Problems for Multivariate Statistics 02409

Compiled by: Bjarne Kjær Ersbøll

September 7, 2010

Problem 7.

Consider the following observations from a $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution:

$$\begin{pmatrix} 5.1 \\ 3.5 \\ 6.3 \end{pmatrix}, \begin{pmatrix} 4.9 \\ 3.0 \\ 5.8 \end{pmatrix}, \begin{pmatrix} 4.6 \\ 3.2 \\ 7.1 \end{pmatrix}, \begin{pmatrix} 5.0 \\ 3.1 \\ 6.3 \end{pmatrix}, \begin{pmatrix} 5.4 \\ 3.6 \\ 6.5 \end{pmatrix}$$

Estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Find the axes in the contour ellipsoide.

Problem 8.

Let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ be normally distributed with parameters:

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 4 \\ 4 & 9 \end{pmatrix}$$

Determine the conditional distribution of $(x_1 \mid x_2)$ and $(x_2 \mid x_1)$. Plot the conditional means as functions of the conditional variables.

Problem E6

A laboratory method gives the possibility of determining the weight % of clay (x_1), "fine" sand (x_2) and "coarse" sand (x_3) in a gravel sample. (This does not determine all types of gravel i.e. $x_1 + x_2 + x_3$ is allways smaller than 100). Such a determination is of great help in finding out what the gravel can be used for.

Below the empirical variance covariance matrix for x_1, x_2, x_3 based on 31 determinations of the composition of gravel from different locations has been given.

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 6.8491 & -5.1952 & -1.5193 \\ -5.1952 & 6.8433 & 1.5384 \\ -1.5193 & 1.5384 & 2.3077 \end{pmatrix}$$

The eigenvalues and eigenvectors for $\hat{\boldsymbol{\Sigma}}$ are estimated at:

$$\lambda_1 = 12.50, \boldsymbol{p_1} = \begin{pmatrix} 0.6917 \\ -0.6917 \\ -0.2075 \end{pmatrix},$$

$$\lambda_2 = 1.65, \boldsymbol{p_2} = \begin{pmatrix} 0.6952 \\ 0.7156 \\ -0.0677 \end{pmatrix},$$

$$\lambda_3 = 1.85, \boldsymbol{p_3} = \begin{pmatrix} -0.1953 \\ 0.0974 \\ -0.9759 \end{pmatrix}.$$

Investigate if it can be assumed that 2 and 3 principal components do not contribute equally in explaining the variation of the material.

Try to give a verbal description of gravel samples where the first principal component is "large" and of samples where it is "small".

Problem E20

In an investigation of the general moral in a school class, Hartshore and May have found, the following correlations between outcome of 3 possibilities which measure the degree of dishonesty in situations involving "copying", "help" and "looking over the shoulder":

$$\begin{pmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 \\ 0.4 & 0,4 & 1 \end{pmatrix}$$

1) Find the linear combination of the possibilities which describes the largest possible fraction of the total variation and give the fraction.

2) Give a verbal interpretation of the linear combination.

Tip. You could consider using the fact that the roots of the equation $x^3 - 3\rho^2 x + 2\rho^3 = 0$ are: $\rho, \rho, -2\rho$.

Problem E11

In an analysis of the development (in time) of the strength of portland cement, correlations between the following 3 variables (based on 102 observations) were estimated.

C_3A	=	the amount of tricalciuminat in the cement in percent.
BLAINE	=	a measure of the specific surface and therefore also a measure of the "fine-grain'edness" of the crushed (powdered)cement. A large BLAINE number corresponds to finely powdered cement.
STYRKE 28	=	the strength of the cement after 28 days of hardening.

The obtained correlations were:

	C_3A	BLAINE	STYRKE 28
C_3A	1	0.192	−0.168
BLAINE	0.192	1	0.320
STYRKE 28	-0.168	0.320	1

It should be mentioned that a number of analyses of the distributions of the variables showed that they can be assumed normally distributed.

Using relevent statistical tests you should now give your opinion on the degree of correlation between C_3A -content and 28-day-strength.

1) Unconditionally.
2) For fixed value of BLAINE.
3) Also, give a (verbal) description and conclusion of the results of the analyses under 1) and 2).

Problem 29

Growth experiment with lucerne.

Experiment.

In the period 1968 – 69, at Højbakkegård, the experimental farm at the Royal Veterinary and Agricultural University of Denmark, a growth experiment with lucerne was performed. The offspring of a total of 176 cross-fertilisations was investigated. Nine variables were measured. The values given in the following for each of the nine variables are averages of between 15 and 20 plants (most values are averages of 20 plants).

Variable	Unit	Explanation
(1) type of growth	rank 1 – 9	1 = lying down 9 = standing
(2) regrowth after winter	-”-	1 = worst 9 = best
(3) ability to creep	-”-	1 = no runners 9 = most runners
(4) vigorousity	-”-	1 = weakest 9 = strongest
(5) time of flowering	-”-	1 = latest 9 = earliest
(6) height of plant	cm	
(7) weight of seed	g per plant	
(8) weight of plant	g per (dried) plant	
(9) percent seed	calculated from 7) and (8)	

For the variables (1) to (5) the unit used is a rank between 1 and 9. This is done because it is difficult to measure these variables directly. Experience shows this gives satisfactory results.

Data. The following table shows a few of the 176 observations.

Plant-group #	type of growth (1)	regrowth after (2)	ability to creep winter (3)	vigorous-ity (4)	time of flowering (5)	height of plant (6)	weight of seed (7)	weight of plant (8)	percent seed (9)
1	4.11	5.00	3.05	6.17	3.67	50.00	3.47	120.10	2.75
2	3.08	4.75	4.75	7.50	5.17	61.50	0.82	111.33	0.75
3	3.12	4.00	3.35	6.53	3.94	55.29	0.86	97.47	0.81
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
176	4.00	4.40	4.60	7.40	2.90	50.00	0.66	153.50	0.44

The experiments were performed in order to answer the following questions:

1) The growth of the plants is described above using 9 variables. If we consider the variables as en entity, which characteristics show the greatest variation?

2) Can one describe the variation of the growth using fewer than 9 variables?

