

Chapter 2

Hypothesis tests for a single population mean when σ is known

“Dubium sapientiae initium
(Doubt is the origin of wisdom)”

René Descartes

In this chapter we will introduce the idea of hypothesis testing (or significance testing, depending on who you ask and the philosophical grounds they work from). Hypothesis testing is a form of statistical inference: using information gleaned from a sample to say something about unknown population parameter(s). For example, if your company produces a food product that is boxed, you are required by Canadian law to state the quantity of the food in the box. If you are incorrect (by more than some pre specified amount), then you could face a fine. How do you test whether your stated quantity (a population mean) is equal to that specific value? We will explore this and other tests in this chapter. If you took COMM 104 in Fall 2019, Winter 2020, or Fall 2020, the first part of this chapter will be review for you. We will cover the p-value method and the critical value method.

2.1 The idea behind hypothesis testing

Suppose a company produces boxes of crackers. They claim that there are 454 g of crackers per box. To determine if the claim is true, an inspector may take a random sample of cracker boxes and weigh the crackers. If the sample mean weight is not significantly different from the claimed weight of 454 g, then the company is free to continue their work. However, if the claim is false (if the sample mean weight is significantly different from the claimed weight of 454 g), then the company could be subject to legal penalties under the Consumer Packaging and Labelling Act ¹.

So, what do we mean by ‘significantly different’? We will come back to that point. In this scenario, the inspector is taking evidence from a sample of boxes (in this case, the sample mean weight $\bar{x}$), and using it to say something about an unknown population parameter (in this case, the population mean weight, μ). The inspector is conducting a form of *statistical inference*.

Definition: *Statistical inference* is the process of using information from a sample to make conclusions (inferences) about unknown population parameter(s).

There are two main branches of statistical inference that you will see in business statistics courses: hypothesis testing and confidence intervals. We will cover confidence intervals later in the course.

In the last chapter, you saw how to find a probability related to a sample mean. In this chapter, we will expand on this idea to conduct statistical inference in the form of a hypothesis test.

A hypothesis test is conducted when you want to know the answer to a question about an unknown population parameter. For example:

- What is the population mean weight of a product?
- Do more than 60% of adults in Saskatchewan plan to vote for political party A?

¹For more details, see: https://laws-lois.justice.gc.ca/eng/regulations/C.R.C.,_c._417/index.html

- Is the market share of brand 1 in Canada greater than the market share of brand 1 in the United States? (We will cover this type of scenario in Chapter 8)

In the first example above, the population parameter of interest was the population mean, μ . In the second two examples, the population parameter of interest was the population proportion, p (more on this in later chapters). We will begin by examining hypothesis tests of the population mean μ , when the population standard deviation, σ , is known ².

Remember, unless you have a *census* (a sample that includes every unit in your population), you cannot calculate the value of a parameter; you have to make do with a statistic. In statistical inference, we use a statistic to say something about a parameter.

We will start with five steps to conduct a hypothesis test: (1) state the null and alternative hypotheses; (2) calculate the test statistic; (3) find the p-value or the critical value; (4) use the p-value or the critical value to determine whether you reject the null hypothesis; (5) conclude in the context of the original research question. We will go through each of these steps one by one, and then put them together through some examples. First let's recap the general idea behind hypothesis testing through an example:

Example 1: Suppose we have a sample of 100 boxes of crackers, and we calculate a sample mean weight of 450 g. The company claims that the population mean weight is 454 g. Is this sample mean so far from the claimed value that we doubt the claim? How can we tell?

Solution: We will come back to this later in the chapter. The idea here is to determine how much we *doubt* the claim, based on the evidence from the sample. We will do this through a hypothesis test. ■

After deciding which question we want to ask, we have to turn it into something that we can test statistically. We have to declare the *null and alternative hypotheses*.

²In reality, the population standard deviation is rarely known. Sometimes, if they have quite a lot of historical data on a process, a company will use a historical standard deviation as a placeholder for the population standard deviation. Usually, though, if σ is unknown, s (the sample standard deviation) is used as its estimate and a different type of hypothesis test is used. We will cover this in Chapter 3.

2.2 The null and alternative hypotheses

Suppose a consumer interest research group thinks that a company is short-changing customers by putting less than 454 g of crackers in each box. Then the group would be interested in trying to show that the population mean weight of the crackers per box is **less than** 454 g.

The alternative hypothesis (sometimes called the research hypothesis) is usually the statement that the experimenter wishes to show. In the scenario above, the alternative hypothesis would be that the population mean weight of the crackers per box is under 454 g.

The null hypothesis is the opposite of the alternative hypothesis. It's often called the *status quo hypothesis* or the *hypothesis of no effect*. In this case, the null hypothesis would be that the population mean weight of the crackers per box is greater than or equal to 454 g.

Usually we write the hypotheses using mathematical notation:

$$H_0 : \mu \geq 454 \text{ g}$$

$$H_a : \mu < 454 \text{ g}$$

where we use H_0 to denote the null hypothesis (sometimes verbalized 'H nought', or 'H null') and H_a to denote the alternative hypothesis (we usually just say 'H a', literally the two letters). The hypothesized value of the population mean is called μ_0 ('mu nought'). Here, $\mu_0 = 454\text{g}$.

2.2.1 Properties of the null and alternative hypotheses

There are two main properties of null and alternative hypotheses:

The null and alternative hypotheses are **mutually exclusive** and
collectively exhaustive

By *mutually exclusive*, we mean that both hypotheses cannot be true at the same time. For example, the population mean weight of crackers per box cannot be both greater than or equal to 454 g (the null hypothesis) *and* less than 454 g. It must be one or the other.

