Exploring Experimental Design: An Excel-Based Simulation Using Steller Sea Lion Behavior

Wendy L. Ryan* and Ezry St. Iago-McRae

Department of Biology, Kutztown University, Kutztown, PA 19530

*Corresponding Author: ryan@kutztown.edu

Abstract: Experimentation is the foundation of science and an important process for students to understand and experience. However, it can be difficult to teach some aspects of experimentation within the time and resource constraints of an academic semester. Interactive models can be a useful tool in bridging this gap. This freely accessible simulation provides a unique opportunity for students to practice designing experiments and analyzing their results. The effects of sample size and variability on the usefulness and accuracy of experimental data are an important component of these exercises. In addition, students can easily repeat their experiments, demonstrating that repetition doesn't necessarily lead to exact replication due to natural variability. Lastly, the simulation provides a range of flexible input categories that allow students to develop their own experimental questions about Steller sea lion behavior, and to explore a range of parameters, including various specific behaviors, the sex of the animals, and various sampling intervals such as hourly, daily, seasonally, or even annually. While this exercise does not replace first-hand experience with experimentation, it provides a good foundation for students to build on as they begin the process of designing and implementing their own research projects.

Key words: Hypothesis testing, sample size, scientific method, Steller sea lion

INTRODUCTION

This exercise has two primary goals and one secondary goal. One primary goal is for students to gain experience designing experiments and testing hypotheses. Although this process is the foundation of science, it can be challenging to teach without defaulting to a static, formulaic description. The use of real data that can be quickly and easily "sampled" by students makes the process of designing experiments and interpreting outcomes more accessible to both students and instructors. The simulation also provides near instantaneous feedback, which is particularly useful when a student's experimental design (i.e. selection of input values) needs to be revised to produce data that will allow accurate evaluation of their hypothesis. The second primary goal is for students to develop basic skills in the analysis and interpretation of graphical data with particular attention to variation and sample size. This is another aspect of designing good experiments that can be difficult to teach in the lab or field because of the practical constraints of repeated sampling and also because it requires some statistical savvy. This simulation can stretch beyond these constraints to provide students with concrete research data to observe and evaluate, thereby gaining valuable insights into a critical component of good research. In addition, since the emphasis is on the graphical representation of the data for both statistical analysis and overall interpretation, students will become proficient in graphical analysis. This is a skill whose value extends into many aspects of their professional and personal lives.

The secondary goal of this activity is to provide a vehicle for students to investigate Steller sea lions and some circumstances under which specific behaviors may vary. This is a unique opportunity for students not only to "observe" the behaviors of a marine mammal, but to "see" how specific behaviors might be influenced by aspects of the animal's environment. The ability to manipulate observation parameters provides students with insights into the complexity of this research and the dynamic responses of these animals to the parameters represented in the simulation. This complexity is visible in spite of the fact that these data were collected under the relatively controlled conditions of an artificial habitat (that is, a large outdoor aquarium).

The simulation makes current research accessible to students, not as a presentation of findings, but as an opportunity to manipulate real data while developing and testing their own hypotheses and drawing their own conclusions. The simulation is composed of a set of macros developed for Microsoft Excel 2010™ using Visual Basic for Applications (VBA). The macros simulate data collection by selecting subsamples of the existing data set based on user inputs and presenting the results in both tabular and graphical format. There are also randomizing components within the data selection macro that may generate a different result with each iteration of the

simulation. The simulation interface allows students to quickly and easily model experiments, while allowing them to repeat the same experiment or adjust the inputs to conduct a different experiment within minutes. The interface design allows students to easily manipulate a variety of inputs including sample size, sex of the animals studied, and timeframe within which the data were collected, including time of day, month, and/or year, etc. Not only will students be able to see how changing the parameters of an experiment can affect their results, but they will also learn to apply some basic statistical analysis to test specific hypotheses.

