



UNIVERSITA' DEGLI STUDI DI VERONA

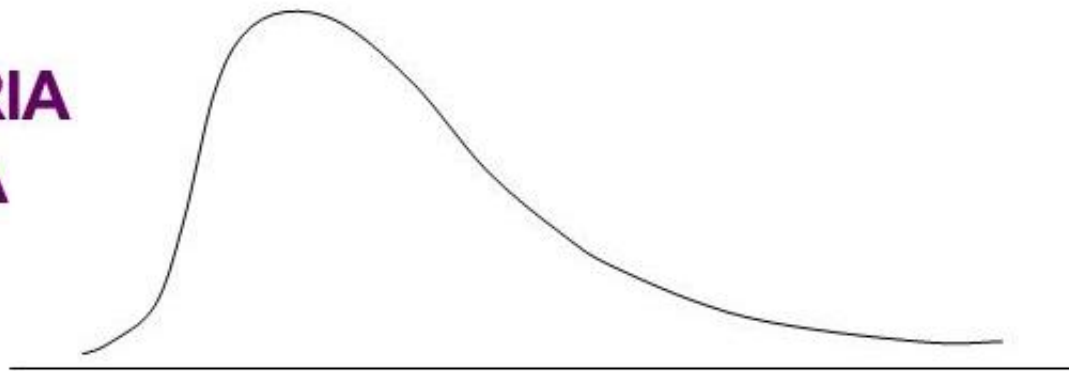
LABORATORIO DI PROBABILITA' E STATISTICA

Docente: Bruno Gobbi

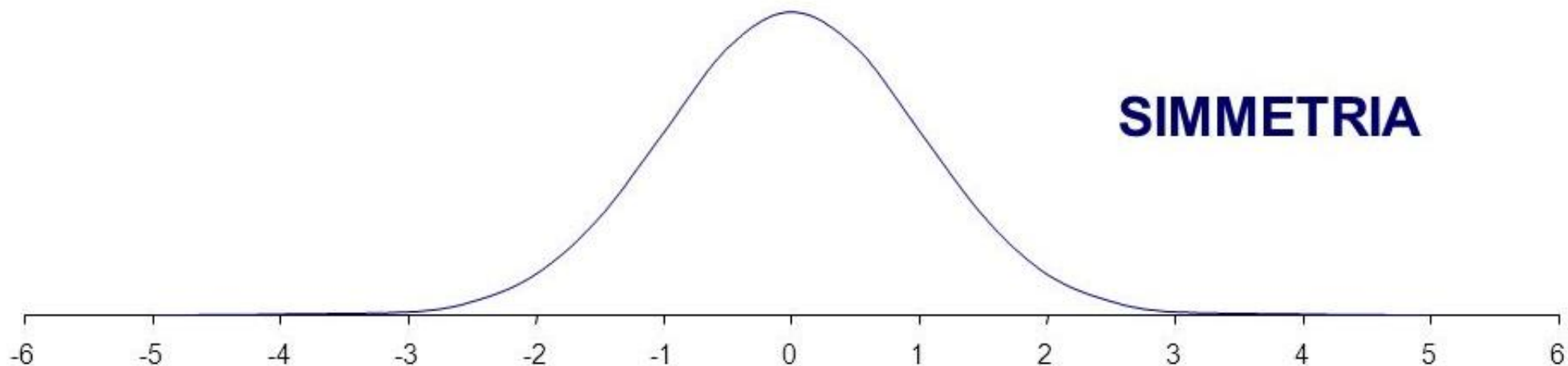
2 - SIMMETRIA, APPIATTIMENTO E
MISURA DELLA CONNESSIONE NELLE
TABELLE A DOPPIA ENTRATA

INDICI DI SIMMETRIA

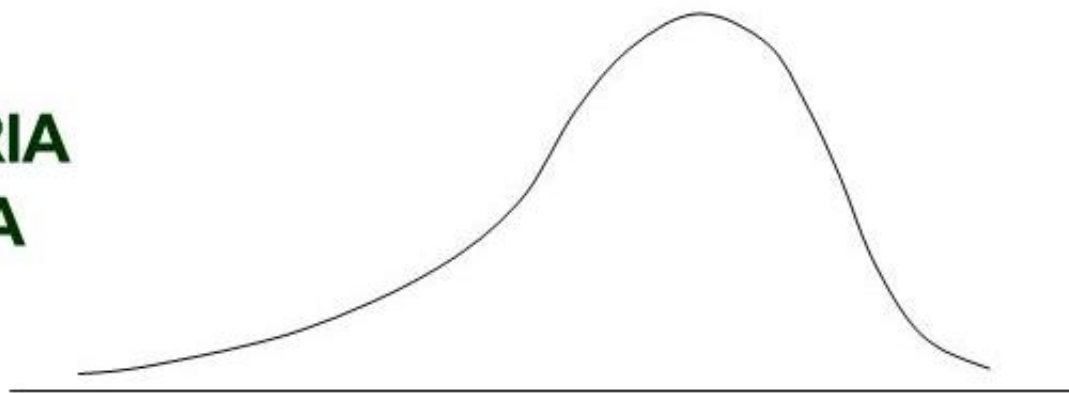
**ASIMMETRIA
POSITIVA**



SIMMETRIA



**ASIMMETRIA
NEGATIVA**



INDICE DI SIMMETRIA γ (gamma) DI FISHER

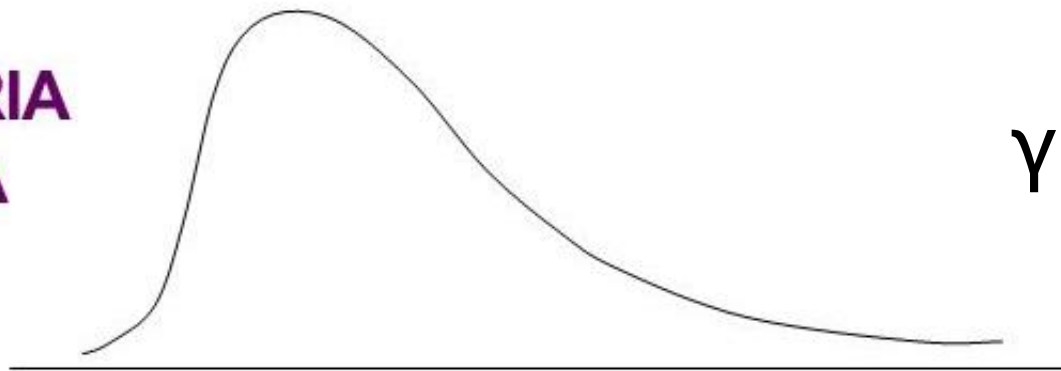
$$\gamma = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3$$

