| | |
|---|---|
| **Iniziato** | mercoledì, 29 gennaio 2025, 10:15 |
| **Stato** | Completato |
| **Terminato** | mercoledì, 29 gennaio 2025, 11:07 |
| **Tempo impiegato** | 52 min. |

Domanda **1**

Parzialmente corretta

Punteggio max.: 1,00

**— Importante —**

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

**— Esercizio —**

Un gruppo di ornitologi sta studiando la distribuzione degli uccelli in due tipi di habitat: una foresta primaria (non disturbata dall'uomo) e una foresta secondaria (ricresciuta dopo un intervento umano). Gli studiosi vogliono confrontare la percentuale di nidi occupati in questi due habitat per valutare la preferenza degli uccelli. I dati sono contenuti nel file seguente:

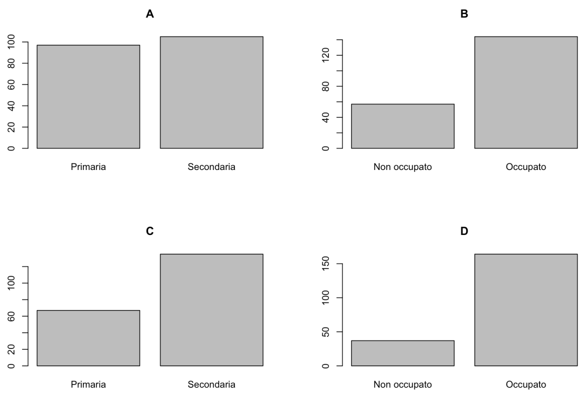
[nidi.RData]

Fare una analisi descrittiva del dataset rispondendo alle domande seguenti.

1. Lo studio comprende [202] ✔ osservazioni e 1 caso mancante che si trova alla riga [33] ✔ del dataset e che riguarda la variabile
   - ○ Foresta
   - ◉ Nido ✔

   La risposta corretta è: Nido

2. La frequenza di osservazioni di foresta **primaria** è pari a [87] ✔ e la corrispondente frequenza relativa è pari a [115] ✖ .
3. La frequenza di osservazioni di foresta **primaria** e con nidi **occupati** è pari a [51] ✔ .
4. Quale dei seguenti grafici è compatibile con i dati a disposizione?




Risposta:

- ○ A
- ○ B
- ◉ D ✔
- ○ C

La risposta corretta è: D

Si vuole dare una risposta quantitativa alla domanda: esiste una associazione tra il tipo di foresta e l'occupazione dei nidi?

5. Per rispondere a questa domanda svolgi
   - ○ un test di ipotesi del chi-quadro per l'indipendenza con H_0: le due variabili non sono indipendenti contro H_1: le due variabili sono indipendenti
   - ○ un test di ipotesi del chi-quadro per la bontà del fit con H_0: la variabile foresta ha distribuzione p_1 = p_2 = 1/2 contro H_1: la variabile foresta non ha la distribuzione indicata
   - ◉ un test di ipotesi del chi-quadro per l'indipendenza con H_0: le due variabili sono indipendenti contro H_1: le due variabili non sono indipendenti ✔

   La risposta corretta è: un test di ipotesi del chi-quadro per l'indipendenza con H_0: le due variabili sono indipendenti contro H_1: le due variabili non sono indipendenti

6. Ottengo un p-value pari a 0 ✔ .
7. Con livello di significatività pari a 0.05, posso affermare che
   - devo rifiutare H_0, quindi le due variabili sono indipendenti ✖
   - non posso rifiutare H_0, quindi le due variabili non sono indipendenti
   - non posso rifiutare H_0, quindi le due sono indipendenti
   - devo rifiutare H_0, quindi le due variabili non sono indipendenti

   La risposta corretta è: devo rifiutare H_0, quindi le due variabili non sono indipendenti

   .
8. La più piccola frequenza attesa è pari a 0,0456 ✖ . Secondo la regola di Cochran
   - devo controllare le frequenze attese e vedere se tutte sono almeno pari a 5
   - i risultati del test non sono affidabili perché la regola è violata ✖
   - i risultati del test sono affidabili perché la regola è rispettata
   - devo controllare le frequenze osservate e vedere se l'80% è almeno pari a 5

   La risposta corretta è: i risultati del test sono affidabili perché la regola è rispettata

   .

