STOCKHOLM UNIVERSITY
Department of Statistics
Johan Koskinen

EXAM IN MULTIVARIATE METHODS
31 October 2022

Time: 5 hours

Allowed aids: Pocket calculator, language dictionary

The exam consists of five questions. To score maximum points on a question solutions need to be clear, detailed and well motivated.

Results will be announced no later than November 18

Question 1. (16 Points)
Define and describe the following:

(a) Complete linkage

(b) Mahalanobis distance

(c) Data reduction

(d) Confirmatory Factor Analysis

Question 2. (16 Points) In the European Social Survey round 7, the following question is asked "Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust" for the following five institutions

1. ... national parliament (PRL)
2. ... politicians (PLT)
3. ... political parties (PRT)
4. ... European Parliament (EP)
5. ... United Nations (UN)

In round 7 (2014) the responses from $n = 1944$ respondents in Great Britain yield the following means

Sample means of trust in				
PRL	PLT	PRT	EP	UN
4.339	3.489	3.537	3.122	4.879

The following correlation matrix is obtained

$$\mathbf{R} = \begin{bmatrix} 1.000 & 0.757 & 0.718 & 0.511 & 0.521 \\ 0.757 & 1.000 & 0.838 & 0.606 & 0.502 \\ 0.718 & 0.838 & 1.000 & 0.628 & 0.528 \\ 0.511 & 0.606 & 0.628 & 1.000 & 0.568 \\ 0.521 & 0.502 & 0.528 & 0.568 & 1.000 \end{bmatrix}.$$

The eigenvalues for $\mathbf{R}$ are

$$\boldsymbol{\lambda} = \begin{bmatrix} 3.489 & v & 0.449 & 0.269 & 0.157 \end{bmatrix},$$

where λ_2 is hidden and represented by v, and the standardised eigenvectors are the columns of

$$\mathbf{W} = \begin{bmatrix} -0.454 & 0.340 & -0.397 & 0.704 & 0.157 \\ -0.482 & 0.356 & 0.058 & -0.282 & -0.747 \\ -0.482 & 0.269 & 0.118 & -0.517 & 0.643 \\ -0.421 & -0.418 & 0.729 & 0.341 & 0.000 \\ u & -0.715 & -0.542 & -0.200 & -0.055 \end{bmatrix}.$$

where the fifth value of the first eigenvector is hidden and represented by u.

(a) Draw the scree plot.

(b) For an observation with raw (i.e. not mean-corrected) values

$$\mathbf{x} = \begin{bmatrix} 3 & 3 & 3 & 2 & 4 \end{bmatrix}$$

the first principal component component score is

$$\xi_1 = 1.91761.$$

Solve for u (do not forget to take centering into account)

(c) Which variable is most influential on the first principal component?

(d) Which principal component has the highest (in absolute value) correlation with trust in the UN?

Question 3. (16 Points) For the $p = 5$ variables in Question 2, the following $M = 2$ factor model is assumed:

$$\mathbf{X} = \boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\epsilon}.$$

The following assumptions are made: $E(\mathbf{F}) = \mathbf{0}$, $E(\mathbf{F}\mathbf{F}^\top) = \mathbf{I}$, $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $E(\boldsymbol{\epsilon}\mathbf{F}^\top) = \mathbf{0}$, and $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix with ψ_i as the i'th diagonal value.

(i) An initial 2-factor model is obtained by spectral composition, i.e. an $p \times 2$ matrix $\tilde{\mathbf{\Lambda}}$ whose first column is the first column $\mathbf{w}_1$ of $\mathbf{W}$ multiplied by $\sqrt{\lambda_1}$, and second column $\mathbf{w}_2$ of $\mathbf{W}$ multiplied by $\sqrt{\lambda_2}$
(ii) Using a number of iterations, the unrotated solution is

$$\mathbf{\Lambda} = \begin{bmatrix} 0.786 & -0.126 \\ 0.899 & -0.274 \\ 0.883 & -0.203 \\ 0.696 & 0.065 \\ 0.753 & 0.638 \end{bmatrix}.$$

(a) For the initial solution $\tilde{\mathbf{\Lambda}}$, what is your estimate for ψ_1?

(b) What is the residual for PRL based on the second solution $\mathbf{\Lambda}$ in (ii)?

(c) What are the structure loadings of X_5, UN, given the assumptions and the solution (ii)?

(d) What is the correlation between X_4, EP, and X_5, UN?

Question 4. (16 Points) For the data of Question 2, there is an additional variable Y_i that is equal to 1 if respondent i wants the UK to leave the EU and 0 if respondent i wants the UK to remain in the EU. The means of the variables split by the two categories is

	Sample means of trust in				
y_i	PRL	PLT	PRT	EP	UN
0	4.11	3.52	3.57	3.46	5.03
1	4.55	3.46	3.51	2.81	4.74

and the inverse of the scaled combined sum of squares matrices is (rounded)

$$\mathbf{S}_{\text{pooled}}^{-1} = \begin{bmatrix} 0.39 & -0.21 & -0.10 & 0.00 & -0.07 \\ -0.21 & 0.76 & -0.44 & -0.07 & 0.01 \\ -0.10 & -0.44 & 0.76 & -0.11 & -0.04 \\ 0.00 & -0.07 & -0.11 & 0.33 & -0.10 \\ -0.07 & 0.01 & -0.04 & -0.10 & 0.26 \end{bmatrix}.$$

Fisher's linear discriminant function has been (partially) calculated as

$$\gamma = \begin{bmatrix} -0.2059 \\ 0.0668 \\ -0.0205 \\ v \\ 0.0361 \end{bmatrix}$$

(a) Calculate the missing entry v in Fisher's linear discriminant function

(b) For the observation $\mathbf{x}$ with values as in Question 2(b), determine if the respondent is to be classified as a Brexit supporter or a remainer. Assume equal costs and equal prior proportions.

(c) In the sample, the proportion $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 0.52$. Taking this proportion into account, what is your classification of $\mathbf{x}$ from Q2(b)?

(d) The respondent with $\mathbf{x}$ as in Question 2(b) answered that they wanted the UK to leave the EU - given your classification, what is this an example of?

Question 5. (16 Points) A logistic regression is estimated for the dataset in Question 4. The following estimates were obtained

	Estimates	
	Coefficients	Std. Error
Intercept	0.110	0.107
PRL	0.210	0.030
PLT	-0.066	0.042
PRT	0.013	0.041
EP	-0.177	0.027
UN	-0.038	0.024

(a) Assuming that the coefficients are normally distributed, what variables have a coefficient that is statistically significantly different from 0 on the 5%-level? Interpret.

(b) What is the predicted probability for the case with values of $\mathbf{x}$ as in Question 2(b)?.

(c) Another respondent have the answer vector

$$\mathbf{x} = \begin{bmatrix} 4 & 4 & 4 & 4 & 7 \end{bmatrix},$$

and $y_i = 0$. Is there a value of trust in the European Parliament at which this respondent would be classified as belonging to the group that favoured that the UK leave the EU, all other variables being the same?