



UNIVERSITA' DEGLI STUDI DI VERONA

LABORATORIO DI PROBABILITA' E STATISTICA

Docente: Bruno Gobbi

8 - ESERCIZI DI RIPASSO FINALE

1 - STATISTICA DESCRITTIVA - LAUREATI

ESERCIZIO 1: La seguente tabella riporta il numero di laureati per facoltà presso l'Università di Verona nel 2008.

Creare una tabella in R che riporti il numero di laureati e in percentuale. Creare un grafico a istogramma per il numero di laureati e uno a torta per le percentuali.

FACOLTA'	LAUREATI
Medicina	755
Economia	754
Lettere	584
Formazione	503
Scienze	263
Altri	842

1 - STATISTICA DESCRITTIVA - LAUREATI

```
> facolta=c("Medicina", "Economia", "Lettere", "Formazione",  
"Scienze", "Altri")
```

```
> nlaureati=c(755, 754, 584, 503, 263, 842)
```

```
> laureati=data.frame(facolta, nlaureati)
```

```
> laureati
```

	facolta	nlaureati
1	Medicina	755
2	Economia	754
3	Lettere	584
4	Formazione	503
5	Scienze	263
6	Altri	842

1 - STATISTICA DESCRITTIVA - LAUREATI

CREIAMO LA COLONNA DELLE PERCENTUALI

```
> tot_laureati=sum(nlaureati)
```

```
> tot_laureati
```

```
[1] 3701
```

```
> perc=nlaureati/tot_laureati
```

```
> perc
```

```
[1] 0.20399892 0.20372872 0.15779519 0.13590921 0.07106188  
0.22750608
```

1 - STATISTICA DESCRITTIVA - LAUREATI

AGGIUNGIAMO LA COLONNA DELLE PERCENTUALI

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

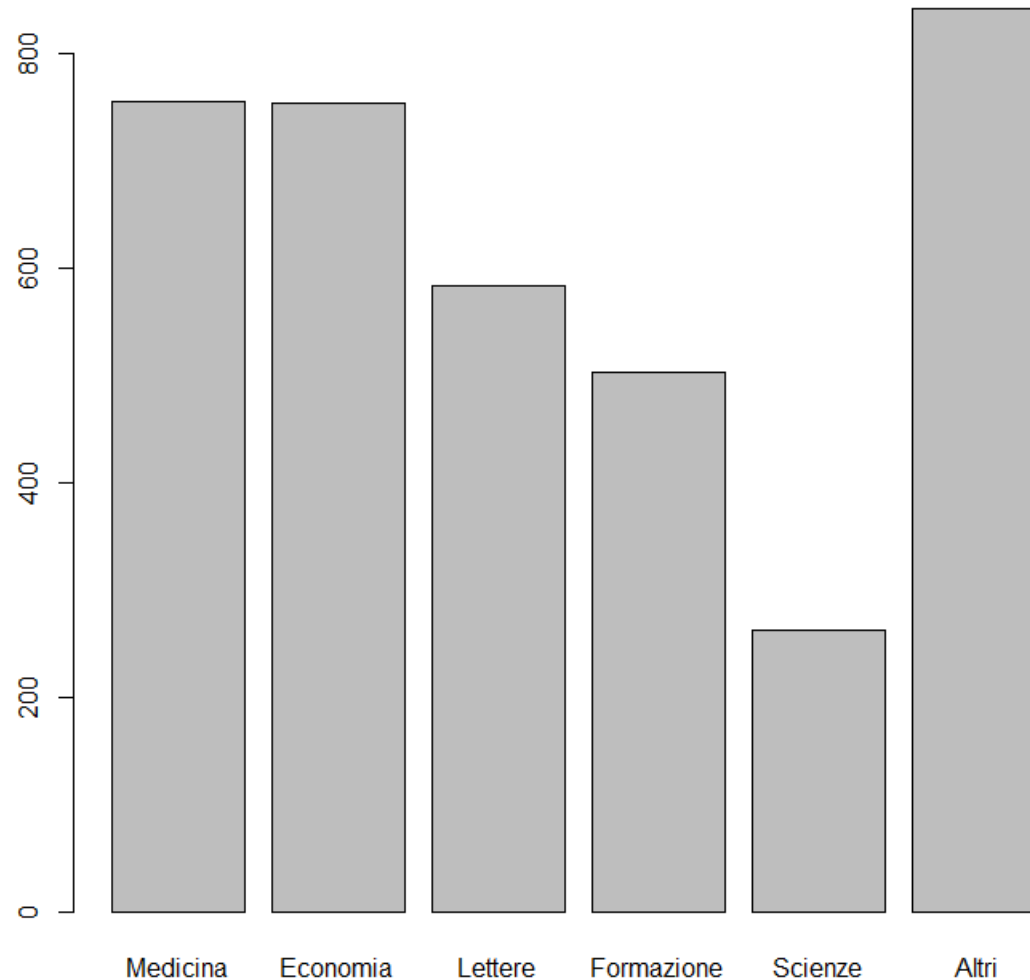
```
> laureati
```

	facolta	nlaureati	perc
1	Medicina	755	0.20399892
2	Economia	754	0.20372872
3	Lettere	584	0.15779519
4	Formazione	503	0.13590921
5	Scienze	263	0.07106188
6	Altri	842	0.22750608

1 - STATISTICA DESCRITTIVA - LAUREATI

GRAFICO DEL NUMERO DI LAUREATI PER FACOLTA'

