Co-funded by the
Erasmus+ Programme
of the European Union

SPECIAL MOBILITY STRAND

QUALITATIVE AND QUANTITATIVE STATISTICAL METHODS IN RISK MANAGEMENT

SNJEŽANA MAKSIMOVIĆ

NOVI SAD 25.02.2020.

Snježana Maksimović[1] Faculty of Architecture, Civil Engineering and Geodesy

Plan of the talk

- *Introduction*
- *The probability theory*
- *Statistical methods*
- *The case of study*

Co-funded by the
Erasmus+ Programme
of the European Union

- *limited non-renewable natural resources (energy, materials) and limited renewable (drinking water, clean air, …).*

- *sustainable development - development that meets the needs of the present without compromising the ability of future generations to meet their own needs.*

- *civil engineering infrastructures are clear: save energy, save non-renewable resources and find out about re-cycling of building materials, do not pollute the air, water or soil with toxic substance and much more.*

Co-funded by the
Erasmus+ Programme
of the European Union

A beneficial engineered facility is understood as:

- *being economically efficient in serving a specific purpose*
- *fulfilling given requirements with regard to the safety of the personnel directly involved with or indirectly exposed to the facility*
- *fulfilling given requirements to limit the adverse effects of the facility on the environment.*

➢ *The task of the engineer is to make decisions or to provide the decision basis for others in order to ensure that engineered facilities are established in such a way as to provide the largest possible benefit.*

Co-funded by the Erasmus+ Programme of the European Union

Example: Feasibility of Hydraulic Power Plant

- *a hydraulic power plant project (involving the construction of a water reservoir in a mountain valley)*
- *the benefit of the hydraulic power plant (monetary income from selling electricity)*
- *the decision problem - compare the costs of establishing, operating and eventually decommissioning the hydraulic power plant with the incomes to be expected during the service life of the plant.*
- *ensured the safety of the personnel involved in the construction and operation of the plant and the safety of third persons.*

Co-funded by the Erasmus+ Programme of the European Union

- *selling electricity will depend on the availability of water, which depends on the future snow and rainfall*

- *the market situation may change and competing energy recourses such as thermal and solar power may cause a reduction of the market price on electricity*

- *the more the capacity the power plant will have, the higher the dam and the larger the construction costs will be ,as a consequence of dam failure the potential flooding will be larger*

- *the safety of the people in a town downstream of the reservoir will also be influencedmby the load carrying capacity of the dam structure*

Co-funded by the
Erasmus+ Programme
of the European Union

Risk - product of consequences and probabilities of dam failure (vary through the life of the power plant)

Careful planning

Questions

- *how large are the acceptable risks?*
- *what is one prepared to invest to obtain a potential benefit?*

The mathematical basis decision theory.

Co-funded by the Erasmus+ Programme of the European Union

If we have one event with potential consequences C, then the risk R is defined

$$R=CP$$

where P is a probability that event will occur.

If we have n events with potential consequences C_i, then the risk R is defined

$$R = \sum_{i=1}^{n} C_i \, P_i$$

where P_i is a probability that event will occur.

Co-funded by the Erasmus+ Programme of the European Union

Theory of probability

Probability theory deals with the study of phenomena whose results cannot be predicted.

- *Cardano and Galileo studying gambling*
- *Pascal and Ferma mathematical basis.*

Definition. *A random experiment is an experiment in wich, independent of the performance conditions, different outcomes occur. A set of all possible outcomes of an experiment we call the space of elementary events* Ω, *elements* $\omega \in \Omega$ *we call elementary events. Every subset* $A \subset \Omega$ *we call an event.*

Co-funded by the
Erasmus+ Programme
of the European Union

Example: *There are seven balls in the box: three red and four white. Determine the probability that from the box, without looking, we pull out a red ball along the assumption that removing each ball is equally possible?*

Solution:

$$\frac{3}{7} = \frac{\text{number of favorable outcomes}}{\text{number of all outcomes}}$$

If a set Ω has finally many equally possible outcomes, than:

Definition. *If m is the number of favorable outcomes of an event A⊂ Ω of a random experiment Ω and n is the number of all possible outcomes of that experiment, then the probability of an event is A is defined*

$$P(A) = \frac{m}{n}.$$

Co-funded by the Erasmus+ Programme of the European Union

Random variables

In addition to the outcome of random experiment, we register the value of a function corresponding to that outcome.

The outcome it can be a number (throwing a metal coin), and sometimes it's not the case.

