Section 1.1 The Simple Linear Regression Model

EXAMPLE 1.1 Prices for Honda Accords

Create a dataframe with **AccordPrice** data.

```
library(Stat2Data)
data("AccordPrice")
```

Look at the structure of the data.

```
str(AccordPrice)
```

```
## 'data.frame':    30 obs. of  3 variables:
##  $ Age    : int  7 4 4 7 9 1 18 2 2 5 ...
##  $ Price  : num  12 17.9 15.7 12.5 9.5 21.5 3.5 22.8 26.8 13.6 ...
##  $ Mileage: num  74.9 53 79.1 50.1 62 4.8 89.4 20.8 4.8 48.3 ...
```

A different way to see the data in Table 1.1 and other variables in our dataframe.

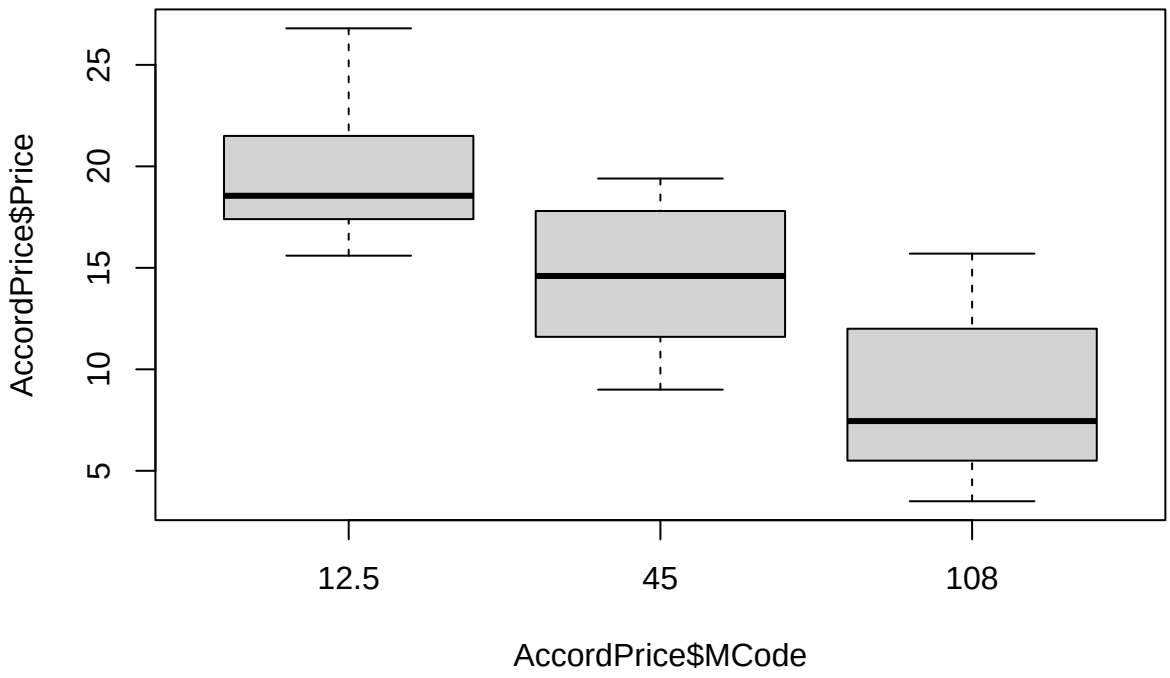
```
head(AccordPrice)
```

```
##   Age Price Mileage
## 1   7  12.0    74.9
## 2   4  17.9    53.0
## 3   4  15.7    79.1
## 4   7  12.5    50.1
## 5   9   9.5    62.0
## 6   1  21.5     4.8
```

FIGURE 1.1 Parallel boxplots for Accords that have been sorted into three groups.

