

Multivariate Analysis for the Behavioral Sciences,
Second Edition (Chapman and Hall/CRC, 2019)

**Examples of Chapter 4:
Multiple Linear Regression**

Kimmo Vehkalahti and Brian S. Everitt

9 October 2018

Contents

Examples	2
Figure 4.1: Time Spent Looking After Car	2
Table 4.1	3
Table 4.2	4
Figure 4.2	5
Table 4.3	6
Figure 4.3	7
Table 4.4: Course Evaluation Data	8
Figure 4.4	9
Table 4.5	10
Table 4.6	12
Figure 4.5	14
Figure 4.6	15
Figure 4.7	16
Figure 4.8	17
Table 4.7: Fecundity of Fruit Flies	18
Table 4.8	19
Table 4.9: A Balanced 2 x 2 Data Set	21
Table 4.10: Unbalanced 2 x 2 Data Set	24

Examples

Figure 4.1: Time Spent Looking After Car

```
Gender <- c(1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0,
           0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1)

Age <- c(55, 43, 57, 26, 22, 32, 26, 29, 40, 30, 34, 44, 49, 22, 34, 47,
        48, 48, 22, 24, 50, 49, 49, 48, 29, 58, 24, 21, 29, 45, 28, 37,
        44, 22, 38, 24, 34, 26, 26, 25)

Extroversion <- c(40, 45, 52, 62, 31, 28, 2, 83, 55, 32, 47, 45, 60, 13, 7, 85,
                 38, 61, 26, 3, 29, 60, 47, 18, 16, 36, 24, 12, 32, 46, 26, 40,
                 46, 44, 3, 25, 43, 41, 42, 36)

Time <- c(46, 79, 33, 63, 20, 18, 11, 97, 63, 46, 21, 71, 59, 44, 30, 80,
         45, 26, 33, 7, 50, 54, 73, 19, 36, 31, 71, 15, 40, 61, 45, 42,
         57, 34, 26, 47, 42, 44, 59, 27)

carwash <- as.data.frame(cbind(Gender, Age, Extroversion, Time))
# Time, Age, Extroversion - points by Gender
pairs(carwash[, c(4,2,3)], panel=function(x,y) text(x,y,Gender))
```

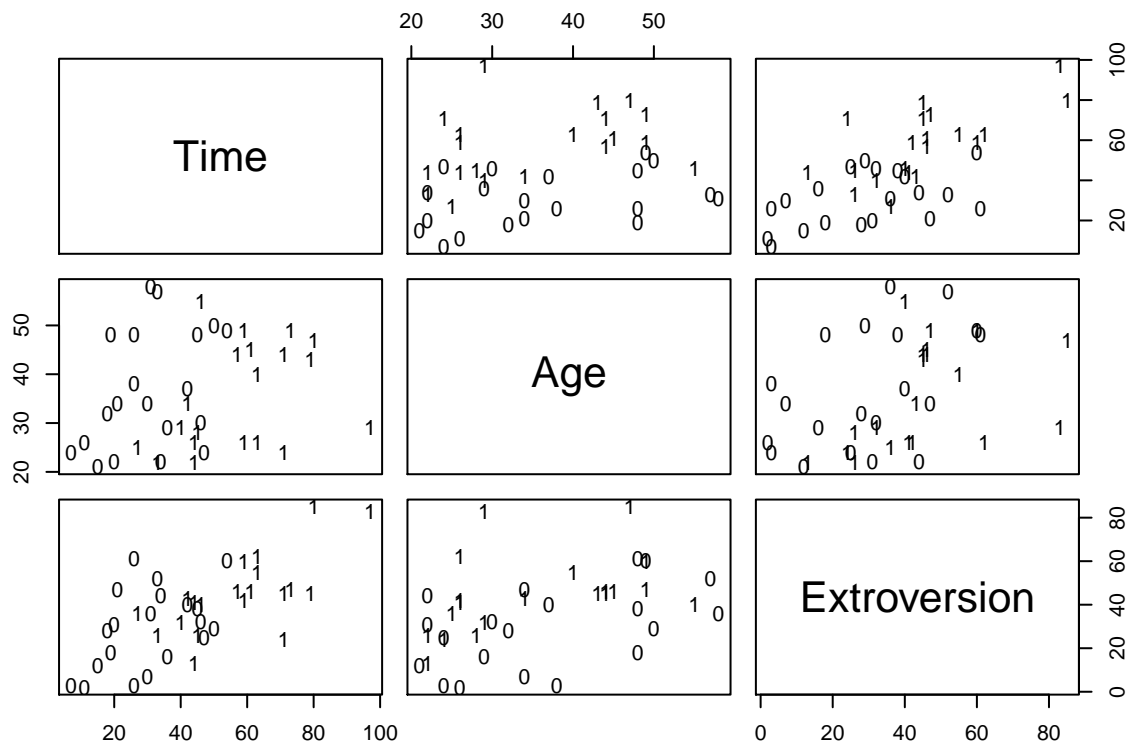


Table 4.1

```
summary(lm(Time ~ Age + Extroversion + Gender, data = carwash))

##
## Call:
## lm(formula = Time ~ Age + Extroversion + Gender, data = carwash)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.981  -9.690  -1.182   8.976  24.746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.3063     7.3153   1.546 0.130956
## Age           0.1556     0.2062   0.754 0.455469
## Extroversion   0.4643     0.1303   3.564 0.001053 **
## Gender        20.0711     4.6514   4.315 0.000119 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.02 on 36 degrees of freedom
## Multiple R-squared:  0.6377, Adjusted R-squared:  0.6076
## F-statistic: 21.13 on 3 and 36 DF,  p-value: 4.569e-08
```

Table 4.2

```
summary(lm(Time ~ Extroversion + Gender, data = carwash))

##
## Call:
## lm(formula = Time ~ Extroversion + Gender, data = carwash)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.193  -9.474  -2.149   10.165   23.918
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   15.6797     4.4365   3.534 0.001118 **
## Extroversion    0.5093     0.1151   4.423 8.24e-05 ***
## Gender         19.1801     4.4727   4.288 0.000124 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.95 on 37 degrees of freedom
## Multiple R-squared:  0.632, Adjusted R-squared:  0.6121
## F-statistic: 31.77 on 2 and 37 DF,  p-value: 9.284e-09
```

