Snježana Maksimović[1]

QUALITATIVE AND QUANTITATIVE STATISTICAL METHODS IN RISK MANAGEMENT

Abstract: In the 21st century construction industry involves more dynamic and uncertain planning than ever before. In the construction management decision makers should follow a systematic and professional approach in a risk management. In this chapter we briefly review some concepts regarding the risk management; give basics in the probability theory and mathematical statistics; give some statistical methods in the risk management. In the end of this chapter we apply some of above statistical methods in a case study of the city of Banja Luka examining the risk of flooding of housing near river beds.

[1] Assistant profesor, Faculty of Architecture, Civil Engineering and Geodesy, Vojvode Stepe Stepanović 77/3, Banja Luka and snjezana.maksimovic@aggf.unibl.org

INTRODUCTION

During the last two decades, there has been a growing awareness that our world only has limited non-renewable natural resources such as energy and materials but also limited renewable resources such as drinking water, clean air etc. A sustainable development is defined as a development "that meets the needs of the present without compromising the ability of future generations to meet their own needs" [4].

In regard to environmental impacts the immediate implications for the planning, design and operation of 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 substances, save or even regain arable land and much more. Engineers play an important role for society as they are directly involved in establishing the basis for societal decision making. The value of decisions is and will always be relative to the objectives as well as the assumed boundary conditions.

The aim of this paper is to present probability theory and statistics in the context of supporting engineering decision making. It is assumed that the reader has only little or no prior knowledge on the subject of probability theory and statistics. The organization of this paper is the following: a first section is an introductory character in which are given basic definitions in the risk management; in the second and the third sections are given basics of probability theory and statistics; finally, in the fourth section, some of the following statistical methods were applied in a case study of the city of Banja Luka examining the risk of flooding of housing near river beds [2], [3].

1. RISK MANAGEMENT PROCESS

Consider as an example the decision problem of exploitation of hydraulic power [4].

Example 1.1. A hydraulic power plant project involving the construction of a water reservoir in a mountain valley is planned. The benefit of the hydraulic power plant is associated only with the monetary income from selling electricity to consumers. *The decision problem* is to 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. It must be ensured that the safety of the personnel involved in the construction and operation of the plant and the safety of third persons.

There are many factors which are important for the building of the power plant:

- 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

Risk respresents product of consequences and probabilities of dam failure. Both consequences and probabilities vary through the life of the power plant.

Careful planning during the first phase of a project is the only way to control the risks associated events (errors, mistakes, accidents, human errors, failures of structural components, extreme load situations ,natural hazards,…).

Finally to evaluate wheter a hydraulic power plant is cost effective we ask the following questions:

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

The mathematical basis for such decision problems is *decision theory*.

Definition 1.1.If we have one event with potential consequences C, then the risk R is defined

$$R=CP \tag{1}$$

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

$$R = \sum_{i=1}^{n} C_i P_i \tag{2}$$

where P_i is a probability that event will occur.

Hazard is a process, phenomena or human activity that may cause loss of life, injury or other health impacts, property damage, social and economic disruption or environmental degradation. Hazards may be single, sequential or combined in their origin and effects. Each hazard is characterized by its location, intensity or magnitude, frequency, and probability [11].

Vulnerability is defined as the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards.

Since we cannot reduce the severity of natural hazards, the main opportunity for reducing risk lies in reducing vulnerability and exposure.

Disaster Risk Reduction is aimed at preventing new and reducing existing disaster risk and managing residual risk, all of which contribute to strengthening resilience and therefore to the achievement of sustainable development. Disaster Risk Reduction strategies and policies define goals and objectives across different timescales, with concrete targets, indicators and time frames.

Disaster Risk Management is the application of disaster risk reduction policies and strategies, to prevent new disaster risks, reduce existing disaster risks, and manage residual risks, contributing to the strengthening of resilience and reduction of losses. Disaster risk management actions can be categorized into; prospective disaster risk management, corrective disaster risk management and compensatory disaster risk management (also referred to as residual risk management).

