Automatic or Manual Transmission Better For MPG?

Executive Summary

Fuel efficiency of a car is a key factor considered when buying cars.In this paper, we look at 1974 Motor Trend data [1] for the purpose of evaluating factors on fuel efficiency. We focused on the effects of automatic vs manual transmission on the gas mileage. We find that we cannot form a conclusion on whether automatic or manual transmission results in superior mileage. The reader is encouraged to consider cars of a weight and number of cylinders if they are seeking better mileage.

Preprocessing the Data

Load data and factorize transmission type and number of cylinders to factors.

```
data(mtcars)
mtcars$am <- factor(mtcars$am, levels=c(0,1), labels=c('Automatic', 'Manual'))
mtcars$cyl <- factor(mtcars$cyl)
```

Explotory Data Analysis

The first step of our analysis is to simply explore the relationship between transmission type and mpg (Appedix: Figure 1). We see that on mtcars, the mean mileage of automatic transmission is 17.147 mpg and the manual transmission is 24.392 mpg.The graph below hints at an increase in mpg when transmission type was manual.This data may have other variables which may play a bigger role in determination of mpg.

Model Selection

We perform an analysis of variance exercise on the data.

```
analysis <- aov(mpg ~ ., data = mtcars)
summary(analysis)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## cyl          2    825     412   61.86 2.7e-09 ***
## disp         1     58      58    8.65  0.0081 **
## hp           1     19      19    2.78  0.1113
## drat         1     12      12    1.79  0.1963
## wt           1     56      56    8.37  0.0090 **
## qsec         1      2       2    0.23  0.6377
## vs           1      0       0    0.05  0.8336
## am           1     17      17    2.49  0.1306
## gear         1      4       4    0.56  0.4618
## carb         1      2       2    0.29  0.5948
## Residuals   20    133       7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Clearly important variables (p-value below 0.05) which determine miles per gallen are the number of cylinders,displacement and weight. We need to refine our linear model by introducing these three confounding variables into the exercise.

To enahncement the first model, we now perform a linear regression with weight, displacement and cylinders as confounding variables in the lm.

```
lm1 <- lm(mpg ~ cyl + disp + wt + am, data = mtcars)
summary(lm1)$coefficients
```

```
##              Estimate Std. Error t value   Pr(>|t|)
## (Intercept) 33.816067    2.91427 11.6036 8.792e-12
## cyl6        -4.304782    1.49236 -2.8846 7.775e-03
## cyl8        -6.318406    2.64766 -2.3864 2.458e-02
## disp         0.001632    0.01376  0.1186 9.065e-01
## wt          -3.249176    1.24910 -2.6012 1.513e-02
## amManual     0.141212    1.32675  0.1064 9.161e-01
```

Coefficient of Displacement has a p-value which allows is to keep the hypothesis that the coefficient is null. We will now observe to see if removing the displacement viable can increase the adjusted R-square of the test.

```
lm1 <- lm(mpg ~ cyl + wt + am, data = mtcars)
summary(lm1)$coefficients
```

```
##              Estimate Std. Error t value   Pr(>|t|)
## (Intercept)  33.7536      2.813 11.9971 2.496e-12
## cyl6         -4.2573      1.411 -3.0167 5.515e-03
## cyl8         -6.0791      1.684 -3.6105 1.228e-03
## wt           -3.1496      0.908 -3.4685 1.771e-03
## amManual      0.1501      1.300  0.1154 9.089e-01
```

Indeed, the adjusted r-squared is higher at 0.81. We choose this is our final model.

Clearly after introducing the number of cylinders and weight as a confounding variable, the ceofficient of the am variable becomes very small with a large p-value. This means that we cannot reject the hypothesis the coefficient of am is 0.

Residual Analysis

Now we plot the residuals to search for discernible patterns (Appendix: figure 2)

There is no discernible pattern when residuals are plotted against the fitted values. The Q-Q plot of the residuals vs a normal we can see the studentized residuals are approximately normal.This confirms that our model is a good fit

Conclusion

While a subtle relationship exists between transmission type and mpg, this is insignificant and we cannot concluded that shifting from automatic to manual gearing will result in a car which gives better mileage. The reader is encouraged to consider the car's weight and number of cylinders instead to determine the mileage of the vehicle.

References

[1] Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391411.

Appendix

Figure 1: mpg comparision with transmission type.

