

Logistic Regression Analysis

Group 5 : MVA

As the main question for our project is derive the Life Expectancy of people in different Countries based on different independent variables. Logistic Regression is a classification technique in which require set of class as 0 or 1 for the values of the dependent variable.

As our dataset is on detecting the factors affecting the life expectancy of a person, it depends on the Life Expectancy as a dependent variable and class of the variable is numeric is nature.

Therefore, we cannot apply logistic regression to our specific model as the values associated with logistic regression for the dependent variable should lie between 0 to 1.

```
54
55 ##### Finding the summary or the distribution of the Dataset #####
56 str(new.life)
57
58 ##### Using the Logistic Regression #####
59
60 logistic_simple <- glm(Life.expectancy ~ ., data=new.life, family="binomial")
61 summary(logistic_simple)
```

43:59 # CLEANING THE DATA R Script

Console Terminal x Jobs x

```
$ thinness.5.9.years : num 2.5 8.2 8.3 8.5 8.8 9 9.2 9.4 9.6 9.7 ...
$ Income.composition.of.resources: num 0.338 0.531 0.527 0.523 0.495 0.488 0.48 0.468 0.454
0.439 ...
$ Schooling : num 5.5 11.4 11.4 11.4 9.4 9 8.5 8.1 7.7 7.2 ...
- attr(*, "na.action")= 'omit' Named int 64 65 66 67 68 69 70 71 72 73 ...
..- attr(*, "names")= chr "64" "65" "66" "67" ...
> logistic_simple <- glm(Life.expectancy ~ ., data=new.life, family="binomial")
Error in eval(family$initialize) : y values must be 0 <= y <= 1
> class(new.life$Life.expectancy)
[1] "numeric"
> |
```

When we run the logistic regression, we get this result in our case as Error. As the values associated with the dependent variable (Life.expectancy) does not belong to 0 or 1 class.

Therefore, we cannot apply logistic regression technique to our analysis.