ORCHESTRATING SIMULATIONS FOR COMMON CORE STATISTICS

DOUG TYSON AND JASON MOLESKY

CMC–S 2015

PALM SPRINGS, CA

Making Inferences & Justifying Conclusions (S-IC)

CCSS.MATH.CONTENT.HSS.IC.B.5
Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

EXPERIMENT 1 – SMELLING PARKINSON'S DISEASE   (RANDOMIZATION TEST FOR A COUNT/PROPORTION)

EXPERIMENT 2 – BONUS OR REBATE?   (RANDOMIZATION TEST FOR A DIFFERENCE IN MEANS)

EXPERIMENT 3 – WARMING UP BODY AND MIND   (RANDOMIZATION TEST FOR A DIFFERENCE IN MEANS)

EXPERIMENT 4 – EFFECTIVE EYE DROPS?   (RANDOMIZATION TEST FOR A DIFFERENCE IN PROPORTIONS)

COPIES OF THIS HANDOUT CAN BE FOUND AT MRTYSONSTATS.COM

INTRODUCTION

As reported by the Washington Post (http://tinyurl.com/SmellPark), Joy Milne of Perth, UK, smelled a "subtle musky odor" on her husband Les that she had never smelled before. At first, Joy thought maybe it was just from the sweat after long hours of work. But when Les was diagnosed with Parkinson's 6 years later, Joy suspected the odor might be a result of the disease.

Scientists were intrigued by Joy's claim and designed an experiment to test her ability to "smell Parkinson's." Joy was presented with 12 different shirts, each worn by a different person, some of whom had Parkinson's and some of whom did not. The shirts were given to Joy in a random order and she had to decide whether each shirt was worn by a Parkinson's patient or not.

1. Why would it be important to know that someone can smell Parkinson's disease?

2. How many correct decisions (out of 12) would you expect Joy make if she couldn't really smell Parkinson's and was just guessing?

3. How many correct decisions (out of 12) would it take to *convince* you that Joy really could smell Parkinson's?

SIMULATING THE EXPERIMENT

Although the researchers wanted to believe Joy, there was a chance that she may not really be able to tell Parkinson's by smell. It's logical to be skeptical of claims that are very different than our experiences. If Joy couldn't really distinguish Parkinson's by smell, then she would just have been guessing which shirt was which. The researchers were not willing to commit time and resources to a larger investigation unless they could be convinced to that Joy's wasn't just guessing. When researchers have a claim that they suspect (or hope) to find evidence against, it's called the **null hypothesis**.
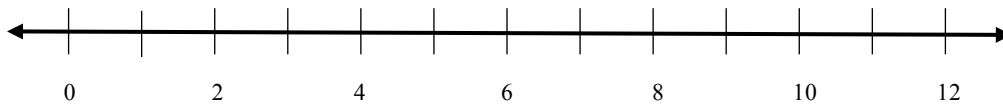
4. What claim were the researchers hoping to find evidence *against*? That is, what was their prior belief (**null hypothesis**) about the ability to smell Parkinson's?

5. What claim were the researchers hoping to find evidence *for*? This is called the **alternative hypothesis** or the **research hypothesis**.

To investigate the idea that Joy was just guessing which shirt was worn by which type of person, we will begin by assuming that the null hypothesis is true.

6. Your instructor will hand you 12 cards (shirts) that have been shuffled into a random order. Don't turn them over yet! On the back of some of them is "Parkinson's" and on the back of others is "No Parkinson's." For each card, guess Parkinson's or No Parkinson's. Once you have made your guess, turn the card over and see if you were correct. Repeat this for each card and record the number of correct identifications (out of 12) below.

Tally of correct identifications	Number of correct identifications	Proportion of correct identifications

7. Create a dotplot of the number of correct identifications with the rest of the class. Record the results below.

8. In the actual experiment, Joy identified 11 of the 12 shirts correctly. Based on the very small-scale simulation by you and your classmates, what proportion of the simulations resulted in 11 or more shirts correctly identified, assuming that the person was guessing?

9. The proportion you just calculated is a crude estimate of a true probability called a **p-value**. How might we improve our estimate of the true probability?

STATISTICAL INFERENCE FROM THE SIMULATION

10. Use the SPA Applet for One Categorical Variable at https://tinyurl.com/SPAapplets to run this simulation 10000 times. Then use that simulation to get a (likely) better estimate of the *p*-value for 11 or more shirts correctly identified, assuming that this person was just guessing. Is it *possible* that Joy correctly identified 11 shirts just by random chance (guessing)? Is it *likely*?

