

Written test, date: 8. December 2009

Course no. : 02409

Course name: Multivariate Statistics “Multivariat Statistik”.

Aids allowed: All usual ones

“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 9 problems. The answers to the 30 questions must be written into the table below.

Problem	1	1	1	1	1	1	2	2	2	2
Question	1.1	1.2	1.3	1.4	1.5	1.6	2.1	2.2	2.3	2.4
Answer										

Problem	2	2	3	3	3	3	3	4	4	4
Question	2.5	2.6	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3
Answer										

Problem	4	4	4	5	5	6	7	8	9	9
Question	4.4	4.5	4.6	5.1	5.2	6.1	7.1	8.1	9.1	9.2
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”) gives 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. Please check that all pages of the exam paper and the enclosure are present.

Problem 1.

Enclosure A with the SAS-program and the corresponding SAS-output belongs to this problem.

The data are part of a study on two species of male flea-beetles: *Chaetocnema concinna* and *Chaetocnema heikertlinger*.

The variables are:

Variable	Description
species	: "Conc" = <i>Chaetocnema concinna</i> , "Heik" = <i>Chaetocnema heikertlinger</i>
number	: observation number within species
x_1	: width of the first joint of the first tarsus in microns (the sum of measurements for both tarsi)
x_2	: the same for the second joint
x_3	: the maximal width of the aedeagus in the fore-part in microns

Option "simple" in the "proc discrim" statement gives various univariate statistics.

Question 1.1.

Mahalanobis' distance between the two groups is:

- 1 $\frac{(193.8-183.1)^2}{267.2}$
- 2 $|-290.6 - (-245.9)|$
- 3 $1 - 0.0323$
- 4 15.07
- 5 9.52
- 6 Don't know.

The problem continues on the next page

Question 1.2.

The usual test statistic for testing a difference in mean for the two groups is distributed as $F(\nu_1, \nu_2)$ where ν_2 equals:

- 1 46
- 2 47
- 3 48
- 4 49
- 5 50
- 6 Don't know.

Question 1.3.

The pooled variance estimate for variable x_1 is:

- 1 267.2
- 2 $\frac{267.2}{193.8}$
- 3 $\frac{147.49+222.13}{2}$
- 4 $\frac{20 \cdot 147.49 + 30 \cdot 222.13}{20+30}$
- 5 $\frac{21 \cdot 147.49 + 31 \cdot 222.13}{21+31}$
- 6 Don't know.

Question 1.4.

A new flea-beetle specimen is collected. It is classified as "Conc" if $\mathbf{d}' \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} > c$.

The constant c is:

- 1 -44.63
- 2 44.63
- 3 15.07
- 4 9.52
- 5 10
- 6 Don't know.

The problem continues on the next page

Question 1.5.

The generalised variance of the pooled covariance matrix is:

- 1 13618.0
- 2 $267.2 \cdot 72.1 \cdot 6.6$
- 3 $e^{\sqrt{15.07}}$
- 4 $12.1^2 + 7.2^2 + 2.2^2 + 14.9^2 + 6.6^2 + 2.3^2$
- 5 50
- 6 Don't know.

Question 1.6.

In order to construct a quadratic classification rule with equal priors based only on variable x_1 we **do not** need the value:

- 1 183.1
- 2 147.5
- 3 201.0
- 4 222.1
- 5 193.8
- 6 Don't know.

Problem 2.

Enclosure B with the SAS-program and the corresponding SAS-output belongs to this problem.

The data are part of a study on a single species of male flea-beetles: *Chaetocnema heikertlinger*.

The variables are:

Variable	Description
number	: observation number
x_1	: width of the first joint of the first tarsus in microns (the sum of measurements for both tarsi)
x_2	: the same for the second joint
x_3	: the maximal width of the aedeagus in the fore-part in microns
x_4	: the front angle of the aedeagus (1 unit = 7.5 degrees)
x_5	: the maximal width of the head between the external edges of the eyes in 0.01mm
x_6	: the aedeagus width from the side in microns

Question 2.1.

We are analysing the correlation matrix. In this case the total variance is:

- 1 1
- 2 2.71
- 3 3.81
- 4 $2.71 \cdot 1.10$
- 5 6
- 6 Don't know.

Question 2.2.

The usual test for the partial correlation $\rho_{12|456}$ has the following number of degrees of freedom under the null hypothesis:

- 1 29
- 2 28
- 3 27
- 4 26
- 5 25
- 6 Don't know.

The problem continues on the next page

Question 2.3.

