OUTPUT

## Output of the Generalized model##

|  |
| --- |
| fit\_reg <- lm(e ~ ., data=f1)  > summary(fit\_reg) # show results  Call:  lm(formula = e ~ ., data = f1)  Residuals:  Min 1Q Median 3Q Max  -4.6542 -0.4909 -0.0006 0.4282 13.3884  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 6.727e-06 1.168e-03 0.006 0.995  car\_at 9.598e-02 1.172e-03 81.929 <2e-16 \*\*\*  car\_engine 5.828e-01 1.217e-03 478.907 <2e-16 \*\*\*  car\_body 4.295e-01 1.266e-03 339.242 <2e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.6962 on 355176 degrees of freedom  Multiple R-squared: 0.5154, Adjusted R-squared: 0.5154  F-statistic: 1.259e+05 on 3 and 355176 DF, p-value: < 2.2e-16 |
|  |
| |  | | --- | |  | |

print(rmse1)

[1] 0.4846261

## Output fuel type based models##  
## Petrol ##  
fit\_petrol <- lm(e ~ car\_at + car\_engine + car\_body, data=petrol)

> summary(fit\_petrol)

Call:

lm(formula = e ~ car\_at + car\_engine + car\_body, data = petrol)

Residuals:

Min 1Q Median 3Q Max

-5.2844 -0.2953 -0.0303 0.2786 10.7572

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.372943 0.001725 216.2 <2e-16 \*\*\*

car\_at 0.099632 0.001533 65.0 <2e-16 \*\*\*

car\_engine 0.651782 0.001150 566.8 <2e-16 \*\*\*

car\_body 0.387746 0.002117 183.2 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5647 on 148978 degrees of freedom

Multiple R-squared: 0.7245, Adjusted R-squared: 0.7245

F-statistic: 1.306e+05 on 3 and 148978 DF, p-value: < 2.2e-16

|  |
| --- |
| p1 <- predict(fit\_petrol, petrol[,3:5])  > p1[1:10]  1 2 3 4 5 6 7 8 9 10  0.4457128 4.9054136 0.6512342 0.5416942 1.7284913 1.4773222 0.6275528 0.7090208 0.8012149 4.8487733  > rmse\_petrol <- mean((petrol$e - p1)^2)  > print(rmse\_petrol)  [1] 0.3188309 |
|  |
| |  | | --- | | > | |

## Diesel##  
diesel <- sqldf("SELECT \* FROM rg WHERE Ft = 'diesel'")

> fit\_diesel <- lm(e ~ car\_at + car\_engine + car\_body, data=diesel)

> summary(fit\_diesel)

Call:

lm(formula = e ~ car\_at + car\_engine + car\_body, data = diesel)

Residuals:

Min 1Q Median 3Q Max

-3.5428 -0.3267 -0.0675 0.2894 7.6157

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.337448 0.001207 -279.60 <2e-16 \*\*\*

car\_at 0.107298 0.001109 96.73 <2e-16 \*\*\*

car\_engine 0.449364 0.001813 247.91 <2e-16 \*\*\*

car\_body 0.685071 0.001210 565.98 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.51 on 200592 degrees of freedom

Multiple R-squared: 0.6612, Adjusted R-squared: 0.6612

F-statistic: 1.305e+05 on 3 and 200592 DF, p-value: < 2.2e-16

> p2 <- predict(fit\_diesel, diesel[,3:5])

> p2[1:10]

1 2 3 4 5 6 7 8 9 10

1.5471327 -0.6867490 -0.5871644 0.9102775 -0.3579458 -0.8349877 0.1684740 3.5232692 1.7712591 2.4617058

> rmse\_diesel <- mean((diesel$e - p2)^2)

> print(rmse\_diesel)

[1] 0.260089

## Biodiesel##  
biodiesel <- sqldf("SELECT \* FROM rg WHERE Ft = 'biodiesel'")

> fit\_biodiesel <- lm(e ~ car\_at + car\_engine + car\_body, data=biodiesel)

> summary(fit\_biodiesel)

Call:

lm(formula = e ~ car\_at + car\_engine + car\_body, data = biodiesel)

Residuals:

1 2 3 4 5

-0.8974 0.3252 0.2189 0.2371 0.1161

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.3753 1.6031 0.858 0.549

car\_at 7.2855 5.9769 1.219 0.437

car\_engine -1.9610 3.5950 -0.545 0.682

car\_body -0.8879 2.7514 -0.323 0.801

Residual standard error: 1.014 on 1 degrees of freedom

Multiple R-squared: 0.6933, Adjusted R-squared: -0.2268

F-statistic: 0.7535 on 3 and 1 DF, p-value: 0.6672

> p3 <- predict(fit\_biodiesel, biodiesel[,3:5])

