

SUGGESTED SOLUTIONS TO ORCHESTRATING SIMULATIONS FOR COMMON CORE STATISTICS

DOUG TYSON AND JASON MOLESKY

CMC-S 2015

PALM SPRINGS, CA

EXPERIMENT 1 – SMELLING PARKINSON’S DISEASE

EXPERIMENT 2 – BONUS OR REBATE?

EXPERIMENT 3 – WARMING UP BODY AND MIND

EXPERIMENT 4 – EFFECTIVE EYE DROPS?

(RANDOMIZATION TEST FOR A COUNT/PROPORTION)

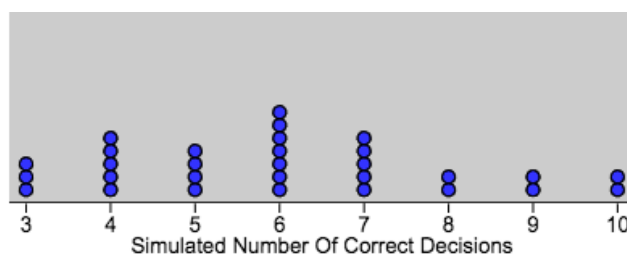
(RANDOMIZATION TEST FOR A DIFFERENCE IN MEANS)

(RANDOMIZATION TEST FOR A DIFFERENCE IN MEANS)

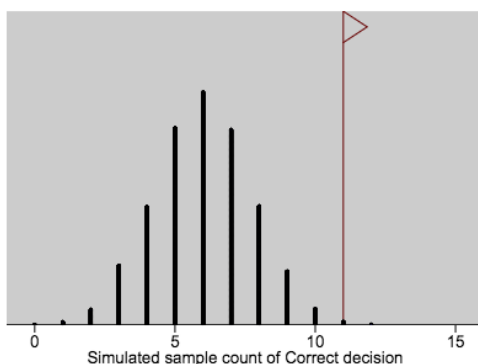
(RANDOMIZATION TEST FOR A DIFFERENCE IN PROPORTIONS)

SMELLING PARKINSON'S DISEASE

- Answers will vary, but one possible answer is that with a particular smell, there may be a chemical compound that is secreted by the body and such a compound may provide a means of early detection for Parkinson's.
- Since she has a 50% chance to guess correctly each time, we would expect 0.50 (12) or 6 correct decisions.
- Answers will vary, but students should only give "high" numbers (numbers they consider much "higher" than 12).
- The null hypothesis is that Joy can't smell Parkinson's disease.
- The alternative hypothesis is that Joy can smell Parkinson's disease.
- Answers will vary.
- Answers will vary, but one possible dotplot would look like the one below. The label on the axis should be "Simulated Number of Correct Decisions."



- Answers will vary, but the correct answer is the number times a simulation resulted in 11 or 12 correct decisions divided by the total number of simulations. In other words, it's the proportion of "dots" that are 11s or 12s. Note that this may be 0! Remember, this is an estimated probability from a simulation. 11 or more correct decisions is not impossible, but it is very unlikely. So, while the exact probability won't truly be 0, it's very close to it.
- We can improve this estimate by running a lot more repetitions/trials of this simulation.
- This simulation results should look similar to the following dotplot. It is definitely possible that Joy is guessing, but not very likely. The applet allows us to calculate the approximate p -

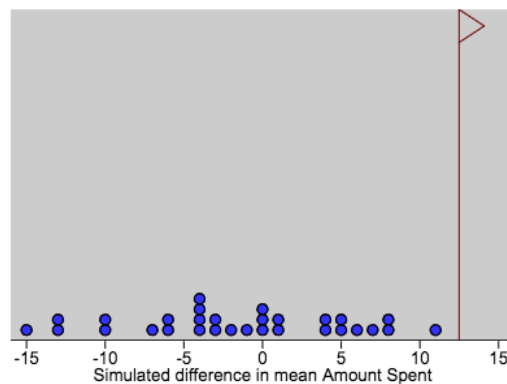


value as 0.0033 in the simulation shown below (33 dots at 11 or 12 out of 10000 repetitions).

