



Two Applets for Teaching Hypothesis Testing

Kady Schneider
Utah State University

Journal of Statistics Education Volume 16, Number 3 (2008),
www.amstat.org/publications/jse/v16n3/schneider.html

Copyright © 2008 by Kady Schneider all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Technology; p -value; Chi-square.

Abstract

Interactive applets have the ability to enhance statistics teaching by providing multiple representations of new concepts and by facilitating experimentation. I introduce two applets that have been developed as aids in illustrating ideas relevant to hypothesis testing and describe how I have used these in my classes.

1. Introduction

Technology plays several fundamental roles in statistical practice. From the crucial roles in data analysis to knowledge dissemination to accessing or gathering data, the field of statistics is in some measure shaped by its tools. While many of these same tools are used in statistics education, in teaching the roles of technology move well beyond manipulation of data. Technology can be used to foster deep understanding of statistical principles by providing students with visual and dynamic presentations of ideas, by creating multiple and varied representations of new concepts, and by making it possible for students to investigate the principles underlying important methodologies. In particular, technology provides powerful tools for illustration and investigation. A variety of interactive statistical applets have been constructed for these purposes. These can be very effective in teaching a wide variety of statistical concepts including probability ([Cannon, 2007](#)), sampling distributions ([Lane, 2006](#)) and confidence intervals ([West, n.d.](#)). Herein, I introduce two applets for use in teaching hypothesis testing. These applets are most appropriate for lower-level undergraduate or high school statistics courses and facilitate investigation of statistical concepts rather than calculation or analysis. The applets can be found at www.math.usu.edu/~schneit/CTIS/ or by following these links: [Chi Square applet](#), [P Value applet](#).

2. Motivation

At Utah State University, we offer course taught from the text [Statistics](#) by [Freedman, Pisani, and Purves](#). The initial interest level of students in the course is usually low as the vast majority enrolls to fulfill a quantitative literacy requirement for the school or a graduation requirement for a particular department or program of study. Many of the students in the course are intimidated by mathematics and more so by

statistics. Using technology for investigation, simulation, or illustration with this audience has the potential to increase both interest in and understanding of the material. However, applications must be chosen carefully since perceived complexity in the tools can exacerbate students' negative attitudes toward the course. Many tools that are effective for teaching statistical concepts in more advanced undergraduate level statistics courses can be daunting in this setting.

The applets described herein have been developed specifically to be used in lower-level undergraduate statistics courses and pre-college curricula. The applets illustrate statistical ideas through simulation, and allow students to investigate relevant ideas to construct a deeper understanding of underlying concepts. With the specified audience in mind, objectives of the development process were to create tools that are simply designed, narrowly focused, and easy-to-use, and to construct tools that facilitate investigation. These applets address issues in hypothesis testing, a topic of difficulty for many students. The first applet works with students' intuitive understanding of probability to introduce the process of using p -values for decision making. The second applet enables the user to investigate the components of a chi-square statistic and to carry out a chi-square test to determine if a die is fair.

3. Applets


3.1 P -value Applet

The concept of a p -value is often difficult for students in introductory statistics classes. Many believe it to be the probability that the null hypothesis is true, others do not attempt to understand it, rather they try to remember a rule "reject if $p < 0.05$ ". Nevertheless, students have good intuition about what makes a thing "too unlikely to be true". This applet guides students to make a decision based on a given probability without introducing formal concepts of hypothesis testing.

The applet consists of two "views", called "test view" ([Figure 1](#)) and "investigate view" ([Figure 2](#)). In the first, the user carries out an experiment to determine if a displayed coin is fair; in the second he can investigate a p -value when the probability of tossing heads is known.

Options Choose view


Test



Is the coin fair?

