Inference on the mean of a population

(STATS201.stats 201 30: Statistical inference)

Anna Helga Jónsdóttir
Sigrún Helga Lund

December 11, 2012

Hypothesis tests for μ

- In this lecture we will discuss hypothesis tests and confidence intervals that apply when making inference on the mean of a population, μ.
- We will use them for example to test the hypothesis that mean precipitation in Reykjavik in june is less then 50 mm, that the average heart rate of men over fifty years old is greater then 99 beats p/minute, that the average number of nights slept in hotels and hostels in june differs from 100000 and so and so forth.
- All hypothesis test that will be discussed have the same null hypothesis, that the mean of the population is equal to a certain value that is called μ_0.

Hypothesis tests for μ

The null hypothesis is written:

$$H_0 : \mu = \mu_0$$

.

- It depends on the direction of the hypothesis test, what conclusions are made if we reject the null hypothesis.
- If the hypothesis test is two sided, we can conclude that the mean of the population, μ, differs from μ_0.
- If it is one sided we can only conclude that it is greater in one case or less in the other case then μ_0 depending on the case.

Hypothesis tests for μ

- Circumstances can be very different when we make inference on the mean of a population and we categorize them into four different cases, but each case is treated differently.
- The decision tree on slide **??** shows which case corresponds to which circumstance , but in order to select the appropriate case we need to answer three questions that are shown on slide **??**.
- The question are about the probability distribution of the population, whether it's variance, σ^2 is known and the size of the population, n. Each case is discussed separately in the following slides

Decision tree

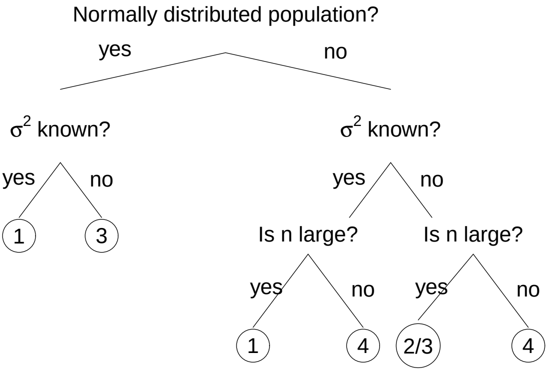

Figure: Decision tree for μ

Decision tree

These three questions are answered in correct order and the answers decide how we trace us down the decision tree.

1. Is the population normally distributed
   This need to be based on prior experience or by looking at the distribution of the sample and conclude from that. It can though be douptful if the sample is small.
2. Is the variance of the **population**, σ^2, known?
   Notice that this is rarely the case, although it may happen that such detailed prior investigations have been made that we can assume that the variance is known.
3. Is the sample large?
   We use the rule of thumb that n is large if $n > 30$. This is not a universal rule though.

Conducting hypothesis tests

Conducting hypothesis tests

1. Decide which hypothesis test is appropriate for our measurements.
2. Decide the α-level.
3. Propose a null hypothesis and decide the direction of the test (one- or two-sided).
4. Calculate the test statistic for the hypothesis test.

5a. See whether the test statistic falls within the rejection interval.

5b. Look at the p-value of the test statistic.

6. Draw conclusions.

Case 1

Case 1 corresponds to:

- When one can assume that the population follows a normal distribution and the variance (σ^2) of the distribution is known.
- when n is large and σ^2 is known, although the population is not normally distributed.

Confidence interval for μ - case 1

Confidence interval for μ - case 1

Lower bound of $1 - \alpha$ confidence interval is:

$$\bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

Upper bound of $1 - \alpha$ confidence interval is:

$$\bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

The confidence interval can thus be written as:

$$\bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

where $\bar{x}$ is the sample mean and σ is the standard deviation of the population. $z_{1-\alpha/2}$ value is found in the standard normal distribution table.

Hypothesis test for μ - case 1

Hypothesis test for μ - case 1

The null hypothesis is:

$$H_0 : \mu = \mu_0$$

The test statistic is:

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

If the null hypothesis is true, the test statistic followsthe standardized normal distribution, or $Z \sim N(0, 1)$.
The rejection areas are:

Alternative hypothesis	Reject H_0 if:
$\mathrm{H}_1 : \mu < \mu_0$	$Z < -z_{1-\alpha}$
$\mathrm{H}_1 : \mu > \mu_0$	$Z > z_{1-\alpha}$
$\mathrm{H}_1 : \mu \neq \mu_0$	$Z < -z_{1-\alpha/2}$ or $Z > z_{1-\alpha/2}$

μ - case 2

Case 2 corresponds to:

- when the sample is large and we do not know the variance of the population. We do not need to assume that the population is normally distributed.

