

LABORATORIO DI PROBABILITA' E STATISTICA

Docente: Bruno Gobbi

Bonus - ESERCIZI DI RIPASSO FINALE

1 - STATISTICA DESCRITTIVA - TABLET

ESERCIZIO 1: La seguente tabella riporta i volumi di vendita (in migliaia di pezzi) dei principali produttori di computer nel secondo trimestre del 2014.

Creare una tabella in R che riporti i volumi di vendita in migliaia di pezzi e in percentuale. Alla fine creare un grafico a istogramma per i volumi di vendita in migliaia e uno a torta per le percentuali.

MARCHIO	VENDITE
Apple	13.300
Samsung	8.500
Lenovo	2.400
Asus	2.300
Acer	1.000
Altri	21.900

1 - STATISTICA DESCRITTIVA - TABLET

```
> marchio=c("Apple", "Samsung", "Lenovo", "Asus", "Acer", "Altri")
```

```
> vendite=c(13300, 8500, 2400, 2300, 1000, 21900)
```

```
> tablet=data.frame(marchio, vendite)
```

```
> tablet
```

	marchio	vendite
1	Apple	13300
2	Samsung	8500
3	Lenovo	2400
4	Asus	2300
5	Acer	1000
6	Altri	21900

1 - STATISTICA DESCRITTIVA - TABLET

CREIAMO LA COLONNA DELLE PERCENTUALI DI VENDITA

```
> tot_vendite=sum(vendite)
```

```
> tot_vendite
```

```
[1] 49400
```

```
> perc=vendite/tot_vendite
```

```
> perc
```

```
[1] 0.26923077 0.17206478 0.04858300 0.04655870 0.02024291  
0.44331984
```

1 - STATISTICA DESCRITTIVA - TABLET

CREIAMO LA COLONNA DELLE PERCENTUALI DI VENDITA

```
> tablet=data.frame(tablet, perc)
```

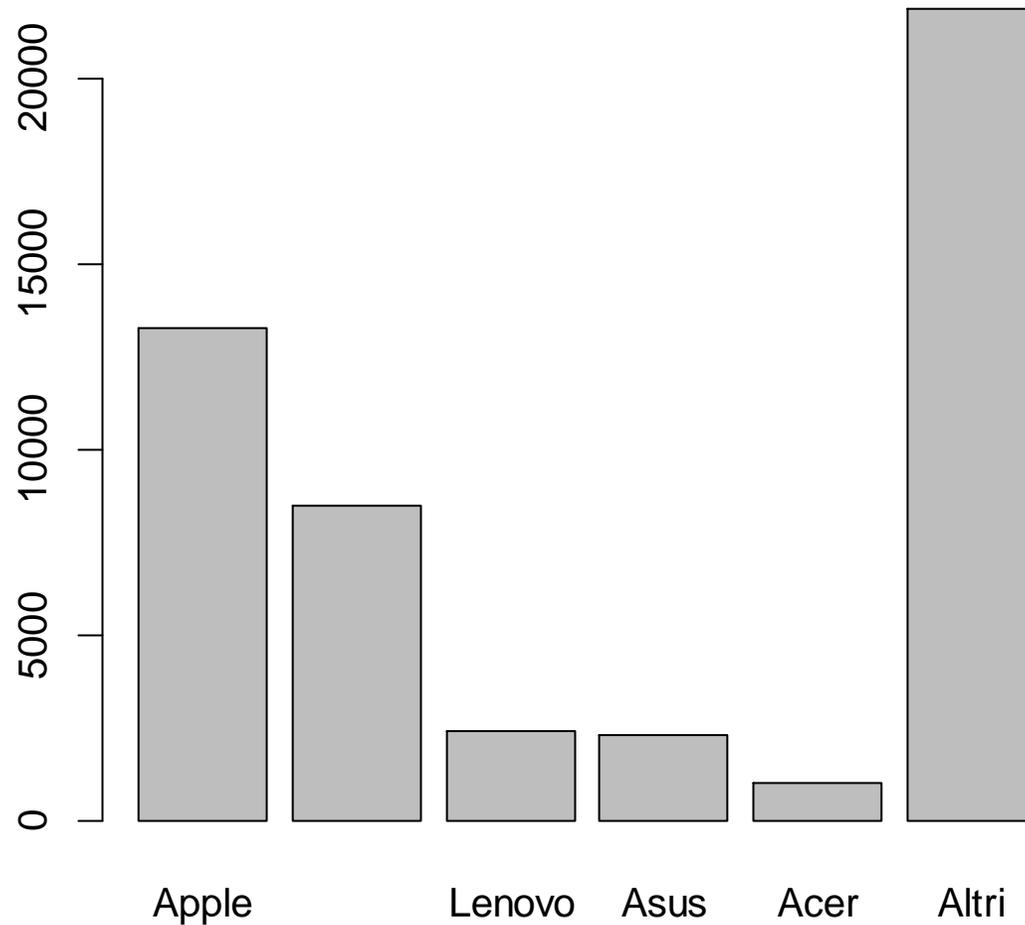
```
> tablet
```

	marchio	vendite	perc
1	Apple	13300	0.26923077
2	Samsung	8500	0.17206478
3	Lenovo	2400	0.04858300
4	Asus	2300	0.04655870
5	Acer	1000	0.02024291
6	Altri	21900	0.44331984

