

WEBWORK PROBLEMS FOR STATISTICS I

An Interactive Qualifying Project proposal to be submitted to the faculty of Worcester Polytechnic Institute in partial fulfillment of the requirements for the Degree of Bachelor of Science

Submitted by:

Zehao Li
Yizhou Xia

Submitted to:

Project Advisor:
Prof. Joseph Petrucci

March 1, 2011

Table of Contents

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: BACKGROUND.....	4
2.1 Obstacles to Learning Statistics.....	4
2.2 Two Learning Models	5
2.2.1 Catrambone’s Subgoal Learning Model	5
2.2.2 Renkl, Atkinson and Maier’s Backward Fading Model	8
2.3 The Role of Homework.....	9
2.3.1 The Benefits of Assigning Homework.....	9
2.3.2 Paper-Based Homework and Web-Based Homework	9
2.3.3 Web-Based Homework Systems	11
2.3.4 The WeBWoRK System.....	12
2.4 Interaction in Learning	14
2.4.1 The Importance of Interaction	14
2.4.2 Types of Interactions	14
2.5 The PG language.....	16
CHAPTER 3: METHODOLOGY	18
3.1 Outline Course Content.....	19
3.2 Identify Present Problems.....	21
3.3 Programming Language	22
3.4 Proposed Improved Homework Problems.....	24
3.4.1 Subgoal Learning Model	24
3.4.2 Forward Fading Model	25
3.4.3 Combining the Two Models.....	26
3.4.4 Implementing the Model.....	27
3.5 Randomized Controlled Experiments	34
3.5.1 Pre-Test.....	35
3.5.2 Treatment Group.....	37
3.5.3 Control Group.....	45
3.5.4 Post-Test.....	47

CHAPTER 4: RESULTS	49
4.1 Conduct of Experiments	49
4.2 Statistical Analysis	51
4.2.1 Two-Way Effects Model	51
4.2.2 Linear Regression.....	52
CHAPTER 5: CONCLUSIONS.....	55
5.1 Accomplishments	55
5.1.1 Packages	55
5.1.2 New WeBWork-Based Instructional Model	55
5.1.3 Implementation	55
5.1.4 Testing	56
5.2 Possible Future Improvements	56
5.2.1 Package.....	56
5.2.2 Problem Set and Experimental Design	57
5.3 In Conclusion	58
APPENDIX	60
A. Package “improvedCompound” Code	60
B. Sample Question Code.....	71
C. Problem Set for Confidence Interval Experiment	78
D. Problem Sets for Hypothesis Test Experiment	83
D.1 Population Mean.....	83
D.2 Population Proportion.....	93
WORK CITED	103

Table of Figures

Figure 1 Project Timeline.....	19
Figure 2 Rudimentary Knowledge Tree for Introductory Statistics.....	20
Figure 3 Example Problem: the First Part.....	28
Figure 4 Example Problem: the Second Part.....	29
Figure 5 Example Problem: the Third Part.....	30
Figure 6 Example Problem: the Fourth Part.....	31
Figure 7 Example Problem: the Fifth Part.....	32
Figure 8 Example Problem: the Hint Part.....	33
Figure 9 Example Problem: the Sixth Part.....	34
Figure 10 Example of Pre-Test Problem, First Question.....	36
Figure 11 Example of Pre-Test Problem, Second Question.....	37
Figure 12 Example of Treatment Group Problem, First Part.....	38
Figure 13 Example of Treatment Group Problem, Second Part.....	39
Figure 14 Example of Treatment Group Problem, the “Go on to Next Part” Button.....	40
Figure 15 Example of Treatment Group Problem, Third Part.....	41
Figure 16 Example of Treatment Group Problem, Fourth Part.....	42
Figure 17 Example of Treatment Group Problem, Fifth Part.....	43
Figure 18 Example of Treatment Group Problem, Hint Part.....	44
Figure 19 Example of Treatment Group Problem, Sixth Part.....	45
Figure 20 Example of Control Group Problem, First Part.....	46
Figure 21 Example of Control Group Problem, Second Part.....	46
Figure 22 Example of Post-Test Problem, First Question.....	47
Figure 23 Example of Post-Test Problem, Second Question.....	48
Figure 24 Value of Factors for Lab of Confidence Interval.....	51
Figure 25 Value of Factors for Lab of Hypothesis Test.....	51
Figure 26 Estimate of β 's.....	52
Figure 27 Post-Test Score vs. Middle Part Score, Treatment Group, Hypothesis Test Lab.....	53
Figure 28 Post-Test Score vs. Middle Part Score, Control Group, Hypothesis Test Lab.....	53
Figure 29 Post-Test Score vs. Middle Part Score, Treatment Group, Confidence Interval Lab.....	54
Figure 30 Problem for Confidence Interval, Population Mean, Pre-Test, Part 1.....	78
Figure 31 Problem for Confidence Interval, Population Mean, Pre-Test, Part 2.....	79
Figure 32 Problem for Confidence Interval, Population Mean, Control Group, Part 1.....	80
Figure 33 Problem for Confidence Interval, Population Mean, Control Group, Part 2.....	81
Figure 34 Problem for Confidence Interval, Population Mean, Post-Test, Part 1.....	82
Figure 35 Problem for Confidence Interval, Population Mean, Post-Test, Part 2.....	82
Figure 36 Problem for Hypothesis Test, Population Mean, Pre-Test, Part 1.....	83
Figure 37 Problem for Hypothesis Test, Population Mean, Pre-Test, Part 2.....	83

Figure 38 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 1.....	84
Figure 39 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 2.....	85
Figure 40 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 3.....	86
Figure 41 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 4.....	87
Figure 42 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 5.....	88
Figure 43 Problem for Hypothesis Test, Population Mean, Treatment Group, Hint Part.....	89
Figure 44 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 6.....	90
Figure 45 Problem for Hypothesis Test, Population Mean, Control Group	91
Figure 46 Problem for Hypothesis Test, Population Mean, Post-Test, Part 1.....	92
Figure 47 Problem for Hypothesis Test, Population Mean, Post-Test, Part 2.....	92
Figure 48 Problem for Hypothesis Test, Population Proportion, Pre-Test, Part 1	93
Figure 49 Problem for Hypothesis Test, Population Proportion, Pre-Test, Part 2	93
Figure 50 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 1	94
Figure 51 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 2	95
Figure 52 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 3	96
Figure 53 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 4	97
Figure 54 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 5	98
Figure 55 Problem for Hypothesis Test, Population Proportion, Treatment Group, Hint Part.....	99
Figure 56 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 6	100
Figure 57 Problem for Hypothesis Test, Population Proportion, Control Group	101
Figure 58 Problem for Hypothesis Test, Population Proportion, Post-Test, Part 1.....	102
Figure 59 Problem for Hypothesis Test, Population Proportion, Post-Test, Part 2.....	102

Chapter 1: Introduction

Science can be described as a “systematic enterprise of gathering knowledge about the world and organizing and condensing that knowledge into testable laws and theories” (Wilson, 1998). This enterprise necessarily requires methods for designing data collection and analyzing the resulting data. The science of statistics has become indispensable in formulating and implementing these methods, and therefore in the conduct of science. As such, statistics is very useful and important for college students in most of the science and engineering subjects. Knowledge of statistics is also important for all citizens. It plays an indispensable role in ensuring the quality of products and the safety and efficacy of medications. Business and government use statistics to make decisions that affect everyone’s lives. Voters must often consider evidence based on statistical methods in deciding civic issues.

However, instructing students in statistics is not an easy task. In the introductory statistics course, students are expected to understand how to identify different types of statistical studies, how to choose, calculate and interpret confidence intervals and how to choose, conduct and interpret hypothesis tests. They should also be able to apply these methods in realistic cases. The greatest difficulty is how to educate them to solve a novel problem. Previous experience of learning similar examples will help students to work out the new problem, but when the new one is a little different from their previous training, they are not usually able to solve it (Bassok, Wu & Olseth, 1995). The ability to figure out new problems depends on which features students learned from the old problems. Those who understand *structural features*, such as the underlying concepts, are more likely to be successful solvers than those who only remember *surface features*, such as the narration of the problems (Chi, Feltovich & Glaser, 1981).

Doing homework problems is an effective way for students to practice problem-solving skills. However, traditional paper-based homework has some shortcomings that limit the efficiency of learning. It takes a long time for instructors to grade homework

assignments, especially for large classes, and students may not be able to have their questions answered when they are working on the homework and receive feedback soon after they finish it. Also, the lag time between assignment and collection does not give instructors timely feedback on student performance. With the development of computer and Internet technology, web-based homework systems can provide a more interactive framework in teaching (Palocsay, Stevens, 2008).

WeBWork is such an online homework platform used mainly for mathematics and science. WeBWork has been used in a number of universities such as the University of Michigan, Johns Hopkins University, Dartmouth College and the University of Rochester. The flexibility of this open source Perl-based system allows implementation of innovative ideas to maximize the efficiency of homework. For example (“WeBWork documentation wiki,” 2010), WeBWork can provide students instant feedback to correct their answers, and give them chance to make multiple attempts at problems. Instructors can get real-time statistics which can help them design customized lesson plans to help students. They can also target areas of weakness in the understanding of individuals or groups of students. By allowing instructors to randomize data values or even the selection of problems seen by students, WeBWork can help discourage unauthorized collaboration.

Just as important as the technology used to present homework problems is the pedagogical design of those problems. Recent studies have proposed models of student learning that can be used to design more effective homework problems. The Subgoal Learning Model (Catrambone, 1998) recommends homework problems with cues to help learners find each step of a solution. This process helps students to remember the structural features which are more applicable than surface features in novel problems. The Backward Fading Model (Renkl, Atkinson and Maier, 2000) suggests homework problems should be given with a series of examples with gradually fading solution steps. This fading process will force students to learn how to apply concepts and formulas in given problems.

Our project mission has been to improve the teaching and learning of statistics by designing more effective homework problems. We have tried to achieve this by combining the latest cognitive research on student learning with the web-based interactive platform provided by WeBWork. By combining both of them, we have aimed to create more efficient and effective homework sets. To accomplish this, our group focused on five major goals:

- Identify the features of effective statistics homework problems and investigate effective ways of web-based implementation. Specifically, we have introduced a new methodology that combines the Subgoal Learning Model with our own modification of the Backward Fading Model, which we call the Forward Fading Model. We have developed a set of homework problems for introductory statistics that incorporate these features.
- Implement these problems in WeBWork
- Design experiments to assess the effectiveness of these problem sets Modify the homework problems based on the results of the experiments and the views of students about the problem sets

The end result is a set of homework problems for statistics that can benefit students and instructors and could serve as templates for future development by instructors and project teams.

Chapter 2: Background

2.1 Obstacles to Learning Statistics

Many students in college have troubles with learning statistical concepts. They often tend to respond to problems involving mathematics in general by falling into a “number crunching” mode, plugging quantities into a computational problem (Ahlgren, 1988). They might memorize the formulas and the steps to follow in familiar, well-defined problems but only seldom appear to get much sense of how concepts can be applied in new situations. More importantly, such shallow learning does not lead to genuine understanding.

There are more difficulties besides the “number crunching” mode. One is the students’ intuitive convictions about statistical phenomena. The second NAEP mathematics assessment produced evidence that students’ intuitive notions of probability seemed to get stronger with age but were not necessarily more correct (Ahlgren, 1988). On the other hand, Fischbein (1975) found decrements in probabilistic performance with increasing age, which he attributed to school experience and to scientific reductionism. Students’ intuitive ideas, presumably formed through their experience, may be reasonable in many of the contexts in which students use them but can be distressingly inconsistent with statistics concepts (Fischbein, 1975).

Another difficulty results from the fundamental difference between statistical thinking and mathematical thinking. Like most sciences, statistics is inductive: statistics starts from the particular and moves to the general. Mathematics is deductive: it moves from the general to the particular. Many beginning statistics students, who have been trained in mathematical thinking for many years, find statistical thinking difficult to master.

2.2 Two Learning Models

Catrambone (1998) and Renkl, Atkinson and Maier (2000) suggest several methods of helping students learn and those methods may be helpful for our project. Catrambone (1998) recommends homework problems with cues to suggest that students separate the solving procedure into a series of steps and work out the problem step by step. The Backward Fading Model proposed by Renkl, Atkinson and Maier (2000) is a teaching approach that consists in presenting the student with an example in an appropriate way.

2.2.1 Catrambone's Subgoal Learning Model

True understanding of a subject, such as statistics, involves an ability to apply it in new and unfamiliar settings. However, students usually have great difficulties solving novel problems. Therefore, they have difficulty in attaining true understanding. Training examples and problems are seldom sufficient preparation for solving novel problems involving several changes to specific examples students have seen, since students tend to remember a solution step by step without understanding the concepts underlying the steps. In this superficial way of learning, students lack of the ability to change the solution procedure and apply their experience in a new setting.

When working on problems, beginning students often focus on surface features which can easily distract them from underlying principles. A student faced with a new problem with surface features similar to a previous one is likely to try to solve the new problem based on the solution to the old one. Since it is not based on underlying concepts, this approach often fails.

According to Richard Catrambone (1998), good problem-solvers break a higher level goal into a hierarchical set of underlying steps. To develop this practice in students, he recommends building students' knowledge in a "meaningful hierarchical structure". This method asks learners to reconsider the problem at a higher level. By using this approach, Catrambone (1998) believes that educators could design better teaching methods to

help students learn the deeper conceptual knowledge and become better problem solvers.

The hierarchical organization in Catrambone's scheme encourages starting problem solving from a high level, such as the goal of finding the variance of some probability distribution. The high level goal will connect to a lower part of the hierarchy such as recalling the variance formula for a discrete distribution. Finally, the students will come up with a detailed solution as the goals become specific.

Catrambone defines a subgoal as representing a meaningful conceptual piece of an overall solution procedure. Working with subgoals usually reduces the complexity of the problem, allowing students to more easily solve difficult problems. Students trained in the subgoal method who try to solve a novel problem will first break the problem into several subgoals, and search their memory for similar subgoals encountered in previous practice. The hierarchical method directs students to find the difficult part of the problems because students will sort the subgoals by level. They will first only look for the high level subgoals, and try to adapt them to the new problem they are working on. In the adaptation process, students will consider lower hierarchical subgoals to get the specific solution.

To make this problem-solving procedure concrete, consider the following problem,

It is believed that a sample taken in a recent TV survey was representative of the American public. Individuals interviewed in the survey were asked whether they approved of Mr. Obama's presidency. Of the 10,000 responses, 5,500 people said "yes." Do the responses suggest that President Obama is doing better than 50-50? Explain your arguments using a confidence interval.

In order to solve this problem, students should first break it into subgoals. For example, which kind of interval they should use? A high level subgoal could be deciding between a confidence interval for one or two populations, and choosing from intervals for population means or proportions. Since there are only two possibilities for each step,

students will focus on only a few decisions and more easily figure out the answer and pick the correct interval.

Having selected the correct type of interval (here, an interval for a population proportion), students will think about a lower level subgoal: what kind of interval should they use; an exact interval or a large-sample interval? The ensuing tasks will involve calculating the numbers used in a large-sample confidence interval for a population proportion, for which students should be able to recall some much lower subgoals, such as how to compute an observed population proportion, the standard deviation of the binomial distribution, the z-value, and the quantiles of the standard normal distribution.

If this problem is presented as a training example, students are expected to obtain the solution based on the above subgoals. They should be able to differentiate between intervals for a population mean and for a population proportion, differentiate between the exact model and the large-sample approximation, and compute some statistical estimators.

Subgoals may apply to more than one problem. For example, the computation of the standard deviation for the binomial distribution can also be a subgoal for a totally different problem as long as binomial distribution and its standard deviation are involved.

Catrambone (1998) believes that directly stating the subgoals to learners is not the best practice, and is sometimes even ineffective. One reason is that students will tend to memorize these rules mechanically and fail to apply them correctly in practice. Catrambone asserts that instructors should embed the subgoals into examples, and let students discover the rules. Therefore, Catrambone proposed the Subgoal Learning Model, which can be summarized as follows:

1. One or more cues suggest to students a set of problem-solving steps for different subgoals.

2. After considering the steps, students will likely try self-explaining why the steps go together, thereby internalizing the concepts and methods involved.

The purpose of a cue is to help students break the problem into subgoals. For example, in the sample problem, asking students to explain why they chose the particular kind of interval would be a cue to isolate the part of how to choose a confidence interval. However, cues should be checked to ensure that students understand them correctly and construct the proper self-explanation.

2.2.2 Renkl, Atkinson and Maier's Backward Fading Model

Traditionally, textbooks and lecture notes have provided students with worked examples and step-by-step solutions. This option has always been available as students did homework with their class notes and textbooks opened. In order to solve problems, students may read worked examples as an aid. In fact, studying examples is considered a valuable way of learning new material. However, in this traditional way, students may develop a dependence on having a worked example at their fingertips, so they may encounter difficulties when they cannot refer to an example. In order to help students develop independent problem solving skills, Renkl, Atkinson and Maier (2000) give a new study model, called the Backward Fading Model.

In the Backward Fading Model the student is set the task of solving a series of problems, each accompanied by a worked example of the same type. For the first problem, the worked example gives the full solution. For the second problem, the example has a single step removed, forcing the student to recall the missing step through self-explanation. As the student works on more problems, the number of steps in the accompanying examples gradually decreases until the student can independently solve the complete problem. In this way, the Backward Fading Model connects example and problem, and helps students make the transition from example learning to independent problem solving skills.

In order to test the effectiveness of this new model, Renkl, Atkinson and Maier (2000) conducted an experiment on two ninth-grade classes. They gave the two classes a pre-test, and applied the Backward Fading Model in one class. In contrast, the other class used their traditional study method. After a period of learning, both classes completed a post-test and Alexander et al. compared their score increases. The result showed that there is a statistically significant improvement of students who used the Backward Fading Model compared with students who used the traditional method.

2.3 The Role of Homework

2.3.1 The Benefits of Assigning Homework

Homework has been an important teaching strategy for a long time (Cooper, 2008). Large scale reviews of educational research show that in all subjects and at all grade levels homework has a positive impact on student learning outcomes (Bonham, Deardorff & Beichner, 2003). Reasons might include:

- Students get deeper understanding of what they have learned by doing homework.
- Homework fosters independent learning and responsible character traits and it develops an interest in learning.
- Homework gives both students and instructors feedback on student progress.
- Homework forces students to learn to use resources such as libraries, reference materials, encyclopedias, and the internet.

2.3.2 Paper-Based Homework and Web-Based Homework

Instructors have traditionally relied on the assignment of paper-based homework to motivate and guide student learning in the hours between meetings. However, the rapid development of computer and internet technologies has introduced new approaches to

teaching and learning. Among these new approaches, online web-based education has become a promising field.

For some instructors, there is uncertainty about how to go about using homework. The need to assign, collect, and grade problem sets places high demands on the instructor's time and may make it difficult for the instructor to monitor student homework performance in large lecture classes. Faced with this situation, many instructors assign "suggested problems" without collecting them, but conventional wisdom says this kind of homework set is less helpful than required and collected homework. Another solution is to give homework grading duties to teaching assistants or peer learning assistants. One downside of this approach is that the instructor cannot monitor student performance closely.

However, web-based delivery of homework offers a possible solution. With web-based homework, students must submit their solutions to homework problems online. And although there is some time cost to instructors in setting up the system and in downloading grades, the time spent with red pen and stacks of student papers is largely eliminated. In addition, instructors can closely monitor the progress of individual students as well as the whole class as the homework assignment is being worked on. This allows them to pinpoint problems with teaching and learning in a timely manner.

