Linear Model Project

Jieqi Tu (jtu22)

11/30/2020

Question 1

In this problem, Let $\mathbf{y}$ denote stiffness (lb/in^2) and $\mathbf{X}$ denote the design matrix. We want to investigate the relationship between stiffness and density(lb/ft^3). Since here we have 30 observations, and we only have one predictor and one outcome, the dimension of $\mathbf{y}$ is 30×1 and the dimension of $\mathbf{X}$ is 30×2. Then we have:

$$\mathbf{y} = \begin{bmatrix} 2622 \\ 22148 \\ 26751 \\ 18036 \\ 96305 \\ \vdots \\ 49499 \\ 25312 \end{bmatrix}_{30\times 1}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 15.0 \\ 1 & 14.5 \\ 1 & 14.8 \\ 1 & 13.6 \\ 1 & 25.6 \\ \vdots & \vdots \\ 1 & 16.7 \\ 1 & 15.4 \end{bmatrix}_{30\times 2}$$

Here the first column in the design matrix allows estimation of the y-intercept and the second column contains the values of our variable (density).
Then we construct a linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Here, we also have:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{30} \end{bmatrix}_{30\times 1}$$

Here we also assume that

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2)$$

Before running into questions, we firstly want to show

$$result1: \mathbf{X}^{'}\mathbf{X} = \begin{bmatrix} 30 & 464.1 \\ 464.1 & 8166.29 \end{bmatrix}$$

$$result2: (\mathbf{X}^{'}\mathbf{X})^{-1} = \begin{bmatrix} 0.2758892 & -0.0156791 \\ -0.0156791 & 0.001013517 \end{bmatrix}$$

$$result3: \mathbf{X}^{'}\mathbf{y} = \begin{bmatrix} 1017405 \\ 19589339 \end{bmatrix}$$

, and

$$result4: s^2 = 165242295.59$$

Below is my code to show these four results.

```
# Import data
x = c(rep(1, 30), 15.0, 14.5, 14.8, 13.6, 25.6, 23.4, 24.4, 23.3, 19.5, 21.2, 22.8, 21.7, 19.8, 21.3, 9
      8.4, 9.8, 11.0, 8.3, 9.9, 8.6, 6.4, 7.0, 8.2, 17.4, 15.0, 15.2, 16.4, 16.7, 15.4)
y = c(2622, 22148, 26751, 18036, 96305, 104170, 72594, 49512, 32207, 48218, 70453, 47661, 38138, 54045,
      17502, 14007, 19443, 7573, 14191, 9714, 8076, 5304, 10728, 43243, 25319, 28028, 41792, 49499, 253

# Create design matrix
x.matrix = matrix(x, ncol = 2, nrow = 30, byrow = F)

# Result 1
result1 = t(x.matrix) %*% x.matrix
result1
```

```
##       [,1]    [,2]
## [1,]  30.0  464.10
## [2,] 464.1 8166.29
```

```
# Result 2
solve(result1)
```

```
##             [,1]         [,2]
## [1,]  0.27588920 -0.015679112
## [2,] -0.01567911  0.001013517
```

```
# Result 3
t(x.matrix) %*% y
```

```
##          [,1]
## [1,]  1017405
## [2,] 19589339
```

```
# Result 4
density = c(15.0, 14.5, 14.8, 13.6, 25.6, 23.4, 24.4, 23.3, 19.5, 21.2, 22.8, 21.7, 19.8, 21.3, 9.5,
            8.4, 9.8, 11.0, 8.3, 9.9, 8.6, 6.4, 7.0, 8.2, 17.4, 15.0, 15.2, 16.4, 16.7, 15.4)
data = data.frame(y, density)
model1 = lm(y~density, data = data)
summary(model1)$sigma^2
```