Se $\gamma = 0 \rightarrow$ allora la distribuzione è simmetrica

Se $\gamma < 0 \rightarrow$ allora la distribuzione è asimmetrica negativa

Se $\gamma > 0 \rightarrow$ allora la distribuzione è asimmetrica positiva

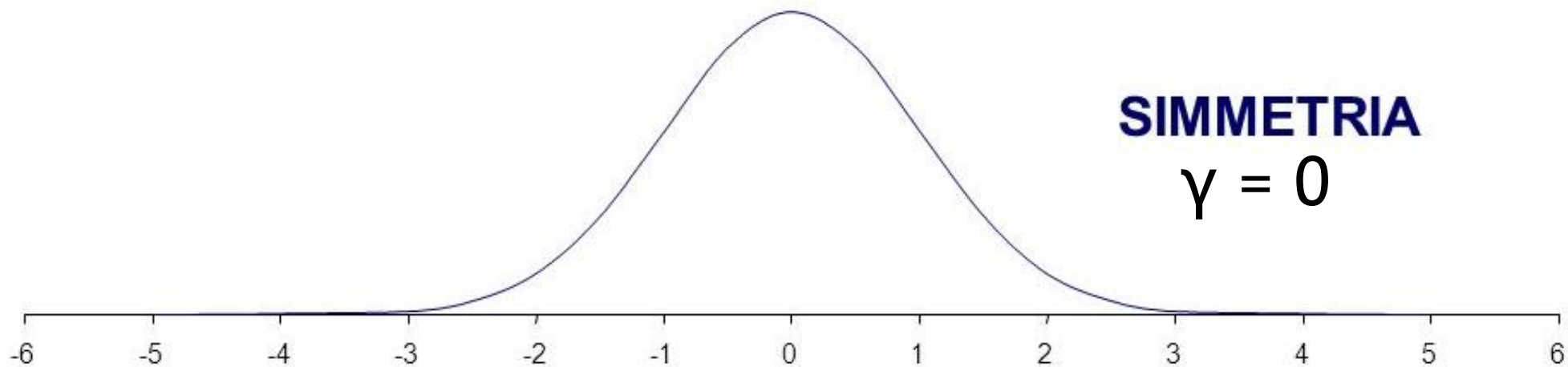
**ASIMMETRIA
POSITIVA**



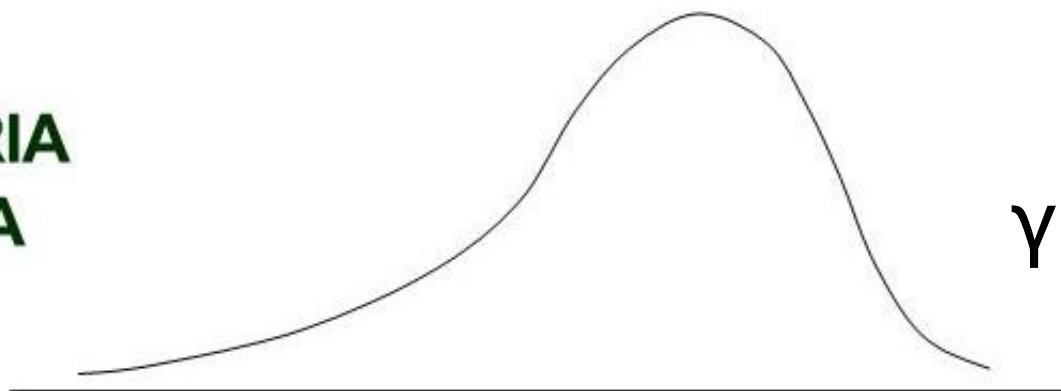
$$\gamma > 0$$

SIMMETRIA

$$\gamma = 0$$



**ASIMMETRIA
NEGATIVA**



$$\gamma < 0$$

SIMMETRIA (O SKEWNESS) IN R

In R esistono diversi pacchetti aggiuntivi che aiutano a calcolare la simmetria di una distribuzione.

ES.

- moment
- e1071
- fUtilities

CREAZIONE DI UNA FUNZIONE PER GAMMA

$$\gamma = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3$$

```
gamma = function(x) {  
  m3 = mean((x-mean(x))^3)  
  skew = m3 / (sd(x)^3)  
  skew  
}
```

{ = AltGr + 7

} = AltGr + 0

NO tastiera numerica

SIMMETRIA (O SKEWNESS) IN R

ES.

$x = c(0, 1, 1, 2, 2, 3, 4, 5)$

Valutare la simmetria di tale distribuzione.

SIMMETRIA (O SKEWNESS) IN R

ES.

$x = c(0, 1, 1, 2, 2, 3, 4, 5)$

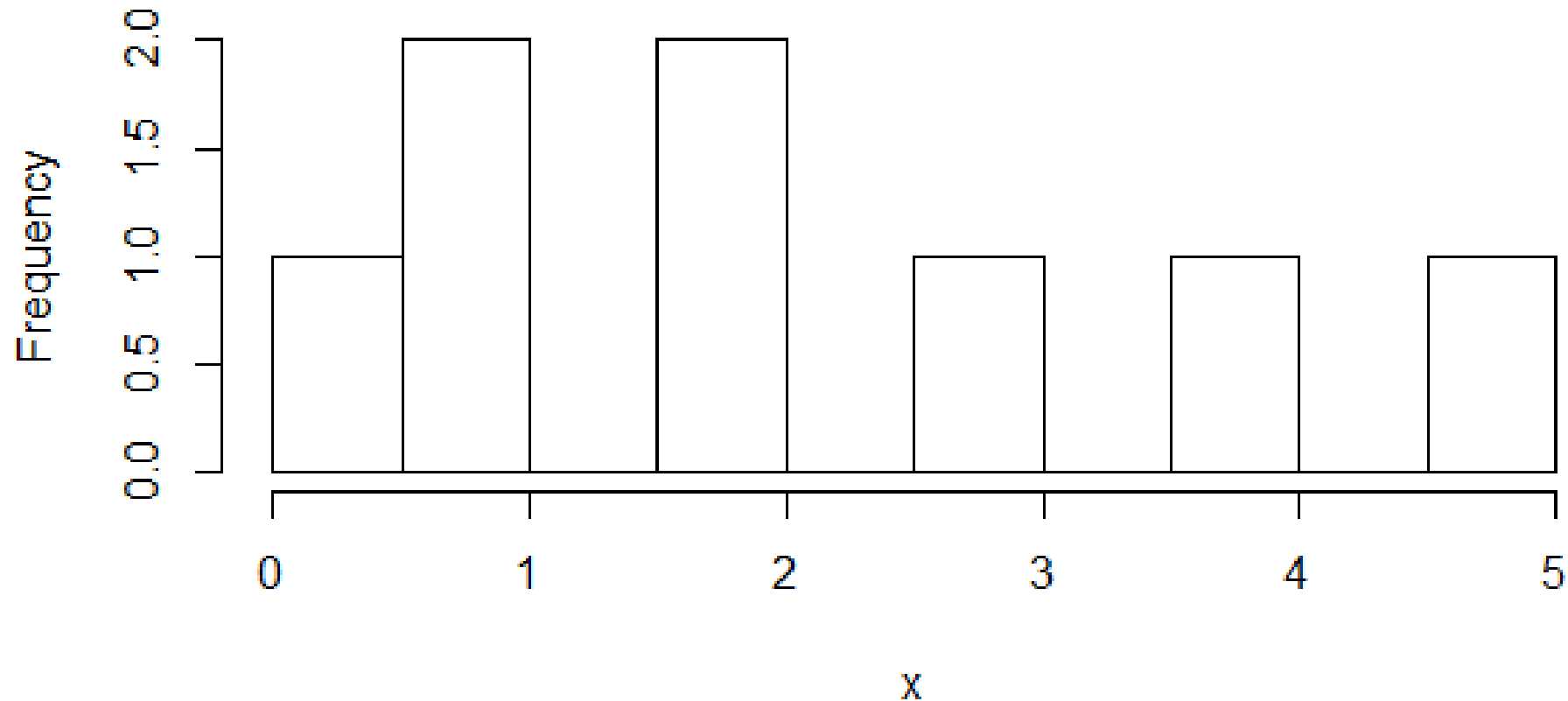
$\text{gamma}(x) = 0.3024528$

C'è asimmetria positiva, la distribuzione presenta una coda più lunga a destra.

