INTRODUZIONE ALLA STATISTICA APPLICATA con esempi in R

http://hpe.pearsoned.it/stefanini

Soluzioni degli esercizi di ricapitolazione Capitolo 4: "Il test delle ipotesi"

F. Frascati F. M. Stefanini

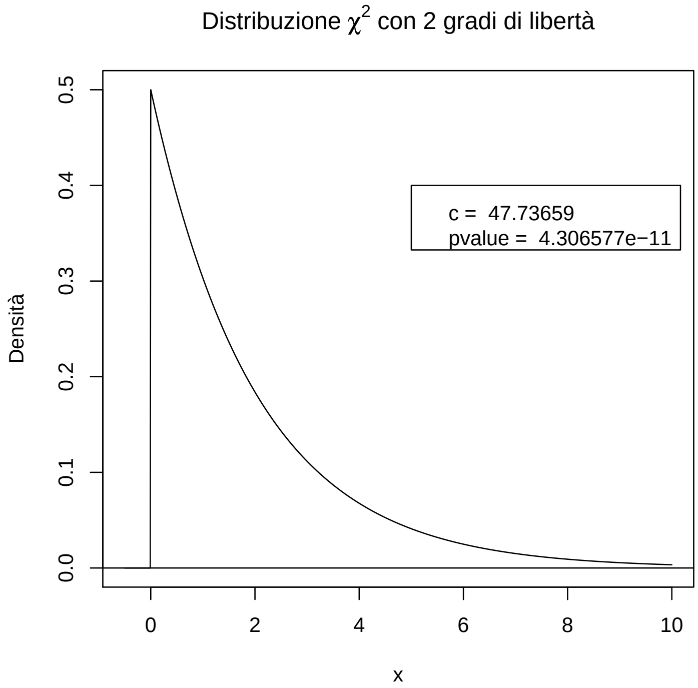
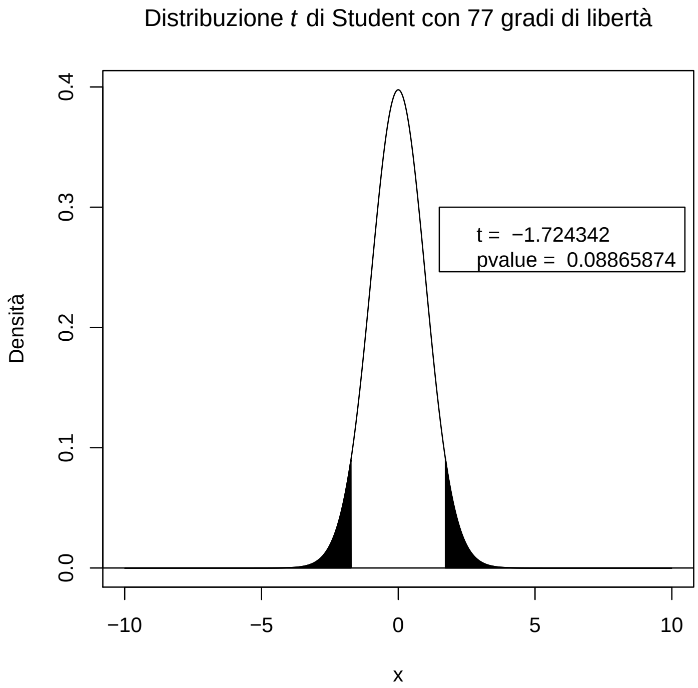
11 gennaio 2008

Esercizio 4.4.1

1) ```
   Y      Freq
   0 0.3122924
   1 0.6877076
   ```

   La percentuale dei capi malati ammonta al 68.771 %.

2) L'ipotesi da testare è l'associazione tra malattia ed età.

   ```
   Statistica test:  47.73659

   pvalue:  4.306577e-11

   Valore critico con ampiezza 0.05 (monolaterale destro):  5.991465

   gradi di libertà:  2

   -> Rifiutare H0 con ampiezza 0.05

   Frequenze osservate:

         X = 0 X = 1 X = 2
   Y = 0     9    38    47
   Y = 1     9    19   179

   Frequenze attese:

             X = 0    X = 1     X = 2
   Y = 0  5.621262 17.80066  70.57807
   Y = 1 12.378738 39.19934 155.42193

   Residui di Pearson:

              X = 0     X = 1     X = 2
   Y = 0  1.4250744  4.787612 -2.806554
   Y = 1 -0.9603207 -3.226248  1.891264
   ```
   ```



Esercizio 4.4.2

1) `H0: mu1 = mu2`

   `H1: mu1 != mu2`

2) `Statistica test:  -1.724342`

   `Valore critico con ampiezza 0.01 (bilaterale): -/+ 2.641198`

   `gradi di libertà:  77`

   `Spooled:  0.3523649`

   `-> Non rifiutare H0 con ampiezza 0.01`

3) `pvalue:  0.08865874`



Esercizio 4.4.3

1) `Statistica test:  17.63636`

```
Valore critico con ampiezza 0.1 (monolaterale destro):  4.60517

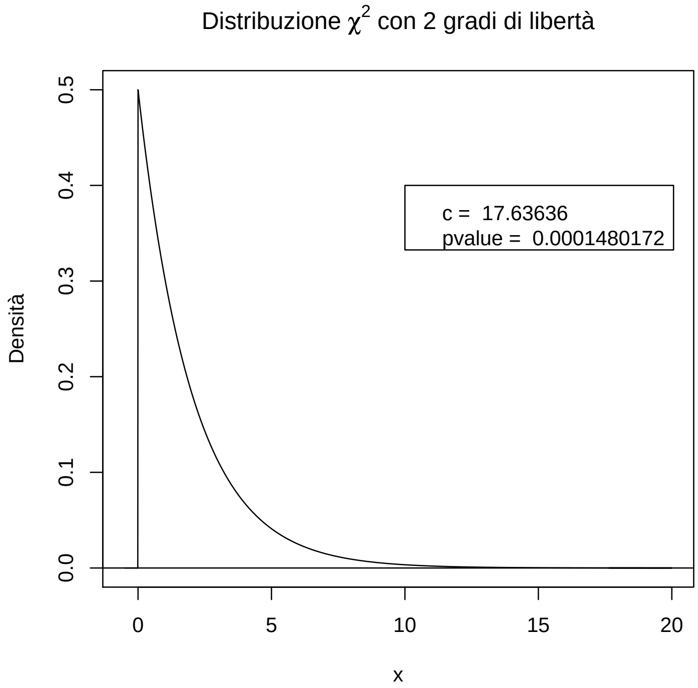
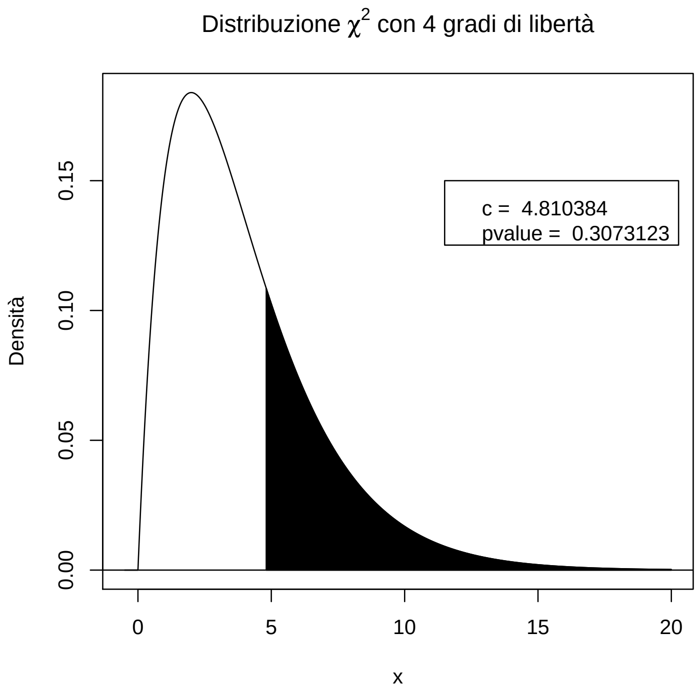
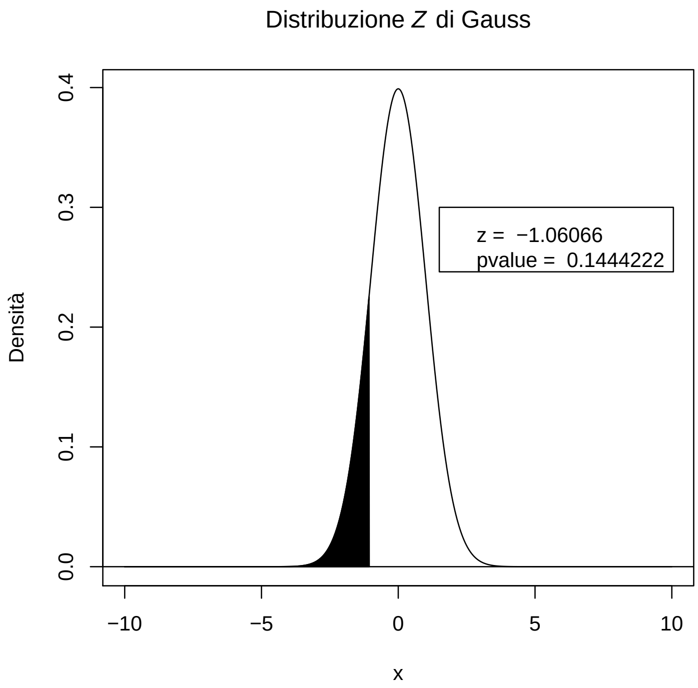
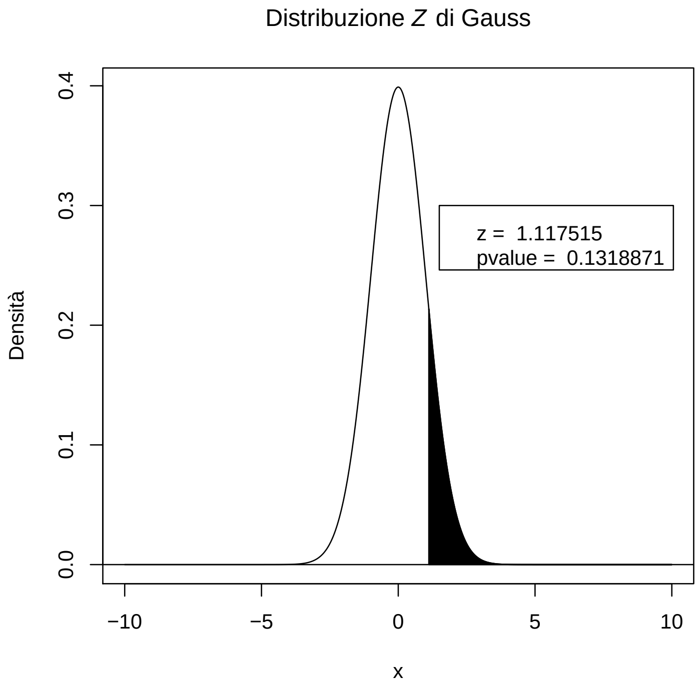
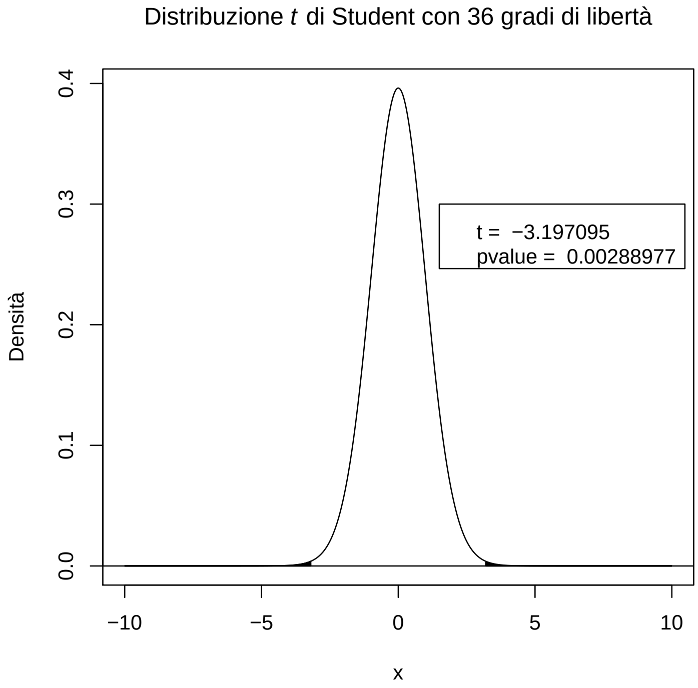
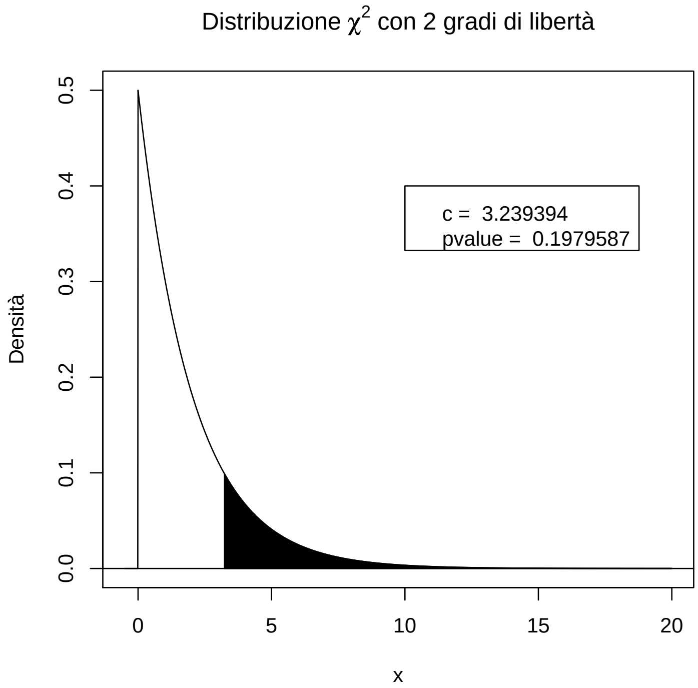
