Statistics II - Hypothesis testing, tests for the mean

Stochastics

Illés Horváth

2021/11/24
(1) Hypothesis testing in general
(2) Structure of a test
(3) Tests for the mean (z-test, t-test)
We are given the information that a pack of sugar contains 1000 g of sugar. We suspect that the sugar factory is cheating and puts less sugar in each pack.
We are given the information that a pack of sugar contains 1000 g of sugar. We suspect that the sugar factory is cheating and puts less sugar in each pack.

How could we check this? Due to the packaging process, the amount of sugar in each pack is random.
We are given the information that a pack of sugar contains 1000 g of sugar. We suspect that the sugar factory is cheating and puts less sugar in each pack.

How could we check this? Due to the packaging process, the amount of sugar in each pack is random.

We take a sample, and based on the sample, make a decision: we either accept or reject the information given. This is called \textit{hypothesis testing}.
We are given the information that a pack of sugar contains 1000 g of sugar. We suspect that the sugar factory is cheating and puts less sugar in each pack.

How could we check this? Due to the packaging process, the amount of sugar in each pack is random.

We take a sample, and based on the sample, make a decision: we either accept or reject the information given. This is called *hypothesis testing*.

Example. We measure 5 bags, and the amount of sugar in each pack is 986, 992, 1003, 976, 968 g respectively. Do we accept that the weight of sugar in a pack has mean 1000 g?
The general setup is as follows. There is an initial hypothesis (information) which we want to test. This is called the *null hypothesis*, and is denoted by $H_0$. We want to make a binary choice whether we accept $H_0$ or not. In the sugar example,

- $H_0$: the weight of sugar in a pack has mean 1000 g.
The general setup is as follows. There is an initial hypothesis (information) which we want to test. This is called the *null hypothesis*, and is denoted by $H_0$. We want to make a binary choice whether we accept $H_0$ or not. In the sugar example,

- $H_0$: the weight of sugar in a pack has mean 1000 g.

The *alternative hypothesis* is the other option, denoted by $H_1$. It is often the opposite of the null hypothesis, but any hypothesis that is disjoint from the null hypothesis works. In the sugar example, a natural choice for $H_1$ is

- $H_1$: the weight of sugar in a pack has mean not equal to 1000 g.
Hypothesis testing

The general setup is as follows. There is an initial hypothesis (information) which we want to test. This is called the null hypothesis, and is denoted by $H_0$. We want to make a binary choice whether we accept $H_0$ or not. In the sugar example,

- $H_0$: the weight of sugar in a pack has mean 1000 g.

The alternative hypothesis is the other option, denoted by $H_1$. It is often the opposite of the null hypothesis, but any hypothesis that is disjoint from the null hypothesis works. In the sugar example, a natural choice for $H_1$ is

- $H_1$: the weight of sugar in a pack has mean not equal to 1000 g.

Another valid choice for $H_1$ is

- $H_1$: the weight of sugar in a pack has mean less than 1000 g.
Then we take a sample, and based on that sample, do a test. A test in general is a calculation based on $H_0$, $H_1$ and the sample that results in a binary choice: either we

- accept $H_0$, or
- we reject $H_0$ and accept $H_1$ instead.

The possibilities are:

- $H_0$ true
- $H_0$ false

- type I error
- type II error

Ideally, a test would have 0 type I error and 0 type II error, but that's not possible due to the randomness of the sample.
Then we take a sample, and based on that sample, do a test. A test in general is a calculation based on $H_0$, $H_1$ and the sample that results in a binary choice: either we

- accept $H_0$, or
- we reject $H_0$ and accept $H_1$ instead.

A test divides the sample space (all possible samples) into two regions: in one region, we accept $H_0$, and in the other region, we reject $H_0$ in favor of $H_1$. 

The possibilities are:

- $H_0$ true ✓
- $H_0$ false ✓

Ideally, a test would have 0 type I error and 0 type II error, but that's not possible due to the randomness of the sample.
Hypothesis testing

Then we take a sample, and based on that sample, do a test. A test in general is a calculation based on $H_0$, $H_1$ and the sample that results in a binary choice: either we

- accept $H_0$, or
- we reject $H_0$ and accept $H_1$ instead.

A test divides the sample space (all possible samples) into two regions: in one region, we accept $H_0$, and in the other region, we reject $H_0$ in favor of $H_1$.

The possibilities are:

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ true</th>
<th>$H_0$ false</th>
</tr>
</thead>
<tbody>
<tr>
<td>we accept $H_0$</td>
<td>✓</td>
<td>type II error</td>
</tr>
<tr>
<td>we reject $H_0$</td>
<td>type I error</td>
<td>✓</td>
</tr>
</tbody>
</table>

Ideally, a test would have 0 type I error and 0 type II error, but that’s not possible due to the randomness of the sample.
Hypothesis testing

Then we take a sample, and based on that sample, do a test. A test in general is a calculation based on $H_0$, $H_1$ and the sample that results in a binary choice: either we

- accept $H_0$, or
- we reject $H_0$ and accept $H_1$ instead.

A test divides the sample space (all possible samples) into two regions: in one region, we accept $H_0$, and in the other region, we reject $H_0$ in favor of $H_1$.

