Sultan Qaboos University
College of Science
Department of Statistics
STAT3336

The most notably causes of traffic accidents: A Shiny-Based Regression Analysis

Manar Alsubhi 137614

Objective:

This study aimed to analyze and model traffic accident data to identify patterns and factors associated with the severity of accidents—specifically injuries and deaths—based on gender, year, and the number of drivers, passengers, and pedestrians involved. By applying statistical methods using the car package and developing an interactive Shiny application, the study seeks to support data-driven decision-making in enhancing road safety and informing targeted prevention strategies.

Theory (Model and Assumptions):

➢ Linear Regression Model:

$$Y = \beta_0 + \beta_1 x_{gender} + \beta_2 x_{indicators} + \beta_3 x_{year} + \epsilon_{ij}$$

- Y : Response variable, the number of individuals involved in accidents (drivers, passengers, or pedestrians) with respect to gender and indicators and year
- Predictors: Year, Indicators(Injures, Death) , gender (male, female)
- β_0 : intercept of the linear model
- $\beta 1$: quantifies the average change in Y associated with a change in gender category.
- $\beta 2$: reflects the difference in average Y between injury and death cases.
- $\beta 3$: indicates how the number of people involved in accidents changes over the years.
- ϵij: the random error term, representing the variation in the response variable not explained by the model. $\epsilon ij \sim N(0,\sigma 2)$

➢ Assumptions:

1.Linearity: Relationship between predictors and response is linear.
2.Independence: Residuals are uncorrelated.
3.Homoscedasticity: Constant variance of residuals.
4.Normality: Residuals are normally distributed.

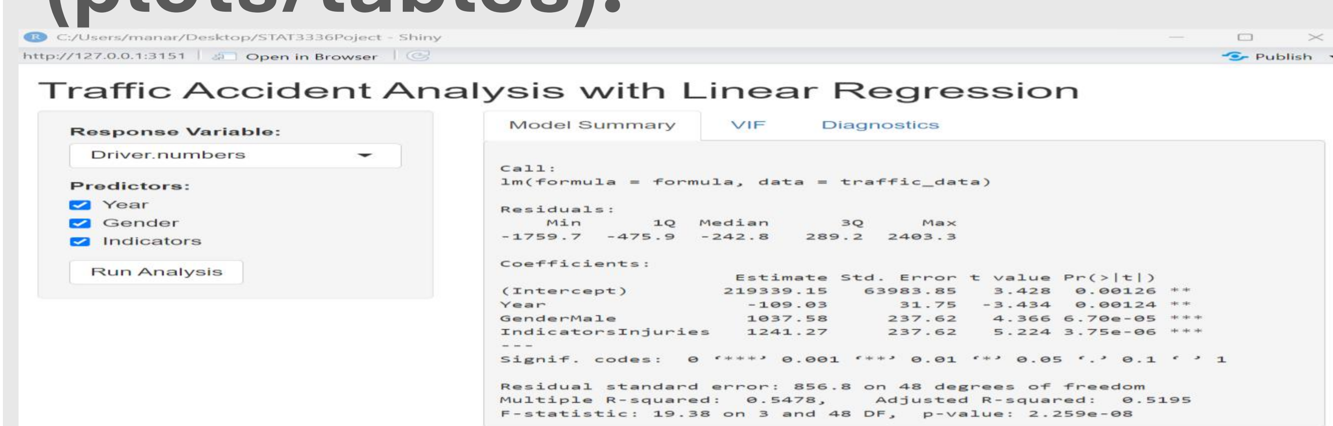
➢ Multicollinearity (VIF):