1 - STATISTICA DESCRITTIVA - TABLET

GRAFICO DEI VOLUMI DI VENDITA

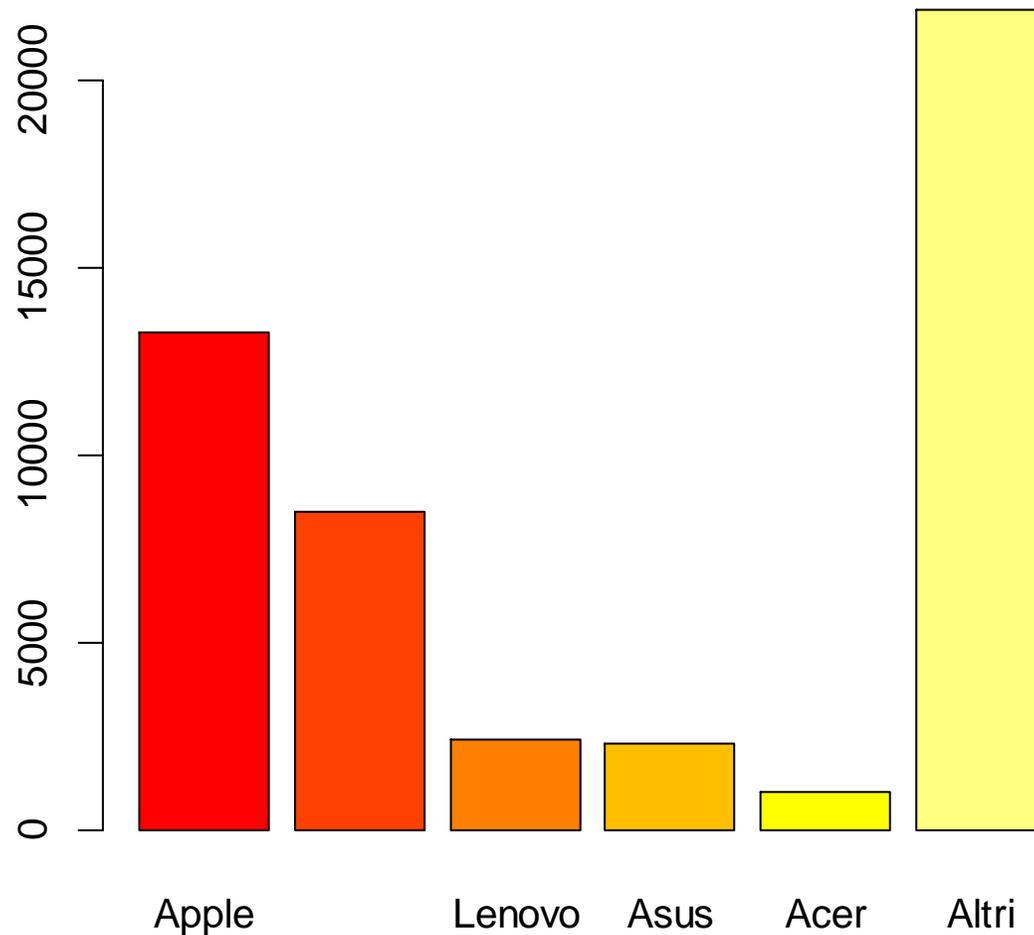
```
> barplot(vendite, names.arg=marchio)
```



1 - STATISTICA DESCRITTIVA - TABLET

GRAFICO DEI VOLUMI DI VENDITA

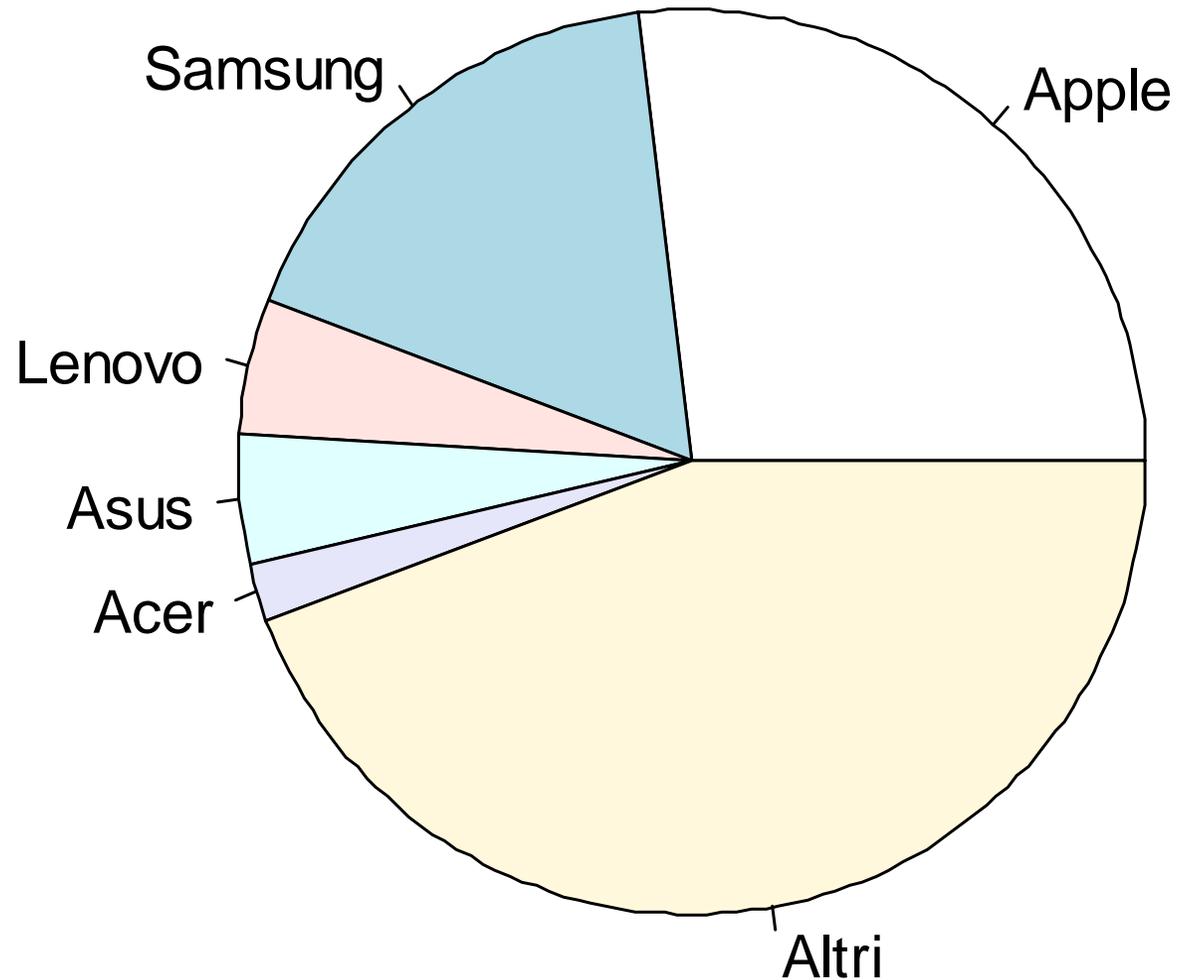
```
> barplot(vendite, names.arg=marchio, col=heat.colors(6))
```



1 - STATISTICA DESCRITTIVA - TABLET

GRAFICO A TORTA DELLE PERCENTUALI DI VENDITA

> pie(perc, labels=marchio)



2 - CURTOSI E APPIATTIMENTO - TABLET

ESERCIZIO 2: Sui dati della tabella precedente calcolare la simmetria e l'appiattimento della distribuzione delle vendite in migliaia utilizzando degli opportuni indici.

