Scientific Research Publishing

Improving Performance and Enhancing Introductory Statistics Using Projects

Joy D'Andrea[1], Rebecca Wooten[2]

[1]University of Florida, Sarasota, USA
[2]New College of Florida, Sarasota, USA
Email: jdandrea@mail.usf.edu
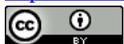
Abstract

Introductory Statistics is a course taught in various community colleges, state colleges, and universities. Implementation of projects in this course has been shown to enhance students' learning; in addition to increasing the ability of educators to assess students' learning outcomes in detail. These projects are often inquiry-based and require a balance between flexibility and efficiency. Maintaining balance has provided the students' opportunities through exploration and learner autonomy. In exploration, a student seeks new methods and options through experimenting. In this paper, we will discuss the benefits of exploration required by projects given in an introductory statistics course. The details of five different projects discussed in this paper—illustrate the practical influence that they could have on higher statistics courses.

1. Types of Teaching

There are two main types of teaching: teacher-centered and student-centered. In the teacher-centered approach, the professor gives direct instructions as the formal authority on the subject matter; this is what you find on many college courses. Student-centered teaching is inquiry-based where the instructor acts as a facilitator and delegate; this includes forms of cooperative learning [1]. Both methods use personal modeling, leading by examples and allowing students to learn by observing and mimicking the outlined process.

Leading by examples is extremely useful in introductory statistics as many students taking these courses are non-mathematically inclined; however, as fu-

ture researchers, having working knowledge of statistical methods and applications is imperative. This is where projects that incorporate student participation are extremely effective. Projects such as creating a dictionary of terms not only exposes students to the first steps of the scientific method where information is gather, but also provides students with the terms necessary to be statistically literate. Beyond the subject of statistics, this project also introduces many students to the basics of word processing without computer experience.

2. Project Enhancements

There are also projects that require student to formulate their own research questions; allow students to work through the process or gather real world data that is of interest to the student; organize why, what and how they gather their own data; organize the resulting information; sort and tally into frequency tables and other descriptive graphics, and create summaries as the start to exploratory data analysis.

For the first project, a dictionary of statistical terms, students are given an un-alphabetized list of 100+ terms of which 40 terms are flagged and must be included among the 50 terms required. This project requires that students be able to extract information from a PDF file; sort and sift out the necessary information; use a word processor to organize the information; present the definitions in alphabetical order and enumerated with cover page, pagination and citations [2].

Another project is the Data Collection and Frequency Organization project. This project emphasizes that the student(s) will understand data collection and construction of organizing the frequency of the data. The students are given a list or a set of objects to experiment in this project. There are two parts of this project: Data Collection and Frequency Organization. The first part of this project requires each student to formulate the idea of why, what, and how they gather their own data; sort and tally their data. The second part of this project requires that each student creates a frequency table of their results (from part one). They are required to organize their data and produce graphical illustrations of their results. Finally, they have to summarize their results and draw any valid conclusions about their data.

The projects progressive become more involved. The M & M Experiments are perfect examples of hands on projects that require student to gather the information, that is, counting their M & Ms before eating them, organize the data, summarize the information obtained, standardize the information, measure the likelihood when compared to the company's claim(s), interpret the results and draw a conclusion. The first experiment outlines the step required for hypothesis testing of proportions and the second covers goodness-of-fit.

On the more serious side, projects that require real world data analysis also requires students to follow the scientific method with statistical rigor, extend the fundamental ideas covered in the M & M experiments and extend the analysis to include an understanding of the collection methods, classification of the variable measures, multiple graphic illustrations of the data, detailed statistical analysis

and performance of hypothesis tests before presenting precise interpretations and valid conclusions [2]. This project uses simplistic real world data, the next outlined project takes the ideas of the scientific method and statistical analysis to the heart of the student's major.

The Hypothesis Construction and Testing project encompasses students to answer a question regarding data that is in their area of interest, for instance, if a student is interested in knowing if there is a difference between the averageamount of coffee people drink. This project varies in levels of novice research statements and advanced research statements. The importance of projects at all levels is the understanding of the subject of introductory statistics and the critical thought skills that the student(s) will gain by the end of the project. For this project, students will create their own research statement of their interest. A short introduction of the research statement is required. They can use a single sample mean or a single proportion. The testing of the research statement will require five parts:

- State the proposed research question (include the level of significance).
- State the hypotheses (null and alternative).
- Determine whether usage of a z-test or t-test is appropriate for 1-sample mean or a 1-sample proportion.
- Calculate the test statistic and determining the p-value.
- Determine whether we reject the null hypothesis or fail to reject the null hypothesis (state why) and interpret results (give a concluding statement).

