

STATISTICA METODOLOGICA

a.a. 2020/2021

Prof.ssa Riviuccio Giorgia

giorgia.riviuccio@uniparthenope.it

8 settembre 2015

Data science e statistica le discipline «sexy»: danno lavoro a 9 laureati su 10

di Alberto Magnani

Il 90% dei laureati lavora in cinque anni

Non serve andare nella Silicon Valley per una stima su quanto offra una laurea nel settore. Gli ultimi dati di AlmaLaurea hanno eletto proprio la statistica, accorpata ad economia, come il quarto gruppo disciplinare più redditizio per gli studenti: tasso di occupazione del 90% a cinque anni dalla laurea e guadagno mensile netto di 1.487 euro. Un'asticella che si spinge al rialzo a seconda del ruolo svolto in azienda e, spesso, del paese di destinazione. Negli Stati Uniti il portale di lavoro CareerCast ha classificato gli “statistician” come la quarta miglior professione del 2015, con reddito medio di 79.191 dollari e stime di crescita del 25,91 per cento. Un ottimismo confermato dai suoi stessi numeri: le ricerche attive sul sito parlano di oltre 3.400 posizioni aperte per professionisti con più specializzazioni, dalla sanità agli organismi governativi. **Il futuro è nel data science**

Proprio il data science, salutato dall'Harvard Business Review come la «professione più sexy del 21 secolo», attinge a piene mani dalla statistica per la sua funzione decisiva: estrarre dati, farne un'analisi, trasformarli in una leva che aumenti l'efficacia del business (e i profitti). La definizione è ampia quanto le prospettive che si possono creare in azienda, a giudicare dal boom di richieste che sta percorrendo il mercato internazionale. Una ricerca di LinkedIn sui mestieri che «a malapena esistevano cinque anni fa» ha registrato un exploit della categoria nelle ricerche del social network: dalle 142 del 2008 alle 4.326. Trenta volte tanto, la quinta crescita più poderosa tra le new entry della piattaforma.

Linguisti, matematici, chimici, fisici, umanisti e psicologi in cima alla classifica dei disoccupati. Il titolo di studio però facilita la ricerca: il diploma e la licenza media non garantiscono l'impiego

Senza lavoro un laureato su cinque

Posti sicuri per esperti di statistica

IN 5 PUNTI

Andrea Scutellà / TORINO

Un 30enne laureato su cinque è disoccupato e un altro svolge un mestiere per cui il titolo non serve. Solo il 61,5% dei "dottori" riesce a far fruttare la sua istruzione nel mondo del lavoro. Sgomberiamo il campo da equivoci: chi ha studiato di più lavora di più e meglio di chi si è fermato prima. Ma il tasso di occupazione dei laureati è in calo. Accade anche - ma non sempre - per l'incremento del loro numero, come sottolinea il dossier dell'Osservatorio dei consulenti del lavoro.

Pesa il gender gap

Nel mondo del lavoro la laurea conta. Chi la con-

segue ha l'8% in più di possibilità di trovare un impiego rispetto ai diplomati, che a loro volta hanno una quotazione più alta del 15,9% di chi ha una licenza media. Ma conta soprattutto il tipo di titolo conseguito: se con Statistica - favorita dal boom della Gig economy - il tasso di disoccupazione è del 3,7% con Lingue sale al 26,8%. Le laureate tra i 30 e 39 anni hanno più difficoltà dei loro colleghi maschi: lavora il 77,3% contro l'87,8% degli uomini. Una differenza che diventa abissale nell'istruzione superiore (61% contro 84,6%) e media (37% contro 57,4%).

2 Più iscritti, meno lavoro

Dal triennio 2012-2014 a quello 2015-2017 il tasso di occupazione dei laureati è diminuito in



Un giovane controlla la bacheca delle offerte di lavoro in un centro per l'impiego. Ma anche per i laureati la strada è in salita

maniera direttamente proporzionale al loro aumento: dello 0,6%. Il numero dei trentenni a lavoro scende in quasi tutti i settori disciplinari, ad eccezione di Statistica (+10,9%), Medicina (+0,5%) e Lettere, Filosofia e Storia - raccolte, discutibilmente, sotto la categoria unica di "Insegnamento" - (+0,6%). In linea con i dati è solo l'aumento degli iscritti alle facoltà di Statistica, con un +5,5%, mentre quello dei medici è sproporzionato: ben 55mila studenti in più rispetto al triennio (+34%). Ma i 15mila iscritti a Psicologia in più dovranno fronteggiare un calo dell'occupazione dell'1,4%, i 19mila delle Scienze Sociali dell'1% e i 21mila di Scienze Economiche dello 0,8%. Un caso a parte è Giurisprudenza, dove le iscrizioni sono calate dell'8,2% insieme al tasso di occupazione (-1,1%).

