Peer-graded Assignment: Regression Models Course Project

1. Instructions

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:

1. Is an automatic or manual transmission better for MPG
2. Quantify the MPG difference between automatic and manual transmissions

2. Dataset Inforation

mtcars is a data frame with 32 observations on 11 (numeric) variables.

1. mpg Miles/(US) gallon
2. cyl Number of cylinders
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 Engine (0 = V-shaped, 1 = straight)
9. am Transmission (0 = automatic, 1 = manual)
10. gear Number of forward gears

	mpg <dbl>	cyl <dbl>	disp <dbl>	hp <dbl>	drat <dbl>	wt <dbl>	qsec <dbl>	vs <dbl>	am <dbl>
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0

Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0

6 rows | 1-10 of 12 columns

3. Exploratory analysis

```
t.test(mtcars[mtcars$am == 0,]$mpg, mtcars[mtcars$am == 1,]$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars[mtcars$am == 0, ]$mpg and mtcars[mtcars$am == 1, ]$mpg
## 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 of x mean of y
##  17.14737  24.39231
```

There is significant difference in mpg between cars with automatic transmission and cars with manual transmission.

```
mdl <- lm(mpg ~ ., mtcars); summary(mdl)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
```

```
## gear            0.65541      1.49326    0.439    0.6652
## carb           -0.19942      0.82875   -0.241    0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

It seems that wt, am, drat, qsec, gear are all high in coefficient with mpg, especially wt. Hence I am going to compare multiple models.

```
anova(lm(mpg ~ am, mtcars),                        #1
      lm(mpg ~ am + wt, mtcars),                   #2
      lm(mpg ~ am + wt + drat, mtcars),            #3
      lm(mpg ~ am + wt + drat + qsec, mtcars),     #4
      lm(mpg ~ am + wt + drat + qsec + gear, mtcars), #5
      lm(mpg ~ wt, mtcars),                        #6
      lm(mpg ~ wt + drat, mtcars),                 #7
      lm(mpg ~ wt + drat + qsec, mtcars),          #8
      lm(mpg ~ wt + drat + qsec + gear, mtcars))   #9
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	30	720.8966	NA	NA	NA	NA
2	29	278.3197	1	442.5769020	68.8054979	8.912768e-09
3	28	266.9876	1	11.3321259	1.7617561	1.959437e-01
4	27	167.8863	1	99.1012597	15.4068400	5.681055e-04
5	26	167.2395	1	0.6467768	0.1005516	7.537011e-01
6	30	278.3219	-4	-111.0824028	4.3173740	8.239000e-03
7	29	269.2413	1	9.0806433	1.4117279	2.455103e-01
8	28	183.5216	1	85.7197187	13.3264703	1.154941e-03
9	27	182.3309	1	1.1906975	0.1851125	6.705613e-01

9 rows

1. Comparing #1 (excluding wt as regressor) with the rest models (including wt as regressor), the difference in RSS (Residual Sum of Squares) is significant, yet P-value is small.
2. Comparing #2~5 (excluding am as regressor) with #6~9 (including am as regressor), the difference in RSS is small.