gradi di libertà:  2

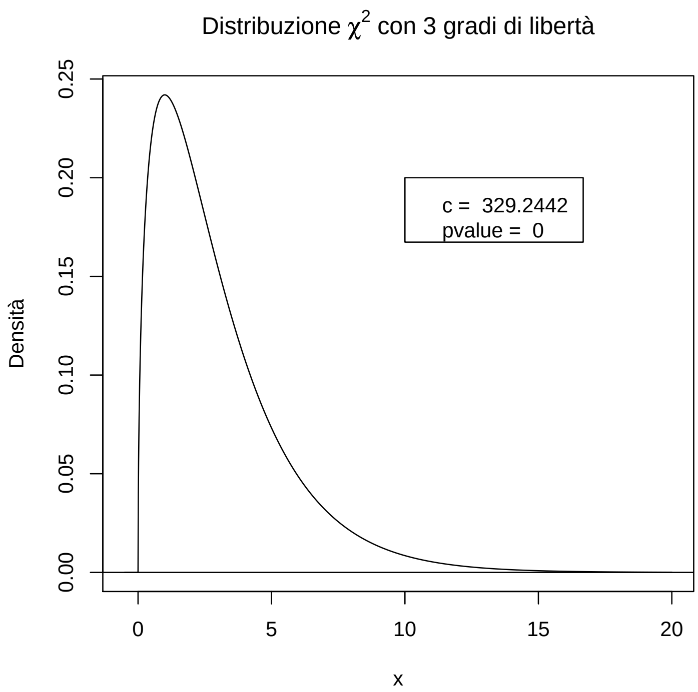
-> Rifiutare H0 con ampiezza 0.1

Frequenze osservate:

 No Imperfezioni Lievi imperfezioni Gravi imperfezioni
             54                 29                  5

Frequenze attese:

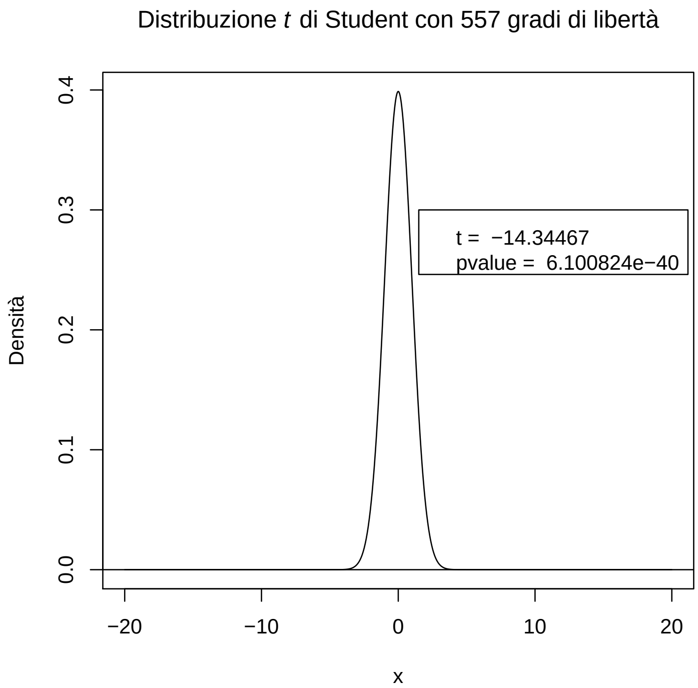
 No Imperfezioni Lievi imperfezioni Gravi imperfezioni
             44                 22                 22
```

