Esercizi sul calcolo delle Probabilità

a cura di Giovanni M. Marchetti

28 febbraio 2019

Questi esercizi hanno la struttura degli esercizi dell'esame scritto di Statistica, cioè sono di 3 tipi: vero/falso, scelta multipla, risposta aperta.

1. Quale delle seguenti affermazioni è sempre vera per qualsiasi coppia di eventi A e B definiti in uno spazio degli eventi elementari S?

   (a) Se l'unione degli eventi A e B è l'insieme vuoto, allora sia A che B sono insiemi vuoti.

   (b) Se gli eventi A e B sono collettivamente esaustivi, allora $A \cup B = \emptyset$.

   (c) Se l'intersezione degli eventi A e B è l'insieme vuoto, allora A e B sono collettivamente esaustivi.

   (d) Se gli eventi A e B sono mutuamente esclusivi e collettivamente esaustivi, allora l'unione di A e B non è necessariamente uguale a S.

   Soluzione. A) Vera: $A \cup B = \emptyset$ non può avvenire se uno dei due eventi contiene degli elementi. B) Falso: la definizione è $A \cup B = S$. C) Falso: $A \cap B = \emptyset$ implica che gli eventi sono disgiunti non esaustivi. D) Falso: in questo caso gli eventi sono una partizione di S e quindi la loro unione è necessariamente uguale a S.

2. Supponiamo di lanciare due dadi e si consideri la somma dei punti dei due dadi. Sia A l'evento "si osserva un numero pari" e B l'evento "si osserva un numero maggiore di 7". Cos'è $\overline{A} \cup \overline{B}$?

   (a) $\{2, 3, 4, 5, 6, 7\}$

   (b) $\{2, 4, 6\}$

   (c) $\{2, 3, 4, 5, 6, 7, 9, 11\}$

   (d) $\{3, 5, 7\}$

   Soluzione. $A = \{2, 4, 6, 8, 10, 12\}$ e quindi $\overline{A} = \{3, 5, 7, 9, 11\}$. Nota che non può avvenire che la somma di due dadi sia 1!

   $B = \{8, 9, 10, 11, 12\}$ quindi $\overline{B} = \{2, 3, 4, 5, 6, 7\}$. Conclusione: $\overline{A} \cup \overline{B} = \{2, 3, 4, 5, 6, 7, 9, 11\}$.

3. Una tavola con due righe, A_1 e A_2, e due colonne, B_1 and B_2, riporta le seguenti probabilità congiunte:

   $$P(A_1 \cap B_1) = 0.10,\ P(A_1 \cap B_2) = 0.30,\ P(A_2 \cap B_1) = 0.05,\ P(A_2 \cap B_2) = 0.55.$$

   Quanto vale $P(B_1)$?

   (a) 0.15

(b) 0.60
(c) 0.40
(d) 0.85

Soluzione. La tavola è

	B_1	B_2
A_1	0.10	1.30
A_2	0.05	0.55

La $P(B_1)$ è il totale della prima colonna cioè $P(B_1) = P(A_1 \cap B_1) + P(A_2 \cap B_1) = 0.10 + 0.05 = 0.15$.

4. In un'indagine recente, i rispondenti sono stati classificati rispetto al sesso, lo stato civile e l'area geografica di residenza. I dati sono sintetizzati nella seguente tabella (nella quale M è usato per indicare il sesso maschile e F per indicare il sesso femminile):

	Single		Sposato		
Area	M	F	M	F	Totale
Nord-est	12	17	22	10	61
Nord-ovest	31	26	7	23	87
Centro	45	33	52	38	168
Sud	34	19	18	13	84
Totale	122	95	99	84	400

Qual è la proporzione di rispondenti non sposati?

(a) 0.543
(b) 0.510
(c) 0.620
(d) 0.305

Soluzione. Interpreta la proporzione come una probabilità. Casi possibili: 400 rispondenti. Casi favorevoli: single M o single F = 122 + 95 = 217. Quindi la proporzione è 217/400 = 0.5425 che arrotondato a 3 cifre è 0.543.

5. La probabilità dell'intersezione di due eventi non può essere superiore alla somma delle loro probabilità. Vero o Falso?

Soluzione. $P(A \cap B) = P(A) + P(B) - P(A \cup B)$ quindi $P(A \cap B) \leq P(A) + P(B)$. Vero.

