American University of Beirut

STAT 210: Elementary Statistics for Sciences

2022–2023 Fall

Siamak Taati

Chapter 10

Estimation and hypothesis testing about two populations

Part 2

Comparing two populations

Comparing two populations

Inference about two populations

Let θ_1 and θ_2 be the values of a certain parameter for two populations.

- Hypothesis testing: In many examples, we need to use statistical evidence to judge whether $\theta_1 = \theta_2$ or not.

  Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta_1 = \theta_2$	H_0 $\theta_1 = \theta_2$	H_0 $\theta_1 = \theta_2$
H_1 $\theta_1 < \theta_2$	H_1 $\theta_1 \neq \theta_2$	H_1 $\theta_1 > \theta_2$

- Estimation: We may also need to estimate the difference $\theta_1 - \theta_2$ (i.e., find a confidence interval for it).

Comparing two populations

Inference about two populations

Let θ_1 and θ_2 be the values of a certain parameter for two populations.

- Hypothesis testing: In many examples, we need to use statistical evidence to judge whether $\theta_1 = \theta_2$ or not.

  Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta_1 - \theta_2 = 0$	H_0 $\theta_1 - \theta_2 = 0$	H_0 $\theta_1 - \theta_2 = 0$
H_1 $\theta_1 - \theta_2 < 0$	H_1 $\theta_1 - \theta_2 \neq 0$	H_1 $\theta_1 - \theta_2 > 0$

- Estimation: We may also need to estimate the difference $\theta_1 - \theta_2$ (i.e., find a confidence interval for it).

Comparing two populations

Inference about two populations

Let θ_1 and θ_2 be the values of a certain parameter for two populations.

- Hypothesis testing: In many examples, we need to use statistical evidence to judge whether $\theta_1 = \theta_2 + a$ or not.

  Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta_1 - \theta_2 = a$	H_0 $\theta_1 - \theta_2 = a$	H_0 $\theta_1 - \theta_2 = a$
H_1 $\theta_1 - \theta_2 < a$	H_1 $\theta_1 - \theta_2 \neq a$	H_1 $\theta_1 - \theta_2 > a$

- Estimation: We may also need to estimate the difference $\theta_1 - \theta_2$ (i.e., find a confidence interval for it).

Comparing two populations

Statistical evidence

The evidence consists in a sample from population #1 and a sample from population #2.

The two samples can either be independent or dependent.

- Independent samples are two independent collections:

  $$X_1, X_2, X_3, \ldots, X_{n_1} \qquad\qquad Y_1, Y_2, Y_3, \ldots, Y_{n_2}$$

  with no connections between them.

- Dependent samples are often in the form of a paired sample:

  $$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots, (X_n, Y_n)\ .$$

Comparing two populations

Statistical evidence

The evidence consists in a sample from population #1 and a sample from population #2.

The two samples can either be independent or dependent.

- Independent samples are two independent collections:

  $$X_1, X_2, X_3, \ldots, X_{n_1} \qquad\qquad Y_1, Y_2, Y_3, \ldots, Y_{n_2}$$

  with no connections between them.

  Example: The "herbal remedy for blood pressure" example.

- Dependent samples are often in the form of a paired sample:

  $$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots, (X_n, Y_n) \ .$$

  Example: The "chewing tobacco and heart rate" example.

Comparing two populations

We focus on the problems of inference about the population means and about the population proportions in two difference populations.

Comparing two populations

Inference about population means

The procedure depends on the answers to the following questions:

Comparing two populations

Inference about population means

The procedure depends on the answers to the following questions:

Q1 Are the samples independent or paired?

Comparing two populations

Inference about population means

The procedure depends on the answers to the following questions:

- Q1 Are the samples independent or paired?
- Q2 Are the samples large or small?

Comparing two populations

Inference about population means

The procedure depends on the answers to the following questions:

- Q1 Are the samples independent or paired?
- Q2 Are the samples large or small?
- Q3 Are the populations (approximately) normally distributed?

Comparing two populations

Inference about population means

The procedure depends on the answers to the following questions:

- Q1 Are the samples independent or paired?
- Q2 Are the samples large or small?
- Q3 Are the populations (approximately) normally distributed?
- Q4 Do we know the population standard deviations?

Comparing two populations

Inference about population means

The procedure depends on the answers to the following questions:

- Q1 Are the samples independent or paired?
- Q2 Are the samples large or small?
- Q3 Are the populations (approximately) normally distributed?
- Q4 Do we know the population standard deviations?
- Q5 If not, are the (unknown) standard deviations the same?

