

Lezione 1

La statistica, i dati, le fonti

PROF. ROBERTO COSTA

SCIENZE DELL'EDUCAZIONE - STATISTICA SOCIALE (305SF)

Informazioni utili

Contatti docente:

Email roberto.costa@deams.units.it

www.units.it/persone/index.php/from/abook/persona/035893

Inizio Lezioni: 22/09/2025

Giorno: lunedì. Orario: 10.00-13.00. Dieci lezioni da 3 ore ciascuna. Totale 30 ore, corso di 6 CFU

Ricevimento: su appuntamento via Teams

Informale: prima e dopo lezioni brevemente

Le slides saranno disponibili entro il venerdì successivo alla lezione.

Cosa devo sapere prima di iniziare?

Nessun corso è necessario prima di poter sostenere questo. Potrebbe anche essere il vostro primo esame!

Le pochissime conoscenze necessarie sono di ambito matematico:

- Le 4 operazioni elementari,
- I simboli matematici di base,
- Come elevare al quadrato e come fare radice quadrata (con il PC o la calcolatrice).

Non abbiate paura dei numeri ;-)

Cosa farò durante il corso?

Ci saranno lezioni teoriche tradizionali, dove comunque cercheremo di interagire.

Ci aiuteremo con degli esercizi in aula con carta & penna.

Proveremo delle applicazioni pratiche: principalmente con il supporto di un foglio elettronico (excel, openoffice,...).

Ogni tanto vi proporrò degli esercizi da fare per la settimana successiva, mi aspetto che ci proviate e che mi chiediate la volta dopo eventuali chiarimenti.

Consiglio il testo: Statistica per le scienze sociali, II^a ed. E. Amaturò e al. (2021, UTET)



Come verrò valutato?

Obbligatorio: verifica finale scritta. Domande di teoria ed esercizi. Si può utilizzare calcolatrice.
Valutazione massima: 30

Facoltativo: stesura di un mini rapporto a gruppi. Lo consiglio vivamente, ma non è obbligatorio.
Si può incrementare la valutazione dello scritto fino ad un massimo di 3 punti.

Entrambe le verifiche saranno poco indirizzate al calcolo e maggiormente alla comprensione.



Gli argomenti del corso

- Introduzione, cos'è la statistica. Come rilevare e trattare in modo statistico i fenomeni sociali
- La classificazione delle variabili statistiche
- Come posso rappresentare le variabili statistiche: le tabelle e i grafici
- La sintesi delle distribuzioni di frequenze: valori centrali e misure di dispersione
- Statistica bivariata: le relazioni tra due caratteri
- Cenni di inferenza statistica
- Il campionamento

Cos'è la statistica?

Troverete molte definizioni di statistica, prendiamo la seguente come riferimento:

«La statistica è la scienza che si occupa dello studio **quantitativo** di **fenomeni collettivi**, normalmente complessi, mediante l'impiego di metodi di riduzione dell'informazione e di analisi di convalida dei risultati in termini di rappresentatività».

Un *modus operandi* per approcciare in modo scientifico la realtà.

Perché studiare statistica

«L'educatore socio pedagogico è un professionista che opera nell'ambito educativo, formativo e pedagogico, in rapporto a qualsiasi attività svolta in modo formale, non formale e informale, in prospettiva di life long learning, in un'ottica di crescita personale e sociale».

Un educatore socio pedagogico informato deve aggiornarsi, conoscere le nuove tecniche di insegnamento e capire le possibili difficoltà di particolari categorie di persone secondo i più recenti studi.

Molto spesso tali studi vengono effettuati tramite raccolta dati e/o esperimenti in entrambi i casi è coinvolta la statistica.

Anche se non sarete voi a condurre sperimentazioni sulla didattica, dovrete comunque essere in grado di interpretare i risultati ottenuti da altri!

Perché studiare statistica

Troverete articoli scientifici come questo, in cui si sostengono tesi, partendo dai risultati di un'indagine quantitativa.

