# FORMULAS FOR STK4040

This collection of formulas may be used at the written exam December 18th, 2009

## A - Vectors and matrices

A.1) For a $n$ x $k$ matrix **A** and a $k$ x $n$ matrix **B** we have $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

A.2) For nonsingular square matrices **A** and **B** we have $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$ and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

A.3) A $k$ x $k$ matrix **Q** is *orthogonal* if $\mathbf{QQ}' = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$, i.e. if $\mathbf{Q}^{-1} = \mathbf{Q}'$.

A.4) For $k$ x $k$ matrices **A** and **B** we have $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$.

A.5) For a diagonal matrix $\mathbf{A} = \text{diag}\{a_{11}, a_{22}, \ldots, a_{kk}\}$ we have $|\mathbf{A}| = \prod_{i=1}^{n} a_{ii}$.

A.6) Let **A** be a symmetric $k$ x $k$ matrix and **x** a $k$ dimensional vector.
Then $\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}x_i x_j$ is denoted a *quadratic form.*

A.7) A symmetric $k$ x $k$ matrix **A** is *nonnegative definite* if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $k$ dimensional vectors $\mathbf{x}$.
It is *positive definite* if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.

A.8) A $k$ x $k$ matrix **A** has *eigenvalue* $\lambda$ with corresponding *eigenvector* $\mathbf{e} \neq \mathbf{0}$ if $\mathbf{Ae} = \lambda\mathbf{e}$.

A.9) An eigenvalue $\lambda$ is a solution to the *characteristic equation* $|\mathbf{A} - \lambda\mathbf{I}| = 0$.

A.10) For a $k$ x $k$ matrix **A** the *trace* is given by $\text{tr}(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$.

A.11) $\text{tr}(\mathbf{A}) = \sum_{i=1}^{k} \lambda_i$ where the $\lambda_i$s are the eigenvalues of **A**.

A.12) For a $m$ x $k$ matrix **B** and a $k$ x $m$ matrix **C,** we have $\text{tr}(\mathbf{BC}) = \text{tr}(\mathbf{CB})$.

A.13) For a positive definite $k$ x $k$ matrix $\mathbf{A}$ we have the *spectral decomposition*

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_k \mathbf{e}_k \mathbf{e}_k'$$

Here $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ are the eigenvalues of $\mathbf{A}$ and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$ are the corresponding orthogonal and normalized eigenvectors (i.e. $\mathbf{e}_i'\mathbf{e}_i = 1$ and $\mathbf{e}_i'\mathbf{e}_j = 0$ for $i \neq j$)

A.14) For a positive definite $k$ x $k$ matrix $\mathbf{A}$, the square root matrix of $\mathbf{A}$ and its inverse are defined as

$$\mathbf{A}^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i}\, \mathbf{e}_i \mathbf{e}_i' \text{ and } \mathbf{A}^{-1/2} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} \mathbf{e}_i \mathbf{e}_i' .$$

A.15) Let $\mathbf{B}$ be a positive definite matrix and $\mathbf{d}$ be a given vector. Then for an arbitrary nonzero vector $\mathbf{x}$, we have $\max_{\mathbf{x} \neq \mathbf{0}} \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d}$ with the maximum attained when $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$ for any constant $c \neq 0$.

A.16) Let $\mathbf{B}$ be a positive definite *p*x*p* matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ and corresponding orthogonal and normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Then

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1 \qquad \text{(attained when } \mathbf{x} = \mathbf{e}_1\text{)}$$

Moreover for $j = 1, 2, \ldots, p-1$

$$\max_{\mathbf{x} \perp \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_j} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_{j+1} \qquad \text{(attained when } \mathbf{x} = \mathbf{e}_{j+1}\text{)}$$

A.17) Given a positive definite *p*x*p* matrix $\mathbf{B}$ and a scalar $b > 0$, we have

$$\frac{1}{|\mathbf{\Sigma}|^b} \exp\left(-\frac{1}{2} \operatorname{tr}\left[\mathbf{\Sigma}^{-1}\mathbf{B}\right]\right) \leq \frac{1}{|\mathbf{B}|^b} (2b)^{pb} e^{-bp}$$

for any positive definite *p x p* matrix $\mathbf{\Sigma}$, with equality if and only if $\mathbf{\Sigma} = (1/2b)\mathbf{B}$.

