MATH1324 Assignment 1

Statistical analysis of Climate data

Student Details

Wing Hang Chan (s3939713)

Matthew Ng Jun Hung (s3963313)

Sukhum Boondecharak (s3940976)

Problem Statement

Many people in Melbourne and Sydney suffered heatstroke in last summer. For example, Rafael Nadal reportedly suffered heat stroke and lost 4 kilograms in Australian Open (Ref 1). Heatstroke (Ref 2) is a condition caused by your body overheating, usually as a result of prolonged exposure to or physical exertion in high temperatures. It is the main reason we have chosen max temperature and solar exposure as variables. High temperatures and high solar exposure may relatively be the cause of body overheating.

Ref 1: Rafael Nadal reportedly suffered heat stroke - https://7news.com.au/sport/tennis/rafael-nadal-reportedly-suffered-heat-stroke-lost-4kg-in-australian-open-win-over-shapovalov-c-5457236

Ref 2: Heatstroke condition - https://www.mayoclinic.org/diseases-conditions/heat-stroke/symptoms-causes/syc-20353581

Load Packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Data

```r
data_mel <- read.csv('Climate Data-Melbourne-1.csv')
data_syd <- read.csv('Climate Data-Sydney-1.csv')

colnames(data_mel)[colnames(data_mel) == 'Maximum.temperature..Degree.C.'] <- 'max_temp'
colnames(data_mel)[colnames(data_mel) == 'solar.exposure'] <- 'solar_exp'

colnames(data_syd)[colnames(data_syd) == 'Maximum.temperature..Degree.C.'] <- 'max_temp'
colnames(data_syd)[colnames(data_syd) == 'solar.exposure'] <- 'solar_exp'

min_temp <- min(min(data_mel$max_temp), min(data_syd$max_temp))
max_temp <- max(max(data_mel$max_temp), max(data_syd$max_temp))

x_lim_max_temp = c(min_temp, max_temp)

min_solar_exp <- min(min(data_mel$solar_exp), min(data_syd$solar_exp))
max_solar_exp <- max(max(data_mel$solar_exp), max(data_syd$solar_exp))

x_lim_solar_exp = c(min_solar_exp, max_solar_exp)
```

Summary Statistics

Calculate mean, median, standard deviation, first and third quartile, interquartile range, minimum and maximum values) of the selected variable grouped by city.

```r
#Statistic for Maximum Temperature in Melbourne

data_mel %>% summarise(Mean = mean(max_temp), Median = median(max_temp),
                       Stand_dev = sd(max_temp),
                       First_quarter = quantile(max_temp, 0.25),
                       Third_quarter = quantile(max_temp,0.75),
                       Interquartile = IQR(max_temp),
                       Min = min(max_temp), Max = max(max_temp))
```

```
##       Mean Median Stand_dev First_quarter Third_quarter Interquartile  Min  Max
## 1 27.01778  26.85  5.596319        22.825        31.675          8.85 14.9 38.4
```

```r
#Statistic for Solar Exposure in Melbourne

data_mel %>% summarise(Mean = mean(solar_exp), Median = median(solar_exp),
                       Stand_dev = sd(solar_exp),
                       First_quarter = quantile(solar_exp, 0.25),
                       Third_quarter = quantile(solar_exp,0.75),
                       Interquartile = IQR(solar_exp),
                       Min = min(solar_exp), Max = max(solar_exp))
```

```
##       Mean Median Stand_dev First_quarter Third_quarter Interquartile Min Max
## 1 21.90556   22.6  7.313407        16.875          28.1        11.225 3.4  32
```

```r
#Statistic for Maximum Temperature in Sydney
```

```r
data_syd %>% summarise(Mean = mean(max_temp), Median = median(max_temp),
                       Stand_dev = sd(max_temp),
                       First_quarter = quantile(max_temp, 0.25),
                       Third_quarter = quantile(max_temp,0.75),
                       Interquartile = IQR(max_temp),
                       Min = min(max_temp), Max = max(max_temp))
```

