Seven

Hypothesis Tests

1 Hypothesis Tests

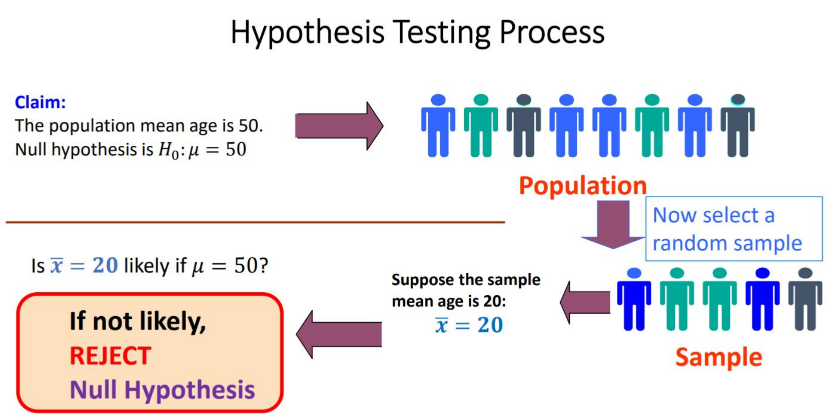
One of the most fundamental technique of statistical inference is the hypothesis test. There are many types of hypothesis tests but all follow the same logical structure, so we begin with hypothesis testing of a population mean.

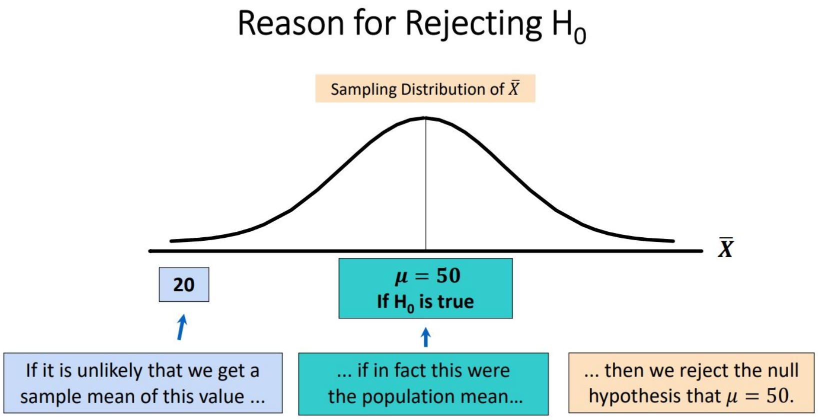
Hypothesis testing begins with a null hypothesis and an alternative hypothesis. Both the null and the alternative hypotheses are statements about a population. In this chapter, that statement will be a statement about the mean(s) of the population(s).

We will illustrate using an example.

Example 7.1 (Mean Age)
We are interested to check if the mean age of a population is $\mu = 50$.

Suppose we have no access to population data. So we take a sample from the population and obtained a sample mean age of $\bar{x} = 20$. Does this gives evidence for or against the hypothesis that $\mu = 50$?





EXAMPLE 7.2 (NUS STUDENTS' IQ)
Consider the statement

"NUS students have higher IQ than the general population (100)."

It is difficult/expensive to ask every NUS student to take an IQ test. So we take a sample.

Suppose the sample average is 110.

- Does that mean we're right?
- What if the sample average is 101? What about 100.1?
- Does the sample size matter?

HOW TO DO A HYPOTHESIS TEST:
There are five main steps to hypothesis testing.

Step 1: Set your competing hypotheses: null and alternative.

Step 2: Set the level of significance.

Step 3: Identify the test statistic, its distribution and the rejection criteria.

Step 4: Compute the observed test statistic value, based on your data.

Step 5: Conclusion. ■

Let us have a closer look at each step.

Step 1: Null Hypothesis vs Alternative Hypothesis

Our goal is to decide between two competing hypotheses.

NULL VS ALTERNATIVE:
In general, we adopt the position of the null hypothesis unless there is overwhelming evidence against it.

The null hypothesis is typically the default assumption, or the conventional wisdom about a population. Often it is exactly the thing that a researcher is trying to show is false.

We usually let the hypothesis that we want to prove be the alternative hypothesis. The alternative hypothesis states that the null hypothesis is false, often in a particular way. ■

The outcome of hypothesis testing is to either reject or fail to reject the null hypothesis.

A researcher would collect data relating to the population being studied and use a hypothesis test to determine whether the evidence against the null hypothesis (if any) is strong enough to reject the null hypothesis in favor of the alternative hypothesis.

We usually phrase the hypotheses in terms of population parameters.

EXAMPLE 7.3 (ONE-SIDED TEST)
Let μ be the average IQ of NUS students. Consider

$$H_0 : \mu = 100 \quad \text{vs} \quad H_1 : \mu > 100.$$

This is an example of a one-sided hypothesis test.

For this alternative hypothesis, we do not care if $\mu < 100$: the goal here is just to show NUS students have IQ higher than 100.

EXAMPLE 7.4 (TWO-SIDED TEST)
Sometimes it is more natural to do a two-sided hypothesis test.

For example, let p be the probability of heads for a particular coin. You want to test if the coin is fair (that is, $p = 0.5$), as it is equally problematic if p was larger or smaller.

Hence you set your hypotheses to be

$$H_0 : p = 0.5 \quad \text{vs} \quad H_1 : p \neq 0.5.$$

Step 2: Level of Significance

For any test of hypothesis, there are two possible conclusions:

- Reject H_0 and therefore conclude H_1;
- Do not reject H_0 and therefore conclude H_0.

Whatever decision is made, there is a possibility of making an error.