One study comparing the effectiveness of paper-based homework and web-based homework in physics classes concluded that there is no substantial difference in student grades between the two methods (Demirci, 2007). However, there is also evidence that students "game" web-based homework and that when they do so, they pick up unproductive "novice-like" habits (Demirci, 2007). For example, some students will recklessly work out first draft answers in order to get the feedback from the computer for a second try. Possible remedies to this problem could be loss of points for excess submissions, limiting the number of submissions allowed, or basing the students' scores on an average of all submissions.

2.3.3 Web-Based Homework Systems

A number of Web-based homework systems are currently available.

WeBWork is an online homework platform used mainly for mathematics and science. This open source Perl-based system has advantages over the customary paper-based homework because of its flexibility, which allows implementation of innovative ideas to maximize the efficiency of homework. For example (“WeBWork documentation wiki,” 2010), WeBWork can provide students with instant feedback to correct their answers, and give them a chance to make multiple attempts until they succeed. Instructors can get real-time statistics which they can use to design customized lesson plans for individual students or the entire class. WeBWork also has features, such as randomization of problem assignment or of data in problems, to discourage dishonest behavior.

Tycho (Tycho), a powerful computer package which enables instructors to put course materials on-line, has two main components, a grade book and a platform for assigning homework. The grade book provides both students and instructor with secure access to student progress in the course from any internet browser.

WebClass (WebClass) is a website providing an interactive environment for class homework and diagnostic testing. The Web Homework System (WHS) distributes homework assignments with immediate feedback for the results of student work as well as providing an authoring and class management environment for the assignments themselves.

ASSISTments (ASSISTments) is a Web-based tutoring program for 4th to 10th grade mathematics. The word “ASSISTments” blends tutoring “assistance” with “assessment” reporting to teachers. ASSISTments increases instructional efficiency by simultaneously testing students and tutoring them on items they get wrong. The system is adaptive in that it can use information about student ability and knowledge to target assistance appropriate to the student. The ASSISTments system gives teachers fine-grained

reporting on roughly 120 skills that the system tracks per grade level. Teachers can use this detailed assessment data to adjust their classroom instruction and pacing. The key point of this system is that students get instant interactions to help on what they have trouble with and instructors can log on to the System and study detailed reports about their students' difficulties and strengths. Unlike WeBWork, however, ASSISTments is not a hands-on tool that instructors can use to develop, implement and manage their own problem sets.

2.3.4 The WeBWork System

2.3.4.1 WeBWork at Worcester Polytechnic Institute

WeBWork is now used by a number of colleges in United States, mostly to assign homework in mathematics, physics and chemistry courses. Over the past three years instructors in the Mathematical Sciences Department at Worcester Polytechnic Institute have chosen the WeBWork system for use in calculus and statistics.

Our group interviewed Professor Bill Farr of WPI's Mathematical Sciences Department to learn about WeBWork and its history at WPI. Professor Farr is the person who introduced the WeBWork system to WPI. We conducted the interview at his office on September 17, 2010.

Professor Farr has been using WeBWork since the spring of 2007. He pointed out that WeBWork is a free and open-source system that gives the instructor total control of the content of homework sets. He also mentioned that WeBWork gives the instructors instant feedback on student performance, which can help them identify difficulties students are having with the course material. Moreover, Professor Farr thought that the instant feedback to the student provided by WeBWork combined with its ability to provide students more than one try, gives students more time to think about the problem and the concepts and formulas behind the problem; with traditional paper-based homework, students do not generally think about problems after finishing since they only can get feedback when instructors return the graded papers.

When asked to compare WeBWork to competing systems, Professor Farr stated that he could not do so since he had never used any other Web-based system for teaching. When asked whether he found any drawbacks to WeBWork, Professor Farr said that WeBWork might not be easy to start with for most instructors who were first time users. The key problem is that if instructors do not like to use existing problem sets shared online, they need to input their own homework sets or modify existing homework which may require learning the WeBWork programming language.

2.3.4.2 WeBWork at Other Universities

A number of studies have been conducted to assess WeBWork's effectiveness. Hauk, Powers, Safer and Segalla (2004) conducted a study to test the effectiveness of WeBWork in Fall 2002 with 644 students enrolled in 19 sections of college algebra. Twelve WeWork-based homework sections were taught by 11 instructors and enrolled 408 students. The control group consisted of seven sections randomly chosen with 236 students and was taught by seven instructors. They also assigned instructors in such way that each of the three instructors who taught multiple sections of the course had at least one control group section and one WeBWork-based homework section. A multiple choice test at the beginning of the term and the same type of test in end of the term were used to evaluate the performance of students. The result showed that there was no statistically significant difference between the control group and the *WeBWork* sections, with or without controlling for SAT scores. Also no statistically significant differences were found when the data were sorted by socio-economic status or ethnicity. However, there was a significant difference between control and treatment section for women with women in treatment section performing better on the post-test, $p=0.045$.

Dedic, Rosenfield, and Ivanov (2008) also did a study on WeBWork's effectiveness. The study was done in the Fall of 2006 with 354 students in nine classes studying Calculus I for social science majors in Concordia University. All classes had the same lectures and assigned homework problems. Three of the nine classes were randomly

assigned to each of three conditions: condition 1, a control group where students had paper and pencil homework; condition 2 in which the homework was submitted in WeBWork; condition 3, which added one hour per week in lab on the WeBWork problems. The researchers considered several measures of student performance and attitude.

The mean of each measure had no statistically significant difference between condition 1 and condition 2 but for each measure there was a statistically significant difference between the means for the first two conditions and that for condition 3. The students in condition 3 showed better achievement and a greater improvement in sense of self-efficacy than those in either condition 2 or 3. This last was especially notable for women.

2.4 Interaction in Learning

2.4.1 The Importance of Interaction

Many educators point out the importance of interaction in high quality online education. For instance, Shale and Garrison (1990) state that interaction is “education at its most fundamental form.” In addition, Palloff and Pratt (1999) argue that “key to the learning process are the interactions among students themselves, the interaction between faculty and students, and the collaboration in learning that results from these interactions.” A sage in the field of distance education, Moore (1992), points out that increasing the interaction between learner and instructor can lead to a smaller transactional distance and more effective learning.

2.4.2 Types of Interactions

A well-recognized classification of interactions was offered by Moore (1989). Her three-part interaction scheme included: learner-instructor, learner-learner, and learner-content interaction.

The first type of interaction considered important by many educators is interaction between the learner and instructor. Instructors usually have a teaching plan and try their best to motivate students to study course materials. The instructors introduce materials, skills or concepts through presentations and demonstrations. They follow up with activities to reinforce learning through practice or reflection. In addition, they provide help and support for students who encounter difficulties. The instructor is especially important in helping students learn to apply new knowledge (Moore, 1989). Even self-motivated students may need the help of the instructor to show the possible range of applications of knowledge and to ensure that the applications are done correctly. It is for real help on concepts and feedback that interaction with an instructor is likely to be most valuable.

The second form of interaction is between one learner and other learners: what Moore (1989) calls inter-learner or learner-learner interaction. Historically, for reasons of convenience and economics, formal classroom instruction has been organized in the learner-instructor mode. However, learner-learner interaction among members of a class can be a valuable, and sometimes essential, learning resource. Since group work is essential for functioning in modern society, Phillips et al. (1988) taught principles and helped students in learning effective group interactions.

Another benefit of inter-learner interaction is competition between learners which drives students to study hard (Moore, 1989). For younger learners, the teaching task of stimulation and motivation will be assisted by peer-group interaction, though this is not as important for most adult and advanced learners, who tend to be self-motivated.

Moore's third type of interaction is interaction between the learner and the content or subject of study. All education heavily involves this type of interaction since this interaction with content will change learner's understanding and ideas. Some learning programs are solely content-interactive in nature; for example, distance learning that relies on one-way communication from written materials. Such a learning program is content-interactive and the learner gets hardly any feedback. According to the findings

of adult education research, the majority of the adult population undertakes self-directed study which only involves learner-content-interaction (Tough, 1971; Penland, 1977; Hiemstra, 1982). Therefore, Educators should design more of this type of interaction for people.

Since Moore (1989) suggested these three interactions, a new type of interaction has arisen (Su, Bonk, Magjuka, Liu & Lee, 2005). Given the technology-mediated nature of online education, learner-interface interaction is considered to be another important type of interaction. Hillman, Willis and Gunawardena (1994) point out that this type of interaction occurs between the learner and the technology used for online education. They further point out that it can be one of most challenging types of interaction due to the fact that people need to adapt to the new technology.

2.5 The PG language

All WeBWork problems are read from text files that are written in a language called PG, for Problem Generation (Release notes for PG 2.3.1), and stored on a WeBWork server. Thus, to create new problems, one has to create or edit a text file and ensure that it is on the server in a location that is accessible to WeBWork.

The PG language is a programming language with a collection of macros written in Perl, and providing some features of the LaTeX document markup language. These macros can input data and mathematical formulas, and compute and output the results. They can provide an interface for displaying problems to students, handle student responses in a number of formats, and evaluate those responses and output answers or other material for students to see.

WeBWork is a Perl-based system, or specifically speaking, the WeBWork system will load some Perl packages to implement their functions. Normally, a problem developer only needs to write problems in the PG language which in turn calls Perl macros to do the input, output, display and processing. Those macros are contained in Perl packages that are loaded when a WeBWork session is started. For self-designed advanced

functions, the WeBWork system permits users to include their own Perl packages. In this project, our group designed new Perl packages to implement the previously mentioned learning models.

Chapter 3: Methodology

The main goal of this project was to improve the effectiveness of homework sets in introductory statistics using Catrambone's Subgoal Learning Model (1998) and Renkl, Atkinson and Maier's Backward Fading Model (2000) implemented through the WeBWork system. Professor Petrucci had developed many homework sets in WeBWork. Those homework sets covered all the topics presented in WPI's introductory statistics sequence MA 2611-12. He had used them for homework assignments in his sections of those courses over the past year and a half.

Our group worked on designing homework sets and implementing them into WeBWork during the second half of A term and the first half of B term. In the second half of B term we did controlled experiments to assess the effectiveness of the new homework sets. Our team had a set of goals that we strived to complete during our project. The general workflow is outlined in Figure 1. The specific goals were:

- Outline the scope of statistical models and concepts of estimation and hypothesis testing that students are expected to understand in the introductory statistics course at WPI, MA 2611
- Identify the major problems students have in learning these concepts
- Learn PG, Perl and LaTeX
- Develop packages to extend WeBWork's capabilities
- Propose improved homework problems in accordance with Subgoal Learning Model and the Backward Fading Model
- Implement the homework problems into the WeBWork system.
- Design and conduct a randomized controlled experiment to test the new homework problems
- Analyze the experiment and draw conclusions.
- Revise the problems in light of the experimental results and observations of student-problem interaction
- Write the final report.

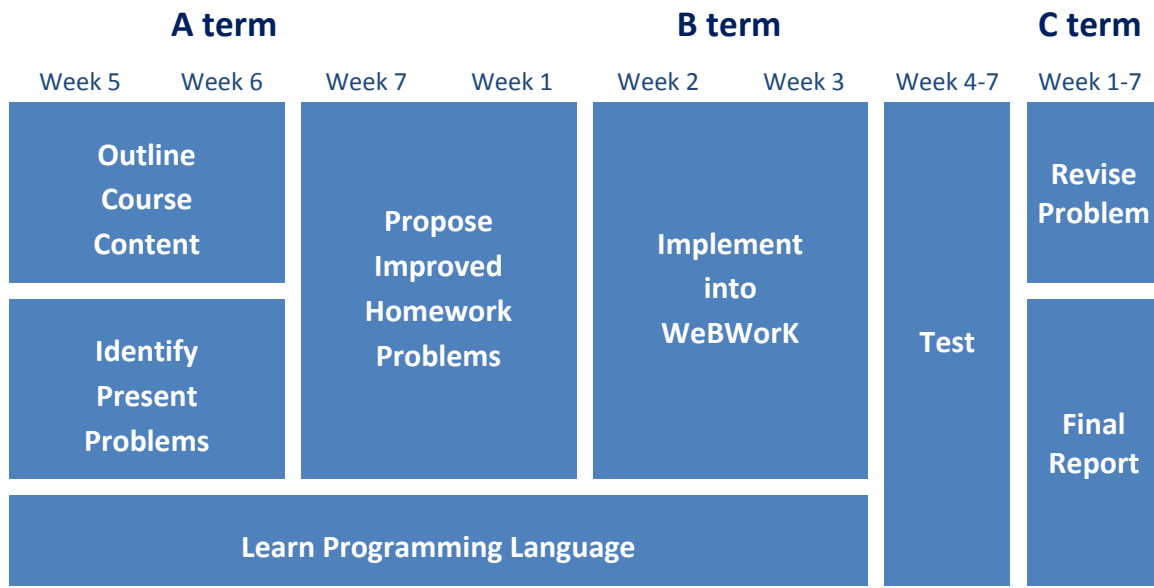


Figure 1 Project Timeline

3.1 Outline Course Content

Because different instructors had various requirements for students in introductory statistics courses, it was necessary to define a specific range of content for the exercises we would design. We first determined an outline of the statistical models and concepts we were going to include in designing new homework problems. In our project, we designed homework exercises for introductory statistics, focusing on the methods of confidence intervals and hypothesis tests for (1) a population mean (2) a population proportion, (3) for comparing two population means and (4) for comparing two population proportions. Because the problems were designed to be mainly used at WPI, we based the content on the syllabus for WPI’s introductory statistics course, MA2611. To complete a detailed outline, we consulted present statistics course instructors at WPI and became aware of their specific expectations for students.

This outline represented information about the possible problems on homework and exams, because the type of homework usually reflected the instructors’ expectations. For example, Professor Petruccelli expected that students should not only be able to compute appropriate answers, but also be able to explain which model to use and

interpret the results. Indeed, comprehension of statistical concepts was more important than calculation, and this fact suggested to us that we should design new exercises to help students improve their understanding. Our group continued collecting information from more instructors to obtain a specific and accurate content outline, which guided us in designing new problems.

In order to visualize the knowledge required for introductory statistics, we built a “Knowledge Tree”. A rudimentary Knowledge Tree is shown in Figure 2. Later, we added more details to this tree in order to list the required knowledge in an organized way.

This information was very important for our project. It showed us how the concepts in introductory statistics are related. We referred to this tree extensively when we tried to propose a better way to help students learn statistics.

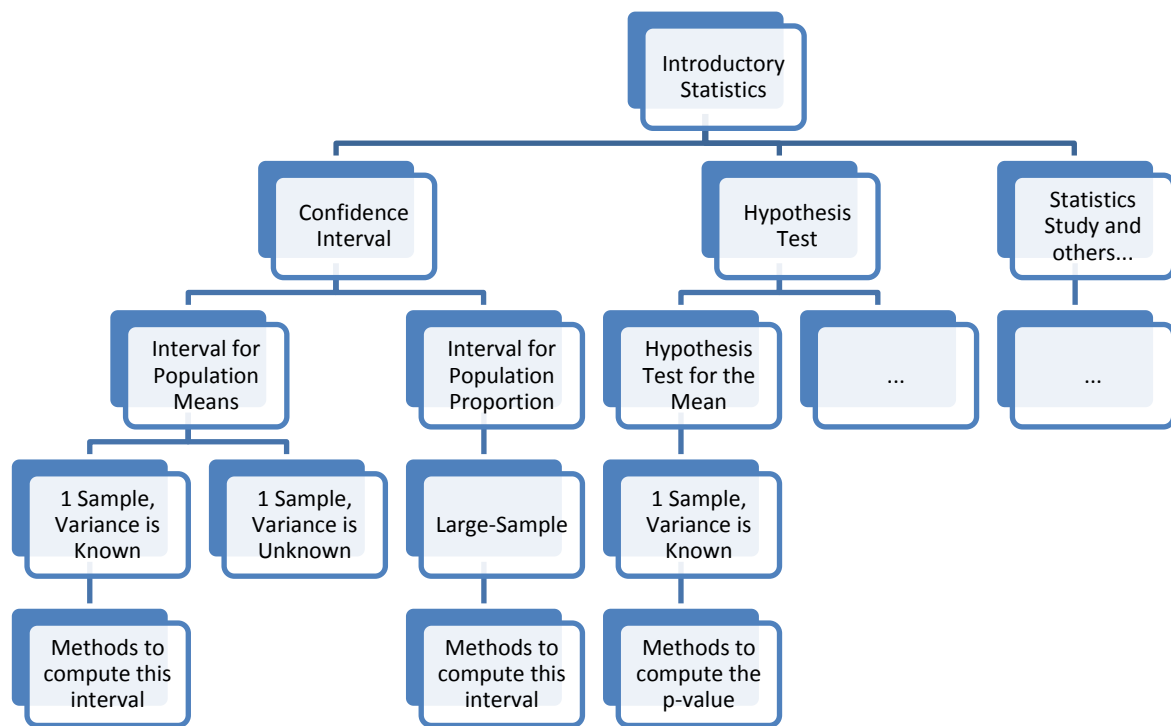


Figure 2 Rudimentary Knowledge Tree for Introductory Statistics

3.2 Identify Present Problems

Instructors' experiences in teaching statistics for many years provided us much useful information to improve the homework problems. Our group interviewed some instructors who had taught introductory statistics to learn the common questions previous students have had. Collecting and analyzing previous student questions was an effective way to have a general idea of students' difficulties in learning statistics.

Our project proceeded by thoroughly discussing the difference between students' performance and instructors' expectations. We tried to understand students' questions and difficulties in detail, and collected specific information about what concepts gave students difficulties and how instructors tried to help them previously.

Also, we classified students' difficulties and located them in the "Knowledge Tree". When students try to work out a homework problem, their solution process should be similar to going down in the knowledge tree. For example, when students try to solve the problem mentioned in section 2.2.1, they are going from "Confidence Interval" to "Large-Sample", then to the "Methods to compute this interval" in this tree. If students don't know how to solve the problem or give a wrong answer, they must have some difficulties in this process. We could ask more questions, such as "Which confidence interval should you use?" to locate their difficulties in the tree. With the location of students' difficulties, we found whether students had questions at a high or low level in the hierarchy. This information was useful when we proposed improved homework problems. We got some information about the common difficulties for students in introductory statistics from statistics professors in the WPI Mathematical Sciences Department, and we also got some information from student-problem interactions during the conduct of our experiments that we can use to revise our WeBWork problems.

3.3 Programming Language

To develop and implement well-designed homework problems in the WeBWork system, we became acquainted with the PG language, which was used to program WeBWork problems. The PG language is used exclusively in the WeBWork system for all processing of text, answer checking, response to students, keeping track of results, etc. In addition, we learned the Perl language to implement our Subgoal Learning Model and Forward Fading Model into the WeBWork system because the WeBWork system did not support the kinds of interactive features these models require. However, we learned that we could endow WeBWork with the requisite functionality by writing a Perl-based package. To learn how to do this, we first found a relatively fundamental Perl package for WeBWork to use as a template.

By developing our own Perl packages for WeBWork, we were able to add more interactions to help students with homework problems, including giving different levels of examples or hints depending on student performance on the original problem. The system could adapt to students' abilities and give proper help, but students would always be able to choose to see the help.

At the beginning of our project, the WeBWork system didn't support many of the features we planned for the new homework problems. For example, the original system could not display different parts of a question step by step, so we were unable to design many-part questions or incorporate sequential, student-selected hints.