MARCHIO	VENDITE
Apple	13.300
Samsung	8.500
Lenovo	2.400
Asus	2.300
Acer	1.000
Altri	21.900

INDICE DI SIMMETRIA γ (gamma) DI FISHER

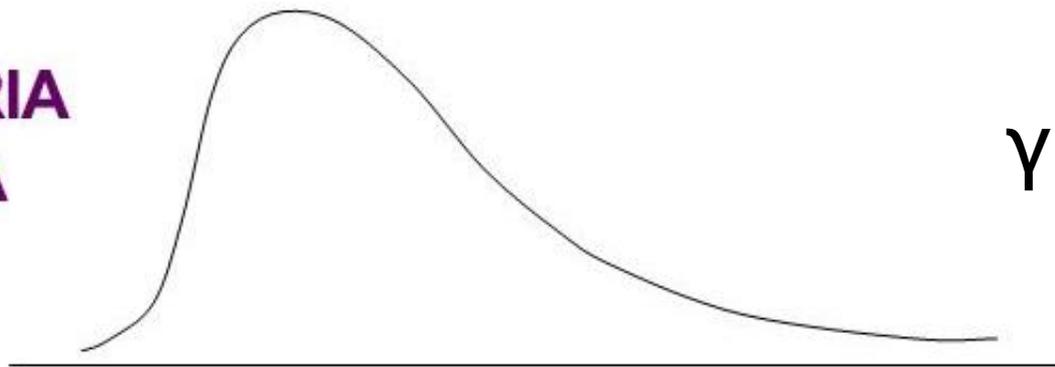
$$\gamma = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3$$

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

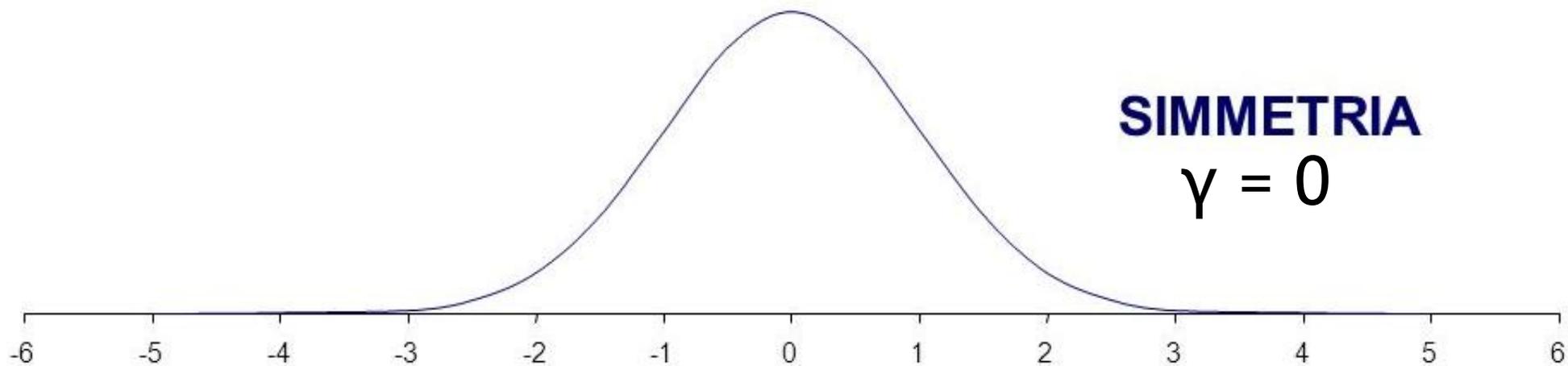
**ASIMMETRIA
POSITIVA**



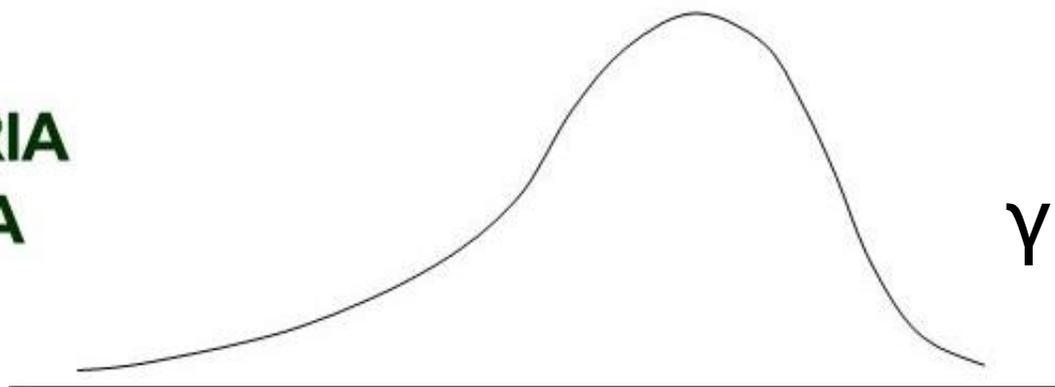
$$\gamma > 0$$

SIMMETRIA

$$\gamma = 0$$



**ASIMMETRIA
NEGATIVA**



$$\gamma < 0$$

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

2 - CURTOSI E APPIATTIMENTO - TABLET

> gamma(vendite) = 0.5788712

IL VALORE DELL'INDICE GAMMA È PARI A 0.5788712. C'È UN'ASIMMETRIA POSITIVA, LA DISTRIBUZIONE PRESENTA UNA CODA PIÙ LUNGA A DESTRA.

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

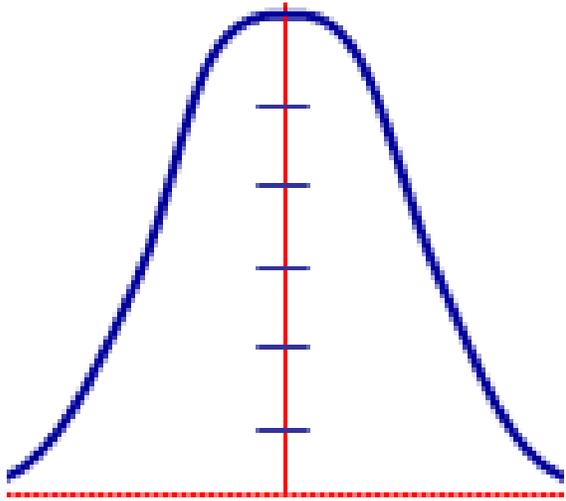
$$\gamma_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3$$

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

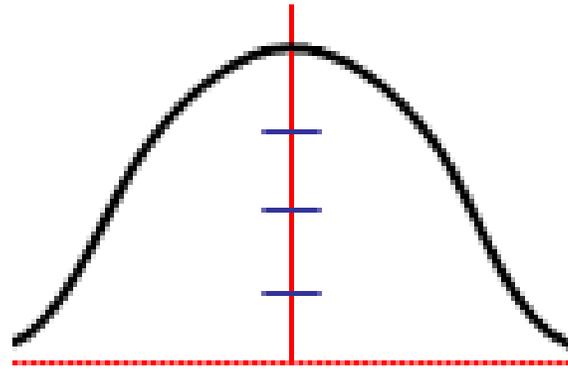
INDICI DI APPIATTIMENTO (CURTOSI)



Leptocurtica

$$\beta > 3$$

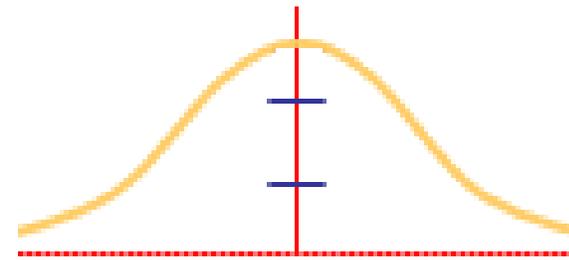
$$\gamma_2 > 0$$



Mesocurtica

$$\beta = 3$$

$$\gamma_2 = 0$$



Platicurtica

$$\beta < 3$$

$$\gamma_2 < 0$$

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  
}
```

2 - CURTOSI E APPIATTIMENTO - TABLET

