Regression Models Course Project

Chiu Wei Chieh

Question 1 : Is an automatic or manual transmission better for MPG?

Question 2 : Quantify the MPG difference between automatic and manual transmissions.

My strategy :
1. Explore the relationship between mpg and other variables.
2. Fit multiple models and diagnose with redidual plots and nested model testing.
3. Select a suitable model and answer the questions.

1. Exploratory

```
library(datasets); data(mtcars)
fitAll <- lm(mpg ~ ., data = mtcars)
summary(fitAll)$coef
```

From the summary of all variables, we have the estimate beta value(absolute value) in decreasing order : wt > am > qsec > drat > gear > vs > carb > cyl > hp > disp

Figure 1 & 2 : Manual transmission seems have better MPG.
Figure 3 : wt and mpg is negative correlation. Automatic transmission cars is heavier.
Figure 4 : qsec and mpg is positive correlation.

2. Fit models and diagnose

Nested model testing 1:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + qsec
## Model 4: mpg ~ factor(am) + wt + qsec + drat
## Model 5: mpg ~ factor(am) + wt + qsec + drat + gear + vs + carb + cyl +
##     hp + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 63.0133 9.325e-08 ***
## 3     28 169.29  1    109.03 15.5240 0.0007497 ***
## 4     27 167.89  1      1.40  0.1993 0.6598757
## 5     21 147.49  6     20.39  0.4839 0.8128394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnose 1: In the mpg ~ am + wt model, am seems have no significant effect on mpg(Figure 5). But mpg ~ am + qsec model shows am has significant effect on mpg(Figure 6). And considering the result of the first nested model testing, our model should include am, wt and qsec.

Nested model testing 2(am, wt, qsec and their interactions) :

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + factor(am):wt
## Model 4: mpg ~ factor(am) + wt + qsec + factor(am):wt
## Model 5: mpg ~ factor(am) + wt + qsec + factor(am):wt + factor(am):qsec
## Model 6: mpg ~ factor(am) + wt + qsec + factor(am):wt + factor(am):qsec +
##     wt:qsec
##   Res.Df    RSS Df Sum of Sq        F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 101.5291 2.751e-10 ***
## 3     28 188.01  1     90.31  20.7180 0.0001191 ***
## 4     27 117.28  1     70.73  16.2261 0.0004609 ***
## 5     26 116.47  1      0.80   0.1841 0.6715862
## 6     25 108.98  1      7.50   1.7196 0.2016573
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnose 2: Compare model with and without am*wt term. The result shows that it has great effect on model. So we want to include this term.

```
##   withAmWt withoutAmWt
## 1    2.936      14.079
```

Diagnose 3: Both am*qsec and wt*qsec has no significant effect on model.

3. Model of choice

fit6 <- lm(mpg ~ factor(am) + wt+ qsec + factor(am)*wt, data = mtcars)

```
summary(fit6)$coef
```

```
##                 Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     9.723053  5.8990407  1.648243 0.1108925394
## factor(am)1    14.079428  3.4352512  4.098515 0.0003408693
## wt             -2.936531  0.6660253 -4.409038 0.0001488947
## qsec            1.016974  0.2520152  4.035366 0.0004030165
## factor(am)1:wt -4.141376  1.1968119 -3.460340 0.0018085763
```

> **Answer 1 : According to our model, manual transmission is better for MPG than automatic transmission.**
>
> **Answer 2 : Holding other variables in constant, the mpg difference between manual and automatic transmission is 14.079 miles/gallon with standard error 3.435.**

Appendix

Figure 1

```
g <- ggplot(data = mtcars, aes(y = mpg, x = factor(am)))
g <- g + geom_violin(aes(fill = factor(am)))
g
```

