

INTRODUZIONE ALLA STATISTICA APPLICATA con esempi in R

<http://hpe.pearsoned.it/stefanini>



Soluzioni degli esercizi di ricapitolazione Capitolo 5: “Introduzione al modello lineare”

F. Frascati

F. M. Stefanini

11 gennaio 2008



Esercizio 5.3.1

La tabella 5.8 contiene degli errori. La tabella corretta corrisponde ai dati sotto riportati (T = Temperatura, V = Vitamina, P = Proteina):

- 1) In termini matriciali $Y = X\beta + \varepsilon$ con Y trasposto definita da P:

```
aov(formula = P ~ T + V + T : V)
```

e la matrice del disegno (due fattori con interazione):

	beta0	T22	T26	T30	V2.5	T22:V2.5	T26:V2.5	T30:V2.5
1	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0
3	1	0	0	0	1	0	0	0
4	1	0	0	0	1	0	0	0
5	1	1	0	0	0	0	0	0
6	1	1	0	0	0	0	0	0
7	1	1	0	0	1	1	0	0
8	1	1	0	0	1	1	0	0
9	1	0	1	0	0	0	0	0
10	1	0	1	0	0	0	0	0
11	1	0	1	0	1	0	1	0
12	1	0	1	0	1	0	1	0
13	1	0	0	1	0	0	0	0
14	1	0	0	1	0	0	0	0
15	1	0	0	1	1	0	0	1
16	1	0	0	1	1	0	0	1

```
attr("assign")
[1] 0 1 1 1 2 3 3 3
attr("contrasts")
attr("contrasts")$T
[1] "contr.treatment"
```

```
attr("contrasts")$V
[1] "contr.treatment"
```

```
2) (Intercept)  3.040
    T22         0.280
    T26         2.200
    T30         2.715
    V2.5        -0.070
    T22:V2.5     0.310
    T26:V2.5    -0.275
    T30:V2.5    -0.200
```

Stima della Varianza: 0.1115938

3) Analysis of Variance Table

Response: P

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
T	3	18.9878	6.3293	56.7169	9.826e-06 ***
V	1	0.0495	0.0495	0.4436	0.5241
T:V	3	0.2049	0.0683	0.6121	0.6260
Residuals	8	0.8927	0.1116		

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

Fteorico con gradi di libertà (3 , 11): 6.21673

Fteorico con gradi di libertà (1 , 11): 9.646034

- 4) La Figura 1 mostra il diagramma Q-Q Plot dei residui del modello finale.
- 5) Il principale commento riguarda la fonte di variazione T che è risultata significativa.

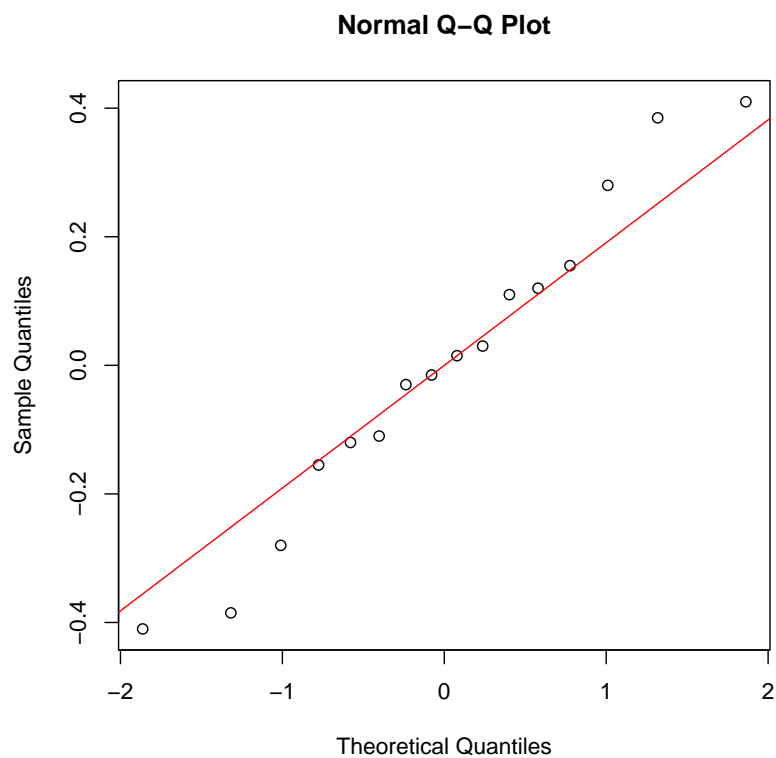


Figura 1: Q-Q Plot

Esercizio 5.3.2

La Figura 2 rappresenta il diagramma di dispersione.

1) β_0 -1.4003750

T 0.2368125

Varianza stimata: 0.1563780

2) β_0 4.505625

V -0.111250

Varianza stimata: 1.434674

3) MODELLO ESTESO:

β_0 -2.5998750

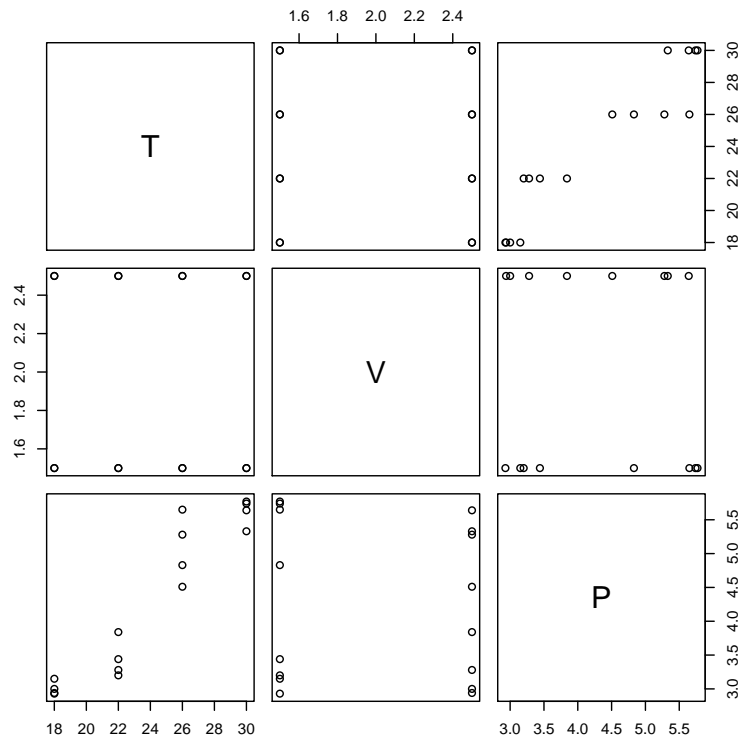


Figura 2: Diagramma di dispersione

T 0.2960625

V 0.5997500

TV -0.0296250

Varianza stimata: 0.1724646

Devianza residua: 2.069575

gradi di libertà: 12

4) 1-alpha: 0.97

Intervallo di confidenza per beta1: (0.1815681 , 0.2920569)

Intervallo di confidenza per beta2: (-0.6053709 , 0.3828709)

5) $E[P \mid T = 27, V = 2] = 4.993563$

6) 1-alpha: 0.95

Intervallo di previsione per la media condizionata: (4.078230 , 5.908895)

Esercizio 5.3.3

1) MODELLO ESTESO:

beta0 65.021826
X1 -2.128530
X2 1.129063
X12 3.192689

Varianza stimata: 6.662895
Devianza residua: 2931.674
gradi di libertà: 440

2) MODELLO RIDOTTO:

beta0 65.021826
X1 -2.128530
X2 1.129063

Varianza stimata: 16.91039
Devianza residua: 7457.481
gradi di libertà: 441

TEST DI IPOTESI:

Contributo parziale: 4525.808
Fempirico: 679.2554
Fteorico: 3.86268
gradi di libertà: (1 , 440)

Rifiutare H_0 con ampiezza 0.05

3) La Figura 3 analizza graficamente i residui. In particolare segue l'evidenza

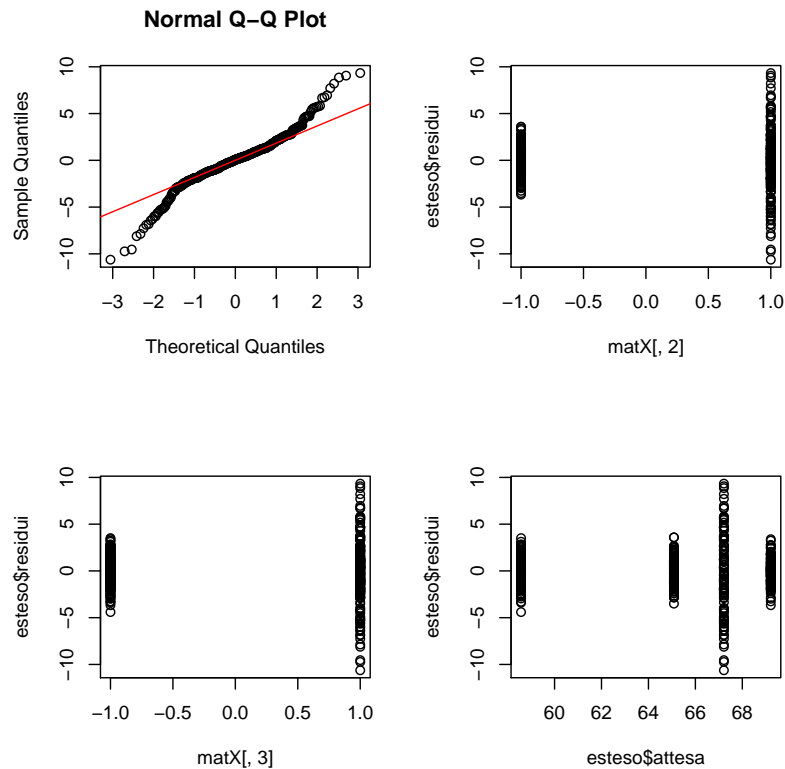


Figura 3: Analisi grafica dei residui

di possibili violazioni delle assunzioni di normalità e di omogeneità della varianza.