A similar project is the full research project which requires students to give a full report on the data set of interest to them. The report is a culmination of all the skills they have acquired doing the previous projects. They must present the write up with a cover page and pagination; they must include the data source and measures included; they are asked to identify the measured variables as response or explanatory, continuous or discrete, numerical or categorical, and the level of measure; they are to include descriptive graphics to help verify the data and obtain an overall understanding of the data's variability, detect ability, shape and tendencies including discussions on lurking information, correlations and confounding variables. Project report must include the basic descriptive statistics, and data analysis that includes but is not limited to hypothesis testing of means, proportions, differences of means, mean differences, difference in proportions, analysis of variance, and regression [3].

Students understanding of the subject material is easier to gauge using projects over homework which can be a short question taken from the text or an in-class exam that is limited by time and computational tools available to students in a standard test environment. Students' reactions may be mixed; however, the majority of students surveyed believe it is a great idea. Students understand the importance of understanding how statistics work as well as the importance of learning technology to aid in the computations required in statistics and increase the value of taking introductory statistics courses. Many students appreciate the time allotted to projects as they are learning technology as well.

3. Methods of Learning

There are four main ways that students learn; there are visual learners who prefer images, maps and other descriptive graphics; auditory learners who can gain a better understanding of the subject material by listening and conversing about the subject matter in forums such as lectures or group discusses. There are students who learn the best from reading and writing. These avid readers take copious notes and are able to creating well organized essays on abstract concepts, and finally there are kinesthetic learners who need hand-on experience before coming to a full understanding of the subject materials, problems solvers that must work through the problem at hand, figuring out the process by trial and error [1]. With these projects, the four learning modes are all incorporated.

Statistics is the art of decision making in the presence of uncertainty (Chris P. Tsokos) and it is the interdisciplinary link between mathematics and all other fields of study (Rebecca D. Wooten). As Math is a language, in order to be fluent in this language, a student must be able to read, write and speak the language of math and understand count techniques introduced in Set Theory. This will help the student to determine the appropriate statistical summaries, understand probability distributions used to measure the likelihood, and interpret the conclusions.

Considering the pros and cons of projects, the pros outweigh the cons in that when executed properly, projects help improve the performance of students in introductory statistics course. Three cons to assigning projects are 1) students may cheat, 2) projects take time to evaluate and grade, and 3) students lack of knowledge with respect to statistical software. However, while cheating on a project may give the student a good grade on that assignment worth 5%-10%, it does not prepare them for the exams worth 15%-25%. Moreover, while some projects take time to grade, projects provide teachers a more detailed understanding of the students statistical knowledge can be assessed. Lastly, while students may not know the statistical software required such as Excel or SPSS, the time allocated for projects allows students' to explore the data and obtain working knowledge of the statistical computer software.

4. Statistical Analysis Using Real World Data

Considering 344 students in Introduction to Statistics I course taught by Dr. Wooten in fall 2015 and spring 2016, there were twelve homework assignments, eight chapter assignments, three projects, three exams and a cumulative final exam. The three projects designed by Dr. Wooten include the Dictionary, the M & M Experiment and the detailed Research Project, Figure 1.

