

Lezione 4

La verifica delle ipotesi sulle differenze tra le medie

Argomenti della lezione:

- **Campioni indipendenti**
- **Campioni dipendenti**

Due campioni indipendenti

Due popolazioni composte da soggetti diversi

Formulazione delle ipotesi

Ipotesi nulla

- **H_0: $\mu_1 = \mu_2$ le medie delle due popolazioni sono uguali**

Ipotesi alternative

- **H_1: $\mu_1 \neq \mu_2$ bidirezionale**
- **H_1: $\mu_1 > \mu_2$ monodirezionale destra**
- **H_1: $\mu_1 < \mu_2$ monodirezionale sinistra**

Modello di riferimento

Distribuzione Campionaria delle Differenze tra le medie di due campioni di numerosità rispettivamente n_1 e n_2

Costruzione

Due popolazioni indipendenti

Campioni di numerosità n_1 e n_2

$\bar{X}_1$ Campione della popolazione 1

$\bar{X}_2$ Campione della popolazione 2

$$d = (\bar{X}_1 - \bar{X}_2)$$

Proprietà

- **Forma: normale**
- **Media:**

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

- **Varianza e Errore Standard:**

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- **Varianza ed errore standard stimati dai campioni:**

$$\hat{\sigma}^2_{\bar{x}_1 - \bar{x}_2} = \frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}$$

- **Test: normale standardizzata**

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

Non occorre conoscere la media delle popolazioni di riferimento

Ipotesi (nulla):

$$\mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$$

Formula per il calcolo di z:

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

n_1 e $n_2 > 30$

Ipotesi di ricerca

Le femmine sono aggressive in misura diversa dai maschi?

Campioni

M:	$n_1 = 62$	$\bar{X}_1 = 5.06$	$s_1 = 2.95$
F:	$n_2 = 69$	$\bar{X}_2 = 4.17$	$s_2 = 2.24$

Ipotesi di ricerca bidirezionale

H_0: $\mu_M = \mu_F$

H_1: $\mu_M \neq \mu_F$

$\alpha = 0.05$

$z_{crit} = \pm\ 1.96$

$$z = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} = \frac{5.06 - 4.17}{\sqrt{\frac{8.70}{61} + \frac{5.00}{68}}} =$$

$$= \frac{0.89}{\sqrt{0.14 + 0.07}} = \frac{0.89}{0.46} = +1.93$$

Si accetta l'ipotesi nulla

Ipotesi di ricerca

I maschi sono più aggressivi delle femmine?

Campioni

M: $n_1 = 62$ $\bar{X}_1 = 5.06$ $s_1 = 2.95$

F: $n_2 = 69$ $\bar{X}_2 = 4.17$ $s_2 = 2.24$

Ipotesi di ricerca monodirezionale

H_0: $\mu_M = \mu_F$

H_1: $\mu_M > \mu_F$

$\alpha = 0.05$

$z_{crit} = +1.64$

Il calcolo di z non cambia

$z = +\ 1.93 > z_{crit} = +1.64$

Si respinge l'ipotesi nulla

n_1 e/o $n_2 \leq 30$

1. si utilizza la "t di Student" con $(n_1 + n_2 - 2)$ gdl
2. media $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

3 errore standard

varianze diverse

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}}$$

varianze uguali

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2-2}\left(\frac{n_2+n_2}{n_1 n_2}\right)}$$

4 test

$$t = \frac{\bar{X}_1-\bar{X}_2-0}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}}$$

Ipotesi di ricerca:

Le femmine sono meno aggressive dei maschi?

Campioni

M: $n_1 = 15$ $\bar{X}_1 = 4.90$ $s_1^2 = 1.93$

F: $n_2 = 15$ $\bar{X}_2 = 3.80$ $s_2^2 = 4.45$

Ipotesi monodirezionale

$H_0: \mu_F = \mu_M$; $H_1: \mu_F < \mu_M$

$\alpha = 0.05$

gdl = 15 + 15 – 2 = 28

$t_{crit} = -1.701$

Varianze uguali

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

$$t = \frac{(3.8-4.9)-0}{\sqrt{\frac{15\times1.93+15\times4.45}{15+15-2}\left(\frac{15+15}{15\times15}\right)}} = -1.63$$

$t = -1.63 > t_{crit}$

Si accetta l'ipotesi nulla

Campioni non indipendenti (correlati)

Coppie di osservazioni:

- ➔ variabile misurata due volte sullo stesso soggetto
- ➔ variabile misurata su coppie "appaiate" di soggetti