VERIFICA GRAFICO SIMMETRIA

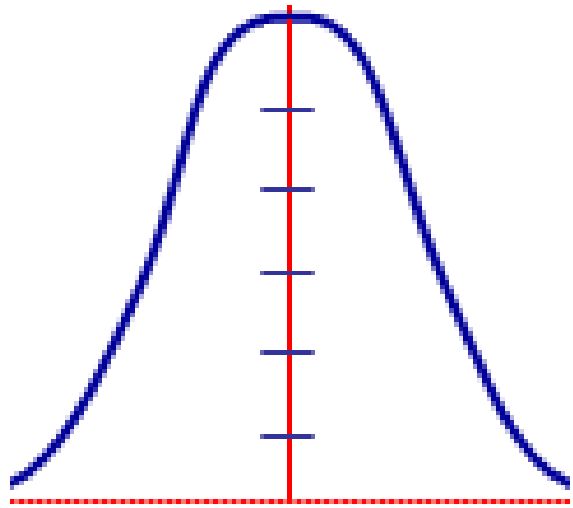
```
hist(x, freq=TRUE, breaks=length(x))
```

Histogram of x

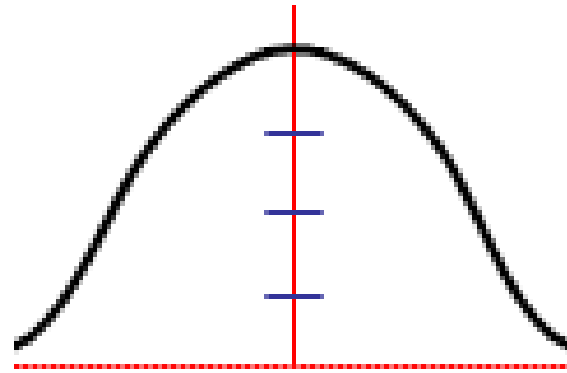


INDICI DI APPIATTIMENTO

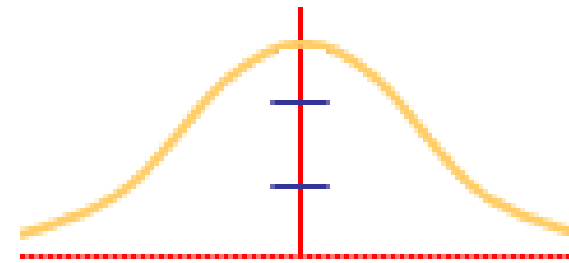
INDICI DI APPIATTIMENTO (CURTOSI)



Leptocurtica



Mesocurtica



Platicurtica

INDICE DI CURTOSI β (beta) DI PEARSON

$$\beta = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4$$

Se $\beta = 3 \rightarrow$ allora la distribuzione è MESOCURTICA

Se $\beta < 3 \rightarrow$ allora la distribuzione è PLATICURTICA

Se $\beta > 3 \rightarrow$ allora la distribuzione è LEPTOCURTICA

INDICE DI CURTOSI γ_2 (gamma2) DI FISHER

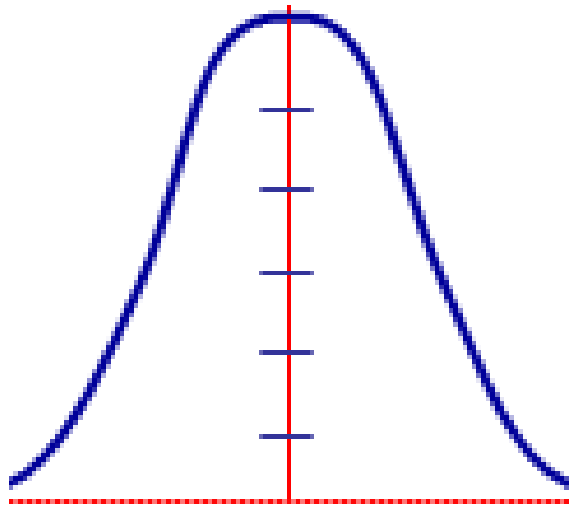
$$\gamma_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3$$

Se $\gamma_2 = 0 \rightarrow$ allora la distribuzione è MESOCURTICA

Se $\gamma_2 < 0 \rightarrow$ allora la distribuzione è PLATICURTICA

Se $\gamma_2 > 0 \rightarrow$ allora la distribuzione è LEPTOCURTICA

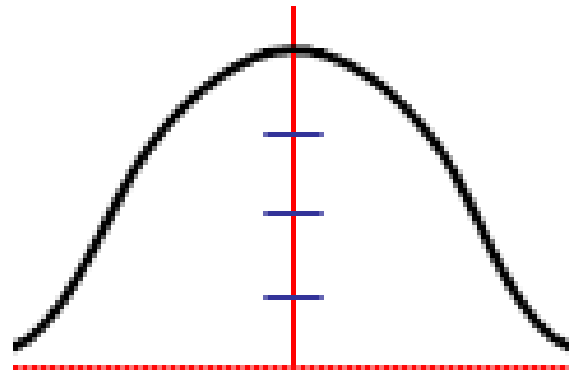
INDICI DI APPIATTIMENTO (CURTOSI)



Leptocurtica

$$\beta > 3$$

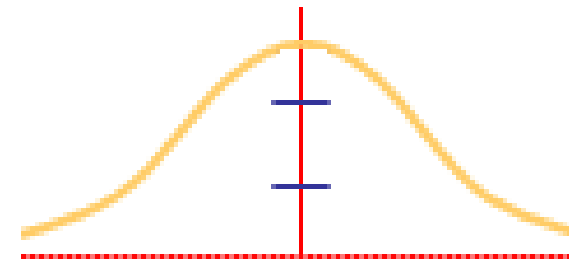
$$\gamma_2 > 0$$



Mesocurtica

$$\beta = 3$$

$$\gamma_2 = 0$$



Platicurtica

$$\beta < 3$$

$$\gamma_2 < 0$$

APPIATTIMENTO (O CURTOSI) IN R

Anche per misurare la curtosi di una distribuzione si possono usare dei pacchetti aggiuntivi, che sono gli stessi per la simmetria:

ES.

- moment
- e1071
- fUtilities

CREAZIONE DI UNA FUNZIONE PER BETA

$$\beta = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4$$

```
beta = function(x) {  
  m4 = mean((x-mean(x))^4)  
  curt = m4/(sd(x)^4)  
  curt  
}
```

APPIATTIMENTO (O CURTOSI) IN R

ES.

$x = c(0, 1, 1, 2, 2, 3, 4, 5)$

Misurare la curtosi di x .

APPIATTIMENTO (O CURTOSI) IN R

ES.

$x = c(0, 1, 1, 2, 2, 3, 4, 5)$

$\text{beta}(x) = 1.569003$

La distribuzione presenta un andamento
“schiacciato” ovvero platicurtico.

CREAZIONE DI UNA FUNZIONE PER GAMMA2

$$\gamma_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3$$

```
gamma2 = function(x) {  
  m4 = mean((x-mean(x))^4)  
  curt = m4 / (sd(x)^4)  
  curt - 3  
}
```

APPIATTIMENTO (O CURTOSI) IN R

ES.

$x = c(0, 1, 1, 2, 2, 3, 4, 5)$

Misurare l'appiattimento di x con γ_2

APPIATTIMENTO (O CURTOSI) IN R

ES.

$x = c(0, 1, 1, 2, 2, 3, 4, 5)$

$\text{gamma2}(x) = -1.430997$

Essendo negativo, γ_2 conferma la forma platicurtica della distribuzione.

