Beginning in Chapter 8, the main focus of the class is hypothesis testing. The book gives you many, many steps to follow. We can summarize it with the following 5 steps. I will expect your hypothesis tests to contain the following 5 steps, whether you follow the traditional method or the \( p \)-value method.

There are three worked out examples at the end.

1. **State your hypotheses.** Be sure to use the correct notation: use “\( p \)” for proportions, “\( \mu \)” for means, and “\( \sigma \)” for standard deviation (“\( \sigma^2 \)” for variance). Then, always put equality in the null hypothesis and one of “\(<\)”, “\(\geq\)”, or “\(\neq\)” in the alternative hypothesis.

   Examples:
   
   “Test the claim that more than 50%...” is written \( p > .5 \), which does not contain equality.
   
   \[
   H_0 : p = .5 \\
   H_1 : p > .5
   \]
   (“more than 50%” does not include 50%; the claim is in the alternative hyp.)

   “Test the claim that at least 50%...” is written \( p \geq .5 \), which contains equality.
   
   \[
   H_0 : p = .5 \\
   H_1 : p < .5
   \]
   (“at least 50%” means 50% or more – so the claim is in our null hyp.)

2. **Set up your decision-making framework.** This involves choosing the correct distribution and possibly finding critical values. To test a claim about a proportion, you’ll always use the normal distribution. Later you learn about the \( t \)-distribution, the \( \chi^2 \)-distribution and the \( F \)-distribution. Pay attention to when you use each of these! Also decide if the test is left-tailed, right-tailed, or two tailed. This depends entirely on the alternative hypothesis: “\(<\)”\(\rightarrow\) left, “\(>\)”\(\rightarrow\) right, and “\(\neq\)”\(\rightarrow\) two-tailed. A sketch here can help immensely. State clearly which distribution you are using, and why.

3. **Do your calculations.** This is where you have a choice, as there are two main ways of conducting hypothesis tests. The calculator emphasizes the \( p \)-value method. The \( p \)-value is the probability of getting a sample at least as unusual as our sample. So we find the area (probability) that our sample creates in one or both tails. The built-in calculator functions do this automatically – you’ll see in the display something like “\(p: 0.045\)”, which means that the \( p \)-value is 0.045 (unusual if our significance level is .05).

   Alternatively, there is the **traditional method.** In this case, we convert the significance level into a critical value using one of the inverse functions or programs. We also calculate a test statistic from our sample data using an appropriate formula (on the “Tables and Formulas” sheet) or STAT\(\rightarrow\)TESTS function.

4. **Make your comparison.** Now we compare our \( p \)-value to the significance level (notice that we are comparing two areas) OR we compare the test statistic to the critical value (depending on which method you are using – the answers in the solutions manual often do both, but it’s enough to do one method).

5. **Make your conclusion.** If it turns out the \( p \)-value is too small, or the test statistic is too extreme, then we must reject our null hypothesis. Otherwise we fail to reject the null. Be sure to also give a statement in terms of the original claim. See the “Formulas and Tables” handouts for more on this.
Example 1: Testing a claim about a proportion.
In a study of smokers who used patch therapy to quit, 39 were smoking one year later and 32 were not smoking one year later. Use a 0.1 significance level to test the claim that among smokers who use the patch, the majority are smoking one year later.

1. Hypotheses. [“the majority” does not imply equality, so this will translate into my alternative hypothesis. That is, the claim is in \( H_1 \) for this problem.]
\[
H_0 : p = .5 \\
H_1 : p > .5 \text{ (claim)}
\]
[Notice I do not use my sample data in the statement of the hypotheses.]

2. Framework. Since this is a test about proportions, I use the normal distribution. The alternative hypothesis contains “>” so it is a right-tailed test.

Sample data: \( \hat{p} = 39/(39 + 32) = .549 \)
(Note that .549 > .5, but is it significantly greater?)

3. Calculations.

\[ p \text{-value method:} \]
Using STAT \( \rightarrow \) TESTS \( \rightarrow \) 1-PropZTest and entering the given data (\( p_0 = .5, x = 39, n = 71 \)) and choosing the correct alt. hyp (\( > p_0 \)), I get \( p \)-value = .203

\[ \text{traditional method} \]
Using the InvNorm distribution or program on the significance level of 0.1 in the right tail gives me the critical value \( z = 1.28 \). The test statistic is \( z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{(39/71) - .5}{\sqrt{(\frac{.5 \cdot .5}{71})}} = .83 \)
[Notice value this also appears in the display of the 1-PropZTest readout! So I don’t actually have to use the formula – this is always true on our calculator!]

