

## MATH 2510: WEEKS 11 AND 12

JOSEPH TIMMER (EDITED BY COLTON GRAINGER)  
NOVEMBER 10, 2019

This was prepared for University of Colorado's MATH 2510 (Introduction to Statistics) in Fall 2019. Please send questions, comments, complaints, and corrections to [colton.grainger@colorado.edu](mailto:colton.grainger@colorado.edu). Any mistakes are my own. This material corresponds to Chapter 8 of *Understandable Statistics* [Brase and Brase, 2018]. Contents

0.

### Definitions

After reading through Brase and Brase Chapter 8, know the definitions of the following.<sup>1</sup>

- (1) null hypothesis
- (2) alternate hypothesis
- (3)  $p$ -value
- (4)  $\alpha$ , the level of significance.
- (5) error type
- (6) test type (determined by  $H_1$ )
- (7)  $z$  (standard normal) tests
- (8)  $t$  (student's  $t$ ) tests
- (9) rejecting the null hypothesis  $H_0$  when  $p \leq \alpha$
- (10) paired versus unpaired data
- (11) tests for which  $H_0 : \mu = 0$  is **always** the null hypothesis

---

<sup>1</sup>“If you are like me, you may also despise the memorization aspect of math. I’ve always disliked how tests force you to memorize. But we need a common and efficient language to speak of complex ideas, and this section is meant to make sure that you can communicate using the language on which English-speaking mathematicians have come to agree. Moreover, I hope that some of these ideas become intuitive enough that you will even be able to guess the correct definition, based on how we have come to know these words.” [Hiro Lee Tanaka, Algebra 122 Midterm Review Notes, 2014]

1.

## Lecture Outlines

**Hypothesis Testing** In Chapter 7, we estimated the value of population parameters (mean and proportion) using confidence intervals. Another method of statistical inference is to make decisions concerning the value of a population parameter, which we do in Chapter 8 with hypothesis testing.

- (1) Suppose that you roll a regular six-sided die 600 times. About how many times would you expect to see a 4 rolled within those 600?
  - (a) If you saw 105 rolls that were 4, would this be surprising... enough to question the fairness of the die?
  - (b) What if you saw 595 rolls that were 4, would this be surprising... enough to question the fairness of the die?
  - (c) Where would you draw the line between “not so surprising” and “surprising”?
- (2) The basic idea in hypothesis testing is to start with an assumption of what “should” happen and to draw a line on what extreme outcomes would be “surprising”. If the random sample indicates a “surprising” result, we have evidence to abandon our assumption... if the random sample indicates a “not so surprising” result, we do not have adequate evidence to abandon our assumption and we must stick with it.
- (3) Note, as with the case of the rolls of the die, the result of the random sample may be very “surprising” (595 of our 600 rolls were 4), but it will never serve as PROOF that our assumption is wrong (as it’s possible that this 595/600 happens with a completely fair die).

### 8.1 Introduction to Statistical Tests

- (1) This section introduces the language and formalizes the concepts of “drawing a line” and interpreting the results of the test.
- (2) When reading, pay close attention to
  - (a) The notation and definition/usage of the null hypothesis and the alternate hypothesis.
  - (b) How we categorize the test as right-tailed, left-tailed, or two-tailed.
  - (c) What the P-value measures and how it is used to draw the conclusion of the test.
  - (d) The usage and meaning of the conventional language of “Reject  $H_0$ ” and “Fail to Reject  $H_0$ ”.

### 8.2 Testing the Mean

- (1) The calculator functions that will be useful are **Z-Test** and **T-Test**. Like **Z-Interval** and **T-Interval**, one is used with  $\sigma$  is known and the other when  $\sigma$  is unknown.
- (2) One may find the “critical regions” method helpful in solidifying the concepts of hypothesis testing, but one will be required to compute and interpret P-values as well. So, this “critical regions” method should be considered a secondary method.

**8.3 Testing the Proportion** The calculator function that will be useful is **1-PropZTest**.

### 8.4 Paired Data

- (1) A discussion of dependent (paired) versus independent data will need to be had. We treat paired data quite differently than independent data.
- (2) A good guideline is the following: If you implement two distinct processes on **the same group** of individuals, then the data is paired.
- (3) When the data can be paired, the null hypothesis is always that the mean of the differences is 0 (there is no difference).
- (4) One will need to create a “new” data set that is the list of differences of the pairs—then use T-Test to complete the test. By typing  $L_1 - L_2$  in the label of the stats editor will generate a difference column.