6. Dati due eventi A e B, se almeno uno di loro è necessariamente vero si dice che A e B sono collettivamente esaustivi. Vero o Falso?

Soluzione. Due eventi sono esaustivi se la loro unione è l'intero spazio campionario. Ossia se l'evento "si verifica A oppure si verifica B" è l'evento certo. Quindi è Vero.

7. In una recente indagine sulla fiducia dei consumatori, 160 rispondenti sono stati classificati in base al loro livello di fiducia e al loro titolo di studio:

```
                  Titolo di studio
Fiducia      Diploma    Laurea       Master
Bassa          13         17           15
Media          27         22           13
Alta           32         14            7
```

Supponiamo di estrarre a caso un consumatore: gli eventi “ha un master” e “ha un alto livello di fiducia” sono statisticamente indipendenti?

(a) No.
(b) Forse.
(c) Sì.
(d) Non ci sono sufficienti informazioni per rispondere.

Soluzione. Sono indipendenti se

$$P(Master \cap Alta) = P(Master)P(Alta)$$

Risulta che i casi possibili sono 160, mentre

$$P(Master \cap Alta) = 7/160,\ P(Master) = \frac{15+13+7}{160} = 35/160,\ P(Alta) = \frac{32+14+7}{160} = 53/160$$

Quindi poiché $7/160 \neq (35/160)(53/160)$ i due eventi non sono indipendenti.

8. La regola empirica si applica a qualsiasi distribuzione, indipendentemente dalla sua forma, come guida per interpretare la distribuzione. Vero o Falso?

   Rileggi la definizione: la regola empirica si applica se la popolazione è simile a una variabile normale (cioè Gaussiana). Falso.

9. Se in una certa prova si ha che $P(A) = 0.7$ può accadere che $P(A \cup B) = 0.5$?

   (a) Solo se A e B sono eventi incompatibili
   (b) Solo se $B = \emptyset$
   (c) SI è possibile
   (d) NO mai

   Soluzione.

   (a) Falsa: per esempio perché se B è disgiunto da A, $P(A \cup B) = P(A) + P(B) \geq 0.7$.
   (b) Falsa: se $B = \emptyset$, $P(A \cup B) = P(A) = 0.7$
   (c) Falsa: vedi D).
   (d) Vera: L’evento $A \cup B$ contiene sempre A e quindi si verifica sempre se si verifica A. Quindi $P(A \cup B) \geq P(A)$ e quindi non può mai succedere che $P(A \cup B) = 0.5 < P(A)$.

10. Se X e Y sono due variabili casuali independenti, quale delle seguenti uguaglianze è vera?

   (a) $E(2X + 3Y) = 5\, E(X + Y)$
   (b) $\mathrm{var}(2X + 3Y) = 2\mathrm{var}(X) + 3\mathrm{var}(Y)$
   (c) $\mathrm{var}(2X + 3Y) = 4\mathrm{var}(X) + 9\mathrm{var}(Y)$
   (d) $E(2X + 3Y) = E(X) + E(Y) + 5$.

Soluzione. A) Falsa: $E(2X+3Y) = 2E(X)+3E(Y)$.

B) Falsa: $\text{var}(2X+3Y) = 4\text{var}(X)+9\text{var}(Y)$

C) Vera.

D) Falsa.

11. (Richiede la conoscenza dell'ipergeometrica). In una partita di 18 autocarri, ce ne sono 4 senza aria condizionata. Se ne estraggono 4 a caso. Qual è la probabilità al più due siano senza aria condizionata?

    Soluzione. L'estrazione è senza ripetizione perché il testo dice la numerosità della popolazione da cui si estrae e il campione è piccolo. Qui (X =) numero di autocarri senza aria condizionata sul 4 estratti. Pertanto, X ha distribuzione ipergeometrica. Quindi:

    $$P(X=3) = \frac{\binom{S}{3}\binom{N-S}{n-3}}{\binom{N}{n}}$$

    e anche

    $$P(X=4) = \frac{\binom{S}{4}\binom{N-S}{n-4}}{\binom{N}{n}}$$

    La probabilità chiesta è $P(X \leq 2) = 1 - P(X=3) - P(X=4)$. Quindi si calcola

    $$\begin{aligned} P(X=3) &= \frac{\binom{4}{3}\binom{14}{1}}{\binom{18}{4}} \\ &= \frac{4 \times 14}{3060} = 0.0183 \end{aligned}$$

    $$\begin{aligned} P(X=4) &= \frac{\binom{4}{4}\binom{14}{0}}{\binom{18}{4}} \\ &= \frac{1 \times 1}{3060} = 0.000326797. \end{aligned}$$

    Quindi $P(X \leq 2) = 1 - 0.0183 - 0.0004 = 0.9813$.