The data used in this simulation were collected during regular observations of selected social behaviors of five (three female, two male) Steller sea lions (*Eumetopias jubatus,* hereafter SSL) housed in two outdoor exhibits at Mystic Aquarium (Mystic, CT). The males were never housed together, but any year-round association is a novel condition for these animals, since in a natural population social interactions would be limited to the rookery during the breeding season (Burkanov, et al., 2011). However, studying marine mammals in a zoological setting allows investigators to control or eliminate a variety of potentially confounding factors, such as migration, foraging, and predator evasion. In an uncovered outdoor exhibit, animals can perceive external environmental cues such as day length, air temperature, and season that can influence their behavior. Monitored behaviors included bite, chase, touch, and butt, as well as vocalizations and interactions with toys available in the exhibits. The presence or absence of animal trainers was also recorded, as direct interactions with humans also occur during training, feeding, and husbandry sessions. The designated behaviors were recorded using a tally-system on data sheets similar to Table 1. An observer monitored each individual SSL for a five-minute interval, rotating among the animals throughout the day, recording each occurrence of the selected behaviors. At times, individuals were taken off display. During these occurrences data collection continued to rotate among the remaining animals.

Some Background on Steller Sea Lions

Students will need a basic introduction to SSL behaviors in order to develop informed hypotheses. The brief overview below is provided as a starting point for this exercise. In an advanced course, students are asked to look further into the available literature in order to support both their hypotheses and conclusions, but when used in a freshman/sophomore level course, the students work largely with the background information in the following paragraphs.

Pinnipeds, semi-aquatic marine mammals, are separated into the families Phocidae, true seals, and Otariidae, sea lions and fur seals. Steller sea lions are the largest member of the Otariidae family, averaging 1,000 kilograms for adult males and 273 kilograms for adult females (Jefferson et al., 2008). They are found in the North Pacific Ocean along coastal regions in Canada and Alaska and extending to Russia and Japan. From the 1970s to the 1990s the SSL population declined by 80% (Calkins, et al., 1999). In 1990 *E. jubatus* was listed as a threatened species, and then in 1997 two distinct reproductive populations were identified and the population west of 144°W (near Cape Suckling, AK) was reclassified as endangered. The eastern population recovered to the point that it was delisted in 2012, while the western population continues to be listed as endangered under the Endangered Species Act (Speegle, 2013). It is hypothesized that the decline of these large predators is due to a combination of diverse factors, including parasites and disease, declining prey diversity due to overfishing or climate change, competition with commercial fisheries for prey, negative interactions with marine debris, and direct mortality due to killer whales and humans (DeMaster, et al., 2006; Trites, 2012). The balance and impact of both natural and anthropogenic factors may change over time in an unpredictable manner, which is often the case in biological systems. The

Table 1. A sample modified data sheet for Steller Sea Lion behaviors.

STELLAR SEA LION DATA SHEET #1: BEHAVIOR   DATE:_____

INDIVIDUAL OBSERVED? _____

Observation Start Time	Vocalization	Interactions					Trainer Present?	Observer's Initials
		Bite	**Touch**	**Chase**	**Butt**	**Toy**		

NOTE: Activities for an individual should be recorded continuously for 5 min using tallies.

simulation exposes students to this fact of real-world experimentation, albeit on a fairly short time-scale; from days to a few years. It also demonstrates the variability inherent in sampling itself, through the use of randomized selection of each data subset from the master data.

There is evidence that the behaviors of both sea lions and seals are influenced by a variety of factors such as day length and season, as well as the age and sex of the individuals. In wild populations of SSLs, interaction rates between individuals peak around the breeding season, from late May through July (Trites, 2012). But do these patterns carry over to animals maintained in artificial habitats? A study of juvenile captive bearded seals (*Erignathus barbatus*) found that only the males developed underwater vocalization, which is in agreement with observations of wild seals and may reflect the harem structure and dominant reproductive roles of male seals. In contrast, vocalizations in the captive seals were less complex, with the differences being attributable to either the immaturity of the seals at the time of the study or other factors associated with their artificial environment (Davies, et al., 2006). Another study of captive animals (Moulton, et al., 1999) examined the haulout duration of harp seals (*Pagophilus groenlandicus*) as a percentage of total daylight hours, finding that the maximum haulout duration of 18.2% occurred during the fall. So, not only did captive animals demonstrate seasonal variation, but changes in the amount of time spent out of the water may be linked to changes in other behaviors. These findings in related species may guide students in deciding upon hypotheses to test using the SSL simulation.