A principal component analysis was therefore performed. This gave the following results:

Eigenvalues:
3.1270 2.3416 1.2166

Cumulated percent of eigenvalues:
0.3475 0.6076 0.7428

Eigenvectors:

	1	2	3	4	5	6	7	8	9
Vector 1	-.0182	.2665	.1731	.3788	.3527	.4516	.3885	.4162	.3204
Vector 2	.0148	-.4346	-.4662	-.3638	.2130	.0518	.4108	-.1734	.4633
Vector 3	.7465	.1993	.3601	-.1011	.2418	-.3056	.1033	-.2611	.1799

This was followed by a factor analysis which resulted in:

Rotated factor matrix (3 factors)		
Variable 1		
0.17201	-0.13516	0.79480
Variable 2		
-0.2246	-0.84236	0.05135
Variable 3		
-0.13652	-0.81818	0.26920
Variable 4		
0.12099	-0.80729	-0.32196
Variable 5		
0.72809	-0.17618	0.07281
Variable 6		
0.56871	-0.34554	-0.56106
Variable 7		
0.93280	0.05221	-0.08480
Variable 8		
0.31501	-0.58527	-0.50336
Variable 9		
0.91295	0.16875	0.03397

	Communalities
Variable	Original
1	0.67956
2	0.71271
3	0.76053
4	0.77002
5	0.56646
6	0.75763
7	0.88004
8	0.69514
9	0.86310

1) Give an interpretation of the factors found.

Problem E44

In some attempts of geological studies by means of ”remote sensing” one has used the measurements from the American Landsat-4 satellite. This satellite measures simultaneously the values of 7 spectral bands in the electromagnetic spectrum of square areas (so-called pixels) of $35m \times 35m$ on the surface of the earth.

There is special interest in differentiating between two types of geology:

bed group 10 (a type of slate) and
bed group 13 (a type of quartsite).

One assumes that 3 of the 7 bands of the satellite are especially suited to differentiate between the two bed groups. They are:

$X_1 = \text{Band } 1 \sim 0.45 - 0.52\,\mu m$
$X_2 = \text{Band } 3 \sim 0.63 - 0.69\,\mu m$
$X_3 = \text{Band } 4 \sim 0.76 - 0.90\,\mu m$

From an area on Ymer Island in East Greenland one has measured 493 pixels from bed group 10 and 73 pixels from bed group 13.

The estimated mean vectors, where μ_1 is from: bed group 10 and μ_2 from bed group 13 and the common variance-covariance matrix are given below. Furthermore other measures are given as a help.

$$\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} 75.9 \\ 45.3 \\ 41.8 \end{pmatrix}, \qquad \hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 78.2 \\ 39.5 \\ 44.1 \end{pmatrix}$$

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 26.9 & 13.5 & 3.9 \\ 13.5 & 17.5 & 12.2 \\ 3.9 & 12.2 & 16.6 \end{pmatrix}$$

$$\hat{\boldsymbol{\Sigma}}^{-1} = \begin{pmatrix} 0.0788 & -0.0987 & 0.0543 \\ -0.0987 & 0.2415 & -0.1547 \\ 0.0543 & -0.1547 & 0.1615 \end{pmatrix}$$

$$\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2) = \begin{pmatrix} -0.88 \\ 1.98 \\ -1.39 \end{pmatrix}$$

$$\hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}_1 = 310.4\,, \qquad \hat{\boldsymbol{\mu}}_2'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}_2 = 396.5$$

1) Are these measurements reasonable to use in an attempt to differentiate between *bed group* 10 and *bed group* 13?
2) Would it be reasonable to drop bands 1 and 3 and use only band 4 in an attempt to differentiate between *bed group* 10 and *bed group* 13?
3) If one uses all the information from all 3 bands, what is then the Bayes-classification rule, assuming equal priors?
4) What is the Bayes-classification rule corresponding to using prior probabilities which are proportional to the number of observations?
5) Using the rule found under 3), what will

$$\boldsymbol{x} = \begin{pmatrix} 74 \\ 43 \\ 40 \end{pmatrix}$$

be classified as?

Problem E9.

A team of medical doctors has been working on finding an easier way of diagnosing two similar but different diseases A and B.

Presently the team is working on the hypothesis that it is possible to distinguish A from B based on 2 relatively simple measurements on a patient. The measures are called x_1 and x_2.

Based on an old but difficult diagnosis technique the team has sure knowledge of 106 patients with the disease A and 81 with the disease B. Measurements of x_1 and x_2 on all these patients gave the following empirical means for x_1 and x_2:

	Disease	
	A	B
	$n_1 = 106$	$n_2 = 81$
$\bar{x}_1$	380	400
$\bar{x}_2$	120	90

and the following empirical variance-covariance matrix common for A and B:

	x_1	x_2
x_1	25	40
x_2	40	100

(x_1 and x_2 are assumed normal).

1) Test if there is a difference between the means of the two frequency distributions $f_1(x_1, x_2 \mid A)$ and $f_2(x_1, x_2 \mid B)$.
2) Determine an optimal decision rule with respect to discriminating between A and B when the maximal probability of misclassification must be minimised.
3) Determine if a patient with measurements $x_1 = 390$ and $x_2 = 100$ is classified as having disease A or B.

Problem 9.

Write the following statistical models as a general linear model.

1) One sided analysis of variance with 3 groups. The number of observations in each group is 2, 4 and 3 respectively.
2) Two sided analysis of variance with 2 rows and 2 columns and 1 observation in each cell.
3) Linear regression, where the line passes through $(0, 0)$. There are n observations in all.

Problem 10.

In an econometric investigation, it is assumed that a certain phenomena y can be described using a polynomial:

$$ax^2 + bx,$$

plus a random error. The variable x is an amount of crowns in thousands. The following corresponding values of y and x have been observed

$$\begin{array}{rrrrr} x: & 0, & 1, & 1, & 2 \\ y: & \frac{1}{2}, & 0, & -\frac{1}{2}, & -2 \end{array}$$

Estimate a and b and their respective errors.

Problem E 19.

While examining the sulphur emission in Europe one has estimated the yearly emission for 3 areas by means of an ordinary "general linear model". These estimates are given together with their standard errors and the correlations between them in the table below.

Area	Emission in $\frac{10^3\ ts\ S}{127^2\ km^2}$	Standard error	Correlation		
Denmark	48	8	1		
East Germany	70	8	0.0	1	
Benelux, Ruhr & South Germany	192	7	0.0	-0.2	1

The estimate of the residual standard deviation was 18. Emissions are independent variables in the model. The observations are measurements of the SO_4-content in the air at a number of measuring stations in Europe, and the model is a linear relationship between emissions and measurements. The estimates are based on approx. 10000 observations.

