Stat 321 - Day 1
Introduction to Probability

Activity 1: Friendly Observers
In a study published in the *Journal of Personality and Social Psychology* (Butler and Baumeister, 1998), researchers investigated a conjecture that having an observer with a vested interest would decrease subjects' performance on a skill-based task. Subjects were given time to practice playing a video game that required them to navigate an obstacle course as quickly as possible. They were then told to play the game one final time with an observer present. Subjects were randomly assigned to one of two groups. One group (A) was told that the participant and observer would each win $3 if the participant beat a certain threshold time, and the other group (B) was told only that the participant would win the prize if the threshold were beaten. The threshold was chosen to be a time that they beat in 30% of their practice turns. It turned out that 3 of the 12 subjects in group A beat the threshold, while 8 of 12 subjects in group B achieved success.

	A: observer shares prize	B: no sharing of prize	Total
Beat threshold	3	8	11
Do not beat threshold	9	4	13
Total	12	12	24

(a) Calculate the sample proportions of success (beating the threshold) for each group.

(b) Do these sample proportions differ in the direction conjectured by the researchers?

(c) Even if there were absolutely no effect of the observer's interest, is it possible to have gotten a difference like this just due to chance variation? Explain.

Simulation:
The reasoning of statistical tests of significance asks how likely the sample results would have been if in fact the observer's incentive had no effect on the subjects' performance. One way to analyze this question is to assume that those 11 subjects who passed the threshold and 13 who did not would have achieved the same outcome regardless of which group they had been assigned to. In other words, we begin by assuming that the observer's interest had no effect, and we will ask how likely it is to obtain the results the researchers found given this assumption.

We can then *simulate* the process of assigning subjects at random to the two groups, just as the researchers did at the beginning of their study. Our focus will be on noting how often we obtain a sample result as extreme as (3 or fewer successes assigned to A) as in the actual sample. Repeating this a large number of times will give us a sense for how unusual the sample result would be to occur by chance alone.

(d) Mark 11 cards as "success" and 13 as "failure", shuffle them well, and randomly deal out 12 to represent the cases assigned to group A. How many of these 12 are successes? Is this result as extreme as in the actual sample?

(e) Repeat this a total of five times, recording your results in the table:

Repetition #	1	2	3	4	5
"successes" assigned to group A					
as extreme as actual sample?					

(f) Combine your results with the rest of the class, forming a *dotplot* of the number of successes randomly assigned to group A.

(g) If the observer's incentive had no effect on the participant's performance, about how many successes would you expect to see in group A? Explain. Is the center of your simulation distribution close to this expected value?

(h) How many repetitions were performed by the class as a whole? How many of them gave a result as extreme as the actual sample (3 or fewer successes in group A)? What proportion of the repetitions is this?

(i) Remember that your random shuffling and dealing assumed that the observer's incentive had no effect on the participant's performance. Based on these simulated results, does it appear that it is very unlikely for random assignment to produce a result as extreme as the actual sample when the observer has no effect?

(j) In light of your answer to the previous question, considering that the actual sample is what the researchers found, would you say that the data provide reasonably strong evidence in support of the researchers' conjecture? Explain.

(k) If it had turned out that only 1 of the successes had been in group A (and 10 in group B), would that sample result have provided strong evidence in favor of the researchers' conjecture? Explain, based on the result of your simulation.

This activity introduces the important idea that *statistical significance* assesses the likeliness of a sample result by asking how often such an extreme result would occur by chance alone. When the sample result is unlikely to occur by chance, it is said to be statistically significant. The long-run proportion of times that a result as extreme as the sample would occur by chance alone is called the *probability* of such a result. This probability is also called the *p-value* of the test. Our simulation has produced an *empirical approximation* of this probability.

Computer Simulation:

We can use the computer to simulate this process of random assignment much more quickly and efficiently. We will do this with a Java applet that you can access from our course webpage (http://statweb.calpoly.edu/rossman/stat321).

(l) Go to the applet and click Randomize. Report how many successes were randomly assigned to Group A. Then click Randomize and report this number again; did you get the same number of successes in Group A?

(m) Turn off the "Animate" feature and change the number of repetitions to 998 (for a total of 1000). Click Randomize. Write a few sentences describing the distribution.

(n) Report the mean number of successes randomly assigned to group A. Is this value close to what you expected? Explain.

(o) Now click on "Show tallies," and record the distribution in the table:

	0	1	2	3	4	5	6	7	8	9	10	11
tally												

(p) Now click on the "Approx p-value" button. Based on these 1000 simulated randomizations, what is the empirical p-value for this study? How does this compare to the empirical p-value from the class card simulation?

(q) Are the sample data fairly unlikely to occur by chance variation alone if the observer's incentive had no effect? Do the sample data provide reasonably strong evidence in favor of the researchers' conjecture? Explain.

(r) If the study had seen two successes in Group A and 9 in Group B, what would the empirical p-value be? Would this constitute stronger or weaker evidence for the researchers' conjecture? Explain why this makes sense.

(s) Now suppose that the experiment had twice as many subjects and that the proportions of success in both groups had turned out exactly the same as in this study (.250 vs. .667). Would you expect the p-value to be smaller, larger, or the same? Explain your thinking.

(t) Click at the bottom of the applet where it indicates a version that allows you to change table entries. Set the number of repetitions to 500 and uncheck the "Animate" box. Make the change suggested (double the counts in every cell) in the table there and click the Randomize button. Report the empirical p-value. Is it smaller or larger than before? Explain why this makes sense.

The reasoning process of this activity typifies that of statistical tests of significance. One starts by assuming that there is no difference between the two experimental groups and then investigates how often the observed data would occur if nothing more than the random assignment of subjects to groups were involved. If the observed data are quite unlikely to arise due to chance, then the data provide evidence against the assumption of no difference between the groups, thus supporting the hypothesis that the treatment does indeed have an effect. There is no precise rule for determining how "unlikely" the data need to be in order to support the research hypothesis, but the most common standard is a probability of .05.

To this point we have approximated this probability through physical and computer simulations. This approximation generally gets closer and closer to the long-term probability as one increases the number of repetitions in the simulation. Another approach to calculating this probability is to use probability theory and the mathematics of counting techniques.