

Iniziato	venerdì, 11 luglio 2025, 10:13
Stato	Completato
Terminato	venerdì, 11 luglio 2025, 11:28
Tempo impiegato	1 ora 15 min.

Domanda **1**  
Parzialmente corretta  
Punteggio max.: 1,00

— Importante —

- Approssimate, se necessario, i risultati alla **quarta cifra decimale**.

— Esercizio —

Un gruppo di ricercatori marini ha raccolto dati di alcune località costiere per studiare la relazione tra la temperatura dell’acqua (in °C) e la densità di fitoplancton (in milioni di cellule per litro), un importante indicatore della salute dell’ecosistema marino. I dati rilevati sono i seguenti:

[fito.RData](#)

Fare una analisi descrittiva del dataset rispondendo alle domande seguenti.

1. Lo studio comprende

509

✓ osservazioni e sono presenti

1

✓ osservazioni mancanti.

2. La densità media di cellule di fitoplancton nelle località con temperatura inferiore (<) a 18.5 gradi centigradi è pari a

3,5491

✓ con deviazione standard pari a

0,6626

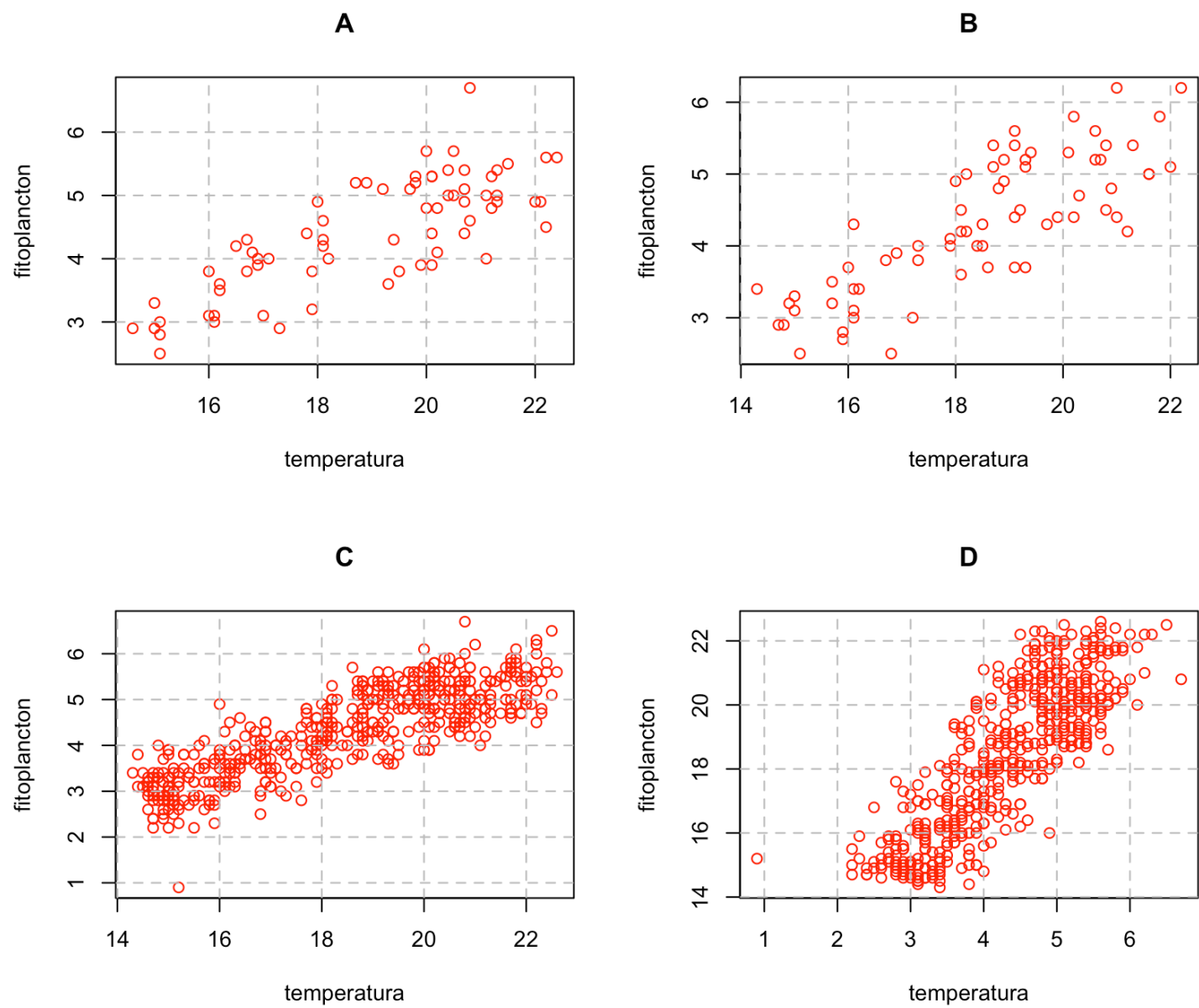
✓

3. Il 5% delle località con temperatura dell’acqua più bassa ha temperatura inferiore a

14,8

✓

4. Quale tra i grafici qui proposti è lo scatteplot delle variabili temperatura e fitoplancton?



Risposta:

- ☐ B
- ☐ A
- ☐ D
- ☒ C ✓

La risposta corretta è: C

Si vuole indagare sulla seguente domanda: la temperatura dell’acqua e la densità di cellule di fitoplancton sono legate da una relazione lineare?

