alcuni esercizi
(unità g-j)

guido masarotto

2 giugno 2004

---

1. 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

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

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 (unità g). 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.

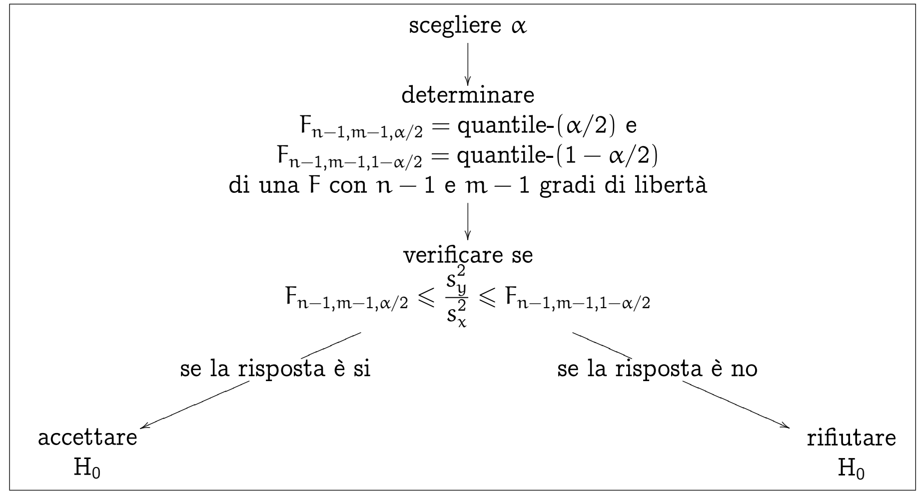

Figura 1: Confronto della varianza di due variabili casuali normali

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

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

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 1. La procedura stilizzata nella figura 1 garantisce che

$$\begin{aligned}
&P(\text{accettare } H_0 \text{ quando } H_0 \text{ è vera}) = \\
&= P\left(F_{n-1,m-1,\alpha/2} \leqslant \frac{s_y^2}{s_x^2} \leqslant F_{n-1,m-1,1-\alpha/2} | \; H_0 \text{ vera}\right) = \\
&= P(F_{n-1,m-1,\alpha/2} \leqslant \text{F con } n-1 \text{ e } m-1 \text{ gradi di libertà} \leqslant F_{n-1,m-1,1-\alpha/2}) = \\
&= 1 - \alpha.
\end{aligned}$$

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

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

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

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

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

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

L'ultima uguaglianza mostra che l'intervallo

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

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 (unità g dei lucidi). 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

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

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).

Tabella 1: Effetto di tre differenti diete sulla crescita dei polli

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

Nota: le varianze campionarie sono state calcolate dividendo per $n-1$.

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[1] ci segnala anche le scarse informazioni che abbiamo sulle varianze in campioni così piccoli.

---

2. 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 tabella 1. 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. Poniamo

$$\begin{aligned} n &= \text{numero complessivo di osservazioni} = 14+9+10 = 33, \\ k &= \text{numero delle diete} = 3. \end{aligned}$$

Quindi calcoliamo,

$$\text{media totale} = \frac{1}{33}(14 \times 218{,}8 + 9 \times 246{,}4 + 10 \times 328{,}9) \approx 259{,}7,$$

$$\begin{aligned} &\text{varianza tra i gruppi} = \\ &= \frac{1}{33}\left[14 \times (218{,}8-259{,}7)^2 + 9 \times (246{,}4-259{,}7)^2 + 10 \times (328{,}9-259{,}7)^2\right] \approx \\ &\approx 2209{,}0 \end{aligned}$$

[1] 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.

varianza entro i gruppi =

$$= \frac{1}{33}[(14-1)\times 2728{,}6+(9-1)\times 2930{,}0+(10-1)\times 2385{,}1]\approx$$
$$\approx\ 2435{,}7,$$

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

$$F_{oss} = \frac{2209{,}0/(3-1)}{2435{,}7/(33-3)} = 13{,}6.$$

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[2] 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. Un risultato così "estremo" è usualmente considerato altamente significativo contro l'ipotesi nulla.

