Fuel Efficiency Analysis

This study is to analyse a regression dataset. In particular, you will be looking at predicting the fuel efficiency of a car (in kilometers per litre) based on characteristics of the car and its engine. This is clearly an important and useful problem. The dataset fuel2017-20.csv contains n = 2, 000 observations on p = 9 predictors obtained from actual fuel efficiency tables for car models available for sale during the years 2017 through to 2020. The target is the fuel efficiency of the car measured in kilometers per litre. The higher this score, the better the fuel efficiency of the car. The data dictionary for this dataset is given at the end.

1) Fit a multiple linear model to the fuel efficiency data using R. Using the results of fitting the linear model, which predictors do you think are possibly associated with fuel efficiency, and why? Which three variables appear to be the strongest predictors of fuel efficiency, and why?

R code used:

#4.1 fueldata <- read.csv("fuel2017-

20.csv")

mult_fit <- lm(Comb.FE ~., fueldata) summary(mult_fit)

```
 Call: lm(formula = Comb.FE ~ ., data =
fueldata)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2229 -0.9985 -0.0975  0.7149 11.4355

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -2.003e+02  7.241e+01  -2.766  0.00573 **
Model.Year                1.074e-01  3.588e-02   2.993  0.00279 **
Eng.Displacement         -1.287e+00  8.674e-02 -14.832  < 2e-16 ***
No.Cylinders              2.569e-03  5.767e-02   0.045  0.96447
AspirationOT             -2.471e-01  6.343e-01  -0.390  0.69692
AspirationSC             -1.015e+00  1.995e-01  -5.089 3.94e-07 ***
AspirationTC             -1.268e+00  1.085e-01 -11.685  < 2e-16 ***
AspirationTS             -1.183e+00  4.215e-01  -2.807  0.00506 **
No.Gears                 -1.745e-01  2.534e-02  -6.888 7.58e-12 ***
Lockup.Torque.ConverterY -7.859e-01  9.506e-02  -8.267 2.48e-16 ***
Drive.SysA               -3.829e-02  1.294e-01  -0.296  0.76725
Drive.SysF                1.512e+00  1.438e-01  10.511  < 2e-16 ***
Drive.SysP               -4.435e-01  2.427e-01  -1.827  0.06781 .
Drive.SysR                9.319e-02  1.243e-01   0.750  0.45349
Max.Ethanol              -6.993e-03  2.490e-03  -2.808  0.00503 **
Fuel.TypeGM               5.696e-01  3.752e-01   1.518  0.12913
```

```
Fuel.TypeGP              5.024e-01  1.163e-01   4.321 1.63e-05 ***
Fuel.TypeGPR             2.066e-01  1.199e-01   1.723  0.08500 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.619 on 1982 degrees of freedom
Multiple R-squared:  0.6639, Adjusted R-squared:  0.661  F-statistic:
230.3 on 17 and 1982 DF,  p-value: < 2.2e-16
```

As from the table above, Eng.displacement, Aspiration and Drive would be three variables being the strongly associated with fuel efficiency. While, other variables such as Model.year, max.ethanol had the moderative effect on fuel efficiency.

2) Would your assessment of which predictors are associated change if you used the Bonferroni procedure with $\alpha = 0.05$?

Question-4 :-

(2.) According to Bonferroni Procedure,
we should reject the null hypothesis, [null hypothesis : Predictor is unassociated.]

only if $p\text{-value} < \frac{\alpha}{p}$ where $p = 10$ in this case.

$\downarrow$

$\rightarrow$ $p\text{-value} < \frac{0.05}{10}$

$\Rightarrow$ $p\text{-value} < 0.005$

* As after checking for p-values of all the variables, we would need to check the p-values of all the variables to compare with 0.005 to believe that a predictor is associated with fuel efficiency. Max Ethenol and No of Gears variables don't pass that test so they don't consider to be associated with fuel Efficiency

3) Describe what effect the year of manufacture (Model.Year) appears to have on the mean fuel efficiency. Describe the effect that the number of gears (No.Gears) variable has on the mean fuel efficiency of the car

(3.) → After fitting the simple linear model,

→ predicting fuel efficiency :

$$\rightarrow E[\text{Fuel\_efficiency}_i] = -2.0025 \times 10^{2} + \underline{1.074 \times 10^{-1} \times \text{Model yr}_i}$$
$$- 1.2865 \times \text{Eng.dis}_i + 2.569 \times 10^{-3} \times \text{No cyl}_i - 2.471 \times \text{Aspi OT}_i$$
$$- 1.015 \times \text{Aspi SC}_i - 1.268 \times \text{Aspi TC}_i - 1.1830 \times \text{Aspi TS}_i -$$
$$\underline{10^{-1} \times 1.7452 \times \text{No.gears}_i} - 7.8596 \times 10^{-1} \times \text{Torque.con}_i - 3.829 \times 10^{-1} \times \text{drive SA}$$
$$+ 1.517 \times \text{drive SF}_i - 4.434 \times 10^{-1} \times \text{drive SP}_i + 9.3192 \times 10^{-2} \times \text{drive SR}_i$$
$$- 6.992 \times 10^{-3} \text{ max.eth}_i + 5.6962 \times 10^{-1} \times \text{Fuel\_Type GM}_i$$
$$+ 5.023 \times 10^{-1} \text{ Fuel.GP}_i + 2.0661 \times 10^{-1} \text{ Fuel.GPR}_i$$

⇒ This model says,

- for every year changes (by 1), the fuel efficiency will be risen by 0.1074 km/l.