Let $h_1, h_2, \cdots, h_{12}$ be the percent scores on the graded homework assignments (graded by Canvas for consistency) and H be the average of these scores; let $c_1, c_2, \cdots, c_8$ be the eight chapter review assignments (graded by Canvas as students were allowed multiple attempts keeping the high score) and C be the
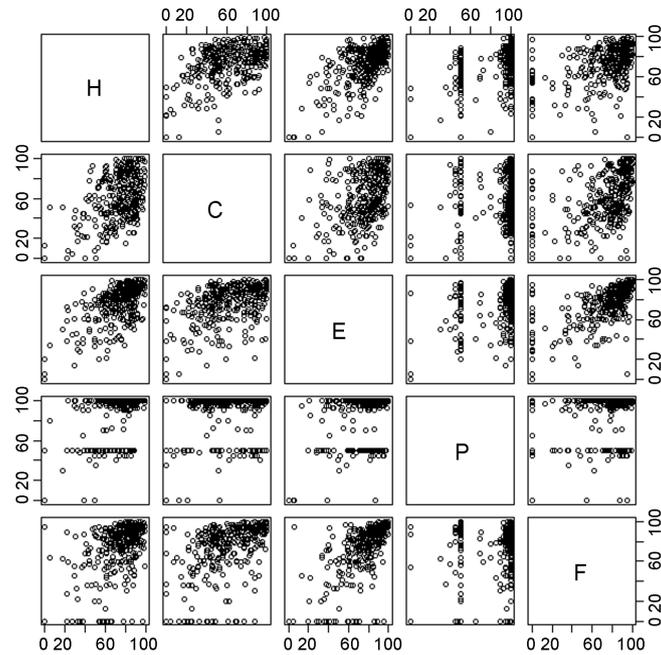
Figure 1. Pairwise comparisons of homework (H), chapter reviews (C), exams (E), projects (P) and final (F).

average of these assignments after conversion to a percent score; let p_1, p_2, p_3 be the percent scores on the projects and P be the average of the best two out of three scores; let e_1, e_2, e_3 be the percent score on the three midterm exams and E be the average of the best two out of three test scores; and finally, let F be the percent score earned on the two-hour cumulative final exam where the projects and exams were graded by TAs using a detailed rubric.

Analysis of the full interactive model (1) indicates the most significantly contributing factor is that of the projects, followed by the interaction between the chapter review assignments and projects as outlined in the text *The Joy of Statistics* (Tsokos, C. P., & Wooten, R. D. [4] [5] [6]).

$$F = \beta_0 + \beta_1 H + \beta_2 C + \beta_3 P + \beta_4 E + \beta_5 HC + \beta_6 HP + \beta_7 CP + \beta_8 HE$$
$$+ \beta_9 CE + \beta_{10} PE + \beta_{11} HCP + \beta_{12} CHE + \beta_{13} HPE + \beta_{14} CPE + \beta_{15} HCPE \tag{1}$$

Other contributing terms include (in order of significance) the average exam grade, the interaction between the projects, chapter review assignments and the exam grades, and lastly, the chapter review assignments (Table 1). Homework was found to be insignificant; that is, all things consider that homework does not significantly affect the score students earn on the final. In fact, the projects were found to be a significant indicator of a student's overall understanding the subject matter. As illustrated in the box plots in Figure 2, the average percent score for each assessment trends downward as the material increases in difficulty. There is a slight increase in homework and chapter reviews after the second exam; and while graded on a rougher scale, the first two project grades are high with best two out of three averaging to 88.72%. The scales for grading homework assignments and chapter reviews varied and were graded to the point; therefore, assignments worth 50 points defined a scale accurate to 2%. Exams (including the

Table 1. Parameter estimates, standard errors, test values and significance of terms included in the full interactive model.

Coefficients:	Estimate	Std. Error	T-value	Pr(T > |t|)	Significance
(*Intercept*)	34.17269	12.04068	2.838	0.004821	**
H	−0.1231	0.64847	−0.19	0.849553	
C	−1.68859	0.91179	−1.852	0.064932	.
P	−192.574	57.97787	−3.322	0.000996	***
E	135.9642	65.6364	2.071	0.039096	*
HC	0.01696	0.01532	1.107	0.269184	
HP	1.66046	1.19236	1.393	0.164688	
CP	5.1844	1.9463	2.664	0.00811	**
HE	−0.96325	1.25851	−0.765	0.444592	
CE	1.64553	1.74168	0.945	0.34546	
PE	174.7754	123.3161	1.417	0.157346	
HCP	−0.04253	0.02833	−1.502	0.13418	
HCE	−0.01239	0.02457	-0.504	0.614529	
HPE	−1.21683	1.98209	−0.614	0.539699	
CPE	−6.4218	3.22103	−1.994	0.047011	*
HCPE	0.05178	0.04151	1.247	0.213134	

Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

Residual standard error: 18.35 *on* 328 *degrees of freedom*

Multiple R-squared: 0.5001, *Adjusted R-squared:* 0.4772

F-statistic: 21.87 *on* 15 *and* 328 *DF, p-value:* <2.2e-16

final) were 200 points graded to the point (scale of 0.5%). However, projects were 100 points with most points worth 5 - 10 points, all or nothing making the scalar rougher (between 5% - 10%).