By means of the model one finds - under certain meteorological conditions - that the measurement at a SO_2-measuring station, measuring SO_2 content in air volumes, which have passed Denmark, crossed the Baltic, crossed East Germany, South Germany and Ruhr can be estimated by the following linear combination of the estimated emissions

$$48 \times 0.4 + 70 \times 0.6 + 192 \times 0.8 = 214.80$$

where the unit on the figures 0.4, 0.6 and 0.8 make the unit on the result equal $SO_2\ \mu g/m^3$ air.

1) Find - assuming 0.4, 0.6 and 0.8 are exact - a 95% confidence interval for the expected value corresponding to 0.4, 0.6 and 0.8.
2) Find a 95% prediction interval for an observation corresponding to the above mentioned values.
3) Comment the two results found under 1) and 2).

Problem 12.

In the law for common schools, § 50 _rules *are given to specify when* the commune is obliged to provide transport to the school.

Here, the communes transport obligation is given as the sum of the pupils distances (as the crow flies) from home to school for those pupils who have the right of transport.

We wish to investigate the following models, which describe a *school* transport-obligation:

$$\begin{array}{llll} I & TF & = & a + b \cdot E \cdot \sqrt{A} + c \cdot E \\ II & TF & = & d + f \cdot E \cdot \sqrt{A} + g \cdot \sqrt{A} \end{array}$$

where:

$$\begin{array}{lll} TF & = & \text{transport obligation} \\ E & = & \text{number of pupils having the right of transport} \\ A & = & \text{area of school district} \end{array}$$

Assume $TF \in N\left(\underline{\mu}, \sigma^2 \cdot \underline{\underline{I}}\right)$

Based on the observations given later do the following (using SAS):

1) Estimate coefficients a and b.
2) Give an estimate for the residual error in model I.
3) Check if c and/or a can be assumed to be 0. First test $c = 0$ and then $a = 0$.
4) Do the same analyses (1-3) for model II, now testing i) $d = 0$ and ii) $g = 0$.
5) Is one of the models better than the other?

School #	*Pupils to transport*	*Areal School District*	*Transport obligation*
1	286	155.0	2 127
2	58	85.0	442
3	75	97.5	516
4	167	196.4	1 458
5	22	24.6	130

Problem E1

As part of large rationalisation investigation the board of a company has been interested in comparing the acidity of a product from 2 of the company's factories. Because of the special product, the pH-measurement has a large error, so the pH-measurment was measured 6 times at each factory. The results were:

	Result of 6 measurements of pH						Sum of measurements	Sum of squares of measurements
Factory A	1.33	1.11	-0.73	0.99	-0.29	1.43	3.82	6.6038
Factory B	-1.05	-0.29	1.29	0.55	0.09	1.83	2.42	6.5102

Since factory B's analysis lab only has an old pH-meter, it can be assumed that the variance of the measurement from factory B are twice as big as the variance of the measurements from factory A.

One is now interested in knowing if the two analysed samples have the same pH-value and secondly if a common value might be 0.

Formulate a suitable model for the analysis of the observations, and perform a statistical analysis in order to solve the above mentioned problems.

Problem 13.

Given the linear model

$$E(Y_i) = \alpha + \beta x_i \qquad \begin{aligned} i &= 1,2,3,4 \\ x_1 &= -2 \\ x_2 &= -1 \\ x_3 &= 1 \\ x_4 &= 2 \end{aligned}$$

$$V(Y_i) = x_i^2 \sigma^2$$

One has the following observations:

$$Y_1 \ldots Y_4 = 3,\ 6,\ 4,\ 5$$

Investigate, assuming normality, the hypothesis $\beta = 0$.

Problem 53

We consider data from exercise 17 and the model from Anderson & Bancrofts book. In the following table parameters are estimated using all the observations and using only the 26 first observations.

	27 obs.	26 obs.	S. E. 27 obs.	S. E. 26 obs.
$\hat{b}_0$	82.1733	82.2390	20.3383	20.7216
$\hat{b}_1$	2.4629	2.3498	4.7221	4.8178
$\hat{b}_2$	-75.3781	-77.3402	39.1439	40.1277
$\hat{b}_3$	1.5839	1.5922	3.1222	3.1810
$\hat{b}_{12}$	-1.3742	0.4259	21.2645	22.0454
$\hat{\sigma}^2$	101.94354	105.81677	—	—

Table 1

Furthermore the standard deviations of the estimates are given in the two cases. Finally, the error of the residual of observation 27 is estimated at 9.3312.

1) Find the predicted value $\hat{y}_{27}$ and determine an estimate of the standard deviation of this value.
2) Find a 95% confidence interval for the expected mean of y_{27} and find a 95% prediction interval corresponding to:

$$(1,\, x_1,\, x_2,\, x_3,\, x_1x_2) = (1.205,\, -0.474,\, 7.6,\, 0.9717)$$

3) Compare these intervals with y_{27}.
4) How much does the predicted value change if observation 27 is deleted?
5) What is DFFITS estimated at?
6) What is DFBETAS for x_1x_2 estimated at?

Problem 17.

Below are given data from an investigation of the amount of vitamin B_2 in beetroot-leaves. The data originate from "Annual Progress Report on the Soils-weather Project, 1948" by J.T. Wakeley, University of North Carolina (Raleigh) Institute of Statistics Mimeo Series 19 (1949). The variables shown in the table are:

X_1 = influx of sun-light in relative gram-calories per minute during the last half day with sun-light. (Coded by division by 100)

X_2 = average ability of the soil (in which the beetroots are grown) to retain moisture. Note: 3 different types of soil were used (special measuring unit).

X_3 = air-temperature in degrees Fahrenheit. (Coded by division by 10)

Y = milli-grams of vitamin B_2 per gram beetroot leaf.

In the book "Statistical Theory and Research" R.L. Anderson and T.A. Bancroft find a regression equation of the form:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + \varepsilon$$

The problem is now to try to construct a regression equation from the data given. You could consider including 2nd order terms (e.g. X_1^2, X_1X_2 etc.). In all it is possible to construct 512 regression equations given three independent variables, their squared values, and all possible products of two variables. Compare the equation you find with the one proposed by Anderson and Bancroft.

