

Esercizi di Statistica della 1^a settimana (Corso di Laurea Triennale in Biologia, Università degli Studi di Padova).

Esercizio 1. Uno scienziato che si occupa di inquinamento ambientale vuole verificare se una sorgente è inquinata. A questo scopo effettua 10 misurazioni indipendenti, con i seguenti risultati del pH:

6.24 6.31 6.28 6.30 6.25 6.26 6.24 6.29 6.22 6.28

1. Calcola media e mediana campionarie.
2. Calcola la varianza e la deviazione standard campionarie.
3. Supponendo che la deviazione standard vera sia uguale a quella campionaria, qual è l'errore standard della media?
4. Calcola il 10° e il 90° percentile.

Esercizio 2. I seguenti dati rappresentano l'assunzione giornaliera di vitamina A da parte di 14 studenti:

31 21 74 95 19 64 108 48 24 13 37 21 78 17

Calcolare:

1. media e mediana di questi dati;
2. la deviazione standard;
3. il primo e il terzo quartile;
4. l'errore standard della media.

Esercizio 3. Una scatola contiene una biglia rossa, una verde e una blu.

1. Descrivi lo spazio Ω degli esiti dell'esperimento che consiste nell'estrarre una biglia, rimetterla nella scatola ed estrarre una seconda volta.
2. Ripeti l'esercizio senza la rimessa della prima biglia.
3. Quanti elementi ha Ω nel primo caso? E nel secondo?
4. Supponiamo di avere 4 biglie di diverso colore invece che tre. Quanti elementi ha Ω sia nel caso che ci sia rimessa sia nel caso che non ci sia?
5. Rispondere alla domanda precedente nel caso le biglie non siano 4 ma n .

Esercizio 4. Siano $\Omega := \{1, 2, 3, 4, 5, 6, 7, 8\}$ con \mathbb{P} uniforme, e definiamo gli eventi $E := \{2, 4, 6, 8\}$, $F := \{1, 2, 3\}$, $G := \{1, 5\}$. Scrivi gli elementi dei seguenti eventi:

$$E \cap F, \quad E \cap G^c, \quad E^c \cap (F \cup G), \quad E \cup (F \cap G)$$

Esercizio 5. Un gruppo di 5 bambini e 10 bambine è in fila in ordine casuale, nel senso che tutti i possibili ordinamenti (o **permutazioni**) si suppongono equiprobabili.

1. Quante sono le possibili permutazioni?
2. Qual è la probabilità che il quarto della fila sia un bambino?
3. E il dodicesimo?
4. Qual è la probabilità che un dato bambino occupi la terza posizione?

Esercizio 6. Si supponga che il 5% degli uomini e lo 0.25% delle donne siano daltonici. Supponiamo all'inizio che vi sia lo stesso numero di uomini e di donne.

1. Qual è la percentuale di daltonici (uomini e donne) nella popolazione totale?
2. Se si sceglie a caso un daltonico, qual è la probabilità che questo sia un uomo?

Supponiamo ora che gli uomini siano il doppio delle donne.

3. Qual è la percentuale di daltonici nella popolazione totale?
4. Se si sceglie a caso un daltonico, qual è la probabilità che stavolta sia una donna?

Esercizio 7. Siano A, B due eventi. Sapendo che $P(A|B) = 0.7$, $P(A|B^c) = 0.3$ e $P(B|A) = 0.6$, calcolare $P(A)$.

Esercizio 8. È stata studiata la relazione tra forma fisica e malattie cardiovascolari in un gruppo di impiegati delle ferrovie di sesso maschile, con questi risultati:

frequenza cardiaca sotto esercizio (battiti/minuto)	percentuale dei lavoratori	mortalità per malattie cardiovascolari in 20 anni
≤ 105	20%	9,1%
106–115	30%	8,7%
116–127	30%	11,6%
> 127	20%	13,2%

Supponiamo che un certo test sia positivo se la frequenza cardiaca sotto esercizio è maggiore di 127 battiti al minuto e negativo altrimenti.

1. Qual è la percentuale totale di mortalità per malattie cardiovascolari in 20 anni?
2. Qual è la probabilità di avere avuto un test positivo tra gli uomini che sono morti nel periodo di 20 anni?
3. Qual è la probabilità di avere avuto un test positivo tra gli uomini che sono sopravvissuti nel periodo di 20 anni?
4. Qual è la probabilità di morte tra gli uomini con un test negativo?

