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

Esercitazione del 08/03/2005
dott. Claudio Conversano

Esercizio 1

Un'indagine effettuata dalla segreteria studenti della facoltà di Economia dell'università di Cassino nel periodo 1993 – 2002 ha rivelato che l'età media degli studenti laureatisi negli ultimi 10 anni è 25 anni, con uno scarto quadratico medio di 3.6 anni. Per i 144 studenti laureatisi nel 2004 l'età media è 24.5 anni. Si può ritenere che i laureati del 2004 sono stati più veloci dei loro colleghi degli anni precedenti nel conseguire la laurea?

SVOLGIMENTO

Si richiede di effettuare un test sulla media conoscendo la varianza del carattere osservato. Poiché dal campione risulta un valore medio inferiore a quello dichiarato le ipotesi da sottoporre a verifica sono le seguenti:

$$H_0 : \mu = 25 \qquad\qquad H_1 : \mu < 25$$

Essendo la dimensione del campione elevata (*n=144*), si può supporre che il carattere osservato abbia una distribuzione normale con uno scarto quadratico medio $\sigma = 3.6$.
La statistica test da considerare è la seguente:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

che ha una distribuzione normale standard. Fissato un livello di significatività $\alpha = 0.05$ è possibile determinare le regioni di accettazione e di rifiuto:

Regione di accettazione: $\{z_i > -z_\alpha\}$; Regione di rifiuto: $\{z_i < -z_\alpha\}$.

Il test è a una coda; il valore critico per questo livello di significatività è $z_{0.05} = -1.645$.
La regola di decisione consiste nel rifiutare l'ipotesi nulla se il valore campionario della statistica test è inferiore a -1.645.
Dai dati a disposizione risulta:

$$n = 144; \qquad \bar{x} = 24.5; \qquad \mu = 25; \qquad \sigma = 3.6.$$

Il valore campionario della statistica test è: $z = \dfrac{24.5 - 25}{3.6/\sqrt{144}} = -1.67$

Al livello di significatività del 5% l'ipotesi nulla viene rifiutata, concludendo che l'età media degli studenti laureati nel 2004 si discosta significativamente da quella degli studenti laureati nel decennio precedente.
Viceversa per un livello di significatività del 1%, essendo $z_{0.01} = -2.33$, l'ipotesi nulla verrebbe accettata.

Esercizio 2

Una ditta produttrice di lampadine assicura che la loro durata nel 98% dei casi è superiore alle 200 ore. In un campione di 90 lampadine, la durata effettiva ha superato le 200 ore in 82 prove. Si sottoponga a verifica, al livello di significatività dell'1%, l'affermazione della ditta.

SVOLGIMENTO

Si richiede di effettuare un test sulla proporzione di successi del carattere osservato. Poiché dal campione risulta un proporzione di successi inferiore a quella ipotizzata le ipotesi da sottoporre a verifica sono le seguenti:

$$H_0 : \frac{X}{N} = 0.98 \qquad H_1 : \frac{X}{N} < 0.98$$

La proporzione di successi $\frac{X}{N}$ si distribuisce come una v.c. di Bernoulli. La dimensione del campione consente di approssimare la distribuzione binomiale delle *n* prove indipendenti alla distribuzione normale.

La statistica test da considerare è la seguente:

$$Z = \frac{\frac{X}{n} - \frac{X}{N}}{\sqrt{\frac{X/N\left(1 - X/N\right)}{n}}} \approx N\left(0,1\right)$$

che ha una distribuzione normale standard. Fissato un livello di significatività $\alpha = 0.01$ è possibile determinare le regioni di accettazione e di rifiuto:

Regione di accettazione: $\left\{z_i > -z_\alpha\right\}$; Regione di rifiuto: $\left\{z_i < -z_\alpha\right\}$.

Il test è a una coda; il valore critico per questo livello di significatività è $z_{0.01} = -2.33$.

La regola di decisione consiste nel rifiutare l'ipotesi nulla se il valore campionario della statistica test è inferiore a -2.33.

