**Statistics 200: Lab Activity for Section 4.2**

**Note: This is another very important lab. Today students are creating randomization distributions and using them to calculate p-values. Wow – p-values! It is very important that all LAs and TAs understand the stat 200 appropriate definition of a p-value, and also the process and theory behind randomization distributions.**

**Note: For stat 200 students, a randomization distribution shows us what kinds of samples statistics we would see by random chance *if the null is true* Students don’t need to know the nitty gritty details about how exactly randomization distributions are constructed, they just need to know that we do so assuming the null hypothesis is true.**

**Note: You should be able to navigate in the randomization section of StatKey with confidence at this point. Make sure you are very familiar with all the software controls before lab begins.**

**Note: A key take-away for students is that smaller p-values indicate more evidence against the null hypothesis and in favor of the alternative. This a very important lecture for students, so if at all possible… GO TO LECTURE!**

**Measuring Evidence with P-values - Learning objectives:**

* Recognize that a randomization distribution shows what is likely to happen by random chance if the null hypothesis is true
* Use technology to create a randomization distribution
* Interpret a p-value as the proportion of samples that would give a statistic as extreme as the observed sample, if the null hypothesis is true
* Distinguish between one-tailed and two-tailed tests in finding p-values
* Find a p-value from a randomization distribution

**Activity 1**:  ***Create a randomization distribution***

This activity is meant to have you participate in the creation of a randomization distribution to understand that it shows a distribution of sample statistics that were created assuming the null hypothesis is true.

Each week during the NFL season, ESPN has a panel of experts predict the results of professional football games. These predictions are then compared across experts to see who is the best at forecasting games. One of these experts is Mike Golic, who played professional football and has had a sports show on ESPN for many years. During the 2020 NFL playoffs (not including the super bowl), Golic made the following predictions:

|  |  |  |  |
| --- | --- | --- | --- |
| **Game** | **Prediction** | **Winner** | **Prediction accuracy** |
| BUF @ HOU | BUF | HOU | Incorrect |
| TEN @ NE | NE | TEN | Incorrect |
| MIN @ NO | NO | MIN | Incorrect |
| SEA @ PHI | SEA | SEA | Correct |
| MIN @ SF | SF | SF | Correct |
| TEN @ BAL | BAL | TEN | Incorrect |
| HOU @ KC | KC | KC | Correct |
| SEA @ GB | GB | GB | Correct |
| TEN @ KC | KC | KC | Correct |
| GB @ SF | SF | SF | Correct |

Are his predictions better than a random 50-50 chance?

1. What are the correct null and alternative hypotheses? Hint – what is p if his predictions are random?

**H0: p = 0.5 (p=proportion of correct) Ha: p > 0.5**

1. What is p-hat when considering this example?

**p-hat = 6/10 = 0.60**

1. To create a randomization distribution, we must determine what the distribution of p-hat is if his prediction is random. We will use virtual coins, which have a true 50% chance of being heads. Go to [justflipacoin.com](file:///C:\Users\dlp245\Dropbox\PSU\Teaching\17-18%20Spring%20200\Labs\justflipacoin.com).

How many times will you need to flip this penny to create one sample statistic for the randomization distribution?

**Note: students should understand that the sample size for a simulated sample should be the same as that of the original sample, just like with bootstrapping.**

**10 times**

1. Now flip the penny that many times. Pretend that getting a heads with the coin is equivalent to Golic making a correct prediction. What was your p-hat?

**(Answers will vary)**

1. Did your sample make as many correct predictions Golic?

**(Answers will vary)**

1. Now we will go big and have [Statkey](http://www.lock5stat.com/StatKey/index.html) create many many more statistics for our randomization distribution. Notice that the null is already set to the correct proportion.

Now generate at least 5000 samples.

1. Where is the randomization distribution centerd?

**Note: the randomization distribution should always be centered at the null value.**

**0.5**

1. Find the p-value. In StatKey, click on the correct tail (right or left), then click on the box along the x-axis. Enter in our original sample statistic (from part 2).