Comparing two populations

Inference about population means

We consider the following scenarios:

1. independent samples, σ_1 and σ_2 known,
   - (a) normal populations
   - (b) large samples
2. independent samples, σ_1 and σ_2 unknown but $\sigma_1 = \sigma_2$,
   - (a) normal populations
   - (b) large samples
3. independent samples, σ_1 and σ_2 unknown and $\sigma_1 \neq \sigma_2$,
   - (a) normal populations
   - (b) large samples
4. paired sample,
   - (a) normal population of pair differences
   - (b) large sample

Comparing two populations

Inference about population means

We consider the following scenarios:

1. independent samples, σ_1 and σ_2 known,
   - (a) normal populations
   - (b) large samples
2. independent samples, σ_1 and σ_2 unknown but $\sigma_1 = \sigma_2$,
   - (a) normal populations
   - (b) large samples
3. independent samples, σ_1 and σ_2 unknown and $\sigma_1 \neq \sigma_2$,
   - (a) normal populations
   - (b) large samples
4. paired sample,
   - (a) normal population of pair differences
   - (b) large sample

Inference about population proportions

We consider the following scenario:

1. large independent samples

Comparing two populations

Inference about population means

We consider the following scenarios:

1. independent samples, σ_1 and σ_2 known,
   (a) normal populations
   (b) large samples
2. independent samples, σ_1 and σ_2 unknown but $\sigma_1 = \sigma_2$,
   (a) normal populations
   (b) large samples
3. independent samples, σ_1 and σ_2 unknown and $\sigma_1 \neq \sigma_2$,
   (a) normal populations
   (b) large samples
4. paired sample,
   (a) normal population of pair differences
   (b) large sample

Inference about population proportions

We consider the following scenario:

1. large independent samples

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

To test the claim about the effectiveness of a type of herbal tea in reducing blood pressure among adults with chronic hypertension, researchers conduct an experiment involving 50 subjects.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

To test the claim about the effectiveness of a type of herbal tea in reducing blood pressure among adults with chronic hypertension, researchers conduct an experiment involving 50 subjects.

Q What are the competing hypotheses?

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

To test the claim about the effectiveness of a type of herbal tea in reducing blood pressure among adults with chronic hypertension, researchers conduct an experiment involving 50 subjects.

Q What are the competing hypotheses?

A H$_0$ $\mu_{\text{t}} = \mu_{\text{c}}$ (the herbal tea is not effective)

H$_1$ $\mu_{\text{t}} < \mu_{\text{c}}$ (the herbal tea is effective)

where

- μ_{t} is the mean blood pressure among the population of adults with chronic hypertension who receive the treatment
- μ_{c} is the mean blood pressure among the population of those who do not receive the treatment.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

To test the claim about the effectiveness of a type of herbal tea in reducing blood pressure among adults with chronic hypertension, researchers conduct an experiment involving 50 subjects.

Q What are the competing hypotheses?

A H$_0$ $\mu_\mathrm{t} - \mu_\mathrm{c} = 0$ (the herbal tea is not effective)

H$_1$ $\mu_\mathrm{t} - \mu_\mathrm{c} < 0$ (the herbal tea is effective)

where

- μ_t is the mean blood pressure among the population of adults with chronic hypertension who receive the treatment
- μ_c is the mean blood pressure among the population of those who do not receive the treatment.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\mathrm{H}_0\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} = 0 \qquad \mathrm{H}_1\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} < 0}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0\text{: } \mu_t - \mu_c = 0 \qquad H_1\text{: } \mu_t - \mu_c < 0}$$

Of the 50 subjects, 27 end up in the treatment group and the remaining 23 in the control group.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\mathrm{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \mathrm{H}_1\text{: } \mu_t - \mu_c < 0}$$

Of the 50 subjects, 27 end up in the treatment group and the remaining 23 in the control group.

- The treatment group will receive one glass of the herbal tea every night for a whole month.
- The control group will receive placebo for the same duration.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} = 0 \qquad \text{H}_1\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} < 0}$$

Of the 50 subjects, 27 end up in the treatment group and the remaining 23 in the control group.

- The treatment group will receive one glass of the herbal tea every night for a whole month.
- The control group will receive placebo for the same duration.

The researchers then plan to measure the blood pressures of the two groups and compare them.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0\text{: } \mu_t - \mu_c = 0 \qquad H_1\text{: } \mu_t - \mu_c < 0}$$

Of the 50 subjects, 27 end up in the treatment group and the remaining 23 in the control group.

- The treatment group will receive one glass of the herbal tea every night for a whole month.
- The control group will receive placebo for the same duration.

The researchers then plan to measure the blood pressures of the two groups and compare them.

Q What is a suitable test statistic?

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\text{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\text{c}}$ be the mean blood pressure of the subjects in the control group.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\mathrm{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\mathrm{c}}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\mathrm{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\mathrm{c}}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$?

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\mathrm{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\mathrm{c}}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$?

A Recall that

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_\mathrm{t}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_\mathrm{c}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_\mathrm{t} - \overline{X}_\mathrm{c}$?