---

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

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

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.

[2] ad esempio quella riportata in "formule e tavole per gli esami"

Schema di soluzione.
(a) Sintesi del test t a due campioni:
- dati: due campioni $y_1,\dots,y_n$ e $x_1,\dots,x_n$ provenienti da due popolazioni;
- normalità e indipendenza:

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

  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).
(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 2.

---

4. 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 (i dati sono nei lucidi, unità g).
Schema di soluzione. Lavoriamo "sotto le ipotesi" e con le notazioni dell'unità g:
(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

$$y_1,\dots,y_n \text{ determinazioni i.i.d. di una } N(\mu,\sigma^2)$$

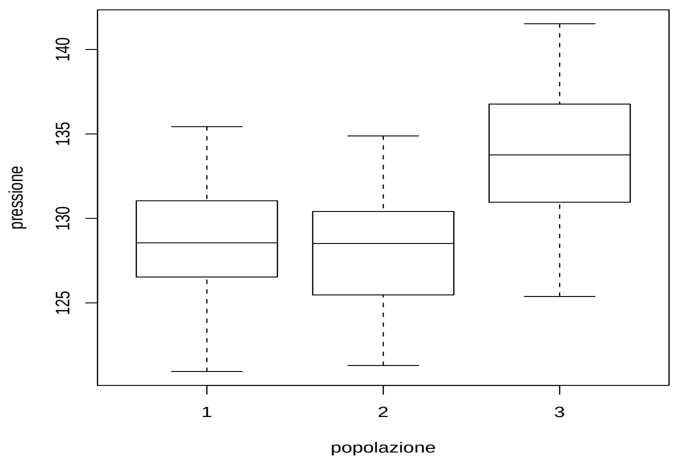

Figura 2: Esempi di diagramma a scatola con baffi

$x_1, \ldots, x_m$ determinazioni i.i.d. di una $N(\eta, \tau^2)$
$y_1, \ldots, y_n$ indipendenti da $x_1, \ldots, x_m$

(iii) varianze uguali[3] $\sigma^2 = \tau^2$.
Sappiamo allora che

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

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

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

[3]visto anche la verifica fatta nel primo esercizio di questo blocco.

Indichiamo con $t_{g,p}$ 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 $t_{g,1-p} = -t_{g,p}$,

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

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

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

ovvero che

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

è un intervallo di confidenza (con livello di copertura $1-\alpha$) per la differenza delle due medie.
Nel caso dei dati che stiamo considerando:

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

Quindi,

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

Ponendo $\alpha = 0{,}10$ troviamo $t_{29,0,95} = 1{,}70$ 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

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

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

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

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 $t_{28,0,95} = 1{,}70$, 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.

---

5. Una industria farmaceutica produce delle pillole che dovrebbero contenere 50mg di una certa sostanza attiva. Durante un controllo, sono state prelevate a caso 8 pillole tra tutte quelle prodotte nei due giorni precedenti (4 per ogni giorno). E' stato poi dosato il contenuto di sostanza attiva per ciascuna pillola. I risultati ottenuti sono stati 49,0, 48,6, 49,3 e 49,1 per le pillole del primo giorno e 49.8, 50.3, 49.9 e 50.2 per quelle del secondo giorno.

(a) Dire, utilizzando una appropriata procedura statistica, se vi sembra plausibile che la media della sostanza attiva contenute nelle pillole del primo giorno controllato sia uguale a 50 (ovvero al valore nominale dichiarato dall'azienda sulle confezioni).

(b) Dire, utilizzando una appropriata procedura statistica, se vi sembra plausibile che la media della sostanza attiva contenuta nelle pillole del primo giorno sia uguale alla media del secondo giorno.