	Do not reject H_0	Reject H_0
H_0 is true	Correct Decision	**Type I error**
H_0 is false	**Type II error**	Correct Decision

DEFINITION 1 (TYPE I VS TYPE II ERROR)
The rejection of H_0 when H_0 is true is called a ***Type I error****.*

Not rejecting H_0 when H_0 is false is called a ***Type II error****.*

DEFINITION 2 (SIGNIFICANCE LEVEL VS POWER)
The probability of making a Type I error is called the ***level of significance****, denoted by α. That is,*

$$\alpha = P(\textit{Type I error}) = P(\textit{Reject } H_0 \mid H_0 \textit{ is true}).$$

Let

$$\beta = P(\textit{Type II error}) = P(\textit{Do not reject } H_0 \mid H_0 \textit{ is false}).$$

We define $1 - \beta = P(\textit{Reject } H_0 \mid H_0 \textit{ is false})$ to be the ***power of the test****.*

REMARK:
The Type I error is considered a serious error, so we want to control the probability of making such an error.

Thus prior to conducting a hypothesis test, we set the significance level α to be small, typically at $\alpha = 0.05$ or 0.01. ■

Step 3: Test Statistic, Distribution and Rejection Region

To test the hypothesis, we first select a suitable test statistic for the parameter under the hypothesis.

The test statistic serves to quantify just how unlikely it is to observe the sample, assuming the null hypothesis is true.

As the significance level α is given, a decision rule can be found such that it divides the set of all possible values of the test statistic into two regions, one being the rejection region (or critical region) and the other, the acceptance region.

Step 4 & 5: Calculation and Conclusion

Once a sample is taken, the value of the test statistic is obtained.

We check if it is within our rejection region.

- If it is, our sample was too improbable assuming H_0 is true, hence we reject H_0.
- If it is not, we did not accomplish anything. We failed to reject H_0 and hence fall back to our original assumption of H_0.

Note that in the latter case, we did not "prove" that H_0 is true. Hence, it is prudent to use the term "fail to reject H_0" instead of "accept H_0."

L-Example 7.1 (Possible Structures of H_0 and H_1)

For a valid set of hypotheses (i.e., H_0 versus H_1), they need to be disjoint. For example "$H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$" is one valid set; "$H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$" is another valid set.

By convention (and theoretically supported), we usually write the null hypothesis in the "equal" form, i.e., $H_0 : \theta = \theta_0$, with θ_0 being a given value. The hypotheses have three possible forms:

- ⋆ $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$; in this case $H_0 : \theta_0 = \theta_0$ in fact means $\theta \leq \theta_0$;
- ⋆ $H_0 : \theta = \theta_0$ versus $H_1 : \theta < \theta_0$; in this case $H_0 : \theta_0 = \theta_0$ in fact means $\theta \geq \theta_0$;
- ⋆ $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

So we need to check the form of H_1 to ensure the real meaning of $H_0 : \theta = \theta_0$.

L-Example 7.2

A certain type of cold vaccine is known to be only 25% effective after a period of 2 years.

In order to determine if a new and somewhat more expensive vaccine is superior in providing protection against the same virus for a longer period of time, 20 people are chosen at random and inoculated with the new vaccine.

If more than 8 of those receiving the new vaccine surpass the 2-year period without contracting the virus, the new vaccine will be considered superior to the one presently in use.

This is equivalent to testing the hypothesis that the binomial parameter for the probability of a success on a given trial is $p = \frac{1}{4}$ against the alternative that $p > \frac{1}{4}$.

In other words, we are testing

$$H_0 : p = \frac{1}{4} \quad \text{vs} \quad H_1 : p > \frac{1}{4}.$$

Let X be the number of individuals who remain free of the virus for at least 2 years.

For the conditions given, we think of the "acceptance region" and "rejection region" as

$$\overbrace{0, 1, 2, \ldots, 7, 8,}^{\text{acceptance region}} \overbrace{9, 10, 11, \ldots, 19, 20}^{\text{rejection region}}$$

The above decision rule has the level of significance given by

$$\begin{aligned}
\alpha &= P(\text{Type I error}) \\
&= P(\text{Reject } H_0 \mid H_0 \text{ is true}) \\
&= P(X > 8 \mid p = \frac{1}{4}) \\
&= \sum_{i=9}^{20} \binom{20}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{20-i} = 0.0409.
\end{aligned}$$

The probability of committing a Type II error is impossible to compute unless we have a specific alternative hypothesis. So let's consider

$$H_0 : p = \frac{1}{4} \quad \text{vs} \quad H_1 : p = \frac{1}{2}.$$

Note that $p = \frac{1}{2}$ satisfies $p > \frac{1}{4}$.

With this,

$$\begin{aligned}
\beta &= P(\text{Type II error}) \\
&= P(\text{Do not reject } H_0 \mid H_1 \text{ is true}) \\
&= P(X \leq 8 \mid p = \frac{1}{2}) \\
&= \sum_{i=0}^{8} \binom{20}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{20-i} = 0.2517.
\end{aligned}$$

2 Hypotheses Concerning the Mean

Let's apply our hypothesis steps to testing a population mean.

Case: Known variance

Let us consider the case where

- the population variance σ^2 is known; AND
- where
  - the underlying distribution is normal; OR
  - n is sufficiently large (say, $n \geq 30$).

Step 1: We set the null and alternatives hypotheses as

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

Note that in this case we are considering a two-sided alternative hypothesis.

Step 2: Set level of significance: α is typically set to be 0.05.

Step 3: Statistic & its distribution:
With σ^2 known and population normal (or $n \geq 30$),

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

When H_0 is true, $\mu = \mu_0$, the above becomes

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1),$$

and will serve as our test statistic.