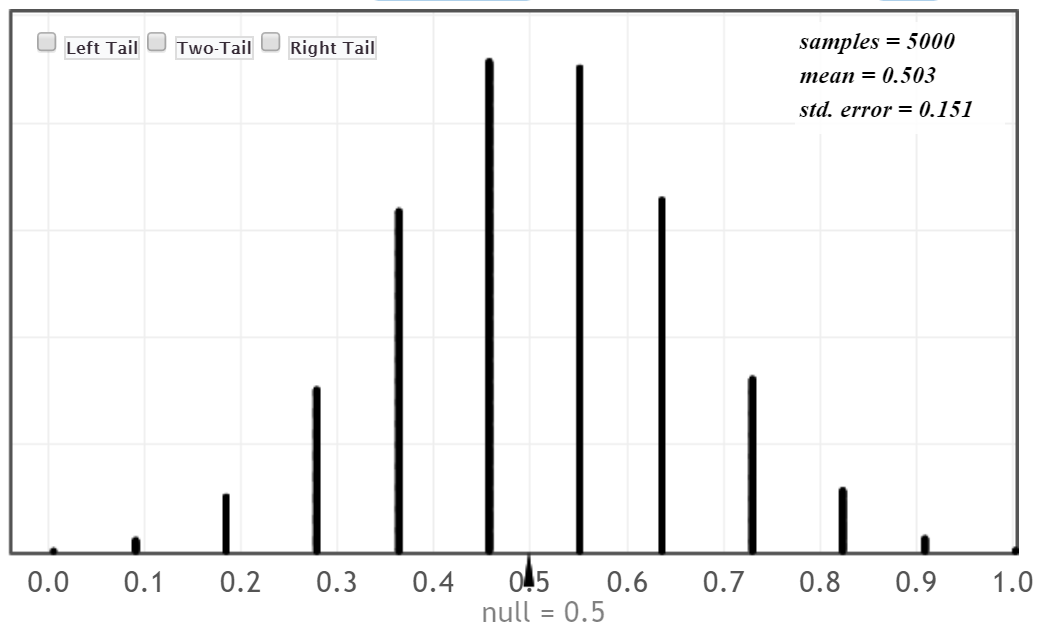
What was the p-value? **0.372 (0.370 – 0.374)**

1. Interpret the p-value in context:

**If Golic is choosing randomly, the chance that he would correctly predict at least 6 out of 10 games is about 0.372.**

**Note: students don’t need to use statkey for the question below, but they certainly could. They will have problems deciding what the correct p-value is for the p-hat. Tell them they need to eyeball it and choose the value that is reasonable. For example, a student should quickly realize that the correct p-hat is .75, but then might get stuck deciding if the p-value is closer to 0.48 or 0.15. Tell them to locate 0.75 on the plot below, and think if the proportion of dots at or above 0.75 is closer to 0.48 or to 0.15.**

1. Let’s say that Golic correctly predicted the Super Bowl, bringing his record to 7 correct of 11. The randomization distribution for this scenario is below:



Using the new data and randomization distribution, what is our sample p-hat and the approximate p-value for testing the same hypothesis we wrote in question 1? Choose the correct answer below.

1. **p-hat = 0.636, p-value = 0.276**
2. p-hat = 0.636, p-value = 0.724
3. p-hat = 1, p-value = 0.007
4. p-hat = 1, p-value = 0.039

**Activity 2: *Where is the middle?***

**This activity is meant to have students understand that the null is always the middle for randomization distributions, and also to have them think about where they would begin shading dots to calculate the p-value. We provide students with a sample answer to replicate.**

**Note: This textbook calculates two-sided p-values by multiplying the smaller tail area by 2. (see example below)**

For the settings below, determine a) where the middle of the randomization will be and b) whether the hypothesis test is right-tailed, left-tailed, or two-tailed. Finally consider c) how to find the p-value.

