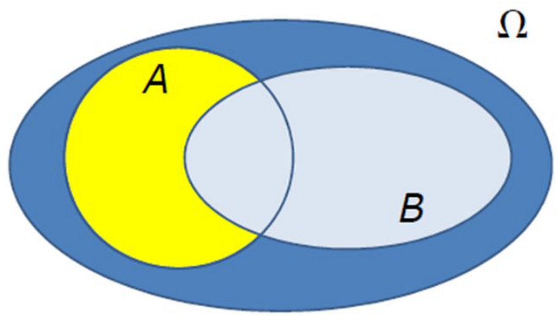
STATISTICA

Tabelle di contingenza
test di indipendenza
test di buon adattamento

Tabelle di contingenza

La probabilità condizionata

A = evento che ci interessa
B = evento "collegato", P(B) ≠ 0



Probabilità condizionata:

$$\mathbf{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

Gli eventi A e B si definiscono (*stocasticamente*) **indipendenti** se

$$P(A|B) = \frac{72}{80} = 0.9 \qquad \mathbf{P(A|B) = P(A)} \qquad P(A) = \frac{90}{100} = 0.9$$

A	B Test positivo (gravidanza)	Test negativo (no gravidanza)	Marginale gravidanza
Incinta	72	18	**90**
Non incinta	5	2	**10**
Marginale test	**80**	**20**	***100***

Tabelle di contingenza

ANTICIPAZIONE

	X=1	X=0	Marginale Y
Y=1	80	5	85
Y=0	3	11	14
Marginale X	83	16	99

Tabella di **contingenza**, o *a doppia entrata*

Analisi dell'**associazione** (assenza di indipendenza) tra le variabili categoriche X e Y

Tabelle di contingenza

tabella a
doppia entrata

(X, Y)

		Y				
		y_1	y_2	…	y_k	TOT
X	x_1					$n_{1\cdot}$
	x_2					$n_{2\cdot}$
	…					
	x_h					$n_{h\cdot}$
	TOT	$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot k}$	n

n_{ij} distribuzione **congiunta**

distribuz. **marginale** di X (somma nelle righe)

distribuz. **marginale** di Y (somma nelle colonne)

Le tabelle di contingenza

Vogliamo vedere se la credenza nell'aldilà è ugualmente diffusa tra uomini e donne. Selezioniamo **un campione casuale di n soggetti** che classifichiamo nelle 2×2 caselle, rispetto alle due variabili:

Le tabelle di contingenza

Vogliamo vedere se la credenza nell'aldilà è **ugualmente diffusa tra uomini e donne**. Selezioniamo **un campione casuale di n soggetti** che classifichiamo nelle 2×2 caselle, rispetto alle due variabili:

la credenza è
indipendente
dal genere

Le tabelle di contingenza

Vogliamo vedere se la credenza nell'aldilà è ugualmente diffusa tra uomini e donne. Selezioniamo **un campione casuale di n soggetti** che classifichiamo nelle 2×2 caselle, rispetto alle due variabili:

(Agresti. p. 21)

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	1127

Se la proporzione di credenti è la stessa nei due generi, allora possiamo dire che non c'è associazione con il genere.

Le tabelle di contingenza

Vogliamo vedere se la credenza nell'aldilà è ugualmente diffusa tra uomini e donne. Selezioniamo **un campione casuale di n soggetti** che classifichiamo nelle 2×2 caselle, rispetto alle due variabili:

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	1127

$$\frac{509}{625} = 0.814$$

$$\frac{398}{502} = 0.793$$

Se la proporzione di credenti è la stessa nei due generi, allora possia

probabilità condizionate di essere credente sono le stesse

$$P(Credente|Maschio) = P(Credente|Femmina)$$

Confronto di proporzioni

variabile risposta

variabile esplicativa

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	
Maschio	398	104	
			$\boxed{1127}$

ipotesi sulla distribuzione dei dati

campionamento multinomiale

Confronto di proporzioni

variabile risposta

variabile esplicativa

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	1127

ipotesi sulla distribuzione dei dati

campionamento multinomiale

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	$\hat{p}_1 = 0.814$	0.186	1
Maschio	$\hat{p}_2 = 0.793$	0.207	1

2 campionamenti binomiali ***indipendenti***

test di cfr di prop.

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\,(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 0.884 \Rightarrow$$ non posso rifiutare l'ip. nulla di uguali proporzioni

Confronto di proporzioni

variabile risposta

variabile esplicativa

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	1127

ipotesi sulla distribuzione dei dati

campionamento multinomiale

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	$\hat{p}_1$=0.814	0.186	1
Maschio	$\hat{p}_2$=0.793	0.207	1

2 campionamenti binomiali ***indipendenti***

test di cfr di prop.

IC asint. $\hat{p}_1 - \hat{p}_2 \mp z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = (-0.026, 0.068)$

Confronto di proporzioni

variabile risposta

variabile esplicativa

(Agresti. p. 27)
(Statistica-5c… p. 9)

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	189	10845	11034
Aspirina	104	10933	11037

stabilito in anticipo

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	0.0171289		1
Aspirina	0.00942285		1

$\hat{p}_1 - \hat{p}_2 = 0.00770602$; IC asint. (95%): $(0.00468775, 0.0107243)$

$p_1 > p_2$ cioè l'aspirina sembra diminuire il rischio di infarto miocardico.

$\frac{p_1}{p_2}$ rischio relativo

$\hat{p}_1 \approx 2\hat{p}_2$!

Rischio relativo

rischio relativo $\frac{p_1}{p_2}$

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	0.0171289		1
Aspirina	0.00942285		1

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.82$$

IC asint. (1-α)%: $$log\left(\frac{\hat{p}_1}{\hat{p}_2}\right) \mp z_{1-\alpha/2}\sqrt{\frac{(1-\hat{p}_1)}{n_1\hat{p}_1} + \frac{(1-\hat{p}_2)}{n_2\hat{p}_2}}$$

$$(0.359792\,, 0.835464) \xrightarrow[\text{(exp)}]{} (1.43303\,, 2.30588)$$

Rischio relativo

rischio relativo $\frac{p_1}{p_2}$

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	0.0171289		1
Aspirina	0.00942285		1

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.82$$

IC asint. (1-α)%: $$log\left(\frac{\hat{p}_1}{\hat{p}_2}\right) \mp z_{1-\alpha/2}\sqrt{\frac{(1-\hat{p}_1)}{n_1\hat{p}_1} + \frac{(1-\hat{p}_2)}{n_2\hat{p}_2}}$$

$$(0.359792\,, 0.835464) \xrightarrow{(\exp)} (1.43303\,, 2.30588)$$

quale più informativo?

$\hat{p}_1 - \hat{p}_2 = 0.00770602;$ IC asint. (95%): $(0.00468775, 0.0107243)$

Rischio relativo

rischio relativo $\frac{p_1}{p_2}$

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	0.0171289		1
Aspirina	0.00942285		1

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.82$$

IC asint. (1-α)%: $$log\left(\frac{\hat{p}_1}{\hat{p}_2}\right) \mp z_{1-\alpha/2}\sqrt{\frac{(1-\hat{p}_1)}{n_1\hat{p}_1} + \frac{(1-\hat{p}_2)}{n_2\hat{p}_2}}$$

$$(0.359792\,, 0.835464) \xrightarrow{(exp)} (1.43303\,, 2.30588)$$

quale più informativo?

$\hat{p}_1 - \hat{p}_2 = 0.00770602;$ IC asint. (95%): $(0.00468775, 0.0107243)$

$$\hat{p}_1 = 0.41770602 \text{ e } \hat{p}_2 = 0.41 \Rightarrow \hat{p}_1 - \hat{p}_2 = 0.00770602 \;\; \boldsymbol{ma} \;\; \frac{\hat{p}_1}{\hat{p}_2} = 1.02;$$

Rischi relativo e *odds ratio*

variabile risposta

variabile esplicativa

	Credenza nell'aldilà	
Sesso	**Sì**	**No**
Femmina	$\hat{p}_1 = 0.814$	0.186
Maschio	$\hat{p}_2 = 0.793$	0.207

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.03$$

≈ 1
indicatore di indipendenza

odds di successo in F

ODDS RATIO

$$\frac{\boxed{\hat{p}_1/(1-\hat{p}_1)}}{\boxed{\hat{p}_2/(1-\hat{p}_2)}} =$$

odds di successo in M

Rischi relativo e *odds ratio*

variabile risposta

variabile esplicativa

	Credenza nell'aldilà	
Sesso	**Sì**	**No**
Femmina	$\hat{p}_1 = 0.814$	0.186
Maschio	$\hat{p}_2 = 0.793$	0.207

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.03$$

≈ 1
indicatore di indipendenza

odds di successo in F

ODDS RATIO

$$\frac{\boxed{\hat{p}_1/(1-\hat{p}_1)}}{\boxed{\hat{p}_2/(1-\hat{p}_2)}} = \frac{\frac{0.814}{0.186}}{\frac{0.793}{0.207}} = 1.142$$

odds di successo in M

Rischi relativo e *odds ratio*

variabile risposta

variabile esplicativa

	Credenza nell'aldilà	
Sesso	**Sì**	**No**
Femmina	$\hat{p}_1$=0.814	0.186
Maschio	$\hat{p}_2$=0.793	0.207

