

CE18B118-CE3030-Assignment-3.R

Monish Kumar

2020-09-29

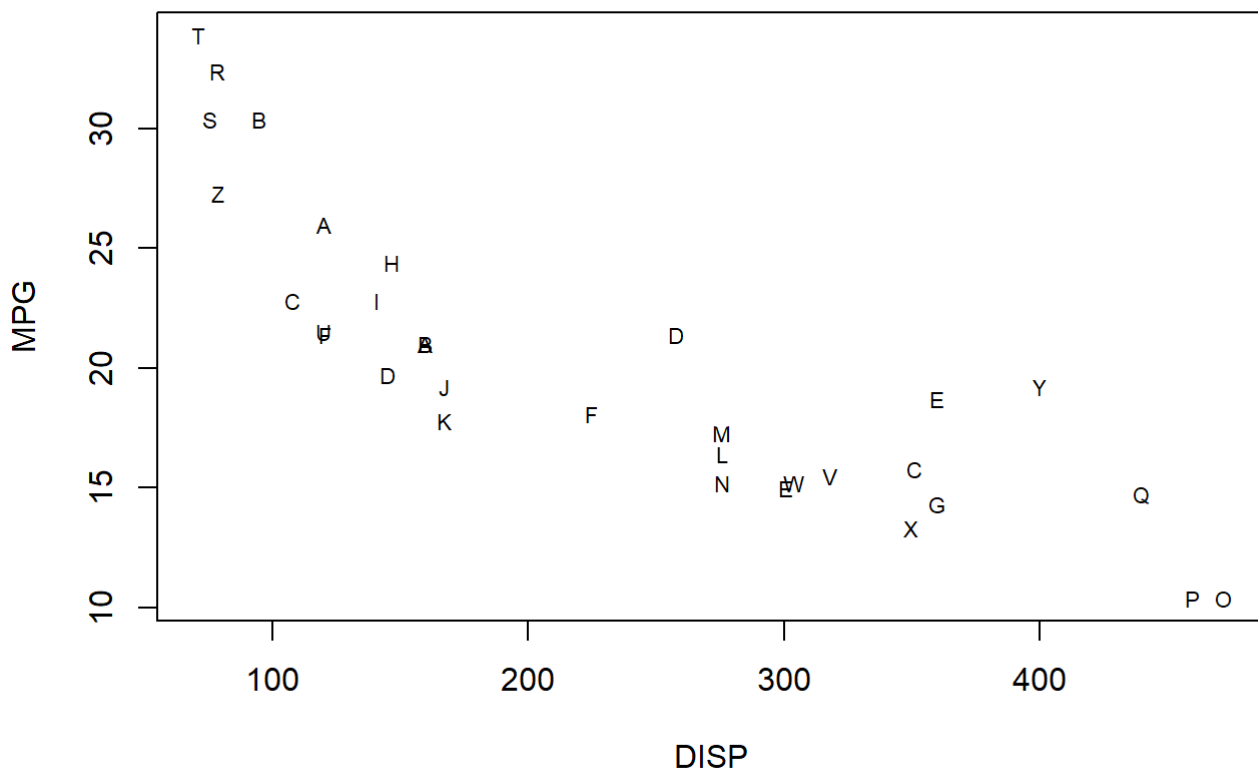
Question No. 1

```
attach(mtcars)
```

```
# Figure 1
```

```
plot(displacement, mpg, pch='.', col = 'white', xlab = "DISP", ylab = "MPG")
```

