Winter 2019
DA 410
Multivariate Analysis
Charlene Cheng

Take Home Final
Exam March 2019

Name (Print):_____

I acknowledge and accept the Honor Code

Signature:_____

Score:_____/100

Instructions:

- Please complete this exam in a Word document, and save it as ***DA410_Final_XXXXX***, where XXXXX is the first five letters of your last name. Make sure you put down the problem # clearly.

This exam is Open Book, Open Notes, No Discussions, No Questions Answered by Instructor.

- .
- Exam must be submitted through Assignment Tool by **11:59pm of Sunday, 3/17/19 (Pacific Time). NO late final will be accepted.**
- Round to the THIRD decimal place, unless otherwise noted in the instruction.
- **PLEASE SHOW ALL YOUR WORK COMPLETELY AND CLEARLY!!! Without sufficient work shown will result at least 50% penalty of the total available credit of the question.**
- ***You may use R, make sure to include the commands and outputs, as well as the interpretations of the outputs.***

☺ Good Luck ☺

You may use R and built-in functions. Make sure to include the commands and outputs, as well as the interpretations of the outputs.

Problem 1:

Suppose our multivariate date have covariance matrix $s = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 8 \end{pmatrix}$

(a) Find the eigenvalues and eigenvectors of S.
(b) Show the percent of variance explained.
(c) Decide how many components to retain.

Show your reason completely and clearly, including all necessary commands and outputs, as well as interpretations/conclusions.

Problem 2:

The correlation matrix given below arises from the scores of 220 boys in six school subjects: (1) French, (2) English, (3) history, (4) arithmetic, (5) algebra, and (6) geometry. Obtain principal component loadings for three factors.

$$\mathbf{R} = \begin{array}{c} \\ \text{French} \\ \text{English} \\ \text{History} \\ \text{Arithmetic} \\ \text{Algebra} \\ \text{Geometry} \end{array} \begin{pmatrix} 1.00 & & & & & \\ 0.44 & 1.00 & & & & \\ 0.41 & 0.35 & 1.00 & & & \\ 0.29 & 0.35 & 0.16 & 1.00 & & \\ 0.33 & 0.32 & 0.19 & 0.59 & 1.00 & \\ 0.25 & 0.33 & 0.18 & 0.47 & 0.46 & 1.00 \end{pmatrix}$$

Show your work completely and clearly, including all necessary commands and outputs, as well as interpretations/conclusions.

Problem 3:

For the Foodstuff Contents data set below, column names are “Food Energy Protein Fat Calcium Iron”
:

(a) Discuss your choice of the number of factors.
(b) Obtain principal component loadings.

(c) Calculate percent of variance explained for each factor, plot the factor scores using appropriate plot(s), and decide how many components to retain.

Show your work completely and clearly, including all necessary commands and outputs, as well as interpretations/conclusions.

```
BB 340 20 28 9 2.6
HR 245 21 17 9 2.7
BR 420 15 39 7 2.0
BS 375 19 32 9 2.5
BC 180 22 10 17 3.7
CB 115 20 3 8 1.4
CC 170 25 7 12 1.5
BH 160 26 5 14 5.9
LL 265 20 20 9 2.6
LS 300 18 25 9 2.3
HS 340 20 28 9 2.5
PR 340 19 29 9 2.5
PS 355 19 30 9 2.4
BT 205 18 14 7 2.5
VC 185 23 9 9 2.7
FB 135 22 4 25 0.6
AR 70 11 1 82 6.0
AC 45 7 1 74 5.4
TC 90 14 2 38 0.8
HF 135 16 5 15 0.5
MB 200 19 13 5 1.0
MC 155 16 9 157 1.8
PF 195 16 11 14 1.3
SC 120 17 5 159 0.7
DC 180 22 9 367 2.5
UC 170 25 7 7 1.2
RC 110 23 1 98 2.6
```

Problem 4:

The data below measures in five variables in comparison of normal patients and diabetics:

x_1 : glucose intolerance
x_2 : insulin response to oral glucose
x_3 : insulin resistance
y_1 : relative weight
y_2 : fasting plasma glucose

(a) Find the canonical correlation between (y_1, y_2) and (x_1, x_2, x_3).
(b) Test the significance of each canonical correlation.

Show your work completely and clearly, including all necessary commands and outputs, as well as interpretations/conclusions.

Patient Number	y_1	y_2	x_1	x_2	x_3
1	.81	80	356	124	55
2	.95	97	289	117	76
3	.94	105	319	143	105
4	1.04	90	356	199	108
5	1.00	90	323	240	143
6	.76	86	381	157	165
7	.91	100	350	221	119
8	1.10	85	301	186	105
9	.99	97	379	142	98
10	.78	97	296	131	94
11	.90	91	353	221	53
12	.73	87	306	178	66
13	.96	78	290	136	142
14	.84	90	371	200	93
15	.74	86	312	208	68
16	.98	80	393	202	102
17	1.10	90	364	152	76
18	.85	99	359	185	37
19	.83	85	296	116	60
20	.93	90	345	123	50
21	.95	90	378	136	47
22	.74	88	304	134	50
23	.95	95	347	184	91
24	.97	90	327	192	124
25	.72	92	386	279	74

Problem 5:

The data consists of mental ability test scores of seventh- and eighth-grade children from two different schools (Pasteur and Grant-White). In our version of the dataset, only 9 out of the original 26 tests are included. A CFA model that is often proposed for these 9 variables consists of three latent variables (or factors), each with three indicators:

- a *visual* factor measured by 2 variables: x1 and x2
- a *textual* factor measured by 4 variables: x3, x4, x5 and x6
- a *speed* factor measured by 3 variables: x7, x8 and x9
- a visual factor and a textual factor have zero correlation

(a) Please draw a figure contains a graphical representation of the three-factor model.
(b) Please write out the corresponding syntax for specifying this model.

Example as below

```
HS.model <- ' visual  =~ x1 + x2 + x3
              textual =~ x4 + x5 + x6
              speed   =~ x7 + x8 + x9 '
```

Problem 6:

Make a conclusion for this class DA 410, make sure you include the following aspects:

(a) How many models you have learnt, use 3 to 5 sentences to explain each of them.
(b) Which one really impressed you when you learnt and why
(c) Which one is your favorite one and why
(d) Select two models out, make a comparison. Show the differences and similarities between them.
(e) If you will build up a project to solve some real problem using one of them, which one you would like to you, and what kind of project you will like to build. (200 words)

The End

Take Home Final
Exam March 2019