Worksheet 9

Author:

Discussants:

===

The following worksheet with give you practice running hypothesis tests for two means. Please answers the following questions and submit a compiled pdf with your answers to Moodle by 11:59pm on Sunday November 11th. Note, the following symbols might be useful: μ, H_0, H_a, μ_{sleep}, $\bar{x}_{control}$, $\neq$, α

Part 1: Lock5 questions to understand hypothesis tests

Lock5 exercise 4.52 (first edition): **Arsentic in Chicken**

A study was done to determine if the mean level of arsenic in chicken meat is above 80 ppb (see page 228 in the first edition of Lock5 or page 265 in the second edition). If a restaurant chain finds significant evidence that the mean arsenic level is above 80, the chain will stop using that supplier of chicken meat. The hypotheses are:

$H_0 : \mu = 80$

$H_0 : \mu > 80$

Where μ represents the mean arsenic level in all chicken meat from that supplier. Samples from two different suppliers are analyzed, and the resulting p-values are given:

Sample from Supplier A: p-value is 0.0003

Sample from Supplier B: p-value is 0.3500

 a) Interpret each p-value in terms of the probability of the results happening by random chance.

 b) Which p-value shows stronger evidence for the alternative hypothesis? What does this mean in terms of arsenic and chickens?

 c) Which supplier, A or B, should the chain get chickens from in order to avoid too high a level of arsenic?

Answers:

 a)

 b)

 c)

Lock5 exercise 4.60: **Smiles and Leniency.** An experiment was conducted to study the effects of smiling on leniency in judging students accused of cheating (see Data 4.2 on page 223 of the first edition of Lock5, and the full data are in the Smiles data set). Let's examine the hypothesis that $H_0 : \mu_s = \mu_n$ vs. $H_a : \mu_s > \mu_n$ and use data to test if there is evidence that average leniency score is higher for smiling students (μ_s) than for students with a neutral expressions (μ_n). A dot plot for the difference in sample means based on 1,000
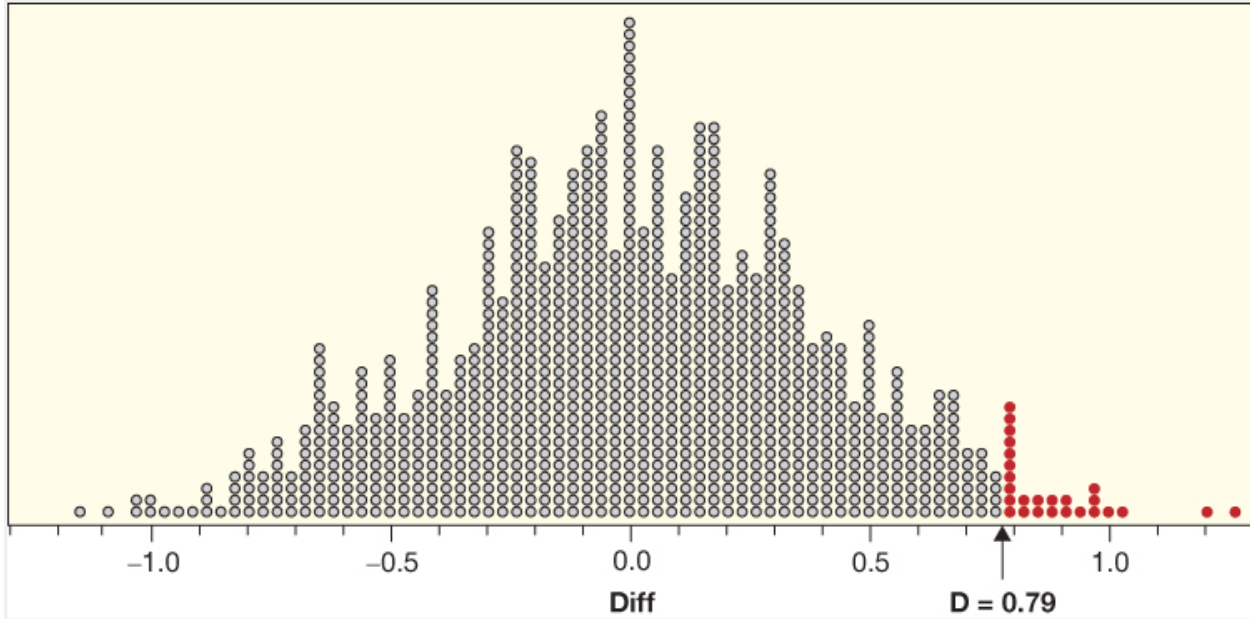
Figure 1: Figure 4.20

random assignments of leniency scores from the original sampel to smile and neutral groups is shown in Figure 4.20.

