Lesson 9
Using Technology

Outline of the Lesson

We have developed two methods for analyzing categorical variables and making statements about the population proportion *p*. In Lesson 7 we learned about hypothesis tests:

Hypothesis test

- The purpose is to use a random sample's proportion to investigate a claim about the proportion for the entire population.
- The investigation is based on a null hypothesis and an alternative hypothesis. The alternative hypothesis can take one of three forms.

$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
$H_a: p \neq p_0$	$H_a: p > p_0$	$H_a: p < p_0$

- Calculate the proportion for the sample, using the formula $\hat{p} = \frac{x}{n}$. Here *x* is the count for the sample, *n* is the sample size, and $\hat{p}$ is the proportion for the sample.
- Calculate the standard error $se = \sqrt{\frac{p_0(1-p_0)}{n}}$.
- Calculate a *z* score using $z = \frac{\hat{p}-p_0}{se}$, and use the *z* score to calculate a *p*-value.
- Make a decision by comparing the *p*-value to the predetermined significance level α.
  - If $p < \alpha$, reject the null hypothesis. There is sufficient evidence to conclude that the alternative hypothesis is true.
  - Otherwise, fail to reject the null hypothesis. There is *not* sufficient evidence to conclude that the alternative hypothesis is true.

Not surprisingly, technology tools (calculators and computer programs) are capable of automating the calculations involved. All we must do is supply the values for X, n, and p_0, and the calculator or computer program will determine the *z* score and the *p*-value. (It will still be up to us to draw the appropriate conclusions.)

Next, in Lesson 8, we learned about confidence intervals:

Confidence interval

- The purpose is to use a random sample's proportion to estimate the proportion for the entire population.
- Calculate the proportion for the sample, using the formula $\hat{p} = \frac{x}{n}$. Here x is the number of "successes," n is the sample size, and $\hat{p}$ is the proportion for the sample.
- Calculate the standard error $se = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- Based on the desired confidence level, use a table to determine an appropriate z*. The margin of error is $z^* \cdot se$.
- The confidence interval is $(\hat{p} - z^* \cdot se, \hat{p} + z^* \cdot se)$, that is $\hat{p} \pm margin\ of\ error$.

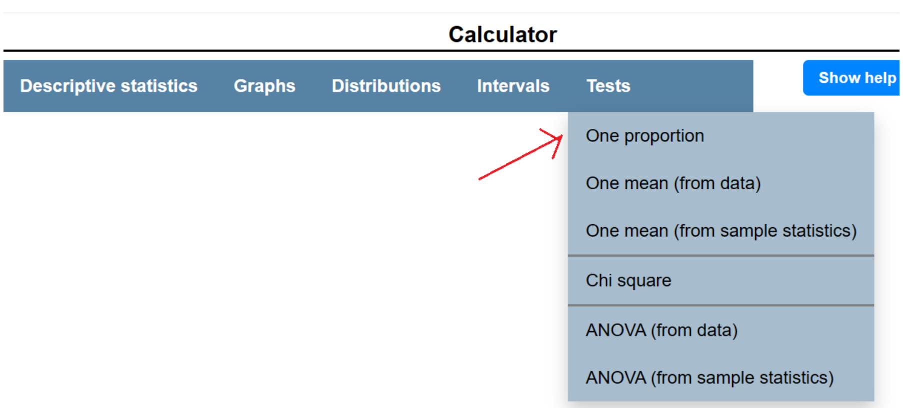
Again, technology can be used to perform the calculations: if we supply X, n, and the desired confidence level, the calculator or computer program will calculate the confidence interval.

The purpose of this lesson is to show you how to use technology – specifically, the online calculators provided by the author of these lessons – to carry out hypothesis test and confidence interval calculations for population proportions. We begin with the calculator designed for solving problems such as those typically encountered in an introductory statistics class[1]. Here again is a link to that calculator:

