Lezione 5

La verifica delle ipotesi sulla forma della distribuzione

Argomenti della lezione:

- Caso di un campione
- Caso di due campioni

Caso di un campione

- Non si stima un parametro
- Esame della forma della distribuzione nella popolazione dai dati del campione

Problema

- Gradimento programmi televisivi
- Campione: 42 telespettatori
- Scelta programma preferito

Variabile "tipo di programma".
Su scala nominale

Distribuzione di frequenza

programma	frequenza
A – culturale	5
B – sportivo	15
C – cartoni	22
totale	42

Ipotesi nulla (H_0)

Nella popolazione da cui è estratto il campione,
le frequenze nelle k categorie sono uguali
(preferenze equidistribuite)

Ipotesi alternativa (H_1)

Nella popolazione da cui è estratto il campione, le frequenze sono diverse

Test del chi quadrato:

Adattamento fra distribuzione teorica (secondo H_0) ed empirica, rilevata sul campione

Confronto fra:

frequenze teoriche (o attese) e osservate (o empiriche)

$$\chi^2 = \sum_{1}^{k} \frac{(f_0 - f_t)^2}{f_t}$$

gdl = (k - 1)

Ipotesi statistiche dell'esempio

H_0: $f_A = f_B = f_C$

(equidistribuzione nella popolazione)

H_1: $(f_A \neq f_B)$ o $(f_A \neq f_C)$ o $(f_B \neq f_C)$

(frequenze diverse nella popolazione)

Calcolo delle frequenze teoriche

Se è vera l'ipotesi nulla:

$f_t = n/k$

quindi

42/3= 14

Scelta di α e individuazione di chi² critico con 2gdl

$\alpha = 0.05$

chi² critico = 5.99

Calcolo del chi qudrato:

$$Chi^2 = \sum_{1}^{3} \frac{(f_0 - f_t)^2}{f_t} =$$

$$= \frac{(5-14)^2}{14} + \frac{(15-14)^2}{14} + \frac{(22-14)^2}{14} =$$

$$= 81/14 + 1/14 + 64/14 =$$

$$= 146/14 = 10.43$$

Decisione

- ➔ $Chi^2 > chi^2$ critico
- ➔ Si respinge H_0

C'è differenza nella preferenza dei programmi televisivi

Celle che danno il maggior contributo alla significatività

Residuo Standardizzato

$$R = \frac{f_0 - f_t}{\sqrt{f_t}}$$

Deve essere maggiore di |2|

Programma	F_0	F_t	? F_t	R
A – culturale	5	14	3.74	-2.41
B – sportivo	15	14	3.74	.27
C – cartoni	22	14	3.74	2.14

Programma più scelto: Cartoni
Programma meno scelto: Culturale

Limiti di applicabilità

- indipendenza
- $k = 2$, $f_t \geq 5$
- $k > 2$, $f_t \geq 1$, $f_t < 5$ in non più del 20%

Caso di due campioni

Incrocio fra 2 variabili categoriali

Variabile 1: Genere: M, F

Variabile 2: Tipo di facoltà: I, P, L

Esiste una relazione tra le due variabili ?

Campione: 84 soggetti
tabella a doppia entrata

	I	P	L	
M	20	6	8	34($n_{1.}$)
F	8	20	22	50($n_{2.}$)
	28 ($n_{.1}$)	26 ($n_{.2}$)	30 ($n_{.3}$)	84 (N)

- **Ipotesi nulla**
  Equivalenza tra le proporzioni nelle popolazioni (assenza di relazione)
- **Ipotesi alternativa**
  Proporzioni diverse nelle popolazioni (relazione significativa)

Le f_t si calcolano per ogni casella tenendo conto dei totali marginali

	B=1	B=2	B=3	
A = 1				($n_{1.}$)
A = 2				($n_{2.}$)
	($n_{.1}$)	($n_{.2}$)	($n_{.3}$)	(N)

frequenza teorica, in ogni cella, che si avrebbe se H_0 fosse vera

A e B indipendenti $\Rightarrow$

$$P[(A = 1) \text{ e } (B = 1)] = = P(A = 1) * P(B = 1)$$

$P(A = 1) = n_{1.}/N$, $P(B = 1) = n_{.1}/N$

$P[(A = 1) \text{ e } (B = 1)] = n_{1.}/N * n_{.1}/N$

Probabilità = Frequenza relativa

Frequenza Assoluta = Probabilità * N

$$f_{11}[(A = 1) \text{ e } (B = 1)] = N * P[(A = 1) \text{ e } (B = 1)] = N*(n_{1.}/N * n_{.1}/N) = n_{1.}n_{.1}/N$$

