Statistical Data Analysis
Exercise session 2

1. Generate $n = 100$ random data points from a bivariate normal distribution with mean $\boldsymbol{\mu} = (2,3)^t$, and variances $s_{11} = 5$ and $s_{22} = 2$. Vary the covariance s_{12} such that the correlation r_{12} between X_1 and X_2 is successively 0, 0.3, 0.6 and 0.9. Make a scatter plot of every situation.

   From now on, always use the data set with $r_{12} = 0.6$.

2. Show (in R) that this correlation corresponds to the covariance between the standardized data.

3. Illustrate that the mean and (empirical) covariance matrix are affine equivariant.

4. Illustrate that the Mahalanobis distances are affine invariant.

5. Given a $(n \times p)$ data matrix $\boldsymbol{X}$ with covariance matrix $\boldsymbol{S}$, we can consider the sphered data $\boldsymbol{X}\boldsymbol{S}^{-1/2}$.

   (a) Show (by hand) that the Mahalanobis distance between two observations is equal to the Euclidean distance between the corresponding sphered observations.

   (b) Consider the bivariate data set again. Make a scatter plot of the original data, the standardized data and the sphered data. On the first figure you add the *tolerance ellipse*:

   $$\{\boldsymbol{x}; (\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = c^2\}$$

   with $c = \sqrt{\chi^2_{p,0.05}}$. You also add the tolerance ellipse to the first figure based on the estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

   $$\{\boldsymbol{x}; (\boldsymbol{x} - \overline{\boldsymbol{x}})^t \boldsymbol{S}^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}}) = c^2\}$$

   On the last figure you add the corresponding *tolerance sphere*.

Useful functions:

1. `scale()`
2. `mahalanobis()`
3. `ellipse()` in package `ellipse`