Section II. Chapter by Chapter Advice

Objectives for this Chapter
Ideas for Teaching this Chapter
Let's Do It! (LDI) Solutions
Think About It (TAI) Solutions
Extra Examples
Projects

Chapter 1 – How to Make a Decision with Statistics

OBJECTIVES FOR THIS CHAPTER

- Introduce the scientific method and the role of statistics in decision making.
- Distinguish between a population and a sample.
- Learn how to formulate and write the null and alternative hypotheses.
- Understand what it means to say data are statistically significant.
- Learn about the two types of error in a statistical decision problem—Type I error and Type II error.
- Understand the meaning of the significance level, *p*-value, and power of a test.
- Discover the effect of the sample size in a decision-making process.
- Understand the difference between *significant* versus *important.*

IDEAS FOR TEACHING THIS CHAPTER

First recall the ideas presented in the **Day 1** section of this manual. Our first lecture involves the randomized seating ice-breaker, details on the class syllabus, either a short video on an overview of statistics or a general discussion of theories, data, and decisions using a couple of distributed articles.

Day 2 might begin with a recap of the ideas from Day 1 along with any announcements (need a syllabus?, labs sections start next week, etc.). We might start with an overview of pages 1 through 6, do the Let's Do It! 1.1 together as a class, and have students work with a partner to do Let's Do It! 1.2. We then demonstrate the ideas from the Package of Balls Example 1.2. We bring a closed package (bag, basket, or box) to class which contains a total of 5 yellow balls. We tell the students that we just bought this package and we were told that it contains 1 yellow ball and 4 blue balls. We ask the students to help assess if this is the case. We then set up our competing theories, explaining the statistical terminology as we do it. We then proceed to mix up the package well and gather some data, one ball at a time (with replacement). We go around the class and ask 5 different students (one at a time) to select a ball from the package (without looking!), note its color and then replace it. Each time we ask them how they feel about the result. Nearly every semester, by the time we get to the 3rd student, and the 3rd yellow ball, students are beginning to laugh. This is our opportunity to talk about the meaning of rejecting and failing to reject the null hypothesis, the idea of outcomes being possible versus likely. We continue with pages 9 through 15 on the concepts of statistical significance and the types of errors.

The central idea for Days 3 and 4 is the **Bag A and Bag B scenario** in Section 1.4. A visual display of an actual bag is not necessary for some, but helpful to others, and fun for all (and easy to make up one on your own). We bring the bag to Day 2 and may introduce the scenario at the end of that day. The bag is a **Bag A**, but this is not told to the students. Our task is to decide which bag it really is. We often will start by looking at pages 16 and 17 -- stating the hypotheses and examining the frequency plots. With the bag in hand, we can get the students involved and gather data. Ask a student to select a voucher (after mixing the bag, and not looking). Share with the class the observed voucher value. Ask the class (or student) what they would decide based on the observed value and why. You might do this a couple times (replacing the selected voucher each time first) -- to get a couple of different voucher values. This can lead into the discussion about extreme values, direction of extreme and decision rules on pages 18 through 23. We try to end Day 3 with pages 23 through 28 which cover the other directions of extreme. Day 4 is reserved primarily for the topic of *p*-value, pages 29-47. The pictures on page 31 and 32 really help students understand the idea of rejecting the null hypothesis if the *p*-value is very small, in particular, less than or equal to α. We introduce the idea of the power of the test and the power function along with the *p*-value, α, and β. On Day 5 we would cover practical significance and the optional $n = 2$ Bag A and Bag B scenario in Section 1.5.

Let's Do It (LDI) Solutions

Let's Do It! 1.1 Fair Die?

If the die is "fair", the true proportion of 5's or 6's should be 1/3. However, a close examination of a real die reveals that the "pips" are made by small indentations into the face of the die. Sides 5 and 6 have more indentations than the other faces, and so these sides should be slightly lighter than the other faces, which suggests that the true proportions of 5's or 6's may be a bit higher than the "fair" value 1/3. State the appropriate null and alternative hypotheses for assessing if the data provide compelling evidence for the competing theory.

H_0: The die is fair, that is, the indentations have no effect, and the proportion of 5's or 6's is **equal to 1/3**
H_1: The die is not fair, that is, the indentations have an effect, and the proportion of 5's or 6's is **greater than 1/3**

Let's Do It! 1.2 Stress Can Cause Sneezes

Studies suggest that stress doubles a person's risk of getting a cold. Acute stress, lasting maybe only a few minutes, can lead to colds. On average, up to 90% of people exposed to a cold virus become infected, meaning the virus multiplies in the body, but only 40% actually become sick. One researcher thinks that the accumulation of stress tips the infected person over into illness. The percentage of people exposed to a cold virus who actually get a cold is 40%. The researcher would like to assess if stress *increases* this percentage. So, the population of interest is people who are under (acute) stress. State the appropriate hypotheses for assessing the researcher's theory regarding this population.

H_0: The percentage of people exposed to a cold virus and have stress who actually get a cold is still just 40%.

H_1: The percentage of people exposed to a cold virus and have stress who actually get a cold is greater than 40%.