By *collectively exhaustive*, we mean that together, both hypotheses cover all possible values of the population parameter. For example, μ can either be greater than or equal to 454 g or less than 454 g. Together, those two possibilities cover all possible values of μ ³.

2.2.2 More examples of null and alternative hypotheses

Say the company thinks that the boxes of crackers are being overfilled, leading to lost revenue. They might be interested in testing whether the boxes contain **more than** the stated 454 g of crackers, on average. That is, they may set up these null and alternative hypotheses:

$$H_0 : \mu \leq 454 \text{ g}$$

$$H_a : \mu > 454 \text{ g}$$

Or suppose they just want to know if their population mean weight per box is **different from** their stated claim of 454g. Then they might test whether μ is different from 454 g.

$$H_0 : \mu = 454 \text{ g}$$

$$H_a : \mu \neq 454 \text{ g}$$

So there are three possible settings for null and alternative hypotheses, summarized in Table 2.1. If H_a has a $>$ sign or a $<$ sign, it is called a one-sided (or one-tailed) alternative hypothesis, because as we will see later only one side of a probability distribution is considered. If H_a has a $\neq$ sign, it is called a two-sided (or two-tailed) alternative hypothesis because both sides are considered.

No matter how the hypotheses are set up, the following statements are always true in this course:

³Hey, math fans: notice this covers the entire real line. We've said nothing about complex numbers, and we won't. But if you saw $\sqrt{i} = -1$ in high school and want to know more, take MATH 379. It'll blow your mind.

Scenario	H_0	H_a	Vocabulary
1	$\leq$	$>$	One-sided (one-tailed, right tailed)
2	$\geq$	$<$	One-sided (one-tailed, left tailed)
3	$=$	$\neq$	Two-sided (two-tailed)

Table 2.1: The three possible scenarios for null and alternative hypotheses.

- The null and alternative hypotheses are opposites;
- The equality always belongs in H_0 , never in H_a .^{4 5}
- The alternative hypothesis is what we want to demonstrate (the thing we are trying to show).
- The hypotheses are always stated in terms of parameters (like μ), never in terms of statistics (like $\bar{x}$).

There is a very good reason for that last statement: you can calculate a statistic exactly. You never have to guess at what value it could be - you can calculate it based on your sample. That means there is never any need to test what value it could be (or what range of values it could take on).

The null hypothesis is assumed to be true at the outset of the experiment. It is up to the experimenter to gather evidence from the sample, and interpret whether there is sufficient evidence to *reject* the null hypothesis. So, once the null and alternative hypotheses are formed, we need a way to express the amount of evidence gathered from the sample. That leads us to the test statistic.

⁴There are some sources that only use an equals sign in H_0 . This is still considered correct, but is not used in this course so as to emphasize that the hypotheses are mutually exclusive and collectively exhaustive.

⁵You may be wondering why μ can be equal to a particular value. We will reject H_0 if μ is *close enough*, in a manner determined statistically, to the hypothesized value, so we are not actually alleging that μ is *exactly* equal to a particular value, just that it's so close to that value as to be indistinguishable from it.

2.3 The test statistic

After formulating the question, we need a way of summarizing the evidence against the null hypothesis. Our first step is to calculate a sample statistic. In a hypothesis test regarding the population mean μ , our sample statistic is the sample mean $\bar{x}$, as this is our best possible estimate of the population mean. Remember we said in the previous chapter that if we are sampling from a normally distributed population (or if $n > 30$), the sample mean has the following probability distribution:

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n}).$$

Also recall that when we calculated probabilities related to the sample mean, we used the following formula to standardize (or normalize) $\bar{X}$:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

In a hypothesis test, we collect a sample, calculate a sample mean $\bar{x}$, and standardize to calculate a z-score ⁶. That z-score is called the *test statistic*, because it's a statistic being used for a hypothesis test. The formula follows the same idea:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}},$$

where z is used instead of Z because we are calculating a particular value of the random variable (a z-score), and μ_0 is the hypothesized value of μ (the value of μ that comes from the null and alternative hypotheses). If the null hypothesis is true, then the test statistic will follow a standard normal distribution.

⁶We only do this if the population standard deviation, σ is known. If it is unknown, we will use a different technique discussed in Chapter 3.

Example 2: Show the first two steps associated with conducting a hypothesis test to determine if the population mean weight of the cracker boxes is less than 454 g. Assume that you have a sample of $n = 9$ boxes and the company knows the weights of the boxes of crackers are normally distributed. They weigh each of the 9 boxes, and find a sample mean weight of $\bar{x} = 450$ g. The company knows from past experience that the population standard deviation (σ) is 4 g.

Solution: Since we want to know if $\mu < 454$ g (from the question statement), we know step 1 - declaring the null and alternative hypotheses - will be:

$$H_0 : \mu \geq 454 \text{ g}$$

$$H_a : \mu < 454 \text{ g}$$

Step 2 is to calculate the test statistic. To do this, we need to show three things: 1) the formula; 2) how we plug in values; 3) the final rounded answer.

$$\begin{aligned} z &= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \\ &= \frac{450 - 454}{4/\sqrt{9}} \\ &= -3.00 \end{aligned}$$

■

Important: z-scores, such as test statistics that have a normal distribution, must always be rounded to 2 decimal places in this course, to match the number of decimal places for z-scores on Table A.4.

Remember, we are trying to gather evidence against the null hypothesis. The null hypothesis is that the population mean weight is 454 g or more. The smallest value that μ can take, if H_0 is true, is 454 g. To see this visually, consider the number line in Figure 2.1.