Figure 4.2

```
#par(mfrow = c(1,1))
gender <- rep("m", length(Gender))
gender[Gender=="0"] <- "f"
x <- seq(0, 90, length = 50)
ym <- 15.68 + 19.18 + 0.51*x
yf <- 15.68 + 0.51*x
plot(Time ~ Extroversion, xlab = "Extroversion score",
      ylab = "Time spent looking after car (mins)", type = "n")
text(Extroversion, Time, labels = gender)
lines(x, ym, lty = 1)
lines(x, yf, lty = 2)
legend("topleft", c("Male", "Female"), lty = 1:2)
```

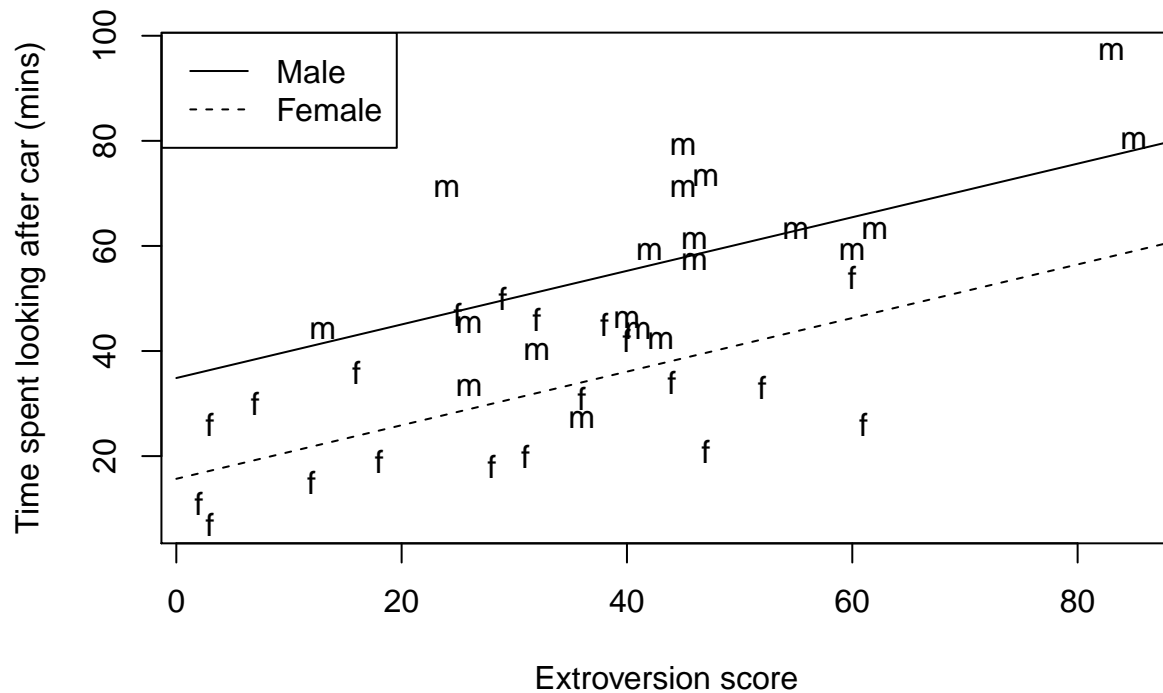


Table 4.3

```
summary(lm(Time ~ Extroversion + Gender + Extroversion:Gender, data = carwash))
```

```
##
## Call:
## lm(formula = Time ~ Extroversion + Gender + Extroversion:Gender,
##     data = carwash)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.807  -9.445  -1.677   10.477   27.183
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      20.0182     5.4560   3.669 0.000782 ***
## Extroversion       0.3607     0.1590   2.268 0.029430 *
## Gender            7.8178     9.5705   0.817 0.419379
## Extroversion:Gender 0.3052     0.2279   1.339 0.188970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.81 on 36 degrees of freedom
## Multiple R-squared:  0.6495, Adjusted R-squared:  0.6203
## F-statistic: 22.23 on 3 and 36 DF,  p-value: 2.548e-08
```

Figure 4.3

```
gender <- rep("m", length(Gender))
gender[Gender==0] <- "f"
x <- seq(0, 90, length = 50)
ym <- 20 + 7.82 + (0.36+0.31)*x
yf <- 20 + 0.36*x
plot(Time ~ Extroversion, xlab = "Extroversion score",
      ylab = "Time spent looking after car (mins)", type = "n")
text(Extroversion, Time, labels = gender)
lines(x, ym, lty = 1)
lines(x, yf, lty = 2)
legend("topleft", c("Male", "Female"), lty = 1:2)
```

