Advances in Biology Laboratory Education
Publication of the Association for Biology Laboratory Education
Volume 42, Article 35, 2022

The Use of Databases to Support Undergraduate Research Experiences

David Mindorff

SUNY Buffalo, Department of Curriculum, Instruction, and the Science of Learning
c/o Lyford Cay International School, Nassau, N7776, The Bahamas,
(dmindorf@buffalo.edu)

There are benefits of undergraduate research experiences, but they can be resource intensive. Computer-based research enabled by databases can achieve similar objectives to bench-based research experiences. Examples of databases that can be used to support open inquiry are outlined.

Keywords: undergraduate research experiences, course-based research experiences, computer-based research experiences

Introduction

Course-based undergraduate research experiences (CUREs) are often those which engage students in an investigation in which the outcome is unknown (Ballen et al, 2017). This can provide students with a better understanding of how scientific knowledge claims are justified in comparison to traditional 'cookbook' laboratory experiences (Brownell et al, 2012).

The International Baccalaureate (IB) diploma programme is a secondary school finishing qualification. Students studying IB diploma biology undertake a practical experience that is an externally moderated self-directed inquiry. It involves an experimental investigation where the research questions are expected to yield results that are unknown before the investigation is conducted. Students are expected to develop a focused research question, develop a methodology that addresses the research question, collect sufficient relevant data, process and interpret the data and form a detailed conclusion to the research question. As this is routinely successful in the secondary school context, it provides a model for what can be achieved in undergraduate courses that are designed to offer students a research experience. The main difference being that undergraduate classes are often much larger than secondary school classes.

Bench-based research investigations can place significant demands on human, material and lab space resources (Kirkpatrick et al, 2019). Computer-based opportunities expand the possibilities and range of CUREs and require less resources.

Importantly, the lack of access to laboratory equipment and physical lab facilities during the period of virtual learning caused by the COVID-19 pandemic created an opportunity to experience the benefit of computer-based research opportunities (Kelly, 2021; Campbell et al, 2020). Instructional innovations suited to virtual instruction are also likely to have a lasting impact on distance learning after the pandemic is over.

There has been a growth in the use of databases and datasets to carry out inquiry tasks in undergraduate science classes for other reasons (Brownell et al, 2012; O'Reilly et al, 2017). There are a number of factors providing motivation for this inclusion such as improving quantitative literacy (Klug et al, 2017) improving problem-solving skills, visualizing concepts through real-world examples (Greengrove et al, 2020) and providing an antidote to the overuse of direct instruction strategies (Kjelvik & Schultheis, 2019). During the COVID-19 pandemic, many IB students turned to databases as the foundation for their data collection.

The author has successfully supported this type of research experience in the context of the

secondary school biology classroom. The student outline indicates some examples of databases that have been successfully used to support independent student research.

Student Outline

Objectives

Develop a focused research question
Develop a methodology that addresses the research question.
Query a database for sufficient relevant data that can support a detailed and valid conclusion related to the research question.
Correctly process and interpret data so that a detailed conclusion to the research question is deduced.

Introduction

A database is a structured collection of information stored on a computer. It can include data in a range of formats including qualitative information, articles, images, or quantitative information. A researcher can employ a database to do several tasks such as adding the results of their research for others to access, extracting subsets of data or querying the database by searching for a particular piece of data. Advances in technology and computing have meant that the rate of creation, storage and publication of data is increasing. One research report tracked the growth in information in bioinformatics databases for example and concluded that the amount of data collecting has a doubling time of between 12 and 24 months (Cook et al, 2016). This fact presents both a challenge and an opportunity. The data is growing faster than it can be 'mined' or analyzed and studied. At the same time, the public availability of many of these databases means that the barriers to entry for students interesting in conducting novel investigations are less than those investigations that are dependent on lab-bench based investigation. Analogous to a database is a large set of data that can be processed in different ways to uncover patterns or test hypotheses. This document provides you with examples of research questions that can be explored using a database providing both an illustration of the possibilities as well as possible protocols that can be replicated.