LDI 1.1 and 1.2

How long?	2-3 minutes each
How might it be done?	Ask students to read through the scenario (or you paraphrase it to the class), complete the hypotheses, and compare with a neighbor.
How important?	We recommend you do at least one of these two and/or bring some in from recent news to share with the class.

Let's Do It! 1.3 Complaints about Chips

The chain decided to test the following hypotheses concerning the true average weight (in ounces) of a bag of such potato chips in the next shipment received from their supplier:

H_0: The average weight is at least 16 ounces. H_1: The average weight is less than 16 ounces.

If there is evidence in favor of the alternative hypothesis, the shipment would be refused and a complaint registered with the supplier. Some bags of chips were selected from the next shipment and the weight of each selected bag was measured. The researcher for the supermarket chain stated that the data were **statistically significant**.

What hypothesis was rejected? **H_0 because statistically significant implies the null hypothesis was rejected.**
Was a complaint registered with the supplier? **Yes.**
Could there have been a mistake? If so, describe it. **Yes, there could have been a mistake. Based on a few bags it was concluded that the average weight of bags in the shipment is less than 16 ounces. However, it is possible that the shipment is fine, we just happened to select a few bags that were on the light side.**

LDI 1.3

How long?	4 – 5 minutes
How might it be done?	Ask students to read through the scenario (or you paraphrase it to the class), complete the questions and discuss with a neighbor.
How important?	Important. This exercise addresses the phrase 'statistically significant' and plants the idea that mistakes can be made in decision-making.

Let's Do It! 1.4 Did Everyone Have an Equal Chance of Surviving?

The *Titanic*, the largest ship that had ever been built up to that time, left Southampton, England toward New York on Wednesday, April 10, 1912. It was carrying many rich and famous people of both England and the United States. Many thought that the *Titanic* could not sink. It made a stop at Cherbourg, in France, where it took on many third-class passengers, and a brief stop at Queenstown, Ireland, and set its course across the Atlantic. On the evening of April 14 it struck an iceberg and in less than hours, by 2:15 A.M. of the next morning, it sank. Only 710 survived among the 2201 passengers and crew. A question of interest may be whether or not everybody had an equal chance of surviving. In the language of the testing our hypotheses might be:

H_0: Men and women had the same chance of surviving.

H_1: Men and women did not have the same chance of surviving.

Let p_M represent the proportion of all men that survived and p_W represent the proportion of all women that survived.

(a) Using the notation for proportions given above, rewrite H_0 and H_1.

H_0: $p_M = p_W$

H_1: $p_M \neq p_W$

(b) Suppose the data obtained were statistically significant. What hypothesis was accepted? H_1

LDI 1.4

How long?	2 minutes
How might it be done?	Ask students to read through the scenario (or you paraphrase it to the class), complete the questions and discuss with a neighbor.
How important?	Important. This exercise addresses the problem of using symbols instead of words, sometimes difficult for students. We also revisit the idea of statistical significance.

Let's Do It! 1.5 Which Error Is Worse?

For each set of hypotheses given below decide whether a Type I error or a Type II error would be more serious. Recall that a Type I error occurs if you reject the null hypothesis H_0 when it is true. A Type II error occurs if you do not reject the null hypothesis H_0 when it is false. Answers may vary, but be prepared to discuss your answer.

(a) H_0: The water is contaminated.

H_1: The water is not contaminated.

Since a Type I error is rejecting H_0 when H_0 is true, we would conclude the water is not contaminated when it is contaminated. Since a Type II error is failing to reject H_0 when H_1 is true, we would think the water is contaminated when it is not contaminated. A <u>Type I error</u> may be more serious as one might drink contaminated water and become ill.

(b) H_0: The parachute works.

H_1: The parachute does not work.

Since a Type I error is rejecting H_0 when H_0 is true, we would conclude the parachute does not work when it does work. Since a Type II error is failing to reject H_0 when H_1 is true, we would think the parachute does work when it does not work. A <u>Type II error</u> may be more serious as we might try to use a parachute that is faulty and may not open when needed.

(c) H_0: The ship is unsinkable.

H_1: The ship is sinkable.

Since a Type I error is rejecting H_0 when H_0 is true, we would conclude the ship is sinkable when it is unsinkable. Since a Type II error is failing to reject H_0 when H_1 is true, we would think the ship is unsinkable when it is sinkable. A <u>Type II error</u> may be more serious as we might take a ride in a ship that may sink.

(d) H_0: The shirt is not washable.
H_1: The shirt is washable.
Since a Type I error is rejecting H_0 when H_0 is true, we would conclude the shirt is washable when it is not. Since a Type II error is failing to reject H_0 when H_1 is true, we would think the shirt is not washable when it is washable. A <u>Type I error</u> may be more serious as we might ruin our shirt when we wash it.

(e) H_0: The value of my investment stock portfolio is going to increase over the next 6 months.
H_1: The value of my investment stock portfolio is not going to increase over the next 6 months.
Since a Type I error is rejecting H_0 when H_0 is true, we would conclude the value of my investment stock portfolio is not going to increase over the next 6 months when it would. I might sell off my investments too early, losing money I could have received if I had waited. Since a Type II error is failing to reject H_0 when H_1 is true, we would conclude the value of my investment stock portfolio is going to increase over the next 6 months when it does not. I would hold on to investments too long, losing money I could have received if I had sold sooner. There is <u>not one clear error</u> that is more serious. It would depend on other things such as how much the value goes up or down, and the cost when I purchased the stock initially.

