PSTAT126_Lab4

R Markdown

https://rmarkdown.rstudio.com/index.html

Chap 3: Multiple Linear Regression

Model

$E[Y|X] = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p,\ Var(Y|X) = \sigma^2$;
Unknown Parameters: βs, σ^2;

Matrix Notation

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & ... & x_{1p} \\ 1 & x_{21} & ... & x_{2p} \\ ... & & & \\ 1 & x_{n1} & ... & x_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ ... \\ \beta_p \end{pmatrix}$$

. So we can write the mean function in matrix terms as $E[Y|X] = X\beta$.
Errors: $e_i = y_i - \mathbf{x}_i'\beta$ and $\mathbf{e} = (e_1, ..., e_n)'$, $\mathbf{x}_i'$ is X's i^{th} row;
Normality assumption: $\mathbf{e}|X \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$;

OLS Estimators

Minimizing the residual sum of squares function $RSS((\beta)) = \sum(y_i - \mathbf{x}_i'\beta)^2 = (Y - X\beta)'(Y - X\beta)$, we get the OLS estimates:
$$\hat{\beta} = (X'X)^{-1}X'Y,$$
if $(X'X)^{-1}$ exists.

```
library(alr4)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

Fuel Consumption

Textbook example. Data introduced in page 15.
Goal: how fuel consumption varies over different US areas?

```
#?fuel2001
fuel <- transform(fuel2001,
    Dlic=1000 * Drivers/Pop,
    Fuel=1000 * FuelC/Pop,
    logMiles=log(Miles),
    Income=Income/1000)
fuel <- subset(fuel, select = -c(Drivers, FuelC, MPC, Pop, Miles))
head(fuel)
```

```
##    Income  Tax      Dlic      Fuel  logMiles
## AL 23.471 18.0 1031.3801 690.2644 11.455720
## AK 30.064  8.0 1031.6411 514.2792  9.519882
## AZ 25.578 18.0  908.5972 621.4751 10.919533
## AR 22.257 21.7  946.5706 655.2927 11.494069
## CA 32.275 18.0  844.7033 573.9129 12.036298
## CO 32.949 22.0  989.6062 616.6115 11.360403
```

```
#Textbook page 58
nrow(fuel) #N=51
```

```
## [1] 51
```

```
summary(fuel)
```

```
##      Income          Tax            Dlic            Fuel
##  Min.   :20.99   Min.   : 7.50   Min.   : 700.2   Min.   :317.5
##  1st Qu.:25.32   1st Qu.:18.00   1st Qu.: 864.1   1st Qu.:575.0
##  Median :27.87   Median :20.00   Median : 909.1   Median :626.0
##  Mean   :28.40   Mean   :20.15   Mean   : 903.7   Mean   :613.1
##  3rd Qu.:31.21   3rd Qu.:23.25   3rd Qu.: 943.0   3rd Qu.:666.6
##  Max.   :40.64   Max.   :29.00   Max.   :1075.3   Max.   :842.8
##     logMiles
##  Min.   : 7.336
##  1st Qu.:10.507
##  Median :11.276
##  Mean   :10.914
##  3rd Qu.:11.634
##  Max.   :12.614
```

```
#Textbook page 17 and 59
pairs(Fuel~Tax + Dlic + Income + logMiles, data=fuel)
```
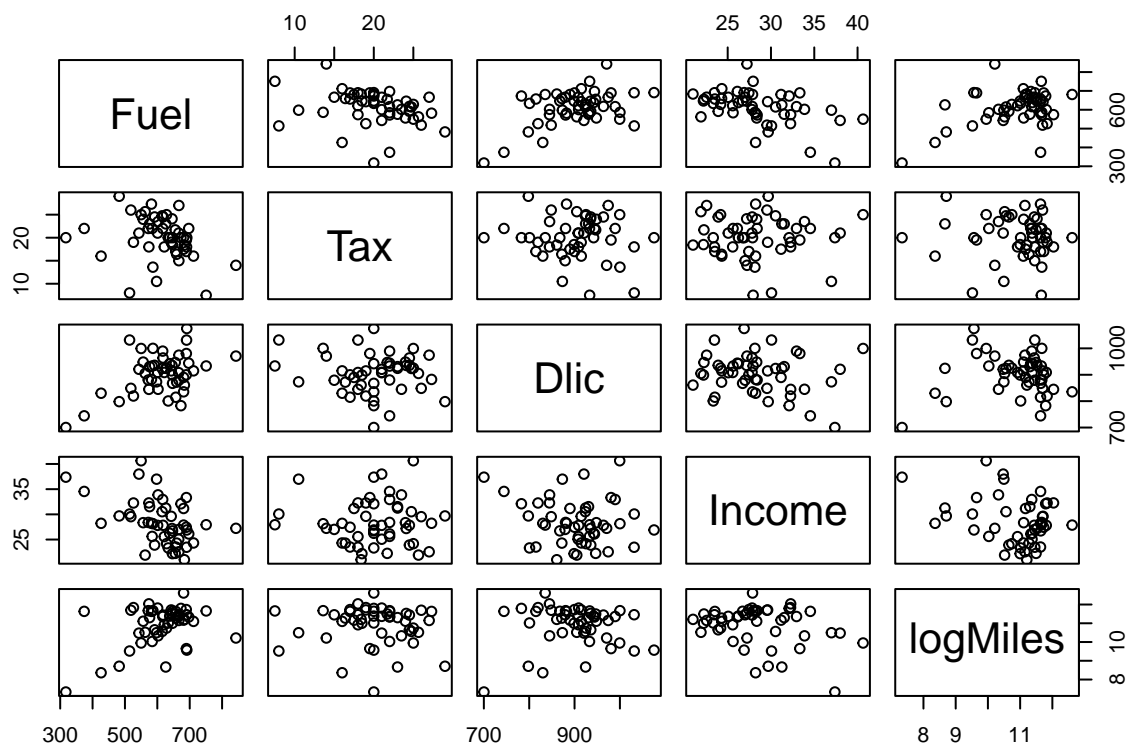
```
#plot(fuel)
```

```
#Textbook page 59 Table 3.2
cor(fuel)
```

```
##               Income         Tax        Dlic       Fuel    logMiles
## Income     1.00000000 -0.01068494 -0.17596063 -0.4644050 -0.29585136
## Tax       -0.01068494  1.00000000 -0.08584424 -0.2594471 -0.04373696
## Dlic      -0.17596063 -0.08584424  1.00000000  0.4685063  0.03059068
## Fuel      -0.46440498 -0.25944711  0.46850627  1.0000000  0.42203233
## logMiles  -0.29585136 -0.04373696  0.03059068  0.4220323  1.00000000
```

Model: Fuel~Tax+Dlic+Income+log(Miles).

```
fit <- lm(Fuel~Tax+Dlic+Income+logMiles, data = fuel)
summary(fit)
```

```
##
## Call:
## lm(formula = Fuel ~ Tax + Dlic + Income + logMiles, data = fuel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -163.145  -33.039    5.895   31.989  183.499
```

```
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 154.1928   194.9062   0.791 0.432938
## Tax          -4.2280     2.0301  -2.083 0.042873 *
## Dlic          0.4719     0.1285   3.672 0.000626 ***
## Income       -6.1353     2.1936  -2.797 0.007508 **
## logMiles     26.7552     9.3374   2.865 0.006259 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 64.89 on 46 degrees of freedom
## Multiple R-squared:  0.5105, Adjusted R-squared:  0.4679
## F-statistic: 11.99 on 4 and 46 DF,  p-value: 9.331e-07
```