Rispondere sia assumendo che non assumendo la normalità della distribuzione della sostanza attiva. Nelle risposta, se serve, è viceversa possibile sempre assumere che la la sostanza attiva contenuta nelle pillole abbia una varianza costante (non dipendente dal giorno di produzione) e media uguale per tutta la produzione di un giorno ma che può variare da giorno a giorno.
Schema di soluzione. Siamo nel contesto dei test t e di Wilcoxon a un campione per la prima domanda e a due per la seconda. Utilizziamo R per fare i calcoli.


```
> y1 <- c(49,48.6,49.3,49.1)
```

Commento sui dati: tutte le 4 pillole del 1° giorno contengono una quantità di sostanza attiva inferiore al valore nominale.

```
> t.test(y1,mu=50)

        One Sample t-test

data:  y1
t = -6.7937, df = 3, p-value = 0.00652
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 48.53156 49.46844
sample estimates:
mean of x
       49
```

Commento sul test t *ad un campione:* il test è altamente significativo; ipotizzando la normalità dei dati i dati sembrano suggerire che la vera media è differente dal valore nominale (50).

```
> wilcox.test(y1,mu=50)

        Wilcoxon signed rank test

data:  y1
V = 0, p-value = 0.125
alternative hypothesis: true mu is not equal to 50
```

Commento sul test di Wilcoxon a un campione: il valore della statistica test è il più piccolo possibile ($V=0$); infatti nessuna delle 4 osservazioni è più grande del valore di μ previsto da H_0; nonostante questo il risultato non è significativo ($p-value = 0{,}125$); dato che nessun risultato più estremo contro l'ipotesi nulla esiste (oltre al simmetrico valore massimo) questo risultato mostra che il test di Wilcoxon ad un campione con 4 osservazioni non è in grado di fornire mai un risultato significativo (i dati sono troppi pochi), e quindi, per contrasto con quanto visto sopra per il test t che può essere meno potente di quest'ultimo; del resto, per evitare di utilizzare la normalità dei dati, utilizza meno informazioni; si osservi ad esempio come in questo caso l'informazione "incorporata" nella statistica test ($V=0$) sia "solamente" che tutte le osservazioni erano risultate più piccole di 50; infatti per qualsiasi quaterna di osservazioni tutte più piccole di 50 otteniamo lo stesso valore della statistica test (sono tutte equivalenti dal punto di vista del test); viceversa il test t di Student "incorpora" anche informazioni sulla distanza delle osservazioni da 50 e sulla loro variabilità interna.

```
> y2 <- c(49.8,50.3,49.9,50.2)
```

Commento sui dati del secondo giorno. La più piccola osservazione del secondo giorno è risultata maggiore della più grande osservazione del primo giorno: "a occhio" sembrerebbe esserci stata una variazione.

```
> t.test(y1,y2,var.equal=TRUE)

        Two Sample t-test

data:  y1 and y2
t = -5.5468, df = 6, p-value = 0.001450
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.5131929 -0.5868071
sample estimates:
mean of x mean of y
    49.00     50.05

> t.test(y1,y2)

        Welch Two Sample t-test

data:  y1 and y2
t = -5.5468, df = 5.748, p-value = 0.001670
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.5181564 -0.5818436
sample estimates:
mean of x mean of y
    49.00     50.05

> wilcox.test(y1,y2)

        Wilcoxon rank sum test

data:  y1 and y2
W = 0, p-value = 0.02857
alternative hypothesis: true mu is not equal to 0
```

Comenti sui test a due campioni: sia assumendo che non assumendo una varianza uguale[4] il test t indica una variazione nella media altamente significativa; significativo è anche il risultato del test di Wilcoxon a due campioni; si osservi tra l'altro che come nel caso precedente anche in questo caso il valore della statistica test di Wilcoxon è ottenibile senza calcoli: nessuna delle osservazioni del primo giorno è maggiore di una osservazioni del secondo giorno e quindi deve risultare $W = 0$.

[4]il caso con varianza disuguale non era comunque richiesto nel testo dell'esercizio.

