1 Interval estimates

Example 1 **(interval for μ - known variance)**

Assume, that the weight of products has normal distribution with the variance $\sigma^2 = 25$. Determine the interval α-I, $\alpha = 0.01$, in which the true weight will lie if we measured a data sample of 25 values with the average 135.

Result

$I = (132.4, 137.5)$ - *both sided*

Example 2 **(interval for μ - unknown variance)**

Assume, that the weight of products has normal distribution. Determine the interval α-I, $\alpha = 0.01$, in which the true weight will lie if we measured the data

x = [136 127 141 129 132 138 143 131]

Result

$I = (127.5, 141.8)$ - *both sided*

2 TH - one sample (parametric)

2.1 Expectation

Example 3 **(TH one expectation, unknown variance)**

On a section of motorway with a recommended speed 80 km/h we monitored the speeds of passing cars. We obtained a dataset of

a) $n = 12$ measurements

b) $n = 120$ measurements

with average $\bar{x} = 84$ and standard deviation $s = 18$. Test H0: the cars maintain the recommended speed. Test on the level $\alpha = 0.05$.

Result

influence of the sample length, BOTH-SIDED *test*

a) pv = 0.458

b) pv = 0.016

Example 4 **(TH one expectation, unknown variance)**

On a section of motorway with a recommended speed of 80 km/h we monitored the speeds of passing cars. We obtained the data

a) x = [98 86 65 92 83 92 85 66 62 82 99 92]

b) x = [98 86 65 92 83 92 85 66 62 82 99 92 89 94 81 88 79 95]

On the level $\alpha = 0.05$ test H0: the drivers in average do not exceed the speed.

Result

Test of expectation with one sample, RIGHT-SIDED

a) pv = 0.181 - they do not exceed

b) pv = 0.039 - they exceed

2.2 Proportion

Example 5 **(TH one proportion)**

On a section of motorway with a speed limit of 80 km/h we monitored the speeds of passing cars and obtained the data

x = [78 86 65 82 83 92 85 66 62 82 79 92]

Test H0 that the ratio of cars exceeding the speed is not greater than 30%

Result

n = 7, all = 12 *(n - number of those who exceed)*

pv = 0.016 - RIGHT-SIDED test (normal approximation)

$I = (0.304, 0.862)$ *-> 0.3 is not in I* $\rightarrow$ *reject*

2.3 Variance

Example 6 **(TH one variance)**

A machine produces rods of a specified length. The accuracy of the machine can be verified by the variance of the lengths. If the variance is greater then the level 50, it is necessary to adjust it. A data sample of lengths has been measured x = [101 104 103 110 108 116 129 98 104 111 115]

On the level 0.05 test if it is necessary to adjust the machine. Test on the level $\alpha = 0.05$.

Result

s2=76.2; RIGHT-SIDED test of variance

pv = 0.123 - not necessary

3 TH - two samples (parametric)

3.1 Expectation

Example 7 **(TH two expectations, independent)**

Two classes are to be compared in the knowledge of English. We randomly selected children from both classes and let them write a test. The results were in the range 0 – 100 (the higher the better).

The results are

x1 = [65 81 38 76 59 58 63 63 78]

x2 = [92 83 81 96 95 42 33 66 79 85 99]

H0: "The first class is not worse than the second one". Test it on the level $\alpha = 0.05$.

Result

LEFT-SIDED TH for 2 expectations, independent samples

Comparison of groups

pv = 0.062 - H0 not rejected

Example 8 **(TH two expectations, paired)**

Teacher insists students are getting worse in English. To test if individual pupils get worse, 10 of them were selected. They wrote a test with results 0 – 100 (higher is better). Next year the same pupils wrote another, similar test. The results are

x1 = [69 88 84 95 100 84 83 79 68 94]

x2 = [70 85 89 99 98 85 83 75 71 98]

Test on the level $\alpha = 0.05$.

