Activity #3a: Permutation / Randomization Methods

Today we're going to learn how to conduct a statistical hypothesis test using methods that, while widely applicable, are rarely taught to students in a traditional statistics course. We're going to learn how to use randomization (or permutation) tests.

Scenario: Most people are right-handed and even the right eye is dominant for most people. Molecular biologists have suggested that late-stage human embryos tend to turn their heads to the right. In a study reported in Nature (2003), German bio-psychologist Onur Güntürkün conjectured that this tendency to turn to the right manifests itself in other ways as well, so he studied kissing couples to see if they tended to lean their heads to the right while kissing. He and his researchers observed couples in public places such as airports, train stations, beaches, and parks. They were careful not to include couples who were holding objects such as luggage that might have affected which direction they turned. For each couple observed, the researchers noted whether the couple leaned their heads to the right or to the left.

Suppose they found that 8 of 12 (66.7%) kissing couples leaned to the right.

Do these results provide strong evidence that kissing couples lean to the right a majority of the time?

Adapted from Concepts of Statistical Inference: A Randomization Based Curriculum by Rossman, Chance, Cobb, and Holcomb NSF/DUE/CCLI#0633349

1) What are some possible reasons why the researcher would have found 8 of the 12 couples leaned to the right. Is it possible to get these results simply by chance?

2) As we learned the first day of class, the first step in conducting a statistical hypothesis test is to write out two competing hypotheses. In general, the null (dull) hypothesis states that nothing interesting happens. What would the null hypothesis be in this study?

Null hypothesis: The probability of a couple leaning to the right when kissing is ____________________

Alternate hypothesis: The probability of a couple leaning to the right when kissing is ____________________

As we learned in Activity 1, we will work under the assumption that the null hypothesis is true.

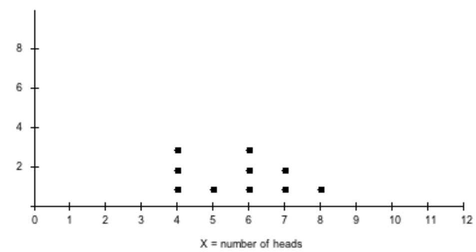
Our null hypothesis states that each of the 12 couples had a 50% chance of leaning to the left or to the right. If this is true, then we would expect to have found 6 of the 12 couples leaning to the right.

We found 8 of the 12 couples leaned to the right -- an "extreme" result under our null hypothesis. Our task is to estimate the probability of observing 8 couples leaning to the right if, in fact, the null hypothesis is true.

3) If the null hypothesis is true -- and the couples just randomly chose to lean either right or left -- then we can simulate this study by flipping a coin 12 times. Arbitrarily, let's say that if the coin comes up heads, it represents a couple leaning to the right. I went ahead and flipped a coin 12 times and found 7 heads. In my mind, that tells me it's certainly possible to have 7 couples (from 12) lean to the right simply by chance.

4) Ok, now I'll admit that I actually didn't slip a coin 12 times. Instead, I went to the following website and used their coin toss simulator: http://statweb.calpoly.edu/bchance/applets/BinomDist3/BinomDist.html

I went ahead and did this 9 more times -- tossing 12 coins each time and recording the number of heads that came up. The following graph displays the number of heads I found in each of the 10 simulations.

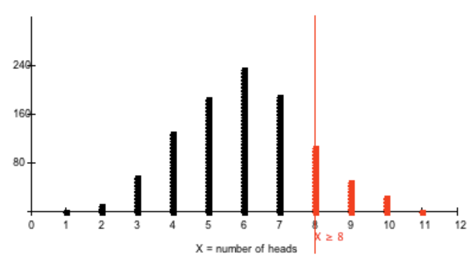


Let's review what we've done thus far. By assuming couples just randomly choose to lean right or left when kissing, we were able to simulate this study by flipping a coin. We conducted 10 simulations of this study and found results that happened just by chance.

Our task, once again, is to estimate the likelihood of observing 8 or more couples leaning to the right (what we actually observed in the study) if our null hypothesis were true.

Based on the graph above, what's our best estimate of this likelihood? We call this our p-value (the probability of observing results as or more extreme than what we actually observed, assuming a true null hypothesis). Since we actually did observe 8 couples leaning to the right, would you be willing to reject our null hypothesis?

