SL 4.11 Statistical Tests: t-test

Introduction

In this presentation we want introduce the t-test.

t-test

There are many types of t-tests. The one that you need to know is designed to compare means of the same variable of two populations. In order to compare these means we take a random sample from each population and compare the means of these samples. So the test we're using is a 2-sample t-test.

It's important to understand that the means of the two samples will almost never be exactly equal to each other.Let us consider an example to illustrate the whole process.

t-test

There are many types of t-tests. The one that you need to know is designed to compare means of the same variable of two populations. In order to compare these means we take a random sample from each population and compare the means of these samples. So the test we're using is a 2-sample t-test.

It's important to understand that the means of the two samples will almost never be exactly equal to each other.Let us consider an example to illustrate the whole process.

t-test

There are many types of t-tests. The one that you need to know is designed to compare means of the same variable of two populations. In order to compare these means we take a random sample from each population and compare the means of these samples. So the test we're using is a 2-sample t-test.

It's important to understand that the means of the two samples will almost never be exactly equal to each other.Let us consider an example to illustrate the whole process.

t-test

There are many types of t-tests. The one that you need to know is designed to compare means of the same variable of two populations. In order to compare these means we take a random sample from each population and compare the means of these samples. So the test we're using is a 2-sample t-test.

It's important to understand that the means of the two samples will almost never be exactly equal to each other. Let us consider an example to illustrate the whole process.

t-test

There are many types of t-tests. The one that you need to know is designed to compare means of the same variable of two populations. In order to compare these means we take a random sample from each population and compare the means of these samples. So the test we're using is a 2-sample t-test.

It's important to understand that the means of the two samples will almost never be exactly equal to each other.Let us consider an example to illustrate the whole process.

Apple tree example

There are two apple trees. One grows in more sunny side of the field, lets call this tree A. So tree A receives more sunlight than tree B. We want to compare the weight of the apples from each tree. We may want to see if tree A produces larger (heavier) apples.

In order to do this we will gather some apples from each tree (these are our 2 samples). We will calculate the means weight of each sample and compare these numbers.

Apple tree example

There are two apple trees. One grows in more sunny side of the field, lets call this tree A. So tree A receives more sunlight than tree B. We want to compare the weight of the apples from each tree. We may want to see if tree A produces larger (heavier) apples.

In order to do this we will gather some apples from each tree (these are our 2 samples). We will calculate the means weight of each sample and compare these numbers.

Apple tree example

There are two apple trees. One grows in more sunny side of the field, let's call this tree A. So tree A receives more sunlight than tree B. We want to compare the weight of the apples from each tree. We may want to see if tree A produces larger (heavier) apples.

In order to do this we will gather some apples from each tree (these are our 2 samples). We will calculate the means weight of each sample and compare these numbers.

Apple tree example

There are two apple trees. One grows in more sunny side of the field, let's call this tree A. So tree A receives more sunlight than tree B. We want to compare the weight of the apples from each tree. We may want to see if tree A produces larger (heavier) apples.

In order to do this we will gather some apples from each tree (these are our 2 samples). We will calculate the means weight of each sample and compare these numbers.

Apple tree example

Suppose the apples from the tree A weigh [in grams]:
82, 90, 80, 88, 87, 85, 83.

The apples from tree B weigh [in grams]:
81, 88, 82, 83, 76, 84, 80, 83, 82, 81.

How can we analyse this data?

Apple tree example

The means are $\bar{x}_A = 85\ g$ and $\bar{x}_B = 82\ g$. What can we conclude? Not much.

Of course the means are not the same. It would be very surprising if they were. Imagine if we picked two samples from the same tree. Would you expect the means of these samples to be exactly the same? No, they would most likely be similar, but not the same.

This is why we use the t-test. The t-test answers the question (loosely speaking): how likely it is that the means differ by this much **if the trees produce apples of the same weight.**

Apple tree example

The means are $\bar{x}_A = 85\ g$ and $\bar{x}_B = 82\ g$. What can we conclude? Not much.

Of course the means are not the same. It would be very surprising if they were. Imagine if we picked two samples from the same tree. Would you expect the means of these samples to be exactly the same? No, they would most likely be similar, but not the same.

This is why we use the t-test. The t-test answers the question (loosely speaking): how likely it is that the means differ by this much **if the trees produce apples of the same weight.**

Apple tree example

The means are $\bar{x}_A = 85\ g$ and $\bar{x}_B = 82\ g$. What can we conclude? Not much.

Of course the means are not the same. It would be very surprising if they were. Imagine if we picked two samples from the same tree. Would you expect the means of these samples to be exactly the same? No, they would most likely be similar, but not the same.

