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)

Motivation

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.

Motivation

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.

Motivation

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*.

Motivation

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?

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.

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.

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.

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.

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.

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:

	H_0 true	H_0 false
we accept H_0	✓	type II error
we reject H_0	type I error	✓

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
we accept H_0	✓	type II error
we reject H_0	type I 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.

Structure of a test

Two trivial tests are as follows.

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.

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.

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.

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

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.

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.

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.

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.

Structure of a test

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.

Structure of a test

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.

Structure of a test

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).

Structure of a test

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 ε. The type II error is not controlled.

Structure of a test

Statistical software often execute tests in the following manner: instead of computing the percentile for a given significance level, they compute the *smallest significance level* p at which H_0 is still accepted for the given sample. This value is known as the *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.

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.

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 μ 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 μ 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:

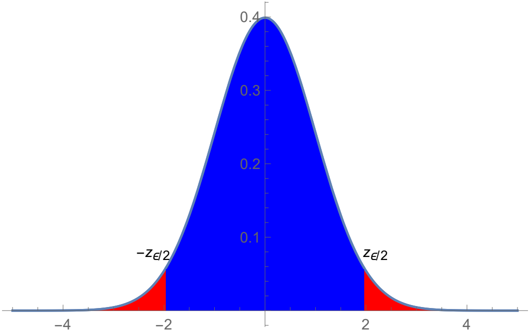
- compute the *statistic* $z = \frac{\bar{x} - \mu}{\sigma}\sqrt{n}$ from the sample;
- compute the *percentile* $z_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2)$ from the table for the standard normal distribution and the significance level, and
- if $z \in [-z_{\varepsilon/2}, z_{\varepsilon/2}]$ holds, we accept H_0; if it does not hold, we reject H_0.

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

If z is close enough to 0, then the difference between z and $z_{\varepsilon/2}$ might be due to randomness, so we accept H_0.

If z is close enough to 0, then the difference between z and $z_{\varepsilon/2}$ 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.



The sugar example

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:

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

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:

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

Finally, we check $z \in [-z_{\varepsilon/2}, z_{\varepsilon/2}]$.

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

holds, so we accept H_0 on a 95% significance level.

Tests 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 μ is a known constant;
- H_1: $m < \mu$.

Tests 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

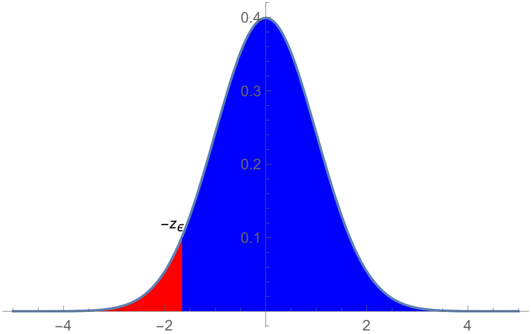
- H_0: $m = \mu$, where μ is a known constant;
- H_1: $m < \mu$.

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 ε probability is put to the left tail.)



Tests 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 μ is a known constant;
- H_1: $m > \mu$.

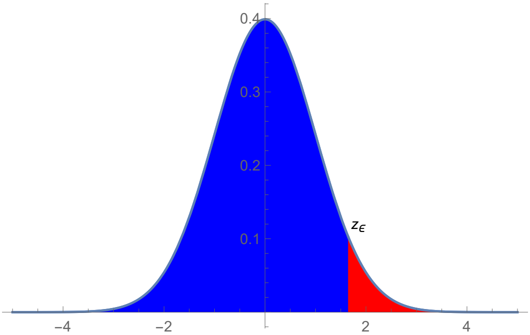
Assume we have an iid sample $X_1, \dots, X_n$ with $\mathbb{D}(X_1) = \sigma$ known and $\mathbb{E}(X_1) = m$ unknown. Let

- H_0: $m = \mu$, where μ 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.



Tests and limit theorems

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.

