Regession interpretation 5 - 7

Oh SukJu

Contents

0.1 5. Diagnostics for multiple regression

```
# page 4
load("/Users/ohsukju/Downloads/data/hills.Rdata")
head(races.table)
```

```
##          Race Distance Climb   Time
## 1 Greenmantle      2.5   650 16.083
## 2    Carnethy      6.0  2500 48.350
## 3 CraigDunain      6.0   900 33.650
## 4      BenRha      7.5   800 45.600
## 5   BenLomond      8.0  3070 62.267
## 6    Goatfell      8.0  2866 73.217
```

```
# page 5
races.lm = lm(Time ~ Distance + Climb, data=races.table)
summary(races.lm)
```

```
##
## Call:
## lm(formula = Time ~ Distance + Climb, data = races.table)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.215  -7.129  -1.186   2.371  65.121
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.992039   4.302734  -2.090   0.0447 *
## Distance     6.217956   0.601148  10.343 9.86e-12 ***
## Climb        0.011048   0.002051   5.387 6.45e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 14.68 on 32 degrees of freedom
## Multiple R-squared:  0.9191, Adjusted R-squared:  0.914
## F-statistic: 181.7 on 2 and 32 DF,  p-value: < 2.2e-16
```

```
X = rnorm(100)
Y = 2 * X + 0.5 + rnorm(100)
cutoff = qt(0.95, 97)
sum(abs(rstudent(lm(Y~X))) > cutoff)
```

```
## [1] 9
```

```
X = rnorm(100)
Y = 2 * X + 0.5 + rnorm(100)
cutoff = qt(0.95, 97)
sum(abs(rstudent(lm(Y~X))) > cutoff)
```

```
## [1] 11
```

```
# page 29
n = nrow(races.table)
cutoff = qt(1 - 0.05 / (2*n), (n-4))
races.table[which(abs(rstudent(races.lm)) > cutoff),]
```

```
##         Race Distance Climb  Time
## 18 KnockHill        3   350 78.65
```

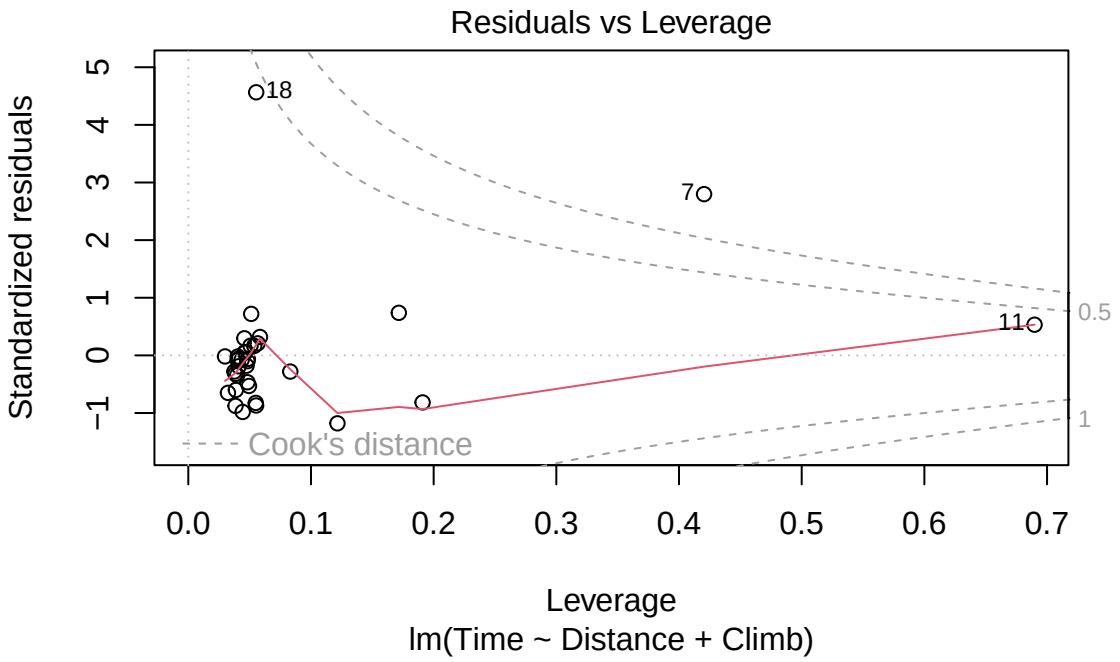
```
library(car)
```

```
## Loading required package: carData
```

```
outlierTest(races.lm)
```

```
##    rstudent unadjusted p-value Bonferroni p
## 18 7.610845         1.3973e-08   4.8905e-07
```

```
# page 30
plot(races.lm, which=5)
```

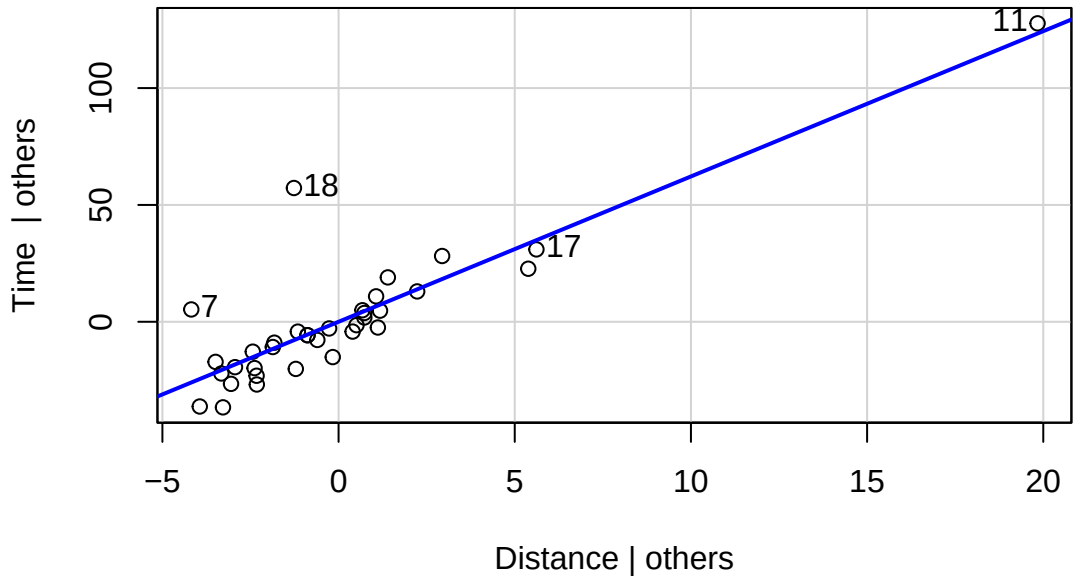



```
# page 31
influence.measures(races.lm)
```

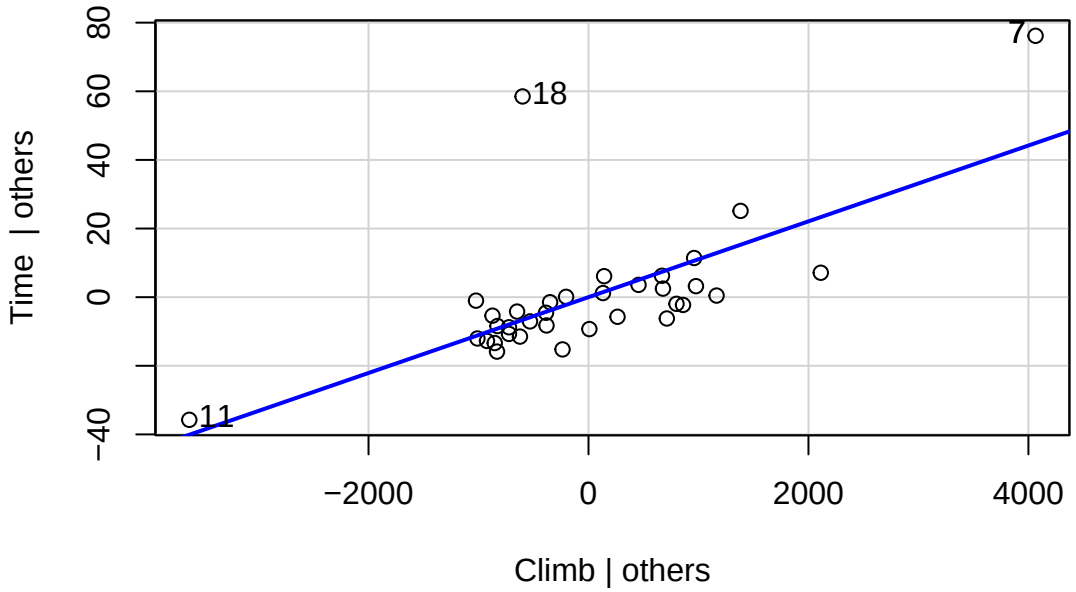
```
## Influence measures of
##   lm(formula = Time ~ Distance + Climb, data = races.table) :
##
##      dfb.1_  dfb.Dstn  dfb.Clmb    dffit  cov.r   cook.d    hat inf
## 1   0.03781 -0.016614 -0.004744  0.03862 1.1595 5.13e-04 0.0538
## 2  -0.05958  0.067215 -0.073396 -0.11956 1.1269 4.88e-03 0.0495
## 3  -0.04858 -0.006707  0.028033 -0.06310 1.1329 1.37e-03 0.0384
## 4  -0.00766 -0.005675  0.008764 -0.01367 1.1556 6.43e-05 0.0485
## 5  -0.05046  0.084709 -0.145005 -0.20947 1.0837 1.47e-02 0.0553
## 6   0.00348 -0.004316  0.007576  0.01221 1.1536 5.13e-05 0.0468
## 7  -0.89065 -0.712774  2.364618  2.69909 0.8178 1.89e+00 0.4204   *
## 8  -0.00844 -0.001648  0.005562 -0.01115 1.1467 4.28e-05 0.0410
## 9  -0.01437  0.000913  0.006161 -0.01663 1.1453 9.52e-05 0.0403
## 10  0.04703  0.013057 -0.036519  0.06399 1.1431 1.41e-03 0.0457
## 11 -0.30118  0.768716 -0.479849  0.78569 3.4525 2.11e-01 0.6898   *
## 12 -0.01149  0.009656 -0.007488 -0.01672 1.1492 9.61e-05 0.0435
## 13 -0.03173 -0.029911 -0.000707 -0.11770 1.0922 4.70e-03 0.0323
## 14  0.11803  0.042034 -0.104884  0.16610 1.1039 9.34e-03 0.0513
## 15 -0.10038  0.057701 -0.022317 -0.11920 1.1062 4.83e-03 0.0388
## 16 -0.01852  0.006789 -0.099862 -0.21135 1.0501 1.49e-02 0.0444
## 17  0.01196 -0.066505  0.034455 -0.08337 1.1908 2.39e-03 0.0831
## 18  1.75827 -0.406545 -0.655934  1.84237 0.0493 4.07e-01 0.0554   *
## 19 -0.15889  0.044311  0.029414 -0.17484 1.0635 1.03e-02 0.0385
## 20  0.00866  0.001424 -0.005946  0.01102 1.1526 4.18e-05 0.0459
## 21  0.04777 -0.010019 -0.019199  0.05032 1.1611 8.70e-04 0.0566
## 22 -0.01889  0.013856 -0.006465 -0.02234 1.1546 1.72e-04 0.0483
```

```
## 23 -0.04131  0.034097 -0.033022 -0.06961 1.1326 1.66e-03 0.0398
## 24  0.07483 -0.046385  0.006428  0.07839 1.1571 2.11e-03 0.0584
## 25  0.03691 -0.012633 -0.008257  0.03808 1.1557 4.99e-04 0.0507
## 26 -0.13772  0.136124 -0.101306 -0.19782 1.0914 1.32e-02 0.0550
## 27 -0.02920 -0.005702  0.019239 -0.03857 1.1431 5.11e-04 0.0410
## 28 -0.04764  0.006936  0.014990 -0.05446 1.1345 1.02e-03 0.0376
## 29 -0.00214  0.000647 -0.000328 -0.00309 1.1338 3.29e-06 0.0299
## 30 -0.08532 -0.007705  0.054838 -0.10362 1.1323 3.67e-03 0.0482
## 31  0.02099  0.170124 -0.373634 -0.44138 1.0960 6.41e-02 0.1216
## 32 -0.02858 -0.008694  0.023275 -0.03931 1.1513 5.31e-04 0.0475
## 33 -0.15823  0.097014  0.155702  0.33384 1.2609 3.77e-02 0.1716
## 34 -0.00356  0.000704  0.001054 -0.00392 1.1461 5.29e-06 0.0403
## 35  0.20872 -0.199048 -0.100907 -0.39445 1.2764 5.24e-02 0.1910
```

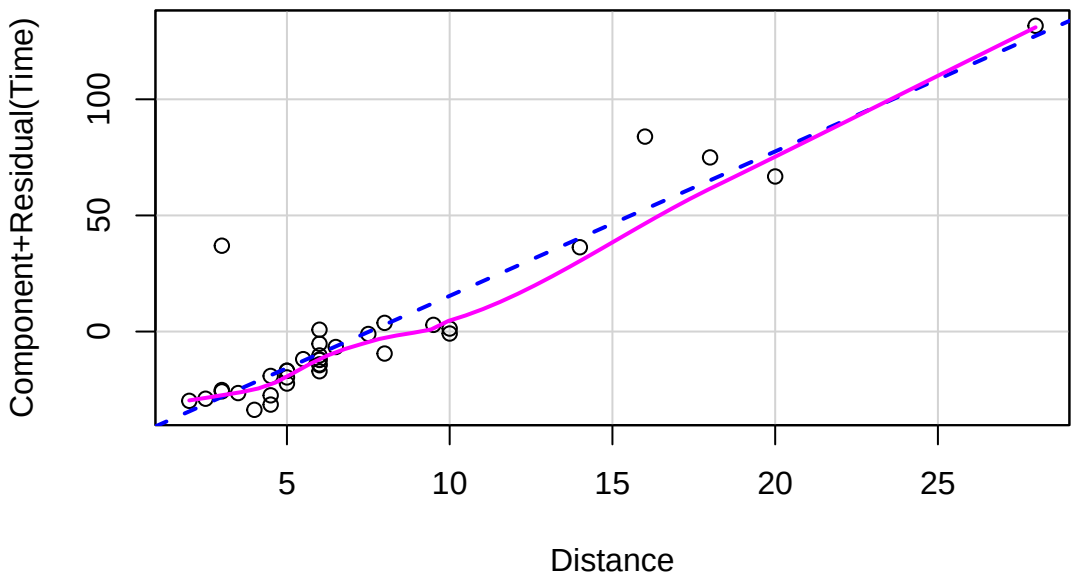
```
# page 34
avPlots(races.lm, 'Distance')
```



