FUTURELAB+

BIOMED

Crowdsourcing Innovations in Biotechnology

Statistically Significant Biomarkers

Developed in partnership with:

Discovery Education and Ignited

In this Lesson Plan:

Print the **Teacher Section** →

Print the **Student Section** →

This document is separated into two sections, For Teachers [T] *and* Student Resources [S], *which can be printed independently.*

Select the appropriate printer icon above to print either section in its entirety.

Follow the tips below in the Range field of your Print panel to print single pages or page ranges:

Single Pages (use a comma): T3, T6

Page Range (use a hyphen): T3–T6

Cover Image
The image shows a cell that has been divided by mitosis.

BIOMED / CROWDSOURCING INNOVATIONS IN BIOTECHNOLOGY

Statistically Significant Biomarkers

DRIVING QUESTION

How do we know whether the trends in biomarker data are significant enough to be considered unique for a disease?

OVERVIEW

We live in a time marked by easy access to information. Throughout an average day, we are bombarded by statements and claims pouring in through our cell phone, radio, social media, television, and of course, the internet. However, not all of that information is factual. Some of it is blatantly misleading with some sources using pseudoscience and skewed statistics to push specific agendas. Deceptive graphs and misrepresented scientific findings can be more difficult to identify than simple inaccurate statements.

In this lesson, students will explore the importance of representative and honest data collection as part of our ethical responsibility. They will also learn how common statistical tests can be used to find relationships in data. The lesson culminates in students assuming the role of a biostatistician—applying this new knowledge to a previously-selected scientific article.

ACTIVITY DURATION

Two class sessions (45 minutes each)

ESSENTIAL QUESTIONS

How do we know if one thing causes another?

How can data make us better scientists?

How can statistics support the sciences?

OBJECTIVES

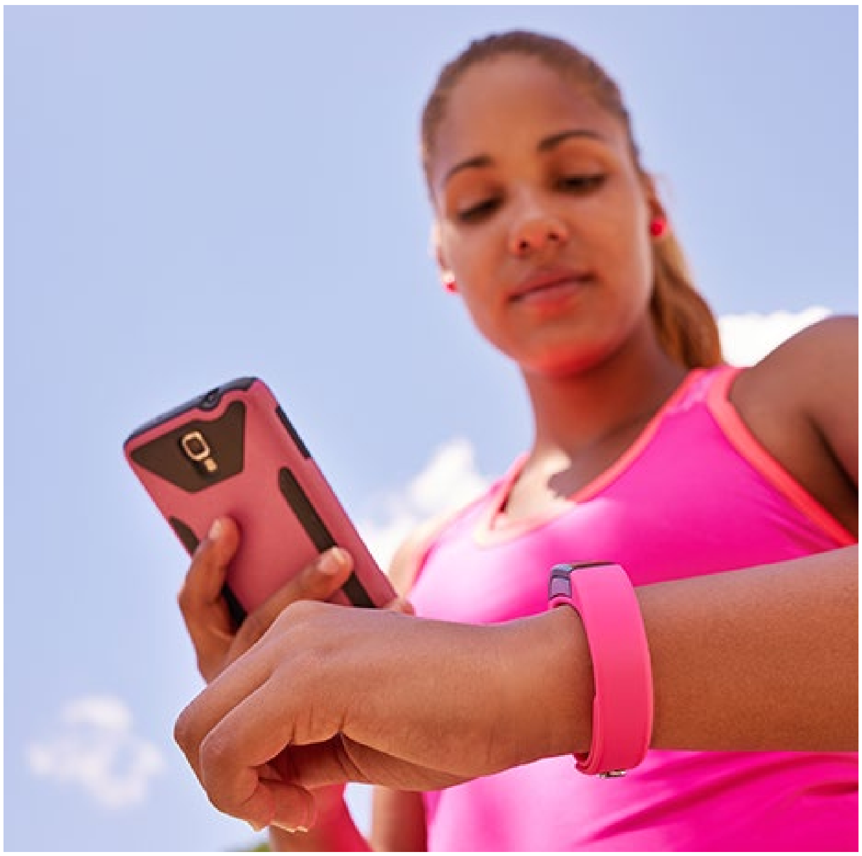
Students will be able to:

Determine which statistical tests may help identify relationships present in data.