A Recall that

- $\overline{X}_\mathrm{t}$ is normal with mean μ_t and standard deviation $\sigma_\mathrm{t}/\sqrt{27}$.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_\text{t}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_\text{c}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_\text{t} - \overline{X}_\text{c}$?

A Recall that

- $\overline{X}_\text{t}$ is normal with mean μ_t and standard deviation $\sigma_\text{t}/\sqrt{27}$.
- $\overline{X}_\text{c}$ is normal with mean μ_c and standard deviation $\sigma_\text{c}/\sqrt{23}$.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\mathrm{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\mathrm{c}}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$?

A Recall that

- $\overline{X}_{\mathrm{t}}$ is normal with mean μ_{t} and standard deviation $\sigma_{\mathrm{t}}/\sqrt{27}$.
- $\overline{X}_{\mathrm{c}}$ is normal with mean μ_{c} and standard deviation $\sigma_{\mathrm{c}}/\sqrt{23}$.
- The sum or difference of two independent normal RVs is again normal.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\mathrm{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\mathrm{c}}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$?

A Recall that

- $\overline{X}_{\mathrm{t}}$ is normal with mean μ_{t} and standard deviation $\sigma_{\mathrm{t}}/\sqrt{27}$.
- $\overline{X}_{\mathrm{c}}$ is normal with mean μ_{c} and standard deviation $\sigma_{\mathrm{c}}/\sqrt{23}$.
- The sum or difference of two independent normal RVs is again normal.

Therefore, $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$ is also normal with

$$\mathbb{E}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] = \qquad \mathbb{SD}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] =$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\mathrm{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\mathrm{c}}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$?

A Recall that

- $\overline{X}_{\mathrm{t}}$ is normal with mean μ_{t} and standard deviation $\sigma_{\mathrm{t}}/\sqrt{27}$.
- $\overline{X}_{\mathrm{c}}$ is normal with mean μ_{c} and standard deviation $\sigma_{\mathrm{c}}/\sqrt{23}$.
- The sum or difference of two independent normal RVs is again normal.

Therefore, $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$ is also normal with

$$\mathbb{E}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] = \mu_{\mathrm{t}} - \mu_{\mathrm{c}} \qquad \mathbb{SD}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] =$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Let $\overline{X}_{\mathrm{t}}$ be the mean blood pressure of the subjects in the treatment group, and $\overline{X}_{\mathrm{c}}$ be the mean blood pressure of the subjects in the control group.

Assuming that the two populations are normal,

Q What is the distribution of $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$?

A Recall that

- $\overline{X}_{\mathrm{t}}$ is normal with mean μ_{t} and standard deviation $\sigma_{\mathrm{t}}/\sqrt{27}$.
- $\overline{X}_{\mathrm{c}}$ is normal with mean μ_{c} and standard deviation $\sigma_{\mathrm{c}}/\sqrt{23}$.
- The sum or difference of two independent normal RVs is again normal.

Therefore, $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$ is also normal with

$$\mathbb{E}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] = \mu_{\mathrm{t}} - \mu_{\mathrm{c}} \qquad \mathbb{SD}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] = \sqrt{\frac{\sigma_{\mathrm{t}}^2}{27} + \frac{\sigma_{\mathrm{c}}^2}{23}}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0:\ \mu_t - \mu_c = 0 \qquad H_1:\ \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_t - \overline{X}_c$ as a test statistic if

I. The two populations are normal, and
II. We know the population standard deviations σ_t and σ_c.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

$\boxed{\text{A}}$ We can use $\overline{X}_t - \overline{X}_c$ as a test statistic if

I. The two populations are normal, and
II. We know the population standard deviations σ_t and σ_c.

The assumption that the two populations are normal is fair.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\mathrm{H}_0\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} = 0 \qquad \mathrm{H}_1\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$ as a test statistic if

I. The two populations are normal, and
II. We know the population standard deviations σ_{t} and σ_{c}.

The assumption that the two populations are normal is fair. Prior knowledge of the population standard deviations is unrealistic.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0\text{: } \mu_t - \mu_c = 0 \qquad H_1\text{: } \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_t - \overline{X}_c$ as a test statistic if

I. The two populations are normal, and
II. We know the population standard deviations σ_t and σ_c.

The assumption that the two populations are normal is fair. Prior knowledge of the population standard deviations is unrealistic.

Nevertheless, for the sake of this example, let us assume that the researchers know (by some miracle) that $\sigma_t = \sigma_c = 7$ mm Hg.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0:\ \mu_t - \mu_c = 0 \qquad H_1:\ \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_t - \overline{X}_c$ as a test statistic if

I. The two populations are normal, and
II. We know the population standard deviations σ_t and σ_c.

Assumption: normal populations with $\sigma_t = \sigma_c = 7\,\text{mm Hg}$.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_t - \overline{X}_c$ as a test statistic if

I. The two populations are normal, and

II. We know the population standard deviations σ_t and σ_c.

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

Q What is the distribution of $\overline{X}_t - \overline{X}_c$ assuming H_0?