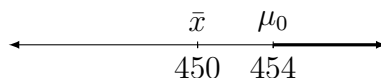


Figure 2.1: A number line showing $\bar{x}$ relative to μ_0 in Example 2. The null hypothesis is true in the bolded region of the number line: this is where $\mu \geq 454$ g. The smallest value that μ can take, if H_0 is true, is 454 g. We use $\mu_0 = 454$ to calculate the test statistic, to be as favourable to the null hypothesis as possible: to ensure that we need the greatest possible evidence to reject it.

This means we would have evidence against the null hypothesis if we see a sample mean that is much less than 454 g, but this is not enough. It is not enough to simply calculate a sample statistic: we need to know how variable that sample statistic is. That is, we need to know if seeing $\bar{x} = 450$ g is typical when we sample $n = 9$ boxes of crackers and the population mean is $\mu = 454$ g. That's where the sampling distribution of the sample mean comes in, and that's why we calculate the test statistic.

So, once the test statistic is calculated, we need a way of knowing whether its value is very likely if the null hypothesis is true. We need to know if this value of z is typically seen if the null hypothesis is correct. That is, is $z = -3.00$ a likely value of Z if the population mean μ is really greater than or equal to 454 g? To answer this question, we must either use the *p-value method* or the *critical value method* (which is also called the *critical value rule*, or the *rejection region method*). We will start with the p-value method.

2.3.1 Back up: why do we use $\mu = \mu_0$ when we calculate the test statistic?

Have another look at Figure 2.1. The sample statistic, $\bar{x}$, is closest to the region in which the null hypothesis is true at the point μ_0 . That is, the null hypothesis is true if μ takes any value at or above 454. So the set of values of μ for which the null hypothesis is true is the set of values at or above 454. If we were to conclude that the null hypothesis is false, when in fact it is true in reality (i.e. if we were to say $\mu < 454$ when in reality $\mu \geq 454$), then we would be making a big mistake. To minimize our chances of making a mistake,

we want to make it as hard as possible to reject H_0 . We reject H_0 when we see a sample mean that is very far away from where we expect it to lie if the null hypothesis is true. So by choosing the point at which $\bar{x}$ and μ are closest (i.e. μ_0), we make it as hard as possible to reject H_0 : we ensure we need a lot of statistical evidence to back up our conclusion.

2.4 The p-value method

We will define a p-value later; for now, think of it as a probability related to a test statistic. Here are our null and alternative hypotheses again, from the example:

$$H_0 : \mu \geq 454 \text{ g}$$

$$H_a : \mu < 454 \text{ g}$$

Remember, we calculated a test statistic of $z = -3.00$.

We want to know the probability that we see $z = -3.00$, or an even smaller value of z , if μ really is at least (i.e. greater than or equal to) 454 g. That is, we want to know the probability that we see a test statistic (z) that gives this much evidence, or even more evidence, against the null hypothesis, if the null hypothesis is really true.

Evidence against the null hypothesis can be summarized as the probability that we see a test statistic as contradictory to the hypothesized value, or even more contradictory, given the null hypothesis is true. That leads us to the definition of a p-value.

Definition: A *p-value* is the probability that you see a test statistic as contradictory to H_0 or even more contradictory, given H_0 is true.

Here, that means (from Table A4):

$$\text{p-value} = P(Z < -3.00) = 0.0013$$

For an illustration, see Figure 2.2.

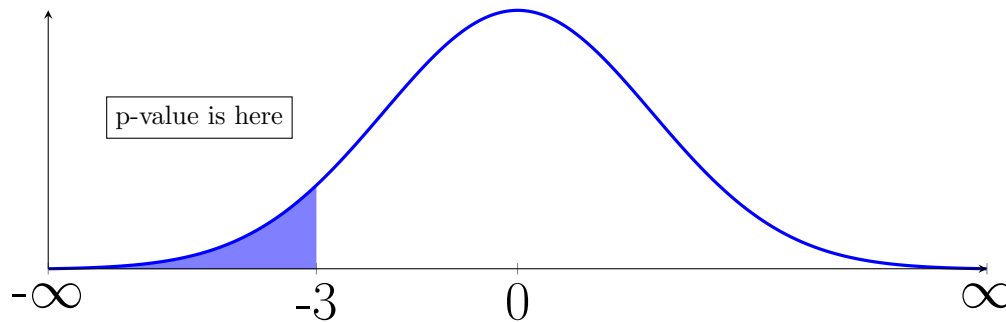


Figure 2.2: Shaded in blue is the p-value, $P(Z < -3.00) = 0.0013$, which means there is a 0.13% chance of seeing a test statistic of -3.00, or even less, if H_0 is true. Since we see a test statistic of -3.00, and we have determined that this is not a probable outcome if μ is really 454 g, we doubt the null hypothesis, so we reject it.

So the probability that we see a test statistic of -3.00 or less, given $\mu \geq 454\text{g}$, is 0.0013 (this is the p-value). That's extremely low! Since this probability is so low, we have *very strong evidence* against the null hypothesis.

Now the question is: how low do we have to go? Before a hypothesis test is set up, the experimenter decides what threshold value (called a *significance level*, denoted α and verbalized 'alpha'), they will consider. If the p-value is less than the significance level, then they will reject H_0 . Unless otherwise specified, the significance level is usually chosen to be 0.05. That is, if the p-value is less than $\alpha = 0.05$, we reject H_0 . This leads us to the general rule of thumb:

If the p-value is less than the significance level α , reject H_0 ,

otherwise do not reject H_0 ⁷.

