Project 5

Marcus McKenzie

5.1

Describe a situation or problem from your job, everyday life, current events, etc., for which a linear regression model would be appropriate. List some (up to 5) predictors that you might use.

As someone who works with robots, it would be interesting to see the relationship between the amount of Roombas sold, for a certain period and the amount spent on advertising over that same amount of time.

Some of the predictor variables for this relationship could be income, location, family size, housing size, and age. Increased advertising would be expected to result in higher sales. As income, family size, house size and age increases the sale of Roombas likely also increases.

5.2

Use regression to predict the observed crime rate in a city with the following data

Load data:

```
data <- rm(list = ls())

#data <- read.table("Documents/OMSCS/Analytics Modeling/Assignments/Assignment4/sucrime.txt", header= T

data <- read.table("uscrime.txt", header=TRUE, stringsAsFactors = FALSE)

head(data)
```

```
##      M So   Ed  Po1  Po2    LF   M.F Pop   NW    U1  U2 Wealth Ineq     Prob
## 1 15.1  1  9.1  5.8  5.6 0.510  95.0  33 30.1 0.108 4.1   3940 26.1 0.084602
## 2 14.3  0 11.3 10.3  9.5 0.583 101.2  13 10.2 0.096 3.6   5570 19.4 0.029599
## 3 14.2  1  8.9  4.5  4.4 0.533  96.9  18 21.9 0.094 3.3   3180 25.0 0.083401
## 4 13.6  0 12.1 14.9 14.1 0.577  99.4 157  8.0 0.102 3.9   6730 16.7 0.015801
## 5 14.1  0 12.1 10.9 10.1 0.591  98.5  18  3.0 0.091 2.0   5780 17.4 0.041399
## 6 12.1  0 11.0 11.8 11.5 0.547  96.4  25  4.4 0.084 2.9   6890 12.6 0.034201
##      Time Crime
## 1 26.2011   791
## 2 25.2999  1635
## 3 24.3006   578
## 4 29.9012  1969
## 5 21.2998  1234
## 6 20.9995   682
```

Create regression model:

```
regress <- lm(Crime ~ ., data)

summary(regress)
```

```
##
## Call:
## lm(formula = Crime ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -395.74  -98.09   -6.69  112.99  512.67
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.984e+03  1.628e+03  -3.675 0.000893 ***
## M            8.783e+01  4.171e+01   2.106 0.043443 *
## So          -3.803e+00  1.488e+02  -0.026 0.979765
## Ed           1.883e+02  6.209e+01   3.033 0.004861 **
## Po1          1.928e+02  1.061e+02   1.817 0.078892 .
## Po2         -1.094e+02  1.175e+02  -0.931 0.358830
## LF          -6.638e+02  1.470e+03  -0.452 0.654654
## M.F          1.741e+01  2.035e+01   0.855 0.398995
## Pop         -7.330e-01  1.290e+00  -0.568 0.573845
## NW           4.204e+00  6.481e+00   0.649 0.521279
## U1          -5.827e+03  4.210e+03  -1.384 0.176238
## U2           1.678e+02  8.234e+01   2.038 0.050161 .
## Wealth       9.617e-02  1.037e-01   0.928 0.360754
## Ineq         7.067e+01  2.272e+01   3.111 0.003983 **
## Prob        -4.855e+03  2.272e+03  -2.137 0.040627 *
## Time        -3.479e+00  7.165e+00  -0.486 0.630708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.1 on 31 degrees of freedom
## Multiple R-squared:  0.8031, Adjusted R-squared:  0.7078
## F-statistic: 8.429 on 15 and 31 DF,  p-value: 3.539e-07
```

Create test datapoint:

```
test <-data.frame(M = 14.0,So = 0, Ed = 10.0, Po1 = 12.0, Po2 = 15.5,LF = 0.640, M.F = 94.0, Pop = 150,


test
```

```
##    M So Ed Po1  Po2   LF M.F Pop  NW   U1  U2 Wealth Ineq Prob Time
## 1 14  0 10  12 15.5 0.64  94 150 1.1 0.12 3.6   3200 20.1 0.04   39
```

Run prediction:

```
predict(regress, test)
```

```
##        1
## 155.4349
```

Create 2nd model for comparison:

```
regress_two <- lm(Crime~M+Ed+Po1+U2+Ineq, data)

summary(regress_two)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -453.44  -98.59  -18.07  106.03  629.64
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5243.74     951.16  -5.513 2.13e-06 ***
## M             101.98      35.32   2.887 0.006175 **
## Ed            203.08      47.42   4.283 0.000109 ***
## Po1           123.31      14.16   8.706 7.26e-11 ***
## U2             91.36      43.41   2.105 0.041496 *
## Ineq           63.49      14.68   4.324 9.56e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 213 on 41 degrees of freedom
## Multiple R-squared:  0.7296, Adjusted R-squared:  0.6967
## F-statistic: 22.13 on 5 and 41 DF,  p-value: 1.105e-10
```

```
predict(regress_two, test)
```

```
##        1
## 1299.626
```

Conclusion

As we can see from the results from the two different predictions, the models do a fairly effective job in predicting the estimates. However, it seems that the second model, which uses smaller amount of predictors, does a better job in predicting the crime.