Topic 8. The Logic and Components of Hypothesis Testing

Cheng Peng

Contents

1 Introduction

We have introduced methods for constructing confidence intervals to estimate population means and proportions under various assumptions. Now, we turn to a new form of statistical inference that allows us to evaluate claims about a population mean or proportion—hypothesis testing.

- In statistics, a hypothesis is a statement or claim about a characteristic of a population. A hypothesis test is a procedure that uses sample data to assess the validity of such a claim regarding a population parameter.

- The fundamental idea is to let data speak truthfully.

- **Rare Event Rule for Inferential Statistics**

  If, under a given assumption (such as the null hypothesis being true), the probability of an observed event occurring is exceptionally small, we conclude that the assumption (i.e., the null hypothesis) is likely incorrect.

Example 1. Suppose someone claims that the average height of students at WCU is at least 80 inches. To assess this claim, we take a random sample of 100 students, representing the WCU student population. The sample mean is found to be 68 inches, with a standard deviation of 15 inches.

2 Steps for Hypothesis Testing: Some Analogies

To better understand the logic and steps for conducting a statistical hypothesis and the types of inevitable errors in the decision process, we use the two analogies that are familiar to us.

Analogies of Statistical Hypothesis Testing		
Medical Diagnostics	**Test of Statistical Significance**	**Court Trial**
The patient experienced the symptoms of some disease and believe he is sick.	Someone claimed about the population parameter.	The prosecutor claimed the defendant committed a crime.
Every incoming patient is assumed to have a disease until proved disease-free	Null Hypothesis: claims the attainable values of the parameter.	Every defendant is assumed to be innocent until proved guilty
Order labs and gathering health info	data (information) collection	Investigation and evidence gathering
summarize clinical evidence	Information Aggregation: test Statistics	Summary of evidence exhibitions
Clinical standards	Statistical Decision Rules	Jury's instruction: cross-examination
Diagnostic decision	Statistical Decision on the null hypothesis	Verdict
Claim a diseased patient to be disease-free	**Type I Error**	Convict an innocent defendant
Claim a disease-free patient to be diseased	**Type II Error**	Acquit a criminal

No matter what decision will be made, there are two possible errors:

1. **Type I Error**: the patient has a disease, but the doctor claims NO disease.

2. **Type II Error**: the patient has NO disease, but the doctor claims a disease.

Next is an introduction video by Prof. Fulkerson

Watch this video first before read the subsequent definitions.

3 Components of Testing Hypothesis

The approach to formulating the process of hypothesis testing in this course is divided into three blocks: a practical claim (also called a practical hypothesis), steps for statistical hypothesis testing, and a conclusion addressing the claim.

The following chart depicts the process for testing a hypothesis.

The claim (hypothesis) about the population (Step 1)
(or practical hypothesis such as a statement about population mean, proportion, variance, etc.)

Statistical Hypothesis Procedure (Steps 2-5)
- Setting-up NULL and alternative hypotheses
- Gathering sample information and select appropriate statistical tables
- Aggregating information from the same and the null hypothesis to define a test statistic.
- Decisioning on the null hypothesis.

Address the question in the original claim (Step 6)
Use the relationship between the statistical alternative hypothesis and the original claim to answer the practical question.

Next, we introduce the components of hypothesis testing based on the above chart.

3.1 The claim (or practical hypothesis)

A claim (or a practical hypothesis) is a statement about the population parameter of interest such as population mean (µ) and proportion (p). It has six possible distinct forms. Using population mean (μ) as an example, we have the following six possible claims.

$$\mu \begin{matrix} > \\ \geq \\ \neq \\ = \\ < \\ \leq \end{matrix} \mu_0$$

where μ_0 is the claimed value. Three of them involve an equal sign ($=, \geq, \leq$) and the other three do not have an equal sign ($\neq, >, <$). For a given problem, we need to identify the claim and write it explicitly.

Some of the keywords are helpful for identifying the claim:

at most is equivalent to **less than or equal to**, i.e., $\leq$;

at least is equivalent to **greater than or equal to**, i.e., $\geq$;

differs implies $\neq$;

majority implies that proportion is greater than 0.5, i.e., $p > 0.5$.

Example 1. A water faucet manufacturer announces that the mean flow rate of a certain type of faucet is at most 2.5 gallons per minute.

What is the claim about the population mean flow rate of the type of faucet?

Solution: we see the keyword **at most** in the claim. Therefore, the claim is $\mu \leq 2.5$.

Example 2. A cereal company advertises that the mean weight of the contents of its 20-ounce size cereal boxes is more than 20 ounces.

What is the claim about the population mean?

Solution: We see the statement has keyword **more than**, the claim is $\mu > 20$.

Example 3. A university publicizes that the proportion of its students who graduate in 4 years is 82%.

What is the claim about the population proportion?

Solution: Key word **is** implies that the claim is: $p = 0.82$. Caution: The proportion MUST be written as a decision in all formulas.

A Cautionary Note: If the claim in the problem is vague, we have to choose an explicit claim in order to set up the null and alternative hypotheses.