Consider a principal component analysis based on the correlation matrix. For the first observation the value of the first principal component is found as:

$$1 \quad \begin{bmatrix} 0.52 \\ 0.47 \\ \vdots \\ 0.26 \end{bmatrix}'$$

$$2 \quad \begin{bmatrix} 0.52 \\ 0.47 \\ \vdots \\ 0.26 \end{bmatrix}' \cdot \begin{bmatrix} 186 \\ 107 \\ \vdots \\ 84 \end{bmatrix} - \begin{bmatrix} 201 \\ 119.3 \\ \vdots \\ 81 \end{bmatrix}$$

$$3 \quad \begin{bmatrix} 0.52 \\ 0.47 \\ \vdots \\ 0.26 \end{bmatrix}' \cdot \begin{bmatrix} 14.9 & 0 & \dots & 0 \\ 0 & 6.6 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 8.9 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 186 \\ 107 \\ \vdots \\ 84 \end{bmatrix} - \begin{bmatrix} 201 \\ 119.3 \\ \vdots \\ 81 \end{bmatrix}$$

$$4 \quad \begin{bmatrix} 0.52 \\ 0.47 \\ \vdots \\ 0.26 \end{bmatrix}' \cdot \begin{bmatrix} 14.9 & 0 & \dots & 0 \\ 0 & 6.6 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 8.9 \end{bmatrix} \cdot \begin{bmatrix} 186 \\ 107 \\ \vdots \\ 84 \end{bmatrix} - \begin{bmatrix} 201 \\ 119.3 \\ \vdots \\ 81 \end{bmatrix}$$

$$5 \quad \begin{bmatrix} 0.52 \\ 0.47 \\ \vdots \\ 0.26 \end{bmatrix}' \cdot \begin{bmatrix} 14.9 & 0 & \dots & 0 \\ 0 & 6.6 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 8.9 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 201 \\ 119.3 \\ \vdots \\ 81 \end{bmatrix}$$

6 Don't know.

Question 2.4.

For the two factor model the varimax rotation corresponds to a rotation of:

1 30.7 degrees

2 75.7 degrees

3 104.3

4 -14.3 degrees

5 14.3 degrees

6 Don't know.

The problem continues on the next page

Question 2.5.

Compare the two factor and the three factor model. The *increase* in the proportion of the variance explained by going from a two factor model to a three factor model is:

- 1 0.33
- 2 0.15
- 3 0.19
- 4 0.18
- 5 0.64
- 6 Don't know.

Question 2.6.

Compare the two factor and the three factor model. The proportion of the variance explained for each variable changes dramatically for one of the variables. Which one?

- 1 x_1
- 2 x_2
- 3 x_6
- 4 x_4
- 5 x_5
- 6 Don't know.

Problem 3.

Enclosure C with the SAS-program and the corresponding SAS-output belongs to this problem.

The data originate from an experiment where a supermarket chain randomly assigned different prices (in US dollars) to a house brand of coffee. The coffee was advertised (advertise=1) in a standard way in all cases. The amount of coffee sold (in pounds) was registered. Some weeks later the experiment was conducted again but without advertising (advertise=0). The squared values of the prices are also included in the data (variable: price2).

The first proc glm corresponds to a regression model with individual intercepts, individual coefficients to the first degree term (price), and individual coefficients to the second degree term (price2). Consider this model as "M".

The second glm corresponds to a simplified regression model. Consider this model as "H".

In some of the following questions the notation in the ANOVA-table on p. 136 in the book is used.

The problem continues on the next page

Question 3.1.

Going from model "M" to model "H" the fraction of variance explained drops by:

- 1 5.697
- 2 2
- 3 0.203
- 4 0.000156
- 5 1356.08
- 6 Don't know.

Question 3.2.

$\| P_M(\mathbf{Y}) - P_H(\mathbf{Y}) \|^2$ equals:

- 1 $3252.47 - 1896.39$
- 2 610.1
- 3 44042.21
- 4 $14.17 - 13.97$
- 5 298.9
- 6 Don't know.

Question 3.3.

$k - r$ equals:

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

The problem continues on the next page

Question 3.4.

$\| \mathbf{Y} - P_M(\mathbf{Y}) \|^2$ equals:

- 1 3211.7
- 2 1903352
- 3 1896.4
- 4 200.7
- 5 0.9983
- 6 Don't know.

Question 3.5.

$n - k$ equals:

- 1 14
- 2 15
- 3 16
- 4 17
- 5 18
- 6 Don't know.

Problem 4.

A supermarket chain randomly assigned different prices (in US dollars) to a house brand of coffee. The amount of coffee sold (in pounds) was registered. The following data corresponds to a subset of the original data in Enclosure C.

