



Assessment of Materials for Engaging Students in Statistical Discovery

Amy G. Froelich
W. Robert Stephenson
Iowa State University

William M. Duckworth
Creighton University

Journal of Statistics Education Volume 16, Number 2 (2008),
www.amstat.org/publications/jse/v16n2/froelich.html

Copyright © 2008 by Amy G. Froelich, W. Robert Stephenson and William M. Duckworth all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Activities, Introductory statistics, Statistical concepts

Abstract

As part of an NSF funded project we developed new course materials for a general introductory statistics course designed to engage students in statistical discovery. The materials were designed to actively involve students in the design and implementation of data collection and the analysis and interpretation of the resulting data. Our overall goal was to have students begin to think like statisticians, to construct ways of thinking about data collection and analysis, to solve problems using data in context. During their development, the materials and related activities were field tested in a small special section of an introductory statistics course for two semesters. This field testing was a ``proof of concept," that is that the materials could work in the laboratory setting and that the materials showed promise for improving students' learning. As a first step in evaluating these materials, students who enrolled in regular sections of the introductory course were used as a comparison group. In this paper, the development and use of the course materials will be discussed briefly. The strategy for evaluating the materials while they were being developed and analysis of students' performance on common assessment questions and the course project will be presented. In addition, the relationship between student attitudes toward statistics and students' performance will be examined.

``Declare the past, diagnose the present, foretell the future; practice these acts.
As to diseases, make a habit of two things - to help, or at least to do no harm."

Hippocrates from *Epidemics*, Bk. I, Sect. XI.

1. Introduction

During the past decade there has been a dramatic change in the way statisticians view statistics education. Statistics educators are focusing more on statistical concepts in a constructivist atmosphere. In general, constructivism encourages students to explore, think and construct mechanisms that help them understand ([Brooks and Brooks 1993](#)). Constructivist approaches to teaching probability and statistics and their relationship to attitudes and performance of students have been studied in recent years. See [Quinn and Weist \(1998\)](#), [Mvududu \(2003\)](#). Efforts to move from lecture (passive) to active learning and from emphasis on procedural to emphasis on conceptual knowledge are throughout the emerging statistics education literature. See [Keeler and Steinhorst \(1995\)](#), [Steinhorst and Keeler \(1995\)](#), [Kvam \(2000\)](#), [Weinberg and Abramowitz \(2000\)](#), and [Anderson-Cook and Dorai-Raj \(2001\)](#). Many of these efforts have come as a result of National Science Foundation (NSF) funded projects to develop hands-on activities, real-world data sets and simulation-based learning ([Cobb 1993](#)).

The NSF grant that funded the development of the new course materials was for a ``proof of concept'' project. The focus of the project was to develop the materials and assess them to show that the concept was viable, that is that the materials could work in the laboratory setting and that the materials showed promise for improving students' learning. Through field testing and assessment we found that the materials worked well in the laboratory setting and that there was some improvement in student performance on the general topic of regression.

The materials were developed for a general introductory statistics course, Stat 101, at Iowa State University. Stat 101 is a four credit course designed for general majors on campus and is comprised of three hours of lecture and two hours of laboratory each week during the semester. Current laboratory activities are very prescribed. For example, the laboratory on the Normal model involves working through standard problems solving for the probability that one gets a value from a Normal model in a particular range or finding a cut-off value associated with a given probability. Similarly, the current lab involving regression looks at the basics of plotting lines and evaluating the strength of a relationship using the correlation coefficient. Students can, and often do, complete the current activities with not much thought given to the underlying statistical concepts.

The new materials we developed are lab activities that actively involve students in the design and implementation of data collection and the analysis and interpretation of the resulting data. Our overall goal was to have students begin to think like statisticians, to construct ways of thinking about data collection and analysis, to solve problems using data in context. Rather than follow steps set down for them, students would take ownership by making decisions about what data to collect and how to organize that data and interpret results within the context of the problem. We had identified several areas where students had struggled in the past and tried to approach those problem areas in new ways in the activities.

One of those problem areas involves the concept of a distribution. Students often struggle with what a histogram is really depicting and how numerical summaries of data relate to the shape and spread of a distribution. A second problem area is the Normal model. Again, many students struggle with the abstract concept of a model for the distribution of a population. Activity #2 was designed to allow students to explore distributions and the Normal model in the context of deciding how to appropriately label a Fun Size bag of M&Ms.

We view the ideas of correlation and regression to be of fundamental importance in an introductory statistics course. Students should understand how to interpret the least squares regression line within the context of the data. This interpretation should go beyond the simple algebraic definition of slope and

intercept to include the statistical idea of variation and that the regression line fits within a meaningful context. Over the years we have seen many students have difficulty with correctly interpreting the slope and intercept, understanding the applications of correlation and regression and the limitations of these procedures. Activities #3, 4 and 5 explore various aspects of correlation and regression.

In the past students have been able to learn the terms related to experimental design but have had difficulty putting the ideas of experimental design into practice when asked to design an experiment to collect data. Activity #5 was developed to allow students to design a simple experiment and analyze the resulting data. Finally, we have noticed students struggle with the course project in Stat 101. This project combines development of an experimental plan with data collection and a full descriptive analysis of the resulting data. The activities were designed to give students more experience in conducting studies of this nature.

Because the activities were being developed and tested for the first time, it was necessary to come up with a plan that would facilitate the further development of the materials, field test the materials, allow for assessment of students, those who used the new materials and those who didn't, while meeting the constraints of teaching over 500 students in the introductory statistics course each semester. Here Hippocrates admonition to ``do no harm" is an important consideration. We did not want to expose large numbers of students to untested materials in their first and, for most, only experience with statistics.

In [Section 2](#) of this paper, we provide a short description of the new course materials, together with learning outcomes, that we developed for an introductory course in statistics. [Section 3](#) discusses the plan for development and preliminary assessment of the materials. [Section 4](#) gives a description of the items used to assess student learning and how they relate to learning objectives for the activities. In [Section 5](#) we give a description of the participants involved in the study. In [Section 6](#) we present the results of the assessments and a statistical analysis of those results. We discuss our findings in [Section 7](#). Finally we indicate directions for further work in [Section 8](#).

2. A Short Description of the New Course Materials

The new course materials developed for the introductory course in statistics consist of new laboratory and classroom activities dealing with data collection, data interpretation, distributions, simple linear regression and statistically designed experiments. Many of the laboratory activities revolve around data collected from Fun Size bags of M&Ms (which do not have a label weight).

In the first activity, students determine the variables that could be used to describe the bags of M&Ms and then collect data on those variables. The learning outcome for this activity is to have students identify categorical and numerical variables that would be helpful in learning more about the bags of M&Ms and to carefully collect data.

In the second activity, students look at the distribution of the weight of the bags of M&Ms using the statistical software package JMP. The learning outcome for this activity is to reinforce the ideas of shape, center, spread and unusual values when describing the distribution of a numerical variable. It also serves to introduce the idea of using a Normal model to represent the distribution of the weights of bags of M&Ms. Students then use the distribution of the weight of the Fun Size bags to determine a reasonable label weight for these bags. Data on larger (labeled) bags of M&Ms are used to further motivate the use of a Normal model for package weights and students discover that the mean weight is not used as the labeled weight of a package. This leads to a discussion of how to use the Normal model to establish a label weight for the Fun Size bags. The learning outcome for this part of Activity # 2 is to see how to use a Normal model to approximate the distribution of a numerical variable and to use that

Normal model in a practical application.

In the third activity, students use simple linear regression to predict the gross weight of a bag based on the number of M&Ms in the bag. The fourth activity has the students look at the regression of gross weight on net weight and of net weight on number to determine if estimates of the slope and intercept from these regression equations would give reasonable estimates of the weight of a single M&M and the weight of a single empty bag. In both activities, students apply the concepts of slope and intercept and discover when they are interpretable within the context of a problem. These activities also highlight the danger of extrapolation.