Example 4. A bakery machine fills boxes with crackers, averaging 454 grams (roughly one pound) of crackers per box. The quality control manager wants to know whether the production process still maintains the quality standard. What is the claim about the mean weight of crackers per box?

Solution: The **claim** is vague. The manager only wants to know whether the production line still maintains the same quality over time. Therefore, the claim could be either $\mu = 454$ or $\mu \neq 454$. No matter what claim is used, we eventually draw a conclusion to justify it.

3.2 Null and Alternative Hypotheses

Definition 1: *Null Hypothesis* - The null hypothesis (H0) is a statement about a single value of a population parameter (such as the mean). It MUST take one of the following forms:

$$H_0 : \mu = \mu_0, \qquad H_0 : \mu \geq \mu_0, \qquad H_0 : \mu \leq \mu_0$$

where μ_0 is the value in the claim to be justified.

Important Observation: "=" must be included in H_0.

Definition 2: *Alternative Hypothesis*: The alternative hypothesis (H_a or H_1) is the statement that must be true if the null hypothesis is false. That is, the corresponding alternative hypothesis of each of the above three null hypotheses must be:

$$H_a : \mu \neq \mu_0, \qquad H_a : \mu < \mu_0, \qquad H_a : \mu > \mu_0$$

Important Observation: H_a must NOT have "="!

Relationship between H_0 and H_a

The null and alternative hypotheses are mutually complementary. This means that only one of the hypotheses is correct. For example, if we reject H_0, then we must accept H_a, and vice versa.

Because of the relationship between H_0 and H_a, there are ONLY three possible pairs of NULL and alternative hypotheses. Next, we use population mean as an example to explicitly write these three possible pairs of H_0 and H_a.
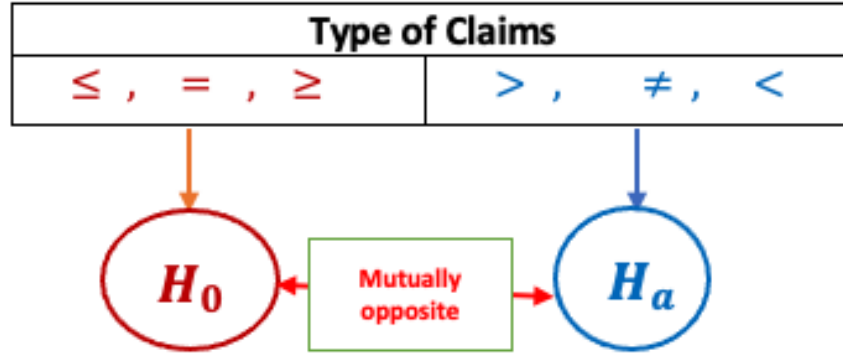
$$H_0 : \mu = \mu_0 \quad v.s. \quad H_a : \mu \neq \mu_0 \ ,$$
$$H_0 : \mu \geq \mu_0 \quad v.s. \quad H_a : \mu < \mu_0,$$

and

$$H_0 : \mu \leq \mu_0 \quad v.s. \quad H_a : \mu > \mu_0.$$

The above relationship is depicted in the following.

Relationship between the claim and H_a.

The following table illustrates the relationship.

The guideline for setting up Ha is given below:

a). If **the original claim** contains an "=" (i.e., $\leq, =, \geq$), we choose the opposite of the original claim (i.e., $>$, $\neq$, $<$) as **the alternative hypothesis** Ha.

b). If **the original claim** doesn't contain an "="(i.e., $>$, $\neq$, $<$), we simply choose the original claim (i.e., $>$, $\neq$, $<$) as **the alternative hypothesis** Ha.

Example 5. A bakery machine fills boxes with crackers, averaging 454 grams (roughly one pound) of crackers per box

(a). If the management of the bakery is concerned about the possibility that the actual average is different from 454 grams, what null hypothesis and what alternative hypothesis should it use to put this to test?

(b). If the management of the bakery is concerned about the possibility that the actual average is at most 454 grams, what null hypothesis and what alternative hypothesis should it use to put this to test?

Solution: We use the relationship between claim, null and alternative hypotheses.

(a). The original claim: "the actual average is different from 454 grams", i.e., $\mu \neq 454$. Since claim $\mu \neq 454$ doesn't have an "=" sign, it is used as H_a. Therefore, the null and alternative hypotheses are specified in the following.

$$H_0 : \mu = 454 \;\; v.s. \;\; H_a : \mu \neq 454.$$

(b). The original claim: " the actual average is at most 454 grams", i.e., $\mu \leq 454$. Since the "=" sign is contained in the claim, the claim $\mu \leq 454$ is used as H_0. Therefore, the null and alternative hypotheses are specified in the following.

$$H_0 : \mu \le 454 \quad v.s \quad H_a : \mu > 454$$

Exercises Read examples 1 - 4 and set up the null and alternative hypotheses for these examples.

Watch the following short video from `tutors.com` to get more examples.

3.3 Test Statistic

Some characteristics of test statistics.

- The **test statistic** is a value computed from the sample data that is used to decide whether **null hypothesis** is rejected.

- The **test statistic** converts the sample statistic (such as sample mean) to a score (such as z-score, t-score, etc.) with the assumption that the null hypothesis is true.

