Applying Statistics in Biotechnology Research Projects

Author's Name: Rocky Ng **Coach Name:** I-Heng McComb
Host Organization: Genentech **ETP Type:** Classroom
Subject/Grade: Biotechnology 10-12th

Abstract

In this project, students will utilize industry-relevant data analysis skills to support their research project. Students will learn different statistical tools such as mean, standard deviation, and Student's t-test to support their findings. First through readings, practices, and experiments, students are able to model and learn from other scientists and assigned lab experiments. As a class, we will analyze examples and use statistical language to explain phenomena.

As a summative assessment, students will design their own research question and implement the research. In pairs, they will present their research project in a scientific poster to faculty and community members. In these scientific posters, one requirement is to support their findings through statistics. Furthermore, students are expected to use appropriate language to communicate the statistics and research findings to different audiences ranging from high school students to professional scientists.

Focal Content & Supporting Practices

CTE Health and Medical Technology Standard A6.1 Apply knowledge of symbols, algebra, and statistics to graphical data presentation.

21st Century Skills and Applications

Creativity and Innovation
In the second-year biotechnology course, students will perform various experiments relating to molecular genetics and recombinant DNA technology. After choosing a previously completed experiment, students will design an experimental variable to investigate. Subsequently, students will formally present their findings in a scientific research poster in a symposium. Furthermore, the students will need to use scientific literature to support their findings in addition to synthesize possible future directions.

Problem Solving
Throughout the research designing process, students will often encounter issues in laboratory and post-laboratory. They will consult peers, faculty, and scientific literature to address the various issues. Oftentime, they will need to plan and optimize their experimental protocol. Lastly, biotechnology students will need to use statistical evidence to support their hypothesis.

Measurable Objective(s)

Students will be able to:
1. Identify the control and experimental variables in their experiment
2. Calculate the standard deviation in their experiment
3. Apply and explain the p-value to their statistical tests such as the Student's t-test

Formative Assessment(s)

Students will be peer reviewing other posters and presenting their own poster to peers. Other formative assessments include practice problems and previous experiments on identifying control variables, calculating p-values, and learning other statistical variables.

Summative Assessment(s)

Students in pairs will design a scientific poster relating to their chosen research question. They will implement statistics in order to generate various graphical outputs. Finally, they will present their scientific posters in a poster symposium in front of faculty and community members. Students will be graded according to a rubric.

Fellowship Description (300-500 words)

In the biopharmaceutical industry like Genentech, large therapeutic molecules such as proteins can be produced through genetic engineering. Gene of interest is inserted into host cells such as *Escherichia coli* and Chinese Hamster Ovary cells in order to express the desired product. Products are then purified and analyzed before commercialization. Throughout the biomanufacturing process, stringent documentation, quality controls, and various statistical tools are used to adhere to federal regulations before delivering the drug to the patient.

After protein purification, analytical testings are needed to ensure therapeutic drugs host cell proteins (HCP) are correctly identified and minimized. HCP impurities can become immunogenic and toxic to the patients. Correctly identifying HCP and improving manufacturing processes will adhere to guidelines set by the regulatory agencies.

One method to identify HCP impurities is liquid chromatography with tandem mass spectrometry (LC-MSMS). Samples containing the desired product and HCP are first separated through high-performance liquid chromatography and then ionized through mass spectrometry. Fragments of different peptides are then collected through the mass spectrometry analyzer. Subsequently, *in silico* analysis is performed by comparing peptide sequences to known universal protein standards (UPS).

With the increasing use of LC-MSMS in HCP detection, a reliable analytical method comparing the differences between samples is needed. Principal component analysis (PCA) is one of the methods to detect data variance across samples by reducing data dimensionality and defining key variables that affect data variability[1]. This fellowship in Genentech Analytical Operations enables the fellow to use *in silico* methods to identify key variables (eg. molecular weight, hydrophobicity, and pI) affecting the peptide mapping. The fellow will employ data analysis and statistical tools to determine which variable should be used when analyzing LC-MSMS data.

