PURDUE UNIVERSITY

IE 33000: Probability and Statistics in Engineering II (Fall 2022)
School of Industrial Engineering, Purdue University

Homework 5

Instruction: There are 4 problems in total.
1) Problems 1 and 3 – 20 points
2) Problems 4 – 40 points

Due November 4, 2022 (11:59 pm)

For all problems, provide both hand-written solutions and R codes (for 30 points bonus), wherever applicable.

Problem 1

Diabetes and obesity are serious health concerns in the United States and many of the developed world. Measuring the amount of body fat a person carries is one way to monitor weight control progress, but measuring it accurately involves either expensive X-ray equipment or a pool in which to dunk the subject. Instead body mass index (BMI) is often used as a proxy for body fat because it is easy to measure: BMI = 703 mass(lb)/(height(in))^2. In a study of 250 men at Brigham Young University, both BMI and body fat were measured. The researchers also measured 13 physical characteristics of each man, including his age (yrs), height (in), and waist size (in.).

(a) Write out the regression model if

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 2.9705 & -4.0042E-2 & -4.1679E-2 \\ -0.04004 & 6.0774E-4 & -7.3875E-5 \\ -0.00417 & -7.3875E-5 & 2.5766E-4 \end{bmatrix} \quad \text{and} \quad (\mathbf{X'y}) = \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

(b) What is the predicted body fat of a man who is 6-ft tall with a 34-in waist?

$$\hat{\beta} = (X'X)^{-1} \cdot X'y = \begin{bmatrix} -6744.13 \\ -0.59 \\ 1.76 \end{bmatrix} \Rightarrow \hat{y} = -6744.13 - 0.59x_1 + 1.76x_2$$

$$\hat{y}_{(6,34)} = -6744.13 - 0.59(72) + 1.76(34)$$
$$= -6726.77$$

Problem 2

The electric power consumed each month by a chemical plant is thought to be related to the average ambient temperature (x_1), the number of days in the month (x_2), the average product purity (x_3), and the tons of product produced (x_4). The past year's historical data are available and are presented below.
(a) Fit a multiple linear regression model to these data.
(b) Estimate σ^2.
(c) Predict power consumption for a month in which x_1 = 75°F, x_2 = 24 days, x_3 = 90%, and x_4 = 98 tons.

y	x_1	x_2	x_3	x_4
240	25	24	91	100
236	31	21	90	95
270	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

a) $$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 25 & 31 & \cdots & 38 \\ 24 & 21 & \cdots & 23 \\ 91 & 90 & \cdots & 89 \\ 100 & 95 & \cdots & 98 \end{bmatrix} \cdot \begin{bmatrix} 25 & 24 & 91 & 100 \\ 31 & 21 & 90 & 95 \\ \vdots & \vdots & \vdots & \vdots \\ 38 & 23 & 89 & 98 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 25 & 31 & \cdots & 38 \\ 24 & 21 & \cdots & 23 \\ 91 & 90 & \cdots & 89 \\ 100 & 95 & \cdots & 98 \end{bmatrix} \cdot \begin{bmatrix} 240 \\ 236 \\ \vdots \\ 261 \end{bmatrix} = \begin{bmatrix} -123 \\ 0.76 \\ 7.52 \\ 2.48 \\ -0.48 \end{bmatrix}$$

$$\hat{y} = -123 + 0.76x_1 + 7.52x_2 + 2.48x_3 - 0.48x_4$$

b) $$\sigma^2 = \frac{SSE}{n-p} = \frac{972}{12-5} = 138.86$$

c) $$\hat{y}_{(75,24,.9,98)} = -123 + 0.76(75) + 7.52(24) + 2.48(0.9) - 0.48(98) = 69.67$$

Problem 3

An article reported on a study that analyzed powdered mixtures of coal and limestone for permittivity. The errors in the density measurement were the response. The data are reported in the following table.

Density	Dielectric Constant	Loss Factor
0.749	2.05	0.016
0.798	2.15	0.02
0.849	2.25	0.022
0.877	2.3	0.023
0.929	2.4	0.026
0.963	2.47	0.028
0.997	2.54	0.031
1.046	2.64	0.034
1.133	2.85	0.039
1.17	2.94	0.042
1.215	3.05	0.045

(a) Fit a multiple linear regression model to these data with the density as the response.
(b) Estimate σ^2.
(c) Use the model to predict the density when the dielectric constant is 2.5 and the loss factor is 0.03.

a) $X'X = \begin{bmatrix} 1 & \cdots & 1 \\ 2.05 & 2.15 \cdots & 3.05 \\ 0.016 & 0.02 \cdots & 0.045 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2.05 & 0.016 \\ 1 & 2.15 & 0.02 \\ \vdots & \vdots & \vdots \\ 1 & 3.05 & 0.045 \end{bmatrix}$ $\quad X'Y = \begin{bmatrix} 1 & \cdots & 1 \\ 2.05 & 2.15 \cdots & 3.05 \\ 0.016 & 0.02 \cdots & 0.045 \end{bmatrix} \cdot \begin{bmatrix} 0.749 \\ \vdots \\ 1.215 \end{bmatrix}$

