**Assignment 3 Submission Document**

**Group Members:**

Md Shoruv Hussain   
Matriculation Number: 5587329

Md Sohel Rana   
Matriculation Number: 5574894

# 1. Research Question & Hypothesis

**Research Question:**

“How do different child-rearing values and parental responsibilities influence respect and love for parents.”

Hypothesis:

There is no significant relationship between child-rearing values (Important child qualities) and respect and love for parents.

There is a significant relationship between child-rearing values (important child qualities) and respect and love for parents.

# 2. Data Overview

The dataset is a comprehensive collection of survey data compiled from the European Values Study (EVS) and the World Values Surveys (WVS). This dataset includes 452 surveys conducted across 115 countries and territories, offering a broad representation of social, cultural, and political values worldwide.

**Data Description**

* Dataset Name: WVS 1981-2022 trend file
* Source: worldvaluessurvey
* Data Dimension: **442473** x **732**
* Timeframe: 1981–2022
* Dimension of Selected Data: **442473** x **25**

## 2.1 Composition of the IVS 1981-2022

|  |  |  |  |
| --- | --- | --- | --- |
|  | **IVS** | **EVS Trend File** | **WVS Trend File** |
| **Survey period** | 1981-2022 | 1981-2017 | 1981-2022 |
| **Number of waves** | 7 | 5 | 7 |
| **Number of cases** | 663.965 | 224.434 | 442.473 |
| **Number of variables** | 838 | 635 | 732 |
| **Countries/ territories** | 120 | 49 | 108 |
| **Number of surveys** | 464 | 160 | 306 |

## 2.3 Selected Key Variables & Justification

Table 1:

|  |  |  |
| --- | --- | --- |
| Variable Name | Description | Role in Analysis |
| A025 | Respect and love for parents | Dependent |
| A001 | Important in life: Family | Independent |
| A005 | Important in life: Work | Independent |
| A006 | Important in life: Religion | Independent |
| A007 | Service to others important in life | Independent |
| A026 | Parents responsibilities to their children | Independent |
| A027 | Important child qualities: good manners | Independent |
| A029 | Important child qualities: independence | Independent |
| A030 | Important child qualities: hard work | Independent |
| A032 | Important child qualities: feeling of responsibility | Independent |
| A034 | Important child qualities: imagination | Independent |
| A035 | Important child qualities: tolerance and respect for other people | Independent |
| A038 | Important child qualities: thrift saving money and things | Independent |
| A039 | Important child qualities: determination perseverance | Independent |
| A040 | Important child qualities: religious faith | Independent |
| A041 | Important child qualities: unselfishness | Independent |
| A042 | Important child qualities: obedience | Independent |
| A047 | Abortion when child physically handicapped | Independent |
| A048 | Abortion when woman not married | Independent |
| A058 | Spend time with friends | Independent |
| A060 | Spend time with people at your church, mosque or synagogue | Independent |
| A064 | Belong to social welfare service for elderly, handicapped or deprived people | Independent |
| A065 | Member: Belong to religious organization | Independent |
| A066 | Member: Belong to education, arts, music or cultural activities | Independent |
| A170 | Satisfaction with your life | Independent |

## 2.4 Data Cleaning & Preprocessing

Handling Missing Values

In the selected dataset, there were a number of missing values. These missing values were imputed using the most frequent values for each variable. This method ensures that the missing data is replaced with the value that occurs most frequently in the respective column, preserving the overall distribution of the data.