Think of α as a threshold value. Suppose $\alpha = 0.05$. If the p-value is less than 5%, then it is less than 5% likely that you see a test statistic as far to the tail of the normal distribution, or even farther, given H_0 is true. So since you do see a test statistic this far to the tail, H_0

⁷It is common practice *not* to say "accept H_0 ", and instead say "do not reject H_0 ", partly because there is still some probability that H_0 is not true.

is likely false and should be rejected. Again, although 0.05 is often used, you could instead have a significance level of, say, 0.01. Make sure to read the question carefully.

Statistical significance vs. practical significance

When the null hypothesis is rejected, we say we have a *statistically significant* result. You may have seen that in political studies, or economics: statistical significance just means that we can say with reasonable certainty (depending on α) that the null hypothesis is false. This differs from a *practically significant* result - one that has an immediate business application, that would change a process or product. It is possible to conduct a hypothesis test for the sake of it, and not have anything actionable come of it. It is also possible to take action in the absence of a hypothesis test if the evidence that a change needs to be made is overwhelmingly obvious. A result may be statistically significant but not practically significant, or it may be practically significant but not statistically significant, or both, or neither.

2.4.1 More examples of p-values

Now, what if instead we had a different setup? Consider what would happen if we had wanted to test whether the boxes were being overfilled instead of underfilled. We would have different null and alternative hypotheses:

$$H_0 : \mu \leq 454 \text{ g}$$

$$H_a : \mu > 454 \text{ g}$$

but the test statistic would be the same:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{450 - 454}{4/\sqrt{9}} = -3.00$$

and the p-value would be:

$$\begin{aligned}
 P(Z > -3.00) &= 1 - P(Z \leq -3.00) \\
 &= 1 - 0.0013 \\
 &= 0.9987
 \end{aligned}$$

This is shown in Figure 2.3.

This is saying that the probability that we see a test statistic of -3.00, or even greater, when the population mean μ is really at least 454 g, is 0.9987. Since this probability is very high (much greater than $\alpha = 0.05$), we can be very confident that the population mean weight of crackers per box is really at least 454 g, so we do not reject H_0 .

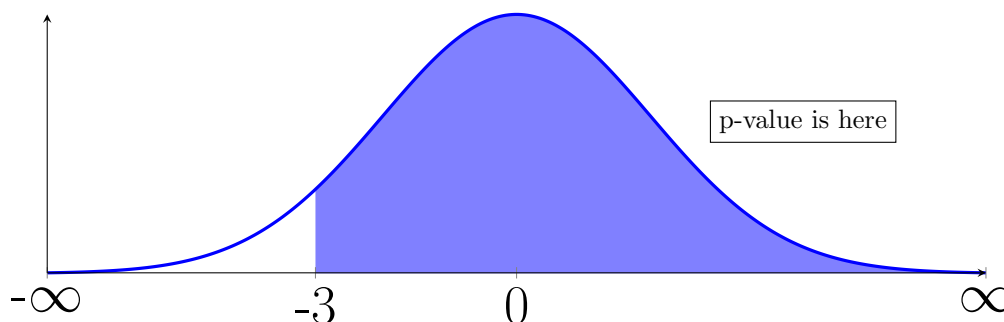


Figure 2.3: Shaded in blue is the p-value, $P(Z > -3.00) = 0.9987$, which means there is a 99.87% chance of seeing a test statistic of -3.00, or even higher, if H_0 is true. Since we see a test statistic of -3.00, and we have determined that this is a probable outcome if $\mu \geq 454$ g, we do not doubt the null hypothesis, so we do not reject it.

Now let's consider one more possibility. Suppose the company were interested in testing whether the population mean weight were different from 454 g. We would have this setup:

$$\begin{aligned}
 H_0 : \mu &= 454 \text{ g} \\
 H_a : \mu &\neq 454 \text{ g}
 \end{aligned}$$

and the test statistic would still be the same:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{450 - 454}{4/\sqrt{9}} = -3.00$$

but now the p-value gets interesting.

If $\mu = 454$ g, then there are two ways in which we would reject the null hypothesis: if we had calculated a test statistic that was very high (giving a small upper tail probability), or very low (giving a small lower tail probability). Another way of saying this is that since we have a $\neq$ sign in the alternative hypothesis, we have to consider both tails of the normal curve. This is visualized in Figure 2.4. In this case, since the p-value is 0.026, which is less than the significance level of 0.05, we would reject the null hypothesis.

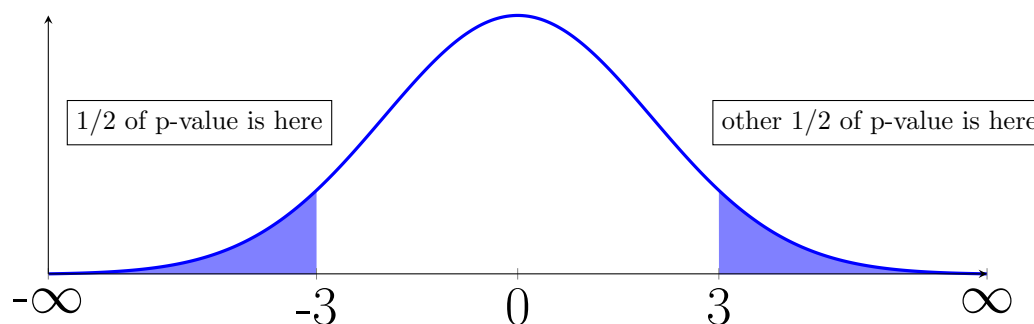


Figure 2.4: Shaded in blue is the p-value. Because we have a $\neq$ sign in H_a , we have to consider what would happen if we calculated a test statistic less than -3.00, or greater than 3.00, because both would give an equal amount of evidence against the null hypothesis. Here, $P(Z < -3.00) = 0.0013$, which is shown on the left. Also, $P(Z > 3.00) = 1 - P(Z \leq 3.00) = 1 - 0.9987 = 0.0013$, shown on the right. Since the normal distribution is symmetric, the area under the curve from $-\infty$ to -3.00 is the same as the area from 3.00 to ∞ : both are 0.0013. So the p-value is $2(0.0013) = 0.0026$.

