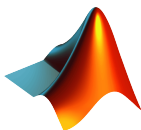


MatLab Programming Fundamentals

guarantor: Maroš Tunák

tel.: 3465

e-mail: maros.tunak@tul.cz



Course objectives

The aim of the course is to acquire basics knowledge and skills of students the MatLab program. At the end of the course students will be able to use MatLab for their own work and will be ready to deepen their programming skills in MatLab.

MatLab Programming Fundamentals

time requirements: 0p+2c

credits: 4

exercises: Monday 10:40-12:15; 12:30-14:05 (B-PC2, Tunák M.)
Tuesday 08:50-10:25; 10:40-12:15 (B-PC2, Tunák M.)

consultation: Wednesday 10:40-12:15 (E-KHT)

Requirements on student/graded credit

- 1 participation in exercises (max. 3 absences)
- 2 elaboration of semester work (after approval of the semester work, you can attend a practical demonstration)
- 3 practical demonstration of acquired skills (there will be 1-2 examples to solve; elaboration time 1 hour; you can use any materials ...)

Content

IS/STAG Syllabus

1. Getting started with Matlab. Working environment, windows, paths, basic commands, variables. Loading, saving and information about variables. Help.
2. Mathematics with vectors and matrices. Creating vectors and matrices. Indexing. Special matrices. Matrix operations. Element by element operations. Relational operations, logical operations, examples and tricks.
3. Control flow. Loops, conditional statements, examples.
4. Script m-files, Function m-files.
5. Visualisation. Two-dimensional graphics. Three-dimensional graphics.
6. Graphical user interface.
- 7.-10. Statistics and Machine Learning Toolbox. Basics of statistical data processing, exploratory data analysis, descriptive statistics, data visualisation, hypothesis testing, confidence intervals, regression analysis, control charts.
- 11.-13. Solution of practical problems in textile and industrial engineering.

Literature

Recommended

MathWorks. *Getting Started with MATLAB*. [Online]. Dostupné z:

<https://www.mathworks.com/help/matlab/getting-started-with-matlab.html>

Study materials

<http://elearning.tul.cz>

Installation

<http://liane.tul.cz/cz/software/MATLAB>

Statistics and Machine Learning Toolbox

Basics of statistical data processing, exploratory data analysis, descriptive statistics, data visualisation, hypothesis testing, confidence intervals, regression analysis, control charts.

Hypothesis Testing

Hypothesis Testing

Hypothesis Testing

Statistical hypothesis - assumption about probability distribution of one or more random variables. The assumption applies to the parameters of distribution of a random variable in the population or can only relate to of the law of distribution of a random variable (to the distribution function, the quantile function or probability density function).

Test statistical hypothesis - a rule based on results obtained from random sample objectively prescribes the decision, if any whether or not the verified hypothesis is rejected. When testing, a distinction is made between the tested null hypothesis H_0 and the alternative hypothesis H_A . The test should decide whether to reject the null hypothesis or not. An alternative hypothesis is the one we accept if we reject the null hypothesis.

To test the null hypothesis, relevant test statistics are constructed. If it falls this statistic into the region of acceptance, the null hypothesis is does not reject. However, if it falls into a critical region, it is zero hypothesis rejected. The probability of a test statistic falling into a critical region is called the significance level α (0.05, 0.01). The critical region can be defined two tailed or one tailed. Two tailed defines if there is no reason why the test statistic should have either only a positive or negative sign. The significance level α is then decomposed on two identical parts of size $\alpha/2$.

Hypothesis Testing

The process of testing hypotheses can be divided into several steps.