12. Se X e Y sono due variabili casuali indipendenti, quale delle seguenti uguaglianze è falsa? A) $\text{cov}(X,Y) = 0$ B) $\text{var}(X-Y) = \text{var}(X)+\text{var}(Y)$ C) $E(X+Y) = E(X)+E(Y)$ D) $\text{cov}(X,Y) = 1$.

    Soluzione. A) Vera: se X e Y sono indipendenti sono anche incorrelate. B) Vera: perché $\text{var}(X-Y) = \text{var}(X)+\text{var}(Y) - 2\text{cov}(X,Y)$ C) Vera. D) Falsa.

13. Un' indagine ha rilevato che il 40% dei controllori di volo ritiene il proprio lavoro molto stressante. Supponi che 12 controllori di volo siano selezionati casualmente. Qual è la probabilità che almeno 2 di loro ritengano il proprio lavoro molto stressante?

    (a) 0.7218

    (b) 0.2806

    (c) 0.9804

    (d) 0.0282

Soluzione. Qui si suppone che selezionare casualmente 12 controllori da una popolazione (di numerosità ignota) sia equivalente a fare 12 prove di Bernoulli indipendenti con probabilità 0.4 di selezionare uno stressato. Il numero X di controllori di volo stressati su 12 è quindi binomiale.

$$X = \# \text{ stressati su } 12 \sim Binomiale(12, p = 0.4)$$

Quindi la probabilità che almeno 2 due di loro siano stressati su 12 è

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)$$

Quindi si usa la formula della probabilità Binomiale:

$$1 - \binom{12}{0}(0.4)^0(0.6)^{12} - \binom{12}{1}(0.4)^1(0.6)^{11} = 1 - 0.6^{12} - 12(0.4)(0.6^11) = 0.9804.$$

14. Data la distribuzione di probabilità

x	0	1	2	3	4	5	6	7
$p(x)$	0.05	0.16	0.19	0.24	0.18	0.11	0.03	0.04

Quali delle seguenti affermazioni è vera?

(a) $P(X \geq 3) = 0.64$
(b) $P(2 < X < 5) = 0.42$
(c) $P(X > 6) = 0.07$
(d) $P(X \leq 6) = 0.93$.

Soluzione.

(a) $P(X \geq 3) = 1 - (0.05 + 0.16 + 0.19) = 0.6$. Falso.
(b) $P(2 < X < 5) = p(3) + p(4) = 0.24 + 0.18 = 0.42$. Vero.
(c) $P(X > 6) = 0.04$. Falso.
(d) $P(X \leq 6) = 1 - p(7) = 1 - 0.04 = 0.96$. Falso.

15. Considera la seguente distribuzione di probabilità della variabile casuale X.

x	0	1	2	3	4	5	6
$p(x)$	0.07	0.19	0.23	0.17	0.16	0.14	0.04

Qual è il valore atteso di X?

(a) 0.46
(b) 1.78
(c) 2.74
(d) 3.02

Soluzione. Abbiamo

$$E(X) = (0)(0.07) + (1)(0.19) + (2)(0.23) + (3)(0.17) + (4)(0.16) + (5)(0.14) + (6)(0.04) = 2.74.$$

16. Se X e Y sono due variabili casuali con $\text{cov}(X,Y) = 0.25$, $\sigma_X^2 = 0.36$, e $\sigma_Y^2 = 0.49$, allora il il coefficiente di correlazione è A) 0.595 B) 0.354 C) 1.417 D) 1.190.

    Soluzione. Abbiamo
$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$
    quindi
$$\rho = \frac{0.25}{\sqrt{0.36 \cdot 0.49}} = 0.595.$$
    Velo pietoso sui coefficienti di correlazione più grandi di 1!