LDI 1.5

How long?	5 – 7 minutes
How might it be done?	Go through part (a) as a class, recalling the definition of the two types of errors and showing how to state each error in the context of the scenario. Then have students work in groups of 2 to 4 to discuss parts (b) to (e). If time permits to only cover a few parts, that is often sufficient. Exercises 1.5 and 1.6 are similar problems which can be assigned as homework.
How important?	Somewhat Important. It helps students to learn to convert a standard term into the context/language of the research question at hand. This is helpful when students are later asked to not only state the decision, but to also give a conclusion (what the decision means) in the context of the research question at hand.

Let's Do It! 1.6 Testing a New Drug

In the previous section, two drugs were being compared, a new drug and a standard drug, to treat a particular disease. The hypotheses were as follows:

H_0: The new drug is as effective as the standard drug.
H_1: The new drug is more effective than the standard drug.

A study is conducted in which the investigator administers the new drug to some number of patients suffering from the disease and the standard drug to another group of patients suffering from the disease. The proportion of cures for both drugs is recorded. What are the two types of errors that you could make when deciding between these two hypotheses?

A Type I error would be: **Conclude that the new drug is better than the standard drug, when it is not.**
A Type II error would be: **Conclude that the new drug is no better than the standard drug, when it is actually better.**

What are the consequences of a Type I error? **The standard drug might be discontinued. The new drug might be prescribed when it could be no better or even worse for some patients.**
What are the consequences of a Type II error? **The new drug might not be made available on the market and thus not prescribed when it could actually be beneficial to some patients.**

Which error might be considered more severe from an ethical point of view? **There are arguments for each error to be considered more severe. The Food and Drug Administration has established guidelines to follow before a drug is made available for use.**

To know the true proportion of patients suffering from the disease that would be cured using the new drug, we would need to administer the new drug to all such patients. However, this is not possible. Why not?
It is not possible to administer the new drug to all patients due to money and time constraints, and the consequences involved if the new drug turns out to be less effective than the standard drug.

LDI 1.6

How long? 7-10 minutes

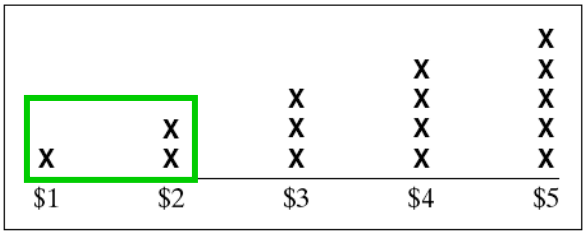
How might it be done? Ask students to read through the scenario (or you paraphrase it to the class), complete the questions and discuss with a neighbor.

How important? Very important. This exercise helps students connect the statistical concepts of Type I and Type II error to the context of a real-life scenario. You might discuss the idea that the null hypothesis is actually the complement of the alternative, namely that the new drug is not more effective than the standard drug. However, if the data supports the alternative because it is unlikely assuming the two drugs are equally effective, than it will be unlikely assuming the new drug is less effective than the standard drug. The idea more simply: if the alternative hypothesis is that the average is more than 10, and if the data are strong enough to support it, then the data will also be strong enough to support that the mean is larger than any number less than 10.

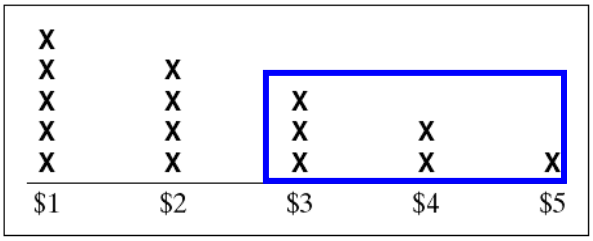
Let's Do It! 1.7 Enlarging the Rejection Region

Recall the **Bag C** and **Bag D** scenario of the previous example.

Frequency Plot of Bag C



Frequency Plot of Bag D



H_0: The shown bag is ***Bag C***.

H_1: The shown bag is ***Bag D***.

Decision rule 1 had a significance level α = 0.067 and a rather large level for β = 0.667.

Enlarge the rejection region and give a new decision rule which will have a smaller level for β.
To do this, you need to find the next most extreme value among the values 2, 3, 4, and 5. What is it? **_2_**

Decision rule 2: Reject H_0 if your selected voucher is **$2 or less**.
The rejection region for Decision Rule 2 is **the set of voucher values less than or equal to $2, namely, $1, and $2**.

The values for α and β corresponding to decision rule 2 are:
α = 3/15 = 0.20 **β = 6/15 = 0.40**
How did these values compare to those for Decision Rule 1? **The level α increased, while the level β decreased.**

LDI 1.7

How long? 7-10 minutes

How might it be done? To be worked on by students after you discuss (or have them read) Example 1.5. Ask students to work with a partner.

