

University of Michigan, Dept of Statistics

Stat 503, Instructor: Long Nguyen

### Homework Assignment 2

(due Thursday, Feb 8, 2017)

in class or Canvas dropbox by 1pm

1. Consider two random vectors  $F \sim N_2(0, I_2)$  (a multivariate normal vector in  $\mathbb{R}^2$ ),  $\epsilon \sim N_4(0, \Psi)$  (a multivariate normal vector in  $\mathbb{R}^4$ ) be defined as follows

$$\begin{aligned} X_1 &= F_1 + \epsilon_1 \\ X_2 &= 2F_1 - F_2 + \epsilon_2 \\ X_3 &= F_1 + 2F_2 + \epsilon_3 \\ X_4 &= 3F_1 - 4F_2 - \epsilon_4 \end{aligned}$$

- (a) What *additional* assumptions must be made to make this a factor analysis model?
  - (b) Assuming  $\Psi = \text{diag}(1, 0.5, 0.5, 2)$  and all assumptions from (a) hold, calculate the matrices  $\Sigma$  and  $\Lambda$ , and write down the factor analysis covariance decomposition.
  - (c) Produce the factor score estimates for data point  $x = (1, 1, 1, 1)$ .
2. **Code-it-yourself MDS.** The figure below depicts airline distances among 10 US cities.
    - (a) Please write your own codes to implement the classical MDS program to produce a 2-dimensional map for these cities. (To be clear: the input to your program is the distance matrix given below, the output is a map with name and location of all cities. No R package is allowed. Please submit your codes).
    - (b) Try one or more choices of distances defined in terms of the given distance matrix and report your findings.

CITIES	ATLA	CHIC	DEN	HOU	L.A.	MIAMI	N.Y.	S.F.	SEAT	WASH D.C.
ATLANTA		587	1212	701	1936	604	748	2139	2182	543
CHICAGO	587		920	940	1745	1188	713	1858	1737	597
DENVER	1212	920		879	831	1726	1631	949	1021	1494
HOUSTON	701	940	879		1374	968	1420	1645	1891	1220
LOS ANGELES	1936	1745	831	1374		2339	2451	347	959	2300
MIAMI	604	1188	1726	968	2339		1092	2594	2734	923
NEW YORK	748	713	1631	1420	2451	1092		2571	2408	205
SAN FRANCISCO	2139	1858	949	1645	347	2594	2571		678	2442
SEATTLE	2182	1737	1021	1891	959	2734	2408	678		2329
WASHINGTON	543	597	1494	1220	2300	923	205	2442	2329	

(B) AIRLINE DISTANCES BETWEEN TEN U.S. CITIES

3. Using the data set in `auto-mpg.data` from Homework 1, perform factor analysis and multidimensional scaling. Consider carefully which variables to include and to omit in each of them. Compare to the results of PCA from Homework 1. Comment on any interesting features of the dataset and/or individual observations you were able to find.

**Instructions** The solution to problem 1 can be either typed up or written by hand. The solution to data analysis questions (Problem 3) may be written as a data analysis report. The report needs to be clear, concise, and to the point. There should be no graphs or tables that are not commented on in the text. Please include your codes (at least the main parts) in a *separate* appendix at the end of the report.