Methods and Suggested Research Questions

Option 1: The Audubon Christmas Bird Count

The Audubon Society of North America has been sponsoring the annual Christmas Bird Count (CBC) in the month of December for over 100 years. The resulting population data over time for hundreds of species is stored in a public database that can be readily accessed online https://netapp.audubon.org/cbcobservation/ .

a. The website contains a bibliography of many scientific research papers that have been published using information stored in the database. Many of these articles were first authored as far back as 50 years ago and could be replicated to update conclusions.
b. The emerald ash borer (*Agrilius plannipennis*) is an invasive alien insect in large parts of Canada and the United States, where it has killed over 100 million trees particularly in the genus *Fraxinus*.
   The emerald ash borer infestation is likely to have an impact on bird populations. Choose one of the hypotheses below and test it using the CBC database, filtering your search by location to focus on the Great Lakes region.
   i. The emerald ash borer infestation provides food for birds that eat insects. Insectivores possibly affected include red-bellied woodpeckers and white-breasted nuthatches showing numerical increases.
   ii. The death of ash trees will reduce the food supply of birds that feed on ash seeds so these birds would decrease in numbers.
   iii. The red-bellied woodpecker and pileated woodpecker are known to use ash trees for cavity nesting. The number of dead trees with hollows may increase providing greater nesting habitat or the loss of mature trees may reduce their numbers.

c. LaSorte and Thompson (2007) used the database to calculate the mean latitude of observations of some bird species over a range of years. A student could use their methodology to determine if there has been any northern range expansion of a bird species as might be predicted due to climate change.

Option 2: Survivorship Curves

Survivorship curves are graphic representation of the number of individuals in a population that can be expected to survive to any specific age. Explanations of the concept of survivorship and guidance as to how to construct the curves can be found in Lanza (2012) and Rauschert (2010). A pdf repository of birth and death date

data taken from cemeteries across the US is housed on the Hamilton College website (https://academics.hamilton.edu/biology/ewilliam/cemetery/default.html#datasets) .

Examples of research questions include:

a. In what ways and to what extent do the survivorship curves of males and females compare in the 1860s?
b. In what ways and to what extent do the survivorship curves of females compare in rural and urban communities in the 1860s?
c. In what ways and to what extent do survivorship patterns change for people born in the 1820s, 1870s, and 1920s?

Option 3: Menstrual Hormone Inquiry

The 'Open Door' biology website (https://www.saburchill.com/IBbiology/ICT/dataprocessing/025.html) published a data set from 20 healthy women volunteers aged 20 to 35 years old The same web-site provides instructions on how to process the data using a spreadsheet. These volunteers had not been taking oral contraceptives for at least 3 months before the start of the study. Daily morning blood samples were taken on the first day of their menstrual cycle and continued until the onset of the next cycle. Levels of six hormones were measured from the blood samples. Possible inquiry questions using the data include:

a. What evidence is there from the data of a negative or positive feedback relationship between any two hormones?
b. Test the null hypothesis that there is no significant difference in the hormone levels between women 20-28 years old and 29-35 years old.
c. Graphical presentations of the variation in concentration of FSH during the menstrual cycle show the peak in concentration occurring either in the follicular phase or during ovulation. Use the data to determine which is the accurate picture.

Option 4: Primate Skeleton Allometry

The website eAnthro (http://www.eanthro.org/) houses a directory of databases eFOSSILS, eLUCY and eSKELETONS. Each of these is a repository of bone photos with scale bars, often taken from multiple angles. Allometry refers to the scaling relationship between the size of a body part and the size of the body as a whole. Each of these recommendations involves reference to allometric measurements.

a. The chimpanzee is relatively quadrupedal compared to humans. How does mode of locomotion affect bone morphology?
   Compare the relative length of the ilium of the chimp with that of *Homo sapiens*.
   Compare the relative arm length of the chimp with that of *Homo sapiens*
   Compare the relative length of leg bones of the chimp with that of *Homo sapiens*
   Compare the ratio of long bone length of the arm and leg in both
   Compare the foot structure of the chimp with the human
   Compare the position of the foramen oval on the skull in the two skeletons.
b. What similarities and differences can be seen between humans and the skeletons of fossil hominids such as *Australopithecus afarensis* that can be used to infer mode of locomotion.
c. Tooth shape and skull morphology varies with the properties of the foods that primates eat (Unger, 2015). Dental form and function and skull morphology can be used to deduce the diets of different extinct and extant primates.