### 8.5 Independent Populations

- (1) Here the null hypothesis is always that the difference of the means (or proportions) is 0 (there is no difference).
- (2) The **2-SampZTest**, **2-SampTTest**, and **2-PropZTest** are the key calculator functions.
- (3) Computations are easy with the calculator... reading and deciphering which function does the job for a specific problem is the tough part.

2.

## Worksheet: Testing the Mean

- (1) If the  $P$ -value in a statistical test is less than or equal to the level of significance for the test, do we reject or fail to reject  $H_0$ ? Does this imply that there IS or IS NOT enough evidence in the data (and the test being used) to justify the rejection of  $H_0$ ?
  
  
  
  
  
  
  
  
  
  
- (2) *Weatherwise* magazine is published in association with the American Meteorological Society. Volume 46, Number 6 has a rating system to classify Nor'easter storms that frequently hit New England states and can cause much damage near the ocean coast. A *severe* storm has an average peak wave height of 16.4 feet for waves hitting the shore. Suppose that a Nor'easter is in progress at the severe storm class rating.
  - (a) Let us say that we want to set up a statistical test to see if the wave action (i.e., height) is dying down or getting worse. What would be the null hypothesis regarding average wave height?
  - (b) If you wanted to test the hypothesis that the storm is getting worse, what would you use for the alternate hypothesis?
  - (c) If you wanted to test the hypothesis that the waves are dying down, what would you use for the alternate hypothesis?
  - (d) Suppose you do not know whether the storm is getting worse or dying out. You just want to test the hypothesis that the average wave height is *different* (either higher or lower) from the severe storm class rating. What would you use for the alternate hypothesis?
  - (e) For each of the tests in parts (b), (c), and (d), would the area corresponding to the  $P$ -value be on the left, on the right, or on both sides of the mean? Explain your answer in each case.

- (3) Gentle Ben is a Morgan horse at a Colorado dude ranch. The mean glucose level for horses should be  $\mu = 85$  mg/100 ml (Reference: *Merck Veterinary Manual*). Over the past 8 weeks, a veterinarian took weekly glucose readings from this horse (in mg/100 ml) and found the sample mean  $\bar{x} = 93.8$  mg/100 ml. Do the data indicate that Gentle Ben has an overall average glucose level higher than 85 mg/100 ml?
- (a) State the appropriate null and alternate hypothesis for this test. Is this a left-tailed, right-tailed, or two-tailed test?
  - (b) If we assume that  $x$  has a normal distribution and that we know from past experience that  $\sigma = 12.5$ , then the corresponding  $P$ -value is about 0.0232. Verify this is correct using the appropriate statistical test. At the  $\alpha = 0.05$  level, do these data indicate that Gentle Ben has an overall average glucose level higher than 85 mg/100 ml? Explain.
- (4) The price-to-earnings (P/E) ratio is an important tool in financial work. A recent copy of the *Wall Street Journal* indicated that the P/E ratio of the entire S&P 500 stock index is  $\mu = 19$ . A random sample of 14 large U.S. banks (J.P. Morgan, Bank of America, and others) had a sample mean of  $\bar{x} \approx 17.1$ . Do these data indicate that the P/E ratio of all U.S. bank stocks is less than 19?
- (a) State the appropriate null and alternate hypothesis for this test. Is this a left-tailed, right-tailed, or two-tailed test?
  - (b) If we assume that  $x$  has a normal distribution and that the sample standard deviation is  $s = 4.52$ , then the corresponding  $P$ -value is about 0.0699. Verify this is correct using the appropriate statistical test. At the  $\alpha = 0.05$  level, do these data indicate that the P/E ratio of all U.S. bank stocks is less than 19?

- (5) *USA Today* reported that the state with the longest mean life span is Hawaii, where the population mean life span is 77 years. A random sample of 20 obituary notices in the *Honolulu Advertiser* gave the following information about life span (in years) of Honolulu residents:

72, 68, 81, 93, 56, 19, 78, 94, 83, 84, 77, 69, 85, 97, 75, 71, 86, 47, 66, 27

Assuming that the life span in Honolulu is approximately normal distributed, does this information indicate that the population mean life span for Honolulu residents is less than 77 years? Use a 5% level of significance.

