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. The data set has some outliers like having 98 kids and education level is 99. Those are unrealistic data. So in our sample, we removed them and analyzed the remaining data. Fist we analyzed the distribution of each important parameters by visualizing them in charts. Then we analyzed the relativeness of the parameters and make a hypothesis. We analyzed further to prove our hypothesis and concluded with the results.

Group Members

1. Nuwani

2. Upekka

3. Miyuru

4.Nilusija

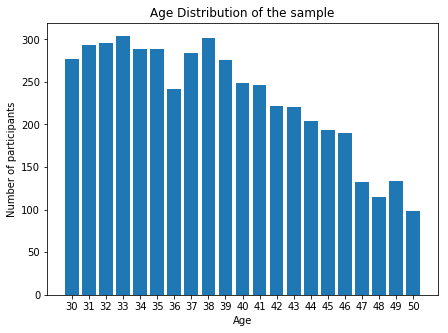
# Initial Analysis

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 |

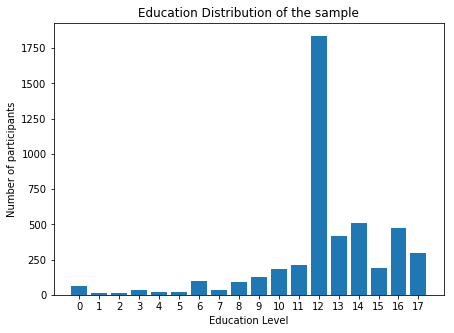
## 2.1 Age

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.



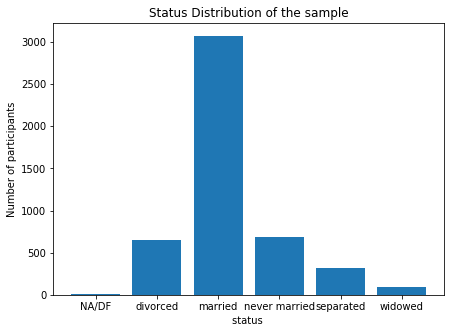
## 2.2 Education

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. In this sample, 1804 people has 12 as their education level. 81.20% of people has 12 or above 12 as their education level.

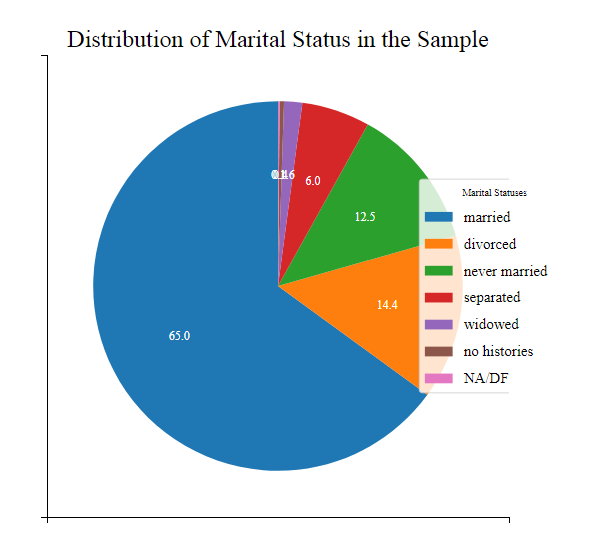


## Marital Status

The figures below show the marital status distribution of the sample. 65% of this sample participants are married. And divorced and never married participants of this sample is nearly equal.

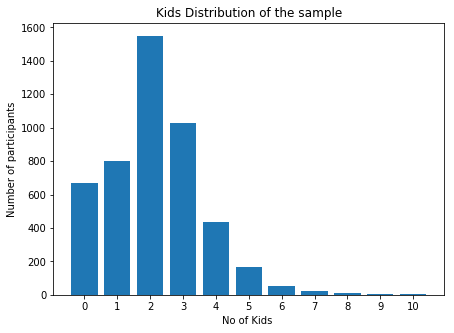


This pie chart show the marital status without preprocessing the data. In the data set, marital status shown as “no histories” has the number of child as 90+. So in the preprocessing step, we have removed those data. But those data may refer the orphange childdrens count. Those are like outliers. We need more information to analyze those data and better to analyze those data separately.



## 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. Most of this people in this sample have two children. The people are willing to have less number of children. May be because of the financial problem and the parents want to give the better education and facilities to their children.



# 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

* The people who are less than 30 years and greater than 50 years are not earning
* Kids are not earning and they don’t have educations as well.

# References

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