THE USE OF THE JIGSAW IN HYPOTHESIS TESTING

Fe EVANGELISTA
University of Wisconsin - Whitewater
Mathematical and Computer Sciences Department
Whitewater, WI 53190
e-mail: evangelf@mail.uww.edu

Clare HEMENWAY
University of Wisconsin - Marathon
518 South 7th Ave., Wausau, WI 54401
e-mail: chemenwa@uwc.edu

ABSTRACT

The Jigsaw is a cooperative learning technique in which the class is first divided into expert groups that are assigned different but related tasks. New "home" groups consisting of one member from each expert group are then formed. Each expert instructs the other members of this new group about what they have learned. In our application of the jigsaw, we use the expert groups to give students the opportunity to study one particular example in-depth. We use the home groups as a way to compare the different examples, and observe their similarities and differences. This gives the students a chance to create their own generalizations. We provide two examples of how the Jigsaw can be used in an introductory statistics class. The first example is for presenting different sampling techniques and a more advanced application is for introducing hypothesis testing. In the latter, we have found it effective to have expert groups use experiments to investigate a specific claim. Our experience indicates that these preliminary concrete activities will provide a smoother transition to understanding the formal and symbolic hypothesis testing framework.

Keywords: cooperative learning, hypothesis test, jigsaw, statistics

1. Introduction

It has been the experience of the authors and others (Rogers, et. al. 2001) that group activities can enhance the learning and understanding of many statistical concepts. Yet many times the students need some structure to guide their explorations. One technique that can be useful when there are a variety of similar tasks is known as the "jigsaw." This technique gets its name from the fact that individuals or groups each study a piece of a project and then the students put the pieces together to get a complete picture of the project. It is reminiscent of many spy stories in which the only way a code may be broken or a fortune recovered is if each individual in the plot contributes their portion of the puzzle.

In the cooperative learning setting, the jigsaw organizes students together in what have been called expert groups. Each expert group does a specific task and then new "home" groups are formed so that each home group includes at least one member from each expert group. Each individual expert is expected to convey their knowledge to the rest of the group members, that is, contribute their piece to the material to be mastered. The method has been primarily used in the elementary school classroom to teach social studies and reading. One of its strengths is that it creates a learning environment that "made it imperative that the children treat each other as resources" (Aronson, et. al., 1978).

The jigsaw is often used when the tasks are different but similar. It has been used in situations such as studying different positions of a certain issue and for sharing the workload of reading several related articles. In mathematics, it has been used to study the properties of complex numbers (Lucas, 2000) and for sharing the calculation tasks in standard statistical procedures (Perkins & Saris, 2001). In our own classes, we have found the technique useful when presenting the proofs of properties of an algebraic structure (for example, the proofs of the properties of logarithms or determinants.) The students often do not recognize the similarity in the structure of the proofs of these distinctive properties. However, when each group is assigned a specific proof and then they teach other students, they tend to more readily see the patterns among them. When used in this manner, perhaps the greatest benefit of employing the jigsaw is that the students notice *for themselves* the differences and similarities in these procedures. In fact, it was the authors' many frustrating experiences trying to point out to students such connections in the hypothesis testing of various parameters that led them to seek out an alternative pedagogy.

This paper details an implementation of the jigsaw to introduce new and somewhat complicated material. In what follows we suggest activities for both the expert groups and the home groups in each of the areas of sampling design and hypothesis testing.

2. Sampling Design

A good introduction to the use of the jigsaw early in the semester is in the study of sampling techniques. For the pieces of the jigsaw, the class is divided into groups and each of the groups studies one of the following sampling techniques: random, systematic, cluster, stratified and convenience sampling. Then they use that particular technique to collect data to estimate, for example, the average number of words per page in a dictionary. Similar activities have been used before to teach sampling techniques (Paranjpe & Shah, 2000), but not by the jigsaw method. New groups consisting of at least one member from each expert group are formed and each "expert" teaches their sampling method to the rest of the group. The puzzle is now complete. As a follow-up exercise to assess individual student's understanding of each of the procedures, the students are assigned to collect data to calculate the average number of advertisements in their favorite magazine using the different sampling methods. This stresses individual accountability and gives us feedback on whether or not the "experts" were successful in teaching other students.

We observed the following advantages in using the jigsaw in this setting. Because of the simplicity of the tasks, students are able to gain experience in studying a concept on their own, work with other students to plan and implement a task (all the work is outside class time), and analyze the results of the task. The students tend to be more focused and pay more attention to explanations given by other students.

