**PROJECT Planning**

|  |  |
| --- | --- |
| **Class Section** | STAT-614-001.2025S Statistic Methods |
| **GROUP MEMBERS** | Qintao Jia, Anna Hyunjung Kim |
| **TITLE** | World Happiness Report |

|  |  |  |
| --- | --- | --- |
| **DATA set** (detailed description of your data set) with list of variables you will use (mentioning units of measurements)  Data source: Sazidthe1. (2023). World Happiness Report (till 2023) [Data set]. Kaggle. <https://www.kaggle.com/datasets/sazidthe1/global-happiness-scores-and-factors>  Data Originally from <https://worldhappiness.report/>   |  |  | | --- | --- | | 1. The World Happiness Report (2020)    1. 153 unique observations    2. 9 columns       1. country       2. region       3. happiness\_score       4. gdp\_per\_capita       5. social\_support       6. healthy\_life\_expectancy       7. freedom\_to\_make\_life\_choice       8. generosity       9. perceptions\_of\_corruption | 1. The World Happiness Report (2015)    1. 158 unique observations    2. 9 columns       1. country       2. region       3. happiness\_score       4. gdp\_per\_capita       5. social\_support       6. healthy\_life\_expectancy       7. freedom\_to\_make\_life\_choice       8. generosity       9. perceptions\_of\_corruption | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Description** | **Type** | **Unit** |
| country | Name of the country | Categorical | N/A |
| region | Region the country belongs to | Categorical | N/A |
| happiness\_score | Average life evaluation score based on the Cantril ladder (0–10 scale) | Quantitative | Score (0–10) |
| gdp\_per\_capita | Log of GDP per capita based on Purchasing Power Parity (PPP) in constant 2021 international dollars | Quantitative | Log(USD, PPP-adjusted) |
| social\_support | National average of the binary response (0 = no, 1 = yes) to the question: “Do you have someone to rely on in times of trouble?” | Quantitative (proportion) | Proportion (0–1) |
| healthy\_life\_expectancy | Life expectancy at birth adjusted for health quality | Quantitative | Years |
| freedom\_to\_make\_life\_choices | National average of responses to: “Are you satisfied with your freedom to choose what you do with your life?” (0 = no, 1 = yes) | Quantitative (proportion) | Proportion (0–1) |
| generosity | Residual from regressing donation behavior on GDP per capita; measures prosocial behavior | Quantitative | Index (no unit) |
| perceptions\_of\_corruption | National average of responses to corruption-related questions (0 = not widespread, 1 = widespread) | Quantitative (proportion) | Proportion (0–1) |

|  |  |
| --- | --- |
| **Research questions** | |
| **Q1** | **Do the happiness scores of people on different continents in 2015 differ significantly** |
| **Q2** | **Have the happiness scores of people in various countries changed between 2015 and 2020 (before and after the COVID-19)** |
| **Q3** | **How do the different explanatory variables influence the happiness score in 2015** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **TENTATIVE Plan of Action (techniques to be used to address each question listed above)**   1. **We will use descriptive statistics and boxplots to explore the data. These summaries will help us understand the distribution, variability, and outliers in the data.** 2. **We will use the Shapiro-Wilk test and Q-Q plots to assess whether the variables are normally distributed (Normality). Moreover, we will use Levene’s test for ANOVA to verify homogeneity of variances. Lastly, Residual vs. fitted plots and Q-Q plots of residuals will be used for linearity and normality of residuals.** 3. **Statistical procedures.**  |  |  | | --- | --- | | **Q1** | **ANOVA** | | **Q2** | **Pair t-test** | | **Q3** | **Multiple regression** | |

|  |  |
| --- | --- |
| **Group members** | **Responsibilities** |
| Qintao Jia | Project plan |
| Anna Hyunjung Kim | Search for dataset, Project plan |