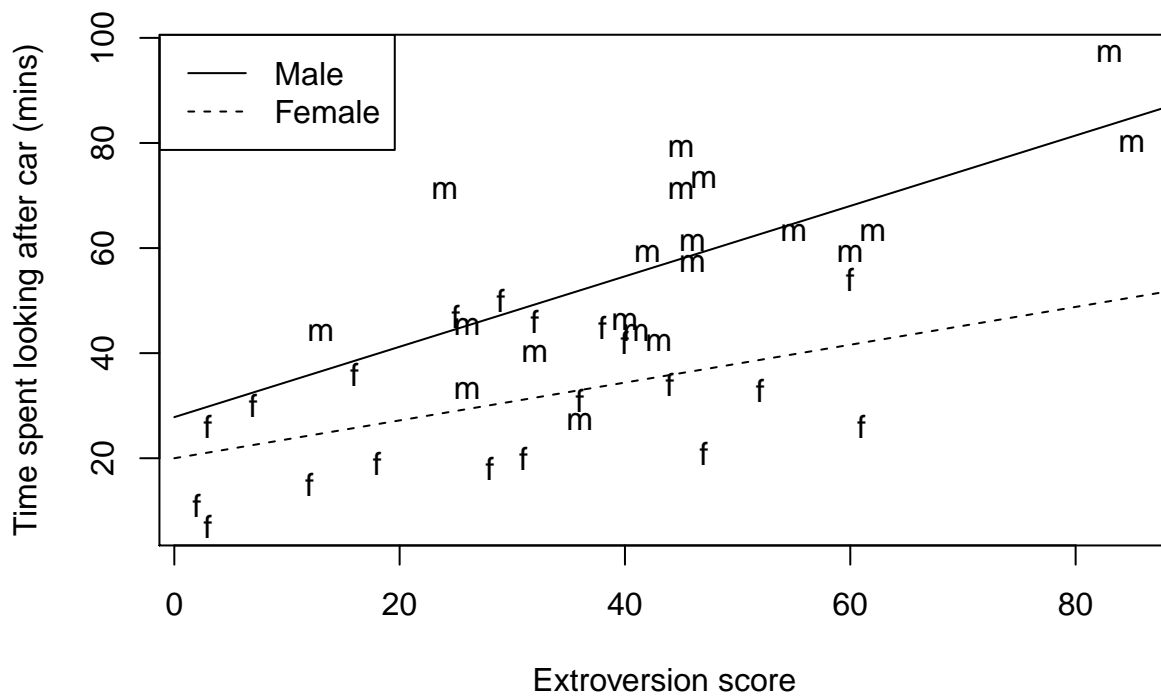


Table 4.4: Course Evaluation Data

```

Overall <- c(3.4, 2.9, 2.6, 3.8, 3.0, 2.5, 3.9, 4.3, 3.8, 3.4,
            2.8, 2.9, 4.1, 2.7, 3.9, 4.1, 4.2, 3.1, 4.1, 3.6,
            4.3, 4.0, 2.1, 3.8, 2.7, 4.4, 3.1, 3.6, 3.9, 2.9,
            3.7, 2.8, 3.3, 3.7, 4.2, 2.9, 3.9, 3.5, 3.8, 4.0,
            3.1, 4.2, 3.0, 4.8, 3.0, 4.4, 4.4, 3.4, 4.0, 3.5)

Teach <- c(3.8, 2.8, 2.2, 3.5, 3.2, 2.7, 4.1, 4.2, 3.7, 3.7,
           3.3, 3.3, 4.1, 3.1, 2.9, 4.5, 4.3, 3.7, 4.2, 4.0,
           3.7, 4.0, 2.9, 4.0, 3.3, 4.4, 3.4, 3.8, 3.7, 3.1,
           3.8, 3.2, 3.5, 3.8, 4.4, 3.7, 4.0, 3.4, 3.2, 3.8,
           3.7, 4.3, 3.4, 4.0, 3.1, 4.5, 4.8, 3.4, 4.2, 3.4)

Exam <- c(3.8, 3.2, 1.9, 3.5, 2.8, 3.8, 3.8, 4.1, 3.6, 3.6,
          3.5, 3.3, 3.6, 3.8, 3.8, 4.2, 4.1, 4.0, 4.3, 4.2,
          4.0, 4.1, 2.7, 4.4, 4.4, 4.3, 3.6, 4.1, 4.2, 3.6,
          4.4, 3.4, 3.2, 3.7, 4.3, 4.1, 3.7, 4.0, 3.6, 4.0,
          3.7, 4.2, 4.2, 4.1, 3.2, 4.5, 4.3, 3.6, 4.0, 3.9)

Knowledge <- c(4.5, 3.8, 3.9, 4.1, 3.5, 4.2, 4.5, 4.7, 4.1, 4.1,
              3.9, 3.9, 4.0, 4.1, 4.5, 4.5, 4.5, 4.5, 4.7, 4.0,
              4.5, 4.6, 3.7, 4.1, 3.6, 4.4, 3.3, 3.8, 4.2, 3.8,
              4.0, 3.1, 4.4, 4.3, 5.0, 4.2, 4.5, 4.5, 4.7, 4.3,
              4.0, 4.2, 3.8, 4.9, 3.7, 4.6, 4.3, 3.5, 4.4, 4.4)

Grade <- c(3.5, 3.2, 2.8, 3.3, 3.2, 3.2, 3.6, 4.0, 3.0, 3.1,
           3.0, 3.3, 3.2, 3.4, 3.7, 3.8, 3.8, 3.7, 4.2, 3.8,
           3.3, 3.2, 3.1, 3.9, 4.3, 2.9, 3.2, 3.5, 3.3, 3.2,
           4.1, 3.5, 3.6, 3.7, 3.3, 3.6, 3.5, 3.4, 3.0, 3.4,
           3.7, 3.8, 3.7, 3.7, 3.3, 4.0, 3.6, 3.3, 4.1, 3.3)

Enroll <- c(21, 50, 800, 221, 7, 108, 54, 99, 52, 47, 73, 25,
            37, 83, 70, 16, 14, 12, 20, 18, 260, 100, 118, 35,
            32, 25, 55, 28, 28, 27, 25, 50, 76, 28, 85, 75, 90,
            94, 65, 100, 105, 70, 49, 64, 700, 27, 15, 40, 18, 90)

courses <- as.data.frame(cbind(Overall, Teach, Exam, Knowledge, Grade, Enroll))
head(courses)

## Overall Teach Exam Knowledge Grade Enroll
## 1      3.4   3.8  3.8      4.5   3.5     21
## 2      2.9   2.8  3.2      3.8   3.2     50
## 3      2.6   2.2  1.9      3.9   2.8    800
## 4      3.8   3.5  3.5      4.1   3.3    221
## 5      3.0   3.2  2.8      3.5   3.2      7
## 6      2.5   2.7  3.8      4.2   3.2    108

```


Figure 4.4

```
pairs(courses, panel=function(x,y) {
  points(x,y)
  abline(lm(y ~ x))
  lines(lowess(y ~ x), lty = 3)
})
```