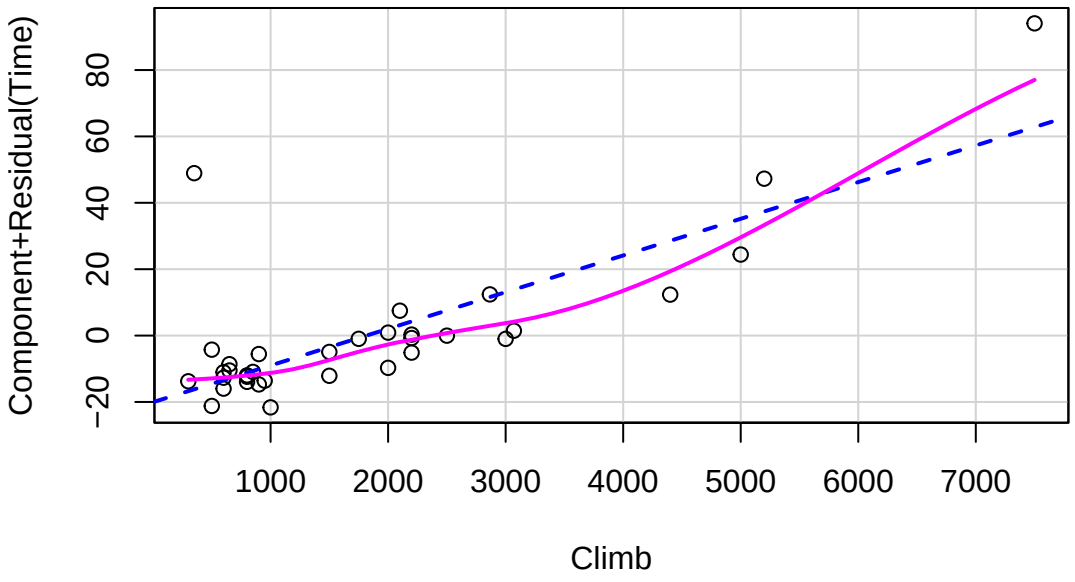

```
avPlots(races.lm, 'Climb')
```



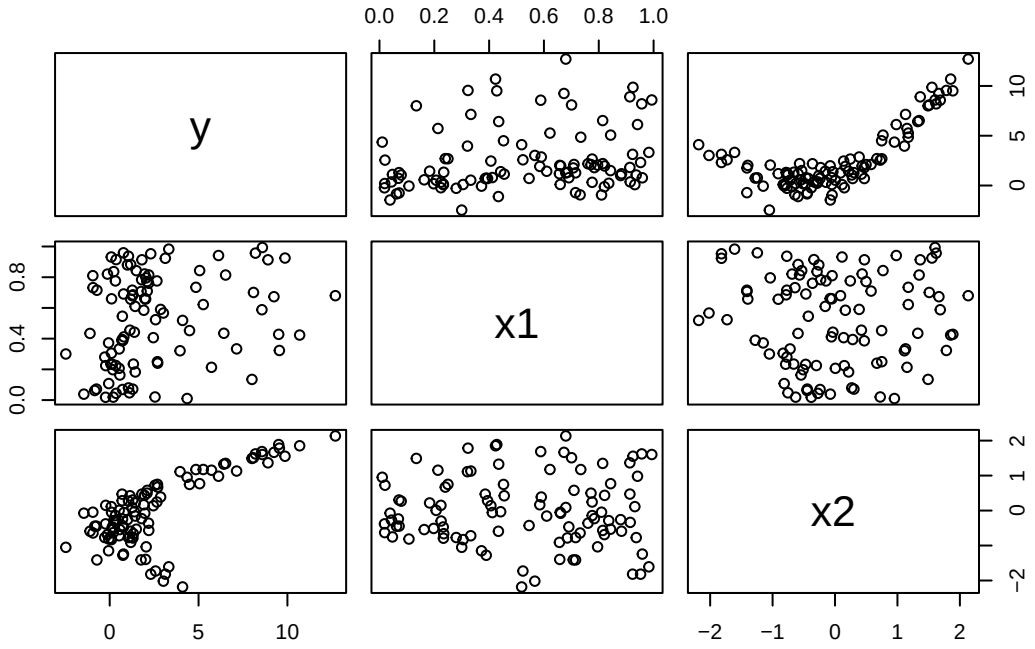

```
# page 36
library(car)
crPlots(races.lm, 'Distance')
```



```
crPlots(races.lm, 'Climb')
```




```
### simulation
set.seed(101)
ns<-100
x1<-runif(ns)
x2<-rnorm(ns)
tr.X.mat<-cbind(rep(1,ns), x1, x2, x2^2)
tr.beta.vec<-c(0.5, 1, 2, 1.5)
ys<-tr.X.mat%*%tr.beta.vec+rnorm(ns)
sim.data<-data.frame(y=ys, x1=x1, x2=x2, x2.sq=x2^2)
pairs(sim.data[,-4])
```




```
lm.sim1<-lm(y~x1+x2, data=sim.data)
summary(lm.sim1)
```

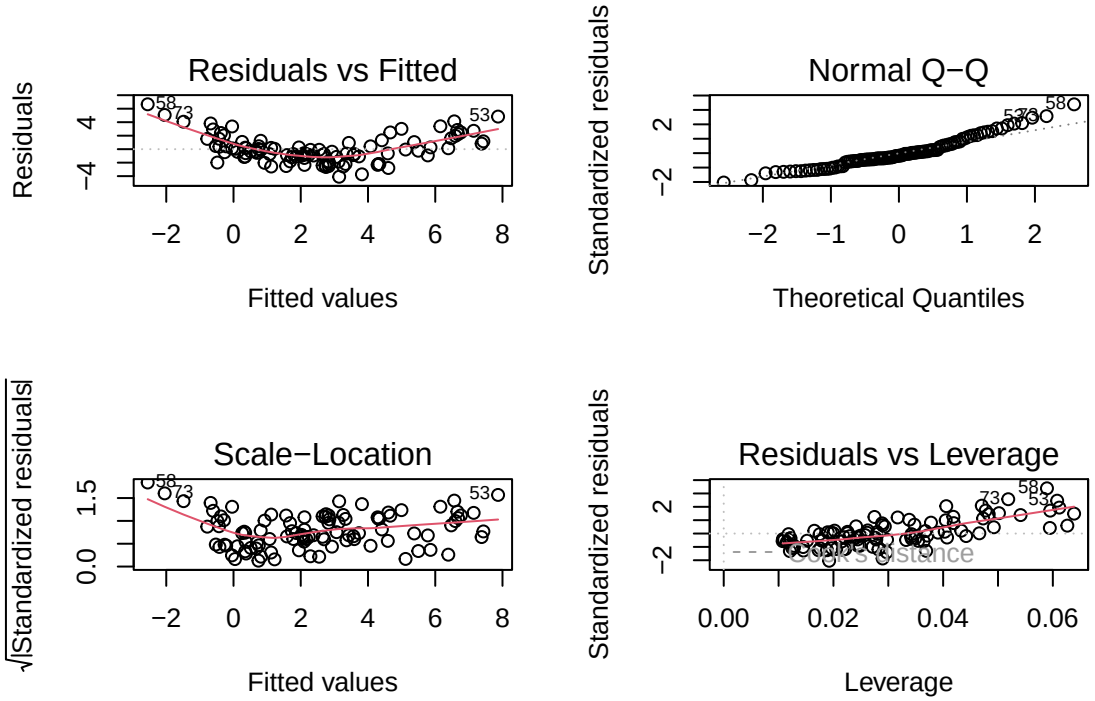
```
##
## Call:
## lm(formula = y ~ x1 + x2, data = sim.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0902 -1.1056 -0.3552  1.0557  6.6538
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.1974     0.4114   2.910 0.004480 **
## x1            2.5349     0.6822   3.716 0.000339 ***
## x2            2.3192     0.2063  11.243  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.025 on 97 degrees of freedom
## Multiple R-squared:  0.5891, Adjusted R-squared:  0.5807
## F-statistic: 69.55 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
library(car)
outlierTest(lm.sim1)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
```

```
##    rstudent unadjusted p-value Bonferroni p
## 58 3.587754         0.00052716     0.052716
```

```
par(mfrow=c(2,2))
plot(lm.sim1)
```



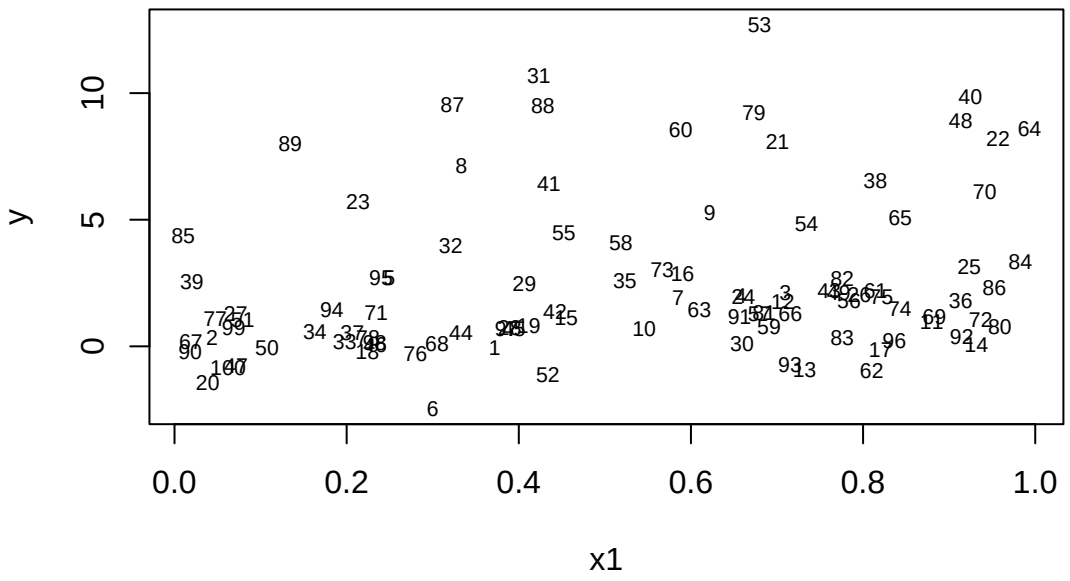

```
influence.measures(lm.sim1)
```

```
## Influence measures of
##   lm(formula = y ~ x1 + x2, data = sim.data) :
##
##        dfb.1_   dfb.x1    dfb.x2    dffit cov.r   cook.d    hat inf
## 1    0.022189 -0.01243 -0.027320  0.03796 1.058 4.85e-04 0.0262
## 2   -0.030226  0.02578  0.004593 -0.03057 1.071 3.15e-04 0.0371
## 3    0.005891 -0.07052 -0.068205 -0.14836 1.011 7.32e-03 0.0176
## 4    0.013678  0.05239 -0.172110  0.21828 1.018 1.58e-02 0.0317
## 5   -0.057473  0.04068 -0.033858 -0.06936 1.051 1.62e-03 0.0244
## 6   -0.116826  0.07792  0.107803 -0.16626 1.028 9.21e-03 0.0272
## 7   -0.017964 -0.01161 -0.010449 -0.05897 1.032 1.17e-03 0.0107
## 8    0.132075 -0.07936  0.145597  0.20912 1.011 1.45e-02 0.0275
## 9   -0.002338 -0.00400 -0.014114 -0.01869 1.058 1.18e-04 0.0258
## 10  -0.018894 -0.00280  0.018397 -0.04738 1.038 7.55e-04 0.0118
## 11   0.050457 -0.11003  0.021764 -0.14552 1.031 7.07e-03 0.0248
## 12  -0.003408  0.06211 -0.148366  0.19262 1.033 1.24e-02 0.0338
## 13   0.013987 -0.08695  0.079734 -0.17341 1.002 9.97e-03 0.0188
## 14   0.134676 -0.26361 -0.028041 -0.32717 0.952 3.48e-02 0.0290
## 15  -0.076132  0.02514 -0.048185 -0.12248 1.007 4.99e-03 0.0125
## 16  -0.010785 -0.00815 -0.014798 -0.03985 1.040 5.34e-04 0.0122
```

