02441 Applied Statistics and Statistical Software

Exercise 5A - Forbes

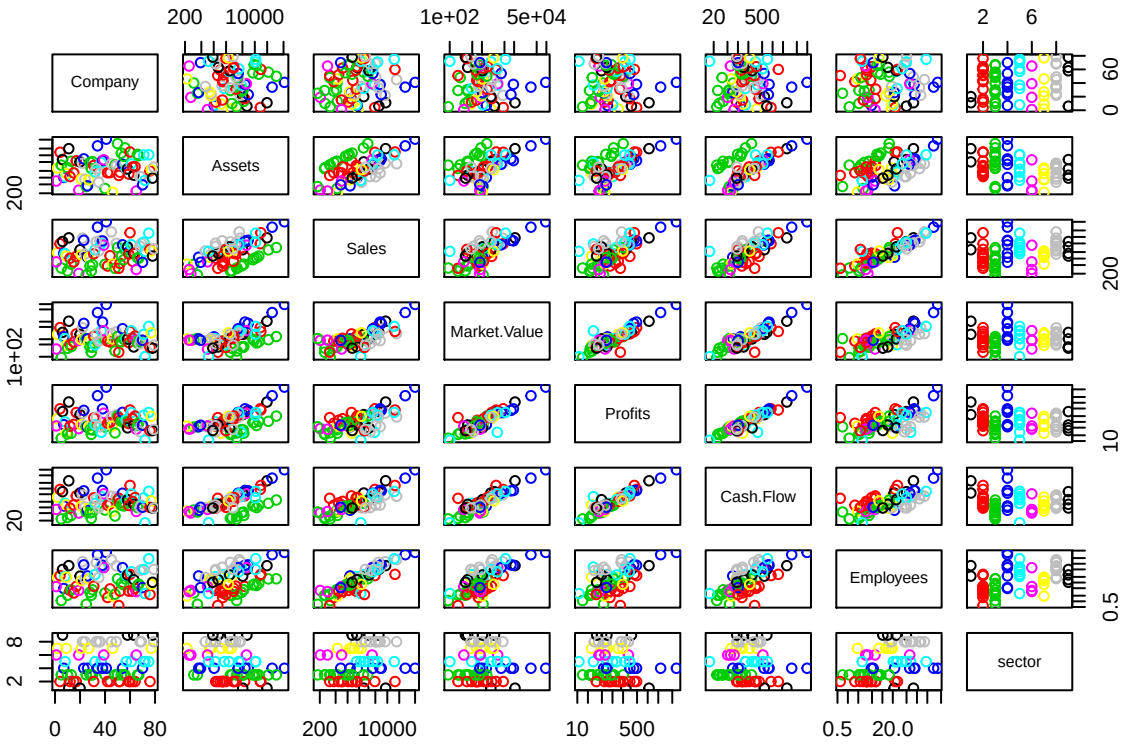
This dataset holds several facts about 77 companies selected from the Forbes 500 list for 1986. This is a 1/10 systematic sample from the alphabetical list of companies. The Forbes 500 includes all companies in the top 500 on any of the criteria, and thus has almost 800 companies in the list. Companies are often interested in how to increase sales. Many of the variables are skewed - a common occurrance with financial data - which suggests that much of the data are better analyzed after taking logarithms. For this study one should take the log of Sales and Assets.

Variable name	Description
Company	Company Name
Assets	Amount of assets (in millions)
Sales	Amount of sales (in millions)
Market.Value	Market Value of the company (in millions)
Profits	Profits (in millions)
Cash.Flow	Cash Flow (in millions)
Employees	Number of employees (in thousands)
Sector	Type of market the company is associated with

1. Investigate the relation between Log(Sales) predicted by Log(Assets)

Let's start with some visualizations. Logarithmic transformations come in handy for all continuous variables. Cashflow and Profits contain negativ values. These are resulting in missing data due to the log transformation. The corresponding warnings are suppressed in the code below.

```
forbes <- read.table("forbes.txt", header = TRUE, sep = '\t')
# Remove rows with negativ CashFlow and Profits to make this exercise
# a bit easier (log transformation of negative values will result in NaNs).
forbes <- forbes[forbes$Cash.Flow>0 & forbes$Profits>0,]
pairs(forbes, col = forbes$sector, log = c(2:7))
```



There is a linear relationship between log(Assets) and log(Sales). However, it also seems that sector is significant (different intercepts; represented by color in pairs plot).

```
lm1 <- lm(log(Sales)~log(Assets), forbes)
summary(lm1)
```

```
##
## Call:
## lm(formula = log(Sales) ~ log(Assets), data = forbes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9484 -0.8641  0.2083  0.6392  1.9710
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.10267    0.79559   3.900 0.000221 ***
## log(Assets)  0.56067    0.09966   5.626 3.66e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.003 on 69 degrees of freedom
## Multiple R-squared:  0.3144, Adjusted R-squared:  0.3045
## F-statistic: 31.65 on 1 and 69 DF,  p-value: 3.658e-07
```

Log(Assets) is significant at a 5% significance level. 31.4% of variance in log(Sales) is explained by the model `lm1`.

2. Include Sector as a discrete factor in the model - what is your conclusion?

```
lm2a <- lm(log(Sales)~log(Assets)+sector, forbes)
lm2b <- lm(log(Sales)~sector+log(Assets), forbes)
anova(lm2a)
```

```
## Analysis of Variance Table
##
## Response: log(Sales)
##             Df Sum Sq Mean Sq F value    Pr(>F)
## log(Assets)  1 31.820  31.820 139.346 < 2.2e-16 ***
## sector       8 55.445   6.931  30.351 < 2.2e-16 ***
## Residuals   61 13.929   0.228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm2b)
```

```
## Analysis of Variance Table
##
## Response: log(Sales)
##             Df Sum Sq Mean Sq F value    Pr(>F)
## sector       8 46.684   5.835  25.555 < 2.2e-16 ***
## log(Assets)  1 40.582  40.582 177.716 < 2.2e-16 ***
## Residuals   61 13.929   0.228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First of all, there is an issue here. The standard `anova` function performs sequential *Type I* tests. Depending on the order of the main effects in the model formula the results may differ when data is unbalanced. The Sum of Squares (SS) of `lm2a` are calculated such that

$SS(log(Assets))$ for Assets and $SS(sector|log(Assets))$ for sector.

Using *Type II* tests you can make sure to control for other main effects when calculating Sum of Squares. In this particular example *Type II* Anova calculates the SS, such that

$SS(log(Assets)|sector)$ for Assets and $SS(sector|log(Assets))$ for sector.

The `Anova` function of the `car` package performs *Type II* tests. It yields similar results to `drop1` and the result does not depend on effect sequence in the model formula.