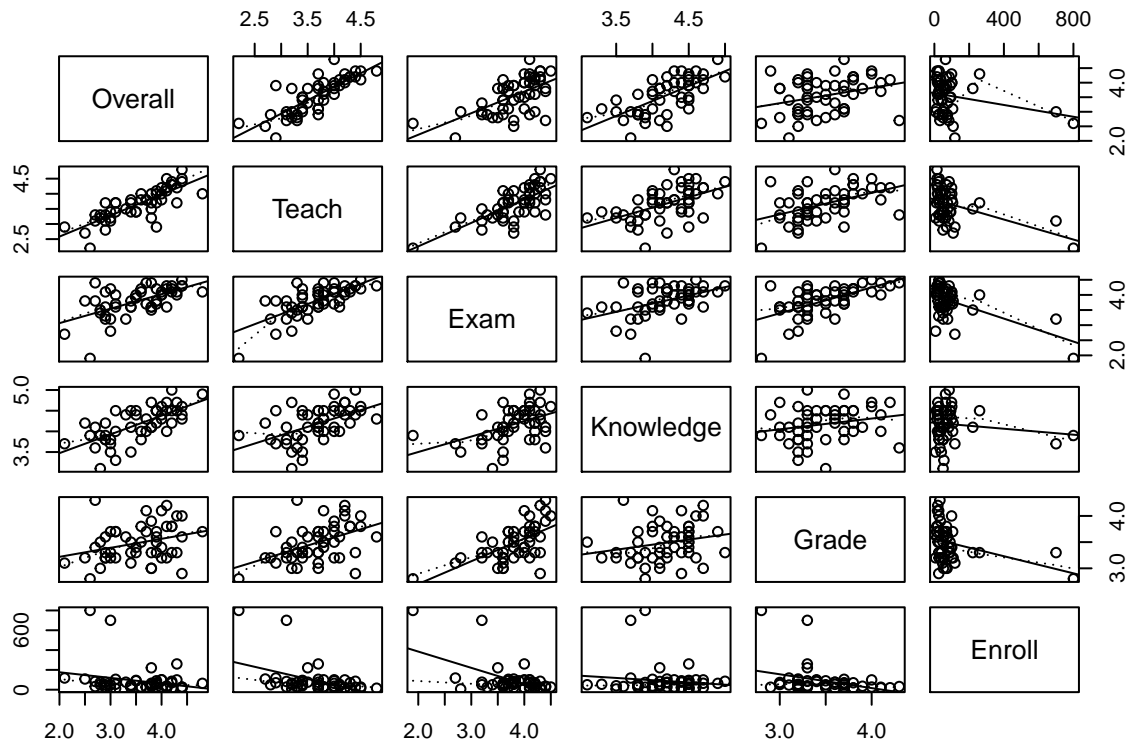


Table 4.5

```
reg <- lm(Overall ~ Teach + Exam + Knowledge + Grade + Enroll)
summary(reg)

##
## Call:
## lm(formula = Overall ~ Teach + Exam + Knowledge + Grade + Enroll)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.69991	-0.16255	0.01672	0.20513	0.82465

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.1951810	0.6311922	-1.894	0.064875 .
Teach	0.7632345	0.1329150	5.742	8.06e-07 ***
Exam	0.1320347	0.1627995	0.811	0.421716
Knowledge	0.4889675	0.1365333	3.581	0.000849 ***
Grade	-0.1842549	0.1654897	-1.113	0.271586
Enroll	0.0005259	0.0003901	1.348	0.184555

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3202 on 44 degrees of freedom
## Multiple R-squared:  0.7555, Adjusted R-squared:  0.7277
## F-statistic: 27.19 on 5 and 44 DF,  p-value: 1.977e-12
```

```

# remove outliers
reg_345 <- lm(Overall[-c(3, 45)] ~ Teach[-c(3, 45)] + Exam[-c(3, 45)] +
             Knowledge[-c(3, 45)] + Grade[-c(3, 45)] + Enroll[-c(3, 45)])
summary(reg_345)

##
## Call:
## lm(formula = Overall[-c(3, 45)] ~ Teach[-c(3, 45)] + Exam[-c(3,
##      45)] + Knowledge[-c(3, 45)] + Grade[-c(3, 45)] + Enroll[-c(3,
##      45)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71525 -0.15273  0.02601  0.20149  0.84484
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.245711    0.643199  -1.937  0.05952 .
## Teach[-c(3, 45)]    0.788694    0.138463   5.696 1.09e-06 ***
## Exam[-c(3, 45)]     0.135633    0.173513   0.782  0.43878
## Knowledge[-c(3, 45)] 0.446810    0.146940   3.041  0.00405 **
## Grade[-c(3, 45)]    -0.162200    0.172142  -0.942  0.35146
## Enroll[-c(3, 45)]    0.001254    0.001072   1.169  0.24898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3248 on 42 degrees of freedom
## Multiple R-squared:  0.7422, Adjusted R-squared:  0.7116
## F-statistic: 24.19 on 5 and 42 DF,  p-value: 2.233e-11

# VIFs
s1 <- summary(lm(Teach ~ Exam + Knowledge + Grade + Enroll))
s2 <- summary(lm(Exam ~ Teach + Knowledge + Grade + Enroll))
s3 <- summary(lm(Knowledge ~ Exam + Teach + Grade + Enroll))
s4 <- summary(lm(Grade ~ Exam + Knowledge + Teach + Enroll))
s5 <- summary(lm(Enroll ~ Exam + Knowledge + Grade + Teach))
#
rsq <- c(s1$r.squared, s2$r.squared, s3$r.squared, s4$r.squared, s5$r.squared)
rsq

## [1] 0.5817980 0.6754306 0.3253565 0.3803019 0.3466162

# rounded values used in the book (no difference in practice):
# rsq <- c(0.58, 0.68, 0.33, 0.38, 0.35)
1/(1 - rsq)

## [1] 2.391189 3.081005 1.482264 1.613689 1.530494

```

Table 4.6

```
step(reg, direction = "backward")

## Start:  AIC=-108.28
## Overall ~ Teach + Exam + Knowledge + Grade + Enroll
##
##           Df Sum of Sq   RSS   AIC
## - Exam      1    0.0674 4.5779 -109.539
## - Grade      1    0.1271 4.6376 -108.891
## <none>                4.5105 -108.281
## - Enroll     1    0.1863 4.6968 -108.257
## - Knowledge   1    1.3148 5.8253 -97.491
## - Teach      1    3.3802 7.8907 -82.317
##
## Step:  AIC=-109.54
## Overall ~ Teach + Knowledge + Grade + Enroll
##
##           Df Sum of Sq   RSS   AIC
## - Grade      1    0.0736 4.6515 -110.742
## - Enroll     1    0.1299 4.7079 -110.139
## <none>                4.5779 -109.539
## - Knowledge   1    1.5185 6.0965 -97.216
## - Teach      1    4.6408 9.2187 -76.539
##
## Step:  AIC=-110.74
## Overall ~ Teach + Knowledge + Enroll
##
##           Df Sum of Sq   RSS   AIC
## - Enroll     1    0.1665 4.8180 -110.983
## <none>                4.6515 -110.742
## - Knowledge   1    1.5236 6.1751 -98.575
## - Teach      1    4.7841 9.4356 -77.377
##
## Step:  AIC=-110.98
## Overall ~ Teach + Knowledge
##
##           Df Sum of Sq   RSS   AIC
## <none>                4.8180 -110.983
## - Knowledge   1    1.7079 6.5259 -97.812
## - Teach      1    5.0529 9.8709 -77.122
##
## Call:
## lm(formula = Overall ~ Teach + Knowledge)
##
## Coefficients:
## (Intercept)      Teach      Knowledge
##      -1.2984      0.7097      0.5383
```

```

final_reg <- lm(Overall ~ Teach + Knowledge)
summary(final_reg)

##
## Call:
## lm(formula = Overall ~ Teach + Knowledge)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.68847 -0.20335  0.00175  0.20174  0.71779
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.2984     0.4773  -2.720 0.009121 **
## Teach         0.7097     0.1011   7.021 7.6e-09 ***
## Knowledge     0.5383     0.1319   4.082 0.000172 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3202 on 47 degrees of freedom
## Multiple R-squared:  0.7388, Adjusted R-squared:  0.7277
## F-statistic: 66.47 on 2 and 47 DF,  p-value: 1.991e-14