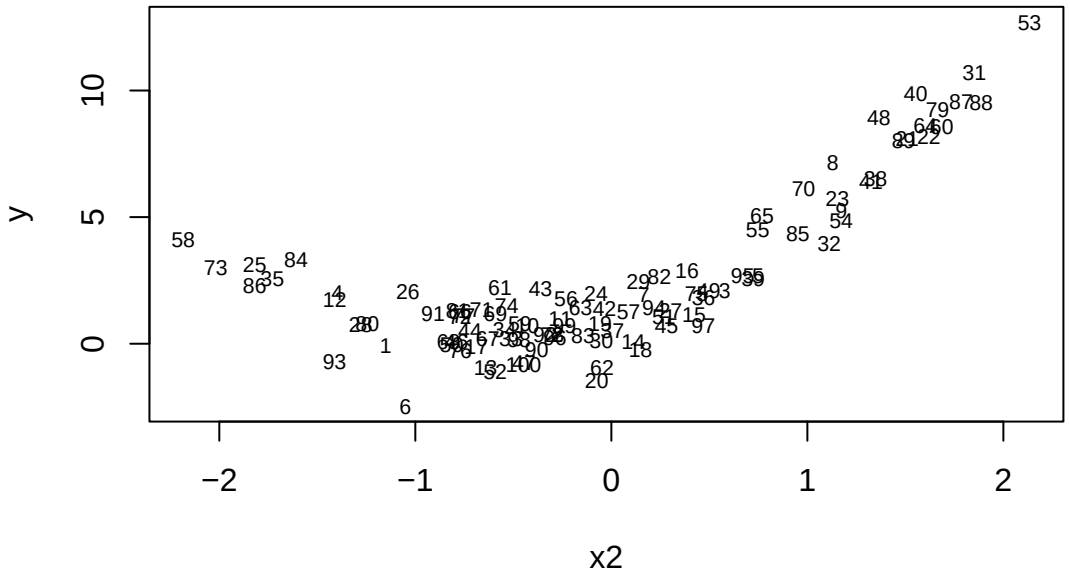
```
## 17   0.033810 -0.09011  0.061653 -0.14295 1.031 6.82e-03 0.0244
## 18  -0.161907  0.11921 -0.017698 -0.16876 1.010 9.46e-03 0.0205
## 19  -0.053699  0.02490  0.003354 -0.06976 1.030 1.63e-03 0.0115
## 20  -0.254221  0.21719  0.011439 -0.25456 1.016 2.14e-02 0.0369
## 21  -0.002643  0.05076  0.129014  0.16093 1.050 8.66e-03 0.0378
## 22  -0.033586  0.06333  0.072118  0.10452 1.091 3.67e-03 0.0595
## 23   0.093850 -0.06932  0.078648  0.12507 1.054 5.25e-03 0.0348
## 24  -0.003349 -0.01633  0.002152 -0.03933 1.040 5.20e-04 0.0121
## 25  -0.135156  0.26764 -0.370408  0.50295 0.976 8.19e-02 0.0612
## 26  -0.018629  0.05662 -0.065091  0.10737 1.050 3.87e-03 0.0290
## 27  -0.072585  0.06076 -0.011938 -0.07378 1.063 1.83e-03 0.0343
## 28   0.069401 -0.03670 -0.100176  0.13139 1.043 5.78e-03 0.0289
## 29  -0.003878  0.00184 -0.000683 -0.00502 1.044 8.50e-06 0.0118
## 30  -0.013091 -0.05997  0.004410 -0.14571 0.989 7.02e-03 0.0121
## 31   0.169417 -0.06877  0.411814  0.47160 0.943 7.16e-02 0.0470
## 32  -0.034538  0.02140 -0.036177 -0.05295 1.058 9.43e-04 0.0276
## 33  -0.024425  0.01863  0.008790 -0.02649 1.057 2.36e-04 0.0249
## 34   0.017500 -0.01379 -0.006277  0.01880 1.061 1.19e-04 0.0280
## 35   0.108037 -0.00636 -0.372189  0.42876 0.941 5.92e-02 0.0406
## 36   0.092712 -0.18773 -0.072789 -0.24607 1.000 2.00e-02 0.0297
## 37  -0.086245  0.06485 -0.000192 -0.08871 1.043 2.64e-03 0.0215
## 38  -0.002502  0.00683  0.009618  0.01359 1.073 6.22e-05 0.0391
## 39  -0.038139  0.03275 -0.014079 -0.04074 1.078 5.59e-04 0.0442
## 40  -0.098416  0.19543  0.230233  0.33222 1.027 3.64e-02 0.0541
## 41   0.039786 -0.01513  0.072214  0.09111 1.054 2.79e-03 0.0294
## 42  -0.031097  0.01183  0.001112 -0.04391 1.037 6.48e-04 0.0108
## 43   0.001001 -0.00402  0.001753 -0.00673 1.050 1.53e-05 0.0175
## 44   0.008666 -0.00537 -0.005970  0.01143 1.052 4.40e-05 0.0195
## 45  -0.093974  0.04690 -0.031584 -0.12170 1.009 4.93e-03 0.0128
## 46   0.008165 -0.00600 -0.004868  0.00977 1.059 3.21e-05 0.0262
## 47  -0.101677  0.08546  0.025736 -0.10495 1.060 3.70e-03 0.0355
## 48  -0.075722  0.15356  0.164347  0.25157 1.041 2.10e-02 0.0473
## 49   0.026024 -0.09321 -0.059011 -0.15618 1.013 8.11e-03 0.0197
## 50   0.032141 -0.02647 -0.015619  0.03581 1.070 4.32e-04 0.0368
## 51  -0.087857  0.07320 -0.012941 -0.08911 1.059 2.67e-03 0.0332
## 52  -0.077656  0.03207  0.060556 -0.12249 1.014 5.00e-03 0.0145
## 53   0.006713  0.14381  0.571867  0.64307 0.905 1.31e-01 0.0608   *
## 54   0.006036 -0.03389 -0.057124 -0.08102 1.056 2.21e-03 0.0298
## 55   0.014555 -0.00485  0.015995  0.02671 1.048 2.40e-04 0.0165
## 56   0.011316 -0.03708  0.008977 -0.05733 1.045 1.11e-03 0.0180
## 57  -0.003510 -0.04857 -0.010128 -0.10524 1.018 3.70e-03 0.0128
## 58   0.193709 -0.01915 -0.818044  0.89780 0.752 2.39e-01 0.0589   *
## 59  -0.000484 -0.02991  0.024661 -0.06678 1.038 1.50e-03 0.0152
## 60   0.030041  0.02399  0.175436  0.20371 1.043 1.38e-02 0.0405
## 61  -0.004510  0.01235 -0.007051  0.01917 1.055 1.24e-04 0.0226
## 62   0.071851 -0.20091  0.004589 -0.29027 0.922 2.72e-02 0.0192
## 63  -0.011511 -0.01339  0.006764 -0.04950 1.036 8.23e-04 0.0110
## 64  -0.054241  0.09746  0.101467  0.15243 1.089 7.80e-03 0.0626
## 65   0.001257 -0.00306 -0.002286 -0.00473 1.061 7.52e-06 0.0280
## 66  -0.000155  0.00158 -0.001920  0.00355 1.053 4.24e-06 0.0200
## 67   0.042369 -0.03667 -0.013986  0.04455 1.077 6.68e-04 0.0433
## 68   0.003073 -0.00202 -0.002269  0.00405 1.055 5.51e-06 0.0226
## 69   0.024781 -0.05360  0.025645 -0.07465 1.055 1.87e-03 0.0278
## 70  -0.009942  0.01918  0.013960  0.02718 1.074 2.49e-04 0.0403
```

```
## 71   0.073083 -0.05356 -0.036584  0.08422 1.048 2.38e-03 0.0241
## 72   0.026194 -0.05076  0.027857 -0.06894 1.065 1.60e-03 0.0349
## 73   0.103907  0.03052 -0.555182  0.61940 0.881 1.20e-01 0.0519   *
## 74   0.013144 -0.03203  0.015554 -0.04679 1.054 7.37e-04 0.0241
## 75   0.044624 -0.11822 -0.055211 -0.17524 1.012 1.02e-02 0.0222
## 76  -0.024166  0.01663  0.015597 -0.03013 1.055 3.06e-04 0.0230
## 77   0.150564 -0.12852 -0.061708  0.16250 1.057 8.84e-03 0.0419
## 78  -0.050285  0.03714  0.011214 -0.05324 1.050 9.53e-04 0.0212
## 79   0.006172  0.06698  0.219812  0.26240 1.025 2.28e-02 0.0419
## 80  -0.001316  0.00246 -0.002103  0.00368 1.082 4.55e-06 0.0466
## 81   0.000258  0.00504 -0.007406  0.01311 1.051 5.79e-05 0.0190
## 82   0.013159 -0.04565 -0.014716 -0.07196 1.041 1.74e-03 0.0179
## 83   0.030622 -0.10653  0.015475 -0.16616 1.000 9.15e-03 0.0173
## 84  -0.150054  0.27134 -0.286544  0.43565 1.000 6.20e-02 0.0595
## 85   0.095723 -0.08252  0.045931  0.10642 1.078 3.81e-03 0.0492
## 86  -0.116245  0.21914 -0.281975  0.39148 1.028 5.04e-02 0.0639
## 87   0.191027 -0.11658  0.323655  0.38864 0.987 4.93e-02 0.0477
## 88   0.113826 -0.04463  0.287030  0.32691 1.016 3.52e-02 0.0483
## 89   0.256563 -0.20394  0.238339  0.35297 1.010 4.10e-02 0.0502
## 90  -0.060129  0.05199  0.012108 -0.06127 1.072 1.26e-03 0.0407
## 91   0.002433  0.00933 -0.019692  0.03083 1.051 3.20e-04 0.0202
## 92   0.078239 -0.15785  0.037867 -0.20269 1.016 1.36e-02 0.0282
## 93   0.001323 -0.01436  0.032751 -0.04274 1.066 6.15e-04 0.0342
## 94  -0.054590  0.04199 -0.007972 -0.05632 1.052 1.07e-03 0.0237
## 95  -0.044059  0.03163 -0.022693 -0.05142 1.053 8.89e-04 0.0239
## 96   0.051418 -0.12821  0.032664 -0.18075 1.007 1.08e-02 0.0217
## 97  -0.115697  0.05956 -0.062399 -0.15546 0.995 8.00e-03 0.0145
## 98  -0.037313  0.02735  0.013240 -0.04093 1.052 5.64e-04 0.0220
## 99  -0.008661  0.00728  0.001194 -0.00875 1.068 2.58e-05 0.0342
## 100 -0.109558  0.09259  0.027498 -0.11297 1.059 4.28e-03 0.0365
```

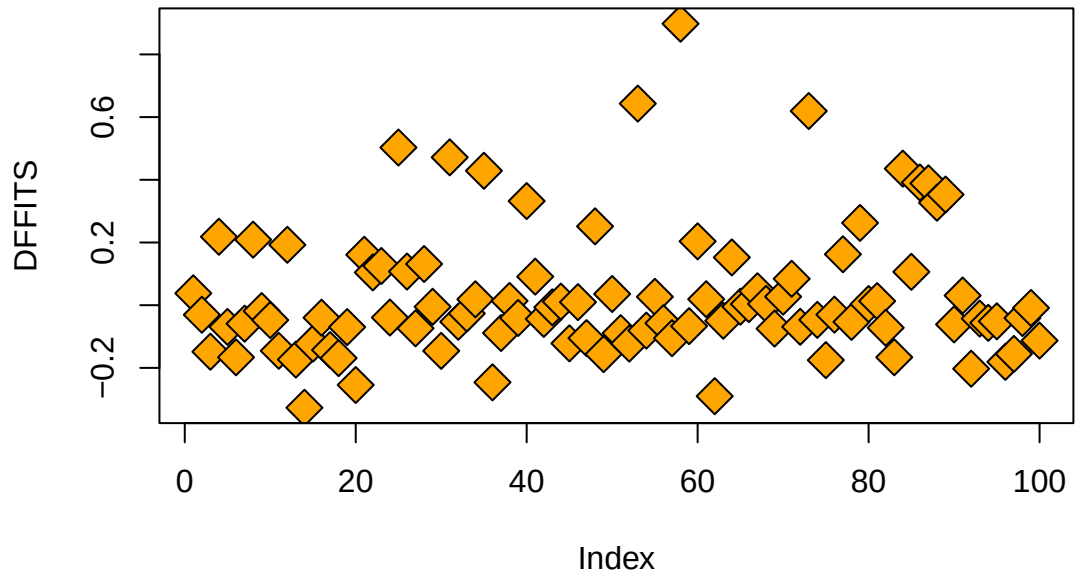
```
with(sim.data, plot(x1,y, type='n'))
with(sim.data, text(x1, y, 1:100, cex=0.7))
```




```
with(sim.data, plot(x2,y, type='n'))
with(sim.data, text(x2, y, 1:100, cex=0.7))
```



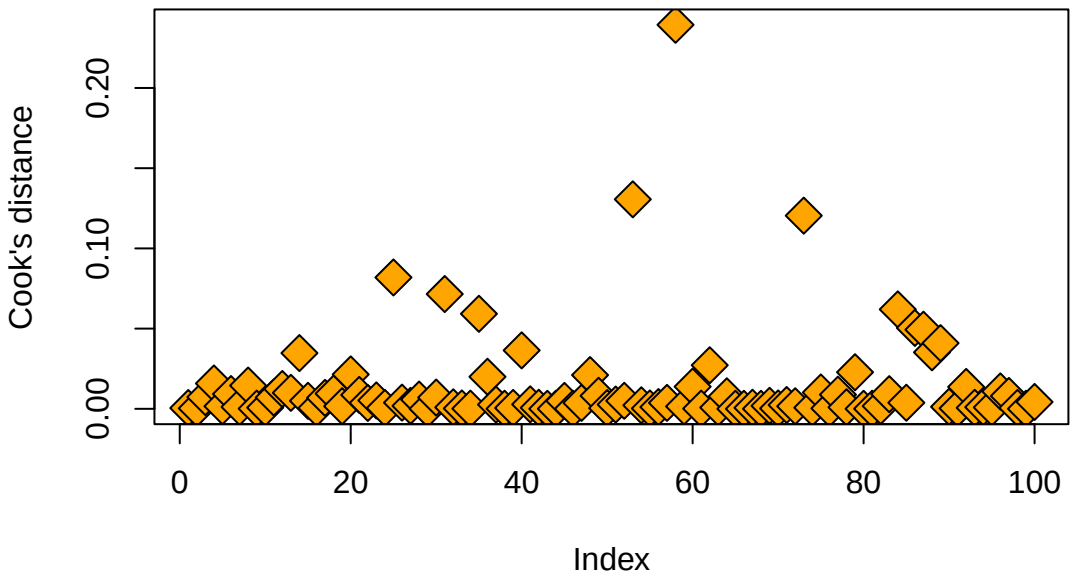
```
plot(dffits(lm.sim1), pch=23, bg='orange', cex=2, ylab="DFFITS")
```




```
sim.data[which(dffits(lm.sim1) > 0.4),]
```

```
##            y        x1        x2    x2.sq
## 25  3.129769 0.9233189 -1.818935 3.308523
## 31 10.688743 0.4233471  1.852148 3.430451
## 35  2.580983 0.5233111 -1.728927 2.989190
## 53 12.697904 0.6797742  2.133486 4.551764
## 58  4.101425 0.5187037 -2.183740 4.768719
## 73  3.020564 0.5667020 -2.018473 4.074235
## 84  3.324928 0.9827948 -1.608631 2.587695
```

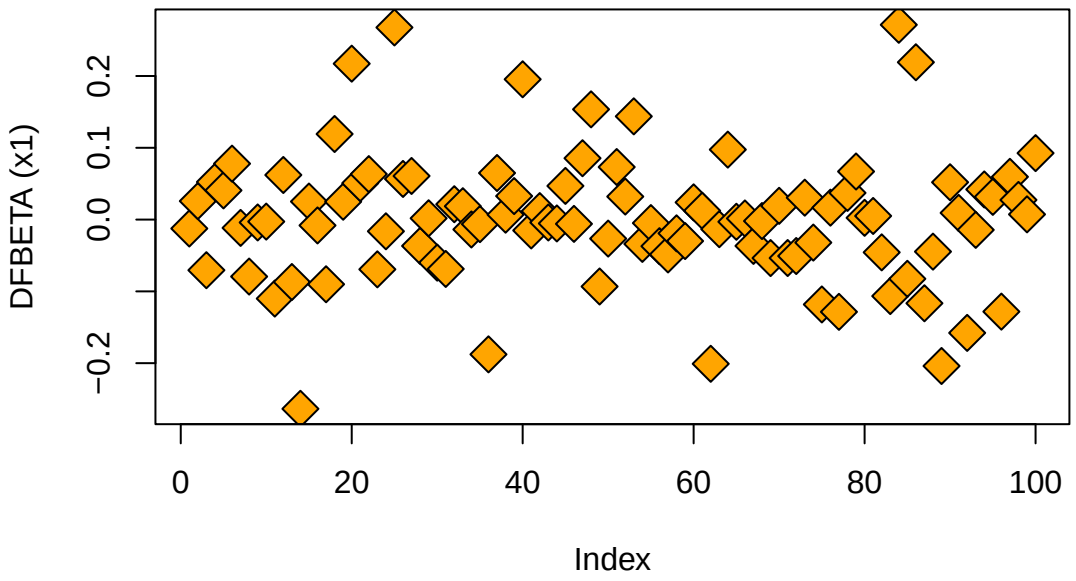
```
plot(cooks.distance(lm.sim1), pch=23, bg='orange', cex=2, ylab="Cook's distance")
```




```
sim.data[which(cooks.distance(lm.sim1) > 0.05),]
```

```
##            y        x1        x2    x2.sq
## 25  3.129769 0.9233189 -1.818935 3.308523
## 31 10.688743 0.4233471  1.852148 3.430451
## 35  2.580983 0.5233111 -1.728927 2.989190
## 53 12.697904 0.6797742  2.133486 4.551764
## 58  4.101425 0.5187037 -2.183740 4.768719
## 73  3.020564 0.5667020 -2.018473 4.074235
## 84  3.324928 0.9827948 -1.608631 2.587695
## 86  2.311289 0.9525859 -1.819132 3.309240
```

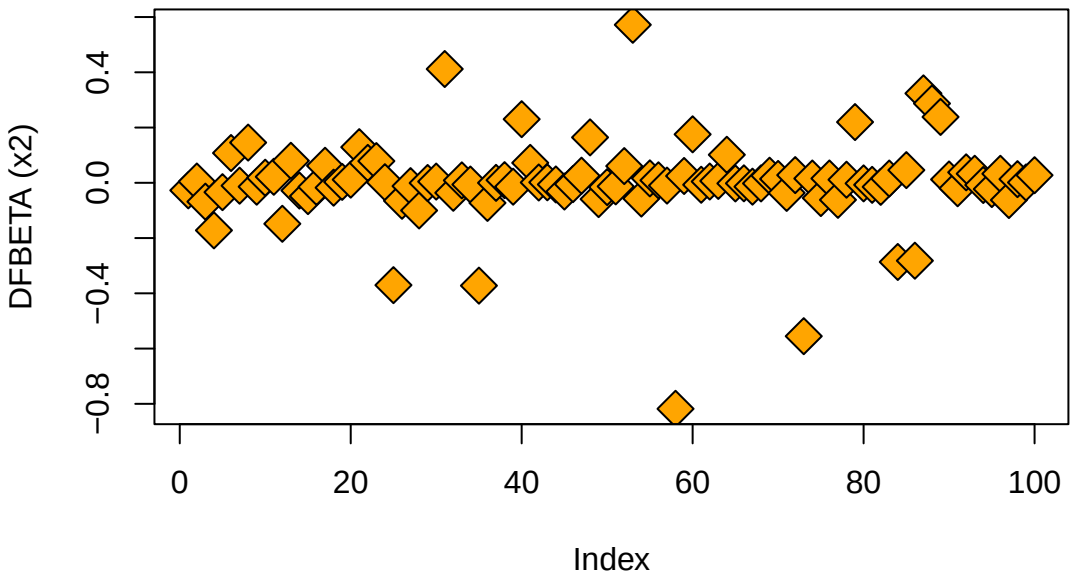
```
plot(dfbetas(lm.sim1)[,'x1'], pch=23, bg='orange', cex=2, ylab="DFBETA (x1)")
```




```
sim.data[which(abs(dfbetas(lm.sim1)[,'x1']) > 0.2),]
```

```
##              y         x1          x2       x2.sq
## 14  0.08862869 0.93163443  0.11245751 0.012646691
## 20 -1.45361649 0.03861056 -0.07482336 0.005598536
## 25  3.12976917 0.92331888 -1.81893450 3.308522718
## 62 -0.94783340 0.81004495 -0.04674302 0.002184910
## 84  3.32492770 0.98279480 -1.60863142 2.587695036
## 86  2.31128875 0.95258585 -1.81913169 3.309240110
## 89  8.00596636 0.13449669  1.49071878 2.222242474
```