However, the open source WeBWork system has much freedom to let the users design what they want to do. The problem design language in WeBWork system, the PG language, is based on a collection of packages written in the Perl scripting language. These packages can tell the WeBWork system how to display the homework problems, give a score, record students' grades, etc. WeBWork also allowed us to load our own packages when designing the problems in order to add more features into WeBWork system. Because the packages could handle HTML codes, they could help the WeBWork system to achieve almost any feature in the common Internet webpage.

display the proper information for the given part. The *improvedCompound* object takes care of maintaining the data as the parts change. In order to handle the scoring of the problem as a whole when only part is showing, the *improvedCompound* object uses its own problem grader to manage the scores, and calls the WeBWork grader when needed. One can specify a different value for the variables above using the `$cp->useGrader()` method.

3.4 Proposed Improved Homework Problems

Due to the limitations of time and number of students, we combined the Subgoal Learning Model and the Forward Fading Model together to design new homework sets.

3.4.1 Subgoal Learning Model

The Subgoal Learning Model (Catrambone, 1998) gave us a possible framework to design new problems. In order to apply this model in introductory statistics suitably, we got some advice of experienced statistics instructors. Our group designed new problems generally on the basis of the Subgoal Learning Model, but in details we made some adjustments with advice from instructors.

In brief, our methods of implementing the Subgoal Learning Model were the following:

- Classify and locate students' difficulties in the knowledge tree
- Point out the hierarchical knowledge according to the locations of difficulties in the knowledge tree
- Propose homework problems that concentrate on their corresponding hierarchical knowledge

We mainly considered two attributes of students' difficulties. The first attribute was the concepts involved, and the second attribute was the type of difficulty such as making a choice or interpreting a result. Each difficulty had both a first attribute and a

second attribute. For example, failing to choose a correct confidence interval included the first attribute about “confidence interval” and the second attribute about “making a choice”, while failing to correctly interpret the p-value involved the first attribute about “hypothesis test” and the second attribute about “interpreting”.

Problems with the same first attributes usually shared the same or similar high level hierarchical knowledge, which was closely related to the high level nodes of the “knowledge tree”. It was necessary to investigate the first attribute in detail, and the tree helped us to build the high level hierarchy. The first attribute for a problem was the underlying concepts and formula for this problem. For example, two problems sharing the same first attribute, might both ask a student to conduct a hypothesis test for one population mean.

The second attribute was more essential when we proposed solutions for students’ problems. The second attribute was about whether students could interpret the problem correctly and choose the proper formula for the problem. For difficulties corresponding to different second attributes, we developed methods different than those we proposed for the first attribute, because, for example, explaining how to make a choice between models was qualitatively different than giving an interpretation of the models. However, there were some similarities between choosing a confidence interval and choosing a hypothesis test model, so a template designed for helping students making a choice could be used to solve both these problems which had different first attributes but the same second attributes.

3.4.2 Forward Fading Model

We devised a modification of the Backward Fading Model that we called the Forward Fading Model. Instead of giving a fading process on examples to help students, the Forward Fading Model gave students more information about solution steps gradually.

For example, for a given problem, students would first see a problem with an isomorphic example without solutions. For students who worked out the problem answer, no extra help was shown on the WeBWorK system. Students who could not find the answer would be able to click a button to ask for hints and the first solution step of the example problem would be given in the WeBWorK system. If the student still could not solve this problem, the system asked students whether they needed the second solution step for example problem. This process would continue until the student solved the problem, or until all solution steps in example problem were shown. Hopefully, by that stage the students were able to work out the problem by following the steps in example problem.

As another model for designing problem sets, the Forward Fading Model needed a process to design problems and combine them with the model. As a first step, we found examples isomorphic to the types of problems the students would be assigned. This was not difficult, as these problems were numerous and standard in introductory statistics. We required problems to have at least three solution steps to ensure they would provide a sufficient level of help. We also required that the fading parts of the solution steps involved key concepts instead of simple calculation.

We also considered design issues to make the homework sets more efficient. First, we consulted instructors and other resources to get information about students' difficulties in understanding each concept. We used the resulting information to address specific areas of student difficulty.

3.4.3 Combining the Two Models

We got some advice from instructors and revised these two models to make them easier to apply. We combined two models in the multi-part problem instead of the traditional WeBWorK problem.

For a multi-part problem, we gave students several problems as given subgoals in a sequence. Those subgoals were keys to understanding and solving the problem. Usually,

subgoal problems are multiple-choice or fill-in-blank problems to help students emphasize concepts they learned.

Our forward fading process was that every subgoal result was shown in subsequent parts of a multi-part problem. That is, when students went to a new part of the problem, all solutions of previous subgoals were shown in the problem statement part as help.

3.4.4 Implementing the Model

While the Forward Fading concept is quite general and can be applied in many settings, our implementation of the forward Fading Model was based on a five step procedure for confidence interval problems and hypothesis test problems that was taught in class by Professor Petruccelli. It therefore served to reinforce the methodology and specific topics and approaches being taught in class. Our implementation was also confined to two specific statistical problems: inference about a population mean and inference about a population proportion.

For a confidence interval mean problem combining the Subgoal Learning Model and the Forward Fading Model, the student would first see the problem statement with a multiple choice problem asking about the scientific goal (Figure 3). In hypothesis testing problems, this problem statement asked about the scientific hypothesis.

WPI sets an expectation of 17 hours work per week for a $\frac{1}{3}$ unit seven week undergraduate course, which equals 119 hours per term. To see if reality meets expectations, a WPI project team took a random sample of 7 $\frac{1}{3}$ unit courses and from each course, randomly selected one student volunteer to follow during the term. Each selected student agreed to keep track of the hours spent on coursework for that course. The totals for the term are below:

90 93 80 88 98 119 86

The project team was interested in determining how the average time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course compares with the 119 hour guideline.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

- A. Estimating the average hours per term spent by WPI students on a $\frac{1}{3}$ unit seven week undergraduate course
- B. Estimating the minimum hours per term spent by WPI students on a $\frac{1}{3}$ unit seven week undergraduate course and comparing with the WPI expectation.
- C. Determining how the average time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course compares with the 119 hour guideline.
- D. Finding out which WPI courses are the most work.

Figure 3 Example Problem: the First Part

If the students gave the correct answer, they were allowed to go to the next part of the problem by clicking a button. If the students chose the wrong answer, they could continue to try to solve problem. Students could click on “go to the next part” button if they got the correct answer. Otherwise, they were forced to go to the next part of the problem after they reached the maximum attempts allowed for this part of the problem. This transition procedure was used repeatedly.

The next part of problem would be the same problem statement as before but with a different question (Figure 4). There was an additional part of the problem statement which was the answer to part 1. This was the forward fading part of the process.

90 93 80 88 98 119 86

The project team was interested in determining how the average time spent by all WPI students on a 1/3 unit seven week undergraduate course compares with the 119 hour guideline.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

Determining how the average time spent by all WPI students on a 1/3 unit seven week undergraduate course compares with the 119 hour guideline.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. What are the measurements in this problem, and what is the Statistical Model?

- A. The hours WPI students spend on 1/3 unit seven week undergraduate courses per term; $N(\mu, \sigma^2)$
- B. The hours WPI students spend on 1/3 unit seven week undergraduate courses; $b(n, p)$
- C. The hours WPI students spend on 1/3 unit seven week undergraduate courses; $N(\sigma, \sigma^2)$
- D. The hours WPI students spend on 1/3 unit seven week undergraduate courses per week; $N(\mu, \sigma^2)$

Figure 4 Example Problem: the Second Part

We used this idea to help students understand the formula and theory. The question of part 2 was about the statistical model (Figure 4). The same transition procedure used to move the students from part 1 to part 2 of the problem was also used to move them from part 2 to part 3 (Figure 5). Part 3 of the problem contained problem statement and also the correct answers to the previous two parts of the problem as the forward fading process.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

Determining how the average time spent by all WPI students on a 1/3 unit seven week undergraduate course compares with the 119 hour guideline.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The hours WPI students spend on 1/3 unit seven week undergraduate courses per term;
 $N(\mu, \sigma^2)$

3. The Model Parameter to Be Estimated:

At this point we examine how to achieve the Scientific Goal in terms of the statistical model. In this problem, the Scientific Goal is to find how many hours, on average, WPI students spend per term on coursework. This suggests that the model parameter we want to estimate is:

- A. the population proportion
- B. the variance
- C. the population mean
- D. something else

Figure 5 Example Problem: the Third Part

Determining how the average time spent by all WPI students on a 1/3 unit seven week undergraduate course compares with the 119 hour guideline.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The hours WPI students spend on 1/3 unit seven week undergraduate courses per term;
 $N(\mu, \sigma^2)$

3. The Model Parameter to Be Estimated:

At this point we examine how to achieve the Scientific Goal in terms of the statistical model. In this problem, To achieve the Scientific Goal, the project team needs to estimate the average time spent by all WPI students on a 1/3 unit seven week undergraduate courses to find how many hours, on average, WPI students spend per term on coursework. This suggests means that the model parameter we want to estimate is: **the Population Mean.**

4.1 Point Estimates:

Obtain a single number estimate of the average number of hours per term spent by WPI students for a 1/3 unit seven week undergraduate course.

$\bar{y} =$

Figure 6 Example Problem: the Fourth Part

We applied the Subgoal Learning Model in such a way that for each problem, we supplied several questions step by step which were the subgoals of the problem. For part 4 (Figure 6), students were expected to find the point estimates for the confidence interval (the test statistic for hypothesis test problems).

At this point we examine how to achieve the Scientific Goal in terms of the statistical model. In this problem, To achieve the Scientific Goal, the project team needs to estimate the average time spent by all WPI students on a 1/3 unit seven week undergraduate courses to find how many hours, on average, WPI students spend per term on coursework. This suggests means that the model parameter we want to estimate is: **the Population Mean.**

4.1 Point Estimates:

The single number estimate of the average number of hours per term spent by WPI students for a 1/3 unit seven week undergraduate course is:

$$\bar{y} = 93.42$$

4.2 Interval Estimates:

Estimate this same quantity using a 95% confidence interval.

$$\boxed{} \leq \mu \leq \boxed{}$$

Note: *You can earn partial credit on this problem.*

Note: *This problem has more than one part.*

You have tried 0 of 3 times in this part.

*Your score for this attempt is for this part only;
your overall score is for all the parts combined.*

Show hint

Figure 7 Example Problem: the Fifth Part

For part 5 (Figure 7), students were expected to give the precise interval estimates of the confidence interval (the p-value for hypothesis test problems).

HINT:

The following is an example problem which is similar to the above problem you are working on. You can look at the solution of the following example and it will help you to solve the original problem.

A soccer club wants to know whether players average more than 75 training hours per month. The coach randomly selected four players and recorded their training hours every day for a month. Suppose the resulting data are 80,70,65,80. In the following, we will construct a 95% confidence interval for the mean.

First, we find the sample mean

$$\bar{y} = \frac{y_1 + y_2 + y_3 + y_4}{n} = \frac{80 + 70 + 65 + 80}{4} = 73.75$$

and the sample standard deviation,

$$\begin{aligned} s &= \sqrt{\frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + (y_4 - \bar{y})^2}{n-1}} \\ &= \sqrt{\frac{(80-73.75)^2 + (70-73.75)^2 + (65-73.75)^2 + (80-73.75)^2}{3}} \\ &= 7.5 \end{aligned}$$

Next, we find the appropriate quantile of the t distribution. Since the sample size is $n = 4$ and the confidence level is $L = 0.95$, the desired quantile is $t_{n-1, (1+L)/2} = t_{3, 0.975} = 3.1824$

The confidence interval is then computed as

Figure 8 Example Problem: the Hint Part

In part 5, students could click on a “hint” button to see a solved problem that was isomorphic to the original problem (Figure 8).

5. Results and Interpretation:

The 95% confidence interval is (90.79, 120.34). What can the project team conclude about how the average time spent by all WPI students on a 1/3 unit seven week undergraduate course compares with the 119 hour guideline?

- A. The interval (90.79, 120.34) lies entirely below the WPI expectation, therefore the project team cannot conclude with 95% confidence that the average level of work expended is below the WPI expectation.
- B. The project team is almost sure all students at WPI spend more than 65 hours per term.
- C. The interval (90.79, 120.34) lies entirely below the WPI expectation, therefore the project team can conclude with 95% confidence that the average level of work expended is below the WPI expectation.
- D. Since 80 is not in the confidence interval, no one at WPI spends more than 80 hours.

Figure 9 Example Problem: the Sixth Part

For part 6 (Figure 9), students needed to answer a question about interpretation of the result.

We treated the whole process as a Forward Fading Model since we asked questions at each step and this process would give students more information step by step.

3.5 Randomized Controlled Experiments

In order to test the effectiveness of our homework problems, we conducted two randomized controlled experiments in WPI's introductory statistics course MA2611 in B term, 2010. The experiments were conducted during two different lab periods: the first on the topic of confidence intervals and the second on hypothesis tests. Each lab period allowed two hours for testing.

In each experiment, there were two factors: type of assistance and statistical model. The main factor of interest was the type of assistance. This factor had two levels, treatment and control. Students in the control group received standard WeBWork homework sets of the type currently used in the course with no extra help; students in

the treatment group received homework sets with the implementation of the combined Subgoal Learning Model and Forward Fading Model. The second factor was statistical model. We prepared two types of problem: one involved statistical inference for a population proportion, and the other statistical inference for a population mean.

There were five lab sections in the course with approximately 25 students in each. The experiment was conducted on two occasions, one week apart. On the first occasion, problems on estimation and confidence intervals were tested; on the second occasion, problems on hypothesis testing were tested. Within each lab section on the first occasion, we randomly assigned students to the four assistance type and statistical model combinations. On the second occasion, we assigned students to the opposite levels of both factors. So, for example, students who were in treatment/mean on the first occasion were assigned to control/proportion for the second. All problems were presented and all responses obtained using the WeBWork system.

During the whole lab period, students were not allowed to consult books, notes or the web because we wanted to reduce extraneous sources of variation, but they were given access to a PDF file containing all required formulas, and a Z-table and t-table they might need for problems.

The experimental protocol was as follows. We used the problem set we designed for confidence interval of population proportion as examples.

3.5.1 Pre-Test

Each student was first given a short pre-test on that lab's material. The pre-test had two problems. The first problem of the pre-test was a fill-in-blank problem for finding the exact confidence interval (first lab occasion: confidence intervals) or the final p-value (second lab occasion: hypothesis tests). The second problem was a multiple-choice problem requiring interpretation of the result of the first problem. All students assigned the same statistical model received the same problems in the pre-test, whether they were in the control or treatment group.

A local Boston magazine wants to know whether most people in Boston support Boston Celtics rather than other teams. They randomly asked 10 people living in Boston, "Which NBA team is your favorite?" 8 of 10 said they supported Boston Celtics, and 2 of 10 people said they liked other teams.

Use a 95% approximate score confidence interval to estimate the proportion of the population that likes Boston Celtics.

$$\boxed{} \leq p \leq \boxed{}$$

Note: *You can earn partial credit on this problem.*

Show correct answers

Preview Answers

Check Answers

Figure 10 Example of Pre-Test Problem, First Question

Figure 10 shows the first problem assigned to students being tested on confidence intervals for proportions. Students in both control and treatment groups received this as the first problem in the pre-test.

A local Boston magazine wants to know whether most people in Boston support Boston Celtics rather than other teams. They randomly asked 11 people living in Boston, "Which NBA team is your favorite?" 11 of 11 said they supported Boston Celtics, and 0 of 11 people said they liked other teams. Based on these results a 95% approximate score confidence interval for the proportion of Bostonians who like the Boston Celtics is $(0.672, 1)$.

If the historical record shows that 90% people in Boston support Boston Celtics, which of the following can the editors of the magazine conclude?

- A. Based on this interval, the editors of the magazine cannot conclude with 95% confidence that the proportion of Bostonians who support Boston Celtics differs from the historical value of 90%.
- B. They cannot conclude anything since the sample size is small compare to the population of Boston.
- C. Since 0.5 lies below our confidence interval, the editors of the magazine are 95% confident that only half of Bostonians support the Boston Celtics.
- D. Based on this interval, the editors of the magazine can conclude with 95% confidence that the proportion of Bostonians who support the Boston Celtics is equal to the historical value of 90%.

Figure 11 Example of Pre-Test Problem, Second Question

Figure 11 shows the second problem assigned to students being tested on confidence intervals for proportions. Students in both control and treatment groups received this as the second problem in the pre-test. Students being tested on confidence intervals for a population mean received analogous pre-test problems (see Appendix A).

3.5.2 Treatment Group

After the pre-test, the student was given the appropriate homework set in WeBWork as practice. Students in the control group worked on a set of problems that were similar to problems in the pre-test, and to what they might encounter in a homework assignment, with no extra help. Students in the treatment group worked on a set of problems that were the same as the problems for the control group, but with

the implementation of the combined Subgoal Learning Model and Forward Fading Model. Students in both groups were allowed two tries.

Historically, a producer of high-quality jewelry has found that 10% of the top-grade diamonds it buys have serious defects after cutting and polishing. To monitor quality, a diamond company inspects a random sample of 10 recently processed diamonds from the latest shipment. They find that 1 have serious defects. Based on these results, the head of the quality division wants to obtain a single number estimate of the proportion of defective diamonds in the shipment. He also wants to find a range of values he is 95% confident will contain this proportion.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

- A. Estimate the proportion of the perfect diamonds in the sample.
- B. Estimate the proportion of defective diamonds in the sample.
- C. Estimate the proportion of defective diamonds in the shipment.
- D. Estimate the mean weight of the diamonds that don't meet the company's expectation.

Figure 12 Example of Treatment Group Problem, First Part

Figure 12 shows the problem given to students in the proportion treatment group.

Students in the mean treatment group received a multiple-part problem about the mean with the same structure as this example. Students in either the proportion or mean control group received the same question as those in their respective treatment group, but presented in the traditional way.

Historically, a producer of high-quality jewelry has found that 10% of the top-grade diamonds it buys have serious defects after cutting and polishing. To monitor quality, a diamond company inspects a random sample of 10 recently processed diamonds from the latest shipment. They find that 1 have serious defects. Based on these results, the head of the quality division wants to obtain a single number estimate of the proportion of defective diamonds in the shipment. He also wants to find a range of values he is 95% confident will contain this proportion.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

Estimate the proportion of defective diamonds in the shipment.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. What are the measurements in this problem, and what is the Statistical Model?

- A. The proportion of defective diamonds in the sample; $N(\sigma, \sigma^2)$
- B. The proportion of defective diamonds in the sample; $N(n, p)$
- C. The number of defective diamonds in the sample; $b(10, p)$
- D. The proportion of perfect diamonds in the sample; $b(10, p)$

Figure 13 Example of Treatment Group Problem, Second Part

Figure 13 displays the second part of our treatment problem for the proportion confidence interval treatment group. Students were expected to choose the correct statistical model.

of the quality division wants to obtain a single number estimate of the proportion of defective diamonds in the shipment. He also wants to find a range of values he is 95% confident will contain this proportion.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

Estimate the proportion of defective diamonds in the shipment.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. What are the measurements in this problem, and what is the Statistical Model?

- A. The proportion of defective diamonds in the sample; $N(\sigma, \sigma^2)$
- B. The proportion of defective diamonds in the sample; $N(n, p)$
- C. The number of defective diamonds in the sample; $b(10, p)$
- D. The proportion of perfect diamonds in the sample; $b(10, p)$

Note: *This problem has more than one part.*

You have tried 1 of 3 times in this part.

