**Assignment 1**

**Author:** Hamna Ashraf, Student # 8826836

**Course:** PROG8431

**Question 1: Summarizing Data**

**Question 1.1: Summarizing Table**

1. Create a table to show the total income by each category of marital status.

**Marital\_Status Income**

1 divorced 47305889

2 married 72416616

3 never 31039000

1. widowed 8788122
2. Which status has the highest total income?

married 72416616

**Question 1.2: Calculate the mean (rounded to two decimal places)**

1. Calculate the mean age of respondents born in Asia:

**Nation Age**

Asia 40.96

Mean age of respondents born in Asia is 40.96

1. Calculate the mean age of respondents born in Asia weighted by the number of children they have:

The weighted average is: 40.61039

**Question 1.3: Table Comparison**

1. Create a table to show the mean score on the political awareness test for males compared to females.

**Gender Score**

1 female 0.0068007663

2 male 0.0007063197

1. Which has a higher score?

Gender with the higher mean score: Female

**Question 1.4: Calculate the 34th and 63rd percentiles of percentage of time taken on the test.**

**34th** and **63rd** percentile**: 0.2** **0.61**

**Question 2: Organizing Data**

NOTE: All charts should have appropriated scaled axes and should be properly labelled.

**Question 2.1 Pie Chart:**

* 1. Create a pie chart showing the number of respondents by Political Affiliation.

A pie chart with a red circle and yellow circle with numbers

Description automatically generated

* 1. Which Political Affiliation contains the most respondents (remember each row of your study file represents one respondent)?

Max number of respondents are from: Conservative

* 1. Which Political Affiliation has the fewest respondents?

Min number of respondents are from: Other

**Question 2.2 Summary Table:**

1. Create a table that shows the percentage of respondents from each Region that are in the Treatment group.

**Nation Group Percentage**

1 Asia 242 22.51163

2 Europe 163 15.16279

3 North America 546 50.79070

4 Southern 124 11.53488

1. Which region has the highest percentage of people in the Treatment group?

**Nation Percentage**

North America 50.7907

1. Which region has the lowest percentage of people in the Treatment group?

**Nation Percentage**

Southern 11.53488

**Question 2.3 Bar Chart**

1. Create a bar chart showing the mean Standardized Test Score on the Political Awareness Test for each Region.

A graph of a number of rectangular objects

Description automatically generated with medium confidence

1. Which Region has the lowest mean score?

**Region Scr**

North America 1.019343

1. Which Region has the highest mean score?

**Region Scr**

Southern 1.062424

**Question 2.4 Histogram**

1. Create a histogram with 5 bins showing the distribution of the percentage of household income going to food.

A green graph with black text

Description automatically generated

1. Which range of values has the highest frequency?

Most people spend 0 to 20 percent of their income on food.

**Question 2.5 Box Plots**

1. Create a sequence of box plots showing the distribution of income separated by marital status.

A graph of a distribution of income by marital status

Description automatically generated

1. According to the charts, which marital status has the highest average income?

**Highest Average Income:**

married 85396.95

1. Which marital status has the lowest average income?

**Lowest Average Income:**

Divorced 63074.52

1. Which marital status has the greatest variability in income?

**Greatest income variability:**

never 44316.79

**Question 2.6 Scatter Plots**

1. Create a histogram for income.

A graph of green bars

Description automatically generated

1. Create a histogram for standardized score.

A graph of a bar graph

Description automatically generated

1. Create a scatter plot showing the relationship between the income and standardized score. (note: income should be on the x-axis, standardized score should be the y-axis)

A graph with blue dots

Description automatically generated

1. What conclusions, if any, can you draw from the chart?

In the chart, we can see that there is no moderate or strong correlation between ‘income’ and ‘standardized score’. If there was a correlation, then data might have been linearly distributed. With an increase in income, there should have been an increase in score. But this trend can’t be seen here in this graph.

Other than that, we can notice that there is a wide range income level.

1. Calculate a correlation coefficient between these two variables. What conclusion you draw from it?

The correlation between Income and Standardized Score is: 0.4568418

**Conclusion**:

As the correlation between the two variables is less than 0.5 therefore, it is proved that both income and standardized score are not related.