The fellowship provides various skills that directly relate to my class. For example, laboratory documentation, reading scientific literature, and data analysis are skills that are explicitly taught in my biotechnology classroom. Furthermore, these technical skills can be used as a lens to teach my students the 21st century soft skills which are essential for students to be successful in any career setting.

[1] Shatat SM, Eltanany BM, Mohamed AA, Al-Ghobashy MA, Fathalla FA, Abbas SS. Coupling of on-column trypsin digestion-peptide mapping and principal component analysis for stability and biosimilarity assessment of recombinant human growth hormone. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2018;1072:105-115. doi:10.1016/j.jchromb.2017.11.007

Fellowship Connection to School/Classroom

South San Francisco is one of the biggest biopharma clusters in the United States, and we need more locally grown workforce for the South San Francisco biotechnology industry. As a biotechnology teacher in South San Francisco High School, I engage community partners like Genentech and Skyline College to provide a three-year biotechnology pathway called Science Garage, where my students learn biotechnology with state-of-the-art equipment.

During the second-year course (Biotechnology 3-4), biotechnology students at South San Francisco High School will be able to develop their research questions. In the past, students had research questions like optimizing DNA extraction protocols, determining bacterial growth curves, and many other experiments. Students are expected to create a scientific poster presentation describing their research. One improvement to student research design is to implement statistics in their data analysis. In the upcoming academic year, students in the second and third-year biotechnology will include statistical evidence to support their research. Particularly, students will be able to employ the Student's t-test in their research design.

The fellowship has a direct connection to my classroom. First, by immersing myself as a research associate in the Analytical Operations, I plan to introduce data analytics to my student's learning as I mentioned above. Secondly, laboratory documentation can be implemented in a paper-bound lab notebook or electronically. I plan to explore an electronic lab notebook system during the 2020-2021 school year. Lastly, I learned there are many career options in the biotechnology field besides scientific researchers. For example, a biotechnology company like Genentech needs science-related talents ranging from statisticians, medical writers, analysts, animal care technicians to operational-related talents like facilities maintenance mechanics, business development leads, event planners, and even lawyers. I plan to expose the different types of viable careers in a biotechnology company to my students. Overall, I plan to bring the workforce skills needed in the biotechnology industry and share them back to my students.

Host Organization Engagement

Genentech Futurelab is a hyper-local scaled science education initiative.[2] With their support, the Science Garage program provides students in South San Francisco Unified School District exposure to biotechnology. Starting in 9th grade, students take biology with a biotechnology unit embedded. If students choose to take biotechnology, they can dual-enroll in our three-year pathway where they receive Skyline College credits as well as high school credits. Furthermore, Genentech provides an annual field trip to all Science Garage students. Finally, Genentech employees are able to volunteer in the Science Garage program in order to engage students and lend a hand in the lab.

Instructional Plan

Day 0: Graphing Introduction (to be implemented before or as a separate unit)		
Objective: Students will be able to use Student's t-tests to describe their experiment.		**Materials**
45 min	As an asynchronous activity, students learn how to calculate mean, median, standard deviation. Students will also learn how to make a scatter plot and column graph in this tutorial. When completed, students submit their Google Sheet to the assigned learning management	Google Sheet tutorial

