Cars-data

Yaswanth Pulavarthi

9/14/2020

## Loading data

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

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.2

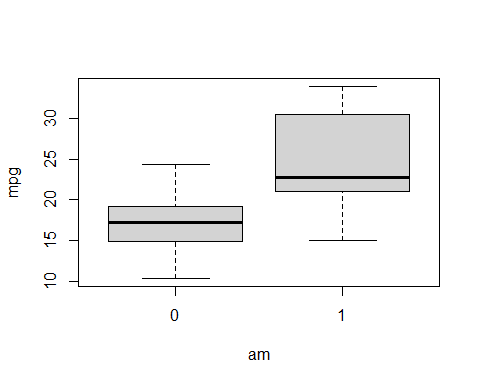
## Processing Data

mtcars$cyl <- as.factor(mtcars$cyl)  
mtcars$vs <- as.factor(mtcars$vs)  
mtcars$am <- as.factor(mtcars$am)  
mtcars$gear <- factor(mtcars$gear)  
mtcars$carb<- factor(mtcars$carb)  
str(mtcars)

## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...  
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...  
## $ qsec: num 16.5 17 18.6 19.4 17 ...  
## $ vs : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...  
## $ am : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...  
## $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...  
## $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...

## Basic Exploratory analysis

boxplot(mpg~am, mtcars)



aggregate(mpg ~ am, data = mtcars, mean)

## am mpg  
## 1 0 17.14737  
## 2 1 24.39231

## single var fit

FitAM <- lm(mpg ~ am, data = mtcars)  
summary(FitAM)

##   
## Call:  
## lm(formula = mpg ~ am, data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.3923 -3.0923 -0.2974 3.2439 9.5077   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.147 1.125 15.247 1.13e-15 \*\*\*  
## am1 7.245 1.764 4.106 0.000285 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.902 on 30 degrees of freedom  
## Multiple R-squared: 0.3598, Adjusted R-squared: 0.3385   
## F-statistic: 16.86 on 1 and 30 DF, p-value: 0.000285

only 35% of variance is explained

## multi variable fit

multiVarFit <- lm(mpg ~., data = mtcars)  
summary(multiVarFit)

##   
## Call:  
## lm(formula = mpg ~ ., data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5087 -1.3584 -0.0948 0.7745 4.6251   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 23.87913 20.06582 1.190 0.2525   
## cyl6 -2.64870 3.04089 -0.871 0.3975   
## cyl8 -0.33616 7.15954 -0.047 0.9632   
## disp 0.03555 0.03190 1.114 0.2827   
## hp -0.07051 0.03943 -1.788 0.0939 .  
## drat 1.18283 2.48348 0.476 0.6407   
## wt -4.52978 2.53875 -1.784 0.0946 .  
## qsec 0.36784 0.93540 0.393 0.6997   
## vs1 1.93085 2.87126 0.672 0.5115   
## am1 1.21212 3.21355 0.377 0.7113   
## gear4 1.11435 3.79952 0.293 0.7733   
## gear5 2.52840 3.73636 0.677 0.5089   
## carb2 -0.97935 2.31797 -0.423 0.6787   
## carb3 2.99964 4.29355 0.699 0.4955   
## carb4 1.09142 4.44962 0.245 0.8096   
## carb6 4.47757 6.38406 0.701 0.4938   
## carb8 7.25041 8.36057 0.867 0.3995   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.833 on 15 degrees of freedom  
## Multiple R-squared: 0.8931, Adjusted R-squared: 0.779   
## F-statistic: 7.83 on 16 and 15 DF, p-value: 0.000124

anova(FitAM,multiVarFit)

## Analysis of Variance Table  
##   
## Model 1: mpg ~ am  
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 30 720.9   
## 2 15 120.4 15 600.49 4.9874 0.001759 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

here using multivarfit we can reject the null hypothesis but their is significant change from fitam

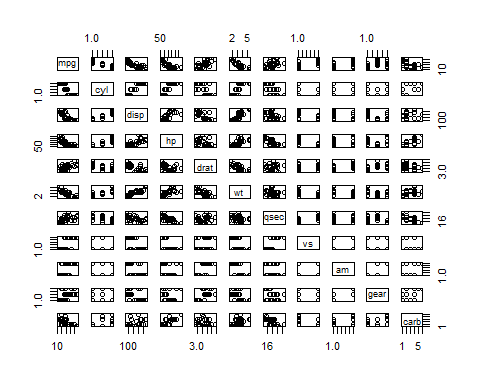
t.test(mtcars$mpg~mtcars$am)

##   
## Welch Two Sample t-test  
##   
## data: mtcars$mpg by mtcars$am  
## 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 in group 0 mean in group 1   
## 17.14737 24.39231

P-value less than 0.05,we reject the null hypothesis that there is no difference in MPG. This concludes that manual transmission works better than automatic.

## slope of variables with respect to mpg

pairs(mtcars)

 ## best fit

bestfit <-step(multiVarFit,direction = "both")

