Technical University of Denmark

Written examination, date: 11. December 2012 *Page 1 of 16 pages Enclosure: 12 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 5 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	4	4	4	5	5	5	5	5	5	5
Question	4.6	4.7	4.8	5.1	5.2	5.3	5.4	5.5	5.6	5.7
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. Please check that all pages of the exam paper and the enclosure are present.

Problem 1

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

The data are observations of monthly beer sales and the highest and lowest temperature for that month. The data are taken from the SAS Documentation. We want to analyze the connection between the beer sales and the temperature measurements by means of regression analyses. We introduce some new independent variables by squaring and multiplying the original independent variables. We suppose that the normal assumptions for regression models are fulfilled.

Question 1.1.

We consider the model with all 5parameters plus the intercept, and we want to test the hypothesis that all variables involving the low temperatures are zero. The value of the test statistic is:

1 ☐ $1.64^2 + 1.72^2 + (-1.71)^2$

2 ☐ $\frac{48825/5}{23835/54}$

3 ☐ $\frac{(25186-23835)/3}{23835/54}$

4 ☐ $\frac{(47473-23835)/52}{23835/54}$

5 ☐ $\frac{(25186-23835)/3}{72659/59}$

6 ☐ Don't know.

The problem continues on the next page

Question 1.2.

The test statistic for the hypothesis described above is – if the hypothesis is true – distributed as:

1 ☐ F(3, 54)

2 ☐ t(3)

3 ☐ F(5,54)

4 ☐ F(3, 59)

5 ☐ t(59)

6 ☐ Don't know.

Question 1.3.

We want – based on model 5 – to reduce the number of parameters with the backward elimination procedure. Which variable will be removed first?

1 ☐ Hightemp

2 ☐ hightsq

3 ☐ Lowtemp

4 ☐ lowtsq

5 ☐ highlow

6 ☐ Don't know.

Question 1.4.

If we remove observation no 9 from the analysis, we obtain a new estimate of the standard deviation $\hat{\sigma}(i)$..The ratio between the old estimate $\hat{\sigma}$and the new estimate $\hat{\sigma}(i)$ is:

1 ☐ -1.040

2 ☐ -1.055

3 ☐ 1.070

4 ☐ 1.085

5 ☐ 1.100

6 ☐ Don't know.

The problem continues on the next page

Question 1.5.

What is the 95% confidence interval for the expected value (mean value) corresponding to observation no 9?

1 ☐ $216.0575 \pm t(58)_{0.975} \times 3.8809$

2 ☐ $216.0575 \pm t(2)_{0.975} \times 20.8650$

3 ☐ $152.500 \pm t(58)_{0.975} \times \sqrt{435.34191}$

4 ☐ $0.18327 \times 19.4^2 \pm t(58)_{0.975} \times 3.8809$

6 ☐ Don't know.

Problem 2

We consider a random variable

$$\begin{bmatrix} Y \\ X_1 \\ X_2 \end{bmatrix}$$

with the dispersion matrix

$$S = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Question 2.1.

We want to determine the linear combination $Z = aX_1 + bX_2$ of the two X-variables that has the maximum correlation with Y. Z is equal to:

1 ☐ $\frac{1}{7}X_1 + \frac{2}{7}X_2$

2 ☐ $\frac{1}{2}X_1 + \frac{1}{2}X_2$

3 ☐ $\frac{3}{7}X_1 + \frac{2}{7}X_2$

4 ☐ $3X_1 + 4X_2$

5 ☐ $4X_1 + 2X_2$

6 ☐ Don't know

The problem continues on the next page

Question 2.2.

The variance of Z is:

1 ☐ 6

2 ☐ $\frac{13}{49}$

3 ☐ 20

4 ☐ $\frac{5}{7}$

5 ☐ $\frac{56}{49}$

6 ☐ Don't know.

Question 2.3.

The squared correlation between Y and Z is:

1 ☐ $\frac{4}{49}$

2 ☐ $\frac{1}{2}$

3 ☐ $\frac{1}{4}$

4 ☐ $\frac{2}{7}$

5 ☐ $-\frac{1}{2}$

6 ☐ Don't know.

