Assignment #3: Applications of Counting Rules; Permutation Tests Name: _____

On this page, you can speed up the calculations by using the hypergeometric calculator:
http://stattrek.com/online-calculator/hypergeometric.aspx

School bus inspector: Suppose you are a safety inspector who must examine 20 school buses in a lot. Unknown to you, 3 of the buses are unsafe

1. If you sample 5 buses, what's the probability that you will find at least one unsafe bus?

2. If you sample 3 buses, what's the probability that you will find at least one unsafe bus?

3. How many buses would you need to sample in order to have a 50% chance of finding all 3 unsafe buses?

Hiring a statistician: Due to the popularity of MATH 300 and MATH 301, SAU needs to hire additional faculty to teach statistics courses. Suppose 20 statisticians apply for these faculty positions.

4. If 5 random applicants are chosen for employment, what's the probability that the 5 best statisticians are selected?

5. If 10 random applicants are chosen for employment, what's the probability that the 5 best statisticians are among those 10 chosen?

6. Suppose we REALLY wanted to hire the 5 best statisticians, no matter how much it cost. If we randomly hire applicants from the pool of 20, how many would we need to hire in order to have a 50% chance of hiring all 5 of the best statisticians?

Randomization/Permutation Tests

Dolphin Therapy (revisited): In assignment #1, we used a simulation approach to analyze data from the following scenario:

Situation: Swimming with dolphins can certainly be fun, but is it also therapeutic for patients suffering from clinical depression? To investigate this possibility, researchers recruited 30 subjects aged 18-65 with a clinical diagnosis of mild to moderate depression. Subjects were required to discontinue use of any antidepressant drugs or psychotherapy four weeks prior to the experiment, and throughout the experiment. These 30 subjects went to an island off the coast of Honduras, where they were randomly assigned to one of two treatment groups. Both groups engaged in the same amount of swimming and snorkeling each day, but one group (the animal care program) did so in the presence of bottlenose dolphins and the other group (outdoor nature program) did not. At the end of two weeks, each subjects' level of depression was evaluated, as it had been at the beginning of the study, and it was determined whether they showed substantial improvement (reducing their level of depression) by the end of the study.

Our research question is: Does the presence of dolphins reduce depression levels?

The following data displays the results of this study:

	Dolphin Group	No dolphin group	Total
Showed improvement	10	3	13
Did not improve	5	12	17
Total	15	15	30

7. Using our simulation methods, we estimated a p-value near 0.0123. Let's calculate this probability directly using counting rules (or the *hypergeometric distribution*, as we're starting to call it).

   Recall that our null hypothesis is an assumption that the dolphin therapy had no effect on depression levels. If that was true, this study simply consisted of:
   - 13 subjects who were going to improve (IMPROVERS) and 17 subjects who were not going to improve
   - The researchers randomly assigning 15 subjects to a dolphin therapy group (that has no effect)

   Let's find the probability of observing 10 or more IMPROVERS if we randomly choose 15 people from this study. We'll have to sum 5 different calculations (the probability of exactly 10 improvers + 11 improvers + 12 + …).

   You can use the following website to speed up the calculations. You already know your answer will be close to 0.0123. http://stattrek.com/online-calculator/hypergeometric.aspx

Nurse Gilbert: For several years in the 1990s, Kristen Gilbert worked as a nurse in the intensive care unit (ICU) of the Veteran's Administration hospital in Northampton, Massachusetts. Over the course of her time there, other nurses came to suspect that she was killing patients by injecting them with the heart stimulant epinephrine. Part of the evidence against Gilbert was a statistical analysis of more than one thousand 8-hour shifts during the time Gilbert worked in the ICU (Cobb and Gelbach, 2005). Here are the data:

	Gilbert Worked Shift	Gilbert did not work	Total
Death occurred on shift	40	34	74
Death did not occur	217	1350	1567
Total	257	1384	1641

8. Were deaths more likely to occur on shifts that Gilbert was working than on shifts when she was not? Show the proportions you used to answer this question.

9. Suppose you're an attorney defending Gilbert. Could you use the "random chance" argument for her defense? Is it possible that deaths are no more likely on her shifts; that it was just the "luck of the draw" that resulted in such a higher percentage of deaths on her shifts?

10. Let's first try the simulation-based approach using this website:
    http://www.rossmanchance.com/applets/ChiSqShuffle.html?dolphins=1

    Fill-in the 2x2 table at the top of the page and select STATISTIC: CELL 1 COUNT near the bottom-left. Then, check the "show shuffle options box" at the top-right and simulate 10,000 shuffles to see how many deaths would occur on Nurse Gilbert's shift just by random chance. Finally, count samples greater than 40 to find the likelihood of observing so many deaths by chance. Write that p-value below:

    P-value: _____

11. Describe how you could use the hypergeometric distribution to calculate the p-value without a simulation. Set-up the first couple formulas you would use in this calculation.

Fish Oil: Four randomly selected subjects were put on a fish-oil diet; 3 other subjects remained on their normal diets. The following table shows the reduction in blood pressure for each subject 6 months after this study began:

Fish Oil	Regular Diet
-5 (an increase of 5)	-20
5	-6
10	12
42	

To simplify our calculations, we can convert this data into ranks (the lowest score is ranked 1, the next-lowest is ranked 2, and so on).

Fish Oil	Regular Diet
3	1
4	2
5	6
7	

12. As we did in class, let's calculate the sum of these scores (the ranks) to serve as our test statistic. Sum these ranks for each group.

    Sum of ranks for fish oil group = _____ Sum of ranks for regular diet group = _____

13. Write out a null hypothesis for this study. How large of a difference do we assume the fish oil will make?

    Null hypothesis: _____

14. Suppose we go back in time and then, again, randomly assign 4 subjects to the fish oil diet and 3 to the regular diet. In how many ways could we do this (with the 7 subjects in this study)?

15. List out those randomizations (or at least the most extreme randomizations) to determine the likelihood of obtaining data *as or more extreme than what we observed in this study*.

16. What conclusions can we make from this study?