Regression Models 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%.

Questions of interest for Motor Trend Magazine:

Q1 "Is an automatic or manual transmission better for 'mpg'"

A1. 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 with a 95 % confidence interval of between 18.49 mpg and 30.29 mpg

Q2 "Quantify the mpg difference between automatic and manual transmissions"

A2. The difference in "mpg" is on average 3.767 mpg and ranges between 3.21 mpg less and 11.28 mpg less for models having automatic transmissions than for those models having manual transmissions predicted at a 95% confidence level.

Technical Environment:

System - session Info; Set the Working Directory; Clean up work space, Record the System & Session Info & date; Check for required packages

```
setwd("~/Desktop/Coursera_R/7_Regression Models/RM_proj_MPG_MotorTrendData")
rm(list=ls())
sessionInfo()
```

R version 3.3.1 (2016-06-21) Platform: x86_64-apple-darwin13.4.0 (64-bit) Running under: OS X 10.11.6 (El Capitan)

locale: [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages: [1] stats graphics grDevices utils datasets methods base

loaded via a namespace (and not attached): [1] magrittr_1.5 formatR_1.4 tools_3.3.1 htmltools_0.3.5 [5] yaml_2.1.13 Rcpp_0.12.7 stringi_1.1.1 rmarkdown_1.0
[9] knitr_1.14 stringr_1.1.0 digest_0.6.10 evaluate_0.9

```
date()
```

[1] "Thu Dec 29 06:12:38 2016"

Raw Data:

```
    Load the data from the dataset package & check
    for missing values. Overview: Motor Trend 'mtcars' data set:
```

```
library(car);library(dplyr);library(MASS);data("mtcars");any(is.na(mtcars))
```

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
```

```
data("mtcars")
for(i in c(2,8:11))mtcars[,i] <- as.factor(mtcars[,i]);
str(mtcars)
```

Descriptive Statistics:

```
library(datasets);library(dplyr);data("mtcars")
head(mtcars,4);mean(mtcars$mpg);sd(mtcars$mpg)
round(t.test(mtcars$mpg ~ am, data = mtcars,
             var.equal =FALSE, paired = FALSE, conf.level = 0.95 )$conf.int,3)
```

Exploratory Analysis:

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

Statistical Modeling:

Multivarite Linear Model Finding Best Fit with Step function:

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

Multivarite Linear Model Finding Best Fit with Step function:

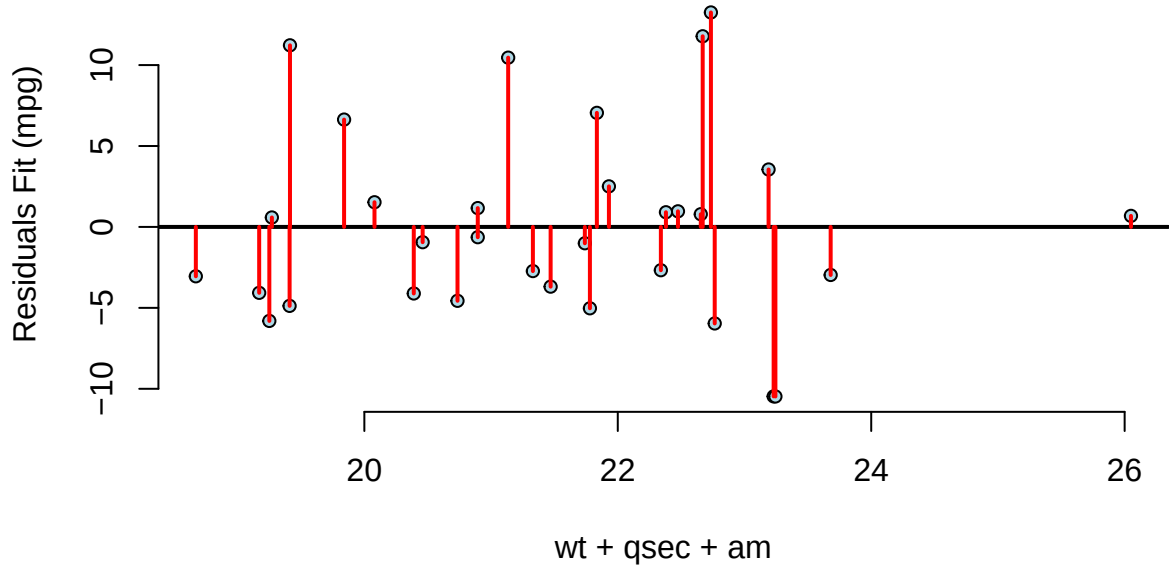
```
library(stats);library(MASS);library(ggplot2)
fstp <- lm(mpg ~ ., data = mtcars)
stp <- step(fstp, trace = FALSE)
coef(summary(stp))
summary(stp)$r.squared
x <- (mtcars$wt + mtcars$qsec + mtcars$am)
par(mfrow = c(1, 1), mar = c(4,4,4,2))
g <- ggplot(mtcars, aes(x = x, y = mpg),)
g <- g + xlab("Model 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
```

Diagnostics:

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

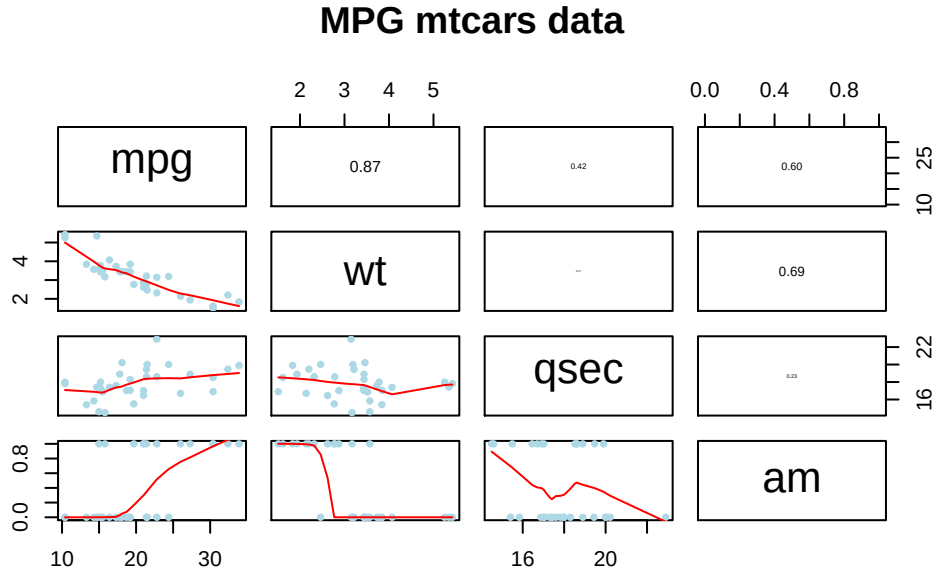
```
H0 = mean(mpg[am==automatic])  =   mean(mpg[am==manual]) (REJECT)
Ha = mean(automatic transmission)mpg  !=  mean(manual transmission)mpg (ACCEPT)
```

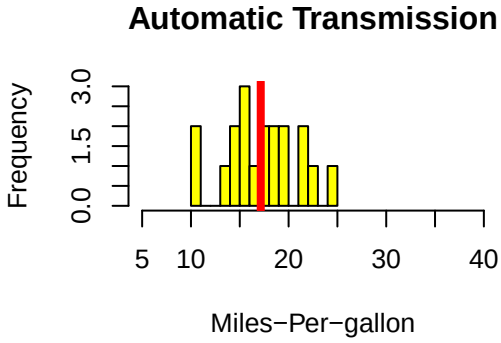
```
t.test(mpg ~ factor(am), paired = FALSE, var.equal=FALSE, data = mtcars)
```

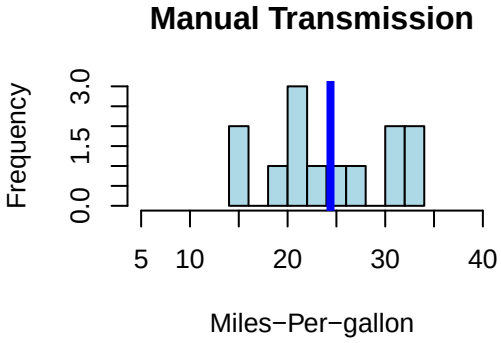
```
Welch Two Sample t-test
```

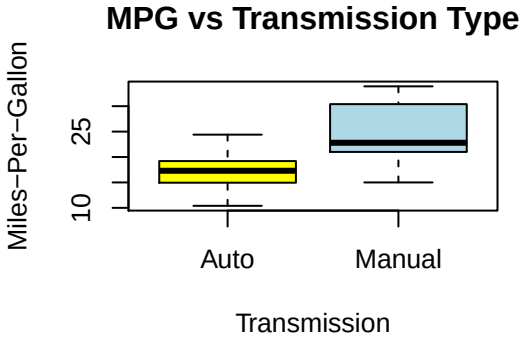
data: mpg by factor(am) t = -3.7671, df = 18.332, p-value = 0.001374 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -11.280194 -3.209684 sample estimates: mean in group 0 mean in group 1 17.14737 24.39231

Appendix A: Figures: Exploratory Analysis









=== END ===