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

james c walmsley

12/1/2016

Executive Summary:
Using a braod range of linear regressoin model variations including the step function to gauge best model fit we idendified a model using (wt + qsec + am) that provides an 84.966 R^2 value inidicating a reasonalby goo fit which practically matches a slightly different fit that was
devloped using the manual nested approach followed by and anove 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%).

```
Problem Statement:
        Backround information, problem statement & questions of interest:
        Background situation:
                As a member of a team of data analysts for the Motor Trend Magazine we have been gi
                a data set called "mtcars" and asked to answer some questions of interest concernin
                differences between automatic and  manual transmissoin types in regards to associat
                mpg or miles per gallon ratings within the given data set.
        Assumptions:
                The given data set (a sample of a larger population) for this analysis consists of
                independent and identically distrubuted random varialbles for 32 subjects (vehicles
                11 observations or variables.

                Questions of interest for Motor Trend Magazine:

                Q1 "Is an automatic or manual transmission better for 'mpg'"
                        or which type of tramsmission is associated with better mph or gas mileage
                        A1. The mean "MPG" rating of all vehicle models including both transmission
                        is 20.091 mpg with a 95% confint of 17.917mpg to 22.263mpg.
                Q2 "Quantify the MPG difference between automatic and manual transmissions"
                        or assuming there is an associated difference in mpg ratings between manual
                        automatic type transmissions then: What is the expected difference and how
                        is this estimate based on the given data?
                        A2. The mean "MPG"  of models with automatic transmisions is 17.147 mpg, an
                        with manual transmisions 24.392 mpg for a difference of 7.24 mpg in favor o
                        manual transmission in the given data set.
```

Analysis Considerations:
Descriptive - any(is.na()), str() & summary(), Exploratory - pairsPlots(), histograms(), boxPlots(), barPlots() QQ_Plots & multiple plots Regression Models Analysis - OLS, SLR, BiVariate Regression, Multivariate Linear Regression, Heatmaps, HCL, PCA, SVD, Mean, T-Test, Z-Test, covariance, OLS, regression to mean (-1), simple linear regression, statistical linear regression, multivariable regression, logit & model selection, adjustments, residuals (predict fit, residual fit (-1)), hatvalues, variation, & dfbetas, R^2, diagnostics; ANOVA, GLMs & Binary GLMs, coeficients, correlation, confidence intervals, Cooks Distance, ChiSq-Test, VIF, binary, binomial, poisson, influence & leverage, Odds & OddRatio, Inferential & Predictive, Causal ~ NA, Mechanistic ~ NA Diagnostics See section on Diagnostics Final Model Selection strategy Beginning with the simmple linear regression using just one predictor (am), Use a bivariate model, Use a multivariate model, Use an

intercept adjusted multivariate model Use a multivariate model removing a suspected key regressor, Use the nested multivariate process Use the step(function) both directions process, Usee different combinations and leave out from
the results of the step(function), Choose the best fit which is understandable and easy to explian

Technical Environment:
Environment: System - session Info Set the Working Directory Record the System & Session Info Check which packages have beeb installed

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) so there values can be used as 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

```
mtcars0 <- mtcars[mtcars$am==0,];mtcars0;t.test(mtcars0$mpg)
```

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

Hornet 4 Drive 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3 1 Hornet Sportabout 18.7 8 360.0 175 3.15 3.440 17.02 0 0 3 2 Valiant 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3 1 Duster 360 14.3 8 360.0 245 3.21 3.570 15.84 0 0 3 4 Merc 240D 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2 Merc 230 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2 Merc 280 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4 Merc 280C 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4 Merc 450SE 16.4 8 275.8 180 3.07 4.070 17.40 0 0 3 3 Merc 450SL 17.3 8 275.8 180 3.07 3.730 17.60 0 0 3 3 Merc 450SLC 15.2 8 275.8 180 3.07 3.780 18.00 0 0 3 3 Cadillac Fleetwood 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3 4 Lincoln Continental 10.4 8 460.0 215 3.00 5.424 17.82 0 0 3 4 Chrysler Imperial 14.7 8 440.0 230 3.23 5.345 17.42 0 0 3 4 Toyota Corona 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1 Dodge Challenger 15.5 8 318.0 150 2.76 3.520 16.87 0 0 3 2 AMC Javelin 15.2 8 304.0 150 3.15 3.435 17.30 0 0 3 2 Camaro Z28 13.3 8 350.0 245 3.73 3.840 15.41 0 0 3 4 Pontiac Firebird 19.2 8 400.0 175 3.08 3.845 17.05 0 0 3 2

```
One Sample t-test
```

data: mtcars0$mpg t = 19.495, df = 18, p-value = 1.497e-13 alternative hypothesis: true mean is not equal to 0 95 percent confidence interval: 15.29946 18.99528 sample estimates: mean of x 17.14737

```
mtcars1 <- mtcars[mtcars$am==1,];mtcars1;t.test(mtcars1$mpg)
```

