# The influence of transmission types on miles per gallon

**Executive summary:** On average, cars with manual transmission run 7.25 additional miles per gallon (MPG). When considering other influences on MPG like weight and horsepower some of the additional miles can be attributed to the fact that cars with manual transmission are generally likely to save fuel due to their other features. After controlling for those, cars with manual transmission still run 2.05 miles per gallon more, on average and holding all other characteristics fixed. The effect is not statistically different from zero, though.

## Exploring the data

The data set consists of 32 observations of 11 variables, namely miles per gallon, the number of cylinders, the displacement, the horsepower, the rear axle ratio, the weight, the quarter mile time, V/S (i.e. engine type), the transmission type, the number of forward gears and the number of carburetors. The cars in the sample were all built in 1973 or 1974. These cars reach on average `20.0906` miles per gallon.

Regarding the difference between cars with automatic and manual transmission calculating the mean MPG shows that cars with manual transmission run on average 7.25 miles per gallon more. Furthermore the two groups do not overlap very much (see Appendix A for a boxplot) and the coefficient on the transmission type (am) is highly significant in a simple regression model (not shown here). This observation serves as a first impression concerning the influence of the transmission type. To make a more reliable statement about the size of the effect a multivariate regression model is estimated in the next chapter.

## Controlling for additional influences on MPG using a multivariate regression model

As mentioned before, the data set includes various other variables which can be assumed to influence MPG. It follows that due to omitting relevant variables the estimate of a 7.25 MPG difference between manual and automatic transmission is biased. Actually, all variables in the data set have considerable correlations with MPG:

```
##     mpg     cyl    disp      hp    drat      wt    qsec      vs      am
##  1.0000 -0.8522 -0.8476 -0.7762  0.6812 -0.8677  0.4187  0.6640  0.5998
##    gear    carb
##  0.4803 -0.5509
```

Accordingly, a model using all other variables except for qsec as predictors for MPG was estimated. The 1/4 mile time was excluded as it can be expected to be determined by other variables like horsepower and does not directly influence MPG. This is confirmed by high correlations among these variables.

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + hp + drat + wt + vs + am + gear +
##     carb, data = mtcars)
##
## Coefficients:
## (Intercept)          cyl         disp           hp         drat
##    28.51488     -0.43381      0.00761     -0.02413      0.72034
##          wt           vs           am         gear         carb
##    -2.63620      1.18116      1.88731      0.51961     -0.44954
```

This model achieves a very high R-squared of 86%, however all coefficients are insignificant. The effect of am on mpg is estimated at 1.89 miles per gallon, i.e. a car with manual transmission runs additional 1.89 miles per gallon on average, holding all other variables fixed.

The diagnostic plots show an obsevation with high leverage, see Appendix C. Furthermore, high correlations among the explanatory variables could be seen in a pairs plot (Appendix B). Therefore the variance inflation factors (VIF) are considered in order to check whether these correlations inflate the standard errors of the variables and lead to high p-values.

```
## $VIF.Eigenvalue
##          VIF eigenvalue
## cyl  14.215    5.53306
## disp 19.858    2.09860
## hp    9.716    0.48656
## drat  3.370    0.26827
## wt   11.266    0.21862
## vs    4.304    0.17857
## am    4.300    0.13267
## gear  5.322    0.05743
## carb  7.338    0.02623
##
## $Stat
##                       Stat              Comments
## Condition Number    210.94    Collinear if CN > 15
## Sum(1/eigenvalue)   79.69 Collinear if Sum >  45
```

Very high VIFs are found for the number of cylinders (cyl), the displacement (disp), and the weight (wt). These variables are all highly correlated with each other, see Appendix B. Of these variables, only weight is kept in the model. This model was estimated and the VIFs checked again.

```
##
## Call:
## lm(formula = mpg ~ hp + drat + wt + vs + am + gear + carb, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.995 -1.520 -0.405  1.457  5.152
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.4422     7.5740    3.23   0.0036 **
## hp           -0.0215     0.0138   -1.56   0.1328
## drat          0.9308     1.5213    0.61   0.5464
## wt           -2.0766     1.0350   -2.01   0.0562 .
## vs            1.3178     1.4561    0.91   0.3744
## am            2.0460     1.8775    1.09   0.2866
## gear          0.7301     1.3536    0.54   0.5946
## carb         -0.7000     0.5854   -1.20   0.2434
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.57 on 24 degrees of freedom
## Multiple R-squared:  0.859,  Adjusted R-squared:  0.818
## F-statistic:   21 on 7 and 24 DF,  p-value: 8.87e-09
```