Pounds	Price
1124	3.0
830	3.4
619	3.8
451	4.2
296	4.6
194	5.0

In the following we choose to analyse the data by means of orthogonal polynomials.

The problem continues on the next page

Question 4.1.

We first consider constructing the orthogonal polynomials directly based on the original values of "price" in dollars. In that case the first order (orthogonal) polynomial is:

- 1 price
- 2 price - 3.5
- 3 price - 4.0
- 4 price + 3.5
- 5 price + 4.0
- 6 Don't know.

Question 4.2.

In order to use the framework in the book we need to perform the following:

- 1 Transform the price by $\frac{\text{price}-3.0}{5.0-3.0}$
- 2 Transform the price by $\frac{\text{price}-3.0}{0.4} + 1$
- 3 Transform the price by $\frac{\text{price}-2.0}{5.0-3.0}$
- 4 Transform the price by $\frac{\text{price}-2.0}{0.4}$
- 5 Transform the price by $\frac{\text{price}-5.0}{0.4} + 1$
- 6 Don't know.

Question 4.3.

In order to conduct the analysis we need the following entry in table 4.1:

- 1 The entry corresponding to $n = 8$
- 2 The entry corresponding to $n = 2$
- 3 The entry corresponding to $n = 7$
- 4 The entry corresponding to $n = 4$
- 5 The entry corresponding to $n = 6$
- 6 Don't know.

The problem continues on the next page

Question 4.4.

A polynomial of order two was chosen.

The estimated residual variance is: 187.6.

The standard errors of the parameter estimates for the individual orthogonal polynomials (using transformed prices) are as follows:

Order	Estimate	Standard error
0	585.7	5.59
1	-91.7	1.64
2	14.1	1.49

The coefficient for the second order orthogonal polynomial is

- 1 significant at level 0.001 but not at level 0.0005
- 2 significant at level 0.005 but not at level 0.001
- 3 significant at level 0.01 but not at level 0.005
- 4 significant at level 0.05 but not at level 0.01
- 5 significant at level 0.1 but not at level 0.05
- 6 Don't know.

The problem continues on the next page

Question 4.5.

A polynomial of order two was chosen.

The estimated residual variance is: 187.6.

The standard errors of the parameter estimates for the individual orthogonal polynomials (using transformed prices) are as follows:

Order	Estimate	Standard error
0	585.7	5.59
1	-91.7	1.64
2	14.1	1.49

We wish to construct a confidence interval for the expected value of a new observation at (transformed) price: p . The constant c used in the construction of a confidence interval for the expected value of a new observation is:

$$1 \square \begin{bmatrix} 1 & p & p^2 \end{bmatrix} \cdot \frac{1}{187.6} \begin{bmatrix} 5.59^2 & 0 & 0 \\ 0 & 1.64^2 & 0 \\ 0 & 0 & 1.49^2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}$$

$$2 \square \begin{bmatrix} 1 & p & p^2 \end{bmatrix} \cdot 187.6 \begin{bmatrix} 5.59^2 & 0 & 0 \\ 0 & 1.64^2 & 0 \\ 0 & 0 & 1.49^2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}$$

$$3 \square \begin{bmatrix} 1 & p & p^2 \end{bmatrix} \cdot \frac{1}{187.6} \begin{bmatrix} 5.59 & 0 & 0 \\ 0 & 1.64 & 0 \\ 0 & 0 & 1.49 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}$$

$$4 \square \begin{bmatrix} 1 & p & p^2 \end{bmatrix} \cdot 187.6 \begin{bmatrix} 5.59 & 0 & 0 \\ 0 & 1.64 & 0 \\ 0 & 0 & 1.49 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}$$

$$5 \square \begin{bmatrix} 1 & p & p^2 \end{bmatrix} \begin{bmatrix} 5.59^2 & 0 & 0 \\ 0 & 1.64^2 & 0 \\ 0 & 0 & 1.49^2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ p \\ p^2 \end{bmatrix}$$

6 Don't know.

The problem continues on the next page

Question 4.6.

A polynomial of order two was chosen.

The estimated residual variance is: 187.6.

The standard errors of the parameter estimates for the individual orthogonal polynomials (using transformed prices) are as follows:

Order	Estimate	Standard error
0	585.7	5.59
1	-91.7	1.64
2	14.1	1.49

In order to make a 95% confidence interval for the expected value of a new observation we need the following percentile from a t-distribution:

- 1 5.841
- 2 3.182
- 3 2.353
- 4 4.303
- 5 2.920
- 6 Don't know.

Problem 5.

Consider the two random vectors: \mathbf{X} and \mathbf{Y} .