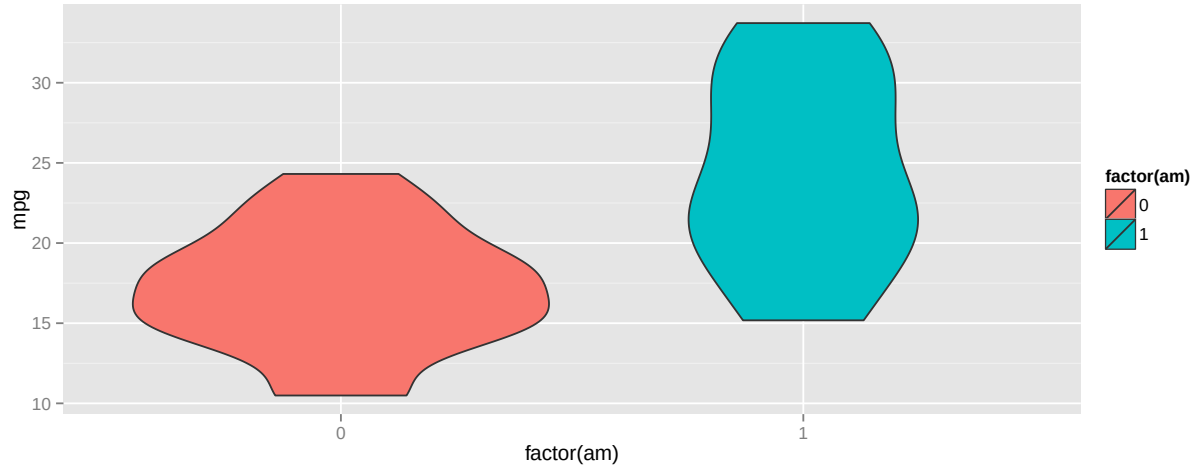

Figure 2

```
g <- ggplot(data = mtcars, aes(y = mpg, x = factor(am)))
g <- g + geom_point(aes(fill = wt)) + geom_smooth(method = lm)
g
```

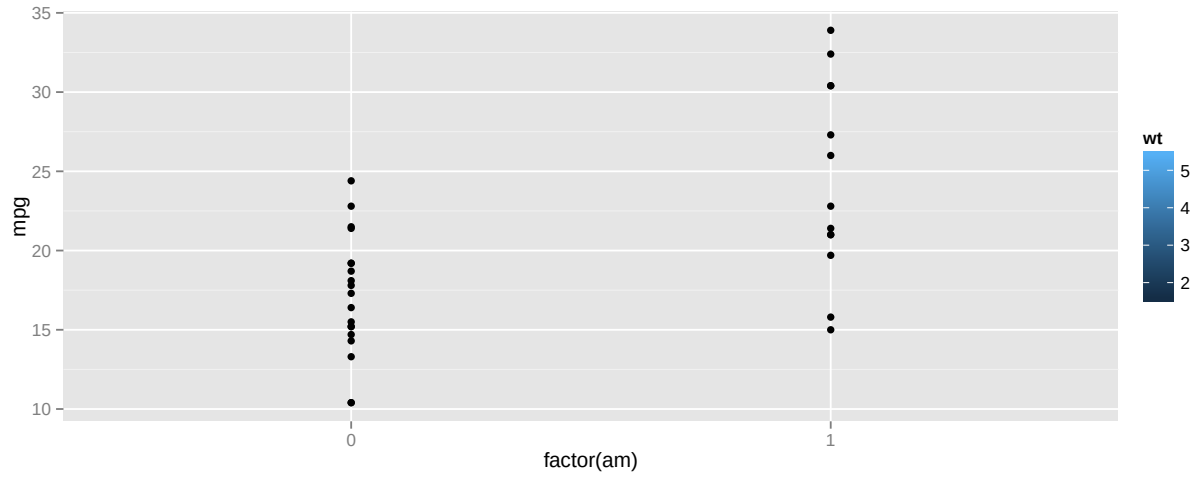

Figure 3

```
g <- ggplot(data = mtcars, aes(y = mpg, x = wt))
g <- g + geom_point(aes(colour = factor(am))) + geom_smooth(method = lm)
g
```

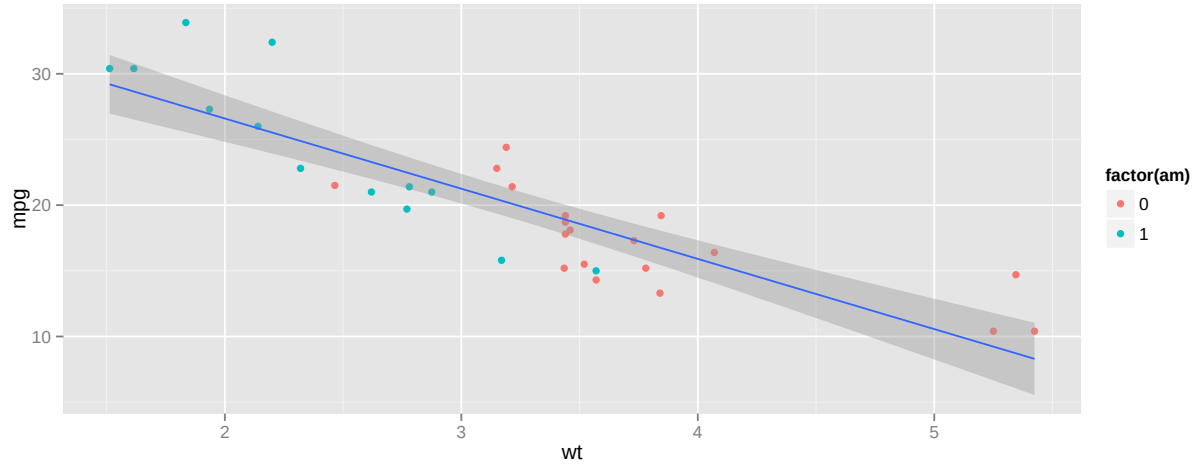

Figure 4

```
g <- ggplot(data = mtcars, aes(y = mpg, x = qsec))
g <- g + geom_point(aes(colour = factor(am))) + geom_smooth(method = lm)
g
```