Statistical calculator

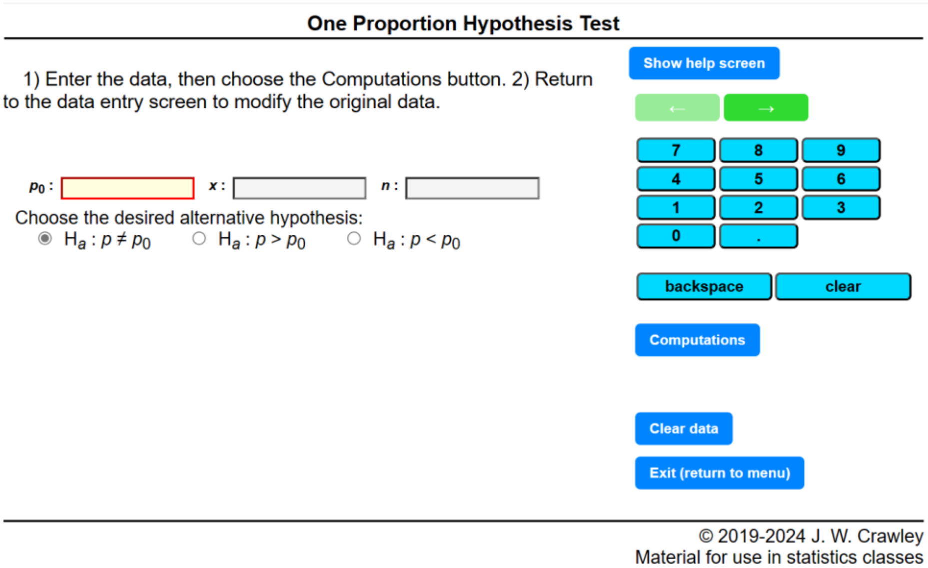
9.1 – Hypothesis Tests with Technology

To carry out a hypothesis test for a proportion, simply click on the submenu option *One proportion* in the *Tests* menu as shown here:



Once we choose this option, we get the following:

[1] In Section 9.3 we will use the datafile-based calculator to analyze categorical variables in a data file.



To complete the process:

- Enter the value of p_0 from the null hypothesis
- Enter the number of successes x
- Enter the sample size n
- Choose whether you want a two-sided or a particular one-sided alternative. By default the alternative hypothesis will be the two-tail form:

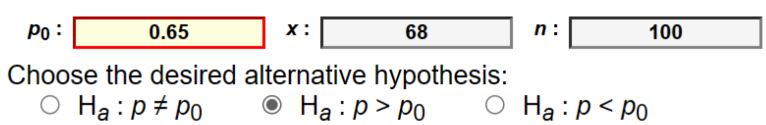
  $$H_z: p \neq p_0$$

  However, you can change that to one of the one-tail tests if you wish.
- Click on *Computations*.

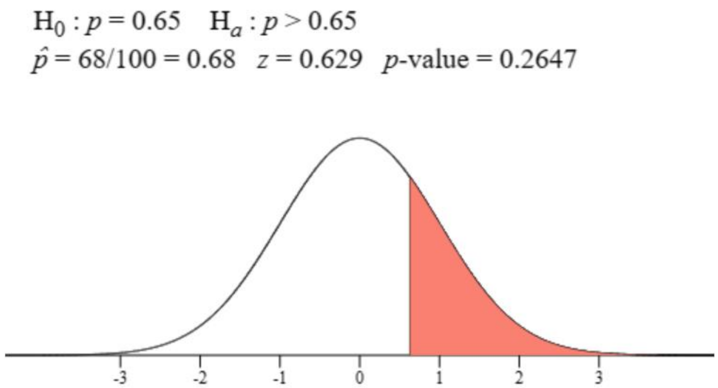
Example: Test the following claim, assuming that 68 of the 100 people surveyed answered "yes" when asked the question. Is this enough evidence to refute the null hypothesis?

$$H_0: p = 0.65$$
$$H_a: p > 0.65$$

Solution: In this problem, $p_0 = 0.65$, $n = 100$ and $x = 68$. We want a one-tail test, since the alternative hypothesis has the form "$p > p_0$". We enter the data as shown below:



When we click *Computations* we obtain these results:



Therefore the *z*-test statistic value (rounded to four places) is 0.6290 and the *p*-value is 0.2647. Since the *p*-value is not small we do not reject the null hypothesis. Notice that the output contains a graphical rendering of the *p*-value.

Exercise 1[2]:

a. Use the calculator to test the claim that a population proportion is 53%. Use a two-tail test. Report the test statistic and the *p*-value, and state your conclusion. In the random sample, 859 of the 1562 surveyed answered "yes."

b. Check your answer by doing the calculation "by hand," that is without the *Tests* menu of the calculator.[3]

The following applets provide additional practice with the calculations. The first limits the problems to two-tail tests, while the second has a mixture of one-tail and two-tail tests. You have used one or both of these applets in Lesson 7; the difference is that this time you should use technology (the *Tests* menu of the calculator) to carry out the calculations.

