

STAT450/550 Multivariate Statistical Analysis

Take-home MT (Due Monday, May 5)

-
- Note:**
1. Late submission will NOT be accepted.
 2. You have to write the project paper as if you are submitting a statistical report to the company you are consulting, assuming that they are not statisticians. Your paper must be concise, precise, and easy to read (please type in a Word Process).
 3. Carefully labeled and captioned outputs (e.g., Table 1, ..., Figure 1,...) should be inside of your paper or attach as appendices. The SAS outputs must be well summarized and organized in tables and figures. Do not turn in the raw SAS outputs. Resize figures and tables. Do NOT turn in any outputs not discussed.
 4. To receive a full credit you must include complete discussions and your paper must be brief and well organized. Do not exceed 5 pages for each part, excluding tables and figures.
-

Part I

Consider the Business School Data below (source: *AMD, J. Lattin et al, 2003*). We are interested in discriminate analysis and classification among the graduates of different schools using the info available. Perform complete DA, Logistic DA, and Canonical DA. Visualize the results and compare the methods.

```
/******  
* DATA B_SCHOOL *  
*****  
ID : Student ID number (1 to 100)  
  
School: Business school attended  
        1 = School A (East Coast)  
        2 = School B (East Coast)  
        3 = School C (Midwest)  
        4 = School D (West Coast)  
  
SI1-SI4: Indicators of School  
  
GPA: Grade point average  
  
GMAT: GMAT score  
  
Major: College Major  
        1 = Humanities / Social Sciences  
        2 = Math/Engineering  
        3 = Business  
        4 = Other  
  
MI1-MI4: Indicators of Major  
  
Gender: 1=Male, 2=Female  
  
EXP: Years of business experience (6=more than 5)  
*/
```

Data B_school;

Input ID School SI1-SI4 GPA GMAT Major MI1-MI4 Gender Exp;

Datalines;

