Stat 250 Gunderson Lecture Notes
5: Learning about a Population Proportion

Part 1: Distribution for a Sample Proportion

To be a statistician is great!! You never have to be "absolutely sure" of something.
Being "reasonably certain" is enough!
-- Pavel E. Guarisma, North Carolina State University

Recall: Parameters, Statistics, and Statistical Inference

Some distinctions to keep in mind:

- **Population** versus **Sample**
- **Parameter** versus **Statistic**
  Population proportion p versus sample proportion $\hat{p}$
  Population mean μ versus sample mean $\overline{X}$

Since we hardly ever know the true population parameter value, we take a sample and use the sample statistic to estimate the parameter. When we do this, the sample statistic may not be equal to the population parameter, in fact, it could change every time we take a new sample. Will the observed sample statistic value be a reasonable estimate? If our sample is a ***RANDOM SAMPLE***, then we will be able to say something about the accuracy of the estimation process.

Statistical Inference: the use of sample data to make judgments or decisions about populations.

The two most common statistical inference procedures are confidence interval estimation and hypothesis testing.

- **Confidence Interval Estimation**: A confidence interval is a range of values that the researcher is fairly confident will cover the true, unknown value of the population parameter. In other words, we use a confidence interval to estimate the value of a population parameter. We have already encountered the idea of a margin of error and using it to form a confidence interval for a population proportion.
- **Hypothesis Testing:** Hypothesis testing uses sample data to attempt to reject a hypothesis about the population. Usually researchers want to reject the notion that chance alone can explain the sample results. Hypothesis testing is applied to population parameters by specifying a null value for the parameter—a value that would indicate that nothing of interest is happening. Hypothesis testing proceeds by obtaining a sample, computing a sample statistic, and assessing how unlikely the sample statistic would be if the null parameter value were correct. In most cases, the researchers are trying to show that the null value is not correct. Achieving statistical significance is

equivalent to rejecting the idea that the observed results are plausible if the null value is correct.

An Overview of Sampling Distributions

The value of a statistic from a **random sample** will vary from sample to sample. So **a statistic is a random variable** and **it will have a probability distribution**. This probability distribution is called the **sampling distribution** of the statistic.

> ***Definition:***
> The distribution of all possible values of a statistic for repeated samples of the same size from a population is called the **sampling distribution of the statistic**.

We will study the sampling distribution of various statistics, many of which will have approximately normal distributions. The general structure of the sampling distribution is the same for each of the five scenarios. The sampling distribution results, along with the ideas of probability and random sample, play a vital role in the inference methods that we continue studying throughout the remainder of the course.

Sampling Distributions for One Sample Proportion

Many responses of interest produce counts rather than measurements -- sex (male, female), political preference (republican, democrat), approve of new proposal (yes, no). We want to learn about a population proportion and we will do so using the information provided from a sample from the population.

Example: Do you work more than 40 hours per week?

A poll was conducted by The Heldrich Center for Workforce Development (at Rutgers University). A probability sample of 1000 workers resulted in 460 (for 46%) stating they work more than 40 hours per week.

Population =

Parameter =

Sample =

Statistic =

Can anyone say how close this observed sample proportion $\hat{p}$ is to the true population proportion p ?

If we were to take another random sample of the same size $n = 1000$, would we get the same value for the sample proportion $\hat{p}$?

So what are the possible values for the sample proportion $\hat{p}$ if we took many random samples of the same size from this population? What would the distribution of the possible $\hat{p}$ values look like? **What is the sampling distribution of $\hat{p}$?**

Aside: Can you Visualize It?
Consider taking your one random sample of size n and computing your one $\hat{p}$ value. (As in the previous example, our one $\hat{p}$ = 460/1000 = 0.46.) Suppose we did take another random sample of the same size, we would get another value of $\hat{p}$, say ___________. Now repeat that process over and over; taking one random sample after another; resulting in one $\hat{p}$ value after another.

Example picture showing the possible values when *n* = _______

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 $\hat{p}$ values

Observations:

How would things change if the sample size n were even larger, say n = _____? Suppose our first sample proportion turned out to be $\hat{p}$ = _________ . Now imagine again repeating that process over and over; taking one random sample after another; resulting in many $\hat{p}$ possible values.

Example picture showing the possible values when *n* = _______

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 $\hat{p}$ values

Observations:

Let's take a closer look at the sample proportion $\hat{p}$. The sample proportion is found by taking the number of "successes" in the sample and dividing by the sample size. So the count variable *X* of the number of successes is directly related to the proportion of successes as $\hat{p} = \frac{X}{n}$.

Earlier we studied the distribution of our first statistic, the count statistic *X* (the number of successes in *n* independent trials when the probability of a success was *p*). We learned about its exact distribution called the **Binomial Distribution**. We also learned when the sample size *n* was large, the distribution of *X* could be approximated with a normal distribution.

Normal Approximation to the Binomial Distribution

If ***X* is a binomial** random variable based on *n* trials with success probability *p*, and ***n* is large**, then the random variable ***X* is also approximately** $N\left(np, \sqrt{np(1-p)}\right)$.

Conditions: The approximation works well when both *np* and *n*(1 – *p*) are at least 10.

So any probability question about a sample proportion could be converted to a probability question about a sample count, and vice-versa.

- If ***n is small***, we would need to convert the question to a count and use the binomial distribution to work it out.

- If ***n is large***, we could convert the question to a count and use the normal approximation for a count, OR use a related normal approximation for a sample proportion (for large *n*).