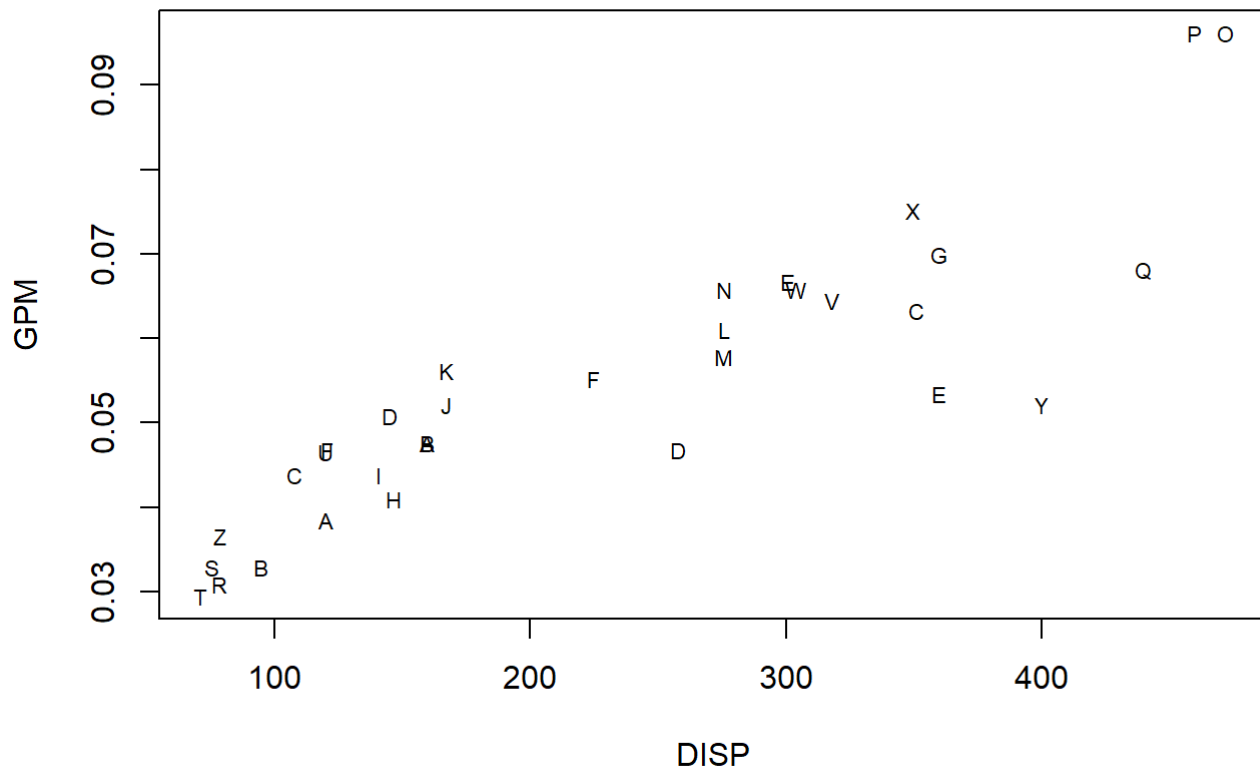
```
text(displacement, mpg, rep(LETTERS, length.out = 32), cex = 0.7)
```



```
# Figure 2
```

```
plot(displacement, 1/mpg, pch='.', col = 'white', xlab = "DISP", ylab = "GPM")
```

```
text(displacement, 1/mpg, rep(LETTERS, length.out = 32), cex = 0.7)
```



```
# Figure 3
gpm = 1/mpg
hp_per_wt = hp/wt

model_1 = lm(gpm ~ wt)
model_2 = lm(hp_per_wt ~ wt)

plot(model_2$residuals, model_1$residuals, pch='.', col = 'white', xlab = " HP/WT . WT", ylab = " GPM . WT")
text(model_2$residuals, model_1$residuals, rep(LETTERS,length.out = 32), cex = 0.7)
```