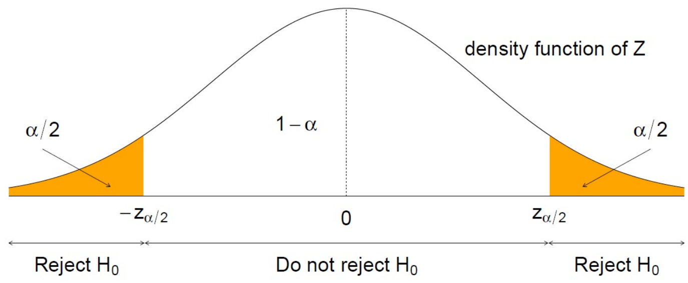
Rejection region:
Intuitively, we should reject H_0 when $\overline{X}$ is too large or too small compared with μ_0.

This is the same as when Z is too large or too small. In theory,

$$P(|Z| > z_{\alpha/2}) = \alpha.$$

Let the observed value of Z be z. Then the rejection region is defined by $|z| > z_{\alpha/2}$, which is

$$z < -z_{\alpha/2} \quad \text{or} \quad z > z_{\alpha/2}.$$



Step 4: Computations: z should be computed from the statistic above based upon the observed sample.

Step 5: Conclusion: check whether z is located within rejection region. If so, reject H_0, otherwise do not reject H_0.

WHERE DID THE VALUE 0.05 COME FROM?

In 1931, in a famous book called The Design of Experiments, *Sir Ronald Fisher discussed the amount of evidence needed to reject a null hypothesis.*

He said that it was situation dependent, but remarked, somewhat casually, that for many scientific applications, 1 out of 20 might be a reasonable value.

Since then, some people — indeed some entire disciplines — have treated the number 0.05 as sacrosanct.

Sir Ronald Fisher (1890 – 1962) was one of the founders of modern Statistics. For a biography of Fisher, browse to
http://www-history.mcs.st-andrews.ac.uk/Biographies/Fisher.html

EXAMPLE 7.5
The director of a factory wants to determine if a new machine A is producing cloths with a breaking strength of 35 kg with a standard deviation of 1.5 kg.

A random sample of 49 pieces of cloths is tested and found to have a mean breaking strength of 34.5 kg. Is there evidence that the machine is not meeting the specifications for mean breaking strength?
Use $\alpha = 0.05$.

Solution:
Note that $n > 30$ and $\sigma = 1.5$.

Let μ be the mean breaking strength of cloths manufactured by the new machine.

Step 1: We test

$$H_0 : \mu = 35 \quad \text{vs} \quad H_1 : \mu \neq 35.$$

Step 2: Set $\alpha = 0.05$.

Step 3: As σ^2 is known and $n \geq 30$,

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

will serve as our test statistic.

Since $z_{\alpha/2} = z_{0.025} = 1.96$, the critical/rejection region is

$$z < -1.96 \quad \text{or} \quad z > 1.96.$$

Step 4: z is computed to be

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{34.5 - 35}{1.5/\sqrt{49}} = -2.3333 < -1.96.$$

Step 5: The observed z value, $z = -2.3333$, falls inside the critical region. Hence the null hypothesis $H_0 : \mu = 35$ is rejected at the 5% level of significance.

One-sided alternatives

Now the above procedures are establish under

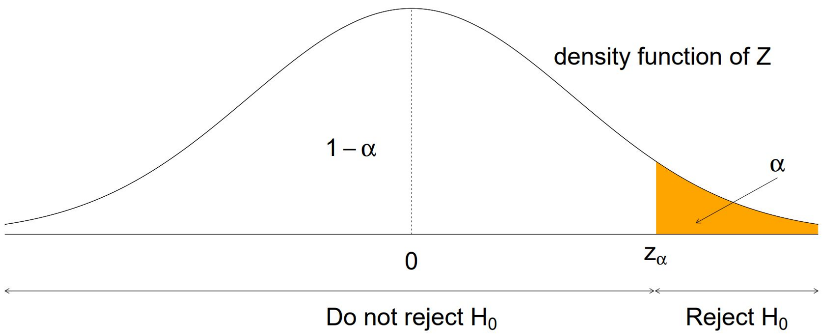
$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

Suppose instead we are considering

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0.$$

Similar steps can be used to address this problem, we only need to do the following changes:

- Step 1: H_1 is replaced with $H_1 : \mu > \mu_0$.
- Step 3: The test statistic and its distribution are kept the same. The rejection region should be replaced with $z > z_\alpha$, since now, we should reject only when $\bar{x}$ (and therefore z) is large.



The case for

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$$

should be self-evident.

HYPOTHESIS TEST FOR THE MEAN: KNOWN VARIANCE:
Consider the case where

- the population variance σ^2 is known; AND
- where
  - the underlying distribution is normal; OR
  - n is sufficiently large (say, $n \geq 30$).

For the null hypothesis $H_0 : \mu = \mu_0$, the test statistics is given by

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

Let z be the observed Z value. For the alternative hypothesis

- $H_1 : \mu \neq \mu_0$, the rejection region is

$$z < -z_{\alpha/2} \quad \text{or} \quad z > z_{\alpha/2}.$$

- $H_1 : \mu < \mu_0$, the rejection region is

$$z < -z_\alpha.$$

- $H_1 : \mu > \mu_0$, the rejection region is

$$z > z_\alpha.$$

■

p-value approach to testing

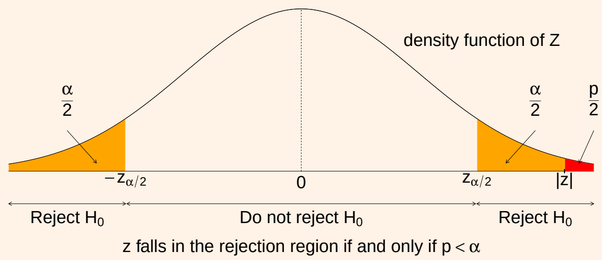
The above technique introduced by Fisher is based on a pre-declared significance level α.

