Homework 8

3.1

a) For the difference in mean times we calculate as follows:

(8 + 10 + 15)/3 – (5 + 9)/2 = 4

b)

c) In order to find the proportion of differences that are as large or larger than our observed we need to resample and our data using R. Code for this specific problem can be found in the appendix to this homework, but the answered I received for this was .2085. Thus, approximately 20% of our simulated values found through our permutation test were as large or larger than our observed. Since the probability that our simulated values would be as large or larger than our observed is 20%, we would assume that our data is reflective of the null distribution, thus there is no difference between the means.

d) For this we again have to simulate, code is in the appendix. I received a value of .2921.

3.3

b) I set this problem up by calculating the mean of the delays in the month of May less the mean of the delays in the month of June. I received an observed value of approximately -5.667. Based on this, we could (but shouldn’t) assume that the mean of delays in the month of June are greater than the mean of the delays in the month of May. Let’s test this to see if this is actually true or only occurring by chance!

After conducting a two-sided permutation test we receive a P-value of approximately .0002. This means that there is a .002% chance that the difference between the means of the delays happens by random chance/occurrence. Thus, we would reject the null, and confidently stake a claim that there is a difference between the two, and that our data is not under the null distribution. Specifically, that the mean delay times in June are greater than the mean delay times in May.

3.4

b) I set this problem up by taking the variance of the United delays less the variance of the American delays. I received an observed difference of approximately 431. After running a permutation test I received a P-value of approximately .1481, thus indicating that there is a 14.81% chance that a random permutation of our data would come up with a value greater than or equal to this value. Thus, we accept the null that there isn’t a difference between the variances.

3.13

a) Since we are dealing with 3 separate populations of fish, I would consider running a test of homogeneity. This is because we want to see if there is a difference in the distribution of rays on the second dorsal fin based on the population they are in.

b)

After conducting the test, I received a P-value of .2347. Thus, I would accept the null hypothesis that the number of rays on the second dorsal fin is similar among fish from different regions.

3.16

a)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Gender | Bush | Didn’t Vote | Gore | Nader | Other |
| Female | 459 | 5 | 492 | 26 | 3 |
| Male | 426 | 5 | 289 | 31 | 13 |

b)

After conducting the test, I received an extremely low P-value of approximately .00001. I would therefore conclude that gender and presidential vote in the year 2000 were not independent!

c) After running a permutation test for independence, I received P-value of approximately .0001, which is in-line and confirms the above chi-square test of independence. It seems that gender and who you voted for in the 2000 presidential election aren’t independent!