Elementi di Psicometria con Laboratorio di SPSS 1

15-La probabilità

vers. 1.0a (26 novembre 2014)

Germano Rossi[1]

germano.rossi@unimib.it

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

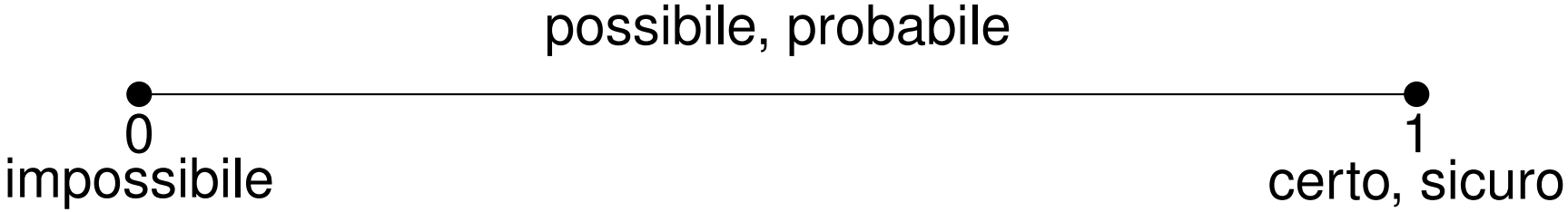
2014-2015

Metodi parametrici/non parametrici

- Metodi parametrici: metodi di inferenza che fanno riferimento ai parametri della popolazione (o meglio alla loro stima). Richiedono variabili che si distribuiscano (nella popolazione e spesso anche nel campione) secondo curve di probabilità predefinite (ad es. la normale)
- Metodi non parametrici: metodi di inferenza che non fanno riferimento ai parametri della popolazione.
  - La probabilità della statistica si può calcolare direttamente
  - La statistica utilizzata ha una propria distribuzione di probabilità
- In genere si usano quando si lavora con variabili qualitative o variabili quantitative di cui si sospetta (o si conosce) la non normalità

Probabilità di un evento

- Nel linguaggio comune, usiamo parole come “sicuro, certo, probabile, possibile, impossibile. . . ” per parlare dell’accadere di un fatto (evento):



- Un evento “sicuro, certo” accadrà sicuramente (e sarà un evento con probabilità uguale a 1, $P(x) = 1$)
- un evento impossibile non accadrà mai (non ci saranno eventi, e la sua probabilità si indica con zero $P(x) = 0$)

Teorie della probabilità

Usiamo il termine probabilità per indicare "*quanta* possibilità vi sia che un evento accada" fra diversi altri possibili (più di 0 e meno di 1, $0 < P(x) < 1$)

Esistono, sostanzialmente, 3 approcci teorici:

- classica: gli eventi possibili sono equiprobabili oppure la loro probabilità è teoricamente calcolabile o conoscibile (è stata sviluppata soprattutto per i giochi d'azzardo)
- frequentista: la probabilità dipende dalla frequenza con cui un certo evento è accaduto nel passato
- soggettiva: il rischio che siamo disposti a correre, la probabilità che associamo (senza fare calcoli espliciti) all'accadere di un evento

Probabilità classica

- Esistono diversi eventi possibili, ma *uno solo può accadere*; **alcuni** eventi possono essere utili ai nostri scopi, tutti gli altri no
- La probabilità si calcola dividendo gli eventi che ci servono rispetto a tutti gli eventi egualmente possibili

$$P(x) = \frac{\text{casi favorevoli}}{\text{casi possibili}} = \frac{f}{n}$$

Esempio

Ottenere un numero superiore o uguale a 5 tirando un dado:

- se il dado non è truccato, la probabilità di ogni numero è identica
- due sono gli eventi utili (5 e 6), 6 gli eventi possibili
- $P(5,6) = 2/6 = 1/3 = 0.33$

Probabilità classica

- La probabilità classica si può esprimere anche come rapporto di probabilità (in inglese, *odd*), ovvero il rapporto fra perdita e successo

$$\frac{\text{casi sfavorevoli}}{\text{casi favorevoli}} = \frac{s}{n-s} = \frac{s}{f}$$

- Ha senso solo all'interno della teoria classica, in cui si assume l'equiprobabilità degli eventi
- È usato soprattutto nel mondo anglosassone e nel mondo delle scommesse

Esempio

Il rapporto di probabilità relativo all'evento "ottenere un 5 o un 6 nel lancio di un dado" è

$$\frac{4}{2}$$

Probabilità frequentista

- Si usa quando gli eventi possibili non hanno la stessa probabilità
- La loro probabilità dipende dalle esperienze passate (frequenza empirica)
- Si chiama “frequentista” perché usiamo la frequenza (o meglio la proporzione) di una categoria come stima della probabilità del verificarsi di quella categoria
- È ampiamente usata dalle assicurazioni