Finally, in the fifth activity, students design an experiment to determine the weight of a single peanut M&M and the weight of an empty bag using regression estimates of the slope and intercept. Through the previous activities the students have discovered difficulties with the observational study of existing packages of M&Ms especially with having the estimated slope and intercept give reasonable estimates of the average weight of an M&M and the average weight of an empty bag. In Activity # 5 they now use the ideas of experimental design to create a study that avoids those difficulties. A more complete description of the activities with learning outcomes and suggestions for extensions can be found in [Froelich and Stephenson \(2008\)](#) and at <http://stated.stat.iastate.edu/NSF/ESSDactivities/index.html>.

3. Development and Assessment Plan

At the beginning of the project we had drafts of the activities but we had not field tested them in a classroom environment. Because we were developing the materials for our general introductory statistics course (Stat 101) we wanted to use students from this course. During a typical semester there are five lecture sections of Stat 101 each with an enrollment of 100 students. Each lecture section is split into laboratory sections of 50 students each. Four of the five lecture sections are taught by experienced graduate teaching assistants. The ten laboratory sections are conducted by first year graduate teaching assistants. Because of the relative inexperience of the laboratory instructors we felt it was unwise to introduce the new materials into the regular laboratory sections until they had been field tested with a smaller group of students preferably with an experienced teacher available during the laboratory sessions. Luckily, in addition to the five regular lecture sections each spring we have a special section of Stat 101. This special section is smaller (limited to 50 students but usually with an enrollment of about 25). Although there is a laboratory assistant, the instructor for this section (one of the authors) is present at and can interact with the students during the laboratory sessions. This is an ideal section, logically, for introducing new activities.

This special section was created, in part, to attract students with good math skills to the discipline of statistics. Students are invited to enroll in this section and invitations go to students with ACT math scores of 27 or higher. As such, students in this section have higher ACT math scores, on average, than students in the regular sections of Stat 101. This poses a problem of finding a suitable comparison group within the regular sections of Stat 101. We decided to have two ``Control" groups for our study. The first consists of students in regular sections of Stat 101 with ACT math scores of 27 and above. This group would at least have a similar ACT math profile compared to the students in the special section. The second ``Control" group consists of students with ACT math scores below 27. We wanted to collect data on these individuals to establish a baseline for the majority of students who currently enroll in Stat 101. Because we needed access to students records we obtained Institutional Review Board approval and only students who signed an informed consent form were included in our research study.

In the first year of the study draft activities were used in the special section of Stat 101. One of the authors was the instructor for this section. This instructor was present at the laboratory sessions where

the new activities were first tried. If problems with the implementation of an activity arose, which happened with some frequency, the instructor was there to deal with the problem and suggest changes in the conduct of the activity for the next time it would be used. The control groups for the first year of the study were selected from the regular sections of the course during the same semester.

The goal for this first year was to establish the feasibility of the activities and look at preliminary assessment of the efficacy of the activities. This is analogous to Phase I - Feasibility and Phase II - Initial Efficacy in medical research, i.e. clinical trials. For a very nice discussion of the relevance of clinical trial methodology in statistics education research see "Using Statistics Effectively in Mathematics Education Research" www.amstat.org/research_grants/pdfs/SMERReport.pdf. Appendix A of that report shows how the medical model and the research model proposed in a RAND (2003) report relate to a proposed framework for education research. Specifically, Phase I/II in the medical model are consistent with the framework's phase - Frame and begin to Examine where small systematic studies are conducted. The next phase in the proposed framework is Examine and Generalize - where larger studies are conducted under varying conditions with proper controls.

One of the drawbacks to the first year's assessment plan is the potential confounding variable of instructor. The special section is taught by an experienced faculty member while the regular sections of Stat 101 are taught by experienced graduate teaching assistants. In order to control for this potential confounding variable, the control groups for the second year of the study included only students enrolled in an experienced faculty member's regular section of Stat 101 in the fall semester. That same faculty member also taught the special section the following spring where the new activities, with revisions from the previous year, were used. The special section again was chosen for the experimental group as additional activities (outside the scope of this paper) were introduced and field tested in the second year. The regular section in fall was not used to do a randomized comparative trial with the revised activities because of the inexperience of the graduate teaching assistants assigned to conduct the laboratory sections. Because many of our graduate students come from undergraduate mathematics programs and have had little or no exposure to the concepts presented in Stat 101, the fall semester is often as much a learning experience for the new graduate teaching assistants as it is for the students enrolled in the course. The goal of the second year was to look more carefully at the initial efficacy of the activities while controlling for the potential instructor effect. The second year's study would still fall into the Phase I/II medical model framework or the Frame and begin to Examine phase.

4. Description of Assessment Materials

The materials used for assessment of student learning consisted of a pretest exam on algebra and basic statistical concepts, common exam questions on the first and second exams during the semester, and the common course group project. In the second year of the project we also included the Survey of Attitudes Toward Statistics (SATS), [Schau, Stevens, Dauphinee and Del Vecchio \(1995\)](http://www.amstat.org/publications/jse/v16n2/froelichpdf.html).

A pretest exam was administered to all students during the first laboratory session of the semester. Questions on the pretest asked students to perform basic algebraic manipulations, calculate such numerical summaries as the mean and median, to describe a histogram of low temperatures for selected cities on a particular day, and to use the boiling point and freezing point of water to develop the equations relating degrees Fahrenheit to degrees Celsius. From the pretest questions, 11 items were assessed dichotomously. If the student answered the question related to an item correctly, the student was determined to have mastered that item. The 11 separate items were combined to form three main skill groups: ability to compute numerical summaries (Pretest Skill 1), ability to interpret a histogram and describe a distribution of numerical values (Pretest Skill 2) and ability in applied algebra (Pretest Skill 3).

Each student in the study completed common exam questions on the first and second exams during the semester. The first exam covered course material on describing and summarizing distributions and working with the Normal model. The second exam covered course material on describing the relationship between two quantitative variables using correlation and linear regression (no inference) and data collection either through sampling or experimentation. The specific problems used in the assessments for Years 1 and 2 of the study can be found in the [Appendix](#). The same problems were not used in Years 1 and 2 because the Year 1 questions and answers had been released and appeared on web sites accessible to Year 2 students. For each assessment question, the authors developed a scoring rubric. The questions were blinded so that scorers would not be aware of group membership while grading. Two of the authors then scored all students' responses to each question. Any discrepancies in the scores obtained by the two authors for a particular student were resolved between the two authors. The agreed upon score for each student on each problem was then recorded.

The common group project required students to design an experiment to analyze the effect of a specific physical aspect of paper helicopters on an observable aspect of the helicopters' flight. A common choice for several groups was to vary the length of the helicopter wings to determine its effect on the flight time of the helicopters. Students were then required to analyze the resulting data using correlation and regression and to make a conclusion about the effect of the change in the helicopters on the change in the helicopters' flight. A scoring rubric, written by the authors, was used to evaluate the group projects. Students had general knowledge of the requirements of the project, but were not given the specific rubric. At the end of each year of the study, the course projects were blinded, randomly ordered and scored. In Year 1, an independent consultant scored all the projects using the project rubric. In Year 2, one of the authors (who was not involved in teaching the Stat 101 courses in Year 2) scored all the projects using the grading rubric. A complete description of the course project appears in the [Appendix](#).

[Table 1](#) summarizes the specific activities, learning objectives and corresponding assessment items.