3. My guess is wt is much higher in coefficient with mpg than other regressors.
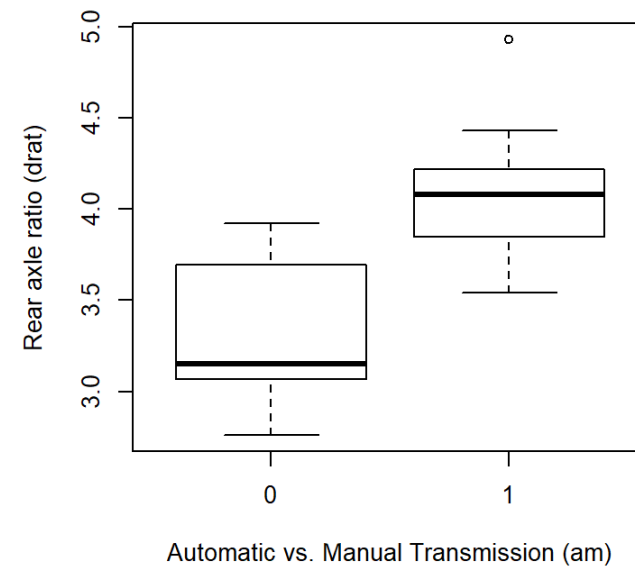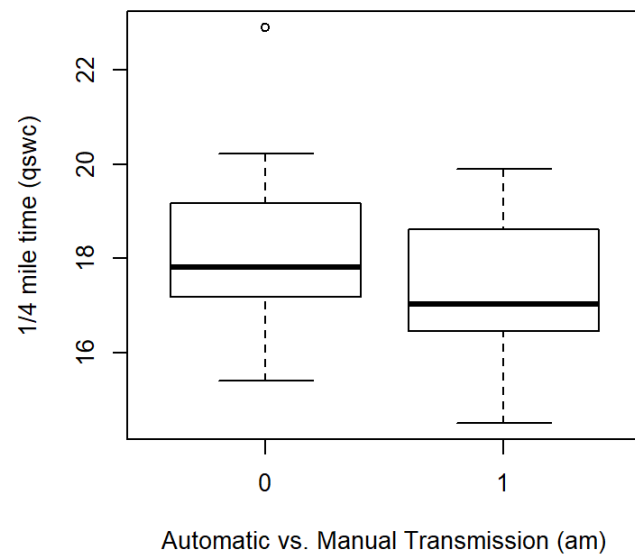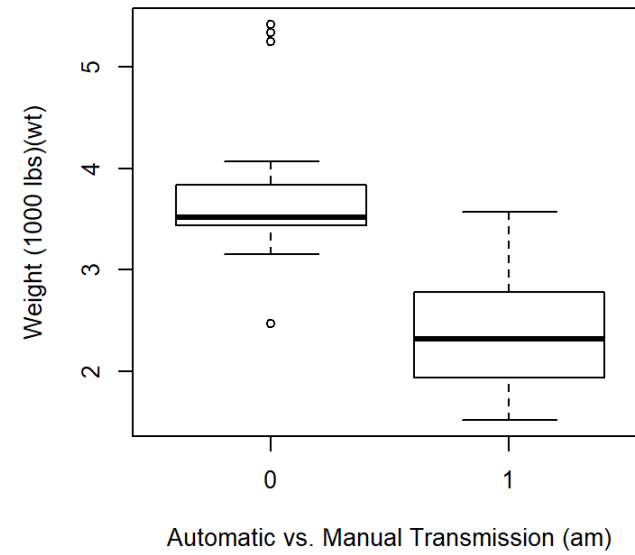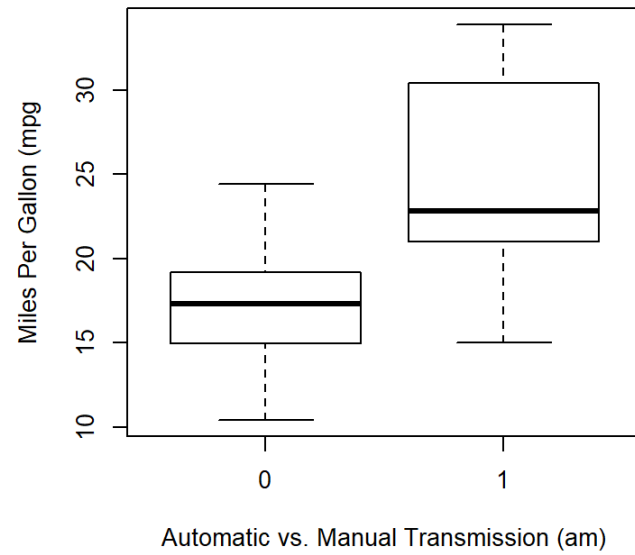
Form the plots above we can see, cars with automatic transmission also weight heavier in general than cars with manual transmission.