6. Una azienda ha aperto un nuovo negozio in una certa città. Prima dell'apertura, si prevedeva che le vendite avrebbero dovuto essere circa seimila euro per settimana. Nelle prime 12 settimane le vendite sono state le seguenti

settimana											
1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°
5846	5945	10946	4795	4677	5723	4946	4826	5690	4716	4666	5973
media delle vendite= 5729,08							varianza delle vendite= 2991111				

la varianza è stata calcolata dividendo per "$n-1$"

Secondo voi la previsione fatte sulle vendite è stata confermata dalle vendite effettive? Rispondere utilizzando sia il test t che il test di Wilcoxon (ambedue nella versione "a un campione") e commentando i risultati.
Schema di soluzione. Il sistema d'ipotesi sotto verifica è

$$H_0 : \mu = 6000 \text{ verso } H_1 : \mu \neq 6000$$

dove μ indica la mediana[5] della distribuzione da cui provengono i dati.
Soluzione via t *di Student.* La statistica test è

$$t_{oss} = \frac{5729{,}08 - 6000}{\sqrt{2991111/12}} \approx -0{,}54.$$

Il valore dovrebbe essere confrontato con quelli "previsti" da una t con 11 gradi di libertà. Consultando rapidamente una tabella dei quantili di questa distribuzione o anche semplicemente ricordandoci che una t è simile ad una normale solo un poco schiacciata concludiamo che il risultato è assolutamente non significativo. Ad esempio, utilizzando la tabella contenuta in "formule e tavole per l'esame" troviamo che

$$\text{liv. sign. osservato} = 2P(t_{11} \geqslant 0{,}54) \approx 0{,}6.$$

Soluzione via test di Wilcoxon a un campione. La statistica test può essere calcolata per semplice ispezione dei dati. Si osservi infatti che

(i) esiste una sola osservazione più grande di 6000;
(ii) questa osservazione (la terza che vale 10946) è decisamente la più lontana in valore assoluto da 6000 (tutte le altre osservazioni sono contenute tra 4666 e 6000); il suo rango dello scarto assoluto da 6000 è quindi 12 (il massimo rango possibile);
(iii) ma allora

$$V = \sum_{\{i: y_i > 6000\}} \text{rango di } |y_i - 6000| = 12.$$

[5]possiamo parlare della mediana anche nel caso del test t visto che il modello di riferimento prevede la normalità dei dati.

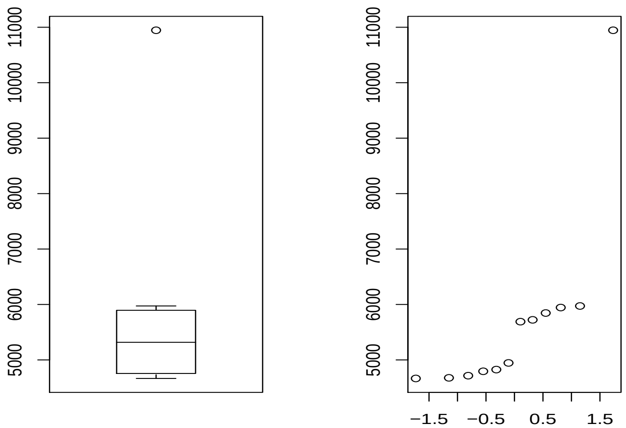

Figura 3: *Boxplot* e *normal probability plot* dei dati sulle vendite

Per la simmetria della distribuzione sotto H_0 di V il livello di significatività osservato vale $2P(V \leqslant 12)$. Dalla tabella sui quantili di V contenuta in "formule e tavole..." vediamo che la mediana di V quando $n = 12$ vale 39. Quindi[6]

$$2P(V \leqslant 12) = 2P(V \geqslant 66).$$

Quindi, dalla stessa tabella troviamo[7] che

$$0{,}02 < \text{liv. sign. osservato} < 0{,}05$$

e che in definitiva il test di Wilcoxon ci fornisce un risultato significativo contro H_0.