17. Quale dei seguenti è un esempio di variabile casuale discreta?

    (a) L'ammontantare di pioggia che cade in un intervallo temporale di 24 ore.
    (b) Il peso di un pacco all'ufficio postale.
    (c) La distanza che puoi percorrere con un pieno di benzina.
    (d) Il numero di vacche in una fattoria.

    Soluzione.

    (a) È una misura (in mm)
    (b) È una misura (in kg)
    (c) È una misura (in km)
    (d) È un conteggio e dunque è una variabile discreta.

18. Sia $X \sim N(17.1, \sigma = 3.2)$. Calcolare $P(15 < X < 20)$.

    (a) 0.5640
    (b) 0.5581
    (c) 0.2546
    (d) 0.1814

    Soluzione. Si ha
$$P(15 < X < 20) = P(X < 20) - P(X < 15)$$
    Standardizzando, si ha
$$P(X < 20) = P(Z < (20 - 17.1)/3.2) = P(Z < 0.91)$$
    dove ho arrotondato a 2 cifre. Analogamente
$$P(X < 15) = P(Z < (15 - 17.1)/3.2) = P(Z < -0.66) = 1 - P(0.66)$$
    Ora dalle tavole si trova che
$$P(Z < 0.91) = 0.8186, \quad P(Z < 0.66) = 0.7454$$
    Quindi la probabilità è $0.8186 - (1 - 0.7454) = 0.5640$.

19. Sia $Z \sim N(0,1)$ e sia c un possibile valore di Z. Trovare c tale che l'area totale tra $-c$ e $+c$ sia 0.9544.

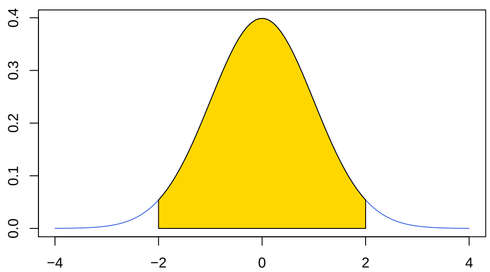
(a) 0.11

(b) 2.50

(c) 2.00

(d) 0.06

Soluzione. Trova c tale che $P(-c < Z < c) = 0.9544$.



Nota che per rispettare la definizione bisogna che

$$P(Z > c) = (1 - 0.9544)/2 = 0.0228$$

Allora $P(Z < c) = 1 - P(Z > c) = 1 - 0.0228 = 0.9772$. Vai sulle tavole della normale e vedi che $c = 2$. Quindi la risposta è C).

20. La probabilità che una persona prenda il raffreddore durante l'inverno è 0.4. Assumendo che 10 persone siano selezionate a caso, qual è la probabilità che esattamente 4 di loro prendano il raffreddore?

    (a) 0.502

    (b) 0.242

    (c) 0.751

    (d) 0.251

    Soluzione. Chiamo p = probabilità che una persona prenda il raffreddore in inverno = 0.4. Se si selezionano "a caso" 10 persone (da una popolazione infinita) si stanno facendo 10 prove di Bernoulli indipendenti con probabilità di successo p. Il numero di persone X con il raffreddore si intende hanno distribuzione X *Binomiale*$(10, p = 0.4)$. Quindi

    $$P(X = 4) = \binom{10}{4} p^4 (1-p)^6 = 210(1-p)^6 p^4 = 0.251.$$

    Risposta D.

21. Se X e Y sono due variabili casuali con $E(X) = 5$, $E(Y) = 6$, $E(XY) = 21$, $\mathrm{var}(X) = 9$ e $\mathrm{var}(Y) = 10$, allora l'associazione lineare tra X e Y è:

    (a) debole e positiva.

    (b) debole e negativa.

    (c) forte e negativa.

    (d) forte e positiva.

Soluzione. L'associazione lineare si misura con il coefficiente di correlazione lineare.

$$\text{cor}(X,Y) = \text{cov}(X,Y)/\sqrt{\text{var}(X)\text{var}(Y)}.$$

dove $\text{cov}(X,Y) = E(XY) - E(X)E(Y) = 21 - (5)(6) = -9$. Quindi

$$\text{cor}(X,Y) = -9/\sqrt{(9)(10)} \simeq -0.95.$$

L'associazione è negativa e forte. Risposta C).