3. Hypothesis Testing

In our experience, we have observed the following difficulties that students have with hypothesis testing. The obstacles occur at both the procedural and conceptual level. These observations are consistent with other studies (Hong, 1992; Albert, 2000).

(a) Inability to distinguish a test of hypothesis situation from other situations such as estimation or finding probabilities.
(b) Failure to recognize the population parameter to be tested and whether more than one population is involved.
(c) Difficulty specifying the null and alternative hypothesis and determining the rejection region.
(d) Confusing the sample and the population, resulting in a weak conceptual understanding of the role of sampling distributions in statistical inference.
(e) Difficulty interpreting their conclusion to reject or not to reject the null hypothesis in the context of the problem.
(f) Poor understanding of the reasoning behind the structure of hypothesis testing even if they can procedurally do all the textbook exercises.

In our opinion, the reasoning behind hypothesis testing is a natural process, one that is common to the students' experiences. This is why many instructors and textbooks use familiar examples such as cards, dice and coins to demonstrate the principles of statistical reasoning (Rossman, 1996; Maxwell, 1994; Eckert, 1994). We think that many of the above noted difficulties develop as the students try to translate their intuitive (and usually correct) notions of hypotheses testing to a formal statistical framework. In trying to learn the steps of hypothesis testing, as well as the terminology and the symbolism, the students tend to lose sight of the intuitive reasoning behind the process.

For this reason, we focus our pedagogy on the affirmation of the students' intuitive notions of hypothesis testing and then on the transference of this knowledge to the components of the formal framework of hypothesis testing. In order to facilitate this transition, we would like our students to first realize that the reasoning behind hypothesis testing is natural and within their range of experiences. Then we introduce the students to the formal procedures of a hypothesis test in a specific setting. Finally, we require our students to reflect on the *process* of hypothesis testing by studying several examples in a variety of accessible contexts.

One can focus lectures to address these concerns and we have observed that most students do see patterns in hypothesis testing, but that they tend to focus on the pattern in the formalism and symbolism rather than on the reasoning behind the formalism. As a result, they tend to view hypothesis testing as a collection of independent algorithms, one for each parameter. We feel that group work would allow the students more opportunity and time to grapple with the issues involved in making the transition to the formal framework.

However, it has been our experience that understanding one example in depth, even in a group setting, is not sufficient to reinforce the students' innate understanding of inference. In particular, we observed that students have trouble applying the knowledge that they gained in studying a specific example to similar situations. We believe that it would be helpful if the students had the chance to work through many similar, but related examples, so that they can identify common patterns, as well as differences. But of course, this is time-consuming.

Our implementation of the jigsaw, as described below, gives students this opportunity, in a relatively short amount of time. By studying a single problem in their expert groups, and then sharing each other's work in the students' home groups, the students are guided to recognize that they are procedurally doing the same kinds of tasks in different settings---thus focusing on the structure and not the technical details. Also, their understanding is reinforced when they communicate their knowledge to other members of the group.

4. Implementation

The class is divided into "expert" groups of three to five people each. Each group is given a worksheet with a story that involves testing a claim. In our statistics classes (typically classes of 25 – 35 students), there will be 6 to 8 groups and two "expert" groups working on identical claims. The claims we use vary in three aspects: the parameter to be tested (proportion or mean), the rejection region (one-tailed or two-tailed), and the conclusion of the test (reject the null hypothesis H_0 or not reject the null hypothesis H_0). A sample worksheet is included in the appendix, and examples of claims we have used are listed below:

1. Is a specific coin fair? (test for a proportion, two-tailed, do not reject H_0)
2. Is a specific coin more likely to show heads? (test for a proportion, one-tailed, reject H_0)
3. Are there about 56 M&M's in a 47.9-g bag? (test for a mean, two-tailed, do not reject H_0)
4. Is the average number of M&M's in a 47.9-g bag greater than 42? (test for a mean, one-tailed, reject H_0)

The numbering system used in the worksheet parallels the formal steps used in hypothesis testing and the wording of the questions in number 3 of worksheet #1 emphasizes the condition under which a hypothesis test is conducted, namely, that the null hypothesis is true.

After completing this part of the activity, the students read a summary of the hypothesis testing procedure. The terminology, symbolic representation and step-by-step procedure are introduced, except that there is no mention of the significance level nor is there a formal presentation of Type I and Type II errors. Then they translate their earlier observations into the formal framework (see sample worksheet #2 in the appendix).