- When we test the claims about a population mean (μ) in a large sample case, we can use the central limit theorem (CLT) to derive the distribution of the **test statistic**.

In other words, a test statistic is an **information aggregator** that consolidates information from multiple sources. For example, the following **test statistic** is used to test a population mean μ.

$$TS = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

The above **test statistic** contains information from the sample ($\bar{x}$, s, and sample size n) and the hypotheses (the claimed population mean μ_0).

Recall that the above **test statistic** is an approximate standard normal distribution if the sample size is large (n > 30 in our course).

An Intuitive Interpretation of TS: TS is a standardized "distance" that measures the discrepancy between the observed value ($\bar{x}$) and the claimed value (μ_0).

For example, if the above test statistic is used to test

$$H_0 : \mu = \mu_0 \quad v.s. \quad H_a : \mu \ne \mu_0.$$

If TS is close to 0, the observed value and the claimed value are close to each other implying that the sample evidence supports the null hypothesis (H_0) and rejects the alternative hypothesis (H_a). Otherwise, if TS is NOT close to 0, there is a difference between the observed value and the claimed value implying that the sample evidence **does not** support the null hypothesis (H_0) but supports the alternative hypothesis (H_a).

A natural question: *how close* is called close?

Example 6. We use the body temperature data collected by the researchers at the University of Maryland with characteristics: $n = 36$, $\bar{x} = 98.20^0 F$, $s = 0.62^0 F$. Find the value of the test statistic for the claim that the population mean is $\mu = 98.6^0 F$.

Solution: Based on the given information and the formula, we calculate the test statistic in the following

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{98.2 - 98.6}{0.62/\sqrt{36}} = -3.87.$$

The "closeness" question: Is $TS = -3.87$ close to 0?

Standard Normal Distribution N(0,1)
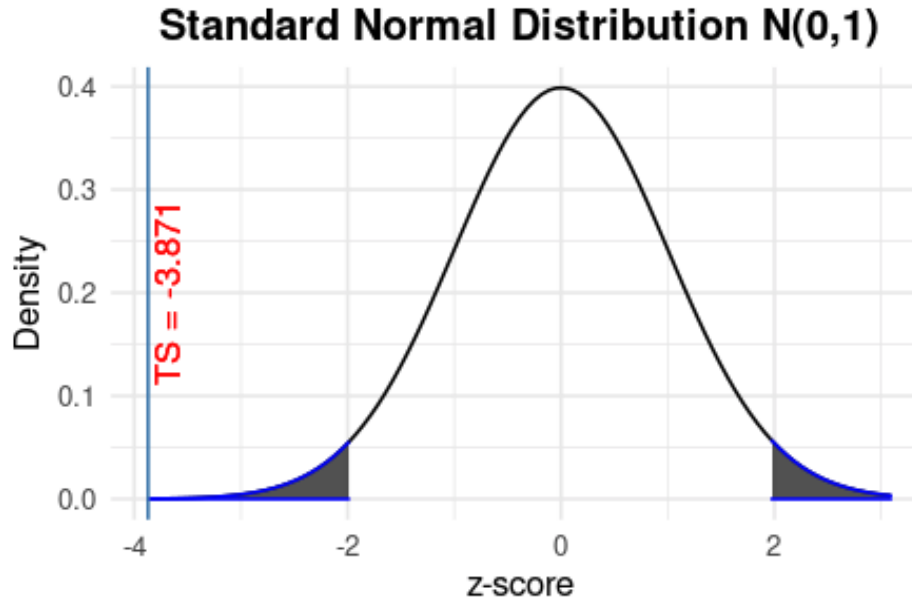
We can see from the above figure that TS is not close to 0 since it is located far away from the center. The next question is how to find a threshold that can tell **closeness**. The threshold is called **critical value**.

Watch the following short video from `tutors.com` to see another example for both small and large samples. Please pay attention to the given conditions

3.4   Significance Level and Critical Value(s)

- The **Rejection Region (RR)** is the set of all values of the score converted from the statistic that causes us to reject the **null hypothesis**. The cut-off value(s) is(are) called Critical Value(s).

- The **significance level** (denoted by α) is the probability that the test statistic will fall into the critical region while the null hypothesis is true. This α is the same α in confidence level $1 - \alpha$ used in constructing confidence intervals.