Definition. *A random variable is a function* $f:\Omega \to R$ *that assigns a real number to each event* $\omega \in \Omega$ *.*

A variable, such as the strength of a concrete or any other material or physical quantity, whose value is uncertain or unpredictable is a random variable.

Co-funded by the
Erasmus+ Programme
of the European Union

Example. *A random variable X can be number of floods in a year or the number of vehicles passing an intersection during a given period.*

Example. *The coin is thrown twice. The space of elemental events is Ω={GG,GP,PG,PP}. Let X be a random variable that represents a number of letters (P) in two throws of coin. Than*

X(GG)=0, X(GP)=1, X(PG)=1, X(PP)=2.

Random variables: discrete (X(Ω) is countable) and continuous (X(Ω) is non-countable).

Definition. *A discrete random variable is a function* $X\colon\Omega \to R$ *that takes values from a countable set* $\{x_1, x_2, \ldots\}$ *with probability* $p_1 = P(X = x_1), p_2 = P(X = x_2), \ldots$

$$X = \begin{pmatrix} x_1 & x_2 & \ldots \\ p_1 & p_2 & \ldots \end{pmatrix}, \quad p_i \geq 0, \ \sum_i p_i = 1. \quad (1)$$

Co-funded by the
Erasmus+ Programme
of the European Union

Definition. *The distribution function of the discrete random variable X is the function F:R →[0,1] defined as*

$$F(x) = P(X \leq x), \qquad x \in R.$$

If X is defined by (1) then

$$F(x) = \sum_{x_i \leq x} p_i$$

Definition. *The expected value of the discrete random variable X defined by (1) is a number*

$$E[X] = \sum_i x_i p_i$$

the variance of X is

$$Var[X] = E[X^2] - E^2[X]. \qquad (2)$$

In practice is used a standard deviation $\sigma = \sqrt{Var[X]}$.

Co-funded by the
Erasmus+ Programme
of the European Union

Example: *Will we invest to I_1 or I_2?*

$$I_1 = \begin{pmatrix} 450\$ & 550\$ \\ 1/2 & 1/2 \end{pmatrix}, I_1 = \begin{pmatrix} 0\$ & 1000\$ \\ 1/2 & 1/2 \end{pmatrix}.$$

Notice that expectted values of I_1 and I_2 are $E[I_1] = E[I_2] = 500\$$. Since that values are equal we find the variance

$$Var[I_1] = \frac{450^2 + 550^2}{2} - 500^2 = 2\,500, \qquad \sigma = 50\$$$

$$Var[I_2] = \frac{1000^2}{2} - 500^2 = 250\,000, \qquad \sigma = 500\$$$

We will decide to investition I_1 .

Co-funded by the
Erasmus+ Programme
of the European Union

Distributions of discrete type

Binomial distribution. *The random variable X has a binomial distribution with parameters n and p, n ∈N, p ∈[0,1], if*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \le k \le n$$

We denote it X~Bin(n,p). If np<10, binomial distribution we approximate by a Poisson distribution.

Poisson distribution. *The random variable X has a Poisson distribution with parameter λ>0 if*

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0,1,2, \ldots$$

We denote it X~Poiss(λ). The Poisson distribution models well the phenomena in which there is a large population in which each member with a low probability gives a point in the process (example-Geiger counter).

Co-funded by the
Erasmus+ Programme
of the European Union

Distributions of continuous type

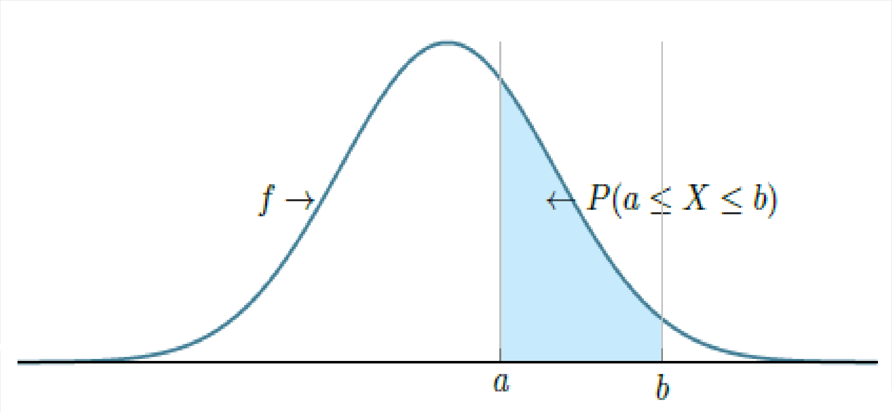
A random variable that can take any value from an interval [a,b] is called a continuous random variable.

