

Written examination, date: 6 December 2016

Page 1 of 17 pages Enclosure: 17 pages

Course name: Multivariate Statistics

Course number: 02409

Aids allowed: All

Exam duration: 4 hours

Weighting: The questions are given equal weight

This exam is answered by:

(name)

(signature)

(study no.)

There is a total of 30 questions for the 6 problems. The answers to the 30 questions must be written into the table below.

Problem	1	1	1	1	1	1	1	1	2	2
Question	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	2.1	2.2
Answer										

Problem	2	2	3	3	3	3	4	4	4	4
Question	2.3	2.4	3.1	3.2	3.3	3.4	4.1	4.2	4.3	4.4
Answer										

Problem	4	4	5	5	5	6	6	6	6	6
Question	4.5	4.6	5.1	5.2	5.3	6.1	6.2	6.3	6.4	6.5
Answer										

The possible answers for each question are numbered from 1 to 6. If you enter a wrong number, you may correct it by crossing the wrong number in the table and writing the correct answer immediately below. If there is any doubt about the meaning of a correction then the question will be considered not answered.

Only the front page must be returned. The front page must be returned even if you do not answer any of the questions or if you leave the exam prematurely. Drafts and/or comments are not considered, only the numbers entered above are registered.

A correct answer gives 5 points, a wrong answer gives – 1 point. Unanswered questions or a 6 (corresponding to “don’t know”) give 0 points. The total number of points needed for a satisfactorily answered exam is determined at the final evaluation of the exam. Especially note that the grade 10 may be given even if only one answer is wrong or unanswered.

Remember to write your name, signature, and study number on the front page.

Please note, that there is one and only one correct answer to each question. Furthermore, some of the possible alternative answers may not make sense. When the text refers to SAS-output, the values may be rounded to fewer decimal places than in the output itself. The enclosures do not necessarily contain all the output generated by the given SAS programs. Please check that all pages of the exam paper and the enclosures are present.

Problem 1.

Enclosure A with SAS program and SAS output belongs to this problem. The data are taken from Lichman, M. (2013): UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>], University of California Irvine, School of Information and Computer Science. The data give the results of chemical analyses of wines grown in the same region in Italy but derived from three different types of grapes. We only use a subset of the data corresponding to the variables

Type:	The variety (cultivar) of grape. A cultivar is a horticulturally derived variety of a plant, as distinguished from a natural variety.
Proline:	The concentration of proline, an amino acid that is used in the biosynthesis of proteins. In plants proline accumulation is a common physiological response to various stresses.
Flavan:	The concentration of flavonoids, a group of plant metabolites thought to provide health benefits through antioxidant effects. They are essential pigments for producing the colors needed to attract pollinating insects.
ColInt:	The color intensity of the wine.
Hue:	The color hue of the wine.

It is not necessary to be familiar with the biological or chemical information provided in the above in order to answer the following questions.

Initially we assume that the dispersion (variance-covariance) matrices for the three populations are the same.

We are firstly interested in, whether the populations given by the three types have different distributions.

Question 1.1.

The distribution of the usual multivariate test statistic for the hypotheses H_0 of no difference between the mean values for the three types will be (under H_0)

- 1 ☐ $U(2, 4, 175)$
- 2 ☐ $U(2, 2, 175)$
- 3 ☐ $U(4, 2, 175)$
- 4 ☐ $U(4, 2, 178)$
- 5 ☐ $U(2, 4, 178)$
- 6 ☐ Don't know.

The problem continues on the next page

Question 1.2.

Which distribution may we use in computing the critical values for the test statistic in the above question?

- 1 ☐ F(4, 344)
- 2 ☐ F(8, 346)
- 3 ☐ F(8, 348)
- 4 ☐ F(8, 344)
- 5 ☐ F(4, 173)
- 6 ☐ Don't know.

Question 1.3.

If we test whether the mean values for ColInt are the same for the three types, then the usual, one-dimensional test statistic becomes

- 1 ☐ 165.69
- 2 ☐ $\frac{551.4/2}{399.9/175}$
- 3 ☐ 5.82664
- 4 ☐ $\frac{4.29 \times 48.07 \times 5198844.33}{399.9}$
- 5 ☐ $\frac{551.4}{399.9 + 551.4}$
- 6 ☐ Don't know.

Question 1.4.

We are now just looking at Type2 and Type3. Proc Discrim computes the squared Mahalanobis distance between the two types. The F-test statistic corresponding to this squared Mahalanobis distance between the two types using all four variables is

- 1 ☐ $\frac{71+48-5}{4 \times (71+48-2)} \frac{71 \times 48}{(71+48)} 25.83438$
- 2 ☐ $\frac{1-19.02435^{0.2}}{19.02435^{0.2}}$
- 3 ☐ $\frac{19.02435/3}{25.83438/12}$
- 4 ☐ $0.1949 \times 28.6387 \times 9.57363$
- 5 ☐ $\frac{71+48-5}{4 \times (71+48-2)} \frac{71 \times 48}{(71+48)} 19.02435$
- 6 ☐ Don't know.

The problem continues on the next page

Question 1.5.

We still only consider Type2 and Type3. We want to investigate the hypothesis that if we use Collnt and Hue in the discrimination between the two types, the variables Flavan and Proline will not yield any further information. The F-statistic corresponding to this hypothesis is

1 ☐ $\frac{71+48-5}{2} \frac{71 \times 48 \times (25.83438 - 9.57363)}{(71+48) \times (71+46) + 71 \times 48 \times 9.57363}$

2 ☐ $\frac{71+48-5}{5 \times (71+48-2)} \frac{71 \times 48}{(71+48)} 25.83438$

3 ☐ $0.1949 \times 28.6387 \times 25.83438$

4 ☐ $\frac{71 \times 48}{(71+48)} \frac{(19.02435 - 9.57363)}{71 \times 48 \times 9.57363}$

5 ☐ $\frac{71+48-5}{2} \frac{71 \times 48 \times (19.02435 - 9.57363)}{(71+48) \times (71+46) + 71 \times 48 \times 9.57363}$

6 ☐ Don't know.

Question 1.6.

We are still only considering Type2 and Type3. Based on measurements on the color related variables Collnt and Hue, we want to classify observation no 142. The observation is classified as Type2 if

1 ☐ $-23.29 + [5.6 \quad 0.7] \begin{bmatrix} 3.50 \\ 30.26 \end{bmatrix} < 0$

2 ☐ $-24.27 + [5.6 \quad 0.7] \begin{bmatrix} 2.11 \\ 39.80 \end{bmatrix} < 0$

3 ☐ $(-24.27 + 23.29) + [5.6 \quad 0.7] \begin{bmatrix} 2.11 - 3.50 \\ 39.80 - 30.26 \end{bmatrix} > 0$

4 ☐ $(-24.27 + 23.29) + [5.6 \quad 0.7] \begin{bmatrix} 2.11 - 3.50 \\ 39.80 - 30.26 \end{bmatrix} < 0$

5 ☐ $-24.27 + [5.6 \quad 0.7] \begin{bmatrix} 2.11 \\ 39.80 \end{bmatrix} > 0$

6 ☐ Don't know.

The problem continues on the next page

Question 1.7.

We now want to find a Bayes solution assuming that the prior probability of Type2 is twice as large as the prior probability of Type3. In the above expressions, > 0 or < 0 should be replaced with $> d$ or $< d$, where d equals

- 1 ☐ -1.0986
- 2 ☐ -0.6931
- 3 ☐ -0.4055
- 4 ☐ -0.2877
- 5 ☐ -0.2231
- 6 ☐ Don't know.

Question 1.8.

We now return to the case with all types and all variables. We consider the number of misclassified observations using cross validation. If we go from a linear discriminant analysis to a quadratic discriminant analysis, the number of misclassified observations is reduced with

- 1 ☐ 1
- 2 ☐ 2
- 3 ☐ 3
- 4 ☐ 4
- 5 ☐ 5
- 6 ☐ Don't know.

