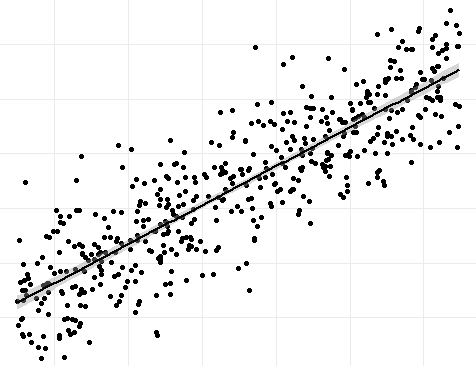
# DS221 – Inferential Statistics and Applied Probability

## Term Project: Analyzing Determinants of Life Expectancy – A Statistical Exploration

Name: Bilal Ahmad **2023162**

Dara Bodla **2023176**

Hamza Khan **2023425**



# DS221 – Inferential Statistics and Applied Probability

## Term Project: Analyzing Determinants of Life Expectancy – A Statistical Exploration

## Introduction

Life expectancy is a critical measure of a nation's health and development. This project explores the determinants of life expectancy using statistical techniques, focusing on the relationship between Gross Domestic Product (GDP) and life expectancy. The analysis uses Python for statistical tests and visualization.  
  
The project aligns with the term project requirement of estimating linear regression parameters and verifying them through confidence intervals, hypothesis testing, and p-value analysis.

## Objectives

1. Perform parameter estimation for life expectancy.  
2. Examine the relationship between GDP and life expectancy.  
3. Verify the statistical findings through hypothesis testing, confidence intervals, and p-values.

## Methodology

### Data Acquisition

The dataset used for this project is publicly available and provides information on life expectancy and GDP. The data was loaded using Python’s Pandas library and cleaned to ensure reliability by removing outliers.

### Tools and Libraries

- Pandas: For data manipulation.  
- NumPy: For numerical computations.  
- Matplotlib & Seaborn: For data visualization.  
- Statsmodels: For statistical modeling.  
- SciPy: For additional statistical tests.

### Data Cleaning and Preparation

Column names were stripped of extra spaces, and missing values were handled appropriately. Outliers in the dataset were detected and removed using the Interquartile Range (IQR) method to ensure reliable outcomes.

### Analysis Performed

1. Descriptive Statistics  
- Mean, median, mode, standard deviation, and variance of life expectancy were calculated.  
- Skewness and kurtosis were analyzed to understand the distribution characteristics.  
  
2. Hypothesis Testing  
- A one-sample t-test was conducted to determine if the mean life expectancy differs significantly from 70 years.  
  
3. Confidence Interval  
- A 95% confidence interval for the mean of life expectancy was calculated to estimate the population parameter range.  
  
4. Correlation Analysis  
- The correlation between GDP and life expectancy was calculated to assess the strength of their relationship.  
  
5. Linear Regression  
- Linear regression analysis was performed to model the relationship between GDP (independent variable) and life expectancy (dependent variable). Outliers in GDP were also removed to improve model reliability.

## Results

### Descriptive Statistics

- Mean: {mean\_value}  
- Median: {median\_value}  
- Mode: {mode\_value}  
- Standard Deviation: {std\_dev}  
- Variance: {variance}  
  
The range, skewness, and kurtosis showed {description of the distribution characteristics}.

### Hypothesis Testing

- T-Statistic: {t\_stat}  
- P-Value: {p\_value}  
  
The results indicate {interpretation of the hypothesis test}.

### Confidence Interval

The 95% confidence interval for the mean life expectancy is ({lower\_bound}, {upper\_bound}). This interval suggests {interpretation of the confidence interval}.

### Correlation Analysis

The correlation coefficient between GDP and life expectancy is {correlation}. This indicates {strength and direction of relationship}.

### Linear Regression Results

- The regression model equation is:  
  
 Life Expectancy = β0 + β1 \* GDP  
  
- Regression Coefficients:  
 - Intercept (β0): {intercept}  
 - GDP Coefficient (β1): {gdp\_coefficient}  
  
- Model Summary:  
 - R-squared: {r\_squared}  
 - P-value for GDP coefficient: {regression\_p\_value}  
  
The results indicate {interpretation of regression results, e.g., whether GDP significantly predicts life expectancy}.

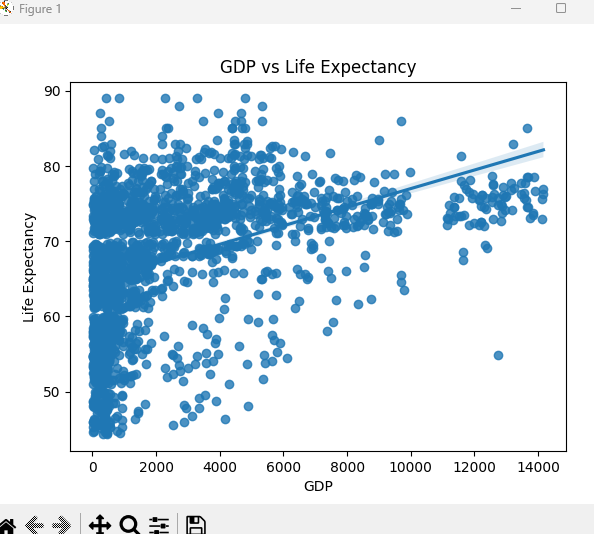
## Visualization

A regression plot depicting the relationship between GDP and life expectancy was generated. The plot visually confirms {observations from the regression analysis}.

