STOCKHOLM UNIVERSITY
Department of Statistics
Ellinor Fackle-Fornius

EXAM IN MULTIVARIATE METHODS
February 18 2019

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 March 4.

Question 1. (16 points)

Define and explain the following concepts:

a) Underidentification
b) Centroid method
c) Deviance
d) False positive rate

Question 2. (16 points)

The weekly rates of return for five stocks (JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell and ExxonMobil) listed on the New York Stock Exchange were determined for the period January 2004 through December 2005. JP Morgan, Citibank and Wells Fargo are bank stocks while Royal Dutch Shell and ExxonMobil are oil stocks. Let $x_1, x_2, ..., x_5$ denote observed weekly rates of return for JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell, and ExxonMobil, respectively. A principal components analysis was performed on the correlation matrix resulting in the following eigenvalues and eigenvectors.

$$\lambda_1 = 2.437, \quad w_1^T = \begin{pmatrix}0.469 & 0.532 & 0.465 & 0.387 & 0.361\end{pmatrix}$$
$$\lambda_2 = 1.407, \quad w_2^T = \begin{pmatrix}-0.368 & -0.236 & -0.315 & 0.585 & 0.606\end{pmatrix}$$
$$\lambda_3 = 0.501, \quad w_3^T = \begin{pmatrix}-0.604 & -0.136 & 0.772 & 0.093 & -0.109\end{pmatrix}$$
$$\lambda_4 = 0.400, \quad w_4^T = \begin{pmatrix}0.363 & -0.629 & 0.289 & -0.381 & 0.493\end{pmatrix}$$
$$\lambda_5 = 0.255, \quad w_5^T = \begin{pmatrix}0.384 & -0.496 & 0.071 & 0.595 & -0.498\end{pmatrix}$$

a) Describe the general objective of principal components analysis.

b) Draw a scree plot and comment the plot.

c) How many PCs are needed to account for at least 80 % of the total variance?

d) Compute the loadings of the variables on the first two PCs. Interpret the first two PCs.

Question 3. (16 points)

Consider again the weekly rates of return stock-data described in Question 2. An exploratory factor analysis was performed on the correlation matrix and 2 factors were extracted. The factor loadings produced after an orthogonal rotation are given in the following table.

Variable	F_1	F_2
x_1	0.763	0.024
x_2	0.821	0.227
x_3	0.669	0.104
x_4	0.118	0.993
x_5	0.113	0.675

a) Describe two types of orthogonal rotations.

b) What are the usual assumptions for the factor model?

c) Compute shared variances between the indicators and each of the common factors.

d) What percentage of the variance of each of the oil stocks is not accounted for by the common factors F_1 and F_2? What is this percentage called?

Question 4. (16 points)

Based on a data set of city crime rates the matrix of distances between 6 cities was found to be

$$\mathbf{D} = \begin{array}{l} \\ \text{Atlanta} \\ \text{Boston} \\ \text{Chicago} \\ \text{Dallas} \\ \text{Denver} \\ \text{Detroit} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Atlanta} & \text{Boston} & \text{Chicago} & \text{Dallas} & \text{Denver} & \text{Detroit} \end{array} \\ \begin{pmatrix} 0 & & & & & \\ 536 & 0 & & & & \\ 516 & 447 & 0 & & & \\ 590 & 833 & 924 & 0 & & \\ 694 & 915 & 1073 & 528 & 0 & \\ 716 & 881 & 972 & 465 & 359 & 0 \end{pmatrix} \end{array}$$

a) Use the Complete-Linkage method to perform a hierarchical clustering of the cities.

b) Draw the dendogram for the hierarchical clustering in a).

c) Name one advantage and one disadvantage of hierarchical clustering methods versus non-hierarchical clustering methods.

Question 5. (16 points)

Samples of steel produced at two different rolling temperatures have been measured on two variables: x_1 = yield point and x_2 = ultimate strength. The data for the two groups of temperatures are presented in the following table and plotted in Figure 1.

Temperature 1		Temperature 2	
x_1	x_2	x_1	x_2
33	60	35	57
36	61	36	59
35	64	38	59
38	63	39	61
40	65	41	63
		43	65
		41	59

