Matlab: Statistics Toolbox

1. Overview
2. In-class exercise

MATLAB: Statistics Toolbox

Overview

Statistics Toolbox Capabilities

- Descriptive statistics
- Statistical visualization
- Probability distributions
- Hypothesis tests
- Linear models
- Nonlinear models
- Multivariate statistics
- Statistical process control
- Design of experiments
- Hidden Markov models

Histograms

```
>> y = [1 3 5 8 2 4 6 7 8 3 2 9 4 3 6 7 4 1
     5 3 5 8 9 6 2 4 6 1 5 6 9 8 7 5 3 4 5 2 9
     6 5 9 4 1 6 7 8 5 4 2 9 6 7 9 2 5 3 1 9 6
     8 4 3 6 7 9 1 3 4 7 5 2 9 8 5 7 4 5 4 3 6
     7 9 3 1 6 9 5 6 7 3 2 1 5 7 8 5 3 1 9 7 5
     3 4 7 9 1]';
```

>> mean(y) → ans = 5.1589

>> var(y) → ans = 6.1726

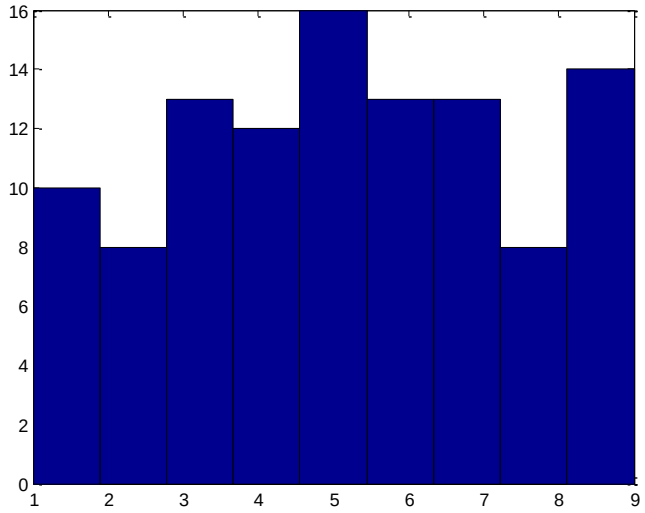
>> std(y) → ans = 2.4845

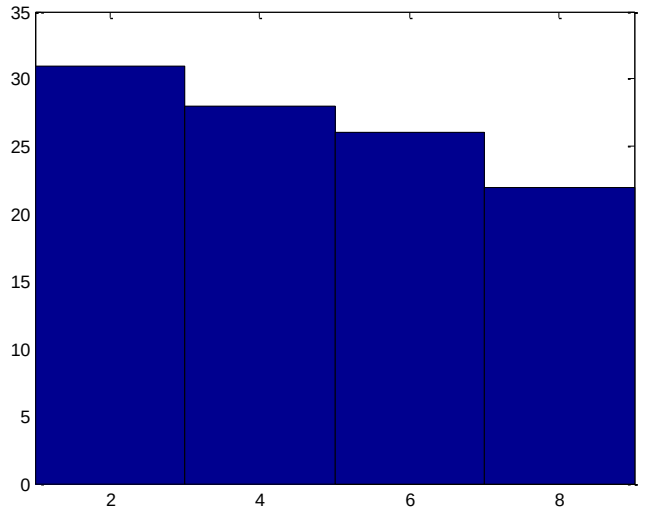
>> hist(y,9) → histogram plot with 9 bins

>> n = hist(y,9) → store result in vector n

>> x = [2 4 6 8]'

>> n = hist(y,x) → create histogram with bin centers specified by vector x





UMASS.

Permutations and Combinations

>> perms([2 4 6]) → all possible permutations of 2, 4, 6

```
6   4   2
6   2   4
4   6   2
4   2   6
2   4   6
2   6   4
```

>> randperm(6) → returns one possible permutation of 1-6

```
5   1   2   3   4   6
```

>> nchoosek(5,4) → number of combinations of 5 things taken 4 at a time without repetitions

```
ans = 5
```

>> nchoosek(2:2:10,4) → all possible combinations of 2, 4, 6, 8, 10 taken 4 at a time without repetitions

```
2   4   6   8
2   4   6  10
2   4   8  10
2   6   8  10
4   6   8  10
```

Probability Distributions

- 21 continuous distributions for data analysis
  - » Includes normal distribution
- 6 continuous distributions for statistics
  - » Includes chi-square and t distributions
- 8 discrete distributions
  - » Includes binomial and Poisson distributions
- Each distribution has functions for:
  - » pdf — Probability density function
  - » cdf — Cumulative distribution function
  - » inv — Inverse cumulative distribution
  - » functionsstat — Distribution statistics function
  - » fit — Distribution fitting function
  - » like — Negative log-likelihood function
  - » rnd — Random number generator