The possibilities are:

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ true</th>
<th>$H_0$ false</th>
</tr>
</thead>
<tbody>
<tr>
<td>we accept $H_0$</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>we reject $H_0$</td>
<td>type I error</td>
<td>✓</td>
</tr>
</tbody>
</table>

Ideally, a test would have 0 type I error and 0 type II error, but that’s not possible due to the randomness of the sample.
Two trivial tests are as follows.
Two trivial tests are as follows.

If we always accept $H_0$, regardless of the sample, then we never make a type I error.
Structure of a test

Two trivial tests are as follows.

If we always accept $H_0$, regardless of the sample, then we never make a type I error.

On the other hand, if we always reject $H_0$, we never make a type II error.
Two trivial tests are as follows.

If we always accept $H_0$, regardless of the sample, then we never make a type I error.

On the other hand, if we always reject $H_0$, we never make a type II error.

Typical tests are “in between”: they accept $H_0$ for some samples and reject $H_0$ for other samples.
Two trivial tests are as follows.

If we always accept $H_0$, regardless of the sample, then we never make a type I error.

On the other hand, if we always reject $H_0$, we never make a type II error.

Typical tests are “in between”: they accept $H_0$ for some samples and reject $H_0$ for other samples.

In general, a test aims to decide if the sample can be considered typical according to $H_0$. If yes, we will accept $H_0$. If not, we reject $H_0$. 
Two trivial tests are as follows.

If we always accept $H_0$, regardless of the sample, then we never make a type I error.

On the other hand, if we always reject $H_0$, we never make a type II error.

Typical tests are “in between”: they accept $H_0$ for some samples and reject $H_0$ for other samples.

In general, a test aims to decide if the sample can be considered typical according to $H_0$. If yes, we will accept $H_0$. If not, we reject $H_0$.

There is a tradeoff: if the acceptance region is larger, then a type I error is less likely and a type II error is more likely; if the acceptance region is smaller, it is the other way round.
The size of the acceptance region is usually given by the *significance level* (also called the confidence level). The significance level of a test is a number between 0 and 100%; for example, a 95% significance level means that we are going to accept $H_0$ if the sample is among the 95% most typical according to $H_0$. Accordingly, the significance level represents the level of trust we have in $H_0$. It does not describe how good the test is.
The size of the acceptance region is usually given by the *significance level* (also called the confidence level). The significance level of a test is a number between 0 and 100%; for example, a 95% significance level means that we are going to accept $H_0$ if the sample is among the 95% most typical according to $H_0$.

Accordingly, the significance level represents the level of trust we have in $H_0$. It does *not* describe how good the test is.
The size of the acceptance region is usually given by the *significance level* (also called the confidence level). The significance level of a test is a number between 0 and 100%; for example, a 95% significance level means that we are going to accept $H_0$ if the sample is among the 95% most typical according to $H_0$.

Accordingly, the significance level represents the level of trust we have in $H_0$. It does *not* describe how good the test is.

Checking if the sample is typical or not is usually done by computing a statistic, then checking whether the value of the statistic is among the 95% most typical values (assuming $H_0$ holds).
Most tests are structured the following way:

- we compute a *statistic* from the sample,
- we compute a *percentile* (also known as a quantile) based on the significance level and the theoretical limit distribution of the statistic assuming $H_0$ holds,
- the outcome of the test is based on the comparison of the statistic and percentile.

The significance level controls the type I error; if the significance level is $1 - \varepsilon$, then the probability of a type I error is $\varepsilon$. The type II error is not controlled.
Statistical software often execute tests in the following manner: instead of computing the percentile for a given significance level, they compute the \textit{smallest significance level} $p$ at which $H_0$ is still accepted for the given sample. This value is known as the \textit{p-value} of the sample. Then this $p$-value can be compared with the significance level directly.

In general, if the $p$-value is high (close to 1), that means that the sample is not very typical according to $H_0$. What $p$-value is considered still acceptable for $H_0$ depends on the application.
The following test is known as the \textit{one-sample, 2-tail z-test}, and it can be used to test the mean of a sample against a fixed value. It is based on the CLT.
The following test is known as the one-sample, 2-tail z-test, and it can be used to test the mean of a sample against a fixed value. It is based on the CLT.

Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1) = \sigma$ known and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m \neq \mu$. 

Tests for the mean — one-sample, 2-tail $z$-test

The following test is known as the *one-sample, 2-tail $z$-test*, and it can be used to test the mean of a sample against a fixed value. It is based on the CLT.

Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1) = \sigma$ known and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m \neq \mu$.

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

- compute the *statistic* $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$ from the sample;
- compute the *percentile* $z_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2)$ from the table for the standard normal distribution and the significance level, and
- if $z \in [-z_{\varepsilon/2}, z_{\varepsilon/2}]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$. 
If \( z \) is close enough to 0, then the difference between \( z \) and \( z_{\epsilon/2} \) might be due to randomness, so we accept \( H_0 \).
If $z$ is close enough to 0, then the difference between $z$ and $z_{\epsilon/2}$ might be due to randomness, so we accept $H_0$.

If $z$ is too far away from 0, then we do not accept that the difference is due to randomness, so we reject $H_0$. 
Example. Back to the sugar example. The sample is

\[ X_1 = 986, \quad X_2 = 992, \quad X_3 = 1003, \quad X_4 = 976, \quad X_5 = 968. \]

Let’s assume \( \sigma = 20 \) is known from the packaging technology.

