Regression and Lab HW 5

Na SeungChan

2023-06-06

Q1

Importing Data

```
b1 <- c(2.60, 31.0, 21)
b2 <- c(2.40, 31.0, 21)
b3 <- c(17.32, 31.5, 24)
b4 <- c(15.60, 31.5, 24)
b5 <- c(16.12, 31.5, 24)
b6 <- c(5.36, 30.5, 22)
b7 <- c(6.19, 31.5, 22)
b8 <- c(10.17, 30.5, 23)
b9 <- c(2.62, 31, 21.5)
b10 <- c(2.98, 30.5, 21.5)
b11 <- c(6.92, 31, 22.5)
b12 <- c(7.06, 30.5, 22.6)
Beverages <- as_tibble(rbind(b1, b2, b3, b4, b5, b6, b7, b8, b9, b10, b11, b12)) %>%
  rename(Carbonation = V1, Temperature = V2, Pressure = V3)
```

```
## Warning: The `x` argument of `as_tibble.matrix()` must have unique column names if
## `.name_repair` is omitted as of tibble 2.0.0.
## i Using compatibility `.name_repair`.
```

(a)

```
fit11 <- lm(Carbonation ~ poly(Temperature, Pressure, degree = 2, raw = TRUE), data = Beverages)
summary(fit11)
```

```
##
## Call:
## lm(formula = Carbonation ~ poly(Temperature, Pressure, degree = 2,
##     raw = TRUE), data = Beverages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76031 -0.32595  0.04094  0.25689  0.95969
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                     2968.7591  2230.2245
## poly(Temperature, Pressure, degree = 2, raw = TRUE)1.0 -187.5829   143.9432
## poly(Temperature, Pressure, degree = 2, raw = TRUE)2.0    3.4640     2.4061
```

```
## poly(Temperature, Pressure, degree = 2, raw = TRUE)0.1  -10.4076    22.1038
## poly(Temperature, Pressure, degree = 2, raw = TRUE)1.1   -1.1758     0.9638
## poly(Temperature, Pressure, degree = 2, raw = TRUE)0.2    1.1424     0.3528
##                                                      t value Pr(>|t|)
## (Intercept)                                            1.331   0.2315
## poly(Temperature, Pressure, degree = 2, raw = TRUE)1.0  -1.303   0.2403
## poly(Temperature, Pressure, degree = 2, raw = TRUE)2.0   1.440   0.2000
## poly(Temperature, Pressure, degree = 2, raw = TRUE)0.1  -0.471   0.6544
## poly(Temperature, Pressure, degree = 2, raw = TRUE)1.1  -1.220   0.2683
## poly(Temperature, Pressure, degree = 2, raw = TRUE)0.2   3.238   0.0177 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6746 on 6 degrees of freedom
## Multiple R-squared:  0.992,  Adjusted R-squared:  0.9854
## F-statistic: 149.2 on 5 and 6 DF,  p-value: 3.305e-06
```

(b)

From summary of fit1, we can get F-statistic 149.2 on (5,6) DF. p-value is 3.305e-06, and regression is significant.

(c)

lack_of_fit test

```
fit13 <- lm(Carbonation ~ Temperature + Pressure, data = Beverages)

anova(fit11, fit13)
```

```
## Analysis of Variance Table
##
## Model 1: Carbonation ~ poly(Temperature, Pressure, degree = 2, raw = TRUE)
## Model 2: Carbonation ~ Temperature + Pressure
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1      6 2.7308
## 2      9 8.6434 -3   -5.9126 4.3304 0.06021 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

no lack of fit.

(d)

t-value for interaction term is -1.220. Interaction term does not significantly contribute to the model.