- 4) L'intercetta diventa la media della combinazione di livelli (cella) -1,-1.

Esercizio 5.3.4

- 1) Eseguire il seguente codice:

```
> enne <- 10000
> minimo <- 1
> massimo <- 10
> x <- runif(enne, minimo, massimo)
> epsilon <- rnorm(enne)
> dataDF <- data.frame(x, epsilon)
```

2) `> dataDF$Y <- 10 + 2 * dataDF$x + dataDF$epsilon`

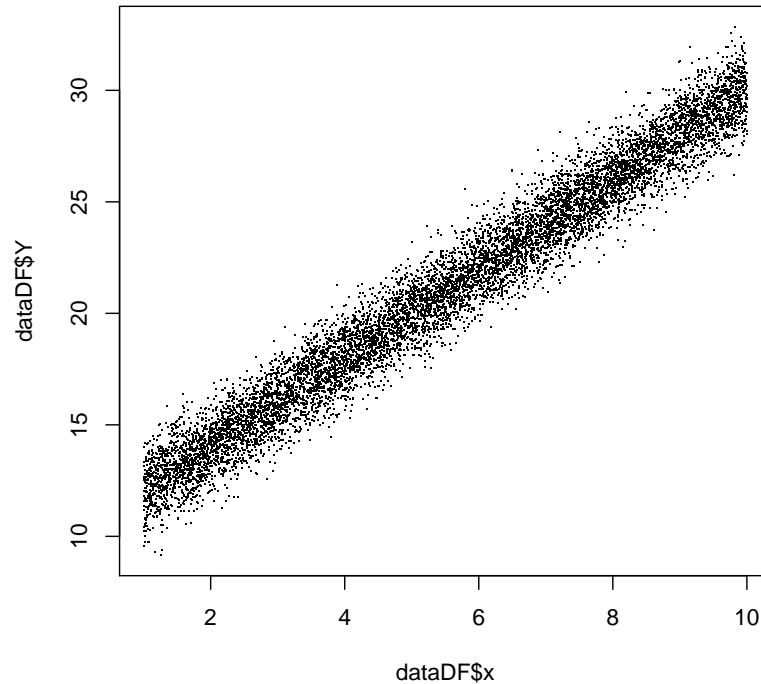


Figura 4: Diagramma di dispersione

La Figura 4 mostra il diagramma di dispersione di Y su x .

- 3) La Figura 5 mostra il diagramma di dispersione e la retta dei minimi quadrati.
- 4) La Figura 6 mostra il diagramma dei residui verso l'attesa di Y e verso i valori di X .
- 5) La Figura 7 mostra l'istogramma delle frequenze relative dei residui e sovrappone una gaussiana che ha media nulla e varianza uguale al valore stimato per σ^2 .
- 6) La Figura 8 mostra il grafico quantile-quantile dei residui.
- 7) Lasciato per esercizio.

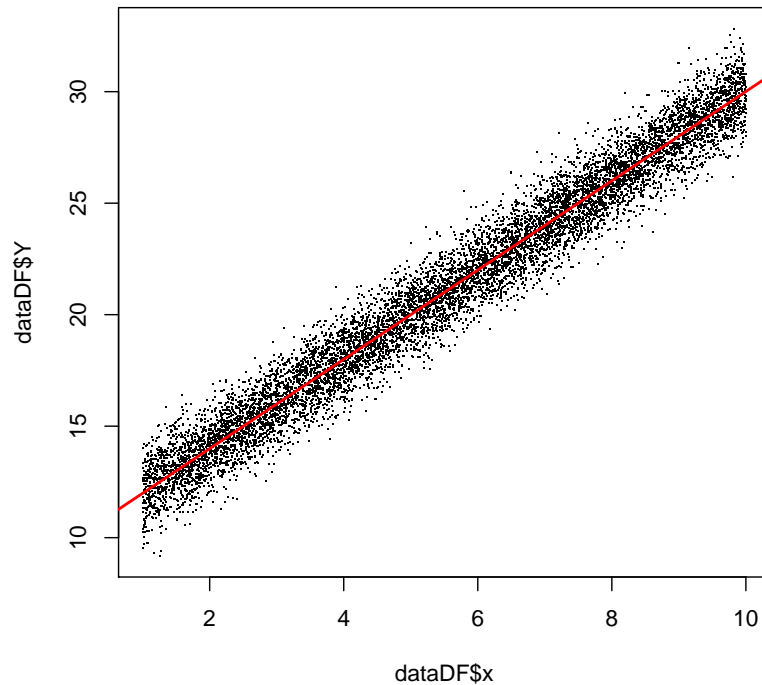


Figura 5: Diagramma di dispersione e retta dei minimi quadrati

Esercizio 5.3.5

1) β_0 0.18

x 1.02

2) $E[Y \mid x = 114.5] = 116.97$

1-alpha: 0.95

Intervallo di confidenza per la media condizionata: (116.7863 , 117.1537)

Esercizio 5.3.6

1) 1-alpha: 0.9

Intervallo di confidenza per β_1 : (0.005497878 , 0.06062366)

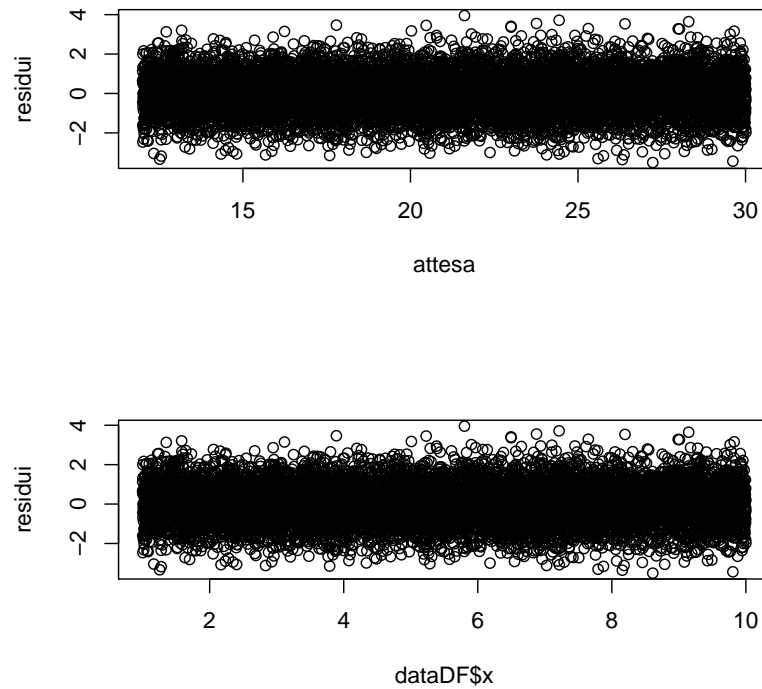


Figura 6: Residui verso attesa e valori di x

2) $E[Y \mid x = 77] = 3.622907$

1-alpha: 0.9

Intervallo di confidenza per la media condizionata: (3.192361 , 4.053453)

