STP 530: Applied Regression Analysis

Name : **Sai Swaroop Reddy Vennapusa**

Homework **8**

Instructor : **Yi Zheng**

Due Date : 31st Oct 2023, 10:30AM

Question 8.19 a. Fit regression model (8.49) and state the estimated regression function.

Answer:

R Code:

```
# Set working directory to the location of the datasets
setwd("/home/swaroop/Downloads/Assignments/STP530/HW8")

# Read the data files
data1.20 <- read.table("CH01PR20.txt", header=FALSE) # Assuming there's a header
data8.15 <- read.table("CH08PR15.txt", header=FALSE) # Assuming there's a header

combined_data <- cbind(data1.20, data8.15)
colnames(combined_data) <- c("X_i1", "Y", "X_i2")


model <- lm(Y ~ X_i1 + X_i2 + X_i1:X_i2, data=combined_data)
summary(model)  # This will provide a summary of the regression including coefficients.
```

R Output:

```
> summary(model)  # This will provide a summary of the regression including coefficients.

Call:
lm(formula = Y ~ X_i1 + X_i2 + X_i1:X_i2, data = combined_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.88212 -0.43809  0.02041  0.33100  1.49878 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.062586   0.242216   0.258    0.797    
X_i1         0.066446   0.002752  24.149   <2e-16 ***
X_i2         0.402667   0.355971   1.131    0.265    
X_i1:X_i2   -0.006259   0.004101  -1.526    0.135    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5769 on 41 degrees of freedom
Multiple R-squared:  0.9599,	Adjusted R-squared:  0.957 
F-statistic: 327.3 on 3 and 41 DF,  p-value: < 2.2e-16
```

The regression model in general form can be represented as:

Y = beta_0 + beta_1 X_{i1} + beta_2 X_{i2} + beta_3 X_{i1}X_{i2} + epsilon

From the summary provided, the estimated regression coefficients ({beta}) are:

- {beta_0} (Intercept) = 0.062586
- {beta_1} for (X_{i1} = 0.066446
- {beta_2} for (X_{i2} = 0.402667
- {beta_3} for (X_{i1}X_{i2} = -0.006259

Therefore, the estimated regression function based on the provided dataset is:

{Y} = 0.062586 + 0.066446X_{i1} + 0.402667X_{i2} - 0.006259X_{i1}X_{i2}

This equation represents the relationship between the number of copies made (`Y`), the size of the firm (`X_i1`), and whether or not the firm is a stock company (`X_i2`), taking into account the interaction between `X_i1` and `X_i2`.

b. Test whether the interaction term can be dropped from the model; control the (alpha) risk at .10. State the alternatives, decision rule, and conclusion. What is the P-value of the test? lf the interaction term cannot be dropped from the model. describe the nature of the interaction effect.

Answer:

Hypothesis testing of the interaction term

1. Assumptions:
   - Residuals are independent and identically distributed (i.i.d.)
   - Residuals follow a normal distribution with mean 0 and variance sigma^2.

2. Hypotheses:
   H_0 : beta_{interaction} = 0 (The interaction term is not significant.)
   H_1 : beta_{interaction} != 0 (The interaction term is significant.)

3. Test-statistic:
Given the regression output, we already have the t-value for the interaction term:
t_{obs} = {beta_{interaction} - 0}/{s{beta_{interaction}}}
Where:
   - beta_{interaction} is the estimated coefficient for the interaction term from the regression output.
   - s{beta_{interaction}} is the standard error of the estimated coefficient for the interaction term from the regression output.

Using the summary output:
t_{obs} = {-0.006259 - 0}/{0.004101} = -1.526

4. P-value:
From the regression output summary, the p-value for the interaction term is already provided as 0.135.

5. Conclusion:
Given the significance level (alpha) of 0.10, and since the p-value (0.135) > alpha, we fail to reject the null hypothesis H_0. There is not enough evidence to suggest that the interaction term is significant at the 0.10 level of significance. This suggests that the interaction term can potentially be dropped from the model.

Nature of the Interaction Effect:
Though statistically insignificant at the 0.10 level, the interaction's negative coefficient (-0.006259) suggests that as X_{i1} increases, the relationship between X_{i2} and Y becomes less positive or more negative (and vice versa). This indicates a diminishing positive effect (or increasing negative effect) of one predictor as the other predictor increases. In practical terms, the impact of one variable on the dependent variable Y depends on the level of the other variable.