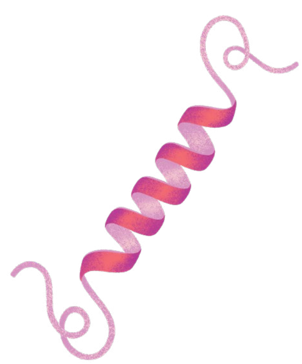
Explain the importance of representative and balanced data collection.

BACKGROUND INFORMATION

Understanding statistical analysis is essential to understanding how wearable medical technology works. It is impossible to talk about data without also examining how that data is interpreted. Predominantly biology-based lessons will teach students about signs of disease and important measurable biomarkers, and technological coursework will introduce students to the engineering aspects of biological technology, but mathematics is the bridge that links the two.



Materials
Introduction to t-tests
Puzzle Experiment
p-value Capture Sheet
Scientific Honesty Paragraph Rubric
Analyzing a Scientific Paper
Analyzing a Scientific Paper Rubric
Design Journal



Pedagogical Framing

Instructional materials are designed to meet national education and industry standards to focus on in-demand skills needed across the full product development life cycle—from molecule to medicine—which will also expose students and educators to the breadth of education and career pathways across biotechnology.

Through this collection, educators are equipped with strategies to engage students from diverse racial, ethnic, and cultural groups, providing them with quality, equitable, and liberating educational experiences that validate and affirm student identity.

Units are designed to be problem-based and focus on workforce skill development to empower students with the knowledge and tools to be the change in reducing health disparities in communities.

SOCIAL-EMOTIONAL LEARNING

Students engage in cooperative learning and use gamification strategies, which aid in building community. Interacting with peers in a collaborative manner meets many social-emotional learning goals and builds a supportive classroom, all while teaching necessary soft skills such as active listening, leadership skills, and compromise.

CULTURALLY AND LINGUISTICALLY RESPONSIVE INSTRUCTION

This lesson includes many cultural and linguistically responsive strategies, including the use of individually chosen puzzles to learn about testing hypotheses, and the “Raise a Righteous Hand” technique for simple exercise of choice. There are opportunities for affirming and validating the contributions of culturally and linguistically diverse people in relevant research and other scientific endeavors.

ADVANCING INCLUSIVE RESEARCH

Students explore the importance of representative and honest data collection. They will examine the historical relationship between the medical community and its treatment of people of color and women. This helps open dialogue around mistrust in the medical community and what we can do to rebuild trust to support populations currently understudied in genomic and clinical research.

COMPUTATIONAL THINKING PRACTICES

In this lesson, students learn about the risks of assuming that patterns in data signify the presence of a relationship. To illustrate this concept, students examine case studies in which faulty analyses of data led to outlandish conclusions.

CONNECTION TO THE PRODUCT LIFE CYCLE

Students will collect data and run a t-test on their results. They will then relate how this type of testing may be beneficial in the **manufacturing** phase of the product life cycle. Afterwards, they will explore the importance of regulation through the story of Yoshihiro Sato, whose numerous fraudulent studies led many other researchers to false conclusions and was discovered through the medical community’s own self-regulation. Students will also discuss how government regulation is working to improve the diversity of clinical trials.

Have you ever wondered…

How can we find connections in data?

Statistical tests can be applied to data in an attempt to demonstrate reasonable proof of a cause and response relationship.

How do scientists determine if a medication or treatment is effective?

Biostatisticians help researchers design experiments that would help confirm or refute their hypotheses. These mathematicians then examine data from clinical trials by using statistical analysis which can help determine whether any changes noted in patients are related to the treatment.

MAKE CONNECTIONS!

How does this connect to the larger unit storyline?

To comprehend how biomarkers are potentially useful for the management and tracking of diseases, students need to have an understanding of the statistical tests that can be used to show relationships within data sets. Wearable medical devices output a lot of data, but without interpretation, it is just meaningless numbers and cannot be considered useful information. At the end of this lesson, students will be able to explain how statisticians attempt to prove causality, as well as apply their understanding of common statistical tests to a scientific article of their choice.

How does this connect to careers?

Biostatisticians play important roles, both in designing experiments and clinical studies, as well as in interpreting their findings. These professionals assist researchers in minimizing unwanted variables that could have an effect on their results. Afterward, using mathematical analysis, biostatisticians examine whether a study's findings indicate a relationship.

