Lesson 7
Hypothesis Tests for Proportions

Outline of the Lesson

In Lesson 6 we laid the foundation for understanding the basic ideas involved in one type of inferential statistics:

> *Evaluate a claim about a population proportion by measuring the corresponding proportion for a sample taken from that population.*

In that section, we concentrated on a very concrete situation, using a sample consisting of *n* rolls of a pair of dice to evaluate a claim about certain probabilities for that pair of dice. The claims we evaluated in those examples were different ways of evaluating the fairness of a pair of dice, including: (1) the probability of rolling a 7 is 1/6; and (2) the probability of rolling an even number is 1/2. The procedure we used to decide whether to keep or discard the dice is called **hypothesis testing**. Carrying out that procedure is referred to as a **hypothesis test.**

In this lesson, we extend these ideas to develop the general procedures, logic, and notation used in this type of hypothesis testing. We will continue to refer to the dice examples from Lesson 6, and in addition, we will revisit this example first described in Lesson 6:

> A recent report by the Centers for Disease Control states that 16.8% of American adults are smokers. The author was recently taking care of family business in the town where his in-laws live, and felt that he was seeing more people smoking than he usually did in his own home town. This raised the question: Is the proportion of smokers in his in-laws' town 16.8%, or is it higher (or possibly lower)?
>
> To answer this question, one possibility would be to ask every adult in the entire town whether or not they are a smoker. However, that would be very time-consuming and very expensive. The approach taken by statisticians to answer this type of question is to conduct a poll, that is: choose a random sample, measure the proportion of smokers in

that sample, and use the result to make a judgement about the proportion in the entire town.

As you will see, you are actually already familiar with the process used to answer questions such as those posed about smoking for that town. In this lesson, we will adapt the calculations and logic used in Lesson 6 to this new setting, and we will learn about the terminology and notation connected with those processes.

Connecting population proportions to probabilities

The key to our discussion is the connection between proportion and probability. For example, consider the question about the proportion of smokers in a certain town. The question the statisticians would ask the people in the sample is this: "Are you a smoker?" For that question, the *population* the pollsters are interested in consists of all adults in that town. The pollsters want to investigate the proportion of that population that would answer "yes" to that question. The basis for what they do is the fact that these two numbers are the same:

- The proportion of adults in that town who would answer "yes."
- The probability that a randomly selected adult from that town would answer "yes."

As a result, the claim the pollsters are investigating can be thought of in two equivalent ways:

- The proportion of smokers in that town is 16.8% (that is, 0.168).
- The probability that a randomly selected adult from that town is a smoker is 0.168.

The pollsters are going to investigate a claim about a probability, just as we investigated claims about probabilities for dice rolls in Lesson 6. The methods we developed for dice probabilities in Lesson 6 carry over quite nicely to investigating general claims about population proportions. Here is a table summarizing the connections between the first dice example and the question about smoking proportion.

	Dice example	**Smoking example**
Claim being investigated	This pair of dice is fair (the probability of rolling 7 is 1/6).	The proportion of smokers in the town is 16.68%.
Conclusions possible	Discard: The dice are loaded. We found evidence the probability *is not* 1/6. Keep: The dice *might be* fair (the evidence isn't enough to conclude that they are loaded).	Reject the claim: We found evidence the proportion *is not* 16.68%. Do not reject the claim: The proportion *might be* 16.68% (the evidence isn't enough to conclude the claim was incorrect).
Type 1 error	Discard fair dice.	Reject a true claim.
Type 2 error	Keep loaded dice.	"Keep" a false claim (that is, fail to reject that false claim).

Important note. Recall from Lesson 6 that when we keep the dice, we aren't claiming that they *are* fair; we are only acknowledging that there wasn't enough evidence to cause us to discard them – they *might be* fair. Similarly, in the general hypothesis testing methodology of this lesson, we have these possible conclusions: we either reject the original claim, or we acknowledge that the original claim *might be* true. We never assert that the original claim *is* true.

Another important note: For our process of rolling the dice to have any meaning, the rolling of the dice had to be random. Similarly, for the process of selecting individuals for the sample to have any meaning, the selection must be random. Provided the pollsters use a simple random sample for the study, the mathematical results are quite predictable and reliable, just as we saw for the dice rolling examples.

Notation: We have seen the use of the variable "p-hat," written as $\hat{p}$, to stand for the proportion in a sample. Statisticians generally use the variable p (without the hat) to stand for the proportion in an entire population. To help you get used to this notation, we will employ it in the remainder of our discussion. Additional notation and terminology is developed in Section 7.3.

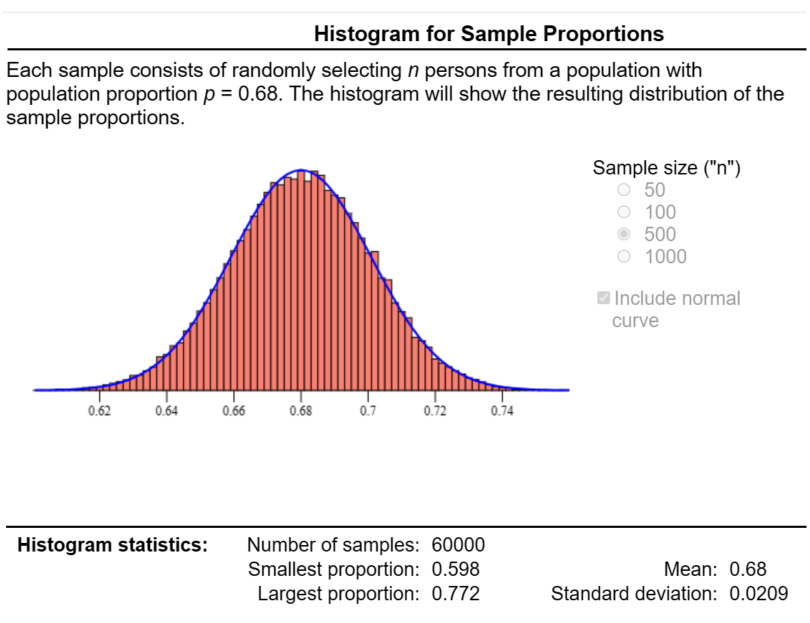
The sampling distribution (review and extension)

As will be true for every hypothesis test strategy we cover, the starting point is to figure out what *should* happen if the claim being investigated is true. This is the so-called ***sampling distribution***.

So, what should happen when we take a simple random sample from a population where the proportion in the entire population is equal to p? The answer (provided n is large enough) is that we expect the proportion in the sample ($\hat{p}$) to be close to the value p – probably not exactly p, but pretty close. If we repeat the experiment over and over, sometimes it will be very close, sometimes not so close. This is illustrated by the applets at this link, which are similar to the ones you used in Lesson 6. (As for the previous apps, the only difference between the two is that the second plots all the histograms, no matter the sample size, on the same scale.)

Histogram of p-hat values (where population proportion is p)
Histogram of p-hat values (where population proportion is p)

Here is a sample run of the first applet obtained by the author, using a sample size of 500 and repeating the experiment for 6000 samples. (Each time the applet is run, it chooses a different value for the population proportion to use. This particular run was based on a population proportion of p = 0.68.)



As was the case for the Lesson 6 sampling distributions, the sampling distribution for sampling proportions appears to be mound-shaped. And, similar to what we learned in Lesson 6, mathematicians have established three facts about this sampling distribution:

- It is approximately normal.
- The mean is the proportion in the claim, which we have labeled with the variable p. In this case, that proportion is 0.68.
- The standard error (standard deviation) can be calculated using the formula $s.e. = \sqrt{\frac{p(1-p)}{n}}$. In this case, that formula yields $s.e. = \sqrt{\frac{0.68(1-0.68)}{500}} = 0.0208614477$.

Exercise 1. Use the applets to experiment. After opening the applet, choose a sample size and use the start sampling / stop sampling buttons to generate about 100,000 samples with an overlaid normal curve. Each time you run the applet, record your answers to these questions:

a. What is the population proportion p for this run?
b. Calculate the standard error for the sampling distribution using the formula $s.e. = \sqrt{\frac{p(1-p)}{n}}$
c. Does the histogram match the normal curve fairly closely?
d. What is the mean for the histogram? Is it close to p?
e. What is the standard deviation for the histogram? Is it close to what you calculated in step b?

