34. **Analysis of Dosage Prediction using Multiple Linear Regression Models in Healthcare Data**

**Abstract:**

This research paper investigates the effectiveness of multiple linear regression models in predicting the dosage of medication based on age, gender, and BMI. We employ statistical tools to examine the relationship between the predictors and the response variable (dosage of medication) and evaluate the models' accuracy and validity using residual plots and correlation matrices. The results highlight the importance of variable selection and model diagnostics in understanding the underlying data patterns and improving prediction accuracy.

**Introduction**

Accurate prediction of medication dosage is critical in healthcare to ensure patient safety and therapeutic efficacy. Traditional regression models, like multiple linear regression, offer a way to understand the relationships between predictors (such as age, gender, and BMI) and the response variable (dosage). This study aims to evaluate the performance of linear regression models in predicting medication dosage and to understand the significance of different predictors through various statistical analyses.

**Methods**

1. **Dataset Description**: The dataset includes variables such as Age, Gender, BMI, and Dosage of Medication (mg). The target variable for prediction is 'Dosage of Medication.'
2. **Statistical Models**:
   * Multiple linear regression models were employed to assess the relationship between the predictors (Age, Gender, and BMI) and the dosage.
   * Model diagnostics included the examination of residuals and predicted values to evaluate model fit and the presence of any biases or patterns.
3. **Correlation Analysis**: A correlation matrix was constructed to examine the strength and direction of relationships among predictors.
4. **Validation Methods**:
   * Variance Inflation Factor (VIF) was calculated to assess multicollinearity among predictors.
   * Residual plots were analyzed to check for any systematic patterns or heteroscedasticity that might indicate model inadequacy.

**Results**

1. **Model Coefficients and Diagnostics**:
   * The linear regression model shows the coefficient estimates for Age (0.111), Gender, and BMI. The results suggest that none of the predictors, except for the intercept, are statistically significant at the 0.05 level, as indicated by their p-values (e.g., p = 0.210 for Age).
2. **Correlation Analysis**:
   * The correlation matrix shows a perfect positive correlation (1) between Age and BMI, indicating redundancy and potential multicollinearity. This finding suggests that including both predictors in the model may not contribute additional explanatory power.
3. **Residual Analysis**:
   * The residual plots reveal a random distribution around zero, suggesting no apparent patterns or trends. This implies that the model errors are homoscedastic and normally distributed, which meets the assumptions of linear regression.
   * However, the spread of residuals indicates a relatively high variance, which may point to a lack of fit or the need for a more complex model.
4. **Actual vs. Predicted Dosage**:
   * The scatter plot of actual vs. predicted dosage demonstrates that the model tends to underpredict and overpredict without a clear pattern, indicating that the linear model may not capture the full complexity of the underlying data relationships.
5. **Variance Inflation Factor (VIF)**:
   * The VIF values are close to 1, indicating that multicollinearity is not a significant issue despite the high correlation between Age and BMI, likely due to the relatively small coefficients associated with each variable.

**Discussion**

The analysis reveals that the linear regression model has limited predictive power for the dosage of medication based on the available predictors (Age, Gender, and BMI). The high p-values for each predictor suggest that none significantly contributes to explaining the variation in dosage, and the correlation matrix confirms a redundancy between Age and BMI.

The residuals' pattern indicates that while the assumptions of homoscedasticity and normality hold, the model may lack the complexity needed to capture all relevant factors influencing dosage. The scatter plot of actual vs. predicted values further supports this by showing widespread dispersion around the predicted line.

**Conclusion**

This study demonstrates the limitations of using simple linear models to predict medication dosage in healthcare settings. The results suggest that other factors not included in the current model may have more significant predictive power. Future research should explore alternative models, such as machine learning techniques or non-linear models, to enhance predictive accuracy.