Assume: $E(\mathbf{X}) = \boldsymbol{\mu}_X$, $D(\mathbf{X}) = \boldsymbol{\Sigma}_{XX}$, $E(\mathbf{Y}) = \boldsymbol{\mu}_Y$, $D(\mathbf{Y}) = \boldsymbol{\Sigma}_{YY}$, $C(\mathbf{X}, \mathbf{Y}) = \boldsymbol{\Sigma}_{XY}$

Furthermore, consider the constant matrices \mathbf{A} and \mathbf{B} and the constant vector \mathbf{c} . All have dimensions so the involved matrix-vector operations exist.

Question 5.1.

$E(\mathbf{AX} + \mathbf{BY} + \mathbf{c})$ equals

- 1 $\mathbf{A}\boldsymbol{\mu}_X + \mathbf{B}\boldsymbol{\mu}_Y$
- 2 $\mathbf{A}\boldsymbol{\mu}_X\mathbf{A}' + \mathbf{B}\boldsymbol{\mu}_Y\mathbf{B}' + \mathbf{c}$
- 3 $\mathbf{A}\mathbf{A}'\boldsymbol{\mu}_X + \mathbf{B}\mathbf{B}'\boldsymbol{\mu}_Y + \mathbf{c}$
- 4 $\mathbf{A}\boldsymbol{\mu}_X + \mathbf{B}\boldsymbol{\mu}_Y + \mathbf{c}$
- 5 $\mathbf{A}\mathbf{B}' + \boldsymbol{\mu}_X\boldsymbol{\mu}_Y'$
- 6 Don't know.

The problem continues on the next page

Question 5.2.

$D(\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Y})$ equals

- 1 $\mathbf{A}\Sigma_{XX} + \mathbf{B}\Sigma'_{XY} + \mathbf{A}\Sigma_{XY} + \mathbf{B}\Sigma_{YY}\mathbf{B}'$
- 2 $\mathbf{A}\Sigma_{XX}\mathbf{A}' + \mathbf{B}\Sigma_{YY}\mathbf{B}' + \mathbf{c}\mathbf{c}'$
- 3 $\mathbf{A}\Sigma_{XX}\mathbf{A}' + \mathbf{B}\Sigma_{YY}\mathbf{B}'$
- 4 $\mathbf{A}\Sigma_{XX}\mathbf{A}' + \mathbf{B}\Sigma'_{XY}\mathbf{A}' + \mathbf{A}\Sigma_{XY}\mathbf{B}' + \mathbf{B}\Sigma_{YY}\mathbf{B}' + \mathbf{c}\mathbf{c}'$
- 5 $\mathbf{A}\Sigma_{XX}\mathbf{A}' + \mathbf{B}\Sigma'_{XY}\mathbf{A}' + \mathbf{A}\Sigma_{XY}\mathbf{B}' + \mathbf{B}\Sigma_{YY}\mathbf{B}'$
- 6 Don't know.

Problem 6.

Consider a linear discriminant analysis with 3 or more groups.

Question 6.1.

An overall assessment (test) for the null hypothesis: "all group means are equal" is best tested using:

- 1 Three Hotellings T^2 in the one-sample case
- 2 Three Hotellings T^2 in the two-sample case
- 3 A multivariate one-sided analysis of variance
- 4 A multivariate two-sided analysis of variance
- 5 A test for equality of the last three eigenvalues of the covariance matrix
- 6 Don't know.

Problem 7.

Question 7.1.

Only one of the following matrices can be a covariance matrix. Which one?

1 $\begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 2 \\ 0 & 2 & 1 \end{bmatrix}$

2 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 2 \\ 0 & 2 & 4 \end{bmatrix}$

3 $\begin{bmatrix} 2 & -3 \\ -3 & 2 \end{bmatrix}$

4 $\begin{bmatrix} -2 & 1 \\ 1 & -1 \end{bmatrix}$

5 $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

6 Don't know.

Problem 8.

Consider two populations: $\pi_1 \simeq N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $\pi_2 \simeq N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$

Assume: $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = 1$

Classify as π_1 if: $x' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2} \boldsymbol{\mu}_1' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 > 0$

Question 8.1.

The probability of misclassification for population 1 is:

1 0.1587

2 0.3085

3 0.5

4 0.6915

5 0.8413

6 Don't know.

Problem 9.

An experiment is conducted with 4 levels of factor A and 5 levels of factor B. For each combination of levels of A and B there is one observation. The outcome is measured using a bivariate variable.

Question 9.1.