In the reduced model, Equation (2), 46.32% of the variance in the grade on the final exam explained by the chapter review assignments (which students were allowed to do multiple times for high score), projects, exams, the interaction between chapter review assignments and projects, and the second order interaction term between chapter review assignments, projects and exam grades, Table 2.

$$F = \beta_0 + \beta_1 C + \beta_2 P + \beta_3 E + \beta_4 CP + \beta_5 CE + \beta_6 P + \beta_7 CPE \qquad (2)$$

The outlined analysis is similar to cluster analysis which indicates using Ward's distance the ranking clusters based on Euclidean distance are Projects (AU/BP value of 100/100), Chapter Assignments (80/69), Homework (83/60) and Exams (53/59), ending with the final. In principle component analysis, each eigenvalue is approximately one; however, it is interesting to know that in the bi-plot of the first two components, the projects pull in nearly orthogonal to the chapter reviews and the other measures are more in-line with the final. That is, in the projected space, the order and degree of separation appear as Projects (90), Homework (60), Exams (40), Final (20) and finally the Chapter Review (0). All the analysis outlined (Regression, Cluster Analysis and Principle Component Analysis) are often used by Dr. Wooten and Dr. D'Andrea in their research (environmental, medical and educational).
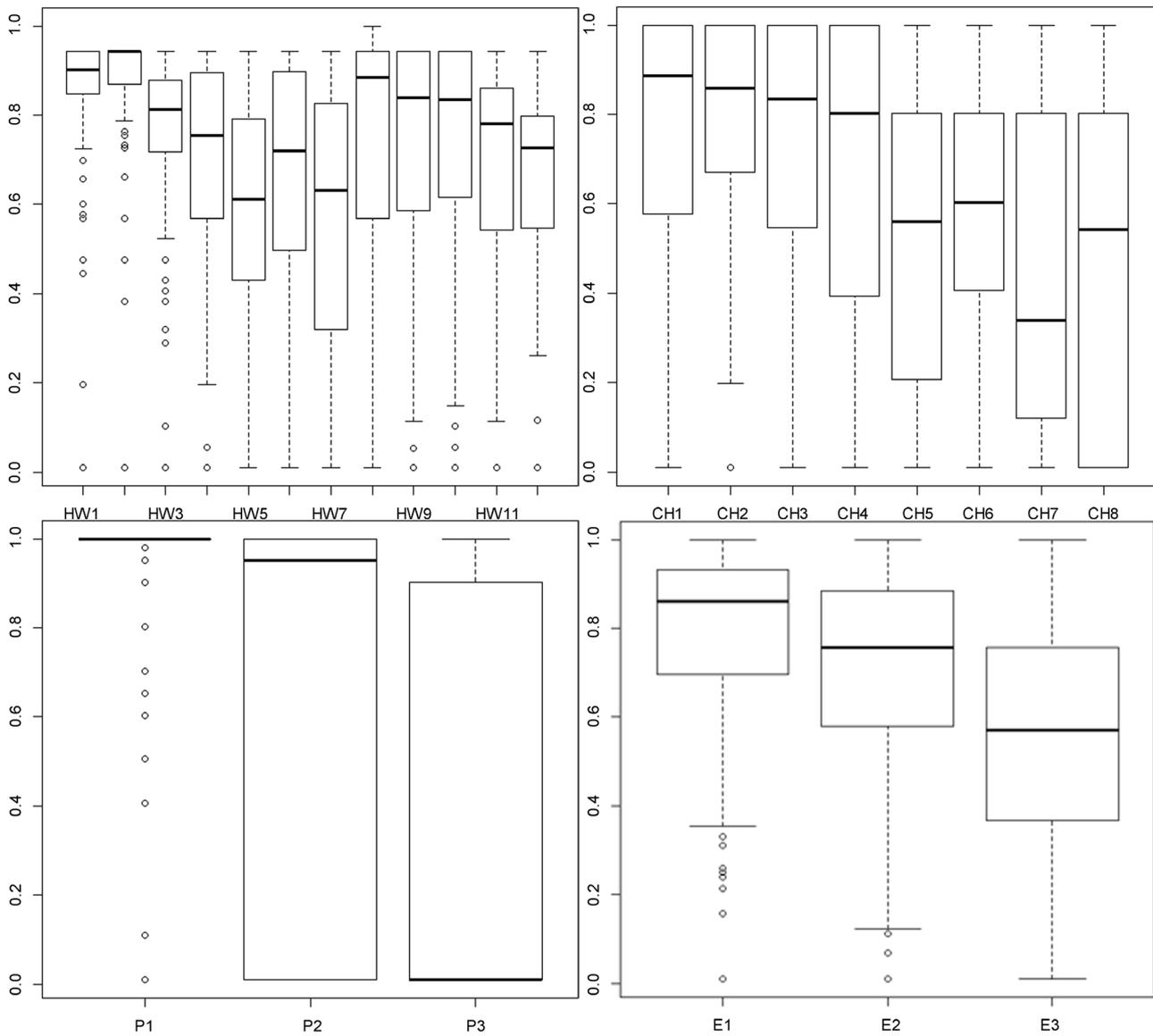
Figure 2. Box plots of the homework, chapter review assignments, projects and exams.

