Probability Theory and Statistics

Lecture 6: Hypothesis testing, part 1

26 April 2025

Lecturer: Batashov Ruslan

Email: rusbatashov@gmail.com

Statistical hypotheses

A statistical hypothesis is a statement about a population distribution. To test a hypothesis means to say whether it is "likely".

Examples of statistical hypotheses:

1. The mean salary in a region is 45 000 RUB
2. Customers equally likely prefer Apple and Samsung

Not statistical hypotheses:

1. The sample mean salary is 50 000 RUB
2. On a certain day, 50 clients of the insurance company filed claims

Exercise – are the following statistical hypotheses?

1. The population distribution is normal
2. The population size is 1 mln. elements

The procedure of hypothesis testing

(1) Always we begin with formulating the 2 hypotheses:

- A null hypothesis (H_0) is considered to be true unless we find statistical evidence against it.
- An alternative hypothesis (H_1) is accepted if we find evidence against the null hypothesis.

Example

$H_0: p = 0.5$ (customers equally likely prefer Apple or Samsung)
$H_1: p \neq 0.5$

Often, H_0 is stated as the absence of some effect, H_1 as the presence of the effect.

(2) We compute necessary numerical quantities from sample data which are needed to make a decision.

For example, compute $\widehat{p}$.

(3) We apply a certain rule to make one of the two conclusions:

- reject the null hypothesis and accept the alternative hypothesis;

OR

- do not reject the null hypothesis.

Example

If $\widehat{p} \notin [0.4,\ 0.6]$, then reject $H_0 : p = 0.5$ and accept $H_1 : p \neq 0.5$.
If $\widehat{p} \in [0.4,\ 0.6]$, then do not reject H_0.

If the null hypothesis is not rejected it may mean that either it is indeed true, or we do not have enough evidence against it.

Simple null hypotheses

Let θ be an unknown parameter of a population distribution.

- A simple null hypothesis can be formulated in the form

$$H_0\colon \theta = \theta_0$$

  where θ_0 is a given number (for example, $H_0\colon p = 0.5$).

- A two-sided alternative hypothesis can be formulated in the form

$$H_1\colon \theta \neq \theta_0$$

For a while, we will consider only simple null hypotheses and two-sided alternative hypotheses.

Two types of test errors

Two types of incorrect decisions (errors) which can occur because a sample is random:

- Type I error, also called false positive
  Reject the null hypothesis when it is true

- Type II error, also called false negative
  Fail to reject the null hypothesis when it is false

Because consequences of type I and type II errors are often different, the goal of designing a test is:

for a given probability of type I error,
find a test with the smallest probability of type II error.

Notation: α = probability of type I error, β = probability of type II error.

Probabilities of type I and type II errors

- Probability of type I error (significance / significance level of a test)

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

  α is a number (because we consider simple null hypotheses).

- Probability of type II error

$$\beta = P(\text{not reject } H_0 \mid H_0 \text{ is false})$$

  β is a function of the parameter in the alternative hypothesis.

- Power of a test

$$power = 1 - \beta = P(\text{reject } H_0 \mid H_0 \text{ is false})$$

Example

A person is suspicuous that his dice is unfair for sides 5 and 6. In order to do so he rolls dice five times. If 5 or 6 do not appear or appear more than 3 times he concludes that dice is unfair. Find

- significance
- type II error and power of the test if dice shows 5 and 6 with probabilities 1/3 and remaining sides with equal probabilities.

Hypotheses testing and Machine Learning (ML)

Confusion matrix helps to assess quality of certain algorithm prediction

H_1 - presence of some effect, H_0 absence of some effect

	Decision: reject H_0	Decision: not reject H_0
H_0 is false	TP (true positive)	FN (false negative)
H_0 is true	FP (false positive)	TN (true negative)

Popular ML metrics:

- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$

Which cells of confusion matrix correspond to α, β, power?

Hypotheses testing and Machine Learning (ML)

Confusion matrix helps to assess quality of certain algorithm prediction

H_1 - presence of some effect, H_0 absence of some effect

	Decision: reject H_0	Decision: not reject H_0
H_0 is false	TP (true positive) power	FN (false negative) β
H_0 is true	FP (false positive) α	TN (true negative)

Popular ML metrics:

- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$ =power

Example

Test whether the coin is fair:

$$H_0\colon p = 0.5, \qquad H_1\colon p \neq 0.5.$$

We reject H_0 if $|\widehat{p} - 0.5| > 0.098$. Find significance level of this test if 100 coin flips have been made.

