Statistica descrittiva

Dati statistici

Domenico De Stefano

a.a. 2021/2022

Indice

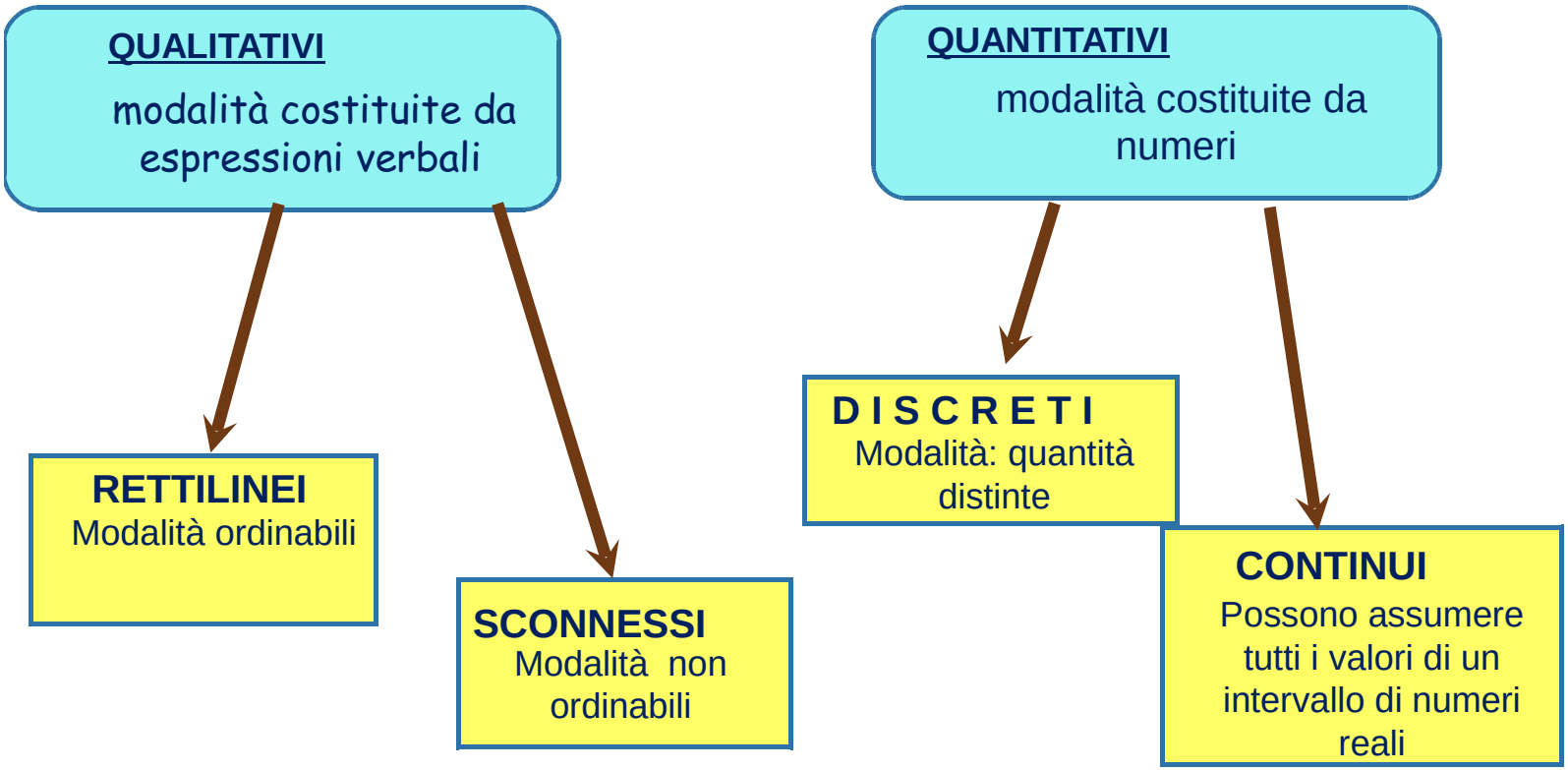
Terminologia elementare

Un dato statistico è il risultato della rilevazione (misurazione/osservazione) di un qualche carattere su un'unità statistica appartenente a una popolazione.