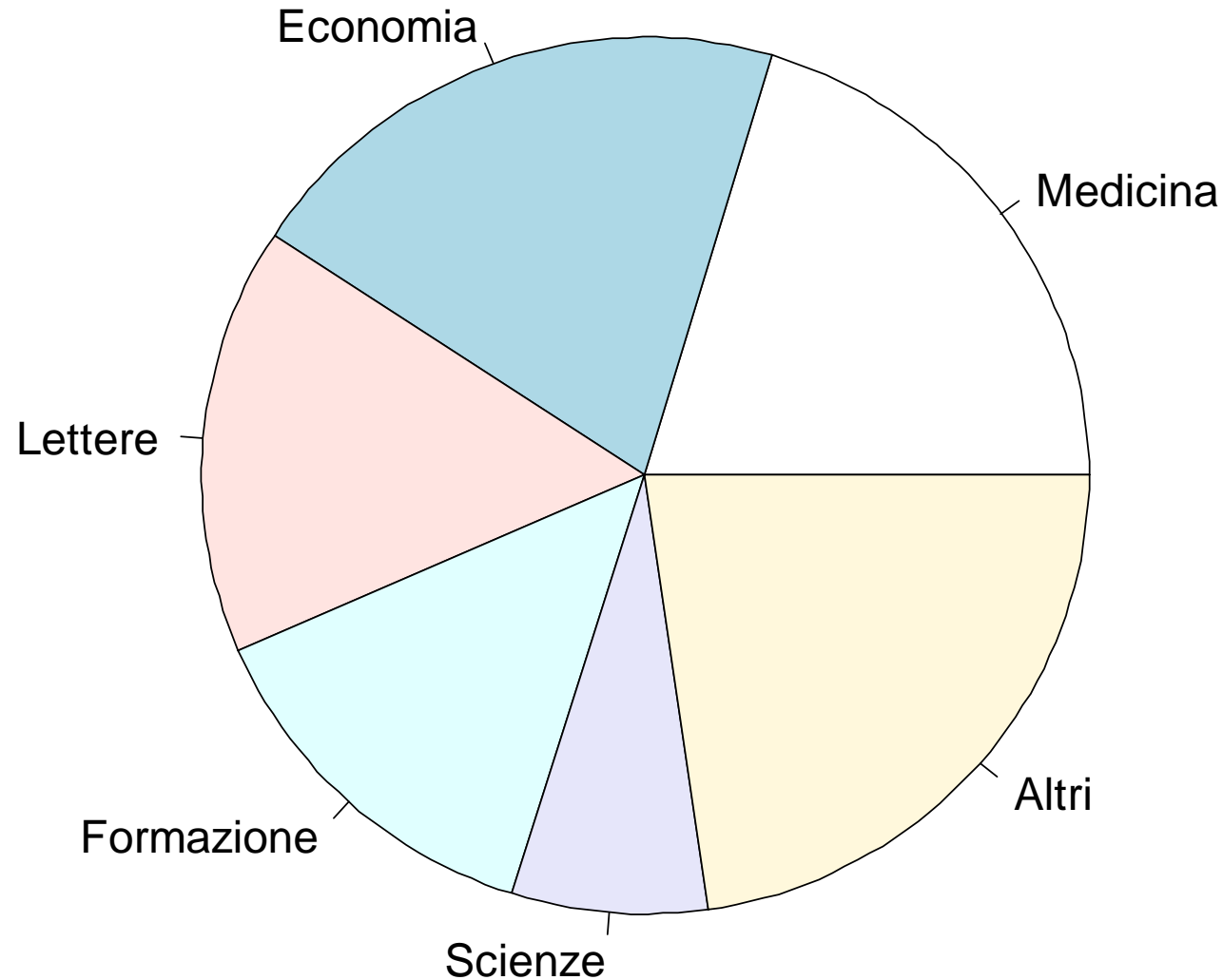
> barplot(nlaureati, names.arg=facolta)



1 - STATISTICA DESCRITTIVA - LAUREATI

GRAFICO A TORTA DELLE PERCENTUALI

> pie(perc, labels=facolta)



2 - CURTOSI E APPIATTIMENTO - laureati

ESERCIZIO 2: Sui seguenti dati calcolare:

- ▶ Media, mediana, primo e terzo quartile
- ▶ Il minimo e il massimo
- ▶ La varianza campionaria

Infine misurare la simmetria e l'appiattimento con degli opportuni indici

N.
50
27
35
42
22
61

2 - CURTOSI E APPIATTIMENTO - laureati

```
> x=c(50, 27, 35, 42, 22, 61)
```

```
# CALCOLO LE STATISTICHE RICHIESTE
```

```
> summary(x)
```

```
Min. 1st Qu. Median Mean 3rd Qu.
```

```
Max.
```

```
22.0 29.0 38.5 39.5 48.0 61.0
```

```
# CALCOLO LA VARIANZA CAMPIONARIA
```

```
> var(x)
```

```
[1] 212.3
```

2 - CURTOSI E APPIATTIMENTO - laureati

> gamma(x) = 0.1997853

IL VALORE DELL'INDICE GAMMA È PARI A 0.1997853. C'È UN'ASIMMETRIA POSITIVA, LA DISTRIBUZIONE PRESENTA UNA CODA PIÙ LUNGA A DESTRA.