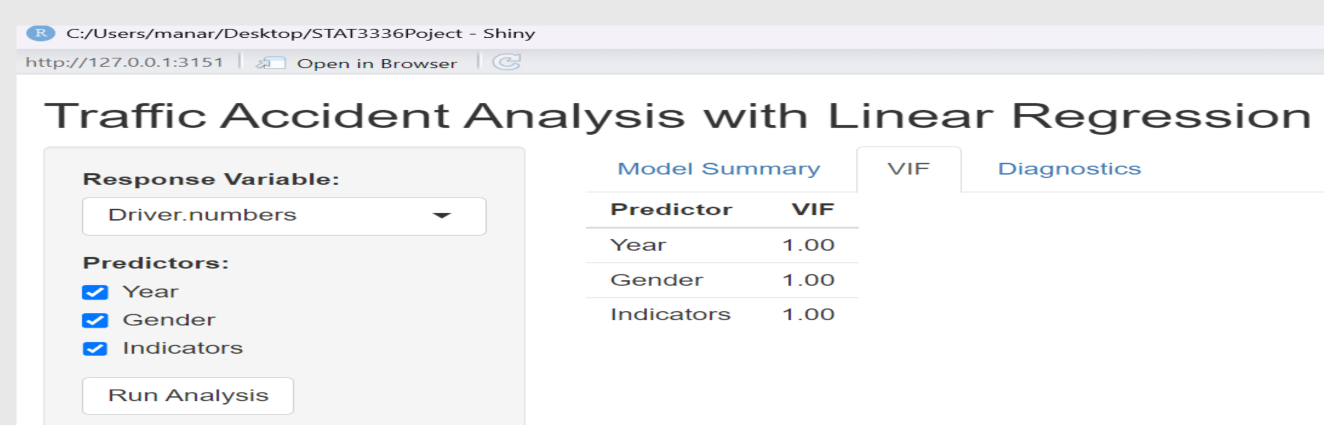
- Variance Inflation Factor (VIF): Measures how much the variance of a coefficient is inflated due to multicollinearity.
- Rule of Thumb: VIF > 5 indicates high multicollinearity.

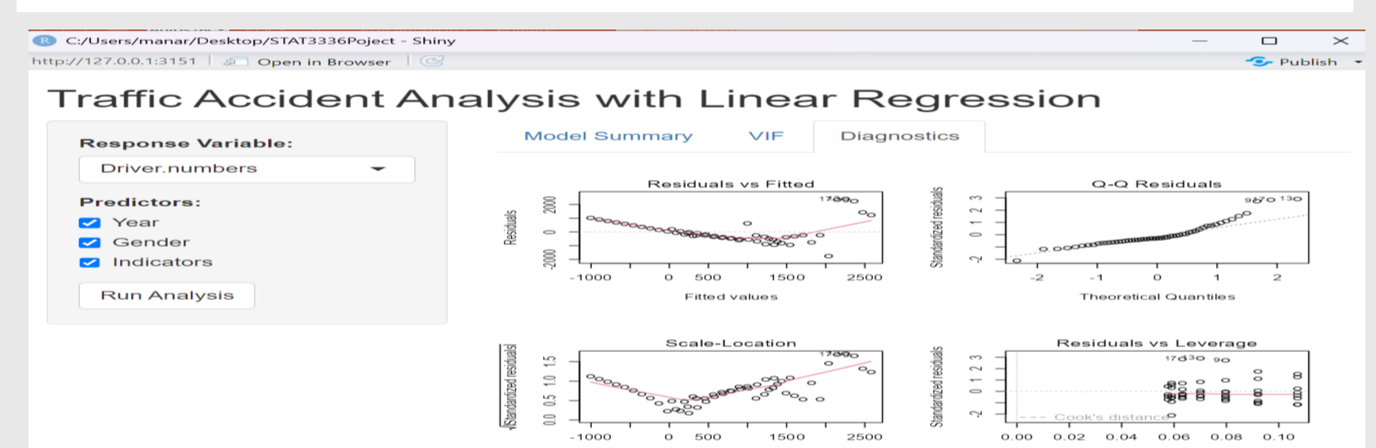
Algorithm:

1. **Import and Preprocess Data**
   - Load dataset by read it from an excel file
   - Store data set in the data frame
2. **Fit Model with Selected Predictors**
   - Use lm() for regression.
   - Allow dynamic predictor selection via Shiny inputs.
3. **Check Multicollinearity**
   - Compute VIF using car::vif().
4. **Generate Outputs**
   - Summary: Coefficients, R-squared, p-values.
   - Plots: Residuals vs Fitted, Q-Q plot.

