

# STATISTICS: HOMEWORK 1

Studente: **Riccardo Vattelli**

Matricola: **1849199**

*Riccardo Vattelli*

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

Differenti formulazioni sono state proposte nel tempo circa la definizione di Statistica.

Secondo Seligman: "Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing and interpreting numerical data collected to throw some light on any sphere of enquiry." [1] King propone la seguente definizione: "The science of statistics is the method of judging collective, natural or social phenomenon from the 'results obtained from the analysis or enumeration or collection of estimates.'" [2] A.L. Bowley la definì, invece, in tal modo: "Science of measurement of the social organism regarded as a whole in all its manifestations." [2] Per il Cambridge Dictionary: "[Statistics is] the science of using information discovered from collecting, organizing, and studying numbers" [3]

La Statistica riveste un ruolo importante, talvolta fondamentale, in svariati settori e campi di studio, coadiuvandone e potenziandone gli sviluppi. Esempi lampanti sono costituiti dalle sfere dell'Economia, del Business and Planning, oppure delle Scienze Naturali e Sociali, fino a quelle dell'Astronomia o dell'Amministrazione statale. [4][5][6]

In Economia vengono impiegati metodi statistici per stabilire relazioni di domanda e richiesta o per distribuire incassi e guadagni o, ancora, per calcolare i tassi di inflazione.

Nel Business and Planning colui che si avvale di conoscenze statistiche riesce a ideare una pianificazione di svariati assets in differenti periodi e secondo differenti modalità per soddisfare esigenze e bisogni individuati a seguito di analisi di mercato. Altre attività quali trasporti, manifattura, assicurazioni, banking, etc. ricorrono a fasi di pianificazione ed esecuzione assistite da risultati statistici.

Nelle Scienze Sociali, per condurre indagini, sondaggi ed esperimenti sovente ci si avvale di tecniche statistiche quali campionamento di un sottoinsieme della popolazione da prendere in esame (Sampling), individuazione di relazioni, trend e comportamenti (Regression, Analysis of variance) e studio dei dati raccolti (Categorical data, Data mining).

Nell'Amministrazione, la Statistica si rivela utile per ciò che concerne decisioni amministrative a vari livelli, in particolare i fondi da stanziare e redistribuire, i salari da regolare o il personale statale da monitorare e dirigere.

In generale, la differenza tra Descriptive Statistics e Inference Statistics è sostanziale. [7][8]

Descriptive Statistics permette essenzialmente di descrivere caratteristiche dei dati, presentati sotto molteplici forme, in maniera tale da fornirne una rappresentazione quantitativa (numerica o grafica), a seguito di un processo di collezionamento e organizzazione delle informazioni di un intero dataset.

Aspetti chiave di questo tipo di Statistica sono rappresentati dalla Measure of Central Tendency (calcolo di moda, media e mediana), Measure of Dispersion or Variability (calcolo di deviazione standard, varianza e range), Measure of Distribution (calcolo di percentuali) e Visualizzazioni (creazione di istogrammi, grafici a torta o a barre verticali/orizzontali).

L'analisi permette di trarre conclusioni sul dataset senza ulteriori inferenze a più ampio raggio.

Inferential Statistics trae conclusioni deduttive dai dati provenienti da un campione (sample, randomico o meno) della popolazione prescelta, così da applicarle all'intero dataset.

Dunque, la principale differenza rispetto alla Descriptive Statistics risiede nel fatto che qualunque attività di Inferential Statistics (ipotesi, esperimenti, sondaggi, previsioni...) viene condotta su un sottoinsieme ragionevole delle unità statistiche, anziché su un dataset potenzialmente impossibile da esaminare nella sua totalità.

Le più rilevanti strategie adottate dalla Inferential Statistics includono Sampling Techniques (selezione di subset appropriati e rilevanti), Tests of significance o Hypothesis tests (processo di formulazione di ipotesi su un campione ed effettiva verifica dell'attendibilità di tali ipotesi sul determinato campione), Correlation Analysis (individuazione di relazioni/correlazioni tra variabili) e Regression Analysis (previsione di relazioni di causalità tra variabili), Confidence Intervals (stima della probabilità di un determinato risultato).

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

Population: insieme completo di elementi da cui estrapolare dati per fini statistici, in cui ciascun componente condivide almeno una caratteristica comune con gli altri (tali elementi possono consistere di persone, oggetti, eventi, attività, ...).

Secondo l'Australian Bureau of Statistics [9]: "A population is any complete group with at least one characteristic in common. Populations are not just people. Populations may consist of, but are not limited to, people, animals, businesses, buildings, motor vehicles, farms, objects or events."

Sample Attribute: anzitutto, un sample (campione) è un sottoinsieme di elementi, individuato nella totalità della popolazione, che contiene dati specificamente caratterizzati, in cui sia assente qualunque forma di condizionamento delle proprietà e dell'affidabilità dei componenti.

