Elementi di Psicometria (con laboratorio software 1)

01-Rappresentazione dei dati (1)

(v. 1.1, 7 marzo 2020)

Germano Rossi[1]
germano.rossi@unimib.it

[1]Dipartimento di Psicologia, Università di Milano-Bicocca

2019-20

Perché studiare statistica?

1. Per comprendere la letteratura specializzata

Teorici

- Ci sono psicologi (i "teorici") che studiano le "regole" con cui "funziona" la *psiche* dell'essere umano
- I *teorici* usano la statistica per avere una buona certezza della bontà delle regole
- La diffusione delle "regole" da parte dei *teorici* avviene tramite comunicazioni a congressi, report, articoli su riviste specializzate e libri
- La maggior parte degli articoli scientifici, includoo analisi dei dati e statistiche

Tecnici o applicativi

- Altri psicologi (i "tecnici") "usano" queste regole negli ambiti più svariati
- I *tecnici* devono essere in grado di capire le regole, come e perché funzionano, quanto ci si può fidare
- I *tecnici* devono conoscere la statistica

Perché studiare statistica?

2. Per comprendere la logica sottostante la ricerca in psicologia
   - La letteratura scientifica presenta i propri risultati seguendo degli schemi (teoria, ipotesi, ricerca, risultati, conclusioni).
   - Sapere come "funziona" la ricerca, ci permette di capire la "letteratura psicologica"
3. Per condurre ricerche in psicologia
   - È possibile che anche voi "possiate fare ricerca" un domani (ad es. con la tesi o proseguendo con una Magistrale e/o il Dottorato)

Cos'è la statistica

- È una branca della matematica
- Il termine statistica deriva da *statista* (uomo di stato) e originariamente indicava le misurazioni (informazioni varie riguardanti il territorio, ad es. numero di persone, grano disponibile...) relative allo Stato di un determinato marchesato, ducato, principato, regno.
- Oggi indica
  - sia la **scienza della statistica**
  - sia i valori che risultano dall'applicazione di alcuni algoritmi di calcolo (che producono gli **indici statistici**)
  - sia i **metodi** e le **tecniche** per calcolarli (ovvero l'analisi dei dati)

Cos'è la statistica

I **metodi statistici** si possono suddividere in

Statistica descrittiva

- Riassume e descrive tramite numeri le caratteristiche principali di un insieme di "misurazioni" (che rappresentano "informazioni") raccolte sulla realtà

Statistica inferenziale

- Permette di stimare la "sicurezza" di alcune statistiche calcolate su un piccolo insieme (campione) rispetto alla popolazione oppure di confrontare tra loro dei gruppi (insiemi) differenti

Definizioni: Costrutto, variabile, costante

Un costrutto è un concetto mentale che viene associato a qualche elemento della realtà per un qualche motivo; in un certo senso è la stessa cosa di un significato, ma un costrutto è un "significato complesso" (ad es. l'amicizia, l'aggressività)

Una variabile è una visione particolare, una parte, un'aspetto di un costrutto (la variabile *persone frequentate* è un'aspetto del costrutto *amicizia*); la *variabile* ha questo nome perché rappresenta qualcosa che può assumere valori diversi

- Una costante è qualcosa che non cambia
- La possibilità di misurare un costrutto (tramite le variabili associate a certi aspetti del costrutto) è oggetto della *teoria della misurazione*

Definizioni: Variabile, valore, punteggio

Variabile è ciò che intendiamo (e possiamo) misurare (ad es. età, genere, corporatura, livello di benessere)

Valore è l'insieme di tutte le caratteristiche di ciò che stiamo misurando. (età: anni, genere: maschio/femmina, corporatura: mingherlino/normale/sovrappeso, benessere: un numero che indica il livello di benessere)

Punteggio è il valore di un caso statistico per una specifica variabile, è il valore che viene attribuito (età: 31, genere: maschio, corporatura: normale, benessere: 43)

Definizioni: Unità statistica, misurazione

Unità statistica è l'unità minima di quello che stiamo studiando; in psicologia è (generalmente) un individuo.
Tuttavia, in ambito statistico, può essere molte cose diverse: un barattolo di tonno, una coppia marito e moglie, un'automobile, una famiglia, un'azienda.
Può essere chiamata anche **caso statistico**

Misurazione è l'operazione di assegnare un *valore* ad una *variabile* in riferimento ad un *caso statistico*

Teoria della misurazione

La Teoria della misurazione è il ponte fra la realtà e le informazioni da sottoporre ad analisi statistica.

