Homework 5: Sections 4.1 - 4.5
KEY

STA209-04: Applied Statistics

Assigned: 03/04/2019
Due: 03/11/2019

Total Possible Points: 55

From the Book:

Questions: 4.26, 4.29, 4.30, 4.77, 4.78 (a - e only), 4.117, 4.119, 4.149, 4.152, 4.176

4.26 [**3 pts**] For a random sample of households in the US, we record annual household income, whether the location is east or west of the Mississippi River, and the number of children. We are interested in determining whether there is a difference in average household income between those east of the Mississippi and those west of the Mississippi.

a) [**2 pts**] Define the relevant parameter(s) and state the null and alternative hypotheses.

For this problem, the relevant parameters are the mean household income east of the Mississippi (μ_{EM}) and west of the Mississippi (μ_{WM}). Alternatively, the relevant parameter may be the difference in these means, $\mu_{EM} - \mu_{WM}$. The null hypothesis is that there is no difference in mean household income between those east and west of the Mississippi ($H_0 : \mu_{EM} - \mu_{WM} = 0$), and the alternative hypothesis is that there is ($H_A : \mu_{EM} - \mu_{WM} \neq 0$)

b) [**1 pts**] What statistic(s) from the sample would we use to estimate the difference?

We would use the difference in sample means, i.e. $\bar{x}_{EM} - \bar{x}_{WM}$.

4.29 [**4 pts**] By some accounts, the first formal hypothesis test to use statistics involved the claim of a lady tasting tea. In the 1920's Muriel Bristol-Roach,a British biological scientist, was at a tea party where she claimed to be able to tell whether milk was poured into a cup before or after the tea. R. A. Fisher, an eminent statistician, was also attending the party. As a natural skeptic, Fisher assumed that Muriel had no ability to distinguish whether the milk or tea was poured first, and decided to test her claim. An experiment was designed in which Muriel would be presented with some cups of tea with the milk poured first, and some cups with the tea poured first.

a) [**2 pts**] In plain English (no symbols), describe the null and alternative hypotheses for this scenario.

The null hypothesis is that Muriel has no ability to distinguish whether the milk or tea was poured first. This corresponds to guessing correctly at a rate consistent with random chance (e.g. 50%). The alternative hypothesis is that Muriel can distinguish whether the milk or tea was poured first. This corresponds to guessing correctly at a rate greater than 50%.

b) [**2 pts**] Let p be the true proportion of times Muriel can guess correctly. State the null and alternative hypothesis in terms of p.

$H_0 : p = 0.50; H_A : p > 0.50$

4.30 [**4 pts**] Studies have shown that omega-3 fatty acids have a wide variety of health benefits. Omega-3 oils can be found in foods such as fish, walnuts, and flaxseed. A company selling milled flaxseed advertises that one tablespoon of the product contains, on average, at least 3800 mg of ALNA, the primary omega-3.

a) [2 pts] The company plans to conduct a test to ensure that there is sufficient evidence that its claim is correct. To be safe, the company wants to make sure that evidence shows the average is higher than 3800 mg. What are the null and alternative hypotheses?

Let μ denote the average amount of ALNA contained in the product. $H_0 : \mu = 3800; H_A : \mu > 3800$

b) [2 pts] Suppose, instead, that a consumer organization plans to conduct a test to see if there is evidence *against* the claim that the product contains an average of 3800 mg per tablespoon. The consumer organization will only take action if it finds evidence that the claim made by the company is false and that the actual average amount of omega-3 is less than 3800 mg. What are the null and alternative hypotheses?

Let μ denote the average amount of ALNA contained in the product. $H_0 : \mu = 3800; H_A : \mu < 3800$

4.77 [2 pts] Using the definition of a p-value, explain why the area in the tail of a randomization distribution is used to compute a p-value.

The p-value is the probability of obtaining a statistic as or more extreme than that observed in your sample. The area in the tail of a randomization distribution is the proportion of all distribution values which are greater (or less) than the specified cutoff. This proportion is the probability under the randomization distribution.