Result

H0: art the same
HA: get worse

LEFT-SIDED TH for 2 expectations, paired samples

Comparison of individuals

pv = 0.19 - cannot reject "get worse"

3.2 Proportion

Example 9 **(TH two proportions - normal approximation)**

At a crossroads, we have written down numbers of cars going straight (S) turning to left (L) and right (R). The measured data are xS=62, xL=39 a xR=46. On the level 0.05 test H0: the ratio of cars going straight is not less then those that turn.

Result

LEFT-SIDED test of 2 proportions

xT=xL+xR=85; all = 147

al = 0.0073 - H0 is rejected

4 Many samples (parametric)

4.1 Anova

Example 10 **(Bartlett, one-way anova)**

For five years, we monitored number of accidents at four crossroads. The results are in the following table.

$X\backslash$year	2000	2001	2002	2003	2004
X_1	2	5	3	1	2
X_2	4	2	5	6	3
X_3	2	2	5	6	4
X_4	3	3	1	4	2

a) At the level 0.05 test the hypothesis: The average number of accidents is equal at all monitored crossroads.

b) On the same level test the hypotheses: The average number of accidents is equal for both the crossroads and the years.

Result

Bartlett: pv = 0.498 (for transposed)

a) pv_X = 0.333

b) pv_X = 0.381;   pv_year = 0.654

5 One sample (nonparametric)

5.1 Median

Example 11   (TH one median - Wilcoxon)

Let X denote the length (in centimeters), of a certain

fish species. We obtained the data set

d = [4.5 3.8 4.9 4.2 4.1 5.2 3.5 3.8 4.5 4.2 ...

4.7 4.2 3.5 3.8 4.5 4.2 4.5 3.8 4.9 4.2...

4.5 4.1 4.5 3.9 4.2 4.2 4.6 4.3 4.1 4.2]

Can we conclude that the median length of the fish species differs

significantly from the length 4.1 centimeters?

Result

Wilcoxon (one sample, BOTH-SIDED) - data are practically uniform

pv = 0.041

Help: Use Wilcoxon for two samples. first is d and second is $ones(d) * 4.1$

6 Two samples (nonparametric)

6.1 Median

Example 12 **(TH two medians - Mann-Whitney)**

Two doctors recommend treating colds with two different methods. The results (number of days of the treatment) are

x1=[5 6 4 4 5 8 5 7 5 6 3 4 7 7 5 6]

x2=[8 4 12 9 8 3 8 15 9 6 4]

Test equality of the methods.

Result

Man-Whitney - BOTH-SIDED

H0: are equal

pv = 0.056 (are not equal)

Example 13 **(TH two medians - Wilcoxon)**

Ten athletes in some sports club were tested with respect to their performance. They all threw the javelin once and then they were subjected to an intense training. After this they threw once more. The measured lengths were

x1=[68 69 75 72 83 88 79 88 76 81]

x2=[71 62 81 70 74 85 82 91 85 82]

The hypothesis HA is: One day of training is not enough to improve their performance. Test on the level 0.05.

Result

Wilcoxon (two samples), RIGHT-SIDED

pv = 0.65 (0.63 or 0.64 normal approximation) - is enough (according to H0)

6.2 Detection of a change

Example 14 **(TH McNemar)**

Some cold medicine has been tested (whether it helps or harms). The health of ten selected people was inspected 0-they are OK, 1-thea have cold. Then the medicine was applied and the health checked once more. The result is

x1 = [0 0 1 0 1 1 0 1 1 1]

x2 = [0 1 1 0 1 0 0 1 0 0]

Test, if the medicine has some effect.

