

Using Seashells to Teach Statistics through Experimental Design

Hans Lemke

University of Maryland, Department of Biology, Bldg. 144 Rm. 1210, College Park MD 20742 USA
(hlemke@umd.edu)

The use of scenario based exercises is beneficial when teaching complex or abstract material such as statistics, a subject that is challenging to learn and difficult to teach. Framing this instruction in the context of experimental design allows students to learn the essential concepts of statistics in a familiar setting. This also helps them to appreciate the practical applications of what they are learning. In our introductory ecology and upper-level experimental design classes we have found seashells to be particularly useful for these lessons as students find them familiar and engaging.

Keywords: Statistics, experimental design

Link to Original Poster: <http://www.ableweb.org/volumes/vol-34/poster?art=55>

Introduction

We have found that placing the instruction of statistics in an experimental design context helps students to more easily understand the material. Seashells have proven to be excellent subjects for these lessons. They are inexpensive, durable, familiar, interesting and easy to adapt to scenario-based exercises. We have successfully used shells in both introductory and graduate level courses.

The learning objectives for the introductory course are data types, hypotheses, mean, standard deviation, and basic hypothesis testing. Scenarios are used to provide a controlled exercise to introduce students to basic concepts in statistics and the mechanics of calculating t and χ^2 tests. This information is spread out over two laboratory exercises. In the first, students learn about continuous data, generate summary statistics (mean \pm standard deviation), create null and alternative hypotheses, and test these with a t test. In the second, students learn about discrete data, create null and alternative hypotheses, and test these with a χ^2 test.

In the graduate course, the learning objectives include data types, hypotheses, sampling, principles of experimental design, and more advanced tests such as ANOVA. The ex-

ercise using dyed *Bullia* shells emphasizes that a researcher's decisions about the measurement and categorization of the shells is relevant to the analysis that they will perform. Students also get to explore the difficulties in sampling and the difference between haphazard and random techniques. By the time the students complete the second exercise, using the assorted small shells, they have learned about hypotheses and are able to design and conduct an experiment. They often struggle with achieving random sampling and this provides an excellent opportunity for discussion. Students tend to start with complex designs and soon learn that a simpler approach is required.

Seashells have proven to be excellent subjects for the design of exercises for teaching statistics and experimental design. For introductory courses, they are inexpensive and durable materials that can be used in well-defined exercises. In a graduate course, the wide variety of shells available has made it relatively easy to replace traditional lectures with active exercises, allowing students to work through the material in a much more engaging manner.

Student Outline

Note: Some tables have been omitted due to space constraints.

Exercise 1.

Describing continuous variation in populations and testing for statistically significant difference between means:

The t test

Scenario

Snails of the genus *Bullia* live in marine habitats, and we have obtained samples from two different populations (Fig. 1). The “natural” population comes from a pristine bay with no obvious human impact. The “impacted” population comes from a nearby bay on which there are a number of factories. Snails in this bay are variously colored because they absorb pigments from the textile factories along the shore. The textile dyes themselves are not considered toxic, but because they are associated with many other substances produced by the mills, you are curious to see whether the size distribution in these two populations is different. In this exercise you will 1) measure and describe shell length in the samples of “natural” and “impacted” snails; and 2) test whether the differences in shell length between “natural” and “impacted” snail populations suggest an influence of environment, or should be attributed to chance.



Figure 1. A set of natural and dyed *Bullia*.

Measure the lengths (in mm, round to the nearest millimeter) of all the natural and impacted snails in your samples, and enter the values in Tables 1 and 2 respectively.

Table 1. Natural snail shell lengths (mm).

Specimen	Length (mm)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Table 2. Impacted snail shell lengths (mm).

Specimen	Length (mm)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Mean Shell Length

Calculate mean shell lengths for the snail samples below:

Mean of the natural snail shell length

$$\bar{x} = \frac{\sum x_i}{n}$$

Mean of the impacted snail shell

$$\bar{x} = \frac{\sum x_i}{n}$$

Standard Deviation

For calculating the standard deviation, Table 3 will be useful. We understand that this can be quickly done on a computer, but just this once you will be required to do it by hand, otherwise it will just seem like voodoo.

