Solution to exercise 3.4

We have measurements $Y_1, \cdots, Y_{10\,000}$ satisfying

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{10\,000} \end{bmatrix} = \boldsymbol{x} \begin{bmatrix} \theta_{DK} \\ \theta_{EGer} \\ \theta_{BEN} \end{bmatrix} + \boldsymbol{\varepsilon} = \boldsymbol{x} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} + \boldsymbol{\varepsilon} = \boldsymbol{x\theta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{x}$ is a known $10\,000 \times 3$ matrix determined from meteorological data, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{10\,000})$ and $\boldsymbol{I}_{10\,000}$ is the 10 000 dimensional identity matrix.

For a particular measuring station we have the observation

$$Y \sim N(\mu, \sigma^2)\,,$$

and we assume that under some specific weather conditions

$$\mu = [0.4 \quad 0.6 \quad 0.8] \begin{bmatrix} \theta_{DK} \\ \theta_{EGer} \\ \theta_{BEN} \end{bmatrix} = \boldsymbol{z}^T \boldsymbol{\theta}\,.$$

In the sequel we shall determine a confidence interval for μ and a prediction interval for Y assuming the same weather conditions.

From the exercise text we have

$$\hat{\sigma}^2 = 18^2 = 324$$

and

$$\widehat{\boldsymbol{\theta}} = \begin{bmatrix} 48 \\ 70 \\ 192 \end{bmatrix}.$$

Since $D(\widehat{\boldsymbol{\theta}}) = \sigma^2 (\boldsymbol{x}^T \boldsymbol{x})^{-1}$, we get

$$\widehat{D}(\widehat{\boldsymbol{\theta}}) = \hat{\sigma}^2 (\boldsymbol{x}^T \boldsymbol{x})^{-1} = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} \\ \hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_3^2 \end{bmatrix}.$$

From the relation between covariances, correlations and standard deviations

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j} \;\Rightarrow\; \hat{\sigma}_{ij} = \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{ij}\,,$$

and using the estimated standard deviations and correlations

$$\begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \hat{\sigma}_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 7 \end{bmatrix} \quad \& \quad \boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.2 \\ 0 & -0.2 & 1 \end{bmatrix},$$

we get

$$\widehat{D}(\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} 64 & 0 & 0 \\ 0 & 64 & -11.2 \\ 0 & -11.2 & 49 \end{bmatrix} .$$

The estimate for μ becomes

$$\hat{\mu} = U = \boldsymbol{z}^T \widehat{\boldsymbol{\theta}} = [0.4 \quad 0.6 \quad 0.8] \begin{bmatrix} 48 \\ 70 \\ 192 \end{bmatrix} = 214.80 .$$

Furthermore

$$V(\hat{\mu}) = V(U) = \boldsymbol{z}^T D(\widehat{\boldsymbol{\theta}}) \boldsymbol{z} = \sigma^2 \boldsymbol{z}^T (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{z} = \sigma^2 c$$

using notation from section 2.1.5. This gives

$$\widehat{V}(\hat{\mu}) = \hat{\sigma}^2 c = 324c .$$

We also have the following expression for the estimated variance of $\hat{\mu} = U$:

$$\widehat{V}(\hat{\mu}) = \boldsymbol{z}^T \widehat{D}(\widehat{\boldsymbol{\theta}}) \boldsymbol{z}$$

$$= [0.4 \quad 0.6 \quad 0.8] \begin{bmatrix} 64 & 0 & 0 \\ 0 & 64 & -11.2 \\ 0 & -11.2 & 49 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.6 \\ 0.8 \end{bmatrix}$$

$$= 53.888$$

From the two expressions we get

$$c = \frac{53.888}{324} = 0.1663 .$$

Therefore

$$\hat{\sigma}\sqrt{c} = 7.3408 \ \& \ \hat{\sigma}\sqrt{c+1} = 19.4393$$

1. According to theorem 2.15, the confidence interval for μ becomes

$$\hat{\mu} \pm t(n-k)_{0.975} \hat{\sigma}\sqrt{c} = 214.80 \pm 1.96 \times 7.3408 = [200.41, 229.19] .$$

2. The prediction interval for Y becomes (using theorem 2.17)

$$\hat{\mu} \pm t(n-k)_{0.975} \hat{\sigma}\sqrt{c+1} = 214.80 \pm 1.96 \times 19.4393 = [176.70, 252.90] .$$

3. We notice that the prediction interval is considerably wider than the confidence interval.

The exercise was inspired by

L.P. Prahm, K. Conradsen, and L.B. Nielsen (1980): Regional Source Quantification model for sulphur oxides in Europe. *Atmospheric Environment*, Volume 14, Issue 9, Pages 1027-1054.