Detailed Report

# Introduction

As a team we analyzed the given data set regarding the Panel Survey of Income Dynamics. The data set consist of 4856 tuples with age, education, earnings, working hours, number of kids and marital status.

Group Members

1.

2.

3.

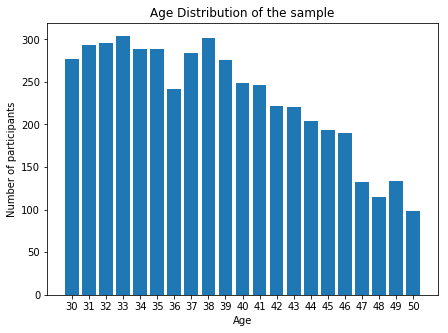
4.

# Initial Analysis

This is the initial analysis on the data set without any pre-processing.

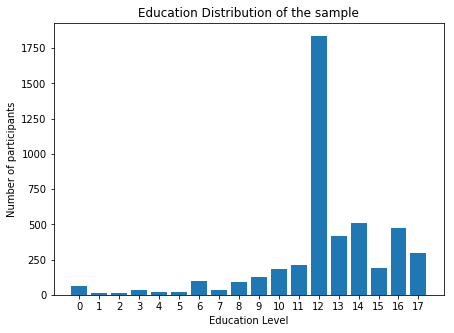
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Min | Max | Median | Mode |
| Age | 30 | 50 | 38 | 33 |
| Education | 0 | 99 | 12 | 12 |
| Earnings | 0 | 240000 | 11000 | 0 |
| Hours | 0 | 5160 | 1517 | 0 |
| Kids | 0 | 99 | 12 | 2 |

## 2.1 Age

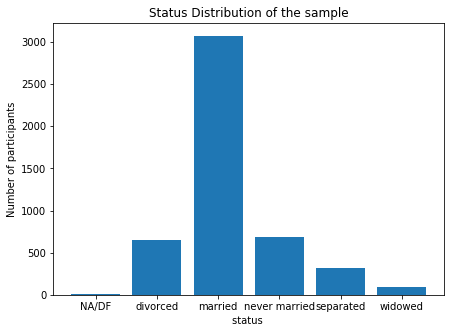
The given data has the age group 30 to 50. The data set is biased. The figure below is the age distribution of the sample.

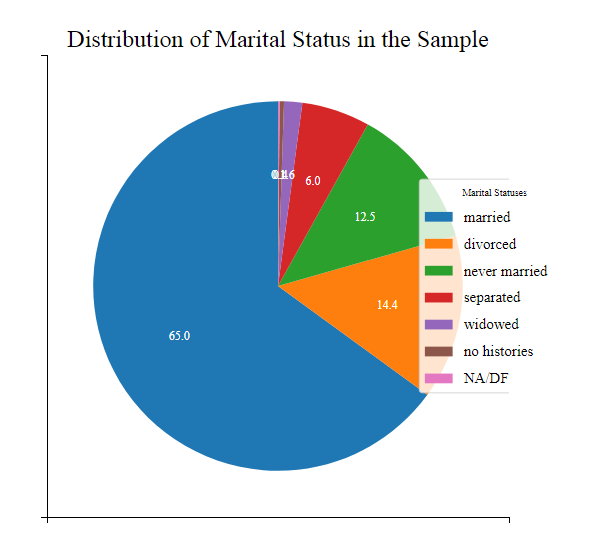
## 2.2 Education

The data set has some uncertain data like education level 98. The figure below is the education distribution of the sample. The uncertain data has been omitted for the graph.



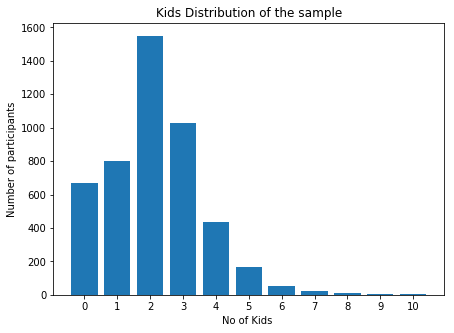
## Marital Status

The figures below show the marital status distribution of the sample.



## Kids

The figures below show the marital kids distribution of the sample. The data set has some uncertain data like having 98 children. Those values were omitted for the graph.



# Secondary Analysis

This is the analysis after removing unrealistic data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | Min | Max | Avg | SD | Median |
| Age | 30 | 50 | 38.46293 | 5.595116 | 38 |
| Education | 0 | 17 | 12.36 | 3.05 | 12 |
| Earnings | 0 | 240000 | 14244.51 | 15985.45 | 11000 |
| Hours | 0 | 5160 | 1235.335 | 947.1758 | 1517 |
| Kids | 0 | 10 | 2.14 | 1.43 | 2 |

# Hypothesis Test

Earning increase with the education level of people who are in the age group of 30 to 50.

# Assumptions

# References

1. https://psidonline.isr.umich.edu/
2. https://psidonline.isr.umich.edu/data/Documentation/UserGuide2015.pdf