# TEST DI AUTOVALUTAZIONE
# INTERVALLI DI CONFIDENZA E TEST



**Metodi statistici per la biologia**

## 1 Parte A

### 1.1

La formula

$$\mu = \overline{x} \pm \frac{s}{\sqrt{n}} t_{n-1, \frac{\alpha}{2}}$$

per l'intervallo di confidenza per la media di un campione Normale è valida

☐ solo se $n$ è grande;

☐ solo se $n$ è piccolo;

☒ per qualsiasi valore di $n$;

☐ solo per valori grandi di $\alpha$.

### 1.2

L'ampiezza dell'intervallo di confidenza di un campione Normale a varianza nota

☐ dipende dal campione di dati, oltre che dalla taglia $n$ del campione;

☒ dipende dalla taglia $n$ del campione, ma non dal valore dei dati;

☐ ha distribuzione normale;

☐ ha distribuzione $t$ di Student.

### 1.3

Sia $\Delta$ la lunghezza dell'intervallo di confidenza al $100(1-\alpha)\%$ per la media di un campione di $n$ dati estratto da una $N(\mu, \sigma^2)$ con $\sigma^2$ nota. Impiegando un campione di $4n$ dati, a parità di $\alpha$ e $\sigma^2$ la lunghezza dell'intervallo di confidenza è

☐ $4\Delta$;

☐ $\frac{\Delta}{4}$;

☒ $\frac{\Delta}{2}$;

☐ bisogna conoscere il valore di $\sigma^2$ per poterlo dire.

### 1.4

Quale delle seguenti è un'ipotesi statistica?

☐ un campione di dati;

☐ un insieme di valori assunti dal campione;

☐ una stima per un parametro incognito di una distribuzione

☒ un'affermazione su un parametro incognito di una distribuzione.

## 1.5

Commettere un errore di I specie nella verifica di un ipotesi significa:

☐ rifiutare l'ipotesi;

☒ rifiutare l'ipotesi quando essa è vera;

☐ accettare l'ipotesi;

☐ accettare l'ipotesi quando essa è falsa.

## 1.6

In un certo di test di verifica di ipotesi la regione critica è del tipo $ST > k$ e vale $P^{H_0}(ST > 10) = 0,05$, dove $P^{H_0}$ denota la probabilità calcolata assumendo vera l'ipotesi nulla $H_0$. Sia $\alpha$ il livello di significatività del test.

☐ se $\alpha = 0.01$ e $T(x_1, \ldots, x_n) = 9$ allora $H_0$ viene rifiutata;

☐ se $\alpha = 0.1$ e $T(x_1, \ldots, x_n) = 12$ allora $H_0$ viene accettata;

☒ se $\alpha = 0.05$, rifiuto $H_0$ se $T(x_1, \ldots, x_n) > 10$;

☐ nessuna delle precedenti è necessariamente vera.

## 1.7

Sia $p$ il valore-$p$ di un test per la verifica dell'ipotesi $H_0$, e sia $\alpha$ il livello del test effettuato. Allora

☒ se $\alpha > p$ l'ipotesi $H_0$ viene rifiutata;

☐ se $\alpha < p$ l'ipotesi $H_0$ viene rifiutata;

☐ $p$ è la probabilità di commettere un'errore nella verifica dell'ipotesi $H_0$;

☐ $p$ è un parametro incognito nella distribuzione del campione.

## 1.8

Sia $p$ il valore-$p$ di un test sulla media di un campione Gaussiano. Allora

☒ $p$ dipende dai dati osservati;

☐ $p$ è sicuramente maggiore di 0.5;

☐ se l'ampiezza del test è $> p$ allora l'ipotesi viene accettata;

☐ nessuna delle precedenti.

## 1.9

La regione critica di un test di verifica dell'ipotesi $H_0$ e‘

☐ il valore del $p$-dei-dati del test;

☐ un insieme di valori del parametro incognito che contrasta con i dati sperimentali;

☒ un insieme di valori che, se assunti dalla statistica-test, conduce al rifiuto di $H_0$;

☐ un insieme di valori del livello di significativita‘ per i quali $H_0$ viene respinta.

## 1.10

L'intervallo di confidenza al 95% per la media di un campione normale contiene il valore 0. Allora

☐ sicuramente anche l'intervallo di confidenza al 90% contiene il valore 0;

☐ l'intervallo di confidenza al 99% non contiene il valore 0;

☒ l'ipotesi $H_0 : \mu = 0$ viene accettata al 5% di significatività;

☐ l'ipotesi $H_0 : \mu = 0$ viene rifiutata all' 1% di significatività.

