**DS201**

**Statistical Programming**

**Assignment 9**

**Ashutosh Kumar Jha**

**12340390**

**DSAI**

**2nd Year**

**Semester 4**

**---------------------------------------------------------------------------**

**Question 1:** Vehicle Features Impact on Fuel Efficiency: A Multiple Linear Regression Analysis

**Introduction:** This report investigates how vehicle characteristics influence fuel efficiency, measured in miles per gallon (MPG). Understanding these relationships is crucial for automotive engineers, manufacturers, and consumers interested in fuel economy. The analysis examines three key predictors: engine size (L), vehicle weight (kg), and horsepower, to determine their individual and combined effects on MPG using multiple linear regression modeling.

**Data:**

The dataset consists of 20 vehicles with the following attributes:

* **Engine Size**: Ranging from 1.3L to 3.5L
* **Weight**: Ranging from 1020 kg to 1700kg
* **Horsepower**: Ranging from 98 to 200
* **MPG**: Ranging from 18 to 39

Initial data exploration revealed strong negative correlations between MPG and all three predictor variables. Engine size (r = -0.86), weight (r = -0.90), and horsepower (r = -0.91) all exhibit inverse relationships with fuel efficiency, suggesting that as these values increase, MPG decreases.

**Methodology:**

I applied multiple linear regression analysis to model the relationship between the three independent variables (engine size, weight, and horsepower) and the dependent variable (MPG). The model takes the form:

MPG = β₀ + β₁(Engine Size) + β₂(Weight) + β₃(Horsepower) + ε

Where:

* β₀ is the intercept
* β₁, β₂, β₃ are the coefficients for each predictor
* ε is the error term

The regression was performed using the Ordinary Least Squares (OLS) method from the statsmodels package in Python. I also conducted residual analysis to verify model assumptions and calculated variance inflation factors (VIF) to check for multicollinearity among predictors.

**Results:**

The multiple regression model yielded the following equation:

MPG = 59.8406 - 1.0617 × Engine\_Size - 0.0108 × Weight - 0.0777 × Horsepower

**Key Statistics:**

* R-squared: 0.892 (89.2% of the variation in MPG is explained by the model)
* Adjusted R-squared: 0.872

**Coefficients and p-values:**

* Intercept: 59.8406 (p-value: <0.0001)
* Engine Size: -1.0617 (p-value: 0.2898)
* Weight: -0.0108 (p-value: 0.0312)
* Horsepower: -0.0777 (p-value: 0.0468)

**Significance at different levels:**

* At α = 0.01: No variables are statistically significant
* At α = 0.05: Weight and Horsepower are statistically significant
* At α = 0.10: Weight and Horsepower are statistically significant

Residual analysis showed reasonably well-behaved residuals with no strong patterns, supporting the validity of the linear regression assumptions.

**Discussion:**

The model demonstrates that all three vehicle characteristics have negative relationships with MPG, which aligns with engineering principles. Heavier vehicles require more energy to move, and more powerful engines typically consume more fuel.

At the conventional significance level (α = 0.05), both weight and horsepower are statistically significant predictors of MPG, while engine size is not. This is interesting because engine size has a strong correlation with MPG in isolation, but its effect becomes statistically insignificant when accounting for the other variables.

The non-significance of engine size may be due to multicollinearity, as engine size is typically correlated with horsepower. The analysis revealed moderate multicollinearity among predictors (VIF values between 5-10), suggesting some redundancy in the information provided by these variables.

For every 100 kg increase in vehicle weight, MPG decreases by approximately 1.08, holding other factors constant. Similarly, for every 10-unit increase in horsepower, MPG decreases by approximately 0.78, all else being equal.

**Conclusion:**

The multiple regression analysis confirms that vehicle weight and horsepower are significant predictors of fuel efficiency. Together with engine size, these variables explain nearly 90% of the variation in MPG, making this model valuable for predicting fuel economy.

For manufacturers aiming to improve fuel efficiency, reducing vehicle weight appears to be a statistically significant approach. While horsepower also significantly impacts MPG, engine size alone may not be the primary focus if weight and power requirements are already optimized.

These findings provide quantitative support for the automotive industry's ongoing efforts to develop lighter vehicles with more efficient power delivery systems. Future research could expand this analysis by incorporating additional variables such as aerodynamic design, transmission type, and driving conditions to create more comprehensive fuel efficiency prediction models.