We are now ready to apply what we have learned, from the dice examples, to the process of sampling from a large population. Although there are a few subtle differences, with proper sampling methods these differences are very minor and do not affect our final results.

7.1 – The Logic of Hypothesis Testing (Two-Tail Tests)

In this section we develop the formal strategy used to carry out a hypothesis test involving a population proportion. That strategy has already been introduced in Lesson 6, in the context of testing dice for fairness, so this section may be viewed as a review of that lesson. In this section, we concentrate on the overall logic of the strategy, using the dice examples as illustration. In Section 7.2 we apply the strategy to the smoking example mentioned above. In Section 7.3 we fill in the details, including terminology, notation, and assumptions.

Claim to be investigated

The purpose of the hypothesis test is to investigate the claim that the population proportion is equal to some particular value. We will reject this claim if we obtain evidence that the population proportion is not equal to this value – evidence either that it is smaller than claimed or that it is larger than claimed.

For example, in Lesson 6 we first focused on the proportion of 7s, which for fair dice should be 1/6 (which is approximately 16.67%, or 0.1667). The manufacturer's claim that the dice are fair translated to this claim to be investigated:

$$p = 0.1667$$

In later examples, we also considered other proportions, including the proportion of even rolls, which should be 1/2 for fair dice. The corresponding claim to be investigated was written as

$$p = 0.5$$

In general, the claim is written in a form similar to these two examples: population proportion is equal to some particular value. This particular value is commonly written as p_0, so our claim takes the form

$$p = p_0$$

where p_0 is some particular proportion.

The sampling distribution

The decision process is based on considering what should happen if the claim is true, the so-called "sampling distribution" for sample proportions. If we take many, many simple random samples of the same size, calculate the sample proportion for each sample, and create a histogram of these sample proportions, the resulting distribution is approximately normal. Moreover, the mean for this distribution is the population proportion in the claim (designated p_0), and the standard deviation (standard error) is given by the formula $\sqrt{\frac{p_0(1-p_0)}{n}}$. We use these formulas to calculate the mean and standard error for the sampling distribution.

Comment: We saw these formulas in Lesson 6 and earlier, but without the subscripts on the variable *p*. In this lesson and what follows, we will use p_0 to emphasize that the calculations are based on the particular number in the claim being investigated.

For example, if we take a sample of size 1000 to investigate the claim that the proportion of even rolls is 0.5, the sampling distribution has these properties:

- Approximately normal
- Mean is the population proportion in the claim, so the mean is 1/2 or 0.5.
- The standard error is calculated as $se = \sqrt{\frac{0.5(1-0.5)}{1000}}$

Calculations based on a sample

To investigate the claim, we obtain a simple random sample from the given population, and we calculate the proportion for that sample. The sample proportion is written using the notation $\hat{p}$ (read "p-hat").

The sampling distribution tells us what we should expect to see if the claim is true, so we can use the sampling distribution to find out if what we see in the sample supports or contradicts the claim.

Samples that fall reasonably close to the proportion in the claim will support that claim. Samples which fall out in the tails of the sampling distribution will contradict the claim. We therefore need some measure of how far out in the tails a particular sample proportion lies.

Because the sampling distribution is approximately normal, we can quantify this by first using the mean and standard error for the sampling distribution to calculate a *z*-score for our sample, then using that *z*-score to calculate a *P*-value. The *z*-score calculation uses the usual formula for a *z*-score, which we can express in words as

$$z = \frac{data\ value - mean}{standard\ deviation}$$

or in symbols as

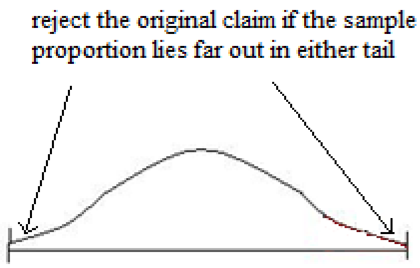
$$z = \frac{\hat{p} - p_o}{se}$$

We then use either Table A or technology to calculate a tail probability (a *P*-value) for this *z*-score. We use a two-tail *P*-value because both unusually small and unusually large sample proportions provide evidence that the claim is incorrect.

The decision: Is there enough evidence to reject the original claim?

In words, the strategy for making a decision is very simple. We will reject the claim if the sample yields a proportion which is nowhere close to the claimed proportion. On the other hand, if the sample's proportion is reasonably close to the claimed proportion, we will acknowledge that that claim might be correct.

In terms of the sampling distribution, this means that we should reject the original claim when we obtain a sample proportion which lies far out in one of the tails of the sampling distribution.



As we have seen, the *P*-value provides a measure of how far out in the tail a particular piece of data lies – the further out in the tail, the smaller the *P*-value. This leads to this summary of the decision-making strategy:

***P*-value**	**Conclusion**	**For dice examples**
Small	Reject claim	Discard dice
Otherwise	Do not reject claim	Keep dice

But, how small is small? As in Lesson 6, our criterion for answering this question is tied to our desire to limit the occurrence of Type 1 errors. Using "less than 0.05" as our criterion for "small" will result in a

5% occurrence of Type 1 errors; similarly, using "less than 0.01" results in a 1% occurrence of Type 1 errors. Although other criteria are certainly possible, these two are the most frequently used by statisticians.

Notes:

1. Which criterion (boundary value) to use – whether 0.05, or 0.01, or perhaps some other choice – must be decided *before* carrying out the study on the sample. It is unethical for a statistician to wait until after calculating the *P*-value to decide on the criterion for smallness.
2. For purposes of this course, you should use "less than 0.05" for your decision if we do not specify some other boundary value.
3. The criterion we use is related to the desire to limit Type 1 errors.

Summary:

- If the calculated *P*-value is less than the chosen boundary value, reject the claim.
- Otherwise, do not reject the claim (acknowledge that the claim *could be* correct). Notice that this conclusion does *not* state that the claim *is* correct, only that it *could be* correct.

7.2 – An Example: Smoking in a Small Town

In this section we apply the general hypothesis test strategy for population proportions to a typical example of the type of situation for which the strategy is appropriate. This example was first described in the previous lesson, and again in the introduction to this lesson.

> A recent report by the Centers for Disease Control states that 16.8% of American adults are smokers. The author was recently taking care of family business in the town where his in-laws live, and felt that he was seeing more people smoking than he usually did in his own home town. This raised the question: Is the proportion of smokers in his in-laws' town 16.8%, or is it higher (or possibly lower)?

To answer this question, statisticians apply inferential statistics – specifically, the hypothesis testing procedure. They take a random sample of the adults in that town, and use the proportion of smokers in that sample to make a statement about the proportion of smokers in the entire town. To illustrate the process, let us suppose the researchers find that 215 out of 1100 adults sampled report being smokers[1]. Using 0.05 as the criterion / boundary value, what conclusion would they draw? (That is, what conclusion would they draw if they wish to limit Type 1 error to 5%?)

Claim to be investigated

The study was triggered by a suspicion that the proportion of smokers in that particular town might not match the 16.8% reported by the CDC for the entire country. We will investigate the claim written symbolically as:

$p = 0.168$ (that is, $p = 16.8\%$)

[1] The examples in these notes generally fall into one of several categories. Some may be taken completely from actual studies reported in the literature. Others are adaptations of actual studies, with the data modified to allow for practice with the methods being studied. Still others are situations which are similar to studies that statisticians do but which are made up by the author. This example falls into the latter category (although the 16.8% figure reported by the CDC is real).

The Form of the Claim to be Investigated

No matter how the situation might be described in the statement of the problem, for a hypothesis test about population proportions the claim to be investigated will always be written as a claim that the population proportion *is equal to* some particular value, that is in the form

$$p = p_0$$

where p_0 is some particular proportion. In this example, the study was triggered by the author's suspicion that the proportion of smokers in that town *is not equal to* 16.8%. Nevertheless, the claim to be investigated is that the proportion *is* 16.8%.

In general, we can describe the claim as some way of saying "no difference" or "no change" or "just as it should be." In this case, the claim states that there is no difference between the smoking rate for this particular town and the rate reported by the CDC for the entire country. Frequently, as in this example, studies are motivated by a researcher's opinion that things are *not* the same, or *not* as they should be. In the dice examples, we started out with a worry/suspicion that some casinos might be using loaded dice, but the claim we investigated was "these dice are fair." The hypothesis test methodology always begins by investigating the claim that the population proportion *is* equal to some specific value. This is true no matter what the researcher's suspicion might be.

