1 Examples solved in Scilab

Examples from the textbook without Scilab program - the last two digits in the examples match the numbers in the textbook.

1.1 Regression analysis

1.1.1 Example

In a factory, dependence of the overall costs y (in thousands of Kč) on the production x have been investigated. The following data have been measured

$x =$ [532 297 378 121 519 613 592 497]
$y =$ [48 32 42 27 45 51 53 48]

a) Using linear regression estimate the costs for the production of 1000 products

b) For which production the costs would be equal to $ 100 000.

Results

line: $y = 0.053x + 19.51$

cost_ 1000 = 73.02

prod_ 100 = 1504.09

1.1.2 Example

A harmful substance leaked into the container with water. Neutralizing agent has been applied and the concentration of the harmful substance has been measured at time instants x. The measured concentrations y are

x= [5 12 20 26 29 38 65 126]
$y =$ [19 17 18 17 17 15 14 7]

Compute the correlation coefficient of linear regression and conclude about its suitability. If suitable, estimate when the concentration will be zero.

Results

Correlation coefficient r = -0.9832531

Line $y = -0.095x + 19.305$

Zero concentration will be at x = 203.58

1.1.3 • Example

At certain process we have measured the data

$x =$ [5 12 20 26 29 38 40 45]
$y =$ [9 7 12 12 27 35 44 76]

Perform the polynomial regression of the order 3 and the exponential regression. Using p-value of the regression decide which type of regression is better.

Results

$pv_{pol} = 0.007$

$pv_{exp} = 0.016$

Polynomial is better.

1.1.4 • Example

At certain process we have measured the data

$x1$ = [15 12 11 9 9 8 5 3]
$x2$ = [3 9 5 11 28 14 32 58]
y = [9 7 22 12 27 31 44 36]

Perform multivariate linear regression and test its suitability using p-value from the ANOVA table.

Results

pv = 0.052

On the level 0.05 is the regression OK.

1.2 Confidence intervals

1.2.1 Example

Assume, that the height of children in the age 10 has normal distribution with the variance 38. Determine the interval α-I, in which the true height will be if we have measured the data sample of the length 12 with the average 127.3. Compute on the level $\alpha = 0.01$.

Results

CI = (122.72, 131.88)

1.2.2 Example

Assume, that the height of children in the age 10 has normal distribution. Determine the interval α-I, in which the true height will be if we have measured the data sample of the length 12 with the average 127.3 and variance 38. Test on the level $\alpha = 0.01$.

Compare with the previous result and justify.

Results

CI = (121.77, 132.83)

1.2.3 • Example

To learn the accuracy of a method for measuring the volume of manganese in the steel, we performed independent measurements of several variances. We would like to know the border for which it holds that only 5% of possibly measured variances will be greater than this border. The measured values are

$x = [4.3\ 2.9\ 5.1\ 3.3\ 2.7\ 4.8\ 3.6]$

Results

The border of variance is 3.2

Sorry, 2.4 is missing

1.3 Parametric tests

1.3.1 Example

At the motorway with recommended speed 80 km/h we monitored the speeds of passing cars and obtained data presented in the following table of values and frequencies

x	0-78	79	80	81	82	83-∞
n	543	32	45	15	8	2

Test the hypothesis H0 that only 3% of drivers exceed the speed 80 km/h. Test on the level 0.05.

Results

$pv = 0.096$

We do not reject.

1.3.2 Example

From a set of steel rods with equal nominal length 6.2 cm, we have selected random choice with the lengths x. The producer guarantees that the standard deviation of the lengths is 0.8 cm. At the significance level 0.05 test the assertion of the producer that the produced rods have the nominal length if the measured lengths are

$x=$ [6.2 7.5 6.9 8.9 6.4 7.1]

Compare the result with the situation when the variance is not known.

Results

$pv_z =$ *0.003,* $pv_t = 0.059$

1.3.3 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 28, it is necessary to perform new setting. A data sample of values, from which the variance is to be determined is as follows

x = [102 113 108 119 114 102 115 119 99 117 108 101]

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

Results

pv = 0.03

On the level 0.05 reject H0: the machine is OK

1.3.4 • Example

Solidity of materials is verified by two methods A and B. The same material has been subdued testing by both methods. The results are

A = [20.1 19.6 20.0 19.9 20.1]
B = [20.9 20.1 20.6 20.5 20.7 20.5]

On the level 0.05 test equality of both methods if the variability of methods is assumed to be equal.

Results

pv = 0.002

Equality is rejected.

1.3.5 • Example

We are going to test if the tire removal on left and right sides of the front wheels of cars is equal. The measured values are

xL= [1.8 1.0 2.2 0.9 1.5]
xP= [1.5 1.1 2.0 1.1 1.4]

Test at the level 0.05.

Results

pv= 0.55, are the same.

1.3.6 Example

On a highway with a recommended speed limit of 80 km/h, the speeds of vehicles traveling toward the city (xT) and away from the city (xF) were monitored. The data measured are

xT = [95 88 71 82 69 75 78 67 77 82 79]
xF = [81 69 75 91 77 76 88 68 91 74 92]

At the level 0.05 test the hypothesis H0: To the town the cars go faster.

Results

1. test of variances: pv = 0.8 - variances are equal.

2. test of speeds: pv = 0.322 - to town are higher is not rejected

1.3.7 Example

During a check of the front lights of cars we have measured the data xL (left light) and xR (right light).

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

The values are distances (in cm) above (positive) and below (negative) of the real level with respect to the optimal level. At the level 0.1 test if

a) the light levels at each car are the same,

b) the left lights are higher then right.

c) the left lights are lower then right.