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.03$$

≈ 1
indicatore di indipendenza

odds di successo in F

ODDS RATIO

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{\frac{0.814}{0.186}}{\frac{0.793}{0.207}} = 1.142$$

odds di successo in M

≈ 1
indicatore di indipendenza

$$\hat{p}_1 = \frac{\text{odds}_1}{\text{odds}_1 + 1}$$

$OR > 1 \Leftrightarrow \hat{p}_1 > \hat{p}_2$

Rischi relativo e *odds ratio*

variabile risposta

variabile esplicativa

	Credenza nell'aldilà	
Sesso	**Sì**	**No**
Femmina	$\hat{p}_1 = 0.814$	0.186
Maschio	$\hat{p}_2 = 0.793$	0.207

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.03$$

≈ 1
indicatore di indipendenza

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = 1.14$$

Rischi relativo e *odds ratio*

variabile risposta

variabile esplicativa

	Credenza nell'aldilà	
Sesso	**Sì**	**No**
Femmina	$\hat{p}_1 = 0.814$	0.186
Maschio	$\hat{p}_2 = 0.793$	0.207

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.03$$

≈ 1
indicatore di
indipendenza

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = 1.14$$

	Infarto miocardico	
Gruppo	**Sì**	**No**
Placebo	0.0171289	0.982871
Aspirina	0.00942285	0.990577

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.82$$

$\gg 1$
indicatore di
dipendenza

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = 1.83$$

Rischi relativo e *odds ratio*

variabile risposta

variabile esplicativa

	Credenza nell'aldilà	
Sesso	**Sì**	**No**
Femmina	$\hat{p}_1$=0.814	0.186
Maschio	$\hat{p}_2$=0.793	0.207

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.03$$

≈ 1
indicatore di indipendenza

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = 1.14$$

	Infarto miocardico	
Gruppo	**Sì**	**No**
Placebo	0.0171289	0.982871
Aspirina	0.00942285	0.990577

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.82$$

$\gg 1$
indicatore di **di**pendenza

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = 1.83$$

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{\hat{p}_1}{\hat{p}_2} \times \boxed{\frac{1-\hat{p}_2}{1-\hat{p}_1}}$$

per $\hat{p}_i$ entrambi vicini a 0

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} \approx \frac{\hat{p}_1}{\hat{p}_2}$$

Rischi relativo e *odds ratio*

variabile risposta

variabile esplicativa

	Credenza nell'aldilà	
Sesso	**Sì**	**No**
Femmina	$\hat{p}_1$=0.814	0.186
Maschio	$\hat{p}_2$=0.793	0.207

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.03$$

≈ 1
indicatore di indipendenza

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = 1.14$$

	Infarto mioca	
Gruppo	**Sì**	
Placebo	0.0171289	0.9
Aspirina	0.00942285	0.9

(regola del pollice)

Hosmer & Lemeshow *Applied Logistic Regression*

P(*successo*) ≤ 0.10

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{\hat{p}_1}{\hat{p}_2} \times \boxed{\frac{1-\hat{p}_2}{1-\hat{p}_1}}$$

per $\hat{p}_i$ entrambi vicini a 0

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} \approx \frac{\hat{p}_1}{\hat{p}_2}$$

Odds ratio

variabile risposta

variabile esplicativa

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	189	10845	11034
Aspirina	104	10933	11037

$$\frac{\hat{p}_1}{\hat{p}_2} = 1.82$$

OR$= 1.83$

	Gruppo		
IM	**Placebo**	**Aspirina**	
Sì	189**/293=0.645**	104 /293	**293**
No	10845**/21778=0.498**	10933/21778	**21778**

$$\frac{\hat{p}_1}{\hat{p}_2} = \frac{0.645}{0.498} = 1.29$$

OR$= 1.83$

OR tratta in modo simmetrico la variabile risposta e quella esplicativa. Indifferente al condizionamento.

Odds ratio

variabile risposta

variabile esplicativa

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	189	10845	11034
Aspirina	104	10933	11037

$\frac{\hat{p}_1}{\hat{p}_2} = 1.82$

OR= 1.83

	Gruppo		
IM	**Placebo**	**Aspirina**	
Sì	189**/293=0.645**	104 /293	**293**
No	10845**/21778=0.498**	10933/21778	**21778**

$\frac{\hat{p}_1}{\hat{p}_2} = 1.29$

OR= 1.83

OR tratta in modo simmetrico la variabile risposta e quella esplicativa. Indifferente al condizionamento.

OR=$\frac{n_{11} \times n_{22}}{n_{12} \times n_{21}}$

$\boldsymbol{n_{11}}$	n_{12}
n_{21}	$\boldsymbol{n_{22}}$

Odds ratio: caso-controllo

variabile risposta

variabile esplicativa

Gruppo	IM	Controllo
Fumatore	172	173
Non fumatore	90	346
	262	≈**262**×**2**

(Agresti. p. 26)

studio retrospettivo

Ogni *caso* (262 donne di età<69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) accoppiato con 2 *controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**

Odds ratio: caso-controllo

variabile risposta

Gruppo	IM	Controllo
	variabile esplicativa	
Fumatore	172	173
Non fumatore	90	346
	262	$\approx$**262**$\times$**2**

(Agresti. p. 26)

Ogni *caso* (262 donne di età<69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) accoppiato con 2 *controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**

le 262x3 u.s. **non** si distribuiscono ***casualmente*** nelle 4 caselle.

$$\frac{262}{262 \times 3} = \frac{1}{3}$$

non è una stima della frazione di infartuati nella popolazione

$$\frac{172}{172 + 173} = 0.5$$

non è la prob. che a una donna venga l'infarto sapendo che fuma.

Odds ratio: caso-controllo

variabile risposta

Gruppo	IM	Controllo
	variabile esplicativa	
Fumatore	172	173
Non fumatore	90	346
	262	**≈262×2**

(Agresti. p. 26)

Ogni *caso* (262 donne di età<69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) accoppiato con 2 *controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**

le 262x3 u.s. **non** si distribuiscono ***casualmente*** nelle 4 caselle.

$$\frac{262}{262 \times 3} = \frac{1}{3}$$

non è una stima della frazione di infartuati nella popolazione

$$\frac{172}{172 + 173} = 0.5$$

non è la prob. che a una donna venga l'infarto sapendo che fuma.

$\frac{\hat{p}_1}{\hat{p}_2}$ non è calcolabile

Odds ratio: caso-controllo

variabile risposta

Gruppo	IM	Controllo
	variabile esplicativa	
Fumatore	172	173
Non fumatore	90	346
	262	$\approx$**262**$\times$**2**

(Agresti. p. 26)

Ogni *caso* (262 donne di età<69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) accoppiato con 2 *controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) $\Rightarrow$ **marginali di colonna fissate.**

le 262x3 u.s. **non** si distribuiscono ***casualmente*** nelle 4 caselle.

$$\frac{262}{262 \times 3} = \frac{1}{3}$$

non è una stima della frazione di infartuati nella popolazione

$$\frac{172}{172 + 173} = 0.5$$

non è la prob. che a una donna venga l'infarto sapendo che fuma.

$\frac{\hat{p}_1}{\hat{p}_2}$ non è calcolabile

$$\boldsymbol{OR = \frac{172 \times 346}{90 \times 173} = 3.82}$$

Odds ratio: caso-controllo

variabile risposta

Gruppo	IM	Controllo
	variabile esplicativa	
Fumatore	172	173
Non fumatore	90	346
	262	≈**262**×**2**

(Agresti. p. 26)

Ogni *caso* (262 donne di età<69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) accoppiato con 2 *controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate.**

le 262x3 u.s. **non** si distribuiscono ***casualmente*** nelle 4 caselle.

$$\frac{262}{262 \times 3} = \frac{1}{3}$$

non è una stima della frazione di infartuati nella popolazione

$$\frac{172}{172 + 173} = 0.5$$

non è la prob. che a una donna venga l'infarto sapendo che fuma.

$\frac{\hat{p}_1}{\hat{p}_2}$ non è calcolabile

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{\frac{172}{262}/\frac{90}{262}}{\frac{173}{519}/\frac{346}{519}} = \frac{172/90}{173/346} = \boldsymbol{OR} = \frac{\mathbf{172 \times 346}}{\mathbf{90 \times 173}} = \mathbf{3.82}$$

Odds ratio: caso-controllo

variabile risposta

variabile esplicativa

Gruppo	IM	Controllo
Fumatore	172	173
Non fumatore	90	346
	262	≈**262**×**2**

(Agresti. p. 26)

Ogni *caso* (donne di età<69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) accoppiato con 2 *controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate** (campionamenti binomiali *dipendenti*).

siccome la probabilità che donne non anziane abbiano IM dovrebbe essere piccola indipendentemente dallo stato di fumatore

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = OR = 3.82 \approx \frac{\hat{p}_1}{\hat{p}_2}$$

Odds ratio: caso-controllo

variabile risposta

variabile esplicativa

Gruppo	**IM**	**Controllo**
Fumatore	172	173
Non fumatore	90	346
	262	$\approx$**262**$\times$**2**

(Agresti. p. 26)

Ogni *caso* (donne di età<69 ricoverate presso unità di cura coronarica in Nord Italia con IM acuto tra il 1983-1988) accoppiato con 2 *controlli* (pazienti ricoverati agli stessi ospedali per altri disordini acuti) ⇒ **marginali di colonna fissate** (campionamenti binomiali *dipendenti*).

siccome la probabilità che donne non anziane abbiano IM dovrebbe essere piccola indipendentemente dallo stato di fumatore

$$\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = OR = 3.82 \approx \frac{\hat{p}_1}{\hat{p}_2}$$

si stima che donne che sono state/sono fumatrici abbiano probabilità di IM pari a circa 4 volte la probabilità di quelle che non han mai fumato.

