Lab 4: Hypothesis Testing Using Randomization

Javier E. Flores

February 27, 2019

Introduction

Throughout recent lectures, we have discussed randomization distributions and how they are used in hypothesis testing. As you may recall, randomization distributions are generated from our sample data under the assumption of some specified null hypothesis. In our initial discussions of null hypotheses, we largely focused on those involving mean differences. In this lab, we will explore how randomization distributions and hypothesis testing proceeds in scenarios with some other common parameters of interest (i.e. single mean, single proportion, difference in proportions, and slope/correlation)

As you work through the lab with your group, you will be asked to answer several questions. Please submit your responses (as a group) in a single, separate document. Include the original questions as well as your group's response in the final submission.

Data

Rather than analyze a single dataset, today we will depart from lab tradition and analyze multiple datasets. Each of these data were pulled from a repository of "miscellaneous" datasets. This repository is owned/managed by Larry Winner, a statistics professor at the University of Florida. Each dataset in this repository was collected from real, but very niche, academic studies. We will be using the following five datasets throughout this lab:

1) Concussion: Counts of concussions among collegiate athletes in five sports for three years by gender. The codebook for these data may be found here.

    Original Study:

    T. Covassin, C.B. Swanik, M.L. Sachs (2003). "Sex Differences and the Incidence of Concussions Among Collegiate Athletes", Journal of Athletic Training, Vol. (38)3, pp238-244

2) Net Profit: Net profit or loss percentages for 352 department stores with sales under $1,000,000 in 1925. The codebook for these data may be found here.

    Original Study:

    McNair (1930) "Margins, Expenses and Profits in Retail Trade in the U.S. as Studied by the Harvard University Bureau of Business Research". The Economic Jounal Vol.40,No.160,pp599-632.

3) Blues Hands: The location, date of birth, hand posture, and thumb style for 93 blues guitarists born between 1874 and 1940. The codebook for these data may be found here.

    Original Study:

    A.M. Cohen (1996). "The Hands of Blues Guitarists," American Music, Vol. 14, #4, pp. 455-479.

4) ESP Skeptic: Data from an experiment that classified subjects as believers or skeptics with respect to psi and then measured successful matches in 50 trials with 5 Zener card symbols (star, waves, square, circle, and cross). The codebook for these data may be found here.

   Original Study:

   L. Storm and M.A. Thalbourne (2005). "The Effect of Change in Pro Attitude on Paranormal Performance: A Pilot Study Using Naive and Sophisticated Skeptics," Journal of Scientific Exploration, Vol.19, #1, pp.11-29.

5) Work Conflict: State level data on industrial conflict, as well as percent union workers, percent working in relatively non-unionized industries, percent working in agriculture, and an indicator of whether the state is a union-shop state. The codebook for these data may be found here.

   Original Study:

   D. Gilbert (1966). "A Statistical Analysis of the Right to Work Conflict", Industrial and Labor Relations Review", Vol 19, #4, pp 533-537.

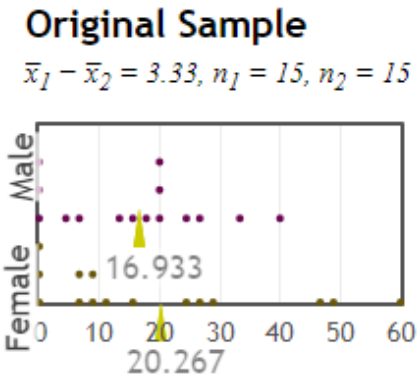
Refresher

Before we explore how randomization is used for different parameters of interest, we'll first walk through an example with which we should be somewhat familiar: hypothesis testing for a difference in means. We'll use the first of our datasets, "Concussion", to answer the question: "Is the average number of concussions in male and female collegiate athletes different?"