Altro metodo per calcolare gamma2:
> $\text{beta}(x) - 3$

VERIFICA GRAFICO APPIATTIMENTO

```
hist(x, freq=TRUE, breaks=length(x))
```

Histogram of x

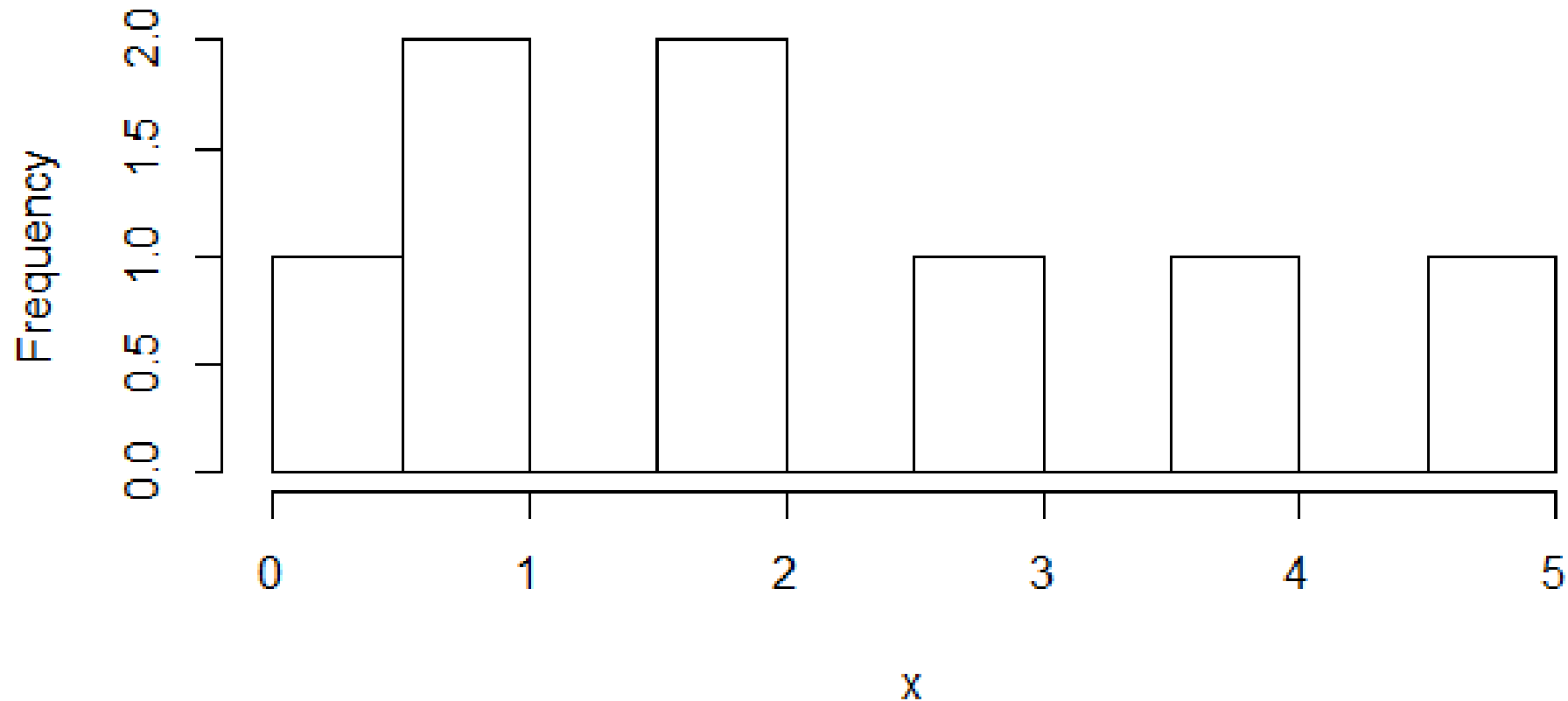


TABELLE A DOPPIA ENTRATA E MISURA DELLA CONNESSIONE FRA DUE FENOMENI

ESEMPIO DI TABELLA A DOPPIA ENTRATA

		CAPELLI	
		BIONDI	NERI
OCCHI	AZZURRI	25	10
	SCURI	15	60

CREAZIONE TABELLA A DOPPIA ENTRATA

**# CREIAMO INNANZITUTTO LA MATRICE DEI DATI
CON IL COMANDO `matrix`:**

```
> colore=matrix(c(25, 10, 15, 60), nrow=2,  
byrow=TRUE)
```

**# `nrow=2` E' IL NUMERO DI RIGHE, `byrow=TRUE`
INDICA CHE I DATI VANNO LETTI PER RIGA**

CREAZIONE TABELLA A DOPPIA ENTRATA

CREIAMO LE ETICHETTE PER LE RIGHE E LE COLONNE:

> occhi=c("azzurri", "scuri")

> capelli=c("biondi", "neri")

CREAZIONE TABELLA A DOPPIA ENTRATA

ASSEGNIAMO LE ETICHETTE ALLA MATRICE CON IL COMANDO:

```
> dimnames(colore)=list(occhi, capelli)
```

```
> colore
```

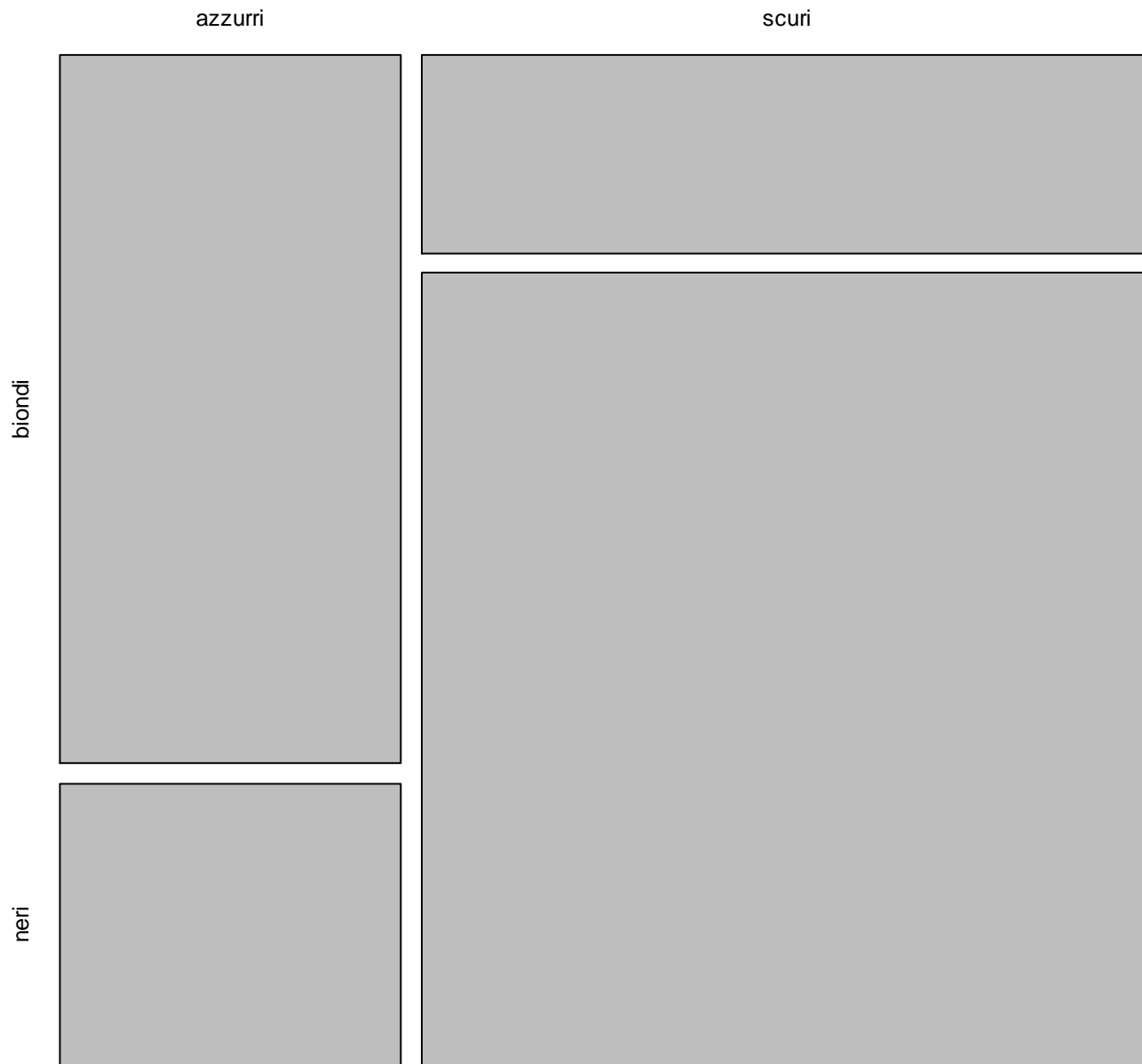
	biondi	neri
azzurri	25	10
scuri	15	60

DISEGNO GRAFICO MOSAICPLOT

DISEGNAMO IL GRAFICO AD AREE CHE RAPPRESENTA
LA TABELLA:

```
> mosaicplot(colore)
```

colore



IL GRAFICO RIPORTA LA RELAZIONE CHE ESISTE FRA I CARATTERI DEGLI OCCHI (AZZURRI O SCURI) E QUELLO DEI CAPELLI (BIONDI O NERI). L'AREA PIÙ GRANDE È QUELLA RELATIVA AGLI OCCHI SCURI E AI CAPELLI NERI, MENTRE SONO POCHI QUELLI CHE HANNO I CAPELLI NERI E GLI OCCHI AZZURRI. QUESTO GRAFICO PERMETTE DI AVERE SUBITO UN'IDEA DEI RAPPORTI DI "FORZA" CHE CI SONO FRA LE VARIABILI.