```
library(car)
```

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

```
Anova(lm2a)
```

```
## Anova Table (Type II tests)
##
## Response: log(Sales)
##             Sum Sq Df F value    Pr(>F)
## log(Assets) 40.582  1 177.716 < 2.2e-16 ***
## sector      55.445  8  30.351 < 2.2e-16 ***
## Residuals   13.929 61
```

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

```
Anova(lm2b)
```

```
## Anova Table (Type II tests)
##
## Response: log(Sales)
##             Sum Sq Df F value    Pr(>F)
## sector      55.445  8  30.351 < 2.2e-16 ***
## log(Assets) 40.582  1 177.716 < 2.2e-16 ***
## Residuals   13.929 61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm2a, test = 'F')
```

```
## Single term deletions
##
## Model:
## log(Sales) ~ log(Assets) + sector
##             Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                   13.929 -95.636
## log(Assets)  1    40.582 54.511  -0.764 177.716 < 2.2e-16 ***
## sector       8    55.445 69.375   2.356  30.351 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm2b, test = 'F')
```

```
## Single term deletions
##
## Model:
## log(Sales) ~ sector + log(Assets)
##             Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                   13.929 -95.636
## sector       8    55.445 69.375   2.356  30.351 < 2.2e-16 ***
## log(Assets)  1    40.582 54.511  -0.764 177.716 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

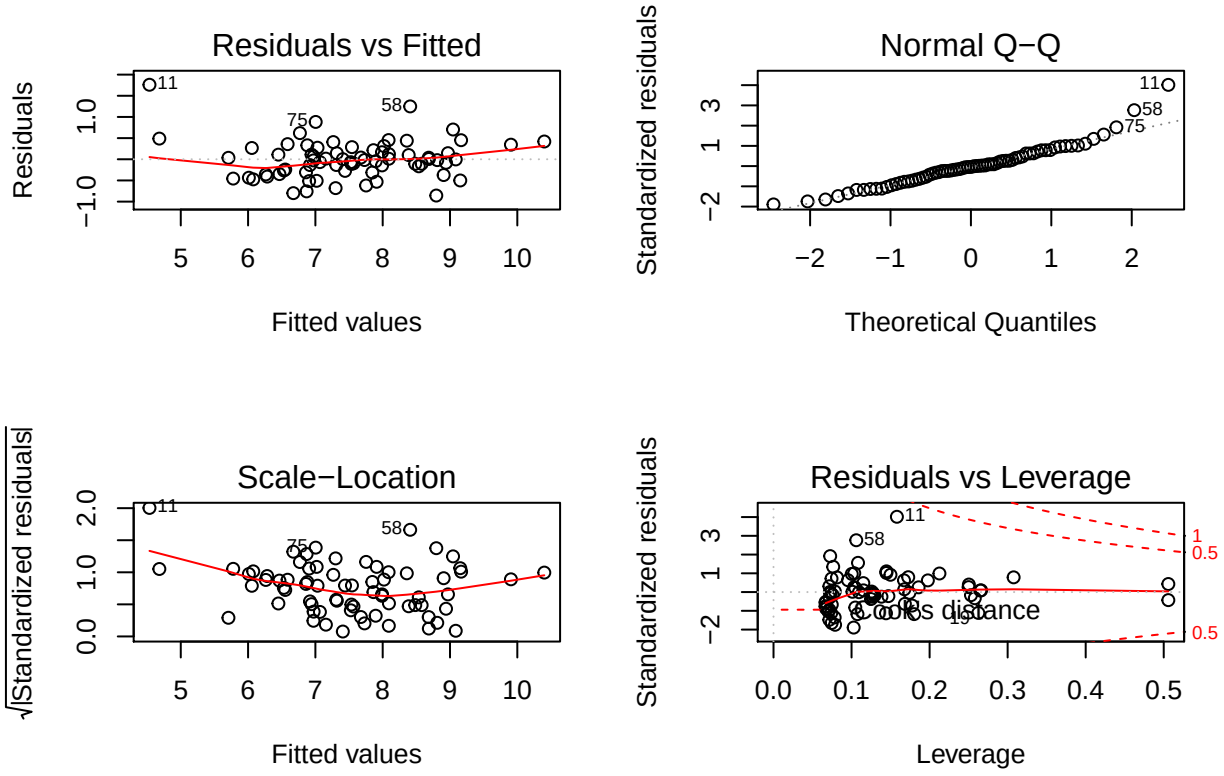
```
summary(lm2a)
```

```
##
## Call:
## lm(formula = log(Sales) ~ log(Assets) + sector, data = forbes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.85688 -0.29117 -0.01791  0.24077  1.76024
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.85366    0.60106   3.084 0.003067 **
## log(Assets)          0.71934    0.05396  13.331  < 2e-16 ***
## sectorEnergy        -0.31491    0.36808  -0.856 0.395587
## sectorFinance       -1.48355    0.36195  -4.099 0.000125 ***
```

```
## sectorHiTech           0.72727    0.38134   1.907 0.061211 .
## sectorManufacturing    0.64875    0.38323   1.693 0.095586 .
## sectorMedical         -0.12279    0.43693  -0.281 0.779638
## sectorOther            0.31251    0.40851   0.765 0.447219
## sectorRetail           1.19386    0.38112   3.133 0.002662 **
## sectorTransportation   0.52312    0.42237   1.239 0.220263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4779 on 61 degrees of freedom
## Multiple R-squared:  0.8624, Adjusted R-squared:  0.842
## F-statistic: 42.46 on 9 and 61 DF,  p-value: < 2.2e-16
```

Both, sector and log(Assets) are significant. 86.2% of variance in log(Sales) is explained by the model `lm2a`. Let's check if the model assumptions are fulfilled.