3 Gli scienziati disoccupati

La maglia nera dei non occupati rispetto agli iscritti spetta a Lingue (26,8%), seguono i trentenni laureati in Scienze, chimiche, fisiche e matematiche (25,4%) e quelli della macro categoria Insegnamento (24,9%), che in senso assoluto ha il maggior numero di disoccupati: 71.691. Al di sopra della media nazionale (19,5%) anche Psicologia

(22,7%), Scienze sociali (22,6%), Giurisprudenza (22,5%) e Architettura (21,4%).

4 Il caso dei sovraistruiti

Diffuso è anche il fenomeno dei laureati sovraistruiti, impiegati in settori lavorativi che non richiederebbero la laurea. I più penalizzati sono i laureati in Scienze sociali con il 33% di laureati che non lavorano in un campo attinente agli studi. Seguono ad una incollatura il 32,6% degli ex studenti di Lingue: solo il 40,7% di loro ha un impiego nel settore. Più staccati i trentenni di Scienze Economiche (28%) e del filone Insegnamento (19,5). Particolare è il caso di Farmacia che si attesta al 15,9%, ma con un incremento del 7% negli ultimi tre anni.

5 Gli stipendi dei laureati

Alla fine gli studi pagano: i laureati trentenni guadagnano in media 1.632 euro, il 30% in più rispetto a chi ha la licenza media e il 20% dei diplomati. Medicina (1.869 euro), Ingegneria (1.850 euro) e Statistica (1.749 euro) sono le lauree meglio retribuite. In fondo alla classifica linguisti (1.420 euro), insegnanti (1.387 euro) e psicologi (1.351 euro). —

“La statistica descrive i metodi per raccogliere e trasformare i dati (le informazioni) in modo da renderli utile supporto per i processi decisionali” (Levin, Krehbiel, Berenson, 2006).

Statistica Descrittiva

Statistica Inferenziale

- In relazione agli obiettivi dell'analisi statistica, si distingue la statistica descrittiva dalla statistica inferenziale.
- Statistica descrittiva
Sintetizza le informazioni raccolte (da una popolazione o da un campione) in modo tale da avere una “descrizione” del fenomeno di interesse.
- Statistica inferenziale
Utilizza le informazioni raccolte da un campione al fine di trarre delle conclusioni sulla popolazione.

FASI:

- ❑ Raccolta dati: indagini statistiche.
- ❑ Presentazione dati: tabelle e grafici.
- ❑ Sintesi e descrizione dati: statistiche descrittive.
- ❑ Analisi dati: esplorazione e previsione.
- ❑ Formulazione ipotesi e confronto con la realtà:
inferenza.

Testi consigliati

- ❑ Levine D.M., Krehbiel T.C., Berenson M.L. (ult. ed), Statistica, Pearson
- ❑ Borra S., Di Ciaccio A. (2014), Statistica (terza edizione), McGraw-Hill

Programma e materiale didattico:

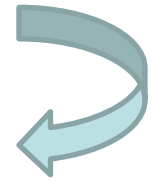
❑ www.docenti.uniparthenope.it

sezione docenti: Riviaccio Giorgia

❑ [Piattaforma e-Learning](#)

Che cos'è la Statistica?

È l'insieme delle tecniche e degli strumenti che consentono di analizzare e sintetizzare in termini quantitativi i **fenomeni collettivi**



il cui studio richiede l'osservazione di un insieme di manifestazioni individuali

Qual è lo scopo della Statistica?

Rilevare (attraverso la raccolta dei dati)
misurare (attraverso la loro elaborazione) e
interpretare (attraverso la loro utilizzazione) le
manifestazioni dei **fenomeni collettivi** che si
realizzano in un determinato istante temporale
o in uno specifico contesto territoriale

Il “raggio d’azione” della Statistica

I campi di applicazione della Statistica sono svariati. Ecco alcuni esempi.

1. Ambito demografico.

Raccolta di informazioni riguardanti nascite, decessi, cambiamenti di residenza nel Comune di Salerno nel quinquennio 1999-2004. Tali informazioni possono essere utilizzate per studiare la struttura per età dei residenti. Inoltre, scopo più pragmatico, per prevedere il numero di iscritti alle scuole nei prossimi anni.

Il “raggio d’azione” della Statistica

2.Ambito aziendale.

Raccolta di informazioni che riguardano il peso (o altre caratteristiche) dei prodotti di un’azienda, al fine di evitare un deterioramento degli standard qualitativi.

Oppure...

Indagine sul grado di soddisfazione dei clienti di un’azienda. I risultati possono essere utilizzati per attuare cambiamenti di politica aziendale.

