Paper presented at the ARTIST Roundtable Conference on Assessment in Statistics
held at Lawrence University, August 1-4, 2004

Class Data Collection and Analysis - Assessment

Nathan Wetzel, University of Wisconsin - Stevens Point
Department of Mathematics and Computing, UW-Stevens Point, Stevens Point, WI 54481

Key Words: Statistics Education, Assessment

Every semester, my sections of Elementary Statistics conduct a statistical investigation from start to finish. Typically, I have an assignment that introduces the topic and asks students to design the data collection. We collect the data, and each student writes a report, which answers a hypothetical client's questions and addresses further research questions and issues. In this paper, I will outline the reasons for doing a class project like this, how the details work out and how to I assess the students' work. I will also report a small sample of student responses.

Reasons for a Class Data Collection and Analysis

For an introductory class, I define statistics to be the art of collecting, analyzing, interpreting and presenting data. Conducting a class data collection and analysis incorporates all four parts of this definition and models the "real" world. In the "real" world, a client poses a question, the client and the statistician work together to design the data collection technique or protocol. The data is then collected and analyzed. Often, the statistician and the client work together to interpret the results in the client's context. Finally a report is written to summarizes the results.

The typical introductory statistics book breaks this process down into discrete parts that are easily understood. The main purpose of a class data collection and analysis is to put all of the parts together. My students also have mentioned additional benefits. (1) Since the data is **their** data, the students have more interest in the results. (2) Since the students collect the data, they gain an appreciation for the difficulty and expense in collecting good data.

How I organize a Class Data Collection and Analysis

Each semester, I choose and topic and I play the role of an imaginary client who asks the class to answer a question or questions. Most textbooks discuss the design of experiments and surveys early in the book. While we are covering that material, I have an assignment that asks "How should we collect the data?" An example is included below.

(*10 points*) We want to design a study of fast food. We have an imaginary client who has asked us to get answers to the following questions.

- How many ounces are in the average McDonald's Large French Fry? Small French Fry?
- How often does the average UWSP student eat fast food? Is there a difference between men and women?

If you have any questions for our client, Professor Wetzel will play the role of the client.

Describe a practical, reasonable, usable way for our class to get data that will help in our investigation. **DO NOT** collect any data. Your description should include:

- a restatement and clarification of the investigation.
- For the French fry part, what measurements(s) to make and how to make them. (we need a way to get the data which allows everyone to make some measurements - I agree that ideally, we would have the same person make all of the measurements, but I want everyone measuring.)
- For the fast food part what questions would you ask students? Include specifics on wording and format.
- at least one additional demographic question (like gender) for which we could compare fast food use.
- how to get random samples.
- any other data that we might want to collect.

We will have 50 students helping collect the data. We will distribute the work equally among these 50. Also, this data collection will cost us all a small amount of money. Arrange your data collection so that we use no more than 100 orders of French fries.

This problem requires thinking and possibly some research. A good answer will get 7 out of 10 points, in order to get full credit, you need to provide a **great** answer.
We will discuss your answers in class and later we will collect data using our consensus answer.
An Example of a bad answer to this question: (this answer would get 2 points - minimal restatement, no specifics, lacks explanation, no extras)
"We are investigating French fries and fast food. I would have everyone go to McDonalds and buy a order of French fries. Then we would weigh it. We would also ask students how often they go to a fast food restaurant."

Rubric for grading this Problem

(*5 points*) Basics: Does the description include all of the required parts including accurate and appropriate use of terminology?

- restatement
- measurements - including specifics
- questions to ask - including specifics
- extra demographic question
- how to get randomize - including specifics
- other data

(*2 points*) Organization: Is the description organized and neatly presented?

Great answers also include some pluses.

Pluses: Does the description include any significant extras?

- Significant improvements to a basic data collection design.
- Extra thought into the specifics of this context - recognizing potential problems and giving solutions.

Minuses:

- Does the description indicate that the student is mimicking a book answer and not considering the context?
- Does the description include a design that would be extremely impractical?
- Does the randomization described introduce a significant confounding variable that was not identified?

In the Fall of 2002, the main question was to compare the mail delivery times for campus and postal mail.

Examples of Pluses: An example of a plus was the response that indicated that we should consider the fact that campus mail is only delivered on weekdays and postal mail is delivered on Saturdays when measuring how long a letter took to reach its destination. Another plus involved finding out that selected Residence Halls are delivery and pickup points for campus mail.

Examples of Minuses: An example of a minus was a response that indicated that we should send letters to random people without considering the difficulty of having this random person tell us when the letter arrived. Another example of a minus was a response that said we shouldn't tell the person mailing the letter whether it was being mailed in campus mail or postal mail. The problem in this case is that the person mailing the letter must put the letter in the correct mail box and campus mail does not need a stamp.