The sampling distribution

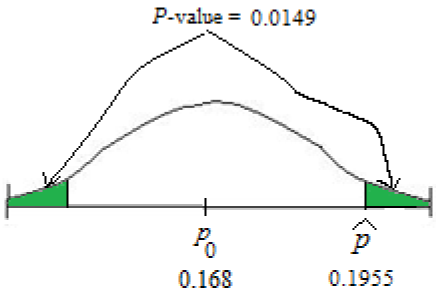
The sample size is n = 1100. Therefore, the sampling distribution is approximately normal, with mean = p_0 = 0.168 and standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.168(1-0.168)}{1100}} = 0.0113$.

Calculations based on a sample

In the sample, 215 out of 1100 adults were smokers. We use this, along with the sampling distribution, to carry out these calculations:

- The sample proportion is $\hat{p} = \frac{215}{1100} = 0.1955$.
- The z-score is $z = \frac{data\ value - mean}{standard\ deviation} = \frac{\hat{p}-p_o}{se} = \frac{0.1955-0.168}{0.0113} = 2.4336$.
- Using technology, the area to the right of this z-score is 0.00747, giving a two-tail P-value of 0.0149. (If you use Table A, or do less rounding, you may obtain a slightly different answer.)

This figure illustrates the situation:



The decision: Is there enough evidence to reject the original claim?

This *P*-value is smaller than 0.05, so we reject the claim. We conclude that the proportion of smokers in this town is not 16.8%. (Based on the sample, we believe it is larger than 16.8%.)

Comment: One way to think of the *P*-value is this: It is the probability, *if the claim is correct*, of obtaining a sample whose sample proportion is at least this far away from the proportion in the claim. That probability is less than 5% (0.05), so we reject the claim.

Was the conclusion correct? When a claim is rejected, there is always the possibility that we have committed a Type 1 error. Perhaps the proportion in the town actually *is* 16.8%, and this just happened to be one of the very unusual samples that might arise from a population proportion of 16.8% (Only 1.49% of all such samples would be as unusual as, or more unusual than, this sample.)

In a real-world study, we never know for sure. We can say, however, that we have used a procedure which limits the occurrence of Type 1 error (rejecting a true claim) to 5%.

Exercise 2: Suppose the researcher had decided, when planning the study, to use 0.01 as the criterion for the decision. Describe the decision that would be made in that case.

Exercise 3: In this exercise, you will carry out the same hypothesis test but using several different hypothetical sample results. For each, show the calculations, and answer the question "Do you reject the claim" using criterion 0.05 and also criterion 0.01.

a. 99 out of 700 are smokers
b. 266 out of 1365 are smokers
c. 68 out of 520 are smokers

	Standard error	$\hat{p}$	z	***P*-value**	**Do you reject claim?**	
					Using 0.05	**Using 0.01**
a.						
b.						
c.						

Here is another example. Except that the context does not involve tossing dice or checking people's smoking habits, the problem is essentially identical to those you have already seen. The steps are the same, using the data supplied in the description of the problem.

Example. Should we believe that a certain population proportion is 0.35? Based on a sample in which 355 people answer yes to the question posed, do an appropriate two-tail test. (The sample consisted of 927 people.) Use 0.05 as the criterion. (That is, you want to keep the probability of a Type 1 error below 5%.) Show all the steps of the process.

Claim being investigated: The population proportion is 0.35. Symbolically,
$p = 0.35$

The sampling distribution, assuming the claim is true.
mean = p_0 (the proportion in the claim) = 0.35

se (standard error) = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.35(1-0.35)}{927}} = 0.0157$

Obtain sample, calculate z-score and P-value.
the sample proportion is $\hat{p} = \frac{355}{927} = 0.3830$
$z = \frac{0.3830-0.35}{0.0157} = 2.1019$
using technology, the two-tail P-value is 0.0356

Decision. We are using 0.05 as the criterion. Since the P-value is less than 0.05, we reject the claim. (Note that a criterion of 0.01 would lead to a different decision in this example.)

The applet at the following link provides additional practice in the calculations for this type of hypothesis test. You will also practice forming a conclusion – reject the claim, or do not reject the claim – based on a specified criterion. For some of the problems, the specified criterion is 0.05 (that is, you wish to keep the Type 1 error rate below 5%). For others, you will use 0.01 as the criterion (to keep the Type 1 error rate below 1%).

Hypothesis tests for population proportions

7.3 – Terminology, Notation, and Assumptions

Assumptions

In this section we will do a quick review of the logic of a two-tail hypothesis test, along with an introduction to additional terminology and notation that is used. Before we begin, however, here are a few "assumptions," that is, conditions which must hold if we wish to use these methods.

- It goes without saying, perhaps, that we are studying a categorical variable – otherwise, it makes no sense to even be talking about proportions.
- The population must be much larger than the sample (typical guidelines vary from specifying at least 10 times as large to at least 20 times as large). For this course, we will generally just assume this to be true – after all, if the entire population is only a little larger than our sample, it would be possible to just ask the question to the entire population.
- The theory is based on taking a simple random sample from the population, or performing a random experiment on a sample taken from the population.

- We will be using the normal distribution in our calculations. For this to be valid, the sample size must be large enough that both these conditions hold:
  - $np_0 \geq 15$
  - $n(1 - p_0) \geq 15$

Example: For the example of Section 7.2:

- The categorical variable is the answer to the question, “Are you a smoker?”
- The sample was size 1100, and we are assuming the town has at least 10 to 20 times the sample size. This would be true for a town of size 25,000 or larger, for example.
- The problem description described using a random sample.
- Since *n* is 1100, and p_0 is 0.168, we have these calculations:
  - $1100(0.168) = 184.8$
  - $1100(1 - 0.168) = 915.2$

  Both numbers are certainly ≥ 15

Hypothesis test logic with terminology and notation

At this point we have completed our description of the calculations involved, and the logic used, for testing hypotheses about population proportions. In the current subsection we will enhance that coverage by introducing some of the terminology and notation that statisticians use when carrying out a hypothesis test about a population proportion. The calculations and logic are the same – the only thing new is the terminology and notation.

Claim to be investigated

The purpose of the hypothesis test is to investigate the claim that the population proportion is equal to some particular value commonly written as p_0. This claim is referred to as the ***null hypothesis***, and the symbol H_0 is used to represent the claim. The null hypothesis is written symbolically in this form:

$$H_0 : p = p_0$$

Along with the null hypothesis, we also write what is called an ***alternative hypothesis***, written as H_a. The alternative hypothesis describes the conclusion we will reach if we reject the null hypothesis. In general, for the two-tail tests we are studying, the alternative hypothesis is of this form:

$$H_a : p \neq p_0.$$

We will reject the null hypothesis (and conclude that the alternative hypothesis is true) if we obtain evidence that the population proportion is not equal to p_0 – evidence either that it is smaller than claimed or that it is larger than claimed.

The purpose of a hypothesis test is to examine whether we believe a claim that a proportion is equal to a particular value. This particular value is written p_0 and the null and alternative hypotheses are written:

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

This particular value p_0 can have a variety of sources.

- In the dice example, it is the theoretical probability for fair dice.
- In the smoking example, it is the nationwide proportion reported by the CDC. We were testing whether that particular town's proportion is the same as the nationwide proportion.
- Another common situation involves testing whether the proportion is still the same as it was at some time in the past.

Note that how a problem is stated does not change the nature of the null hypothesis – it always makes the claim that the proportion *is* equal to some stated value. For the dice example, one person might say the Gaming Commission is investigating whether the dice are fair; another might describe it as testing whether the dice are loaded. In either case, the null hypothesis is that $p = 0.5$.

Examples: For the dice example based on the proportion of even rolls, the null and alternative hypotheses would be

$H_0: p = 0.5$
$H_a: p \neq 0.5$

For the smokers example, we have

$H_0: p = 0.168$
$H_a: p \neq 0.168$

The sampling distribution

The logic of hypothesis testing always relies on the sampling distribution. If the null hypothesis *is* true, what should we expect when we take a sample? If we know what to expect, we can decide whether our sample falls into the "this is what we expected to happen" category or into the "wow, that doesn't seem consistent with the null hypothesis" category.