This is why we use the t-test. The t-test answers the question (loosely speaking): how likely it is that the means differ by this much **if the trees produce apples of the same weight.**

Apple tree example

The means are $\bar{x}_A = 85\ g$ and $\bar{x}_B = 82\ g$. What can we conclude? Not much.

Of course the means are not the same. It would be very surprising if they were. Imagine if we picked two samples from the same tree. Would you expect the means of these samples to be exactly the same? No, they would most likely be similar, but not the same.

This is why we use the t-test. The t-test answers the question (loosely speaking): how likely it is that the means differ by this much **if the trees produce apples of the same weight**.

Apple tree example

The means are $\bar{x}_A = 85\ g$ and $\bar{x}_B = 82\ g$. What can we conclude? Not much.

Of course the means are not the same. It would be very surprising if they were. Imagine if we picked two samples from the same tree. Would you expect the means of these samples to be exactly the same? No, they would most likely be similar, but not the same.

This is why we use the t-test. The t-test answers the question (loosely speaking): how likely it is that the means differ by this much **if the trees produce apples of the same weight.**

Apple tree example

The means are $\bar{x}_A = 85\ g$ and $\bar{x}_B = 82\ g$. What can we conclude? Not much.

Of course the means are not the same. It would be very surprising if they were. Imagine if we picked two samples from the same tree. Would you expect the means of these samples to be exactly the same? No, they would most likely be similar, but not the same.

This is why we use the t-test. The t-test answers the question (loosely speaking): how likely it is that the means differ by this much **if the trees produce apples of the same weight**.

Apple tree example

The means are $\bar{x}_A = 85\ g$ and $\bar{x}_B = 82\ g$. What can we conclude? Not much.

Of course the means are not the same. It would be very surprising if they were. Imagine if we picked two samples from the same tree. Would you expect the means of these samples to be exactly the same? No, they would most likely be similar, but not the same.

This is why we use the t-test. The t-test answers the question (loosely speaking): how likely it is that the means differ by this much **if the trees produce apples of the same weight**.

t-test

We start be writing down the null and alternative hypothesis:

- H_0 The trees produce apples of the same weight. In other words the mean all apples produced by tree A is the same as the mean of all trees produced by tree B: $\mu_A = \mu_B$.
- H_1 Tree A produces larger (heavier) apples than tree B: $\mu_A > \mu_B$

Then we need to choose the significance level for our test. Let's say we choose significance level to be 5%.

Now we will use the GDC.

t-test

We start be writing down the null and alternative hypothesis:

- H_0 The trees produce apples of the same weight. In other words the mean all apples produced by tree A is the same as the mean of all trees produced by tree B: $\mu_A = \mu_B$.
- H_1 Tree A produces larger (heavier) apples than tree B: $\mu_A > \mu_B$

Then we need to choose the significance level for our test. Let's say we choose significance level to be 5%.

Now we will use the GDC.

t-test

We start be writing down the null and alternative hypothesis:

- H_0 The trees produce apples of the same weight. In other words the mean all apples produced by tree A is the same as the mean of all trees produced by tree B: $\mu_A = \mu_B$.
- H_1 Tree A produces larger (heavier) apples than tree B: $\mu_A > \mu_B$

Then we need to choose the significance level for our test. Let's say we choose significance level to be 5%.

Now we will use the GDC.

t-test

We start be writing down the null and alternative hypothesis:

- H_0 The trees produce apples of the same weight. In other words the mean all apples produced by tree A is the same as the mean of all trees produced by tree B: $\mu_A = \mu_B$.
- H_1 Tree A produces larger (heavier) apples than tree B: $\mu_A > \mu_B$

Then we need to choose the significance level for our test. Let's say we choose significance level to be 5%.

Now we will use the GDC.

t-test GDC

Go to STAT and input the data into lists. List 1 should contain weights of apples from tree A, and list 2 weights of apples from tree B.

Now go TESTS and choose 2-SAMPLE t-test.Now we want:

Data: List (and you want to pick the appropriate lists below),

$\mu_A > \mu_B$

Pooled: ON.

Now press CALCULATE.

You should get $p = 0.0410$.

t-test GDC

Go to STAT and input the data into lists. List 1 should contain weights of apples from tree A, and list 2 weights of apples from tree B.

Now go TESTS and choose 2-SAMPLE t-test. Now we want:

Data: List (and you want to pick the appropriate lists below).
$\mu_A > \mu_B$
Pooled: ON.

Now press CALCULATE.

You should get $p = 0.0410$.

t-test GDC

Go to STAT and input the data into lists. List 1 should contain weights of apples from tree A, and list 2 weights of apples from tree B.