3) Varianza stimata: 0.1309715

4) MODELLO RIDOTTO:

beta0 1.07722762

x 0.03306077

Varianza stimata: 0.1309715

Devianza residua: 0.6548573

gradi di libertà: 5

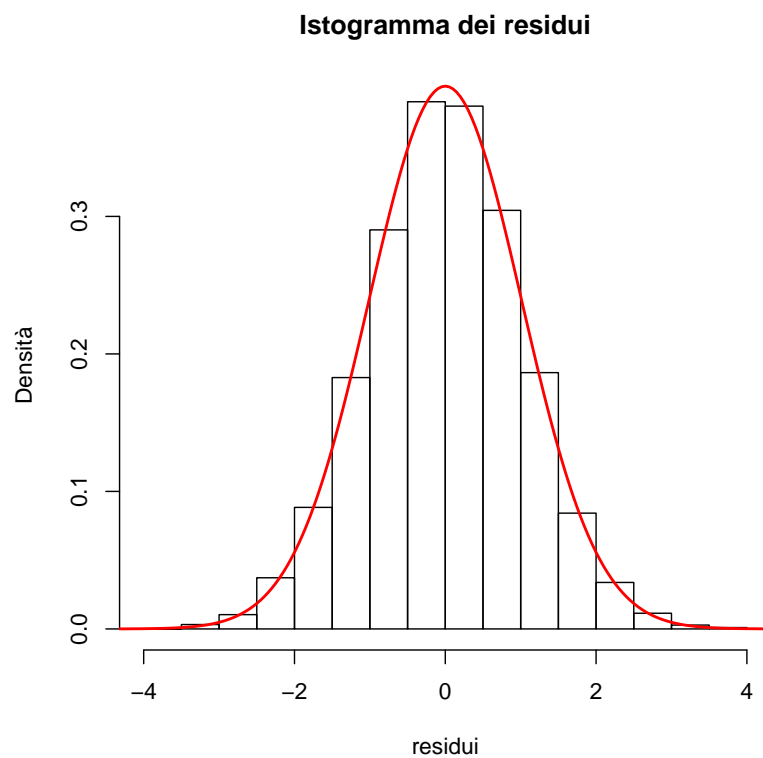


Figura 7: Istogramma dei residui ed approssimazione normale

MODELLO ESTESO:

beta0 14.97864147

x -0.40504439

x2 0.00337004

Varianza stimata: 0.01465150

Devianza residua: 0.05860599

gradi di libertà: 4

TEST DI IPOTESI:

Contributo parziale: 0.5962513

Fempirico: 40.69559

Fteorico: 7.708647

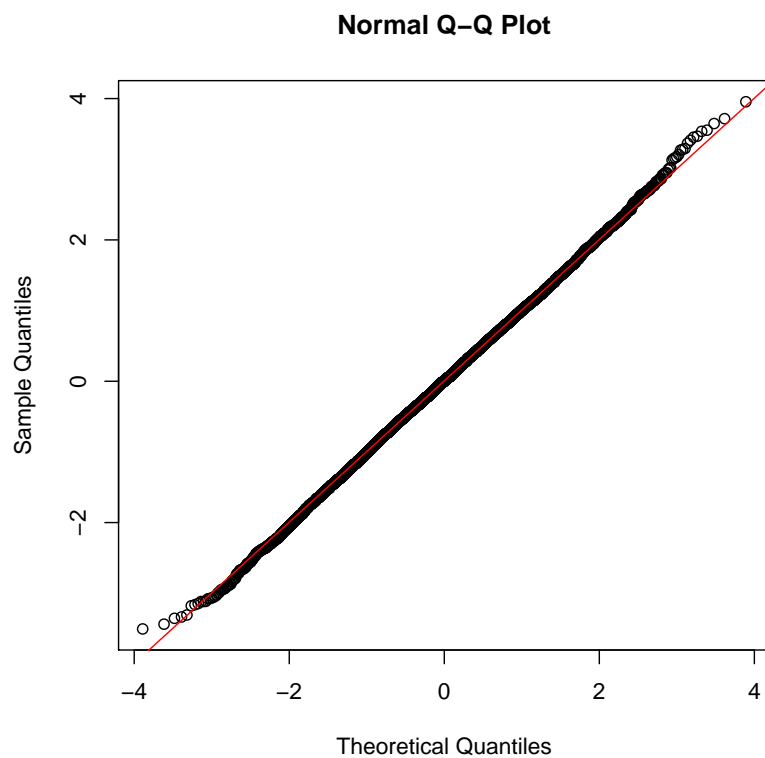


Figura 8: Grafico Q-Q Plot dei residui

gradi di libertà: (1 , 4)

Rifiutare H_0 con ampiezza 0.05

Esercizio 5.3.7

1) La Figura 9 rappresenta il diagramma di dispersione.

2) $Y = \text{beta0} + \text{beta1} * x + \text{epsilon}$

3) $\text{beta0} = 58.53235$

$x = 16.89275$

Varianza stimata: 0.8962629

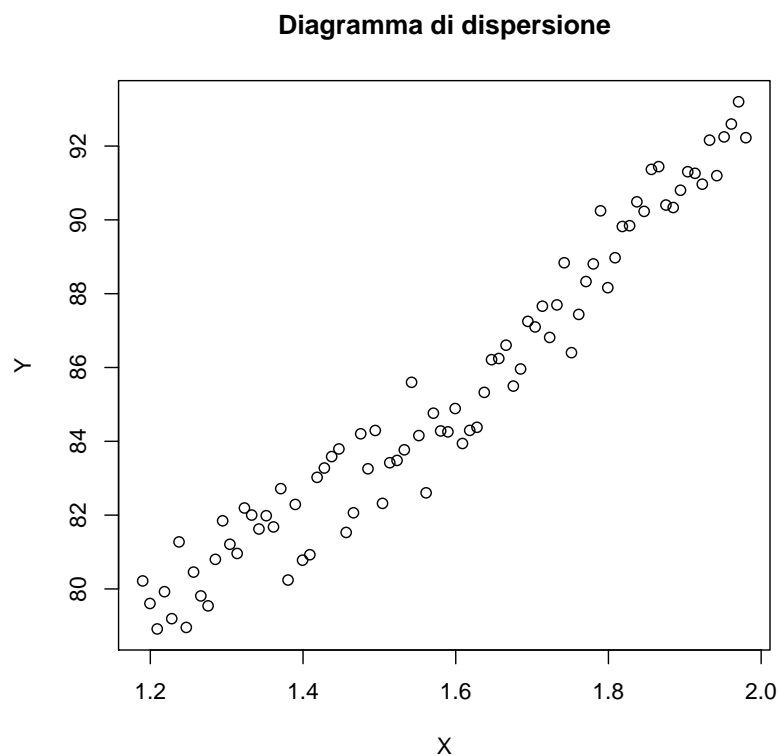


Figura 9: Diagramma di dispersione

- 4) La Figura 10 vede sovrapposto al grafico nel punto (1) il valore dell'attesa condizionata.
- 5) Questa domanda chiede di effettuare un test di ipotesi per saggiare l'introduzione di un termine di secondo grado.

MODELLO RIDOTTO:

beta0 84.27645

x -16.29554

x2 10.46949

Varianza stimata: 0.6494712

Devianza residua: 52.60717

gradi di libertà: 81

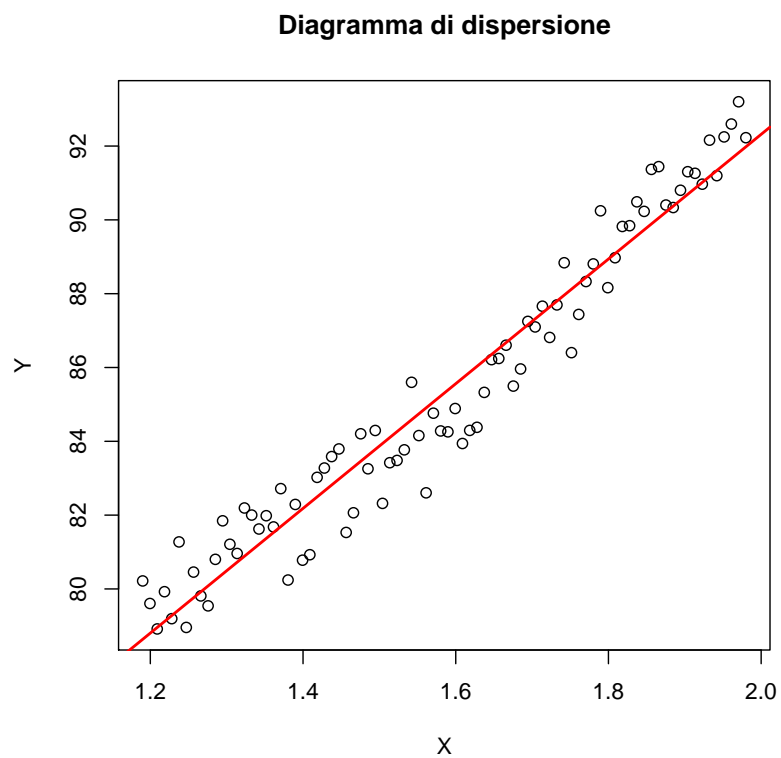


Figura 10: Attesa condizionata

MODELLO ESTESO:

beta0 82.1218445

x -12.1095708

x2 7.7944811

x3 0.5625682

Varianza stimata: 0.6575587

Devianza residua: 52.60469

gradi di libertà: 80

TEST DI IPOTESI:

Contributo parziale: 0.002474975

Fempirico: 0.003763885

Fteorico: 3.960352

gradi di libertà: (1 , 80)

Non rifiutare H_0 con ampiezza 0.05

E uno di terzo?