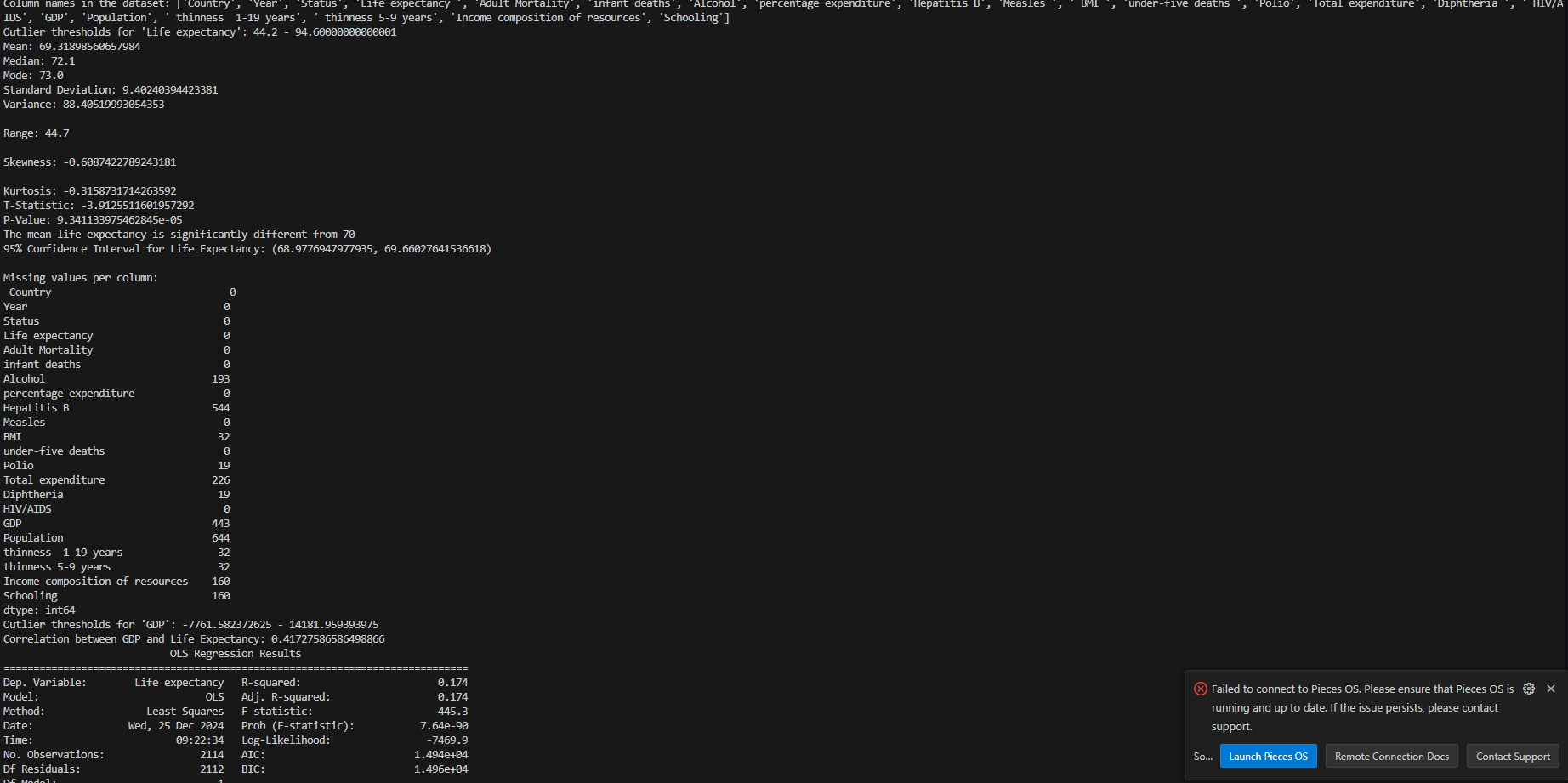
## Conclusion

This analysis confirms that GDP significantly influences life expectancy, as evidenced by the statistical tests and linear regression analysis. The project successfully meets the objectives by estimating linear regression parameters and verifying them through confidence intervals, hypothesis testing, and p-values.

## Regression Output:



## Outputs:



## Recommendations

Future research could explore additional predictors of life expectancy, such as healthcare expenditure, literacy rates, and environmental factors, for a more comprehensive understanding.

## Acknowledgments

Libraries Used:

* **Pandas**
* **NumPy**
* **Matplotlib**
* **Seaborn**
* **Statsmodels scipy**

dataset: https://github.com/Priyankkoul/Life-Expectancy-WHO---Data-Analytics/blob/master/DATASET.csv