

Practice Problem for Data Science Exam #2

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

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

The data are shown below along with the means for each group.

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				

Mean number correct: 27.839

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: 29.926

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

4. 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.

5. 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.)

6. The following is output from the JMP Randomization Test Add-In. There are three graphs. Note that the lower right graph appeared when we clicked on the highlighted black dot on the left graph (near the value $x=5$). Use the output to answer the following questions:
 - a. What does the graph on the left show? (ignore the red area for now)
 - b. What does the graph on the upper right show?
 - c. What does the graph on the lower right show?
 - d. What is the p-value for the hypothesis test? Interpret this value.
 - e. Do the researchers have convincing evidence against the null hypothesis and for the alternative hypothesis? Explain.

