

Levine, Krehbiel, Berenson

Statistica II ed.

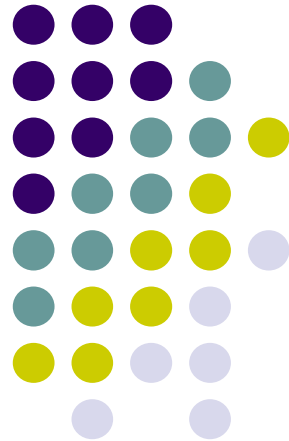
© 2006 Apogeo

Capitolo 11

Test chi-quadro

Insegnamento: Statistica
Corsi di Laurea Triennale in Economia

Facoltà di Economia, Università di Ferrara
Docenti: Dott. S. Bonnini, Dott.ssa Susanna Ragazzi



Argomenti

- Test Chi-quadrato per la differenza tra due proporzioni (campioni indipendenti)
- Test Chi-quadrato per la differenza tra più di due proporzioni
- Test Chi-quadrato per l'indipendenza

Test Chi-quadrato per la differenza tra due proporzioni

- Il problema della verifica di ipotesi sulla differenza tra due proporzioni (test Z) può essere affrontato anche con una procedura alternativa basata su una statistica test la cui distribuzione tende ad approssimarsi con una **distribuzione chi-quadrato** (χ^2). I risultati ottenuti saranno del tutto equivalenti a quelli dalla statistica Z
- Se siamo interessati a confrontare le proporzioni di casi che presentano una certa caratteristica in due gruppi indipendenti possiamo costruire una **tabella a doppia entrata** (o **di contingenza**) di dimensioni 2×2 nella quale sono riportati il numero (o le percentuali) di successi e insuccessi nei due gruppi

Test Chi-quadrato per la differenza tra due proporzioni

Variabile di riga	Variabile di colonna (gruppi)		Totale
	1	2	
Successi	X_1	X_2	X
Insuccessi	$n_1 - X_1$	$n_2 - X_2$	$n - X$
Totale	n_1	n_2	n

dove:

- X_1 = numero di successi nel gruppo 1
- X_2 = numero di successi nel gruppo 2
- $n_1 - X_1$ = numero di insuccessi nel gruppo 1
- $n_2 - X_2$ = numero di insuccessi nel gruppo 2
- $X = X_1 + X_2$ = numero totale di successi
- $n - X = (n_1 - X_1) + (n_2 - X_2)$ = numero totale di insuccessi
- n_1 = ampiezza campionaria del gruppo 1
- n_2 = ampiezza campionaria del gruppo 2
- $n = (n_1 + n_2)$ = ampiezza campionaria totale

- Ad esempio

Tornerà nell'albergo	Albergo		Totale
	Beachcomber	Windsurfer	
Sì	163	154	317
No	64	108	172
Totale	227	262	489

Test Chi-quadrato per la differenza tra due proporzioni

- Per verifica l'ipotesi nulla secondo cui non c'è differenza tra le due proporzioni

$$H_0: \pi_1 = \pi_2$$

contro l'alternativa

$$H_1: \pi_1 \neq \pi_2$$

si può considerare la statistica χ^2

Statistica test χ^2 per la differenza tra due proporzioni

$$\chi^2 = \sum_{\text{tutte le celle}} \frac{(f_0 - f_e)^2}{f_e} \quad (11.1)$$

- La statistica χ^2 si ottiene calcolando per ogni cella della tabella di contingenza la differenza al quadrato fra la frequenza osservata (f_0) e quella attesa (f_e), divisa per f_e , e sommando quindi il risultato ottenuto per ogni cella

Test Chi-quadrato per la differenza tra due proporzioni

- Per calcolare la frequenza attesa si deve tener conto del fatto che se l'ipotesi nulla è vera la proporzione di successi e insuccessi è la stessa nei due gruppi e le proporzioni campionarie dovrebbero differire solo per effetto del caso. In questo caso per stimare il parametro π conviene utilizzare una combinazione delle due frequenze campionarie, indicata con \bar{p}