4.78 [7 pts] You roll a die 60 times and record the sample proportion of 5's, and you want to test whether the die is biased to give more 5's than a fair die would ordinarily give. To find the p-value for your sample data, you create a randomization distribution of proportions of 5's in many simulated samples of size 60 with a fair die.

a) [2 pts] State the null and alternative hypotheses.

Let p represent the probability of rolling a 5. $H_0 : p = 1/6; H_A : p \neq 1/6$.

b) [2 pts] Where will the center of the distribution be? Why?

The center of this distribution will be at 1/6 since this is the most likely value under the null hypothesis.

c) [1 pts] Give an example of a sample proportion for which the number of 5's obtained is *less* than what you would expect in a fair die.

Anything pretty far below 1/6 would be reasonable. One example would be 1/12.

d) [1 pts] Will your answer in part (c) lie on the left or the right of the center of the randomization distribution?

It would lie to the left of the center of the randomization distribution.

e) [1 pts] To find the p-value for your answer to part (c), would you look at the left, right, or both tails?

Since we are interested in determining if the number of 5's is less than what would be expected under a fair dice, we are performing a one-sided hypothesis test. Therefore, we would look at the left tail of the randomization distribution only.

4.117 [4 pts] This exercise addresses lizard behavior in response to fire ants. The red imported fire ant, *Solenopsis invicta*, is native to South America, but has an expansive invasive range, including much of the southern United States (invasion of this ant is predicted to go global). In the United States, these ants occupy similar habitats as fence lizards. The ants eat the lizards and the lizards eat the ants, and in either scenario the venom from the fire ant can be fatal to the lizard. A study explored the question of whether lizards learn to adapt their behavior if their environment has been invaded by the fire ants. The researchers selected lizards from an uninvaded habitat (eastern Arkansas) and lizards from an invaded habitat (southern Alabama, which has been invaded for more than 70 years) and exposed them to fire ants. They measured how long it takes each lizard to flee and the number of twitches each lizard does. The data are stored in FireAnts.

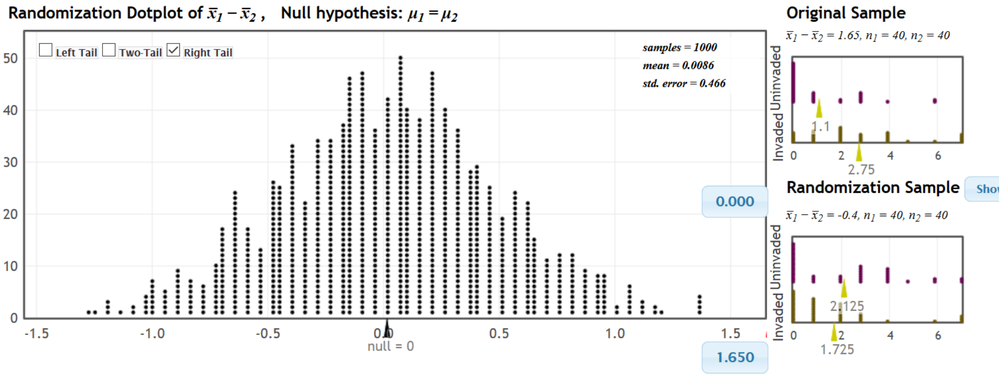
If lizards adapt their behavior to the fire ants, then lizards from the invaded habitats should twitch more than lizards from the uninvaded habitats when exposed to red imported fire ants (twitching helps to repel the ants). Test this hypothesis. The variable *Twitches* is the number of twitches exhibited by each lizard in the first minute after exposure.

a) **[2 pts]** State the null and alternative hypotheses.

Let μ_{IH} and μ_{UH} correspond to the mean number of twitches by lizards in invaded and univaded habitats, respectively. $H_0 : \mu_{IH} - \mu_{UH} = 0; H_A : \mu_{IH} - \mu_{UH} > 0;$

b) **[1 pts]** Use technology to calculate the p-value.