Problem 2.

Consider a random variable

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

with dispersion (variance-covariance) matrix

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}.$$

Question 2.1.

The partial correlation $\rho_{24|3}$ is

- 1 ☐ 0
- 2 ☐ ρ
- 3 ☐ ρ^2
- 4 ☐ $1 - \rho^2$
- 5 ☐ $\rho/\sqrt{1 + \rho^2}$
- 6 ☐ Don't know.

Question 2.2.

The squared multiple correlation $\rho_{1|23}^2$ is

- 1 ☐ 0
- 2 ☐ ρ
- 3 ☐ ρ^2
- 4 ☐ $1 - \rho^2$
- 5 ☐ $\rho/\sqrt{1 + \rho^2}$
- 6 ☐ Don't know.

The problem continues on the next page

Question 2.3.

The squared multiple correlation $\rho_{2|34}^2$ is

- 1 ☐ 0
- 2 ☐ ρ
- 3 ☐ ρ^2
- 4 ☐ $1 - \rho^2$
- 5 ☐ $\rho/\sqrt{1 + \rho^2}$
- 6 ☐ Don't know.

Question 2.4.

The partial correlation $\rho_{12|34}$ is (*hint: it is suggested that you calculate $\rho_{12|34}$ directly in the conditional distribution rather than using the usual formula*).

- 1 ☐ 0
- 2 ☐ ρ
- 3 ☐ ρ^2
- 4 ☐ $1 - \rho^2$
- 5 ☐ $\rho/\sqrt{1 + \rho^2}$
- 6 ☐ Don't know.

Problem 3.

For a person undergoing a training program, the strength was measured every other day for two weeks. The data obtained were (the days were coded from -3 to 3 , and 90 was subtracted from the measured strengths). The data are taken from a larger data set presented in Littell, Freund & Spector (1991): SAS System for Linear Models, SAS Institute INC.

Time	Strength
-3	-3
-2	-1
-1	1
0	0
1	1
2	2
3	2

We assume that the observed strength data follow a standard general linear model of the form

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \end{bmatrix}$$

where the $\varepsilon_1, \dots, \varepsilon_7$ are assumed to be normally distributed with mean 0 and common variance σ^2 . On matrix form we get

$$Y = X\theta + \varepsilon.$$

In the sequel you may use that

$$(Y - X\hat{\theta})'(Y - X\hat{\theta}) = 3 \frac{19}{28}$$

where $\hat{\theta} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$ is the maximum likelihood estimator for θ .

The problem continues on the next page

Question 3.1.

The value of \hat{a} is

1 ☐ $\frac{2}{7}$

2 ☐ $\frac{1}{4}$

3 ☐ $\frac{1}{7}$

4 ☐ $\frac{3}{4}$

5 ☐ $\frac{1}{28}$

6 ☐ Don't know.

Question 3.2.

The value of $\hat{\beta}$ is

1 ☐ $\frac{2}{7}$

2 ☐ $\frac{1}{4}$

3 ☐ $\frac{1}{7}$

4 ☐ $\frac{3}{4}$

5 ☐ $\frac{1}{28}$

6 ☐ Don't know.

The problem continues on the next page

Question 3.3.

The mean predicted value at time = 0 is $[1 \ 0] \hat{\theta}$. The estimated variance of this quantity is

1 ☐ $\sqrt{3\frac{19}{28}}$

2 ☐ $19.42857/5$

3 ☐ $\frac{2}{7}$

4 ☐ $\frac{1}{5} \times 3\frac{19}{28}$

5 ☐ $\frac{1}{35} \times 3\frac{19}{28}$

6 ☐ Don't know.

Question 3.4.

If we want to determine a confidence interval for the mean predicted value given in Question 3.3, we must find quantiles in an appropriate t-distribution. The degrees of freedom for this t-distribution are

1 ☐ 3

2 ☐ 4

3 ☐ 5

4 ☐ 6

5 ☐ 7

6 ☐ Don't know.

Problem 4.

The data here are sub-sampled digital images from the Thematic Mapper (TM) scanner onboard the Landsat series of satellites. We have data covering the greater Copenhagen area from two time points, 27 June 1986 and 18 August 2005. The data are geometrically co-registered so there is time-wise pixel-correspondance, i.e., pixel (i, j) in 1986 represents the same area on the ground as pixel (i, j) in 2005. The variables are digital numbers representing reflected sunlight from the surface of the Earth measured in six wavelength bands, namely visible blue, green and red, and three infrared bands.

Enclosure B gives SAS code and output from a principal component analysis on twelve variables, namely the six spectral bands from time point 1, 1985 (variables are named X1-X6), and the six spectral bands from time point 2, 2005 (variables are named Y1-Y6).

The data were originally downloaded from saccess.dk.

Question 4.1.

The variance of the first principal component is

- 1 ☐ 82.0487²
- 2 ☐ 1.0000
- 3 ☐ 9.0163
- 4 ☐ 0.7514
- 5 ☐ 7.3863
- 6 ☐ Don't know.

Question 4.2.

If we want to retain at least 95 % of the total variation in the original 12 variables, we need to retain how many principal components?

- 1 ☐ 1
- 2 ☐ 2
- 3 ☐ 3
- 4 ☐ 4
- 5 ☐ 11
- 6 ☐ Don't know.

The problem continues on the next page

Question 4.3.

A good detector of change over time will measure the contrast between all variables in the 1986 and the 2005 data. Which principal component is the best change detector?

- 1 ☐ 1
- 2 ☐ 2
- 3 ☐ 3
- 4 ☐ 4
- 5 ☐ 11
- 6 ☐ Don't know.

Question 4.4.

Which principal component best reflects (is proportional to) the overall mean of all 12 original variables?

- 1 ☐ 1
- 2 ☐ 2
- 3 ☐ 3
- 4 ☐ 4
- 5 ☐ 11
- 6 ☐ Don't know.

Question 4.5.

What is the test statistic for testing whether the eight smallest eigenvalues are equal?

- 1 ☐ $9.0163 \times 1.6300 \times 0.8746 \times 0.1491$
- 2 ☐ 3954
- 3 ☐ 2944
- 4 ☐ 9016
- 5 ☐ 2923
- 6 ☐ Don't know.

The problem continues on the next page

Question 4.6.

What is the number of degrees of freedom associated with the test statistic for testing whether the eight smallest eigenvalues are equal?

- 1 ☐ 12
- 2 ☐ 88
- 3 ☐ 44
- 4 ☐ 70
- 5 ☐ 35
- 6 ☐ Don't know.

Problem 5.

The data here are the same as the data used in Problem 4. Enclosure C gives SAS code and (some of the) output from a canonical correlation analysis where one set of variables is the six spectral bands from 1985 (variables are named X1-X6), and the other set is the six spectral bands from 2005 (variables are named Y1-Y6).

Question 5.1.

The largest (also known as the leading) canonical correlation between the two data sets is

- 1 ☐ 0.9100
- 2 ☐ 0.9041
- 3 ☐ 0.9540
- 4 ☐ 1
- 5 ☐ 96.0150 %
- 6 ☐ Don't know.

Question 5.2.

How many canonical correlations are significantly different from zero?

- 1 ☐ 6
- 2 ☐ 5
- 3 ☐ 4
- 4 ☐ 3
- 5 ☐ 2
- 6 ☐ Don't know.

Question 5.3.

The canonical variates corresponding to the second largest canonical correlation from both time points are calculated largely as

- 1 ☐ The mean level of all variables.
- 2 ☐ A contrast between the first variable and the remaining variables.
- 3 ☐ A contrast between the first two variables and the remaining variables.
- 4 ☐ A contrast between the first three variables and the remaining variables.
- 5 ☐ A contrast between the first four variables and the remaining variables.
- 6 ☐ Don't know.

Problem 6.