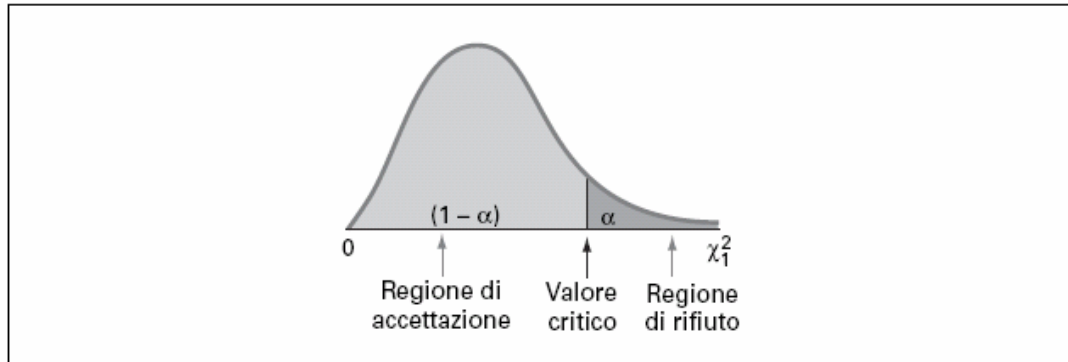
Calcolo della proporzione globale di successi

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{X}{n} \quad (11.2)$$

- Per calcolare la frequenza attesa f_e per le celle relative al successo (prima riga) si dovrà moltiplicare l'ampiezza campionaria n_1 e n_2 (totali di colonna) per \bar{p}

Test Chi-quadrato per la differenza tra due proporzioni

- Analogamente, per calcolare la frequenza attesa f_e per le celle relative all'insuccesso (seconda riga) si dovrà moltiplicare l'ampiezza campionaria n_1 e n_2 di ciascuno dei due gruppi per $(1 - \bar{p})$
- La statistica test introdotta nell'equazione (11.1) si distribuisce approssimativamente secondo una distribuzione chi-quadrato con 1 grado di libertà



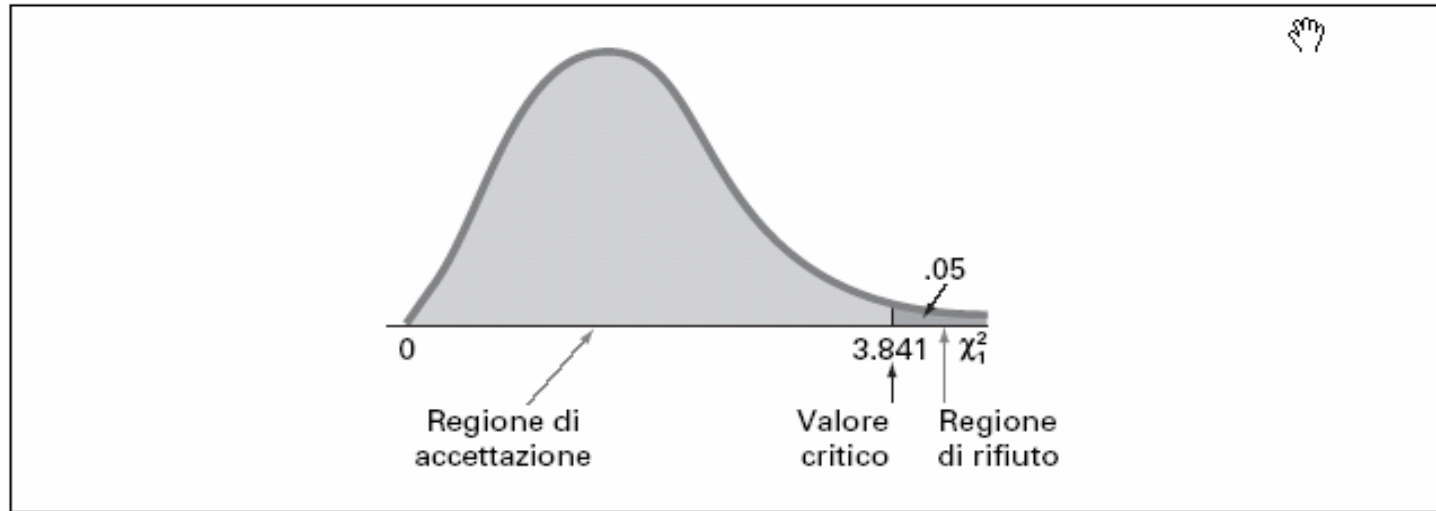
- Fissato α , l'ipotesi nulla dovrà essere rifiutata se il valore osservato della statistica χ^2 è maggiore del valore critico χ^2_U di una distribuzione χ^2 con 1 grado di libertà