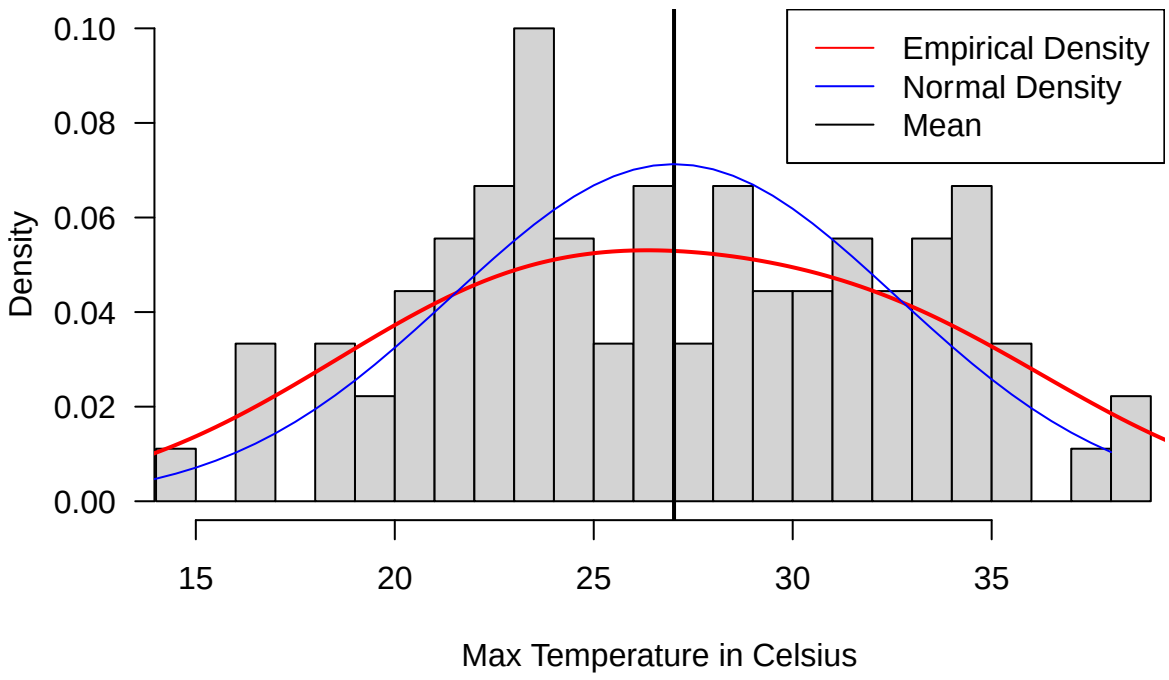
```
##       Mean Median Stand_dev First_quarter Third_quarter Interquartile  Min  Max
## 1 27.01222  26.85  3.584355        24.325          29.2         4.875 19.7 37.8
```

```r
#Statistic for Solar Exposure in Sydney

data_syd %>% summarise(Mean = mean(solar_exp), Median = median(solar_exp),
                       Stand_dev = sd(solar_exp),
                       First_quarter = quantile(solar_exp, 0.25),
                       Third_quarter = quantile(solar_exp,0.75),
                       Interquartile = IQR(solar_exp),
                       Min = min(solar_exp), Max = max(solar_exp))
```