**Example 7. (continuation of Example 6)** Assume we null and alternative hypotheses are

$$H_0 : \mu = 98.6 \ \ v.s. \ \ H_a : \mu \neq 98.6.$$

we have found that $TS = -3.87$. Under Ho, we have the following density curve
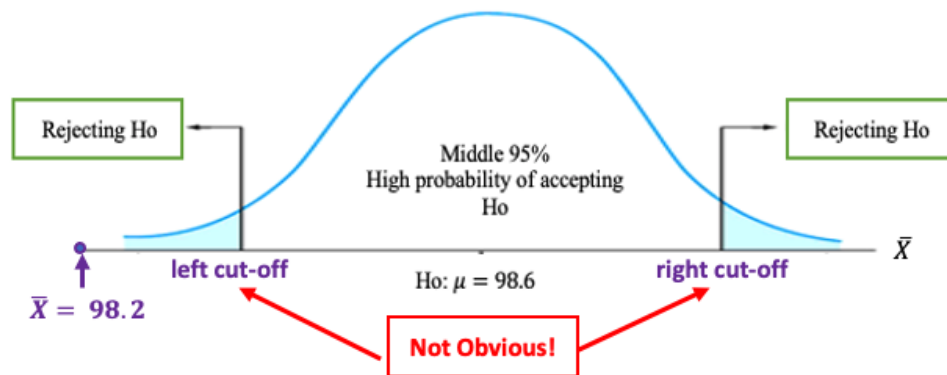
Since left and right cut-offs are unknown, we still cannot tell whether $\bar{X} = 98.2^0 F$ is close to $H_0 : \mu_0 = 98.6^0 F$. However, we can easily see from the standardized test statistic on the density curve of the standard normal curve is NOT close to 0 (see the following figure).
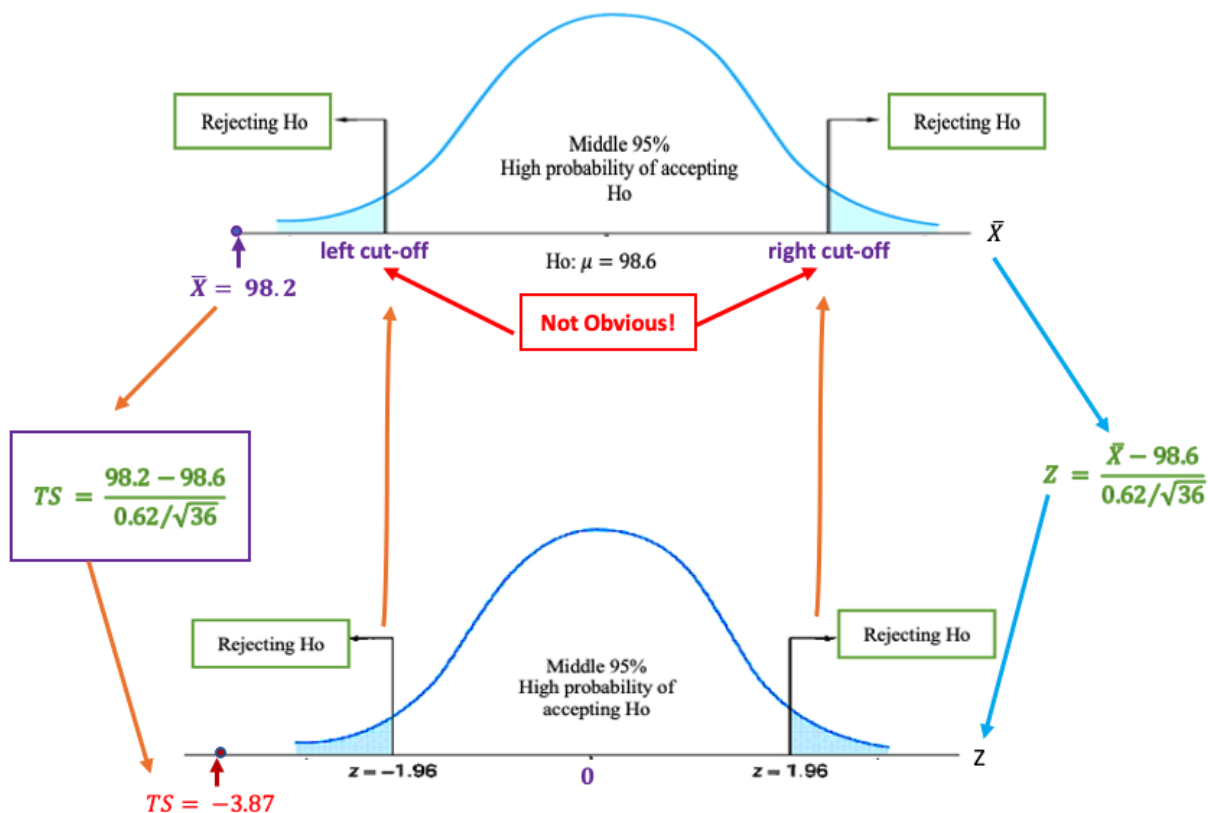
As we have discussed in the confidence interval, all numbers in interval $(-1.96, 1.96)$ are considered to be close to 0. Therefore, $TS = -3.87$ is NOT close to 0. This implies that $\bar{x} = 98.2^0 F$ is NOT close to $H_0 : \mu_0 = 98.6^0 F$. Since the bottom panel of the above figure is independent of the specific normal distribution, we use the distribution of TS to define rejection region(s) and critical value(s) for a given significance level. In other words, for testing

$$H_0 : \mu = 98.6 \ \ v.s. \ \ H_a : \mu \neq 98.6.$$

at **significance level** $\alpha = 0.05$,

Rejection Region (RR): consists of two tail regions.

9

Critical Values (CV): are the values that are defined in such a way that both tail areas are equal to $\alpha/2 = 0.05/2 = 0.025$. That is, the critical value for this testing hypothesis is $CV = \pm 1.96$ which can be found in the normal table.

