Probability and Statistics

Raluca Muresan

West University of Timisoara

raluca.muresan@e-uvt.ro

Overview

1. Hypothesis Testing
    - Preliminaries
    - Z-test
    - T tests: unknown σ
    - χ^2 test for variance/standard deviation
    - Comparison of two variances
    - Hypothesis testing and confidence intervals

Prelimiaries

What have we done so far:

- described data using numerical characteristics and statistical graphs
- estimated population parameters point-wise and using confidence intervals

Next, another thing we can do is **testing hypothesis**, i.e verifying statements, claims, conjectures.

For example, using a random sample we can verify whether

- the average number of current users increased by 2000 this year
- the average connection speed is 54 Mbps, as claimed by the internet service provider
- service times have Gamma distribution

What are we testing?

- H_0 = hypothesis (the null hypothesis)
- H_A = alternative (the alternative hypothesis)

which are simply two mutually exclusive statements.

• Each test consists in rejecting H_0 in favor of H_A or failing to reject it.

• H_0 is always an **equality, absence of an effect or relation**

• In order to reject H_0, we need **significant evidence** in favor of H_A (provided by a sample).

- Hypothesis testing is similar to a criminal trial.

Type I and type II errors

The decision of rejecting or not the H_0 could be wrong due to the **sampling error**.

	Result of the test	
	Reject H_0	**Accept H_0**
H_0 **is true**	*Type I error*	correct
H_0 **is false**	correct	*Type II error*

Definition

A **type I error** occurs when we reject the true null hypothesis.
A **type II error** occurs when we accept the false null hypothesis.

A type I error is often considered more dangerous and undesired than a type II error. Making a type I error can be compared with convicting an innocent defendant or sending a patient to a surgery when (s)he does not need one.

Definition

Probability of a type I error is the **significance level** of a test,

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

The power of a test is

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

Significance levels used: 0.1, 0.05 or 0.01.

In order to define a **statistical test** we have to:

- specify the hypothesis H_0 and the alternative H_A
- choose a level of significance α
- compute the **test statistic** (T) from the sample; the test statistic has a known distribution if H_0 is true
- define the **acceptance region** and the **rejection region**
- formulate the **conclusion**: if the test statistic falls in the rejection area, H_0 is rejected; otherwise it is accepted

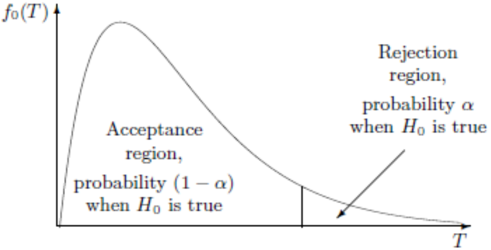


FIGURE 9.6: *Acceptance and rejection regions.*

$$P(T \in \text{ rejection region } | H_0) = \alpha.$$

To avoid type II errors, we choose such a rejection region that will likely cover the test statistic T in case if the alternative H_A is true. This maximizes the power of our test because we'll rarely accept H_0 in this case. (examples!)

• There are two types of tests: **one-tailed tests** and **two-tailed tests**
One-tailed tests:

$$H_0 : \theta = \theta_0$$

$$\begin{cases} H_A : \theta < \theta_0, & \mathcal{R} = (-\infty, t_\alpha) \\ H_A : \theta > \theta_0, & \mathcal{R} = (t_\alpha, \infty) \end{cases}$$

Two-tailed tests:

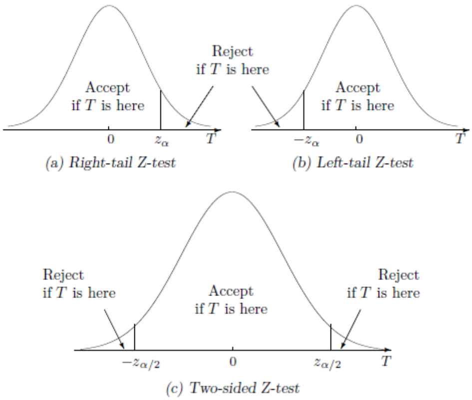
$$H_0 : \theta = \theta_0$$

$$H_0 : \theta \neq \theta_0,$$

$$\mathcal{R} = (-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty))$$

Z-test

- used when the test statistic (Z) has a Standard Normal Distribution



• Now consider testing a hypothesis about a population parameter θ.