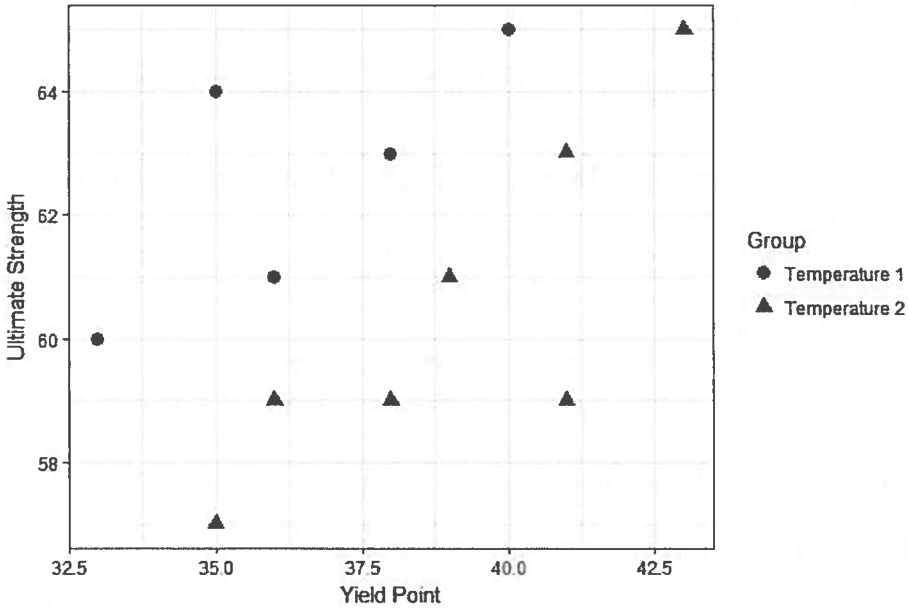


Figure 1: Ultimate strength and yield point for steel rolled at two temperatures

From the data, we calculate

$$\bar{\mathbf{x}}_1 = \begin{pmatrix} 36.4 \\ 62.6 \end{pmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{pmatrix} 39.0 \\ 60.4 \end{pmatrix}, \quad \mathbf{SSCP}_1 = \begin{pmatrix} 29.2 & 16.8 \\ 16.8 & 17.2 \end{pmatrix}, \quad \mathbf{SSCP}_2 = \begin{pmatrix} 50 & 40 \\ 40 & 45.7 \end{pmatrix}.$$

a) Comment on the discrimination provided by the two variables individually as well as jointly.

b) Calculate the pooled covariance matrix $\mathbf{S}_{\text{pooled}}$.

c) Calculate Fisher's linear discriminant function for these data.

Formula Sheet for the Exam in Multivariate Methods

Vectors and matrices

- Length of a vector $\mathbf{a} = (a_1, a_2, \ldots, a_p)$

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_p^2}$$

- Determinant of a 2×2 matrix $\mathbf{A}$

$$\det(\mathbf{A}) = |\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$$

- Inverse of a 2×2 matrix $\mathbf{A}$

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

- Eigenvalues are the roots of the characteristic equation

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{0}$$

  For each eigenvalue the solution to

$$(\mathbf{A} - \lambda \mathbf{I})\,\mathbf{x} = \mathbf{0}$$

  gives the associated eigenvector $\mathbf{x}$

Distances

- Euclidean

$$D_{ik} = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{kj})^2}$$

- Statistical

$$SD_{ik} = \sqrt{\sum_{j=1}^{p} \left(\frac{x_{ij} - x_{kj}}{s_j}\right)^2}$$

- Mahalanobis

$$MD_{ik} = \sqrt{(\mathbf{x}_i - \mathbf{x}_k)^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_k)}$$

For $p = 2$

$$MD_{ik} = \sqrt{\frac{1}{1-r^2}\left[\frac{(x_{i1} - x_{k1})^2}{s_1^2} + \frac{(x_{i2} - x_{k2})^2}{s_2^2} - \frac{2r\,(x_{i1} - x_{k1})\,(x_{i2} - x_{k2})}{s_1 s_2}\right]}$$

Mean-correction and covariance

- Mean-corrected data

$$\underset{(n\times p)}{\mathbf{X}_m} = \{x_{ij}\} = \{X_{ij} - \bar{X}_j\}$$

- Covariance

$$\underset{(p\times p)}{\mathbf{S}} = \{s_{ij}\} = \left\{\frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{n-1}\right\} = \frac{\mathbf{SSCP}}{df} = \frac{1}{n-1}\mathbf{X}_m^T\mathbf{X}_m$$

Group Analysis

- Total sum of squares and cross products

$$\mathbf{SSCP}_{\text{total}} = \mathbf{SSCP}_{\text{within}} + \mathbf{SSCP}_{\text{between}}$$

- Pooled within-group sum of squares and cross products

$$\mathbf{SSCP}_{\text{within}} = \sum_{\ell=1}^{g} \mathbf{SSCP}_{\ell}$$

- Pooled covariance matrix

$$\mathbf{S}_{\text{pooled}} = \frac{\mathbf{SSCP}_{\text{within}}}{n-g}$$

- Between-group sum of squares and cross products

$$\mathbf{SSCP}_{\text{between}} = \mathbf{SSCP}_{\text{total}} - \mathbf{SSCP}_{\text{within}}$$

For $g = 2$ groups

$$\mathbf{SSCP}_{\text{between}} = \frac{n_1 n_2}{n_1 + n_2} \left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right) \left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)^T$$