22. Ci siamo recentemente iscritto ad un golf club. Il numero di volte che si presume di giocare a golf in un mese è una variabile casuale con media 10 e deviazione standard 2.2. Si assume di pagare una quota sociale di 500 euro al mese e di pagare una quota addizionale di 50 euro per ogni partita di golf giocata. Qual è la deviazione standard della quota media mensile da pagare al Club?

    (a) 110

    (b) 324

    (c) 180

    (d) 230

    Soluzione. Il numero di volte che si gioca è aleatorio e lo chiamiamo X. Se in un mese pago 500 euro più 50 a partita al mese pago $500 + 50X$ euro. In media perciò pago

    $$500 + 50E(X) = 500 + (50)(10) = 1000 \text{ euro.}$$

    La varianza è

    $$\text{var}(X) = \text{var}(500 + 50X) = 2500\,\text{var}(X) = 2500(2.2^2).$$

    La deviazione standard è $\sigma(X) = (50)(2.2) = 110$ euro.

23. Se X e Y sono due variabili casuali qualsiasi, quali delle seguenti ugualianze non è sempre vera?

    (a) $E(X+Y) = E(X) + E(Y)$

    (b) $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

    (c) $\text{var}(4X+5Y) = 16\text{var}(X) + 25\text{var}(Y) + 40\text{cov}(X,Y)$

    (d) $E(4X+5Y) = 4E(X) + 5E(Y)$

    Soluzione. Chiede quale è Falsa. A) Vera. B) Falsa: manca il termine $+2\text{cov}(X,Y)$. C) Vera. D) Vera.

    Risposta B).

24. Si assuma che il tempo di attesa per sedersi ad un ristorante sia approssimabile con una variabile casuale normale con media 15 minuti e scostamento quadratico medio di 4.75 minuti. Calcolare la probabilità di dover aspettare almeno 20 minuti prima di sedersi.

    (a) 0.1469

    (b) 0.3531

    (c) 0.6761

(d) 0.1761.

Soluzione. Tempo $= X \sim N(15, \sigma = 4.75)$. La probabilità di dover aspettare almeno 20 minuti prima di sedersi è

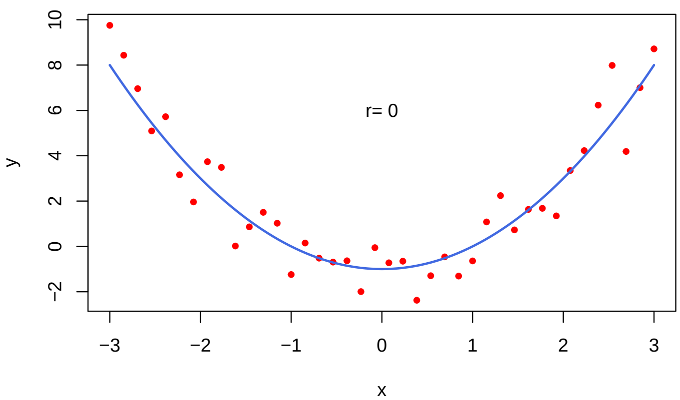
$$P(X \geq 20) = P[Z > (20 - 15)/4.75] = P(Z > 1.05) = 1 - P(Z \leq 1.05) = 1 - 0.8531 = 0.1469.$$

25. Data la variabile casuale $Y = a + bX$, segue che $\mu_Y = b\mu_X$. Vero o Falso?

    Soluzione. Falso. $E(Y) = E(a + bX) = a + b\mu_X$.

26. Se il coefficiente di correlazione lineare è nullo, allora non c'è nessun tipo di relazione tra le due variabili. Vero o Falso?

    Soluzione. Falso. Ci può essere una relazione non lineare. Vedi Figura.



27. Il lancio di una moneta rappresenta un esperimento di binomiale solo se la moneta è bilanciata, cioè se $p = 0.5$.

    Soluzione. Falso ovviamente.

28. Se X eY sono due variabili casuali con $\sigma_X^2 = 3.25$, $\sigma_Y^2 = 5.8$, e $\text{cov}(X,Y) = 14.703$, allora il coefficiente di correlazione lineare è $\rho = 0.78$.