Facciamo un salto in R

	Infarto miocardico	
Gruppo	**Sì**	**No**
Placebo	189	10845
Aspirina	104	10933

$\hat{p}_1 - \hat{p}_2 = 0.00770602;$

> prop.test()

2-sample test for equality of proportions with continuity correction

data: z
X-squared = 24.4291, df = 1, p-value = 7.71e-07
alternative hypothesis: two.sided
95 percent confidence interval: **di** $\boldsymbol{p_1 - p_2}$
0.004597134 0.010814914
sample estimates:
prop 1 prop 2
0.01712887 0.00942285

Script4b.R

Le tabelle di contingenza

		Y				
		Lieve o assente	**Moderato**	**Grave**	**Molto Grave**	TOT
X	x_1	27	20	9	6	62
	x_2	66	63	53	44	226
	x_3	34	61	50	73	218
	x_4	25	53	43	109	230
TOT		152	197	155	232	736

COSA SIGNIFICA INDIPENDENZA??

Le tabelle di contingenza

Vogliamo vedere se la credenza nell'aldilà è ugualmente diffusa tra uomini e donne. Selezioniamo **un campione casuale di n soggetti** che classifichiamo nelle 2×2 caselle, rispetto alle due variabili:

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	1127

$$\frac{509}{625} = 0.814$$

$$\frac{398}{502} = 0.793$$

Se la proporzione di credenti è la stessa nei due generi, allora possia

probabilità condizionate di essere credente sono le stesse

l'unico modo per assicurarsi prob. condizionate tutte uguali è quello corrispondente all'**indipendenza delle variabili**

Ripassino veloce

Gli eventi A e B sono **indipendenti se**

$$P(A \cap B) = P(A)P(B)$$

Le variabili casuali X e Y, discrete o categoriche, sono **indipendenti se**

$$\boxed{P(X = a \ \cap Y = b) = \ P(X = a)P(Y = b)}$$

per qualunque coppia di valori (a, b)

Le tabelle di contingenza

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	n =1127

ferme restando le distribuzioni marginali, quale valore dovrei avere al posto di 509 nell'ipotesi di indipendenza?

Consideriamo la prima cella: un individuo scelto a caso nel campione può finire (1) o meno (0) nella cella: $(X_1, \ldots, X_n)$ i.i.d. $X_i \sim b(P(\text{Sì} \cap F))$.
Il numero atteso di individui nella prima cella è

$$\boldsymbol{n^*_{11}} = \boldsymbol{E(S_{11})} = n \times P(\text{Sì} \cap F) = \boldsymbol{n} \times [\boldsymbol{P(S\text{ì})} \times \boldsymbol{P(F)}] = 1127 \times \frac{907}{1127} \times \frac{625}{1127} = 502.99$$

$S_{11} \sim Binom(n, \boldsymbol{P(S\text{ì})} \times \boldsymbol{P(F)})$ **se c'è indipendenza**.

Le tabelle di contingenza

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	n =1127

ferme restando le distribuzioni marginali, quale valore dovrei avere al posto di 509 nell'ipotesi di indipendenza?

$$\boldsymbol{n^*}_{\mathbf{11}} = \boldsymbol{E}(\boldsymbol{S}_{\mathbf{11}}) = n \times P(\text{Sì} \cap F) = \boldsymbol{n} \times [\boldsymbol{P}(\boldsymbol{S}\text{ì}) \times \boldsymbol{P}(\boldsymbol{F})] = 1127 \times \frac{907}{1127} \times \frac{625}{1127} = 502.99$$

$\boldsymbol{n^*}_{\boldsymbol{ij}}$ frequenze *attese* o *teoriche* nell'ipotesi di **indipendenza**

$$\boldsymbol{n^*}_{\boldsymbol{ij}} = \boldsymbol{n} \times \frac{\boldsymbol{n_{i\cdot}}}{\boldsymbol{n}} \times \frac{\boldsymbol{n_{\cdot j}}}{\boldsymbol{n}} \Leftrightarrow n^*_{ij} = \frac{n_{i\cdot} \times n_{\cdot j}}{n}$$

L'indice del chi-quadrato

frequenze **osservate**

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	n =1127

frequenze **attese** sotto **indip.**

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	502.99	122.01	625
Maschio	404.00	98.00	502
	907	220	$\boldsymbol{n}$=1127

L'indice del chi-quadrato

frequenze **osservate**

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	n =1127

frequenze **attese** sotto **indip.**

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	502.99	122.01	625
Maschio	404.00	98.00	502
	907	220	$\boldsymbol{n}$=1127

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(n_{ij} - n^*{}_{ij}\right)^2}{n^*{}_{ij}}$$

L'indice del chi-quadrato

frequenze **osservate**

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	509	116	625
Maschio	398	104	502
	907	220	n =1127

$\chi^2 = 0$ indica assenza totale di associazione, o **indipendenza**

frequenze **attese** sotto **indip.**

	Credenza nell'aldilà		
Sesso	**Sì**	**No**	
Femmina	502.99	122.01	625
Maschio	404.00	98.00	502
	907	220	$\boldsymbol{n}$=1127

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(n_{ij} - n^*{}_{ij}\right)^2}{n^*{}_{ij}}$$

L'indice del chi-quadrato

Perfetta interdipendenza tra le due variabili

	Credenza aldilà	
Sesso	**Sì**	**No**
Femmina	0	220
Maschio	907	0

Perfetta dipendenza della Credenza dal Gruppo.

	Credenza aldilà	
Gruppo	**Sì**	**No**
F adulta	0	220
M adulto	100	0
Bambino	807	0

L'indice del chi-quadrato

Indicazione di dipendenza tra le due variabili

	Tipo di vacanza		
N. figli	Estero	Mare	Montagna
0	120	0	30
1	20	0	80
≥ 2	0	100	10

L'indice del chi-quadrato

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(n_{ij} - n^*{}_{ij}\right)^2}{n^*{}_{ij}} \qquad \mathbf{0 \leq \chi^2 \leq n \times \min(r-1, c-1)}$$

$$\tilde{\chi}^2 = \frac{\chi^2}{n \times \min(r-1, c-1)} \qquad 0 \leq \tilde{\chi}^2 \leq 1$$

$\chi^2 = \tilde{\chi}^2 = 0 \iff$ **mancanza di associazione**
indipendenza

$\chi^2 = n \times min(r-1, c-1) \iff \tilde{\chi}^2 = 1 \iff$ **massima associazione**

Il test del chi-quadrato

Sotto l'ipotesi di indipendenza tra le due variabili:

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-n^*{}_{ij}\right)^2}{n^*{}_{ij}} \sim \chi^2((r-1)(c-1))$$

Rifiuto H_0 a livello di significatività α se l'indice χ^2 è superiore al quantile $\chi^2{}_{(r-1)(c-1);\,1-\alpha}$

$\boxed{n^*{}_{ij} \geq 5}$

test asintotico,
non-parametrico!

Il test del chi-quadrato

Sotto l'ipotesi di indipendenza tra le due variabili:

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-n^*{}_{ij}\right)^2}{n^*{}_{ij}} \sim \chi^2((r-1)(c-1))$$

Rifiuto H_0 a livello di significatività α se l'indice χ^2 è superiore al quantile $\chi^2{}_{(r-1)(c-1);\,1-\alpha}$

$n^*{}_{ij} \geq 5$

test asintotico, non-parametrico!

Prendiamo i dati della nostra patologia e identifichiamo le coppie di variabili per cui è interessante testare l'indipendenza.

Il test del chi-quadrato con R

```
> chisq.test(x)

        Pearson's Chi-squared test with Yates' continuity correction

data:  x
X-squared = 0.42293, df = 1, p-value = 0.5155

Warning message:
In chisq.test(x) :
  L'approssimazione al Chi-quadrato potrebbe esse
```

"a logical indicating whether to apply continuity correction when computing the test statistic for 2 by 2 tables: one half is subtracted from all |O - E| differences; however. the correction will not be bigger than the differences themselves. No correction is done if simulate.p.value = TRUE".

