FIT2086 Assignment 3

Ian Wong 30612616

04/10/2021

Loading Libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Question 1

1.

```
concrete <- read_csv("concrete.ass3.2021.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Cement = col_double(),
##   Blast.Furnace.Slag = col_double(),
##   Fly.Ash = col_double(),
##   Water = col_double(),
##   Superplasticizer = col_double(),
##   Coarse.Aggregate = col_double(),
##   Fine.Aggregate = col_double(),
##   Age = col_double(),
##   Strength = col_double()
## )
```

```r
fit <- lm(Strength ~ ., data=concrete)
summary(fit)
```

```
##
## Call:
## lm(formula = Strength ~ ., data = concrete)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.635  -7.318   0.826   7.156  34.187
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -3.019699  54.861392  -0.055  0.95615
## Cement              0.117578   0.018286   6.430 6.79e-10 ***
## Blast.Furnace.Slag  0.095221   0.021604   4.408 1.57e-05 ***
## Fly.Ash             0.069974   0.026894   2.602  0.00985 **
## Water              -0.188562   0.082386  -2.289  0.02296 *
## Superplasticizer    0.424494   0.201844   2.103  0.03650 *
## Coarse.Aggregate    0.008365   0.018829   0.444  0.65726
## Fine.Aggregate      0.016529   0.022908   0.722  0.47126
## Age                 0.127542   0.011233  11.354  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.76 on 241 degrees of freedom
## Multiple R-squared:  0.6213, Adjusted R-squared:  0.6087
## F-statistic: 49.41 on 8 and 241 DF,  p-value: < 2.2e-16
```

The fitted linear model is:

$$\begin{aligned}\widehat{Strength}_i = &-3.02 + 0.12Cement_i + 0.10Blast.Furnace.Slag_i + 0.07Fly.Ash_i - 0.19Water_i \\ &+ 0.42Superplasticizer_i + 0.01Coarse.Aggregate_i + 0.02Fine.Aggregate_i + 0.13Age_i\end{aligned}$$

The predictors that are possibly associated with the compressive strength of concrete are **Cement, Blast.Furnace.Slag, Fly.Ash, Water, Superplasticizer**, and **Age**. This is because compressive strength is most likely determined by the amount of each component substance within the concrete mixture.

The three variables that appear to be the strongest predictors of compressive strength appear to be:

- **Age,**
- **Cement,** and
- **Blast Furnace Slag**

This is based on the p-values of these three predictors being the smallest compared to the other predictors according to the information gathered from `summary()`.

2.

If we were to use the Bonferroni procedure with $\alpha = 0.05$ and $p = 8$, we would instead reject the null hypothesis only if:

$$\begin{aligned} \text{p-value} &< \frac{\alpha}{p} \\ &< \frac{0.05}{8} \\ &< 0.00625 \end{aligned}$$

Therefore, we would need to see p-values of < 0.00625 to believe that a predictor is associated with the compressive strength of concrete. **Cement, Blast.Furnace.Slag, and Age** are the only predictors that are associated with compressive strength afterwards.

3.

An extra kg of Cement in a m^3 mixture would increase the mean compressive strength of concrete by 0.117578.
An extra day of waiting after pouring the concrete would increase the mean compressive strength of concrete by 0.127542.