Bioethicists study the ethics of advancing science. Just because we can do something, should we? Is this new practice equitable? Who is benefiting and who is being left behind? Many topics in science can be controversial, so bioethicists help us navigate the larger concepts of science and attempt to ensure its impact is for the greater good.

How does this connect to our world?

The search for better health is not a new one, and now through wearable devices, individuals are better able to monitor different aspects of their well-being. There are currently dozens of different devices being used everyday, ranging from insulin pumps and portable cardiac monitors, to more consumer-driven products used to track levels of activity, O_2 saturation, and heart rate. With the influx of new data and new machines aimed at keeping the population thriving, this market is only expected to grow. The world of biotechnology requires analytical minds that can help set parameters to determine what is normal and what is concerning, individuals who can relate biomarker data to real-life implications, and dedicated people who hope to make the world a better place through mathematics.

Day 1

LEARNING OUTCOMES

Students will be able to:

Define correlation and causation.

Create media that explains the difference between correlation and causation.

Explain the purpose of a t-test.

Conduct their own experiment to collect data.

Interpret their experiment's results for statistical significance.

COMPUTATIONAL THINKING IN ACTION

In this lesson, students practice the computational thinking strategy of finding patterns to examine real-world case studies and recognize that patterns in data do not always signify relationships.

Procedure

Teacher Note > *Today, you will start by looking at the difference between correlation and causation. Despite the well-known phrase, "correlation does not imply causation," it is still a leap that is often seen in the media, and can have dire effects on the world of health sciences. As you progress through this unit, students will learn how biostatisticians use mathematics in an attempt to prove causation.*

Whole Group (10 minutes)

1 Pose this scenario to the class, "Online, a scientist states that while studying fertility in a poorer country in southeast Asia, she noted that owning a microwave led to having fewer children."

Ask the class to consider "Does owning a microwave decrease fertility?"

2 Give students a few minutes to write down their thoughts, and then have them move into groups of two to four to discuss their answers.

3 As a class consider: "Would more information make it easier to decide? What types of information would help them come to a decision?"

Teacher Note > *This scenario is related to a real study in which the actual correlation was that people who could afford a microwave could also afford birth control.*

4 Have the class watch a video on how correlation does not equate to causation, *The danger of mixing up causality and correlation: Ionica Smeets at TEDxDelft*. Make note of the information about the speaker and that this is an international talk.

Teacher Note > *Ask students to make note of at least two studies that mistakenly linked two factors as causation when there was just a correlation. This is an exceedingly important conversation for high schoolers to have and a new analytical lens they can apply.*

Continues next page >

Day 1

Continued

Procedure

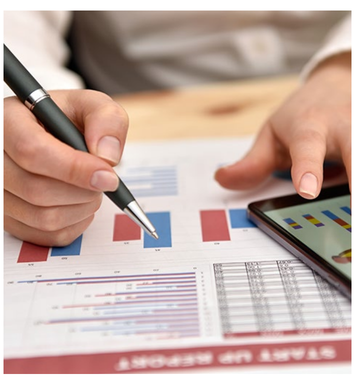
5	Guide the class to work together through discussion to come up with definitions of both **correlation** and **causation** with guidance. Ideas from students can be typed into a slide and projected or simply written on the board.

Teacher Note > *Correlation is when the value of two things seem to move in relation to each other, i.e. as ice cream sales increase so do drownings. Correlation can also be when those values move in opposite directions (negative correlation), such as when the temperatures decrease, sales of winter coats increase. Causation is simply when one thing causes another; it is much more difficult to prove. Just because two things seem to happen in succession does not mean one thing has caused the other. (*View video ahead of time to ensure it is appropriate for all students in your class)*

INDUSTRY AND CAREER CONNECTION

To conduct this experiment with success, students need to demonstrate organizational skills and precise data collection skills, which are key skills for a biostatistician. Students will need to keep clear records and manage their time wisely.

Small Group (10 minutes)	
1	Let students work in pairs to read the *Introduction to t-tests Capture Sheet*.
2	After reading the introduction, review the example with the class. Ask students to collaborate with their partners to identify the null and alternative hypotheses.
3	Allow each student pair time to match and compare answers with other pairs.



