Technical University of Denmark

Written examination, date: 10. December 2013 *Page 1 of 16 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	2	2	2	2	2
Question	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5
Answer										

Problem	3	3	3	3	3	4	4	4	4	4
Question	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5
Answer										

Problem	5	5	5	5	6	6	6	6	6	6
Question	5.1	5.2	5.3	5.4	6.1	6.2	6.3	6.4	6.5	6.6
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.

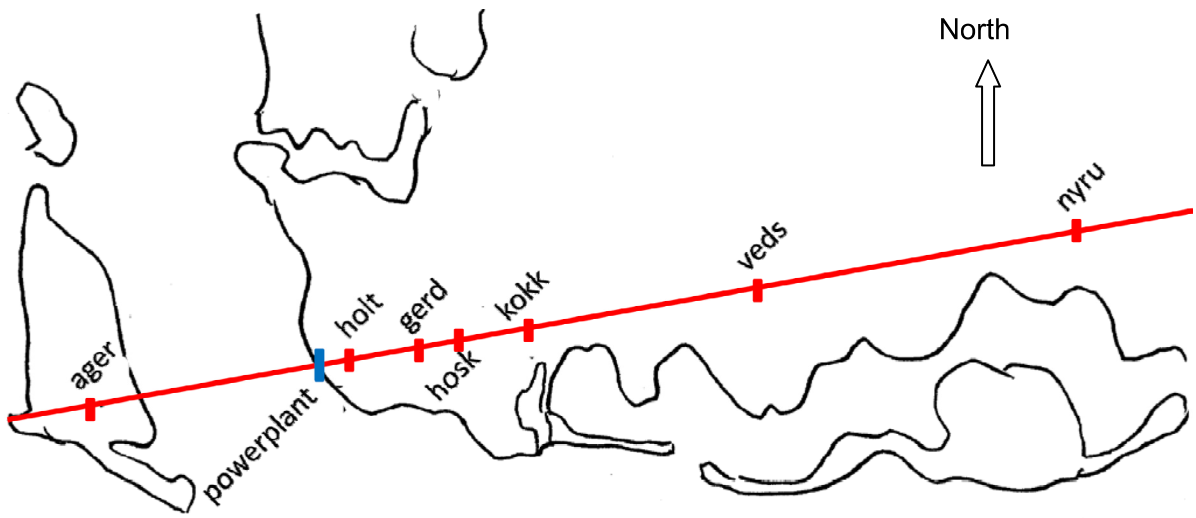
A correct answer gives 5 points, a wrong answer gives – 1 point. Unanswered questions or a 6 (corresponding to “don’t know”) give 0 points. The total number of points needed for a satisfactorily answered exam is determined at the final evaluation of the exam. Especially note that the grade 10 may be 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 measurements are estimated correlations between 109 observed values of sulphur dioxide (SO_2) at 7 measurement stations in Southern Zealand with code names: ager, holt, gerd, hosk, kokk, veds, and nyru . The measurement stations are placed on a line parallel to the dominating wind direction and going through a power plant emitting large amounts of sulphur dioxide. The distance between the two stations furthest apart (ager and nyru) is 20 kms. The situation is depicted below.



We are interested in analyzing the correlation structure between the measurements at the different sites.

The problem continues on the next page

Question 1.1.

We call the estimated eigenvalues of the correlation matrix $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_7$ .The arithmetic and the geometric average of the smallest 6 eigenvalues are respectively

$$\frac{1}{6}\sum_{i=2}^{7} \hat{\lambda}_i = 0.22461$$

$$\left\{\prod_{i=2}^{7} \hat{\lambda}_i\right\}^{\frac{1}{6}} = 0.10749$$

The usual test statistic for testing the hypothesis that the smallest 6 eigenvalues are equal against all alternatives is

1 ☐ $\frac{(0.22461-0.10749)^2}{0.10749}$

2 ☐ $\left[\frac{5.65231}{0.22461}\right]^2$

3 ☐ $-109 \times 6 \times \{\ln(0.10749) - \ln(0.22461)\}$

4 ☐ $-109 \times \{\ln(0.10749) - \ln(0.22461)\}$

5 ☐ $\left[\frac{0.22461}{0.10749}\right]^{109}$

6 ☐ Don’t know

Question 1.2.

The degrees of freedom in the test statistic considered above is:

1 ☐ 10

2 ☐ 20

3 ☐ 30

4 ☐ 40

5 ☐ 50

6 ☐ Don’t know.

The problem continues on the next page

Question 1.3.

We now consider a factor analysis with two factors of the data considered above. Consider the following statements on interpretation of an arbitrary factor

A. The factor basically represents the mean level of all stations.
B. The factor basically represents the mean level of all stations east of the power plant.
C. The factor basically represents a contrast between the measurements west and east of the power plant.
D. The factor basically represents the mean level of all stations except the one closest to the power plant.
E. The factor basically represents the mean level at the station closest to the power plant.

For (VARIMAX rotated factor 1, VARIMAX rotated factors 2) the following characterization is adequate:

1 ☐ (E, B)

2 ☐ (D, C)

3 ☐ (A, C)

4 ☐ (D, C)

5 ☐ (D, E)

6 ☐ Don’t know.

Question 1.4.

What fraction of the total variance will be explained by the two VARIMAX rotated factors?

1 ☐ 0.9044

2 ☐ 0.9053

3 ☐ 0.9044^2

4 ☐ $(4.9643^2 + 1.3730^2)/7$

5 ☐ $0.8075^2 + 0.0979^2$

6 ☐ Don’t know.

The problem continues on the next page

Question 1.5.

What fraction of the variation at the station closest to the power plant (holt) is explained by the first VARIMAX rotated factor?

1 ☐ $1 - 0.96125$

2 ☐ 0.9970

3 ☐ 0.9970^2

4 ☐ 0.27020^2

5 ☐ 0.6851^2

6 ☐ Don't know.

Problem 2

We consider a random variable

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

with mean value and dispersion matrix respectively equal to

$$\mu = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}.$$

Furthermore we consider the random variables

$$U = X + Y + Z$$

$$V = 2X - Y - Z$$

The problem continues on the next page

Question 2.1.

The mean value of the two-dimensional random variable $\begin{bmatrix} U \\ V \end{bmatrix}$ is

1 ☐ $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

2 ☐ $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$

3 ☐ $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$

4 ☐ $\begin{bmatrix} 0 \\ 3 \end{bmatrix}$

5 ☐ $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$

6 ☐ Don't know

Question 2.2.

The variance of U is:

1 ☐ 3

2 ☐ $3 + 3\rho$

3 ☐ 6

4 ☐ $6\rho^2$

5 ☐ $3 + 6\rho$

6 ☐ Don't know.

Question 2.3.

The variance of V is:

1 ☐ $3 + 6\rho$

2 ☐ $2\rho^2$

3 ☐ 6ρ

4 ☐ $6 - 6\rho$

5 ☐ $3 + 3\rho$

6 ☐ Don't know.

The problem continues on the next page

Question 2.4.

The covariance between U and V is:

1 ☐ 1
2 ☐ 0
3 ☐ -1
4 ☐ ρ
5 ☐ $-\rho$
6 ☐ Don't know.

Question 2.5.

The conditional mean $E\left(\begin{bmatrix} X \\ Y \end{bmatrix} \middle| Z = z\right)$ is equal to:

1 ☐ $1 + \rho(z - 1)$

2 ☐ $\begin{bmatrix} 1 + \rho(z - 1) \\ 1 + \rho(z - 1) \end{bmatrix}$

3 ☐ $\begin{bmatrix} \rho(z - 1) \\ \rho(z - 1) \end{bmatrix}$

4 ☐ $\begin{bmatrix} 0 \\ 3 \end{bmatrix}$

5 ☐ $3 + (1 + 2\rho)z$

6 ☐ Don't know.