11. An interesting side note is that Joy's one "mistake" really wasn't a mistake. The shirt was worn by a person who supposedly didn't have Parkinson' even though Joy claimed that she could smell the telltale smell on that shirt. That person called the experimenters 8 months after the experiment and reported that he had just been diagnosed with Parkinson's disease. That meant that Joy correctly identified 12 out of 12 shirts. What is the approximate *p*-value for 12 shirts correctly identified, assuming that this person was just guessing?

Note: A small p-value is considered strong evidence against the null hypothesis and in favor of the alternative hypothesis. But how small is small? As a rule of thumb, statisticians generally agree that *p*-values below 0.05 provide pretty strong evidence against the null hypothesis. Observed results with small *p*-values are said to be **statistically significant**.

DEEPER MATHEMATICAL CONNECTIONS

12. The true theoretical probability to get *k* successes in *n* trials when there is a true probability *p* of a success on each trial is given by the **binomial probability formula**: $\binom{n}{k} p^k (1-p)^{n-k}$. Compute the exact theoretical probability to get 11 or more successes in 12 trials when the true probability of success is 0.5. (*Hint*: calculate the probability for 11 successes and then do another calculation for 12 successes and then add these together.)

BONUS OR REBATE?

INTRODUCTION

Are people more likely to spend money if it is called a bonus or if it is called a rebate? Researchers Nicholas Epley (University of Chicago), Dennis Mak (Harvard), and Lorraine Idson (Harvard) investigated this question with a series of statistical investigations. One of those investigations was an experiment with 47 student volunteers from Harvard. Each student was called to a laboratory and was given $50 with no strings attached. However, 22 of the 47 students were randomly assigned to be told "you are receiving this *tuition rebate* because our lab has a surplus of funds," that "we will contact you in one week to ask you some questions about your *tuition rebate*," and that they should ask the experimenter "if they have any questions about this *tuition rebate*." The other 25 students were given identical instructions, except that the words *tuition rebate* were replaced with *bonus income*. After one week, the students were asked to report how much of the $50 they spent. Many thanks to Nick Epley for providing these data!

Amount spent in dollars (Bonus income)				
0	0	0	0	0
0	0	0	0	10
10	10	20	25	26
30	30	40	50	50
50	50	50	50	50

Mean amount spent: $22.04

Amount spent in dollars (Tuition Rebate)				
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	10	20	30	50
50	50			

Mean amount spent: $9.55

1. Calculate the observed difference in the mean amount spent between these two groups (bonus – rebate).

One way to think of the null hypothesis in this case is to imagine that a student was going to spend the amount she spent no matter if she was told it was a rebate or a bonus. If the bonus/rebate wording made no difference in the amount a student would spend, then the difference in the mean amounts (bonus – rebate) would be the result of the random shuffling of subjects into their groups.

2. Is it *possible* that a difference in means of $12.49 or more is just the result of the random shuffling of subjects into their groups?

3. Is it *likely* that a difference in means of $12.49 or more is just the result of the random shuffling of subjects into their groups? Actually, that's not a fair question. Why can't you really answer this question yet?

SIMULATING THE EXPERIMENT

4. Your instructor will give you 47 cards, one for each student. On one side of each card is the dollar amount the student spent. Let's investigate whether the chance of the random assignment is likely to produce a difference in means of $12.49 or more extreme assuming that

   _____.

   Shuffle your cards and deal them into two piles: 25 cards for the bonus group and 22 cards for the rebate group. Record the difference in means (bonus – rebate).

5. Combine your difference in means with that of your classmates to produce a dotplot of your results. Based on this small-scale simulation, does a difference in means as big or bigger than $12.49 seem likely to happen just by chance? Justify your answer.

6. How could you get a better idea of the probability that random chance would produce a difference in means as big or bigger than the observed difference?

STATISTICAL INFERENCE FROM THE SIMULATION

Go to the StatKey website (http://lock5stat.com/statkey/) and use the Test for a Difference in Means applet to run a simulation of the difference in mean amount spent. Use at least 10000 repetitions of the random assignment.

7. Does random chance seem like a reasonable explanation for a difference in means of $12.49 or greater? Support your answer with an approximate *p*-value.