5) Rather than just conducting 10 simulations, I went ahead and conducted 1000 simulations. The following graph displays the number of heads I found in each of the 1000 simulations. Highlighted in red are the 183 simulations that resulted in 8 or more heads (out of 12 tosses).



Based on these 1000 simulations, estimate the p-value for this study. What conclusion would you make? Do couples have a tendency to lean to the right when kissing? Would your conclusion change if we had observed 9 of the 12 couples leaning to the right?

6) In reality, the researchers didn't observe only 12 couples. They actually observed 80 couples leaning to the right out of 124 total couples. Use the coin toss simulator to estimate the p-value for this study. What conclusions can you make?

Situation: A company has developed a new drug they believe makes people run faster. To test this claim, they randomly sampled 8 individuals. Four of the subjects were randomly assigned to receive the drug; the other four subjects received a placebo. The researchers then had these eight people run a race. They observed the following results:

Drug Group: Finished in 1st, 2nd, 4th, and 5th
Placebo Group: Finished in 3rd, 6th, 7th, and 8th place

Can we conclude the drug does make people run faster?

7) Why were subjects randomly selected for this study? Why were they randomly assigned to treatment groups? What's the difference between an observational and experimental study?

8) What are some possible reasons why the drug group performed better than the placebo group?

Looking ahead: The places in which the subjects finished the race represent an ordinal level of measurement. The order of the values (1st, 2nd, 3rd place, etc.) have meaning, but the intervals between the places are not necessarily equal.

9) Write out the null and alternate hypotheses. Remember that the null hypothesis states that nothing interesting happens (the treatment did not impact the results).

Null hypothesis:

Alternate hypothesis:

10) Based on the results from the race, did the drug group outperform the placebo group? How did you determine this? Restate the hypotheses in mathematical notation.

Null hypothesis:

Alternate hypothesis:

As we learned in Activity 1, we will work under the assumption that the null hypothesis is true.

The null hypothesis states that the drug had no impact on the results of the study. If this is true, then the performance of the runners would not have changed if they had been randomly assigned to either group.

For example, we would expect the fastest runner to still finish in first place if he had taken the drug or placebo.

Let's pretend that we're able to go back in time and randomly assign the subjects to groups (again). Since we're assigning subjects randomly, we'd expect that some of the subjects would be assigned to different groups. We are still assuming, however, that the group assignments have no impact on their performance.

11) The following tables show 4 different ways subjects could have been randomly assigned to treatment groups (4 randomizations).

Randomization 1	
Drug	Placebo
1	
2	
	3
4	
5	
	6
	7
	8
Sum = 12	Sum = 24

Randomization 2	
Drug	Placebo
	1
2	
3	
4	
	5
	6
	7
8	
Sum = 17	Sum = 19

Randomization 3	
Drug	Placebo
1	
2	
3	
4	
	5
	6
	7
	8
Sum = 10	Sum = 26

Randomization 4	
Drug	Placebo
1	
	2
	3
	4
	5
6	
7	
8	
Sum = 22	Sum = 14

In which of these randomizations would you conclude that the drug group outperformed the placebo group?

12) This study randomly assigned 8 subjects to 2 groups. How many different ways could we randomly assign 8 subjects into 2 groups? Is each one of these randomizations equally likely to occur?

We observed an extreme result in which the drug group outperformed the placebo group.

Since we can calculate the total number of possible randomizations and we know each randomization is equally likely to occur, we can calculate the probability of observing results as (or more) extreme as we did.

In other words, we're assuming the drug has no effect on performance. If that's true, we're going to calculate the likelihood of observing our results (the drug group finishing 1st, 2nd, 4th, and 5th place, or better).

13) To do this, we could list every possible randomization and find those randomizations that yield a sum of 12 or less for the drug group. Obviously, this would be rather time-consuming. An alternate method is to try to list out only the most extreme randomizations that yield a sum of 12 or less for the drug group.

The rows in the following table show possible randomizations. In each row, write 4 X's to indicate which subjects could have been assigned to the drug group. In the final column, calculate the sum of their ranks (places finished). The first two rows are completed.

	1st	2nd	3rd	4th	5th	6th	7th	8th	Sum
Randomization 1	X	X	X	X					10
Randomization 2	X	X	X		X				11
Randomization 3									
Randomization 4									
Randomization 5									
(continue as necessary)									

14) Assuming the drug has no effect on performance, what was the likelihood of observing results as extreme (or more extreme) as the results we did obtain?

