Data Generation

2023-07-21

```
world_hap <- read.csv("world_happiness_2016.csv")
```

```
world_hap2 <- world_hap[,c(-5,-6)]
```

Deleting the irrelevant attributes from the dataframe that we will not be using.

```
world_hap3 <- rbind(world_hap2,
                    list("Vatican City", "Western Europe", 133, 4.139, 1.1,
                         1.05, 0.8, 0.0001, 0.0002, 0.4, 2.6))
```

Adding an observation to our dataset.

```
world_hap4 <- world_hap3[c(-133),]
```

Deleting a specific observation that does not have a value for one of our predictors.

```
world_hap5 <- world_hap4[c(-140),]
```

```
world_hap6 <- world_hap5
```

```
#Western Europe, Central and Eastern Europe -> Europe
world_hap6$Region <- ifelse(world_hap6$Region == "Western Europe" |
                              world_hap6$Region == "Central and Eastern Europe",
                            "Europe", world_hap6$Region)

#Eastern Asia, Southeastern Asia, Southern Asia -> Asia
world_hap6$Region <- ifelse(world_hap6$Region == "Eastern Asia" |
                              world_hap6$Region == "Southeastern Asia" |
                              world_hap6$Region == "Southern Asia", "Asia",
                            world_hap6$Region)
world_hap6$Region <- ifelse(world_hap6$Region == "Australia and New Zealand",
                            "Oceania", world_hap6$Region)

#Middle East and Northern Africa, Sub-Saharan Africa -> Africa and Middle East
world_hap6$Region <- ifelse(world_hap6$Region == "Middle East and Northern Africa"
                            | world_hap6$Region == "Sub-Saharan Africa",
                            "Africa and Middle East", world_hap6$Region)


#North America, Latin America and Caribbean -> Americas
world_hap6$Region <- ifelse(world_hap6$Region == "North America" |
                              world_hap6$Region == "Latin America and Caribbean"
                            , "Americas", world_hap6$Region)
```

This part will be explained in the diagnosis section of the project. This is here to be able to create our response variable.

```
X <- model.matrix(~world_hap6$Region + world_hap6$Trust..Government.Corruption. + log(world_hap6$Freedo
betas <- c(beta0 = 3.2, beta1 = 3.4, beta2 = 3.1, beta3 = 3.4, beta4= 3.2,
           beta5 = 2.1, beta6 = 0.7)

y <- X%*%betas + rnorm(length(world_hap6$Region), 0, 0.5)
```

```
world_hap5$Skepticism <- y
```