The Relationship Between Transmission Type and Fuel Economy

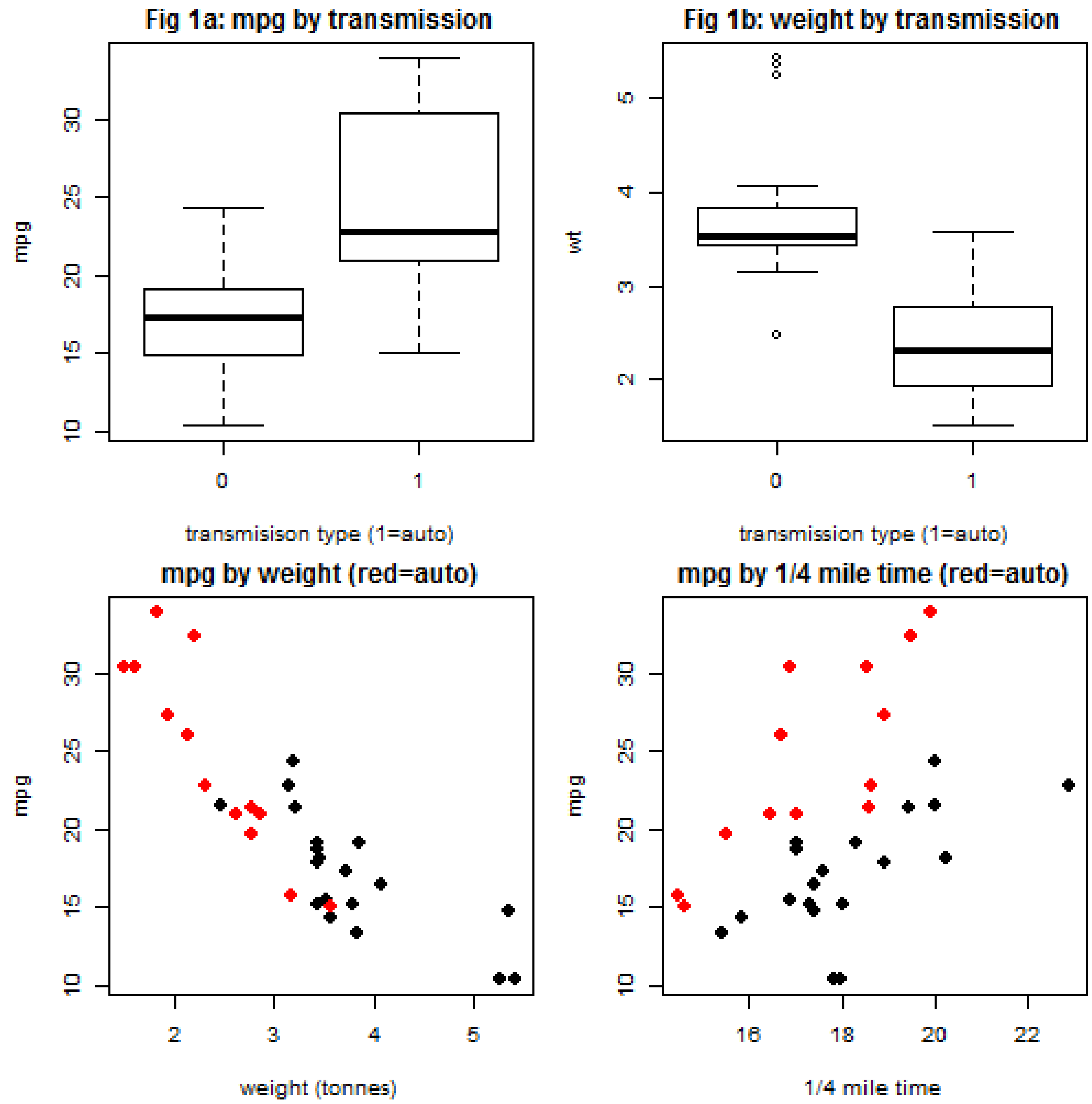
Executive Summary

This report estimates the impact of transmission type on average fuel economy, using the a dataset of 32 vehicles in the mtcars data package. The analysis identifies a statistically significant association between fuel economy and transmission type, supporting automatic vehicles as possessing greater fuel economy when controlling for the weight of the vehicle and its ¼ mile time.