Dai dati a disposizione risulta:

$$n = 90\,; \qquad X/n = 82/90 = 0.91\,; \quad X/N = 0.98\,.$$

Il valore campionario della statistica test è: $z = \frac{0.91 - 0.98}{\sqrt{\frac{0.98 \cdot 0.02}{90}}} = -4.73$

Al livello di significatività del 1% l'ipotesi nulla viene rifiutata.

Esercizio 3

Sono stati selezionati due campioni di studenti del I e del II anno di corso, il primo di dimensione 200 ed il secondo di dimensione 300. Nel primo campione il 56% aveva già sostenuto l'esame di Statistica 1, mentre per il secondo campione tale percentuale era pari al 48%. A livello di significatività del 5% possiamo affermare che c'è differenza tra la percentuale di studenti del I anno e quelli del II anno che hanno sostenuto l'esame di Statistica 1?

SVOLGIMENTO

Si richiede di effettuare un test sulla differenza tra proporzioni di successi non conoscendo le stesse quantità nelle due popolazioni dalle quali i due campioni provengono. In tal caso, per poter effettuare il test, è necessario ipotizzare che le due proporzioni siano uguali $\left(\frac{X_1}{N_1}=\frac{X_2}{N_2}=\frac{X_1+X_2}{N_1+N_2}\right)$.

Le ipotesi da sottoporre a verifica sono le seguenti:

$$H_0:\frac{X_1}{N_1}-\frac{X_2}{N_2}=0 \qquad\qquad H_1:\frac{X_1}{N_1}-\frac{X_2}{N_2}\neq 0$$

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:

$$Z=\frac{\left(\frac{X}{n_1}-\frac{X}{n_2}\right)-\left(\frac{X}{N_1}-\frac{X}{N_2}\right)}{\sqrt{\frac{X_1+X_2}{N_1+N_2}\left(1-\frac{X_1+X_2}{N_1+N_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\approx N(0,1)$$

che ha una distribuzione normale standard. La stima della proporzione di successi della popolazione $\frac{X_1+X_2}{N_1+N_2}$ è costituita dalla proporzione di successi campionaria: $\frac{X_1+X_2}{n_1+n_2}$.
Fissato un livello di significatività $\alpha=0.05$ è possibile determinare le regioni di accettazione e di rifiuto:

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

Dai dati a disposizione risulta:

$$n_1=300;\qquad n_2=200;\qquad \frac{X_1}{N_1}=0.56;\qquad \frac{X_2}{N_2}=0.48;\qquad \frac{X_1+X_2}{n_1+n_2}=\frac{300\cdot 0.56+200\cdot 0.48}{500}=0.528.$$

Il test è a due code; i valori critici della distribuzione normale standard per questo livello di significatività sono $z_{0.025}=-1.96$ e $z_{0.975}=1.96$.
La regola di decisione consiste nel rifiutare l'ipotesi nulla se il valore campionario della statistica test è inferiore a -1.96 o superiore a 1.96.
Il valore campionario della statistica test è:

$$z = \frac{0.56 - 0.48}{\sqrt{0.528(1-0.528)\left(\frac{1}{300}+\frac{1}{200}\right)}} = 1.75$$

Al livello di significatività del 5% l'ipotesi nulla viene accettata, concludendo che non c'è differenza tra la percentuale di studenti del I anno e quelli del II anno che hanno sostenuto l'esame di Statistica 1.

Esercizio 4
Il rendimento annuale di certi titoli azionari è distribuito in modo normale con scarto quadratico medio pari a 0.25. L'esame di un campione di 20 titoli azionari fornisce uno scarto quadratico medio campionario corretto di 0.32. L'apparente aumento dello scarto quadratico medio è significativo al livello del 5%? E a livello dell'1%?

