VERIFICA DI IPOTESI SULLA DIFFERENZA TRA DUE MEDIE

Psicometria 1 - Lezione 12
Lucidi presentati a lezione

AA 2000/2001 dott. Corrado Caudek

Il caso più comune di disegno sperimentale è quello in cui i soggetti vengono assegnati in maniera casuale a due gruppi.

Un gruppo viene sottoposto al trattamento sperimentale mentre l'altro gruppo funge da controllo.

In queste circostanze, il problema diventa quello di stabilire se il trattamento sperimentale ha avuto un effetto, ovvero se c'è una differenza statisticamente significativa tra le medie dei due gruppi.

Per risolvere questo problema è necessario conoscere la distribuzione campionaria della differenza tra le medie di due campioni.

La distribuzione campionaria della media di un campione casuale estratto da una popolazione normale con media μ e varianza σ^2: la media si distribuisce normalmente con media μ e varianza s^2/n.

Le medie del gruppo sperimentale e di controllo ($\overline{Y}_1$ e $\overline{Y}_2$) possono essere considerate come due variabili aleatorie normalmente distribuite aventi medie uguali a μ_1 e μ_2, e varianze uguali a s_1^2/n_1 e s_2^2/n_2 (dove n_1, n_2 sono le dimensioni dei due campioni).

In base alle proprietà del valore atteso della differenza di due variabili aleatorie, il valore atteso della differenza tra le medie dei due campioni è uguale a

$$E\left(\overline{Y}_1 - \overline{Y}_2\right) = E\left(\overline{Y}_1\right) - E\left(\overline{Y}_2\right) = m_1 - m_2$$

Se le due medie sono indipendenti, la varianza di $\overline{Y}_1$ - $\overline{Y}_2$ diventa uguale a

$$V\left(\overline{Y}_1 - \overline{Y}_2\right) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

Se è possibile assumere che le popolazioni di provenienza hanno varianze uguali, $s_1^2 = s_2^2 = s^2$, la varianza della differenza tra le medie di due campioni diventa

$$V\left(\overline{Y}_1 - \overline{Y}_2\right) = \frac{s^2}{n_1} + \frac{s^2}{n_2}$$

$$= s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

L'errore standard della differenza tra le medie di due campioni sarà

$$s_{\bar{Y}_1 - \bar{Y}_2} = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Se s è conosciuto, quindi, le ipotesi sulla differenza tra le medie di due campioni possono essere sottoposte a verifica usando la statistica

$$z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (m_1 - m_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

con $z \sim N(0, 1)$.

Se la varianza della popolazione non è conosciuta,
deve essere stimata usando le varianze dei campioni.

Date le varianze di due campioni, una stima priva di errore sistematico della deviazione standard della popolazione da cui i campioni sono tratti è data da

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

Quando s non è noto, dunque, le ipotesi sulla differenza tra due medie vengono sottoposte a verifica usando la statistica

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (m_1 - m_2)}{\hat{s}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

distribuita come t con $(n_1 + n_2 - 2)$ gradi di libertà.

Esempio 11.1. Due metodi di insegnamento (A e B) vengono confrontati esaminando i punteggi ottenuti alla fine del corso dai soggetti assegnati in maniera casuale al gruppo A oppure al gruppo B (maggiore il punteggio, migliori le prestazioni).

Alla luce dei dati che sono stati ottenuti, è possibile concludere che i due metodi di insegnamento producono risultati significativamente diversi? Sia α = 0,05.

Metodo A	Metodo B
$n_1 = 9$	$n_1 = 12$
$\overline{Y}_1 = 77$	$\overline{Y}_2 = 89$
$s_1 = 12$	$s_2 = 9$

$$H_0 : m_1 - m_2 = 0$$

$$H_a : m_1 - m_2 \neq 0$$

$$t = \frac{(\overline{Y}_1 - \overline{Y}_2) - 0}{\sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

con $n = (n_1 + n_2 - 2) = 19$ gradi di libertà.