Table 3. Calculations of Standard Deviation for Shell Lengths.

A. Natural Shell Sample			
	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

$$\sum(x_i - \bar{x})^2 = \underline{\hspace{2cm}}$$

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}} = \underline{\hspace{2cm}}$$

Table 3. Calculations of Standard Deviation for Shell Lengths.

B. Impacted Shell Sample			
	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

$$\Sigma(x_i - \bar{x})^2 = \underline{\hspace{2cm}} \qquad s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n - 1}} = \underline{\hspace{2cm}}$$

- For either the natural or impacted populations, do you think your estimate of the mean shell length of the population will be exactly right (i.e., the same as the true population mean)? Why or why not?
- If you were to take another sample (e.g., measure a different sample of snails from the same locations), do you think you would get exactly the same sample mean? Why or why not?
- Would measuring a much larger number of snails (e.g., 2,000) give you greater confidence in your estimate?
- If the standard deviation was really large, would that increase or decrease your confidence in your estimate of the population mean? Why?

Now that you have calculated mean length for samples of snails from different locations, it’s time to return to our original question: “Are they different?” Well, in a trivial sense they almost certainly will be. As we’ve already seen, even if you take two samples from the same population the sample means are likely to differ due to chance. This is increased with smaller samples or larger standard deviations. But what about the effects of being from a pristine or an impacted environment? At what point do we decide that chance is not an adequate explanation for the difference in means? This requires statistical testing. We begin with a null hypothesis (H_0)—essentially a hypothesis that the factors we are investigating have no effect.

Write the null hypothesis for our investigation of the effect of source on shell length below:

H_0 :

This would mean that snails have the same length regardless of source. If, when we compare our two sample means, we determine that the probability of getting the differences we see just by chance (i.e., assuming they came from populations with the same mean) is very low, we reject our null hypothesis, and accept our alternative hypothesis (H_a).

H_a : The samples come from populations having different means.

How unlikely must it be for the samples to have come from populations with the same mean before we reject our null hypothesis? In most cases, scientists will reject the null hypothesis if the probability of it being true is less than 5% (0.05)—this is referred to as the significance level (α). If the chance of getting sample means this different if drawn from populations with the same mean is less than α , the differences are considered to be *statistically significant*.

Briefly distinguish “difference” and “statistically significant difference.”

If our sample means differ by a given amount, how do we determine the likelihood of this difference being due entirely to chance? What factors influence this likelihood? There are three factors that are important for us:

1. *The difference between the means.* The more different the two means are, the more likely we are to reject the null hypothesis that they came from populations having the same mean (other things being equal).
2. *The amount of variation.* If the populations from which the samples were drawn have little variation, we wouldn't expect the sample mean to differ very much. In reality, we don't know the standard deviation of the population, but we can estimate it based on the standard deviation in our samples.
3. *Sample size.* As we've already discussed, the larger the sample size, the less likely we are to have sample means that differ greatly from the true population mean (or from each other).

How do we calculate the probability of getting the differences we see? We will use a statistical test called a t-test. The formula for calculating the t-statistic is:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Note: This formula assumes that the two populations have equal standard deviations and that we have fairly large samples. We won't explore this complication here, but you will in statistics courses. The subscripts 1 and 2 refer to the two treatments we are considering (e.g., male vs. female, Anabaena presence vs. absence). The larger the t-statistic is, the more likely you are to reject the null hypothesis. A large t-statistic, not surprisingly, can come from having a large difference between the two sample means, or small standard deviations, or a combination of the two.

Calculate the t-statistic for your comparison of Natural vs. Impacted snail shell length:

$$t_{\text{calc}} =$$

To determine if the difference is statistically significant, you must compare your calculated t-value to a table of critical t-values (t_{crit}). We've seen that the difference between means, and the amount of variation, was accounted for when you calculated your t-value, but what about sample size? You would expect that the larger the sample size the more likely you would be to reject the null hypothesis for a given t-value.