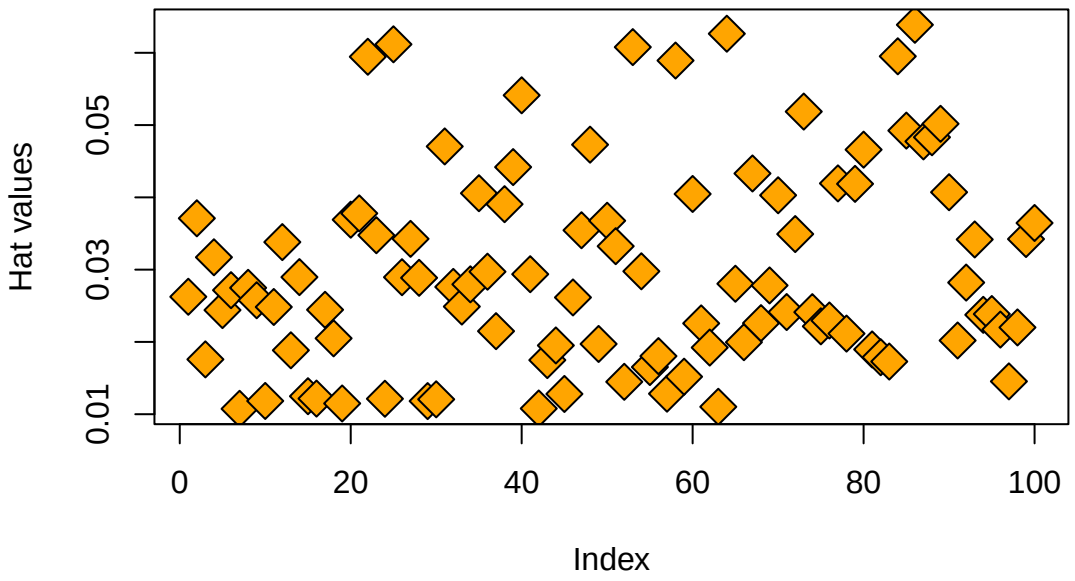
```
plot(dfbetas(lm.sim1)[,'x2'], pch=23, bg='orange', cex=2, ylab="DFBETA (x2)")
```




```
sim.data[which(abs(dfbetas(lm.sim1)[,'x2']) > 0.4),]
```

```
##            y        x1        x2    x2.sq
## 31 10.688743 0.4233471  1.852148 3.430451
## 53 12.697904 0.6797742  2.133486 4.551764
## 58  4.101425 0.5187037 -2.183740 4.768719
## 73  3.020564 0.5667020 -2.018473 4.074235
```

```
plot(hatvalues(lm.sim1), pch=23, bg='orange', cex=2, ylab='Hat values')
```



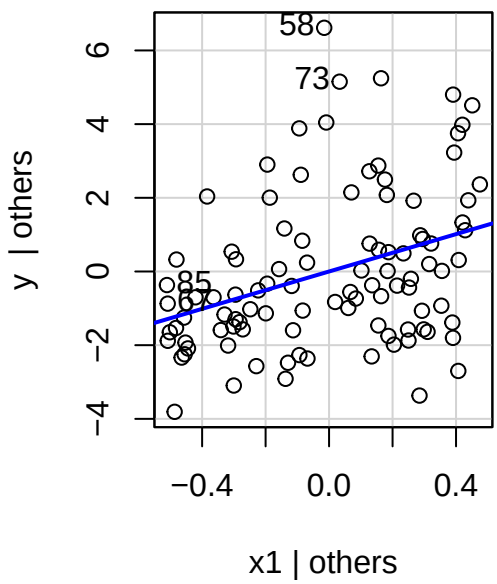

```
sim.data[which(hatvalues(lm.sim1) > 0.05),]
```

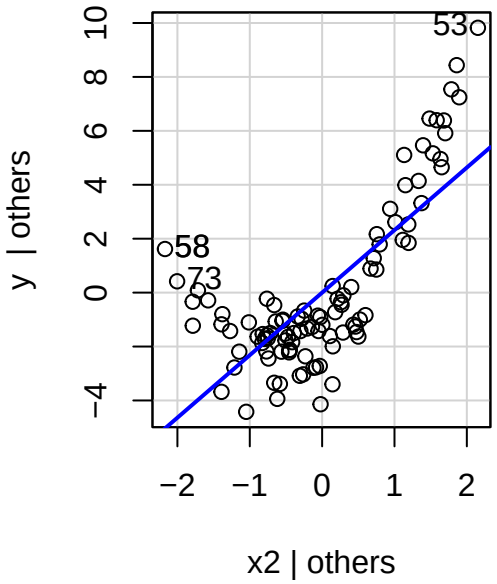
```
##            y        x1        x2    x2.sq
## 22  8.200058 0.9568375  1.619937 2.624196
## 25  3.129769 0.9233189 -1.818935 3.308523
## 40  9.865225 0.9248044  1.552549 2.410409
## 53 12.697904 0.6797742  2.133486 4.551764
## 58  4.101425 0.5187037 -2.183740 4.768719
## 64  8.591778 0.9934096  1.602242 2.567181
## 73  3.020564 0.5667020 -2.018473 4.074235
## 84  3.324928 0.9827948 -1.608631 2.587695
## 86  2.311289 0.9525859 -1.819132 3.309240
## 89  8.005966 0.1344967  1.490719 2.222242
```

```
summary(sim.data)
```

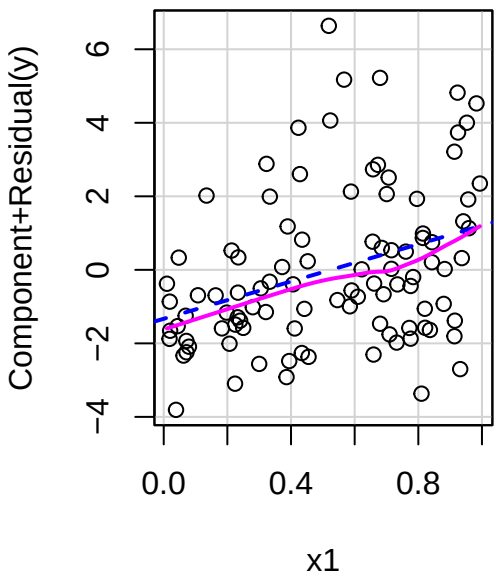
```
##        y                 x1                x2                x2.sq        
##  Min.   :-2.4678   Min.   :0.01019   Min.   :-2.18374   Min.   :0.000029  
##  1st Qu.: 0.3617   1st Qu.:0.24741   1st Qu.:-0.66815   1st Qu.:0.108606  
##  Median : 1.4097   Median :0.57578   Median :-0.11075   Median :0.461068  
##  Mean   : 2.4999   Mean   :0.52502   Mean   :-0.01223   Mean   :0.964403  
##  3rd Qu.: 3.1786   3rd Qu.:0.77783   3rd Qu.: 0.68362   3rd Qu.:1.420308  
##  Max.   :12.6979   Max.   :0.99341   Max.   : 2.13349   Max.   :4.768719
```

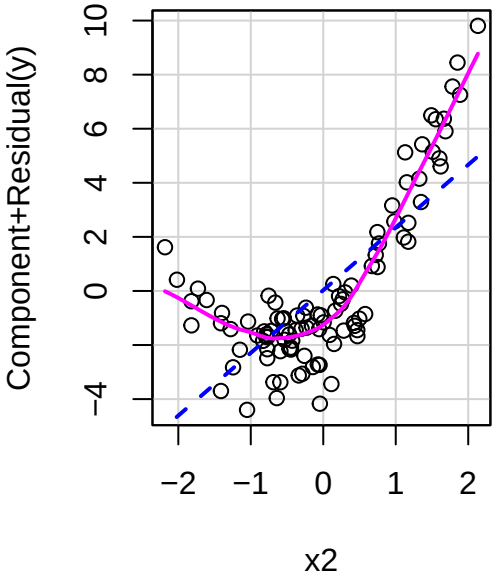
```
par(mfrow=c(1,2))
avPlots(lm.sim1, 'x1')
avPlots(lm.sim1, 'x2')
```






```
crPlots(lm.sim1, 'x1')
crPlots(lm.sim1, 'x2')
```





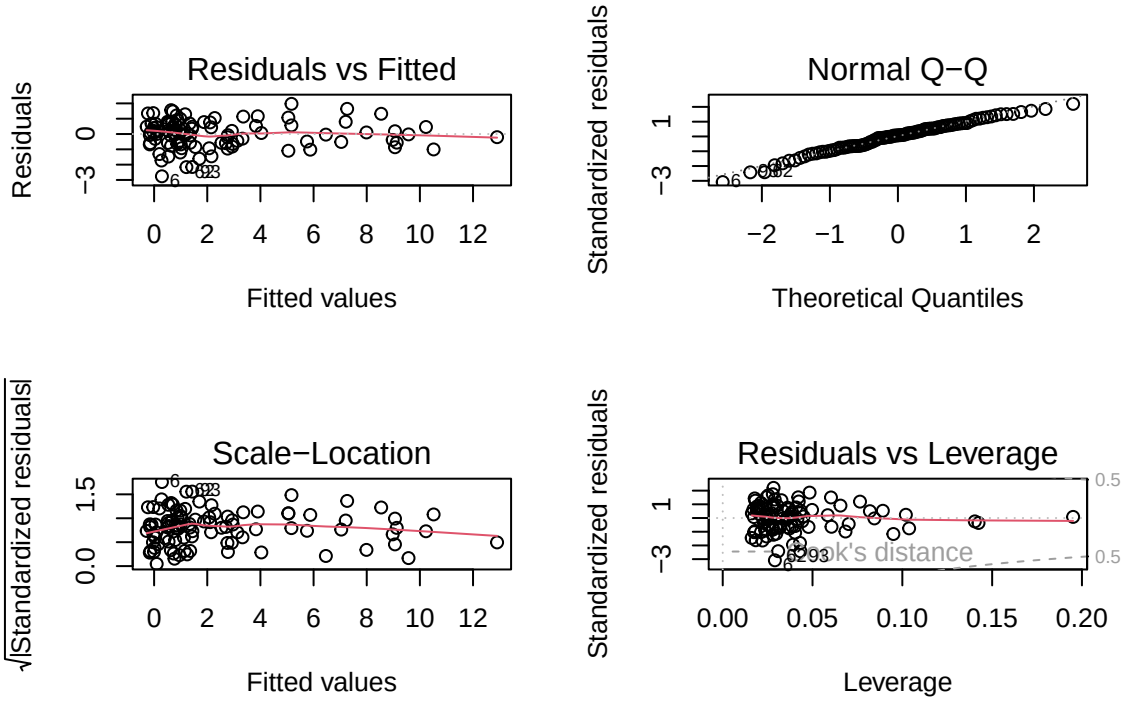
```r
## with the true model
lm.sim2<-lm(y~x1+x2+x2.sq, data=sim.data)
summary(lm.sim2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x2.sq, data = sim.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.76152 -0.58416  0.06429  0.63388  1.97134
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.38208    0.18910   2.020 0.046119 *
## x1           1.14424    0.31398   3.644 0.000435 ***
## x2           2.09221    0.09322  22.445  < 2e-16 ***
## x2.sq        1.59963    0.08138  19.657  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9082 on 96 degrees of freedom
## Multiple R-squared:  0.9182, Adjusted R-squared:  0.9157
## F-statistic: 359.4 on 3 and 96 DF,  p-value: < 2.2e-16
```

```r
outlierTest(lm.sim2)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 6 -3.234252          0.0016782      0.16782
```

```r
par(mfrow=c(2,2))
plot(lm.sim2)
```



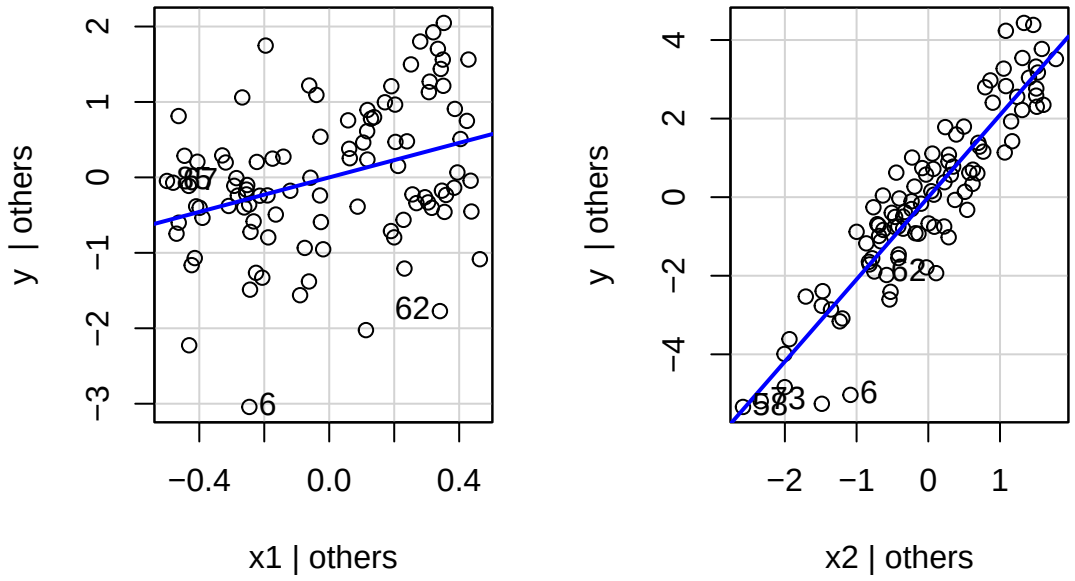

```
influence.measures(lm.sim2)
```

```
## Influence measures of
##   lm(formula = y ~ x1 + x2 + x2.sq, data = sim.data) :
##
##        dfb.1_    dfb.x1    dfb.x2  dfb.x2.s    dffit cov.r   cook.d    hat inf
## 1   -0.052242  0.042605  0.080733 -3.84e-02 -0.11309 1.056 3.22e-03 0.0297
## 2    0.084745 -0.065198 -0.010425 -1.69e-02  0.08568 1.076 1.85e-03 0.0386
## 3   -0.011197 -0.073452 -0.065216  7.28e-02 -0.14310 1.031 5.13e-03 0.0237
## 4   -0.007984  0.015864 -0.119458  7.56e-02  0.15974 1.060 6.41e-03 0.0409
## 5   -0.065994  0.041992 -0.039449  1.22e-02 -0.07926 1.059 1.58e-03 0.0250
## 6   -0.339644  0.279029  0.365658 -1.42e-01 -0.55931 0.707 7.12e-02 0.0290   *
## 7    0.027152  0.021597  0.015778 -4.86e-02  0.07377 1.050 1.37e-03 0.0190
## 8    0.218774 -0.154363  0.253177  6.47e-02  0.38356 0.872 3.53e-02 0.0283   *
## 9   -0.009030 -0.016595 -0.065108 -7.90e-03 -0.08853 1.057 1.97e-03 0.0260
## 10   0.019573  0.007299 -0.011595 -2.28e-02  0.04385 1.055 4.85e-04 0.0162
## 11  -0.001597  0.007360 -0.000548 -5.52e-03  0.00997 1.081 2.51e-05 0.0358
## 12  -0.009454  0.014175 -0.059651  3.74e-02  0.08113 1.083 1.66e-03 0.0430
## 13  -0.003094 -0.134024  0.092892  9.66e-02 -0.24928 0.952 1.53e-02 0.0222
## 14   0.079671 -0.301566 -0.054855  2.23e-01 -0.39015 0.947 3.71e-02 0.0430
## 15  -0.086924  0.006747 -0.054681  7.36e-02 -0.13799 1.015 4.76e-03 0.0175
## 16   0.040254  0.034606  0.043295 -7.07e-02  0.11740 1.032 3.46e-03 0.0191
## 17   0.020120 -0.094450  0.051617  5.06e-02 -0.14439 1.040 5.23e-03 0.0279
## 18  -0.202585  0.115554 -0.031119  8.64e-02 -0.21436 0.990 1.14e-02 0.0245
## 19   0.006736 -0.001383  0.000304 -5.31e-03  0.00912 1.061 2.10e-05 0.0174
## 20  -0.398043  0.300347  0.005472  9.51e-02 -0.39862 0.924 3.86e-02 0.0391
## 21  -0.002541  0.004753  0.016862  9.92e-03  0.02481 1.091 1.56e-04 0.0450
## 22   0.099150 -0.125236 -0.158481 -9.63e-02 -0.26688 1.075 1.78e-02 0.0684
## 23   0.081459 -0.070543  0.071131  2.70e-02  0.12238 1.065 3.77e-03 0.0365
```