*Your score for this attempt is for this part only;
your overall score is for all the parts combined.*

Go on to Next Part

Figure 14 Example of Treatment Group Problem, the “Go on to Next Part” Button

Figure 14 shows the “Go to the Next Pat” button that appeared when the student solved the problem.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

Estimate the proportion of defective diamonds in the shipment.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The number of defective diamonds in the sample; $b(10, p)$

3. The Model Parameter to Be Estimated:

At this point we examine how to achieve the Scientific Goal in terms of the statistical model. In this problem, the Scientific Goal is to estimate the proportion of defective diamonds in the shipment. This suggests that the model parameter we want to estimate is:

- A. the population mean
- B. the variance
- C. the population proportion
- D. something else

Figure 15 Example of Treatment Group Problem, Third Part

The third part of our treatment problem for the proportion confidence interval treatment group is shown in Figure 15.

1. The Scientific Goal:

The Scientific Goal is the reason for doing the experiment or study. In this problem, the Scientific Goal is:

Estimate the proportion of defective diamonds in the shipment.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The number of defective diamonds in the sample; $b(10, p)$

3. The Model Parameter to Be Estimated:

At this point we examine how to achieve the Scientific Goal in terms of the statistical model. In this problem, to achieve the Scientific Goal, the company wants to find out the proportion of defective diamonds and compares it with the historical value of 10%. This suggests means that the model parameter we want to estimate is: **the population proportion, p .**

4.1 Point Estimate:

Obtain a single number estimate of the population proportion.

$p =$

Figure 16 Example of Treatment Group Problem, Fourth Part

Figure 16 presents the fourth treatment problem for the proportion confidence interval treatment group. Students in the control group started their control problem with this question and they didn't receive the previous three subgoal problems.

The number of defective diamonds in the sample, $Y \sim \text{Bin}(10, p)$

3. The Model Parameter to Be Estimated:

At this point we examine how to achieve the Scientific Goal in terms of the statistical model. In this problem, to achieve the Scientific Goal, the company wants to find out the proportion of defective diamonds and compares it with the historical value of 10%. This suggests means that the model parameter we want to estimate is: **the population proportion, p .**

4.1 Point Estimate:

Obtain a single number estimate of the population proportion: $p = 0.1$

4.2 Interval Estimates:

Estimate this same quantity using a 95% approximate score confidence interval.

$\leq p \leq$

Note: *You can earn partial credit on this problem.*

Note: *This problem has more than one part.
You have tried 0 of 3 times in this part.
Your score for this attempt is for this part only;
your overall score is for all the parts combined.*

Show hint

Figure 17 Example of Treatment Group Problem, Fifth Part

Figure 17 shows the fifth treatment problem for the proportion confidence interval treatment group. Notice the “Show hint” button in the bottom of Figure 17.

To monitor product quality, the quality control unit of a manufacturing company randomly selects 20 products and 1 of them is defective. Construct a 95% approximate score confidence interval for the population proportion of defective products.

The point estimate is $\hat{p} = \frac{1}{20} = 0.05$

To construct the confidence interval, we first find the adjusted sample proportion. Since the confidence level is $L = 0.95$,

$$\tilde{p} = \frac{y+0.5(z_{(1+L)/2})^2}{n+(z_{(1+L)/2})^2} = \frac{1+0.5(1.96^2)}{20+1.96^2} = 0.1225$$

The confidence interval is then computed as

$$\begin{aligned} & (\tilde{p} - z_{(1+L)/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}, \tilde{p} + z_{(1+L)/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}) \\ &= (0.1225 - 1.96 \sqrt{\frac{0.1225(1-0.1225)}{20}}, 0.1225 + 1.96 \sqrt{\frac{0.1225(1-0.1225)}{20}}) \\ &= (-0.0211, 0.2661) \end{aligned}$$

Since the population proportion is always nonnegative, the confidence interval is $(0, 0.2661)$.

Figure 18 Example of Treatment Group Problem, Hint Part

Figure 18 shows the screen displayed when students in the treatment group clicked on the “show hint” button. This is the complete solution to a problem isomorphic to the one they were being asked to solve.

4.2 Interval Estimates:

The 95% confidence interval is:
(0, 0.4639)

5. Results and Interpretation:

If the company expects the proportion of defective diamonds in all its shipments to be equal to the historical value of 0.10, and the 95% confidence interval for the proportion of defective diamonds in the shipment is (0, 0.4639), what can the company conclude about the proportion of defective diamonds in the latest shipment?

- A. Based on this interval, the company can conclude that the proportion of defective diamonds in the shipment is above the historical value.
- B. Based on this interval, the company can conclude that the proportion of defective diamonds in the shipment equals the historical value.
- C. Based on this interval, the company can conclude that the proportion of defective diamonds in the shipment is below the historical value.
- D. Based on this interval, the company cannot conclude that the proportion of defective diamonds in the shipment differs from the historical value.

Figure 19 Example of Treatment Group Problem, Sixth Part

The last part of the treatment problem is shown in Figure 19. To successfully finish the problem, students had to choose the correct interpretation of the confidence interval.

3.5.3 Control Group

Figures 20 and 21 show the control problem for the proportion confidence interval control group. The problem is the same as that given the treatment group, but the control group was not given the subgoal problems provided the treatment group.

(1 pt)

Historically, a producer of high-quality jewelry has found that 10% of the top-grade diamonds it buys have serious defects after cutting and polishing. To monitor quality, a diamond company inspects a random sample of 10 recently processed diamonds from the latest shipment. They find that 2 have serious defects. Based on these results, the head of the quality division wants to find a range of values he is 95% confident will contain the proportion of defective diamonds in this shipment.

Construct a 95% approximate score confidence interval for this proportion.

$$\boxed{} \leq p \leq \boxed{}$$

Figure 20 Example of Control Group Problem, First Part

Historically, a producer of high-quality jewelry has found that 10% of the top-grade diamonds it buys have serious defects after cutting and polishing. To monitor quality, a diamond company inspects a random sample of 10 recently processed diamonds from the latest shipment. They find that 1 have serious defects. Based on these results, the head of the quality division computes a 95% approximate score confidence interval for the true proportion of defective diamonds in the shipment: $(0, 0.463)$. If the company expects the proportion of defective diamonds in all its shipments to be equal to the historical value of 0.10, what can the company conclude about the proportion of defective diamonds in the latest shipment?

- A. Based on this interval, the company cannot conclude that the proportion of defective diamonds in the shipment differs from the historical value.
- B. Based on this interval, the company can conclude that the proportion of defective diamonds in the shipment equals the historical value.
- C. Based on this interval, the company can conclude that the proportion of defective diamonds in the shipment is below the historical value.
- D. Based on this interval, the company can conclude that the proportion of defective diamonds in the shipment is below the historical value.

Figure 21 Example of Control Group Problem, Second Part

3.5.4 Post-Test

For the last part in each lab section, all students were given a post-test with two problems. Problems in the post-test were the same for all students and of the same types as the pre-test problems. Students were given one try.

A local Boston magazine wants to know whether most people in Boston support Boston Celtics rather than other teams. They randomly asked 10 people living in Boston, "Which NBA team is your favorite?" 8 of 10 said they supported Boston Celtics, and 2 of 10 people said they liked other teams.

Use a 95% approximate score confidence interval to estimate the proportion of the population that likes Boston Celtics.

$\leq p \leq$

Note: *You can earn partial credit on this problem.*

Show correct answers

Figure 22 Example of Post-Test Problem, First Question

A local Boston magazine wants to know whether most people in Boston support Boston Celtics rather than other teams. They randomly asked 11 people living in Boston, "Which NBA team is your favorite?" 11 of 11 said they supported Boston Celtics, and 0 of 11 people said they liked other teams. Based on these results a 95% approximate score confidence interval for the proportion of Bostonians who like the Boston Celtics is $(0.672, 1)$.

If the historical record shows that 90% people in Boston support Boston Celtics, which of the following can the editors of the magazine conclude?

- A. Based on this interval, the editors of the magazine cannot conclude with 95% confidence that the proportion of Bostonians who support Boston Celtics differs from the historical value of 90%.
- B. They cannot conclude anything since the sample size is small compare to the population of Boston.
- C. Since 0.5 lies below our confidence interval, the editors of the magazine are 95% confident that only half of Bostonians support the Boston Celtics.
- D. Based on this interval, the editors of the magazine can conclude with 95% confidence that the proportion of Bostonians who support the Boston Celtics is equal to the historical value of 90%.

Figure 23 Example of Post-Test Problem, Second Question

Figures 22 and 23 show the problem types in the post-test. Those two problems were in the same form as students had in the pre-test.

Grades on the pre-test and post-test were recorded in WeBWork. Student performances on the learning materials were also recorded for data analysis. We analyzed the data using a two-way effects model with a set of responses: post-test score minus pre-test score to see if there was a difference in mean response between treatment and control. We also explored the effect of covariates such as gender, major, class year and pre-test score.

The confidence interval problems of population mean for treatment group have already been shown in 3.4.4, and the corresponding pre-test, post-test, and control group problems are attached in Appendix C. We also attached all the hypothesis test problems in Appendix D.

Chapter 4: Results

4.1 Conduct of Experiments

We conducted our first experiment on December 1, 2010. There were five sections of approximately 25 students each. The subject material for the experiment conducted that day was confidence intervals. We both stayed in the lab for the entire ten hour duration of the experiment. During this first lab, we found several deficiencies in our problem set and experimental design.

The first issue was that many students in the treatment group did not finish the problem set and did not realize that they hadn't finished it since they didn't know that they needed to click on a "go to next part" button to get to the next part of the multi-part problem. Professor Petruccelli did address this at the beginning of the each lab section but some students still got confused by the structure of the new type of problem. We realized that as a result the data we obtained might be hard to analyze without bias and that the results obtained might be inaccurate.

The second issue was that a few students did not treat our problem set seriously. Instead, they just gave answers to get the task over with quickly. For example, we observed a few students who left just after the professor left and a few students who just randomly chose answers to finish the lab. This led us to conclude that some data was not helpful and that we needed to remove it from our data set to get a more accurate result. After communicating with the professor, we decided to try to identify these cases with the help of WeBWork's tracking information, and to remove them when we analyzed the data.

The third issue was that students worked on problem set without following the order of the problem set. That is, some students might first work on a post-test problem and do the pre-test problem last after finishing the treatment or control problem. This

might bias the results since someone might get full score on pre-test and low score on post-test.

Based on these deficiencies, we revised our problem set for the second lab. First, we made the problem set one multi-part problem for both control and treatment group. This forced all students, not just those in the treatment group, to navigate by clicking on “go to the next part” buttons. Also, students were forced to work on the problems in order since there was only one multi-part problem. To minimize frivolous responses, we kept close track of student performance, looking for any evidence of spurious answers.

After these improvements, we conducted our second experiment on December 8, 2010. In this experiment students worked on a hypothesis testing problem set. Both control and treatment group were only given one multi-part problem which contained pre-test, treatment (control) and post-test. We again stayed in lab for the duration.

After the second lab, we started to sort the data and analyze it. We first looked through each student’s scores. There were two problems we found. First, there were some faulty data as we mentioned above. These data were removed from the analysis. Second, we realized that our grading system for the second occasion (hypothesis test problems) might have biased the results. For these problems, in order to get the score of each of the three parts, we used a special score weight on each part where the total score of problem set was a 3-digit number. The first digit represented the score of the pre-test. The second digit represented the score of the treatment or control part. The third digit represented the score of the post-test. Therefore, if this number is viewed as a single total score, the score of the pre-test counts more than 80 percent of the total score. Since students worked on the pre-test first, they could see their percentage of score achieved out of the total score after they finished the pre-test and then, despite the fact that they had been told their score would not count toward their course grade, they might simply stop working after realizing that they had attained a high total score for the problem set. This might greatly reduce the score of the post-test, which was not something we had anticipated.

4.2 Statistical Analysis

4.2.1 Two-Way Effects Model

We used a two-way effects model to test if there was any difference in student scores associated with statistical model or type of assistance. In this model, we considered as the response, the increase in student scores from pre to post test.

We fit the following two-way model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}, i = 1,2, j = 1,2, k = 1, \dots, n_{ij}$$

where Y_{ijk} is the post-pre difference in scores for the k -th student having assistance i and statistical model j , μ is the overall mean post-pre difference, α_i is the effect of assistance i , β_j is the effect of statistical model j , $\alpha\beta_{ij}$ is their interaction, and ε_{ijk} is a random error term.

		Estimate	Standard Error
Assistance Type α_i	Control	-0.072	0.646
	Treatment	0.072	0.565
Problem Type β_j	Mean	0.046	0.503
	Proportion	-0.046	1.012
Interactions $\alpha\beta_{ij}$	Control * Mean	-0.024	0.514
	Control * Proportion	0.026	0.759
	Treatment * Mean	0.014	0.489
	Treatment * Proportion	-0.021	0.647

Figure 24 Value of Factors for Lab of Confidence Interval

		Estimate	Standard Error
Assistance Type α_i	Control	-0.012	0.497
	Treatment	0.010	0.619
Problem Type β_j	Mean	-0.119	0.480
	Proportion	0.100	0.608
Interactions $\alpha\beta_{ij}$	Control * Mean	0.095	0.515
	Control * Proportion	-0.082	0.500
	Treatment * Mean	-0.077	0.458
	Treatment * Proportion	0.083	0.704

Figure 25 Value of Factors for Lab of Hypothesis Test

Figures 24 and 25 give effects estimates and standard errors for the two-way effects model. While the estimated effect of treatment is positive (0.07237 for the confidence interval experiments, 0.01039 for the hypothesis test experiments), neither effect was statistically significant (p-values 0.1511 and 0.4412, respectively). In fact, no effects in the model were statistically significant.

4.2.2 Linear Regression

One of the goals of our experiments was to test whether our subgoal learning /forward fading problems were more effective than traditional problems in helping students who did not understand the materials improve their scores. To investigate this, we looked only at those students who got one or both pre-test questions wrong. Almost every one of these students got exactly 1 out of 2 problems correct in the pre-test. For these students we decided to look at the relation between the score on the treatment or control problem (whichever was applicable) and the post-test score. Specifically, we fit a simple linear regression of post-score on treatment problem score or control problem score. For the lab 1 data, we fit a regression for students in the treatment group and students in the control group.

$$y = \beta_0 + \beta_1 T + \beta_2 G + \beta_3 TG + \varepsilon$$

In this linear regression model, y is the post-test score, T is the score, on the treatment or control problem and $G = \{0,1\}$ is the type of treatment that $G = 1$ stands for treatment group and $G = 0$ stands for control group. The estimates of β 's are

	Estimate	Standard Error
β_0	0.5912	0.1805
β_1	0.1640	0.0550
β_2	-0.6633	0.4173
β_3	0.0895	0.3104

Figure 26 Estimate of β 's

In Figure 27 and Figure 28, the x-axis represents the score in the treatment part or control part, and y-axis represents the post-test.

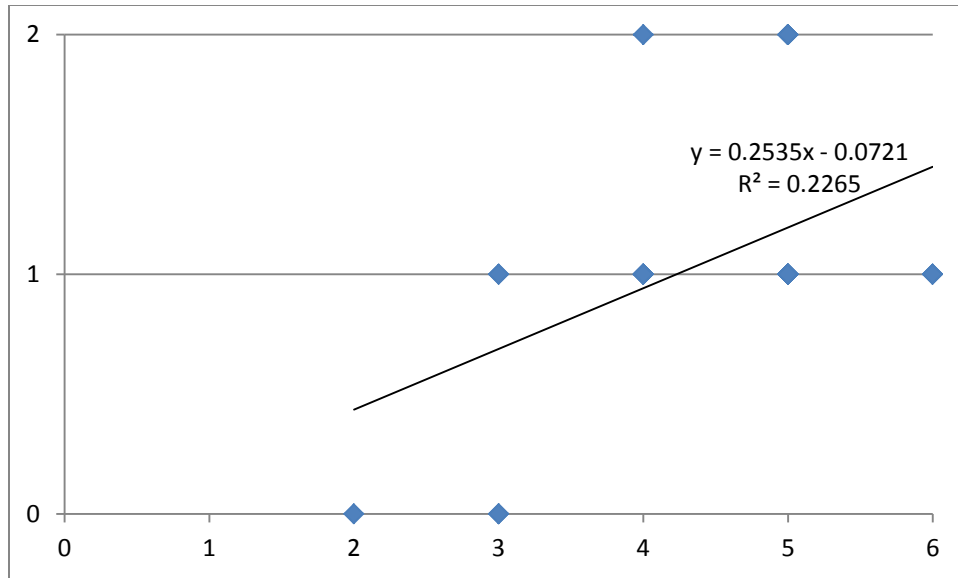


Figure 27 Post-Test Score vs. Middle Part Score, Treatment Group, Hypothesis Test Lab

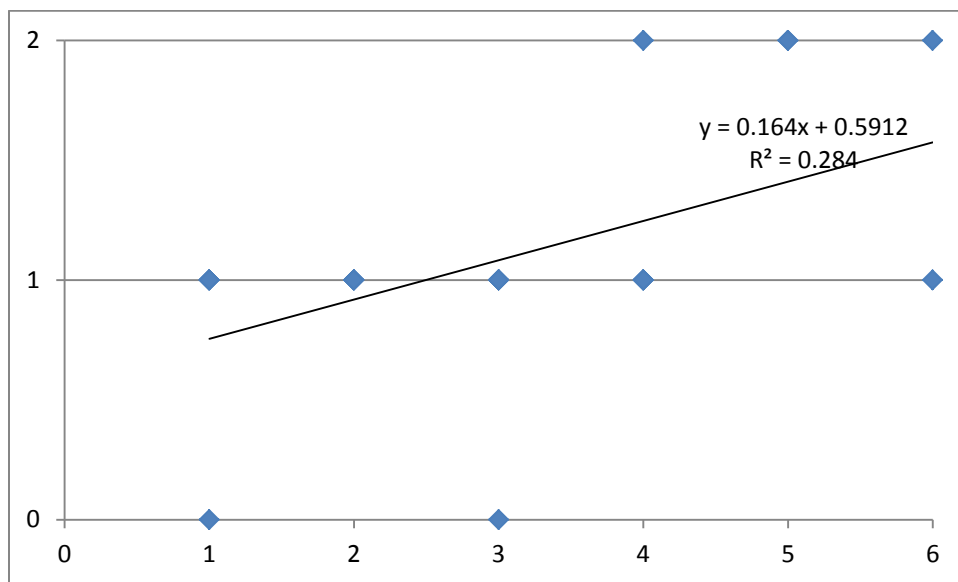


Figure 28 Post-Test Score vs. Middle Part Score, Control Group, Hypothesis Test Lab

Both graphs show that for students who did not score perfectly on the pre-test there is a positive association between score on the treatment or control problem, and score on the post-test. The slope of the fitted equation between treatment and post-test

score is greater than the slope of the fit equation between control and post-test score. The 95% confidence interval for the slope difference β_3 is (0.0273, 0.1517). Therefore, we estimate that for those students deficient in understanding (as indicated by a low score on the pre-test) the subgoal learning/forward fading treatment increases performance by between 2.7 and 15.14 percent of the score on the treatment problem relative to the increase obtained by the control group.

We also used the regression model for data from our confidence interval experiment. Because in this experiment, the numbers of questions in the middle part for the treatment group and for the control group are different, we cannot compare them. However, we did regress post-test score on treatment score for the treatment group.