The usual test statistic for the hypothesis: "no overall effect of factor A" is distributed as $U(\nu_1, \nu_2, \nu_3)$ where ν_1 is:

1 1

2 2

3 3

4 4

5 5

6 Don't know.

Question 9.2

The usual test statistic for the hypothesis: "no overall effect of factor A" is distributed as $U(\nu_1, \nu_2, \nu_3)$ where ν_2 is:

1 1

2 2

3 3

4 4

5 5

6 Don't know.

Nov 24, 09 11:26

EnclA.sas

Page 1/1

```
/* encla.sas   Crted: 22-11-09 08:55 by BKE. Updt: 22-11-09 09:02 */
/* Purpose: */

title1 'Enclosure A - data from two species of flea-beetle';

title2 'Print of data';
proc print data=stat2.flea beetlea;
run;

title2 'Discriminant analysis';
proc discrim data=stat2.flea beetlea simple wcov pool=yes;
class species;
var x1 x2 x3;
run;
```

Enclosure A - data from two species of flea-beetle
Discriminant analysis

The DISCRIM Procedure

Observations	52	DF Total	51
Variables	3	DF Within Classes	50
Classes	2	DF Between Classes	1

Class Level Information

species	Variable Name	Frequency	Weight	Proportion	Prior Probability
Conc	Conc	21	21.0000	0.403846	0.500000
Heik	Heik	31	31.0000	0.596154	0.500000

Enclosure A - data from two species of flea-beetle
Discriminant analysis

The DISCRIM Procedure
Within-Class Covariance Matrices

Variable	species = Conc,	DF = 20	x1	x2	x3
x1	147.4904762	66.6380952	18.5261905	11.5452381	4.9904762
x2	66.6380952	51.2476190	11.5452381	18.5261905	4.9904762
x3	18.5261905	11.5452381	4.9904762	18.5261905	11.5452381

Variable	species = Heik,	DF = 30	x1	x2	x3
x1	222.1333333	63.4000000	22.6000000	7.9096774	5.5161290
x2	63.4000000	44.1591398	7.9096774	22.6000000	5.5161290
x3	22.6000000	7.9096774	5.5161290	22.6000000	7.9096774

Enclosure A - data from two species of flea-beetle
Print of data

Obs	species	number	x1	x2	x3
1	Conc	1	191	131	53
2	Conc	2	185	134	50
3	Conc	3	200	137	52
4	Conc	4	173	127	50
5	Conc	5	171	118	49
6	Conc	6	160	118	47
7	Conc	7	188	134	54
8	Conc	8	186	129	51
9	Conc	9	174	131	52
10	Conc	10	163	115	47
11	Conc	11	190	143	52
12	Conc	12	174	131	50
13	Conc	13	201	130	51
14	Conc	14	190	133	53
15	Conc	15	182	130	51
16	Conc	16	184	131	51
17	Conc	17	177	127	49
18	Conc	18	178	126	53
19	Conc	19	210	140	54
20	Conc	20	182	121	51
21	Conc	21	186	136	56
22	Heik	1	186	107	49
23	Heik	2	211	122	49
24	Heik	3	201	114	47
25	Heik	4	242	131	54
26	Heik	5	184	108	43
27	Heik	6	211	118	51
28	Heik	7	217	122	49
29	Heik	8	223	127	51
30	Heik	9	208	125	50
31	Heik	10	199	124	46
32	Heik	11	211	129	49
33	Heik	12	218	126	49
34	Heik	13	203	122	49
35	Heik	14	192	116	49
36	Heik	15	195	123	47
37	Heik	16	211	122	48
38	Heik	17	187	123	47
39	Heik	18	192	109	46
40	Heik	19	223	124	53
41	Heik	20	188	114	48
42	Heik	21	216	120	50
43	Heik	22	185	114	46
44	Heik	23	178	119	47
45	Heik	24	187	111	49
46	Heik	25	187	112	49
47	Heik	26	201	130	54
48	Heik	27	187	120	47
49	Heik	28	210	119	50
50	Heik	29	196	114	51
51	Heik	30	195	110	49
52	Heik	31	187	124	49

Enclosure A - data from two species of flea-beetle
 Discriminant analysis

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \text{COV}^{-1} (\bar{X}_i - \bar{X}_j)$$

Generalized Squared Distance to species

From species	Conc	Heik
Conc	0	15.07245
Heik	15.07245	0

Linear Discriminant Function

$$\text{Constant} = -.5 \bar{X}_j' \text{COV}^{-1} \bar{X}_j \quad \text{Coefficient Vector} = \text{COV}^{-1} \bar{X}_j$$