• Suppose that its estimator $\hat{\theta}$ has Normal distribution, at least approximately, and we know $E(\hat{\theta})$ and $Var(\hat{\theta})$ if the hypothesis is true.

• Then the test statistic

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}}$$

has Standard Normal distribution and is called a **Z-statistic**.

• Z-tests can be used for the **population mean, proportion, comparing two population means and proportions.**

Z-test about the population mean

The standard deviation of the population σ is **known**.
Hypothesis:

$$H_0 : \mu = \mu_0$$

$$\begin{cases} H_A : \mu < \mu_0, & \mathcal{R} = (-\infty, z_\alpha) \\ H_A : \mu > \mu_0, & \mathcal{R} = (z_\alpha, \infty) \\ H_A : \mu \neq \mu_0 & \mathcal{R} = (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty) \end{cases}$$

Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Example 1. The number of concurrent users for some internet service provider has always averaged 5000 with a standard deviation of 800. After an equipment upgrade, the average number of users at 100 randomly selected moments of time is 5200. Does it indicate, at a 5% level of significance, that the mean number of concurrent users has increased? Assume that the standard deviation of the number of concurrent users has not changed.

Z-test about a population proportion

Hypothesis:

$$H_0 : p = p_0$$

$$\begin{cases} H_A : p < p_0, & \mathcal{R} = (-\infty, z_\alpha) \\ H_A : p > p_0, & \mathcal{R} = (z_\alpha, \infty) \\ H_A : p \neq p_0 & \mathcal{R} = (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty) \end{cases}$$

Test statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

Example 2. We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it. The manufacturer claims that at most one in 10 items in the shipment is defective. At the 1% level of significance, do we have sufficient evidence to disprove this claim?

Z-test comparing proportions of two populations

Hypothesis:

$$H_0 : p_1 = p_2$$

$$\begin{cases} H_A : p_1 < p_2, & \mathcal{R} = (-\infty, z_\alpha) \\ H_A : p_1 > p_2, & \mathcal{R} = (z_\alpha, \infty) \\ H_A : p_1 \neq p_2 & \mathcal{R} = (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty) \end{cases}$$

Test statistic:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n} + \frac{1}{m})}} \sim N(0,1), \;\; \hat{p} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m}$$

Example 3. A quality inspector finds 10 defective parts in a sample of 500 parts received from manufacturer A. Out of 400 parts from manufacturer B, she finds 12 defective ones. A computer-making company uses these parts in their computers and claims that the quality of parts produced by A and B is the same. At the 5% level of significance, do we have enough evidence to disprove this claim?

One-sample t tests I

The population standard deviation σ is **unknown**.
Hypothesis:

$$H_0 : \mu = \mu_0$$

$$\begin{cases} H_A : \mu < \mu_0, & \mathcal{R} = (-\infty, t_\alpha) \\ H_A : \mu > \mu_0, & \mathcal{R} = (t_\alpha, \infty) \\ H_A : \mu \neq \mu_0 & \mathcal{R} = (-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty) \end{cases}$$

Test statistic:

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

has a Student t distribution with n-1 degrees of freedom, where s is the sample standard deviation

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}.$$

One-sample t tests II

Example 4. (Unauthorized use of a computer account, continued). A long-time authorized user of the account makes 0.2 seconds between keystrokes. One day, the data in Example 9.19 on p. 260 are recorded as someone typed the correct username and password. At a 1% level of significance, is this an evidence of an unauthorized attempt?

.24, .22, .26, .34, .35, .32, .33, .29, .19, .36, .30, .15, .17, .28, .38, .40, .37, .27

T-tests for comparison of two means (Two-sample t test) I

The population standard deviations σ_1, σ_2 are **unknown, but equal**.
Hypothesis:

$$H_0 : \mu_1 = \mu_2$$

$$\begin{cases} H_A : \mu_1 < \mu_2, & \mathcal{R} = (-\infty, t_\alpha) \\ H_A : \mu_1 > \mu_2, & \mathcal{R} = (t_\alpha, \infty) \\ H_A : \mu_1 \neq \mu_2 & \mathcal{R} = (-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty) \end{cases}$$