Il test del chi-quadrato con R

```
> chisq.test(x)

        Pearson's Chi-squared test with Yates' continuity correction

data:  x
X-squared = 0.42293, df = 1, p-value = 0.5155

Warning message:
In chisq.test(x) :
  L'approssimazione al Chi-quadrato potrebbe esse
```

"a logical indicating whether to apply continuity correction when computing the test statistic for 2 by 2 tables: one half is subtracted from all |O - E| differences; however. the correction will not be bigger than the differences themselves. No correction is done if simulate.p.value = TRUE".

```
> chisq.test(x, correct=F)

 Pearson's Chi-squared test

data:  x
X-squared = 1.0234, df = 1, p-value = 0.3117

Warning message:
In chisq.test(x, correct = F) :
  L'approssimazione al Chi-quadrato potrebbe essere inesatta
```

Il test del chi-quadrato con R

```
> chisq.test(x)
        Pearson's Chi-squared test with Yates' continuity correction

data:  x
X-squared = 0.42293, df = 1, p-value = 0.5155

Warning message:
In chisq.test(x) :
  L'approssimazione al Chi-quadrato potrebbe esse
```

"a logical indicating whether to apply continuity correction when computing the test statistic for 2 by 2 tables: one half is subtracted from all |O - E| differences; however. the correction will not be bigger than the differences themselves. No correction is done if simulate.p.value = TRUE".

```
> chisq.test(x, correct=F)
  Pearson's Chi-squared test

data:  x
X-squared = 1.0234, df = 1, p-value = 0.3117
```

```
> chisq.test(x, simulate.p.value = TRUE)
  Pearson's Chi-squared test with simulated p-value (based on 2000
        replicates)

data:  x
X-squared = 1.0234, df = NA, p-value = 0.4523
```

Il test del chi-quadrato con R

```
> chisq.test(x)

        Pearson's Chi-squared test with Yates' continuity correction

data:  x
X-squared = 0.42293, df = 1, p-value = 0.5155

Warning message:
In chisq.test(x) :
  L'approssimazione al Chi-quadrato potrebbe esse
```

"a logical indicating whether to apply continuity correction when computing the test statistic for 2 by 2 tables: one half is subtracted from all |O - E| differences; however. the correction will not be bigger than the differences themselves. No correction is done if simulate.p.value = TRUE".

```
> chisq.test(x, correct=F)

 Pearson's Chi-squared test

data:  x
X-squared = 1.0234, df = 1, p-value = 0.3117
```

```
> fisher.test(x)
```

```
> chisq.test(x, simulate.p.value = TRUE)

  Pearson's Chi-squared test with simulated p-value (based on 2000
        replicates)

data:  x
X-squared = 1.0234, df = NA, p-value = 0.4523
```

Il test del chi-quadrato con R

La nostra patologia:

- associazione tra patologia (0/1) e genere/X2c

n_{ij}	Patologia	
Sesso	**Sì**	**No**
F		
M		

```
> fisher.test(x)
```

```
> chisq.test(x, simulate.p.value =TRUE,B=500000)
```

oppure:

- associazione tra gravità della patologia (0-3) e genere

n_{ij}	Patologia			
Sesso	**0**	**1**	**2**	**3**
F				
M				

Il test del chi-quadrato con R

$$\sum_{i=1,\dots,r}\sum_{j=1,\dots,c}\frac{(n_{ij}-n^*_{ij})^2}{n^*_{ij}}$$

Se si rifiuta l'ipotesi di indipendenza, è interessante capire il perchè.

$p \approx 0.02$	Gravità della patologia			
X2c	**0**	**1**	**2**	**3**
0	12 9.2	6 5.3	8 7.6	6 9.9
1	0 2.8	1 1.7	2 2.4	7 3.1

$$\frac{n_{ij}-n^*_{ij}}{\sqrt{n^*_{ij}}}$$

residui

> test1$residuals

```
[0]  0.9449112   0.2886751  0.1380131  -1.240716
[1] -1.6903085  -0.5163978 -0.2468854   2.219461
```

Il test del chi-quadrato con R

$$\sum_{i=1,\dots,r} \sum_{j=1,\dots,c} \frac{(n_{ij} - {n^*}_{ij})^2}{{n^*}_{ij}}$$

$$\frac{n_{ij} - {n^*}_{ij}}{\sqrt{{n^*}_{ij} \left(1 - \frac{n_{i+}}{n}\right)\left(1 - \frac{n_{+j}}{n}\right)}}$$

residui **standardizzati**
sotto H_0 sono asint. gauss. std.

Se si rifiuta l'ipotesi di indipendenza, è interessante capire il perchè.

$p \approx 0.02$	Gravità della patologia			
X2c	**0**	**1**	**2**	**3**
0	12 9.2	6 5.3	8 7.6	6 9.9
1	0 2.8	1 1.7	2 2.4	7 3.1

```
> test1$stdres
```

```
[0]   2.291288    0.6480741    0.324037    -3.06001
[1]  -2.291288   -0.6480741   -0.324037     3.06001
```

Il test del chi-quadrato con R

$$\sum_{i=1,\dots,r}\sum_{j=1,\dots,c}\frac{(n_{ij}-n^*{}_{ij})^2}{n^*{}_{ij}}$$

Se si rifiuta l'ipotesi di indipendenza, è interessante capire il perchè.

$$\frac{n_{ij}-n^*{}_{ij}}{\sqrt{n^*{}_{ij}\,\left(1-\frac{n_{i+}}{n}\right)\left(1-\frac{n_{+j}}{n}\right)}}$$

residui **standardizzati**
sotto H_0 sono asint. gauss. std.

$p \approx 0.02$	Gravità della patologia			
X2c	**0**	**1**	**2**	**3**
0	12 9.2	6 5.3	8 7.6	6 9.9
1	0 2.8	1 1.7	2 2.4	7 3.1

\> test1$**stdres**

[0]	**2.291288**	0.6480741	0.324037	**-3.06001**
[1]	**-2.291288**	-0.6480741	-0.324037	**3.06001**

Esempio

Obstructive sleep apnea (OSA) is a common condition in which there are intermittent partial (viz., hypopneas) and complete (viz., apnea) limitations in airflow, with associated hypoxia and sympathetic arousals, during sleep. Because polysomnography, the standard test for diagnosing OSA, is expensive and time-consuming, questionnaires have been developed to identify persons with OSA. The Berlin questionnaire (BQ) reliably identifies middle-aged and older persons in the community who are at high-risk for OSA.

Esempio

Obstructive sleep apnea (OSA) is a common condition in which there are intermittent partial (viz., hypopneas) and complete (viz., apnea) limitations in airflow, with associated hypoxia and sympathetic arousals, during sleep. Because polysomnography, the standard test for diagnosing OSA, is expensive and time-consuming, questionnaires have been developed to identify persons with OSA. The Berlin questionnaire (BQ) reliably identifies middle-aged and older persons in the community who are at high-risk for OSA.

Studio retrospettivo su una popolazione italiana:

	OSA	Non OSA	Marginale BQ
BQ=low risk	79	294	**373**
BQ=high risk	73	287	**360**
Marginale OSA	**152**	**581**	**733**

Esempio

Obstructive sleep apnea (OSA) is a common condition in which there are intermittent partial (viz., hypopneas) and complete (viz., apnea) limitations in airflow, with associated hypoxia and sympathetic arousals, during sleep. Because polysomnography, the standard test for diagnosing OSA, is expensive and time-consuming, questionnaires have been developed to identify persons with OSA. The Berlin questionnaire (BQ) reliably identifies middle-aged and older persons in the community who are at high-risk for OSA.

Studio retrospettivo su una popolazione italiana:

	OSA	Non OSA	Marginale BQ
BQ=low risk	79	294	**373**
BQ=high risk	73	287	**360**
Marginale OSA	**152**	**581**	**733**

$$\chi^2 = 0.044, p - value = 0.83!$$

Esempio

Obstructive sleep apnea (OSA) is a common condition in which there are intermitt neas) and complete (viz., apnea) limitations ed hypoxia and sympathetic arousals, d ysomnography, the standard test for diagnos nd time-consuming, questionnaires have been persons with OSA. The Berlin questionnaire (BQ) reliably ntifies middle-aged and older persons in the community who are at -risk for OSA.

INDIPENDENZA praticamente perfetta!!!