Creating groups to plot low, middle, and high mileage

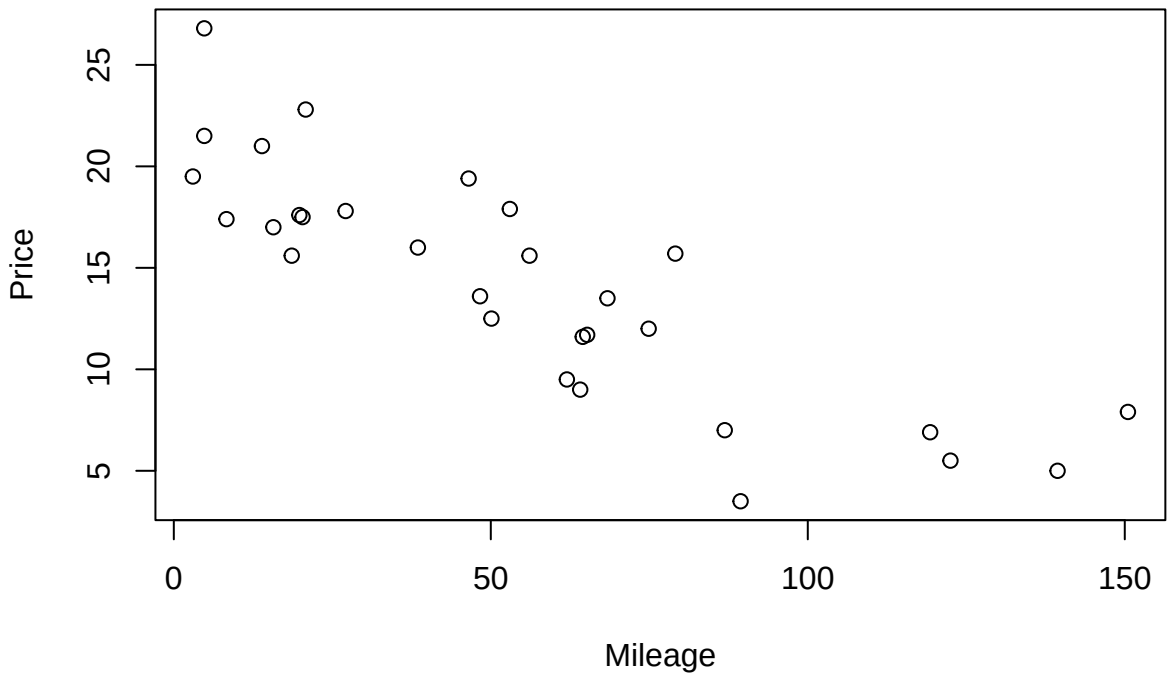
```
AccordPrice$MCode=rep(NA, nrow(AccordPrice))
AccordPrice$MCode=ifelse(AccordPrice$Mileage<25, 12.5, AccordPrice$MCode)
AccordPrice$MCode=ifelse(AccordPrice$Mileage>=25 & AccordPrice$Mileage<=65, 45, AccordPrice$MCode)
AccordPrice$MCode=ifelse(AccordPrice$Mileage>65, 108, AccordPrice$MCode)
boxplot(AccordPrice$Price~AccordPrice$MCode)
```



EXAMPLE 1.2 Honda Accord prices-choosing a model

FIGURE 1.2 Scatterplot of Honda Accord Price versus Mileage

```
plot(Price~Mileage, data=AccordPrice)
```



EXAMPLE 1.3 Honda Accord prices-fitting a model FIT

Find the least-squares regression line.

```
regmodel=lm(Price~Mileage, data=AccordPrice)
summary(regmodel)
```

```
##
## Call:
## lm(formula = Price ~ Mileage, data = AccordPrice)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.5984 -1.8169 -0.4148  1.4502  6.5655
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.8096     0.9529   21.84  < 2e-16 ***
## Mileage      -0.1198     0.0141   -8.50 3.06e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.085 on 28 degrees of freedom
## Multiple R-squared:  0.7207, Adjusted R-squared:  0.7107
## F-statistic: 72.25 on 1 and 28 DF,  p-value: 3.055e-09
```

The fitted values and residuals for all of the observations in the data frame are saved in the regmodel object.

```
regmodel$fitted
```

```
##         1         2         3         4         5         6         7         8
## 11.835698 14.459580 11.332488 14.807034 13.381272 20.234515 10.098425 18.317524
##         9        10        11        12        13        14        15        16
## 20.234515 15.022696 15.238357 20.450177 13.129667 19.815174 17.562709 18.377430
##        17        18        19        20        21        22        23        24
## 12.614476 10.397955 13.081742  2.777915 12.997874 14.088163  4.107827 19.144227
##        25        26        27        28        29        30
## 18.581111 18.928565 16.196853 18.437336  6.516047  6.132649
```