- formulation of null hypothesis H_0 and alternative hypothesis H_A according to the nature of the problem
- choice of significance level α
- choice of test statistics, i.e. function of random sample values with a known probability distribution in case of the null hypothesis is valid
- determination of the critical region of the test characteristic on the basis of its probability distribution and significance level
- obtaining a random sample, calculation of test statistics and their quantiles, which form the critical region
- decision whether
 - reject hypothesis H_0 and accept H_A if the calculated value of test statistics will fall into the critical region
 - do not reject hypothesis H_0 if the calculated test value characteristics do not fall into the critical region

Hypothesis Tests for One Sample

A one sample ttest

ttest

Let x_1, \dots, x_n sample from distribution $N(\mu, \sigma^2)$, where $n > 1$ and $\sigma^2 > 0$ is unknown. We test the null hypothesis against the alternative

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0$$

Test statistics:

$$T = \frac{\bar{x} - \mu_0}{s} \sqrt{(n)}$$

Hypothesis H_0 is rejected at the significance level α if it is valid

$$|T| \geq t_{n-1}(1 - \alpha/2)$$

If the inequality does not apply, then we do not reject H_0 . Critical fields of hypothesis tests $H_0 : \mu = \mu_0$ against various alternatives H_A for significance level α

H_0	H_A	testing characteristics	critical region
$\mu = \mu_0$	$\mu > \mu_0$	$T = \frac{\bar{x} - \mu_0}{s} \sqrt{(n)}$	$T \geq t_{n-1}(1 - \alpha)$
	$\mu < \mu_0$		$T < t_{n-1}(\alpha)$
	$\mu \neq \mu_0$		$ T \geq t_{n-1}(1 - \alpha/2)$

Hypothesis Tests for One Sample

- **Example:** When inspecting the packaging machine to be filled packages weighing 1 kg, the following deviations were found during the exact weighing of 5 packages: $-3, 2, -2, 0, -1$ It is necessary to find out whether the automaton has a systematic deviation from the required value.¹

Solution: We will consider individual deviations as realizations of independent random variables with distribution $N(\mu, \sigma^2)$, where σ^2 is an unknown parameter. It is necessary to test the hypothesis $H_0 : \mu = 0$ (i.e. that the deviations fluctuate around zero and are not systematically shifted to either positive or negative values). We have:

$$n = 5, \mu_0 = 0$$

$$T = \frac{-0.8 - 0}{1.9235} \sqrt{5} = -0.9300$$

Since $0.9300 \geq t_4(0.975) = 2.7764$ does not apply to inequality, the data obtained do not contradict the assumption that the automaton does not have a systematic deviation (we do not reject H_0).

¹ Anděl, J., Matematická statistika, SNTL Praha, 1985

Hypothesis Tests for One Sample

```
>> x=[-3 2 -2 0 -1]
x =
    -3     2    -2     0    -1
>> [h,p,ci,stats]=ttest(x)
h =
     0
p =
    0.4050
ci =
   -3.1884    1.5884
stats =
  struct with fields:
    tstat: -0.9300
       df: 4
       sd: 1.9235
```

- The result h is 1, if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise.
- p -value is the probability of observing a test statistic as extreme as, or more extreme than, the observed value under the null hypothesis (small values of p cast doubt on the validity of the null hypothesis.)
- ci - confidence interval for the true population mean, returned as a two-element vector containing the lower and upper boundaries of the 100% (1- α) confidence interval.
- $stats$ - $tstat$ - value of the test statistic, df - degrees of freedom of the test, sd - estimated population standard deviation

Hypothesis Tests for One Sample

A one sample variance test

varitest

Let x_1, \dots, x_n sample from distribution $N(\mu, \sigma^2)$, where $n > 1$ and $\sigma^2 > 0$ is unknown. We test the null hypothesis against the alternative

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

$$H_A : \sigma^2 \neq \sigma_0^2$$

Test statistics:

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

Hypothesis H_0 is rejected at the significance level α if it is valid

$$\chi_{n-1}^2(\alpha/2) \leq \chi^2 \leq \chi_{n-1}^2(1 - \alpha/2)$$

If the inequality does not apply, then we do not reject H_0 . Critical field of hypothesis tests $H_0 : \sigma^2 = \sigma_0^2$ against various alternatives H_A for significance level α

H_0	H_A	testing characteristics	critical region
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 \geq \chi_{n-1}^2(1 - \alpha)$
	$\sigma^2 < \sigma_0^2$		$\chi^2 < \chi_{n-1}^2(\alpha)$
	$\sigma^2 \neq \sigma_0^2$		$\chi_{n-1}^2(\alpha/2) \leq \chi^2 \leq \chi_{n-1}^2(1 - \alpha/2)$