> p3[1:10]

1 2 3 4 5 <NA> <NA> <NA> <NA> <NA>

0.3373864 1.4547982 -0.1789435 -0.4571001 -0.1961411 NA NA NA NA NA

> rmse\_biodiesel <- mean((biodiesel$e - p3)^2)

> print(rmse\_biodiesel)

[1] 0.20574

## e85##  
e85 <- sqldf("SELECT \* FROM rg WHERE Ft = 'e85'")

> fit\_e85 <- lm(e ~ car\_at + car\_engine + car\_body, data=e85)

> summary(fit\_e85)

Call:

lm(formula = e ~ car\_at + car\_engine + car\_body, data = e85)

Residuals:

Min 1Q Median 3Q Max

-0.68862 -0.23173 0.06405 0.26383 2.35594

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.54878 0.02370 23.157 < 2e-16 \*\*\*

car\_at 0.13620 0.03185 4.276 2.74e-05 \*\*\*

car\_engine 0.64661 0.02552 25.338 < 2e-16 \*\*\*

car\_body 0.76430 0.05366 14.242 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3335 on 243 degrees of freedom

Multiple R-squared: 0.8974, Adjusted R-squared: 0.8962

F-statistic: 708.7 on 3 and 243 DF, p-value: < 2.2e-16

> p4 <- predict(fit\_e85, e85[,3:5])

> p4[1:10]

1 2 3 4 5 6 7 8 9 10

0.44964523 0.86257647 3.18058099 0.24658624 0.27697297 0.05507486 0.84274193 0.84274193 0.93780304 0.84274193

> rmse\_e85 <- mean((e85$e - p4)^2)

> print(rmse\_e85)

[1] 0.1094011

## LPG ##  
lpg <- sqldf("SELECT \* FROM rg WHERE Ft = 'lpg'")

> fit\_lpg <- lm(e ~ car\_at + car\_engine + car\_body, data=lpg)

> summary(fit\_lpg)

Call:

lm(formula = e ~ car\_at + car\_engine + car\_body, data = lpg)

Residuals:

Min 1Q Median 3Q Max

-2.2604 -0.3376 -0.0283 0.2680 12.9990

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.53521 0.02230 24.00 <2e-16 \*\*\*

car\_at 0.14858 0.01345 11.05 <2e-16 \*\*\*

car\_engine 0.80537 0.02764 29.13 <2e-16 \*\*\*

car\_body 0.39624 0.01907 20.78 <2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5394 on 1918 degrees of freedom

Multiple R-squared: 0.4024, Adjusted R-squared: 0.4014

F-statistic: 430.5 on 3 and 1918 DF, p-value: < 2.2e-16

> p5 <- predict(fit\_lpg, lpg[,3:5])

> p5[1:10]

1 2 3 4

-0.47223605 0.14820128 -0.51416860 0.11193494

5 6 7 8

0.11998863 0.08326106 -0.49413844 -0.51416860

9 10

-0.48997325 -0.48997325

> rmse\_lpg <- mean((lpg$e - p5)^2)

> print(rmse\_lpg)

[1] 0.2903638

## Biomithane##  
summary(fit\_ng\_biomethane)

Call:

lm(formula = e ~ car\_at + car\_engine + car\_body, data = ng\_biomethane)

Residuals:

Min 1Q Median 3Q Max

-2.5348 -0.2689 -0.0496 0.2044 1.6172

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.330753 0.020922 -15.81 <2e-16 \*\*\*

car\_at 0.067871 0.007847 8.65 <2e-16 \*\*\*

car\_engine 0.423881 0.027406 15.47 <2e-16 \*\*\*

car\_body 0.692633 0.012783 54.18 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4294 on 1366 degrees of freedom

Multiple R-squared: 0.71, Adjusted R-squared: 0.7093

F-statistic: 1115 on 3 and 1366 DF, p-value: < 2.2e-16

> p6 <- predict(fit\_ng\_biomethane, ng\_biomethane[,3:5])

> p6[1:10]

1 2 3 4 5

-0.8187739 -0.7490530 -0.7421267 -0.8187739 -0.7517405

6 7 8 9 10

-0.7352003 -0.7186601 -0.8346354 -0.8346354 -0.7718408

> rmse\_ng\_biomethane <- mean((ng\_biomethane$e - p6)^2)

> print(rmse\_ng\_biomethane)

[1] 0.1838843