alcuni esercizi risolti o quasi di inferenza statistica I a.a. 2005/06

guido masarotto

23 maggio 2006

guido masarotto
facoltà di scienze statistiche
università di padova
e-mail: guido.masarotto@unipd.it

ISTRUZIONI PER L'USO

- La numerazione delle unità corrisponde a quella dei lucidi delle lezioni.
- Tavole dei quantili delle distribuzioni di interesse possono essere trovate in "formule e tavole" disponibile nella pagina del corso sul sito della Facoltà. Consiglio gli studenti a cercare di svolgere gli esercizi tenendo, in prima battuta, a disposizione solo questo documento e facendo solo successivamente ricorso a lucidi e libri. Infatti, durante l'esame scritto, è possibile consultare solamente "formule e tavole".
- Tutti gli esercizi "potrebbero" essere proposti durante un compito d'esame e, infatti, quasi tutti provengono da compiti d'esame proposti negli anni passati. In alcuni rari casi (ad esempio l'esercizio 55) la soluzione proposta richiede però più calcoli e grafici di quanti si possono svolgere durante il compito. Sono inseriti qui nella speranza di riuscire a chiarire meglio il discorso. Altre volte la soluzione presentata è schematica e non mostra tutti i dettagli di calcolo. Credo però sia buona pratica ricostruirli.
- Durante la soluzione di un esercizio può essere buona pratica seguire il seguente schema:

1. Descrivere la popolazione di riferimento (quanto è desumibile dal testo, ovvero, quando si fa riferimento ad un problema concreto -inventato o meno che sia); Ad esempio,

   La popolazione di riferimento è l'insieme di tutte le lastre di metallo che il processo produttivo che stiamo considerando potrebbe produrre quando è nello stato attuale. Si tratta di una popolazione virtuale e infinita.

2. Precisare il modello probalistico di riferimento. Ad esempio,

   Per condurre le analisi, supporremmo che le misure ottenute siano determinazioni indipendenti ed identicamente distribuite di una normale di media μ ignota e varianza σ^2 nota ed uguale a 0,01.

3. Nel caso si calcoli un intervallo di confidenza, precisare, prima di svolgere i calcoli, il livello e la formula utilizzata. Ad esempio,

   Un intervallo di confidenza per la media di livello $1-\alpha$ può essere calcolato come $\overline{y} \pm z_{1-\alpha/2}\sigma_0/\sqrt{n}$. Useremmo $\alpha = 0{,}05$.

4. Nel caso si conduca un test descrivere il sistema di ipotesi, la statistica test e la distribuzione sotto H_0. Ad esempio,

   Il sistema di ipotesi è

   $$H_0 : \mu = 14 \text{ verso } H_1 : \mu \neq 14.$$

   La statistica test che useremo è $\sqrt{n}(\overline{y} - 14)/\sigma_0$ che, nel caso sia vera l'ipotesi nulla si distribuisce come un normale standard.

5. Infine, non dimenticarsi mai di trarre le conclusioni di quanto fatto e di esprimerle nei termini del problema originale. Ad esempio,

   Le analisi condotte suggeriscono che il processo produttivo non era ben tarato nel momento in cui sono stati rilevati i dati.

Seguire lo schema infatti dovrebbe aiutarvi nella soluzione e sopratutto spingervi a rivedere le cose che vi sono risultate più "esoteriche". Non sarebbe male inoltre che teneste presente questo schema anche durante la prova d'esame.

Unità B

ESERCIZIO 1. In Inghilterra, dal 1200 fino a quando le monete sono state coniate in metallo prezioso, l'onestà del *Master of Mint* (il responsabile/appaltatore della zecca reale) è stata controllata mediante un controllo campionario, la cosidetta *Trial of Pyx* (*pyx* è un recipiente sacro). Il *Master of Mint*, nel caso la prova indicasse la sua disonestà, era soggetto a conseguenze non del tutto "gradevoli" (in alcuni secoli anche la condanna a morte).
I dettagli per le ghinee d'oro nel 1799 erano:
- il peso nominale di una ghinea era di 128 grani (360 grani fanno un oncia);
- 100 ghinee venivano state estratte casualmente tra tutte quelle prodotte durante l'intero anno (due funzionari reali si recavano in giorni scelti casualmente disseminati durante tutto l'anno presso la zecca e sceglievano a caso una o più monete della produzione di quel giorno);
- le ghinee "estratte" venivano man mano conservate nel *pyx*;
- alla fine dell'anno il recipiente con le 100 ghinee veniva pesato;
- il *Master of Mint* "passava" la *Trial of Pyx* se il peso delle ghinee "estratte" era uguale al peso atteso più o meno 1/400 del peso atteso stesso.

Si ritiene che con la tecnologia dell'epoca il peso di ogni singola moneta si distribuisse come una normale di media controllabile dal *Master of Mint* e varianza unitaria. Inoltre, i pesi reali di differenti monete possono essere considerati indipendenti.
(a) Calcolare la probabilità che un *Master of Mint* onesto sopravvivesse alla *Trial of Pyx*.
(b) Calcolare la probabilità che un *Master of Mint* disonesto e che avesse deciso di rubare mediamente 0,3 grani d'oro per ogni ghinea prodotta venisse scoperto.
(c) Implicitamente quale test statistico utilizzava il re per verificare l'onesta del suo *Master of Mint*?
(d) Nel contesto del test delineato al punto precedente che cosa sono e come si chiamano le due probabilità calcolate ai primi due punti dell'esercizio?

Schema di soluzione.

Poniamo

$$\begin{aligned}
n &= 100 = \text{numero ghinee pesate},\\
y_i &= \text{peso della i-sima ghinea}, i = 1, \dots, n,\\
s &= \sum_{i=1}^{n} y_i = \text{peso totale delle delle ghinee estratte},
\end{aligned}$$

$$\begin{aligned}
\overline{y} &= \frac{s}{n} = \text{peso medio delle ghinee estratte},\\
\mu &= \text{peso medio delle ghinee fissato dal } \textit{Master of Mint}.\\
\mu_0 &= 128 = \text{peso "nominale" di una ghinea}\\
\mu_1 &= 127{,}7 = \text{peso medio di una ghinea se il } \textit{Master}\\
&\qquad \text{cerca di "rubare" 0,3 grani d'oro per moneta}
\end{aligned}$$

Il *Master* passa la prova se

$$n\mu_0 - \frac{n\mu_0}{400} \leq s \leq n\mu_0 + \frac{n\mu_0}{400}.$$

Dividendo per n[1] possiamo riscrivere la "regola" utilizzata dal re come

"se $|\overline{y} - \mu_0| \leq \frac{\mu_0}{400}$ allora il *Master of Mint* viene dichiarato onesto".

Ricordiamoci che sappiamo che

$$\overline{y} \sim N\left(\mu, \frac{1}{n}\right) \quad \Leftrightarrow \quad \sqrt{n}(\overline{y} - \mu) \sim N(0,1). \tag{1}$$

Per rispondere alle prime due domande dobbiamo calcolare

P(il *master* venga dichiarato onesto quando fissa la media a μ)

ponendo nel caso della prima domanda $\mu = \mu_0$ e nel caso della seconda $\mu = \mu_1$. Ci conviene quindi calcolare la probabilità di sopra una volta per tutte per un valore di μ qualsiasi:

$$\begin{aligned}
&\text{P(il } \textit{Master} \text{ viene dichiarato onesto se } \mu \text{ è la media)} =\\
&= \mathrm{P}\left(\mu_0 - \frac{\mu_0}{400} \leq \overline{y} \leq \mu_0 + \frac{\mu_0}{400}\right) =\\
&= \mathrm{P}\left(\mu_0 - \mu - \frac{\mu_0}{400} \leq \overline{y} - \mu \leq \mu_0 - \mu + \frac{\mu_0}{400}\right) =\\
&= \mathrm{P}\left(\sqrt{n}\left(\mu_0 - \mu - \frac{\mu_0}{400}\right) \leq \sqrt{n}(\overline{y} - \mu) \leq \sqrt{n}\left(\mu_0 - \mu + \frac{\mu_0}{400}\right)\right) =\\
&= \mathrm{P}\left(\sqrt{n}\left(\mu_0 - \mu - \frac{\mu_0}{400}\right) \leq N(0,1) \leq \sqrt{n}\left(\mu_0 - \mu + \frac{\mu_0}{400}\right)\right) =\\
&= \Phi\left(\sqrt{n}\left(\mu_0 - \mu + \frac{\mu_0}{400}\right)\right) - \Phi\left(\sqrt{n}\left(\mu_0 - \mu - \frac{\mu_0}{400}\right)\right) =
\end{aligned}$$

Alla terza riga ho diviso per lo scarto quadratico medio di $\overline{y}$[2] in maniera tale da poter utilizzare la (1) nei passaggi successivi. Nell'ultima riga $\Phi(\cdot)$ indica al solito la funzione di ripartizione di una $N(0,1)$.

[1] visto che a lezione abbiamo lavorato con le "medie" e non con i totali continuo a lavorare così
[2] ovvero per $1/\sqrt{n}$

(a) Un *Master* onesto cerca di fissare μ uguale μ_0. Sperando (per la sua testa!) che non sbagli, troviamo

$$\begin{aligned} &\mathrm{P}(\text{un } M. \text{ onesto viene dichiarato onesto}) = \\ &= \Phi\left(\frac{128\sqrt{100}}{400}\right) - \Phi\left(-\frac{128\sqrt{100}}{400}\right) = \Phi(3{,}2) - \Phi(-3{,}2) = 2\Phi(3{,}2) - 1. \end{aligned}$$

L'ultimo passaggio è una conseguenza del fatto che la simmetria della $N(0,1)$ ci permette di scrivere

$$\Phi(-x) = 1 - \Phi(x).$$

Da una tabella dei quantili di una normale standard possiamo osservare che

$$3{,}2 > 3{,}09 = \text{quantile } 0{,}999 \text{ di una } N(0,1).$$

Ma allora

$$\Phi(3{,}2) = \mathrm{P}(N(0,1) \leq 3{,}2) > \mathrm{P}(N(0,1) \leq 3{,}09) = 0{,}999$$

e quindi la probabilità cercata è maggiore di $2 \times 0.999 - 1 = 0{,}998$, ovvero un *Master of Mint* onesto rischia, al più, "di perdere la sua testa" una volta ogni 500 anni.
Utilizzando R per calcolare la probabilità troviamo

```
> 2*pnorm(3.2)-1
[1] 0.9986257
```

(b) In questo caso μ è posto dal *Master of Mint* uguale a $\mu_1 = \mu_0 - 0,3 = 127{,}7$. Ovvero $\mu_0 - \mu = 0{,}3$. Quindi

$$\begin{aligned} &\mathrm{P}(\text{il } \mathit{Master} \text{ viene dichiarato onesto}) = \\ &= \Phi\left(\sqrt{100}\left(0{,}3 + \frac{128}{400}\right)\right) - \Phi\left(\sqrt{100}\left(0{,}3 - \frac{128}{400}\right)\right) = \\ &= \Phi(6{,}2) - \Phi(-0{,}2) \end{aligned}$$

Sappiamo che

$$\Phi(6{,}2) \approx 1.$$

Inoltre, utilizzando la tabella dei quantili della normale e ricordandoci della simmetria della distribuzione, troviamo

$$\Phi(-0{,}2) = 1 - \Phi(0{,}2) \approx 1 - 0{,}58 = 0{,}42$$

e quindi

$$\mathrm{P}(\mathit{Master} \text{ disonesto la faccia franca}) = \Phi(6{,}2) - \Phi(-0{,}2) \approx 1 - 0{,}42 = 0{,}58.$$

In R


```
> pnorm(6.2)-pnorm(-0.2)
[1] 0.5792597
```

La probabilità che "venga scoperto" è perciò, approssimativamente, $1 - 0{,}58 = 0{,}42$.
(c) Il re vuole verificare il sistema di ipotesi

$$\begin{cases} H_0 : \mathit{Master} \text{ onesto} \\ H_1 : \mathit{Master} \text{ disonesto} \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Il contesto e il test utilizzato sono quelli sulla media di una normale di varianza nota considerato nella seconda parte dell'unità B. Quello che sui lucidi è indicato con h[3] è in questo caso posto uguale a $\sqrt{n}\mu_0/400 = 3{,}2$. Questo, per quando visto al punto (a), garantisce che

$$0{,}998 < \mathrm{P}(\text{accettare } H_0 \text{ quando } H_0 \text{ è vera}) < 1.$$

Il livello di significatività del test adottato dal re quindi è inferiore a 0.002.
(d) Come appena notato la probabilità calcolata al punto (a) è

$$\mathrm{P}(\text{accettare } H_0 \text{ quando } H_0 \text{ è vera})$$

che possiamo anche indicare come la probabilità di non commettere un errore di I tipo.
Al punto (b) dell'esercizio abbiamo viceversa calcolato la probabilità che il test non commetta un errore di II tipo quando la vera media è uguale a 127,7.
Si osservi inoltre che se $\gamma(\mu)$ indica la funzione di potenza allora

$$\begin{aligned} &\text{probabilità punto (a)} = 1 - \gamma(128) \\ &\text{probabilità punto (a)} = \gamma(127{,}7) \end{aligned}$$

Esercizio 2. Si assuma che la pressione sistolica media di un adulto sano sia 120 (mm Hg) e lo scarto quadratico medio 5,6.
Assumendo che la pressione sia indipendente tra gli individui e che abbia distribuzione normale, calcolare la probabilità che
(a) selezionando un individuo sano a caso questi abbia pressione superiore 125;
(b) scegliendo a caso 4 individui, la media della loro pressione sistolica sia superiore a 125;
(c) scegliendo a caso 25 individui, la media della loro pressione sistolica sia superiore a 125;
Inoltre esprimere in forma verbale quello che queste probabilità ci raccontano sulla media campionaria.

[3] la soglia con cui confrontare il valore della statistica test

Schema di soluzione.

Ricordiamo che se $y_1, \dots, y_n$ sono determinazioni indipendenti di una variabile casuale normale di media μ e varianza σ^2 allora

$$\overline{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \sim N(\mu, \sigma^2/n).$$

Le probabilità ricercate sono quindi, per $n = 1, 4, 25$,

$$\begin{aligned} P(\overline{y}_n > 125) &= P(N(120, 5{,}6^2/n) > 125) = 1 - P(N(120, 5{,}6^2/n) \leq 125) = \\ &= 1 - \Phi\left(\frac{125-120}{5{,}6/\sqrt{n}}\right) \end{aligned}$$

e quindi valgono[4]
− $n = 1$: $1 - \Phi(0{,}89) \approx 1 - 0{,}8125 = 0{,}1875$;
− $n = 4$: $1 - \Phi(1{,}79) \approx 1 - 0{,}963 = 0{,}037$;
− $n = 25$: $1 - \Phi(4{,}46) < 1 - 0{,}9999 = 0{,}0001$.
Queste probabilità ci raccontano come, all'aumentare della numerosità campionaria, diminuisca la probabilità che la media campionaria si "allontani troppo" dalla vera media.

Esercizio 3. Si supponga che la media e lo scarto quadratico del colesterolo in individui sani tra i 18 e i 25 valgano, rispettivamente, 150 e 25.
Calcolare, almeno approssimativamente, la probabilità che la media dei valori rilevati indipendentemente su 100 individui estratti a caso sia compresa tra 149 e 151. Ricalcolare la medesima probabilità supponendo che siano state estratte 1000 e 10000 persone. Infine, esprimere verbalmente quello che le probabilità calcolate ci raccontano sulle proprietà della media aritmetica.

Schema di soluzione.

Il calcolo esatto non è possibile visto che non viene data la distribuzione del colesterolo (ma solo i suoi primi momenti).
Le numerosità da considerare ($n = 100, 1000, 10000$) sono però sufficentemente elevate da permettere l'utilizzo del teorema del limite centrale. Calcoleremo le probabilità richieste supponendo quindi, che almeno approssimativamente,

$$\frac{\overline{y}_n - 150}{25/\sqrt{n}} \sim N(0,1)$$

dove, al solito, $\overline{y}_n$ è la media di $y_1, \dots, y_n$.

[4]approssimativamente, uso deliberatamente solo la tavola dei quantili interpolando "ad occhio"; lo studente provi a rifare i calcoli utilizzando R.

Le probabilità richieste possono essere calcolate come

$$\begin{aligned} P(149 \leq \overline{y}_n \leq 151) &= P\left(\frac{149-150}{25/\sqrt{n}} \leq \frac{\overline{y}_n - 150}{25/\sqrt{n}} \leq \frac{151-150}{25/\sqrt{n}}\right) \approx \\ &\approx P\left(\frac{149-150}{25/\sqrt{n}} \leq N(0,1) \leq \frac{151-150}{25/\sqrt{n}}\right) = \\ &= \Phi\left(\frac{\sqrt{n}}{25}\right) - \Phi\left(-\frac{\sqrt{n}}{25}\right) \end{aligned}$$

e valgono[5]
− $n = 100$: $\Phi(0{,}4) - \Phi(-0{,}4) \approx 0{,}31$;
− $n = 1000$: $\Phi(1{,}26) - \Phi(-1{,}26) \approx 0{,}79$;
− $n = 1000$: $\Phi(4) - \Phi(-4) > 0{,}9998$.
Le probabilità appena calcolate fanno vedere come la probabilità che la media campionaria "commetta" un errore superiore a 1 diminuisce all'aumentare della numerosità campionaria. Ovvero, esemplifica come al crescere di n la distribuzione campionaria della media campionaria si "concentri" intorno al vero parametro.

Esercizio 4. In una recente ispezione in un ospedale, è stata misurato il rumore (in decibel) in 74 corridoi e stanze di ricovero. La media delle misure ottenute è stata 61,3 e lo scarto quadratico medio campionario 7,8. Calcolare un intervallo di confidenza al 90% per la media dei decibel a cui sono esposti i ricoverati dell'ospedale considerato.

Schema di soluzione.

La distribuzione delle singole osservazioni non è data. Però, vista la numerosità campionaria, possiamo, dal teorema del limite centrale, calcolare degli intervalli di confidenza approssimati per la media. Poniamo

$$\begin{aligned} n &= 74 = \text{numero delle osservazioni}, \\ \overline{y} &= 63{,}3 = \text{media campionaria}, \\ s &= 7{,}8 = \text{scarto quadratico medio campionario}, \\ \alpha &= 0{,}1 = \text{probabilità che l'intervallo non includa il vero valore}, \\ z_{1-\alpha/2} &= 1{,}64 = \text{quantile “appropriato” di una normale standard}. \end{aligned}$$

Allora l'intervallo di confidenza richiesto è

$$\overline{y} \pm z_{1-\alpha/2}\frac{s}{\sqrt{n}} = 63{,}3 \pm 1{,}64\frac{7{,}8}{\sqrt{74}} \approx [61{,}8\ ;\ 64{,}8].$$

[5]calcolate "a occhio" dalla tavola dei quantili di una normale standard.

Quindi, con alta probabilità (approssimativamente 90%), la media del rumore a cui sono esposti i pazienti (e i medici e gli infermieri) dell'ospedale considerato è compresa tra i 61,8 e i 64,8 decibel.

Esercizio 5. Per misurare la concentrazione di Pb (in $\mu g g^{-1}$) si procede nella seguente maniera:

(i) il materiale originario viene diviso in n *pezzettini*;
(ii) su ciascun *pezzettino* viene misurata la concentrazione utilizzando uno strumento appropriato;
(iii) la stima della concentrazione di Pb viene infine calcolata facendo la media aritmetica delle n misure ottenute.

Sapendo che gli errori commessi dallo strumento utilizzato si distribuiscono (almeno approssimativamente) come delle normali di media zero e scarto quadratico medio 0,2, dire quanto deve essere grande n affinchè almeno 9 volte su 10

$$|(\text{stima della concentrazione}) - (\text{vera concentrazione})| < 0{,}05.$$

Schema di soluzione.

Poniamo

$$\begin{aligned}
\mu &= \text{vera concentrazione} \\
y_i &= \text{misura sull'i-simo pezzettino} \\
\overline{y} &= \sum_{i=1}^{n} y_i/n = \text{stima concentrazione} \\
\sigma &= 0{,}2 = \text{scarto quadratico medio dell'errore di una singola misura} \\
\delta &= 0{,}05 = \text{massimo errore da commettere 9 volte su 10} \\
1-\alpha &= 0{,}9 = \text{probabilità desiderata di un errore inferiore a } \delta.
\end{aligned}$$

Quello che viene richiesto è di determinare n in maniera tale che

$$\mathrm{P}(|\overline{y} - \mu| < \delta) = 1 - \alpha.$$

Ma allora, per definizione di intervallo di confidenza,

$$\overline{y} \pm \delta$$

è un intervallo di confidenza di livello $1-\alpha$ per μ. Sappiamo che questi, nel presente contesto, sono del tipo

$$\overline{y} \pm \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}}.$$

Quindi, affinchè sia soddisfatta la condizione desiderata dobbiamo scegliere n in maniera tale che

$$\frac{z_{1-\alpha/2}\sigma}{\sqrt{n}} = \delta.$$

Risolvendo in n questa equazione troviamo

$$n = \left(\frac{z_{1-\alpha/2}\sigma}{\delta}\right)^2.$$

Con i dati del problema, osservando nella tabella dei quantili di una $N(0,1)$ che $z_{0{,}95} = 1{,}645$, troviamo

$$n = \left(\frac{1{,}645 \times 0{,}2}{0{,}05}\right)^2 \approx 43{,}3.$$

Ovviamente un numero frazionale di misure non possiamo farle. Osservando che più piccolo è n più è piccolo l'intervallo possiamo scegliere $n = 44$. Infatti, così facendo ci garantiamo che

$$\mathrm{P}(|\overline{y} - \mu| < \delta) \geq 1 - \alpha.$$

Esercizio 6. Per misurare un certo parametro ematico sono a disposizione due diverse metodiche, chiamiamole A e B. Sapendo che:

(i) le misure prodotte da ambedue le metodiche hanno media μ dove μ indica il vero valore del parametro[6] che si sta misurando;
(ii) la varianza degli errori della seconda misura è 10 volte la varianza degli errori della prima misura;
(iii) il costo (reagenti, tempo dell'operatore,...) di una misura con la prima metodica è 8 volte quello di una misura con la seconda metodica;
(iv) è possibile, per ambedue le metodiche, dividere il campione di sangue di un individuo in n "campioncini" e utilizzare la media delle misure ottenute nei vari campioncini per stimare il vero valore del parametro.

si dica se è più conveniente utilizzare
- la metodica A o
- la metodica B in questo caso utilizzando la possibilità (iv) per ottenere misurazioni con la stessa varianza.

Schema di soluzione.

La varianza della media di n misure è

$$\frac{\text{varianza singola misura}}{n}.$$

[6]che ovviamente dipende dall'individuo e dal momento del prelievo di sangue.

Quindi per ottenere stime con la stessa variabilità dobbiamo, per ogni misura con la metodica A, fare 10 misure con la metodica B. Conviene quindi utilizzare A visto che il suo costo è "solo" 8 quello di B.