Esempio

Se su 100 compiti di un esame, 15 sono insufficienti; in base alla probabilità frequentista, la probabilità di superare quell’esame sarebbe di 85/100=.85

Probabilità soggettiva

- È la stima che ciascuno di noi fa, sull'accadere di un evento su cui non abbiamo informazioni sicure
- Normalmente si usa nelle scommesse: "quanto scommetti che. . . " oppure nei giudizi: "mi consigli di fare. . . ", "devo fare così. . . o così. . . "

Esempio

- Decidere di affrontare un'esame universitario, dipende dalla probabilità soggettiva di superarlo.
- Accellerare quando un semaforo diventa giallo, dipende dalla stima soggettiva (probabilità) di riuscire a passare prima che diventi rosso.

Regole della probabilità

- Una volta trovata la probabilità di un evento (classica, frequentista o soggettiva) le regole che si applicano sono le stesse
- La somma delle probabilità di tutti gli eventi alternativi (e possibili) è (dev'essere) pari a 1 (uno solo è l'evento che può accadere)
- La probabilità di ciascuno degli eventi possibili è **la distribuzione di probabilità** di quel tipo di evento
- Le regole sono:
  - addizionale (regola OR)
  - moltiplicativa (regola AND)

Regola addizionale (OR, o)

La probabilità che accada un evento A oppure un evento B è pari a:

- somma delle singole probabilità meno la probabilità dell'intersezione: $P(A) + P(B) - P(A \cap B)$
- *se A e B sono mutualmente escludentisi fra di loro* cioè se $P(A \cap B) = 0$ (l'intersezione dei due eventi è nulla), allora:

$$P(A) + P(B)$$

Esempio

Probabilità di ottenere un numero pari con un lancio del dado
$P(2) + P(4) + P(6) - P(2 \cap 4 \cap 6) = 1/6 + 1/6 + 1/6 - 0 = 3/6 = 1/2$

Regola addizionale (OR, o)

- *se A e B non sono mutualmente escludentisi*: $P(A) + P(B) - P(A \cap B)$ (esistono eventi che cadono in entrambe le categorie)

Esempio

Probabilità di ottenere almeno un 5 o un 6 con 2 dadi

1. P(A=5,6 dado 1) = 2/6
2. P(B=5,6 dado 2) = 2/6
3. P(A∩ B=entrambi) = 4/36 $\Longrightarrow$
4. P(A) + P(B) - P(A∩ B) = 2/6 + 2/6 - 4/36 = (12+12-4) / 36 = 20/36

11	21	31	41	51	61
12	22	32	42	52	62
13	32	33	43	53	63
14	24	34	44	54	64
15	25	35	45	55	65
16	26	36	46	56	66

Regola addizionale (OR, o)

Esercizio

Lanciando 4 monete qual è la probabilità di ottenere almeno 3 Teste? [3 o 4 T]

Soluzione

$P(3t) + P(4t) = .25 + .0625 = .3125$

Esercizio

Non più di 1C?

Soluzione

$P(0C) + P(1C)$

Regola addizionale (OR, o)

Esercizio

Estraendo una carta da un mazzo di 40 carte, qual è la probabilità che sia una figura oppure una carta di cuori?

Soluzione

$P(\text{figura}) + P(\text{cuori}) - P(\text{figura di cuori}) = 12/40 + 10/40 - 3/40 = 19/40$

Regola moltiplicativa (AND, e)

- La probabilità che accadano **contemporaneamente** 2 eventi (*fra loro indipendenti* oppure *con re-immissione*) è pari al prodotto delle singole probabilità

$$P(AeB) = P(A) \times P(B)$$

Esempio

Lanciando 4 monete, qual è la probabilità di ottenere 4 teste simultaneamente?

$$1/2 * 1/2 * 1/2 * 1/2 = .5 * .5 * .5 * .5 = .5^4 = .0625$$

Regola moltiplicativa (AND, e)

Esempio

E la probabilità di rispondere giustamente ad un questionario a scelta multipla di 30 domande? (4 scelte)

$$\frac{1}{4}^{30} = 8.67e - 19$$

- Se gli eventi **non sono indipendenti fra loro**, si usa la probabilità condizionata
- Ovvero la probabilità di un evento condizionato dall'accadere di un'altro evento: $P(A/B)$

Applicare entrambe le regole

- Probabilità di 3 teste e 1 croce
- P(t) = P(c) = .5
- AND - 4 eventi simultanei = $.5^4 = .0625$
- OR - 4 possibili combinazioni (tttc, ttct, tctt, cttt) =

$$.0625 + .0625 + .0625 + .0625 = .0625 * 4 = .25$$

Probabilità condizionata

[Non presente nel libro]

