Esercitazione 5 del corso di Statistica (parte 2)

Dott.ssa Paola Costantini

25 Maggio 2012

Esercizio 1

Per verificare l'efficacia di un corso di statistica vengono confrontati i rendimenti medi di due campioni di studenti di ampiezza $n_1 = n_2 = 50$, di cui uno costituito da studenti frequentanti l'altro da studenti non frequentanti. Tali valori risultano essere pari rispettivamente a 28 e 25. Sapendo che le popolazioni da cui sono stati estratti i campioni sono distribuite normalmente con varianze pari a 16, costruire un intervallo di confidenza al 95% per la differenza tra le medie delle due popolazioni.

Soluzione

Intervallo di confidenza per la v.c. differenza tra medie $(\mu_1 - \mu_2)$

Popolazione normale, varianze note

$$\overline{x_1} = 28 \qquad \sigma_1^2 = \sigma_2^2 = 16$$
$$\overline{x_2} = 25 \qquad \mathrm{n}_1 = \mathrm{n}_2 = 50$$
$$\alpha = 0{,}05$$
$$\mathrm{z}_{\alpha/2} = 1{,}96$$

$$IC_{0,95} = (28-25) \pm 1{,}96\sqrt{2 \cdot \frac{16}{50}} = 3 \pm 1{,}568 = [1{,}432; 4{,}568]$$

$$P(1{,}432 \leq |\mu_1 - \mu_1| \leq 4{,}568) = 0{,}95$$

L'intervallo di confidenza non contiene lo 0: questo vuol dire che una differenza nulla tra le medie è poco probabile, ossia che esiste una differenza *significativa* (significativamente diversa da 0) tra il rendimento medio dei frequentanti e quello dei non frequentanti, il che a sua volta vuol dire che il corso è efficace.

Esercizio 2

Per valutare se esiste o meno una differenza tra il reddito medio di uomini e donne si estraggono due campioni di 25 donne e 20 uomini, da cui risulta un reddito medio di 9 (migliaia di euro) con s.q.m. di 3 per gli uomini. Costruire un intervallo di confidenza per la differenza tra i redditi medi al livello di significatività del 99%.

Soluzione

Intervallo di confidenza per la v.c. differenza tra medie $(\mu_1 - \mu_2)$

<u>Popolazione non nota, varianze non note, campioni piccoli</u>

$\overline{x_1} = 9 \quad s_1 = 2 \quad s_2 = 3$

$\overline{x_2} = 15 \quad \mathrm{n}_1 = 25 \quad \mathrm{n}_2 = 20$

$\alpha = 0{,}01$

Ottenuta la stima della varianza comune, attraverso lo stimatore:

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

L'intervallo di confidenza è costruito come:

$$IC_{1-\alpha} = \left|\overline{x_1} - \overline{x}_2\right| \pm t_{\alpha/2;n1+n2-2}\, s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$s^2 = \frac{(25-1)\cdot 4 + (20-1)\cdot 9}{25+20-2} = 6{,}21$$

$$t_{\alpha/2;n1+n2-2} = t_{0,005;43} = 2{,}695$$

$$IC_{0,99} = |9-15| \pm 2{,}695 \cdot 6{,}21\sqrt{\frac{1}{25} + \frac{1}{20}} = 6 \pm$$

$$P(0{,}9792 \le |\mu_1 - \mu_2| \le 11{,}0207$$

L'intervallo di confidenza non contiene lo 0: questo vuol dire che esiste una differenza *significativa* (significativamente diversa da 0) tra il reddito medio di donne e uomini (il sesso incide sul livello medio di reddito).

Esercizio n. 3

In passato la lunghezza media delle pannocchie di grano è stata uguale a 27 cm con $\sigma^2 = 24$. Si vuole sottoporre a test l'ipotesi che le pannocchie di un determinato anno abbiano una lunghezza media diversa, sulla base di un campione di 20 elementi con un $\alpha = 0{,}04$.