Problem 3

Enclosure B belongs to this problem. The data are color measurements of different meat samples including raw meat from cattle, pork, and turkey, and processed meat as sausages, coocked ham, forcemeat etc. Color parameters have been assessed by means of a Minolta CR-300 colorimeter (M1) and a VideometerLab multispectral imaging device (M2). The color components measured are based on the definition of the CIELAB space – the lightness L* varying from black (zero) to white (100), a* varying from green (negative values) to red (positive values), and b* varying from yellow (positive values) to blue (negative values). In the sequel we omit the asterisk (*). The L, a, b measurements obtained using the two methods are called LMi, aMi, and bMi, i=1, 2.We are now interested in comparing the results obtained by using the two different instruments. Understanding the nature of these measurements is not crucial for solving the present problem.

The problem continues on the next page

Question 3.1.

At first we consider how well bM1 is predicted from bM2.
Below (,) corresponds to an open ended interval, [,] to a closed interval The fraction of the variation in bM1 that is explained by bM2 lies in the interval:

1 ☐ [0, 0.2]

2 ☐ (0.2, 0.4]

3 ☐ (0.4, 0.6]

4 ☐ (0.6, 0.8]

5 ☐ (0.8, 1]

6 ☐ Don't know.

Question 3.2.

The test statistic for testing whether the correlation between bM1 and bM2 is equal to 0 is:

1 ☐ $\frac{0.4876}{\sqrt{1-0.4876^2}}$

2 ☐ $\frac{0.4876}{0.9953+0.8824}$

3 ☐ $\frac{0.4876^2}{1-0.4876^2}$

4 ☐ $\frac{0.4876^2}{0.9953^2+0.8824^2}$

5 ☐ $\frac{0.4876}{\sqrt{1-0.4876^2}}\sqrt{58}$

6 ☐ Don't know.

Question 3.3.

Under the null hypothesis the test statistic related to the test described in Question 3.2 is distributed as:

1 ☐ t(58)

2 ☐ t(59)

3 ☐ F(1,59)

4 ☐ F(2,58)

5 ☐ $\chi^2(59)$

6 ☐ Don't know.

The problem continues on the next page

Question 3.4.

The canonical variables for the L, a, b values obtained by the two methods M1 and M2 have very similar structures. Which of the following statements describes the structure of canonical variable number 1 best:

1 ☐ The variable is basically equal to a.

2 ☐ The variable is basically a contrast between the a and the b values.

3 ☐ The variable is basically a contrast between the L and the a value.

4 ☐ The variable is basically a contrast between the L and the b values.

5 ☐ The variable is basically the overall average of the L, a, b values

6 ☐ Don’t know.

Question 3.5.

The largest correlation between the b-value measured by method M2 (bM2) and the canonical variables based on method M1 lies in the interval:

1 ☐ [0, 0.2]

2 ☐ (0.2, 0.4]

3 ☐ (0.4, 0.6]

4 ☐ (0.6, 0.8]

5 ☐ (0.8, 1]

6 ☐ Don’t know.

Problem 4

Enclosure C belongs to this problem. The data are measurements on the relation between the level of prostate-specific antigen and a number of clinical measures. The 97 males investigated were operated for prostate cancer. The data were originally published in Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N (1989): Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate II. Radical prostatectomy treated patients, Journal of Urology 16: 1076–1083.

The problem continues on the next page

The variables are:

- Lcavol = logarithm of cancer volume
- Lweight = logarithm of prostate weight
- age in years
- lbph = logarithm of the amount of benign prostatic hyperplasia
- svi = seminal vesicle invasion
- Lcp = logarithm of capsular penetration
- gleason = a numeric vector
- pgg45 = percent of Gleason score 4 or 5
- lpsa = response, the logarithm of the level of prostate-specific antigen (PSA)

As mentioned, we are now interested in predicting the logarithm of the level of prostate-specific antigen (PSA) that is elevated in men with prostate cancer. We first run a regression analysis with the 8 clinical variables as independent variables

Question 4.1.

The 95% confidence interval for the expected value of the first observation is:

1 ☐ $[0.8229 - t(96)_{0.975} \times 0.1991,\ 0.8229 + t(96)_{0.975} \times 0.1991]$

2 ☐ $\left[0.8229 - t(88)_{0.975} \times \sqrt{0.48930},\ 0.8229 + t(88)_{0.975} \times \sqrt{0.48930}\right]$

3 ☐ $[-0.4308 - t(88)_{0.975} \times 0.1991, -0.4308 + t(88)_{0.975} \times 0.1991]$

4 ☐ $[0.8229 - t(88)_{0.975} \times 0.1991,\ 0.8229 + t(88)_{0.975} \times 0.1991]$

5 ☐ $[-0.4308 - 1.96 \times 0.1991, -0.4308 + 1.96 \times 0.1991]$

6 ☐ Don't know.

Question 4.2.

The 95% prediction interval for the first observation is:

1 ☐ $\left[0.8229 - t(88)_{0.975} \times \sqrt{0.48930},\ 0.8229 + t(88)_{0.975} \times \sqrt{0.48930}\ \right]$

2 ☐ $\left[0.8229 - t(88)_{0.975} \times \sqrt{0.1991^2 + 0.48930},\ 0.8229 + t(88)_{0.975} \times \sqrt{0.1991^2 + 0.48930}\ \right]$

3 ☐ $[0.8229 - t(88)_{0.975} \times (1 + 0.1991),\ 0.8229 + t(88)_{0.975} \times (1 + 0.1991)]$

4 ☐ $\left[-0.4308 - t(88)_{0.975} \times \sqrt{0.1991^2 + 0.48930}, -0.4308 + t(88)_{0.975} \times \sqrt{0.1991^2 + 0.48930}\ \right]$

5 ☐ $\left[0.8229 - t(88)_{0.975} \times \sqrt{1 + 0.1991},\ 0.8229 + t(88)_{0.975} \times \sqrt{1 + 0.1991}\right]$

6 ☐ Don't know.

The problem continues on the next page

Question 4.3.

We now remove the observations one at a time. The observation that - when removed- will cause the most substantial change in its predicted value is number:

1 ☐ 47

2 ☐ 38

3 ☐ 39

4 ☐ 69

5 ☐ 95

6 ☐ Don't know.

Question 4.4.

The number of observations that have as well an extreme RStudent residual as an extreme (i.e. large) leverage is:

1 ☐ 0

2 ☐ 2

3 ☐ 4

4 ☐ 6

5 ☐ 8

6 ☐ Don't know.

Question 4.5.

We now consider the stepwise selection of variables. In steps one and two the variables lcavol and lweight were included. The next variable to be included is:

1 ☐ age

2 ☐ lbph

3 ☐ svi

4 ☐ gleason

5 ☐ pgg45

6 ☐ Don't know.

The problem continues on the next page

Problem 5

We consider the model

$$\begin{bmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \\ Y_3 & Z_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 & \varphi_1 \\ \varepsilon_2 & \varphi_2 \\ \varepsilon_3 & \varphi_3 \end{bmatrix}$$

where the error terms $\begin{bmatrix} \varepsilon_i \\ \varphi_i \end{bmatrix}$, $i = 1,2,3,$ are independent and normally distributed $N_2(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is the unknown dispersion matrix. We assume that we obtained the following observations

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

Question 5.1.

The maximum likehood estimator for $\begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$ becomes

1 ☐ $\begin{bmatrix} 2 & 0.5 \\ 2 & -0.5 \end{bmatrix}$

2 ☐ $\begin{bmatrix} 2 & 2 \\ 0.5 & -0.5 \end{bmatrix}$

3 ☐ $\begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix}$

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

5 ☐ $\begin{bmatrix} -1 & -1 \\ 0.3 & 0.3 \end{bmatrix}$

6 ☐ Don't know

The problem continues on the next page

Question 5.2.

We now want to test the hypothesis

$$H_0 : (\alpha, \gamma) = (2,2)$$

against all alternatives. This hypothesis may also be written

$$H_0\text{: } \boldsymbol{A}\begin{bmatrix}\alpha & \gamma \\ \beta & \delta\end{bmatrix}\boldsymbol{B}' = \boldsymbol{C} \text{ against } H_1\text{: } \boldsymbol{A}\begin{bmatrix}\alpha & \gamma \\ \beta & \delta\end{bmatrix}\boldsymbol{B}' \neq \boldsymbol{C}$$

Here the matrix ***A*** is:

1 ☐ $\begin{bmatrix}1 & 0 \\ 0 & 0\end{bmatrix}$

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

3 ☐ $\begin{bmatrix}0 & 1\end{bmatrix}$

4 ☐ $\begin{bmatrix}1 & 0\end{bmatrix}$

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

6 ☐ Don’t know.