Table 1: Activities, Learning Objectives, and Assessment Items

Activity	Learning Outcomes	Assessment Items	
		Year 1	Year 2
#2	Distributions and Numerical Summaries	Exam 1: Q1 & Q2	Exam 1: Q1 & Q2
#2	The Normal Model	Exam 1: Q3 & Q4	Exam 1: Q3 & Q4
#3	Regression and correlation	Exam 2: Q1	Exam 2: Q1
#5	Experimental design	Exam 2: Q2	Exam 2: Q2
#3, #5	Connect experimental design with regression and correlation	Project	Project

For the Year 2 study, the SATS ([Schau, et al., 1995](#)) was used to assess student attitudes toward statistics. Each of the 36 items on the SATS is designed to measure one of six different components of student attitudes toward statistics:

- Affect - Students' feelings toward statistics. A high score on this component indicates that students like learning statistics, they enjoy taking a statistics course and they are not afraid of or nervous about learning statistics.
- Cognitive Competence - Students' attitudes about their intellectual knowledge and skills when applied to statistics. A high score on this component indicates that students believe they have the ability and the mathematical aptitude to learn statistics.
- Value - Students' attitudes about the usefulness, relevance, and worth of statistics in personal and professional life. A high score on this component indicates that students understand the usefulness of statistics in the curriculum of their chosen major, in their future profession, and in their daily lives.
- Difficulty - Students' attitudes about the difficulty of statistics as a subject. Higher scores in this area indicate a belief that statistics is not difficult, while lower scores indicate that students find statistics to be more difficult.
- Interest - Students' level of individual interest in statistics. Higher scores on this component indicate that students are interested on a personal level in learning statistics and communicating statistical information.
- Effort - Amount of work the student expends to learn statistics. A high score on this component indicates that students intend to work and study hard during the statistics class and will complete all assignments and attend all class sessions.

Each item is measured on a Likert-like scale from 1 to 7 and is scored so that a higher response on the question indicates a more positive attitude toward statistics. A student's score on each of the six attitude components is the mean score on the items within that component. The SATS includes a pre-course version and a post-course version that differ only in terms of the tense of the question. The pre-course version of the SATS was given to all students during the first two weeks of the semester and the post-course was administered during the last week of the semester.

5. Description of the Study Subjects

The special section of Statistics 101 was offered during the spring semester in both Year 1 and Year 2. The students in this section were those who responded to a special letter of invitation to enroll in this section. Over 100 freshmen and sophomores with ACT Math Scores of 27 or higher and with majors that require Statistics 101 (such as biology, sociology, psychology, statistics, etc.) or with open majors in the College of Liberal Arts and Sciences were invited to register for the special section for each year of the study. Of these invited students, 20 students in Year 1 and 16 students in Year 2 ultimately signed up for and completed the course. All students enrolled in the special sections of the course agreed to participate in the study and formed the Year 1 and Year 2 Experimental Groups.

During Year 1 of the study, students from four of the regular sections of Stat 101 were invited to participate during spring. Each of these four sections was taught by a different graduate teaching assistant with at least one semester of experience teaching Stat 101. 377 students in the regular sections completed the course with a passing grade of D- or better. Of these students, 199 had agreed to participate in the study. Student characteristics: ACT Math score, ACT English score, ACT Composite score, High School Percentile Rank, Cumulative College GPA and Total Number of Hours Completed, were obtained from the university's registrar's office for all 199 students. 41 students of the 199 had ACT Math scores of 27 or higher. The 39 students with high math ACT scores who completed all assignments throughout the semester were designated the High Math Control Group (Control: H M). The remaining 158 students, whose ACT Math scores were 26 or lower, were candidates for the Regular Control Group. Due to limited resources for assessment, the number of students included in the Regular Control Group was reduced. The first reduction was made by removing the 39 students in the Regular

Control Group who had completed the course project with at least one of the students from the High Math Control Group. A random sample of 50 students was selected from the 119 remaining students. Two of these 50 students chosen for the sample did not complete all assessments during the semester and were dropped from the study. Thus, the final Regular Control Group (Control: Reg) contained 48 students.

During Year 2 of the study, students from a regular section of Statistics 101 for the fall semester, taught by the same instructor as the special section in the following spring semester, were used to form the control groups. 88 students in this section ultimately completed the semester course with a passing grade of D- or higher. Of these students, 72 had agreed to participate in the study at the beginning of the semester. 10 of these 72 students failed to complete one or more of the assessments throughout the semester and were dropped from the study. Student characteristics: ACT Math score, ACT English score, ACT Composite score, High School Percentile Rank, Cumulative College GPA and Total Number of Hours Completed, were obtained from the university's registrar's office for all students in the study. Six out of the remaining 62 students from the regular section of Statistics 101 did not have ACT Math scores on record. These students were dropped from the study, leaving 56 students in the control groups. These remaining students were divided into two groups based on their ACT Math scores. The 17 students with ACT Math Scores of 27 or higher were placed into the High Math Control Group and the other 39 students were placed in the Regular Control Group.

For both years we needed to check to see if the Experimental Group and the High Math Control Group were indeed similar in terms of ACT Math scores. [Table 2](#) displays the numbers of students participating, means and standard deviations for ACT Math scores for the three groups in both years.

Table 2: ACT Math Scores

Year 1					
Group		Number	Mean	Std. Dev.	
Experimental	A	20	29.95	1.701	
Control: H M	A	39	28.97	2.134	
Control: Reg	B	48	21.90	3.068	
Year 2					
Group		Number	Mean	Std. Dev.	
Experimental	A	16	30.00	2.191	
Control: H M	A	17	29.35	1.539	
Control: Reg	B	39	21.74	3.354	

Groups with different letters have differences in mean values that are statistically significant. In both years, the Experimental Group has a slightly higher mean ACT Math score than the High Math Group

but this difference is not statistically significant. In both years, the Regular Group does have a mean ACT Math score that is lower than either of the other two groups and this difference is statistically significant.

6. Assessment Results

Two different types of measures were used to assess student performance for both the Year 1 and Year 2 studies. The first was an overall measure of course performance consisting of common items on examinations. The overall measure was the total score on all examination items. The second type of measures used the score for each assessment (subsets of common exam questions linked to specific activities and the course project) to assess whether student performance among the three groups varied with the concepts covered on that assessment. For each measure of student performance, an ANOVA was used to determine if the mean scores of the three groups (Experimental, High Math Control and Regular Control) were significantly different and if so, the ordering of the means for the three groups. An ANCOVA was then used to determine if any student characteristics, such as ACT Math score, ACT English score, Cumulative College GPA, Total Hours Earned, High School Percentile Rank, and score on the three pretest skills (Pretest Skill 1, 2, and 3) were significant covariates in the model. Instead of dividing the students into the three membership groups (Experimental, High Math Control and Regular Control Groups) as in the ANOVA, the students were divided into just two groups (Experimental and Control) and the deciding factor of the division of the Control Group (ACT Math score) was included as a covariate in the ANCOVA model. If the overall ANCOVA model was significant, the final model was determined by selecting the model with the highest R^2 value having all included variables significant at the 5% level.

In addition, in Year 2, an ANOVA was used to determine if the mean scores of the three groups on the six components of student attitudes toward statistics (SATS) were significantly different. An ANCOVA was then used to determine if there was a significant difference in changes in attitudes (post - pre) between the three groups when including the pre-course attitude as a covariate. Finally, we included the six pre-course and six post-course attitude values as covariates in the ANCOVA models to determine if any aspects of attitudes were significant in predicting performance on the course assessments after controlling for student characteristics and group membership.

6.1 Year 1 Results

[Table 3](#) below gives the means and standard deviations for the overall measure of student performance during the Year 1 study for the Experimental, High Math Control and Regular Control Groups.