Esercizio 7. Vi recate all'aeroporto di Venezia per prendere un volo per Boston. La durata prevista del volo è di 8 ore. Al *check-in* oltre alla carta di imbarco vi danno un depliant patinato pieno di foto e di informazioni sull'aeromobile su cui volerete. In fondo ad una pagina, scritto con un carattere piccolo piccolo, trovate la seguente frase:

> *La distribuzione delle ore di funzionamento senza guasti di una componente fondamentale del motore ha densità*
>
> $$f(x;\vartheta) = \begin{cases} 0 & se\ x < 0 \\ \vartheta e^{-\vartheta x} & se\ x \geq 0 \end{cases}.$$
>
> *L'esatto valore di ϑ non è noto ma con i dati in nostro possesso possiamo dire che* [0,00001 ; 0,09] *è un intervallo di confidenza per ϑ con probabilità di copertura pari a* 0,99.

Vi imbarcate lo stesso?

Schema di soluzione.

A noi ovviamente interessa che l'aereo arrivi a Boston senza guasti ovvero che, se indichiamo con T il tempo di funzionamento di quella componente del motore,

$$\pi(\vartheta) = \mathrm{P}(T \geq 8) = \int_8^{+\infty} f(x;\vartheta)dx = e^{-\vartheta 8}$$

sia sufficentemente grande.
E' facile vedere che $\pi(\vartheta)$ è una funzione monotona decrescente in ϑ. Quindi

$$\mathrm{P}(0{,}0001 \leq \vartheta \leq 0{,}09) = 0{,}99$$

implica che

$$\mathrm{P}(e^{-0{,}09\times 8} \leq e^{-\vartheta 8} \leq e^{-0{,}00001}) = 0{,}99.$$

Ovvero, quello che ci sta raccontando il depliant è che i dati disponibili per l'azienda indicano che con alta probabilità (99%) la probabilità di non avere un guasto è compresa tra

$$e^{-0{,}09\times 8} \approx 0{,}4867$$

e

$$e^{-0{,}00001\times 8} \approx 0{,}9999.$$

A questo punto decidete voi cosa fare. Ma la risposta più simpatica, quando ho proposto l'esercizio in un compito d'esame, è stata di una studentessa che ha scritto

> *Ho deciso di prendere il Titanic. E' più sicuro. E poi, chissà, con tutti quei principi e conti...*

Esercizio 8. . Siano $y_1, \ldots, y_n$ delle determinazioni indipendenti di una variabile casuale normale di media μ e varianza σ_0^2 nota e si considerino le seguenti tre situazioni:
(i) $n = 5$, $\sigma_0 = 10$, $\overline{y} = 1$,
(ii) $n = 1000$, $\sigma_0 = 10$, $\overline{y} = 1$,
(iii) $n = 5$, $\sigma_0 = 0{,}1$, $\overline{y} = 1$
dove, al solito, $\overline{y}$ indica la media campionaria delle n osservazioni.
In tutti e tre i casi
(a) si calcoli un intervallo di confidenza al 95% per la vera media;
(b) Si consideri il sistema di ipotesi

$$\mathrm{H}_0 : \mu = 0 \text{ verso } \mathrm{H}_1 : \mu \neq 0$$

e si dica se l'usuale test sulla media di una variabile casuale normale di varianza nota porta all'accettazione o al rifiuto dell'ipotesi nulla quando il livello di significatività viene posto uguale a 0,01;
(c) infine, si spieghi intuitivamente perchè il test suggerisce conclusioni differenti nonostante in tutte e tre le situazioni la media campionaria, ovvero la stima della media, sia uguale.

Schema di soluzione.

(a) L'intervallo di confidenza richiesto è

$$\overline{y} \pm z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}$$

dove $1-\alpha$ indica il livello di copertura richiesto e z_p è il quantile-p di una normale standard. Nel nostro caso

$$1 - \alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow z_{1-\alpha/2} = z_{0{,}975} = 1{,}96.$$

Nei tre casi quindi otteniamo
(i) $1 \pm 1{,}96 \times 10/\sqrt{5} = [-7{,}76; 9{,}76]$;
(ii) $1 \pm 1{,}96 \times 10/\sqrt{1000} = [0{,}38; 1{,}62]$;
(iii) $1 \pm 1{,}96 \times 0{,}1/\sqrt{5} = [0{,}91; 1{,}09]$.

(b) La statistica test è

$$z = \frac{\sqrt{n}\overline{y}}{\sigma_0}$$

e il test suggerisce di accettare l'ipotesi nulla quando

$$z_{1-\alpha/2} \le z \le z_{1-\alpha/2}.$$

Osserviamo che

$$\alpha = 0{,}01 \Rightarrow 1 - \alpha/2 = 0{,}995 \Rightarrow z_{1-\alpha/2} = z_{0{,}995} = 2{,}58.$$

Nelle tre situazioni descritte troviamo
(i) $z = \sqrt{5}/10 \approx 0{,}22$ e quindi accettiamo H_0;
(ii) $z = \sqrt{1000}/10 \approx 3{,}16$ e quindi rifiutiamo H_0;
(iii) $z = \sqrt{5}/0{,}1 \approx 22{,}36$ e quindi rifiutiamo H_0.
(c) Le conclusioni del test sono differenti nei tre casi. Si osservi tra l'altro che nulla cambierebbe se si considerassero valori di α differenti (ad esempio, $\alpha = 0{,}05$ o $\alpha = 0{,}1$).
In tutti e tre i casi la differenza tra la stima della media, $\overline{y}$, e la media ipotizzata dall'ipotesi nulla, 0, è uguale. La differenza infatti vale infatti sempre 1.
Nei tre casi è differente però la precisione con cui $\overline{y}$ stima la vera media. Infatti, lo scarto quadratico medio di $\overline{y}$ vale, come sappiamo, $\sigma_0/\sqrt{n}$ e quindi è tanto più piccolo quanto più σ_0 è piccolo e/o n è grande.
Non è quindi sorprendente che la stessa differenza possa risultare ancora "compatibile" con H_0 quando l'errore di stima è grande ma "incompatibile" con H_0 quando l'errore di stima è piccolo.
Del resto questo è anche quello che ci raccontano i tre intervalli di confidenza: nel primo ma non negli altri il valore $\mu = 0$ è incluso, ovvero, con la precisione con cui è stimato μ nel primo caso non possiamo escludere dai "valori plausibili" per μ il valore $\mu = 0$; viceversa, questo lo possiamo fare negli altri due casi.

Esercizio 9. Siano $(y_1, x_1), \ldots, (y_n, x_n)$ n determinazioni indipendenti tratte da una variabile casuale bivariata (X, Y) e si indichi con r_n in coefficiente di correlazione calcolato con queste osservazioni. Si indichi viceversa con ρ il coefficiente di correlazione esistente tra (Y, X) (ovvero il coefficiente di correlazione "nella popolazione").
(a) Sotto ipotesi deboli, che però non precisiamo, è possibile far vedere che quando ρ è uguale a zero la distribuzione di

$$z_n = \frac{r_n\sqrt{n-2}}{\sqrt{1-r_n^2}}$$

può essere approssimata, se n è sufficientemente grande, da una distribuzione normale standard[7]. Utilizzare questo risultato per costruire un test per il sistema d'ipotesi

$$\begin{aligned} H_0 &: \rho = 0 \\ H_1 &: \rho \neq 0 \end{aligned}$$

ovvero per verificare l'ipotesi che nella popolazione non esista correlazione.
(b) Si supponga che $n = 100$ e che per certi dati sia risultato $r_n = 0{,}17$. A quali conclusione arrivate utilizzando il test delineato al punto precedente?
(c) Si supponga ora che $n = 1000$ e che il valore osservato di r_n continui ad essere 0,17. Applicando il test ora visto quale sarebbe la risposta?
(d) Spiegare il motivo delle eventuali differenze trovate nei due punti precedenti.
Nota. *L'esercizio da un lato vuole fornire uno strumento utile (come verificare la presenza di correlazione tra due variabili) e dall'altro illustrare come conoscendo una statistica test appropriata e conoscendone la distribuzione campionaria "sotto" l'ipotesi nulla si è in grado di costruire autonomamente un test statistico.*

Schema di soluzione.

(a) Per il sistema d'ipotesi dato è intuitivamente plausibile pensare di utilizzare r_n come statistica test. Valori di r_n troppo lontani da zero (positivi o negativi) saranno ovviamente da interpretare come evidenza contro H_0.
Si ponga ora $f(x) = x/\sqrt{1-x^2}$ e si osservi che
(i) $f(0) = 0$;
(ii) $f(x) = -f(-x)$ ovvero che la funzione, come si usa dire, è dispari;
(iii) $f(x)$ è monotona crescente se $-1 < x < 1$; infatti, derivando troviamo

$$\frac{df(x)}{dx} = \frac{1}{\sqrt{1-x^2}} + \frac{x^2}{(1-x^2)^{\frac{3}{2}}} > 0 \text{ se } -1 < x < 1.$$

Quindi un test che rifiuta per $|r_n|$ troppo grande è equivalente ad un test che rifiuta per $|z_n|$ troppo grande. Possiamo perciò "ridefinire" la statistica test decidendo di utilizzare z_n al posto di r_n. Fatta questa scelta la meccanica del test diventa quella descritta nelle unità B e C con riferimento alla media della normale (con varianza nota) e alla probabilità di successo di una binomiale. Infatti la distribuzione sotto H_0 della statistica prescelta è anche in questo caso (solo asintoticamente) una $N(0, 1)$.
(b) Calcoliamo

$$z_n = \frac{r_n\sqrt{n-2}}{\sqrt{1-r_n^2}} = \frac{0{,}17\sqrt{100-2}}{\sqrt{1-0{,}17^2}} \approx 1{,}71.$$

[7] l'approssimazione è considerata ragionevole se $n > 50$.

Questo valore va confrontato con i valori "attesi" per una $N(0,1)$. Utilizzando un test con un livello di significatività prefissato e pari ad α il valore osservato va comparato con i quantili $1-\alpha/2$ di una normale standard. Alcuni valori "canonici" sono:

$$\alpha = 0{,}1 \Rightarrow z_{1-\alpha/2} = z_{0{,}95} = 1{,}64 \Rightarrow \text{ rifiuto H}_0 \tag{2}$$
$$\alpha = 0{,}05 \Rightarrow z_{1-\alpha/2} = z_{0{,}975} = 1{,}96 \Rightarrow \text{ accetto H}_0 \tag{3}$$
$$\alpha = 0{,}01 \Rightarrow z_{1-\alpha/2} = z_{0{,}995} = 2{,}58 \Rightarrow \text{ accetto H}_0 \tag{4}$$
$$\tag{5}$$

Come si vede siamo in una zona ai "confini della significatività". La conclusione può essere una dubbiosa accettazione di H_0.
(c) In questo caso

$$z_n = \frac{r_n\sqrt{n-2}}{\sqrt{1-r_n^2}} = \frac{0{,}17\sqrt{1000-2}}{\sqrt{1-0{,}17^2}} \approx 5{,}45.$$

In questo caso, per qualsiasi α ragionevole, il test ci suggerisce di rifiutare H_0.
(d) Si veda l'esercizio 8.

Esercizio 10. Rispondere alle seguenti domande spiegando brevemente il perchè delle risposte.
(a) Se un test statistico ci porta ad accettare l'ipotesi nulla, possiamo considerare dimostrato che l'ipotesi nulla è vera?
(b) Se un test statistico ci porta a rifiutare l'ipotesi nulla, possiamo considerare dimostrato che l'ipotesi nulla è falsa?
(c) All'aumentare dell livello di copertura "$1-\alpha$" aumenta anche l'ampiezza dell'intervallo di confidenza sulla media di una normale con varianza nota?
(d) E' vero che in un test statistico,

$$\mathrm{P}(\text{errore di I tipo}) = 1 - \mathrm{P}(\text{errore di II tipo})?$$

(e) Si consideri il test sulla media di una distribuzione normale discusso nell'Unità B. Si supponga che il test venga applicato in un momento in cui il processo è in controllo (la media degli spessori è uguale a $14mm$). Se campionassimo non 5 ma 100 lastre, la probabilità di commettere un errore di primo tipo diminuirebbe?

Schema di soluzione.

(a) No. Purtroppo, per effetto del caso potremmo essere incappati in un errore di II tipo (accettare H_0 quando H_0 è falsa). Quello che possiamo dire è che le informazioni contenute nei dati non permettono di rifiutare l'ipotesi nulla. E' un po' come in un tribunale. Se un indiziato viene dichiarato non colpevole non possiamo dire con certezza che non lo è. Quello che però possiamo dire è che l'evidenza processuale è compatibile con la sua innocenza.
(b) No. Purtroppo, per effetto del caso potremmo essere incappati in un errore di I tipo (rifiutare H_0 quando H_0 è vera). La costruzione che abbiamo visto permette di controllare non di eliminare la probabilità di commettere questo errore. Ritornando al paragone con un tribunale, è un po' come quando viene condannato un innocente: l'evidenza processuale (nel nostro caso i dati) sono contro di lui ma non era stato lui a commettere il "delitto".
(c) Si. L'intervallo di confidenza è

$$\overline{y} \pm z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}$$

e ovviamente, al crescere è $1-\alpha$ cresce $z_{1-\alpha/2}$ e quindi l'intervallo si allarga. Lasciando perdere considerazioni troppo tecniche per il livello a qui siamo, è quello che capita praticamente sempre. Del resto sembra scontato che un intervallo che contiene il 99% delle volte un parametro ignoto debba essere più lungo di un intervallo che contiene lo stesso parametro "solo" il 50% delle volte (almeno se sono tutti e due costruiti in maniera sensata!).
(d) No. Le probabilità dei due errori sono addirittura calcolate sotto due ipotesi differenti sul mondo (quella di I tipo supponendo H_0 vera, quella di II ipotizzando vera H_1).
(e) No. Nei test tipo quello sviluppato nell'unità B la probabilità di errore di I tipo è prefissata (e indicata con α). Non dipende dalla numerosità campionaria.

Esercizio 11. L'Istituto Superiore di Sanità ha stimato che le spese a carico del Sistema Sanitario Nazionale per la riabilitazione di un paziente che ha avuto un ictus è di 42372 euro. L'amministrazione di una ASL, per verificare se i costi nella ASL erano in linea con la media nazionale, ha raccolto le informazioni sul costo della riabilitazione di 64 pazienti. Il costo medio è risultato pari a 44143 euro con uno scarto quadratico medio (campionario) di 9156 euro.
(a) Determinare un intervallo di confidenza con livello di copertura 99% per la vera media dei costi nell'ASL considerata.
(b) La differenza tra il costo medio nazionale e il costo medio stimato nell'ASL è significativa per $\alpha = 0{,}01$?
(c) Le risposte precedenti sono coerenti tra loro? Ovvero ci raccontano la stessa storia?

Schema di soluzione.

Il testo dell'esercizio non menziona la normalità dei costi ed inoltre siamo in presenza di uno scarto quadratico medio stimato. La numerosità è però tale che è possibile utilizzare le procedure per l'inferenza sulla medie basate sul teorema del limite centrale.
Sia per l'intervallo di confidenza che per il test ci serve il quantile-0,995 di una normale standard. Da una tavola dei quantili della distribuzione normale troviamo che vale 2,58.
(a) L'intervallo di confidenza è

$$\overline{y} \pm z_{1-\alpha/2}\frac{s}{\sqrt{n}} = 44143 \pm 2{,}58\frac{9156}{\sqrt{64}} = [41189{,}54\ ;\ 47096{,}46].$$

(b) La statistica test è

$$z = \frac{\overline{y} - \mu_0}{\dfrac{s}{\sqrt{n}}} = \frac{44143 - 42372}{\dfrac{9156}{\sqrt{64}}} \approx 1{,}55.$$

Per verificare se la differenza tra la stima della media e la media ipotizzata è significativa al 1% dobbiamo confrontare questo valore con il quantile-0,995 di una normale standard. Ovviamente

$$-2{,}58 = -z_{0{,}995} \leq z = 1{,}55 \leq z_{0{,}995} = 2{,}58$$

quindi la differenza osservata non è significativa al 1%.
(c) Certo. I risultati sono coerenti. Infatti
- l'intervallo di confidenza include la media nazionale al suo interno; ovvero, ci sta raccontando che con i dati disponibili, 42372 potrebbe anche essere la vera media dei costi nell'ASL considerata;
- il test dichiara "non significativa" la differenza tra la media campionaria osservata e la media nazionale; ovvero ci dice che non possiamo escludere che la 42372 sia il vero costo medio.

Unità C-D

Esercizio 12. Nella contea di *Chiare Acque* da anni viene condotta un'indagine per stimare la proporzione di castelli abitati da fate. L'indagine è campionaria: ogni anno, sono considerati 100 castelli estratti a caso e per ogni castello viene rilevato se ci abitano o meno delle fate.

1. Sapendo che quest'anno il numero di castelli con fate nel campione è risultato pari a 38, costruire un intervallo che includa la vera proporzione con una probabilità (almeno approssimativamente) uguale a 0,90.
2. In un vecchio libro, il duca di *Dolci Acque* ha scritto che circa il 50% dei castelli della contea di *Chiare Acque* è abitato da fate. Utilizzare un appropriato test statistico per verificare se i dati sono compatibili con questa affermazione.
3. La vicina contea di *Fresche Acque* vuole da quest'anno condurre una indagine analoga. Durante una riunione il ciambellano addetto ai castelli fatati dice: "*La nostra contea ha il doppio di castelli della contea di* Chiare Acque. *Quindi, affinchè la stima della nostra indagine abbia la stessa precisione dobbiamo ogni anno estrarre* 200 *castelli*". Siete d'accordo con il ciambellano?

Schema di soluzione.

Certamente la densità dei castelli, fatati o meno, è altissima nelle contee di *Chiare e Fresche Acque*. Quindi, o esattamente se l'estrazione è fatta con reintroduzione o approssimativamente se l'estrazione avviene senza reintroduzione, possiamo supporre che per quanto ci riguarda siano nel *Reame della Binomiale*. Ovvero, posto

$$\begin{aligned} y &= \text{numero castelli fatati estratti} \\ n &= \text{numero castelli estratti e ispezionati} \\ \vartheta &= \text{percentuale di castelli fatati nella contea} \\ \hat{\vartheta} &= \frac{y}{n} = \text{stima della percentuale di castelli fatati nella contea} \end{aligned}$$

assumeremmo che

$$y \sim Bi(n, \vartheta).$$

(a) La prima domanda richiede di calcolare, quando $y = 38$ e $n = 100$ un intervallo di confidenza che includa con probabilità uguale (almeno approssimativamente) a 90% il vero valore di ϑ. Sappiamo dall'unità C che una soluzione è l'intervallo

$$\hat{\vartheta} \pm z_{0,95}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}}$$

che nel nostro caso diventa

$$0{,}38 \pm 1{,}645\sqrt{\frac{0{,}38(1-0{,}38)}{100}} \approx [0{,}30\ ;\ 0{,}46].$$

(b) La seconda domanda richiede di verificare, utilizzando gli stessi dati di prima, il sistema di ipotesi

$$\begin{cases} H_0 : \vartheta = 0{,}5 \ \ (= \vartheta_0) \\ H_1 : \vartheta \neq 0{,}5 \ \ (\neq \vartheta_0) \end{cases}$$

Calcoliamo la statistica test usuale

$$(\hat{\vartheta} - \vartheta_0)/\sqrt{\vartheta_0(1-\vartheta_0)/n} = (0{,}38 - 0{,}5)/\sqrt{0{,}5 \times 0{,}5/100} = -2{,}4.$$

Il valore ottenuto deve essere confrontato con i valori previsti da una $N(0,1)$, distribuzione che descrive i valori che ci attendiamo per la statistica test quando è vera H_0.

− *Livello di significatività prefissato.* Le probabilità di errore di I tipo (α) usualmente considerate in un approccio di questo tipo e i quantili corrispondenti sono

$$\begin{aligned} \alpha = 0{,}1 &\Rightarrow z_{1-\alpha/2} = 1{,}645 \\ \alpha = 0{,}05 &\Rightarrow z_{1-\alpha/2} = 1{,}960, \\ \alpha = 0{,}01 &\Rightarrow z_{1-\alpha/2} = 2{,}576; \end{aligned}$$

Il valore osservato per la statistica ci porterebbe quindi a rifiutare H_0 per $\alpha = 0{,}1$ e $\alpha = 0{,}05$ e ad accettare H_0 se fissiamo $\alpha = 0{,}01$. Risultati di questo tipo sono normalmente descritti come significativi (ma non altamente significativi) contro H_0.

− *p-value.* Può essere calcolato come

$$2\mathrm{P}(N(0,1) > 2{,}4).$$

Utilizzando R lo calcoliamo come

```
> 2*(1-pnorm(2.4))
[1] 0.01639507
```


Se abbiamo a disposizione solo una tabella dei quantili di una normale possiamo osservare, ad esempio, che

$$\begin{aligned} \mathrm{P}(N(0,1) \geq 2{,}326) &= 0{,}01 \\ \mathrm{P}(N(0,1) \geq 2{,}576) &= 0{,}005 \end{aligned}$$

e, visto che ovviamente

$$\mathrm{P}(N(0,1) \geq 2{,}576) < \mathrm{P}(N(0,1) \geq 2{,}4) < \mathrm{P}(N(0,1) \geq 2{,}326)$$

concludere che

$$0{,}01 < (\text{livello significatività osservato}) < 0{,}02.$$

La interpretazione è la stessa di prima: i valori ottenuti fanno "sospettare parecchio" di H_0 ma non proprio "escluderla".
In definitiva i risultati suggeriscono che la situazione dei castelli fatati nella contea di *Chiare Acque* dovrebbe essere cambiata dagli anni del viaggio e del relativo libro del Duca di *Dolci Acque*. L'evidenza a questo proposito non è però fortissima.
(c) Il ciambellano è fuori strada. Quando lo incontreremo ricordiamoci di fargli osservare cose del tipo:

(i) la distribuzione dell'errore di stima della indagine dipende dalla percentuale di castelli fatati nella sua contea (ϑ) e dal numero di castelli "ispezionati" (n);

(ii) come conseguenza anche l'ampiezza degli intervalli di confidenza dipende solamente da $\hat{\vartheta}$ (e quindi indirettamente da ϑ), da n e ovviamente anche dalla copertura desiderata (α).

Nessuna traccia di dipendenza dal numero complessivo di castelli che, quindi, sembra irrilevante per scegliere un valore appropriato per n.
[**nota importante**] Le affermazioni precedenti dipendono in maniera cruciale dal tipo di campionamento adottato (e in questo caso ipotizzato). Sono vere nel caso di un campionamento con reintroduzione. Ma non nel caso di un campionamento senza reintroduzione[8]

Esercizio 13. Sia y una determinazione di una variabile casuale binomiale con numero di prove uguale a n e probabilità di successo uguale ϑ. Dire in quali di queste situazioni è possibile considerare l'approssimazione normale alla distribuzione binomiale:

[8] si pensi alla situazione in cui ci siano 100 castelli nella contea; se "campioniamo con reintroduzione 100 castelli" li prendiamo tutti e quindi non abbiamo nessun errore di stima; viceversa se ne "campioniamo con reintroduzione 100 castelli" ma ce ne sono 1000 nella contea qualche errore possiamo commetterlo; quindi la distribuzione dell'errore di stima nei due casi non può essere la stessa.