The p-value is 0.000


Randomization Dotplot of $\bar{x}_1 - \bar{x}_2$, Null hypothesis: $\mu_1 = \mu_2$
□ Left Tail □ Two-Tail ☑ Right Tail
samples = 1000
mean = 0.0086
std. error = 0.466
50
40
30
20
10
0
-1.5
-1.0
-0.5
0.0
null = 0
0.5
1.0
1.5
0.000
1.650
Original Sample
$\bar{x}_1 - \bar{x}_2 = 1.65, n_1 = 40, n_2 = 40$
Invaded Uninvaded
1.1
2.75
0
2
4
6
Randomization Sample
Show
$\bar{x}_1 - \bar{x}_2 = -0.4, n_1 = 40, n_2 = 40$
Invaded Uninvaded
2.125
1.725
0
2
4
6

c) **[1 pts]** What (if anything) does this p-value tell you about lizards and fire ants?

Given this p-value, which provides extremely compelling evidence against our null hypothesis, we may conclude that lizards have adapted their behavior to the invasion of fire ants. Specifically, lizards from invaded habitats twitch more often than those lizards from uninvaded habitats so as to repel the fire ants which they know to be dangerous.

4.119 **[7 pts]** Could owning a cat as a child be related to mental illness later in life? Toxoplasmosis is a disease transmitted primarily through contact with cat feces, and has recently been linked with schizophrenia and other mental illnesses. Also, people infectedwith Toxoplasmosis tend to like cats more and are 2.5 times more likely to get in a car accident, due to delayed reaction times. The CDC estimates that about 22.5% of Americans are infected with Toxoplasmosis (most have no symptoms), and this prevalence can be as high as 95% in other arts of the world. A study randomly selected 262 people registered with the National Alliance for the Mentally ILL (NAMI), almost all of whom had schizophrenia, and for each person selected, chose two people from families without mental illness who were the same age, sex, and socioeconomic status as the person selected from NAMI. Each participant was asked whether or not they owned a cat as a child. The results showed that 136 of the 262 people in the mentally ill group had owned a cat, while 220 of the 522 people in the not mentally ill group had owned a cat.

a) **[2 pts]** This is known as a case-control study, where *cases* are selected as people with a specific disease or trait, and controls are chosen to be people without the disease or trait being studied. Both cases and controls are then asked about some variable from their past being studied as a potential risk factor. This is particularly useful for studying rare diseases (such as schizophrenia), because the design ensures a sufficient sample size of people with the disease. Can case-control studies such as this be used to infer a causal relationship between the hypothesized risk factor (e.g., cat ownership) and the disease (e.g., schizophrenia)? Why or why not?

Case-control studies may not be used to infer a causal relationship. This study design is inherently observational in nature and is subject to confounding and other sources of bias (e.g. recall bias).

b) **[1 pts]** In case-control studies, controls are usually chosen to be similar to the cases. For example, in this study each control was chosen to be the same age, sex, and socioeconomic status as the corresponding case. Why choose controls who are similar to cases?

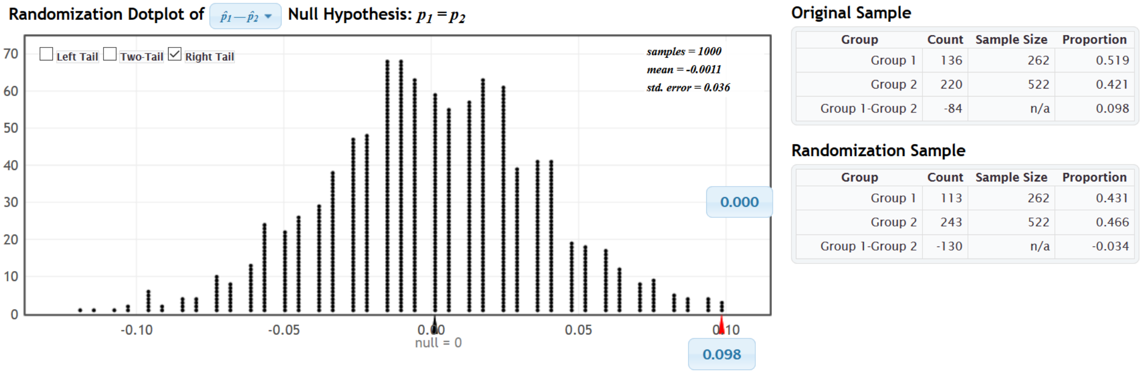
Controls are chosen to be similar to cases so as to minimize the potential confounding bias of the study. Typically this "matching" is done on variables that are known to be associated with the outcome of interest.

