Università degli Studi di Cassino, Anno accademico 2004-2005
Corso di Statistica 2, Prof. M. Furno

Esercitazione del 01/03/2005
dott. Claudio Conversano

Esercizio 1
Si consideri il seguente campione casuale semplice estratto da una popolazione normale di media μ e varianza 16:

33.2, 27.8, 30.8, 39.6, 36.2, 43.4, 36.6, 28.8.

Calcolare l'intervallo di confidenza per μ al livello di confidenza $1 - \alpha = 0.99$.

SVOLGIMENTO

Si richiede di costruire l'intervallo di confidenza per la media avendo osservato un carattere X la cui distribuzione è normale e di cui è nota anche la varianza.
Per costruire l'intervallo di confidenza per μ bisogna considerare la v.c. media campionaria che ha una distribuzione normale con media μ e varianza σ^2/n. La v.c. standardizzata $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ ha una distribuzione normale con media 0 e varianza 1.
E' possibile costruire un intervallo simmetrico intorno a Z individuando due valori $-z_{\alpha/2}$ e $+z_{\alpha/2}$ tali che $P(-z_{\alpha/2} \leq Z \leq +z_{\alpha/2}) = 1 - \alpha$. Tale relazione equivale a:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq +z_{\alpha/2}\right) = P\left(\bar{X} - z_{\alpha/2}\,\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1 - \alpha$$

Dai dati a disposizione risulta:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(33.2 + 27.8 + 30.8 + 39.6 + 36.2 + 43.4 + 36.6 + 28.8) = 34.55$$

Poichè $1 - \alpha = 0.99$ si avrà $\alpha/2 = 0.005$ e $|z_{\alpha/2}| = 2.575$. La probabilità di cui sopra può essere riscritta nel seguente modo:

$$P\left(-2.575 \leq \frac{34.55 - \mu}{4/\sqrt{8}} \leq +2.575\right) = 0.99\,.$$

L'intervallo di confidenza per μ al 99% risulta pari a:

$$\left[34.55 - 2.575 \cdot 4/\sqrt{8}\,, 34.55 + 2.575 \cdot 4/\sqrt{8}\right] = [30.91, 38.19]$$

Esercizio 2
E' stato scoperto che l'aggiunta di un determinato composto chimico ad un certo fertilizzante permette di aumentare la quantità di fragole ottenibile da ogni singola unità di fertilizzante. Un esperimento viene condotto, introducendo il composto e misurando poi la quantità di fragole prodotta da 5 unità di fertilizzante. I risultati (in grammi di fragole per unità di fertilizzante) sono i seguenti:

229, 255, 280, 203, 229.

Stimare un intervallo per la media della quantità di fragole ottenuta dai 5 fertilizzanti utilizzati, ad un livello di confidenza del 99%.

SVOLGIMENTO

Si richiede di costruire l'intervallo di confidenza per la media avendo osservato un carattere X di cui non è nota la varianza.
Come regola pratica, se la dimensione del campione è minore di 30 non si può invocare il teorema del limite centrale e sfruttare l'approssimazione alla distribuzione normale, ma occorre utilizzare la distribuzione campionaria della media facendo riferimento alla v.c. t di Student con *n-1* gradi di libertà.
In questo caso occorre verificare che la distribuzione della popolazione sia normale. Se questa ipotesi non è soddisfatta il test può non essere significativo.
Nel caso in oggetto per costruire l'intervallo di confidenza per la media bisogna comunque ipotizzare che, per la popolazione, il carattere osservato X abbia una distribuzione normale.
Un intervallo di confidenza è ricavato dalla seguente relazione:

$$P\left(-t_{\alpha/2} \le \frac{\bar{X}-\mu}{s/\sqrt{n}} \le +t_{\alpha/2}\right) = P\left(\bar{X} - t_{\alpha/2}\, s/\sqrt{n} \le \mu \le \bar{X} + t_{\alpha/2}\, s/\sqrt{n}\right) = 1-\alpha$$

in cui il valore $\left|t_{\alpha/2}\right|$ è il quantile di una distribuzione *t* di Student con *n-1* gradi di libertà.
Dai dati a disposizione risulta:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}\left(229+255+280+203+229\right) = 239.2$$

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\bar{x}\right)^2} = \sqrt{\frac{1}{5}\left(229^2+255^2+280^2+203^2+229^2\right) - \left(239.2\right)^2} = 29.3$$