Hypothesis Tests for Two Samples

Paired t-test

ttest

Let's have a random sample of $(y_1, z_1), \dots, (y_n, z_n)$ from some a two-dimensional distribution whose vector of mean values is (μ_1, μ_2) . We want to test the null hypothesis against the alternative

$$H_0 : \mu_1 - \mu_2 = \Delta$$

$$H_A : \mu_1 - \mu_2 \neq \Delta$$

where Δ s a given number. We put $x_1 = y_1 - z_1, \dots, x_n = y_n - z_n$. Quantities (x_1, \dots, x_n) are independent and we assume that $x_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$. Apparently $\mu = \mu_1 - \mu_2$. If these prerequisites are met, it is actually a test

$$H_0 : \mu = \Delta$$

$$H_A : \mu \neq \Delta$$

This actually translates the task into a one-sample t-test. First, from x_1, \dots, x_n calculate \bar{x}, s^2 . We reject the hypothesis H_0 at the level of significance α , if it is valid

$$\frac{|\bar{x} - \Delta|}{s} \sqrt{(n)} \geq t_{n-1}(1 - \alpha/2)$$

Hypothesis Tests for Two Samples

The paired t-test is used in situations where we have two quantities measured on each of the n objects. Individual objects can usually be considered independent, but measurements on the same object are not (e.g. properties before material treatment and after material treatment of the same samples - if the mean value of differences does not differ significantly from zero, it means that the effect of material processing is not statistically significant for the observed property).

- **Example:** We need to decide whether a car of certain marks, when the car's geometry is set correctly, are driven by both front tires just as fast. Therefore, 6 new cars were selected and after a certain time it was found out how many mm their right and left front tires went down. ²

car	1	2	3	4	5	6
right tire used by	1.8	1.0	2.2	0.9	1.5	1.6
left tire used by	1.5	1.1	2.0	1.1	1.4	1.4
difference	0.3	-0.1	0.2	-0.2	0.1	0.2

²Anděl, J., Matematická statistika, SNTL Praha, 1985

Hypothesis Tests for Two Samples

Solution: We will consider the differences in the used tires as the realization of independent random variables with a distribution $N(\mu, \sigma^2)$, where σ^2 is unknown. If both front tires roll at the same speed, the hypothesis applies $H_0 : \mu = 0$.

$$\frac{|\bar{x} - \Delta|}{s} \sqrt{(n)} = \frac{0.0833}{0.1941} \sqrt{(6)} = 1.0518$$

Since $1.051 < t_5(0.975) = 2.5706$, based on the data obtained, the hypothesis that both front tires run at the same speed cannot be rejected.

Hypothesis Tests for Two Samples

```
>> r=[1.8 1 2.2 0.9 1.5 1.6];  
>> l=[1.5 1.1 2 1.1 1.4 1.4];  
>> [h,p,ci,stats] = ttest(r,l)  
h =  
    0  
p =  
    0.3411  
ci =  
   -0.1203    0.2870  
stats =  
    struct with fields:  
  
    tstat: 1.0518  
    df: 5  
    sd: 0.1941
```

Hypothesis Tests for Two Samples

Two sample t-test

ttest2

Let us have two samples of the range n_1 a n_2 from distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ and sample averages \bar{x}_1, \bar{x}_2 , the variances σ_1^2 and σ_2^2 are unknown. We test the null hypothesis against the alternative

$$H_0 : \mu_1 = \mu_2$$

$$H_A : \mu_1 \neq \mu_2$$

(a) If $\sigma_1^2 = \sigma_2^2$ (which is verified by F - test) the test criterion is:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Hypothesis H_0 is reject at the α significance level if it is valid

$$|T| \geq t_{n_1+n_2-2}(1 - \alpha/2)$$

If the inequality does not apply, then we do not reject H_0 .