(a) $n = 10$ e $\vartheta = 0{,}05$;
(b) $n = 10$ e $\vartheta = 0{,}95$;
(c) $n = 100$ e $\vartheta = 0{,}05$;
(d) $n = 100$ e $\vartheta = 0{,}95$;
(e) $n = 10$ e $\vartheta = 0{,}5$;
(f) $n = 10000$ e $\vartheta = 0{,}9999$;
(g) $n = 100000$ e $\vartheta = 0{,}0001$;
(h) $n = 30$ e $\vartheta = 0{,}23$.

Schema di soluzione.

L'approssimazione normale alla distribuzione binomiale viene considerata ragionevole se[9]

$$n \min(\vartheta, 1-\vartheta) \geq 5.$$

Quindi, rispondendo si e no per le situazioni suggerite: (a) no; (b) no; (c) si; (d) si; (e) si; (f) no; (g) si; (h) si.

Esercizio 14.

In un recente rapporto la Polizia Stradale ha indicato che negli ultimi 80 incidenti del "sabato sera" con almeno un morto, 57 possono essere messi in relazione con un eccessivo consumo di alcool.
(a) Calcolare un intervallo di confidenza al 90% per la percentuale di incidenti mortali del "sabato sera" legati all'alcool.
(b) Riassumere, in forma verbale, quello che ci racconta l'intervallo.

Schema di soluzione.

(a) Siamo nel contesto di un campionamento binomiale (almeno se gli 80 incidenti possono essere considerati indipendenti). Poniamo,

$$\begin{aligned}\hat{\vartheta} &= \text{stima della proporzione di incidenti legati all'alcool} = \\ &= \frac{\text{numero di incidenti legati all'alcool}}{\text{numero totale di incidenti}} = \\ &= \frac{57}{80} = 0{,}7125\end{aligned}$$

e $\alpha = 0{,}1$.
L'intervallo di confidenza è

$$\hat{\vartheta} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}} = 0{,}7125 \pm 1{,}64\sqrt{\frac{0{,}7125 \times (1-0{,}7125)}{80}} \approx [0{,}63\ ;\ 0{,}80].$$

[9] se la quantità a sinistra è maggiore di 1 l'approssimazione diventa quasi accettabile; è comunque meglio limitarsi a considerare il caso scritto nel testo.

(b) Una possibile descrizione verbale è

> *La proporzione di incidenti legati ad un consumo eccessivo di alcool è compresa, con alta probabilità (circa 90%), tra il 63% e l'80%.*

Si noti che parliamo di circa 90% e non di 90% perchè l'intervallo è basato su alcune approssimazioni (la binomiale è approssimata da una normale con varianza è stimata).

Esercizio 15.

Un ricercatore ha formulato un modello teorico che prevede che il 40% di certe cellule dovrebbero godere di una particolare proprietà.
Due biologi conducono un esperimento per verificare il modello del loro collega. Ambedue isolano un certo numero di cellule del tipo considerato e poi contano, tra le cellule ottenute, quelle che godono della proprietà di interesse. I risultati sono riassunti nella seguente tabella.

	cellule considerate (n)	cellule con la proprietà di interesse (y)
primo biologo	20	10
secondo biologo	2000	1000

Assumendo che si possa far riferimento ad una distribuzione binomiale,
(a) verificare che utilizzando l'usuale test sulla probabilità di successo di una binomiale i due biologi arrivano a conclusioni differenti;
(b) spiegare verbalmente perchè questo accade nonostante i dati di ambedue i biologi portino ad una medesima stima della probabilità di trovare una cellula con la proprietà di interesse.
Al punto (a) rispondere calcolando (anche solo in maniera approssimata) e commentando il livello di significatività osservato.

Schema di soluzione.

(a) Il sistema d'ipotesi da verificare è

$$\mathrm{H}_0 : \vartheta = 0{,}4 \text{ verso } \mathrm{H}_1 : \vartheta \neq 0{,}4$$

dove ϑ indica la probabilità che una cellula del tipo considerato abbia la proprietà di interesse.
L'usuale statistica test è

$$z = \frac{\frac{y}{n} - 0{,}4}{\sqrt{\frac{0{,}4(1-0{,}4)}{n}}}$$

e il livello di significatività osservato è

$$P(N(0,1) \leq -z) + P(N(0,1) \geq z) = 2P(N(0,1) \geq z).$$

Con i dati dei due biologi otteniamo

- primo biologo: $z = (0{,}5-0{,}4)/\sqrt{0{,}4 \times 0{,}6/20} \approx 0{,}91$; consultando delle tavole dei quantili di una normale standard, troviamo che 0,91 è compreso tra il quantile 0,815 e il quantile 0,820 ovvero che

$$0{,}815 < P(N(0,1) < 0{,}91) < 0{,}825;$$

  perciò il livello di significatività osservato è compreso tra 0,36 e 0,37; il valore è grande; accettiamo H_0: i risultati sperimentali ottenuti dal primo biologo sono compatibili con quanto predetto dal modello teorico ;
- secondo biologo: $z = (0{,}5 - 0{,}4)/\sqrt{0{,}4 \times 0{,}6/2000} \approx 9{,}13$; questo valore è oltre tutti i quantili usualmente tabulati di una normale standard; in questo caso il livello di significatività osservato è ≈ 0; rifiutiamo quindi H_0: i dati del secondo biologo contratticono quanto predetto dal modello teorico.

(b) La stima della probabilità che un cellula del tipo considerato abbia la proprietà di interesse vale 0,5 in ambedue gli studi.
Nel primo studio sono state però considerate pochissime cellule (20). L'errore della stima è perciò piuttosto grande e tenendo conto di questo non risulta possibile escludere che la vera probabilità sia 0,4.
La numerosità delle cellule considerate dal secondo biologo è stata decisamente più grande (2000). Con questa numerosità otteniamo stime molto più precise della vera probabilità e, difatti, siamo in grado di escludere che possa essere 0,4.
Questa differenza può essere fatta toccare con mano se si calcolano, nei due casi, degli intervalli di confidenza. Ad esempio[10] degli intervalli di confidenza di livello 95% valgono in questo caso:

	da	a
dati del primo biologo	0,28	0,72
dati del secondo biologo	0,48	0,52

Il primo include 0,4 tra i "valori di ϑ plausibili", il secondo no.

ESERCIZIO 16. In una recente indagine condotta negli Stati Uniti è stato rilevato, tra le altre cose, che 40 delle 90 famiglie intervistate possedeva almeno una pistola o un'altra arma da fuoco.
Determinare un intervallo di confidenza al 95% per la proporzioni di famiglie statunitensi che possiede un'arma da fuoco.

[10] lo studente volenteroso lo verifichi.

Schema di soluzione.

Facciamo riferimento ad una distribuzione binomiale. Allora l'intervallo di confidenza è

$$\hat{\vartheta} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}} \approx 0{,}444 \pm 1{,}96\sqrt{\frac{0{,}444 \times (1-0{,}444)}{90}} = [0{,}34\ ;\ 0{,}55].$$

ESERCIZIO 17. L'American Automobile Association ha riportato che la percentuale di incidenti negli Stati Uniti attribuibili ad errori del guidatore è 54%. L'Automobile Club Italiano, sezione veneta, ha analizzato le cause di 100 incidenti in cui erano stati coinvolti suoi soci. Il numero di incidenti attribuibili al guidatore è risultato pari a 47. Sulla base di questi dati ha rilasciato un comunicato stampa in cui si dichiarava che

> *I guidatori veneti sono più bravi dei guidatore degli Stati Uniti. Infatti, mentre negli Stati Uniti, più del 50% degli incidenti è dovuto ad errori di guida, nel Veneto questa percentuale è inferiore al 50%.*

Questa dichiarazione è scorretta per almeno due motivi. Spiegare quali.
Nota. Tolta la percentuale del 54% riferita dall'AAA il resto dell'esercizio è inventato. Quindi non si offenda nessuno.

Schema di soluzione.

1° motivo. Il primo errore sta nel generalizzare a tutti i guidatori veneti delle conclusioni riferite ad incidenti in cui sono coinvolti soci ACI. Non tutti i guidatori sono soci ACI. E potrebbero esserci delle differenze tra i due gruppi (soci e non soci dell'ACI). Ad esempio, i soci ACI potrebbero essere le persone più attente/interessate alla guida e alle automobili e quindi, magari, anche più esperte.
2° motivo. Il secondo errore è che non si è tenuto conto del fatto che erano a disposizione solo informazioni campionarie.
Ad esempio, facendo riferimento alla binomiale, un intervallo di confidenza al 95% per la probabilità che un incidente in cui sia coinvolto un socio veneto dell'ACI è

$$0{,}47 \pm 1{,}96\sqrt{\frac{0{,}47 \times 0{,}53}{100}} = [0{,}37\ ;\ 0{,}57].$$

Quindi la proporzione di indidenti dovuti a causa umana potrebbe essere più grande del 50% e anche più grande dell'analoga proporzione trovata negli Stati Uniti.

Alternativamente, ma arrivando alla stessa conclusione, si osservi che i dati a disposizione "sono compatibili" con l'ipotesi che la percentuale di incidenti con causa umana sia nella popolazione da cui provengono i dati uguale a quella statunitense. Sempre facendo riferimento alla binomiale, la statistica test è infatti

$$z = \frac{0{,}47 - 0{,}54}{\sqrt{\dfrac{0{,}54 \times 0{,}46}{100}}} \approx -1{,}40.$$

Questo valore, da confrontare con quelli "possibili per una $N(0,1)$", non ci permette di rifiutare[11] l'ipotesi che la vera percentuale nel Veneto (per i soci ACI) valga 0,54.

ESERCIZIO 18. Prima di un referendum sono state intervistate 2605 persone estratte a caso tra gli aventi diritto al voto. Di questi, 1207 hanno dichiarato di non avere intenzione di andare a votare. Sulla base di questi dati quale delle seguenti affermazioni fareste:
(A) il *quorum* sarà certamente raggiunto;
(B) è molto plausibile che il *quorum* venga raggiunto;
(C) non posso concludere se il *quorum* verrà o non verrà raggiunto.
(D) è poco plausibile che il *quorum* venga raggiunto.
Rispondere utilizzando un intervallo di confidenza.

Schema di soluzione.

Possiamo pensare di essere nel contesto di un campionamento di tipo binomiale. Il parametro di interesse (la probabilità di successo della binomiale che indicheremo con ϑ) è in questo caso la percentuale di elettori che si recheranno a votare.
Scegliamo di utilizzare un intervallo di confidenza con probabilità di copertura pari al 99%.

$$\hat{\vartheta} = \frac{y}{n} = \frac{1207}{2605} \approx 0{,}463$$

$$\alpha = 0{,}01 \Rightarrow z_{1-\alpha/2} = 2{,}576$$

$$z_{1-\alpha/2}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}} \approx 2{,}576\sqrt{\frac{0{,}463(1-0{,}463)}{2605}} \approx 0{,}025$$

[11]discutelo per completare l'esercizio sia utilizzando un α fisso che calcolando il livello di significatività osservato.

$$\hat{\vartheta} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}} \approx 0{,}463 \pm 0{,}025 = [0{,}438\ ;\ 0{,}488]$$

L'intervallo di confidenza indica che i valori per la percentuale di votanti "plausibili" sulla base dei dati sono tutti più piccoli del 50%. Difficile dire con certezza che cosa accadrà. Però sembra poco plausibile sulla base di questi dati che il *quorum* venga raggiunto (affermazione D).

ESERCIZIO 19. Dall'agenzia che conduce le indagini di mercato per l'azienda in cui lavorate ricevete un rapporto contenente la seguente frase:

> "Sulla base di <numero illegibile> interviste telefoniche, possiamo dire che la percentuale di donne tra i 18 e i 25 anni interessate al nuovo prodotto che state per immettere sul mercato è compresa tra il 24% e il 32% con una probabilità pari al 90%."

Secondo voi, quante interviste telefoniche sono state fatte?

Schema di soluzione.

E' verosimile che sia stato utilizzato un intervallo di confidenza basato sulla binomiale del tipo

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

Il punto centrale di questo intervallo è $\hat{\vartheta}$. Quindi nel caso in esame deve essere

$$\hat{\vartheta} = \frac{0{,}24 + 0{,}32}{2} = 0{,}28.$$

Inoltre il livello di copertura dell'intervallo è 90%. Quindi,

$$\alpha = 0{,}1 \Rightarrow z_{1-\alpha/2} = 1{,}645.$$

Sfruttando la semiampiezza dell'intervallo dato possiamo scrivere

$$z_{1-\alpha/2}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}} = \frac{0{,}32 - 0{,}24}{2} = 0{,}04$$

dove l'unica incognita rimasta dopo le considerazioni precedenti è proprio n. Risolvendo quindi per n troviamo

$$n = \left(\frac{z_{1-\alpha/2}}{0{,}04}\right)^2 \hat{\vartheta}(1-\hat{\vartheta}) = \left(\frac{1{,}645}{0{,}04}\right)^2 0{,}28(1-0{,}28) \approx 350.$$

Esercizio 20. Un gruppo di medici vuole stimare l'efficacia di un protocollo di terapia recentemente proposto per curare una certa patologia. Ha quindi deciso di utilizzare il nuovo protocollo per i prossimi pazienti e di rilevare su di essi, dopo un tempo appropriato, il carattere dicotomico "guarito" o "non guarito".
Si indichi con n il numero di pazienti che "entreranno" nello studio e con y il numero di pazienti che guarirà. Si assuma inoltre che $y \sim Bi(n, \vartheta)$ dove ϑ denota la probabilità che un paziente trattato guarisca e si ponga

$$\hat{\vartheta} = \frac{y}{n}$$

(ovvero $\hat{\vartheta}$ è l'usuale stima di ϑ calcolata dai dati).
Si determini n in maniera tale che

$$\mathrm{P}\left(\left|\hat{\vartheta} - \vartheta\right| \leq 0{,}02\right) \geq 0{,}99.$$

Schema di soluzione.

Definiamo

$$\begin{aligned} \delta &= 0{,}02, \\ \alpha &= 0{,}01. \\ z_{1-\alpha/2} &= 2{,}576 = \begin{pmatrix} \text{percentile } 1-\alpha/2 \\ \text{di una } N(0,1) \end{pmatrix} \end{aligned}$$

Il problema chiede di determinare la numerosità campionaria (n) in maniera tale che l'errore di stima ($\hat{\vartheta} - \vartheta$) sia più piccolo, in valore assoluto, di δ con probabilità maggiore di $1 - \alpha$.
Ricordando quello che sappiamo sugli intervalli di confidenza per la probabilità di successo di una binomiale quanto richiesto accade se

$$z_{1-\alpha/2}\sqrt{\frac{\vartheta(1-\vartheta)}{n}} \leq \delta.$$

Isolando n nella disequazione precedente troviamo

$$n \geq \left(\frac{z_{1-\alpha/2}}{\delta}\right)^2 \vartheta(1-\vartheta).$$

L'ultima disequazione deve essere soddisfatta per ogni θ quindi deve risultare

$$n \geq \left(\frac{z_{1-\alpha/2}}{\delta}\right)^2 \sup_{\vartheta \in [0,1]} \vartheta(1-\vartheta).$$

E' facile verificare che

$$\sup_{\vartheta \in [0,1]} \vartheta(1-\vartheta) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}.$$

Infatti

$$\frac{d\vartheta(1-\vartheta)}{d\vartheta} = 1 - 2\vartheta = \begin{cases} > 0 \text{ se } 0 \leq \vartheta < \frac{1}{2} \\ = 0 \text{ se } \vartheta = \frac{1}{2} \\ < 0 \text{ se } \frac{1}{2} < \vartheta \leq 1 \end{cases}$$

e quindi $\vartheta(1-\vartheta)$ è crescente tra 0 e 1/2 e decrescente tra 1/2 e 1 ovvero ha un massimo quando $\vartheta = 1/2$.
In definitiva troviamo

$$n \geq \left(\frac{z_{1-\alpha/2}}{\delta}\right)^2 \frac{1}{4} = \left(\frac{2{,}576}{0{,}02}\right)^2 \frac{1}{4} \approx 4147{,}4.$$

e poichè, per considerazioni sia etiche che di tempo/costo, è meglio limitare il più possibile il numero di pazienti coinvolti sembra naturale scegliere la numerosità campionaria più bassa tra quelle che garantiscono la precisione richiesta ovvero porre $n = 4148$.
[nota] Il problema non precisava possibili valori per ϑ. Lo abbiamo risolto quindi "difendendoci" rispetto alla situazione "meno favorevole". Spesso nelle applicazioni esistono delle informazioni a priori su ϑ che possono essere utilizzate per determinare la numerosità campionaria. Ad esempio se ci aspetta che $\vartheta \approx 0{,}85$ potremmo porre

$$n \approx \left(\frac{z_{1-\alpha/2}}{\delta}\right)^2 0{,}85(1-0{,}85) = \left(\frac{2{,}576}{0{,}02}\right)^2 0{,}85(1-0{,}85) \approx 2115.$$

Ovviamente procedendo in questa maniera non siamo sicuri di riuscire a rispettare con certezza la condizione richiesta però, come si può vedere anche dall'esempio numerico, si può arrivare ad un valore di n inferiore.

Esercizio 21. Sia y una determinazione di una variabile casuale binomiale con numero di prove uguale a n e probabilità di successo ϑ. Si supponga che n sia sufficentemente grande per poter utilizzare l'approssimazione normale.

(a) Si costruisca un test appropriato per verificare il sistema di ipotesi *unilaterale*

$$\mathrm{H}_0 : \vartheta \leq \vartheta_0 \text{ verso } \mathrm{H}_1 : \vartheta > \vartheta_0$$

dove ϑ_0 è un valore assegnato.

(a) Si applichi il test alla situazione in cui $y = 558$, $n = 1000$ e $\vartheta_0 = 0{,}5$.

Schema di soluzione.

(a) Nel caso *bilaterale* abbiamo considerato la statistica test

$$z = \frac{\hat{\vartheta} - \vartheta_0}{\sqrt{\dfrac{\vartheta_0(1-\vartheta_0))}{n}}}$$

dove $\hat{\vartheta} = y/n$ indica la stima della probabilità di successo.
z sembra una statistica adeguata anche nel caso di una ipotesi unilaterale. Infatti ci aspettiamo che sotto H_1 assuma valori più grandi di quanto accade sotto H_0 e quindi z sembra essere in grado di discriminare tra le due ipotesi.
Si osservi che alla "frontiera" tra le due ipotesi, ovvero quando $\vartheta = \vartheta_0$, z si distribuisce (approssimativamente) come una $N(0,1)$.
Possiamo "interpretare" i valori di z nella seguente maniera:

- valori più piccoli di quelli che ci "aspettiamo" da una $N(0,1)$ (ad esempio, $z = -32$): verosimilmente $\vartheta < \vartheta_0$ quindi dovremmo accettare H_0;
- valori "possibili" per una $N(0,1)$ (ad esempio $z = \pm 1$): il vero valore di ϑ potrebbe anche essere ϑ_0; quindi non possiamo escludere che l'ipotesi nulla sia vera; anche in questo caso accetteremmo H_0;
- valori più grandi di quelli che ci "aspettiamo" da una $N(0,1)$ (ad esempio: $z = +32$): i dati ci stanno suggerendo che ϑ dovrebbe essere maggiore di ϑ_0; in questo caso dovremmo quindi rifiutare H_0.

Questa disamina suggerisce[12] di accettare l'ipotesi nulla quando

$$z \leq h$$

scegliendo la soglia h uguale ad un "quantile alto" di una $N(0,1)$ ovvero a $z_{1-\alpha}$ per qualche valore di α piccolo ma non nullo.
(b) Nel caso indicato troviamo $\hat{\vartheta} = 558/1000 = 0{,}558$. Quindi

$$z = \frac{\hat{\vartheta} - \vartheta_0}{\sqrt{\dfrac{\vartheta_0(1-\vartheta_0))}{n}}} = \frac{0{,}558 - 0{,}5}{\sqrt{\dfrac{0{,}5(1-0{,}5))}{1000}}} \approx 3{,}67.$$

Questo valore è più grande di quelli che ci si aspetta di "estrarre" da una $N(0,1)$.
In particolare si osservi che, utilizzando la soglia h proposta prima per alcuni "usuali" valori di α,

$$\alpha = 0{,}05 \Rightarrow z_{1-\alpha} = z_{0.95} \approx 1{,}65$$
$$\alpha = 0{,}01 \Rightarrow z_{1-\alpha} = z_{0.99} \approx 2{,}33$$

arriviamo a rifiutare l'ipotesi nulla.

[12] ed è quello che comunemente si fa in un sistema di ipotesi unilaterale

Esercizio 22. Calcolare (anche solo in maniera approssimata) il livello di significatività osservato per il test condotto nell'unità B sugli spessori di 5 lastre di metallo.

Schema di soluzione.

Sappiamo che[13]

- la statistica test utilizzata vale 6,75;
- ci aspettiamo valori della statistica intorno allo zero quando è vera H_0 e lontani da zero, sia verso $-\infty$ che verso $+\infty$, quando è vera H_1.
- la distribuzione sotto H_0 della statistica test è una normale standard;

In questa situazione il livello di significatività osservato è

$$P(N(0,1) \leq 6{,}75) + P(N(0,1) \geq 6{,}75) = 2P(N(0,1) \geq 6{,}75).$$

Da una tavola dei quantili della normale standard, troviamo che 6,75 è più grande del quantile-0,999 di questa distribuzione. Quindi,

$$P(N(0,1) \geq 6{,}75) < 0{,}001$$

e perciò il livello cercato è minore di 0,002.
Un valore così piccolo è altamente significativo contro l'ipotesi formulata. In definitiva, nel momento in cui sono stati rilevati gli spessori delle 5 lastre il processo produttivo, verosimilmente, non era ben tarato.

Esercizio 23. Si consideri il diagramma di dispersione mostrato nella figura 1. Calcolando il livello di significatività osservato del test descritto nell'esercizio 9 in quale dei seguenti intervalli

$$[-1; -0{,}5) \quad [-0{,}5; -0{,}05) \quad [-0{,}05; 0) \quad [0; 0{,}05) \quad [0{,}05; 0{,}5) \quad [0{,}5; 1]$$

vi aspettate che cada?

Schema di soluzione.

Il livello di significatività osservato non può certamente essere negativo (è una probabilità). Questo esclude i primi tre intervalli.
Inoltre, il diagramma mostra chiaramente la presenza di una correlazione lineare. Ci aspettiamo quindi che il test la segnali. Il valore della statistica test che ci attendiamo sarà quindi più grande dei valori "prevedibili" da una normale standard. Per questo motivo non ci aspettiamo che il livello di significatività osservato cada in uno degli ultimi due intervalli.

[13] vedi lucidi delle lezioni.