- For each year increasing [0-10], the fuel efficiency will be decreased by 0.17452 km/l.

R code used:

```
#4.3 mult_fit$coefficients

#4.4 mult_fit_bic <- step(mult_fit, k =
log(length(fueldata$Comb.FE)))  summary(mult_fit_bic)
mult_fit_bic$coefficients
```

4) Use the stepwise selection procedure with the BIC penalty to prune out potentially unimportant variables. Write down the final regression equation obtained after pruning.

(4.)

→ The final regression equation after BIC penalty,

→ $E[\text{Fuel\_efficiency}_i] = -2.0965 \times 10^2 + 0.1119 \times [\text{model yr}_i]$

$- 1.2534 \times [\text{Eng\_displace}_i] - 0.1014 \times [\text{AspiOT}_i]$

$- 0.7208 \times [\text{AspiSC}]_i - 1.0926 \times [\text{AspiTC}_i] - 1.1003 \times [\text{AspiTS}]_i$

$- 0.16063 \times [\text{No. Gears}_i] - 0.7999 \times [\text{Torque con}]_i$

$+ 0.071841 \times [\text{drive SA}_i] + 1.5447 \times [\text{Drive SF}]_i$

$- 0.54536 \times [\text{drive SP}_i] + 0.1688 \times [\text{Drive SR}_i]$

$- 0.008183 \times [\text{max\_Eth}]_i$

model year - increase ↑

drive system All wheel, Front wheel, rare wheel ↑
(good condition)

5) If we wanted to improve the fuel efficiency of our car, what does this BIC model suggest we could do?

As it can be seen below in the image, to increase fuel efficiency, we have to make sure we use the car launched in the latest year and newer the car, better the fuel efficiency. We also have to make sure that front, back and other wheels are in the good condition such that fuel doesn't get too wasted so there are the factors need to be considered.

6) Imagine that you are looking for a new car to buy to replace your existing car. Load the dataset fuel2017-20.test.csv. The characteristics of the new car that you are looking at are given by the first row of this dataset.

(a) Use your BIC model to predict the mean fuel efficiency for this new car. Provide a 95% confidence interval for this prediction.

R code used: #4.6 fueldatatesting <- read.csv("fuel2017-

20.test.csv", header=TRUE) mult_fit_testing <- lm(Comb.FE ~.,

fueldatatesting) mult_fit_testing_bic <- step(mult_fit_testing, k =

log(length(fueldatatesting$Comb.FE)))

predict(mult_fit_testing_bic , interval = "confidence", level = 0.95)

mean= 12.34 km/l

cli= (7.35,17.33)

b) The current car that you own has a mean fuel efficiency of 8.5km/l (measured over the life time of your ownership). Does your model suggest that the new car will have better fuel efficiency than your current car?

As the current efficienty is 8.5 km/l which is in the confidence interval found as above so the fitted model may have better fuel efficiency than the previous one.

Variable name	Description	Values
`Model.Year`	Year of sale	$2017 - 2020$
`Eng.Displacement`	Engine Displacement (litres, l)	$0.9 - 8.4$
`No.Cylinders`	Number of Cylinders	$3 - 16$
`Aspiration`	Engine Aspiration (Oxygen intake)	`N`: Naturally*
		`OT`: Other
		`SC`: Supercharged
		`TC`: Turbocharged
		`TS`: Turbo+supercharged
`No.Gears`	Number of Gears	$1 - 10$
`Lockup.Torque.Converter`	Lockup torque converter present?	`N`* and `Y`
`Drive.Sys`	Drive System	`4`*: 4-wheel drive
		`A`:All-wheel
		`F`:Front-wheel
		`P`:Part-time 4-wheel
		`R`:Rear-wheel
`Max.Ethanol`	Maximum % of Ethanol allowed	$10 - 85$
`Fuel.Type`	Type of Fuel	`G`*: Regular Unleaded
		`GM`: Mid-grade Unleaded Recommended
		`GP`: Premium Unleaded Recommended
		`GPR`: Premium Unleaded Required
`Comb.FE`	Fuel Efficiency (km/l)	$4.974 - 26.224$