

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Metodi Statistici per la Ricerca Sociale

Analisi dell'associazione tra variabili categoriali

1. La seguente tabella mostra la distribuzione doppia di frequenza di un collettivo di 125 operai disoccupati, rispetto all'età e alla durata del periodo di disoccupazione (in mesi):

Durata della disoccupazione (in mesi)	Età	
	< 35	≥ 35
1 – 7	36	21
8 – 14	30	5
15 – 30	7	10
> 30	7	9

- (a) Determinare la distribuzione di frequenze marginale assoluta del carattere durata del periodo di disoccupazione
- (b) Determinare le distribuzioni relative condizionate e percentuali condizionate del carattere “durata del periodo di disoccupazione” per classe di età.

Soluzione:

- (a) Distribuzione di frequenze marginale assoluta del carattere durata del periodo di disoccupazione:

Durata della disoccupazione (in mesi)	n_i
1 – 7	57
8 – 14	35
15 – 30	17
> 30	16

- (b) Distribuzioni relative condizionate e percentuali condizionate del carattere “durata del periodo di disoccupazione” per classe di età.

Durata della disoccupazione (in mesi)	Età	
	< 35	≥ 35
1 – 7	0.450 (45%)	0.467 (46.7%)
8 – 14	0.375 (37.5%)	0.111 (11.7%)
15 – 30	0.0875 (8.75%)	0.222 (22.2%)
> 30	0.0875 (8.75%)	0.200 (20.0%)

2. Un'industria di cioccolata ha confezionato i cioccolatini prodotti in 9 diversi tipi di scatola, che si differenziano per colore e per forma. Viene data la distribuzione doppia di frequenze delle scatole vendute in un mese, per colore e forma:

Forma	Colore			Totale
	Rosso	Verde	Blu	
Quadrata	800	1600	600	3000
Rettangolare	300	250	2200	2750
Esagonale	1950	200	300	2450
Totale	3050	2050	3100	8200

Valutare al livello di significatività $\alpha = 0.05$ se colore e forma delle scatole di cioccolatini sono statisticamente indipendenti.

Soluzione: Ipotesi:

H_0 : Colore e forma delle scatole di cioccolatini sono statisticamente indipendenti

H_a : Colore e forma delle scatole di cioccolatini non sono statisticamente indipendenti

Regione critica: $RC_{0.05} = \chi^2 > 9.4877$.

Valore osservato della statistica test: $\chi^{2,oss} = 5273.86$

Decisione: Si rifiuta H_0 al livello di significatività del 5%

3. Nell'ambito di una ricerca sulla relazione tra lettura e TV, vengono rilevati due fenomeni: il numero di libri letti mensilmente (X) e le ore settimanali passate davanti alla TV (Y). I risultati sono rappresentati nella seguente tabella:

Numero Libri letti (X)	Ore di TV (Y)				Totale
	6	7	8	9	
0	2	4	6	2	14
1	4	6	5	4	19
2	5	2	2	3	12
Totale	11	12	13	9	45

(a) Costruire le distribuzioni di frequenza relative condizionate. Interpretare.

(b) Valutare al livello di significatività $\alpha = 0.05$ se esiste una associazione tra i due fenomeni

Soluzione:

(a) Distribuzioni di frequenza relative condizionate (e marginali)

N. Libri letti (X)	Ore di TV (Y)					Totale	N. Libri letti (X)	Ore di TV (Y)					Totale
	6	7	8	9				6	7	8	9		
0	0.182	0.333	0.462	0.222	0.311	0.311	0	0.143	0.286	0.429	0.143	1	
1	0.364	0.500	0.385	0.444	0.422	0.422	1	0.211	0.316	0.263	0.211	1	
2	0.455	0.167	0.154	0.333	0.267	0.267	2	0.417	0.167	0.167	0.250	1	
Totale	1	1	1	1	1	1	Totale	0.244	0.267	0.289	0.200	1	

(b) Ipotesi:

H_0 : X e Y sono statisticamente indipendenti

H_a : X e Y non sono statisticamente indipendenti

Regione critica: $RC_{0.05} = \chi^2 > 12.59$.