Small Group (25 minutes)	
1	Organize students into groups of four or five, and assure every group has a device.
2	Groups will challenge other groups in completing online puzzles. You can access free puzzles on *this site*. Allow each group to select one puzzle to use for the length of the experiment. Each group should have a different puzzle. Do not set a time limit.

Continues next page >

Day 1

Continued

Procedure

3 Prompt groups to come up with their own hypotheses to test, including the independent and dependent variables. (i.e., We hypothesize that juniors will solve the puzzle faster than sophomores. We hypothesize that people from the left side of the class will solve the puzzle faster than the right side of the class.)

Teacher Note > *Ask students to stick to benign elements, such as location in class, grade classification, if someone is wearing tennis shoes, etc. Students should not use gender, religion, ethnicity, or race in these examples.*

4 Let students alternate between testing classmates on how quickly they can solve the puzzle and taking classmates' puzzles. In groups of four, they should be able to essentially swap devices and take them simultaneously.

5 Ask students to maintain clear records of testing times using the *Puzzle Experiment Capture Sheet*.

6 Direct students to this website *GraphPad QuickCalcs: t-test calculator* where an online calculator feature will run a t-test to determine whether there was any significant difference between their two sets of data.

Teacher Note > *Students will leave the page on its default settings (50 rows and an unpaired t-test), use the bottom left data section to add headers, enter their testing data, and then select "Calculate Now."*

7 After completing the online calculation, have students interpret their t-test results at the bottom of their experimental data sheet. Allow them to review and revisit the results after students have studied p-value, and determine if the difference between the two groups is statistically significant or not.

8 During the last five minutes of class, lead a discussion as to how t-tests may be beneficial in the manufacturing portion of the product life cycle in which we are seeking to ensure the quality of the product. Use the *Stand and Share* discussion protocol to provide for some simple movement. When participants have a solution, answer, or comment, they stand. When all who want to answer have stood, the facilitator asks each for his or her input. Once a student has given it, that student can sit down. Time required: two to three minutes to solve the issue, plus a minute or less per person in the discussion.

Day 2

LEARNING OUTCOMES

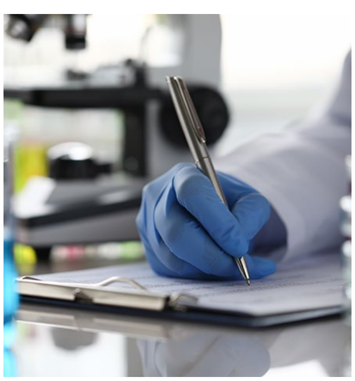
Students will be able to:

Explain the purpose of both confidence intervals and p-values.

Construct a paragraph explaining the importance of honest scientific results.

CULTURALLY AND LINGUISTICALLY RESPONSIVE INSTRUCTION

The read aloud and follow along provides scaffolding for language development and reading fluency for all students, but is especially beneficial for CLD students in a small group setting.



Procedure

Whole Group (5 minutes)

1 Open class using the participation protocol *Raise a Righteous Hand* to answer one of the questions: "What does it mean to have confidence in something?" or "Is having confidence more important than being right?" Take two to three answers from students.

2 Let students know that they are going to explore the final two statistical tests, the confidence interval and p-values (p-value to be defined later in lesson). Share with them that confidence intervals tell us how certain we are that a data point will fall within a selected range. The higher the confidence interval (common values are 90%, 95%, and 99%), the more certain we are that the data point lies within that range.

3 Show the video *Understanding Confidence Intervals: Statistics Help* Pause video at 0:41, 1:07, and 2:49 to allow students to process the information or jot down the key ideas. Make note of the *information about the speaker* and that this is an international channel.

Small Group (10 minutes)

1 In groups of two to three, students will take turns reading aloud (while others follow along silently) the article *"Election polls are 95% confident but only 60% accurate, Berkeley Haas study finds."*

2 Ask students to respond to the following "Explain how confidence intervals are used in election polling" on sticky notes, and place the sticky notes that the group decides are most relevant in a parking lot on the board.

Whole Group (5 minutes)

Project video *What Is A p-value?—Clearly Explained*. Guide students to write down key points on the *p-value Capture Sheet* with the end goal of defining "p-value."