Two-tail hypothesis tests for proportions

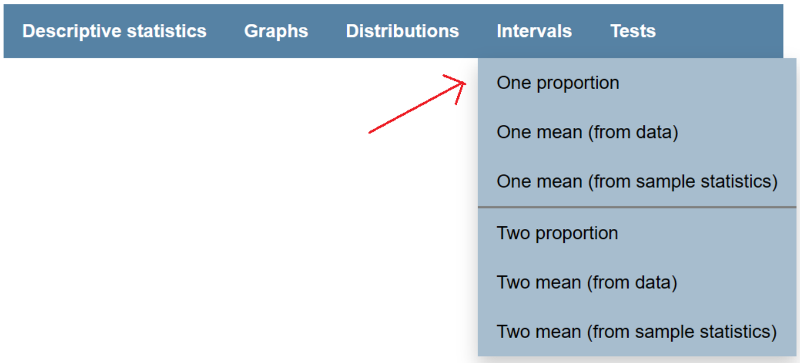
One-tail and two-tail hypothesis tests for proportions

9.2 – Confidence Intervals with Technology

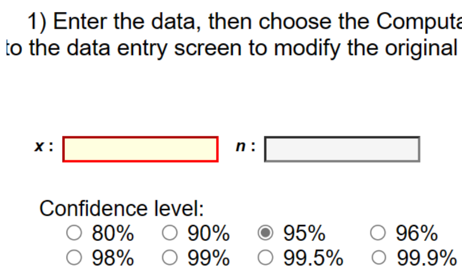
To calculate a confidence interval for a population proportion, simply click on the submenu option *One proportion* in the *Intervals* menu as shown here:

[2] Solutions to the exercises may be found at the end of the lesson.

[3] Throughout this lesson we will consistently put the phrase *by hand* in quotes for at least two reasons. First, we certainly use a standard calculator for most of the calculations. Second, when doing hypothesis tests we frequently use statistics-based technology to calculate the *p*-value (the alternative would be to use a table such as Table A). So, in terms of the calculator supplied with these lessons, "by hand" simply means without using the ***Intervals*** or ***Tests*** menu options.



Once we choose this option, we get the following:

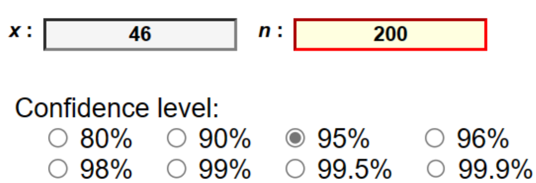


To complete the process, we:

- Enter the number of successes x
- Enter the sample size n
- Use the radio button to choose the confidence level (the default is 95%)
- Click on *Computations*.

Example: Suppose that in a random sample of 200 people in a certain district, 46 said that they would vote for a particular candidate in an upcoming election. Find a 95% confidence interval to estimate the population proportion of all people in this district that would vote for this candidate.

Solution: In this problem x=46, n=*200*, and the confidence level is 95%. We enter the data as shown below:



When we click *Computations* we obtain these results:

Point estimate ($\hat{p}$) = 46/200 = 0.2300

95% confidence interval: (0.1717, 0.2883)

Details:
Standard error = 0.0298
z^* = 1.960
Margin of error = 0.0583

The 95% confidence interval is (.1717,.2883). This means that we are 95% confident that between 17.17% and 28.83% of people in this district would vote for this candidate. The calculator also shows the $\hat{p}$ value, 23%, as well as some of the details that would be involved in a calculation "by hand."

Exercise 2:

a. Use the calculator to calculate a 99% confidence interval if 433 persons in a random sample of size 923 answered "yes" to the question posed.
   - Report the answer as an interval, in a form similar to (42.31%, 48.75%)
   - Report the answer in a form similar to 45.53% ± 3.22%

b. Check your answer by doing the calculation "by hand," that is without using the *Intervals* menu of the calculator.

The following applet provides additional practice in calculating confidence intervals with varying confidence levels. You have used this applet in Lesson 8; the difference is that this time you should use technology to carry out the calculations.