Tests and limit theorems

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.

Tests and limit theorems

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.

Tests and limit theorems

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.

Tests for the mean – two-sample, 2-tail z-test

In the two-sample, 2-tail z-test, we have 2 separate samples:

- $X_1, \dots, X_n$ has known deviation σ_1 and unknown mean m_1;
- $Y_1, \dots, Y_m$ has known deviation σ_2 and unknown mean m_2.

Tests for the mean – two-sample, 2-tail z-test

In the two-sample, 2-tail z-test, we have 2 separate samples:

- $X_1, \dots, X_n$ has known deviation σ_1 and unknown mean m_1;
- $Y_1, \dots, Y_m$ has known deviation σ_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: $m_1 \neq m_2$;

Tests for the mean – two-sample, 2-tail z-test

In the two-sample, 2-tail z-test, we have 2 separate samples:

- $X_1, \ldots, X_n$ has known deviation σ_1 and unknown mean m_1;
- $Y_1, \ldots, Y_m$ has known deviation σ_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.

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 σ_1 and unknown mean m_1;
- $Y_1, \ldots, Y_m$ has known deviation σ_2 and unknown mean 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 σ_1 and unknown mean m_1;
- $Y_1, \ldots, Y_m$ has known deviation σ_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 σ_1 and unknown mean m_1;
- $Y_1, \ldots, Y_m$ has known deviation σ_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.

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?

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 σ. 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 σ. 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 σ. 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?

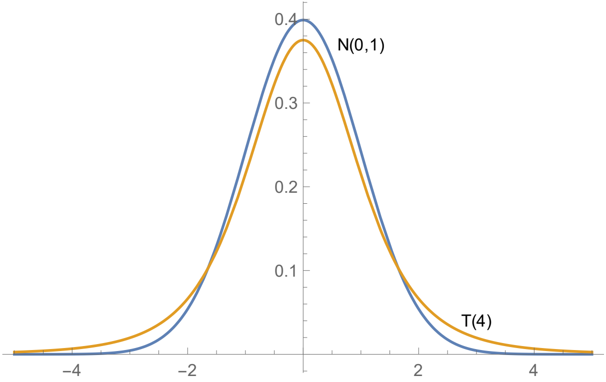
A natural idea is to use the corrected sample deviation s_n^* instead of σ. The result is the so-called t-test.

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:



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 μ is a known constant;
- H_1: $m \neq \mu$.

Tests for the mean – one-sample, 2-tail t-test

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

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

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 μ 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_ε 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.

Tests for the mean – one-sample, 1-tail t-test

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

- H_0: $m = \mu$, where μ 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_ε 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.

Problem 2

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 μ 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%?

Problem 2

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%?

Problem 2

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?

- σ 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.

Problem 2

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?

- σ 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: $\mu \neq 25$.

Problem 2

(a) We compute the statistic:

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

Problem 2

(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.

Problem 2

(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.

Problem 2

(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%?

Problem 2

(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

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

does not hold, so we reject H_0 on a 95% significance level.

Problem 4

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?

Problem 4

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. σ 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

$$(s_n^*)^2 = \frac{1}{5-1}\big((7.7 - 7.6)^2 + (8.1 - 7.6)^2 + (7.7 - 7.6)^2 + (7.7 - 7.5)^2 + (7.0 - 7.6)^2\big) = 0.16,$$

so

$$s_n^* = 0.4,$$

and the statistic is

$$t = \frac{\bar{x} - \mu}{s_n^*}\sqrt{n} = \frac{7.6 - 7.2}{0.4}\sqrt{5} = 2.236.$$

Problem 4

The percentile is the 95% 2-tail quantile of the t-distribution with degree of freedom $n - 1 = 4$:

$$t_{\varepsilon/2} = 2.776.$$

Problem 4

The percentile is the 95% 2-tail quantile of the t-distribution with degree of freedom $n-1=4$:

$$t_{\varepsilon/2} = 2.776.$$

Then the comparison

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

holds, so we accept H_0 on a 95% significance level.