# 3. Descriptive Statistics & Visualizations

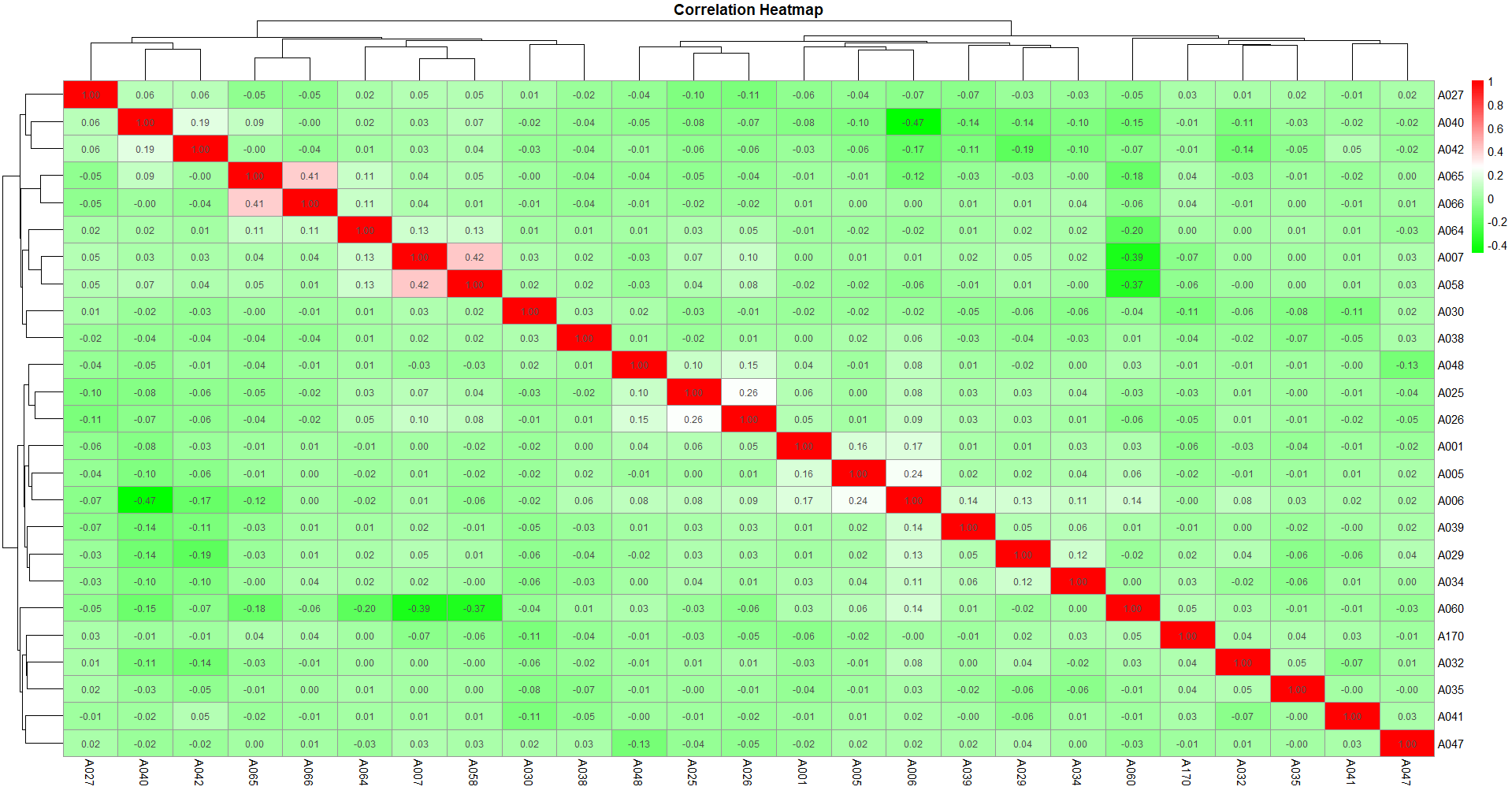
## 3.1 Descriptive Statistics of the Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Frequencies | | |
| A025 | | 1 | 419439 |
| 2 | 23034 |
| A001 | | 1 | 401312 |
| 2 | 36193 |
| 3 | 3909 |
| 4 | 1059 |
| A005 | | 1 | 290981 |
| 2 | 112539 |
| 3 | 26107 |
| 4 | 12846 |
| A006 | | 1 | 225174 |
| 2 | 95215 |
| 3 | 72278 |
| 4 | 49806 |
| A007 | | 1 | 414116 |
| 2 | 21171 |
| 3 | 5776 |
| 4 | 1410 |
| A026 | | 1 | 401888 |
| 2 | 31929 |
| 3 | 8656 |
| A027 | | 0 | 5391 |
| 1 | 388557 |
| A029 | | 0 | 238400 |
| 1 | 204073 |
| A030 | | 0 | 193869 |
| 1 | 248604 |
| A032 | | 0 | 138620 |
| 1 | 303853 |
| A034 | | 0 | 351064 |
| 1 | 303853 |
| A035 | | 0 | 351064 |
| 1 | 91409 |
| A038 | | 0 | 286099 |
| 1 | 156374 |
| A039 | | 0 | 286835 |
| 1 | 155638 |
| A040 | | 0 | 273008 |
| 1 | 169465 |
| A041 | | 0 | 310430 |
| 1 | 132043 |
| A042 | | 0 | 273318 |
| 1 | 169155 |
| A047 | | 0 | 6119 |
| 1 | 436354 |
| A048 | | 0 | 435614 |
| 1 | 6859 |
| A058 | | 1 | 414955 |
| 2 | 17384 |
| 3 | 6938 |
| 4 | 3196 |
| A060 | | 1 | 16086 |
| 2 | 8682 |
| 3 | 9219 |
| 4 | 408486 |
| A064 | | 0 | 438312 |
| 1 | 4161 |
| A065 | | 0 | 395431 |
| 1 | 47042 |
| A066 | | 0 | 416229 |
| 1 | 26244 |
| A170 | | 5 | 60272 |
| 6 | 48722 |
| 7 | 66430 |
| 8 | 86720 |
| 9 | 44383 |
| 10 | 61717 |
| Other | 74229 |
|  | | Total | 442,473 |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | | |

## 3.2 Correlation among the variables

This correlation heatmap visualizes the relationships between different variables, with red indicating positive correlations and green indicating negative correlations. The intensity of the color corresponds to the strength of the correlation, and the values within the cells provide the correlation coefficients. The dendrograms along the top and left sides reveal clustering patterns among the variables based on their correlation profiles. The heatmap can be used to identify variables that tend to move together or in opposite directions, offering insights into potential dependencies within the dataset.



