Elementi di Psicometria con Laboratorio di SPSS 1

01-Perché serve la statistica?
vers. 1.2a (9 ottobre 2014)

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

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

2014-2015

Introduzione

- Immaginiamo un mondo nel quale tutto sia uguale e nel quale le persone siano tra loro identiche in tutto e per tutto
- Se non ci sono diversità, tutto è uguale a tutto
- Dal punto di vista matematico tutte queste caratteristiche rappresentano delle costanti

- Fortunatamente il mondo non è così: la variabilità è una caratteristica essenziale della vita e del mondo sociale nel quale viviamo
- la statistica studia e interpreta questa variabilità
- la variabile è una qualunque caratteristica che viene misurata su diversi “individui” e che assume valori diversi

Cos'è la statistica

- Il termine statistica deriva da *statista* (uomo di stato) e originariamente indicava le misurazioni dello Stato.
- 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

La **statistica** come scienza si può suddividere in

- Statistica descrittiva (il primo passo) vs. statistica inferenziale (il passo successivo)
- 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

Cos'è la statistica

In altre parole

1. Statistica descrittiva
   - Fornisce modi per riassumere le informazioni raccolte da una moltitudine di fonti
   - Organizza in tabelle delle informazioni raccolte nella ricerca
   - Tramite formule semplici, trasforma dati complessi in indici che ne descrivono numericamente i principali aspetti
2. Statistica inferenziale
   - Non sempre si può fare un censimento della **popolazione**
   - Utilizziamo un **campione** estratto dalla popolazione
   - L'inferenza statistica stima la sicurezza con la quale possiamo generalizzare informazioni *da un campione all'intera popolazione*
3. Tecniche di esplorazione di dati
   - Mettono in evidenza le tendenze di una vasta quantità di dati
   - Semplificano, esplorano e riducono i dati

Cos'è la statistica

La **statistica** come scienza distingue fra

- Popolazione vs. campione
  - popolazione è l'insieme di tutti i "soggetti/oggetti" che si vorrebbe studiare
  - campione è l'insieme dei "soggetti/oggetti" (estratto dalla popolazione) che si studiano veramente
  - indice statistico è la sintesi "matematica" di un certo pensiero logico che viene applicato alle informazioni (variabili) raccolte su un campione (in particolare gli indici descrittivi)
  - parametro è il corrispettivo dell'indice statistico, ma calcolato/stimato sulla popolazione

Cos'è la statistica

La **statistica** come scienza distingue fra

- Popolazione vs. campione
  - popolazione è l'insieme di tutti i "soggetti/oggetti" che si vorrebbe studiare
  - campione è l'insieme dei "soggetti/oggetti" (estratto dalla popolazione) che si studiano veramente
  - indice statistico è la sintesi "matematica" di un certo pensiero logico che viene applicato alle informazioni (variabili) raccolte su un campione (in particolare gli indici descrittivi)
  - parametro è il corrispettivo dell'indice statistico, ma calcolato/stimato sulla popolazione

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 mai all'interno di un costrutto
- La possibilità di misurare un costrutto (tramite le variabili associate a certi aspetti del costrutto) è oggetto della *teoria della misurazione*

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 uno dei prodotti in offerta (un succo di frutta):

a) una signora piuttosto grassa 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.

Abbiamo raccolto delle informazioni (misurazioni) di alcune caratteristiche (variabili) su 3 persone (unità statistiche o casi statistici)
Sull'insieme degli individui possiamo incrociare diversi sistemi classificatori.