- \( H_0: \ m = 1000; \)
- \( H_1: \ m \neq 1000. \)

Test \( H_0 \) against \( H_1 \) on a 95% significance level.
The sugar example

First we compute the statistic:

\[
z = \frac{\bar{x} - \mu}{\sigma\sqrt{n}} = \frac{985 - 1000}{10} \sqrt{5} = -1.677.
\]
The sugar example

First we compute the statistic:

\[ z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{985 - 1000}{10} \sqrt{5} = -1.677. \]

Then we compute the percentile:

\[ z_{\epsilon/2} = \Phi^{-1}(0.975) = 1.96. \]
The sugar example

First we compute the statistic:

\[ z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{985 - 1000}{10} \sqrt{5} = -1.677. \]

Then we compute the percentile:

\[ z_{\frac{\epsilon}{2}} = \Phi^{-1}(0.975) = 1.96. \]

Finally, we check \( z \in [-z_{\frac{\epsilon}{2}}, z_{\frac{\epsilon}{2}}] \).

\[ -1.677 \in [-1.96, 1.96] \]

holds, so we accept \( H_0 \) on a 95% significance level.
Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1) = \sigma$ known and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m < \mu$. 

To test $H_0$ against $H_1$ on a $1 - \epsilon$ significance level, we do the following:

1. Compute the statistic $z = \bar{x} - \mu / \sigma \sqrt{n}$ from the sample;
2. Compute the percentile $z_{\epsilon} = \Phi^{-1}(1 - \epsilon)$ from the table for the standard normal distribution and the significance level, and if $z \in \left[ -z_{\epsilon}, \infty \right)$ holds, we accept $H_0$; if it does not hold, we reject $H_0$ in favor of $H_1$. 

Testing for the mean – one-sample, 1-tail z-test
Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1) = \sigma$ known and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m < \mu$.

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

- Compute the statistic $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$ from the sample;
- Compute the percentile $z_\varepsilon = \Phi^{-1}(1 - \varepsilon)$ from the table for the standard normal distribution and the significance level, and
- If $z \in [-z_\varepsilon, \infty)$ holds, we accept $H_0$; if it does not hold, we reject $H_0$ in favor of $H_1$. 

Tests for the mean – one-sample, 1-tail z-test
If $z$ is too far away from 0 to the left, then we do not accept that the difference is due to randomness, so we reject $H_0$ in favor of $H_1$. (On the figure, the entire $\varepsilon$ probability is put to the left tail.)
Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1) = \sigma$ known and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m > \mu$. 

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

1. Compute the statistic $z = \bar{x} - \mu / \sigma \sqrt{n}$ from the sample;
2. Compute the percentile $z_\varepsilon = \Phi^{-1}(1 - \varepsilon)$ from the table for the standard normal distribution and the significance level, and if $z \in (-\infty, z_\varepsilon]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$ in favor of $H_1$. 
Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1) = \sigma$ known and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m > \mu$.

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

- compute the statistic $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$ from the sample;
- compute the percentile $z_\varepsilon = \Phi^{-1}(1 - \varepsilon)$ from the table for the standard normal distribution and the significance level, and
- if $z \in (-\infty, z_\varepsilon]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$ in favor of $H_1$. 
If $z$ is too far away from 0 to the right, then we do not accept that the difference is due to randomness, so we reject $H_0$ in favor of $H_1$. 
The z-tests are all based on the CLT, which guarantees that the statistic 
\[ z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} \] is close to \( N(0,1) \) if \( H_0 \) holds.
The z-tests are all based on the CLT, which guarantees that the statistic $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$ is close to $N(0,1)$ if $H_0$ holds.

Most tests in general have a limit theorem in the background. It might be different from the CLT, and the limit distribution may also be different from $N(0,1)$, in which case we need to take the percentile from another distribution.
The $z$-tests are all based on the CLT, which guarantees that the statistic $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$ is close to $\text{N}(0,1)$ if $H_0$ holds.

Most tests in general have a limit theorem in the background. It might be different from the CLT, and the limit distribution may also be different from $\text{N}(0,1)$, in which case we need to take the percentile from another distribution.

That said, most tests have the same structure:

- compute a statistic from the sample;
- compute a percentile from the table of the relevant distribution and the significance level, and
- compare the statistic and the percentile to accept or reject $H_0$. 
The $z$-tests are all based on the CLT, which guarantees that the statistic $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$ is close to $N(0,1)$ if $H_0$ holds.

Most tests in general have a limit theorem in the background. It might be different from the CLT, and the limit distribution may also be different from $N(0,1)$, in which case we need to take the percentile from another distribution.

That said, most tests have the same structure:

- compute a statistic from the sample;
- compute a percentile from the table of the relevant distribution and the significance level, and
- compare the statistic and the percentile to accept or reject $H_0$.

In general, we will focus on this structure and not the limit theorem behind the test.
In the two-sample, 2-tail $z$-test, we have 2 separate samples:

- $X_1, \ldots, X_n$ has known deviation $\sigma_1$ and unknown mean $m_1$;
- $Y_1, \ldots, Y_m$ has known deviation $\sigma_2$ and unknown mean $m_2$. 
In the two-sample, 2-tail $z$-test, we have 2 separate samples:
- $X_1, \ldots, X_n$ has known deviation $\sigma_1$ and unknown mean $m_1$;
- $Y_1, \ldots, Y_m$ has known deviation $\sigma_2$ and unknown mean $m_2$.