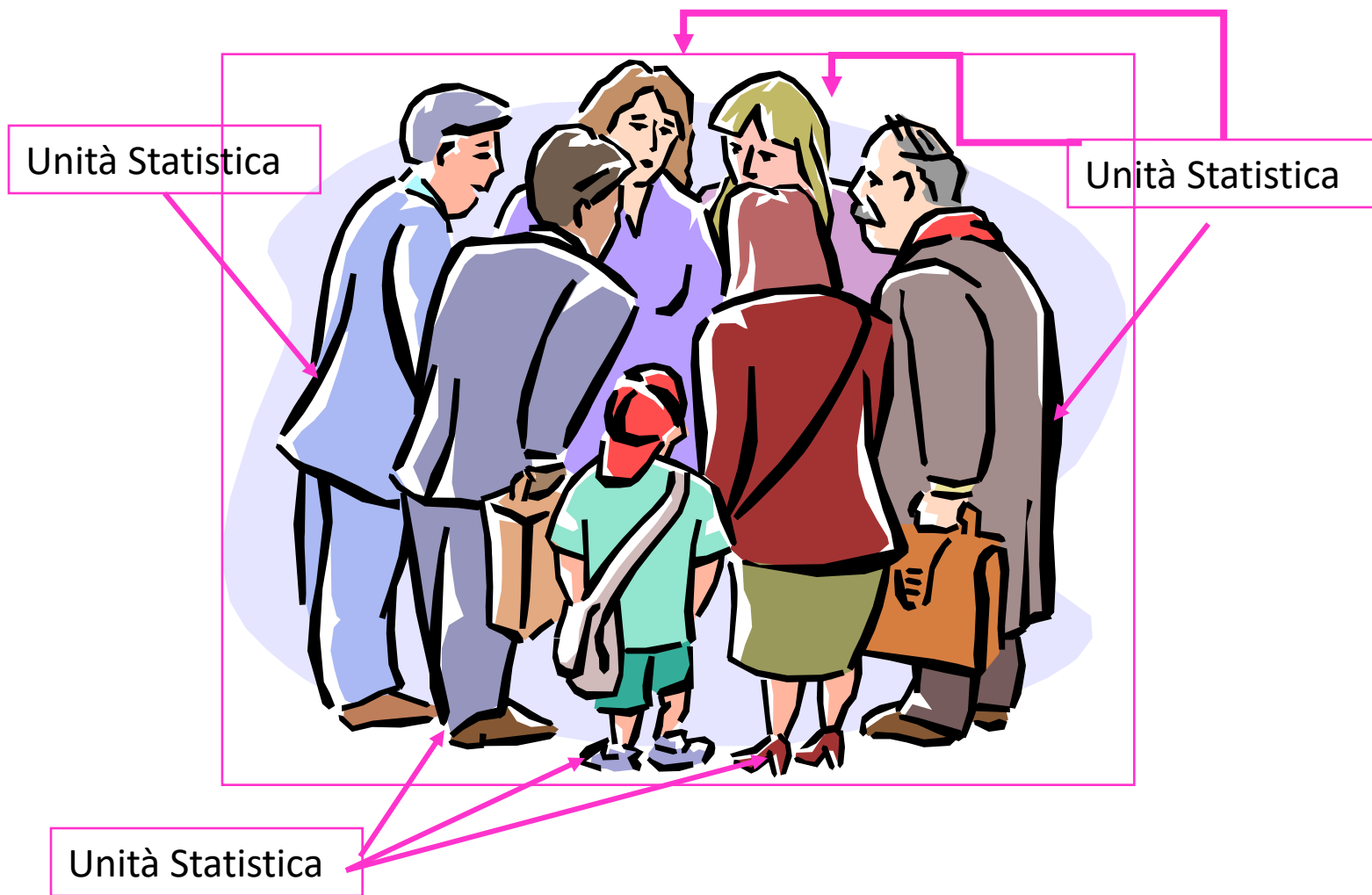
IN ALTRE PAROLE... SI TRATTA DI STUDIARE L'ANDAMENTO DI FENOMENI COINVOLGENTI UN INSEME DI UNITA', ACCOMUNATI DA UNA O PIU' CARATTERISTICHE...

...CHE SI IDENTIFICANO COME UN **GRUPPO** O, UTILIZZANDO UN LINGUAGGIO PIÙ APPROPRIATO ALLA STATISTICA, COME UN ...



...COLLETTIVO!

INTRODUZIONE ALLA STATISTICA



COLLETTIVO: 7 UNITA'

Cosa sono questi nuovi termini?

- L'**unità statistica** è l'unità elementare su cui vengono osservati i caratteri oggetto di studio, ad esempio: il residente, il prodotto finito o il cliente.
- Un insieme di unità statistiche omogenee rispetto ad una o più caratteristiche costituisce un **collettivo statistico**, **definito anche "popolazione"**. Esempio di popolazione: l'insieme degli individui residenti nel Comune di Salerno nel periodo 1999-2004 oppure l'insieme dei prodotti finiti dell'azienda ics oppure l'insieme dei clienti della società ipsilon.

Altri termini, che vedremo nel dettaglio successivamente, sono:

- **Campione**: sottoinsieme della popolazione selezionato per la raccolta delle informazioni quando la popolazione è molto estesa. Le unità della popolazione che rientrano nel campione sono dette unità campionarie
- **Parametro**: misura di sintesi che descrive una caratteristica dell'insieme delle unità statistiche (nella **popolazione**). Esempi di parametro: variazione annua dei residenti oppure quota di prodotti con un peso inferiore a quello dichiarato sulla confezione oppure il grado di soddisfazione medio.
- **Statistica**: misura di sintesi che descrive una caratteristica dell'insieme delle unità campionarie (nel campione).