For proportions, the sampling distribution is approximately $N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$. That is, it is approximately normal, with mean equal to the claimed proportion and standard deviation (standard error) given by the formula $se = \sqrt{\frac{p_0(1-p_0)}{n}}$. If we obtain a sample whose proportion lies far out in either tail of this normal distribution, we will reject the null hypothesis.

Calculations based on a sample

After we take our sample, there are three steps, although of course they can be combined.

1. Calculate the sample proportion $\hat{p}$.
2. Calculate the *z*-score using $z = \frac{\hat{p}-p_0}{se}$, that is "what the sample says, minus what the null hypothesis claimed, divided by the standard error." (This *z*-score is frequently referred to as a ***test statistic***. It is a statistic calculated from the sample, and which is used to test the null hypothesis. In what follows we will frequently refer to this calculation as calculating the test statistic.)
3. Use the test statistic (that is, the *z*-score) to measure how far out in the tails of the sampling distribution this sample lies. This is given by the two-tail *P*-value.

The decision: Is there enough evidence to reject the original claim?

Samples that lie far out in the tails of the sampling distribution have small *P*-values. So if the *P*-value is small, we reject the null hypothesis. How small is small enough? The criterion we use is based on our desire to limit the occurrence of Type 1 error. The most common criteria used are 0.05 and 0.01.

This criterion is referred to as the ***significance level***, and it is symbolized using the Greek letter alpha, written α. So a common value is $\alpha = 0.05$, and another common value is $\alpha = 0.01$. The significance level indicates the probability of making a Type 1 error.

- If *P*-value is less than α, **reject** the null hypothesis, and conclude that the alternative hypothesis is true.

- Otherwise, **fail to reject** the null hypothesis – it might be true, there is not enough evidence to conclude that the alternative hypothesis is true.

Example: A researcher investigating a claim about a population proportion samples 983 people from the population, and 380 of those people answer yes to the question she poses. Evaluate the claim that the population proportion is 37.2%, using a two tail hypothesis test. At a 0.01 significance level, what does the researcher conclude?

Solution: The first step is to write the null hypothesis and the alternative hypothesis:

$$H_0: p = 0.372$$
$$H_a: p \neq 0.372$$

The decision will be based on the sampling distribution, which is approximately normal with:

Mean = p_0 = 0.372

Standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{.372(1-.372)}{983}} = 0.0154$

We calculate the sample proportion – that is, the proportion obtained by asking the question to the sample:

$$\hat{p} = \frac{380}{983} = 0.3866$$

Notice that the sample proportion is definitely NOT exactly equal to the 37.2% (0.372) in the null hypothesis. However, our job is to decide whether this difference is due to just the randomness of the sampling process, or whether it indicates that the null hypothesis should be rejected. To make this decision, we calculate two additional values: the *z* score (test statistic) for this particular sample proportion, and the two-tail *P*-value for that *z* score. A small *P*-value will indicate that this sample lies out in the tail of the sampling distribution, and we will reject the null hypothesis. If the *P*-value is not small, we will not reject the null hypothesis. Here are the calculations (using technology to determine the *P*-value):

$$z = \frac{\hat{p}-p_0}{se} = \frac{0.3866-0.372}{0.0154} = 0.9481$$

P-value = 0.3431

Because our *P*-value (0.3431) is **not** less than the significance level (0.01), we **do not** reject the null hypothesis. We conclude that the population proportion **could be** 37.2%.

Comment: We used the normal distribution in our calculations. For this to be valid, the sample size must be large enough that both these conditions hold:

- $np_0 \geq 15$
- $n(1-p_0) \geq 15$

Here are the calculations that justify this use of the normal distribution:

- $np_0 = 983(0.372) = 365.676$
- $n(1-p_0) = 983(1-0.372) = 617.324$

Both these numbers are definitely ≥ 15, so our use of the normal distribution is justified.

Exercise 4: Should we believe that a certain population proportion is 0.695? Based on a sample in which 755 people answer yes to the question posed, carry out an appropriate two-tail test. (The sample consisted of 1149 people.)

Note: The problem as stated does not specify the significance level you should use. As noted above, when this happens you should use α = 0.05.

The applet at the following link provides additional practice similar to Exercise 4.

Hypothesis tests (calculations and conclusions)

The next applet is similar to the previous applet, but the calculations are done for you – it is up to you to write the null and alternative hypotheses, and to identify the correct conclusion.

Hypothesis tests (conclusions only)

7.4 – Stating the Conclusion, in the Context of the Problem

Our conclusion in general takes one of two forms: (1) We *reject* the null hypothesis; or (2) We *fail to reject* the null hypothesis. One caution we must always keep in mind is that *fail to reject* does not mean the same thing as *accept*. When we fail to reject the null hypothesis, we acknowledge that it *might be* true, which is different from saying that it *is* true. The conclusion we write frequently includes the significance level. For example, we might write sentences such as these:

- At significance level $\alpha = 0.05$, we reject the null hypothesis; or
- We are unable to reject the null hypothesis at significance level 0.01.

This way of writing the conclusion is generic; it applies to any problem. However, we also want to learn how to express the conclusion in terms of the original problem – that is, in terms of dice, or in terms of smokers, etc. The easiest way to do this is to write the conclusion in terms of the alternative hypothesis, as in these sample templates. First, if we reject the null hypothesis:

- At significance level $\alpha = 0.05$ we are able to conclude that ___verbal description of what the alternative hypothesis states__________.
- There was enough evidence, at significance level 0.05, to show that ___verbal description of what the alternative hypothesis states__________.

When we fail to reject the null hypothesis, we simply negate the statement we would make if we rejected the null hypothesis, as in these templates.

- At significance level $\alpha = 0.05$ we are **un**able to conclude that ___verbal description of what the alternative hypothesis states__________.
- There was **not** enough evidence, at significance level 0.05, to show that ___verbal description of what the alternative hypothesis states__________.

Example. In our smoking example of this section and the previous section, the researcher is studying the smoking rate in a certain small town, with the following null and alternative hypotheses:

$$H_0: p = 0.168$$

$$H_a: p \neq 0.168$$

In the sample, 215 out of 1100 adults were smokers, giving a sample proportion of 0.1955, a *z*-score is of 2.4336, and P-value of 0.0149. Write the conclusion, in the context of the study, for these situations:

a. Using significance level 0.05.

Solution: Since 0.0149 is less than 0.05, we reject the null hypothesis, and write:

At significance level $\alpha = 0.05$ we are able to conclude that the proportion of smokers in that town is not 16.8%.

In fact, since the sample proportion was 19.55%, we can go one step further and conclude that the proportion of smokers in that town is *greater than* 16.8%.

b. Using significance level 0.01.

Solution: Since 0.0149 is **not** less than 0.05, we **do not** reject the null hypothesis, and write:

There was not enough evidence, at significance level 0.01, to show that the proportion of smokers in that town is not 16.8%.

That is, the proportion **might be** 16.8%, we cannot conclude that it isn't 16.8%.

Exercise 5: Researchers compare the success rate for a new drug to the known success rate (73%) for an existing drug, using the following null and alternative hypotheses:

$H_0: p = 0.73$

$H_a: p \neq 0.73$

a. If they reject the null hypothesis, how will they write the conclusion in the context of the study?

b. How will they write the conclusion if they do not reject the null hypothesis?

Comment: There is another way statisticians like to write the conclusion which uses the word *significant* or a similar word. For example, if they reject the null hypothesis in the situation of the previous example (on smoking in a small town), they might write, "The smoking rate in that town is significantly different from the CDC's reported 16.8% for the entire nation." If they did not reject the null hypothesis, they would just insert the word *not* as shown here: "The smoking rate in that town is **not** significantly different from the CDC's reported 16.8% for the entire nation."

In this manner of writing the conclusion, it is important to note that the word "significant" simply implies that the difference between the sample and the claim was large enough to reject the claim for the entire population. That is, *significant* implies the null hypothesis was rejected, and *not significant* implies that the null hypothesis was not rejected.

Exercise 6: For the situation described in exercise 5, use the word "significant" to write the conclusion if:

a. they reject the null hypothesis

b. they do not reject the null hypothesis

The applet at the following link provides additional practice interpreting the description of a study, including formulating conclusions based on the data provided.

Formulating conclusions and other interpretation practice

7.5 – More Practice

If you have done all the exercises and used the suggested applets, you have extensive practice with the individual components of hypothesis testing for proportions. In this section, we present complete examples incorporating all the steps of the hypothesis test.