4.

```r
fit.sw.bic = step(fit, k = log(length(concrete$Strength)))
```

```
## Start:  AIC=1228.65
## Strength ~ Cement + Blast.Furnace.Slag + Fly.Ash + Water + Superplasticizer +
##     Coarse.Aggregate + Fine.Aggregate + Age
##
##                      Df Sum of Sq   RSS    AIC
## - Coarse.Aggregate    1      22.9 27948 1223.3
## - Fine.Aggregate      1      60.3 27986 1223.7
## - Superplasticizer    1     512.5 28438 1227.7
## - Water               1     607.0 28532 1228.5
## <none>                            27925 1228.7
## - Fly.Ash             1     784.4 28710 1230.0
## - Blast.Furnace.Slag  1    2251.1 30176 1242.5
## - Cement              1    4790.8 32716 1262.7
## - Age                 1   14938.4 42864 1330.2
##
## Step:  AIC=1223.33
## Strength ~ Cement + Blast.Furnace.Slag + Fly.Ash + Water + Superplasticizer +
##     Fine.Aggregate + Age
##
##                      Df Sum of Sq   RSS    AIC
## - Fine.Aggregate      1      45.2 27993 1218.2
## - Superplasticizer    1     493.0 28441 1222.2
## <none>                            27948 1223.3
## - Fly.Ash             1    1103.3 29052 1227.5
## - Water               1    2183.4 30132 1236.6
## - Blast.Furnace.Slag  1    4992.6 32941 1258.9
```

```
## - Cement              1   10431.9 38380 1297.1
## - Age                 1   14935.7 42884 1324.8
##
## Step:  AIC=1218.22
## Strength ~ Cement + Blast.Furnace.Slag + Fly.Ash + Water + Superplasticizer +
##     Age
##
##                      Df Sum of Sq   RSS    AIC
## - Superplasticizer    1     579.2 28573 1217.8
## <none>                            27993 1218.2
## - Fly.Ash             1    1171.2 29165 1222.9
## - Water               1    3114.2 31108 1239.1
## - Blast.Furnace.Slag  1    6439.1 34433 1264.5
## - Age                 1   14911.7 42905 1319.5
## - Cement              1   15434.2 43428 1322.5
##
## Step:  AIC=1217.81
## Strength ~ Cement + Blast.Furnace.Slag + Fly.Ash + Water + Age
##
##                      Df Sum of Sq   RSS    AIC
## <none>                            28573 1217.8
## - Fly.Ash             1    2951.8 31524 1236.9
## - Water               1    8037.5 36610 1274.3
## - Blast.Furnace.Slag  1   10339.2 38912 1289.5
## - Age                 1   15590.4 44163 1321.2
## - Cement              1   20238.3 48811 1346.2
```

```
summary(fit.sw.bic)
```

```
##
## Call:
## lm(formula = Strength ~ Cement + Blast.Furnace.Slag + Fly.Ash +
##     Water + Age, data = concrete)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.555  -6.267   0.967   7.530  31.840
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        40.225834   7.948050   5.061 8.21e-07 ***
## Cement              0.114250   0.008691  13.146  < 2e-16 ***
## Blast.Furnace.Slag  0.094203   0.010025   9.396  < 2e-16 ***
## Fly.Ash             0.076221   0.015181   5.021 9.94e-07 ***
## Water              -0.293329   0.035406  -8.285 7.90e-15 ***
## Age                 0.128814   0.011164  11.538  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.82 on 244 degrees of freedom
## Multiple R-squared:  0.6125, Adjusted R-squared:  0.6045
## F-statistic: 77.13 on 5 and 244 DF,  p-value: < 2.2e-16
```

After pruning, the final regression equation is:

$$\widehat{Strength}_i = 40.23 + 0.11Cement_i + 0.09Blast.Furnace.Slag_i + 0.08Fly.Ash_i - 0.29Water_i + 0.13Age_i$$

5.

a.

```
exampleData <- data.frame(Cement = 491,
                          Blast.Furnace.Slag = 26,
                          Fly.Ash=26,
                          Water = 210,
                          Superplasticizer = 3.9,
                          Coarse.Aggregate = 882,
                          Fine.Aggregate = 699,
                          Age = 28)

pred <- predict(fit.sw.bic, newdata=exampleData, interval="confidence", level=0.95)
pred
```

```
##        fit      lwr      upr
## 1 42.76111 38.73784 46.78438
```

The predicted mean compressive strength for this mix is 42.761111. The lower confidence interval is 38.737839 and the upper confidence interval is 46.7843829.

b.

As the mix of concrete that the engineer is currently using has a greater mean compressive strength than the predicted mean compressive strength of the new concrete mix, the model suggests that the new mix is worse than the current mix.

Question 2.

1.

Using the