Winter 2019
DA 410
Multivariate Analysis
Charlene Cheng

Take Home Midterm
Exam February 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_Midterm_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, 2/10/19 (Pacific Time).**
- 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.**
- Some questions do allow you to use R, while some do not -- you would have to solve the questions by hand/formula/calculator (using R in this case will result at least 50% penalty of the total available credit of the question). Make sure you read the instructions carefully.

☺ Good Luck ☺

In Problem 1 through Problem 3: You are NOT allowed to use R

Problem 1:

The columns of $A = \begin{pmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{pmatrix}$ are mutually orthogonal.

(a) Normalize the column of *A* by dividing each column by its length; denote the resulting matrix by *C*.
(b) Show that *C* is an orthogonal matrix.

Note: you are NOT allowed to use R in this question.

Problem 2:

$$A = \begin{pmatrix} 3 & 2 \\ -1 & 3 \end{pmatrix}, B = \begin{pmatrix} 0 & 4 \\ 6 & 2 \end{pmatrix}$$

(a) Find *AB.*
(b) Find *BA*.

Note: you are NOT allowed to use R in this question.

Problem 3:

$$A = \begin{pmatrix} 3 & 5 & 2 \\ 7 & 5 & 6 \end{pmatrix}, B = \begin{pmatrix} 4 & 3 & -2 \\ -5 & 8 & 3 \end{pmatrix}$$

(a) Find $A + B$ and $A - B$.
(b) Find $A'A$ and AA'.
(c) Find $(A + B)'$ and $A' + B'$.

Note: you are NOT allowed to use R in this question.

In Problem 4 through Problem 8: You may use R and built-in fuctions. Make sure to include the commands and outputs, as well as the interpretations of the outputs.

Problem 4:

In the following table, we have a comparison of four reagents. The first reagent is the one presently in use and the other three are less expensive reagents that we wish to compare with the first. All four reagents are used with a blood sample from each patient.

	Reagent 1			Reagent 2			Reagent 3			Reagent 4		
Subject	y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
1	8.0	3.96	12.5	8.0	3.93	12.7	7.9	3.86	13.0	7.9	3.87	13.2
2	4.0	5.37	16.9	4.2	5.35	17.2	4.1	5.39	17.2	4.0	5.35	17.3
3	6.3	5.47	17.1	6.3	5.39	17.5	6.0	5.39	17.2	6.1	5.41	17.4
4	9.4	5.16	16.2	9.4	5.16	16.7	9.4	5.17	16.7	9.1	5.16	16.7
5	8.2	5.16	17.0	8.0	5.13	17.5	8.1	5.10	17.4	7.8	5.12	17.5
6	11.0	4.67	14.3	10.7	4.60	14.7	10.6	4.52	14.6	10.5	4.58	14.7
7	6.8	5.20	16.2	6.8	5.16	16.7	6.9	5.13	16.8	6.7	5.19	16.8
8	9.0	4.65	14.7	9.0	4.57	15.0	8.9	4.58	15.0	8.6	4.55	15.1
9	6.1	5.22	16.3	6.0	5.16	16.9	6.1	5.14	16.9	6.0	5.21	16.9
10	6.4	5.13	15.9	6.4	5.11	16.4	6.4	5.11	16.4	6.3	5.07	16.3
11	5.6	4.47	13.3	5.5	4.45	13.6	5.3	4.46	13.6	5.3	4.44	13.7
12	8.2	5.22	16.0	8.2	5.14	16.5	8.0	5.14	16.5	7.8	5.16	16.5
13	5.7	5.10	14.9	5.6	5.05	15.3	5.5	5.02	15.4	5.4	5.05	15.5
14	9.8	5.25	16.1	9.8	5.15	16.6	8.1	5.10	13.8	9.4	5.16	16.6
15	5.9	5.28	15.8	5.8	5.25	16.4	5.7	5.26	16.4	5.6	5.29	16.2
16	6.6	4.65	12.8	6.4	4.59	13.2	6.3	4.58	13.1	6.4	4.57	13.2
17	5.7	4.42	14.5	5.5	4.31	14.9	5.5	4.30	14.9	5.4	4.32	14.8
18	6.7	4.38	13.1	6.5	4.32	13.4	6.5	4.32	13.6	6.5	4.31	13.5
19	6.8	4.67	15.6	6.6	4.57	15.8	6.5	4.55	16.0	6.5	4.56	15.9
20	9.6	5.64	17.0	9.5	5.58	17.5	9.3	5.50	17.4	9.2	5.46	17.5

The three variables measured for each reagent are y_1=white blood count, y_2=red blood count, and y_3=hemoglobin count.

Compare the four reagents using all four MANOVA tests. State each hypotheses clearly, and interpret the results.