Now go TESTS and choose 2-SAMPLE t-test.Now we want:

Data: List (and you want to pick the appropriate lists below).
$\mu_A > \mu_B$
Pooled: ON.

Now press CALCULATE.

You should get $p = 0.0410$.

t-test GDC

Go to STAT and input the data into lists. List 1 should contain weights of apples from tree A, and list 2 weights of apples from tree B.

Now go TESTS and choose 2-SAMPLE t-test.Now we want:

Data: List (and you want to pick the appropriate lists below),
$\mu_A > \mu_B$
Pooled: ON.

Now press CALCULATE.

You should get $p = 0.0410$.

t-test GDC

Go to STAT and input the data into lists. List 1 should contain weights of apples from tree A, and list 2 weights of apples from tree B.

Now go TESTS and choose 2-SAMPLE t-test.Now we want:

Data: List (and you want to pick the appropriate lists below),
$\mu_A > \mu_B$
Pooled: ON.

Now press CALCULATE.

You should get $p = 0.0410$.

t-test GDC

Go to STAT and input the data into lists. List 1 should contain weights of apples from tree A, and list 2 weights of apples from tree B.

Now go TESTS and choose 2-SAMPLE t-test.Now we want:

Data: List (and you want to pick the appropriate lists below),
$\mu_A > \mu_B$
Pooled: ON.

Now press CALCULATE.

You should get $p = 0.0410$.

t-test GDC

Because $p < 0.05$, so p is smaller than the chosen significance level, we conclude that there is evidence to **reject** H_0. So it is likely that tree A indeed produces heavier apples.

The above test was an example of a **one-tailed** t-test. It is one-tailed because we've only considered one alternative hypothesis, namely $\mu_A > \mu_B$. The alternative hypothesis for a two-tailed test would be $\mu_A \neq \mu_B$.

In general it's quite obvious which of the two (one-tailed or two-tailed) test we want to perform. If we want to test if the quantity has increased, is larger, has improved (or decreased, is smaller, has worsened) etc. we use one-tailed test. If we want to test if the quantity has changed, is different, then we use a two-tailed test.

The above test was an example of a **one-tailed** t-test. It is one-tailed because we've only considered one alternative hypothesis, namely $\mu_A > \mu_B$. The alternative hypothesis for a two-tailed test would be $\mu_A \neq \mu_B$.

In general it's quite obvious which of the two (one-tailed or two-tailed) test we want to perform. If we want to test if the quantity has increased, is larger, has improved (or decreased, is smaller, has worsened) etc. we use one-tailed test. If we want to test if the quantity has changed, is different, then we use a two-tailed test.

The above test was an example of a **one-tailed** t-test. It is one-tailed because we've only considered one alternative hypothesis, namely $\mu_A > \mu_B$. The alternative hypothesis for a two-tailed test would be $\mu_A \neq \mu_B$.

In general it's quite obvious which of the two (one-tailed or two-tailed) test we want to perform. If we want to test if the quantity has increased, is larger, has improved (or decreased, is smaller, has worsened) etc. we use one-tailed test. If we want to test if the quantity has changed, is different, then we use a two-tailed test.

The above test was an example of a **one-tailed** t-test. It is one-tailed because we've only considered one alternative hypothesis, namely $\mu_A > \mu_B$. The alternative hypothesis for a two-tailed test would be $\mu_A \neq \mu_B$.

In general it's quite obvious which of the two (one-tailed or two-tailed) test we want to perform. If we want to test if the quantity has increased, is larger, has improved (or decreased, is smaller, has worsened) etc. we use one-tailed test. If we want to test if the quantity has changed, is different, then we use a two-tailed test.

The underlying assumption of the t-test is that the quantity we're measuring is **normally distributed**. This is likely to be true for weight of apples (and many, many other quantities).

If the two populations we want to compare have the same standard deviation (again this is plausible for apples from two different trees), then we use a **pooled** test. In IB examination you always want to use the pooled test.

The underlying assumption of the t-test is that the quantity we're measuring is **normally distributed**. This is likely to be true for weight of apples (and many, many other quantities).

If the two populations we want to compare have the same standard deviation (again this is plausible for apples from two different trees), then we use a **pooled** test. In IB examination you always want to use the pooled test.

The underlying assumption of the t-test is that the quantity we're measuring is **normally distributed**. This is likely to be true for weight of apples (and many, many other quantities).

If the two populations we want to compare have the same standard deviation (again this is plausible for apples from two different trees), then we use a **pooled** test. In IB examination you always want to use the pooled test.

t-test - exam question

Ms Calhoun measures the heights of students in her mathematics class. She is interested to see if the mean height of male students, μ_1, is the same as the mean height of female students, μ_2. The information is recorded in the table.