Il test è a due code e la regione critica è definita da

$|t| < t_{\alpha/2} = 2.093$. Il valore osservato della statistica test

$$t = \frac{(77-89)-0}{\sqrt{\left(\frac{(9-1)\times 12^2 + (12-1)\times 9^2}{9+12-2}\right)\left(\frac{1}{9}+\frac{1}{12}\right)}} = -2.6244$$

cade all'interno della regione di rifiuto e quindi l'ipotesi nulla può essere rigettata. E' dunque possibile concludere che, alla fine del corso, i soggetti dei due gruppi ottengono punteggi significativamente diversi.

IMPORTANZA DELLE ASSUNZIONI PER IL TEST *t* DI STUDENT

Le procedure di verifica delle ipotesi sulla media di un campione sono basate sull'assunzione che la statistica $\sqrt{n}\left(\overline{Y}-m\right)/s$ venga calcolata utilizzando le informazioni fornite da un campione estratto da una ***popolazione normale***.

Se la grandezza del campione non è estremamente piccola, però, anche violazioni piuttosto marcate dell'assunzione della normalità della popolazione hanno effetti trascurabili nel caso di test bidirezionali.

Gli effetti sono invece più marcati nel caso di test monodirezionali.

Le procedure di verifica di ipotesi sulle differenze fra medie si dimostrano anch’esse relativamente insensibili alle violazioni dell'assunzione di normalità.

Il test sulla differenza fra due medie si dimostra relativamente robusto anche alle violazioni dell'assunzione di omogeneità delle varianze ($s_1^2 = s_2^2$) quando n_1 e n_2 sono uguali (o quasi uguali).

Se invece le popolazioni da cui sono estratti i campioni non hanno varianze uguali, e se la grandezza dei campioni è molto diversa, allora diventa necessario introdurre una correzione nel calcolo dei gradi di libertà:

$$n = \frac{\left(\hat{s}_1^2 + \hat{s}_2^2\right)^2}{\left(\hat{s}_1^2\right)^2 \big/ (n_1 + 1) + \left(\hat{s}_2^2\right)^2 \big/ (n_2 + 1)} - 2$$

VERIFICA DI IPOTESI SULLA DIFFERENZA TRA LE MEDIE DI CAMPIONI DIPENDENTI

La verifica di ipotesi sulle differenze tra le medie è basata sull'assunzione che i campioni siano indipendenti.

Talora, però, è necessario confrontare le medie di campioni dipendenti, come nel caso dei cosiddetti “campioni appaiati” (ovvero quei campioni composti, ad esempio, da mariti e mogli), oppure nel caso dei punteggi forniti dagli stessi soggetti prima e dopo la somministrazione di un trattamento.

In queste circostanze, la differenza tra le medie dei campioni continua a fornire una stima priva di errore sistematico della differenza tra le medie delle due popolazioni:

$$E(\bar{Y}_1 - \bar{Y}_2) = m_1 - m_2$$

Nel caso di campioni dipendenti, però, la varianza della distribuzione campionaria della differenza delle medie non sarà quella indicata dalla formula considerata in precedenza ma assumerà invece il valore:

$$V\left(\overline{Y}_1 - \overline{Y}_2\right) = s^2_{\overline{Y}_1} + s^2_{\overline{Y}_2} - 2s_{\overline{Y}_1\overline{Y}_2}$$

Nel caso di campioni dipendenti, quindi, se la covarianza tra $\overline{Y}_1$ e $\overline{Y}_2$ non è conosciuta, l'errore standard della differenza tra le medie non può essere calcolato e la procedura di inferenza statistica descritta in precedenza non può essere applicata.

Questa difficoltà, però, può essere superata considerando i dati come se provenissero da un unico campione di differenze (D_i) di punteggi, anziché da due campioni dipendenti.

Una volta ricodificati i dati in questo modo, l'ipotesi di assenza di differenza tra i due gruppi è equivalente all'ipotesi nulla $H_0 : E(D) = 0$ e può essere sottoposta a verifica calcolando la statistica

$$t = \frac{D_i - E(D)}{s_D \big/ \sqrt{n}}$$

distribuita come t con $(n-1)$ gradi di libertà (dove n è il numero di coppie di osservazioni).