The problem continues on the next page

Question 2.4.

We now assume that S is obtained as the empirical dispersion matrix based on 13 independent observations of $[Y, X_1, X_2]'$. Then the value of the usual statistic for testing whether the maximum correlation is equal to 0 against all alternatives is

1 ☐ 1

2 ☐ 2

3 ☐ 3

4 ☐ 4

5 ☐ 5

6 ☐ Don't know.

Question 2.5.

The test statistic obtained above should be compared with quantiles in the following distribution:

1 ☐ -t(12)

2 ☐ -F(1, 12)

3 ☐ $\chi^2(12)$

4 ☐ F(2,10)

5 ☐ t(10)

6 ☐ Don't know.

Problem 3

Enclosure B belongs to this problem. The data are measurements of the subsurface laser scattering from two pieces of lard. Various features are extracted from measurements taken at an experimental set up at DTU. The first feature is the slope of the log-log curve of a measured intensity profile, the second feature extracted is the y-axis interception of the log-log curve. These features have been determined at many wavelengths, but we only consider a few of those. The end goal of the experiments is to assess whether such measurements can be used to distinguish lard from boars with high hormone concentrations from lard from boars wit low hormone concentrations. The presence of hormones may influence the taste of the pig meat.

In the analysis below, we have limited ourselves to consider the intercept feature derived at two wavelengths, 500 and 750 nm.

The problem continues on the next page

Understanding the nature of the measurements is not crucial for solving the present problem.

Question 3.1.

At first we are interested in estimating the correlation between the two variables int500 and int750. The estimate is:

1 ☐ 1.0

2 ☐ 0.9

3 ☐ 0.8

4 ☐ 0.7

5 ☐ 0.6

6 ☐ Don't know.

Question 3.2.

Hotellings T^2statistic for testing whether the mean values for the two pieces are the same is

1 ☐ 85.57199

2 ☐ 85.57199^2

3 ☐ $\sqrt{85.57199}$

4 ☐ $\frac{110}{21} \times 85.57199$

5 ☐ $\frac{110}{19} \times 85.57199$

6 ☐ Don't know.

Question 3.3.

The degrees of freedom for the test statistic related to the test described in Question 3.2 are:

1 ☐ (2, 21)

2 ☐ (2, 17)

3 ☐ (3, 21)

4 ☐ (19, 21)

5 ☐ (2, 18)

6 ☐ Don't know.

The problem continues on the next page

Question 3.4.

The generalized variance for the within-class dispersion (covariance) matrix is:

1 ☐ $0.0000524870 + 0.0000465714$

2 ☐ $e^{-21.54737}$

3 ☐ $\left(\sqrt{0.0000524870} + \sqrt{0.0000465714}\right)^2$

4 ☐ $(-21.54737)^2$

5 ☐ $0.0000524870 \times 0.0000465714$

6 ☐ Don't know.

Question 3.5.

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 ☐ -85275

2 ☐ 0

3 ☐ 85.57199

4 ☐ 2933

5 ☐ 262

6 ☐ Don't know.

Problem 4

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

The data are of the same type as described in the introduction to Problem 3. Now we are however, considering as well slope as intercept variables, and we consider wavelengths 500, 580, 750. 830, and 910 nm. We are interested in analyzing the correlation structure between the different variables.

Understanding the nature of the measurements is not crucial for solving the present problem.

The problem continues on the next page

Question 4.1.

If we want to describe at least 95% of the total variation, what is the smallest number of principal component we should use?

1 ☐ 1

2 ☐ 2

3 ☐ 3

4 ☐ 4

5 ☐ 5

6 ☐ Don't know

Question 4.2.

The fraction of the variation of slo910 that is explained by the first principal component may be written as $c\lambda_1$ where c is:

1 ☐ 0.1

2 ☐ 0.2

3 ☐ 0.3

4 ☐ 0.4

5 ☐ 0.5

6 ☐ Don't know.

The problem continues on the next page

Question 4.3.

Consider the following statements on interpretation of an arbitrary PC (principal component):

A. The component is a contrast between the short wavelength slope and long wavelength intercept variables.
B. The component is roughly an average of all measurements.
C. The component is a contrast between the short wavelength and the long wavelength variables.
D. The component is a contrast between the slope and the intercept variables.