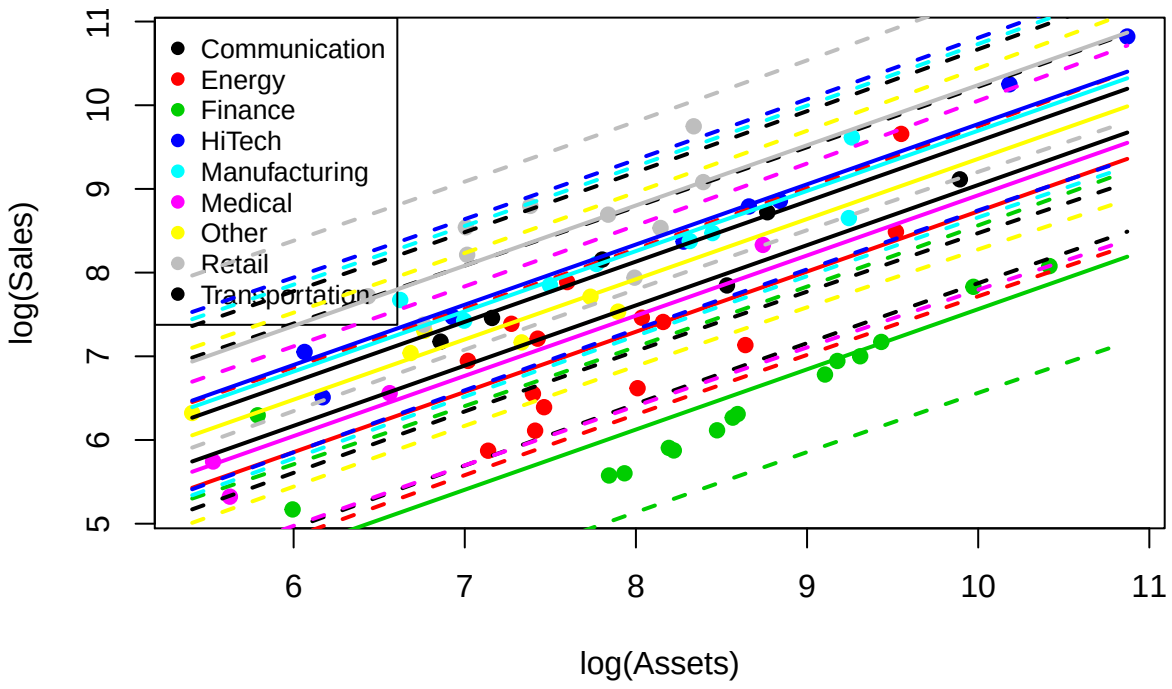
```
par(mfrow=c(2,2))
plot(lm2a)
```



Let's make a plot of the results.

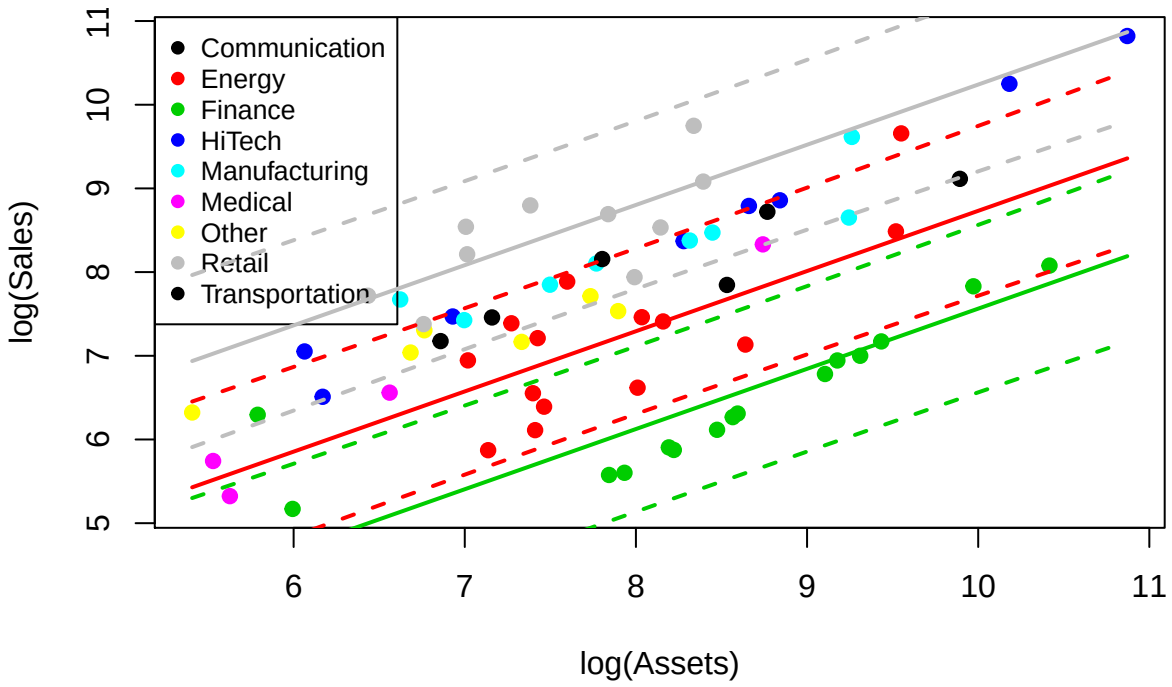
```
plot(log(Sales)~log(Assets), forbes, col = forbes$sector, pch = 19)
legend("topleft", legend = levels(forbes$sector), col = 1:nlevels(forbes$sector),
       pch = 19, cex = 0.8)
new_data <- seq(min(forbes$Assets), max(forbes$Assets), length.out = 100)
```

```
new_data_grid <- expand.grid(Assets = new_data, sector = levels(forbes$sector))
color_codes <- as.character(levels(forbes$sector))

plot_confidence <- function(sector){
  matlines(log(new_data), predict(lm2a, new_data_grid[new_data_grid$sector==color_codes[sector],],
                                  interval = "prediction"), lty = c(1,2,2), lw = 2,
           col = c(sector, sector, sector))
}

for (sector in c(1:9)) {
  plot_confidence(sector)
}
```



The plot above shows 95% prediction intervals for all sectors, but it is not very informative due to information overflow. Let's only plot the prediction intervals for three sectors. In the R chunk below we we use the `sapply` command. `apply` type functions can be viewed as a substitute to the loop and are often computationally more efficient.

```
plot(log(Sales)~log(Assets), forbes, col = forbes$sector, pch = 19)
legend("topleft", legend = levels(forbes$sector), col = 1:nlevels(forbes$sector),
       pch = 19, cex = 0.8)
lapply(c(2,3,8), plot_confidence)
```




```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
```