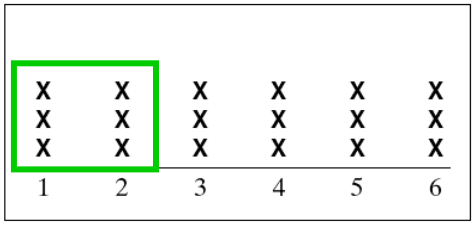
How important? Important. This exercise reinforces a different direction of extreme than the original Bag A and Bag B. It also provides another example of the trade-off between the chances of committing the two types of mistakes -- enlarging the rejection region will increase the level for α and decrease the level for β.

Let's Do It! 1.8 The Case of the Asymmetric Hypotheses

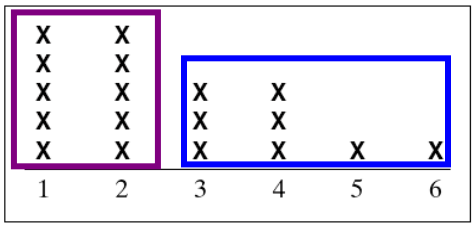
The frequency plots *A* and *B* below provide the models for a fair die and a die loaded in favor of outcomes 1 and 2.For example, the chance of obtaining a "2" when you roll a fair die once is 3/18 and the chance of obtaining a "2" when you roll the loaded die once is 5/18.
Suppose you have a die and you do not know if it is the fair die or the loaded die. You will be allowed to roll the die once and based on the outcome of your roll you have to decide between the following hypotheses:

H_0: The die is fair.
H_1: The die is loaded (in favor of 1 and 2).

A: Fair Die



B: Loaded Die



(a) The direction of extreme for this test is
(circle one)
one-sided to the right
one-sided to the left
two-sided
can't tell

(b) Consider the following decision rule:
Reject *H*0 if you roll a 2 or "*more extreme*" value.
Clearly indicate on the graphs and compute the values of α, β and the power.
α = 6/18.
β = 8/18.
power = 10/18

(c) Suppose the die is rolled and the result was statistically significant at the significance level α. Then the decision made was (circle one)
The die is fair.
The die is loaded (in favor of 1 and 2).

LDI 1.8

How long? 5 minutes
How might it be done? Ask students to work with a partner.
How important? Important. This exercise reinforces a different direction of extreme than the original Bag A and Bag B. It also provides another example of the trade-off between the chances of committing the two types of mistakes – and the idea that $\alpha + \beta \neq 1$.

Let's Do It! 1.9 Chromium Supplements

Sixteen men were monitored in the study. Half got the chromium supplements and the other half got a fake substitute. At the end of the training program, both groups had become stronger, but there was no statistically meaningful difference in strength gains, the study reported. Consider the following hypotheses:

H_0: No difference in strength gains between the two groups
H_1: Chromium group had a higher strength gain

Based on the stated results, which hypothesis was supported by the data? $\boldsymbol{H_0}$

Would the *p*-value for testing the above hypotheses have been somewhat small? Explain.
No, since the null hypothesis was supported, the *p*-value must not have been very small. A *p*-value that is not small indicates that the observed data, or data more extreme, are somewhat likely if the null hypothesis H_0 is true. In particular, the *p*-value must have been larger than α.

LDI 1.9

How long? 5-7 minutes

How might it be done? Emphasize the definition of a *p*-value and the reasoning behind a smaller *p*-value provides more evidence against the null hypothesis. Then ask students to read through the scenario (or you paraphrase it to the class), complete the questions and discuss with a neighbor.

How important? Important. This exercise helps students connect the decision made as stated in an article to the statistical measure used to make that decision, the *p*-value. It is a nice example in that the decision was to stay with the null hypothesis of no difference. We often see newspapers focus on reporting studies in which the new theory was supported.

Let's Do It! 1.10 Three Studies

The following table summarizes the hypotheses and results for three different studies.

	Null Hypothesis	*Alternative Hypothesis*	*p-value*
Study A	The true average lifetime is $\geq$ 54 months.	The true average lifetime is $<$ 54 months.	0.0251
Study B	The average time to relief for all Treatment I users is equal to the average time to relief for all Treatment II users.	The average time to relief for all Treatment I users is not equal to the average time to relief for all Treatment II users.	0.0018
Study C	The true proportion of adults who work two jobs is ≤ 0.33.	The true proportion of adults who work two jobs is > 0.33.	0.3590

(a) For which study do the results show the most support for the null hypothesis? Explain.
Study C, since a smaller *p*-value represents stronger evidence for the alternative hypothesis and Study C reported the largest *p*-value.

(b) Suppose Study A concluded that the data supported the alternative hypothesis that the true average lifetime is less than 54 months, but in fact the true average lifetime is greater than or equal to 54 months. In our statistical language, would this be called a Type I error or a Type II error?
Type I error, rejecting the null hypothesis when it is true.

(c) If the results of Study A are "statistically significant," which hypothesis is supported? $\boldsymbol{H_1}$

(d) For each study, determine if the rejection region would have been one-sided to the right, one-sided to the left, or two-sided. (Circle your answer.)

Study A: one-sided to the right **one-sided to the left** two-sided

Study B: one-sided to the right one-sided to the left **two-sided**

Study C: **one-sided to the right** one-sided to the left two-sided

LDI 1.10

How long? 5-7 minutes

How might it be done? Ask students to read through the scenario (or you paraphrase it to the class), complete the questions and discuss with a neighbor.