Flip 1 time Flip 5 times Flip 10 times

Coin Toss Results



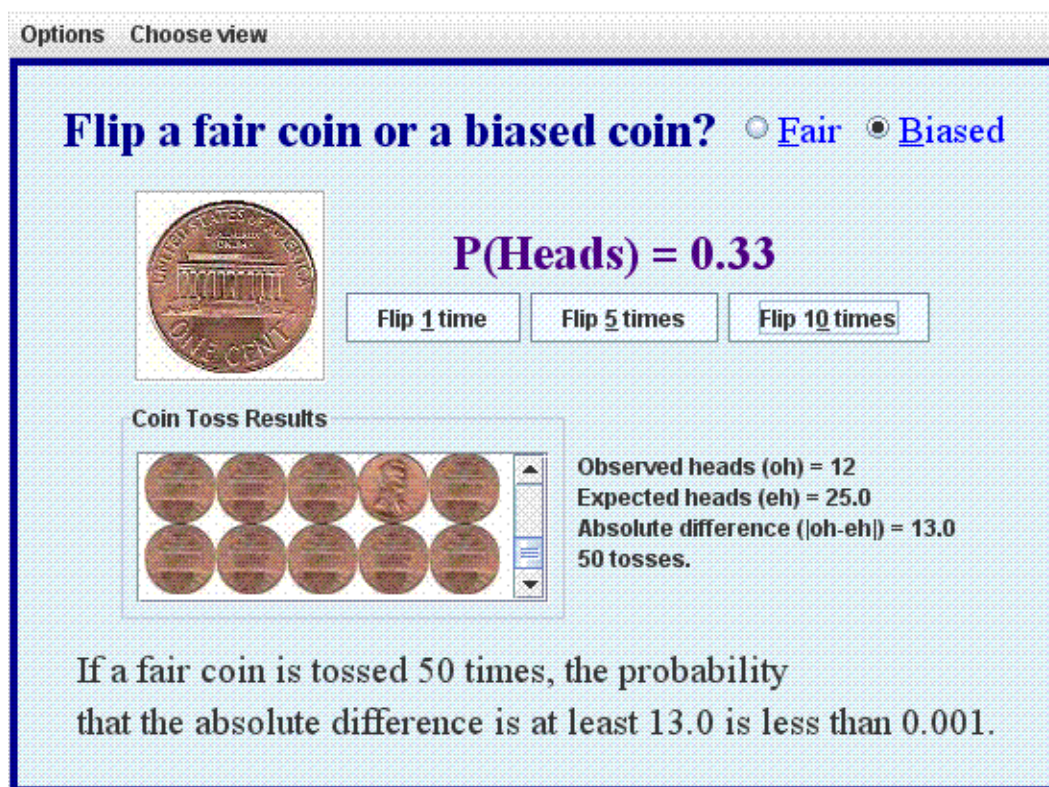
Observed heads (oh) = 28
Expected heads (eh) = 25.0
Absolute difference (|oh-eh|) = 3.0
50 tosses.

If a fair coin is tossed 50 times, the probability that the absolute difference is at least 3.0 is 0.4798.

Hypothesis: The coin is fair. Accept Reject

New coin

Figure 1: The test view of the p -value applet.



Options Choose view

Flip a fair coin or a biased coin? Fair Biased

P(Heads) = 0.33

Flip 1 time Flip 5 times Flip 10 times

Coin Toss Results

Observed heads (oh) = 12
Expected heads (eh) = 25.0
Absolute difference (|oh-eh|) = 13.0
50 tosses.

If a fair coin is tossed 50 times, the probability that the absolute difference is at least 13.0 is less than 0.001.

Figure 2: The investigate view of the p -value applet.

Guidelines for an activity in the test view are provided in the instructions. The user is directed to devise a plan for determining if the coin is fair; he should decide how many times he plans to toss the coin and decide, under the assumption that the coin is fair, how unlikely the result must be for him to decide that the coin is biased. The results of all coin tosses are displayed.