The test aims to compare the means of the two samples:
- $H_0$: $m_1 = m_2$;
- $H_1$: $m_1 \neq m_2$;
In the two-sample, 2-tail $z$-test, we have 2 separate samples:
- $X_1, \ldots, X_n$ has known deviation $\sigma_1$ and unknown mean $m_1$;
- $Y_1, \ldots, Y_m$ has known deviation $\sigma_2$ and unknown mean $m_2$.

The test aims to compare the means of the two samples:
- $H_0$: $m_1 = m_2$;
- $H_1$: $m_1 \neq m_2$;

The test itself is the following:
- from the sample, compute the statistic
  \[
  z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}};
  \]
- from the table for the standard normal distribution and the significance level $1 - \varepsilon$, compute the percentile
  $z_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2)$;
- if $z \in [-z_{\varepsilon/2}, z_{\varepsilon/2}]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$. 

Stochastics  Illés Horváth  Statistics II - Hypothesis testing, tests for the mean
In the two-sample, 1-tail $z$-test, we have 2 separate samples:

- $X_1, \ldots, X_n$ has known deviation $\sigma_1$ and unknown mean $m_1$;
- $Y_1, \ldots, Y_m$ has known deviation $\sigma_2$ and unknown mean $m_2$. 
In the two-sample, 1-tail z-test, we have 2 separate samples:

- $X_1, \ldots, X_n$ has known deviation $\sigma_1$ and unknown mean $m_1$;
- $Y_1, \ldots, Y_m$ has known deviation $\sigma_2$ and unknown mean $m_2$.

The test aims to compare the means of the two samples:

- $H_0$: $m_1 = m_2$;
- $H_1$: $m_1 > m_2$. 
Tests for the mean – two-sample, 1-tail $z$-test

In the two-sample, 1-tail $z$-test, we have 2 separate samples:
- $X_1, \ldots, X_n$ has known deviation $\sigma_1$ and unknown mean $m_1$;
- $Y_1, \ldots, Y_m$ has known deviation $\sigma_2$ and unknown mean $m_2$.

The test aims to compare the means of the two samples:
- $H_0$: $m_1 = m_2$;
- $H_1$: $m_1 > m_2$;

The test itself is the following:
- from the sample, compute the statistic
  \[ z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; \]
- from the table for the standard normal distribution and the significance level $1 - \varepsilon$, compute the percentile
  \[ z_\varepsilon = \Phi^{-1}(1 - \varepsilon); \]
- if $z \in (-\infty, z_\varepsilon]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$ in favor of $H_1$. 
The $z$-test always assumes that the deviation is known. What can we do when the deviation is not known?

A natural idea is to use the corrected sample deviation $s^*$ instead of $\sigma$. The result is the so-called $t$-test.

The main difference to the $z$-test is that when $n$ is small, $t = \bar{x} - \mu \frac{s^*}{\sqrt{n}}$ has Student $t$-distribution or simply $t$-distribution instead of being close to $N(0,1)$. Accordingly, the percentile will come from the $t$-distribution.

There is a separate $t$-distribution for each $n$, so we also have to keep track of the degree of freedom: if the sample size is $n$, then we need to take the percentile from the $t$-distribution with degree of freedom $n - 1$.

As $n \to \infty$, the $t$-distribution converges to $N(0,1)$, so for large values of $n$, the $z$-test and $t$-test are almost identical.
Tests for the mean – $t$-test

The $z$-test always assumes that the deviation is known. What can we do when the deviation is not known?

A natural idea is to use the corrected sample deviation $s_n^*$ instead of $\sigma$. The result is the so-called $t$-test.
Tests for the mean – t-test

The z-test always assumes that the deviation is known. What can we do when the deviation is not known?

A natural idea is to use the corrected sample deviation $s_n^*$ instead of $\sigma$. The result is the so-called t-test.

The main difference to the z-test is that when $n$ is small, $t = \frac{\bar{x} - \mu}{s_n^*} \sqrt{n}$ has Student $t$-distribution or simply $t$-distribution instead of being close to $N(0,1)$. Accordingly, the percentile will come from the $t$-distribution.
Tests for the mean – $t$-test

The $z$-test always assumes that the deviation is known. What can we do when the deviation is not known?

A natural idea is to use the corrected sample deviation $s_n^*$ instead of $\sigma$. The result is the so-called $t$-test.

The main difference to the $z$-test is that when $n$ is small, $t = \frac{\bar{x} - \mu}{s_n^*} \sqrt{n}$ has Student $t$-distribution or simply $t$-distribution instead of being close to $N(0,1)$. Accordingly, the percentile will come from the $t$-distribution.

There is a separate $t$-distribution for each $n$, so we also have to keep track of the degree of freedom: if the sample size is $n$, then we need to take the percentile from the $t$-distribution with degree of freedom $n - 1$. 

Tests for the mean – $t$-test

The $z$-test always assumes that the deviation is known. What can we do when the deviation is not known?

A natural idea is to use the corrected sample deviation $s^*_n$ instead of $\sigma$. The result is the so-called $t$-test.

The main difference to the $z$-test is that when $n$ is small,

$$t = \frac{\bar{x} - \mu}{s^*_n} \sqrt{n}$$

has Student $t$-distribution or simply $t$-distribution instead of being close to $N(0,1)$. Accordingly, the percentile will come from the $t$-distribution.