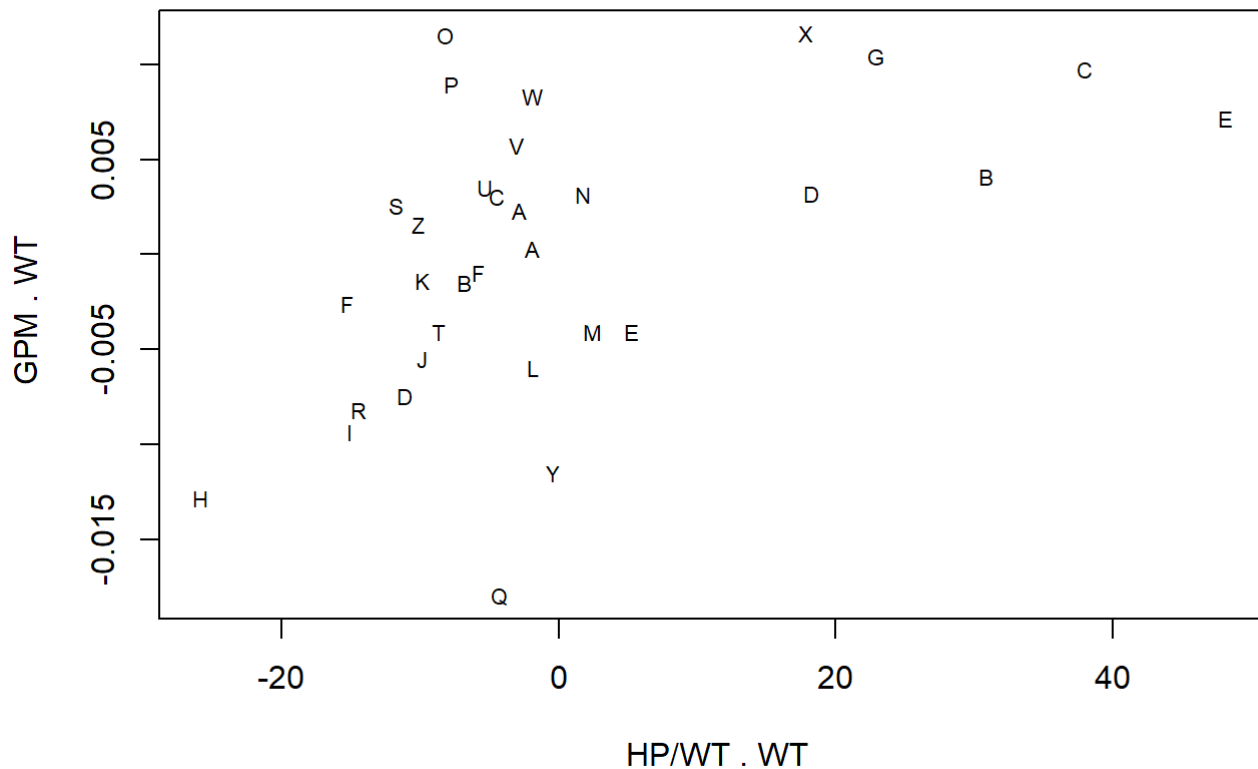


Figure 4 and Table 2

gpm = 100/mpg

hp_per_wt = hp/wt

gpm_model = lm(gpm ~ wt + hp_per_wt)