Fonti dei dati

Come reperire le informazioni di cui si ha bisogno?

Si indicano 3 principali possibilità:

1. Dati già pubblicati (SISTAN)
2. Esperimento
3. Sondaggio

Sistema Statistico Nazionale (SISTAN)

Le rilevazioni statistiche di interesse pubblico (le cosiddette statistiche ufficiali) vengono stabilite nel Programma Statistico Nazionale (PSN), predisposto annualmente dall'Istat ed approvato da Governo e Parlamento.

La sua realizzazione è affidata al Sistema Statistico Nazionale (SISTAN), composto da una rete di Uffici di statistica operanti all'interno delle diverse istituzioni pubbliche e coordinati dall'Istituto Nazionale di Statistica (ISTAT).

SISTAN

Gli organi del SISTAN sono:

-ISTAT

- Uffici delle Camere di Commercio e dei Comuni (+ 100.000 abitanti)
- Uffici di statistica centrali e periferici delle amministrazioni dello Stato e delle aziende autonome coordinati dall'Istat
- Uffici di statistica delle regioni e delle province autonome
- Uffici di statistica delle province, delle prefetture e dei comuni (-100.000 abitanti) e Aziende Sanitarie Locali (ASL)
- Uffici di statistica di enti pubblici
- Altri enti individuati con decreto dal Presidente del Consiglio (ISAE)

Funzioni dell' Istat

- Predisposizione del PSN
- Esecuzione di censimenti e di altre rilevazioni
- Ricerca e studio dei risultati delle rilevazioni
- Pubblicazione e diffusione dei dati delle proprie rilevazioni e degli uffici del SISTAN che non possono provvedervi direttamente
- Esecuzione di rilevazioni per enti o privati
- Promozione di studi e ricerche statistiche
- Coordinamento attività e assistenza tecnica degli enti del SISTAN
- Predisposizione nomenclature e metodologie di base per la classificazione e rilevazione dei fenomeni di carattere demografico, sociale e economico
- Informatizzazione archivi e formazione addetti

Principali pubblicazioni dell'Istat

- **Annuario statistico italiano:** raccolta di tavole statistiche relative a tutti i fenomeni oggetto di rilevazione (dalla popolazione al sistema economico, dall'ambiente alle condizioni delle famiglie, ecc.).
- **Rapporto annuale sulla situazione del Paese:** risultati di approfondite analisi sulla realtà economica, sociale ed ambientale del paese.
- **Banche dati on line:** censimenti, indicatori territoriali, previsioni congiunturali, commercio estero, ecc...

Le rilevazioni e le relative informazioni di sintesi possono essere fornite a livello locale o a livello nazionale, attraverso indagini censuarie o campionarie, attraverso interviste dirette o indirette

Le indagini statistiche

- a) Censimenti: rilevazione delle informazioni su tutte le unità del collettivo d'interesse. Nelle rilevazioni statistiche ufficiali come l'indagine nazionale sui bilanci familiari, le unità di rilevazione sono costituite da tutte le famiglie italiane. SVANTAGGI: **costoso, lungo.**

- b) Campionamenti: rilevazione dei dati su una parte delle unità statistiche che compongono il collettivo. Nell'indagine sui consumi a livello regionale viene intervistata solo una parte del totale delle famiglie residenti in quella regione. SVANTAGGI: **errori.**

Il campionamento

Il campionamento è la procedura con la quale si seleziona un campione.

L'utilizzo di un campione risponde a tre fondamentali esigenze:

- **Risparmio di tempo**
- **Risparmio di costi**
- **Praticità nella gestione**

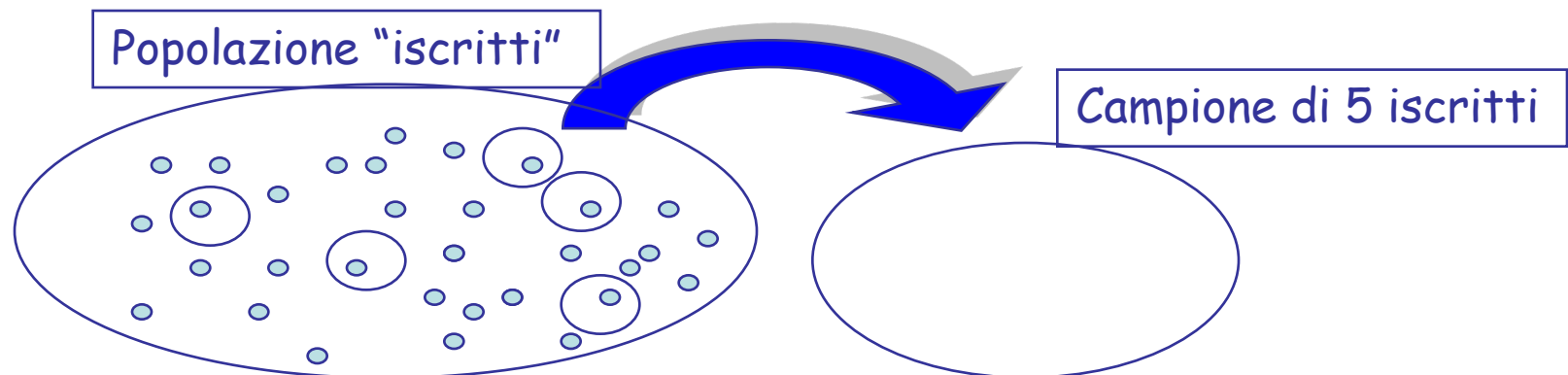
Per poter indagare le abitudini di consumo nei bar degli iscritti all'Università Parthenope si hanno due strade alternative:

- chiedere ad ognuno di essi cosa preferisce abitualmente consumare!?!?

Troppo faticoso in termini di tempo!

...oppure...

- ...è possibile intervistare UN CAMPIONE di studenti (5 ad esempio), estratto dalla popolazione "iscritti all'Università Parthenope", che si possa ritenere rappresentativo dell'intera popolazione



...In che senso rappresentativo? ...Come scegliere quel campione?

Un campione è da considerarsi rappresentativo se la proporzione di unità nel campione che possiedono certe caratteristiche è approssimativamente uguale alla proporzione di unità nella popolazione con le stesse caratteristiche.

Si distinguono campionamenti:

- **NON PROBABILISTICI**

- A scelta ragionata
- Per quote

- **PROBABILISTICI**

Campioni non probabilistici

Si parla di campione non probabilistico (o a scelta ragionata) quando

- la selezione delle unità campionarie viene affidata al rilevatore o ad un esperto del fenomeno
- autoselezione del campione (sondaggi su Internet)

Detto **non probabilistico** perché non si prende in considerazione la probabilità associata ad ogni unità di appartenenza al campione.

CAMPIONAMENTO

Campionamento
non probabilistico

```
graph TD; A["Campionamento non probabilistico"] --- B["A scelta ragionata  
(autoselezione / scelta dell'intervistatore)"]; A --- C["Per quote  
(scelta dell'interv. nel rispetto dei vincoli di quota)"];
```

A scelta ragionata
(autoselezione /
scelta dell'intervistatore)

Per quote
(scelta dell'interv.
nel rispetto dei
vincoli di quota)

Campioni probabilistici

Si parla di campione probabilistico quando le unità campionarie sono scelte in considerazione della probabilità di appartenenza al campione.

Si suddividono in:

- **Campionamento casuale semplice**
- **Campionamento sistematico**
- **Campionamento stratificato**
- **Campionamento a grappolo**
- **Campionamento a due stadi**

Il campionamento casuale semplice

Nel campionamento casuale semplice tutte le unità della popolazione hanno la stessa probabilità di far parte del campione.

Si indica con

N = numerosità della popolazione

n = numerosità del campione

Bisogna selezionare (estrarre) n elementi dagli N elementi della popolazione.

Popolazione e campione (esempi)

Problema da affrontare	Popolazione	Campione
Il preside vuole valutare il rendimento universitario degli iscritti alla Facoltà di Economia	Tutti gli studenti iscritti alla Facoltà di Economia	Un gruppo di (per esempio) 100 studenti della Facoltà di Economia
Si vuole conoscere l'orientamento di voto in vista delle elezioni	Tutti gli aventi diritto al voto	Un gruppo di (diciamo) 1000 elettori
Un'azienda per il Turismo vuole monitorare l'offerta ricettiva locale	Tutti gli esercizi ricettivi (alberghi, campeggi,...) della zona	Un gruppo di (ad esempio) 60 esercizi ricettivi locali

Errori nelle indagini campionarie

- Distorsione nella selezione del campione o errore di copertura

Lista non aggiornata

- Mancata risposta

- Errore di campionamento

Differenza tra campione e popolazione

- Errore di misurazione

Errori nel questionario o influenza dell'intervistatore

Somministrazione del questionario

- Intervista faccia a faccia: costosa, ma assicura una risposta più precisa alle domande del questionario.
- Intervista postale: economica, ma il grado di copertura è generalmente basso.
- Intervista telefonica: economica, presenta qualche problema di copertura. Oggi, però, è la preferita!
- Indagini via web: autoselezione del campione o invio questionari via mail, economiche ma troppo lente e danno luogo a errori di mancate risposte. In fase di studio.

Elaborazione del questionario

...Come si raccolgono tutte le informazioni ottenute dal questionario...???

