STA5075: Practical 3

Jessica Stow (STWJES003@myuct.ac.za)

2025-02-04

Question 1

Find all rows in airquality that have missing values.

```r
which(is.na(airquality)) # identify which rows have missing values
```

```
##  [1]    5  10  25  26  27  32  33  34  35  36  37  39  42  43  45  46  52  53  54
## [20]   55  56  57  58  59  60  61  65  72  75  83  84 102 103 107 115 119 150 158
## [39]  159 164 180 249 250 251
```

Question 2

Find mean, sd, min, max for each of temperature and ozone level.

```r
# Temperature
mean(airquality$Temp) # mean temperature
```

```
## [1] 77.88235
```

```r
sd(airquality$Temp) # standard deviation of temperature
```

```
## [1] 9.46527
```

```r
min(airquality$Temp) # minimum temperature
```

```
## [1] 56
```

```r
max(airquality$Temp) # maximum temperature
```

```
## [1] 97
```

```r
# Ozone level
mean(airquality$Ozone, na.rm = TRUE) # mean Ozone level
```

```
## [1] 42.12931
```

```r
sd(airquality$Ozone, na.rm = TRUE) # standard deviation of Ozone level
```

```
## [1] 32.98788
```

```r
min(airquality$Ozone, na.rm = TRUE) # minimum Ozone level
```

```
## [1] 1
```

```r
max(airquality$Ozone, na.rm = TRUE) # maximum Ozone level
```

```
## [1] 168
```

Question 3

The cars data (an R data set) contains two variables: speed and distance to stop. Fit a simple linear regression model to these data, i.e. find the B estimates, using the equation above, and matrix calculations in R.

```r
x <- cbind(1, cars$speed) # x matrix
y <- cars$dist # y matrix

solve(t(x)%*%x)%*%(t(x)%*%y) # y intercept = -17.579095, coefficient of speed = 3.932409
```

```
##               [,1]
## [1,] -17.579095
## [2,]   3.932409
```

Question 4

Check that you get the same B estimates as when fitting the linear regression model using lm() in R.

```r
m1 <- lm(dist ~ speed, data = cars)
m1$coefficients
```

```
## (Intercept)        speed
##   -17.579095     3.932409
```

Question 5

Mean life expectancy in South Africa? **53.99317.**

```r
library(gapminder)
sa <- gapminder[gapminder$country == "South Africa", ] # create subset of dataset for South Africa
mean(sa$lifeExp) # mean life expectancy in South Africa = 53.99317
```

```
## [1] 53.99317
```