3.5 Statistical Decision on Ho

Once we found the critical value and the value of the test statistic, we can make a statistical decision using the following rule:

- If TS falls into the rejection region (RR), we **reject** null hypothesis **Ho**;

- If TS does NOT fall into the rejection region (RR), we **fail to reject** null hypothesis **Ho**.

3.6 Conclusion about the Claim

We use the relationship between **the claim**, **null hypothesis**, and **alternative hypothesis** to conclude the claim (supporting or rejecting the claim.)
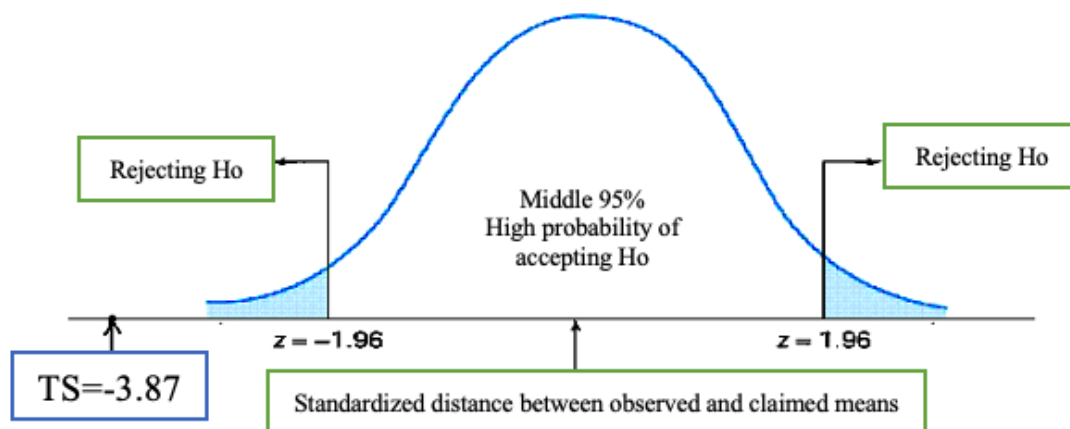
Example 8. (continuation of Example 6) The claim is *the population mean is* $\mu = 98.6^0 F$. Recall that null and alternative hypotheses are

$$H_0 : \mu = 98.6 \ \ v.s. \ \ H_a : \mu \neq 98.6.$$

and

$$TS = -3.87.$$

For this specific hypothesis test. The rejection region (RR) has two parts located on each tail (see the figure below).

Apparently, the test statistic $TS = -3.87$ falls into the rejection. Therefore, we reject the null hypothesis $H_0 : \mu = 98.6$. Since the null hypothesis and the claim are identical, the sample does not have evidence to support the claim.

4   Formal Steps for Hypothesis Testing

Based on the components discussed in the previous section, we formulate the following **six-step procedure** for testing the hypothesis about the mean or proportion of a population.

Step 1: Identify the claim about a population parameter such as the population mean or proportion.

Step 2: Set up null and alternative hypotheses.

Step 3: Evaluate the test statistic.

Step 4: Find the critical value(s) based on the given significance level (α) and specify the rejection region (RR).

Step 5: Make a statistical decision: reject or fail to reject H_0.

Step 6: Draw a conclusion about the population parameter in the claim.

These steps will be used in all examples of hypothesis testing problems throughout the course.

5   Types of Hypothesis Tests

In examples 6 - 8, we test the null hypothesis $H_0 : \mu_0 = 98.6$. This means that the null hypothesis should be rejected if the test statistic TS is far from zero (in either a positive or negative direction), the rejection region has two parts: one on the left tail and one on the right tail. This type of test is called the **two-tailed** test.
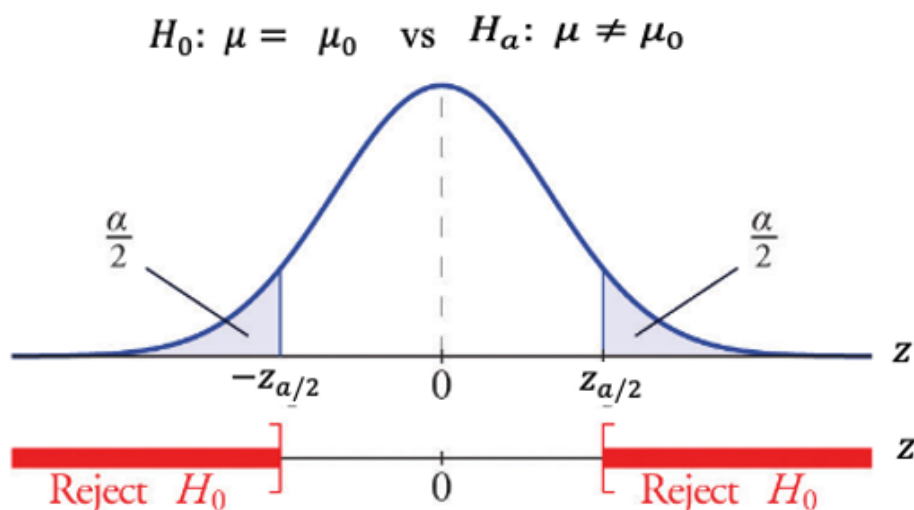
The form of the alternative hypothesis tells the type of test. For example, for a two-tailed test, the alternative hypothesis test takes the form of $H_0 : mu \neq \mu_0$. Note that $\neq$ means either $>$ or $<$. That is, **the direction of the inequality sign points to the location of the rejection region (RR)**.