3. Are other variables significant?

```
lm3a <- lm(log(Sales)~(log(Assets)+log(Employees)+log(Cash.Flow)+
                         log(Profits)+log(Market.Value))*sector, forbes, na.action = )
Anova(lm3a)
```

```
## Note: model has aliased coefficients
##       sums of squares computed by model comparison
```

```
## Anova Table (Type II tests)
##
## Response: log(Sales)
##                           Sum Sq Df F value    Pr(>F)
## log(Assets)               0.1327  1  1.8326 0.1879311
## log(Employees)            0.9360  1 12.9309 0.0013869 **
## log(Cash.Flow)            0.0904  1  1.2486 0.2744506
## log(Profits)              0.0365  1  0.5036 0.4844883
## log(Market.Value)         0.0210  1  0.2896 0.5952301
## sector                    3.4021  8  5.8751 0.0002869 ***
## log(Assets):sector        0.7837  5  2.1654 0.0904546 .
## log(Employees):sector     0.7707  5  2.1294 0.0950223 .
## log(Cash.Flow):sector     1.1097  5  3.0662 0.0271135 *
## log(Profits):sector       0.8852  5  2.4459 0.0617695 .
```

```
## log(Market.Value):sector 0.1728  5  0.4775 0.7895251
## Residuals                1.8096 25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm3a)
```

```
##
## Call:
## lm(formula = log(Sales) ~ (log(Assets) + log(Employees) + log(Cash.Flow) +
##     log(Profits) + log(Market.Value)) * sector, data = forbes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36129 -0.05652  0.00000  0.06643  0.51219
##
## Coefficients: (8 not defined because of singularities)
##                                    Estimate Std. Error t value
## (Intercept)                       -10.65222   21.31733  -0.500
## log(Assets)                         1.18820    3.42516   0.347
## log(Employees)                     -3.82151    7.24437  -0.528
## log(Cash.Flow)                      4.02263    6.75290   0.596
## log(Profits)                       -0.74641    0.72544  -1.029
## log(Market.Value)                  -0.17727    0.80261  -0.221
## sectorEnergy                       16.50212   21.37247   0.772
## sectorFinance                      13.97121   21.41017   0.653
## sectorHiTech                       11.73447   21.39026   0.549
## sectorManufacturing                15.04533   21.36807   0.704
## sectorMedical                       8.39602   24.74028   0.339
## sectorOther                        -3.70877   31.28291  -0.119
## sectorRetail                       16.03921   19.90827   0.806
## sectorTransportation                6.49592    9.63048   0.675
## log(Assets):sectorEnergy           -0.82495    3.45429  -0.239
## log(Assets):sectorFinance          -1.16113    3.42851  -0.339
## log(Assets):sectorHiTech            0.29838    3.53057   0.085
## log(Assets):sectorManufacturing    -0.68998    3.43245  -0.201
## log(Assets):sectorMedical          -0.14030    5.25602  -0.027
## log(Assets):sectorOther            -0.23308    3.49189  -0.067
## log(Assets):sectorRetail           -1.61685    3.46798  -0.466
## log(Assets):sectorTransportation   -0.35736    0.62002  -0.576
## log(Employees):sectorEnergy         4.70485    7.24765   0.649
## log(Employees):sectorFinance        4.19797    7.24835   0.579
## log(Employees):sectorHiTech         3.82270    7.24655   0.528
## log(Employees):sectorManufacturing  4.29748    7.25189   0.593
## log(Employees):sectorMedical        2.78894    6.99533   0.399
## log(Employees):sectorOther          5.03365    7.39198   0.681
## log(Employees):sectorRetail         4.49161    7.13917   0.629
## log(Employees):sectorTransportation      NA         NA      NA
## log(Cash.Flow):sectorEnergy        -2.82369    6.77741  -0.417
## log(Cash.Flow):sectorFinance       -2.18709    6.79354  -0.322
## log(Cash.Flow):sectorHiTech        -6.69059    7.04549  -0.950
## log(Cash.Flow):sectorManufacturing -4.75398    6.77000  -0.702
## log(Cash.Flow):sectorMedical       -2.21363    7.26272  -0.305
## log(Cash.Flow):sectorOther         -6.54255    7.27923  -0.899
## log(Cash.Flow):sectorRetail        -2.43601    6.60501  -0.369
```

```
## log(Cash.Flow):sectorTransportation           NA         NA      NA
## log(Profits):sectorEnergy                -0.59094    0.87133  -0.678
## log(Profits):sectorFinance               -0.27130    1.04558  -0.259
## log(Profits):sectorHiTech                 2.77955    1.95598   1.421
## log(Profits):sectorManufacturing          0.85105    0.74433   1.143
## log(Profits):sectorMedical                     NA         NA      NA
## log(Profits):sectorOther                 -2.61723    4.70904  -0.556
## log(Profits):sectorRetail                      NA         NA      NA
## log(Profits):sectorTransportation              NA         NA      NA
## log(Market.Value):sectorEnergy           -0.23077    0.88875  -0.260
## log(Market.Value):sectorFinance          -0.03707    0.87224  -0.042
## log(Market.Value):sectorHiTech            0.19214    1.10948   0.173
## log(Market.Value):sectorManufacturing     0.41673    0.87463   0.476
## log(Market.Value):sectorMedical                NA         NA      NA
## log(Market.Value):sectorOther             5.79333    7.19698   0.805
## log(Market.Value):sectorRetail                 NA         NA      NA
## log(Market.Value):sectorTransportation         NA         NA      NA
##                                        Pr(>|t|)
## (Intercept)                               0.622
## log(Assets)                               0.732
## log(Employees)                            0.602
## log(Cash.Flow)                            0.557
## log(Profits)                              0.313
## log(Market.Value)                         0.827
## sectorEnergy                              0.447
## sectorFinance                             0.520
## sectorHiTech                              0.588
## sectorManufacturing                       0.488
## sectorMedical                             0.737
## sectorOther                               0.907
## sectorRetail                              0.428
## sectorTransportation                      0.506
## log(Assets):sectorEnergy                  0.813
## log(Assets):sectorFinance                 0.738
## log(Assets):sectorHiTech                  0.933
## log(Assets):sectorManufacturing           0.842
## log(Assets):sectorMedical                 0.979
## log(Assets):sectorOther                   0.947
## log(Assets):sectorRetail                  0.645
## log(Assets):sectorTransportation          0.570
## log(Employees):sectorEnergy               0.522
## log(Employees):sectorFinance              0.568
## log(Employees):sectorHiTech               0.602
## log(Employees):sectorManufacturing        0.559
## log(Employees):sectorMedical              0.694
## log(Employees):sectorOther                0.502
## log(Employees):sectorRetail               0.535
## log(Employees):sectorTransportation          NA
## log(Cash.Flow):sectorEnergy               0.681
## log(Cash.Flow):sectorFinance              0.750
## log(Cash.Flow):sectorHiTech               0.351
## log(Cash.Flow):sectorManufacturing        0.489
## log(Cash.Flow):sectorMedical              0.763
## log(Cash.Flow):sectorOther                0.377
```

```
## log(Cash.Flow):sectorRetail               0.715
## log(Cash.Flow):sectorTransportation          NA
## log(Profits):sectorEnergy                 0.504
## log(Profits):sectorFinance                0.797
## log(Profits):sectorHiTech                 0.168
## log(Profits):sectorManufacturing          0.264
## log(Profits):sectorMedical                   NA
## log(Profits):sectorOther                  0.583
## log(Profits):sectorRetail                    NA
## log(Profits):sectorTransportation            NA
## log(Market.Value):sectorEnergy            0.797
## log(Market.Value):sectorFinance           0.966
## log(Market.Value):sectorHiTech            0.864
## log(Market.Value):sectorManufacturing     0.638
## log(Market.Value):sectorMedical              NA
## log(Market.Value):sectorOther             0.428
## log(Market.Value):sectorRetail               NA
## log(Market.Value):sectorTransportation       NA
##
## Residual standard error: 0.269 on 25 degrees of freedom
## Multiple R-squared:  0.9821, Adjusted R-squared:  0.9499
## F-statistic: 30.51 on 45 and 25 DF,  p-value: 7.238e-14
```