For (principal component 1, principal component 2) the following characterization is adequate:

1 ☐ (D, B)

2 ☐ (D, C)

3 ☐ (A, B)

4 ☐ (A, C)

5 ☐ (C, D)

6 ☐ Don't know.

Question 4.4.

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

1 ☐ 0.78165

2 ☐ 0.9450

3 ☐ 0.78165^2

4 ☐ $(4.8452152^2 + 4.6050509^2)/100$

5 ☐ $0.80511^2 + 0.36531^2$

6 ☐ Don't know.

The problem continues on the next page

Question 4.5.

How much of the variation of int500 is explained by the two VARIMAX rotated factors?

1 ☐ $0.16135^2 + 0.94880^2$

2 ☐ 0.926253^2

3 ☐ $\frac{1}{2}(0.16135^2 + 0.94880^2)$

4 ☐ $1 - (0.16135^2 + 0.94880^2)$

5 ☐ $1 - \frac{1}{2}(0.16135^2 + 0.94880^2)$

6 ☐ Don't know.

Question 4.6.

Consider the following description of the correlation between an arbitrary factor and the original variables:

A. Large values of the factor correspond primarily to large values of the slope variables.
B. Large values of the factor correspond basically to large values of all variables.
C. Large values of the factor correspond to small values of the slope variables and large values of the intercept variables.
D. Large values of the factor correspond primarily to large values of the intercept variables

The outcome of the VARIMAX rotation ($(F_1, F_2) \rightarrow (RotF_1, RotF_2)$ corresponds to the following change in interpretation:

1 ☐ (A, B) → (C, D)

2 ☐ (C, B) → (A, D)

3 ☐ (B, D) → (C, A)

4 ☐ (D, A) → (B, C)

5 ☐ (A, C) → (B, D)

6 ☐ Don't know.

Question 4.7.

We are now considering the relation between the slope variables and the intercept variables by looking at uncorrelated linear combinations of the slope variables and uncorrelated linear combinations of the intercept variables. We limit the investigation to the shortest wavelength 500 nm and the longest wavelength 910 nm.

The problem continues on the next page

The maximum correlation between a linear combination of the slope variables and a linear combination of the intercept variables is

1 ☐ 0.999013

2 ☐ 0.37736560

3 ☐ 0.999507

4 ☐ $\sqrt{0.999507 \times 0.789072}$

5 ☐ $\sqrt{0.999451}$

6 ☐ Don't know.

Question 4.8.

We now look at the canonical structure, i.e the correlations between the original variables and the canonical variables. The slo500 variable has its maximum correlation with

1 ☐ The first canonical variable of the slope variables.

2 ☐ The second canonical variable of the slope variables.

3 ☐ The first canonical variable of the intercept variables.

4 ☐ The second canonical variable of the intercept variables.

5 ☐ slo910 .

6 ☐ Don't know.

Problem 5

We consider the model

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

or

$$\boldsymbol{Y} = \boldsymbol{x\theta} + \boldsymbol{\epsilon}$$

We assume as usual that $\boldsymbol{\varepsilon} \in N_4(\mathbf{0}, \sigma^2 \boldsymbol{I})$. Suppose that we obtained the following values of $\boldsymbol{Y}$

$$\boldsymbol{y} = \begin{bmatrix} 2 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

Question 5.1.

The maximum likehood estimator $\widehat{\boldsymbol{\theta}}$ becomes

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

2 ☐ $\begin{bmatrix} 4 \\ 20 \end{bmatrix}$

3 ☐ $\begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$

4 ☐ $\begin{bmatrix} 1.5 \\ 0.3 \end{bmatrix}$

5 ☐ $\begin{bmatrix} 2.5 \\ 0.3 \end{bmatrix}$

6 ☐ Don't know

The problem continues on the next page

Question 5.2.

We have that

$$(\boldsymbol{y} - \boldsymbol{x}\,\widehat{\boldsymbol{\theta}})'(\boldsymbol{y} - \boldsymbol{x}\,\widehat{\boldsymbol{\theta}}) = 3.2$$

From this we obtain that the unbiased estimator $\hat{\sigma}^2$ is equal to:

1 ☐ 3.2

2 ☐ 1.6

3 ☐ 1.0667

4 ☐ 0.8

5 ☐ 0.64

6 ☐ Don't know.