4. Fitting models

```
coef(lm(mpg ~ am, mtcars))
```

```
## (Intercept)          am
##   17.147368    7.244939
```

```
coef(lm(mpg ~ am + wt, mtcars))
```

```
## (Intercept)          am          wt
## 37.32155131 -0.02361522 -5.35281145
```

```
coef(lm(mpg ~ am + wt + qsec, mtcars))
```

```
## (Intercept)          am          wt        qsec
##    9.617781    2.935837   -3.916504    1.225886
```

```
mdl <- lm(mpg ~ am + wt + qsec + drat, mtcars); summary(mdl)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec + drat, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3046 -1.6260 -0.6634  1.2097  4.6626
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.6277     8.2103   0.929 0.361095
## am            2.5729     1.6225   1.586 0.124446
## wt           -3.8040     0.7592  -5.010 2.96e-05 ***
## qsec          1.1958     0.2995   3.992 0.000452 ***
## drat          0.6429     1.3551   0.474 0.639003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.494 on 27 degrees of freedom
## Multiple R-squared:  0.8509, Adjusted R-squared:  0.8288
## F-statistic: 38.52 on 4 and 27 DF,  p-value: 8.673e-11
```

```
par(mfrow = c(2,2)); plot(mdl)
```
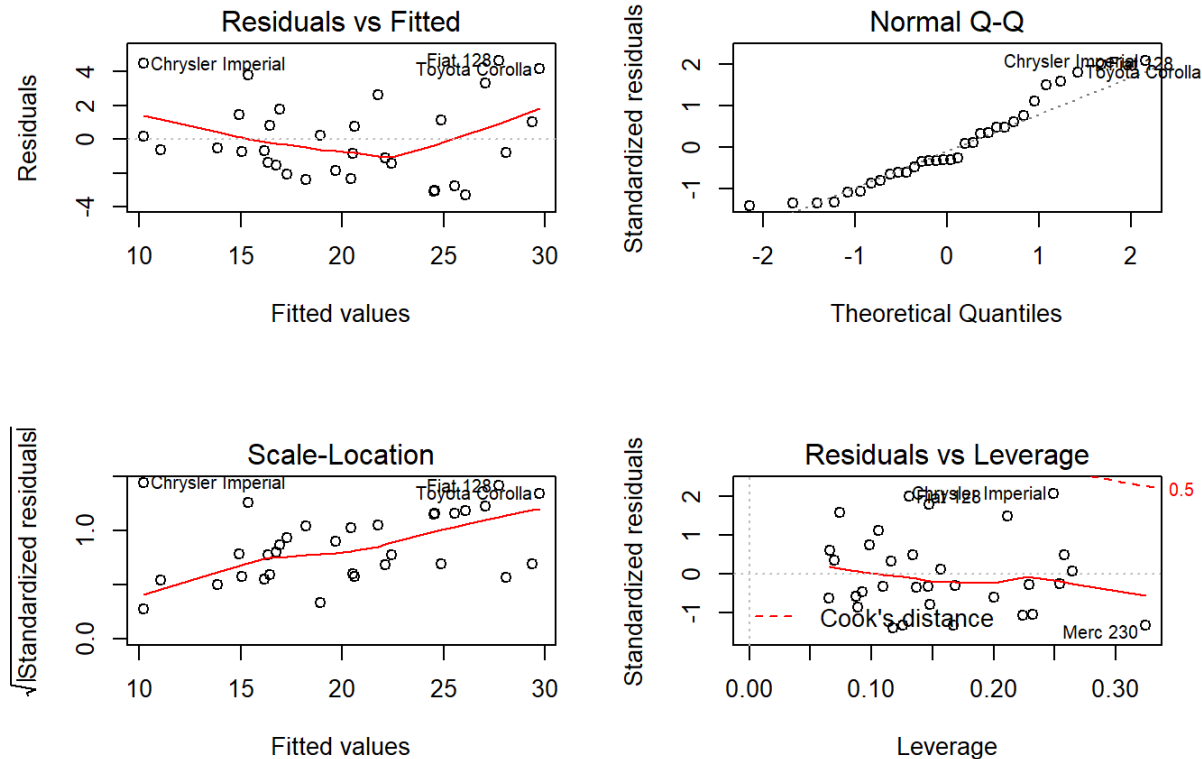

If we add different regressors into the linear model, and we will have different coefficients for am.

5. Summary

1. In general, cars with manual transmission are **7.24 higher** in **MPG(Mile per Gallon)** than cars with automatic transmission.
2. However, if weight is also considered, we will find that the former are **0.02 lower** in MPG than the latter; if weight and qsec(1/4 mile time) also considered, **2.93 higher**; if weight, qsec, and drat(drat Rear axle ratio) also considered, **2.57 higher**.
3. Hence we **cannot** conclude that an automatic transmission is better for MPG than a manual one.