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0\text{: } \mu_t - \mu_c = 0 \qquad H_1\text{: } \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_t - \overline{X}_c$ as a test statistic if

I. The two populations are normal, and

II. We know the population standard deviations σ_t and σ_c.

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

Q What is the distribution of $\overline{X}_t - \overline{X}_c$ assuming H_0?

A Normal with parameters

$$\mathbb{E}[\overline{X}_t - \overline{X}_c] =$$

$$\mathbb{SD}[\overline{X}_t - \overline{X}_c] =$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0:\ \mu_t - \mu_c = 0 \qquad H_1:\ \mu_t - \mu_c < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_t - \overline{X}_c$ as a test statistic if

I. The two populations are normal, and

II. We know the population standard deviations σ_t and σ_c.

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

Q What is the distribution of $\overline{X}_t - \overline{X}_c$ assuming H_0?

A Normal with parameters

$$\mathbb{E}[\overline{X}_t - \overline{X}_c] = 0$$

$$\mathbb{SD}[\overline{X}_t - \overline{X}_c] =$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\mathrm{H}_0\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} = 0 \qquad \mathrm{H}_1\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} < 0}$$

Q What is a suitable test statistic?

A We can use $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$ as a test statistic if

I. The two populations are normal, and

II. We know the population standard deviations σ_{t} and σ_{c}.

Assumption: normal populations with $\sigma_{\mathrm{t}} = \sigma_{\mathrm{c}} = 7\,\mathrm{mm\ Hg}$.

Q What is the distribution of $\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}$ assuming H_0?

A Normal with parameters

$$\mathbb{E}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] = 0$$

$$\mathbb{SD}[\overline{X}_{\mathrm{t}} - \overline{X}_{\mathrm{c}}] = \sqrt{\frac{7^2}{27} + \frac{7^2}{23}} \approx \boxed{1.9863\,\mathrm{mm\ Hg}}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\mathrm{H}_0\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} = 0 \qquad \mathrm{H}_1\text{: } \mu_{\mathrm{t}} - \mu_{\mathrm{c}} < 0}$$

Assumption: normal populations with $\sigma_{\mathrm{t}} = \sigma_{\mathrm{c}} = 7\,\mathrm{mm\ Hg}$.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0$$

Assumption: normal populations with $\sigma_t = \sigma_c = 7\,\text{mm Hg}$.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

The above tables show the result of the experiment.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0:\ \mu_t - \mu_c = 0 \qquad H_1:\ \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A The observed value of the test statistic is

$$\bar{x}_t - \bar{x}_c = 134.55 - 138.90 = \boxed{-4.35 \text{ mm Hg}}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A The observed value of the test statistic is

$$\bar{x}_t - \bar{x}_c = 134.55 - 138.90 = \boxed{-4.35 \text{ mm Hg}}$$

Using the applet for the normal distribution,

$$\text{p-value} = \mathbb{P}(\bar{X}_t - \bar{X}_c \leq -4.35 \mid \text{H}_0) \approx$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = 7\,\text{mm Hg}$.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A The observed value of the test statistic is

$$\bar{x}_t - \bar{x}_c = 134.55 - 138.90 = \boxed{-4.35\,\text{mm Hg}}$$

Using the applet for the normal distribution,

$$\text{p-value} = \mathbb{P}(\bar{X}_t - \bar{X}_c \leq -4.35 \mid \text{H}_0) \approx \boxed{0.01426}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A The observed value of the test statistic is

$$\bar{x}_t - \bar{x}_c = 134.55 - 138.90 = \boxed{-4.35 \text{ mm Hg}}$$

Using the applet for the normal distribution,

$$\text{p-value} = \mathbb{P}(\overline{X}_t - \overline{X}_c \leq -4.35 \mid H_0) \approx \boxed{0.01426}$$

Since p-value $\leq \alpha$, the researchers should reject the null hypothesis.

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = 7\,\text{mm Hg}$.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A The observed value of the test statistic is

$$\bar{x}_t - \bar{x}_c = 134.55 - 138.90 = \boxed{-4.35\,\text{mm Hg}}$$

Using the applet for the normal distribution,

$$\text{p-value} = \mathbb{P}(\overline{X}_t - \overline{X}_c \leq -4.35 \mid \text{H}_0) \approx \boxed{0.01426}$$

Since p-value $\leq \alpha$, the researchers should reject the null hypothesis.

Conclusion: The evidence is statistically significant in favor of the effectiveness of the herbal tea.