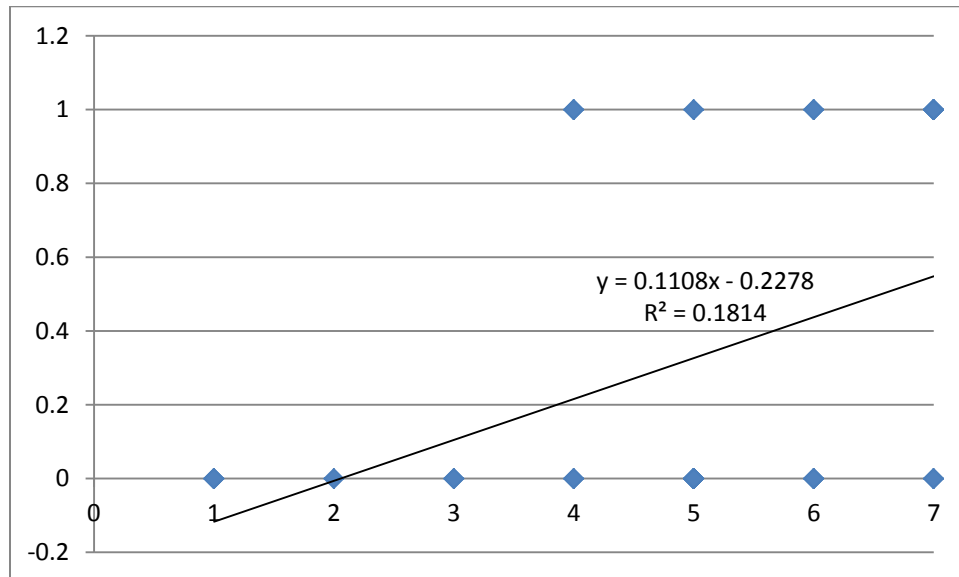


Figure 29 Post-Test Score vs. Middle Part Score, Treatment Group, Confidence Interval Lab

Figure 29 is a linear regression between treatment score and post-test score. It indicates the positive relation between the treatment problem score and post-test score. A 95% confidence interval for the slope is (0.023, 0.199).

Chapter 5: Conclusions

5.1 Accomplishments

In this project, we accomplished the following:

5.1.1 Packages

We made a package that allowed multi-part problems with hints and more interactions in WeBWork. For our project, we used Perl and the Html language to revise an existing package to extend the ability of having multiple-part problems with extra hints where necessary. This improvement gave the opportunity for more interactions between students and the material and provided instant help to students who didn't understand the materials.

5.1.2 New WeBWork-Based Instructional Model

We also developed an instructional model combining the Subgoal Learning Model of Catrabone and our own modification of Renkl et. al.'s Backward Fading Model, which we call the Forward Fading Model, to make effective multi-part problems for use as tutorials. The package is flexible enough to allow instructors to implement other instructional models to fit their own needs.

5.1.3 Implementation

We used our package to implement the new instructional model. We created problem sets for confidence intervals and hypothesis tests which incorporated the subgoal Learning/Forward Fading model with worked examples.

5.1.4 Testing

We designed experiments for testing the effectiveness of our new instructional materials in helping students learn confidence interval and hypothesis test material. From the experiment results, we found that there were no statistically significant differences in score improvement between students in the control group and students in treatment group for either the confidence interval or hypothesis test material. However, for students who scored 50% or less on the pre-test, we found,

1. A significant positive relationship between the score on the instructional material (either treatment or control) and the increase in post-pre score (as a proportion of instructional material score);
2. The increase for those in the treatment group was greater than for those in the control group;
3. This result held for both the confidence interval and hypothesis test material.

5.2 Possible Future Improvements

From our IQP experience and advice from Professor Petrucci, we have some ideas for improving our package, problem set, and testing methods.

5.2.1 Package

Our package enables the implementation of new features in WeBWork problems, but as we learned, to be effectively used in our experiments, these features had to be employed carefully. When we tried to implement multi-part problems in our experiments, we found that students were unfamiliar with the navigation through the problems. For example, many students in the treatment group didn't know how to get to the next part of the problem for the treatment problem. This led to students not completing parts of the experiments or to doing problems out of sequence.

Our multi-part problem requires relatively long code which is made of many repeats of the same code. This makes writing problems tedious even for experienced PG language programmers. It would be beneficial to build into the package the ability to make these repeated displays automatically. Also, the scoring system for multi-part problems in our package allows the instructor to set different scores for each part, but only reports the total score in WeBWork. Instructors who want to see the score of each problem part may need to make some adjustments to our package.

Here are some ways our package might be improved:

1. Develop adaptive functionality: For a multi-part problem, the system could choose the next problem from a set of problems based on the correctness of the previous problem for a student. This process is called adaptive and presents suitable problems based on student performance.
2. Packages should be made for easily integrating graphics, video and other applications.
3. In multiple choice problems, the package might be modified to allow targeted hints or help based on the wrong answer chosen.

5.2.2 Problem Set and Experimental Design

The design and conduct of our experiments gave us good experience in how to design a good statistical experiment. First, our experiment involved students as subjects which, as is well known, often results in noisy data. We tried to reduce the noise by having a relatively large sample but some students didn't treat our experiment seriously and as a result for those students the data didn't represent the real improvement on their understanding of materials.

Some failures in the experiment were due to our lack of foresight. For example, in the first session (confidence intervals), we structured successive parts of the experiment so that students could do them out of order: for example, they could do the post-test

before the treatment/control part of the experiment. We did remedy this in the second session (hypothesis tests) by forcing a sequence on the problems.

Our experience leads us to suggest ways to improve future experiments. First, to make students treat the experiments seriously we would, with the instructor's agreement, tell students that the score of some randomly-selected problems in the experiment would be counted as part of their final grade in the course. However, to avoid treating students unequally, we would only count the pre-test score toward their final grade. Second, we would design the experiments more carefully and try to avoid adjustments of lab design between two labs.

5.3 In Conclusion

Our most important conclusion is about our problem set efficacy. In our analysis of the experimental data, we did not find a statistically significant difference in mean score improvement between students in the treatment group and control group. However, we found that students who did not do well in the pre-test and did well in the treatment part in the treatment group had a greater improvement of their score compared with students in the control group who did not do well in the pre-test and did well in the control part.

On the other hand, the package we developed gives instructors a lot of freedom to create their problems. They can set the maximum allowed tries for each multi-part problem and they can add a hint part for each part of a problem. The hint part could be anything that professors want to add: for example, an isomorphic problem with solutions. Our new type problem could become a tutorial for students when they work on a problem involving new materials. Students involved in distance learning or self-study could get help and interaction directly from the hint part or from the forward fading process we mentioned in the methodology chapter. Instructors also could give supplementary materials, such as a page from a textbook or a website link as a hint. We

hope that instructors could create suitable problems with our package with more interactions.

With improved packages and problem sets, we feel that students can learn more efficiently, and we hope that the results of our project will make a positive contribution toward this goal.

Appendix

A. Package “improvedCompound” Code

```
sub _improvedCompound_init {}; # don't reload this file

#####
#
# This package implements a method of handling multi-part problems
# that show only a single part at any one time. The students can
# work on one part at a time, and then when they get it right (or
# under other circumstances determined by the professor), they can
# move on to the next part. Students cannot return to earlier parts
# once they have been completed. The score for problem as a whole is
# made up from the scores on the individual parts, and the relative
# weighting of the various parts can be specified by the problem
# author.
#
# To use the improvedCompound library, use
#
#   loadMacros("improvedCompound.pl");
#
# at the top of your file, and then create a improvedCompound object
# via the command
#
#   $cp = new improvedCompound(options)
#
# where '$cp' is the name of a variable that you will use to
# refer to the improved compound problem, and 'options' can include:
#
#   parts => n           The number of parts in the problem.
#                       Default: 1
#
#   weights => [n1,...,nm] The relative weights to give to each
#                       part in the problem. For example,
#                       weights => [2,1,1]
#                       would cause the first part to be worth 50%
#                       of the points (twice the amount for each of
#                       the other two), while the second and third
#                       part would be worth 25% each. If weights
#                       are not supplied, the parts are weighted
#                       by the number of answer blanks in each part
#                       (and you must provide the total number of
#                       blanks in all the parts by supplying the
#                       totalAnswers option).
#
#   totalAnswers => n    The total number of answer blanks in all
#                       the parts put together (this is used when
#                       computing the per-part scores, if part
#                       weights are not provided).
#
#   saveAllAnswers => 0 or 1 Usually, the contents of named answer blanks
#                       from previous parts are made available to
#                       later parts using variables with the
#                       same name as the answer blank. Setting
#                       saveAllAnswers to 1 will cause ALL answer
#                       blanks to be available (via variables
#                       like $ANSWER1, and so on).
#                       Default: 0
#
#   parserValues => 0 or 1 Determines whether the answers from previous
#                       parts are returned as MathObjects (like
#                       those returned from Real(), Vector(), etc)
```

```

# or as strings (the unparsed contents of the
# student answer). If you intend to use the
# previous answers as numbers, for example,
# you would want to set this to 1 so that you
# would get the final result of any formula
# the student typed, rather than the formula
# itself as a character string.
# Default: 0
#
# nextvisible => type Tells when the "go on to the next part" option
# is available to the student. The possible
# types include:
#
# 'ifCorrect' next is available only when
# all the answers are correct.
#
# 'Always' next is always available
# (but remember that students
# can't go back once they go
# on.)
#
# 'Never' next is never allowed (the
# problem will control going
# on to the next part itself).
#
# Default: 'ifCorrect'
#
# nextStyle => type Determines the style of "next" indicator to
# display (when it is available).
# The type can be one of:
#
# 'CheckBox' a checkbox that allows the students
# to go on to the next part when they
# submit their answers.
#
# 'Button' a button that submits their answers
# and goes on to the next part.
#
# 'Forced' forces the student to go on to the
# next part the next time they submit
# answers.
#
# 'HTML' allows you to provide an arbitrary
# HTML string of your own.
#
# Default: 'Checkbox'
#
# nextLabel => string Specifies the string to use as the label for the
# checkbox, the name of the button, the text of
# the message indicating the name of the button,
# the text of that the next submit will move to
# next part, or the HTML string, depending on the
# setting of nextStyle above.
#
# nextNoChange => 0 or 1 Since the students must submit their answers
# again to go on to the next part, it is possible
# for them to change their answers before they
# submit, and if nextvisible is 'ifCorrect' they
# might go on to the next without having correct
# answers stored. This option lets you control
# whether the answers are checked against the
# previous ones before going on to the next part.
# If the answers don't match, a warning is issued
# and they are not allowed to move on.
# Default: 1
#
# allowReset => 0 or 1 Determines whether a "Go back to the first part"
# checkbox is provided on parts 2 and later.
# This is intended for the professor during
# testing of the problem (otherwise it would be
# impossible to go back to earlier parts).
#

```

```

#                                     Default: 0
#
#   resetLabel => string           The string used to label the reset checkbox.
#
# Once you have created a improvedCompound object, you can use $cp->part to
# determine the part that the student is working on, and use 'if' statements
# to display the proper information for the given part. The improvedCompound
# object takes care of maintaining the data as the parts change. (See the
# improvedCompound.pg file for an example of a improved compound problem.)
#
# In order to handle the scoring of the problem as a whole when only part is
# showing, the improvedCompound object uses its own problem grader to manage
# the scores, and calls your own grader from there. The default is to use
# the one that was installed before the improvedCompound object was created,
# or avg_problem_grader if none was installed. You can specify a different
# one using the $cp->useGrader() method (see below). It is important that
# you NOT call install_problem_grader() yourself once you have created the
# improvedCompound object, as that would disable the special grader, causing
# the improved compound problem to fail to work properly.
#
# You may call the following methods once you have a improvedCompound:
#
#   $cp->part                       Returns the part the student is working on.
#   $cp->part(n)                     Sets the part to be part n, as long as the
#                                   student has finished the preceding parts.
#                                   If not, the part is set to the highest
#                                   one the student hasn't completed, and he
#                                   can work up to the given part. (The
#                                   nextVisible option is set to 'ifCorrect' if
#                                   it was 'Never' so that students can go on
#                                   once they finish the earlier parts.)
#
#   $cp->useGrader(code_ref)        Supplies your own grader to use in
#                                   place of the default one. For example:
#                                   $cp->useGrader(~&std_problem_grader);
#
#   $cp->score                       Returns the (weighted) score for this part.
#                                   Note that this is the score shown at the
#                                   bottom of the page on which the student
#                                   pressed submit (not the score for the answers
#                                   the student is submitting -- that is not
#                                   available until after the body of the problem
#                                   has been created).
#
#   $cp->scoreRaw                     Returns the unweighted score for this part.
#
#   $cp->scoreOverall                 Returns the overall score for the problem
#                                   so far.
#
#   $cp->addAnswers(list)            Make additional answer blanks be available
#                                   from one part to another. E.g.,
#                                   $cp->addAnswers('AnSwEr1');
#                                   would make the first unnamed blank be
#                                   available in later parts as well.
#                                   (This command should be issued only
#                                   when the part containing the
#                                   given answer blank is displayed.)
#
#   $cp->nextCheckbox(label)         Returns the HTML string for the "go on to next
#                                   part" checkbox so you can use it in the body
#                                   of the problem if you wish. This should not
#                                   be inserted when the $displayMode is 'Tex'.
#                                   If the label is not given or is blank, the
#                                   default label is used.
#
#   $cp->nextButton(label)           Returns the HTML string for the "go on to next
#                                   part" button so you can use it in the body of
#                                   the problem if you wish. This should not be
#                                   inserted when the $displayMode is 'Tex'. If
#                                   the label is not given or is blank, the
#                                   default label is used.
#

```

```

#
#   $cp->nextForces(label)   Returns the HTML string for the forced "go on
#                           to next part" so you can use it in the body
#                           of the problem if you wish. This should not
#                           be inserted when the $displayMode is 'Tex'.
#                           If the label is not given or is blank, the
#                           default label is used.
#
#   $cp->reset               Go back to part 1, clearing the answers
#                           and score. (Best used when debugging
#                           problems.)
#
#   $cp->resetCheckbox(label) Returns the HTML string for the reset checkbox
#                           so that you can provide one within the body
#                           of the problem if you wish. This should not
#                           be inserted when the $displayMode is 'Tex'.
#                           If the label is not given or is blank, the
#                           default label will be used.
#
#####

package improvedCompound;

#
# The state data that is stored between invocations of
# the problem.
#
our %defaultStatus = (
    part => 1,                # the current part
    answers => "",           # answer labels from previous parts
    new_answers => "",       # answer labels for THIS part
    ans_rule_count => 0,     # the ans_rule count from previous parts
    new_ans_rule_count => 0, # the ans_rule count from THIS part
    score => 0,              # the (weighted) score on this part
    total => 0,              # the total on previous parts
    raw => 0,                # raw score on this part
    hint => 0,               # if the hint display to students
    tries => -1,            # num of tries students already used
);

#
# Create a new instance of the improved Compound Problem and initialize
# it. This includes reading the status from the previous
# parts, defining the variables from the answers to previous parts,
# and setting up the grader so that the current data can be saved.
#
sub new {
    my $self = shift; my $class = ref($self) || $self;
    my $cp = bless {
        parts => 1,
        totalAnswers => undef,
        weights => undef,    # array of weights per part
        saveAllAnswers => 0,  # usually only save named answers
        parserValues => 0,    # make Parser objects from the answers?
        nextVisible => "ifCorrect", # or "Always" or "Never"
        nextStyle => "Checkbox", # or "Button", "Forced", or "HTML"
        nextLabel => undef,    # checkbox text or button name or HTML
        nextNoChange => 1,    # true if answer can't change for new part
        allowReset => 0,      # true to show "back to part 1" button
        resetLabel => undef,   # label for reset button
        hint => 0,
        hintLabel => undef,    # label for hint button
        maxtries => 3,
        grader => $main::PG_FLAGS{PROBLEM_GRADER_TO_USE} ||
            \&main::avg_problem_grader,
        @_,
        status => $defaultStatus,
    }, $class;
    die "You must provide either the totalAnswers or weights"
        unless $cp->{totalAnswers} || $cp->{weights};
}

```

```

$cp->getTotalWeight if $cp->{weights};
main::loadMacros("Parser.pl") if $cp->{parserValues};
$cp->reset if $cp->{allowReset} && $main::inputs_ref->{_reset};
$cp->getStatus;
$cp->initPart;

if ($main::inputs_ref->{_hint} || ($main::inputs_ref->{submitAnswers} &&
    $main::inputs_ref->{submitAnswers} eq "Show hint") )
{
    $cp->{status}->{hint} = 1 ;
    $cp->{status}->{tries} = $cp->{status}->{tries} - 1;
}
return $cp;
}

#
# Compute the total of the weights so that the parts can
# be properly scaled.
#
sub getTotalWeight {
    my $self = shift;
    $self->{totalWeight} = 0; $self->{totalAnswers} = 1;
    foreach my $w (@{$self->{weights}}) {$self->{totalWeight} += $w}
    $self->{totalWeight} = 1 if $self->{totalWeight} == 0;
}

#
# Look up the status from the previous invocation
# and see if we need to go on to the next part.
#
sub getStatus {
    my $self = shift;
    main::RECORD_FORM_LABEL("_next");
    main::RECORD_FORM_LABEL("_status");
    $self->{status} = $self->decode;
    $self->{isNew} = $main::inputs_ref->{_next} ||
        ($main::inputs_ref->{submitAnswers} &&
            $main::inputs_ref->{submitAnswers} eq ($self->{nextLabel} ||
                "Go on to Next Part"));
    if ($self->{isNew}) {
        $self->checkAnswers;
        $self->incrementPart unless $self->{nextNoChange} &&
            $self->{answersChanged};
    }
}

#
# Initialize the current part by setting the ans_rule
# count (so that later parts will get unique answer names),
# installing the grader (to save the data), and setting
# the variables for previous answers.
#
sub initPart {
    my $self = shift;
    $main::ans_rule_count = $self->{status}{ans_rule_count};
    main::install_problem_grader(\&improvedCompound::grader);
    $main::PG_FLAGS{improvedCompound} = $self;
    $self->initAnswers($self->{status}{answers});
}

#
# Look through the list of answer labels and set
# the variables for them to be the associated student
# answer. Make it a Parser value if requested.
# Record the value so that it will be available
# again on the next invocation.
#
sub initAnswers {
    my $self = shift; my $answers = shift;
    foreach my $id (split(/;/,$answers)) {
        my $value = $main::inputs_ref->{$id};
    }
}