MODELLO RIDOTTO:

beta0 58.53235

x 16.89275

Varianza stimata: 0.8962629

Devianza residua: 73.49356

gradi di libertà: 82

MODELLO ESTESO:

beta0 84.27645

x -16.29554

x2 10.46949

Varianza stimata: 0.6494712

Devianza residua: 52.60717

gradi di libertà: 81

TEST DI IPOTESI:

Contributo parziale: 20.88639

Fempirico: 32.15907

Fteorico: 3.958852

gradi di libertà: (1 , 81)

Rifiutare H_0 con ampiezza 0.05

6) La Figura 11 rappresenta l'analisi dei residui per il modello finale.

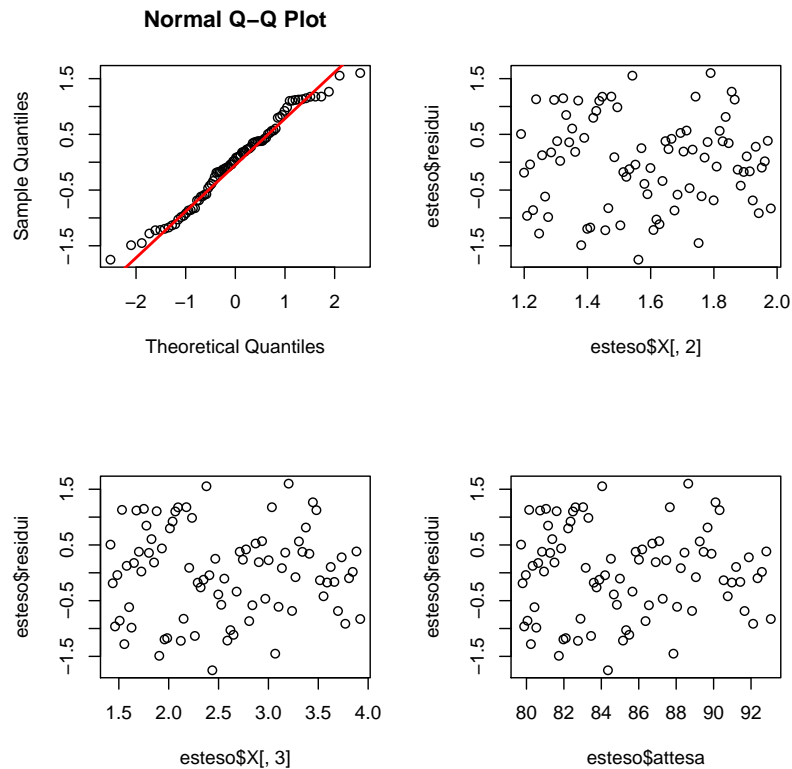


Figura 11: Analisi dei residui

Esercizio 5.3.8

1) `aov(formula = Y ~ dose)`

TEST DI IPOTESI:

Contributo parziale: 2.48

Fempirico: 124

Fteorico: 5.143253

gradi di libertà: (2 , 6)

Rifiutare H_0 con ampiezza 0.05

2) Si rifiuta decisamente l'ipotesi nulla di uguaglianza tra medie di popolazione con ampiezza $\alpha = 0.05$. La conclusione per per $\alpha = 0.01$ sarebbe stata la stessa:

TEST DI IPOTESI:

Contributo parziale: 2.48

Fempirico: 124

Fteorico: 10.92477

gradi di libertà: (2 , 6)

Rifiutare H_0 con ampiezza 0.01

3) Test di ipotesi per il primo contrasto:

TEST DI IPOTESI:

Contributo parziale: 0.06

Fempirico: 6

Fteorico: 5.987378

gradi di libertà: (1 , 6)

Rifiutare H_0 con ampiezza 0.05

Test di ipotesi per il secondo contrasto:

TEST DI IPOTESI:

Contributo parziale: 2.42

Fempirico: 141.1667

Fteorico: 5.591448

gradi di libertà: (1 , 7)

Rifiutare H_0 con ampiezza 0.05

4) β_0 1.133333

X 0.600000

5) La Figura 12 mostra diagrammati i valori osservati e la retta dei minimi quadrati.

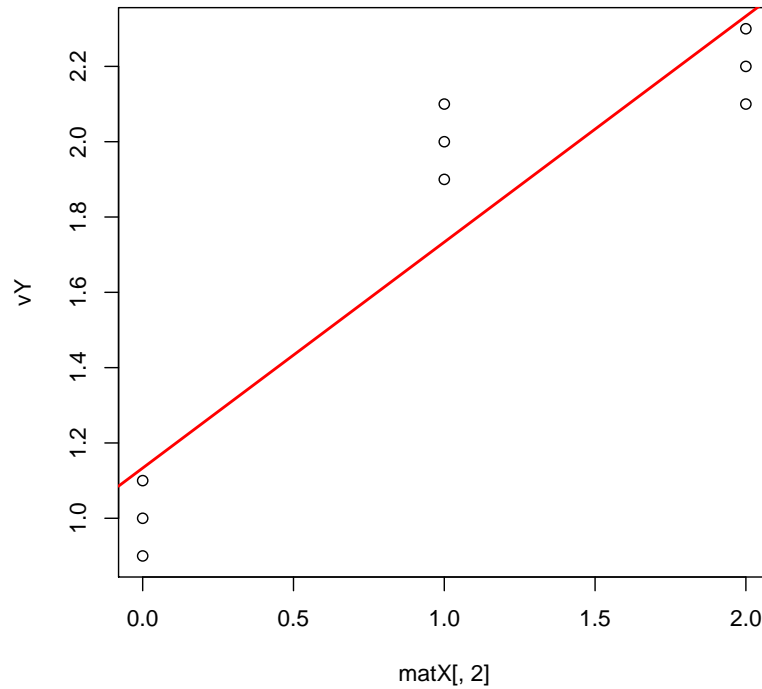


Figura 12: Valori osservati e retta dei minimi quadrati

Esercizio 5.3.9

1) Sintesi descrittive di X:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
122.2	122.7	123.1	123.1	123.5	124.0

Varianza stimata: 0.2292867

Sintesi descrittive di X condizionatamente a tempo = 1:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
122.2	122.7	123.1	123.1	123.5	124.0

Varianza stimata: 0.2129011

Sintesi descrittive di X condizionatamente a tempo = 2:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
122.3	122.7	123.2	123.2	123.6	124.0

Varianza stimata: 0.25954

Sintesi descrittive di X condizionatamente a $\text{tempo} = 3$:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
122.2	122.6	122.9	123.0	123.3	124.0

Varianza stimata: 0.2089716

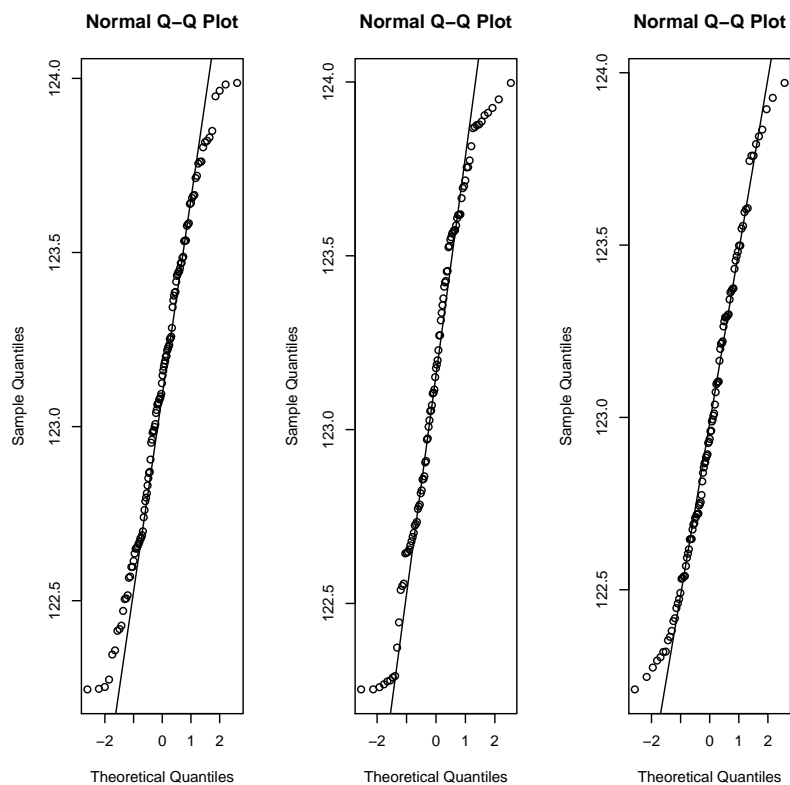


Figura 13: Normalità di X

- 2) La Figura 13 permette di studiare graficamente la normalità delle distribuzioni condizionate al tempo.