## 1.11

In un test per la verifica dell'ipotesi $H_0 : \mu \geq 1$ per un campione normale, il valore-$p$ calcolato sui dati è $p = 0.03$. Quale delle seguenti affermazioni è *sicuramente* vera?

☒ sulla base degli stessi dati, $H_0$ viene rifiutata a livello di significatività $\alpha = 0.05$;

☐ la probabilità di rifiutare $H_0$ è 0.97;

☐ i dati sono in sostanziale accordo con l'ipotesi $H_0$;

☐ sulla base degli stessi dati, $H_0$ viene accettata a livello di significatività $\alpha = 0.15$.

## 1.12

In un test per la verifica dell'ipotesi $H_0 : \mu = \mu_0$ per un campione normale, la statistica del test è $T$. Nel caso in cui $H_0$ è vera, si ha che $P(|ST| \geq 2) = 0.04$. Si raccolgono dei dati in corrispondenza dei quali $ST = 2$. Allora

☐ $H_0$ viene rifiutata al 1%;

☐ $H_0$ viene accettata al 10%;

☒ $H_0$ viene rifiutata al 7%;

☐ $H_0$ viene accettata al 7%.

## 1.13

Si vuole stimare la media $\mu$ di una variabile con distribuzione normale di varianza incognita. Calcolando l'intervallo di confidenza al 95% su un campione di taglia 20, si ottiene $I_1 = [4.21, 4.56]$. Si ricalcola quindi l'intervallo di confidenza al 95% su un secondo campione, sempre di taglia 20, ottendendo $I_2 = [4.59, 4.73]$. Da questi dati si può concludere che:

☐ la media $\mu$ è certamente compresa tra 4.56 e 4.59

☐ la media $\mu$ cade in uno dei due intervalli $I_1$, $I_2$ con probabilità almeno del 10%

☒ la media $\mu$ potrebbe non appartenere a nessuno dei due intervalli $I_1$, $I_2$

☐ è stato commesso un errore nelle misurazioni o nei calcoli, perché $I_1$ e $I_2$ non possono essere disgiunti

## 1.14

Effettuando un test d'ipotesi di livello di significatività $\alpha$ sono sicuro che

- ☐ $P(\text{accetto } H_0 \text{ quando è corretta}) \leq \alpha$
- ☐ $P(\text{accetto } H_0 \text{ quando è sbagliata}) \leq \alpha$
- ☒ $P(\text{rifiuto } H_0 \text{ quando è corretta}) \leq \alpha$
- ☐ $P(\text{rifiuto } H_0 \text{ quando è sbagliata}) \geq \alpha$

## 1.15

Si consideri un test di verifica di ipotesi su un parametro incognito $\mu$, per l'ipotesi $H_0 : \mu \geq 0$. Quale delle seguenti affermazioni è corretta?

- ☐ La regione critica del test è $\{\mu$ tali che $\mu < 0\}$.
- ☒ La regione critica del test è l'insieme dei valori della statistica test per cui $H_0$ viene rifiutata.
- ☐ La regione critica del test è l'insieme dei valori della statistica test per cui $H_0$ viene accettata.
- ☐ La regione critica del test è $\{\mu$ tali che $\mu \geq 0\}$.

## 1.16

Viene effettuato un test per verificare una certa ipotesi nulla $H_0$. Se il valore-$p$, calcolato sui dati osservati, vale 0.73, si può concludere che:

- ☒ i dati sono in accordo con $H_0$
- ☐ i dati sono in forte disaccordo con $H_0$
- ☐ $H_0$ è vera
- ☐ $H_0$ è falsa

## 1.17

Un'associazione di consumatori ha il sospetto che il contenuto medio $\mu$ di grassi in un prodotto dietetico sia maggiore del valore dichiarato $\mu_0$. Si fissi un livello di significatività $\alpha$, e si denoti con $H_0$ l'ipotesi che viene verificata a livello $\alpha$. In quale dei seguenti casi possiamo concludere che il sospetto sia fondato?