Test Chi-quadrato per la differenza tra due proporzioni

Albergo	Beachcomber		Windsurfer		Totale
	Osservate	Attese	Osservate	Attese	
Tornerà nell'albergo					
Sì	163	147.16	154	169.84	317
No	64	79.84	108	92.16	172
Totale	227	227.00	262	262.00	489

f_0	f_e	$(f_0 - f_e)$	$(f_0 - f_e)^2$	$(f_0 - f_e)^2 / f_e$
163	147.16	15.84	250.9056	1.705
154	169.84	-15.84	250.9056	1.477
64	79.84	-15.84	250.9056	3.143
108	92.16	15.84	250.9056	<u>2.723</u>
				9.048

Test Chi-quadrato per la differenza tra due proporzioni



Area nella coda di destra

Gradi di libertà	.995	.9905	.025	.01	.005
1			...	3.841	5.024	6.635	7.879
2	0.010	0.020	...	5.991	7.378	9.210	10.597
3	0.072	0.115	...	7.815	9.348	11.345	12.838
4	0.207	0.297	...	9.488	11.143	13.277	14.860
5	0.412	0.554	...	11.071	12.833	15.086	16.750

Fonte: Tavola E.4.

Test Chi-quadrato per la differenza tra due proporzioni

	A	B	C	D	E	F	G
1	Analisi della soddisfazione dei clienti						
2							
3	Frequenze osservate						
4	Albergo						
5	Tornerà nell'albergo	Beachcomber	Windsurfer	Totale		f_o-f_e	
6	Si	163	154	317		16,84458	-15,8446
7	No	64	108	172		-15,8446	15,84458
8	Totale	227	262	489			
9							
10	Frequenze attese						
11	Albergo						
12	Tornerà nell'albergo	Beachcomber	Windsurfer	Totale		(f_o-f_e)²/f_e	
13	Si	147,1554	169,8446	317		1,706024	1,47812
14	No	79,8446	92,1554	172		3,144243	2,72421
15	Totale	227	262	489			
16							
17	Dati						
18	Livello di significatività	0,05					
19	Numero di righe	2					
20	Numero di colonne	2					
21	Gradi di libertà	1					
22							
23	Risultati						
24	Valore critico	3,8415					
25	Statistica test chi-quadro	9,0526					
26	p-Value	0,0026					
27	Rifiutare l'ipotesi nulla						
28							
29	L'assunzione sulle frequenze attese						
30	è soddisfatta.						

Test Chi-quadrato per la differenza tra più di due proporzioni

- Il test χ^2 opportunamente generalizzato può essere utilizzato per confrontare le proporzioni di più popolazioni indipendenti. Supponiamo di voler verificare l'ipotesi nulla secondo cui le proporzioni di c popolazioni sono uguali

$H_0: \pi_1 = \pi_2 = \dots = \pi_c$ contro l'alternativa

$H_1: \text{non tutte le } \pi_j \text{ sono uguali tra loro (con } j=1, \dots, c)$

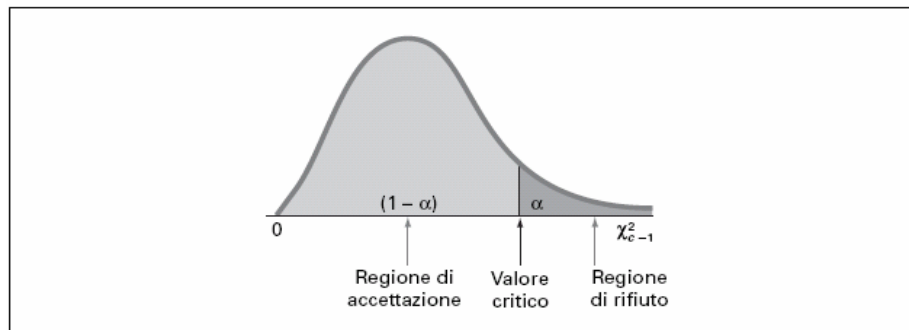
- Per risolvere questo problema dovremo costruire una tabella di contingenza di due righe (successo e insuccesso) che avrà un numero di colonne pari a c
- La statistica test sarà la stessa dell'equazione (11.1), dove la frequenza attesa viene calcolata dalla stima di $\pi = \pi_1 = \pi_2 = \dots = \pi_c$ che in questo caso si ottiene come combinazione delle c frequenze campionarie

Test Chi-quadrato per la differenza tra più di due proporzioni

Calcolo della proporzione globale di successi

$$\bar{p} = \frac{X_1 + X_2 + \dots + X_c}{n_1 + n_2 + \dots + n_c} = \frac{X}{n} \quad (11.3)$$

