Examples from Statistics

Regression ———————————

```
Example 1
In a factory, dependence of the overall costs 'n'
(in thousands of Kč) on the production 'p' has been
investigated. The following data have been measured
p = [532 297 378 121 519 613 592 497];
n = [ 48  32  42  27  45  51  53  48];
a) Using linear regression estimate the costs for
   the production 1000 products
b) For which produciton the costs would be equal
   to $ 100 000.

Results
 param  = 0.0535136, 19.510024
 cost_1000  = 73.023636
 prod_100  = 1504.1028


Example 3
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
xi = [5 12 20 26 29 38 65 126];
yi = [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
 Correletion coefficient
  r  = -0.9832531
 Parameters
  p  = -0.0948295, 19.305033
 Zero concentration at x_zero
  x_zero  = 203.57626


Example 5
At certain process we have measured the data
xi = [5 12 20 26 29 38 40 45];
yi = [9 7 12 12 27 35 44 76];
Perform the polynomial regression of the order 'k' and
the exponential regression. Using prediction errors
decide which type of regression is better.
```

```
k  = 3;

Results
 SEp  = 0.0365208
 SEe  = 0.1113473
 Polynomial is better.
```

```
Example 6
At certain process we have measured the data
x1i = [15 12 11  9  9  8  5  3]';
x2i = [3   9  5 11 28 14 32 58]';
yi = [9 7 22 12 27 31 44 36]';
Perform multivariate linear regression and test its
suitability. Use test for independence of residuals.

Results
 SE  = 0.2799241
 Regression looks good.
```