There is a separate $t$-distribution for each $n$, so we also have to keep track of the degree of freedom: if the sample size is $n$, then we need to take the percentile from the $t$-distribution with degree of freedom $n - 1$.

As $n \to \infty$, the $t$-distribution converges to $N(0,1)$, so for large values of $n$, the $z$-test and $t$-test are almost identical.
Comparison of the pdf of $N(0,1)$ and the pdf of the $t$-distribution with degree of freedom 4:
Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1)$ unknown and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m \neq \mu$. 

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

1. Compute the statistic $t = \bar{x} - \mu \frac{s}{\sqrt{n}}$ from the sample;
2. Take the percentile $t_{\varepsilon/2}$ from the table for the $t$-distribution with degree of freedom $n - 1$ and significance level $1 - \varepsilon$;
3. If $t \in [-t_{\varepsilon/2}, t_{\varepsilon/2}]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$. 

Stochastics Illés Horváth

Statistics II - Hypothesis testing, tests for the mean
Tests for the mean – one-sample, 2-tail $t$-test

Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1)$ unknown and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m \neq \mu$.

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

- compute the statistic $t = \frac{\bar{x} - \mu}{s_n^* \sqrt{n}}$ from the sample;
- take the percentile $t_{\varepsilon/2}$ from the table for the $t$-distribution with degree of freedom $n - 1$ and significance level $1 - \varepsilon$, and
- if $t \in [-t_{\varepsilon/2}, t_{\varepsilon/2}]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$. 
Tests for the mean – one-sample, 1-tail $t$-test

Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1)$ unknown and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m > \mu$. 

For the formulas of the two-sample $t$-tests, we refer to the table of hypothesis testing.
Tests for the mean – one-sample, 1-tail $t$-test

Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1)$ unknown and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m > \mu$.

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

- compute the statistic $t = \frac{\bar{x} - \mu}{s_n^*} \sqrt{n}$ from the sample;
- take the percentile $t_\varepsilon$ from the table for the $t$-distribution with degree of freedom $n - 1$ and significance level $1 - \varepsilon$, and
- if $t \in (-\infty, t_\varepsilon/2]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$. 

For the formulas of the two-sample $t$-tests, we refer to the table of hypothesis testing.
Assume we have an iid sample $X_1, \ldots, X_n$ with $\mathbb{D}(X_1)$ unknown and $\mathbb{E}(X_1) = m$ unknown. Let

- $H_0$: $m = \mu$, where $\mu$ is a known constant;
- $H_1$: $m > \mu$.

To test $H_0$ against $H_1$ on a $1 - \varepsilon$ significance level, we do the following:

- compute the statistic $t = \frac{\bar{x} - \mu}{s_n^*} \sqrt{n}$ from the sample;
- take the percentile $t_\varepsilon$ from the table for the $t$-distribution with degree of freedom $n - 1$ and significance level $1 - \varepsilon$, and
- if $t \in (-\infty, t_\varepsilon/2]$ holds, we accept $H_0$; if it does not hold, we reject $H_0$.

For the formulas of the two-sample $t$-tests, we refer to the table of hypothesis testing.
A company produces cement in packs with 25kg nominal weight. Due to the packaging process, the amount of cement in a single pack has deviation 0.5kg, but the expectation $\mu$ is unknown. We examine 25 packs, and the mean of the cement inside turns out to be 24.82kg.

(a) Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 95%?

(b) Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 90%?

(c) Assume the deviation for each pack is only 0.3kg. Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 95%?
Solution.

(a) Do we accept the hypothesis \( H_0 \) that \( \mu = 25 \) against the hypothesis \( H_1 \) that \( \mu \neq 25 \) with a confidence level 95%?
Solution.

(a) Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 95%?

What test to apply?

- $\sigma$ is known, so we apply a $z$-test;
- we need to test the mean of one sample against a fixed value, so it’s a one-sample $z$-test;
- $H_1$ is $\mu \neq 1000$, so it’s a one-sample 2-tail $z$-test.
Solution.

(a) Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 95%?

What test to apply?

- $\sigma$ is known, so we apply a $z$-test;
- we need to test the mean of one sample against a fixed value, so it’s a one-sample $z$-test;
- $H_1$ is $\mu \neq 1000$, so it’s a one-sample 2-tail $z$-test.

The null hypothesis and alternative hypothesis are

- $H_0$: $\mu = 25$,
- $H_1$: $\mu \neq 25$. 
(a) We compute the statistic:

\[ z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{24.82 - 25}{0.5} \sqrt{25} = -1.8. \]
(a) We compute the statistic:

\[ z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{24.82 - 25}{0.5} \sqrt{25} = -1.8. \]

We compute the percentile:

\[ z_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2) = \Phi^{-1}(0.975) = 1.96. \]
(a) We compute the statistic:

\[ z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{24.82 - 25}{0.5} \sqrt{25} = -1.8. \]

We compute the percentile:

\[ z_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2) = \Phi^{-1}(0.975) = 1.96. \]

Then we do the comparison:

\[ z = -1.8 \in [-z_{\varepsilon/2}, z_{\varepsilon/2}] = [-1.96, 1.96] \]

holds, so we accept \( H_0 \) on a 95% significance level.
(b) Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 90%?
(b) Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 90%?

