Determining confidence: sex and statistics

Terry L Derting[1], Diane Ebert-May[2], Janet Hodder[3], and Everett P Weber[2]

In this article we demonstrate the use of a primary research paper as an assessment tool, helping to determine how well students understand a biological concept and their ability to interpret statistics. The article by Willingham (p 309–313) investigates the effects of temperature in conjunction with endocrine disrupting chemicals (EDCs) on sex determination in turtles. We have designed instruction and assessments on the subject of sex determination, a topic that is conceptually uncomplicated for most students, and statistical analyses which present them with some challenges. We have made the assumption that students have learned about mechanisms of sex determination among animal taxa, know how to develop and test hypotheses, and have a basic understanding of natural selection and fitness.

Student goals

- Apply understanding of sex determination to the consequences of altered sex ratios in animal populations.
- Demonstrate understanding of statistical testing and skills in interpreting data used in a research paper.

Instructor goals

- Use primary literature as a source of information about biological topics and as an assessment tool.
- Implement an active learning strategy to help students understand the concept of statistical testing and significance.

Engage – content

Begin the class with a question for students to discuss in their groups: "In some coastal areas, well-meaning individuals dig up eggs laid by sea turtles on beaches and rebury them further inland where the eggs are better protected. What impact(s) do you think this has on the sexual development of these turtles?"

After selected groups report out, the instructor summarizes the discussion, adding information about sex determination and EDCs. The topic of EDCs is of interest to students, as the mechanisms and ubiquity of the effects are easily understood and are of personal relevance. Although controversy exists regarding the links between endocrine disruptors and negative impacts on human health, it is evident that these compounds are present on a global scale, with high levels occurring in the blood or body fats of humans and wildlife. This introduction leads to an exploration of the use of statistics in these types of studies.

[1]*Murray State University,* [2]*Michigan State University,* [3]*University of Oregon*

Explore – statistics and data

Two objectives guide students' exploration of statistics: (1) connect data interpretation and confidence level with statistical testing, and (2) connect statistical analyses of data with support or rejection of hypotheses.

The instructor polls the class, using computer-based personal response systems, "clickers" (Brewer 2004), or hand-held cards to provide real-time displays of responses to questions like the following, using five choices (eg 100%, 95–85%, 75–65%, 55–45%, < 40%):

"What level of confidence does an engineer need to have in a new material for building bridges for public use before the bridges are actually built with that material?" Students usually say 100%. After a discussion that perfection is never possible, groups of students decide what level of confidence they think is acceptable for these and similar items. These types of questions lead naturally to a discussion of the effects of sample size and data variability on confidence in results.

Sample size and confidence levels

Students then discuss the following scenario to explore determination of sample size:

"Suppose you are the head of a drug-testing team. You have a pool of 10 000 people on whom to test the drug. What sample size of people will you use – 10, 100, or 1000? You may assume that one individual by chance alone responds unpredictably to the drug. Explain your answer in terms of your confidence that the drug effects on all individuals measured are truly representative of the drug."

In groups, students calculate the impact of one anomalous outcome within a sample of 10, 100, or 1000 and relate their solutions to thinking about sample size and confidence levels. Why would they/would they not use a smaller or larger sample size? Students report out answers.

P-values

The next hurdle for students is in understanding what *P*-values mean in relation to hypothesis testing. Most students view hypotheses as absolutes – ie right or wrong – and have difficulty understanding statistical significance. The instructor builds on the previous activity by explaining the meaning of a *P*-value and that 0.05, which is equivalent to a 95% confidence level, is a value traditionally used to indicate actual rather than chance effects of treatments if the null hypothesis is true.

Explain – analysis

Students work in groups to explore the concept of statistical significance using *P*-values. Groups work with one of two datasets, representing the occurrence of cricket frogs by gonadal sex in relation to time period and/or geographic region (Figures 1a, b). Students interpret the data and consider the results in terms of corresponding *P*-values and hypothesis testing.