A chi credere? I due test forniscono risultati contrastanti. La figura 3 mostra il *boxplot* delle vendite. E' chiaro che l'unica osservazione maggiore di 6000 si configura come una osservazione anomala (= generata da un meccanismo differente da quello che ha generato le altre osservazioni)

Uno degli effetti dell'osservazione anomala si concretizza in una esplosione dello scarto quadratico medio campionario (con la terza osservazione vale ovviamente $\sqrt{2991111} \approx 1729{,}5$, se la eliminiamo, ovvero se calcoliamo lo scarto quadratico medio delle 11 osservazioni differenti dalla 3° otteniamo $\approx 566{,}7$ ovvero circa un terzo). Lo scarto quadratico medio campionario entra nel denominatore di t_{oss} e

[6] ci si ricordi che la distribuzione di V è discreta visto V assume solo valori interi.

[7] 66 è compreso tra il quantile-0,975 e il quantile-0,99 della distribuzione di V quando $n = 12$.

quindi, più è grande più "sposta" la statistica test verso lo zero ovvero verso la non significatività.

La statistica del test di Wilcoxon è viceversa più "resistente" rispetto a fenomeni di questo tipo e quindi anche da questo punto di vista fornisce risultati più affidabili (una singola osservazione non può cambiare di troppo i risultati).

Questa non "robustezza", come si usa dire, della t di Student e in realtà di tutte le procedure basate su di una assunzione di normalità diventa evidente se si ripete l'analisi senza la terza osservazione. Il livello di significatività del test t a un campione diventa $< 0{,}01$ e quindi il test diventa significativo Il valore di significatività osservato del test di Wilcoxon a un campione rimane piccolo e quindi niente cambia nelle conclusioni suggerite da questo test (la statistica test diventa ovviamente zero in questo caso visto che nessuna delle osservazioni rimanenti è maggiore di 6000).

---

7. Durante un certo studio clinico è stata rilevata la pressione arteriosa in 30 individui sani di sesso maschile e in 25 individui sani di sesso femminile. Indichiamo con $(x_1, \dots, x_{30})$ le osservazioni ottenute sui maschi e con $(y_1, \dots, y_{25})$ 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 4-(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,

(a)

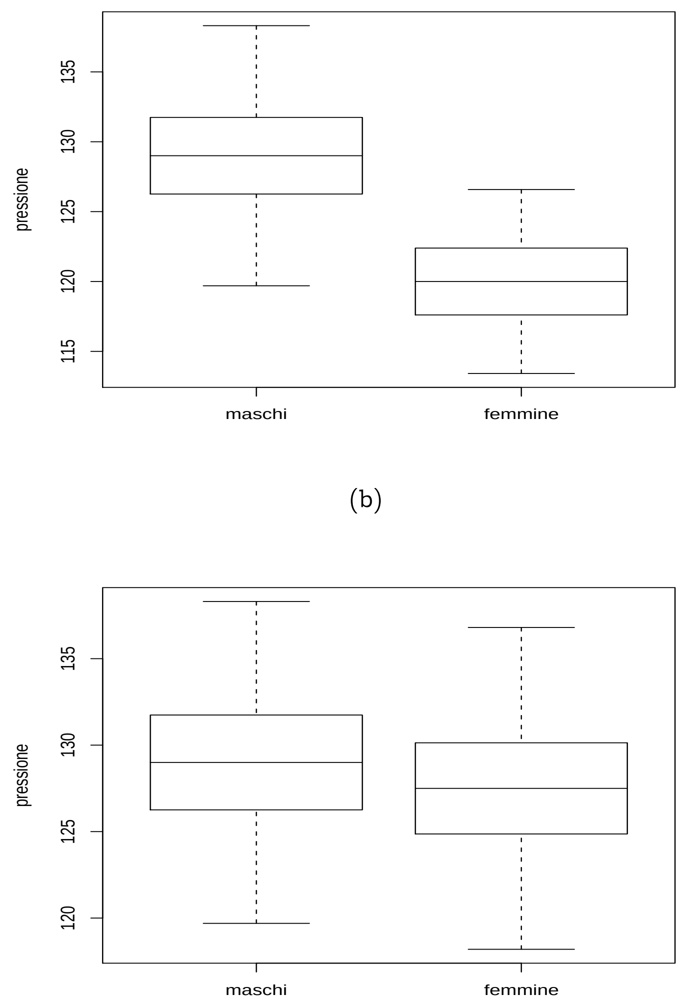

Figura 4: Due coppie di diagrammi a scatola con baffi

"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 4-(b).