SVOLGIMENTO

Si richiede di effettuare un test sulla varianza. Poiché dal campione risulta una varianza diversa da quella dichiarata, le ipotesi da sottoporre a verifica sono le seguenti:

$$H_0 : \sigma^2 = 0.25 \qquad H_1 : \sigma^2 > 0.25$$

La statistica test da considerare è la seguente:

$$\chi^2 = \frac{(n-1)\tilde{s}^2}{\sigma^2} \approx \chi^2_{n-1}$$

che ha una distribuzione chi-quadro con *n-1* gradi di libertà.
Fissato un livello di significatività $\alpha = 0.05$ è possibile determinare le regioni di accettazione e di rifiuto:

Regione di accettazione: $\left\{\chi^2 < \chi^2_\alpha\right\}$; Regione di rifiuto: $\left\{\chi^2 > \chi^2_\alpha\right\}$.

Dai dati a disposizione risulta:

$$n-1 = 20-1 = 19; \qquad \tilde{s} = 0.32; \qquad \sigma = 25.$$

Il test è a una coda; il valore critico per questo livello di significatività è $\chi^2_{0.05,19} = 30.144$. La regola di decisione consiste nel rifiutare l'ipotesi nulla se il valore campionario della statistica test è superiore a 30.144.

Il valore campionario della statistica test è: $\chi^2 = \dfrac{19 \cdot 0.32^2}{0.25^2} = 31.13$

Al livello di significatività del 5% l'ipotesi nulla viene rifiutata, concludendo che l'apparente aumento dello scarto quadratico medio è significativo.

Fissato un livello di significatività $\alpha = 0.01$, il valore critico per questo livello di significatività è $\chi^2_{0.01,19} = 36.191$. In questo caso l'ipotesi nulla può essere accettata.

Esercizio 5
Si considerano le reazioni di 280 pazienti ai trattamenti di tre terapie (A, B e C) per la cura di una certa malattia. La distribuzione dell'esito della terapia è la seguente:

Esito	**Terapia**			
	A	**B**	**C**	**Totale**
Paziente guarito	73	81	84	238
Paziente non guarito	10	21	11	42
Totale	83	102	95	280

Si vuole sottoporre a verifica, a livello di significatività del 5%, l'ipotesi che le tre terapie abbiano lo stesso effetto sui pazienti.

SVOLGIMENTO

Si richiede di effettuare un test per verificare se esiste indipendenza tra i due caratteri osservati. Le ipotesi da sottoporre a verifica sono le seguenti:

$$H_0 : n_{ij} = \hat{n}_{ij} \quad \forall i, j \qquad\qquad H_1 : n_{ij} \neq \hat{n}_{ij}$$

La statistica test da considerare è la seguente:

$$\chi = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{\left(n_{ij} - \hat{n}_{ij}\right)^2}{\hat{n}_{ij}} \approx \chi^2_{[(k-1)(h-1)]}$$

che ha una distribuzione Chi quadro con gradi di libertà $g_1 = k-1$ e $g_2 = h-1$.
Fissato un livello di significatività $\alpha = 0.05$ è possibile determinare le regioni di accettazione e di rifiuto:

Regione di accettazione: $\left\{\chi < \chi^2_{\alpha;[(k-1)(h-1)]}\right\}$; Regione di rifiuto: $\left\{\chi > \chi^2_{\alpha;[(k-1)(h-1)]}\right\}$.

Dai dati a disposizione risulta:

$k-1 = 2-1 = 1$; $\qquad h-1 = 3-1 = 2$;

Tabella delle frequenze teoriche:

Esito	**Terapia**		
	A	**B**	**C**
Paziente guarito	70.55	86.70	80.75
Paziente non guarito	12.45	15.30	14.25

Il test è a una coda; il valore critico per questo livello di significatività è $\chi^2_{0.05;2} = 5.991$.
La regola di decisione consiste nel rifiutare l'ipotesi nulla se il valore campionario della statistica test è superiore a 5.991.

Il valore campionario della statistica test è:

$$\chi^2 = \frac{(73-70.55)^2}{70.55} + \frac{(81-86.7)^2}{86.7} + \frac{(84-80.75)^2}{80.75} + \frac{(10-12.45)^2}{12.45} +$$

$$+ \frac{(21-15.30)^2}{15.30} + \frac{(11-14.25)^2}{14.25} = 3.938$$

Al livello di significatività del 5% l'ipotesi nulla viene accettata, concludendo che le tre terapie hanno lo stesso effetto sui pazienti.