The Stat 250 formula card summarizes this related normal approximation as follows:

Sample Proportions

$$\hat{p} = \frac{x}{n}$$

Mean

$$E(\hat{p}) = \mu_{\hat{p}} = p$$

Standard Deviation

$$\text{s.d.}(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Sampling Distribution of $\hat{p}$

If the sample size n is large enough (namely, $np \geq 10$ and $n(1-p) \geq 10$)

then $\hat{p}$ is *approximately* $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.

Let's put this result to work in our next Try It! Problem.

Try It! Do Americans really vote when they say they do?

To answer this question, a telephone poll was taken two days an election. From the 800 adults polled, 56% reported that they had voted. However, it was later reported in the press that, in fact, only 39% of American adults had voted. Suppose the 39% rate reported by the press is the correct population proportion. Also assume the responses of the 800 adults polled can be viewed as a random sample.

a. Sketch the sampling distribution of $\hat{p}$ for a random sample of size n = 800 adults.

b. What is the approximate probability that a sample proportion who voted would be 56% or larger for a random sample of 800 adults?

c. Does it seem that the poll result of 56% simply reflects a sample that, by chance, voted with greater frequency than the general population?

More on the Standard Deviation of $\hat{p}$

The standard deviation of $\hat{p}$ is given by: s.d.($\hat{p}$) = $\sqrt{\dfrac{p(1-p)}{n}}$

This quantity would give us an idea about how far apart a sample proportion $\hat{p}$ and the true population proportion p are likely to be.

We can ***interpret*** **this standard deviation** as **approximately** the **average distance** of the possible $\hat{p}$ value**s** (for repeated samples of the same size n) from the **true population proportion** p.

In practice when we take a random sample from a large population, we only know the sample proportion. We generally would not know the true population proportion *p*. So we could not compute the standard deviation of $\hat{p}$.

However we can use the sample proportion in the formula to have an estimate of the standard deviation, which is called the **standard error of** $\hat{p}$.

The standard error of $\hat{p}$ **is given by: s.e.(** $\hat{p}$ **) =** $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

This quantity is an *estimate* of the standard deviation of $\hat{p}$ **.**

So we can ***interpret*** **this standard error** as ***estimating***, **approximately**, the **average distance** of the possible $\hat{p}$ value**s** (for repeated samples of the same size *n*) from the **true population proportion** *p*.

Moreover, we can use this standard error to create a range of values that we are very confident will contain the true proportion *p*, namely, $\hat{p}$ ± (a few)s.e.($\hat{p}$).

This is the basis for confidence interval for the true proportion *p*, discussed next

Try It! Love at first sight?

In a random sample of *n* = 500 adults, 300 stated they believe in love at first sight.

a. Estimate the population proportion of adults that believe in love at first sight.

b. Find the corresponding standard error of for the estimate in part a and use this standard error to provide an interval estimate for the population proportion *p*, with 95% confidence.

Stat 250 Gunderson Lecture Notes
5: Learning about a Population Proportion

Part 2: Estimating Proportions with Confidence

Big Idea of Confidence Intervals: Use sample data to estimate a population parameter.

Recall some of the language and notation associated with the estimation process.

- **Population and Population Parameter**
- **Sample and Sample Statistic** ***(sample estimate or point estimate)***

The **sample estimate** provides our best guess as to what is the value of the population parameter, but it is not 100% accurate.

The **value of the sample estimate will vary** from one sample to the next. The values often vary around the population parameter and the standard deviation give an idea about how far the sample estimates tend to be from the true population proportion on average.

The **standard error of the sample estimate** provides an idea of how far away it would tend to vary from the parameter value (on average).

The **general format** for a confidence interval estimate is given by:

Sample estimate ± (a few) standard errors

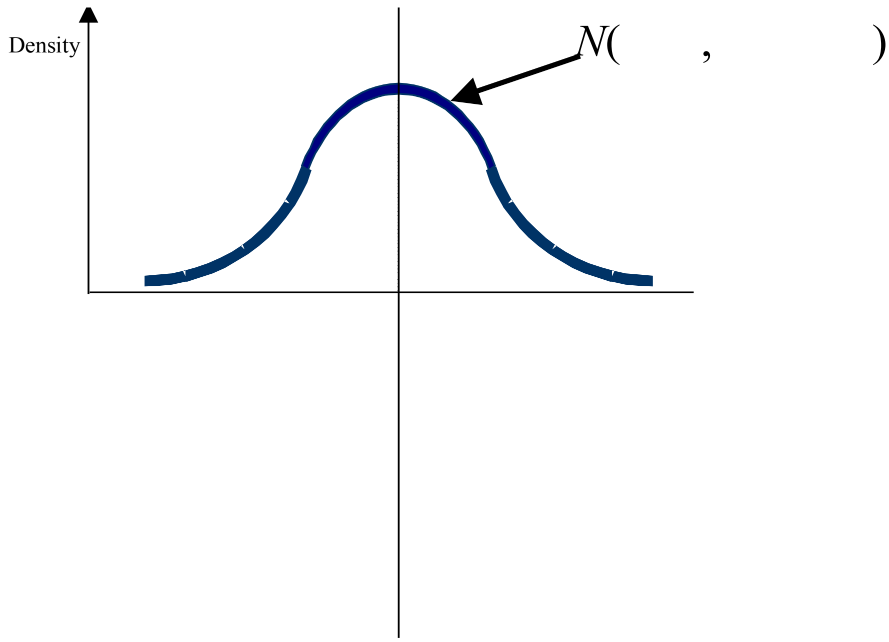
The "**few**" or number of standard errors we go out each way from the sample estimate will depend on how confident we want to be.