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 - laureati

```
> beta(x)
```

```
[1] 1.273843
```

IL VALORE DELL'INDICE BETA E' PARI A
1.273843 LA DISTRIBUZIONE APPARE
SCHIACCIATA, PLATICURTICA

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  
}
```

3 - STATISTICHE E BOXPLOT - JOHNSON

ESERCIZIO 3: Utilizzando la base dati già presente in R relativamente ai dividendi trimestrali dell'azione della Johnson & Johnson, aerei fra il 1960 e il 1980 (nome del database: "JohnsonJohnson"), 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 - JOHNSON

```
> summary(JohnsonJohnson)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.440	1.248	3.510	4.800	7.132	16.200

```
> var(JohnsonJohnson)
```

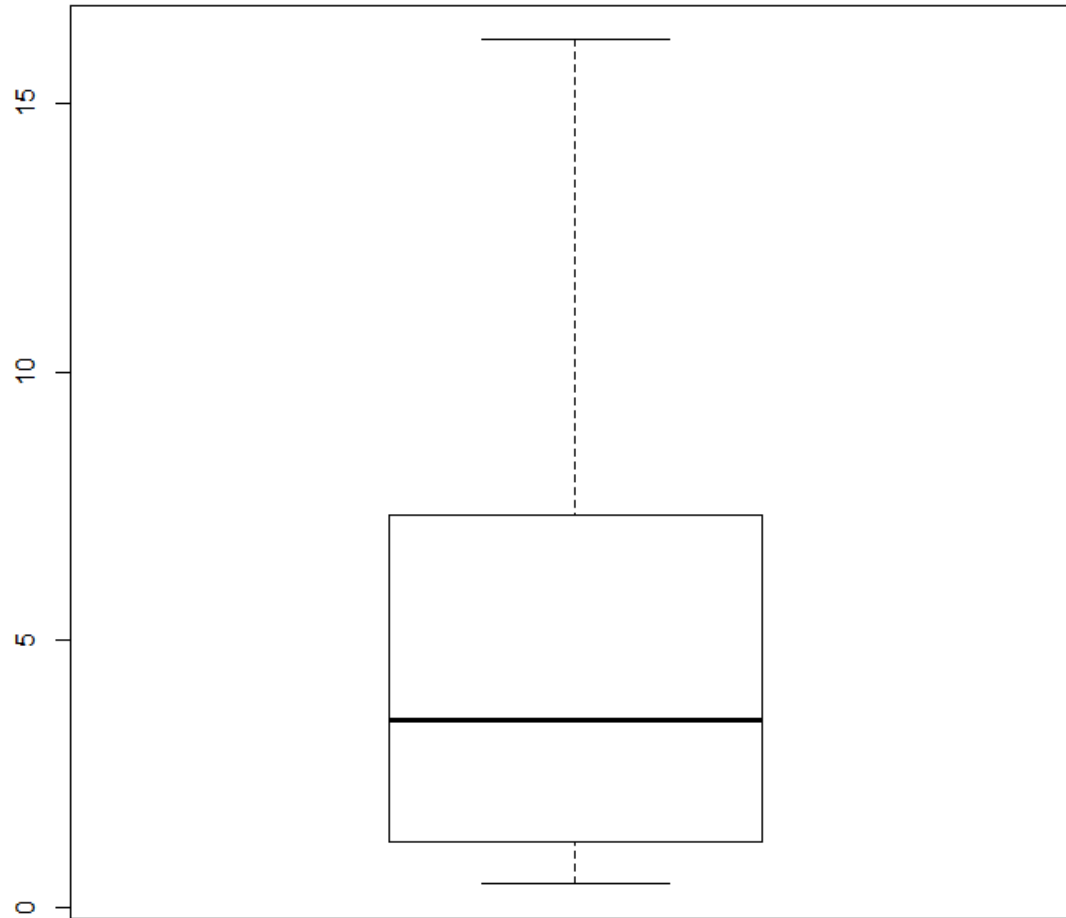
```
[1] 18.57602
```

```
> length(JohnsonJohnson)
```

```
[1] 84
```

3 - STATISTICHE E BOXPLOT - JOHNSON

> boxplot(JohnsonJohnson)



ES. REDDITO E CILINDRATA AUTO

ESERCIZIO 4: La seguente tabella riporta la distribuzione per classi di reddito e per cilindrata d'auto di un campione di 10.000 persone.

Disegnare il grafico a mosaico e valutare la connessione ad un livello di significatività dell'1%.

		Cilindrata auto			
		Fino a 1200	Da 1200 a 2000	Oltre 2000	TOTALE
Reddito	Sotto 15 mila euro	2.100	1.150	450	3.700
	Da 15 a 50 mila euro	1.400	3.300	1.100	5.800
	Oltre 50 mila euro	50	150	300	500
	TOTALE	3.550	4.600	1.850	10.000

g.d.l.	Livello di 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. REDDITO E CILINDRATA AUTO

```
> classi=matrix(c(2100, 1150, 450, 1400, 3300, 1100, 50, 150, 300), nrow=3, byrow=TRUE)
```

```
> classiReddito=c("Sotto 15 mila", "Da 15 a 50 mila", "Oltre 50 mila")
```

```
> classiCilindrata=c("Fino a 1200", "Da 1200 a 2000", "Oltre 2000")
```

```
> dimnames(classi)=list(classiReddito, classiCilindrata)
```

```
> classi
```

	Fino a 1200	Da 1200 a 2000	Oltre 2000
Sotto 15 mila	2100	1150	450
Da 15 a 50 mila	1400	3300	1100
Oltre 50 mila	50	150	300

```
> mosaicplot(classi)
```

ES. REDDITO E CILINDRATA AUTO



ES. REDDITO E CILINDRATA AUTO

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

```
> testchiq
```

Pearson's Chi-squared test

```
data: classi
```

```
X-squared = 1676.496, df = 4, p-value < 2.2e-16
```

**# POICHE' IL VALORE CALCOLATO DEL CHI-
QUADRATO E' 1676.496, 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**

g.d.l.	Livello di 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. REDDITO E CILINDRATA AUTO

CALCOLIAMO IL VALORE DELLA STATISTICA V DI CRAMER

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

```
> chiquadrato
```

```
X-squared
```

```
1676.496
```

```
> N = sum(classi)
```

```
> N
```

```
[1] 10000
```

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

```
> V
```

```
X-squared
```

```
0.2895251
```

**# IL V DI CRAMER E' PARI A 0,2895251 E QUESTO RISULTATO
PORTA AD AFFERMARE CHE C'È UNA BASSA CONNESSIONE
FRA I DUE FENOMENI**

REGRESSIONE LINEARE: GIRASOLI

ESERCIZIO 5:

I risultati dell'utilizzo di un fertilizzante su una coltura di girasoli ha portato alla crescita di piante come riportato in tabella.