In recent years, researchers and experts have been developing methods to conduct the assessment of hazards, vulnerability, and coping capacities; as well as techniques to combine such assessments in order to present them in risk map format. Such maps are essential in developing strategies to reduce the level of existing risks, and as a way to avoid a generation of new risks due to underlying social and economic risk drivers. Since the beginning of the 1990s, the United Nations has been promoting efforts to change the paradigm of disasters, advocating for the incorporation of disaster risk reduction efforts worldwide as a way to reduce the effects of natural hazards on vulnerable communities.

2. PROBABILITY THEORY

The theory of probability deals with the study of phenomena whose results cannot be predicted.The foundations of probability theory were laid by Cardano and Galileo studying gambling, and Pascal and Ferma are the laws of gambling set on a mathematical basis. Nowadays the theory of probability is away from gambling it is connected with natural phenomena and processes in the production. In the probability theory, mathematical models of real life are studied occurrence, while in statistics, a sampling method establishes a connection between the actual phenomena and the corresponding models. The statistics are therefore closer to reality than the probability. We can also say that the statistics is applied probability [5], [6].

Definition 2.1. A random experiment is an experiment in which, 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.

Example 2.1. 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 2.2. If m is the number of favorable outcomes of an event $A \subset \Omega$ 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

$$P(A) = \frac{m}{n}. \qquad (3)$$

2.1. Random variables

When we perform an experiment, in addition to outcomes of that experiment, we register the value of a function corresponding to that outcome (to the outcome of the experiment we assign a real number see Example 2.3). Depending on the experiment, sometimes the outcome can be a number (throwing a metal coin) and sometimes it's not the case.

Definition 2.3. A random variable is a function $f: \Omega \rightarrow R$ 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.

Example 2.2. 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 2.3. The coin is thrown twice. The space of elemental events is $\Omega=\{GG,GP,PG,PP\}$. Let X be a random variable that represents a number of letters (P) in two throws of coin. Than

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

We have two basic types of random variables: *discrete* when the set $X(\Omega)$ is countable and *continuous* then the set $X(\Omega)$ is noncountable.

Remark 2.1. There are random variables that are neither discrete or continuous, but we will not study them separately.

2.1.1.Discrete random variables

Definition 2.4. A discrete random variable is a function $X: \Omega \rightarrow R$ that takes values from a countable set $\{x_1, x_2, \ldots\}$ with probabilities $p_1 = P(X = x_1)$, $p_2 = P(X = x_2), \ldots$

All variables x_i and all probabilities p_i, $i \in N$, make the law of probability distribution of a discrete random vaiable and we write

$$X = \begin{pmatrix} x_1 & x_2 & \ldots \\ p_1 & p_2 & \ldots \end{pmatrix}, \quad p_i \geq 0, \quad \sum_i p_i = 1. \tag{4}$$

Definition 2.4. The distribution function of the discrete random variable X is the function $F: R \rightarrow [0,1]$ defined as

$$F(x) = P(X \leq x), \quad x \in R. \tag{5}$$

If the random variable X is defined by (4) then its distribution fuction is

$$F(x) = \sum_{x_i \leq x} p_i \tag{6}$$

Definition 2.5. The expected value of the discrete random variable X defined by (4) is a number

$$E[X] = \sum_i x_i p_i \tag{7}$$

and the variance of X is

$$Var[X] = E[X^2] - E^2[X]. \tag{8}$$

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

Below are some of the most important distributions of discrete type.

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

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \leq k \leq n \tag{9}$$

The random variable X with binomial distribution we denote by $X \sim Bin(n,p)$. If $np < 10$, binomial distribution we approximate by a Poisson distribution.

Poisson distribution. The random variable *X* has a Poisson distribution with parameter $\lambda>0$ if

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\dots \tag{10}$$

The random variable *X* with a Poisson distribution we denote by $X\sim Poiss(\lambda)$. The Poisson distibution 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).

Geometric distribution. The random variable *X* has a geometric distribution with parameter $p\in(0,1)$ if

$$P(X=k)=(1-p)^{k-1}p,\qquad k\in N \tag{11}$$

The random variable *X* with geometric distribution we denote $X\sim Geo(p)$.