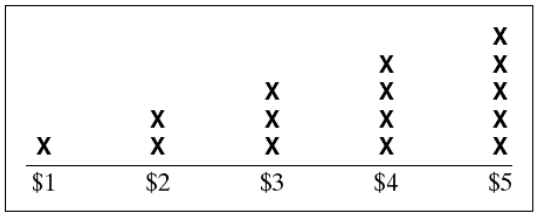
How important? Important. This exercise helps students to view the *p*-value as a measure of the weight of the evidence based on the data. It also revisits errors, the meaning of statistically significance, and the various directions of extreme. All this can be done before you actually start computing a *p*-value (which comes next).

Let's Do It! 1.11 *p*-value for a One-Sided Rejection Region to the Left

Consider again the two identical bags, Bag C and Bag D, from Example 1.5. You are allowed to select just one voucher from the shown bag and must decide whether or not to reject H_0.

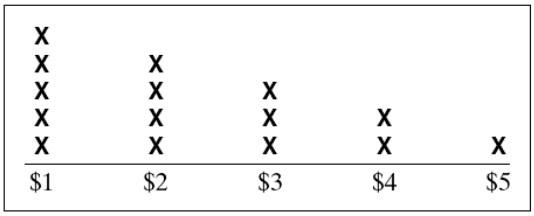
H_0: The shown bag is ***Bag C***.

Frequency Plot of Bag C



H_1: The shown bag is ***Bag D***.

Frequency Plot of Bag D



(a) Suppose now that the observed voucher value is \$2. Find the corresponding *p*-value.
The *p*-value is the chance of getting \$2 or less from Bag C, which is 3/15 = 0.20.

For the following significance levels, are the data statistically significant?

Significance level α	*Circle one*	
0.10	Yes	**No**
0.05	Yes	**No**
0.01	Yes	**No**

(b) Suppose now that the observed voucher value is \$1. Find the corresponding p-value.
The *p*-value is the chance of getting \$1 or less from Bag C, which is 1/15 = 0.067.

For the following significance levels, are the data statistically significant?

Significance level α	*Circle one*	
0.10	**Yes**	No
0.05	Yes	**No**
0.01	Yes	**No**

LDI 1.11

How long? 4-5 minutes

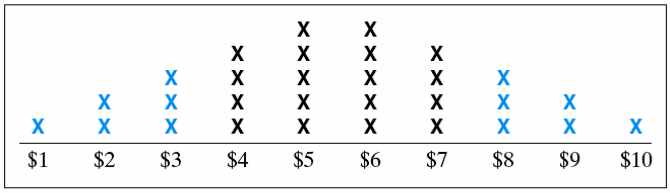
How might it be done? To be worked on by students after you discuss (or have them read) Example 1.7. Ask students to work with a partner.

How important? Important. This exercise reinforces a different direction of extreme than the original Bag A and Bag B. It also provides more examples of computing the *p*-value based on the observed voucher value and comparing it to a significance level α for assessing statistical significance.

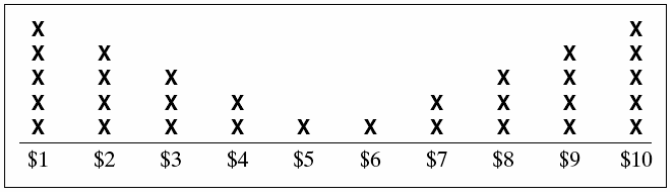
Let's Do It! 1.12 *p*-value for a Two-Sided Rejection Region

Consider again the two identical bags, Bag E and Bag F, from Example 1.6. You are allowed to select just one voucher from the shown bag and must decide whether or not to reject H_0.
H_0: The shown bag is ***Bag E***.
H_1: The shown bag is ***Bag F***.

Frequency Plot of Bag E



Frequency Plot of Bag F



(a) Suppose the observed voucher value is \$2. Find the corresponding *p*-value.
The *p*-value is the chance of getting \$2 or less plus the chance of getting \$8 or more from Bag E, which is 6/30 = 0.20.

For the following significance levels, are the data statistically significant?

Significance level α	*Circle one*	
0.10	Yes	**No**
0.05	Yes	**No**
0.01	Yes	**No**

(b) Suppose now that the observed voucher value is \$10. Find the corresponding *p*-value.
The *p*-value is the chance of getting \$10 or more plus the chance of getting \$1 or less from Bag E, which is 2/30 = 0.067.

For the following significance levels, are the data statistically significant?

Significance level α	*Circle one*	
0.10	**Yes**	No
0.05	Yes	**No**
0.01	Yes	**No**

LDI 1.12

How long? 4-5 minutes

How might it be done? To be worked on by students after you discuss (or have them read) Example 1.8. Ask students to work with a partner.

How important? Important. This exercise reinforces the two-sided direction of extreme. It follows very nicely from the previous Let's do it exercise. It also provides more examples of computing the *p*-value based on the observed voucher value and comparing it to a significance level α for assessing statistical significance.