c) **[1 pts]** For this study, calculate the relevant difference in proportions; proportion of cases (those with schizophrenia) who owned a cat as a child minus the proportion of controls (no mental illness) who owned a cat as a child.

$136/262 - 220/522 = 0.0976$

d) **[1 pts]** For testing the hypothesis that the proportion of cat owners is higher in the schizophrenic group then the control group, use technology to generate a randomization distribution and calculate the p-value.

The p-value is 0.000


Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$
Null Hypothesis: $p_1 = p_2$
□ Left Tail □ Two-Tail ☑ Right Tail
samples = 1000
mean = -0.0011
std. error = 0.036
70
60
50
40
30
20
10
0
-0.10
-0.05
0.00
null = 0
0.05
0.10
0.000
0.098

Original Sample

Group	Count	Sample Size	Proportion
Group 1	136	262	0.519
Group 2	220	522	0.421
Group 1-Group 2	-84	n/a	0.098

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	113	262	0.431
Group 2	243	522	0.466
Group 1-Group 2	-130	n/a	-0.034

e) **[2 pts]** Do you think this provides evidence that there is an association between owning a cat as a child and developing schizophrenia? Why or why not?

Yes. The p-value suggests that, under the null hypothesis of no difference in proportion, observing this proportion difference in our sample is highly unlikely.

4.149 **[6 pts]** *Newscientist.com* ran the headline ”Breakfast Cereals Boost Chances of Conceiving Boys,” based on an article which found that women who eat breakfast cereal before becoming pregnant are significantly more likely to conceive boys. The study used a significance level of $\alpha = 0.01$. The researchers kept track of 133 foods and, for each food, tested whether there was a difference in the proportion conceiving boys between women who ate the food and women who didn't. Of all the foods, only breakfast cereal showed a significant difference.

a) **[2 pts]** If none of the 133 foods actually have an effect on the gender of a conceived child, how many (if any) of the individual tests would you expect to show a significant result just by random chance? Explain. (*Hint*: Pay attention to the significance level.)

We would expect $133 * 0.1 = 1.33 \approx 1$ test to show a significant result. Recall that the significance level specifies our Type I error rate. With $\alpha = 0.01$, we expect a Type I error to be committed once for every hundred tests performed.

b) **[2 pts]** Do you think the researchers made a Type I error? Why or why not?

It is highly likely that a Type I error rate was committed given that our rejection rate is compatible with the Type I error rate. In other words, we've found 1 significant result out of 133 tests while we expect to commit 1 Type I error in performing 100 tests.

c) **[2 pts]** Even if you could somehow ascertain that the researchers did not make a Type I error, that is, women who eat breakfast cereal are actually more likely to give bith to boys, should you believe the headline ”Breakfast Cereals Boost Chances of Conceiving Boys”? Why or why not?

The headline should not be believed. This study is not randomized and is therefore subject to confounding bias. We cannot claim that eating cereal is the direct cause for an increased chance of conceiving a boy. Rather, assuming a Type I error was not committed, we can only say the two are associated.