Definition. *A random variable* $X:\Omega \to R$ *is a continuous if there exists a continuous function* $f: R \to R$ *such that f(x)≥0 for every x and* $\int_{-\infty}^{\infty} f(x)dx = 1$ *by which we can express*

$$P(a < X < b) = \int_a^b f(x)dx$$

Function f is a function density of distribution. A distribution function F is a primitive function od density f

$$F(x) = \int_{-\infty}^{x} f(t)dt$$



Co-funded by the Erasmus+ Programme of the European Union

Definition. *The expected value of continuous random variable X with density f is a number* $E[X]=\int_{-\infty}^{\infty} xf(x)dx$*, the variance is defined by (2).*

Uniform distribution. *A continuous random variable X has a uniform distribution on [a,b] if its function of density is*

$$f(x)=\begin{cases}\frac{1}{b-a}, & x\in[a,b]\\ 0, & otherwise\end{cases}$$

We denote it X~U(a,b). A discrete uniform random variable is distributed

$$X=\begin{pmatrix} x_1 & x_2 & \dots & x_n\\ 1/n & 1/n & \dots & 1/n\end{pmatrix}$$

Connection with Poisson distribution. *If Poisson proces has n points in [a,b], their locations are distributed independently each with a uniform distribution on [a,b].*

Co-funded by the
Erasmus+ Programme
of the European Union

Exponential distributions. *A continuous random variable X has a exponential distribution* with the parameter λ *if a function of density of X is*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

We denote it X~Exp(λ). The exponential distribution is used as a model for the time between two faults of a device, the time between the arrivals of persons in mass services (banks, shops ...), the time between phone calls,…

<u>Connection with Poisson distribution</u>. The time between the random events in Poisson process is distributed by the exponential distribution.

Normal distribution. *A continuous random variable X is a normally distributed with parameters μ and $\sigma^2 > 0$ if its density is*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; x \in R.$$

We denote it X~N(μ, σ^2).

Co-funded by the Erasmus+ Programme of the European Union

Normal distribution - when we wait for a queue in one of the hypermarkets, when we pour milk into a coffee milk particles are normally distributed before filling all the volume, the student's achievement in classes is normally distributed, also the weight and height of people,…

If $X \sim N(\mu, \sigma^2)$*, then* $Z=(X-\mu)/\sigma \sim N(0,1)$ *(standardized normal distribution). A density and distribution function of Z are*

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx.$$

Many discrete distribution can be approximated by normal distribution. If in binomial distribution n>50 and np>10, than

$$Bin(n,p) \approx N(np, np(1-p)).$$

Co-funded by the
Erasmus+ Programme
of the European Union

Logaritmic normal distribution. *A continuous random variable X is a logaritmic normally distributed with parameters* μ *and* $\sigma^2 > 0$ *if its density is*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}, \ x > 0.$$

<u>Modeling with logaritmic normal distribution:</u> the cow's milk production, rainfall, maximum water flow rate in the river during the year, an amount of personal income,…

Gamma distribution. *A continuous random variable X has a gamma distribution with parameters* $\lambda, \alpha > 0$ *if its density*

$$f(x) = \begin{cases} \dfrac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Connection with exponential distribution: $\alpha = 1$

<u>Modeling with gamma distribution:</u> modeling of waiting time, model for financial losses or insurance claims, in wireless comunication as model of multistage weakening of power signal …

Co-funded by the Erasmus+ Programme of the European Union

χ^2 ***distribution.*** *This distribution is the special case of Gamma distribution when λ=1/2, α=n/2, n $\in$N.*
The application in mathematical statistics (χ^2 test).

Student's t-distribution. *A continuous random variable X is has a Student's distribution with n degrees of freedom if its density is*

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\left(\frac{n+1}{2}\right)}{\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad , x \in R$$

When n$\rightarrow\infty$, then Student's t distribution converges to standardized normal distribution. Student's t distribution is used in various statistical estimation problems where the goal is to estimate an unknown parameter, such as the mean in an environment where data are viewed with additional errors.

Multidimensional random variables

Monitor multiple values of random variables in parallel (when monitoring the quality of ceramic)

Definition. *A function* $X = (X_1, X_2, \dots, X_n) \colon \Omega \to R^n$ *is an n- dimensional random variable.*
For n=2 - a two-dimensional random variable.