30	29	16	19	25	23	18	17	29	30
29	30	23	27	22	16	28	24	26	30

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = 24{,}55$$

IPOTESI	$H_0 : \mu = 27$ $H_1 : \mu \neq 27$
STATISTICA TEST	$X_{test} = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}}$
VALORI CRITICI	$\pm Z_{\alpha/2} \to \pm Z_{0,02} = \pm 2{,}05375$
REGOLA DI DECISIONE	Accetto H_0 se $-2{,}05375 \leq v_{test} \leq +2{,}05375$
VALORE TEST	$V_{test} = \dfrac{24{,}55 - 27}{4{,}9\sqrt{20}} = -2{,}24$
DECISIONE	$-2{,}24 < -2{,}05375$ si rifiuta l'ipotesi H_0

Esercizio n 4

Una ditta produttrice di batterie per cellulari pubblicizza i propri prodotti garantendo una durata media di 18 ore con uno scarto quadratico medio di 0,5. Poiché ha ricevuto parecchi reclami da parte dei clienti che sostengono che la durata è inferiore, la ditta decide di effettuare una prova di durata su un campione di 10 batterie, ottenendo un tempo medio di accensione di 17,7 ore.

a) sulla base di tale risultato come può la ditta verificare la validità della sua affermazione riguardante la durata media garantita.

Svolgimento

Si richiede di effettuare un test sulla media conoscendo la varianza del carattere osservato, con un livello di significatività del 5%.

IPOTESI	$H_0 : \mu = 18$ $H_1 : \mu < 18$
STATISTICA TEST	$X_{test} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
VALORI CRITICI	$z_{0,05} = \phi(0,05)$ = 1-0,05 =0,95, che sulle tavole è uguale a 1,645 $(-1,645 \to \text{coda sinistra})$
REGOLA DI DECISIONE	Accetto H_0 se $V_{test} > -1,645$
VALORE TEST	$V_{test} = \frac{17,7-18}{0,5\sqrt{10}} = -1,896$
DECISIONE	-1,896 < -1,645 si rifiuta l'ipotesi H_0

Esercizio 5

Un economista del Ministero degli Esteri desidera verificare se gli accordi di negoziazione tra Italia e Giappone siano rispettati. In particolare egli sospetta che i produttori giapponesi fissino un prezzo più basso per i prodotti venduti sul mercato italiano rispetto a quello usato sul mercato interno, ostacolando al contempo le importazioni di prodotti italiani con forti ostacoli di tipo burocratico. Si interessa in particolare al mercato dell'auto e vuole testare l'ipotesi che prezzi più alti siano applicati in Giappone rispetto all'Italia per le autovetture di produzione giapponese. Esamina a tal fine due campioni relativi a pratiche di acquisto di tali autovetture nello stesso periodo di tempo (50 per il mercato italiano e 30 per il mercato giapponese). Convertendo i prezzi di vendita in Giappone usando il cambio corrente Yen/Euro, ottiene i risultati elencati nella seguente tabella:

	ITALIA	GIAPPONE
AMPIEZZA CAMPIONE	50	30
MEDIA CAMPIONARIA	€16545	€17243

Siano inoltre noti i seguenti valori per le rispettive popolazioni di riferimento

	ITALIA	GIAPPONE
DEVIAZIONE STANDARD	€1989	€1843

a) Costruire un test d'ipotesi usando un livello di significatività α=0,05

I due campioni sono selezionati in maniera indipendente dalle due popolazioni. Le rispettive ampiezze campionarie n_1 e n_2 sono sufficientemente grandi affinché sia $\bar{x}_1$ che $\bar{x}_2$ siano distribuite approssimativamente come normali.