- ☐ $H_0 : \mu \leq \mu_0$, e $H_0$ viene accettata
- ☒ $H_0 : \mu \leq \mu_0$, e $H_0$ viene rifiutata
- ☐ $H_0 : \mu \geq \mu_0$, e $H_0$ viene accettata
- ☐ $H_0 : \mu \geq \mu_0$, e $H_0$ viene rifiutata

## 1.18

Si vuole stimare la percentuale $p$ di membri di una popolazione che hanno una data caratteristica. Viene selezionato un campione di 100 individui, di cui 15 hanno la data caratteristica. Quale delle seguenti affermazioni è *vera*?

- ☐ l'ipotesi $H_0 : p = 0.20$ viene rifiutata per qualsiasi livello di significatività;
- ☐ l'ipotesi $H_0 : p \geq 0.20$ viene rifiutata per qualsiasi livello di significatività;
- ☒ l'ipotesi $H_0 : p \leq 0.20$ viene accettata per qualsiasi livello di significatività;
- ☐ l'ipotesi $H_0 : p \leq 0.10$ viene rifiutata per qualsiasi livello di significatività;

### 1.19

In un test per la verifica dell'ipotesi $H_0$ viene usata la statistica test $ST$, e la regione critica è $\{ST > 10\}$. Supponiamo che $H_0$ sia vera e che i dati osservati $(x_1, x_2, \ldots, x_n)$ siano tali che $ST = 2$. Allora

☐ si commette un errore di prima specie

☐ si commette un errore di seconda specie

☒ non si commette alcun errore

☐ nessuna delle precedenti

### 1.20

Vogliamo stimare la media incognita $\mu$ di una variabile con legge normale di varianza nota pari a 1. Usando un campione $x_1, \ldots, x_n$ estratto da tale variabile otteniamo l'intervallo di confidenza bilatero al 95% pari a $[0.25, 0.59]$. Se ora, *usando lo stesso campione*, calcoliamo l'intervallo di confidenza bilatero al 99%, quale dei seguenti risultati è il solo possibile?

☐ $[0.31, 0.53]$

☒ $[0.19, 0.65]$

☐ $[0.31, 0.65]$

☐ $[0.19, 0.53]$

### 1.21

Effettuando un test sulla media di un campione normale con varianza incognita, il campione di dati cade all'interno della regione critica per il livello di significatività 1%. Allora

☒ l'ipotesi $H_0$ viene rifiutata al 5% di significatività

☐ l'ipotesi $H_0$ viene accettata all'1% di significatività

☐ il $p$-value del test è maggiore di 0.01

☐ l'ipotesi $H_0$ è falsa

## 2 Parte B

### 2.1

Per rilevare il grado di apprezzamento della mensa universitaria, viene intervistato un campione di 100 studenti, 67 dei quali dichiarano di apprezzare la qualità del cibo servito in mensa. Si indichi con $p$ la probabilità che uno studente scelto a caso tra tutti gli iscritti apprezzi il cibo servito in mensa.

a) Usando i dati forniti dal problema, si determini l'intervallo di confidenza al 95% per $p$.

b) Quanto deve essere grande il campione di studenti intervistati, affinché l'ampiezza dell'intervallo di confidenza al 95% per $p$ sia più piccola di 0.04?

**Soluzione.**

a) I dati forniti sono $n = 100$ e $\overline{x} = \frac{67}{100} = 0.67$. Posto $\alpha = 0.05$, dato che $z_{\alpha/2} = z_{0.025} = 1.96$ l'intervallo di confidenza cercato vale

$$\left[\overline{x} - \sqrt{\frac{\overline{x}(1-\overline{x})}{n}} z_{\alpha/2} \ , \ \overline{x} + \sqrt{\frac{\overline{x}(1-\overline{x})}{n}} z_{\alpha/2}\right]$$
$$= \left[0.67 - \sqrt{\frac{0.67(1-0.67)}{100}} 1.96 \ , \ 0.67 + \sqrt{\frac{0.67(1-0.67)}{100}} 1.96\right]$$
$$= [0.67 - 0.047 \cdot 1.96 \ , \ 0.67 + 0.047 \cdot 1.96] \ = \ [0.58 \, , \, 0.76] \, .$$

b) Dato che $\sqrt{x(1-x)} \leq 1/2$ per ogni $x \in [0, 1]$, per l'ampiezza $I$ dell'intervallo di confidenza si ha

$$I \ = \ 2\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} \, z_{\alpha/2} \ \leq \ \frac{z_{\alpha/2}}{\sqrt{n}} \, ,$$

per cui per avere $I \leq 0.04$ basta imporre che

$$\frac{z_{\alpha/2}}{\sqrt{n}} \ \leq \ 0.04 \qquad \Longleftrightarrow \qquad n \ \geq \ \left(\frac{z_{\alpha/2}}{0.04}\right)^2 \ = \ \left(\frac{1.96}{0.04}\right)^2 \ = \ 2401 \, .$$