Question 5.3.

The estimated variance $\hat{V}(\hat{\beta})$ is equal to

1 ☐ 0.08

2 ☐ 0.32

3 ☐ 0.1067

4 ☐ 3.2

5 ☐ 0.16

6 ☐ Don't know.

The problem continues on the next page

Question 5.4.

The correlation between $\hat{\alpha}$ and $\hat{\beta}$ is:

1 ☐ -0.5

2 ☐ -0.25

3 ☐ 0

4 ☐ 0.25

5 ☐ 0.5

6 ☐ Don't know.

Question 5.5.

We now assume that we besides Y also observe a random variable Z giving the model

$$\begin{bmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \\ Y_3 & Z_3 \\ Y_4 & Z_4 \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \\ \varepsilon_{31} & \varepsilon_{32} \\ \varepsilon_{41} & \varepsilon_{42} \end{bmatrix}$$

with the usual assumptions. We now want to test the hypothesis

$$H_0 : \beta = \delta = 0$$

against all alternatives. This hypothesis may also be written

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

Here the matrix $\boldsymbol{A}$ is:

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

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

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

4 ☐ $[0 \quad 1]$

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

6 ☐ Don't know.

The problem continues on the next page

Question 5.6.

The matrix $\boldsymbol{B}$ is:

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

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

3 ☐ $[0 \quad 1]$

4 ☐ $[1 \quad 1]$

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

6 ☐ Don’t know.

Question 5.7.

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

1 ☐ U(1, 1, 1)

2 ☐ U(1, 1, 2)

3 ☐ U(1, 2, 1)

4 ☐ U(2, 1, 1)

5 ☐ U(2, 1, 2)

6 ☐ Don’t know.

Enclosure A – SAS program

```
data b;
set sasuser.beer;
hightsq=hightemp*hightemp; lowtsq=lowtemp*lowtemp;
highlow=hightemp*lowtemp;
run;

Title 'Beer sales data';
proc print data=b;
run;

proc reg;
model sales = hightemp lowtemp hightsq lowtsq highlow;
model sales = hightemp hightsq;
model sales = hightsq/r influence;
run;
```


Enclosure A – SAS output

Beer sales data

Obs	Month	Sales	HighTemp	LowTemp	hightsq	lowtsq	highlow
1	1	130.2	3.5	-3.2	12.25	10.24	-11.20
2	2	137	6.9	0.8	47.61	0.64	5.52
3	3	171.4	11.5	2.9	132.25	8.41	33.35
4	4	178.5	13	4.7	169.00	22.09	61.10
5	5	211.9	16.9	8.6	285.61	73.96	145.34
6	6	227.4	19.9	10.3	396.01	106.09	204.97
7	7	253.8	22.7	13	515.29	169.00	295.10
8	8	254.6	22.1	13.2	488.41	174.24	291.72
9	9	152.5	19.4	11.5	376.36	132.25	223.10
10	10	177.2	13.9	7.1	193.21	50.41	98.69
11	11	161.8	8.8	0.8	77.44	0.64	7.04
12	12	179.3	8.6	3.1	73.96	9.61	26.66
13	13	156.8	3.7	-2.8	13.69	7.84	-10.36
14	14	130.4	8.4	2.9	70.56	8.41	24.36
15	15	166.1	8.8	2.5	77.44	6.25	22.00
16	16	205.4	14.2	5.7	201.64	32.49	80.94
17	17	209.1	16	8.5	256.00	72.25	136.00
18	18	235.4	17.8	10.6	316.84	112.36	188.68
19	19	227.3	20.7	12.4	428.49	153.76	256.68
20	20	203.9	20.3	10.2	412.09	104.04	207.06
21	21	182.5	17.5	10.3	306.25	106.09	180.25
22	22	200.4	15	6.5	225.00	42.25	97.50
23	23	161	9.8	4.1	96.04	16.81	40.18
24	24	195.9	7.1	2.7	50.41	7.29	19.17
25	25	142.8	8.2	2.7	67.24	7.29	22.14
26	26	161.8	8.4	1.9	70.56	3.61	15.96
27	27	173.3	11.8	3.8	139.24	14.44	44.84
28	28	185.6	12.4	5	153.76	25.00	62.00
29	29	192.1	16	8.9	256.00	79.21	142.40
30	30	223.8	17.4	10.3	302.76	106.09	179.22

