Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d'accordo con la frase "L'aborto è una questione privata delle donne e dovrebbe essere lasciata loro la decisione senza l'intervento del governo". 868 degli intervistati si sono detti d'accordo.

a) Fornire una stima della proporzione di maschi americani d'accordo con la frase in oggetto, ed il relativo IC(98%).
b) Verificare al livello di sign. del 5% se la percentuale nell'intera popolazione di maschi favorevoli sia superiore al 70%.

c) Gli uomini sono stati intervistati da intervistatori di entrambi i sessi, secondo la seguente tabella:

	Genere dell'intervistatore	
	Maschio	Femmina
Uomini d'accordo	560	308
Uomini in disaccordo	240	92

Verificare al livello dell'1% l'omogeneità delle risposte rispetto al genere dell'intervistatore.

Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d'accordo con la frase "L'aborto è una questione privata delle donne e dovrebbe essere lasciata loro la decisione senza l'intervento del governo". 868 degli intervistati si sono detti d'accordo.

a) Fornire una stima della proporzione di maschi americani d'accordo con la frase in oggetto, ed il relativo IC(98%).

b) Verificare al livello di sign. del 5% se la percentuale nell'intera popolazione di maschi favorevoli sia superiore al 70%.

c) Gli uomini sono stati intervistati da intervistatori di entrambi i sessi, secondo la seguente tabella

	G	re
		a
Uomini d'accordo		
Uomini in disaccordo	24	92

Sinonimo di indipendenza

Verificare al livello dell'1% l'**omogeneità** delle risposte rispetto al genere dell'intervistatore.

Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d'accordo con la frase "L'aborto è una questione privata delle donne e dovrebbe essere lasciata loro la decisione senza l'intervento del governo". 868 degli intervistati si sono detti d'accordo.

a) Fornire una stima della proporzione di maschi americani d'accordo con la frase in oggetto, ed il relativo IC(98%).

$$\hat{p}_n = \frac{868}{1200} = \mathbf{0.723} \qquad IC(1-\alpha) = \left(\hat{p}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right)$$

$$0.723 \mp z_{\frac{0.02}{2}}\sqrt{\frac{0.723 \times 0.277}{1200}} = 0.723 \mp 2.3263 \times 0.013 \Rightarrow (0.69\,,0.75)$$

Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d'accordo con la frase "L'aborto è una questione privata delle donne e dovrebbe essere lasciata loro la decisione senza l'intervento del governo". 868 degli intervistati si sono detti d'accordo.

a) Fornire una stima della proporzione di maschi americani d'accordo con la frase in oggetto, ed il relativo IC(98%).

$$\hat{p}_n = \frac{868}{1200} = 0.723 \qquad IC(1-\alpha) = \left(\hat{p}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right)$$

$$0.723 \mp z_{\frac{0.02}{2}}\sqrt{\frac{0.723 \times 0.277}{1200}} = 0.723 \mp 2.3263 \times 0.013 \Rightarrow (0.69\,, 0.75)$$

Al livello di confidenza del 98% la percentuale di maschi americani d'accordo con la frase in oggetto varia tra il 69% ed il 75%.

Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d'accordo con la frase "L'aborto è una questione privata delle donne e dovrebbe essere lasciata loro la decisione senza l'intervento del governo". 868 degli intervistati si sono detti d'accordo.

b) Verificare al livello di sign. del 5% se la percentuale nell'intera popolazione di maschi favorevoli sia superiore al 70%.

$$\hat{p}_n = \frac{868}{1200} = 0.723 \qquad H_0 : p = 0.7, \qquad H_1 : p > 0.7$$

$$\frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0 \times (1 - p_0)}{1200}}} = \frac{0.723 - 0.7}{\sqrt{\frac{0.7 \times 0.3}{1200}}} = 1.739 \;?>?\; z_\alpha = z_{0.05} = 1.6449$$

Si rifiuta l'ipotesi nulla che $p \leq 0.7$ al livello del 5%

Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d'accordo con la frase "L'aborto è una questione privata delle d... ...nza l'inte... ...i d'acc...

b) V... ...l'intera p...

Attenzione: avevamo accennato alla corrispondenza tra IC e test.

L'analogo dell'IC di cui al punto a) sarebbe il test

$$H_0 : p = 0.7 \ \ vs \ \ H_1 : p \neq 0.7$$

al livello di significatività del 2%.