## 2.2

Una casa farmaceutica vuole determinare il tempo medio di risposta dell'organismo all'assunzione di un nuovo antidolorifico per i dolori articolari. Il farmaco viene provato su 10 individui, i quali iniziano ad avvertire effetti positivi dopo i tempi che seguono, misurati in minuti:

35 24 51 42 40 33 37 41 44 29

Si assuma che il tempo di risposta abbia distribuzione normale di media $\mu$ e varianza $\sigma^2$, entrambe incognite. Determinare intervalli di confidenza al 95% e al 99% per $\mu$.
**Soluzione**. Abbiamo: $\overline{x} = 37.6$, $s = 7.78$, $t_{0.025,9} = 2,26$, $t_{0.005,9} = 3.25$. Gli intervalli di confidenza al 95% e al 99% per la media sono, rispettivamente

$$\left[\overline{x} - \frac{s}{\sqrt{10}} t_{0.025,9}, \overline{x} + \frac{s}{\sqrt{10}} t_{0.025,9}\right] = [32.04, 43.16]$$
$$\left[\overline{x} - \frac{s}{\sqrt{10}} t_{0.005,9}, \overline{x} + \frac{s}{\sqrt{10}} t_{0.005,9}\right] = [29.6, 45.59]$$

## 2.3

È in corso uno studio clinico per la sperimentazione di un nuovo farmaco che si sospetta possa avere come effetto indesiderato un abbassamento della pressione arteriosa. Viene misurata la pressione sistolica di 20 soggetti in trattamento con il nuovo farmaco ottenendo media campionaria di 105 mm/Hg e deviazione standard campionaria di 10 mm/Hg. Gli studiosi ritengono che se il farmaco induce una pressione sistolica media inferiore ai 110 mm/Hg, sia inopportuno metterlo in commercio. Sulla base dei dati a disposizione quale sarebbe la vostra decisione? (testare a livello di significatività $\alpha = 0.01$).
**Soluzione.** Verifichiamo l'ipotesi $H_0 : \mu \geq 110$, dove $\mu$ è la pressione arteriosa media. $H_0$ viene rifiutata se

$$st := \frac{\overline{x} - 110}{10}\sqrt{20} < -t_{0.01,19}.$$

Essendo $t = -2.24$ e $t_{0.01,19} = 2.86$ l'ipotesi viene accettata, cioè i dati non sono sufficienti a concludere, per un test all'1%, che la pressino media scenda sotto i 110 mm/Hg. Tuttavia va considerato che il valore di $t$ è assai vicino alla soglia, e dunque i dati inducono un ragionevole sospetto.

## 2.4

Si vuole determinare, all'interno di una certa popolazione di alcune decine di milioni di individui, la percentuale $p$ di portatori sani di una determinata malattia. Dalle analisi su 1000 individui, e' risultato che 41 di essi sono portatori sani. Determinare un intervallo di confidenza per $p$ di livello 0.99.
**Soluzione**. Si tratta di determinare l'intervallo di confidenza per il parametro di un campione Bernoulliano, dato dalla formula

$$\overline{x} \pm \frac{\sqrt{\overline{x}(1-\overline{x})}}{\sqrt{n}} z_{\alpha/2}.$$

In questo caso $\alpha = 0.01$, $\overline{x} = 41/1000 = 0.041$, $z_{0.005} = 2.58$, $n = 1000$ .....

## 2.5

Nella sperimentazione di un farmaco si e' interessati a misurare il tempo impiegato dall'organismo ad assorbire i suoi principi attivi. Si assuma che tale tempo di assorbimento abbia distribuzione normale, di media $\mu$ e varianza $\sigma^2$ entrambe incognite. Una sperimentazione su 20 volontari ha prodotto un tempo di assorbimento medio campionario $\overline{x} = 35.2$ min., e una deviazione standard campionaria di 12.4 min.

a. Determinare un intervallo di confidenza per $\mu$ di livello di confidenza 0.95.

b. Effettuare un test, con livello di significativita' $\alpha = 0.01$, sull'ipotesi $H_0 : \mu \geq 40$.