...Si inseriscono i dati in una tabella, definita “**MATRICE DI DATI**”

Una matrice di dati ha per riga gli individui intervistati e per colonna le caratteristiche rilevate su ognuno di essi

Le caratteristiche vengono definite in statistica con il termine “**caratteri**”

Matrice dei dati

CARATTERI
QUANTITATIVI

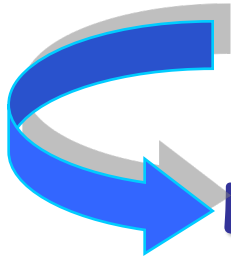
CARATTERI
QUALITATIVI

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	72	55
Marcelli F.	32	M	Diploma	Studente	63	78
Petrone A.	51	F	Diploma	Disoccupato	64	21

I modi di essere dei caratteri, in statistica si definiscono **modalità!**

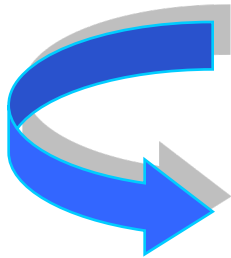
Tipologia di caratteri

Qualitativi



Le possibili **modalità** di risposta sono espresse sotto forma di **parole**

Quantitativi



Le possibili **modalità** di risposta sono espresse sotto forma di **numeri** (interi o decimali)

Tipologia di caratteri

Qualitativi

Utilizzate internet nel tempo libero?

- Si
- No

Modalità di
risposta non
ordinabili

**CARATTERE QUALITATIVO SCONNESSO O
SU SCALA **NOMINALE****

Quanto lo utilizzate?

- poco
- abbastanza
- tanto

Modalità di risposta che
possono prevedere un
ordine logico-naturale

**CARATTERE QUALITATIVO ORDINATO O SU
SCALA **ORDINALE****

Tipologia di caratteri

Quantitativi

Quanti libri avete letto nell'ultimo anno?

- 1-3
- 4-6
- 7-10

→ Numeri interi, derivano da un processo di conteggio

CARATTERE QUANTITATIVO DISCRETO

Quanto spendete in media al mese per le vostre telefonate?

- 25-30
- 30-50
- 50-80

→ Numeri decimali, derivano da un processo di misurazione.

CARATTERE QUANTITATIVO CONTINUO

Classificazione dei caratteri statistici

Alcune volte un carattere **continuo** può essere trattato come se fosse **discreto** a causa del metodo di misurazione (esempio: il PIL misurato in milioni di euro senza l'uso di decimali)

Un carattere quantitativo si dice **trasferibile** quando la sua intensità può essere trasferita da un'unità all'altra (esempi: reddito, fatturato)

Classificazione dei caratteri (autovalutazione)

All'uscita di una libreria, supponete di venire coinvolti in un sondaggio dove vi viene chiesto:

- Se avete comprato qualcosa
- Se sì, quanto avete speso (*somma spesa*)
 quanti libri avete acquistato (*numero di libri*)
 come avete pagato (*mezzo di pagamento*)
- Qual è la vostra condizione lavorativa o non (*attività principale*)

Per ciascun carattere, stabilire se è qualitativo (nominale o ordinale) oppure quantitativo (discreto o continuo)

Classificazione dei caratteri (autovalutazione)

Supponete che un cliente di una banca abbia fornito le seguenti informazioni per la richiesta di un mutuo per la casa:

- Spese mensili: 1427 euro
- Numero di impieghi negli ultimi 10 anni: 1
- Reddito da lavoro annuale familiare: 80000 euro
- Stato civile: coniugato
- Numero di figli: 1
- Residenza: Napoli
- Proprietà di altre case: no

Per ciascun carattere, stabilire se è qualitativo (nominale o ordinale) oppure quantitativo (discreto o continuo)

Matrice dei dati

È immediato notare che alcune informazioni, lette per colonna, si ripetono. Ad es. leggendo la colonna riferita al sesso:

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	72	55
Marcelli F.	32	M	Diploma	Studente	63	78
Petrone A.	51	F	Diploma	Disoccupato	64	21

M si ripete 3 volte!!!

F si ripete 2 volte!!!

I diplomati sono 3 e i laureati sono 2!

Elaborazione del questionario

Si può parlare in questo caso di frequenze con cui le **modalità** di ogni carattere si ripetono

Ma, in ogni caso...le tabelle non possono essere commentate in maniera così semplicistica, è necessario esprimere le osservazioni in modo più opportuno riscrivendo i dati in tabelle ed esplicitando le **frequenze** (“**assolute**”), cioè il numero di volte in cui ogni caratteristica (modalità) si ripete

Vengono create, pertanto, delle nuove sottotabelle dalla precedente e i FENOMENI (CARATTERI) SONO ESPRESSI SOTTO FORMA **DI DISTRIBUZIONE DI FREQUENZA**

Tabelle e frequenze assolute

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	72	55
Marcelli F.	32	M	Diploma	Studente	63	78
Petrone A.	51	F	Diploma	Disoccupato	64	21



Sesso
M
F
M
M
F

SUCCESIONE DI DATI



Sesso	frequenze
M	3
F	2

**DISTRIBUZIONE DI
FREQUENZA**

Tabelle e frequenze assolute

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	72	55
Marcelli F.	32	M	Diploma	Studente	63	78
Petrone A.	51	F	Diploma	Disoccupato	64	21



Sesso	frequenze
M	3
F	2

Titolo di studio	frequenze
Laurea	2
Diploma	3

Peso (kg)	frequenze
72	2
55	1
63	1
64	1

Leggendo le nuove tabelle perdiamo le informazioni relative ai singoli individui: prima era chiaro che Rossi era un uomo di 32 anni, laureato, occupato, con un peso pari a 72 kg e un punteggio di 65...ora sappiamo che nel nostro collettivo intervistato (5 persone) 2 sono laureati, 3 sono maschi...