2.1.2. Distributions of continuous type

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

Definition 2.6. A random variable $X\colon \Omega\to R$ is a continuous if there exists a continuous function $f\colon R\to R$ such that $f(x)\geq 0$ for every *x* and

$$\int_{-\infty}^{\infty} f(x)dx=1 \tag{12}$$

by which we can express probability that random variable *X* is in *(a,b)* (Figure 1)

$$P(a<X<b)=\int_a^b f(x)dx \tag{13}$$

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

$$F(x)=\int_{-\infty}^{x} f(t)dt \tag{14}$$

Remark 2.2. Notice that a discrete random variable has not a function density, nor a continuous random variable has not the law of the distribution propability. But both have a function of distribution.

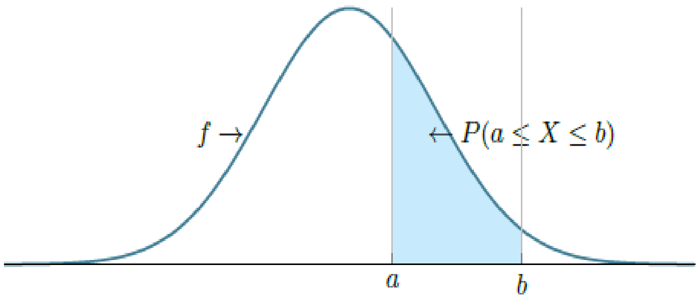

Figure 1. Density of continuous distribution

Definition 2.7. The expected value of continuous random variable X with density f is a number

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx, \tag{15}$$

the variance is defined by (8).

Below are some of the most important distributions of continuous type.

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

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b] \\ 0, & otherwise \end{cases} \tag{16}$$

The random variable X with uniform distribution we denote by $X \sim U(a,b)$. A discrete uniform random variable is distributed

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ 1/n & 1/n & \dots & 1/n \end{pmatrix} \tag{17}$$

The uniform distribution is connected with the Poisson distribution. If a Poisson proces (collection of random variables) has n points in $[a,b]$, their locations are distributed independently each with a uniform distribution on $[a,b]$.

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

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{18}$$

The random variable X with exponential distribution we denote by $X \sim Exp(\lambda)$. 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,…

The exponential distribution is connected with the Poisson distribution. The time between the random events in the Poisson proces 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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in R. \qquad (19)$$

The random variable X with exponential distribution we denote by $X\sim N(\mu,\sigma^2)$.

Normal distribution we find 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 normaly distributed, also the weight and height of people,…

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

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx. \qquad (20)$$

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

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

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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}, x > 0. \qquad (21)$$

Modeling with logaritmic normal distribution: 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

$$f(x) = \begin{cases} \dfrac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad (22)$$

If we put $\alpha=1$ in (22) we obtain an exponential distribution, so the exponential distribution is a particular case of Gamma distribution.

One can use a Gamma distribution for modeling of waiting time, as a model for financial losses or insurance claims, in wireless comunication as model of multistage weakening of power signal …

χ^2 ***distribution.*** This distribution is the special case of Gamma distribution when $\lambda=1/2$, $\alpha=n/2$, $n\in N$.

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

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\left(\frac{n+1}{2}\right)}{\left(\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, x \in R \qquad (23)$$

When $n\to\infty$, 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 value in an environment where data are viewed with additional errors.

2.2. Multidimensional random variables

In practice, it is often necessary to monitor multiple values of random variables in parallel. For example, when monitoring the quality of ceramic tiles produced, all three dimensions are controlled.

Definition 2.8. A function $X=(X_1,X_2,\ldots,X_n)\colon \Omega\to R^n$ is an n-dimensional random variable.

For $n=2$, we have a two-dimensonal random variable.

Definition 2.9. Random variables X and Y are independent if

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y), \quad x, y \in R. \qquad (24)$$

For expected value of $X+Y$ it holds that $E[X+Y]=E[X]+E[Y]$. Does it hold for $Var[X+Y]$?

The answer is no, since

$$Var[X+Y]=Var[X]+Var[Y]+2\,E[(X-E[X])(Y-E[Y])].$$

The expresion $E[(X-E[X])(Y-E[Y])]$ shows how X and Y affect each other. This is motivation for the following definition.

