1 Examples from probability

1.1 Example

The data file is stored in the file `dataDesc1.txt`.

Determine: (i) average, (ii) variance, (iii) standard deviation, (iv) minimum, maximum, (v) mode, (vi) median.

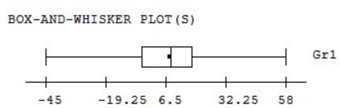
Results

(i) *8.2;* (ii) *335.67 (332.32);* (iii) *18.32 (19.23);* (iv) *-45, 58* (v) *6;* (vi) *8.5*

1.2 Example

For the data in data file `dataDesc1.txt` determine the Box and Whiskers graph and explain what does it represent.

Results



1.3 Example

The data file is stored in the file `dataDesc2.txt`. Determine the sample covariance matrix.

Results

```
COVARIANCE MATRIX (POPULATION)

            x1      x2      x3
   x1  {   3.94    0.04   -0.96 }
   x2  {   0.04    4.47    0.37 }
   x3  {  -0.96    0.37    4.52 }

COVARIANCE MATRIX (SAMPLE)

            x1      x2      x3
   x1  {   3.98    0.04   -0.97 }
   x2  {   0.04    4.52    0.37 }
   x3  {  -0.97    0.37    4.56 }
```

1.4 Example

A. Consider the experiment A "throwing a dice".

- Write the sample space for this experiment.

- What is the probability of the event $J = J_1 \cup J_2$ where J_1 ="odd number" and J_2 ="less then 5"

B. Let B be a new conditional experiment "throwing a dice on condition J_3" where J_3 ="greater then 3".

- Write the sample space for this experiment.
- Write the space of events for this experiment.
- Write the probability function associated with this experiment.
- Write the probability of obtaining the result 1 or 2.
- Write the probability of obtaining the result 5 or 6.

Results

A

- $\{1, 2, 3, 4, 5, 6\}$
- $P(J) = \frac{2}{6} = \frac{1}{3}$

B

- $\{4, 5, 6\}$
- $\{\emptyset, \{4\}, \{5\}, \{6\}, \{4, 5\}, \{4, 6\}, \{5, 6\}, \{4, 5, 6\}\}$
- 4,5,6| 1/3, 1/3, 1/3
- P=0
- P=2/3

1.5 Example

Random variable X has density function

$$f(x) = -|x - 1| + 1$$

for $x \in (0, 2)$. It is zero for $x \leq 0$ and one for $x \geq 2$.

- Verify that it is a density function

- Determine its expectation and variance.

Results

- *int =1*
- *exp = 1 (symmetry), variance = 1/6*

1.6 Example

Probability function of X is given by the table

x	1	2	3	4
$f(x)$	0.2	$5k$	$0.5-3k$	k

Determine the constant k.

Results

$k = 0.1$

1.7 Example

Random vector $[X, Y]$ has joint probability function given by the table

$x \backslash y$	1	2	3
1	0.2	0.1	0.2
2	0.1	0.3	0.1

Determine the marginal $f(y)$ and the conditional probability function $f(y|x)$.

Results

f(y): 0.3, 0.4, 0.3

f(x/y)

0.4, 0.2, 0.4

0.2, 0.6, 0.2

1.8 Example

Write probability function of the binomial distribution of random variable X with parameters p and n.

For $p = 0.3$ and $n = 5$ determine the probability $P\left(X \leq 3\right)$

Results

$f\left(x\right) = \binom{n}{k} p^k \left(1 - p\right)^{n-k}$, $k = 0, 1, ..., n$

P=0.969

2 Examples from theory of statistics

2.1 Example

The population is

$$\{3, 2, 1, 2, 1, 3, 2, 2, 2, 1, 3, 3, 1, 2, 3, 2, 2, 3, 1, 3\}$$

We made a sample realization

$$\{3, 1, 3, 2, 2\}$$

a) Write probability function of the population

b) Determine population expectation and sample average.

c) What is the point estimate of the expectation?