Data Analysis: Getting Started
It might be also be necessary to include a short treatment of statistics in the background information for students. The fundamental challenge of working with quantitative data is that there is variability inherent in the measures that scientists make for virtually any variable that can be measured, including the behavioral data in this simulation. Two closely timed observation periods of a single individual would most likely result in different counts for many if not all of the identified behaviors. These differences result from differences between observers and inherent variability in the activities of the animals from moment to moment.

In graphs and tables, scientists typically report the mean ($\bar{x}$), or arithmetic average, of the data they have collected from a population or from a sample of that population. If a scientist measures the number of vocalizations during a 5-minute focal period, the average frequency of vocalization is often the primary variable of interest. If the number of vocalizations were 5, 0, and 10 in three different 5-minute periods, then the mean would be the sum of all three counts divided by the number of focal periods: $\bar{x} = (5 + 0 + 10)/3 = 5$ vocalizations/5 min with a sample size n=3.

Standard Deviation (SD) is a way of expressing the variability inherent in the data and can be used by scientists to decide what sample size is needed to accurately estimate the mean for a given population. There is a formula for calculating the Standard Deviation, but in this case, SD is one of the outputs of the Excel simulation. Often, the mean and standard deviation of a sample are incorporated into a graph with the mean represented as the actual data point, in this case the height of the bars on the graph, and SD represented as an "error bar". The error bar then provides insight into the variation within the data set, since about 68% of the individual values in the sample will usually fall within one SD of the sampled mean. Increasing the sample size will give you a more accurate estimate of the variation within the population and a better estimate of the actual mean for the population.

The Standard Error (SE) is based on both the standard deviation and the number of data values in your sample. The SE expresses the variability of the estimate of the mean for the data values that have been collected. In essence, the SE indicates that if a sample of the same size were collected repeatedly from the same original population, then most (68%) means from the repeated samples would lie within one SE of the true mean of the entire population from which the samples were drawn. Thus, the larger the sample size, the closer the estimate will be to the true mean, and therefore the smaller the SE. In other words, the SE is inversely proportional to the square root of the sample size, so that if the sample size is quadrupled, our estimate of the population's true mean is about twice as accurate (because the scatter of such sample means around the population's true mean will be half as broad).

The SE is especially useful because it can be used to test for statistical differences between sampled means. If the difference between the means of two samples is "statistically significant", it means that the difference between the means is unlikely to have been caused by random sampling error; instead, the means are different probably because the sampled populations are truly different. If your goal is to use the mean and error bars from your results as the foundation for preliminary statistical analysis, then the appropriate error to use is SE, and not SD. For example, to test the hypothesis that male SSLs vocalize more frequently than female SSLs, you would need to sample the numbers of vocalizations of males and females over certain identical time periods, and then graph the two means with SE error bars to show the reliability of the estimated means. To determine whether the sample means for males and females are statistically significantly different, you could check whether the SE bars around the two means overlap (Motulsky, 1995). If the SE bars

around the two means overlap or have a gap between them less than the average length of the two error bars, as in Figure 3A, then the means are not significantly different. That is, the difference between the means is likely due to random sampling error. But if there is a gap between the two SE bars at least as long as the average of the two error bars, as in Figure 3B (remember, the error bars extend both directions from their respective means), then the two means may be significantly different, although the appropriate statistical test would be needed to determine this with certainty (Cumming and Finch, 2005).

METHODS

Activity 1. Hypothesis testing.