La pedagogia nell'era digitale: il ruolo dei social media
nella promozione dell'attività motoria
*Pedagogy in the digital age: the role of social media
in promoting physical activity*



Patrizia Belfiore

Università degli Studi di Napoli "Parthenope" – patrizia.belfiore@uniparthenope.it

ABSTRACT

The literature of different areas of study reports that social media, especially new social media (Facebook), plays an important role for society in the third millennium: consequently, this issue requires the continuous attention of the pedagogical community as well. Facebook has had great potential to increase knowledge and perception of the concept of well-being and motor education, but its role is still being defined. The aim of the paper is to demonstrate, through an online survey, how and how much Facebook promotes well-being by encouraging the practice of motor activity. Statistical data analysis, used in an integrated and complementary approach, was multiple match analysis, cluster analysis, and cross-tabs. The final sample consisted of 1,820 Facebook users. Social media websites are widely used not

La statistica è intorno a noi

La statistica la usiamo, anche inconsciamente, molto di frequente...

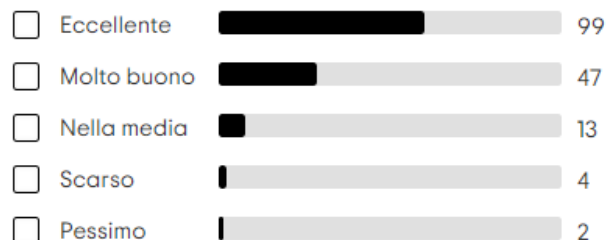
Punteggi e recensioni

4,5  231 recensioni

Recensioni (231)

Controlliamo le recensioni ⓘ

Valutazione



Tipo di viaggiatore

- ☐ Famiglie
- ☐ In coppia
- ☐ Da solo
- ☐ Affari
- ☐ Amici

Periodo dell'anno

- ☐ Mar-Mag
- ☐ Giu-Ago
- ☐ Set-Nov
- ☐ Dic-Feb

Scopri i commenti dei viaggiatori:

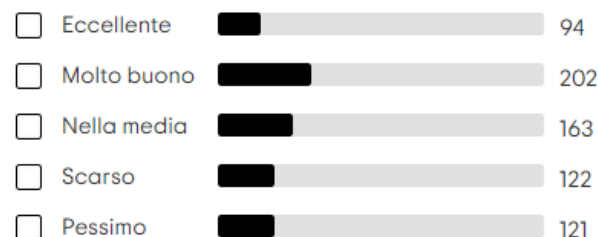
Punteggi e recensioni

3,0  813 recensioni

Recensioni (813)

Controlliamo le recensioni ⓘ

Valutazione



Tipo di viaggiatore

- ☐ Famiglie
- ☐ In coppia
- ☐ Da solo
- ☐ Affari
- ☐ Amici

Periodo dell'anno

- ☐ Mar-Mag
- ☐ Giu-Ago
- ☐ Set-Nov
- ☐ Dic-Feb

Scopri i commenti dei viaggiatori:

Poniamoci alcune domande

Qual è la fonte delle informazioni?

Come vengono raccolti i dati?

Sulla base di cosa faccio la mia scelta?

Come posso leggere i dati?

Come analizzare un fenomeno sociale

L'attività scientifica in generale si basa sull'osservazione empirica della realtà.

Distinguiamo tra:

- Ricerca quantitativa (uso di tecniche più strutturate e basata su un numero più ampio di osservazioni)
- Ricerca qualitativa (più orientata alla comprensione di un fenomeno che alla quantificazione)

La prima è molto più strutturata rispetto alla ricerca qualitativa.

Nelle prossime slides parleremo di ricerca quantitativa.

Le fasi di una ricerca

- ✓ Progettazione o disegno della ricerca
- ✓ Raccolta/rilevazione dei dati
- ✓ Analisi dei dati
- ✓ Diffusione dei risultati

What is GIGO?



Progettazione o disegno della ricerca

- ✓ Definire operativamente i concetti della ricerca;
- ✓ Definire gli oggetti della ricerca (le unità di rilevazione);
- ✓ Stabilire se condurre una rilevazione su tutte le unità o su una parte (rilevazione censuaria/campionaria);
- ✓ Stabilire le modalità di passaggio dal piano teorico a quello empirico,
- ✓ Costruire lo strumento di rilevazione (ad es. un questionario);
- ✓ Prefigurare la fase di analisi dei dati.

Raccolta dei dati

- ✓ E' la cosiddetta fase sul campo di una ricerca.
- ✓ Si raccolgono i dati presso le unità di rilevazione.
- ✓ Il risultato è una matrice dei dati, ovvero una tabella dove in riga troviamo le risposte dei singoli e in colonna troviamo le risposte alle singole variabili.
- ✓ Nell'incrocio tra un riga e una colonna troviamo un dato.

