

Advanced Applied Statistics Homework 2 Solutions

Joshua D. Ingram

2022-09-13

1. Linear Model Theory

Consider the following linear model for the response vector \mathbf{y} :

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

where we assume

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

and hence

$$\boldsymbol{\mu} = E[\mathbf{y}] = X\boldsymbol{\beta}$$

Note that to multiply two matrices A and B in R, you write `A %*% B`, and to get the inverse of a matrix A, you write `solve(A)`. To get the transpose of a matrix A, use `t(A)`.

```
# Relevant objects created in HW1

# Create response vector y
y <- matrix(c(2,4,3), nrow = 3, ncol = 1)
# Create model matrix X
X <- matrix(c(1, 1, 1, 1, 2, 3), nrow = 3)
# fit simple linear regression model
fit <- lm(y~ X[,2])
# Create Hat-matrix
H <- X %*% solve(t(X)%*%X) %*% t(X)
```

1. (practical) Compute \mathbf{r} directly (i.e., using $\mathbf{r} = (I - H)\mathbf{y}$) for our little dataset and compare it to what you get from the R command `residuals(fit)`.

Answer:

```
# Manually calculate residuals with H
I <- diag(3)
r <- (I - H) %*% y
print(r)
```

```
##      [,1]
## [1,] -0.5
## [2,]  1.0
## [3,] -0.5
```

```
# Residuals from `residuals()`
print(residuals(fit))
```

```
##      1      2      3
## -0.5  1.0 -0.5
```

The residuals manually calculated with H are the same as those given by the `residuals()` function.

m. (theoretical) Show that $E[\mathbf{r}] = \mathbf{0}$.

Answer:

$$E[\mathbf{r}] = E[(I - H)\mathbf{y}] = (I - H)E[\mathbf{y}] = (I - H)X\boldsymbol{\beta} = X\boldsymbol{\beta} - HX\boldsymbol{\beta}$$

$$X\boldsymbol{\beta} - X(X'X)^{-1}X'X\boldsymbol{\beta} = X\boldsymbol{\beta} - X\boldsymbol{\beta} = \mathbf{0}.$$

n. (theoretical) Show that $\text{Var}[\mathbf{r}] = \sigma^2(I - H)$.

Answer:

$$\text{Var}[\mathbf{r}] = \text{Var}[(I - H)\mathbf{y}] = (I - H)\text{Var}[\mathbf{y}](I - H)' = (I - H)\sigma^2 I(I - H)'$$

$$= \sigma^2(I - H)(I - H)' = \sigma^2(I - H).$$

$(I - H)(I - H)' = (I - H)$ due to properties of symmetry and idempotency.

o. (theoretical) Are the residuals independent, like the components of the error term $\boldsymbol{\epsilon}$? Why or why not?

Answer:

The residuals are not independent. The off-diagonals of the variance-covariance matrix for the residuals are non-zero due to the hat-matrix H .

p. (theoretical) Are the residuals homoscedastic (i.e., do they have the same variance), like the components of the error term $\boldsymbol{\epsilon}$? Why or why not?

Answer:

The residuals are not homoscedastic, they are heteroscedastic unlike the components of the error term $\boldsymbol{\epsilon}$. The off-diagonals of the variance-covariance matrix of the residuals depend on those of the hat-matrix H , and for each i th residual this will differ.

2. Read the “Review of Simple Linear Regression” handout under “R handouts” under “Files”.

Many of the steps should look familiar, and with what we went over in class, they may make even more sense. Please let me know in class on Thursday if you have any questions!

3. Read the R handout on “Multiple Regression: Model Building and Inference” up to page 12, i.e., including Section 4.

There is even mention of the likelihood there, and the value of the maximized log-likelihood. All this should make a bit more sense now that we have introduced the likelihood.