Valore osservato della statistica test: $\chi^{2,oss} = 4.776$

Decisione: Non si rifiuta H_0 al livello di significatività del 5%

4. In una città per un anno sono stati osservati giornalmente la condizione meteorologica e il livello di traffico automobilistico. Con tali dati si è costruita la seguente tabella a doppia entrata:

Condizione meteorologica	Livello di traffico			Totale
	Basso	Medio	Alto	
Sereno	84	26	11	121
Variabile	29	98	29	156
Pioggia	7	26	55	88
Totale	120	150	95	365

- (a) Valutare al livello di significatività $\alpha = 0.01$ se esiste una associazione tra i due fenomeni utilizzando il test χ^2
- (b) Calcolare i residui standardizzati per studiare la struttura dell'associazione. Commentare i risultati.

Soluzione:

- (a) Ipotesi:

H_0 : Condizione meteorologica e livello del traffico sono statisticamente indipendenti

H_a : Condizione meteorologica e livello del traffico non sono statisticamente indipendenti

Regione critica: $RC_{0.01} = \chi^2 > 13.2767$.

Valore osservato della statistica test: $\chi^{2,oss} = 169.18$

Decisione: Si rifiuta H_0 al livello di significatività del 1%

- (b) Residui standardizzati:

Condizione meteorologica	Livello di traffico		
	Basso	Medio	Alto
Sereno	10.466	-5.362	-5.193
Variabile	-5.020	7.288	-2.798
Pioggia	-5.713	-2.528	8.951

5. La seguente tabella riporta i dati relativi a un gruppo di ragazzi di 14 anni di cui alcuni hanno una storia di bronchite prima dei 5 anni.

	Bronchite	Non Bronchite	Totale
Tosse	26	44	70
Non tosse	247	1002	1249
Totale	273	1046	1319

- (a) Studiare, usando l'Odds Ratio, l'associazione tra la storia di bronchite e la tosse ricorrente.
- (b) Costruire un intervallo di confidenza per l'Odds Ratio al livello di confidenza del 99%.
- (c) Commentare i risultati.

Soluzione:

- (a) $\hat{OR} = 2.4$

- (b) $IC_{0.99}(OR) = (1.234; 4.657)$

6. La seguente tabella riporta i dati di una agenzia assicurativa riguardo la gravità degli incidenti stradali e il sesso del conducente.

	Grave	Lieve	Totale
Maschi	198	83	281
Femmine	107	192	299
Totale	305	275	580

- (a) Valutare il grado di associazione tra sesso del conducente e gravità dell'incidente tramite l'OR e costruire il relativo intervallo di confidenza (al livello di confidenza del 95%).
- (b) Sottoporre a test l'ipotesi di indipendenza tra sesso del conducente e gravità dell'incidente utilizzando il test χ^2 (utilizzare $\alpha = 0.05$).
- (c) Calcolare i residui standardizzati.
- (d) Commentare i risultati.

Soluzione:

(a) $\hat{OR} = 4.28$, $IC_{0.95}(OR) = (3.02; 6.07)$

(b) Ipotesi:

H_0 : Sesso del conducente e gravità dell'incidente sono statisticamente indipendenti

H_a : Sesso del conducente e gravità dell'incidente non sono statisticamente indipendenti

Regione critica: $RC_{0.05} = \chi^2 > 3.84$.

Valore osservato della statistica test: $\chi^{2,oss} = 69.86$

Decisione: Si rifiuta H_0 al livello di significatività del 5%

(c) Residui standardizzati

	Grave	Lieve
Maschi	8.358	-8.358
Femmine	-8.358	8.358

Esercizi dal libro di testo: Capitolo 8

8.1 8.3 8.5 8.6 8.8 8.9 8.10 8.13 8.15 8.16 8.19 8.20 8.21 8.22 8.24
8.33 8.34 8.37 8.38(a, b, c) 8.45