Linear Discriminant Function for species

Variable	Conc	Heik
Constant	-290.58071	-245.94708
x1	-0.65325	-0.26363
x2	1.87991	1.32505
x3	8.92101	7.91420

Enclosure A - data from two species of flea-beetle
 Discriminant analysis

The DISCRIM Procedure
 Simple Statistics

Total-Sample

Variable	N	Sum	Mean	Variance	Standard Deviation
x1	52	10076	193.76923	267.20060	16.3463
x2	52	6421	123.48077	72.09766	8.4910
x3	52	2591	49.82692	6.57730	2.5646

species = Conc

Variable	N	Sum	Mean	Variance	Standard Deviation
x1	21	3845	183.09524	147.49048	12.1446
x2	21	2722	129.61905	51.24762	7.1587
x3	21	1076	51.23810	4.99048	2.2339

species = Heik

Variable	N	Sum	Mean	Variance	Standard Deviation
x1	31	6231	201.00000	222.13333	14.9041
x2	31	3699	119.32258	44.15914	6.6452
x3	31	1515	48.87097	5.51613	2.3486

Pooled Covariance Matrix Information

Covariance Matrix Rank	3
Natural Log of the Determinant of the Covariance Matrix	9.51915

Enclosure A - data from two species of flea-beetle
Discriminant analysis

The DISCRIM Procedure
Classification Summary for Calibration Data: STAT2.FLEABEETLEA
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X - \bar{X}_j)' \text{COV}_j^{-1} (X - \bar{X}_j)$$

Posterior Probability of Membership in Each species

$$\text{Pr}(j|X) = \frac{\exp(-.5 D_j^2(X)) / \sum_k \exp(-.5 D_k^2(X))}{\sum_k \exp(-.5 D_k^2(X)) / \sum_k D_k^2(X)}$$

Number of Observations and Percent Classified into species

From species	Conc	Heik	Total
Conc	21	0	21
	100.00	0.00	100.00
Heik	2	29	31
	6.45	93.55	100.00
Total	23	29	52
	44.23	55.77	100.00
Priors	0.5	0.5	

Error Count Estimates for species

	Conc	Heik	Total
Rate	0.0000	0.0645	0.0323
Priors	0.5000	0.5000	

Nov 24, 09 11:28 **EnclB.sas** Page 1/1

```
/* enclb.sas Crted: 22-11-09 08:57 by BKE. Updt: 22-11-09 10:06 */
/* Purpose: */

title1 'Enclosure B - data from a species of flea-beetle';
title2 'Print of data';
proc print data=stat2.flea beetleb;

title2 'Factor analysis - two factors';
proc factor data=stat2.flea beetleb nfactor=2 simple corr eigenvectors rotate=varim
var x1-x6;
run;

title2 'Factor analysis - three factors';
proc factor data=stat2.flea beetleb nfactor=3 eigenvectors rotate=varimax reorder;
var x1-x6;
run;
```

Enclosure B - data from a species of flea-beetle
 Factor analysis - two factors

The FACTOR Procedure

Means and Standard Deviations from 31 Observations

Variable	Mean	Std Dev
x1	201.00000	14.904138
x2	119.32258	6.645234
x3	48.87097	2.348644
x4	124.64516	4.622758
x5	14.29032	1.101319
x6	81.00000	8.929352

Correlations

	x1	x2	x3	x4	x5	x6
x1	1.00000					
x2	0.64014	1.00000				
x3	0.64563	0.50679	1.00000			
x4	0.44075	0.38472	0.52371	1.00000		
x5	0.26603	0.04599	0.00208	-0.06421	1.00000	
x6	0.22141	0.19324	0.20186	0.28344	0.12880	1.00000

Enclosure B - data from a species of flea-beetle
 Factor analysis - two factors

Obs	species	number	x1	x2	x3	x4	x5	x6
1	Heik	1	186	107	49	120	14	84
2	Heik	2	211	122	49	123	16	95
3	Heik	3	201	114	47	130	14	74
4	Heik	4	242	131	54	131	16	90
5	Heik	5	184	108	43	116	16	75
6	Heik	6	211	118	51	122	15	90
7	Heik	7	217	122	49	127	15	73
8	Heik	8	223	127	51	132	16	84
9	Heik	9	208	125	50	125	14	88
10	Heik	10	199	124	46	119	13	78
11	Heik	11	211	129	49	122	13	83
12	Heik	12	218	126	49	120	15	85
13	Heik	13	203	122	49	119	14	73
14	Heik	14	192	116	49	123	15	90
15	Heik	15	195	123	47	125	15	77
16	Heik	16	211	122	48	125	14	73
17	Heik	17	187	123	47	129	14	75
18	Heik	18	192	109	46	130	13	90
19	Heik	19	223	124	53	129	13	82
20	Heik	20	188	114	48	122	12	74
21	Heik	21	216	120	50	129	15	86
22	Heik	22	185	114	46	124	15	92
23	Heik	23	178	119	47	120	13	78
24	Heik	24	187	111	49	119	16	66
25	Heik	25	187	112	49	119	14	55
26	Heik	26	201	130	54	133	13	84
27	Heik	27	187	120	47	121	15	86
28	Heik	28	210	119	50	128	14	68
29	Heik	29	196	114	51	129	14	86
30	Heik	30	195	110	49	124	13	89
31	Heik	31	187	124	49	129	14	88

