# Examination Multivariate Statistical Methods

Linköpings Universitet, IDA, Statistik

| | |
|---|---|
| Course code and name: | 732A97 Multivariate Statistical Methods |
| Date: | 2018/01/09, 8–12 |
| Examinator: | Krzysztof Bartoszek phone 013–281 885 |
| Allowed aids: | Pocket calculator |
| | Table with common formulae and moment generating functions (distributed with the exam) |
| | Table of integrals (distributed with the exam) |
| | Table with distributions from Appendix in the course book (distributed with the exam) |
| | One double sided A4 page with own hand written notes |
| Grades: | A= $[19 - \infty)$ points |
| | B= $[17 - 19)$ points |
| | C= $[14 - 17)$ points |
| | D= $[12 - 14)$ points |
| | E= $[10 - 12)$ points |
| | F= $[0 - 10)$ points |
| Instructions: | Write clear and concise answers to the questions. |

## Problem 1 (6p)

You are given the data matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \end{bmatrix}$$

(a 1p) Calculate, $\mathbf{S}$, the sample covariance matrix of $\mathbf{X}$.
(b 2p) Calculate the eigenvalues and eigenvectors of $\mathbf{S}$.
(c 1p) Calculate the total and generalized variances associated with $\mathbf{S}$.
(d 1p) What percantage of the variance do each of the principle components explain?
(e 1p) Provide an interpretation of the principal components.

## Problem 2 (4p)

Let $\mathbf{A}$ and $\mathbf{B}$ be two square matrices of the same dimensions. Assume that both are invertible, i.e. $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ exist.
(a 2p) Show that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
(b 2p) Assume furthermore that $\mathbf{A}$ is eigendecomposable. Represent $\mathbf{A}^{-1}$ in terms of $\mathbf{A}$'s eigendecomposition and prove that this representation is indeed correct.

## Problem 3 (4p)

Let the random vector $\mathbb{R}^4 \ni \vec{X} = (X_1, X_2, X_3, X_4)^T$ be normally distributed with expectation vector $\vec{\mu} = (3, 4, 0, 2)^T$ and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 & 1 \\ -2 & 8 & 3 & 0 \\ 0 & 3 & 10 & -3 \\ 1 & 0 & -3 & 10 \end{bmatrix}.$$

(a 2p) Write all pairs of variables that are independent. Justify your choice.
(b 2p) Is $3X_1 + X_2$ correlated with $2(X_3 + X_4)$? Justify your answer.

## Problem 4 (6p)

You are provided with the following distributional results.

- Let $\mathbb{R}^p \ni \vec{X} \sim \mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$, then

$$(\vec{X} - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{X} - \vec{\mu}) \sim \chi^2_p,$$

- Let $\mathbb{R}^p \ni \overline{x}$ be the sample mean of $n$ normal observations and $\mathbf{S}$ the sample covariance. If the population expectation is $\vec{\mu}$, then

$$(\overline{x} - \vec{\mu})^T \left(\frac{1}{n}\mathbf{S}\right)^{-1} (\overline{x} - \vec{\mu}) \sim \frac{(n-1)p}{n-p} F_{p,n-p},$$

- If we have two independent samples, both of dimension $p$, first of size $n_1$ from $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma}_1)$ and second of sizse $n_2$ from $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma}_2)$, then denoting by $\overline{x}_1$, $\mathbf{S}_1$ and $\overline{x}_2$, $\mathbf{S}_2$ the respective sample averages and covariances

  –
$$(\overline{x}_1 - \overline{x}_2)^T (\frac{1}{n_1}\mathbf{\Sigma}_1 + \frac{1}{n_2}\mathbf{\Sigma}_2)^{-1} (\overline{x}_1 - \overline{x}_2) \sim \chi^2_p,$$

  – if $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$
$$(\overline{x}_1 - \overline{x}_2)^T (\frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2)^{-1} (\overline{x}_1 - \overline{x}_2) \sim \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p,n_1+n_2-p-1},$$

– if $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$ and $n$ is large, then approximately

$$(\overline{x}_1 - \overline{x}_2)^T(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2)^{-1}(\overline{x}_1 - \overline{x}_2) \sim \chi^2_p.$$

In a medical study the content of two different substances in the blood are studied. The substances can be treated as a two–dimensional random vector $\vec{X} = (X_1, X_2)^T$ which can be assumed to be normally distributed. 50 people are randomly chosen to have these three variables measured. The result is: sample mean vector $\overline{x} = (2, 2.5)^T$ and sample covariance matrix

$$\mathbf{S} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

with inverse

$$\mathbf{S}^{-1} = \frac{1}{3}\begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix}$$

and eigendecomposition

$$\mathbf{S} = \begin{bmatrix} 0.29 & -0.957 \\ 0.957 & 0.29 \end{bmatrix}\begin{bmatrix} 4.303 & 0 \\ 0 & 0.697 \end{bmatrix}\begin{bmatrix} 0.29 & -0.957 \\ 0.957 & 0.29 \end{bmatrix}^{-1}.$$

(a 2p) Perform a test at the 5% significance level if the expectation vector of the substances is $\vec{\mu} = (2, 2)^T$.
(b 3p) Sketch a 95% confidence ellipse for the mean vector. Does $\vec{\mu} = (2, 2)^T$ lie in it? Mark it on the graph.