The ‘options’ menu allows the user to decide whether to conduct a one-sided or a two-sided hypothesis test. If a one-sided test is chosen, a biased coin will always favor tails. The one-sided option is included because the statement of the probability is more straight-forward. For low-level courses, it may be beneficial to begin with a discussion of the one-sided test as students are developing the basic understanding of the decision making process. It would then be natural a follow-up to discuss with students the probability that should be computed if a unfair coin will always be biased toward tails or if the direction of the bias is unknown. This could be used to lead students into thinking about two-sided vs. one-sided tests.

As the user tosses the coin, the exact probability of obtaining the observed results or more extreme results is displayed. The user can accept or reject the hypothesis "the coin is fair" based on the given probability. After the user has made his determination, the true probability of tossing heads is displayed.

Students can use the applet to investigate how the decision making process is affected by increasing the number of coin tosses. The user is asked to consider questions such as ‘for you to reject the hypothesis, how unlikely must the observed results be for a fair coin?’, ‘is it possible to get a very small p -value if the coin is fair?’, and ‘how are your decisions affected by increasing the number of times you flip the coin?’

In the second part of the applet, the investigate view, the user can decide whether to flip a fair or a biased coin and the probability of tossing heads is displayed. Knowing this probability, the user can investigate properties of the p -value. In particular, he is asked to consider the following questions: ‘If the coin is tossed 100 times and the probability of getting the observed number of heads or a more extreme number of heads (under the assumption that the coin is fair) is large, must the coin be fair?’ and ‘If the coin is tossed 100

times and the probability of getting the observed number of heads or a more extreme number of heads (under the assumption that the coin is fair) is very small, is it possible that the coin is fair?

3.2 Chi-square Applet

While the chi-square test for multiple categories is not a test that is particularly difficult, the test statistic itself is often more of a black box to students than a z -statistic or a t -statistic. Like the p -value applet, the chi-square applet has a 'test view' in which the user implements a virtual experiment to make a decision about a stated hypothesis and an investigate view to explore components of the test statistic.

In the test view (Figure 3), the user rolls a die to make a determination about the hypothesis "the die is fair". Each of the faces of the die is displayed along with the number of times that face was observed, the number of times it was expected, and the difference between the observed and expected counts. The total number of rolls, the chi-square statistic, and the p -value are also reported to facilitate making a decision regarding the hypothesis. The probabilities associated with each of the die faces are given once the hypothesis has been rejected or accepted. This is a straightforward implementation of the chi-square test.