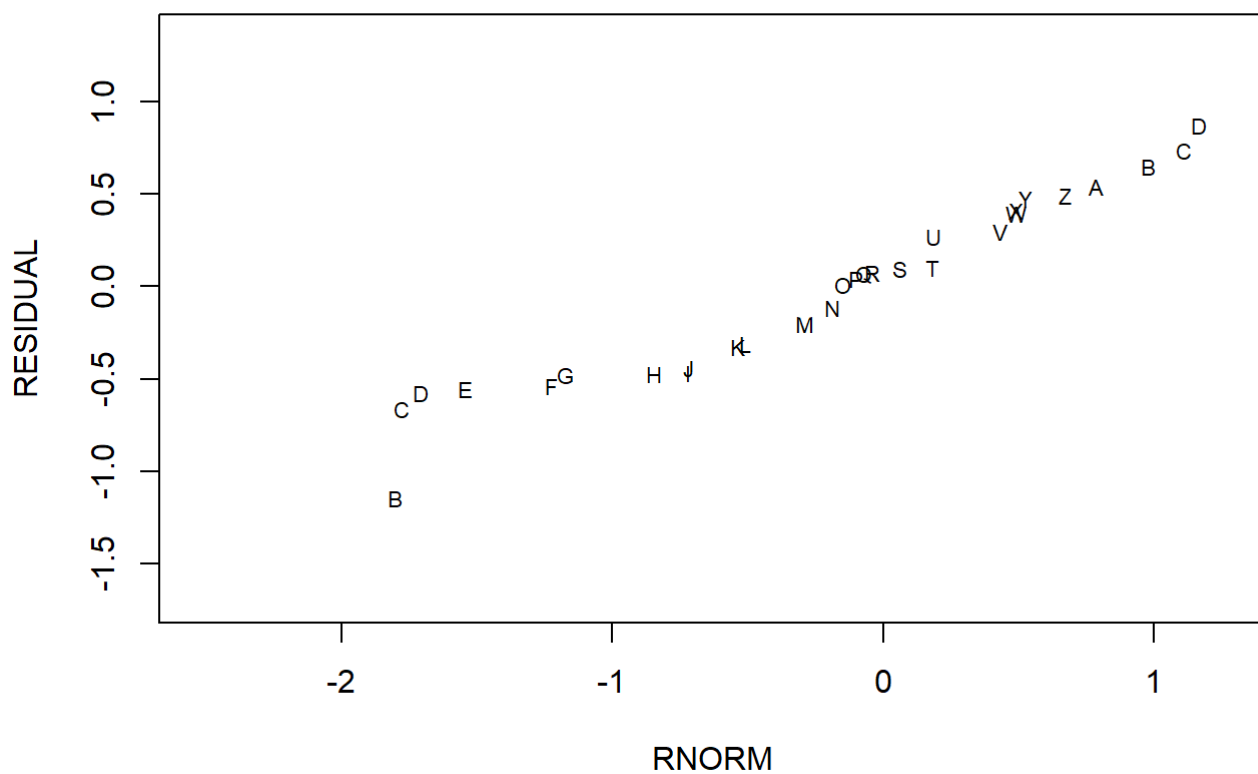
summary(gpm_model)

```
##
## Call:
## lm(formula = gpm ~ wt + hp_per_wt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.69714 -0.46822  0.05312  0.42744  1.35097
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.401534   0.512044  -0.784  0.43929
## wt          1.472176   0.121554  12.111 7.24e-13 ***
## hp_per_wt    0.023997   0.007302   3.286 0.00266 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6612 on 29 degrees of freedom
## Multiple R-squared:  0.8484, Adjusted R-squared:  0.8379
## F-statistic: 81.13 on 2 and 29 DF,  p-value: 1.322e-12
```

```
shapiro.test(gpm_model$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data: gpm_model$residuals
## W = 0.98071, p-value = 0.82
```

```
plot(sort(rnorm(length(gpm))),sort(gpm_model$residuals), pch='.', col = 'white', xlab = "RNORM
M", ylab = "RESIDUAL")
text(sort(rnorm(length(gpm))),sort(gpm_model$residuals), rep(LETTERS,length.out = 32), cex =
0.7)
```



```
# Model with MPG as dependent variable:

wt_inverse = 1/wt
wt_per_hp = wt/hp