1	3	0	0	1	0	3.6	650	1	1	0	0	0	2	3
2	3	0	0	1	0	3.7	680	3	0	0	1	0	1	2
3	4	0	0	0	1	3.8	710	1	1	0	0	0	1	2
4	4	0	0	0	1	3.7	720	1	1	0	0	0	2	6
5	2	0	1	0	0	3.9	680	1	1	0	0	0	1	4
6	4	0	0	0	1	3.8	680	4	0	0	0	1	1	3
7	4	0	0	0	1	3.8	670	1	1	0	0	0	2	2
8	1	1	0	0	0	3.9	660	1	1	0	0	0	2	2
9	2	0	1	0	0	3.7	690	1	1	0	0	0	1	2
10	4	0	0	0	1	3.3	650	2	0	1	0	0	1	4
11	4	0	0	0	1	3.7	690	3	0	0	1	0	1	2
12	3	0	0	1	0	3.9	610	1	1	0	0	0	1	2
13	2	0	1	0	0	3.9	670	2	0	1	0	0	1	3
14	2	0	1	0	0	3.4	610	3	0	0	1	0	2	3
15	2	0	1	0	0	3.8	660	3	0	0	1	0	1	3
16	2	0	1	0	0	4.0	670	2	0	1	0	0	2	4
17	1	1	0	0	0	3.6	630	3	0	0	1	0	1	5
18	1	1	0	0	0	4.0	710	3	0	0	1	0	1	2
19	4	0	0	0	1	4.0	680	4	0	0	0	1	2	2
20	3	0	0	1	0	3.8	700	4	0	0	0	1	1	2
21	2	0	1	0	0	3.8	650	3	0	0	1	0	1	2
22	2	0	1	0	0	3.5	640	2	0	1	0	0	1	2
23	3	0	0	1	0	3.6	640	1	1	0	0	0	2	4
24	2	0	1	0	0	3.9	670	3	0	0	1	0	1	2
25	2	0	1	0	0	3.6	630	4	0	0	0	1	1	4
26	1	1	0	0	0	3.8	650	3	0	0	1	0	1	3
27	3	0	0	1	0	3.8	660	1	1	0	0	0	2	3
28	4	0	0	0	1	4.0	720	2	0	1	0	0	1	2
29	2	0	1	0	0	4.0	700	2	0	1	0	0	2	6
30	2	0	1	0	0	3.9	670	2	0	1	0	0	1	4
31	1	1	0	0	0	4.0	670	4	0	0	0	1	1	2
32	2	0	1	0	0	3.7	690	2	0	1	0	0	1	2
33	1	1	0	0	0	3.6	630	1	1	0	0	0	2	2
34	1	1	0	0	0	3.9	700	3	0	0	1	0	2	2
35	3	0	0	1	0	3.5	640	1	1	0	0	0	1	2
36	3	0	0	1	0	3.8	650	3	0	0	1	0	2	2
37	3	0	0	1	0	3.4	640	4	0	0	0	1	1	2
38	2	0	1	0	0	3.7	690	2	0	1	0	0	1	5
39	3	0	0	1	0	3.9	620	4	0	0	0	1	1	3
40	4	0	0	0	1	3.7	700	2	0	1	0	0	2	2
41	1	1	0	0	0	3.8	670	1	1	0	0	0	1	2
42	1	1	0	0	0	4.0	720	1	1	0	0	0	1	2
43	1	1	0	0	0	3.5	620	1	1	0	0	0	1	2
44	4	0	0	0	1	4.0	710	1	1	0	0	0	2	2
45	4	0	0	0	1	3.7	670	3	0	0	1	0	1	2
46	4	0	0	0	1	3.7	680	2	0	1	0	0	2	2
47	1	1	0	0	0	3.7	640	3	0	0	1	0	1	4
48	1	1	0	0	0	3.8	660	4	0	0	0	1	1	2
49	1	1	0	0	0	3.7	660	1	1	0	0	0	1	2
50	4	0	0	0	1	3.6	680	4	0	0	0	1	2	2
51	1	1	0	0	0	3.9	700	2	0	1	0	0	2	3
52	4	0	0	0	1	3.7	670	2	0	1	0	0	1	2
53	1	1	0	0	0	3.7	640	1	1	0	0	0	2	6
54	1	1	0	0	0	3.7	650	1	1	0	0	0	1	6
55	3	0	0	1	0	3.8	660	1	1	0	0	0	1	3
56	1	1	0	0	0	3.6	630	1	1	0	0	0	1	2