Test statistics and acceptance/rejection regions

A test statistic is a numerical characteristic computed for a sample, which is used to make a decision on whether to reject H_0.

- A rejection region (RR): a region of values of the test statistic where H_0 is rejected.
- An acceptance region (AR): a region of values of the test statistic where H_0 is not rejected.

Acceptance and rejection regions are found from the sampling distribution of a test statistic when H_0 is true.

In example above

$$AR = \{|\widehat{p} - 0.5| \leq 0.098\},$$
$$RR = \{|\widehat{p} - 0.5| > 0.098\}$$

Standard test for a population proportion

Suppose the population has the Bernoulli distribution and we want to test

$$H_0 : p = p_0, \qquad H_1 : p \neq p_0.$$

Let test statistic be

$$z_{st} = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

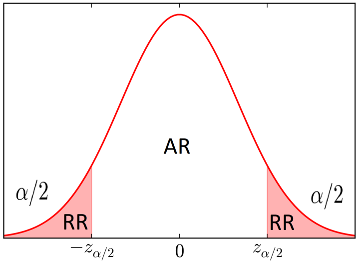
The test of significance α:

$$\text{reject } H_0 \text{ if } |z_{st}| > z_{\alpha/2}.$$

Conditions: $np_0 \geqslant 10$ and $n(1-p_0) \geqslant 10$.

$z_{\alpha/2}$ above is called z_{cr} and rejection region becomes $RR = \{|z_{st}| > z_{cr}\}$.

$$RR = \{|Z_{st}| > Z_{\alpha/2}\}.$$



Reading

Newbold, Carlson, Thorne: § 9.1 – 9.4 (skip the material related to p-values and one-sided alternatives)

Mann: § 9.1 – 9.4 (also skip the material related to p-values)

Probability Theory and Statistics

Lecture 6: Hypothesis testing, part 2

26 April 2025

Lecturer: Batashov Ruslan

Email: rusbatashov@gmail.com

Review of the previous lecture: examples

Example 1: Rejection region is given. What is α?

A researcher wants to verify the statement that approximately 10% of people are left-handed.

In order to do that researcher takes a sample of 200 people. If number of left-handed people is less than 10 or greater than 30 researcher concludes that the statement is false.

What is type I error (significance level) of proposed test?

Example 2: Significance level is given. What is RR?

A researcher wants to verify the statement that approximately 10% of people are left-handed.

In a sample of 200 people, he finds that 25 are left-handed. What can be said about the statement based on the sample data?

Use the 10% significance level.

Example 3

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ iid with unknown μ and known $\sigma = 1$. Suppose that based on a sample of $n = 25$ elements we want to test

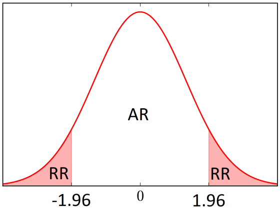
$$H_0 \colon \mu = 0, \qquad H_1 \colon \mu \neq 0.$$

Let

$$Z_{st} = \frac{\overline{x}}{\sigma / \sqrt{n}}$$

Consider the test: reject H_0 if $|Z_{st}| > 1.96$

do not reject H_0 if $|Z_{st}| \leqslant 1.96$;

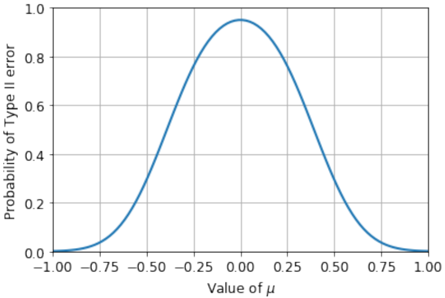


- The probability of type I error is 5%:

$$\alpha = P(|Z| > 1.96) = 0.05$$

(because if H_0 is true, then $Z \sim N(0, 1)$).

- The probability of type II error depends on the actual value of μ, which under H_1 can be any value except zero.



Exercise: how was the above graph obtained?