- unità statistica: il caso individuale componente del collettivo statistico;
- carattere (o variabile): ogni aspetto elementare oggetto di rilevazione nelle unità statistiche del collettivo;
- modalità di un carattere: i diversi modi con cui questo si presenta nelle unità statistiche del collettivo
- supporto: insieme (teorico) delle modalità di un carattere.

Nel seguito, i termini carattere e variabile verranno usati in modo interscambiabile.

Tipi di carattere



Variabili qualitative

- Una variabile è qualitativa se le modalità che si presentano sono espresse in forma verbale;
  - una variabile qualitativa è sconnessa se le sue modalità non implicano una graduazione (mutabile sconnessa);
  - una variabile qualitativa è ordinale se le sue modalità implicano una graduazione;
- le modalità possono essere predefinite a priori;
- a volte, nelle indagini, le modalità vengono desunte a posteriori dalla descrizione dettagliata che il rilevatore fa dello stato della singola unità relativamente al carattere in questione.

Esempio: qualitativa sconnessa

Ti è piaciuta l'ultima edizione del Festival di Sanremo?

- L'ho visto e mi è piaciuto
- L'ho visto e non mi è piaciuto
- Non l'ho visto

Esempio: qualitativa sconnessa 2

Qual è il tuo genere letterario preferito?

- Comico/umoristico
- Fantascienza
- Fantasy
- Giallo/noir/thriller
- Psicologico
- Romantico
- Storico
- Altro

Esempio: qualitativa sconnessa 2 (2)

Però c'era anche una domanda successiva che recitava

Se hai risposto "altro" alla domanda precedente puoi precisare qui per

la quale si sono osservate le seguenti risposte

Saggio storico/ Trattato filosofico
Sportivo/Biografico
biografie
urban fantasy
Avventura
Saggio scientifico

Si sono dunque ridefinite le modalità in modo da includere anche le risposte trovate in "altro"

Esempio: qualitativa sconnessa 2 (3)

Qual è il tuo genere letterario preferito?

- Comico/umoristico
- Fantascienza
- Fantasy ←*include Urban fantasy*
- Giallo/noir/thriller
- Psicologico
- Romantico
- Storico
- Saggistica ←*include Saggio storico/ Trattato filosofico; Saggio scientifico*
- Biografico ←*include Sportivo/Biografico; biografie*
- Avventura ←*include Avventura*

Esempio: qualitativa sconnessa 2 (4)

Si noti che le scelte fatte non sono neutre e possono essere discutibili

---

I criteri da tenere presenti sono

- rendere con la dovuta precisione le risposte degli intervistati
- mantenere un certo livello di aggregazione (mettere insieme cose molto simili) in modo da non frammentare troppo (n risposte diverse su n unità alla fine non sarebbero utili)

e sono ovviamente conflittuali.

Quale sia la scelta migliore dipende dal problema (ovviamente) ma anche dal numero di osservazioni.

Esempio: qualitativa ordinale

Quanto frequentemente bevi birra?

- Mai
- Raramente
- Qualche volta
- Spesso
- Ogni giorno
- Più volte al giorno

Variabili quantitative

- Una variabile è quantitativa se le modalità che si presentano sono espresse in forma numerica;
  - una variabile quantitativa è discreta se l'insieme delle sue modalità è finito oppure numerabile (detto in altri termini, se la quantità che rappresenta varia "a salti"). Spesso la loro rilevazione è frutto di un **conteggio**;
  - una variabile quantitativa è continua se l'insieme delle sue modalità è un intervallo, limitato o illimitato. Spesso la loro rilevazione è frutto di una **misurazione**;

  NB. Per la limitata precisione utilizzabile nel rilevare le misure, la distinzione tra variabile discreta e continua è convenzionale.
- e rispetto alle operazioni che è ragionevole fare
  - una variabile è intervallare se ha senso fare differenze tra valori ma non c'è uno zero naturale e non ha senso fare rapporti
  - una variabile è rapportabile se ha senso fare rapporti tra valori (c'è uno zero naturale)

Esempi: variabili quantitative

- Discreta rapportabile
  Quante volte sei stato al cinema negli ultimi tre mesi?
- Continua rapportabile
  Qual è la tua altezza (in centimetri)?
- Discreta non rapportabile
  In che anno sei nato?
- Continua non rapportabile
  Temperatura esterna

Esercizio

Che tipo di carattere è il prefisso telefonico?