Analizzare la relazione fra i fenomeni 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.

Mg di fertilizzante	Mm di crescita
52	1000
78	1545
64	1430
39	820
44	945
55	1100

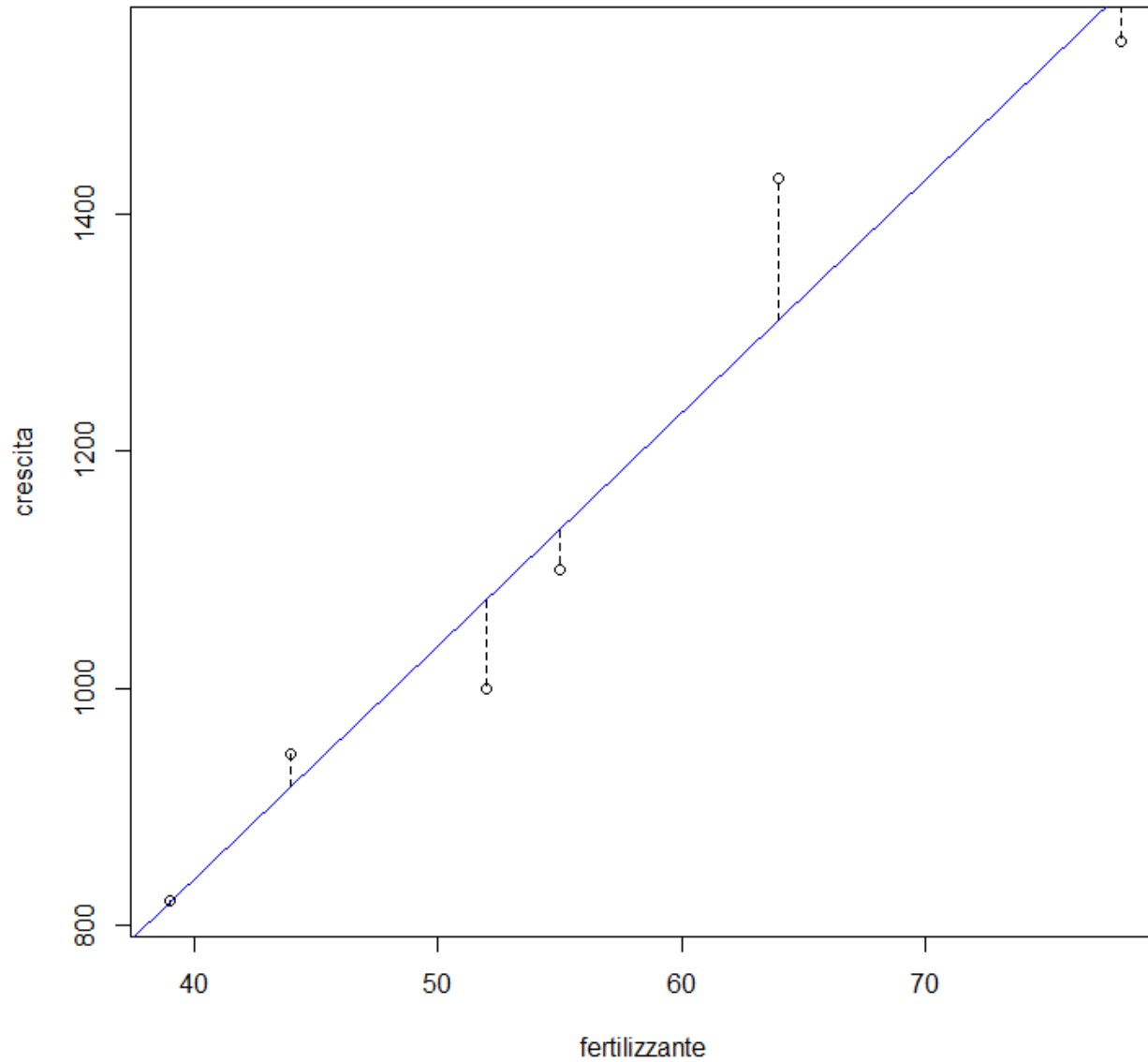
ES. STUDIO RELAZIONE FERTILIZZANTE-GIRASOLI

```
> fertilizzante=c(52, 78, 64, 39, 44, 55)
> crescita=c(1000, 1545, 1430, 820, 945, 1100)
> plot(fertilizzante, crescita)
> rettafertilizzante=lm(crescita~fertilizzante)
> abline(rettafertilizzante, col="blue")
> segments(fertilizzante, fitted(rettafertilizzante), fertilizzante,
crescita, lty=2)
> title(main="Regressione lineare fra fertilizzante e crescita")
```

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

ES. STUDIO RELAZIONE FERTILIZZANTE-GIRASOLI

Regressione lineare fra fertilizzante e crescita



ES. STUDIO RELAZIONE FERTILIZZANTE-GIRASOLI

> summary (rettafertilizzante)

Call:

lm(formula = crescita ~ fertilizzante)

Residuals:

1	2	3	4	5	6
-74.327	-41.577	119.250	1.798	28.289	-33.433

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	49.826	137.718	0.362	0.73581
fertilizzante	19.702	2.424	8.128	0.00125 **