Livelli di misura

Riassumiamo in una tabella (simile ad un piano cartesiano, in insiemistica) le informazioni (i sistemi classificatori) che abbiamo:

	Genere	corporatura	confezioni	conto
a	F	grassa	12	180
b	M	media	2	50
c	F	mingherlina	4	40

- Il **genere** si limita a distinguere categorie fra loro diverse
- La **corporatura** oltre a distinguere le categorie, permette anche di metterle in ordine (da magro a grasso), ma non è possibile *quantificare* la differenza fra “mingherlina” e “normale” o fra “media” e “grassa”
- Le variabili **confezioni** e **conto** usano anche un’*unità di misura* 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
- I livelli di misura possono essere classificati anche secondo altre modalità

qualitative vs. quantitative
discrete vs. continue
nominali vs. ordinali vs. intervallo/rapporto

Attenzione

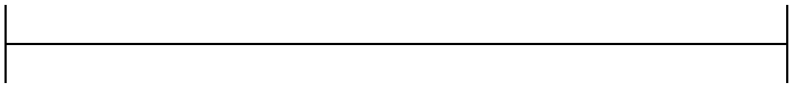
- Excel e SPSS possono lavorare con numeri e con stringhe (testo)
- Excel e SPSS non sanno nulla delle scale di misura
- Se sono numeri, li trattano come numeri

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
- 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**

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
- 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
- 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

- 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 appartenga contemporaneamente ai poveri e ai ricchi

- 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
- 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

- 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)

Variabili dicotomiche e binomiali

- 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**
- Molte variabili possono essere trasformate in dicotomiche (se il voto >=18, 1="esame superato" altrimenti 0="esame non superato")
- In molte situazioni una variabile dicotomica si comporta matematicamente come una variabile a intervallo.

Ogni volta che sarà possibile usare una variabile dicotomica come se fosse a intervallo, lo indicherò

Variabili ordinali quantitative

- Di per sé le variabili ordinali non sono quantitative
- Ma in alcuni contesti “umani” hanno caratteristiche molto vicine a quelle quantitative

1. Variabili a ranghi
2. Variabili misurate su una dimensione “psicologica”

Variabili ordinali a ranghi

- Un rango è semplicemente un punteggio assegnato in ordine crescente (o decrescente) in modo che casi con lo stesso rango abbiano lo stesso punteggio (ex equo)
- Esempio: i vincitori di una gara; l'ordine di arrivo può essere considerato come un rango
- Dal momento che si considera la posizione e che viene misurata con numeri interi progressivi, viene considerata **quantitativa**
- L'uso dei ranghi verrà affrontato meglio nella **Statistica non parametrica**

Variabili ordinali tipo “Likert”

- Misurano un atteggiamento su un continuum (in genere, disaccordo-accordo, niente-molto, brutto-bello)
- Nel rispondere tendiamo ad equi-spaziare fra il significato su un’estremità (disaccordo, niente, brutto) e l’altra (usando quindi una sorta di unità di misura personale)
- Misurate con almeno 5 “gradini” (0-4 o 1-5)
- Il punteggio che scegliamo è un’approssimazione all’intero legata allo strumento di misura

	max disac					max accor
Spesso mi sento fortemente legato a un potere più grande di me	1	2	3	4	5	6
Spesso provo un senso di connessione con la natura	1	2	3	4	5	6

Livello variabile

	Vostro libro	**Io e altri**
Categoriale o qualitativa	Nominale	Nominale
		Ordinale
Punteggio o quantitativa	Ordinale o a ranghi	Ordinale a ranghi o continua approssimata
	Intervallo	Intervallo
	Rapporto	Rapporto

La classificazione in un livello di misura (tipo NOIR) dipende dalle caratteristiche della variabile. Noi “misuriamo” un concetto (o costrutto), perciò la variabile deve rispettare le caratteristiche del costrutto. Ovviamente è meglio avere poche variabili qualitative e molte variabili quantitative, ma non bisogna forzare le caratteristiche della realtà per farle diventare ciò che vogliamo noi.

