

# Homework 2 Report

Santhoshi Penmetsa

## Setup

Read in the data with `read_csv()` and store the data as an R object named `dataset`. Check the data to make sure all of the expected observations and variables are there.

```
## Load the data and any necessary packages here.
dataset = read.csv('maacs.csv.gz')
```

## Part 1

We will first consider the relationship between FEV1 and age. In general, it is expected that as children get older (and hence, larger in size), their FEV1 values should get higher.

Consider the statement “FEV1 values in children are higher in older children relative to younger children”.

Write a function in R that takes the `dataset` object as an argument and returns `TRUE` if the statement above is true for the dataset and `FALSE` otherwise.

NOTE: In order to write this function, you will need to translate the statement above into something that can be checked with the data. There are many ways in which you can do that translation correctly and you only need to pick one way here.

NOTE: For this part, do not use any plots.

```
# Linear regression model
model <- lm(fev1 ~ age, data = dataset)
slope <- coef(model)["age"]
slope
```

```
##           age
## 0.1711623
```

```
# Correlation between the two variables
cor(dataset$fev1, dataset$age, use = "complete.obs")
```

```
## [1] 0.8059696
```

```
## Write your function here
fev1_age <- function(dataset) {
  model <- lm(fev1 ~ age, data = dataset)
  slope <- coef(model)["age"]

  if (slope > 0) {
    result <- TRUE
  } else {
    result <- FALSE
  }

  return(result)
}
```

```
fev1_age(dataset)
```

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

## Part 2

Fit a linear regression model with FEV1 as the outcome and age as a predictor.

How much does FEV1 change for a 1-year increase in the child's age?

```
## Add your code here
```

Write your data analysis statement interpreting the regression model here:

## Part 3

Develop **three** supporting premises derived from the data that support the statement you wrote in Part 2. These can be plots, other summary statistics, or model results.

NOTE: At least one supporting premise should use a plot.

```
## Add your code here
```

Write the three supporting premise statements here:

- 1.
- 2.
- 3.

## Part 4

For each of the supporting premises above, write a function that takes the **dataset** object as an argument and returns **TRUE** if the supporting premise statement above is true for the dataset and **FALSE** otherwise.

For statements involving plots, instead of returning **TRUE** or **FALSE**, you function should do two things:

1. Produce the plot that is used in the statement
2. Produce a hypothetical version of the plot in the event that the statement is true. This can be done using simulated data or by simply hand drawing a plot.

```
## Function for supporting premise statement 1
```

```
## Function for supporting premise statement 2
```

```
## Function for supporting premise statement 3
```

Execute each of your function and show that they produce the expected output.

## Part 5

Describe one alternative to the primary statement “FEV1 values in children are higher in older children relative to younger children”.

Create a fault tree for the alternative outcome describing how the alternative outcome could be realized in the data even if the primary statement were true.

Your fault tree should be created as a separate image and does not need to be created in R. Upload the image of the fault tree to Canvas.