Hypothesis Tests for Two Samples

(b) If $\sigma_1^2 \neq \sigma_2^2$ the test criterion is:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Hypothesis H_0 is reject at the α significance level if it is valid

$$|T| \geq \frac{\frac{s_1^2}{n_1} t_{n_1-1}(1 - \alpha/2) + \frac{s_2^2}{n_2} t_{n_2-1}(1 - \alpha/2)}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

If the inequality does not apply, then we do not reject H_0 .

We use a two-sample t-test in cases where, for example, n patients are tested for the effect of A and other m patients are tested for the effect of B . The purpose of the experiment is to determine whether the effects of both drugs are the same.

Hypothesis Tests for Two Samples

- **Example:** 11 piglets of the same age were randomly divided into two groups. There were 6 of them in the first group and these piglets were fed diet A, in the second there were 5 of them and they were fed diet B. After six months they were calculated average daily gains in dkg. Results were obtained: ³

diet A		62	54	55	60	53	58
diet B		52	56	49	50	51	

It is necessary to find out whether both diets are equally effective.

³ Anděl, J., Matematická statistika, SNTL Praha, 1985

Hypothesis Tests for Two Samples

Solution: We will consider the average daily increments in the first group as sample $N(\mu_1, \sigma_1^2)$, increments in the second group will be considered as sample $N(\mu_2, \sigma_2^2)$ where $\sigma_1^2 = \sigma_2^2$ is unknown. The hypothesis that both diets are equally effective can be expressed as $H_0 : \mu_1 = \mu_2$.

$$T = \frac{57 - 51.6}{\sqrt{(6-1)12.8 + (5-1)7.3}} \sqrt{\frac{6 * 5(6 + 5 - 2)}{6 + 5}}$$

Since $2.7712 \geq t_9(0.975) = 2.2622$, we reject the hypothesis that both diets are equally effective.

Hypothesis Tests for Two Samples

```
>> A=[62 54 55 60 53 58];  
>> B=[52 56 49 50 51];  
>> [h,p,ci,stats] = ttest2(A,B,[],[],'equal')  
h =  
    1  
p =  
    0.0217  
ci =  
    0.9920    9.8080  
stats =  
    struct with fields:  
  
    tstat: 2.7712  
    df: 9  
    sd: 3.2180
```

Hypothesis Tests for Two Samples

Two sample F-test

vartest2

Let's have two random samples of n_1 and n_2 from the distribution $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ a sample variances s_1^2 a s_2^2 .

We test the null hypothesis against the alternative

$$\begin{aligned}H_0 : \sigma_1^2 &= \sigma_2^2 \\H_A : \sigma_1^2 &\neq \sigma_2^2\end{aligned}$$

Test criterion:

$$F = \max\left(\frac{s_1^2}{s_2^2}, \frac{s_2^2}{s_1^2}\right)$$

is constructed so that the larger of the variances is in the numerator. If $s_1^2 > s_2^2$ critical value holds

$$F_{n_1-1, n_2-1}(1 - \alpha/2)$$

is determined from the Fischer distribution tables for $n_1 - 1$ and $n_2 - 1$ degrees of freedom. Otherwise, the order of the degrees of freedom is reversed. If

$F > F_{n_1-1, n_2-1}(1 - \alpha/2)$, we reject H_0 and accept H_A .

Hypothesis Tests for Two Samples

- **Example:** Based on the data given in the previous example on the two-sample t test, it is to be decided whether both base files have the same variance. It is assumed that each of these basic files has a normal distribution: ⁴

diet A	62	54	55	60	53	58
diet B	52	56	49	50	51	

Solution:

$$\sigma_A^2 = 12.8$$

$$\sigma_B^2 = 7.3$$

$$1.7534 > F_{5,4}(0.975) = 9.3645$$

We do not reject the H_0 hypothesis of the agreement of variances.

