Aditya Parey

ALY 6010 Summer 2024

practice module 6

Statical Outputs:

A screenshot of a computer screen

Description automatically generated

Dummy Variables and Linear Regression Model

A screenshot of a computer

Description automatically generated

Multiple Linear Regression Plot for lung capacity vs height by gender:

A graph with red and blue dots

Description automatically generated

Regression Lines for lung capacity vs height of males:

A graph with blue dots

Description automatically generated

Regression lines for lung capacity vs height of females:

A graph with a line and a blue line

Description automatically generated

Interpretation:

Subsets and Lines: We can effectively establish two subsets (male and female) by creating dummy variables for gender. The multiple regression line scatterplot (one for each gender) illustrates how gender influences lung capacity in relation to other variables such as height.

Effect on Comprehension: We can see that the gender differences in the association between the predictors (height, age) and the dependent variable (lung\_cap) by adding the categorical variable as a dummy.

Variations in Regression Lines: The slope and intercept of the regression lines for men and women may vary. This suggests that there are gender-specific differences in the association between lung capacity and variables such as age and height.

Impact on understanding: We can learn more about how gender, a categorical variable, modifies the relationship between the predictors and the outcome by examining distinct regression lines for each subset. This method gives you a more detailed picture of the data and lets you determine whether predictor effects vary depending on the group.

Conclusion:

By highlighting potential disparities across groups, subsetting by a categorical variable enables a greater understanding of how that variable effects the link between predictors and the outcome.