Exploratory Data Analysis

There is relatively strong correlation between mpg and all variables including transmission type (rho=.599). Manual transmission vehicles in the mtcars dataset had worse average fuel economy than automatic vehicles, with mean mpg of 17.15, and 24.39 respectively. Manual vehicles, however, were also heavier on average, with a mean weight of 3769 pounds, compared to 2411 pounds for automatic vehicles (see figure 1).

```
# load the data
library(datasets)
DF <- mtcars
# plot mpg by weight and by transmission type.
par(mfrow = c(2, 2), mar = c(4, 4, 2, 1))
boxplot(mpg ~ am, data = DF, ylab = "mpg", xlab = "transmisison type (1=auto)",
    main = "Fig 1a: mpg by transmission")
boxplot(wt ~ am, data = DF, ylab = "wt", xlab = "transmission type (1=auto)",
    main = "Fig 1b: weight by transmission")
plot(mpg ~ wt, pch = 20, col = 1 * am + 1, cex = 2, xlab = "weight (tonnes)",
    ylab = "mpg", main = "mpg by weight (red=auto)", data = DF)
plot(mpg ~ qsec, pch = 20, col = 1 * am + 1, cex = 2, xlab = "1/4 mile time",
    ylab = "mpg", main = "mpg by 1/4 mile time (red=auto)", data = DF)
```



Limitations and Possible Improvements

Correlations between the variables indicate high potential for multicollinearity in the regression model. In particular, the number of cylinders, displacement, horsepower and weight are highly correlated. Ex-ante, each of these variables could potentially have a mechanistic effect on mpg. A more thorough research approach exploring the mechanical process through which these variables interact with mpg was beyond the scope of this exercise, but would likely yield a more robust model grounded in real-world vehicle mechanics. Instead, a stepwise model selection method was used, employing backwards elimination based on the Bonferroni criteria to eliminate unecessary regressors. Given the high correlations between the potential regressors, the potential confounding impact of high multicollinearity of the regressors should be noted. A paired sample including automatic and manual versions of the **same** model would represent a much more effective dataset for exmaination and would reduce problems associated with unmeasured variables (e.g. the different quality and model of various mpg-influencing parts such as tires, etc).

Model Selection

As a starting point, an explanatory model with all variables in the dataset was run and examined.

```
summary(lm(mpg ~ ., DF))
```

```
##
## Call:
## lm(formula = mpg ~ ., data = DF)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3.45  -1.60  -0.12   1.22   4.63
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.3034    18.7179    0.66    0.518
## cyl          -0.1114     1.0450   -0.11    0.916
## disp          0.0133     0.0179    0.75    0.463
## hp           -0.0215     0.0218   -0.99    0.335
## drat          0.7871     1.6354    0.48    0.635
## wt           -3.7153     1.8944   -1.96    0.063 .
## qsec          0.8210     0.7308    1.12    0.274
## vs            0.3178     2.1045    0.15    0.881
## am            2.5202     2.0567    1.23    0.234
## gear          0.6554     1.4933    0.44    0.665
## carb         -0.1994     0.8288   -0.24    0.812
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.807
## F-statistic: 13.9 on 10 and 21 DF,  p-value: 3.79e-07
```

A backwards elimination stepwise procedure was used to determine a simple linear model explaining vehicle mpg. (Code not shown for this process). To determine relevant regressors, an initial linear model was fit using all variables in the dataset (see output below). Following this, unecessary variables were excluded one-by-one starting with the least significant, by comparing the t-values with the Bonferonni cutoff value. The process yielded an explanatory model of mpg, using weight, ¼ mile time and transmission type as dependent variables. The ex-ante expectation is that ¼ mile captures a combination of other variables important for fuel economy, such as horsepower and engine displacement. A further exploration of a nested model incorporating interaction effects between weight and transmission type was examined, which supported the inclusion of the interaction term. The inclusion is based on the underlying assumption that the influence of weight and transmission type on mpg is not a simple additive relationship.