Soluzioni su <http://www.math.unipd.it/~vargiolu/Statistica/>

Soluzioni

Esercizio 1. Per trovare la mediana (cioè il 50esimo percentile) e il 10mo e il 90esimo percentile, riordiniamo i dati:

6.22 6.24 6.24 6.25 6.26 6.28 6.28 6.29 6.30 6.31

1. La media campionaria è

$$\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i = \frac{62.67}{10} = 6.267$$

La mediana è definita come il numero x tale che $\mathbb{P}\{X \leq x\} = 0.5$. Siccome tutte le informazioni che abbiamo sulla distribuzione sono quelle del campione, dobbiamo utilizzare quelle. Dal campione notiamo che qualsiasi numero compreso tra 6.26 e 6.28 soddisfa la definizione di mediana; se dobbiamo sceglierne uno, un buon candidato può essere $\frac{6.26+6.28}{2} = 6.27$.

Volendo utilizzare la regola di Ross (2008), pag. 86, bisogna calcolare $np = 10 \cdot 0.5 = 5$, che è un intero, per cui bisogna prendere la media tra il 5° e il 6° valore, ossia $\frac{6.26+6.28}{2} = 6.27$.

2. La varianza è

$$s_X^2 = \frac{1}{10-1} \sum_{i=1}^{10} (X_i - \bar{X})^2 = 0.000841$$

e la deviazione standard campionaria è quindi $s_X = \sqrt{s_X^2} = \sqrt{0.000841} = 0.029$.

3. L'errore standard della media è dato da

$$s_{\bar{X}} = \frac{s_x}{\sqrt{10}} = 0.00917$$

4. Analogamente a quanto fatto per la mediana, possiamo prendere come 10mo e 90esimo percentile qualsiasi numero compreso rispettivamente tra 6.22 e 6.24 e tra 6.30 e 6.31 (quindi ad esempio $\frac{6.22+6.24}{2} = 6.23$ per il 10° percentile e $\frac{6.30+6.31}{2} = 6.305$ per il 90° percentile).

Volendo utilizzare la regola di Ross (2008), pag. 86, bisogna calcolare $np = 10 \cdot 0.1 = 1$ e $np = 10 \cdot 0.9 = 9$, che sono interi, per cui bisogna prendere la media tra il 1° e il 2° valore, ossia $\frac{6.22+6.24}{2} = 6.23$, per il 10° percentile, e la media tra il 9° e il 10° valore, ossia $\frac{6.30+6.31}{2} = 6.305$ per il 90° percentile.

Esercizio 2.

1. Consideriamo i dati come la realizzazione di un campione X_1, \dots, X_{14} di taglia 14 estratto da una popolazione più grande. Per trovare la media, basta applicare la formula:

$$\bar{X} = \frac{1}{14} \sum_{i=1}^{14} X_i = 46,43$$

Per trovare la mediana (cioè il 50esimo percentile), riordiniamo i dati:

13 17 19 21 21 24 31 37 48 64 74 78 95 108

La mediana è definita come il numero x tale che $\mathbb{P}\{X \leq x\} = 0,5$. Siccome tutte le informazioni che abbiamo sulla distribuzione sono quelle del campione, dobbiamo utilizzare quelle. Dal campione notiamo che qualsiasi numero compreso tra 31 e 37 soddisfa la definizione di mediana; se dobbiamo sceglierne uno, un buon candidato può essere $\frac{37-31}{2} = 34$.

Volendo utilizzare la regola di Ross (2008), pag. 86, bisogna calcolare $np = 14 \cdot 0,5 = 7$, che è un intero, per cui bisogna prendere la media tra il 7° e l'8° valore, ossia $\frac{37-31}{2} = 34$.

2. Abbiamo:

$$s_X^2 = \frac{1}{12-1} \sum_{i=1}^{12} (X_i - \bar{X})^2 = 1008,81$$

e quindi $s_X = \sqrt{s_X^2} = \sqrt{1008,81} = 31,74$.

3. Per calcolare il primo e il terzo quartile, notiamo che questi dovrebbero essere numeri tali da essere maggiori o uguali di $1/4$ (risp. $3/4$) dei dati e minori o uguali di $3/4$ (risp. $1/4$) dei dati. Dato che $14/4 = 3,5$, ciò è possibile solo scegliendo come primo quartile il quarto dato in ordine crescente (quindi 21), e come terzo quartile l'undicesimo dato in ordine crescente (quindi 74).

Volendo utilizzare la regola di Ross (2008), pag. 86, bisogna calcolare $np = 14 \cdot 0,25 = 3,5$ e $np = 14 \cdot 0,75 = 10,5$, che non sono interi; bisogna allora determinare i più piccoli interi maggiori di questi due indici, ossia 4 e 11, e in corrispondenza di quelli ci saranno i percentili cercati (quindi 21 per il primo quartile e 74 per il terzo).