Table 3: Year 1 Overall Measure of Student Performance (54 pts max)

Group		Number	Mean	Std. Dev.
Experimental	A	20	40.35	5.29
Control: H M	A	39	39.90	5.60
Control: Reg	B	48	31.23	8.11

The null hypothesis of equal means among the three groups was rejected with a P-value <0.0001. On average, the Experimental and High Math Control Groups both scored significantly higher than the Regular Control Group. Although the overall mean for the Experimental Group was higher than that of the High Math Control Group, this difference was not statistically significant at the 5% level.

The final ANCOVA model for the Year 1 overall measure was highly significant with a P-value <0.0001. The factor of group membership (Experimental vs. Control) was not significant in the model. However, the covariates ACT Math score and Cumulative College GPA were highly significant in the model. This is consistent with the ANOVA model in [Table 3](#). In the ANOVA, students with high ACT Math scores performed better on average than students with lower ACT Math scores. Summaries of the final ANCOVA model are given in [Table 4](#).

Table 4: Year 1 ANCOVA Results for Overall Measure of Student Performance

R ²	Significant Variable	Coefficient	P-value
55.9%	ACT Math	0.7439	<0.0001
	Cumulative College GPA	5.5152	<0.0001
	Pretest Skill 1	1.7416	0.0205

Overall performance on common exam questions during the Year 1 study was significantly related to ACT Math score but not significantly different for students exposed or not exposed to the new course materials. The new materials appear to ``at least do no harm.''

Even though the overall student performance was not significantly different for students exposed or not exposed to the new course materials, student performance on particular aspects covered during the semester could have differed for those students exposed to the new materials compared to those not exposed. To look at student performance on particular aspects of statistics, the scores on questions relating to learning outcomes for the specific activities were analyzed separately. The means and standard deviations of these scores for the Experimental, High Math Control and Regular Control Groups during Year 1 are in [Tables 5, 6, 7 and 8](#). All ANOVAs were statistically significant with P-values less than or equal to 0.0002. Subsequent comparisons of group means was done using a Least Significant Difference approach with an individual comparison alpha level of 0.05. Groups with different letters have differences in means that are statistically significant.

Table 5: Year 1 - Distributions & Numerical Summaries (Exam 1, Questions 1 and 2 combined, out of 18 points).

Group		Number	Mean	Std. Dev.
Experimental		B	20	13.45
Control: H M	A		39	16.05
Control: Reg		B	48	13.33
				4.184

We were very concerned about the results on the assessment questions dealing with distributions and numerical summaries. In this instance the Experimental Group was comparable to the Regular Control Group and significantly lower than the High Math Control Group. Closer examination of responses revealed that the Experimental Group missed the connection between the shape of the distribution and appropriate summary measures (e.g. symmetric shape - sample mean and sample standard deviation, asymmetric shape - five number summary or median and interquartile range). Although mentioned in lecture this idea was not reinforced by what students were doing in Activity #2. This lead us to revise the activity so as to reinforce this idea.

Table 6: Year 1 - Normal model (Exam 1, Questions 3 and 4 combined, out of 6 points).

Group		Number	Mean	Std. Dev.
Experimental	A		20	4.60
Control: H M	A		39	4.15
Control: Reg		B	48	2.04
				1.701

Average scores on questions dealing with the Normal model were similar for the Experimental and High Math Control Groups. These groups scored significantly higher, on average, than the Regular Control Group. Being able to solve Normal model questions relies on basic algebra skills and the ability to use the table of the standard normal distribution. These skills are more developed in students with higher ACT math scores.

Table 7: Year 1 - Regression (Exam 2, Question 1, out of 20 points)

Group			Number	Mean	Std. Dev.
Experimental	A		20	14.20	3.172
Control: H M		B	39	11.77	3.232
Control: Reg		C	48	9.21	3.984

The exam question on regression showed the largest differences between the groups in terms of average scores. The Experimental Group had the highest average score followed by the High Math Control Group with the Regular Control Group having the lowest average score. Examination of student responses revealed that the Experimental Group did better on interpretations of regression coefficients and R^2 . The High Math Control Group got lower average scores due to lower scores on these interpretations. The Regular Control Group tended to have difficulties with some of the calculations as well as the interpretations. All students in Stat 101 see the appropriate interpretations in lecture and in homework assignments. The fact that the Experimental Group did significantly better, on average, on this assessment was encouraging to us, indicating that Activities #3, 4 and 5 might be of some help.

Table 8: Year 1 - Experiments (Exam 2, Question 2, out of 10 points)

Group			Number	Mean	Std. Dev.
Experimental	A		20	8.10	1.373
Control: H M	A		39	7.92	1.707
Control: Reg		B	48	6.65	1.564

The questions dealing with experimentation yield similar results to those dealing with the Normal model. There was no statistically significant difference between the Experimental and High Math Control Groups average scores. Both of these groups had scores that were significantly higher than the Regular Control Group.

Analysis of Covariance models were also run on each of the individual assessments. Summaries of the ANCOVA models are given in [Table 9](#). All final ANCOVA models were highly significant with a P-value <0.0001 . Group membership (Experimental vs. Control) and the covariates ACT Math score and Cumulative College GPA were all significant at the 0.1% level in the final ANCOVA model for Distributions and Numerical Summaries. The coefficient for the Group membership variable indicates that the Experimental Group had lower scores on this assessment than the Control Groups once you adjusted for ACT Math and Cumulative College GPA. This is consistent with what we saw in the Analysis of Variance. For the scores on the Regression question, the final ANCOVA group membership was also statistically significant at the 5% level. However, the sign of the coefficient indicates that the

Experimental Group scored better, on average, than the Control Groups once the other variables are taken into account. Both of these results are consistent with what we saw in the Analysis of Variance. For scores on questions involving the Normal model and Experiments, group membership was not statistically significant. Again, these results are consistent with the results of the Analysis of Variance.

Table 9: Year 1 - ANCOVA Models

Assessment	R ²	Significant Variable	Coefficient	P-value
Distributions Numerical Summaries	29.3%	Group(C-E) ACT Math Cummulative College GPA	1.5496 0.2925 1.7616	0.0001 0.0006 0.0010
Normal Model	45.8%	ACT Comp ACT Math Pretest Skill 2	0.1665 0.1210 0.6120	0.0028 0.0151 0.0374
Regression	46.9%	Group(C-E) ACT Math Cumulative College GPA Pretest Skill 1	-0.8791 0.2130 2.6132 1.1296	0.0374 0.0080 <0.0001 0.0070
Experiments	18.9%	Cumulative College GPA Pretest Skill 3	0.7444 0.2619	0.0068 0.0057

In all but one of the ANCOVA models, Cumulative College GPA was a highly significant covariate. Three of the four ANCOVA models also included a significant Pretest Skill (Pretest Skill 2 for the Normal model, Pretest Skill 1 for Regression and Pretest Skill 3 for Experiments). The significance of Pretest Skill 2 (ability to describe distributions using histograms) on student performance on the Normal model was not surprising. However, it was surprising this Pretest Skill did not show up in the ANCOVA model for Distributions and Numerical Summaries. For the Regression questions, the results of the ANCOVA indicate that even after controlling for ACT Math Score and Cumulative College GPA, the ability to compute numerical summaries (Pretest Skill 1) was still significantly related to students' scores.

For the course group project, students were randomly assigned to project groups in both the regular sections and special section of the course. There were a total of 7 projects from the Experimental Group, 28 projects from the High Math Group and 18 projects from the Regular Control Group. The 28 projects in the High Math Control Group were completed by groups containing only one or two students from the High Math Control Group. The rest of the project group members in the High Math Control Group were students with ACT Math scores below 27. The 18 projects from the Regular Control Group contained only students with ACT Math scores below 27. [Table 10](#) gives the means and standard deviations for the project scores for the three groups.