Based on this analysis, mpg was examined using the following model specification:

mpg = beta0 + (beta1 * wt) + (beta2 * qsec) + (beta3 * am) + (beta 4 * am * wt) + resid

where wt is vehicle weight in thousands of pounds, am is a dummy variable for transmission type (1=automatic) and qsec is the vehicle's ¼ mile time.

Outliers

In lieu of a real-world analysis of the vehicles to determine which should be excluded as true outliers, vehicles with a cook's distance exceeding the cutoff of 4/(n-k-2) were excluded from analysis. One outlier vehicle was removed, the Chrysler Imperial (cooks=.225). In a more thorough analysis, real world criteria for exclusion would have been implemented - for example, by excluding vehicles which represent less than x% of vehicle sales or which meet a formal classification as sports cars; however for the purpose of this assignment, cook's distance method was deemed sufficient.

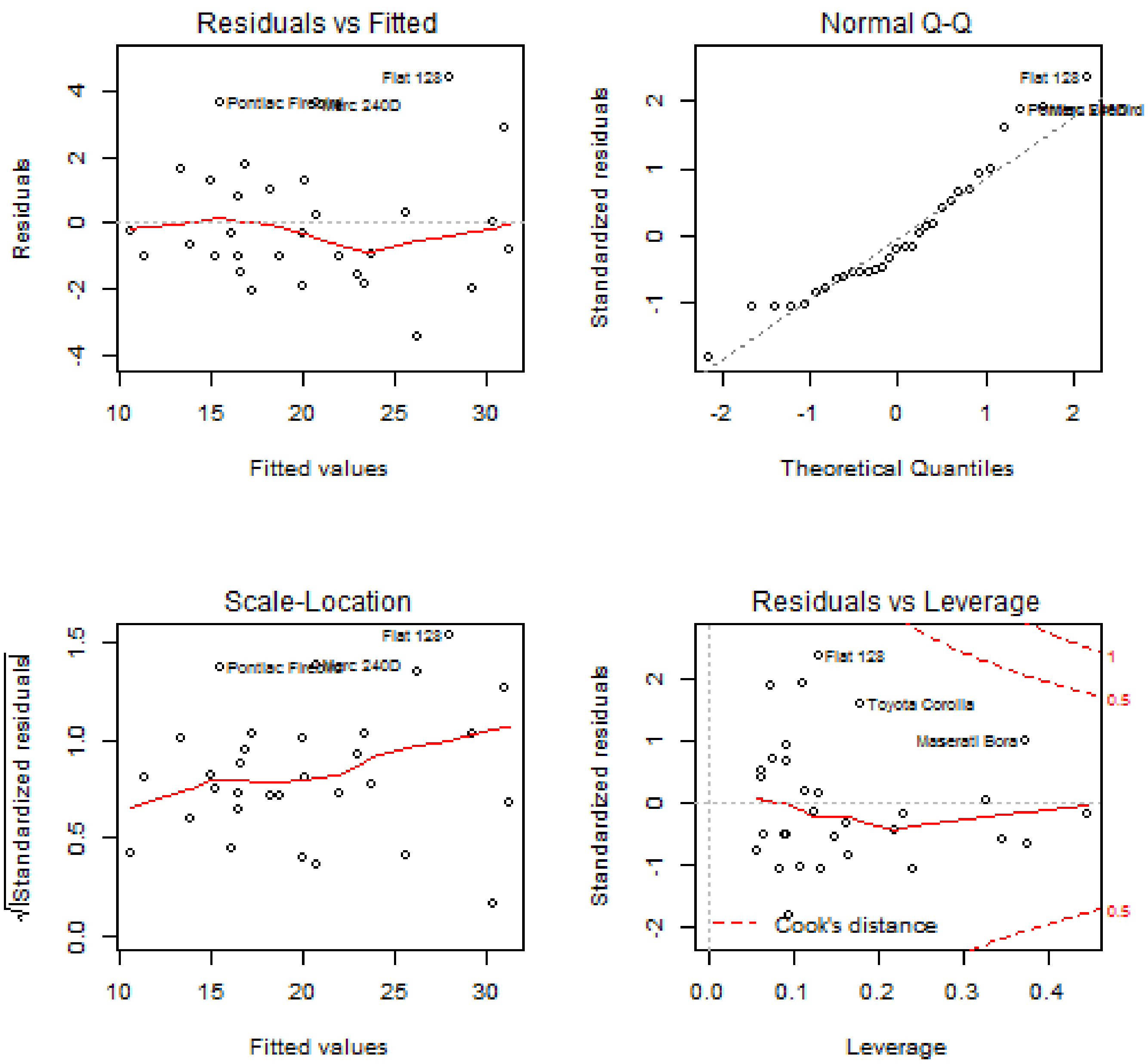
```
# Code below is for performing Cook's test for outliers
# cooks<-4/(nrow(DF)-4-2) fit<-lm(mpg~wt+qsec+am+am*wt,DF)
# cooks.distance(fit)[cooks.distance(fit)>cooks]

# Remove the outlier Chrysler Imperial (based on Cook's test)
DF <- DF[rownames(DF) != "Chrysler Imperial", ]
```

