ESS212 HW4

Yiting Yu

2024-03-17

Yiting's Github repository url

Problem 1

$$y_i = ax_i^2 + bx_i + c + e_i \quad e_i \sim N(0, \sigma^2)$$

$$\hat{\beta} = \begin{bmatrix} c \\ b \\ a \end{bmatrix} \mathbf{A} = \begin{bmatrix} \mathbf{1} & \mathbf{t} & \mathbf{t^2} \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ . & . & . \\ . & . & . \\ 1 & t_{10} & t_{10}^2 \end{bmatrix}$$

$$\hat{\beta} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{A}\hat{\beta}$$

$$\sigma^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{\Sigma_{i=1}^n e_i^2}{n-k}$$

$$H = \sigma^2(\mathbf{A}'\mathbf{A})^{-1}$$

The following R codes are based on the above calculation process.

```r
t <- seq(1:10)
y <- c(-2.73,-2.71,-2.65,-0.87,-3.10,-1.03,0.63,1.46,5.90,8.38)

n <- length(y)
k <- 3  # the number of parameters

# the design matrix
A <- cbind(1,t,t^2)

# estimate parameters
beta <- solve(t(A)%*%A) %*% t(A)%*%y

# the error vector
residuals <- y - A%*%beta

# estimate residual variance
sigma2 <- sum(residuals^2)/(n-k)

# covariance matrix
cov_matrix <- solve(t(A)%*%A)*sigma2

# standard errors of the estimates
std_errors <- sqrt(diag(cov_matrix))
```

```r
# prediction for y(t = 12)
A_new <- cbind(1, 12, 12^2)
y_pred <- A_new %*% beta

# standard error of the prediction
pred_var <- A_new %*% cov_matrix %*% t(A_new)
pred_std_error <- sqrt(diag(pred_var))

# results
cat("Estimated parameters: ", beta, "\n")
```

```
## Estimated parameters:  -1.0255 -1.310614 0.2223864
```

```r
cat("Standard errors: ", std_errors, "\n")
```

```
## Standard errors:  1.194821 0.4990075 0.04421019
```

```r
cat("Predicted y for x = 12: ", y_pred, "\n")
```

```
## Predicted y for x = 12:  15.27077
```

```r
cat("Standard error of the prediction: ", pred_std_error, "\n")
```

```
## Standard error of the prediction:  1.70034
```

The estimated a is 0.2223864, with a standard error of 0.04421019; the estimated b is -1.310614, with a standard error of 0.4990075; and the estimated c is -1.0255, with a standard error of 1.194821.

The estimated value of $y(t = 12)$ is 15.27077, with a standard error of 1.70034.

Problem 2

$$y_i = mx_i + b + e_i \quad e_i \sim N(0, \sigma^2)$$

In this problem, we only know the error of the first two measurements. Although the last three measurements of the error are unknown, they are assumed to have a constant variance. Therefore, I used the average of the errors in the first two measurements to estimate the error in the last three measurements. Then, I used Weighted Least Squares (WLS) to adjust for heteroscedasticity across observations.

$$\hat{\beta} = \begin{bmatrix} b \\ m \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x} \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \\ 1 & t_5 \end{bmatrix}$$

To find the estimates that minimize the weighted sum of squared residuals, we set the derivative of the objective function to zero to solve for $\hat{\beta}$.

$$
\begin{aligned}
\frac{\partial}{\partial \beta}(\mathbf{y}-\mathbf{X}\hat{\beta})'\mathbf{W}(\mathbf{y}-\mathbf{X}\hat{\beta}) &= 0 \\
\mathbf{X}'\mathbf{W}\mathbf{X}\hat{\beta} &= \mathbf{X}'\mathbf{W}\mathbf{y} \\
\hat{\beta} &= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \\
\mathbf{e} &= \mathbf{y}-\mathbf{X}\hat{\beta} \\
\sigma^2 &= \frac{\Sigma_{i=1}^{n} w_i e_i^2}{n-k} \\
H &= \sigma^2(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}
\end{aligned}
$$