Totale delle frequenze

Sesso	frequenze
M	3
F	2
totale	5

Titolo di studio	frequenze
Laurea	2
Diploma	3

Peso (kg)	frequenze
72	2
55	1
63	1
64	1

Il totale generico delle persone intervistate (unità statistiche) e quindi la somma delle frequenze assolute viene indicata con il simbolo "n"

Nella prima tabella $n = 5$

Nella seconda tabella $n = 5$

Nella terza $n = 5$

Frequenze relative e percentuali

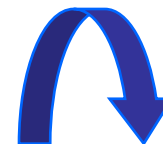
È possibile, poi, trarre informazioni più dettagliate che ci consentano di descrivere meglio l'andamento delle 3 variabili (peso, titolo di studio e sesso)

Ad es., per il sesso si ha:

Sesso	frequenze
M	3
F	2
totale	5



Sesso	frequenze	frequenze relative
M	3	$3/5=0.6$
F	2	$2/5=0.4$
totale	5	1



3 su 5 sono maschi e 2 su 5 sono donne

Sesso	frequenze	frequenze relative	percentuali
M	3	$3/5=0.6$	$3/5*100=60$
F	2	$2/5=0.4$	$2/5*100=40$
totale	5	1	100

Per esprimere meglio il concetto:

Il 60% delle persone intervistate è di sesso maschile il restante 40% è di sesso femminile

Frequenze relative e percentuali

La frequenza relativa viene calcolata come rapporto tra la frequenza assoluta e la frequenza totale n

Ad esempio: la prima frequenza assoluta è $0.6 = 3/5$

Sesso	frequenze	frequenze relative
M	3	$3/5=0.6$
F	2	$2/5=0.4$
totale	5	1

Mentre, la frequenza percentuale è uguale alla frequenza relativa moltiplicata per 100

Sesso	frequenze	frequenze relative	percentuali
M	3	$3/5=0.6$	$3/5*100=60$
F	2	$2/5=0.4$	$2/5*100=40$
totale	5	1	100

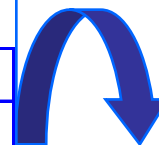
Frequenze relative e percentuali

E, per il titolo di studio, si ha:

Titolo di studio	frequenze
Laurea	2
Diploma	3



Titolo di studio	frequenze	frequenze relative
Laurea	2	$2/5=0.4$
Diploma	3	$3/5=0.6$
		1



2 su 5 sono laureati e
3 su 5 sono diplomati

Titolo di studio	frequenze	frequenze relative	percentuali
Laurea	2	$2/5=0.4$	$2/5*100=40$
Diploma	3	$3/5=0.6$	$3/5*100=60$
totale	5	1	100

Per esprimere meglio il concetto:

Il 40% delle persone intervistate è laureato, il restante 60% è diplomato

Frequenze relative e percentuali

E, per il peso, si ha:

Peso (kg)	frequenze	frequenze relative	percentuali %
72	2	$2/5=0.4$	40
55	1	$1/5=0.2$	20
63	1	0.2	20
64	1	0.2	20
totale	5	1	100

Il 40% delle persone intervistate pesa 72 Kg

Mentre il restante 60% meno di 72 Kg e precisamente un 20% pesa 55 Kg, un altro 20% pesa 63 kg ...

Frequenze cumulate

Le frequenze cumulate consentono di determinare il numero di unità statistiche (costituenti il collettivo) che presentano una modalità inferiore o uguale ad un determinato valore

Consistono in una somma progressiva delle frequenze assolute corrispondenti alle modalità ordinate in modo crescente:

Peso (kg)	frequenze	frequenze cumulate
55	1	1
63	1	2
64	1	3
72	2	5
totale	5	

Il valore "3" rappresenta il numero di coloro che presentano un peso inferiore o uguale a 64 Kg

Frequenze cumulate relative

Esempio:

salari settimanali	dipendenti	frequenze cumulate
250	15	15
320	25	40
340	10	50
360	16	66
400	8	74
Totale	74	

66 è il numero di dipendenti che presentano un salario settimanale inferiore o uguale a 360 euro

