

Activity: Basic Concepts in Statistics and Probability

Concepts: Introduction to basic concepts of probability and statistics including randomness, probability and empirical probability distributions, the reasoning of statistical tests of significance, simulation as a tool.

Scenario: Friendly Observers I

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

- Calculate the sample proportions of success (beating the threshold) for each group.
- Do these sample proportions differ in the direction conjectured by the researchers?
- 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.
- If the observers had no effect on the students' performance then how would you expect the 11 winners to be distributed among the two groups?

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. Once we know some probability theory and have reviewed counting techniques we will determine this by a method other than simulation. First let's try simulation.

(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) 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?

(h) 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?

(i) 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.

(j) If it the researchers data showed only 1 of the 11 successes had been in group A (and thus 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 value is so special that we give it a name: the *p-value*. Our simulation has produced an *empirical approximation* of this p-value. Once we have reviewed counting and learned some probability theory we will be able to compute the actual p-value for this data.

Computer Simulation:

We could use the computer to simulate this process of random assignment much more quickly and efficiently. The following will step you through how to use Minitab to perform the simulation. In particular you will learn how to write a Minitab macro to repeatedly run a simulation.

First make sure you have a Minitab prompt (i.e. MTB>) in your session window. If you don't have a prompt then go to Edit|Preferences. The Preferences window should appear. Click on Session Window, then click Select. The Session Window Preferences window should appear. Click the Enable button for Command Language and click on the OK button. Now click the Save button in the Preferences window. A prompt should now appear in your session window.

- (k) Use Minitab to simulate one random assignment of the 11 winners and 12 losers to groups A and B by first entering the data:

```
MTB> set c1
DATA> 11(1) 13(0)
DATA> end
```

with 1 representing a winner and 0 denoting a loser. Note that you now have eleven ones and thirteen zeros in column c1. Then take a random sample of 12 of these 23 subjects to form group A:

```
MTB> sample 12 c1 c2
```

Here Minitab has randomly selected 12 of the subjects in column c1 and placed them into column c2 (which represents the subjects assigned to group A). Finally, calculate the number of winners randomly assigned to group A:

```
MTB> sum c2
```

Count the number of winners in column c2. It should be the same as the number you just computed via Minitab. Record how many “winning” subjects were randomly assigned to group A in this repetition.

- (l) Use Minitab to keep track of these results from repetition to repetition by first setting up and initializing a counter variable and then collecting your first result into a new column:

```
MTB> let k1=1
MTB> let c3(k1)=sum(c2)
```

Here k1 counts the number of samples and c3 stores the number of “winning” subjects randomly assigned to group A. You might want to name this column to make sure that it is clearly identified:

```
MTB> name c3 'numAwins'
```

Note: You can also name c3 by clicking on the row above one and typing in the name. Then increment the counter:

```
MTB> let k1=k1+1
```

and take another random sample of 12 of these 23 subjects:

```
MTB> sample 12 c1 c2
```

and again calculate the number of winners randomly assigned to group A and continue to collect them:

```
MTB> let c3(k1)=sum(c2)
```

(m) Use Minitab to repeatedly take samples in this manner by incrementing the counter and then repeating the sampling and summing commands. In other words, you can repeatedly copy and paste the following commands:

```
MTB> let k1=k1+1
```

```
MTB> sample 12 c1 c2
```

```
MTB> let c3(k1)=sum(c2)
```

(n) Do this for a total of 25 repetitions. Then use Minitab to produce a tally of the results :

```
MTB> tally c3
```

and to compute the descriptive statistics for the results (note: you can also use the Stats pull down menu here):

```
MTB> describe c3
```

What are the mean and standard deviation for your empirical distribution?

Now produce a dotplot and histogram of the distribution:

```
MTB> %Dotplot c3 (It may be easier to choose "Dotplot" from the "Graph" menu.)
```

```
MTB> hist c3
```

Show the mean and standard deviation on your distribution histogram.

(o) Is this distribution similar to what you found with the card simulation? In how many of these 25 random assignments were 3 or fewer winners assigned to group A?

While this Minitab simulation is more efficient than shuffling and dealing cards, we can make it run much more quickly by writing a macro (a series of Minitab commands).

(p) Create a Minitab macro that allows for quickly this random assignment for these data a large number of times. To do this first re-initialize the counter and clear the output columns:

```
MTB> let k1=1
```

```
MTB> erase c2 c3
```

Then copy into a text file (open Notepad under "Accessories") the following commands:

```
sample 12 c1 c2
```

```
let c3(k1)=sum(c2)
```

```
let k1=k1+1
```

Save the text file as "friendly.mtb". Be sure to use the ".mtb" extension and put quotes around the file name, and remember where you put it. (Save it to your disk if you have one.) Then within Minitab, select File> Other Files> Run an Exec..., tell it to execute 1000 times, click on "Select file" and choose the "friendly.mtb" file that you just created.

- (q) When the macro has finished running, ask Minitab for a tally, descriptive statistics, and a histogram of the results (remember that the results represent the number of winners randomly assigned to group A under the assumption that the observer's incentive has no effect). Write a few sentences describing the distribution, and record the distribution in the table:

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

- (r) In how many of these 1000 repetitions were the results as extreme as in the researchers' actual data? Based on this larger simulation, what is the approximate probability of obtaining a sample result as extreme as the researchers' actual data (note: this probability is also called the *p-value* of this test)? Are the sample data pretty 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.

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 answer is that 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 to have a *p-value* less than .05.

To this point we have approximated this *p-value* through physical and computer simulations, i.e. we have computed an empirical approximation of this *p-value*. 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.