

INTRODUZIONE

Iniziando un Corso di Statistica è d'obbligo definire il contenuto della Statistica moderna e comprendere di cosa si tratti.

La Statistica si occupa della raccolta, presentazione ed elaborazione numerica delle informazioni, per agevolare l'analisi dei dati ed i processi decisionali.

La Statistica è una tecnica speciale per lo studio dei fenomeni di massa o collettivi, intendendo per fenomeni di massa o collettivi i fenomeni descritti da una massa o collezione di osservazioni.

Trattasi, quindi, di una tecnica di analisi delle informazioni e non di una scienza come l'Economia, la Sociologia, la Fisica o altre. Essa più che basarsi su postulati, assiomi e definizioni rappresenta una raccolta di metodi attraverso i quali si cerca di analizzare una massa di informazioni relative allo stesso fenomeno, per trarre da tale analisi gli elementi che possano far comprendere quanto il fenomeno analizzato rispecchi nelle sue diverse manifestazioni quelle regole e leggi comportamentali che sono alla base delle scienze nel cui ambito il fenomeno viene studiato.

Un secondo elemento va evidenziato in questa definizione e cioè che l'analisi statistica di un fenomeno in tanto ha un senso in quanto il fenomeno

stesso abbia la possibilità di manifestarsi e si manifesti secondo modalità tra loro diverse e soprattutto in un gran numero di casi diversi.

La sua iniziale connotazione di strumento da applicare alle osservazioni concrete, con il tempo si è arricchita di ricerche a contenuto teorico che, nella fase preliminare di studio, potevano in qualche modo prescindere dal riferimento diretto alla realtà per dedicarsi allo studio di problemi anche astratti, ma che hanno sempre condotto immancabilmente a sviluppi di carattere applicativo.

Questa integrazione tra origini legate allo studio della realtà fenomenica e successivi sviluppi centrati su speculazioni di carattere teorico, più o meno astratto, è stata possibile mutuando alcuni sistemi e tecniche di indagine anche da altre discipline come la Matematica ed il Calcolo delle Probabilità, che hanno consentito di stabilire opportune relazioni tra caratteristiche statistiche basate su un dato insieme di elementi (o popolazione di unità statistiche) e caratteristiche analoghe determinate in base ad una parte soltanto (o campione) di unità statistiche appartenenti a quell'insieme complessivo.

Pertanto, l'intera materia trattata dalla Statistica moderna si può distinguere in due parti fondamentali:

- la *Statistica descrittiva*, la quale comprende l'*insieme dei metodi che riguardano la raccolta, la presentazione e la sintesi di un insieme di dati per descriverne le caratteristiche essenziali*;

- la *Statistica inferenziale*, che comprende l'*insieme dei metodi con cui si possono elaborare i dati dei campioni per dedurre omogeneità o differenze nelle caratteristiche analizzate, al fine di estendere le conclusioni alla popolazione da cui quei campioni sono estratti*.

Per questa ragione, il Corso inizia con la disamina preliminare di alcune definizioni e regole di calcolo combinatorio, la cui conoscenza costituisce la base per riuscire a misurare correttamente le probabilità con le quali gli eventi studiati possono presentarsi, si presentano o si sono presentati. Questo primo insieme di nozioni costituisce il nucleo portante del primo capitolo.

Nel secondo capitolo, invece, sono fornite le regole fondamentali del calcolo delle probabilità, le quali servono appunto a determinare non solo quali sono le condizioni che debbono essere soddisfatte affinché un certo

evento si manifesti, ma anche in qual modo si può misurare l'aspettativa dell'evento stesso, cioè il grado del suo verificarsi sotto condizioni prestabilite.

Nel terzo capitolo si cominciano ad affrontare le prime nozioni di Statistica attraverso lo studio degli elementi basilari della statistica descrittiva, che vanno dalla determinazione delle *scale di misura* e dalla rappresentazione grafica dei fenomeni, alla illustrazione delle caratteristiche specifiche più importanti considerate nella Statistica per descrivere sinteticamente una distribuzione, come le misure di tendenza centrale o *valori centrali* (medie ferme o analitiche e medie lasche o di posizione), la *variabilità* (momenti, deviazione standard) e la *forma* delle distribuzioni (asimmetria, appiattimento).

Nel capitolo quarto sono state descritte alcune tra le distribuzioni principali, discrete e continue, correntemente utilizzate per le analisi statistiche; per ognuna di esse vengono determinati i principali parametri caratteristici già richiamati in termini generali nel terzo capitolo e si esaminano anche le interconnessioni reciproche tra alcune di esse, ossia le condizioni che permettono di derivare da una di esse ognuna delle altre.