2) `pvalue:  0.0001480172`



Esercizio 4.4.4

1) Statistica test: 4.810384

   pvalue: 0.3073123

   Valore critico con ampiezza 0.01 (monolaterale destro): 13.27670

   gradi di libertà: 4

   -> Non rifiutare H0 con ampiezza 0.01

   Frequenze osservate:

	Brembana	Seriana	Torta
bassa	57	56	81
media	84	133	141
alta	50	66	86

   Frequenze attese:

	Brembana	Seriana	Torta
bassa	49.14324	65.61008	79.24668
media	90.68700	121.07427	146.23873
alta	51.16976	68.31565	82.51459

Residui di Pearson:

	Brembana	Seriana	Torta
bassa	1.1207580	-1.1864289	0.1969562
media	-0.7021970	1.0838246	-0.4332061
alta	-0.1635274	-0.2801643	0.3836972



Esercizio 4.4.5

1) H0: mu1 = mu2

   H1: mu1 < mu2

2) `Statistica test:  -1.060660`

`pvalue:  0.1444222`

`Valore critico con ampiezza 0.1 (monolaterale sinistro):  -1.281552`

`-> Non rifiutare H0 con ampiezza 0.1`



Esercizio 4.4.6

1) `1-alpha:  0.91`

`Intervallo di confidenza: ( 154.3447 , 158.1895 )`

2) `H0: mu = 155`
`H1: mu > 155`

`Statistica test:  1.117515`

Valore critico con ampiezza 0.09 (monolaterale destro): 1.695398

-> Non rifiutare H0 con ampiezza 0.09

3) pvalue: 0.1318871



Esercizio 4.4.7

1) H0: mu1 = mu2
   H1: mu1 != mu2

   Statistica test: -3.197095

   Valore critico con ampiezza 0.1 (bilaterale): -/+ 1.688298

   gradi di libertà: 36

   -> Rifiutare H0 con ampiezza 0.1

2) pvalue: 0.002889770



Esercizio 4.4.8

1) Statistica test: 3.239394

Valore critico con ampiezza 0.05 (monolaterale destro): 5.991465

gradi di libertà: 2

-> Non rifiutare H0 con ampiezza 0.05

Frequenze osservate:

Bassa qualità	Media qualità	Alta qualità
27	58	25

Frequenze attese:

```
Bassa qualità Media qualità Alta qualità
           22            55           33
```

2) `pvalue:  0.1979587`



Esercizio 4.4.9

1) `Statistica test:  329.2442`

`pvalue:  0`

`Valore critico con ampiezza 0.05 (monolaterale destro):  7.814728`

`gradi di libertà:  3`

`-> Rifiutare H0 con ampiezza 0.05`

`Frequenze osservate:`

```
 X = 0 X = 1 X = 2 X = 3
    12    14    94     0
```

```
Frequenze attese:

    X = 0    X = 1    X = 2   X = 3
 48.30291 43.95565 19.99982 7.74163
```



Esercizio 4.4.10

1)
```
H0: mu1 = mu2
H1: mu1 != mu2

Statistica test:  -14.34467

Valore critico con ampiezza 0.01 (bilaterale): -/+ 2.584685

gradi di libertà:  557
```

```
Spooled:  4.018948

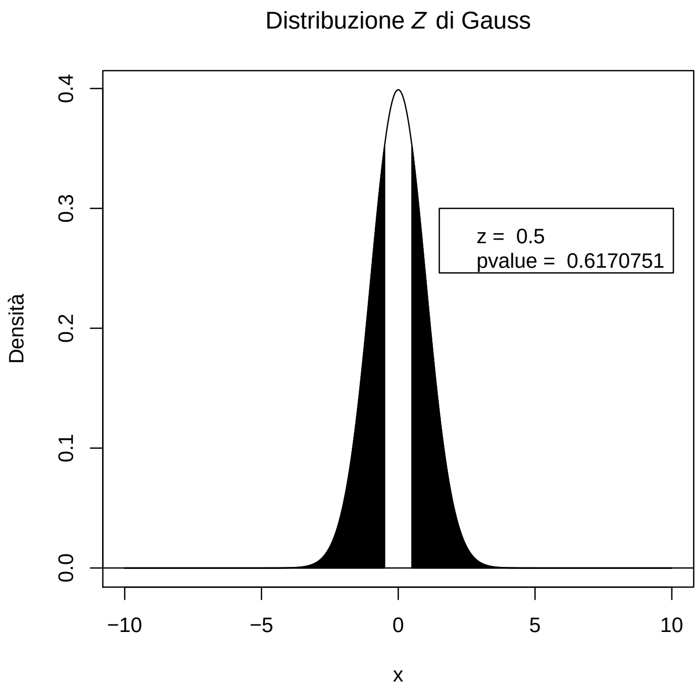
-> Rifiutare H0 con ampiezza 0.01
```

2) `pvalue:  6.100824e-40`



Esercizio 4.4.11

1) ```
H0: p = 0.2
H1: p != 0.2

Statistica test: 0.5

pvalue: 0.6170751

Valore critico con ampiezza 0.1 (bilaterale): -/+ 1.644854

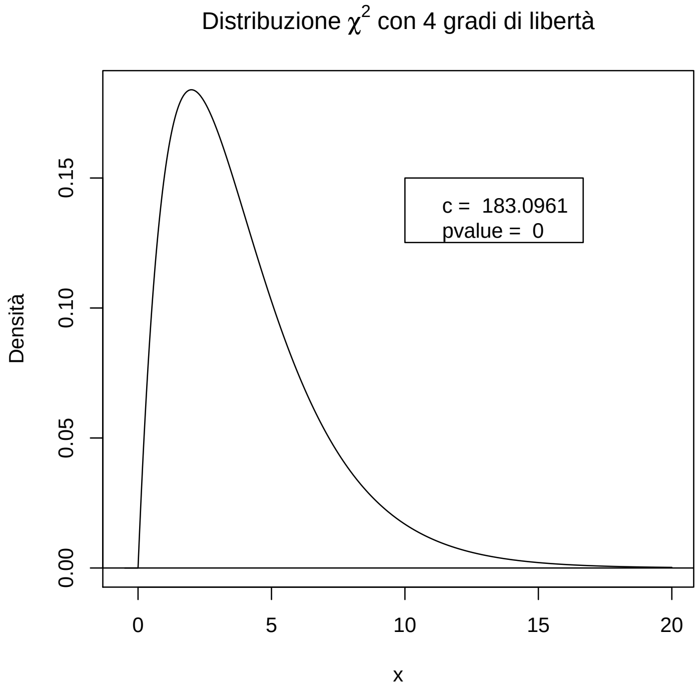
-> Non rifiutare H0 con ampiezza 0.1
```
```