Example: Has the proportion of smokers among adults in the Unites States changed since the 1960s, when it was reported to be 44%? Researchers randomly sample 2075 such adults, and find that 857 are smokers. Carry out a hypothesis test based on this data, using significance level 0.05.

Solution: The first step is to write the null and alternative hypotheses. The null hypothesis will be of the form $p = p_0$, and will represent the idea "no change" or "no difference." That is, it will represent the claim that nothing has changed since the 1960s. The alternative hypothesis will state that the null hypothesis is false.

$H_0: p = 0.44$
$H_a: p \neq 0.44$

Note that these are claims about the population: *all adults in the United States.* In words, the null hypothesis claims that the proportion of smokers among all adults in the United States is still 44%.

To examine these hypotheses, we calculate the proportion of smokers for our sample. If that proportion is "fairly close" to 44%, we will conclude that the population proportion may still be 44%. If not, we will conclude that the population proportion is no longer 44%.

Sample proportion = $\hat{p} = \frac{857}{2075} = 0.4130$

The notion of "fairly close" is based on the sampling distribution, as described in previous sections. Here are the calculations.

Sampling distribution:

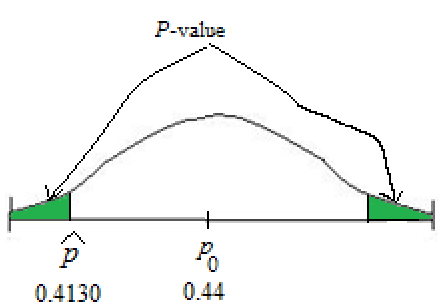
Mean = $p_0 = 0.44$

Standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.44(1-0.44)}{2075}} = 0.0109$

We know the sampling distribution is approximately normal, provided both np_0 and $n(1 - p_0)$ are at least 15. Here are the calculations that justify the use of the normal distribution:

$np_0 = 2075(0.44) = 913$
$n(1 - p_0) = 2075(1 - 0.44) = 1162$

This figure illustrates the situation. The calculated sample proportion is certainly less than the 44% in the null hypothesis. To measure how close it is to the null hypothesis we calculate the illustrated two-tail P-value.



The first step in this calculation is to find the *z*-score. We then use technology to find the two-tail P-value corresponding to this *z*-score.

$z = \frac{\hat{p} - p_0}{se} = \frac{0.4130 - 0.44}{0.0109} = -2.4771$
P-value = 0.0132

Since this P-value is less than the chosen significance level of 0.05, we reject the null hypothesis. We have evidence that the alternative hypothesis is true. We write our conclusion in terms of the alternative hypothesis: ***We have evidence, at α = 0.05, that the proportion of smokers among U.S. adults is no longer 44%.*** Or, using the "significant" terminology: ***The proportion of smokers among U.S. adults differs significantly from the 44% that was reported in the 1960s.***

Example: An existing medication for a certain medical condition is known to provide a cure for 73% of patients treated with the medication. In a study of a newly developed medicine, 951 of the 1250 patients in the study are cured. Carry out an appropriate hypothesis test using significance level 0.01.

Solution: As usual, the null hypothesis will be of the form $p = p_0$, and will represent the idea "no change" or "no difference." In this case, "no difference" implies that the cure rate for the new medicine is no different from that of the existing medication.

$H_0 : p = 0.73$

The alternative hypothesis will state that the null hypothesis is false.

$H_a : p \neq 0.73$

Note that these are claims about the population. But what is that population? Well, the sample consists of 1250 patients treated with the new medication. This sample must have been drawn from a population consisting of all patients who have been (or will be in the future) treated with this new medication. In words, the null hypothesis claims that the proportion of cures among all patients treated with the new medication will be 73% (the same as for the existing medication).

The hypothesis testing method we are using is based on a normal sampling distribution; to justify the use of the normal distribution we calculate both np_0 and $n(1 - p_0)$ – they should be at least 15 to justify the methods we use. Here are the pertinent calculations:

$np_0 = 1250(0.73) = 912.5$
$n(1 - p_0) = 1250(1 - 0.73) = 337.5$

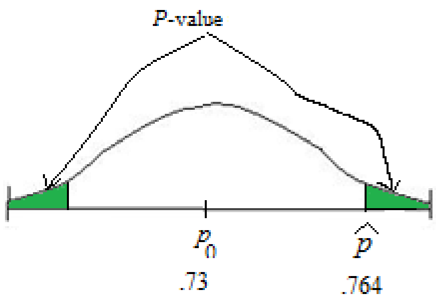
Since both numbers are at least 15, we move on to the calculations for the hypothesis test. First, we calculate the sample proportion – that is, the proportion of cures for the sample:

Sample proportion = $\hat{p} = \frac{951}{1250} = 0.7608$

The test will be based on the sampling distribution whose mean and standard error (that is, standard deviation) are given by these calculations.

Mean = $p_0 = 0.73$
Standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.73(1-0.73)}{1250}} = 0.0126$

This figure illustrates the situation. The calculated sample proportion is certainly more than the 73% in the null hypothesis. To measure how close it is to the null hypothesis we calculate the *z*-score and use that to find the illustrated two-tail P-value (using technology or Table A).



$z = \frac{\hat{p}-p_0}{se} = \frac{0.7608-0.73}{0.0126} = 2.4444$

P-value = 0.0145

Since this P-value is **not** less than the chosen significance level of 0.01, we are unable to reject the null hypothesis (or, put another way, we don't have enough evidence to support the alternative hypothesis). Although the proportion of cures *in this particular sample* was certainly more than 73%, that fact might be just a feature of the inherent randomness of selecting samples – it is conceivable that within the entire population of **all** persons treated with this new medicine the cure rate is 73%.

As always, we write our conclusion in terms of the alternative hypothesis. ***There is not enough evidence, at α = 0.01, to conclude that the proportion of cures for the new medicine is any different from the 73% cure rate for the old medicine.*** Or, using the "significant" terminology: ***The proportion of cures for the new medicine is not significantly different from the 73% cure rate for the old medicine.*** Or, more simply: ***Researchers found no significant difference between the cure rates for the two medicines.***

Summary of hypothesis testing: Although the order in which we describe the calculations may differ somewhat from example to example, in general a hypothesis test for proportions involves the following steps:

- State the null and alternative hypotheses. Remember that the null hypothesis always has the form $H_0: p = p_0$.
- Check assumptions. Some of these are implicit in the statement of the problem, and are therefore not always mentioned – for example, we must be dealing with a categorical variable, and the sample must be randomly chosen from a much larger population. In terms of calculations, this step involves verifying that both np_0 and $n(1-p_0)$ are at least 15.
- Describe the sampling distribution on which the conclusion will be based. The calculations are

  Mean = p_0

  Standard error = $se = \sqrt{\frac{p_0(1-p_0)}{n}}$
- Find the sample proportion:

  $\hat{p} = \frac{count\ for\ sample}{n}$

- Calculate the test statistic (the *z*-score):
  $$z = \frac{\hat{p} - p_0}{se}$$
- Use the test statistic and technology (or Table A) to calculate a two-tail *P*-value.
- Compare the *P*-value to the significance level α to arrive at a conclusion. If the *P*-value is less than α, we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis and acknowledge that the claim in the null hypothesis ***might*** be correct.

Exercise 7: In a large company, 45% of the non-management employees are female, and 55% are male. Management is starting a class to train future managers, and claims to have randomly chosen the participants from the available pool. It turns out that 47 of the 110 chosen are female. Use a hypothesis test to examine management's claim of having randomly chosen the participants.

Exercise 8: A newspaper article claims that 38% of the population of a large city favors a stricter drug control law. To investigate this claim, a statistics student randomly samples 200 residents of the city and finds that 80 favor the stricter law. Carry out an appropriate hypothesis test.