- La statistica test dell'equazione (11.1) si distribuisce approssimativamente secondo una distribuzione chi-quadrato con $(2-1) \times (c-1) = (c-1)$ gradi di libertà
- Fissato α , l'ipotesi nulla dovrà essere rifiutata se il valore osservato della statistica χ^2 è maggiore del valore critico χ^2_U di una distribuzione χ^2 con $(c-1)$ gradi di libertà



Test Chi-quadrato per la differenza tra più di due proporzioni

Esempio: tabella di contingenza 2x3 relativa alla soddisfazione dei clienti di 3 alberghi

	Albergo			
Tornerà nell'albergo	Golden Palm	Palm Royale	Palm Princess	Totale
Sì	128	199	186	513
No	<u>88</u>	<u>33</u>	<u>66</u>	<u>187</u>
Totale	216	232	252	700

Frequenze attese

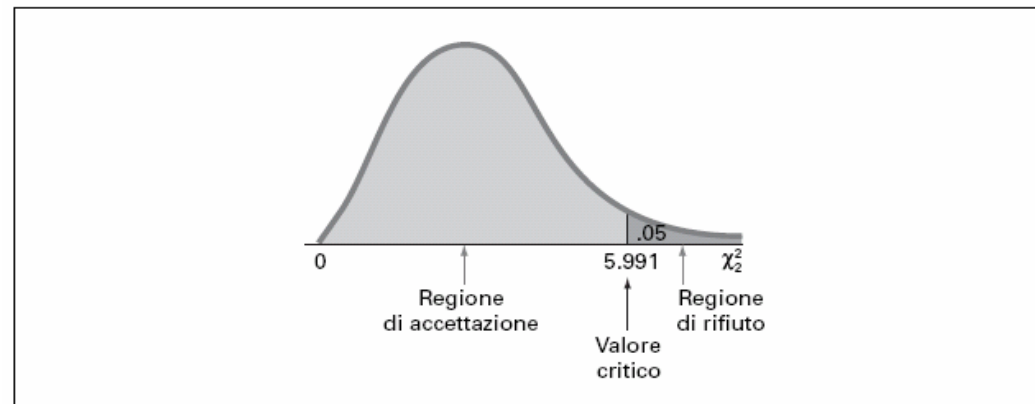
	Albergo			
Tornerà nell'albergo	Golden Palm	Palm Royale	Palm Princess	Totale
Sì	158.30	170.02	184.68	513
No	<u>57.70</u>	<u>61.98</u>	<u>67.32</u>	<u>187</u>
Totale	216.00	232.00	252.00	700

Test Chi-quadrato per la differenza tra più di due proporzioni

Calcolo della statistica test χ^2 per l'esempio relativo alla soddisfazione dei clienti di 3 alberghi

f_0	f_e	$(f_0 - f_e)$	$(f_0 - f_e)^2$	$(f_0 - f_e)^2 / f_e$
128	158.30	-30.30	918.0900	5.800
199	170.02	28.98	839.8404	4.940
186	184.68	1.32	1.7424	0.009
88	57.70	30.30	918.0900	15.911
33	61.98	-28.98	839.8404	13.550
66	67.32	-1.32	1.7424	<u>0.026</u>
				40.236

Regione di rifiuto e di accettazione del test χ^2 (2 gradi di libertà) per la differenza tra tre proporzioni al livello di significatività $\alpha=0.05$



Test Chi-quadrato per la differenza tra più di due proporzioni

Foglio di Microsoft Excel con i calcoli necessari per calcolare valore critico e p -value (test chi-quadro, 3 gruppi)

	A	B	C	D	E	F	G	H	I
1	Analisi della soddisfazione dei clienti (3 alberghi)								
2									
3	Frequenze osservate								
4		Albergo				Calcoli			
5	Tornerà nell'albergo	Golden Palm	Palm Royale	Palm Princess	Totale		fo-fe		
6	Si	128	199	186	513	-30,2971	28,97714	1,32	
7	No	88	33	66	187	30,29714	-28,9771	-1,32	
8	Totale	216	232	252	700				
9									
10	Frequenze attese								
11		Albergo							
12	Tornerà nell'albergo	Golden Palm	Palm Royale	Palm Princess	Totale		(fo-fe)²/fe		
13	Si	158,2971	170,0229	184,68	513	5,798695	4,9386	0,009435	
14	No	57,7029	61,9771	67,32	187	15,90766	13,54814	0,026882	
15	Totale	216	232	252	700				
16									
17	Dati								
18	Livello di significatività	0,05							
19	Numero di righe	2							
20	Numero di colonne	3							
21	Gradi di libertà	2							
22									
23	Risultati								
24	Valore critico	5,9915							
25	Statistica test chi-quadro	40,2284							
26	p-Value	0,0000							
27	Rifiutare l'ipotesi nulla								
28									
29	L'assunzione sulle frequenze attese								
30	è soddisfatta.								