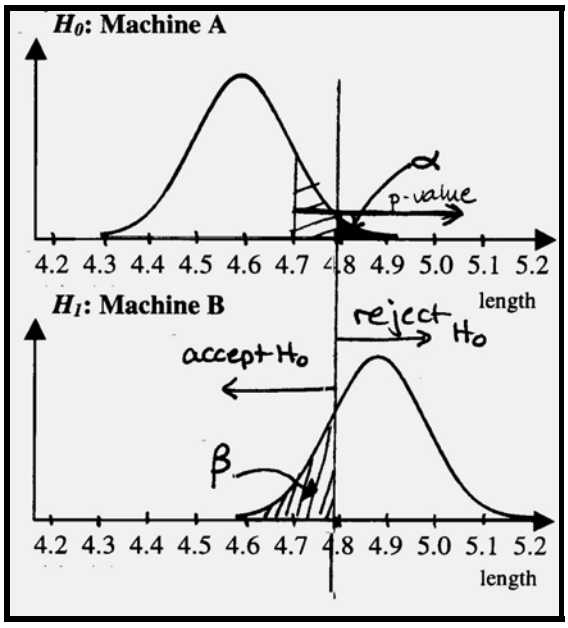
Let's Do It! 1.13 Machine A or Machine B?

Machine A makes parts whose lengths average around 4.6 mm. Machine B makes parts whose lengths vary similarly but around a higher average of 4.9 mm. Suppose you have a box of parts that you believe are from Machine A, but you're not sure. You will test the hypotheses by randomly selection one part from the box and measuring it.

H_0: The parts are from Machine A.

H_1: The parts are from Machine B.

The hypotheses for the distribution of length of part (in mm) are presented at the right as the smoothed versions of the frequency plots.



(a). The direction of extreme is: (circle your answer)
one-sided to the right
one-sided to the left
two-sided

(b) Consider the decision rule: Reject H_0 if the selected part length is 4.8 mm or more extreme.
(i) In the picture shade in the region that corresponds to α, the significance level and clearly label the region with α.
(ii) In the picture shade in the region that corresponds to β, the chance of a Type II error and clearly label the region with β.

(c) The selected part length is 4.7 mm. In the picture shade in the region that corresponds to the *p*-value and clearly label the region with *p*-value.

(d). Is the selected result in part (c) statistically significant? (circle your answer) Yes **No**
Explain your answer. **The *p*-value is greater than α.**

LDI 1.13

How long? 4 – 5 minutes

How might it be done? As much of this chapter is new to students, it can be helpful to talk through the scenario together as a class first. Perhaps coming to consensus for part (a) that the direction of extreme is to the right. You might ask students: "What values would show more support for the alternative model (Machine B) and less support for the null model (Machine A)? Values that are too large?, too small?, both directions? You could draw in the "line" that marks the cut-off value in the decision rule and draw the arrows for the rejection and acceptance regions. Then have them work with a partner and the definitions of a, b, and *p*-value to do parts (b), (c) , and (d).

How important? Important. Here is another example where students can "see" the various components of the decision making process, without having to compute the quantities (no math/formulae, just definitions and concepts).

Let's Do It! 1.14 Finding a 5% Level Decision Rule

Suppose the specified significance level is $\alpha = 0.05$. Decision Rules 1 and 2 both resulted in actual significance levels that were less than 0.05. Continue to enlarge the rejection region to find a decision rule with a 5% significance level.

The decision rule will be: Reject H_0 if the average of the two vouchers is **$45 or more**

This rule results in a significance level of $\alpha = {(1+2+4)}/{190} = 0.0368$

The corresponding level for $\beta = {(20+17+9+4+2+1)}/{190} = {53}/{190} = 0.2789$

Note: Using $40 as the cut-off value would yield a significance level of $\alpha = {(1+2+4+9)}/{190} = 0.0842$ which exceeds the desired level of 0.05.

LDI 1.14

How long?	5-7 minutes
How might it be done?	Ask students to work with a partner. The small table located on page 46 can be used to do the counting for determining the levels. When you gather the class back together, be sure everyone has stated the same rule and understand why enlarging the rejection region to include the $40 voucher will not work.
How important?	Important. This exercise helps students to see the connection between a pre-determined fixed level of significance and the decision rule corresponding to that level.

Think About It (TAI) Solutions

Page 14:

Think about it

If the consequences of making a Type I error are considered very serious, why not set the chance of making a Type I error to zero?

To achieve a value of zero for a Type I error, you would never reject the null hypothesis. Your decision rule would be always fail to reject H_0, and you would never support any new, alternative theory. We have to be willing to accept some small chance of making an error.

Page 17:

Think about it

What if the voucher you select is $60? Would this observation lead you to think the shown bag is **Bag A** or **Bag B**?

Bag B

Why? **Although the $60 voucher value is possible from both bags, it is more likely to occur from Bag B (7/20 chance) than from Bag A (1/20 chance).**

How would you answer these questions if the voucher you select is $10?

We might think that the shown bag is Bag A, because a $10 is more likely to occur from Bag A (7/20 chance) than from Bag A (1/20 chance).

Page 33:

Think about it

The significance level is α =0.10, the chance of committing a Type I error. The corresponding decision rule is: Reject H_0 if the selected voucher is \$50 or more. A voucher is selected and it turns out to be \$60. Your decision is to reject the null hypothesis and conclude that the data are statistically significant at the 10% level.