Question 5.3.

The matrix ***B*** is:

1 ☐ $\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}$

2 ☐ $\begin{bmatrix}1 \\ 0\end{bmatrix}$

3 ☐ $\begin{bmatrix}0 & 1\end{bmatrix}$

4 ☐ $\begin{bmatrix}1 \\ 1\end{bmatrix}$

5 ☐ $\begin{bmatrix}1 & 1 \\ 1 & 1\end{bmatrix}$

6 ☐ Don’t know.

The problem continues on the next page

Question 5.4.

If the hypothesis H_0 is true then the distribution of the usual test statistic is:

1 ☐ U(2, 2, 2)

2 ☐ U(1, 1, 2)

3 ☐ U(2, 1, 1)

4 ☐ U(2, 2, 1)

5 ☐ U(1, 1, 1)

6 ☐ Don't know.

Problem 6

Enclosure D belongs to this problem. The data are the same as described in Problem 3 and Enclosure B. However, here we only consider measurements taken with the VideometerLab multispectral imaging device (M2). Our primary interest is to see whether we may distinguish between raw meat (pr=0) and processed meat (pr=1) by means of the color variables LM2, aM2, and bM2.

Question 6.1.

Hotellings T^2statistic for testing whether the mean values of the color measurements for raw and for processed meat are the same is:

1 ☐ 62.79699

2 ☐ $\frac{35\times 25}{35+25}13.37960$

3 ☐ -142.78879+135.31413

4 ☐ $\frac{35}{60}13.37960$

5 ☐ $\frac{35}{60}62.79699$

6 ☐ Don't know.

The problem continues on the next page

Question 6.2.

The constant in the usual discriminant function for distinguishing between the two populations (cf the boxed formula on p 261 in the lecture notes) is:

1 ☐ -10.0

2 ☐ -7.5

3 ☐ -5.0

4 ☐ -2.5

5 ☐ 0.0

6 ☐ Don't know.

Question 6.3.

We now assume that the prior probability for raw meat (pr=0) is twice as large as the probability for processed meat (pr=1). If we still want to use a Bayes classifier then the constant in Question 6.2 should be decreased by

1 ☐ 0.6931

2 ☐ 0.3010

3 ☐ 1.0986

4 ☐ 1.7918

5 ☐ 1.3956

6 ☐ Don't know.

Question 6.4.

The number of misclassified samples if we use cross validation is:

1 ☐ 1

2 ☐ 2

3 ☐ 3

4 ☐ 4

5 ☐ 0

6 ☐ Don't know.

The problem continues on the next page

Question 6.5.

The test statistic for testing whether LM2 provides additional information in discriminating between raw and processed meat compared to just using aM2 and bM2 is:

1 ☐ $\frac{91.46590}{62.79699}$

2 ☐ $\frac{91.46590/57}{62.79699/56}$

3 ☐ $\frac{56\times35\times25\times(13.37960-12.76396)}{60\times58+35\times25\times12.76396}$

4 ☐ $\frac{56\times35\times25\times(91.46590-62.79699)}{60\times58+35\times25\times62.79699}$

5 ☐ $\frac{13.37960/57}{12.76396/56}$

6 ☐ Don’t know.

Question 6.6.

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

1 ☐ U(2,2,58)

2 ☐ t(58)

3 ☐ t(56)

4 ☐ F(2,58)

5 ☐ F(1,56)

6 ☐ Don’t know.

Enclosure A – SAS program

```
data stigsnaes (type=corr);
infile cards missover;
input _type_ $ _name_ $ ager holt gerd hosk kokk veds nyru;
cards;
N     .     109   109   109   109   109   109   109
CORR ager   1
CORR holt   0.541  1
CORR gerd   0.760  0.429  1
CORR hosk   0.935  0.560  0.788  1
CORR kokk   0.939  0.545  0.774  0.981  1
CORR veds   0.914  0.480  0.763  0.949  0.968  1
CORR nyru   0.812  0.453  0.689  0.852  0.899  0.911  1
;
ods graphics on;
proc princomp data=stigsnaes;
run;
proc factor data=stigsnaes
  score
  rotate=varimax
  plots=all
  nfactors=2;
run;
ods graphics off;
```


Enclosure A – SAS output

The PRINCOMP Procedure

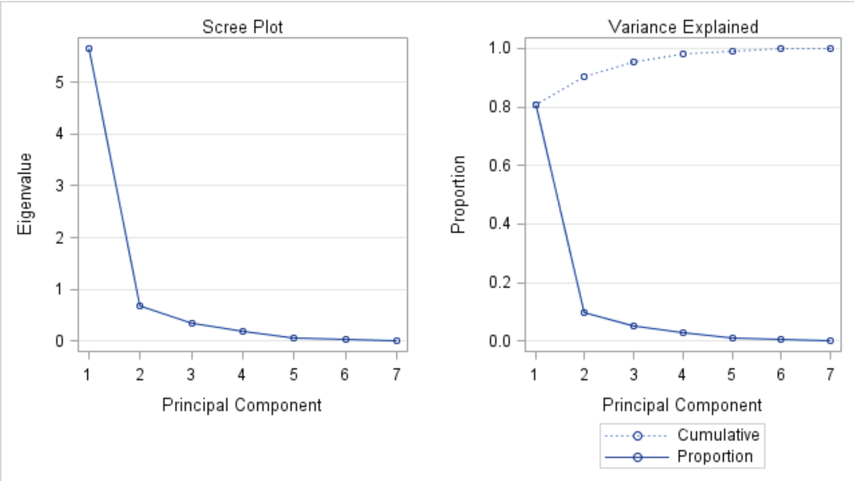
Observations	109
Variables	7

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	5.65231177	4.96724774	0.8075	0.8075
2	0.68506403	0.33305053	0.0979	0.9053
3	0.35201350	0.16261735	0.0503	0.9556
4	0.18939615	0.11855387	0.0271	0.9827
5	0.07084227	0.03310281	0.0101	0.9928
6	0.03773947	0.02510667	0.0054	0.9982
7	0.01263280		0.0018	1.0000

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
ager	0.399796	-.037747	-.041095	-.550362	0.728201	-.000003	0.062242
holt	0.255957	0.956464	0.006752	0.121149	0.001097	0.070265	-.001158
gerd	0.350684	-.131117	0.895955	0.234814	0.031046	0.015457	-.027429
hosk	0.410819	-.040349	-.049428	-.288180	-.496142	-.428667	0.560532
kokk	0.414107	-.076497	-.150489	-.130252	-.269114	-.237280	-.808905
veds	0.407158	-.169269	-.188282	0.004673	-.247584	0.836443	0.095732
nyru	0.381622	-.174436	-.367410	0.726121	0.298129	-.234798	0.132983



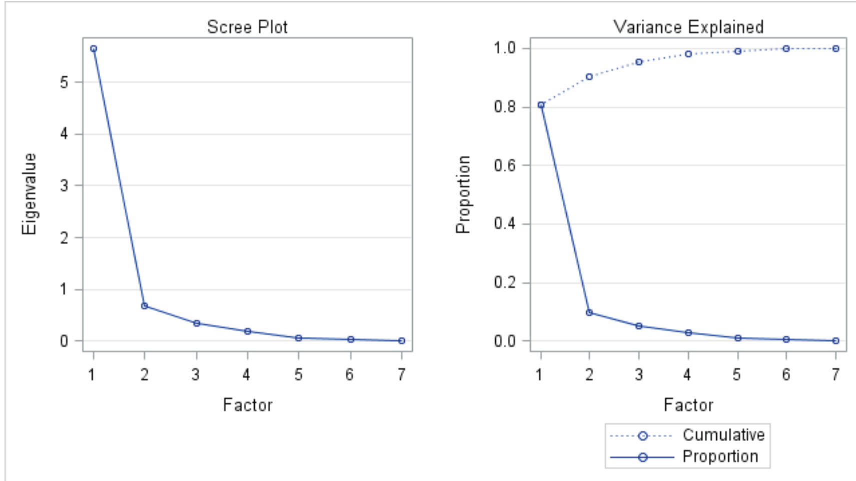
The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total= 7 Average = 1				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	5.65231177	4.96724774	0.8075	0.8075
2	0.68506403	0.33305053	0.0979	0.9053
3	0.35201350	0.16261735	0.0503	0.9556
4	0.18939615	0.11855387	0.0271	0.9827
5	0.07084227	0.03310281	0.0101	0.9928
6	0.03773947	0.02510667	0.0054	0.9982
7	0.01263280		0.0018	1.0000