X_1	X_2	X_3	Y
1.76	0.070	7.8	110.4
1.55	0.070	8.9	102.8
2.73	0.070	8.9	101.0
2.73	0.070	7.2	108.4
2.56	0.070	8.4	100.7
2.80	0.070	8.7	100.3
2.80	0.070	7.4	102.0
1.84	0.070	8.7	93.7
2.16	0.070	8.8	98.9
1.98	0.020	7.6	96.6
0.59	0.020	6.5	99.4
0.80	0.020	6.7	96.2
0.80	0.020	6.2	99.0
1.05	0.020	7.0	88.4
1.80	0.020	7.3	75.3
1.80	0.020	6.5	92.0
1.77	0.020	7.6	82.4
2.30	0.020	8.2	77.1
2.03	0.474	7.6	74.0
1.91	0.474	8.3	65.7
1.91	0.474	8.2	56.8
1.91	0.474	6.9	62.1
0.76	0.474	7.4	61.0
2.13	0.474	7.6	53.2
2.13	0.474	6.9	59.4
1.51	0.474	7.5	58.7
2.05	0.474	7.6	58.0

Problem E24

In the production of “ftalsyreanhydrid” we use - among other things - “naftalen” to start of with. If one wants a maximal productivity in a chemical reactor, it is of prime importance, that the feeding speed of naftalen is chosen correctly. At too small a speed one gets to small a production, and at too large a speed the process can be “dampened” so that the production-yield again drops.

In 7 experiments the following corresponding observations of feeding speed and productivity per hour were found. Both variables are given in a relative scale.

Feeding speed	5	20	35	50	65	80	95
Productivity/hour	10	30	40	55	70	65	50

1) Determine a suitable relationship between feeding speed and productivity.
2) Estimate - using the result from 1) the feeding speed which will maximise the productivity.
3) Find a confidence interval for this productivity. It can be assumed that the feeding speed is non-stochastic in nature.

Problems E50+E51+E52

In all problems we consider a data set published by Hermon Bumpus in 1898. After a severe storm on Feb. 1. 1898 a number of dying sparrows were taken to the biological laboratory at Browns University, Rhode Island. About half of the birds did not survive the impact of the storm and Bumpus used the incidence to study the effect of "natural selection" (in Darwin's sense) among birds.

The data set, which we will use in the following is only part of Bumpus' original data.

The data set comprises measurements of 5 parameters of 49 female sparrows together with an indicator of if they survived or not. A listing of data and the results of some analyses are found in the enclosure. The variable names are:

BIRD = bird number
SURVIVE = survived (Y = yes, N = no)
TOTAL = total length
ALAR = wing span
BEAK HEAD = length of beak and head
HUMERUS = length of upper arm"
STERNUM = length of breast bone

All lengths in mm.

E50

In this problem we mostly consider output from "proc reg".

1a) Give the analysed model in a notation, which corresponds to the one used in the "statistic 2" book.

1b) Give the necessary assumptions in order to determine a least squares solution from the given output.

1c) Give the necessary assumptions in order to determine a maximum likelihood solution from the given output.

2a) Give, using these assumptions, the unbiased estimate of the residual variance and the maximum likelihood estimate of the other parameters in the model.

2b) Give the residual for observation 3, and indicate how the estimate of the residual variance in 2a) is found.

2c) Give the multiple correlation coefficient, and indicate how it is found.

3) Give the observation, which has the most critical value of Cook's D, and do a statistical evaluation of if is is extreme.

E51

In this problem we mostly consider output from "proc discrim". The option "simple" requests the computation of simple statistics (e.g. means and variances) for all data and per class.

Given a new storm of the same type and strength as that of Feb. 1. 1898, it would be nice to be able to classify a bird in one of two groups: "will survive" and "will not survive".

1a) Using the information collected by Bumpus we now want a Bayes classification rule for linear discriminant analysis, where the prior probabilities are assumed equal.

1b) Using the rule found in 1a) we now want to classify observation 49.

2) Perform a test to show if it is reasonable to use the given measurements to discriminate between surviving and non-surviving birds.

3) We assume that ALAR is sufficient to perform a classification. Using the loss-function below we now want a (linear) minimax classification rule. (it is sufficient to find "c" with 1 decimal.)

Loss-function:

		We choose: will survive	We choose: will not survive
Truth:	will survive	0	1
	will not survive	5	0

E52

In this problem we mostly consider output from the 2 versions of "proc princomp".

1a) Briefly describe the difference between analysing the data by means of principal components on the variance-covariance matrix and on the correlation matrix.

1b) Give a (brief) description of the consequences it has for the first eigenvector in the two analysed cases.

2a) Consider the analysis performed on the variance-covariance matrix. Give for observation 1 the value of the first principal component. Give the variance of the first principal component.

2b) Consider the analysis performed on the correlation matrix. Give for observation 1 the value of the first principal component. Give the variance of the first principal component.

3a) Consider the analysis performed on the variance-covariance matrix. Perform a test for if the 3 smallest eigenvalues could be equal.

3b) Give a (brief) interpretation of principal component 1 and 2 in this case.

```
*******************************************************************
*                                                                 *
*  SAS - program til eksamen i 0411 Statistik 2, 2. januar 1991.  *
*                                                                 *
*******************************************************************;

proc print data=sasdata.bumpus;
  var bird survive total alar beakhead humerus sternum;
title1 'Datalist. Alle laengder er i mm. (All lengths in mm.)';
title2 'BIRD    =fugl nummer (bird number)                           ';
title3 'SURVIVE =overlevet, Y=ja N=nej (survived, Y=yes N=no)        ';
title4 'TOTAL   =total laengde (total length)                        ';
title5 'ALAR    =laengde af vinger udstrakt (wing span)              ';
title6 'BEAKHEAD=laengde af naeb og hoved (length of beak and head)  ';
title7 'HUMERUS =laengde af "overarmsknoglen" (length of "upper arm")';
title8 'STERNUM =laengde af brystben (length of breastbone)          ';

proc reg data=sasdata.bumpus;
  model total=alar beakhead humerus sternum / covb r influence;
title1 'proc reg';

proc discrim data=sasdata.bumpus
  simple wcov wcorr pcov pcorr pool=yes;
  class survive;
  var total alar beakhead humerus sternum;
title1 'proc discrim';

proc princomp data=sasdata.bumpus cov;
  var total alar beakhead humerus sternum;
title1 'proc princomp - 1';

proc princomp data=sasdata.bumpus;
  var total alar beakhead humerus sternum;
title1 'proc princomp - 2';
```



```
          Datalist. Alle laengder er i mm. (All lengths in mm.)            1
      BIRD    =fugl nummer (bird number)
      SURVIVE =overlevet, Y=ja N=nej (survived, Y=yes N=no)
      TOTAL   =total laengde (total length)
      ALAR    =laengde af vinger udstrakt (wing span)
      BEAKHEAD=laengde af naeb og hoved (length of beak and head)
      HUMERUS =laengde af "overarmsknoglen" (length of "upper arm")
      STERNUM =laengde af brystben (length of breastbone)
```