CALCOLO DEL CHI-QUADRATO

- ▶ Il test del chi-quadrato consiste in un test che mette a confronto le seguenti due ipotesi:
- ▶ **ipotesi nulla H_0** : afferma che c'è indipendenza fra i due fenomeni;
- ▶ **ipotesi alternativa H_1** : che invece dice che c'è una connessione fra i caratteri.

CALCOLO DEL CHI-QUADRATO

- ▶ In R il test del chi-quadrato viene condotto molto semplicemente con il comando: **chisq.test**
- > testchiq=**chisq.test**(colore)
- > testchiq
- ▶ Nel caso di tabelle 2x2, il **chisq.test** applica una correzione, quella di Yates. Se si desidera non usarla, occorre specificare l'opzione `correct=FALSE`

```
> testchiq=chisq.test(colore, correct=FALSE)
```

```
> testchiq
```


CALCOLO DEL CHI-QUADRATO

```
> testchiq
```

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

```
X-squared = 25.0983, df = 1, p-value =  
5.448e-07
```

- ▶ **“X-squared”** è il chi-quadrato calcolato
- ▶ **“df”** sono i degrees of freedom, i gradi di libertà, dati dal prodotto:
$$df = (n. \text{ Righe} - 1) * (n. \text{ Colonne} - 1)$$
- ▶ **“p-value”** è il livello di significatività. Questo valore deve essere inferiore al 5% (ovvero 0,05) per considerare valido il risultato trovato con il test.

TAVOLA DEL CHI-QUADRATO

	alpha (significatività)	
g.d.l.	1%	5%
1	6,64	3,84
2	9,21	5,99
3	11,35	7,82
4	13,28	9,49
5	15,09	11,07
6	16,81	12,59
7	18,48	14,07
8	20,09	15,51
9	21,67	16,92
10	23,21	18,31

CONFRONTO DEL CHI-QUADRATO CALCOLATO CON LA SOGLIA TEORICA

- ▶ Il valore del chi quadrato (X-squared) così calcolato va confrontato con un valore teorico per poter accettare o meno l'ipotesi nulla H_0 .
- ▶ In particolare le soglie critiche del chi-quadrato con 1 g.d.l. (grado di libertà) sono:
 - ▶ **3.84** per un livello di significatività del **5%**
 - ▶ **6.64** per un livello di significatività dell'**1%**
- ▶ Questi valori sono le soglie oltre le quali si rifiuta l'ipotesi nulla sbagliando rispettivamente al massimo nel 5% dei casi o solo nell'1%.

CONFRONTO CHI-QUADRATO CON SOGLIA TEORICA

3.84 per un livello di significatività del 5% e 1 g.d.l.

6.64 per un livello di significatività dell'1% e 1 g.d.l.

- ▶ In questo caso abbiamo 25.0983, che è abbondantemente superiore non solo a 3.84, che è la soglia critica per sbagliare al massimo nel 5% dei casi, ma addirittura a 6.64, che è la soglia critica oltre la quale si rifiuta l'ipotesi nulla di indipendenza sbagliando solo nell'1% dei casi.
- ▶ Quindi **il test rifiuta l'ipotesi nulla H_0 di indipendenza e quindi conferma che al 99% c'è connessione fra i fenomeni.**

CALCOLO DEL "V" DI CRAMER

- ▶ Una volta che abbiamo rilevato che c'è una connessione fra i 2 fenomeni, possiamo misurare quanto sono connessi fra di loro con un opportuno indice, il **V di Cramer**.
- ▶ Questo indicatore assume:
 - ▶ valore 0 nel caso di **perfetta indipendenza**;
 - ▶ valore 1 quando invece c'è la **massima connessione** fra i due fenomeni.

CALCOLO DEL "V" DI CRAMER

- ▶ Per calcolare il V di Cramer bisogna usare la seguente formula:

$$V = \sqrt{\frac{\chi^2}{N * (\min(\text{righe}, \text{colonne}) - 1)}}$$

- ▶ χ^2 = valore della variabile chi-quadrato ricavato dal test chi quadrato (**\$statistic**)
- ▶ N = numero totale di casi (**N=sum(colore)**)
- ▶ $\min(\text{righe}, \text{colonne}) - 1$ = si sceglie il minore fra il numero delle righe e delle colonne; quindi si sottrae 1 (ES. tab. 2 righe e 3 colonne: si sceglie 2, quindi si toglie 1: 2-1=1)

- ▶ ES. tabella 2 x 2:

	BIONDI	NERI
AZZURRI	25	10
SCURI	15	60

- ▶ n. righe = 2
- ▶ n. colonne = 2
- ▶ In questo caso il numero di righe e di colonne è lo stesso, quindi scelgo 2.
- ▶ Da 2 sottraggo 1: $2-1 = 1$

▶ ALTRO ES. tabella 4 x 3:

	ALPHA	BETA	GAMMA
TIPO A	25	10	12
TIPO B	15	60	48
TIPO AB	22	10	36
TIPO 0	21	6	12

▶ n. righe = 4

▶ n. colonne = 3

▶ In questo caso il minore fra il numero di righe e di colonne è 3.

▶ Da 3 sottraggo 1: $3-1 = 2$

CALCOLO DEL "V" DI CRAMER

PER CALCOLARE IL COEFFICIENTE V DI CRAMER,
DEVO QUINDI PRIMA RICAVARMI LE SINGOLE
COMPONENTI: IL CHI QUADRATO E LA
NUMEROSITA' TOTALE "N"

IL VALORE DEL CHI-QUADRATO SI RICAVA DA:

```
> chiquadrato= testchiq$statistic
```

```
> chiquadrato
```

```
X-squared
```

```
25.09833
```

CALCOLO DEL "V" DI CRAMER

IL TOTALE DI ELEMENTI PRESENTI SI OTTIENE
COSÌ:

```
> N = sum(colore)
```

```
> N
```

```
[1] 110
```

CALCOLO DEL "V" DI CRAMER

IL VALORE DEL CHI QUADRATO SI CALCOLA COSI':

> V=sqrt(chiquadrato / (N*(2-1)))

> V

X-squared

0.4776679

CALCOLO DEL "V" DI CRAMER

ESEMPI DI COMMENTI AL V DI CRAMER:

- ▶ DA 0 A 0,2: BASSA CONNESSIONE
- ▶ DA 0,2 A 0,4: DISCRETA CONNESSIONE
- ▶ DA 0,4 A 0,6: BUONA CONNESSIONE
- ▶ DA 0,6 IN SU: ALTA CONNESSIONE (VARIABILI RIDONDANTI)

CALCOLO DEL "V" DI CRAMER

IL VALORE DEL V DI CRAMER (0.4776679) PORTA A RITENERE CHE C'E' UNA BUONA CONNESSIONE FRA I DUE CARATTERI "COLORE DEI CAPELLI" E "COLORE DEGLI OCCHI", NEL SENSO CHE E' CORRETTO IPOTIZZARE CHE DI SOLITO AD UN CERTO COLORE DEI CAPELLI CORRISPONDE UN CERTO COLORE DEGLI OCCHI.