$\hat{p}_n =$

$$\frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0 \times (1 - p_0)}{1200}}} = \frac{\ldots}{\sqrt{\frac{0.7 \times 0.3}{1200}}} = 1.739 \, ? > ? \, z_\alpha = z_{0.05} = 1.6449$$

Si rifiuta l'ipotesi nulla che $p \leq 0.7$ al livello del 5%

Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d’accordo con la frase “L’aborto è una questione privata delle donne e dovrebbe essere lasciata loro la decisione senza l’intervento del governo”. 868 degli intervistati si sono detti d’accordo.

c) Gli uomini sono stati intervistati da intervistatori di entrambi i sessi, secondo la seguente tabella:

	Genere dell’intervistatore	
	Maschio	Femmina
Uomini d’accordo	560	308
Uomini in disaccordo	240	92

Verificare al livello dell’1% l’omogeneità delle risposte rispetto al genere dell’intervistatore.

Esempio p. 402 (con variazioni)

Ad un campione casuale di 1200 maschi americani è stato chiesto se sono d'accordo con la frase "L'aborto è una questione privata delle donne e dovrebbe essere lasciata loro la decisione senza l'intervento del governo". 868 degli intervistati si sono detti d'accordo.

c) Gli uomini sono stati intervistati da intervistatori di entrambi i sessi, secondo la seguente tabella:

Freq. attese	**Genere dell'intervistatore**	
	Maschio	Femmina
Uomini d'accordo	560|| **578.7**	308|| **289.3**
Uomini in disaccordo	240|| **221.3**	92|| **110.7**

$\chi^2 = 6.53 ? > ? \ \chi^2(1)_{0.01} = 6.6349 \Rightarrow$ Non possiamo rifiutare l'ipotesi nulla di omogeneità (indipendenza dal genere) al livello dell'1%, però……

Il **test** del chi-quadrato, II

Lanciamo 120 volte un dado ottenendo la distribuzione di frequenza seguente:

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. n_j	20	30	20	25	15	10

c'è abbastanza evidenza nei dati per concludere che il dado non è equilibrato?

Il **test** del chi-quadrato, II

Se il dado è equilibrato in 120 lanci ci aspettiamo:

p_j	1/6	1/6	1/6	1/6	1/6	1/6
Fr. att. n^*_j	**20**	**20**	**20**	**20**	**20**	**20**

$$\boldsymbol{n^*_j} = np_j$$

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. n_j	20	30	20	25	15	10

Il **test** del chi-quadrato, II

Se il dado è equilibrato in 120 lanci ci aspettiamo:

p_j	1/6	1/6	1/6	1/6	1/6	1/6
Fr. att. n^*_j	**20**	**20**	**20**	**20**	**20**	**20**

$$\boldsymbol{n^*}_j = np_j$$

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. n_j	20	30	20	25	15	10

Rifiutiamo al livello α l'ipotesi che il dado sia equilibrato se

$$\sum_{i=1}^{c} \frac{(\boldsymbol{n^*}_j - n_j)^2}{\boldsymbol{n^*}_j} > \chi^2(c-1)_\alpha$$

Il **test** del chi-quadrato, II

Se il dado è equilibrato in 120 lanci ci aspettiamo:

p_j	1/6	1/6	1/6	1/6	1/6	1/6
Fr. att. $n^*{}_j$	**20**	**20**	**20**	**20**	**20**	**20**

$$\boldsymbol{n^*}{}_j = np_j$$

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. n_j	20	30	20	25	15	10

Rifiutiamo al livello α l'ipotesi che il dado sia equilibrato se

$$\sum_{i=1}^{c} \frac{(\boldsymbol{n^*}{}_j - n_j)^2}{\boldsymbol{n^*}{}_j} > \chi^2(c-1)_\alpha$$

$$\frac{(20-20)^2}{20} + \frac{(20-30)^2}{20} + \frac{(20-20)^2}{20} + \frac{(20-25)^2}{20} + \frac{(20-15)^2}{20} + \frac{(20-10)^2}{20} = 12.5$$

Il **test** del chi-quadrato, II

Se il dado è equilibrato in 120 lanci ci aspettiamo:

p_j	1/6	1/6	1/6	1/6	1/6	1/6
Fr. att. n^*_j	**20**	**20**	**20**	**20**	**20**	**20**

$\boldsymbol{n^*}_j = np_j$

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. n_j	20	30	20	25	15	10

Rifiutiamo al livello α l'ipotesi che il dado sia equilibrato se

$$\sum_{i=1}^{c} \frac{(\boldsymbol{n^*}_j - n_j)^2}{\boldsymbol{n^*}_j} > \chi^2(c-1)_\alpha$$

$$\frac{(20-20)^2}{20} + \frac{(20-30)^2}{20} + \frac{(20-20)^2}{20} + \frac{(20-25)^2}{20} + \frac{(20-15)^2}{20} + \frac{(20-10)^2}{20} = 12.5$$

$$> \chi^2(5)_{0.05} = 11.0705$$

Il **test** del chi-quadrato, II

Campione *casuale* $X_1, \dots, X_n \text{ i.i.d}$

$$H_0 : \quad X_i \sim F, \text{discreta} \qquad H_1 : \text{un'altra qualunque}, \neq F$$