Confidence interval calculations

9.3 – Data File Analysis, Part 3

As we mentioned earlier, the author of the lessons has provided two calculators for your use. In this section we continue our description of the second calculator, which is designed to work with data files similar to those that might be created by spreadsheet software. The data file we use is the same used in Lessons 2 and 3. You should have it saved to your own computing device, but in any case here again is a link to that file:

First day survey

As a reminder, the file contains student responses to a first day survey containing these questions:

1. What is your gender? (M) Male (F) Female
2. What is your class year? (FR) Freshman (SO) Sophomore (JR) Junior (SR) Senior
3. How many states have you visited?
4. Do you currently smoke? (Y) Yes (N) No
5. How tall are you (in inches)?
6. How many days per week do you read a newspaper?

In the Data File Analysis, Part 1 (in Lesson 2), we used statistics (mean, etc.) and graphs to examine several of the variables individually. In Part 2 (in Lesson 3) we examined associations between two of the variables in the file. again using statistics (mean, etc.) and graphs as our tools. Now we will use the calculator to study proportions for a categorical variable in the file – using hypothesis tests and confidence intervals as our tools. To begin, open the data file calculator using the following link, then use the *Load vertical file* button to load the data file containing the student responses.

Data file calculator

NOTES:

1. The survey for this datafile was administered to two classes of an introductory statistics course for non-majors at a state-owned university. Accordingly, we will take our population to be all students who take that particular course at that particular university. We will treat those two classes of students as being representative of that population, although of course the students surveyed were not obtained from a random sample. Any interpretation of our results should be analyzed with this in mind – the mathematics is based on having a random sample from the population.
2. The survey in the datafile was administered to a fairly small group – only 63 students. Typical data file analysis would generally involve a much larger set of data. Because the file is small, it may happen that the conditions for carrying out our analyses are not met. For instance, in the example that follows, $np_0 = 63(0.15) = 9.45$, which is not greater than or equal to 15. This should also be kept in mind when interpreting the results. In any case, however, this file is useful for illustrating the mechanics of data file analysis.

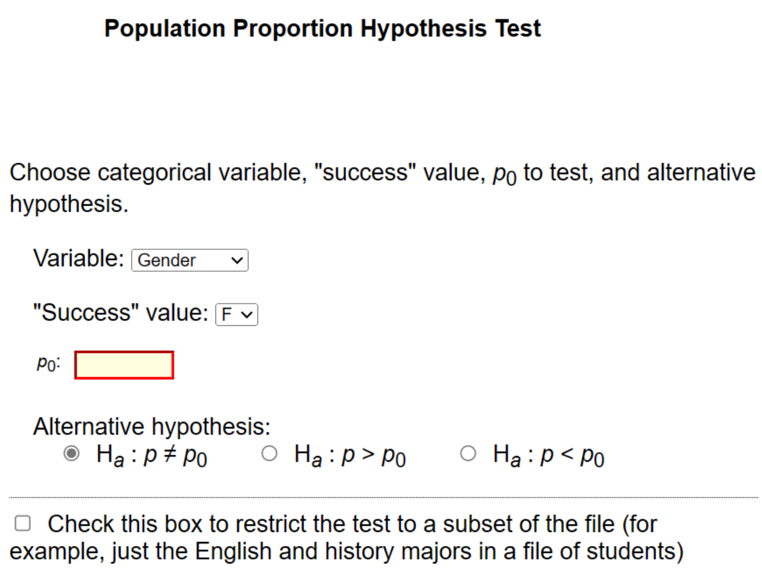
Example. A recent article claims that 15% of all college students smoke. Test to see whether this applies to all students who take that course at that university.

Solution. The hypotheses we are testing are:

$H_0 : p = 0.15$
$H_a : p \neq 0.15$

where p is the proportion of smokers in the population consisting of all students who take that course at that university.