> «Misurare è assegnare un modello formale ad oggetti ed eventi secondo una regola, una qualunque regola, purché sia consistente» (Stevens)

Afferma che le misurazioni effettuate devono mantenere le caratteristiche di ciò che viene misurato.

Ad es. il sesso biologico è una qualità dell'individuo che non è esprimibile con dei numeri, che non permette somme o sottrazioni...

Nel momento in cui "misuriamo" dobbiamo essere consapevoli del livello di misura della variabile.

Livelli di misura

Ipotizziamo di essere al supermercato, in coda alla cassa. Davanti a noi abbiamo 3 persone e notiamo che tutte e tre hanno acquistato anche un prodotto in offerta (un succo di frutta):

a) una signora grassottella con un carrello zeppo di prodotti e 12 confezioni in offerta
b) un signore di corporatura media con 2 confezioni
c) un'altra signora mingherlina con 4 confezioni

Quando arrivano alla cassa, prendiamo nota di quanto spendono.

Livelli di misura

Abbiamo raccolto delle informazioni (misurazioni) su 3 persone (unità statistiche o casi statistici)

- Abbiamo delle variabili (genere, corporatura, confezioni, conto)
- Abbiamo dei valori possibili (maschio/femmina, mingherlina..., dei numeri)
- Abbiamo dei punteggi (F, grassottella, 12...)
- Possiamo costruire una tabella a doppia entrata
- Le colonne contengono "variabili"; le righe "unità statistiche"; le celle contengono i "valori" di una certa "variabile"

Livelli di misura

Id	Sesso	corporatura	confezioni in offerta	conto
a	F	grassottella	12	180
b	M	media	2	50
c	F	mingherlina	4	40

- Il **sesso** si limita a distinguere categorie diverse (M/F)
- La **corporatura** oltre a distinguere le categorie, permette anche di metterle in ordine (da magro a grassottella), ma non è possibile *quantificare* la differenza fra "mingherlina" e "normale" o fra "media" e "grassottella"
- Le variabili **confezioni** e **conto** usano anche un'*unità di misura* fissa e permettono di quantificare le differenze fra i diversi valori (fra 2 e 4 confezioni c'è una differenza di 2)

Livelli di misura

- I diversi modi di misurare si chiamano livelli di misura
- Possono essere classificati anche con altre (diverse) modalità

categorie	vs.	punteggi
qualitative	vs.	quantitative
discrete	vs.	ontinue
nominali, ordinali	vs.	intervallo/rapporto

Attenzione

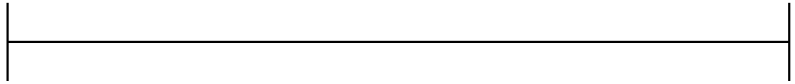
- Excel e SPSS possono lavorare con numeri e con stringhe (testo)
- Excel e SPSS non sanno nulla delle scale di misura
- Se sono numeri, li trattano come numeri; se sono testi, come testi

Qualitative vs. quantitative

- Un livello di misura è qualitativo quando tiene conto solo e soltanto di una qualche **qualità** di ciò che si sta misurando (ad. es. il colore dei capelli, la razza di appartenenza, il titolo di studio raggiunto)
- Le scale qualitative vengono anche chiamate categoriali perché i loro valori sono categorie
- Un livello di misura è quantitativo quando è possibile usare i numeri per indicare i diversi livelli di ciò che si sta misurando (ad. l'età, l'altezza di una persona, il numero di risposte giuste ad un questionario) e il numero esprime "effettivamente" un'**unità di misura**
- Le scale quantitative utilizzano punteggi

Discrete vs. continue

- Un livello di misura è continuo quando un valore utilizzato "scivola" nell'altro, cioè quando, fra una qualunque misurazione e l'altra, vi sono infinite possibili misurazioni (ad es. l'età). Il valore che viene misurato è sempre un'*approssimazione del valore reale* (l'approssimazione dipende dalla sensibilità dello strumento di misura)



- Un livello di misura è discreto quando un valore utilizzato è completamente separato da qualunque altro e non vi sono valori intermedi; le variabili qualitative sono sempre discrete, mentre quelle quantitative sono discrete se non è possibile effettuare una misurazione intermedia fra due valori contigui (ad es. numero di persone che entrano in un negozio)

Nominali vs. ordinali vs. intervallo/rapporto

- È il sistema di classificazione più usato ed è stato proposto da Stevens
- Un livello di misura è nominale quando è soltanto possibile dare un nome (o etichetta) alle categorie di ciò che si sta misurando (ad. es. il colore dei capelli)
- Un livello di misura è ordinale quando è possibile ordinare i diversi livelli di categorie (ad. per il titolo di studio, la maturità è superiore alla licenza media, la laurea è superiore alla maturità, ...) in un modo qualunque
- Un livello di misura è ad intervallo/rapporto quando i valori assunti dalla variabile possono essere espressi tramite numeri che fanno riferimento ad una specifica **unità di misura** (anni di studio, altezza in cm, ...)