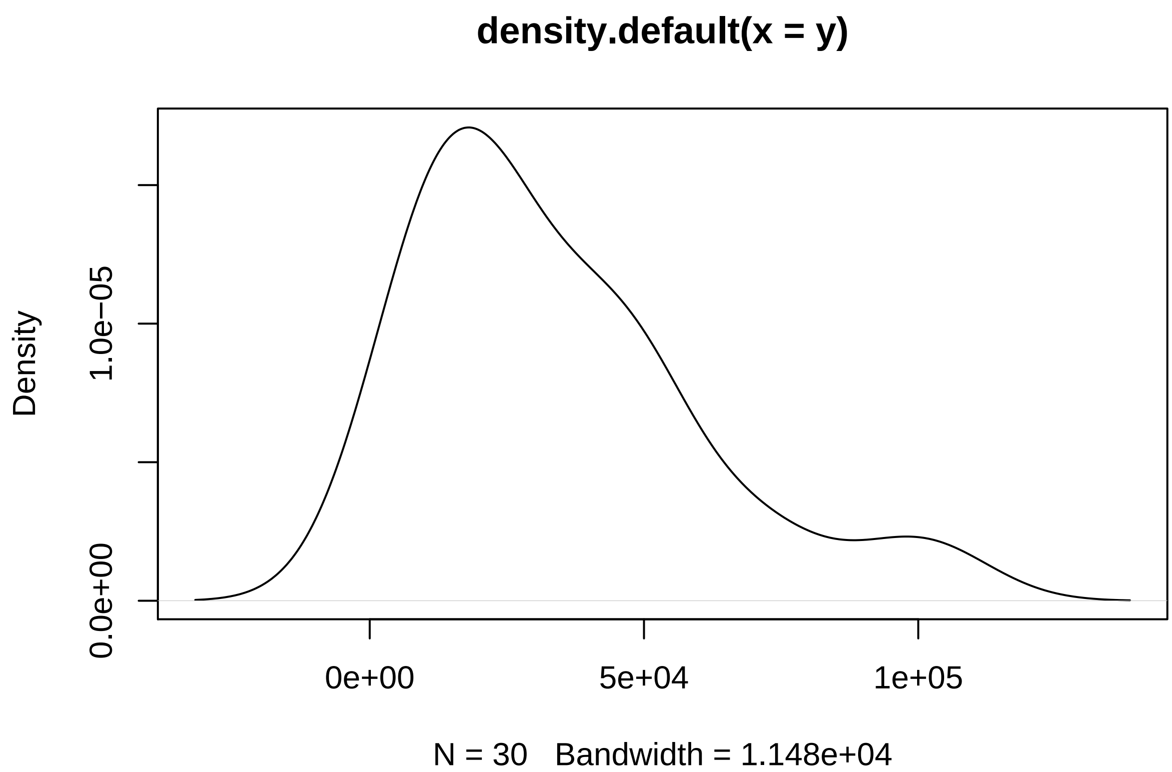
```
## [1] 165242296
```

All these four results are confirmed.

(a)

Estimate $\boldsymbol{\beta}$.

```r
# Check the distribution of y (it is approximately normally-distributed)
plot(density(y))
```




```r
# Fit a linear model
model1 = lm(y~density, data = data)
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ density, data = data)
##
## Residuals:
##    Min      1Q Median     3Q    Max
## -29458  -8036   1905   4394  39313
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -26452.4     6751.9  -3.918 0.000524 ***
## density       3902.1      409.2   9.535 2.72e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 12850 on 28 degrees of freedom
## Multiple R-squared:  0.7645, Adjusted R-squared:  0.7561
## F-statistic: 90.92 on 1 and 28 DF,  p-value: 2.724e-10
```

From the result, we could see that, the estimates of $\boldsymbol{\beta}$ are: $\beta_0 = -26452.4, \beta_1 = 3902.1$. The p-values for these two estimates are both less than 0.001, which indicates that they are significant different from 0.

(b)

Find a point estimate for the mean stiffness reading when the density of the particleboard is 10 lb/ft^3. Find a 95% confidence interval on this mean reading.

```
newdata = data.frame(
  density = 10
)
predict(model1, newdata = newdata, interval = "confidence")
```

```
##        fit      lwr     upr
## 1 12568.87 5925.235 19212.5
```

So the point estimate for the stiffness reading when density = 10 is 12568.87 lb/in^2. The 95% confidence interval of the point estimate is [5925.235, 19212.5].

(c)

Find a point estimate for the stiffness reading of an individual particleboard whose density is 10 lb/ft^3. Find a 95% prediction interval on the stiffness reading for such a board.

```
newdata = data.frame(
  density = 10
)
predict(model1, newdata = newdata, interval = "prediction")
```

```
##        fit      lwr      upr
## 1 12568.87 -14587.9 39725.63
```

The point estimate is the same as what has been shown in part (b). The 95% prediction interval for this individual is [-14587.9, 39725.63]

(d)

Find a 95% confidence interval on the slope of the regression line.

```
# Calculate the confidence interval for the slope of the regression line
confint(model1, 'density', level = 0.95)
```

```
##           2.5 %   97.5 %
## density 3063.84 4740.413
```

The 95% confidence interval on the slope of the regression line is [3063.87, 4740.413].

(e)

Find a 95% joint confidence region on the pair of parameters (β_0, β_1)

```
# Calculate joint 95% confidence region
joint.ci.bonf(model1, conf=0.95)
```

```
##                  2.5 %     97.5 %
## (Intercept) -42444.015 -10460.772
## density       2932.865   4871.387
```