Using the information provided above, along with additional literature research as needed (or desired on the part of the instructor), students should develop an experimental question regarding the frequency of interactive behaviors for SSLs with respect to some variation in time. For example, students might expect the frequency of behaviors to change throughout the year or at different times during the day, but then they would need to predict either an increase or a decrease in these behaviors. Once students have an experimental question, they will need to state a testable hypothesis. Using a standard format for the hypothesis: If …. (Independent variable/predictor) … then …. (Dependent variable/response) should provide a conditional relationship that can be supported or falsified by the available data. The students are now ready to run the Excel simulation. Throughout the simulation, we will refer to the experimental population as Group Y and the reference population Group Z. Using the INTERFACE sheet (see Figure 1), the Number of Data Points for both Groups Y and Z should be set at 100 for this activity, since we do not want sample size to influence this experiment. Students should then change the input values for Group Y as appropriate for their research question. For this experiment students should be manipulating an input value for time based on the month, year, and/or time range for the data as highlighted in Figure 1. They will also need to consider what changes, if any, are needed to make the Group Z data set so that it can serve as the reference or control condition.

Figure 1. Image of the INTERFACE sheet of the Excel simulation highlighting the input values representing an aspect of time (1) that may be changed in order to test a question for Activity 1 and reinforcing that students should select 100 data points for both Group Y and Z (2).

Once students have selected all of their input values, they will need to click on the generate buttons at the bottom of each column on the interface tab to run the simulation (Figure 1). Both buttons must be clicked each time they want to re-run the simulation. The results will then be available under the GRAPH sheet, in both graphical and tabular format, although this exercise focuses on the graphical representations. For Activity 1, students should select Standard Deviation (SD) for the error bars from the box at the top of the GRAPH sheet. They should try running the simulation a few times to confirm that the results do indeed change! Once this has been confirmed, have students run the simulation several times. Although one might expect repeated samples from the same population to be similar, they may in fact differ due to random variation (see Figure 2). For the purpose of this exercise, choosing results that are noticeably different will make it easier for students to understand some of the principles of experimental design and data analysis that will be discussed; however, it would be appropriate to also impress upon students that when running an actual experiment all of their results must be included and evaluated.

Student Assignment

Copy the resulting graphs from two different runs of the simulation without changing any of the input values. In order to insert a graph into a Microsoft Word document right click on the graph and then do a copy/paste. Paste the graph as a picture so that it doesn't update as you continue to use the simulation. Include a short, descriptive caption for each figure, which is not only appropriate but will help you to easily identify and track all your graphs.
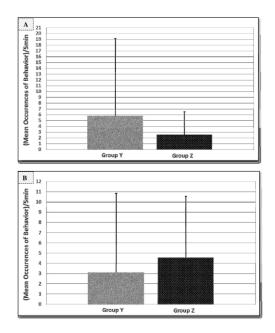
Figure 2. Sample graphs showing the average frequency of interactions for all individuals based on a comparison of activity from January to May (Group Y) vs January through December (Group Z) for two generated runs of the simulation (A and B). The error bars represent +/- 1 SD. Note that students should pay particular attention to the range of values for the y-axis as the scale may vary between their graphs.

Attempt to explain any differences between your graphs, in light of the fact that you didn't make any changes to the input values. What is responsible for the differences, if any, between your graphs? Review your two graphs, focusing only on the mean values. Indicate whether the results from each experiment seem to support or refute your hypothesis regarding SSL interactions and why. Are there differences between the two graphs such that one supports your hypothesis more strongly than the other or even that only one of the graphs supports your hypothesis? Attempt to explain any apparent contradictions in the results from your two simulated experiments. Is there a "take home" message from these results with respect to experimental design?

Now examine your graphed data again, but consider the error bars associated with each mean. First, identify the property that they represent, and describe what it conveys about the data. In combination with the mean values, can these be used to estimate whether there is a statistically significant difference between the two groups in your experiment? Why or why not?

Lastly, identify one additional research question that is raised by your results and clearly explain whether or not this new question can be examined using the current simulation.

Activity 2. Experimental design and sample size.

Now students should examine what other information is available within the simulation and the range of selections that can be made using the categories available on the INTERFACE sheet. Note that all of the parameters on the INTERFACE sheet can be changed; some are drop-down selections and others are manual entries. Based on observations, curiosity, and additional literature research as needed (or desired by the instructor), have students develop an experimental question about the interactive behaviors of SSLs. There are many possible questions and combinations for this data set, so the ideas should be as diverse and different as the number of different students in the class. Once students have an experimental question, they again need to develop a testable hypothesis before running the Excel simulation.