La matrice dei dati

- ✓ In riga troviamo tutte le risposte di una unità di rilevazione.
- ✓ In colonna le risposte di tutte le unità di rilevazione a una variabile.
- ✓ Nell'incrocio tra un riga e una colonna troviamo un dato, ovvero la risposta di una unità di rilevazione a una variabile.

unità	età	titolo di studio	categoria	retribuzione mensile netta
1	24	secondario	impiegato	1500
2	35	terziario	quadro	1950
3	49	secondario	operaio	1450
4	57	secondario	quadro	2300
5	52	secondario	impiegato	1650
6	21	secondario	operaio	1200
7	39	terziario	dirigente	2700
8	44	terziario	quadro	2200
9	31	secondario	impiegato	1550
10	60	primario	operaio	1500
11	28	secondario	impiegato	1550
12	22	secondario	operaio	1200
13	36	secondario	operaio	1300
14	41	secondario	impiegato	1600
15	52	terziario	dirigente	3100
16	55	primario	operaio	1400
17	61	primario	operaio	1550
18	20	secondario	operaio	1200
19	18	secondario	operaio	1200
20	42	secondario	impiegato	1600
21	36	terziario	quadro	2100
22	37	terziario	impiegato	1600
23	48	primario	operaio	1550
24	55	secondario	operaio	1700
25	29	secondario	operaio	1550

La matrice dei dati

Usando una terminologia statistica, se raccolgo m variabili su n unità otterrò una matrice dei dati con $n \times m$ celle.

La j esima variabile dell' i esima unità di rilevazione si indica con $x_{i,j}$

		variabili						
unità statistiche		v_1	v_2	v_3	...	v_j	...	v_m
	u_1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$				
	u_2	$x_{2,1}$						
	u_3	$x_{3,1}$						
	...							
	u_i					$x_{i,j}$		
	...							
	u_n							$x_{n,m}$

Il dato

Il **dato** si definisce come: «l'esito di un processo di rilevazione delle informazioni alla luce dell'apparato metodico e concettuale di una certa disciplina». (Statistica per le scienze sociali, E. Amaturio et al.).

Abbiamo visto che è il contenuto di una singola cella di una matrice dei dati.

Per poter parlare di dato dev'esserci una **definizione operativa**, ovvero quell'insieme di regole per poter trasformare i possibili stati di una variabile in un numero o descrizione da inserire nella matrice.

In alcuni casi può essere semplice, come ad es. statura: «la misura del corpo umano eretto in centimetri», in altri può essere più complessa, pensiamo ad es, alla classe sociale.

Definizione operativa

Certi fenomeni si possono misurare direttamente, come ad esempio l'età, altri necessitano di una fase detta **operazionalizzazione**, ovvero di essere trasformati in una o più domande, perché altrimenti il concetto o il fattore in sé non è misurabile.

Pensiamo ad esempio ad alcuni concetti come povertà, benessere, salute, ...

Proviamo a ragionare assieme al concetto di povertà.

Cosa significa «essere poveri»

La povertà può essere definita come una condizione umana caratterizzata da privazione continua e cronica delle risorse, capacità, scelte, sicurezza e potere necessari per poter godere di uno standard di vita adeguato ed altri diritti civili, culturali, economici, politici e sociali.

Una famiglia si può definire povera se la sua spesa per consumi è inferiore o uguale al valore monetario di un paniere di beni e servizi considerati essenziali per evitare gravi forme di esclusione sociale.

Cosa significa «essere poveri»

L'Istat ha due misure di povertà: la povertà assoluta e quella relativa.

A differenza delle misure di povertà relativa, che individuano la condizione di povertà nello svantaggio di alcuni soggetti rispetto agli altri, la povertà assoluta si riferisce all'incapacità di acquisire i beni e i servizi necessari a raggiungere uno standard di vita ritenuto “minimo accettabile” nel contesto di appartenenza (paniere di povertà assoluta).

Tre livelli dei dati

Microdati: costituiscono l'unità elementare di informazione, ovvero un dato sulla singola unità statistica (ad es. il singolo voto ad un esame).

Macrodati: sono unità complesse di informazione, costituite dall'elaborazione e dalla sintesi dei microdati (ad es. la media dei voti di una sessione di esame).

Metadati: sono i «dati sui dati». Forniscono informazioni dettagliate sui dati e sul loro processo di produzione per consentire una corretta lettura e interpretazione. (ad. es. periodo e modalità di raccolta, definizioni operative, ecc.).