Table 10: Year 1 Project Results (out of 50 points)

Group			Number	Mean	Std. Dev.
Experimental	A		7	46.36	2.48
Control: H M		B	28	40.38	5.82
Control: Reg		C	18	36.25	8.36

As with the individual assessments, the null hypothesis of equal means between the three groups was rejected with P-value of 0.0037. The Experimental Group had the highest score on average, followed by the High Math Control Group and then the Regular Control Group. All pairs of comparisons among the three groups were statistically significant. We were encouraged by the performance of the Experimental Group on this assessment as, unlike the exam questions, the project requires students to put together ideas from various topics into a coherent statistical investigation.

6.2 Year 2 Results

[Table 11](#) below gives the means and standard deviations of the overall measure of student performance on common exam questions during the Year 2 study for the Experimental, High Math Control and Regular Control Groups.

Table 11: Year 2 Overall Measure of Student Performance (out of 45 points)

Group			Number	Mean	Std. Dev.
Experimental	A		16	40.06	2.48
Control: H M		B	17	34.71	5.88
Control: Reg		C	39	27.99	7.86

The null hypothesis of equal means between the three groups was rejected with a P-value <0.0001. The Experimental Group scored higher on average than the High Math Control Group which scored higher on average than the Regular Control Group. These differences were statistically significant at the 5% level.

The final ANCOVA model for the Year 2 overall measure was highly significant with a P-value <0.0001. Again, the results of the ANCOVA are consistent with the results from the ANOVA. Group membership was significant in the ANCOVA even after you controlled for ACT Math score and Cumulative College GPA. The sign of the coefficient indicates that, on average, the students in the Experimental Group performed better than students in the Control Groups even after controlling significant covariates. Summary values of the final ANCOVA model for the overall measure of students'

performance for the Year 2 study are given in [Table 12](#).

Table 12: Year 2 ANCOVA Results on Overall Measure of Student Performance

R²	Significant Variable	Coefficient	P-value
60.53%	Group(C-E) ACT Math Cumulative College GPA	-1.8602 0.7858 3.9522	0.0474 <0.0001 0.0015

Unlike in Year 1, the results from Year 2 indicate that students exposed to the new course materials did significantly better than those who did not use the new materials when using the overall measure of performance on common exam questions. In addition, after controlling for the significant covariates, the factor of group membership was significant at the 5% level. As in Year 1, ACT Math score and Cumulative College GPA turned out to be statistically significant covariates.

To look at differences in student performance over the course of the semester on items tied to the activities, separate analyses of scores on common exam questions dealing with Distributions and Numerical Summaries, the Normal model, Regression and Experiments was performed. [Table 13](#) contains the means and standard deviations of the scores for the Experimental, High Math Control and Regular Control Groups for these areas, respectively.

Table 13: Year 2 - ANOVA Results

Distributions and Numerical Summaries (Exam 1, Questions 1 & 2, out of 16 points)				
Group		Number	Mean	Std. Dev.
Experimental	A	16	13.66	1.62
Control: H M	A	17	12.53	2.54
Control: Reg	B	39	9.28	3.22
Normal Model (Exam 1, Questions 3 & 4, out of 10 points)				
Group		Number	Mean	Std. Dev.
Experimental	A	16	9.72	0.55
Control: H M	A	17	8.76	2.36
Control: Reg	B	39	6.90	2.98
Regression (Exam 2, Question 1, out of 12 points)				
Group		Number	Mean	Std. Dev.
Experimental	A	16	10.38	1.16
Control: H M	B	17	7.53	2.05
Control: Reg	B	39	6.51	2.42
Experiments (Exam 2, Question 2, out of 7 points)				
Group		Number	Mean	Std. Dev.
Experimental	A	16	6.31	0.57
Control: H M	A	B	17	5.88
Control: Reg	B	39	5.29	1.29

For each assessment, the null hypothesis of equal means for the three groups was rejected with P-values

of <0.0001 , 0.0006 , <0.0001 , and 0.0064 for the Distributions and Numerical Summaries, Normal Model, Regression and Experiments assessments, respectively. With all of these analyses, the Regular Control Group had significantly lower mean scores than the Experimental Group. The Experimental Group had higher average scores than the High Math Control Group on each of the four assessments. However, for all except the assessment question on Regression, there was no statistically significant difference between the Experimental Group and the High Math Control Group. The accumulation of differences was enough to create the statistically significant difference between these two groups when looking at the overall measure of performance on common exam questions.

The ANCOVA results mirror the findings above. All final ANCOVA models were statistically significant with P-values <0.0001 , <0.0001 , 0.0002 , and <0.0001 for the Distributions and Numerical Summaries, Normal Model, Regression and Experiments assessments, respectively. The only time Group membership was statistically significant in the ANCOVA model was for the question on Regression. The sign of the coefficient was consistent with the fact that the Experimental Group scored higher, on average, than the Control Groups on this assessment question. The three other assessment areas produced ANCOVA models that did not include a Group membership variable. Summary values for the final ANCOVA models for the three exams are given in [Table 14](#).

Table 14: Year 2 - ANCOVA Results}

Assessment	R ²	Significant Variable	Coefficient	P-value
Distributions and Numerical Summaries	58.1%	ACT Math Pretest Skill 3	0.3836 0.5443	<0.0001 0.0142
Normal Model	34.4%	Cummulative College GPA Pretest Skill 3	1.2015 0.7739	0.0102 <0.0001
Regression	52.6%	Group (C-E) Cumulative College GPA	-1.3614 1.9991	<0.0001 <0.0001
Experiments	21.8%	ACT Math Cumulative College GPA	0.0642 0.5534	0.0229 0.0147

For the course group project, students were randomly assigned to project groups in both the regular and special section of the course. There were 6 projects completed from the Experimental Group. From the two control groups, there were a total of 21 different projects. From prior experience, high math ability students play a significant role in the completion of the group course project. Therefore, a project was assigned as a High Math Control Group project if the project was completed by at least one student from the High Math Control Group. This assignment also matches the way projects were treated in the Year 1 study, in that only Regular Control Group students contributed to the projects in their own group, while both High Math and Regular Control Group members contributed to projects in the High Math Control Group. This division resulted in 10 projects for the High Math Control Group and 11 projects for the Regular Control Group in the Year 2 study. [Table 15](#) gives the mean and standard deviations for the course project for the three groups (Experimental, High Math Control, Regular Control).

Table 15: Year 2 Project Results (out of 50 points)

Group	Number		Mean	Std. Dev.
Experimental	A		6	38.67
Control: H M	A		10	31.80
Control: Reg	A		11	33.91

Unlike in Year 1, there were no significant differences in means among the three groups (P -value = 0.2734). The scores on the project were generally lower in Year 2 compared to Year 1. Also, except for the Regular Control Group, there was quite a bit more variation in scores for the Year 2 project compared to that in Year 1. The magnitude of the differences between groups was substantial but finding statistically significant differences was hampered by the larger variation and smaller sample sizes.

The SATS was used to assess differences in student attitudes about statistics at the beginning of the semester. The means and standard deviations of the scores on the six components of the SATS from the three groups is given in [Table 16](#) below.

