205615894_stats101a_hw6

Takao

2/14/2022

Problem 1

(a)

```
matXtX <- matrix(c(1.17991, -7.30982e-3, 7.3006e-4, -7.30982e-3, 7.9799e-5, -1.2371e-4, 7.3006e-4, -1.2
matXty <- matrix(c(220, 36768, 9965), nrow = 3)
matXtX %*% matXty
```

```
##             [,1]
## [1,] -1.91221386
## [2,]  0.09308919
## [3,]  0.25334232
```

The regression coefficient in the model specified is -1.9122139, 0.0930892, 0.2533423. y hat = 0.093089(x1) + 0.253232(x2) -1.91221

(b)

```
yhat1 <- 0.093089*200 + 0.253232*50 -1.91221
yhat1
```

```
## [1] 29.36719
```

The predicted value of the absorption index y when x1 = 200 and x2 = 50 is 29.36719.

(c)

```
var1 <- 1 * matXtX
var1
```

```
##             [,1]         [,2]         [,3]
## [1,]  1.17991000 -0.007309820  0.000730060
## [2,] -0.00730982  0.000079799 -0.000123713
## [3,]  0.00073006 -0.000123710  0.000465760
```

The variances of the estimated coefficients are 1.17991, -0.0073098, 7.3006×10^{-4}, -0.0073098, 7.9799×10^{-5}, -1.2371×10^{-4}, 7.3006×10^{-4}, -1.23713×10^{-4}, 4.6576×10^{-4}

Problem 2

```r
electric_data <- read.csv("electric.csv", header = TRUE)
head(electric_data)
```

```
##     y x1 x2 x3  x4
## 1 240 25 24 91 100
## 2 236 31 21 90  95
## 3 270 45 24 88 110
## 4 274 60 25 87  88
## 5 301 65 25 91  94
## 6 316 72 26 94  99
```

(a)

```r
mlr.model1 <- lm(y~x1 + x2 + x3 + x4, data = electric_data)
mlr.model1
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4, data = electric_data)
##
## Coefficients:
## (Intercept)           x1           x2           x3           x4
##   -123.1312       0.7573       7.5188       2.4831      -0.4811
```

Estimated regression coefficients: x1: 0.7573 x2: 7.5188 x3: 2.4831 x4: -0.4811 intercept: -123.1312 y hat = 0.7573(x1) + 7.5188(x2) + 2.4831(x3) - 0.4811(x4) - 123.1312

(b)

Estimated regression coefficient: x2: 7.5188 The electric power consumed each month by a chemical plant increases 7.5188 Megawatts per hour (MWh) when the number of days in the month increases by 1 day, when all other factors that influence the electric power consumption are fixed.

x4: -0.4811 The electric power consumed each month by a chemical plant decreases 0.4811 Megawatts per hour (MWh) when the product produced increases by 1 ton, when all other factors that influence the electric power consumption are fixed.

(c)

```r
e <- mlr.model1$residuals # Extract the residuals.
n <- length(e)
p <- 4
est.variance1 <- sum(e^2)/(n- p - 1)
print(est.variance1)
```

```
## [1] 138.9234
```

Estimation of the error variance is 138.9234078.

(d)

```
summary(mlr.model1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4, data = electric_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.098  -9.778   1.767   6.798  13.016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -123.1312   157.2561  -0.783    0.459
## x1             0.7573     0.2791   2.713    0.030 *
## x2             7.5188     4.0101   1.875    0.103
## x3             2.4831     1.8094   1.372    0.212
## x4            -0.4811     0.5552  -0.867    0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.79 on 7 degrees of freedom
## Multiple R-squared:  0.852,  Adjusted R-squared:  0.7675
## F-statistic: 10.08 on 4 and 7 DF,  p-value: 0.00496
```

The standard errors of the regression coefficients: x1: 0.2791 x2: 4.0101 x3: 1.8094 x4: 0.5552 intercept: 157.2561

As we can see, the standard errors of the regrssion coefficients are fairly distinct, which indicates that not all of the model parameters estimated with the same precision.

(e)

```
# y hat = 0.7573(x1) + 7.5188(x2) + 2.4831(x3) - 0.4811(x4) - 123.1312

powerconsumption <- 0.7573*75 + 7.5188*24 + 2.4831*90 - 0.4811*98 - 123.1312
powerconsumption
```

```
## [1] 290.4487
```

The predicted power consumption for a month in which x1 = 75 degrees farenheit, x2 = 24 days, x3 = 90 percent, and x4 = 98 tons is 290.4487 MWh.

(f)

```
summary(mlr.model1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4, data = electric_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.098  -9.778   1.767   6.798  13.016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -123.1312   157.2561  -0.783    0.459
## x1             0.7573     0.2791   2.713    0.030 *
## x2             7.5188     4.0101   1.875    0.103
## x3             2.4831     1.8094   1.372    0.212
## x4            -0.4811     0.5552  -0.867    0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.79 on 7 degrees of freedom
## Multiple R-squared:  0.852,  Adjusted R-squared:  0.7675
## F-statistic: 10.08 on 4 and 7 DF,  p-value: 0.00496
```

At a level of $\alpha = 0.05$, we draw the following conclusions from the t-tests:

1. The average ambient temperature has a significant association with the electric power consumed each month by a chemical plant, when the predictors number of days in the month, product purity, and tons of product produced are in the model.
2. The number of days in the month does not have a significant association with the electric power consumed each month by a chemical plant, when the predictors ambient temperature, product purity, and tons of product produced are in the model.
3. The average product purity does not have a significant association with the electric power consumed each month by a chemical plant, when the predictors ambient temperature, number of days in the month, and tons of product produced are in the model.
4. The tons of product produced does not have a significant association with the electric power consumed each month by a chemical plant, when the predictors, ambient temperature, number of days in the month, and product purity are in the model.

(g)

```r
confint(mlr.model1, level = 0.95)
```

```
##                     2.5 %      97.5 %
## (Intercept) -494.98273363 248.7202411
## x1             0.09734663   1.4172315
## x2            -1.96364640  17.0012143
## x3            -1.79543852   6.7615956
## x4            -1.79391356   0.8316431
```

Confidence Intervals: x1: (0.09734663, 1.4172315) x2: (-1.96364640, 17.0012143) x3: (-1.79543852, 6.7615956) x4: (-1.79391356, 0.8316431) Intercept: (-494.98273363, 248.7202411)

As we can see in the various confidence interval, x1 is the only CI that does not contain the value "0" in the interval, which indicates validation to the conclusion we made in part (f).