In Table I of critical values (omitted for space from this handout), you will see three columns. The left column contains a value called degrees of freedom - this is how sample size influences the likelihood of rejecting your null hypothesis. The degrees of freedom can be calculated for your study as $d.f. = n_1 + n_2 - 2$

Calculate your d.f. for Natural vs. Impacted shell length $d.f. =$

To determine to which tabled t-value you need to compare your calculated t-value, you will need to know the degrees of freedom and the significance level (0.05 in most cases, 0.01 if you really want to reduce the probability of rejecting a true null hypothesis to 1%). For example, if you measured *Selenastrum* densities in 10 microcosms each from the two treatments, and your level of significance is 0.05, you would reject the null hypothesis if your calculated t was greater than t_{crit} . Notice that the greater your degree of freedom, the smaller your calculated t-value needs to be to reject your null hypothesis.

Based on your samples, is there a statistically significant difference between the lengths of snails from different sources?

Exercise 2. Difference versus statistically significant difference in discrete data: The Chi-square test

The statistical reasoning described above for continuous data also applies to discrete data (as in the example with coin flips), and the procedures for testing for statistical significance are similar (the detailed background is a little less intuitive, though, so we won't explore it here). You form a null hypothesis, design an appropriate experiment, collect data, and statistically analyze your data to determine whether the differences from your expectation are attributable to chance (accept the null hypothesis) or are unlikely to be due to chance (reject the null hypothesis). There are many types of statistical tests for discrete data—we will use a very common one called the χ^2 (or Chi-square) test.

With discrete distributions, mean values and standard deviations are not very meaningful because the values fall into only a

few discrete categories. For example:

- *Natural selection.* *Drosophila* either survive or they don't. There are only two possible values for any individual fly: alive or dead.
- *Inheritance.* Wing variation either is inherited according to Mendelian proportions or it is not. In this case, there are only a limited number of possible genotypes and phenotypes that can result from a cross.
- *Population genetics.* Phenotypic and genotypic frequencies either conform to the expectations of the Hardy-Weinberg equilibrium, or they do not. Again, only a few discrete genotypes and phenotypes are possible.

The basic design of a Chi-square test involves the comparison of discrete frequency distributions. Our null hypothesis is always that our observed values are the same as our predicted values. The basic calculation of chi-square is as follows:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O = observed number of counts and E = expected number of counts. Note: You use the actual counts, *not* the proportions or percent of observed or expected frequencies!

In many cases you will have some reason to expect particular relative frequencies when a particular variable is sampled. The simplest example is a coin-toss experiment. We normally expect $\frac{1}{2}$ heads and $\frac{1}{2}$ tails. If we tossed the coin 20 times, we would expect 10 heads and 10 tails. Now suppose we did toss a coin 20 times and obtained 9 heads and 11 tails.

The calculated $\chi^2 = (9-10)^2/10 + (11-10)^2/10 = 0.20$.

The number of degrees of freedom for the goodness of fit test is: d.f. = C - 1, where C is the number of categories being studied.

Since there are two categories (heads and tails) this means there is 2 - 1, or 1 degree of freedom. We use Table II of critical values (omitted for space from this handout) at the end of this exercise to look up the critical value at the selected probability level of 0.05 and at one degree of freedom. We find that the χ^2 crit value is 3.84. Since χ^2 calc < χ^2 crit, we accept the null hypothesis. You can use this type of test for a single sample, with any number of categories, provided you already have the expected relative frequencies.



Figure 2. A set of ark shells.

Imagine the following scenario. Suppose you are an ecologist studying the biology of ark clams (Fig. 2). Your years of study have established that in healthy populations with normal birth and death rates, the ark clam population should consist of 60% one-year-olds, 30% two-year olds, and 10% three-year-olds. These clams do not live past age three, and we are disregarding age 0 because they are too hard to count. These different ages form discrete size classes as indicated in the table below. In a nearby bay, where a salmon farm has been operating for several years, residents have been reporting fewer large ark clams and ask you to investigate. Are ark clams dying at younger ages? Is the distribution of ark clam ages different from a typical ark clam population? You have a sample of ark clams from this population with which to answer this question. Spread your ark clams out before you and measure them across the longest length to assign them to the appropriate age class. Enter the numbers in the appropriate column in the Table 4.