Enclosure B - data from a species of flea-beetle
 Factor analysis - two factors

The FACTOR Procedure
 Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.96917	0.24638
2	-0.24638	0.96917

Rotated Factor Pattern

	Factor1	Factor2
x3	0.83955	0.02068
x1	0.78964	0.38036
x2	0.76379	0.13627
x4	0.76041	-0.11508
x5	-0.07557	0.94806
x6	0.34771	0.35212

Variance Explained by Each Factor

	Factor1	Factor2
	2.6165937	1.1997179

Final Communality Estimates: Total = 3.816312

x1	x2	x3	x4	x5	x6
0.76821148	0.60194896	0.70526671	0.59146884	0.90452123	0.24489435

Enclosure B - data from a species of flea-beetle
 Factor analysis - two factors

The FACTOR Procedure
 Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 6 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	2.71448658	1.61266159	0.4524	0.4524
2	1.10182498	0.19353518	0.1836	0.6361
3	0.90828981	0.32705803	0.1514	0.7874
4	0.58123178	0.14502894	0.0969	0.8843
5	0.43620284	0.17823882	0.0727	0.9570
6	0.25796402	0.0430	0.0430	1.0000

2 factors will be retained by the NFACTOR criterion.

Eigenvectors

	1	2
x1	0.52138	0.16585
x2	0.46967	-0.05345
x3	0.49695	-0.17797
x4	0.43010	-0.28473
x5	0.09732	0.89308
x6	0.25720	0.24350

Factor Pattern

	Factor1	Factor2
x1	0.85901	0.17409
x3	0.81876	-0.18681
x2	0.77382	-0.05611
x4	0.70862	-0.29888
x6	0.42375	0.25560
x5	0.16034	0.93745

Variance Explained by Each Factor

	Factor1	Factor2
	2.7144866	1.1018250

Final Communality Estimates: Total = 3.816312

x1	x2	x3	x4	x5	x6
0.76821148	0.60194896	0.70526671	0.59146884	0.90452123	0.24489435

Nov 24, 09 11:30		EncIB.III		Page 6/6	
Enclosure B - data from a species of flea-beetle					
Factor analysis - three factors					
The FACTOR Procedure					
Rotation Method: Varimax					
Orthogonal Transformation Matrix					
	1	2	3		
1	0.93992	0.32993	0.08772		
2	-0.15063	0.17020	0.97383		
3	-0.30636	0.92854	-0.20967		
Rotated Factor Pattern					
	Factor1	Factor2	Factor3		
x1	0.85927	0.07638	0.29833		
x3	0.83578	0.12296	-0.08404		
x2	0.80752	0.02834	0.06234		
x4	0.63696	0.40753	-0.27961		
x6	0.10448	0.95713	0.11135		
x5	0.05286	0.08102	0.95666		
Variance Explained by Each Factor					
	Factor1	Factor2	Factor3		
2.5083723	1.1105019	1.1057272			
Final Communality Estimates: Total = 4.724601					
x1	x2	x3	x4	x5	x6
0.83317209	0.65677251	0.72070557	0.64998248	0.92455654	0.93941218