**Soluzione**. a. L'intervallo di confidenza è fornito da

$$35.2 \pm \frac{12.4}{\sqrt{20}} t_{0.025,19}.$$

b. Rifiuto $H_0$ se

$$st := \frac{35.2 - 40}{12.4}\sqrt{20} < -t_{0.05,19}.$$

Essendo $st = -1.73$ e $t_{0.05,19} = 2.09$, $H_0$ viene accettata.

## 2.6

Uno strumento di misura viene impiegato per determinare una grandezza incognita $\mu$. Lo strumento fornisce un valore aleatorio con distribuzione normale di media $\mu$ e varianza 0.01. Vengono effettuate 10 misure ed i valori osservati sono

2.1, 2.2, 1.9, 1.8, 2.3, 2.2, 1.7, 1.8, 2.1, 1.9

a) Determinare un intervallo di confidenza al livello 0.95 per la media $\mu$.

b) Si determini il valore-$p$ del test per l'ipotesi $H_0 : \mu = 2.2$. Quali conclusioni si possono trarre?

**Soluzione**. a) Essendo $\overline{x} = 2$ e $z_{0,025} = 1.96$, $\sigma^2 = 0.01$, l'intervallo di confidenza è

$$\left[\overline{x} - \frac{\sigma}{\sqrt{10}} z_{0,025}, \overline{x} + \frac{\sigma}{\sqrt{10}} z_{0,025}\right] = [1,94, 2.06].$$

b) Il $p$-value $\bar{\alpha}$ soddisfa la relazione

$$z_{\bar{\alpha}/2} = \left|\frac{\overline{x} - 2.2}{0.1}\sqrt{10}\right| = 6.32,$$

cioè

$$\bar{\alpha} = 2[1 - \Phi(6.32)] \simeq 0.$$

Si può senz'altro concludere che i dati sono in forte contraddizione con $H_0$.

## 2.7

In un Videopoker la probabilita‘ che un giocatore vinca dovrebbe essere 3/100. Un controllore vuole verificare se il gioco sia stato truccato. Ripete il gioco 1000 volte, vincendone 26.

a. Verificare l'ipotesi che il gioco non sia stato truccato, con livello di significativita‘ $\alpha = 0.05$

b. Calcolare il valore-$p$ del test al punto a.

**Soluzione**. a. Si tratta di una verifica di ipotesi per un campione Bernoulliano, con $n = 1000$ e $\overline{x} = 0.026$. Possiamo applicare il test asintotico, essendo $n\overline{x}(1-\overline{x}) > 5$. L'ipotesi di equità del gioco ($H_0 : p = 0.03$) viene respinta a livello 0.05 se

$$\left| \frac{0.026 - 0.03}{\sqrt{0.03 \times 0.97}} \sqrt{1000} \right| = 0.74 > z_{0.025}.$$

Essendo $z_{0.025} = 1.96$, l'ipotesi viene accettata: a questo livello di significatività i dati non sono sufficienti a concludere che che il gioco sia truccato.

b. Il $p$-value è

$$\overline{\alpha} = 2(1 - \Phi(0.74)) = 0.46.$$

Per nessun livello ragionevole di significatività l'ipotesi viene respinta.

## 2.8

Secondo un modello teorico, una certa popolazione di batteri in particolari condizioni ha una probabilità 0.4 di sopravvivere oltre 10 ore, cioè si estingue con probabilità 0.6. Vengono preparate in laboratorio 50 copie di tale popolazione, al fine di stimare la probabilità $p$ di estinzione. Di queste 50, 33 si esinguono entro 10 ore.

a. Sulla base di questi dati determinare il valore-$p$ del test relativo all'ipotesi $H_0 : p = 0.6$;

b. L'ipotesi $H_0$ viene rifiutata o accettata a livello di significatività 0.05?

c. Determinare un intervallo di confidenza per $p$ al 95%.

d. Se avessimo voluto stimare $p$ al 95% con una precisione di $\pm 0.01$, quante copie della popolazione avremmo dovuto preparare in laboratorio?