Enclosure D with SAS program and SAS output belongs to this problem. The data are taken from “A.N. Christensen et al, Injectable silver nanosensors: in vivo dosimetry for external beam radio-therapy using positron emission tomography, *Nanoscale*, 2016, 8, 11002-11011”. The data are:

FBP: PET-scan measurements of radioactivity of silver in gels inside a phantom, Y
scan: scan number,
position: position of the gel inside the phantom,
gel: the number of the gel,
corrDose: radiation dose given to the silver in the phantom, $corrDose_l$

It is not necessary to understand the physical properties to answer the questions.

We model the FBP as a function of the other variables using a GLM and have the following model and hypothesis:

$$M: E(Y_{ijkl}) = \mu + scan_i + position_j + gel_k + \beta \times corrDose_l$$

$$H: E(Y_{ijkl}) = \tau + \gamma \times corrDose_l,$$

where $scan_i$, $position_j$, and gel_k are suitable parameters describing the effects from the different scans, positions and gels, μ and τ are intercept parameters, and finally β and γ are slope parameters describing the dependence on $corrDose$.

Question 6.1.

What is the usual test-statistic for M vs H:

- 1 ☐ $\frac{(0.99336-0.982407)/11}{0.982407/43}$
- 2 ☐ $\frac{(0.36471-0.36071)/(12-1)}{0.36071/1}$
- 3 ☐ $\frac{(0.00644-0.00214)/(5-1)}{0.00244/(12-1)}$
- 4 ☐ $\frac{(0.00644-0.00244)/(43-32)}{0.00244/32}$
- 5 ☐ $\frac{(0.00644-0.00244)/(43-32)}{0.00644/(45-2)}$
- ☐ Don't know.

The problem continues on the next page

Question 6.2.

The distribution of the statistic above under the null hypothesis is:

- 1 ☐ $F(11, 44)$
- 2 ☐ $F(4, 11)$
- 3 ☐ $F(11, 43)$
- 4 ☐ $F(11, 32)$
- 5 ☐ $F(4, 45)$
- 6 ☐ Don't know.

Question 6.3.

Assume that the hypothesis H is true. Then the 95% confidence interval for γ is equal to

- 1 ☐ $0.0110837 \pm t(43)_{0.975} \times 0.0002258$
- 2 ☐ $0.0110837 \pm t(43)_{0.975} \times \sqrt{0.36715474}$
- 3 ☐ $0.0110837 \pm t(43)_{0.975} \times \sqrt{1 + 0.0002258}$
- 4 ☐ $0.0110837 \pm t(32)_{0.975} \times 0.002437160$
- 5 ☐ $0.0110837 \pm t(43)_{0.975} \times 0.00014971$
- 6 ☐ Don't know.

Question 6.4.

We consider model M and assume that the studentized residuals follow a standardized normal distribution. For how many of the studentized residuals is the probability of getting an outcome whose absolute value is more extreme than the actual one, smaller than 5%:

- 1 ☐ 1
- 2 ☐ 2
- 3 ☐ 3
- 4 ☐ 4
- 5 ☐ 5
- 6 ☐ Don't know.

The problem continues on the next page

Question 6.5.

Considering model M, which observation has the largest overall influence on the parameter estimates:

- 1 ☐ 26
- 2 ☐ 31
- 3 ☐ 39
- 4 ☐ 40
- 5 ☐ 44
- 6 ☐ Don't know.

Enclosure A – SAS Program

```
data wine; set sasuser.wine; run; Title 'Wine dataset'; proc print data=wine; run;

Title 'GLM Analysis on Wine Data Set';
proc glm data=wine plots=none;
class type;
model Collnt Hue Flavan Proline=type/nouni;
manova h=_all_ / printe printh;
run;

Title 'Discriminant Analysis on Wine Data Set, three classes';
Title2 'pool=yes';
proc discrim data=wine method=normal pool=yes crossvalidate;
class type;
var Collnt Hue Flavan Proline;
run;

Title2 'pool=test';
proc discrim data=wine method=normal pool=test crossvalidate;
class Type;
var Collnt Hue Flavan Proline;
run;

Title 'Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set';
data wine23;
set wine;
if type = 'Type2' or type = 'Type3';
run;

Title2 'Variables used are Collnt, Hue, Flavan, Proline';
proc discrim data=wine23 method=normal pool=yes crossvalidate;
class Type;
var Collnt Hue Flavan Proline;
run;

Title2 'Variables used are Collnt, Hue';
proc discrim data=wine23 method=normal pool=yes crossvalidate;
class Type;
var Collnt Hue;
run;
```

Enclosure A – SAS output

Wine dataset

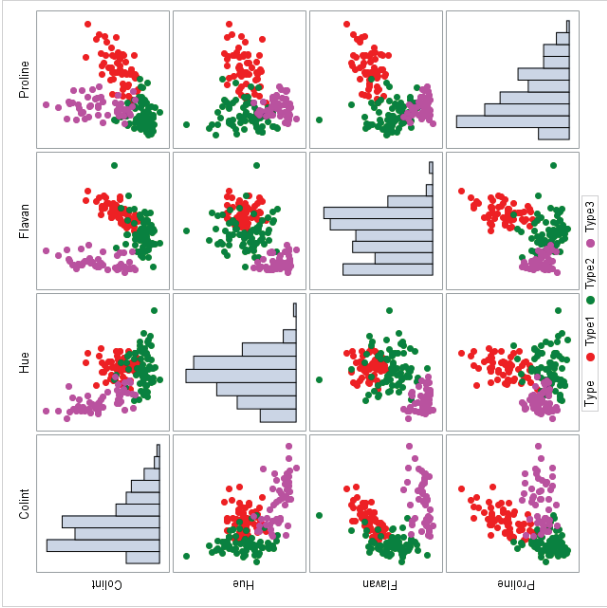
Obs	Type	Collnt	Hue	Flavan	Proline
1	Type1	5.6400	1.040	3.06	1065
2	Type1	4.3800	1.050	2.76	1050
3	Type1	5.6800	1.030	3.24	1185
4	Type1	7.8000	0.860	3.49	1480
5	Type1	4.3200	1.040	2.69	735
6	Type1	6.7500	1.050	3.39	1450
7	Type1	5.2500	1.020	2.52	1290
8	Type1	5.0500	1.060	2.51	1295
9	Type1	5.2000	1.080	2.98	1045
10	Type1	7.2200	1.010	3.15	1045
11	Type1	5.7500	1.250	3.32	1510
12	Type1	5.0000	1.170	2.43	1280
60	Type2	1.9500	1.050	0.57	520
61	Type2	3.2700	1.250	1.09	680
62	Type2	5.7500	0.980	1.41	450
63	Type2	3.8000	1.230	1.79	630
64	Type2	4.4500	1.220	3.10	420
65	Type2	2.9500	1.450	1.75	355
66	Type2	4.6000	1.190	2.65	678
67	Type2	5.3000	1.120	3.18	502
68	Type2	4.6800	1.120	2.00	510
69	Type2	3.1700	1.020	1.30	750
70	Type2	2.8500	1.280	1.28	718
71	Type2	3.0500	0.906	1.02	870
131	Type3	4.1000	0.760	1.25	630
132	Type3	5.4000	0.740	1.22	530
133	Type3	5.7000	0.660	1.09	560
134	Type3	5.0000	0.780	1.20	600
135	Type3	5.4500	0.750	0.58	650
136	Type3	7.1000	0.730	0.66	695
137	Type3	3.8500	0.750	0.47	720
138	Type3	5.0000	0.820	0.60	515
139	Type3	5.7000	0.810	0.48	580
140	Type3	4.9200	0.890	0.60	590
141	Type3	4.6000	0.770	0.50	600
142	Type3	5.6000	0.700	0.50	780

The first 12 observations from each type are shown

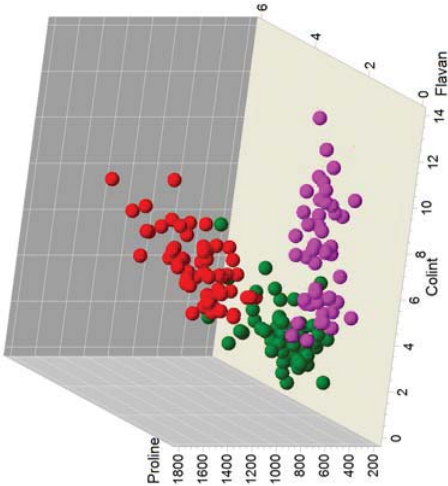
The SAS program for generating this output is not shown, and the programs for generating the scatter plots on the next page are not shown.

