**Unit 1 GROUP ASSIGNMENT: Simulation Methods**

Source: Allan Rossman: http://statweb.calpoly.edu/arossman/stat325/notes.html

**Situation #1: Contagious yawns?**

Are yawns contagious? Conventional wisdom says yes: when you see someone else yawn, you're prone to feel sleepy and yawn yourself. How many times have you caught yourself in this situation, or noticed it in someone else? But will this hypothesis withstand a scientific test? Will data support this claim?

The folks at *MythBusters* investigated this issue by using a two-way mirror and a hidden camera. 50 subjects sat in a booth, accompanied only by an experimental attendee. For some of the subjects, the attendee yawned (planting a yawn "seed"), while for other subjects the attendee did not yawn. The researchers decided in advance, with a random mechanism, which subjects would receive the yawn seed and which would not. As time passed, the researchers watched to see which subjects yawned. They found that 10 of 34 subjects who had been given a yawn seed actually yawned themselves, compared with 4 of 16 subjects who had not been given a yawn seed. These data are summarized in the following 2×2 table:

| | Yawn seed planted | Yawn seed not planted | Total |
|---|---|---|---|
| Subject yawned | 10 | 4 | 14 |
| Subject did not yawn | 24 | 12 | 36 |
| Total | 34 | 16 | 50 |

Do the data appear to support the claim that yawns are contagious? We'll investigate this in this activity.

1) At first glance, it might appear that yawns are contagious because 10 people yawned in the "seed" group and only 4 yawned in the "no seed" group. Upon closer inspection, though, we see that the seed group had 34 subjects while the no seed group had only 16 subjects. Rather than comparing the number of yawns in each group, let's compare the proportion of subjects who yawned in each group. Calculate these proportions.

   Proportion of subjects in the seed group who yawned = _____

   Proportion of subjects in the no seed group who yawned = _____

These proportions support the idea that yawning is contagious, but is it possible that this difference in proportions could have happened even if yawning weren't contagious? It is possible that this difference is simply due to the random nature of putting subjects into groups (we just happened to assign people who tend to yawn to the seed group).

Look at the total number of subjects (14) who yawned in the study. It's possible that these 14 subjects were going to yawn regardless of which group they were assigned to. But, if this is true, how likely was it that 10 of those 14 yawners were randomly placed into the seed group? How likely were we to find 10 or more yawners in the seed group?

If the answer to that question is that it was unlikely to have randomly assigned 10 of 14 yawners to the seed group, then we would have evidence supporting the conclusion that yawns are contagious. Why? Because otherwise, we would have to believe that a rare event just happened to occur in this experiment.

So the key question now is how do we determine the likelihood of observing 10 of 14 yawners in the seed group if we assume that yawns are not contagious (and the 14 yawners were just randomly assigned to the two groups)? We'll call this assumption the *null model*.

We'll use simulation methods in an attempt to find this likelihood. We'll pretend that we can go back in time and, once again, randomly assign 14 yawners and 36 non-yawners into two groups. We'll assign 34 of those 50 subjects into the seed group and the remaining 34 subjects to the no-seed group.

Take 50 small pieces of paper to represent our 50 subjects. We're going to assume that 14 of the subjects are yawners, so write "YAWN" on 14 of those pieces of paper. The remaining 36 blank pieces of paper will represent our non-yawners. Now shuffle

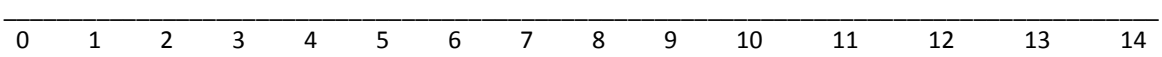
your pieces of paper and blindly choose 34 of them. These 34 pieces of paper represent the 34 subjects who were randomly assigned to the seed group. The remaining 16 pieces of paper represent the 16 subjects assigned to the no-seed group.

So go ahead and do this. Select 34 pieces of paper at random and count the number of times you chose a YAWN subject. Record that number in the first blank cell of the table below. Repeat this process 9 more times – put all 50 pieces of paper in a pile, blindly choose 34 of them, and count the number of YAWNS you chose. Use the table to record the number of YAWNS you chose in each of the 10 replications.

| Randomization | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| # of yawners | | | | | | | | | | |

2) Now share your results with the other members of your group. Let's create a simple visualization of everyone's results. Use the axis below to create a dotplot of the number of yawners randomly assigned to the seed group. So, for example, if you found 3, 7, 12, 11, 8, 1, 0, 8, 4, and 5 YAWNs in your 10 randomizations, you would put a dot on the line above 3, 7, 12, 11, etc. on the axis below. Since you observed 8 YAWNs in 2 randomizations, you would have two dots stacked vertically above the 8.