The difference here is that the percentile corresponding to a 90% significance level is

$$z_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2) = \Phi^{-1}(0.95) = 1.65,$$

and now

$$z = -1.8 \in [-z_{\varepsilon/2}, z_{\varepsilon/2}] = [-1.65, 1.65]$$

does not hold anymore, so on a 90% significance level, $H_0$ is rejected.
(c) Assume the deviation for each pack is only 0.3kg. Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 95%?

If $\sigma = 0.3$, then the statistic is now $z = \bar{x} - \mu / \sigma \sqrt{n} = 24.82 - 25 / 0.3 \sqrt{25} = -3$, and $z = -3 \notin [-z_{\alpha/2}, z_{\alpha/2}] = [-1.96, 1.96]$, does not hold, so we reject $H_0$ on a 95% significance level.
(c) Assume the deviation for each pack is only 0.3kg. Do we accept the hypothesis $H_0$ that $\mu = 25$ against the hypothesis $H_1$ that $\mu \neq 25$ with a confidence level 95%?

If $\sigma = 0.3$, then the statistic is now

\[ z = \frac{\bar{x} - \mu}{\sigma \sqrt{n}} = \frac{24.82 - 25}{0.5} \sqrt{25} = -3, \]

and

\[ z = -3 \in [-z_{\alpha/2}, z_{\alpha/2}] = [-1.96, 1.96] \]

does not hold, so we reject $H_0$ on a 95% significance level.
We measure the concentration of salt in a dilution. We obtain the following sample after 5 measurements: (g/l): 7.7, 8.1, 7.7, 7.5, 7.0. Previously, someone stated that the concentration is 7.2 g/l. Do we accept this on a 95% confidence level against the hypothesis that the concentration is not equal to 7.2 g/l? And what about the following sample: 7.5, 7.4, 7.3, 7.4, 7.5?
We measure the concentration of salt in a dilution. We obtain the following sample after 5 measurements: (g/l): 7.7, 8.1, 7.7, 7.5, 7.0. Previously, someone stated that the concentration is 7.2 g/l. Do we accept this on a 95% confidence level against the hypothesis that the concentration is not equal to 7.2 g/l? And what about the following sample: 7.5, 7.4, 7.3, 7.4, 7.5?

Solution. $\sigma$ is unknown, so it’s a $t$-test; $H_1$ says $c \neq 7$, so it’s a one-sample, 2-tail $t$-test. The concentration is denoted by $c$.

- $H_0$: $c = 7$;
- $H_1$: $c \neq 7$. 
Problem 4

The sample mean is

\[ \bar{x} = \frac{7.7 + 8.1 + 7.7 + 7.5 + 7.0}{5} = 7.6, \]

and

\[ (s_n^*)^2 = \frac{1}{5 - 1} \left( (7.7 - 7.6)^2 + (8.1 - 7.6)^2 + (7.7 - 7.6)^2 + (7.7 - 7.5)^2 + (7.0 - 7.6)^2 \right) = 0.16, \]
Problem 4

The sample mean is

\[ \bar{x} = \frac{7.7 + 8.1 + 7.7 + 7.5 + 7.0}{5} = 7.6, \]

and

\[ (s_n^*)^2 = \frac{1}{5 - 1} \left( (7.7 - 7.6)^2 + (8.1 - 7.6)^2 + (7.7 - 7.6)^2 + \\ (7.7 - 7.5)^2 + (7.0 - 7.6)^2 \right) = 0.16, \]

so

\[ s_n^* = 0.4, \]

and the statistic is

\[ t = \frac{\bar{x} - \mu}{s_n^* \sqrt{n}} = \frac{7.6 - 7.2}{0.4 \sqrt{5}} = 2.236. \]
The percentile is the 95% 2-tail quantile of the $t$-distribution with degree of freedom $n - 1 = 4$:

$$t_{\epsilon/2} = 2.776.$$
The percentile is the 95% 2-tail quantile of the $t$-distribution with degree of freedom $n - 1 = 4$:

$$t_{ε/2} = 2.776.$$

Then the comparison

$$t = 2.236 \in [-t_{ε/2}, t_{ε/2}] = [-2.776, 2.776]$$

holds, so we accept $H_0$ on a 95% significance level.
For the second sample 7.5, 7.4, 7.3, 7.4, 7.5,

$$\bar{x} = 7.42, \quad s_n^* = 0.0837,$$
For the second sample 7.5, 7.4, 7.3, 7.4, 7.5,

\[ \bar{x} = 7.42, \quad s_n^* = 0.0837, \]

and the statistic is

\[ t = \frac{\bar{x} - \mu}{s_n^*} \sqrt{n} = \frac{7.42 - 7.2}{0.0837} \sqrt{5} = 5.880, \]
For the second sample 7.5, 7.4, 7.3, 7.4, 7.5,

\[ \bar{x} = 7.42, \quad s_n^* = 0.0837, \]

and the statistic is

\[ t = \frac{\bar{x} - \mu}{s_n^* \sqrt{n}} = \frac{7.42 - 7.2}{0.0837 \sqrt{5}} = 5.880, \]

and the comparison

\[ t = 5.880 \in \left[ -t_{\varepsilon/2}, t_{\varepsilon/2} \right] = [-2.776, 2.776] \]

does not hold, so based on the second sample, we reject \( H_0 \) on a 95% significance level.
Problem 5

A company wants to motivate its employees to increase productivity. The company tests two different methods: method A is to increase the salary of people, and method B is to improve the work environment. The change in productivity was measured for all 6 employees with both methods:

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>salary incr.</td>
<td>-0.2</td>
<td>0.3</td>
<td>3.6</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

(a) Test on a 95% confidence level whether improving the work environment increases productivity or not. (What is the null hypothesis?)