To put the pieces of the jigsaw together, the class forms new "home" groups with at least one person from each expert group. Each "expert" teaches the rest of the group about their problem, procedure and results. One advantage of duplication of group assignments is that there are two experts for that problem in the home group, thereby reducing the possibility of having an ineffective "expert." To complete the picture of this statistical jigsaw puzzle, the students are asked to synthesize their results by noting the similarities and differences in their various tasks. Specifically, they are asked to address the following:

(a) Explain your problem and solution to the other members of the group.

(b) What similarities do you observe in all of your activities and results?
(c) What differences do you observe?

At the end of the group activities, the instructor leads the students in a class discussion to summarize their findings and address any misconceptions or missed conceptions. The amount of class time needed to complete these activities is about 1 ½ fifty-minute class periods.

5. Development of our Implementation

Our current implementation of the jigsaw is a result of two refinements of previous implementations, and is still undergoing revisions. In our initial attempt at adopting this strategy, we assumed (incorrectly) that the students' intuitive notions were well-developed and so we concentrated on the transition to the formal framework. At that time, the activities of the expert groups started with a worksheet similar to our current worksheet #2. In addition, the experiments were more complicated – some involving the difference of two population parameters – and there was no replication of the individual problems among the expert groups. Also, the types of problems were too varied for the students to readily see the common themes, and they focused on the experiments rather than on the structure of the tasks. The result was a situation that was more time-consuming and frustrating (for both students and instructor) than was anticipated. The exercises were completed with much instructor input to both the expert and home groups.

Despite the problems encountered, we were encouraged to persist with this approach, albeit with modifications. In the second round, we had the students first use the jigsaw technique to learn sampling designs so as to familiarize themselves with the technique in a relatively easy setting. Also, we simplified the experiments and focused the student's attention on the informal aspects and natural logic of hypothesis testing. This time, we felt we were more successful in developing the students' confidence in their natural problem-solving abilities. However, we were not satisfied in how the activities bridged the gap to the formal structure and terminology of statistical inference

Our combined experiences have led us to develop the particular version that is discussed in this paper and we will continue to modify the exercises and worksheets as needed. We have also developed computer simulation activities for two of the experiments to enhance our students' understanding of how a decision criterion is chosen to reject the null hypothesis. These explorations leads the students to the concepts of significance level, Type I and Type II errors.

6. Summary

As noted in the introduction, the jigsaw has most often been applied at the elementary school level and in the areas of reading and social science. In statistics, it is most often used to divide up a tedious calculation task, for example, having each expert group find the summary statistics for a particular sample to be used in an application of an ANOVA (Perkins and Saris, 2001). In this paper, we go beyond these usual ways of using the jigsaw. We use the technique in a statistics class to introduce a concept (hypothesis testing), by giving the students a chance to study different examples of this concept, in order to lead them to make their own generalizations. Their observations form the basis of further classroom discussion. We feel that in this way, a concept, procedure or formula would be more grounded in the students' experience.

REFERENCES

-Aronson, E., Blaney, N., Stephan, C., Sikes, J., Snapp, M., 1978, *The Jigsaw Classroom*, Beverly Hills, CA: Sage Publications.

-Albert, J., 2000, "Using a sample survey project to assess the teaching of statistical inference", *Journal of Statistics Education*, [Online], **2**(1). (http://www.amstat.org/publications/jse/v2n1/)

-Eckert, S., 1994, "Teaching Hypothesis Testing with Playing Cards: A Demonstration", *Journal of Statistics Education*, [Online], **2**(1). (http://www.amstat.org/publications/jse/v2n1/)

-Hong, E., O'Neil Jr., H., 1992, "Instructional Strategies to Help Learners Build Relevant Mentail Models in Inferential Statistics, *Journal of Educational Psychology*, **84**, 150-159.

-Lucas, C., 2000, "Jigsaw Lesson for Operations of Complex Numbers," *Primus*, **10**, 219-224.

-Maxwell, N., 1994, "A coin-flipping exercise to introduce the p-value", *Journal of Statistics Education*, [Online], **2**(1). (http://www.amstat.org/publications/jse/v2n1/)

-Paranjpe, S., Shah, A., 2000, "How many words in a dictionary? Innovative laboratory teaching of sampling techniques", *Journal of Statistics Education*, [Online], **8**(2). (http://www.amstat.org/publications/jse/v2n1/)

-Perkins, D., Saris, R., 2001, "A Jigsaw Classroom Technique for Undergraduate Statistics Courses", *Teaching of Psychology*, **28**, 111-113.

-Rogers, E., Reynolds, B., Davidson, N., Thomas, A. (eds.), 2001, *Cooperative Learning in Undergraduate Mathematics: Issues That Matter and Strategies That Work*, Washington, DC: Mathematical Association of America.