1. To test H0: μ = 45 vs Ha: μ > 45 using sample data with = 53.7:
2. Where will the randomization distribution be centered?

**45**

1. Is this a left-tail test, a right-tail test, or a two-tail test?

**Right tailed**

1. How can we find the p-value once we have the randomization distribution?

**Find the proportion of randomization statistics that are to the right of the sample statistic of 53.7.**

Example answer: Find the proportion of randomization statistics that are to the left of the sample statistic of 43.7. (use this as a guide when answering questions 2.c and 3.c).

1. To test H0: p1 = p2 vs Ha p1 ≠ p2 using sample data with = 0.35:
2. Where will the randomization distribution be centered?

**At zero**

1. Is this a left-tail test, a right-tail test, or a two-tail test?

**Two tailed**

1. How can we find the p-value once we have the randomization distribution?

**Find the proportion of randomization statistics that are to the right of the sample statistic of -0.35 , then multiply this proportion by two to account for the other tail.**

1. To test H0: ρ = 0 vs Ha:  ρ < 0 (rho) using sample data with r = -0.13:

a. Where will the randomization distribution be centered?

**At zero**

b. Is this a left-tail test, a right-tail test, or a two-tail test?

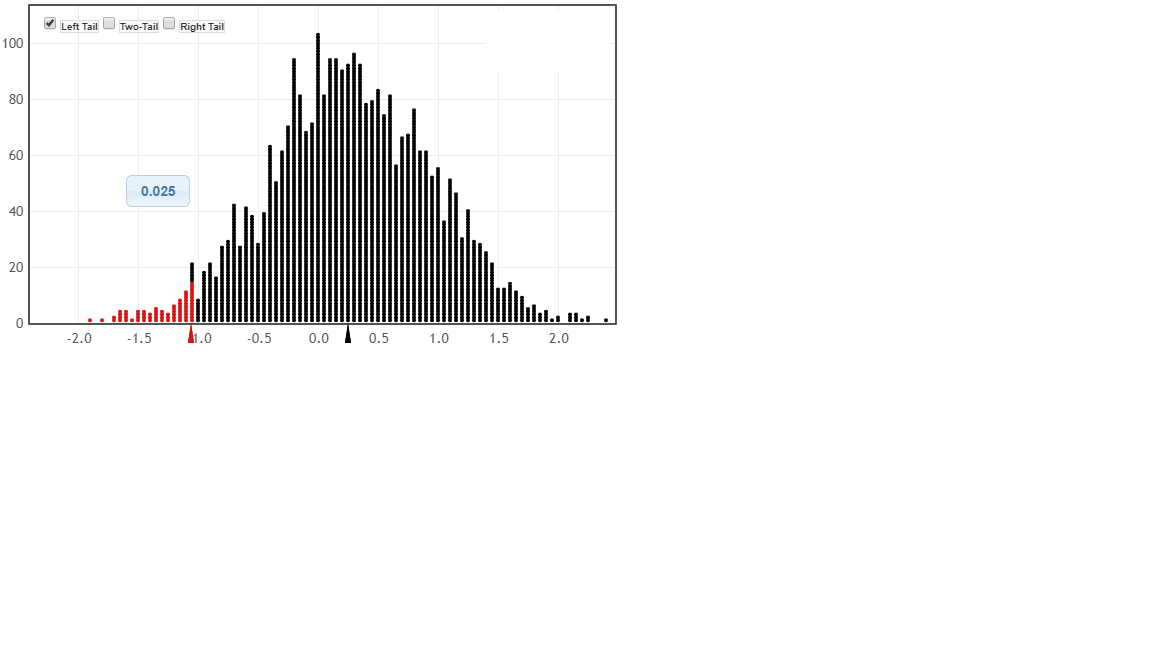
**Left tailed**

1. How can we find the p-value once we have the randomization distribution?

**Find the proportion of randomization statistics that are to the left of the sample statistic of -0.13.**

**Note: again, the question below should be completed only using the material below. This is an exam-type question that illustrates the kind of output students might see on an exam. Students will need to synthesize knowledge to choose the correct answer. First, they’ll have to decide on the null hypothesis based on the middle of the randomization distribution (narrowing choices down to c and d), then they have to decide between the last two options depending on the shading which corresponds to the alternative hypothesis.**