(a) numerica, continua
(b) numerica, discreta
(c) categorica
(d) ordinale

Indice

Matrice dei dati

Dati del questionario:

variabile ↓

Stu.	sesso	Sanremo	···	sonno	studio
1	maschio	Non l'ho visto	···	8	2
2	femmina	L'ho visto e mi è piaciuto	···	6	30
3	maschio	Non l'ho visto	···	9	5
4	femmina	Non l'ho visto	···	8	25
⋮	⋮	⋮	⋮	⋮	⋮
52	femmina	Non l'ho visto	···	8	20

← unità statistica

Genere	Anno	Altezza	Peso	OreStudio	OreSonno	VotoMaturita	EconPol	DiffStat	Libri
Femmina	Primo	160	47.00	12	10	71	S	3	2
Femmina	Primo	169	54.00	4	7	70	S	3	4
Femmina	Primo	173	58.00	4	8	75	No	3	3
Femmina	Primo	170	73.00	15	8	78	S	4	0
Femmina	Primo	167	60.00	27	9	81	S	4	3
Femmina	Primo	157	54.00	10	7	72	No	3	2
Femmina	Terzo	165	56.00	15	8	77	S	4	6
Maschio	Terzo	182	66.00	30	8	75	S	2	10
Femmina	Primo	173	58.00	18	9	85	No	3	5
Femmina	Primo	172	50.00	24	8	84	No	5	2
Femmina	Primo	170	50.00	24	9	81	No	5	2
Femmina	Primo	169	62.00	2	8	60	S	4	0
Maschio	Oltre il terzo	180	78.00	20	8	80	S	3	3
Femmina	Oltre il terzo	164	54.00	20	8	100	No	3	0
Femmina	Primo	168	59.00	14	8	72	S	4	0
Maschio	Terzo	160	65.00	24	9	67	S	4	1
Femmina	Primo	168	60.00	14	8	74	No	3	2
Femmina	Terzo	165	95.00	20	8	76	S	4	3
Maschio	Terzo	180	61.00	15	7	80	S	3	0
Femmina	Primo	172	70.00	15	8	60	No	4	6
Maschio	Primo	180	60.00	24	8	78	No	5	2
Maschio	Primo	188	86.40	5	7	80	No	2	1
Maschio	Primo	180	68.00	7	8	74	S	3	4
Maschio	Terzo	175	80.00	10	7	75	No	5	0
Maschio	Oltre il terzo	180	67.00	13	6	62	S	3	1
Femmina	Terzo	165	52.00	7	8	70	No	4	0
Maschio	Primo	173	66.00	10	8	62	No	4	1
Femmina	Primo	165	54.00	20	7	68	No	3	2
Maschio	Primo	179	68.00	30	8	80	S	3	0
Maschio	Terzo	183	95.00	10	6	64	S	5	2
Maschio	Terzo	186	93.00	20	7	74	S	5	4
Femmina	Primo	169	61.00	30	8	75	No	3	0
Femmina	Primo	165	52.00	60	7	68	S	4	10
Femmina	Primo	153	50.00	35	8	82	No	3	4
Femmina	Oltre il terzo	169	68.00	8	7	85	S	4	3

L'agente arancio

- L'agente arancio è una mistura erbicida ampiamente usata nella guerra del Vietnam.
- L'agente arancio è stato collegato a patologie cancerose in molti studi epidemiologici.
- Per studiare l'assorbimento della diossina, nel 1987 le concentrazioni di diossina (in parti per trilione) vennero misurate nel plasma di veterani (soldati di terra) dell'esercito USA.
- Il campione era così composto
  - campione (non casuale) di veterani del Vietnam che servirono nel 1967-1968
  - campione (non casuale) di veterani che servirono in USA e Germania nel 1965-1971

L'agente arancio (vets): matrice dei dati

SUBJECT	DIOXIN	VETERAN
1	0	VIETNAM
2	0	VIETNAM
3	0	VIETNAM
4	0	VIETNAM
5	0	VIETNAM
.....	....	.....
739	9	OTHER
740	9	OTHER
741	10	OTHER
742	11	OTHER
743	15	OTHER

Ciliegi neri (`trees`)

Per 31 alberi di un parco sono stati rilevati

- il volume di legno,
- il diametro (rilevato a una prefissata altezza),
- l'altezza, riportati.

