1 Test (statistics)

1.1 Example

The population is

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

We made a sample realization

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

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)

x	1	2	3
$f(x)$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{3}{12}$

, b) $E = 1.92$, $\bar{x} = 2$, $\hat{\theta} = 2$

1.2 Example

A population is described by normal distribution with expectation 3 and variance 1. We are going to make a sample of the length 10 and construct sample average. How will the sample average be distributed?

Result: Normally with expectation 3 and variance 0.1 - why?

1.3 Example

Show that sample average is unbiased estimate of expectation of Bernoulli distribution.

Hint: Show, that it holds $E[T] = \theta$ for Bernoulli.

1.4 Example

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

$$T = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

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

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

Result: $\hat{\theta} = 3.375$

1.5 Example

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

$$f(T) = \frac{1}{5} \text{ for } \theta \in (3,\, 8)$$

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

Result: (i) CI=(3.25, 7.75), (ii) CI=$(3.5,\, \infty)$, (iii) CI=$(-\infty,\, 7.5)$.

1% of the area is separated by $5/100 = 0.05$. For both-sided CI we need to add to the lower border 5% and to subtract the same from the upper border. Left sided CI has 10% added to the lower border and ∞ at the upper border. Right-sided is vice versa.

1.6 Example

For given data

$$x = \{1, 3, 5, 8, 15, 27, 39, 58\}$$

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

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

a) Write the resulting equations

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

Result:

a)

(i) $y = 44.93x - 231.47$; $pv = 4.4 \cdot 10^{-6}$

(ii) $y = 0.4x^2 + 22.3x - 90$; $pv = 1.7 \cdot 10^{-6}$

(iii) $y = -0.02x^3 + 1.76x^2 - 7.04x + 10.32$; $pv = 10^{-10}$

b) Cubic is the best regression. The smaller pv is the better regression.

1.7 Example

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

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

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

Result

$pv = 6.48 \cdot 10^{-5}$ - linear is suitable

$$y = 29.73x - 56463.4$$

1.8 Example

We suppose that traffic intensities 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	318	219	247	462	357	296	361
P_2	152	143	151	163	146	110	121	132
P_3	51	62	43	54	65	46	51	55

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

Solution

$$y = 0.11P_1 - 0.04P + 0.17P_3 - 2.88$$

$pv = 0.044$ on the level 5% the regression is not suitable.