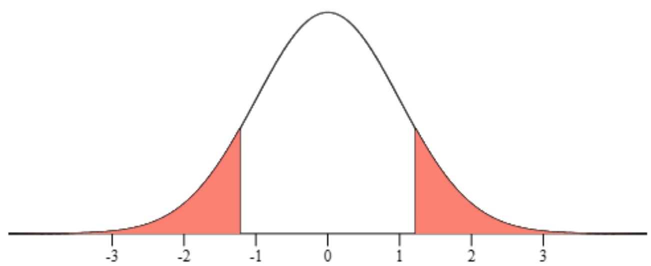
Use menu option *Tests*, submenu option *One proportion hypothesis test*, to obtain the following screen:



Use the pull-down lists to select *Smoke* as the variable, and *Y* as the corresponding “success” value. Fill in 0.15 as the value for p_0; the default is correct for the alternative hypothesis. When you click *Computations* you should obtain this result:

Hypothesis test for proportion of "successes"
Variable: *Smoke*
"Success" value: Y

$H_0 : p = 0.15$
$H_a : p \neq 0.15$
$\hat{p} = 6/63 = 0.0952$
p-value = 0.2235



If desired, you can check the box *Include some calculation details*, which will add the following details to the result screen:

Details: (*se* is standard error)
Sample size: $n = 63$
$se = \sqrt{p_0(1 - p_0) / n} = 0.045$
$z = (\hat{p} - p_0) / se = -1.2173$

We can see that the sample proportion was 9.52%, yielding a test statistic (z-score) of -1.2173 and a p-value of 0.2235. We can write our conclusion as: “There is not enough evidence to conclude that the proportion of smokers among all students who take that particular class at that university is different from the national value of 15%.” Although the sample proportion was quite a bit lower than 15%, it was not enough lower to assert a difference for the entire population.

Example. Create a 99% confidence interval for the proportion of sophomores

Solution. Use menu option *Intervals*, submenu option *One proportion hypothesis test*, to obtain the following screen:

Population Proportion Confidence Interval

Choose categorical variable, "Success" value, and confidence level.

Variable: Gender

"Success" value: F

Confidence level:
○ 80% ○ 90% ◉ 95% ○ 96%
○ 98% ○ 99% ○ 99.5% ○ 99.9%

☐ Check this box to restrict the interval to a subset of the file (for example, just the English and history majors in a file of students)

Use the pull-down lists to select *Class_Year* as the variable, and *SO* as the corresponding "success" value. Use the radio buttons to change the confidence level to 99%. When you click *Computations* you should obtain this result:

Confidence interval for proportion of "successes"
Variable: *Class_Year*
"Success" value: SO

Point estimate ($\hat{p}$) = 27/63 = 0.4286
99% confidence interval: (0.2680, 0.5892)

If desired, you can check the box "Include some calculation details, which will add the following details to the result screen:

Details: (*se* is standard error, m.e. is margin of error)
Sample size: $n = 63$
$se = \sqrt{\hat{p}(1-\hat{p})/n} = 0.0623$
$z^* = 2.576$
m.e. = $(se)(z^*) = 0.1606$
Interval = $(\hat{p} - \text{m.e.}, \hat{p} + \text{m.e.})$

We can write the conclusion as: "I am 99% confident that between 26.80% and 58.92% of all students at that university who take that class are sophomores."

Exercise 3:

a. At that university 52% of the students are female. A researcher suspects that proportion does not carry over to that particular introductory statistics class. Use the data file to test the researcher's claim.
b. Calculate a 90% confidence interval for the proportion of sophomores in the population.

Exercise 4: You may have noticed that, like the menu options we used in Lesson 2, these menu options provide the ability to restrict the analysis to a subset of the entire data file. Use this ability to answer the following.

a. Repeat Exercise 3(a), but restricted to the students who are NOT freshmen.
b. Repeat Exercise 3(b), but restricted to the non-smokers.

Exercise 5: In Exercise 16 of Lesson 2 you created a data file containing the data presented originally in Lesson 1. This data was collected in a statistics course at a public university. Based on the list of majors that require that course, a researcher wants to investigate several characteristics of the population of students who take that course. Use that data file for the following:

a. Test the claim that 50% of the students who take the course are Democrats.
b. Estimate the percentage of the underclassmen (freshmen/sophomores) who have a job, using a 90% confidence interval.

Solutions to Exercises

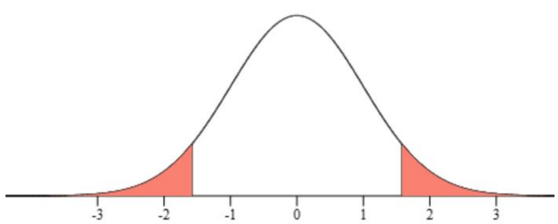
1:

a. Use the calculator to test the claim that a population proportion is 53%. Use a two-tail test. Report the test statistic and the *p*-value, and state your conclusion. In the random sample, 859 of the 1562 surveyed answered "yes."