L'obiettivo dell'analisi è prevedere il volume ligneo sulla base di diametro e altezza.

diametro (in pollici)	altezza (in piedi)	volume del legno (in piedi^3)
8.3	70	10.3
8.6	65	10.3
8.8	63	10.2
10.5	72	16.4
10.7	81	18.8
10.8	83	19.7
11.0	66	15.6
11.0	75	18.2
11.1	80	22.6
11.2	75	19.9
11.3	79	24.2
11.4	76	21.0
11.4	76	21.4
11.7	69	21.3
12.0	75	19.1
12.9	74	22.2
12.9	85	33.8
13.3	86	27.4
13.7	71	25.7
13.8	64	24.9
14.0	78	34.5
14.2	80	31.7
14.5	74	36.3
16.0	72	38.3
16.3	77	42.6
17.3	81	55.4
17.5	82	55.7
17.9	80	58.3
18.0	80	51.5
18.0	80	51.0
20.6	87	77.0

Esempio: effetto del fumo sul peso dei neonati (babies)

Per 32 neonati si sono rilevati

- il peso alla nascita (in grammi),
- la durata della gravidanza (in settimane),
- la condizione rispetto al fumo della madre (S/N).

Interessa valutare se esista una relazione tra il peso alla nascita dei neonati e la durata della gravidanza e se questa relazione cambi rispetto alla condizione madre fumatrice/non fumatrice.

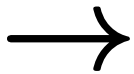
Peso	Durata gravidanza	Fumo
2940	38	S
2420	36	S
2760	39	S
2440	35	S
3301	42	S
2715	36	S
3130	39	S
2928	39	S
3446	42	S
2957	39	S
2580	38	S
3500	42	S
3200	41	S
3346	42	S
3175	41	S
2740	38	S
3130	38	N
2450	34	N
3226	40	N
2729	37	N
3410	40	N
3095	39	N
3244	39	N
2520	35	N
3523	41	N
2920	38	N
3530	42	N
3040	37	N
3322	39	N
3459	40	N
2619	35	N
2841	36	N

Gli stessi dati...

È molto spesso comodo codificare la condizione di fumatrice della madre con un numero (tipo: 1 fumatrice, 0 non fumatrice), anzichè con una lettera (S/N).

Ovviamente, i codici 0 ed 1 usati per le due condizioni sono ancora da considerarsi come etichette dei due gruppi.

Peso	Durata gravidanza	Fumo
2940	38	S
2420	36	S
2760	39	S
2440	35	S
3301	42	S
2715	36	S
3130	39	S
2928	39	S
3446	42	S
2957	39	S
2580	38	S
3500	42	S
3200	41	S
3346	42	S
3175	41	S
2740	38	S
3130	38	N
2450	34	N
3226	40	N
2729	37	N
3410	40	N
3095	39	N
3244	39	N
2520	35	N
3523	41	N
2920	38	N
3530	42	N
3040	37	N
3322	39	N
3459	40	N
2619	35	N
2841	36	N



Peso	Durata gravidanza	Fumo
2940	38	1
2420	36	1
2760	39	1
2440	35	1
3301	42	1
2715	36	1
3130	39	1
2928	39	1
3446	42	1
2957	39	1
2580	38	1
3500	42	1
3200	41	1
3346	42	1
3175	41	1
2740	38	1
3130	38	0
2450	34	0
3226	40	0
2729	37	0
3410	40	0
3095	39	0
3244	39	0
2520	35	0
3523	41	0
2920	38	0
3530	42	0
3040	37	0
3322	39	0
3459	40	0
2619	35	0
2841	36	0

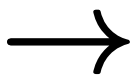
Livelli di fosfato nel plasma (`cholesterol`)

Misurazioni del livello di fosfato inorganico (mg/dl) nel plasma di soggetti obesi iperglicemici (OI), obesi non iperglicemici (ON) e di controllo (C) a un'ora dalla somministrazione di un test standard per l'assorbimento del glucosio.

