Appendix: Regression Models Course Project

Tamer Koksal

1. Variable Descriptions:

Variable Name	Variable Description
mpg	Miles per gallon (US)
cyl	Number of cylinders
disp	Displacement / Engine size (in cubic inches)
hp	Horsepower
drat	Rear axle ratio
wt	Weight
qsec	Qurarter mile time
vs	Engine shape (V-engine (0) or Straight-engine(1))
am	Transmission type (automatic (0) or manual (1))
gear	Number of forward gears
crab	Number of carburetor barrels

2. Exploratory Analysis

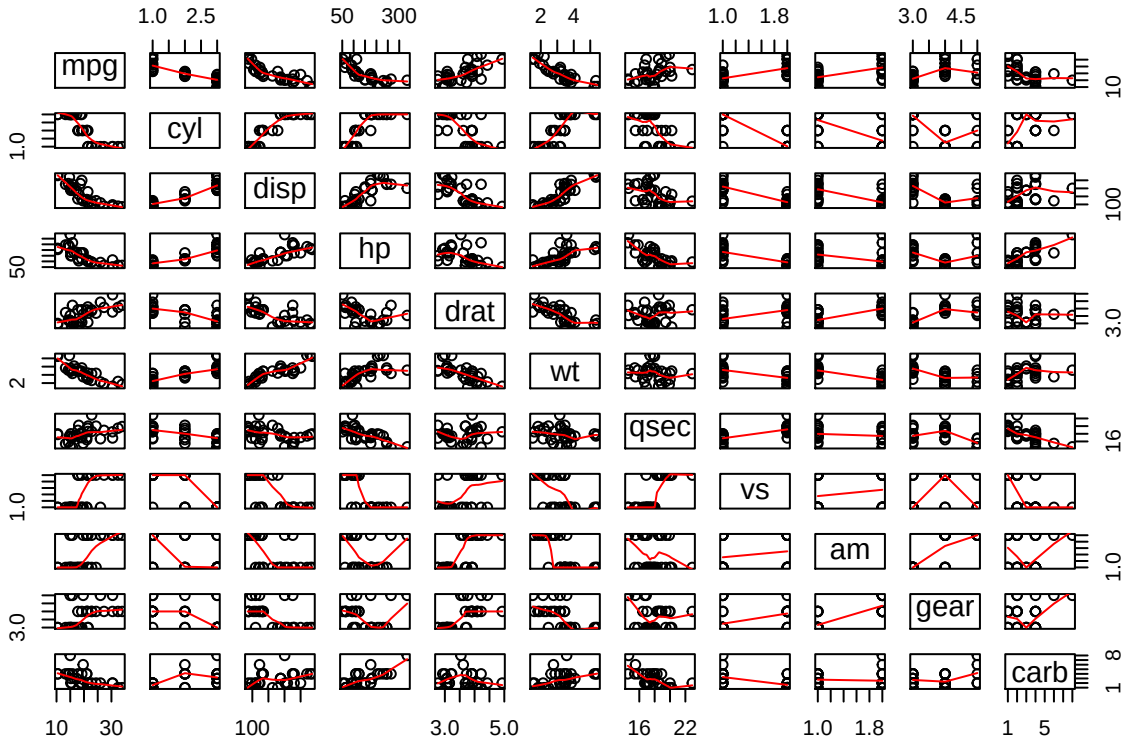
```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
library(dplyr)
mcars <- mtcars
mcars <- mutate(mcars, cyl = factor(cyl, ordered = TRUE),
                vs = factor(vs, labels = c("V", "S")),
                am = factor(am, labels = c("Auto", "Manual")))
```

3. Pairwise comparison (scatter plots) of variables

```
pairs(mcars, panel = panel.smooth)
```



4. Regression of `mpg` on all of the variables

```
fit <- lm(mpg ~ ., data=mcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4734 -1.3794 -0.0655  1.0510  4.3906
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.81222   15.42303   1.155   0.2617
## cyl.L        1.15784    3.05169   0.379   0.7084
## cyl.Q        2.02412    1.38827   1.458   0.1604
## disp         0.01391    0.01740   0.799   0.4334
## hp          -0.04613    0.02712  -1.701   0.1045
## drat         0.02635    1.67649   0.016   0.9876
## wt          -3.80625    1.84664  -2.061   0.0525 .
## qsec         0.64696    0.72195   0.896   0.3808
```

```
## vsS          1.74739    2.27267   0.769   0.4510
## amManual     2.61727    2.00475   1.306   0.2065
## gear         0.76403    1.45668   0.525   0.6057
## carb         0.50935    0.94244   0.540   0.5948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.582 on 20 degrees of freedom
## Multiple R-squared:  0.8816, Adjusted R-squared:  0.8165
## F-statistic: 13.54 on 11 and 20 DF,  p-value: 5.722e-07
```