Continues next page >

Day 2

Continued

Procedure

Small Group (15 minutes)

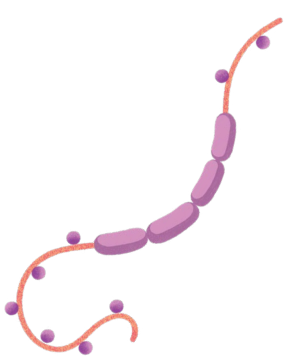
Organize students into groups of two to three, and ask students to take turns reading aloud (while others follow along silently) the article on Yoshihiro Sato and his fraudulent studies. Sato's false data had far-reaching consequences in the medical research community. It led others to spend years and resources investigating false leads, as well as deeply skewing large meta-analyses performed around the world. While reading, students will continue their notes on the same capture sheet from the last activity.

For more information, see: *Researcher at the center of an epic fraud remains an enigma to those who exposed him*.

Individual Work (10 minutes)

1. Once they complete the reading, allow students to independently construct a well-detailed paragraph on the importance of honesty in scientific research and how statistics can reveal inconsistencies in published works. Ask "How did oversight within the scientific community lead to this exposure? How could it have done better?"

2. If time allows, students may read their paragraphs to the class and discuss their findings.



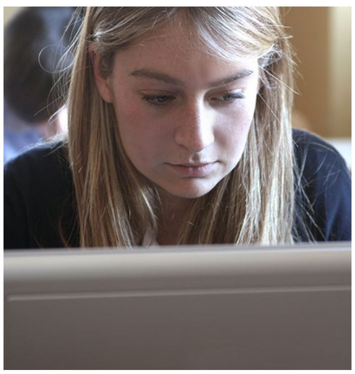
Extension

LEARNING OUTCOMES

Students will be able to:

Analyze a scientific article's statistical claims.

Apply their knowledge of statistics to real world research.



CULTURALLY AND LINGUISTICALLY RESPONSIVE INSTRUCTION

This activity provides for personal voice since choice of topic has been allowed. Emphasis of educational content over creative design is important for modeling standard academic work, but allowing for some personal choice in design can validate and affirm their cultural ideals. This allows the whole assignment to become a bridge between the academic and the cultural.

Procedure

Individual Work (45 minutes)	
1	Guide students to return to the article that they previously selected for their Lesson 7 final project and utilize what they have learned about statistics during the week to investigate its statistical claims. Ask them to use the *Analyzing a Scientific Paper Capture Sheet* to collate their thoughts.
2	Guide students to create a presentation of their choosing to detail their findings. Students can use presentation boards, videos, etc. Students should use the capture sheet and *Analyzing a Scientific Paper Rubric* as a guide to complete their projects. Remind them that the educational content in the presentation must be approved before any decorative design elements can be added. Call out work time in 10- or 15-minute intervals to help with time management.

National Standards

Next Generation Science Standards	**ETS1.B: Developing Possible Solutions** When evaluating solutions, it is important to take into account a range of constraints, including cost, safety, reliability, and aesthetics, and to consider social, cultural, and environmental impacts.
	Science and Engineering Practices Using Mathematics and Computational Thinking Use mathematical, computational, and/or algorithmic representations of phenomena or design solutions to describe and/or support claims and/or explanations. Planning and Carrying Out Investigations Make directional hypotheses that specify what happens to a dependent variable when an independent variable is manipulated. Obtaining, Evaluating, and Communicating Information Critically read scientific literature adapted for classroom use to determine the central ideas or conclusions and/or to obtain scientific and/or technical information to summarize complex evidence, concepts, processes, or information presented in a text by paraphrasing them in simpler but still accurate terms.
	Crosscutting Concepts Scale, Proportion, and Quantity Algebraic thinking is used to examine scientific data and predict the effect of a change in one variable on another (e.g., linear growth vs. exponential growth).
Career and Technical Education (CTE)	**A2.4** Understand the critical need for ethical policies and procedures for institutions engaged in biotechnology research and product development.
	A5.1: Use the Internet and World Wide Web to collect and share scientific information.
	A6.1: Apply knowledge of symbols, algebra, and statistics to graphical data presentation.

Introduction to t-tests

ANSWER KEY Do not share with students

Directions
In the following scenarios, write down the Null Hypothesis and the Alternative Hypothesis.

Daniel thinks wearing his lucky socks helps his baseball team win games, but Sophia disagrees. They decide to record how many wins his team has both when he is wearing his lucky socks and when he is not.