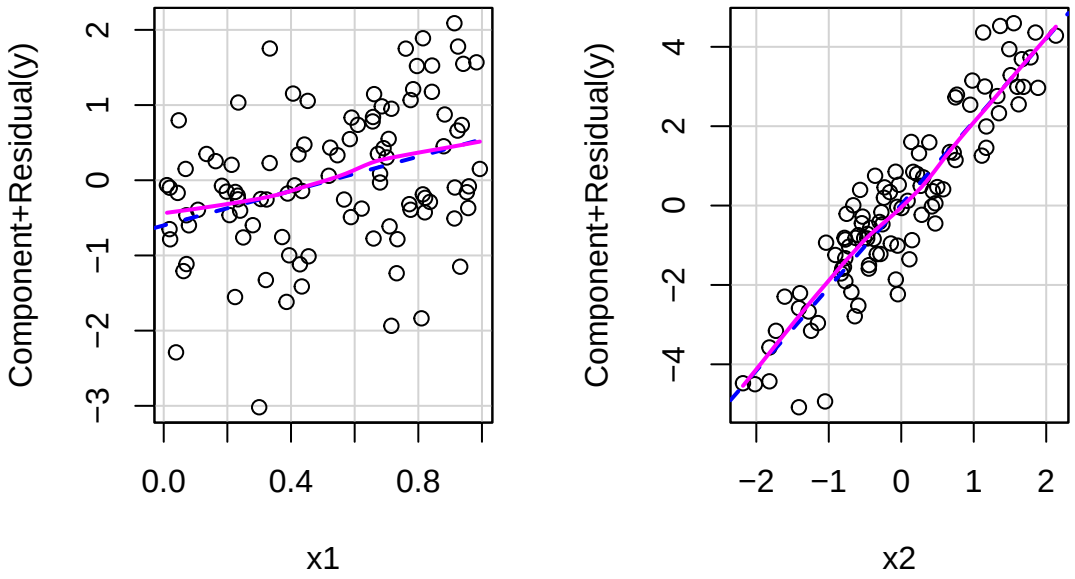
```
## 24   0.033589  0.073693  0.006543 -1.07e-01  0.16263 1.013 6.60e-03 0.0213
## 25  -0.027216  0.020626 -0.051310  5.04e-02  0.07962 1.158 1.60e-03 0.1020   *
## 26  -0.040281  0.119834 -0.141357  2.72e-03  0.23443 0.994 1.36e-02 0.0290
## 27   0.011431 -0.008390  0.002161 -2.85e-03  0.01161 1.082 3.41e-05 0.0365
## 28  -0.001662  0.001594  0.003376 -2.10e-03 -0.00462 1.082 5.40e-06 0.0364
## 29   0.143531 -0.030892  0.035230 -1.12e-01  0.19377 0.974 9.28e-03 0.0178
## 30  -0.032099 -0.068656 -0.009004  1.00e-01 -0.15253 1.019 5.81e-03 0.0213
## 31   0.017176 -0.042656  0.089982  1.14e-01  0.16530 1.131 6.88e-03 0.0891   *
## 32  -0.124273  0.088514 -0.135926 -3.22e-02 -0.20888 1.008 1.09e-02 0.0283
## 33   0.037252 -0.025309 -0.012043 -7.80e-03  0.04027 1.068 4.09e-04 0.0259
## 34   0.118853 -0.086014 -0.039516 -1.89e-02  0.12775 1.049 4.10e-03 0.0286
## 35   0.002556 -0.025374 -0.103447  1.06e-01  0.14903 1.124 5.60e-03 0.0819
## 36   0.044635 -0.165615 -0.069335  1.13e-01 -0.21974 1.036 1.21e-02 0.0403
## 37  -0.017023  0.010092 -0.000890  6.90e-03 -0.01775 1.069 7.96e-05 0.0253
## 38   0.025813 -0.052504 -0.079622 -2.17e-02 -0.11918 1.071 3.58e-03 0.0404
## 39  -0.046280  0.038757 -0.017469  2.26e-03 -0.05018 1.089 6.36e-04 0.0443
## 40  -0.131595  0.180174  0.234101  1.22e-01  0.38256 1.009 3.61e-02 0.0603
## 41  -0.002704  0.001881 -0.005770 -2.77e-03 -0.00825 1.078 1.72e-05 0.0331
## 42   0.056780 -0.005937  0.004613 -5.06e-02  0.08326 1.043 1.74e-03 0.0171
## 43   0.000448  0.163500 -0.039135 -1.50e-01  0.26880 0.953 1.77e-02 0.0253
## 44   0.053634 -0.030307 -0.035353 -8.02e-03  0.07055 1.053 1.25e-03 0.0198
## 45  -0.096271  0.024791 -0.036299  6.93e-02 -0.12799 1.023 4.10e-03 0.0181
## 46   0.011879 -0.008793 -0.007254  1.99e-04  0.01462 1.070 5.40e-05 0.0262
## 47  -0.124896  0.097096  0.028743  1.89e-02 -0.12914 1.062 4.19e-03 0.0362
## 48  -0.136925  0.235010  0.263857  6.31e-02  0.42382 0.947 4.38e-02 0.0484
## 49   0.002103 -0.069088 -0.043245  6.10e-02 -0.11304 1.053 3.21e-03 0.0278
## 50   0.015938 -0.013727 -0.008232  1.53e-03  0.01864 1.082 8.78e-05 0.0370
## 51  -0.018899  0.013702 -0.003282  4.96e-03 -0.01917 1.081 9.28e-05 0.0356
## 52  -0.122707  0.031769  0.079548  5.97e-02 -0.18700 0.970 8.64e-03 0.0161
## 53   0.015649  0.002595 -0.048141 -7.44e-02 -0.09877 1.210 2.46e-03 0.1404   *
## 54   0.015983 -0.080395 -0.139689 -6.34e-03 -0.20088 1.017 1.01e-02 0.0298
## 55   0.098846 -0.017470  0.104204 -5.18e-02  0.17253 0.993 7.39e-03 0.0182
## 56  -0.004294  0.110886 -0.008915 -1.03e-01  0.17304 1.027 7.48e-03 0.0278
## 57  -0.002569 -0.007405 -0.002334  1.00e-02 -0.01521 1.066 5.85e-05 0.0226
## 58  -0.002698 -0.007977 -0.023991  3.34e-02  0.03996 1.295 4.03e-04 0.1951   *
## 59   0.004535  0.018678 -0.009539 -1.96e-02  0.03811 1.062 3.67e-04 0.0207
## 60   0.001411  0.005824 -0.101457 -9.32e-02 -0.16176 1.091 6.58e-03 0.0606
## 61  -0.032922  0.197395 -0.081623 -1.30e-01  0.29777 0.944 2.17e-02 0.0279
## 62   0.024710 -0.297152 -0.028258  2.72e-01 -0.44321 0.837 4.66e-02 0.0309   *
## 63   0.031070  0.034208 -0.002304 -6.38e-02  0.09866 1.041 2.45e-03 0.0189
## 64   0.048163 -0.062045 -0.070318 -3.94e-02 -0.12047 1.112 3.66e-03 0.0701
## 65  -0.037814  0.160389  0.117886 -9.05e-02  0.23997 1.004 1.43e-02 0.0327
## 66   0.001953  0.057691 -0.058551 -3.15e-02  0.12048 1.036 3.64e-03 0.0214
## 67   0.106540 -0.091271 -0.035390 -1.92e-03  0.11437 1.077 3.29e-03 0.0433
## 68   0.000236 -0.000157 -0.000179  6.16e-06  0.00032 1.067 2.59e-08 0.0226
## 69  -0.019756  0.070573 -0.025063 -3.99e-02  0.09666 1.067 2.35e-03 0.0335
## 70  -0.076254  0.183072  0.133325 -5.59e-02  0.25422 1.024 1.61e-02 0.0423
## 71   0.209805 -0.143771 -0.100856 -2.53e-02  0.24257 0.969 1.45e-02 0.0244
## 72  -0.017173  0.044037 -0.020459 -1.66e-02  0.05858 1.080 8.66e-04 0.0380
## 73   0.011166  0.022108  0.093238 -1.17e-01 -0.14670 1.209 5.43e-03 0.1424   *
## 74  -0.023256  0.111875 -0.038192 -7.28e-02  0.16062 1.039 6.46e-03 0.0304
## 75   0.010035 -0.078005 -0.038171  6.39e-02 -0.11596 1.059 3.38e-03 0.0318
## 76  -0.042532  0.028633  0.027501  1.56e-03 -0.05394 1.062 7.34e-04 0.0230
## 77   0.284041 -0.249273 -0.122137  1.74e-02  0.31891 0.989 2.51e-02 0.0421
```

```
## 78   0.030633 -0.018405 -0.005068 -1.08e-02  0.03252 1.066 2.67e-04 0.0238
## 79  -0.004974  0.004615  0.032541  2.72e-02  0.05105 1.106 6.58e-04 0.0585
## 80   0.055155 -0.088522  0.084364 -2.28e-02 -0.14547 1.076 5.32e-03 0.0478
## 81   0.009093  0.053156 -0.065470 -3.06e-02  0.12734 1.030 4.06e-03 0.0201
## 82  -0.000953  0.093579  0.035298 -9.16e-02  0.14940 1.040 5.59e-03 0.0287
## 83   0.000903 -0.080798 -0.000415  7.87e-02 -0.12799 1.047 4.11e-03 0.0278
## 84  -0.137941  0.148189 -0.218856  1.62e-01  0.34509 1.063 2.96e-02 0.0764
## 85   0.114911 -0.104579  0.055849  1.37e-02  0.13505 1.081 4.59e-03 0.0497
## 86   0.093012 -0.073771  0.163230 -1.59e-01 -0.25603 1.136 1.65e-02 0.1038   *
## 87  -0.001836  0.003130 -0.004587 -5.60e-03 -0.00850 1.139 1.82e-05 0.0844   *
## 88  -0.033489  0.095682 -0.202260 -2.65e-01 -0.37845 1.088 3.57e-02 0.0950
## 89   0.123424 -0.144410  0.126301  1.16e-01  0.24003 1.078 1.44e-02 0.0656
## 90  -0.016002  0.012842  0.002881  2.35e-03 -0.01635 1.088 6.75e-05 0.0416
## 91   0.009480  0.031461 -0.062795 -7.89e-03  0.10094 1.043 2.56e-03 0.0203
## 92   0.025259 -0.093181  0.011423  6.34e-02 -0.12178 1.068 3.73e-03 0.0387
## 93   0.067329 -0.099449  0.387725 -2.42e-01 -0.52911 0.847 6.64e-02 0.0432   *
## 94   0.056732 -0.035340  0.010336 -2.09e-02  0.05897 1.066 8.77e-04 0.0272
## 95  -0.011971  0.007484 -0.006296  2.87e-03 -0.01387 1.069 4.86e-05 0.0249
## 96   0.013894 -0.090812  0.010275  7.29e-02 -0.13000 1.054 4.25e-03 0.0317
## 97  -0.166934  0.051239 -0.091936  1.02e-01 -0.22375 0.951 1.23e-02 0.0184
## 98   0.020369 -0.012782 -0.006230 -5.52e-03  0.02223 1.067 1.25e-04 0.0234
## 99   0.144058 -0.107513 -0.015120 -3.28e-02  0.14551 1.056 5.32e-03 0.0361
## 100 -0.144466  0.113377  0.033133  2.08e-02 -0.14931 1.057 5.60e-03 0.0372
```

```
par(mfrow=c(1,2))
avPlots(lm.sim2, 'x1')
avPlots(lm.sim2, 'x2')
```



```r
crPlots(lm.sim2, 'x1')
crPlots(lm.sim2, 'x2')
```

