Problem 1.1 (MATLAB)
Covariance matrix and eigenvalue decomposition

For a multivariate distribution is given the population covariance matrix

$$\Sigma = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix}$$

Perform an eigenvalue decomposition of Σ, that is find the eigenvalues and corresponding eigenvectors.

Verify the following properties of the eigenvalues

- $|\Sigma| = \prod_{i=1}^{3} \lambda_i$
- $\mathrm{trace}(\Sigma) = \sum_{i=1}^{3} \lambda_i$

Find the matrices $\Sigma^{1/2}$ and $\Sigma^{-1/2}$ by using

- the eigenvalue decomposition
- built-in square root commands

Problem 1.2 (MATLAB)
Multivariate visualization and descriptive statistics

The datafile '***dataset_problem_1_2.dat***' contains national track records for women in 54 countries for the running disciplines 100m [s], 200m [s], 400m [s], 800m [min], 1500m [min], 3000m [min], marathon [min].

- Calculate the descriptive statistics $\bar{x}, S, R$ and comment.
- Make plots of the multivariate scatter matrix plus marginal histograms and boxplots and comment.
- Calculate the generalized sample variances $|S|$ and $|R|$ and verify the expression $|S| = \left(\prod_{i=1}^{p} s_{ii}\right)|R|$
- Comment on magnitudes.

Convert the time-data to speed measured in meters per second.

For the calculated speed-data repeat the above points.

Problem 1.3 (Theoretical)
Properties of moments for multivariate distributions

Prove the following connection between the population covariance matrix, Σ, and the population correlation matrix, ρ (where V is the diagonal matrix consisting of marginal variances, see textbook)

- $V^{-1/2}\,\Sigma\,V^{-1/2} = \rho$
- $\Sigma = V^{1/2}\,\rho\,V^{1/2}$

Using the just proven expression, prove the relation between population generalized variances

- $|\Sigma| = \left(\prod_{i=1}^{p} \sigma_{ii}\right)|\rho|$

Using the eigenvalue decomposition, $\Sigma = P\,\Lambda\,P^T$, prove the property

- $|\Sigma| = \prod_{i=1}^{p} \lambda_i$

Consider a bivariate distribution with fixed marginal variances, σ_{11} and σ_{22}.

- Write the population covariance matrix, Σ, and the generalized variance, $|\Sigma|$ as function of the correlation coefficient, ρ.
- What is the minimum and maximum values of $|\Sigma|$ and for which values of ρ are they achieved (explain) ?

Consider the multivariate population covariance matrix

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Find the eigenvalues and eigenvectors of Σ
- Find Σ^{-1}
- Find the eigenvalues and eigenvectors of Σ^{-1}

Problem 1.4 (MATLAB + theoretical)
Kernel density pdf estimation

The datafile '***dataset_problem_1_4.mat***' contains a random sample of 100 observed x-values of an unknown, univariate probability distribution, $f(x)$.

Using Kernel Density Estimation with

- various kernel functions, e.g. gaussian and uniform
- various, different bandwidths

make KDE estimates, $\hat{f}(x)$.

The datafile also contains vectors xplot and f, that are corresponding x- and f-values of the true pdf $f(x)$.
Use these vectors to make a plot of f(x) for comparison with the estimates, $\hat{f}(x)$.

The Kernel density estimator for a true, unknown pdf, $f(x)$ is (see literature)

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)$$

Before the observations, x_i, are actually found/used, they are random variables and so is then the estimator, $\hat{f}(x)$ also.

Derive the expected value of the estimator, $E\left[\hat{f}(x)\right]$ and thereby show that it can be expressed as a convolution between the true pdf, $f(x)$ and the used kernel function, $K(x)$.

Show that the Mean Square Error of the estimator can be written as

$$MSE(x) \stackrel{\text{def}}{=} E\left[\left(\hat{f}(x) - f(x)\right)^2\right] = \text{ squared bias error + estimation variance}$$

where

- squared bias error = $\left\{E\left[\hat{f}(x)\right] - f(x)\right\}^2$
- estimation variance = $Var\left[\hat{f}(x)\right]$