- According to the simulation above, the approximate p -value is 3/10000 or 0.0003.
- $P(\text{exactly 11 successes}) \approx 0.0029$. $P(\text{exactly 12 successes}) \approx 0.0002$. Therefore, $P(11 \text{ or more successes in 12 trials}) \approx 0.0031$.

BONUS OR REBATE?

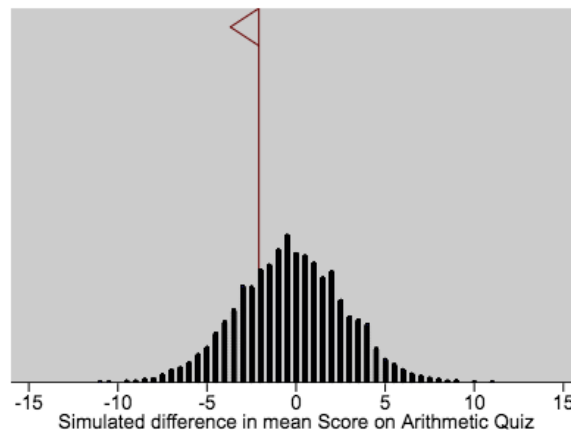
1. $\$22.04 - \$9.55 = \$12.49$
2. Yes, it's possible!
3. I can't answer that question because I don't know what kind of differences in means to expect from the random shuffling of subjects into groups. I can't tell if \$12.49 would be a surprising result from the chance of the random assignment.
4. Fill in the blank with "assuming that the bonus/rebate wording makes no difference in the amount a student would spend."
Differences in means will vary.
5. Dotplots will vary, but one possible dotplot is shown to the right. A difference in means this big or bigger doesn't seem likely to happen by chance because it didn't even occur in 30 trials.



6. I could run this simulation many, many more times. (But I don't want to keep shuffling cards!)
7. Simulated p -values will vary. One simulation of 10000 trials gave an approximate p -value of $198/10000 \approx 0.0198$.
8. Yes, I do have evidence that the bonus/rebate wording causes a difference in the amount students will spend. The groups spent very different amounts – a difference that is very unlikely to be due to the random chance. Thus, I think there is another reason for the extreme difference in means. The random assignment should have roughly balanced out all other variables between the groups, so the only systematic difference between the groups is the bonus/rebate wording. The evidence suggests that this wording is the cause of the difference in the amount spent.
9. These students were volunteers. Statistically speaking, I should only generalize these results to the population of people like those in the study, but it's unclear exactly who those people would be because the students were not a random sample of any particular population.

WARMING UP BODY AND MIND

1. Random assignment tends to roughly balance out the groups. General academic ability, athletic ability, sleep habits, health habits, etc should be similar for the two groups.
2. The null hypothesis is that exercise makes no difference in academic performance on an arithmetic test.
3. Well, since the group should be pretty similar, and if exercise makes no difference in academic performance, then I'd expect the difference in the mean number of correct answers to be near 0.
4. $27.839 - 29.926 = -2.087$. (Note that negative values actually provide evidence in favor of the alternative in this case.)
5. The alternative hypothesis is that exercise improves academic performance on an arithmetic test.
6. Yes, this value does give *some* evidence that exercise improves academic performance. The difference in means is negative, which tells me that the students who sat in desks scored lower on average than the students who ran stairs.
7. I would take 58 cards and write the score on the arithmetic test for each person on one of the cards. Then, I'd shuffle the cards well to mimic the random assignment. I'd deal out 31 cards into one pile to represent the students who sat at a desk for 60 seconds. The other 27 cards would go into a second pile to represent the students who ran stairs for 60 seconds. Then I'd find the mean score of each group and calculate the difference in means (desk – stairs). I'd put this difference on a dotplot. Then I'd repeat this shuffling, dealing, calculating, and plotting for many, many trials/repetitions.
8. Answers will vary. For the simulation shown below consisting of 10000 trials, the approximate probability (p -value) is $2432/10000 \approx 0.2432$.



