Portfolio Project Module 8

Cassidy Halpin

MIS446

June 9th 2024

R Script:

# Name: Cassidy Halpin

# Course: MIS446

# Module: 8

# Load and inspect the mtcars dataset

data(mtcars) # Load the dataset

head(mtcars) # Display the first few rows

# Create a subset with the variables of interest

mtcars\_subset <- mtcars[, c("qsec", "hp", "wt", "cyl", "carb", "disp")]

# Display the names of the variables in the subset

print(names(mtcars\_subset))

# Check the linear relationship using correlation

correlation\_matrix <- cor(mtcars\_subset)

print(correlation\_matrix)

# Visualize pairwise relationships using scatter plot matrix

pairs(mtcars\_subset, main = "Scatter Plot Matrix of Selected Variables")

# Build the multiple linear regression model

model <- lm(qsec ~ hp + wt + cyl + carb + disp, data = mtcars\_subset)

# Summarize the model to explore coefficients and statistics

summary(model)

# Model diagnostics

# Set up the layout for diagnostic plots

par(mfrow = c(2, 2))

# Plot diagnostic plots

plot(model)

Sys.time()

Sys.getenv("username")

# Create a subset with the variables of interest

mtcars\_subset <- mtcars[, c("qsec", "wt", "carb", "disp")]

# Build the second multiple linear regression model

# Using 'qsec' as the dependent variable and 'wt', 'carb', 'disp' as predictors

model\_refined <- lm(qsec ~ wt + carb + disp, data = mtcars\_subset)

# Summarize the refined model to explore coefficients and statistics

summary(model\_refined)

# Make predictions using the refined model

# New data for prediction

new\_data <- data.frame(

wt = c(2, 3, 4, 5),

carb = c(2, 4, 4, 6),

disp = c(100, 200, 300, 400)

)

# Predict the 'qsec' values for the new data

predicted\_qsec <- predict(model\_refined, newdata = new\_data)

# Combine the new data with the predicted 'qsec' values

predicted\_results <- cbind(new\_data, predicted\_qsec)

# Print the prediction results

print(predicted\_results)

# Plot the residuals to check the fit

par(mfrow = c(2, 2)) # Set up plotting area

plot(model\_refined) # Plot diagnostic plots for the refined model

Sys.time()

Sys.getenv("username")

Data Frame Creation and Correlation Matrix

A screenshot of a computer

Description automatically generated

Linear Regression Model 1 and Predictors

A screenshot of a computer

Description automatically generated

Linear Regression Model 2 and Predictors

A screenshot of a computer

Description automatically generated

Predicted Values

A screenshot of a computer

Description automatically generated

Explanation of Work:

First loaded the 'mtcars' dataset from R's built-in datasets. Then created a subset containing only 'qsec', 'wt', 'carb', and 'disp' variables. Built a linear model using 'qsec' as the dependent variable and 'wt', 'carb', 'disp' as predictors. Summarized the model to check the significance and coefficients of the predictors. Prepared new data for prediction with the specified 'wt', 'carb', and 'disp' values.Then used the 'predict' function to estimate 'qsec' for the new data.

The initial multiple linear regression model was built using qsec as the dependent variable and hp, wt, cyl, carb, and disp as predictors.

summary(initial\_model) provides the coefficients, standard errors, t-values, and p-values for each predictor, along with overall model statistics such as R-squared and F-statistic.

The refined model was built using qsec as the dependent variable and only wt, carb, and disp as predictors

Interpretation of Results:

Initial Model

**Coefficients**: These indicate the change in the dependent variable (qsec) for a one-unit change in the predictor, holding all other predictors constant.

**p-values**: These indicate each predictor is statistically significant. Typically, a p-value less than 0.05 is considered significant and the p-value is 1.449e-08

**Significant Predictors**: Based on the p-values, we identified the significant predictors (wt, carb, and disp) for building the refined model.

Refined model

**Coefficients**:

Intercept: The expected value of qsec when all predictors are zero = 16.89

wt: The effect of a one-unit increases in weight on the quarter-mile time (qsec). A positive coefficient means that as weight increases, the qsec value increases, indicating a slower acceleration.

carb: The effect of a one-unit increases in the number of carburettors on qsec. A positive coefficient suggests that more carburettors are associated with a slower quarter-mile time.

disp: The effect of a one-unit increases in displacement on qsec. A positive coefficient suggests that higher displacement is associated with a slower quarter-mile time.

**p-values**: the p-value is equal to 1.905e-09 making it statistically significant.

New Data Predictions

These predictions indicate the expected quarter-mile times for vehicles with the specified combinations of weight, number of carburettors, and displacement.

For instance, increasing weight generally results in an increase in qsec, indicating slower acceleration. Similarly, increasing the number of carburettors and displacement also results in higher qsec values, meaning slower acceleration.

The addition of creating new data predictions helped define the expected results of the regression models. The stepwise motion of moving from a simple model to more complex and then predicting new data made understanding the mechanisms of the statistics simpler.