Fino a quest'ultimo capitolo, la trattazione della materia ubbidisce allo scopo di fornire gli strumenti di base per approfondire la conoscenza che abbiamo di un dato fenomeno, che nel corso dell'analisi viene considerato come se fosse isolato rispetto a tutti gli altri. Inoltre, l'interesse viene portato esclusivamente su fenomeni che si manifestano assumendo modalità diverse di natura quantitativa o esprimibili secondo una scala di numeri cardinali. Sono cioè tralasciate tutte le estensioni della tecnica statistica alle serie ordinali (ad esempio le serie cicliche come dati riferiti ai giorni della settimana oppure alle ore del giorno oppure alle stagioni nell'anno; le serie storiche come quelle di dati riferiti agli anni di calendario) ed a quelle sconnesse (ad esempio i dati riferiti al colore degli occhi oppure ai tipi di reato oppure alla tipologia delle opere pubbliche e così via).

Ma giunti a questo punto, si manifesta immediatamente la necessità di spingere oltre l'illustrazione delle tecniche, per considerare anche i problemi che si pongono se si vuole comprendere come i comportamenti di un

dato fenomeno possano essere determinati dal modo in cui si manifestano due o

più altri fenomeni (caso della dipendenza) da cui il primo può dipendere e quelli relativi a fenomeni che possono mutualmente influenzarsi nel loro modo di manifestarsi (caso della interdipendenza).

Di questi aspetti specifici si fanno carico i successivi capitoli.

Anzitutto, per completare lo studio degli aspetti principali di natura descrittiva che consentono di conoscere una distribuzione statistica, si affronta il problema della rappresentazione analitica delle variabili statistiche (capitolo V). Questo tema diviene particolarmente rilevante sia se si vuole ricercare una distribuzione teorica che ben si adatti alle osservazioni effettive del fenomeno investigato sia quando vi sia l'esigenza di tentare una interpretazione delle relazioni funzionali tra due o più caratteri tramite la verifica dell'adattamento ai dati rilevati di un modello interpretativo teorico formulato, in prima approssimazione, dal ricercatore (dipendenza e interdipendenza in senso matematico).

Quindi, prima di affrontare l'analisi della dipendenza e della interdipendenza in senso statistico si introduce il tema dell'inferenza statistica (capitolo VI), cioè di come ci si deve comportare quando per conoscere una data variabile o per comprendere le caratteristiche ed il funzionamento di certe relazioni tra variabili statistiche siamo costretti a ragionare basandoci *non* sull'esame di tutte le unità statistiche che formano la popolazione oggetto di studio (universo), ma solo tenendo conto delle informazioni che provengono da una parte più o meno modesta di tali unità (campione).

A questa parte della materia che cerca di illustrare dei criteri da adottare per ridurre al minimo la possibilità di fare scelte erronee sulla base dei risultati emersi dalle analisi, segue una illustrazione specifica di come si deve operare per valutare l'attendibilità della stima dei parametri caratteristici di una popolazione di unità statistiche, quando tale stima sia il risultato dell'esame di una porzione soltanto delle unità di quella popolazione (capitolo VII).

Quindi, una volta stabilite le modalità del passaggio dalle statistiche relative ad un dato carattere e tratte da un campione ai corrispondenti

parametri che lo stesso carattere ha nell'universo da cui il campione stesso

proviene, si affronta il problema dell'inferenza statistica applicata all'analisi delle relazioni tra due o più variabili statistiche. Questa parte del corso prende in considerazione sia il problema della programmazione degli esperimenti sia quello della dipendenza (regressione, analisi della varianza) e della interdipendenza tra variabili (correlazione, analisi della covarianza), non limitatamente al caso delle relazioni tra due variabili, ma anche con una attenzione specifica al caso di più variabili in generale (capitoli VIII e IX).

Nel capitolo X si affronta poi il problema del confronto tra distribuzioni, facendo riferimento tanto a distribuzioni di caratteri quantitativi quanto a distribuzioni di caratteri qualitativi. In riferimento a quest'ultimo aspetto, quindi, si introduce la nozione di associazione e si illustrano le tavole di contingenza.

Seguono, infine, altri tre capitoli (capitoli XI, XII e XIII) interamente dedicati alla statistica non parametrica: il capitolo XI dedicato ai metodi non parametrici per uno o due campioni; il capitolo XII relativo ai metodi non parametrici per più campioni ed il capitolo XIII che tratta delle misure di tendenza, associazione e correlazione non parametrica.