Scala nominale

- Sistema empirico classificatorio
- 2 elementi possono appartenere a categorie uguali (=) o diverse ($\neq$) (relazione di uguaglianza o di equivalenza)
- ogni elemento appartiene ad una ed ad una sola categoria
- le categorie possono usare qualsiasi tipo di etichetta (simboli, testo o numeri); anche se si usano i numeri, questi non sono numeri, ma testo (quindi 1 significa "uno")

Esempio

Il **genere** (maschio, femmina), la **professione** (chirurgo, operaio, ...), il **partito politico** (PDL, PD, IDV, M5S, ...)

Scala ordinale

- Oltre alla caratteristiche precedenti (uguaglianza, disuguaglianza)
- 2 elementi diversi possono essere ordinati fra di loro, così che si può dire che uno dei due è minore dell'altro oppure è maggiore (relazione d'ordine)
- le categorie possono usare etichette che permettano di esprimere la relazione d'ordine (testo [a, b, c...] e numeri)

Esempio

Lo **status socio-economico** (povero, borghese, ricco), il **livello scolastico** (elementari, medie, superiori, laurea, dottorato), le **classi di età** (neonato, bambino, pre-adolescente, adolescente, giovane, giovane-adulto, adulto, adulto-anziano, anziano)

Scala a intervalli

- Oltre alle caratteristiche precedenti (uguaglianza, disuguaglianza, ordine)
- Fra 2 categorie esiste teoricamente un intervallo che è esprimibile come multiplo di un intervallo unitario (o intervalli equivalenti)
- All'intervallo unitario è possibile riferire qualunque intervallo fra 2 categorie
- Esiste un punto-origine relativo
- Le categorie possono usare solo etichette numeriche (valori numerici positivi e negativi)

Esempio

La **temperatura** (0 gradi, 15 gradi, 25 gradi), la maggior parte dei **test psicologici** (QI, STAI, DPI, ...)

Scala a rapporti

- Oltre alla caratteristiche precedenti (uguaglianza, disuguaglianza, ordine, intervallo unitario)
- Il punto origine è assoluto (0 significa totale assenza della caratteristica misurata)
- Le categorie possono usare solo etichette numeriche (solo valori numerici positivi)

Esempio

L'**età**, l'**altezza**, il **peso**, la maggior parte delle variabili fisiche

Riepilogo

Relazione	**Scala di misura**			
	Qualitative		*Quantitative*	
	Nominale	*Ordinale*	*Intervallo*	*Rapporto*
Equivalenza	$=, \neq$	$=, \neq$	$=, \neq$	$=, \neq$
Ordine		$<, >$	$<, >$	$<, >$
Intervallo			$+, -$	$+, -$
Rapporto				$\times, \div$

Implicazioni 1

- Ogni unità statistica può avere uno e un solo valore per ogni caratteristica misurata

Esempio

Se ho rilevato il reddito (scala a rapporti), non posso avere una persona che (contemporaneamente) guadagni sia 500 euro sia 60.000 euro l'anno

- Ogni livello di misurazione superiore include le caratteristiche di quelli inferiori
- La scala "nominale" ha il minimo di informazione
- La scala "a rapporto" ha il massimo di informazione

Implicazioni 2

- È possibile abbassare il livello di una misurazione, perdendo informazioni (ad es. da una scala a intervallo ad una ordinale)

Esempio

Posso trasformare l'età (misurata in anni dalla nascita) in una scala ordinale (fasce di età)

Implicazioni 3

- Non è possibile alzare il livello di una misurazione, perché non posso aggiungere un'informazione che non ho raccolto in precedenza

Esempio

Se ho rilevato la fascia d'età (scala ordinale), non posso passare all'età in anni (scala a rapporti)

- Se l'informazione aggiuntiva è compresa nella scala ordinale, allora posso alzare il livello

Esempio

Se ho rilevato la scolarità tramite i livelli scolastici (elementare, media inf., media superiore...), posso passare al numero di anni trascorsi a scuola (scala a rapporti, elementare=5, m.inf=8, ...)