Problem

Reeder *et al.* (2005) studied museum specimens of cricket frogs, looking for relationships between EDC use, population declines, and the occurrence of intersexuality. Specimens were examined from three regions, each differing in human population density and land use, over five time periods which differed in relation to EDC use. After explaining the purpose of the studies, groups write a null hypothesis(es) based on the study and proceed as follows:

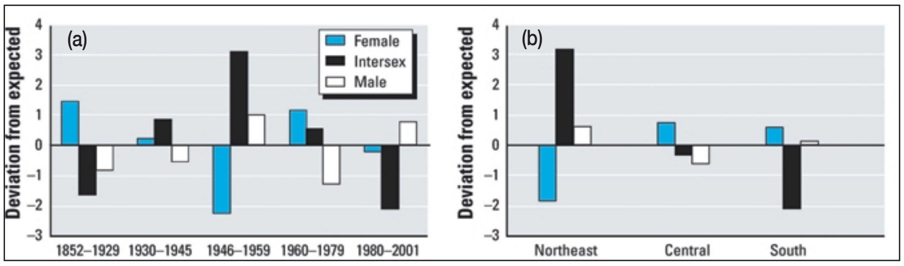

Figure 1. *Deviations of observed from expected values of cricket frog sex (a) by time period and (b) by region. Reproduced from Reeder* et al. *(2005) by permission from* Environmental Health Perspectives.

Group 1 (Figure 1a)

- Decide whether EDCs have primarily a feminizing or masculinizing effect on cricket frogs, justifying the answer in terms of the data.
- Explain what the authors mean by: "The proportion of specimens in each gonadal sex class differed significantly among the time periods of collection ($P < 0.001$)".

Group 2 (Figure 1b)

- Explain what the authors mean by: "The proportion of specimens in each gonadal sex class differed significantly among the geographical regions ($P < 0.001$)".
- Provide an ecological explanation for the statement, "In the 1990s, very few museum specimens of cricket frogs were available from regions that previously had the most elevated intersex rates".

The instructor picks groups to present their responses. Students are asked whether they think the results support or fail to support their group's stated hypothesis and to rate their confidence (strong, average, weak) in their response.

In a minute paper (Angelo and Cross 1993), students explain: (1) how *P*-values relate to confidence, and (2) why they stated that the results support or fail to support their group's stated hypothesis, justifying their answer in terms of the probability that the results occurred by chance alone.

Active homework – summative assessment

Willingham's paper is read for homework and questions are used to assess students' understanding of sex determination and the use of statistics. The instructor can evaluate all or several of the questions and should post one response that meets the highest level criteria for all questions. Questions can also be used in quizzes.

Content questions

- Explain the physiological mechanism whereby atrazine (and/or temperature) alters the sex of turtles.
- Explain two ways in which the shift in sex ratio between the control group and the atrazine–warm temperature group (see Figure 1 in Willingham paper) might influence a population.
- In what ways would population impacts differ if the effect of the treatments was an increase in the proportion of males, rather than females, in a population?
- What is meant by: "Size at hatching has been implicated as a fitness factor in reptiles, including turtles"?

Statistical questions

- Explain how the statistical results support or fail to support the hypothesis. In your explanation, show your understanding of the meaning of a *P*-value and confidence level.
- How would your confidence change if Willingham used a *P*-value of 0.1 instead of 0.05 to determine statistical differences among measurements of the experimental groups? Explain.
- What is the importance of statistical analyses in reaching conclusions based on evidence?

Final note

The answers to the homework questions provide an assessment of students' ability to apply the understanding gained in this class. The next steps will depend on how the instructor processes the answers and what is learned from these answers. The following class period may be a time to reinforce the content or statistical ideas that students found difficult.

Acknowledgements

We thank the National Science Foundation for their long-term support of the FIRST project, Faculty Institutes for Reforming Science Teaching (DUE 0088847). All Pathways papers are peer reviewed.

References

Angelo TA and Cross KP. 1993. Classroom assessment techniques. San Francisco, CA: Jossey-Bass.

Brewer CA. 2004. Near real-time assessment of student learning and understanding in biology courses. *Bioscience*. **54**: 1034–39.

Reeder AL, Ruiz MO, Pessier A, *et al.* 2005. Intersexuality and the cricket frog decline: historic and geographic trends. *Environ Health Perspec* **113**: 261–65. (http://ehp.niehs.nih.gov/members/2004/7276/7276.html).