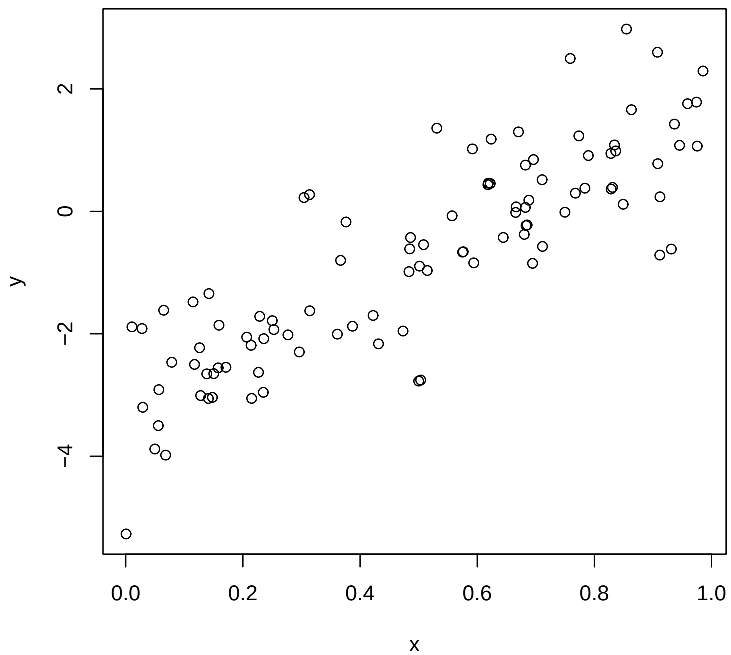

Figura 1: Diagramma di dispersione

In definitiva l'intervallo $[0; 0{,}05]$ dovrebbe contenere il livello di significatività osservato.
[**nota per i curiosi**] il livello di significatività calcolato con i dati nella figura è dell'ordine di 10^{-60}.

Esercizio 24. Solo chi non sa cos'è il livello di significatività osservato può commettere un errore di I tipo se, applicando un certo test statistico, ottiene un livello di significatività osservato uguale a 0,52?

Schema di soluzione.

E' vero. Con un livello di significatività così "grande" è abbastanza assurdo rifiutare H_0. Quindi non si può commettere un errore di I tipo (rifiutare H_0 quando doveva essere accettata).

Esercizio 25. Se conducendo un certo test statistico si perviene ad un livello di significatività osservato uguale a 0,000000007 solo uno statistico "inesperto" potrebbe commettere un errore di II tipo?

Schema di soluzione.

E' vero. Con un livello di significatività così piccolo uno statistico "esperto" rifiuta H_0. Quindi non può commettere un errore di II tipo (accettare H_0 quando doveva essere rifiutata).

Esercizio 26. Quando, conducendo un certo test statistico si perviene ad un livello di significatività osservato minore di 0,00001, è impossibile commettere un errore di II tipo?

Schema di soluzione.

Con un livello di significatività così piccolo normalmente si rifiuta H_0. Quindi, purtroppo, è possibile commettere un errore di II tipo (rifiutare H_0 quando doveva essere accettata). L'affermazione perciò è falsa.

Unità E-F

Esercizio 27. La seguente tabella di contingenza mostra come 319 studenti universitari di varie università si distribuiscono sulla base delle due variabili "Tipo di maturità" e "Numero di esami superati durante il 1° anno".

	Esami superati		
Maturità	$0-1$	$2-5$	>5
Classica	10	67	31
Scientifica	4	52	36
Altre	14	65	40

(a) Analizzare la dipendenza esistente tra le due variabili con gli strumenti a voi noti.
(b) Calcolare una stima della probabilità che, a prescindere dalla maturità di provenienza, uno studente faccia più di 5 esami. Calcolare inoltre un intervallo di confidenza per la medesima probabilità con un livello di copertura uguale a 0,95.

Schema di soluzione.

(a) Completiamo la tabella di contingenza con le distribuzioni marginali e calcoliamo le frequenze relative condizionate "di riga" (visto che quello che interessa è capire se il tipo di maturità influenza l'esito) e le frequenze attese in ipotesi di indipendenza.

– Tabella completata con le distribuzioni marginali

	Esami superati			
Maturità	$0-1$	$2-5$	>5	totale
Classica	10	67	31	108
Scientifica	4	52	36	92
Altre	14	65	40	119
totale	28	184	107	319

– Distribuzioni del numero di esami superati: marginale e condizionate al tipo di maturità (frequenze relative)

	Esami superati			
Maturità	$0-1$	$2-5$	>5	totale
Classica	0,09	0,62	0,29	1
Scientifica	0,04	0,57	0,39	1
Altre	0,12	0,55	0,34	1
totale	0,09	0,58	0,34	1

Commento: Gli studenti provenienti dallo Scientifico inclusi nel campione sembrano essere "andati" leggermente meglio degli altri (solo un 4% nella categoria $0-1$ esami contro una frequenza marginale del 9%, 39% nella categoria "con più di 5 esami" dove la frequenza marginale è del 34%). Gli studenti provenienti dal Classico vanno leggermente meglio di quelli provenienti da "altre" maturità per quanto riguarda la proporzione di studenti con più di 1 esame ma peggio per quanto concerne la proporzione di studenti "ottimi" (più di 5 esami).

– Frequenze attese in ipotesi di indipendenza

	Esami superati			
Maturità	$0-1$	$2-5$	>5	totale
Classica	9,48	62,29	36,23	108
Scientifica	8,08	53,07	30,86	92
Altre	10,45	68,64	39,92	119
totale	28	184	107	319

Commento. I commenti sono gli stessi di quelli fatti guardando alla precedente tabella. Ad esempio, guardando alla riga degli studenti provenienti dallo Scientifico vediamo che, se ci fosse stata indipendenza tra maturità e risultato al 1° anno, (i) ci saremmo aspettati di trovare 8 studenti con 1 o meno esami mentre ne abbiamo osservati solo 4 (cioè la metà) e (ii) viceversa, abbiamo trovato 36 studenti con più di 5 esami nel campione quando ce ne saremmo attesi 30/31.

La dipendenza appena descritta nel campione è una "manifestazione" di una situazione di dipendenza anche nella popolazione da cui il campione è stato estratto? Oppure la dipendenza osservata può essere più banalmente un semplice effetto del caso? Proviamo, per rispondere a queste domande, a considerare il seguente sistema d'ipotesi

$$\begin{cases} H_0 : \begin{pmatrix}\text{tra “tipo di maturità” e “numero di esami} \\ \text{al 1° anno” esiste indipendenza}\end{pmatrix} \\ H_1 : (\text{qualche forma di dipendenza esiste}) \end{cases}$$

La popolazione di riferimento di interesse è ovviamente qualcosa del tipo "tutti gli studenti universitari al primo anno dell'anno in cui è stata condotta l'indagine". Calcoliamo innanzitutto l'X^2 di Pearson

$$X^2 = \frac{(10-9{,}48)^2}{9{,}48} + \cdots + \frac{(40-39{,}92)^2}{39{,}92} = 5{,}48$$

Questo valore sotto H_0 si distribuisce (approssimativamente) come un χ^2 di Pearson con 4 gradi di libertà[14]. Possiamo discutere i risultati in varie maniere alternative ma nella sostanza (e nelle conclusioni) equivalenti:

[14] la più piccola della frequenze attese è circa 8 quindi possiamo fare riferimento a questo risultato asintotico.

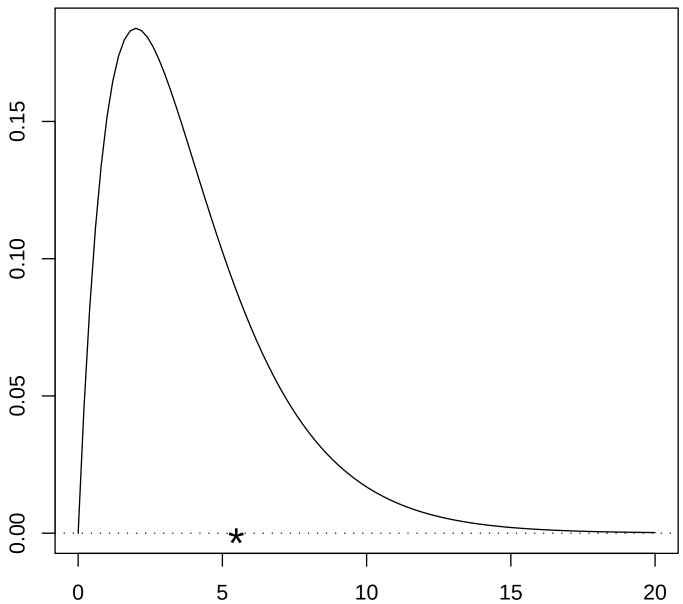

Figura 2: Densità di un χ^2 con 4 gradi di libertà. La stellina sull'asse delle ascisse indica il valore calcolato di X^2.

(i) da una tabella dei quantili di un χ^2 troviamo

$$\begin{aligned} \text{quantile-0,75 di un } \chi^2 \text{ con 4 gradi di libertà} &= 5{,}39 \\ \text{quantile-0,90 di un } \chi^2 \text{ con 4 gradi di libertà} &= 7{,}78 \end{aligned}$$

quindi il valore osservato (5,48) non è particolarmente grande rispetto ai valori attesi sotto l'ipotesi di indipendenza (si veda anche la figura 2).

(ii) dato che per questo test "lontano da H_0" equivale a "lontano da 0", dalle informazioni di prima sui quantili, troviamo

$$0{,}10 < (\text{livello di significatività osservato}) < 0{,}25.$$

Questi valori sono sufficientemente grandi per non "permetterci" di dubitare dell'ipotesi nulla [15].

(iii) pensando ad un approccio con livello di significatività prefissato dovremmo confrontare il valore osservato di X^2 con il quantili $1-\alpha$ (con α piccolo e scelto dall'utilizzatore) di un χ^2 con 4 gradi di libertà, e "rifiutare" l'ipotesi nulla se il valore calcolato dai dati è maggiore del percentile prescelto. Valori usuali per α e i relativi quantili $1-\alpha$ sono

$$\begin{aligned} \alpha = 0{,}10 &\Leftrightarrow \text{quantile} - 0{,}90 = 7{,}78 \\ \alpha = 0{,}05 &\Leftrightarrow \text{quantile} - 0{,}95 = 9{,}49 \\ \alpha = 0{,}01 &\Leftrightarrow \text{quantile} - 0{,}99 = 13{,}28 \end{aligned}$$

Il valore osservato è più piccolo di tutti questi quantili. Quindi, anche utilizzando questo approccio "accettiamo" H_0.

In conclusione: la dipendenza osservata nel campione di 319 studenti tra esito all'università e tipo di scuola superiore di provenienza potrebbe essere dovuta semplicemente al caso; sulla base di questi dati, non abbiamo elementi per pensare che nella popolazione di studenti da cui sono stati estratti gli studenti il "numero di esami superati" dipenda dalla "scuola media di provenienza".

(b) Senza tener conto della scuola di provenienza, abbiamo che $y = 107$ studenti tra gli $n = 319$ considerati hanno durante il primo anno superato più di 5 esami. Possiamo pensare di essere nel contesto di un campionamento binomiale[16] ovvero che

$$y \sim Bi(n, \vartheta)$$

dove ϑ è la probabilità di superare più di 5 esami al primo anno.

La stima di ϑ è

$$\hat{\vartheta} = \frac{y}{n} = \frac{107}{309} \approx 0{,}33.$$

Un intervallo di confidenza (probabilità che includa il vero ϑ approssimativamente uguale a $1-\alpha = 0{,}95$) è

$$\hat{\vartheta} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}} = 0{,}33 \pm 1{,}96\sqrt{\frac{0{,}33(1-0{,}33}{319}} \approx 0{,}33 \pm 0{,}05 = [0{,}28 - 0{,}38].$$

[15] il valore calcolato con R è

```
> 1-pchisq(5.48,4)
[1] 0.2414931
```

[16] il campionamento in situazioni di questo tipo è sempre "senza reinserimento" ma la popolazione di riferimento è composta da qualche centinaio di migliaia di studenti.

Esercizio 28. Tra gli effetti collaterali di un certo farmaco c'è il bruciore di stomaco. Durante uno studio, a 200 pazienti sono state somministrate pillole in cui il farmaco era mescolato con una sostanza tampone A, ad altri 100 pazienti pillole comprendenti la sostanza tampone B ed, infine, ad ulteriori 250 pazienti pillole comprendenti la sostanza tampone C. La percentuale di pazienti che hanno manifestato bruciori di stomaco è stata rispettivamente 12% nel gruppo che aveva ricevuto A, 17% nel gruppo B e 14% nel gruppo C.

(a) Dire, utilizzando un appropriato test statistico, se le differenze osservate tra i tre gruppi possono essere attribuite al caso.

(b) Costruire un grafico che mostri per ogni sostanza tampone sia la stima che un intervallo di confidenza al 90% per la probabilità di avere dei bruciori.

Schema di soluzione.

(a) Dalle informazioni fornite nel testo del problema possiamo ricostruire la sequente tabella di contingenza

	sostanza tampone			
bruciore	A	B	C	totale
no	176	83	215	474
si	24	17	35	76
totale	200	100	250	550

La prima domanda chiede di verificare se, sulla base dei dati, ci aspettiamo che ci sia una differente capacità delle tre sostanze di "tamponare" gli effetti indesiderati del farmaco considerato. Differenti "probabilità di tamponare" equivalgono ovviamente alla dipendenza tra "somministrazione della sostanza tampone" e "presenza di bruciore". Possiamo perciò utilizzare il test X^2 di Pearson, calcolato a partire dalla tabella di sopra, per rispondere alla domanda:

(i) frequenze attese in ipotesi di indipendenza

	sostanza tampone			
bruciore	A	B	C	totale
no	172,36	86,18	215,45	474
si	27,64	13,82	34,55	76
totale	200	100	250	550

(ii) il valore della statistica test è

$$X^2 = \frac{(176 - 172{,}36)^2}{172{,}36} + \cdots + \frac{(35 - 34{,}55)^2}{34{,}55} = 1{,}41;$$

(iii) il valore osservato di X^2 va confrontato con i "valori prevedibili" per un χ^2 di Pearson con 2 gradi di libertà;

(iv) in una tabella dei percentili di questa distribuzione troviamo che

$$1{,}41 \approx (\text{mediana di un } \chi^2 \text{ con 2 gradi di libertà}) = 1{,}39;$$

il valore osservato è perciò più o meno al centro dei valori che ci aspettiamo di osservare quando non esiste dipendenza[17].

In conclusione: i dati suggeriscono che non ci sono differenze di efficacia tra le tre sostanze tampone.

(b) Calcoliamo gli intervalli di confidenza facendo riferimento al procedimento visto per una distribuzione binomiale

$$\hat{\vartheta} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\vartheta}(1-\hat{\vartheta})}{n}}$$

osservando che in questo caso le stime ($\hat{\vartheta}$) e le numerosità (n) sono indicate separatamente per le tre sostanze tampone nel testo e che, sempre nel testo, è richiesto di porre

$$1 - \alpha = 0{,}9 \Rightarrow 1 - \alpha/2 = 0{,}95 \Rightarrow z_{1-\alpha/2} = 1{,}645.$$

Calcolo degli intevalli di confidenza:
sostanza A. $0{,}12 \pm 1{,}645\sqrt{0{,}12(1-0{,}12)/200} = [0{,}08 - 0{,}16]$;
sostanza B. $0{,}17 \pm 1{,}645\sqrt{0{,}17(1-0{,}17)/100} = [0{,}11 - 0{,}23]$;
sostanza C. $0{,}14 \pm 1{,}645\sqrt{0{,}14(1-0{,}14)/250} = [0{,}10 - 0{,}18]$;
Il grafico richiesto diventa quindi "qualcosa" simile a quello della figura 3. Si osservi come gli intevalli di confidenza siano sovrapposti tra di loro. Questo di nuovo è una indicazione che i dati osservati nello studio sono compatibili con una efficacia uguale delle tre sostanze.

Esercizio 29. La sordità (di origine genetica) è purtroppo diffusa tra le gatte bianche europee. Uno zoologo sostiene che il carattere è più diffuso tra le gatte bianche con gli occhi arancioni che tra le gatte bianche con gli occhi celesti. A prova della sua affermazione riporta la seguente tabella contingenza

	colore degli occhi	
sorda	arancioni	celesti
si	17	11
no	8	14

(a) Vi sembra condivisibile l'affermazione dello zoologo?

(b) Determinare, separatamente per i due gruppi, un intervallo di confidenza al 95% per la frazione (nella popolazione di tutte le gatte bianche da cui le gatte utilizzate nello studio provengono) sorde.

[17] alternativamente ed equivalentemente discutere il risultato utilizzando il livello di significatività osservato o via "accetto/rifiuto".

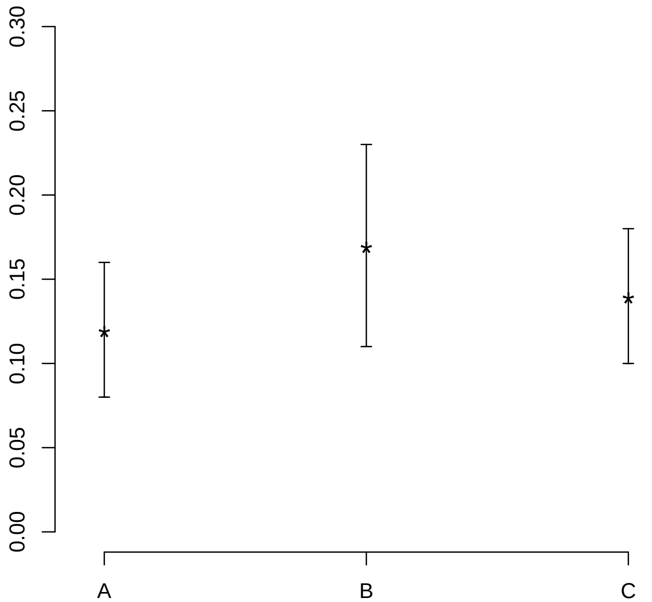

Figura 3: Tre sostanze tampone: intervalli di confidenza per le proporzioni di pazienti con bruciori

Schema di soluzione.

(a) Le differenze delle percentuali nel campione di gatte sorde tra i due colori sono molto grandi (68% tra le gatte con occhi arancioni, 44% tra le gatte con occhi celesti). Però se si calcola X^2 (= 2,92) e si confronta il valore trovato con la distribuzione attesa sotto H_0 (χ^2 con 1 grado di libertà) si conclude a favore di una dubbiosa accettazione (o di un dubbioso rifiuto) dell'ipotesi che la sordidà non dipenda dal colore degli occhi ($p - value \approx 0{,}08$ ad esempio). Le differenze richiamate prima potrebbero perciò anche essere dovute al caso. Probabilmente il nostro zoologo dovrebbe ripetere l'esperimento "guardando gli occhi a più gatte" (la numerosità è piccola e quindi gli errori attesi sono grandi);
(b) Calcoliamo gli intervalli di confidenza facendo riferimento alla distribuzione binomiale. Il risultato è per le gatte con gli occhi arancioni $\approx [0{,}48 - 0{,}83]$ e per le gatte gli occhi celesti $\approx [0{,}27 - 0{,}63]$. Il commento è simile a prima: i dati indicano che le gatte con occhi celesti potrebbero essere meno esposte alla sordità; però non sono sufficienti per escludere la possibilità che la proporzione di gatte sorde nei due gruppi sia la stessa (ad esempio, i due intervalli di confidenza presentano dei valori comuni). Questa "incapacità di discriminare" è certamente legata alla scarsa numerosità campionaria (si guardi l'ampiezza degli intervalli di confidenza).

Esercizio 30. Nel reparto Controllo di Qualità di una azienda si vuole sapere se la frequenza con cui vengono presentati dei reclami al *call-center* varia nei diversi giorni della settimana. Per farlo ha selezionato a caso una settimana e rilevato il numero di reclami presentati nei vari giorni. I dati ottenuti sono

giorno	lunedì	martedì	mercoledì	giovedì	venerdì
reclami	30	32	38	32	25

Secondo voi le differenze osservate nei vari giorni sono significative?

Schema di soluzione.

Siamo nel *reame* di X^2 e χ^2 utilizzati per verificare l'adattamento di un modello teorico.
Assumendo che le frequenze osservate $(O_1, \ldots, O_5) = (30, 32, 38, 32, 25)$ sia una determinazione di una variabile casuale Multinomiale$(157, (p_1, \ldots, p_5))$ l'ipotesi nulla sotto verifica è

$$\mathrm{H}_0 : p_1 = \cdots = p_5 = \frac{1}{5}$$

verso l'ipotesi alternativa

$$\mathrm{H}_1 : \text{almeno una delle uguaglianze previste da } \mathrm{H}_0 \text{ è falsa.}$$

Le frequenze attese sotto H_0 sono

giorno	lunedì	martedì	mercoledì	giovedì	venerdì
reclami attesi	31,4	31,4	31,4	31,4	31,4

La statistica test è

$$X^2 = \frac{(30-31{,}4)^2}{31{,}4} + \cdots + \frac{(25-31{,}4)^2}{31{,}4} \approx 2{,}78.$$

Questo valore va confrontato con quelli previsti da una distribuzione χ^2 con 4 gradi di libertà. Ovviamente, sono i valori più grandi di quelli previsti da questa distribuzione a costituire evidenza contro l'ipotesi nulla. Il valore 2,78 è più piccolo della mediana di un $\chi^2(4)$ che vale 3,36. Il livello di significatività osservato

$$P(\chi^2(4) \geq 2{,}78)$$

è quindi maggiore di 0,5. I dati non contengono quindi nessuna evidenza contro l'ipotesi che la frequenza dei reclami dipenda dal giorno della settimana.

Esercizio 31.

Per verificare se il consumo di alcol dipende dal genere del soggetto sono stati raccolti i dati riassunti nella seguente tabella.

	consumo di bevande alcoliche		
	nullo	moderato	elevato
uomini	139	232	191
donne	97	248	107

E' possibile dire con questi dati che il consumo è differente tra gli uomini e le donne? Rispondere assumendo che le 1014 persone appartenenti al campione siano stati scelti casualmente tra tutti gli abitanti di Padova con età maggiore di 18 anni.

Schema di soluzione.

Innanzitutto, vista l'ultima frase, si deve osservare che i risultati ottenibili analizzando questi dati sono estendibili solamente agli abitanti di Padova di età superiore ai 18 anni.

– *Frequenze attese:*

	consumo di bevande alcoliche		
	nullo	moderato	elevato
uomini	130,80	266,04	165,16
donne	105,20	213,96	132,84

– *Statistica test:*

$$X^2 = \frac{(139-130{,}80)^2}{130{,}80} + \cdots + \frac{(107-132{,}84)^2}{132{,}84} \approx 19{,}99.$$

– *Discussione del risultato:* il valore della statistica test è più grande del quantile-0,999 di una variabile casuale χ^2 con 2 gradi di libertà; quindi il livello di significatività osservata è minore di 0,001; la differenza osservata le donne e gli uomini appartenenti nel campione non sembra quindi dovuta solo al caso ma viceversa manifestazione di una reale differenza di comportamento nella popolazione di riferimento.

