ICPSR Regression II - Problem Set 2

Ziyuan Gao

July 2024

1 Question 1

1.1

A is 3x3 matrix.

1.2

$$A + B = \begin{pmatrix} 5+1 & 4+1 & 8+2 \\ -3+2 & 1+0 & 2+9 \\ 8+5 & 0+8 & 4+7 \end{pmatrix}$$
$$= \begin{pmatrix} 6 & 5 & 10 \\ -1 & 1 & 11 \\ 13 & 8 & 11 \end{pmatrix}$$

1.3

$$BA = \begin{pmatrix} 18 & 5 & 18 \\ 82 & 8 & 52 \\ 57 & 28 & 84 \end{pmatrix}$$

1.4

$$CA = \begin{pmatrix} 55 & 6 & 40 \\ 83 & 19 & 74 \end{pmatrix}$$

2 Question 2

2.1

$$X' = \begin{pmatrix} 1 & 1 & 1 \\ 4 & 3 & 9 \\ 2 & 5 & 0 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 21 & 23 & 37 \\ 23 & 35 & 28 \\ 37 & 28 & 82 \end{pmatrix}$$

2.2

$$X'y = \begin{pmatrix} 12 \\ 61 \\ 24 \end{pmatrix}$$

3 Question 3

$$[M]_{ij} = (-1)^{i+j}\det(M_{j,i})$$

$$= \begin{pmatrix} |M_{1,1}| & -|M_{2,1}| & |M_{3,1}| \\ -|M_{1,2}| & |M_{2,2}| & -|M_{3,2}| \\ |M_{1,3}| & -|M_{2,3}| & |M_{3,3}| \end{pmatrix}$$

$$= \begin{pmatrix} -27 & 12 & 46 \\ 0 & 0 & -1 \\ 6 & -3 & -10 \end{pmatrix}$$

$$\det(M) = -3$$

$$M' = \frac{1}{det(M)}[M]_{ij} = \begin{pmatrix} 9 & -4 & -\frac{46}{3} \\ 0 & 0 & \frac{1}{3} \\ -2 & 1 & \frac{10}{3} \end{pmatrix}$$

$$MM' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I$$

4 Question 4

$$e = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ e_n \end{pmatrix}$$

$$e' = \begin{pmatrix} e_1 & e_2 & e_3 & ... & e_n \end{pmatrix}$$

$$ee' = \begin{pmatrix} e_1e_1 & e_1e_2 & e_1e_3 & ... & e_1e_n \\ e_2e_1 & e_2e_2 & e_2e_3 & ... & e_2e_n \\ e_3e_1 & e_3e_2 & e_3e_3 & ... & e_3e_n \\ ... & ... & ... & ... & ... \\ e_ne_1 & e_ne_2 & e_ne_3 & ... & e_ne_n \end{pmatrix}$$

$$e'e = \left(e_1^2 + e_2^2 + e_3^2 ... e_n^2\right)$$

5 Question 5

5.1

$$T = \begin{pmatrix} 8 \\ 13 \\ 5 \end{pmatrix}$$

$$C = \begin{pmatrix} 5 \\ 9 \\ 2 \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 5 \\ 1 & 9 \\ 1 & 2 \end{pmatrix}$$

$$X' = \begin{pmatrix} 1 & 1 & 1 \\ 5 & 9 & 2 \end{pmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{pmatrix} 2.54 \\ 1.15 \end{pmatrix}$$

$$For\ T_i = b_1 + b_2 C_i + e_i : \quad b_1 = 2.54, b_2 = 1.15$$

5.2

$$\hat{y_1} = 5 * 1.15 + 2.54 = 8.29$$

$$\hat{y_2} = 9 * 1.15 + 2.54 = 12.89$$

$$\hat{y_3} = 2 * 1.15 + 2.54 = 4.84$$

5.3

$$e_1 = 8 - 8.29 = -0.29$$

$$e_2 = 13 - 12.89 = 0.11$$

$$e_3 = 5 - 4.84 = 0.16$$

5.4

$$\text{SSR} = \sum_{i=1}^{n} e_i^2 = (-0.29)^2 + (0.11)^2 + (0.16)^2 = 0.1218$$

$$\text{SEE} = \frac{SSR}{\text{df}} = sqrt(\frac{0.12}{3-2}) = 0.3490$$

5.5

$$\text{Var}(\hat{\beta}) = SSR(X^T X)^{-1} = \begin{pmatrix} 0.1811 & -0.0263 \\ -0.0263 & 0.0049 \end{pmatrix}$$