2.5 Interpreting the results of a hypothesis test

Before we discuss the critical value method, it would be helpful to see some full examples of hypothesis tests. That means we need to discuss the final step. What does rejecting the

null hypothesis mean in a business context?

In general, it is not enough to simply state that you reject the null hypothesis. You must interpret the result of the hypothesis test (reject H_0 or do not reject H_0) in the *context* of the original research question.

In the first example, with a $<$ sign in H_a , our interpretation was: “The population mean weight (μ) of the crackers in a box is likely less than 454g. The boxes are being underfilled.”

Based on this result, the company might change something in the process of boxing the crackers. We have used statistics to make a business-related decision: to change a process. However, it has not yet told us *how* to change it. That’s often an engineering question!

2.6 More examples of hypothesis tests using the p-value method

What follows are full worked examples of hypothesis tests. In this course, to keep things simple, always follow these five steps when you conduct a hypothesis test. We will expand on these later. These steps will be listed in each of the examples that follow regarding p-values:

1. State the null and alternative hypotheses.
2. Calculate the test statistic.
3. Find the p-value.
4. State whether you reject the null hypothesis (H_0), or not, based on the p-value (if the p-value is less than α , reject H_0).
5. Conclude in the context of the original research question.

For your convenience, a summary table of the p-value to calculate in each possible scenario is given in Table 2.2.

Sign in H_a	P-value (note test statistic is denoted z)	Example figure	Worked example
$<$	$P(Z < z)$	Figure 2.2	Example 3a
$>$	$P(Z > z)$	Figure 2.3	Example 3b
$\neq$	$2P(Z > z)$	Figure 2.4	Example 3c

Table 2.2: A summary of the p-value to calculate for each possible sign in the alternative hypothesis. Note that the sign in H_a matches the sign in the probability calculation, and the symbol $|\cdot|$ means ‘absolute value’. For example, if $z = -2.00$, then $|z| = |-2.00| = 2.00$, and if $z = 2.00$, then $|z| = |2.00| = 2.00$.

Example 3: Suppose $\bar{x} = 10$, $\sigma = 3.5$, and $n = 25$.

- Conduct a hypothesis test to determine if μ is less than 12.
- Conduct a hypothesis test to determine if μ is greater than 9.
- Conduct a hypothesis test to determine if μ is different from 9.5.

Solution

a) (1) $H_0 : \mu \geq 12$ $H_a : \mu < 12$

(2) $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10 - 12}{3.5/\sqrt{25}} = -2.86$

(3) $P(Z < -2.86) = 0.0021$

(4) Since the p-value (0.0021) is less than $\alpha = 0.05$, we reject H_0 .

(5) We have enough evidence to reject the hypothesis that the population mean is greater than or equal to 12.

b) (1) $H_0 : \mu \leq 9$ $H_a : \mu > 9$

(2) $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10 - 9}{3.5/\sqrt{25}} = 1.43$

(3)

$$\begin{aligned} P(Z > 1.43) &= 1 - P(Z \leq 1.43) \\ &= 1 - 0.9236 \\ &= 0.0764 \end{aligned}$$

(4) Since the p-value (0.0764) is greater than $\alpha = 0.05$, we do not reject H_0 .

(5) We do not have enough evidence to reject the null hypothesis that the population mean is less than or equal to 9.

c) (1) $H_0 : \mu = 9.5$ $H_a : \mu \neq 9.5$

$$(2) \quad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10 - 9.5}{3.5/\sqrt{25}} = 0.71$$

(3)

$$\begin{aligned} 2P(Z \geq 0.71) &= 2(1 - P(Z < 0.71)) \\ &= 2(1 - 0.7611) \\ &= 0.4778 \end{aligned}$$

(4) Since the p-value (0.4788) is greater than $\alpha = 0.05$, we do not reject H_0 .

(5) We do not have enough evidence to reject the null hypothesis that the population mean is equal to 9.5.

■

Note: The interpretation in part (c) does *not* mean that $\mu = 9.5$; it only means that based on the evidence that we have gathered from the sample, we can't reject the idea that μ could be equal to 9.5.

Sign in H_a	Critical value	Example figure	Worked example
$<$	$z_\alpha = z_{0.05} = -1.645$	Figure 2.5	Example 4a
$>$	$z_\alpha = z_{0.05} = 1.645$	Figure 2.7	Example 4b
$\neq$	$z_{\alpha/2} = z_{0.025} = \pm 1.96$	Figure 2.8	Example 4c

Table 2.3: A summary of the critical value to which to compare the test statistic, z , for each possible sign in the alternative hypothesis. The critical value is z_α for a one-sided alternative hypothesis, and $z_{\alpha/2}$ for a two-sided alternative hypothesis. Note in the second column, we are assuming $\alpha = 0.05$. If instead we had, say, $\alpha = 0.01$, then we would have different critical values. More on this later.

2.7 The critical value method

The *critical value method*, also called the *critical value rule* or *rejection region method*, does not rely on a probability calculated based on a test statistic; rather, it relies on a value along the horizontal axis of the distribution of the test statistic.