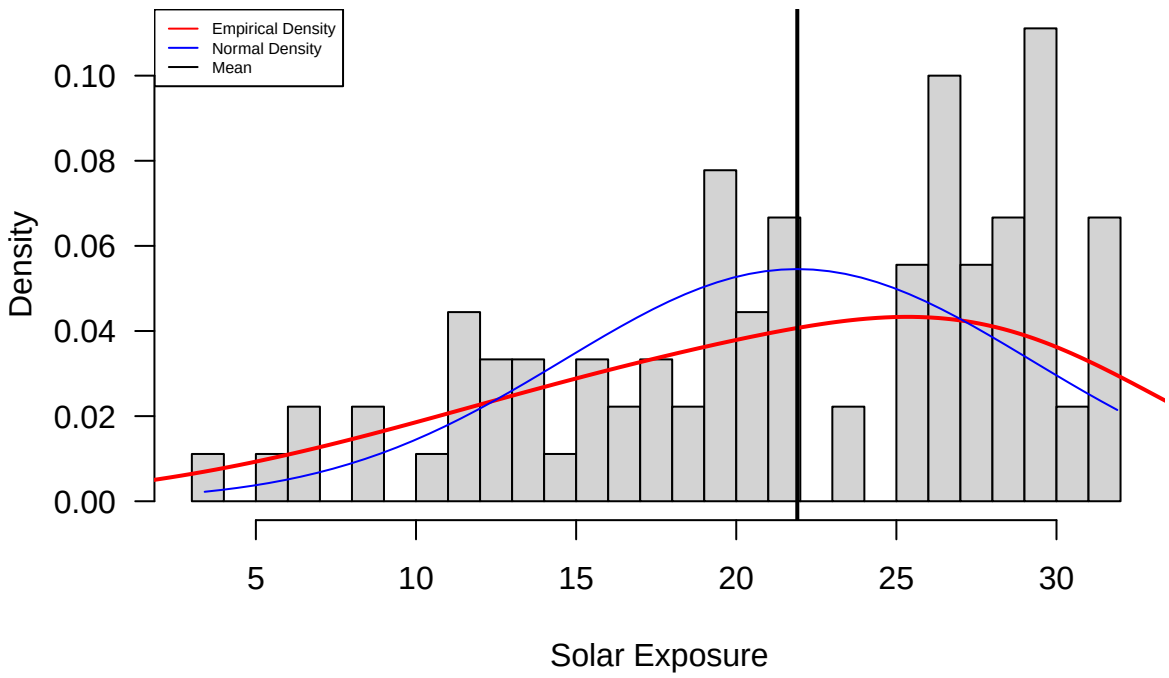
```
##       Mean Median Stand_dev First_quarter Third_quarter Interquartile Min  Max
## 1 20.08889  20.65  7.320457        15.875        25.875            10   3 32.2
```

Distribution Fitting

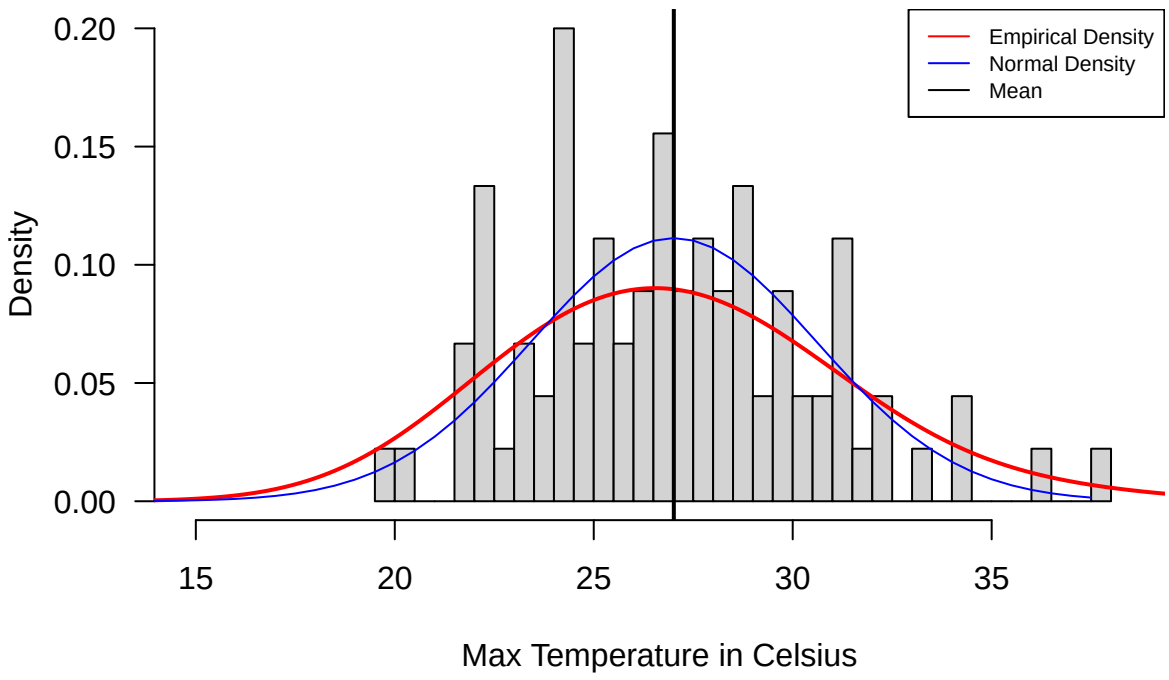
```r
#Distribution of variables
#Histogram of Melbourne's Max Temperature
# par(mfrow = c(1, 1))
#par(mfrow = c(2, 2))
hist(data_mel$max_temp,
     breaks = 30,
     main = "1: Histogram of Melbourne's Max Temperature",
     xlab = "Max Temperature in Celsius"  ,
     xlim = x_lim_max_temp,
     prob = T,
     las = 1)
lines(density(data_mel$max_temp, adjust = 2), col = "Red", lwd = 2)
abline(v=mean(data_mel$max_temp),col='black',lw=2)

#Normal Distribution Data Fitting of Melbourne's Max Temperature
#Blue Line
lines(seq(0,
          max(data_mel$max_temp), by=.5),
      dnorm(seq(0,
                max(data_mel$max_temp), by=.5),
            mean(data_mel$max_temp),
            sd(data_mel$max_temp)), col="Blue")

legend(x = 'topright',
       legend = c("Empirical Density","Normal Density","Mean"),
       col = c('red',"Blue","black"),
       lty = c(1,1,1))
```



