Regression Models Course Project - Motor Trend Data = 'mtcars' Miles Per Gallon Analysis

james c walmsley

12/1/2016

Executive Summary:

```
Interpretation of Results
```

Using linear regression model variations, including the step function, to gauge model fit, we identified model fstp <- lm(mpg ~ (wt + qsec + am), data= mtcars) that provides an 84.966 R^2 value indicating a reasonalby good fit and another model fit fnm6 <- lm(mpg ~ I(cyl + disp + hp + drat + wt + factor(am), data = mtcars) devloped using the multivariate nested approach followed by and anova table test to check for multicollinearity which also produced a reasonably good fit with an R^2 value of 85.13%. It must also be noted that vehicle weight is highly correlated (-86.77%) with mpg ratings and transmission type is relatively highly correlated with vehicle wieght at (69%).

```
Questions of interest for Motor Trend Magazine:

Q1 "Is an automatic or manual transmission better for 'mpg'"
        or which type of transmission is associated with better mpg or gas mileage ratings?
A1. The mean "mpg" rating of all vehicle models including both transmission types
is 20.09 mpg with a 95% confidence interval of 17.92 mpg to 22.26 mpg.

Q2 "Quantify the mpg difference between automatic and manual transmissions"
What is the expected difference in mpg rating and how accurate
is this estimate based on the given data?

A2. The mean "mpg" of models with automatic transmisions is 17.15 mpg, with a 95%
confidence interval of between 14.85 mpg to 19.44 mpg and vehicles with manual
transmisions have a mean of 24.39 mpg for a difference of 7.24 mpg with a 95 %
confidence interval of between 18.49 mpg and 30.29 mpg
```

Technical Environment:

```
System - session Info; Set the Working Directory; Record the System & Session Info; Check for requi
```

Raw Data:

```
Clean up work space, import the data & check for missing values
Overview: Motor Trend 'mtcars' data set:
```

A data frame with 32 observations on 11 variables.

[, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders (4,6,8) [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (1000 lbs) [, 7] qsec 1/4 mile time [, 8] vs V/S (0 = vee-block, 1 = straight-block) [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears (3:5) [,11] carb Number of carburetors (1:4,6,8)

Processed Data:

```
Factor columns 2 & 8:11 (cyl,vs,am,gear,carb) into levels
```

Descriptive Statistics:

```
library(datasets);library(dplyr);data("mtcars")
head(mtcars,4);mean(mtcars$mpg);sd(mtcars$mpg)
```

```
            mpg cyl disp  hp drat    wt  qsec vs am gear carb
```

Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4 Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4 Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1 Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1 [1] 20.09062 [1] 6.026948

```
round(t.test(mtcars$mpg)$conf.int,3)
```

[1] 17.918 22.264 attr(,“conf.level”) [1] 0.95

Exploratory Analysis:

```
See Appendix A. Figures (pairs-plot, histogram, box-plot)
```

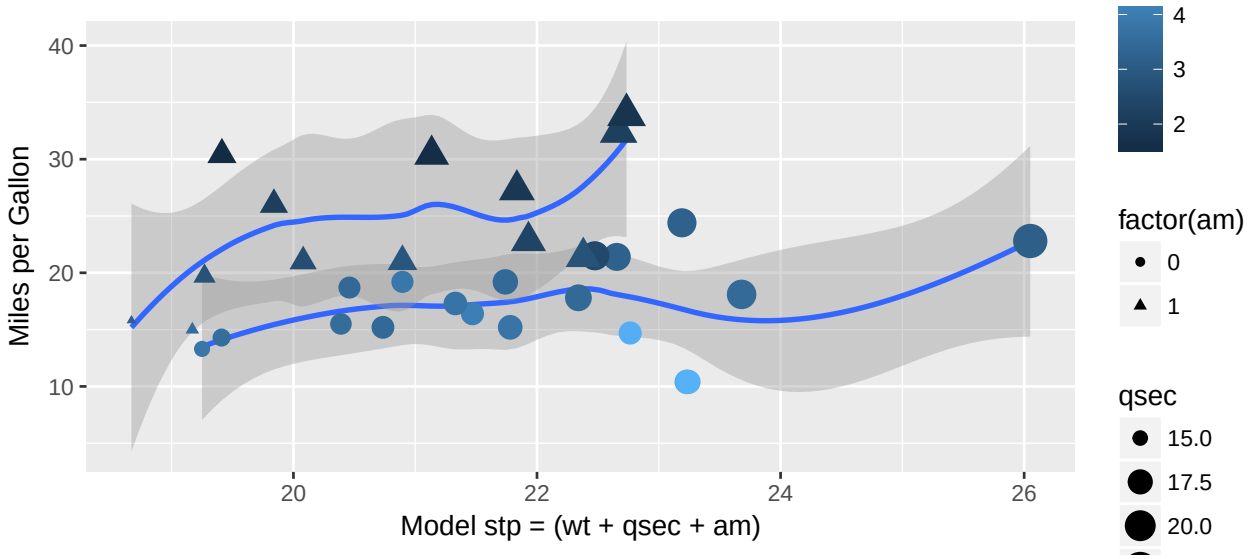
Statistical Modeling:

```
Multivarite Linear Model Finding Best Fit with Step function:
```