Studio retrospettivo su una po olazione italiana:

	OSA	Non OSA	Marginale BQ
BQ=low risk	79 / 77	294 / 296	**373**
BQ=high risk	73 / 75	287 / 285	**360**
Marginale OSA	**152**	**581**	**733**

$$\chi^2 = 0.044, p - value = 0.83!$$

Test d'indipendenza

variabile dipendente **ordinale**

variabile esplicativa **ordinale**

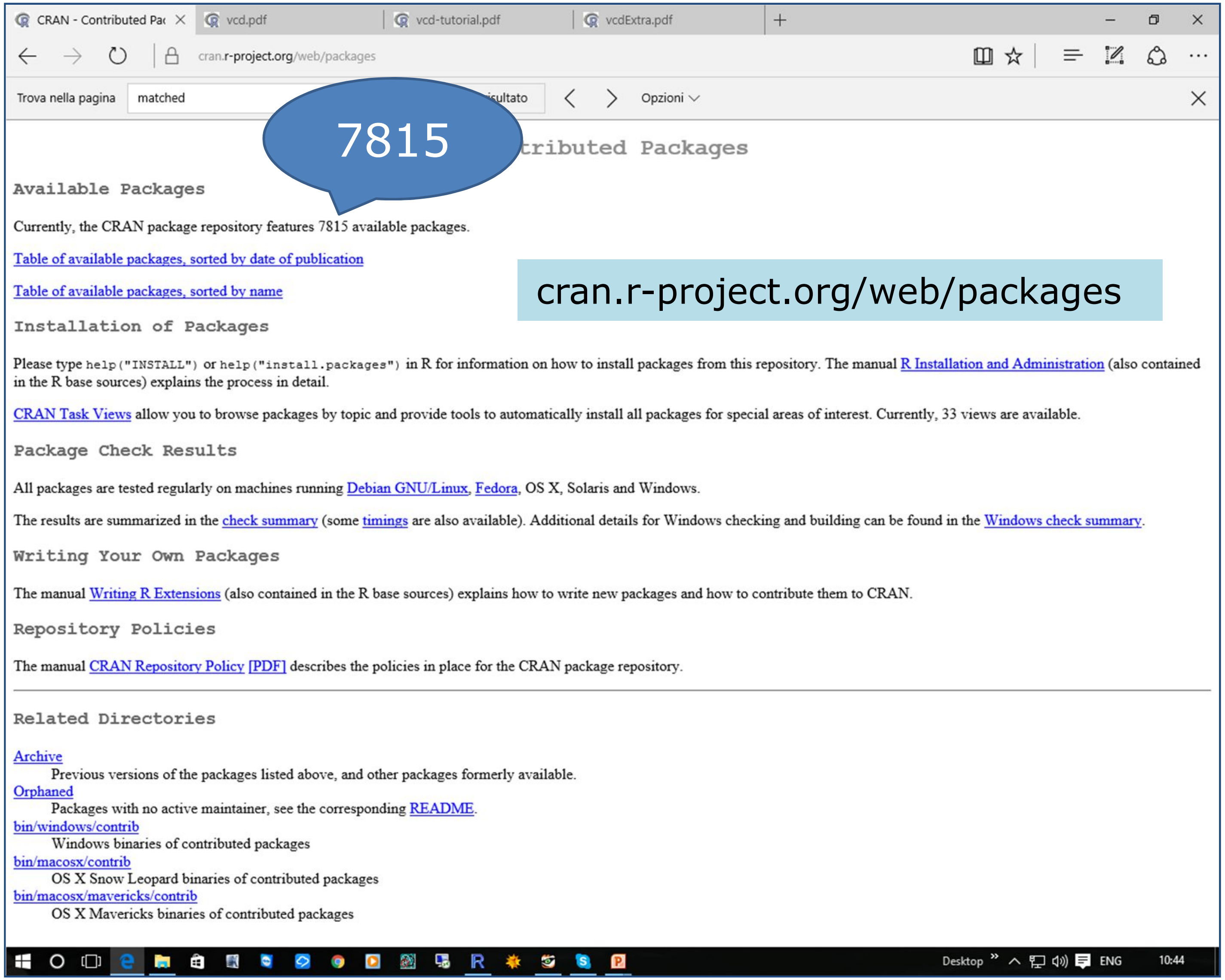
	OSA			
X1c	**Ind1** $\leq \gamma$	**Ind1** $> \gamma$	**Tot.**	**%**
15-25	3	7	10	0.70
25-30	5	10	15	0.67
30-35	1	4	5	0.80
> 35	9	14	23	0.61

Il test del χ^2 tratta tutte le variabili come nominali e non permette di vedere trend.

Si può fare un test basato su una correlazione tra gli *scores* dei dati: v. Agresti (p. 34). R-package "coin"

Digressione

Packages



...tributed Packages

Available Packages

Currently, the CRAN package repository features 7815 available packages.

Table of available packages, sorted by date of publication

Table of available packages, sorted by name

Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this repository. The manual R Installation and Administration (also contained in the R base sources) explains the process in detail.

CRAN Task Views allow you to browse packages by topic and provide tools to automatically install all packages for special areas of interest. Currently, 33 views are available.

Package Check Results

All packages are tested regularly on machines running Debian GNU/Linux, Fedora, OS X, Solaris and Windows.

The results are summarized in the check summary (some timings are also available). Additional details for Windows checking and building can be found in the Windows check summary.

Writing Your Own Packages

The manual Writing R Extensions (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

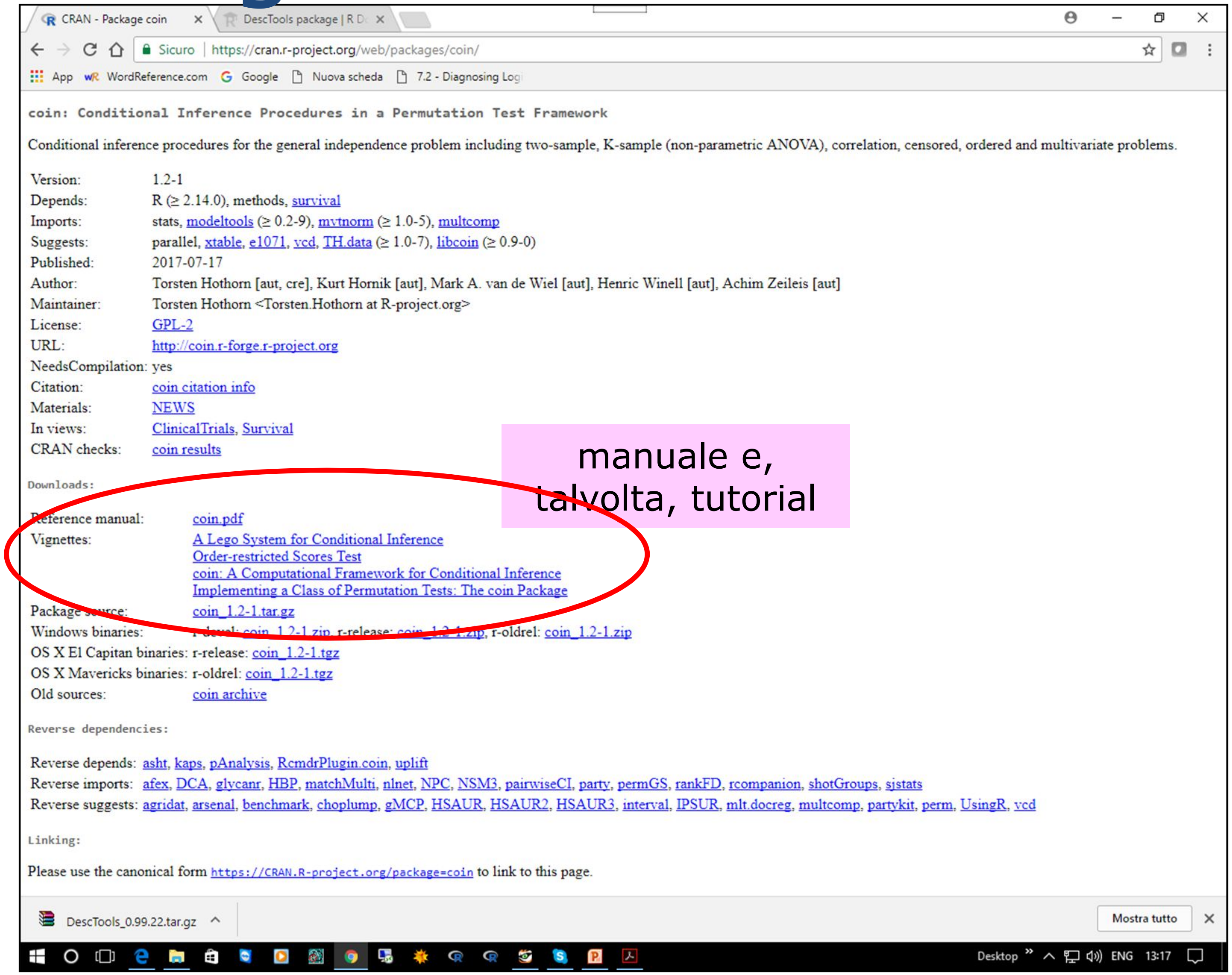
Repository Policies

The manual CRAN Repository Policy [PDF] describes the policies in place for the CRAN package repository.

Related Directories

Archive
Previous versions of the packages listed above, and other packages formerly available.
Orphaned
Packages with no active maintainer, see the corresponding README.
bin/windows/contrib
Windows binaries of contributed packages
bin/macosx/contrib
OS X Snow Leopard binaries of contributed packages
bin/macosx/mavericks/contrib
OS X Mavericks binaries of contributed packages

Digressione



coin: Conditional Inference Procedures in a Permutation Test Framework

Conditional inference procedures for the general independence problem including two-sample, K-sample (non-parametric ANOVA), correlation, censored, ordered and multivariate problems.