Beer sales data

Obs	Month	Sales	HighTemp	LowTemp	hightsq	lowtsq	highlow
31	31	258.5	20.6	13.5	424.36	182.25	278.10
32	32	241.3	23.4	14.4	547.56	207.36	336.96
33	33	202.2	19	10.8	361.00	116.64	205.20
34	34	154.9	12.6	6.1	158.76	37.21	76.86
35	35	149.3	10.8	4.4	116.64	19.36	47.52
36	36	194.4	6.4	1.3	40.96	1.69	8.32
37	37	128.9	4.3	-0.5	18.49	0.25	-2.15
38	38	150.7	7	1.4	49.00	1.96	9.80
39	39	173.4	9.3	1.6	86.49	2.56	14.88
40	40	187	11.8	3.5	139.24	12.25	41.30
41	41	214	16.4	7.8	268.96	60.84	127.92
42	42	252	21.2	12.2	449.44	148.84	258.64
43	43	239.2	20.8	13.2	432.64	174.24	274.56
44	44	231	20.7	12.9	428.49	166.41	267.03
45	45	158.5	18.7	10.5	349.69	110.25	196.35
46	46	161.9	13.7	6.7	187.69	44.89	91.79
47	47	155.2	7.7	0.7	59.29	0.49	5.39
48	48	175.2	6.9	1.3	47.61	1.69	8.97
49	49	139.2	9	3.5	81.00	12.25	31.50
50	50	141.8	9.5	3.2	90.25	10.24	30.40
51	51	195.9	11.8	4.5	139.24	20.25	53.10
52	52	196.4	13.4	5	179.56	25.00	67.00
53	53	192.6	17.9	9.4	320.41	88.36	168.26
54	54	220.9	19.1	11.4	364.81	129.96	217.74
55	55	200.8	20.5	12.7	420.25	161.29	260.35
56	56	242.7	21.8	13.6	475.24	184.96	296.48
57	57	165.2	17.5	9.4	306.25	88.36	164.50
58	58	174.4	13	5.9	169.00	34.81	76.70
59	59	159.5	9.9	5.2	98.01	27.04	51.48
60	60	167.9	3.8	-2.5	14.44	6.25	-9.50

Beer sales data

The REG Procedure

Model: MODEL1

Dependent Variable: Sales

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	48825	9764.94047	22.12	<.0001
Error	54	23835	441.38310		
Corrected Total	59	72659			

Root MSE	21.00912	R-Square	0.6720
Dependent Mean	186.90167	Adj R-Sq	0.6416
Coeff Var	11.24073		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	275.33286	86.57127	3.18	0.0024
HighTemp	1	-45.36688	28.25263	-1.61	0.1142
LowTemp	1	54.92091	33.58710	1.64	0.1078
hightsq	1	4.06330	2.30003	1.77	0.0829
lowtsq	1	5.86001	3.40376	1.72	0.0909
highlow	1	-9.49010	5.55574	-1.71	0.0933

Beer sales data

The REG Procedure

Model: MODEL2

Dependent Variable: Sales

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	47473	23736	53.72	<.0001
Error	57	25186	441.86809		
Corrected Total	59	72659			

Root MSE	21.02066	R-Square	0.6534
Dependent Mean	186.90167	Adj R-Sq	0.6412
Coeff Var	11.24691		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	153.07466	16.50905	9.27	<.0001
HighTemp	1	-1.01261	2.67433	-0.38	0.7064
hightsq	1	0.21932	0.09685	2.26	0.0274

Beer sales data

The REG Procedure

Model: MODEL3

Dependent Variable: Sales

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	47410	47410	108.90	<.0001
Error	58	25250	435.34191		
Corrected Total	59	72659			