Definition 2.10. Covariance of random variables X and Y is defined by

$$Cov(X,Y)=E[(X-E[X])(Y-E[Y])]. \qquad (25)$$

If Cov(X,Y)>0, than random variables X and Y are positive correlated; Cov(X,Y)<0 means that X and Y are negative correlated and if Cov(X,Y)=0, than random variables X and Y are noncorelated . If X and Y are independent, than Cov(X,Y)=0 and

$$E[XY]=E[X]E[Y]\ , \qquad Var[X+Y]=Var[X]+Var[Y]. \tag{26}$$

A random variable

$$X^* = \frac{X-E[X]}{\sqrt{Var[X]}} \tag{27}$$

is called a standardized random variable.

Definition 2.11. A covariance of standardized random variables X^* and Y^* is called a correlation coefficient and defined by

$$Cov(X^*,Y^*) = \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]}\sqrt{Var[Y]}}. \tag{28}$$

3. MATHEMATICAL STATISTICS

The mathematical statistics is closely realated to the probability theory. The study of random phenomena implies the possibility of measuring various data. These tada are called *statistical data*. The task of mathematical statistics is to choose from registrated data the probability which the best describe the accidental phenomenon being observed.

Descriptive statistics play an important role in engineering risk analysis as a standardized basis for assessing and documenting data obtained for the purpose of understanding and representing uncertainties in risk assessment.

Definition 3.1. 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 $X(\omega), Y(\omega), Z(\omega),\ldots$ (or simpler $X,Y,Z,\ldots$).

Example 2.1. The population is a set of products of one factory. The feature of each product is, for example, its price.

In the most cases it is complicated to register a feature for each element of the population, so we register feature only on the one part of the population which we call *a sample*.

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 (there is no hierarchy e.g. nationality)
- quantitative (numerical).

If the sample has a large number of elements we arrange registered data and display them graphically. If a feature is discrete, registrated data we grouping into classes, and with continuous feature registrated data is grouped at intervals.

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 a dot diagram, a frequency histogram, a boxplot and a line diagram.

We say that $(X_1,X_2,\dots,X_n)$ is a *simple random sample* if random variables $X_1,X_2,\dots,X_n$ are independent with the same distribution.

Definition 3.2. Let $(X_1,X_2,\dots,X_n)$ be a simple random sample of population with a feature X and a function $f\colon R^n \to R$. A random variable $U=h(X_1,X_2,\dots,X_n)$ is called statistics.

Most important statistics are:

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

A sample $(x_1,x_2,\dots,x_n)$ is a realisation of n-dimensional random variable. The parameter t is the value which depends only from the sample, that is $t=h(x_1,x_2,\dots,x_n)$. Therefore, the arameter t is a realisation of a random variable

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

Example 2.2. The statistics $\overline{X_n}$ is an estimation of μ, since statistics $\overline{S_n}^2$ is an estimation of σ^2.

We usually use the following estimates to estimate an unknown parameter:

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

3.1. Statistical tests

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

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

Most often we want to reach a conclusion about the parameters of the basic set (mean value and variance) based on a random sample. The first step in the process of reaching a conclusion is to set up appropriate statistical hypotheses or assumptions which are tested.

Definition 3.3. 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.

At the beginning we set null hypothesis which we denoted by H_0. The null hypotesis is the one we are testing. The oposite hypotesis is an alternative hypotesis which we denote by H_1.

Definition 3.4.The procedure for testing of null hypothesis based of a realized sample is called a statistical test.

We will consider two types of hypotheses:

- Parametric hypotheses that represent an assumption about the value of someone from the parameters of the known (assumed) distribution.
- Non-parametric hypotheses representing the assumption of the type of distribution population, determination of dependencies (two or more) features, equality testing,….

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

The critical region (rejection region) $C \subset R^n$ is the set of all those points for which hypothesis H_0 is rejected. By type of critical region, statistical tests are divided into:

- one-tailed
- two-tailed

If a realised value of sample $(x_1, x_2, \ldots, x_n) \in C$, 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, so we have two types of errors:

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

We denote the probability of an error of the type 1 by α. The number α is called *the level of significance* and that value represents a risk in making decisions.

Remark. We often choose for the null hypothesis the one for which the data suggests that be rejected, because the rejection is usually more reliable than acceptance.