Comparing two means: indep. samples & σ_1, σ_2 known

Covered scenarios:

I. The samples are large.

II. The populations are (approximately) normal.

Tests about μ_1 and μ_2 (p-value approach)

$$\boxed{\mathrm{H}_0: \ \mu_1 - \mu_2 = a \qquad \mathrm{H}_1: \ \mu_1 - \mu_2 < a}$$

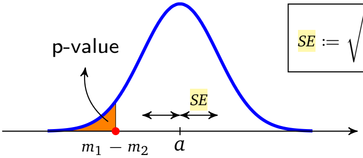
We can use the difference $\overline{X}_1 - \overline{X}_2$ as the test statistic.

Assuming either I or II, at significance level α,

$$\text{p-value} = \mathbb{P}(\overline{X}_1 - \overline{X}_2 \leq m_1 - m_2 \mid \mathrm{H}_0)$$

where m_1 and m_2 are the observed values of the sample means.

We reject H_0 if p-value $\leq \alpha$.



$$\boxed{SE := \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Covered scenarios:

I. The samples are large.

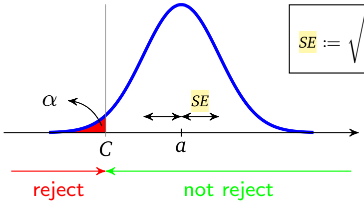
II. The populations are (approximately) normal.

Tests about μ_1 and μ_2 (rejection region approach)

$$\boxed{H_0: \ \mu_1 - \mu_2 = a \qquad H_1: \ \mu_1 - \mu_2 < a}$$

We can use the difference $\overline{X}_1 - \overline{X}_2$ as the test statistic.

Assuming either I or II, at significance level α, we reject H_0 if $\overline{X}_1 - \overline{X}_2 \leq C$, where C is the critical value indicated below.



$$\boxed{SE := \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Covered scenarios:

I. The samples are large.

II. The populations are (approximately) normal.

Confidence interval for $\mu_1 - \mu_2$

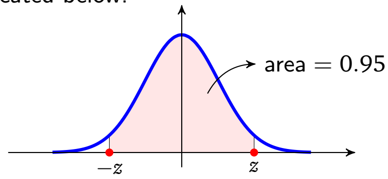
Assuming either I or II, the 95% confidence interval for $\mu_1 - \mu_2$ is

$$(m_1 - m_2) \pm z \cdot SE$$

where m_1 and m_2 are the observed values of the sample means,

$$SE := \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

and z is the value indicated below.



Standard normal distribution

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A Since the two populations are normal,

$$95\%\text{-CI} = (\bar{x}_t - \bar{x}_c) \pm z \cdot SE$$
$$=$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A Since the two populations are normal,

$$
\begin{aligned}
95\%\text{-CI} &= (\bar{x}_t - \bar{x}_c) \pm z \cdot SE \\
&= -4.35 \pm \qquad\qquad \times
\end{aligned}
$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A Since the two populations are normal,

$$\begin{aligned} 95\%\text{-CI} &= (\bar{x}_t - \bar{x}_c) \pm z \cdot SE \\ &= -4.35 \pm 1.95996 \times \end{aligned}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Assumption: normal populations with $\sigma_t = \sigma_c = 7\,\text{mm Hg}$.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A Since the two populations are normal,

$$\begin{aligned} 95\%\text{-CI} &= (\bar{x}_t - \bar{x}_c) \pm z \cdot SE \\ &= -4.35 \pm 1.95996 \times 1.9863 \end{aligned}$$

Comparing two means: indep. samples & σ_1, σ_2 known

Example (Herbal remedy for high blood pressure)

Assumption: normal populations with $\sigma_t = \sigma_c = 7$ mm Hg.

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A Since the two populations are normal,

$$\begin{aligned}
95\%\text{-CI} &= (\bar{x}_t - \bar{x}_c) \pm z \cdot SE \\
&= -4.35 \pm 1.95996 \times 1.9863 \\
&= \boxed{-4.35 \pm 3.89 \text{ mm Hg}}
\end{aligned}$$

Comparing two populations

Inference about population means

We consider the following scenarios:

1. independent samples, σ_1 and σ_2 known,
   (a) normal populations
   (b) large samples
2. independent samples, σ_1 and σ_2 unknown but $\sigma_1 = \sigma_2$,
   (a) normal populations
   (b) large samples
3. independent samples, σ_1 and σ_2 unknown and $\sigma_1 \neq \sigma_2$,
   (a) normal populations
   (b) large samples
4. paired sample,
   (a) normal population of pair differences
   (b) large sample

Inference about population proportions

We consider the following scenario:

1. large independent samples

Comparing two populations

Inference about population means

We consider the following scenarios:

1. independent samples, σ_1 and σ_2 known,
   (a) normal populations
   (b) large samples
2. independent samples, σ_1 and σ_2 unknown but $\sigma_1 = \sigma_2$,
   (a) normal populations
   (b) large samples
3. independent samples, σ_1 and σ_2 unknown and $\sigma_1 \neq \sigma_2$,
   (a) normal populations
   (b) large samples
4. paired sample,
   (a) normal population of pair differences
   (b) large sample

Inference about population proportions

We consider the following scenario:

1. large independent samples

Comparing two means: indep. samples & σ_1, σ_2 unknown

Covered scenarios:

I. The samples are large.

II. The populations are (approximately) normal.

Tests about μ_1 and μ_2

$$\boxed{H_0: \ \mu_1 - \mu_2 = a \qquad H_1: \ \mu_1 - \mu_2 < a}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Covered scenarios:

I. The samples are large.

II. The populations are (approximately) normal.

Tests about μ_1 and μ_2

$$\boxed{H_0: \ \mu_1 - \mu_2 = a \qquad H_1: \ \mu_1 - \mu_2 < a}$$

Choice of statistic

Comparing two means: indep. samples & σ_1, σ_2 unknown

Covered scenarios:

I. The samples are large.

II. The populations are (approximately) normal.

Tests about μ_1 and μ_2

$$\boxed{H_0: \ \mu_1 - \mu_2 = a \qquad H_1: \ \mu_1 - \mu_2 < a}$$

Choice of statistic

When σ_1 and σ_2 are unknown, we cannot use $\overline{X}_1 - \overline{X}_2$ as the test statistic.

Indeed, assuming H_0, the difference $\overline{X}_1 - \overline{X}_2$ is still (approximately) normally distributed, but we do not have access to its standard deviation $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Covered scenarios:

I. The samples are large.

II. The populations are (approximately) normal.

Tests about μ_1 and μ_2

$$\boxed{\text{H}_0\text{: } \mu_1 - \mu_2 = a \qquad \text{H}_1\text{: } \mu_1 - \mu_2 < a}$$

Choice of statistic

When σ_1 and σ_2 are unknown, we cannot use $\overline{X}_1 - \overline{X}_2$ as the test statistic.

Indeed, assuming H$_0$, the difference $\overline{X}_1 - \overline{X}_2$ is still (approximately) normally distributed, but we do not have access to its standard deviation $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

In this case, the appropriate choice of the test statistic depends on whether $\sigma_1 = \sigma_2$ or $\sigma_1 \neq \sigma_2$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 = \sigma_2$)

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 = \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 = \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 = \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.

Choice of statistic (when $\sigma_1 = \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 = \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

If $\sigma_1 = \sigma_2$, then,

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 = \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

If $\sigma_1 = \sigma_2$, then,

$$T := \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has the t distribution with $df = n_1 + n_2 - 2$ degrees of freedom,

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 = \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

If $\sigma_1 = \sigma_2$, then,

$$T := \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has the t distribution with $df = n_1 + n_2 - 2$ degrees of freedom, where

$$S_p := \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

is the pooled standard deviation of the two samples.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

Then,

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

Then,

$$T := \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

has the t distribution with

$$df =$$

degrees of freedom.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Choice of statistic (when $\sigma_1 \neq \sigma_2$)

Suppose

- $\overline{X}_1$ and S_1 are the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1 and standard deviation σ_1.
- $\overline{X}_2$ and S_2 are the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2 and standard deviation σ_2.
- The two samples are independent.
- Either n_1 and n_2 are large, or the populations are normal.

Then,

$$T := \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

has the t distribution with

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 / (n_1 - 1) + \left(\frac{S_2^2}{n_2}\right)^2 / (n_2 - 1)}$$

degrees of freedom.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Covered scenarios:

I. The samples are large.

II. The populations are (approximately) normal.

Tests about μ_1 and μ_2 (p-value approach)

$$\boxed{\mathrm{H}_0\colon\ \mu_1 - \mu_2 = a \qquad \mathrm{H}_1\colon\ \mu_1 - \mu_2 < a}$$

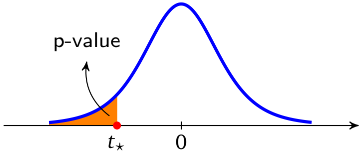
We can use T as the test statistic. [Choose the right version!]

Assuming either I or II, at significance level α,

$$\text{p-value} = \mathbb{P}(T \leq t_\star \mid \mathrm{H}_0)$$

where $t_\star$ is the observed value of the test statistic T.

We reject H_0 if p-value $\leq \alpha$.



t-distribution with df depending on the scenario

Comparing two means: indep. samples & σ_1, σ_2 unknown

Covered scenarios:

I. The samples are large.

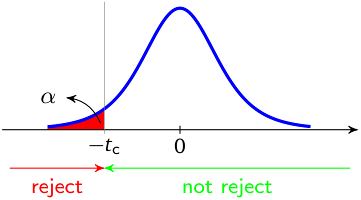
II. The populations are (approximately) normal.