# 4. Linear Regression Analysis

After fitting the linear regression model, we evaluated its overall performance using the R-squared (R²) value, which indicates how well the model explains the variation in the dependent variable. In our case, the R² value was 0.08921 meaning that the model explained 8.921% of the variance in the outcome. A higher R² indicates a better fit, suggesting that the model does a good job of capturing the relationship between the predictors and the dependent variable.

## 4.1 Coefficients and Interpretation

In this linear regression analysis, several predictors were evaluated to understand their relationship with the dependent variable. The results show that many of the variables have a significant impact, with their coefficients either positively or negatively affecting the outcome. For instance, the intercept is estimated at 0.990, indicating that when all other predictors are zero, the dependent variable is expected to be around 0.99. Among the predictors, A026 has a large positive effect with a coefficient of 0.132, suggesting that for every one-unit increase in A026, the dependent variable increases by 0.132. On the other hand, A027 shows a negative effect with a coefficient of -0.051, meaning that as A027 increases, the dependent variable decreases by 0.051.

The standard errors of the estimates are generally small, indicating precise estimates for most variables. For example, A026 has a very small standard error of 0.000887, suggesting that its estimate is highly reliable. The t-values, which reflect the ratio of each coefficient to its standard error, are all high, with A026 having a t-value of 148.776, demonstrating its strong significance.

Regarding statistical significance, most variables show highly significant results with p-values less than 2e-16, such as A026, A048, and A001, which all have three asterisks (\*\*\*), indicating their strong influence on the outcome. Variables like A029 and A032 have p-values of 0.0186 and 0.0019, respectively, showing that they are still significant, though their effects are somewhat less pronounced. In total, the results suggest that the majority of the predictors are significant, with a mix of positive and negative relationships with the dependent variable.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Coefficients** | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** | **Significance** |
| (Intercept) | 0.990002 | 0.004788 | 206.76 | < 2e-16 | \*\*\* |
| A026 | 0.132007 | 0.000887 | 148.77 | < 2e-16 | \*\*\* |
| A027 | -0.05122 | 0.000993 | -51.57 | < 2e-16 | \*\*\* |
| A029 | 0.001573 | 0.000669 | 2.352 | 0.018651 | \* |
| A030 | -0.01611 | 0.000661 | -24.37 | < 2e-16 | \*\*\* |
| A032 | -0.00219 | 0.000704 | -3.108 | 0.001883 | \*\* |
| A034 | 0.010103 | 0.000806 | 12.54 | < 2e-16 | \*\*\* |
| A035 | -0.00265 | 0.000688 | -3.853 | 0.000117 | \*\*\* |
| A038 | -0.01461 | 0.000676 | -21.60 | < 2e-16 | \*\*\* |
| A039 | -0.00307 | 0.000684 | -4.49 | 7.14E-06 | \*\*\* |
| A040 | -0.01631 | 0.000764 | -21.36 | < 2e-16 | \*\*\* |
| A041 | -0.00679 | 0.000708 | -9.59 | < 2e-16 | \*\*\* |
| A042 | -0.01499 | 0.000696 | -21.53 | < 2e-16 | \*\*\* |
| A047 | -0.02965 | 0.002765 | -10.72 | < 2e-16 | \*\*\* |
| A048 | 0.111625 | 0.002642 | 42.255 | < 2e-16 | \*\*\* |
| A001 | 0.017882 | 0.00092 | 19.447 | < 2e-16 | \*\*\* |
| A005 | -0.00538 | 0.000452 | -11.91 | < 2e-16 | \*\*\* |
| A006 | 0.006553 | 0.000364 | 18.011 | < 2e-16 | \*\*\* |
| A007 | 0.025353 | 0.001059 | 23.946 | < 2e-16 | \*\*\* |
| A058 | 0.003615 | 0.000915 | 3.952 | 7.76E-05 | \*\*\* |
| A060 | -0.00691 | 0.000591 | -11.693 | < 2e-16 | \*\*\* |
| A064 | 0.041245 | 0.003404 | 12.116 | < 2e-16 | \*\*\* |
| A065 | -0.02573 | 0.001166 | -22.073 | < 2e-16 | \*\*\* |
| A066 | -0.00949 | 0.00149 | -6.369 | 1.90E-10 | \*\*\* |
| A170 | -0.00129 | 0.000135 | -9.579 | < 2e-16 | \*\*\* |

# 5. Conclusion