Factor Analysis

- For the two-factor model

$$Var(x) = \lambda_1^2 + \lambda_2^2 + Var(\epsilon) + 2\lambda_1\lambda_2\phi$$

$$Cor(x, \xi_1) = \lambda_1 + \lambda_2\phi$$

$$Cor(x_j, x_k) = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} + \left(\lambda_{j1}\lambda_{k2} + \lambda_{j2}\lambda_{k1}\right)\phi$$

- RMSR for EFA

$$RMSR = \sqrt{\frac{\sum_{i=1}^{p} \sum_{j=i+1}^{p} res_{ij}^2}{p(p-1)/2}}$$

- RMSR for CFA

$$RMSR = \sqrt{\frac{\sum_{i=1}^{p} \sum_{j=i}^{p} (s_{ij} - \hat{\sigma}_{ij})^2}{p(p+1)/2}}$$

Two-Group Discriminant Analysis

- Maximize

$$\lambda = \frac{\gamma^T \mathbf{B} \gamma}{\gamma^T \mathbf{W} \gamma}$$

- Fisher's linear discriminant function

$$\gamma^T = (\mu_1 - \mu_2)^T \boldsymbol{\Sigma}^{-1}$$

- Wilks' Λ

$$\Lambda = \frac{|\mathbf{SSCP}_w|}{|\mathbf{SSCP}_t|}$$

$$F = \left(\frac{1-\Lambda}{\Lambda}\right)\left(\frac{n_1 + n_2 - p - 1}{p}\right) \sim F\left(p, n_1 + n_2 - p - 1\right)$$

- Classification based on decision theory: assign the observation to group 1 if

$$Z \geq \frac{\bar{Z}_1 + \bar{Z}_2}{2} + \ln\left[\frac{p_2 C(1|2)}{p_1 C(2|1)}\right]$$

Logistic regression

- Odds of the event $Y = 1$

$$odds = \frac{p}{1-p}$$

  where

$$p = P(Y = 1)$$

- Probability of the event $Y = 1$ as a function of the explanatory variables

$$P(Y = 1|X_1, X_2, \ldots, X_k) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)}}$$

Quadratic equation

- The roots of the quadratic equation $ax^2 + bx + c$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Department of Statistics

Stockholms universitet

Correction sheet

Date: 18/02/2019

Room: Laduvikssalen

Exam: Multivariate methods (eng)

Course: Multivariate methods (eng)

Anonymous code: 0014-LFF

☑ I authorise the anonymous posting of my exam, in whole or in part, on the department homepage as a sample student answer.

NOTE! ALSO WRITE ON THE BACK OF THE ANSWER SHEET

Mark answered questions

1	2	3	4	5	6	7	8	9	Total number of pages
X	X	X	X	X					8
Teacher's notes 14	15	14	14	14					

GC

Points	Grade	Teacher's sign.
71		

SU, DEPARTMENT OF STATISTICS

Room: LA Anonymous code: 0014-LFF Sheet number: 1

Question 1

a) Underidentification

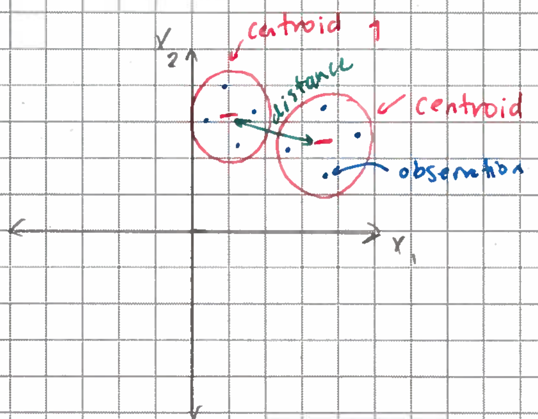
When a model is underidentified, parameters cannot be estimated and the fit of the model cannot be assessed, due to a lack of information. This problem can arise in confirmatory factor analysis, when

$$\underbrace{\frac{p(p+1)}{2}}_{\text{number of elements in covariance matrix}} < \underbrace{q}_{\text{number of parameters to estimate}}$$

4

b) Centroid method

The centroid method is a hierarchical clustering method for cluster analysis, where the distance between clusters is calculated from a centroid (as the average of the observations in the cluster).



4

c) Deviance

Deviance is a goodness-of-fit statistic that can be used to test if a hypothesized model fits the data.

$$\text{Deviance} = -2\log\left[\frac{L\,(\text{fitted model})}{L\,(\text{Saturated model})}\right]$$

maximum model where # parameters = # observations

R

The statistic is approximately χ^2-distributed with n- eg e of freedom

R

The difference in deviance between two models can e used to compare models. The difference will also be approximately χ^2-distributed & the degrees of freedom will be equal to the difference in degrees of freedom for the two models

R

4

d) False positive rate

The false positive rate is a measure of classification e.) a discriminant analysis or logistic regression. A low false positive rate is desirable, since this means that the share of pred ed events that are truly non-events is low. high !