However, there is another underlying question for this particular activity, specifically, to what extent will the results differ based on sample size? So, in addition to an experimental hypothesis, students should predict the outcome of their experiment when using a very small sample size as compared to when using a larger one.

First, change the Number of Samples (or the sample size) to be used for both Group Y and Z on the INTERFACE sheet. The simulation interface allows for the selection of a sample size from 1-1096 (the maximum number of data points); however, depending on the experimental question and subsets of the data used, the actual number of available data points may be much smaller. Therefore, for this exercise it would be best to select a particularly small sample size first (15 or less), then one in the mid-range of the options (less than 30), and lastly one using all available data (up to 1096 data points). This ensures the largest sample size will be clearly bigger than the small and medium samples. In the example (Figure 3), the experimental question is whether the presence of a trainer would affect the frequency of behaviors for the SSLs. However, since we are also considering the impact of sample size and how it might affect the results, sample size for both groups was initially set at 11 (small) and then increased to 25 (moderate).

Once the experimental parameter inputs have been selected, and an initial (small) sample size has been set for both groups, students should click the generate buttons on the INTERFACE sheet, and view their results under the GRAPH sheet. Students should use error bars representing Standard Error (SE) for this activity.

Student Assignment

It is suggested that you observe the results of a number of different iterations of the simulation before proceeding as this will provide valuable
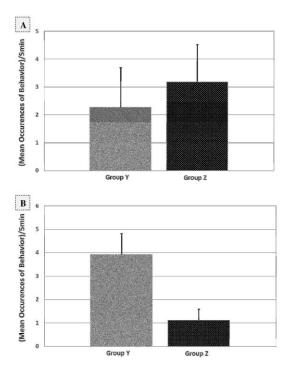
Figure 3. Sample graphs showing the average frequency of vocalization for all individuals in the presence (Group Y) and absence (Group Z) of a trainer based on a small sample of 11 (A) and then a moderate sample size of 25 (B). The error bars represent +/- 1 SE.

insight into the impact that a small sample size can have on your experimental results. Copy the resulting graphs from two different runs of the simulation, both using the same input values for your experimental question and a small sample size. In order to insert a graph into a Microsoft Word document follow the same process used in Activity 1, and again be sure to include a short, descriptive caption for each figure.

Now, return to the INTERFACE sheet and select a moderate sample size for both groups, but don't change any other experimental inputs. Once you have done this, use the generate buttons and view the graphical results. Insert two additional graphs into your Word file representing the results of your experiment using a moderate sample size.

Return to the INTERFACE sheet one last time and select the maximum sample size (1096) for both groups. Again, don't change any other inputs. Once you have done this, use the generate buttons and view the graphical results, and insert one additional graph into your Word file representing the results of your experiment using a larger sample size.

Why did you run only one simulation with the larger sample size? Hint: Rerun the simulation several times to help you with this.

Now, use the graphs to evaluate your hypothesis. Do the results using the small sample sizes support your hypothesis? The moderate sample sizes? The

large sample size? For each sample size, record your conclusion and justify/support your findings.

Are the results of your experiment statistically significant or not? How did you determine this? Hint: You should not have to do any additional calculations to come to a preliminary conclusion here.
What conclusions would the researchers on this project come to if they asked your experimental question? Explain your reasoning.
What insights did you gain regarding the impact of sample size on experimental outcomes? How will the results of this exercise influence your experimental design for future research projects?

DISCUSSION

The background information provided here can be shared with students in a discussion format or as a handout. If time and resources permit, students can also be encouraged to expand their knowledge of pinnipeds, and specifically SSLs, by doing their own literature search. A Google search for "Steller sea lions" will generate links to reputable websites (NOAA, The Marine Mammal Center, etc.) with additional, basic information on SSLs appropriate for most undergraduate students, while Google Scholar will provide links to published literature, although very few articles address the behavior of these animals.