You rejected H_0. Could you have made a mistake? circle one **Yes** No

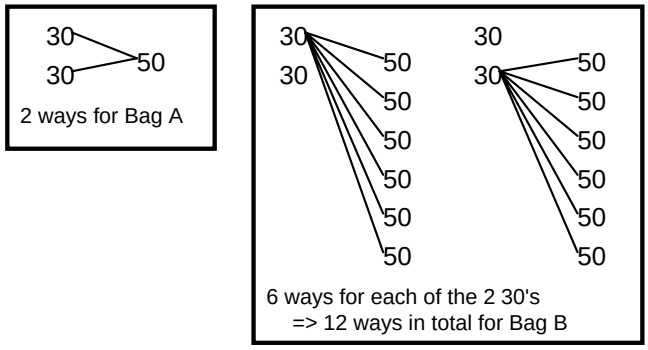
What type of mistake could you have made? circle one **Type I error** Type II error

What is the chance that you have made a mistake? **Many students might initially say the chance is 0.10, since the significance level of 0.10 is the chance of making a Type I error. Making this mistake here is OK as long as a discussion follows regarding the reason why the answer is either 0 or 1, so students can learn from their mistake and correct their thinking for the future.**

Page 53:

Think about it

The pair of 30 and 50 can occur in two ways from **Bag A** and in 12 ways from **Bag B**.
Why are these the number of occurrences?



Page 57:

Think about it

What if the average of the two selected vouchers is \$55?
Would this observation lead you to think the shown bag is **Bag A** or **Bag B**? **Bag B**
Why? **Although an average voucher value of \$55 is possible from both bags, it is more likely to occur from Bag B (42/190 chance) than from Bag A (1/190 chance).**

How would you answer these questions if the average of the two selected vouchers is \$15?
We might think that the shown bag is Bag A, because a \$15 is more likely to occur from Bag A (42/190 chance) than from Bag B (1/190 chance).

Page 60:

Think about it

What are the possible values for the average of n = 2 vouchers selected from Bag A? **The possible averages from bag A are: -495, -490,…… 50, and 55 (See page 56)**
What are the possible values for the average of n = 2 vouchers selected from Bag B? **The possible averages from bag B are: 15, 20, 25, …..520, 525, and 530 (See page 56)**
How do these possible values compare to those for each bag when the sample size was just n = 1?
The are 190 different averages, many more than the 20 for n = 1, but they are less spread.

Page 61:

Think about it

Suppose the significance level was set at α =0.05, that is, the chance of a Type I error to occur was set at 0.05. Two vouchers were then selected and the observed average turns out to be \$55. Since the *p*-value for this observed result is 0.0158, which is less than the significance level of α =0.05, you would reject the null hypothesis and conclude that the data are statistically significant at the 5% level.

You rejected H_0. Could you have made a mistake? circle one **Yes** No

What type of mistake could you have made? circle one **Type I error** Type II error

What is the chance that you have made a mistake? **Hopefully students will remember this concept of *before* versus *after* discussed in the *n* = 1 case and correctly answer 'the chance is either 0 or 1.'**

Page 63:

Think about it

Case 1

The population is the 100 students in the Engineering 101 class. The hypotheses to be tested are:

H_0: The class consists of all males (100% males).

H_1: The class does not consist of all males. (<100% males).

The decision rule is, Reject H_0 if you observe at least 1 female in your sample.
Suppose the class actually consists of 2 females and 98 males, that is, H_1 is actually true.

(a) If you take a sample of size n = 2, how likely would it be to obtain a female in your sample and thus reject the null hypothesis?

Not Very Likely Very Likely

(b) If you take a sample of size n = 80, how likely would it be to obtain a female in your sample and thus reject the null hypothesis?

Not Very Likely **Very Likely**

With a large enough sample, even this small difference (100% males under the null hypothesis versus the true percentage of 98% males in the population) can be found to be statistically significant.

Page 63:

Think about it

Case 2

The population is the 100 students in the Engineering 101 class. The hypotheses to be tested are:

H_0: The class consists of all males (100% males).

H_1: The class does not consist of all males. (<100% males).

The decision rule is, Reject H_0 if you observe at least 1 female in your sample.
Suppose the class actually consists of 30 females and 70 males, so again, H_1 is actually true.

(a) If you take a sample of size n = 80, how likely would it be to obtain a female in your sample and thus reject the null hypothesis?

Not Very Likely **Very Likely**

(b) If you take a sample of size n = 2, how likely would it be to obtain a female in your sample and thus reject the null hypothesis?

Not Very Likely Very Likely

If the sample size is too small, even a large difference (100% males under the null hypothesis versus the true percentage of 70% males in the population) may not be statistically significant.

Chapter 1 Extra Examples

Example 1.A: On-Line Instruction

Jerald Schutte, a sociology professor at California State University at Northridge, ran an experiment to assess the value of on-line instruction (Source: 'A professor divides his class in two to test value of on-line instruction.' *The Chronicle of Higher Education*, 21 February 1997). He randomly divided his statistics class into two groups. One-half completed a web-based course. The other half took the course in a traditional classroom setting. In the web-based course, problems were assigned by e-mail, students collaborated in small groups and consulted with the professor only through on-line chat-rooms. These virtual students only went to the classroom to take the midterm and final exams. On these exams, the virtual students out-performed the traditional students by an average of 20%.