Table 4. Size of ark clams in a population.

Age (yrs)	Size range (cm)	Typical proportion	Actual number in sample	Expected number in sample
1	< 3.8	60%		
2	3.8-5.1	30%		
3	>5.1	10%		

Write the null hypothesis for our investigation of the effect of salmon farming on ark shell size distribution below:

H_0 :

This is the same as saying that ark shells will have the same age distribution in areas with a salmon farm as in areas without. If, when we compare our sample (the shells in front of you) to what we expect (based on previous studies), we determine that the probability of getting the differences we see just by chance is very low, we reject our null hypothesis, and accept our alternative hypothesis (H_a).

Observed Numbers

How many did you find?

Age 1 (= O_1)? _____

Age 2 (= O_2)? _____

Age 3 (= O_3)? _____

Expected Numbers

We know from previous studies that the typical proportion of ages 1, 2, and 3 are 60%, 30%, and 10%, respectively. What are your expected numbers? (Remember—use actual numbers!)

Expected number of age 1 (= E_1)? _____

Expected number of age 2 (= E_2)? _____

Expected number of age 3 (= E_3)? _____

Calculate the Chi-square statistic for your comparison of age class distribution in clams:

$\chi^2_{\text{calc}} =$

d.f. =

As with the previous example, we will use a level of significance (α) of 0.05. Use Table II (omitted for space from this hand-out) to determine the critical value of the Chi-square for your comparison of Age class distribution in clams

$\chi^2_{\text{crit}} =$

Based on your sample, is there a statistically significant difference in the proportion of clam age classes near the salmon farm versus a typical population?

Exercise 3.**Introduction to experimental design using dyed *Bullia***

On the first day of class, students are given two boxes of dyed *Bullia* (Fig. 3) and are asked to devise ways to determine if these two populations are different. Students actively engage in the discovery of many of the problems that arise when designing and conducting research. After they have had time to work through it on their own, the groups compare their results. This allows the instructors to introduce technical terms (e.g. continuous and discrete data) and lead a discussion of the problems with generating true random samples.



Figure 3. Sets of *Bullia* for comparison.

Exercise 4.**Advanced experimental design and analysis using assorted small shells**

After students are more familiar with sampling and basic tests, they are given two bags of assorted tiny shells (Fig. 4) and asked to devise an experiment to determine if they come from similar populations. There are dozens of different species represented and thousands of shells in each sample making counting everything impossible. Each group designs and conducts an experiment. This usually results in performing numerous preliminary studies to finalize their design. An interesting situation that arises is that groups often arrive at different answers. This provides a great opportunity for discussion about research finding the “right” answer.



Figure 4. One of the sets of small shells.

Notes for the Instructor

Sources for shells

If you are fortunate to be near the water, shells can be collected in the wild. They can also be scrounged from the corners of biology department storerooms or purchased from commercial sources. Tourist shops and craft stores are also good places to look. For the best selection, a web search for “shell wholesaler” will provide numerous places to buy large quantities of specimens. We have purchased most of ours from U.S. Shell Inc. (www.ussshell.com).

Prep notes and sample data

Describing continuous variation in populations

When setting up sets of shells, it is not realistic to make all of the sets have identical sizes distributions. The easiest way to construct the sets is to use Tukey’s Quick Test. This simple test is based on the number of non-overlapping values in the data. If there is significant difference based on this test, the t test will find one as well. It may be helpful to consider students and instructors who are color blind when constructing the sets.

Natural shells (mm): 49, 47, 49, 32, 39, 34, 47, 36, 40, 49
mean \pm sd: 42.2 ± 6.75

Impacted shells (mm): 46, 54, 61, 50, 45, 64, 53, 43, 50, 47
mean \pm sd: 51.3 ± 6.86
 $t_{\text{calc}} = 2.99$, d.f. = 18, $p < 0.05$

Difference versus statistically significant difference in discrete data

To avoid confusion arising from measurement error, we have found that it is helpful to use shells that fit neatly into each category, in this case at least 2 mm from a border. This means that there is often a large number of shells that are not usable in an order (Table 5).