```

Figure 4.5

```
rawresid <- final_reg$residuals
infl.reg <- lm.influence(final_reg)

r.std <- rawresid/sqrt(0.3202^2 * (1 - infl.reg$hat))
r.del <- rawresid/sqrt(infl.reg$sigma^2 * (1 - infl.reg$hat))

par(mfrow=c(2,2))
boxplot(r.std, ylab = "Standardised residual")
qqnorm(r.std, main = "")
boxplot(r.del, ylab = "Deletion residual")
qqnorm(r.del, main = "")
```

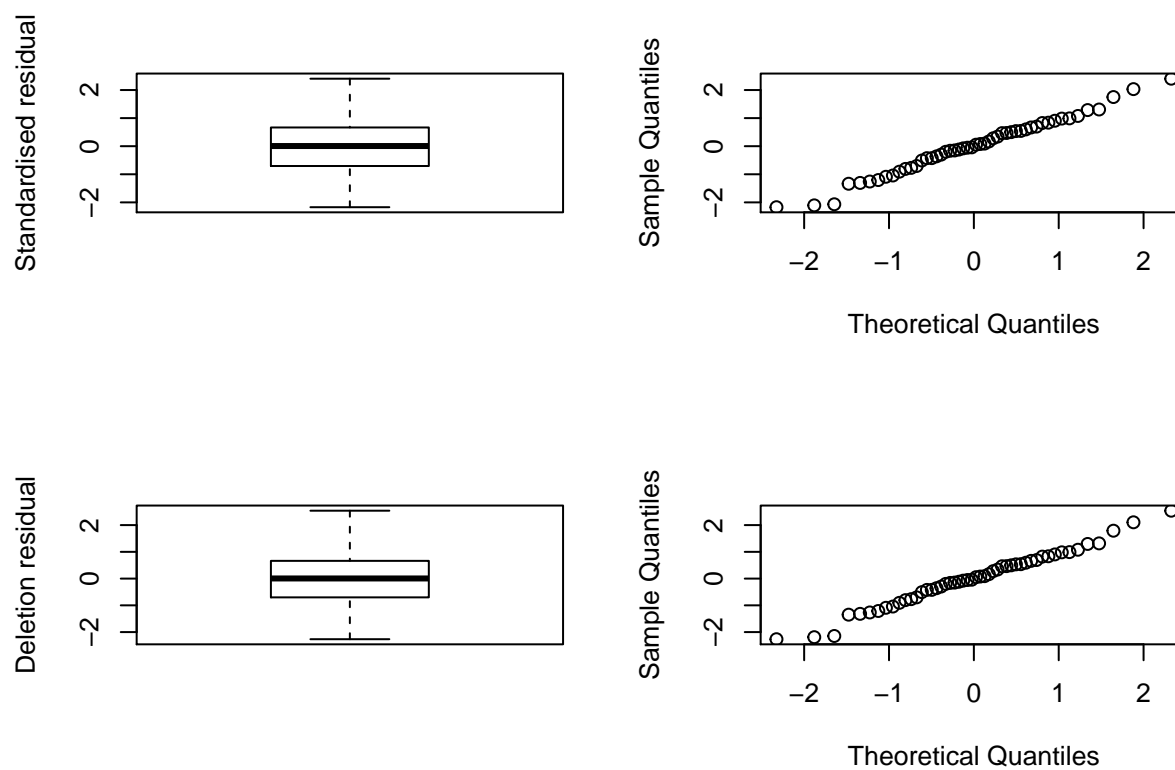


Figure 4.6

```
par(mfrow = c(3,1), mar = c(4, 4, 2, 0.1))
lab <- "Standardised residual"
plot(predict(final_reg), r.std, xlab = "Fitted Value of Overall", ylab = lab)
plot(Teach, r.std, xlab = "Teacher rating", ylab = lab)
plot(Knowledge, r.std, xlab = "Knowledge of Subject Rating", ylab = lab)
```