The following video explains the three types of hypothesis tests,

Next we will give more detail graphical explanations of the three types of tests.

5.1   Two-tailed Test

The following figure shows the locations of the rejection regions of a two-tailed test. The form of the alternative hypothesis $H_a : \mu \neq \mu_0$ implies that TS is not close to 0. Therefore, the rejection region of H_0 is **evenly split into right and left sub-rejection-regions.**

both tail areas in the density curve are equal to half of the significance level.

5.2   Right-tailed Test

When the alternative hypothesis takes the form of $H_a : \mu > \mu_0$, the corresponding test is called a **right-tailed test** with the inequality sign in H_a pointing to the right. Consequently, there is only **one** critical value on the right tail of the density curve.
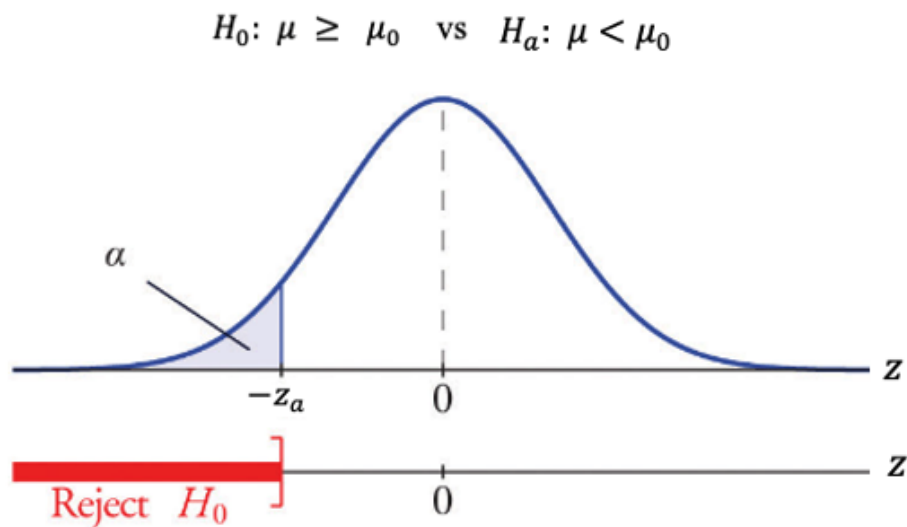
$$H_0: \mu \leq \mu_0 \quad \text{vs} \quad H_a: \mu > \mu_0$$
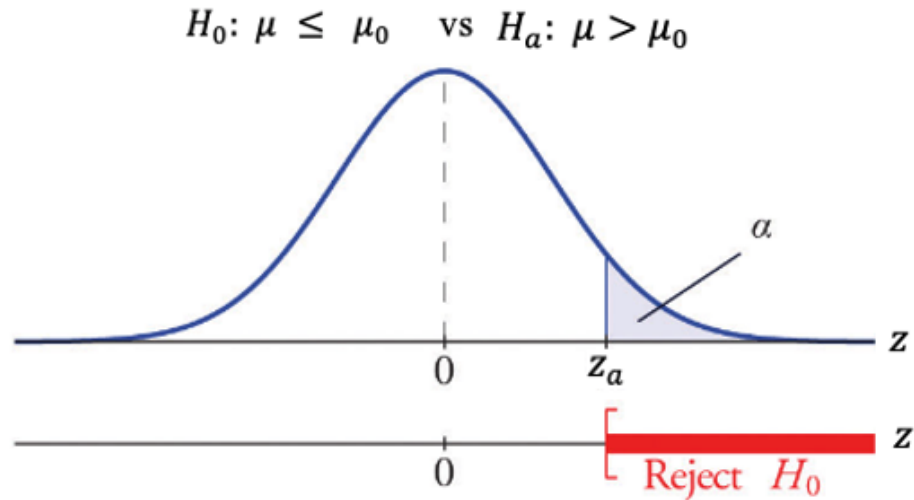
This is intuitive since the null hypothesis of a right-tailed test takes the form of $H_0 : \mu \leq \mu_0$. This implies that H_0 is supported if the observed **test statistic** is far left. Otherwise, we reject the null hypothesis H_0. Consequently, there are **two** critical values: one on the right tail and one on the left tail of the density curve.

The right tail area of the density curve is equal to the significance level.

5.3   Left-tailed Test

When the alternative hypothesis takes the form of $H_a : \mu < \mu_0$, the corresponding test is called a **left-tailed test** with the inequality sign in H_a pointing to the left. Consequently, there is only **one** critical value on the left tail of the density curve.

$$H_0: \mu \geq \mu_0 \quad \text{vs} \quad H_a: \mu < \mu_0$$

Similarly, the null hypothesis of a right-tailed test takes the form of $H_0 : \mu \geq \mu_0$. This implies that H_0 is supported if the observed **test statistic** is far right. Otherwise, we reject the null hypothesis H_0.