$$s_{\beta_1} = sqrt(0.1811) = 0.4255$$

$$s_{\beta_2} = sqrt(0.0049) = 0.0703$$

5.6

$$t_1 = \frac{\beta_1}{s_{\beta_1}} = \frac{2.54}{0.4255} = 5.9694$$

$$t_2 = \frac{\beta_2}{s_{\beta_2}} = \frac{1.15}{0.0703} = 16.3585$$

5.7

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{0.1218}{32.6} = 0.9963$$

6 Question 6

Below is R code for linear model:

```
# Function to calculate beta_hat
calculate_beta_hat <- function(t,x){
  c <- cbind(1, x)
  c_t <- t(c)
  beta_hat <- solve(c_t %*% c) %*% c_t %*% t
  return(beta_hat)
}

# Function to calculate residuals
calculate_residual <- function(t,x,beta_hat){
  c <- cbind(1, x)
  y_hat <- c %*% beta_hat
  residuals <- t - y_hat
  return(residuals)
}

# Function to calculate the sum of squared residuals (SSR)
calculate_SSR <- function(residuals){
  SSR <- sum(residuals^2)
  return(SSR)
}

# Function to calculate the standard error of estimate (SEE)
calculate_SEE <- function(SSR,df){
  SEE <- sqrt(SSR/df)
  return(SEE)
}

# Function to calculate variance of beta
calculate_var_b <- function(SSR,x){
  c <- cbind(1, x)
  c_t <- t(c)
  var_b <- SSR * solve(c_t %*% c)
  return(var_b)
}

# Function to calculate total sum of squares (SST)
calculate_SS_tot <- function(t){
  t_mean <- mean(t)
  y_e <- (t - t_mean)^2
  SST <- sum(y_e)
  return(SST)
}

# Function to calculate t-statistics for beta
calculate_t_beta <- function(beta_hat, var_b){
  SE_beta <- sqrt(diag(var_b))
  t_beta <- beta_hat / SE_beta
  return(t_beta)
}
```

```
# Data input
x <- matrix(c(39.4, 40.1, 44.3, 38.2, 48.4, 41.9, 45.9, 41.2,  # Average Age
              5511.8, 4855.2, 3825.5, 5600.6, 3974.4, 3847.2, 5081.2, 4382.9),
            nrow = 8, byrow = FALSE)

t <- matrix(c(29.3, 31.8, 44.3, 27.2, 57.6, 39.7, 53.8, 32.6), nrow = 8, byrow = FALSE)

# Calculations
SS_tot <- calculate_SS_tot(t)
df <- nrow(t) - 3
beta_hat <- calculate_beta_hat(t, x)
residuals <- calculate_residual(t, x, beta_hat)
SSR <- calculate_SSR(residuals)
SEE <- calculate_SEE(SSR, df)
var_b <- calculate_var_b(SSR, x)
t_beta <- calculate_t_beta(beta_hat, var_b)

# Output results
cat("Beta_hat:\n", beta_hat, "\n")
cat("SSR:\n", SSR, "\n")
cat("SEE:\n", SEE, "\n")
cat("var_b:\n", var_b, "\n")
cat("Standard Errors of Beta:\n", sqrt(diag(var_b)), "\n")
```