Today, there is little reason to stick to the arbitrary 1% or 5% levels that Fisher suggested. We can instead use the idea of the p-value.

DEFINITION 3 (p-VALUE)
*The **p-value** is the probability of obtaining a test statistic at least as extreme ($\leq$ or $\geq$) than the observed sample value, given H_0 is true.*
It is also called the observed level of significance.

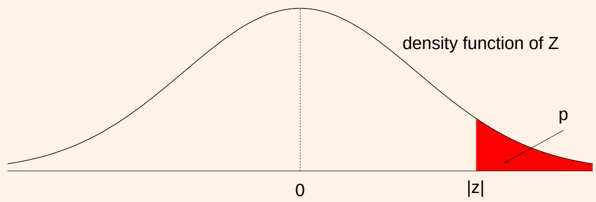
p-VALUE FOR HYPOTHESIS TESTS:
Suppose our computed test statistic was z. For a two sided test, a "worse" result would be if $Z > |z|$ or $Z < -|z|$, in other words, $|Z| > |z|$.
So the p-value is given by

$$p\text{-value} = P(|Z| > |z|) = 2P(Z > |z|) = 2P(Z < -|z|)$$



For the alternative hypothesis $H_1 : \mu < \mu_0$, the p-value is $P(Z < -|z|)$.
That is, only the area in the left tail is used.

For the alternative hypothesis $H_1 : \mu > \mu_0$, the p-value is $P(Z > |z|)$.
That is, only the area in the right tail is used.



One−sided test: p−value is represented by the shaded area

∎

REJECTION CRITERIA USING p-VALUE:
We see that the p-value is smaller than the significance level *if and only if* our test statistic is in the rejection region.

Thus our rejection criteria would be

- If p-value $< \alpha$, reject H_0; else
- If p-value $\geq \alpha$, do not reject H_0. ∎

REMARK:
In practice, it is better to report the p-value than to indicate whether H_0 is rejected.

- The p-values of 0.049 and 0.001 both result in rejecting H_0 when $\alpha = 0.05$, but the second case provides much stronger evidence.
- p-values of 0.049 and 0.051 provide, in practical terms, the same amount of evidence about H_0.

Most research articles report the p-value rather than a decision about H_0. From the p-value, readers can view the strength of evidence against H_0 and make their own decision, if they want to. ∎

EXAMPLE 7.6 (MIDTERM EXAM SCORE)
Recall the midterm exam scores example in an earlier chapter. The data obtained are

$$20, 19, 24, 22, 25.$$

We were told that the exam scores are approximately normal.

The lecturer announced that the variance of the exam score over the class is 5 (just believe that this is the truth). Test at $\alpha = 0.01$ significance level whether the average midterm score is different from 16.

Solution:
Let μ be the average midterm score for the whole class.

Step 1: $H_0 : \mu = 16 \quad \text{vs} \quad H_1 : \mu \neq 16.$

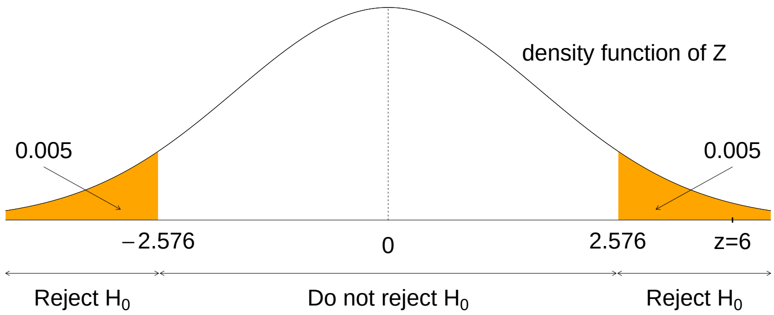
Step 2: Choose $\alpha = 0.01$.

Step 3: In this example $\sigma = \sqrt{5}$ is known, data are normal, and $n = 5$. Therefore the test statistic and its distribution is

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

Now $z_{\alpha/2} = z_{0.005} = 2.576$. Thus the rejection region is

$$z < -2.576 \quad \text{or} \quad z > 2.576.$$



Step 4: $z = (22-16)/(\sqrt{5}/\sqrt{5}) = 6 > 2.576.$

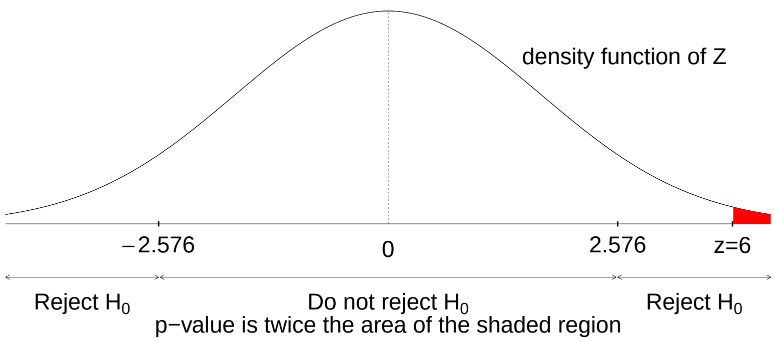
Step 5: As $z = 6$ falls in rejection region, H_0 is rejected.