Multivarite Linear Model Finding Best Fit with Step function:

```
        Estimate Std. Error   t value     Pr(>|t|)
```

(Intercept) 9.617781 6.9595930 1.381946 1.779152e-01 wt -3.916504 0.7112016 -5.506882 6.952711e-06 qsec 1.225886 0.2886696 4.246676 2.161737e-04 am 2.935837 1.4109045 2.080819 4.671551e-02 [1] 0.8496636



Diagnostics:

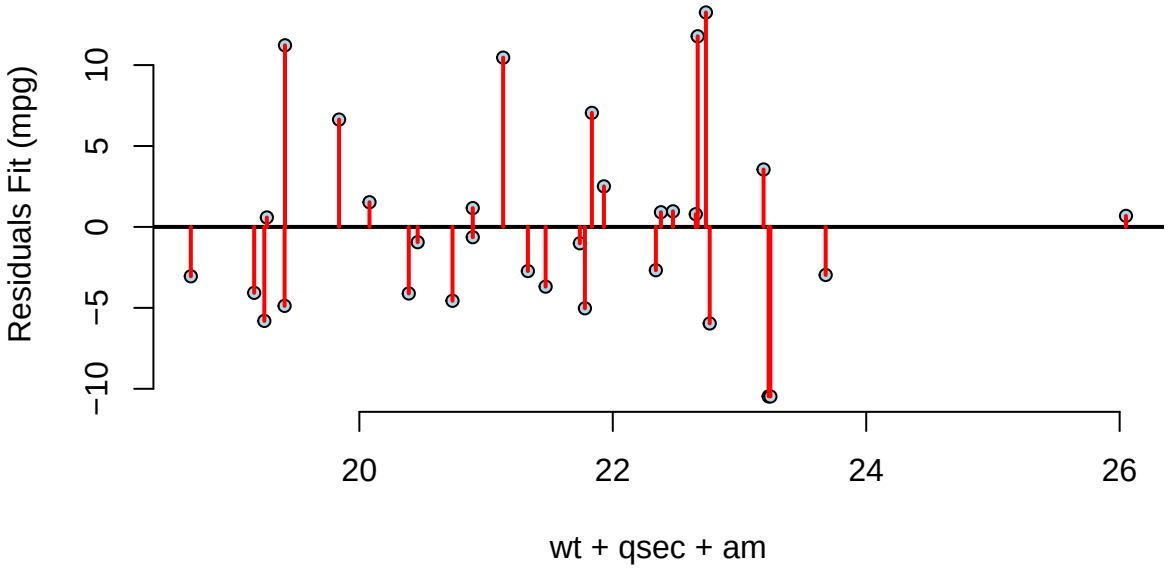
```
x <- (mtcars$wt + mtcars$qsec + as.numeric(mtcars$am))
y <- mtcars$mpg
n <- length(y)
fw <- lm(y ~ x, data = mtcars)
coef(summary(fw))
```

```
##               Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) 10.5846719 14.3078992 0.7397782 0.4651847
## x            0.4427088  0.6644599 0.6662686 0.5103310
```

```
e <- resid(fw);sum(e)
```

```
## [1] 5.995204e-15
```

```
#par(mfrow = c(1,1), mar = c(4,4,4,3)) # set margin
plot(x,e,
     xlab = "wt + qsec + am",
     ylab = "Residuals Fit (mpg)",
     bg = "lightblue",
     col = "black", cex = .85, pch = 21, frame = FALSE)
abline(h = 0, lwd = 2)
for(i in 1:n)
        lines(c(x[i], x[i]), c(e[i], 0), col = "red", lwd = 2)
```




```
sum(e)
```

```
## [1] 5.995204e-15
```

Hypothesis Test:

H0 = mean(mpg[am==automatic]) = mean(mpg[am==manual]) (REJECT)
Ha = mean(automatic transmission)mpg != mean(manual transmission)mpg (ACCEPT)

```
t.test(mpg ~ factor(am), paired = FALSE, var.equal=FALSE, data = mtcars)
```

```
Welch Two Sample t-test
```

data: mpg by factor(am) t = -3.7671, df = 18.332, p-value = 0.001374 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -11.280194 -3.209684 sample estimates: mean in group 0 mean in group 1 17.14737 24.39231 # Inference & Prediction:

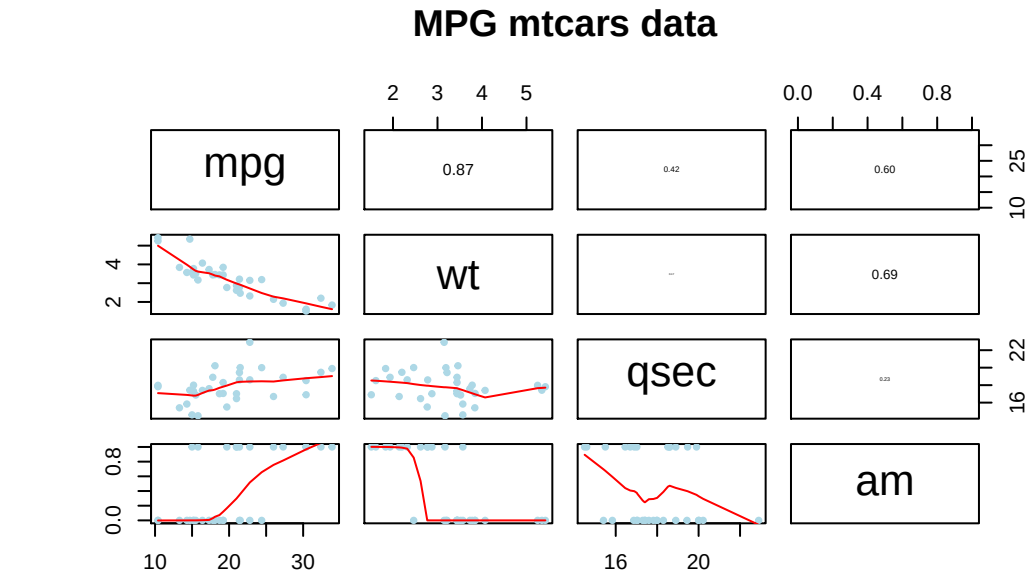
```
data("mtcars")
levels(mtcars$wt) <- 2
fw <- lm(mpg ~ wt, data = mtcars)
new.weights <- c(1.750, 2.125, 2.750, 3.475, 4.125, 4.700)
predict(fw, newdata = data.frame(wt = new.weights))
```

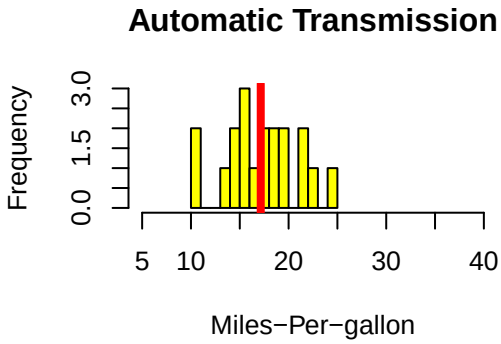
Interpretation of Results:

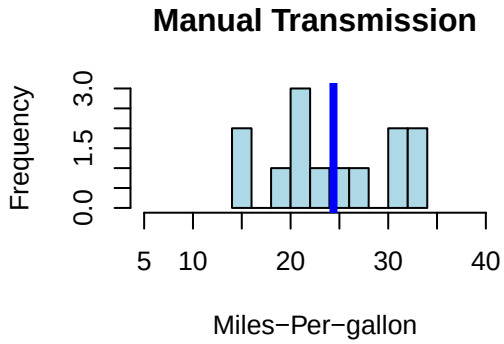
See Executive Summary

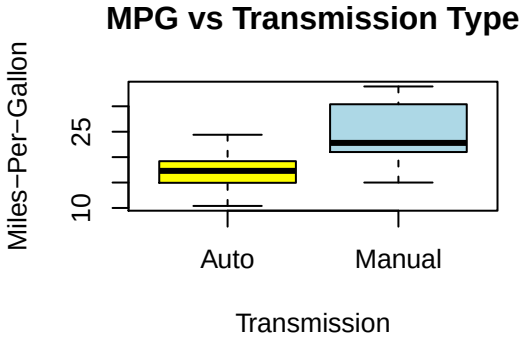
Appendix A: Figures: Exploratory Analysis

Appendix A: Figures: Exploratory Analysis









=== END ===