The most commonly used parametric tests are:

- Z-test (one or two samples),
- t-test (one or two samples),
- analysis of variance,
- tests of hypotheses involving the variance, …

The most commonly used non-parametric tests are:

- Pearson χ^2 test,
- Kolmogorov-Smirnov test,
- Mann Whitney U-test,
- Kruscal Wallis H-test,…

The procedure for testing is the following:

1. Set the hull hypotesis H_0;
2. Set the alternative hypotesis H_1;
3. Determine the test statistic and determine a realized value of the test statistic;
4. Based on significance level α define a critical region for the test statistic;
5. Verify if computed value of the test statistic is within or outside the critical region.

4. STATISTICAL METHODS IN RISK MANAGEMENT

Bosnia and Herzegovina is susceptible to many natural disasters such as floods, landslides, droughts, fires and earthquakes. The floows in May 2014, as the worst floods in the last 120 years, hit a quarter of its territory. Heavy rainfall led to the discharge of several rivers from their beds, especially Vrbas, Vrbanja, Bosna and Sava, as well as Drina, Una, Sana and their tributaries, causing sudden and extremely large floods in numerous places in the Bosnia Valley, with the largest spills exceeding a return period of 500 years [1]. These floods and landslides caused damage to the infrastructure, an agriculture, public institutions and the local economy, estimated at 1.67 billion $ and affected about one million people [8].

It turned out that the population was unprepared for such situations. It was a reminder that much more attention should be paid to measures to protect and prevent future natural disasters. Rebuilding the damaged and recreating the living conditions of the flood-affected population was a much greater financial challenge for the country than investing in conservation measures. Some of flood protection measures that were subsequently implemented are: a functional wastewater system has been developed to prevent the return of wastewater back to homes in the future, river beds have been expanded and strengthened, dams and embankments and bridges have been strengthened [8].

Flooding also affected the housing sector. In Bosnia and Herzegovina were affected 8 103 620 square meters of built-up living by floods in 2014 and 283 777 inhabitants. In the city Banja Luka were affected 312 930 square meters of built-up living space by floods and 11 327 people were threatened [9].

The major limiting factors for flood risk assessment for the housing sector in BiH are follows:

- a deficiency of available flood hazard maps, especially for the Republic of Srpska,
- population estimate using two different methods, depending on data availability:
  - calculation of population density in specific flood hazard maps where housing units are missing,
  - calculating the number of inhabitants by attributing a certain number of persons to each of the types of housing units present in the floodplain,
- a deficiency of a hydrodynamic model which is why the hydrological method was used.

4.1. The case of study: the Banja Luka city

We analyzed 38 plots and objects as well as the population living there, in two Banja Luka settlements (Cesma and Budžak) that have the highest risk of flooding and that were flooded in May 2014 (some and more times). The distance of these objects from the riverbeds of rivers Vrbas and Vrbanja was observed and their floodplain was analyzed.

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 [7]. We use descriptive statistics and Mann Withneu U test, since data are not normaly distributed. If the distribution is not normal, most often we use Mann Withneu U test and Kruscal Wallis test for the decision making

Figure2 shows rainfall data for months: March, April and May during the period 2014-2019 in the territory of the Banja Luka city [10].

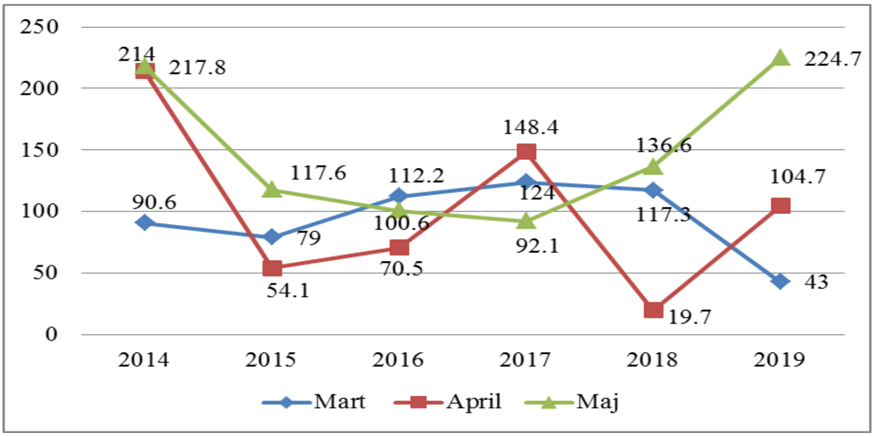

Figure 2. Rainfall 2014-2019 for critical months (l / m^2)

In our sample, 35 housing units were flooded in May 2014, of which 8 housing units were flooded multiple times in the last 10 years, while only 3 housing units were not flooded.