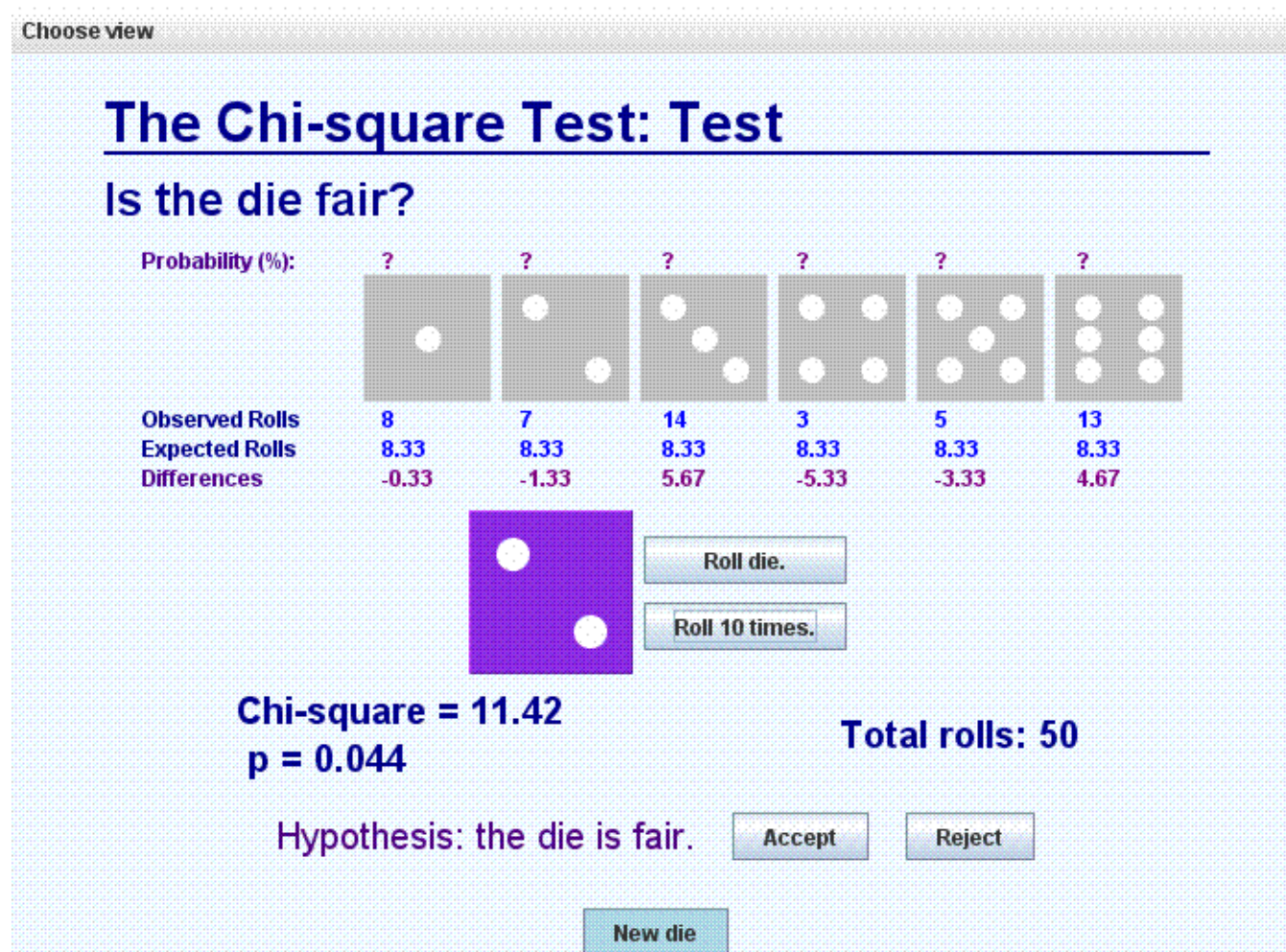


Figure 3: The test view of the chi-square applet.

Alternatively, in the 'investigate view' (Figure 4), the user can choose whether to roll a fair or a loaded die; the probabilities of observing each of the die faces are given based on the selection. The only other information that is displayed by default is the expected number of rolls for each die face and the total number

of rolls. The user then has the options to display the chi-square statistic, the p -value, the observed counts, the differences between of observed and expected counts, and the 'scaled differences' (the squared differences divided by the expected counts - the summands of the chi-square statistic). Under this construction, a user is able to investigate the behavior of the components of the test statistic with knowledge about the underlying face probabilities and to examine the effects on the test statistic. For instance, a user who chooses to use a loaded die and to display the scaled differences will observe that a die face with a probability that is very different from fair will usually contribute much more to the chi-square statistic than a face with a probability closer to fair. By allowing selected information to be displayed, the applet facilitates examination of the components individually and avoids distractions of too much information.