Si rifiuta l'ipotesi nulla se:

$$\boxed{\sum_{i=1}^{c} \frac{(n^*{}_j - n_j)^2}{n^*{}_j} > \chi^2(c-1)_\alpha}$$

Test di buon adattamento (*goodness of fit test*)

Esercizio 1

In una certa regione, l’84% degli automobilisti in un anno non ha nessun incidente, il 14% ne ha uno solo, il 2% ne ha due o più. Su un campione casuale di 400 avvocati, 308 non hanno avuto alcun incidente nell’ultimo anno, 66 ne hanno avuto uno solo e 26 ne hanno avuti due o più. Sulla base di questi dati è possibile affermare che gli avvocati non abbiano lo stesso profilo di incidenti del resto della popolazione?

Esercizio 1

In una certa regione, l'84% degli automobilisti in un anno non ha nessun incidente, il 14% ne ha uno solo, il 2% ne ha due o più. Su un campione casuale di 400 avvocati, 308 non hanno avuto alcun incidente nell'ultimo anno, 66 ne hanno avuto uno solo e 26 ne hanno avuti due o più. Sulla base di questi dati è possibile affermare che gli avvocati non abbiano lo stesso profilo di incidenti del resto della popolazione?

Esito	**0**	**1**	**≥2**
Fr. n_j	308	66	26

p_j	**0.84**	**0.14**	**0.02**
Fr. att.	**336**	**56**	**8**

$n^*_j = 400p_j$

Esercizio 1

In una certa regione, l'84% degli automobilisti in un anno non ha nessun incidente, il 14% ne ha uno solo, il 2% ne ha due o più. Su un campione casuale di 400 avvocati, 308 non hanno avuto alcun incidente nell'ultimo anno, 66 ne hanno avuto uno solo e 26 ne hanno avuti due o più. Sulla base di questi dati è possibile affermare che gli avvocati non abbiano lo stesso profilo di incidenti del resto della popolazione?

Esito	**0**	**1**	**≥2**
Fr. $\boldsymbol{n_j}$	308	66	26

$H_0 \; : \; n_j = n^*_j$ per ogni j

$H_1 \; : \; n_j \neq n^*_j$ per almeno un j

p_j	**0.84**	**0.14**	**0.02**
Fr. att.	**336**	**56**	**8**

$$n^*_j = 400 p_j$$

$$\sum_{i=1}^{c} \frac{(n_j - n^*_j)^2}{e_j} > \chi^2(c-1)_\alpha$$

Esercizio 1

In una certa regione, l'84% degli automobilisti in un anno non ha nessun incidente, il 14% ne ha uno solo, il 2% ne ha due o più. Su un campione casuale di 400 avvocati, 308 non hanno avuto alcun incidente nell'ultimo anno, 66 ne hanno avuto uno solo e 26 ne hanno avuti due o più. Sulla base di questi dati è possibile affermare che gli avvocati non abbiano lo stesso profilo di incidenti del resto della popolazione?

Esito	**0**	**1**	**≥2**
Fr. n_j	308	66	26

p_j	**0.84**	**0.14**	**0.02**
Fr. att.	**336**	**56**	**8**

$H_0 \;:\; n_j = n^*_j$ per ogni j

$H_1 \;:\; n_j \neq n^*_j$ per almeno un j

$$n^*_j = 400 p_j$$

$$\sum_{i=1}^{c} \frac{(n_j - n^*_j)^2}{e_j} > \chi^2(c-1)_\alpha$$

$$\frac{(308-336)^2}{336} + \frac{(66-56)^2}{56} + \frac{(26-8)^2}{8} = 44.62 \;\; ?>? \; \chi^2(2)_{0.01} = 9.21$$

Esercizio di compito

La seguente tabella riporta la ripartizione per fasce d'età delle donne americane non sposate con figli nel 1986 (U.S. National Center for Health Statistics, *Vital Statistics of the United States*):

Fascia d'età	**%**
≤ 14	1.1
15-|19	32.0
20-|24	36.0
25-|29	18.9
≥ 30	12.0

Da un recente campione di 1000 nati da donne non sposate risulta che 42 delle madri hanno 14 anni o meno, 403 tra i 15 ed i 19 anni, 315 tra 20 e 24 anni, 150 tra 25 e 29 anni e 90 ne hanno almeno 30. Si può sostenere che la distribuzione sia variata?