Alternatively, we can use the p-value approach.

Note that the p-value is given, using a computer, as

$$2P(Z > 6) = 1.973175 \times 10^{-9},$$

which is smaller than $\alpha = 0.01$. So we reject H_0.



We can use our knowledge of the sampling distribution to determine the test statistic for other situations.

HYPOTHESIS TEST FOR THE MEAN: UNKNOWN VARIANCE:
Consider the case where

- the population variance σ^2 is unknown; AND
- the underlying distribution is normal.

For the null hypothesis $H_0 : \mu = \mu_0$, the test statistics is given by

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}.$$

Let t be the observed T value. For the alternative hypothesis

- $H_1 : \mu \neq \mu_0$, the rejection region is
$$t < -t_{n-1,\alpha/2} \quad \text{or} \quad t > t_{n-1,\alpha/2}.$$
- $H_1 : \mu < \mu_0$, the rejection region is
$$t < -t_{n-1,\alpha}.$$
- $H_1 : \mu > \mu_0$, the rejection region is
$$t > t_{n-1,\alpha}.$$ ■

REMARK:
When $n \geq 30$, we can replace t_{n-1} by Z, the standard normal distribution. ■

L-EXAMPLE 7.3 (MIDTERM EXAM SCORE II)
Continuing from the previous example. Let's say the lecturer didn't announce the variance, that is, σ is unknown.

The data given has $\bar{x} = 22$ and $s = 2.55$.

Test again at $\alpha = 0.01$ significance level whether the average midterm score is different from 16.

Solution:
Since σ is unknown, we need to perform a t-test.

Again, we let μ be the average midterm score for the whole class.

Step 1: $H_0 : \mu = 16 \quad \text{vs} \quad H_1 : \mu \neq 16.$

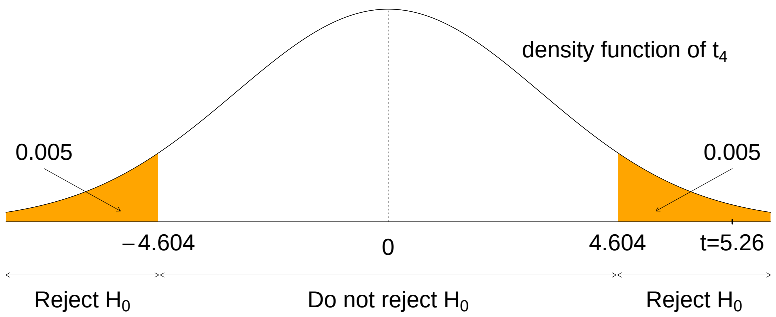
Step 2: Choose $\alpha = 0.01$.

Step 3: Since σ is unknown, data are normal, and $n = 5$, the test statistics is

$$T = \frac{\overline{X} - 16}{S/\sqrt{n}} \sim t_{n-1} = t_4.$$

Now $t_{n-1,\alpha/2} = t_{4,0.005} = 4.604$. So the rejection region is

$$t < -4.604 \quad \text{or} \quad t > 4.604.$$



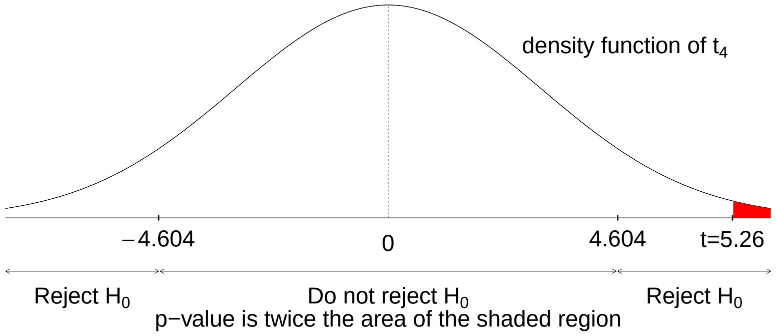
Step 4: $t = (22 - 16)/(2.55/\sqrt{5}) = 5.26$.

Step 5: Since $t = 5.26$ falls within the rejection region, we reject H_0.

Alternatively, the p-value can be found to be

$$2P(t_4 > 5.26) = 0.0063,$$

which is smaller than $\alpha = 0.01$, so we reject H_0.



L-EXAMPLE 7.4 (DEPARTMENT STORE)

A department store manager determines that a new billing system will be cost-effective only if the mean monthly account is more than \$170. It is known that the accounts has standard deviation \$65.

A random sample of 400 monthly accounts is drawn, for which the sample mean is \$178. Can we conclude that the new system will be cost-effective at 5% level of significance?

Solution:

Let μ be the mean monthly account.

Step 1: We test

$$H_0 : \mu = 170 \quad \text{vs} \quad H_1 : \mu > 170.$$

Step 2: Choose $\alpha = 0.05$.

Step 3: Since n is large and σ is known, we use the test statistic:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Under H_0, by the Central Limit Theorem, we have, $Z \sim N(0,1)$.

At a 5% significance level ($\alpha = 0.05$), we get

$$z_\alpha = z_{0.05} = 1.645.$$

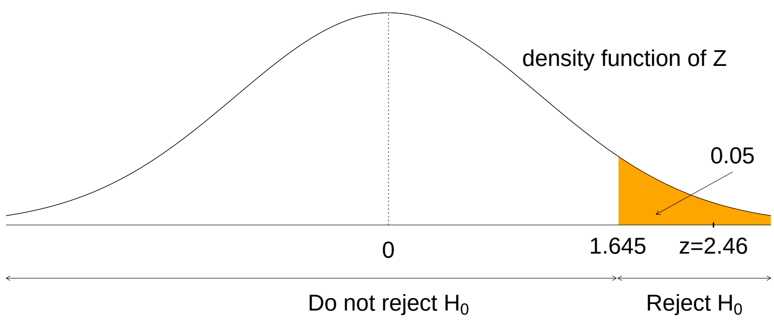
Step 4: We are given that

$$n = 400, \quad \bar{x} = 178, \quad \sigma = 65, \quad \alpha = 0.05$$

and so

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{178 - 170}{65/\sqrt{400}} = 2.46 > z_\alpha = 1.645.$$

Step 5: Therefore, we reject the null hypothesis and conclude that the mean monthly account is more than \$170.