**---------------------------------------------------------------------------**

**Question 2:** Parents' Heights Influence on Son's Height: Multiple Linear Regression Analysis

**Introduction:** This report investigates the relationship between parents' heights and their son's height using multiple linear regression analysis. The study aims to determine the extent to which a son's height can be predicted from his parents' heights and to test the theory of regression toward the mean—the statistical phenomenon where extreme values tend to move closer to the average in subsequent measurements.

**Data:**

The dataset consists of height measurements (in inches) from 10 families:

* Father's heights range from 60 to 74 inches (average: 66.8 inches)
* Mother's heights range from 61 to 69 inches (average: 65.2 inches)
* Son's heights range from 63.6 to 70.1 inches (average: 67.01 inches)

The data shows positive correlations between son's height and both father's height (r = 0.93) and mother's height (r = 0.86).

**Methodology:**

I used multiple linear regression to model the relationship between the independent variables (father's and mother's heights) and the dependent variable (son's height). The model takes the form:

Son's Height = β₀ + β₁(Father's Height) + β₂(Mother's Height) + ε

Where:

* β₀ is the intercept
* β₁ is the coefficient for father's height
* β₂ is the coefficient for mother's height
* ε is the error term

To test for regression toward the mean, I performed one-tailed hypothesis tests to determine whether the regression coefficients are significantly less than 1. The hypotheses were:

H₀: β₁ ≥ 1 vs. H₁: β₁ < 1 (for father's height) H₀: β₂ ≥ 1 vs. H₁: β₂ < 1 (for mother's height)

I also analyzed the residuals to check the model's assumptions and validity.

**Results:**

The multiple regression model yielded the following equation:

Son's Height = 14.1741 + 0.4007 × Father's Height + 0.4292 × Mother's Height

**Key Statistics:**

* R-squared: 0.885 (88.5% of the variation in son's height is explained by parents' heights)
* Adjusted R-squared: 0.852

**Coefficients and p-values:**

* Intercept: 14.1741 (p-value: 0.0235)
* Father's Height: 0.4007 (p-value: 0.0061)
* Mother's Height: 0.4292 (p-value: 0.0271)

**Tests for Regression toward the Mean:**

* Father's Height Coefficient < 1: p-value = 0.0012 (significant at α = 0.05)
* Mother's Height Coefficient < 1: p-value = 0.0096 (significant at α = 0.05)

The residual plots showed no clear patterns, indicating that the linear model is appropriate.

**Discussion:  
  
1. Interpretation of Coefficients**

The regression coefficients indicate that:

* For each additional inch in the father's height, the son's height increases by approximately 0.40 inches, holding the mother's height constant.
* For each additional inch in the mother's height, the son's height increases by approximately 0.43 inches, holding the father's height constant.
* Both parents appear to contribute similarly to their son's height, with the mother's height having a slightly larger coefficient.

### **2. Evidence of Regression toward the Mean**

The analysis strongly supports the concept of regression toward the mean in height inheritance. Both coefficients (0.40 for father and 0.43 for mother) are significantly less than 1, confirming that sons of unusually tall or short parents tend to be closer to the average height than their parents.

For example:

* For parents both at 62 inches (below average), the predicted son's height is 64.77 inches, which is closer to the average son height (67.01 inches) than the parents are to their respective averages.
* For tall parents (father: 72 inches, mother: 68 inches), the predicted son's height is 69.74 inches, which again shows movement toward the mean.

The average "regression ratio" (how much of the parents' deviation from average is inherited by the son) is approximately 0.83, indicating that a son inherits about 83% of his parents' deviation from the mean height.

**Conclusion:**

This analysis confirms that both parents' heights are significant predictors of their son's height, together explaining about 88.5% of the variation. The data strongly supports the theory of regression toward the mean in height inheritance, with sons of unusually tall or short parents tending to be closer to average height than their parents.

The model provides insights into genetic inheritance patterns and could be used for height predictions. However, the relatively small sample size (10 families) should be considered when generalizing these findings. Future research could expand this analysis by including daughters, considering environmental factors such as nutrition, and investigating other hereditary traits to better understand generational patterns in human development.

These findings have implications for both genetic counseling and understanding the complex interplay of genetics and regression toward the mean in human traits.

Code: [12340390 Ashutosh Asg9](https://colab.research.google.com/drive/1S54gpNBSS_ZUQZiNY0y-aovRcjGQ-r71?usp=sharing)