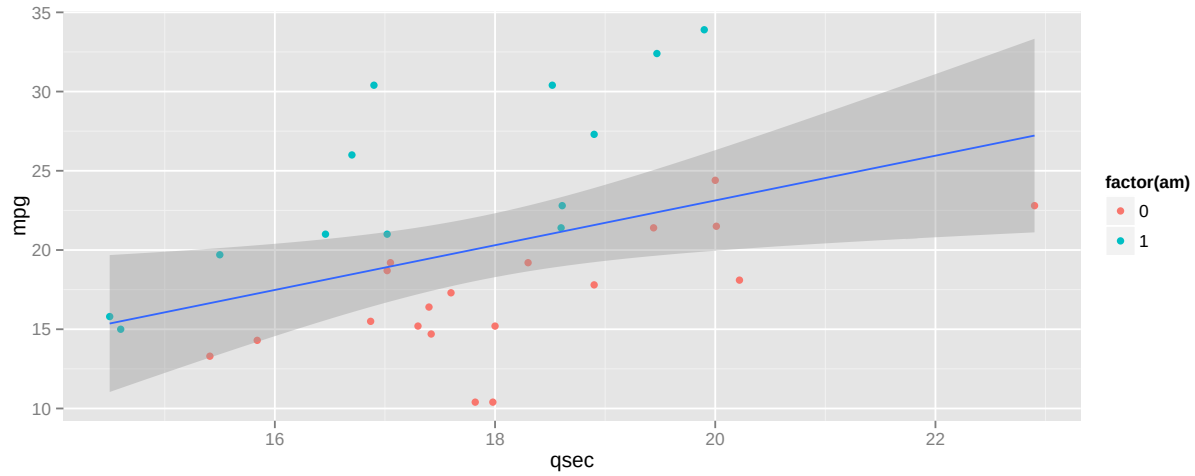

Figure 5

```
dataResid <- data.frame(y = resid(lm(mpg ~ wt, data=mtcars)), x = resid(lm(am ~ wt, data=mtcars)))
g <- ggplot(data = dataResid, aes(y = y, x = x))
g <- g + geom_point() + geom_smooth(method = lm)
g
```

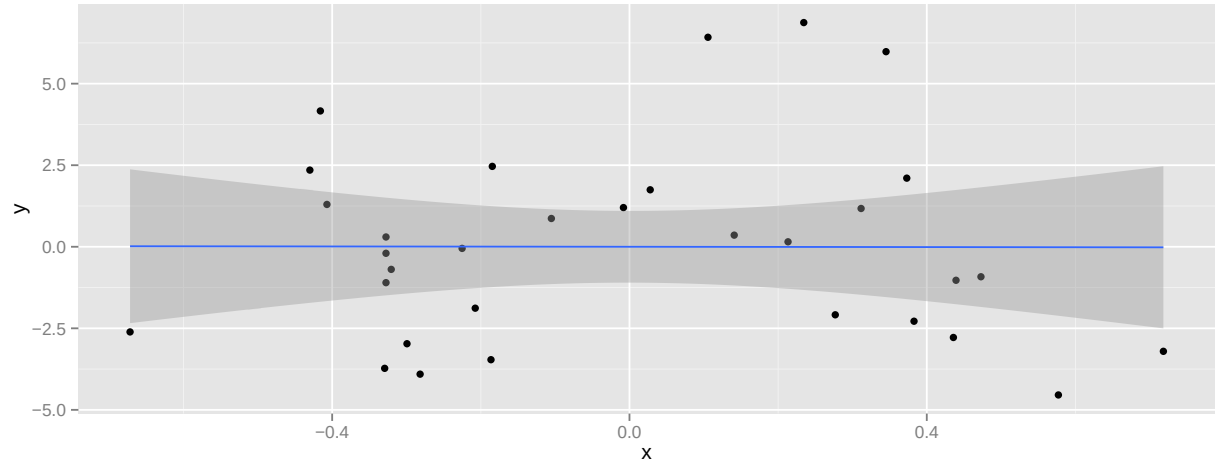

Figure 6

```
dataResid <- data.frame(y = resid(lm(mpg ~ qsec, data=mtcars)), x = resid(lm(am ~ qsec, data=mtcars)))
g <- ggplot(data = dataResid, aes(y = y, x = x))
g <- g + geom_point() + geom_smooth(method = lm)
g
```