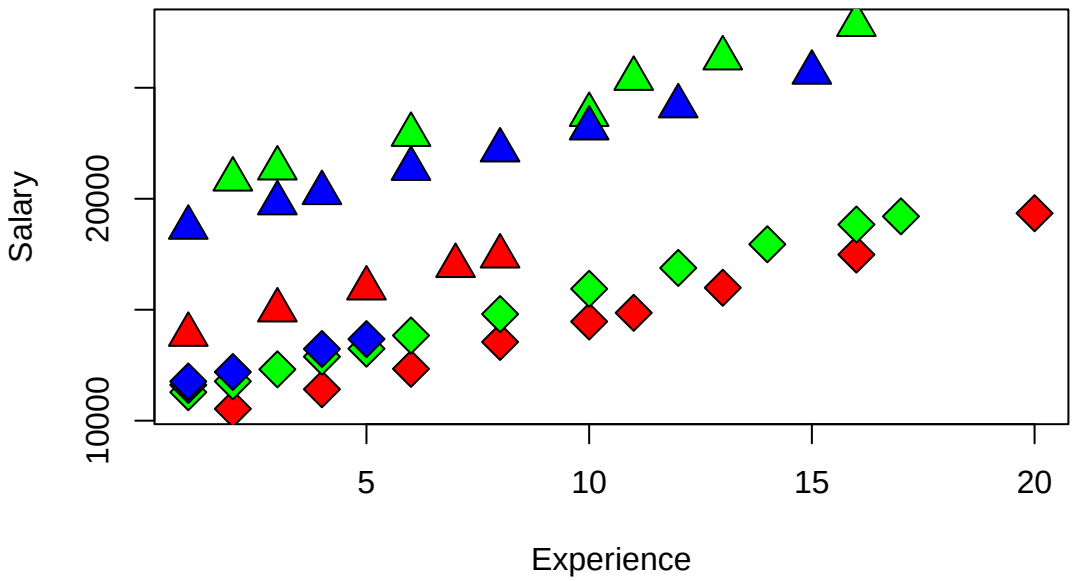


0.2 6. Interactions and qualitative variables

```r
# page 4
load("/Users/ohsukju/Downloads/data/salary.Rdata")
salary.table$E <- factor(salary.table$E)
salary.table$M <- factor(salary.table$M)
head(salary.table$E)
```

```
## [1] 1 3 3 2 3 2
## Levels: 1 2 3
```

```r
# page 5
plot(salary.table$X, salary.table$S, type='n', xlab='Experience', ylab='Salary')
colors <- c('red', 'green', 'blue')
symbols <- c(23,24)
for (i in 1:3) {
 for (j in 0:1) {
  subset <- as.logical((salary.table$E == i) * (salary.table$M == j))
  points(salary.table$X[subset], salary.table$S[subset], pch=symbols[j+1], bg=colors[i], cex=2)
 }
}
```




```
# page 9
salary.lm <- lm(S ~ E + M + X, salary.table)
summary(salary.lm)
```

```
##
## Call:
## lm(formula = S ~ E + M + X, data = salary.table)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1884.60  -653.60    22.23   844.85  1716.47
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8035.60     386.69  20.781  < 2e-16 ***
## E2           3144.04     361.97   8.686 7.73e-11 ***
## E3           2996.21     411.75   7.277 6.72e-09 ***
## M1           6883.53     313.92  21.928  < 2e-16 ***
## X             546.18      30.52  17.896  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1027 on 41 degrees of freedom
## Multiple R-squared:  0.9568, Adjusted R-squared:  0.9525
## F-statistic: 226.8 on 4 and 41 DF,  p-value: < 2.2e-16
```

```
salary.lm2<- lm(S ~ (E + M)* X, salary.table)
summary(salary.lm2)
```

```
##
## Call:
## lm(formula = S ~ (E + M) * X, data = salary.table)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2278.03  -545.02   -77.97   517.26  1839.76
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7466.44     550.48  13.563 3.84e-16 ***
## E2           4190.38     659.11   6.358 1.84e-07 ***
## E3           4132.04     679.25   6.083 4.38e-07 ***
## M1           6354.01     546.37  11.629 4.38e-14 ***
## X             614.40      52.98  11.597 4.75e-14 ***
## E2:X         -147.31      69.40  -2.123   0.0404 *
## E3:X         -208.63      95.70  -2.180   0.0355 *
## M1:X          118.30      69.35   1.706   0.0962 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 981.5 on 38 degrees of freedom
## Multiple R-squared:  0.9634, Adjusted R-squared:  0.9567
## F-statistic:   143 on 7 and 38 DF,  p-value: < 2.2e-16
```

```
anova(salary.lm2, salary.lm)
```

```
## Analysis of Variance Table
##
## Model 1: S ~ (E + M) * X
## Model 2: S ~ E + M + X
##   Res.Df      RSS Df Sum of Sq      F  Pr(>F)
## 1     38 36607751
## 2     41 43280719 -3  -6672969 2.3089 0.09187 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# page 10
head(model.matrix(salary.lm))
```

```
##   (Intercept) E2 E3 M1 X
## 1           1  0  0  1 1
## 2           1  0  1  0 1
## 3           1  0  1  1 1
## 4           1  1  0  0 1
## 5           1  0  1  0 1
## 6           1  1  0  1 2
```

```
head(model.frame(salary.lm))
```

```
##       S E M X
## 1 13876 1 1 1
```

```
## 2 11608 3 0 1
## 3 18701 3 1 1
## 4 11283 2 0 1
## 5 11767 3 0 1
## 6 20872 2 1 2
```

```
# page 13
model_XE = lm(S~ E + M + X + X:E, salary.table)
print(summary(model_XE))
```

```
##
## Call:
## lm(formula = S ~ E + M + X + X:E, data = salary.table)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2013.04  -634.68   -16.71   615.66  2014.14
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7256.28     549.49  13.205 5.65e-16 ***
## E2           4172.50     674.97   6.182 2.90e-07 ***
## E3           3946.36     686.69   5.747 1.16e-06 ***
## M1           7102.45     333.44  21.300  < 2e-16 ***
## X             632.29      53.19  11.888 1.53e-14 ***
## E2:X         -125.51      69.86  -1.797   0.0801 .
## E3:X         -141.27      89.28  -1.582   0.1216
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1005 on 39 degrees of freedom
## Multiple R-squared:  0.9606, Adjusted R-squared:  0.9546
## F-statistic: 158.6 on 6 and 39 DF,  p-value: < 2.2e-16
```

```
# page 14
anova(salary.lm, model_XE)
```

```
## Analysis of Variance Table
##
## Model 1: S ~ E + M + X
## Model 2: S ~ E + M + X + X:E
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
## 1     41 43280719
## 2     39 39410680  2   3870040 1.9149  0.161
```

```
# page 15
model.matrix(model_XE)[10:20,]
```

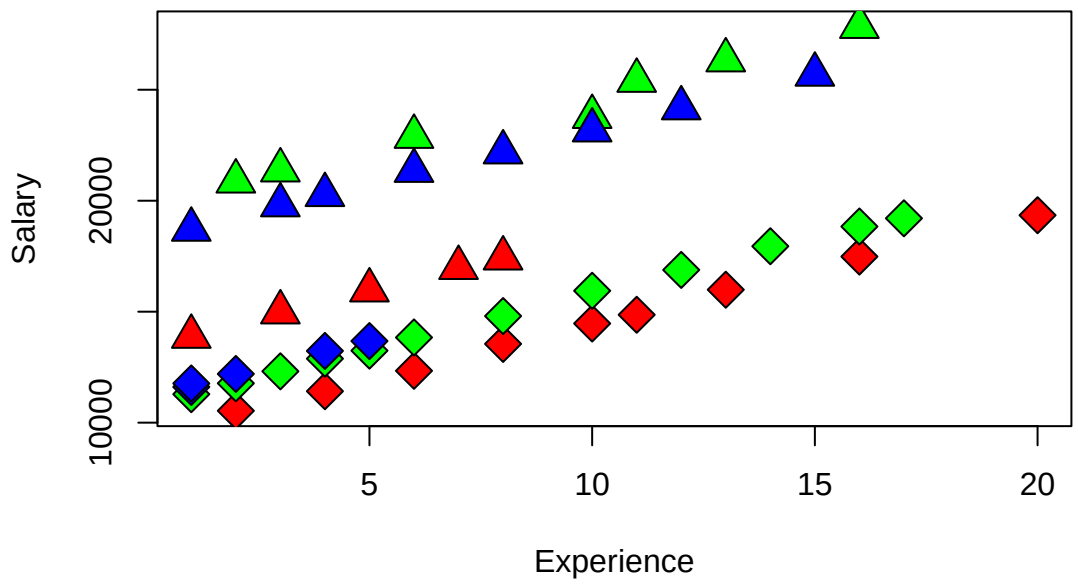
```
##    (Intercept) E2 E3 M1 X E2:X E3:X
## 10           1  1  0  0 3    3    0
## 11           1  0  0  1 3    0    0
## 12           1  1  0  1 3    3    0
```

```
## 13           1  0  1  1 3    0    3
## 14           1  0  0  0 4    0    0
## 15           1  0  1  1 4    0    4
## 16           1  0  1  0 4    0    4
## 17           1  1  0  0 4    4    0
## 18           1  1  0  0 5    5    0
## 19           1  0  1  0 5    0    5
## 20           1  0  0  1 5    0    0
```

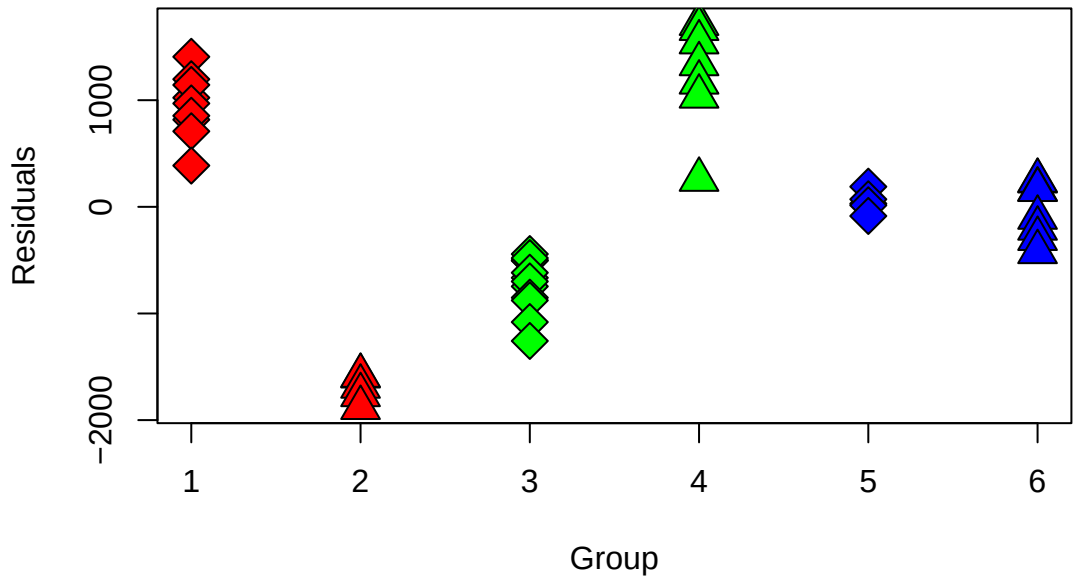
```
head(model.matrix(model_XE))
```

```
##   (Intercept) E2 E3 M1 X E2:X E3:X
## 1           1  0  0  1 1    0    0
## 2           1  0  1  0 1    0    1
## 3           1  0  1  1 1    0    1
## 4           1  1  0  0 1    1    0
## 5           1  0  1  0 1    0    1
## 6           1  1  0  1 2    2    0
```

```
# page 18
plot(salary.table$X, salary.table$S, type='n', xlab='Experience', ylab='Salary')
colors <- c('red', 'green', 'blue')
symbols <- c(23,24)
for (i in 1:3) {
 for (j in 0:1) {
  subset <- as.logical((salary.table$E == i) * (salary.table$M == j))
  points(salary.table$X[subset], salary.table$S[subset], pch=symbols[j+1], bg=colors[i], cex=2)
 }
}
```



```
r = resid(salary.lm)
k = 1
plot(salary.table$X, r, xlim=c(1,6), type='n', xlab='Group', ylab='Residuals')
for (i in 1:3) {
 for (j in 0:1) {
  subset <- as.logical((salary.table$E == i) * (salary.table$M == j))
  points(rep(k, length(r[subset])), r[subset], pch=symbols[j+1], bg=colors[i], cex=2)
  k = k+1
 }
}
```




```
# page 19
interaction.plot(salary.table$E, salary.table$M, r, type='b', col=c('red','blue'), lwd=2, pch=c(23,24))
```


```
# page 21
model_EM = lm(S ~ X + M + E + E:M, salary.table)
summary(model_EM)
```

```
##
## Call:
## lm(formula = S ~ X + M + E + E:M, data = salary.table)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -928.13  -46.21   24.33   65.88  204.89
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9472.685     80.344  117.90   <2e-16 ***
## X            496.987      5.566   89.28   <2e-16 ***
## M1          3981.377    101.175   39.35   <2e-16 ***
## E2          1381.671     77.319   17.87   <2e-16 ***
## E3          1730.748    105.334   16.43   <2e-16 ***
## M1:E2       4902.523    131.359   37.32   <2e-16 ***
## M1:E3       3066.035    149.330   20.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 173.8 on 39 degrees of freedom
## Multiple R-squared:  0.9988, Adjusted R-squared:  0.9986
## F-statistic:  5517 on 6 and 39 DF,  p-value: < 2.2e-16
```

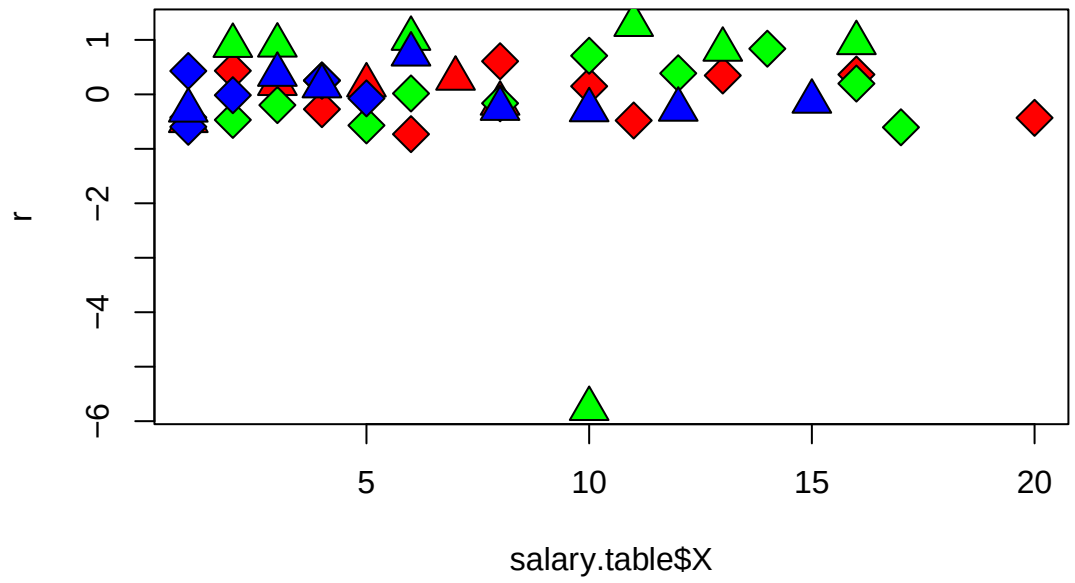
```r
# page 22
anova(salary.lm, model_EM)
```

```
## Analysis of Variance Table
##
## Model 1: S ~ E + M + X
## Model 2: S ~ X + M + E + E:M
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1     41 43280719
## 2     39  1178168  2  42102552 696.84 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# page 23
head(model.matrix(model_EM))
```

```
##   (Intercept) X M1 E2 E3 M1:E2 M1:E3
## 1           1 1  1  0  0     0     0
## 2           1 1  0  0  1     0     0
## 3           1 1  1  0  1     0     1
## 4           1 1  0  1  0     0     0
## 5           1 1  0  0  1     0     0
## 6           1 2  1  1  0     1     0
```

```r
# page 24
r = rstandard(model_EM)
plot(salary.table$X, r, type='n')
for (i in 1:3) {
 for (j in 0:1) {
  subset <- as.logical((salary.table$E == i) * (salary.table$M == j))
  points(salary.table$X[subset], r[subset], pch=symbols[j+1], bg=colors[i], cex=2)
 }
}
```




```
# page 25
library(car)
outlierTest(model_EM)
```

```
##     rstudent unadjusted p-value Bonferroni p
## 33 -14.95083         1.6769e-17    7.714e-16
```

```
subset(salary.table, (E==2)&(M==1))
```