Parametric tests ———————————————-

```
Example 7
Assume, that the height of children in the age 10 has normal distribution
with the variance "sig2". Determine the interval al-I, in which the true
height will be it we have measured the data sample of the length "n" with
the average "mx".
The interval determine as a) both-sided,  b) rihgt-sided.
sig2 = 38;
n    = 12;
mx   = 127.3;
al   = .01;

Results
 CI_O  = 122.71628   131.88372
 CI_L  = 123.16023   Inf
 CI_P  = -Inf   131.43977
```

```
Example 8
Assume, that the height of children in the age 10 has normal distribution.
Determine the interval al-I, in which the true height will be it we have
measured the data sample of the length "n" with the average "mx".
The interval determine as a) both-sided,  b) rihgt-sided.
n    = 12;
mx   = 127.3;
s2x  = 38;
```

```
al   = .01;

Results
 CI_O  = 121.77318   132.82682
 CI_L  = 122.46314   Inf
 CI_P  = -Inf   132.13686


Example 9
To learn the accuracy of a method for measuring the volume
of manganese in the steel we performed independent measurements
of several samples. We would like to know the border for which it holds
that only 5% of possibly measured variances will be greater than
the value of the true variance of the method.
The measured values are
x = [4.3 2.9 5.1 3.3 2.7 4.8 3.6];

Results
 The border is 2.7509


Example 10
At the motorway with recommended speed 80 km/h we monitored
the speeds of passing cars and obtained data 'xi' and 'ni'
(values and frequencies). Determine both-sided 'al'-interval
for variance of the speeds.
xi = [60 70 75 80 85 90 110];
ni = [ 3 27 36 29 25 31   8];
al = .05;

Results
  CI  = 40.620085   474.35049


Example 11
At the motorway with recommended speed 80 km/h we monitored
the speeds of passing cars and obtained data 'x'
Determine both-sided 'al'-interval for the ratio of drivers
that exceed the recommended speed by more than 'r' km/h.
x = [78 86 65 92 83 92 85 66 42 82 ...
99 92 75 81 66 76 89 76 97 76 ...
75 56 76 78 96 77 86 79 86 93];
al= .05;
r = 3;

Results
  CI_both  = 0.2246955   0.5753045
  CI_both  = 0.   0.5471202
```

```
Example 12
From a set of steel rods with equal nominal length we have chosen
random choice with the lengths 'x'. The producer guarantees that
the variance of the lengths is 'sig2'. At the
significance level 'al' test the assertion of the producer
that the nominal length of the rods id 'd'.
x    = [6.2 7.5 6.9 8.9 6.4 7.1];
d    = 6.5;
sig2 = .8;
al   = .05;

Results
 pv  = 0.0678892


Example 13
From a set of steel rods with equal nominal length we have chosen
random choice with the lengths 'x'. At the
significance level 'al' test the assertion of the producer
that the nominal length of the rods id 'd'.
x  = [6.2 7.5 6.9 8.9 6.4 7.1];
d  = 6;
al = .05;

Results
 pv  = 0.1534113


Example 15
The accuracy of setting of certain machine can be verified
according to the variance of the products. If the variance is
greater then the level 'sig2', it is necessary to perform new setting.
A data sample has been meaasured with values 'mx' and
frequencies of values 'nx'. On the level 'al' test if it is necessary
to set the machine.
nx   = [ 5  12  32  11   8   3];
mx   = [95 100 105 110 115 120];
sig2 = 28;
al   = .05;

Results
 pv  = 0.078373


Example 16
Solidity of materials is verified by two methods A and B.
The same material has been subdued testing by both methods.
The results are 'xA' a 'xB'. On the level 'al' test equality of
```

```
both methods. The variability of methods is assumed to be equal.
xA = [20.1 19.6 20.0 19.9 20.1];
xB = [20.9 20.1 20.6 20.5 20.7 20.5];
al = .05;

Results
 pv  = 0.0024367


Example 17
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' a 'xP'.
Test at the level 'al'.
xL = [1.8 1.0 2.2 0.9 1.5];
xP = [1.5 1.1 2.0 1.1 1.4];
al = .05;

Results
pv  = 0.5528894


Example 18
At the motorway with recommended speed 80 km/h we monitored
the speeds of passing cars and obtained data 'x'.
At the level 'al' test the hypothesis: The ratio of drivers
that exceed the recommended speed by more than 'r' km/h is not
greater than 'P'%
x = [78 86 65 92 83 92 85 66 42 82 ...
99 92 75 81 66 76 89 76 97 76 ...
75 56 76 78 96 77 86 79 86 93];
al= .05;
r = 3;
P = 20;

Results
pv  = 0.0030849


Example 19
At the motorway with recommended speed 80 km/h we monitored
the speeds of cars going into the town and from the town. We
obtained data 'rD' (speeds into), 'nD' (frequencies into) and
'rZ' (speeds from) 'nZ' (frequencies from). At the level 'al'
test the hypothesis: From the town the cars go more quickly.
nD = [ 5 11 17 65 98 73 79  63   3];
rD = [65 70 75 80 85 90 95 100 110];
nZ = [ 8 22 13 71 48 64 89  24   5];
rZ = [65 70 75 80 85 90 95 100 110];
al = .01;
```

```
Results
 pv  = 0.0195199


Example 20
During a check of the front lights of cars we have measured
the data 'xL' (left light) and 'xP' (right light). The values
are distances (in cm) above (positive) and below (negative) of the
real level with respect to the optimal level. At the level 'al'
test if the light levels at each car are the same.
xP = [-3   5  16  9 -8 -2  23  5  -6 -3];
xL = [-5 -12  22 -3 -9  1  -1  2 -13 -5];
al = .1;

Results
pv  = 0.0749262


Example 21
During a check of the front lights of cars we have measured
the data 'xL' (left light) and 'xP' (right light). The values
are distances (in cm) above (positive) and below (negative) of the
real level with respect to the optimal level. At the level 'al'
test the hypothesis: Left lights are higher than right.
xP = [-3   5  16  9 -8 -2  23  5  -6 -3];
xL = [-5 -12  22 -3 -9  1  -1  2 -13 -5];
al = .1;

Results
 pv  = 0.0374631


Example 22
At a crossroads we have written down numbers of cars going
stright (R) turning to left (L) anf right (P). The measured
data are  'xR', 'xL' a 'xP'.
On the level 'al' test assertion that the ratio of cars
going straight is equal to those that are turning .
xR = 62;
xL = 39;
xP = 46;
al = .1;

Results
 pv  = 0.060902


Example 23
```

```
At a crossroads we have written down numbers of cars going
stright (R) turning to left (L) anf right (P). The measured
data are  'xR', 'xL' a 'xP'.
On the level 'al' test assertion that the ratio of cars
going straight is smaller than that of cars turning .
xR = [82 78 92 83 99 97];
xL = [29 42 34 38 45 34];
xP = [31 44 36 54 31 24];
al = .05;

Results
 pv  = 1.130D-69
```