57	3	0	0	1	0	4.0	680	4	0	0	0	1	1	2
58	4	0	0	0	1	3.8	690	4	0	0	0	1	2	2
59	2	0	1	0	0	4.0	710	3	0	0	1	0	2	2
60	1	1	0	0	0	3.7	690	4	0	0	0	1	2	2
61	2	0	1	0	0	3.8	660	2	0	1	0	0	1	2
62	2	0	1	0	0	3.7	640	3	0	0	1	0	1	4
63	4	0	0	0	1	3.3	680	2	0	1	0	0	1	5
64	1	1	0	0	0	3.4	670	1	1	0	0	0	1	5
65	1	1	0	0	0	3.5	640	3	0	0	1	0	1	2
66	3	0	0	1	0	3.4	650	2	0	1	0	0	2	3
67	4	0	0	0	1	3.9	700	2	0	1	0	0	2	3
68	4	0	0	0	1	3.9	660	4	0	0	0	1	1	2
69	1	1	0	0	0	3.6	630	1	1	0	0	0	1	2
70	4	0	0	0	1	3.6	680	3	0	0	1	0	1	4
71	3	0	0	1	0	3.9	640	2	0	1	0	0	1	2
72	2	0	1	0	0	4.0	680	1	1	0	0	0	1	2
73	1	1	0	0	0	3.5	660	1	1	0	0	0	2	2
74	4	0	0	0	1	3.6	690	1	1	0	0	0	1	3
75	2	0	1	0	0	3.6	650	3	0	0	1	0	2	3
76	1	1	0	0	0	3.9	660	3	0	0	1	0	2	5
77	4	0	0	0	1	3.8	670	2	0	1	0	0	1	2
78	3	0	0	1	0	4.0	630	1	1	0	0	0	2	2
79	4	0	0	0	1	3.8	700	4	0	0	0	1	1	2
80	2	0	1	0	0	3.4	620	2	0	1	0	0	1	3
81	2	0	1	0	0	3.7	640	3	0	0	1	0	1	2
82	3	0	0	1	0	3.7	650	3	0	0	1	0	1	2
83	4	0	0	0	1	3.9	710	3	0	0	1	0	1	2
84	2	0	1	0	0	3.5	620	4	0	0	0	1	1	2
85	4	0	0	0	1	3.7	650	1	1	0	0	0	1	5
86	3	0	0	1	0	3.5	680	3	0	0	1	0	2	4
87	2	0	1	0	0	3.8	680	1	1	0	0	0	1	6
88	2	0	1	0	0	3.7	670	3	0	0	1	0	1	5
89	2	0	1	0	0	3.8	680	2	0	1	0	0	1	2
90	1	1	0	0	0	3.8	660	1	1	0	0	0	2	4
91	3	0	0	1	0	3.6	660	1	1	0	0	0	1	2
92	4	0	0	0	1	3.6	680	1	1	0	0	0	1	5
93	2	0	1	0	0	3.9	660	2	0	1	0	0	2	4
94	4	0	0	0	1	3.5	640	4	0	0	0	1	1	3
95	4	0	0	0	1	3.7	670	2	0	1	0	0	1	6
96	1	1	0	0	0	3.4	650	3	0	0	1	0	2	2
97	2	0	1	0	0	3.6	640	4	0	0	0	1	1	2
98	1	1	0	0	0	3.6	640	3	0	0	1	0	1	3
99	4	0	0	0	1	3.6	680	4	0	0	0	1	1	4
100	4	0	0	0	1	3.3	610	1	1	0	0	0	2	2

;

Part II

Consider the National Track Records Data for Women and for Men (*AMSA, Johnson and Wichern*). For each data, cluster the countries using both hierarchical (Single linkage and Complete linkage) and nonhierarchical (K-means) methods. Use the Euclidean distances as measures of (dis)similarity. Compare the results from two data sets. Perform principal components analyses on the variables and plot the observations in the space of the component(s) and label the points according to the cluster to which they were assigned. Note that if the cluster analysis was performed on standardized variables, you should use PCA on the correlation matrix.

Data women;

Group="Women";

Input x1-x7 country \$;

Label x1="100m" x2="200m" x3="400m" x4="800m" x5="1500m" x6="3000m"

x7="Marathon";

Datalines;