1. `nrow(dati)`
   ,
   `which(is.na(dati),arr.ind = TRUE)[1]`
   ,
   `dati[which(is.na(dati),arr.ind = TRUE)[1],]`
2. `table(dati, useNA = "always")`
3. Come sopra.
4. Disegno
   `barplot(table(dati[,2]), main = "D")`
5. Svolgo un test di ipotesi del chi-quadro per l'indipendenza che ha H_0: le due variabili sono indipendenti contro H_1: le due variabili non sono indipendenti.
6. `chisq.test(table(dati))`
   e
   `chisq.test(table(dati))$p.value`
7. Devo rifiutare H_0 perché il p-value è inferiore a 0.05, quindi le due variabili non sono indipendenti.
8. `min(chisq.test(table(dati))$expected)`
   . La regola di Cochran è soddisfatta.

Domanda **2**
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 1 e varianza 1.
Determinare:

1. La $\mathbb{P}(\{X=0\} \cup \{X=1\})$ 0 ✔
2. $\mathbb{P}(X \leq 2.5) + \mathbb{P}(X \geq 3) - \mathbb{P}(X < 2.5)$ 0,0228 ✔
4. La probabilità che X sia maggiore di 0.5 sapendo che X è minore di 6. 0,6915 ✔

Soluzione:

1. 0 (zero)

2. 1-pnorm(3,1,1) = 0.0228

3. Usiamo la definizione di prob. condizionata:
( pnorm(6,1,1) - pnorm(0.5,1,1) ) / pnorm(6,1,1) = 0.6915

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

La media campionaria è [una statistica] ✔ . Al crescere della taglia campionaria la sua media [rimane costante] ✔ . Al crescere della taglia campionaria la sua varianza [diminuisce] ✔ . Per taglia campionaria grande ha distribuzione approssimativamente [normale] ✔ . Questo ultimo risultato è l'enunciato [del teorema del limite centrale] ✔ .

| | | |
|---|---|---|
| normale | rimane costante | del teorema del limite centrale |
| una statistica | uniforme | una quantità pivotale |
| una costante | della legge dei grandi numeri | diminuisce |
| aumenta | esponenziale | |

La risposta corretta è:
La media campionaria è [una statistica]. Al crescere della taglia campionaria la sua media [rimane costante]. Al crescere della taglia campionaria la sua varianza [diminuisce]. Per taglia campionaria grande ha distribuzione approssimativamente [normale]. Questo ultimo risultato è l'enunciato [del teorema del limite centrale].

Domanda **4**

Parzialmente corretta

Punteggio max.: 1,00

— Importante —

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

— Esercizio —

Sia $X$ una variabile aleatoria Poisson(10) (cioè di parametro $lambda = 10$)

Calcolare:

1. la probabilità che "la variabile aleatoria valga 2 oppure assuma un valore nell'insieme $\{7, 8, 9\}$" 0,3278 ✗
2. la $\mathbb{P}(X \geq 5)$ 0,9707 ✓
3. la $\mathbb{P}(X \leq 3 | X \neq 0)$ 0,0103 ✓

NB: può essere utile la funzione di R "ppois"

1.

dpois(2,10) + dpois(7,10) + dpois(8,10) + dpois(9,10) = 0.3301

```
2.

1 - ppois(4,10) = 0.9707

3.
```

P(X <=3 | X <> 0) = P(X <=3 , X <> 0) / P(X <> 0) = P(1 <= X <= 3) / (1 - P(X = 0) )

= ( ppois(3,10) - ppois(0,10) ) / (1 - dpois(0,10)) = 0.0103

Domanda **5**

Parzialmente corretta

Punteggio max.: 1,00

**— Importante —**

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

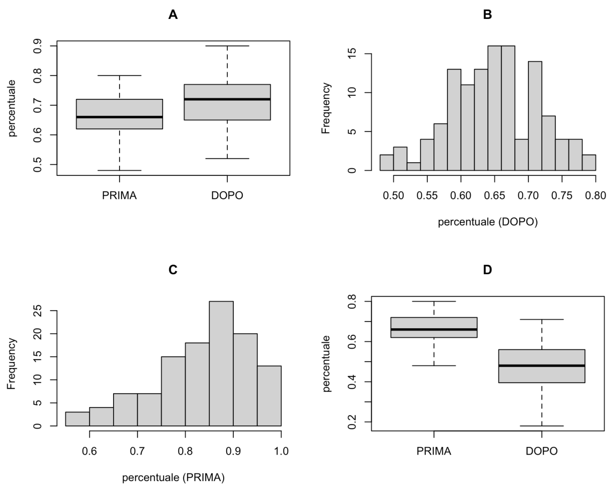
**— Esercizio —**

Un allenatore di pallavolo vuole verificare se un nuovo programma di allenamento migliora la precisione delle battute dei giocatori. Viene misurata la percentuale di battute efficaci per ciascun giocatore prima e dopo 6 settimane di allenamento seguendo il nuovo programma. I risultati sono raccolti nel dataset seguente:

[pallavolo.RData]

Fare una analisi descrittiva del dataset rispondendo alle domande seguenti.

1. Lo studio comprende [120] ✔ giocatori. La percentuale media di battute efficaci dei giocatori dopo le 6 settimane di nuovo programma di allenamento è pari a [0,7115] ✔ e la deviazione standard è pari a [0,084] ✔ .
2. Nel dataset ci sono [68] ✔ giocatori che realizzano una percentuale di battute efficaci dopo le 6 settimane di nuovo programma di allenamento superiore (>) a 0.7.
3. Il 5% dei giocatori realizzano una percentuale di battute efficaci dopo le 6 settimane di nuovo programma di allenamento superiore a [0,8305] ✔ .
4. Quale dei seguenti grafici è compatibile con i dati a disposizione?




Risposta:

- ○ D
- ◉ A ✔
- ○ B
- ○ C

La risposta corretta è: A

Si vuole dare una risposta quantitativa alla domanda: il nuovo programma di allenamento funziona? Ovvero, mediamente dopo il nuovo programma di allenamento la percentuale di battute efficaci dei giocatori è aumentata?

5. Per rispondere a questa domanda calcoli
   - ○ un test di ipotesi per la differenza di medie, per campioni indipendenti, con ipotesi alternativa H_1: mu_PRIMA > mu_DOPO
   - ○ un test di ipotesi per la differenza di medie, per campioni indipendenti, con ipotesi nulla H_0: mu_PRIMA = mu_DOPO
   - ◉ un test di ipotesi per la media delle differenze, per campioni appaiati, con ipotesi nulla H_0: mu_PRIMA < mu_DOPO ✖
   - ○ un test di ipotesi per la media delle differenze, per campioni appaiati, con ipotesi alternativa H_1: mu_PRIMA < mu_DOPO

   La risposta corretta è: un test di ipotesi per la media delle differenze, per campioni appaiati, con ipotesi alternativa H_1: mu_PRIMA < mu_DOPO

6. Ottengo un p-value pari a 0 ✔ .
7. Con livello di significatività pari a 0.01, posso affermare che
   - devo rifiutare l'ipotesi nulla ✔
   - non posso rifiutare l'ipotesi nulla

   La risposta corretta è: devo rifiutare l'ipotesi nulla

   .
8. Con livello di significatività pari a 0.01, posso affermare che
   - Non c'è sufficiente evidenza per affermare che la percentuale di battute efficaci è mediamente aumentata dopo il nuovo programma di allenamento
   - Non c'è sufficiente evidenza per affermare che il programma di allenamento è efficace
   - nessuna di queste affermazioni è corretta
   - C'è sufficiente evidenza per affermare che la percentuale di battute efficaci è mediamente aumentata dopo il nuovo programma di allenamento ✔

   La risposta corretta è: C'è sufficiente evidenza per affermare che la percentuale di battute efficaci è mediamente aumentata dopo il nuovo programma di allenamento

   .

1. `nrow(dati)`
   e
   `mean(dati$DOPO)`
   e
   `sd(dati$DOPO)`
2. `sum(dati$DOPO > 0.7)`
3. `quantile(dati$DOPO, 0.9)`
4. `boxplot(dati$PRIMA, dati$DOPO, ylab = "percentuale", main = "A", names=c("PRIMA","DOPO"))`
5. Svolgo un test di ipotesi per la media della differenza delle percentuali: **PRIMA - DOPO**, con ipotesi alternativa H_1: mu_PRIMA < mu_DOPO".
6. `t.test(dati$PRIMA, dati$DOPO, alternative = "less", paired = TRUE)`
7. Rifiuto H_0 perché il p-value è inferiore a 0.01.
8. Posso affermare che il nuovo allenamento è efficace, ovvero la percentuale media è aumentata.