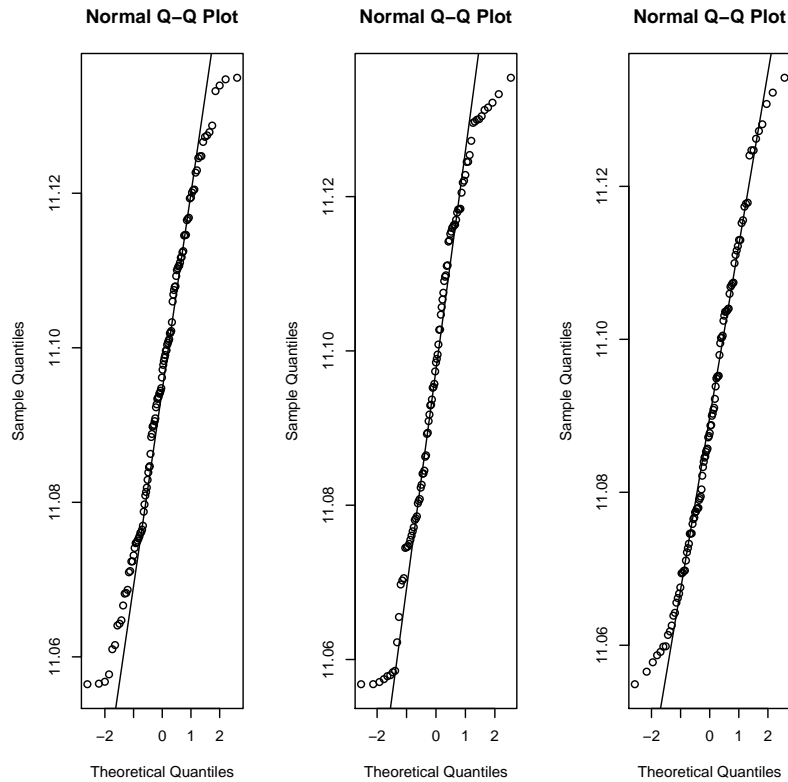


Figura 14: Normalità della trasformata $Y = \sqrt{X}$

- 3) La Figura 14 studia la normalità (condizionata alla variabile **tempo**) della trasformata $Y = \sqrt{X}$ mentre la Figura 15 si occupa della trasformata $Z = \log(X)$.
- 4) Test per effetto del tempo di prelievo sulla scala di $Z = \log(X)$.

TEST DI IPOTESI:

Contributo parziale: 9.97e-05

Fempirico: 3.3461

Fteorico: 4.6773

gradi di libertà: (2 , 297)

Non rifiutare H_0 con ampiezza 0.01

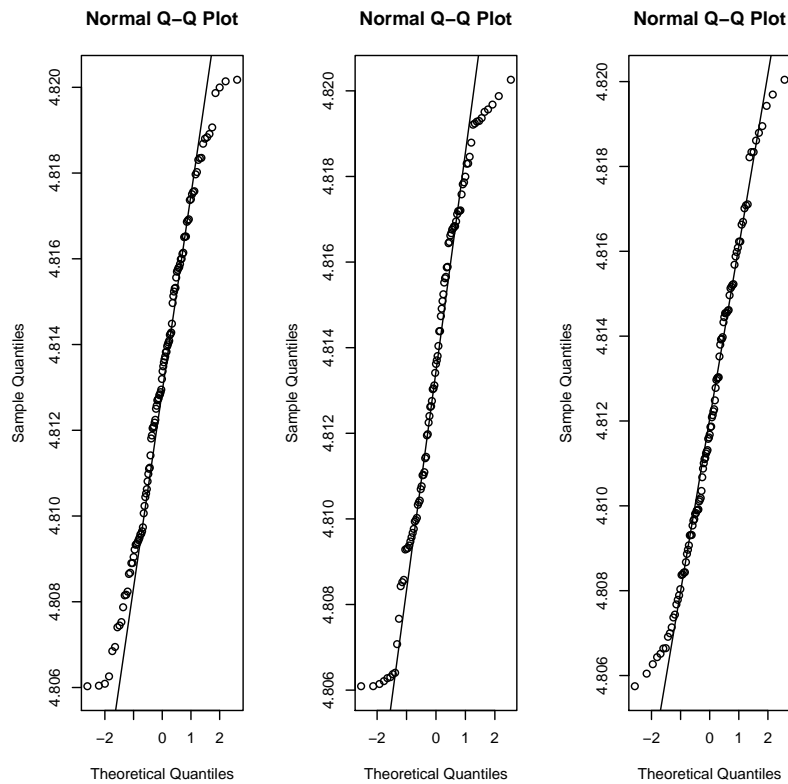


Figura 15: Normalità della trasformata $Z = \log(X)$

Esercizio 5.3.10

1) `beta0` 41.15476
`secondi` 0.74159

2) La Figura 16 rappresenta graficamente l'analisi dei residui.

3) MODELLO RIDOTTO:
`beta0` 41.15476
`secondi` 0.74159

Varianza stimata: 1.9649

Devianza residua: 27.509

gradi di libertà: 14

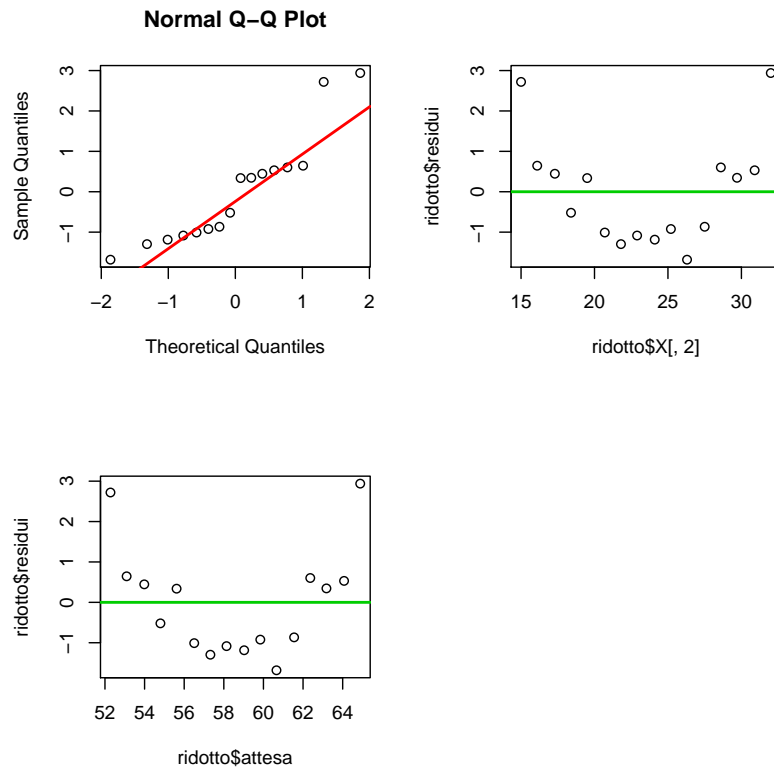


Figura 16: Analisi dei residui

MODELLO ESTESO:

beta0 67.189001

secondi -1.589403

secondi2 0.049596

Varianza stimata: 0.32676

Devianza residua: 4.2479

gradi di libertà: 13

TEST DI IPOTESI:

Contributo parziale: 23.261

Fempirico: 71.185

Fteorico: 9.0738

gradi di libertà: (1 , 13)

Rifiutare H_0 con ampiezza 0.01

4) TEST DI IPOTESI:

Contributo parziale: 36.179

Fempirico: 0.62455

Fteorico: 3.4903

gradi di libertà: (3 , 12)

Non rifiutare H_0 con ampiezza 0.05

5) $E[\text{tempera} \mid \text{secondi} = 250, \text{secondi}^2 = 250^2] = 2769.6$

6) $1-\alpha$: 0.95

Intervallo di previsione per la media condizionata: (53.274 , 55.915)

Esercizio 5.3.11

1)	flavo	polifenoli
Min.	: 5.00	Min. : 5.2
1st Qu.:	6.92	1st Qu.:13.2
Median :	7.50	Median :15.7
Mean :	8.03	Mean :14.7
3rd Qu.:	8.20	3rd Qu.:16.6
Max.	:25.70	Max. :19.3

2) β_0 17.19321

X -0.30537

3) La Figura 17 rappresenta il diagramma di dispersione e la retta dei minimi quadrati.