4.152 [8 pts] Exercise 4.119 on page 303 revealed an association between owning a cat as a child and developing schizophrenia later in life. Many people enjoy cats as pets, so this conclusion has profound implications and could change pet ownership habits substantially. However, because of the chance for false positives (TypeI errors) and potential problems with generalizability, good scientific conclusions rarely rest on a foundation of just one study. Because of this, significant results often require *replication* with follow up studies before they are truly trusted. If study results can be replicated, especially in a slightly different setting, they become more trustworthy, and if results can not be replicated, suspicions of a Type I error (significant results by random chance) or a lack of generalizability from the setting of the initial study may arise. In fact, the paper cited in Exercise 4.119 actually provided three different datasets, all from different years (1982, 1992, and 1997) and with different choices for choosing the control group. The sample proportions for each dataset, with the sample sizes in the denominator, are given in Table 4.13.

Table 4.13

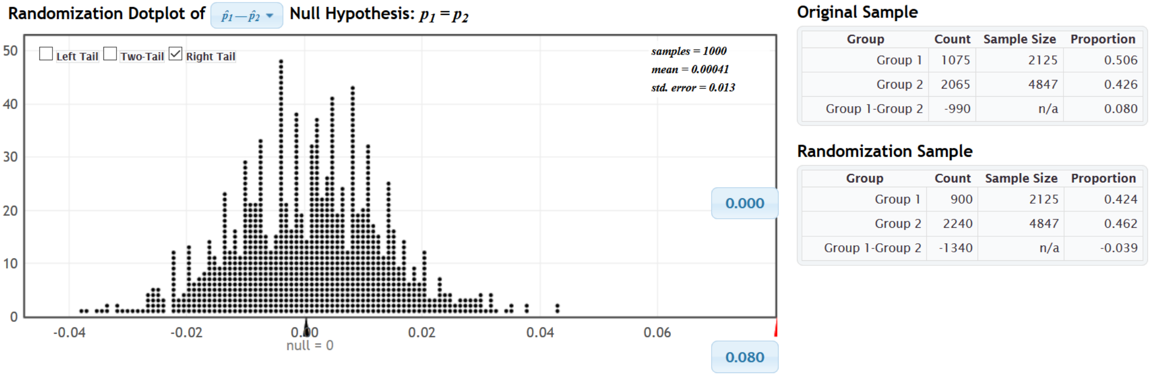
Year	**Proportion of Schizophrenics who owned cats as children**	**Proportion of controls who owned cats as children**
1982 Data (Analyzed in 2015)	$1075/2125 = 0.506$	$2065/4847 = 0.426$
1992 Data	$84/165 = 0.509$	$65/165 = 0.394$
1997 Data	$136/262 = 0.519$	$220/522 = 0.421$

a) [1 pts] As we know, statistics vary from sample to sample naturally, so it is not surprising that the sample proportions differ slightly from year to year. However, does the relative consistency of the sample proportions affect the credibility of any single dataset?

No. Imagining a sampling distribution for the proportion, extremes are possible. If a single dataset yields a proportion that is not consistent with the more commonly observed sample proportion(s), it is not necessarily the case that those data are without credibility.

b) [3 pts] Use technology to calculate the p-value for each dataset, testing the alternative hypothesis that the proportion of cat owners is higher among schizophrenics.

The figures below are presented in an order consistent with the row ordering of the datasets. The corresponding p-values are 0.000, 0.021, and 0.003.