Tests about μ_1 and μ_2 (rejection region approach)

$$\boxed{\mathrm{H}_0\text{: } \mu_1 - \mu_2 = a \qquad \mathrm{H}_1\text{: } \mu_1 - \mu_2 < a}$$

We can use T as the test statistic. [Choose the right version!]

Assuming either I or II, at significance level α, we reject H_0 if $T \leq -t_c$, where $-t_c$ is the critical value indicated below.



t-distribution with df depending on the scenario

Comparing two means: indep. samples & σ_1, σ_2 unknown

Covered scenarios:

I. The samples are large.

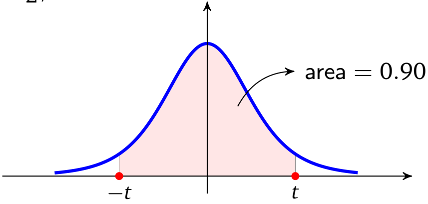
II. The populations are (approximately) normal.

Confidence interval for $\mu_1 - \mu_2$

Assuming either I or II, the 90% confidence interval for $\mu_1 - \mu_2$ is

$$(m_1 - m_2) \pm t \cdot SE$$

where m_1 and m_2 are the observed values of the sample means, $SE = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ or $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ depending on whether $\sigma_1 = \sigma_2$ or $\sigma_1 \neq \sigma_2$, and t is the value indicated below.



t-distribution with df depending on the scenario

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

H$_0$: $\mu_t - \mu_c = 0$ H$_1$: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{H_0\text{: } \mu_t - \mu_c = 0 \qquad H_1\text{: } \mu_t - \mu_c < 0}$

$\boxed{\textit{Assumption}\text{: normal populations with } \sigma_t = \sigma_c = ?}$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is a suitable test statistic?

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

H_0: $\mu_t - \mu_c = 0$ $\quad$ H_1: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is a suitable test statistic?

A Since $\sigma_t = \sigma_c$ is unknown, we can use

$$T := \frac{(\overline{X}_t - \overline{X}_c) - (\mu_t - \mu_c)}{S_p\sqrt{\frac{1}{27} + \frac{1}{23}}}$$

where S_p is the pooled standard deviation of the two samples.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{\mathrm{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \mathrm{H}_1\text{: } \mu_t - \mu_c < 0}$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the distribution of T assuming H_0?

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the distribution of T assuming H_0?

A t distribution with $df = 27 + 23 - 2 = \boxed{48}$ degrees of freedom.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

H_0: $\mu_t - \mu_c = 0$ $\quad$ H_1: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of the pooled standard deviation?

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

H$_0$: $\mu_t - \mu_c = 0$ H$_1$: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of the pooled standard deviation?

A We have

$$s_p = \sqrt{\frac{(27-1){s_t}^2 + (23-1){s_c}^2}{27+23-2}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

H_0: $\mu_t - \mu_c = 0$ $\qquad$ H_1: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of the pooled standard deviation?

A We have

$$s_p = \sqrt{\frac{(27-1){s_t}^2 + (23-1){s_c}^2}{27+23-2}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{H_0}$: $\mu_t - \mu_c = 0$ $\quad$ $\boxed{H_1}$: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of the pooled standard deviation?

A We have

$$s_p = \sqrt{\frac{(27-1)s_t{}^2 + (23-1)s_c{}^2}{27+23-2}}$$

$$= \sqrt{\frac{(27-1) \times 5.54^2 + (23-1) \times 7.82^2}{27+23-2}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

H$_0$: $\mu_t - \mu_c = 0$ H$_1$: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of the pooled standard deviation?

A We have

$$
\begin{aligned}
s_p &= \sqrt{\frac{(27-1)s_t^2 + (23-1)s_c^2}{27+23-2}} \\
&= \sqrt{\frac{(27-1) \times 5.54^2 + (23-1) \times 7.82^2}{27+23-2}} \\
&\approx \boxed{6.682275\,\text{mm Hg}}
\end{aligned}
$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of T?

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0:\ \mu_t - \mu_c = 0 \qquad H_1:\ \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of T?

A The observed value is

$$t = \frac{(\bar{x}_t - \bar{x}_c) - (\mu_t - \mu_c)}{s_p \sqrt{\frac{1}{27} + \frac{1}{23}}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{H_0: \mu_t - \mu_c = 0 \qquad H_1: \mu_t - \mu_c < 0}$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of T?

