REGRESSION MODELS PROJECT *Lawrence Lau*

Summary

Given the mtcars dataset, we use a linear regression and verify the results with a t.test to see that manual transmissions average 7.24 miles per gallon more than automatic transmissions. We then use covariate adjustment and multiple regression models to explore other variables that could be factors, and after finding a best model to hold all else equal we discover manual transmissions are more efficient by 1.49 mpg .

Exploratory Data Analysis

In Appendix 1, Boxplot - Transmission Type vs MPG, the avg mpg difference between automatics (17.15) and manuals (24.39) is 7.24 mpg. Let's run a hypothesis test to verify there really is a difference.

```
## [1] 0.001373638
```

With a p-value of less than .05, we reject the null hypothesis that the average miles per gallon are the same in automatics and manuals. So there really is a difference. But as r-squared shows, that only explains ~36% of the variance. There must be other factors.

```
summary(lm(mpg~am))$r.squared
```

```
## [1] 0.3597989
```

Models Analyses

We fit each individual predictor variable against outcome MPG and rank their effect using the coefficient.

```
fit2 <- lm(mpg~., data=mtcars); fit2$coefficients
```

```
## (Intercept)         cyl        disp          hp        drat          wt
## 12.30337416 -0.11144048  0.01333524 -0.02148212  0.78711097 -3.71530393
##        qsec          vs          am        gear        carb
##  0.82104075  0.31776281  2.52022689  0.65541302 -0.19941925
```

The greatest net effect variables, wt, carb, cyl, vs, gear, drat, and qsec, are taken and after excluding any with correlation > .5 with am, fit them into a new regression with predictor am against mpg. First, correlations.

```
##        am/wt    am/carb    am/cyl     am/vs   am/gear   am/drat    am/qsec
## 1 -0.6924953 0.05753435 -0.522607 0.1683451 0.7940588 0.7127111 -0.2298609
```

Next, we'll leave out any pairs with correlation > .5 and run covariate model tests.

```
##  (Intercept)          am           wt  (Intercept)          am
##  37.32155131  -0.02361522  -5.35281145  23.14583626   7.65311904
##         carb  (Intercept)           am          cyl  (Intercept)
##  -2.19174787  34.52244254   2.56703470  -2.50095764  14.59444444
##           am           vs  (Intercept)           am         qsec
##   6.06666667   6.92936508 -18.88928081   8.87633094   1.98186968
```

We'll finally run an ANOVA test starting with predictor am and outcome mpg, then add variables individually in order of influence (most to least based on above coefficients).

```
Mod1 <- lm(mpg~am); Mod2 <- lm(mpg~am+wt); Mod3 <- lm(mpg~am+wt+vs); Mod4 <- lm(mpg~am+wt+vs+cyl)
anova <- anova(Mod1, Mod2, Mod3, Mod4)
anova[6]
```

```
##      Pr(>F)
## 1
## 2 < 2.2e-16 ***
## 3  0.006176 **
## 4  0.062218 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value crosses above the .05 threshold after Model 3, which means the variables in that model, am+wt+vs (transmission type, weight, and cylinder alignment) gives us our best model. A residual plot of that model (reference APPENDIX 2, Residual Plot, lm(mpg~am+wt+vs)) confirms constant variance (no heteroskedasticity) and a normal distribution. It's a good model fit.

```
## (Intercept)          am          wt          vs
##   30.078722    1.491345   -3.784454    3.615040
```

R-Squared

```
## [1] 0.8078982
```

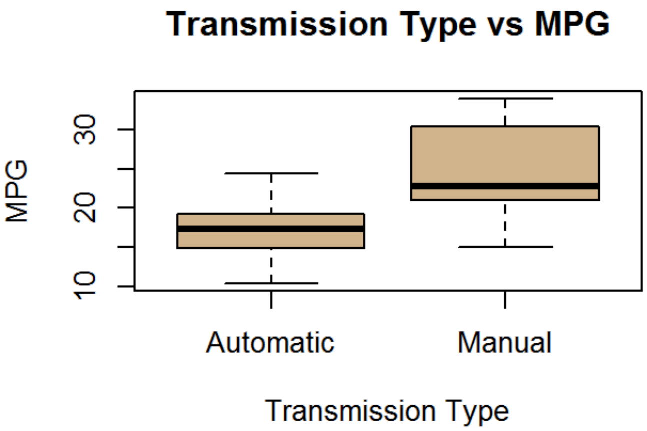
Our best model coefficient shows us that we can expect a car with a manual transmission to average 1.49 mpg better than its counterpart with an automatic transmission. The R-squared number of .8079 means this model explains ~81% of the variance- in other words, we still have about a 19% rate of uncertainty.

Conclusion

Our initial exploratory data analysis suggested that cars with manual transmissions are more efficient by 7.24 mpg. While a hypothesis test proved there was a difference, the r-squared number showed the transmission type was responsible for only ~36% of variance. To find the true difference in mpg caused by transmission type, all other variables were regressed individually as a predictor variable against mpg, and the variables that had the largest net effect were selected to run a correlation test against am (transmission type). Any pairs that had a correlation > .5 (suggesting the pairs were dependent or otherwise related) were thrown out. The remaining variables were then individually added as predictors to a model with predictor am against outcome mpg. An ANOVA test was run with those models to see which model was the best (last model for which p-value < .05). The resulting model was a regression with the predictor variables am (transmission type) + wt (weight) + vs (cylinder alignment) against outcome mpg. The coefficients 1.49(am), -3.78(wt), and 3.62(vs) give us our best estimates that ***1.)*** a standard/manual transmission is more efficient than an automatic by 1.49 mpg , ***2.)*** for every 1,000 pounds heavier a car gets it loses 3.78 mpg and ***3.)*** an inline/straight cylinder shape gets 3.62 mpg more than a v-shaped arrangement. The intercept of 30.079 tells us the average car averages 30.079 miles per gallon. These numbers explain about 81% of the variance and leave 19% uncertain.

APPENDIX 1 - Boxplot, Transmission Type vs MPG

```
par(mfrow=c(2,2))
boxplot(mpg~am, data = mtcars, col = c("tan", "tan"), xlab = "Transmission Type", ylab = "MPG", main
= "Transmission Type vs MPG", xaxt="n")
axis(side=1, at=c(0,1)+1, labels=c("Automatic", "Manual"))
```



APPENDIX 2 - Residual Plot, lm(mpg~am+wt+vs)

```
par(mfrow = c(2,2))
plot(lm(mpg~am+wt+vs))
```