    Soluzione. Questo esercizio ha un problema. Infatti non esistono due variabili con quelle varianze e quella covarianza. Questo si vede subito perché il coefficiente di correlazione è

    $$\rho = \frac{14.703}{\sqrt{3.25 \cdot 5.8}} = 3.39!!$$

    Quindi la risposta è FALSO.

    NOTA: il coefficiente di correlazione tornerebbe se fosse $\sigma_X = 3.25$, $\sigma_Y = 5.8$.

29. Se X e Y sono due variabili casuali correlate, allora $E(X + Y) = E(X) + E(Y) + \text{cov}(X,Y)$. Vero o Falso?

    Soluzione. Falso è $E(X + Y) = E(X) + E(Y)$.

30. Se Z è una variabile casuale normale standardizzata, un valore negativo di Z indica che la deviazione standard di Z è negativa. Vero o Falso?

    Soluzione. Falso. La deviazione standard di Z è 1.

31. La media e la mediana sono le stesse per una variabile casuale uniforme. Vero o Falso?

    Soluzione. Vero: è una variabile simmetrica.

32. La forma della distribuzione normale è determinata dalla deviazione standard σ. Una diminuzione di σ riduce l'altezza della curva e appiattisce la curva lungo l'asse delle ascisse. Un incremento di σ aumenta l'altezza della curva e riduce l'appiattimento della curva. Vero o Falso?

    Soluzione. Falso: è esattamente l'opposto.

33. Se due variabili casuali X e Y sono indipendenti, allora

    $$P(y \mid x) = P(x), \text{ e } P(x \mid y) = P(y) \text{ per ogni } x, y.$$

    Vero o Falso?

    Soluzione. Falso: la definizione è

    $$P(y \mid x) = P(y), \text{ e } P(x \mid y) = P(x) \text{ per ogni } x, y.$$

34. Se X è una variabile casuale binomiale con $n = 5$ e $p = 0.2$, allora il valore atteso è 1. Vero o Falso?

    Soluzione. Vero: $E(X) = np = (5)(0.2) = 1$.

35. Dato un campione casuale di dimensione $n = 4900$ da una popolazione Bernoulliana con $p = 0.5$ sia X il numero di successi

    (a) la $P(X > 2480)$
    (b) la $P(X < 2390)$
    (c) la $P(2445 < X < 2520)$
    (d) Il numero di successi che ha probabiità 0.1 di NON essere superato.
    (e) Il numero di successi che ha probabilità 0.09 di ESSERE superato.

    Soluzione.

    Si usa l'approsimazione normale $X \approx N(\mu = np, \sigma^2 = npq) = N(\mu = 2450, \sigma = 35)$. Quindi

    (a) $P(X > 2480) = P(Z > (2480-2450)/35) = P(Z > 0.86) = 1-P(Z \leq 0.86) = 1-0.8051 = 0.1949$
    (b) $P(X < 2390) = P(Z < (2390 - 2450)/35) = P(Z < -1.71) = 1 - P(Z < 1.71) = 1 - 0.9564 = 0.0436$
    (c) $P(2445 < X < 2520) = P((2445 - 2450)/35 < Z < (2520 - 2450)/35) = P(-0.14 < Z < 2) = P(Z < 2)-P(Z < -0.14) = 0.9772-(1-P(Z < 0.14)) = 0.9772-(1-0.5557) = 0.5329$
    (d) Trova k tale $P(X < k) = 0.1$ ossia $P(Z < (k - 2450)/35) = 0.1$. Si cerca sulla Tavola 2, ultima riga il quantile che lascia A DESTRA 0.1 e poi gli si cambia segno. Fare il grafico! Si trova sulla Tavola 2, ultima riga si vede che $z_{0.1} = 1.282$. Quindi $(k-2450)/35 = -1.282$ e quindi

        $$k = 2450 - 1.282 \cdot 35 = 2405.13.$$

    (e) Trova k tale che $P(X > k) = 0.09$ ossia $P(X \leq k) = 0.91$. Standardizzo e ottengo

        $$P(Z \leq (k - 2450)/35)) = 0.91$$

        Il quantile 0.91 stra tra 1.34 e 1.35 quindi interpolando è 1.355. Allora

        $$(k - 2450)/35) = 1.355, \text{ cioè } k = 2450 + 1.355 \cdot 35 = 2497.43.$$