Test Chi-quadrato per l'indipendenza

- Se si considera una tabella di contingenza con r righe c colonne il procedimento del test χ^2 può essere generalizzato per verificare l'indipendenza tra due variabili categoriali X e Y
- In questo contesto le ipotesi nulla e alternativa sono
 H_0 : le due variabili categoriali sono indipendenti ($X \perp Y$)
(non sussistono relazioni tra le due variabili)
 H_1 : le due variabili categoriali sono dipendenti ($X \not\perp Y$)
(sussiste una relazione tra le due variabili)

- Il test si basa ancora una volta sull'equazione (11.1)

$$\chi^2 = \sum_{\text{tutte le celle}} \frac{(f_o - f_e)^2}{f_e}$$

- La regola decisionale consiste nel rifiutare H_0 se il valore osservato della statistica χ^2 è maggiore del valore critico χ^2_U della distribuzione χ^2 con $(r-1) \times (c-1)$ gdl

Test Chi-quadrato per l'indipendenza

- Pure se presentano delle analogie, la differenza fondamentale tra il test chi-quadrato per le proporzioni e per l'indipendenza riguarda lo schema di campionamento:
 - Nel confronto tra proporzioni siamo di fronte a campioni estratti da popolazioni indipendenti
 - Nel test di indipendenza abbiamo un solo campione su cui rileviamo due variabili qualitative che possono assumere r e c modalità distinte
- Nel caso di test chi-quadrato per l'indipendenza è possibile semplificare il calcolo delle frequenze attese applicando la seguente regola:

$$f_e = \frac{\text{totale di riga} \times \text{totale di colonna}}{n}$$

Test Chi-quadrato per l'indipendenza

Esempio: tabella della frequenze osservate con riferimento al principale motivo di insoddisfazione e all'albergo

Motivo di insoddisfazione	Albergo			Totale
	Golden Palm	Palm Royale	Palm Princess	
Prezzo	23	7	37	67
Posizione dell'albergo	39	13	8	60
Sistemazione nelle camere	13	5	13	31
Altro	<u>13</u>	<u>8</u>	<u>8</u>	<u>29</u>
Totale	88	33	66	187

Frequenze attese

Motivo di insoddisfazione	Albergo			Totale
	Golden Palm	Palm Royale	Palm Princess	
Prezzo	31.53	11.82	23.65	67
Posizione dell'albergo	28.24	10.59	21.18	60
Sistemazione nelle camere	14.59	5.47	10.94	31
Altro	13.65	5.12	10.24	29
Totale	88.00	33.00	66.00	187