Results

a)

1 2 3

(5 8 7)/20 = 0.25 0.4 0.35

b) 2.1

c) 2.2

2.2 Example

A population is described by normal distribution with expectation 10 and variance 10. We are going to make a sample of the length 10 and construct sample average. How will the sample average be distributed?

Results

Normally with expectation 10 and variance 1 - why?

2.3 Example

Consider an experiment of flipping a coin with sides 1 and 2. We performed 10 flips and obtained the data

$$\{1, 1, 0, 1, 0, 0, 0, 1, 0, 0\}.$$

Write evolution of the point estimates of the probabilities $P(1)$ and $P(2)$.

Results

1, 1, 0.666, 0.75, 0.6, 0.5, 0.428,0.5, 0.444, 0.4

2.4 Example

For a point estimate of the parameter θ we constructed the statistics T

$$T = \left(\prod_{i=1}^{n} x_i^2 \right)^{\frac{1}{n}}$$

Determine the value of the point estimate $\hat{\theta}$ if the sample realization is

$$\{1, 2, 1, 1, 2, 3, 1, 1\}$$

Result: $\hat{\theta} = 1.364$

2.5 Example

We have defined a statistics T for estimation of unknown parameter θ. It has uniform distribution

$$f\,(T) = \frac{1}{8} \text{ for } \theta \in (0,\, 8)$$

Determine (i) both-sided, (ii) left-sided and (iii) right-sided confidence interval for the estimate $\hat{\theta}$ on the level of significance $\alpha = 0.1$.

Results

a) $(0.4,\, 0.76)$

b) $(0.8,\, \infty)$

c) $(-\infty,\, 0.72)$

2.6 Example

For given data

$$x = \{1, 3, 5, 8, 10, 12, 15, 17\}$$

$$y = \{2, 8, 20, 55, 250, 800, 1501, 2521\}$$

determine (i) linear, (ii) quadratic and (iii) cubic regression.

a) Write the resulting regression equations.

b) On the basis of p-value P(>F) determine which regression is the best one.

Result:

a)

(i) $y = 142.2x - 617.7$;

(ii) $y = 16.8x^2 - 159x + 259.2$;

(iii) $y = 0.82x^3 - 5.5x^2 + 2.7x + 14.8$;

b) Quadratic is the best regression. The smaller pv of the validity of the model.

2.7 Example

The production of certain factory in selected years is listed in the table

year of production	1997	2000	2007	2012	2018	2019	2020	2021
prod. (in thousands)	2900	3000	3200	3300	3500	3600	3550	3700

a) Verify linear course of the production and write the regression line.

b) If the linearity is verified, estimate the production in the year 2025.

Result

a) $pv = 3.14 \cdot 10^{-6}$ *- linear is suitable*

$$y = 30.6x - 58242.9$$

b) prod(2025) $\rightarrow$ *3722.1 thousands*

2.8 Example

We suppose that traffic intensity (y) at certain point of traffic micro-region linearly depend on intensities at three other points P_1, P_2 and P_3. Measurements gave us data from the following table

y (intensity)	155	210	132	201	254	169	212	179
P_1	234	218	219	247	162	357	296	361
P_2	252	143	121	163	246	110	121	102
P_3	41	29	49	34	45	41	51	58

Write equation of linear regression and decide if the regression is suitable.

Results

$$y = -0.107P_1 + 0.103P - 0.252P_3 + 211.671$$

$pv = 0.89$. *On the level* 5% *the regression is not suitable.*

3 Examples (i) from hypotheses testing

3.1 Example

On a road with recommended speed 80 km/h we monitored the speeds of passing cars and obtained data (in km/h)

$$\{83\ 93\ 78\ 82\ 76\ 95\ 80\ 89\ 94\ 72\ 81\ 87\ 81\ 85\ 76\}$$

a) Use W/S test for verifying normality of the speeds.

b) Test the hypothesis H0 that the average speed is 80 km/h. Test on the level 0.05.

c) Test the hypothesis H0 that the average speed is less then 80 km/h. The level is 0.05;

Results

a) Normality - OK pv = 0.48

b) H0 is not rejected, pv = 0.075

c) H0 is rejected, pv = 0.037

3.2 Example

On a highway with recommended speed 100 km/h we monitored the speeds of cars going into the town and from the town. We obtained data From (from the town) and Into (into the town)

a) At the level 0.01 test the hypothesis H0: "Both the average speeds are equal" against HA: "From the town the cars go more quickly".