The left tail area of the density curve is equal to the significance level.

Example 9. A professor wants to investigate the effectiveness of a new pedagogical method. Assume the average score for the population is 80 before the experiment. With a new training method, the professor believes that the score might change. Professor tested randomly 36 students' scores. The average score of the sample is 88 and the standard deviation is 10. With a 5% significance level, is there enough evidence to suggest the average score changed?

Solution: We follow the 6-step procedure to perform the hypothesis to answer the above question.

Step 1: Since the statement **'professor believes the mean score *might change*'** does not specifically mention the direction of the change, we can take either one of the two forms: $\mu = 80$ or $\mu \neq 80$.

In this solution, we take the form: $\mu \neq 80$.

Step 2: Since the claim $\mu \neq 80$ has no '=' in it, it will be the alternative hypothesis and the opposite of it will be the null hypothesis. More specifically,

$$H_0 : \ \mu = 80 \ \ v.s. \ \ H_a : \mu \neq 80$$

This is a two-tailed test. Therefore, there will be two rejection regions.

Step 3: The test statistic is defined by

$$TS = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{88 - 80}{10/\sqrt{36}} = 4.8.$$

Step 4: Since $n = 36 > 30$, by the central limit theorem (CLT), the test statistic is approximately normally distributed. We use the given confidence level $\alpha = 0.05$ and the normal table to find the two critical values: $\pm z_{0.05/2} = \pm z_{0.025} \pm 1.96$. All above information is labeled in the following figure.
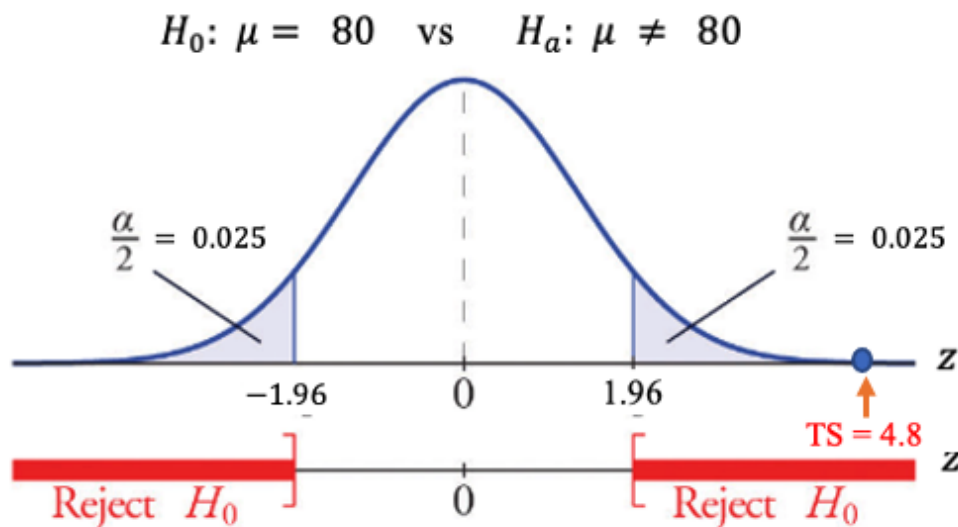
Step 5: Since $TS = 4.8$ is in the rejection region (right-hand side), the null hypothesis is rejected.

Step 6: We conclude that the population score has changed after the training program.

Critical values of Z-tests with commonly used significance level: Three commonly used significance levels are $\alpha = 0.10, 0.05$ and 0.01. Their corresponding critical values associated with different types of tests are summarized in the following table.

$$H_0: \mu = 80 \quad \text{vs} \quad H_a: \mu \neq 80$$
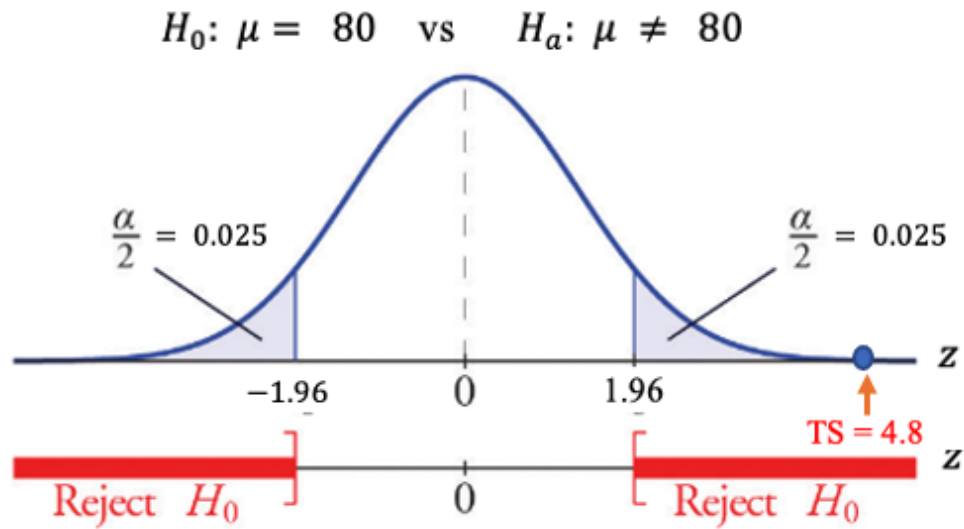
The following screenshot shows the solution generated by ISLA