- (a) State the null and alternate hypothesis.
- (b) What sampling distribution should be used? Explain.
- (c) Is this a right-tailed, a left-tailed, or two-tailed test? Find the  $P$ -value.
- (d) Will you reject or fail to reject the null hypothesis? Explain and interpret this conclusion.

- (6) *Weatherize* is a magazine published by the American Meteorological Society. One issue gives a rating system used to classify Nor'easter storms that frequently hit New England and can cause much damage near the ocean. A severe storm has an average peak wave height of  $\mu = 16.4$  feet for waves hitting the shore. Suppose that a Nor'easter is in progress at the severe storm class rating. Peak wave heights are usually measured from land (using binoculars) off fixed cement piers. Suppose that a reading of 36 waves showed an average wave height of  $\bar{x} = 17.3$  feet. Previous studies of severe storms indicate that  $\sigma = 3.5$  feet. Does this information suggest that the storm is (perhaps temporarily) increasing above the severe rating? Use  $\alpha = 0.01$ . (Note that although this problem has not itemized out the steps, like the previous problems on this worksheet, a complete solution will include all such steps.)

3.

**Worksheet: Testing the Proportion**

- (1) Women athletes at the University of Colorado, Boulder, have a long-term graduation rate of 67% (Source: *Chronicle of Higher Education*). Over the past several years, a random sample of 38 women athletes at the school showed that 21 eventually graduated. Does this indicate that 67% is now an overestimate for the population proportion of women athletes who graduate from the University of Colorado, Boulder? Use a 5% level of significance.
- (a) State the null and alternate hypothesis.
  - (b) What sampling distribution should be used? Explain.
  - (c) Is this a right-tailed, a left-tailed, or two-tailed test? Find the  $P$ -value.
  - (d) Will you reject or fail to reject the null hypothesis? Explain and interpret this conclusion.
- (2) A large survey of countries, including the United States, China, Russia, France, Turkey, Kenya, and others, indicated that most people prefer the color blue. In fact, about 24% of the population claim blue as their favorite color. (Reference: Study by J.Bunge and A.Freeman-Gallant, Statistics Center, Cornell University). Suppose a random sample of  $n = 56$  college students were surveyed and  $r = 12$  of them said that blue is their favorite color. Does this information imply that the color preference of all college students is different (either way) from that of the general population? Use  $\alpha = 0.05$ .
- (a) State the null and alternate hypothesis.
  - (b) What sampling distribution should be used? Explain.
  - (c) Is this a right-tailed, a left-tailed, or two-tailed test? Find the  $P$ -value.
  - (d) Will you reject or fail to reject the null hypothesis? Explain and interpret this conclusion.

- (3) *Symposium* is part of a larger work referred to as Plato's *Dialogues*. Wishart and Leach (in *Computer Studies of Humanities and Verbal Behavior*, Vol. 3, pp. 90-99) found that about 21.4% of five-syllable sequences in *Symposium* of the type in which four are short and one is long. Suppose an antiques store in Athens has a very old manuscript that the owner claims is part of Plato's *Dialogues*. A random sample of 493 five-syllable sequences from this manuscript showed that 136 were of the type four short and one long. Do the data indicate that the population proportion of this type of five-syllable sequence is higher than that found in Plato's *Symposium*? Use  $\alpha = 1\%$ . (Make sure that you are showing all relevant steps to the hypothesis test.)
- (4) Suppose a hypothesis test is executed and the  $P$ -value is found.
- (a) If the  $P$ -value is such that you can reject  $H_0$  at the 1% level of significance, can you always reject  $H_0$  at the 5% level of significance too? Explain.
  - (b) If the  $P$ -value is such that you can reject  $H_0$  at the 5% level of significance, can you always reject  $H_0$  at the 1% level of significance too? Explain.

4.

**Worksheet: Paired Data**

On each of the following problems, be sure have an answer to each of the following:

- State the null and alternate hypothesis.
  - Determine if it is a left-tailed, a right-tailed, or a two-tailed test.
  - Determine which sampling distribution should be used and explain why.
  - Determine the  $P$ -value for the test.
  - Draw a conclusion on whether you will reject or fail to reject the null hypothesis.
  - Interpret and explain what that conclusion tells you within the context of that data.
- (1) For a random sample of 20 data pairs, the sample mean of the difference was 2. The sample standard deviation of the differences was 5. Assume that the distribution of the differences is (approximately) normally distributed. At the 5% level of significance, test the claim that the population mean of the differences is positive.
- (a) Why is it appropriate to use a Students'  $t$  distribution for the sample test statistic? What degrees of freedom are used?
- (b) State the null and alternate hypotheses for this test.
- (c) Compute the  $P$ -value.
- (d) Does the  $P$ -value indicate that we reject or fail to reject the null hypothesis? Explain and interpret what the result then tells you.