disp_inverse = 1/disp
# All other variables are highly correlated with wt/wt_inverse

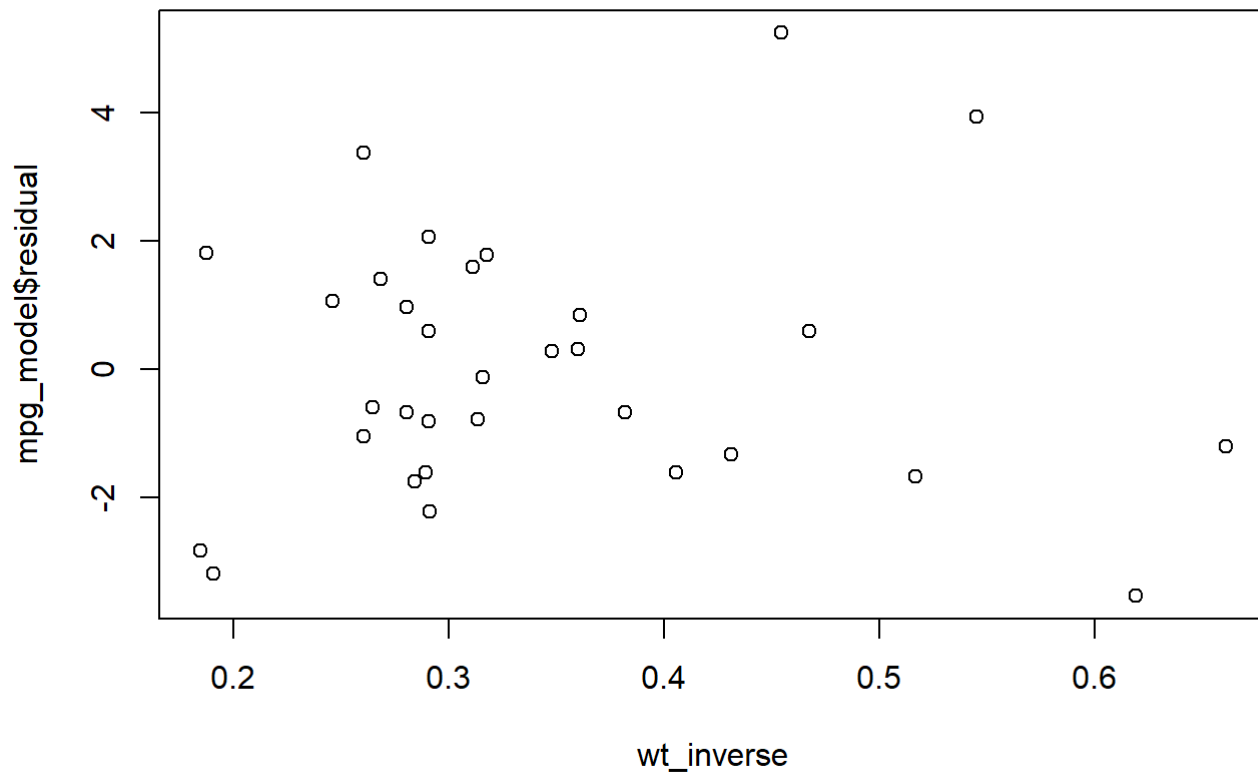
mpg_model = lm(mpg ~ wt_inverse + wt_per_hp )
summary(mpg_model)
```

```
##
## Call:
## lm(formula = mpg ~ wt_inverse + wt_per_hp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5407 -1.4045 -0.3689  1.1423  5.2369
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.923      1.589   -0.581    0.566
## wt_inverse     44.435      3.215   13.819 2.73e-14 ***
## wt_per_hp     236.657     48.502    4.879 3.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.091 on 29 degrees of freedom
## Multiple R-squared:  0.8874, Adjusted R-squared:  0.8796
## F-statistic: 114.3 on 2 and 29 DF,  p-value: 1.765e-14
```

```
shapiro.test(mpg_model$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data:  mpg_model$residuals
## W = 0.97233, p-value = 0.5659
```

```
plot(wt_inverse, mpg_model$residual)      # to check if mean = 0
```



```
plot(mpg, mpg_model$residuals)           # to check if variance constant
```