5. Model Selection

```
# Stepwise Backward Regression
step(fit)
```

```
## Start:  AIC=69.67
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - drat  1    0.0016 133.32 67.666
## - gear  1    1.8339 135.16 68.103
## - carb  1    1.9472 135.27 68.129
## - vs    1    3.9408 137.26 68.597
## - disp  1    4.2608 137.58 68.672
## - cyl   2   14.2508 147.57 68.915
## - qsec  1    5.3532 138.68 68.925
## <none>              133.32 69.665
## - am    1   11.3619 144.69 70.282
## - hp    1   19.2853 152.61 71.989
## - wt    1   28.3207 161.64 73.829
##
## Step:  AIC=67.67
## mpg ~ cyl + disp + hp + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - gear  1    1.8575 135.18 66.108
## - carb  1    2.2247 135.55 66.195
## - vs    1    4.0705 137.40 66.628
## - disp  1    4.3401 137.66 66.691
## - qsec  1    5.3765 138.70 66.931
## - cyl   2   16.2318 149.56 67.342
## <none>              133.32 67.666
## - am    1   11.7008 145.03 68.358
## - hp    1   20.7664 154.09 70.298
## - wt    1   29.2443 162.57 72.012
##
## Step:  AIC=66.11
## mpg ~ cyl + disp + hp + wt + qsec + vs + am + carb
##
##        Df Sum of Sq    RSS    AIC
## - vs    1     4.250 139.43 65.099
```

```
## - carb  1     4.808 139.99 65.227
## - disp  1     4.895 140.08 65.247
## - qsec  1     4.918 140.10 65.252
## - cyl   2    17.095 152.28 65.919
## <none>              135.18 66.108
## - am    1    16.829 152.01 67.863
## - hp    1    19.891 155.07 68.501
## - wt    1    33.543 168.73 71.201
##
## Step:  AIC=65.1
## mpg ~ cyl + disp + hp + wt + qsec + am + carb
##
##        Df Sum of Sq    RSS    AIC
## - carb  1     2.898 142.33 63.757
## - disp  1     4.214 143.65 64.052
## - cyl   2    13.993 153.43 64.160
## <none>              139.43 65.099
## - qsec  1    10.717 150.15 65.469
## - am    1    14.361 153.79 66.236
## - hp    1    15.649 155.08 66.503
## - wt    1    36.334 175.77 70.510
##
## Step:  AIC=63.76
## mpg ~ cyl + disp + hp + wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## - disp  1     1.651 143.98 62.126
## - cyl   2    11.107 153.44 62.162
## - qsec  1     8.078 150.41 63.524
## <none>              142.33 63.757
## - hp    1    15.403 157.73 65.046
## - am    1    17.424 159.75 65.453
## - wt    1    40.707 183.04 69.807
##
## Step:  AIC=62.13
## mpg ~ cyl + hp + wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## - cyl   2    16.085 160.07 61.515
## - qsec  1     7.044 151.03 61.655
## <none>              143.98 62.126
## - hp    1    15.443 159.42 63.387
## - am    1    16.566 160.55 63.611
## - wt    1    52.932 196.91 70.145
##
## Step:  AIC=61.52
## mpg ~ hp + wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## - hp    1     9.219 169.29 61.307
## <none>              160.07 61.515
## - qsec  1    20.225 180.29 63.323
## - am    1    25.993 186.06 64.331
## - wt    1    78.494 238.56 72.284
```

```
##
## Step:  AIC=61.31
## mpg ~ wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## <none>              169.29 61.307
## - am    1    26.178 195.46 63.908
## - qsec  1   109.034 278.32 75.217
## - wt    1   183.347 352.63 82.790
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mcars)
##
## Coefficients:
## (Intercept)           wt         qsec     amManual
##       9.618       -3.917        1.226        2.936
```

```
fitbackward <- lm(mpg ~ am + wt + qsec, data=mcars)
summary(fitbackward)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec, data = mcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## amManual      2.9358     1.4109   2.081 0.046716 *
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

