Using Simulated Surveys to Teach Statistics: A Preliminary Report

Marc Isaacson[1]
[1]Augsburg College, 2211 Riverside Avenue, Minneapolis, MN 55112

Abstract
One aspect of statistical literacy is the ability to read and interpret survey data presented in tabular and graphical formats. In addition, some may include the tasks of data collection and the creation of such tables and graphs under the auspices of statistical literacy. Unfortunately, properly collecting survey data may take a great deal of time. This paper presents a computer simulation program that allows students to select their topic of interest and the associated survey questions. Given student input, the simulation program generates a random sample of realistic survey data which allow students to do traditional data analysis and interpretation. Preliminary results are presented along with student feedback.

Key Words: Statistical Literacy, Survey sample, Simulated Survey, Student Project, Realistic Data

1. Background

The development of teaching materials related to statistical literacy has been occurring at Augsburg College for more than a decade and well-documented in the literature. See Schield (2004a and 2004b) for more details. Augsburg College has also been a leader in the movement to adopt quantitative reasoning as part of its liberal arts graduation requirements. In 2006, the college expanded its existing QR graduation skill to include a quantitative foundations designated course. In addition, students are desired to take a quantitative applications course within their chosen major. As a part of the designation for the quantitative application skills, these courses must include a student project applying the necessary quantitative tools. By the college definition, these Quantitative Application (QA) projects must include the following characteristics:

- **Student Generated Inquiry** – The project must revolve around some central question generated by the student. Thus, no two student projects are necessarily the same.

- **Build upon foundational quantitative skills** – As part of the general education program revisions, additional prerequisite courses were designated as Quantitative Foundations (QF). These courses must include course content which focuses on developing skills in 2 of the following fours areas in depth and cover a third area for breadth.
    1. Mathematical Relationships – graphical, symbolic and numerical representations; proportions, percents, estimation.
    2. Statistical Relationships – data analysis (including graphical analysis), elementary probability.
    3. Algebraic Relationships – modeling, functions, algebraic representations.
    4. Logical Analysis – deductive reasoning, fallacies, arguments, counter examples.

- **Significant portion of grade**: - To qualify for the QA designation, the student generated project must comprise a significant portion (typically >25%) of the course grade.

- **Written or Oral presentation of results** - Besides the obvious quantitative emphasis, students are also encouraged to integrate the written and oral communication skills acquired in their other general education courses with the quantitative work done in these QA courses.

In the Department of Business Administration, there are three different courses which qualify for the Quantitative Application (QA) designation. The traditional statistics course (BUS 379) is a natural course for a quantitative project. In addition, Augsburg began offering a course titled Statistical Literacy for Managers (BUS 264) in 2006. This course focuses on the statistical skills that managers need to be able to read and interpret statistics. This includes a mix of descriptive statistics, probability, exploratory data analysis and statistical literacy materials. In

addition, the department also provides support to a Statistical Literacy / Quantitative Reasoning course in General Studies. This course (GST 200) is designed for those liberal arts students who do not have a required QR course within their major; i.e. music, history, etc. For more information on the content / goals of these two statistical literacy courses, see Isaacson (2005).

2. Challenges

2.1 Varied Student Audiences

One of the issues in developing a QA project is the very distinct audiences within the three courses. Typically, the traditional statistics course (BUS 379) is required to be taken by the Accounting and Finance majors within the Business Department. The remaining majors (MIS, Marketing, and Management) are recommended to take the statistical literacy for managers course (BUS 264). Then, there are the liberal arts majors in the General Studies Statistical Literacy course (GST 200). All three groups of students have varying levels of comfort with quantitative materials. While most accounting and finance majors relish the idea of manipulating numbers, the other business majors and liberal arts students are typically less inclined to work with numbers. In addition, there are not only differences in students but also course formats. All three of the described courses are taught in both the traditional day school as well as our adult weekend college programs.

2.2 Course Content / Time Availability

One of the major challenges in the implementation of this new quantitative reasoning skills system is that it must be integrated into an already full curriculum. Since these courses existed previously, the addition of the quantitative applications project must require some modification of current course content. The goal is to create a project which can meet the desired goals and minimize the amount of material needed to be removed from the course.

2.3 Lack of Second Semester Course

Among other departments on campus, the QA projects are more easily integrated into the major through an existing two course research methods sequence. This certainly allows for more time to devote to building foundational quantitative skills while leaving extended periods of time to do a project in the second semester. Currently, the Business related majors have a number of course requirements which make it infeasible to add a second statistics course. This means that any QA project must be integrated into the existing one semester courses.

3. Proposed Solution – Survey Simulation

After considering a number of alternatives, the proposed solution to integrating a QA project into these three courses involved the use of simulation. With surveys and market research being central to many of the applications of statistics in business, a multi-part project involving surveys and simulation was developed.

The goal of the survey simulation project is to immerse students to the entire process of statistical practice in surveys. This includes all of the steps from the development and documentation of student research questions to the communication of statistical analysis in written or oral format.

In order to do this in a one semester introductory course, creativity was required to make this project a possibility. While survey data collection may have some experiential benefits, it is most certainly very time consuming. In order to maintain the project to fit within a single term, the collection of survey data has been automated through the use of computer simulation / random number generation. Through some guidance with student parameters, the computer simulates "realistic" data for the student to analyze. By going from writing survey questions to analyzing survey data, we are able to maintain the timeline to complete the project within the semester schedule. In addition, the use of simulated data has other benefits which will be discussed later in this paper.

4. Details of Survey Simulation Project

Designed to be integrated into a semester length statistics course, the survey simulation project was designed to be completed in 6 steps which could be spread across the duration of the course. Each of these steps provides a

milestone to ensure that students are making adequate progress towards completion of the project. In addition, these separate milestones allow an instructor to provide feedback to students on each component to verify that each step is completed and they are correctly prepared to work on the next task. In short, these project milestones include:

4.1 Write Proposal for Research Topic

In this step, students must articulate their topic of interest for the overall survey. In addition, they begin to document their initial thoughts regarding the populations of interest, hypothetical sampling methodologies, and possible important variables to gather in the survey. Since the survey data are generated by computer simulation, students have much greater latitude in their choice of survey populations.

4.2 Formulate Hypothetical Data Collection Plan

Once students have chosen their topic of interest, the next step is to put together a hypothetical data collection plan which includes a number of items. The most important item in this step is the actual writing of the survey questions. Students are asked to complete a template regarding survey questions similar to the one listed below:

Question	Actual Question	Question Type	Expected Value	Min	Max
One		Binomial		0	1
Two		Binomial		0	1
Three		Binomial		0	1
Four		Binomial		0	1
Five		5-point Likert or Multiple Choice		1	5
Six		5-point Likert or Multiple Choice		1	5
Seven		Continuous			
Eight		Continuous			

Figure 1: Survey Question Template for Hypothetical Data Collection Plan

In this template, students must write out their actual survey questions. Due to the use of computer simulated data and for purposes of standardization, students are asked to write out a total of 8 questions comprising of a mix of binomial, multiple choice, Likert scale or continuous questions.

In addition to the actual survey questions, the students must also provide some context for the computer random number generation program. As a result, students must provide some estimate of the expected value for the first six questions. For the last two questions where the answer to the survey question is a number, students must provide an estimate of expected value along with some definition of the minimum and maximum allowable values.

These point and range estimates serve two purposes in the project. First, these provide the computer with some guidance regarding the distribution of the data. This is important in generating random data which can be seen as realistic by the student. In addition, these estimates in component two can be used later in the project for the purposes of hypothesis testing. While they may not know what a hypothesis test is at the time they complete hypothetical data collection plan (component #2), they will have covered that material in most classes by the time that they get to the inferential statistics (component #5) section later in the course.

4.3 Simulate Survey Data

In this project component, the simulated survey data is generated given the inputs from the previous component. Typically, this has been done by the instructor as the data generation program was being refined. Currently, the random number generation is done by the use of a pre-programmed Excel spreadsheet which produces 125 survey results in a single iteration. Larger sample sizes can be easily obtained through multiple iterations. Through the use of VBA macros and spreadsheet formulas, student inputs are converted into a set of individualized randomly generated survey responses. Even if two students had the same topics, the same questions and the same distribution parameter estimates, they would have different data sets to analyze in the last portion of the project. While students provide some inputs on expected values and such, the data generation program is programmed to vary some of these slightly and create some relationships / correlations between variables. This forces students to actually investigate their data sets in exploratory data analysis (component 4) while also making it possible that their initial hypothesis may be rejected when doing inference (component 5).

4.4 Exploratory Data Analysis / Descriptive Statistics

Once the business students receive their unique detailed data file from the instructor, the task shifts to the application of traditional exploratory data analysis and descriptive statistics. Since we are using simulated "realistic" data, students must correctly "read" their results in order to prepare an accurate summary of the results in their written conclusions. As mentioned earlier, there is some built-in variation in the distributions produced such that students can't simply assume that their data will match their initial input values. Thus, this task provides students with the real-life experience of reading and interpretation of survey results, even if computer simulated.

