Verifica di fine parte IV · Contare, prevedere, misurare la dipendenza

Tempo: 60 minuti. Punteggio totale: 10 punti (punteggio indicato per ogni esercizio). Non è ammessa la calcolatrice: i calcoli sono pensati per essere svolti a mano, e le probabilità si possono lasciare in forma di frazione, con un valore approssimato quando è utile per il commento. Scrivi tutti i passaggi e dichiara sempre quale modello di conteggio usi: un risultato senza procedimento vale metà punteggio.

Test (0,5 punti ciascuno)

1. ● *(0,5 punti · cap. 9)* Quanti sono gli anagrammi, anche privi di senso, della parola TORTA?
   - **A.** 120
   - **B.** 60
   - **C.** 30
   - **D.** 24

2. ● *(0,5 punti · cap. 9)* Quale uguaglianza è vera?
   - **A.** $\binom{10}{3} = \binom{9}{3} + \binom{9}{4}$
   - **B.** $\binom{10}{3} = 3! \cdot D_{10,3}$
   - **C.** $\binom{10}{3} = \binom{10}{7}$
   - **D.** $\binom{10}{0} = 0$

3. ●● *(0,5 punti · cap. 10)* Si lanciano due dadi. Siano A = «la somma dei punti è 7» e B = «il primo dado dà 3». Allora:
   - **A.** A e B sono indipendenti
   - **B.** A e B sono incompatibili
   - **C.** $P(A \mid B) = \dfrac{1}{36}$
   - **D.** $P(A \cap B) = \dfrac{1}{3}$

4. ● *(0,5 punti · cap. 11)* Tra due caratteri quantitativi x e y, rilevati sulle stesse unità, il coefficiente di correlazione lineare è $r = -0{,}95$. Si può affermare che:
   - **A.** un aumento di x provoca una diminuzione di y
   - **B.** la retta di regressione di y su x ha pendenza $-0{,}95$
   - **C.** tra i due caratteri non c'è alcun legame, perché il coefficiente è negativo
   - **D.** i punti del diagramma a dispersione sono vicini a una retta con pendenza negativa

Calcolare

5. ●● *(1 punto · cap. 9)* Alle elezioni dei rappresentanti degli studenti si presentano 12 candidati: 7 del biennio e 5 del triennio. a) In quanti modi si possono eleggere 3 rappresentanti, che hanno tutti lo stesso ruolo? b) In quanti di questi casi è eletto almeno uno studente del triennio? c) In quanti modi si possono scegliere invece, tra gli stessi 12, un presidente, un vicepresidente e un segretario? Spiega perché i risultati di a) e di c) sono diversi.

Tabella V4.1 Colazione e risultato in una verifica di 400 studenti (dati inventati).

	Sufficiente	Insufficiente	Totale
Fa colazione	180	60	240
Non fa colazione	90	70	160
Totale	270	130	400

6. ●● *(1 punto · cap. 10)* In un liceo si è chiesto a 400 studenti se fanno colazione prima di venire a scuola, e si è registrato il risultato della verifica del giorno (Tabella V4.1). Si sceglie a caso uno studente; siano C = «fa colazione» e S = «ha preso la sufficienza». a) Calcola $P(S)$. b) Calcola $P(S|C)$ e $P(C|S)$, e spiega a parole perché sono diverse. c) Stabilisci con il calcolo se gli eventi S e C sono indipendenti.

Tabella V4.2 Genere di film preferito da 200 spettatori di un cinema, per fascia d'età (dati inventati).

	Commedia	Giallo	Documentario	Totale
Sotto i 30 anni	48	22	10	80
Dai 30 anni in su	42	48	30	120
Totale	90	70	40	200

7. ●● *(1 punto · cap. 11)* La Tabella V4.2 riporta il genere di film preferito da 200 spettatori. a) Scrivi la distribuzione marginale del genere preferito in frequenze relative percentuali. b) Scrivi le distribuzioni condizionate del genere preferito nelle due fasce d'età, in percentuale. c) Calcola le frequenze teoriche f' nell'ipotesi di indipendenza e indica in quali caselle la frequenza osservata se ne discosta di più. I due caratteri sono indipendenti?