```
> beta(vendite)
```

```
[1] 1.518738
```

IL VALORE DELL'INDICE BETA E' PARI A
1.518738 LA DISTRIBUZIONE APPARE
SCHIACCIATA, PLATICURTICA

```
> beta(vendite)-3
```

```
[1] -1.481262
```

3 - STATISTICHE E BOXPLOT - AIR PASSENGERS

ESERCIZIO 3: Utilizzando la base dati già presente in R relativamente al numero di passeggeri aerei fra il 1949 e il 1960 (nome del database: "AirPassengers"), calcolare:

- Media
- Mediana
- Primo e terzo quartile
- Minimo e Massimo
- Varianza campionaria
- Numero di elementi del database

Infine disegnare il grafico boxplot della serie storica.

3 - STATISTICHE E BOXPLOT - AIR PASSENGERS

```
> summary(AirPassengers)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
104.0	180.0	265.5	280.3	360.5	622.0

```
> var(AirPassengers)
```

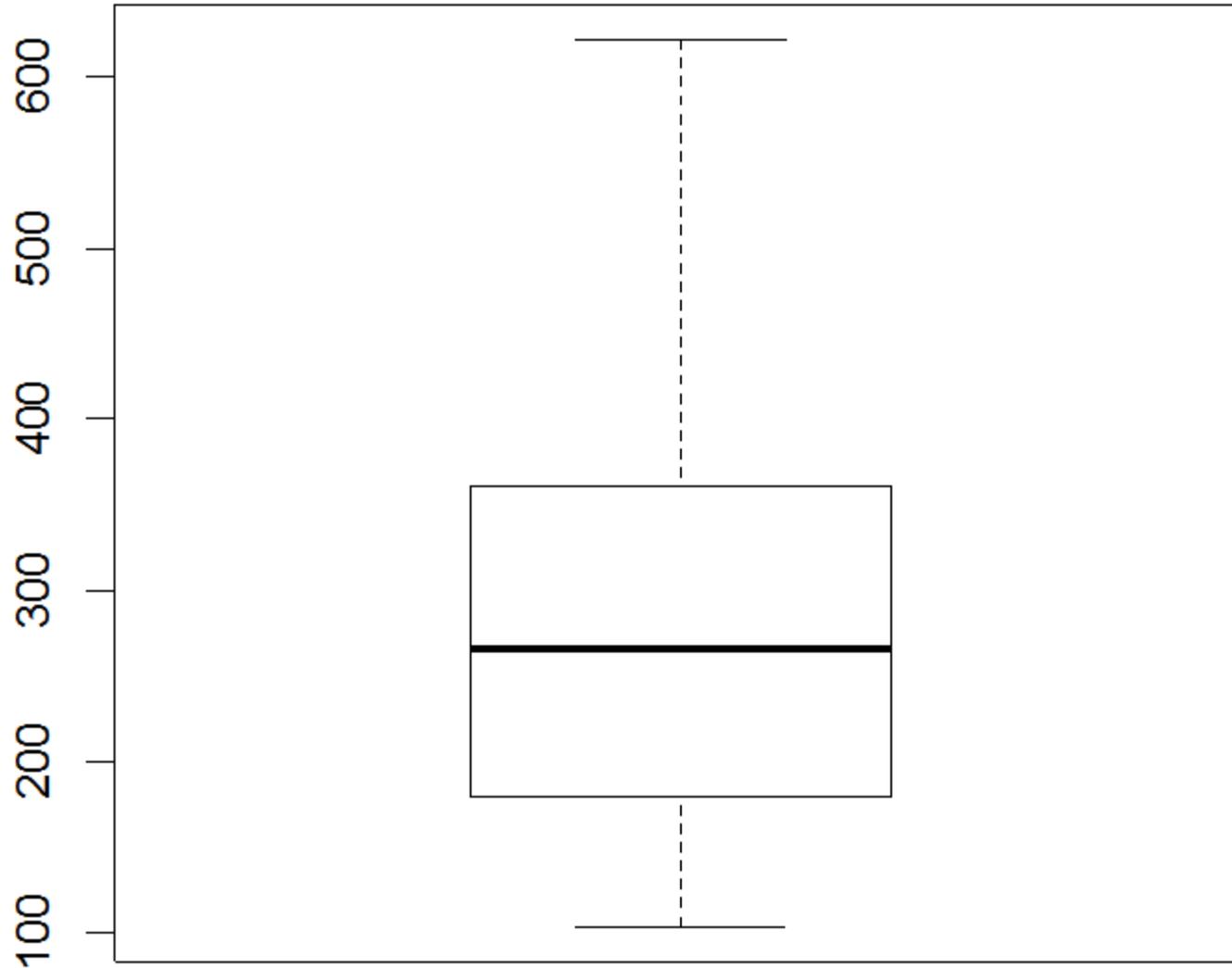
```
[1] 14391.92
```

```
> length(AirPassengers)
```

```
[1] 144
```

3 - STATISTICHE E BOXPLOT - AIR PASSENGERS

```
> boxplot(AirPassengers)
```



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

ESERCIZIO 4: La seguente tabella riporta la distribuzione dei sistemi operativi di computer e smartphone di un campione di 1.000 persone. Valutare la connessione ad un livello di significatività dell'1%.

		S.O. smartphone			
		Windows	iOS	Android	TOTALE
S.O. 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

ES. STAGE E ASSUNZIONE (CASO NORMALE)

ESERCIZIO 5 A: 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 E ASSUNZIONE (CASO NORMALE)

```
> 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 E ASSUNZIONE (CASO NORMALE)

```
> 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 $(r-1)*(c*1)=(2-1)*(2-1)$

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 E ASSUNZIONE (CASO NORMALE)

CALCOLIAMO IL VALORE DELLA STATISTICA V DI CRAMER

```
> chiquadrato=testchisq$statistic  
> chiquadrato  
X-squared  
58.46617
```

IL TOTALE DI ELEMENTI PRESENTI SI OTTIENE IN QUESTO MODO:

```
> N = sum(stage_lavoro)  
> N  
[1] 200
```

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

```
> 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. STAGE E ASSUNZIONE (CASO LIMITE 1)

ESERCIZIO 5 B: 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'	100	0	100
	NO	0	100	100
	Totale	100	100	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 E ASSUNZIONE (CASO LIMITE 1)

```
> stage_lavoro=matrix(c(100, 0, 0, 100), nrow=2, byrow=TRUE)
```

```
> dimnames(stage_lavoro)=list(stage, lavoro)
```

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