Un attributo si riferisce alla qualità di una certa caratteristica osservabile in un'entità, che distingue questa da un'altra. Attribute Sampling è definibile come "the method of measuring quality that consists of noting the presence (or absence) of some characteristic (attribute) in each of the units under consideration and counting how many units do (or do not) possess it" [10]

Variable: è una qualunque quantità contabile o misurabile, il cui valore può variare nello spazio e nel tempo tra le data units che compongono l'insieme. Esistono varie tipologie di variabili, ad esempio variabili numeriche (discrete o continue), variabili categoriche/qualitative, variabili randomiche, variabili dipendenti, indipendenti o di controllo.

Level of measurement: proprio delle variabili e utilizzato per distinguere variabili misurate con differenti proprietà. Esistono quattro levels of measurement basilari, classificati in base al grado di precisione: Nominal, Ordinal, Interval e Ratio. [11][12]

Nominal comprende variabili distinte tra loro per differenze qualitative (ad esempio: genere, partito d'appartenenza).

Ordinal organizza le variabili secondo una certa relazione d'ordine od ordinamento di misurazioni (ad esempio: classifica in una gara, grado di soddisfazione)

Interval pone le variabili in una relazione d'ordine e, al contempo, specifica anche la distanza tra di esse in un intervallo.

Ratio, similmente a Interval, colloca le variabili in una relazione d'ordine, specificandone la distanza in un determinato intervallo, ma include nella scala anche lo zero (punto di zero assoluto).

Dataset: collezione organizzata di dati. In base alla tipologia di variabili considerate, è possibile classificare in vario modo i tipi di dataset (Numerical se i dati in esame sono numeri, Categorical se le informazioni riguardano caratteristiche di una persona o di un oggetto, Bivariate e Multivariate qualora vengano scelte ogni volta, rispettivamente, una coppia di variabili o molteplici variabili).

### **3) Briefly describe the main sampling methods**

Nella scelta cruciale di campioni rappresentativi dell'intero gruppo, vengono adoperati diversi Sampling Methods. Una prima fondamentale distinzione tra di essi è rintracciabile nell'approccio probabilistico o non-probabilistico al campionamento, ovvero Probability Sampling (selezione randomica del sample) e Non-Probability Sampling (selezione non probabilistica sulla base di specifici criteri). [13][14]

Nell'ambito del Probability Sampling vengono generalmente proposti i seguenti metodi principali:

- Simple Random Sampling, in cui ciascun elemento della popolazione ha eguale probabilità di essere campionato;
- Systematic Sampling, secondo cui, similmente al Simple Random Sampling, vengono scelti randomicamente elementi, con la differenza, però, che vengono identificati intervalli regolari nei quali avviene tale scelta;
- Stratified Sampling, per cui la popolazione è suddivisa in strati, ovvero gruppi caratterizzati e distinti tra loro significativamente per qualche caratteristica (ogni gruppo ha singoli componenti scelti randomicamente);
- Cluster Random Sampling, nel quale, analogamente allo Stratified Sampling, sono individuati sottogruppi, ma ogni sottogruppo deve possedere caratteristiche simili a quelle del campione e deve essere campionato per intero randomicamente (anziché venir presi i singoli componenti);

Per ciò che concerne il Non-Probability Sampling, esistono diversi approcci basati su criteri non-randomici (pur tendenzialmente pronti a condizionamenti/bias), tra cui:

- Convenience Sampling, in cui la scelta del campione è effettuata secondo la maggiore facilità d'accesso agli elementi da parte di chi conduce il campionamento (influenzabile in tal modo da sampling bias e selection bias, ad esempio);
- Voluntary Response Sampling, secondo cui i partecipanti (umani) inclusi nel campione vengono direttamente contattati da chi conduce il campionamento e volontariamente accettano di farne parte;
- Quota Sampling, simile allo Stratified Sampling, per cui sono individuati in maniera predeterminata non-randomica sottogruppi (strati) mutuamente esclusivi e vengono campionate unità fino al raggiungimento di una quota fissata;
- Snowball Sampling, caratterizzato da una reclutazione di partecipanti attraverso altri partecipanti;
- Purposive/Judgement Sampling, basato sul campionamento ragionato e, generalmente, mirato e specifico da parte di chi conduce il sampling, secondo il proprio giudizio dettato dall'esperienza;

#### **4) Briefly describe the main experiment designs**

Experimental Design o Design of Experiments (DOE) identifica il processo di ideazione e definizione di un esperimento statistico controllato, quantitativo e oggettivo, volto al collezionamento di dati per specifici obiettivi di ricerca (dimostrazione o confutazione di ipotesi precedentemente formulate). La raccolta dei dati permette di individuare e comprendere relazioni causali tra variabili d'interesse (o fattori di studio), una volta portate a termine le fasi principali dell'esperimento (formulazione di osservazioni, domande e ipotesi, applicazione di metodi e ottenimento di risultati). [15]