8. ●● *(1 punto · cap. 10)* Un test ha 5 domande, ciascuna con 4 risposte di cui una sola esatta. Uno studente che non ha studiato risponde a caso a tutte, in modo indipendente. a) Calcola la probabilità che dia esattamente 3 risposte esatte. b) La sufficienza richiede almeno 3 risposte esatte: con quale probabilità la ottiene? c) Completa la tabella delle probabilità per 0, 1, …, 5 risposte esatte e trova il numero più probabile di risposte esatte; confrontalo con np.

Dimostrare

9. ●● *(1 punto · cap. 9)* a) Dimostra per induzione che, per ogni $n \geq 1$, $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \ldots + n(n+1) = \dfrac{n(n+1)(n+2)}{3}$. b) Una compagna ha controllato al computer l'uguaglianza per tutti gli n da 1 a 10 000 e sostiene che basta come dimostrazione. Spiega perché ha torto, con un esempio tratto dal capitolo 9, e in che cosa il principio di induzione matematica si distingue dall'induzione delle scienze sperimentali.

Problemi

10. ●● *(1 punto · cap. 10)* Per una malattia (dati inventati) la prevalenza nella popolazione è del 4%; un test ha sensibilità dell'80% e specificità del 95%. a) Rappresenta la situazione con le frequenze naturali su 10 000 persone (albero o tabella a doppia entrata). b) Una persona scelta a caso risulta positiva: calcola la probabilità che sia malata, con le frequenze naturali e poi con la formula di Bayes, e commenta il risultato. c) La persona ripete un secondo test dello stesso tipo, indipendente dal primo, e risulta di nuovo positiva. Qual è ora la probabilità che sia malata?

11. ●● *(1 punto · cap. 11)* Cinque studenti (dati inventati) indicano le ore x dedicate allo studio nella settimana prima di una prova e il punteggio y ottenuto, su 20 punti: (1; 8), (2; 10), (3; 12), (4; 16), (5; 14). a) Calcola le medie $\bar{x}$ e $\bar{y}$, la covarianza e le deviazioni standard σ_x e σ_y (dividendo per n), poi il coefficiente r. b) Scrivi la retta di regressione di y su x e calcola i residui. c) Che punteggio prevede la retta per 3,5 ore di studio? E per 10 ore? Commenta le due previsioni. d) Si può concludere che ogni ora di studio in più fa guadagnare 1,8 punti? Motiva la risposta.

Sfida

12. ●●● *(1 punto · cap. 9–10)* Due scatole identiche contengono 10 lampadine ciascuna: nella prima 2 sono difettose, nella seconda 4. Si sceglie una scatola a caso e se ne estraggono contemporaneamente 3 lampadine. a) Calcola, per ciascuna scatola, la probabilità che tra le 3 lampadine ce ne sia esattamente una difettosa. b) Calcola la probabilità che ci sia esattamente una lampadina difettosa. c) Tra le 3 estratte ce n'è esattamente una difettosa: qual è la probabilità che vengano dalla seconda scatola? d) Se invece nessuna delle 3 fosse difettosa, quale sarebbe la probabilità che vengano dalla prima scatola? Confronta le due informazioni: quale dice di più sulla scatola?

Soluzioni

1. B. Le 5 lettere contengono la T due volte e le altre una volta: le permutazioni con ripetizione sono $\frac{5!}{2!} = \frac{120}{2} = 60$. $120 = 5!$ conterebbe come diversi i due anagrammi che si ottengono scambiando le due T, che invece sono la stessa parola; $30 = \frac{5!}{2! \cdot 2!}$ supporrebbe due lettere ripetute.

2. C. Per la simmetria dei coefficienti binomiali $\binom{10}{3} = \binom{10}{10-3} = \binom{10}{7} = 120$. La formula di Stifel dà $\binom{10}{3} = \binom{9}{2} + \binom{9}{3}$: la somma in A vale $84 + 126 = 210 = \binom{10}{4}$. In B la relazione è rovesciata: $\binom{10}{3} = \frac{D_{10,3}}{3!} = \frac{720}{6}$. Infine $\binom{10}{0} = 1$.

3. A. I 36 esiti sono equiprobabili. $A = \{(1; 6), (2; 5), (3; 4), (4; 3), (5; 2), (6; 1)\}$, quindi $P(A) = \frac{6}{36} = \frac{1}{6}$; $P(B) = \frac{1}{6}$; $A \cap B = \{(3; 4)\}$, quindi $P(A \cap B) = \frac{1}{36} = P(A) \cdot P(B)$: gli eventi sono indipendenti (sapere che il primo dado dà 3 non cambia la probabilità di fare 7: $P(A|B) = \frac{1}{6}$). Non sono incompatibili, perché possono verificarsi insieme; D somma le probabilità invece di moltiplicarle.