```
> testchiq
```

```
data: stage_lavoro
```

```
X-squared = 196.02, df = 1, p-value < 2.2e-16
```

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

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

```
> V
```

```
0.99
```

QUI C'E' LA MASSIMA CONNESSIONE, NEL SENSO CHE QUANDO UNO STUDENTE FA LO STAGE, VIENE SEMPRE ASSUNTO E VICEVERSA.

IL CHI-QUADRATO E' MOLTO ALTO (196.02) E DI CONSEGUENZA IL V DI CRAMER E' VICINISSIMO A 1 (0.99)

ES. STAGE E ASSUNZIONE (CASO LIMITE 2)

ESERCIZIO 5 C: 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'	50	50	100
	NO	50	50	100
	Totale	100	100	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 E ASSUNZIONE (CASO LIMITE 2)

```
> stage_lavoro=matrix(c(50, 50, 50, 50), nrow=2, byrow=TRUE)
```

```
> dimnames(stage_lavoro)=list(stage, lavoro)
```

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

```
> testchiq
```

```
data: stage_lavoro
```

```
X-squared = 0, df = 1, p-value = 1
```

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

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

```
> V
```

```
0
```

NEL CASO DI EQUIDISTRIBUZIONE, NON C'E' NESSUNA CONNESSIONE, NEL SENSO CHE I DUE FENOMENI NON SEMBRANO AVERE ALCUN EFFETTO L'UNO SULL'ALTRO. CHE UNO STUDENTE FACCIA O MENO LO STAGE, NON SEMBRA CAMBIARE LE SUE POSSIBILITA' DI ESSERE ASSUNTO. IL CHI-QUADRATO E' PARI A ZERO E DI CONSEGUENZA LO E' ANCHE IL V DI CRAMER.

REGRESSIONE LINEARE: Alpha - Beta

ESERCIZIO 6:

Analizzare la relazione fra i fenomeni Alpha e Beta utilizzando la regressione lineare, disegnando il grafico, calcolando i parametri della retta interpolante, i residui con grafico, il coefficiente di correlazione lineare e giudicandone la bontà di accostamento.

Alpha	Beta
120	2000
102	1555
114	1650
118	1742
140	2218
135	1921

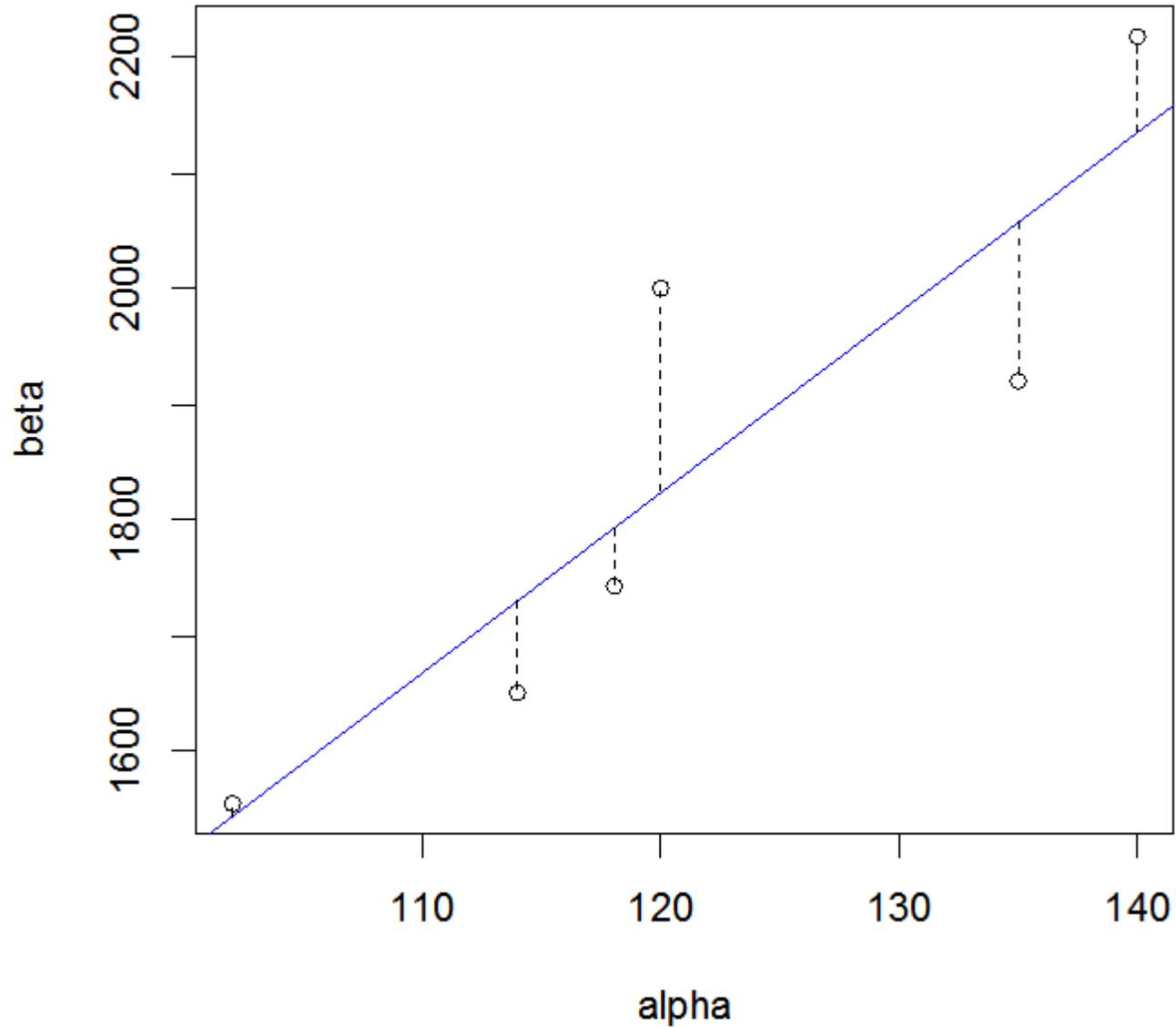
ES. STUDIO RELAZIONE alpha - beta

```
> alpha=c(120, 102, 114, 118, 140, 135)
> beta=c(2000, 1555, 1650, 1742, 2218, 1921)
> plot(alpha, beta)
> rettaalpha=lm(beta~alpha)
> abline(rettaalpha, col="blue")
> segments(alpha, fitted(rettaalpha), alpha, beta, lty=2)
> title(main="Regressione lineare fra alpha e beta")
```

Per scrivere la tilde ~ in
Ubuntu premere:
ALT GR + ì

ES. STUDIO RELAZIONE alpha - beta

Regressione lineare fra alpha e beta



ES. STUDIO RELAZIONE alpha - beta

```
> summary (rettaalpha)
```

Call:

```
lm(formula = beta ~ alpha)
```

Residuals:

1	2	3	4	5	6
175.66	10.62	-81.02	-51.23	82.60	-136.63

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-42.029	501.106	-0.084	0.9372
alpha	15.553	4.102	3.792	0.0192 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 128.1 on 4 degrees of freedom

Multiple R-squared: 0.7823, Adjusted R-squared: 0.7279

F-statistic: 14.38 on 1 and 4 DF, p-value: 0.01924

ES. STUDIO RELAZIONE alpha - beta

I PARAMETRI TROVATI SONO $a=-42.029$ E $b=15.553$

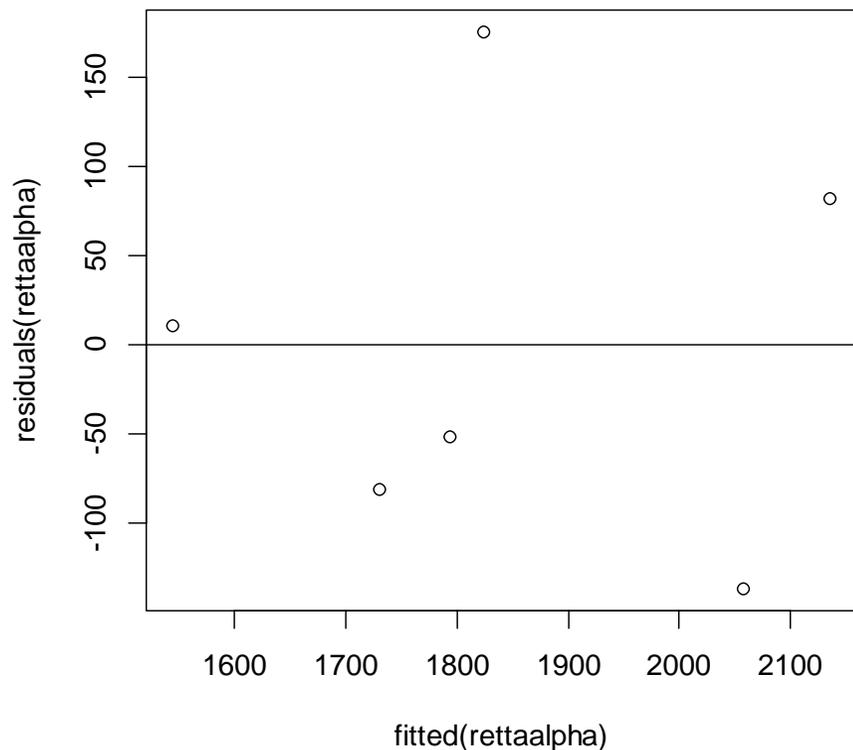
QUINDI IL MODELLO TEORICO SARA':

$$Y' = -42.029 + 15.553 \cdot \alpha$$

EFFETTIAMO L'ANALISI DEI RESIDUI

> plot(fitted(rettaalpha), residuals(rettaalpha))

> abline(0, 0)



L'analisi dei residui conferma che questi si distribuiscono in maniera uniforme e apparentemente casuale attorno all'asse zero, quindi si può confermare l'ipotesi di distribuzione casuale degli stessi, con media nulla e incorrelazione.

ES. STUDIO RELAZIONE alpha - beta

CALCOLIAMO IL COEFFICIENTE DI CORRELAZIONE LINEARE:

```
> R=cor(alpha, beta)
```

```
> R
```

```
[1] 0.8845007
```

C'E' UNA FORTE RELAZIONE LINEARE DIRETTA FRA LE DUE VARIABILI

CALCOLIAMO IL COEFFICIENTE DI DETERMINAZIONE FACENDO IL QUADRATO DI R PER GIUDICARE LA BONTA' DI ACCOSTAMENTO:

```
> R2=R^2
```

```
> R2
```

```
[1] 0.7823414
```

IL MODELLO TEORICO USATO SI ADATTA ABBASTANZA BENE AI VALORI OSSERVATI

ESERCIZIO 7

Si ipotizzi di estrarre una carta da un mazzo regolare di 52 e di voler verificare quante volte su 100 tentativi viene selezionata una figura di Fiori.

Utilizzare una opportuna variabile aleatoria e rappresentarla graficamente.

LA FUNZIONE `dbinom(k, n, p)`

CREO IL VETTORE DEI k

```
> k=c(0:100)
```

**# CALCOLO LE PROBABILITA' DELLA
BINOMIALE CON LA FUNZIONE `dbinom`**

```
> fiori=dbinom(k, 100, 13/52)
```

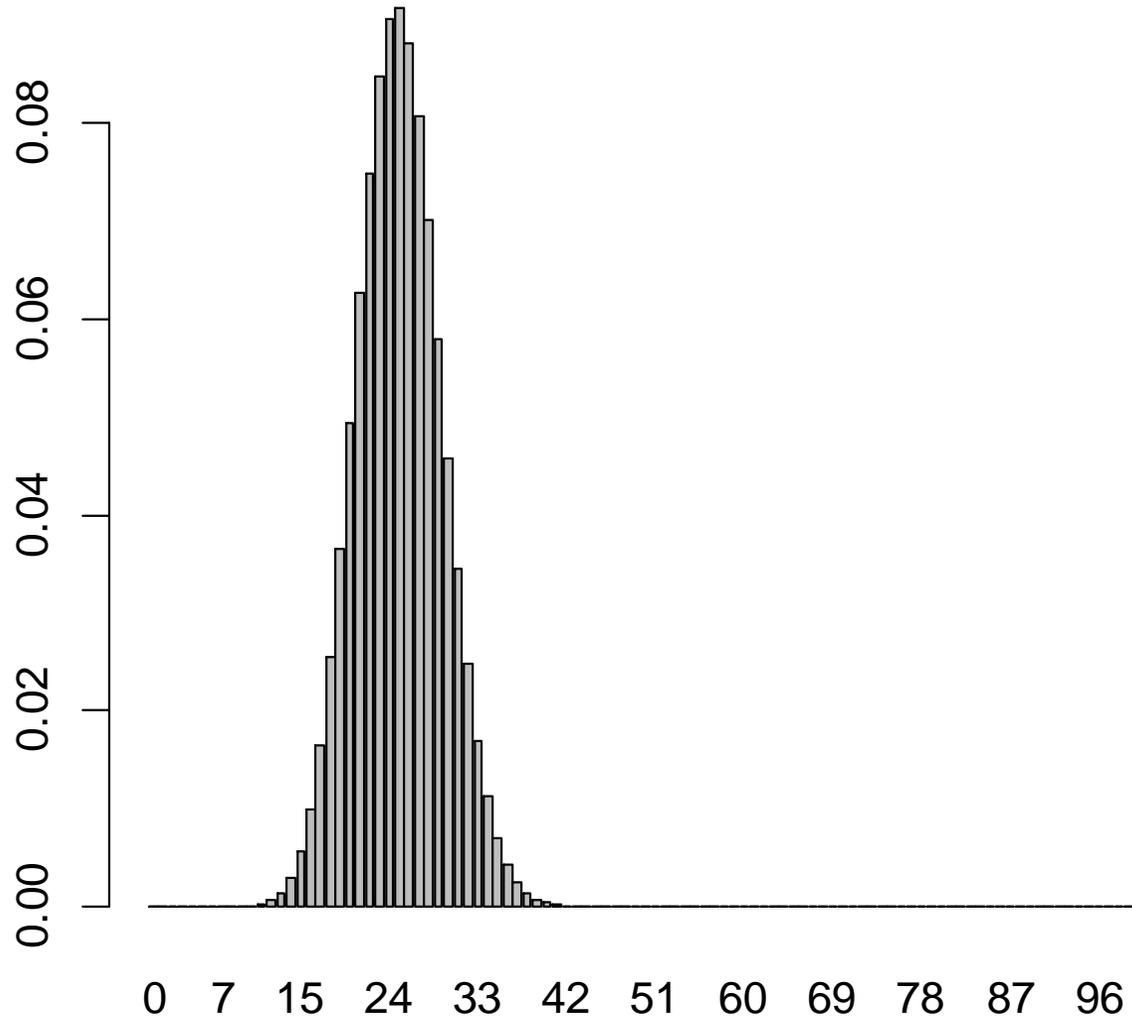
```
> fiori
```