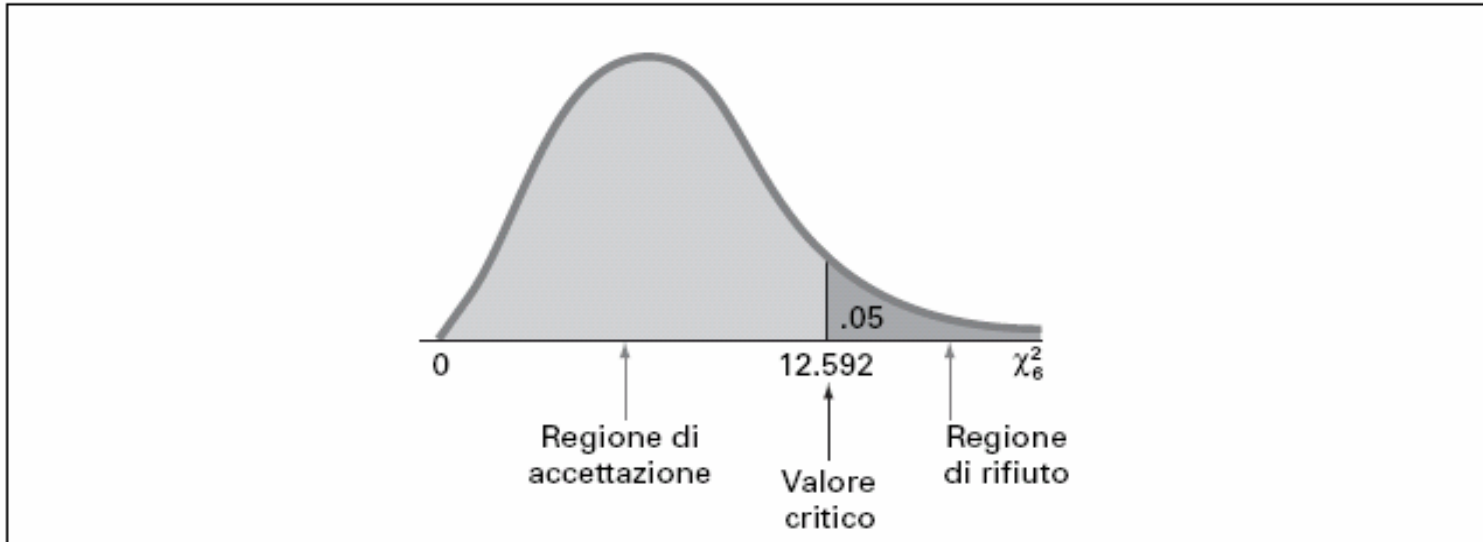
Test Chi-quadrato per l'indipendenza

Calcolo della statistica χ^2 per il test di indipendenza

Cella	f_0	f_e	$(f_0 - f_e)$	$(f_0 - f_e)^2$	$(f_0 - f_e)^2/f_e$
Prezzo/Golden Palm	23	31.53	-8.53	72.7609	2.308
Prezzo/Palm Royale	7	11.82	-4.82	23.2324	1.966
Prezzo/Palm Princess	37	23.65	13.35	178.2225	7.536
Posizione/Golden Palm	39	28.24	10.76	115.7776	4.100
Posizione/Palm Royale	13	10.59	2.41	5.8081	0.548
Posizione/Palm Princess	8	21.18	-13.18	173.7124	8.202
Stanza/Golden Palm	13	14.59	-1.59	2.5281	0.173
Stanza/Palm Royale	5	5.47	-0.47	0.2209	0.040
Stanza/Palm Princess	13	10.94	2.06	4.2436	0.388
Altro/Golden Palm	13	13.65	-0.65	0.4225	0.031
Altro/Palm Royale	8	5.12	2.88	8.2944	1.620
Altro/Palm Princess	8	10.24	-2.24	5.0176	<u>0.490</u>
					27.402

Test Chi-quadrato per l'indipendenza

Regione di rifiuto e di accettazione del test χ^2 per l'indipendenza nell'esempio sulla soddisfazione dei clienti (al livello di significatività 0.05 con 6 gradi di libertà)



Test Chi-quadrato per l'indipendenza

Foglio di Microsoft Excel con i calcoli necessari per la verifica dell'ipotesi di indipendenza tra motivo di insoddisfazione e albergo

	A	B	C	D	E	F	G	H	I
1	Motivi di insoddisfazione e albergo								
2									
3	Frequenze osservate								
4		Albergo				Calcoli			
5	Motivo di insoddisfazione	Golden Palm	Palm Royale	Palm Princess	Totale	fo-fe			
6	Prezzo	23	7	37	67	-8,5294	-4,8235	13,3529	
7	Posizione	39	13	8	60	10,7547	2,4118	-13,1765	
8	Stanza	13	5	13	31	-1,5882	-0,4706	2,0588	
9	Altro	13	8	8	29	0,6471	2,8824	-2,2353	
10	Totale	88	33	66	187				
11									
12	Frequenze attese								
13		Albergo				(fo-fe) ² /fe			
14	Motivo di insoddisfazione	Golden Palm	Palm Royale	Palm Princess	Totale				
15	Prezzo	31,5294	11,8235	23,5471	67	2,3074	1,9678	7,5401	
16	Posizione	28,2353	10,5882	21,1765	60	4,1040	0,5483	8,1987	
17	Stanza	14,5882	5,4706	10,9412	31	0,1729	0,0405	0,3974	
18	Altro	13,6471	5,1176	10,2353	29	0,0307	1,6234	0,4882	
19	Totale	88	33	66	187				
20									
21	Dati								
22	Livello di significatività	0,05							
23	Numero di righe	4							
24	Numero di colonne	3							
25	Gradi di libertà	6							
26									
27	Risultati								
28	Valore critico	12,5916							
29	Statistica test chi-quadro	27,4104							
30	p-Value	0,00012							
31	Rifiutare l'ipotesi nulla								
32									
33	<i>L'assunzione sulle frequenze attese</i>								
34	<i>è soddisfatta.</i>								

Riepilogo