Table 2. Parameter estimates, standard errors, test values and significance of terms included in the full interactive model.

Coefficients:	Estimate	Std. Error	T-value	Pr(T > |t|)	Significance
(Intercept)	4.936	7.751	0.637	0.52470	
C	−0.2277	0.1642	−1.387	0.16644	
P	−0.3182	0.1207	−2.637	0.00876	**
E	1.089	0.1387	7.852	7.852 5.49e-14	***
CP	0.0168	0.003211	3.325	0.00098	***
CPE	−0.00006951	0.00002352	−2.956	0.00334	**

Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

Residual standard error: 18.73 on 338 degrees of freedom

Multiple R-squared: 0.4632, Adjusted R-squared: 0.4552

F-statistic: 58.32 on 5 and 338 DF, p-value: <2.2e-16

5. Future Research Projects

Dr. D'Andrea and Dr. Wooten are originally Mathematicians with degrees emphasizing Applied Statistics; and they have worked with the Educational departments at USF in Sarasota and USF main campus to develop courses in mathematics and statistics for middle school and high school educators. Dr. D'Andrea and Dr. Wooten are working on a text book tentatively titled The Core of Mathematics in Education K-12 which will include hands on projects designed to teach students the art and language of Mathematics.

To address the issue of understanding computer applications in general, Dr. Wooten is working on a text book entitled *Using Technology: Practical Applications of Computers* which uses projects to teach computers skills such as word processing, graphics, and spreadsheets to middle school and high school students. Dr. Wooten is working on the Gamification of Statistics, a 3-dimensional virtual world that teaches statistical concepts-data collection, probability, and hypothesis testing (parametric). This game allows students to earn badges upon the completion of projects.

6. Conclusion

In conclusion, projects not only give students hands-on experience with statistics, it also provides students opportunities to learn how statistics is applied to their area of interest, introduces the statistical software to them and strengthens their decision-making abilities. For the teacher, it provides insight to the students understanding beyond what grades measured using short homework problems or time-limited in-class exams. It can also provide information about how each student learns and provide clarity on which types of learners the students are. When the instructor has a sense of clarification on how their students learn, then they can provide the sufficient environment for their students to be successful.

References

[1] Van de Walle, J.A. and Bay-Williams, J.M. (2013) Elementary and Middle School Mathematics: Teaching Developmentally. 8th Edition, Pearson, New York, 20-35.

[2] Wooten, R.D. and Tsokos, C.P. (2015) Probability and Statistics Connections for Middle School Educators. Pro-Copy, Tampa.

[3] Wooten, R.D. (2015). An Introduction to Statistics Using the Statistical Platform R. Pro-Copy, Tampa.

[4] Tsokos, C.P. and Wooten, R.D. (2011) Student Handbook for the Joy of Statistics: Learning with Real World Data. Kendall Hunt Publishing Company, Dubuque.

[5] Tsokos, C.P. and Wooten, R.D. (2013) The Joy of Statistics: Learning with Real World Data. 2nd Edition, Kendall Hunt Publishing Company, Dubuque.

[6] Tsokos, C.P. and Wooten, R.D. (2016) The Joy of Finite Mathematics: The Language and Art of Math. Elsevier, Inc., Amsterdam.

Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/

Or contact ajcm@scirp.org