Filippo Olimpieri 1933529
Homework 1

1)What is Statistics and its relationship with other disciplines. Difference between Descriptive and Inferential Statistics.

Statistics is a discipline that is concerned with collecting and analyzing data to a specific purpose.
Statistics is related to many others disciplines, like for example is related to:

1. Mathematics: Statistics is grounded in mathematics, and many of its concepts and methods are rooted in probability theory, linear algebra, and calculus.

2. Social Sciences: Statistics is widely used to analyze data related to society and human behavior in general.

3. Computer Science: Statistics is used for data analysis and for machine learning.

4. Economics: Statistics is extensively used in economic analysis to study markets, economic trends, and financial data.

The difference between Descriptive and Inferential Statistics lies in their respective purposes.
Descriptive Statistics aims to describe a set of statistical units without making any guesses.
Inferential Statistics instead deals with describing a sample, which should be as representative as possible (achieved through sampling techniques and experimental design) to make guesses about the entire population.

In Inferential Statistics, we only have information about a small sample from the population, and by studying this sample, we attempt to make inferences about the entire population.
In Descriptive Statistics, we have complete knowledge of the entire population, and our goal is simply to provide a description.

2)Describe the concepts of Population, Sample, Attribute, Variable, Level of measurement and Dataset.

In Statistics, the Population refers to the entire group of individuals, items, or data units that are the subject of interest for a particular study. The population is the complete set from which a sample is typically selected.

A Sample is a subset of a larger population. It is a representative group of statistical units that are selected from the population for the purpose of drawing conclusions about the entire population.
Saying that the sample has to be representative means that it has to accurately reflect the characteristics and variations present in the larger group.

In statistics, an attribute is a characteristic shared by all the statistical units of interest. The attribute can also be called a feature.

In statistics, a Variable is a characteristic, or quantity that can take on different values. Variables are used to represent and measure various aspects of data or phenomena in a study.
It can be independent if it is manipulated in the experiment or dependent if it is observed during the experiment.
Variables can take different values, for example they can be boolean, numeric, multi-value and so on.

In statistics, the level of measurement refers to the nature and characteristics of the values that a variable can take on. There are four primary levels of measurement:

1. Nominal Level: data are categorized into distinct, non-numeric categories or labels, so you can only determine whether two categories are the same or different.

2. Ordinal Level: data can be ranked or ordered, but the differences between values are not meaningful or consistent. Ordinal data represent ordered categories, but the intervals between categories are not uniform.

3. Interval Level: data can be ranked and the intervals between values are consistent and meaningful, but there is no true zero point.

4. Ratio Level: data can be ranked, the intervals between values are consistent and meaningful, and there is a true zero point, which indicates the absence of the attribute being measured.

In statistics, a Dataset refers to a collection of data, typically organized into a structured format for analysis and study. The dataset can be represented using a table, in which the statistical units represent the rows, the attributes or features represent the columns and the variables are the different values possible for each cell of the table.

3)Briefly describe the main sampling methods

The main sampling methods are the followings:

1. Simple random sampling: each individual is chosen entirely by chance and each member of the population has an equal probability of being selected.

2. Systematic sampling: the statistical units are chosen at certain intervals from a fixed starting point.

3. Stratified sampling: the statistical units are divided in groups based on some shared characteristics, then are randomly selected with the same number of units from each group.

4)Briefly describe the main experiment designs.

The main experiment designs are the followings:

1. Independent measures: different groups of participants are exposed to different experimental conditions or treatments. Each group experiences only one of the experimental conditions, and the goal is to compare the results or outcomes across these groups to determine if there are any significant differences.
There are multiple groups (at least two) that are considered during this experiment.
Participants are randomly assigned to one of the experimental groups to reduce the bias and each group is tested separately under its assigned condition.

2. Repeated Measures Design: the same group of participants is exposed to all experimental conditions or treatments. In this design, each participant experiences every condition being tested in different sequences. This approach is in contrast to the "Independent Measures Design".
After data collection, statistical analyses are conducted to compare the results within the same group of participants.

3. Matched Pairs Design: It's a variation of the Repeated Measures Design and involves matching each participant in one group with a participant in another group based on certain characteristics. The goal is to reduce individual differences and increase the precision of the comparison between the groups.
Participants are initially paired based on specific characteristics that are relevant to the research question.
After matching, one member of each pair is randomly assigned to one group while the other member is assigned to the other group. This random assignment helps ensure that any differences observed between the groups are due to the treatment or condition being studied and not the initial differences between participants.
Both groups are exposed to the same experimental conditions or treatments, and data are collected on the outcome variable of interest.

Bibliography:

-https://en.wikipedia.org/wiki/Design_of_experiments
-https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/methods-of-sampling-population
-https://www.treccani.it/enciclopedia/probabilita-e-statistica_%28Enciclopedia-Italiana%29/
-https://www.simplypsychology.org/experimental-designs.html