Null Hypothesis	Alternative Hypothesis
There is no significant difference between the number of wins when Daniel does and does not wear his lucky socks.	There is a significant difference between the number of wins when Daniel does and does not wear his lucky socks.

Mr. Compher's 9th grade Biology class hypothesizes that the pond behind the school will have more frogs this year because there have been heavy rains.

Null Hypothesis	Alternative Hypothesis
There is no significant difference between the number of frogs in the pond last year and this year.	There is a significant difference between the number of frogs in the pond last year and this year.

Continues next page >

Introduction to t-tests

ANSWER KEY | **Do not share with students**

Continued

Ezra is training for the school's annual 5K race. He decides to improve his time. He is going to run four times a week instead of two.

Null Hypothesis	Alternative Hypothesis
There is no significant difference between Ezra's race time when he runs four times a week instead of two.	There is a significant difference between Ezra's race time when he runs four times a week instead of two.

Smalltown School District wants to encourage more students to take science electives, so it decides to invest in new science lab equipment.

Null Hypothesis	Alternative Hypothesis
There is no significant difference between the number of students taking science electives before and after purchasing new equipment.	There is a significant difference between the number of students taking science electives before and after purchasing the new equipment.

Maria wants to get an A in mathematics this year and decides that forming a study group will help her achieve her goal.

Null Hypothesis	Alternative Hypothesis
There is no significant difference between her grades when studying with and without a study group.	There is a significant difference between her grades when studying with and without a study group.

p-value Capture Sheet

ANSWER KEY | **Do not share with students**

Directions

Respond to the questions below after viewing the video and reading the article.

Video Response

1. Write down any statements that you believe might be beneficial to defining "p-value".

p-value is an abbreviation for probability value.

The p-value is a number between 0 and 1.

The lower the number is, the stronger the evidence is for something to be true.

The p-value is the fraction of a chance something occurred through random noise, or coincidence.

2. Following the video, how would you define a p-value?

p-value is the probability of obtaining the observed difference in the outcome measure, given that no difference exists between two populations.

Tide of Lies Response

1. How does this article describe p-values?

p-value measures baseline characteristics and provides the ability to confirm that the groups are equally distributed so that any results of the study are not attributed to chance. In comparing treatment and control groups, scientists calculate p-values to measure the similarities between the two groups, which are randomly selected.

2. How did Sato's reported p-values indicate fraud?

Sato's reports consistently produced strong results but the likelihood of him being able to enroll and evaluate the number of patients is not possible. Additionally, it was found that out of more than 500 variables, more than half were above 0.8 which just should not happen.

FUTURELAB+

Introduction to t-tests

Today, we are going to learn our second statistical test, the t-test. When biostatisticians analyze data from an experiment or clinical trial, the first step is to determine if the proposed treatment or intervention had any effect at all. A t-test examines two sets of data to see if there is a "significant difference" between them. When working in medical science, this might mean something such as comparing blood pressure data from a group given a new medication to blood pressure data from a group given a placebo. If there is a "significant difference," this would signal any changes seen in the experimental group, such as improved blood pressure values, could be attributed to the new medication.

When talking about t-tests, we must also learn two key vocabulary terms: *Null Hypothesis* and *Alternative Hypothesis.*

Null Hypothesis: There is no significant difference between the two groups.

Alternative Hypothesis: There is a significant difference between the two groups.

Directions

In the following scenarios, write down the Null Hypothesis and the Alternative Hypothesis.

Daniel thinks wearing his lucky socks helps his baseball team win games, but Sophia disagrees. They decide to record how many wins his team has both when he is wearing his lucky socks and when he is not.

Null Hypothesis	Alternative Hypothesis

Mr. Compher's 9th grade Biology class hypothesizes that the pond behind the school will have more frogs this year since there have been heavy rains.

Null Hypothesis	Alternative Hypothesis

Continues next page >

FUTURELAB+

Introduction to t-tests

Continued

Ezra is training for the school's annual 5K race. He decides to improve his time. He is going to run four times a week instead of two.

Null Hypothesis	Alternative Hypothesis

Smalltown School District wants to encourage more students to take science electives, so it decides to invest in new science lab equipment.

Null Hypothesis	Alternative Hypothesis

Maria wants to get an A in mathematics this year and decides that forming a study group will help her achieve her goal.