[2] https://www.gene.com/good/local-initiatives/science-education/science-garage

Day 1: Introduction		
45 min	Using a virtual meeting platform, the teacher will present various statistical terms (confidence interval, standard deviation, Student's t test) to students. While students are watching the videos, pause and provide work time for the problems. As homework, students should watch *The Origin of Species: The Beak of the Finch*, a 16-minute documentary. Students will submit three words that they learned from the documentary to a Google Form. Teacher will then compile the list of words and generate a WordCloud by importing the CSV file.	Copy of Slides *The Origin of Species: The Beak of the Finch*
Day 2		
45 min	Teacher will start the discussion by showing the word cloud. Example words that might strike out would be beak size, Galapagos finches, and evolution. These words should be familiar to students since all biotechnology students have taken biology. Teacher will then synchronously analyze finches data using HHMI student handout. Scaffolding occurs to demonstrate examples on how to calculate descriptive statistics variables (sample mean, sample standard deviation, sample size) and t test. Teacher will generate a histogram showing the mean and standard deviation of the body mass, beak depth, wing length, and tarsus of survivors and non-survivors. Asynchronous Homework* Using HHMI Spreadsheet Tutorial 4: t-test, students will perform the t-test on beak size measurements of Galapágos finches. Synchronous instructions will include guidance on document access and setting up the manual example. Students should do this as an asynchronous formative assessment. Students will submit this to the specified learning management system.	HHMI student handout Student's t-test Tutorial on Google Sheet
Day 3		
45 mins	As a warm up, the teacher reviews the Tutorial 4 with students. The synchronous time is to review pitfalls and Google Sheet formula. As a formative assessment for homework, provide a problem with a different data set and ask students to evaluate the data using the Student's t-test. Assessment is taken and graded on Google Classroom.	Cricket Data for Formative Assessment
Final Project: Scientific Poster		
2-3 weeks	Students will design an experiment based on a performed experiment from the lab manual. The experimental design should include a control and experimental variable. Throughout the weeks, research teams will meet with teachers and Genentech volunteers to go over the research design. Due to 2020-2021 limited campus access, experiments are most likely limited to at-home lab kits and plant-based. For example, students can evaluate tomato plant growth with different fertilizer brands. Then, students will use t-test to evaluate their research hypothesis.	Designing a Scientific Poster Presentation rubric Poster example without statistics An exemplar example figure

	Students will then present their results in a poster symposium hosted by the school district and/or Genentech. Students will also be graded according to a rubric.	with Student t-test.

In-Person Enhancements

- In-person coaching
- Laboratory experiments can be implemented
- Posters will be printed

Supply List

References

Genentech. (n.d.). Retrieved July 08, 2020, from https://www.gene.com/

Shatat SM, Eltanany BM, Mohamed AA, Al-Ghobashy MA, Fathalla FA, Abbas SS. Coupling of on-column trypsin digestion-peptide mapping and principal component analysis for stability and biosimilarity assessment of recombinant human growth hormone. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2018;1072:105-115. doi:10.1016/j.jchromb.2017.11.007

Strode, P. (2020, April 30). Evolution in Action: Data Analysis. Retrieved July 08, 2020, from https://www.biointeractive.org/classroom-resources/evolution-action-data-analysis

Keywords

Biotechnology
Statistics
Student T Test

Links to Files in this ETP

Introductory slides:
https://docs.google.com/presentation/d/1PeKoek97HtWXu5_Xdve9J4rvBctGfeHgEE7SZ1qlYWo/edit?usp=sharing

HHMI student handout:
https://drive.google.com/file/d/1LKaHAxxfUELmeznc9Hqtq74gcfmM0Zo7/view?usp=sharing

Student's t-test Tutorial on Google Sheet:
https://docs.google.com/spreadsheets/d/1jCxOGE-b4wMUKao08McFJKoNU7gECfBjVKzxSMH-Po0/edit?usp=sharing

Designing a Scientific Poster:
https://docs.google.com/document/d/1lS7RNdzB663j5ZVeqKMEi2Cr_pQ1YhHloE33lSZ-dAE/edit?usp=sharing

Presentation Rubric:
https://drive.google.com/file/d/1yCZ37oSOWjiIy-Ob4UFi4OG1iScPM1gi/view?usp=sharing

Poster Example without Statistics:
https://drive.google.com/file/d/1C_CkjwKh5QaSzpELZny1Y6tK3_QRA4ei/view?usp=sharing

An exemplar example figure with Student t-test:
https://drive.google.com/file/d/1xFJzAfNRze7FQPv2jNl_ktEQ50fz4YK0/view?usp=sharing