Standard tests for a population mean

1. z-test for a population mean (known σ)

Conditions: $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ iid with known σ; or any population distribution with known σ and sample size $n \geqslant 40$.

For a sample of size n, given number μ_0 and significance level α, we test the hypothesis

$$H_0\colon \mu = \mu_0, \qquad H_1\colon \mu \neq \mu_0.$$

How to test this hypothesis? What rejection region should we choose?

Example

It is reported that the lake contains 0.5g of salt per 1 liter with population standard deviation 0.25g.

In order to check this statement 50 samples of water were chosen and the mean amount of salt in a sample of one liter was 0.57g. Is the report of the salt content correct at 5% significance level?

$$H_0\colon \mu = \mu_0, \qquad H_1\colon \mu \neq \mu_0.$$

Let

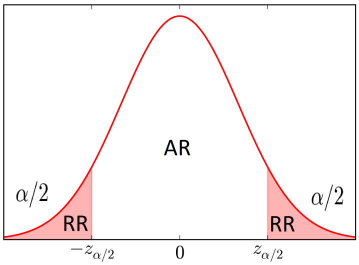
$$z_{st} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

The test of significance α:

$$\text{reject } H_0 \text{ if } |z_{st}| > z_{\alpha/2}.$$

Conditions: normality of population distribution or $n \geqslant 40$.

$$RR = \{|Z_{st}| > Z_{\alpha/2}\}.$$



2. t-test for a population mean (unknown σ)

$$H_0 \colon \mu = \mu_0, \qquad H_1 \colon \mu \neq \mu_0.$$

Let

$$t_{st} = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

The test of significance α:

$$\text{reject } H_0 \text{ if } |t_{st}| > t_{\alpha/2}(n-1).$$

Condition: normal population distribution.

If $(n-1) > 100$, it is possible to use $z_{\alpha/2}$ instead of $t_{\alpha/2}(n-1)$.

Example

An owner of a company says that the company, on average, has a monthly profit of 1 million.

During the last 24 months, the mean monthly profit was 950 thousand with standard deviation 50 thousand.

Verify the statement at the 5% significance level. Assume that the profit distribution is normal.

Test statistics and acceptance/rejection regions

A test statistic is a numerical characteristic computed for a sample, which is used to make a decision on whether to reject H_0.

Statistics used in the standard tests

- The z-statistic used in the test for a population mean (known σ)

$$z_{st} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

- The z-statistic used in the test for a population proportion

$$z_{st} = \frac{\widehat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

- The t-statistic used in the test for a population mean (unknown σ)

$$t_{st} = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

Reading

Newbold, Carlson, Thorne: § 9.1 – 9.4 (except p-values).

Mann: § 9.1 – 9.4.

Probability Theory and Statistics

Lecture 6: Hypothesis testing, part 3

26 April 2025

Lecturer: Batashov Ruslan

Email: rusbatashov@gmail.com

Confidence intervals and hypothesis testing

The test procedure for $H_0\colon \mu = \mu_0$ against $H_1\colon \mu \neq \mu_0$ at a significance level α is equivalent to the following one:

1. For a given sample, construct the confidence interval $\mu = \overline{x} \pm moe$ (moe = margin of error) with confidence level $1 - \alpha$.

2. If $\mu_0 \notin [\overline{x} - moe, \overline{x} + moe]$, then reject H_0. Otherwise do not reject H_0.

$$\boxed{\mu_0 \in CI_{1-\alpha}(\mu) \Longleftrightarrow don't\ reject\ H_0\ at\ sign.\ level\ \alpha}$$

Hypothesis testing for population variance

Suppose the population distribution is $N(\mu, \sigma^2)$ with unknown μ, σ. For given σ_0, consider the hypothesis

$$H_0\colon \sigma = \sigma_0,$$
$$H_1\colon \sigma \neq \sigma_0.$$

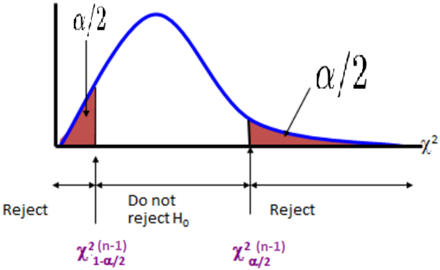
How to test this hypothesis for given significance level α?

