Theory part

**- What are statistics and its relationship with other disciplines**

Statistics is a term that comes from German (statistik) and means “description of a state”. It is a science that deals with data in various ways: collection, organization, processing/analysis, reporting/presentation, interpretation.

Although statistics can be defined a science such as physics and chemistry, it differs from these one because it does not use the scientific method to investigate phenomena: it is rather a sistematic body of knowledge used to study populations (groups of entities like people, animals, cars) through the collection and a statistical analysis of data.

Statistics and mathematics have a common item that consists of the formal mathematical language. Statistics use estensively the theory of probability (from mathematics).

The main goal of statistics is to derive knowledge from the information extracted from data: this process starts from a bunch of data (set of statistical units) from which a dataset is created (after deciding what attributes to observe and defining the “scale” of measurement of these ones), then different techniques can be used to “summarize” (to reduce complexity) the important aspects of the dataset and have a better understanding of the scenario studied.

References:

Statistics – Wikipedia: <https://en.wikipedia.org/wiki/Statistics>

**- Describe the concepts of variable, attribute, population, sample and dataset**

The words “variable,” “attribute,” “population,” “sample,” and “dataset” are all concepts used in statistics and data analysis.

* **Variable**: A variable is a characteristic or attribute that can be measured or observed. It can take on different values for different individuals or objects in a population or sample. For example, age, height, weight, and income are all examples of variables.
* **Attribute**: An attribute is a specific value or category that a variable can take on. For example, if the variable is “gender,” then the attributes could be “male” and “female.”
* **Population**: A population is the entire group of individuals or objects that we are interested in studying. For example, if we are interested in studying the heights of all people in the world, then the population would be all people in the world.
* **Sample**: A sample is a subset of the population that we actually measure or observe. For example, if we want to estimate the average height of people in the world, it would be impractical to measure everyone’s height. Instead, we could take a sample of people and use that to estimate the average height of the population.
* **Dataset**: A dataset is a collection of data that has been organized into some kind of structure. It can be thought of as a table with rows and columns, where each row represents an individual or object and each column represents a variable. Datasets are often used for statistical analysis and machine learning.

**- Briefly describe the main sampling techniques**

Statistical sampling is a technique used in statistics to extract a subset of elements (samples) from a population, to estimate the properties of the entire population. There are several sampling techniques, including:

**Random sampling**: is a statistical sampling technique in which each unit of the population has an equal chance of being drawn as sample. In other words, each element of the population has an equal chance of being selected as part of the sample. This method is considered the fairest and most impartial of the sampling techniques. Random sampling can be divided into two categories: simple random sampling and stratified sampling. In the first case, each unit of the population has the same probability of being drawn as a sample, while in the second case, the population is divided into homogeneous groups (strata) and a random extraction is carried out within each stratum.

References:

<https://it.wikipedia.org/wiki/Campionamento_statistico>

<https://www.okpedia.it/campionamento-statistico>

**Probability sampling**: is a statistical sampling technique in which each unit of the population has a non-zero probability of being drawn as sample. In other words, each element of the population has a chance of being selected as part of the sample. This type of sampling guarantees the representativeness of the sample, while in non-probability samples the results of survey cannot be generalized.

Probability sampling can be carried out according to the following methods:

Simple random sampling: each unit of the population has the same probability of being drawn as sample.

Stratified random sampling: the population is divided into homogeneous groups (strata) and a random extraction is carried out within each stratum.

Systematic sampling: a random element is chosen as the starting point and the other elements are chosen at regular intervals.

Cluster sampling: the population is divided into groups (clusters) and a random extraction of the clusters is carried out, followed by the random extraction of the elements within the selected clusters.

References:

<https://it.wikipedia.org/wiki/Campionamento_probabilistico>

<https://elearning.unite.it/pluginfile.php/189099/mod_resource/content/1/IIIa%20campionamento%2020_21.pdf>

**Reasoned sampling:** is a statistical sampling technique in which the choice of the sample profile is completely left to those who prepare the survey plan, on the basis of a substantive theory that accounts for the phenomenon under study. In other words, the population units are chosen based on specific criteria, such as the opinion of experts or literature, in order to select only those that best respond to the research objectives. Reasoned sampling is a very quick and inexpensive technique, since the selection of the sample is very streamlined. However, the sample size is usually set based on purely convenience criteria and does not guarantee the representativeness of the sample. Can be used when some characteristics of the target population or the object of the investigation are known a priori, so that the sampling units are selected in a targeted way. This technique can be useful when you want to obtain specific information on the population, but there is no need to generalize the results of the survey.

References:

<https://it.wikipedia.org/wiki/Campionamento_ragionato>

<https://www.docenti.unina.it/webdocenti-be/allegati/materiale-didattico/62086>

**Quota sampling**: is a statistical sampling technique in which the population is divided into homogeneous groups (quotas) based on one or more characteristics, such as age, sex, profession, etc. Subsequently, a random extraction is carried out within each quota, in order to obtain a representative sample of the total population. Is a non-probability sampling technique, as the population units are not selected randomly. However, this technique ensures that the sample is representative of the total population, as the quotas are chosen to reflect the proportions of the population for the characteristics considered.

References:

<https://it.wikipedia.org/wiki/Campionamento_per_quote>

https://www.greelane.com/it/scienza-tecnologia-matematica/scienze-sociali/quota-sampling-3026728/

**- Briefly describe the main experiment designs**

Experimental design refers to the design of an experiment, i.e. choosing how to assign participants to different groups in an experiment. There are several types of experimental designs, including:

Independent measures experimental design: participants are divided into two or more groups, each of which is subjected to a different experimental condition.

Repeated measures experimental design: participants are subjected to all experimental conditions, to evaluate the effects of different conditions within the same individual.

Experimental design with equivalent groups: participants is divided into homogeneous groups, to guarantee that the groups are equivalent for some characteristics relevant to experiment.

In addition, there are also other types of experimental designs, such as the longitudinal design and the transverse design. The experimental design is a fundamental phase of scientific research, since it allows us to control extraneous variables and evaluate the effects of the independent variables on the dependent variables 2. Furthermore, the experimental design allows us to establish causal relationships between the variables involved in the experiment.

References:

<https://www.simplypsychology.org/experimental-designs.html>

<https://www.scribbr.com/methodology/experimental-design/>

Practical part

- Download Visual Studio

- Write a program in C# or VB.NET that creates a window containing a single line, point, circle and triangle

- Write a program in JavaScript or TypeScript that creates a window containing a single line, point, circle and triangle