   Complete the dotplot and record the total number of randomizations included in your table.




   Total number of randomizations (10 x the number of people in your group) = _____

3) At this point, we've randomized and repeated our experiment (through simulation). Now we need to decide if we can reject the null model (that yawning is not contagious). Based on the dotplot you just completed, in how many randomizations did your group observe 10 or more YAWNs? Divide this number by the total number of randomizations from your group to get the proportion of randomizations with 10 or more YAWNS.

   Proportion of randomizations with 10+ YAWNs = _____

4) That proportion represents the likelihood, based on your simulation, that we would have observed 10 or more YAWNs in the seed group if yawning were not contagious. So, can we conclude yawning is contagious? Briefly explain your answer.

   _____

   _____

   _____

5) Let's repeat this process using an applet. Go to http://www.rossmanchance.com/applets/Yawning/Yawning.html to work with the applet. At the top of the applet, you'll see the actual data from this yawning study. You'll see a blank 2x2 table below this. The simulated results will appear in this blank table.

   Uncheck the "animate" box (to speed things up) and click **Randomize**. This will run one randomization of the study, selecting 34 cards and counting the number of YAWNs (displayed in green). It will also put a green dot on the graph representing the results from that randomization. Click **Randomize** again and you will see a second randomization added to the graph.

   Now click **Reset** to reset all the results. Go ahead and change the **number of randomizations** to **1000** and click **randomize**. This runs 1000 randomizations.

   Look at the resulting dotplot. In approximately what proportion of the 1000 randomizations were the results as (or more) extreme than the results from the actual study (10 or more yawners in the seed group)? How does this compare to the results you obtained in your simulation? You do not need to write out your answers to these questions.

6) A mathematical analysis reveals that if we had conducted an infinite number of randomizations, approximately 51.3% of them would have contained 10 or more yawners in the seed group. How does this compare to your simulation results?

   This proportion is called a p-value. It represents the probability of observing data as (or more) extreme than what we actually observed in our study under the assumption that yawning is not contagious.

   From your simulation, the applet, and the mathematical analysis, you should have reached the conclusion that the data observed in the study (10 yawners) were not at all surprising under the null model (assuming yawning isn't contagious). If yawning isn't contagious, it's quite plausible that the random assignment process alone could have produced 10 yawners in the seed group and only 4 in the no-seed group. So we have no reason to reject out null model. In other words, the data do not provide convincing evidence to believe that yawning is contagious (that the results weren't just due to the luck of the draw).

7) Would you say we found a statistically significant difference in the proportion of yawners between our seed and no-seed groups?

   Answer: _____

**Extra Credit -- Situation #2: Dr. Flipper**

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 (Antonioli and Reveley, 2005).

8) Were subjects in this study randomly selected from a population? Were subjects randomly assigned to groups?

Random selection? _____ Random assignment? _____

The researchers found that 10 of 15 subjects in the dolphin therapy group showed improvement, compared to 3 of 15 subjects in the control group. Organize these results in the following 2x2 table:

| | Dolphin Therapy | Control Group | Total |
|---|---|---|---|
| Showed improvement | | | |
| Did not show improvement | | | |
| Total | | | 30 |

9) Calculate the *conditional proportions* of subjects who improved in each group. That is, what proportion of subjects in each group showed improvement?

Proportion in the dolphin group showing improvement = _____

Proportion in the control group showing improvement = _____

10) Let's assume a null model is true. That is, let's assume the dolphin therapy had no impact on improvement (in other words, we have 13 subjects who would have shown improvement regardless of which group they were assigned to). If this assumption is true, how likely were we to have observed 10 or more subjects in the dolphin therapy group showing improvement? Briefly describe the process you used to estimate this likelihood. Based on your estimation, do you think dolphin therapy is beneficial? Did you find a statistically significant difference between the dolphin therapy and control groups?

**Extra Credit -- Situation #3: 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 working on shift | Gilbert not working on shift | Total |
|---|---|---|---|
| Death occurred on shift | 40 | 34 | 74 |
| Death did not occur on shift | 217 | 1350 | 1567 |
| Total | 257 | 1384 | 1641 |

11) 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.

12) 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?

In an attempt to answer this question, try the applet at: http://www.rossmanchance.com/applets/Friendly1/friendly1.html

Fill-in the 2x2 table at the top of the page, uncheck the animate box, and try 1000 replications.

Based on the resulting graph, write out your conclusions.

Use http://statweb.calpoly.edu/bchance/stat217/labs/lab7/Randomization/Randomization.html in the future.