Table 5. Sample data for ark shell sizes.

Age (yrs)	Size range (cm)	Typical proportion	Actual # in sample	Expected # in sample
1	< 3.8	60%	9	17
2	3.8-5.1	30%	14	8
3	>5.1	10%	5	3

$$\chi^2_{\text{calc}} = 9.598, \text{ d.f.} = 2, p < 0.05$$

Introduction to experimental design using dyed Bullia

For this exercise we used a large number of dyed *Bullia* shells, though the actual species used is not important. For maximum effectiveness, the shells should vary in size and color, allowing students to devise experiments based on discrete or continuous data. In addition, there should be large enough quantities so that counting all of the shells in unreal-

istic. The sets of shells are haphazardly constructed, with no “correct” answer in mind. We conduct this exercise on the first day of class and the protocols that the students devise are usually severely flawed. This is used as a springboard for discussion about subjects like data type and random sampling. At this point, they do not do any statistical analysis.

Advanced experimental design and analysis using assorted small shells

This experiment uses an assortment of tiny (3-10 mm) shells available in 1 kg packages from U.S. Shell (Small Natural India Mix). A bag contains thousands of shells in at least 15 species. This provides a large variety and quantity of shells and 1 kg is used for each treatment.

Example of a Student Generated Procedure and Data (Fig.5):

1. Empty the bag of shells into a rectangular container (39 cm wide x 54.75 cm long x 8 cm deep).
2. Divide the area into a 3 x 3 grid. Label the plots on the grid 1 through 9, starting in the upper left corner and proceeding left to right.
3. Generate random integer through random.org to select test plot sample.
4. Fashion a 10.75 cm x 9.5 cm paper mask to cover the randomly selected plot.
5. Separate shells based on shape and rib pattern.
6. Results were: $\chi^2_{\text{calc}} = 35.99$ d.f. = 13 $p < 0.001$

Type	B	A
Ivory Ribbed clam	15	11
Dark Ribbed clam	20	19
Darkish Elliptical clam	62	56
Ivory Unribbed clam	5	10
White/Brown conical snail	67	44
Tight Spical Snail	45	81
Bulbous Spical Snail	61	68
Cinnabon Snail	13	6
Rough Conehead Snail	19	20
Smooth Conehead Snail	12	22
Discus Spical Snail	118	118
Round Snail	76 76	133
Other snail	1	1
other clam	0	1

Figure 5. Student generated data collection table.

Acknowledgements

Thanks to Jeffrey Jensen and Bretton Kent for helping with the development of these exercises and the continuous use and refinement in their classes, Michael Keller for his assistance and guidance, and Kaci Thompson and HHMI for providing funding to allow me present these exercises at ABLE. In addition, I would like to thank the many ABLE members who provided feedback and suggestions at the meeting.

About the Author

Hans is the Lab Coordinator for Principles of Biology II at the University of Maryland. He holds a B.A. in Biology from St. Mary's College of Maryland, an M.S. in Entomology from the University of Maryland, and an M.D.E. in Distance Education from University of Maryland, University College. He currently teaches labs for both semesters of introductory biology and experimental design. Current research focuses on educational outcomes in laboratory and online settings and a survey of tiny Miocene shark and ray fossils found along the Chesapeake Bay.

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/>.

Papers published in *Tested Studies for Laboratory Teaching: Peer-Reviewed Proceedings of the Conference of the Association for Biology Laboratory Education* 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.

Citing This Article

Lemke, H. 2013. Using Seashells to Teach Statistics through Experimental Design. Pages 450-459 in *Tested Studies for Laboratory Teaching*, Volume 34 (K. McMahon, Editor). Proceedings of the 34th Conference of the Association for Biology Laboratory Education (ABLE), 499 pages. <http://www.ableweb.org/volumes/vol-34/?art=55>

Compilation © 2013 by the Association for Biology Laboratory Education, ISBN 1-890444-16-2. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner.

ABLE strongly encourages individuals to use the exercises in this proceedings volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.