ES. FARMACO

Ipotizziamo di avere i risultati di un test sull'efficacia di un nuovo farmaco su N=400 pazienti.

		EFFETTO	
		MIGLIORAMENTO	PEGGIORAMENTO
TRATTAMENTO	FARMACO	250	50
	PLACEBO	50	50

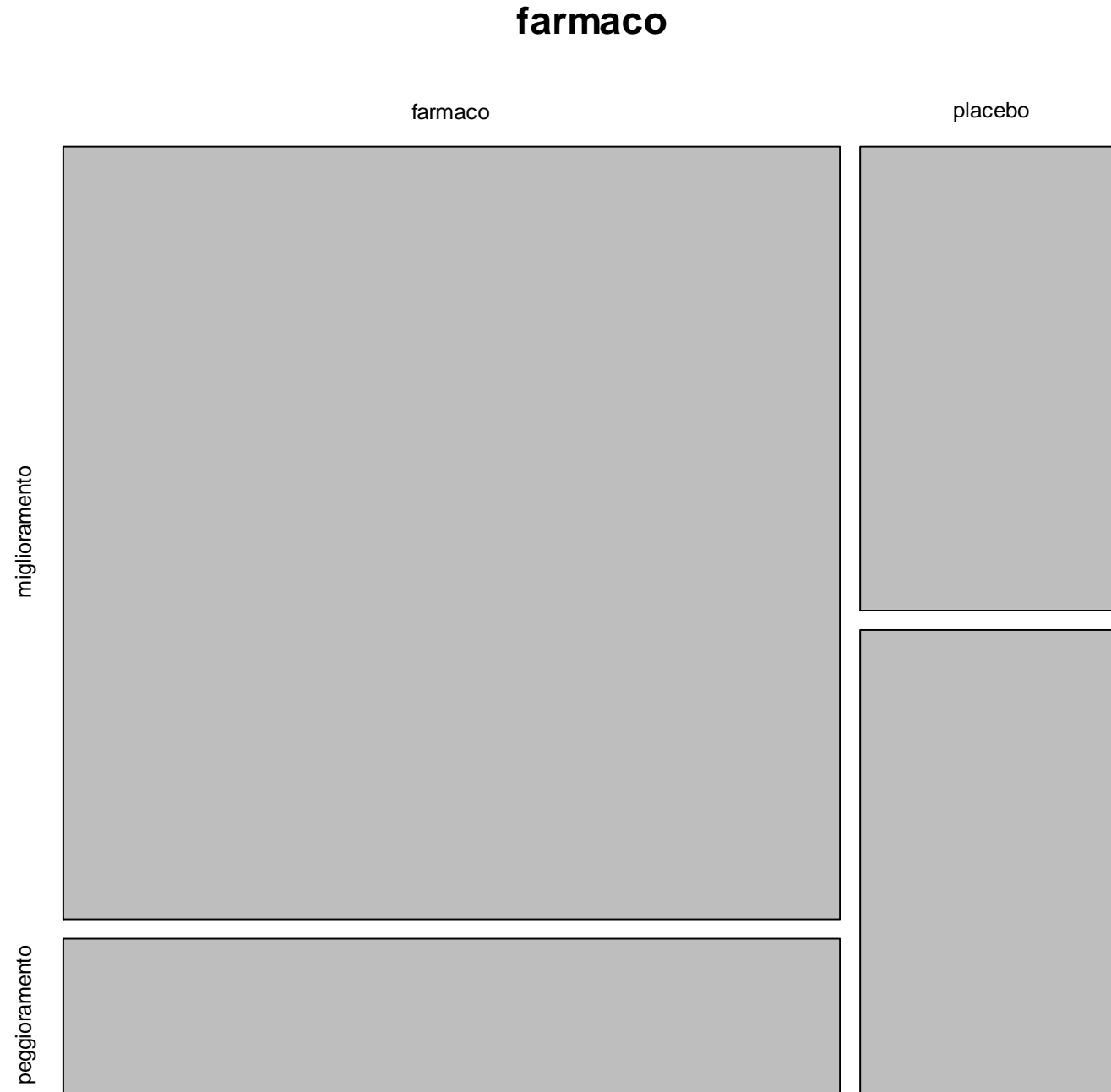
g.d.l.	alpha (significatività)	
	1%	5%
1	6,64	3,84
2	9,21	5,99
3	11,35	7,82
4	13,28	9,49
5	15,09	11,07
6	16,81	12,59
7	18,48	14,07
8	20,09	15,51
9	21,67	16,92
10	23,21	18,31

ES. FARMACO

```
> farmaco=matrix(c(250, 50, 50, 50), nrow=2,  
byrow=TRUE)  
> trattamento=c("farmaco", "placebo")  
> effetto=c("miglioramento", "peggioramento")  
> dimnames(farmaco)=list(trattamento, effetto)  
> farmaco
```

	miglioramento	peggioramento
farmaco	250	50
placebo	50	50

```
> mosaicplot(farmaco)
```



ES. FARMACO

CALCOLIAMO IL TEST DEL CHI QUADRATO

```
> testchiq=chisq.test(farmaco)
> testchiq
```

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

data: farmaco

X-squared = 42.6844, df = 1, p-value = 6.432e-11

POICHE' IL VALORE CALCOLATO DEL CHI-QUADRATO E' 42.6844, BEN SUPERIORE ALLA SOGLIA CRITICA DI 6.64 VALIDO ALL'1% CON 1 GRADO DI LIBERTA' (G.D.L.)(O DF=DEGREES OF FREEDOM NELL'OUTPUT), SI RIFIUTA L'IPOTESI NULLA DI INDIPENDENZA E SI CONFERMA LA CONNESSIONE FRA I FENOMENI

CALCOLO IL V DI CRAMER

```
> chiquadrato= testchisq$statistic
```

```
> chiquadrato
```

```
X-squared
```

```
42.68444
```

IL TOTALE DI ELEMENTI PRESENTI E':

```
> N = sum(farmaco)
```

```
> N
```

```
[1] 400
```

```
> V=sqrt( chiquadrato / (N*(2-1)) )
```

```
> V
```

```
X-squared
```

```
0.3266667
```

**# IL RISULTATO PORTA AD AFFERMARE CHE C'È UNA DISCRETA
CONNESSIONE FRA I DUE FENOMENI**

ES. SOPRAVVISSUTI DEL TITANIC

La tabella riporta i sopravvissuti e i deceduti fra i passeggeri del STAGE a seconda della classe di appartenenza.

