

# Regression Models Project - Motor Trend Data 'mtcars'

## Miles Per Gallon Analysis

*james c walmsley*

*12/1/2016*

### Executive Summary:

## Add after completing analysis

### Analysis considerations:

- Descriptive
  - `any(is.na)`
  - `head(data)`
  - `str(data)`
  - `summary(data)`
- Exploratory
  - Pairs
  - Histograms
  - Boxplots
  - QQ plots
- OLS Ordinary least squares
  - General least squares for linear equations
- Regression to the mean - Simple linear regression
- Statistical linear regression
  - Basic - w additive Gaussian error
  - Interpretation of regression coefficients (intercept, slope)
  - Regression - prediction
- Residuals
  - Residual variation
  - Influence
  - Leverage
  - Estimate residual variation
  - R squared
- Regression inference
  - Parameters
  - Confidence intervals
  - Prediction
- Multivariate regression analysis
  - Linear models
  - Two variable simple linear regression (additive) / (multiplicative)
  - Summary coefficients
  - Fitted values, residuals and residual variation
  - Summary coefficients
  - Model Adjustment
- GLMs
  - Linear
  - Logistic

```

Poisson
Binary GLMs
    Odds
    Fitting
Poisson
    Count data
Predictive ~ NA
Causal ~ NA
Mechanistic ~ NA

```

## Accessing data:

Getting the data:

```

library(datasets)
data("mtcars")
head(mtcars)

```

```

##           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
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant        18.1   6  225 105 2.76 3.460 20.22  1  0    3    1

```

## Processing 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 tran

```

```

data("mtcars")
mtcars$vs <- factor(mtcars$vs, labels = c("V", "S")); mtcars$am <- factor(mtcars$am, labels = c("A", "M"),
str(mtcars); head(mtcars)

```

```

## '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

## Statistical modeling, regression & model fit:

```

Assumptions:
  A
  B
  C
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
GLMs

```

## Inference:

```

Assumptions:
  A Variables other than transmission type may influence MPG ratings
  B
  C
State H0 & Ha hypothesis'here
Set Seed, if required
One or Two Sided Test
Power / Alpha
Beta = (1 - Alpha)
Confidence Intervals (.95 one sided, .975 two sided)
Standard Error

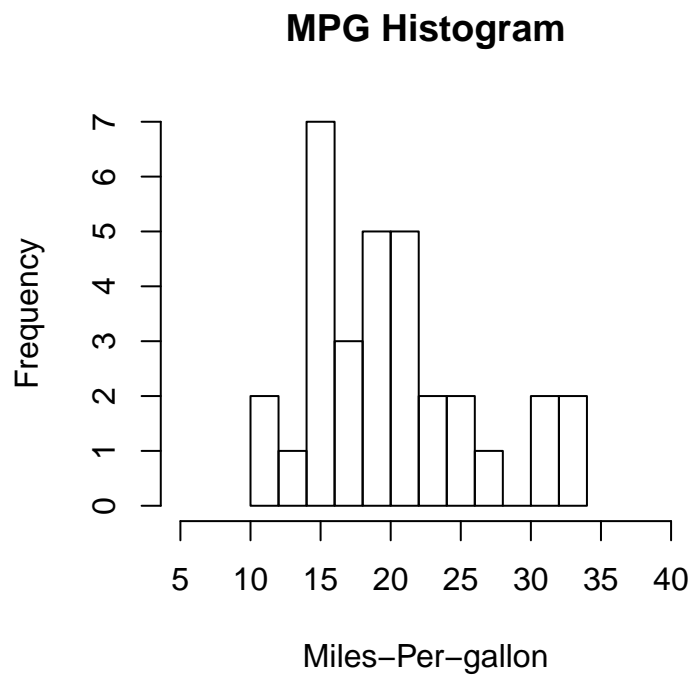
```

Variance  
student's T-score  
Z-score  
p-Values  
Residual Plots with diagnostics see Appendix

## What are some possible alternative analyses?

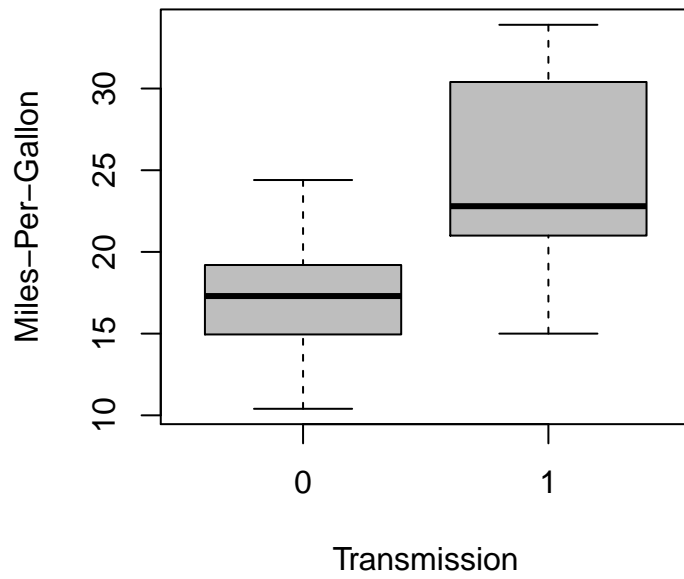
A  
B

Histograms



Box Plots

**Boxplot MPG**



QQ Plots

Residuals

==== END ====