The following R codes are based on the above calculation process.

```r
x <- c(0,1,2,3,4)
y <- c(0.0434,1.0343,-0.2588,3.68622,4.3188)
err_known <- c(0.1, 0.1)

# estimate the error of the last three measurements
var_known <- mean(err_known^2)
weights <- c(1/err_known^2, rep(1/var_known,3))

# WLS
W <- diag(weights)

# the design matrix
X <- cbind(1,x)

# estimate parameters
beta <- solve(t(X)%*%W%*%X) %*% t(X)%*%W%*%y

# estimate residuals and residual variance
residuals <- y - X%*%beta
sigma2 <- sum(weights*residuals^2)/(length(y)-ncol(X))

# covariance matrix
cov_matrix <- sigma2*solve(t(X)%*%W%*%X)

# standard errors of the coefficients
std_errors <- sqrt(diag(cov_matrix))

# results
cat("Coefficients:\n")
```

```
## Coefficients:
```

```r
print(beta)
```

```
##         [,1]
##   -0.475760
## x  1.120272
```

```r
cat("\nStandard Errors:\n")
```

```
##
## Standard Errors:
```

```r
print(std_errors)
```

```
##                       x
## 1.0253275 0.4185882
```

The estimated m is 1.120272, with a standard error of 0.4185882; the estimated b is -0.475760, with a standard error of 1.0253275.

Problem 3

$$y_{i,j} = e_{i,j}exp(-bt_j) \quad log(e_{i,j}) \sim N(\mu, \sigma^2)$$
$$log(y_{i,j}) = log(e_{i,j}) - bt_j$$

$$\hat{\beta} = \begin{bmatrix} \mu \\ -b \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{t} \end{bmatrix} = \begin{bmatrix} 1 & t_{11} \\ 1 & t_{12} \\ . & . \\ . & . \\ 1 & t_{45} \end{bmatrix}$$

$$\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'}\mathbf{log}(\mathbf{y})$$
$$\mathbf{e} = \mathbf{log}(\mathbf{y}) - \mathbf{A}\hat{\beta}$$
$$\sigma = \sqrt{\frac{\mathbf{e'e}}{n-k}} = \sqrt{\frac{\Sigma_{i=1}^{n} e_i^2}{n-k}}$$
$$H = \sigma^2(\mathbf{X'X})^{-1}$$

The following R codes are based on the above calculation process.

```r
y <- c(3.75, 0.93, 0.38, 0.05, 0.04,
       0.36, 0.32, 0.11, 0.15, 0.03,
       0.58, 0.67, 0.12, 0.05, 0.08,
       2.06, 1.01, 0.60, 0.11, 0.06)
log_y <- log(y)

# the design matrix X
t <- rep(c(1,2,3,4,5), times=4)  # repeated for each y_ij
X <- cbind(1,t)

# estimate parameters
beta <- solve(t(X)%*%X) %*% t(X)%*%log_y

# residuals and estimate sigma
n <- 20
k <- 2
residuals <- log_y - X%*%beta
sigma_est <- sqrt(sum(residuals^2)/(n-k))

# covariance matrix
cov_beta <- solve(t(X)%*%X)*sigma_est^2

# standard errors for beta
se_beta <- sqrt(diag(cov_beta))

# estimates and standard errors for mu and b
mu_est <- beta[1]
```

```r
b_est <- -beta[2]  # change the sign
mu_se <- se_beta[1]
b_se <- se_beta[2]

# results
cat("mu estimate:", mu_est, "with SE:", mu_se, "\n")
```

```
## mu estimate: 1.063857 with SE: 0.3605393
```

```r
cat("b estimate:", b_est, "with SE:", b_se, "\n")
```

```
## b estimate: 0.839466 with SE: 0.1087067
```

```r
cat("sigma estimate:", sigma_est, "\n")
```

```
## sigma estimate: 0.6875214
```

The estimated μ is 1.063857, with a standard error of 0.36053932; the estimated b is 0.839466, with a standard error of 0.108706; the estimated σ is 0.6875214.