7.6 – One-Tail Tests

Hypotheses for one-tail tests

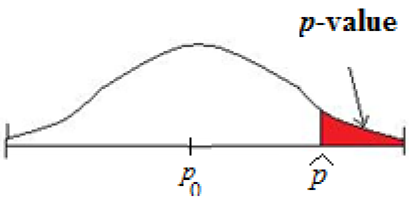
It occasionally happens that the researcher, *prior to taking the sample,* has a preconception of whether the proportion in the null hypothesis is too low or too high. In this case, the researcher may not use the neutral alternative hypothesis

$$H_a: p \neq p_0$$

Instead, the alternative hypothesis will reflect the belief of the researcher. For example, a researcher who believes the stated proportion is too low will write:

$$H_0: p = p_0$$
$$H_a: p > p_0$$

In this case, only samples whose proportion is unusually *large* will be considered to support the alternative hypothesis. Therefore, the *p*-value will be calculated as the probability of obtaining a sample proportion as large as, or larger than, that obtained in the study, as shown in this diagram:

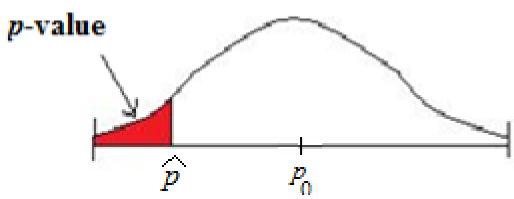


This is a one-tail (right-tail) *p*-value.

On the other hand, if the researcher believes the stated proportion is too high, he or she will write

$$H_0: p = p_0$$
$$H_a: p < p_0$$

and will calculate a one-tail (left-tail) *p*-value as depicted by this diagram:



Comment: We use a one-tail test only if there is some indication that the researcher has a *preconception* that the stated claim is too low or too high. This happens **before** the sample is taken. We do not use the sample's results to decide to do a one-tail test. For a professional statistician, this is never an issue, since the null and alternative hypotheses are established before the actual sampling begins. In this course, the entire study is described in a single paragraph, and if you are not careful you can let the stated results of the survey influence your choice of alternative hypothesis. This is never appropriate: you should use a one-tail test only if the problem statement tells you something about the researcher's *preconceptions*.

Note: There is a movement in the statistical community to do away with, or at least greatly restrict, the use of one-tail tests. The complete discussion of this issue is beyond the scope of this course. In this course, you should use a one-tail test if the problem statement tells you something about the researcher's preconceptions. Otherwise you should use a two-tail test.

Examples: For each of the following situations, describe the null and alternative hypotheses that would be used. Will the researcher calculate a right-tail, left-tail, or two-tail *P*-value?

1. A researcher believes that meditation improves clairvoyance – for example, that immediately following a 30-minute period of meditation your ability to anticipate whether a card from a standard deck of cards is red or black will be improved.

2. A researcher is investigating the claim that 10% of all drivers in a large city fail to wear a seatbelt. In a random sample of drivers in that city, she finds that 8% are not wearing a seatbelt.

3. A new non-addictive pain medication has been developed. The pharmaceutical company developing it hopes that as a side-effect it will reduce the incidence of heart attacks in persons taking the medication. Assume the probability of having a heart attack in the general population is 0.0025.

Solutions:

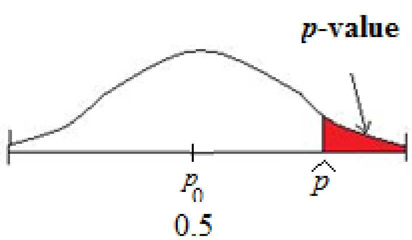
1. The null hypothesis always takes the form "no difference," or "no change," or "no effect." In this case, the null hypothesis asserts that the probability of guessing correctly will be the same as it is without the meditation. Since half the cards in a standard deck of cards are red and half are black, a random guess has a 50% chance of being correct. Thus the null hypothesis is

$$H_0: p = 0.5$$

The statement of the problem indicates a preconception that the ability to guess correctly will be improved by meditation. That is, the researcher believes the probability of a correct guess will be better than 50%. Thus the alternative hypothesis is

$$H_a: p > 0.5$$

Only samples whose sample proportion is larger than 0.5 will support the alternative hypothesis, so the researcher will calculate a right-tail *P*-value, as illustrated here:



2. The null hypothesis states that there is no difference between the percentage in the city and the percentage that has been claimed:

$$H_0: p = 0.1$$

The statement of the problem does not indicate any preconception on the part of the researcher, so we will use a two-tail test and a two-tail *P*-value. The alternative hypothesis is

$$H_a: p \neq 0.1$$

Caution*:* Students are frequently misled by the reported 8% result from the study; since 8% is less than 10%, they might write $H_a: p < 0.1$ as the alternative hypothesis. The only time you should use a one-tail test is if there is a ***pre-conception*** (that is, a belief ***before*** the data is collected) on the part of the researcher.

3. The null hypothesis states that there is "no effect," that the incidence of heart attacks for the population of people taking this drug is / will be the same as for the general population:

$$H_0: p = 0.0025$$

The statement of the problem indicates a preconception that taking the drug will reduce the incidence ("the pharmaceutical company … hopes"). For purposes of this course, you should use this preconception to write the alternative hypothesis for a one-tail (left-tail) test using a left-tail *P*-value:

$$H_a: p < 0.0025$$

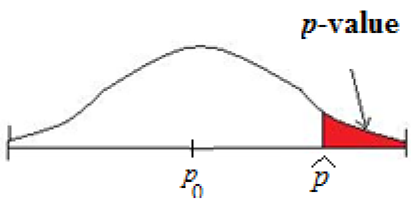
We note, however, that this is a prime example of the type of situation in which professional statisticians today would instead use a two-tail test. The reason is simple – although the company hopes the percentage will be lowered, it could also be raised. Medical ethics would require the pharmaceutical company to investigate both possibilities. A two-tail test is designed to uncover differences from the proportion in the null hypothesis, *in either direction*, and thus this is the type of test which should be carried out.

Calculations for one-tail tests

Once you have chosen to do a one-tail test, and have written the null and alternative hypotheses, the remainder of the process is identical to that for a two-tail test, except that you will calculate a one-tail *p*-value rather than a two-tail *p*-value. As we noted earlier, if the alternative hypothesis has the form

$$H_a: p > p_0$$

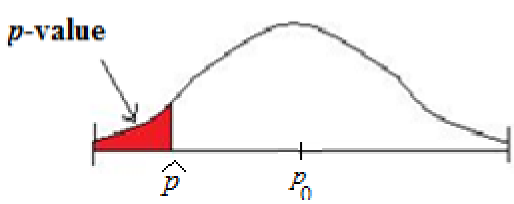
the conclusion will be based on a right-tail *p*-value as shown in this diagram:



Similarly, for alternative hypotheses of the form

$$H_a: p < p_0$$

the decision will be based on a left-tail *p*-value as depicted by this diagram:



Example. Consider the researcher who believes that meditation improves clairvoyance, who has chosen these hypotheses:

$$H_0: p = 0.5$$
$$H_a: p > 0.5$$

He carries out an experiment using a large number of subjects trying to anticipate the color of a large number of cards from standard decks of cards following a period of meditation. Out of 5000 card, the subjects guess correctly for 2587 of the cards. Carry out the appropriate hypothesis test.

Solution. First, the size of the experiment is certainly large enough to use our methods:

$$np_0 = 5000(0.5) = 2500$$
$$n(1 - p_0) = 5000(1 - 0.5) = 2500$$

Both numbers are larger than 15. We therefore use a sampling distribution which is approximately normal, with

Mean = $p_0 = 0.5$
Standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{5000}} = 0.0071$

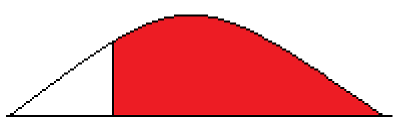
We calculate the sample proportion and the test statistic (*z* score), then use the *z* score to find the right-tail *p*-value using technology:

Sample proportion = $\hat{p} = \frac{2587}{5000} = 0.5174$
$z = \frac{\hat{p}-p_0}{se} = \frac{0.5174-0.5}{0.0071} = 2.4507$
P-value = 0.0071

At either α = 0.01 or α = 0.05, we would reject the null hypothesis, and conclude that for the population of person who have meditated 30 minutes prior to trying to guess cards, the proportion of correct guesses is indeed more than 50%.[2]

Comments:

1. If the researcher expects the sample proportion to be larger than the null hypothesis claim, and it is actually smaller, we *could* calculate a *p*-value. This diagram illustrates the situation:



However, there is really no need. Clearly the *p*-value will not be small; in fact, it will be greater than 50% (0.50).

A similar comment applies for the opposite situation, where the alternative hypothesis claims that p is less than p_0 but the sample proportion is greater than p_0.