Enclosure A – SAS output

Scatter Plots of Wine Data Set



3D-Scatter Plot of Wine Data Set



Enclosure A – SAS output

GLM Analysis on Wine Data Set

The GLM Procedure
Multivariate Analysis of Variance

Class Level Information		
Class Levels Values		
Type	3	Type1 Type2 Type3

Number of Observations Read	178
Number of Observations Used	178

E = Error SSCP Matrix				
	Colint	Hue	Flavan	Proline
Colint	399.86153923	-7.17434139250	168.70287611916	394286
Hue	-7.1743413924	2.853382151	-0.736111231	785.92797709
Flavan	50.168702876	-0.736111231	148.073815009599	49720811
Proline	11916.394286785	92.797709599	49720811	5198844.3273

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > r				
DF = 175	Colint	Hue	Flavan	Proline
Colint	1.000000	-0.173314	0.361846	0.261359
Hue	-0.173314	0.0214	<.0001	0.0005
Flavan	0.361846	-0.051286	0.4991	0.166509
Proline	0.261359	0.166509	0.037921	1.000000

H = Type III SSCP Matrix for Type				
	Colint	Hue	Flavan	Proline
Colint	551.41600164	-41.76718541	-120.821549728929	449698
Hue	-41.76718541	4.962019829922	6986196572223	1115622
Flavan	-120.821549722	698619657128	5223900526914	708915
Proline	28929.4496982223	111562226914	70891512353664	645

Enclosure A – SAS output

GLM Analysis on Wine Data Set

The GLM Procedure

Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where H = Type III SSCP Matrix for Type E = Error SSCP Matrix					
Characteristic Roots		Characteristic Vector V'E=1			
Percent		Collint	Hue	Flavan	Proline
6.33452716	74.12	-0.03239428	0.12582573	0.12634616	0.00019837
2.21152965	25.88	0.02148890	-0.14503160	-0.01085625	0.00034193
0.00000000	0.00	0.03204651	0.46288477	-0.03395846	-0.00008436
0.00000000	0.00	0.02681555	-0.04696832	0.08254642	-0.00023419

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall Type Effect
H = Type III SSCP Matrix for Type
E = Error SSCP Matrix

S=2 M=0.5 N=85						
Statistic	Value	F Value	Num DF	Den DF	Pr > F	
Wilks' Lambda	0.04245374	165.69	8	344	<.0001	
Pillai's Trace	1.55228046	149.95	8	346	<.0001	
Hotelling-Lawley Trace	8.54605681	183.11	8	243.4	<.0001	
Roy's Greatest Root	6.33452716	273.97	4	173	<.0001	
NOTE: F Statistic for Roy's Greatest Root is an upper bound.						
NOTE: F Statistic for Wilks' Lambda is exact.						

Enclosure A – SAS output

Discriminant Analysis on Wine Data Set, three classes

pool=yes

The DISCRIM Procedure

Total Sample Size	178	DF Total	177
Variables	4	DF Within Classes	175
Classes	3	DF Between Classes	2

Number of Observations Read	178
Number of Observations Used	178

Class Level Information				
Variable Type	Name	Frequency	Weight	Prior Probability
Type1	Type1	59	59.0000	0.331461
Type2	Type2	71	71.0000	0.398876
Type3	Type3	48	48.0000	0.269663

Pooled Covariance Matrix Information		
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix	
4	5.82664	

Generalized Squared Distance to Type			
From Type	Type1	Type2	Type3
Type1	0	14.70495	41.16067
Type2	14.70495	0	5.83438
Type3	41.16067	5.83438	0

Linear Discriminant Function for Type			
Variable	Type1	Type2	Type3
Constant	-55.56174	-34.54157	-26.87120
Collint	1.09267	1.04056	3.65203
Hue	41.86352	44.71313	32.42958
Flavan	10.01349	7.07965	-0.56962
Proline	0.02757	0.00753	0.00800

Enclosure A – SAS output

Discriminant Analysis on Wine Data Set, three classes
pool=yes

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WINE
Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type1	Type2	Type3	Total
Type1	56	3	0	59
	94.92	5.08	0.00	100.00
Type2	5	65	1	71
	7.04	91.55	1.41	100.00
Type3	0	1	47	48
	0.00	2.08	97.92	100.00
Total	61	69	48	178
	34.27	38.76	26.97	100.00
Priors	0.33333	0.33333	0.33333	

Error Count Estimates for Type				
	Type1	Type2	Type3	Total
Rate	0.05080	0.08450	0.02080	0.0521
Priors	0.33330	0.33330	0.3333	

Classification Summary for Calibration Data: WORK.WINE
Cross-validation Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type1	Type2	Type3	Total
Type1	56	3	0	59
	94.92	5.08	0.00	100.00
Type2	5	65	1	71
	7.04	91.55	1.41	100.00
Type3	0	1	47	48
	0.00	2.08	97.92	100.00
Total	61	69	48	178
	34.27	38.76	26.97	100.00
Priors	0.33333	0.33333	0.33333	

Enclosure A – SAS output

Discriminant Analysis on Wine Data Set, three classes
pool=yes

Error Count Estimates for Type				
	Type1	Type2	Type3	Total
Rate	0.0508	0.0845	0.0208	0.0521
Priors	0.3333	0.3333	0.3333	

Discriminant Analysis on Wine Data Set, three classes
pool=test

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
215.376877	20	<.0001

Since the Chi-Square value is significant at the 0.1 level, the within covariance matrices will be used in the discriminant function.
Reference: Morrison, D.F. (1976) Multivariate Statistical Methods p252.

Classification Summary for Calibration Data: WORK.WINE
Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type1	Type2	Type3	Total
Type1	59	0	0	59
	100.00	0.00	0.00	100.00
Type2	3	68	0	71
	4.23	95.77	0.00	100.00
Type3	0	0	48	48
	0.00	0.00	100.00	100.00
Total	62	68	48	178
	34.83	38.20	26.97	100.00
Priors	0.33333	0.33333	0.33333	

Error Count Estimates for Type				
	Type1	Type2	Type3	Total
Rate	0.00000	0.04230	0.00000	0.0141
Priors	0.33330	0.33330	0.3333	

Enclosure A – SAS output

Discriminant Analysis on Wine Data Set, three classes
pool=test

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WINE
Cross-validation Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type1	Type2	Type3	Total
Type1	58	1	0	59
	98.31	1.69	0.00	100.00
Type2	3	67	1	71
	4.23	94.37	1.41	100.00
Type3	0	1	47	48
	0.00	2.08	97.92	100.00
Total	61	69	48	178
	34.27	38.76	26.97	100.00
Priors	0.33333	0.33333	0.33333	

Error Count Estimates for Type				
	Type1	Type2	Type3	Total
Rate	0.0169	0.0563	0.0208	0.0314
Priors	0.3333	0.3333	0.3333	

Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set
Variables used are Collnt, Hue, Flavan, Proline

The DISCRIM Procedure

Generalized Squared Distance to Type			
From Type	Type	Type2	Type3
Type2		0	19.02435
Type3		19.02435	0

Linear Discriminant Function for Type			
Variable	Type2	Type3	
Constant	-35.26743	-29.64616	
Collnt	1.21230	3.09822	
Hue	37.60283	27.82026	
Flavan	6.92688	1.47243	
Proline	0.02437	0.02577	

