context] is equal to μ_o
  - Symbols: H_0: $\mu = \mu_o$
- Alternative Hypothesis:
  - Words: The population mean of [context] is not equal to μ_o
  - Symbols: H_1: $\mu \neq \mu_o$

Test Statistic: $TS = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

Note:

- If TS < -3.5 OR TS > 3.5, we reject H_0
- If p-value < 0.1, we reject H_0 (calculator needed to find p-value)

For Example...

1. A tech company asserts that the average battery life of their new smartphone model is different from the industry standard of 20 hours. A sample of 100 smartphones is tested, revealing an average battery life of 19.5 hours with a standard deviation of 2.5 hours.

Statistical Question?

Parameter of Interest?

Null Hypothesis?

Alternative Hypothesis?

Test Statistic?

Conclusion?

One-Population Proportion Hypothesis Test

Assumptions: Data is random; $np \geq 10$ & $n(1 - p) \geq 10$

Hypotheses for Two-Tailed Test:

- Null Hypothesis:
  - Words: The population proportion of [context] is equal to p_o
  - Symbols: H_0: $p = p_o$
- Alternative Hypothesis:
  - Words: The population proportion of [context] is not equal to p_o
  - Symbols: H_1: $p \neq p_o$

Test Statistic: $TS = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$

Note:

- If TS < -3.5 OR TS > 3.5, we reject H_0
- If p-value < 0.1, we reject H_0 (calculator needed to find p-value)

For example...

1. A company proposes that 30% of its customers are satisfied with its customer service. A random sample of 200 customers is surveyed, and 100 express satisfaction.

Statistical Question?

Parameter of Interest?

Null Hypothesis?

Alternative Hypothesis?

Test Statistic?

Conclusion?

Now, in R! (Dr. Williams)

Output

Visualization

Analysis

Let's Practice (Complete on your own and come to Learning Hours if questions!)

1. A fitness tracker company claims that the average number of steps taken per day by users wearing their device is different from the industry standard of 10,000 steps. A sample of 150 users is selected, and their average daily step count is found to be 9,600 steps with a standard deviation of 1,200 steps.

Statistical Question?

Parameter of Interest?

Null Hypothesis?

Alternative Hypothesis?

Test Statistic?

Conclusion?

2. An online retailer asserts that 40% of its customers make repeat purchases within one month of their initial purchase. A sample of 500 customers is surveyed, and 220 of them make repeat purchases within the specified time frame.

Statistical Question?

Parameter of Interest?

Null Hypothesis?

Alternative Hypothesis?

Test Statistic?

Conclusion?

Video 2: Confidence Interval (Mean, Proportion)

Purpose of Statistics: Make inferences about a population from a sample

---

One-Population Mean Confidence Interval

- Critical Value: Number of standard errors from the parameter needed to achieve a certain level of confidence (CV)
- Margin of Error: ME = CV * SE
- Upper Bound: UB = PE + ME
- Lower Bound: LB = PE - ME

Assumptions: Data is random; observations are independent

Point Estimate: $\bar{x}$

Standard Error: $\frac{s}{\sqrt{n}}$

For example...

1. An online retailer wants to estimate the average amount spent by customers in a week. A random sample of 100 customers shows an average spending of \$150 with a standard deviation of \$30. Calculate a 95% confidence interval for the population mean. (The critical value for a 95% confidence interval is 1.96)

Statistical Question?

Parameter of Interest?

Margin of Error?

Upper Bound?

Lower Bound?

Interpretation of Confidence Interval?

One-Population Proportion Confidence Interval

- Critical Value: Number of standard errors from the parameter needed to achieve a certain level of confidence (CV)
- Margin of Error: ME = CV * SE
- Upper Bound: UB = PE + ME
- Lower Bound: LB = PE - ME

Assumptions: Data is random; $np \geq 10$ & $n(1 - p) \geq 10$

Point Estimate: $\hat{p}$

Standard Error: $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

For example...

1. A company wants to estimate the proportion of customers who prefer their new product packaging. A survey of 200 customers shows that 120 of them prefer the new packaging. Calculate a 90% confidence interval for the population proportion. (The critical value for a 90% confidence interval is 1.6449)

Statistical Question?

Parameter of Interest?

Margin of Error?

Upper Bound?

Lower Bound?

Interpretation of Confidence Interval?

Now, in R! (Dr. Williams)

Output

Visualization

Analysis

Let's Practice (Complete on your own and come to Learning Hours if questions!)