15) Would you say our results were likely to have happened if, in fact, the drug does not impact performance? What can you conclude from this study?

Situation: I once missed 11 straight free throws over the course of a few basketball games. Suppose a shooting coach called me and claimed that shooting free throws underhanded is better than shooting them the traditional way. To demonstrate this, this coach randomly selected six subjects from the street. He had 3 subjects shoot 15 free throws underhanded and the other 3 subjects shoot 15 free throws the traditional way.

Here are the number of free throws made by each subject:

Underhand Group: 10 14 8 (total = 32)

Traditional Group: 6 12 9 (total = 27)

From these results, can we conclude that the underhand method is better than the traditional method?

16) State the null and alternate hypotheses for this study. Which hypothesis appears to be correct, based on our observed data?

Null hypothesis:

Alternate hypothesis:

17) Before we run our test and make our conclusion, let's look at the possible decisions we can make. In reality, one of two things can be true: either the underhand method is better or it isn't. We also will make one of two conclusions: we will decide the underhand method is better or it isn't.

The following table displays the possible realities and decisions. Which of the cells are errors? Which are correct decisions? Describe the results from each cell in the table.

		Reality	
		Null hypothesis is true	Null hypothesis is false
Decision	Null hypothesis is true		
	Null hypothesis is false (We reject the null)		

Recall our data: Underhand Group: 10 14 8 (total = 32)
Traditional Group: 6 12 9 (total = 27)

18) If the null hypothesis is true, the underhand method is not any better than the traditional method. Assuming this is true, the subjects would have made the same number of free throws regardless of which group they were assigned to. For example, if the first subject were assigned to the traditional group, he still would have made 10 free throws.

How many different ways could we have randomly assigned six subjects into two treatment groups?

19) Using the following table of randomizations, determine the likelihood of observing results as or more extreme that what we did observe.

	6	8	9	10	12	14	Sum
Rand. 1	X	X	X				23
Rand. 2	X	X		X			24
Rand. 3	X	X			X		26
Rand. 4	X	X				X	28
Rand. 5	X		X	X			25
Rand. 6	X		X		X		27
Rand. 7	X		X			X	29
Rand. 8	X			X	X		28
Rand. 9	X			X		X	30
Rand. 10	X				X	X	32
Rand. 11		X	X	X			27
Rand. 12		X	X		X		29
Rand. 13		X	X			X	31
Rand. 14		X		X	X		30
Rand. 15		X		X		X	32
Rand. 16		X			X	X	34
Rand. 17			X	X	X		31
Rand. 18			X	X		X	33
Rand. 19			X		X	X	35
Rand. 20				X	X	X	36

20) What do you conclude?

Assigned Problems: Choose one of the following problems to complete by ____________________

1) Major League Baseball has faced controversy regarding alleged steroid use by its players. Suppose we randomly test 9 players for steroids at the end of the season. 5 of the players test positive for steroids, while the other 4 test clean. The homerun totals for these players are displayed below:

Tested positive for steroids:	16	32	37	45	58
Tested clean:	0	9	12	46	

To simplify things a bit, let's convert the data to ranks (the lowest homerun total is ranked 1; the highest is ranked 9):

Tested positive for steroids:	4	5	6	7	9
Tested clean:	1	2	3	8	

a) Can you think of any reasons why we would NOT want to convert the data to ranks and then run our analysis?

b) Using the methods we learned in this activity, test to see if players on steroids hit more homeruns than players not on steroids. Write out your hypotheses, estimate the p-value, and write out any conclusions you can make.

2) Four randomly selected subjects were put on a fish-oil diet. Three other individuals remained on their normal diets. After six months, the reduction in their blood pressure levels were as follows:

Fish Oil:	10	5	42	-5
Regular Diet:	-6	12	-20	

To simplify the calculations, let's convert the data to ranks (greatest decrease in blood pressure is ranked #1):

Fish Oil:	5	4	7	3
Regular Diet:	2	6	1	

To simplify the calculations, let's convert the data to ranks (greatest decrease in blood pressure is ranked #1):

a) Can you think of any reasons why we would NOT want to convert the data to ranks and then run our analysis?

b) Using the methods we learned in this activity, test to see if fish oil reduces blood pressure. Write out your hypotheses, estimate the p-value, and write out any conclusions you can make.