5. Posso affermare che

- ☐ Lo scatterplot fa pensare che le due variabili non siano correlate
- ☐ Lo scatterplot fa pensare che le due variabili siano negativamente correlate
- ☒ Lo scatterplot fa pensare che le due variabili siano positivamente correlate ✓
- ☐ Nessuna affermazione è vera

La risposta corretta è: Lo scatterplot fa pensare che le due variabili siano positivamente correlate

6. La correlazione campionaria tra le due variabili è pari a

0,8306

- ✓ ed è significativamente (con livello di significatività 0.01)
- ☒ non nulla ✓
- ☐ nulla

La risposta corretta è: non nulla

Infatti, il p-value del test è pari a

0



7. Stimo la retta di regressione lineare. Il modello lineare corretto

- ☐ deve contenere solo il coefficiente angolare, l'intercetta deve essere posta uguale a 0
- ☒ deve contenere sia il coefficiente angolare sia l'intercetta ✓
- ☐ nessuna delle affermazioni è vera

La risposta corretta è: deve contenere sia il coefficiente angolare sia l'intercetta

8. Il modello ha  $R^2$  pari a

0,5



1. `nrow(fito)` e poi `sum(is.na(fito))`
2. `mean(fito$fitoplancton[fito$temperatura < 18.5], na.rm = T)` e `sd(fito$fitoplancton[fito$temperatura < 18.5], na.rm = T)`
3. `quantile(fito$temperatura,0.05,na.rm = T)`
4. `plot(fito$temperatura,fito$fitoplancton)`
5. Guardando lo scatterplot, le due variabili sembrano positivamente correlate
6. `cor.test(fito$temperatura, fito$fitoplancton) -> test` e poi `test$estimate` e `test$p.value`
7. `lm(fitoplancton ~ temperatura,data = fito)` servono sia intercetta che coefficiente angolare infatti i p-value dei test per la nullità dei parametri sono entrambi  $< 0.01$ .
8. `summary(lm(fitoplancton ~ temperatura,data = fito)) -> summodello` e `summodello$r.squared`

Domanda **2**  
Risposta corretta  
Punteggio max.: 1,00

— Importante —

Approssimate, se necessario, i risultati alla quarta cifra decimale.

— Esercizio —

L'urna  $U_1$  contiene una **proporzione** 0.3 di palline bianche e l'urna  $U_2$  una proporzione 0.5 di palline bianche. Si estraggono **con reimbussolamento** 4 palline da  $U_1$  e 6 da  $U_2$ . Tutte le palline estratte vengono sistemate in una terza urna  $U_3$ . Sia  $X$  la **proporzione** di palline bianche nell'urna  $U_3$ . Calcolare,

1. il valore atteso  $\mathbb{E}X$

0,42



2. la varianza  $\mathbb{V}\text{ar}X$

0,0234



**NB: può essere utile ricordare la distribuzione Binomiale**

Chiamiamo  $B_1$  e  $B_2$  rispettivamente il numero di palline bianche estratte dalle due urne  $U_1$  e  $U_2$ . Si ha che  $B_1 \sim \text{Bin}(4, 0.3)$  e  $B_2 \sim \text{Bin}(6, 0.5)$ . Inoltre  $B_1$  e  $B_2$  sono indipendenti. La proporzione di palline bianche nell'urna  $U_3$  è

$$X = \frac{B_1 + B_2}{10}.$$

Quindi

$$\mathbb{E}X = (\mathbb{E}B_1 + \mathbb{E}B_2)/10 = (4 \times 0.3 + 6 \times 0.5)/10 = 0.42$$

$$\mathbb{V}\text{ar}X = (\mathbb{V}\text{ar}B_1 + \mathbb{V}\text{ar}B_2)/10^2 = (4 \times 0.3 \times 0.7 + 6 \times 0.5 \times 0.5) / 100 = 0.0234$$

Domanda **3**  
Risposta corretta  
Punteggio max.: 1,00

— Importante —

Approssimate, se necessario, i risultati alla quarta cifra decimale.