Table 16: Year 2 SATS Survey Results

SATS Component	Group	Pre-course				Post-course			
				Mean	Std. Dev			Mean	Std. Dev.
		A	B			A	B		
Affect	Experimental	A		5.04	1.02	A		5.43	1.11
	Control: H M	A		4.91	0.79	A		5.25	0.92
	Control: Reg	B		4.03	1.02	B		3.97	1.27
Cognitive Competence	Experimental	A		5.57	0.71	A		5.83	0.93
	Control: H M	A		5.84	0.75	A		5.98	0.67
	Control: Reg	B		4.97	0.90	B		4.69	1.20
Value	Experimental	A		5.74	0.77	A		5.42	0.88
	Control: H M	B		5.11	0.93	A		5.15	1.14
	Control: Reg	B		4.94	0.89	B		4.52	1.01
Difficulty	Experimental	A		3.95	0.62	A		4.28	0.72
	Control: H M	A		4.27	0.63	A		4.44	0.77
	Control: Reg	A		3.89	0.61	A		3.86	1.00
Interest	Experimental	A		5.36	1.08	A		5.02	1.42
	Control: H M	B		4.53	1.14	A		4.56	1.43
	Control: Reg	B		4.19	1.09	B		3.73	1.34
Effort	Experimental	A	B	6.25	0.94	A	B	5.75	0.90
	Control: H M	A	B	5.66	0.75	A	B	5.33	0.69
	Control: Reg	A		6.31	0.69	A		6.04	1.10

There were significant differences in mean attitudes between the three groups on five of the six components on both the pre-course and post-course SATS. The Experimental and High Math Control Groups had significantly higher mean scores than the Regular Control Group on the Affect and Cognitive Competence components of the pre-course SATS. Many students, especially at the beginning of a semester, see the introductory statistics course as essentially a mathematics course. It is not surprising then that students with strong mathematics backgrounds would have a better feeling about learning statistics (Affect) and would be more confident in their abilities to learn the course materials (Cognitive Competence). By the end of the course, the same patterns emerged with slightly higher means scores for the Experimental and High Math Groups and slightly lower mean scores for the Regular Group. The Experimental Group had significantly higher mean scores on the Value and Interest component of the pre-course SATS than either control group. This result is not unexpected given the nature of the Experimental Group. These students were interested in learning statistics and valued the subject enough to enroll in a special section of the course. By the end of the semester, the difference between the High Math Group and Experimental Group was still present but no longer statistically significant. Interestingly, the Experimental Group and the Regular Control Group had statistically higher mean scores for the amount of Effort they would spend on learning statistics and on the course itself than the students in the High Math Control Group. This result may be attributed to the fact that the

Regular Control Group expected to work harder on the course possibly due to perceived lack of mathematics ability while the Experimental Group expected to work harder possibly due to enrollment in a special section of the course. Again, by the end of the semester, the difference between the High Math Group and the Experimental Group had narrowed and was no longer statistically significant. The means for Difficulty are consistent with those for Effort, both pre- and post-course, but no statistically significant differences are seen among any of the groups.

Students responses on both the pre-course and post-course versions of the SATS were used to look at potential differences in the change in attitudes over the course of the semester between the three groups. Six ANCOVA models, one for each component of the SATS, were used to look at differences in changes in attitudes with the pre-course attitude score on the component included as a covariate in the model. Differences in the change in attitudes in the three groups were statistically significant at the 5% level only for the Affect and Cognitive Competence components. In both cases, the significant change occurred between the Regular Control Group and the other two groups. And in both cases, the sign of the change was negative, indicating a more negative change in attitudes on these two components across the semester for the Regular Control Group.

Finally, student responses on the SATS were used to determine if attitudes toward statistics were significantly related to overall performance in the course after controlling for group membership and student characteristics. Using the overall measure of student performance on the common exam questions, student mean scores from the six components of the pre course SATS and the 6 components of the post course SATS were added to the list of potential variables that could be included in an ANCOVA model. Only one covariate, post course Cognitive Competence was a statistically significant addition to the ANCOVA model (P-value of 0.0006 for full versus reduced model). The coefficients and corresponding P-values for the SATS ANCOVA model are given in [Table 17](#).

Table 17: Year 2 ANCOVA Model with SATS

Significant Variable	Coefficient	P-value
Group(C-E)	-2.0865	0.0168
ACT Math	0.4703	0.0094
Cumulative College GPA	3.2214	0.0052
Post Cognitive Competence	2.2945	0.0006

7. Discussion

The laboratory activities we developed had the overall goal of getting students in the first course in statistics to begin to think more like a statistician. By this we mean that students should begin to see data in context and statistical thinking as a way to answer questions about the world around us. An important part of beginning to think like a statistician involves asking questions about the data. Why should we collect data? What data should we collect? How should the data be collected? Students should also see that there is often more than one way to analyze data. Different analyses can lead to different answers and students should recognize why those answers are different. With this in mind we developed our laboratory activities to revolve around data collected at the beginning of the course. By revisiting the

data in activities throughout the first half of the course we depart from the usual presentation of a new data set for each new method. We believe this helps students see statistics as a way of learning about the world around them as opposed to a laundry list of techniques and methods.

>From the point of view of a ``proof of concept," this study was a success. We were able to develop new lab activities, field test and make adjustments to them, so as to ``do no harm." The preliminary results on assessing student learning are especially encouraging for the topic of regression. Students exposed to the new activities performed better on the assessment of regression compared to students that were not exposed even after adjusting for significant covariates. This may be due, in part, to an additional laboratory activity (Activity #4) that tries to get students to think about what might be reasonable estimates for the y-intercept and slope coefficient given the context of the explanatory and response variables.

In analyzing the statistical results above, several patterns emerge. Students with strong math backgrounds performed better on average than the other students. In every case, the Regular Control Group performed no better and often times significantly worse on average than either the High Math Control Group or the Experimental Group. Particularly at the beginning of the semester, students with high math abilities appear to use their previous mathematics skills to help them learn concepts about distributions. The questions dealing with distributions and numerical summaries require students to have basic skills in statistics along with good numerical literacy. It is not surprising then that students with stronger mathematics backgrounds would score higher on these questions.

Except for the Year 1 assessment of distributions and numerical summaries, the Experimental Group scored, on average, the same as or in many cases significantly higher than the High Math Control Group. The performance of students in the special section was noticeably lower on the questions which required students to describe characteristics of a distribution (Questions 1 & 2 - Year 1 in the [Appendix](#)). Many times, students in the special section left out some of the characteristics of the distribution and thus scored lower on the problem as a whole. This result could be due to the activity or to differences in the emphasis of the instructors when teaching this material. In the Year 2 study, which controls for instructor differences and used a revised Activity #2, the mean scores of the High Math Control Group and the Experimental Group on assessment of distributions and numerical summaries are not statistically different.

In both Year 1 and Year 2 of the study, the Experimental Group performed much better, on average, on the assessment of regression than either of the control groups. Regression concepts, such as slope and intercept, appear many times in high school algebra courses. However, in teaching this material as it relates to regression in statistics, our experience has been that students, even ones with strong math backgrounds, do not easily transfer their previous mathematical knowledge to this new application. The new course materials put a great deal of emphasis on regression and correlation concepts. The students in the special section of the course were exposed to these concepts repeatedly over the course of several labs. While the amount of class time used to cover these topics between the two sections was approximately the same, the regular section only completed one lab on regression and correlation as opposed to two labs for the special section.

While the differences in performance on the regression assessment carried over into the course group project in the Year 1 study, no significant difference in performance between the three groups was found for the projects in the Year 2 study. The observed mean differences between the three groups were roughly equal between Years 1 and 2. However, the standard deviations of experimental and high math control groups were higher and the number of projects was lower in Year 2 than in Year 1. These differences lead to a lower power for the Year 2 study. Also, in studying the Year 2 projects, we found that students had difficulties in understanding the concept of replication - having several experimental

units in each treatment group. Many project groups in both the special and the regular section simply conducted the experiment by applying the treatment to each experimental unit several times. The labs on experimentation used in the two sections of Statistics 101 have since been rewritten to help students better understand the idea of replication within an experiment. The effect of these revisions on project scores has not yet been tested. Finally, in discussing these results, we have discovered that the two instructors for the special section in this study approach the concept of experimental units differently when teaching the course. This discussion has lead to more consistent instruction of this topic.