Esercizio 32.

Uno zoologo ha seguito 48 stambecchi adulti classificandoli in tre differenti momenti della giornata sulla base delle attività svolte. I risultati sono riassunti nella seguente tabella

	numero di stambecchi che stanno		
momento della giornata	mangiando	dormendo	svolgendo attività sociali
mezzora dopo l'alba	6	25	17
mezzogiorno	14	29	5
tramonto	13	15	20

Le differenze delle distribuzioni degli stambecchi tra le varie attività nei tre momenti considerati sono significative utilizzando $\alpha = 0{,}01$?

Schema di soluzione.

Si tratta di un esempio di test di omogeneità (uguaglianza) tra tre distribuzioni multinomiali.

– *Frequenze attese:*

	numero di stambecchi che stanno		
momento della giornata	mangiando	dormendo	svolgendo attività sociali
mezzora dopo l'alba	11	23	14
mezzogiorno	11	23	14
tramonto	11	23	14

– *Statistica test:*

$$X^2 = \frac{(6-11)^2}{11} + \cdots + \frac{(20-14)^2}{14} \approx 16{,}98.$$

– *Discussione di X^2.* Il valore di X^2 va confrontato con quelli "previsti" da una distribuzione χ^2 con 4 gradi di libertà. Il quantile-0,99 di questa distribuzione vale 13,28. Dato che, ovviamente, 16,98 > 13,28 le differenze sono significative all'1%.

– *Conclusione.* I dati suggeriscono che le attività svolte dipendono dal momento della giornata.

Unità G-I

Esercizio 33. La figura 4 mostra il *normal probability plot* di un certo insieme di dati.
(a) Commentare il grafico.
(b) Ipotizzare un possibile valore (o un intervallo di valori) del livello di significatività del test di Shapiro-Wilks quando applicato a questi dati.

Schema di soluzione.

(a) Anche trascurando le due/tre osservazioni più piccole (ai "bordi" la varianza dei punti disegnati nel *normal probability plot* è elevata) il grafico mostra una chiara curvatura: i dati non sembrano essere stati generati da una distribuzione normale.
(b) Se il test di Shapiro-Wilks fa il suo "lavoro" dovrebbe in questo caso segnalarci la non normalità dei dati e quindi risultare altamente significativo. Quindi un possibile intervallo di valori è

$$\text{livello di significatività osservato} \leq 0{,}01.$$

Nota. Il vero livello di significatività per i dati del grafico vale, all'incirca 10^{-7}.

Esercizio 34. La figura 5 mostra il *normal probability plot* di un certo insieme di dati.
(a) Commentare il grafico.
(b) Ipotizzare un possibile valore (o un intervallo di valori) del livello di significatività del test di Shapiro-Wilks quando applicato a questi dati.

Schema di soluzione.

(a) La linearità del grafico sembra buona. Il grafico sembra indicare che i dati considerati provengono da una popolazione normale.
(b) Il test non dovrebbe risultare significativo e anzi permeterci di accettare senza patemi l'ipotesi di normalità. Diciamo quindi

$$\text{livello di significatività osservato} \geq 0{,}2.$$

Nota. Il vero livello di significatività per i dati del grafico vale, all'incirca 0,69.

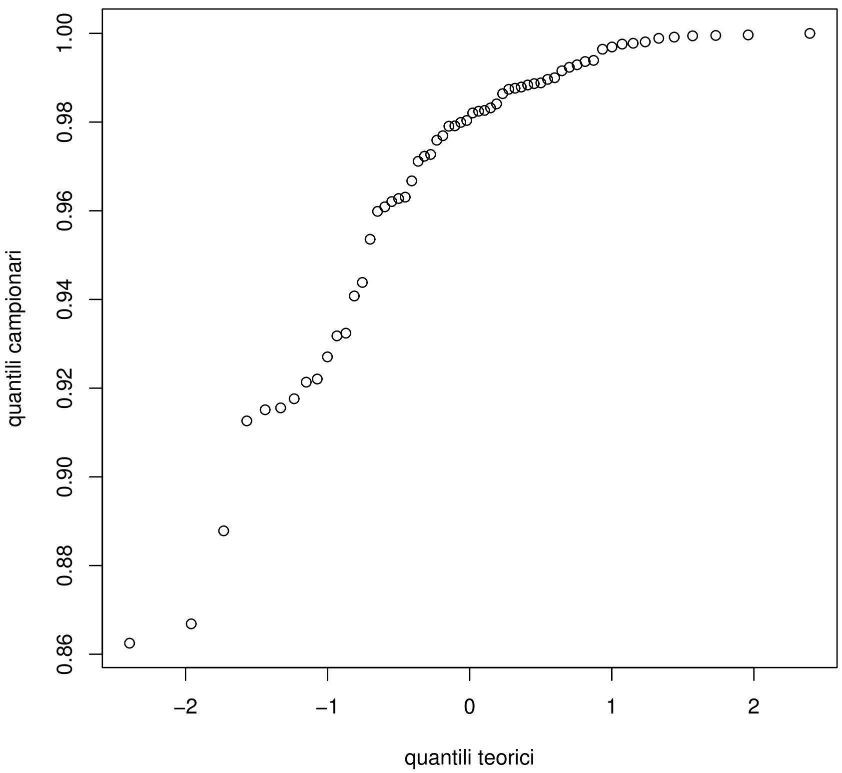

Figura 4: *Normal probability plot* di un insieme di dati.

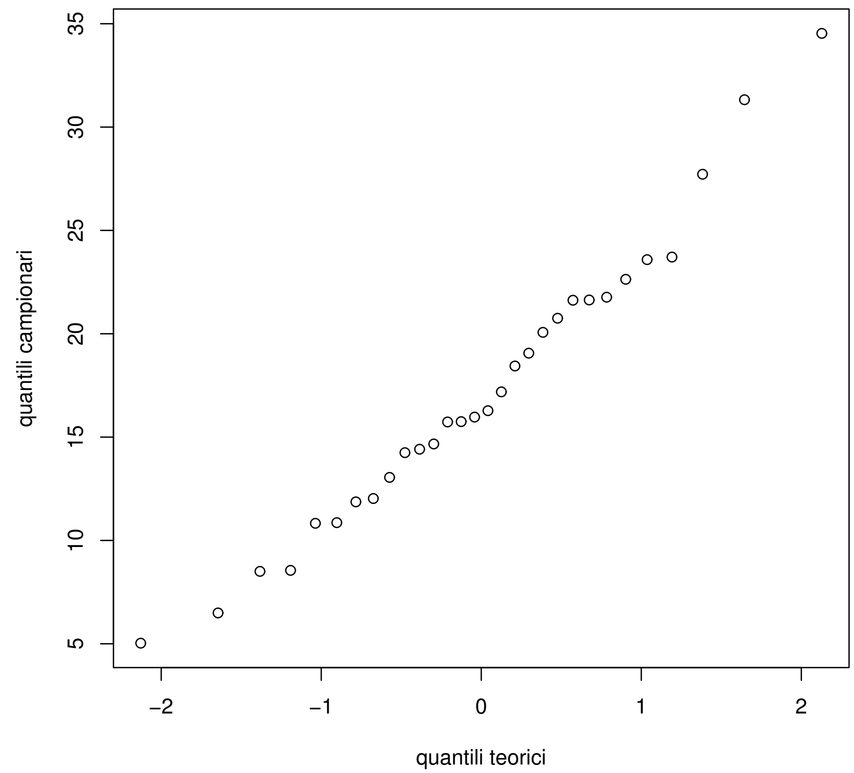

Figura 5: *Normal probability plot* di un insieme di dati.

Esercizio 35. Il test di Shapiro-Wilk è stato applicato a due insiemi di dati numerici univariati, chiamiamoli l'insieme A e quello B. Il livello di significatività osservato è risultato $\approx 0{,}65$ per l'insieme A e $< 10^{-5}$ per l'insieme B.

(a) disegnare, separatamente per i due gruppi, dei *normal probability plot* compatibili con l'informazione data;

(b) disegnare, su di uno stesso piano cartesiano dei diagrammi a scatola con baffi compatibili con le informazioni formite.

Schema di soluzione.

Il test di Shapiro-Wilk verifica l'ipotesi che la distribuzione di un certo insieme di dati sia normale di media e varianza arbitrarie. Quindi, l'informazione che viene data nel testo del problema è:

(i) il gruppo A ha una distribuzione normale (o almeno così sembra dai dati);

(ii) la distribuzione dei dati del gruppo B non è normale (o almeno così sembra dai dati).

I due grafici devono quindi riflettere questa informazione. Un esempio è mostrato nelle figure 6-8. I due insiemi di dati tra l'altro danno luogo ai $p-value$ menzionati nel testo del problema. La non normalità nel gruppo B si nota dalla non linearità del *normal probability plot* e dalla asimmetria del diagramma a scatola con baffi.

Esercizio 36. A volte capita, nei compiti di esame, che qualche studente scriva

> *Se la variabile che stiamo considerando ha distribuzione normale allora i punti del* normal probability plot *sono allineati intorno alla bisettrice del* 1^0 *e del* 3^0 *quadrante.*

Spiegate a questi studenti perchè questa affermazione è, in generale, errata.

Schema di soluzione.

Se la variabile che stiamo considerando ha una distribuzione normale di media μ e varianza σ^2, i punti del *normal probability plot* dovrebbero allinearsi intorno alla retta

$$y = \mu + \sigma x.$$

Questa retta coincide con la bisettrice menzionata solo se $\mu = 0$ e $\sigma = 1$, ovvero se la distribuzione della variabile è una normale standard. In generale però è una qualsiasi retta con coefficente angolare positivo (visto che l'intercetta μ può essere qualsiasi mentre il coefficente angolare coincide con lo scarto quadratico medio σ che non è mai negativo).

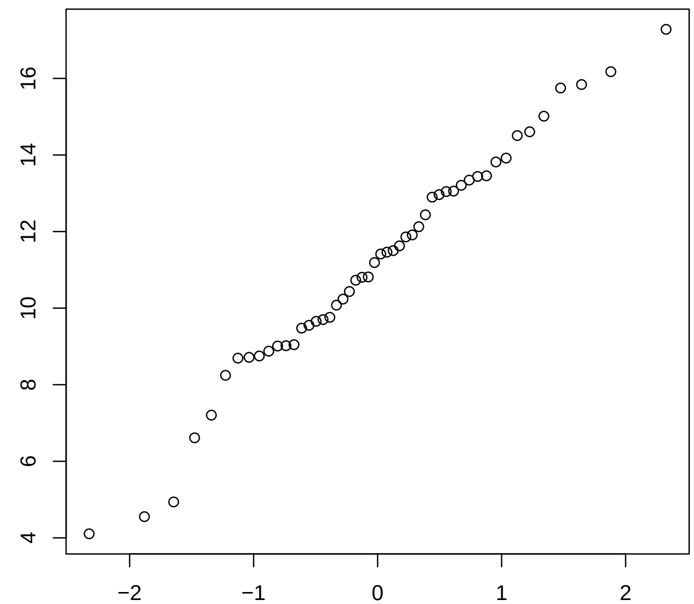

Figura 6: *Normal probability plot* di un ipotetico gruppo A

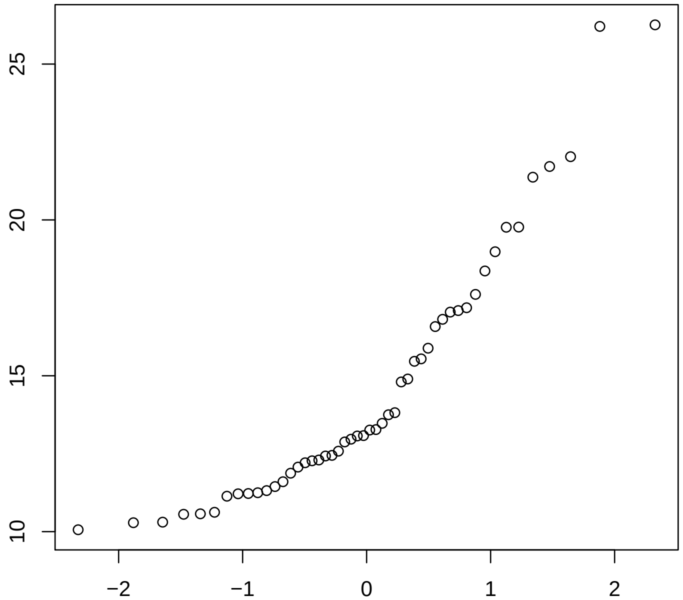

Figura 7: *Normal probability plot* di un ipotetico gruppo B

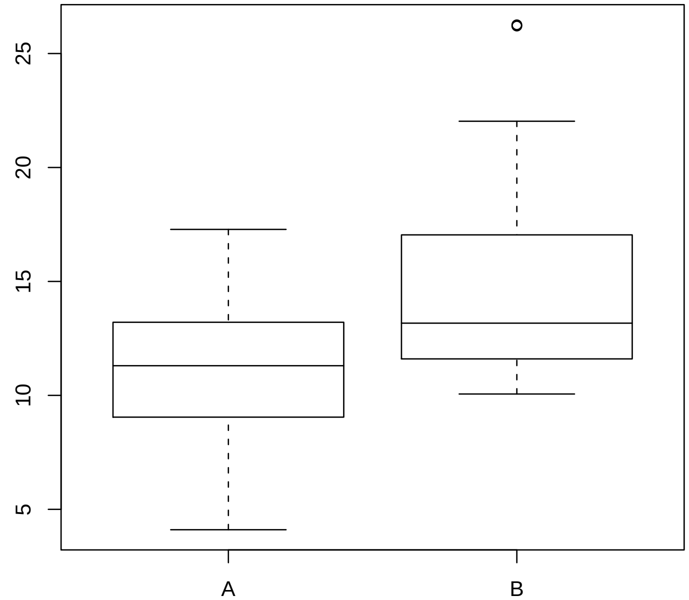

Figura 8: Diagramma a scatola con baffi dei dati del gruppo A e B

ESERCIZIO 37. In un test statistico con livello di significatività prefissato ed uguale ad α se aumentiamo α aumenta o diminuisce la probabilità di errore di I tipo?

Schema di soluzione.

Beh, α è la probabilità di errore di I tipo. Quindi...

ESERCIZIO 38. In un compito d'esame di Inferenza I in cui si richiedeva, tra l'altro, di condurre un test sulla media, $y_1, \ldots, y_n$ rappresentavano i dati osservati e $\overline{y}$ la loro media (ovvero era la media campionaria). Uno studente scrive il sistema d'ipotesi come

$$\mathrm{H}_0 : \overline{y} = 0 \text{ verso } \mathrm{H}_1 : \overline{y} \neq 0.$$

Spiegare perchè lo studente meriterebbe di essere bocciato senza neanche guardare il resto del suo elaborato?

Schema di soluzione.

In primo luogo, il sistema d'ipotesi scritto sopra è veramente poco interessante. Uno studente può anche annoiarsi da pazzi ma, ancora di più se è così, non sembra una idea particolarmente brillante mettersi a fare delle cose inutili. Infatti, nel contesto descritto, $\overline{y}$ lo conosciamo[18]. Quindi, non è che abbiamo bisogno di condurre un test per dire se $\overline{y}$ è uguale o meno allo zero.
In secondo luogo, i metodi, le tecniche e le idee dell'inferenza statistica cercano di farsi raccontare dal campione quello che sa sulla popolazione di riferimento, ovvero, sullo spicchio di mondo da cui "provengono" le osservazioni. Di questo non c'è traccia nel sistema di ipotesi scritto sopra che è "tutto all'interno del campione". Quindi lo studente sembra proprio non sapere bene cosa sta facendo!

ESERCIZIO 39. Una macchina dovrebbe produrre pezzi il cui peso nominale è di $50g$. La media e la varianza campionaria dei pesi di 16 pezzi, scelti casualmente nella produzione di un particolare giorno, sono risultate uguali rispettivamente a $44g$ e $25g^2$. Sapendo che gli scostamenti dei pesi effettivi dal peso medio si distribuiscono normalmente, dire se la macchina in questione è stata ben tarata. Rispondere sia utilizzando un test statistico che mediante il calcolo di un intervallo di confidenza (diciamo al 99%).

[18]riandando al primo esempio che abbiamo fatto sarebbe la media degli spessori delle 5 lastre.

Schema di soluzione.

Siamo ovviamente nel *reame* del test t ad un campione.
Il sistema di ipotesi da verificare è

$$\begin{cases} H_0 : (\text{la macchina produce pezzi mediamente di peso } 50g) \\ H_1 : (H_0 \text{ è falsa}). \end{cases}$$

Le osservazioni disponibili sono i pesi di $n = 16$ pezzi. I singoli pesi non li conosciamo ma ne conosciamo, e per il resto questo sarà sufficiente, la media e la varianza campionaria, indichiamoli con $\overline{y}$ e s^2.
L'usuale statistica test è

$$t_{oss} = \frac{\overline{y} - 50}{\sqrt{s^2/n}} = \frac{44 - 50}{\sqrt{25/16}} \approx -4{,}8.$$

Questo valore deve essere confrontato con i valori "previsti" da una t di Student con $n - 1 = 16 - 1 = 15$ gradi di libertà. Il valore osservato della statistica è, se considerato in valore assoluto, più grande dei quantili di questa distribuzione usualmente considerati come "soglie di accettazione" per il test. Infatti, se indichiamo con $t_p(g)$ il quantile-p di una t con g gradi di libertà, troviamo

$$\begin{array}{lll} \alpha = 0{,}1 & \Leftrightarrow \ 1 - \alpha/2 = 0{,}95 & \Leftrightarrow \ t_{0{,}95}(15) = 1{,}75 \\ \alpha = 0{,}05 & \Leftrightarrow \ 1 - \alpha/2 = 0{,}975 & \Leftrightarrow \ t_{0{,}975}(15) = 2{,}13 \\ \alpha = 0{,}01 & \Leftrightarrow \ 1 - \alpha/2 = 0{,}995 & \Leftrightarrow \ t_{0{,}995}(15) = 2{,}95. \end{array}$$

Utilizzando l'ultimo valore possiamo anche immediatamente dire che il livello di significatività osservato è minore di 0,01.
In conclusione: il valore osservato della statistica test è caduto in una regione dove non ci attendiamo che cada se è vera H_0. Quindi possiamo concludere che i dati suggeriscono che la macchina non sia stata ben tarata.
Alla stessa conclusione possiamo arrivare utilizzando un intervallo di confidenza per la media dei pesi dei pezzi prodotti dalla macchina quando è tarata come ora. L'intervallo (con una copertura pari a $1 - \alpha = 0{,}99$) è

$$\overline{y} \pm t_{1-\alpha/2}(n-1)\sqrt{\frac{s^2}{n}} = 44 \pm 2{,}95\sqrt{\frac{25}{16}} \approx [40{,}3\ ;\ 47{,}7].$$

L'intervallo di confidenza non include il peso nominale di $50g$. Quindi questo peso "non sembra plausibile" sulla base dei dati.

ESERCIZIO 40. Otto uomini sani sono stati sottoposti a una prova di sforzo. Il numero medio di pulsazioni alla fine dell'esercizio è risultato essere uguale a 126 battiti al minuto e la varianza campionaria uguale a 17.
(a) Calcolare un intervallo di confidenza al 99% per la media dei battiti di un individuo sano assumendo l'indipendenza e la normalità dei dati.
(b) Esprimere verbalmente quello che l'intervallo ci racconta.

Schema di soluzione.

(a) Poniamo

$$\begin{aligned} n &= 8 = \text{numerosità del campione}, \\ \overline{y} &= 126 = \text{media campionaria}, \\ s^2 &= 17 = \text{varianza campionaria}, \\ t_{0{,}995}(7) &= 3{,}50 = \text{quantile “appropriato” di una } t \text{ con 7 gradi di libertà}. \end{aligned}$$

L'intervallo richiesto è

$$\overline{y} \pm t_{0{,}995}(7)\sqrt{\frac{s^2}{n}} = 126 \pm 3{,}50\sqrt{\frac{17}{8}} \approx [120{,}9\ ;\ 131{,}1].$$

(b) *"Il numero medio di pulsazioni dopo l'esecuzione dell'esercizio non lo possiamo determinare esattamente; però, sulla base dei dati raccolti, possiamo dire che con grande probabilità (99%) è compreso tra* 121,9 *e* 131,1 *battiti al minuto."*

Esercizio 41.

Per capire se l'ordine alla nascità tra due gemelli dipende dal peso, per dieci coppie di gemelli è stata calcolata la differenza (in Kg).

$$D = \begin{pmatrix}\text{peso alla nascità} \\ \text{primo gemello nato}\end{pmatrix} - \begin{pmatrix}\text{peso alla nascità se-} \\ \text{condo gemello nato}\end{pmatrix}$$

La media e la varianza (calcolata dividendo per $n-1$) delle 10 differenze ottenute valgono rispettivamente 0,15 e 1,21. Supponendo che sia possibile assumere che la distribuzione di D sia normale,

(a) verificare l'ipotesi che l'ordine di nascità non è influenzato dal peso del gemello (condurre il test sia utilizzando un livello di significatività prefissato uguale a 0,90 che calcolando e commentando il livello di significatività osservato);

(b) quello che è stato usato al punto precedente è un test a un campione, a due campioni, per dati appaiati,. . . ?

Schema di soluzione.

(a) La situazione è molto simile a quanto visto nell'unità in cui è stato presentato il test t ad un campione (dati sulle ore di *extra-sonno*).
L'ipotesi nulla sotto verifica è

$$H_0 : \text{l'ordine alla nascita non è influenzata dal peso}$$

che, "traduciamo" in

$$H_0 : \text{la media della distribuzione normale “che genera” } D \text{ è nulla.}$$

La statistica test è

$$t_{oss} = \frac{\sqrt{n}D}{s} = \frac{\sqrt{10} \times 0{,}15}{\sqrt{1{,}21}} \approx 0{,}43.$$

La distribuzione sotto H_0 è una t di Student con 9 gradi di libertà.
Volendo prefissare il livello di significatività a $1-\alpha = 0{,}9$ dobbiamo per prima cosa determinare il percentile $0{,}95 = 1-\alpha/2$ di una t con 9 gradi di libertà. Da una tabella dei quantili troviamo che vale 1,83. Poichè

$$-1{,}83 \leq t_{oss} \leq 1{,}83$$

concludiamo favore di H_0.
Il livello di significatività osservato in questo caso vale

$$2 \times \mathrm{P}(|t \text{ con 9 gradi di libertà}| > 0{,}43).$$

Dalle tabelle troviamo che è maggiore di 0,5. Difficile quindi "sospettare" H_0 sulla base dei dati.
(b) Quello che abbiamo applicato è un test t per dati appaiati che equivale ad utilizzare un test t ad un campione alle differenze tra due osservazioni provenienti dalla stessa unità statistica.
In questo caso:
– *unità statistica* (i)*:* coppia di gemelli;
– *prima misura sulla coppia* (y_i)*:* peso del primo gemello;
– *seconda misura sulla coppia* (x_i)*:* peso del secondo gemello;
– *differenza delle due misure:* $D_i = y_i - x_i$.