4) Distanza di *Cook* per ogni unità campionaria:

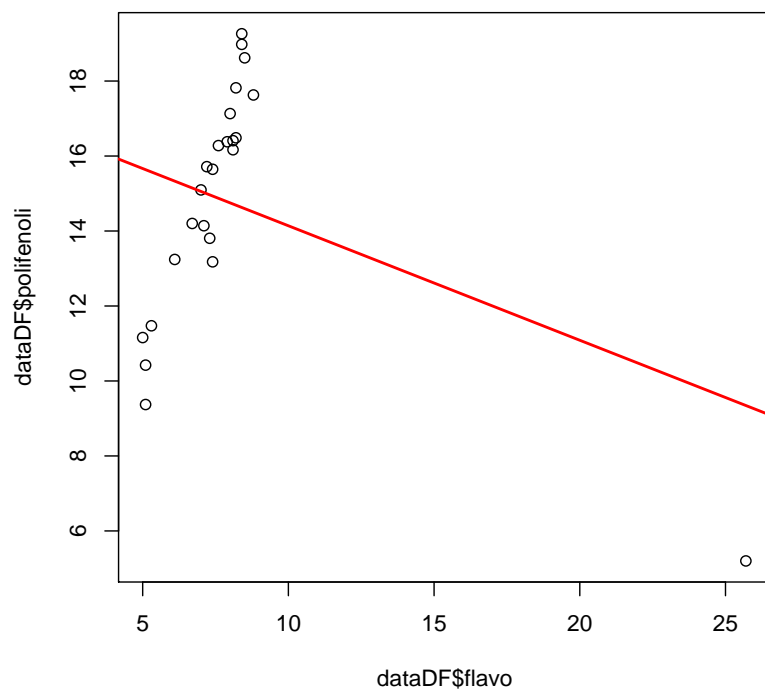


Figura 17: Diagramma di dispersione e retta dei minimi quadrati

1	2	3	4	5	6	7
1.8273e-03	2.2416e-02	1.1146e+02	9.9115e-02	5.8020e-02	1.2258e-02	7.1306e-03
8	9	10	11	12	13	14
3.0608e-03	2.2221e-03	3.6190e-02	2.1611e-02	7.6282e-02	1.2059e-03	5.6402e-03
15	16	17	18	19	20	21
1.1541e-03	4.2081e-02	3.9481e-06	4.6154e-03	6.2879e-03	4.7747e-02	4.4124e-03
22	23	24				
1.2492e-02	1.4311e-01	6.9731e-03				

La Figura 18 rappresenta il grafico della distanza di *Cook*.

5) La Figura 19 mostra l'analisi dei residui.

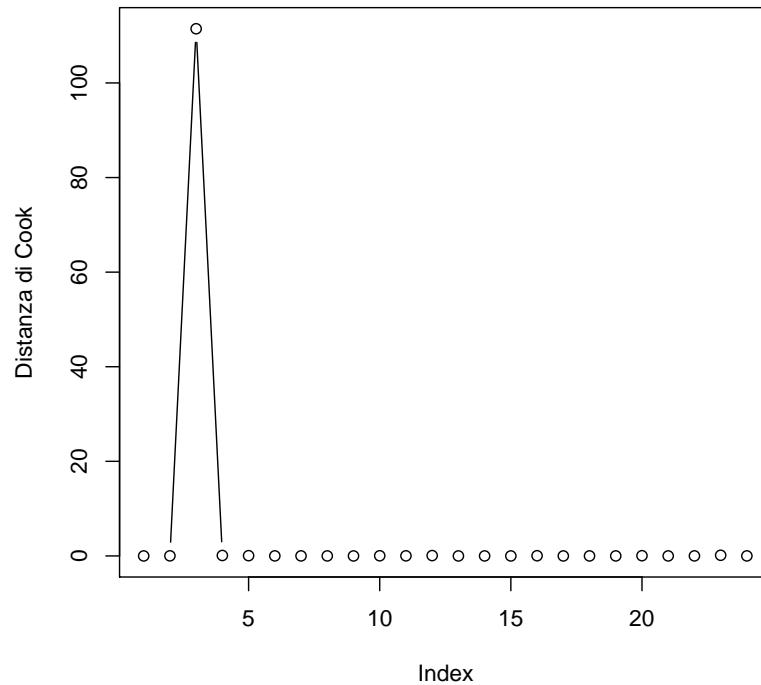


Figura 18: Distanza di *Cook*

Esercizio 5.3.12

- 1) Da svolgere.
- 2) Da svolgere.
- 3) Da svolgere.

Esercizio 5.3.13

- 1) Parametrizzazione a trattamento (corner):

	(Intercept)	A2	A3	S2	A2:S2	A3:S2
1	1	0	0	0	0	0
2	1	0	0	0	0	0

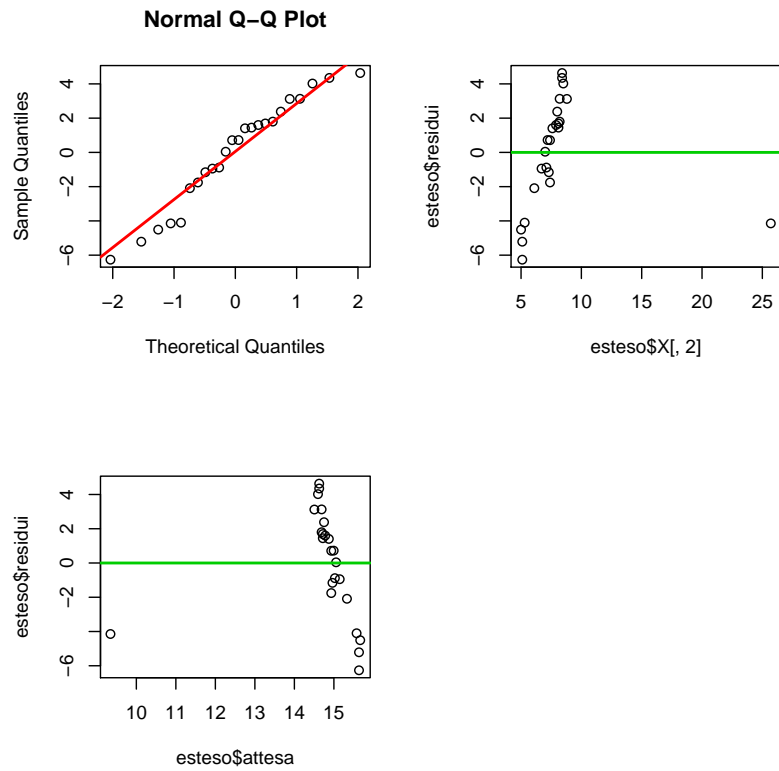


Figura 19: Analisi dei residui

3	1	0	0	0	0	0
4	1	0	0	1	0	0
5	1	0	0	1	0	0
6	1	0	0	1	0	0
7	1	1	0	0	0	0
8	1	1	0	0	0	0
9	1	1	0	0	0	0
10	1	1	0	1	1	0
11	1	1	0	1	1	0
12	1	1	0	1	1	0
13	1	0	1	0	0	0
14	1	0	1	0	0	0
15	1	0	1	0	0	0
16	1	0	1	1	0	1

```

17          1  0  1  1      0      1
18          1  0  1  1      0      1
attr("assign")
[1] 0 1 1 2 3 3
attr("contrasts")
attr("contrasts")$A
[1] "contr.treatment"

attr("contrasts")$S
[1] "contr.treatment"

```

Parametrizzazione a media a somma nulla:

	(Intercept)	A1	A2	S1	A1:S1	A2:S1
1	1	1	0	1	1	0
2	1	1	0	1	1	0
3	1	1	0	1	1	0
4	1	1	0	-1	-1	0
5	1	1	0	-1	-1	0
6	1	1	0	-1	-1	0
7	1	0	1	1	0	1
8	1	0	1	1	0	1
9	1	0	1	1	0	1
10	1	0	1	-1	0	-1
11	1	0	1	-1	0	-1
12	1	0	1	-1	0	-1
13	1	-1	-1	1	-1	-1
14	1	-1	-1	1	-1	-1
15	1	-1	-1	1	-1	-1
16	1	-1	-1	-1	1	1
17	1	-1	-1	-1	1	1
18	1	-1	-1	-1	1	1

```

attr("assign")
[1] 0 1 1 2 3 3

```

```
attr("contrasts")
attr("contrasts")$A
[1] "contr.sum"
```

```
attr("contrasts")$S
[1] "contr.sum"
```

Parametrizzazione a media di cella:

	codeA1S1	codeA1S2	codeA2S1	codeA2S2	codeA3S1	codeA3S2
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1
7	1	0	0	0	0	0
8	0	1	0	0	0	0
9	0	0	1	0	0	0
10	0	0	0	1	0	0
11	0	0	0	0	1	0
12	0	0	0	0	0	1
13	1	0	0	0	0	0
14	0	1	0	0	0	0
15	0	0	1	0	0	0
16	0	0	0	1	0	0
17	0	0	0	0	1	0
18	0	0	0	0	0	1

```
attr("assign")
[1] 1 1 1 1 1 1
attr("contrasts")
attr("contrasts")$code
[1] "contr.treatment"
```

2) MODELLO RIDOTTO:

(Intercept) 32.63889

A2 0.93333

A3 3.58333

S2 4.42222

Varianza stimata: 18.937

Devianza residua: 265.11

gradi di libertà: 14

MODELLO ESTESO:

(Intercept) 35.06667

A2 -0.70000

A3 -2.06667

S2 -0.43333

A2:S2 3.26667

A3:S2 11.30000

Varianza stimata: 13.639

Devianza residua: 163.67

gradi di libertà: 12

TEST DI IPOTESI:

Contributo parziale: 101.45

Fempirico: 3.7191

Fteorico: 6.9266

gradi di libertà: (2 , 12)