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

Mazda RX4 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4 4 Mazda RX4 Wag 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4 4 Datsun 710 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1 Fiat 128 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1 Honda Civic 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2 Toyota Corolla 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1 Fiat X1-9 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1 Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5 2 Lotus Europa 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2 Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5 4 Ferrari Dino 19.7 6 145.0 175 3.62 2.770 15.50 0 1 5 6 Maserati Bora 15.0 8 301.0 335 3.54 3.570 14.60 0 1 5 8 Volvo 142E 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2

```
One Sample t-test
```

data: mtcars1$mpg t = 14.262, df = 12, p-value = 6.909e-09 alternative hypothesis: true mean is not equal to 0 95 percent confidence interval: 20.66593 28.11869 sample estimates: mean of x 24.39231

```
c6 <- mtcars$mpg[mtcars$cyl==6];c6
```

[1] 21.0 21.0 21.4 18.1 19.2 17.8 19.7

```
c4 <- mtcars$mpg[mtcars$cyl==4];c4;t.test(c4,c6,var.equal = TRUE)
```

[1] 22.8 24.4 22.8 32.4 30.4 33.9 21.5 27.3 26.0 30.4 21.4

```
Two Sample t-test
```

data: c4 and c6 t = 3.8952, df = 16, p-value = 0.001287 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 3.154286 10.687272 sample estimates: mean of x mean of y 26.66364 19.74286

```
c6 <- mtcars$mpg[mtcars$cyl==6];c6
```

[1] 21.0 21.0 21.4 18.1 19.2 17.8 19.7

```
c8 <- mtcars$mpg[mtcars$cyl==8];c8;t.test(c6,c8,var.equal = TRUE)
```

[1] 18.7 14.3 16.4 17.3 15.2 10.4 10.4 14.7 15.5 15.2 13.3 19.2 15.8 15.0

```
Two Sample t-test
```

data: c6 and c8 t = 4.419, df = 19, p-value = 0.0002947 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 2.443809 6.841905 sample estimates: mean of x mean of y 19.74286 15.10000

Exploratory Analysis:
See Appendix A. Figures (pairs-plot, histogram, box-plot)

Statistical Modeling:
Multivarite Linear Model Finding Best Fit with Step function:

```r
library(stats);library(MASS)
fstp <- lm(mpg ~ ., data = mtcars)
stp <- step(fstp, trace = FALSE)
coef(summary(stp))
summary(stp)$r.squared
```

Preliminary findings: # Quesions of interest: & interpretation of results: A Revisit the Question - Considering all regressors: A. Is an automatic or manual transmission better for mpg The results of using multiple linear regression techniques
sugggest that manual transmissions are associated with better mpg ratings than automatic transmissions B. Quantify the MPG difference between automatic and manual transmissions On average manual transmissions provides 24.39 mpg which is 7.24 mpg more than the 17.15 mpg average of the automatic transmission models B Primary result A. Are any other regressors significantly correlated with mpg rating? a. model fnm6 = factor(am) + cyl + disp + hp + drat + wt this model has an R^2 value of 85.13% B. Further testing a. using the step function in both directions selects wt, qsec and am as good predictors with an 84.96% R^2 value indicating very good predictability using this set of regresssors

```
C Direction, Magnitude, Uncertainty
        A.
```

Multivarite Linear Model Finding Best Fit with Step function:

```r
library(stats);library(MASS);library(ggplot2)
fstp <- lm(mpg ~ ., data = mtcars)
stp <- step(fstp, trace = FALSE)
coef(summary(stp))
```