R

$$\text{False positive rate} = \frac{\text{Number of false positives}}{\text{Number of true positives} + \text{nr. of false positives}}$$

$$\underbrace{\text{\# of true neg. / non-events}}_{\text{\# of non-events}}$$

2

14

SU, DEPARTMENT OF STATISTICS

Room: LA Anonymous code: 0014-LFF Sheet number: 2

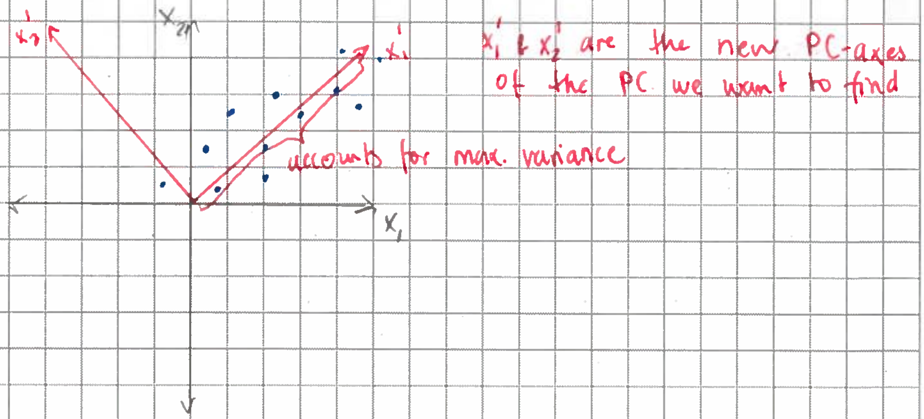
Question 2

$p = 5$

PCA on correlation matrix $\Rightarrow$ standardized data

Eigenvalues	Eigenvectors				
$\lambda_1 = 2.437$	$w_1^T = ($0.469	0.532	0.465	0.387	0.361$)$
$\lambda_2 = 1.407$	$w_2^T = ($-0.368	-0.236	-0.315	0.585	0.606$)$
$\lambda_3 = 0.501$	$w_3^T = ($-0.604	-0.136	0.772	0.093	-0.109$)$
$\lambda_4 = 0.400$	$w_4^T = ($0.363	-0.629	0.289	-0.381	0.493$)$
$\lambda_5 = 0.255$	$w_5^T = ($0.384	-0.496	0.071	0.595	-0.498$)$

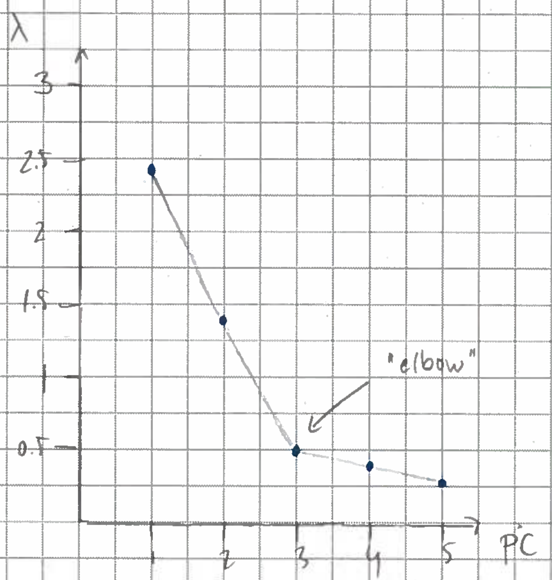
a) The general objective of Principal Components analysis is to find principal components that account for as much variation in the data as possible.



3

b) A scree plot plots the Eigenvalues (i.e. variances) of the PCs against the PCs

Scree plot:



R

The characteristic "elbow" of the scree plot is at PC_3

4

c) Looking for: PCs to account for 80% of total variance

The sum of variance of the original variables equals the sum of variance of the principal components (i.e. the eigenvalues):

$$\boxed{\sum s_i^2 = \sum \lambda_i}$$

s_i^2: jth variable; λ_i: ith PC

share of total variance for each PC will thus be: $\boxed{\frac{\lambda_i}{\sum \lambda_i}}$

R

Total variance $(\sum \lambda_i) = 2.437 + 1.407 + 0.501 + 0.400 + 0.255 = 5$

PC	Share of tot variance
1	$2.437/5 = 0.4874 = 48.74\%$
2	$1.407/5 = 0.2814 = 28.14\%$
3	$0.501/5 = 0.1002 = 10.02\%$
4	$0.400/5 = 0.08 = 8\%$
5	$0.255/5 = 0.051 = 5.1\%$

Jointly account for 86.9% of total variance

Answer: 3 PCs are needed to account for at least 80% of tot. variance

R 3

SU, DEPARTMENT OF STATISTICS

Room: LA   Anonymous code: 0014-LFF   Sheet number: 3

Cont. question 2

d) Looking for loadings of the variables of the first two PCs

Loadings: $l_{ij} = \frac{w_{ij} \sqrt{\lambda_i}}{s_j}$ ← std for ith PC

weight from eigenvector   ← std for jth variable

R

$s_j = 1$ due to standardization

$l_{11} = 0.469 \cdot \sqrt{2.437} \approx 0.7322$ $l_{21} = -0.368 \cdot \sqrt{1.407} \approx -0.4365$