Obs	bird	survive	total	alar	beakhead	humerus	sternum
1	1	Y	156	245	31.6	18.5	20.5
2	2	Y	154	240	30.4	17.9	19.6
3	3	Y	153	240	31.0	18.4	20.6
4	4	Y	153	236	30.9	17.7	20.2
5	5	Y	155	243	31.5	18.6	20.3
6	6	Y	163	247	32.0	19.0	20.9
7	7	Y	157	238	30.9	18.4	20.2
8	8	Y	155	239	32.8	18.6	21.2
9	9	Y	164	248	32.7	19.1	21.1
10	10	Y	158	238	31.0	18.8	22.0
11	11	Y	158	240	31.3	18.6	22.0
12	12	Y	160	244	31.1	18.6	20.5
13	13	Y	161	246	32.3	19.3	21.8
14	14	Y	157	245	32.0	19.1	20.0
15	15	Y	157	235	31.5	18.1	19.8
16	16	Y	156	237	30.9	18.0	20.3
17	17	Y	158	244	31.4	18.5	21.6
18	18	Y	153	238	30.5	18.2	20.9
19	19	Y	155	236	30.3	18.5	20.1
20	20	Y	163	246	32.5	18.6	21.9
21	21	Y	159	236	31.5	18.0	21.5
22	22	N	155	240	31.4	18.0	20.7
23	23	N	156	240	31.5	18.2	20.6
24	24	N	160	242	32.6	18.8	21.7
25	25	N	152	232	30.3	17.2	19.8
26	26	N	160	250	31.7	18.8	22.5
27	27	N	155	237	31.0	18.5	20.0
28	28	N	157	245	32.2	19.5	21.4
29	29	N	165	245	33.1	19.8	22.7
30	30	N	153	231	30.1	17.3	19.8
31	31	N	162	239	30.3	18.0	23.1
32	32	N	162	243	31.6	18.8	21.3
33	33	N	159	245	31.8	18.5	21.7
34	34	N	159	247	30.9	18.1	19.0
35	35	N	155	243	30.9	18.5	21.3
36	36	N	162	252	31.9	19.1	22.2
37	37	N	152	230	30.4	17.3	18.6
38	38	N	159	242	30.8	18.2	20.5
39	39	N	155	238	31.2	17.9	19.3
40	40	N	163	249	33.4	19.5	22.8
41	41	N	163	242	31.0	18.1	20.7
42	42	N	156	237	31.7	18.2	20.3
43	43	N	159	238	31.5	18.4	20.3
44	44	N	161	245	32.1	19.1	20.8
45	45	N	155	235	30.7	17.7	19.6
46	46	N	162	247	31.9	19.1	20.4
47	47	N	153	237	30.6	18.6	20.4
48	48	N	162	245	32.5	18.5	21.1
49	49	N	164	248	32.3	18.8	20.9

proc reg 2

The REG Procedure
Model: MODEL1
Dependent Variable: total

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	406.26511	101.56628	19.04	<.0001
Error	44	234.71448	5.33442		
Corrected Total	48	640.97959			

Root MSE	2.30964	R-Square	0.6338
Dependent Mean	157.97959	Adj R-Sq	0.6005
Coeff Var	1.46198		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	25.60197	16.88961	1.52	0.1367
alar	1	0.35140	0.10571	3.32	0.0018
beakhead	1	1.21323	0.66745	1.82	0.0759
humerus	1	-0.58482	1.12423	-0.52	0.6055
sternum	1	0.97033	0.42841	2.26	0.0285

Covariance of Estimates

Variable	Intercept	alar	beakhead	humerus	sternum
Intercept	285.2589583	-1.10429589	-3.541039328	4.2988145962	0.6408910102
alar	-1.10429589	0.0111748217	-0.014025139	-0.057202396	-0.00455056
beakhead	-3.541039328	-0.014025139	0.4454838653	-0.351527088	-0.028636126
humerus	4.2988145962	-0.057202396	-0.351527088	1.2638823741	-0.133422954
sternum	0.6408910102	-0.00455056	-0.028636126	-0.133422954	0.1835346196

proc reg 3

The REG Procedure
Model: MODEL1
Dependent Variable: total

Output Statistics

Obs	Dep Var total	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual
1	156.0000	159.1066	0.5245	-3.1066	2.249	-1.381
2	154.0000	155.3713	0.6991	-1.3713	2.201	-0.623
3	153.0000	156.7771	0.4184	-3.7771	2.271	-1.663
4	153.0000	155.2714	0.5929	-2.2714	2.232	-1.018
5	155.0000	158.0299	0.4586	-3.0299	2.264	-1.338
6	163.0000	160.3904	0.5541	2.6096	2.242	1.164
7	157.0000	155.5649	0.5417	1.4351	2.245	0.639
8	155.0000	159.0748	0.9913	-4.0748	2.086	-1.953
9	164.0000	161.7266	0.6803	2.2734	2.207	1.030
10	158.0000	157.1988	0.9327	0.8012	2.113	0.379
11	158.0000	158.3826	0.6531	-0.3826	2.215	-0.173
12	160.0000	158.0901	0.5656	1.9099	2.239	0.853
13	161.0000	161.1008	0.6037	-0.1008	2.229	-0.0452
14	157.0000	158.7558	0.8061	-1.7558	2.164	-0.811
15	157.0000	155.0259	0.7179	1.9741	2.195	0.899
16	156.0000	155.5444	0.4454	0.4556	2.266	0.201
17	158.0000	159.5799	0.5245	-1.5799	2.249	-0.702
18	153.0000	155.8758	0.6107	-2.8758	2.227	-1.291
19	155.0000	153.9786	0.9568	1.0214	2.102	0.486
20	163.0000	161.8498	0.7872	1.1502	2.171	0.530
21	159.0000	157.0853	0.7983	1.9147	2.167	0.883
22	155.0000	157.5934	0.5484	-2.5934	2.244	-1.156
23	156.0000	157.5007	0.4246	-1.5007	2.270	-0.661
24	160.0000	160.2545	0.7213	-0.2545	2.194	-0.116
25	152.0000	153.0421	0.8451	-1.0421	2.149	-0.485
26	160.0000	162.7501	0.9392	-2.7501	2.110	-1.303
27	155.0000	155.0822	0.6785	-0.0822	2.208	-0.0372
28	157.0000	160.1230	0.8121	-3.1230	2.162	-1.444
29	165.0000	162.3009	1.0834	2.6991	2.040	1.323
30	153.0000	152.3896	0.8098	0.6104	2.163	0.282
31	162.0000	158.2362	1.3769	3.7638	1.854	2.030
32	162.0000	159.0046	0.4157	2.9954	2.272	1.318
33	159.0000	160.5136	0.5864	-1.5136	2.234	-0.678
34	159.0000	157.7386	1.1966	1.2614	1.975	0.639
35	155.0000	158.3308	0.5986	-3.3308	2.231	-1.493
36	162.0000	163.2290	0.9059	-1.2290	2.125	-0.578
37	152.0000	151.2378	0.9479	0.7622	2.106	0.362
38	159.0000	157.2572	0.5416	1.7428	2.245	0.776
39	155.0000	155.3479	0.6741	-0.3479	2.209	-0.158
40	163.0000	164.3429	0.9346	-1.3429	2.112	-0.636
41	163.0000	157.7524	0.5349	5.2476	2.247	2.336
42	156.0000	156.3980	0.6137	-0.3980	2.227	-0.179
43	159.0000	156.3898	0.4951	2.6102	2.256	1.157
44	161.0000	159.6534	0.5801	1.3466	2.236	0.602
45	155.0000	154.0952	0.5993	0.9048	2.231	0.406
46	162.0000	159.7254	0.7217	2.2746	2.194	1.037
47	153.0000	154.9266	0.8475	-1.9266	2.149	-0.897
48	162.0000	160.7807	0.7661	1.2193	2.179	0.560
49	164.0000	161.2227	0.6640	2.7773	2.212	1.255

