Aditya Parey

ALY 6010 Summer 2024

Module practice 5

Statistical Outputs:

A screenshot of a computer screen

Description automatically generated

Correlation Matrix:

A computer code with numbers and symbols

Description automatically generated

A screenshot of a graph

Description automatically generated

Key Findings:

Height and Lung Capacity: Lung capacity and height have a significant positive connection (0.75), indicating that lung capacity tends to increase with height.   
  
Lung Capacity and Age: There is a positive substantial correlation (0.70) between lung capacity and age, lung capacity normally increases with age.   
  
Height and Age: Height increases with age, the correlation between age and height is moderate (0.55).

Why Is There a 5-Variable Limit?

Correlation charts are diagnostic tools, and in order to prevent complexity and overpopulation, they should only include five variables. A chart with too many variables may be more difficult to understand, less clear, and may contain conclusions that are not accurate because of the overemphasis on petty or unimportant correlations.

Regression Model:

A screenshot of a computer

Description automatically generated

Key findings:

Model Fit:

* Residual Standard Error: 1.092
* Multiple R-squared: 0.8321
* Adjusted R-squared: 0.8319
* F-statistic: 3583 on 1 and 723 DF
* p-value: < 2.2e-16

Model Significance: Low p-value (< 2.2e-16) indicate that the model is highly significant. This implies that lung capacity can be strongly predicted by height.

Interpretation of R-squared: R-squared value of 0.8321 suggests that height accounts for about 83.21% of the variation in lung capacity. This implies that these two variables have a very strong correlation.  
  
Interpretation of the Coefficient: A positive coefficient (0.337) for height shows that lung capacity rises with height. Therefore, it is statistically significant.

Difference between regression and correlation:

The linear link between two variables is measured by correlation, which also indicates its direction and strength. It implies neither causation nor the ability to anticipate one variable based on another.  
  
By quantifying the link between dependent and independent variables, regression goes beyond correlation. It tests theories on causal linkages and gives an equation to predict the dependent variable based on the independent variable or variables.