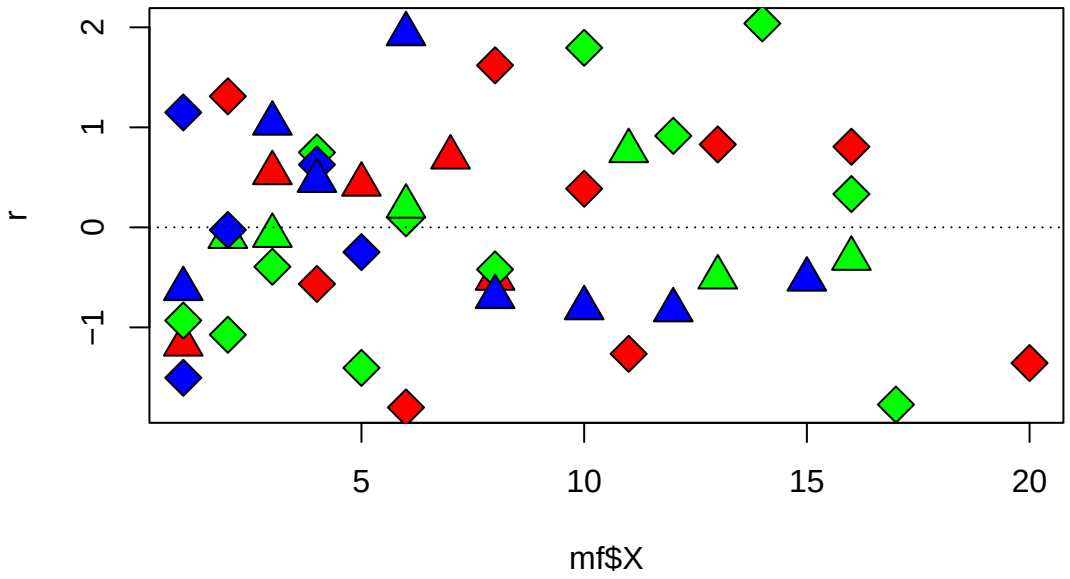
```
##        S  X E M
## 6  20872  2 2 1
## 12 21371  3 2 1
## 24 22884  6 2 1
## 33 23780 10 2 1
## 34 25410 11 2 1
## 39 26330 13 2 1
## 42 27837 16 2 1
```

```
# page 26
subs33 = c(1:length(salary.table$S))[-33]
salary.lm33 = lm(S ~ E + X + M, data=salary.table, subset=subs33)
model_EM33 = lm(S ~ E + X + E:M + M, data=salary.table, subset=subs33)
anova(salary.lm33, model_EM33)
```

```
## Analysis of Variance Table
##
## Model 1: S ~ E + X + M
```

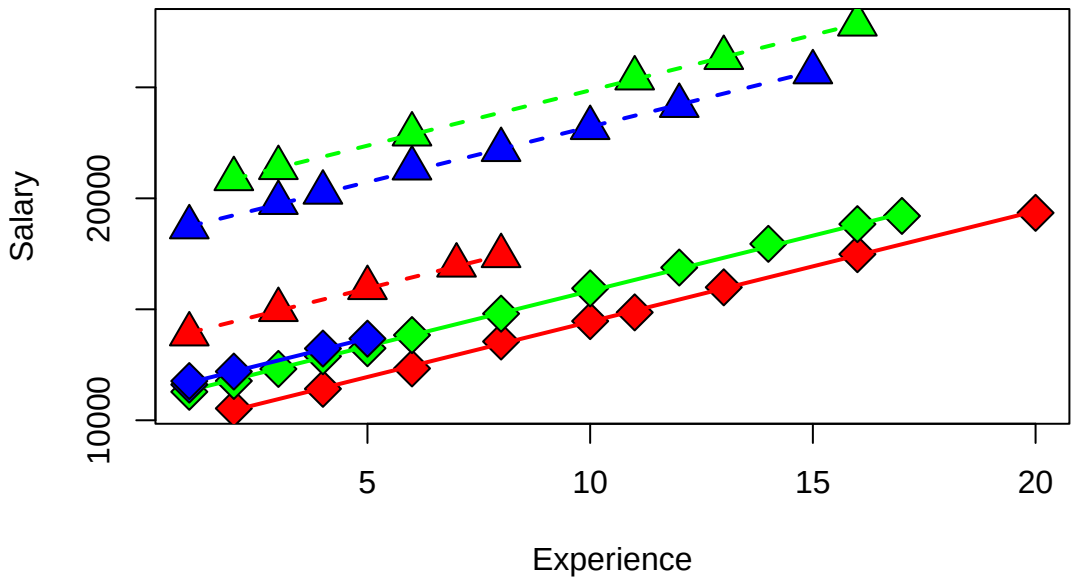
```
## Model 2: S ~ E + X + E:M + M
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1     40 43209096
## 2     38   171188  2  43037908 4776.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# page 27
r = rstandard(model_EM33)
mf = model.frame(model_EM33)
plot(mf$X, r, type='n')
for (i in 1:3) {
 for (j in 0:1) {
  subset <- as.logical((mf$E == i) * (mf$M == j))
  points(mf$X[subset], r[subset], pch=symbols[j+1], bg=colors[i], cex=2)
 }
}
abline(h=0, lty=3)
```



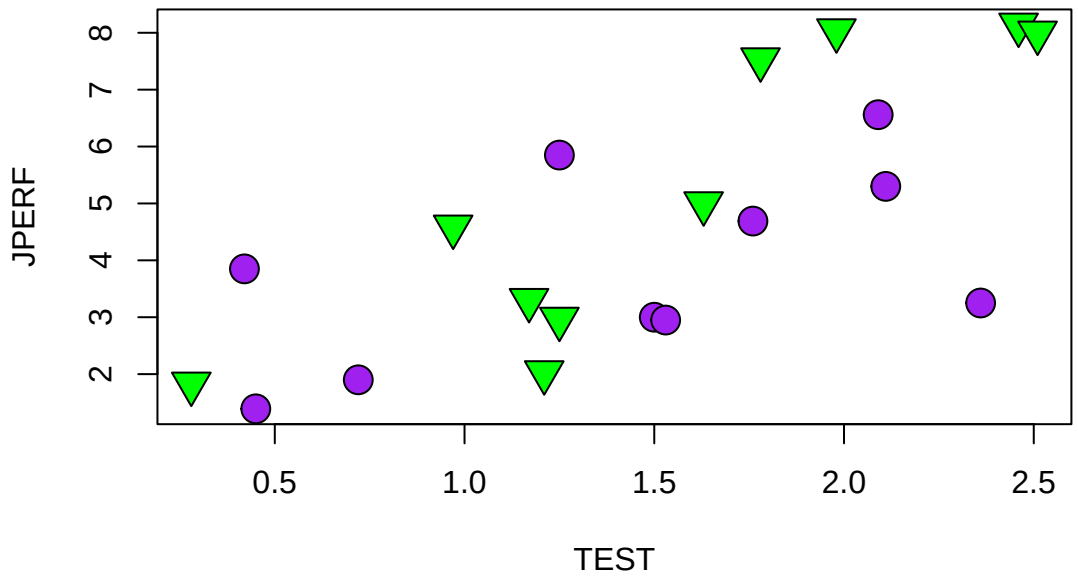

```
# page 28
salaryfinal.lm = lm(S ~ X + E * M, salary.table, subset=subs33)
mf = model.frame(salaryfinal.lm)
plot(mf$X, mf$S, type='n', xlab='Experience', ylab='Salary')
colors <- c('red', 'green', 'blue')
ltys <- c(2,3)
symbols <- c(23,24)
for (i in 1:3) {
 for (j in 0:1) {
```

```
  subset <- as.logical((mf$E == i) * (mf$M == j))
  points(mf$X[subset], mf$S[subset], pch=symbols[j+1], bg=colors[i], cex=2)
  lines(mf$X[subset], fitted(salaryfinal.lm)[subset], lwd=2, lty=ltys[j], col=colors[i])
 }
}
```




```
# page 30
U = salary.table$S - salary.table$X * model_EM$coef['X']
interaction.plot(salary.table$E, salary.table$M, U, type='b', col=c('red','blue'), lwd=2, pch=c(23,24))
```

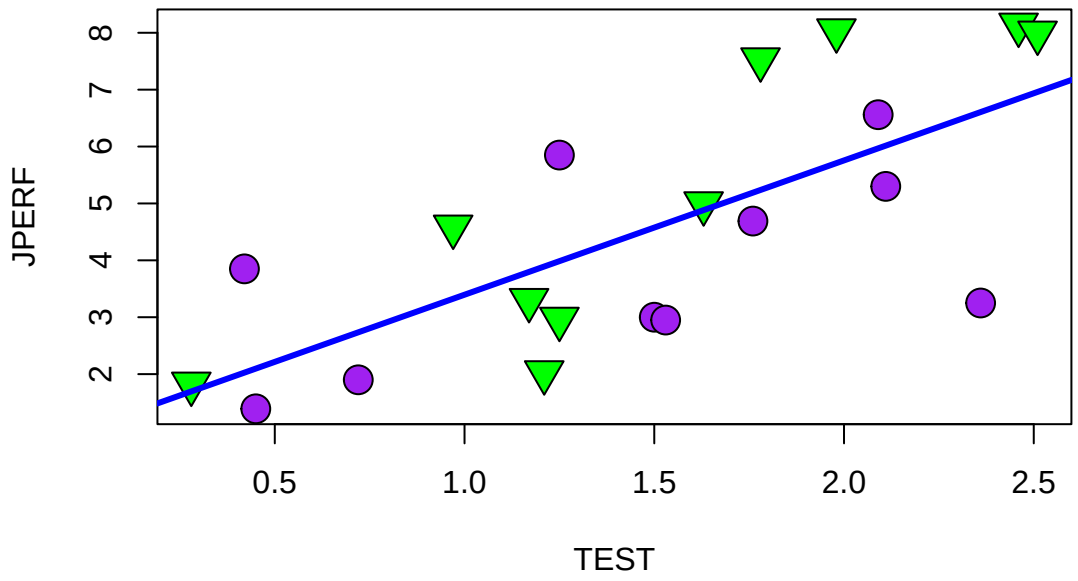

```
# page 31
load("/Users/ohsukju/Downloads/data/jobtest.Rdata")
jobtest.table$ETHN <- factor(jobtest.table$ETHN)
attach(jobtest.table)
plot(TEST, JPERF, type='n')
points(TEST[(ETHN == 0)], JPERF[(ETHN == 0)], pch=21, cex=2, bg='purple')
points(TEST[(ETHN == 1)], JPERF[(ETHN == 1)], pch=25, cex=2, bg='green')
```




```
# page 33
jobtest.lm1 <- lm(JPERF ~ TEST, jobtest.table)
print(summary(jobtest.lm1))
```

```
##
## Call:
## lm(formula = JPERF ~ TEST, data = jobtest.table)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3558 -0.8798 -0.1897  1.2735  2.3312
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.0350     0.8680   1.192 0.248617
## TEST          2.3605     0.5381   4.387 0.000356 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.591 on 18 degrees of freedom
## Multiple R-squared:  0.5167, Adjusted R-squared:  0.4899
## F-statistic: 19.25 on 1 and 18 DF,  p-value: 0.0003555
```

```
plot(TEST, JPERF, type='n')
points(TEST[(ETHN == 0)], JPERF[(ETHN == 0)], pch=21, cex=2, bg='purple')
points(TEST[(ETHN == 1)], JPERF[(ETHN == 1)], pch=25, cex=2, bg='green')
abline(jobtest.lm1$coef, lwd=3, col='blue')
```

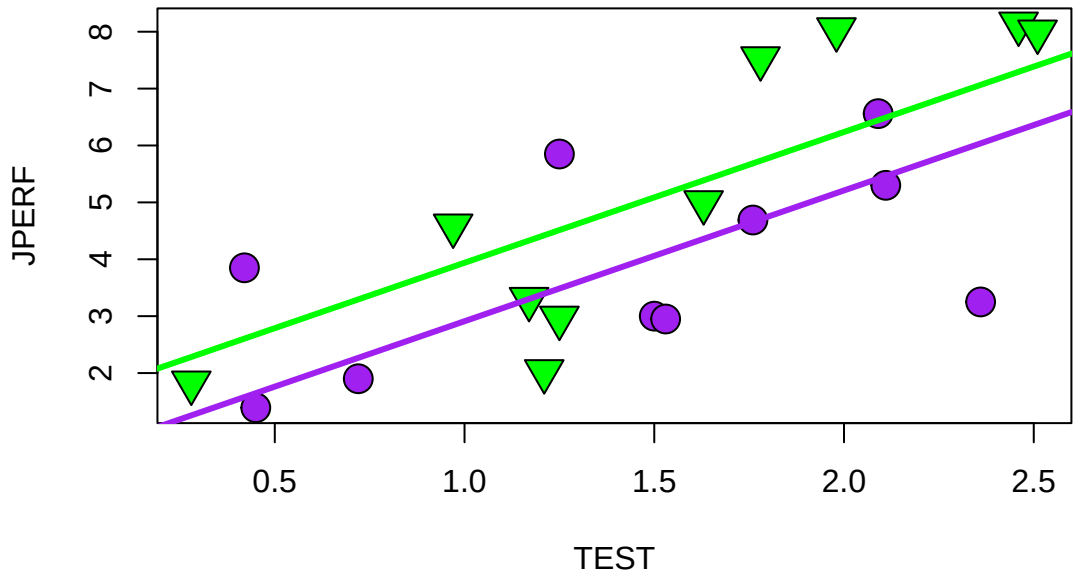



```
# page 34
jobtest.lm2 = lm(JPERF ~ TEST + ETHN)
print(summary(jobtest.lm2))
```

```
##
## Call:
## lm(formula = JPERF ~ TEST + ETHN)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7872 -1.0370 -0.2095  0.9198  2.3645
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6120     0.8870   0.690 0.499578
## TEST          2.2988     0.5225   4.400 0.000391 ***
## ETHN1         1.0276     0.6909   1.487 0.155246
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.54 on 17 degrees of freedom
## Multiple R-squared:  0.5724, Adjusted R-squared:  0.5221
## F-statistic: 11.38 on 2 and 17 DF,  p-value: 0.0007312
```

```
plot(TEST, JPERF, type='n')
points(TEST[(ETHN == 0)], JPERF[(ETHN == 0)], pch=21, cex=2, bg='purple')
points(TEST[(ETHN == 1)], JPERF[(ETHN == 1)], pch=25, cex=2, bg='green')
abline(jobtest.lm2$coef['(Intercept)'], jobtest.lm2$coef['TEST'], lwd=3, col='purple')
```