	Genere	Età	Anni esperienza lavorativa
A	M	43	14
B	M	36	6
C	F	56	28
D	F	47	12
E	F	28	1
F	F	37	6
G	M	46	5
H	F	30	2
I	M	28	2
J	F	26	1

	F	M	
A(<=10)	4	3	7
B(>10)	2	1	3
	6	4	10

Esempio

Qual è la probabilità che, avendo estratto una femmina, questa abbia più di 10 anni di esperienza? (eventi dipendenti)

$2/6$ ovvero

$P(B/F) = P(B \cap F)/P(F)$

$$\frac{2}{10}/\frac{6}{10} = \frac{2}{10}\frac{10}{6} = \frac{2}{6}$$

Probabilità condizionale

- Probabilità di ottenere un certo evento dopo che si è verificato un primo evento:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{f_{A \cap B}}{f_B}$$

Esempio

Probabilità che, un asso estratto da un mazzo di 52 carte, sia di cuori (eventi *indipendenti*)
P(A=asso)=4/52=0.077
P(B=cuori)=13/52=0.25
$P(A \cap B) = 1/52 = 0.0192$
P(A/B)=(1/52) / (13/52) = 1/13 = .077

Eventi indipendenti

- Due eventi sono indipendenti fra loro se:

$$P(A/B) = P(A)$$

e se

$$P(B/A) = P(B)$$

ovvero se la probabilità del primo evento condizionata dal secondo è uguale a quella del primo

Esempio

$$P(A \cap B) = 1/52$$

P(A=asso)=4/52=1/13
P(A/B)=(1/52) / (13/52) = 1/13

P(B=cuori)=13/52=1/4
P(B/A)=(1/52) / (4/52) = 1/4

Probabilità e significatività

- Quando parliamo del 95% centrale di una distribuzione e del 5% estremo stiamo usando la probabilità
- Dal momento che 1 è un evento certo e 0 un evento impossibile, tutti i valori intermedi (da 0.000000000001 a 0.99999999999) sono probabilità intermedie di un qualunque evento espresse in forma di proporzione (1 su x)
- Dal momento che le % sono (in genere) più facili da capire e gestire, si usano le % al posto delle proporzioni.
- Hai fini pratici basta spostare di due posizioni la virgole
- a destra per passare dalle % alle proporzioni (25% = .25)
- a sinistra per passare dalle proporzioni alle percentuali (.05 = 5%)

Probabilità e campionamento

- Quando usiamo il principio della distribuzione campionaria, usiamo il ri-inserimento (cioè, dopo aver estratto un campione lo "ri-buttiamo nella popolazione) affinché ogni campione estratto abbia la stessa identica probabilità
- ma anche quando selezioniamo un campione, la singola osservazione estratta viene poi "ri-buttata" nella popolazione
- Quando però si estrae un campione di ricerca, non si usa il reinserimento, perché se la popolazione è grande, la differenza nelle probabilità delle osservazioni estratte successivamente sarebbe solo di poco inferiore... (1/10mila=0.0001; 1/(10mila-1)=0.00010001)

Implicazione 1

Ripetizioni di test per valutare la significatività all'interno di una ricerca

- Per ogni test statistico, uso un $\alpha = .05$
- Questo significa che ho il 5% di probabilità che il mio test (e quindi la mia scelta dell'ipotesi alternativa) sia risultato tale per effetto del caso
- Se ripeto questo test diverse volte, allora la mia probabilità di sbagliare viene moltiplicata per il numero delle ripetizioni
- Se faccio 20 test, almeno 1 (pari al 5% di 20) potrebbe essere significativo **assolutamente** per caso

Implicazione 1

Per ovviare a questo problema:

- fare solo le verifiche di ipotesi delle “ipotesi” della ricerca (vale a dire **non fare tutti i confronti** possibili immaginabili)
- se è necessario fare più confronti, usare dei valori di α più piccoli (ad es. con il criterio di Bonferroni)
- si divide l’α per il numero di confronti (2 confronti $= .05/2 = .025 = 2.5\%$; 4 confronti $= .05/4 = .0125 = 1.25\%$ e così via)
- Nella realtà, sono pochi gli articoli/report di ricerca che adottano questa tecnica

Implicazione 2

Significatività fra studi diversi

- Replicare uno studio significa in entrambi i casi correre un rischio del 5
- Ma se entrambi sono significativi, la loro significatività congiunta è pari al prodotto delle significatività
- Se il primo studio è significativo al 5% e anche il secondo, la loro significatività congiunta è di $.05 * .05 = .0025 = 0.25\%$ quindi se più ricerche simili risultano significative, diminuisce la probabilità che l'ipotesi alternativa sia errata e aumenta la probabilità che sia effettivamente vera