There are five key aspects to this assignment. First, it is very open-ended and there are lots of good answers. This is intentional and although it is difficult both for the students and the instructor, it is fairly realistic. Second, the data that will be collected will have both a quantitative and categorical components. When the analysis is done, this means that they will need to study more than one chapter of the text. Thirdly, there is a pretty clear basic data collection method that could be used, but there are also clear extras that should be included. Fourth, although the grading has a subjective component, I have been fairly clear about the expectations. The rubric for this assignment details the pluses and minuses and gives some of the good and bad answers from previous semesters. Finally, this is only about 1.5% of the grade. It is valuable enough, but not "high-stakes."

After this assignment is collected, I score their work. The day that I hand back the graded work, I plan a 15 - 20 minute discussion of the great ideas. I talk through most of the great ideas and extras that students had written on their papers. This serves two purposes. First, it serves as a review of some of the themes from the chapters that we had just covered and helps them match the generalities with specifics. And second, it helps convince those that got a "good" score of 7 out of 10, that there were answers that were much better than theirs. In the section on Student Reaction, I will give some details on how students felt about their scores.

Next, I decide how we are really going to collect the data. I make some decisions that reduce the cost, that make sure that all of the students are involved in the data collection, and that make it more practical. I plan 30 minutes of another class period to discuss my decisions and to decide the details on the other issues. For example, in the McDonalds and fast food example, I decided that we would only visit *one* of the local McDonalds. I also decided that our definition of fast food would not included pizza restaurants. As a class, we decided the specific wording of the survey questions, the time and dates of the survey and visits to McDonalds, etc.

I compile and enter the data into Minitab. I do my best to correctly enter the data as it is on the surveys or data sheets. Almost always there are outliers and I have a short assignment to find the outliers and decide whether to omit them in the analysis, keep them, or correct them. In the McDonalds example, we were supposed to measure in grams, however, some measured in ounces. We converted these measurements. One measurement of the mass of the empty bag was four times larger than the others. In this case, we omitted this data.

Finally, I have a final class assignment that has two parts. The first part gives step by step instructions for answering very specific data analytic questions. For example, "Find a 95% confidence interval for the average mass of a small order of French fries. McDonalds claims that the serving size is 71 grams. Is this consistent with our data?" Or, "Use a hypothesis test to test whether there is a difference between fries ordered in the evening or in the afternoon." In this part of the assignment I want them to provide the Minitab output and/or the calculations for the problem. The second part of this assignment is to write a report to the client. Here is an example.

> (*22 points*) Analyze all interesting portions of the data and write a short report (at least 2 pages, double spaced, typed, not including graphs) to our client summarizing our data collection, your analysis and your interpretation. Write the report to a client who understands p-values, but is not interested in the details of the computations. Your client understands all of the graphical methods that we have looked at. In other words, don't include MINITAB hypothesis test output, but do include graphical summaries.
> In addition, include a short letter that we could send to McDonalds.
> Notice that this part is the main part of the assignment. 'Good' answers will receive 14 points, you must do a great job to receive full credit. In other words, I want YOU to ANALYZE the data. This part is open ended and in order to do a great job, you will need to look at MINITAB output that I did not explicitly tell you to get. You will also need to THINK AND CARE about the data. Answer the questions that the client should have asked or that you saw interesting results,

Again, for this assignment, I provide examples of good and great written reports.

Finally, every semester, I change the assignment slightly to focus on a different topic. I have used surveys about the Atomic bomb, campus parking, cell phone use, credit cards use, and Y2K worries. We have gathered data on pulse rates, backpack weights, volume in Coke products, readability of Time magazine, chocolate chip cookie weights, time given in campus parking meters, monthly phone bills, and the time for a letter to reach its destination.

Assessment

I have always believed in having students write in mathematics and statistics classes. When a student writes a mathematical proof, I tend to grade on the correctness of the mathematics and the logic. I can justify taking off points for mathematical errors, gaps in logic, etc. However, when I assigned "report" writing, there are often no mathematical errors, nor gaps in logic, etc. Yet, I could look at the papers and say that some were clearly better than others. I struggled for a long time looking for a grading system that is perceived as being fair, that is fair, that would give higher scores for better papers, yet would not be too time consuming. The assessment system that I use has four parts that are key to achieving these goals.

1. *The students know from the beginning that this problem will be assessed differently.* I have found that many are very satisfied to get a score of 70% because they knew that their solution was average.
2. *The students have access to solution to similar problems.* Every semester, I choose a different topic, so solutions from past semesters do not "give it all away." The sample "great" solution is fantastic, so it sets a high standard. This helps with the students who want to do exactly enough work to get an "A" and no more.
3. *The assessment is done and the problem is handed back in a short time period.* This helps keep the topic current for the students.
4. *The students who did not receive full credit have a chance to see what kinds of solution did receive full credit.* The timing of this often serves as a review of the chapter, but it also automatically answers most of the questions that the "point hungry" students ask and encourages more questions from the "knowledge hungry" students.

