Inferenza statistica

A. Inferenza statistica per un solo campione
B. Inferenza statistica per due campioni
C. Inferenza statistica per più di due campioni

Inferenza statistica per un solo campione

Inferenza statistica per un singolo campione

A. Inferenza sulla media di una popolazione con:

- Varianza nota (la statistica test utilizzata è la Z)
- Varianza non nota (la statistica test utilizzata è la t)

B. Inferenza sulla varianza (la statistica test utilizzata è χ^2)

C. Inferenza sulla proporzione (la statistica test utilizzata è Z)

Inferenza statistica sulla media per un solo campione

Inferenza sulla media di una popolazione con varianza nota

Si supponga che x sia una variabile casuale con media sconosciuta μ e varianza nota σ^2.

Si intende verificare l'ipotesi che la media sia uguale a un valore standard, ad esempio μ_0.

L'ipotesi può essere formalmente scritta come:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Inferenza sulla media di una popolazione con varianza nota

La procedura per verificare questa ipotesi è quella di selezionare un campione casuale di n osservazioni sulla variabile casuale x calcolare la statistica test:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Si rifiuta H_0 $|Z_0| > Z_{\alpha/2}$, dove $Z_{\alpha/2}$ è il punto percentuale superiore della distribuzione normale standardizzata.

Questa procedura viene talvolta chiamata test Z per un campione.

Inferenza sulla media di una popolazione con varianza nota: esempio

Si supponga che il produttore di certe confezioni voglia assicurarsi che il loro peso sia 75 kg.

Da precedenti esperienze sa che la deviazione standard è pari a 8 kg.

Il sistema di ipotesi potrebbe essere il seguente:

$H_0 : \mu = 75$ kg

$H_1 : \mu \neq 75$ kg

Inferenza sulla media di una popolazione con varianza nota: esempio

Viene estratto un campione casuale di 25 confezioni il cui peso medio è di 79,25 kg. Siccome la varianza della popolazione è nota allora il valore della statistica test è:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{79{,}25 - 75}{\frac{8}{\sqrt{25}}} = 2{,}66$$

Ad un livello α = 0,05, $Z_{(0,025)}$= 1,96

Poiché Z = 2,66 > 1,96, si rifiuta l'ipotesi nulla.

C'è evidenza sufficiente per affermare che il peso (in kg) di certe confezioni sia mediamente più alto di quello stabilito.

Inferenza sulla media di una popolazione con varianza non nota

Si supponga che x sia una variabile casuale normale con media μ e varianza σ^2 sconosciute.

Si intende verificare l'ipotesi che la media sia uguale a un valore standard, ad esempio μ_0.

L'ipotesi può essere formalmente scritta come:

$$H_0 \colon \mu = \mu_0$$

$$H_1 \colon \mu \neq \mu_0$$

Inferenza sulla media di una popolazione con varianza non nota

Poiché la varianza non è nota è necessario supporre che la variabile casuale sia normalmente distribuita.

Il presupposto della normalità è necessario per sviluppare formalmente il test statistico anche se moderate deviazioni dalla normalità non influiranno seriamente i risultati.

Inferenza sulla media di una popolazione con varianza non nota

Poiché σ^2 è sconosciuto può essere stimato tramite s^2.

La distribuzione di riferimento per questa statistica test è la distribuzione t con n-1 gradi di libertà

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Per un fissato livello di significatività l'ipotesi nulla H_0 si rifiuta se $|(t_0)| > t_{\frac{\alpha}{2},n-1}$ dove $t_{\frac{\alpha}{2},n-1}$ denota il punto percentuale superiore dello distribuzione t con n-1 gradi di libertà.

Inferenza sulla media di una popolazione con varianza non nota: esempio

Si supponga che il produttore di tubi di plastica voglia assicurarsi che il loro diametro sia 25 cm.

Il sistema di ipotesi potrebbe essere il seguente:

$H_0 : \mu = 25$ cm

$H_1 : \mu \neq 25$ cm

Inferenza sulla media di una popolazione con varianza non nota: esempio

Viene estratto un campione casuale di 15 tubi di plastica il cui diametro medio è di 28,5 cm e la deviazione standard 4 cm.