Esempio. Supponiamo di verificare l'efficacia di un programma dietetico. Il peso di 10 individui viene misurato prima e dopo il completamento della dieta. Possiamo concludere che la dieta è efficace sulla base dei dati riportati nella tabella? Sia $\alpha = 0{,}01$.

Soggetti	*Prima*	*Dopo*	*D*
A	98	96	2
B	103	97	6
C	78	81	-3
D	84	82	2
E	77	75	2
F	68	69	-1
G	84	81	3
H	79	80	-1
I	91	85	6
L	89	88	1
Medie	85,1	83,4	1,7

L'ipotesi nulla è $H_0 : \overline{D} = 0$ e l'ipotesi sostantiva è $H_a : \overline{D} > 0$ (con D uguale alla differenza tra il peso prima e dopo la dieta). Il test è dunque ad una coda.

Con $\alpha = 0{,}01$ e $n = (n-1) = 9$ gradi di libertà, la regione critica è $t > t_a$, dove $t_a = 2{,}821$.

La variabile D ha media 1,7 e deviazione standard uguale a 2,9078.

Il valore della statistica test diventa

$$t = \frac{\overline{D} - m}{s_D \big/ \sqrt{n}} = \frac{1{,}7 - 0}{2{,}9078 \big/ \sqrt{10}} = 1{,}8488$$

e non cade all'interno della regione di rifiuto ($t_9 < 2{,}821$). I dati della tabella, quindi, non offrono evidenze sufficienti per concludere che il programma dietetico è efficace.

Confronti tra più di due campioni

Molto spesso lo sperimentatore si trova nella situazione di volere confrontare le medie di più di due campioni.

Supponiamo che ci siano 3 gruppi. Lo sperimentatore potrebbe decidere di eseguire 3 test *t* per confrontare tra loro le medie di tutti i gruppi. Questo approccio, però, sarebbe del tutto inadeguato.

Supponiamo che il primo e il secondo gruppo vengano confrontati con un test *t* scegliendo $\alpha = .05$. In queste circostanze lo sperimentatore è consapevole che, nel 5% dei casi, *l'ipotesi nulla verrà rifiutata quando è effettivamente vera*.

Usando lo stesso livello di significatività, lo sperimentatore potrebbe confrontare il secondo e il terzo gruppo e il primo e il terzo gruppo.

Il problema di questo approccio è che, qualora venissero eseguiti tutti e tre questi test la probabilità di commettere un errore del I tipo non sarebbe più uguale a .05.

Quale è la probabilità di commettere un errore del I tipo quando vengono eseguiti *tutti e tre* i test descritti in precedenza?

Consideriamo l'eventualità di commettere un errore del I tipo come un processo bernoulliano con probabilità di un "successo" uguale a .05. Dunque,

$$p(\text{nessun errore}) = \binom{3}{0}(.05)^0(.95)^3 = .86$$

Di conseguenza, la probabilità di commettere almeno un errore del I tipo sarà uguale a:

$$p(\text{uno o più errori}) = 1 - .86 = .14$$

Se eseguiamo 3 test *t*, quindi, la probabilità di commettere almeno un errore del I tipo è quasi 3 volte superiore alla probabilità di commettere un errore del I tipo quando viene eseguito un singolo test *t*.

In generale, se un numero J di test indipendenti viene eseguito usando per ciascuno il livello di significatività $\alpha = .05$, allora la probabilità di commettere almeno un errore del I tipo è:

$$p(\text{uno o più errori di tipo I}) = 1 - (1 - a)^J$$

Supponiamo, per esempio, di avere 8 gruppi e di esaminare le differenze tra tutte le coppie di questi gruppi. Quante coppie possibili possiamo esaminare?

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{8!}{2!(8-2)!} = 28$$

La probabilità di commettere almeno un errore del primo tipo sarà dunque:

$$p(\text{uno o più errori di tipo I}) = 1 - (1 - .05)^{28} = .76$$

Con 8 gruppi, quindi, la probabilità di compiere almeno un errore del primo tipo è .76.

Nel caso di 20 gruppi, questa probabilità è quasi uguale a 1.