The "**how confident**" we want to be is referred to as the **confidence level**. This level reflects how confident we are in the *procedure*. Most of the intervals that are made will contain the truth about the population, but occasionally an interval will be produced that does not contain the true parameter value. Each interval either contains the population parameter or it doesn't. The confidence level is the percentage of the time we expect the *procedure* to produce an interval that does contain the population parameter.

Confidence Interval for a Population Proportion *p*

Goal: we want to learn about a population proportion $\hat{p}$. **How?** We take a random sample from the population and estimate *p* with the resulting sample proportion $\hat{p}$. Let's first recall how those many possible values for the sample proportion would vary, that is, the sampling distribution of the statistic $\hat{p}$.

Sampling Distribution of $\hat{p}$: If the sample size *n* is large and $np \geq 10$ and $n(1-p) \geq 10$, then $\hat{p}$ is approximately $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.



1. Consider the following interval or range of values and show it on the picture.

$$p \pm 2\sqrt{\frac{p(1-p)}{n}} \Rightarrow \left(p - 2\sqrt{\frac{p(1-p)}{n}}, p + 2\sqrt{\frac{p(1-p)}{n}}\right)$$

2. What is the probability that a (yet to be computed) sample proportion $\hat{p}$ will be in this interval (within 2 standard deviations from the true proportion *p*)?

3. Take a possible sample proportion $\hat{p}$ and consider the interval

$$\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}} \Rightarrow \left(\hat{p} - 2\sqrt{\frac{p(1-p)}{n}}, \hat{p} + 2\sqrt{\frac{p(1-p)}{n}}\right)$$

Show this range on the normal distribution picture above.

4. Did your first interval around your first $\hat{p}$ contain the true proportion *p*?
Was it a 'good' interval? ___________

5. Repeat steps 3 and 4 for other possible values of $\hat{p}$.

Big Idea:

- Consider all possible random samples of the same large size *n*.
- Each possible random sample provides a possible sample proportion value.
  If we made a histogram of all of these possible $\hat{p}$ values it would look like the normal distribution on the previous page.
- About 95% of the possible sample proportion $\hat{p}$ values will be in the interval $p \pm 2\sqrt{\frac{p(1-p)}{n}}$; and for each one of these sample proportion $\hat{p}$ values, the interval $p \pm 2\sqrt{\frac{p(1-p)}{n}}$ will contain the population proportion *p*.
- Thus about 95% of the intervals $\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}}$ will contain the population proportion *p*.

Thus, an initial 95% confidence interval for the true proportion *p* is given by:

$$\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}}$$

The Dilemma: When we take our one random sample, we can compute the sample proportion $\hat{p}$, but we can't construct the interval $\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}}$ because we don't know the value of *p*.

The Solution: Replace the value of *p* in the standard deviation with the estimate $\hat{p}$, that is use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ called ____________________**.**

An <u>approximate</u> 95% confidence interval (CI) for the population proportion *p* is:

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Note: The ± part of the interval $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is called the 95% <u>margin of error</u>.

Note: The **<u>approximate</u>** is due to the multiplier of '2' being used.
We will learn about other multipliers, including the exact 95% multiplier value later.

Try It! Getting Along with Parents

In a Gallup Youth Survey n = 501 randomly selected American teenagers were asked about how well they get along with their parents. One survey result was that 54% of the sample said they get along "VERY WELL" with their parents.

a. The sample proportion was found to be 0.54. Give the standard error for the sample proportion and use it to complete the sentence that interprets the standard error in terms of an average distance.

We would estimate the average distance between the possible ______ values

(from repeated samples) and _____________________ to be about 0.022.

b. Compute a 95% confidence interval for the population proportion of teenagers that get along very well with their parents.

c. Fill in the blanks for the typical interpretation of the confidence interval in part b:

"Based on this sample, with 95% confidence, we would estimate that

somewhere between ________ and _______ of *all* American teenagers think they get

along very well with their parents."

d. Can we say the probability that the above (already observed) interval (___________ , ___________) contains the population proportion p is 0.95?

That is, can we say $P($________ $\le p \le$ _________$) = 0.95$?

e. Can we say that 95% of the time the population proportion p will be in the interval computed in part b?

Just what does the 95% confidence level mean? Interpretation

The phrase **confidence level** is used to describe the likeliness or chance that a yet-to-be constructed interval will actually contain the true population value. However, we have to be careful about how to interpret this level of confidence if we have already completed our interval.

The population proportion *p* is not a random quantity, it does not vary - once we have "looked" (computed) the actual interval, we cannot talk about probability or chance for this particular interval anymore. The 95% **confidence level applies to the procedure**, not to an individual interval; it applies "before you look" and not "after you look" at your data and compute your interval.

Try It! Getting Along with Parents

In the previous Try It! you computed a 95% confidence interval for the population proportion of teenagers that get along very well with their parents in part (b). This was based on a random sample of n = 501 American teenagers. You interpreted the interval in part (c). Write a sentence or two that ***interprets the confidence <u>level</u>***.

The interval we found was computed with a method which if repeated over and over ...