Enclosure A – SAS output

Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set
Variables used are Collnt, Hue, Flavan, Proline

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WINE23
Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type2	Type3	Total	
Type2	71	0	71	
	100.00	0.00	100.00	
Type3	1	47	48	
	2.08	97.92	100.00	
Total	72	47	119	
	60.50	39.50	100.00	
Priors	0.5	0.5		

Error Count Estimates for Type			
Type2	Type3	Total	
Rate	0.00000	0.02080	0.0104
Priors	0.50000	0.50000	

Classification Summary for Calibration Data: WORK.WINE23
Cross-validation Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type2	Type3	Total	
Type2	70	1	71	
	98.59	1.41	100.00	
Type3	1	47	48	
	2.08	97.92	100.00	
Total	71	48	119	
	59.66	40.34	100.00	
Priors	0.5	0.5		

Error Count Estimates for Type			
Type2	Type3	Total	
Rate	0.0141	0.02080	0.0175
Priors	0.50000	0.50000	

Enclosure A – SAS output

Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set
Variables used are Collnt, Hue

The DISCRIM Procedure

Generalized Squared Distance to Type			
From Type	Type2	Type3	
Type2	0	9.57363	
Type3	9.57363	0	

Linear Discriminant Function for Type		
Variable	Type2	Type3
Constant	-24.27489	-23.28922
Collnt	2.11025	3.50475
Hue	39.79642	30.25656

Classification Summary for Calibration Data: WORK.WINE23
Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type			
From Type	Type2	Type3	Total
Type2	70	1	71
	98.59	1.41	100.00
Type3	5	43	48
	10.42	89.58	100.00
Total	75	44	119
	63.03	36.97	100.00
Priors	0.5	0.5	

Error Count Estimates for Type			
	Type2	Type3	Total
Rate	0.0141	0.1042	0.0591
Priors	0.5000	0.5000	

Enclosure A – SAS output

Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set
Variables used are Collnt, Hue

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.WINE23
Cross-validation Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type			
From Type	Type2	Type3	Total
Type2	69	2	71
	97.18	2.82	100.00
Type3	7	41	48
	14.58	85.42	100.00
Total	76	43	119
	63.87	36.13	100.00
Priors	0.5	0.5	

Error Count Estimates for Type			
	Type2	Type3	Total
Rate	0.0282	0.1458	0.0870
Priors	0.5000	0.5000	

Enclosure B – SAS Program

```
data cph;
input X1-X6 Y1-Y6;
cards;
128 112 96 122 122 85 129 113 98 127 129 89
107 126 202 195 174 173 103 132 201 195 200 180
9 13 31 55 50 47 2 2 20 23 39 42
(more data lines)
;

title 'Landsat TM data over Copenhagen, subsample';
title2 'Data from saccess.dk, 1986-06-27 and 2005-08-18';
proc princomp data=cph;
var X1-X6 Y1-Y6;
run;
```

Enclosure B – SAS output

Landsat TM data over Copenhagen, subsample
Data from saccess.dk, 1986-06-27 and 2005-08-18
The PRINCOMP Procedure

Observations	1000
Variables	12

Simple Statistics						
	X1	X2	X3	X4	X5	X6
Mean	96.01500000	80.33700000	73.66700000	96.34200000	97.30300000	71.56100000
Std	82.04868531	66.19848692	52.57800450	46.63414822	44.45651328	42.62040330

Simple Statistics						
	Y1	Y2	Y3	Y4	Y5	Y6
Mean	75.04500000	90.36800000	86.82300000	99.60900000	102.3920000	82.66900000
Std	65.71873345	79.26296230	60.24622211	60.61529785	47.7888702	42.25011848

Correlation Matrix												
	X1	X2	X3	X4	X5	X6	Y1	Y2	Y3	Y4	Y5	Y6
X1	1.0000	0.7926	0.4086	0.5211	0.6266	0.3222	0.8308	0.8976	0.7314	0.7821	0.7637	0.5916
X2	0.7926	1.0000	0.7860	0.8376	0.8673	0.7032	0.8505	0.7958	0.6897	0.7625	0.7730	0.6458
X3	0.4086	0.7860	1.0000	0.9841	0.9366	0.9540	0.5439	0.4941	0.5643	0.6290	0.6570	0.6275
X4	0.5211	0.8376	0.9841	1.0000	0.9715	0.9279	0.6288	0.5845	0.6276	0.6991	0.7236	0.6642
X5	0.6266	0.8673	0.9366	0.9715	1.0000	0.8873	0.6897	0.6647	0.6803	0.7523	0.7821	0.6948
X6	0.3222	0.7032	0.9540	0.9279	0.8873	1.0000	0.4436	0.4119	0.5193	0.5713	0.6008	0.6063
Y1	0.8308	0.8505	0.5439	0.6288	0.6897	0.4436	1.0000	0.8128	0.5923	0.6693	0.7076	0.5215
Y2	0.8976	0.7958	0.4941	0.5845	0.6647	0.4119	0.8128	1.0000	0.8743	0.8953	0.8776	0.7473
Y3	0.7314	0.6897	0.5643	0.6276	0.6803	0.5193	0.5923	0.8743	1.0000	0.9830	0.9470	0.9334
Y4	0.7821	0.7625	0.6290	0.6991	0.7523	0.5713	0.6693	0.8953	0.9830	1.0000	0.9684	0.9205
Y5	0.7637	0.7730	0.6570	0.7236	0.7821	0.6008	0.7076	0.8776	0.9470	0.9684	1.0000	0.9246
Y6	0.5916	0.6458	0.6275	0.6642	0.6948	0.6063	0.5215	0.7473	0.9334	0.9205	0.9246	1.0000

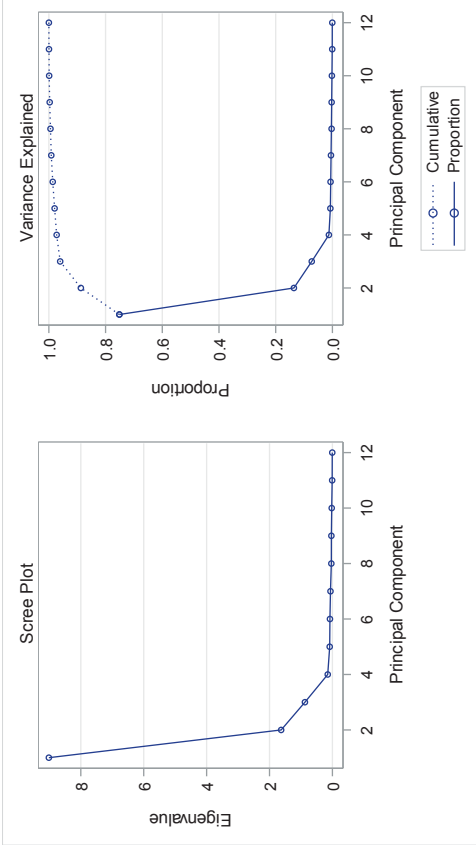
Enclosure B – SAS output

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	9.01626784	7.38629813	0.7514	0.7514
2	1.62996970	0.75537952	0.1358	0.8872
3	0.87459018	0.72547429	0.0729	0.9601
4	0.14911589	0.06258133	0.0124	0.9725
5	0.08653455	0.00778286	0.0072	0.9797
6	0.07875169	0.01712586	0.0066	0.9863
7	0.06162583	0.02716816	0.0051	0.9914
8	0.03445767	0.00126063	0.0029	0.9943
9	0.03319704	0.01084316	0.0028	0.9970
10	0.02235388	0.01438586	0.0019	0.9989
11	0.00796802	0.00280031	0.0007	0.9996
12	0.00516771		0.0004	1.0000