From={115 98 111 105 129 96 81 108 114 118}
Into={121 114 82 95 134 99 105 107 112 76}

b) Determine both data averages and explain the result.

Results

a) $pv = 0.34$ - H0 is not rejected

b) $\mu_F = 107.5$ and $\mu_I = 104.4$

3.3 Example

During a check of the front lights of cars we have measured the data 'xL' (left light) and 'xR' (right light). The values are deviations (in cm) above (positive) and below (negative) the optimal level. At the level 0.1 test the hypothesis H0: "Are equal" against HA: "Left lights are higher than right". The data measured are

xR = {-3 5 16 9 -8 -2 23 5 -6 -3}
xL = {-5 -12 22 -3 -9 1 -1 2 -13 -5}

Results

$pv = 0.96$ *do not reject.*

3.4 Example

The accuracy of setting of certain machine can be verified according to the variance of its products. If the variance is greater then the level 80, it is necessary to perform new setting. The following data sample has been measured

x = {258 215 225 229 235 228 231 225 242 222}

On the level 0.05 test if it is necessary to set the machine.

Results

TH for variance, right-sided, pv=0.065

3.5 Example

Consumption of cars is measured by two methods A and B. The same car has been subdued to measuring by both methods. The results are

A = {4.3 6.2 6.8 6.7 5.1 5.0}
B = {4.8 5.3 5.2 5.8 5.3 5.9}

On the level 0.05 test equality of both methods if the consumption is assumed to be normally distributed.

Results

TH for 2 expectations, paired samples, equal variances, both-sided, pv=0.48, H0 is not rejected.

3.6 Example

During the police check, the front tyre pressure has been checked. For six selected cars the following pressures have been registered

Lt= {2.6 2.2 1.5 1.6 2.2 2.1}
Rt= {2.2 2.5 1.8 1.8 2.5 2.0}

a) Test H0: “The pressures are equal” against HA: “they are different”.

b) Test H0: “The pressures are equal” against HA: “Left tyres have higher pressure”.

Test at the level 0.05; the normality is assumed.

Results

a) TH for 2 expectations, paired, both-sided, pv=0.44

b) TH for 2 expectations, paired, right-sided, pv=0.78

3.7 Example

At the intersection, we recorded the number of cars going straight (S), turning left (L) and right (R). We obtained the following data $x_L = 39$, $xR = 46$ and $x_S = 62$. On the significance level 0.1 test the hypothesis H0: “The ratio of cars going straight ahead is equal to those who are turning” against HA: “There is more cars that go straight then those who turn.”

Results

pS=62; pT=85;

TH for 2 proportions, right-sided, pv=0.99

3.8 Example

A factory produces steel rods of the nominal length 50 cm. We have randomly chosen 10 rods with the lengths

$$x = \{50\ 51\ 50\ 50\ 51\ 50\ 52\ 50\ 51\ 51\}$$

Test H0: “The produced rods have the nominal length” against HA: “The length of produced rods is greater than the nominal length”.

Test on the level 0.01.

Results

TH expectation, right-sided, pv=0.012 - do not reject H0

4 Examples (ii) from hypotheses testing

4.1 Example

We monitor speeds of three race cars. Randomly, we measured their speeds and got the following data

x1= {231 158 223 197 185 194}
x2= {285 263 238 199 221 236}
x3= {241 222 231 295 208 201}
x4= {254 267 241 224 278 200}

At the level 0.05 test the equality of the average speeds of the cars. The variances are assumed to be equal. Verify the entry requirements.