Test statistic:

$$T = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

has a Student t distribution with $m + n - 2$ degrees of freedom, where s_p is the pooled standard deviation (s_x, s_y are the sample standard deviations)

$$s_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}.$$

T-tests for comparison of two means (Two-sample t test) II

Example 5. (CD writer and battery life). Does a CD writer consume extra energy, and therefore, does it reduce the battery life on a laptop? The data collected is the following: eighteen users without a CD writer worked an average of 5.3 hours with a standard deviation of 1.4 hours; other twelve, who used their CD writer, worked an average of 4.8 hours with a standard deviation of 1.6 hours. Consider a level of significance $\alpha = 0.1$

Two-sample t test I

If the two standard deviations σ_1 and σ_2 are **not equal**, then the test statistic is

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

which has a Student t distribution with ν degrees of freedom,

$$\nu = \left(\frac{s_X^2}{n} + \frac{s_Y^2}{m} \right)^2 \Big/ \left(\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)} \right),$$

s_X and s_Y are the sample standard deviations.

Example 6. (Comparison of two servers, continued). A certain computer algorithm is executed 30 times on server A and 20 times on server B with the following results. Is server A faster? Formulate and test the hypothesis at a level $\alpha = 0.05$.

Two-sample t test II

	Server A	Server B
Sample mean	6.7 min	7.5 min
Sample standard deviation	0.6 min	1.2 min

χ^2 test for variance/standard deviation

Hypothesis:

$$H_0 : \sigma^2 = \sigma_0^2$$

$$\begin{cases} H_A : \sigma^2 < \sigma_0^2, & \mathcal{R} = (0, \chi^2_{\alpha}) \\ H_A : \sigma^2 > \sigma_0^2, & \mathcal{R} = (\chi^2_{1-\alpha}, \infty) \\ H_A : \sigma^2 \neq \sigma_0^2 & \mathcal{R} = (0, \chi^2_{\alpha/2}) \cup (\chi^2_{1-\alpha/2}, \infty) \end{cases}$$

Test statistic:

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

has a χ^2 distribution with n-1 degrees of freedom, where s^2 is the sample variance.

Example 7. A sample of 6 measurements

$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

is collected from a Normal distribution with mean μ and standard deviation σ. Test if the value of the standard deviation is $\sigma = 2.2$ with a level of significance $\alpha = 5\%$.

Example 8. Installation of a certain hardware takes random time with a standard deviation of 5 minutes. A manager questions this assumption. Her pilot sample of 40 installation times has a sample standard deviation of $s = 6.2$ min, and she says that it is significantly different from the assumed value of $\sigma = 5$ min. Do you agree with the manager? Conduct the suitable test of a standard deviation.

Comparison of two variances I

Hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$\begin{cases} H_A : \sigma_1^2 < \sigma_2^2, & \mathcal{R} = (0, F_\alpha) \\ H_A : \sigma_1^2 > \sigma_2^2, & \mathcal{R} = (F_{1-\alpha}, \infty) \\ H_A : \sigma_1^2 \neq \sigma_2^2 & \mathcal{R} = (0, F_{\alpha/2}) \cup (F_{1-\alpha/2}, \infty) \end{cases}$$

Test statistic:

$$F = \frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2}$$

has a F distribution with $n-1, m-1$ degrees of freedom, where s_x^2, s_y^2 are the sample variances.

Example 9. A data channel has the average speed of 180 Megabytes per second. A hardware upgrade is supposed to improve stability of the data transfer while maintaining the same average speed. Stable data transfer rate implies low standard deviation. After the upgrade, the instantaneous speed of data transfer, measured at 16 random instants, yields a standard deviation of 14 Mbps. Records show that the standard deviation was 22 Mbps before the upgrade, based on 27 measurements at random times. Test whether the stability wasimproved with a significance level $\alpha = 5\%$.

Hypothesis testing and confidence intervals I

We can conduct two-sided tests using nothing but the confidence intervals!

A level α Z-test of $H_0 : \theta = \theta_0$ vs $H_A : \theta \neq \theta_0$
accepts the null hypothesis

if and only if

a symmetric $(1-\alpha)100\%$ confidence Z-interval for θ contains θ_0.

What about one sided tests? Answer: we proceed in a similar manner, but how?

Example 7. A sample of 6 measurements

$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

is collected from a Normal distribution with mean μ and standard deviation σ. Test whether $\mu = 6$ against a two-sided alternative HA : $\mu \neq 6$ at the 5% level of significance.

The End