In the current project, students are required to examine their "realistic" computer generated data in several different ways. With Augsburg business students taking a prerequisite course in spreadsheet problem solving, it is natural that this project would integrate the use of spreadsheet software to analyze their data. Summary descriptive statistics are required for each question / variable. In addition, they are to create graphical summaries of the responses to all eight questions in the survey. Also, students are required to investigate the inter-relationship between questions through the use of pivot tables, appropriate graphs, or both. Thus if a student has questions both on whether respondents are college graduates or not and their gender, the student should be able to summarize and discuss the observed differences (if any) between males and females in regards to college graduation rates. With eight questions and sixty-four possible combinations, it is impossible for students to examine / summarize all of these possible relationships given the time constraints of a one semester course. As a result, it is typically recommended that they select a minimum of 4 or more relationships between questions that might be most central to their topic of interest.

4.5 Inferential Statistics

Upon completion of the exploratory data analysis and descriptive statistics component, the business students can then proceed in the project to utilize their newly acquired knowledge of inferential statistics with their detailed project data. At a minimum, students report the confidence intervals associated with the results of their 8 survey questions. In addition, they have the information available to conduct a hypothesis test of their initial expected values against the actual obtained sample statistics. Lastly, by looking at combinations of questions they can again investigate difference between groups across questions. As in the previously mentioned example, they could test whether the observed differences in college graduation rates among males and females was statistically significant given the sample data and sample size. Again, there are many possible combinations of inferential statistics that an instructor may or may not want to emphasize. With the simulated survey data and the initial expected value statements, instructors can tailor this section of the project to fit the depth / topics of inferential statistics covered in their introductory course.

4.6 Written Final Report

The final and cumulative work of this semester long project is the final executive summary written by the student. This document is meant to pull together all of the work done on the project into one final document. In past courses, this component included a minimum 2-3 page executive summary which synthesizes everything from the topic of interest to the important findings and conclusions. I have also required students to write a 2-3 page reflection statement on their learning experience throughout the project. This forces students to think about where they have come from and the amount of material that they have integrated into this final project. In addition, students submit all of the previous 5 components as an appendix. This leads to a substantial document which averages around 15 to 20 pages.

Depending on the class size, instructor preferences, etc., the use of the simulated survey project could be used in other formats. With the individual nature of the topics, students could also be required to present their findings in an oral format. For large classes where this might be infeasible, another option would be for students to create a poster session to display / share their results. In addition, these projects could also be done as a group project if so desired.

5. Benefits of Simulated Survey Project

The simulated survey project has a number of benefits for use in the introductory statistics course. One of the first is that this simulated survey project allows for a greater level of student involvement. Since the student has a personal stake in the individually selected research topic, they develop a greater level of attachment to the project. In a course where student attitudes towards the course material are usually negative, this opportunity for greater student buy-in may help improve student attitudes. Currently, Augsburg is participating in a national survey of student attitudes towards statistics to help research this hypothesis. See Schau (2003) for more details on the assessment of student attitudes towards statistics.

Another obvious benefit to the simulated survey project is the time savings gained through the compression of the data collection cycle. Data collection is traditionally a very expensive task in either time or resources. In an effort to include this survey project experience into an already full introductory statistics course, this simulated survey keeps the focus of the student on the entire survey process rather than on the actual data collection. With the appropriate preparation of the hypothetical data collection plan, computer generated survey data can be generated in a matter of seconds.

One other benefit of using a simulated survey project is the expansion of possible topics / populations that students might choose to investigate. Since students are allowed to hypothetically sample from populations which might normally be too distant or difficult to obtain, their world of possible project topics is greatly expanded. If students were to collect actual data, many students would be forced to choose populations on campus from which to sample which may or may not be applicable for the topic they wish to investigate.

Besides opening up the possibilities of topics, the use of simulated surveys also avoids the default to bad sampling practice. While nearly all introductory statistics texts focus on the importance of getting a random and representative sample of your chosen population, this is often more difficult than imagined for most introductory students. If tasked with actual data collection, many students default to surveying their friends, family and others close to them. This convenience sample may generate the desired number of surveys but is far from anything resembling a random and representative sample. This use of a simulated survey project avoids that issue which might be later addressed in a later statistics or research methods course.

6. Preliminary Classroom Feedback

After having developed this survey simulation project over the last few semesters, there are several key findings which are important to consider. With the individual nature of the survey projects, there is an increased load on the instructor for reading / assessing these projects and their associated milestones. While they are all unique in content, the template format of the project and survey does aid in this matter. While this format has worked for relatively small classes (<25 students), it could be prohibitive in larger sections. With larger class sizes, it might be more feasible to do the project as a group activity.