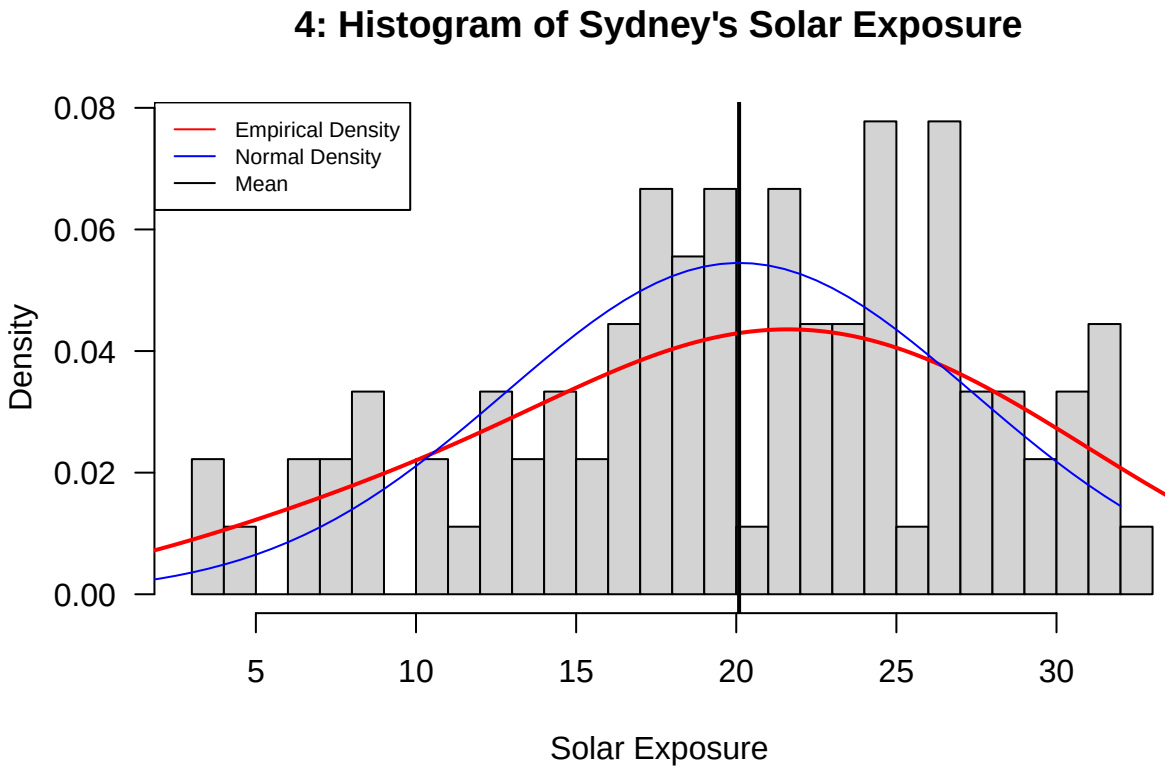

```
#Histogram of Melbourne's Solar Exposure
hist(data_mel$solar_exp,
     breaks = 30,
     main = "2: Histogram of Melbourne's Solar Exposure",
     xlab = "Solar Exposure",
     xlim = x_lim_solar_exp,
     prob = T,
     las = 1)
lines(density(data_mel$solar_exp, adjust = 2), col = "Red", lwd = 2)
abline(v=mean(data_mel$solar_exp),col='black',lw=2)

#Normal Distribution Data Fitting of Melbourne's Solar Exposure
lines(seq(min(data_mel$solar_exp),
          max(data_mel$solar_exp), by=.5),
      dnorm(seq(min(data_mel$solar_exp),
                max(data_mel$solar_exp), by=.5),
            mean(data_mel$solar_exp),
            sd(data_mel$solar_exp)), col="Blue")

legend(x = 'topleft',
       legend = c("Empirical Density","Normal Density","Mean"),
       col = c('red',"Blue","Black"),
       lty = c(1,1,1),
       cex = 0.5)
```




```
#Histogram of Sydney's Max Temperature
hist(data_syd$max_temp,
     breaks = 30,
     main = "3: Histogram of Sydney's Max Temperature",
     xlab = "Max Temperature in Celsius",
     xlim = x_lim_max_temp,
     prob = T,
     las = 1)
lines(density(data_syd$max_temp, adjust = 2), col = "Red", lwd = 2)
abline(v=mean(data_syd$max_temp),col='black',lw=2)

#Normal Distribution Data Fitting of Sydney's Maximum Temperature
lines(seq(0,
          max(data_syd$max_temp), by=.5),
      dnorm(seq(0,
                max(data_syd$max_temp), by=.5),
            mean(data_syd$max_temp),
            sd(data_syd$max_temp)), col="Blue")

legend(x = 'topright',
       legend = c("Empirical Density","Normal Density","Mean"),
       col = c('red',"Blue","Black"),
       lty = c(1,1,1),
       cex = 0.7)
```




```
#Histogram of Sydney's Solar Exposure
hist(data_syd$solar_exp,
     breaks = 30,
     main = "4: Histogram of Sydney's Solar Exposure",
     xlab = "Solar Exposure",
     xlim = x_lim_solar_exp,
     prob = T,
     las = 1)
lines(density(data_syd$solar_exp, adjust = 2), col = "Red", lwd = 2)
abline(v=mean(data_syd$solar_exp),col='black',lw=2)

#Normal Distribution Data Fitting of Sydney's Solar Exposure
lines(seq(0,
          max(data_syd$solar_exp), by=.5),
      dnorm(seq(0,
                max(data_syd$solar_exp), by=.5),
            mean(data_syd$solar_exp),
            sd(data_syd$solar_exp)), col="Blue")


legend(x = 'topleft',
       legend = c("Empirical Density","Normal Density","Mean"),
       col = c('red',"Blue","black"),
       lty = c(1,1,1),
       cex = 0.7)
```



Interpretation

As graph 1 and 3 show a flatter curve compared with normal distribution curve, it may be because we got a small size of data set. There is no evidence that maximum temperature in both countries relates to a high chance of people suffering from heatstroke with the balance distribution. However, there is a finding with analysing data of solar exposure. It shows both cities had much sunshine last summer as the mean value is at the right hand side of the graph. There is a higher chance to have higher solar exposure than normal expected in both Melbourne and Sydney. It may explain why many people suffered from heatstroke last summer.