Esercizio 4.4.12

1) Statistica test: 183.0961

```
pvalue:  0

Valore critico con ampiezza 0.01 (monolaterale destro):  13.27670

gradi di libertà:  4

-> Rifiutare H0 con ampiezza 0.01

Frequenze osservate:

 (136,146] (146,161] (161,177] (177,193] (193,208] (208,224]
        11        29         9        21        80        25

Frequenze attese:

 (136,146] (146,161] (161,177] (177,193] (193,208] (208,224]
  16.02883  34.49198  46.05131  37.61902  23.06671  17.74215
```



Esercizio 4.4.13

1) Abbinando ad ogni trattamento un livello del fattore B si ottiene la tabella ANOVA:

```
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value Pr(>F)
code       5 151.964  30.393  1.5032 0.2602
Residuals 12 242.620  20.218
```

2) Una soluzione si può ottenere ricorrendo ad opportuni contrasti per le medie di trattamento. Per saggiare l'effetto dell'esposizione al sole il vettore dei coefficienti vale:

```
-1
 1
-1
```

```
  1
 -1
  1
```

Il vettore delle medie vale:

```
[1] 33.30000 32.30000 36.83333 36.00000 39.63333 40.06667
```

Da cui il valore del contrasto -1.4. La SS dovuta al contrasto vale 0.98 che diviso per $QM(e)$ porta ad un valore empirico di $F = 0.07185$. Per saggiare gli effetti dovuti all'altitudine si componga la matrice dei contrasti di nome `matContra`:

```
  2  0
  2  0
 -1  1
 -1  1
 -1 -1
 -1 -1
```

che sono ortogonali, infatti:

```
t(matContra) %*% matContra
```

```
 12 0
  0 4
```

I valori stimati delle due componenti del contrasto sono:

```
 -21.333333
  -6.866667
```

Pertanto la devianza associata al contrasto vale 149.14111 e $QM(e)$ del contrasto vale 74.57056 a cui segue un valore empirico di F pari a 5.46745. Un contrasto che saggia l'interazione si ottiene con la matrice ottenuta dal prodotto dei due precedenti contrasti, vettore a vettore:

```
-2  0
 2  0
 1 -1
-1  1
 1  1
-1 -1
```

che porta alla stima:

```
-1.600000
-1.266667
```

e alla devianza del contrasto 1.84333 e alla QM del contrasto 0.92167. Segue che il valore empirico di F vale 0.06758.
Riepilogando i contrasti sottoforma di tabella ANOVA si ha:

```
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value  Pr(>F)
A          2  41.468  20.734  1.5202 0.25795
S          1  88.002  88.002  6.4523 0.02593 *
A:S        2 101.448  50.724  3.7191 0.05536 .
Residuals 12 163.667  13.639
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Esercizio 4.4.14

1)
```
 Y=0       Freq
   0 0.16956802
   1 0.09219858
  10 0.15022566
 100 0.29400387
5000 0.29400387
```