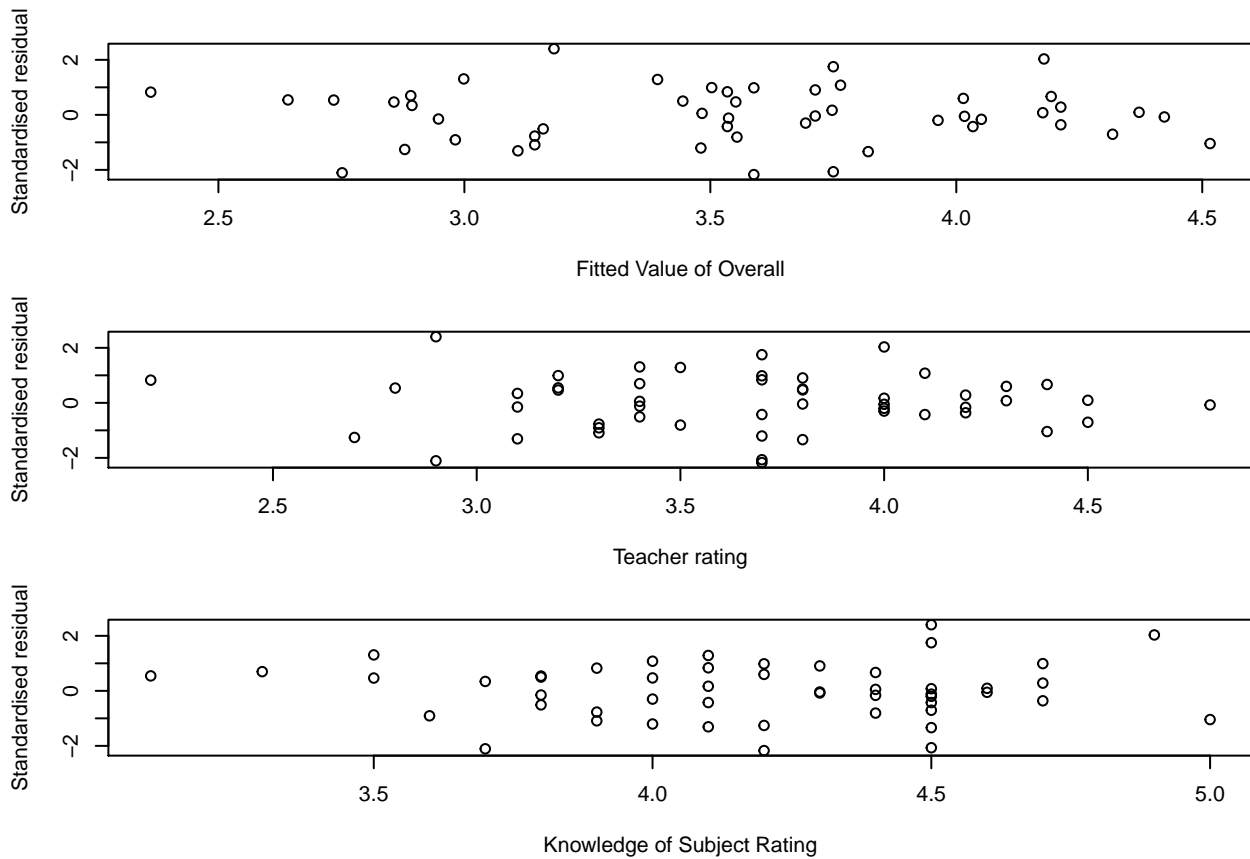


Figure 4.7

```
par(mfrow = c(3,1), mar = c(4, 4, 2, 0.1))
lab <- "Deletion residual"
plot(predict(final_reg), r.del, xlab = "Fitted Value of Overall", ylab = lab)
plot(Teach, r.del, xlab = "Teacher rating", ylab = lab)
plot(Knowledge, r.del, xlab = "Knowledge of Subject Rating", ylab = lab)
```

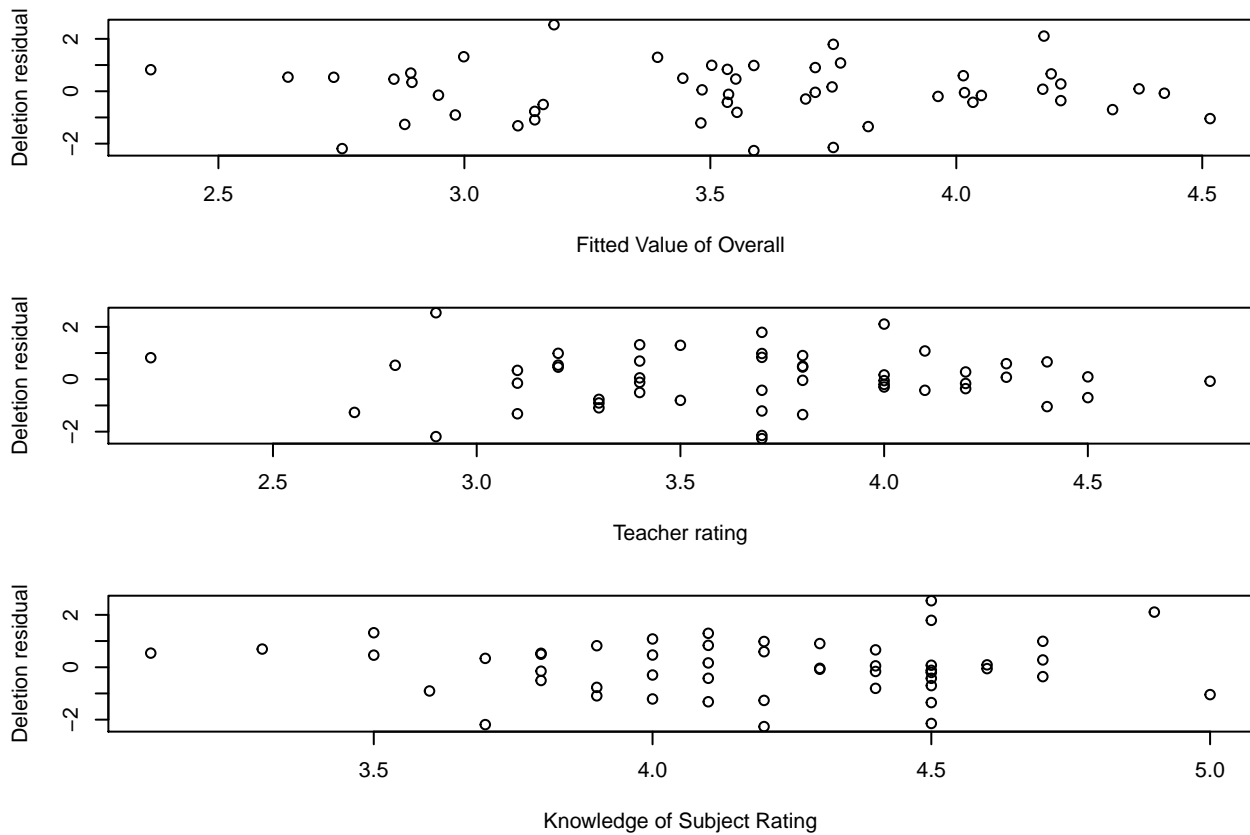


Figure 4.8

```
par(mfrow = c(1,1))
plot(1:50, infl.reg$hat, xlab = "Observation Number", ylab = "Cook's Distance")
segments(1:50, rep(0,50), 1:50, infl.reg$hat)
```