```
## $VIF.Eigenvalue
##        VIF eigenvalue
## hp   4.207    3.7710
## drat 3.109    2.0656
## wt   4.819    0.4621
## vs   2.531    0.2594
## am   4.125    0.1731
## gear 4.687    0.1518
## carb 4.201    0.1170
##
## $Stat
##                    Stat            Comments
## Condition Number  32.23   Collinear if CN > 15
## Sum(1/eigenvalue) 27.68 Collinear if Sum >  35
```
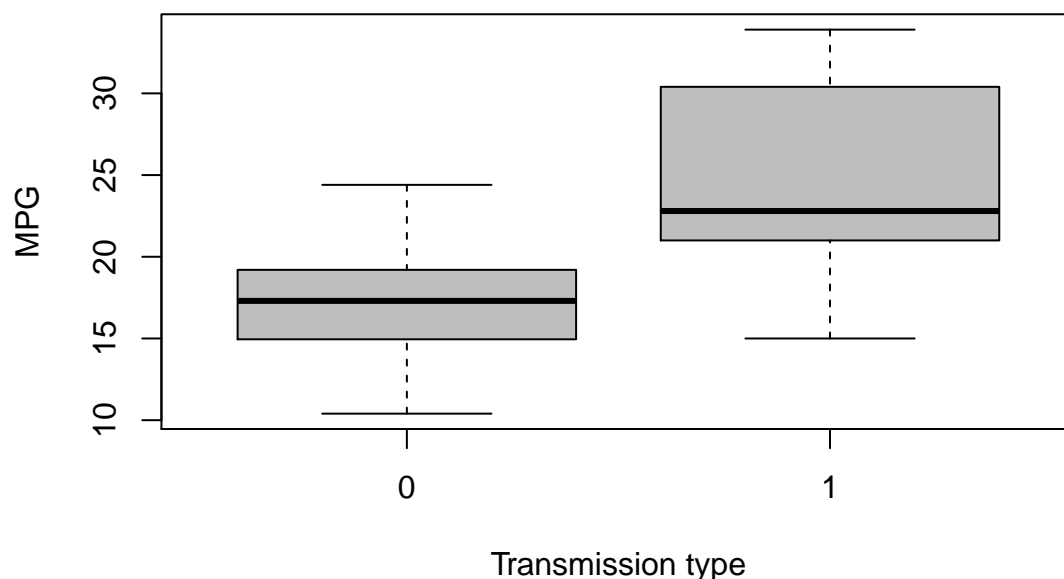
This time the intercept and the coefficient on weight are significantly different from zero at about 95% confidence (please refer to the table for exact p-values). The coefficient on am changes to 2.05 and the R-squared is virtually as high as before. All VIFs are quite low and the diagnostic plots show no obvious problems with this specification, see Appendix D. It could perhaps be suspected that there is heteroscedasticity as the variance of