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

Residual standard error: 76.48 on 4 degrees of freedom

Multiple R-squared: 0.9429, Adjusted R-squared: 0.9286

F-statistic: 66.06 on 1 and 4 DF, p-value: 0.001247

ES. STUDIO RELAZIONE FERTILIZZANTE-GIRASOLI

I PARAMETRI TROVATI SONO $a=49.826$ E $b=19.702$

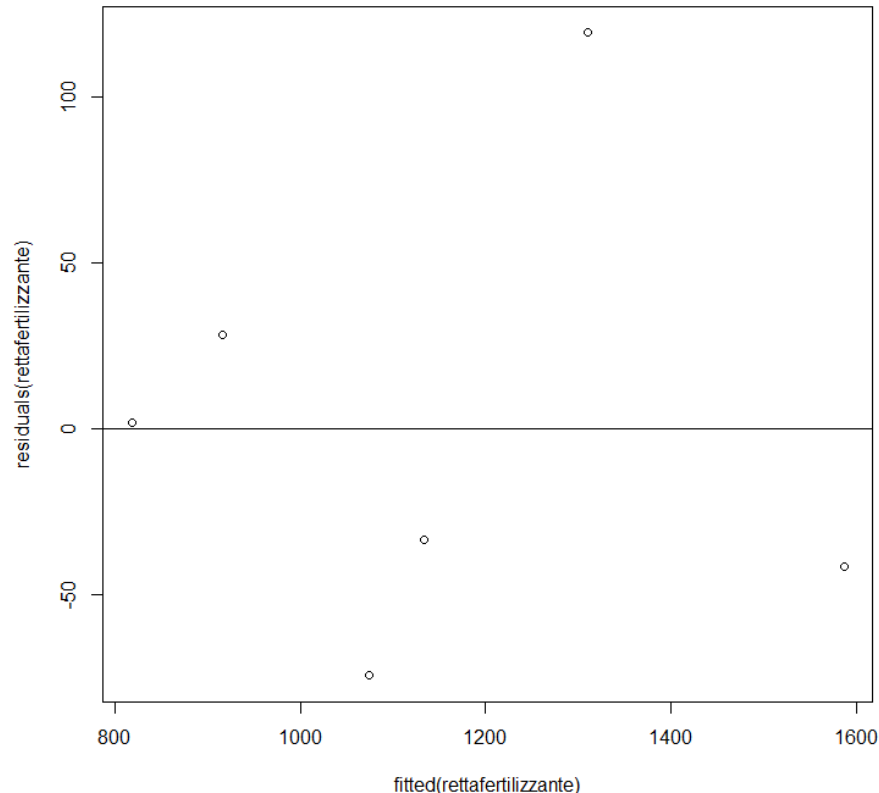
QUINDI IL MODELLO TEORICO SARA':

$Y' = 49.826 + 19.702 * \text{fertilizzante}$

EFFETTO L'ANALISI DEI RESIDUI

> `plot(fitted(rettafertilizzante), residuals(rettafertilizzante))`

> `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 FERTILIZZANTE-GIRASOLI

CALCOLO IL COEFFICIENTE DI CORRELAZIONE LINEARE:

> R=cor(fertilizzante, crescita)

> R

[1] 0.9710327

R E' PARI A 0,97103217 E CONFERMA CHE C'E' UNA FORTE RELAZIONE LINEARE DIRETTA FRA LE DUE VARIABILI

CALCOLO IL COEFFICIENTE DI DETERMINAZIONE:

> R2=R^2

> R2

[1] 0.9429044

R2 E' PARI A 0,9429044 QUINDI IL MODELLO TEORICO USATO SI ADATTA MOLTO BENE AI VALORI OSSERVATI

ESERCIZIO 6

Studiare la distribuzione di probabilità relativa ad un numero qualsiasi della roulette su 100 tentativi (tenendo conto che ci sono 37 possibili risultati).

Utilizzare una opportuna variabile aleatoria e rappresentarla graficamente.

ESERCIZIO 6

CREO IL VETTORE DEI k

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

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

```
> roulette=dbinom(k, 100, 1/37)
```

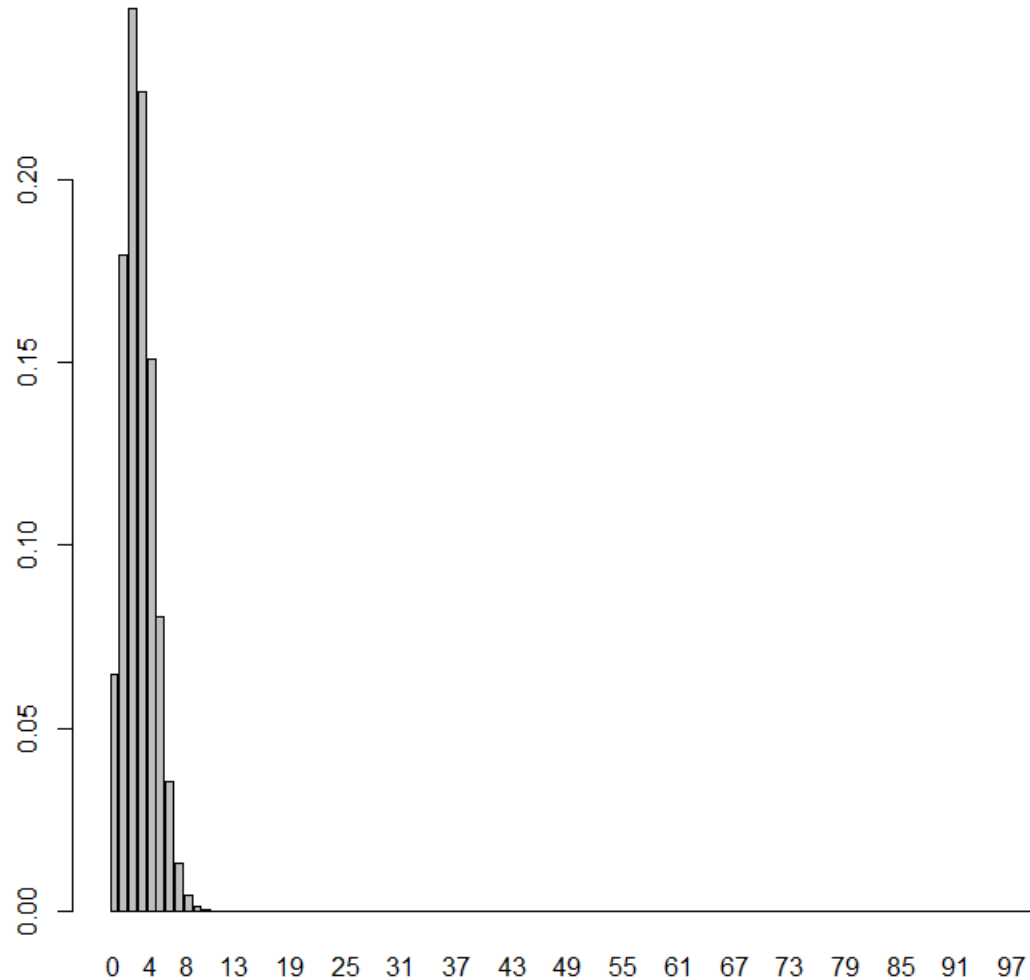
```
> roulette
```

```
[1] 6.457697e-02 1.793805e-01 2.466482e-01  
2.238104e-01 1.507611e-01 8.040594e-02
```

```
...
```

