Multiple Linear Regression (MLR) Analysis

Eduardo Martinez

9/12/2021

Contents

File Background Information:

- Type: My Solutions to a Homework Problem
- Course: Applied Statistics / Regression (MATH-564)
- Level: Graduate Course
- Institution: Illinois Institute of Technology (IIT)
- Semester: Fall, 2021
- Course Textbook: Chatterjee, S. & Hadi, A. S. (2012). *Regression Analysis by Example.* (Fifth Edition).

R Packages Used:

```
library(knitr, quietly = T)
library(kableExtra, quietly = T, warn.conflicts = F)
library(tidyverse, quietly = T, warn.conflicts = F)
```

Problem Instructions:

Consider the *Supervisor Performance Data* in Table 3.3 on page 60 of the TEXT.

(1) Estimate the regression coefficients vector $\hat{\beta}$.
(2) Verify that $\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$.
(3) Now consider $p = 2$ with X_3 and X_4 being the only two predictors used. The model becomes

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + \epsilon.$$

Use the 3-step method described on page 63 to obtain the coefficient for X_3, and compare it with the coefficient of X_3 by regressing Y on X_3 and X_4 using the 2-predictor model above. Are they the same? Explain why or why not.

The Data:

```
table3.3 <- read_tsv("Table3.3.txt", show_col_types = F)
LaTeX_tab <- cbind(Row = 1:30, table3.3)
colnames(LaTeX_tab) <- c("Row", "$Y$", "$X_1$", "$X_2$", "$X_3$", "$X_4$", "$X_5$", "$X_6$")
kbl(LaTeX_tab, booktabs = T, escape = F, align = "c") %>%
  kable_classic() %>%
  kable_styling(latex_options = c("condensed", "striped", "HOLD_position"), font_size = 11) %>%
  column_spec(1:8, width = "15mm") %>%
  add_header_above(header = c(" "=1, "\\\\textbf{Table 3.3} $~$ \\\\text{Supervisor Performance Dat
                   font_size = 12, escape = F)
```

	Table 3.3 Supervisor Performance Data						
Row	Y	X_1	X_2	X_3	X_4	X_5	X_6
1	43	51	30	39	61	92	45
2	63	64	51	54	63	73	47
3	71	70	68	69	76	86	48
4	61	63	45	47	54	84	35
5	81	78	56	66	71	83	47
6	43	55	49	44	54	49	34
7	58	67	42	56	66	68	35
8	71	75	50	55	70	66	41
9	72	82	72	67	71	83	31
10	67	61	45	47	62	80	41
11	64	53	53	58	58	67	34
12	67	60	47	39	59	74	41
13	69	62	57	42	55	63	25
14	68	83	83	45	59	77	35
15	77	77	54	72	79	77	46
16	81	90	50	72	60	54	36
17	74	85	64	69	79	79	63
18	65	60	65	75	55	80	60
19	65	70	46	57	75	85	46
20	50	58	68	54	64	78	52
21	50	40	33	34	43	64	33
22	64	61	52	62	66	80	41
23	53	66	52	50	63	80	37
24	40	37	42	58	50	57	49
25	63	54	42	48	66	75	33
26	66	77	66	63	88	76	72
27	78	75	58	74	80	78	49
28	48	57	44	45	51	83	38
29	85	85	71	71	77	74	55
30	82	82	39	59	64	78	39

My Solutions:

(1) Estimate the regression coefficients vector $\hat{\beta}$.

```
fit3.3 <- lm(Y ~., data = table3.3)

LaTeX_vars <- c("(Intercept)", paste("$X_", 1:6, "$", sep = ""))
Beta_labs <- paste("$\\beta_", 0:6, "$", sep = "")
Coef_LaTeX <- tibble(Variable = LaTeX_vars, Coefficeint = Beta_labs, Estimate = round(coef(fit3.3),

kbl(Coef_LaTeX, booktabs = T, escape = F, align = "c", linesep = "") %>%
  kable_classic() %>%
  kable_styling(latex_options = c("striped", "HOLD_position"), font_size = 11) %>%
  column_spec(1:3, width = "1.35in") %>%
  add_header_above(header = c("Estimated Regression Coefficeints"=3), font_size = 12, bold = T)
```

Estimated Regression Coefficeints		
Variable	Coefficeint	Estimate
(Intercept)	β_0	10.787
X_1	β_1	0.613
X_2	β_2	-0.073
X_3	β_3	0.320
X_4	β_4	0.082
X_5	β_5	0.038
X_6	β_6	-0.217

---

(2) Verify that $\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$.

```
sum1 <- sum(fit3.3$fitted.values)
sum2 <- sum(table3.3$Y)
```

$$\sum_{i=1}^{n} \hat{y}_i = 1939 \quad \text{and} \quad \sum_{i=1}^{n} y_i = 1939 \quad \checkmark$$

---

(3) Now consider $p = 2$ with X_3 and X_4 being the only two predictors used. The model becomes

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + \epsilon.$$

Use the 3-step method described on page 63 to obtain the coefficient for X_3, and compare it with the coefficient of X_3 by regressing Y on X_3 and X_4 using the 2-predictor model above. Are they the same? Explain why or why not.

Step 1:

```
fit_X4 <- lm(Y ~ X4, data = table3.3)
coefX4 <- round(coefficients(fit_X4), 4)
```

$$\hat{Y} = 19.9778 + 0.6909\ X_4$$

Step 2:

```
fit_X3 <- lm(X3 ~ X4, data = table3.3)
coefX3 <- round(coefficients(fit_X3), 4)
```

$$\hat{X}_3 = 9.6481 + 0.7228\ X_4$$

Step 3:

```
fit_eYX4 <- lm(fit_X4$residuals ~ fit_X3$residuals)
coef_eYX4 <- round(coefficients(fit_eYX4), 4)
```

$$\hat{e}_{Y \cdot X_4} = 0 + 0.4321\ e_{X_3 \cdot X_4}$$

MLR (Two-Predictor) Model:

```
fit_X3X4 <- lm(Y~ X3 + X4, data = table3.3)
coefX3X4 <- round(coefficients(fit_X3X4), 4)
```

$$\hat{Y} = 15.8091 + 0.4321\ X_3 + 0.3786\ X_4$$

Conclusion

The estimated coefficient, $\hat{\beta}_3$, for X_3 was equal to 0.4321 when applying the 3-step procedure of SLR models and when using a MLR model.

As a result, $\hat{\beta}_3$ can be found using either a series of SLR models or the associated MLR model.