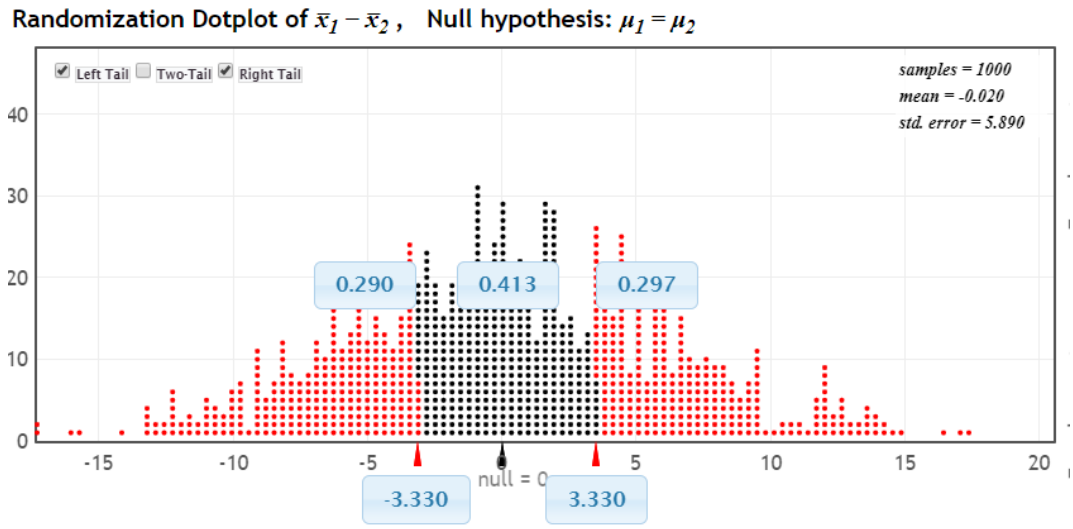
Q1) State the null and alternative hypotheses that would be appropriate for this research question.

Since we will be performing a test for a difference in means, we need to select this option in StatKey under the section titled, "Randomization Hypothesis Tests". After doing so, we then need to input our data using the "Edit Data" button. When copying your data into this field, be sure to only include the columns "Gender" and "Concussion".

After successfully importing the data, you should see in the top right corner a graph accompanied by sample statistics that is labeled "Original Sample":



Using the figure above, we see that StatKey is reporting a difference in mean of 3.33. Be sure to pay close attention to what StatKey defines as $\bar{x}_1$ and $\bar{x}_2$. Here, the females are defined as the first group ($\bar{x}_1$) and the males as the second. Based off of this sample statistic, it would appear that females have (on average) more concussions than males! Let's see if this conclusion holds after going through our testing procedure.



Remember that the p-value is defined as the probability of obtaining a statistic (the mean difference in this case) as or more extreme than what was observed in your original sample. Since our original sample had a mean difference of 3.33, we set the cutoff points on the figure above to be -3.33 and 3.33. To compute the p-value, we then add the areas to the left and right of these cutoffs: 0.290+0.297 = 0.587.

Q2) Given this p-value, what would you conclude? State your conclusion in the context of the original research question. Be sure to state whether we reject or fail to reject the null hypothesis described in **Q1)**.

Before proceeding to subsequent sections, I'd like to remind you that performing a formal hypothesis test requires that you:

1) Clearly state your null and alternative hypotheses
2) State your observed test statistic and compare it to some reference distribution (i.e. the randomization distribution)
3) Provide a p-value along with any conclusions reached in the context of the original research question.

Randomization Testing: Single Mean

Recall that for the difference in mean between two groups, we permuted group labels in order to generate samples to form the randomization distribution. This was done in order to maintain consistency with our null hypothesis, which was that there is no group difference. When our interest is in a single mean, it is easy to see why we can no longer rely on this method to generate a randomization distribution. Without group labels, there is no permutation to be done!

Keep in mind that, in the most general sense, the randomization distribution is a form of sampling distribution generated under a specified null hypothesis. For a single mean, the specified null hypothesis is whether or not your sample mean is equal to some **null value**. As an example, let's say we were interested in the class performance on an exam. We want to know whether the average score is 80%. For this example, the null value would be 80%.

Knowing how the null hypothesis differs for single means, it would then make sense that the randomization distributed is generated differently. For the single mean case, the randomization distribution is generated by first shifting the original sample data points such that their average becomes the null value. These shifted values are then sampled *with replacement* (similar to the bootstrap) to create the randomization distribution.

Q3) Why are the shifted data sampled with replacement?

Q4) The next few questions require the use of the "Net Profit" dataset. Consider the following research question: "Among small department stores in 1925 (sales under $1,000,000) was there, on average, a net profit different from 0?"

- Using proper statistical notation, state the null and alternative hypotheses.
- Input the necessary data and make sure that the null hypothesis is correctly specified in StatKey.
- Generate 10 randomized samples one at a time and track the standard deviation of each in the table below.

Randomization Sample	Standard Deviation
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Q5) How do the standard deviations of each randomization sample compare to that of your original sample? How would you expect them to compare? Why?

Q6) Conduct a *two-sided randomization test* at the $\alpha = 0.05$ level. State your p-value, whether you reject or fail to reject your null hypothesis, and the conclusion you reached in the context of the original research question.

