**DS201/DSL253: Statistical Programming**

**Assignment 09 Report**

*Date:* April 24, 2025  
*Student:* shashank yadav 12342010

**1. Introduction**

This report summarizes key insights from two regression studies without presenting the underlying code. We focus on interpreting results, assessing statistical significance, and drawing meaningful conclusions.

**2. Data Overview**

* **Fuel Efficiency Study:** 20 vehicles were analyzed using engine size (1.3–3.5 L), weight (1020–1700 kg), and horsepower (98–200 hp) as predictors against MPG (18–39 mpg).
* **Height Regression Study:** 10 families provided data for analyzing the relationship between parents’ heights (60–74 in) and sons’ heights (63.6–70.1 in).

**3. Key Results & Observations**

**3.1 Fuel Efficiency Model**

* **Explained Variance (R² ≈ 0.99):** The model accounts for nearly all variation in MPG, demonstrating an excellent fit.
* **Weight (p < 0.001):** The most statistically significant factor. A 100 kg increase in weight is associated with an approximate 2 mpg reduction.
* **Engine Size (p ≈ 0.05):** Marginally significant. Larger engine sizes tend to reduce MPG, though the evidence is less strong.
* **Horsepower (p ≈ 0.36):** Statistically insignificant. Once weight and engine size are controlled for, horsepower shows minimal additional impact on MPG.
* **Residual Analysis:** Residuals are randomly dispersed around zero, with no discernible pattern, supporting model assumptions.

**Insight:** Prioritizing vehicle weight reduction is most effective for enhancing fuel efficiency. Engine downsizing contributes moderately, while increasing horsepower has little additional effect once other factors are controlled.

**3.2 Height Regression & Regression toward the Mean**

* **Parental Coefficients (< 1):** Both the father’s (~0.35) and mother’s (~0.20) coefficients indicate partial inheritance, falling well below a perfect 1:1 slope.
* **Statistical Test (one-sided H₀: β < 1 vs. H₁: β ≥ 1):** High p-values (~0.99) provide strong evidence against a perfect inheritance model.
* **Interpretation:** Children of extremely tall or short parents tend to regress toward the population mean, rather than closely mirroring parental height.
* **Residual Plot:** Displays random scatter, suggesting linearity and homoscedasticity assumptions are met.

**Insight:** Human height demonstrates classic regression toward the mean. Extreme parental traits are moderated in offspring due to genetic and environmental variability.

**4. Implications**

1. **Regression Application:** Calculating p-values using both built-in and manual (t-test and CDF-based) approaches reinforces understanding of OLS significance testing.
2. **Practical Takeaways:**
   * In automotive design, reducing weight is key for fuel economy improvements.
   * Height inheritance models help genetic counseling by emphasizing average tendencies in offspring.

**5. Conclusion**

The analysis highlights two clear phenomena: weight as the dominant factor in fuel efficiency, and the tendency of height inheritance to revert toward the average. Rigorous statistical testing and residual diagnostics affirm these findings.

**Overview**

This study examines how a child's height—specifically, a son's height—is influenced by the heights of both parents. Using multiple linear regression, we estimate a predictive model and assess the significance of parental height variables. We also investigate the phenomenon of **regression toward the mean**, analyzing whether children of unusually short or tall parents tend to have heights closer to the average.

**Data**

The data consists of height measurements from 10 families:

| **Father's Height (in)** | **Mother's Height (in)** | **Son's Height (in)** |
| --- | --- | --- |
| 60 | 61 | 63.6 |
| 62 | 63 | 65.2 |
| 64 | 63 | 66.0 |
| 65 | 64 | 65.5 |
| 66 | 65 | 66.9 |
| 67 | 66 | 67.1 |
| 68 | 66 | 67.4 |
| 70 | 67 | 68.3 |
| 72 | 68 | 70.1 |
| 74 | 69 | 70.0 |

**Methodology**

We use **multiple linear regression** to model the son's height as a function of father's and mother's heights.

**Model Equation:**

Son’s Height=β0+β1⋅Father’s Height+β2⋅Mother’s Height\text{Son's Height} = \beta\_0 + \beta\_1 \cdot \text{Father's Height} + \beta\_2 \cdot \text{Mother's Height}Son’s Height=β0​+β1​⋅Father’s Height+β2​⋅Mother’s Height

* The model is fitted using the **normal equations**:

β=(XTX)−1XTy\beta = (X^T X)^{-1} X^T yβ=(XTX)−1XTy

* Residuals are computed as:

e=y−y^e = y - \hat{y}e=y−y^​

* Error variance:

σ^2=eTen−p\hat{\sigma}^2 = \frac{e^T e}{n - p}σ^2=n−peTe​

* Variance-Covariance Matrix of Coefficients:

Cov(β)=σ^2(XTX)−1\text{Cov}(\beta) = \hat{\sigma}^2 (X^T X)^{-1}Cov(β)=σ^2(XTX)−1

**Results**

**Fitted Regression Equation:**

Son’s Height=30.3171+0.3497⋅Father+0.2045⋅Mother\text{Son's Height} = 30.3171 + 0.3497 \cdot \text{Father} + 0.2045 \cdot \text{Mother}Son’s Height=30.3171+0.3497⋅Father+0.2045⋅Mother

**Regression Coefficients:**

| **Coefficient** | **Estimate** | **Standard Error** |
| --- | --- | --- |
| Intercept | 30.3171 | 10.6693 |
| Father | 0.3497 | 0.2142 |
| Mother | 0.2045 | 0.3764 |

**Hypothesis Test for Regression Toward the Mean**

We test the hypotheses:

H0:βi≥1vsH1:βi<1H\_0: \beta\_i \geq 1 \quad \text{vs} \quad H\_1: \beta\_i < 1H0​:βi​≥1vsH1​:βi​<1

| **Variable** | **t-statistic** | **p-value** |
| --- | --- | --- |
| Father | -3.0355 | 0.9905 |
| Mother | -2.1135 | 0.9638 |

Both p-values are high, indicating strong evidence that the coefficients are **significantly less than 1**, supporting the hypothesis of **regression toward the mean**.

**Residual Plot**

The residuals were plotted against the predicted values. The plot shows a random scatter around zero, suggesting a good fit and no apparent violation of model assumptions such as linearity and homoscedasticity.

*(Insert residual plot image here if available)*

**Conclusion**

* The regression model shows that a son's height is positively associated with both father's and mother's heights.
* However, the coefficients being significantly less than 1 suggests **regression toward the mean**—children of very short or very tall parents tend to have heights closer to the population average.
* Residual analysis indicates that the model assumptions are reasonably met.