Example

Temperature of an object is determined by measurement device which as stated in documentation has standard deviation $1°C$. At 5% significance level test validity of this claim if 25 measurements $t_1, \ldots, t_{25}$ have been made and

$$\sum_{i=1}^{25} t_i = 250.5, \ \sum_{i=1}^{25} t_i^2 = 2539.05$$

Assume normality of distribution.



Suppose the population distribution is $N(\mu, \sigma^2)$ with unknown μ, σ. For given σ_0, consider the hypothesis

$$H_0\colon \sigma = \sigma_0,$$
$$H_1\colon \sigma \neq \sigma_0.$$

The test statistic (the χ^2-statistic)

$$\chi^2_{st} = \frac{(n-1)\cdot s^2}{\sigma_0^2}.$$

Then the rejection region at a significance level α:

$$RR = \{\chi^2_{st} < \chi^2_{1-\frac{\alpha}{2}}(n-1) \cup \chi^2_{st} > \chi^2_{\frac{\alpha}{2}}(n-1)\}$$

P-values

The p-value of a test for a given sample is the smallest significance level at which the null hypothesis can be rejected for this sample.

Example

$X_1 \ldots, X_{10} \sim N(\mu, 10^2)$. Test $H_0 \colon \mu = 0$ against $H_1 \colon \mu \neq 0$.

Suppose we obtain a sample:

$$4,\ 2,\ 15,\ -13,\ 2,\ -2,\ 9,\ 9,\ 20,\ 17$$

($n = 10$, $\overline{x} = 6.3$, $z_{st} = 1.99$).

At what significance level can we reject H_0?

- We can reject H_0 at the 5% significance level, because $|z_{st}| > 1.96$
- We cannot reject H_0 at the 1% significance level, because $|z_{st}| < 2.58$.
- ......

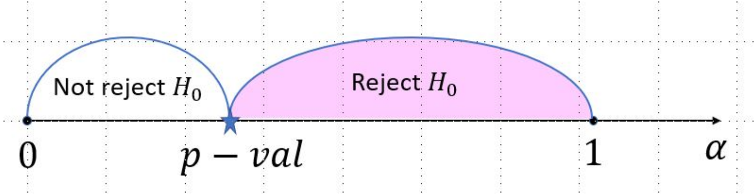
Example (continued)

The minimum significance level at which we can reject H_0 can be found as

$$\alpha_* = P(|Z| > |z_{st}|)$$

where $Z \sim N(0, 1)$.

This special significance level α_* is called the p-value.



The p-value of a test shows "how confident" we are in that the null hypothesis can be rejected

A p-value is always a number between 0 and 1. Smaller p-values mean that we have stronger evidence against H_0.

The standard interpretation of p-values

- $p \leqslant 0.01$: very strong evidence against H_0
- $0.01 < p \leqslant 0.05$: strong evidence against H_0
- $0.05 < p \leqslant 0.1$: moderate evidence against H_0
- $p > 0.1$: not enough evidence to reject H_0

Computation of p-values

- The z-test for a population mean (σ is known)

$$p - val = P(|Z| > |z_{st}|), \quad \text{where } Z \sim N(0,1) \text{ and } z_{st} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

- The z-test for a population proportion

$$p{-}val = P(|Z| > |z_{st}|), \quad \text{where } Z \sim N(0,1) \text{ and } z_{st} = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

- The t-test for a population mean (σ is unknown)

$$p - val = P(|T| > |t_{st}|), \quad \text{where } T \sim t(n-1) \text{ and } t_{st} = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

Often it is convenient to use the equivalent formulas:
$p - val = 2P(Z > |z_{st}|)$ or $p - val = 2P(T > |t_{st}|)$

Example

A researcher wants to verify the statement that approximately 10% of people are left-handed.

In a sample of 200 people, he finds that 25 are left-handed. What can be said about the statement based on the sample data?

Example

An owner of a company says that the company, on average, has a monthly profit of 1 million.

During the last 24 months, the mean monthly profit was 950 thousand with standard deviation 50 thousand.

Verify the owner's statement. Assume that the profit distribution is normal.

Reading

Newbold, Carlson, Thorne: § 9.1 – 9.4 (p-values), 9.6 (the test for σ^2).

Mann: § 9.1 – 9.4.