When looking at the summary of model `lm3a` it is clear that some coefficients could not be estimated (Coefficients: (8 not defined because of singularities)). This is due to perfect multicolinearity or said differently, some of the coefficients can be expressed as a linear combination of others. When this happens you might also get a warning stating that coefficients are aliased.

To fix this problem use the `alias` function to find out which coefficients are affected.

```
alias(lm3a)
```

```
## Model :
## log(Sales) ~ (log(Assets) + log(Employees) + log(Cash.Flow) +
##     log(Profits) + log(Market.Value)) * sector
##
## Complete :
##                                        (Intercept)
## log(Employees):sectorTransportation                    843/169
## log(Cash.Flow):sectorTransportation          36374081/12095508
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                            72875/3097
## log(Profits):sectorTransportation              -2608497/133736
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                     22326/87919
## log(Market.Value):sectorTransportation      352317227/116471077
##                                        log(Assets)
## log(Employees):sectorTransportation               -11214/11849
## log(Cash.Flow):sectorTransportation               -101954/92861
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                           -2462/2327
## log(Profits):sectorTransportation                 -37381/665302
```

```
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                   -42341615/21079183
## log(Market.Value):sectorTransportation               158741/214512
##                                          log(Employees)
## log(Employees):sectorTransportation                           1
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                           371813/45315
## log(Profits):sectorTransportation                  -163921/19978
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -12105/5741
## log(Market.Value):sectorTransportation                17254/8183
##                                          log(Cash.Flow)
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                           1
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                        -1090707/153104
## log(Profits):sectorTransportation                 2526976/354715
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                     568117/228017
## log(Market.Value):sectorTransportation             -240465/96512
##                                          log(Profits)
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                                     1
## log(Profits):sectorTransportation                             0
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                                0
## log(Market.Value):sectorTransportation                        0
##                                          log(Market.Value)
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                                     0
## log(Profits):sectorTransportation                             0
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                                1
## log(Market.Value):sectorTransportation                        0
##                                          sectorEnergy
## log(Employees):sectorTransportation                    -843/169
## log(Cash.Flow):sectorTransportation            -9704934/3227191
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                          -156433/6648
## log(Profits):sectorTransportation        91616462715/4697118401
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                     -13315/52434
## log(Market.Value):sectorTransportation        -16365182/5410097
##                                          sectorFinance
## log(Employees):sectorTransportation                    -843/169
## log(Cash.Flow):sectorTransportation            -16964213/5641126
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                          -156433/6648
## log(Profits):sectorTransportation      204128547449/10465542196
```

```
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                       -13315/52434
## log(Market.Value):sectorTransportation          -19824544/6553713
##                                            sectorHiTech
## log(Employees):sectorTransportation                     -843/169
## log(Cash.Flow):sectorTransportation              -9704934/3227191
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                          -156433/6648
## log(Profits):sectorTransportation            2454017359/125815926
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -13315/52434
## log(Market.Value):sectorTransportation          -19824544/6553713
##                                            sectorManufacturing
## log(Employees):sectorTransportation                     -843/169
## log(Cash.Flow):sectorTransportation              -9704934/3227191
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                          -156433/6648
## log(Profits):sectorTransportation              967463228/49601231
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -13315/52434
## log(Market.Value):sectorTransportation          -19824544/6553713
##                                            sectorMedical
## log(Employees):sectorTransportation                     -843/169
## log(Cash.Flow):sectorTransportation           -99494995/33085166
## log(Profits):sectorMedical                               7185/2641
## log(Profits):sectorRetail                     -262536255/10000849
## log(Profits):sectorTransportation             1908835420/97864791
## log(Market.Value):sectorMedical                  71053997/4069157
## log(Market.Value):sectorRetail                       -157438/8887
## log(Market.Value):sectorTransportation        -68786515/22739846
##                                            sectorOther
## log(Employees):sectorTransportation                     -843/169
## log(Cash.Flow):sectorTransportation            -21622465/7190139
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                          -156433/6648
## log(Profits):sectorTransportation           6102732410/312883251
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -13315/52434
## log(Market.Value):sectorTransportation       -128928022/42621775
##                                            sectorRetail
## log(Employees):sectorTransportation                     -843/169
## log(Cash.Flow):sectorTransportation              -9704934/3227191
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                          -156433/6648
## log(Profits):sectorTransportation          10738314672/550546637
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -13315/52434
## log(Market.Value):sectorTransportation          -18094863/5981905
##                                            sectorTransportation
## log(Employees):sectorTransportation                     -843/169
## log(Cash.Flow):sectorTransportation              -9704934/3227191
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                           -971614/91981
## log(Profits):sectorTransportation                19115593/2924130
```

```
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                       207809/198959
## log(Market.Value):sectorTransportation           -6101734/1411343
##                                         log(Assets):sectorEnergy
## log(Employees):sectorTransportation                     4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                               2462/2327
## log(Profits):sectorTransportation                    76261/1357283
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                  66206473/32959970
## log(Market.Value):sectorTransportation             -498686/673891
##                                         log(Assets):sectorFinance
## log(Employees):sectorTransportation                     4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                               2462/2327
## log(Profits):sectorTransportation                    76261/1357283
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                  60818372/30277579
## log(Market.Value):sectorTransportation             -498686/673891
##                                         log(Assets):sectorHiTech
## log(Employees):sectorTransportation                     4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                               2462/2327
## log(Profits):sectorTransportation                   113642/2022585
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail               373633461/186008212
## log(Market.Value):sectorTransportation             -498686/673891
##                                         log(Assets):sectorManufacturing
## log(Employees):sectorTransportation                     4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                               2462/2327
## log(Profits):sectorTransportation                    76261/1357283
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail               204267001/101691480
## log(Market.Value):sectorTransportation             -498686/673891
##                                         log(Assets):sectorMedical
## log(Employees):sectorTransportation                     4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                           -43764/86764247
## log(Profits):sectorRetail                           1560546/1474273
## log(Profits):sectorTransportation                    113642/2022585
## log(Market.Value):sectorMedical                       -55982/16383
## log(Market.Value):sectorRetail                           9137/1684
## log(Market.Value):sectorTransportation             -498686/673891
##                                         log(Assets):sectorOther
## log(Employees):sectorTransportation                     4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                               2462/2327
## log(Profits):sectorTransportation                    76261/1357283
```

```
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                1005964395/500805356
## log(Market.Value):sectorTransportation             -498686/673891
##                                          log(Assets):sectorRetail
## log(Employees):sectorTransportation                    4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                               2462/2327
## log(Profits):sectorTransportation                      76261/1357283
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                373633461/186008212
## log(Market.Value):sectorTransportation             -498686/673891
##                                          log(Assets):sectorTransportation
## log(Employees):sectorTransportation                    4309/4553
## log(Cash.Flow):sectorTransportation                  210983/192166
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                            782311/1389520
## log(Profits):sectorTransportation                       393/713
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                         1698/31097
## log(Market.Value):sectorTransportation              1857817/1530226
##                                          log(Employees):sectorEnergy
## log(Employees):sectorTransportation                          -1
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                           -163921/19978
## log(Profits):sectorTransportation                    371813/45315
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                         17254/8183
## log(Market.Value):sectorTransportation              -12105/5741
##                                          log(Employees):sectorFinance
## log(Employees):sectorTransportation                          -1
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                           -163921/19978
## log(Profits):sectorTransportation                    371813/45315
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                         17254/8183