The instructor can choose to structure this activity in many different configurations, using a single part of the exercise as a standalone activity or using the material in its entirety. In addition, Activity 2 can easily be expanded by requiring students to explore the impact of specific categories from the interface tab, such as sex, specific behaviors, etc. Students could work on this activity as collaborative research teams or individually, possibly even as out-of-class projects. Lastly, the exercise can be very interactive with discussion and assessment after each part of the activity. This would provide students with the benefit of learning more about pinnipeds, the development of hypotheses, and designing experiments.

In reviewing the outcomes of the students' tests of their hypotheses, emphasis should be placed on the process used and not on the accuracy of their predictions, as false hypotheses that are ultimately refuted by the evidence—so-called "negative results"—are as useful as correct hypothesis that are confirmed by the evidence, as long as the hypotheses lead to effective tests in each case. Students should be able to support both their predictions and conclusions with rational arguments and evidence, going beyond whether they were right or not. Some justifications that students might use to explain why their predictions are not confirmed by the outcomes of the simulation include:
- recognizing the high level of variability in the data due to the limited size of the data set and

somewhat sporadic sampling over four years, such that few significant differences would be expected;

- acknowledging that any particular random subsampling of data within the simulation could return an "outlier" result which may incorrectly support or reject the student's hypothesis, while realizing that this mimics some of the confounding aspects of data collection and experimental design;
- challenging the limited amount of data on SSL behaviors available in the published literature as a basis for making accurate predictions.

The simulation and data set described in this activity can be accessed for classroom use via the faculty webpage for W. Ryan at Kutztown University **http://faculty.kutztown.edu/ryan/.**

REFERENCES

BURKANOV, V., E. GURARIE, A. ALTUKHOV, E. MAMAEV, P. PERMYAKOV, A. TRUKHIN, J. WAITE, AND T. GELATT. 2011. Environmental and biological factors influencing maternal attendance patterns of Steller sea lions (*Eumetopias jubatus*) in Russia. *J. Mammalogy* 92(2):352 – 366.

CALKINS, D. G., MALLISTER, D. C., PITCHER, K. W., AND G. W. PENDLETON. 1999. Steller Sea Lion Status and Trend in Southeast Alaska: 1979-1997. *Marine Mammal Science* 15:462-477.

CUMMING, G., AND S. FINCH. 2005. Inference by Eye – Confidence Intervals and How to Read Pictures of Data. *American Psychologist* 60(2): 170-180.

DAVIES, C. E., KOVACS, K. M., AND C. LYDERSEN 2006. Development of Display Behavior in Young Captive Bearded Seals. *Marine Mammal Science* 22(4): 952–965.

DeMASTER, D. P., TRITES, A. W., CLAPHAM, P., MIZROCH, S., WADE, P., et al. 2006. The sequential megafaunal collapse hypothesis: Testing with existing data. *Progress in Oceanography* 68:329-342.

JEFFERSON, T. A., WEBBER, M. A., AND R. L. PITMAN. 2008. Marine Mammals of the World: A Comprehensive Guide to Their Identification. London: Academic Press.

MOTULSKY, H. 1995. The link between error bars and statistical significance. Accessed from https://egret.psychol.cam.ac.uk/statistics/local_copies_of_sources_Cardinal_and_Aitken_ANOVA/errorbars.htm on 18 November 2015.

MOULTON, V. MILLER, E., AND H. OCHOA ACUNA. 2000. Haulout behaviour of captive harp seals (Pagophilus groenlandicus): Incidence, seasonality, and relationships to weather. *Applied Animal Behaviour Sciences* 65(4):367-378.

SPEEGLE, J. 2013. NOAA removes the eastern Steller sea lion from the Endangered Species Act list. Accessed from http://alaskafisheries.noaa.gov/newsreleases/2013/easternssl102313.htm on 22 November 2015.

TRITES, ANDREW. 2012. *STELLER SEA LION BIOLOGY fast facts.* Accessed from http://www.marinemammal.org/biology/steller-sea-lion/ on 16 November 2015.