Siccome la varianza della popolazione non è nota allora il valore della statistica test è:

$$T = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{28{,}5 - 25}{\frac{4}{\sqrt{15}}} = 3{,}39$$

Ad un livello α = 0,025; $t_{(14;\ 0,025)} = 2{,}145$

Poiché t= 3,39 > 2,145 si rifiuta l'ipotesi nulla.

C'è evidenza sufficiente per affermare che il diametro medio dei tubi di plastica sia mediamente più alto di quello stabilito.

Inferenza statistica sulla varianza per un solo campione

Inferenza sulla varianza di una popolazione

Supponiamo di voler verificare l'ipotesi che la varianza di una distribuzione normale equivale a una costante σ_0^2.

Il sistema di ipotesi è il seguente:

$H_0 : \sigma^2 = \sigma_0^2$

$H_1 : \sigma^2 \neq \sigma_0^2$

La statistica del test per questa ipotesi è

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

dove s^2 è la varianza del campione calcolata da un campione casuale di n osservazioni.

Inferenza sulla <u>varianza</u> di una popolazione

Per un fissato livello di significatività l'ipotesi nulla H_0 si rifiuta se $\chi_0^2 > \chi^2_{\frac{\alpha}{2};\, n-1}$ o se $\chi_0^2 < \chi^2_{1-\frac{\alpha}{2};\, n-1}$

Dove $\chi^2_{\frac{\alpha}{2};\, n-1}$ e $\chi^2_{1-\frac{\alpha}{2};\, n-1}$ sono i punti percentuali superiore $(\alpha/2)$ e inferiore (1-$(\alpha/2)$) della distribuzione chi-quadro con n-1 gradi di libertà.

Inferenza sulla varianza di una popolazione

Se viene specificata un'ipotesi alternativa unilaterale:

$H_1 : \sigma^2 < \sigma_0^2$

Si rifiuterà se: $\chi_0^2 < \chi^2_{1-\alpha;\, n-1}$

Se invece viene specificata un'ipotesi alternativa unilaterale:

$H_1 : \sigma^2 > \sigma_0^2$

Si rifiuterà se: $\chi_0^2 > \chi^2_{\alpha;\, n-1}$

Inferenza sulla varianza di una popolazione

Questo test è molto utile in molte applicazioni di miglioramento della qualità e dei processi.

Ad esempio si consideri una normale variabile casuale con media μ e varianza σ^2.

Se σ^2 è inferiore o uguale a un valore ad esempio σ_0^2 allora la variabilità naturale intrinseca del processo rientra nei requisiti di progettazione e di conseguenza quasi tutta la produzione sarà conforme alle specifiche.

Inferenza sulla varianza di una popolazione

Invece se σ^2 supera σ_0^2 la variabilità naturale del processo supererà i limiti delle specifiche determinando un'alta percentuale di produzione non conforme.

In altre parole la capacità di processo è direttamente correlata alla variabilità del processo.

Questo test può essere utilizzato per analizzare varie altre situazioni simili e formano la base per una procedura di monitoraggio o controllo per la variabilità del processo.

Inferenza sulla varianza di una popolazione: esempio

Un'azienda produttrice di tubi di metallo riscontra una varianza dei tubi prodotti pari a 96 cm.

A seguito dell'introduzione di un nuovo processo si estrae un campione di 30 tubi per verificare se c'è stato un decremento significativo del diametro e si riscontra una varianza campionaria corretta pari a 85 cm.

Decidere ad un livello di significatività α=0,05 se il decremento nella varianza dei diametri è significativo supponendo che la distribuzione dei diametri sia normale.

Inferenza sulla varianza di una popolazione: esempio

Il sistema di ipotesi potrebbe essere il seguente:

$H_0 : \sigma^2 = 96$ cm

$H_1 : \sigma^2 < 96$ cm

Il test da utilizzare è il χ^2 con 29 (30-1) gradi di libertà.

$\chi^2 = \frac{(30-1)*85}{96} = 25{,}677$

Ad un livello $\alpha = 0{,}05$ e 28 gdl $\chi^2_{(0,95;\ 29)} = 17{,}71$

Poiché $\chi^2 = 25{,}677 > 17{,}71$ non si rifiuta l'ipotesi nulla il decremento nella varianza dei diametri non è significativo.

Inferenza statistica sulla proporzione per un solo campione

Inferenza sulla proporzione di una popolazione

Supponiamo di voler verificare l'ipotesi che la proporzione di p una popolazione equivale a un valore standard ad esempio p_0.