Model Output & Diagnostics (plots/tables):







Interpretation and Key Findings:

- Significant yearly decrease of ≈109 drivers in accidents (p=0.00124) , and suggests improving safety measures over time
- Males involved in ≈1,038 more accidents than females (p<0.001) , Could reflect gender differences in driving exposure or risk-taking
- Injury incidents involve ≈1,241 more drivers than fatal crashes (p<0.001), so may indicate that more vehicles involved in non-fatal collisions, and underreporting of single-vehicle fatal accidents
- Residual patterns suggest ,potential omitted variables (e.g., traffic volume, weather) , and possible non-linear relationships , $R^2 = 0.55$ indicates 45% of variance remains unexplained

Shiny UI Screenshot & Code Highlights:

➢ **Code Highlights:**

1. Key UI Components

- **Dynamic Input Selection**:
  User can select one response variable (drivers, passengers, or pedestrians) , And choose the one or more predictor variable (gender, indicators , year)

```r
# Define UI
ui <- fluidPage(
  titlePanel("Traffic Accident Analysis with Linear Regression"),
  sidebarLayout(
    sidebarPanel(
      selectInput("response", "Response Variable:",
                  choices = c("Driver.numbers", "Passengers.numbers", "Pedestnion.numbers")),
      checkboxGroupInput("predictors", "Predictors:",
                  choices = c("Year", "Gender", "Indicators")),
      actionButton("run", "Run Analysis")
    ),
```

- **Tabbed Output Display**:
  To organize results into tabs for clarity

```r
    mainPanel(
      tabsetPanel(
        tabPanel("Model Summary", verbatimTextOutput("summary")),
        tabPanel("VIF", tableOutput("vif")),
        tabPanel("Diagnostics", plotOutput("diagnostics"))
      )
    )
  )
)
```

2. Server Logic

- **Reactive Model Fitting**: Dynamically constructs the regression formula based on user inputs.

```r
# Define Server
server <- function(input, output) {
  model <- eventReactive(input$run, {
    req(input$predictors)
    formula <- as.formula(paste(input$response, "~", paste(input$predictors, collapse = "+")))
    lm(formula, data = traffic_data)
  })
```

- **VIF Calculation**:
  Checks multicollinearity only if ≥2 predictors are selected.
  Uses car::vif() for multicollinearity checks.

```r
  output$summary <- renderPrint({
    summary(model())
  })

  output$vif <- renderTable({
    if (length(input$predictors) > 1) {
      vif_data <- car::vif(model())
      data.frame(Predictor = names(vif_data), VIF = vif_data)
    } else {
      data.frame(Note = "Select at least 2 predictors for VIF.")
    }
  })
```

➢ **User-Friendly UI**:
  •Dropdowns and checkboxes for intuitive variable selection.
  •Action button (actionButton) to control analysis execution.

➢ **Error Handling**:
  •req(input$predictors) ensures the model runs only when predictors are selected.
  •Conditional VIF output prevents errors with single predictors.

➢ **Diagnostic Integration**:
  •Uses car::vif() for multicollinearity checks.
  •Base R plot.lm() for automated model diagnostics.

- **Diagnostic Plots**:Generates residuals vs fitted, Q-Q, scale-location, and leverage plots.

```r
  output$diagnostics <- renderPlot({
    par(mfrow = c(2, 2))
    plot(model())
  })
}

# Run App
shinyApp(ui = ui, server = server)
```

➢ **Shiny UI Screenshot:**