Here is the resulting calculator output:

$H_0 : p = 0.53 \quad H_a : p \neq 0.53$

$\hat{p} = 859/1562 = 0.5499 \quad z = 1.5787 \quad p\text{-value} = 0.1144$



Test statistic $z = 1.5787$, p-value = 0.1144, do not reject null hypothesis.

b. Check your answer by doing the calculation "by hand," that is without the *Tests* menu of the calculator.

$\hat{p} = \frac{859}{1562} = 0.5499 \qquad se = \sqrt{\frac{.53(1-.53)}{1562}} = 0.0126$

$z = \frac{0.5499-0.53}{0.0126} = 1.5794 \qquad p\text{-value} = 0.1142$

Note that answers are slightly different due to rounding at each step.

2:

a. Use the calculator to calculate a 99% confidence interval if 433 persons in a random sample of size 923 answered "yes" to the question posed.

Here is the resulting calculator output:

Point estimate ($\hat{p}$) = 433/923 = 0.4691

99% confidence interval: (0.4268, 0.5114)

Details:
Standard error = 0.0164
$z^* = 2.576$
Margin of error = 0.0423

- Report the answer as an interval, in a form similar to (42.31%, 48.75%)
  (42.68%, 51.14%)
- Report the answer in a form similar to 45.53% ± 3.22%
  46.91% ± 4.23% (The margin of error is shown in the "Details" portion of the output; it can also be found by subtracting the sample proportion from the upper end of the interval: 51.14% – 46.91%)

b. Check your answer by doing the calculation "by hand," that is without using the *Intervals* menu of the calculator.

$\hat{p} = \frac{433}{923} = 0.4691 \qquad se = \sqrt{\frac{.4691(1-.4691)}{923}} = 0.0164$

$me = 2.576 \cdot se = 0.0422$

Interval: (46.91% – 4.22%, 46.91% + 4.22%) = (42.69%, 51.13%)

Note that answers are slightly different due to rounding at each step.

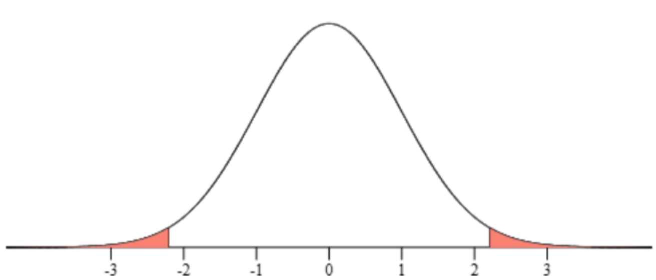
3:

a. At that university 52% of the students are female. A researcher suspects that proportion does not carry over to that particular introductory statistics class. Use the data file to test the researcher's claim.

Hypothesis test for proportion of "successes"
Variable: *Gender*
"Success" value: F

$H_0 : p = 0.52$
$H_a : p \neq 0.52$
$\hat{p} = 24/63 = 0.381$
p-value = 0.0272

Details: (se is standard error)
Sample size: $n = 63$
$se = \sqrt{p_0(1-p_0)/n} = 0.0629$
$z = (\hat{p} - p_0)/se = -2.2091$



At significance level α = 0.05, we do have evidence (p-value 0.0272) to support the researcher's claim that the proportion of females among all the students at that university who take that class is not 52%. (It appears to be lower than 52%).

b. Calculate a 90% confidence interval for the proportion of sophomores in the population.

Confidence interval for proportion of "successes"
Variable: *Class_Year*
"Success" value: SO

Point estimate ($\hat{p}$) = 27/63 = 0.4286
90% confidence interval: (0.3260, 0.5311)

Details: (*se* is standard error, m.e. is margin of error)
Sample size: $n = 63$
$se = \sqrt{\hat{p}(1-\hat{p})/n} = 0.0623$
$z^* = 1.645$
m.e. = $(se)(z^*) = 0.1026$
Interval = $(\hat{p} - \text{m.e.}, \hat{p} + \text{m.e.})$

We are 90% confident that for the population of all students who take that class, the proportion of sophomores is between 32.60% and 53.11%. Put another way, the proportion is 42.86% with a margin of error of 10.26%.