```
fun_mean <- function(x){
    return(
        data.frame(y=round(mean(x),3),label=paste("mean =",round(mean(x),3)))
        )}

ggplot(mtcars,aes(x=am,y=mpg)) +
    geom_boxplot(aes(fill=am)) +
    stat_summary(fun.y = "mean", geom="point",colour="darkred", size=3) +
    stat_summary(fun.data = fun_mean, geom="text", vjust=-0.7)
```
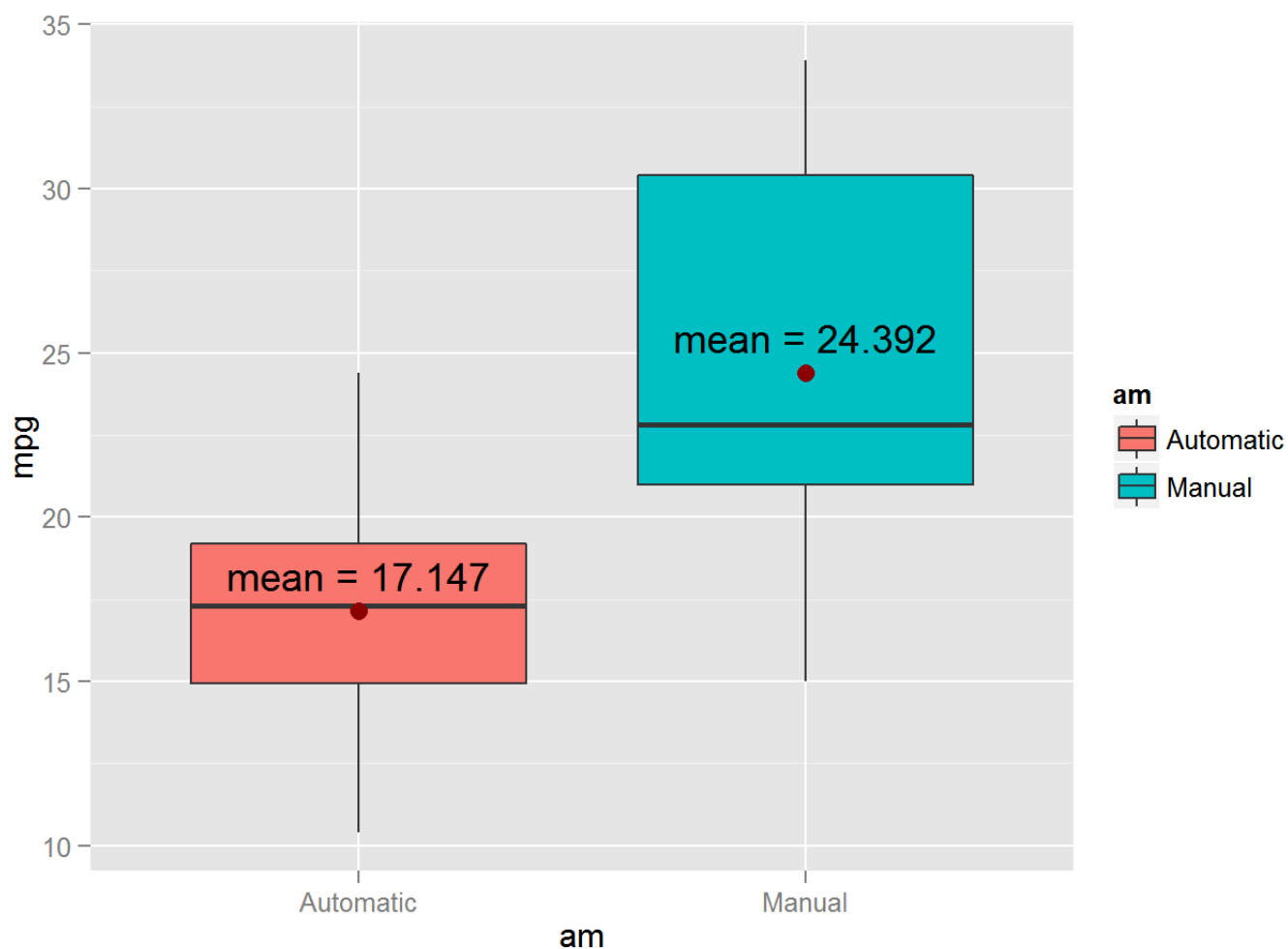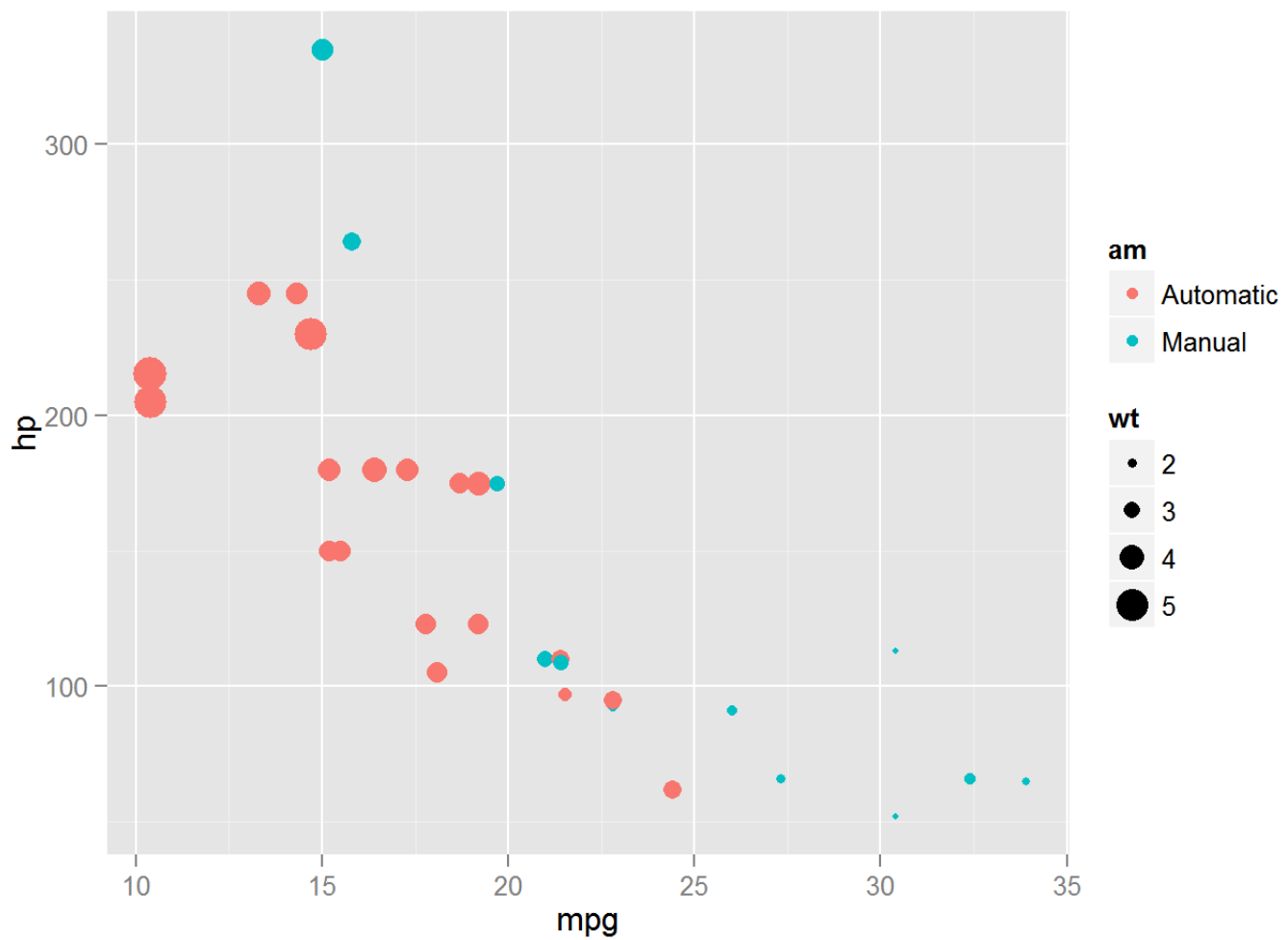

Figure 2: Residual analysis.

```
par(mfcol = c(1,2))
plot(fitted(lm1), resid(lm1))
abline(h = 0)
qqnorm(rstudent(lm1))
qqline(rstudent(lm1))
```


Figure 3: mpg comparision with respect to weight and horsepower.

```
ggplot(data=mtcars,aes(x=mpg,y=hp,
                        colour=am,size=wt)) + geom_point()
```

Figure 4: Pair analysis of all variables.

```
key.variables<- c(1,2,3,4,6,9)
ggpairs(mtcars[,key.variables],
        upper = list(continuous = "density", combo = "box"),
        lower = list(continuous = "points", combo = "dot"))
```