```
# Check for interaction effect:
fitint <- update(fitbackward, mpg ~ am + wt + qsec + am*wt)
anova(fitbackward, fitint)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + wt + qsec
## Model 2: mpg ~ am + wt + qsec + am:wt
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     28 169.29
```

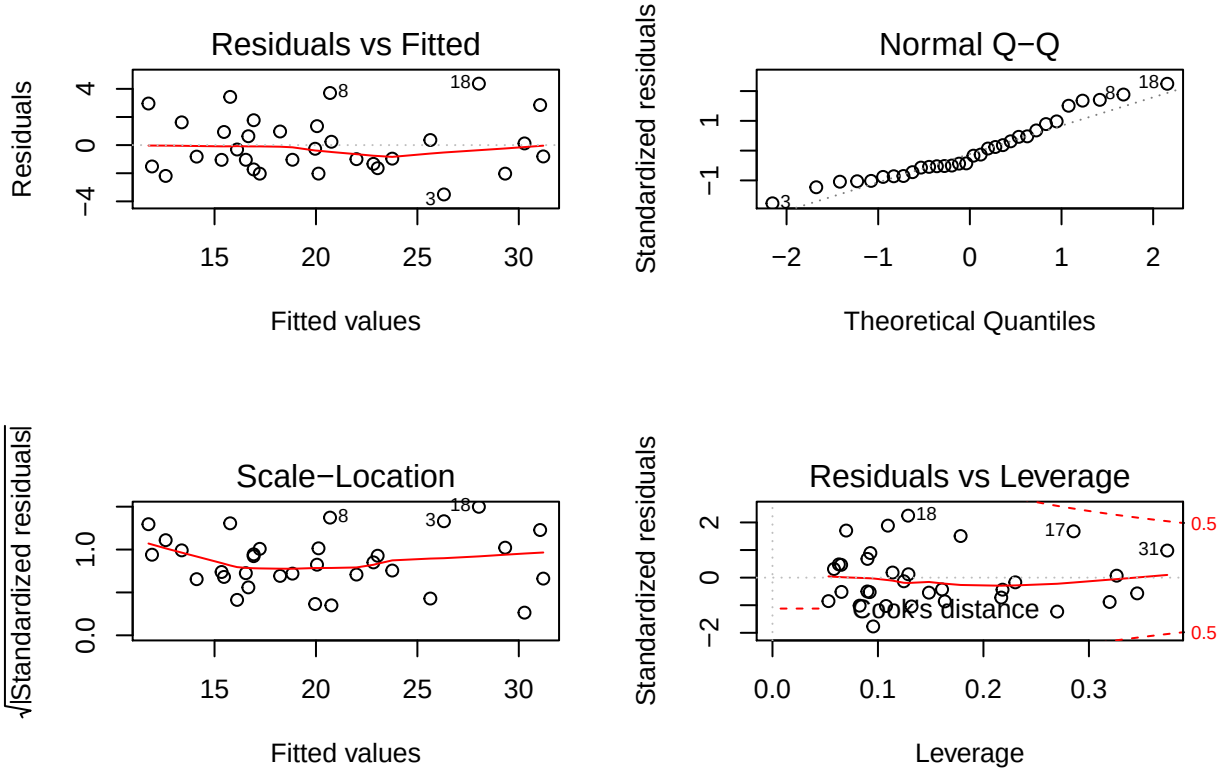
```
## 2     27 117.28  1     52.01 11.974 0.001809 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fitint)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec + am:wt, data = mcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5076 -1.3801 -0.5588  1.0630  4.3684
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.723      5.899   1.648 0.110893
## amManual      14.079      3.435   4.099 0.000341 ***
## wt            -2.937      0.666  -4.409 0.000149 ***
## qsec           1.017      0.252   4.035 0.000403 ***
## amManual:wt   -4.141      1.197  -3.460 0.001809 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.084 on 27 degrees of freedom
## Multiple R-squared:  0.8959, Adjusted R-squared:  0.8804
## F-statistic: 58.06 on 4 and 27 DF,  p-value: 7.168e-13
```

6. Regression Dignostics

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


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage