Matematica 3 · Parte V · Statistica a due variabili · versione 1.0
Verifica: Verifica di fine parte V · pagine 381–382 del volume completo · estratto per lo studio

Verifica di fine parte V · Statistica a due variabili

Tempo: 60 minuti. Punteggio totale: 10 punti (punteggio indicato per ogni esercizio). È ammessa la calcolatrice scientifica. Scrivi tutti i passaggi: un risultato senza procedimento vale metà punteggio.

Test (0,5 punti ciascuno)

1. ● *(0,5 punti · cap. 14)* In una tabella a doppia entrata con N = 120, una casella sta in una riga con totale 40 e in una colonna con totale 30. La sua frequenza teorica è:
   - **A.** 12
   - **B.** 10
   - **C.** 70
   - **D.** $\frac{1}{3}$

2. ● *(0,5 punti · cap. 14)* Se per due caratteri quantitativi si trova $r = 0$, si può concludere che:
   - **A.** i due caratteri sono indipendenti
   - **B.** i punti stanno su una retta orizzontale
   - **C.** la covarianza è uguale a 1
   - **D.** non c'è un legame lineare, ma può esservi un legame di altro tipo

3. ● *(0,5 punti · cap. 14)* La retta di regressione di un insieme di dati è $y = 2x + 1$ e la media delle x è 3. La media delle y è:
   - **A.** 7
   - **B.** 3
   - **C.** 6
   - **D.** non si può calcolare

4. ●● *(0,5 punti · cap. 14)* Per stimare il tempo che gli abitanti di una città passano sui mezzi pubblici, quale campione è un campione casuale semplice?
   - **A.** i passeggeri di un autobus scelto a caso
   - **B.** chi risponde a un questionario sul sito del comune
   - **C.** 500 persone estratte a caso dall'elenco di tutti i residenti
   - **D.** tutti i residenti di un quartiere scelto a caso

Calcolare

5. ● *(1 punto · cap. 14)* A 150 studenti (dati inventati) si chiede se fanno colazione al mattino e se a metà mattina si sentono stanchi. Tra i 90 che fanno colazione, 18 si sentono stanchi; tra i 60 che non la fanno, 27 si sentono stanchi. a) Scrivi la tabella a doppia entrata con le distribuzioni marginali. b) Calcola la percentuale di stanchi in ciascun gruppo.

6. ●● *(1 punto · cap. 14)* Per la tabella dell'esercizio precedente calcola le frequenze teoriche, l'indice χ^2 e l'indice V di Cramér. I due caratteri sono indipendenti?

7. ●● *(1,5 punti · cap. 14)* Una famiglia registra in cinque giorni la temperatura massima x (°C) e il consumo di acqua y (litri); i dati sono inventati: $x = 14, 18, 22, 26, 30$; $y = 180, 200, 230, 240, 270$. Calcola le medie, la covarianza e il coefficiente di correlazione lineare.

8. ●● *(1 punto · cap. 14)* Per i dati dell'esercizio precedente trova la retta di regressione di y su x, prevedi il consumo in un giorno con massima 24 °C e calcola il residuo del giorno con 26 °C.

Problemi

9. ●● *(1,5 punti · cap. 14)* Per 8 auto usate dello stesso modello (dati inventati) l'età x (anni) ha media 4 e deviazione standard 2, il prezzo y (migliaia di euro) ha media 12 e deviazione standard 3, e $r = -0{,}9$. a) Trova la retta di regressione del prezzo sull'età. b) Che prezzo prevede per un'auto di 6 anni? c) E per una di 14 anni? Commenta. d) Quale quota della varianza dei prezzi spiega l'età?

10. ●● *(1 punto · cap. 14)* **Dimostra** che, se a ogni valore y_i si aggiunge la stessa costante c, la retta di regressione di y su x mantiene la stessa pendenza e il suo termine noto aumenta di c.

Sfida

11. ●●● *(1 punto · cap. 14)* Una tabella 2×2 ha totali di riga 50 e 50, totali di colonna 60 e 40, e la casella in alto a sinistra contiene a. a) Scrivi le altre caselle in funzione di a e trova i valori possibili di a. b) Dimostra che $\chi^2 = \dfrac{(a - 30)^2}{6}$. c) Per quale a i caratteri sono indipendenti? Qual è il valore massimo di V? Perché non può essere 1?