Try It! Completing a Graduate Degree

A researcher has taken a random sample of n = 100 recent college graduates and recorded whether or not the student completed their degree in 5 years or less. Based on these data, a 95% confidence interval for the population proportion of all college students that complete their degree in 5 years or less is computed to be (0.62, 0.80).

a. How many of the 100 sampled college graduates completed their degree in 5 years or less?

b. Which of the following statements gives a valid interpretation of this 95% confidence level? Circle all that are valid.

   i. There is about a 95% chance that the population proportion of students who have completed their degree in 5 years or less is between 0.62 and 0.80.

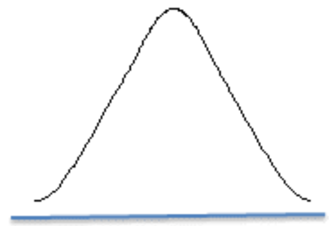
   ii. If the sampling procedure were repeated many times, then approximately 95% of the resulting confidence intervals would contain the population proportion of students who have completed their degree in 5 years or less.

   iii. The probability that the population proportion p falls between 0.62 and 0.80 is 0.95 for repeated samples of the same size from the same population.

What about that Multiplier of 2?

The exact multiplier of the standard error for a 95% confidence level would be 1.96, which was rounded to the value of 2. Where does the 1.96 come from? Use the standard normal distribution, the $N(0, 1)$ distribution at the right and Table A.1.

Researchers may not always want to use a 95% confidence level. Other common levels are 90%, 98% and 99%.



Using the same idea for confirming the value of 1.96, find the correct multiplier if the confidence level were 90%.

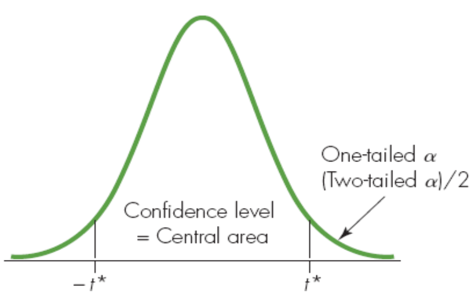
The generic expression for this multiplier when you are working with a standard normal distribution is given by z^*.

Here are a few other multipliers for a population proportion confidence interval.

Confidence Level	90%	95%	98%	99%
Multiplier z*	1.645	1.96 (or about 2)	2.326	2.576

Now, the easiest way to find multipliers is to actually look ahead a bit and make use of Table A.2. Look at the df row marked *Infinite* degrees of freedom and you will find the z* values for many common confidence levels. Check it out!

Table A.2 t^* Multipliers for Confidence Intervals and Rejection Region Critical Values



	Confidence Level						
df	*.80*	*.90*	*.95*	*.98*	*.99*	*.998*	*.999*
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
				...			
100	1.29	1.66	1.98	2.36	2.63	3.17	3.39
1000	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Infinite	1.281	1.645	1.960	2.326	2.576	3.090	3.291

From Utts, Jessica M. and Robert F. Heckard. Mind on Statistics, Fourth Edition. 2012. Used with permission.

When the confidence level increases, the value of the multiplier increases. So the width of the confidence interval also increases. In order to be more confident in the procedure (have a procedure with a higher probability of producing an interval that will contain the population value, we have to sacrifice and have a wider interval. The formula for a confidence interval for a population proportion p is summarized next.

Confidence Interval for a Population Proportion *p*: $\hat{p} \pm z^{*}\text{s.e.}(\hat{p})$

where $\hat{p}$ is the sample proportion and z^* is the appropriate multiplier

and s.e.($\hat{p}$) = $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ is the standard error of the sample proportion.

Conditions:

1. The sample is a randomly selected sample from the population. However, available data can be used to make inferences about a much larger group if the data can be considered to be representative with regard to the question(s) of interest.
2. The sample size *n* is large enough so that the normal curve approximation holds $np \geq 10$ and $n(1-p) \geq 10$

Try It! A 90% CI for *p*

A random sample of *n* = 501 American teenagers resulted in 54% stating they get along very well with their parents. The standard error for this estimate was found to be 2.2%. The 95% confidence interval for the population proportion of teenagers that get along very well with their parents went from 49.6% to 58.4%. The corresponding 90% confidence interval would go from 50.4% to 57.6%, which is indeed narrower (but still centered around the estimate of 54%).

The Conservative Approach

From the general form of the confidence interval, the margin of error is given as:

$$\text{Margin of error} = z^* \text{ s.e.}(\hat{p}) = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

For any fixed sample size *n*, this margin of error will be the largest when $\hat{p}$ = ½ = 0.5. Think about the function $\hat{p}(1-\hat{p})$. So using ½ for $\hat{p}$ in the above margin of error expression we have:

$$z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = z^*\sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}} = \frac{z^*}{2\sqrt{n}}$$

By using this margin of error for computing a confidence interval, we are being *conservative*. The resulting interval may be a little wider than needed, but it will not err on being too narrow. This leads to a corresponding **conservative confidence interval** for a population proportion.

Conservative Confidence Interval for a Population Proportion *p*

$$\hat{p} \pm \frac{z^*}{2\sqrt{n}}$$

where $\hat{p}$ is the sample proportion and z^* is the appropriate multiplier.

Earlier we saw the margin of error for a proportion was given as $1/\sqrt{n}$. This is actually a **95% conservative margin of error**. What happens to the conservative margin of error in the box above when you use z^* = 2 for 95% confidence?

Choosing a Sample Size for a Survey

The choice of a sample size is important in planning a survey. Often a sample size is selected (using the conservative approach) that such that it will produce a desired margin of error for a given level of confidence. Let's take a look at the conservative margin of error more closely.