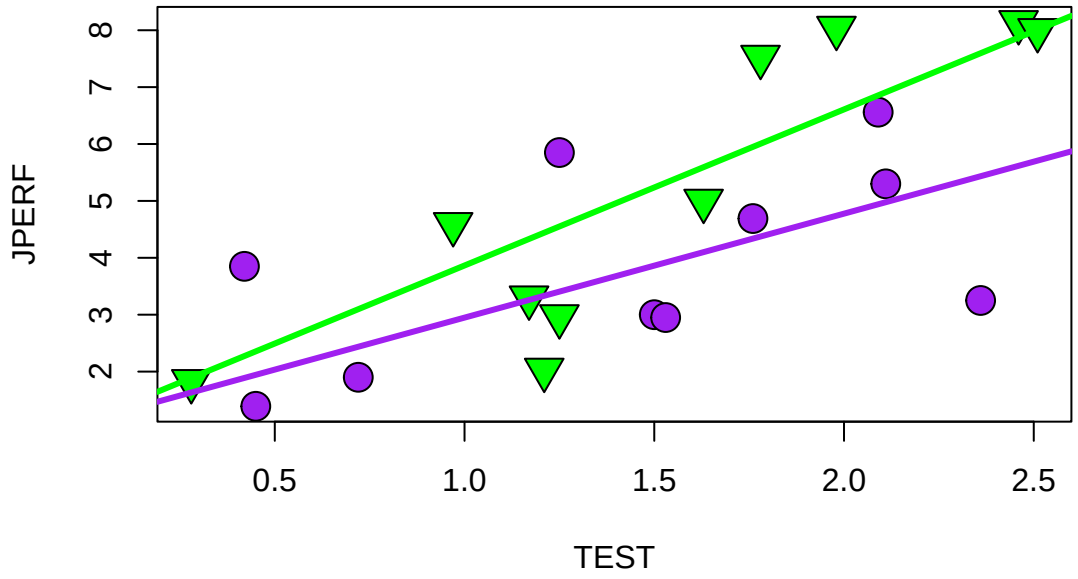
```r
abline(jobtest.lm2$coef['(Intercept)'] + jobtest.lm2$coef['ETHN1'], jobtest.lm2$coef['TEST'], lwd=3, col
```




```r
# page 35
jobtest.lm3 = lm(JPERF ~ TEST + TEST:ETHN)
print(summary(jobtest.lm3))
```

```
##
## Call:
## lm(formula = JPERF ~ TEST + TEST:ETHN)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.41100 -0.88871 -0.03359  0.97720  2.44440
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.1211     0.7804   1.437  0.16900
## TEST          1.8276     0.5356   3.412  0.00332 **
## TEST:ETHN1    0.9161     0.3972   2.306  0.03395 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.429 on 17 degrees of freedom
## Multiple R-squared:  0.6319, Adjusted R-squared:  0.5886
## F-statistic: 14.59 on 2 and 17 DF,  p-value: 0.0002045
```

```
plot(TEST, JPERF, type='n')
points(TEST[(ETHN == 0)], JPERF[(ETHN == 0)], pch=21, cex=2, bg='purple')
points(TEST[(ETHN == 1)], JPERF[(ETHN == 1)], pch=25, cex=2, bg='green')
abline(jobtest.lm3$coef['(Intercept)'], jobtest.lm3$coef['TEST'], lwd=3, col='purple')
abline(jobtest.lm3$coef['(Intercept)'], jobtest.lm3$coef['TEST'] + jobtest.lm3$coef['TEST:ETHN1'], lwd=
```




```
# page 36
head(model.matrix(jobtest.lm3))
```

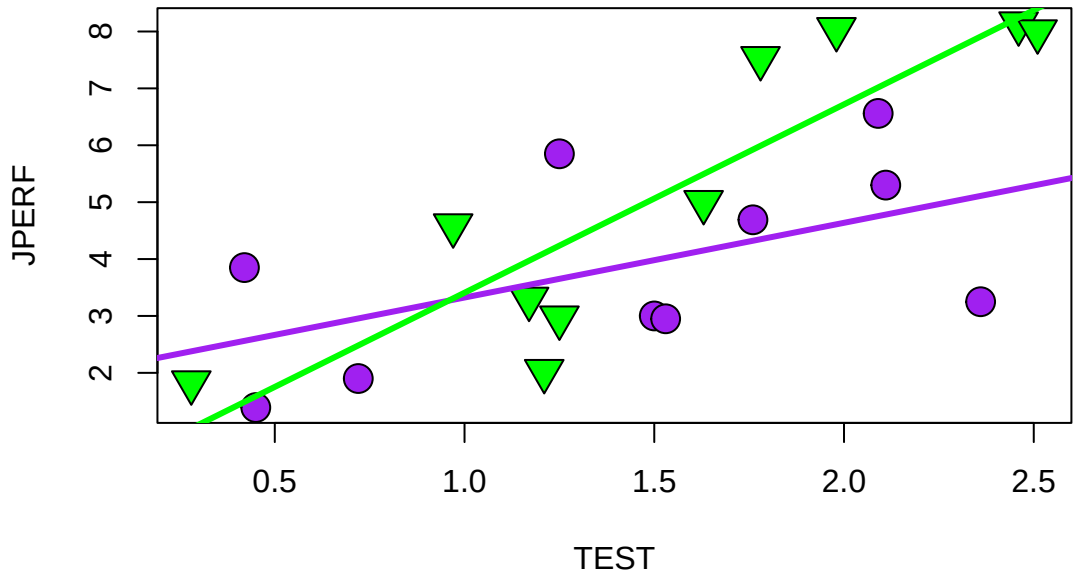
```
##   (Intercept) TEST TEST:ETHN1
## 1           1 0.28       0.28
## 2           1 0.97       0.97
## 3           1 1.25       1.25
## 4           1 2.46       2.46
## 5           1 2.51       2.51
## 6           1 1.17       1.17
```

```
# page 37
jobtest.lm4 = lm(JPERF ~ TEST * ETHN)
print(summary(jobtest.lm4))
```

```
##
## Call:
## lm(formula = JPERF ~ TEST * ETHN)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.0734 -1.0594 -0.2548  1.2830  2.1980
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.0103     1.0501   1.914   0.0736 .
## TEST          1.3134     0.6704   1.959   0.0677 .
## ETHN1        -1.9132     1.5403  -1.242   0.2321
## TEST:ETHN1    1.9975     0.9544   2.093   0.0527 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.407 on 16 degrees of freedom
## Multiple R-squared:  0.6643, Adjusted R-squared:  0.6013
## F-statistic: 10.55 on 3 and 16 DF,  p-value: 0.0004511
```

```
plot(TEST, JPERF, type='n')
points(TEST[(ETHN == 0)], JPERF[(ETHN == 0)], pch=21, cex=2, bg='purple')
points(TEST[(ETHN == 1)], JPERF[(ETHN == 1)], pch=25, cex=2, bg='green')
abline(jobtest.lm4$coef['(Intercept)'], jobtest.lm4$coef['TEST'], lwd=3, col='purple')
abline(jobtest.lm4$coef['(Intercept)'] + jobtest.lm4$coef['ETHN1'],
jobtest.lm4$coef['TEST'] + jobtest.lm4$coef['TEST:ETHN1'], lwd=3, col='green')
```




```
# page 38
head(model.matrix(jobtest.lm4))
```

```
##   (Intercept) TEST ETHN1 TEST:ETHN1
## 1           1 0.28     1       0.28
## 2           1 0.97     1       0.97
```

```
## 3           1 1.25     1       1.25
## 4           1 2.46     1       2.46
## 5           1 2.51     1       2.51
## 6           1 1.17     1       1.17
```

```
# page 39
anova(jobtest.lm1, jobtest.lm4)
```

```
## Analysis of Variance Table
##
## Model 1: JPERF ~ TEST
## Model 2: JPERF ~ TEST * ETHN
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     18 45.568
## 2     16 31.655  2    13.913 3.5161 0.05424 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# page 40
anova(jobtest.lm1, jobtest.lm2)
```

```
## Analysis of Variance Table
##
## Model 1: JPERF ~ TEST
## Model 2: JPERF ~ TEST + ETHN
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     18 45.568
## 2     17 40.322  1    5.2468 2.2121 0.1552
```

```
# page 41
anova(jobtest.lm3, jobtest.lm4)
```

```
## Analysis of Variance Table
##
## Model 1: JPERF ~ TEST + TEST:ETHN
## Model 2: JPERF ~ TEST * ETHN
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     17 34.708
## 2     16 31.655  1    3.0522 1.5427 0.2321
```

```
# page 42
anova(jobtest.lm1, jobtest.lm3)
```

```
## Analysis of Variance Table
##
## Model 1: JPERF ~ TEST
## Model 2: JPERF ~ TEST + TEST:ETHN
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     18 45.568
## 2     17 34.708  1    10.861 5.3196 0.03395 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# page 43
anova(jobtest.lm2, jobtest.lm4)
```

```
## Analysis of Variance Table
##
## Model 1: JPERF ~ TEST + ETHN
## Model 2: JPERF ~ TEST * ETHN
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     17 40.322
## 2     16 31.655  1    8.6661 4.3802 0.05265 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
detach(jobtest.table)
```

0.3 7. ANOVA

```r
### salary data
load("/Users/ohsukju/Downloads/data/salary.Rdata")
```

```r
## page 2
salary.table$E = factor(salary.table$E)
salary.table$M = factor(salary.table$M)
salary.lm = lm(S ~ X + E + M, salary.table)
head(model.matrix(salary.lm))
```

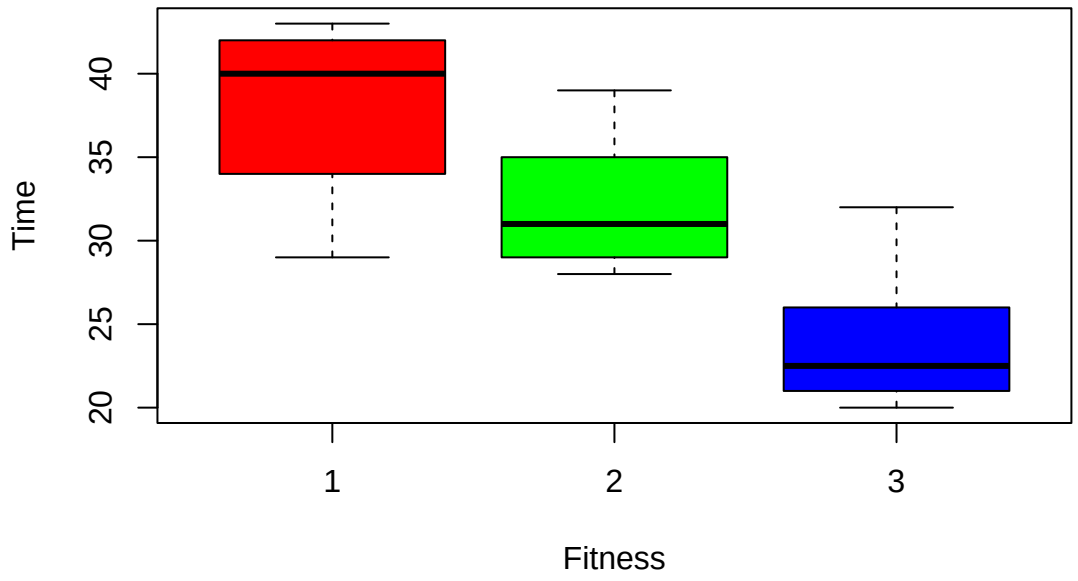
```
##   (Intercept) X E2 E3 M1
## 1           1 1  0  0  1
## 2           1 1  0  1  0
## 3           1 1  0  1  1
## 4           1 1  1  0  0
## 5           1 1  0  1  0
## 6           1 2  1  0  1
```

```r
## recovery time
load("/Users/ohsukju/Downloads/data/ANOVA.Rdata")
```

```r
## page 3
rehab.table$Fitness <- factor(rehab.table$Fitness)
head(rehab.table)
```

```
##   Fitness Time
## 1       1   29
## 2       1   42
## 3       1   38
## 4       1   40
## 5       1   43
## 6       1   40
```

```r
## page 4
attach(rehab.table)
boxplot(Time ~ Fitness, col=c('red','green','blue'))
```




```r
## page 8
rehab.lm <- lm(Time ~ Fitness)
summary(rehab.lm)
```

```
##
## Call:
## lm(formula = Time ~ Fitness)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -9.0   -3.0   -0.5    3.0    8.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   38.000      1.574  24.149  < 2e-16 ***
## Fitness2      -6.000      2.111  -2.842  0.00976 **
## Fitness3     -14.000      2.404  -5.824 8.81e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.451 on 21 degrees of freedom
## Multiple R-squared:  0.6176, Adjusted R-squared:  0.5812
## F-statistic: 16.96 on 2 and 21 DF,  p-value: 4.129e-05
```

```r
## page 9
print(predict(rehab.lm, list(Fitness=factor(c(1,2,3)))))
```

```
##  1  2  3
## 38 32 24
```

```r
c(mean(Time[Fitness == 1]), mean(Time[Fitness == 2]), mean(Time[Fitness == 3]))
```

```
## [1] 38 32 24
```

```r
## page 10
head(model.matrix(rehab.lm))
```

```
##   (Intercept) Fitness2 Fitness3
## 1           1        0        0
## 2           1        0        0
## 3           1        0        0
## 4           1        0        0
## 5           1        0        0
## 6           1        0        0
```

```r
## page 11
anova(rehab.lm)
```

```
## Analysis of Variance Table
##
## Response: Time
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Fitness    2    672  336.00  16.962 4.129e-05 ***
## Residuals 21    416   19.81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
F = 336.00 / 19.81
pval = 1 - pf(F, 2, 21)
print(data.frame(F,pval))
```

```
##          F         pval
## 1 16.96113 4.129945e-05
```

```r
## page 15
head(model.matrix(rehab.lm))
```

```
##   (Intercept) Fitness2 Fitness3
## 1           1        0        0
## 2           1        0        0
## 3           1        0        0
## 4           1        0        0
## 5           1        0        0
## 6           1        0        0
```

```r
detach(rehab.table)
```

```r
### Job example
## page 35
jobtest.table$ETHN <- factor(jobtest.table$ETHN)
jobtest.lm = lm(JPERF ~ TEST * ETHN, jobtest.table)
summary(jobtest.lm)
```

```
##
## Call:
## lm(formula = JPERF ~ TEST * ETHN, data = jobtest.table)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0734 -1.0594 -0.2548  1.2830  2.1980
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.0103     1.0501   1.914   0.0736 .
## TEST          1.3134     0.6704   1.959   0.0677 .
## ETHN1        -1.9132     1.5403  -1.242   0.2321
## TEST:ETHN1    1.9975     0.9544   2.093   0.0527 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.407 on 16 degrees of freedom
## Multiple R-squared:  0.6643, Adjusted R-squared:  0.6013
## F-statistic: 10.55 on 3 and 16 DF,  p-value: 0.0004511
```

```r
## page 36
anova(jobtest.lm)
```

```
## Analysis of Variance Table
##
## Response: JPERF
##           Df Sum Sq Mean Sq F value    Pr(>F)
## TEST       1 48.723  48.723 24.6266 0.0001412 ***
## ETHN       1  5.247   5.247  2.6519 0.1229524
## TEST:ETHN  1  8.666   8.666  4.3802 0.0526501 .
## Residuals 16 31.655   1.978
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
## page 37
anova(lm(JPERF ~ TEST, jobtest.table), lm(JPERF ~ TEST + ETHN, jobtest.table))
```

```
## Analysis of Variance Table
##
## Model 1: JPERF ~ TEST
## Model 2: JPERF ~ TEST + ETHN
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     18 45.568
## 2     17 40.322  1    5.2468 2.2121 0.1552
```

```r
## page 38
anova(lm(JPERF ~ 1, jobtest.table), lm(JPERF ~ TEST, jobtest.table))
```

```
## Analysis of Variance Table
##
## Model 1: JPERF ~ 1
## Model 2: JPERF ~ TEST
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     19 94.291
## 2     18 45.568  1    48.723 19.246 0.0003555 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
## page 29
library(car)
Anova(jobtest.lm, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: JPERF
##             Sum Sq Df F value  Pr(>F)
## (Intercept)  7.251  1  3.6647 0.07363 .
## TEST         7.594  1  3.8385 0.06775 .
## ETHN         3.052  1  1.5427 0.23211
## TEST:ETHN    8.666  1  4.3802 0.05265 .
## Residuals   31.655 16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
## page 30
summary(jobtest.lm)
```

```
##
## Call:
## lm(formula = JPERF ~ TEST * ETHN, data = jobtest.table)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0734 -1.0594 -0.2548  1.2830  2.1980
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.0103     1.0501   1.914   0.0736 .
## TEST          1.3134     0.6704   1.959   0.0677 .
## ETHN1        -1.9132     1.5403  -1.242   0.2321
## TEST:ETHN1    1.9975     0.9544   2.093   0.0527 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.407 on 16 degrees of freedom
## Multiple R-squared:  0.6643, Adjusted R-squared:  0.6013
## F-statistic: 10.55 on 3 and 16 DF,  p-value: 0.0004511
```