Model Results & Diagnostics

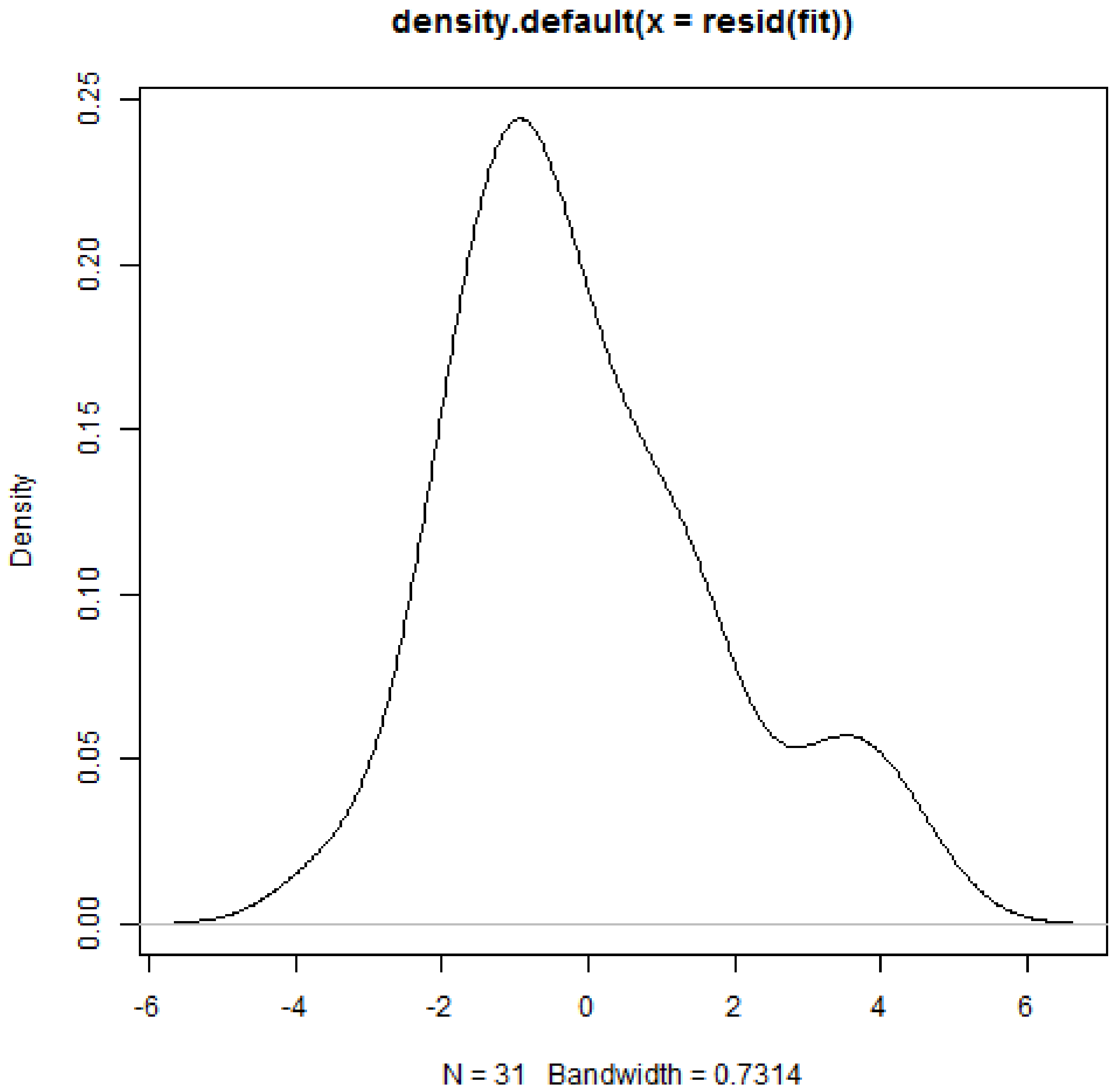
```r
# main model regression (outliers removed)
summary(fit <- lm(mpg ~ wt + qsec + am + am * wt, DF))
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am + am * wt, data = DF)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.472 -1.042 -0.357  1.122  4.425
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12.479      5.905    2.11  0.04434 *
## wt            -3.564      0.737   -4.84  5.1e-05 ***
## qsec           0.983      0.244    4.03  0.00043 ***
## am            12.064      3.510    3.44  0.00199 **
## wt:am         -3.580      1.198   -2.99  0.00607 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.01 on 26 degrees of freedom
## Multiple R-squared:  0.904,  Adjusted R-squared:  0.889
## F-statistic: 61.3 on 4 and 26 DF,  p-value: 7.34e-13
```

```r
# Final model is mpg = beta0 + beta1*wt + beta2*qsec +beta3*am + beta4*wt*am
# + error
par(mfrow = c(2, 2))
plot(fit)
```




```
par(mfrow = c(1, 1))
plot(density.default(x = resid(fit)))
```



Conclusions

A statistically significant (p-value = 0.002) relationship between mpg and transmission type was found, controlling for the effects of weight and ¼ mile time. There was a strong overall model fit, as demonstrated by the adjusted R-squared value of 0.89. The model residuals are also homoskedastic and approximately normal, though with a slight positive skew.

As this model includes an interaction term for transmission type, the interpretation of the effect of transmission type on mpg is dependent on the given value of weight. The marginal impact of transmission type, all else equal, is given by (beta3+beta4*wt), where wt is a given weight of interest. Given that beta3 is positive and beta4 is negative, there exists a weight at which the overall impact of transmission type “flips”. Based on this model specification, the results suggest a net fuel economy gain for automatic vehicles weighing less than 3,369 lbs (all else equal). At the mean vehicle weight (3148 lbs), the model suggests that the marginal impact of transmission type is 0.79 additional mpg for automatic vehicles. (12.064 - 3.580 * 3.148 = 0.79)