```
[1] 3.207202e-13 1.069067e-11 1.763961e-10  
1.920758e-09 1.552613e-08
```

```
...
```

DISEGNO IL GRAFICO

```
> barplot(fiori, names.arg=k)
```



ESERCIZIO 7a

Sui dati dell'esercizio precedente, calcolare la probabilità di selezionare una carta di Cuori un numero di volte pari o inferiore a 25 su 100 lanci.

LA FUNZIONE pbinom

**# CALCOLO LA PROBABILITÀ DI OTTENERE
“CUORI” UN NUMERO DI VOLTE COMPRESO
FRA 0 E 25**

```
> cuori25p=pbinom(25, 100, 13/52)
```

```
> cuori25p
```

```
[1] 0.5534708
```

ESERCIZIO 7b

Sui dati dell'esercizio precedente, calcolare:

- ▶ La probabilità di ottenere “Quadri” un numero di volte maggiore di 40
- ▶ La probabilità di ottenere “Picche” un numero di volte compreso fra 20 e 30
- ▶ La probabilità di ottenere una carta appartenente ai segni rossi (Cuori o Quadri) 50 volte

ESERCIZIO 7b

**# CALCOLO LA PROBABILITÀ DI OTTENERE
“QUADRI” PIU’ DI 40 VOLTE**

```
> 1-pbinom(40, 100, 13/52)
```

```
[1] 0.0003239654
```

OPPURE:

```
pbinom(40, 100, 13/52, lower.tail=FALSE)
```

```
[1] 0.0003239654
```

ESERCIZIO 7b

**# CALCOLO LA PROBABILITÀ DI OTTENERE
“PICCHE” FRA 20 E 30 VOLTE**

> picche30p = pbinom(30, 100, 13/52)

> picche19p = pbinom(19, 100, 13/52)

> picche_da_20_a_30 = picche30p - picche19p

> picche_da_20_a_30

[1] 0.7966824

LA PROBABILITA' E' DEL 79,66824%

ESERCIZIO 7b

CALCOLO LA PROBABILITÀ DI OTTENERE UN
SEGNO “ROSSO” 50 VOLTE

```
> rosso50=dbinom(50, 100, 26/52)
```

```
[1] 0.07958924
```

ESERCIZIO 7c

Sulla distribuzione di probabilità precedente relativa all'estrazione di una carta di Fiori da un mazzo di 52 carte, calcolare:

- ▶ Il valore mediano
- ▶ Il primo e il terzo quartile
- ▶ Il valore corrispondente al 70% della distribuzione

ESERCIZIO 7c

CALCOLO IL VALORE MEDIANO

```
> fiori_mediana=qbinom(0.5, 100, 13/52)
```

```
> fiori
```

```
[1] 25
```

ESERCIZIO 7c

```
# CALCOLO IL PRIMO E IL TERZO QUARTILE
```

```
> fiori_1quart=qbinom(0.25, 100, 13/52)
```

```
> fiori_1quart
```

```
[1] 22
```

```
> fiori_3quart=qbinom(0.75, 100, 13/52)
```

```
> fiori_3quart
```

```
[1] 28
```

ESERCIZIO 7c

**# CALCOLO IL VALORE CORRISPONDENTE AL
70% DELLA DISTRIBUZIONE**

```
> fiori_70p=qbinom(0.70, 100, 13/52)
```

```
> fiori_70p
```

```
[1] 27
```

ESERCIZIO 8

Ipotizziamo di avere dei dati distribuiti come una normale con media 40 cm e deviazione standard 7 cm (si consiglia asse delle X da 0 a 90).

Disegnare il grafico e calcolare:

- ▶ probabilità $x=60$
- ▶ probabilità di $x \leq 55$
- ▶ probabilità di $x > 30$

ESERCIZIO 8

CREO INNANZITUTTO L'ASSE DELLE X