9. In this case, the researchers do not have *convincing* evidence against the null hypothesis. It's not all that uncommon for random chance to produce a difference in means as low or lower than -2.087 if exercise had no effect on academic performance. To have convincing evidence, I would like to see a very low p -value (lower than 0.05 is the standard).

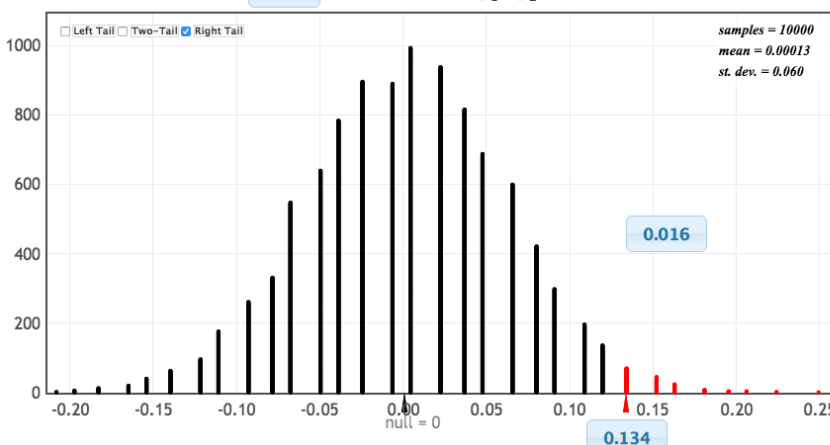
EFFECTIVE EYE DROPS

- Here is the completed table:

		Type of drops		
		AzaSite	Placebo	Total
Final status	Cleared	82	74	156
	Not cleared	48	75	123
	Total	130	149	279

- Proportion of Azasite group cleared = $82/130 \approx 0.631$
Proportion of placebo group cleared = $74/149 \approx 0.497$
- Difference in proportions (Azasite – placebo) = 0.134
- Yes there is some evidence. The difference in proportions is 0.134, which means that the Azasite's cure rate was 13.4 percentage points higher than the placebo's cure rate.
- To answer that question, I need to know if random chance would be likely to produce a difference in proportions as big or bigger than 0.134.
- Null hypothesis: There is no difference in the effectiveness of the Azasite and the placebo in clearing up conjunctivitis.
Alternative hypothesis: Azasite is more effective than the placebo in clearing up conjunctivitis. (Note: it is in fact cheating to form the alternative hypothesis after peeking at the data. One presumes that the company that was testing Azasite believed it to be better than a placebo treatment before they ever ran the experiment. Otherwise, why would they waste their time and money?)
- Take 279 cards and write “cleared” on each of 156 cards to represent the patients that would have been cured no matter which treatment they received (we’re assuming Azasite and placebo are equally effective for the simulation). Write “not cleared” on each of the remaining 123 cards. Shuffle the cards well and deal them into two piles: one pile of 130 cards for the Azasite group and one pile of 149 cards for the placebo group. Calculate the proportion of each group that was “cleared” and find the difference in proportions (Azasite – placebo). Put a dot on a dotplot for the difference in proportions. Repeat the shuffling, dealing, calculating, and plotting for many, many trials. In the end, calculate the fraction of the trials that resulted in a difference in proportions of 0.134 or more.
- Here is one simulation of 10000 trials. The approximate p -value is 0.016.

Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$ Null Hypothesis: $p_1 = p_2$



Original Sample

Group	Count	Sample Size	Proportion
Group 1	82	130	0.631
Group 2	74	149	0.497
Group 1-Group 2	8	n/a	0.134

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	77	130	0.592
Group 2	79	149	0.530
Group 1-Group 2	-2	n/a	0.062

9. Assuming that the two treatments are equally effective, I would not say that a difference in proportions of 0.134 is a reasonable occurrence just due to random chance because it would only occur about 1.6% of the time.
10. Since random assignment should have roughly balanced out the two groups before the experiment began, the only other reasonable explanation is that the Azasite is more effective at clearing up conjunctivitis.
11. A cause-and-effect conclusion is justified because this was a controlled experiment that implemented random assignment.