Version:	1.2-1
Depends:	R (≥ 2.14.0), methods, survival
Imports:	stats, modeltools (≥ 0.2-9), mvtnorm (≥ 1.0-5), multcomp
Suggests:	parallel, xtable, e1071, vcd, TH.data (≥ 1.0-7), libcoin (≥ 0.9-0)
Published:	2017-07-17
Author:	Torsten Hothorn [aut, cre], Kurt Hornik [aut], Mark A. van de Wiel [aut], Henric Winell [aut], Achim Zeileis [aut]
Maintainer:	Torsten Hothorn <Torsten.Hothorn at R-project.org>
License:	GPL-2
URL:	http://coin.r-forge.r-project.org
NeedsCompilation:	yes
Citation:	coin citation info
Materials:	NEWS
In views:	ClinicalTrials, Survival
CRAN checks:	coin results

Downloads:

Reference manual:	coin.pdf
Vignettes:	A Lego System for Conditional Inference Order-restricted Scores Test coin: A Computational Framework for Conditional Inference Implementing a Class of Permutation Tests: The coin Package
Package source:	coin_1.2-1.tar.gz
Windows binaries:	r-devel: coin_1.2-1.zip, r-release: coin_1.2-1.zip, r-oldrel: coin_1.2-1.zip
OS X El Capitan binaries:	r-release: coin_1.2-1.tgz
OS X Mavericks binaries:	r-oldrel: coin_1.2-1.tgz
Old sources:	coin archive

Reverse dependencies:

Reverse depends:	asht, kaps, pAnalysis, RcmdrPlugin.coin, uplift
Reverse imports:	afex, DCA, glycanr, HBP, matchMulti, nlnet, NPC, NSM3, pairwiseCI, party, permGS, rankFD, rcompanion, shotGroups, sjstats
Reverse suggests:	agridat, arsenal, benchmark, choplump, gMCP, HSAUR, HSAUR2, HSAUR3, interval, IPSUR, mlt.docreg, multcomp, partykit, perm, UsingR, vcd

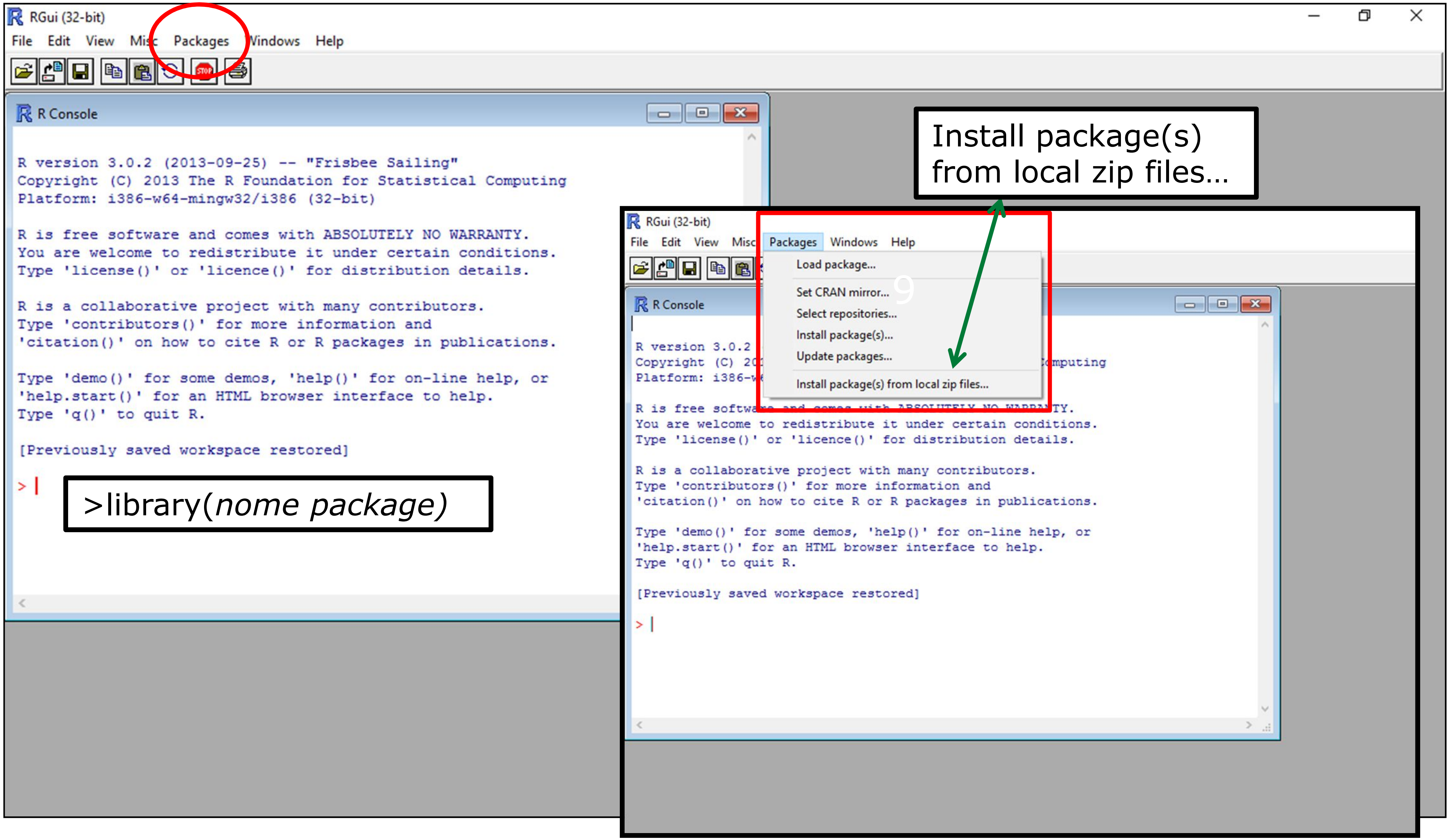
Linking:

Please use the canonical form https://CRAN.R-project.org/package=coin to link to this page.