- (2) A random sample of nine banks shows their deposits (in billions of dollars) 3 years ago and their deposits (also in billions) today. At the  $\alpha = 0.05$  significance level, can we support claim that the average deposit for the banks is greater today than it was 3 years ago?

Bank	1	2	3	4	5	6	7	8	9
3 years ago	11.42	8.41	3.98	7.37	2.28	1.10	1.00	0.90	1.35
Today	11.69	9.44	6.53	5.58	2.92	1.88	1.78	1.50	1.22

- (a) What assumption(s) will we need to make to use a paired  $t$ -Test?
- (b) State the null and alternate hypotheses for this test. Be sure to state how  $\mu_d$  is defined.
- (c) Type the data from 3 years ago into the  $L_1$  column and the data from today into  $L_2$ . By selecting  $L_3$  and typing the formula  $L_3 = L_1 - L_2$ , what are the 9 differences from each data pair?
- (d) Compute the test statistic  $t$  and the corresponding  $P$ -value.
- (e) What conclusion do you make based on this hypothesis test?

- (3) Is fishing better from a boat or from the shore? Pyramid Lake is located on the Paiute Indian Reservation in Nevada. Presidents, movies stars, and people who just want to catch fish go to Pyramid Lake for really large cutthroat trout. Let row  $B$  represent hours per fish caught fishing from the shore, and let row  $A$  represent hours per fish caught using a boat. The following data are paired by month from October through April (Source: *Pyramid Lake Fisheries*, Paiute Reservation, Nevada). Use a 1% level of significance to test if there is a difference in the population mean hours per fish caught using a boat compared with fishing from the shore.

	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.
Shore	1.6	1.8	2.0	3.2	3.9	3.6	3.3
Boat	1.5	1.4	1.6	2.2	3.3	3.0	3.8

- (4) A veterinary nutritionist developed a diet for overweight dogs. The total volume of food consumed remains the same, but one-half of the dog food is replaced with a low-calorie “filler”, such as canned green beans. Six overweight dogs were randomly selected from her practice and were put on this program. Their initial weights were recorded, and they were weighed again 4 weeks later. At 0.01 level of significance, can it concluded that the diet was effective?

Before	42	53	48	65	40	52
After	39	45	40	58	42	47

5.

**Worksheet: Independent Populations**

On each of the following problems,

- State the null and alternate hypothesis.
- Determine if it is a left-tailed, a right-tailed, or a two-tailed test.
- Determine which sampling distribution should be used and explain why.
- Determine the  $P$ -value for the test.
- Draw a conclusion on whether you will reject or fail to reject the null hypothesis.
- Interpret and explain what that conclusion tells you within the context of that data.

Note that you will first need to determine whether you are looking at paired data or independent samples.

- (1) Based on information from *The Denver Post*, a random sample of  $n_1 = 12$  winter days in Denver gave a sample mean pollution index of  $\bar{x}_1 = 43$ . Previous studies show that  $\sigma_1 = 21$ . For Englewood (a suburb of Denver), a random sample of  $n_2 = 14$  winter days gave a sample mean pollution index of  $\bar{x}_2 = 36$ . Previous studies show that  $\sigma_2 = 15$ . Assume the pollution index is normally distributed in both Englewood and Denver. Do these data indicate that the mean population pollution index of Englewood is different (either way) from that of Denver in the winter? Use a 1% level of significance.

- (2) A study of fox rabies in southern Germany gave the following information about different regions and the occurrence of rabies in each region (B. Sayers et al., "A Pattern Analysis Study of a Wildlife Rabies Epizootic," *Medical Informatics*, Vol. 2, pp. 11-34). Based on information from this article, a random sample of  $n_1 = 16$  locations in region I gave the following information about the number of cases of fox rabies near that location.

$x_1$ : **Region I Data**      1 8 8 8 7 8 8 1 3 3 3 2 5 1 4 6

A second random sample of  $n_2 = 15$  locations in region II gave the following information about the number of cases of fox rabies near that location.

