Lab 11

David Wiley / Duy Truong

March 24, 2019

Using Data from Problem 3.11:

a.

Determine if (x1,x2,x3,x4,x5) = (411, 22.5, 14.2, 40.3, 4.07) is an interpolation or extrapolation point. Should you be concerned in your previous estimation and prediction procedure?

```
# DATA

dat=read.csv("/home/david/Documents/2019 Spring/Applied Regression/Labs_HW/Data_Sets2/Data Sets/Appendi
y = dat$y
x1 = dat$x1
x2 = dat$x2
x3 = dat$x3
x4 = dat$x4
x5 = dat$x5

fit = lm(y~x1+x2+x3+x4+x5, dat)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.250  -4.438   0.125   5.250   9.500
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.208e+01  1.889e+01   2.757 0.020218 *
## x1           5.556e-02  2.987e-02   1.860 0.092544 .
## x2           2.821e-01  5.761e-02   4.897 0.000625 ***
## x3           1.250e-01  4.033e-01   0.310 0.762949
## x4           8.774e-17  2.016e-01   0.000 1.000000
## x5          -1.606e+01  1.456e+00 -11.035  6.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.065 on 10 degrees of freedom
## Multiple R-squared:  0.9372, Adjusted R-squared:  0.9058
## F-statistic: 29.86 on 5 and 10 DF,  p-value: 1.055e-05
```

```
dat
```

```
##     x1 x2 x3 x4   x5  y
## 1  415 25  5 40 1.28 63
## 2  550 25  5 40 4.05 21
## 3  415 95  5 40 4.05 36
```

```
## 4  550 95  5 40 1.28 99
## 5  415 25 15 40 4.05 24
## 6  550 25 15 40 1.28 66
## 7  415 95 15 40 1.28 71
## 8  550 95 15 40 4.05 54
## 9  415 25  5 60 4.05 23
## 10 550 25  5 60 1.28 74
## 11 415 95  5 60 1.28 80
## 12 550 95  5 60 4.05 33
## 13 415 25 15 60 1.28 63
## 14 550 25 15 60 4.05 21
## 15 415 95 15 60 4.05 44
## 16 550 95 15 60 1.28 96
```

```
x = as.matrix(cbind(1,dat[,1:5]))
H = x%*%solve(t(x)%*%x)%*%t(x)

eh = diag(H)

hmax = max(eh)

x01 = c(1, 411, 22.5, 14.2, 40.3, 4.07)

x01_point = t(x01)%*%solve(t(x)%*%x)%*%x01
```

Our hmax is 0.375 and the new point is 0.3715987. Since it is below the max we can say it is an interpolation point. We shouldn't be concerned with our previous estimiation and prediction procedure because this point isn't necessarily a data point used to fit the model. It's just a point that fits within the ellipsoid that encompasses our given data points.

b.

Is multi-collinearity present?

```
library(car)
vif(fit)
```

```
## x1 x2 x3 x4 x5
##  1  1  1  1  1
```

Our values are all 1 which is less than 10 as well as 5. Therefore, we don't have a concern for multi-collinearity.