Il test che descriveremo si basa sull'approssimazione normale della binomiale.

Se viene estratto dalla popolazione un campione casuale di n elementi e x elementi nel campione appartengono alla classe associata a p, il sistema di ipotesi da testare è il seguente:

$H_0 : p = p_0$

$H_1 : p \neq p_0$

Inferenza sulla proporzione di una popolazione: Esempio

Un’azienda che assembla computer rileva difetti di assemblaggio nel 20% dei casi.

Si decide di utilizzare un nuovo procedimento e si esegue l’osservazione su un campione di 100 computer assemblati estratti da un lotto di 1.000 per essi si rileva che 18 presentano difetti.

Ad un livello di significatività α=0,03 è significativo il decremento nella proporzione di difetti?

Inferenza sulla proporzione di una popolazione: Esempio

Il sistema di ipotesi potrebbe essere il seguente:

$H_0 : p = 0,2$

$H_1 : p < 0,2$

La variabile standardizzata per p=18/100=0,18 è:

$$Z = \frac{P - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}\frac{N-n}{N-1}}} = \frac{0,18 - 0,2}{\sqrt{\frac{0,2(0,8)}{100}\frac{1.000-100}{1.000-1}}} = -0.53$$

Ad un livello $\alpha = 0,03$, Z = -1,88

Poiché -0,53 > -1,88 non si rifiuta l'ipotesi nulla; il decremento nella proporzione di difetti non è significativo.

Inferenza statistica per due campioni

Inferenza statistica per due campioni

Finora i test hanno riguardato la stima di un singolo parametro della popolazione (media varianza o proporzione).

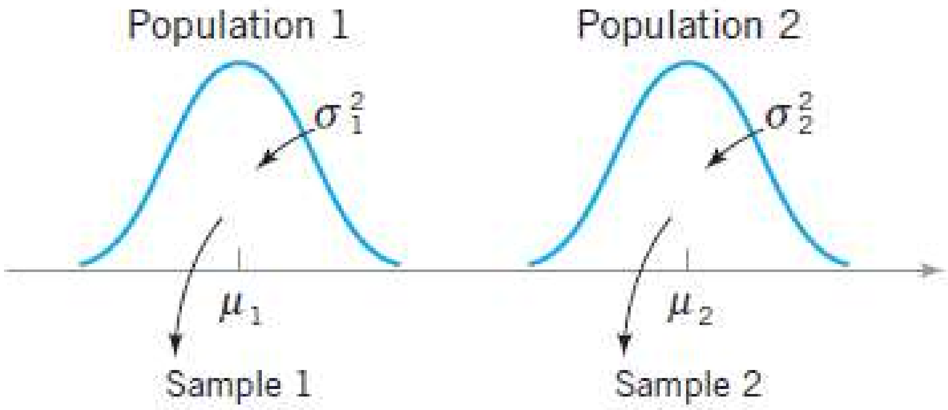
Supponiamo di voler confrontare due popolazioni in relazione ad un fenomeno di interesse.

Ad esempio si vuole confrontare la produzione di due macchinari.

Inferenza statistica per due campioni

x_{11} $x_{12}...x_{1n_1}$ è un campione casuale di n_1 osservazioni dalla popolazione 1, mentre x_{21} $x_{22}...x_{2n_2}$ dalla popolazione 2.

Supponiamo i due campioni siano tra loro indipendenti e che entrambe le popolazioni siano normali.



Inferenza statistica per due campioni

A. Inferenza sulla differenza tra le <u>medie</u> di 2 popolazioni con:

- Varianze note (la statistica test utilizzata è la z)
- Varianze non note (la statistica test utilizzata è la t)

B. Inferenza sul rapporto tra <u>varianze</u> (la statistica test utilizzata è F)

C. Inferenza sulla differenza tra <u>proporzioni</u> (la statistica test utilizzata è Z)

Inferenza statistica sulla differenza tra medie di 2 campioni

Inferenza sulla differenza tra le medie con varianze note

Consideriamo inferenze statistiche sulla differenza tra le medie di 2 popolazioni $(\mu_1 - \mu_2)$ quando le varianze delle due popolazioni sono note.