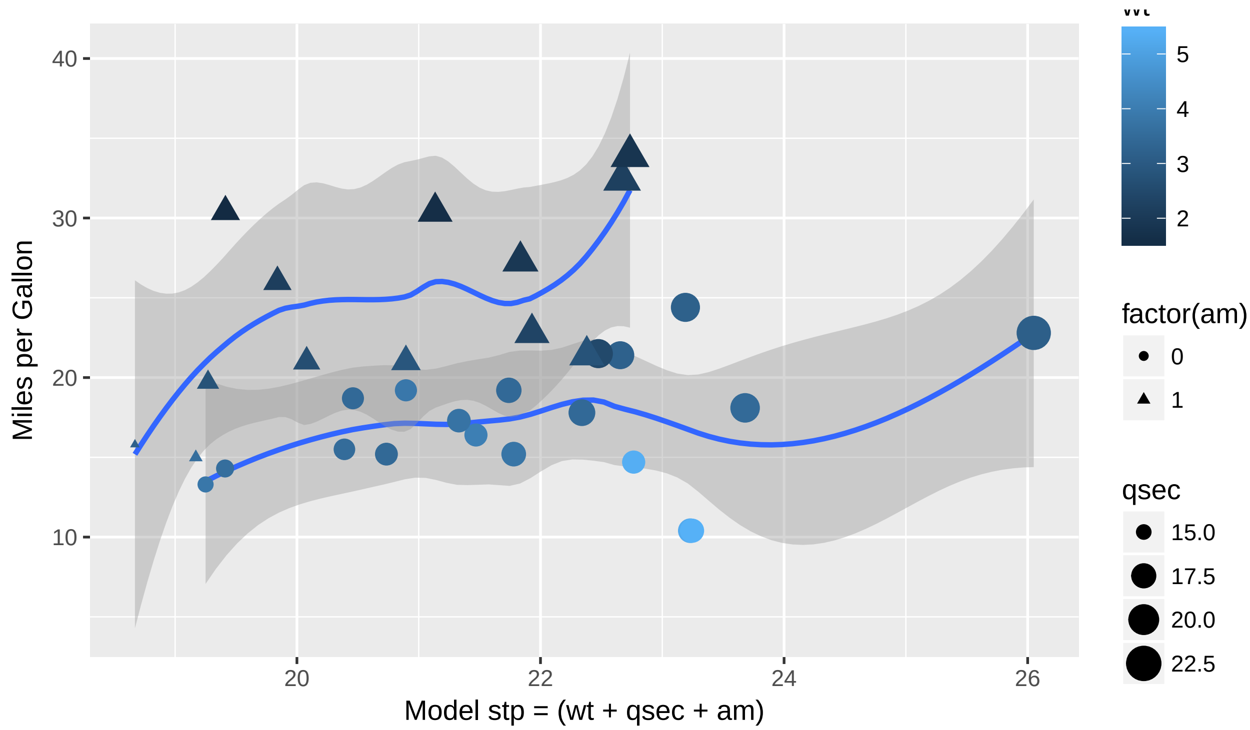
```
         Estimate Std. Error   t value     Pr(>|t|)
```

(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

```r
summary(stp)$r.squared
```

[1] 0.8496636

```r
par(mfrow = c(1, 1), mar = c(4,4,4,2))
g <- ggplot(mtcars, aes(x = (wt + qsec + am), y = mpg),)
g <- g + xlab("Model stp = (wt + qsec + am)")
g <- g + ylab("Miles per Gallon")
g <- g + geom_smooth(aes(method = "lm", shape = factor(am)))
g <- g + geom_point(aes(shape = factor(am), size=qsec, colour=wt))
g
```



D Context A. It should be noted that vehicle weight has a strong negative correlation to mpg ratings (-86.76%) and the weight of vehicle models with manual transmissions range from 1.513tons to 3.570 tons and the weight of vehicles with automatic transmissions range from 2.465 tons to 5.424 tons

```
E Implications - Congruence with existing knowledge?
        A. Sedan, Sports, Luxury
        Generally accepted expectations of mpg ratings are that sports and luxury models
        typically will have lower mpg ratings than sedans
```

Diagnostics:
Diagnostic tests were conducted on model results in accordance with the plan for analysis considerations. Besides several vehicle models exhibiting leverage on the model fit at the high end of the qsec, and weight scales results were generally as expected with faster and or heavier vehilces of both manual and automatic transmissoin types getting lower mpg ratings and slower lighter vehicle models of both transmission types exhibiting better or higher mpg ratings.

Hypothesis Test: ? H0 = mean(automatic transmission)mpg = mean(manual transmission)mpg Ha = mean(automatic transmission)mpg != mean(manual transmission)mpg

Inference & Prediction: ?

Interpretation of Results: #

Appendix A. Figures: need to review again possibly condense to one plot