Poiché $1 - \alpha = 0.99$ si avrà $\alpha/2 = 0.005$ e $\left|t_{\alpha/2}\right| = 4.604$. La probabilità di cui sopra può essere riscritta nel seguente modo:

$$P\left(-4.604 \le \frac{239.2-\mu}{29.3/\sqrt{5}} \le +4.604\right) = 0.99 .$$

L'intervallo di confidenza per μ al 99%, assumendo che X abbia una distribuzione normale, risulta pari a:

$$\left[239.2 - 4.604 \cdot 29.2/\sqrt{5}\,,\, 239.2 + 4.604 \cdot 29.2/\sqrt{5}\right] = \left[179.08, 299.32\right]$$

Esercizio 3

Supponiamo che un analista finanziario voglia valutare la sensibilità delle variazioni dei rendimenti dei titoli dell'indice S&P500 alle variazioni dei rendimenti dell'indice stesso. A tal proposito misura in un certo istante la correlazione tra i rendimenti di 45 titoli azionari componenti lo S&P500 ed i rendimenti dello stesso indice S&P500 osservati nell'arco di un anno su base settimanale.

Il valore medio del coefficiente di correlazione tra i rendimenti dei 45 titoli considerati e i rendimenti dell'indice S&P500 è risultato pari a 0.62 con una varianza pari a 0.36. Stimare il valore medio della correlazione tra i rendimenti di tutti i 500 titoli dell'indice S&P500 ed i rendimenti dell'indice stesso ad un livello di confidenza del 90%.

SVOLGIMENTO

Si richiede di costruire l'intervallo di confidenza per la media avendo osservato un carattere *X* di cui non è nota la varianza. In questo caso, a differenza dell'esercizio precedente, la dimensione del campione è maggiore di 30 (*n=45*) per cui è possibile ipotizzare che la distribuzione campionaria della media approssimi una distribuzione normale.

E' possibile costruire un intervallo simmetrico intorno a *Z* individuando due valori $-z_{\alpha/2}$ e $+z_{\alpha/2}$ tali che $P\left(-z_{\alpha/2} \le Z \le +z_{\alpha/2}\right) = 1-\alpha$. Tale relazione equivale a:

$$P\left(-z_{\alpha/2} \le \frac{\overline{X}-\mu}{s/\sqrt{n}} \le +z_{\alpha/2}\right) = P\left(\overline{X} - z_{\alpha/2}\, s/\sqrt{n} \le \mu \le \overline{X} + z_{\alpha/2}\, s/\sqrt{n}\right) = 1-\alpha$$

Dai dati a disposizione risulta:

$n = 45$; $\overline{x} = 0.62$; $s = 0.36$.

Poiché $1 - \alpha = 0.90$ si avrà $\alpha/2 = 0.05$ e $\left|z_{\alpha/2}\right| = 1.645$. La probabilità di cui sopra può essere riscritta nel seguente modo:

$$P\left(-1.645 \le \frac{0.62-\mu}{0.6/\sqrt{45}} \le +1.645\right) = 0.90 .$$

L'intervallo di confidenza per μ al 90% risulta pari a:

$$\left[0.62 - 1.645\cdot 0.6/\sqrt{45}\,, 0.62 - 1.645\cdot 0.6/\sqrt{45}\right] = \left[0.473, 0.767\right].$$

Esercizio 4

I dati esposti nella tabella seguente mostrano n = 10 misurazioni della temperatura minima (in gradi centigradi) rilevata nel comune di Cassino durante il mese di aprile 2004.

Rilevazione	Temperatura	Rilevazione	Temperatura
1	5.507	6	5.527
2	5.506	7	5.504
3	5.500	8	5.490
4	5.497	9	5.500
5	5.506	10	5.497

Al fine di poter giudicare la precisione delle misurazioni, nell'ipotesi che i dati provengano da una popolazione normale e che la temperatura minima media rilevata per 5 anni durante il mese di aprile sia di 5.5 gradi centigradi, si costruisca un intervallo di confidenza al 95% per la varianza.

SVOLGIMENTO

Si richiede di costruire l'intervallo di confidenza per la varianza avendo osservato un carattere X la cui distribuzione è normale e di cui è nota anche la media.

Per costruire l'intervallo di confidenza per σ^2 bisogna considerare la statistica nS^2/σ^2, che ha una distribuzione chi-quadro *n* gradi di libertà.