3 Two-sided Tests and Confidence Intervals

In this section, we establish that the two-sided hypothesis test procedure is equivalent to finding a $100(1-\alpha)\%$ confidence interval for μ.

We illustrate using Case III: normal population, small n, unknown σ.

Once again, consider

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

The $100(1-\alpha)\%$ confidence interval for μ in this case is given by

$$\left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right).$$

If the $100(1-\alpha)\%$ confidence interval contains μ_0, we will have

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}.$$

Rearranging the above inequality, we obtain

$$-t_{\alpha/2} \leq \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \leq t_{\alpha/2}.$$

This means that the computed test statistic $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ satisfies

$$-t_{\alpha/2} \leq t \leq t_{\alpha/2}.$$

Note that the rejection region for this case is

$$t < -t_{\alpha/2} \quad \text{or} \quad t > t_{\alpha/2}.$$

This means that when the confidence interval contains μ_0, H_0 will not be rejected at level α.

Similarly, when the confidence interval does not contain μ_0, then

$$t > t_{\alpha/2} \quad \text{or} \quad t < -t_{\alpha/2}.$$

Thus t falls within the rejection region and so H_0 will be rejected.

Therefore confidence intervals can be used to perform two-sided tests.

EXAMPLE 7.7 (MIDTERM EXAM SCORE III)
Back to Example 7.6, regarding midterm exam scores. Assume that the lecturer did not announce the variance, i.e., σ is unknown.

The student constructed a 99% ($\alpha = 0.01$) confidence interval for the average score of students for the midterm:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 22 \pm 4.604 \times \frac{2.55}{\sqrt{5}} = (16.75, 27.25).$$

The interval does not contain 16, so the following test of hypothesis should be rejected at $\alpha = 0.01$:

$$H_0 : \mu = 16 \quad \text{vs} \quad H_1 : \mu \neq 16.$$

What about

$$H_0 : \mu = 17 \quad \text{vs} \quad H_1 : \mu \neq 17?$$

L-EXAMPLE 7.5
A study based on a sample size of 36 reported a mean of 87 with a margin of error of 10 for 95% confidence.

Give the 95% confidence interval for the population mean μ.

You are then asked to test the hypothesis that $\mu = 80$ against a two sided alternative at $\alpha = 0.05$. What is your conclusion?

Solution:
The 95% confidence interval for μ is given as

$$\bar{x} \pm E = 87 \pm 10 = (77, 97).$$

The 95% confidence interval contains the the value 80 so there is no evidence to reject the null at $\alpha = 0.05$.

4 TESTS COMPARING MEANS: INDEPENDENT SAMPLES

Suppose two independent samples are drawn from two populations with means μ_1 and μ_2. We are interested in testing

$$H_0 : \mu_1 - \mu_2 = \delta_0$$

against a suitable alternative hypothesis.

COMPARING MEANS: INDEPENDENT SAMPLES I:
(A) Consider the case where

- the population variances σ_1^2 and σ_2^2 are known; AND
- where
  - the underlying distributions are normal; OR
  - n_1, n_2 are sufficiently large (say, $n_1 \geq 30, n_2 \geq 30$).

For the null hypothesis $H_0 : \mu_1 - \mu_2 = \delta_0$, the test statistics is given by

$$Z = \frac{(\overline{X} - \overline{Y}) - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

(B) Consider the case where

- the population variances σ_1^2 and σ_2^2 are unknown; AND
- n_1, n_2 are sufficiently large (say, $n_1 \geq 30, n_2 \geq 30$). ■

For the null hypothesis $H_0 : \mu_1 - \mu_2 = \delta_0$, the test statistics is given by

$$Z = \frac{(\overline{X} - \overline{Y}) - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1).$$

The rejection regions or p-values can be established similarly as before.

REJECTION REGIONS AND p-VALUES:
For the null hypothesis $H_0 : \mu_1 - \mu_2 = \delta_0$, and specified alternative H_1, the rejection regions and p-values are given below.

H_1	Rejection Region	p-value
$\mu_1 - \mu_2 > \delta_0$	$z > z_\alpha$	$P(Z > \lvert z \rvert)$
$\mu_1 - \mu_2 < \delta_0$	$z < -z_\alpha$	$P(Z < -\lvert z \rvert)$
$\mu_1 - \mu_2 \neq \delta_0$	$z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$	$2P(Z > \lvert z \rvert)$

■

EXAMPLE 7.8
Analysis of a random sample consisting of $n_1 = 20$ specimens of cold-rolled steel to determine yield strengths resulted in a sample average strength of $\bar{x} = 29.8$ ksi.

A second random sample of $n_2 = 25$ two-side galvanized steel specimens gave a sample average strength of $\bar{y} = 34.7$ ksi.

Assuming that the two yield strength distributions are normal with $\sigma_1 = 4.0$ and $\sigma_2 = 5.0$, does the data indicate that the corresponding true average yield strengths μ_1 and μ_2 are different?
Use $\alpha = 0.01$.