proc reg 4

The REG Procedure
Model: MODEL1
Dependent Variable: total

Output Statistics

Obs	-2-1 0 1 2	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS
1	| **| |	0.021	−1.3959	0.0516	0.9477	−0.3255
2	| *| |	0.008	−0.6186	0.0916	1.1815	−0.1965
3	| ***| |	0.019	−1.6981	0.0328	0.8384	−0.3128
4	| **| |	0.015	−1.0180	0.0659	1.0661	−0.2704
5	| **| |	0.015	−1.3510	0.0394	0.9487	−0.2737
6	| |** |	0.017	1.1687	0.0576	1.0180	0.2888
7	| |* |	0.005	0.6349	0.0550	1.1330	0.1532
8	| ***| |	0.172	−2.0206	0.1842	0.8737	−0.9602
9	| |** |	0.020	1.0307	0.0868	1.0873	0.3177
10	| | |	0.006	0.3755	0.1631	1.3186	0.1657
11	| | |	0.001	−0.1708	0.0800	1.2152	−0.0503
12	| |* |	0.009	0.8502	0.0600	1.0979	0.2148
13	| | |	0.000	−0.0447	0.0683	1.2038	−0.0121
14	| *| |	0.018	−0.8080	0.1218	1.1847	−0.3009
15	| |* |	0.017	0.8973	0.0966	1.1318	0.2934
16	| | |	0.000	0.1988	0.0372	1.1598	0.0391
17	| *| |	0.005	−0.6983	0.0516	1.1180	−0.1628
18	| **| |	0.025	−1.3012	0.0699	0.9943	−0.3567
19	| | |	0.010	0.4816	0.1716	1.3183	0.2192
20	| |* |	0.007	0.5253	0.1162	1.2293	0.1904
21	| |* |	0.021	0.8812	0.1195	1.1650	0.3246
22	| **| |	0.016	−1.1604	0.0564	1.0190	−0.2836
23	| *| |	0.003	−0.6567	0.0338	1.1045	−0.1228
24	| | |	0.000	−0.1147	0.0975	1.2411	−0.0377
25	| | |	0.007	−0.4806	0.1339	1.2610	−0.1890
26	| **| |	0.067	−1.3141	0.1654	1.1039	−0.5849
27	| | |	0.000	−0.0368	0.0863	1.2276	−0.0113
28	| **| |	0.059	−1.4630	0.1236	1.0040	−0.5495
29	| |** |	0.099	1.3350	0.2200	1.1740	0.7091
30	| | |	0.002	0.2792	0.1229	1.2675	0.1045
31	| |**** |	0.454	2.1076	0.3554	1.0645	1.5649
32	| |** |	0.012	1.3299	0.0324	0.9478	0.2434
33	| *| |	0.006	−0.6733	0.0645	1.1379	−0.1768
34	| |* |	0.030	0.6342	0.2684	1.4637	0.3842
35	| **| |	0.032	−1.5149	0.0672	0.9273	−0.4065
36	| *| |	0.012	−0.5741	0.1538	1.2761	−0.2448
37	| | |	0.005	0.3583	0.1684	1.3291	0.1612
38	| |* |	0.007	0.7727	0.0550	1.1080	0.1864
39	| | |	0.000	−0.1558	0.0852	1.2228	−0.0475
40	| *| |	0.016	−0.6315	0.1637	1.2810	−0.2794
41	| |**** |	0.062	2.4668	0.0536	0.6116	0.5873
42	| | |	0.000	−0.1768	0.0706	1.2027	−0.0487
43	| |** |	0.013	1.1616	0.0460	1.0075	0.2550
44	| |* |	0.005	0.5979	0.0631	1.1488	0.1552
45	| | |	0.002	0.4018	0.0673	1.1805	0.1079
46	| |** |	0.023	1.0376	0.0976	1.0986	0.3413
47	| *| |	0.025	−0.8947	0.1346	1.1822	−0.3529
48	| |* |	0.008	0.5552	0.1100	1.2163	0.1952
49	| |** |	0.028	1.2640	0.0827	1.0190	0.3794