11.61	22.94	54.50	2.15	4.43	9.79	178.52	argentin
11.20	22.35	51.08	1.98	4.13	9.08	152.37	australi
11.43	23.09	50.62	1.99	4.22	9.34	159.37	austria
11.41	23.04	52.00	2.00	4.14	8.88	157.85	belgium
11.46	23.05	53.30	2.16	4.58	9.81	169.98	bermuda
11.31	23.17	52.80	2.10	4.49	9.77	168.75	brazil
12.14	24.47	55.00	2.18	4.45	9.51	191.02	burma
11.00	22.25	50.06	2.00	4.06	8.81	149.45	canada
12.00	24.52	54.90	2.05	4.23	9.37	171.38	chile
11.95	24.41	54.97	2.08	4.33	9.31	168.48	china
11.60	24.00	53.26	2.11	4.35	9.46	165.42	columbia
12.90	27.10	60.40	2.30	4.84	11.10	233.22	cookis
11.96	24.60	58.25	2.21	4.68	10.43	171.80	costa
11.09	21.97	47.99	1.89	4.14	8.92	158.85	czech
11.42	23.52	53.60	2.03	4.18	8.71	151.75	denmark
11.79	24.05	56.05	2.24	4.74	9.89	203.88	domrep
11.13	22.39	50.14	2.03	4.10	8.92	154.23	finland
11.15	22.59	51.73	2.00	4.14	8.98	155.27	france
10.81	21.71	48.16	1.93	3.96	8.75	157.68	gdr
11.01	22.39	49.75	1.95	4.03	8.59	148.53	frg
11.00	22.13	50.46	1.98	4.03	8.62	149.72	gbni
11.79	24.08	54.93	2.07	4.35	9.87	182.20	greece
11.84	24.54	56.09	2.28	4.86	10.54	215.08	guatemal
11.45	23.06	51.50	2.01	4.14	8.98	156.37	hungary
11.95	24.28	53.60	2.10	4.32	9.98	188.03	india
11.85	24.24	55.34	2.22	4.61	10.02	201.28	indonesi
11.43	23.51	53.24	2.05	4.11	8.89	149.38	ireland
11.45	23.57	54.90	2.10	4.25	9.37	160.48	israel
11.29	23.00	52.01	1.96	3.98	8.63	151.82	italy
11.73	24.00	53.73	2.09	4.35	9.20	150.50	japan
11.73	23.88	52.70	2.00	4.15	9.20	181.05	kenya
11.96	24.49	55.70	2.15	4.42	9.62	164.65	korea
12.25	25.78	51.20	1.97	4.25	9.35	179.17	dprkorea
12.03	24.96	56.10	2.07	4.38	9.64	174.68	luxembou
12.23	24.21	55.09	2.19	4.69	10.46	182.17	malaysia
11.76	25.08	58.10	2.27	4.79	10.90	261.13	mauritiu
11.89	23.62	53.76	2.04	4.25	9.59	158.53	mexico
11.25	22.81	52.38	1.99	4.06	9.01	152.48	netherla
11.55	23.13	51.60	2.02	4.18	8.76	145.48	nz
11.58	23.31	53.12	2.03	4.01	8.53	145.48	norway
12.25	25.07	56.96	2.24	4.84	10.69	233.00	png
11.76	23.54	54.60	2.19	4.60	10.16	200.37	philippi

11.13	22.21	49.29	1.95	3.99	8.97	160.82	poland
11.81	24.22	54.30	2.09	4.16	8.84	151.20	portugal
11.44	23.46	51.20	1.92	3.96	8.53	165.45	rumania
12.30	25.00	55.08	2.12	4.52	9.94	182.77	singapor
11.80	23.98	53.59	2.05	4.14	9.02	162.60	spain
11.16	22.82	51.79	2.02	4.12	8.84	154.48	sweden
11.45	23.31	53.11	2.02	4.07	8.77	153.42	switzerl
11.22	22.62	52.50	2.10	4.38	9.63	177.87	taipei
11.75	24.46	55.80	2.20	4.72	10.28	168.45	thailand
11.98	24.44	56.45	2.15	4.37	9.38	201.08	turkey
10.79	21.83	50.62	1.96	3.95	8.50	142.72	usa
11.06	22.19	49.19	1.89	3.87	8.45	151.22	ussr
12.74	25.85	58.73	2.33	5.81	13.04	306.00	wsamoa

;

Data men;

Group="Men";

Input x1-x8 country \$;

Label x1="100m" x2="200m" x3="400m" x4="800m" x5="1500m" x6="3000m"

x7="10000m" x8="Marathon";

Datalines;

