Activity 3a: Permutation Test / Fisher's Randomization Test

Today we're going to learn a technique that's usually taught at the end of a student's second or third statistics class. Just like we did in Activity 1, we're going to learn how to conduct a statistical hypothesis test. The tests we learn today are called *permutation tests*.

1) What's the difference between *probability* and *statistics*? What is statistical inference?

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?

2) 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?

3) 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.

4) As we learned in Activity 1, the first step in conducting a statistical test is to write-out two competing hypotheses. In general, the *null hypothesis* states that nothing interesting happens (the treatment did not impact the results). The *alternate hypothesis* states that our treatment had an effect on the results. Write out the hypotheses for this study:

Null Hypothesis:

Alternate Hypothesis:

5) 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.

The following table shows 4 different ways subjects could have been randomly assigned to treatment groups (4 randomizations)

Randomization #1		**Randomization #2**		**Randomization #3**		**Randomization #4**	
Drug	Placebo	Drug	Placebo	Drug	Placebo	Drug	Placebo
1			1	1		1	
2		2		2			2
	3	3		3			3
4		4		4			4
5			5		5		5
	6		6		6	6	
	7		7		7	7	
	8	8			8	8	
Sum = 12	Sum = 24	Sum = 17	Sum = 19	Sum = 10	Sum = 26	Sum = 22	Sum = 14

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

6) 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).

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.

	1	2	3	4	5	6	7	8	Sum
Randomization 1	**X**	**X**	**X**	**X**					**10**
Randomization 2	**X**	**X**	**X**		**X**				**11**
Randomization 3									
Randomization 4									
Randomization 5									
(continue as necessary)									

7) 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?

8) 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?

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

Null hypothesis:

Alternate hypothesis:

10) 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	
		H_0 is true	**H_1 is true**
Decision	**H_0 is true**	Correct Decision	Beta
	H_1 is true	Alpha	Correct Decision Power

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

11) 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?

12) 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
Traditional	X	X	X				23
Traditional	X	X		X			24
Traditional	X	X			X		26
Traditional	X	X				X	28
Traditional	X		X	X			25
Traditional	X		X		X		27
Traditional	X		X			X	29
Traditional	X			X	X		28
Traditional	X			X		X	30
Traditional	X				X	X	32
Traditional		X	X	X			27
Traditional		X	X		X		29
Traditional		X	X			X	31
Traditional		X		X	X		30
Traditional		X		X		X	32
Traditional		X			X	X	34
Traditional			X	X	X		31
Traditional			X	X		X	33
Traditional			X		X	X	35
Traditional				X	X	X	36

13) What do you conclude?

14) We could have also changed our data into ranks and run the same procedure. Can you think of reasons why we might want to do this? Can you think of reasons why we would not want to do this?

Data	Free Throws Made			Total
Underhand	10	14	8	32
Traditional	6	12	9	27

Ranks	Free Throws Made			Total
Underhand	4	6	2	12
Traditional	1	5	3	9

QUESTIONS: DUE _____

1) There is currently a controversy regarding baseball players and alleged steroid use. Suppose we randomly test nine players for steroids. Five players test positive for steroids and four players test clean. The homerun totals for these players are as follows:

	Homeruns				
Steroid	16	32	37	45	58
Clean	0	9	12	46	

Convert this data into ranks and test to see if players on steroids hit more homeruns than players not on steroids.

Ranks	Homeruns				
Steroid	4	5	6	7	9
Clean	1	2	3	8	

Situation: 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:

	Reduction in blood pressure				
Fish Oil	10	5	42	-5	
Regular	-6	12	-20		

Convert this data into ranks and test to see if fish oil reduces blood pressure.

Ranks	Reductino in blood pressure				
Fish Oil					
Regular					