Eigenvalues							
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
X1	0.265382	-.344758	0.341909	-.542694	0.225310	0.417245	0.262426
X2	0.304236	0.048394	0.337080	-.044573	-.761246	0.263464	-.249231
X3	0.274342	0.434404	0.000887	-.001131	-.037508	-.140833	-.208649
X4	0.293249	0.356342	0.052250	-.070095	0.164756	-.083087	-.294205
X5	0.305738	0.264857	0.089019	-.182918	0.413683	0.072865	-.178829
X6	0.253710	0.477851	-.088451	-.036464	-.098089	-.056597	0.767310
Y1	0.265175	-.157106	0.531964	0.690452	0.195213	-.045306	0.112292
Y2	0.291196	-.328656	0.100210	-.109384	-.140987	-.682585	0.168448
Y3	0.294705	-.235797	-.354534	-.104648	-.119393	-.184028	-.044614
Y4	0.310159	-.197091	-.247105	-.107818	0.015822	-.089014	-.167410
Y5	0.312982	-.161072	-.213937	0.187527	0.272425	0.088748	-.153441
Y6	0.285988	-.107409	-.474810	0.342581	-.117167	0.451223	0.153737

Enclosure B – SAS output

Eigenvectors					
	Prin8	Prin9	Prin10	Prin11	Prin12
X1	0.201407	-.024140	0.241094	0.072036	0.033266
X2	-.257100	0.042231	-.090476	-.003678	-.048798
X3	0.298727	-.101839	0.413721	0.220889	0.593032
X4	0.272541	-.059147	0.181582	-.170100	-.720482
X5	-.250859	-.207321	-.667211	0.058482	0.179070
X6	-.179396	0.237796	0.002316	-.059657	-.065209
Y1	0.192174	0.207850	-.099236	0.043122	0.026085
Y2	-.096008	-.498470	0.016830	-.113453	-.012796
Y3	0.229510	0.356028	-.241625	0.639643	-.154384
Y4	0.114894	0.468531	-.075725	-.675043	0.248861
Y5	-.678689	0.070818	0.450246	0.135731	-.052431
Y6	0.249124	-.491192	-.086058	-.110056	-.007234



Enclosure C – SAS Program

```
data cph;
input X1-X6 Y1-Y6;
cards;
128 112 96 122 122 85 129 113 98 127 129 89
107 126 202 195 174 173 103 132 201 195 200 180
9 13 31 55 50 47 2 2 20 23 39 42
(more data lines)
;

proc cancorr data=cph all
vprefix=CV1986 _ vname='1986-06-27 data'
wprefix=CV2005_ wname='2005-08-18 data';
var X1-X6;
with Y1-Y6;
title 'Landsat TM data over Copenhagen, subsample';
title2 'Data from saccess.dk, 1986-06-27 and 2005-08-18';
run;
```

Some of the outputs have been omitted.

Enclosure C – SAS output

*Landsat TM data over Copenhagen, subsample
Data from saccess.dk, 1986-06-27 and 2005-08-18
The CANCORR Procedure*

1986-06-27 data	6
2005-08-18 data	6
Observations	1000

Means and Standard Deviations		
Variable	Mean	Standard Deviation
X1	96.015000	82.048685
X2	80.337000	66.198487
X3	73.667000	52.578005
X4	96.342000	46.634148
X5	97.303000	44.456513
X6	71.561000	42.620403
Y1	75.045000	65.718733
Y2	90.368000	79.262962
Y3	86.823000	60.246222
Y4	99.609000	60.615298
Y5	102.392000	47.788870
Y6	82.669000	42.250118

Enclosure C – SAS output

Correlations Among the Original Variables

Correlations Among the 1986-06-27 data						
	X1	X2	X3	X4	X5	X6
X1	1.0000	0.7926	0.4086	0.5211	0.6266	0.3222
X2	0.7926	1.0000	0.7860	0.8376	0.8673	0.7032
X3	0.4086	0.7860	1.0000	0.9841	0.9366	0.9540
X4	0.5211	0.8376	0.9841	1.0000	0.9715	0.9279
X5	0.6266	0.8673	0.9366	0.9715	1.0000	0.8873
X6	0.3222	0.7032	0.9540	0.9279	0.8873	1.0000

Correlations Among the 2005-08-18 data						
	Y1	Y2	Y3	Y4	Y5	Y6
Y1	1.0000	0.8128	0.5923	0.6693	0.7076	0.5215
Y2	0.8128	1.0000	0.8743	0.8953	0.8776	0.7473
Y3	0.5923	0.8743	1.0000	0.9830	0.9470	0.9334
Y4	0.6693	0.8953	0.9830	1.0000	0.9684	0.9205
Y5	0.7076	0.8776	0.9470	0.9684	1.0000	0.9246
Y6	0.5215	0.7473	0.9334	0.9205	0.9246	1.0000

Correlations Between the 1986-06-27 data and the 2005-08-18 data						
	Y1	Y2	Y3	Y4	Y5	Y6
X1	0.8308	0.8976	0.7314	0.7821	0.7637	0.5916
X2	0.8505	0.7958	0.6897	0.7625	0.7730	0.6458
X3	0.5439	0.4941	0.5643	0.6290	0.6570	0.6275
X4	0.6288	0.5845	0.6276	0.6991	0.7236	0.6642
X5	0.6897	0.6647	0.6803	0.7523	0.7821	0.6948
X6	0.4436	0.4119	0.5193	0.5713	0.6008	0.6063

Enclosure C – SAS output

Canonical Correlation Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.953951	0.953511	0.002847	0.910022
2	0.645910	0.641986	0.018439	0.417200
3	0.399448	0.390246	0.026590	0.159559
4	0.316757	0.313096	0.028464	0.100335
5	0.226966	.	0.030009	0.051514
6	0.029380	.	0.031611	0.000863

Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)			
	Eigenvalue	Difference	Proportion Cumulative
1	10.1138	9.3980	0.9041
2	0.7159	0.5260	0.0640
3	0.1899	0.0783	0.0170
4	0.1115	0.0572	0.0100
5	0.0543	0.0534	0.0049
6	0.0009		0.0001

Enclosure C – SAS output

Page 5 of 9

Canonical Correlation Analysis

Test of H0: The canonical correlations in the current row and all that follow are zero

	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	0.03757517	134.00	36	4341.4	<.0001
2	0.41760357	38.96	25	3675.5	<.0001
3	0.71654635	21.80	16	3025.1	<.0001
4	0.85258356	18.15	9	2412	<.0001
5	0.94766752	13.51	4	1984	<.0001
6	0.99913684	0.86	1	993	0.3546

Multivariate Statistics and F Approximations

S=6 M=-0.5 N=493

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.03757517	134.00	36	4341.4	<.0001
Pillai's Trace	1.63949190	62.23	36	5958	<.0001
Hotelling-Lawley Trace	11.18622263	306.59	36	2874.7	<.0001
Roy's Greatest Root	10.11381837	1673.84	6	993	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Enclosure C – SAS output

Page 6 of 9

Canonical Correlation Analysis

Raw Canonical Coefficients for the 1986-06-27 data						
	CV1986_1	CV1986_2	CV1986_3	CV1986_4	CV1986_5	CV1986_6
X1	0.0067206085	-0.006592568	0.0233141076	0.014802686	-0.007265146	-0.008745519
X2	0.0047827102	-0.010879715	-0.03955786	0.0024522392	0.0131749062	0.0111678395
X3	-0.005456504	-0.007643262	0.0375081455	0.0505007172	0.0611903953	-0.148806669
X4	0.0124374356	0.0173960484	-0.04633181	-0.048729623	-0.173733984	0.105827791
X5	0.0007958029	0.0184702953	0.0176109879	-0.068087599	0.0989430239	-0.004354097
X6	-0.003974868	0.008428175	0.0133965666	0.0486575539	0.0049728433	0.0664256848