Recall that in the p-value method, we reject the null hypothesis if the p-value is less than α . The value of z that puts α in the tail is called the *critical value*. It's a z-score!

Definition: A *critical value* is a value on the horizontal axis of the probability distribution of a test statistic that does one of three possible things depending on the sign in H_a :

- puts α in the left hand tail (if there is a $<$ sign in H_a);
- puts α in the right hand tail (if there is a $>$ sign in H_a), or
- puts $\alpha/2$ in each tail (if there is a $\neq$ sign in H_a).

To distinguish the critical value z-score from the test statistic z-score, we call the critical value z_α ('zed alpha') for a one-tailed alternative hypothesis, and $z_{\alpha/2}$ ('zed alpha over 2') for a two-tailed H_a . We will go through each of these in turn.

2.7.1 When H_a has a $<$ sign

When the alternative hypothesis has a $<$ sign, as in Figure 2.5, we will reject the null hypothesis if the p-value is less than 0.05. So our critical value $z_\alpha = z_{0.05}$ will be the

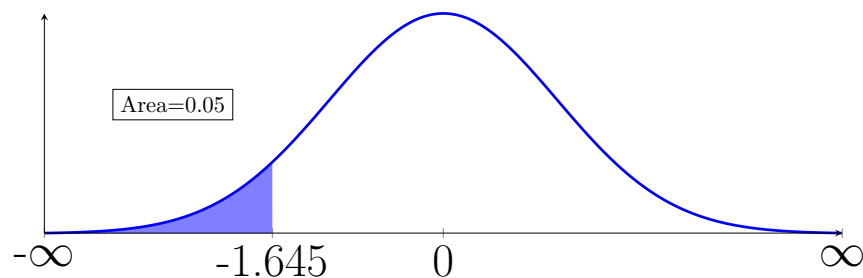


Figure 2.5: If H_a has a $<$ sign, then we are interested in the left hand tail of the standard normal distribution. $P(Z < -1.645) = 0.05$, so if we were to calculate a test statistic less than -1.645 , we would reject the null hypothesis. We call the region of the number line from $-\infty$ to -1.645 the *rejection region* or *critical region* of the test.

z-score that puts exactly 0.05 in the left tail.

To find that z-score, we look up 0.0500 in the middle of Table A4 (where the probabilities are). However, we see quickly that this value does not exist. See Figure 2.6.

<i>z</i>	.03	.04	.05
-3.4	0.0003	0.0003	0.0003
-3.3	0.0004	0.0004	0.0004
-3.2	0.0006	0.0006	0.0006
-1.8	0.0336	0.0329	0.0322
-1.7	0.0418	0.0409	0.0401
-1.6	0.0516	0.0505	0.0495

Figure 2.6: The value 0.0500 does not exist in Table A4; instead we have to average 0.0495 and 0.0505 to get it, so we get both associated z-scores: -1.64 and -1.65 , and average them to get the final z-score (the critical value) $z_{0.05} = -1.645$. Note portions of the table have been removed for ease of visualization.

However, 0.0495 exists, and so does 0.0505. Notice these are both 0.0005 away from the value we want: one is over, and one is under.

By reading ‘out’ to the column and the row, we see $P(Z < -1.64) = 0.0505$ and $P(Z < -1.65) = 0.0495$.

If we average these two values, we get: $(0.0505 + 0.0495)/2 = 0.0500$, which is exactly the probability we want.

We do this with the corresponding z-scores to get the critical value: $(-1.64 + (-1.65))/2 = -1.645$.

So the critical value is $z_{0.05} = -1.645$ ⁸.

2.7.2 When H_a has a $>$ sign

When the alternative hypothesis has a $>$ sign, as in Figure 2.7, we reject the null hypothesis if the area under the curve to the *right* of the test statistic is less than 0.05. Similarly, we see 0.9505 is the probability that Z takes a value less than $z = 1.65$, and 0.9495 is the probability that Z takes a value less than $z = 1.64$. We can average again: $(0.9505 + 0.9495)/2 = 0.9500$, and average the z-scores: $(1.65 + 1.64)/2 = 1.645$, so the area in the right hand tail when $z = 1.645$ is 0.05. That is, $P(Z > 1.645) = 0.05$, so $z_{0.05} = 1.645$.

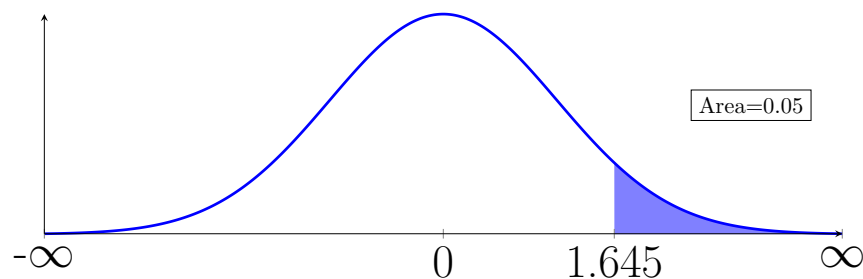


Figure 2.7: If H_a has a $>$ sign, then we are interested in the right hand tail of the standard normal distribution. $P(Z > 1.645) = 0.05$, so if we were to calculate a test statistic greater than 1.645, we would reject the null hypothesis. We call the region of the number line from 1.645 to ∞ the *rejection region* or *critical region* of the test.

Common Question: Why can we say both $z_{0.05} = -1.645$ and $z_{0.05} = 1.645$?

⁸Some of you might be thinking: hang on, this is a linear interpolation, but we have a nonlinear function - why can we do this? The answer is that this average provides a reasonable enough approximation to be getting on with, but you're correct, it's not quite exact.