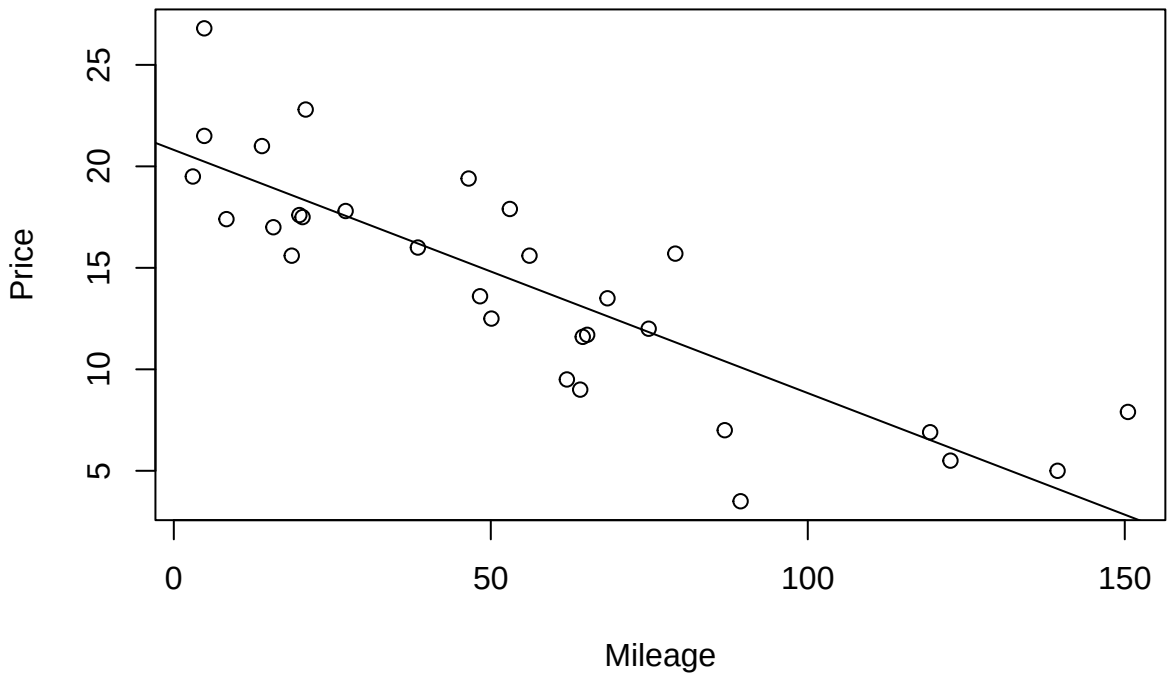
```
regmodel$residuals
```

```
##          1          2          3          4          5          6          7
##  0.1643021  3.4404204  4.3675123 -2.3070342 -3.8812720  1.2654845 -6.5984246
##          8          9         10         11         12         13         14
##  4.4824757  6.5654845 -1.4226957  4.1616428 -0.9501770 -4.1296669 -2.4151737
##         15         16         17         18         19         20         21
##  0.2372910 -0.8774303  0.8855244 -3.3979545 -1.4817422  5.1220854 -1.2978738
##         22         23         24         25         26         27         28
##  1.5118375  0.8921728  1.8557732 -2.9811106 -1.9285653 -0.1968528 -0.8373363
##         29         30
##  0.3839526 -0.6326491
```

FIGURE 1.3 Linear regression to predict Accord Price based on Mileage

Adding the regression line to a scatterplot

```
plot(Price~Mileage, data=AccordPrice)
abline(regmodel)
```



EXAMPLE 1.4 Accord price with centered mileages

Centering a predictor and refitting the regression line

```
AccordPrice$MileageC=AccordPrice$Mileage-50
regmodelC=lm(Price~MileageC, data=AccordPrice)
summary(regmodelC)
```

```
##
## Call:
## lm(formula = Price ~ MileageC, data = AccordPrice)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.5984 -1.8169 -0.4148  1.4502  6.5655
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.8190     0.5668   26.14  < 2e-16 ***
## MileageC     -0.1198     0.0141   -8.50 3.06e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.085 on 28 degrees of freedom
## Multiple R-squared:  0.7207, Adjusted R-squared:  0.7107
## F-statistic: 72.25 on 1 and 28 DF,  p-value: 3.055e-09
```

FIGURE 1.4 Plots comparing regressions to predict Accord prices

Orignal plot is as above, here's code for the plot with centered data.

```
plot(Price~MileageC, data=AccordPrice)
abline(regmodelC)
```