Transformation

오태환

2020-05-27

0) Importing Data

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

```
## -- Attaching packages ----------------------------------- tidyverse 1.3.0 --
```

```
## √ ggplot2 3.3.0     √ purrr   0.3.3
## √ tibble  2.1.3     √ dplyr   0.8.4
## √ tidyr   1.0.2     √ stringr 1.4.0
## √ readr   1.3.1     √ forcats 0.5.0
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'purrr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

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

```
library(fBasics)
```

```
## Warning: package 'fBasics' was built under R version 3.6.3
```

```
## Loading required package: timeDate
```

```
## Loading required package: timeSeries
```

```
## Warning: package 'timeSeries' was built under R version 3.6.3
```

```
mice_data = read_csv("C:/Users/dhxog/Desktop/ESC4-1/Final_Project/ESC20SPRING_team4/raw_data/im
puted_data_mice.csv")
```

```
## Parsed with column specification:
## cols(
##    .default = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
knn_data = read_csv("C:/Users/dhxog/Desktop/ESC4-1/Final_Project/ESC20SPRING_team4/raw_data/imp
uted_data_knn.csv")
```

```
## Parsed with column specification:
## cols(
##    .default = col_double()
## )
## See spec(...) for full column specifications.
```

```
mean_data = read_csv("C:/Users/dhxog/Desktop/ESC4-1/Final_Project/ESC20SPRING_team4/raw_data/im
puted_data_mean.csv")
```

```
## Parsed with column specification:
## cols(
##    .default = col_double()
## )
## See spec(...) for full column specifications.
```

1) Transformation

1-a) Mice Imputated Data

```
for(i in 1:(ncol(mice_data)-1)){
  skew_score = skewness(mice_data[,i])
  if(skew_score > 1 | skew_score < -1){
    if(min(mice_data[,i]) <= 0){
    mice_data[,i] = log(mice_data[,i] - min(mice_data[,i]) + 2)
    }
    else{
      mice_data[,i] = log(mice_data[,i])
    }
  }
}

write.csv(mice_data, "transfromed_data_mice.csv", row.names = FALSE)
```

왜도가 절대값 1 기준으로 큰 것들은 log transformation으로 왜도를 줄여주었다.

1-b) KNN Imputed Data

```r
for(i in 1:(ncol(knn_data)-1)){
  skew_score = skewness(knn_data[,i])
  if(skew_score > 1 | skew_score < -1){
    if(min(knn_data[,i]) <= 0){
    knn_data[,i] = log(knn_data[,i] - min(knn_data[,i]) + 2)
    }
    else{
      knn_data[,i] = log(knn_data[,i])
    }
  }
}

write.csv(knn_data, "transfromed_data_knn.csv", row.names = FALSE)
```

1-c) MEAN Imputed Data

```r
for(i in 1:(ncol(mean_data)-1)){
  skew_score = skewness(mean_data[,i])
  if(skew_score > 1 | skew_score < -1){
    if(min(mean_data[,i]) <= 0){
    mean_data[,i] = log(mean_data[,i] - min(mean_data[,i]) + 2)
    }
    else{
      mean_data[,i] = log(mean_data[,i])
    }
  }
}

write.csv(mean_data, "transfromed_data_mean.csv", row.names = FALSE)
```