4. D. $r = -0{,}95$ è vicino a -1: il legame lineare è forte e decrescente, cioè i punti si dispongono vicino a una retta con pendenza negativa. A confonde correlazione e causalità; B è falsa perché la pendenza è $m = r \cdot \frac{\sigma_y}{\sigma_x}$, che vale $-0{,}95$ solo se $\sigma_x = \sigma_y$; C è falsa: il segno di r indica il verso del legame, non la sua assenza.

5. a) L'ordine non conta e non ci sono ripetizioni: combinazioni, $C_{12,3} = \frac{12 \cdot 11 \cdot 10}{3!} = 220$. b) Conviene contare il complementare: i gruppi formati solo da studenti del biennio sono $C_{7,3} = \frac{7 \cdot 6 \cdot 5}{3!} = 35$, quindi quelli con almeno uno studente del triennio sono $220 - 35 = 185$. (Controllo per casi disgiunti: $5 \cdot \binom{7}{2} + \binom{5}{2} \cdot 7 + \binom{5}{3} = 105 + 70 + 10 = 185$.) c) Ora i ruoli sono diversi e l'ordine conta: disposizioni semplici, $D_{12,3} = 12 \cdot 11 \cdot 10 = 1320$. Ogni gruppo di tre eletti di a) si può ripartire nei tre ruoli in $3! = 6$ modi, e infatti $220 \cdot 6 = 1320$.

6. a) $P(S) = \frac{270}{400} = \frac{27}{40} = 0{,}675$. b) $P(S|C) = \frac{180}{240} = \frac{3}{4} = 0{,}75$: tra chi fa colazione, i tre quarti hanno la sufficienza. $P(C|S) = \frac{180}{270} = \frac{2}{3} \approx 0{,}67$: tra chi ha la sufficienza, due su tre fanno colazione. Le due probabilità hanno lo stesso numeratore (i 180 studenti che fanno colazione e hanno la sufficienza) ma restringono lo spazio campionario a gruppi diversi: nel primo caso ai 240 che fanno colazione, nel secondo ai 270 sufficienti. c) $P(S|C) = 0{,}75 \neq P(S) = 0{,}675$: gli eventi non sono indipendenti. Equivalentemente, $P(S \cap C) = \frac{180}{400} = 0{,}45$, mentre $P(S) \cdot P(C) = 0{,}675 \cdot 0{,}6 = 0{,}405$. (La dipendenza non dice che la colazione sia la causa del voto: potrebbero contare abitudini diverse dei due gruppi.)

7. a) Su 200 spettatori: commedia $\frac{90}{200} = 45\%$, giallo $\frac{70}{200} = 35\%$, documentario $\frac{40}{200} = 20\%$ (totale 100%). b) Sotto i 30 anni (su 80): commedia $\frac{48}{80} = 60\%$, giallo $\frac{22}{80} = 27{,}5\%$, documentario $\frac{10}{80} = 12{,}5\%$. Dai 30 anni in su (su 120): commedia $\frac{42}{120} = 35\%$, giallo $\frac{48}{120} = 40\%$, documentario $\frac{30}{120} = 25\%$. c) $f' = \frac{(\text{totale di riga}) \cdot (\text{totale di colonna})}{200}$. Sotto i 30 anni: $\frac{80 \cdot 90}{200} = 36$, $\frac{80 \cdot 70}{200} = 28$,

$\frac{80 \cdot 40}{200}$ =16; dai 30 anni in su: 54, 42, 24 (le righe e le colonne hanno gli stessi totali della tabella osservata). Gli scarti tra osservate e teoriche sono +12, −6, −6 nella prima riga e −12, +6, +6 nella seconda: la differenza maggiore è nella colonna della commedia (48 contro 36 tra i giovani, 42 contro 54 tra gli altri). Le distribuzioni condizionate sono diverse e le frequenze osservate non coincidono con quelle teoriche: i due caratteri non sono indipendenti; i più giovani preferiscono più spesso la commedia.