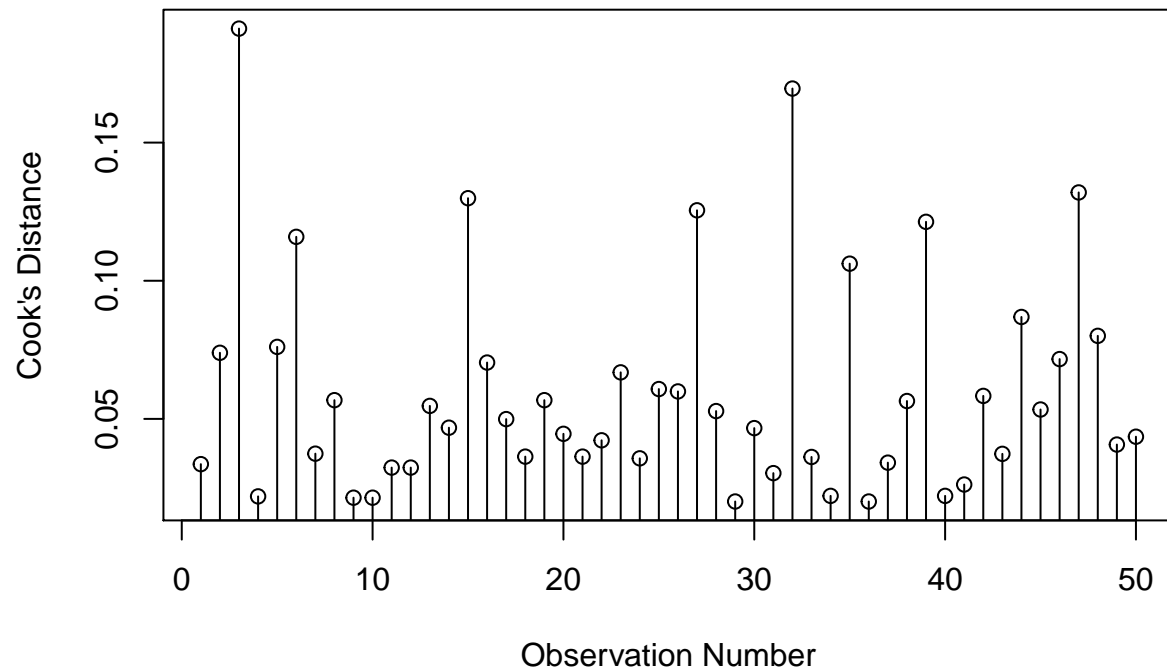


Table 4.7: Fecundity of Fruit Flies

```
flies <- read.table("data/fruitfly.txt", header = T)
flies
```

##		RS	SS	NS
##	1	12.8	38.4	35.4
##	2	21.6	32.9	27.4
##	3	14.8	48.5	19.3
##	4	23.1	20.9	41.8
##	5	34.6	11.6	20.3
##	6	19.7	22.3	37.6
##	7	22.6	30.2	36.9
##	8	29.6	33.4	37.3
##	9	16.4	26.7	28.2
##	10	20.3	39.0	23.4
##	11	29.3	12.8	33.7
##	12	14.9	14.6	29.2
##	13	27.3	12.2	41.7
##	14	22.4	23.1	22.6
##	15	27.5	29.4	40.4
##	16	20.3	16.0	34.4
##	17	38.7	20.1	30.4
##	18	26.4	23.3	14.9
##	19	23.7	22.9	51.8
##	20	26.1	22.5	33.8
##	21	29.5	15.1	37.9
##	22	38.6	31.0	29.5
##	23	44.4	16.9	42.4
##	24	23.2	16.1	36.6
##	25	23.6	10.8	47.4

Table 4.8

```

n <- 25
x1 <- c(rep(1,n), rep(0,n), rep(-1,n))
x2 <- c(rep(0,n), rep(1,n), rep(-1,n))
response <- c(flies$RS, flies$SS, flies$NS)

flies2 <- cbind(x1, x2, response)
head(flies2)

##      x1 x2 response
## [1,]  1  0    12.8
## [2,]  1  0    21.6
## [3,]  1  0    14.8
## [4,]  1  0    23.1
## [5,]  1  0    34.6
## [6,]  1  0    19.7

tail(flies2)

##      x1 x2 response
## [70,] -1 -1    33.8
## [71,] -1 -1    37.9
## [72,] -1 -1    29.5
## [73,] -1 -1    42.4
## [74,] -1 -1    36.6
## [75,] -1 -1    47.4

reg0 <- lm(response ~ 1)

reg <- lm(response ~ x1 + x2)
summary(reg)

##
## Call:
## lm(formula = response ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.472  -5.764  -0.728   4.436  24.872
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   27.419     1.024   26.784  <2e-16 ***
## x1            -2.163     1.448   -1.494   0.1396
## x2            -3.791     1.448   -2.618   0.0108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.866 on 72 degrees of freedom
## Multiple R-squared:  0.194, Adjusted R-squared:  0.1716
## F-statistic: 8.666 on 2 and 72 DF, p-value: 0.0004244
# Residual sum of squares:
sum(reg$residuals^2)

```

```
## [1] 5659.022
```

```
anova(reg0, reg)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: response ~ 1
```

```
## Model 2: response ~ x1 + x2
```

```
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1      74 7021.2
```

```
## 2      72 5659.0  2    1362.2 8.6657 0.0004244 ***
```

```
## ---
```

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

Table 4.9: A Balanced 2 x 2 Data Set

```
x1 <- c(rep(1,8), rep(-1,8))
x2 <- c(rep(1,4), rep(-1,4), rep(1,4), rep(-1,4))
response <- c(23, 25, 27, 29, 26, 32, 30, 31, 22, 23, 21, 21, 37, 38, 40, 35)
```

```
cbind(x1, x2, response)
```

```
##      x1 x2 response
## [1,]  1  1      23
## [2,]  1  1      25
## [3,]  1  1      27
## [4,]  1  1      29
## [5,]  1 -1      26
## [6,]  1 -1      32
## [7,]  1 -1      30
## [8,]  1 -1      31
## [9,] -1  1      22
## [10,] -1  1      23
## [11,] -1  1      21
## [12,] -1  1      21
## [13,] -1 -1      37
## [14,] -1 -1      38
## [15,] -1 -1      40
## [16,] -1 -1      35
```

```
reg0 <- lm(response ~ 1)
reg1 <- lm(response ~ x1)
summary(reg1)
```

```
##
## Call:
## lm(formula = response ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.625 -5.312  0.125  4.438 10.375
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   28.750      1.610   17.855 4.97e-11 ***
## x1            -0.875      1.610   -0.543  0.595
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.441 on 14 degrees of freedom
## Multiple R-squared:  0.02066,    Adjusted R-squared:  -0.0493
## F-statistic: 0.2953 on 1 and 14 DF,  p-value: 0.5954
```

```
# Residual sum of squares:
sum(reg1$residuals^2)
```

```
## [1] 580.75
```

```

anova(reg0, reg1)

## Analysis of Variance Table
##
## Model 1: response ~ 1
## Model 2: response ~ x1
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      15 593.00
## 2      14 580.75  1      12.25 0.2953 0.5954

#
reg2 <- lm(response ~ x1 + x2)
summary(reg2)

##
## Call:
## lm(formula = response ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.750 -2.750 -0.375  2.750  6.000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28.7500     0.9818   29.283 2.96e-13 ***
## x1           -0.8750     0.9818   -0.891 0.389011
## x2           -4.8750     0.9818   -4.965 0.000258 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.927 on 13 degrees of freedom
## Multiple R-squared:  0.6619, Adjusted R-squared:  0.6099
## F-statistic: 12.72 on 2 and 13 DF,  p-value: 0.0008687

sum(reg2$residuals^2)

## [1] 200.5

anova(reg1, reg2)

## Analysis of Variance Table
##
## Model 1: response ~ x1
## Model 2: response ~ x1 + x2
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      14 580.75
## 2      13 200.50  1      380.25 24.655 0.0002583 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#
reg3 <- lm(response ~ x1 + x2 + x1:x2)
summary(reg3)

##
## Call:
## lm(formula = response ~ x1 + x2 + x1:x2)