proc reg 5

The REG Procedure
Model: MODEL1
Dependent Variable: total

Output Statistics

	-------------------------DFBETAS-------------------------				
Obs	Intercept	alar	beakhead	humerus	sternum
1	0.1662	−0.2278	−0.0091	0.0979	0.1228
2	−0.0348	−0.1015	0.0918	0.0269	0.0740
3	−0.1220	0.0239	0.1737	−0.1196	0.0229
4	−0.1151	0.0208	−0.0779	0.1572	−0.0297
5	0.0586	−0.0737	0.0323	−0.0665	0.1711
6	−0.1772	0.1239	−0.0213	0.0678	−0.1200
7	0.0848	−0.0602	−0.0678	0.0963	−0.0492
8	0.0728	0.4946	−0.8320	0.1921	−0.0527
9	−0.2424	0.0877	0.1607	−0.0238	−0.1088
10	0.0798	−0.0924	−0.0819	0.0969	0.0828
11	−0.0158	0.0185	0.0152	−0.0096	−0.0375
12	−0.0274	0.1039	−0.1332	0.0615	−0.0702
13	0.0046	0.0022	0.0003	−0.0064	−0.0011
14	0.0814	−0.0041	0.0048	−0.1684	0.2336
15	0.1139	−0.1758	0.1630	0.0064	−0.1064
16	0.0234	−0.0093	−0.0011	−0.0062	0.0003
17	0.0369	−0.0747	0.0450	0.0543	−0.0923
18	−0.2099	0.0478	0.2521	−0.0921	−0.1255
19	0.1324	−0.0904	−0.1482	0.1728	−0.0428
20	−0.1073	0.0578	0.1165	−0.1316	0.0726
21	0.1026	−0.1397	0.1454	−0.1420	0.1986
22	0.0097	−0.0657	−0.1247	0.2207	−0.0569
23	−0.0022	−0.0029	−0.0598	0.0640	0.0005
24	0.0083	0.0155	−0.0280	0.0064	−0.0098
25	−0.1112	0.0226	−0.0302	0.1047	−0.0295
26	0.2886	−0.3948	0.1666	0.2139	−0.3150
27	−0.0060	0.0063	0.0033	−0.0083	0.0048
28	0.0710	0.1978	0.1016	−0.4547	0.1109
29	−0.1099	−0.4123	0.1550	0.3754	0.1568
30	0.0813	−0.0302	−0.0039	−0.0313	0.0161
31	0.4187	0.0727	−0.6537	−0.3599	1.3893
32	−0.0178	−0.0240	−0.0794	0.1091	0.0425
33	0.0796	−0.0871	−0.0291	0.1093	−0.0913
34	−0.1318	0.3058	−0.0705	−0.1003	−0.2216
35	−0.0272	−0.1454	0.3021	−0.0461	−0.1446
36	0.1543	−0.1809	0.0868	0.0365	−0.0673
37	0.0998	−0.0558	0.0412	−0.0243	−0.0678
38	0.0068	0.1096	−0.1026	−0.0317	−0.0066
39	−0.0053	−0.0040	−0.0185	0.0147	0.0299
40	0.1689	0.0266	−0.1515	0.0218	−0.0937
41	−0.0441	0.3705	−0.1325	−0.3194	0.0926
42	−0.0119	0.0255	−0.0330	0.0076	0.0089
43	0.0915	−0.1529	0.0805	0.0726	−0.0882
44	−0.0558	−0.0098	0.0052	0.0869	−0.0775
45	0.0607	−0.0178	0.0151	−0.0288	−0.0302
46	−0.1478	0.1033	−0.0657	0.1575	−0.2217
47	−0.2026	0.1648	0.2220	−0.2858	0.0500
48	−0.1105	0.0565	0.1502	−0.1301	−0.0002
49	−0.2967	0.2353	0.1392	−0.1368	−0.1135

Sum of Residuals	0
Sum of Squared Residuals	234.71448
Predicted Residual SS (PRESS)	302.12567

proc discrim 6

The DISCRIM Procedure

Observations	49	DF Total	48
Variables	5	DF Within Classes	47
Classes	2	DF Between Classes	1

Class Level Information

survive	Variable Name	Frequency	Weight	Proportion	Prior Probability
N	N	28	28.0000	0.571429	0.500000
Y	Y	21	21.0000	0.428571	0.500000

```
/><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>></
```

proc discrim 7

The DISCRIM Procedure
Within-Class Covariance Matrices

survive = N, DF = 27

Variable	total	alar	beakhead	humerus	sternum
total	15.06878307	17.19047619	2.24285714	1.74603175	2.93068783
alar	17.19047619	32.55026455	3.39788360	2.95026455	4.06560847
beakhead	2.24285714	3.39788360	0.72841270	0.46955026	0.55902116
humerus	1.74603175	2.95026455	0.46955026	0.43443122	0.50588624
sternum	2.93068783	4.06560847	0.55902116	0.50588624	1.32099206

---

survive = Y, DF = 20

Variable	total	alar	beakhead	humerus	sternum
total	11.04761905	9.10000000	1.55666667	0.87000000	1.28619048
alar	9.10000000	17.50000000	1.91000000	1.31000000	0.88000000
beakhead	1.55666667	1.91000000	0.53133333	0.18900000	0.23966667
humerus	0.87000000	1.31000000	0.18900000	0.17600000	0.13250000
sternum	1.28619048	0.88000000	0.23966667	0.13250000	0.57490476

```
/><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>></
```

proc discrim 8

The DISCRIM Procedure

Pooled Within-Class Covariance Matrix, DF = 47

Variable	total	alar	beakhead	humerus	sternum
total	13.35764944	13.74772036	1.95086120	1.37325228	2.23090172
alar	13.74772036	26.14589666	2.76474164	2.25227964	2.71003040
beakhead	1.95086120	2.76474164	0.64454914	0.35016717	0.42312563
humerus	1.37325228	2.25227964	0.35016717	0.32446049	0.34699848
sternum	2.23090172	2.71003040	0.42312563	0.34699848	1.00350811

proc discrim 9

The DISCRIM Procedure
Within-Class Correlation Coefficients / Pr > |r|

survive = N

Variable	total	alar	beakhead	humerus	sternum
total	1.00000	0.77620	0.67698	0.68242	0.65687
		<.0001	<.0001	<.0001	0.0001
alar	0.77620	1.00000	0.69782	0.78455	0.62001
	<.0001		<.0001	<.0001	0.0004
beakhead	0.67698	0.69782	1.00000	0.83470	0.56989
	<.0001	<.0001		<.0001	0.0015
humerus	0.68242	0.78455	0.83470	1.00000	0.66779
	<.0001	<.0001	<.0001		0.0001
sternum	0.65687	0.62001	0.56989	0.66779	1.00000
	0.0001	0.0004	0.0015	0.0001	