Enclosure A – SAS output



Factor Pattern		
	Factor1	**Factor2**
ager	0.95050	-0.03124
holt	0.60853	0.79165
gerd	0.83374	-0.10852
hosk	0.97671	-0.03340
kokk	0.98452	-0.06332
veds	0.96800	-0.14010
nyru	0.90729	-0.14438

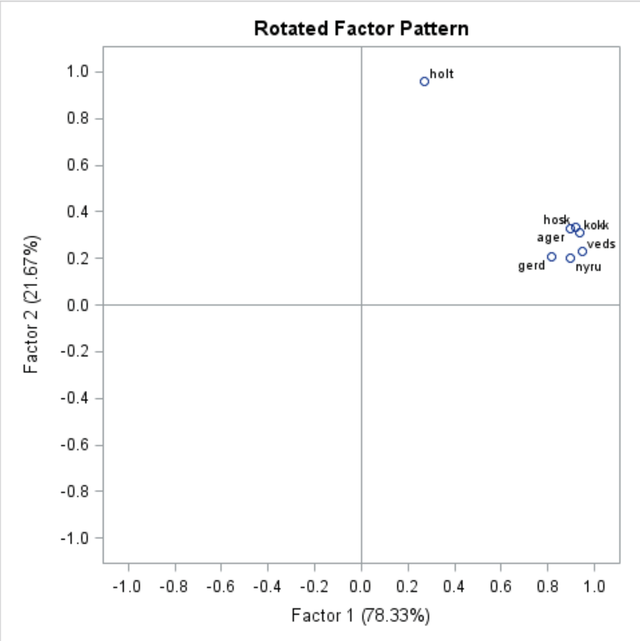
Variance Explained by Each Factor	
Factor1	**Factor2**
5.6523118	0.6850640

Final Community Estimates: Total = 6.337376						
ager	**holt**	**gerd**	**hosk**	**kokk**	**veds**	**nyru**
0.90442272	0.99701777	0.70689647	0.95506869	0.97329344	0.95665485	0.84402187

Enclosure A – SAS output

The FACTOR Procedure

Rotation Method: Varimax



Enclosure B – SAS program

```
Title 'First 10 Observations in LabM1M2';

proc print data=sasuser.LabM1M2 (obs=10);
run;

Title 'Canonical Correlation Analysis on LabM1M2';
proc cancorr data=sasuser.LabM1M2 all ;
var LM1 aM1 bM1;
with LM2 aM2 bM2;
run;
```

First 10 Observations in LabM1M2

Obs	LM1	aM1	bM1	LM2	aM2	bM2
1	54.3450	5.9700	6.3700	59.4775	14.9929	15.4351
2	56.6825	6.6200	6.6700	60.1689	14.9891	14.9946
3	57.5375	6.9525	7.0225	61.3140	14.4471	14.3082
4	58.0600	6.5875	6.9850	61.2555	13.7060	14.6464
5	58.4475	6.6950	7.1275	61.2011	14.5520	14.5022
6	44.5500	12.7400	7.2650	45.2058	16.7109	13.2196
7	44.8900	12.7975	7.8075	45.5458	19.4897	15.4711
8	44.5625	10.7575	6.7650	44.8497	18.3109	13.9600
9	47.8600	8.3300	6.3350	49.5087	17.7987	15.3426
10	50.7450	8.6075	7.1875	53.0712	15.7536	15.2295

Canonical Correlation Analysis on LabM1M2

The CANCORR Procedure

VAR Variables	3
WITH Variables	3
Observations	60

Means and Standard Deviations		
Variable	Mean	Standard Deviation
LM1	54.145375	14.017547
aM1	11.938750	7.210135
bM1	9.295125	3.034791
LM2	55.323785	15.391321
aM2	15.085795	8.185637
bM2	14.220339	2.628269

Canonical Correlation Analysis on LabM1M2

The CANCORR Procedure

Correlations Among the Original Variables

Correlations Among the VAR Variables			
	LM1	aM1	bM1
LM1	1.0000	-0.7975	0.2039
aM1	-0.7975	1.0000	0.2035
bM1	0.2039	0.2035	1.0000

Correlations Among the WITH Variables			
	LM2	aM2	bM2
LM2	1.0000	-0.9208	-0.5042
aM2	-0.9208	1.0000	0.4376
bM2	-0.5042	0.4376	1.0000

Correlations Between the VAR Variables and the WITH Variables			
	LM2	aM2	bM2
LM1	0.9953	-0.9249	-0.5061
aM1	-0.8226	0.8824	0.4813
bM1	0.1555	-0.2166	0.4876

Canonical Correlation Analysis on LabM1M2

The CANCORR Procedure

Canonical Correlation Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)			
					Eigenvalue	Difference	Proportion	Cumulative
1	0.998139	0.998035	0.000484	0.996280	267.8523	236.1000	0.8932	0.8932
2	0.984616	0.984411	0.003975	0.969468	31.7524	31.4776	0.1059	0.9991
3	0.464315	.	0.102122	0.215588	0.2748		0.0009	1.0000

	Test of H0: The canonical correlations in the current row and all that follow are zero				
	Likelihood Ratio	ApproximateF Value	Num DF	Den DF	Pr > F
1	0.00008908	660.09	9	131.57	<.0001
2	0.02394975	150.20	4	110	<.0001
3	0.78441179	15.39	1	56	0.0002

Multivariate Statistics and F Approximations					
S=3 M=-0.5 N=26					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00008908	660.09	9	131.57	<.0001
Pillai's Trace	2.18133657	49.74	9	168	<.0001
Hotelling-Lawley Trace	299.87958514	1776.10	9	81.733	<.0001
Roy's Greatest Root	267.85234967	4999.91	3	56	<.0001
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					

Canonical Correlation Analysis on LabM1M2

The CANCORR Procedure

Canonical Correlation Analysis

Raw Canonical Coefficients for the VAR Variables			
	V1	V2	V3
LM1	0.0529120838	-0.122766458	0.0694656593
aM1	-0.039505401	-0.236177729	0.1686194591
bM1	0.0355850852	0.4098960971	0.1209124579

Raw Canonical Coefficients for the WITH Variables			
	W1	W2	W3
LM2	0.0465305326	-0.091000828	0.1407859243
aM2	-0.039221557	-0.220045512	0.2207753858
bM2	0.0155767032	0.2854238639	0.3371007214

Standardized Canonical Coefficients for the VAR Variables			
	V1	V2	V3
LM1	0.7417	-1.7209	0.9737
aM1	-0.2848	-1.7029	1.2158
bM1	0.1080	1.2439	0.3669

Standardized Canonical Coefficients for the WITH Variables			
	W1	W2	W3
LM2	0.7162	-1.4006	2.1669
aM2	-0.3211	-1.8012	1.8072
bM2	0.0409	0.7502	0.8860

Canonical Correlation Analysis on LabM1M2

The CANCORR Procedure

Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables			
	V1	V2	V3
LM1	0.9909	-0.1093	0.0790
aM1	-0.8543	-0.0774	0.5139
bM1	0.2012	0.5467	0.8128

Correlations Between the WITH Variables and Their Canonical Variables			
	W1	W2	W3
LM2	0.9911	-0.1204	0.0561
aM2	-0.9626	-0.1833	0.1997
bM2	-0.4607	0.6682	0.5842

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables			
	W1	W2	W3
LM1	0.9890	-0.1076	0.0367
aM1	-0.8527	-0.0762	0.2386
bM1	0.2009	0.5382	0.3774