Problem 5:

The table below displays scores on math, English, and art tests for 5 students. Note that data from the table is represented in matrix *A*, where each column in the matrix shows scores on a test and each row shows scores for a student:

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

(a) Calculate the sample covariance matrix S
(b) Calculate the sample correlation matrix R.
(c) Now let's define $z = -2y_1 + 3y_2 + y_3$, where $y_1$denotes Math scores, y_2 denotes English scores, and y_3 denotes Art scores. Find the sample mean vector $\bar{z}$ and the sample variance s_z^2.

Problem 6:

Use the beetle data, do the following:

(a) Find the classification function and cutoff point.
(b) Find the classification table using the nearest neighbor method by setting k = 3.
(c) Calculate misclassification rate.

Haltica oleracea					*Haltica carduorum*				
Experiment Number	y_1	y_2	y_3	y_4	Experiment Number	y_1	y_2	y_3	y_4
1	189	245	137	163	1	181	305	184	209
2	192	260	132	217	2	158	237	133	188
3	217	276	141	192	3	184	300	166	231
4	221	299	142	213	4	171	273	162	213
5	171	239	128	158	5	181	297	163	224
6	192	262	147	173	6	181	308	160	223
7	213	278	136	201	7	177	301	166	221
8	192	255	128	185	8	198	308	141	197
9	170	244	128	192	9	180	286	146	214
10	201	276	146	186	10	177	299	171	192
11	195	242	128	192	11	176	317	166	213
12	205	263	147	192	12	192	312	166	209
13	180	252	121	167	13	176	285	141	200
14	192	283	138	183	14	169	287	162	214
15	200	294	138	188	15	164	265	147	192
16	192	277	150	177	16	181	308	157	204
17	200	287	136	173	17	192	276	154	209
18	181	255	146	183	18	181	278	149	235
19	192	287	141	198	19	175	271	140	192
					20	197	303	170	205

Problem 7:

Use the above beetle data, do the following:

(a) Use LDA by setting probability of 50% and 50% to train model.
(b) Predict new observation (189,245,138,164).
(c) Calculate misclassification rate.

Problem 8:

The following table contains data from O'Sullivan and Mahan with measurements of blood glucose levels on three occasions for 30 women. The y's represent fasting glucose measurements on the three occasions; the x's are glucose measurements 1 hour after sugar intake. Find the mean vector and covariance matrix for all six variables and partition them into $\begin{pmatrix} \bar{y} \\ \bar{x} \end{pmatrix}$, and $S = \begin{pmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{pmatrix}$.

Blood Glucose Measurements on Three Occasions

Fasting			One Hour After Sugar Intake		
y_1	y_2	y_3	x_1	x_2	x_3
62	75	68	116	130	91
74	64	70	109	101	103
64	71	66	77	102	130
73	70	64	115	110	109
68	67	75	76	85	119
69	82	74	72	133	127
60	67	61	130	134	121
70	74	78	150	158	100
66	74	78	150	131	142
83	70	74	99	98	105
68	66	90	119	85	109
78	63	75	164	98	138
103	77	77	160	117	121
77	68	74	144	71	153
66	77	68	77	70	109
70	70	72	114	93	122
75	65	71	77	82	89
91	74	93	114	93	122
66	75	73	77	70	109
75	82	76	153	132	115
74	71	66	143	105	100
76	70	64	114	113	129
74	82	76	73	106	116
74	71	66	116	81	77
67	70	64	63	87	70
67	90	86	63	87	70
78	77	80	105	132	80
64	71	69	83	94	133
71	76	80	81	87	86
63	73	71	120	89	59

Problem 9:

Various aspects of economic cycles were measured for consumer goods and producer goods by Tintner. The variables are

y_1=length of cycle

y_2=percentage of rising prices

y_3=cyclical amplitude

y_4=rate of change

The data for several items are given in the following table:

Consumer Goods					Producer Goods				
Item	y_1	y_2	y_3	y_4	Item	y_1	y_2	y_3	y_4
1	72	50	8	0.5	1	57	57	12.5	0.9
2	66.5	48	15	1.0	2	100	54	17	0.5
3	54	57	14	1.0	3	100	32	16.5	0.7
4	67	60	15	0.9	4	96.5	65	20.5	0.9
5	44	57	14	0.3	5	79	51	18	0.9
6	41	52	18	1.9	6	78.5	53	18	1.2
7	34.5	50	4	0.5	7	48	50	21	1.6
8	34.5	46	8.5	1.0	8	155	44	20.5	1.4
9	24	54	3	1.2	9	84	64	13	0.8
					10	105	35	17	1.8

Use Hotelling's T^2 test to test for a difference in the mean measurements vector of the Consumers Goods and the mean vector of the Producer Goods. State each hypotheses clearly, and interpret the results.