---

survive = Y

Variable	total	alar	beakhead	humerus	sternum
total	1.00000	0.65447	0.64251	0.62392	0.51036
		0.0013	0.0017	0.0025	0.0181
alar	0.65447	1.00000	0.62637	0.74644	0.27744
	0.0013		0.0024	0.0001	0.2234
beakhead	0.64251	0.62637	1.00000	0.61805	0.43364
	0.0017	0.0024		0.0028	0.0495
humerus	0.62392	0.74644	0.61805	1.00000	0.41654
	0.0025	0.0001	0.0028		0.0603
sternum	0.51036	0.27744	0.43364	0.41654	1.00000
	0.0181	0.2234	0.0495	0.0603	

The DISCRIM Procedure

Pooled Within-Class Correlation Coefficients / Pr > |r|

Variable	total	alar	beakhead	humerus	sternum
total	1.00000	0.73564 <.0001	0.66486 <.0001	0.65964 <.0001	0.60933 <.0001
alar	0.73564 <.0001	1.00000	0.67348 <.0001	0.77329 <.0001	0.52907 0.0001
beakhead	0.66486 <.0001	0.67348 <.0001	1.00000	0.76571 <.0001	0.52612 0.0001
humerus	0.65964 <.0001	0.77329 <.0001	0.76571 <.0001	1.00000	0.60812 <.0001
sternum	0.60933 <.0001	0.52907 0.0001	0.52612 0.0001	0.60812 <.0001	1.00000

The DISCRIM Procedure
Simple Statistics

Total-Sample

Variable	N	Sum	Mean	Variance	Standard Deviation
total	49	7741	157.97959	13.35374	3.6543
alar	49	11825	241.32653	25.68282	5.0678
beakhead	49	1542	31.45918	0.63163	0.7948
humerus	49	905.00000	18.46939	0.31842	0.5643
sternum	49	1021	20.82653	0.98282	0.9914

---

survive = N

Variable	N	Sum	Mean	Variance	Standard Deviation
total	28	4436	158.42857	15.06878	3.8819
alar	28	6764	241.57143	32.55026	5.7053
beakhead	28	881.40000	31.47857	0.72841	0.8535
humerus	28	516.50000	18.44643	0.43443	0.6591
sternum	28	583.50000	20.83929	1.32099	1.1493

---

survive = Y

Variable	N	Sum	Mean	Variance	Standard Deviation
total	21	3305	157.38095	11.04762	3.3238
alar	21	5061	241.00000	17.50000	4.1833
beakhead	21	660.10000	31.43333	0.53133	0.7289
humerus	21	388.50000	18.50000	0.17600	0.4195
sternum	21	437.00000	20.80952	0.57490	0.7582

Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
5	0.96243

proc discrim 12

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

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

Generalized Squared Distance to survive

From survive	N	Y
N	0	0.23531
Y	0.23531	0

Linear Discriminant Function

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

Linear Discriminant Function for survive

Variable	N	Y
Constant	-1349	-1336
total	5.69222	5.53690
alar	8.02457	7.99808
beakhead	25.31740	25.22455
humerus	-38.63704	-37.60456
sternum	-10.87363	-10.80431

proc discrim 13

The DISCRIM Procedure
Classification Summary for Calibration Data: SASDATA.BUMPUS
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

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

Posterior Probability of Membership in Each survive

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \, \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into survive

From survive	N	Y	Total
N	19	9	28
	67.86	32.14	100.00
Y	8	13	21
	38.10	61.90	100.00
Total	27	22	49
	55.10	44.90	100.00
Priors	0.5	0.5	

Error Count Estimates for survive

	N	Y	Total
Rate	0.3214	0.3810	0.3512
Priors	0.5000	0.5000	

proc princomp - 1 14

The PRINCOMP Procedure

Observations	49
Variables	5

Simple Statistics

	total	alar	beakhead	humerus	sternum
Mean	157.9795918	241.3265306	31.45918367	18.46938776	20.82653061
StD	3.6542772	5.0678223	0.79475320	0.56428571	0.99137436

Covariance Matrix

	total	alar	beakhead	humerus	sternum
total	13.35374150	13.61096939	1.92206633	1.33061224	2.19221939
alar	13.61096939	25.68282313	2.71360544	2.19770408	2.65782313
beakhead	1.92206633	2.71360544	0.63163265	0.34226616	0.41464711
humerus	1.33061224	2.19770408	0.34226616	0.31841837	0.33937075
sternum	2.19221939	2.65782313	0.41464711	0.33937075	0.98282313

Total Variance 40.969438776

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	35.3257569	30.7032976	0.8622	0.8622
2	4.6224593	3.9915416	0.1128	0.9751
3	0.6309178	0.3181342	0.0154	0.9905
4	0.3127836	0.2352626	0.0076	0.9981
5	0.0775211		0.0019	1.0000

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5
total	0.536501	0.828100	-.156491	-.040210	0.017652
alar	0.829015	-.550512	-.057744	-.069022	-.039642
beakhead	0.096496	0.033562	0.237515	0.897627	-.356953
humerus	0.074352	-.014595	0.203245	0.307241	0.926581
sternum	0.100304	0.099234	0.935123	-.305760	-.110219

proc princomp - 2 15

The PRINCOMP Procedure

Observations	49
Variables	5

Simple Statistics

	total	alar	beakhead	humerus	sternum
Mean	157.9795918	241.3265306	31.45918367	18.46938776	20.82653061
StD	3.6542772	5.0678223	0.79475320	0.56428571	0.99137436

Correlation Matrix

	total	alar	beakhead	humerus	sternum
total	1.0000	0.7350	0.6618	0.6453	0.6051
alar	0.7350	1.0000	0.6737	0.7685	0.5290
beakhead	0.6618	0.6737	1.0000	0.7632	0.5263
humerus	0.6453	0.7685	0.7632	1.0000	0.6066
sternum	0.6051	0.5290	0.5263	0.6066	1.0000

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	3.61597834	3.08447427	0.7232	0.7232
2	0.53150408	0.14507953	0.1063	0.8295
3	0.38642455	0.08485903	0.0773	0.9068
4	0.30156552	0.13703801	0.0603	0.9671
5	0.16452751		0.0329	1.0000

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5
total	0.451799	0.050721	-.690470	0.420414	-.373909
alar	0.461681	-.299564	-.340548	-.547863	0.530080
beakhead	0.450542	-.324572	0.454493	0.606296	0.342792
humerus	0.470739	-.184684	0.410935	-.388278	-.651667
sternum	0.397675	0.876489	0.178456	-.068872	0.192434