Null Hypothesis	Alternative Hypothesis

FUTURELAB+

Puzzle Experiment Capture Sheet

Directions
With your group, select one puzzle to use for the length of the experiment. Come up with your own hypotheses to test. Maintain clear records of testing times as you alternate with other groups to test your experiment.

We hypothesize that ______________________________ (Group 1)

will complete the puzzle faster than ______________________________ (Group 2)

Group 1 data:

Group 2 data:

How would you interpret the findings from your t-test?

FUTURELAB+

p-value Capture Sheet

Directions
Respond to the questions below after viewing the video and reading the article.

Video Response

1. Write down any statements that you believe might be beneficial to defining "p-value".

2. Following the video, how would you define a p-value?

Tide of Lies Response

1. How does this article describe p-values?

2. How did Sato's reported p-values indicate fraud?

FUTURELAB+

Scientific Honesty Paragraph Rubric

Score	3	2	1
Overall	Paragraph is well edited without any obvious errors.	Paragraph has minor spelling and grammatical errors.	Paragraph has multiple spelling and grammatical errors.
Connection to Statistics	Student provides a good explanation for how statistics can reveal scientific dishonesty.	Student makes some attempt at demonstrating how statistics can reveal scientific dishonesty.	Student does not mention statistics.
Explanation on Importance	Thorough explanation of why scientific honesty is important is provided and supported with details and examples.	Some effort made to explain why scientific honesty is important, but limited details or examples provided.	No statement about why scientific honesty is important.
Final Score			

FUTURELAB+

Analyzing a Scientific Paper

Directions

Returning to your scientific paper selected in Lesson 7, today we are going to examine how the author utilized statistical analysis and what those tests say about their results. Find five instances from your paper in which conclusions were statistically significant, including the p-value.

1. p-value

1a. Does your paper report their p-value? If so, what is it?

1b. What is the purpose of a p-value?

1c. If listed, what does the p-value say about the data collected in this study?

2. Statistical Significance of p-value

2a. Was the p-value statistically significant?

2b. How do we know it was significant?

2c. What are the implications of this finding?

Continues next page >

FUTURELAB+

Analyzing a Scientific Paper

Continued

3. Statistically Significant

3a. What finding was statistically significant?

3b. How do we know it was significant?

3c. What are the implications of this finding?

4. Statistically Significant

4a. What finding was statistically significant?

4b. How do we know it was significant?

4c. What are the implications of this finding?

Continues next page >

FUTURELAB+

Analyzing a Scientific Paper

Continued

5. Statistically Significant

5a. What finding was statistically significant?

5b. How do we know it was significant?

5c. What are the implications of this finding?

6. Statistically Significant

6a. What finding was statistically significant?

6b. How do we know it was significant?

6c. What are the implications of this finding?

Next Step

Now, create a report that clearly communicates how you've applied your knowledge of statistics to your selected paper. It is your choice how you decide to present your findings, but be sure to include information on all five statistically significant findings.

FUTURELAB+

Analyzing a Scientific Paper Rubric

Score	3	2	1
Overall Appearance	Report appears professional with minimal to no errors present.	Report includes some spelling and grammatical errors, with some effort made toward appearance.	Appearance of report is sloppy and rushed.
Interpretation of p-value	Example is present and the explanation of its significance demonstrates an understanding of statistics.	Example is included but explanation does not demonstrate a clear understanding of statistics.	Example and explanation are missing.
Second Statistically Significant Finding	Example is present and the explanation of its significance demonstrates an understanding of statistics.	Example is included but explanation does not demonstrate a clear understanding of statistics.	Example and explanation are missing.
Third Statistically Significant Finding	Example is present and the explanation of its significance demonstrates an understanding of statistics.	Example is included but explanation does not demonstrate a clear understanding of statistics.	Example and explanation are missing.
Fourth Statistically Significant Finding	Example is present and the explanation of its significance demonstrates an understanding of statistics.	Example is included but explanation does not demonstrate a clear understanding of statistics.	Example and explanation are missing.
Fifth Statistically Significant Finding	Example is present and the explanation of its significance demonstrates an understanding of statistics.	Example is included but explanation does not demonstrate a clear understanding of statistics.	Example and explanation are missing.
Final Score			