Option 5

Digital Coyote (https://wikieducator.org/Digital_Coyote) is a collection of 96 skull images from a broad range of North American locations.

a. Students can test hypotheses of whether there are regional variations in various skull dimensions. Measurements can be made from images using digital analysis software such as ImageJ. Vu et al (2018) published a research report comparing regional differences in skull size using the database. Possible questions that could be addressed by the dataset include whether there is a correlation between latitude and skull size.

b. Is there a correlation between mean average January temperature (using a database such as ClimateCharts.net) and skull size?
c. Is there a correlation between biome type and skull size?

Notes for the Instructor

Students often need support in coming up with meaningful research questions. The instructor could limit the exercise by providing students with the research question otherwise leaving the student to follow the remaining stages of scientific investigation. This still leaves open the possibility for the kind of significant student input and decision making that characterizes the scientific endeavor.

Support is often needed in terms of the procedures unique to different databases. Support may also be needed with the manipulation of data within a spreadsheet, the construction of correct visual representations and the use of basic statistics. In forming their conclusions, students often need to be reminded about the distinction between correlation and cause and effect.

A challenge is to provide support with diverse software such as Excel, Numbers or Google Sheets. It is recommended to require the use of Sheets, as it is free and there are online tutorials for most operations. It is important that whatever statistical package is used, students use the same one so that the instructor need only have facility with one package.

Cited References

Ballen CJ, Blum JE, Brownell S, Hebert S, Hewlett J., Klein, J. R., ... & Cotner, S. (2017). A call to develop course-based undergraduate research experiences (CUREs) for nonmajors courses. CBE—Life Sciences Education, 16(2).

Brownell SE, Kloser MJ, Fukami T, Shavelson R. Undergraduate Biology Lab Courses: Comparing the Impact of Traditionally Based" Cookbook" and Authentic Research-Based Courses on Student Lab Experiences. Journal of College Science Teaching. 2012 Mar 1;41(4).

Cook CE, Bergman MT, Finn RD, Cochrane G, Birney E, Apweiler R. The European Bioinformatics Institute in 2016: data growth and integration. Nucleic acids research. 2016 Jan 4;44

Kirkpatrick C, Schuchardt A, Baltz D, Cotner S. Computer-based and bench-based undergraduate research experiences produce similar attitudinal outcomes. CBE—Life Sciences Education. 2019;18(1):ar10.

Klug JL, Carey CC, Richardson DC, Darner Gougis R. Analysis of high-frequency and long-term data in undergraduate ecology classes improves quantitative literacy. Ecosphere. 2017 Mar;8(3):e01733.

Lanza J. Demography from physical cemeteries, “virtual cemeteries,” and census data. Teaching Issues and Experiments in Ecology. 2012;8...

Sorte FA, III FR. Poleward shifts in winter ranges of North American birds. Ecology. 2007 Jul;88(7):1803-12.

O'reilly CM, Gougis RD, Klug JL, Carey CC, Richardson DC, Bader NE, Soule DC, Castendyk D, Meixner T, Stomberg J, Weathers KC. Using large data sets for open-ended inquiry in undergraduate science classrooms. BioScience. 2017 Dec 1;67(12):1052-61.

Rauschert ES. (2010). Survivorship curves. The Nature Education Knowledge Project, 3(10).

Russell SH, Hancock MP, & McCullough, J. (2007). Benefits of undergraduate research experiences. Science 316 (5824), 548-549.

Ungar PS (2015) Primate Teeth and Plant Fracture Properties. Nature Education Knowledge 6(7):3

Vu CM, Piniewski DW, McLean OA, McCabe DJ. Use of point-and-shoot photography to compare regional differences in *Canis latrans* (Coyote) skull size. Northeastern Naturalist. 2018 May;25(2):319-32.

About the Author

David Mindorff is a biology instructor and school head at Lyford Cay International School in The Bahamas as well as a PhD student in the department of Curriculum, Instruction and the Science of Learning at SUNY, Buffalo.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit http://www.ableweb.org/.

Advances in Biology Laboratory Education is the peer-reviewed publication of the conference of the Association for Biology Laboratory Education. Published articles and extended abstracts are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board. Published abstracts are evaluated and selected by a committee prior to presentation at the conference.

Citing This Article

David Mindorff. 2022. The Use of Databases to Support Undergraduate Research Experiences. Article 35 In: Boone E and Thuecks S, eds. *Advances in biology laboratory education*. Volume 42. Publication of the 42nd Conference of the Association for Biology Laboratory Education (ABLE). https://doi.org/10.37590/able.v42.art35