**Soluzione**. a. Si tratta di uno $z$-test per un campione Bernoulliano. Posto

$$st = \frac{\overline{x} - 0.6}{\sqrt{0.6 \cdot 0.4}} \sqrt{50} = \frac{\frac{33}{50} - 0.6}{\sqrt{0.6 \cdot 0.4}} \sqrt{50} = 0.866.$$

Il $p$-value del relativo test è dato da

$$2(1 - \Phi(|st|)) = 0.3865.$$

b. Essendo il livello di significatività minore del $p$-value, $H_0$ viene accettata.

c.

$$p = \frac{33}{50} \pm \frac{\sqrt{\frac{33}{50}\frac{17}{50}}}{\sqrt{50}} z_{0.025} = 0.66 \pm 1.13$$

d. Di dovrebbe avere

$$\frac{1}{2\sqrt{n}} z_{0.025} \leq 0.01$$

cioè

$$\sqrt{n} \geq 97.998 \;\Rightarrow\; n \geq 9603.65 \;\Rightarrow\; n \geq 9604.$$

Avremmo dunque dovuto preparare 9604 copie!

## 2.9

Di 70 studenti universitari scelti a caso, 37 sono miopi.

a. Determinare un intervallo di confidenza al 95% per la percentuale di studenti miopi.

b. È noto che, nella popolazione generale, i miopi sono il 45%. Questi dati consentono di concludere che la miopia tra gli studenti sia più diffusa che nella popolazione generale (usare un test al 5%)?

c. Quanti studenti avremmo dovuto scegliere per essere certi che la semi-ampiezza dell'intervallo di confidenza fosse al massimo di 2 punti percentuali?

**Soluzione**. a.

$$\overline{x} \pm \frac{\sqrt{\overline{x}(1-\overline{x})}}{\sqrt{n}} z_{0.025} = 0.528571429 \pm 0.116938791$$

b. Verifichiamo l'ipotesi $H_0 : p \leq 0.45$, dove $p$ è la percentuale di studenti miopi. Rifiutiamo $H_0$ se

$$st = \frac{\overline{x} - 0.45}{\sqrt{0.45 \times 0.55}/\sqrt{n}} > z_{0.05}.$$

Essendo $st = 1.321374945$ e $z_{0.05} = 1.644853476$, $H_0$ viene accettata: i dati non consentono di concludere che la miopia tra gli studenti sia più diffusa che nella popolazione generale.

c. Notiamo che la semiampiezza dell'intervallo di confidenza è

$$\frac{\sqrt{\overline{x}(1-\overline{x})}}{\sqrt{n}} z_{0.025} \leq \frac{1}{2\sqrt{n}} z_{0.025},$$

dove abbiamo usato il fatti che $\overline{x}(1-\overline{x}) \leq 1/4$ per ogni possibile valore di $\overline{x}$. Dunque, per essere certi che la semi-ampiezza dell'intervallo di confidenza fosse al massimo di 2 punti percentuali avremmo dovuto avere

$$\frac{1}{2\sqrt{n}} z_{0.025} \leq 0.02 \iff \sqrt{n} \geq \frac{z_{0.025}}{0.04} = 49 \iff n \geq 2401.$$

## 2.10

Viene misurato il livello di diossina nell'aria in una certa area urbana, in 8 giorni distinti. I dati ottenuti, in $ng/m^3$, sono

0.161 0.083 0.142 0.156 0.103 0.091 0.162 0.113

Si ritiene che una zona sia pericolosa se il livello medio di diossina supera 0.16 $ng/m^3$. Dai dati si può concludere che l'area in esame sia sicura? (Si effettui un test al 5%)

**Soluzione**. Effettuiamo un $t$-test per $H_0 : \mu \geq 0.16$. La statistica test è

$$st = \frac{\overline{x} - 0.16}{S/\sqrt{8}} = -2.915.$$

Inoltre $t_{7,0.5} = 1.895$. Essendo $|st| > t_{7,0.5}$, $H_0$ viene rifiutata: a livello di significativit'a del 5%, i dati contraddicono l'ipotesi che la zona sia pericolosa: possiamo dunque concludere che la zona sia sicura.

## 2.11

In uno dei suoi famosi esperimenti, Mendel esaminò 580 piselli, osservando che esattamente 152 di questi erano gialli. Mendel aveva ipotizzato che la probabilità $p$ che uno di piselli fosse giallo fosse 0.25. Ritenete che i dati osservati contraddicano l'ipotesi di Mendel in modo significativo? (Calcolare il $p$-value del test eseguito)

**Soluzione**. Effettuando un test sull'ipotesi $H_0 : p = 0.25$, si calcola la statistica test

$$\frac{\overline{x} - 0.25}{\sqrt{0.25 \cdot 0.75}} \sqrt{580} = \frac{\frac{152}{580} - 0.25}{\sqrt{0.25 \cdot 0.75}} \sqrt{580} \simeq 0.67.$$

Il $p$-value del test è (se $Z \sim N(0,1)$)

$$2(1 - P(Z \leq 0.67)) \simeq 0.5.$$

L'elevato valore del $p$-value indica che i dati non sono affatto in contraddizione con l'ipotesi di Mendel.