the residuals seems to appreciate with larger fitted values. A white test of constant variance fails to reject this null hypothesis, though (not shown here).

Thus, accepting this model as the best of the considered models, the MPG advantage of manual transmission over automatic transmission is estimated at 2.05 miles per gallon, all other variables fixed. It should be noted though that the coefficient is not significantly different from zero, so after controlling for the other influences it cannot be concluded with certainty that manual transmission leads to a higher MPG.
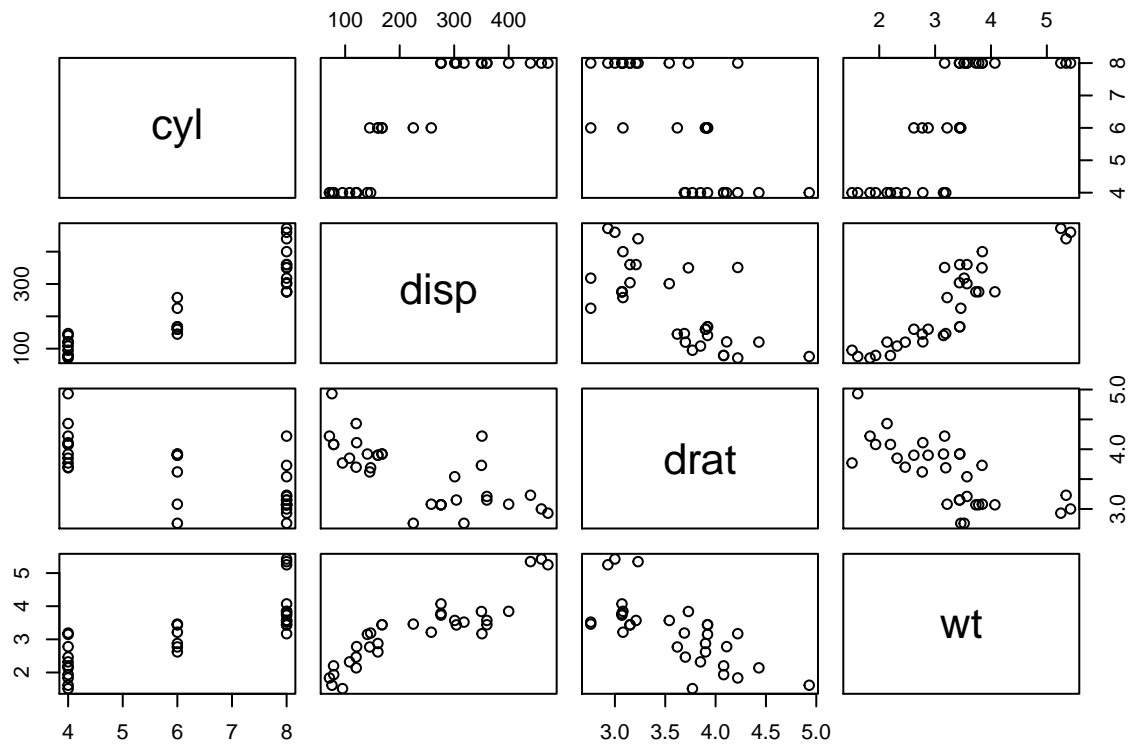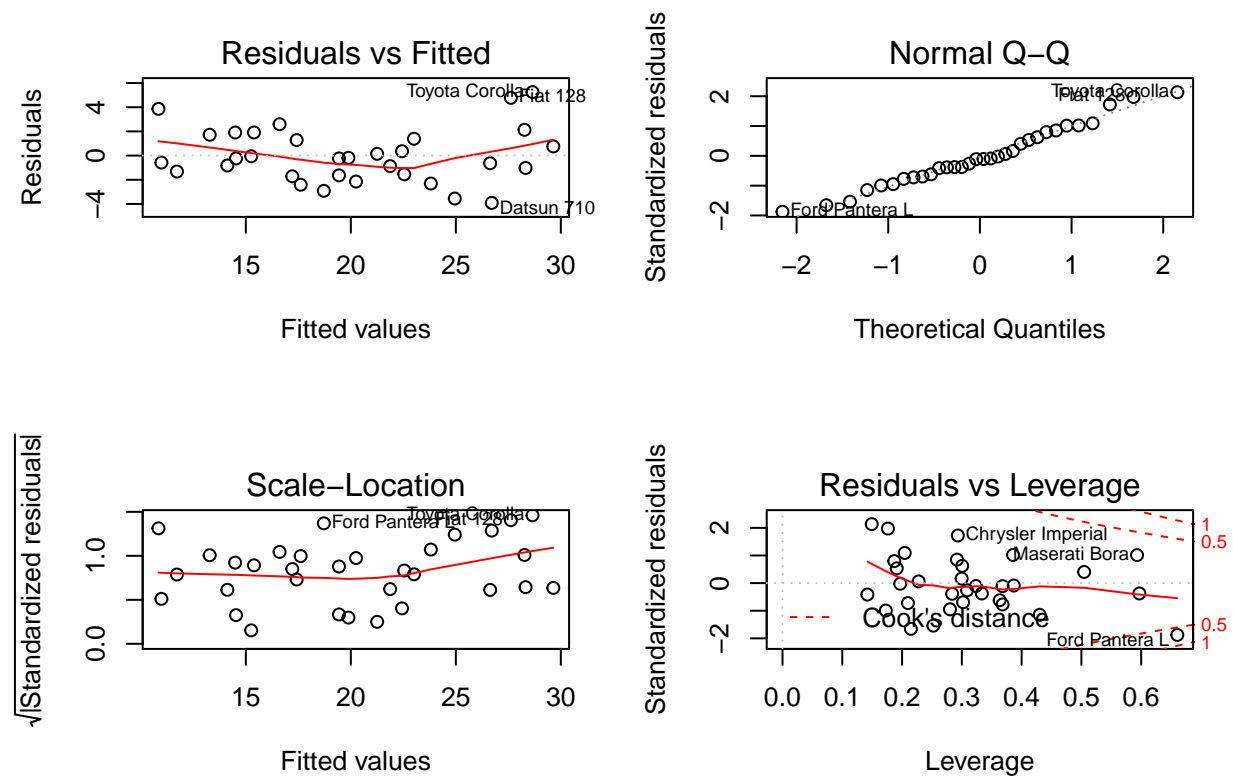
**Appendix**

**Appendix A**

## MPG for automatic (0) and manual (1) transmission

## Appendix B



## Appendix C

**Appendix D**



Residuals vs Fitted

Normal Q–Q

Scale–Location

Residuals vs Leverage

5