## B - Random vectors and matrices

B.1) For random matrices $\mathbf{X}$ and $\mathbf{Y}$ of the same dimension, and $\mathbf{A}$ and $\mathbf{B}$ matrices of constants, we have that

$$E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$$
$$E(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A}\,E(\mathbf{X})\,\mathbf{B}$$

B.2) For a random vector $\mathbf{X}$ with mean vector $\boldsymbol{\mu}$, the *covariance matrix* is given by

$$\operatorname{Cov}(\mathbf{X}) = E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\}$$

B.3) For a random vector $\mathbf{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ and a matrix $\mathbf{C}$ of constants, we have that $E(\mathbf{C}\mathbf{X}) = \mathbf{C}E(\mathbf{X}) = \mathbf{C}\boldsymbol{\mu}$ and $\operatorname{Cov}(\mathbf{C}\mathbf{X}) = \mathbf{C}\mathbf{\Sigma}\mathbf{C}'$.

B.4) Let $\boldsymbol{\rho}$ denote the correlation matrix. Then

$$\mathbf{V}^{1/2}\boldsymbol{\rho}\mathbf{V}^{1/2} = \boldsymbol{\Sigma} \qquad \text{and} \qquad \boldsymbol{\rho} = \mathbf{V}^{-1/2}\,\boldsymbol{\Sigma}\,\mathbf{V}^{-1/2},$$

where $\mathbf{V}^{1/2}=\text{diag}\left(\sqrt{\sigma_{11}},\sqrt{\sigma_{22}},\ldots,\sqrt{\sigma_{kk}}\right)$ is the standard deviation matrix.

## C - The multivariate normal distribution and related distributions

C.1) A $p$-variate random vector $\mathbf{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is multivariate normally distributed, $\mathbf{X} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, if its density takes the form

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}}\,\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

C.2) $\mathbf{X} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ if and only if $\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu},\ \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for all $p$-dimensional vectors **a.**

C.3) Let $\mathbf{X} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, then the probability is $1-\alpha$ that $\mathbf{X}$ takes values in the ellipsoid $\left\{\mathbf{x}\colon (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \le \chi_p^2(\alpha)\right\}$

C.4) Let $\mathbf{X} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, and let $\mathbf{A}$ be a $q$x$p$ matrix and $\mathbf{d}$ a $q$-dimensional vector. Then $\mathbf{AX}+\mathbf{d} \sim N_q(\mathbf{A}\boldsymbol{\mu}+\mathbf{d},\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

C.5) Assume that $\begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{bmatrix} \sim N_p\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$, where $\mathbf{X}_1$ is a $q$ dimensional vector.

Then $\mathbf{X}_1|\mathbf{X}_2=\mathbf{x}_2 \sim N_q(\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$

C.6) Assume that $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n$ are independent, $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}_j,\boldsymbol{\Sigma})$, and let $\mathbf{V}_1 = c_1\mathbf{X}_1+c_2\mathbf{X}_2+\cdots+c_n\mathbf{X}_n$ and $\mathbf{V}_2 = b_1\mathbf{X}_1+b_2\mathbf{X}_2+\cdots+b_n\mathbf{X}_n$, where the $c_j$'s and $b_j$'s are constants. Then

$$\begin{bmatrix}\mathbf{V}_1\\ \mathbf{V}_2\end{bmatrix} \sim N_p\left(\begin{bmatrix}\sum_{j=1}^n c_j\boldsymbol{\mu}_j\\ \sum_{j=1}^n b_j\boldsymbol{\mu}_j\end{bmatrix},\begin{bmatrix}\mathbf{c}'\mathbf{c}\,\boldsymbol{\Sigma} & \mathbf{b}'\mathbf{c}\,\boldsymbol{\Sigma}\\ \mathbf{b}'\mathbf{c}\,\boldsymbol{\Sigma} & \mathbf{b}'\mathbf{b}\,\boldsymbol{\Sigma}\end{bmatrix}\right)$$

C.7) Assume that $\mathbf{Z}_1,\ldots,\mathbf{Z}_m$ are i.i.d. $N_p(\mathbf{0},\boldsymbol{\Sigma})$. The distribution of $\sum_{j=1}^m \mathbf{Z}_j\mathbf{Z}_j'$ is called the *Wishart distribution* with $m$ degrees of freedom (and $p$x$p$ covariance matrix $\boldsymbol{\Sigma}$), denoted $W_{p,m}(\boldsymbol{\Sigma})$.

C.8) Properties of the Wishart distribution:

- If $\mathbf{W}_1 \sim W_{p,m_1}(\boldsymbol{\Sigma})$ and $\mathbf{W}_2 \sim W_{p,m_2}(\boldsymbol{\Sigma})$ are independent, then $\mathbf{W}_1+\mathbf{W}_2 \sim W_{p,m_1+m_2}(\boldsymbol{\Sigma})$
- If $\mathbf{W} \sim W_{p,m}(\boldsymbol{\Sigma})$ and $\mathbf{C}$ is a $q$x$p$ matrix, then $\mathbf{CWC}' \sim W_{q,m}(\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$

C.9) Let $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{W} \sim W_{p,m}(\mathbf{\Sigma})$ be independent.