## log(Market.Value):sectorTransportation              -12105/5741
##                                          log(Employees):sectorHiTech
## log(Employees):sectorTransportation                          -1
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                           -163921/19978
## log(Profits):sectorTransportation                    371813/45315
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                         17254/8183
## log(Market.Value):sectorTransportation              -12105/5741
##                                          log(Employees):sectorManufacturing
## log(Employees):sectorTransportation                          -1
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                           -163921/19978
## log(Profits):sectorTransportation                    371813/45315
```

```
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                          17254/8183
## log(Market.Value):sectorTransportation                 -12105/5741
##                                          log(Employees):sectorMedical
## log(Employees):sectorTransportation                             -1
## log(Cash.Flow):sectorTransportation                              0
## log(Profits):sectorMedical                                1613/2682
## log(Profits):sectorRetail                              -1326566/150635
## log(Profits):sectorTransportation                        371813/45315
## log(Market.Value):sectorMedical                          392260/114339
## log(Market.Value):sectorRetail                             -5245/3967
## log(Market.Value):sectorTransportation                   -12105/5741
##                                          log(Employees):sectorOther
## log(Employees):sectorTransportation                             -1
## log(Cash.Flow):sectorTransportation                              0
## log(Profits):sectorMedical                                       0
## log(Profits):sectorRetail                              -163921/19978
## log(Profits):sectorTransportation                        371813/45315
## log(Market.Value):sectorMedical                                  0
## log(Market.Value):sectorRetail                           17254/8183
## log(Market.Value):sectorTransportation                   -12105/5741
##                                          log(Employees):sectorRetail
## log(Employees):sectorTransportation                             -1
## log(Cash.Flow):sectorTransportation                              0
## log(Profits):sectorMedical                                       0
## log(Profits):sectorRetail                              -163921/19978
## log(Profits):sectorTransportation                        371813/45315
## log(Market.Value):sectorMedical                                  0
## log(Market.Value):sectorRetail                           17254/8183
## log(Market.Value):sectorTransportation                   -12105/5741
##                                          log(Cash.Flow):sectorEnergy
## log(Employees):sectorTransportation                              0
## log(Cash.Flow):sectorTransportation                             -1
## log(Profits):sectorMedical                                       0
## log(Profits):sectorRetail                              2526976/354715
## log(Profits):sectorTransportation                     -1090707/153104
## log(Market.Value):sectorMedical                                  0
## log(Market.Value):sectorRetail                           -240465/96512
## log(Market.Value):sectorTransportation                   568117/228017
##                                          log(Cash.Flow):sectorFinance
## log(Employees):sectorTransportation                              0
## log(Cash.Flow):sectorTransportation                             -1
## log(Profits):sectorMedical                                       0
## log(Profits):sectorRetail                              2526976/354715
## log(Profits):sectorTransportation                     -1090707/153104
## log(Market.Value):sectorMedical                                  0
## log(Market.Value):sectorRetail                           -240465/96512
## log(Market.Value):sectorTransportation                   568117/228017
##                                          log(Cash.Flow):sectorHiTech
## log(Employees):sectorTransportation                              0
## log(Cash.Flow):sectorTransportation                             -1
## log(Profits):sectorMedical                                       0
## log(Profits):sectorRetail                              2526976/354715
## log(Profits):sectorTransportation                     -1090707/153104
```

```
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                       -240465/96512
## log(Market.Value):sectorTransportation               568117/228017
##                                          log(Cash.Flow):sectorManufacturing
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                          -1
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                             2526976/354715
## log(Profits):sectorTransportation                    -1090707/153104
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -240465/96512
## log(Market.Value):sectorTransportation               568117/228017
##                                          log(Cash.Flow):sectorMedical
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                          -1
## log(Profits):sectorMedical                             153757/2141490
## log(Profits):sectorRetail                            10230029/1450623
## log(Profits):sectorTransportation                    -1090707/153104
## log(Market.Value):sectorMedical                       4336617/4319816
## log(Market.Value):sectorRetail                           -21102/6037
## log(Market.Value):sectorTransportation                568117/228017
##                                          log(Cash.Flow):sectorOther
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                          -1
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                             2526976/354715
## log(Profits):sectorTransportation                    -1090707/153104
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -240465/96512
## log(Market.Value):sectorTransportation               568117/228017
##                                          log(Cash.Flow):sectorRetail
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                          -1
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                             2526976/354715
## log(Profits):sectorTransportation                    -1090707/153104
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                       -240465/96512
## log(Market.Value):sectorTransportation               568117/228017
##                                          log(Profits):sectorEnergy
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                                    -1
## log(Profits):sectorTransportation                             0
## log(Market.Value):sectorMedical                               0
## log(Market.Value):sectorRetail                                0
## log(Market.Value):sectorTransportation                        0
##                                          log(Profits):sectorFinance
## log(Employees):sectorTransportation                           0
## log(Cash.Flow):sectorTransportation                           0
## log(Profits):sectorMedical                                    0
## log(Profits):sectorRetail                                    -1
## log(Profits):sectorTransportation                             0
```

```
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                               0
## log(Market.Value):sectorTransportation                       0
##                                         log(Profits):sectorHiTech
## log(Employees):sectorTransportation                          0
## log(Cash.Flow):sectorTransportation                          0
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                                   -1
## log(Profits):sectorTransportation                            0
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                               0
## log(Market.Value):sectorTransportation                       0
##                                         log(Profits):sectorManufacturing
## log(Employees):sectorTransportation                          0
## log(Cash.Flow):sectorTransportation                          0
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                                   -1
## log(Profits):sectorTransportation                            0
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                               0
## log(Market.Value):sectorTransportation                       0
##                                         log(Profits):sectorOther
## log(Employees):sectorTransportation                          0
## log(Cash.Flow):sectorTransportation                          0
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                                   -1
## log(Profits):sectorTransportation                            0
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                               0
## log(Market.Value):sectorTransportation                       0
##                                         log(Market.Value):sectorEnergy
## log(Employees):sectorTransportation                          0
## log(Cash.Flow):sectorTransportation                          0
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                                    0
## log(Profits):sectorTransportation                            0
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                              -1
## log(Market.Value):sectorTransportation                       0
##                                         log(Market.Value):sectorFinance
## log(Employees):sectorTransportation                          0
## log(Cash.Flow):sectorTransportation                          0
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                                    0
## log(Profits):sectorTransportation                            0
## log(Market.Value):sectorMedical                              0
## log(Market.Value):sectorRetail                              -1
## log(Market.Value):sectorTransportation                       0
##                                         log(Market.Value):sectorHiTech
## log(Employees):sectorTransportation                          0
## log(Cash.Flow):sectorTransportation                          0
## log(Profits):sectorMedical                                   0
## log(Profits):sectorRetail                                    0
## log(Profits):sectorTransportation                            0
```

```
## log(Market.Value):sectorMedical                                0
## log(Market.Value):sectorRetail                                -1
## log(Market.Value):sectorTransportation                         0
##                                       log(Market.Value):sectorManufacturing
## log(Employees):sectorTransportation                            0
## log(Cash.Flow):sectorTransportation                            0
## log(Profits):sectorMedical                                     0
## log(Profits):sectorRetail                                      0
## log(Profits):sectorTransportation                              0
## log(Market.Value):sectorMedical                                0
## log(Market.Value):sectorRetail                                -1
## log(Market.Value):sectorTransportation                         0
##                                       log(Market.Value):sectorOther
## log(Employees):sectorTransportation                            0
## log(Cash.Flow):sectorTransportation                            0
## log(Profits):sectorMedical                                     0
## log(Profits):sectorRetail                                      0
## log(Profits):sectorTransportation                              0
## log(Market.Value):sectorMedical                                0
## log(Market.Value):sectorRetail                                -1
## log(Market.Value):sectorTransportation                         0
```