My grading system is not complicated. If the solution is a basic solution, the student receives about 70%. I receive lots of basic solutions. For example, in the 10 point assignment shown on the first page, many students wrote a solution that essentially said that we should send students to McDonalds and order a Small or Large order of fries and then find the mass. They often had some aspect of randomization in the design. Either to go at random times or randomly order a Small or a Large, or to randomly choose one of the three local McDonalds. I considered this to be a basic solution for this part of the problem. Many went on to included **extras** in their solution. Some gave specific details on the measurement. For example, "to keep the scale clean, we should get the fries 'to go' and put the whole bag on the scale and find the mass. Next, we should empty out the fries and find the mass of the empty bag." Others, specified whether we got the fries salted or not. One student found the times of the shift changes at McDonalds and indicated that we should get some orders for each shift. Some got the nutritional information from McDonalds and found what the serving size is supposed to be for a Small and a Large order.

For the survey portion of this assignment, the basic answer needed to describe a reasonably representative sample of our student body as well as specific wording in a survey. Extras included a formal definition of what counts as fast food and careful wording of the questions.

In this case, the whole description was worth 10 points. A basic solution for both parts received 7 points. One or two extras received 8 points, three of four extras got 8 points, and

five or more extras got 10 points.

As I grade, I put a "B" in the margin for the basic parts, I put a "+" for any extras and a "−" for any portions of the solution that are minuses. A student receives a minus for mimicking a book answer and not considering the context, for a design that would be extremely impractical, or for a randomization that introduces a significant confounding variable that was not identified. A specific example of a minus is the responses that indicated that we should survey students *at fast food restaurants* to ask them how often they go to a fast food restaurants.

Student Reaction

During the Spring of 2004, I solicited response from my students regarding the assignment that had them design the study of fast food (from the first page of this paper). I asked three questions

1. When you were working on the problem, did you feel that you knew what the expectations were for a 10 point score? If not, explain what may have helped your understanding? (Results: 39 out of 45 said yes.)
2. Were you surprised by the score that you received on the problem? Please explain why or why not. (Results: 24 out of 40 said no.)
3. Did the short class discussion about this problem help you understand this material? Please explain. (Results: 40 out of 44 said yes.)

In general, the students reacted favorably to this assessment. One student, responded the he/she was not surprised, *"I included more than required, but I didn't go as above and beyond as possible, so my score of 8 is appropriate."* He/she went on in the third question to indicate the the class discussion helped her/him realize *"different aspects important to the study that I had not given thought to."* Another student said *"I was not surprised by my score. I knew what was expected for a 10 and I completed those expectations."* Some were surprised by the score. Most of these response indicated that they did learn from the experience. For example, one student said, *"I was surprised, I thought I would have had a higher score because I had detail and extras but was o.k. because some other people had some more good things I didn't think of."* This student went on in the third question to indicate that the class discussion *"made me understand that there is a lot to think about and plan to ensure variables don't effect [sic] what it is we want to test."*

Final Comments

First, I have included the actual assignments, data and results at `www.uwsp.edu/math/wetzel/Artist`

Second, a recent draft of the Guideline for Assessment and Instruction in Statistics Education (GAISE) report has six recommendations. The Data Collection and Analysis (DCA) class projects that I have described fit well into this framework. The first recommendation is to *Emphasize statistical literacy and develop statistical thinking* and the texts gives an analogy to carpentry.

> In week 1 of the carpentry (statistics) course we learned to use various kinds of planes (summary statistics). In week 2 we learned to use different kinds of saws (graphs). Then we learned about using hammers (confidence intervals). Later we learned about the characteristics of different types of wood (tests). By the end of the course we had covered many aspects of carpentry (statistics). But I wanted to learn how to build a table (collect and analyze data to answer a question) and I never learned how to do that.

The DCA project that I have described is "building the table." The other recommendations of the GAISE draft are:

2. *Use real data.* Not only does the DCA project use real data, it will provide new sets of real data for your future classes.
3. *Stress conceptual understanding rather than mere knowledge of procedures.* In the DCA project, we use the computer for the calculations. In addition, even before you collect the data you can refer to the data that you will collect to motivate certain aspects of important concepts.
4. *Foster active learning in the classroom.* Although the DCA project is not in the classroom, it is certainly active learning, especially with regard to the difficulty of collecting good data.
5. *Use technology for developing concepts and analyzing data.* See point 3 above.
6. *Use assessments to improve and evaluate student learning.* To be honest, until Spring 2004, I did not think about how critical the after assessment discussion was to the learning process. However, as I described in the Student Reaction section, many students did not learn how much thought should go into the data collection design until *after* the assignment was done and assessed.

In summary, the DCA project described is consistent with GAISE. The assessment of the student work can be done fairly, consistently and quickly. The careful reader may have noticed that I consistently refer to this idea as being helpful for *me* and *my students*. Although I believe that it will be helpful for many other educators and students, I do not believe that it will be better for everyone. This project takes class time that could be spent on other issues. This project takes extra work for the instructor and the students. An instructor who implements this *must* be excited about the particulars of the data collection. If not, it will fail miserably. In other words, if you want to be a "great" educator, you will need to go above and beyond the basics.