Results

a) pv = 0.075; On the level 0.1 we reject equality.

b) pv = 0.03; Left are higher is rejected.

c) pv = 0.96; Right are higher is not rejected.

1.3.8 • Example

At the intersection, we recorded the number of cars going straight (S), turning left (L), and turning right (R). The measured data are $xS = 62$, xL =39 and $xR = 46$. At level 0.1, test the claim that the ratio of cars

a) going straight is equal to those that are turning,

b) going straight are less than those turning.

Results

a) pv = 0.0073 - equality is rejected

b) pv = 0.0036 - "straight is less" is rejected

1.4 ANOVA

1.4.1 Example

We monitor three machines. Randomly, we measure their productions per hour

$x1$= [53 55 49 58 52 61 56 55]
$x2$= [49 56 52 45 51 56 44 51]
$x3$= [52 53 52 54 55 53 53 52]

At the level 0.05 test the equality of their production.

Results

pv= 0.054

1.4.2 Example

For one month in the years 1999 2000 2001 2002 2003, we monitored number of accidents at four crossroads. The results are as follows

C1 C2 C3 C4

1999: [3 5 2 1 3]

2000: [6 2 5 3 4]

2001: [3 2 1 1 2]

2002: [4 1 1 2 2]

2003: [4 2 5 5 6]

At the level 0.01 test hypothesis H0: The average number of accidents is equal at all monitored crossroads.

Results

One-way anova pv = 0.021

Two-ways anova Equality in crossroads pv = 0.0195

Equality in time pv = 0.275

1.4.3 • Example

The factory manufactures certain products whose weight must be constant. It uses four machines for production. Samples of products were taken from all machines to test the uniformity of product weight. 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=[30.2 35.1 34.2 36.3 30.8 35.6 35.2]
x4=[39.1 34.3 38.6 34.5 36.4 36.1]

Test the equality of the product weights on all four machines.

Results

pv = 0.036

1.5 Nonparametric tests

1.5.1 Example

At a crossroads we have written down numbers of passing cars. The lengths of monitoring were $d = [15\ 10\ 20\ 35\ 10\ 50]$ and the measured numbers $x = [71\ 56\ 98\ 121\ 44\ 271]$. At the level 0.05 test the hypothesis that the cars go uniformly (the same number per time unit).

Results

pv = 0.002

1.5.2 Example

The following data are frequencies (f) of incidents at certain big factory at time intervals (i)

i: 8-10h. 10-12h. 12-13h. 13-17h.

f: 2 7 1 16

At the level 0.05 test the hypothesis that the accidents occur uniformly.

Results

pv= 0.13

1.5.3 • Example

A connection between color of eyes and hair has been investigated. In a collected data sample we obtained the following frequencies

eyes\hair	light	brown	dark
blue	90	75	55
gray	96	136	88
brown	108	135	119

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

Results

pv= 0.017

1.5.4 Example

Two operators O1 and O2 alternate regularly in production of certain articles. The produced articles are checked for quality (either 1 or 2). The following data have been measured

O1: 1 2 1 1 2 2 2 1 2 1 1 1 2 1 2 2 2 1 2 1 2

O2: 2 2 1 2 1 1 2 2 2 1 2 2 1 2 1 2 2 2 1 1 2

At the level 0.05 test the assertion that both the operators are with respect to the production quality independent.

Results

pv= 0.466

1.5.5 • Example

Two doctors recommend curing a cold with two different methods. The results (number of days of the treatment) are $x1$ and $x2$. Test equality of the methods.

$x1$=[5 8 7 8 4 5 5 6 9 3 5 8 6 8 7 5 8 5 7 5 6 8 4 7 7 5 6]
$x2$=[3 4 9 5 4 9 9 8 3 3 5 3 6 4 5 6 2 2 3 4 2 3]

Results

pv=0.006

1.5.6 Example

Eight sportsmen in a certain sports club were tested with respect to their performance. All of them threw a javelin once and then they were subdued to intensive training. Then they threw once more. The measured lengths were $x1$ and $x2$. The zero hypothesis is that one day of training is not enough to improve their performance. Test on the level 0.05.

x1=[68 81 69 72 66 91 98 89 75 68 69 75 72 83 88 79 88 76 81 85 79]
x2=[79 62 70 75 68 81 85 94 71 62 81 70 74 85 82 91 85 82 83 73 71]

Results

pv = 0.238

1.5.7 Example

Test if mice and stags have equally long front legs. The measured values are

x1=[135 123 3.1 2.5 98 124 131 3.4 2.8 128 154 135 2.9 137 2.7 3.0 3.2 131 2.8 148]
x2=[136 121 2.9 2.6 101 121 130 3.5 2.9 126 162 141 2.8 142 2.9 2.8 3.0 132 3.1 151]

Results

pv=0.404

1.5.8 • Example

Tree inspectors are to evaluate functionality of five fast food stands. Each inspector evaluates each stand. The result is at the table (rows correspond to inspectors, columns columns to stands). Evaluation marks are 1,2,..,10. The mark 10 is the best one. Test if the quality of the stands is equal.

Table

10 8 3 9 7
8 7 5 9 10
8 9 5 7 6

Results

pv= 0.155

1.5.9 • Example

Let x denote the length (in centimeters), of a certain fish species. We obtained the data set

x=[5.0 3.9 8.2 7.5 2.8 4.1 5.4 2.6 1.7 4.3 7.4 4.1 5.2 ...
9.3 2.7 3.4 5.9 4.3 9.4 8.2 4.8 3.3 4.7 5.3 4.2 4.0].

Can we conclude that the median length of the fish species differs significantly from 4.1 centimeters?

Results

pv = 0.052