```
> x=seq(0, 90, 0.01)
```

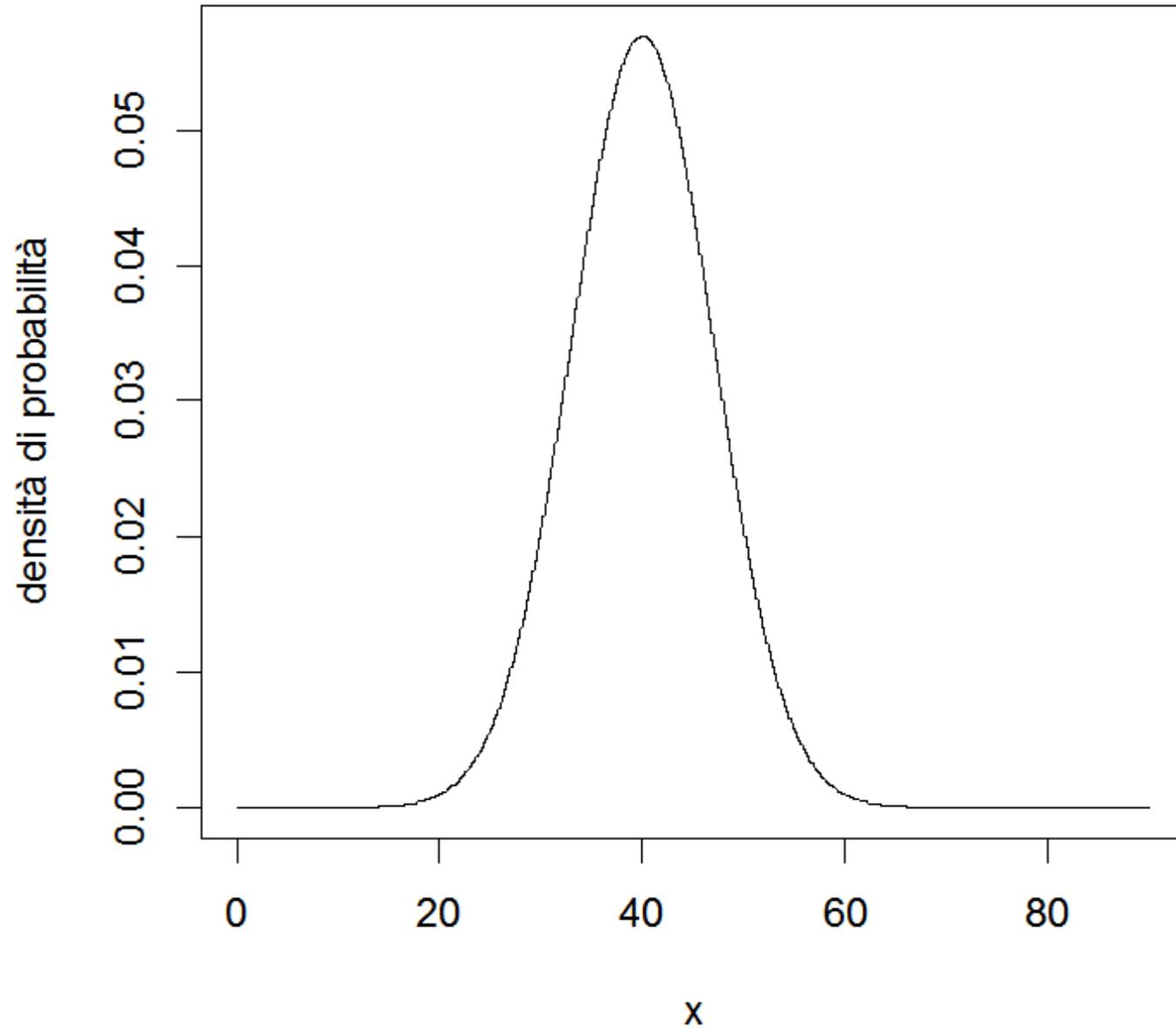
CREO LA DISTRIBUZIONE NORMALE

```
> normale=dnorm(x, 40, 7)
```

CREO IL GRAFICO

```
> plot(x, normale, type = "l", xlab="x", ylab =  
"densità di probabilità")
```

ESERCIZIO 8



ESERCIZIO 8

**# PER CONOSCERE LA PROBABILITA'
DI $x = 60$:**

```
> dnorm(60, 40, 7)
```

```
[1] 0.0009620142
```

ESERCIZIO 8

PER CONOSCERE LA PROBABILITA'
DI $x \leq 55$:

```
> pnorm(55, 40, 7)
```

```
[1] 0.9839377
```

ESERCIZIO 8

**# PER CONOSCERE LA PROBABILITA'
DI $x > 30$:**

```
> pnorm(30, 40, 7, lower.tail=FALSE)
```

```
[1] 0.9234363
```

ESERCIZIO 8a

Sui dati dell'esercizio precedente calcolare:

- ▶ probabilità fra 42 e 52
- ▶ il valore mediano
- ▶ il primo e il terzo quartile

ESERCIZIO 8a

PER CONOSCERE LA PROBABILITA' FRA 42 E 52 CM:

```
> pnorm(52, 40, 7, lower.tail=TRUE) -  
pnorm(42, 40, 7, lower.tail=TRUE)
```

```
[1] 0.3443103
```

ESERCIZIO 8a

LA MEDIANA E':

```
> qnorm(0.5, 40, 7)
```

```
[1] 40
```

IL PRIMO QUARTILE CORRISPONDE AL 25% DELLA DISTRIBUZIONE:

```
> qnorm(0.25, 40, 7)
```

```
[1] 35.27857
```

IL TERZO QUARTILE CORRISPONDE AL 75% DELLA DISTRIBUZIONE:

```
> qnorm(0.75, 40, 7)
```

```
[1] 44.72143
```

ESERCIZIO 9

IL NUMERO MEDIO DI ORE DEDICATE ALLO SPORT OGNI MESE RILEVATE PER UN ANNO SU UN CAMPIONE DI 10 PERSONE E' RISULTATO PARI AL SEGUENTE VETTORE:

$\text{sport} = c(12, 15, 9, 3, 10, 12, 8, 25, 7, 8)$

- 1) VERIFICARE L'IPOTESI CHE IL NUMERO MEDIO DI ORE DI SPORT SIA PARI A 11 (AL LIVELLO DI CONFIDENZA DEL 99%).
- 2) INDICARE ANCHE L'INTERVALLO DI CONFIDENZA PER LA MEDIA.

ESERCIZIO 9

```
> t.test(sport, mu=11, alternative="two.sided",  
conf.level=0.99)
```

One Sample t-test

data: sport

t = -0.0533, df = 9, p-value = 0.9587

alternative hypothesis: true mean is not equal to 11

99 percent confidence interval:

4.801806 16.998194

sample estimates:

mean of x

10.9

ESERCIZIO 9

1) POICHE' IL LIVELLO DI SIGNIFICATIVITA' (0.01) E' MINORE DEL P-VALUE CALCOLATO (0.9587) SI ACCETTA L'IPOTESI NULLA

2) L'INTERVALLO DI CONFIDENZA PER LA MEDIA E' COMPRESO FRA 4.801806 E 16.998194

ESERCIZIO 10

SI IPOTIZZI DI AVER RILEVATO IL N. DI ORE DI STRAORDINARIO MENSILI SVOLTE DA UN GRUPPO DI LAVORATORI ITALIANI E DA UN CORRISPONDENTE GRUPPO DI COLLEGHI CINESI E CHE LA VARIANZA DELLE DUE POPOLAZIONI SIA UGUALE. IL LIVELLO DI CONFIDENZA E' IL 95%. VERIFICARE L'IPOTESI CHE LE MEDIE SIANO UGUALI.

ITALIA	CINA
12	30
15	25
14	8
27	19
21	6

ESERCIZIO 10

```
straord.ita = c(12, 15, 14, 27, 21)
```

```
straord.cin = c(30, 25, 8, 19, 6)
```

```
t.test(straord.ita, straord.cin,  
var.equal=TRUE, conf.level=0.95)
```

ESERCIZIO 10

Two Sample t-test

data: straord.ita and straord.cin

$t = 0.0369$, $df = 8$, $p\text{-value} = 0.9715$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-12.30356 12.70356

sample estimates:

mean of x mean of y

17.8 17.6

POICHE' L'ALPHA TEORICO (OSSIA IL LIVELLO DI SIGNIFICATIVITA') E' 0.05 ED E' INFERIORE AL p-value CALCOLATO DI 0.9715, SI ACCETTA L'IPOTESI NULLA DI UGUAGLIANZA FRA LE MEDIE