(Conservative) Margin of Error = *m* = $\dfrac{z^*}{2\sqrt{n}}$

Solving this expression for the **sample size** n we have: $n = \left(\dfrac{z^*}{2m}\right)^2$

If this does is not a whole number, we would round up to the next largest integer.

Try It! Coke versus Pepsi

A poll was conducted in Canada to estimate *p*, the proportion of Canadian college students who prefer Coke over Pepsi. Based on the sampled results, a 95% conservative confidence interval for *p* was found to be (0.62, 0.70).

a. What is the margin of error for this interval?

b. What sample size would be necessary in order to get a conservative 95% confidence interval for *p* with a margin of error of 0.03 (that is, an interval with a width of 0.06)?

c. Suppose that the same poll was repeated in the United States (whose population is 10 times larger than Canada), but four times the number of people were interviewed. The resulting 95% conservative confidence interval for *p* will be:
   - twice as wide as the Canadian interval
   - 1/2 as wide as the Canadian interval
   - 1/4 as wide as the Canadian interval
   - 1/10 as wide as the Canadian interval
   - the same width as the Canadian interval

Using Confidence Intervals to Guide Decisions

Think about it: A value that is not in a confidence interval can be rejected as a likely value of the population proportion. A value that is in a confidence interval is an "acceptable" possibility for the value of a population proportion.

Try It! Coke versus Pepsi

Recall the poll conducted in Canada to estimate *p*, the proportion of Canadian college students who prefer Coke over Pepsi. Based on the sampled results, a 95% conservative confidence interval for *p* was found to be (0.62, 0.70). Do you think it is reasonable to conclude that a majority of Canadian college students prefer Coke over Pepsi? Explain.

Population Proportion	
Parameter	p
Statistic	$\hat{p}$
Standard Error $\text{s.e.}(\hat{p}) = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$	
Confidence Interval $\hat{p} \pm z^* \text{s.e.}(\hat{p})$ **Conservative Confidence Interval** $\hat{p} \pm \dfrac{z^*}{2\sqrt{n}}$	
Large-Sample *z*-Test $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	
Sample Size $n = \left(\dfrac{z^*}{2m}\right)^2$	

Additional Notes

A place to … jot down questions you may have and ask during office hours, take a few extra notes, write out an extra problem or summary completed in lecture, create your own summary about these concepts.

Stat 250 Gunderson Lecture Notes
5: Learning about a Population Proportion

Part 3: Testing about a Population Proportion

We make decisions in the dark of data. -- *Stu Hunter*

Overview of Testing Theories

We have examined statistical methods for estimating the population proportion based on the sample proportion using a confidence interval estimate. Now we turn to methods for testing theories about the population proportion. The hypothesis testing method uses data from a sample to judge whether or not a statement about a population is reasonable or not. We want to test theories about a population proportion and we will do so using the information provided from a sample from the population.

Basic Steps in Any Hypothesis Test

Step 1: Determine the null and alternative hypotheses.

Step 2: Verify necessary data conditions, and if met, summarize the data into an appropriate test statistic.

Step 3: Assuming the null hypothesis is true, find the *p*-value.

Step 4: Decide whether or not the result is statistically significant based on the *p*-value.

Step 5: Report the conclusion in the context of the situation.

Formulating Hypothesis Statements

Many questions in research can be expressed as which of two statements might be correct for a population. These two statements are called the **null** and the **alternative hypotheses**.

The **null hypothesis** is often denoted by $\mathbf{H_0}$**,** and is a statement that there is **no effect, no difference, that nothing has change or nothing is happening**. The null hypothesis is usually referred to as the ***status quo***.

The **alternative hypothesis** is often denoted by $\mathbf{H_a}$**,** and is a statement that **there is a relationship, there is a difference, that something has changed or something is happening**.

Usually the researcher hopes the data will be strong enough to reject the null hypothesis and support the **new theory** in the alternative hypothesis.

It is important to remember that the null and alternative hypotheses are statements about a population parameter (not about the results in the sample). Finally, there will often be a **"*direction of extreme*"** that is indicated by the alternative hypothesis. To see these ideas, let's try writing out some hypotheses to be put to the test.

Try It! Stating the Hypotheses and defining the parameter of interest

1. About 10% of the human population is left-handed. Suppose that a researcher speculates that artists are more likely to be left-handed than are other people in the general population.

    H_0:

    let ________ =____________________________________

    H_a: parameter = written description

    Direction:

2. Suppose that a pharmaceutical company wants to be able to claim that for its newest medication the proportion of patients who experience side effects is less than 20%.

    H_0:

    let ________ =____________________________________

    H_a: parameter = written description

    Direction:

3. The US Census reports that 48% of households have no children. A random sample of 500 households will be taken to assess if the population proportion has changed from the Census value of 0.48.

    H_0:

    let ________ =____________________________________

    H_a: parameter = written description

    Direction:

Notes:

1. When the alternative hypothesis specifies a single direction, the test is called a **one-sided or one-tailed hypothesis test**. In practice, most hypothesis tests are one-sided tests because investigators usually have a particular direction in mind when they consider a question.

2. When the alternative hypothesis includes values in both direction from a specific standard, the test is called a **two-sided or two-tailed hypothesis test**.

3. A generic null hypothesis could be expressed as **H_0: population parameter = null value,** where the null value is the specific number the parameter equals if the null hypothesis is true. In all of the examples above, the population parameter is *p*, the population proportion. Example 1 above has the null value of 10% or 0.10.