Esercizio 42.

Un ricercatore vuole sapere se un certo parametro, rilevabile con un esame del sangue, è differente tra fumatori e non fumatori. Per fare questo seleziona casualmente 10 non fumatori e 10 fumatori e ad ognuno misura il parametro sotto indagine. Alcune misure di sintesi sono riportate nella seguente tabella

	media campionaria	scarto quadratico medio campionario
non fumatori	90,7	5,4
fumatori	87,2	2,8

E' possibile con questi dati dire che la media del parametro di interesse è differente per i fumatori che per i non fumatori? Nella risposta si può assumere che le distribuzioni del parametro di interesse all'interno dei due gruppi siano normali ma con varianza non necessariamente uguale.

Schema di soluzione.

Viste le informazioni date quello siamo nel contesto di un test t a due campioni da condurre utilizzando la correzione di Welch.
Il sistema di ipotesi è

$$\mathrm{H}_0 : \mu = \eta \text{ verso } \mathrm{H}_1 : \mu \neq \eta$$

dove μ e η indicano le medie del parametro ematico di interesse nei fumatori e nei non fumatori.
La statistica test vale

$$\frac{90{,}7 - 87{,}2}{\sqrt{\dfrac{5{,}4^2}{10} + \dfrac{2{,}8^2}{10}}} \approx 1{,}82.$$

La statistica, se è vera H_0, si distribuisce approssimativamente come una t di Student con gradi di libertà uguali all'intero più piccolo di

$$\frac{\left(\dfrac{5{,}5^2}{10} + \dfrac{2{,}8^2}{10}\right)^2}{\dfrac{1}{9}\left(\dfrac{5{,}4^2}{10}\right)^2 + \dfrac{1}{9}\left(\dfrac{2{,}8^2}{9}\right)^2} \approx 13{,}51$$

ovvero, a 13. Uno sguardo alle tavole mostra che 1,82 è compreso tra il quantile 0,95 e il quantile 0,975 di una t con 13 gradi di libertà. Quindi,

$$\text{liv. sign. oss.} \approx 2P(t(13) \geq 1{,}82) \text{ è compreso tra } 0{,}05 \text{ e } 0{,}10.$$

I risultati non sono quindi significativi al 5% ma lo sono all'10%. Siamo quindi in una zona usualmente definita al margine della significatività: i dati ci stanno suggerendo, con molti dubbi, di accettare l'ipotesi nulla. In questo caso verosimilmente la difficoltà di discriminare tra le due ipotesi è legato alla bassa numerosità campionaria.

Esercizio 43. La pressione arteriosa è stata rilevata (in maniera campionaria) in tre popolazioni con abitudini alimentari differenti. I dati disponibili hanno quindi la forma

$$\begin{aligned}
(y_1, \ldots, y_n) &= \text{pressione rilevata su } n \text{ individui della } 1^\circ \text{ popolazione} \\
(x_1, \ldots, x_m) &= \text{pressione rilevata su } m \text{ individui della } 2^\circ \text{ popolazione} \\
(z_1, \ldots, z_h) &= \text{pressione rilevata su } h \text{ individui della } 3^\circ \text{ popolazione}
\end{aligned}$$

Per verificare se la pressione arteriosa media è uguale nelle tre popolazioni è stato utilizzato il test t a due campioni applicandolo separatamente a tutte le

tre possibili coppie di popolazioni. I livelli di significatività ottenuti sono stati i seguenti:

$$\begin{aligned}
1^\circ \text{ popolazione verso } 2^\circ \text{ popolazione: liv. sign. osservato} &= 0{,}622 \\
1^\circ \text{ popolazione verso } 3^\circ \text{ popolazione: liv. sign. osservato} &< 0{,}001 \\
2^\circ \text{ popolazione verso } 3^\circ \text{ popolazione: liv. sign. osservato} &< 0{,}001
\end{aligned}$$

(a) Richiamare brevemente il test t a due campioni.
(b) Commentare i risultati ottenuti. In particolare i dati suggeriscono un effetto delle differenti diete sulla pressione?
(c) Disegnare su di uno stesso grafico degli ipotetici diagrammi a scatola con baffi (uno per le "y", uno per "x" e uno per le "z") compatibili con le informazioni fornite.

Schema di soluzione.

(a) Sintesi del test t a due campioni:
– dati: due campioni $y_1, \ldots, y_n$ e $x_1, \ldots, x_n$ provenienti da due popolazioni;
– normalità e indipendenza:

$$\begin{aligned}
&y_1, \ldots, y_n \text{ determinazioni i.i.d. di una } N(\mu, \sigma^2) \\
&x_1, \ldots, x_m \text{ determinazioni i.i.d. di una } N(\eta, \tau^2) \\
&y_1, \ldots, y_n \text{ indipendenti da } x_1, \ldots, x_m;
\end{aligned}$$

una versione del test (quella esatta e classica) prevede inoltre l'uguaglianza delle due varianze ($\sigma^2 = \tau^2$).
– sistema di ipotesi da verificare:

$$H_0 : \mu = \eta \text{ verso } H_1 : \mu \neq \eta$$

o versioni unilaterali.
– statistica test:

$$t_{oss} = \frac{\overline{y} - \overline{x}}{\text{“s.e. numeratore”}}$$

dove "s.e. numeratore" è una stima dello scarto quadratico medio della differenza delle due medie (la stima è differente a seconda se le varianze possono essere assunte o meno uguali tra di loro).
– La distribuzione della statistica test è, quando le due medie sono uguali, ovvero se è vera H_0, esattamente nel caso di varianze uguali approssimativamente nel caso di varianze differenti una t di Student con gradi di libertà appropriati.
(b) L'esito dei test suggerisce che la media della pressione non differisce tra gli individui della 1° e della 2° popolazione (il confronto non è significativo). Viceversa, i dati suggeriscono delle differenze tra gli individui di queste due popolazioni e quelli della terza popolazione (i risultati sono altamente significativi).

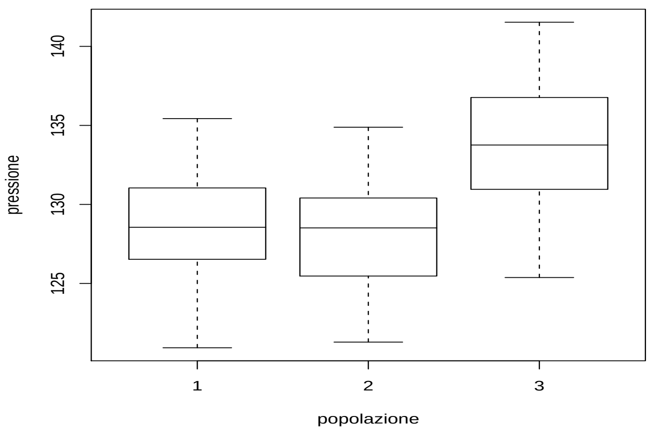

Figura 9: Esempi di diagramma a scatola con baffi

(c) Il grafico deve mostrare differenze in posizione per il terzo gruppo di dati e richiamare in qualche maniera la normalità dei dati (quindi i boxplot devono essere grossomodo simmetrici). Un esempio è mostrato nella figura 9.

Esercizio 44. Svolgere in dettaglio i calcoli dei test t a due campioni e per dati appaiati menzionati nell'unità dei lucidi dedicata alla "coltivazione delle fragole"

Schema di soluzione.

I dati sono

i	1	2	3	4	5	6	7	8	9	10
y_i	216,7	149,9	136,9	211,2	171,4	138,0	127,6	160,3	153,8	150,4
x_i	196,8	108,9	134,7	195,6	160,2	141,9	114,1	130,6	116,2	150,2
z_i	19,9	41,0	2,2	15,6	11,2	-3,9	13,5	29,7	37,6	0,2

dove i indica l'appezzamento, y_i (x_i) indica la quantità di fragole prodotta nel sotto-appezzamento dell'appezzamento i-simo in cui è stato utilizzato il fertilizzante A (B) e $z_i = y_i - x_i$ indica la differenza di prodotto nei due sotto-appezzamenti.

Le quantità che ci servono sono

$$\begin{aligned}
n &= \text{numerosità dei due campioni} = 10 \\
\overline{y} &= \text{media campionaria delle "}y\text{"} = \frac{1}{10}(216{,}7 + \cdots + 150{,}4) = 161{,}62 \\
s_y^2 &= \text{var. camp. "}y\text{"} = \frac{1}{9}[(216{,}7 - 161{,}62)^2 + \cdots + (150{,}4 - 161{,}2)^2] = 915{,}57 \\
\overline{x} &= \text{media campionaria delle "}x\text{"} = \frac{1}{10}(196{,}8 + \cdots + 150{,}2) = 144{,}92 \\
s_x^2 &= \text{var. camp. "}x\text{"} = \frac{1}{9}[(196{,}8 - 144{,}92)^2 + \cdots + (150{,}2 - 144{,}92)^2] = 990{,}24 \\
s^2 &= \text{stima varianza di } x \text{ e } y \text{ quando si assume che sia uguale nei due gruppi} = \\
&= \frac{1}{18}(9s_y^2 + 9s_x^2) = 952{,}90 \\
\overline{z} &= \text{media campionaria delle "}z\text{"} = \frac{1}{10}(19{,}9 + \cdots + 0{,}2) = 16{,}7 \\
s_z^2 &= \text{var. camp. "}z\text{"} = \frac{1}{9}[(19{,}9 - 16{,}7)^2 + \cdots + (0{,}2 - 16{,}7)^2] = 239{,}45
\end{aligned}$$

(a) *t a due campioni: varianze uguali.* La statistica test è

$$\frac{\overline{y} - \overline{x}}{s\sqrt{\dfrac{1}{n} + \dfrac{1}{n}}} = \frac{16{,}7}{13{,}81} = 1{,}21.$$

La distribuzione sotto H_0 è (se valgono le ipotesi di normalità, varianze uguali e indipendenza tra e nei gruppi) una t di Student con $n + n - 2 = 18$ gradi di libertà. Dalla tavola dei quantili, troviamo che 1,33 è il quantile-0,90 di questa distribuzione. Perciò

$$\text{liv. sign. oss.} = 2P(t(18) \geq 1{,}21) > 2P(t(18) \geq 1{,}33) = 0{,}2.$$

Questo test "dichiara" quindi non significativa la differenza tra le medie.
(b) *t a due campioni: varianze non uguali.* La statistica test è

$$\frac{\overline{y} - \overline{x}}{\sqrt{\dfrac{s_y^2}{n} + \dfrac{s_x^2}{n}}} = \frac{16{,}7}{14{,}07} = 1{,}19.$$

La distribuzione sotto H_0 (se valgono le ipotesi di normalità e di indipendenza tra e nei gruppi) può essere approssimata con una t di Student con gradi di libertà uguali all'intero più piccolo

$$\frac{\left(\frac{s_y^2}{n}+\frac{s_x^2}{n}\right)^2}{\frac{1}{n-1}\left(\frac{s_y^2}{n}\right)^2+\frac{1}{n-1}\left(\frac{s_x^2}{n}\right)^2}=17{,}97$$

ovvero, con una t di Student con 17 gradi di libertà. Anche in questo caso, una scorsa veloce ad una tabella dei quantili della t permette di affermare che il livello di significatività osservato è maggiore del 20%. La conclusione quindi è uguale a quella ottenuta nel caso precedente.
(c) *t per dati appaiati.* La statistica test è

$$\frac{\overline{z}}{\sqrt{\frac{s_z^2}{n}}}=\frac{16{,}7}{4{,}89}=3{,}41.$$

La distribuzione sotto H_0 è (se vale l'ipotesi di normalità e indipendenza nei gruppi) una t di Student con $n-1=9$ gradi di libertà. Dalla tavola dei quantili, troviamo che 3,25 è il quantile-0,995 di questa distribuzione. Perciò

$$\text{liv. sign. oss.} = 2P(t(9)\geq 3{,}41) < 2P(t(9)\geq 3{,}25)=0{,}01.$$

Questo test "dichiara" quindi altamente significativa la differenza tra le medie.
Osservazione. Si noti come il numeratore delle tre statistiche test, che "misura" la differenza media tra le quantità prodotte nei sotto-appezzamenti in cui era stato utilizzato il primo fertilizzante e quelli in cui era stato utilizzato il secondo fertilizzante, sia sempre la stessa. Quello che cambia nei tre test è la maniera in cui questa differenza è standardizzata, ovvero, la stima posta al denominatore dello scarto quadratico medio della differenza al numeratore. Nei primi due casi, lo scarto quadratico medio fa riferimento alla relazione

$$\text{var}\{\overline{y}-\overline{x}\}=\text{var}\{\overline{y}\}+\text{var}\{\overline{x}\}=\frac{\text{var}\{``y"\}}{n}+\frac{\text{var}\{``x"\}}{n}$$

che vale solamente nel caso in cui le "y" siano indipendenti dalle "x" (ipotesi che come visto nei lucidi non sembra in questo caso valere). Nel terzo caso (t per dati appaiati) viceversa la varianza delle differenze viene stimata dalle differenze stesse. Quindi non dipende dall'assunzione di indipendenza tra le "y" e le "x".

Esercizio 45. Durante un certo studio clinico è stata rilevata la pressione arteriosa in 12 individui sani di sesso maschile e in 15 individui sani di sesso

femminile. Indichiamo con $(x_1,\dots,x_{12})$ le osservazioni ottenute sui maschi e con $(y_1,\dots,y_{15})$ le osservazioni fatte sulle femmine. Una analisi preliminare dei dati ha indicato che ci si trovava in una situazione in cui per verificare se la media della pressione arteriosa dei maschi è differente dalla media delle femmine poteva essere utilizzato il test t a due campioni.
(a) Quali caretteristiche dei dati potrebbero essere state analizzate per arrivare alla conclusione che poteva essere utilizzato il test t a due campioni?
(b) Supponendo che il livello di significatività osservato del test t calcolato con i dati in questione sia risultato uguale a 0,003 disegnare, affiancati su di uno stesso piano cartesiano, due ipotetici *boxplot*, uno per le "x", l'altro per le "y", compatibili con le informazioni date.
(c) Ripetere il grafico richiesto alla domanda (b) ipotizzando però questa volta che il livello di significatività osservato valga 0,31.

Schema di soluzione.

(a) Il test t si basa sull'assunzione di normalità dei dati. Quindi certamente questa caratteristica è stata indagata separatamente per i due gruppi (ad esempio utilizzando un *normal probability plot* e/o il test di Shapiro-Wilk). Potrebbe poi essere stata confrontata la variabilità delle "y" e delle "x" per capire se utilizzare la versione del test che si basa sull'uguaglianza delle due varianze oppure quella di Welch.
(b) Il livello di significatività osservato è molto piccolo: i dati suggeriscono una differenza nella medie delle due popolazioni. Quindi devono essere disegnati due *boxplot* che, vista la domanda di prima, "richiamino la normalità" (in particolare siano grossomodo "simmetrici") ma con posizione differente; un possibile esempio è mostrato nella figura 10-(a).
(c) In questo caso il livello di significatività osservato è abbastanza grande: non possiamo sulla base dei dati dubitare del fatto che la pressione media sia uguale nei due gruppi. Quindi dobbiamo disegnare due *boxplot* che, anche in questo caso, "richiamino la normalità" e che abbiano una posizione non di molto differente tra di loro (ma neanche esattamente uguale, visto che se le due medie campionarie fossero esattamente uguali il livello di significatività osservato varrebbe zero). Un possibile esempio è mostrato nella figura 10-(b).

Esercizio 46. Calcolare un intervallo di confidenza (livello di copertura 90%) per la differenza tra le medie delle lunghezze delle uova di cuculo deposte nei nidi di pettirossi e scriccioli viste nei lucidi.

Schema di soluzione.

Lavoriamo "sotto le ipotesi" e con le notazioni dei lucidi:

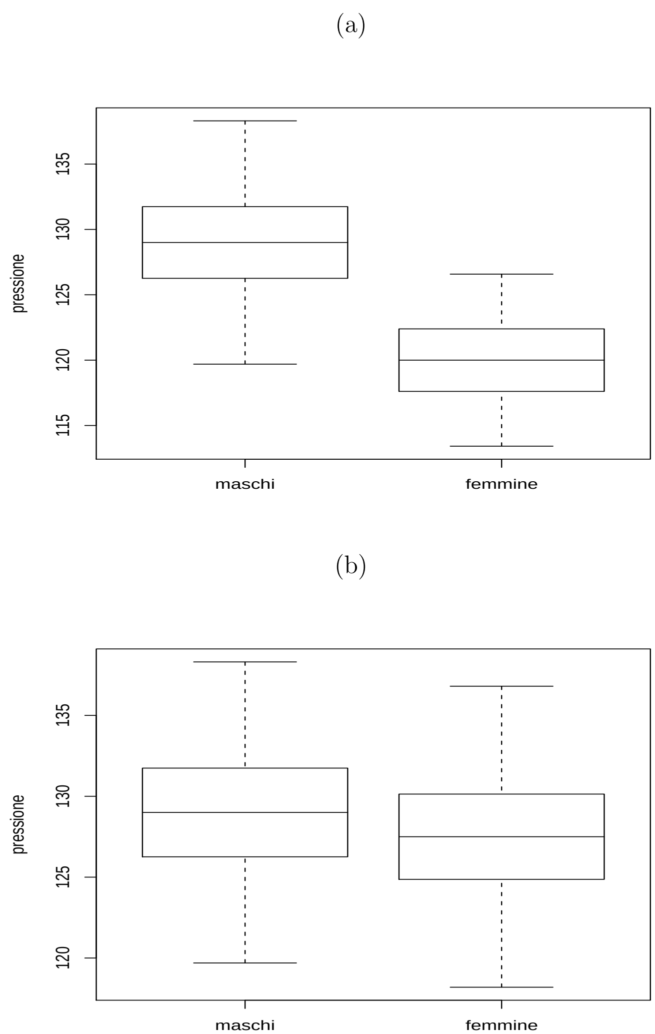

Figura 10: Due coppie di diagrammi a scatola con baffi

(i) due insiemi di dati

$$y_1, \dots, y_n \text{ e } x_1, \dots, x_n$$

provenienti da due differenti popolazioni;

(ii) normalità e indipendenza

$$\begin{aligned} &y_1, \dots, y_n \text{ determinazioni i.i.d. di una } N(\mu, \sigma^2) \\ &x_1, \dots, x_m \text{ determinazioni i.i.d. di una } N(\eta, \tau^2) \\ &y_1, \dots, y_n \text{ indipendenti da } x_1, \dots, x_m \end{aligned}$$

(iii) varianze uguali $\sigma^2 = \tau^2$.

Sappiamo allora che

$$\frac{(\overline{y} - \overline{x}) - (\mu - \eta)}{s\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim t \text{ di Student con } n + m - 2 \text{ gradi di libertà}$$

dove

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,\ \overline{x} = \frac{1}{m}\sum_{i=1}^{n} x_i \text{ e } s^2 = \frac{1}{n+m-2}\left((n-1)s_y^2 + (m-1)s_x^2\right)$$

con

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 \text{ e } s_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \overline{x})^2,$$

Indichiamo con $t_p(g)$ il quantile-p di una t di Student con g gradi di libertà. Allora, ricordandoci anche della simmetria delle t e quindi del fatto che $t_{1-p}(g) = -t_p(g)$,

$$\mathrm{P}\left(-t_{1-\alpha/2}(n+m-2) \le \frac{(\overline{y} - \overline{x}) - (\mu - \eta)}{s\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \le t_{1-\alpha/2}(n+m-2)\right) = 1 - \alpha.$$

Riscrivendo le due disuguaglianze in maniera tale da isolare $\mu - \eta$, troviamo che

$$(\overline{y} - \overline{x}) \pm t_{1-\alpha/2}(n+m-2)s\sqrt{\frac{1}{n} + \frac{1}{m}}$$

è un intervallo di confidenza (con livello di copertura $1-\alpha$) per la differenza delle due medie.

Nel caso dei dati che stiamo considerando:

$$\begin{aligned} &n = 16 \quad \overline{y} \approx 22{,}47 \quad s_y^2 \approx 0{,}46 \\ &m = 15 \quad \overline{x} \approx 21{,}13 \quad s_x^2 \approx 0{,}55 \end{aligned}$$

Quindi,

$$s \approx \sqrt{(15 \times 0{,}46 + 14 \times 0{,}55)/29} \approx 0{,}71.$$

Ponendo $\alpha = 0{,}10$ troviamo $t_{0,95}(29) = 1{,}70$ e quindi l'intervallo di confidenza richiesto è

$$(22{,}47 - 21{,}13) \pm 1{,}70 \times 0{,}71\sqrt{\frac{1}{16} + \frac{1}{15}} \approx [0{,}91; 1{,}77].$$

L'adattamento delle uove di cuculo a quelle dell'uccello "ospite" sembra quindi essersi concretizzato in una diminuzione della lunghezza che con alta probabilità (90%) i dati ci suggeriscono essere tra $0{,}91mm$ e $1{,}77mm$.

Osservazione: Nell'ipotesi di varianze non uguali si può procedere essenzialmente nella stessa maniera utilizzando "la correzione di Welch". In questo caso il risultato di partenza è che la distribuzione di

$$\frac{(\overline{y} - \overline{x}) - (\mu - \eta)}{\sqrt{\dfrac{s_y^2}{n} + \dfrac{s_x^2}{m}}}$$

può essere approssimata con quella di una t di Student con

$$\text{gradi di libertà} \approx \frac{\left(\dfrac{s_y^2}{n} + \dfrac{s_x^2}{m}\right)^2}{\dfrac{1}{n-1}\left(\dfrac{s_y^2}{n}\right)^2 + \dfrac{1}{m-1}\left(\dfrac{s_x^2}{m}\right)^2}.$$

Indichiamo con g i gradi di libertà calcolati con l'ultima formula (e magari arrotondati all'intero più vicino). Allora, possiamo scrivere che

$$\mathrm{P}\left(-t_{1-\alpha/2}(g) \leq \frac{(\overline{y} - \overline{x}) - (\mu - \eta)}{\sqrt{\dfrac{s_x^2}{n} + \dfrac{s_x^2}{m}}} \leq t_{1-\alpha/2}(g)\right) \approx 1 - \alpha$$

da cui, isolando al solito $\mu - \eta$, troviamo che

$$(\overline{y} - \overline{x}) \pm t_{1-\alpha/2}(g)\sqrt{\frac{s_y^2}{n} + \frac{s_x^2}{m}}$$

è un intervallo di confidenza (con livello di copertura approssimativamente uguale ad $1 - \alpha$) per la differenza delle due medie.

Nel caso in esame, abbiamo visto nei lucidi, che $g \approx 28$. Quindi, osservando che al livello di approssimazione con cui stiamo lavorando anche $t_{0,95}(28) = 1{,}70$, un intervallo di confidenza al 90% per la differenza delle due medie può essere calcolato come

$$(22{,}47 - 21{,}13) \pm 1{,}70\sqrt{\frac{0{,}46}{16} + \frac{0{,}55}{15}} \approx [0{,}91; 1{,}77].$$

Come si vede, nessuna differenza nei primi due decimali del risultato. In questo caso assumere o non assumere le varianze uguali non produce nessun cambiamento.