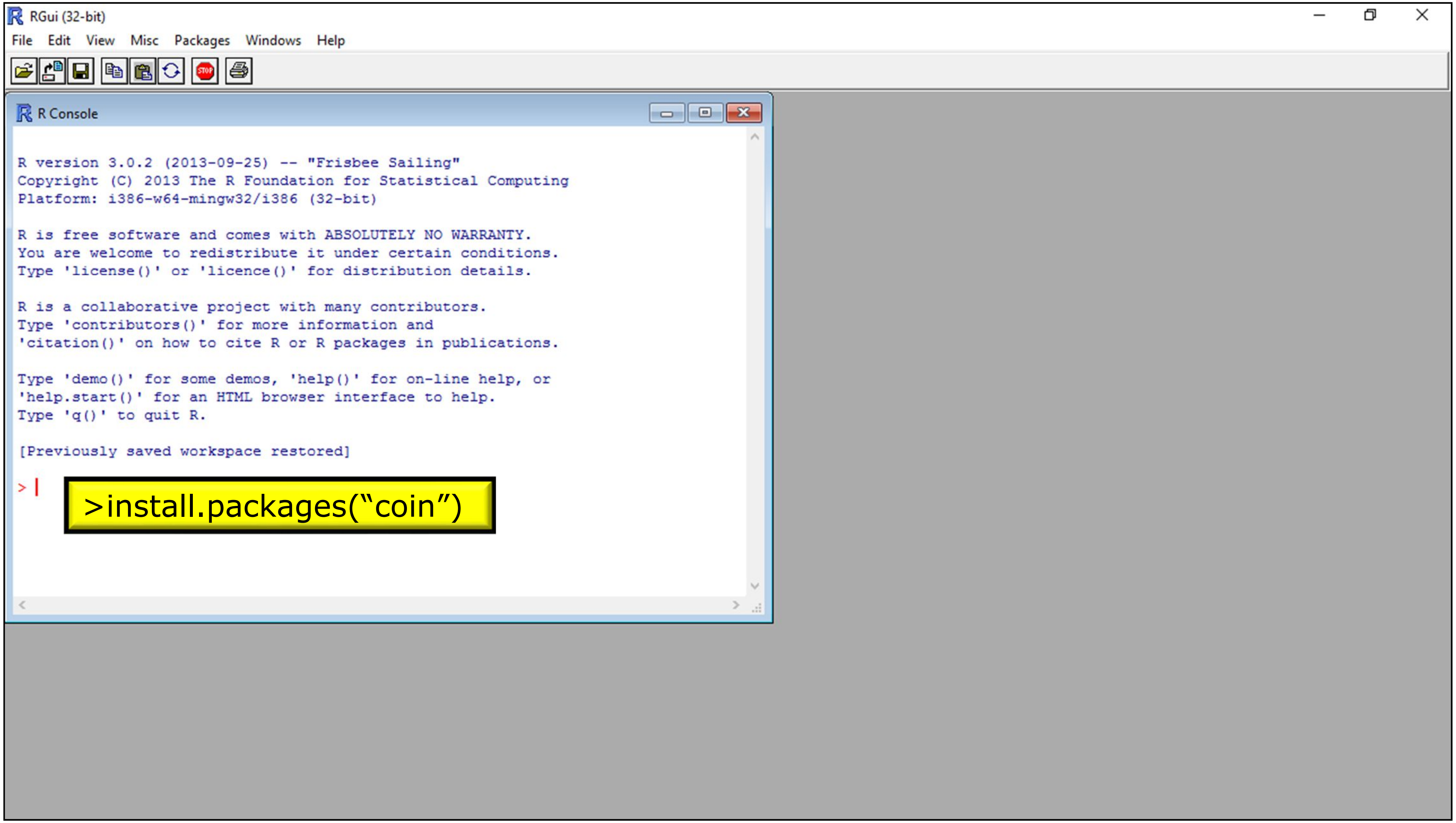
manuale e, talvolta, tutorial

https://cran.r-project.org/web/packages/coin/

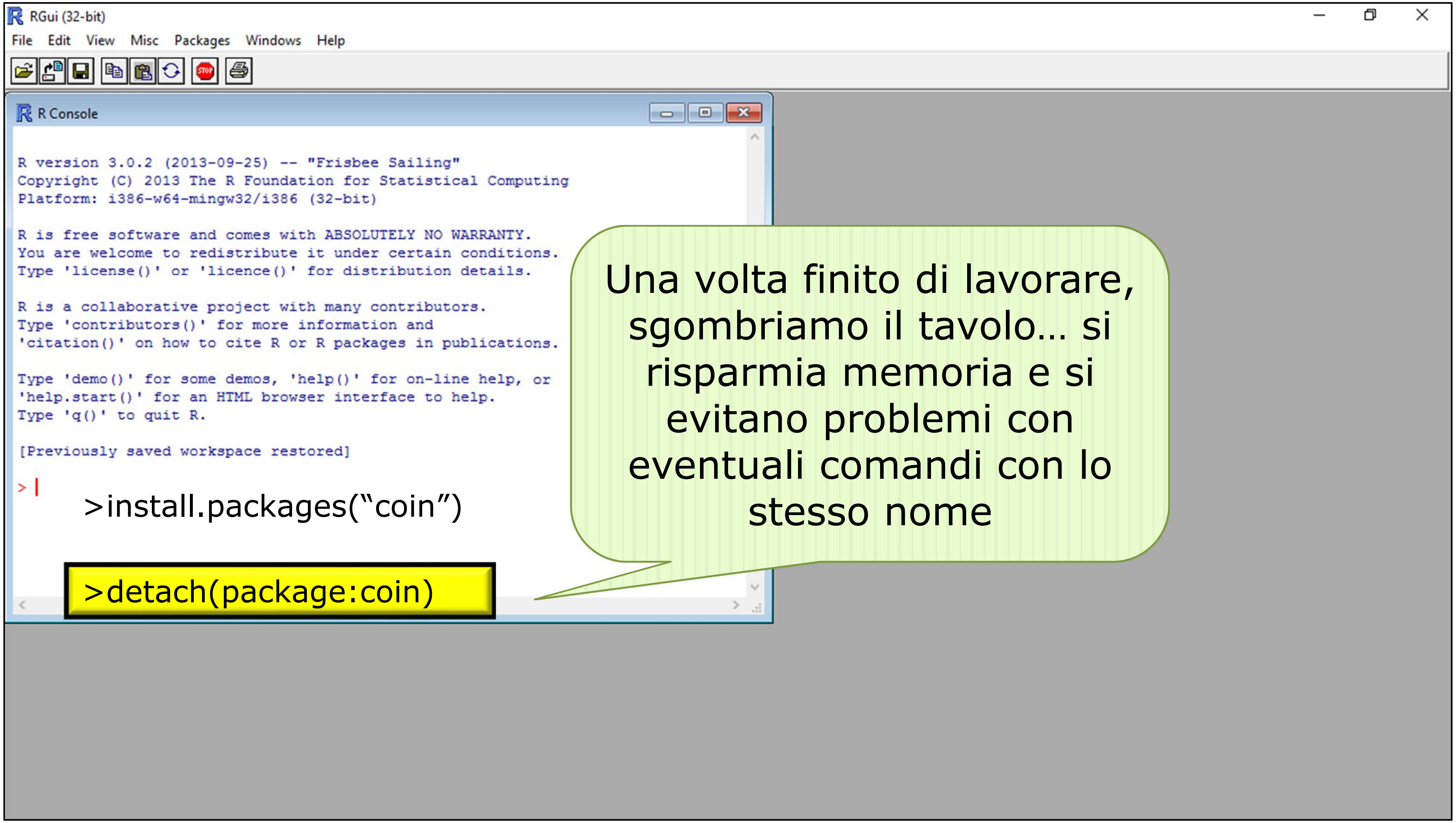
Digressione



Digressione



Digressione



>install.packages("coin")

>detach(package:coin)

Una volta finito di lavorare, sgombriamo il tavolo… si risparmia memoria e si evitano problemi con eventuali comandi con lo stesso nome

Test di buon adattamento

Esempio classico: il dado è equilibrato?

Esito	1	2	3	4	5	6
Fr.	**20**	**30**	**20**	**25**	**15**	**10**
F.r. oss.	**20/120**	**30/120**	**20/120**	**25/120**	**15/120**	**10/120**

$n = 120$

LASCIARLO DI COMPITO A FAVORE DELL'ESEMPIO SEGUENTE, POISSON.

Test di buon adattamento

Esempio classico: il dado è equilibrato?

Esito	1	2	3	4	5	6	
Fr.	**20**	**30**	**20**	**25**	**15**	**10**	$n = 120$
F.r. oss.	**20/120**	**30/120**	**20/120**	**25/120**	**15/120**	**10/120**	
F.r. att.	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	
Fr. att.	**20**	**20**	**20**	**20**	**20**	**20**	$e_j = np_j$

Test di buon adattamento

Esempio classico: il dado è equilibrato?

Esito	1	2	3	4	5	6
Fr.	**20**	**30**	**20**	**25**	**15**	**10**
F.r. oss.	**20/120**	**30/120**	**20/120**	**25/120**	**15/120**	**10/120**
F.r. att.	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**
Fr. att.	**20**	**20**	**20**	**20**	**20**	**20**

$n = 120$

$e_j = np_j$

$\sum_{j=1,\dots,c} \frac{(n_j - e_j)^2}{e_j} \sim \chi^2(c-1),$

$12.5 \quad > \quad 11.0705$

rifiutiamo l'ipotesi che il dado sia equilibrato al livello del 5%

Test di buon adattamento

Esempio classico: il dado è equilibrato?

Esito	1	2	3	4	5	6
Fr.	**20**	**30**	**20**	**25**	**15**	**10**
F.r. oss.	**20/120**	**30/120**	**20/120**	**25/120**	**15/120**	**10/120**
F.r. att.	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**
Fr. att.	**20**	**20**	**20**	**20**	**20**	**20**

$n = 120$

$e_j = np_j$

$$\sum_{j=1,\dots,c} \frac{(n_j - e_j)^2}{e_j} \sim \chi^2(c-1),$$

$$12.5 \quad > \quad 11.0705$$

rifiutiamo l'ipotesi che il dado sia equilibrato al livello del 5%

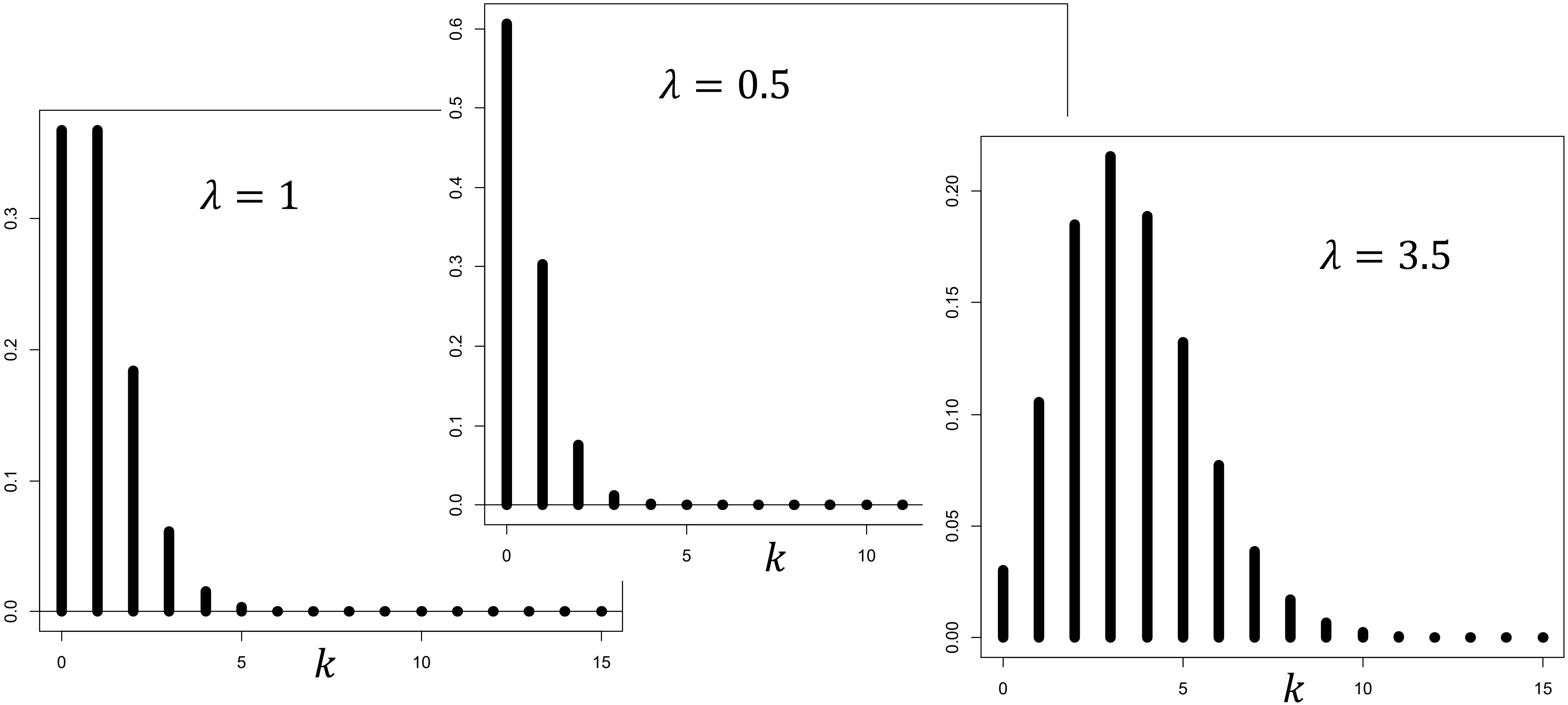
```
> chisq.test(c(20,30,20,25,15,10))

        Chi-squared test for given probabilities

data:  c(20, 30, 20, 25, 15, 10)
X-squared = 12.5, df = 5, p-value = 0.02854
```