Discussion:

- What might be the appropriate hypotheses for assessing if on-line instruction is more effective than learning in the traditional classroom setting?
- What information is missing from the above summary that would be needed to assess if the 20% difference in averages is indeed convincing?

Example 1.B: Cost of Off-Campus Housing

The Office of New Student Programs is planning to update one of their brochures that includes information about the cost of attending college. Part of the cost may involve off-campus housing and currently the brochure states that the average rent of a 1-bedroom apartment within 1-mile of the campus is $420 per month. A study was conducted to assess if this figure should be increased. The hypotheses were:

H_0: The average rent is $420 H_1: The average rent is more than $420

It was reported that the data were statistically significant at the 10% significance level.

As an intern of this office you are asked to interpret the results.

Discussion:

- Explain the conclusion of the study to someone who has not had a statistics course.
- Comment on which error could have been made and the consequences of that error.
- The *p*-value was reported to be 0.08. Explain what this value of 0.08 means.

Example 1.C: Cholera vaccine proves promising

Read the following summary regarding a new vaccine (Ann Arbor News, January 24, 1997).

Discussion:

- What might be the appropriate hypotheses for assessing if the vaccine is effective at protecting against cholera?
- Compute the percent of those vaccinated who developed cholera. Compute the percent of those in the unvaccinated group who developed cholera. What is the difference in these percentages?
- Do you think this difference is statistically significant? Do you think this difference is practically important?
- What questions might you ask in order to assess if the difference in cholera rates is indeed convincing?

VIETNAM

Cholera vaccine proves promising

Vietnam has produced a vaccine for cholera that could help combat the disease in poor parts of the world where it is most prevalent, a medical journal reported Friday. Tests of the vaccine have produced encouraging results, according to a report in the new issue of The Lancet. Advance copies of the magazine were available Friday. The vaccine was developed at the National Institute of Hygiene and Epidemiology in Hanoi, Vietnam, then administered in two doses to 67,395 people in the central Vietnamese city of Hue. The following year 37 people who received the vaccine developed cholera, compared with 92 in an unvaccinated group of 67,058.

Chapter 1 Projects

Project 1.A: More on the significance level. What does a significance level of 10% mean?

In the What's in the Bag? example, you were asked to decide between the following two theories:

H_0: The bag is **Bag A**.

H_1: The bag is **Bag B**.

When you were allowed to select only one voucher, one possible decision rule was:

Decision rule: Reject H_0 **if** ... your selected voucher is $\geq$\$50

This decision rule is based on setting the significance level at 10%. Just what does this value of 10% mean? Calculate the power of the test.

"If the null hypothesis is true (the bag is Bag A),

... the chance of getting a voucher of \$50 or more is only 10%,

... we would expect to get a voucher of \$50 or more about 10% of the time in many repetitions."

Let's Do It! The significance level of 0.10

All groups need to have a Bag A. When the significance level was 0.10, the cut-off value was \$50.

Each group will repeat the following procedure a total of 20 times:

(1) Mix the contents of the bag well.

(2) Select one voucher from the bag.

(3) Record a tally mark (𝍸 /) in the correct row in the table below according to whether the selected voucher has a value $\geq$\$50 or < \$50.

(4) Put the selected voucher back in the bag. Go to step (1).

Your Group Results: **TALLY**

Selected Voucher was $\geq$\$50:

Selected Voucher was < \$50:

How many times was the selected voucher $\geq$\$50 ? _____

Combine the results across all groups: **Total number of repetitions: _____**

How many times was the selected voucher $\geq$\$50 ? _____

What proportion of times (out of all repetitions) was the selected voucher $\geq$\$50 ? _____

How does this proportion compare to the significance level of 0.10?

Project 1.B: More on the *p*-value. What does a *p*-value of say 0.038 mean?

In the What's in the Bag? example, you were asked to decide between the following two theories:

H_0: The bag is **Bag A**.
H_1: The bag is **Bag B**.

You collected your own data: Observed or Selected voucher = _____

You found the *p*-value based on your observed data: P-value = _____

Just what does this *p*-value mean? If the *p*-value was 0.038 ...
"If the null hypothesis is true (the bag is Bag A),
... the chance of getting a voucher of greater than or equal to the one you observed is only 3.8%,
... we would expect to get a voucher greater than or equal to the one you observed about 3.8% of the time in many repetitions."

Let's Do It! The *p*-value

All groups need to have a Bag A. Your observed voucher was = _____

Each group will repeat the following procedure a total of 20 times:
(1) Mix the contents of the bag well.
(2) Select one voucher from the bag.
(3) Record a tally mark (/) in the correct row in the table below according to whether the selected voucher has a value ≥ your originally observed voucher or < your originally observed voucher.
(4) Put the selected voucher back in the bag. Go to step (1).

Your Group Results:	**TALLY**
Selected Voucher was ≥ your originally observed voucher:	
Selected Voucher was < your originally observed voucher:	

How many times was the selected voucher ≥ your originally observed voucher? _____

Combine the results across all groups: **Total number of repetitions: _____**

How many times was the selected voucher ≥ your originally observed voucher? _____

What proportion of times was the selected voucher ≥ your originally observed voucher?

How does this proportion compare to your *p*-value?