Results

Bartlett: pv=0.96

ANOVA, pv=0.056 - equality is not rejected

4.2 Example

From two classes 1 and 2 several children were tested how long they need to solve an example from math. The following data (in minutes) have been measured

x1={8 14 9 6 4 18 7 9 4 7 9 16 9 3}
x2={15 4 15 9 14 16 13 7 26 14 16 4 18 14 18 13 7 17 4 15 13}

The populations from which the data have been measured cannot be assumed normal. Using Man-Whitney test if in both classes the children solve the task with the same speed. Test on the level 0.05

Results

TH two medians, Mann Whitney, independent, both-sided, pv=0.04 - H0 is rejected.

4.3 Example

We are interested whether the children from a chosen class are improving in math. To that end, we selected 18 children last year and had them solve certain example. This year we have asked the same children to solve another example similarly difficult. The following data (the time of solving in minutes and in the same order) have been measured

x1={12 10 14 5 6 9 7 10 11 12 9 8 5 3 8 9 15 7}
x2={10 12 14 8 7 7 9 12 9 15 9 8 6 7 8 12 18 6}

The populations from which the data have been measured cannot be assumed normal. On the level 0.05 test if the children individually improve. For the test use a) Wilcoxon, b) sign tests. Comment the results.

Results

TH two medians,

a) Wilcoxon, paired, left-sided pv=0.025 - H0 is rejected

b) Sign-test, left-sided, pv=0.09 - H0 is not rejected (is less strict)

4.4 Example

The association between weight and height in children was investigated. The following data sample was obtained (frequencies of combinations of both these features measured values)

weight (kg) \ height (m)	less than 1.2	between 1.2 and 1.5	more than 1.5
less than 12	19	42	43
between 20 and 30	44	69	45
more than 30	42	39	31

At the level 0.05 test the hypothesis that the color of eyes and hair are independent.

Results

TH Chi2 test of independence, pv=0.0139 - H0 is rejected
Setting the data in Statext: {} {} {}

4.5 Example

Tree supervisors are evaluating functionality of five fast tea services. Each inspector evaluates each service with marks 1,2,..,10 (10 being the best one). Test if the quality of the services is equal. The data are in the table

supervisor \ service	1	2	3	4	5
1	6	8	4	8	9
2	3	4	2	5	6
3	7	9	9	6	9

Results

TH Friedman test, pv=0.157. Equality of treatments of individual services is not rejected
Setting: {} {} {}, data set for subject (each data set is measured by one subject)

4.6 Example

In certain crossroads we have written down numbers of passing cars. The measured data are

$$d = \{15\ 10\ 20\ 35\ 10\ 50\}$$
$$x = \{71\ 56\ 98\ 121\ 44\ 271\}$$

where 'd' is length of monitoring and 'x' - number of cars.

At the level 0.05 test the hypothesis that the cars go uniformly.

Hint: The expected frequencies must be constructed manually. Make sure the totals of the observed and expected frequencies are the same. $\left[nd = \sum d;\ nx = \sum x;\ E = \frac{d}{nd} nx\right]$

Results

E={70.82 47.21 94.43 165.25 47.21 236.08}

pv = 0.0019 - the uniformity is rejected

4.7 Example

A factory produces some products whose weight must be constant. It uses four machines for the production. A sample of products has been taken from all machines to test equality of the product weights. The measured values are

x1={39.4 34.8 35.6 35.1 35.8}
x2={34.4 34.2 35.1 31.1 32.5 33.8}
x3={36.2 35.1 38.2 36.3 35.8 35.6 35.2}
x4={37.1 34.3 38.6 34.5 36.4 36.1}

Test the equality on the level 0.01, if the normality cannot be assumed .

Results

Kruskal-Wallis, pv=0.0127

4.8 Example

Athletes in javelin throwing in certain sport club were tested with respect to their performance before and after an intensive training. They threw the javelin once before training and once after it. The measured lengths are

x1={68 81 69 72 66 91 98 89 75 68 69 75 72 83 88 79 88 76 81 85}
x2={79 82 70 75 68 99 95 84 81 82 81 79 74 85 92 91 85 82 83 73} ;

On the level 0.05 and 0.01 test H0: An intense training close before the race does not help. The normality cannot be assumed.

Results

Wilcoxon, left-sided, pv=0.014; on 0.05 - reject; on 0.01 - do not reject