Anova ————————————————

```
Example 24
We monitor three machines. Randomly, we measure their productions
per hour 'x1', 'x2' and 'x3. At the level 'al', test the equality
of their production.
al = 0.05;
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];

Results
  pv  = 0.0541908


Example 25
For one month, we monitor number of accidents at five crossroads.
The results are in the following table.
---------------------------------
rok:     1999 2000 2001 2002 2003
num_1      3    5    2    1    3
num_2      6    2    5    3    4
num_3      3    2    1    1    2
num_4      4    1    1    2    2
num_5      4    2    5    5    6
--------------------------------
At the level 'al' test hypothesis: The average number of accidents
is equal at all monitored crossroads.
al = 0.01;

Results
One-way anova
pv  = 0.0207414

Two-ways anova
```

Equality in crossoads 0.0195312

Equality in time 0.2749788

Nonparametric tests ——————————

```
Example 26
At a crossroads we have written down numbers of passing cars.
The measured data are 'd' - length of monitoring and
'x' - number of cars
d = [15 10 20 35 10 50];
x = [71 56 98 121 44 271];
At the level 'al' test the hypothesis that the cars go uniformly
(the same number per time unit).
al = .05;

Results
 pv  = 0.001917


Example 27
The following data are frequencies of incidents at certain
big factory
---------------------------------------------------
time interval        8-10h. 10-12h. 12-13h. 13-17h.
number of accidents   2       7       1       16
---------------------------------------------------
At the level 'al' test the hypothesis that the accidents
occur uniformly.
al = .05;

Results
 pv  = 0.1286455


Example 28
A connection between color of eyes and hair has been investigated.
I 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 'al' test the hypothesis that the color of eyes
and hair are independent.
al = .05;
```

```
Results
 pv  = 0.0167358


Example 29
Two operators alternate regularly at two machines. The produced
products are checked for quality. Each product is assigned by
the machine (S) and operator (O). The following data have
been measured
---------------------------------------------------
machine   1 2 1 1 2 2 2 1 2 1 1 1 2 1 2 2 2 1 2 1 2
operator  2 2 1 2 1 1 2 2 2 1 2 2 1 2 1 2 2 2 1 1 2
---------------------------------------------------
At the level 'al' test the assertion that the machines and
operators are with respect to the production quality independent.
al = .05;

Results
 pv  = 0.4663938


Example 30
Two doctors recommend curing the 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];
x2=[3 4 9 5 4 9 9 8];

Results
 Kr  = 24.
 U  = 48.5
 pv = 1.


Example 31
Eight sportsmen in certain sport 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 hypothesis is that one day of training is not enough to
improve their performance. Test on the level 'al'.

x1=[68 81 69 72 66 91 98 89 75 68];
x2=[79 62 70 75 68 81 85 94 71 62];

Results
 kr = 8.1
 W  = 36.
 pv = 1.
```

Example 32
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];
x2=[136 121 2.9 2.6 101 121 130 3.5 2.9 126];
```

Results

```
 kr = 8.1
 W  = 34.
 pv = 1.
```

Example 33
Tree inspectors are to evaluate functionality of five fast food stands. Each inspector evaluates each stand. The result is the table Tab: rows correspond to inspectors, columns columns to stands. Evaluation is 1,2,..,10. 10 is the best.
Test if the quality of the stans is equal.

```
Tab=[10 8 3 9 7
8  7 5 9 10
8  9 5 7  6];
```

Results

```
 W  = 9.487729   Inf
 T  = 6.6666667
 pv = 0.1545873
```

Example 34
A factory produces some products whose weight must be constant. For the production it uses four machines. A sample of products has been taken from all machines to test equality of the product weights. The measured values are 'x1', 'x2', 'x3' and 'x4'. Test the equality.

```
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];
```

Results

```
 pv  = 0.0335338
```