E' possibile costruire un intervallo simmetrico intorno a σ^2 individuando due valori $\chi^2_{(\alpha/2);n}$ e $\chi^2_{(1-\alpha/2);n}$ tali che $P\left(\chi^2_{(\alpha/2);n} \le nS^2/\sigma^2 \le \chi^2_{(1-\alpha/2);n}\right) = 1-\alpha$. Tale relazione equivale a:

$$P\left(\chi^2_{(\alpha/2);n} \le \frac{nS^2}{\sigma^2} \le \chi^2_{(1-\alpha/2);n}\right) = P\left(\frac{nS^2}{\chi^2_{(\alpha/2);n}} \le \sigma^2 \le \frac{nS^2}{\chi^2_{(1-\alpha/2);n}}\right) = 1-\alpha$$

Dai dati a disposizione risulta:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}\left(5.507+5.506+5.5+5.497+5.506+5.527+5.504+5.49+5.5+5.497\right) = 5.5034$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\overline{x}\right)^2 =$$
$$= \frac{1}{10}\left(5.507^2+5.506^2+5.5^2+5.497^2+5.506^2+5.527^2+5.504^2+5.49^2+5.5^2+5.497^2\right) - 5.5034^2 =$$
$$= 0.00008684$$

Poiché 1 − α = 0.95 si avrà $\alpha/2 = 0.025$, $\chi^2_{(0.025);10} = 20.48$ e $\chi^2_{(0.975);10} = 3.25$. La probabilità di cui sopra può essere riscritta nel seguente modo:

$$P\left(20.48 \le \frac{10\cdot 0.00008684}{\sigma^2} \le 3.25\right) = 0.95$$

L'intervallo di confidenza per σ^2 al 95% risulta pari a:

$$\left[\frac{10\cdot 0.00008684}{20.48},\frac{10\cdot 0.00008684}{3.25}\right]=\left[0.0000480, 0.0003028\right].$$

Esercizio 5
Un progetto di ricerca tra l'università di Cassino e l'università di Stanford prevedeva la realizzazione di un nuovo software statistico. Esso venne sottoposto a prove di verifica e venne chiesto ad alcuni esperti di valutare l'affidabilità del prodotto. Su 48 esperti interpellati, 36 ritennero che il nuovo software fosse affidabile. Trovare un intervallo di confidenza al 95% per la vera frazione di esperti che ritenne che il nuovo software fosse affidabile.

SVOLGIMENTO

Si richiede di costruire l'intervallo di confidenza per la proporzione di successi avendo osservato un campione di dimensione $n=48$.
Per costruire l'intervallo di confidenza per la proporzione di successi X/N bisogna considerare ciascuna prova come la realizzazione di una v.c. di Bernoulli di parametro X/n (nel caso in esame X/n rappresenta la probabilità che un esperto dichiari il software affidabile). Le n prove considerate permettono di definire la v.c. proporzione di successi nel campione che, per n elevato ($n > 30$) ha una distribuzione normale con media X/n e varianza $\frac{X/n(1-X/n)}{n}$. La v.c. standardizzata $Z=\left(\frac{X}{n}-\frac{X}{N}\right)\Big/\sqrt{\frac{X/n(1-X/n)}{n}}$ ha una distribuzione normale con media 0 e varianza 1.
E' possibile costruire un intervallo simmetrico intorno a Z individuando due valori $-z_{\alpha/2}$ e $+z_{\alpha/2}$ tali che $P\left(-z_{\alpha/2}\le Z\le +z_{\alpha/2}\right)=1-\alpha$. Tale relazione equivale a:

$$P\left(-z_{\alpha/2}\le\frac{X/n-\frac{X}{N}}{\sqrt{\frac{X/n(1-X/n)}{n}}}\le +z_{\alpha/2}\right)=$$

$$=P\left(X/n-z_{\alpha/2}\sqrt{\frac{X/n(1-X/n)}{n}}\le\frac{X}{N}\le X/n+z_{\alpha/2}\sqrt{\frac{X/n(1-X/n)}{n}}\right)=1-\alpha$$

Dai dati a disposizione risulta:

$n=48$; $X/n=36/48=0.75$

Poiché $1-\alpha = 0.95$ si avrà $\alpha/2=0.025$, e $\left|z_{\alpha/2}\right|=1.96$. La probabilità di cui sopra può essere riscritta nel seguente modo:

$$P\left(-1.96\le\frac{0.75-\frac{X}{N}}{\sqrt{\frac{0.75(1-0.75)}{48}}}\le +1.96\right)=0.95.$$

L'intervallo di confidenza per X/N al 95% risulta pari a:

$$\left[0.75-1.96\sqrt{\frac{0.75(1-0.75)}{48}},0.75+1.96\sqrt{\frac{0.75(1-0.75)}{48}}\right]=[0.628,0.872].$$

Esercizio 6

Da un'indagine campionaria condotta a Cassino, è risultato che l'altezza media dei bambini iscritti alla classe I elementare era uguale a 104.5 cm. (con scarto quadratico medio campionario corretto uguale a 8.41), mentre l'altezza media delle bambine iscritte alla stessa classe era uguale a 101.9 cm. (con scarto quadratico medio campionario corretto uguale a 5.61).

