

Statistica

Lezione 1

Bice Cavallo



Indice

- ◆ Introduzione
- ◆ Caratteri, unità statistiche e collettivo
- ◆ Classificazione dei caratteri statistici
- ◆ Suddivisione in classi di un carattere quantitativo

Introduzione

- ◆ La **statistica** è la scienza che ha come fine lo studio di un "collettivo", per descrivere e prevedere un fenomeno.
- ◆ Libro di testo: STATISTICA metodologie per le scienze economiche e sociali. Simone Borra, Agostino Di Ciaccio. McGraw-Hill

Introduzione

- ◆ La **statistica descrittiva** ha come scopo quello di **sintetizzare** i dati attraverso
 - ◆ i suoi strumenti grafici (diagrammi a barre, a torta, istogrammi, boxplot)
 - ◆ e indici (indicatori statistici, indicatori di posizione come la media, di variazione come la varianza e la concentrazione, di correlazione, ecc.)che descrivono gli aspetti salienti dei dati osservati.

Introduzione

- ◆ **La statistica inferenziale** ha come obiettivo, invece, quello di fare affermazioni, con una possibilità di errore controllata, riguardo la natura teorica (la legge probabilistica) del fenomeno che si osserva. La conoscenza di questa natura permetterà poi di fare previsioni.

Introduzione

- ◆ Principali metodologie per la **rilevazione** e l'**acquisizione** di dei dati.
- ◆ Per mezzo di **un'indagine** totale, in cui tutte le unità della popolazione vengono rilevate, si può pervenire ad una completa descrizione del fenomeno oggetto di studio.
- ◆ Limitarsi a osservare un piccolo sottoinsieme della popolazione di interesse (**campionamento**)

Caratteri, unità statistiche e collettivo

- ◆ La statistica analizza in termini quantitativi i **fenomeni collettivi**, ossia i fenomeni il cui studio richiede l'osservazione di un insieme di manifestazioni individuali. Es.
 - ◆ Consumo di un determinato bene in un periodo fissato
 - ◆ Il reddito di un insieme di individui
 - ◆ Il peso di un gruppo di oggetti o di persone
 - ◆ ecc.

Caratteri, unità statistiche e collettivo

Nome	Età	Sesso	Titolo di studio	Attività	Peso (kg)	Punteggio esercizi
Rossi M.	32	M	laurea	occupato	72	65
Bianchi G.	39	F	laurea	occupato	55	55
Nicoletti C.	46	M	diploma	disoccupato	79	53
Marcelli F.	28	M	diploma	studente	63	78
Petrone A.	51	F	diploma	casalinga	64	21

Tabella 1.2.1 – *Un piccolo insieme di osservazioni.*

- ◆ Ogni riga corrisponde a un individuo (**unità statistica**) del quale sono state rilevate i seguenti **caratteri**:
 - ◆ Nome (lo identifica in maniera univoca), Età, Sesso, etc.
- ◆ Ogni carattere, assume in corrispondenza di ogni individuo, una determinata **modalità**
 - ◆ E.s. il carattere Peso assume la modalità 72 in corrispondenza dell'individuo Rossi M.
- ◆ Le modalità possono essere:
 - ◆ **numeriche** (es. età, peso, punteggio esercizi)
 - ◆ **non-numeriche** (es. sesso, titolo di studio, attività)

Caratteri, unità statistiche e collettivo

Si definisce **unità statistica** l'unità elementare su cui vengono osservati i caratteri oggetto di studio.



NOTA A volte la corretta definizione dell'unità statistica richiede una definizione convenzionale molto dettagliata che elimini possibili ambiguità. Per definire le *famiglie*, per esempio, possiamo utilizzare la definizione fornita dall'ISTAT (censimento 2001, ma anche altre definizioni sono possibili): *Per famiglia si intende un insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, coabitanti e aventi dimora abituale nello stesso comune (anche se non sono ancora iscritte nell'Anagrafe della popolazione residente del comune medesimo). Una famiglia può essere costituita anche da una sola persona.*

Un insieme di unità statistiche omogenee rispetto a una o più caratteristiche costituisce un **collettivo** statistico o una **popolazione**.

- ◆ Es. di unità statistiche: studenti, abbonati alla TV, automobili vendute, abitazioni
- ◆ Es. di collettivo statistico: la popolazione residente a Roma al censimento del 2001, le automobili vendute in un certo periodo.

Classificazione dei caratteri statistici

- ◆ Un carattere può assumere modalità differenti in corrispondenza delle diverse unità statistiche del collettivo. Le modalità debbono essere **esaustive e non sovrapposte**.
 - ◆ Esaustive: le modalità elencate devono rappresentare tutti i possibili modi di manifestarsi del carattere.
 - ◆ Non sovrapposte: ad ogni unità si può associare solo una modalità

Classificazione dei caratteri statistici

- ◆ Quando le modalità sono espresse numericamente, il carattere è detto **quantitativo**, altrimenti è detto **qualitativo**.
- ◆ Un carattere qualitativo viene distinto in:
 - ◆ Carattere **sconnesso** (o con scala nominale), se date due modalità è possibile affermare solo se queste sono diverse o uguali (es. sesso, attività, luogo di nascita, religione, colore degli occhi)
 - ◆ Carattere **ordinato** (o con scala ordinale), se date due modalità è possibile solo dare un ordine, specificando che una precede l'altra (es. titolo di studio)

Classificazione dei caratteri statistici

- ◆ Caratteri **ordinati rettilinei** che possiedono cioè una modalità iniziale ed una finale
- ◆ Caratteri **ordinati ciclici** che, avendo le modalità legate da un ordine naturale di successione, non hanno però vere e proprie modalità iniziali e finali, le quali possono essere fissate solo in modo convenzionale (es. direzione del vento, mese di nascita, giorno della settimana).