Implicazioni

Per stabilire il livello di misurazione di una variabile devo considerare:

- la natura della variabile stessa
- il modo in cui è stata rilevata
- quali sono le relazioni valide all'interno della variabile
- **Attenzione**: il nome della variabile potrebbe non indicare la sua scala di misura (se non è un nome codificato)

Esempio

Il **reddito** è una variabile quantitativa che, in teoria, oscilla fra 0 e $+\infty$; se rilevo il valore esatto del reddito, sarà misurata a livello *rapporto* (unità di misura, origine assoluta); se rilevo delle fasce di reddito, sarà misurata a livello *ordinale* (non esiste un'unità di misura)

Varabili dicotomiche o binomiali, scale a ranghi

- Le variabili **binomiali** sono variabili che possono assumere 2 soli valori possibili (ad es. maschio/femmina, vero/falso, ...) sono variabili misurate a livello nominale e possono essere associate a simboli, testo o numeri.
- Se codifichiamo la variabile binomiale usando i valori numerici 0 o 1, diventa **dicotomica**
- In molte situazioni una variabile dicotomica si comporta matematicamente come una variabile a intervallo.

Scale a ranghi: i numeri indicano solo un ordine, ma...

per niente probabile	1	2	3	4	5	6	7	8	9	10	molto probabile

Scala a ranghi

- Le scale a ranghi sono scale in cui i numeri rappresentano un ordine (I premio, II arrivato, III scelta...)
- Sono variabili misurate a livello ordinale (fra il 1º posto e il 2º posto c'è lo stesso intervallo che c'è fra il 3º e il 4º? no!)
- Tuttavia, si è visto che se l'intervallo è abbastanza ampio (da 7 in su) noi abbiamo la tendenza a equi-spaziare
- Se poi usiamo anche una rappresentazione equi-spaziata allora la tendenza è ancora maggiore
- Per cui possono essere trattate sia come ordinali sia come intervallo
- In particolare le "scale Likert" (ps. sociale)

per niente probabile	1	2	3	4	5	6	7	8	9	10	molto probabile

Scale "Likert"

- Sono scale ampiamente usate in psicologia (soprattutto psicologia sociale)
- Sviluppate per misurare atteggiamenti e opinioni
- Possono essere pensate come una misurazione su una scala bipolare

max disac-cordo	1	2	3	4	5	6	7	8	9	max accordo

Riconoscere i livelli di misura

- Fra i livelli di misura, la tipologia più utilizzata è l'ultima.
- Come possiamo riconoscere a quale livello stiamo misurando?

1. I valori di ciò che stiamo misurando permettono valori diversi?
   - No, allora è una costante. Non è una variabile!
2. Prendiamo due qualunque valori possibili di quella variabile
3. Possiamo dire che sono uguali o sono diversi?
   - Sì, allora siamo almeno a livello nominale
4. Possiamo in qualche modo ordinare questi valori?
   - Sì, allora siamo almeno a livello ordinale
5. Possiamo identificare un'unità di misura specifica?
   - Sì, allora siamo almeno a livello intervallo
6. L'unità di misura ha un punto d'origine assoluto?
   - No, allora siamo a livello intervallo
   - Sì, allora siamo a livello a rapporto

Esercizio

	Descrizione variabile	N O I R
1	Genere (maschi, femmine)	
2	Anno di nascita	
3	Età	
4	Titolo di studio (licenza elementare, medie...)	
5	Anni di studio completati	
6	Tipo di mezzo di trasporto utilizzato x raggiungere il luogo di lavoro	
7	Ore passate su un mezzo di trasporto x raggiungere il luogo di lavoro	
8	Quante volte alla settimana mangi fuori casa (considerando solo i due pasti principali)	
9	Quoziente di Intelligenza	

Esercizio

	Descrizione variabile	N O I R
10	Quanto sei soddisfatto della tua vita in generale? 0) per niente; 1) poco; 2) abbastanza; 3) molto; 4) moltissimo	
11	Spesso i partiti politici vengono classificati di destra o di sinistra (e considerati come opposti) o di centro. Su una scala che va da sinistra a destra, come giudichi il tuo orientamento politico (1=sinistra – 10=destra)	
12	Su una scala da 1 a 10, quanto sei d'accordo che gli extracomunitari vadano espulsi dal paese?	