## Start: AIC=76.4  
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb  
##   
## Df Sum of Sq RSS AIC  
## - carb 5 13.5989 134.00 69.828  
## - gear 2 3.9729 124.38 73.442  
## - am 1 1.1420 121.55 74.705  
## - qsec 1 1.2413 121.64 74.732  
## - drat 1 1.8208 122.22 74.884  
## - cyl 2 10.9314 131.33 75.184  
## - vs 1 3.6299 124.03 75.354  
## <none> 120.40 76.403  
## - disp 1 9.9672 130.37 76.948  
## - wt 1 25.5541 145.96 80.562  
## - hp 1 25.6715 146.07 80.588  
##   
## Step: AIC=69.83  
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear  
##   
## Df Sum of Sq RSS AIC  
## - gear 2 5.0215 139.02 67.005  
## - disp 1 0.9934 135.00 68.064  
## - drat 1 1.1854 135.19 68.110  
## - vs 1 3.6763 137.68 68.694  
## - cyl 2 12.5642 146.57 68.696  
## - qsec 1 5.2634 139.26 69.061  
## <none> 134.00 69.828  
## - am 1 11.9255 145.93 70.556  
## - wt 1 19.7963 153.80 72.237  
## - hp 1 22.7935 156.79 72.855  
## + carb 5 13.5989 120.40 76.403  
##   
## Step: AIC=67  
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am  
##   
## Df Sum of Sq RSS AIC  
## - drat 1 0.9672 139.99 65.227  
## - cyl 2 10.4247 149.45 65.319  
## - disp 1 1.5483 140.57 65.359  
## - vs 1 2.1829 141.21 65.503  
## - qsec 1 3.6324 142.66 65.830  
## <none> 139.02 67.005  
## - am 1 16.5665 155.59 68.608  
## - hp 1 18.1768 157.20 68.937  
## + gear 2 5.0215 134.00 69.828  
## - wt 1 31.1896 170.21 71.482  
## + carb 5 14.6475 124.38 73.442  
##   
## Step: AIC=65.23  
## mpg ~ cyl + disp + hp + wt + qsec + vs + am  
##   
## Df Sum of Sq RSS AIC  
## - disp 1 1.2474 141.24 63.511  
## - vs 1 2.3403 142.33 63.757  
## - cyl 2 12.3267 152.32 63.927  
## - qsec 1 3.1000 143.09 63.928  
## <none> 139.99 65.227  
## + drat 1 0.9672 139.02 67.005  
## - hp 1 17.7382 157.73 67.044  
## - am 1 19.4660 159.46 67.393  
## + gear 2 4.8033 135.19 68.110  
## - wt 1 30.7151 170.71 69.574  
## + carb 5 13.0509 126.94 72.095  
##   
## Step: AIC=63.51  
## mpg ~ cyl + hp + wt + qsec + vs + am  
##   
## Df Sum of Sq RSS AIC  
## - qsec 1 2.442 143.68 62.059  
## - vs 1 2.744 143.98 62.126  
## - cyl 2 18.580 159.82 63.466  
## <none> 141.24 63.511  
## + disp 1 1.247 139.99 65.227  
## + drat 1 0.666 140.57 65.359  
## - hp 1 18.184 159.42 65.386  
## - am 1 18.885 160.12 65.527  
## + gear 2 4.684 136.55 66.431  
## - wt 1 39.645 180.88 69.428  
## + carb 5 2.331 138.91 72.978  
##   
## Step: AIC=62.06  
## mpg ~ cyl + hp + wt + vs + am  
##   
## Df Sum of Sq RSS AIC  
## - vs 1 7.346 151.03 61.655  
## <none> 143.68 62.059  
## - cyl 2 25.284 168.96 63.246  
## + qsec 1 2.442 141.24 63.511  
## - am 1 16.443 160.12 63.527  
## + disp 1 0.589 143.09 63.928  
## + drat 1 0.330 143.35 63.986  
## + gear 2 3.437 140.24 65.284  
## - hp 1 36.344 180.02 67.275  
## - wt 1 41.088 184.77 68.108  
## + carb 5 3.480 140.20 71.275  
##   
## Step: AIC=61.65  
## mpg ~ cyl + hp + wt + am  
##   
## Df Sum of Sq RSS AIC  
## <none> 151.03 61.655  
## - am 1 9.752 160.78 61.657  
## + vs 1 7.346 143.68 62.059  
## + qsec 1 7.044 143.98 62.126  
## - cyl 2 29.265 180.29 63.323  
## + disp 1 0.617 150.41 63.524  
## + drat 1 0.220 150.81 63.608  
## + gear 2 1.361 149.66 65.365  
## - hp 1 31.943 182.97 65.794  
## - wt 1 46.173 197.20 68.191  
## + carb 5 5.633 145.39 70.438

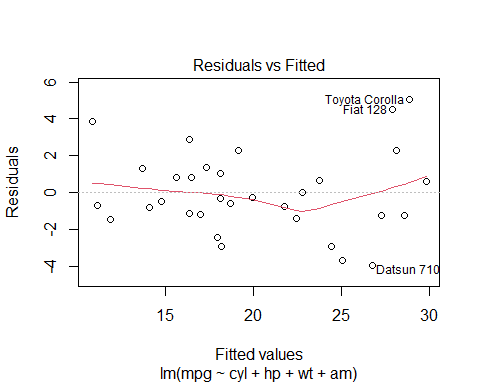
summary(bestfit)

##   
## Call:  
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.9387 -1.2560 -0.4013 1.1253 5.0513   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 33.70832 2.60489 12.940 7.73e-13 \*\*\*  
## cyl6 -3.03134 1.40728 -2.154 0.04068 \*   
## cyl8 -2.16368 2.28425 -0.947 0.35225   
## hp -0.03211 0.01369 -2.345 0.02693 \*   
## wt -2.49683 0.88559 -2.819 0.00908 \*\*   
## am1 1.80921 1.39630 1.296 0.20646   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.41 on 26 degrees of freedom  
## Multiple R-squared: 0.8659, Adjusted R-squared: 0.8401   
## F-statistic: 33.57 on 5 and 26 DF, p-value: 1.506e-10

In this bestfit model 86.59% of variance is explained. So, we choose this model for comparision.

## residual plot

plot(bestfit, which=1)



# summary

by this, we can clearly say that manual transmission is better than automatic. and, manual mpg is greater than automatic by 1.8 times. However, we need to consider other vaiables than transmission variable.