The ANCOVA results are consistent with the ANOVA results. The results of significant mean differences in scores on the assessments between the experimental and control students from the ANOVA analyses are still present even when controlling for other student characteristics. The variable Cumulative College GPA is significant in all but one ANCOVA model. Controlling for other student characteristics and their group membership (either Control or Experimental) students with higher GPAs are scoring higher on average on these assessments. ACT Math is significant in four of five ANCOVA models from Year 1, and in three of five ANCOVA models from Year 2. Students with higher ACT Math scores tend to do better on these assessment items.

In terms of attitudes towards statistics, as measured by SATS, there was no significant change in the patterns of means for the three groups from pre- to post-course. According to the ANCOVA models, only Affect and Cognitive Competence changed, for the worse, for the Regular Control Group. Again, the new activites ``did no harm" for the experimental group.

8. Summary and Conclusions

The process of development and field testing of the activities presented many challenges but there were some encouraging results. The design of the study was influenced by many factors. Because the activities were to be tested for the first time, we wanted to present them in a manner that would enable us to address and correct problems immediately. The special section of Stat 101 provided us with a research laboratory where we could do this. Given the poor performance of the students in the special section on the assessment of distributions and numerical summaries the first year, we were able to make adjustments to Activity #2 and to address the deficiency in additional instruction in the special section. Had we tried this activity on a larger group of students in the regular sections of Stat 101 we may have done harm that could not be corrected as easily.

Field testing the activities in the special section solved a logistical problem but introduced the problem of finding a comparable control group. Having a control group with a similar ACT Math profile addressed this difficulty. Due to constraints on teaching assignments for Year 1 of the study, group membership (either experimental or control) was confounded with course instructor. The special section of the course was taught by an experienced professor while the regular sections of the course were taught by relatively inexperienced graduate teaching assistants. Differences in teaching styles, methods, and emphasis between the five instructors could affect student learning. While this limitation is not present in Year 2 of the study (the same instructor taught all students), the regular and special sections of Statistics 101 were structured differently in both years. The enrollments of the special sections were around 20 students while the enrollments in the regular sections were around 100. The lab section for the special sections was led by both the course instructor and a graduate student teaching assistant while the labs for the regular sections were each run by a single graduate student teaching assistant.

Now that the development and initial assessment has been performed and we are satisfied that the new materials will not do harm we plan to proceed to a randomized clinical trial (Phase II) so as to include a proper randomized control group. Because of the restrictions mentioned earlier, we intend to conduct

this trial in a different introductory statistics course where the laboratory is taught by the course instructor, rather than a graduate teaching assistant. There are multiple sections of this course that will act as blocks in our design. Students from sections of this introductory statistics course who agree to participate will be randomly assigned to treatment groups. One group will use the activities we developed in lab and the other group will use the current laboratory activities. With the availability of the ARTIST materials (www.gen.umn.edu/artist) we hope to use these in future assessment so as to be able to compare performance with other statistics students at different institutions.

Acknowledgment

This material is based upon work supported by the National Science Foundation, DUE # 0231322. We would also like to thank Dr. Carl Lee of Central Michigan University for evaluating the course projects for Year 1 of this study. We would also like to thank the Associate Editor and two anonymous referees whose thoughtful comments have greatly improved this paper. Earlier versions of the assessment of the Year 1 and Year 2 studies have appeared in the *ASA Proceedings of the Section on Statistical Education*.

Appendix

First Exam Questions

Year 1

1. The table below gives the daily high temperature recorded at the Des Moines Airport for the month of January 2004.

Day	Temp								
1	52	7	26	13	35	19	16	25	27
2	60	8	31	14	40	20	23	26	21
3	34	9	26	15	38	21	46	27	15
4	21	10	31	16	34	22	19	28	5
5	11	11	49	17	34	23	53	29	2
6	16	12	45	18	27	24	25	30	2
								31	14

- a. Make a stem and leaf plot of the daily high temperatures for the month of January 2004.
- b. Describe the distribution of the daily high temperatures for the month of January 2004. (Calculations are not needed for your description).

c. Which numerical summaries would you report for these data? Do not calculate these values, but briefly explain your choice.

2. Another measure of center is the midrange.

$$\text{midrange} = (\text{min} + \text{max})/2$$

For a sample of size 20, which is affected more by a single outlier, the mean or the midrange? Explain your answer.

3. The Environmental Protection Agency (EPA) estimates fuel economy for automobile models. Assume the distribution of fuel economy is normally distributed with a mean of 24.8 mpg and a standard deviation of 6.2 mpg for highway driving.

- What percent of all cars will have fuel economies greater than 27 mpg?
- The worst 5% of all cars will have fuel economies less than what amount?

4. First-time Freshman attending Iowa State University in Fall 2003 had a mean ACT Composite Score of 24.6 points. The first quartile of ACT Composite scores was 21.7. If the ACT Composite scores follow a normal distribution, what is the standard deviation of these scores?

Year 2

1. Short Answer

- A data set contains 10 observations that have a median of 22. One of the 10 observations is changed from 5 to 3. What is the new median?
- A data set contains 10 observations that have a mean of 22. One of the 10 observations is changed from 10 to 20. What is the new mean?
- A data set contains 10 observations that have a standard deviation of 3 and an interquartile range of 2.5. Five points are added to each of the 10 observations. What are the new standard deviation and IQR?
- A data set contains 10 observations that have a mean of 20 and a median of 22. Five points are added to each of the 10 observations. What are the new mean and median?
- A data set contains 10 observations that have a standard deviation of 0. What is the range of the data?
- Lilac House, a Bed & Breakfast on Market Street in Mackinac Island, MI, has 5 rooms. The most expensive room is \$120 a night and the cheapest room is \$60 a night. What are possible prices for the other three rooms so that the median price of a room is \$60 and the mean price of a room is \$75 a night?

2. Use the JMP output to answer the following questions about the distribution of the number of tornadoes in Iowa per year for the years 1953-2003.

- a. Describe the distribution of the number of tornadoes per year in Iowa.
 - b. Give two reasons why the mean is larger than the median for these data.
 - c. Which numerical summaries of center and spread are most appropriate for these data? Give the values of these summaries and explain your answer.
3. The tensile strength of paper is measured in pounds per square inch (psi). The tensile strength of the paper produced by a particular company has a normal distribution with a mean of 35 psi.
 - a. Currently, 25% of the paper produced by the company has a tensile strength less than the minimum requirement of 30 psi. What is the standard deviation of the tensile strength of the paper produced by this company?
 - b. The company would like to achieve its goal of having only 5% of the paper produced below the minimum tensile strength of 30 psi. If the standard deviation of the tensile strength of the paper is 5 psi, what does the mean tensile strength of the paper have to be to achieve the company's goal?
4. Control Groups: The Mental Development Index (MDI) of the Bayley Scales of Infant Development is a standardized measure used in longitudinal follow-up of high-risk infants. Scores on the MDI have a normal distribution with a mean of 100 and a standard deviation of 16.
 - a. What proportion of children will have a MDI score above 80?
 - b. What is the MDI score so that 90% of all children will be above that score?

Experimental Group: Intelligence Quotient (IQ) scores are normally distributed with a mean of 100 points and a standard deviation of 15.

- a. What percent of people will have an IQ score more than 95?
- b. To belong to the group MENSA, a person is required to have an IQ score in the top 2% of the population. What is the minimum IQ score required to belong to the group MENSA?

Second Exam Questions

Year 1

1. Can consumption of wine help reduce the number of deaths from heart attacks? Yearly wine consumption (liters of alcohol from drinking wine, per person) and yearly death rates from heart disease (deaths per 100,000 people) for 9 randomly selected European countries are obtained. Consult the JMP output entitled Wine Consumption and Heart Disease.
 - a. What is the least squares regression equation relating Death Rate to Wine Consumption?
 - b. Give an interpretation of the intercept within the context of the problem.
 - c. Give an interpretation of the slope within the context of the problem.