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables			
	V1	V2	V3
LM2	0.9893	-0.1185	0.0261
aM2	-0.9608	-0.1805	0.0927
bM2	-0.4598	0.6579	0.2712


```
Title 'Prostate Data. First 20 observations';
proc print data=sasuser.prostate (obs=20);
run;

ods graphics on;

Title 'Regression with Influence and Residuals Diagnostics';
proc reg data=sasuser.prostate;
model lpsa = lcavol lweight age lbph svi lcp gleason pgg45 /
  influence r;
run;

Title 'Stepwise Regression with at most 2 steps';
proc reg data=sasuser.prostate;
model lpsa = lcavol lweight age lbph svi lcp gleason pgg45 /
  selection=stepwise maxstep=2;
run;

Title 'Partial Correlation Analysis';
proc corr data=sasuser.prostate;
partial lcavol lweight;
run;

ods graphics off;
```

Prostate Data. First 20 observations

Obs	patno	lcavol	lweight	age	lbph	svi	lcp	gleason	pgg45	lpsa
1	1	-0.57982	2.76946	50	-1.38629	0	-1.38629	6	0	-0.43078
2	2	-0.99425	3.31963	58	-1.38629	0	-1.38629	6	0	-0.16252
3	3	-0.51083	2.69124	74	-1.38629	0	-1.38629	7	20	-0.16252
4	4	-1.20397	3.28279	58	-1.38629	0	-1.38629	6	0	-0.16252
5	5	0.75142	3.43237	62	-1.38629	0	-1.38629	6	0	0.37156
6	6	-1.04982	3.22883	50	-1.38629	0	-1.38629	6	0	0.76547
7	7	0.73716	3.47352	64	0.61519	0	-1.38629	6	0	0.76547
8	8	0.69315	3.53951	58	1.53687	0	-1.38629	6	0	0.85442
9	9	-0.77653	3.53951	47	-1.38629	0	-1.38629	6	0	1.04732
10	10	0.22314	3.24454	63	-1.38629	0	-1.38629	6	0	1.04732
11	11	0.25464	3.60414	65	-1.38629	0	-1.38629	6	0	1.26695
12	12	-1.34707	3.59868	63	1.26695	0	-1.38629	6	0	1.26695
13	13	1.61343	3.02286	63	-1.38629	0	-0.59784	7	30	1.26695
14	14	1.47705	2.99823	67	-1.38629	0	-1.38629	7	5	1.34807
15	15	1.20597	3.44202	57	-1.38629	0	-0.43078	7	5	1.39872
16	16	1.54116	3.06105	66	-1.38629	0	-1.38629	6	0	1.44692
17	17	-0.41552	3.51601	70	1.24415	0	-0.59784	7	30	1.47018
18	18	2.28849	3.64936	66	-1.38629	0	0.37156	6	0	1.49290
19	19	-0.56212	3.26767	41	-1.38629	0	-1.38629	6	0	1.55814
20	20	0.18232	3.82538	70	1.65823	0	-1.38629	6	0	1.59939

Regression with Influence and Residuals Diagnostics

The REG Procedure

Model: MODEL1

Dependent Variable: lpsa

Number of Observations Read	97
Number of Observations Used	97

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	84.85924	10.60741	21.68	<.0001
Error	88	43.05842	0.48930		
Corrected Total	96	127.91766			

Root MSE	0.69950	R-Square	0.6634
Dependent Mean	2.47839	Adj R-Sq	0.6328
Coeff Var	28.22400		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.18156	1.32057	0.14	0.8910
lcavol	1	0.56434	0.08783	6.43	<.0001
lweight	1	0.62202	0.20090	3.10	0.0026
age	1	-0.02125	0.01108	-1.92	0.0585
lbph	1	0.09671	0.05791	1.67	0.0985
svi	1	0.76167	0.24118	3.16	0.0022
lcp	1	-0.10605	0.08987	-1.18	0.2412
gleason	1	0.04923	0.15534	0.32	0.7521
pgg45	1	0.00446	0.00437	1.02	0.3100