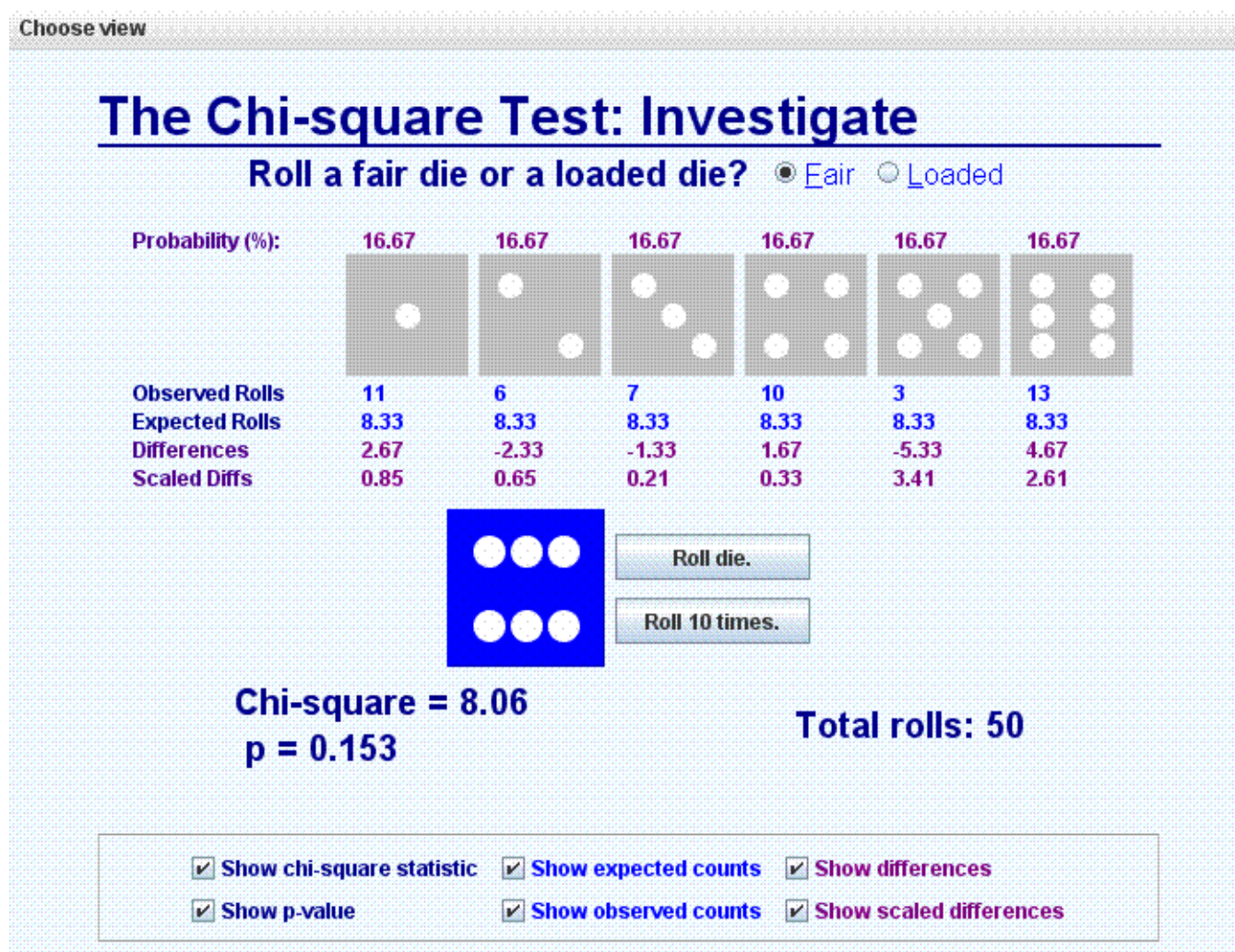


Figure 4: The investigate view of the chi-square applet.

4. Conclusion

I have used these applets in my classes in a variety of ways: for in-class examples, for homework assignments, and as a reference for extra help with hypothesis testing. In particular, I have used the p -value applet in an introduction to hypothesis testing to help students to gain an informal and intuitive sense of the process. We work with the applet before formally introducing the concepts of null and alternative hypotheses, test statistics, and p -values. I have found that students readily grasp the idea of rejecting the hypothesis that the coin is fair when the displayed probability is small, consequently this becomes a reference point for us. It is common for students to remember that "5%" relates to the rule for determining

whether or not we reject the null hypothesis but to forget on which side of 5% to reject. I am able to refer them back to the applet-motivated discussion to help them to think through the logic they used to determine whether or not a coin is fair and to relate that to a general condition for rejecting the null hypothesis.