Root MSE	20.86485	R-Square	0.6525
Dependent Mean	186.90167	Adj R-Sq	0.6465
Coeff Var	11.16354		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	147.08298	4.67065	31.49	<.0001
hightsq	1	0.18327	0.01756	10.44	<.0001

Enclosure A – SAS output

The REG Procedure
Model: MODEL3
Dependent Variable: Sales. Obs. no 1-30

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D	RStudent	Hat Diag H
1	130.2000	149.3280	4.4966	-19.1280	20.375	-0.939	0.021	-0.9378	0.0464
2	137.0000	155.8084	4.0166	-18.8084	20.475	-0.919	0.016	-0.9174	0.0371
3	171.4000	171.3202	3.0798	0.0798	20.636	0.00387	0.000	0.003832	0.0218
4	178.5000	178.0554	2.8239	0.4446	20.673	0.0215	0.000	0.0213	0.0183
5	211.9000	199.4263	2.9489	12.4737	20.655	0.604	0.004	0.6006	0.0200
6	227.4000	219.6591	4.1363	7.7409	20.451	0.379	0.003	0.3757	0.0393
7	253.8000	241.5194	5.8863	12.2806	20.017	0.613	0.016	0.6102	0.0796
8	254.6000	236.5931	5.4708	18.0069	20.135	0.894	0.030	0.8927	0.0687
9	152.5000	216.0579	3.8809	-63.5579	20.501	-3.100	0.172	-3.3649	0.0346
10	177.2000	182.4923	2.7266	-5.2923	20.686	-0.256	0.001	-0.2538	0.0171
11	161.8000	161.2753	3.6450	0.5247	20.544	0.0255	0.000	0.0253	0.0305
12	179.3000	160.6375	3.6864	18.6625	20.537	0.909	0.013	0.9074	0.0312
13	156.8000	149.5919	4.4764	7.2081	20.379	0.354	0.003	0.3510	0.0460
14	130.4000	160.0144	3.7275	-29.6144	20.529	-1.443	0.034	-1.4564	0.0319
15	166.1000	161.2753	3.6450	4.8247	20.544	0.235	0.001	0.2329	0.0305
16	205.4000	184.0372	2.7076	21.3628	20.688	1.033	0.009	1.0332	0.0168
17	209.1000	193.9997	2.7782	15.1003	20.679	0.730	0.005	0.7272	0.0177
18	235.4000	205.1498	3.2115	30.2502	20.616	1.467	0.026	1.4824	0.0237
19	227.3000	225.6117	4.5843	1.6883	20.355	0.0829	0.000	0.0822	0.0483
20	203.9000	222.6061	4.3545	-18.7061	20.405	-0.917	0.019	-0.9154	0.0436
21	182.5000	203.2090	3.1141	-20.7090	20.631	-1.004	0.011	-1.0038	0.0223
22	200.4000	188.3184	2.6971	12.0816	20.690	0.584	0.003	0.5806	0.0167
23	161.0000	164.6841	3.4334	-3.6841	20.580	-0.179	0.000	-0.1775	0.0271
24	195.9000	156.3215	3.9803	39.5785	20.482	1.932	0.071	1.9805	0.0364
25	142.8000	159.4060	3.7680	-16.6060	20.522	-0.809	0.011	-0.8067	0.0326
26	161.8000	160.0144	3.7275	1.7856	20.529	0.0870	0.000	0.0862	0.0319
27	173.3000	172.6013	3.0222	0.6987	20.645	0.0338	0.000	0.0336	0.0210
28	185.6000	175.2623	2.9154	10.3377	20.660	0.500	0.002	0.4971	0.0195
29	192.1000	193.9997	2.7782	-1.8997	20.679	-0.0919	0.000	-0.0911	0.0177
30	223.8000	202.5694	3.0838	21.2306	20.636	1.029	0.012	1.0294	0.0218

Enclosure B – SAS program

```
Title 'Reflectance Measurements on Lard';

proc print data=sasuser.multi;
var piece int500 int750;
run;

proc discrim data=sasuser.multi method=normal pool=yes wcov
pcov distance crossvalidate crosslist crosslisterr;
class piece;
var int500 int750;
run;
```

