

STP 530: Applied Regression Analysis
Name : Sai Swaroop Reddy Vennapusa
Homework 6
Instructor : Yi Zheng
Due Date : 3rd Oct 2023, 10:30AM

Question 7.5. Refer to Patient satisfaction Problem 6.15.

(a) Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_2 ; with X_1 given X_2 ; and with X_3 , given X_2 and X_1 .

Answer:

R code:

```
# Read the dataset from the given path
data <- read.table("/home/swaroop/Downloads/Assignments/STP530/HW5/CH06PR15.txt", quote="\"", comment.char="", header=FALSE)

# Rename the columns for ease of reference
colnames(data) <- c("Y", "X1", "X2", "X3")

# Full Model with all predictors
model_full <- lm(Y ~ X1 + X2 + X3, data = data)

# Model with only X2
model_X2 <- lm(Y ~ X2, data = data)

# Model with X1 given X2
model_X1_given_X2 <- lm(Y ~ X2 + X1, data = data)

# Model with X3 given X1 and X2
model_X3_given_X1_X2 <- lm(Y ~ X2 + X1 + X3, data = data)

# Compute SSR for each model
SSR_X2 = sum((fitted(model_X2))^2)
SSR_X1_given_X2 = sum((fitted(model_X1_given_X2))^2) - SSR_X2
SSR_X3_given_X1_X2 = sum((fitted(model_X3_given_X1_X2))^2) - sum((fitted(model_X1_given_X2))^2)
SSR_full = sum((fitted(model_full))^2)

# Compute SSE for the full model
SSE_full = sum(residuals(model_full)^2)

# Total sum of squares
SSTO = sum((data$Y - mean(data$Y))^2)

# Compute df for each model
df_X2 = 1
df_X1_given_X2 = 1
df_X3_given_X1_X2 = 1
df_full = 3
df_error = nrow(data) - 4

# Display ANOVA table
anova_table <- data.frame(
  Source_of_Variation = c("Regression", "X2", "X1|X2", "X3|X1,X2", "Error", "Total"),
  SS = c(SSR_full, SSR_X2, SSR_X1_given_X2, SSR_X3_given_X1_X2, SSE_full, SSTO),
  df = c(df_full, df_X2, df_X1_given_X2, df_X3_given_X1_X2, df_error, nrow(data) - 1),
  MS = c(SSR_full/df_full, SSR_X2/df_X2, SSR_X1_given_X2/df_X1_given_X2, SSR_X3_given_X1_X2/df_X3_given_X1_X2,
  SSE_full/df_error, SSTO/(nrow(data) - 1))
)

print(anova_table)
```

R Output:

```
> print(anova_table)
  Source_of_Variation      SS df      MS
1      Regression 183473.1593   3 61157.7198
2              X2 179212.9557   1 179212.9557
3      X1|X2      3896.0441   1   3896.0441
4      X3|X1,X2    364.1595   1    364.1595
5          Error  4248.8407  42   101.1629
6          Total 13369.3043  45   297.0957
```

(b) Test whether X3 can be dropped from the regression model given that X1 and X2 are retained. Use the F* test statistic and level of significance .025. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Answer:

Step 1: Assumptions

The errors, epsilon , are assumed to be independent and identically distributed (i.i.d. following a normal distribution with mean 0 and constant variance σ^2 .

Step 2: Hypotheses

- Null Hypothesis H_0 : $\beta_3 = 0$ (i.e., X_3 does not significantly contribute to the model)
- Alternative Hypothesis H_1 : $\beta_3 \neq 0$

Step 3: Test-statistic

The test statistic F(obs) is computed as:

$$F(\text{obs}) = (\text{SSR}(X_3|X_1, X_2)/1) / (\text{SSE}(X_1, X_2, X_3)/(n-4))$$

Given from the ANOVA table:

$$\text{SSR}(X_3|X_1, X_2) = 364.1595$$

$$\text{SSE}(X_1, X_2, X_3) = 4248.8407$$

$$n = 45$$

Plugging in the values:

$$F(\text{obs}) = \{364.1595\} / \{101.1629\} = 3.60$$

Step 4: P-value

R code:

```
anova_results <- anova(model_X1_given_X2, model_X3_given_X1_X2)
print(anova_results)

p_value <- 1 - pf(3.60, 1, 41)
p_value
|
```

R Output:

```
> print(anova_results)
Analysis of Variance Table

Model 1: Y ~ X2 + X1
Model 2: Y ~ X2 + X1 + X3
    Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      43 4613.0
2      42 4248.8  1    364.16 3.5997 0.06468 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> p_value <- 1 - pf(3.60, 1, 41)
> p_value
[1] 0.06483751
```

Step 5: Conclusion

The P-value of 0.06483751 is greater than the pre-determined significance level of 0.025.