Definition. *Random variables X and Y are independent if*

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y), \qquad x, y \in R.$$

For expected value of X+Y it holds that E[X+Y]=E[X]+E[Y]. Does it hold for Var[X+Y]?
Answer: No!

Co-funded by the
Erasmus+ Programme
of the European Union

Definition. *Covariance of random variables X and Y is defined by*

$$Cov(X,Y)=E[(X-E[X])(Y-E[Y])].$$

If Cov(X,Y)=0, than random variables X and Y are non-corelated . If X and Y are independent, than Cov(X,Y)=0 and

$$E[XY]=E[X]E[Y] \qquad Var[X+Y]=Var[X]+Var[Y].$$

A random variable $X^* = \frac{X-E[X]}{\sqrt{Var[X]}}$ *is called a standardized random variable.*

Definition. *Covariance of standardized random variables* X^* *and* Y^* *is called a correlation coefficient and defined by*

$$Cov(X^*,Y^*) = \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]}\sqrt{Var[Y]}}.$$

Co-funded by the
Erasmus+ Programme
of the European Union

Basics of mathematical statistics

The study of random phenomena - measuring various data (statistical data).

What is the task of mathematical statistics?

Definition. *A set Ω considered in mathematical statistics is called a population or a general set. A function that every $\omega \in \Omega$ assigns a real number is called the feature and it is denoted by X, Y, Z,… (random variable in probability theory).*

Example. *The population is a set of products of one factory. The feature of each product is, for example, its price.*

sample

Co-funded by the
Erasmus+ Programme
of the European Union

The statistical study of a some feature involves three stages:

- *statistical observation,*
- *grouping and arranging data*
- *processing and analysis of results.*

The observed feature can be:

- *qualitatively: binary - there are two choices e.g. smoker and non-smoker; ordinarily - there is a hierarchy e.g. level of education; nominally - no hierarchy e.g. nationality*
- *quantitative (numerical).*

registered data display graphical

- ➢ *discrete, registered data grouping into classes,*
- ➢ *continuous feature registered data is grouped at intervals.*

Co-funded by the
Erasmus+ Programme
of the European Union

How we graphically displayed a qualitative feature? The usual choice is a pie graph and a bar chart.

How we graphically displayed a numerical feature? The usual choice is dot diagram, frequency histogram, boxplot and line diagram.

Definition. *Let* $(X_1, X_2, \dots, X_n)$ *be a simple random sample of population with a feature X and a function* $f: R^n \to R$ *. A random variable* $U = h(X_1, X_2, \dots, X_n)$ *is called statistics.*

Most important statistics:

- *mean value of sample* $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$
- *variance of sample* ${S_n}^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X_n})^2$
- *repaired variance of sample* $\overline{S_n}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X_n})^2$

Co-funded by the
Erasmus+ Programme
of the European Union

A sample $(x_1, x_2, \ldots, x_n)$ is a realization of n- dimensional random variable.

Parameter t is the value which depends only from the sample

$$t = h(x_1, x_2, \ldots, x_n)\ .$$

Parameter t is a realization of a random variable

$$T=h(X_1, X_2, \ldots, X_n).$$

Example. *Statistics $\overline{X_n}$ is an estimation of μ, since statistics $\overline{S_n}^2$ is an estimation of σ^2.*

Estimations :

- *Dotted estimations (maximum likelihood method)*
- *Interval estimations* (confidence intervals)

Co-funded by the Erasmus+ Programme of the European Union

Statistical tests

An application of statistical methods in laboratories and manufacturing facilities is associated with conclusion about:

- *does analytical method have systematic errors*
- *do two measurement methods differ in accuracy,*
- *which of two technological processes is better*

- *to make conclusions about parameters of basic set based on a random sample.*

Definition. *Any assumption about the characteristics of the basic set expressed in the form of a statement of distribution (one or more) features is called the statistical hypothesis.*

- *null hypothesis* H_0
- *alternative hypotesis* H_1.

Co-funded by the Erasmus+ Programme of the European Union

Definition. *The procedure for testing of null hypothesis based of a realized sample is called a statistical test.*

Hypotheses:

- *Parametric hypotheses*
- *Non-parametric hypotheses*

Procedure of testing a parametric hypothesis is called a parametric test, and non-parametric hypotheses non-parametric test.

