

Regression Models Project - Motor Trend Data 'mtcars'

Miles Per Gallon Analysis

james c walmsley

12/1/2016

EXECUTIVE SUMMARY

Add after completion of analysis

Problem Statement & Define the Question(s)

Assuming you work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars;

They are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome).

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"
```

Planned Approach

Experimental Design Considerations

```
Correlation  
Deviance  
Descriptive  
    is.na  
    str  
    summary  
Exploratory  
Simple linear comparisons  
Multivariate  
    Additive  
    Multiplicative  
    Missing  
    Steped  
Coefficients  
Residuals  
    Influence  
    Leverage  
Inferential  
    Null Hypothesis
```

```

Alternative Hypothesis
Power or Alpha
Confidence Interval = .95, one or two sided?
pValue
R^2
Predictive >
Causal ~ NA
Mechanistic ~ NA

```

Software Environment

System / session Info:

```

## 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

```

Access the Data:

Raw Data overview:

Motor Trend 'mtcars' data set

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am
## Mazda RX4	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Mazda RX4 Wag	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Datsun 710	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Hornet 4 Drive	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Hornet Sportabout	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Valiant	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Duster 360	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Merc 240D	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Merc 230	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Merc 280	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Merc 280C	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Merc 450SE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Merc 450SL	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Merc 450SLC	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Cadillac Fleetwood	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

## Lincoln Continental	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Chrysler Imperial	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Fiat 128	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Honda Civic	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Toyota Corolla	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Toyota Corona	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Dodge Challenger	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## AMC Javelin	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Camaro Z28	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Pontiac Firebird	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Fiat X1-9	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Porsche 914-2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Lotus Europa	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Ford Pantera L	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Ferrari Dino	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Maserati Bora	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## Volvo 142E	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
##	gear	carb							
## Mazda RX4	FALSE	FALSE							
## Mazda RX4 Wag	FALSE	FALSE							
## Datsun 710	FALSE	FALSE							
## Hornet 4 Drive	FALSE	FALSE							
## Hornet Sportabout	FALSE	FALSE							
## Valiant	FALSE	FALSE							
## Duster 360	FALSE	FALSE							
## Merc 240D	FALSE	FALSE							
## Merc 230	FALSE	FALSE							
## Merc 280	FALSE	FALSE							
## Merc 280C	FALSE	FALSE							
## Merc 450SE	FALSE	FALSE							
## Merc 450SL	FALSE	FALSE							
## Merc 450SLC	FALSE	FALSE							
## Cadillac Fleetwood	FALSE	FALSE							
## Lincoln Continental	FALSE	FALSE							
## Chrysler Imperial	FALSE	FALSE							
## Fiat 128	FALSE	FALSE							
## Honda Civic	FALSE	FALSE							
## Toyota Corolla	FALSE	FALSE							
## Toyota Corona	FALSE	FALSE							
## Dodge Challenger	FALSE	FALSE							
## AMC Javelin	FALSE	FALSE							
## Camaro Z28	FALSE	FALSE							
## Pontiac Firebird	FALSE	FALSE							
## Fiat X1-9	FALSE	FALSE							
## Porsche 914-2	FALSE	FALSE							
## Lotus Europa	FALSE	FALSE							
## Ford Pantera L	FALSE	FALSE							
## Ferrari Dino	FALSE	FALSE							
## Maserati Bora	FALSE	FALSE							
## Volvo 142E	FALSE	FALSE							

##	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
## 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
```

Pre-Process the data:

```
Transformations;
  1 factor variables 8:11;
  2 change variable labels in columns 8 & 9;
    a Note for column header 8 = vs; V = V block motor, & S = Straight block motor;
    b Note for column header 9 = am; A = automatic transmission = A, & M = manual trans

## 'data.frame':   32 obs. of  11 variables:
## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl : num   6  6  4  6  8  6  8  4  4  6 ...
## $ disp: num  160 160 108 258 360 ...
## $ hp  : num  110 110  93 110 175 105 245  62  95 123 ...
## $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt  : num   2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num   16.5 17 18.6 19.4 17 ...
## $ vs  : Factor w/ 2 levels "V","S": 1 1 2 2 1 2 1 2 2 2 ...
## $ am  : Factor w/ 2 levels "A","M": 2 2 2 1 1 1 1 1 1 1 ...
## $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
## $ carb: Factor w/ 6 levels "1","2","3","4",...: 4 4 1 1 2 1 4 2 2 4 ...

##           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  M    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02 V  M    4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61 S  M    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44 S  A    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 V  A    3    2
## Valiant        18.1   6  225 105 2.76 3.460 20.22 S  A    3    1
```

Exploratory Analysis

```
Histograms
Boxplots
Rug
Barplots
Scatterplots
Multiple plots
Graphing - base, lattice, ggplot2
ABline (h/v)
Confidence intervals
```

- Standard error
- Variance
- Fitted lines

- State the H0 & Ha hypothesis here
 - Comparisons
 - Causality?
 - Multivariate
 - Nested Analysis
 - Summaries
 - Boxplots
 - Histograms
 - Rug
 - Barplot
 - ABline (h/v)
 - Scatterplot
 - Multiple scatter plots
 - Graphing - base, lattice, ggplot2
 - Heatmap
 - K-Means
 - Dimension Reduction
 - PCA
 - SVD
 - Figures: Exploratory

Statistical Modeling, Regression & Model Fit

- Simple Linear Regression
- Multivariate Linear Regression
 - lm - simple
 - lm - multivariate
 - lm - nested
 - lm - remove the intercept (-1)
 - lm - step function
- Coefficients / Slope
- Standard Error
- T-Vales
- pValues
- Residuals
 - Leverage
 - Influence
- Confidence Intervals
- Residuals
- Hatvalues
- dfbetas
- Influence Measures
- Anova
 - Chisq
- Ancova
- GLM
 - ?
 - ?

Assumptions Main:

A
B
C

Preliminary Findings: Questions of Interest: & Interpretation of Results;

A
B
C

Inference

Hypothesis testing
Set Seed, if required
One or Two Sided Test
Power / Alpha
Beta = (1 - Alpha)
Confidence Intervals (.95 one sided, .975 two sided)
Standard Error
Variance
student's T-score
Z-score
p-Values
Residual Plots with diagnostics see Appendix

Conclusions / Recommendations

A
B
C

Challenge the results ?
Measures of uncertainty 'e'

What are some possible alternative analyses?

???

Appendix A

Plots with Code
Pairs
Histograms

Box Plots
QQ Plots
Fitted
Residuals
Residuals vs Fitted

=== END ===