SVOLGIMENTO

IPOTESI	$H_0 : \mu_1 - \mu_2 = 0$ $H_1 : \mu_1 - \mu_2 < 0$
STATISTICA TEST	$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{(\bar{x}_1 - \bar{x}_2)}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx Z$
VALORI CRITICI	$\alpha = 0{,}05 \rightarrow Z_\alpha = -1{,}645$
REGOLA DI DECISIONE	$\text{Accetto } \mathrm{H}_0 \text{ se } \mathrm{V}_{\text{test}} > -1{,}645$
VALORE TEST	$= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(16545 - 17243)}{\sqrt{\frac{1989^2}{50} + \frac{1843^2}{30}}} = -1{,}59$
DECISIONE	-1,59 < -1,645 si accetta l'ipotesi H_0

Esercizio 6

Al fine di verificare le spese trimestrali per la tenuta di c/c bancari, la Banca d'Italia esamina le spese di tenuta conto praticate da 8 banche italiane riscontrando una spesa trimestrale media di 32 Euro con una varianza campionaria pari a 1.8.
Allo stesso tempo, procede ad effettuare lo stesso tipo di rilevazione presso 10 banche straniere operanti in Italia, rilevando una spesa trimestrale media di 32.9 Euro con una varianza campionaria pari a 1.7.
Si vuole verificare se le spese trimestrali medie di tenuta conto sono le stesse presso i due tipi di banche.

SVOLGIMENTO

Si richiede di effettuare un test sulla differenza tra medie non conoscendo le varianze del carattere osservato nelle due popolazioni. In tal caso, per poter effettuare il test, è necessario ipotizzare che le due varianze siano uguali $\left(\sigma_1^2 = \sigma_2^2 = \sigma^2\right)$.
Le ipotesi da sottoporre a verifica sono le seguenti:

$$H_0 : \mu_1 - \mu_2 = 0 \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

Per poter effettuare il test bisogna anche ipotizzare che il carattere osservato abbia una distribuzione normale presso le due popolazioni.
La statistica test da considerare è la seguente:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx t_{(n_1+n_2-2)}$$

che ha una distribuzione *t* di Student con n_1+n_2-2.
Fissato un livello di significatività $\alpha = 0.05$ è possibile determinare le regioni di accettazione e di rifiuto:

Regione di accettazione: $\left\{-t_{\alpha/2} < t_i < t_{1-\alpha/2}\right\}$; Regione di rifiuto: $\left\{t_i < -t_{\alpha/2}\right\} \cup \left\{t_i > t_{1-\alpha/2}\right\}$.

Dai dati a disposizione risulta:

$$n_1 = 8; \quad n_2 = 10; \quad \bar{x}_1 = 32; \quad \bar{x}_2 = 32.9; \quad s_1^2 = 1.8 \quad s_2^2 = 1.7.$$

Il test è a due code; i valori critici della distribuzione *t* di Studente per questo livello di significatività sono $t_{0.025} = -2.12$ e $t_{0.975} = 2.12$.
La regola di decisione consiste nel rifiutare l'ipotesi nulla se il valore campionario della statistica test è inferiore a -2.12 o superiore a 2.12.

Il valore campionario della statistica test è: $t = \frac{32 - 32.9}{\sqrt{\frac{8(1.8)+10(1.7)}{8+10-2}\left(\frac{1}{8} + \frac{1}{10}\right)}} = -1.35$

Al livello di significatività del 5% l'ipotesi nulla viene accettata, concludendo che le spese trimestrali medie di tenuta conto sono le stesse presso i due tipi di banche.

Esercizio 6

Tipo di contratto / Utilizzo della laurea	Stabile	Atipico	Inserimento/ formazione	Senza contratto	Tot
In misura elevata	5	3	1	0	9
In misura ridotta	2	5	1	0	8
Per niente	2	4	0	2	8
totale	9	12	2	2	25

Verificare ad un livello di significatività del 90% se i due caratteri sono indipendenti

IPOTESI	$H_0 = \sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij} = 0$
	$H_1 = \sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij} > 0$
STATISTICA TEST	$\sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij}$
VALORI CRITICI	$\chi^2_{0,1,6}$ 6=(r-1)*(c-1) = 10,645
REGOLA DI DECISIONE	Accetto H_0 se $V_{test} < 10,645$
VALORE TEST	$\chi^2 = =$ **7,4**
DECISIONE	7,4 < 10,645 si accetta l'ipotesi H_0

Accetto l'ipotesi nulla, quindi i due caratteri sono indipendenti. Il valore test è diverso da zero, ma non significativamente diverso da zero

.