The critical area $C \subset R^n$*:*

- *one-tailed*
- *two-tailed*

Co-funded by the Erasmus+ Programme of the European Union

If a realised value of sample is in C, then H_0 is rejected.

An acceptance of hypothesis H_0 based on the sample from the basic set does not mean that it is it is correct, it just means that the sample does not contradict the hypothesis.

By testing hypotheses there is a risk that the conclusion of the test is incorrect:

- *Type 1- H_0 is correct, but based on the sample is rejected (α)*
- *Type 2- H_0 is not correct, but based on the sample is accepted.*

How we choose null hypothesis?

Co-funded by the Erasmus+ Programme of the European Union

Parametric tests:
- *Z-test,*
- *t-test,*
- *analysis of Variance,*
- *tests of hypotheses involving the variance, …*

Non-parametric tests:
- *Pearson* χ^2 *test,*
- *Kolmogorov-Smirnov test,*
- *Mann Whitney U-test,*
- *Kruscal Wallis H-test,…*

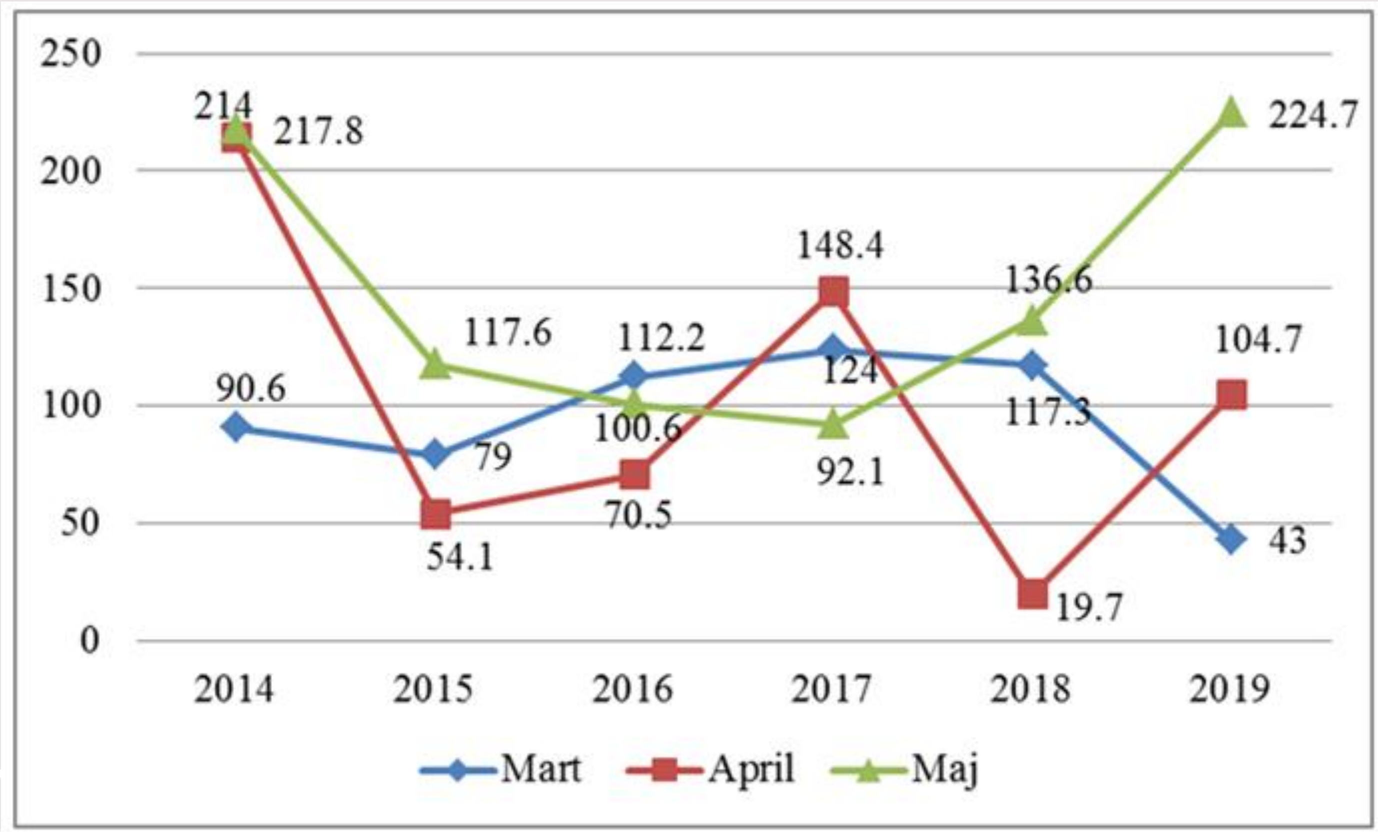
Co-funded by the
Erasmus+ Programme
of the European Union