8. Prove ripetute con $n = 5$ e $p = \frac{1}{4}$. a) $P(3) = C_{5,3} \cdot (\frac{1}{4})^3 \cdot (\frac{3}{4})^2 = 10 \cdot \frac{1}{64} \cdot \frac{9}{16} = \frac{90}{1024} = \frac{45}{512} \approx 0{,}088$. b) $P(4) = 5 \cdot \frac{1}{256} \cdot \frac{3}{4} = \frac{15}{1024}$, $P(5) = \frac{1}{1024}$; quindi $P(\text{almeno 3}) = \frac{90 + 15 + 1}{1024} = \frac{106}{1024} = \frac{53}{512} \approx 0{,}10$: rispondendo a caso la sufficienza arriva circa una volta su dieci. c) Con denominatore 1024 i numeratori, per 0, 1, 2, 3, 4, 5 risposte esatte, sono $3^5 = 243$, $5 \cdot 3^4 = 405$, $10 \cdot 3^3 = 270$, $10 \cdot 3^2 = 90$, $5 \cdot 3 = 15$, 1, con somma 1024 ✓. Il numero più probabile di risposte esatte è 1 (probabilità $\frac{405}{1024} \approx 0{,}40$), vicino a $np = 5 \cdot \frac{1}{4} = 1{,}25$.

9. a) Sia $P(n)$ l'uguaglianza da dimostrare. *Passo base* ($n = 1$): il primo membro vale $1 \cdot 2 = 2$, il secondo $\frac{1 \cdot 2 \cdot 3}{3} = 2$ ✓. *Passo induttivo*: supponiamo vera $P(n)$ e aggiungiamo ai due membri il termine successivo $(n + 1)(n + 2)$:

$1 \cdot 2 + \ldots + n(n + 1) + (n + 1)(n + 2) =$

$\frac{n(n + 1)(n + 2)}{3} + (n + 1)(n + 2)$

$= (n + 1)(n + 2) \cdot \frac{n + 3}{3} = \frac{(n + 1)(n + 2)(n + 3)}{3}$, che è proprio $P(n + 1)$, cioè la formula con $n + 1$ al posto di n. Per il principio di induzione l'uguaglianza vale per ogni $n \geq 1$. c.v.d. b) Diecimila verifiche, per quanto numerose, riguardano un numero finito di casi e non dicono nulla sul caso successivo: manca il passo induttivo. Il numero $n^2 + n + 41$ è primo per tutti gli n da 0 a 39, ma non per $n = 40$ ($1681 = 41^2$). L'induzione matematica è una deduzione: il passo base e il passo induttivo, dimostrati, garantiscono la proprietà per tutti i naturali, perché ogni naturale si raggiunge da quello iniziale con un numero finito di passi; le *Indicazioni* parlano di «invarianza delle leggi del pensiero». L'induzione delle scienze sperimentali generalizza da molte osservazioni a una legge, presupponendo l'«invarianza delle leggi dei fenomeni»: la conclusione resta sempre smentibile da un'osservazione futura (il problema posto da Hume).

10. a) Su 10 000 persone: malate $4\% \cdot 10\,000 = 400$, di cui positive $80\% \cdot 400 = 320$ e negative 80; sane 9600, di cui positive (falsi positivi) $5\% \cdot 9600 = 480$ e negative 9120. Totale dei positivi: 320 + 480 = 800. b) Con le frequenze naturali $P(M|+) = \frac{320}{800} = \frac{2}{5} = 0{,}4$. Con Bayes: $P(M|+) = \frac{0{,}04 \cdot 0{,}8}{0{,}04 \cdot 0{,}8 + 0{,}96 \cdot 0{,}05} = \frac{0{,}032}{0{,}032 + 0{,}048} = \frac{0{,}032}{0{,}08} = 0{,}4$. Anche con un test abbastanza accurato, una persona positiva è malata solo nel 40% dei casi: la malattia è poco diffusa e i falsi positivi, il 5% di una folla di 9600 sani, sono più numerosi dei veri positivi. c) La probabilità a posteriori 0,4 diventa la nuova probabilità a priori: $\frac{0{,}4 \cdot 0{,}8}{0{,}4 \cdot 0{,}8 + 0{,}6 \cdot 0{,}05} = \frac{0{,}32}{0{,}35} = \frac{32}{35} \approx 0{,}91$. Con le frequenze: degli 800 positivi, 320 malati danno di nuovo positivo in $80\% \cdot 320 = 256$ casi, 480 sani in $5\% \cdot 480 = 24$: $\frac{256}{280} = \frac{32}{35}$ ✓.