A The observed value is

$$t = \frac{(\bar{x}_t - \bar{x}_c) - (\mu_t - \mu_c)}{s_p\sqrt{\frac{1}{27} + \frac{1}{23}}}$$

$$= \frac{(134.55 - 138.90) - 0}{6.682275\sqrt{\frac{1}{27} + \frac{1}{23}}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{H_0: \mu_t - \mu_c = 0 \qquad H_1: \mu_t - \mu_c < 0}$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the observed value of T?

A The observed value is

$$t = \frac{(\bar{x}_t - \bar{x}_c) - (\mu_t - \mu_c)}{s_p\sqrt{\frac{1}{27} + \frac{1}{23}}} = \frac{(134.55 - 138.90) - 0}{6.682275\sqrt{\frac{1}{27} + \frac{1}{23}}} \approx \boxed{-2.294171}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A Using the applet for the t distribution,

$$\text{p-value} = \mathbb{P}(T \leq -2.294171 \mid H_0) \approx$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A Using the applet for the t distribution,

$$\text{p-value} = \mathbb{P}(T \leq -2.294171 \mid \text{H}_0) \approx \boxed{0.0131}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_\text{t} - \mu_\text{c} = 0 \qquad \text{H}_1\text{: } \mu_\text{t} - \mu_\text{c} < 0}$$

Assumption: normal populations with $\sigma_\text{t} = \sigma_\text{c} = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A Using the applet for the t distribution,

$$\text{p-value} = \mathbb{P}(T \leq -2.294171 \mid \text{H}_0) \approx \boxed{0.0131}$$

Since p-value $\leq \alpha$, the researchers should reject the null hypothesis.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{H_0: \ \mu_t - \mu_c = 0 \qquad H_1: \ \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q At significance level 5%, how should the researchers conclude?

A Using the applet for the t distribution,

$$\text{p-value} = \mathbb{P}(T \leq -2.294171 \mid H_0) \approx \boxed{0.0131}$$

Since p-value $\leq \alpha$, the researchers should reject the null hypothesis.

Conclusion: The evidence is statistically significant in favor of the effectiveness of the herbal tea.

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{H_0\text{: } \mu_t - \mu_c = 0 \qquad H_1\text{: } \mu_t - \mu_c < 0}$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0}$$

$$\boxed{\textit{Assumption}\text{: normal populations with } \sigma_t = \sigma_c = ?}$$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A We have,

$$95\%\text{-CI} = (\bar{x}_t - \bar{x}_c) \pm t \cdot s_p \sqrt{\frac{1}{27} + \frac{1}{23}}$$

$$=$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$\boxed{\mathrm{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \mathrm{H}_1\text{: } \mu_t - \mu_c < 0}$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A We have,

$$95\%\text{-CI} = (\bar{x}_t - \bar{x}_c) \pm t \cdot s_p \sqrt{\frac{1}{27} + \frac{1}{23}}$$

$$= -4.35 \pm \qquad \times \qquad \times \sqrt{\frac{1}{27} + \frac{1}{23}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$\boxed{\text{H}_0\text{: } \mu_t - \mu_c = 0 \qquad \text{H}_1\text{: } \mu_t - \mu_c < 0}$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A We have,

$$\text{95\%-CI} = (\bar{x}_t - \bar{x}_c) \pm t \cdot s_p \sqrt{\frac{1}{27} + \frac{1}{23}}$$

$$= -4.35 \pm 2.01063 \times \qquad \times \sqrt{\frac{1}{27} + \frac{1}{23}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

$$H_0: \mu_t - \mu_c = 0 \qquad H_1: \mu_t - \mu_c < 0$$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A We have,

$$95\%\text{-CI} = (\bar{x}_t - \bar{x}_c) \pm t \cdot s_p \sqrt{\frac{1}{27} + \frac{1}{23}}$$

$$= -4.35 \pm 2.01063 \times 6.682275 \times \sqrt{\frac{1}{27} + \frac{1}{23}}$$

Comparing two means: indep. samples & σ_1, σ_2 unknown

Example (Herbal remedy for high blood pressure)

H_0: $\mu_t - \mu_c = 0$ H_1: $\mu_t - \mu_c < 0$

Assumption: normal populations with $\sigma_t = \sigma_c = ?$

	1	2	3	...	27	mean	sd
treatment group	135.2	129.7	138.4	...	135.2	134.55	5.54

	1	2	3	...	23	mean	sd
control group	134.8	155	144.5	...	130.1	138.90	7.82

Q What is the 95% confidence interval for $\mu_t - \mu_c$?

A We have,

$$\begin{aligned} 95\%\text{-CI} &= (\bar{x}_t - \bar{x}_c) \pm t \cdot s_p \sqrt{\frac{1}{27} + \frac{1}{23}} \\ &= -4.35 \pm 2.01063 \times 6.682275 \times \sqrt{\frac{1}{27} + \frac{1}{23}} \\ &= \boxed{-4.35 \pm 3.81 \text{ mm Hg}} \end{aligned}$$