Hw06ST430Yu

Haozhe (Jerry) Yu

2023-11-10

Question 1

A researcher studied the effects of the charge rate and temperature on the life of a new type of power cell in a preliminary small-scale experiment. The charge rate (Xl) was controlled at three levels (0.6, 1.0, and 1.4 amperes) and the ambient temperature (X2) was controlled at three levels (l0, 20, 30°C). Factors pertaining to the discharge of the power cell were held at fixed levels. The life of the power cell (Y) was measured in terms of the number of discharge-charge cycles that a power cell underwent before it failed.

The researcher was not sure about the nature of the response function in the range of the factors studied. Hence, the researcher decided to fit the second-order polynomial regression model

```
data <- read.table("Datasets/battery.txt", header=FALSE)
names(data) <- c("cycles","rate","temp")
attach(data)
```

##a. Find the correlation matrix and report any high correlation between predictor variables.

```
cor(data)
```

```
##             cycles       rate      temp
## cycles  1.0000000 -0.5555349 0.7512159
## rate   -0.5555349  1.0000000 0.0000000
## temp    0.7512159  0.0000000 1.0000000
```

The correlation between cycles and temp is 0.7512159. This is high and could be a sign of multicollinearity.

##b. Fit a full model (Shown above) and report the overall F value and individual t-values. Do you suspect any multicollinearity problem?

```
mod1<-lm(cycles~rate+temp+I(rate^2)+I(temp^2)+ I(rate*temp))
summary(mod1)
```

```
##
## Call:
## lm(formula = cycles ~ rate + temp + I(rate^2) + I(temp^2) + I(rate *
##     temp))
##
## Residuals:
##       1       2       3       4       5       6       7       8       9      10
## -21.465   9.263  12.202  41.930  -5.842 -31.842  21.158 -25.404 -20.465   7.263
```

```
##      11
##  13.202
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     337.7215   149.9616   2.252   0.0741 .
## rate           -539.5175   268.8603  -2.007   0.1011
## temp              8.9171     9.1825   0.971   0.3761
## I(rate^2)       171.2171   127.1255   1.347   0.2359
## I(temp^2)        -0.1061     0.2034  -0.521   0.6244
## I(rate * temp)    2.8750     4.0468   0.710   0.5092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.37 on 5 degrees of freedom
## Multiple R-squared:  0.9135, Adjusted R-squared:  0.8271
## F-statistic: 10.57 on 5 and 5 DF,  p-value: 0.01086
```

Yes I do. The overall p value for the ANOVA is < 0.05, but each of the individual regression coefficient's p values are more than 0.05. This is a sign of multicollinearity. Additionaly, this is a polynomial regression that has not been centered so by definition it will have structural multicollinearity.

c. We can remove the high correlation between explanatory variables and their powers by centering.

```r
rate.code <- (rate-mean(rate))/0.4
temp.code <- (temp-mean(temp))/10
cor(cbind(rate.code,temp.code,rate.code^2,temp.code^2))
```

```
##                rate.code temp.code
## rate.code  1.000000e+00         0 -4.042173e-16 -1.994753e-17
## temp.code  0.000000e+00         1  0.000000e+00  0.000000e+00
##           -4.042173e-16         0  1.000000e+00  2.666667e-01
##           -1.994753e-17         0  2.666667e-01  1.000000e+00
```

In this new correlation matrix I do not obervs any high correlations and therefore signs of multicollinearity.

d.Fit a new full model with the scaled new predictor variables and report the estimated regression function

```r
mod2<-lm(cycles~rate.code+temp.code+I(rate.code^2)+I(temp.code^2)+I(rate.code*temp.code))
summary(mod2)
```

```
##
## Call:
## lm(formula = cycles ~ rate.code + temp.code + I(rate.code^2) +
##     I(temp.code^2) + I(rate.code * temp.code))
##
```

```
## Residuals:
##       1       2       3       4       5       6       7       8       9      10
## -21.465   9.263  12.202  41.930  -5.842 -31.842  21.158 -25.404 -20.465   7.263
##      11
##  13.202
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                162.84      16.61   9.805 0.000188 ***
## rate.code                  -55.83      13.22  -4.224 0.008292 **
## temp.code                   75.50      13.22   5.712 0.002297 **
## I(rate.code^2)              27.39      20.34   1.347 0.235856
## I(temp.code^2)             -10.61      20.34  -0.521 0.624352
## I(rate.code * temp.code)    11.50      16.19   0.710 0.509184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.37 on 5 degrees of freedom
## Multiple R-squared:  0.9135, Adjusted R-squared:  0.8271
## F-statistic: 10.57 on 5 and 5 DF,  p-value: 0.01086
```

```r
summary(mod2)$coeff[1,1]
```

```
## [1] 162.8421
```

Cycles = 162.8421053 + -55.8333333rate.code + 75.5temp.code + 27.3947368rate.code^2 + -10.6052632temp.code^2 + 11.5[rate.code * temp.code]