The Logic of Hypothesis Testing: What if the Null is True?

Think about a jury trial …

H_0: The defendant is _______________ **H_a:** The defendant is _______________

We assume that the null hypothesis is true until the sample data conclusively demonstrate otherwise. We assess whether or not the observed data are consistent with the null hypothesis (allowing reasonable variability). If the data are "unlikely" when the null hypothesis is true, we would reject the null hypothesis and support the alternative theory.

The Big Question we ask: If the null hypothesis is true about the population, what is the probability of observing sample data like that observed (or more extreme)?

Reaching Conclusions about the Two Hypotheses

We will be deciding between the two hypotheses using data. The data is assumed to be a **random sample** from the population under study.

The data will be summarized via a ________________________________. In many cases the test statistic is a *standardized statistic* that measures the distance between the sample statistic and the null value in standard error units.

$$\textbf{Test Statistic} = \frac{\textbf{Sample Statistic} - \textbf{Null Value}}{\textbf{(Null) Standard Error}}$$

In fact, our first test statistic will be a z-score and we are already familiar with what makes a z-value unusual or large.

With the test statistic computed, we quantify the compatibility of the result with the null hypothesis with a probability value called the *p*-value.

The ***p*-value** is computed by assuming the null hypothesis is true and then determining the probability of a result as extreme (or more extreme) as the observed test statistic in the direction of the alternative hypothesis.

Notes:

(1) The *p*-value is a probability, so it must be between 0 and 1. It is really a conditional probability – the probability of seeing a test statistic as extreme or more extreme than observed **given (or conditional on)** the null hypothesis is true.

(2) The *p*-value is **not** the probability that the null hypothesis is true.

The ____________________ the *p*-value,
the stronger the evidence is **AGAINST H_0** (and in favor of **H_a**).

Common Convention: Reject **H_0** if the *p*-value is ___________________________.

This borderline value is called the ____________________ and denoted by _____.

When the *p*-value is ≤ α, we say the result is ______________________________.

Common levels of significance are: _________________________________

Two Possible Results:

- The ***p*-value is ≤ α,** so we **reject H_0**
  and say the **results are statistically significant** **at the level α**.
  We would then write a real-world conclusion to explain what 'rejecting H_0' translates to in the context of the problem at hand.

- The ***p*-value is > α**, so we **fail to reject H_0**
  and say the **results are not statistically significant at the level α.**
  We would then write a real-world conclusion to explain what 'failing to reject H_0' translates to in the context of the problem at hand.

Be careful: we say "fail to reject H_0" and not "accept H_0" because the data do not prove the null hypothesis is true, rather the data were not convincing enough to support the alternative hypothesis.

Testing Hypotheses About a Population Proportion

In the context of testing about the value of a population proportion *p*, the possible hypotheses statements are:

1. H_0**: __________________** **versus** H_a**: ____________________**

2. H_0**: __________________** **versus** H_a**: ____________________**

3. H_0**: __________________** **versus** H_a**: ____________________**

Where does p_0 come from? Sometimes the null hypothesis is written as $H_0: p = p_0$ as we compute the *p*-value assuming the null hypothesis is true, that is, we take the population proportion to be the null value p_0.

The sample data will provide us with an estimate of the population proportion *p*, namely the sample proportion $\hat{p}$. For a large sample size, the distribution for the sample proportion will be:

If we have a normal distribution for a variable, then we can **standardize** that variable to compute probabilities, as long as you have the mean and standard deviation for that statistic. In testing, we assume that the null hypothesis is true, that the population proportion $p = p_0$. So the standardized *z*-statistic for a sample proportion in testing is:

$z =$

If the null hypothesis is true, this *z*-test statistic will have approximately a __________________.

The standard normal distribution will be used to compute the p-value for the test.

Try It! Left-handed Artists
About 10% of the human population is left-handed. Suppose that a researcher speculates that artists are more likely to be left-handed than are other people in the general population. The researcher surveys a random sample of 150 artists and finds that 18 of them are left-handed. Perform the test using a 5% significance level.

Step 1: Determine the null and alternative hypotheses.

H_0: ____________________ H_a: ______________________

where the parameter _____ represents ___________________________

____

Note: The direction of extreme is

Step 2: Verify necessary data conditions, and if met, summarize the data into an appropriate test statistic.

- The data are assumed to be a random sample.
- Check if $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.

Observed test statistic: $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

Step 3: Assuming the null hypothesis is true, find the *p*-value.
The *p*-value is the probability of getting a test statistic as extreme or more extreme than the observed test statistic value, assuming the null hypothesis is true. Since we have a **one-sided test to the right**, toward the larger values …

p-value = probability of getting a *z* test statistic as large or larger than observed, assuming the null hypothesis is true.
=

Step 4: Decide if the result is statistically significant based on the *p*-value.

Step 5: Report the conclusion in the context of the situation.

Aside: The researcher chooses the level of significance α **before** the study is conducted. In our Left-Handed Artists example we had H_0: p = 0.10 versus H_a: p > 0.10.