## 2.12

Una ditta che produce cartucce ad inchiostro per stampanti vuole stimare la vita media (cioè il numero medio di fogli dopo i quali le cartucce si esauriscono). A tal scopo estrae un campione di 8 cartucce e registra i seguenti dati.

$$1115,\ 1097,\ 1120,\ 1095,\ 1105,\ 1075,\ 1079,\ 1114$$

1. Supponendo che la popolazione delle cartucce abbia una vita con distribuzione normale e deviazione standard $\sigma = 16.67$ determinare un intervallo di confidenza al 95 per cento per la media $\mu$.
2. Determinare un intervallo di confidenza al 95 per cento per la media $\mu$ relativa al medesimo campione, supponendo di non conoscere la deviazione standard $\sigma$.

**Soluzione**.

1. $\overline{x} = 1100$, $z_{0.025} \simeq 1.96$. Pertanto, l'intervallo di confidenza è dato da
$$\overline{x} \pm \frac{\sigma}{\sqrt{8}} z_{0.025} \simeq 1100 \pm 11.55.$$
2. Se non conosciamo la deviazione standard, calcoliamo la deviazione standard campionaria $s \simeq 16.67$. In questo caso l'intervallo di confineza è dato da (essendo $t_{7,0.025} \simeq 2.36$)
$$\overline{x} \pm \frac{s}{\sqrt{8}} t_{7,0.025} \simeq 1100 \pm 13.91.$$

## 2.13

Una compagnia di assicurazioni deve eseguire uno studio per stimare gli indennizzi pagati a seguito di incidenti automobilistici senza lesioni alle persone. Da studi precedenti è emerso che si può assumere che tali importi abbiano distribuzione normale con media $\mu$ incognita e deviazione standard nota, e pari a 900 euro. Su un nuovo campione casuale di 100 incidenti è stato osservato un indennizzo medio pari a 5562 euro.

a) Determinare un intervallo di confidenza di livello 94% per il parametro $\mu$.

b) Verificate l'ipotesi $H_0 : \mu = 5500$ contro $H_1 : \mu \neq 5500$ al livello 6%.

c) In realtà la compagnia ritiene che l'intervallo di confidenza costruito al punto 1 non sia sufficientemente preciso. Decide quindi di condurre uno studio più vasto, cioè su un campione più numeroso. Determinate il numero minimo di casi da esaminare affinché la lunghezza dell'intervallo di confidenza per $\mu$ non superi i 300 euro.

**Soluzione**.

a) L'intervallo è
$$\overline{x} \pm \frac{\sigma}{\sqrt{n}} z_{0.03} = 5562 \pm \frac{900}{10} 1.88 = 5562 \pm 167.27.$$

b)
$$st = \frac{\overline{x} - 5500}{900/10} \simeq 0.69 < z_{0.03} \simeq 1.88.$$

Pertanto $H_0$ viene accettata.

c) Dev'essere
$$\frac{900}{\sqrt{n}} 1.88 \leq 150$$

da ciò si ricava $n \geq 128$

## 2.14

Secondo un modello teorico, un certo tipo di filamenti molecolari ha una lunghezza con distribuzione normale, di media 27 nm, e deviazione standard 6 nm.

1) Assumendo la validità di una tale modello, qual è la probabilità di osservare un filamento con una lunghezze superiore ai 35 nm?