Solution:
Let μ_1 and μ_2 be the mean strength of cold-rolled steel and two-side galvanized steel respectively.

Step 1: Note that $\delta_0 = 0$ in this example. So the hypotheses are

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Step 2: Set $\alpha = 0.01$.

Step 3: Test statistic and its distribution is given below:

$$Z = \frac{(\overline{X} - \overline{Y}) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0,1).$$

Note that $z_{\alpha/2} = z_{0.005} = 2.5782$. Thus the rejection region is

$$z > 2.5782 \quad \text{or} \quad z < -2.5782.$$

Step 4: Plug in the data,

$$z = \frac{(29.8 - 34.7) - 0}{\sqrt{\frac{16}{20} + \frac{25}{25}}} = -3.652 < -2.5782 = -z_{\alpha/2}.$$

Step 5: Since $z = -3.652$ falls inside the critical region, hence $H_0 : \mu_1 = \mu_2$ is rejected at the 1% level of significance. We conclude that the sample data strongly suggest that the true average yield strength for cold-rolled steel differs from that for galvanized steel.

Alternatively, we can compute the p-value to be

$$2 \times P(Z < -3.652) = 0.00026 < 0.01 = \alpha.$$

Thus we reject the null hypothesis at $\alpha = 0.01$ level.

L-EXAMPLE 7.6 (ELECTRICAL USAGE II)
As a baseline for a study on the effects of changing electrical pricing for electricity during peak hours, July usage during peak hours was obtained for $n_1 = 45$ homes with air-conditioning and $n_2 = 55$ homes without. The summarized results are provided below

	Samples		
population	Size	Mean	Variance
With	45	204.4	13,825.3
Without	55	130.0	8,632.0

Perform a hypothesis test at $\alpha = 0.05$ that the mean on-peak usage for homes with air-conditioning is higher than that for homes without.

Solution:
Let μ_1 and μ_2 be the mean on-peak usage for homes with and without air-conditioning respectively.

Step 1: Again we have $\delta_0 = 0$. So we test

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0.$$

Step 2: Set $\alpha = 0.05$.

Step 3: Test statistic and its distribution is given below:

$$Z = \frac{(\overline{X} - \overline{Y}) - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx N(0, 1).$$

The rejection region is $z > z_{0.05} = 1.645$.

Step 4: Plug in the data,

$$z = \frac{204.4 - 130.0 - 0}{\sqrt{\frac{13{,}825.3}{45} + \frac{8632.0}{55}}} = 3.45.$$

Step 5: We reject H_0 since $z = 3.45 > z_{0.05} = 1.645$.

COMPARING MEANS: INDEPENDENT SAMPLES II:
Consider the case where

- the population variances σ_1^2 and σ_2^2 are unknown but equal;
- the underlying distributions are normal;
- n_1, n_2 are small (say, $n_1 < 30, n_2 < 30$). ■

For the null hypothesis $H_0 : \mu_1 - \mu_2 = \delta_0$, the test statistics is given by

$$Z = \frac{(\overline{X} - \overline{Y}) - \delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

L-EXAMPLE 7.7 (COAL SPECIMENS)

The following are measurements of heat-producing capacity (millions of calories per ton) of sample specimens of coal from two mines:

```
Mine 1: 8260 8130 8350 8070 8340
Mine 2: 7950 7890 7900 8140 7920 7840
```

The sample summary statistics are

$$\bar{x} = 8230, \quad s_1 = 125.5, \quad \bar{y} = 7940, \quad s_2 = 104.5.$$

Assume that both populations are normal with equal variance. Test at $\alpha = 0.01$ level if the means between these two mines are different.

Solution:

Let μ_1 and μ_2 be the mean heat-producing capacity for the two mines.

Step 1: $H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq 0.$

Step 2: $\alpha = 0.01.$

Step 3: We are given that $s_1 = 125.5, s_2 = 104.5$ and that the equal variance assumption holds. The test statistics is

$$T = \frac{(\overline{X} - \overline{Y}) - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

Since $t_{n_1+n_2-2,\alpha/2} = t_{9,0.005} = 3.250$, the rejection region is

$$t < -3.250 \quad \text{or} \quad t > 3.250.$$

Step 4: Plug in everything, we get

$$\begin{aligned} s_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\ &= \frac{(5-1)\times 125.5^2 + (6-1)\times 104.5^2}{5+7-2} = 13066.92, \end{aligned}$$

and

$$t = \frac{(8230 - 7940) - 0}{\sqrt{13066.92} \times \sqrt{\frac{1}{5} + \frac{1}{6}}} = 4.18963.$$

Step 5: Since $t = 4.18963$ is in the rejection region, we reject H_0.

5 Tests Comparing Means: Paired Data

Comparing means with matched-pairs data is easy. We merely use methods we have already learned for single samples.

Comparing Means: Paired Data:
For paired data, define $D_i = X_i - Y_i$.

For the null hypothesis $H_0 : \mu_D = \mu_{D_0}$, the test statistics is given by

$$T = \frac{\overline{D} - \mu_{D_0}}{S_D/\sqrt{n}}.$$

- If $n < 30$ and the population is normally distributed then

$$T \sim t_{n-1}.$$

- If $n \geq 30$, then

$$T \sim N(0,1).$$

Example 7.9 (Treating catalyst surfaces)
Prof X developed a new procedure for treating catalyst surfaces which he claims will result in a significant enhancement in the number of active sites.

The number of active sites can be determined by absorption of H_2 gas.

Prof X tested each sample before and after the treatment and obtained the following H_2 uptake in terms of mmol/g.