2. As always, if we reject the null hypothesis, it means we accept the alternative hypothesis. If we fail to reject the null hypothesis, we acknowledge that the null hypothesis *might* be true.

[2] As noted earlier, some examples in these notes are based loosely on the types of studies which statisticians carry out, but contain data which has been created by the author for illustrative purposes. This example falls in that category.

Exercise 9: For the clairvoyance study, show the calculations and write the conclusion if the study had found that 2048 out of 4000 cards were guessed correctly.

Exercise 10: For the clairvoyance study, what would the researcher conclude if 1493 out of 3000 cards were guessed correctly?

Exercise 11: A researcher is doing a study with the following null and alternative hypotheses:

$H_0: p = 0.27$

$H_a: p < 0.27$

The sample, of size 1027, yields a sample proportion of 240/1027 = 0.2337. What conclusion should the researcher report? Use these steps:

a. Calculate the standard error, $se = \sqrt{\frac{p_0(1-p_0)}{n}}$.

b. Calculate the z score, $z = \frac{\hat{p}-p_0}{se}$.

c. Use technology (or Table A) to find the *p*-value (a left-tail one-tail *p*-value).

d. What is the conclusion, using significance level $\alpha = 0.01$?

The applet at the following link provides additional practice with the calculations for both one-tail and two-tail tests.

Proportion hypothesis test calculations

The next two applets provide additional practice in interpretation for both one-tail and two-tail tests.

Formulating hypotheses and drawing conclusions
Interpretation practice

7.7 – Comments on the Methodology

We began this discussion, in Lesson 6, with two examples. The first dealt with evaluating the fairness of dice by examining the proportion of 7s rolled. The second dealt with the proportion of smokers in a certain town. We are now in a position to answer the questions posed in connection with those first examples.

1. How is it possible to draw conclusions about a group that is larger than the group you actually questioned?

   As we have demonstrated by using the dice applets, there is a certain predictability in carrying out random samples. This is true whether the sampling consists of rolling a pair of dice or of asking people questions. As a result, by making a suitable decision on how far the sample proportion must be from the claimed proportion for us to reject the claim, we can consistently control the likelihood of rejecting a true claim. This is the basis for the logic of hypothesis testing.

2. Is this process legitimate? If so, what precautions must we take in interpreting the result?

   Yes, the process is legitimate, provided we keep a few things in mind. First and most important, the sampling must be *random*. In the case of the dice, the randomness is implicit in the action of throwing the dice. In the case of surveying people, it can be very difficult to obtain true randomness. For example, some people refuse to answer surveys, and some people don't have phones – just two of the many obstacles to obtaining a truly random sample. Ethical professional pollsters utilize a variety of methods for reducing the bad effects of these difficulties, but there is always the possibility of obtaining a faulty result due to lack of randomness in the polling.

   A related problem occurs when a sample is taken from one group, but the results are reported for another group. A recent example occurred when a medical study was done in Singapore, but the results were reported as if the sample had been taken from the entire population of the world. For that particular study, it is possible that the proportion for Singapore and the proportion for the entire world were the same – but it is not obvious on the surface. (Imagine a presidential election poll which samples from only one state but reports the results as valid for the entire country! This would obviously be problematic.)

   Even assuming a perfectly random sampling methodology, with the sample taken from the same population we want to report on, we must take care to hear the results correctly. People in general tend to hear only the conclusion, such as "The proportion of smokers in that town is significantly different from the CDC's reported 16.8% for the nation as a whole." They are unaware, when hearing this, that the method used has a certain probability (perhaps 5%) of generating a Type 1 error. Perhaps worse, they hear the word "significant" as synonymous with "large" when in fact that word is only a statistician's way of stating that a null hypothesis was rejected.

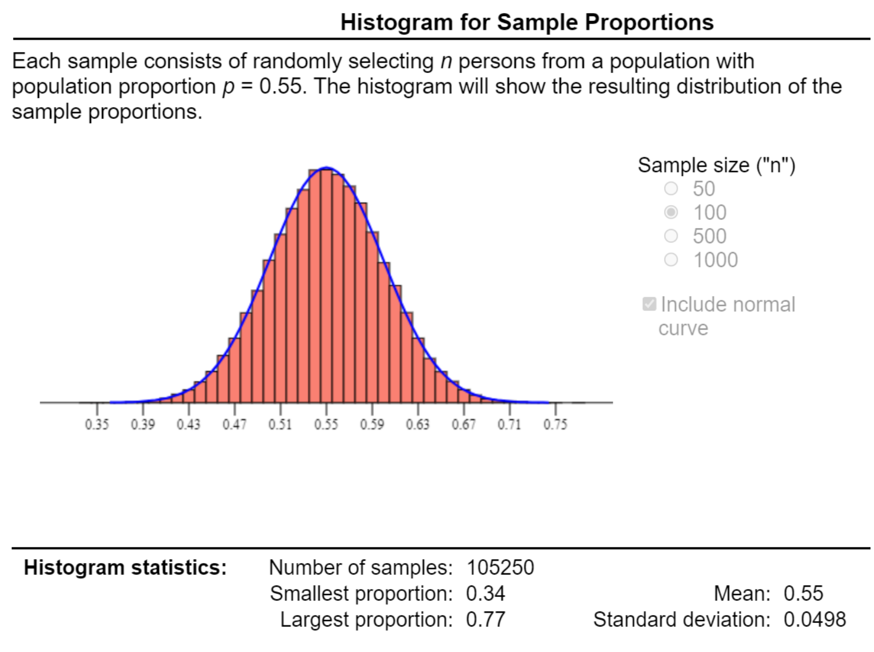
3. What exactly does the phrase "not very close to 16.8%" mean? More generally, exactly how do statisticians make their decision?

   We have covered the technical calculations involved in detail in these lessons. The thing you need to keep in mind as you hear about statistical studies is that the conclusion is always based on some measure of closeness. If the conclusion implies that a null hypothesis has been rejected, there will be a known probability that a Type 1 error has occurred. If, on the other hand, the conclusion implies that the null hypothesis was not rejected, there is in general an unknown probability of having made a Type 2 error.

Solutions to Exercises

1. Use the applets to experiment. After opening the applet, choose a sample size and use the start sampling / stop sampling buttons to generate about 100,000 samples with an overlaid normal curve.

Here is one of the author's runs, using the second applet. Your results should be similar.



Each time you run the applet, record your answers to these questions:

a. What is the population proportion p for this run? 0.55

b. Calculate the standard error for the sampling distribution using the formula $s.e. = \sqrt{\frac{p(1-p)}{n}}$ 0.0497493719

c. Does the histogram match the normal curve fairly closely? yes

d. What is the mean for the histogram? Is it close to p? 0.55. yes

e. What is the standard deviation for the histogram? Is it close to what you calculated in step b? 0.0498. yes

2: Suppose the researcher had decided, when planning the study, to use 0.01 as the criterion for the decision. Describe the decision that would be made in that case.

The P-value is not smaller than 0.01, so we do not reject the claim. We are unable to conclude that the proportion of smokers in this town is not 16.8%. Notice that this is not the same as saying the proportion *is* 16.8%; we simply acknowledge that it *could be* 16.8%.

3: In this exercise, you will carry out the same hypothesis test but using several different hypothetical sample results. For each, write the claim being investigated, show the calculations, and give the final decision (reject the claim, or fail to reject the claim).

- **a.** 99 out of 700 are smokers
- **b.** 266 out of 1365 are smokers
- **c.** 68 out of 520 are smokers

	Standard error	$\hat{p}$	z	*P*-value	Do you reject claim? Using 0.05	Using 0.01
a.	$\sqrt{\frac{.168(1-.168)}{700}}=0.0141$	$\frac{99}{700}=0.1414$	$\frac{.1414-.168}{.0141}=-1.8865$	**0.0592**	**No**	**No**
b.	$\sqrt{\frac{.168(1-.168)}{1365}}=0.0101$	$\frac{266}{1365}=0.1949$	$\frac{.1949-.168}{.0101}=2.6634$	**0.0077**	**Yes**	**Yes**
c.	$\sqrt{\frac{.168(1-.168)}{520}}=0.0164$	$\frac{68}{520}=0.1308$	$\frac{.1308-.168}{.0164}=-2.2683$	**0.0233**	**Yes**	**No**

4: Should we believe that a certain population proportion is 0.695? Based on a sample in which 755 people answer yes to the question posed, carry out an appropriate two-tail test. (The sample consisted of 1149 people.)