In a recent course, I followed up the in-class discussion motivated by the p -value applet with a homework assignment based on the test view of the applet (a summary is displayed in [Exhibit 1](#)). The assignment was intended to reinforce the decision making process and was given after the terms *null hypothesis* and *p-value* had been defined. The students first created a rule for accepting or rejecting the hypothesis that the coin is fair, then carried out a number of tests, each time making a determination based on their rule. Follow-up questions were designed to stimulate students to consider the role of chance in the decision making process.

- Create a rule that you will use to determine whether or not the coin is fair. Describe your rule below.
- Toss the coin 100 times, then click on the appropriate button to ‘accept’ or ‘reject’ the null hypothesis according to the decision rule you described above. In the table, record the number of heads observed, the p -value, whether you accept or reject the null hypothesis, and the true status of the coin (fair/biased). Repeat the experiment 10 times – be sure to get a new coin each time.

Trial	Observed Heads	Probability (p -value)	Decision (Accept/Reject)	Conclusion (Fair/Biased)	True Status (Fair/Biased)
1					
:	:	:	:	:	:
10					

- Were you ever wrong about whether or not the coin was fair?
- Suppose your decision rule is ‘I reject the null hypothesis when the p -value is less than 7%’. Is it possible to get a p -value less than 7% when the coin is truly fair or a p -value greater than 7% when the coin is truly biased? Explain.

Exhibit 1: Excerpts from a homework assignment based on the p -value applet.

The homework assignment brought to light an interesting misconception in a few of the students. Some answered ‘no’ to the question “Suppose your decision rule is ‘I reject the null hypothesis when the p -value is less than 7%’. Is it possible to get a p -value less than 7% when the coin is truly fair or a p -value greater than 7% when the coin is truly biased?” These students explained their answer by pointing out that they had never been wrong with their chosen decision rule. This shows that while they could use a p -value to make a decision, they lacked understanding of the actual meaning of the p -value. This issue could probably be addressed by a better introduction to the applet or through additional questioning in the homework assignment. However, it served us well as a starting point for further discussion of the p -value and its interpretation.

The parallel formats of the chi-square and p -value applets emphasizes the commonalities and highlights overarching principles of hypothesis testing. The applets have been well received by my students. The majority of students who completed the p -value assignment reported that they felt that had learned from the applet, that they enjoyed using the applet, and that they would like to use applets to learn other concepts and subjects. I have found that using these applets and others helps to focus the students attention on new material and stimulates their interest while illuminating the targeted principle. The hypothesis testing applets have an obvious advantage over experiments with concrete coins or dice in that the probabilities of the coin sides or dice faces are not known beforehand. Furthermore, since computational distractions are removed, (e.g. in both of the applets p -values are recomputed instantaneously every time the coin is flipped or the die is rolled) the focus of a lesson can remain on the interpretation of the probability rather than the calculations needed to obtain it.

References

Cannon L, Dorward J, Heal R and Edwards L. (2007) "Spinners". (*National Library of Virtual Manipulatives*). <http://nlvm.usu.edu/> (Accessed: 2008, January 31)

Freedman D, Pisani P, and Purves R. (2007) *Statistics* 4th Ed. New York: WW Norton and Company, Inc.

Lane D. (2006) "Sampling Distribution". (*Rice Virtual Lab in Statistics*.)
http://www.ruf.rice.edu/~lane/stat_sim/sampling_dist/ (Accessed: 2008, January 31)

West W. (n.d.) "Confidence Interval Applets". (*Applets for the Cybergnostics project*)
<http://www.stat.tamu.edu/~west/applets/ci.html> (Accessed: 2008, January 31)

Kady Schneider, Ph.D.
Utah State University
Department of Mathematics and Statistics
3900 Old Main Hill
Logan, UT 84322
kady.schneider@usu.edu

[Volume 16 \(2008\)](#) | [Archive](#) | [Index](#) | [Interactive Computing Archive](#) | [Data Archive](#) | [Resources](#) | [Editorial Board](#) | [Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)