Script4b.R

Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$$X : \Omega \rightarrow \{0,1,2,\dots\} \qquad P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad \lambda > 0$$

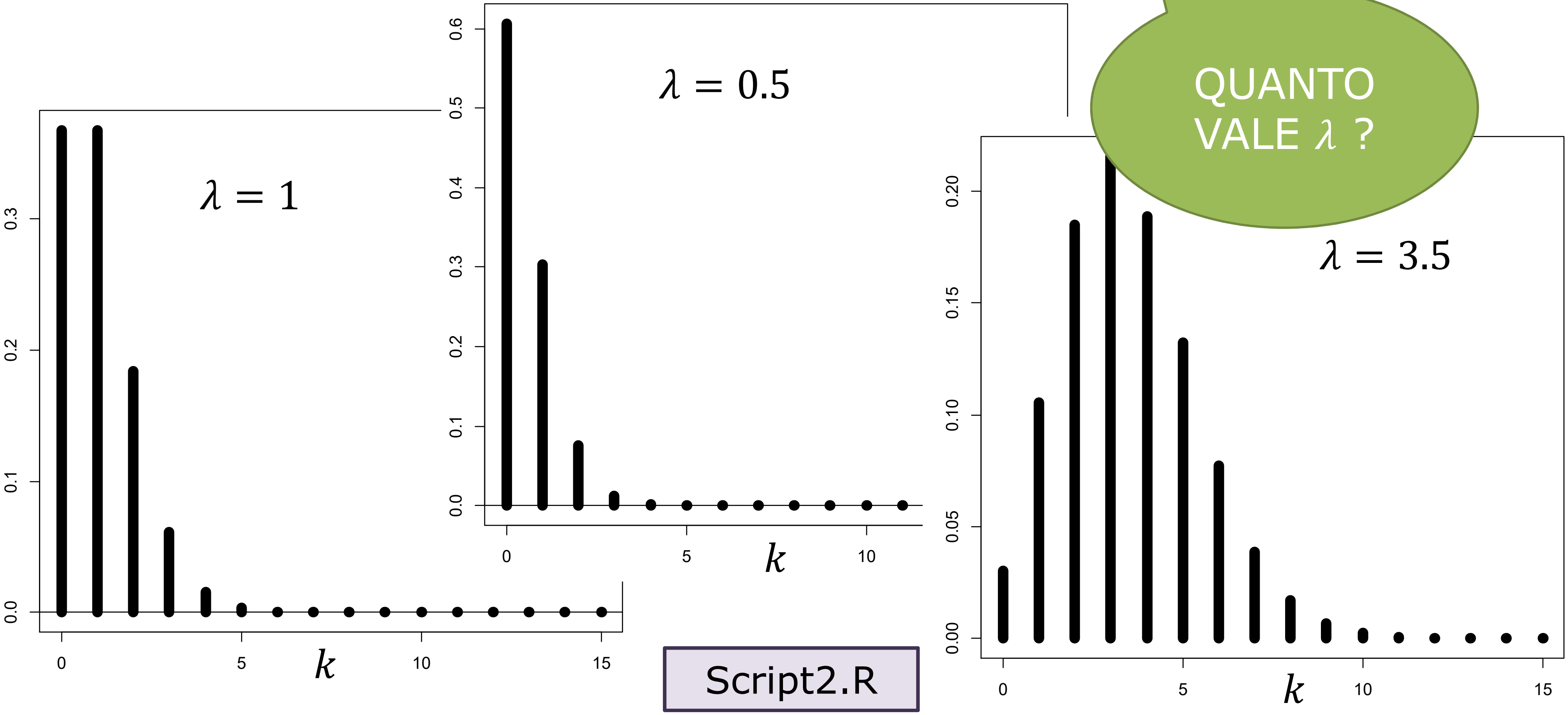


Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$$X : \Omega \to \{0,1,2,\dots\} \qquad \boldsymbol{P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}} \qquad \boldsymbol{\lambda > 0}$$

QUANTO VALE λ ?



Script2.R

Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$X:\ \Omega \to \{0,1,2,\dots\}$ $\quad$ $\boldsymbol{P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad \lambda>0}$

$E(X)=Var(X)=\lambda$ $\quad$ $\hat{\lambda}=\bar{x}_n$

Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$$X:\ \Omega \to \{0,1,2,\dots\} \qquad P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \qquad \lambda>0$$

$$\boxed{E(X)=Var(X)=\lambda} \qquad \boxed{\hat{\lambda}=\bar{x}_n}$$

x_i	n_i
0	2
1	11
2	4
3	8
4	4
7	1
	$n=30$

$$\bar{x}_n=\frac{0\times 2+1\times 11+\cdots+7\times 1}{30}=2.2$$

Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$$X : \Omega \to \{0,1,2,\dots\} \qquad P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad \lambda > 0$$

$$\boxed{E(X) = Var(X) = \lambda} \qquad \boxed{\hat{\lambda} = \bar{x}_n}$$

x_i	n_i	p_i	$e_i = np_i$
0	2	0.111	3.3
1	11	0.244	7.3
2	4	0.268	8.0
3	8	0.197	5.9
4	4	0.108	3.2
7	1	0.005	0.2
	$n = 30$		

$$\bar{x}_n = \frac{0 \times 2 + 1 \times 11 + \cdots + 7 \times 1}{30} = 2.2$$

Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$$X : \Omega \rightarrow \{0,1,2,\dots\} \qquad P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \lambda > 0$$

$$\boxed{E(X) = Var(X) = \lambda} \qquad \boxed{\hat{\lambda} = \bar{x}_n}$$

x_i	n_i	p_i	$e_i = np_i$
0	2	0.111	3.3
1	11	0.244	7.3
2	4	0.268	8.0
3	8	0.197	5.9
4	4	0.108	3.2
≥ 5	1	0.072	2.2
$n = 30$			

$$\bar{x}_n = \frac{0 \times 2 + 1 \times 11 + \cdots + 7 \times 1}{30} = 2.2$$

Ci sono tutti i valori possibili 5,6,8,9,… che non ho osservato!

Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$$X : \Omega \to \{0,1,2,\dots\} \qquad P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \lambda > 0$$

$$\boxed{E(X) = Var(X) = \lambda} \qquad \boxed{\hat{\lambda} = \bar{x}_n}$$

x_i	n_i	p_i	$e_i = np_i$
0	2	0.111	3.3
1	11	0.244	7.3
2	4	0.268	8.0
3	8	0.197	5.9
4	4	0.108	3.2
≥ 5	1	0.072	2.2
	$n = 30$		

$$\bar{x}_n = \frac{0 \times 2 + 1 \times 11 + \cdots + 7 \times 1}{30} = 2.2$$

$$\boxed{\sum_{j=1,\dots,c} \frac{(n_j - e_j)^2}{e_j} \sim \chi^2(c - 1 - 1),}$$

Test di buon adattamento

I dati seguono una distribuzione di Poisson?

$$X : \Omega \to \{0,1,2,\dots\} \qquad P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \lambda > 0$$

$$\boxed{E(X) = Var(X) = \lambda} \qquad \boxed{\hat{\lambda} = \bar{x}_n}$$

x_i	n_i	p_i	$e_i = np_i$
0	2	0.111	3.3
1	11	0.244	7.3
2	4	0.268	8.0
3	8	0.197	5.9
4	4	0.108	3.2
≥ 5	1	0.072	2.2
	$n = 30$		

$$\bar{x}_n = \frac{0 \times 2 + 1 \times 11 + \cdots + 7 \times 1}{30} = 2.2$$

$$\boxed{\sum_{j=1,\dots,c} \frac{(n_j - e_j)^2}{e_j} \sim \chi^2(c - 1 - 1),}$$

5.989

9.488

qchisq(0.95,4)

Non possiamo rifiutare l'ipotesi $Poisson$ al livello del 5%, e la stima del parametro è 2.2

Dal nostro *test*

3. Il *p-value* (*p*-valore) è

Un test	Una probabilità	NON SAPREI

Dal nostro *test*

3. Il *p-value* (*p*-valore) è

Un test	Una probabilità	NON SAPREI