ISLA: ONE SAMPLE NORMAL TEST FOR POPULATION MEAN

Data Source
- summarized statistics
- raw data

sample mean ($\bar{x}$)
88

sample standard deviation (s)
10

sample size (n)
36

Claimed Value (μ_0)
80

Claim Type
not equal to

Significance level α
0.01 0.05   0.2

0.01 0.03 0.05 0.07 0.09 0.11 0.13 0.15 0.17 0.190.2

I.S.L.A
Interactive Stats Learning Apps

Report bugs to C. Peng

Solution: This normal test is based on the Central Limit Theorem CLT. Since the sample size is larger than 30, the test result is reliable.

Given sample information: $n = 36$, $\bar{x} = 88$, $s = 10$.

Step 1: Identify the claim of the population mean (μ_0).
The given information indicates that the claim is: μ_0 is not equal to 80.

Step 2: Set up the null and alternative hypotheses.
Based on the claim, the null and alternative hypotheses are given by $H_0 : \mu = 80$ and $H_1 : \mu \neq 80$.

Step 3: Evaluate the test statistic.
The test statistic is defined to be: $TS = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{88 - 80}{10/\sqrt{36}} = 4.8$

Step 4: Find the critical value and calculate the p-value.
Based on the significance level, we found the critical values to be : $\pm z_{\alpha/2} = \pm z_{0.025} = \pm 1.96$
The p-value is can be found as p-value ≈ 0.

Step 5: Make a statistical decision on H_0.
At the 5% significance level, we reject the null hypothesis. (p-value< 0.001).

Step 6: Draw conclusion [justify the claim in step 1].
At the 5% significance level, we conclude the alternative hypothesis.

Standard Normal Distribution N(0,1)
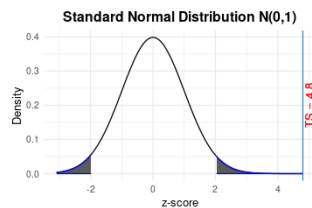
Watch the following short video from `tutors.com` gives additional example in hypothesis.

You are encouraged to use the ISLA app to reproduce the results given in the video.

6   Use of Technology

The ISLA App for normal tests is at https://wcu-peng.shinyapps.io/oneMean-z-Test/. The following example shows how to use the app to check your manual work.

The above discussions are based on the case of large sample size. However, if population is normal and population variance (or standard deviation) is known, the Z-test is used for testing population mean. Next, we use an ISLA app in the following example.

Example 10: Portia bought five of the same racket at an online auction site for the following prices

    155, 179, 175, 175, 161

Assuming that the auction prices of rackets are normally distributed with known standard deviation 10, determine whether there is sufficient evidence in the sample, at the 5% level of significance, to conclude that the **average price** of the racket is **less than** \$179 if purchased at an online auction.

Solution: we are given that $\sigma = 10$. The claim $\mu < 179$. Note that the sample mean is $\bar{x} = 169$, sample size $n = 5$, and significance level $\alpha = 0.05$. Since the population is normal with known standard deviation. A normal test is applied to this example.

We input the related information to the app and obtain the generated the solution in the following screenshot.

ISLA: One Sample Normal Test for Population Mean

Data Source
- ○ summarized statistics
- ● raw data

sample mean $(\bar{x})$

169

sample standard deviation (s)

10

sample size (n)

5

Claimed Value (μ_0)

179

Claim Type

less than ▼

Significance level α
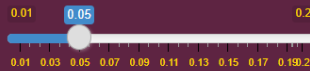
0.01   0.05   0.2

0.01 0.03 0.05 0.07 0.09 0.11 0.13 0.15 0.17 0.190.2

I.S.L.A
Interactive Stats Learning Apps

Report bugs to C. Peng

Solution: This normal test is based on the Central Limit Theorem CLT. Since the sample size is less than 30, we assume that the population is normal and the population variance is known.

Given sample information: $n = 5$, $\bar{x} = 169$, $s = 10$.

Step 1: Identify the claim of the population mean (μ_0).
The given information indicates that the claim is: μ_0 is less than 179.

Step 2: Set up the null and alternative hypotheses.
Based on the claim, the null and alternative hypotheses are given by $H_0 : \mu = 179$ and $H_1 : \mu < 179$.

Step 3: Evaluate the test statistic.
The test statistic is defined to be: $TS = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{169 - 179}{10/\sqrt{5}} = -2.236$

Step 4: Find the critical value and calculate the p-value.
Based on the significance level, we found the critical value$s$ to be : $-z_\alpha = -z_{0.05} = -1.645$
The p-value is can be found as p-value ≈ 0.013.

Step 5: Make a statistical decision on H_0.
At the 5% significance level, we reject the null hypothesis. $(p\text{-value} = 0.013)$.

Step 6: Draw conclusion [justify the claim in step 1].
At the 5% significance level, we conclude the alternative hypothesis.

Standard Normal Distribution N(0,1)