I presupposti sono:

- ✓ x_{11} x_{12}...x_{1n_1} è un campione casuale di n_1 osservazioni dalla popolazione 1
- ✓ x_{21} x_{22}...x_{2n_2} è un campione casuale di n_2 osservazioni dalla popolazione 2
- ✓ Le due popolazioni sono indipendenti
- ✓ Entrambe le popolazioni sono normali

Inferenza sulla differenza tra le medie con varianze note

Per formulare il test di ipotesi sulla differenza tra le due medie $(\mu_1 - \mu_2)$ verrà utilizzata la seguente statistica:

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Inferenza sulla differenza tra le medie con varianze note: esempio

Per ridurre il tempo di essicazione di una certa vernice un progettista deve scegliere tra la formulazione chimica standard e quella ottenuta aggiungendole un nuovo ingrediente che dovrebbe ridurre il tempo di essicazione.

L'esperienza suggerisce che la deviazione standard del tempo di essicazione è 8 minuti e che tale parametro di variabilità interna non dovrebbe essere influenzato dall'aggiunta del nuovo ingrediente.

Inferenza sulla differenza tra le medie con varianze note: esempio

10 provini sono verniciati con la prima formulazione e altri 10 con la seconda seguendo sempre un ordine casuale.

I tempi di asciugatura medi dei due campioni sono rispettivamente x_1=121 min e x_2= 112 min.

Quale conclusione può trarre il progettista circa l'efficacia del nuovo ingrediente usando α = 0,05?

Inferenza sulla differenza tra le medie con varianze note: esempio

Il sistema di ipotesi è il seguente:

$H_0 : \mu_1 - \mu_2 = 0$ ovvero $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 - \mu_2 > 0$ $H_1 : \mu_1 > \mu_2$

La statistica test è pari a

$$Z = \frac{\overline{x}_1 - \overline{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{121 - 1}{\sqrt{\frac{(8)^2}{10} + \frac{(8)^2}{10}}} = 2,52$$

Ad un livello $\alpha = 0,05$; Z = 1,645.

Poiché 2,52 > 1,645 si rifiuta l'ipotesi nulla; l'aggiunta di un nuovo ingrediente alla vernice riduce significativamente i tempi di essiccazione.

Inferenza sulla differenza tra le medie con varianze non note

Ora estendiamo i risultati precedenti alla differenza nelle medie di due distribuzioni quando le varianze di entrambe le distribuzioni sono sconosciute.

Se le dimensioni del campione sono superiori a 30 è possibile utilizzare le normali procedure di distribuzione nel caso precedente.

Inferenza sulla differenza tra le medie con varianze non note

Tuttavia, quando vengono prelevati piccoli campioni si assumerà che le popolazioni siano normalmente distribuite e i test di ipotesi si baseranno sulla distribuzione t.

Ciò si collega al caso di inferenza sulla media di un singolo campione con varianza sconosciuta.

Inferenza sulla differenza tra le medie con varianze non note

Si devono distinguere 2 situazioni:

1. varianze non note ma uguali (ossia $\sigma_1^2 = \sigma_2^2 = \sigma^2$)
2. varianze non note e diverse

Inferenza sulla differenza tra le medie con varianze non note ma uguali

Supponiamo di avere due popolazioni normali e indipendenti, con medie non note e varianze non note ma uguali; si intende verificare

$$H_0: \mu_1 - \mu_2 = \Delta_0$$
$$H_1: \mu_1 - \mu_2 \neq \Delta_0$$

Inferenza sulla differenza tra le medie con varianze non note ma uguali

Sia x_{11} x_{12}...x_{1n_1} un campione casuale di n_1 osservazioni dalla popolazione 1 e x_{21} x_{22}...x_{2n_2} un campione casuale di n_2 osservazioni dalla popolazione 2.

Siano $\bar{x}_1$ $\bar{x}_2$ s_1^2 s_2^2 le medie e le varianze campionarie.

Il valore atteso della differenza nelle medie campionarie è

$$\bar{x}_1 - \bar{x}_2 = E\,(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$

In questo modo la differenza tra le medie campionarie è uno stimatore non distorto della differenza tra le medie.

Inferenza sulla differenza tra le medie con varianze non note ma uguali

La varianza di $\bar{x}_1 - \bar{x}_2$ è

$$V(\bar{x}_1 - \bar{x}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Sembra ragionevole combinare le due varianze del campione s_1^2 e s_2^2 per formare uno stimatore della varianza totale σ^2.