```
fit12 <- lm(Carbonation ~ poly(Temperature, degree = 2, raw = TRUE) + poly(Pressure, degree = 2, raw = TRUE), data
summary(fit12)
```

```
##
## Call:
## lm(formula = Carbonation ~ poly(Temperature, degree = 2, raw = TRUE) +
##     poly(Pressure, degree = 2, raw = TRUE), data = Beverages)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -0.8209 -0.4707  0.1394  0.3155  0.8991
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                          2781.5067  2301.2131   1.209
## poly(Temperature, degree = 2, raw = TRUE)1 -158.7518   146.8575  -1.081
## poly(Temperature, degree = 2, raw = TRUE)2    2.5748     2.3717   1.086
## poly(Pressure, degree = 2, raw = TRUE)1     -33.5184    11.7782  -2.846
## poly(Pressure, degree = 2, raw = TRUE)2       0.8423     0.2616   3.220
##                                       Pr(>|t|)
## (Intercept)                             0.2660
## poly(Temperature, degree = 2, raw = TRUE)1   0.3155
## poly(Temperature, degree = 2, raw = TRUE)2   0.3136
## poly(Pressure, degree = 2, raw = TRUE)1      0.0248 *
## poly(Pressure, degree = 2, raw = TRUE)2      0.0147 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6978 on 7 degrees of freedom
## Multiple R-squared:   0.99,  Adjusted R-squared:  0.9843
## F-statistic:   174 on 4 and 7 DF,  p-value: 4.402e-07
```

Adjusted R-Squared also does not changed significantly.

(e)

Fit the multiple linear regression model.

```
summary(fit13)
```

```
##
## Call:
## lm(formula = Carbonation ~ Temperature + Pressure, data = Beverages)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3640 -0.7462  0.1358  0.6475  1.3165
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -151.6265    21.9720  -6.901 7.06e-05 ***
## Temperature    1.8774     0.7822   2.400   0.0399 *
## Pressure       4.5205     0.2963  15.258 9.73e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.98 on 9 degrees of freedom
## Multiple R-squared:  0.9747, Adjusted R-squared:  0.9691
## F-statistic: 173.7 on 2 and 9 DF,  p-value: 6.47e-08
```

quadratic term for Pressure is significant, but for Temperature intercept isn’t significant.

Q2

Answer is on the 1st page of pdf file.

Q3

Data importing

```
dfq3 <- na.omit(MPV::table.b3)
```

Trans AM, Astre has NA in x3. I just dropped all the data in Trnas AM and Astre, but additional analysis is needed for justification of droping those two data.

(a)

x6, x7, x11 are categorical variable, so those variables are not included in correlation matrix.

```
dfq3 %>%
  select(x1, x2, x3, x4, x5, x8, x9, x10) %>%
  cor()
```

```
##            x1         x2         x3         x4         x5         x8
## x1   1.0000000  0.9408473  0.9891628 -0.3469725 -0.6720903  0.8623681
## x2   0.9408473  1.0000000  0.9643592 -0.2898995 -0.5509642  0.8027387
## x3   0.9891628  0.9643592  1.0000000 -0.3259992 -0.6728661  0.8641224
## x4  -0.3469725 -0.2898995 -0.3259992  1.0000000  0.4137808 -0.3041503
## x5  -0.6720903 -0.5509642 -0.6728661  0.4137808  1.0000000 -0.5613315
## x8   0.8623681  0.8027387  0.8641224 -0.3041503 -0.5613315  1.0000000
## x9   0.7974811  0.7105117  0.7881284 -0.3781736 -0.4534470  0.8831512
## x10  0.9515520  0.8878810  0.9434871 -0.3584588 -0.5798617  0.9554541
##            x9        x10
## x1   0.7974811  0.9515520
## x2   0.7105117  0.8878810
## x3   0.7881284  0.9434871
## x4  -0.3781736 -0.3584588
## x5  -0.4534470 -0.5798617
## x8   0.8831512  0.9554541
## x9   1.0000000  0.8994711
## x10  0.8994711  1.0000000
```

some data pairs like (x1, x2), (x1, x3), (x3, x10) show large correlation. There could be some multicolinearity and large VIFs.

(b)

```
lmq3 <- lm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11, data = dfq3)
vif(lmq3)
```

```
##         x1         x2         x3         x4         x5         x6         x7
## 119.487804  42.800811 149.234409   2.060036   7.729187   5.324730  11.761341
##         x8         x9        x10        x11
##  20.917632   9.397108  85.744344   5.145052
```

variables with VIF > 10 : x1, x2, x3, x7, x8, x10, x11.

This model shows serious multicolinearity problems.