Raw Canonical Coefficients for the 2005-08-18 data						
	CV2005_1	CV2005_2	CV2005_3	CV2005_4	CV2005_5	CV2005_6
Y1	0.0060754891	-0.000311154	-0.024603335	0.0058868396	-0.005059966	0.0226986301
Y2	0.003973438	-0.021172077	0.0121333224	0.0086587368	0.0171146531	-0.032610609
Y3	-0.011535187	-0.033085711	0.0125711308	0.011810573	0.0075190195	0.118672467
Y4	0.0196730614	0.0341934149	-0.009457269	-0.010930644	-0.087538537	-0.079816455
Y5	-0.000724941	0.0222463395	0.0276196232	-0.070552892	0.062916829	0.0197705041
Y6	-0.004050679	0.0161714137	-0.02228827	0.0656234814	0.02333460728	-0.043411169

Enclosure C – SAS output

Canonical Correlation Analysis

Standardized Canonical Coefficients for the 1986-06-27 data						
CV1986_1	CV1986_2	CV1986_3	CV1986_4	CV1986_5	CV1986_6	
X1	0.5514	-0.5409	1.9129	1.2145	-0.5961	-0.7176
X2	0.3166	-0.7202	-2.6187	0.1623	0.8722	0.7393
X3	-0.2869	-0.4019	1.9721	2.6552	3.2173	-7.8240
X4	0.5800	0.8112	-2.1606	-2.2725	-8.1019	4.9352
X5	0.0354	0.8211	0.7829	-3.0269	4.3987	-0.1936
X6	-0.1694	0.3592	0.5710	2.0738	0.2119	2.8311

Standardized Canonical Coefficients for the 2005-08-18 data						
CV2005_1	CV2005_2	CV2005_3	CV2005_4	CV2005_5	CV2005_6	
Y1	0.3993	-0.0204	-1.6169	0.3869	-0.3325	1.4917
Y2	0.3149	-1.6782	0.9617	0.6863	1.3566	-2.5848
Y3	-0.6950	-1.9933	0.7574	0.7115	0.4530	7.1496
Y4	1.1925	2.0726	-0.5733	-0.6626	-5.3062	-4.8381
Y5	-0.0346	1.0631	1.3199	-3.3716	3.0067	0.9448
Y6	-0.1711	0.6832	-0.9417	2.7726	0.9864	-1.8341

Enclosure C – SAS output

Canonical Structure

Correlations Between the 1986-06-27 data and Their Canonical Variables						
CV1986_1	CV1986_2	CV1986_3	CV1986_4	CV1986_5	CV1986_6	
X1	0.9550	-0.2229	0.1915	0.0152	0.0119	0.0348
X2	0.9256	0.1794	-0.2816	0.1419	0.1058	-0.0225
X3	0.6296	0.7211	-0.1530	0.1860	0.0085	-0.1599
X4	0.7274	0.6616	-0.1262	0.0933	-0.0462	-0.0798
X5	0.7999	0.5880	-0.0347	-0.0057	0.1124	-0.0240
X6	0.5268	0.7764	-0.0831	0.3179	0.0872	0.0636

Correlations Between the 2005-08-18 data and Their Canonical Variables						
CV2005_1	CV2005_2	CV2005_3	CV2005_4	CV2005_5	CV2005_6	
Y1	0.9281	-0.0692	-0.3275	-0.0171	0.1286	0.0992
Y2	0.9412	-0.1382	0.2511	0.1426	0.1077	0.0053
Y3	0.7965	0.2094	0.4480	0.2843	-0.0057	0.2003
Y4	0.8675	0.2556	0.3613	0.1975	-0.0492	0.1008
Y5	0.8627	0.3270	0.3113	0.1003	0.1645	0.1216
Y6	0.6895	0.4488	0.3335	0.4240	0.1452	0.1054

Enclosure C – SAS output

Canonical Structure

Correlations Between the 1986-06-27 data and the Canonical Variables of the 2005-08-18 data						
	CV2005_1	CV2005_2	CV2005_3	CV2005_4	CV2005_5	CV2005_6
X1	0.9110	-0.1440	0.0765	0.0048	0.0027	0.0010
X2	0.8829	0.1159	-0.1125	0.0449	0.0240	-0.0007
X3	0.6006	0.4658	-0.0611	0.0589	0.0019	-0.0047
X4	0.6939	0.4273	-0.0504	0.0295	-0.0105	-0.0023
X5	0.7631	0.3798	-0.0139	-0.0018	0.0255	-0.0007
X6	0.5026	0.5015	-0.0332	0.1007	0.0198	0.0019

Correlations Between the 2005-08-18 data and the Canonical Variables of the 1986-06-27 data						
	CV1986_1	CV1986_2	CV1986_3	CV1986_4	CV1986_5	CV1986_6
Y1	0.8853	-0.0447	-0.1308	-0.0054	0.0292	0.0029
Y2	0.8979	-0.0892	0.1003	0.0452	0.0245	0.0002
Y3	0.7599	0.1352	0.1790	0.0900	-0.0013	0.0059
Y4	0.8275	0.1651	0.1443	0.0625	-0.0112	0.0030
Y5	0.8230	0.2112	0.1243	0.0318	0.0373	0.0036
Y6	0.6577	0.2899	0.1332	0.1343	0.0329	0.0031

Enclosure D – SAS Program

```
Title 'The AgNP data set';
proc print data=AgNP;
run;

Title 'Model M';
Proc GLM data=AgNP;
class scan position gel;
model FBP = scan position gel corrDose;
run;

title 'Model H';
proc GLM data=AgNP ;
model FBP = corrDose;
run;
```

The SAS programs for reading the data and for generating the influence statistics are not shown. Furthermore, some of the outputs has been omitted.

Enclosure D – SAS Output

The AgNP data set

Obs	scan	Posi- tion	gel	corrDose	FBP
1	1	1	1	2.0186	0.01367
2	1	2	2	1.9866	0.02543
3	1	3	3	1.9965	0.02454
4	1	4	4	2.0397	0.01659
5	1	5	5	0.0000	-0.01325
6	2	1	1	2.0190	0.02909
7	2	2	2	1.9871	0.02252
8	2	3	3	1.9970	0.03902
9	2	4	4	2.0402	0.02545
10	2	5	5	0.0000	0.02176
11	3	1	6	2.0144	0.02901
12	3	2	7	1.9840	0.03208
13	3	3	8	1.9884	0.02083
14	3	4	9	2.0253	0.02047
15	3	5	5	0.0000	-0.00033
16	1	1	6	9.8833	0.11614
17	1	2	7	9.7431	0.11877
18	1	3	8	9.7201	0.10500
19	1	4	9	9.9245	0.10739
20	1	5	5	0.0000	0.00765
21	2	1	6	9.8833	0.11987
22	2	2	7	9.7431	0.11064
23	2	3	8	9.7201	0.11972
24	2	4	9	9.9245	0.11255
25	2	5	5	0.0000	0.00288
26	3	1	1	9.8745	0.08778
27	3	2	2	9.7305	0.10687
28	3	3	3	9.7352	0.12881
29	3	4	4	9.9704	0.10293
30	3	5	5	0.0000	0.01054
31	1	1	1	20.6310	0.19444
32	1	2	2	20.4520	0.24854
33	1	3	3	20.4120	0.24268
34	1	4	4	21.0180	0.23534
35	1	5	5	0.0000	0.00096
36	2	1	1	20.5970	0.23095
37	2	2	2	20.4180	0.23505
38	2	3	4	20.3780	0.20731
39	2	4	3	20.9830	0.26776
40	2	5	5	0.0000	0.01904
41	3	1	6	20.6480	0.23500
42	3	2	7	20.4780	0.23695
43	3	3	8	20.3710	0.22203
44	3	4	9	20.9250	0.24142
45	3	5	5	0.0000	0.00922

Enclosure D – SAS Output

Model M

Class Level Information	
Class	Levels/Values
scan	3 2 3
position	5 1 2 3 4 5
gel	9 1 2 3 4 5 6 7 8 9