If only 12 LH artists in our sample,

we would have $\hat{p}$ = 0.08, we would certainly not reject H_0

With our 18 LH artists in our sample,

our $\hat{p}$ = 0.12, z=0.82, *p*-value=0.206 and our decision was to fail to reject H_0

What if we had 20 LH artists in our sample,

our $\hat{p}$ = 0.133, our z=1.36, *p*-value=0.0868 and our decision would be______________

What if we had 22 LH artists in our sample,

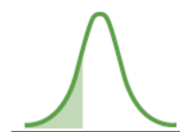
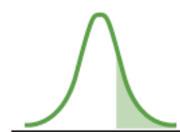
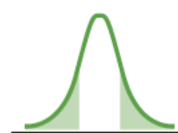
our $\hat{p}$ = 0.147, our z=1.91, *p*-value=0.0284 and our decision would be______________

What if we had 24 LH artists in our sample,

our $\hat{p}$ = 0.16, our z=2.45, *p*-value=0.007, and our decision would be ______________

Selecting the level of significance is like drawing a line in the sand – separating when you will reject H_0 and when there the evidence would be strong enough to reject H_0.

This first test was a one-sided test to the right. How is the *p*-value found for the other directions of extreme? The table below provides a nice summary.

Statement of H_a		*p*-Value Area	Normal Curve Region
$p < p_0$	(less than)	Area to the left of z (even if $z > 0$)	
$p > p_0$	(greater than)	Area to the right of z (even if $z < 0$)	
$p \neq p_0$	(not equal)	$2 \times$ area to the right of $\|z\|$	

Try It! Households without Children

The US Census reports that 48% of households have no children. A random sample of 500 households was taken to assess if the population proportion has changed from the Census value of 0.48. Of the 500 households, 220 had no children. Use a 10% significance level.

Step 1: Determine the null and alternative hypotheses.

H_0: ***p = 0.48*** H_a: ***p ≠ 0.48***

where the parameter *p* represents the population proportion of all households *today* that have no children. Note: The direction of extreme is **two-sided**.

Step 2: Verify necessary data conditions, and if met, summarize the data into an appropriate test statistic.

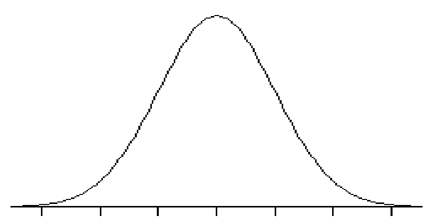
- The data are assumed to be a random sample.
- Check if $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.

Observed test statistic: $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

Step 3: Assuming the null hypothesis is true, find the *p*-value.

The *p*-value is the probability of getting a test statistic as extreme or more extreme than the observed test statistic value, assuming the null hypothesis is true. Since we have a **two-sided test**, both large and small values are "extreme".

Sketch the area that corresponds to the *p*-value.



Compute the *p*-value:

Step 4: Decide if the result is statistically significant based on the *p*-value.

Step 5: Report the conclusion in the context of the situation.

What if n is small?

Goal: we still want to learn about a population proportion p. We take a random sample of size n where n is small (i.e. $np_0 < 10$ or $n(1 - p_0) < 10$). If the sample size is small, we have to go back to the exact distribution for a count X, called the binomial distribution.

> If X has the ***binomial distribution*** $Bin(n, p)$, then
>
> $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0,1,2,\ldots,n$ where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
>
> and the mean of $X = \mu = np$ and the standard deviation of $X = \sqrt{np(1-p)}$

We will use the binomial probability formula for computing the exact p-value.

Testing Hypotheses about a Population Proportion p when n is small

With a small sample size, we will do a Binomial test.

> ***Small-Sample Binomial Test for the population proportion p***
>
> To test the hypothesis H_0: $p = p_0$ we compute the count test statistic
>
> X = the number of successes in the sample of size n
> which has the $Bin(n, p_0)$ distribution when H_0 is true.
>
> This $Bin(n, p_0)$, distribution is used to compute the p-value for the test.

Try It! New Treatment

A group of 10 subjects with a disease are treated with a new treatment. Of the 10 subjects, 9 showed improvement. Test the claim that “a majority” of people using this treatment improved using a 5% significance level. Let p be the true population proportion of people who improve with this treatment.

State the hypotheses: H_0**:** ____________________ H_a**:** __________________

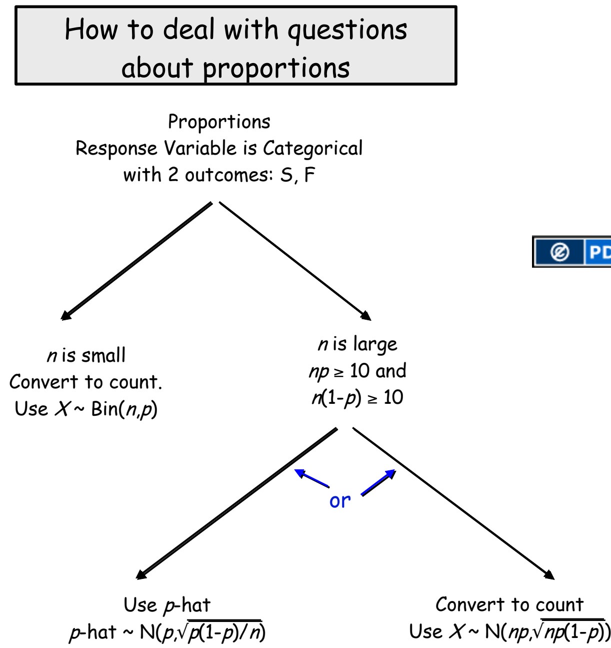
The observed test statistic value is just ____________________

p-value =

At the 5% significance level, we would ______________________ and conclude:

Let's revisit the flow chart for working on problems that deal with a population proportion.

When the sample size is large we can use the large sample normal approximation for computing probabilities about a sample proportion, for testing hypotheses about a population proportion (based on the resulting sample proportion), and for computing a confidence interval estimate for the value of a population proportion (again using the sample proportion as the point estimate).