$l_{12} = 0.532 \cdot \sqrt{2.437} \approx 0.8305$ $l_{22} = -0.236 \cdot \sqrt{1.407} \approx -0.2799$

$l_{13} = 0.465 \cdot \sqrt{2.437} \approx 0.7259$ $l_{23} = -0.315 \cdot \sqrt{1.407} \approx -0.3736$

$l_{14} = 0.387 \cdot \sqrt{2.437} \approx 0.6041$ $l_{24} = 0.585 \cdot \sqrt{1.407} \approx 0.6939$

$l_{15} = 0.361 \cdot \sqrt{2.437} \approx 0.5636$ $l_{25} = 0.606 \cdot \sqrt{1.407} \approx 0.7188$

R

A higher loading indicates that the variable is influential when forming the PC. A rule-of-thumb cut-off value of 0.5 means that all variables were influential in forming PC_1 & variable 4 & 5 were influential in forming PC_2. The loadings are all positive.

Since the variables in question are rate of returns for stocks, one can view PC_1 as overall high return for stocks or a well performing overall stock market. PC_2 can be viewed as high oil stock returns, or a well-performing oil sector.

5

15

SU, DEPARTMENT OF STATISTICS

Room: LA Anonymous code: 001Y-LFF Sheet number: 4

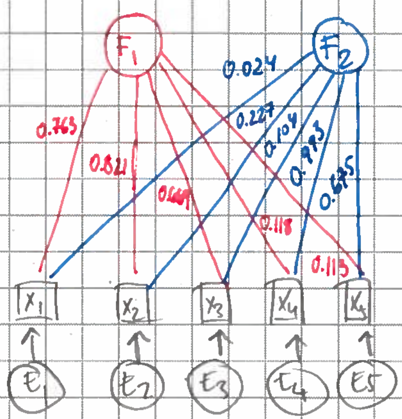
Question 3

EFA on correlation matrix

2 factors

Orthogonal rotation → no correlation between F_1 & F_2

Variable	F_1	F_2
x_1	0.763	0.024
x_2	0.821	0.227
x_3	0.669	0.104
x_4	0.118	0.993
x_5	0.113	0.675



quartimax

a) Two types of orthogonal rotations are varimax & orthomax.

They both aim to simplify the loading matrix: PCs !

Loadings $\begin{pmatrix} \vdots & \vdots \\ \vdots & \vdots \end{pmatrix}$

Varimax tries to simplify the columns of the matrix while the orthomax tries to simplify the rows of the matrix. Varimax rotation often results in PCs only loading high on one factor.

!?

2

b) The usual assumptions for the factor model are three-fold:

① The means of the indicators, the common factors, and the unique factors are equal to zero: $\boxed{E(X_i), E(F_j), E(\epsilon_i) = 0}$

② The variances of the indicators and the common factors are equal to 1: $\boxed{var(X_i), var(F_j) = 1}$

③ The unique factors are uncorrelated amongst themselves & uncorrelated w. the common factors: $\boxed{corr(\epsilon_i, F_j), corr(\epsilon_i, \epsilon_k) = 0}$ where $i \neq k$

4

c) Looking for shared variances between the indicators & each of the common factors.

Shared variance is equal to the squared structure loading, i.e. $\boxed{cor(X_i, F_j)^2}$ R

When there is no correlation between the common factors, ($\phi = 0$) the structure loading is equal to the pattern loading

$\boxed{cor(X_i, F_j) = \lambda_{ij}}$ R

This gives $cor(X_i, F_j)^2 = \lambda_{ij}^2$

Indicator	λ_{i1}^2	λ_{i2}^2
X_1	$0.763^2 \approx 0.5822$	$0.024^2 \approx 0.0006$
X_2	$0.821^2 \approx 0.6740$	$0.227^2 \approx 0.0515$
X_3	$0.669^2 \approx 0.4476$	$0.104^2 \approx 0.0108$
X_4	$0.118^2 \approx 0.0139$	$0.993^2 \approx 0.9860$
X_5	$0.113^2 \approx 0.0128$	$0.675^2 \approx 0.4556$

R

Answer: See table above for shared variances

4

d) Looking for percentage of variance of X_4 & X_5 not accounted for by F_1 & F_2.

$\boxed{var(X_i) = \lambda_1^2 + \lambda_2^2 + var(\varepsilon_i) + 2\lambda_1\lambda_2\phi}$, but $\phi = 0$ & $var(X) = 1$ (by assumption) gives: $1 = \lambda_1^2 + \lambda_2^2 + var(\varepsilon_i)$

$\rightarrow \underbrace{1 - \lambda_1^2 - \lambda_2^2}_{\text{variance not accounted for by } F_1 \,\&\, F_2} = var(\varepsilon_i) =$ variance of unique factor

R

For X_4: $1 - 0.0139 - 0.9860 = 0.0001 = var(\varepsilon_4) \rightarrow 0.01\%$ (since tot. var. = 1)

X_5: $1 - 0.0128 - 0.4556 = 0.5316 = var(\varepsilon_5) \rightarrow 53.16\%$ (— " —)

R

Answer: The percentage not accounted for by common factors is 0.01% for X_4 & 53.16% for X_5. This is called the unique factor variance.