Sample No.	Before treatment (X)	After treatment (Y)	Difference (D)
1	165	172	7
2	146	189	43
3	174	168	−6
4	186	176	−10
5	147	198	51
6	153	184	31
7	132	188	56
8	175	197	22

The summary statistics for the variable D are $\overline{d} = 24.25$ and $s_D = 25.34$.

Has the treatment resulted in an increase in the number of active sites on the catalyst surfaces? Assume normality, and test at $\alpha = 0.05$ level.

Solution:
Note that in such a setup the two samples are not independent, and so the two sample t-test does not apply.

Define $D_i = Y_i - X_i$, where X_i and Y_i are the "before treatment" and "after treatment" readings.

The question is now reduced to:

Do the data give any evidence that $\mu_D > 0$?

Step 1: We set the null and alternative to be

$$H_0 : \mu_D = 0 \quad \text{vs} \quad H_1 : \mu_D > 0.$$

Step 2: Set $\alpha = 0.05$.

Step 3: We use the paired t-test with the test statistics

$$T = \frac{\overline{D} - 0}{S_D/\sqrt{n}}.$$

The rejection region is $t > t_{7,0.05} = 1.895$.

Step 4: The observed t value is

$$t = \frac{\overline{d} - 0}{s_D/\sqrt{n}} = \frac{24.25 - 0}{25.34/\sqrt{8}} = 2.70 > 1.895.$$

Step 5: Since $t = 2.70 > t_{7,0.05} = 1.895$, we reject H_0 and conclude that there is evidence that treatment of catalysts increases the number of active sites.

As an aside, the p-value is

$$P(t_7 > t) = P(t_7 > 2.70) = 0.0153,$$

which is smaller than 0.05.

L-EXAMPLE 7.8 (WATER TREATMENT)
A state law requires municipal waste water treatment plants to monitor their discharges into rivers and streams. A treatment plant could choose to send its samples to a commercial laboratory of its choosing.

Concern over this self-monitoring led a civil engineer to design a matched pairs experiment. Exactly the same bottle of effluent cannot be sent to two different laboratories. To match "identical" as closely as possible, she would take a sample of effluent in a large sample bottle and pour it back and forth over two open specimen bottles.

When they were filled and capped, a coin was flipped to see if the one on the right was sent to commercial laboratory or the state laboratory.

This process was repeated 11 times. The results, for the response suspended solids (SS) are

Sample	1	2	3	4	5	6	7	8	9	10	11
Commercial lab	27	23	64	44	30	75	26	124	54	30	14
State lab	15	13	22	29	31	64	30	64	56	20	21
Difference $X_i - Y_i$	12	10	42	15	-1	11	-4	60	-2	10	-7

The summary statistics for $D = X_i - Y_i$ are

$$\overline{d} = 13.27, s_D^2 = 418.61.$$

Conduct a hypothesis test to check if the SS from the commercial lab is higher than those from state lab at significance level 0.05. Assume a normal distribution for the population.

Solution:
We shall test

$$H_0 : \mu_D = 0 \quad \text{vs} \quad H_1 : \mu_D > 0.$$

The test statistics is

$$T = \frac{\overline{D} - 0}{S_D/\sqrt{n}},$$

and the rejection region is $t > t_{10,0.05} = 1.812$.

Computations gives the observed test statistics as

$$t = \frac{\overline{d} - 0}{\sqrt{418.61/11}} = 2.15 > 1.812.$$

Since $t = 2.15 > t_{10,0.05} = 1.812$, we reject H_0 and conclude that the response from commercial lab is higher than those from the state lab.

L-EXAMPLE 7.9 (MEAN RELATED INFERENCE)
Assume that we are to make inference, including constructing the confidence interval or perform a hypothesis test, concerning a mean-related parameters θ. A general procedure is as follows:

- ✓ Step 1: Look for an estimator $\widehat{\theta}$ for θ, e.g., $\overline{X}$ for μ.
- ✓ Step 2: Derive the formula for $\text{var}(\widehat{\theta})$.
- ✓ Step 3: We then construct

$$T = \frac{\widehat{\theta} - \theta}{\sqrt{V}}. \tag{7.1}$$

For V, we consider the following possibilities.

- ⋆ If $\text{var}(\widehat{\theta})$ does not depend on any unknown parameter, e.g., when σ^2 is known, $\text{var}(\overline{X}) = \sigma^2/n$, we set $V = \text{var}(\widehat{\theta})$. The statistic T given in (7.1) (approximately) follows the $N(0,1)$ distribution, when the data are normal or the sample size is sufficiently large.

⋆ If $\text{var}(\widehat{\theta})$ contains some other unknown parameters, e.g., σ^2, we replace the parameter with its estimator, e.g., S^2 can be used to replace σ^2, and result in $\widehat{\text{var}}(\widehat{\theta})$. We set $V = \widehat{\text{var}}(\widehat{\theta})$. The distribution of T given in (7.1) has two possibilities:

(1) The sample size is sufficiently large; then $T \sim N(0,1)$ approximately.

(2) If the sample size is small, but the observations are normally distributed, then $T \sim t_{df}$, where df is the degrees of the freedom of the parameter estimated in $\text{var}(\widehat{\theta})$.

Now, statistical inference for θ can be done based on T and its associated distribution discussed above.

- Construct $(1-\alpha)$ confidence interval:

$$\widehat{\theta} \pm M\sqrt{V},$$

  where M, based on the distribution of T, is either $z_{\alpha/2}$ or $t_{df,\alpha/2}$.

- Test statistic for hypothesis test with $H_0 : \theta = \theta_0$ is given by

$$T = \frac{\widehat{\theta} - \theta_0}{\sqrt{V}},$$

  with its distribution resulted from above.