In its current format as a written executive summary, this activity does highlight the need for better instruction of students in writing with numbers. As has been highlighted by the work done on the QUIRK program at Carleton College, students do not always write effectively when numbers should or could be involved. See Lutsky (2006). In this particular instance, the students have all of the relevant detail and summary data. Yet, many of them have a hard time transitioning from the numerical calculations / results to constructing well written English statements without resorting to simply listing out a bunch of statistics. While this is a larger issue that can't be solved within a single introductory statistics course, it is an important quantitative literacy skill that college students should possess. In the Augsburg BUS 264 and GST 200 courses, specific emphasis and training are provided in the use of numbers in writing descriptions and comparisons in ordinary English.

For students, the reflections statements included in the project are generally quite favorable. For many of them, this semester long project is an excellent opportunity to synthesize all of the materials covered in an introductory statistics course. While the details of any individual topic / chapter may seem a bit abstract at the time, the process of going from a research question to the written summary of an analyzed data set creates a framework for nearly all of the topics taught in a first statistics course. Here are excerpts of comments from three separate students who completed the project:

> "The survey process was good practice and brought together what we had learned throughout the semester."
>
> "I feel like the course project gave me a great idea on the processes of a survey and converting it into meaningful results. Although I didn't get to see some of the real results I would've liked to see concerning my survey, I thought it was a great way to learn without going through the hassle and added time of conducting a real life survey."
>
> "The course project was also a good learning tool for a variety of reasons. I feel this project allowed me to analyze and interpret statistical results based on a topic I was interested in. It is one thing to do homework assignments on examples of business situations but is a whole different thing to allow someone to pick a topic they are interested in. Analyzing and interpreting the statistical results was a lot more fun with the topic I chose. I also feel the project allowed me to become for comfortable with Microsoft Excel."

7. Conclusions / Further Recommendations

This simulated survey project has met many of the proposed goals within the Department of Business Administration at Augsburg. On one level, it has been successful in achieving the QA designation at Augsburg for students majoring in the various business disciplines. In addition, it also provides a meaningful, motivational and realistic experience for students in introductory statistics courses.

Given its successes, there are certainly opportunities for future improvement including:

7.1 Summary Data Generation Option

While the current simulated survey project has been successful in Augsburg course with an emphasis on data manipulation / calculations, the GST 200 course has a very different audience of liberal arts majors. With a combination of weaker mathematical skills and no prerequisite computer courses, the current project as written would be difficult to implement. One natural modification of the project is to focus on summary rather than detail data. Given that the current data is generated in Microsoft Excel, this integration of random data generation and statistical summarization could be achieved with some additional programming.

7.2 Improved Data Realism

While the current version of the survey simulation data generator has been adequate, there are several possible modifications which could improve the realism of the computer generated data. One of these possible improvements would be better definition / integration of the known correlations between variables. The current program does some attempt at correlations but is far from a fully-defined correlation matrix. In addition, a more flexible template would allow students to mix and match the various question types and allow for more than the current 8 questions in the template.

References:

Isaacson, Marc (2005) Statistical Literacy – An Online Course at Capella University. *ASA Proceedings of the Section of Statistical Education*, Pg 2244-2252. See www.StatLit.org/pdf/2005IsaacsonASA.pdf.

Lutsky, Neil (2006) Quirks of Rhetoric: A Quantitative Analysis of Quantitative Reasoning in Student Writing. *ASA Proceedings of the Section of Statistical Education*, Pg. 2319-2322 See www.StatLit.org/pdf/2006LutskyASA.pdf.

Schau, Candace (2003) Student's Attitudes: The "Other" Important Outcome in Statistics Education. Pg 3673-3681. *ASA Proceedings of the Section of Statistical Education.* See www.StatLit.org/pdf/2003ASASchau.pdf.

Schield, Milo (2004a) Statistical Literacy Curriculum Design. *2004 IASE Curriculum Design Roundtable*. See www.StatLit.org/pdf/2004SchieldIASE.pdf.

Schield, Milo (2004b) Statistical Literacy and Liberal Education at Augsburg College. 2004 Summer Issue of *Peer Review*. See www.StatLit.org/pdf/2004SchieldAACU.pdf.

Acknowledgements

This paper was presented by Marc Isaacson at the 2008 ASA JSM. This paper is part of the W.M. Keck Statistical Literacy project: a project "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts." The author is available at isaacson@augsburg.edu.