		ESITO	
		DECEDUTI	SOPRAVVISSUTI
CLASSE	PRIMA	122	203
	SECONDA	167	118
	TERZA	528	178
	EQUIPAGGIO	673	212

g.d.l.	alpha (significatività)	
	1%	5%
1	6,64	3,84
2	9,21	5,99
3	11,35	7,82
4	13,28	9,49
5	15,09	11,07
6	16,81	12,59
7	18,48	14,07
8	20,09	15,51
9	21,67	16,92
10	23,21	18,31

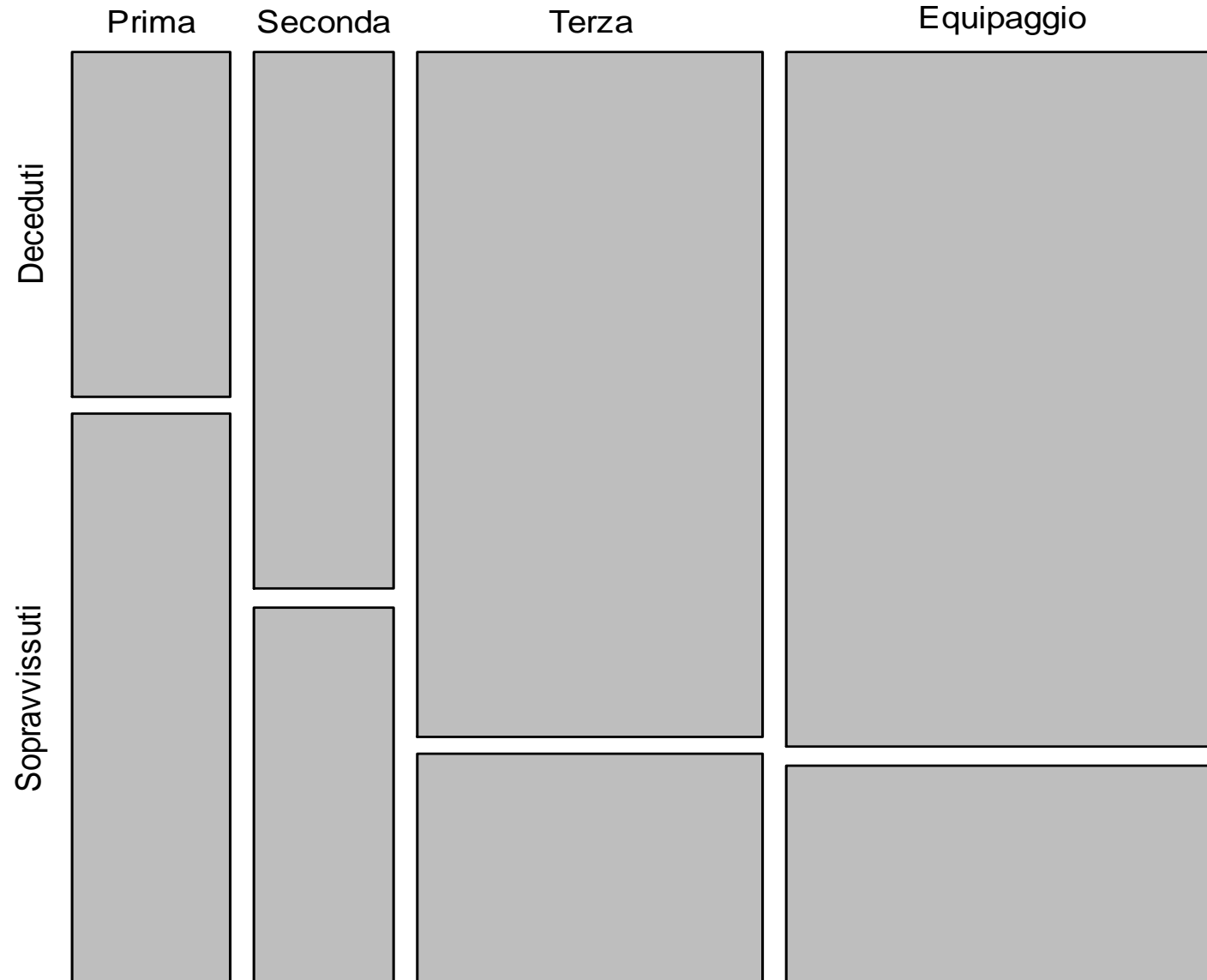
ES. TITANIC

```
> titanic=matrix(c(122, 203, 167, 118, 528, 178, 673,
  212), nrow=4, byrow=TRUE)
> classe=c("prima", "seconda", "terza", "equipaggio")
> esito=c("deceduti", "sopravvissuti")
> dimnames(titanic)=list(classe, esito)
> titanic
```

	Deceduti	Sopravvissuti
Prima	122	203
Seconda	167	118
Terza	528	178
Equipaggio	673	212

```
> mosaicplot(titanic)
```

titanic



ES. TITANIC

CALCOLIAMO IL TEST DEL CHI QUADRATO

```
> testchiq=chisq.test(titanic)
> testchiq
```

Pearson's Chi-squared test

data: STAGE

X-squared = 190.4011, df = 3, p-value < 2.2e-16

POICHE' IL VALORE CALCOLATO DEL CHI-QUADRATO E' 190.4011, BEN SUPERIORE ALLA SOGLIA CRITICA DI 11.35 VALIDO ALL' 1% CON 3 GRADI DI LIBERTA', SI RIFIUTA L'IPOTESI NULLA DI INDIPENDENZA E SI CONFERMA LA CONNESSIONE FRA I FENOMENI, OVVERO FAR PARTE DELLA PRIMA, SECONDA, TERZA CLASSE O DELL'EQUIPAGGIO FACEVA DIFFERENZA FRA LA VITA E LA MORTE. I GRADI DI LIBERTA' SONO 3 PERCHE' DATI DA $(R-1)*(C*1)=(4-1)*(2-1)$

CALCOLO IL V DI CRAMER

```
> chiquadrato= testchisq$statistic
```

```
> chiquadrato
```

```
X-squared
```

```
190.4011
```

IL TOTALE DI ELEMENTI PRESENTI E':

```
> N = sum(titanic)
```

```
> N
```

```
[1] 2201
```

CALCOLO IL V DI CRAMER CONSIDERANDO IL MINORE FRA IL N. DI RIGHE E COLONNE

```
> V=sqrt( chiquadrato / (N*(2-1)) )
```

```
> V
```

```
X-squared
```

```
0.2941201
```

**# IL RISULTATO PORTA AD AFFERMARE CHE C'È UNA DISCRETA
CONNESSIONE FRA I DUE FENOMENI**

ES. STAGE E ASSUNZIONE (CASO NORMALE)

Si vuole verificare se esiste una relazione fra il fatto di svolgere uno stage presso un importante istituto di credito e la successiva eventuale assunzione. Sono stati così presi in considerazione 200 ragazzi così distribuiti:

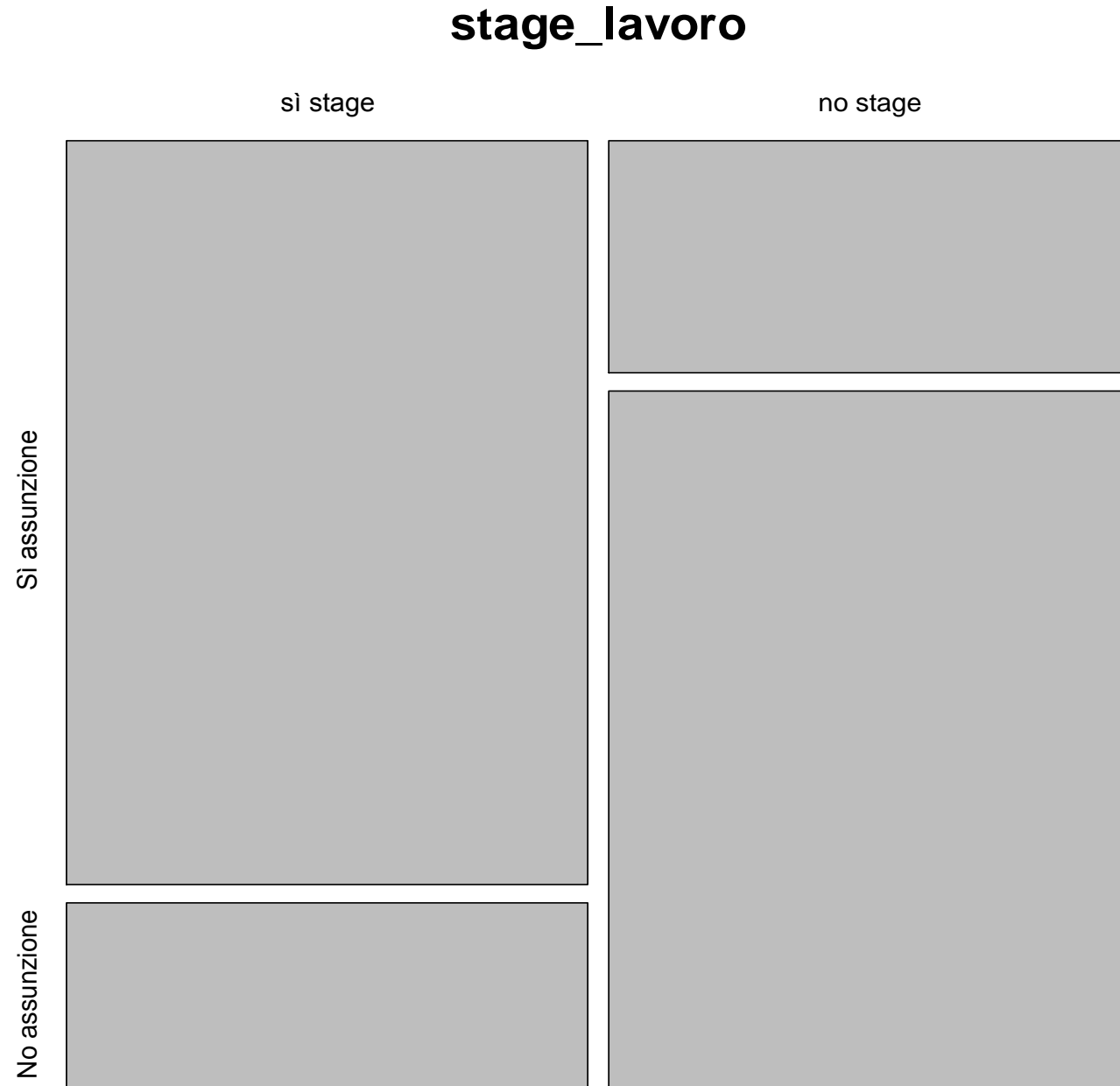
		ASSUNZIONE?		
		SI'	NO	Totale
STAGE?	SI'	80	20	100
	NO	25	75	100
	Totale	105	95	200

g.d.l.	alpha (significatività)	
	1%	5%
1	6,64	3,84
2	9,21	5,99
3	11,35	7,82
4	13,28	9,49
5	15,09	11,07
6	16,81	12,59
7	18,48	14,07
8	20,09	15,51
9	21,67	16,92
10	23,21	18,31

ES. STAGE

- ▶ `> stage_lavoro=matrix(c(80, 20, 25, 75), nrow=2, byrow=TRUE)`
- ▶ `> stage=c("sì stage", "no stage")`
- ▶ `> lavoro=c("Sì assunzione", "No assunzione")`
- ▶ `> dimnames(stage_lavoro)=list(stage, lavoro)`
- ▶ `> stage_lavoro`
- ▶ Sì assunzione No assunzione
- ▶ sì stage 80 20
- ▶ no stage 25 75

```
> mosaicplot(stage_lavoro)
```



ES. STAGE

CALCOLIAMO IL TEST DEL CHI QUADRATO

```
> testchiq=chisq.test(stage_lavoro)
> testchiq
```

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

```
data: stage_lavoro
```

```
X-squared = 58.4662, df = 1, p-value = 2.068e-14
```

POICHE' IL VALORE CALCOLATO DEL CHI-QUADRATO E' 58.4662, BEN SUPERIORE ALLA SOGLIA CRITICA DI 6.64 VALIDO ALL'1%, SI RIFIUTA L'IPOTESI NULLA DI INDIPENDENZA E SI CONFERMA LA CONNESSIONE FRA I FENOMENI, OVVERO FARE UNO STAGE COMPORTA MAGGIORI PROBABILITA' DI ESSERE ASSUNTI. I GRADI DI LIBERTA' SONO 1 PERCHE' DATI DA $(R-1)*(C*1)=(2-1)*(2-1)$

CALCOLO IL V DI CRAMER

```
> chiquadrato= testchisq$statistic
```

```
> chiquadrato
```

```
X-squared
```

```
58.46617
```

IL TOTALE DI ELEMENTI PRESENTI E':

```
> N = sum(stage_lavoro)
```

```
> N
```

```
[1] 200
```

```
> V=sqrt( chiquadrato / (N*(2-1)) )
```

```
> V
```

```
X-squared
```

```
0.5406763
```

**# IL RISULTATO PORTA AD AFFERMARE CHE C'È UNA BUONA
CONNESSIONE FRA I DUE FENOMENI**

ES. S.O. PC E S.O. SMARTPHONE

La seguente tabella riporta la distribuzione dei sistemi operativi di computer e smartphone di un campione di 1.000 persone.

		SO smartphone			
		Windows	iOS	Android	TOTALE
SO computer	Windows	100	180	320	600
	Mac OS	60	120	50	230
	Linux	50	60	60	170
	TOTALE	210	360	430	1.000

g.d.l.	alpha (significatività)	
	1%	5%
1	6,64	3,84
2	9,21	5,99
3	11,35	7,82
4	13,28	9,49
5	15,09	11,07
6	16,81	12,59
7	18,48	14,07
8	20,09	15,51
9	21,67	16,92
10	23,21	18,31

ES. S.O. PC E S.O. SMARTPHONE

```
> SO=matrix(c(100, 180, 320, 60, 120, 50, 50, 60, 60), nrow=3,  
byrow=TRUE)
```

```
> SOpc=c("Windows", "Mac OS", "Linux")
```

```
> SOsmart=c("Windows", "iOS", "Android")
```

```
> dimnames(SO)=list(SOpc, SOsmart)
```

```
> SO
```

	Windows	iOS	Android
Windows	100	180	320
Mac OS	60	120	50
Linux	50	60	60

```
> mosaicplot(SO)
```

ES. S.O. PC E S.O. SMARTPHONE

```
> testchiq=chisq.test(SO)
> testchiq
```

Pearson's Chi-squared test

data: SO

X-squared = 78.0887, df = 4, p-value = 4.424e-16

POICHE' IL VALORE CALCOLATO DEL CHI-QUADRATO E' 78.0887, BEN SUPERIORE ALLA SOGLIA CRITICA DI 13,28 VALIDO ALL'1% PER 4 G.D.L., SI RIFIUTA L'IPOTESI NULLA DI INDIPENDENZA E SI CONFERMA LA CONNESSIONE FRA I FENOMENI, OVVERO AVERE IL COMPUTER CON UN CERTO SISTEMA OPERATIVO INFLUENZA LA SCELTA DEL SISTEMA OPERATIVO DELLO SMARTPHONE. I GRADI DI LIBERTA' SONO 4 PERCHE' DATI DA $(r-1)*(c*1)=(3-1)*(3-1)$

g.d.l.	alpha (significatività)	
	1%	5%
1	6,64	3,84
2	9,21	5,99
3	11,35	7,82
4	13,28	9,49
5	15,09	11,07
6	16,81	12,59
7	18,48	14,07
8	20,09	15,51
9	21,67	16,92
10	23,21	18,31

ES. S.O. PC E S.O. SMARTPHONE

CALCOLIAMO IL VALORE DELLA STATISTICA V DI CRAMER

```
> chiquadrato=testchisq$statistic  
> chiquadrato  
X-squared  
78.08871
```

IL TOTALE DI ELEMENTI PRESENTI SI OTTIENE IN QUESTO MODO:

```
> N = sum(SO)  
> N  
[1] 1000
```

SI SCEGLIE IL MINORE FRA IL NUMERO DI RIGHE E DI COLONNE E SI SOTTRAE 1

```
> V=sqrt( chiquadrato / (N*(3-1)) )  
> V  
X-squared  
0.1975965
```

IL RISULTATO PORTA AD AFFERMARE CHE C'È UNA BASSA CONNESSIONE FRA I DUE FENOMENI