Nov 24, 09 11:30		EncIB.III		Page 5/6	
Enclosure B - data from a species of flea-beetle					
Factor analysis - three factors					
The FACTOR Procedure					
Initial Factor Method: Principal Components					
Prior Communality Estimates: ONE					
Eigenvalues of the Correlation Matrix: Total = 6 Average = 1					
	Eigenvalue	Difference	Proportion	Cumulative	
1	2.71448658	1.61266159	0.4524	0.4524	
2	1.10182498	0.19353518	0.1836	0.6361	
3	0.90828981	0.32705803	0.1514	0.7874	
4	0.58123178	0.14502894	0.0969	0.8843	
5	0.43620284	0.17823882	0.0727	0.9570	
6	0.25796402	0.0430	0.0430	1.0000	
3 factors will be retained by the NFACTOR criterion.					
Eigenvectors					
	1	2	3		
x1	0.52138	0.16585	-0.26743		
x2	0.46967	-0.05345	-0.24568		
x3	0.49695	-0.17797	-0.13038		
x4	0.43010	-0.28473	0.25381		
x5	0.09732	0.89308	-0.14852		
x6	0.25720	0.24350	0.87444		
Factor Pattern					
	Factor1	Factor2	Factor3		
x1	0.85901	0.17409	-0.25487		
x3	0.81876	-0.18681	-0.12425		
x2	0.77382	-0.05611	-0.23414		
x4	0.70862	-0.29888	0.24190		
x5	0.16034	0.93745	-0.14155		
x6	0.42375	0.25560	0.83338		
Variance Explained by Each Factor					
	Factor1	Factor2	Factor3		
2.7144866	1.1018250	0.9082898			
Final Communality Estimates: Total = 4.724601					
x1	x2	x3	x4	x5	x6
0.83317209	0.65677251	0.72070557	0.64998248	0.92455654	0.93941218

Nov 24, 09 11:28

EnclC.sas

Page 1/1

```
/* enclc.sas   Crted: 22-11-09 13:04 by BKE. Updt: 22-11-09 17:57 */
/* Purpose: */

title1 'Enclosure C - coffee data';

title2 'Print of data';
proc print data=stat2.coffee;
run;

title2 'Model "M"';
proc glm data=stat2.coffee;
class advertise;
model pounds=advertise price(advertise) price2(advertise);
run;

title2 'Model "H"';
proc glm data=stat2.coffee;
class advertise;
model pounds=advertise price price2 / solution;
run;
```

Enclosure C - coffee data
Model "M"

The GLM Procedure

Dependent Variable: pounds

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1903352.075	380670.415	1896.39	<.0001
Error	16	3211.744	200.734		
Corrected Total	21	1906563.818			

R-Square 0.998315
Coeff Var 2.322287
Root MSE 14.16806
pounds Mean 610.0909

Source	DF	Type I SS	Mean Square	F Value	Pr > F
advertise	1	27090.182	27090.182	134.96	<.0001
price(advertise)	2	1832219.682	916109.841	4563.80	<.0001
price2(advertise)	2	44042.211	22021.105	109.70	<.0001

Enclosure C - coffee data
Model "H"

The GLM Procedure

Class Level Information

Class	Levels	Values
advertise	2	0 1

Number of Observations Read 22
Number of Observations Used 22

Enclosure C - coffee data
Print of data

Obs	pounds	price	advertise	price2
1	1190	3.0	1	9.00
2	1033	3.2	1	10.24
3	897	3.4	1	11.56
4	789	3.6	1	12.96
5	706	3.8	1	14.44
6	595	4.0	1	16.00
7	512	4.2	1	17.64
8	433	4.4	1	19.36
9	395	4.6	1	21.16
10	304	4.8	1	23.04
11	243	5.0	1	25.00
12	1124	3.0	0	9.00
13	974	3.2	0	10.24
14	830	3.4	0	11.56
15	702	3.6	0	12.96
16	619	3.8	0	14.44
17	529	4.0	0	16.00
18	451	4.2	0	17.64
19	359	4.4	0	19.36
20	296	4.6	0	21.16
21	247	4.8	0	23.04
22	194	5.0	0	25.00

Enclosure C - coffee data
Model "M"

The GLM Procedure

Class Level Information

Class	Levels	Values
advertise	2	0 1

Number of Observations Read 22
Number of Observations Used 22

Enclosure C - coffee data
Model "H" 5

The GLM Procedure

Dependent Variable: pounds

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1903053.160	634351.053	3252.47	<.0001
Error	18	3510.658	195.037		
Corrected Total	21	1906563.818			

R-Square	Coeff Var	Root MSE	pounds Mean
0.998159	2.289093	13.96555	610.0909

Source	DF	Type I SS	Mean Square	F Value	Pr > F
advertise	1	27090.182	27090.182	138.90	<.0001
price	1	1832208.768	1832208.768	9394.18	<.0001
price2	1	43754.210	43754.210	224.34	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
advertise	1	27090.18182	27090.18182	138.90	<.0001
price	1	91776.81727	91776.81727	470.56	<.0001
price2	1	43754.21037	43754.21037	224.34	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	4439.681818 B	132.8897009	33.41	<.0001
advertise 0	-70.181818 B	5.9549299	-11.79	<.0001
advertise 1	0.000000 B			
price	-1466.202214	67.5904553	-21.69	<.0001
price2	126.238345	8.4282879	14.98	<.0001

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.