Lo stimatore totale di σ^2 è definito

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Inferenza sulla differenza tra le medie con varianze non note ma uguali

Quindi s_p^2 ha $n_1 + n_2 - 2$ gradi di libertà.

La statistica test

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Ha una distribuzione standardizzata Z (0,1).

Inferenza sulla differenza tra le medie con varianze non note ma uguali

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Sostituendo a σ la sua stima campionaria s_p , si ottiene la distribuzione t con $n_1 + n_2 - 2$ gradi di libertà.

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Inferenza sulla differenza tra le medie con varianze non note ma uguali: esempio

Sono stati analizzati due catalizzatori per determinare come essi influenzano il rendimento medio di un processo chimico.

Il catalizzatore 1 è quello attualmente in uso, mentre il catalizzatore 2 potrebbe essere adottato in quanto più economico.

Si esegue un esperimento nell'impianto pilota.

Inferenza sulla differenza tra le medie con varianze non note ma uguali: esempio

Numero	Catalizzatore 1	Catalizzatore 2
1	91,50	89,19
2	94,18	90,95
3	92,18	90,46
4	95,39	93,21
5	91,79	97,19
6	89,07	97,04
7	94,72	91,07
8	89,21	92,75

Catalizzatore 1

Media : 92,255

Scarto quadratico medio: 2,39

Catalizzatore 2

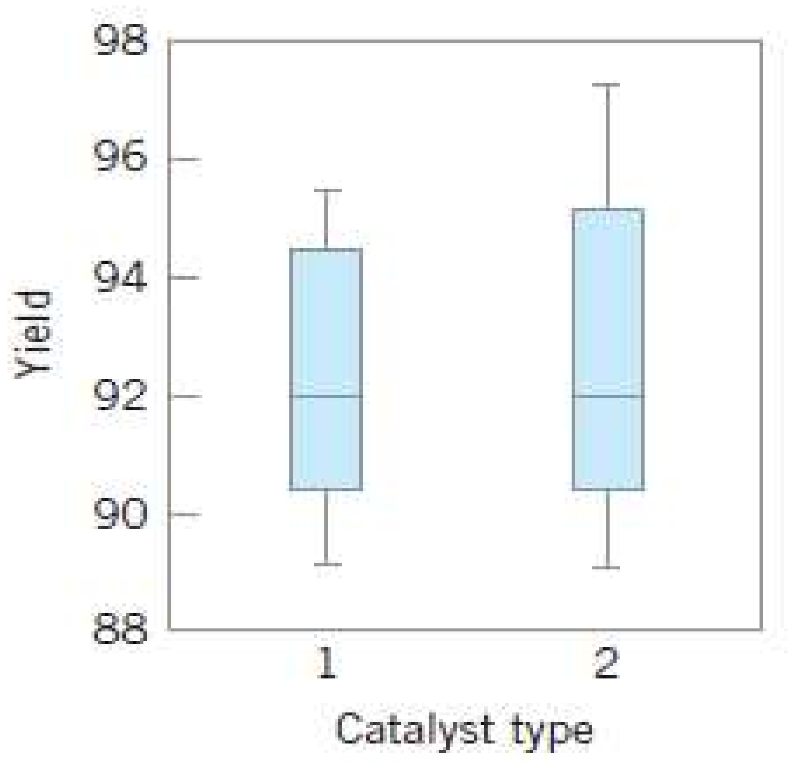
Media : 92,733

Scarto quadratico medio: 2,98

C’è differenza tra i rendimento medi dei 2 catalizzatori?

Inferenza sulla differenza tra le medie con varianze non note ma uguali: esempio

I box plot comparativi dei dati dei rendimenti per i due tipi di catalizzatori indicano che non vi è alcuna differenza evidente nella mediana deli due campioni, sebbene il secondo campione abbia una dispersione o una varianza leggermente maggiore.



Inferenza sulla differenza tra le medie con varianze non note ma uguali: esempio

Il sistema di ipotesi è il seguente:

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(8-1)2{,}39^2 + (8-1)2{,}98^2}{8+8-2} = 7{,}30 \qquad S_p = \sqrt{S_p^2} = \sqrt{7{,}30} = 2{,}70$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{95{,}255 - 92{,}733}{2{,}70\sqrt{\frac{1}{8}+\frac{1}{8}}} = -0{,}35$$

Con $\alpha = 0{,}05$; $t = 2{,}145$

Poiché $-2{,}145 < -0{,}35 < 2{,}145$

L'ipotesi nulla non può essere rifiutata: al livello del 5% non c'è sufficiente evidenza empirica per concludere che il catalizzatore 2 abbia un rendimento medio diverso da quello del catalizzatore 1.