```

```

$value = '' unless defined($value);
if ($self->{parservalues}) {
    my $parser = Parser::Formula($value);
    $parser = Parser::Evaluate($parser) if $parser && $parser->isConstant;
    $value = $parser if $parser;
}
${"main::$id"} = $value unless $id =~ m/$main::ANSWER_PREFIX/o;
$value = quoteHTML($value);
main::TEXT(qq!<input type="hidden" name="$id" value="$value" />!);
main::RECORD_FORM_LABEL($id);
}
}

#
# Look to see if any answers have changed on this
# invocation of the problem.
#
sub checkAnswers {
    my $self = shift;
    foreach my $id (keys(%{$main::inputs_ref})) {
        if ($id =~ m/^\previous_(.*)$/) {
            if ($main::inputs_ref->{$id} ne $main::inputs_ref->{$1}) {
                $self->{answersChanged} = 1;
                $self->{isNew} = 0 if $self->{nextNoChange};
                return;
            }
        }
    }
}

#
# Go on to the next part, updating the status
# to include the data from the old part so that
# it will be properly preserved when the next
# part is showing.
#
sub incrementPart {
    my $self = shift;
    my $status = $self->{status};
    if ($status->{part} < $self->{parts}) {
        $status->{tries} = -1;
        $status->{part}++;
        $status->{answers} .= ';' if $status->{answers};
        $status->{answers} .= $status->{new_answers};
        $status->{ans_rule_count} = $status->{new_ans_rule_count};
        $status->{total} += $status->{score};
        $status->{score} = $status->{raw} = 0;
        $status->{new_answers} = '';
    }
}

#####

#
# Encode all the status information so that it can be
# maintained as the student submits answers. Since this
# state information includes things like the score from
# the previous parts, it is "encrypted" using a dumb
# hex encoding (making it harder for a student to recognize
# it as valuable data if they view the page source).
#
sub encode {
    my $self = shift; my $status = shift || $self->{status};
    my @data = (); my $data = '';
    foreach my $id (main::lex_sort(keys(%defaultStatus))) {
        push(@data,$status->{$id})
    }
    foreach my $c (split(//,join('|',@data))) {$data .= toHex($c)}
    return $data;
}

```

```

#
# Decode the data and break it into the status hash.
#
sub decode {
    my $self = shift; my $status = shift || $main::inputs_ref->{_status};
    return {%defaultStatus} unless $status;
    my @data = (); foreach my $hex (split(/(.)/,$status)) {
        push(@data,fromHex($hex)) if $hex ne '';
    }
    @data = split('\|',join(',',@data)); $status = {%defaultStatus};
    foreach my $id (main::lex_sort(keys(%defaultStatus))) {
        $status->{$id} = shift(@data)
    }
    return $status;
}

#
# Hex encoding is shifted by 10 to obfuscate it further.
# (shouldn't be a problem since the status will be made of
# printable characters, so they are all above ASCII 32)
#
sub toHex {main::spf(ord(shift)-10,"%X")}
sub fromHex {main::spf(hex(shift)+10,"%c")}

#
# Make sure the data can be properly preserved within
# an HTML <INPUT TYPE="HIDDEN"> tag.
#
sub quoteHTML {
    my $string = shift;
    $string =~ s/&/\&/g; $string =~ s/"/\"/g;
    $string =~ s/>/\>/g; $string =~ s/</\</g;
    return $string;
}

#####

#
# Set the grader for this part to the specified one.
#
sub useGrader {
    my $self = shift;
    $self->{grader} = shift;
}

#
# Make additional answer blanks from the current part
# be preserved for use in future parts.
#
sub addAnswers {
    my $self = shift;
    $self->{extraAnswers} = [] unless $self->{extraAnswers};
    push(@{$self->{extraAnswers}},@_);
}

#
# Go back to part 1 and clear the answers and scores.
#
sub reset {
    my $self = shift;
    if ($main::inputs_ref->{_status}) {
        my $status = $self->decode($main::inputs_ref->{_status});
        foreach my $id (split(/;/,$status->{answers})) {
            delete $main::inputs_ref->{$id}
        }
    }
    foreach my $id (1..$status->{ans_rule_count})
    {
        delete $main::inputs_ref->{"${main::QUIZ_PREFIX}${main::ANSWER_PREFIX}$id"}
    }
}

```

```

    $main::inputs_ref->{_status} = $self->encode(\%defaultStatus);
    $main::inputs_ref->{_next} = 0;
}

#
# Return the HTML for the "Go back to part 1" checkbox.
#
sub resetCheckbox {
    my $self = shift;
    my $label = shift ||
        "<b>Go back to Part 1</b> (when you submit your answers).";
    my $par = shift; $par = ($par ? $main::PAR : '');
    qq!$par<input type="checkbox" name="_reset" value="1" />$label!;
}

#
# Return the HTML for the "next part" checkbox.
#
sub hintButton {
    my $self = shift;
    my $label = quoteHTML(shift || "Show hint");
    my $par = shift;
    $par = ($par ? $main::PAR : '');
    $par . qq!<input type="submit" name="submitAnswers" value="$label"!
        . q!onclick="document.getElementById('_next').value=1" />!;
}

#
# Return the HTML for the "next part" checkbox.
#
sub nextCheckbox {
    my $self = shift;
    my $label = shift ||
        "<b>Go on to next part</b> (when you submit your answers).";
    my $par = shift; $par = ($par ? $main::PAR : '');
    $self->{nextInserted} = 1;
    qq!$par<input type="checkbox" name="_next" value="next" />$label!;
}

#
# Return the HTML for the "next part" button.
#
sub nextButton {
    my $self = shift;
    my $label = quoteHTML(shift || "Go on to Next Part");
    my $par = shift; $par = ($par ? $main::PAR : '');
    $par . qq!<input type="submit" name="submitAnswers" value="$label" !
        . q!onclick="document.getElementById('_next').value=1" />!;
}

#
# Return the HTML for when going to the next part is forced.
#
sub nextForced {
    my $self = shift;
    my $label = shift ||
        "<b>Submit your answers again to go on to the next part.</b>";
    $label = $main::PAR . $label if shift;
    $self->{nextInserted} = 1;
    qq!$label<input type="hidden" name="_next" id="_next" value="Next" />!;
}

#
# Return the raw HTML provided
#
sub nextHTML {shift; shift}

#####

#

```



```

# Return the current part, or try to set the part to the given
# part (returns the part actually set, which may be earlier if
# the student didn't complete an earlier part).
#
sub part {
  my $self = shift; my $status = $self->{status};
  my $part = shift;
  return $status->{part} unless defined $part &&
    $main::displayMode ne 'TeX';
  $part = 1 if $part < 1; $part = $self->{parts} if $part > $self->{parts};
  if ($part > $status->{part} && !$main::inputs_ref->{_noadvance}) {
    unless ((lc($self->{nextVisible}) eq 'ifcorrect' && $status->{raw} < 1) ||
      lc($self->{nextVisible}) eq 'never') {
      $self->initAnswers($status->{new_answers});
      $self->incrementPart; $self->{isNew} = 1;
    }
  }
  if ($part != $status->{part}) {
    main::TEXT('<input type="hidden" name="_noadvance" value="1" />');
    $self->{nextVisible} = 'IfCorrect' if lc($self->{nextVisible}) eq 'never';
  }
  return $status->{part};
}

#
# Return the various scores
#
sub score {shift->{status}{score}}
sub scoreRaw {shift->{status}{raw}}
sub scoreOverall {
  my $self = shift;
  return $self->{status}{score} + $self->{status}{total};
}

#####
#
# The custom grader that does the work of computing the scores
# and saving the data.
#
sub grader {
  my $self = $main::PG_FLAGS{improvedCompound};

  my @answers = keys(%{$_[0]});
  my $weight = scalar(@answers)/$self->{totalAnswers};
  $weight = $self->{weights}[$self->{status}{part}-1]/$self->{totalWeight}
    if $self->{weights} && defined($self->{weights}[$self->{status}{part}-1]);
  @answers = grep(!/$main::ANSWER_PREFIX/o,@answers)
    unless $self->{saveAllAnswers};
  push(@answers,@{$self->{extraAnswers}}) if $self->{extraAnswers};
  my $space =
    '';

  #
  # Call the original grader, but put back the old recorded_score
  # (the grader will have updated it based on the score for the PART,
  # not the problem as a whole).
  #
  my $oldScore = ($_[1])->{recorded_score};
  my ($result,$state) = &{$self->{grader}}(@_);
  $state->{recorded_score} = $oldScore;

  #
  # Update that state information and encode it.
  #
  my $status = $self->{status};
  $status->{raw} = $result->{score};
  $status->{score} = $result->{score}*$weight;
  $status->{new_ans_rule_count} = $main::ans_rule_count;

```

```

$status->{new_answers} = join(';', @answers);
$status->{tries} = $status->{tries} + 1 if !$main::inputs_ref-
>{previewAnswers};
my $data = quoteHTML($self->encode);

#
# Update the recorded score
#
my $newScore = $status->{total} + $status->{score};
$state->{recorded_score} = $newScore if $newScore > $oldScore;
$state->{recorded_score} = 0 if $self->{allowReset} &&
    $main::inputs_ref->{_reset};

#
# Add the improvedCompound message and data
#
$result->{type} = "improvedCompound ($result->{type})";
$result->{msg} .= '</i><p><b>Note:</b> <i>' if $result->{msg};
$result->{msg} .= 'This problem has more than one part.'
    . '<br/>'. $space.'You have tried '$status->{tries}.' of '
    . '$self->{maxtries}.' times in this part.'
    . '<br/>'. $space
    . '<small>Your score for this attempt is for this part only;</small>'
    . '<br/>'. $space
    . '<small>your overall score is for all the parts combined.</small>'
    . qq!<input type="hidden" name="_status" value="$data" />!;

#
# Warn if the answers changed when they shouldn't have
#
$result->{msg} .=
    '<p><b>You may not change your answers when going on to the next part!</b>'
    if $self->{nextNoChange} && $self->{answersChanged} &&
        ($self->{maxtries} > $status->{tries});

#
# Include the "hint" button.
#
$result->{msg} .= '<br/>'. $self->hintButton($self->{hintLabel}, $par)
    if $self->{hint} && !$self->{status}->{hint};

#
# Include the "next part" checkbox, button, or whatever.
#
my $par = 1;
if ($self->{parts} > $status->{part} && !$main::inputs_ref->{previewAnswers})
{
    if (lc($self->{nextVisible}) eq 'always' ||
        (lc($self->{nextVisible}) eq 'ifcorrect' && $result->{score} >= 1)) {
        my $method = "next". $self->{nextStyle}; $par = 0;
        $result->{msg} .= $self->$method($self->{nextLabel}, 1). '<br/>';
    }
}

if ($result->{score} > 0)
{
    $result->{msg} .= $self->nextForced(" ", 0). '<br/>';
}

if (!$self->{maxtries} > $status->{tries}) && $result->{score} < 1)
{
    $result->{msg} .=
        $main::PAR . "<b>You have reached the maximum tries
        . </b>". $self->nextForced(). '<br/>';
}

```

```

#
# Add the reset checkbox, if needed
#
$result->{msg} .= $self->resetCheckbox($self->{resetLabel},$par)
    if $self->{allowReset} && $status->{part} > 1;

#
# Make sure we don't go on unless the next button really is checked
#
$result->{msg} .= '<input type="hidden" name="_next" value="0" />'
    unless $self->{nextInserted};

if ($self->{parts} == $self->{status}->{part})
{
    $result->{msg} = '</br><font size="8" color="red">Congratulations! </br>
        You have finished all the problems!</font>';
    $result->{msg} .= $self->resetCheckbox($self->{resetLabel},$par)
        if $self->{allowReset} && $status->{part} > 1;
}

return ($result,$state);
}

```

B. Sample Question Code

In this appendix, we provide the code we created for a confidence interval problem.

Codes for other problems are included in the supplement.

```
##KEYWORDS confidence interval, population mean, treatment
DOCUMENT();
# This should be the first executable line in the problem.

loadMacros(
  "PG.pl",
  "PGstandard.pl",
  "improveCompound.pl",
  "Parser.pl",
  "PGunion.pl",
  "PGcourse.pl",
  "PGbasicmacros.pl",
  "PGchoicemacros.pl",
  "PGanswermacros.pl",
  "PGnumericalmacros.pl",
  "PGstatisticsmacros.pl"
);

TEXT(beginproblem);
BEGIN_PROBLEM();

#####

$isProfessor = ($studentLogin eq 'dpvc' || $studentLogin eq 'professor');

# Start an improved compound problem. See the improveCompound.pl
# file for more details about the parameters you can supply.

$cp = new improvedCompound(
  parts => 6,           # the total number of parts in this problem
  totalAnswers => 6,   # total answers in all parts combined
  parserValues => 1,   # make parser objects from student answers
  allowReset => $isProfessor, # professors get Reset button for testing
  nextStyle => "Button",
);
$part = $cp->part;      # look up the current part

#####
# Set the quantities and answers of all parts
#####

$ybare = 0;
$se = 0;

for ($i=0;$i<4;$i++){
  $ye[$i] = random(41000,43000,1);
  $ybare = $ybare + $ye[$i];
}
$ybare = $ybare/4;

for ($i=0;$i<4;$i++){
  $se = $se + ($ye[$i] - $ybare)*($ye[$i] - $ybare);
}
$se = sqrt($se/3);

$deltae = 3.1824* $se / sqrt(4);
$te=($ybare-44000)/$deltae;
$pe=uprob($te);
```

```

$ybar = 0;
$s = 0;
for ($i=0;$i<7;$i++){
  $y[$i] = random(80,120,1);
  $ybar = $ybar + $y[$i];
}
$ybar = $ybar/7;

for ($i=0;$i<7;$i++){
  $s = $s + ($y[$i] - $ybar)*($y[$i] - $ybar);
}
$s = sqrt($s/6);

$delta = 2.4469 * $s / sqrt(7);
$t=($ybar-119)/$delta;
$p=1-uprob($t);

$ybaro = 0;
$sso = 0;

for ($i=0;$i<7;$i++){
  $yo[$i] = random(40,60,1);
  $ybaro = $ybaro + $yo[$i];
}
$ybaro = $ybaro/7;

for ($i=0;$i<7;$i++){
  $sso = $sso + ($yo[$i] - $ybaro)*($yo[$i] - $ybaro);
}
$sso = sqrt($sso/6);
$deltao = 2.4469* $sso / sqrt(7);
$to=($ybaro-120)/$deltao;
$po=1-uprob(($ybaro-120)/$deltao);

#####
#
# Part 1
#

if ($part == 1) {
$newProblemSeed = random(0,1000*$part,1);
$PG_random_generator->srnd($newProblemSeed);

BEGIN_TEXT
The breaking strengths of four large metal pins used in building
construction randomly chosen from a large production lot are (in psi).
$PAR
\[ \begin{array}{cccc}
$y[0] & $y[1] & $y[2] & $y[3] \cr
\end{array} \]

$PAR
The head of the quality unit expects the mean breaking strength of pins
to be greater than 44000, so he wants to test the hypotheses:
$BR
H_{0}:\(\mu\)=44000
$BR
H_{a}:\(\mu\)>44000 where \(\mu\) stands for the mean strength of the pins
in the lot.
$PAR
Compute the p-value for this hypothesis test.
p = \{ans_rule(10)\}
$BR
END_TEXT
ANS(num_cmp($pe, tol=>.0001, tolType=>"absolute"));
}

#####
#

```

```

# Part 2
#
if ($part == 2) {
$newProblemSeed = random(0,1000*$part,1);
$PG_random_generator->srnd($newProblemSeed);
$cmc = new_multiple_choice();
$cmc->qa(
    "what conclusion can the head of the quality make about the null
    hypothesis at the 0.05 level?",
    "Since the p-value is greater than 0.05, he cannot reject the null
    hypothesis."
);
$cmc->extra(
    "Since the p-value is less than 0.05, he can reject the null
    hypothesis.",
    "Since the p-value is less than 0.05, he cannot reject the null
    hypothesis.",
    "He cannot get the precise p-value,no conclusion can be made."
);
BEGIN_TEXT
The breaking strengths of four large metal pins used in building
construction randomly chosen from a large production lot are (in psi).
$PAR
\[ \begin{array}{cccc}
$y[0] & $y[1] & $y[2] & $y[3] \end{array} \] \cr
\end{array} \]
$PAR
The head of the quality unit expects the mean breaking strength of
pins to be greater than 44000,so he wants to test the hypotheses:
$BR
H_{0}:\(\mu\)=44000
$BR
H_{a}:\(\mu\)>44000 where \(\mu\) stands for the mean strength of
the pins in the lot.
$PAR
$PAR
\{$cmc -> print_q \}
\{$cmc -> print_a \}
END_TEXT
ANS(radio_cmp($cmc->correct_ans));
}

#####
#
# Part 3
#
if ($part == 3) {
$newProblemSeed = random(0,1000*$part,1);
$PG_random_generator->srnd($newProblemSeed);

$cmc = new_multiple_choice();
$cmc->qa(
    "The Scientific Hypothesis is the hypothesized outcome of the
    experiment or study. In this problem, the Scientific Hypothesis is:",
    "The average time spent by all WPI students on a 1/3 unit seven
    week undergraduate course is lower than the 119 hour guideline."
);
$cmc->extra(
    "The maximun time spent by WPI students on a 1/3 unit seven
    week undergraduate course is greater than the 119 hour guideline.",

```

```

"The average time spent by all WPI students on a 1/3 unit seven week
undergraduate course is equal to the 119 hour guideline.",
"The maximum time spent by all WPI students on a 1/3 unit seven week
undergraduate course is equal than the 119 hour guideline.");

$cme = new_multiple_choice();
$cme->qa(
  "The Statistical Model is the distribution of the population
of measurements that are being taken. What are the measurements
in this problem, and what is the Statistical Model?",
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses per term; \(\mathcal{N}(\mu, \sigma^2)\)"
);
$cme->extra(
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses per week; \(\mathcal{N}(\mu, \sigma^2)\)",
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses; \(\mathcal{N}(\sigma, \sigma^2)\)",
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses; \(\mathcal{B}(n, p)\)"
);

$cmd = new_multiple_choice();
$cmd->qa(
  "The Statistical Model is the distribution of the population of
measurements that are being taken. What are the measurements in
this problem, and what is the Statistical Model?",
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses per term; \(\mathcal{N}(\mu, \sigma^2)\)"
);
$cmd->extra(
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses per week; \(\mathcal{N}(\mu, \sigma^2)\)",
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses; \(\mathcal{N}(\sigma, \sigma^2)\)",
  "The hours WPI students spend on 1/3 unit seven week undergraduate
courses; \(\mathcal{B}(n, p)\)"
);

$cmh = new_multiple_choice();
$cmh->qa(
  "What is the Statistical Hypotheses for this problem? ",
  "H_{0}: \(\mu\)=44000",
  "H_{a}: \(\mu\)>44000 where \(\mu\) stands for the average time
spent by all WPI students on a 1/3 unit seven week undergraduate
course."
);
$cmh->extra(
  "H_{0}: \(\mu\)=44000",
  "H_{a}: \(\mu\)<44000 where \(\mu\) stands for the average time spent
by all WPI students on a 1/3 unit seven week undergraduate course.",
  "H_{0}: \(\mu\)=44000",
  "H_{a}: \(\mu\)<44000 where \(\mu\) stands for the minimum time spent
by all WPI students on a 1/3 unit seven week undergraduate course."
);
$cmh->makeLast(
  "H_{0}: \(\mu\)=44000",
  "H_{a}: \(\mu\)>44000 where \(\mu\) stands for the maximum time
spent by all WPI students on a 1/3 unit seven week undergraduate
course."
);