DISEGNO IL GRAFICO

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



ESERCIZIO 6a

Sui dati dell'esercizio precedente, calcolare la probabilità di ottenere 12 un numero pari o inferiore a 5 volte su 100 lanci.

ESERCIZIO 6a

**# CALCOLO LA PROBABILITÀ DI OTTENERE
12 UN NUMERO PARI O INFERIORE A 5 VOLTE
SU 100 LANCI**

```
> n12_5p=pbinom(5, 100, 1/37)
```

```
> n12_5p
```

```
[1] 0.945583
```


ESERCIZIO 6b

Sui dati dell'esercizio precedente, calcolare la probabilità di ottenere un numero fra 0 e 12 venti volte su 100 lanci.

ESERCIZIO 6b

**# CALCOLO LA PROBABILITÀ DI OTTENERE
UN NUMERO FRA 0 E 12 VENTI VOLTE SU 100
LANCI**

```
> n_0_12_20volte=dbinom(20, 100, 13/37)
```

```
> n_0_12_20volte
```

```
[1] 0.0004024809
```

ESERCIZIO 6c

Sui dati dell'esercizio precedente, descrivere la probabilità che la pallina si fermi su una casella colorata di rosso su 100 lanci.

ESERCIZIO 6c

CALCOLO LA PROBABILITÀ DI OTTENERE ROSSO SU 100 LANCI

```
> rosso=dbinom(k, 100, 18/37)
```

```
> [1] 1.135501e-29 1.075738e-27 5.044646e-26  
1.561185e-24 3.586618e-23 6.523869e-22  
9.785803e-21
```

```
...
```

ESERCIZIO 6d

Sui dati dell'esercizio precedente, descrivere la probabilità che la pallina si fermi su un numero pari su 100 lanci.

ESERCIZIO 6d

**# CALCOLO LA PROBABILITÀ DI OTTENERE UN
NUMERO PARI SU 100 LANCI**

```
> pari=dbinom(k, 100, 18/37)
```

```
> [1] 1.135501e-29 1.075738e-27 5.044646e-26  
1.561185e-24 3.586618e-23 6.523869e-22  
9.785803e-21
```

```
...
```

ESERCIZIO 6e

Sui dati dell'esercizio precedente, descrivere la probabilità che la pallina si fermi sullo zero su 100 lanci.

ESERCIZIO 6e

CALCOLO LA PROBABILITÀ DI OTTENERE ZERO SU 100 LANCI

```
> zero=dbinom(k, 100, 1/37)
```

```
[1] 6.457697e-02 1.793805e-01 2.466482e-01  
2.238104e-01 1.507611e-01 8.040594e-02
```

```
...
```


ESERCIZIO 7

La produzione di una nuova APU prevede che i macchinari preposti producano ogni 100.000 unità due pezzi difettosi ($\lambda=2$).

Descrivere con una opportuna variabile aleatoria la probabilità di avere un numero di pezzi difettosi compreso da 0 a 10 e rappresentarla graficamente.

LA FUNZIONE $\text{dpois}(k, \lambda)$

CREO IL VETTORE DEI k

```
> k=c(0:10)
```

CALCOLO LE PROBABILITA' DELLA POISSON CON LA FUNZIONE dpois

```
> poisson=dpois(k, 2)
```