Regression with Influence and Residuals Diagnostics

The REG Procedure

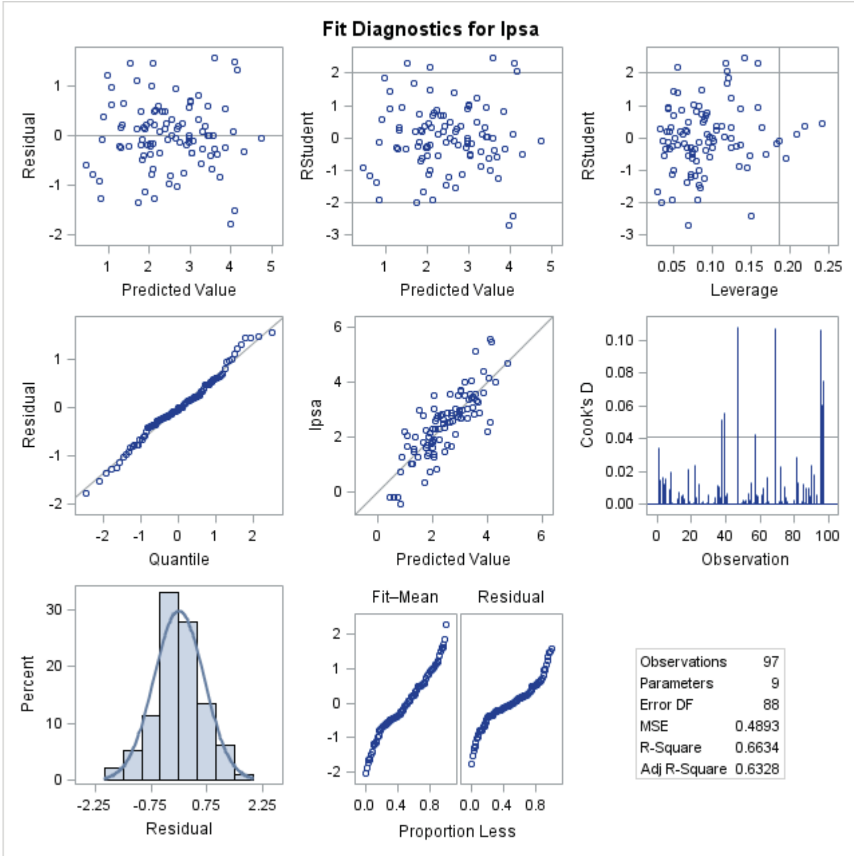
Model: MODEL1
Dependent Variable: lpsa

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Student Residual	Cook's D	RStudent	Hat Diag H	DFFITS
1	-0.4308	0.8229	0.1991	-1.2537	-1.870	0.034	-1.8970	0.0810	-0.5633
2	-0.1625	0.7613	0.1826	-0.9238	-1.368	0.015	-1.3749	0.0681	-0.3718
3	-0.1625	0.4416	0.2667	-0.6041	-0.934	0.016	-0.9335	0.1454	-0.3850
4	-0.1625	0.6200	0.1954	-0.7825	-1.165	0.013	-1.1675	0.0780	-0.3397
5	0.3716	1.7315	0.1308	-1.3600	-1.979	0.016	-2.0132	0.0350	-0.3832
6	0.7655	0.8434	0.2026	-0.0779	-0.116	0.000	-0.1157	0.0839	-0.0350
7	0.7655	1.9002	0.1196	-1.1347	-1.646	0.009	-1.6628	0.0292	-0.2885
8	0.8544	2.1330	0.1536	-1.2786	-1.874	0.020	-1.9012	0.0482	-0.4280
9	1.0473	1.2546	0.2212	-0.2073	-0.312	0.001	-0.3108	0.1000	-0.1036
10	1.0473	1.2953	0.1402	-0.2480	-0.362	0.001	-0.3601	0.0402	-0.0737
11	1.2669	1.4943	0.1489	-0.2273	-0.333	0.001	-0.3309	0.0453	-0.0721
12	1.2669	0.8861	0.2088	0.3809	0.570	0.004	0.5683	0.0891	0.1777
13	1.2669	2.0414	0.1555	-0.7744	-1.136	0.007	-1.1374	0.0494	-0.2593
14	1.3481	1.8363	0.1969	-0.4882	-0.727	0.005	-0.7254	0.0792	-0.2128
15	1.3987	2.0705	0.1569	-0.6718	-0.985	0.006	-0.9853	0.0503	-0.2268
16	1.4469	1.8613	0.1906	-0.4143	-0.616	0.003	-0.6135	0.0743	-0.1737
17	1.4702	1.3088	0.1795	0.1614	0.239	0.000	0.2375	0.0659	0.0631
18	1.4929	2.4625	0.2018	-0.9696	-1.448	0.021	-1.4569	0.0832	-0.4389
19	1.5581	1.3340	0.2459	0.2241	0.342	0.002	0.3405	0.1236	0.1279
20	1.5994	1.7793	0.1540	-0.1799	-0.264	0.000	-0.2623	0.0485	-0.0592
21	1.6390	2.0107	0.1381	-0.3717	-0.542	0.001	-0.5398	0.0390	-0.1087
22	1.6582	2.6799	0.2039	-1.0217	-1.527	0.024	-1.5387	0.0850	-0.4689
23	1.6956	1.0855	0.1473	0.6101	0.892	0.004	0.8911	0.0443	0.1919
24	1.7138	2.5248	0.1890	-0.8110	-1.204	0.013	-1.2073	0.0730	-0.3388
25	1.7317	1.8111	0.1473	-0.0795	-0.116	0.000	-0.1155	0.0443	-0.0249
26	1.7664	1.9682	0.1881	-0.2017	-0.299	0.001	-0.2979	0.0723	-0.0832
27	1.8001	2.0108	0.2458	-0.2107	-0.322	0.002	-0.3201	0.1235	-0.1202
28	1.8165	1.6931	0.1816	0.1234	0.183	0.000	0.1816	0.0674	0.0488
29	1.8485	2.0089	0.2542	-0.1604	-0.246	0.001	-0.2448	0.1320	-0.0955
30	1.8946	2.2497	0.2691	-0.3551	-0.550	0.006	-0.5478	0.1480	-0.2283
31	1.9242	2.0754	0.1643	-0.1512	-0.222	0.000	-0.2211	0.0552	-0.0534
32	2.0082	1.8770	0.1419	0.1312	0.192	0.000	0.1905	0.0411	0.0395
33	2.0082	1.8469	0.2276	0.1613	0.244	0.001	0.2426	0.1059	0.0835
34	2.0215	1.3806	0.1446	0.6409	0.936	0.004	0.9358	0.0427	0.1976
35	2.0477	1.0842	0.1579	0.9635	1.414	0.012	1.4221	0.0510	0.3296
36	2.0857	2.8465	0.1876	-0.7609	-1.129	0.011	-1.1309	0.0719	-0.3147
37	2.1576	1.9371	0.3270	0.2204	0.356	0.004	0.3547	0.2185	0.1876
38	2.1917	0.9815	0.2432	1.2102	1.845	0.052	1.8713	0.1209	0.6940
39	2.2138	3.9802	0.1831	-1.7664	-2.617	0.056	-2.7092	0.0685	-0.7349
40	2.2773	1.7906	0.2049	0.4867	0.728	0.006	0.7257	0.0858	0.2223

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Student Residual	Cook's D	RStudent	Hat Diag H	DFFITS
41	2.2976	2.0239	0.3431	0.2737	0.449	0.007	0.4470	0.2406	0.2516
42	2.3076	2.2434	0.1741	0.0642	0.0948	0.000	0.0943	0.0620	0.0242
43	2.3273	2.1043	0.1523	0.2230	0.327	0.001	0.3250	0.0474	0.0725
44	2.3749	2.4604	0.2152	-0.0855	-0.128	0.000	-0.1277	0.0946	-0.0413
45	2.5217	2.3872	0.1475	0.1345	0.197	0.000	0.1956	0.0445	0.0422
46	2.5533	2.5528	0.1585	0.000521	0.00076	0.000	0.000760	0.0514	0.0002
47	2.5688	4.0781	0.2715	-1.5093	-2.341	0.108	-2.4040	0.1506	-1.0124
48	2.5688	2.6374	0.1337	-0.0686	-0.0999	0.000	-0.0993	0.0365	-0.0193
49	2.5915	2.7372	0.2578	-0.1457	-0.224	0.001	-0.2229	0.1359	-0.0884
50	2.5915	2.1155	0.1569	0.4760	0.698	0.003	0.6963	0.0503	0.1603
51	2.6568	2.4174	0.2279	0.2393	0.362	0.002	0.3601	0.1061	0.1241
52	2.6776	3.0356	0.1833	-0.3580	-0.530	0.002	-0.5281	0.0686	-0.1434
53	2.6844	2.1671	0.2136	0.5173	0.777	0.007	0.7749	0.0932	0.2485
54	2.6912	3.0410	0.1902	-0.3497	-0.520	0.002	-0.5174	0.0739	-0.1461
55	2.7047	3.2745	0.2567	-0.5698	-0.876	0.013	-0.8745	0.1347	-0.3450
56	2.7180	2.9202	0.1819	-0.2022	-0.299	0.001	-0.2978	0.0676	-0.0802
57	2.7881	1.6902	0.2424	1.0978	1.673	0.042	1.6907	0.1200	0.6244
58	2.7942	2.3238	0.2111	0.4704	0.705	0.006	0.7033	0.0911	0.2226
59	2.8064	2.2223	0.1730	0.5841	0.862	0.005	0.8605	0.0612	0.2197
60	2.8124	2.6750	0.1708	0.1374	0.203	0.000	0.2015	0.0596	0.0507
61	2.8420	2.3703	0.2112	0.4717	0.707	0.006	0.7053	0.0912	0.2234
62	2.8536	3.5225	0.2015	-0.6689	-0.999	0.010	-0.9986	0.0830	-0.3003
63	2.8536	2.9779	0.2995	-0.1243	-0.197	0.001	-0.1956	0.1833	-0.0927
64	2.8820	3.7081	0.2096	-0.8261	-1.238	0.017	-1.2416	0.0898	-0.3900
65	2.8820	2.5535	0.1842	0.3285	0.487	0.002	0.4846	0.0694	0.1323
66	2.8876	2.7020	0.1265	0.1855	0.270	0.000	0.2683	0.0327	0.0493
67	2.9205	2.9797	0.2242	-0.0592	-0.0894	0.000	-0.0889	0.1027	-0.0301
68	2.9627	3.0701	0.1782	-0.1074	-0.159	0.000	-0.1579	0.0649	-0.0416
69	2.9627	1.5141	0.2789	1.4486	2.258	0.107	2.3133	0.1589	1.0056
70	2.9730	2.9755	0.2475	-0.002537	-0.0039	0.000	-0.003855	0.1252	-0.0015
71	3.0131	3.2932	0.1900	-0.2801	-0.416	0.002	-0.4141	0.0737	-0.1168
72	3.0374	2.0505	0.2057	0.9868	1.476	0.023	1.4861	0.0865	0.4573
73	3.0564	2.8182	0.2339	0.2382	0.361	0.002	0.3595	0.1118	0.1275
74	3.0750	3.4811	0.3083	-0.4061	-0.647	0.011	-0.6446	0.1943	-0.3165
75	3.2753	3.6611	0.2173	-0.3858	-0.580	0.004	-0.5781	0.0965	-0.1890
76	3.3375	3.6070	0.2118	-0.2695	-0.404	0.002	-0.4023	0.0917	-0.1278
77	3.3928	3.3086	0.2135	0.0842	0.126	0.000	0.1257	0.0932	0.0403
78	3.4356	3.4262	0.2377	0.009359	0.0142	0.000	0.0141	0.1155	0.0051
79	3.4579	3.4483	0.2174	0.009569	0.0144	0.000	0.0143	0.0966	0.0047
80	3.5130	3.1836	0.2085	0.3295	0.493	0.003	0.4913	0.0888	0.1534