Esercizio 47. Per capire se una dieta ricca di pesce può portare ad una riduzione del colesterolo si è proceduto nella seguente maniera:

- si sono "reclutati" 200 individui adulti di sesso maschile senza segni clinici di malattia;
- i 200 individui sono stati suddivisi casualmente in due gruppi di 100 individui ciascuno;
- al primo gruppo è stato chiesto di seguire per 6 mesi una dieta prefissata in cui veniva consumato del pesce una volta ogni due settimane;
- al secondo gruppo, viceversa, è stato chiesto di seguire per 6 mesi la stessa dieta ma sostituendo un giorno si e uno no la fonte di proteine con del pesce;
- dopo i 6 mesi è stato misurato il colesterolo per ogni individuo.

Le medie e le varianza campionarie delle misure ottenute sono riportate nella seguente tabella.

	media campionaria	varianza campionaria
dieta con "poco pesce"	210,1	37,4
dieta con "parecchio pesce"	196,8	13,5

La differenza delle due campionarie può essere *completamente* essere attribuita al caso, ovvero all'errore campionario? Rispondere sia utilizzando un test che calcolando un intervallo di confidenza.

Schema di soluzione.

La dimensione dei due campioni è sufficentemente elevata per poter far riferimento alla distribuzione asintotica della differenza delle due medie. In particolare, possiamo svolgere l'esercizio assumendo che

$$\frac{(\overline{y} - \overline{x}) - (\mu - \eta)}{\sqrt{\dfrac{s_y^2}{n} + \dfrac{s_x^2}{n}}} \text{ sia (approssimativamente) distribuito come una } N(0,1)$$

dove
· μ e η indicano le vere medie del colesterolo nei due gruppi dopo 6 mesi di dieta;
· $\overline{y}$ e s_y^2 indicano la media e la varianza campionaria delle misure ottenute dal gruppo "dieta con poco pesce";
· $\overline{x}$ e s_x^2 sono le analoghe quantità riferite al gruppo "dieta con parecchio pesce";
· $n = 100$ indica la numerosità di ciascuno dei due gruppi.

Si osservi che, vista la non piccola differenza esistente tra le due varianze campionarie, abbiamo deciso di lavorare senza assumere che le varianze dei due gruppi siano uguali.

(a) *Intervallo di confidenza.* Indichiamo con z_p il quantile-p di una normale standard. Se vale l'approssimazione di qui sopra possiamo scrivere

$$P\left(-z_{1-\alpha/2} \leq \frac{(\overline{y}-\overline{x})-(\mu-\eta)}{\sqrt{\frac{s_y^2}{n}+\frac{s_x^2}{n}}} \leq z_{1-\alpha/2}\right) \approx 1-\alpha.$$

Quindi, riscrivendo le due disugliaze in maniera tale di "isolare" $\mu-\eta$, otteniamo che

$$(\overline{y}-\overline{x}) \pm z_{1-\alpha/2}\sqrt{\frac{s_y^2}{n}+\frac{s_x^2}{n}}$$

è un intervallo di confidenza di livello approssimativamente uguale ad $1-\alpha$ per la differenza tra le due medie.

Con i dati, scegliendo $\alpha = 0{,}05$ e quindi ponendo $z_{1-\alpha/2} = z_{0{,}975} = 196$, troviamo

$$(210{,}1 - 196{,}8) \pm 1{,}96\sqrt{\frac{37{,}4}{100}+\frac{13{,}5}{100}} = [11{,}9\ ;\ 14{,}7].$$

Commento. L'intervallo non contiene lo zero e quindi i dati sembrano indicare che esiste una differenza reale tra le due medie ed, in particolare che la prima dieta, quella con "poco pesce", risulti associata a livelli medi più elevati del colesterolo.

(b) *Test d'ipotesi.* Il sistema d'ipotesi è ovviamente

$$\mathrm{H}_0 : \mu = \eta \text{ verso } \mathrm{H}_1 : \mu \neq \eta.$$

Una possibile statistica test è

$$z = \frac{\overline{y}-\overline{x}}{\sqrt{\frac{s_y^2}{n}+\frac{s_x^2}{n}}}.$$

Per quanto detto, la distribuzione di z quando è vera H_0 è approssimativamente una normale standard. Con i dati

$$z = \frac{210{,}1 - 196{,}8}{\sqrt{\frac{37{,}4}{100}+\frac{13{,}5}{100}}} = 18{,}6.$$

Questo valore è molto lontano da quanto ci si aspetterebbe sotto H_0(il livello di significatività osservato è praticamente nullo e 18,6 è molto più grande dei quantili di una $N(0,1)$ relativi agli usuali livelli di significatività).

Commento. I dati non sembrano compatibili con l'ipotesi che le due medie siano uguali.

Esercizio 48.

E' possibile dimostrare che se $y_1, \ldots, y_n$ sono n determinazioni indipendenti ed identicamente distribuite di una variabile casuale $N(\mu, \sigma^2)$ allora

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2 \text{ con } n-1 \text{ gradi di libertà}$$

dove s^2 indica la varianza campionaria.

(a) Utilizzare questo risultato per costruire un test per verificare il sistema di ipotesi

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

dove σ_0^2 indica un valore preassegnato per la varianza.

(b) Utilizzare il test costruito al punto precedente per verificare se è credibile che la varianza della distribuzione che ha generato i 5 spessori considerati nell'unità B sia 0,01.

Schema di soluzione.

(a) Si consideri la statistica

$$\xi = \frac{(n-1)s^2}{\sigma_0^2}.$$

Se H_0 è "vera" ci aspettiamo

$$s^2 \approx \sigma_0^2 \Leftrightarrow \frac{s^2}{\sigma_0^2} \approx 1,$$

e quindi che ξ "cada" intorno a $n-1$. Viceversa se H_0 è falsa ci aspettiamo che la distribuzione di ξ si sposti verso valori più [grandi,piccoli] di $n-1$ a seconda se σ^2 sia più [grande,piccolo] di σ_0^2. Possiamo quindi usare ξ come statistica test. Per quanto enunciato nel testo la distribuzione di ξ sotto l'ipotesi nulla è un χ^2 con $n-1$ gradi di libertà. Il valore calcolato dai dati di ξ dovrà quindi essere confrontato con i valori "attesi" per questa distribuzione. Ad esempio, volendo utilizzare un test con livello di significatività prefissato, possiamo pensare di utilizzare la procedura stilizzata nella figura 11. Infatti, la procedura descritta scegliere α

↓

determinare
$h =$ quantile-$(\alpha/2)$ e
$k =$ quantile-$(1-\alpha/2)$
di un χ^2 con $n-1$ gradi di libertà

↓

calcolare $\xi = \dfrac{(n-1)s^2}{\sigma_0^2}$

↓

verificare se
$h \leq \xi \leq k$

se la risposta è si → accettare H_0

se la risposta è no → rifiutare H_0

Figura 11: Test sulla varianza di una distribuzione normale

garantisce che

$$\begin{aligned} &\mathrm{P}(\text{accettare } H_0 \text{ quando } H_0 \text{ è vera}) = \\ &= \mathrm{P}(h \leq \xi \leq k \text{ quando } H_0 \text{ è vera}) = \\ &= \mathrm{P}(h \leq (\chi^2 \text{ con } n-1 \text{ gradi di libertà}) \leq k) = 1-\alpha. \end{aligned}$$

(b) Le misure da considerare sono

$$14{,}33 \quad 14{,}19 \quad 14{,}39 \quad 14{,}43 \quad 14{,}17.$$

In questo caso, n vale 5,

$$\sum_{i=1}^{n} y_i = 71{,}51 \qquad \sum_{i=1}^{n} y_i^2 = 1022{,}791$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right) = \frac{1}{4}\left(1022{,}791 - \frac{71{,}51^2}{5}\right) = 0{,}0137$$

e quindi

$$\xi = \frac{(n-1)s^2}{\sigma_0^2} = \frac{4 \times 0{,}0137}{0{,}01} = 5{,}48.$$

Scegliendo $\alpha = 0{,}1$ troviamo

$$\begin{aligned} h &= \text{quantile-0,05 di un } \chi^2 \text{ con 4 gradi di libertà} = 0{,}71 \\ h &= \text{quantile-0,95 di un } \chi^2 \text{ con 4 gradi di libertà} = 9{,}49. \end{aligned}$$

Il valore di ξ calcolato dai dati cade ampiamente all'interno dell'intervallo definito da questi due valori. Concludiamo quindi a favore di H_0.

Esercizio 49. Siano $(y_1, \ldots, y_n)$ delle determinazioni indipendenti di una variabile casuale con distribuzione normale di media μ e varianza σ^2 e $(x_1, \ldots, x_m)$ delle determinazioni indipendenti di una variabile casuale con distribuzione normale di media η e varianza τ^2.
Si ponga inoltre

$$\begin{aligned} \overline{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad & s_y^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2, \\ \overline{x} &= \frac{1}{m}\sum_{i=1}^{m} x_i, \qquad & s_x^2 &= \frac{1}{n-1}\sum_{i=1}^{m}(x_i - \overline{x})^2. \end{aligned}$$

E' possibile dimostrare che

$$\frac{s_y^2/\sigma^2}{s_x^2/\tau^2} \sim F(n-1, m-1)$$

dove $F(g_1, g_2)$ indica una variabile casuale F di Snedecor con g_1 gradi di libertà al numeratore e g_2 al denominatore.
Utilizzare questo risultato per costruire
(a) un test statistico per verificare il sistema di ipotesi

$$H_0 : \sigma^2 = \tau^2 \text{ verso } H_1 : \sigma^2 \neq \tau^2$$

(b) un intervallo di confidenza di livello $1-\alpha$ per il rapporto delle due varianze (ovvero per σ^2/τ^2).

Inoltre applicare il test e calcolare l'intervallo di confidenza (fissando $\alpha = 0{,}01$) ai dati delle lunghezze delle uova di cuculo. Nel farlo può essere utile conoscere i seguenti quantili di una $F(15, 14)$

p	0,001	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995	0,999
F_p	0,18	0,24	0,28	0,35	0,41	0,5	1	2,01	2,46	2,95	3,66	4,25	5,85

F_p (seconda riga) è il quantile-p (p è riportato nella prima riga) di una F di Snedecor con 15 gradi di libertà al numeratore e 14 al denominatore.

scegliere α

$\downarrow$

determinare
$F_{\alpha/2}(n-1, m-1) =$ quantile-$(\alpha/2)$ e
$F_{1-\alpha/2}(n-1, m-1) =$ quantile-$(1-\alpha/2)$
di una F con $n-1$ e $m-1$ gradi di libertà

$\downarrow$

verificare se
$$F_{\alpha/2}(n-1, m-1) \leq \frac{s_y^2}{s_x^2} \leq F_{1-\alpha/2}(n-1, m-1)$$

se la risposta è si $\swarrow$ accettare H_0

se la risposta è no $\searrow$ rifiutare H_0

Figura 12: Confronto della varianza di due variabili casuali normali

Schema di soluzione.

(a) E' chiaro che

se	ci aspettiamo che $\frac{s_y^2}{s_x^2}$, tendelzialmente, risulti
$\sigma^2 < \tau^2$	< 1
$\sigma^2 = \tau^2$	≈ 1
$\sigma^2 > \tau^2$	> 1

Possiamo quindi utilizzare s_y^2/s_x^2 come statistica test. Sotto H_0, per il risultato menzionato nel testo del problema risulta

$$\frac{s_y^2}{s_x^2} \sim F \text{ con } n-1 \text{ gradi di libertà al numeratore e } m-1 \text{ al denominatore.}$$

Il valore calcolato dai dati dovrà quindi essere confrontato "con i valori prevedibili per questa distribuzione". Un possibile approccio del tipo accetto/rifiuto è mostrato nella figura 12. La procedura stilizzata nella figura 12 garantisce che

$$\begin{aligned}
&\mathrm{P}(\text{accettare } H_0 \text{ quando } H_0 \text{ è vera}) = \\
&= \mathrm{P}\left(F_{\alpha/2}(n-1, m-1) \leq \frac{s_y^2}{s_x^2} \leq F_{1-\alpha/2}(n-1, m-1) \middle| \, H_0 \text{ vera}\right) = \\
&= \mathrm{P}(F_{\alpha/2}(n-1, m-1) \leq F(n-1, m-1) \leq F_{1-\alpha/2}(n-1, m-1)) = \\
&= 1 - \alpha.
\end{aligned}$$

(b) Sempre utilizzando il risultato dato nel testo del problema possiamo scrivere

$$\mathrm{P}\left(F_{\alpha/2}(n-1, m-1) \leq \frac{s_y^2/\sigma^2}{s_x^2/\tau^2} \leq F_{1-\alpha/2}(n-1, m-1)\right) = 1 - \alpha.$$

Dividendo tutti i termini delle diseguaglianze per s_y^2/s_x^2 troviamo

$$\mathrm{P}\left(\frac{F_{\alpha/2}(n-1, m-1)}{s_y^2/s_x^2} \leq \frac{1}{\sigma^2/\tau^2} \leq \frac{F_{1-\alpha/2}(n-1, m-1)}{s_y^2/s_x^2}\right) = 1 - \alpha.$$

Prendendo termine a termine il reciproco (e ricordandoci che così facendo dobbiamo invertire la direzione delle diseguaglianze) troviamo

$$\mathrm{P}\left(\frac{1}{F_{1-\alpha/2}(n-1, m-1)} \frac{s_y^2}{s_x^2} \leq \frac{\sigma^2}{\tau^2} \leq \frac{1}{F_{\alpha/2}(n-1, m-1)} \frac{s_y^2}{s_x^2}\right) = 1 - \alpha.$$

L'ultima uguaglianza mostra che l'intervallo

$$\left[\frac{1}{F_{1-\alpha/2}(n-1, m-1)} \frac{s_y^2}{s_x^2}; \frac{1}{F_{\alpha/2}(n-1, m-1)} \frac{s_y^2}{s_x^2}\right]$$

costituisce un intervallo di confidenza di livello di copertura $1-\alpha$ per il rapporto tra le due varianze.
Applichiamo ora questi risultati ai dati sui cuculi. Conveniamo, come nei lucidi, che "y" indica le uova di cuculo deposte in nidi di pettirosso e "x" le uove di cuculo deposte in nidi di scricciolo. I dati che ci interessano sono

$$\begin{aligned}
n &= 16, & s_y^2 &\approx 0{,}46 \\
m &= 15, & s_y^2 &\approx 0{,}55
\end{aligned}$$

Il rapporto delle due stime delle varianze vale quindi

$$\frac{s_y^2}{s_x^2} \approx \frac{0{,}46}{0{,}55} \approx 0{,}84.$$

Dalla tabella data dei quantili di un F con 15 e 14 gradi di libertà vediamo che questo valore è compreso tra il quantile-0,10 e la mediana. Il valore osservato cade quindi in una regione in cui "ce lo aspettiamo sotto H_0". I dati suggeriscono perciò di non dubitare dell'ipotesi nulla. Alla stessa conclusione ovviamente arriviamo anche utilizzando un approccio "accetto/rifiuto". Ad esempio, se scegliamo $\alpha = 0{,}05$ allora

$$F_{15,14,0{,}025} = 0{,}35 \text{ e } F_{15,14,0{,}975} = 2{,}95$$

e perciò “accettiamo H_0”. In definitiva, le differenze osservate nelle due stime non sono significative, ovvero sono sufficientemente contenute da poter esser attribuite al caso (=all'errore campionario).
Sulla base di quando visto un intervallo di confidenza (livello di copertura 0,99) è

$$\left[\frac{1}{4{,}25}0{,}84; \frac{1}{0{,}24}0{,}84\right] \approx [0{,}20; 3{,}5].$$

L'intervallo di confidenza ovviamente include 1 ovvero ci segnala che le varianze nei due gruppi potrebbero essere uguali. In realtà l'ampiezza dell'intervallo[19] ci segnala anche le scarse informazioni che abbiamo sulle varianze in campioni così piccoli.

[19]ad esempio, l'intervallo ci segnala come “possibile sulla base dei dati” sia il caso in cui σ^2 è un terzo di τ^2 che quello in cui σ^2 è tre volte τ^2.

Unità J

Esercizio 50. Per verificare tre diferenti linee di attacco per l'ipertensione 18 persone di sesso maschile, leggermente sovrappeso (Body Mass Index tra 25 e 30), con abitudini sedentarie (niente attività sportiva, spostamenti prevalentemente in macchina,...) e con problemi di ipertensione (pressione sistolica maggiore di $140 mmHg$) sono state scelte casualmente tra gli abitanti di Padova. Sempre casualmente sono state poi suddivise in tre gruppi:
- il primo gruppo (5 persone) ha seguito uuna terapia farmacologica;
- il secondo gruppo (7 persone) ha seguito una dieta prefissata;
- il terzo gruppo (6 persone) ha seguito la dieta del secondo gruppo ma ha anche svolto regolarmente delle attività fisiche preordinate e controllate dai ricercatori.

La pressione sistolica è stata misurata sia all'inizio che dopo 3 mesi dall'ingresso nello studio. La seguente tabella mostra, per ognuno dei 18 partecipanti allo studio, la differenza tra la pressione iniziale e quella rilevata dopo 3 mesi.

	soggetto						
gruppo	1	2	3	4	5	6	7
solo farmaco	21	20	7	11	16		
solo dieta	-9	13	1	2	24	6	9
dieta+esercizio fisico	19	18	21	8	8	18	

(a) A chi sono generalizzabili questi risultati?
(b) Perchè secondo voi è stata utilizzata la differenza tra la pressione “iniziale” e la pressione “finale” e non semplicemente quest'ultima?
(c) Verificare la significatività delle differenze tra le medie utilizzando l'analisi della varianza ad un criterio.
(d) Nel condurre l'analisi della varianza vi dovrebbe essere venuto un dubbio sul fatto che le ipotesi su cui si basa non siano completamente soddisfatte nel caso in esame. Spiegare perchè.

Schema di soluzione.

(a) La popolazione di riferimento per lo studio è descritta nel testo dell'esercizio: maschi, leggermente sovrappeso, con abitudini sedentarie e con problemi di ipertensione residenti a Padova. I risultati sulla differenza o meno di efficacia delle tre linee di attacco considerate che ci possiamo far raccontare dai dati sono generalizzabili direttamente solamente a questi individui. Estensioni ad altri gruppi di individui costituiscono pure ipotesi.

(b) La pressione "finale", ovvero dopo tre mesi di terapia, risente ovviamente sia dell'effetto della terapia che della pressione "iniziale". Ad esempio, per un gioco sfortunato del caso, potrebbe accadere di assegnare al gruppo "solo farmaco" i 5 pazienti con la pressione "iniziale" più elevata. In questo caso, potremmo arrivare alla fine a concludere che la terapia farmacologica è inferiore anche nelle situazioni in cui è migliore o paragonabile alle altre semplicemente perchè l'effetto delle condizioni di inizio non è stato "ribaltato" completamente dalla terapia. Considerando le differenze tra "inizio" e "fine" si cerca viceversa di evitare che i risultati dipendano da differenti stati iniziali dei pazienti.
(c) Calcoliamo, separatamente per i tre gruppi, la media e la varianza campionaria.

gruppo	media campionaria	varianza campionaria
solo farmaco	15,0	35,5
solo dieta	6,6	107,6
dieta+esercizio fisico	15,3	33,5

Procediamo con l'analisi della varianza ad un criterio.

$$\begin{aligned}
&\text{media totale} = \frac{1}{18}(5 \times 15{,}0 + 7 \times 6{,}6 + 6 \times 15{,}3) \approx 11{,}8,\\
&\text{devianza tra i gruppi} =\\
&\quad = 5 \times (15{,}0 - 11{,}8)^2 + 7 \times (6{,}6 - 11{,}8)^2 + 6 \times (15{,}3 - 11{,}8)^2 = 313{,}98\\
&\text{devianza entro i gruppi} =\\
&\quad = (5-1) \times 35{,}5 + (7-1) \times 107{,}6 + (6-1) \times 33{,}5 = 955{,}10\\
&\text{statistica test } (F_{oss}) =\\
&\quad = \frac{313{,}98/(3-1)}{955{,}10/(18-3)} \approx 1{,}64.
\end{aligned}$$

Il valore calcolato della statistica test va confrontato con i valori "previsti" da una F di Snedecor con 2 gradi di libertà al numeratore e 18 al denominatore. Una scorsa alla tavola dei quantili di questa distribuzione mostra che 1,64 è compreso tra il quantile-0,75 e il quantile-0,90. Il livello di significatività osservato, che in questo caso è

$$P(F(2, 15) \geq 1{,}64),$$

è quindi a sua volta compreso tra 0,10 e 0,25. Con questi valori tipicamente si accetta l'ipotesi nulla. Ovvero, le differenze tra le medie dei gruppi non sembrano significative[20] se si utilizza l'analisi della varianza ad un criterio[21].
(d) Le ipotesi su cui si basa l'analisi della varianza ad un criterio sono:
- la normalità della distribuzione del fenomeno di interesse all'interno dei gruppi;
- l'uguaglianza delle varianze all'interno dei gruppi;
- l'indipendenza di tutte le osservazioni (tra e nei gruppi).

La varianza campionaria del gruppo "solo dieta" è tre volte quella degli altri due gruppi. L'ipotesi di varianze uguali è quindi sospetta.

[20] non lo sono, ad esempio, ne all'1%, ne al 5% e e neanche al 10%.
[21] per calcolare il livello di significatività osservato con R dovremmo usare il comando `1-pf(1.64,2,15)`. Il valore otenuto 0,2269149 conferma, ovviamente, quanto detto.

Esercizio 51. Un certo gruppo di polli è stato suddiviso casualmente in tre gruppi. Successivamente i tre gruppi sono stati alimentati con tre diete differenti (in realtà le tre diete avevano una base comune ma si differenziavano per un singolo componente). Alcuni dati di sintesi sono riportati nella seguente tabella.

dieta	numero di polli alimentati con la dieta	media campionaria degli incrementi di peso dopo 6 settimane	varianza campionaria degli incrementi di peso dopo 6 settimane
A	14	218,8	2728,6
B	9	246,4	2930,0
C	10	328,9	2385,1

I pesi sono stati rilevati in grammi.
Ipotizzando che sia possibile assumere che valgano le ipotesi su cui fonda l'analisi della varianza ad un criterio, verificare se le differenze negli incrementi di peso possano essere attribuiti al caso.

Schema di soluzione.