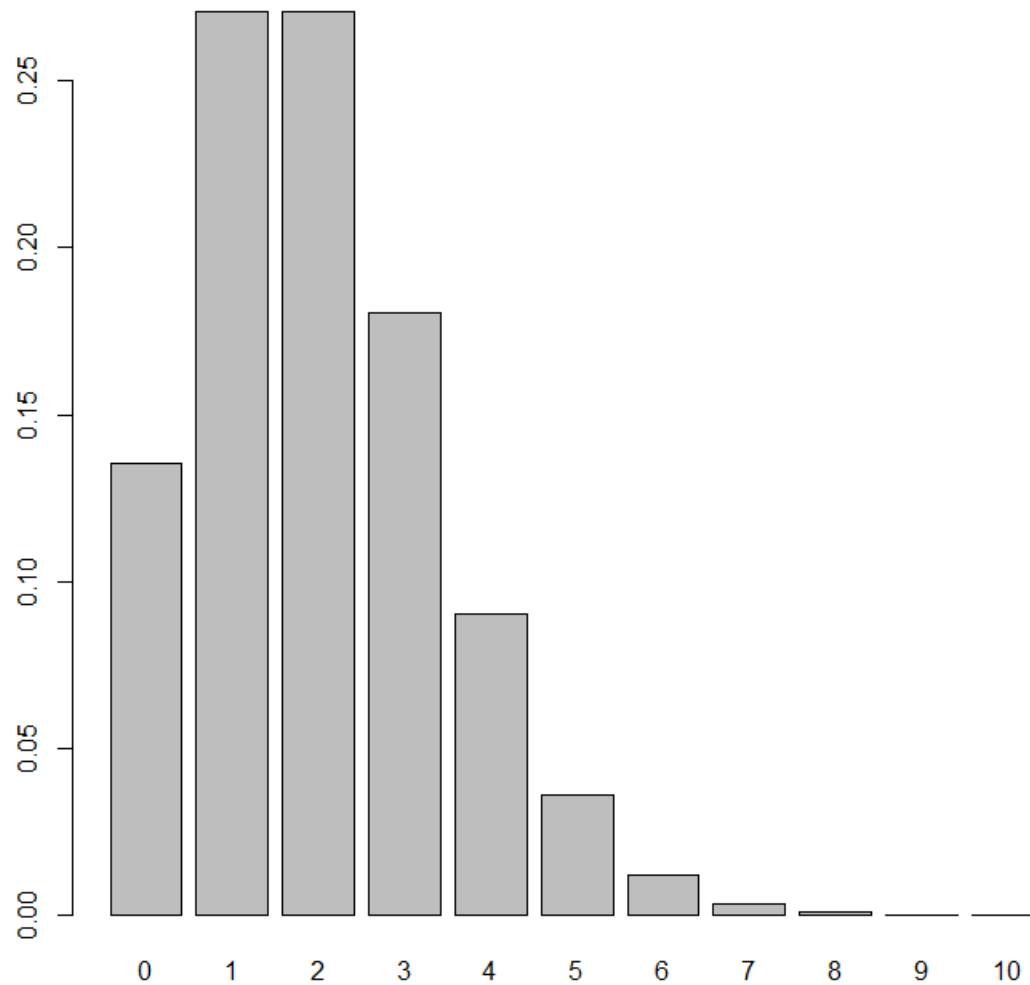
```
> poisson
```

```
[1] 1.353353e-01 2.706706e-01 2.706706e-01  
1.804470e-01 9.022352e-02 3.608941e-02  
1.202980e-02
```

```
[8] 3.437087e-03 8.592716e-04 1.909493e-04  
3.818985e-05
```

DISEGNO IL GRAFICO

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



ESERCIZIO 7a

Sui dati dell'esercizio precedente calcolare:

- ▶ La probabilità di $k \leq 3$
- ▶ La probabilità di $k > 3$

ESERCIZIO 7a

CALCOLO LA PROBABILITA' DI $k \leq 3$:

```
> ppois(3, 2)
```

```
[1] 0.8571235
```

ESERCIZIO 7a

CALCOLO LA PROBABILITA' DI $k > 3$:

```
> 1-ppois(3, 2)
```

```
[1] 0.1428765
```

OPPURE:

```
> ppois(3, 2, lower.tail=FALSE)
```

```
[1] 0.1428765
```

ESERCIZIO 7b

Sui dati dell'esercizio precedente calcolare:

- ▶ Il valore mediano
- ▶ Il valore corrispondente al 75% della distribuzione

ESERCIZIO 7b

CALCOLO IL VALORE MEDIANO:

```
> qpois(0.5, 2)
```

```
[1] 2
```

CALCOLO IL VALORE CORRISPONDENTE AL 75% DELLA DISTRIBUZIONE:

```
> qpois(0.75, 2)
```

```
[1] 3
```


ESERCIZIO 8

Ipotizziamo di avere dei dati distribuiti come una normale con media 300 e deviazione standard 55 (si consiglia asse delle X da 0 a 600).

Disegnare il grafico e calcolare:

- ▶ probabilità $x=400$
- ▶ probabilità di $x \leq 200$
- ▶ probabilità di $x > 500$