Le considerazioni precedenti sono valide se i test sono indipendenti. Se questa condizione non è soddisfatta, la probabilità di commettere un errore del I tipo è ancora maggiore. Questa è la ragione per cui *l'uso del test t per confronti multipli tra medie non è adeguato.*

Correzione di Bonferroni

Bonferroni ha dimostrato che la probabilità di commettere almeno un errore del primo tipo non può assumere un valore maggiore del prodotto tra il livello di significatività α e il numero *C* dei confronti che vengono eseguiti:

$$1-\left(1-a\right)^{C} \leq Ca$$

(se $C\alpha > 1$, la probabilità viene ovviamente fissata a 1).

Per non inflazionare la probabilità di errore del primo tipo (ponendo, ad esempio, $\alpha = 0{,}05$), in base alla *correzione di Bonferroni* è dunque sufficiente usare per ciascun confronto un livello di significatività pari a

$$a = 0{,}05/C$$.

ESERCIZI

E1 Uno studio si occupa della percezione della figura parentale in bambini provenienti da famiglie normali e in bambini istituzionalizzati. I bambini venivano posti ad una certa distanza dallo schermo di un computer e sul monitor venivano presentate figure diverse. Le figure avevano tutte la stessa grandezza, anche se i bambini non lo sapevano. Ai bambini veniva chiesto di fornire un stima numerica della grandezza delle figure.

Lo sperimentatore ipotizza che gli orfani forniscono stime della grandezza parentale maggiori di quelle fornite dai bambini provenienti da famiglie normali. I dati dei due gruppi sono:

<u>Orfani:</u> media = 1.82, s = .7, n = 125
<u>Controllo:</u> media = 1.61, s = .9, n = 150

Sia α = .05. Che cosa si può concludere?

E2 Due campioni casuali di soggetti vengono confrontati per studiare gli effetti del ritmo di presentazione di un lista di item sulla rievocazione. Una lista di 50 parole viene presentata a ciascun gruppo e i soggetti vengono istruiti a rievocare quanti più item possibili.
Nella condizione sperimentale gli items vengono presentati con un ritmo di .5 al secondo mentre, nella condizione di controllo la presentazione è di 2 item al secondo.

I soggetti asssegnati in maniera casuale al gruppo sperimentale e di controllo ammontano a 6 per il gruppo sperimentale e a 8 per quello di controllo. I risultati per i due gruppi sono:

$$n_1=6, \text{ media}_1=32, \; s_1^2=12,$$
$$n_2=8, \text{ media}_2=27, \; s_2^2=10.5$$

Lo sperimentatore può assumere che le due popolazioni di punteggi sono normali con uguale varianza. Sia $\alpha = .01$. Che cosa si può concludere?

E3 Sia il gruppo 1 il gruppo di controllo e il gruppo 2 il gruppo sperimentale. Sulle base dei dati riportati di seguito è possibile concludere che il gruppo sperimentale differisce significativamente da quello di controllo? Sia α = .05. Quale assunzione è necessario formulare per applicare le opportune procedure di inferenza statistica?

$$Y_1 = 65.5 \qquad s_1^2 = 5.55 \qquad n_1 = 10$$

$$Y_2 = 69 \qquad s_2^2 = 7.78 \qquad n_2 = 10$$

E4 Uno sperimentatore studia la possibilità che la dieta denominata "*Q*" abbia effetti diversi negli uomini e nelle donne. Lo sperimentatore ipotizza che, nelle prime due settimana di dieta, le donne perdano più peso degli uomini. Per tenere sotto controllo variabili estranee, lo sperimentatore seleziona un campione casuale di 15 coppie (mariti e mogli). Ciascuna di queste coppie viene sottoposto alla dieta *Q*. Sia $\alpha = .05$. Che cosa è possibile concludere dai dati raccolti?

Perdita di peso (in pounds)

Mariti	Mogli
5	2.7
3.3	4.4
4.3	3.5
6.1	3.7
2.5	5.6
1.9	5.1
3.2	3.8
4.1	3.5
4.5	5.6
2.7	4.2
7	6.3
1.5	4.4
3.7	3.9
5.2	5.1
1.9	3.4