```

```

BEGIN_TEXT
$PAR
WPI sets an expectation of 17 hours work per week for a 1/3 unit seven week
undergraduate course, which equals 119 hours per term. To see if reality
meets expectations, a WPI project team took a random sample of 7 1/3 unit
courses and from each course, randomly selected one student volunteer to
follow during the term. Each selected student agreed to keep track of the
hours spent on coursework for that course. The totals for the term are
below:
$PAR

\[\begin{array}{ccccccc}
$y[0] & $y[1] & $y[2] & $y[3] & $y[4] & $y[5] & $y[6] \cr
\end{array}\]

$PAR
The project team was interested in determining whether the average time
spent by all WPI students on a 1/3 unit seven week undergraduate course is
lower than the 119 hour guideline.
$PAR

$BBOLD 1. The Scientific Hypothesis: $EBOLD
$PAR
\{$cmc -> print_q \}
\{$cmc -> print_a \}
$PAR
$BBOLD 2. The Statistical Model: $EBOLD
$PAR
\{$cme -> print_q \}
\{$cme -> print_a \}
$PAR
$PAR
$BBOLD 3. The Statistical Hypotheses: $EBOLD
$PAR\
\{$cmd -> print_q \}
\{$cmd -> print_a \}
$BBOLD 4.1 The Test Statistic: $EBOLD
$PAR
Obtain a single number estimate of the average number of hours per term
spent by WPI students for a 1/3 unit seven week undergraduate course.
$PAR
\(\bar{y}\) = \{ans_rule(10)\}
$PAR
$BBOLD 4.2 The p-value: $EBOLD
$PAR
Obtain the p-value for the Statistical Hypotheses above:
$BR
p-value=\{ans_rule(10)\}.
$PAR
$BBOLD 5. Results and Interpretation: $EBOLD
$PAR
\{$cmh -> print_q \}
\{$cmh -> print_a \}

END_TEXT

ANS(radio_cmp($cmc->correct_ans));
ANS(radio_cmp($cme->correct_ans));
ANS(radio_cmp($cmd->correct_ans));
ANS(radio_cmp($cmh->correct_ans));
ANS(num_cmp($ybar, tol=>.0001, tolType=>"absolute"));
ANS(num_cmp($p, tol=>.0001, tolType=>"absolute"));
}

#####
#
# Part 4
#

```



```

if ($part == 4) {

$newProblemSeed = random(0,1000*$part,1);
BEGIN_TEXT

It has
been hypothesized that individuals suffering from Alzheimer's disease
may spend less than the normal amount of time per night in the deeper
stages of sleep. To test this hypothesis researchers measured the time,
in minutes, spent in stage IV sleep (the deepest level of sleep)
for a sample of seven Alzheimer's patients (data below). In
healthy individuals, the average time spent in stage IV sleep is 120
minutes per night. The data are:

\[ \begin{array}{ccccccc}
$y[0] & $y[1] & $y[2] & $y[3] & $y[4] & $y[5] & $y[6] \end{array} \cr
\end{array} \]

$PAR
The researchers are interested in whether the mean of time spent in
stage IV sleep of Alzheimer's patients is lower than 120 minutes, so
he wants to test the hypotheses:$PAR
$BR
 $H_0: \mu = 120$ 
$BR
 $H_a: \mu < 120$  where  $\mu$  stands for the mean of time spent in
stage IV sleep of Alzheimer's patients.
$PAR
Compute the p-value for this hypothesis test.
 $p = \text{ans\_rule}(10)$ 
$BR

END_TEXT
ANS(num_cmp($po, tol=>.0001, tolType=>"absolute"));
}

#####
#
# Part 5
#

if ($part == 5) {
$a = floor(($ybar - $delta)*100)/100;
$b = floor(($ybar + $delta)*100)/100;
$ybar2 = floor(100*$ybar)/100;

$newProblemSeed = random(0,1000*$part,1);
$cmc = new_multiple_choice();
$cmc->qa(
  "what conclusion can the researchers make about the null hypothesis
  at the 0.05 level?",
  "Since the p-value is greater than 0.05, they cannot reject the null
  hypothesis."
);
$cmc->extra(
  "Since the p-value is less than 0.05, they can reject the null
  hypothesis.",
  "Since the p-value is less than 0.05, they cannot reject the null
  hypothesis.",
  "They cannot get the precise p-value, no conclusion can be made."
);
}
BEGIN_TEXT

```

It has been hypothesized that individuals suffering from Alzheimer's disease may spend less than the normal amount of time per night in the deeper stages of sleep. To test this hypothesis researchers measured the time, in minutes, spent in stage IV sleep (the deepest level of sleep) for a sample of seven Alzheimer's patients (data below). \$PAR

```
\[ \begin{array}{ccccccc}
$y[0] & & $y[1] & & $y[2] & & $y[3] & & $y[4] & & $y[5] & & $y[6] \cr
\end{array} \]
```

\$PAR
The researchers are interested in whether the mean of time spent in stage IV sleep of Alzheimer's patients is lower than 120 minutes, so he wants to test the hypotheses:\$PAR

\$BR
 $H_0: \mu = 120$

\$BR
 $H_a: \mu < 120$ where μ stands for the mean of time spent in stage IV sleep of Alzheimer's patients.

\$BR
\$PAR
Be given the observed value of standard test statistics:

```
t=$t0
$PAR
\{$cmc -> print_q \}
\{$cmc -> print_a \}
```

```
END_TEXT
ANS(radio_cmp($cmc->correct_ans));
}
#####
```

```
END_PROBLEM();
ENDDOCUMENT(); # This should be the last executable line in the problem.
```

C. Problem Set for Confidence Interval Experiment

In this appendix, we provide screen shots of the pre-test, post-test and control confidence interval problems for the population mean. Screen shots of other confidence interval problems can be found in Sections 3.4.4 and 3.5.

(1 pt) The breaking strengths of four large metal pins used in building construction randomly chosen from a large production lot are (in psi).

41079 41336 40977 41480

The head of the quality unit wants a 95% confidence interval for the mean strength of the pins in the lot.

$\leq \mu \leq$

Note: *You can earn partial credit on this problem.*

Show correct answers

You have attempted this problem 0 times.
This homework set is closed.

Figure 30 Problem for Confidence Interval, Population Mean, Pre-Test, Part 1

(1 pt) The breaking strengths of four large metal pins used in building construction randomly chosen from a large production lot are (in psi).

40056 41425 40088 40273

The head of the quality unit wants a 95% confidence interval for the mean strength of the pins in the lot. He computes it and the confidence interval is (39426, 41494).

If the head of the quality unit expects the mean breaking strength of pins to be at least 44000, what can he conclude?

- A. Since the confidence interval lies entirely below the expectation of 44000, he is 95% confident that the mean breaking strength of those pins doesn't meet the expectation.
- B. Since 44000 is contained in the confidence interval, he is 95% confident that the mean breaking strength of those pins meets the expectation.
- C. Since the confidence interval lies entirely below the expectation of 44000, he is 95% confident that the mean breaking strength of those pins meets the expectation.
- D. He can be sure that all pins don't meet the expectation.

Show correct answers

Preview Answers

Check Answers

Figure 31 Problem for Confidence Interval, Population Mean, Pre-Test, Part 2

WPI sets an expectation of 17 hours work per week for a $1/3$ unit seven week undergraduate course, which equals 119 hours per term. To see if reality meets expectations, a WPI project team took a random sample of 7 $1/3$ unit courses and from each course, randomly selected one student volunteer to follow during the term. Each selected student agreed to keep track of the hours spent on coursework for that course. The totals for the term are below:

100 118 113 102 94 112 89

The project team was interested in determining how the average time spent by all WPI students on a $1/3$ unit seven week undergraduate course compares with the 119 hour guideline.

Construct the 95% confident interval for the mean of studying hours per term.

$\leq \mu \leq$

Note: *You can earn partial credit on this problem.*

Show correct answers

Figure 32 Problem for Confidence Interval, Population Mean, Control Group, Part 1

WPI sets an expectation of 17 hours work per week for a $\frac{1}{3}$ unit seven week undergraduate course, which equals 119 hours per term. To see if reality meets expectations, a WPI project team took a random sample of $7 \frac{1}{3}$ unit courses and from each course, randomly selected one student volunteer to follow during the term. Each selected student agreed to keep track of the hours spent on coursework for that course. The totals for the term are below:

90 114 113 80 118 90 107

The project team was interested in determining how the average time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course compares with the 119 hour guideline. The 95% confidence interval computed from these data is (88.01, 115.41).

Based on this interval, what can the project team conclude about how the average time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course compares with the 119 hour guideline?

- A. The confidence interval lies entirely below the WPI expectation, therefore the project team can conclude with 95% confidence that the average level of work expended is below the WPI expectation.
- B. They can be almost sure all students at WPI spend more than 65 hours per term.
- C. Since 80 is not in the confidence interval, no one at WPI spends less than 80 hours.
- D. The confidence interval lies entirely below the WPI expectation, therefore the project team cannot conclude with 95% confidence that the average level of work expended is below the WPI expectation.

Figure 33 Problem for Confidence Interval, Population Mean, Control Group, Part 2

(1 pt) It has been hypothesized that individuals suffering from Alzheimer's disease may spend less than the normal amount of time per night in the deeper stages of sleep. To test this hypothesis researchers measured the time, in minutes, spent in stage IV sleep (the deepest level of sleep) for a sample of seven Alzheimer's patients (data below). In healthy individuals, the average time spent in stage IV sleep is 120 minutes per night. Using these data, compute a 95 percent confidence interval for the mean time spent in stage IV sleep for Alzheimer's patients.

45 40 44 50 50 60 51

Construct a 95% confidence interval for the mean time of stage IV sleep.

$\leq \mu \leq$

Note: You can earn partial credit on this problem.

Show correct answers

Figure 34 Problem for Confidence Interval, Population Mean, Post-Test, Part 1

(1 pt) It has been hypothesized that individuals suffering from Alzheimer's disease may spend less than the normal amount of time per night in the deeper stages of sleep. To test this hypothesis researchers measured the time, in minutes, spent in stage IV sleep (the deepest level of sleep) for a sample of seven Alzheimer's patients (data below).

60 44 51 60 54 60 54

Using these data, the 95 percent confidence interval for the mean time spent in stage IV sleep for Alzheimer's patients is (49.1, 60.2). Given the fact that in healthy individuals, the average time spent in stage IV sleep is 120 minutes per night, which following can the researchers conclude?

- A. Since 120 is above the confidence interval, they can be 95% confident that the mean time Alzheimer's patients spend in stage IV sleep is less than the normal average.
- B. They cannot conclude anything since the sample size is smaller than 10.
- C. Since 50 is contained in the confidence interval, they can be 95% confident that no Alzheimer's patients spend less than 50 minutes in stage IV sleep
- D. Since 120 is not contained in the confidence interval, they can be 95% confident that on average Alzheimer's patients spend 120 minutes in stage IV sleep.

Show correct answers

Figure 35 Problem for Confidence Interval, Population Mean, Post-Test, Part 2

D. Problem Sets for Hypothesis Test Experiment

In this appendix, we provide screen shots of the hypothesis test problems we created.

D.1 Population Mean

(262 pts) **The link to the distribution calculator:** http://www.tutor-homework.com/statistics_tables/statistics_tables.html#t. **You need it at for all problems. Right click it and open it in a new tab.**

Pretest question 1 The breaking strengths of four large metal pins used in building construction randomly chosen from a large production lot are (in psi).

44979 44002 45774 45001

The head of the quality unit expects the mean breaking strength of pins to be greater than 44000, so he wants to test the hypotheses:
 $H_0 : \mu = 44000$ $H_a : \mu > 44000$, where μ is the mean strength of the pins in the lot.

Compute the p-value for this hypothesis test. p =

Figure 36 Problem for Hypothesis Test, Population Mean, Pre-Test, Part 1

(262 pts) **Pretest question 2** The breaking strengths of four large metal pins used in building construction randomly chosen from a large production lot are (in psi).

44979 44002 45774 45001

The head of the quality unit expects the mean breaking strength of pins to be greater than 44000, so he wants to test the hypotheses:
 $H_0 : \mu = 44000$, $H_a : \mu > 44000$ where μ is the mean strength of the pins in the lot.

If the p-value is 0.0406, what conclusion can the head of the quality make about the null hypothesis at the 0.05 level?

- A. He cannot get the precise p-value, no conclusion can be made.
- B. Since the p-value is less than 0.05, he cannot reject the null hypothesis.
- C. Since the p-value is less than 0.05, he can reject the null hypothesis.
- D. Since the p-value is greater than 0.05, he cannot reject the null hypothesis.

Figure 37 Problem for Hypothesis Test, Population Mean, Pre-Test, Part 2

WPI sets an expectation of 17 hours work per week for a $\frac{1}{3}$ unit seven week undergraduate course, which equals 119 hours per term. To see if reality meets expectations, a WPI project team took a random sample of 7 $\frac{1}{3}$ unit courses and from each course, randomly selected one student volunteer to follow during the term. Each selected student agreed to keep track of the hours spent on coursework for that course. The totals for the term are below:

99 119 105 113 96 120 106

The project team was interested in determining whether the average time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course is lower than the 119 hour guideline.

1. The Scientific Hypothesis:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

- A. The average time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course is equal to the 119 hour guideline.
- B. The maximum time spent by WPI students on a $\frac{1}{3}$ unit seven week undergraduate course is greater than the 119 hour guideline.
- C. The maximum time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course is equal than the 119 hour guideline.
- D. The average time spent by all WPI students on a $\frac{1}{3}$ unit seven week undergraduate course is lower than the 119 hour guideline.

Figure 38 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 1

1. The Scientific Hypothesis:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

The average time spent by all WPI students on a 1/3 unit seven week undergraduate course is lower than the 119 hour guideline.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. What are the measurements in this problem, and what is the Statistical Model?

- A. The hours WPI students spend on 1/3 unit seven week undergraduate courses per week; $N(\mu, \sigma^2)$
- B. The hours WPI students spend on 1/3 unit seven week undergraduate courses; $b(n, p)$
- C. The hours WPI students spend on 1/3 unit seven week undergraduate courses; $N(\sigma, \sigma^2)$
- D. The hours WPI students spend on 1/3 unit seven week undergraduate courses per term; $N(\mu, \sigma^2)$

Figure 39 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 2

1. The Scientific Hypothesis:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

The average time spent by all WPI students on a 1/3 unit seven week undergraduate course is lower than the 119 hour guideline.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The hours WPI students spend on 1/3 unit seven week undergraduate courses per term; $N(\mu, \sigma^2)$

3. The Statistical Hypotheses:

What are the Statistical Hypotheses for this problem?

- Ⓐ $H_0 : \mu = 119, H_a : \mu > 119$ where μ is the average time spent by all WPI students on a 1/3 unit seven week undergraduate course.
- Ⓑ $H_0 : \mu \leq 119, H_a : \mu > 119$ where μ is the minimum time spent by all WPI students on a 1/3 unit seven week undergraduate course.
- Ⓒ $H_0 : \mu = 119, H_a : \mu < 119$ where μ is the average time spent by all WPI students on a 1/3 unit seven week undergraduate course.
- Ⓓ $H_0 : \mu = 119, H_a : \mu > 119$ where μ is the maximum time spent by all WPI students on a 1/3 unit seven week undergraduate course.

Figure 40 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 3

1. The Scientific Hypothesis:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

The average time spent by all WPI students on a 1/3 unit seven week undergraduate course is lower than the 119 hour guideline.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The hours WPI students spend on 1/3 unit seven week undergraduate courses per term; $N(\mu, \sigma^2)$

3. The Statistical Hypotheses:

$H_0 : \mu = 119, H_a : \mu < 119$ where μ is the average time spent by all WPI students on a 1/3 unit seven week undergraduate course.

4.1 The Test Statistic:

Obtain a single number estimate of the average number of hours per term spent by WPI students for a 1/3 unit seven week undergraduate course.

$\bar{y} =$

Figure 41 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 4

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The hours WPI students spend on 1/3 unit seven week undergraduate courses per term; $N(\mu, \sigma^2)$

3. The Statistical Hypotheses:

$H_0 : \mu = 119, H_a : \mu < 119$ where μ is the average time spent by all WPI students on a 1/3 unit seven week undergraduate course.

4.1 The Test Statistic:

The single number estimate of the average number of hours per term spent by WPI students for a 1/3 unit seven week undergraduate course is:

$$\bar{y} = 108.28$$

4.2 The p-value:

Obtain the p-value for the Statistical Hypotheses above:

p-value=

Note: *This problem has more than one part.*

You have tried 0 of 3 times in this part.

Your score for this attempt is for this part only;

your overall score is for all the parts combined.

Show hint

Figure 42 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 5

month. The coach randomly selected four players and recorded their training hours every day for a month. Suppose the resulting data are 80,70,65,80. The coach was interested in determining whether the average training time spent by players per month is lower than the 75 hours.

The Statistical Hypotheses:

$H_0 : \mu = 75, H_a : \mu < 75$ where μ stands for the average training time spent by players per month.

Compute the p-value and interpret the result at 0.05 level.

First, we find the sample mean

$$\bar{y} = \frac{y_1+y_2+y_3+y_4}{n} = \frac{80+70+65+80}{4} = 73.75$$

and the sample standard deviation,

$$\begin{aligned} s &= \sqrt{\frac{(y_1-\bar{y})^2+(y_2-\bar{y})^2+(y_3-\bar{y})^2+(y_4-\bar{y})^2}{n-1}} \\ &= \sqrt{\frac{(80-73.75)^2+(70-73.75)^2+(65-73.75)^2+(80-73.75)^2}{3}} \\ &= 7.5 \end{aligned}$$

Next, we find the observed value of the standardized test statistic t^* . Since the sample size is $n = 4$, $t^* = \frac{\bar{y}-\mu}{s/\sqrt{n}} = \frac{73.75-75}{7.5/\sqrt{4}} = -0.333$

So the p-value is then computed as

$$\Pr(t_{n-1} \leq t^*) = \Pr(t_3 \leq -0.333) = 0.3804$$

Since p-value = 0.3804 > 0.05, the coach cannot reject null hypothesis.

Figure 43 Problem for Hypothesis Test, Population Mean, Treatment Group, Hint Part

per term; $N(\mu, \sigma^2)$

3. The Statistical Hypotheses:

$H_0 : \mu = 119, H_a : \mu < 119$ where μ is the average time spent by all WPI students on a 1/3 unit seven week undergraduate course.

4.1 The Test Statistic:

The single number estimate of the average number of hours per term spent by WPI students for a 1/3 unit seven week undergraduate course is:

$$\bar{y} = 108.28$$

4.2 The p-value:

p-value=0.011646.

5. Results and Interpretation:

If the p-value is 0.0116, what conclusion can the project team make about the null hypothesis at the 0.05 level?

- A. They cannot get the precise p-value, no conclusion can be made.
- B. Since the p-value is greater than 0.05, they cannot reject the null hypothesis.
- C. Since the p-value is less than 0.05, they cannot reject the null hypothesis.
- D. Since the p-value is less than 0.05, they can reject the null hypothesis.

Figure 44 Problem for Hypothesis Test, Population Mean, Treatment Group, Part 6

- Ⓐ $H_0 : \mu = 119, H_a : \mu < 119$ where μ is the minimum time spent by all WPI students on a 1/3 unit seven week undergraduate course.
- Ⓑ $H_0 : \mu = 119, H_a : \mu > 119$ where μ is the average time spent by all WPI students on a 1/3 unit seven week undergraduate course.
- Ⓒ $H_0 : \mu = 119, H_a : \mu < 119$ where μ is the average time spent by all WPI students on a 1/3 unit seven week undergraduate course.
- Ⓓ $H_0 : \mu = 119, H_a : \mu < 119$ where μ is the maximum time spent by all WPI students on a 1/3 unit seven week undergraduate course.

4.1 The Test Statistic:

Obtain a single number estimate of the average number of hours per term spent by WPI students for a 1/3 unit seven week undergraduate course.

$\bar{y} =$

4.2 The p-value:

Obtain the p-value for the Statistical Hypotheses above:

p-value= .

5. Results and Interpretation:

Suppose the project team select another sample, and find a p-value of 0.016. What conclusion can they make about the null hypothesis at the 0.05 level?