Per ogni cella:

$f_{11} = n_{1.}n_{.1}/N$

$f_{12} = n_{1.}n_{.2}/N$

$f_{13} = n_{1.}n_{.3}/N$

$f_{21} = n_{2.}n_{.1}/N$

$f_{22} = n_{2.}n_{.2}/N$

$f_{23} = n_{2.}n_{.3}/N$

Per ogni casella la frequenza teorica f_t si calcola moltiplicando i corrispondenti totali marginali di riga e di colonna e dividendo il prodotto per il totale generale

$$f_t = \frac{m_r x n_c}{N}$$

Calcolo della frequenza teorica f_{11}

	B=1	B=2	B=3	
A = 1	11.33			34
A = 2				
	28			84

$f_{11} = (34 * 28)/ 84$

Calcolo della frequenza teorica f_{12}

	B=1	B=2	B=3	
A = 1		10.52		34
A = 2				
		26		84

$f_{12} = (34 * 26)/ 84$

Calcolo della frequenza teorica f_{13}

	B=1	B=2	B=3	
A = 1			12.14	34
A = 2				
			30	84

$f_{13} = (34 * 30)/ 84$

Calcolo della frequenza teorica f_{21}

	B=1	B=2	B=3	
A = 1				
A = 2	16.67			50
	28			84

$f_{21} = (50 * 28)/ 84$

Calcolo della frequenza teorica f_{22}

	B=1	B=2	B=3	
A = 1				
A = 2		15.48		50
		26		84

$f_{22} = (50 * 26)/ 84$

Calcolo della frequenza teorica f_{23}

	B=1	B=2	B=3	
A = 1				
A = 2			17.86	50
			30	84

f_{23}= (50 * 30)/ 84

Frequenze teoriche: Tabella finale

	I	P	L	
M	11.33	10.52	12.14	34 ($n_{1.}$)
F	16.67	15.48	17.86	50 ($n_{2.}$)
	28 ($n_{.1}$)	26 ($n_{.2}$)	30 ($n_{.3}$)	84 (N)

α = 0.05
gdl = (r-1) x (c-1)
gdl = (2-1) x (3-1) = 2

chi² critico = 5.99

Calcolo del chi²

$$\text{Chi}^2 = \sum_{1}^{k} \frac{(f_0 - f_t)^2}{f_t}$$

f_0	f_t	(f_0-f_t)	$(f_0-f_t)^2$	$(f_0-f_t)^2/f_t$
20	11.33	8.67	75.17	6.63
6	10.52	-4.52	20.43	1.94
8	12.14	-4.14	17.14	1.41
8	16.67	-8.67	75.17	4.51
20	15.58	4.52	20.43	1.32
22	17.86	4.14	17.14	.96

chi² =16.77

Decisione

- ➔ Chi² > chi² critico
- ➔ Rifiuto l'ipotesi nulla

C'è una relazione significativa tra le due variabili

Residui standardizzati

Cella	f_0	f_t	R
M - I	20	11.33	2.57
M - P	6	10.52	-1.40
M - L	8	12.14	-1.19
F - I	8	16.67	-2.13
F - P	20	15.58	1.15
F - L	22	17.86	.98

Formula abbreviata per tabelle 2 x 2

A	B	A+B
C	D	C+D
A+C	B+D	N

$$chi^2 = \frac{N(AD-BC)^2}{(A+B)(C+D)(A+C)(E+D)}$$

Esempio

- Campione: 216 studenti
  - → 90 di Psicologia
  - → 126 di Ingegneria
- Variabile: favorevoli o contrari al numero chiuso

	FAV	CONT	
PSI	55	35	90
ING	55	71	126
	110	106	216

$$chi^2 = \frac{216(55x71-35x55)^2}{90x126x110x106} =$$

$$chi^2 = \frac{216(3905-1925)^2}{90x126x110x106} = 6.40$$

α = .05
gdl = 1
chi² critico = 3.84

chi² calcolato > chi² critico

C'è relazione significativa tra le due variabili categoriali

CONCLUSIONE

- Ipotesi sulla forma della distribuzione
- Bontà dell'adattamento
- Test su uno e su due campioni