4

14

SU, DEPARTMENT OF STATISTICS

Room: LA Anonymous code: 0014-LFF Sheet number: 5

Question 4

$$D = \begin{array}{c|cccccc} & Atl. & Bos. & Chi. & Dal & Den. & Det \\ \hline Atlanta & 0 & & & & & \\ Boston & 536 & 0 & & & & \\ Chicago & 516 & 447 & 0 & & & \\ Dallas & 590 & 833 & 924 & 0 & & \\ Denver & 694 & 915 & 1073 & 528 & 0 & \\ Detroit & 716 & 881 & 972 & 465 & 359 & 0 \end{array}$$

closest

a) Looking for: Complete-linkage clustering of cities

Algorithm: ① start w/ single observation-clusters

② calculate distances in line w/ method

③ merge 2 closest clusters

④ recalculate distances according to method

⑤ repeat ③ & ④ until only 1 cluster remains

R

Complete linkage maximizes the distance of the cluster

$\boxed{\max(d_{ij})}$

Because the distances are already calculated in the matrix above, we do not need to calculate the distance.

R

The closest observations are Denver & Detroit ($D=359$) so these will merge into cluster (Denver, Detroit).

R

Recalculating distances: $\max(d_{Atlanta\,(Denver,\,Detroit)}) = \max(d_{Atlanta, Denver}, d_{Atlanta, Detroit})$

$= \max(694, 716) = 716$

$\max(d_{Boston\,(Denver,\,Detroit)}) = \max(915, 881) = 915$

$\max(d_{Chicago\,(Denver,\,Detroit)}) = \max(1073, 972) = 1073$

$\max(d_{Dallas\,(Denver,\,Detroit)}) = \max(528, 465) = 528$

R

New matrix:

$$D = \begin{array}{l} \\ \text{Atlanta} \\ \text{Boston} \\ \text{Chicago} \\ \text{Dallas} \\ \text{(Denver, Detroit)} \end{array} \begin{pmatrix} \text{Atl.} & \text{Bos.} & \text{Chi.} & \text{Dallas} & \text{(D, D)} \\ 0 & & & & \\ 536 & 0 & & & \\ 516 & 447 & 0 & & \\ 590 & 833 & 924 & 0 & \\ 716 & 915 & 1073 & 528 & 0 \end{pmatrix}$$

The closest observations are Boston & Chicago ($D = 447$) & will merge into (Boston, Chicago)

R

Recalculating distances: $\max(d_{Atlanta\,(Boston, Chicago)}) = \max(d_{Atlanta, Boston}, d_{Atlanta, Chicago})$

$= \max(536, 516) = 536$

$\max(d_{Dallas\,(Boston, Chicago)}) = \max(833, 924) = 924$

$\max(d_{(Denver, Detroit)(Boston, Chicago)}) = \max(915, 1073) = 1073$

New matrix:

$$D = \begin{array}{l} \\ \text{Atlanta} \\ \text{(Boston, Chicago)} \\ \text{Dallas} \\ \text{(Denver, Detroit)} \end{array} \begin{pmatrix} \text{Atl} & \text{(B C)} & \text{Dal.} & \text{(D D)} \\ 0 & & & \\ 536 & 0 & & \\ 590 & 924 & 0 & \\ 716 & 1073 & 528 & 0 \end{pmatrix}$$

R

The closest observations are Dallas & (Denver, Detroit) & will merge into (Dallas, Denver, Detroit).

R

Recalculating distances: $\max(d_{Atlanta\,(D,D,D)}) = \max(d_{Atlanta, Dallas}, d_{Atlanta\,(Dallas, Detroit)})$

$= \max(590, 716) = 716$

$\max(d_{(B,C)(D,D,D)}) = \max(d_{(B,C)\,Dallas}, d_{(BC)\,(Dallas, Detroit)})$

$= \max(924, 1073) = 1073$

New matrix:

$$D = \begin{array}{l} \\ \text{Atlanta} \\ \text{(Boston, Chicago)} \\ \text{(Dallas, Denver, Detroit)} \end{array} \begin{pmatrix} \text{Atl.} & \text{(B C)} & \text{(D D D)} \\ 0 & & \\ 536 & 0 & \\ 716 & 1073 & 0 \end{pmatrix}$$

R

The closest observations are Atlanta & (Boston, Chicago) & will merge into (Atlanta, Boston, Chicago).