(b) Test on a 95% confidence level whether increasing the salary increases productivity or not.

(c) Test on a 95% confidence level whether increasing the salary increases productivity more than improving the work environment.
A company wants to motivate its employees to increase productivity. The company tests two different methods: method A is to increase the salary of people, and method B is to improve the work environment. The change in productivity was measured for all 6 employees with both methods:

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>salary incr.</td>
<td>-0.2</td>
<td>0.3</td>
<td>3.6</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

(a) Test on a 95% confidence level whether improving the work environment increases productivity or not. (What is the null hypothesis?)

(b) Test on a 95% confidence level whether increasing the salary increases productivity or not.

(c) Test on a 95% confidence level whether increasing the salary increases productivity more than improving the work environment.
Solution.

(a) Test on a 95% confidence level whether improving the work environment increases productivity or not. (What is the null hypothesis?)
Solution.

(a) Test on a 95% confidence level whether improving the work environment increases productivity or not. (What is the null hypothesis?)

We do a one-sample, 1-tail $t$-test for the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Problem 5

Solution.

(a) Test on a 95% confidence level whether improving the work environment increases productivity or not. (What is the null hypothesis?)

We do a one-sample, 1-tail $t$-test for the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The mean $m$ is unknown; we want to test $m = 0$ against $m > 0$. $H_0$ always contains equality and $H_1$ contains inequality:

- $H_0$: $m = 0$;
- $H_1$: $m > 0$. 

\[ \bar{x} = 0.9, \quad s^* = 0.2, \quad n = 6, \quad \text{so the statistic is} \]

\[ t = \frac{\bar{x} - \mu}{s^* \sqrt{n}} = \frac{0.9 - 0}{0.2 \sqrt{6}} = 1.06. \]
Problem 5

Solution.

(a) Test on a 95% confidence level whether improving the work environment increases productivity or not. (What is the null hypothesis?)

We do a one-sample, 1-tail $t$-test for the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The mean $m$ is unknown; we want to test $m = 0$ against $m > 0$. $H_0$ always contains equality and $H_1$ contains inequality:

- $H_0$: $m = 0$;
- $H_1$: $m > 0$.  

$\bar{x} = 0.9$, $s^*_n = 0.2$,

so the statistic is

$$t = \frac{\bar{x} - \mu}{s^*_n} \sqrt{n} = \frac{0.9 - 0}{0.2} \sqrt{6} = 10.06.$$
(a) The percentile is the 95% 1-tail quantile of the $t$-distribution with degree of freedom $n - 1 = 5$: 

$$t_\varepsilon = 2.015.$$
(a) The percentile is the 95% 1-tail quantile of the $t$-distribution with degree of freedom $n - 1 = 5$:

$$t_ε = 2.015.$$ 

The comparison

$$t = 10.06 \in (-\infty, t_ε] = (-\infty, 2.015)$$

does not hold, so we reject $H_0$ in favor of $H_1$ on a 95% significance level; that is, we conclude that improving the work environment increases productivity.
(a) The percentile is the 95% 1-tail quantile of the $t$-distribution with degree of freedom $n - 1 = 5$:

$$t_{\varepsilon} = 2.015.$$  

The comparison

$$t = 10.06 \in (-\infty, t_{\varepsilon}] = (-\infty, 2.015)$$

does not hold, so we reject $H_0$ in favor of $H_1$ on a 95% significance level; that is, we conclude that improving the work environment increases productivity.

In general, rejecting $H_0$ on a high significance level is a strong statement.
(b) Test on a 95% confidence level whether increasing the salary increases productivity or not.
(b) Test on a 95% confidence level whether increasing the salary increases productivity or not.

We do a one-sample, 1-tail $t$-test for the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>salary incr.</td>
<td>-0.2</td>
<td>0.3</td>
<td>3.6</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Once again, $H_0: \mu = 0$; $H_1: \mu > 0$. Now $\bar{x} = 1.1$, $s^* \sqrt{n} = 1.439$, so the statistic is $t = \frac{\bar{x} - \mu}{s^* \sqrt{n}} = \frac{1.1 - 0}{1.439 \sqrt{6}} = 1.872$. 
(b) Test on a 95% confidence level whether increasing the salary increases productivity or not.

We do a one-sample, 1-tail $t$-test for the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>salary incr.</td>
<td>-0.2</td>
<td>0.3</td>
<td>3.6</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Once again,

- $H_0$: $m = 0$;
- $H_1$: $m > 0$.

Now

$$\bar{x} = 1.1, \quad s^*_n = 1.439,$$

so the statistic is

$$t = \frac{\bar{x} - \mu}{s^*_n \sqrt{n}} = \frac{1.1 - 0}{1.439 \sqrt{6}} = 1.872.$$
(b) The comparison

\[ t = 1.872 \in (-\infty, t_\alpha] = (-\infty, 2.015) \]

now holds, so we accept \( H_0 \) on a 95% significance level, and conclude that increasing the salary does not increase productivity significantly.
(b) The comparison

\[ t = 1.872 \in (-\infty, t_{\alpha}] = (-\infty, 2.015) \]

now holds, so we accept \( H_0 \) on a 95% significance level, and conclude that increasing the salary does not increase productivity significantly.