Inferenza sulla differenza tra le medie con varianze non note e diverse

In molte situazioni non si può assumere che le varianze incognite siano uguali.

In questo caso la statistica test cui si ricorre è

$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Inferenza sulla differenza tra le medie con varianze non note e diverse

Questa statistica test si distribuisce approssimativamente come una t di Student con un numero di gradi di libertà dato da:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}{\frac{(S_1^2/n_1)^2}{n_1 + 1} + \frac{(S_2^2/n_2)^2}{n_2 + 1}} - 2$$

Inferenza sulla differenza tra le medie con varianze non note e diverse

Pertanto, nel caso in cui le varianze sono diverse le ipotesi sulle differenze nelle medie di due distribuzioni normali vengono verificate come nel caso delle varianze uguali tranne che viene utilizzato come statistica test una approssimazione della statistica t (t_0^*) e $n_1 + n_2 - 2$ viene sostituito da v nel determinare i gradi di libertà per il test.

Dati appaiati (Paired Data)

Finora abbiamo assunto che i due campioni utilizzati nei test siano indipendenti.

In alcune applicazioni vengono rilevati dati accoppiati.

Per misurare la resistenza alla trazione della fibra sintetica vengono utilizzati due diversi tipi di macchine.

Si intende determinare se le due macchine producono o meno gli stessi valori medi di resistenza alla trazione.

Dati appaiati (Paired Data)

Vengono selezionati casualmente 8 campioni di fibra e viene eseguita una misurazione della resistenza utilizzando ciascuna macchina su ciascun campione.

tipo di fibra	Macchinario 1	Macchinario 2	Differenza
1	74	78	-4
2	76	79	-3
3	74	75	-1
4	69	66	3
5	58	63	-5
6	71	70	1
7	66	66	0
8	65	67	-2

Dati appaiati (Paired Data)

I dati di questo esempio sono stati accoppiati per evitare che la differenza tra i campioni di fibra (che potrebbe essere sostanziale) influisca sul test sulla differenza tra le macchine.

La procedura del test consiste nell'ottenere le differenze della coppia di osservazioni su ciascuno degli n campioni $d_j = x_{1j} - x_{2j}$ e quindi testare l'ipotesi che la media della differenza μ_d sia zero.

Dati appaiati (Paired Data)

Testare H_0: $\mu_d = 0$ equivale testare l'ipotesi che H_0: $\mu_1 = \mu_2$

Di conseguenza il test è semplicemente il test t per un campione.

E la statistica test è:

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

Dove $\bar{d}$ rappresenta la media di tutte le differenze.

Dati appaiati (Paired Data)

Tornando all'esempio $\bar{d} = -\frac{11}{8} = -1{,}38$

$$s_d^2 = 7{,}13$$

$$t_0 = \frac{-1{,}38}{2{,}67/\sqrt{8}} = -1{,}46$$

Con α = 0,05, $t_{0{,}025{,}7} = 2{,}365$.

Si può concludere che non ci sono prove evidenti per indicare che le due macchine differiscono nelle loro misurazioni della resistenza alla trazione media.

Inferenza statistica sulla differenza tra varianze di 2 campioni

Inferenza sulle varianze di 2 distribuzioni

Supponiamo di verificare l'ipotesi che le varianze di due distribuzioni normali indipendenti siano uguali.

$$H_0\colon \sigma_1^2 = \sigma_2^2$$

$$H_1\colon \sigma_1^2 \neq \sigma_2^2$$

La statistica test per questa coppia di ipotesi è

$$F_0 = \frac{s_1^2}{s_2^2}$$

Inferenza sulle proporzioni di 2 distribuzioni

Supponiamo di verificare l'ipotesi che le proporzioni di due distribuzioni normali indipendenti siano uguali.

$$H_0\colon p_1 = p_2$$

$$H_1\colon p_1 \neq p_2$$

La statistica test per questa coppia di ipotesi è

$$Z = \frac{\widehat{p_1} - \widehat{p_2} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$