D. Metzler, M. Hutzenthaler

MULTIVAR STATS IN ECOL AND GENETICS — EXERCISES, SHEET 4

---

1. A new cultivar of *Vitis vinifera* (common grape wine) is hoped for having better resistance against *Erysiphe necator* (powdery mildew). To measure this, both the number of affected leaves have benn counted and the grape yield per plant has been weighted (gram). You will find the data in the file `grapes.txt` on the web page.

- Is the number of affected leaves significantly different for the two groups? Use the t-test.
- Is the grape yield per plant significantly different for the two groups? Use the t-test.
- Apply the T^2-test for the null hypothesis that the bivariate data have the same mean in both groups.
- Discuss the original hypthesis that the new cultivar has better resistance.

2. To get some exercise with the multivariate normal distribution in `R`:

- Sample 100 two-dimensional normal distributed vectors with mean vector $(2, 0)^t$ and covariance matrix

$$\Sigma = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}$$

- Plot this two-dimensional data cloud
- Calculate the sample correlation between first and second coordinate.
- Calculate the true correlation of the coordinates of the above normal distribution.

3. Assume that the random Vector $X \in \mathbb{R}^d$ is multivariate normal distributed $\mathcal{N}_d(\mu, \Sigma)$ with mean vector $\mu = (1, 2)^t$ and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}. \tag{1}$$

Define the matrix

$$M = \begin{pmatrix} 2 & -1 \\ 1 & 4 \\ -2 & 1 \end{pmatrix}. \tag{2}$$

What is then the distribution of the random vector $M \cdot X$?

4. The density

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{(x-\mu)^t \Sigma^{-1}(x-\mu)}{2}\right)$$

for $x \in \mathbb{R}^d$ is not fully explicit (S^{-1} and $\det(\Sigma)$) and is in general laborious to calculate. In $d = 2$, there is a more explicit formula which we now derive. Recall that $|det(A)|$ is the volume of the set $\{A \cdot x \colon |x_i| \leq 1 \text{ for } i = 1..d\}$. In $d = 2$ we have that the explicit formula

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

for all $a, b, c, d \in \mathbb{R}$. Moreover the inverse matrix is (assuming $ad - bc \neq 0$)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

as can be easily checked. Let us write $\sigma_1^2 := \mathrm{Var}(X_1)$ and $\sigma_2^2 := \mathrm{Var}(X_2)$ for the variances and $\rho := \mathrm{Cor}(X_1, X_2) \in (-1, 1)$ for the correlation of the two coordinates (so $\mathrm{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$). Derive an explicit formula for $f(x)$ in terms of σ_1, σ_2, ρ in the case $d = 2$ and $\mu = (0, 0)^t$.