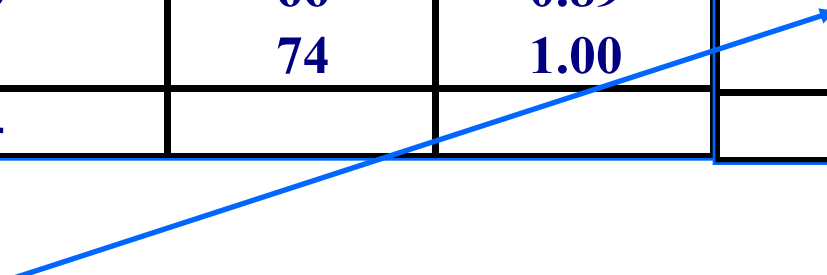
Per trasformare le frequenze cumulate in frequenze relative è necessario dividere ogni frequenza cumulata per il totale $N = 74$. Così, ad esempio, la terza frequenza cumulate relativa è pari a $50/74 = 0.68$ e cioè 50 su 74 percepiscono un salario inferiore o uguale a 340 euro

Frequenze cumulate percentuali

Anche le frequenze cumulate possono essere trasformate in relative e poi in percentuali:

Esempio:

salari settimanali	dipendenti	frequenze cumulate	freq. Cumulate relative	freq. Cumulate percentuali
250	15	15	0.20	$0.20 * 100 = 20\%$
320	25	40	0.54	54.05%
340	10	50	0.68	67.57%
360	16	66	0.89	89.19%
400	8	74	1.00	100.00%
Totale	74			



89 è la percentuale di dipendenti che presentano un salario settimanale inferiore o uguale a 360 euro

...ricapitolando...

Prezzo al Kg di pasta rilevato in diverse città:
DISTRIBUZIONE DI FREQUENZA

I dati possono essere:

- Qualitativi (nominali o ordinali)
- Quantitativi (discreti o continui)

Possono inoltre essere espressi in forma di successione di dati o in distribuzioni di frequenza!

Prezzo al Kg di pasta rilevato in diverse città:
SUCCESSIONE DI VALORI



2.3 2.8 2.9 3.3 3.5
3.6 3.7 3.8 4.1 4.2
4.4 4.6 4.6 4.7 4.7
4.8 5.1 5.3 5.6 5.8

PREZZO	FREQ
2.3	1
2.8	1
2.9	1
3.3	1
3.5	1
3.6	1
3.7	1
3.8	1
4.1	1
4.2	1
4.4	1
4.6	2
4.7	2
4.8	1
5.1	1
5.3	1
5.6	1
5.8	1

DISTRIBUZIONE DI FREQUENZA CON MODALITA' SUDDIVISE IN CLASSI

Un elenco di dati quantitativi, come quello appena rappresentato, può risultare **poco chiaro** ad un'analista

Per un'analisi più efficace si consiglia di formare una tabella dove le modalità sono suddivise in classi.

Ad ogni classe è associata la frequenza assoluta, ovvero il numero delle volte in cui si osservano dati contenuti in quella classe.

Ma bisogna scegliere il numero delle classi per individuarne l'ampiezza

...Vediamo in che modo...

Si calcola il range, dove

$$\text{Range} = \text{Val. max} - \text{Val. min}$$

E, quindi, si individua l'ampiezza dell'intervallo (approssimazione per eccesso) come

$$\text{Ampiezza} = \frac{\text{range}}{\text{numero classi}}$$

Si calcola il range, dove

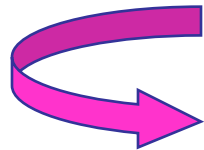
$$\text{Range} = \text{Val. max} - \text{Val. min}$$

Nel nostro caso:

$$\text{Range} = 5.8 - 2.3 = 3.5$$

E, quindi, si individua l'ampiezza dell'intervallo (approssimazione per eccesso) come

$$\text{Ampiezza} = \frac{\text{range}}{\text{numero classi}}$$



$$\text{Ampiezza} = \frac{3.5}{4} = 0.875 \approx 1$$

<i>Prezzo al Kg di pasta</i>	<i>Frequenze (assolute)</i>
(2 -3]	3
(3 -4]	5
(4 -5]	8
(5 -6]	4
<i>Totale</i>	<i>20</i>

Simbologia

$N \rightarrow$ totale delle frequenze

$X_i \rightarrow$ modalità i -esima del carattere X

$n_i \rightarrow$ frequenze assolute della classe i

$f_i \rightarrow$ frequenze relative della classe i

$p_i \rightarrow$ frequenze percentuali della classe i

$a_i \rightarrow$ ampiezza della classe della classe i

$N_i \rightarrow$ frequenze cumulate assolute della classe i

$F_i \rightarrow$ frequenze cumulate relative della classe i

$P_i \rightarrow$ frequenze cumulate percentuali della classe i

Una classe indicata con

$$(x_i - x_{i+1}]$$

implica che x_i non appartiene alla classe mentre x_{i+1} appartiene alla classe.

Nel seguito, se non diversamente specificato, la classe $x_i - x_{i+1}$ significa $(x_i - x_{i+1}]$.