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Student Residual	Cook's D	RStudent	Hat Diag H	DFFITS
81	3.5160	2.0708	0.1638	1.4453	2.125	0.029	2.1695	0.0548	0.5226
82	3.5308	2.9572	0.2585	0.5736	0.882	0.014	0.8813	0.1365	0.3504
83	3.5653	3.5686	0.2211	-0.003279	-0.0049	0.000	-0.004913	0.0999	-0.0016
84	3.5709	3.3765	0.2838	0.1945	0.304	0.002	0.3026	0.1646	0.1343
85	3.5877	2.6523	0.1637	0.9354	1.375	0.012	1.3825	0.0548	0.3329
86	3.6310	3.8673	0.1954	-0.2363	-0.352	0.001	-0.3501	0.0780	-0.1018
87	3.6801	3.0136	0.2019	0.6665	0.995	0.010	0.9951	0.0833	0.3001
88	3.7124	3.0111	0.1929	0.7012	1.043	0.010	1.0435	0.0761	0.2994
89	3.9843	4.2979	0.2874	-0.3136	-0.492	0.005	-0.4896	0.1688	-0.2206
90	3.9936	3.1919	0.2475	0.8017	1.225	0.024	1.2289	0.1252	0.4649
91	4.0298	3.4285	0.2782	0.6013	0.937	0.018	0.9362	0.1582	0.4058
92	4.1296	4.0529	0.3200	0.0767	0.123	0.000	0.1226	0.2093	0.0631
93	4.3851	3.8446	0.1970	0.5406	0.805	0.006	0.8038	0.0793	0.2360
94	4.6844	4.7474	0.3016	-0.0630	-0.0998	0.000	-0.0992	0.1859	-0.0474
95	5.1431	3.5854	0.2636	1.5577	2.404	0.106	2.4730	0.1420	1.0060
96	5.4775	4.1592	0.2411	1.3183	2.008	0.060	2.0436	0.1188	0.7504
97	5.5829	4.0984	0.2392	1.4845	2.258	0.075	2.3136	0.1169	0.8418

Sum of Residuals	0
Sum of Squared Residuals	43.05842
Predicted Residual SS (PRESS)	52.50892



Stepwise Regression with at most 2 steps

The REG Procedure

Model: MODEL1

Dependent Variable: lpsa

Stepwise Selection: Step 1

Variable lcavol Entered: R-Square = 0.5394 and C(p) = 27.4062

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	69.00287	69.00287	111.27	<.0001
Error	95	58.91478	0.62016		
Corrected Total	96	127.91766			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.50730	0.12194	94.76079	152.80	<.0001
lcavol	0.71932	0.06819	69.00287	111.27	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable lweight Entered: R-Square = 0.5955 and C(p) = 14.7473

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	76.17548	38.08774	69.19	<.0001
Error	94	51.74218	0.55045		
Corrected Total	96	127.91766			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-0.81344	0.65309	0.85393	1.55	0.2160
lcavol	0.65154	0.06693	52.15694	94.75	<.0001
lweight	0.66472	0.18414	7.17261	13.03	0.0005

Bounds on condition number: 1.0854, 4.3417

The stepwise method terminated because the maximum number of steps have been executed.

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	lcavol		1	0.5394	0.5394	27.4062	111.27	<.0001
2	lweight		2	0.0561	0.5955	14.7473	13.03	0.0005

Enclosure C – SAS output

Partial Correlation Analysis

The CORR Procedure

2 Partial Variables:	lcavol lweight
8 Variables:	patno age lbph svi lcp gleason pgg45 lpsa

Simple Statistics								
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Partial Variance	Partial Std Dev
lcavol	97	1.35001	1.17862	130.95093	-1.34707	3.82100		
lweight	97	3.62894	0.42841	352.00744	2.37491	4.78038		
patno	97	49.00000	28.14546	4753	1.00000	97.00000	349.06133	18.68318
age	97	63.86598	7.44512	6195	41.00000	79.00000	48.75766	6.98267
lbph	97	0.10036	1.45081	9.73449	-1.38629	2.32630	1.70734	1.30665
svi	97	0.21649	0.41399	21.00000	0	1.00000	0.12421	0.35244
lcp	97	-0.17937	1.39825	-17.39846	-1.38629	2.90417	1.08477	1.04152
gleason	97	6.75258	0.72213	655.00000	6.00000	9.00000	0.43059	0.65619
pgg45	97	24.38144	28.20403	2365	0	100.00000	659.43836	25.67953
lpsa	97	2.47839	1.15433	240.40353	-0.43078	5.58293	0.55045	0.74192

Pearson Partial Correlation Coefficients, N = 97 Prob > |r| under H0: Partial Rho=0								
	patno	age	lbph	svi	lcp	gleason	pgg45	lpsa
patno	1.00000	-0.06206	0.05978	0.32979	0.12385	0.17426	0.24514	0.90114
		0.5502	0.5650	0.0011	0.2318	0.0912	0.0166	<.0001
age	-0.06206	1.00000	0.25349	-0.00625	-0.02423	0.22950	0.21881	-0.11171
	0.5502		0.0132	0.9520	0.8157	0.0253	0.0331	0.2812
lbph	0.05978	0.25349	1.00000	-0.13663	-0.02089	0.12027	0.09134	0.09264
	0.5650	0.0132		0.1867	0.8407	0.2456	0.3787	0.3719
svi	0.32979	-0.00625	-0.13663	1.00000	0.49820	0.11578	0.29520	0.31621
	0.0011	0.9520	0.1867		<.0001	0.2638	0.0037	0.0018
lcp	0.12385	-0.02423	-0.02089	0.49820	1.00000	0.33356	0.50942	0.12579
	0.2318	0.8157	0.8407	<.0001		0.0010	<.0001	0.2245
gleason	0.17426	0.22950	0.12027	0.11578	0.33356	1.00000	0.69544	0.11766
	0.0912	0.0253	0.2456	0.2638	0.0010		<.0001	0.2561
pgg45	0.24514	0.21881	0.09134	0.29520	0.50942	0.69544	1.00000	0.18733
	0.0166	0.0331	0.3787	0.0037	<.0001	<.0001		0.0691
lpsa	0.90114	-0.11171	0.09264	0.31621	0.12579	0.11766	0.18733	1.00000
	<.0001	0.2812	0.3719	0.0018	0.2245	0.2561	0.0691	

Enclosure D – SAS program

```
Title 'Meatcolor data. First 40 observations';
proc print data=sasuser.LabM2 (obs=40);
run;

Title 'Discriminating between Unprocessed and Processed Meat
Products, 3 variables';
proc discrim data=sasuser.LabM2 method=normal pool=yes
   distance crossvalidate list crosslist;
class pr;
var LM2 aM2 bM2;
run;

Title 'Discriminating between Unprocessed and Processed Meat
Products, 2 variables';
proc discrim data=sasuser.LabM2 method=normal pool=yes
  distance;
class pr;
var aM2 bM2;
run;
```

Meatcolor data. First 40 observations

Obs	pr	LM2	aM2	bM2
1	0	59.4775	14.9929	15.4351
2	0	60.1689	14.9891	14.9946
3	0	61.3140	14.4471	14.3082
4	0	61.2555	13.7060	14.6464
5	0	61.2011	14.5520	14.5022
6	0	45.2058	16.7109	13.2196
7	0	45.5458	19.4897	15.4711
8	0	44.8497	18.3109	13.9600
9	0	49.5087	17.7987	15.3426
10	0	53.0712	15.7536	15.2295
11	0	33.0189	21.0710	12.3302
12	0	32.7802	21.7498	13.7257
13	0	31.7667	23.3476	14.7705
14	0	32.5476	21.1010	12.2997
15	0	32.2941	22.4689	13.5786
16	0	44.3855	25.7215	17.7811
17	0	46.6934	25.4414	18.5128
18	0	46.0370	25.2091	17.8593
19	0	43.2264	24.7617	17.2595
20	0	42.6608	25.0321	18.0836
21	0	35.4062	26.5550	17.3420
22	0	35.7442	26.0395	17.2176
23	0	34.6227	26.1196	16.5106
24	0	37.3099	26.7337	17.9072
25	0	35.2981	25.5879	16.5803
26	0	32.8172	26.0475	16.2892
27	0	33.1742	24.4444	14.9070
28	0	34.5928	25.1339	15.8314
29	0	33.9748	26.4634	16.5587
30	0	32.0530	25.3456	16.6212
31	0	57.1911	15.6569	14.1820
32	0	55.6153	16.1582	14.1163
33	0	60.6763	15.6528	14.4895
34	0	56.4277	15.8317	13.0373
35	0	54.1861	15.3322	13.3447
36	1	68.4445	9.5632	10.4036
37	1	69.0008	9.6144	10.2625
38	1	67.3075	9.9379	10.7644
39	1	67.7341	9.6124	10.1346
40	1	68.3540	9.2178	10.5672

