

Regression Models Project - Motor Trend Data ‘mtcars’ Miles Per Gallon Analysis

james c walmsley

12/1/2016

I. Executive Summary:

Add after completing analysis

II. Problem Statement & Questions to Answer:

Assuming I work for Motor Trend, a magazine about the automobile industry. Looking at the data set
They are interested in exploring the relationship between a set of variables and the miles per gallon
They are particularly interested in the following two questions:
Q1 "Is an automatic or manual transmission better for 'mpg'?"
Q2 "Quantify the MPG difference between automatic and manual transmissions"

Grading - Criteria (remove on completion of work)!!!

Did the student interpret the coefficients correctly?

Did the student do some exploratory data analyses?

Did the student fit multiple models and detail their strategy for model selection?

Did the student answer the questions of interest or detail why the question(s) is (are) not answerable?

Did the student do a residual plot and some diagnostics?

Did the student quantify the uncertainty in their conclusions and/or perform an inference correctly?

Was the report brief (about 2 pages long) for the main body of the report and no longer than 5 with supporting appendix of figures?

Did the report include an executive summary?

YES Was the report done in Rmd (knitr) with pdf output?

III. Analysis Considerations:

- Descriptive
- Exploratory
- OLS Ordinary least squares
- Regression to the mean - Simple linear regression
- Statistical linear regression
- Residuals
- Regression inference
- Multivariable regression analysis
- Adjustments
- Residuals, variation, diagnostics
- Multiple variables & model selection
- GLMs
- Binary GLMs
- Count data

IV. Software Environment:

System - session Info:

```
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
```

V. Accessing Data:

Getting the data:

VI. Raw Data Overview:

Motor Trend 'mtcars' data set:

```
any(is.na(mtcars)); colnames(mtcars)
```

```
## [1] FALSE
```

```
## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"  
## [11] "carb"
```

VII. Processing Data:

```
Transformations;  
  1 factor variables 8:11;  
  2 change variable labels in columns 8 & 9;  
    a Note; for column header 8 = vs; variable names = V-block, & S-block;  
    b Note; for column header 9 = am; variable names = Automatic = A, & Manual = M;
```

```
##           mpg cyl disp  hp drat   wt  qsec     vs   am gear carb  
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4  
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4  
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61 S-block Manual    4    1
```

VIII. Exploratory Analysis:

Add narrative here!!
See Appendix A, Figures 1:4

IX. Statistical Modeling, Regression & Model Fit:

```
Assumptions:  
  A Multivariate intercorrelation exists  
  B  
  C  
Simple Linear Regression  
See Appendix A, Figures 5:?
```

```
fslrm <- lm(mpg ~ factor(am), data = mtcars); fslrm
```

```
##  
## Call:  
## lm(formula = mpg ~ factor(am), data = mtcars)  
##  
## Coefficients:  
##      (Intercept)  factor(am)Manual  
##          17.147           7.245
```

```
coef(summary(fslrm))
```

```
##              Estimate Std. Error  t value    Pr(>|t|)  
## (Intercept)   17.147368   1.124603  15.247492 1.133983e-15  
## factor(am)Manual  7.244939   1.764422   4.106127 2.850207e-04
```

X.Preliminary Findings: Questions of Interest: & Interpretation of Results:

A
B
C

XII. Conclusions / Recommendations:

A
B
C

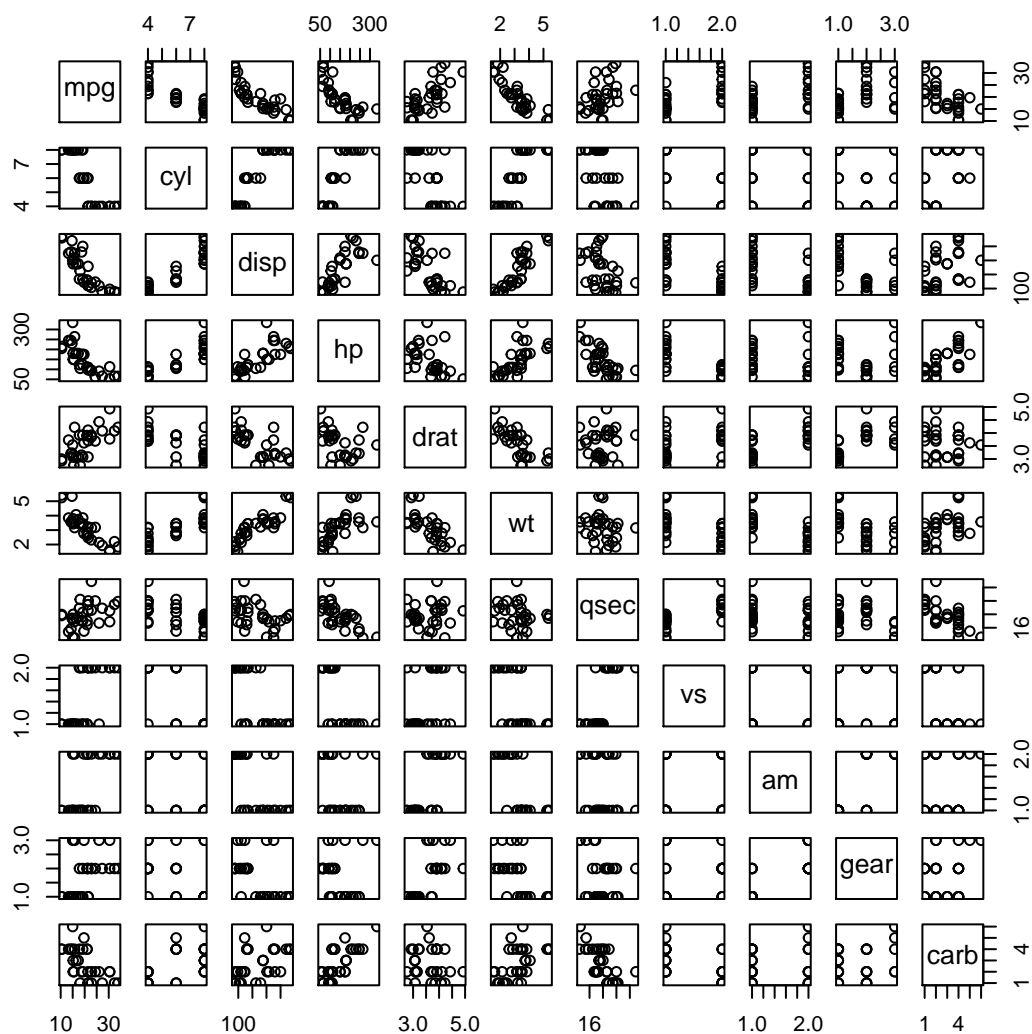
- 1 Challenge the results ?
 - 2 Measures of uncertainty 'e'
-

XIII. Are there other alternative analyses?

A
B

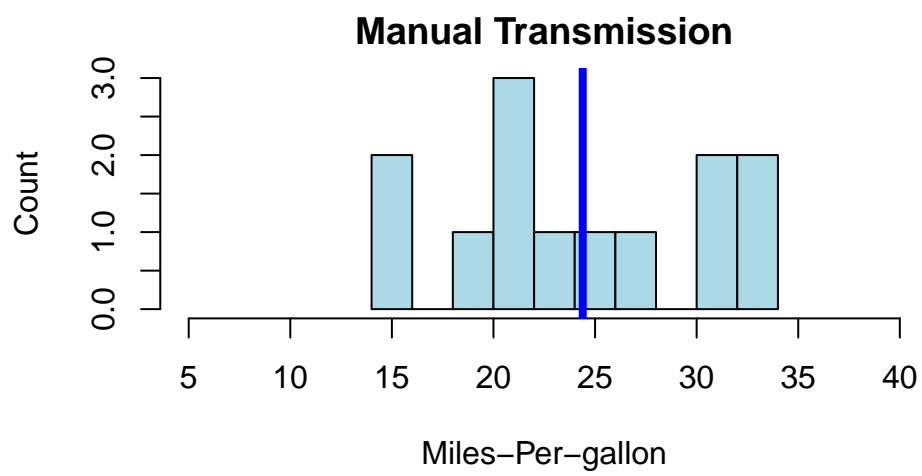
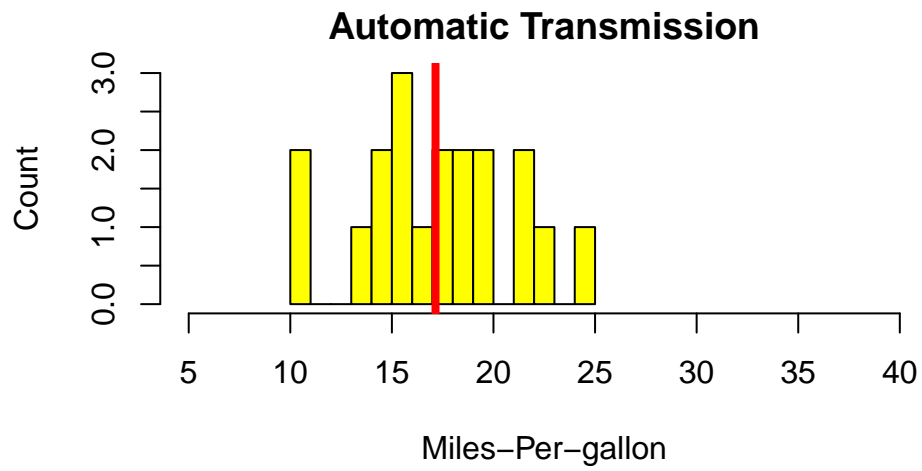
XIV. Appendix A, “Visual Analysis Plots”