Answer: The 0.05 refers to the appropriate tail area of the normal distribution. When we say $z_{0.05} = -1.645$, we're referring to the left tail: there is an area of 0.05 between $-\infty$ and -1.645 . When we say $z_{0.05} = 1.645$, we're referring to the right tail: there is an area of 0.05 between 1.645 and ∞ . The notation z_α , whatever α happens to be, is just the shorthand for 'the appropriate one-tailed critical value'.

2.7.3 When H_a has a 'not equal to' sign

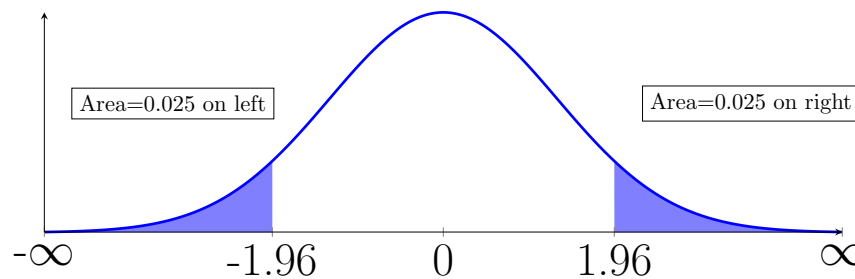


Figure 2.8: If H_a has a $\neq$ sign, then we are interested in both tails of the standard normal distribution. $P(Z > 1.96) = 0.025$, and $P(Z < -1.96) = 0.025$, so if we were to calculate a test statistic greater than 1.96 or less than -1.96, we would reject the null hypothesis. We call the region of the number line from $-\infty$ to -1.96 and from 1.96 to ∞ the *rejection region* or *critical region* of the test.

Finally, when we have a $\neq$ sign in the alternative hypothesis, as in Figure 2.8, we reject the null hypothesis if the area under the left of the negative test statistic, plus the area under the right of the positive version of the test statistic, is less than 0.05. Since we have two areas of interest here, we have two critical values: one negative, and one positive. Remember the normal curve is symmetric, and $0.05/2 = 0.025$, so we want to find the value of z such that $P(Z < -z) = 0.025$ and $P(Z > z) = 0.025$. Now this, we can see directly in the table! If you look up 0.0250, you see $-z = -1.96$, and if you look up $1 - 0.0250 = 0.9750$, you see $z = 1.96$. So our two critical values are 1.96 and -1.96. This means that if your test statistic z lies between -1.96 and 1.96, you will not reject the null hypothesis.

Now we'll re-do Example 3 (calling it Example 4) using the critical value method instead of the p-value method. Here are the steps again as a refresher, when we use the critical value method instead of the p-value method.

1. State the null and alternative hypotheses.
2. Calculate the test statistic.
3. Find the critical value.
4. State whether you reject the null hypothesis (H_0), or not (if the test statistic is further to the appropriate tail than the critical value, reject H_0).
5. Conclude in the context of the original research question.

Example 4: Suppose $\bar{x} = 10$, $\sigma = 3.5$, and $n = 25$.

- (a) Conduct a hypothesis test to determine if μ is less than 12.
- (b) Conduct a hypothesis test to determine if μ is greater than 9.
- (c) Conduct a hypothesis test to determine if μ is different from 9.5.

Solution

(a) 1) $H_0 : \mu \geq 12$ $H_a : \mu < 12$

2) $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10 - 12}{3.5/\sqrt{25}} = -2.86$

3) $z_\alpha = z_{0.05} = -1.645$.

- 4) Since the test statistic (-2.86) lies in the rejection region (the region of the number line from $-\infty$ to -1.645), we reject H_0 .

- 5) We have enough evidence to reject the hypothesis that the population mean is greater than or equal to 12.

(b) 1) $H_0 : \mu \leq 9 \quad H_a : \mu > 9$

2) $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10 - 9}{3.5/\sqrt{25}} = 1.43$

3) $z_\alpha = z_{0.05} = 1.645$

- 4) Since the test statistic (1.43) does not lie in the rejection region (the region from 1.645 to ∞), we do not reject H_0 .

- 5) We do not have enough evidence to reject the null hypothesis that the population mean is less than or equal to 9.

(c) 1) $H_0 : \mu = 9.5 \quad H_a : \mu \neq 9.5$

2) $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10 - 9.5}{3.5/\sqrt{25}} = 0.71$

3) Since we have a two-sided alternative hypothesis, $z_{\alpha/2} = z_{0.025} = \pm 1.96$

- 4) Since the test statistic (0.71) does not lie in the rejection region (from $-\infty$ to -1.96, and from 1.96 to ∞), we do not reject H_0 .

- 5) We do not have enough evidence to reject the null hypothesis that the population mean is equal to 9.5. ■

Common question: How do I choose which method to use?

Answer: In this course, if you are asked to use a specific method, then please use it. If you are not asked to use a specific method, you are free to choose your favourite.

Common question: Suppose I reject H_0 using the p-value method. Would I also have rejected H_0 if I used the critical value method?

Answer: Yes, both methods are equivalent. We cover both because each method is used in industry and academia.

Common question: Should I always assume $\alpha = 0.05$?

Answer: Yes, unless you're told otherwise in the question. The value $\alpha = 0.05$ is the default value used in industry and academia.

2.8 More examples

In this section, we will go through three more examples. Now that you're familiar with the equations involved in this type of hypothesis testing, we will consider some more practical business scenarios to illustrate why hypothesis testing is used in a business setting. In these examples, I will illustrate *exactly* how much work needs to be shown (for an assignment question, for example), and not go into any extra detail.