```
Y=1      Freq
   0 0.1437047
   1 0.1595254
  10 0.1832564
 100 0.2702703
5000 0.2432432
```

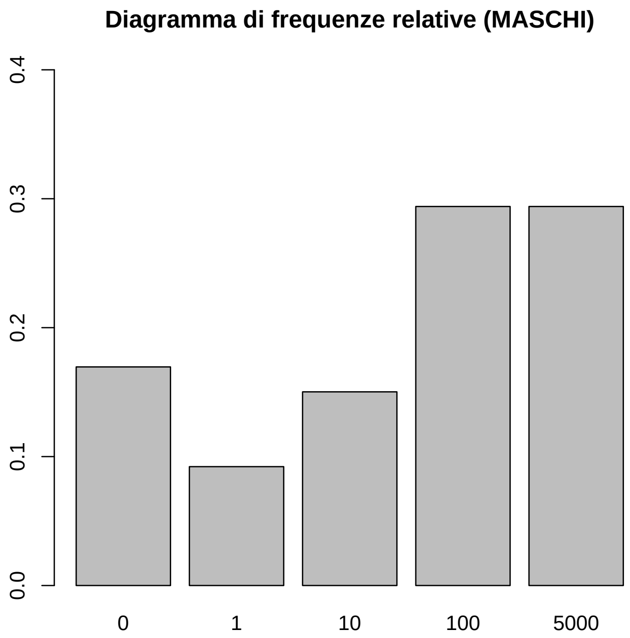


Figura 1: Barplot maschi

La Figura 1 rappresenta il diagramma di frequenze relative per i soli maschi mentre la Figure 2 quello per le sole femmine.

2) La distribuzione della preferenza dei soli maschi è bimodale (100 e 5000 sono i valori che presentano frequenza assoluta maggiore) mentre quella delle sole femmine è unimodale (100 è il valore che presenta frequenza assoluta maggiore).

3) `Statistica test:  44.87662`

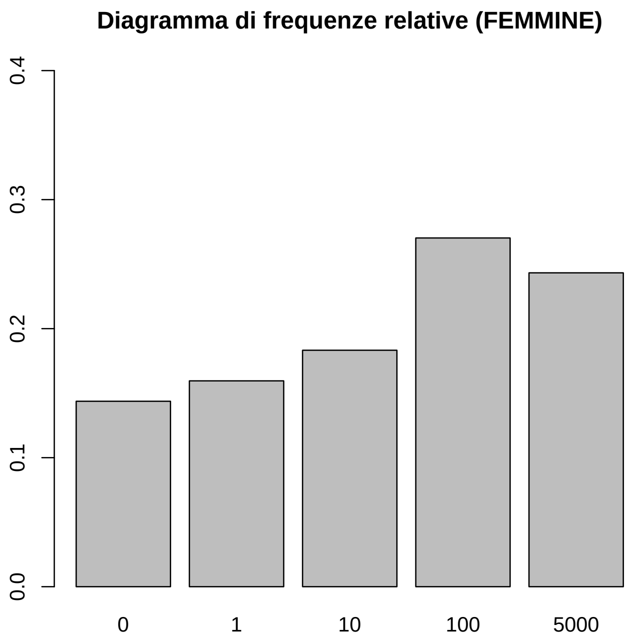
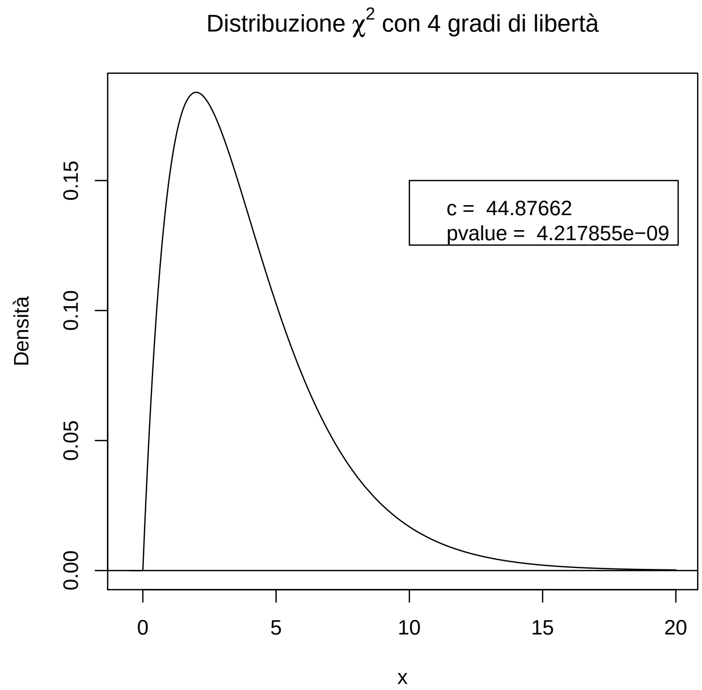


Figura 2: Barplot femmine

```
Valore critico con ampiezza 0.01 (monolaterale destro):  2.584685

gradi di libertà:  4

-> Rifiutare H0 con ampiezza 0.01

Frequenze osservate:

      X = 0 X = 1 X = 10 X = 100 X = 5000
Y = 0   263   143    233     456      456
Y = 1   218   242    278     410      369

Frequenze attese:

         X = 0    X = 1   X = 10  X = 100 X = 5000
Y = 0 243.1653 194.6333 258.3315 437.7986 417.0714
Y = 1 237.8347 190.3667 252.6685 428.2014 407.9286
```

```
Residui di Pearson:

          X = 0     X = 1    X = 10    X = 100  X = 5000
Y = 0  1.271967 -3.701020 -1.576057  0.8698982  1.906179
Y = 1 -1.286142  3.742265  1.593621 -0.8795926 -1.927422
```

4) `pvalue:  4.217855e-09`



Esercizio 4.4.15

1) La Figura 3 rappresenta il Boxplot della risposta per ogni specie.

2) `stima puntuale varianza per specie Setosa:  0.03015918`

   `stima puntuale varianza per specie Versicolor:  0.2208163`

   `stima puntuale varianza per specie Virginica:  0.3045878`

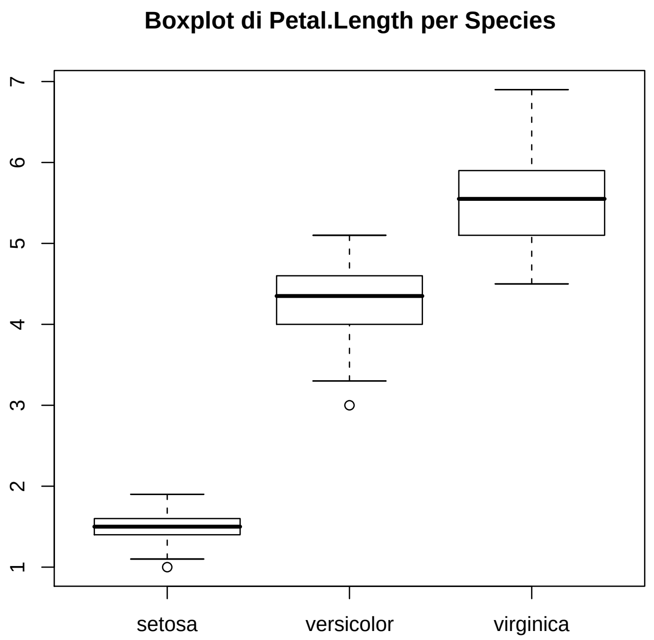


Figura 3: Barplot per specie

3) `Sintesi descrittiva per specie Setosa:`

```
minimo quant0.25   mediana quant0.75   massimo
   1.0       1.4       1.5       1.6       1.9
```

`Sintesi descrittiva per specie Versicolor:`

```
minimo quant0.25   mediana quant0.75   massimo
   3.0       4.0       4.3       4.6       5.1
```

`Sintesi descrittiva per specie Virginica:`

```
minimo quant0.25   mediana quant0.75   massimo
   4.5       5.1       5.5       5.9       6.9
```

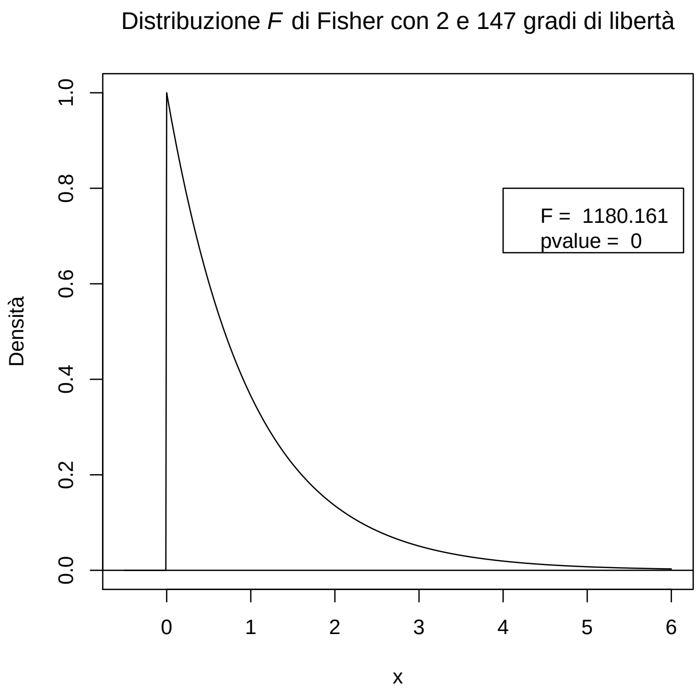
4) `1-alpha:  0.95`

`Intervallo di confidenza per specie Setosa: ( 0.02104456 , 0.04683264 )`

```
Intervallo di confidenza per specie Versicolor: ( 0.1540819 , 0.3428943 )

Intervallo di confidenza per specie Virginica: ( 0.2125361 , 0.4729786 )
```

5) H0: mu1 = mu2 = mu3
H1: almeno una mu diversa dalle altre

```
Statistica test:  1180.161

Valore critico con ampiezza 0.05 (monolaterale destro):  3.057621

gradi di libertà: ( 2 , 147 )

-> Rifiutare H0 con ampiezza 0.05
```



6) media campionaria per specie Setosa: 1.462

```
media campionaria per specie Versicolor:  4.26

media campionaria per specie Virginica:  5.552
```

Le medie campionarie sono molto differenti tra loro (quella di Virginica è circa 4 volte quella di Setosa) a parità di dimensione campionaria (50). Questo fa propendere per il rifiuto dell'ipotesi nulla. L'ipotesi di uguaglianza tra le varianze condizionate non sembra comunque trovare un buon riscontro nei dati.