Classificazione dei caratteri statistici

- ◆ I caratteri quantitativi vengono distinti in:
 - ◆ Carattere quantitativo **con scala ad intervalli**. Non esiste uno zero assoluto, naturale e non arbitrario. (Es. la temperatura misurata in gradi centigradi. Infatti lo zero utilizzato è uno zero convenzionale e non uno zero assoluto, non ha senso dire “la temperatura 40° è due volte più calda che 20° ”, es. $0^{\circ} = 32F$, $20^{\circ} = 68F$, $40^{\circ} = 104F$).
 - ◆ Carattere quantitativo **con scala di rapporti** (es. peso, reddito, età, lunghezza).

Classificazione dei caratteri statistici

- ◆ I caratteri quantitativi vengono distinti anche in:
 - ◆ Carattere **quantitativo discreto** se l'insieme delle modalità assumibili può essere messo in corrispondenza biunivoca con un sottoinsieme dei numeri interi (quindi le modalità assumibili sono in numero finito o al più numerabile). Es. numero di figli, numero di pezzi prodotti, voto ad un esame.
 - ◆ Carattere **quantitativo continuo** se l'insieme delle modalità assumibili può essere messo in corrispondenza biunivoca con un sottoinsieme dei numeri reali. Es. peso, altezza.

Classificazione dei caratteri statistici

- ◆ Un carattere quantitativo è detto **trasferibile**, se ha senso immaginare che un'unità statistica possa cedere tutto o parte del carattere posseduto a un'altra unità statistica.
 - ◆ Es. reddito, numero di operai sono trasferibili
 - ◆ Es. peso, età non sono trasferibili

Classificazione dei caratteri statistici

Tipi di carattere

- quantitativo

- ┌ *scala a intervalli*
- └ *scala di rapporti*
- ┌ *discreto*
- └ *continuo*
- ┌ *trasferibile*
- └ *non trasferibile*

- qualitativo

- *sconnesso (scala nominale)*
- *ordinato (scala ordinale)*
 - *rettilineo*
 - *ciclico*

Suddivisione in classi di un carattere quantitativo

- ◆ Se il carattere che si vuole analizzare presenta **moltissime modalità distinte**, si possono avere notevoli difficoltà nella comprensione dei dati osservati. Es. settore di attività, altezza
- ◆ Accorpamento delle modalità:
 - ◆ Se il carattere è quantitativo si definisce **suddivisione in classi** del carattere l'operazione consistente nel suddividere l'insieme dei possibili valori **in intervalli tra loro disgiunti**

Suddivisione in classi di un carattere quantitativo

- ◆ Con tale operazione il carattere quantitativo passa da un livello di misura su scala di intervalli o scala di rapporti a **un livello ordinale**
- ◆ Si ottiene **immagine sintetica** della distribuzione empirica osservata anche se a prezzo di una certa **perdita di informazione**
- ◆ Es. il carattere “Età in anni compiuti” (che può assumere solo valori interi):
 - ◆ 0-5, 6-10, 11-14, 15-18, ecc.. (cioè $\{0,1,2,3,4,5\}$, $\{6,7,8,9,10\}$, $\{11,12,13,14\}$, $\{15,16,17,18\}$, ecc.)

Suddivisione in classi di un carattere quantitativo

- ◆ E' opportuno definire le classi in modo tale che:
 - ◆ Il loro numero sia abbastanza piccolo da fornire una sintesi adeguata ma sufficientemente grande da mantenere l'informazione con un livello accettabile di dettaglio
 - ◆ Siano tra loro disgiunte
 - ◆ Comprendano tutte le possibili modalità del carattere
 - ◆ Abbiano, se possibile, la stessa ampiezza
- ◆ Es. classi di diversa ampiezza
 - ◆ Il carattere "Età in anni compiuti": 0-5 (età prescolare), 6-10 (scuola elementare), 11-14 (scuola media), 15-19 (scuola superiore), 20-24 (università), 25-30 (inserimento nel mondo del lavoro), ecc. (cioè $\{0,1,2,3,4,5\}$, $\{6,7,8,9,10\}$, $\{11,12,13,14\}$, $\{15,16,17,18,19\}$, $\{20,21,22,23,24\}$, $\{25,26,27,28,29,30\}$, ecc.)
 - ◆ Si può notare che le classi sono state scelte considerando un criterio esterno al carattere.
 - ◆ Scelta che fornisce alle classi particolare interesse dal punto di vista interpretativo

Suddivisione in classi di un carattere quantitativo

- ◆ Nello stabilire gli estremi delle classi occorre tener presente che ognuna delle determinazioni del carattere deve essere compresa in una ed una sola classe.
- ◆ Nel caso di **caratteri quantitativi continui**, è necessario includere nella classe uno solo dei due estremi dell'intervallo, es:
 - ◆ Intervalli aperti a destra e chiusi a sinistra: $[130,150[$, $[150, 170[$, $[170, 190[$
 - ◆ Intervalli chiusi a destra e aperti a sinistra: $]130,150]$, $]150, 170]$, $]170, 190]$