Statistical Data Analysis Sheet 06 - Group: SDAK

December 14, 2021

1 Exercise2

Contributors:

Gunjan Mishra (gunjan.mishra@uni-potsdam.de)
Niklas Kaspareit (niklas.kaspareit@uni-potsdam.de)
Keerthan Shetty (keerthan.shetty@uni-potsdam.de)
Jan Holger Lemcke (jan.lemcke@uni-potsdam.de)

Show that $E[\hat{\epsilon}] = 0$ and determine $Cov(\hat{\epsilon})$ for the linear regression problem with $\hat{\beta} = (X^TX)^{-1}X^Ty$.

$$
\begin{aligned}
\hat{\epsilon} &= y - \hat{y} \\
&= y - X\hat{\beta} && |\hat{y} = X\hat{\beta} \\
&= y - X(X^TX)^{-1}X^Ty && |\hat{\beta} = (X^TX)^{-1}X^Ty \\
&= (I_n - X(X^TX)^{-1}X^T)y && |\text{Taking y common from right} \\
&= (I_n - H)y && |H = X(X^TX)^{-1}X^T \\
&= (I_n - H)(X\beta + \epsilon) && |y = X\beta + \epsilon \\
&= (I_n - H)X\beta + (I_n - H)\epsilon \\
&= X\beta - HX\beta + (I_n - H)\epsilon && |\text{multiplicative identity of matrix; } XI_n\beta = X\beta \\
&= X\beta - X(X^TX)^{-1}X^TX\beta + (I_n - H)\epsilon && |H = X(X^TX)^{-1}X^T \\
&= X\beta - XI_n\beta + (I_n - H)\epsilon && |(X^TX)^{-1}X^TX = I_n; I_n \text{ is an identity matrix} \\
&= X\beta - X\beta + (I_n - H)\epsilon && |XI_n\beta = X\beta \\
&= 0 + (I_n - H)\epsilon \\
&= (I_n - H)\epsilon
\end{aligned}
$$

$\hat{\epsilon}$ is directly proportional to ϵ, which is dependent on the value of H.

Calculating the Expectation of $\hat{\epsilon}$

$$E[\hat{\epsilon}] = E[(I_n - H)\epsilon] \qquad |\text{E[aA] = aE[A]; a is constant and A is R.V.}$$

$$
\begin{aligned}
&= (I_n - H)E[\epsilon] && |(I_n - H) \text{ is not a RV, it is a constant matrix} \\
&= (I_n - H) \cdot 0 && |\epsilon \sim N(0, \sigma^2 I_n) \\
&= 0
\end{aligned}
$$

According to lecture 11, slide 8: The residuals are zero on average. This matches with the above result.

$$\frac{1}{n}\sum_{n=1}^{n} \hat{\epsilon_i} = 0$$

Our assumption on the error term ϵ is that E[ϵ]=0. Which is same as the expected value of residuals.

Calculating the co-variance of $\hat{\epsilon}$

$$cov[\hat{\epsilon}] = cov[(I_n - H)\epsilon]$$

Since $(I_n - H)$ is a constant matrix, we can use the below property of co-variance:

$$cov[Ax] = Acov[x]A^T \qquad |\text{where A is a constant matrix and x is a RV}$$

$$
\begin{aligned}
cov[\hat{\epsilon}] &= (I_n - H)cov[\epsilon](I_n - H)^T \\
&= (I_n - H)cov[\epsilon](I_n - H) && |\text{Since } (I_n - H) \text{ is symmetric} \\
&= (I_n - H)\sigma^2 I_n(I_n - H) && |cov[\epsilon] = \sigma^2 I_n, \epsilon \sim N(0, \sigma^2 I_n) \\
&= \sigma^2(I_n - H)(I_n - H) && |I_n A = A \text{ and } \sigma^2 \text{ is a constant} \\
&= \sigma^2(I_n - H) && |\text{Since } (I_n - H) \text{ is idempotent}
\end{aligned}
$$

Assumption on error term is $var[\epsilon_i] = \sigma^2$, a constant. Since error terms are i.i.d, $cov[\epsilon_i, \epsilon_j] = 0$. Thus $cov[\epsilon] = \sigma^2 I_n$, where σ^2 is a constant. But the $var[\hat{\epsilon}]$ is not just σ^2, it is dependent on X.

From above result we can see co variance of residual is dependent on projection matrix H. Thus $cov[\hat{\epsilon}_i, \hat{\epsilon_j}]$ is not zero, it is dependent on the values of X (H matrix depends on X). Thus $\hat{\epsilon_i}$ and $\hat{\epsilon_j}$ are correlated where $i \neq j$.

The residuals ($\hat{\epsilon_i}$) is orthogonal to the independent variables X_i and the predicted values $\hat{y_i}$. Thus the cov($\hat{\epsilon}$,X) and cov($\hat{\epsilon}, \hat{y}$) is 0.

$$
\begin{aligned}
\hat{\epsilon} &= (I_n - H)\epsilon \\
X^T\hat{\epsilon} &= X^T(I_n - H)\epsilon && |\text{Multiplying by } X^T \text{ on both sides} \\
&= (X^T I_n - X^T H)\epsilon \\
&= (X^T - X^T X(X^T X)^{-1}X^T)\epsilon && |H = X(X^T X)^{-1}X^T \text{ and } X^T I_n = X^T \\
&= (X^T - X^T)\epsilon && |X^T X(X^T X)^{-1} = I_n \text{ and } I_n X^T = X^T \\
&= 0
\end{aligned}
$$

Similarly, the dot product of H and $\hat{\epsilon}$ is also zero, which means H and $\hat{\epsilon}$ are orthogonal to each, which means $\hat{\epsilon}$ is parallel to $I_n - H$.