Discriminating between Unprocessed and Processed Meat Products, 3 variables

The DISCRIM Procedure

Total Sample Size	60	DF Total	59
Variables	3	DF Within Classes	58
Classes	2	DF Between Classes	1

Number of Observations Read	60
Number of Observations Used	60

Class Level Information

pr	Variable Name	Frequency	Weight	Proportion	Prior Probability
0	0	35	35.0000	0.583333	0.500000
1	1	25	25.0000	0.416667	0.500000

Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	8.00043

Discriminating between Unprocessed and Processed Meat Products, 3 variables

The DISCRIM Procedure

Squared Distance to pr		
From pr	0	1
0	0	13.37960
1	13.37960	0

F Statistics, NDF=3, DDF=56 for Squared Distance to pr		
From pr	0	1
0	0	62.79699
1	62.79699	0

Prob > Mahalanobis Distance for Squared Distance to pr		
From pr	0	1
0	1.0000	<.0001
1	<.0001	1.0000

Generalized Squared Distance to pr		
From pr	0	1
0	0	13.37960
1	13.37960	0

Linear Discriminant Function for pr		
Variable	0	1
Constant	-142.78879	-135.31413
LM2	2.56181	2.69962
aM2	4.83832	4.24537
bM2	4.56789	4.05674

Discriminating between Unprocessed and Processed Meat Products, 3 variables

The DISCRIM Procedure

Classification Results for Calibration Data: SASUSER.LABM2
Resubstitution Results using Linear Discriminant Function

Posterior Probability of Membership in pr					
Obs	From pr	Classified into pr		0	1
1	0	0		0.7518	0.2482
2	0	0		0.6869	0.3131
3	0	1	*	0.4889	0.5111
4	0	1	*	0.4249	0.5751
5	0	0		0.5331	0.4669
6	0	0		0.9508	0.0492
7	0	0		0.9967	0.0033
8	0	0		0.9871	0.0129
9	0	0		0.9837	0.0163
10	0	0		0.9119	0.0881
11	0	0		0.9989	0.0011
12	0	0		0.9996	0.0004
13	0	0		0.9999	0.0001
14	0	0		0.9989	0.0011
15	0	0		0.9998	0.0002
16	0	0		1.0000	0.0000
17	0	0		1.0000	0.0000
18	0	0		1.0000	0.0000
19	0	0		1.0000	0.0000
20	0	0		1.0000	0.0000
21	0	0		1.0000	0.0000
22	0	0		1.0000	0.0000
23	0	0		1.0000	0.0000
24	0	0		1.0000	0.0000
25	0	0		1.0000	0.0000
26	0	0		1.0000	0.0000
27	0	0		1.0000	0.0000
28	0	0		1.0000	0.0000
29	0	0		1.0000	0.0000
30	0	0		1.0000	0.0000

Table continues

Posterior Probability of Membership in pr					
Obs	From pr	Classified into pr		0	1
31	0	0		0.7643	0.2357
32	0	0		0.8399	0.1601
33	0	0		0.7008	0.2992
34	0	0		0.6901	0.3099
35	0	0		0.7252	0.2748
36	1	1		0.0027	0.9973
37	1	1		0.0024	0.9976
38	1	1		0.0047	0.9953
39	1	1		0.0027	0.9973
40	1	1		0.0024	0.9976
41	1	1		0.0002	0.9998
42	1	1		0.0004	0.9996
43	1	1		0.0037	0.9963
44	1	1		0.0029	0.9971
45	1	1		0.0033	0.9967
46	1	1		0.0011	0.9989
47	1	1		0.0015	0.9985
48	1	1		0.0007	0.9993
49	1	1		0.0011	0.9989
50	1	1		0.0018	0.9982
51	1	1		0.0045	0.9955
52	1	1		0.0046	0.9954
53	1	1		0.0047	0.9953
54	1	1		0.0045	0.9955
55	1	1		0.0045	0.9955
56	1	1		0.0002	0.9998
57	1	1		0.0002	0.9998
58	1	1		0.0002	0.9998
59	1	1		0.0001	0.9999
60	1	1		0.0001	0.9999

* Misclassified observation

Enclosure D – SAS output

Discriminating between Unprocessed and Processed Meat Products, 3 variables

The DISCRIM Procedure

Classification Results for Calibration Data: SASUSER.LABM2

Cross-validation Results using Linear Discriminant Function

Posterior Probability of Membership in pr					
Obs	From pr	Classified into pr		0	1
1	0	0		0.7177	0.2823
2	0	0		0.6451	0.3549
3	0	1	*	0.4328	0.5672
4	0	1	*	0.3724	0.6276
5	0	1	*	0.4775	0.5225
6	0	0		0.9376	0.0624
7	0	0		0.9964	0.0036
8	0	0		0.9854	0.0146
9	0	0		0.9828	0.0172
10	0	0		0.9088	0.0912
11	0	0		0.9984	0.0016
12	0	0		0.9996	0.0004
13	0	0		0.9999	0.0001
14	0	0		0.9985	0.0015
15	0	0		0.9997	0.0003
16	0	0		1.0000	0.0000
17	0	0		1.0000	0.0000
18	0	0		1.0000	0.0000
19	0	0		1.0000	0.0000
20	0	0		1.0000	0.0000
21	0	0		1.0000	0.0000
22	0	0		1.0000	0.0000
23	0	0		1.0000	0.0000
24	0	0		1.0000	0.0000
25	0	0		1.0000	0.0000
26	0	0		1.0000	0.0000
27	0	0		1.0000	0.0000
28	0	0		1.0000	0.0000
29	0	0		1.0000	0.0000
30	0	0		1.0000	0.0000

Table continues

Posterior Probability of Membership in pr					
Obs	From pr	Classified into pr		0	1
31	0	0		0.7511	0.2489
32	0	0		0.8336	0.1664
33	0	0		0.6502	0.3498
34	0	0		0.6669	0.3331
35	0	0		0.7098	0.2902
36	1	1		0.0032	0.9968
37	1	1		0.0029	0.9971
38	1	1		0.0055	0.9945
39	1	1		0.0032	0.9968
40	1	1		0.0028	0.9972
41	1	1		0.0003	0.9997
42	1	1		0.0005	0.9995
43	1	1		0.0057	0.9943
44	1	1		0.0046	0.9954
45	1	1		0.0054	0.9946
46	1	1		0.0014	0.9986
47	1	1		0.0021	0.9979
48	1	1		0.0009	0.9991
49	1	1		0.0014	0.9986
50	1	1		0.0026	0.9974
51	1	1		0.0052	0.9948
52	1	1		0.0053	0.9947
53	1	1		0.0054	0.9946
54	1	1		0.0051	0.9949
55	1	1		0.0051	0.9949
56	1	1		0.0002	0.9998
57	1	1		0.0002	0.9998
58	1	1		0.0002	0.9998
59	1	1		0.0002	0.9998
60	1	1		0.0002	0.9998

* Misclassified observation

Discriminating between Unprocessed and Processed Meat Products, 2 variables

The DISCRIM Procedure

Total Sample Size	60	DF Total	59
Variables	2	DF Within Classes	58
Classes	2	DF Between Classes	1

Number of Observations Read	60
Number of Observations Used	60

Class Level Information					
pr	Variable Name	Frequency	Weight	Proportion	Prior Probability
0	0	35	35.0000	0.583333	0.500000
1	1	25	25.0000	0.416667	0.500000

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
2	4.52175

Squared Distance to pr		
From pr	0	1
0	0	12.76396
1	12.76396	0

F Statistics, NDF=2, DDF=57 for Squared Distance to pr		
From pr	0	1
0	0	91.46590
1	91.46590	0

Prob > Mahalanobis Distance for Squared Distance to pr		
From pr	0	1
0	1.0000	<.0001
1	<.0001	1.0000

Generalized Squared Distance to pr		
From pr	0	1
0	0	12.76396
1	12.76396	0

Linear Discriminant Function for pr		
Variable	0	1
Constant	-36.41494	-17.18799
aM2	1.20604	0.41770
bM2	3.09173	2.50118