Soluzioni

1. B. $40 \cdot 30 : 120 = 10$.

2. D. r misura solo il legame lineare.

3. A. La retta passa per il baricentro: $\bar{y} = 2 \cdot 3 + 1 = 7$.

4. C. Solo l'estrazione a caso dall'elenco completo dà a ogni gruppo di abitanti la stessa probabilità; gli altri metodi favoriscono chi usa i mezzi, chi frequenta il sito o un solo quartiere.

5. a) Colazione sì: stanchi 18, non stanchi 72 (totale 90). Colazione no: stanchi 27, non stanchi 33 (totale 60). Totali di colonna: stanchi 45, non stanchi 105; N = 150. b) Stanchi: 18 : 90 = 20% tra chi fa colazione, 27 : 60 = 45% tra chi non la fa.

6. Frequenze teoriche: $90 \cdot 45 : 150 = 27$; $90 \cdot 105 : 150 = 63$; $60 \cdot 45 : 150 = 18$; $60 \cdot 105 : 150 = 42$. Tutte le differenze valgono ±9: $\chi^2 = 81 \cdot (1/27 + 1/63 + 1/18 + 1/42) = 75/7 \approx 10{,}71$. $V = \sqrt{(10{,}71 : 150)} \approx 0{,}27$. Non sono indipendenti: c'è un'associazione debole-moderata (le distribuzioni condizionate, 20% e 45%, sono diverse).

7. $\bar{x} = 110 : 5 = 22$; $\bar{y} = 1120 : 5 = 224$. Scarti di x: −8, −4, 0, 4, 8; di y: −44, −24, 6, 16, 46. Σ prodotti = 352 + 96 + 0 + 64 + 368 = 880, cov = 880 : 5 = 176. $\Sigma(x - \bar{x})^2 = 160$, $\Sigma(y - \bar{y})^2 = 4920$. $r = 880 : \sqrt{(160 \cdot 4920)} \approx 0{,}992$: correlazione positiva fortissima.

8. $m = 880 : 160 = 5{,}5$ L/°C; $q = 224 - 5{,}5 \cdot 22 = 103$. Retta: $y = 5{,}5x + 103$. Per $x = 24$: $y = 235$ L. Per $x = 26$ il valore previsto è 246, il residuo $240 - 246 = -6$ L.

9. a) $m = r \cdot \sigma_y : \sigma_x = -0{,}9 \cdot 3 : 2 = -1{,}35$; $q = 12 + 1{,}35 \cdot 4 = 17{,}4$. Retta: $y = -1{,}35x + 17{,}4$. b) $y = 17{,}4 - 8{,}1 = 9{,}3$, cioè circa 9300 €. c) $y = 17{,}4 - 18{,}9 = -1{,}5$: un prezzo negativo è assurdo; 14 anni è probabilmente fuori dall'intervallo dei dati (estrapolazione), dove il calo non può restare lineare. d) $r^2 = 0{,}81$: l'81%.

10. I nuovi dati sono $y_i + c$, con media $\bar{y} + c$; gli scarti $(y_i + c) - (\bar{y} + c) = y_i - \bar{y}$ non cambiano, quindi la covarianza non cambia, e nemmeno σ_x^2: la pendenza resta m. Il nuovo termine noto è $(\bar{y} + c) - m\bar{x} = q + c$. c.v.d.

11. a) Prima riga a e 50 − a; seconda riga 60 − a e a − 10. Tutte ≥ 0: $10 \le a \le 50$. b) Le frequenze teoriche sono 30, 20, 30, 20; le differenze tra osservate e teoriche valgono a − 30, −(a − 30), −(a − 30), a − 30. $\chi^2 = (a - 30)^2 \cdot (1/30 + 1/20 + 1/30 + 1/20) = (a - 30)^2 \cdot 1/6$. c) Indipendenza per $a = 30$. $V = \sqrt{(\chi^2 : 100)} = |a - 30| : \sqrt{600}$, massimo per $a = 10$ o $a = 50$: $V = 20 : \sqrt{600} \approx 0{,}82$. Non può essere 1 perché le marginali sono diverse (50–50 e 60–40): una colonna non può coincidere con una riga, quindi nessuna modalità di un carattere determina quella dell'altro.