Normal Distribution Functions

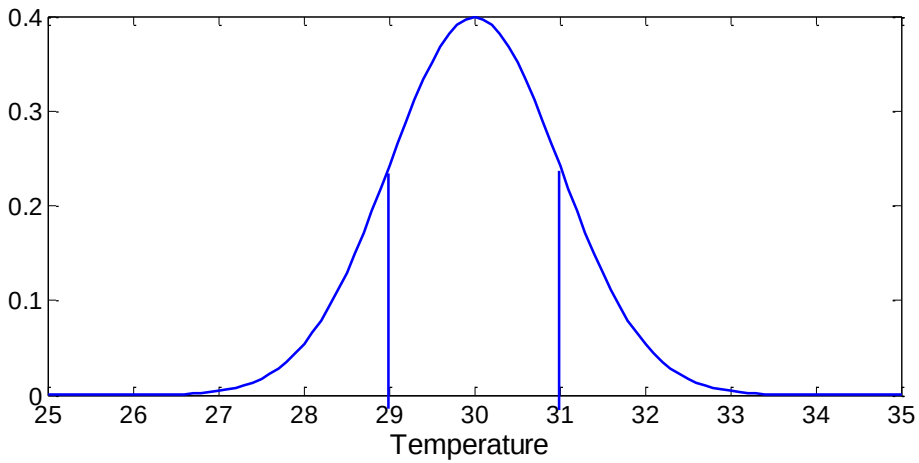
- normpdf – probability distribution function
- normcdf – cumulative distribution function
- norminv – inverse cumulative distribution function
- normstat – mean and variance
- normfit – parameter estimates and confidence intervals for normally distributed data
- normlike – negative log-likelihood for maximum likelihood estimation
- normrnd – random numbers from normal distribution

Normal Distribution Examples

- Normal distribution: normpdf(x,mu,sigma)
  - » normpdf(8,10,2) → ans = 0.1210
  - » normpdf(9,10,2) → ans = 0.1760
  - » normpdf(8,10,4) → ans = 0.0880
- Normal cumulative distribution: normcdf(x,mu,sigma)
  - » normcdf(8,10,2) → ans = 0.1587
  - » normcdf(12,10,2) → ans = 0.8413
- Inverse normal cumulative distribution: norminv(p,mu,sigma)
  - » norminv([0.025 0.975],10,2) → ans = 6.0801 13.9199
- Random number from normal distribution: normrnd(mu,sigma,v)
  - » normrnd(10,2,[1 5]) → ans = 9.1349 6.6688 10.2507 10.5754 7.7071

Normal Distribution Example

- The temperature of a bioreactor follows a normal distribution with an average temperature of 30°C and a standard deviation of 1°C. What percentage of the reactor operating time will the temperature be within +/-0.5°C of the average?
- Calculate probability at 29.5°C and 30.5°C, then calculate the difference:
  - » p=normcdf([29.5 30.5],30,1)
  - p = [0.3085 0.6915]
  - » p(2) – p(1)
  - 0.3829



- The reactor temperature will be within +/- 0.5°C of the average ~38% of the operating time

UMASS.

Confidence Intervals

>> [muhat,sigmahat,muci,sigmaci] = normfit(data,alpha)

- data: vector or matrix of data
- alpha: confidence level = 1-alpha
- muhat: estimated mean
- sigmahat: estimated standard deviation
- muci: confidence interval on the mean
- sigmaci: confidence interval on the standard deviation

>> [muhat,sigmahat,muci,sigmaci] = normfit([1.25 1.36 1.22 1.19 1.33 1.12 1.27 1.27 1.31 1.26],0.05)

muhat = 1.2580

sigmahat = 0.0697

muci = 1.2081

1.3079

sigmaci = 0.0480

0.1273

UMASS.

MATLAB: Statistics Toolbox

In-class Exercise

Membrane Quality

- A membrane manufacturer sells membranes with three different pore sizes
- The average pore size of the membranes supposedly meet the following specifications (in microns):
  - » Small pore membranes: $\mu = 50$, $s = 2.5$
  - » Medium pore membranes: $\mu = 75$, $s = 5$
  - » Large pore membranes: $\mu = 125$, $s = 10$
- The Excel spreadsheet membranes.xls contains the average pore size measured for 25 membranes of each pore size
- Determine if the specifications are satisfied to a 95% confidence level