Note that we reached this conclusion despite the higher \( \bar{x} \) (0.9 for the first sample and 1.1 for the second sample); this is essentially due to the much higher \( s_n^* \) (0.2 for the first sample and 1.439 for the second sample).
(b) The comparison

\[ t = 1.872 \in (-\infty, t_{\epsilon}] = (-\infty, 2.015) \]

now holds, so we accept \( H_0 \) on a 95% significance level, and conclude that increasing the salary does not increase productivity significantly.

Note that we reached this conclusion despite the higher \( \bar{x} \) (0.9 for the first sample and 1.1 for the second sample); this is essentially due to the much higher \( s_n^* \) (0.2 for the first sample and 1.439 for the second sample).

The \( t \)-test tests how big the difference \( \bar{x} - \mu \) is \( \textit{relative to the sample variance} \); for a larger \( s_n^* \), we might accept \( H_0 \) even for a larger average difference, as it might be due to randomness of the sample.
(c) Test on a 95% confidence level whether increasing the salary increases productivity more than improving the work environment.
(c) Test on a 95% confidence level whether increasing the salary increases productivity more than improving the work environment.

Now we have the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>salary incr.</td>
<td>-0.2</td>
<td>0.3</td>
<td>3.6</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

What kind of test do we do?
(c) Test on a 95% confidence level whether increasing the salary increases productivity more than improving the work environment.

Now we have the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>salary incr.</td>
<td>-0.2</td>
<td>0.3</td>
<td>3.6</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

What kind of test do we do?

It would be natural to do a two-sample $t$-test, but it is better to do a one-sample $t$-test for the difference of the two samples.
(c) Test on a 95% confidence level whether increasing the salary increases productivity more than improving the work environment.

Now we have the sample

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>work env. impr.</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>salary incr.</td>
<td>-0.2</td>
<td>0.3</td>
<td>3.6</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

What kind of test do we do?

It would be natural to do a two-sample $t$-test, but it is better to do a one-sample $t$-test for the difference of the two samples.

The reason is that the two samples are not coming from two entirely different sources, as they were conducted on the same set of employees.
(c) This brings in extra randomness due to the employees; however, we are not interested in the employees, we only want to compare the two methods. Doing a one-sample $t$-test for the difference of the two samples cancels out the extra randomness due to the employees.
(c) This brings in extra randomness due to the employees; however, we are not interested in the employees, we only want to compare the two methods. Doing a one-sample $t$-test for the difference of the two samples cancels out the extra randomness due to the employees.

A two-sample test would be justified in the case when the two methods are tested on two different groups of employees.
(c) This brings in extra randomness due to the employees; however, we are not interested in the employees, we only want to compare the two methods. Doing a one-sample $t$-test for the difference of the two samples cancels out the extra randomness due to the employees.

A two-sample test would be justified in the case when the two methods are tested on two different groups of employees.

The difference sample is

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A - B$</td>
<td>1.4</td>
<td>0.7</td>
<td>-2.8</td>
<td>-0.8</td>
<td>1.0</td>
<td>-0.7</td>
</tr>
</tbody>
</table>
(c) This brings in extra randomness due to the employees; however, we are not interested in the employees, we only want to compare the two methods. Doing a one-sample \( t \)-test for the difference of the two samples cancels out the extra randomness due to the employees.

A two-sample test would be justified in the case when the two methods are tested on two different groups of employees.

The difference sample is

<table>
<thead>
<tr>
<th>employee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A - B )</td>
<td>1.4</td>
<td>0.7</td>
<td>-2.8</td>
<td>-0.8</td>
<td>1.0</td>
<td>-0.7</td>
</tr>
</tbody>
</table>

- \( H_0: m = 0; \)
- \( H_1: m < 0 \) (as we want to test whether the salary increases productivity more than the work environment improvement).
(c) \n\n\bar{x} = -0.2, \quad s^*_n = 1.538, \quad \text{and the statistic is} \n\n\begin{align*}
t &= \frac{\bar{x} - \mu}{s^*_n \sqrt{n}} = \frac{-0.2 - 0}{1.538 \sqrt{6}} = -0.318.
\end{align*}
Problem 5

(c)

\[ \bar{x} = -0.2, \quad s_n^* = 1.538, \]

and the statistic is

\[ t = \frac{\bar{x} - \mu}{s_n^*} \sqrt{n} = \frac{-0.2 - 0}{1.538} \sqrt{6} = -0.318. \]

The percentile is the 95% 1-tail quantile of the \( t \)-distribution with degree of freedom \( n - 1 = 5 \):

\[ t_{\alpha} = 2.015. \]
(c) \[
\bar{x} = -0.2, \quad s_n^* = 1.538,
\]
and the statistic is
\[
t = \frac{\bar{x} - \mu}{s_n^* \sqrt{n}} = \frac{-0.2 - 0}{1.538 \sqrt{6}} = -0.318.
\]

The percentile is the 95% 1-tail quantile of the $t$-distribution with degree of freedom $n - 1 = 5$:
\[
t_{\varepsilon} = 2.015.
\]

The comparison
\[
t = -0.318 \in [-t_{\varepsilon}, \infty) = [-2.015, \infty)
\]
holds, so we accept $H_0$ on a 95% significance level, and conclude that increasing the salary does not increase productivity more than improving the work environment.