Result

McNemar (table in Categorical/Count nominal data)

pv = 0.62 - not reject H0: no change in samples

6.3 Chi-Square tests

Example 15 **(Chi2 independence)**

Two operators O_1 and O_2 work in parallel in production. The products are checked for quality (1 good, 2-worse). The following data have been measured

O_1 [1 1 1 2 1 2 2 2 1 2 1 2 1 2 1 1 2 2 2 1 2]

O_2 [2 2 1 2 1 2 2 2 1 1 2 1 1 2 2 2 1 2 2 1 2]

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

Result

Chi-Square independence (table in Categorical/Count nominal data)

H0: are independent

pv = 0.28 (0.53 corrected) - are independent

Example 16 **(Chi2 homogeneity)**

In a study of the television viewing habits of children, a psychologist selects a random sample of 300 first graders - 100 boys and 200 girls. Each child is asked which of the following TV programs they like best: The Lone Ranger, Sesame Street, or The Simpsons. Results are shown in the contingency table below.

	Lone Ranger	Sesame Street	The Simpsons
Boys	50	30	20
Girls	50	80	70

Do the boys' preferences for these TV programs differ significantly from the girls' preferences? Use a 0.05 level of significance.

Result

(Deals with frequencies of discrete variable), H0 - are the same

pv $= 6.3 \cdot 10^{-5}$ *- the preferences are not the same*

Example 17 **(Chi2 goodness)**

The following data are frequencies of incidents at certain big factory

time interval [hour]	8-10	10-12	12-13	13-18
number of accidents	5	8	4	12

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

Result

Chi-Square goodness (sum of o and e must be equal)

$i = [2\ 2\ 1\ 5]$ $o = [5\ 8\ 4\ 12]$ $e = sum(o)*i/sum(i) = [5.8\ 5.8\ 2.9\ 14.5]$

H0: is uniform

pv = 0.62 - is uniform

7 Many samples (nonparametric)

7.1 Nonparametric anova

Example 18 **(Bartlett, Kruskal-Wallis)**

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=[35.6 35.1 35.8 39.4 34.8]

x2=[32.5 33.8 34.4 34.2 35.1 31.1]

x3=[36.3 30.8 35.6 35.2 30.2 35.1 34.2]

x4=[34.5 36.4 36.1 39.1 34.3 38.6]

Test the equality on condition that the data cannot be assumed normal.

Result

Bartlett - pv = 0.74

Test of H0: are equal; pv = 0.033 - are not equal

further analysis - Descriptive/Box-and-Whiskers

8 Other tests

8.1 Independence

Example 19 **(Pearson)**

We would like to build a model $y = f(x)$ from the data

x = [2 5 6 8 15 21 25]

y = [12 32 41 50 115 500 650]

Test whether the data x and y are are dependent at all.

Result

Parametric/Pearson correlation

pv = 0.0017 - are dependent (H0: independent)

Example 20 **(Spearman)**

We have decided, that it is not sure, if the data prom the previous example can be considered normal. We use still a nonparametric test.

We would like to build a model $y = f(x)$ from the data

x = [2 5 6 8 15 21 25]

y = [12 32 41 50 115 500 650]

Test whether the data x and y are are dependent at all.

Result

Nonparametric/Correlation/Spearman ...

coefficient = 1 →Perfect correlation (agrees with the previous)

Example 21 (test of normality)

We measured the speed of cars at a given point on the road and got the data

x = [69 82 79 55 85 80 91 88 69 45 57 82 69 98]

We would like to test whether the average speed is 80 km/h but we do not know if the test should be parametric or nonparametric. Test it.

Result

H0: they are normal

pv = 0.36 - is normal

Shapiro-Wilk W test pv = 0.95; 0.64

Kolmogorov-Smirnov $p > 0.1$

W/S test $p > 0.05$

Example 22 (test of association)

In six schools, similar classes (same year group and same number of children) were selected. Here, the number of children with excellent performance in math and English was found.

Math = [5 8 3 4 6 9]

English = [10 6 8 3 5 11]

Test whether the good performance in the subjects are associated.

Result

H0: they are associated.

chi2 - pv = 0.57 - yes, they are associated (show also in linear regression)