4. L'errore standard della media è

$$s_{\bar{X}} = \frac{s_X}{\sqrt{14}} = 8,48$$

Esercizio 3.

1. Possiamo rappresentare l'estrazione di una biglia rossa, verde, blu, rispettivamente con le lettere R , V e B . Allora abbiamo

$$\begin{aligned}\Omega &= \{(i, j) \mid i, j = R, V, B\} = \{R, V, B\}^2 = \\ &= \{(R, R), (R, V), (R, B), (V, R), (V, V), (V, B), (B, R), (B, V), (B, B)\}\end{aligned}$$

2. Con le stesse convenzioni del punto 1., abbiamo

$$\begin{aligned}\Omega &= \{(i, j) \mid i, j = R, V, B, i \neq j\} = \{R, V, B\}^2 \setminus \{(R, R), (V, V), (B, B)\} = \\ &= \{(R, V), (R, B), (V, R), (V, B), (B, R), (B, V)\}\end{aligned}$$

3. Nel primo caso $|\Omega| = 9$, nel secondo $|\Omega| = 6$.

4. Supponiamo di aggiungere una biglia gialla (G). Allora nel caso di estrazione con rimessa abbiamo $\Omega = \{R, V, B, G\}^2$, e quindi $|\Omega| = 4^2 = 16$. Nel caso di estrazione senza rimessa, invece, abbiamo $\Omega = \{R, V, B, G\}^2 \setminus \{(R, R), (V, V), (B, B), (G, G)\}$, e $|\Omega| = 4^2 - 4 = 12$.

5. Nel caso di estrazione con rimessa abbiamo $|\Omega| = n^2$, mentre nel caso di estrazione senza rimessa abbiamo $|\Omega| = n^2 - n = n(n - 1)$.

Esercizio 4. Abbiamo:

$$\begin{aligned}E \cap F &= \{2\}, \\ E \cap G^c &= \{2, 4, 6, 8\} \quad (= E \setminus G), \\ E^c \cap (F \cup G) &= \{1, 3, 5, 7\} \cap \{1, 2, 3, 5\} = \{1, 3, 5\}, \\ E \cup (F \cap G) &= E \cup \{1\} = \{1, 2, 4, 6, 8\}\end{aligned}$$

Esercizio 5. Se assegnamo ai 5 bambini i numeri $\{1, \dots, 5\}$ e alle 10 bambine i numeri $\{6, \dots, 10\}$, allora tutti i possibili ordinamenti di bambini e bambine sono costituiti da $\Omega =$ tutti i possibili ordinamenti dei numeri $\{1, \dots, 15\}$.

1. Abbiamo $|\Omega| = 15!$ ($\simeq 1.307 \cdot 10^{12}$).
2. Definiamo l'evento $A =$ "il quarto della fila è un bambino". Allora A è costituito dagli ordinamenti in cui nel quarto posto ci sia un bambino (5 scelte possibili) e negli altri posti ci siano gli altri 4 bambini in totale su 14, in un qualsiasi ordine (possibili scelte: $14!$). La probabilità viene quindi ad essere

$$\mathbb{P}(A) = \frac{5 \cdot 14!}{15!} = \frac{5}{15} = \frac{1}{3} = 0.3333$$

3. Definiamo l'evento $B =$ "il dodicesimo della fila è un bambino". Allora B è costituito dagli ordinamenti in cui nel dodicesimo posto ci sia un bambino (5 scelte possibili) e negli altri posti ci siano gli altri 4 bambini in totale su 14, in un qualsiasi ordine (possibili scelte: $14!$). La probabilità viene quindi ad essere ancora

$$\mathbb{P}(B) = \frac{5 \cdot 14!}{15!} = \frac{5}{15} = \frac{1}{3} = 0.3333$$

4. Definiamo l'evento $C =$ "il terzo della fila è un certo bambino". Allora C è costituito dagli ordinamenti in cui nel terzo posto ci sia quel certo bambino (1 scelta possibile) e negli altri posti ci siano gli altri 4 bambini in totale su 14, in un qualsiasi ordine (possibili scelte: $14!$). La probabilità viene quindi ad essere

$$\mathbb{P}(C) = \frac{1 \cdot 14!}{15!} = \frac{1}{15} = 0.0666$$

Esercizio 6. Definiamo gli eventi

$$D := \{\text{daltonico}\}, \quad M := \{\text{uomo}\}, \quad F := \{\text{donna}\},$$

dove ovviamente abbiamo $M \cup F = \Omega$, $M \cap F = \emptyset$. Allora i dati si possono riscrivere così:

$$\mathbb{P}(D|M) = 0.05, \quad \mathbb{P}(D|F) = 0.0025.$$

1. Abbiamo come ipotesi che $\mathbb{P}(M) = \mathbb{P}(F) = 0.5$, quindi entrambi gli eventi sono non trascurabili. Allora usando la formula della probabilità totale calcoliamo

$$\mathbb{P}(D) = \mathbb{P}(D|M)\mathbb{P}(M) + \mathbb{P}(D|F)\mathbb{P}(F) = 0.05 \cdot 0.5 + 0.0025 \cdot 0.5 = 0.02625$$