4. Comparison.

\[ p \text{-value method:} \]
.203 > .1
[We compare our \( p \)-value to the significance level. Remember this means we are comparing probabilities or areas.]

Our sample is not unusual.

Here is an updated sketch:
(stretched a bit for detail)

\[ \text{traditional method} \]
.83 < 1.28
[We compare our test statistic to the critical value. These are both values on the horizontal axis.]

Our sample is not unusual.

5. Conclusion. Fail to reject \( H_0 \). There is not sufficient evidence to support the claim and to conclude \( p > .5 \). We cannot conclude that the majority of patch users are smoking a year after treatment.
Example 2: Testing a claim about a mean.
A sample of 40 new baseballs had a bounce height mean of 92.67 in. and a SD of 1.79 in. Use a .05 sig. level to determine whether there is sufficient evidence to support the claim that the new balls have bounce heights with a mean different from 92.84 in. (a previous test figure).

1. Hypotheses. [“…mean different from 92.84 in.” does not imply equality, so this will translate into my alternative hypothesis. That is, the claim is in \( H_1 \) for this problem.]

\[
H_0 : \mu = 92.84 \\
H_1 : \mu \neq 92.84 \text{ (claim)}
\]

[Our sample data: \( n = 40, \overline{x} = 92.67, s = 1.79 \)]

2. Framework. Since this is a test about means where the population standard deviation is unknown I use the t-distribution. The alternative hypothesis contains “\( \neq \)” so it is a two-tailed test.

[In a two-tailed test, the area in each tail is the same.]

3. Calculations.

\textbf{p-value method:}

Using STAT \( \rightarrow \) TESTS \( \rightarrow \) TTest, entering the given data (\( \mu_0 = 92.84, \overline{x} = 92.67, Sx = 1.79, n = 40 \)) and choosing the correct alt. hyp, I get \( p = .5515 \)

\textbf{traditional method}

I could use the formula to calculate my (t) \textbf{test statistic}, or I can follow the steps at left and see in the display: \( t = -.600656 \). Using the TINVRS83 program or the t-table, I convert the significance level of .05 to the \textbf{critical values} \(+/-2.023\)

4. Comparison.

\textbf{p-value method:} 

\( .5515 > .05 \)

[We compare our \( p \)-value to the significance level. (Remember here we are comparing probabilities or areas.)]

Our sample is not unusual.

\textbf{traditional method}

\( -2.023 < -.6007 < 2.023 \)

[We compare our test statistic to the critical values.]

Our sample is not unusual.

Here is an updated sketch: (stretched a bit for detail)

5. Conclusion. Fail to reject \( H_0 \). There is not sufficient evidence to support the claim and to conclude \( \mu \neq 92.84 \) in. It does not appear that the new baseballs have a different mean bounce height.
Example 3: Testing a claim about a standard deviation.
Tests in the past have had scores with a standard deviation equal to 14.1. A recent class has 27 scores with a standard deviation of 9.3. Use a 0.01 significance level to test the claim that this current class has less variation than past classes.

1. Hypotheses. [“…has less variation” does not imply equality, so this will translate into my alternative hypothesis. That is, the claim is in $H_1$ for this problem.]

   $H_0 : \sigma = 14.1$
   $H_1 : \sigma < 14.1$ (claim)

   [Sample data info: $n = 27$ and $s = 9.3$]

2. Framework. Since this is a test about stand. dev., I use the $\chi^2$ distribution. The alternative hypothesis contains “$<$” so it is a left-tailed test.

3. Calculations.

   $p$-value method:

   Get the test stat at right (11.311), then use
   $2^{\text{nd}} \rightarrow \text{DISTR} \rightarrow \chi^2 \text{cdf}(-99999, 11.311, 26)$ to get
   the $p$-value of .00556
   (See p.435 in 7-6.)

   traditional method

   Using the CHINV83 program or table A-4 on
   the significance level of 0.01 in the left tail
   (0.99 in the right) gives me a critical value of
   12.198. The test statistic is
   $\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{26 \times 9.3^2}{14.1^2} = 11.311$

4. Comparison.

   $p$-value method:  
   .00556 < .01
   Our sample is unusual.

   traditional method

   $11.311 < 12.198$
   Our sample is unusual.

5. Conclusion. Reject $H_0$. There is sufficient evidence to support the claim and to conclude $\sigma < 14.1$. The result suggests that the current class has less variation than past classes.