Let's remove all the affected interaction effects from our model (as shown in the alias table above). Afterwards we can reduce the model using backwards selection. Again `drop1` helps us not to make mistakes during this process.

```
lm3b <- update(lm3a, ~.-log(Market.Value):sector-log(Profits):sector-
                 log(Cash.Flow):sector-log(Employees):sector)
drop1(lm3b, test = "F")
```

```
## Single term deletions
##
## Model:
## log(Sales) ~ log(Assets) + log(Employees) + log(Cash.Flow) +
##     log(Profits) + log(Market.Value) + sector + log(Assets):sector
##                    Df Sum of Sq     RSS      AIC F value   Pr(>F)
## <none>                           8.2636 -108.708
## log(Employees)      1   1.93403 10.1976  -95.777 11.4681 0.001403 **
## log(Cash.Flow)      1   0.12489  8.3885 -109.643  0.7405 0.393683
## log(Profits)        1   0.04449  8.3081 -110.327  0.2638 0.609816
## log(Market.Value)   1   0.00000  8.2636 -110.708  0.0000 0.997347
## log(Assets):sector  8   0.92658  9.1902 -117.163  0.6868 0.701003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm3c <- update(lm3b, ~.-log(Market.Value)), test = "F")
```

```
## Single term deletions
##
## Model:
## log(Sales) ~ log(Assets) + log(Employees) + log(Cash.Flow) +
##     log(Profits) + sector + log(Assets):sector
##                    Df Sum of Sq     RSS      AIC F value   Pr(>F)
## <none>                           8.2636 -110.708
## log(Employees)      1   2.15058 10.4142  -96.285 13.0124 0.000714 ***
```

```
## log(Cash.Flow)      1   0.13596  8.3996 -111.550  0.8227 0.368750
## log(Profits)        1   0.05035  8.3139 -112.277  0.3047 0.583428
## log(Assets):sector  8   0.92700  9.1906 -119.159  0.7011 0.688913
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm3d <- update(lm3c, ~.-log(Assets):sector), test = "F")
```

```
## Single term deletions
##
## Model:
## log(Sales) ~ log(Assets) + log(Employees) + log(Cash.Flow) +
##     log(Profits) + sector
##                Df Sum of Sq     RSS     AIC F value    Pr(>F)
## <none>                      9.1906 -119.16
## log(Assets)     1    1.1712 10.3618 -112.64  7.3912 0.0086308 **
## log(Employees)  1    2.1742 11.3648 -106.08 13.7208 0.0004749 ***
## log(Cash.Flow)  1    0.1950  9.3856 -119.67  1.2308 0.2718254
## log(Profits)    1    0.0214  9.2120 -121.00  0.1349 0.7147754
## sector          8    3.7795 12.9701 -110.70  2.9815 0.0072313 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm3e <- update(lm3d, ~.-log(Profits)), test = "F")
```

```
## Single term deletions
##
## Model:
## log(Sales) ~ log(Assets) + log(Employees) + log(Cash.Flow) +
##     sector
##                Df Sum of Sq    RSS     AIC F value   Pr(>F)
## <none>                      9.212 -121.00
## log(Assets)     1    1.1655 10.377 -114.54  7.4646 0.008288 **
## log(Employees)  1    2.4092 11.621 -106.50 15.4300 0.000227 ***
## log(Cash.Flow)  1    0.3360  9.548 -120.45  2.1520 0.147691
## sector          8    4.1974 13.409 -110.34  3.3604 0.003084 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop1(lm3f <- update(lm3e, ~.-log(Cash.Flow)), test = "F")
```

```
## Single term deletions
##
## Model:
## log(Sales) ~ log(Assets) + log(Employees) + sector
##                Df Sum of Sq    RSS      AIC F value    Pr(>F)
## <none>                      9.548 -120.451
## log(Assets)     1    1.8946 11.443 -109.599 11.9057  0.001031 **
## log(Employees)  1    4.3814 13.929  -95.636 27.5328 2.126e-06 ***
## sector          8    6.7310 16.279  -98.569  5.2872 4.960e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Everything else is siginificant. Let's look at our final model `lm3f`.