R

SU, DEPARTMENT OF STATISTICS

Room: LA Anonymous code: 0014-LFF Sheet number: 6

Cont. question 4

cont. a) Recalculating distances: $\max(d_{(ABC)(DDD)}) = \max(d_{Atlanta,(DDD)}, d_{(BC),(DDD)})$
$= \max(716, 1073) = 1073$

New matrix:

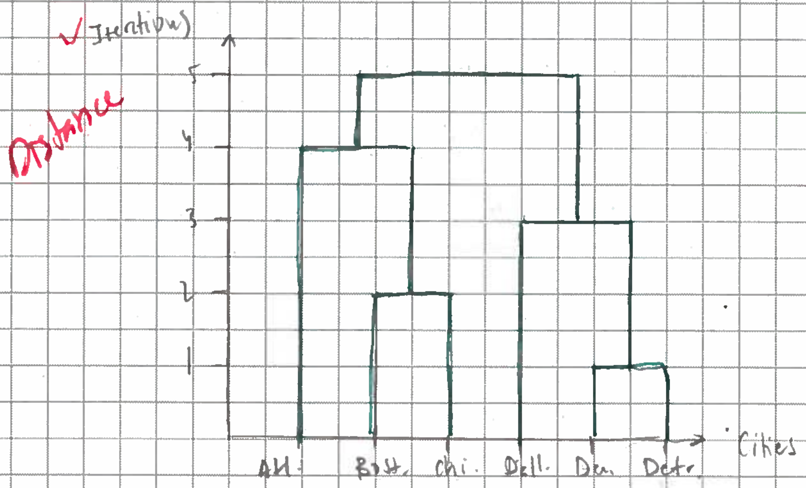
$$D = \begin{array}{c|cc} & (ABC) & (DDD) \\ (\text{Atlanta, Boston, chicago}) & 0 & 1073 \\ (\text{Dallas, Denver, Detroit}) & 1073 & 0 \end{array}$$

R

The final cluster is to merge all observations.

10

b) Dendogram of clustering:



Mergers
1. (Denver, Detroit)
2. (Boston, chicago)
3. (Dallas, (Denver, Detroit))
4. (Atlanta, (Boston, chicago))
5. ((Atlanta, Boston, Chicago) (Dalles, Denver, Detroit))

2

c) One advantage of hierarchical clustering methods is that they will not give you different clustering solutions depending on initial choice of clusters, as can happen with non-hierarchical clustering methods. One disadvantage is that the solution is sensitive to the choice of distance computation method (average linkage, complete linkage, centroid method etc.)

2

14

SU, DEPARTMENT OF STATISTICS

Room: CA Anonymous code: 0014-LFF Sheet number: 7

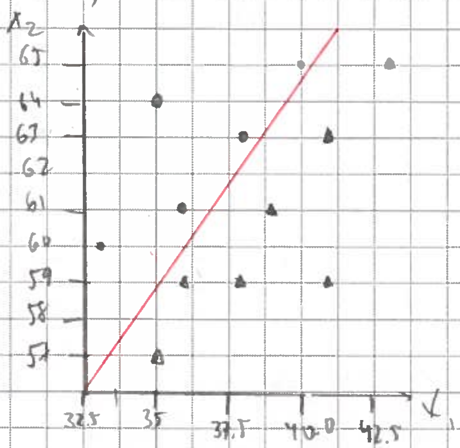
Question 5

$p = 2$; $g = 2$; $n_1 = 5$; $n_2 = 7$

Temperature 1 X_1	Temperature 1 X_2	Temperature 2 X_1	Temperature 2 X_2
33	60	35	57
36	61	36	59
35	64	38	59
38	63	39	61
40	65	41	63
		43	65
		41	59

$\bar{X}_1 = \begin{pmatrix} 36.4 \\ 62.6 \end{pmatrix}$ $\bar{X}_2 = \begin{pmatrix} 39.0 \\ 60.4 \end{pmatrix}$ $SSCP_1 = \begin{pmatrix} 29.2 & 16.8 \\ 16.8 & 17.2 \end{pmatrix}$ $SSCP_2 = \begin{pmatrix} 50 & 40 \\ 40 & 45.7 \end{pmatrix}$

a) Simply by looking at the graph it does not seem like the two variables discriminate well individually but that they would discriminate quite well jointly