ESERCIZIO 8

CREO INNANZITUTTO L'ASSE DELLE X

```
> x=seq(0, 600, 0.01)
```

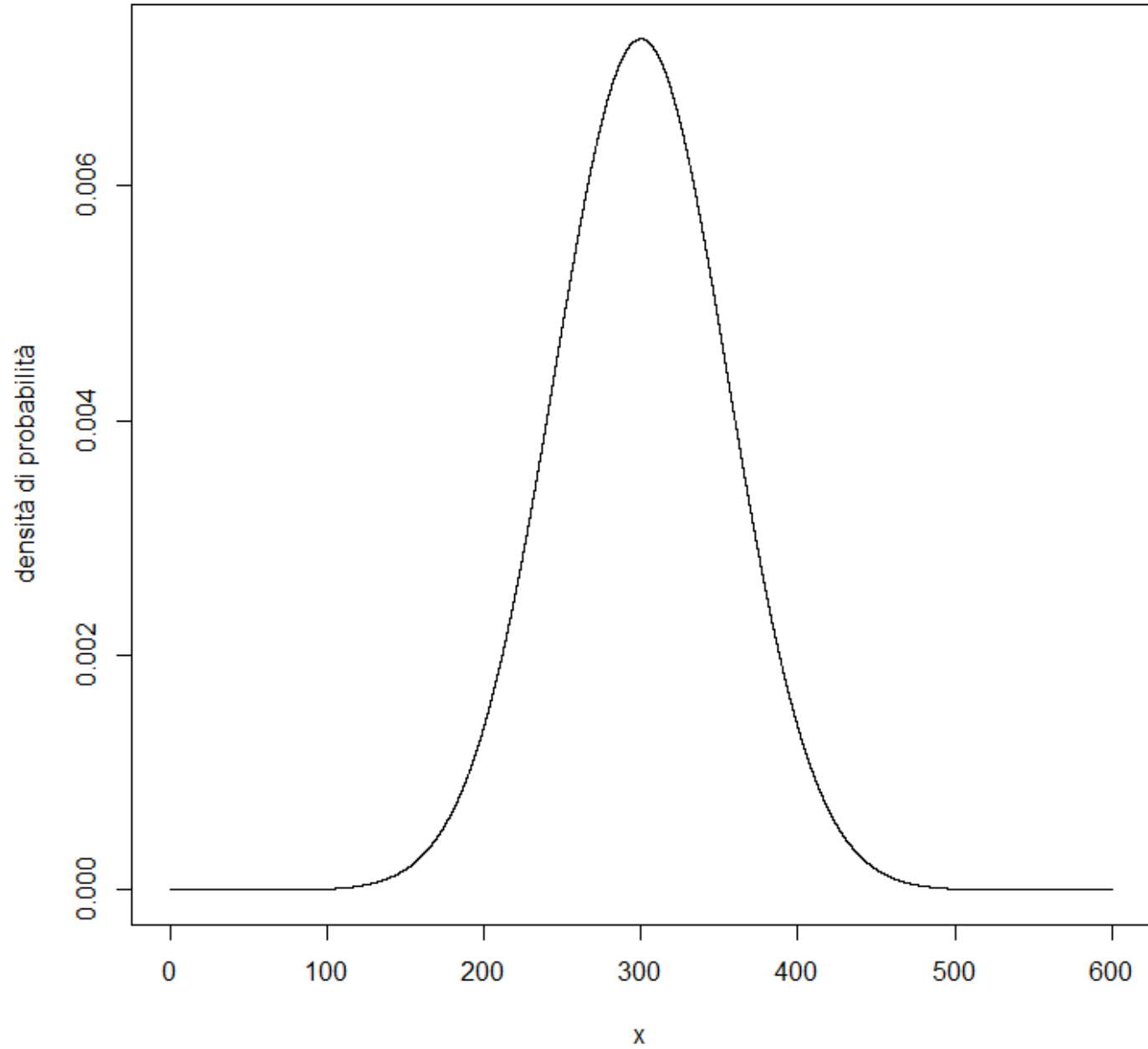
CREO LA DISTRIBUZIONE NORMALE

```
> normale=dnorm(x, 300, 55)
```

CREO IL GRAFICO

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

ESERCIZIO 8



ESERCIZIO 8

PER CONOSCERE LA PROBABILITA'
DI $x = 400$:

```
> dnorm(400, 300, 55)
```

```
[1] 0.00138901
```

ESERCIZIO 8

PER CONOSCERE LA PROBABILITA'
DI $x \leq 200$:

```
> pnorm(200, 300, 55)
```

```
[1] 0.03451817
```

ESERCIZIO 8

**# PER CONOSCERE LA PROBABILITA'
DI $x > 500$:**

```
> pnorm(500, 300, 55, lower.tail=FALSE)  
[1] 0.000138257
```

ESERCIZIO 8a

Sui dati dell'esercizio precedente calcolare:

- ▶ probabilità fra 315 e 520
- ▶ il valore mediano
- ▶ il primo e il terzo quartile

ESERCIZIO 8a

PER CONOSCERE LA PROBABILITA' FRA 315
E 520 CM:

```
> pnorm(520, 300, 55, lower.tail=TRUE) -  
pnorm(315, 300, 55, lower.tail=TRUE)
```

```
[1] 0.3924998
```


ESERCIZIO 8a

LA MEDIANA E':

```
> qnorm(0.5, 300, 55)
```

```
[1] 300
```

IL PRIMO QUARTILE CORRISPONDE AL 25% DELLA
DISTRIBUZIONE:

```
> qnorm(0.25, 300, 55)
```

```
[1] 262.9031
```

IL TERZO QUARTILE CORRISPONDE AL 75% DELLA
DISTRIBUZIONE:

```
> qnorm(0.75, 300, 55)
```

```
[1] 337.0969
```

ESERCIZIO 9

IL NUMERO MEDIO DI KG DI FRUTTA CONSUMATI OGNI ANNO RILEVATE SU UN CAMPIONE DI 10 PERSONE E' RISULTATO PARI AL SEGUENTE VETTORE:

frutta=c(3, 1, 0.5, 2, 1.5, 1, 1.5, 4, 0.5, 2)

- 1) VERIFICARE L'IPOTESI CHE IL NUMERO MEDIO DI KG DI FRUTTA CONSUMATI SIA PARI A 1,5 (AL LIVELLO DI CONFIDENZA DEL 99%).
- 2) INDICARE ANCHE L'INTERVALLO DI CONFIDENZA PER LA MEDIA.

ESERCIZIO 9

```
> t.test(frutta, mu=1.5, alternative="two.sided",  
conf.level=0.99)
```

One Sample t-test

data: frutta

t = 0.5695, df = 9, p-value = 0.583

alternative hypothesis: true mean is not equal to 1.5

99 percent confidence interval:

0.5586953 2.8413047

sample estimates:

mean of x

1.7

ESERCIZIO 9

1) POICHE' IL LIVELLO DI SIGNIFICATIVITA' (0.01) E' MINORE DEL P-VALUE CALCOLATO (0.583) SI ACCETTA L'IPOTESI NULLA

2) L'INTERVALLO DI CONFIDENZA PER LA MEDIA E' COMPRESO FRA 0.5586953 E 2.8413047

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