First, we state the null and alternative hypotheses:

$H_0: p = 0.695$

$H_a: p \neq 0.695$

The decision will be based on the sampling distribution, which is normal with:

Mean = p_0 = 0.695

Standard error = $\sqrt{\frac{.695(1-.695)}{1149}} = 0.0136$

Here are the calculations based on the sample:

$\hat{p} = \frac{755}{1149} = 0.6571$

$z = \frac{0.6571-0.695}{0.0136} = -2.7868$

P-value = 0.0054

Conclusion:

Because 0.0054 < 0.05, we reject the null hypothesis, and conclude that the population proportion is not 0.695. (Notice that we would have reached the same conclusion if we had used α = 0.01.)

Note: Here are the calculations that justify the use of the normal distribution; both numbers calculated are definitely ≥ 15.

1149(0.965) = 798.555

1149(1 – 0.965) = 350.445

5: Researchers compare the success rate for a new drug to the known success rate (73%) for an existing drug, using the following null and alternative hypotheses:

$H_0: p = 0.73$

$H_a: p \neq 0.73$

a. If they reject the null hypothesis, how will they write the conclusion in the context of the study? We are able to conclude that the success rate for the new drug is not 73%. Or perhaps: The success rate for the new drug differs from the 73% success rate of the established drug.

b. How will they write the conclusion if they do not reject the null hypothesis? There was not enough evidence to show that the success rate for the new drug is not 73%. Or perhaps: We could not conclude that the success rate for the new drug is any different from that of the established drug.

6: For the situation described in exercise 4, use the word "significant" to write the conclusion if:

a. they reject the null hypothesis. The success rate for the new drug differs significantly from the 73% rate for the existing drug. (Or perhaps: Researchers found a significant difference between the new drug and the existing drug.)

b. they do not reject the null hypothesis. The success rate for the new drug does not differ significantly from the 73% rate for the existing drug. (Or perhaps: Researchers found no significant difference between the new drug and the existing drug.)

7: In a large company, 45% of the non-management employees are female, and 55% are male. Management is starting a class to train future managers, and claims to have randomly chosen the participants from the available pool. It turns out that 47 of the 110 chosen are female. Use a hypothesis test to examine management's claim of having randomly chosen the participants.

- State the null and alternative hypotheses.
  $H_0: p = 0.45$ (In words: in the population of all persons that would ever be chosen for the class, 45% would be female – that is, the probability of a female being chosen is 45%.)
  $H_a: p \neq 0.45$
- Check assumptions. Verify that both np_0 and $n(1-p_0)$ are at least 15.
  np_0= 110(0.45) = 49.5
  $n(1-p_0) = 110(1-0.45) = 60.5$
- Describe the sampling distribution on which the conclusion will be based. The calculations are
  Mean = $p_0 = 0.45$
  Standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.45(1-0.45)}{110}} = 0.0474$
- Find the sample proportion:
  $\hat{p} = \frac{count\ for\ sample}{n} = \frac{47}{110} = 0.4273$
- Calculate the test statistic (the z-score), and use the test-statistic to calculate a two-tail P-value.
  $z = \frac{\hat{p}-p_0}{se} = \frac{0.4273-0.45}{0.0474} = -0.4789$
  Using technology, P-value = 0.6320
- Compare the *P*-value to the significance level α to arrive at a conclusion.
  The problem does not state a significance level, so we use α = 0.05. Since 0.6320 is *not* less than 0.05, we fail to reject the null hypothesis and acknowledge that the claim in the null hypothesis ***might*** be correct. (Note that the conclusion would be the same at α = 0.01.)
  In the context of the problem: At α = 0.05, there is no evidence that the long-term proportion of females that would be chosen differs from the 45% of females in the company.

8: A newspaper article claims that 38% of the population of a large city favor a stricter drug control law. To investigate this claim, a statistics student randomly samples 200 residents of the city and finds that 90 favor the stricter law. Carry out an appropriate hypothesis test.

- State the null and alternative hypotheses.
  $H_0: p = 0.38$ (In words: in the entire city, 38% favor the stricter drug control law)
  $H_a: p \neq 0.38$
- Check assumptions. Verify that both np_0 and $n(1-p_0)$ are at least 15.
  np_0= 200(0.38) = 76
  $n(1-p_0) = 200(1-0.38) = 124$
- Describe the sampling distribution on which the conclusion will be based. The calculations are
  Mean = $p_0 = 0.38$
  Standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.38(1-0.38)}{200}} = 0.0343$
- Find the sample proportion:
  $\hat{p} = \frac{count\ for\ sample}{n} = \frac{90}{200} = 0.4500$
- Calculate the test statistic (the *z*-score), and use the test-statistic to calculate a two-tail P-value.
  $z = \frac{\hat{p}-p_0}{se} = \frac{0.4500-0.38}{0.0343} = 2.0408$
  Using technology, P-value = 0.0413
- Compare the *P*-value to the significance level α to arrive at a conclusion.
  The problem does not state a significance level, so we use α = 0.05. Since 0.0413 is less than 0.05, we reject the null hypothesis. (Note that the conclusion would be different at α = 0.01.)
  In the context of the problem: At α = 0.05, there is evidence that the proportion of persons in the city who favor stricter drug control laws is *not* the 38% claimed in the newspaper article. (Based on the sample, we believe it is higher than 38%.)

9: For the clairvoyance study, show the calculations and write the conclusion if the study had found that 2048 out of 4000 cards were guessed correctly.

- Check assumptions. Verify that both np_0 and $n(1-p_0)$ are at least 15.
  np_0= 4000(0.5) = 2000
  $n(1-p_0) = 4000(1-0.5) = 2000$
- Describe the sampling distribution on which the conclusion will be based. The calculations are
  Mean = $p_0 = 0.5$
  Standard error = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{4000}} = 0.0079$
- Find the sample proportion:
  $\hat{p} = \frac{count\ for\ sample}{n} = \frac{2048}{4000} = 0.5120$
- Calculate the test statistic (the *z*-score), and use the test-statistic to calculate a right-tail P-value.
  $z = \frac{\hat{p}-p_0}{se} = \frac{0.5120-0.5}{0.0079} = 1.5190$
  Using technology, P-value = 0.0644
- Compare the *P*-value to the significance level α to arrive at a conclusion.
  The problem does not state a significance level, so we use α = 0.05. Since 0.0644 is not less than 0.05, we do not reject the null hypothesis.
  In the context of the problem: At α = 0.05, there is not enough evidence to conclude that the proportion of correct guesses for person who have meditated exceeds 50%. More briefly, we conclude that meditation does not significantly improve clairvoyance.

10: For the clairvoyance study, what would the researcher conclude if 1493 out of 3000 cards were guessed correctly?

We could carry out calculations similar to those in Exercise 8, but there really is no need to. Since 1493 is less than half of 3000, this study does not support the alternative hypothesis. The right-tail *p*-value, if calculated, would be large – in fact, more than 0.5 = 50%. We fail to reject the null hypothesis.

11. A researcher is doing a study with the following null and alternative hypotheses:

$$H_0: p = 0.27$$

$$H_a: p < 0.27$$

The sample, of size 1027, yields a sample proportion of 240/1027 = 0.2337. What conclusion should the researcher report? Use these steps:

a. Calculate the standard error, $se = \sqrt{\frac{p_0(1-p_0)}{n}}$. $se = \sqrt{\frac{0.27(1-0.27)}{1027}} = 0.0139.$

b. Calculate the z score, $z = \frac{\hat{p}-p_0}{se}$. $z = \frac{0.2337-0.27}{0.0139} = -2.6115$

c. Use technology (or Table A) to find the *p*-value (a left-tail one-tail *p*-value). 0.0045

d. What is the conclusion, using significance level $\alpha = 0.01$? Since the *p*-value is less than 0.01, the researcher rejects the null hypothesis, and writes something like this: "At significance level α = 0.01, we found evidence that the population's proportion is less than 27%." Another way to state this would be to state that the population proportion is significantly less than 27%.

Note: if the P-value had been, for example, 0.0317, the conclusion would have been something like this: "At α = 0.01 we were unable to conclude that the population's proportion is less than 27%." Another way to state this would be to state that the population proportion is **not** significantly less than 27%.