In pratica...

Microdati: io peso 70 kg e sono alto 1,76 m.

Macrodati: il 44,4% delle persone di 18 anni e più in Italia è sovrappeso o obesa.

Metadati: Proporzione (standardizzata con la popolazione europea al 2013) di persone di 18 anni e più in sovrappeso o obese sul totale delle persone di 18 anni e più. L'indicatore fa riferimento alla classificazione dell'Organizzazione mondiale della sanità (Oms) dell'Indice di Massa corporea (Imc: rapporto tra il peso, in Kg, e il quadrato dell'altezza, in metri). Fonte: Istat, Indagine Aspetti della vita quotidiana. Anno 2021.

In pratica...

QUESTIONARIO PER LA VALUTAZIONE DEL SENSO DELL'EQUILIBRIO



Ha sofferto/soffre di vertigini o perdita dell'equilibrio?

☐ SÌ ☐ NO

Se sì, ne è stata individuata la causa?

☐ SÌ ☐ NO

Se non è stata individuata la causa: i disturbi erano/sono associati a perdita di coscienza, stato confusionale, senso di obnubilamento?

☐ SÌ ☐ NO

Ha disturbi dell'udito (acufeni, ipoacusia)?

☐ SÌ ☐ NO

La vertigine era/è

☐ soggettiva ☐ oggettiva

I movimenti del capo la peggioravano/peggiorano

☐ SÌ ☐ NO

Durata della vertigine

☐ minuti ☐ ore/giorni

VALUTAZIONE CLINICA

- Manovra di Romberg

☐ positiva ☐ negativa

- Prova indice-naso

☐ positiva ☐ negativa

LA PARTECIPAZIONE AL MERCATO DEL LAVORO PER CLASSI DI ETÀ

PROSPETTO 4. POPOLAZIONE PER CLASSI DI ETÀ E CONDIZIONE PROFESSIONALE
Ottobre 2020, dati destagionalizzati

	Valori assoluti (migliaia di unità)	Variazioni congiunturali				Variazioni tendenziali	
		ott20 set20	ott20 set20	ago-ott20 mag-lug20	ago-ott20 mag-lug20	ott20 ott19	ott20 ott19
		(assolute)	(percentuali)	(assolute)	(percentuali)	(assolute)	(percentuali)
15-24 ANNI							
Occupati	971	-24	-2,4	+18	+1,9	-115	-10,6
Disoccupati	422	+2	+0,6	+19	+4,7	+6	+1,5
Inattivi	4.476	+18	+0,4	-47	-1,0	+102	+2,3
25-34 ANNI							
Occupati	3.946	+64	+1,6	+49	+1,3	-114	-2,8
Disoccupati	705	+4	+0,5	+39	+5,9	+47	+7,2
Inattivi	1.768	-77	-4,2	-121	-6,3	-27	-1,5
35-49 ANNI							
Occupati	9.134	-11	-0,1	+29	+0,3	-290	-3,1
Disoccupati	822	+13	+1,6	+12	+1,5	-18	-2,1
Inattivi	2.517	-32	-1,3	-114	-4,3	+15	+0,6
50 ANNI E PIU'							
Occupati	8.791	-43	-0,5	+19	+0,2	+45	+0,5
Disoccupati	530	-8	-1,5	+50	+10,4	+7	+1,3
Inattivi	17.852	+78	+0,4	-17	-0,1	+302	+1,7
Inattivi 50-64 anni	4.811	+65	+1,4	-6	-0,1	+166	+3,6

Nelle indagini socio-demografiche convivono attualmente due distinti approcci alla rilevazione della condizione professionale: l'approccio diretto, che rileva la condizione professionale degli individui conformemente alla loro dichiarazione (condizione dichiarata) e l'approccio indiretto (ma sarebbe forse meglio definirlo semi-indiretto), nel quale la condizione professionale viene determinata valutando, oltre a quanto dichiarato dagli individui, anche alcune caratteristiche delle relazioni che questi hanno con il mercato del lavoro (condizione ricostruita).

Le fonti dei dati

I dati possiamo raccoglierli direttamente (**dati primari**), in altri casi possiamo utilizzare quelli raccolti da altri (**dati secondari**).

La principale fonte di dati secondari in Italia è l'Istat, l'istituto nazionale di statistica, cui fa capo il Sistan, il sistema statistico nazionale.