— Esercizio —

Sia  $X$  una variabile aleatoria distribuita come una Normale di media 0 e varianza 4.  
Determinare:

1 La probabilità che  $X$  sia minore di 1.5.

0,7734



2 La probabilità che  $X$  sia maggiore di 3.

0,0668



3 La probabilità che  $X$  sia maggiore di 3 sapendo che  $X$  è minore di 6.

0,0655



Soluzione:

1. `pnorm(1.5,0,2)`  
0.7733726 approssimato a 0.7734

2. `1-pnorm(3,0,2)`  
0.0668072 approssimato a 0.0668

3. Usiamo la definizione di prob. condizionata:  
( `pnorm(6,0,2) - pnorm(3,0,2)` ) / `pnorm(6,0,2)`  
0.06554578 approssimato a 0.0655

Domanda **4**  
Risposta corretta  
Punteggio max.: 1,00

Consideriamo una variabile aleatoria  $X$  con funzione di  di probabilità  $f(x) = c \exp^{-cx}$  per  $x \in \mathbb{R}_+$  con  $c$  positivo, e zero altrove.  $X$  è una variabile aleatoria distribuita  con parametro  $c$  ed è  .

Risposta corretta.

Consideriamo una variabile aleatoria  $X$  con funzione di densità di probabilità  $f(x) = c \exp^{-cx}$  per  $x \in \mathbb{R}_+$  con  $c$  positivo, e zero altrove.  $X$  è una variabile aleatoria distribuita esponenzialmente con parametro  $c$  ed è continua.  
La risposta corretta è:  
Consideriamo una variabile aleatoria  $X$  con funzione di [densità] di probabilità  $f(x) = c \exp^{-cx}$  per  $x \in \mathbb{R}_+$  con  $c$  positivo, e zero altrove.  $X$  è una variabile aleatoria distribuita [esponenzialmente] con parametro  $c$  ed è [continua].

Domanda **5**  
Parzialmente corretta  
Punteggio max.: 1,00

— Importante —

- Approssimate, se necessario, i risultati alla **quarta cifra decimale**.

— Esercizio —

Un ente turistico delle Alpi ha condotto un’indagine su alcuni escursionisti per studiarne le preferenze e le abitudini. Per ciascun escursionista sono state resigtrate due variabili: la preferenza di percorso e il numero di escursioni effettuate nell’ultimo mese. I dati sono raccolti nel seguente file:

[escursioni.RData](#)

Fare una analisi descrittiva del dataset rispondendo alle domande seguenti.

1. Lo studio comprende

162

✓ osservazioni delle quali

71

✓ sono riferite ad escursionisti che preferiscono i sentieri facili. La proporzione di escursionisti che preferiscono i sentieri facili è pari a

0,4383

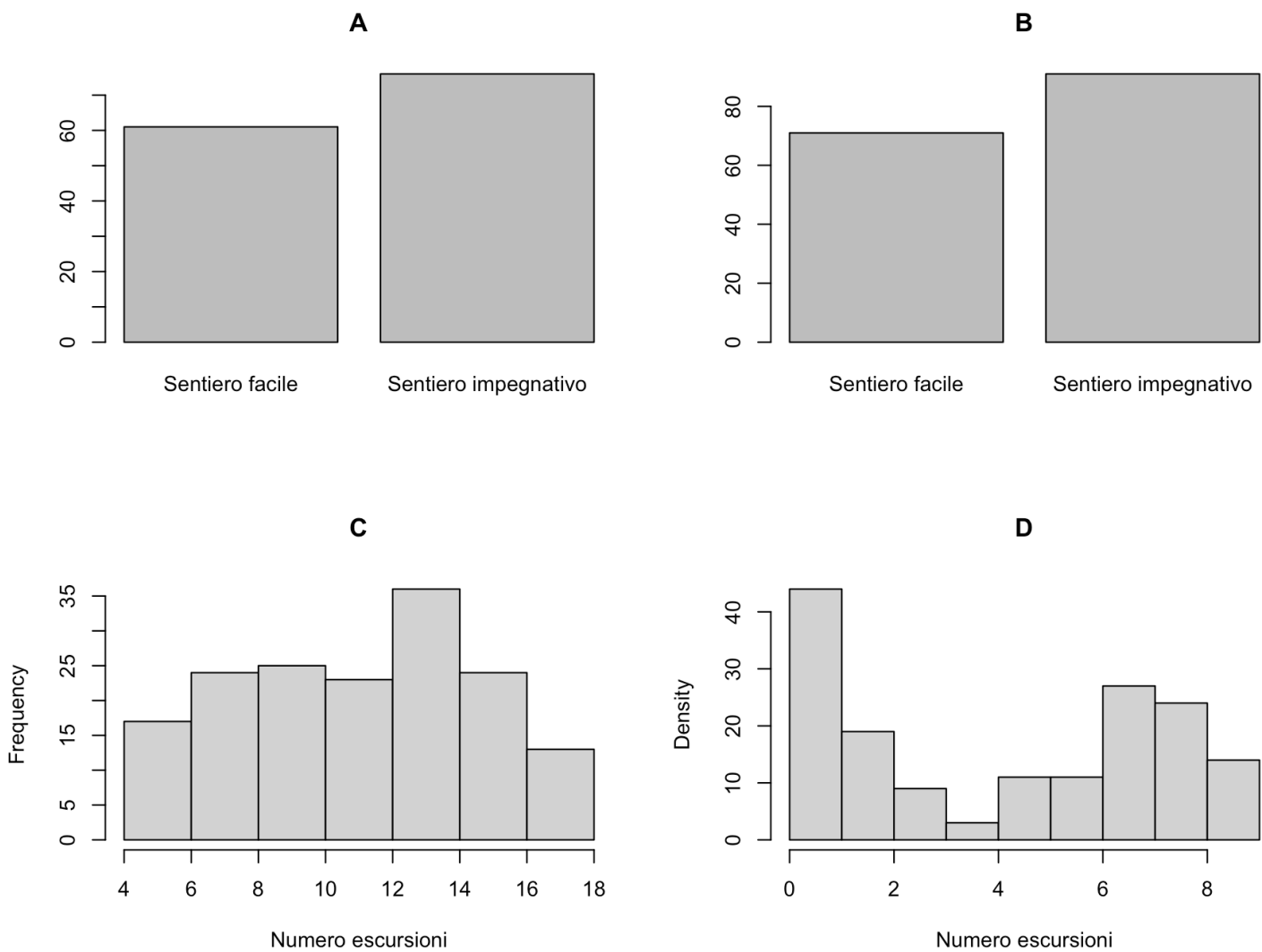
✓ .