Number of Observations Read	45
Number of Observations Used	45

The GLM Procedure

Dependent Variable: FBP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	0.36471758	0.03039313	399.06	<.0001
Error	32	0.00243716	0.00007616		
Corrected Total	44	0.36715474			

R-Square	Coeff Var	Root MSE	FBP Mean
0.9933628	7.44324	0.008727	0.099802

Source	DF	Type I SS	Mean Square	F Value	Pr > F
scan	20.00049585	0.00024793	3.226	0.0517	
position	40.09838848	0.02459712	322.96	<.0001	
gel	50.00311489	0.00062298	8.18	<.0001	
corrDose	10.26271836	0.26271836	3449.51	<.0001	

Source	DF	Type III SS	Mean Square	F Value	Pr > F
scan	20.00053439	0.00026719	3.51	0.0419	
position	10.00055901	0.00055901	7.34	0.0107	
gel	50.00287399	0.00057480	7.55	<.0001	
corrDose	10.26271836	0.26271836	3449.51	<.0001	

Enclosure D – SAS Output

Model H

The GLM Procedure

Dependent Variable: FBP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.36071708	0.36071708	2409.39	<.0001
Error	43	0.00643765	0.00014971		
Corrected Total	44	0.36715474			

R-Square	Coeff Var	Root MSE	FBP Mean
0.9824661	2.25996	0.012236	0.099802

Source	DF	Type I SS	Mean Square	F Value	Pr > F
corrDose	10.36071708	0.36071708	2409.39	<.0001	

Source	DF	Type III SS	Mean Square	F Value	Pr > F
corrDose	10.36071708	0.36071708	2409.39	<.0001	

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	0.0039255847	0.00267248	1.47	0.1491
corrDose	0.0110837163	0.00022580	49.09	<.0001

Enclosure D – SAS Output

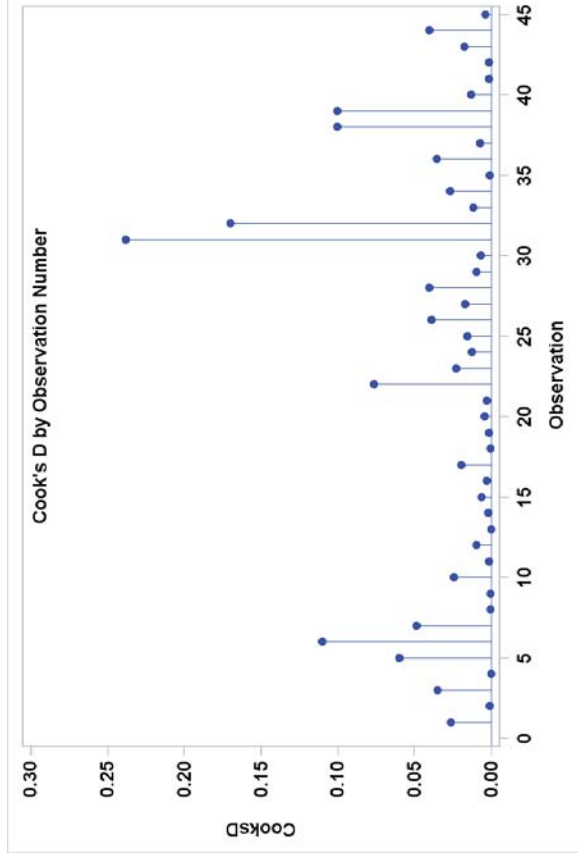
Some influence statistics

Obs	scan	Posi- tion	Corr- Dose	FBP	Residual	Rstudent	CooksD	DFFITS
1	1	1	2.0186	0.01367	0.007048	0.94786	0.02647	0.58567
2	1	2	1.9866	0.02543	0.001404	0.18616	0.00105	0.11483
3	1	3	1.9965	0.02454	-0.007504	-1.02920	0.03498	0.67496
4	1	4	2.0397	0.01659	0.000341	0.04606	0.00007	0.03032
5	1	5	0.0000	-0.01325	-0.016442	-2.16650	0.06007	-0.93328
6	2	1	2.0190	0.02909	0.014385	2.02970	0.11019	1.25387
7	2	2	1.9871	0.02252	-0.009598	-1.30637	0.04885	-0.80562
8	2	3	1.9970	0.03902	-0.001118	-0.15089	0.00078	-0.09894
9	2	4	2.0402	0.02545	0.001115	0.15067	0.00078	0.09917
10	2	5	0.0000	0.02176	0.010488	1.32391	0.02444	0.57031
11	3	1	2.0144	0.02901	0.001400	0.19059	0.00131	0.12865
12	3	2	1.9840	0.03208	0.003853	0.52637	0.00990	0.35468
13	3	3	1.9884	0.02083	-0.000144	-0.01953	0.00001	-0.01315
14	3	4	2.0253	0.02047	-0.001676	-0.22834	0.00189	-0.15437
15	3	5	0.0000	-0.00033	-0.005359	-0.66395	0.00655	-0.28915
16	1	1	9.8833	0.11614	0.002191	0.29802	0.00317	0.20009
17	1	2	9.7431	0.11877	0.005442	0.74582	0.01956	0.50076
18	1	3	9.7201	0.10500	-0.000773	-0.10506	0.00039	-0.07054
19	1	4	9.9245	0.10739	-0.001437	-0.19527	0.00136	-0.13112
20	1	5	0.0000	0.00765	0.004461	0.55045	0.00442	0.23712
21	2	1	9.8833	0.11197	-0.002162	-0.29414	0.00309	-0.19749
22	2	2	9.7431	0.11064	-0.010771	-1.51655	0.07664	-1.01823
23	2	3	9.7201	0.11972	0.005864	0.80484	0.02271	0.54037
24	2	4	9.9245	0.11255	-0.004360	-0.59561	0.01256	-0.39992
25	2	5	0.0000	0.00288	-0.008388	-1.04815	0.01563	-0.45152
26	3	1	9.8745	0.08778	-0.008724	-1.17737	0.03903	-0.71662
27	3	2	9.7305	0.10687	-0.005781	-0.77063	0.01714	-0.46910
28	3	3	9.7352	0.12881	0.008195	1.12322	0.04035	0.72725
29	3	4	9.9704	0.10293	-0.004037	-0.54498	0.00979	-0.35285
30	3	5	0.0000	0.01054	0.005510	0.68292	0.00692	0.29741
31	1	1	20.6310	0.19444	-0.020758	-3.18228	0.23801	-1.99416
32	1	2	20.4520	0.24854	0.017584	2.57796	0.17013	1.61305

Enclosure D – SAS Output

Obs	scan	Posi- tion	Corr- Dose	FBP	Residual	Rstudent	CooksD	DFFITS
33	1	3	20.4120	0.24268	0.004261	0.57990	0.01163	0.38484
34	1	4	21.0180	0.23534	0.006414	0.88120	0.02693	0.58957
35	1	5	0.0000	0.00096	-0.002232	-0.27445	0.00111	-0.11823
36	2	1	20.5970	0.23095	0.008050	1.09161	0.03574	0.68366
37	2	2	20.4180	0.23505	-0.003608	-0.48173	0.00715	-0.30125
38	2	3	20.3780	0.20731	-0.003834	-0.77440	0.10030	-1.13470
39	2	4	20.9830	0.26776	-0.003834	-0.77440	0.10030	-1.13470
40	2	5	0.0000	0.01904	0.007771	0.96868	0.01342	0.41729
41	3	1	20.6480	0.23500	-0.001428	-0.19615	0.00147	-0.13603
42	3	2	20.4780	0.23695	0.001477	0.20272	0.00156	0.14043
43	3	3	20.3710	0.22203	-0.004947	-0.68348	0.01748	-0.47275
44	3	4	20.9250	0.24142	0.007472	1.04478	0.04059	0.72747
45	3	5	0.0000	0.00922	0.004190	0.51768	0.00400	0.22545

Some influence statistics



Some influence statistics

