

Multi-Dataset SAS Data Analysis & Modeling

Questions:

- Q1. Use the SAS built-in dataset SASHELP.HEART to create vertical bar graphs for the average cholesterol level of male and female patients. Before creating the vertical bar graph, compute the average cholesterol levels for each sex using an appropriate SAS procedure. Customize the bar graph by assigning an appropriate title and using different bar colors than the default.
- Q2. Using appropriate SAS procedure, calculate the Body Mass Index of all the students using the formula $BMI = (\text{weight in pounds} / (\text{height in inches})^2) \times 703$ and dataset SAS.CLASS which is in SAS's internal library. Assume that the weight present is in pounds and height in inches.
- Q3. Use the SASHELP.BASEBALL dataset to examine the relationship between Number of At Bats (nAtBat) and Number of Hits (nHits). Use an appropriate SAS procedure to visualize and test this relationship. Justify your method and conduct a suitable hypothesis test. Interpret the result.
- Q4. Use BMIMEN dataset present in SAS's internal library (SASHELP.BMIMEN), to estimate a simple linear regression model with BMI as Y variable and AGE as x variable. Interpret the result.
- Q5. Estimate a multiple linear regression model using SASHELP.CARS dataset to figure out if there is any relationship between MSRP (Y variable) and each of the X variables - EngineSize, Horsepower, MPG_CITY, and, MPG_Highway. Is this a good model to explain variation in car prices? Why?
- Q6. For the model in Q5, to check if there is multicollinearity between MSRP and X variables - EngineSize, Horsepower, MPG_CITY, and MPG_Highway, by using appropriate SAS procedure. Address the multicollinearity issue in this model by using Backward Elimination method.