---

8. Rispondere si o no alle seguenti domande.

(a) Se un test statistico, condotto con un approccio accetto/rifiuto, ci porta a rifiutare l'ipotesi nulla, possiamo considerare dimostrato che l'ipotesi nulla è falsa?

(b) All'aumentare dell livello di copertura "$1-\alpha$" di un intervallo di confidenza aumenta anche l'ampiezza dell'intervallo stesso?

(c) E' vero che in un test statistico,

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

(d) 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?

(e) 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?

(f) Si supponga che per un certo insieme di dati, il livello di significatività osservato di un test sulla media di una normale sia risultato uguale a 0,18. Se con gli stessi dati conducessimo un test in "modalità accetto/rifiuto" usando un livello di significatività (α) pari a 0,05, l'esito potrebbe essere di rifiutare H_0?

Schema di soluzione.

(a) 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. Ad esempio in un approccio accetto/rifiuto la probabilità di errore di I tipo, indicata quasi sempre con α, viene fissata a priori, può essere piccola ma mai nulla.

(b) Si. Nei casi che abbiamo visto questo è stato sempre vero. Ad esempio, l'intervallo di confidenza di livello $1-\alpha$ per la media di una normale (ignota) la varianza è, senza stare a richiamare il significato dei vari simboli oramai visti mille volte,

$$\overline{y} \pm t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$$

e ovviamente, al crescere è $1-\alpha$ cresce $t_{n-1,1-\alpha/2}$ 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!).
(c) No. Le probabilità dei due errori sono addirittura calcolate sotto due ipotesi differenti (quella di I tipo supponendo H_0 vera, quella di II ipotizzando vera H_1).
(d) No. Nei test accetto/rifiuto tipo quello sviluppato nell'unità B la probabilità di errore di I tipo è prefissata (e indicata con α). Non dipende dalla numerosità campionaria.
(e) Si. 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).
(f) No. Supponiamo che il test sia stato basato sulla t di Student (ovvero che la varianza della normale fosse ignota). Indichiamo inoltre con n il numero delle osservazioni. Allora quello che sappiamo, di nuovo senza richiamare la solita notazione, è che

$$\text{livello significatività osservato} = 2P(t_{n-1} \geqslant |t_{oss}|) = 0{,}18.$$

Quindi, per la simmetria delle t di Student

$$P(t_{n-1} \geqslant |t_{oss}|) = 0{,}09.$$

Ma allora $|t_{oss}| < t_{n-1,0.975}$ visto che

$$P(t_{n-1} \geqslant t_{n-1,0.975}) = 0{,}025$$

e quindi in un test accetto/rifiuto condotto con $\alpha = 0{,}05$ accetteremmo di sicuro l'ipotesi nulla.
Osservazione importante. Quanto visto per la t vale sempre. Ovvero, in un approccio accetto/rifiuto verrà sempre accetta l'ipotesi nulla se viene utilizzato un valore di α, la probabilità di errore di I tipo, minore del livello di significatività osservato. O equivalentemente

- sia $\hat{\alpha}$ il livello di significatività osservato di un certo test calcolato con certi dati;
- allora un test accetto/rifiuto caratterizzato da

$$P(\text{accettare } H_0 \text{ quando } H_0 \text{ è vera}) = 1 - \alpha$$

  può essere basato sulla regola

$$\begin{array}{ccc} & \text{verificare se } \hat{\alpha} \geqslant \alpha & \\ \text{se la risposta è si} & & \text{se la risposta è no} \\ \swarrow & & \searrow \\ \text{accettare } H_0 & & \text{rifutare } H_0 \end{array}$$