Simple Regression in R

ECO 6416

2022-09-12

Contents

Here are all the packages needed to get started.

```
library(readxl)

sessionInfo()
```

```
## R version 4.2.1 (2022-06-23 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] readxl_1.4.1
##
## loaded via a namespace (and not attached):
##  [1] digest_0.6.29    cellranger_1.1.0 magrittr_2.0.3   evaluate_0.15
##  [5] rlang_1.0.3      stringi_1.7.8    cli_3.3.0        rstudioapi_0.14
```

```
##  [9] rmarkdown_2.14   tools_4.2.1      stringr_1.4.0    xfun_0.31
## [13] yaml_2.3.5       fastmap_1.1.0    compiler_4.2.1   htmltools_0.5.2
## [17] knitr_1.39
```

1 Data

We will be using the First Class Mail Volume as the dataset

```
mail <- read_excel("../Data/First Class Mail.xlsx")
```

2 Cross Section

2.1 Scatterplots

We can also visualize the relationship with a scatterplot.

```
plot(mail$PopUSA, mail$FirstClVol,
     main = "Scatterplot of Population and First Class Mail",
     xlab = "Population",
     ylab = "Mail Volume")
```

2.2 Regression

```
model_1 <- lm(FirstClVol ~ PopUSA, data = mail)
model_1
```

```
##
## Call:
## lm(formula = FirstClVol ~ PopUSA, data = mail)
##
## Coefficients:
## (Intercept)       PopUSA
##     -7.5008       0.3059
```

2.3 Thing 1

```
newPrediction <- data.frame(PopUSA = 300)

predict(model_1,
        newdata = newPrediction,
        interval = "confidence",
        level = .8)
```

```
##        fit   lwr      upr
## 1 84.27946 79.95 88.60893
```

2.4 Thing 2

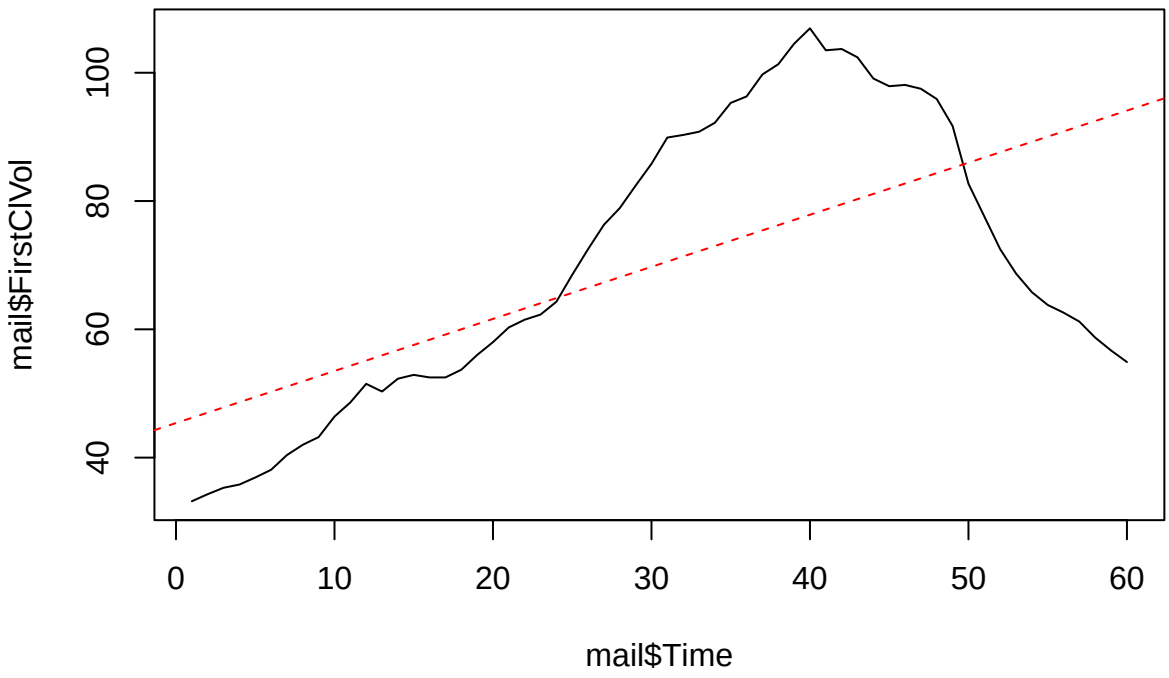
```
model_1$coefficients[2]*10
```

```
##   PopUSA
## 3.059342
```

3 Time Series

3.1 Graph

```
plot(mail$Time,mail$FirstClVol, type = "l")
abline(lm(FirstClVol ~ Time, data = mail),lty = 2, col = "red")
```



3.2 Regression

```
ts_model <- lm(FirstClVol ~ Time, data = mail)

summary(ts_model)
```

```
##
## Call:
## lm(formula = FirstClVol ~ Time, data = mail)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.200 -11.044  -3.453  15.998  29.058
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  45.3886     4.6117   9.842 5.58e-14 ***
## Time          0.8119     0.1315   6.174 7.02e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.64 on 58 degrees of freedom
## Multiple R-squared:  0.3966, Adjusted R-squared:  0.3862
## F-statistic: 38.12 on 1 and 58 DF,  p-value: 7.017e-08
```

3.3 Thing 1

```
newForecast <- data.frame(Time = 62)

predict(ts_model,newdata = newForecast)
```

```
##        1
## 95.72336
```

3.4 Thing 2

```
ts_model$coefficients[2]*.25
```

```
##      Time
## 0.2029627
```