4. Here is a randomization distribution and p-value calculation based on a sample statistic of -1.07.



Select the hypothesis set that could correspond to this randomization distribution and p-value calculation:

* 1. H0: p1 = p2 vs Ha p1 ≠ p2
  2. **H0: μ = 0 .24 vs Ha:  μ < 0.24**
  3. H0: p = 0.3 vs Ha:  p < 0.3
  4. H0: ρ = 0.3 vs Ha:  ρ ≠ 0.3

***Activity 3: Use StatKey to create a randomization distribution and find a p-value.***

In a study to compare on time arrivals of airlines, 1000 Delta flights and 1000 United flights were randomly selected from the month of December in the US. For each flight, the difference between the actual and scheduled arrival time was recorded (so a negative time means the flight was early). We wish to see whether this data provides evidence that Delta has a better arrival record than United (or, more precisely, that the mean difference of times of Delta is significantly lower than the mean of United) **Group 1:** Delta times and **Group 2:** United times.

1. Is this an experiment or an observational study? What are the cases? What are the variables?

**observational study**

**Cases are the 2000 flights**

**Predictor variable - Airline**

**Response variable – Difference in actual and scheduled arrival times**

1. State the null and alternative hypotheses for this test. Define any parameters used.

**H0: (μ1 = μ2 ) more helpful μ1 - μ2 = 0**

**Ha: μ1 - μ2 < 0**

**μ1 = mean time for Delta**

**μ2 = mean time for United**

1. (The data on Arrival Time is one of the available datasets in StatKey, under Test for a Difference in Means.) Use STATKEY to create a randomization distribution for this test using at least 4,000 samples. Use the randomization distribution to indicate whether each of the following possible differences in means is (i) very likely to occur just by random chance, (ii) relatively unlikely to occur but might occur occasionally, or (iii) very unlikely to ever occur just by random chance:

**–7(iii - v. unlikely) 1 (i - v. likely) –4 (ii - rel. unlikely)**

**–0.5(i - v. likely) 6 (iii - v. unlikely)**

**Note: this should be answered when only considering statement in Ho. Want to keep it separate from question asked in part 8.**

1. What is the observed difference in means from the Original Sample?

Give notation and the value of the sample statistic.

**x̅1 − x̅2 = -12.38**

1. Where does the sample statistic lie in the randomization distribution? Is it likely or unlikely to occur just by random chance?

**Very unlikely**

1. Use the sample statistic to find the p-value. Is it large or small?

**Very small -- 0.000**

1. Complete the interpretation for the p-value:
2. **If Delta and United flights are \_\_\_\_equally\_\_\_\_ on time, then the chance that we see a sample statistic of \_\_\_-12.38\_\_\_\_\_ or any statistic \_\_\_\_\_smaller\_\_\_\_\_\_ is \_\_\_\_\_0.00\_\_\_\_\_\_\_\_\_\_.**

Blank 1 options: (equally, not equally) Blank 3 options: (larger, smaller)

8. Use your randomization distribution from part 3 to match the sample statistics below with the closest p-values. Note that is possible to do this without performing the actual calculation.

**Sample statistic p-value**

**-1 -> 0.285**

**1.5 -> 0.804**

**3.25 -> 0.968**

**4.4 -> 0.995**

**Note: Students historically have a lot of difficulty with this type of question (matching sample statistics to p-values). If they just calculate it with StatKey they won’t have issue, but you should reinforce the idea that they should be able to do this without technology.**

**Remind them what the null and alternative hypotheses are, particularly that the alternative is right-tailed (mu1 - mu2 > 0). This means that larger/more positive differences provide more evidence against the null, so they should correspondingly have a smaller p-value!**