⁴ Anděl, J., Matematická statistika, SNTL Praha, 1985

Hypothesis Tests for Two Samples

```
>> A=[62 54 55 60 53 58];  
>> B=[52 56 49 50 51];  
>> [h,p,ci,stats] = vartest2(A,B)  
h =  
    0  
p =  
    0.6063  
ci =  
    0.1872    12.9541  
stats =  
    struct with fields:  
  
    fstat: 1.7534  
    df1: 5  
    df2: 4
```

Confidence Intervals

- confidence interval - the interval in which the actual value of the given parameter Θ will be located with the specified probability $(1 - \alpha)$
- confidence interval limits L_1 and L_2
- confidence interval covers unknown estimated parameter Θ with preselected probability $(1 - \alpha)$ (confidence level)
- $P(L_1 < \Theta < L_2) = (1 - \alpha)$
- the probability that the actual value of Θ will be within the random limits of L_1 and L_2 is equal to $(1 - \alpha)$

The following applies to confidence intervals:

- the larger the sample range n , the narrower the confidence interval
- the more accurate the estimate and the smaller the variance, the narrower the confidence interval
- the higher the statistical certainty $(1 - \alpha)$, the wider the confidence interval
- the specified confidence intervals L_1 and L_2 are two-sided
- in practice, also one-sided interval is used, left-handed $< L_1, \infty)$, or right-sided $(-\infty, L_2 >$

Confidence intervals for μ with unknown σ^2

Confidence intervals for μ with unknown σ^2

normfit

Let x_1, \dots, x_n be a selection from $N(\mu, \sigma^2)$ and let $n > 1$, $\sigma^2 > 0$. The task is to construct a two-sided confidence interval for μ at the level $1 - \alpha$. Because quantity

$$\frac{\bar{x} - \mu}{s} \sqrt{n}$$

has Student's t -distribution, the probability can be expressed

$$P(-t_{n-1}(1 - \alpha/2) \leq \frac{\bar{x} - \mu}{s} \sqrt{n} \leq t_{n-1}(1 - \alpha/2)) = 1 - \alpha$$

where $t_{n-1}(1 - \alpha/2)$ je $100(1 - \alpha/2)\%$ quantile of Student's distribution with $n - 1$ degrees of freedom. After adjustment, we get $100(1 - \alpha/2)\%$ the mean interval of the value in the form

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

For larger sample ranges $n > 30$ you can use the quantile $t(1 - \alpha/2)$ quantile of the normalized normal distribution

Confidence intervals for μ with unknown σ^2

Random variable

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

has χ^2 distribution with $(n-1)$ degrees of freedom. For $100(1-\alpha/2)\%$ two-sided variance confidence interval σ^2 then holds

$$\frac{(n-1)s^2}{\chi_{(n-1)}^2(1-\alpha/2)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{(n-1)}^2(\alpha/2)}$$

where $\chi_{(n-1)}^2(1-\alpha/2)$ is upper and $\chi_{(n-1)}^2(\alpha/2)$ lower quantile of the distribution χ^2 .

Confidence intervals for μ with unknown σ^2

- **Example:** Calculate the interval estimates ($\alpha = 0.05$) of the mean value and standard deviation for $n = 200$ random numbers generated from the normal distribution with $\mu = 50$ and $\sigma = 5$.

Solution:

```
>> x=normrnd(50,5,1,200)
x = Columns 1 through 6
    52.5762    51.3070    45.2926    49.1883    49.2697    47.3399
    ...
Columns 195 through 200
    53.4958    51.3482    52.4714    42.5844    44.8987    47.7650

>> [muHat,sigmaHat,muCI,sigmaCI] = normfit(x,0.05)
muHat =
    49.4929
sigmaHat =
    4.9028
muCI =
    48.8092    50.1765
sigmaCI =
    4.4648    5.4369
```

List of functions

Command	Description
» <code>ttest(x)</code>	one sample t-test
» <code>vartest(x)</code>	one sample variance test
» <code>ttest(x,y)</code>	paired t-test
» <code>ttest2(x,y)</code>	two sample t-test
» <code>vartest2(x,y)</code>	two sample F-test
» <code>normfit</code>	estimates of normal distribution parameter

Examples for practice

Examples for practice

- 1 Example:** In the data file `data_01.txt` the results of the utility property measurements for 6 types of samples are obtained. Each type of sample was measured 10 times. Calculate the interval estimates of the mean value for each sample and plot on a graph.

Solution