$$StandardErrorsofBeta:$$

$$s_{\beta_1} = 36.75726$$

$$s_{\beta_2} = 0.6302174$$

$$s_{\beta_3} = 0.003019045$$

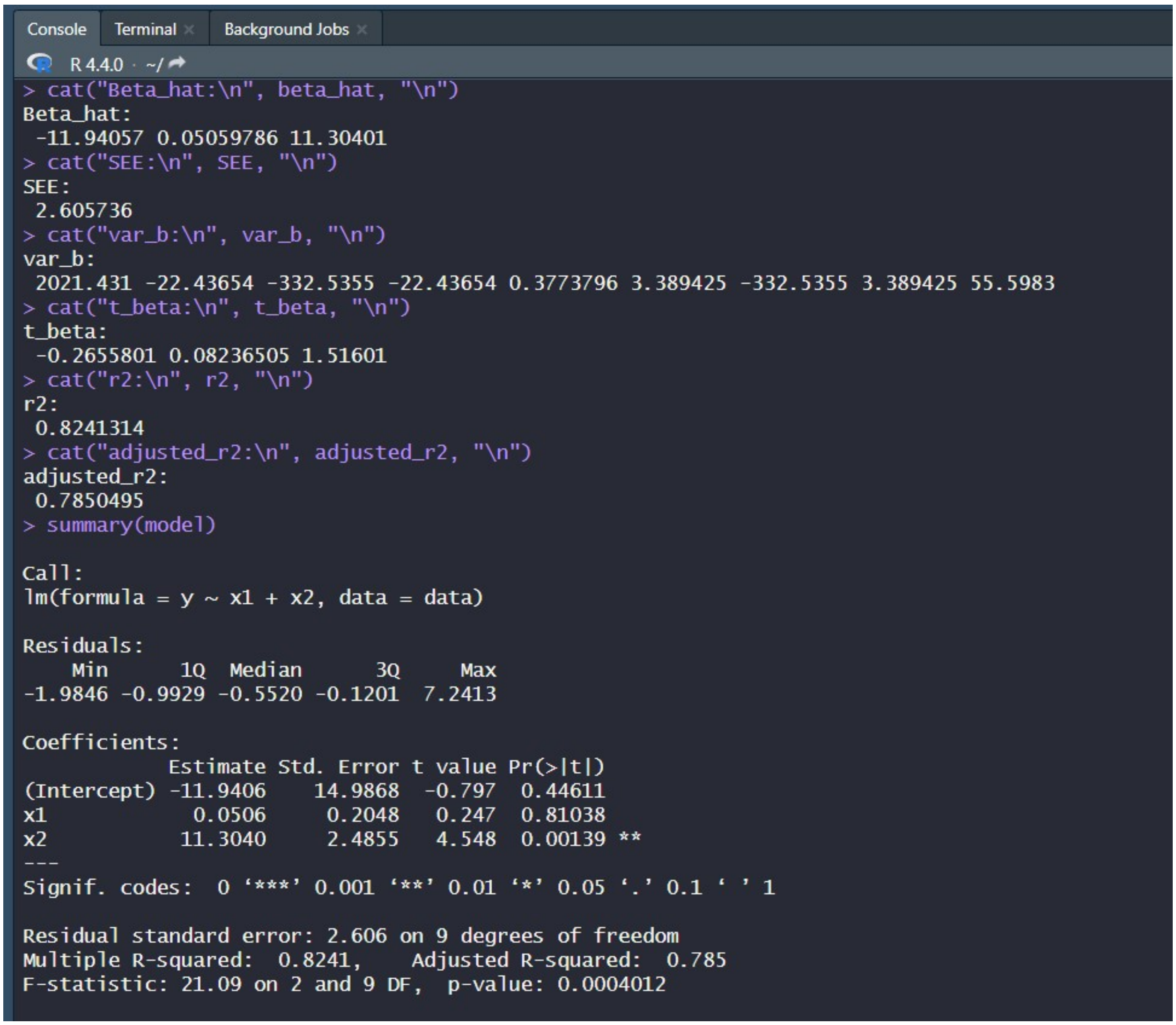
$$TstatsforBeta:$$

$$\beta_1 = -3.008724$$

$$\beta_2 = 5.382235$$

$$\beta_3 = 0.4448892$$

7 Question 7



```
Console   Terminal   Background Jobs
R 4.4.0 · ~/
> cat("Beta_hat:\n", beta_hat, "\n")
Beta_hat:
 -11.94057 0.05059786 11.30401
> cat("SEE:\n", SEE, "\n")
SEE:
 2.605736
> cat("var_b:\n", var_b, "\n")
var_b:
 2021.431 -22.43654 -332.5355 -22.43654 0.3773796 3.389425 -332.5355 3.389425 55.5983
> cat("t_beta:\n", t_beta, "\n")
t_beta:
 -0.2655801 0.08236505 1.51601
> cat("r2:\n", r2, "\n")
r2:
 0.8241314
> cat("adjusted_r2:\n", adjusted_r2, "\n")
adjusted_r2:
 0.7850495
> summary(model)

Call:
lm(formula = y ~ x1 + x2, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9846 -0.9929 -0.5520 -0.1201  7.2413

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.9406    14.9868  -0.797  0.44611
x1            0.0506     0.2048   0.247  0.81038
x2           11.3040     2.4855   4.548  0.00139 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.606 on 9 degrees of freedom
Multiple R-squared:  0.8241,    Adjusted R-squared:  0.785
F-statistic: 21.09 on 2 and 9 DF,  p-value: 0.0004012
```

Figure 1: summary of the R results