11. a) $\bar{x} = \frac{15}{5} = 3$, $\bar{y} = \frac{60}{5} = 12$. Gli scarti sono −2, −1, 0, 1, 2 per x e −4, −2, 0, 4, 2 per y. $\text{cov}(x, y) = \frac{8 + 2 + 0 + 4 + 4}{5} = \frac{18}{5} = 3{,}6$; $\sigma_x^2 = \frac{4 + 1 + 0 + 1 + 4}{5} = 2$, $\sigma_y^2 = \frac{16 + 4 + 0 + 16 + 4}{5} = 8$, quindi $\sigma_x = \sqrt{2}$, $\sigma_y = \sqrt{8} = 2\sqrt{2}$ e $r = \frac{3{,}6}{\sqrt{2} \cdot 2\sqrt{2}} = \frac{3{,}6}{4} = 0{,}9$: correlazione lineare positiva forte. b) $m = \frac{3{,}6}{2} = 1{,}8$, $q = \bar{y} - m\bar{x} = 12 - 5{,}4 = 6{,}6$: $y = 1{,}8x + 6{,}6$. I valori previsti sono 8,4; 10,2; 12; 13,8; 15,6 e i residui −0,4; −0,2; 0; 2,2; −1,6, con somma 0 ✓. c) Per $x = 3{,}5$: $1{,}8 \cdot 3{,}5 + 6{,}6 = 12{,}9$ punti, una previsione ragionevole perché 3,5 è dentro l'intervallo dei dati (interpola-

zione). Per x =10: 18 + 6,6 =24,6 punti, più del massimo possibile (20): è un'estrapolazione lontana dai dati, dove nulla garantisce che il legame resti lineare. d) No. La pendenza 1,8 descrive come varia in media il punteggio passando da uno studente a un altro che ha studiato un'ora in più, in un campione di soli cinque studenti; non è la misura dell'effetto di un'ora in più su una stessa persona. Una correlazione, anche forte, non prova una relazione di causa: potrebbero contare variabili confondenti, come l'interesse per la materia o la preparazione precedente.

12. a) I casi possibili sono $C_{10,3} = 120$ per entrambe le scatole. Prima scatola (2 difettose, 8 buone): $\frac{\binom{2}{1}\cdot\binom{8}{2}}{120} = \frac{2\cdot 28}{120} = \frac{56}{120} = \frac{7}{15}$. Seconda scatola (4 difettose, 6 buone): $\frac{\binom{4}{1}\cdot\binom{6}{2}}{120} = \frac{4\cdot 15}{120} = \frac{60}{120} = \frac{1}{2}$. b) Indicate con H_1 e H_2 le due scatole ($P(H_1) = P(H_2) = \frac{1}{2}$) ed E l'evento «esattamente una difettosa», per la probabilità totale $P(E) = \frac{1}{2}\cdot\frac{7}{15} + \frac{1}{2}\cdot\frac{1}{2} = \frac{7}{30} + \frac{1}{4} = \frac{14}{60} + \frac{15}{60} = \frac{29}{60}$. c) Per Bayes $P(H_2|E) = \frac{1}{4} : \frac{29}{60} = \frac{1}{4}\cdot\frac{60}{29} = \frac{15}{29} \approx 0{,}52$ (e $P(H_1|E) = \frac{14}{29}$, somma 1 ✓). d) Con nessuna difettosa: prima scatola $\frac{\binom{8}{3}}{120} = \frac{56}{120} = \frac{7}{15}$, seconda $\frac{\binom{6}{3}}{120} = \frac{20}{120} = \frac{1}{6}$. La probabilità totale di nessuna difettosa è $\frac{1}{2}\cdot\frac{7}{15} + \frac{1}{2}\cdot\frac{1}{6} = \frac{7}{30} + \frac{1}{12} = \frac{19}{60}$, quindi $P(H_1|\text{nessuna difettosa}) = \frac{7}{30} : \frac{19}{60} = \frac{7}{30}\cdot\frac{60}{19} = \frac{14}{19} \approx 0{,}74$. Trovare una sola difettosa lascia le due scatole quasi alla pari (0,52 contro 0,48), perché l'evento ha probabilità simili nelle due scatole ($\frac{7}{15}$ e $\frac{1}{2}$); trovarne nessuna sposta molto di più la probabilità verso la prima scatola, perché l'evento è quasi tre volte più probabile lì che nella seconda: un'informazione conta tanto più quanto più diverse sono le sue verosimiglianze.