Thus, we **fail to reject H0**.

There isn't enough evidence to suggest that X3 contributes significantly to the regression model when X1 and X2 are already included.

Question 7.6: Refer to Patient satisfaction Problem 6.15. Test whether both X2 and X3 can be dropped from the regression model given that X1 is retained. Use alpha .025. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Answer:

Step 1: Assumptions

epsilon is independently and identically distributed following a normal distribution with mean 0 and variance σ^2 .

Step 2: Hypotheses

Full Model (m.F): $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

Reduced Model (m.R): $Y = \beta_0 + \beta_1 X_1 + \epsilon$

Null hypothesis, H0: $\beta_2 = \beta_3 = 0$ (Both X2 and X3 do not contribute significantly to the model when X1 is included.)

Alternative hypothesis, H1: At least one of β_2 or β_3 is not equal to 0.

Step 3: Test-statistic

$F_{\{obs\}} = \{\text{reduction in residual sum of squares due to predictors in the full model but not in the restricted model}\} / \{\text{mean square error of the full model}\}$

Step 4: P-value:

R code:

```
# Question 7.6
m.F <- model_X3_given_X1_X2

# Reduced model only with X1
m.R <- lm(Y ~ X1, data = data)

# Perform the F-test comparing the two models
test <- anova(m.R, m.F)
p_value <- test$`Pr(>F)`[2]
p_value
|
rsq::rsq.partial(objF = m.F, objR = m.R)
```

R Output:

```
> p_value
[1] 0.02216118
> rsq::rsq.partial(objF = m.F, objR = m.R)
$adjustment
[1] FALSE

$variables.full
[1] "X2" "X1" "X3"

$variables.reduced
[1] "X1"

$partial.rsq
[1] 0.1658989
```

The obtained P-value is 0.02216118.

Step 5: Conclusion

Given that our P-value = 0.02216118 is less than the predetermined significance level $\alpha = 0.025$, we reject the null hypothesis H_0 .

In context, this means that we have sufficient evidence to conclude that at least one of the predictors X_2 or X_3 contributes significantly to the model when X_1 is already included.

Question 7.9. Refer to Patient satisfaction Problem 6.15. Test whether, $\beta_1 = -1.0$ and, $\beta_2 = 0$; use $\alpha = .025$. State the alternatives, full and reduced models, decision rule, and conclusion.

Answer:

Step 1: State the Hypotheses

Null hypothesis, H_0 :

$$\beta_1 = -1.0$$

$$\beta_2 = 0$$

The coefficients for X_1 and X_2 in the model are -1.0 and 0, respectively.

Alternative hypothesis, H_1 :

At least one of the coefficients β_1 or β_2 is not equal to its hypothesized value.

Step 2: Specify the Full and Reduced Models

Full Model (m.F):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Reduced Model (m.R):

First adjust the response variable:

$$Y.R = Y - (-1.0 \text{ times } X_1)$$

The model becomes:

$$Y.R = \beta_0 + \beta_3 X_3 + \epsilon$$

Step 3: Compute the Test Statistic

R Code:

```
# Question 7.9
# Adjusting for the specified hypotheses
# The full model
data.full <- data
data.full$Y <- data.full$Y
m.F <- lm(Y ~ X1 + X2 + X3, data = data.full)

# The reduced model
data.reduced <- data
data.reduced$Y <- with(data, Y + X1) # Adjusting the response for the beta1 = -1.0 hypothesis.
m.R <- lm(Y ~ X3, data = data.reduced) # No need to include X2 as beta2 = 0

# The F-test comparing the two models
test_result <- anova(m.R, m.F)
print(test_result)

# Effect size: partial R-squared
partial_r2 <- rsq::rsq.partial(objF = m.F, objR = m.R)
print(partial_r2)
```

R Output:

```
> print(test_result)
Analysis of Variance Table

Model 1: Y ~ X3
Model 2: Y ~ X1 + X2 + X3
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     44 4427.7
2     42 4248.8  2    178.81 0.8838 0.4208
> # Effect size: partial R-squared
> partial_r2 <- rsq::rsq.partial(objF = m.F, objR = m.R)
> print(partial_r2)
$adjustment
[1] FALSE

$variables.full
[1] "X1" "X2" "X3"

$variables.reduced
[1] "X3"

$partial.rsq
[1] 0.5647479
```

Step 4 : P-Value

From the ANOVA table:

F-statistic (F) = 0.8838

P-value (Pr(>F)) = 0.4208

Step 5: Conclusion

Given that the p-value is 0.4208, which is greater than the significance level of 0.025, we fail to reject the null hypothesis. This means that there's insufficient evidence to suggest that the coefficients β_1 or β_2 are different from their hypothesized values of -1.0 and 0, respectively.