Table 1: The distance of objects from the river bed in relation to the parcel size

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

Table 1 shows that in the observed sample, the most objects are on plots of size from 300 to 500 m2 (17) and 7 objects are 0-50 m and 7 objects 51-100 m away from the river bed. In our sample (38 plot) there is one housing unit on each plot.

Considering the size of plots by settlements in the observed sample, the most plots are 300-500 m^2, 41,40% in the settlement Budžak and 55,60% in the settlement Česma. Large plots (over 1000 m2) were only in the settlement Budžak (3 plots in total).

Out of the observed 38 objects in two settlements, only three were not flooded, most were flooded in May 2014 and 8 were flooded several times in the last 10 years. Of the three non-flooded objects, one is located 50 m from the river bed, one at a distance of 300 m and one at a distance of 2000 m.

Table 2: Estimated damage rates in flooded households

SETTLM	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
Total	19771.4	35	8755.1	20000.0	10000	50000

The estimated amount of damage in the Česma settlement averaged 20 926 KM and in the Budžak settlement 15 857 KM (Table 2). Looking at the amount of damage per settlement, no statistically significant difference was obtained (U = 74,500, z = -1,350, p = 0.177) between the settlement Česma (N = 27, Md = 20,000) and the settlement Budžak (N = 8, Md = 17,500).

Table 3. Population awareness of flood protection measures

		POPULATION AWARENESS			
		YES	NO	PARTIAL	Sum
PLOT	ČESMA	7	8	14	29
	BUDŽAK	0	4	5	9
Total		7	12	19	38

From Table 3 we conclude that half of the population (50%) living in endangered areas are partially informed about flood protection measures.

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.

Based of all of the above, we conclude that it is necessary to put pressure on appropriate state institutions to undertake larger measures of protection, such as to organize lectures or courses for the population in order to inform themselves about appropriate flood protection measures.
In the future it is necessary to take into account the facilities being built on the banks of the rivers, giving building permits for such facilities and assessing the risk of construction.

5. REFERENCES

[1] Imamomić, A. (2015). Causes of floods in the Bosna river sasin with special reference to the floods in May 2014, *Proceedings of the Symposium on flood risk management and mitigation*, 131-145.

[2] Đurđević, N., Kosić-Jeremić, S., Maksimović, S. (2019). Analysis of settlements inscreased by floods in the area of Banja Luka, *Proceedings of XLVI International Symposium on Operations Researc SYM-OP-IS 2019*, 571-576.

[3] Đurđević, N., Maksimović, S., Kosić-Jeremić, S., Kolaković, S. (2020), Vulnerability of caosal settlements to floods Banja Luka case of study, *Proceedings of the Faculty of Technical Sciences*, Novi Sad, 380-382.

[4] Feber, M. H. (2012). *Statistics and Probability Theory*, New York: Springer.

[5] Jakšić, S., Maksimović, S. (2020). *Probability and Statistics - theoretical basis and examples*, Banja Luka: Faculty of Architecture, Civil Engineering and Geodesy.

[6] Kattegoda, N.T., Rosso, R. (2008). *APPLIED STATISTICS FOR CIVIL AND ENVIRONMENTAL ENGINEERS*, Singapore: Utopia Press Pte Ltd

[7] Palant, J. (2009). *SPSS Priručnik za preživljavanje*, Mikro knjiga: Beograd.

[8] http://www.ba.undp.org/content/bosnia_and_herzegovina/bs/home/operations/projects/response_to_floods/support-to-flood-recovery-and-risk-mitigation-in-bosnia-and-herz.html

[9] http://www.msb.gov.ba/PDF/HRA_BHS_Final21122015.pdf

[10] https://rhmzrs.com/

[11] http://www.un-spider.org/risks-and-disasters/disaster-risk-management

6. QUESTIONS

1. What is the risk?
2. What is the random variable? Explain the diference between dicrete and continuous random variables.
3. What is the purpose of descriptive statistics?
4. What is the difference between parametric and non-parametric statistical tests? Give some examples of parametric and non-parametric statistical tests.
5. What tests we use most often in decision making in the risk management?