2) In una serie di esperimenti, vengono osservati $n = 120$ diversi filamenti. I metodi di misura consentono solo di dire che 34 di essi hanno lunghezza maggiore di 35 nm, mentre gli altri hanno una lunghezza minore di 34 nm. Questi dati sono compatibili con il modello teorico? (Effettuare un test al 5%)

**Soluzione**.

1) Sia $X \sim N(27, 36)$ e $Z \sim N(0, 1)$. Allora
$$P(X > 35) = P\left(\frac{X - 27}{6} > \frac{35 - 27}{6}\right) = P(Z > 1.333) = 1 - P(Z \leq 1.333) \simeq 0.091$$

2) Effettuiamo un test sulla proporzione $p$ di filamenti osservati di lunghezza maggiore di 35 nm, per l'ipotesi $H_0 : p = 0.091$. La statistica test è
$$st = \frac{\overline{x} - 0.091}{\sqrt{0.091(1 - 0.091)}} \sqrt{n} = \frac{\frac{34}{120} - 0.091}{\sqrt{0.091(1 - 0.091)}} \sqrt{120} \simeq 7.326.$$

Poichè $|st| > z_{0.025} \simeq 1.96$, $H_0$ viene rifiutata: i dati non sono compatibili con il modello teorico.

## 2.15

Un famoso quotidiano californiano ha recentemente pubblicato i risultati di un sondaggio, dal quale risulta che la percentuale dei residenti di Los Angeles favorevoli alla costruzione di una nuova centrale nucleare nella zona è del 35% con un margine di errore di ±3.305%. Assumiamo che che tale margine di errore corrisponda ad un livello di confidenza del 95%.

a) Spiegare il metodo usato per elaborare i dati del sondaggio.

b) Qual'è la numerosità del campione utilizzato?

**Soluzione**.

a) È stata utilizzata la formula per l'intervallo di confidenza per una proporzione:
$$\hat{p} \pm \frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} z_{0.025},$$

dove, in questo caso, $\hat{p} = 0.35$ e $z_{0.05} \simeq 1.96$.

b) Dev'essere

$$0.03305 = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} z_{0.025}.$$

Sostituendo i dati di cui sopra e risolvendo per $n$, si trova $n = 800$.

## 2.16

Prima che venisse introdotto l'obbligo delle cinture di sicurezza, i feriti in incidenti d'auto trascorrevano in media 1.39 giorni nei reparti di terapia intensiva. Dopo l'introduzione dell'obbligo, su un campione casuale di 123 feriti in incidenti d'auto, il tempo medio (campionario) trascorso in terapia intensiva è risultato di 0.83 giorni, con una deviazione standard campionaria di 0.16 giorni. Questi dati dimostrano l'efficacia dell'introduzione dell'obbligo di indossare le cinture di sicurezza? (Eseguire un test all'1%.

**Soluzione**. Assumiamo la normalità della distribuzione, ed eseguiamo un $t$-test sulla media $\mu$, per verificare l'ipotesi $H_0 : \mu \geq 1.39$ (ossia l'ipotesi di inefficacia delle cinture di sicurezza). La statistica test è

$$st = \frac{\overline{x} - 1.39}{s/\sqrt{123}} = \frac{0.83 - 1.39}{0.16/\sqrt{123}} \simeq -38.81.$$

Per un test all'1%, rifiutiamo $H_0$ se $ST < -t_{122,0.01} \simeq -2.36$. Pertanto $H_0$ viene rifiutata: i dati dimostrano l'efficacia dell'introduzione dell'obbligo di indossare le cinture di sicurezza.

## 2.17

In un'azienda vinicola si produce un cabernet la cui gradazione alcolica è, per ragioni storiche, ritenuta essere in media pari a 12.5. A seguito di una campionatura casuale di otto bottiglie si registrano le seguenti gradazioni:

$$12.1,\ 12.2,\ 12.7,\ 12,\ 12.8,\ 12.9,\ 12.65,\ 12.75$$

1. Supponendo che la gradazione alcolica abbia distribuzione normale con deviazione standard $\sigma = 0.15$ determinare se, applicando un livello di significatività pari a 0.05, i dati sono consistenti con la gradazione media "storica".

2. Si calcoli il valore-$p$ associato ai dati, dandone la relativa interpretazione.

**Soluzione**.

1. Verifichiamo l'ipotesi $H_0 : \mu = 12.5$. Abbiamo $\overline{x} = 12.5125$. La statistica test è

$$st = \sqrt{8}\frac{\overline{x} - 12.5}{\sigma} = 0.236.$$

$H_0$ viene rifiutata al 5% se $|st| > z_{0.025} = 1.96$. Pertanto $H_0$ viene accettata: i dati sono consistenti con la gradazione media "storica".

2. 

$$\text{valore} - p = 2P(Z > |st|) = 2[1 - P(Z \leq 0.236)] \simeq 0.8142.$$

Il valore alto del valore $p$ indica un'ottimo accordo dei dati con $H_0$.