10.39	20.81	46.84	1.81	3.70	14.04	29.36	137.72	argentin
10.31	20.06	44.84	1.74	3.57	13.28	27.66	128.30	australi
10.44	20.81	46.82	1.79	3.60	13.26	27.72	135.90	austria
10.34	20.68	45.04	1.73	3.60	13.22	27.45	129.95	belgium
10.28	20.58	45.91	1.80	3.75	14.68	30.55	146.62	bermuda
10.22	20.43	45.21	1.73	3.66	13.62	28.62	133.13	brazil
10.64	21.52	48.30	1.80	3.85	14.45	30.28	139.95	burma
10.17	20.22	45.68	1.76	3.63	13.55	28.09	130.15	canada
10.34	20.80	46.20	1.79	3.71	13.61	29.30	134.03	chile
10.51	21.04	47.30	1.81	3.73	13.90	29.13	133.53	china
10.43	21.05	46.10	1.82	3.74	13.49	27.88	131.35	columbia
12.18	23.20	52.94	2.02	4.24	16.70	35.38	164.70	cookis
10.94	21.90	48.66	1.87	3.84	14.03	28.81	136.58	costa
10.35	20.65	45.64	1.76	3.58	13.42	28.19	134.32	czech
10.56	20.52	45.89	1.78	3.61	13.50	28.11	130.78	denmark
10.14	20.65	46.80	1.82	3.82	14.91	31.45	154.12	domrep
10.43	20.69	45.49	1.74	3.61	13.27	27.52	130.87	finland
10.11	20.38	45.28	1.73	3.57	13.34	27.97	132.30	france
10.12	20.33	44.87	1.73	3.56	13.17	27.42	129.92	gdr
10.16	20.37	44.50	1.73	3.53	13.21	27.61	132.23	frg
10.11	20.21	44.93	1.70	3.51	13.01	27.51	129.13	gbni
10.22	20.71	46.56	1.78	3.64	14.59	28.45	134.60	greece
10.98	21.82	48.40	1.89	3.80	14.16	30.11	139.33	guatemal
10.26	20.62	46.02	1.77	3.62	13.49	28.44	132.58	hungary
10.60	21.42	45.73	1.76	3.73	13.77	28.81	131.98	india
10.59	21.49	47.80	1.84	3.92	14.73	30.79	148.83	indonesi
10.61	20.96	46.30	1.79	3.56	13.32	27.81	132.35	ireland
10.71	21.00	47.80	1.77	3.72	13.66	28.93	137.55	israel
10.01	19.72	45.26	1.73	3.60	13.23	27.52	131.08	italy
10.34	20.81	45.86	1.79	3.64	13.41	27.72	128.63	japan
10.46	20.66	44.92	1.73	3.55	13.10	27.38	129.75	kenya
10.34	20.89	46.90	1.79	3.77	13.96	29.23	136.25	korea
10.91	21.94	47.30	1.85	3.77	14.13	29.67	130.87	dprkorea
10.35	20.77	47.40	1.82	3.67	13.64	29.08	141.27	luxembou
10.40	20.92	46.30	1.82	3.80	14.64	31.01	154.10	malaysia
11.19	22.45	47.70	1.88	3.83	15.06	31.77	152.23	mauritiu
10.42	21.30	46.10	1.80	3.65	13.46	27.95	129.20	mexico
10.52	20.95	45.10	1.74	3.62	13.36	27.61	129.02	netherla
10.51	20.88	46.10	1.74	3.54	13.21	27.70	128.98	nz

10.55	21.16	46.71	1.76	3.62	13.34	27.69	131.48	norway
10.96	21.78	47.90	1.90	4.01	14.72	31.36	148.22	png
10.78	21.64	46.24	1.81	3.83	14.74	30.64	145.27	philippi
10.16	20.24	45.36	1.76	3.60	13.29	27.89	131.58	poland
10.53	21.17	46.70	1.79	3.62	13.13	27.38	128.65	portugal
10.41	20.98	45.87	1.76	3.64	13.25	27.67	132.50	rumania
10.38	21.28	47.40	1.88	3.89	15.11	31.32	157.77	singapor
10.42	20.77	45.98	1.76	3.55	13.31	27.73	131.57	spain
10.25	20.61	45.63	1.77	3.61	13.29	27.94	130.63	sweden
10.37	20.46	45.78	1.78	3.55	13.22	27.91	131.20	switzerl
10.59	21.29	46.80	1.79	3.77	14.07	30.07	139.27	taipei
10.39	21.09	47.91	1.83	3.84	15.23	32.56	149.90	thailand
10.71	21.43	47.60	1.79	3.67	13.56	28.58	131.50	turkey
9.93	19.75	43.86	1.73	3.53	13.20	27.43	128.22	usa
10.07	20.00	44.60	1.75	3.59	13.20	27.53	130.55	ussr
10.82	21.86	49.00	2.02	4.24	16.28	34.71	161.83	wsamoa

;