**Assignment 1**

**Author:** Vidhya Venugopal, 8902970

**Course:** PROG8430

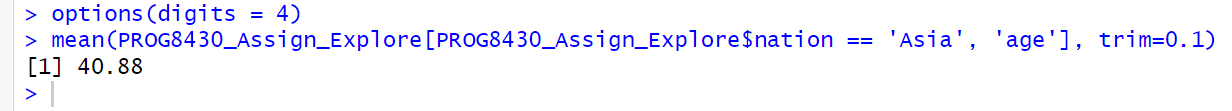
1. Summarizing Data
   1. Summary Table:
      1. Total Income by each category of marital status:

A picture containing text, screenshot, font

Description automatically generated

1.1-b Each row represents the marital status along with the total income in each category of marital status. The total income is highest in category of marital status ‘married’ (72416617).

* 1. Calculate Mean
     1. Mean age of responders born in Asia:



* + 1. Mean age of respondents born in Asia weighted by the number of children they have:

A picture containing text, screenshot, font, line

Description automatically generatedA screenshot of a computer

Description automatically generated with medium confidence

1.3 Table Comparison

1.3-a Mean score on the political awareness test for males compared to females:

A close-up of a computer screen

Description automatically generated with low confidence

1.3-b Each row represents the gender along with the mean score on the political awareness test in each category of gender. The mean score is highest in ‘female’ (0.0068008).

1.4 34th & 63rd percentiles of percentage of time taken on the test:

A picture containing text, font, screenshot, line

Description automatically generated

1. Organizing Data

2.1 Pie Chart

2.1-a Pie chart representing the number of respondents by Political Affiliation:

A screenshot of a computer program

Description automatically generated with medium confidence

A picture containing text, diagram, screenshot, font

Description automatically generated

2.1-b Conservative Political Affiliation contains the most respondents (701).

2.1-c Other Political Affiliation has the fewest respondents (220).

2.2 Summary Table

2.2-a Percentage of respondents from each region that are in the Treatment group:

A screenshot of a computer

Description automatically generated with medium confidence

2.2-b ‘North America’ (Control – 0.25802 & Treat – 0.25755) region has the highest percentage of people in the Treatment Group.

2.2-c ‘Southern’ (Control – 0.04670 & Treat – 0.05849) region has the lowest percentage of people in the Treatment Group.

2.3 Bar Chart

2.3-a Mean score on the Political Awareness Test for each region:

A screenshot of a computer code

Description automatically generated with low confidence

A picture containing text, screenshot, font, diagram

Description automatically generated

Mean score on the Standardized Test for each region:

A screenshot of a computer code

Description automatically generated with low confidence

A picture containing text, screenshot, font, number

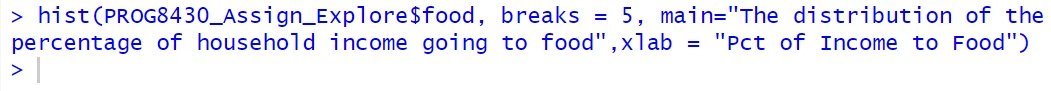
Description automatically generated

2.3-b ‘North America’ Region is the lowest mean score (1.019).

2.3-c ‘Southern’ Region is the highest mean score (1.062).

2.4 Histogram

2.4-a Histogram with 5 bins showing the distribution of the percentage of household income going to food.



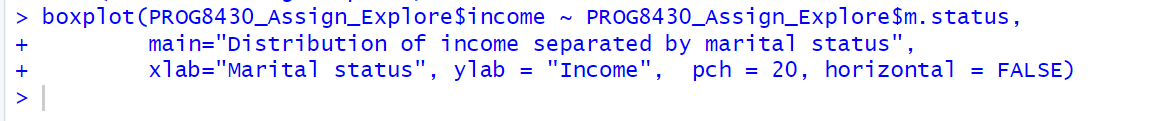
A picture containing text, screenshot, diagram, line

Description automatically generated

2.4-b ‘0.0 – 0.2’ range of values has the highest frequency.

2.5 Box Plots:

2.5-a A sequence of box plots showing the distribution of income separated by marital status:



A picture containing text, screenshot, diagram, rectangle

Description automatically generated

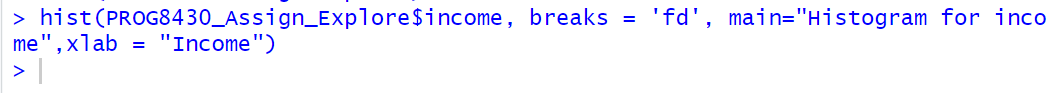
2.5-b ‘Married’ (170000) category of marital status has the highest average income.

2.5-c ‘Never’ & ‘divorced’ (0) category of marital status has the lowest average income.

2.5-d ‘Married’ category of marital status has the greatest variability in income.

2.6 Scatter Plots

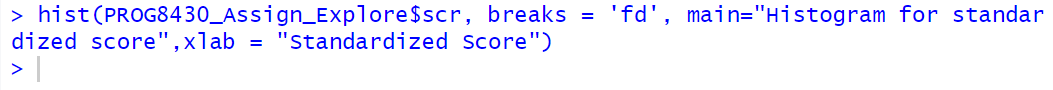
2.6-a Histogram for Income



A picture containing text, screenshot, diagram, font

Description automatically generated

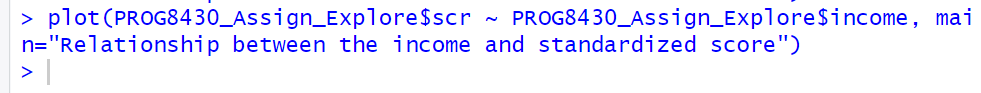
2.6-b Histogram for Standardized score



A picture containing text, diagram, line, plot

Description automatically generated

2.6-c Relationship between the income and standardized score



A picture containing text, screenshot, diagram

Description automatically generated

2.6-d Scatter plot represents the relationship between the income and standardized score. According to the above graph, the x & y values come closer forming the linear straight line and the y values started from the origin, and there is only very light increase. Therefore, this is Negative correlation.

2.6-e Based on the calculated correlation coefficient between the income and standardized score. Since the calculated correlation coefficient falls under this coefficient category < 0.50, it is a Negative & Weak linear association between the income and standardized score variables.