Male height (cm)	150	148	143	152	151	149	147	
Female height (cm)	148	152	154	147	146	153	152	150

At the 10% level of significance, a t-test was used to compare the means of the two groups. The data is assumed to be normally distributed and the standard deviations are equal between the two groups.

(a) (i) State the null hypothesis.

(ii) State the alternative hypothesis. [2]

(b) Calculate the p-value for this test. [2]

(c) State, giving a reason, whether Ms Calhoun should accept the null hypothesis. [2]

t-test - exam question

First note two things:

- the data is normally distributed (so the t-test can be conducted),
- the standard deviations are equal (we will use **pooled** test).

Note that this will always be the case.

Do we want to use a one-tailed or two-tailed test? Two-tailed! We only want to know if the means are the same or different.

t-test - exam question

First note two things:

- the data is normally distributed (so the t-test can be conducted),
- the standard deviations are equal (we will use **pooled** test).

Note that this will always be the case.

Do we want to use a one-tailed or two-tailed test? Two-tailed! We only want to know if the means are the same or different.

t-test - exam question

First note two things:

- the data is normally distributed (so the t-test can be conducted),
- the standard deviations are equal (we will use **pooled** test).

Note that this will always be the case.

Do we want to use a one-tailed or two-tailed test? Two-tailed! We only want to know if the means are the same or different.

t-test - exam question

First note two things:

- the data is normally distributed (so the t-test can be conducted),
- the standard deviations are equal (we will use **pooled** test).

Note that this will always be the case.

Do we want to use a one-tailed or two-tailed test? Two-tailed! We only want to know if the means are the same or different.

t-test - exam question

First note two things:

- the data is normally distributed (so the t-test can be conducted),
- the standard deviations are equal (we will use **pooled** test).

Note that this will always be the case.

Do we want to use a one-tailed or two-tailed test? Two-tailed! We only want to know if the means are the same or different.

t-test - exam question

The null hypothesis is that the means are the same $\mu_1 = \mu_2$. The alternative hypothesis is that the means are different $\mu_1 \neq \mu_2$.

Now we put the lists into the GDC. Remember to put alternative hypothesis as $\mu_1 \neq \mu_2$ and pooled test ON.

You should get $p = 0.296$. So we have that $p > 0.1$, the p-value is greater than the significance level (10%), so there is no evidence to reject the null hypothesis.

t-test - exam question

The null hypothesis is that the means are the same $\mu_1 = \mu_2$. The alternative hypothesis is that the means are different $\mu_1 \neq \mu_2$.

Now we put the lists into the GDC. Remember to put alternative hypothesis as $\mu_1 \neq \mu_2$ and pooled test ON.

You should get $p = 0.296$. So we have that $p > 0.1$, the p-value is greater than the significance level (10%), so there is no evidence to reject the null hypothesis.

t-test - exam question

The null hypothesis is that the means are the same $\mu_1 = \mu_2$. The alternative hypothesis is that the means are different $\mu_1 \neq \mu_2$.

Now we put the lists into the GDC. Remember to put alternative hypothesis as $\mu_1 \neq \mu_2$ and pooled test ON.

You should get $p = 0.296$. So we have that $p > 0.1$, the p-value is greater than the significance level (10%), so there is no evidence to reject the null hypothesis.

t-test - exam question

The null hypothesis is that the means are the same $\mu_1 = \mu_2$. The alternative hypothesis is that the means are different $\mu_1 \neq \mu_2$.

Now we put the lists into the GDC. Remember to put alternative hypothesis as $\mu_1 \neq \mu_2$ and pooled test ON.

You should get $p = 0.296$. So we have that $p > 0.1$, the p-value is greater than the significance level (10%), so there is no evidence to reject the null hypothesis.

t-test - exam question

The null hypothesis is that the means are the same $\mu_1 = \mu_2$. The alternative hypothesis is that the means are different $\mu_1 \neq \mu_2$.

Now we put the lists into the GDC. Remember to put alternative hypothesis as $\mu_1 \neq \mu_2$ and pooled test ON.

You should get $p = 0.296$. So we have that $p > 0.1$, the p-value is greater than the significance level (10%), so there is no evidence to reject the null hypothesis.

t-test - exam question

The null hypothesis is that the means are the same $\mu_1 = \mu_2$. The alternative hypothesis is that the means are different $\mu_1 \neq \mu_2$.

Now we put the lists into the GDC. Remember to put alternative hypothesis as $\mu_1 \neq \mu_2$ and pooled test ON.

You should get $p = 0.296$. So we have that $p > 0.1$, the p-value is greater than the significance level (10%), so there is no evidence to reject the null hypothesis.