Then $\mathbf{Z}' \left( \frac{\mathbf{W}}{m} \right)^{-1} \mathbf{Z}$ is distributed as $\frac{m \cdot p}{m - p + 1} F_{p,m-p+1}$.

## D – Estimation for multivariate distributions

D.1) Assume that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are i.i.d. with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$.

Unbiased estimators for $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are $\bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{X}_j$ and $\mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} \left( \mathbf{X}_j - \bar{\mathbf{X}} \right) \left( \mathbf{X}_j - \bar{\mathbf{X}} \right)'$.

D.2) Assume that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are i.i.d. $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Then we have the following results

- $\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent
- $\bar{\mathbf{X}} \sim N_p \left( \boldsymbol{\mu}, \frac{1}{n} \mathbf{\Sigma} \right)$
- $(n-1)\mathbf{S} \sim W_{p,n-1}(\mathbf{\Sigma})$

## E – Principal components

E.1) Let $\mathbf{X}$ be a *p*-variate random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$.

- The first population principal component is the linear combination $\mathbf{a}_1'\mathbf{X}$ that maximizes $\mathrm{Var}(\mathbf{a}_1'\mathbf{X})$ subject to $\mathbf{a}_1'\mathbf{a}_1 = 1$.
- The second population principal component is the linear combination $\mathbf{a}_2'\mathbf{X}$ that maximizes $\mathrm{Var}(\mathbf{a}_2'\mathbf{X})$ subject to $\mathbf{a}_2'\mathbf{a}_2 = 1$ and $\mathrm{Cov}(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0$.
- Etc

E.2) Let $\mathbf{X}$ a *p*-variate random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, and assume that $\mathbf{\Sigma}$ has eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$.
Then we have the following results:

- The *i*-th population principal component is $Y_i = \mathbf{e}_i'\mathbf{X}$
- $\mathrm{Var}(Y_i) = \lambda_i$
- $\sum_{i=1}^{p} \mathrm{Var}(X_i) = \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$
- $\mathrm{corr}(Y_i, X_k) = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$

## F – Factor analysis

F.1) Let $\mathbf{X}$ a $p$-variate random vector with mean vector $\boldsymbol{\mu}$. The orthogonal factor model assumes that

$$\mathbf{X}-\boldsymbol{\mu}=\mathbf{L}\mathbf{F}+\boldsymbol{\varepsilon}$$

where $\mathbf{L}=\{l_{ij}\}$ is a $p$ x $m$ matrix of *factor loadings*, $\mathbf{F}=\left[F_1,F_2,\ldots,F_m\right]'$ is a $m$-dimisional vector of *common factors*, and $\boldsymbol{\varepsilon}=\left[\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_p\right]'$ is a $p$-vector of errors (*specific factors*).

F.2) The unobservable random vectors $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ in the factor model satisfy:

- $E(\mathbf{F})=\mathbf{0}$
- $\text{Cov}(\mathbf{F})=\mathbf{I}$
- $E(\boldsymbol{\varepsilon})=\mathbf{0}$
- $\text{Cov}(\boldsymbol{\varepsilon})=\boldsymbol{\Psi}=\text{diag}\{\psi_1,\psi_2,\ldots,\psi_p\}$
- $\text{Cov}(\boldsymbol{\varepsilon},\mathbf{F})=\mathbf{0}$

## G – Discrimination and classification

G.1) We have two populations, denoted $\pi_1$ and $\pi_2$, with prior probabilities $p_1$ and $p_2$. If a random vector $\mathbf{X}$ is selected from $\pi_1$ it has density $f_1(\mathbf{x})$, while it has density $f_2(\mathbf{x})$ if it is selected from $\pi_2$. The cost of misclassifying an observation from $\pi_2$ as coming from $\pi_1$ is $c(1|2)$, while the cost is $c(2|1)$ for misclassifying an observation from $\pi_1$ as coming from $\pi_2$. Then the expected cost of misclassification is minimized if an observation $\mathbf{x}$ is allocated to $\pi_1$ provided that

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\geq\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}$$

and allocated to $\pi_2$ otherwise.

G.2) If the densities $f_i(\mathbf{x}),\ i=1,2,$ are multivariate normal with mean vectors $\boldsymbol{\mu}_i$ and common covariance matrix $\boldsymbol{\Sigma}$, then the expected cost of misclassification is minimized if an observation $\mathbf{x}$ is allocated to $\pi_1$ provided that

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\mathbf{x}_0-\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\geq\ln\left[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\right]$$

and allocated to $\pi_2$ otherwise.