Pairs



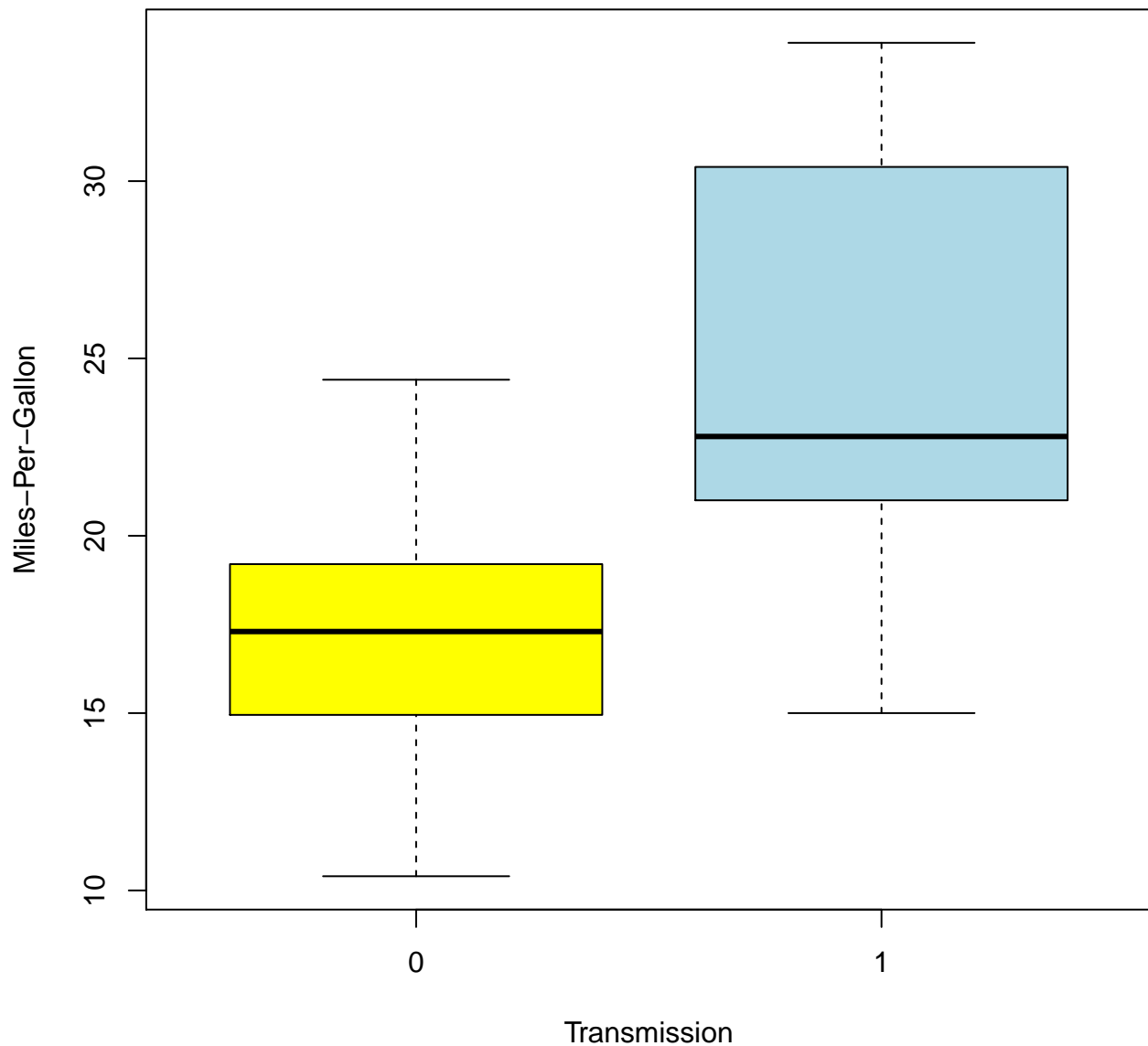
Histograms

##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	V-block	Manual	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	V-block	Manual	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	S-block	Manual	4	1



Box Plots

Automatic vs Manual Transmission Miles Per Gallon



QQ Plots

Single Variable Linear Model Regression Plot

```
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

#To be inserted

#To be inserted

Residuals plot

Residuals vs Fitted

Generalized Linear Models

=== END ===