$\hat{y} = 0.111 + 0.407x_1 + 2.108x_2$

b) $\sigma^2 = \dfrac{SSE}{n-p}$

$= \dfrac{0.060624}{11-3}$

$= 0.000078$

c) $\hat{y}_{(2.5,0.3)} = 0.407 \cdot 2.5 + 2.108 \cdot 0.3 - 0.111$

$= 1.54$

Problem 4

A study was performed on the wear of a bearing and its relationship to x_1 = oil viscosity and x_2 = load. The following data were obtained.

y	x_1	x_2
293	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

(a) Fit a multiple linear regression model to these data.
(b) Estimate σ^2.
(c) Use the model to predict wear when x_1 = 25 and x_2 = 1000.
(d) Fit a multiple linear regression model with an interaction term to these data.
(e) Estimate σ^2 for this new model. How did these quantities change? Does this tell you anything about the value of adding the interaction term to the model?
(f) Use the model in part (d) to predict when x_1 = 25 and x_2 = 1000. Compare this prediction with the predicted value from part (c).

a) $$X^{-1}X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1.6 & 15.5 & \cdots & 40.0 \\ 851 & 816 & \cdots & 1115 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1.6 & 851 \\ 1 & 15.5 & 816 \\ \vdots & \vdots & \vdots \\ 1 & 40.0 & 1115 \end{bmatrix} = \begin{bmatrix} 6 & 155.1 & 6398 \\ 155.1 & 5624.81 & 178309.6 \\ 6398 & 178309.6 & 7036496 \end{bmatrix}$$

$$X^{-1}Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1.6 & 15.5 & \cdots & 40.0 \\ 851 & 816 & \cdots & 1115 \end{bmatrix} \cdot \begin{bmatrix} 293 \\ 230 \\ \vdots \\ 125 \end{bmatrix} = \begin{bmatrix} 383.8 \\ -3.64 \\ -0.11 \end{bmatrix}$$

$$\hat{y} = 383.8 - 3.64x_1 - 0.11x_2$$

c) $$\hat{y}_{(25,1000)} = 383.8 - 3.64(25) - 0.11(1000) = 182.8$$

b) $$\sigma^2 = \frac{SSE}{n-p} = 152.5$$

d) $$\hat{y} = x_1 + x_2 + (x_1 \cdot x_2)$$

* used R to solve this

$$\hat{y} = 483.97 - 7.66x_1 - 0.22x_2 + 0.004(x_1 \cdot x_2)$$

e) $\sigma^2 = 146.89$

f) $$\hat{y}_{(25,1000)} = 483.97 - 7.66(25) - 0.22(1000) + 0.004(25 \cdot 1000) = 172.5$$

```
##Problem 1
XX <- c(2.9705, -4.0042 * 10^-2, -4.1679*10^-2, -0.04004, 6.0774*10^-4,
        -7.3875 * 10^-5, -0.00417, -7.3875 * 10^-5, 2.5766 * 10^-4)
XX <- matrix(XX.data,nrow=3,ncol=3,byrow=TRUE)
XY <- c(4757.9, 334335.8, 179706.7)
XY <- matrix(XY.data, nrow = 3, ncol = 1,byrow=TRUE)
#a
yhat = XX * XY
#b
yhat6_34 <- yhat[1,1] + (yhat[2,1] * 6) + (yhat[3,1] * 34)

##Problem 2
y <- c(240, 236, ..., 261)
x1 <- c(25, 31, ..., 38)
x2 <- c(0.016, 0.02, ..., 0.045)
linreg <- lm(y~x1+x2)
summary(linreg)


##Problem 3
y=c(0.749,0.798,0.849,0.877,0.929,0.963,0.997,1.046,1.133,1.17,1.215)
x1=c(2.05,2.15,2.25,2.3,2.4,2.47,2.54,2.64,2.85,2.94,3.05)
x2=c(0.016,0.02,0.022,0.023,0.026,0.028,0.031,0.034,0.039,0.042,0.045)
model=lm(y~x1+x2)
summary(model)
yhat2.5_0.3 = -0.1105 + 2.5*0.41 +0.3 * 2.108

##Problem 4
y = c(293, 230, 172, 91,113,125)
x1 = c(1.6, 15.5, 22.0, 43.0, 33.0, 40.0)
x2 = c(851, 816, 1058, 1201, 1357, 1115)
model1=lm(y~x1+x2)
summary(model1)
yhat25_1000 = 383.8 - 25*3.64 - 0.11 * 1000
model2 = lm(y~x1+x2+x1*x2)
summary(model2)
yhat25_1000 = 483.97 - 25*7.66 - 0.22 * 1000 + 0.0041 * 25 *1000
```