Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Statistik III/Advanced Statistics
SS 2026

Exercises

Exercise 9. A psychologist administers three subtests of an intelligence test. The first subtest covers verbal skills, while the other two subtests cover mathematical skills. The results of the three subtests are represented by a 3-vector $\mathbf{x}$. The psychologist is interested in the overall score, which is the sum of the three individual results, and a second value indicating whether the respondent's aptitude is more in the area of language or arithmetic. This value is calculated by subtracting the sum of the results in the two arithmetical subtests from the result in the verbal subtest. To avoid negative values, 20 is added to this difference. These two numbers are combined into a 2-vector $\mathbf{y}$.

a) Show that $\mathbf{y}$ can be obtained through an affine transformation $\mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$ by determining the matrix $\mathbf{A}$ and the vector $\mathbf{b}$.

b) Compute the $\mathbf{y}$ values for a subject with results

$$\mathbf{x}' = (7, 6, 2)$$

in the three subtests?

Exercise 10. Which of these matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -4 & 2 \\ 2 & 4 \end{pmatrix},$$

cannot be a covariance matrix?

Exercise 11. Let the (5×3)-matrix

$$\mathbf{X} = \begin{pmatrix} 2 & 1 & 8 \\ 3 & 0 & 5 \\ 3 & 1 & 4 \\ 4 & 1 & 1 \\ 3 & 2 & 2 \end{pmatrix}$$

represent the data of 5 subjects in 3 variables. Calculate by hand and using R

a) the mean vector $\bar{\mathbf{x}}$ and the covariance matrix $\mathbf{S}$;

b) the inverse $\mathbf{S}^{-1}$ of the covariance matrix;

c) the Mahalanobis distance between the data vector of subject 1 (column vector formed by 1st row of $\mathbf{X}$) and the mean vector $\bar{\mathbf{x}}$.

Exercise 12. The file `cheddar.txt` contains concentrations of various chemicals in 30 samples of mature cheddar cheese and a subjective measure of taste for each sample. The columns "Acetic", "H2S," and "Lactic" provide the concentration of acetic acid, hydrogen sulfide (both on a logarithmic scale), and lactic acid (Moore, D.S., & McCabe, G.P. (1989). Introduction to the Practice of Statistics. New York: W.H. Freeman.)

Read the data into R and define a data matrix $\mathbf{X}$ that stores the respective concentrations in three columns. Perform the following calculations in R (see also R code in "Basics of Linear Algebra").

a) Compute the mean vector $\bar{\mathbf{x}}$ and the covariance matrix $\mathbf{S}$.

b) Determine the diagonal matrix $\mathbf{\Lambda}$ of eigenvalues and the matrix $\mathbf{G}$ of eigenvectors for $\mathbf{S}$.

c) Show by calculation that $\mathbf{G}'\mathbf{S}\mathbf{G}$ equals $\mathbf{\Lambda}$.

d) Compute the total sample variance based on $\mathbf{S}$ or $\mathbf{\Lambda}$. What do you find? What is the reason for this observation?

e) Compute the generalized sample variance for $\mathbf{S}$ and compare the result with the product of the eigenvalues. What do you find? What is the reason for this observation?

Note: Read the data, for example, by

```
setwd("C:/...")  # provide path to data folder
dat <- read.table("cheddar.txt", header = TRUE)
```