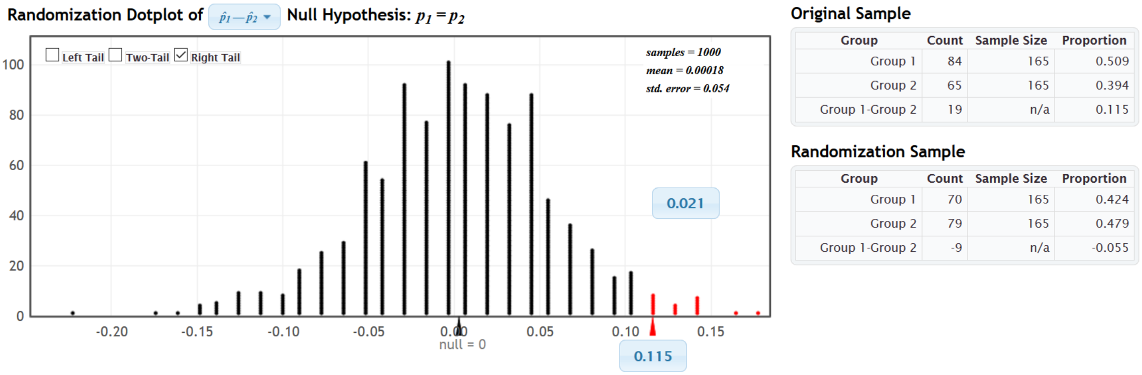


Original Sample

Group	Count	Sample Size	Proportion
Group 1	1075	2125	0.506
Group 2	2065	4847	0.426
Group 1-Group 2	-990	n/a	0.080

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	900	2125	0.424
Group 2	2240	4847	0.462
Group 1-Group 2	-1340	n/a	-0.039

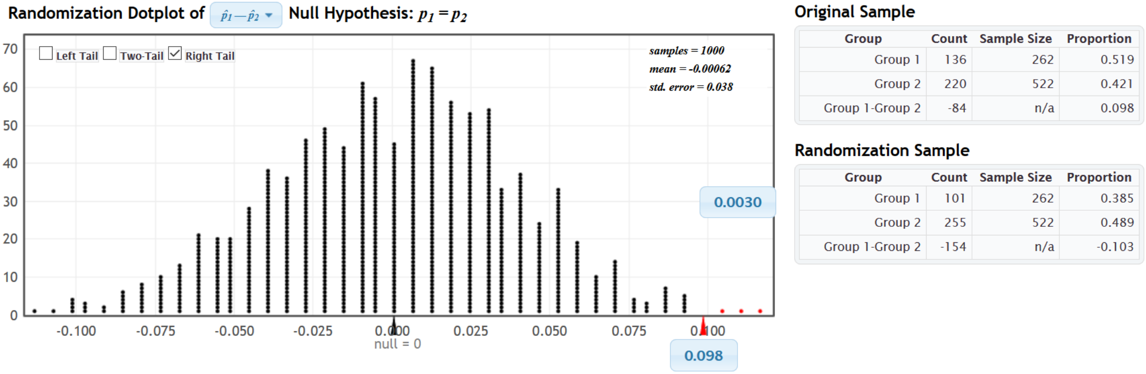


Original Sample

Group	Count	Sample Size	Proportion
Group 1	84	165	0.509
Group 2	65	165	0.394
Group 1-Group 2	19	n/a	0.115

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	70	165	0.424
Group 2	79	165	0.479
Group 1-Group 2	-9	n/a	-0.055



Original Sample

Group	Count	Sample Size	Proportion
Group 1	136	262	0.519
Group 2	220	522	0.421
Group 1-Group 2	-84	n/a	0.098

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	101	262	0.385
Group 2	255	522	0.489
Group 1-Group 2	-154	n/a	-0.103

c) **[2 pts]** Do all datasets yield significant results? Should this increase or decrease potential suspicions that the significance of any single study may have been just a Type I error?

All tests yield significant results. This should decrease suspicions that the significance of any single study was attributed to Type I error.

d) **[2 pts]** Why is the p-value lowest for the 1982 data, even though this dataset yields the smallest difference in proportions? Similarly, why is the p-value highest for the 1992 data, even though this data yielded the largest difference in proportions?

The sample size for the 1982 data is the largest of all three studies. As such, you'll notice that the standard error is the smallest here. With a small standard error, your distribution is pulled closer towards the center which reduces the amount of area in the extreme tails. The opposite occurs with smaller samples. This is why the 1992 data had the largest p-value. It's sample size was the smallest of all three studies and therefore it's standard error was the highest. As a result, the randomization distribution is more spread out, and there are more datapoints in the extreme tail ends.

4.176 **[10 pts]** Exercise 4.102 on page 298 describes a study in which a random sample of 24 adults are divided equally into two groups and given a list of 24 words to memorize. During a break, one group takes a 90-minute nap while another group is given a caffeine pill. The response variable of interest is the number of words participants are able to recall following the break. We are testing to see if there is a difference in the average number of words a person can recall depending on whether the person slept or ingested caffeine. The data are shown in Table 4.17 and are available in SleepCaffeine.

a) **[2 pts]** Define any relevant parameter(s) and state the null and alternative hypotheses.