1. Estimate the average number of hours college students spend studying per week. A random sample of 50 students shows an average study time of 15 hours with a standard deviation of 3 hours. Calculate a 95% confidence interval for the population mean. (The critical value for a 95% confidence interval is 1.96)

Statistical Question?

Parameter of Interest?

Margin of Error?

Upper Bound?

Lower Bound?

Interpretation of Confidence Interval?

2. Estimate the proportion of college students who prefer fast food over home-cooked meals. A survey of 250 college students shows that 180 of them prefer fast food. Calculate a 99% confidence interval for the population proportion. (The critical value for a 99% confidence interval is 2.575)

Statistical Question?

Parameter of Interest?

Margin of Error?

Upper Bound?

Lower Bound?

Interpretation of Confidence Interval?

Video 3: Types of Errors (I, II), Types of Distributions, and Confidence Interval Manipulations

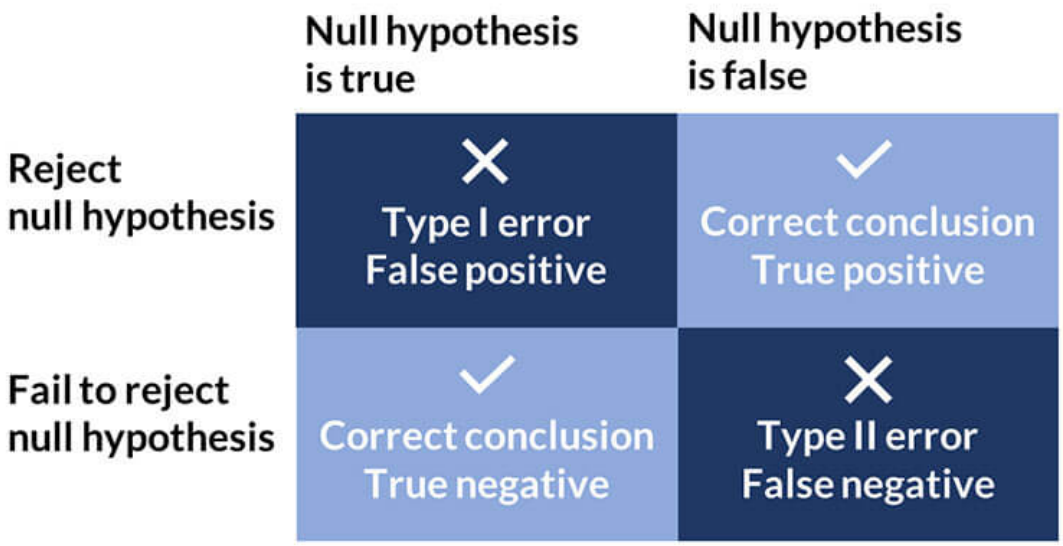
Purpose of Statistics: Make inferences about a population from a sample

---

Hypothesis Test Errors

Type I Error: Rejecting a null hypothesis that is true

Type II Error: Failing to reject a null hypothesis that is false



For example...

1. In a clinical trial, if the company wrongly concludes that a new drug reduces blood pressure when it doesn't, it's a Type I error. This error leads to wasted resources, unnecessary side effects, and misleading conclusions.

2. If a quality control test incorrectly accepts a batch of products as meeting the quality standards when, in fact, they do not meet the standards, this would be a Type II error. As a result, defective products may enter the market, leading to potential customer dissatisfaction, product recalls, and damage to the company's reputation.

Let's Practice

State whether the conclusion could entail a Type I or Type II Error.

1. Based on the results of the entrance exam, an admission office concludes that a student is likely to succeed in their program. However, it's later discovered that the student struggles academically and eventually drops out of the program.

Type of Error:

2. After analyzing the data from a semester-long study, a professor concludes that the new teaching method significantly improves student performance. However, upon further investigation, it's discovered that the conclusion was based on a statistical error, and there is no significant improvement in student performance with the new teaching method.

Type of Error:

Types of Distributions

If $n < 30$, s is given, or nothing is stated about the distribution, use:

- t-distribution

Otherwise, use:

- z-distribution

Confidence Interval Manipulations

Factors that affect Confidence Intervals:

- Sample size n
- Standard Deviation s
- Standard Error
- Level of Confidence

Let's Practice

State whether the confidence interval shrinks or widens from a change to one of the factors.

1. What happens to the interval as n increases?
2. What happens to the interval as level of confidence decreases?
3. What happens to the interval as standard error increases?
4. What happens to the interval as s decreases?