Sapendo che il campione dei bambini e quello delle bambine hanno entrambi numerosità $n = 120$ e che i due campioni provengono da popolazioni aventi varianza uguale, determinare l'intervallo di confidenza al 90 % per $\mu_F - \mu_M$, in cui μ_M e μ_F indicano l'altezza media per i bambini e per le bambine.

SVOLGIMENTO

Si richiede di costruire l'intervallo di confidenza per la differenza tra medie di un certo carattere X ipotizzando di osservare due campioni indipendenti provenienti da due diverse popolazioni aventi stessa varizanza (incognita).

Indichiamo con $\overline{X}_M$ e $\overline{X}_F$ le v.c. distribuzioni campionarie delle medie nei due campioni considerati.

Poiché entrambi i campioni hanno numerosità elevata, le rispettive medie campionarie sono distribuite come v.c. Normali, cioè:

$$\overline{X}_M \approx N\left(\mu_M, \frac{s_M^2}{n_M - 1}\right); \qquad \overline{X}_F \approx N\left(\mu_F, \frac{s_F^2}{n_F - 1}\right).$$

Queste due v.c. sono indipendenti poiché si riferiscono a campioni indipendenti. Quindi, per la proprietà riproduttiva della v.c. Normale, risulta:

$$\overline{X}_F - \overline{X}_M \approx N\left(\mu_F - \mu_M, s_{M,F}^2 = \frac{(n_F - 1)s_F^2 + (n_M - 1)s_M^2}{n_F + n_M - 2}\right) \Rightarrow Z = \frac{\left(\overline{X}_F - \overline{X}_M\right) - \left(\mu_F - \mu_M\right)}{s_{M,F}\sqrt{\frac{1}{n_F} + \frac{1}{n_M}}} \approx N(0,1)$$

E' possibile costruire un intervallo simmetrico intorno a Z individuando due valori $-z_{\alpha/2}$ e $+z_{\alpha/2}$ tali che $P\left(-z_{\alpha/2} \le Z \le +z_{\alpha/2}\right) = 1 - \alpha$. Tale relazione equivale a:

$$P\left(-z_{\alpha/2} \le Z \le +z_{\alpha/2}\right) = P\left(-z_{\alpha/2} \le \frac{\left(\overline{X}_F - \overline{X}_M\right) - \left(\mu_F - \mu_M\right)}{s_{M,F}\sqrt{\frac{1}{n_F} + \frac{1}{n_M}}} \le +z_{\alpha/2}\right) =$$

$$= P\left(\left(\overline{X}_F - \overline{X}_M\right) - z_{\alpha/2}s_{M,F}\sqrt{\frac{1}{n_F} + \frac{1}{n_M}} \le \left(\mu_F - \mu_M\right) \le \left(\overline{X}_F - \overline{X}_M\right) + z_{\alpha/2}s_{M,F}\sqrt{\frac{1}{n_F} + \frac{1}{n_M}}\right) = 1 - \alpha$$

Dai dati a disposizione risulta:

$$n_F = n_M = 120; \qquad \overline{x}_F - \overline{x}_M = 104.5 - 101.9 = 2.6;$$

$$s_{M,F}^2 = \frac{(n_F - 1)s_F^2 + (n_M - 1)s_M^2}{n_F + n_M - 2} = s_{M,F}^2 = \frac{(120-1)8.41 + (120-1)5.61}{120 + 120 - 2} = 7.01;$$

$$s_{M,F} = \sqrt{7.01} = 2.648$$

Poiché $1 - \alpha = 0.90$ si avrà $\alpha/2 = 0.05$, e $\left|z_{\alpha/2}\right| = 1.645$. La probabilità di cui sopra può essere riscritta nel seguente modo:

$$P\left(-1.645 \le Z \le +1.645\right) = P\left(-1.645 \le \frac{2.6-\left(\mu_F - \mu_M\right)}{2.648 \cdot \sqrt{\dfrac{1}{120}+\dfrac{1}{120}}} \le +1.645\right) = 0.90$$

L'intervallo di confidenza per $\left(\mu_F - \mu_M\right)$ al 90% risulta pari a:

$$\left[2.6-1.645 \cdot 2.648 \cdot \sqrt{\frac{1}{120}+\frac{1}{120}}, 2.6-1.645 \cdot \sqrt{\frac{1}{120}+\frac{1}{120}}\right] = \left[2.037, 3.162\right].$$