$x_2$ : **Region II Data**      1 1 3 1 4 8 5 4 4 4 2 2 5 6 9

Does this information indicate that there is a difference (either way) in the mean number of cases of fox rabies between the two regions? Use a 5% level of significance. (Assume the distribution of rabies cases in both regions is mound-shaped and approximately normal.)

- (a) Find the values of  $\bar{x}_1$  and  $s_1$  for Region I and the values of  $\bar{x}_2$  and  $s_2$  for Region II.

- (b) State the null and alternate hypotheses.

- (c) What sampling distribution should be used, and why?

- (d) Determine the  $P$ -value.

- (e) According to your result, will you reject or fail to reject the null hypothesis? Explain and interpret what the result then tells you.

- (3) The following data are based on information from the Regis University Psychology Department. In an effort to determine if rats perform certain tasks more quickly if offered large rewards, the following experiment was performed:

–On day 1, a group of three rats was given a reward of one food pellet each time they ran a maze. A second group of rats was given five food pellets each time they ran the maze.

–On day 2, the groups were reversed, so the first group now got five food pellets for running the maze and the second group got only one pellet for running the same maze.

The average time in seconds for each rat to run the maze 30 times are shown in the following table. Do these times indicate that rats receiving larger rewards tend to run the maze in less time? Use a 5% level of significance.

Rat	A	B	C	D	E	F
Time with one food pellet	3.6	4.2	2.9	3.1	3.5	3.9
Time with five food pellets	3.0	3.7	3.0	3.3	2.8	3.0

- (4) Generally speaking, would you say that most people can be trusted? A random sample of  $n_1 = 250$  people in Chicago ages 18–25 showed that  $r_1 = 45$  said yes. Another random sample of  $n_2 = 280$  people in Chicago ages 35–45 showed that  $r_2 = 71$  said yes (based on information from the *National Opinion Research Center*, University of Chicago). Does this indicate that the population proportion of trusting people in Chicago is higher for the older group? Use  $\alpha = 0.05$ .

- (5) In her book *Red Ink Behaviors*, Jean Hollands reports on leading Silicon Valley companies. An “intimidator” is an employee who tries to stop communication, sometimes sabotages others, and, above all, likes to listen to him- or herself talk. Let  $x_1$  be a random variable representing productive hours per week lost by peer employees of an intimidator. A “stressor” is an employee with a hot temper that leads to unproductive tantrums in corporate society. Let  $x_2$  be a random variable representing productive hours per week lost by peer employees of a stressor. Assume that population distributions of time lost due to intimidators and time lost due to stressors are each (approximately) normally distributed. Assuming that the variables  $x_1$  and  $x_2$  are independent, do the data indicate that the population mean time lost due to stressors is greater than the population mean time lost due to intimidators? Use a 5% level of significance.

$x_1$ (intimidator)	8	3	6	2	2	5	2	7
$x_2$ (stressor)	3	3	10	7	6	2	5	8

- (6) For a random sample of 20 data pairs, the sample mean of the difference was 2. The sample standard deviation of the differences was 5. Assume that the distribution of the differences is (approximately) normally distributed. At the 1% level of significance, test the claim that the population mean of the differences is positive.
- Why is it appropriate to use a Students'  $t$  distribution for the sample test statistic? What degrees of freedom are used?
  - State the null and alternate hypotheses for this test.
  - Compute the  $P$ -value.
  - Does the  $P$ -value indicate that we reject or fail to reject the null hypothesis? Explain and interpret what the result then tells you.
- (7) REM (rapid eye movement) sleep is sleep during which most dreams occur. Each night a person has both REM and non-REM sleep. However, it is thought that children have more REM sleep than adults (*Secrets of Sleep*, Dr. A. Borbely). Assume that REM sleep time is normally distributed for both children and adults. A random sample of  $n_1 = 10$  children (9 years old) showed that they had an average REM sleep time of  $\bar{x}_1 = 2.8$  hours per night. From previous studies, it is known that  $\sigma_1 = 0.5$  hour. Another random sample of  $n_2 = 10$  adults showed that they had an average REM sleep time of  $\bar{x}_2 = 2.1$  hours per night. Previous studies show that  $\sigma_2 = 0.7$  hour. Do these data indicate that, on average, children tend to have more REM sleep than adults? Use a 1% level of significance.
- State the null and alternate hypotheses.
  - What sampling distribution should be used, and why?
  - Determine the  $P$ -value.
  - According to your result, will you reject or fail to reject the null hypothesis? Explain and interpret what the result then tells you.