a) The difference in sample means for the original samples is $D = \bar{x}_s - \bar{x}_n = 4.91 - 4.12 = 0.79$ (as shown in Figure 4.20). What is the p-value for the one-tained test? Hint: There are 27 dots in the tail beyond 0.79.

b) We can also consider the test with a two-tailed alternative, $H_0 : \mu_s = \mu_n$ vs. $H_a : \mu_s \neq \mu_n$, where we make no assumption in advance on whether smiling helps or discourages leniency (see Example 4.3 on page 223 of the first edition of Lock5). How would the randomization distribution in Figure 4.20 chance for this test? How would the p-value change?

Answers:

a)

b)

Lock5 exercise 4.82 (first edition): Preliminary note: For this next question we will use the Neyman-Pearson paradigm of hypothesis testing where we will strict decision that a result is *statistically significant* if a p-value is less than a particular value. We will discuss the Neyman-Pearson paradigm more soon. For the questions below if the p-value is less than 0.05 we will reject the null hypothesis (as we will discuss in class we will say that we set $\alpha = 0.05$ so if the p-value is less than 0.05 we will reject the null hypothesis).

Sleep or Caffeine for Memory? The consumption of caffeine to benefit alertness is a common activity practiced by 90% of adults in North America. Often caffeine is used in order to replace the need for sleep. One recent study compares students' ability to recall memorized information after either the consumption of caffeine or a brief sleep. A random sample of 35 adults (between the ages of 18 and 39) were randomly divided into three groups and verbally given a list of 24 words to memorize. During a break, one of the groups takes a nap for an hour and a half, another group is kept awake and then given a caffeine pill an hour prior to the testing, and a third group is given a placebo. The response variable of interest is the number of words participants are able to recall following the break.

The summary statistics for the three groups are in Table 4.8. We are interested in testing whether there is evidence of a difference in average recall ability between any two of the treatments. Thus we have three possible tests between different pairs of groups: Sleep vs Caffeine, Sleep vs Placebo, and Caffeine vs Placebo.

a) In the test comparing the sleep group to the caffeine group, the p-value is 0.003. What is the conclusion of the test? In the sample, which group had better recall ability? According to the test results, do you think sleep is really better than caffeine for recall ability?

b) In the test comparing sleep group to the placebo group, the p-value is 0.06. What is the conclusion of the test using a 5% significance level? How strong is the evidence of a difference in mean recall ability between these two treatments?

c) In the test comparing the caffeine group to the placebo group, the p-value is 0.22. What is the conclusion of the test? In the sample, which group had better recall ability? According to the test results, would we be justified in concluding that caffeine impairs recall ability?

d) According to this study, what should you do before an exam that asks you to recall information?

Table 4.8: Effect of sleep and caffeine on memory

Group	Sample size	Mean	Standard deviation
Sleep	12	51.25	3.3
Caffeine	12	12.25	3.5
Placebo	11	13.70	3.0

Answers:

a)

b)

c)

d)

Lock5 exercise 4.130 (first edition): ** Effect of Sleep and Caffeine on Memory** Continuing with the analysis of sleep on memory, let's now run the actual hypothesis test to see if there really is statistically significant between the group that got *sleep and the group that got caffeine* (we will use R instead of Statskey to do the analyses). Please answer the following questions:

a) Define any relevant parameter(s) and state the null and alternative hypotheses using words and symbols (step 1).

b) What assumption do we make in creating the randomization distribution?

c) What statistic will we record for each of the simulated samples to create the null distribution? What is the value of that statistic for the observed sample (step 2)?

d) Where will the null distribution be centered?

e) Find one point on the null distribution by randomly dividing the 24 data values into two groups. Describe how you divided 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 results.

f) Use R to create a null distribution (step 3). Estimate the p-value for the observed difference in means given in part (c).

g) Do the results seem statistical significant (step 4)? What does this mean in this context (step 5)?

Answers:

a)

b)

c)

d)

e)

f)

g)

```r
# here is the data from the study
sleep_condition <- c(14, 18, 11, 13, 18, 17, 21, 9, 16, 17, 14,
    15)
caffeine_condition <- c(12, 12, 14, 13, 6, 18, 14, 16, 10, 7,
    15, 10)


# compute the observed statistic


# compute one randomized statistic for part e



# use a for loop to compute the full null distribution can
# get the p-value for part f





# plot the null distribution



# calculate the p-value
```

Reflection

How did this worksheet go? Is the concept of hypothesis testing clear? Do you understand how to apply the code to calculate whether there is a significant difference between population means?