-Rossman, A., 1996, *Workshop Statistics*, New York: Springer-Verlag.

WORKSHEET #1: Introduction to Hypothesis Testing

Paul and Joshua are again squabbling over what TV show to watch. Their mom suggested that they flip a specific coin – if it comes up heads, Paul would choose the show; if it comes up tails, Joshua would decide. Both boys were hesitant to agree, thinking that the other one had an advantage. Mom says: " It's a shiny new coin – I am sure it is a fair coin."

1. **Hypotheses**

   This scenario involves two claims.

   (a) What is the mother's claim? (This claim is the prevailing view about any new coin.)

   (b) The boys have a different, or alternative viewpoint. What is their claim?

2. **Experiment Design**

   The boys ask their mother if they can prove to her that their claim is correct. Design an experiment that the boys can perform and write your procedure below. DO NOT conduct the experiment yet.

3. **Decision Criterion**

   The mother agrees to the boys' experiment but she insists that they must all agree beforehand on what criteria they will use to decide if the coin is fair. They discuss the following situations. Imagine that you are part of the discussion, and answer the questions below.

   (a) What results would you expect if the coin is fair?

   (b) What results would you expect if the coin is not fair?

   (c) If the coin is fair, is it possible to get 80% or more heads? Is it likely to get 80% or more heads?

   (d) If the coin is fair, is it possible to get 10% or fewer heads? Is it likely to get 10% or fewer heads?

   (e) If the coin is fair, is it possible to get 45% or fewer heads? Is it likely to get 45% or fewer heads?

   A **decision criterion** is a method to decide whether a claim is valid before conducting an experiment. An example of a decision criterion is the following:

   Paul and Joshua are correct in saying that the coin is not fair if the experiment shows 55% or more heads or 45% or fewer heads.

   In your group, develop a decision criterion for your experiment that is acceptable to all members of the group, and write it below.

   At this point, consult with your instructor before proceeding.

4. **Gathering the Evidence**

   Conduct the experiment and record your results below.

5. **Decision**

   Based on your results from #4 and your decision criterion from #3, is the boys' claim that the coin is unfair supported? Is the mother's claim that the coin is fair supported?

WORKSHEET #2: Terminology and Framework of a Statistical Test of Hypothesis

A formal statistical test of a hypothesis has several components.

1. Hypotheses. Every hypothesis testing situation has two competing hypotheses or claims. One is called the *null hypothesis* and is denoted H_0. This hypothesis represents the prevailing view or the status quo. Others may believe that the null hypothesis is not true. Their viewpoint is a competing, or *alternative hypothesis*, denoted by H_1. In the situation in the first worksheet, Paul and Joshua disagree with their mother's claim. Using this formal terminology for competing claims, state in ordinary language the null and alternative hypotheses for your experiment. Then restate these hypotheses in terms of the binomial parameter *p*, the proportion of heads in *n* tosses.

	Words	Symbols
H_0:		
H_1:		

2. Experiment Design. Now that you know the claim that you (acting as Paul and Joshua's representatives) want to establish, you need to collect data to support your alternative hypothesis. However, until you can support your alternative hypothesis with evidence, you must conduct your experiment under the assumption that the null hypothesis is true.

3. Rejection Region. The term *rejection region* refers to what was called the decision criterion in worksheet #1. This region is a range of what you believe to be unlikely values obtained from your experiment if the null hypothesis were indeed true. In other words, it is a range of values obtained from your experiment, which would convince most people to reject the null, or prevailing hypothesis (this is why it is called the rejection region), in favor of your claim H_1. In your experiment, the rejection region would be those values of $\hat{p}$ that you think would be unlikely if indeed the null hypothesis is true.

What is the rejection region for your experiment?

4. Test Statistic. The test statistic is the evidence obtained from an experiment that will be used to try and refute, or reject, the null hypothesis. Generally, you compute the sample statistic that corresponds to the population parameter used to state the null and alternative hypotheses. The population parameter of interest in your experiment is the proportion or percentage of heads when a coin is flipped. The corresponding sample statistic in your problem is thus the percentage of heads observed in the sample. What is the value of the test statistic $\hat{p}$ from your experiment?

5. Decision. Lastly, a decision is made based on whether or not the value of your test statistic falls in the rejection region given in #3. If it falls in this region, then you would reject the null hypothesis and support the alternative claim. If it does not, then you cannot reject the null hypothesis. In your test, will the null hypothesis be rejected? If so, state your conclusion in ordinary language. If the null hypothesis is not rejected, state this conclusion in ordinary language.

Whatever your decision, there is a chance that your conclusion is not correct. Explain how this can happen.