Be careful! One can always calculate the variance of the **sample** but the variance of the **population** is rarely known!

μ - case 2

As the variance of the population is not known, we use the variance of the sample to estimate the variance of the population with

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

In order to find the standard deviation of the sample, we take the square root of the variance

$$s = \sqrt{s^2}.$$

Confidence interval for μ - case 2

Confidence interval for μ - case 2

Lower bound of $1-\alpha$ confidence interval:

$$\bar{x} - z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

Upper bound of $1-\alpha$ confidence interval:

$$\bar{x} + z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

The confidence interval can thus be written as:

$$\bar{x} - z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

where $\bar{x}$ og s are the sample mean and standard deviation. $z_{1-\alpha/2}$ value is found in the standard normal distribution table.

Hypothesis test for μ - case 2

The null hypothesis is:

$$H_0 : \mu = \mu_0$$

The test statistic is:

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

If the null hypothesis is true, the test statistic follows the standardized normal distribution, or $Z \sim N(0, 1)$.
The rejection areas are:

Alternative hypothesis	Reject H_0 if:
$\text{H}_1 : \mu < \mu_0$	$Z < -z_{1-\alpha}$
$\text{H}_1 : \mu > \mu_0$	$Z > z_{1-\alpha}$
$\text{H}_1 : \mu \neq \mu_0$	$Z < -z_{1-\alpha/2}$ or $Z > z_{1-\alpha/2}$

μ - case 3

Case 3 corresponds to two cases:
In both cases, the variance (σ^2) of the **population**, to which the sample belongs, unknown. On the other hand, we either need to assume that:

- the population is normally distributed
- or that we have many measurements in our sample (then the population does not have to be normally distributed). This is the same as case 2.

When calculating confidence intervals and conducting hypothesis test in this case, one uses the t-distribution.

Of the overlap of case 2 and 3

- Notice that when case 2 (which uses z-test) is valid, one can successfully use case 3 instead (which uses t-test). This is because when the number of degrees of freedom is large, the t-distribution is similar to the normal distribution.
- T-test, unlike z-tests, are built in most statistical software and therefore more used.
- If we are doing calculations by hand it is often better to use z-tests because then we can easily calculate p-values.

Confidence interval for μ - case 3

Confidence interval for μ - case 3

Lower bound of $1-\alpha$ confidence interval:

$$\bar{x} - t_{1-\alpha/2,(n-1)} \cdot \frac{s}{\sqrt{n}}.$$

Upper bound of $1-\alpha$ confidence interval:

$$\bar{x} + t_{1-\alpha/2,(n-1)} \cdot \frac{s}{\sqrt{n}}.$$

The confidence interval can thus be written as:

$$\bar{x} - t_{1\alpha/2,(n-1)} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2,(n-1)} \cdot \frac{s}{\sqrt{n}}$$

where $\bar{x}$ og s are the mean and the standard deviation of the sample and $t_{1-\alpha/2,(n-1)}$ value is found in the t-table.

Hypothesis test for μ - case 3

Hypothesis test for μ - case 3

The null hypothesis is:

$$H_0 : \mu = \mu_0$$

The test statistic is:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

If the null hypothesis is true, the test statistic follows t-distribution with $(n-1)$ degree of freedom, or $T \sim t_{(n-1)}$.

The rejection areas are:

Alternative hypothesis	Reject H_0 if:
$\mathrm{H}_1 : \mu < \mu_0$	$T < -t_{1-\alpha,(n-1)}$
$\mathrm{H}_1 : \mu > \mu_0$	$T > t_{1-\alpha,(n-1)}$
$\mathrm{H}_1 : \mu \neq \mu_0$	$T < -t_{1-\alpha/2,(n-1)}$ or $T > t_{1-\alpha/2,(n-1)}$

μ - case 4

In case 4 one can neither use z-test nor t-test unless further approximations are used. In these cases one can do one of the following:

- Transform the data
- Use nonparametric tests
- Check whether the population follows some other known distribution and use tests that are applicable for them.