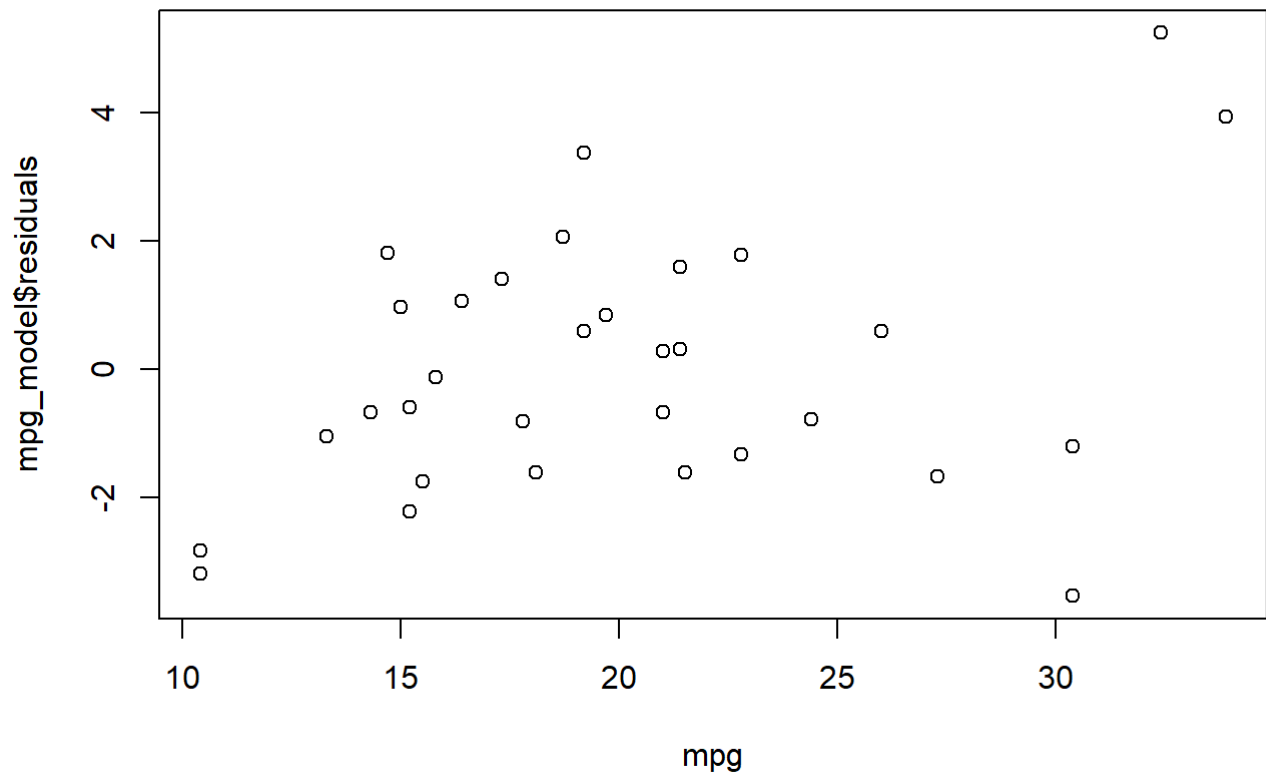
```
##### Question No. 2 #####
```

```
library(mlogit)
```

```
## Loading required package: dfidx
```

```
##
## Attaching package: 'dfidx'
```

```
## The following object is masked from 'package:stats':
##
##   filter
```



```

swissmetro = read.table('swissmetro-renamed.dat',header = TRUE)
swissmetro = swissmetro[swissmetro$CHOICE != 0,]
swissmetro = swissmetro[swissmetro$AGE != 6,]

swissmetro$CHOICE[swissmetro$CHOICE == 1] = 'TRAIN'
swissmetro$CHOICE[swissmetro$CHOICE == 2] = 'SM'
swissmetro$CHOICE[swissmetro$CHOICE == 3] = 'CAR'

# Creating dummy variables:

# LUGGAGE

swissmetro$LUGGAGE_none[swissmetro$LUGGAGE == 0] = 1
swissmetro$LUGGAGE_none[swissmetro$LUGGAGE != 0] = 0

swissmetro$LUGGAGE_one_piece[swissmetro$LUGGAGE == 1] = 1
swissmetro$LUGGAGE_one_piece[swissmetro$LUGGAGE != 1] = 0

# AGE

swissmetro$is_Old[swissmetro$AGE >= 5] = 1
swissmetro$is_Old[swissmetro$AGE < 5] = 0

# TICKET

swissmetro$is_Annual_Ticket[swissmetro$TICKET == 6 | swissmetro$TICKET == 7] = 1
swissmetro$is_Annual_Ticket[swissmetro$TICKET != 6 & swissmetro$TICKET != 7] = 0

# HEADWAY_diff (TRAIN & SM)

swissmetro$HEADWAY_diff = swissmetro$HE.SM - swissmetro$HE.TRAIN

# Defining Model:

swissmetro.wide = mlogit.data(swissmetro, varying = c(19,26,22, 20,27,23),
                             shape = "wide", choice = "CHOICE")

summary(mlogit(CHOICE ~ TT + CO | SURVEY +
              LUGGAGE_none + LUGGAGE_one_piece +
              is_Old + is_Annual_Ticket + GA + HEADWAY_diff
              , data = swissmetro.wide, reflevel = "CAR"))