Dal sito di Istat è possibile scaricare una grande quantità di dati, sotto forma di microdati e macrodati. Sono disponibili anche i metadati, per un uso più corretto e consapevole dei dati.

Esistono svariate fonti di dati, come ad es. gli istituti privati di ricerca, le università, ecc., ma a volte facciamo ricorso a dati prodotti non per finalità statistiche, ma per altri motivi (ad es. **dati amministrativi**).

Pensiamo ai **big data**.

I big data – le 3 V

Si definisce col termine **big data** una raccolta di dati così estesa in termini di **v**olume, **v**elocità e **v**arietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore.

I recenti progressi tecnologici hanno ridotto esponenzialmente il costo di archiviazione e di elaborazione dei dati, rendendo più facile e meno costoso archiviare più dati rispetto al passato.

Con un volume crescente di big data, diventato sempre più economico e accessibile, si possono conoscere i fenomeni sociali, economici, ecc. in modo più approfondito e prendere decisioni e fare valutazioni più accurate e precise.

L'analisi dei dati

Per analisi dei dati si intende «la scelta delle tecniche statistiche più appropriate in base alle variabili a disposizione del ricercatore e ai suoi obiettivi conoscitivi».

Distinguiamo due grandi obiettivi conoscitivi:

- descrivere, esplorare, sintetizzare i dati raccolti;
- inferire i risultati, ottenuti su un numero ristretto di unità, all'intera popolazione.

L'analisi dei dati

Parleremo di **statistica descrittiva** laddove l'obiettivo è di ridurre i dati e rappresentarli in forma sintetizzata attraverso valori, coefficienti, tabelle, grafici.

Laddove invece si voglia risalire alle caratteristiche della popolazione da cui sono state estratte le unità di rilevazione utilizzeremo la **statistica inferenziale**, che si basa sulla teoria delle probabilità e consente di passare dall'osservazione di un campione a tutta la popolazione.

L'analisi dei dati

A seconda del numero di variabili prese in considerazione si parla di:

Analisi univariata se analizziamo le variabili di un fenomeno una alla volta.

Analisi bivariata se l'obiettivo è quello di studiare il legame tra due variabili.

Analisi multivariata se vogliamo studiare il legame tra tre o più variabili.

L'analisi dei dati

Attenzione! Il ruolo del ricercatore è fondamentale:

[Spurious correlations: 15 examples - DataScienceCentral.com](https://www.data-science-central.com/spurious-correlations-15-examples/)

«I numeri, a torturarli a lungo, confessano qualsiasi cosa»

Ronald Harry Coase (Willesden, 29 dicembre 1910 – Chicago, 3 settembre 2013), economista inglese naturalizzato statunitense, vincitore del premio Nobel per l'economia nel 1991.

La diffusione dei risultati

Restituzione dei principali risultati della ricerca e le conclusioni a cui si è giunti.

In considerazione del pubblico a cui ci si rivolgerà cambieranno le modalità di restituzione: verranno prodotti grafici, tabelle e diagrammi con le relative interpretazioni.



Per lunedì prossimo...

Italia, tasso disoccupazione scende al 7,8%. Ma salgono inattivi

**Lavoro, in Italia si è invertita la rotta:
occupazione in forte calo, ma non tra gli
under 25. Cosa succederà quest'inverno?**

In questi titoli sono citati 3 termini: disoccupazione, occupazione, inattività. Provate a ricostruire, attraverso i metadati messi a disposizione dall'Istat, il significato di questi termini e come vengono calcolati questi indicatori?

Esercitiamoci

Cos'è il dato statistico?

- A. Il contenuto di una cella della matrice dei dati, ovvero lo stato di un caso su una determinata proprietà rilevata
- B. Un'unità elementare su cui vengono rilevate le variabili oggetto delle analisi statistiche
- C. Un sottoinsieme della popolazione, composto da unità da estrarre dalla stessa e rappresentative dell'intera popolazione.

Esercitiamoci

«Insieme di tecniche statistiche che hanno come obiettivo quello di ridurre i dati e rappresentarli in forma sintetica attraverso distribuzioni, valori caratteristici e coefficienti e grafici». Si tratta della definizione di?

- A. Statistica descrittiva
- B. Statistica inferenziale

Esercitiamoci

Il questionario è:

- A. Una tecnica di raccolta dei dati basata sull'interrogazione di soggetti.
- B. Una tecnica di elaborazione dei dati.