Distribuzione Campionaria delle Medie delle Differenze

Costruzione

- Due popolazioni dipendenti
- Campioni di numerosità n
- Distribuzione bivariata dei due punteggi X e Y

Differenze per ogni soggetto

$$d_i = X_i - Y_i$$

Media delle differenze tra le coppie

$$\bar{d} = \frac{\sum_{1}^{n} d_i}{n}$$

Proprietà

- Si fa riferimento alla t di Student con n-1 gdl
- Media: $\mu_{\bar{d}} = 0$ assumendo H_0 come vera
- Varianza e Errore Standard:

$$s^2_{\bar{d}} = \frac{\sigma^2_d}{n} \qquad s_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}}$$

- Varianza e errore standard stimata dal campione:

$$s^2_{\bar{d}} = \frac{s^2_d}{n-1} \qquad s_{\bar{d}} = \frac{s_d}{\sqrt{n-1}}$$

dove

$$s_d = \sqrt{\frac{\sum_{1}^{n} (d-\bar{d})^2}{n}}$$

è lo s.q.m. del campione

- Test: t di Student

$$t = \frac{\bar{d} - \mu_{\bar{d}}}{s_{\bar{d}}}$$

L'ipotesi nulla è che la media delle differenze ($\mu_{\bar{d}}$)sia = 0

ESEMPIO

La visione di un filmato informativo sull'AIDS migliora l'atteggiamento verso i sieropositivi ?

Campione di N = 8 soggetti

Scala di atteggiamento somministrata prima e dopo aver visto un filmato

I dati ottenuti sono i seguenti:

sogg	Prima	Dopo
1	20	20
2	17	15
3	15	10
4	12	10
5	20	18
6	18	19
7	17	19
8	20	16
media	17.4	15.9

→ $H_0: \mu_{\bar{d}} = 0$ cioè $\mu_p = \mu_d$

$H_1: \mu_{\bar{d}} > 0$ cioè $\mu_p > \mu_d$

→ per $\alpha = 0.05$ e
$(n-1) = (8-1) = 7$ gdl,
$t_{crit} = 1.895$ (monodirezionale)

Calcolo di t

sogg	Prima	Dopo	$d_i=X_i-Y_i$
1	20	20	0
2	17	15	2
3	15	10	5
4	12	10	2
5	20	18	2
6	18	19	-1
7	17	19	-2
8	20	16	4
media	17.4	15.9	1.5

$$\bar{d} = \frac{\sum d_i}{N} = \frac{12}{8} = 1.5$$

$$s_d = \sqrt{\frac{\sum d_i^2}{N} - \bar{d}^2} = \sqrt{\frac{58}{8} - 1.5^2} = 2.24$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{N-1}} = \frac{2.24}{\sqrt{7}} = 0.84$$

$$t = \frac{\bar{d} - 0}{s_{\bar{d}}} = \frac{1.5}{0.84} = 1.786$$

Poiché $t < t_{crit}$ (1.895) accetto H_0

La visione del film non migliora l'atteggiamento

Verifica delle ipotesi sulle varianze

Campioni "1" e "2"

Varianze s_1^2 e s_2^2

I campioni provengono da popolazioni con varianze uguali?

Questo può influire sul calcolo della t di Student

Distribuzione campionaria del rapporto tra le varianze

Forma: F di Fisher
[n_1-1 e n_2-1 gdl]

$$F = \frac{s^2_{max}}{s^2_{min}}$$

Rapporto F ottenuto dai dati campionari

$$F = \frac{s_1^2 \frac{n_1}{n_1-1}}{s_2^2 \frac{n_2}{n_2-1}}$$

ESEMPIO

Dati dell'esempio sulla t di Student per gruppi indipendenti

$n_1 = 15$ e $n_2 = 15$

$s_1^2 = 1.93$ e $s_2^2 = 4.45$

1. formulazione delle ipotesi
2. $H_0: \sigma_1^2 = \sigma_2^2 \qquad H_1: \sigma_1^2 \neq \sigma_2^2$
3. scelta del livello critico: $\alpha = 0.05$

$F_{crit}(14,14) = 2.48$

4. calcolo di F

$$F = \frac{4.45 \times \frac{15}{15-1}}{1.93 \times \frac{15}{15-1}} = 2.30$$

5. decisione

Poiché $F < F_{crit}$, accetto l'ipotesi nulla. I campioni provengono da popolazioni con varianza uguale

CONCLUSIONE

- Ipotesi su due popolazioni
- Campioni indipendenti
- Campioni correlati
- Ipotesi sulle medie e sulle varianze