Let μ_{nap} and μ_{pill} denote the average number of recalled words in the nap group and caffeine pill groups respectively. $H_0 : \mu_{nap} - \mu_{pill} = 0; H_A : \mu_{nap} - \mu_{pill} \neq 0$

b) **[1 pts]** What assumption do we make in creating the randomization distribution?

We assume that the average number of recalled words in either group is the same. Consequently, it should not matter which observations are labeled as belonging to the nap group or caffeine pill group.

c) [**2 pts**] What statistic will we record for each of the simulated samples to create the randomization distribution? What is the value of that statistic for the observed sample?

The statistic we will record is the difference in sample means, $\bar{x}_{nap} - \bar{x}_{pill}$. The value of this statistic for the observed sample is 3.

d) [**1 pts**] Where will the randomization distribution be centered?

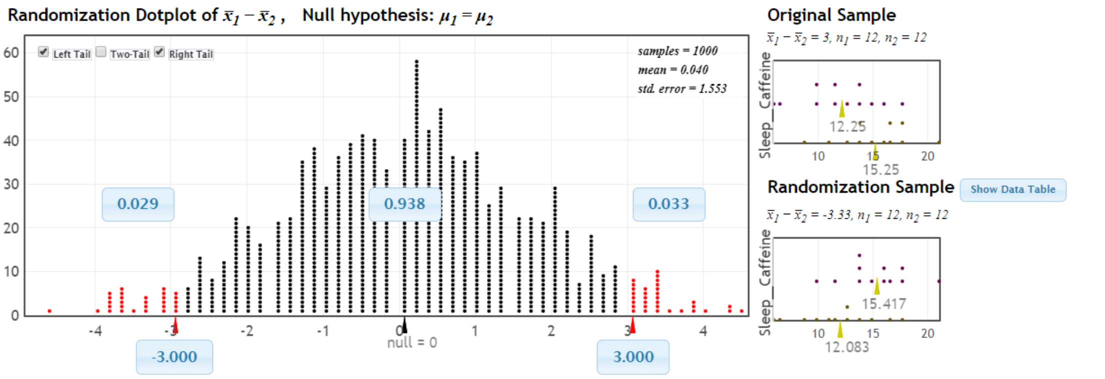
The randomization distribution will be centered at 0, which is the hypothesized null value.

e) [**2 pts**] Find one point on the randomization distribution by randomly dividing the 24 data values into two groups. Describe how you divide the data into two groups and show the values in each group for the simulated sample. Compute the sample mean in each group and compute the difference in the sample means for this simulated result.

Answers will vary. Check that a description of how the simulated data were obtained is provided. Also check that the simulated data itself is provided. Lastly, check for the sample means of each group and the difference in means.

f) [**1 pts**] Use *StatKey* or other technology to create a randomization distribution. Estimate the p-value for the observed difference in means given in part (c).

Using StatKey, the obtained p-value is 0.062.



Randomization Dotplot of $\bar{x}_1 - \bar{x}_2$, Null hypothesis: $\mu_1 = \mu_2$

☑ Left Tail ☐ Two-Tail ☑ Right Tail

samples = 1000
mean = 0.040
std. error = 1.553

Original Sample

$\bar{x}_1 - \bar{x}_2 = 3, n_1 = 12, n_2 = 12$

Randomization Sample

$\bar{x}_1 - \bar{x}_2 = -3.33, n_1 = 12, n_2 = 12$

g) [**1 pts**] At a significance level of $\alpha = 0.01$, what is the conclusion of the test? Interpret the result in context.

At a significance level of $\alpha = 0.01$, we would fail to reject the null hypothesis. There is insufficient evidence to support the conclusion that the average number of recalled words is different between individuals who napped and those who were provided with a caffeine pill.