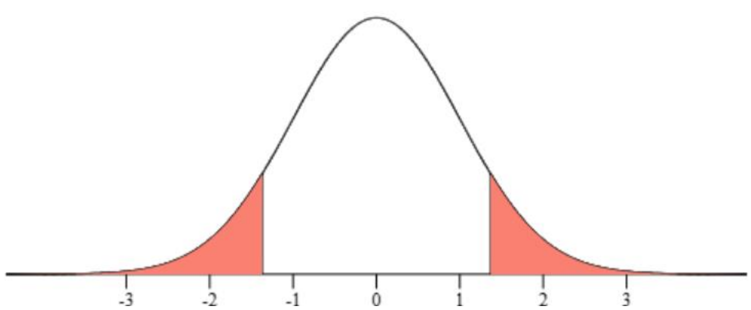
4: You may have noticed that, like the menu options we used in Lesson 2, these menu options provide the ability to restrict the analysis to a subset of the entire data file. Use this ability to answer the following.

a. Repeat Exercise 3(a), but restricted to the students who are NOT freshmen.

Hypothesis test for proportion of "successes"
Variable: *Gender*
"Success" value: F
Restricted to records where *Class_Year* is one of:
JR, SO, SR

$H_0 : p = 0.52$
$H_a : p \neq 0.52$
$\hat{p} = 24/56 = 0.4286$
p-value = 0.1709

Details: (*se* is standard error)
Sample size: $n = 56$
$se = \sqrt{p_0(1-p_0)/n} = 0.0668$
$z = (\hat{p} - p_0)/se = -1.3695$



At significance level α = 0.05, we do NOT have evidence (p-value 0.1709) to support the researcher's claim that the proportion of females among all the *non-freshman* students at that university who take that class is not 52%. Notice that it is lower (42.86%) in the sample, but not enough to draw a conclusion about the entire population.

b. Repeat Exercise 3(b), but restricted to the non-smokers.

Confidence interval for proportion of "successes"
Variable: *Class_Year*
"Success" value: SO
Restricted to records where *Smoke* equals:
N

Point estimate ($\hat{p}$) = 26/57 = 0.4561
90% confidence interval: (0.3476, 0.5647)

Details: (*se* is standard error, m.e. is margin of error)
Sample size: $n = 57$
$se = \sqrt{\hat{p}(1-\hat{p})/n} = 0.0660$
$z^* = 1.645$
m.e. = $(se)(z^*) = 0.1085$
Interval = $(\hat{p} - \text{m.e.}, \hat{p} + \text{m.e.})$

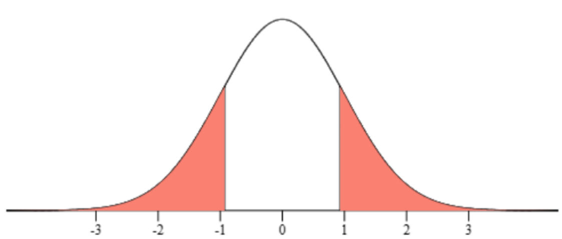
We are 90% confident that for the population of all non-smokers who take that class, the proportion of sophomores is between 34.76% and 56.47%.

5: In Exercise 16 of Lesson 2 you created a data file containing the data presented originally in Lesson 1. This data was collected in a statistics course at a public university. Based on the list of majors that require that course, a researcher wants to investigate several characteristics of the population of students who take that course. Use that data file for the following:

a. Test the claim that 50% of the students who take the course are Democrats.

Hypothesis test for proportion of "successes"
Variable: *Political Party*
"Success" value: D

$H_0 : p = 0.5$
$H_a : p \neq 0.5$
$\hat{p} = 12/29 = 0.4138$
p-value = 0.3532



There is NOT have evidence (*p*-value 0.3532) to reject the claim that the proportion in the population is 50%.

b. Estimate the percentage of the underclassmen (freshmen/sophomores) who have a job, using a 90% confidence interval.

Confidence interval for proportion of "successes"
Variable: *Job*
"Success" value: Yes
Restricted to records where *Class Yr* is one of:
Freshman, Sophomore

Point estimate ($\hat{p}$) = 10/23 = 0.4348
90% confidence interval: (0.2648, 0.6048)

In the population consisting of all the freshman and sophomore students who take that particular course, we are 90% confident that between 26.48% and 60.48% have a job.