MTH 441: Lab 9 - Time Delivery Data Analysis

Jiyanshu Dhaka 220481

2024-10-15

Introduction

This document presents the analysis of the Time Delivery Data, splitting the data into estimation and prediction datasets, proposing regression models, comparing coefficients, and calculating the PRESS statistics and R^2 values.

Load Necessary Libraries

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
library(broom)
library(readxl)  # Library to read Excel files
```

Load the Data

```
# Load the dataset (ensure the file path is correct)
data <- read_excel('TimeDeliveryData.xlsx')  # Use read_excel for Excel files

# Check the structure and column names of the data
str(data)          # Display structure of the dataset
```

```
## tibble [25 × 3] (S3: tbl_df/tbl/data.frame)
##  $ Y : num [1:25] 16.7 11.5 12 14.9 13.8 ...
##  $ X1: num [1:25] 7 3 3 4 6 7 2 7 30 5 ...
##  $ X2: num [1:25] 560 220 340 80 150 330 110 210 1460 605 ...
```

```
colnames(data)    # Display all column names
```

```
## [1] "Y"  "X1" "X2"
```

```
# Split the data into two equal parts for estimation and prediction
n <- nrow(data)
split_index <- floor(n / 2)  # Use floor to ensure index is an integer
estimation_data <- data[1:split_index, ]
prediction_data <- data[(split_index + 1):n, ]
```

Propose Two Regression Models

```
# Model 1
model1 <- lm(Y ~ X1 + X2, data = estimation_data)

# Model 2
model2 <- lm(Y ~ X1, data = estimation_data)  # Simpler model

# Summary of the models
summary(model1)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2, data = estimation_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7077 -1.1826  0.1314  1.3163  4.5857
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.741239   1.251702   1.391   0.1976
## X1          1.867842   0.241182   7.745 2.87e-05 ***
## X2          0.013521   0.004882   2.770   0.0218 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.843 on 9 degrees of freedom
## Multiple R-squared:  0.9826, Adjusted R-squared:  0.9788
## F-statistic: 254.7 on 2 and 9 DF,  p-value: 1.198e-08
```

```
summary(model2)
```

```
##
## Call:
## lm(formula = Y ~ X1, data = estimation_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0081 -2.0785 -0.3808  2.3554  6.8038
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3843     1.5881   1.501    0.164
## X1            2.4624     0.1419  17.347 8.59e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.671 on 10 degrees of freedom
## Multiple R-squared:  0.9678, Adjusted R-squared:  0.9646
## F-statistic: 300.9 on 1 and 10 DF,  p-value: 8.585e-09
```

Compare Regression Coefficients

```
# Coefficients for estimation data
coefficients_model1_estimation <- coef(model1)
coefficients_model2_estimation <- coef(model2)

# Fit the models on the prediction data
model1_pred <- lm(Y ~ X1 + X2, data = prediction_data)
model2_pred <- lm(Y ~ X1, data = prediction_data)

# Coefficients for prediction data
coefficients_model1_prediction <- coef(model1_pred)
coefficients_model2_prediction <- coef(model2_pred)

# Output coefficients
coefficients_model1_estimation
```

```
## (Intercept)          X1          X2
##  1.74123927  1.86784158  0.01352073
```

```
coefficients_model1_prediction
```

```
## (Intercept)          X1          X2
## 4.378376825 1.473818390 0.009963281
```

```
coefficients_model2_estimation
```

```
## (Intercept)          X1
##    2.384310    2.462383
```

```
coefficients_model2_prediction
```

```
## (Intercept)          X1
##    5.480103    1.794190
```

Compute PRESS Statistics and R^2

```
# Function to compute PRESS
compute_press <- function(model, data) {
  y <- data$Y  # Update to use the actual response variable name
  X <- model.matrix(model)

  press <- 0
  for (i in 1:nrow(data)) {
    y_pred <- predict(model, newdata = data[-i, ])
    press <- press + (y[i] - y_pred)^2
  }
  return(press)
}

# Compute PRESS for both models
press_model1 <- compute_press(model1, estimation_data)
press_model2 <- compute_press(model2, estimation_data)

# Calculate R^2 for both models
r2_model1 <- summary(model1)$r.squared
r2_model2 <- summary(model2)$r.squared

# Output results
press_model1
```

```
##         1         2         3         4         5         6         7         8
##  4200.158  6070.410  5653.747  6054.598  5087.769  4549.393  7050.305 36594.396
##         9        10        11
## 11611.302  8137.225  4582.970
```

```
press_model2
```

```
##         1         2         3         4         5         6         7         8
##  4357.507  6258.694  6223.225  5591.741  4699.021  4470.107  7090.452 34983.586
##         9        10        11
## 11596.465  8728.555  4787.000
```

```
r2_model1
```

```
## [1] 0.9826364
```

```
r2_model2
```

```
## [1] 0.9678375
```

Conclusion

In this analysis, we proposed two regression models, compared their coefficients between estimation and prediction datasets, and computed the PRESS statistics and R^2 values. The results indicate… (add your analysis here).