Non rifiutare H_0 con ampiezza 0.01

3) Attesa: 33.572

1-alpha: 0.95

Intervallo di previsione per la media condizionata: (29.172 , 37.972)

Esercizio 5.3.14

1) Sintesi descrittive di `caloria`:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
76.3	79.1	80.8	82.6	86.0	95.5

Varianza stimata: 21.817

2) La variabile `fibra` trasformata nel fattore `fibraQ`:

[23,24]	(24,25]	(25,26]	(26,27]
57	57	56	57

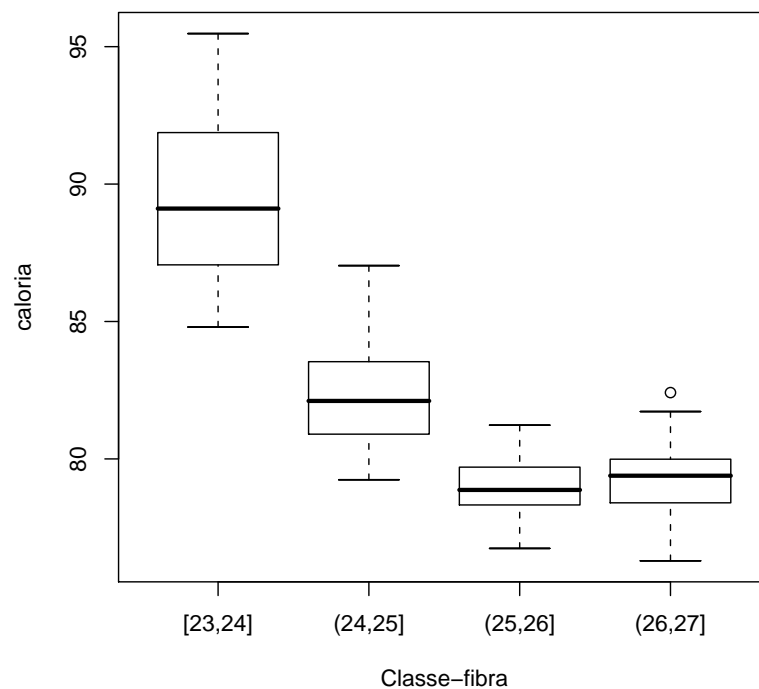


Figura 20: Boxplot di `caloria` condizionata a `fibraQ`

3) La Figura 20 rappresenta il Boxplot di *caloria* condizionata a *fibraQ*.

4) Si osservano forti differenze nelle distribuzioni condizionate.

5) `lm(caloria~(fibra+grass+carboidrat)^3)`

6) Call:

```
lm(formula = caloria ~ fibra + grass + carboidrat + grass:carboidrat,
    data = dataDF)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.55283	-0.50901	-0.00642	0.60093	2.82081

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.15e+04	8.35e+02	37.71	<2e-16 ***
<i>fibra</i>	3.01e+00	9.85e-01	3.06	0.0025 **
<i>grass</i>	-3.55e+02	9.25e+00	-38.37	<2e-16 ***
<i>carboidrat</i>	-2.65e+02	6.93e+00	-38.25	<2e-16 ***
<i>grass:carboidrat</i>	2.99e+00	7.72e-02	38.69	<2e-16 ***

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

Residual standard error: 0.893 on 222 degrees of freedom

Multiple R-Squared: 0.964, Adjusted R-squared: 0.963

F-statistic: 1.49e+03 on 4 and 222 DF, p-value: <2e-16

7) La Figura 21 mostra graficamente l'analisi dei residui.

8) $E[\text{caloria} \mid \text{fibra} = 26.5, \text{grass} = 89.5, \text{carboidrat} = 120] = 82.864$

1-alpha: 0.98

Intervallo di previsione per la media condizionata: (79.293 , 86.435)

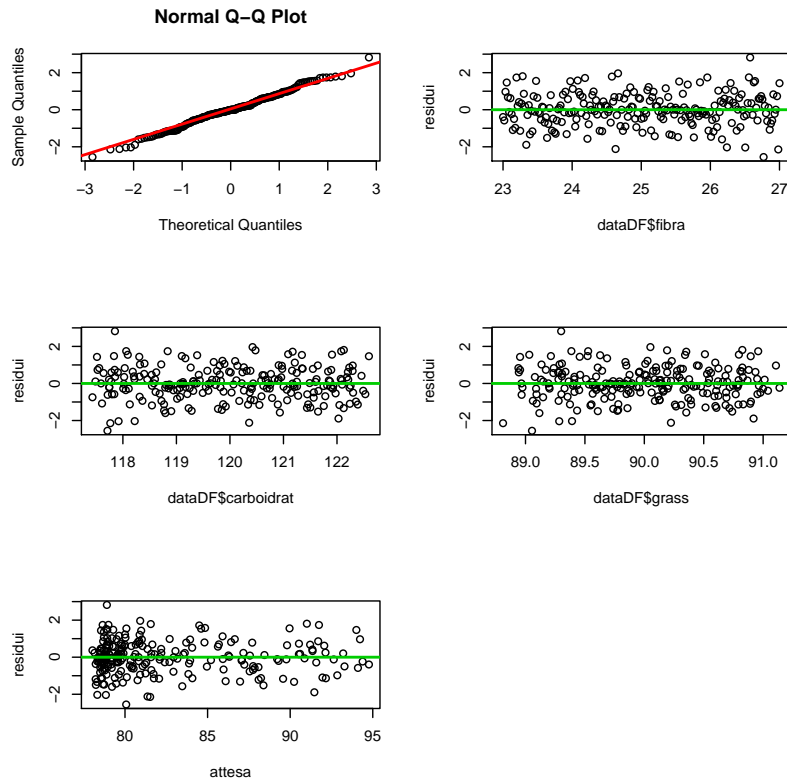


Figura 21: Analisi dei residui

Esercizio 5.3.15

1) La Figura 22 rappresenta il diagramma di dispersione delle variabili quantitative con punti colorati in dipendenza del valore assunto dalla Z .

2) $\text{beta0} -29.8681$
 $X \quad 2.9474$

$$E[Y \mid x = 38.5] = 83.607$$

$1-\alpha:$ 0.95

Intervallo di previsione per la media condizionata: (83.191 , 84.022)

3) La Figura 23 mostra il diagramma di dispersione originale e la retta dei minimi quadrati ottenuta al punto precedente.

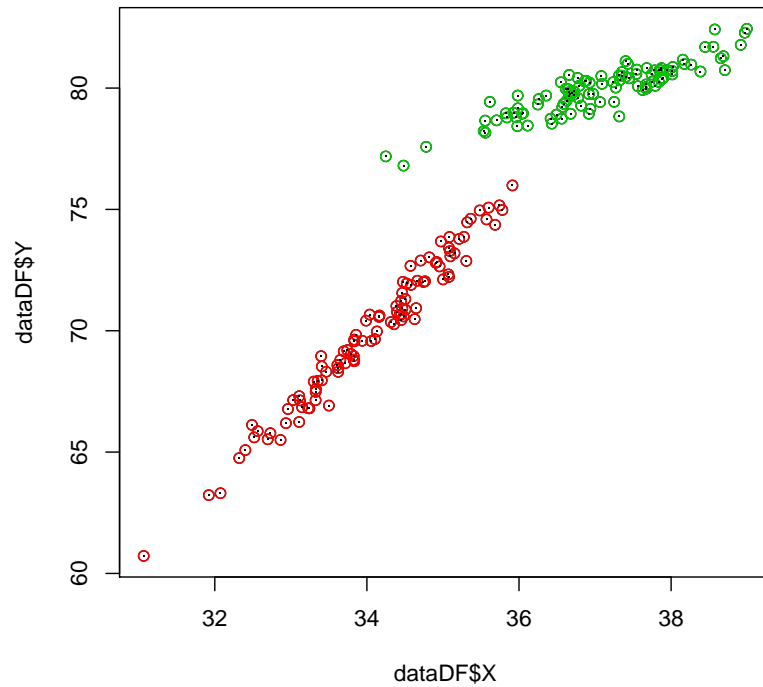


Figura 22: Diagramma di dispersione delle variabili quantitative

4) Modello con interazione:

```
beta0 -35.6644
X      3.1009
Z      79.7610
XZ     -2.1338
```

Modello comune:

```
beta0 -29.8681
X      2.9474
```

Regressione specifica gruppo Z = 0:

```
beta0 -35.6644
X      3.1009
```

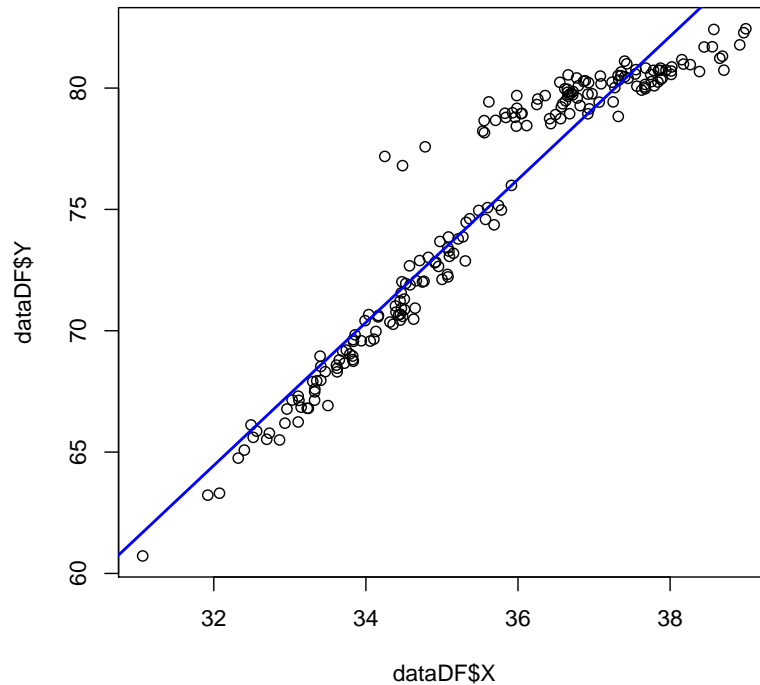


Figura 23: Diagramma di dispersione e retta dei minimi quadrati

Regressione specifica gruppo $Z = 1$:

```
beta0 44.09665
X      0.96717
```

- 5) La Figura 24 rappresenta il diagramma di dispersione delle variabili quantitative con punti colorati in dipendenza del valore assunto dalla Z e sovrainposte le due rette di regressione ottenute al punto precedente.

- 6) MODELLO RIDOTTO:

```
beta0 0.48496
X      2.04018
Z      3.83876
```

Varianza stimata: 1.3072

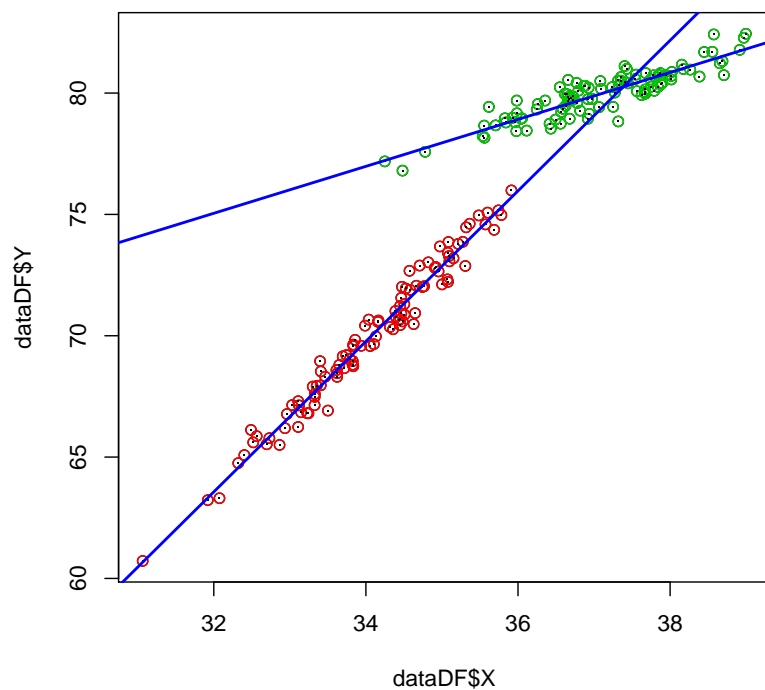


Figura 24: Diagramma di dispersione e rette di regressione

Devianza residua: 257.51
 gradi di libert : 197

MODELLO ESTESO:

beta0 -35.6644

X 3.1009

Z 79.7610

XZ -2.1338

Varianza stimata: 0.23526

Devianza residua: 46.111

gradi di libert : 196

TEST DI IPOTESI:

Contributo parziale: 211.4
 Fempirico: 898.59
 Fteorico: 3.8893
 gradi di libertà: (1 , 196)

Rifiutare H_0 con ampiezza 0.05

7) `lm(formula = Y ~ X + Z + X : Z)`

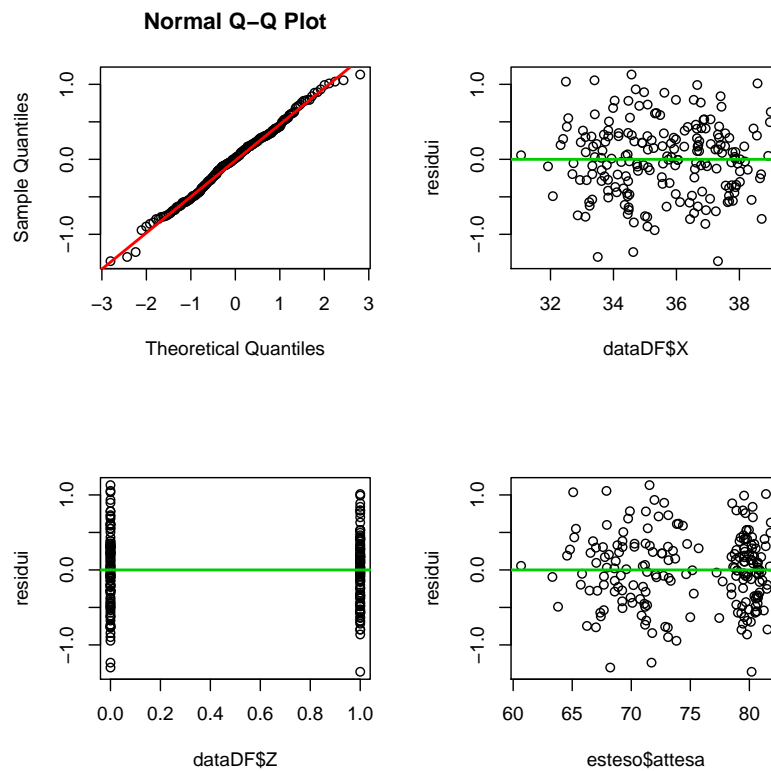


Figura 25: Analisi dei residui

8) La Figura 25 mostra graficamente l'analisi dei residui del modello selezionato.

Esercizio 5.3.16

```

1) (Intercept)                4.821797
   conservante1                0.797862
   correttore1                 0.033750
   esaltatore1                -0.155252
   conservante1:correttore1    -0.305104
   conservante1:esaltatore1    -0.380706
   correttore1:esaltatore1     0.045901

```

2) Posto $\alpha = 0.1$, termina con il modello che ha intercetta, **conservante** e **esaltatore**:

```
lm(formula = Ph ~ conservante + esaltatore)
```

3) Caricare il *package* `car` attraverso il comando `library(package="car")` .

```
Anova Table (Type III tests)
```

```
Response: pH
```

```

              Sum Sq Df F value Pr(>F)
(Intercept)  194.7   1 1053.08 <2e-16 ***
conservante    1.2   1   6.72  0.017 *
esaltatore     0.6   1   3.38  0.080 .
Residuals     3.9  21

```

```
---
```

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

4) La Figura 26 mostra graficamente l'analisi dei residui del modello finale.

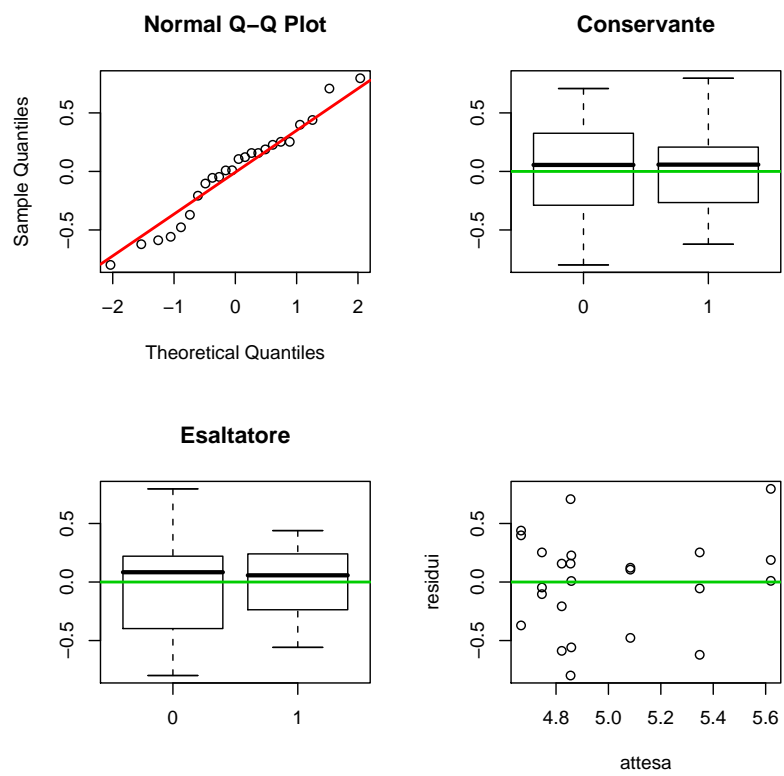


Figura 26: Analisi dei residui