Calcoliamo,

$$\begin{aligned}
&\text{media totale} = \frac{1}{33}(14 \times 218{,}8 + 9 \times 246{,}4 + 10 \times 328{,}9) \approx 259{,}7,\\
&\text{devianza tra i gruppi} =\\
&\quad = 14 \times (218{,}8 - 259{,}7)^2 + 9 \times (246{,}4 - 259{,}7)^2 + 10 \times (328{,}9 - 259{,}7)^2 \approx\\
&\quad \approx 72897{,}7\\
&\text{devianza entro i gruppi} =\\
&\quad = (14-1) \times 2728{,}6 + (9-1) \times 2930{,}0 + (10-1) \times 2385{,}1 \approx\\
&\quad \approx 80377{,}7
\end{aligned}$$

ed infine la statistica test su cui è basata l'analisi della varianza ad un criterio

$$F_{oss} = \frac{72897{,}7/(3-1)}{80377{,}7/(33-3)} \approx 13{,}6.$$

Quest'ultimo valore deve essere confrontato con "quelli prevedibili" da una F di Snedecor con 2 gradi di libertà al numeratore e 30 al denominatore. Consultando una tabella dei quantili di una F troviamo che questo valore è "grande" rispetto ai valori previsti. Ad esempio il quantile-0,999 di questa distribuzione è 8,77. Ovvero, in assenza di differenze nell'efficacia delle diete, ci aspettiamo di osservare un valore così grande meno di una volta ogni mille replicazioni dell'esperimento (il livello di significatività osservato è minore di 1/1000). Un risultato così "estremo" è usualmente considerato altamente significativo contro l'ipotesi nulla.

Unità K

Esercizio 52. Per verificare se inseminare le nuvole con nitrato d'argento aumenta la quantità di pioggia "rilasciata" si è rilevata la quantità di pioggia per 5 nuvole inseminate e per 5 non inseminate[22]. I dati sono ottenuti sono stati i seguenti:

	nuvola				
gruppo	1	2	3	4	5
senza inseminazione	244,3	163,0	147,8	29,1	11,5
con inseminazione	489,1	430,0	242,5	129,0	31,4

L'incremento, visibile ad occhio dai dati, della pioggia nelle nuvole inseminate è significativo al 5%?
Rispondere assumendo l'indipendenza di tutte le osservazioni ma non la normalità dei dati.

Schema di soluzione.

Siamo interessati al confronto "in posizione" tra due gruppi di osservazioni. Visto che ci viene richiesto di procedere senza assumere la normalità dei dati e che la numerosità campionaria è piccola utilizzeremo il test di Wilcoxon a due campioni tenendo presente che ci viene chiesto di verificare una ipotesi unilaterale (la pioggia è superiore per le nuvole inseminate).
Per il calcolo della statistica test è comodo riorganizzare i dati come nella seguente tabella:

osservazioni ordinate	con o senza inseminazione	rango
11,4	senza	1
29,1	senza	2
31,4	con	3
129,0	con	4
147,8	senza	5
163,0	senza	6
242,5	con	7
244,3	senza	8
430,0	con	9
489,1	con	10

[22] in realtà si tratta di un sottoinsieme di dati effettivamente rilevati; nell'esperimento originale, condotto in Australia, erano stati considerate 52 nuvole e ne erano state inseminate 26 scelte casualmente.

La somma dei ranghi delle osservazioni provenienti dalle nuvole inseminate è quindi

$$s = 3 + 4 + 7 + 9 + 10 = 33.$$

La statistica test vale

$$W = s - \frac{\text{(num. nuvole inseminate)(num. nuvole inseminate - 1)}}{2} = 33 - \frac{5 \times 4}{2} = 23.$$

Calcolo e discussione del livello di significatività osservato. Da una tavole dei quantili della distribuzione di W calcolata ipotizzando che i due gruppi abbiano la stessa distribuzione troviamo che 22 è quantile-0,975 di questa distribuzione. Il livello di significatività osservato, che in questo caso visto che l'ipotesi è unilaterale è

$$P(W \geq 23) = P(W > 22)$$

vale quindi 0,0275. L'incremento osservato della pioggia nelle nuvole "inseminate" è quindi significativo al 5%.

Discussione utilizzando un livello di significatività uguale al 5%. In questo caso (ipotesi unilaterale) rifiutiamo per valori "troppo grandi" della statistica test. Quindi, fissando $\alpha = 0{,}05$, dobbiamo confrontare il valore calcolato di W con il quantile-0,95 della distribuzione di W calcolata ipotizzando che i due gruppi abbiano la stessa distribuzione, ovvero che l'incremento osservato sia dovuto solamente al caso. Da una tavole dei quantili di questa distribuzione troviamo che il quantile desiderato vale 20. Il valore calcolato, 23, è maggiore e quindi l'incremento è significativo al 5%.

Esercizio 53.

Per confrontare due gruppi di osservazioni, ciascuno di numerosità uguale a 6, è stata calcolata la statistica del test di Wilcoxon ottenendo come risultato 14. Sapendo che il sistema di ipotesi sotto verifica è

- H_0: la distribuzione del fenomeno di interesse è uguale nei due gruppi
- H_1: la distribuzione del primo gruppo è spostata verso l'alto o verso il basso rispetto a quella nel secondo gruppo

dire se i risultati sono significativi al 5%.

Schema di soluzione.

Si indichi con $W_{oss} = 14$ il valore osservato della statistica test. Tenendo conto che il sistema d'ipotesi è bilaterale il risultato ottenuto è significativo al 5% se

$$\text{o } W_{oss} < W_{0{,}025}(6,6) \text{ o } W_{oss} > W_{0{,}975}(6,6)$$

dove $W_p(n,m)$ indica il quantile-p della distribuzione sotto l'ipotesi nulla quando le numerosità dei due gruppi siano rispettivamente n e m.
Da una tavola dei quantili della statistica test troviamo che

$$W_{0.975}(6,6) = 30.$$

La distribuzione sotto H_0 è simmetrica intorno a

$$\frac{\text{num. primo gruppo)(num. secondo gruppo)}}{2}$$

ovvero in questo caso intorno a

$$\frac{6 \times 6}{2} = 18.$$

Quindi

$$W_{0{,}025}(6,6) = 18 - (W_{0{,}975(6{,}6)} - 18) = 18 - (30 - 18) = 6.$$

Infatti, la simmetria della distribuzione implica che la distanza del quantile-0,025 dalla mediana è uguale alla distanza del quantile-0,975 dalla mediana.
Il valore osservato, 14, è ovviamente compreso tra 6 e 30. I risultati non sono quindi significativi al 5%.

Esercizio 54.

Per confrontare le osservazioni di due gruppi è stato utilizzato il test di Wilcoxon a due campioni ottenendo un livello di significatività osservato minore di 0,001. Sapendo che il sistema di ipotesi sotto verifica era

- H_0: la distribuzione del fenomeno di interesse è uguale nei due gruppi
- H_1: la distribuzione del primo gruppo è spostata verso l'alto rispetto a quella nel secondo gruppo

disegnare, su di uno stesso piano cartesiano, degli ipotetici diagrammi a scatola con baffi (boxplot) delle osservazioni del primo e del secondo gruppo compatibili con questo risultato.

Schema di soluzione.

Un valore così piccolo del livello di significatività osservato indica che i dati ci stanno suggerendo di accettare H_1. Quindi, le osservazioni provenienti dal primo gruppo nel campione sono risultate "più grandi" rispetto alle osservazioni provenienti dal secondo gruppo. Una ipotetica coppia di boxplot in cui questo accade, e quindi compatibile con le informazioni date nel testo, è disegnato nella figura 13.

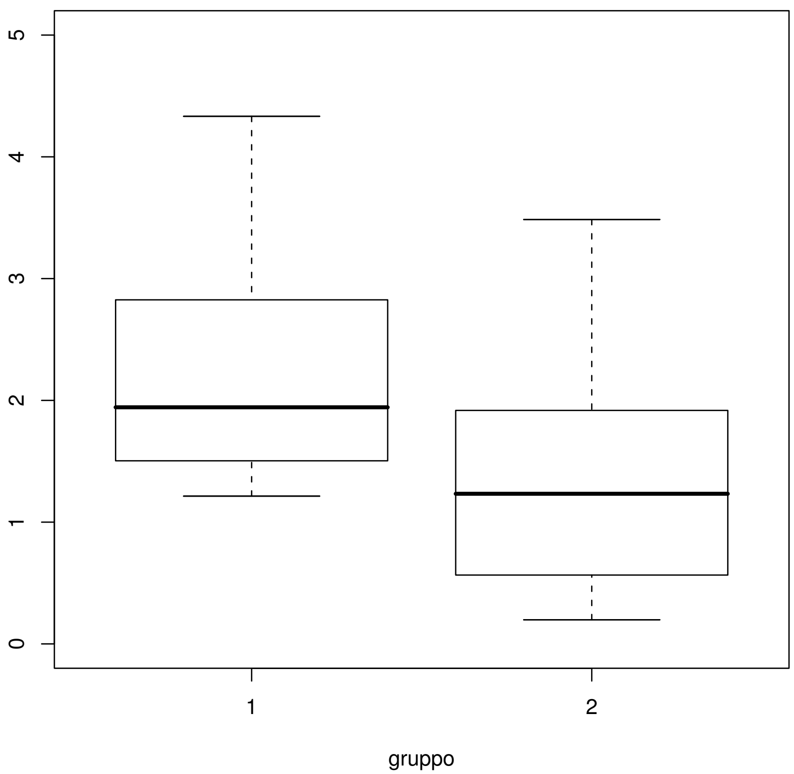

Figura 13: Due ipotetici diagrammi a scatola con baffi.

Esercizio 55. Due differenti diete, chiamiamole A e B, sono state prescritte a 20 pazienti obesi (per la precisione, a 10 pazienti scelti a caso è stata prescritta la prima dieta, agli altri 10 la seconda). La riduzione di peso (in Kg) dopo 12 mesi è risultata

	paziente									
gruppo	1	2	3	4	5	6	7	8	9	10
dieta A	-30	23	24	26	28	29	30	31	32	33
dieta B	10	11	14	15	16	17	18	20	21	22

Applicando il test t di Student a due campioni a questi dati troviamo che la differenza in media tra i risultati delle due diete non è significativa (livello di significatività osservato uguale a 0,33. Viceversa, utilizzando il test di Wilcoxon a due campioni troviamo un livello di significatività uguale a 0,0015 che quindi risulta altamente significativo.
(a) Guardando i dati si spieghi perchè questa discordanza di risultati e si dica quale dei due test dovrebbe essere più affidabile.
(b) Come esprimereste verbalmente i risultati presentati?

Schema di soluzione.

(a) La risposta in breve è che la prima osservazione per la dieta A è anomala. Tutti gli altri 19 pazienti hanno avuto una riduzione di peso. Viceversa, un paziente trattato con la dieta A è aumentato di addirittura $30kg$ (forse non ha seguito la dieta!). Si osservi inoltre che gli altri 9 pazienti del gruppo “A” hanno avuto tutti una riduzione di peso superiore ai 10 pazienti del gruppo “B”. Ovvero, se si trascura il “paziente anomalo” i dati sembrano chiaramente indicare che esiste una differenza di efficacia tra le due diete e che la dieta A è migliore. La situazione è illustrata nella figura 14 che mostra i boxplot delle osservazioni nei due gruppi.
Purtroppo, la statistica test utilizzata nel test t dipende dalle medie e dalle varianze campionarie. Queste quantità sono fortemente influenzate da osservazioni anomale. Ad esempio, la media e la varianza campionarie di tutte le “riduzioni di peso” dei pazienti che hanno utilizzato la dieta A valgono, rispettivamente, 22,6 e 352,5. Se escludiamo la prima osservazione (ovvero il paziente che è aumentato di peso) otteniamo viceversa 28,4 e 12,3. Si osservi in particolare l’enorme diminuzione della varianza. La non significatività del test t è legata a questo. Ad esempio, se confrontiamo i due gruppi escludendo però l’osservazione anomala otteniamo un livello di significatività osservato minore di 0,00001 ovvero le differenze tra i due gruppi sono altamente significative. In una certa misura, questo ha a che fare con il fatto che il test t, almeno quando la numerosità è così piccola, è basato su di una assunzione di normalità e che, una distribuzione normale non

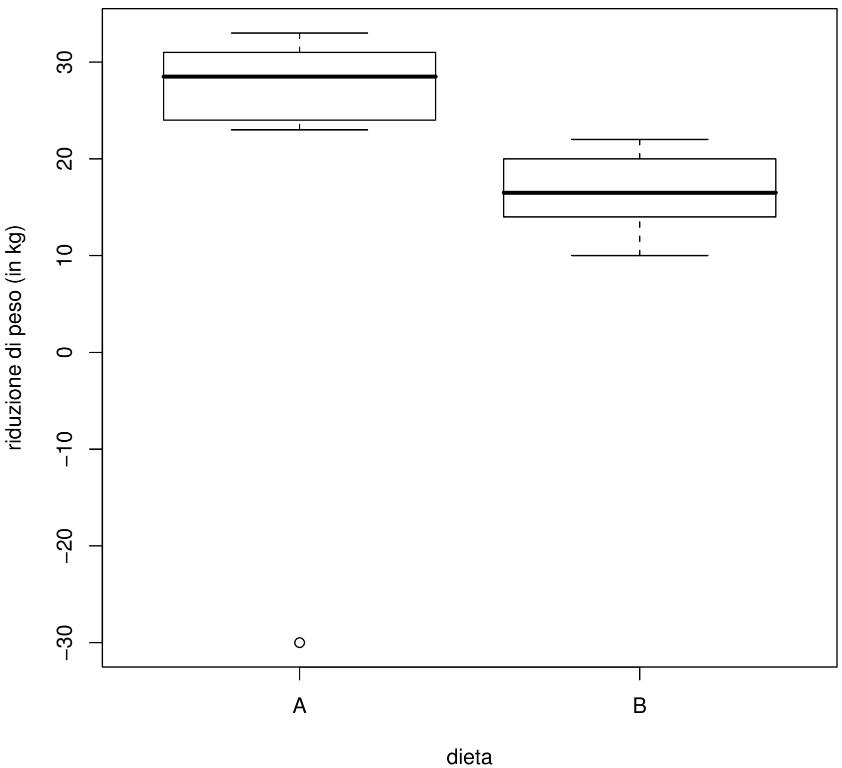

Figura 14: Confronto tra due diete. Diagramma a scatola con baffi delle osservazioni dei due gruppi. Si osservi come il boxplot per il gruppo “dieta A” metta in luce la presenza di una osservazione anomala. Ed inoltre, come, se non si tiene presente questa osservazione, il grafico suggerisca una migliore efficacia per la “dieta A”.

è in grado di spiegare eventuali valori anomali. Si vedano al proposito le figure 15, 16 e 17.

Però in questo caso il problema non è solo la non normalità. Il punto è che il test t confronta medie e che le medie non sono una misura di posizione appropriata nel caso di valori anomali. Si riguardi ad esempio la media di tutte le osservazioni “A”. Il suo valore, 22,6, è più piccolo di tutte le osservazioni “buone” (ovvero le 9 positive). Quindi non ci racconta dove queste osservazioni sono posizionate. Nello stesso tempo è lontana dall'osservazione “cattiva”. Quindi, ci racconta poco anche rispetto a quest'ultima.

Il test di Wilcoxon è più “resistente” rispetto alle osservazioni anomale: la statistica test dipende solo dai ranghi delle osservazioni non dalle osservazioni stesse. Quindi ad esempio, se anche una osservazione “andasse ad infinito”, ovvero fosse enormemente grande, il test ne risentirebbe di molto poco. Ad esempio, in questo caso, il livello di significatività eliminando la prima osservazione del gruppo “A” risulta uguale a 0,00002, ovviamente differente rispetto al valore riportato nel testo che era basato su tutte le osservazioni ma non in maniera tale da farci cambiare le conclusioni.

In definitiva, i risultati suggeriti dal test di Wilcoxon sembrano più affidabili. La non significatività del test t riportata nel testo dell'esercizio non sembra quindi essere una caratteristica del gruppo ma semplicemente dipendere dalla risposta di un solo paziente.

(a) Una possibile presentazione dei risultati è

> *I dati raccolti indicano che la differenza tra i risultati ottenibili con le due diete è significativa (p-value del test di Wilcoxon a due campioni* $< 0{,}01$*). In particolare, dei* 10 *pazienti che hanno utilizzato la dieta A,* 9 *hanno avuto una riduzione di peso superiore a tutte le riduzioni di peso rilevate sui* 10 *pazienti che hanno utilizzato la dieta B. Quindi, lo studio suggerisce una maggiore efficacia della dieta A. Deve però essere notato che un paziente a cui era stata somministrata la dieta A non ha reagito minimamente alla dieta e anzi è aumentato di* $30kg$ *durante il periodo di osservazione. E quindi possibile che ci siano dei pazienti “resistenti” a questa dieta.*

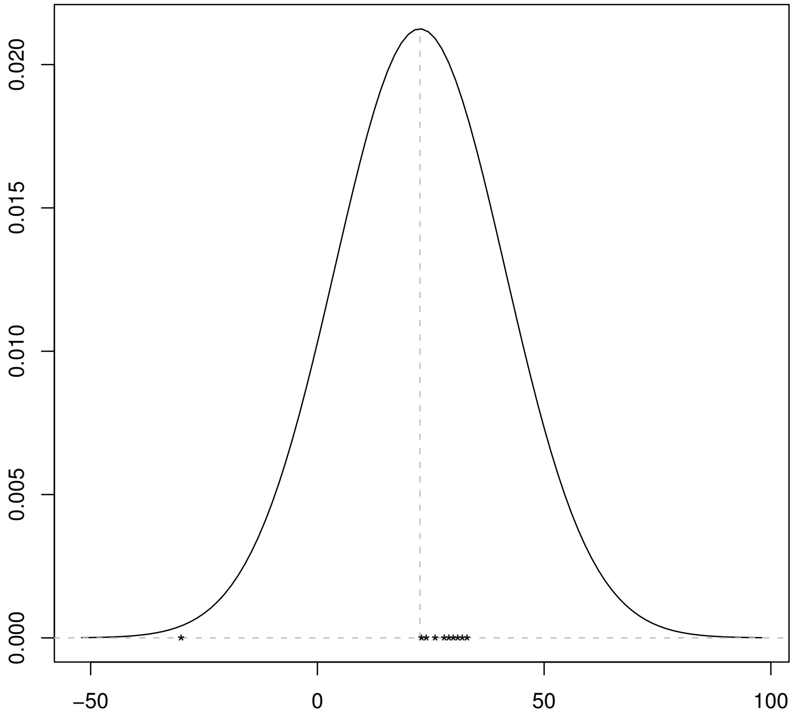

Figura 15: Confronto tra due diete. La densità disegnata è quella di una normale con media e varianza uguali alla media e varianza campionaria di tutte le osservazioni del gruppo dieta “A”. Gli asterischi sull’asse delle ascisse sono disegnati in corrispondenza delle osservazioni. Si osservi come dalla distribuzione disegnata sia difficile aspettarci le osservazioni. In particolare, le 9 osservazioni “buone” sono troppo vicine rispeto a quanto ci si aspetterebbe ed inoltre sono tutte più grandi della media.

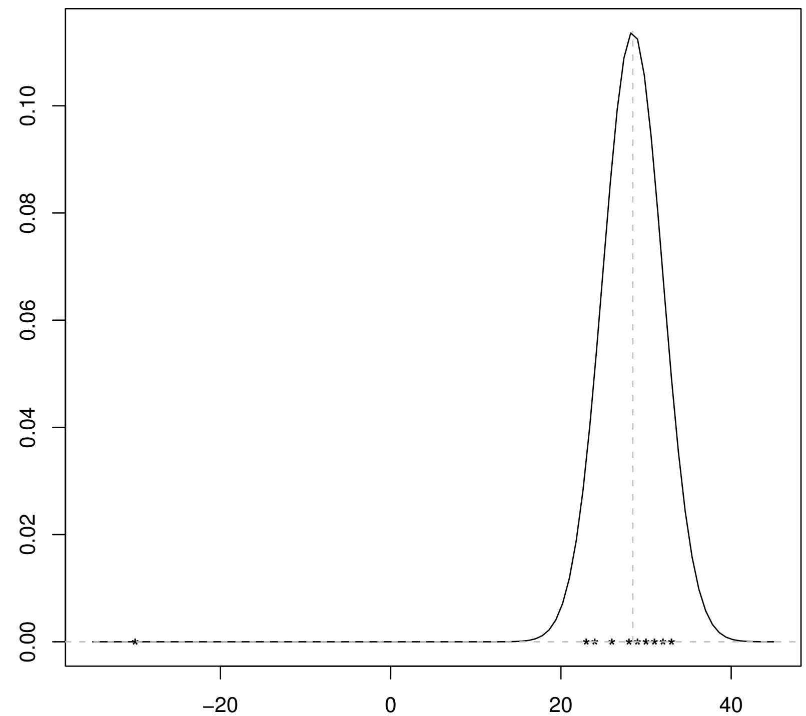

Figura 16: Confronto tra due diete. La densità disegnata è quella di una normale con media e varianza uguali alla media e varianza campionaria delle 9 osservazioni “buone” del gruppo dieta “A”. Gli asterischi sull’asse delle ascisse sono disegnati in corrispondenza delle osservazioni. La densità disegnata non è in grado si spiegare l’osservazione anomala.

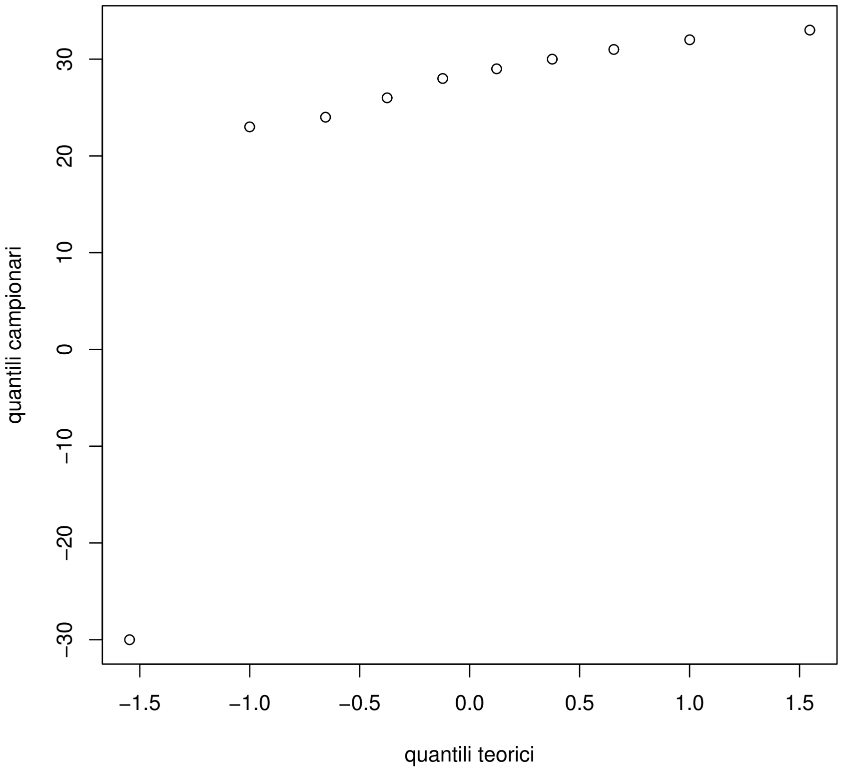

Figura 17: Confronto tra due diete. *Normal probability plot* delle osservazioni del gruppo dieta “A”. Si osservi l’effetto dell’osservazione anomala.