```



```
##
## Call:
## mlogit(formula = CHOICE ~ TT + CO | SURVEY + LUGGAGE_none + LUGGAGE_one_piece +
##       is_Old + is_Annual_Ticket + GA + HEADWAY_diff, data = swissmetro.wide,
##       reflevel = "CAR", method = "nr")
##
## Frequencies of alternatives:choice
##      CAR      SM   TRAIN
## 0.28758 0.58039 0.13203
##
## nr method
## 6 iterations, 0h:0m:1s
## g'(-H)^-1g = 1.78E-07
## gradient close to zero
##
## Coefficients :
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept):SM      2.0093e+00  2.1288e-01   9.4386 < 2.2e-16 ***
## (Intercept):TRAIN    1.4831e+00  2.4829e-01   5.9732 2.326e-09 ***
## TT                  -9.4241e-03  3.8754e-04 -24.3179 < 2.2e-16 ***
## CO                  -8.7919e-04  7.2796e-05 -12.0774 < 2.2e-16 ***
## SURVEY:SM           -3.1166e+00  1.0227e-01 -30.4735 < 2.2e-16 ***
## SURVEY:TRAIN        -4.4178e+00  1.2857e-01 -34.3613 < 2.2e-16 ***
## LUGGAGE_none:SM      9.4299e-01  2.2762e-01   4.1429 3.430e-05 ***
## LUGGAGE_none:TRAIN    1.0226e+00  2.6737e-01   3.8244 0.0001311 ***
## LUGGAGE_one_piece:SM  6.1718e-01  2.2351e-01   2.7614 0.0057559 **
## LUGGAGE_one_piece:TRAIN 7.5353e-01  2.5344e-01   2.9732 0.0029472 **
## is_Old:SM           -6.0781e-01  9.8950e-02  -6.1426 8.118e-10 ***
## is_Old:TRAIN         1.0674e+00  1.2234e-01   8.7245 < 2.2e-16 ***
## is_Annual_Ticket:SM  8.3101e-01  1.8883e-01   4.4009 1.078e-05 ***
## is_Annual_Ticket:TRAIN 1.6236e+00  2.6784e-01   6.0618 1.346e-09 ***
## GA:SM               3.3977e+00  4.2473e-01   7.9995 1.332e-15 ***
## GA:TRAIN            2.6485e+00  4.2092e-01   6.2921 3.132e-10 ***
## HEADWAY_diff:SM     -1.0774e-03  6.3953e-04  -1.6847 0.0920445 .
## HEADWAY_diff:TRAIN    7.2629e-03  1.0348e-03   7.0185 2.243e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -7944
## McFadden R^2: 0.21216
## Likelihood ratio test : chisq = 4278.5 (p.value = < 2.22e-16)
```