2. Se consideriamo solo gli escursionisti che preferiscono i sentieri facili, il numero medio di escursioni effettuate nell’ultimo mese è pari a

1,3803

✓ .

3. Quale dei seguenti grafici è compatibile con i dati a disposizione?



Risposta:

- ☐ D  
☐ A

☒ B ✓

☐ C

La risposta corretta è: B

Si vuole dare una risposta quantitativa alla domanda: posso dire che più della metà degli escursionisti preferisce i sentieri impegnativi?

4. Per rispondere a questa domanda calcoli

- ☐ un test di ipotesi per la proporzione di escursionisti che preferisce i sentieri impegnativi con  $H_0: p = 0.5$  contro  $H_1: p \neq 0.5$
- ☒ un test di ipotesi per la proporzione di escursionisti che preferisce i sentieri impegnativi con  $H_0: p = 0.5$  contro  $H_1: p > 0.5$  ✓
- ☐ un test di ipotesi per la proporzione di escursionisti che preferisce i sentieri impegnativi con  $H_0: p = 0.5$  contro  $H_1: p < 0.5$
- ☐ nessuna di queste affermazioni è vera

La risposta corretta è: un test di ipotesi per la proporzione di escursionisti che preferisce i sentieri impegnativi con  $H_0: p = 0.5$  contro  $H_1: p > 0.5$

5. Ottengo un p-value pari a

0,1352

✗

6. Con livello di significatività pari a 0.01, posso affermare che

- ☒ non posso rifiutare l'ipotesi nulla ✓
- ☐ devo rifiutare l'ipotesi nulla

La risposta corretta è: non posso rifiutare l'ipotesi nulla

7. Con livello di significatività pari a 0.01, posso affermare che

- ☐ Al più il 50% degli escursionisti preferisce i sentieri impegnativi
- ☒ Il 50% degli escursionisti preferisce i sentieri impegnativi ✓
- ☐ Più del 50% degli escursionisti preferisce i sentieri impegnativi
- ☐ Nessuna delle affermazioni è vera

La risposta corretta è: Il 50% degli escursionisti preferisce i sentieri impegnativi

1. `nrow(escursioni)` e `table(escursioni$Preferenza)` e poi `table(escursioni$Preferenza)/sum(table(escursioni$Preferenza))`

2. `mean(escursioni$"Numero escursioni"[escursioni$Preferenza == "Sentiero facile"])`

3. `barplot(table(escursioni$Preferenza))`

4. Svolgo un test di ipotesi per la proporzione di escursionisti che preferisce i sentieri impegnativi con  $H_0: p = 0.5$  contro  $H_1: p > 0.5$

5. `table(escursioni$Preferenza)` e poi `binom.test(91,162, p = 0.5, alternative = "greater")` -> test e infine `test$p.value`

6. Non posso rifiutare  $H_0$  perché il p-value non è minore di 0.01.

7. Posso affermare  $H_0$  ovvero che la metà degli escursionisti preferisce i sentieri impegnativi.