Customized Random Designs, Randomized Block Designs e Observational Studies costituiscono i tre principali designs of experiment utilizzati. [16]

Customized Random Designs prevedono l'assegnazione di unità/partecipanti a gruppi mediante processi randomici, così da garantire un'eguale probabilità di partecipazione in un qualunque gruppo e livellare potenziali differenze sistematiche in e tra tali gruppi;

Randomized Block Designs consistono nella creazione di blocchi in cui raggruppare unità omogenee contraddistinte da una o più caratteristiche specifiche comuni (nuisance factors o nuisance variables), così da ridurre l'impatto e il margine d'errore, poi assegnate randomicamente a gruppi sperimentali;

Observational Studies o Quasi-Experimental Designs vengono impiegati qualora non fosse possibile assegnare randomicamente elementi a gruppi, perciò i partecipanti vengono osservati nei loro gruppi naturali d'appartenenza (precedentemente individuati e definiti dalla ricerca);

Esistono, inoltre, ulteriori approcci di experimental design, quali Between-Subjects Experimental Designs, in base ai quali i soggetti vengono assegnati solamente ad un particolare gruppo ed esposti ad una condizione di trattamento; Matched Pairs Experimental Designs, una variante dei Between-Subjects per cui soggetti con caratteristiche simili vengono accoppiati (matched pair) e, successivamente, assegnati randomicamente l'uno ad un treatment group e l'altro ad un control group, nel tentativo di livellare differenze preesistenti tra i gruppi; Within-Subjects Experimental Designs / Repeated Measures, in cui i partecipanti sono esposti ad ogni condizione sperimentale (tutti i soggetti in tutte le condizioni).

# References

- [1] Dr. Sanjeev Kumar, Associate Professor, Department of Economics, Chaudhary Charan Singh University, Meerut. *Introduction to Statistics: Definition, Nature, Importance and Limitations*  
[https://ccsuniversity.ac.in/bridge-library/pdf-23/E\\_Content\\_INTRODUCTION%20TO%20STATISTICS.pdf](https://ccsuniversity.ac.in/bridge-library/pdf-23/E_Content_INTRODUCTION%20TO%20STATISTICS.pdf)
- [2] Purva Joshi. *Statistics: Definitions, Characteristics and Classes*  
<https://www.biologydiscussion.com/biostatistics-2/statistics-definition-characteristics-and-classes-biostatistics/47440>
- [3] Cambridge Dictionary. *Definitions of 'Statistics'*  
<https://dictionary.cambridge.org/dictionary/english/statistics>
- [4] Petros Maravelakis, Department of Business Administration, University of Pireaus, Pireaus, Greece. *The use of statistics in social sciences*. Journal of Humanities and Applied Social Sciences, ISSN: 2632-279X, 30 October 2019  
<https://www.emerald.com/insight/content/doi/10.1108/JHASS-08-2019-0038/full/html>
- [5] eMathZone. *Importance of Statistics in Different Fields*  
<https://www.emathzone.com/tutorials/basic-statistics/importance-of-statistics-in-different-fields.html>
- [6] Pooja Mehta. *Relation of Statistics with other Sciences*  
<https://www.economicsdiscussion.net/articles/relation-of-statistics-with-other-sciences/2323>
- [7] Bradley University. *What's the difference between Descriptive and Inferential Statistics?*  
<https://onlinedegrees.bradley.edu/blog/whats-the-difference-between-descriptive-and-inferential-statistics/>
- [8] Illiyas Sha. *Descriptive Statistics: Definitions, Types, Examples*  
<https://www.analyticsvidhya.com/blog/2021/06/descriptive-statistics-a-beginners-guide/>
- [9] Australian Bureau of Statistics. *Population*  
<https://www.abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/population>
- [10] American Society for Quality. *Attribute & Variable Sampling Plans and Inspection Procedures*  
<https://asq.org/quality-resources/sampling/attributes-variables-sampling>
- [11] Portland State University. *Types of scales & levels of measurement*  
<https://web.pdx.edu/~newsomj/pa551/lecture1.htm>
- [12] Statistics Solutions. *Data Levels of Measurement*  
<https://www.statisticssolutions.com/dissertation-resources/descriptive-statistics/data-levels-of-measurement/>

[13] Shona McCombes. *Sampling Methods | Types, Techniques & Examples*. Published on September 19, 2019. Revised on June 22, 2023.

<https://www.scribbr.com/methodology/sampling-methods/>

[14] Khan Academy. *What are sampling methods?*

<https://www.khanacademy.org/math/statistics-probability/designing-studies/sampling-methods-stats/a/sampling-methods-review>

[15] Shelley Watts. *Experimental Design in Statistics*. Updated: 07/19/2022.

<https://study.com/learn/lesson/experimental-design-statistics-uses-process-examples.html>

[16] Jim Frost, Statistics by Jim. *Experimental Design: Definition and Types*.

<https://statisticsbyjim.com/basics/experimental-design/>