Esercizio 1

	Descrizione variabile	N O I R
1	Genere (maschi, femmine)	
2	Anno di nascita	
3	Età	
2	Stato civile	
3	Professione	
4	Titolo di studio (licenza elementare, medie. . .)	
5	Anni di studio completati	
6	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 i due pasti principali)	
9	Per lavarti, usi: 1) la doccia; 2) la vasca da bagno	

Esercizio 1

	Descrizione variabile	N O I R
1	Genere (maschi, femmine)	N(B)
2	Anno di nascita	I
3	Età	R
2	Stato civile	N
3	Professione	N
4	Titolo di studio (licenza elementare, medie. . .)	O
5	Anni di studio completati	R
6	Mezzo di trasporto utilizzato x raggiungere il luogo di lavoro	N
7	Ore passate su un mezzo di trasporto x raggiungere il luogo di lavoro	R
8	Quante volte alla settimana mangi fuori casa (considerando i due pasti principali)	R
9	Per lavarti, usi: 1) la doccia; 2) la vasca da bagno	N

Esercizio 2

	Descrizione variabile	N O I R
10	Mediamente, quante docce fai in una settimana (non importa quanto durano)	
11	Rispetto ai tuoi orientamenti sessuali, ti consideri: 1) eterosessuale; 2) omosessuale; 3) bisessuale; 4) transsessuale; 5) altro	
12	Quanto sei soddisfatto della tua vita in generale? 0) per niente; 1) poco; 2) abbastanza; 3) molto; 4) moltissimo	
13	Ammontare delle spese per le vacanze	
14	Numero di tatuaggi	
15	Ampiezza di ogni tatuaggio	
16	Il tatuaggio è colorato? (si, no)	
17	Hai piercing? 1) si 2) no	
18	Se si, quanti piercing hai?	
19	Quoziente di Intelligenza	

Esercizio 2

	Descrizione variabile	N O I R
10	Mediamente, quante docce fai in una settimana (non importa quanto durano)	R
11	Rispetto ai tuoi orientamenti sessuali, ti consideri: 1) eterosessuale; 2) omosessuale; 3) bisessuale; 4) transsessuale; 5) altro	N
12	Quanto sei soddisfatto della tua vita in generale? 0) per niente; 1) poco; 2) abbastanza; 3) molto; 4) moltissimo	O*
13	Ammontare delle spese per le vacanze	R
14	Numero di tatuaggi	R
15	Ampiezza di ogni tatuaggio	R
16	Il tatuaggio è colorato? (si, no)	N(B)
17	Hai piercing? 1) si 2) no	N(B)
18	Se si, quanti piercing hai?	R
19	Quoziente di Intelligenza	I

Esercizio 3

	Descrizione variabile	N O I R
20	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)	
21	Su una scala da 1 a 10, quanto sei d'accordo che gli extra-comunitari vadano espulsi dal paese?	
22	"Un alunno picchia chi non sa difendersi." 0=no; 1=qualche volta; 2=si	
23	A suo parere, il bullismo si manifesta principalmente tra compagni di classe: a) dello stesso sesso, b) di sesso diverso, c) entrambi	
24	Quali forme di bullismo ritiene maggiormente diffuse? a) verbale (umiliazioni, ingiurie, ricatti, minacce), b) fisica (percosse, furti, costrizioni), c) indiretta (ritorsioni, esclusione, mettere in giro storie).	

Esercizio 3

	Descrizione variabile	N O I R
20	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)	O*
21	Su una scala da 1 a 10, quanto sei d'accordo che gli extra-comunitari vadano espulsi dal paese?	O*
22	"Un alunno picchia chi non sa difendersi." 0=no; 1=qualche volta; 2=si	O*
23	A suo parere, il bullismo si manifesta principalmente tra compagni di classe: a) dello stesso sesso, b) di sesso diverso, c) entrambi	N
24	Quali forme di bullismo ritiene maggiormente diffuse? a) verbale (umiliazioni, ingiurie, ricatti, minacce), b) fisica (percosse, furti, costrizioni), c) indiretta (ritorsioni, esclusione, mettere in giro storie).	N