When the sample size is small we use the binomial distribution to compute probabilities about a sample proportion or to test hypotheses about a population proportion (based on the resulting sample count of successes). We did not discuss the small sample confidence interval for a population proportion using the binomial distribution.

Sample Size, Statistical Significance, and Practical Importance

The size of the sample affects our ability to make firm conclusions based on that sample. With a small sample, we may not be able to conclude anything. With large samples, we are more likely to find statistically significant results even though the actual size of the effect is very small and perhaps unimportant. The phrase **statistically significant** only means that the data are strong enough to reject the null hypothesis. The *p*-value tells us about the statistical significance of the effect, but it does not tell us about the size of the effect.

Consider testing H_0: $p = 0.5$ versus H_a: $p > 0.5$ at $\alpha = 0.05$.
Case 1: 52 successes in a sample of size $n = 100$ → $\hat{p}$ = **0.52**

- Test Statistic: z = (0.52 - 0.50) / $\sqrt{}$ [0.5(1-0.5)/**100**] = 0.4
- *p*-value = $P(Z \geq 0.4) = 0.34$. So we would fail to reject H_0.
- An increase of only 0.02 beyond 0.50 seems inconsequential (not significant).

Case 2: 520 successes in a sample of size $n = 1000$ → $\hat{p}$ = **0.52**

- Test Statistic: z = (0.52 - 0.50) / $\sqrt{}$ [0.5(1-0.5)/**1000**] = 1.26
- *p*-value = $P(Z \geq 1.26) = 0.104$. So we would again fail to reject H_0.
- Here an increase of 0.02 beyond 0.50 is approaching significance.

Case 3: 5200 successes in a sample of size n = 10,000 → $\hat{p}$ = **0.52**

- Test Statistic: z = (0.52 - 0.50) / $\sqrt{}$ [0.5(1-0.5)/**10,000**] = 4.0
- *p*-value = $P(Z \geq 4) = 0.00003$. So we would certainly reject H_0.
- Here an increase of 0.02 beyond 0.50 is very significant!

Small samples make it very difficult to demonstrate much of anything. Huge sample sizes can make a practically unimportant difference statistically significant. Key: determine appropriate sample sizes so findings that are *practically important* become *statistically significant*.

What Can Go Wrong: Two Types of Errors

We have been discussing a statistical technique for making a decision between two competing theories about a population. We base the decision on the results of a random sample from that population. There is the possibility of making a mistake. In fact there are two types of error that we could make in hypothesis testing.

- Type 1 error = rejecting H_0 when H_0 is true
- Type 2 error = failing to reject H_0 when H_a is true

In statistics we have notation to represent the probabilities that a testing procedure will make these two types of errors.

P(Type 1 error) =
P(Type 2 error) =

There is another probability that is of interest to researchers – if there really is something going on (if the alternative theory is really true), what is the probability that we will be able to detect it (be able to reject H_0)? This probability is called the **power of the test** and is related to the probability of making a Type 2 error.

Power = P(rejecting H_0 when H_a is true)
= 1 – P(failing to reject H_0 when H_a is true)
= 1 – P(Type 2 error)
= 1 – β.

So we can think of power as the probability of advocating the "new theory" given the "new theory" is true. Researchers are generally interested in having a test with high power. One dilemma is that the best way to increase power is to increase sample size *n* (see comments below) and that can be expensive.

Comments:

1. In practice we want to protect the status quo so we are most concerned with _____.
2. Most tests we describe have the ...

3. Generally, for a fixed sample size *n*, ...

4. Ideally we want the probabilities of making a mistake to be small, we want the power of the test to be large. However, these probabilities are properties of the procedure (the proportion of times the mistake would occur if the procedure were repeated many times) and not applicable to the decision once it is made.

5. Some factors that influence the power of the test …
   - Sample size: larger sample size leads to higher power.
   - Significance level: larger α leads to higher power.
   - Actual parameter value: a true value that falls further from the null value (in the direction of the alternative hypothesis) leads to higher power (however this is not something that the researcher can control or change).

Simple Example

H_0: Basket has 9 Red and 1 White
H_a: Basket has 4 Red and 6 White

Data: 1 ball selected at random from the basket.

What is the most reasonable Decision Rule?

Reject the null hypothesis if the ball is ____________________

With this rule, what are the chances of making a mistake?

P(Type 1 error) =

P(Type 2 error) =

What is the **power** of the test?

Suppose a ball is now selected from the basket and it is observed and found to be WHITE. What is the decision?

You just made a decision, could a mistake have been made?
If so, which type?

What is the probability that this type of mistake *was made*?

Note: The ***Decision Rule*** stated in the simple example resembles the **rejection region approach to hypothesis testing**. We will focus primarily on the *p*-value approach that is used in reporting results in journals.

Population Proportion
Parameter p
Statistic $\hat{p}$
Standard Error $\text{s.e.}(\hat{p}) = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
Confidence Interval $\hat{p} \pm z^* \text{s.e.}(\hat{p})$ **Conservative Confidence Interval** $\hat{p} \pm \dfrac{z^*}{2\sqrt{n}}$
Large-Sample *z*-Test $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
Sample Size $n = \left(\dfrac{z^*}{2m}\right)^2$

Additional Notes

A place to … jot down questions you may have and ask during office hours, take a few extra notes, write out an extra problem or summary completed in lecture, create your own summary about these concepts.