Q7) Conduct a *one-sided randomization test* at the $\alpha = 0.10$ level for whether the average net profit is greater than 0. State the appropriate null and alternative hypotheses, your p-value, whether you reject or fail to reject your null hypothesis, and the conclusion you reached in the context of this new research question.

Randomization Testing: Single Proportion

Once again, with a "new" parameter of interest comes a new and different way of generating the appropriate randomization distribution. Here, the randomization distribution is generated by simulating weighted "coin flips" to reflect the proportion assumed under the null hypothesis. For example, suppose we wanted to determine the proportion of left-handed students at Grinnell. We might hypothesize that this proportion is 15% (this would be our null hypothesis). When generating our randomization distribution, each simulated "coin flip" would then have a 15% chance of recording a success, which would be defined in this example as an individual being left-handed. Repeating this simulated "coin flip" several times generates the randomization distribution.

Q8) The next couple of questions require the use of the "Blues Hands" dataset. Consider the following research question: "Is the extended hand posture the most popular (i.e. used by more than half) among blues musicians between 1874 and 1940?"

- Using proper statistical notation, state the null and alternative hypotheses.
- Using Minitab, create a new binary categorical variable that identifies blues musicians who use the extended hand posture.

Q9) Input the appropriate data into StatKey and conduct a *one-sided randomization test* at the $\alpha = 0.05$ level. State your p-value, whether you reject or fail to reject your null hypothesis, and the conclusion you reached in the context of the original research question.

Randomization Testing: Difference in Proportions

Perhaps not surprisingly, the randomization distribution for a difference in proportions is generated the same way as for a difference in means. The one difference is that for each randomization sample generated, a difference in proportions is computed as opposed to a difference in means.

Q10) The next couple of questions require the use of the "ESP Skeptic" dataset. Consider the following research question: "Are the proportions of successful matches different between ESP skeptics and believers?"

- Using proper statistical notation, state the null and alternative hypotheses.
- Using Minitab, calculate the sum of "Count" separately for believers and non believers. This represents the total number of correct guesses for each group. Count the number of subjects in each group and multiply by 50. This represents the total number of trials for each group.

Q11) Use the "Edit Data" feature in StatKey to import the total correct guesses and total number of trials for each group. Conduct a *two-sided randomization test* at the $\alpha = 0.05$ level. State your p-value, whether you reject or fail to reject your null hypothesis, and the conclusion you reached in the context of the original research question.

Randomization Testing: Slope/Correlation

The last parameter(s) we'll be discussing ate the slope and correlation coefficients. For these parameters, creating the randomization distribution involves reassigning response values to different values of the explanatory variable. This mimics the randomization methods used for testing the difference in means and proportions, but rather than permute group labels, we permute response values. A demonstration is provided in the tables below:

Original Sample

Explanatory (X)	**Response (Y)**
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4
x_5	y_5

Randomization Sample

Explanatory (X)	**Response (Y)**
x_1	y_2
x_2	y_4
x_3	y_1
x_4	y_5
x_5	y_3

Q12) The next couple of questions require the use of the "Work Conflict" dataset. Consider the following research question: "Is the degree of unionization predictive of the degree of work conflict within a state?" State the explanatory and response variables.

Q13) Recall that both correlation and regression can be used to evaluate the relationship between two quantitative variables. For each of these methods, state the appropriate null and alternative hypotheses.

Q14) Input the appropriate data into StatKey and conduct a *two-sided randomization test* at the $\alpha = 0.05$ level. State your p-value, whether you reject or fail to reject your null hypothesis, and the conclusion you reached in the context of the original research question. Do this for both the correlation and regression coefficients.

Q15) Using any one of these datasets or any other found in the "miscellaneous" dataset repository, formulate a research question that can be answered using any one of the testing methods covered in this lab. Be sure to clearly state:

- which dataset you are using
- your research question
- which testing procedure you feel is appropriate to answer the question
- your null and alternative hypotheses
- your p-value
- whether you reject or fail to reject your null hypothesis
- the conclusion you reached in the context of the original research question

Challenge (Optional)

Repeat the analysis from **Q15)** in R. Include a histogram of the randomization distribution. Shade the area(s) of this histogram that are used to compute the p-value. Be sure to thoroughly comment your code with explanations of what tasks the written lines execute.

You may work on this challenge as a group, but I expect individuals to turn in separate copies of their own code and generated statistics/figures.

Turning in this challenge may prompt a brief assessment in which you will be asked to state the code used to generate any one of the results produced.

This challenge may be turned in and redeemed for credit at any point prior to the second exam, which will tentatively be scheduled one week after we return from Spring Break