MSMS 206 : Practical 01

Ananda Biswas

March 7, 2025

Question

Fit a multiple linear regression model for the following data-set and obtain the following results.

(a) estimate of the regression coefficients and σ,

(b) confidence interval of the regression coefficients,

(c) coefficient of determination,

(d) adjusted coefficient of determination.

y	x_1	x_2
16.68	7	560
11.5	3	220
12.03	3	340
14.88	4	80
13.75	6	150
18.11	7	330
8	2	110
17.83	7	210
79.2	30	1460
21	10	215
13.5	4	255
19.75	6	462
24	9	448
29	10	776
15.35	6	220
19	7	132
9.5	3	36
35.1	17	770
17.9	10	140
52.32	26	810
18.75	9	450
19.83	8	635
10.75	4	150
21.5	5	605
40.33	16	688

We fit a multiple linear regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.

```r
raw_data <- read.csv('https://raw.githubusercontent.com/sakunisgithub/data_sets/refs/heads/master/msc_semester_2/tripti_madam_practical_01_data.csv')
```

```r
dim(raw_data)

## [1] 25  3
```

```r
names(raw_data)

## [1] "Y"   "X_1" "X_2"
```

```r
model1 <- lm(Y ~ X_1 + X_2, data = raw_data)
model_summary <- summary(model1)
```

```r
model_summary

##
## Call:
## lm(formula = Y ~ X_1 + X_2, data = raw_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.776 -0.659  0.164  1.173  7.387
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.326534   1.096063   2.123 0.045279 *
## X_1         1.614726   0.170574   9.466 3.24e-09 ***
## X_2         0.014414   0.003615   3.987 0.000623 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.254 on 22 degrees of freedom
## Multiple R-squared:  0.9597,Adjusted R-squared:  0.956
## F-statistic: 261.9 on 2 and 22 DF,  p-value: 4.561e-16
```

✎ Estimates of regression coefficients are

```r
estimates <- model_summary$coefficients[, 'Estimate']
estimates

## (Intercept)         X_1         X_2
##  2.32653410  1.61472628  0.01441393
```

✎ Estimate of error standard deviation σ is

```
sigma_hat <- model_summary$sigma
sigma_hat

## [1] 3.254133
```

Now we shall calculate the confidence intervals of the regression coefficients and the intercept. $100(1-\alpha)\%$ confidence interval for $\beta_j \ \forall j = 0(1)2$ is given by

$$\left(\hat{\beta}_j - t_{\frac{\alpha}{2},n-p} \cdot \mathrm{se}(\hat{\beta}_j), \hat{\beta}_j + t_{\frac{\alpha}{2},n-p} \cdot \mathrm{se}(\hat{\beta}_j)\right).$$

where n is total number of observations and p is total number of parameters in the model.

```
std_errors <- model_summary$coefficients[, 'Std. Error']

t_tabulated <- qt(0.025, 22, lower.tail = FALSE)

CI_lower <- estimates - t_tabulated * std_errors
CI_upper <- estimates + t_tabulated * std_errors
```

✎ 95% confidence interval for β_0 is

```
CI_lower[1]; CI_upper[1]

## (Intercept)
##   0.0534379
## (Intercept)
##     4.59963
```

✎ 95% confidence interval for β_1 is

```
CI_lower[2]; CI_upper[2]

##      X_1
## 1.260978
##      X_1
## 1.968475
```

✎ 95% confidence interval for β_2 is

```
CI_lower[3]; CI_upper[3]

##         X_2
## 0.006916106
##        X_2
## 0.02191175
```

Now we shall calculate the confidence interval of the error variance σ^2. $100(1-\alpha)\%$ confidence interval for σ^2 is given by

$$\left(\frac{\text{RSS}}{\chi^2_{\frac{\alpha}{2},n-p}}, \frac{\text{RSS}}{\chi^2_{1-\frac{\alpha}{2},n-p}}\right).$$

```r
RSS <- sum(model_summary$residuals^2) # Residual Sum of Squares
```

✎ 95% confidence interval for σ^2 is

```r
RSS / qchisq(0.025, 22, lower.tail = FALSE) # lower bound

## [1] 6.333929

RSS / qchisq(1-0.025, 22, lower.tail = FALSE) # upper bound

## [1] 21.21286
```

✎ 95% confidence interval for σ is

```r
sqrt(RSS / qchisq(0.025, 22, lower.tail = FALSE)) # lower bound

## [1] 2.51673

sqrt(RSS / qchisq(1-0.025, 22, lower.tail = FALSE)) # upper bound

## [1] 4.605742
```

✎ R^2 for the model is

```r
model_summary$r.squared

## [1] 0.9596944
```

✎ Adjusted R^2 for the model is

```r
model_summary$adj.r.squared

## [1] 0.9560302
```