- d. Use the least square regression to predict the death rate from heart disease for another European country, France, which has a wine consumption of 9.1 liters per person per year.
- e. The actual death rate from heart disease for France is 71 deaths per 100,000 people. What is the residual for France?
- f. What is the value of R^2 for these data? Give an interpretation of this value.
- g. What is the value of r , the correlation between wine consumption and death rate from heart attack?
- h. Describe what you see in the plot of residuals and what this tells you about the relationship between wine consumption and death rate from heart disease.
- i. Based on this study and statistical analysis, if people in a country like France were to drink less wine would the death rate from heart disease go up?

2. Students in an introductory statistics class were asked to design an experiment to determine the relationship between the height of a ramp and the distance a ball rolls. One group decided to use 5 different ramp heights: 6, 9, 12, 15 and 18 inches. For simplicity, the first six trials were completed with the ramp height set at 6 inches; the next six trials were completed with the ramp height set at 9 inches, etc. until all 30 trials for the experiment were completed. Each of the 30 trials was conducted with a different randomly chosen ball. The same member of the group let go of a ball at the same place on the ramp each time. To make sure the ramp did not move between trials, the location of the bottom of the ramp was marked and the ramp was reset to the same place before each trial. The same group members were responsible for marking the location at which the ball stopped rolling and then measuring this distance from the end of the ramp.

- a. What are the experimental units?
- b. What is the response variable?
- c. What is the explanatory variable?
- d. How many levels of the explanatory variable were used in the experiment?
- e. How many trials were conducted at each level of the factor?
- f. Name two ways the group used the principle of control in their experiment.
- g. Which principle did the group fail to use correctly in their experiment, replication or randomization? Explain what the group did wrong and how you would fix their mistake.

Year 2

1. Use the JMP output to answer the following questions. The data is the length and width in centimeters of a sample of butter clams.

- a. What is the response variable in this regression?

- b. What is the least squares regression equation relating the length of butter clams to their width?
 - c. Give an interpretation of the slope within the context of the problem.
 - d. Use the least squares regression to predict the length of a butter clam who has a width of 7 cm.
 - e. The data includes a clam whose width is 7 cm and length is 9.5 cm. Find the residual for this clam.
 - f. What is the value of R^2 for these data? Give an interpretation of this value in the context of the problem.
 - g. What is the value of r , the correlation between the width and length of a butter clam?
 - h. Describe the residual plot. Do you see any problems with the residual plot? If yes, what effect will these problems have on your linear regression of the width and length of a butter clam?
2. An ultramarathon is a foot race that is longer than 26.2 miles. Doctors have found that people who run an ultramarathon are at increased risk for developing respiratory infections after the race. Doctors believe that taking vitamin C the 10 days before and the 10 days after the race would reduce the incidence of respiratory infections in the ultramarathon runners. To test their hypothesis, 20 people were selected to receive either vitamin C or a placebo. Ten days after the race, the two groups were studied to determine how many of the runners in each group developed a respiratory infection.
 - a. Why is this study an experiment?
 - b. What are the experimental units?
 - c. What is the response variable?
 - d. What is the factor?
 - e. What are the treatments?
 - f. Name one thing the experimenter should use for control in this experiment.

Group Course Project Description

The focus of this project is on designing an experiment and using correlation and regression to analyze the resulting data. Your experiment will involve a paper helicopter. A prototype of a paper helicopter is provided. There are many ways to evaluate the flight of the paper helicopter. There are also many factors that may affect that flight. The object of this project is to investigate the relationship between a single factor you can manipulate that may affect the flight of the paper helicopter and a single measurement of some characteristic of the helicopter's flight. To do so, you should:

1. Phrase a hypothesis about the relationship between a numerical characteristic you can manipulate on the helicopter and a numerical characteristic describing the flight of the helicopter.
2. Identify your explanatory and response variables for the experiment. Indicate how you will measure the response variable.
3. Decide how you are going to design an experiment to investigate your hypothesis. You must have at least 5 levels of your factor.
4. Run the experiment and collect the data. This will require you to construct paper helicopters and fly them. You can make copies of the prototype helicopter. You must have a minimum of 30 data points.
5. Analyze your data. Remember, the focus of the statistical analysis is on correlation and regression. Turning in computer output is not enough. You must interpret the results of any analysis you do.
6. Write a final report. Your report should include sections on your hypothesis and the explanatory and response variables, the design of your experiment, the data, the statistical analysis and its interpretation and your conclusion stating what you have learned about the hypothesis from your data.

Grades will be determined on:

1. How well you used the ideas of Chapter 13 to collect your data. [20 pts]
2. Relevance and completeness of the summary of the data. [20 pts]
3. Appropriateness of your conclusions. [5 pts]
4. Clarity of the final report. [5 pts]

References

Anderson-Cook, C. M. and Dorai-Raj, S. (2001), "An Active Learning In-Class Demonstration of Good Experimental Designs," *Journal of Statistics Education* [Online], 9(1)
www.amstat.org/publications/jse/v9n1/anderson-cook.html

Brooks, J.G. and Brooks, M.G. (1993), *In Search of Understanding: The Case for Constructivist Classrooms*, Association for Supervision and Curriculum Development, Alexandria, Virginia.

Cobb, G.W. (1993), "Reconsidering Statistics Education: A National Science Foundation Conference," *Journal of Statistics Education* [On-line], 1(1). www.amstat.org/publications/jse/v1n1/cobb.html

Froelich, A.G. and Stephenson, W.R. (2008), "How Much Does an M&M Weigh?" submitted for publication.

Keeler, C. M., and Steinhorst, R. K. (1995), "Using Small Groups to Promote Active Learning in the

Introductory Statistics Course: A Report from the Field," *Journal of Statistics Education* [Online], 3(2) www.amstat.org/publications/jse/v3n2/keeler.html

Kvam, P.H. (2000), "The effect of active learning methods on student retention in engineering statistics," *The American Statistician*, 54(2) , 136-140

Mvududu, N. (2003) "A Cross-Cultural Study of the Connection Between Students' Attitudes Toward Statistics and the Use of Constructivist Strategies in the Course," *Journal of Statistics Education* [Online], 11(3) www.amstat.org/publications/jse/v11n3/mvududu.html

Quinn, R.J., and Wiest, L.R. (1998), "A constructivist approach to teaching permutations and combinations," *Teaching Statistics*, 20, 75-77.

Schau, C., Stevens, J., Dauphinee, T. L., and Del Vecchio, A. (1995), "The development and validation of the Survey of Attitudes Toward Statistics," *Educational and Psychological Measurement*, 55(5), 868-875.

Steinhorst, R. K., and Keeler, C. M. (1995), "Developing Material for Introductory Statistics Courses from a Conceptual, Active Learning Viewpoint," *Journal of Statistics Education* [Online], 3(3) www.amstat.org/publications/jse/v3n3/steinhorst.html

Weinberg, S. L., and Abramowitz, S. K. (2000), "Making General Principles Come Alive in the Classroom Using an Active Case Studies Approach," *Journal of Statistics Education* [Online], 8(2) www.amstat.org/publications/jse/secure/v8n2/weinberg.cfm

Amy G. Froelich
Department of Statistics
Iowa State University
Ames, IA 50011-1210
amyf@iastate.edu

W. Robert Stephenson
Department of Statistics
Iowa State University
Ames, IA 50011-1210
wrstephe@iastate.edu

William M. Duckworth
Department of Decision Sciences
Creighton University
Omaha, NE 68178
williamduckworth@creighton.edu