2. Usando la formula di Bayes, la probabilità richiesta è:

$$\mathbb{P}(M|D) = \frac{\mathbb{P}(D|M)\mathbb{P}(M)}{\mathbb{P}(D)} = \frac{0.05 \cdot 0.5}{0.02625} = 0.9523$$

3. Stavolta abbiamo come ipotesi che $\mathbb{P}(M) = \frac{2}{3}$, $\mathbb{P}(F) = \frac{1}{3}$, e anche stavolta i due eventi sono non trascurabili. Allora usando la formula della probabilità totale calcoliamo

$$\mathbb{P}(D) = \mathbb{P}(D|M)\mathbb{P}(M) + \mathbb{P}(D|F)\mathbb{P}(F) = 0.05 \cdot 0.666 + 0.0025 \cdot 0.333 = 0.03416$$

4. Usando la formula di Bayes, la probabilità richiesta stavolta è:

$$\mathbb{P}(F|D) = \frac{\mathbb{P}(D|F)\mathbb{P}(F)}{\mathbb{P}(D)} = \frac{0.0025 \cdot 0.333}{0.03416} = 0.02439$$

Esercizio 7. L'esercizio si può fare in più modi. Uno di questi è il seguente: siccome $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$, abbiamo

$$\frac{P(B)}{P(A)} = \frac{P(B|A)}{P(A|B)} = \frac{0.6}{0.7} = \frac{6}{7}$$

inoltre, siccome $P(A \cap B^c) = P(B^c|A)P(A) = P(A|B^c)P(B^c)$, abbiamo

$$\frac{P(B^c)}{P(A)} = \frac{P(B^c|A)}{P(A|B^c)} = \frac{1 - P(B|A)}{P(A|B^c)} = \frac{1 - 0.6}{0.3} = \frac{4}{3}$$

Si ricava quindi

$$\frac{1}{P(A)} = \frac{P(B) + P(B^c)}{P(A)} = \frac{6}{7} + \frac{4}{3} = \frac{46}{21}$$

e infine $P(A) = \frac{21}{46} = 0.4565$.

Esercizio 8. Per un generico impiegato, definiamo gli eventi

$$\begin{aligned} A_i &:= \{\text{l'impiegato appartiene alla classe } i\}, \quad i = 1, 2, 3, 4 \\ B &= \{\text{l'impiegato muore entro 20 anni per cause cardiovascolari}\} \end{aligned}$$

Allora la seconda colonna si interpreta come $\mathbb{P}(A_i)$, $i = 1, 2, 3, 4$, mentre la terza si interpreta come $\mathbb{P}(B|A_i)$, $i = 1, 2, 3, 4$.

1. Applicando la formula della probabilità totale, si ha

$$\mathbb{P}(B) = \sum_{i=1}^4 \mathbb{P}(B|A_i)\mathbb{P}(A_i) = 0.2 \times 0.091 + 0.3 \times 0.087 + 0.3 \times 0.116 + 0.2 \times 0.132 = 0.105$$

2. Usando la formula di Bayes si ha

$$\mathbb{P}(A_4|B) = \frac{\mathbb{P}(B|A_4)\mathbb{P}(A_4)}{\mathbb{P}(B)} = \frac{0.132 \times 0.2}{0.105} = 0.250$$

3. Usando ancora la formula di Bayes e ricordando che la probabilità condizionale nel primo argomento ha le stesse proprietà di una probabilità, si ha

$$\mathbb{P}(A_4|B^c) = \frac{\mathbb{P}(B^c|A_4)\mathbb{P}(A_4)}{\mathbb{P}(B^c)} = \frac{(1 - 0.132) \times 0.2}{1 - 0.105} = 0.194$$

4. Usando la formula di Bayes di nuovo, si ha:

$$\mathbb{P}(B|A_4^c) = \frac{\mathbb{P}(A_4^c|B)\mathbb{P}(B)}{\mathbb{P}(A_4^c)} = \frac{(1 - 0.250) \times 0.105}{1 - 0.2} = 0.0098$$