Problem 4

For the second sample $7.5, 7.4, 7.3, 7.4, 7.5,$

$$\bar{x} = 7.42, \qquad s_n^* = 0.0837,$$

Problem 4

For the second sample $7.5, 7.4, 7.3, 7.4, 7.5$,

$$\bar{x} = 7.42, \qquad s_n^* = 0.0837,$$

and the statistic is

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

Problem 4

For the second sample 7.5, 7.4, 7.3, 7.4, 7.5,

$$\bar{x} = 7.42, \qquad s_n^* = 0.0837,$$

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

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

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:

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9
salary incr.	-0.2	0.3	3.6	1.4	-0.1	1.6

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:

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9
salary incr.	-0.2	0.3	3.6	1.4	-0.1	1.6

(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.

Problem 5

Solution.

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

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

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9

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

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9

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: $m > 0$.

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

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9

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: $m > 0$.

$$\bar{x} = 0.9, \qquad s_n^* = 0.2,$$

so the statistic is

$$t = \frac{\bar{x} - \mu}{s_n^*}\sqrt{n} = \frac{0.9 - 0}{0.2}\sqrt{6} = 10.06.$$

Problem 5

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

$$t_\varepsilon = 2.015.$$

Problem 5

(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.

Problem 5

(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.

Problem 5

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

Problem 5

(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

employee	1	2	3	4	5	6
salary incr.	-0.2	0.3	3.6	1.4	-0.1	1.6

Problem 5

(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

employee	1	2	3	4	5	6
salary incr.	-0.2	0.3	3.6	1.4	-0.1	1.6

Once again,

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

Now

$$\bar{x} = 1.1, \qquad s_n^* = 1.439,$$

so the statistic is

$$t = \frac{\bar{x} - \mu}{s_n^*}\sqrt{n} = \frac{1.1 - 0}{1.439}\sqrt{6} = 1.872.$$

Problem 5

(b) The comparison

$$t = 1.872 \in (-\infty, t_\varepsilon] = (-\infty, 2.015)$$

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

Problem 5

(b) The comparison

$$t = 1.872 \in (-\infty, t_\varepsilon] = (-\infty, 2.015)$$

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).

Problem 5

(b) The comparison

$$t = 1.872 \in (-\infty, t_\varepsilon] = (-\infty, 2.015)$$

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 *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.

Problem 5

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

Problem 5

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

Now we have the sample

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9
salary incr.	-0.2	0.3	3.6	1.4	-0.1	1.6

What kind of test do we do?

Problem 5

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

Now we have the sample

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9
salary incr.	-0.2	0.3	3.6	1.4	-0.1	1.6

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.

Problem 5

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

Now we have the sample

employee	1	2	3	4	5	6
work env. impr.	1.2	1.0	0.8	0.6	0.9	0.9
salary incr.	-0.2	0.3	3.6	1.4	-0.1	1.6

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.

Problem 5

(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.

Problem 5

(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.

Problem 5

(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

employee	1	2	3	4	5	6
$A - B$	1.4	0.7	-2.8	-0.8	1.0	-0.7

Problem 5

(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

employee	1	2	3	4	5	6
$A - B$	1.4	0.7	-2.8	-0.8	1.0	-0.7

- H_0: $m = 0$;
- H_1: $m < 0$ (as we want to test whether the salary increases productivity more than the work environment improvement).

Problem 5

(c)

$$\bar{x} = -0.2, \qquad s_n^* = 1.538,$$

and the statistic is

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

Problem 5

(c)

$$\bar{x} = -0.2, \qquad s_n^* = 1.538,$$

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 $n - 1 = 5$:

$$t_\varepsilon = 2.015.$$

Problem 5

(c)

$$\bar{x} = -0.2, \qquad s_n^* = 1.538,$$

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.