Example 5: CookieCorp is a company that sells boxes of cookies. The quality assurance team at CookieCorp wants to ensure that the amount of sugar per cookie is higher than their main competitor, to appeal to their target demographic. Their competitor reports 8.2 g of sugar per cookie according to the label on the box. Of course, this is not the true amount of sugar per cookie: this is the population mean amount of sugar per cookie, and the company has a responsibility to keep the actual amount of sugar per cookie as close as possible to that value. The quality assurance team at CookieCorp obtains a sample of 100 of their cookies and finds an average sugar content of 8.21 g. They know the population standard deviation is 0.2 g. Conduct an appropriate hypothesis test using the p-value method.

Solution:

1. $H_0 : \mu \leq 8.2 \text{ g}; \quad H_a : \mu > 8.2 \text{ g}$
2. $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{8.21 - 8.2}{0.2/\sqrt{100}} = 0.5$
3. $P(Z > 0.5) = 1 - P(Z \leq 0.5) = 1 - 0.6915 = 0.3085$
4. Since the p-value (0.3085) is greater than the significance level ($\alpha = 0.05$), we do not reject the null hypothesis.
5. The population mean per-cookie sugar content of CookieCorp's cookies is likely not greater than that of their competitor. ■

Example 6: A company that creates point-of-sale (POS) systems is interested in designing a new system that will reduce the mean speed it takes for a transaction to complete. The old system completes transactions in an average of 2.5 seconds. The quality assurance team obtains a sample of 5 transactions using the new system (yes, this is a rather small number. We will discuss how to choose a sample size later in the course). The data are given below. Assume $\sigma = 0.25$ seconds. Conduct an appropriate hypothesis test.

Transaction #	1	2	3	4	5
Time (seconds)	2.04	2.12	3.12	3.1	2.5

Solution:

- $$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2.04 + 2.12 + 3.12 + 3.1 + 2.5}{5} = 2.576$$
1. $H_0 : \mu \geq 2.5 \text{ s}; \quad H_a : \mu < 2.5 \text{ s}$
 2. $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.576 - 2.5}{0.25/\sqrt{5}} \approx 0.68$
 3. $P(Z < 0.68) = 0.7517$

4. Since the p-value (0.7517) is greater than the significance level ($\alpha = 0.05$), we do not reject the null hypothesis.
5. It is not likely that the new POS reduces the population mean speed of transaction. ■

Example 7: A video game company is interested in testing whether there is a difference in total game play time when optional quests are added. Without the optional quests, the population mean time to complete the game is 35 hours. They recruit 25 game testers and find that their average time to complete the game with the optional quests included is 37 hours. They know the population standard deviation is 1.5 hours. Conduct an appropriate hypothesis test using the critical value method. Use a significance level of 0.01.

Solution:

1. $H_0 : \mu = 35\text{hr}; \quad H_a : \mu \neq 35\text{hr}$
2. $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{37 - 35}{1.5/\sqrt{25}} \approx 6.67$
3. $z_{\alpha/2} = z_{0.01/2} = z_{0.005} = \pm 2.575$
4. Since the test statistic (6.67) lies in the rejection region (the region on a number line from $-\infty$ to -2.575 and from 2.575 to ∞), we reject the null hypothesis.
5. It is likely that adding the optional quests increases the population mean time to complete the game. ■

2.8.1 Quick discussion of the examples

In examples 5 and 6, since the significance level was not given, we assumed it was 0.05. In example 6, we were not given the value of the sample mean, but we know how to calculate it, so we do (since we need it for the test statistic).

In example 5, the significance level was $\alpha = 0.01$, so since we had a two-tailed alternative hypothesis, we had to go to Table A4 and look up $0.01/2 = 0.005$ in the middle of the table. Since that value does not exist, but is bounded by 0.0049 (corresponding to a z-score of -2.58) and 0.0051 (corresponding to a z-score of -2.57), we average the z-scores and the critical value is -2.575. The rejection region is therefore the region on a number line from $-\infty$ to -2.575, and from 2.575 to ∞ . An illustration of this is provided in Figure 2.9.

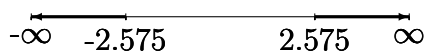


Figure 2.9: The critical value, ± 2.575 , provides the boundary for the rejection region (bold, from $-\infty$ to -2.575 and from 2.575 to ∞). If the test statistic lies in the rejection region (that is, if the test statistic is greater than 2.575 or less than -2.575), reject H_0 .

Notice in every example the units were stated in the null and alternative hypotheses. This is always necessary if units are given in the question.

2.9 Chapter summary

- The process of statistical inference refers to using information from a sample to say something about unknown population parameter(s).
- A hypothesis test is a type of statistical inference procedure.
- The null and alternative hypotheses summarize what you are trying to test.
 - The alternative hypothesis H_a contains the research question.
 - The null hypothesis H_0 is the opposite of the alternative hypothesis.
- There are five steps to conducting a hypothesis test (so far - in the next chapter, we'll introduce more):
 1. State the null and alternative hypotheses.
 2. Calculate the test statistic. The formula in this chapter was:
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$
 3. Find the p-value or the critical value.
 4. State whether you reject the null hypothesis or not (if the p-value is less than α , reject H_0 . If the test statistic lies further to the appropriate tail than the critical value, reject H_0).
 5. Conclude in the context of the original research question.
- The p-value is the probability that we see a test statistic as extreme, or more extreme, than the one we see, given H_0 is true.
- The critical value is the z-score decides the boundary of the rejection region. If the test statistic lies in the rejection region, reject H_0 .