OI	ON	C
2.3	3.0	3.0
4.1	4.1	2.6
4.2	3.9	3.1
4.0	3.1	2.2
4.6	3.3	2.1
4.6	2.9	2.4
3.8	3.3	2.8
5.2	3.9	3.4
3.1		2.9
3.7		2.6
3.8		3.1
		3.2

OI	ON	C
2.3	3.0	3.0
4.1	4.1	2.6
4.2	3.9	3.1
4.0	3.1	2.2
4.6	3.3	2.1
4.6	2.9	2.4
3.8	3.3	2.8
5.2	3.9	3.4
3.1		2.9
3.7		2.6
3.8		3.1
		3.2

Gli stessi dati possono essere rappresentati in forma di matrice.



Invece di avere una variabile e tre gruppi di osservazioni avremo due variabili di cui la seconda rappresenta il gruppo.

Fosfato	Tipo di paziente
2.3	OI
4.1	OI
4.2	OI
4.0	OI
4.6	OI
4.6	OI
3.8	OI
5.2	OI
3.1	OI
3.7	OI
3.8	OI
3.0	ON
4.1	ON
3.9	ON
3.1	ON
3.3	ON
2.9	ON
3.3	ON
3.9	ON
3.0	C
2.6	C
3.1	C
2.2	C
2.1	C
2.4	C
2.8	C
3.4	C
2.9	C
2.6	C
3.1	C
3.2	C

Titanic


LA DOMENICA DEL CORRIERE

Supplemento settimanale illustrato del nuovo CORRIERE DELLA SERA - Spedizione in abbonamento postale - Gruppo 2°

Anno 61 — N. 7 — 15 Febbraio 1959 — L. 40.—

Come il "Titanic". La motonave danese "Hans Hedtoft" di 2875 tonnellate, mentre si trovava a circa 37 miglia a sud di Capo Farvel (Groenlandia), è stata investita da un gigantesco iceberg. Di ritorno dal suo viaggio inaugurale, essa incrocia rotta su Copenhagen con 55 passeggeri e 40 uomini di equipaggio. L'ultimo messaggio diceva: "Affondiamo lentamente. Chiediamo assistenza immediata". Navi da costa e da pesca si sono dirette verso la "Hedtoft", numerosi aerei decollati da basi della Groenlandia hanno perlustrato la zona da cui era stato segnalato il sinistro. Ma invano. Della motonave nessuna traccia. A distanza di 47 anni si è rinnovata la tragedia del "Titanic". (Disegno di Walter Molino)

Il transatlantico britannico *RMS Titanic* affonda a seguito della collisione con un *iceberg* nella notte tra il 14 e il 15 aprile 1912.

Delle 2201 persone a bordo tra passeggeri ed equipaggio, sopravvivono solo 711.

Tra le polemiche che seguono al naufragio c'è chi sostiene che i passeggeri di III classe vennero trascurati nelle operazioni di evacuazioni, dando preferenza ai "ricchi".

Titanic

	Deceduto	Sopravv.
I Cl.	122	203
II Cl.	167	118
III Cl.	528	178
Equipaggio	673	212

I dati a disposizione sono riassumibili in una tabella a doppia entrata, in cui si riporta il numero di sopravvissuti e di deceduti a seconda della classe di appartenenza.

Il sospetto per cui i passeggeri di III classe vennero trascurati si traduce nel dire che le due caratteristiche osservate: sopravvivenza e classe, sono legate.

Il disastro del Titanic

Nome	Passeggero (tipologia)	Sopravvivenza
nome 1	II	sopravvissuto
nome 2	III	non sopravvissuto
nome 3	I	non sopravvissuto
⋮	⋮	⋮
nome 2201	equipaggio	sopravvissuto