The case of study

- *Housing units of two Banja Luka settlements Česma and Budžak, which were flooded in May of 2014 (total 38 housing units).*
- *A survey questionnaire was created which was filled by the inhabitants of these settlements, population data, plot size, distance of the object on the plot of the river bed, height of flood damage.*
- *The distance of these objects in relation to the river was also analyzed, flooding and damage that occurred.*
- *The obtained results were presented through descriptive statistics and adequate statistical tests in the analytical-software package SPSS v.23.*

Co-funded by the
Erasmus+ Programme
of the European Union

Rainfall for critical months 2014-2019

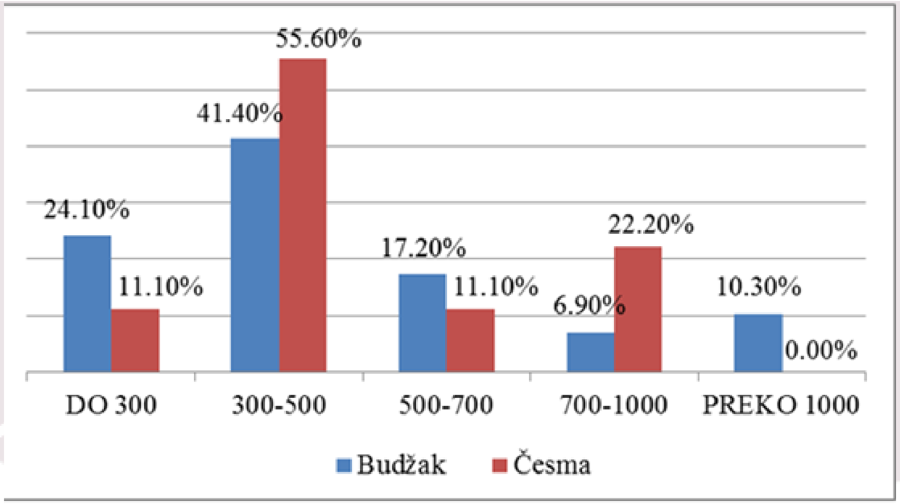


Co-funded by the Erasmus+ Programme of the European Union

The distance of objects from the river bed

		0-50	51-100	101-200	>200	Sum
PLOT	< 300 m²	2	3	1	2	8
	300-500 m²	7	7	2	1	17
	500-700 m²	2	1	2	1	6
	700-1000 m²	0	1	2	1	4
	> 1000 m²	1	0	1	1	3
Sum		12	12	8	6	38

Distribution of plots by settlements



Co-funded by the Erasmus+ Programme of the European Union

The amount of damage by settlements

SETTL.	Mean value	N	Std. Dev.	Median	Min	Max
ČESMA	20925.9	27	9396.9	20000.0	10000	50000
BUDŽAK	15875.0	8	4673.3	17500.0	10000	20000
Sum	19771.4	35	8755.1	20000.0	10000	50000

There is no statistically significant difference in the amount of damage per settlements ($U=74.500$, $z=-1.350$, $p=0.177$).

		POPULATION AWARNESS			
		YES	NO	PARTIALLY	SUM
SETTL.	CESMA	7	8	14	29
	BUDZAK	0	4	5	9
Total		7	12	19	38

Co-funded by the
Erasmus+ Programme
of the European Union

		THE STATE HAS TAKEN THE NECESSARY MEASURES				
		PARTIALLY	LITTLE	NO	I DON'T KNOW	SUM
SETTL.	CESMA	7	1	20	1	29
	BUDZAK	2	3	4	0	9
SUM		9	4	24	1	38

By the opinion of the surveyed population, 24 (63%) believe that the state has not taken the necessary measures to protect against floods. There was no answer YES.

Co-funded by the Erasmus+ Programme of the European Union

K-FORCE

Co-funded by the Erasmus+ Programme of the European Union

Thank you
for your attention

Contact info about the presenter:
snjezana.maksimovic@aggf.unibl.com

Knowledge **FO**r **R**esilient so**C**i**E**ty