- Ⓐ Since the p-value is less than 0.05, they can reject the null hypothesis.
- Ⓑ Since the p-value is less than 0.05, they cannot reject the null hypothesis.
- Ⓒ Since the p-value is greater than 0.05, they cannot reject the null hypothesis.
- Ⓓ They cannot get the precise p-value, no conclusion can be made.

Figure 45 Problem for Hypothesis Test, Population Mean, Control Group

(262 pts) **Posttest question 1** It has been hypothesized that individuals suffering from Alzheimer's disease may spend less than the normal amount of time per night in the deeper stages of sleep. To test this hypothesis researchers measured the time, in minutes, spent in stage IV sleep (the deepest level of sleep) for a sample of seven Alzheimer's patients (data below). In healthy individuals, the average time spent in stage IV sleep is 120 minutes per night. The data are:

98 120 126 114 115 120 120

The researchers are interested in whether the mean of time spent in stage IV sleep of Alzheimer's patients is lower than 120 minutes, so they want to test the hypotheses:

$H_0 : \mu = 120, H_a : \mu < 120$ where μ is the mean of time spent in stage IV sleep of Alzheimer's patients.

Compute the p-value for this hypotheses test. p =

Figure 46 Problem for Hypothesis Test, Population Mean, Post-Test, Part 1

(262 pts) **Posttest question 2** It has been hypothesized that individuals suffering from Alzheimer's disease may spend less than the normal amount of time per night in the deeper stages of sleep. To test this hypothesis researchers measured the time, in minutes, spent in stage IV sleep (the deepest level of sleep) for a sample of seven Alzheimer's patients (data below).

98 120 126 114 115 120 120

The researchers are interested in whether the mean of time spent in stage IV sleep of Alzheimer's patients is lower than 120 minutes,so he wants to test the hypotheses:

$H_0 : \mu = 120, H_a : \mu < 120$ where μ is the mean of time spent in stage IV sleep of Alzheimer's patients.

If the p-value is 0.1479, what conclusion can the researchers make about the null hypothesis at the 0.05 level?

- A. Since the p-value is less than 0.05, they cannot reject the null hypothesis.
- B. Since the p-value is greater than 0.05, they cannot reject the null hypothesis.
- C. They cannot get the precise p-value, so no conclusion can be made.
- D. Since the p-value is less than 0.05, they can reject the null hypothesis.

Figure 47 Problem for Hypothesis Test, Population Mean, Post-Test, Part 2

D.2 Population Proportion

Noted television personality Stuart Smalley claims: "I'm good enough, I'm smart enough, and gosh darn it, people like me." To ascertain whether the last part of the statement is true, a national polling organization asked 100 individuals randomly sampled from the American public, "Gosh darn it, do you like Stuart?" Of the 100 sampled, 50 said they liked him and 50 said they didn't.

The national polling organization expects the proportion of people like Stuart in America to be greater than 0.5, so he wants to test the hypotheses:

$H_0 : p = 0.5$, $H_a : p > 0.5$ where p stands for the proportion of people like Stuart in America.

Compute the p-value for a large sample hypothesis test. $p =$

Figure 48 Problem for Hypothesis Test, Population Proportion, Pre-Test, Part 1

Noted television personality Stuart Smalley claims: "I'm good enough, I'm smart enough, and gosh darn it, people like me." To ascertain whether the last part of the statement is true, a national polling organization asked 100 individuals randomly sampled from the American public, "Gosh darn it, do you like Stuart?" Of the 100 sampled, 50 said they liked him and 50 said they didn't.

The national polling organization expects the proportion of people like Stuart in America to be greater than 0.5, so he wants to test the hypotheses:

$H_0 : p = 0.5$, $H_a : p > 0.5$ where p stands for the proportion of people like Stuart in America.

If the p-value is 0.5398, what conclusion can the national polling organization make about the null hypothesis at the 0.05 level?

- A. They cannot get the precise p-value, so no conclusion can be made.
- B. Since the p-value is less than 0.05, they can reject the null hypothesis.
- C. Since the p-value is greater than 0.05, they cannot reject the null hypothesis.
- D. Since the p-value is less than 0.05, they cannot reject the null hypothesis.

Figure 49 Problem for Hypothesis Test, Population Proportion, Pre-Test, Part 2

Historically, a producer of high-quality jewelry has found that 10% of the top-grade diamonds it buys have serious defects after cutting and polishing. To monitor quality, a diamond company inspects a random sample of 200 recently processed diamonds from the latest shipment. They find that 18 have serious defects. Based on these results, the head of the quality division wants to obtain a single number estimate of the proportion of defective diamonds in the shipment. He also wants to know whether the proportion of defective diamonds in the shipment is different from the historical value of 10% .

1. The Scientific Goal:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

- A. The proportion of defective diamonds in the shipment differs from the historical value of 10%.
- B. The proportion of defective diamonds in the shipment is greater than the historical value of 10%.
- C. The proportion of defective diamonds in the sample differs from the historical value of 10%.
- D. The proportion of defective diamonds in the shipment is smaller than the historical value of 10%.

Figure 50 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 1

Historically, a producer of high-quality jewelry has found that 10% of the top-grade diamonds it buys have serious defects after cutting and polishing. To monitor quality, a diamond company inspects a random sample of 200 recently processed diamonds from the latest shipment. They find that 18 have serious defects. Based on these results, the head of the quality division wants to obtain a single number estimate of the proportion of defective diamonds in the shipment. He also wants to know whether the proportion of defective diamonds in the shipment is different from the historical value of 10% .

1. The Scientific Hypothesis:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

The proportion of defective diamonds in the shipment differs from the historical value of 10%.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. What are the measurements in this problem, and what is the Statistical Model?

- A. The proportion of defective diamonds in the sample; $N(n, p)$
- B. The number of defective diamonds in the sample; $b(10, p)$
- C. The proportion of perfect diamonds in the sample; $b(10, p)$
- D. The proportion of defective diamonds in the sample; $N(\sigma, \sigma^2)$

Figure 51 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 2

1. The Scientific Hypothesis:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

The proportion of defective diamonds in the shipment differs from the historical value of 10%.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The number of defective diamonds in the sample; $b(10, p)$

3. The Statistical Hypotheses:

What is the Statistical Hypotheses for this problem?

- A. $H_0 : p = 0.1, H_a : p \neq 0.1$ where p stands for the proportion of defective diamonds in the shipment.
- B. $H_0 : p = 0.1, H_a : p < 0.1$ where p stands for the proportion of defective diamonds in the shipment.
- C. $H_0 : p = 0.1, H_a : p > 0.1$ where p stands for the proportion of defective diamonds in the shipment.
- D. $H_0 : p = 0.1, H_a : p \neq 0.1$ where p stands for the proportion of defective diamonds in the sample.

Figure 52 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 3

1. The Scientific Hypothesis:

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

The proportion of defective diamonds in the shipment differs from the historical value of 10%.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The number of defective diamonds in the sample; $b(10, p)$

3. The Statistical Hypotheses:

$H_0 : p = 0.1, H_a : p \neq 0.1$ where **p** stands for the proportion of defective diamonds in the shipment.

4.1 Point Estimate:

Obtain a single number estimate of the population proportion.

$\hat{p} =$

Figure 53 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 4

The Scientific Hypothesis is the hypothesized outcome of the experiment or study. In this problem, the Scientific Hypothesis is:

The proportion of defective diamonds in the shipment differs from the historical value of 10%.

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The number of defective diamonds in the sample; $b(10, p)$

3. The Statistical Hypotheses:

$H_0 : p = 0.1, H_a : p \neq 0.1$ where **p stands for the proportion of defective diamonds in the shipment.**

4.1 Point Estimate:

Obtain a single number estimate of the population proportion: $\hat{p} = 0.09$

4.2 The p-value:

Obtain the p-value for the a large sample test of the above hypotheses:

p-value= .

Figure 54 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 5

4.2 The p-value:

Obtain the p-value for the Statistical Hypotheses above:

p-value= .

HINT:

The following is an example problem which is similar to the above problem you are working on. You can look at the solution of the following example and it will help you to solve the original problem.

To monitor product quality, the quality control unit of a manufacturing company randomly selects 200 products and 11 of them are defective. The head of the quality wants to know whether the proportion of the products in the sample differs from 0.05.

He constructs the statistical hypotheses: $H_0 : p = 0.05$, $H_a : p \neq 0.05$ where p stands for the proportion of defective products in the population. Compute the p-value and interpret the result at the level of 0.05.

We have the sample size $n=200$, $y^*=11$ and $p_0=0.05$, then

$$Z_u^* = \frac{y^* - np_0 - 0.5}{\sqrt{np_0(1-p_0)}} = \frac{11 - 0.05 * 200 - 0.5}{\sqrt{200 * 0.05 * (1 - 0.05)}} = 0.0162$$

$$p_+ = \Pr(N(0, 1) \geq Z_u^*) = \Pr(N(0, 1) \geq 0.0162) = 0.4935$$

$$Z_i^* = \frac{y^* - np_0 + 0.5}{\sqrt{np_0(1-p_0)}} = \frac{11 - 0.05 * 200 + 0.5}{\sqrt{200 * 0.05 * (1 - 0.05)}} = 0.48666$$

$$p_- = \Pr(N(0, 1) \geq Z_i^*) = \Pr(N(0, 1) \leq 0.48666) = 0.6868$$

$$\text{Since } p_{\pm} = 2 \min(0.4935, 0.6868) = 0.987$$

Figure 55 Problem for Hypothesis Test, Population Proportion, Treatment Group, Hint Part

2. The Statistical Model:

The Statistical Model is the distribution of the population of measurements that are being taken. In this problem, the Statistical Model is:

The number of defective diamonds in the sample; $b(10, p)$

3. The Statistical Hypotheses:

$H_0 : p = 0.1, H_a : p \neq 0.1$ where **p** stands for the proportion of defective diamonds in the shipment.

4.1 Point Estimates:

Obtain a single number estimate of the population proportion: $\hat{p} = 0.09$

4.2 The p-value:

p-value=0.723674.

5. Results and Interpretation:

If the p-value is 0.7236, what conclusion can the head of the quality make about the null hypothesis at the 0.05 level?

- A. Since the p-value is greater than 0.05, he cannot reject the null hypothesis.
- B. he cannot get the precise p-value, so no conclusion can be made.
- C. Since the p-value is less than 0.05, he can reject the null hypothesis.
- D. Since the p-value is less than 0.05, he cannot reject the null hypothesis.

Figure 56 Problem for Hypothesis Test, Population Proportion, Treatment Group, Part 6

What is the Statistical Hypotheses for this problem?

- Ⓐ $H_0 : p = 0.1, H_a : p \leq 0.1$ where p stands for the proportion of defective diamonds in the shipment.
- Ⓑ $H_0 : p = 0.1, H_a : p \neq 0.1$ where p stands for the proportion of defective diamonds in the shipment.
- Ⓒ $H_0 : p = 0.1, H_a : p > 0.1$ where p stands for the proportion of defective diamonds in the shipment.
- Ⓓ $H_0 : p = 0.1, H_a : p \leq 0.1$ where p stands for the proportion of defective diamonds in the sample.

4.1 The Test Statistic:

Obtain a single number estimate of the proportion of defective diamonds in the shipment.

$\hat{p} =$

4.2 The p-value:

Obtain the p-value for a large sample test of the above statistical hypotheses:

p-value= .

5. Results and Interpretation:

Suppose the head of quality selects another sample, and finds a p-value of 0.56. What conclusion can he make about the null hypothesis at the 0.05 level?

- Ⓐ Since the p-value is less than 0.05, he cannot reject the null hypothesis.
- Ⓑ He cannot get the precise p-value, so no conclusion can be made.
- Ⓒ Since the p-value is greater than 0.05, he cannot reject the null hypothesis.
- Ⓓ Since the p-value is less than 0.05, he can reject the null hypothesis.

Figure 57 Problem for Hypothesis Test, Population Proportion, Control Group

A local Boston magazine wants to know whether most people in Boston support Boston Celtics rather than other teams. They randomly asked 1000 people living in Boston, "Which NBA team is your favorite?" 924 of 1000 said they supported Boston Celtics, and 76 of 1000 people said they liked other teams.

The editors of the magazine are interested in whether the proportion of people support Boston Celtics in sample is greater than 0.9, so he wants to test the hypotheses:

$H_0 : p = 0.9, H_a : p > 0.9$ where p stands for the proportion of people support Boston Celtics in sample.

Compute the p-value for a large sample hypothesis test. $p =$

Figure 58 Problem for Hypothesis Test, Population Proportion, Post-Test, Part 1

A local Boston magazine wants to know whether most people in Boston support Boston Celtics rather than other teams. They randomly asked 1000 people living in Boston, "Which NBA team is your favorite?" 924 of 1000 said they supported Boston Celtics, and 76 of 1000 people said they liked other teams.

The editors of the magazine are interested in whether the proportion of people support Boston Celtics in sample is greater than 0.9, so he wants to test the hypotheses:

$H_0 : p = 0.9, H_a : p > 0.9$ where p stands for the proportion of people support Boston Celtics in sample.

If the p-value is 0.0066, what conclusion can the editors of the magazine make about the null hypothesis at the 0.05 level?

- A. Since the p-value is less than 0.05, they can reject the null hypothesis.
- B. Since the p-value is less than 0.05, they cannot reject the null hypothesis.
- C. They cannot get the precise p-value, so no conclusion can be made.
- D. Since the p-value is greater than 0.05, they can reject the null hypothesis.

Figure 59 Problem for Hypothesis Test, Population Proportion, Post-Test, Part 2

Work Cited

Ahlgren, J. G., & Andrew. (1988). Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research. *Journal for Research in Mathematics Education* , 44-63.

ASSISTments. (n.d.). Retrieved 9 14, 2010, from Wikipedia:
http://teacherwiki.assistment.org/wiki/index.php/Main_Page

Bassok, M., Wu, L. L., & Olseth, K. L. (1995). *Judging a book by its cover: Interpretative effects of content on problem-solving transfer. Memory & Cognition*, 23, 354-367

Bonham, S. W., Deardorff, D. F., Beichner, R. J. (2003). Comparison of Student Performance Using Web and Paper-Based Homework in College-Level Physics. *Journal Of Research In Science Teaching* , vol.40.1050-1071.

Bude Su, Curtis J. Bonk, Richard J. Magjuka, Xiaojing Liu & Seung-hee Lee. (2005). The Importance of Interaction in Web-based Education: A Program-level Case Study of Online MBA Courses. *Journal of Interactive Online Learning* , Volume 4.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology* , 127(4), 355-376.

Chi, Feltovich & Glaser. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121-152.

Cooper, H. (2008). *A Brief History of Homework in the United States*. Reston, VA: National Council of Teachers of Mathematics.

Dedic, H., S. Rosenfield, I. Ivanov. (2008). Online assessments and interactive classroom sessions: a potent prescription for ailing success rates in Social Science Calculus. *MERLOT Journal of Online Learning and Teaching*, Vol 4, No. 4

Demirci, N. (2007). University students' Perceptions of Web-based vs. Paper-based Homework in a General Physics Course. *Eurasia Journal of Mathematics, Science & Technology Education* , 3(1), 29-34.

Fischbein. (1975). *The Intuitive Sources of Probabilistic Thinking in Children (Synthese Library) (No. 85)*. Reidel, Netherlands: Kluwer Academic Print on Demand.

Fynnewer, H. (2008). A Comparison of the Effectiveness of Web-based and Paper-based Homework for General Chemistry. *The Chemical Educator*, 13, 264-269.

Garrison, D. R., Shale, D. (1990). *Education at a Distance: From issues to practice*. Melbourne, FL: Krieger.

Hauk, S., Powers, A. Safer, A. Segalla. (2004). Impact of the web-based homework program WeBWork on student performance in moderate enrollment college algebra courses. Available: <http://hopper.unco.edu/faculty/personal/hauk/segalla/WBWquan.pdf>

Hiemstra, R. E. (1982). *Self-Directed Adult Learning: Some Implications for Practice*. Syracuse, NY: Adult Education Program, School of Education.

Hillman, D. C., Willis, D. J., Gunawardena, C.N. (1994). Learner-interface interaction in distance education: An extension of contemporary models and strategies for practitioners. *The American Journal of Distance Education* , 8(2),30-42.

Moore, M. G. (1989). Three types of interaction. *The American Journal of Distance Education* , 3(2),1-6.

Palloff, R.M., Pratt, K (1999). *Building learning communities in cyberspace*. San Francisco: Jossey-Bass Publishers.

Palocsay, S. W., Stevens, S. P. (2008). *A study of the Effectiveness of Web-Based Homework in Teaching Undergraduate Business Statistics*. *Decision Sciences Journal of Innovative Education*, Volume 6, 213-232.

Penland, P. (1977). *Individual Self-planned Learning in America*. Pittsburgh, PA: Graduate School of Library and Information Sciences, University of Pittsburgh.

Phillips, G. M., Santoro, G. M., & Kuehn, S. A. (1988). The use of computer mediated communication in training students in group problem-solving and decision making techniques. *The American Journal of Distance Education* , 2(1),38-51.

Problem Authoring Background Information. (2009, 11 5). Retrieved 9 18, 2010, from Wikipedia: http://webwork.maa.org/wiki/Problem_Authoring_Background_Information

Release notes for PG 2.3.1. (n.d.). Retrieved 10 10, 2010, from WeBWork: http://webwork.maa.org/wiki/Release_notes_for_PG_2.3.1#Introduction

Renkl, A., Atkinson, R.K., & Maier, U.H. (2000). From studying examples to solving problems: Fading worked-out solution steps helps learning. In L. Gleitman & A.K. Joshi (Eds.), *Proceeding of the 22nd Annual Conference of the Cognitive Science Society* (pp. 393–398). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. retrieved February 6, 2010 from HYPERLINK "http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.6816" <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.6816>

Salden, R., Alevan, V., Renkl, A., & Schwonke, R. (2008). Worked examples and tutored problem solving: redundant or synergistic forms of support? In C. Schunn (Ed.). *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*. New York: Lawrence Erlbaum , 659-664.

Schwonke, R., Wittwer, J., Aleven, V., Salden, R., Krieg, C., & Renkl, A. (2007). Can tutored problem solving benefit from faded worked-out examples? *Proceedings of European Cognitive Science Conference* (pp. 59-64). New York: Lawrence Erlbaum.

Statistics homework help. (n.d.). Retrieved 9 7, 2010, from ClassOf1:
www.TutorVista.com/statistics-help

Su, B., Bonk, C. J., Magjuka, R. J., Liu, X. J., & Lee, S. H. (2005). The Importance of Interaction in Web-based Education: A Program-level Case Study of Online MBA Courses. *Journal of Interactive Online Learning*, Volume 4.

Tough, A. (1971). *The Adult's Learning Projects: A Fresh Approach to Theory and Practice in Adult Learning*. Toronto, Canada: Ontario Institute for Studies in Education.

Tycho: Our web-based homework system. (n.d.). Retrieved 9 7, 2010, from University of Illinois:
<http://research.physics.illinois.edu/per/Tycho.html>

WebClass. (n.d.). Retrieved 9 7, 2010, from WebClass:
<https://www.mathclass.org/mc/default.aspx>

WeBWork. (n.d.). Retrieved 9 7, 2010, from Wikipedia:
http://webwork.maa.org/wiki/Main_Page

Wilson, E. O. (1998). *Consilience, quoted in Robert Park, Voodoo Science*. Oxford: Oxford University Press.