```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.7500 -0.8125  0.2500  1.2500  3.0000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28.7500     0.5425  52.999 1.34e-15 ***
## x1           -0.8750     0.5425  -1.613  0.13272
## x2           -4.8750     0.5425  -8.987 1.12e-06 ***
## x1:x2         3.0000     0.5425   5.530  0.00013 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.17 on 12 degrees of freedom
## Multiple R-squared:  0.9047, Adjusted R-squared:  0.8809
## F-statistic: 37.98 on 3 and 12 DF,  p-value: 2.102e-06
sum(reg3$residuals^2)

## [1] 56.5
anova(reg2, reg3)

## Analysis of Variance Table
##
## Model 1: response ~ x1 + x2
## Model 2: response ~ x1 + x2 + x1:x2
##   Res.Df  RSS Df Sum of Sq    F    Pr(>F)
## 1      13 200.5
## 2      12  56.5  1      144 30.584 0.0001298 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 4.10: Unbalanced 2 x 2 Data Set

```
x1 <- c(rep(1,21), rep(-1,21))
x2 <- c(rep(1,8), rep(-1,13), rep(1,8), rep(-1,13))
response <- c(23, 25, 27, 29, 30, 27, 23, 25,
              26, 32, 30, 31, NA, NA, NA, NA, NA, NA, NA, NA, NA,
              22, 23, 21, 21, 19, 23, 17, NA, 37, 38, 40, 35, 39,
              35, 38, 41, 32, 36, 40, 41, 38)

cbind(x1, x2, response)
```

```
##      x1 x2 response
## [1,]  1  1      23
## [2,]  1  1      25
## [3,]  1  1      27
## [4,]  1  1      29
## [5,]  1  1      30
## [6,]  1  1      27
## [7,]  1  1      23
## [8,]  1  1      25
## [9,]  1 -1      26
## [10,] 1 -1      32
## [11,] 1 -1      30
## [12,] 1 -1      31
## [13,] 1 -1      NA
## [14,] 1 -1      NA
## [15,] 1 -1      NA
## [16,] 1 -1      NA
## [17,] 1 -1      NA
## [18,] 1 -1      NA
## [19,] 1 -1      NA
## [20,] 1 -1      NA
## [21,] 1 -1      NA
## [22,] -1  1      22
## [23,] -1  1      23
## [24,] -1  1      21
## [25,] -1  1      21
## [26,] -1  1      19
## [27,] -1  1      23
## [28,] -1  1      17
## [29,] -1  1      NA
## [30,] -1 -1      37
## [31,] -1 -1      38
## [32,] -1 -1      40
## [33,] -1 -1      35
## [34,] -1 -1      39
## [35,] -1 -1      35
## [36,] -1 -1      38
## [37,] -1 -1      41
## [38,] -1 -1      32
## [39,] -1 -1      36
## [40,] -1 -1      40
## [41,] -1 -1      41
```



```
## [42,] -1 -1      38
reg0 <- lm(response ~ 1)
reg1 <- lm(response ~ x1)
summary(reg1)

##
## Call:
## lm(formula = response ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.800  -4.333   2.167   5.450   9.200
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   29.567     1.294   22.858 <2e-16 ***
## x1            -2.233     1.294   -1.727  0.0945 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.085 on 30 degrees of freedom
## (10 observations deleted due to missingness)
## Multiple R-squared:  0.09039,    Adjusted R-squared:  0.06007
## F-statistic: 2.981 on 1 and 30 DF,  p-value: 0.09453
# Residual sum of squares:
sum(reg1$residuals^2)

## [1] 1505.867
anova(reg0, reg1)

## Analysis of Variance Table
##
## Model 1: response ~ 1
## Model 2: response ~ x1
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      31 1655.5
## 2      30 1505.9  1    149.63 2.981 0.09453 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
reg2 <- lm(response ~ x1 + x2)
summary(reg2)

##
## Call:
## lm(formula = response ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3025  -3.0300  -0.1663   3.1749   6.6513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 29.6663      0.7387 40.158 < 2e-16 ***
## x1          -0.3406      0.7762 -0.439  0.664
## x2          -5.9769      0.7530 -7.938 9.39e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.046 on 29 degrees of freedom
## (10 observations deleted due to missingness)
## Multiple R-squared:  0.7133, Adjusted R-squared:  0.6935
## F-statistic: 36.07 on 2 and 29 DF,  p-value: 1.358e-08
```

```
sum(reg2$residuals^2)
```

```
## [1] 474.6513
```

```
anova(reg1, reg2)
```

```
## Analysis of Variance Table
##
## Model 1: response ~ x1
## Model 2: response ~ x1 + x2
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      30 1505.87
## 2      29  474.65  1    1031.2 63.005 9.387e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#
reg3 <- lm(response ~ x1 + x2 + x1:x2)
summary(reg3)
```

```
##
## Call:
## lm(formula = response ~ x1 + x2 + x1:x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.6923 -1.7335  0.3077  2.1429  3.8750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28.6061      0.4905  58.315 < 2e-16 ***
## x1          -0.6686      0.4905  -1.363  0.184
## x2          -5.1150      0.4905 -10.427 3.79e-11 ***
## x1:x2         3.3025      0.4905   6.732 2.61e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.544 on 28 degrees of freedom
## (10 observations deleted due to missingness)
## Multiple R-squared:  0.8905, Adjusted R-squared:  0.8788
## F-statistic: 75.91 on 3 and 28 DF,  p-value: 1.46e-13
```

```
sum(reg3$residuals^2)
```

```
## [1] 181.2514
```

```
anova(reg2, reg3)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: response ~ x1 + x2
```

```
## Model 2: response ~ x1 + x2 + x1:x2
```

```
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
```

```
## 1      29 474.65
```

```
## 2      28 181.25  1      293.4 45.325 2.613e-07 ***
```

```
## ---
```

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