Automatic vs Manual Transmission on MPG

Regression Models Course Project

Xinchun Tian

Context

Exerpt from Regression Models Course Project.

Suppose I work for Motor Trend, a magazine about the automobile industry. I was asked to Look at a data set of a collection of cars in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome), particularly interested in answering the following two questions:
1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.

Let's load the "mtcars"" data and examine the first couple of rows.

```
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
```

The data look clean and there are no missing values as checked by "is.na"" is 0.

The "am" column tells one car is automatic (0) or manual (1).

```
autocars           <- mtcars[mtcars$am==0,]
manualcars         <- mtcars[mtcars$am==1,]
autocars_mpgmean   <- mean(autocars$mpg)
autocars_mpgvar    <- var(autocars$mpg)
manualcars_mpgmean <- mean(manualcars$mpg)
manualcars_mpgvar  <- var(manualcars$mpg)
n1                 <- dim(autocars)[1]
n2                 <- dim(manualcars)[1]
```

The automatic and manual samples are independent samples. The MPG mean and variance for automatic and manual cars are $\bar{X}_{automatic}$=17.15, $s^2_{automatic}$=14.7, $\bar{X}_{manual}$=24.39, s^2_{manual}=38.03. There are 19 cars in automatic sample, and 13 cars in manual sample.

The pooled variance estimator and standard deviation estimator can be calculated as

```
vp <- (autocars_mpgvar*(n1-1) + manualcars_mpgvar*(n2-1))/(n1+n2-2)
sp <- sqrt(vp)
```

which are 24.03 and 4.9.

The 95% confidence MPG interval of (manual - automatic) can be caluclated as:

```
interval <- manualcars_mpgmean - autocars_mpgmean +
            c(-1,1) * qt(0.975, n1+n2-2) * sp * (1/n1 + 1/n2)^0.5
```

Because the interval [3.64, 10.85] is entirely above zero, therefore the **conclusion is that manual transmission is better for MPG**.

Consider the linear model

$$Y_i = \beta_0 + X_{i1}\beta_1 + \epsilon_i, \tag{1}$$

where each X_{i1} is either automatic (0) or manual (1) so that it is a 1 if measurement i is in manual group and 0 in automatic group. Let us use the linear model fit to quantify the change in means between the automatic and manual groups. The automatic group is chosen as the reference category.

```
summary(lm(mpg~am, data=mtcars))$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

The above table shows that the coefficient for "am" which is 7.24 is the estimated increase in MPG from group 0 or automatic to group 1 or manual. The following plot shows the comparison of the means with 1 standard deviation intervals for automatic (black) and manual (red).