```
Anova(lm3f)
```

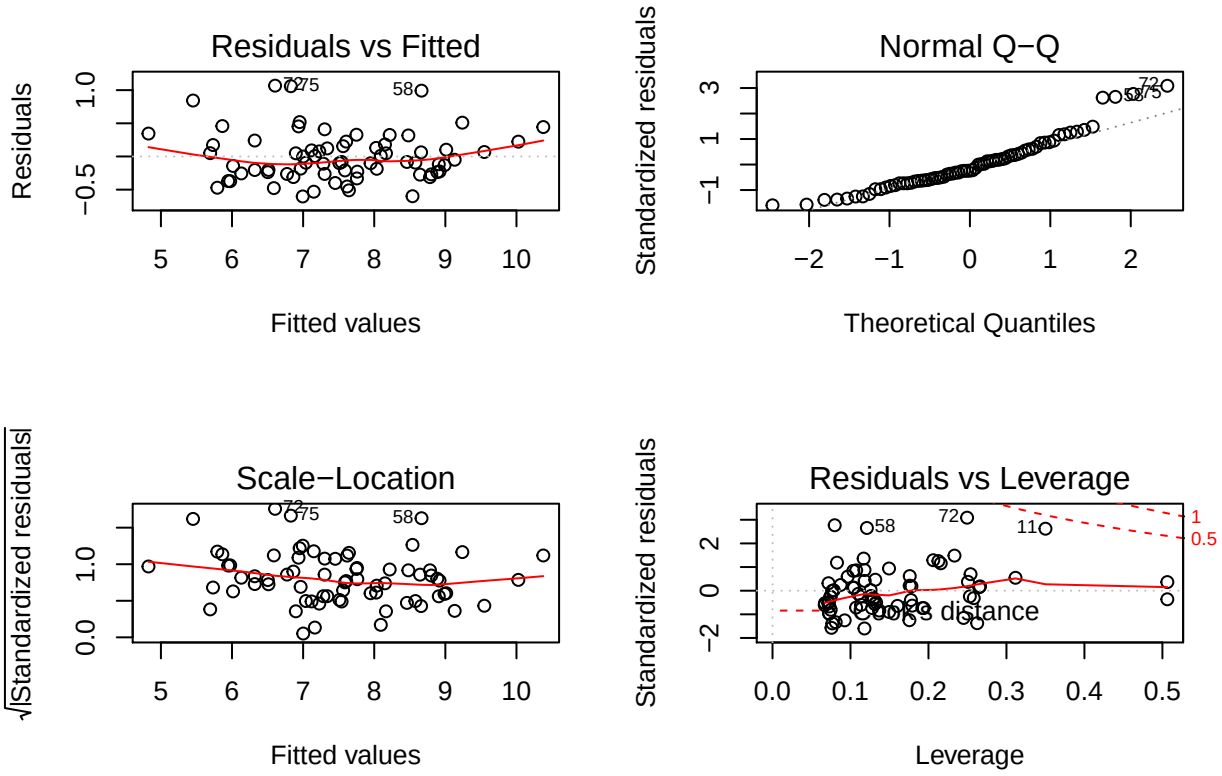
```
## Anova Table (Type II tests)
##
## Response: log(Sales)
##                Sum Sq Df F value    Pr(>F)
## log(Assets)    1.8946  1 11.9057  0.001031 **
## log(Employees) 4.3814  1 27.5328 2.126e-06 ***
## sector         6.7310  8  5.2872 4.960e-05 ***
## Residuals      9.5480 60
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm3f)
```

```
##
## Call:
## lm(formula = log(Sales) ~ log(Assets) + log(Employees) + sector,
##     data = forbes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60248 -0.24602 -0.08277  0.17644  1.06718
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.76251    0.61976   6.071 9.35e-08 ***
## log(Assets)           0.31052    0.09000   3.450  0.00103 **
## log(Employees)        0.49786    0.09488   5.247 2.13e-06 ***
## sectorEnergy          0.29063    0.32823   0.885  0.37945
## sectorFinance        -0.62102    0.34398  -1.805  0.07603 .
## sectorHiTech          0.25597    0.33077   0.774  0.44205
## sectorManufacturing   0.31084    0.32633   0.953  0.34466
## sectorMedical        -0.38089    0.36805  -1.035  0.30487
## sectorOther           0.10317    0.34335   0.300  0.76484
## sectorRetail          0.34654    0.35679   0.971  0.33531
## sectorTransportation  0.23088    0.35696   0.647  0.52024
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3989 on 60 degrees of freedom
## Multiple R-squared:  0.9056, Adjusted R-squared:  0.8899
## F-statistic: 57.59 on 10 and 60 DF,  p-value: < 2.2e-16
```

90.6% Variance in log(Sales) is explained by model `lm3f`. Let's look at our model diagnostics.

```
par(mfrow=c(2,2))
plot(lm3f)
```



One could argue that the outliers 72 and 11 should be removed due to their high Cook's distances. Proceed with re-modelling thereafter and check model assumptions again...