8. Do you have strong evidence that the wording (bonus/rebate) *causes* a difference in the amount students will spend?

9. To what population would you feel comfortable generalizing this result? Why?

INTRODUCTION

Mr. Tyson's former students Natasha and John, both student athletes, wondered if physical activity was beneficial to academic performance. To test their hypothesis, they randomly assigned 58 volunteers to one of two treatments: sitting at a desk for 60 seconds or running stairs for 60 seconds. After the assigned treatment, each participant took an arithmetic test consisting of 100 questions over 60 seconds.

1. Why were participants randomly assigned to groups before being given their treatments?

2. What was the null hypothesis for this experiment? (What did Natasha and John hope to find evidence against?)

3. If this null hypothesis was true, what would you expect the difference in the mean number of correct answers to be (desk – stairs)?

4. The data are shown below along with the means for each group. Compute the difference in the mean number correct (desk – stairs).

Number correct (desk)				
8	13	16	18	18
19	19	19	22	23
24	25	25	25	27
29	29	30	30	30
31	33	34	34	35
35	36	37	44	44
51				

Number correct (stairs)				
8	11	12	15	18
18	21	22	23	23
25	28	30	32	32
34	36	37	37	37
38	38	39	40	42
53	59			

Mean number correct: 27.839   Mean number correct: 29.926

5. What alternative hypothesis were Natasha and John hoping to find evidence for?

6. Recall that the observed difference in the mean number of correct answers (desk – stairs) is –2.087. Does this value give *some* evidence that exercise is beneficial to academic performance? Explain.

SIMULATING THE EXPERIMENT

Even if physical activity made no difference in academic performance (the null hypothesis) was true, we wouldn't expect random chance to perfectly balance out the students in the two treatment groups. We would expect them to be close to balanced, though. To determine if the difference in the mean number of correct answers (desk – stairs) is more unusual than we might expect due to chance, let's simulate many different random assignments.

7. Assume that the treatment (desk or stairs) does not make a difference in the number of correct answers a student will give. How could we use cards to simulate the difference in means (desk – stairs) that we would expect to see when subjects are randomly assigned to treatment groups? (We won't be doing this simulation.)

8. Use the SPA Applet for One Quantitative Variable at https://tinyurl.com/SPAapplets to run a simulation of at least 10000 repetitions. What is the (approximate) probability to see a difference in means as low or lower than the one that the researchers observed, just by chance?

STATISTICAL INFERENCE FROM THE SIMULATION

9. Do the researchers have convincing evidence against the null hypothesis and for the alternative hypothesis? Explain.

INTRODUCTION

Mr. Tyson was prescribed the antibiotic drug AzaSite (azithromycin ophthalmic solution 1%) for his eyes. He puts one drop in each eye before going to be each night to help heal some scar tissue in his eyes. On the pamphlet in the AzaSite box, he found details of the clinical trial for this drug. Patients suffering conjunctivitis were randomly assigned to receive either the AzaSite drops or placebo drops that didn't contain the active medication. After one week, 82 of the 130 patients treated with the AzaSite had the conjunctivitis completely cleared compared to 74 of the 149 patients that saw the conjunctivitis cleared.

1. Organize the given counts into the following two-way table.

Type of drops

Final status		AzaSite	Placebo	Total
	Cleared			
	Not cleared			
	Total			

2. Calculate the **conditional proportion** of the AzaSite group that had their conjunctivitis cleared. Do likewise for the placebo group.

3. Calculate the difference in these proportions (AzaSite – placebo).

4. Is there *some* evidence that the AzaSite is more effective than the placebo at clearing conjunctivitis? Why or why not?

5. Is there *convincing* evidence that the AzaSite is more effective than the placebo at clearing conjunctivitis? (Why can't you really answer this question – at least for now?)

6. State the null and alternative hypotheses for this experiment.

SIMULATING THE EXPERIMENT

One possible explanation for the observed difference in proportions is that the 156 patients whose conjunctivitis was cleared would have had the same result no matter which type of drops they received.

7. Explain in detail how you could use cards to simulate the difference in proportions that would result from random assignment under this null hypothesis mentioned above.

8. Use the Randomization Test for a Difference in Proportions applet on the StatKey website (http://lock5stat.com/statkey/) to run a simulation to estimate the *p*-value for an observed difference in proportions as large or larger than 0.134, assuming that the null hypothesis is true.

STATISTICAL INFERENCE FROM THE SIMULATION

9. Based on the *p*-value from your simulation, would you say that it is reasonable to believe that a difference of 0.134 or larger would occur *just by the chance of the random assignment*?

10. What other reason might there be for such a large difference in proportions between these two groups?

11. Is a cause-and-effect conclusion justified in this situation? Explain.