Comparing $SSCP_{within}$ & $SSCP_{between}$ also supports this view

$$SSCP_{within} = \sum_{l=1}^{g} SSCP_l = SSCP_1 + SSCP_2 = \begin{pmatrix} 29.2 & 16.8 \\ 16.8 & 17.2 \end{pmatrix} + \begin{pmatrix} 50 & 40 \\ 40 & 45.7 \end{pmatrix}$$

$$= \begin{pmatrix} 79.2 & 56.8 \\ 56.8 & 62.9 \end{pmatrix}$$

R

$$SSCP_{between} = \frac{n_1 n_2}{n_1 + n_2} \cdot (\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)^T = \frac{5 \cdot 7}{5+7} \cdot \left[\begin{pmatrix} 36.4 \\ 62.6 \end{pmatrix} - \begin{pmatrix} 39.0 \\ 60.4 \end{pmatrix}\right]\left[\begin{pmatrix} 36.4 \\ 62.6 \end{pmatrix} - \begin{pmatrix} 39.0 \\ 62.6 \end{pmatrix}\right]^T$$

$$\begin{pmatrix} 36.4 \\ 62.6 \end{pmatrix} - \begin{pmatrix} 39.0 \\ 60.4 \end{pmatrix} = \begin{pmatrix} -2.6 \\ 2.2 \end{pmatrix}$$

$$\begin{pmatrix} -2.6 \\ 2.2 \end{pmatrix}^T = \begin{pmatrix} -2.6 & 2.2 \end{pmatrix} \rightarrow \underset{(2\times1)}{\begin{pmatrix} -2.6 \\ 2.2 \end{pmatrix}} \underset{(1\times2)}{\begin{pmatrix} -2.6 & 2.2 \end{pmatrix}}$$

$$= \underset{(2\times2)}{\begin{pmatrix} -2.6\,(-2.6) & 2.2 \cdot (-2.6) \\ -2.6 \cdot 2.2 & 2.2 \cdot 2.2 \end{pmatrix}} = \begin{pmatrix} 6.76 & -5.72 \\ -5.72 & 4.84 \end{pmatrix}$$

$$\rightarrow SSCP_{between} = \frac{35}{12} \cdot \begin{pmatrix} 6.76 & -5.72 \\ -5.72 & 4.84 \end{pmatrix} \approx \begin{pmatrix} 19.717 & -16.683 \\ -16.683 & 14.117 \end{pmatrix}$$

Comparing $SSCP_{between} = \begin{pmatrix} 19.717 & -16.683 \\ -16.683 & 14.117 \end{pmatrix}$ and $SSCP_{within} = \begin{pmatrix} 79.2 & 56.8 \\ 56.8 & 62.9 \end{pmatrix}$

Shows that the dispersion is much greater within the groups than between the groups and because of this the individual variables might not discriminate too well.

3

b) Looking for: pooled covariance matrix (S_{pooled})

$$S_{pooled} = \frac{SSCP_{within}}{n-g} = \frac{1}{13-2} \cdot SSCP_{within} = \frac{1}{11} \cdot \begin{pmatrix} 79.2 & 56.8 \\ 56.8 & 62.9 \end{pmatrix}$$

✓

$$\approx \begin{pmatrix} 7.2 & 5.164 \\ 5.164 & 5.718 \end{pmatrix}$$

3

c) Looking for: Fisher's linear discriminant function.

Fisher's linear discriminant function: $\gamma^T = (\mu_1 - \mu_2)^T \Sigma^{-1}$

Estimating using $\bar{x}_1, \bar{x}_2$ & S_{pooled} $\rightarrow \hat{\gamma}^T = \underset{(1\times2)}{(\bar{x}_1 - \bar{x}_2)^T} \underset{(2\times2)}{S_{pooled}^{-1}}$

R

Since before: $(\bar{x}_1 - \bar{x}_2)^T = \underset{(1\times2)}{(-2.6 \quad 2.2)}$

R

SU, DEPARTMENT OF STATISTICS

Room: LA Anonymous code: 0014-LFF Sheet number: 8

Cont. question 5

Inverse of a 2×2 matrix: $A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$

$$\rightarrow S_{pooled}^{-1} = \frac{1}{7.2 \times 5.718 - 5.164^2} \cdot \begin{pmatrix} 5.718 & -5.164 \\ -5.164 & 7.2 \end{pmatrix}$$

$$\approx \frac{1}{14.503} \cdot \begin{pmatrix} 5.718 & -5.164 \\ -5.164 & 7.2 \end{pmatrix}$$

R

$$\approx \begin{pmatrix} 0.394 & -0.355 \\ -0.355 & 0.496 \end{pmatrix}$$

Discriminant function: $\hat{\gamma}^T = \underset{(1 \times 2)}{(-2.6 \quad 2.2)} \underset{(2 \times 2)}{\begin{pmatrix} 0.394 & -0.355 \\ -0.355 & 0.496 \end{pmatrix}}$

$$\overset{\checkmark}{=} (-2.6) \cdot 0.394 + (\overset{2.2}{-2.6}) \cdot (-0.355) = -0.1014$$

$$(\underset{-2.6}{2.2} \cdot (-0.355)) + 2.2 \cdot 0.496 = 0.3102$$

$$= \underset{(1 \times 2)}{(-0.1014 \quad 0.3102)}$$

Giving $Z = \hat{\gamma}^T \cdot X = -0.1014 x_1 + 0.3102 x_2$

Answer: Fisher's linear discriminant function is $\hat{\gamma}^T = (-0.1014 \quad 0.3102)$,

i.e.: $Z = -0.1014 x_1 + 0.3102 x_2$.

8

14