Enclosure B – SAS output

Reflectance Measurements on Lard

Obs	Piece	int500	int750
1	1	2.8820	2.7760
2	1	2.8840	2.7749
3	1	2.8904	2.7762
4	1	2.8930	2.7773
5	1	2.8914	2.7783
6	1	2.8934	2.7774
7	1	2.8851	2.7693
8	1	2.8877	2.7706
9	1	2.8838	2.7666
10	1	2.8816	2.7647
11	2	2.8805	2.7938
12	2	2.8810	2.7910
13	2	2.8800	2.7920
14	2	2.8806	2.7935
15	2	2.8798	2.7922
16	2	2.8688	2.7872
17	2	2.8747	2.7897
18	2	2.8747	2.7897
19	2	2.8755	2.7921
20	2	2.8555	2.7703
21	2	2.8586	2.7730

The DISCRIM Procedure

Total Sample Size	21	DF Total	20
Variables	2	DF Within Classes	19
Classes	2	DF Between Classes	1

Number of Observations Read	21
Number of Observations Used	21

Class Level Information					
Piece	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	1	10	10.0000	0.476190	0.500000
2	2	11	11.0000	0.523810	0.500000

Enclosure B – SAS output

The DISCRIM Procedure

Within-Class Covariance Matrices

Piece = 1, DF = 9		
Variable	int500	int750
int500	0.0000205338	0.0000153109
int750	0.0000153109	0.0000242134

Piece = 2, DF = 10		
Variable	int500	int750
int500	0.0000812449	0.0000713132
int750	0.0000713132	0.0000666936

Pooled Within-Class Covariance Matrix, DF = 19		
Variable	int500	int750
int500	0.0000524870	0.0000447858
int750	0.0000447858	0.0000465714

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
2	-21.54737

Squared Distance to Piece		
From Piece	1	2
1	0	85.57199
2	85.57199	0

F Statistics, NDF=xx, DDF=xx for Squared Distance to Piece		
From Piece	1	2
1	0	212.32147
2	212.32147	0

Prob > Mahalanobis Distance for Squared Distance to Piece		
From Piece	1	2
1	1.0000	<.0001
2	<.0001	1.0000

Enclosure B – SAS output

The DISCRIM Procedure

Linear Discriminant Function for Piece		
Variable	1	2
Constant	-85144	-85406
int500	23405	20472
int750	37038	40171

The DISCRIM Procedure

Classification Summary for Calibration Data: SASUSER.MULTI

Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Piece			
From Piece	1	2	Total
1	10 100.00	0 0.00	10 100.00
2	0 0.00	11 100.00	11 100.00
Total	10 47.62	11 52.38	21 100.00
Priors	0.5	0.5	

Error Count Estimates for Piece			
	1	2	Total
Rate	0.0000	0.0000	0.0000
Priors	0.5000	0.5000	

Enclosure B – SAS output

The DISCRIM Procedure

Classification Results for Calibration Data: SASUSER.MULTI

Cross-validation Results using Linear Discriminant Function

Posterior Probability of Membership in Piece				
Obs	From Piece	Classified into Piece	1	2
1	1	1	1.0000	0.0000
2	1	1	1.0000	0.0000
3	1	1	1.0000	0.0000
4	1	1	1.0000	0.0000
5	1	1	1.0000	0.0000
6	1	1	1.0000	0.0000
7	1	1	1.0000	0.0000
8	1	1	1.0000	0.0000
9	1	1	1.0000	0.0000
10	1	1	1.0000	0.0000
11	2	2	0.0000	1.0000
12	2	2	0.0000	1.0000
13	2	2	0.0000	1.0000
14	2	2	0.0000	1.0000
15	2	2	0.0000	1.0000
16	2	2	0.0000	1.0000
17	2	2	0.0000	1.0000
18	2	2	0.0000	1.0000
19	2	2	0.0000	1.0000
20	2	2	0.0000	1.0000
21	2	2	0.0000	1.0000

Enclosure B – SAS output

The DISCRIM Procedure

Classification Summary for Calibration Data: SASUSER.MULTI

Cross-validation Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Piece			
From Piece	1	2	Total
1	10	0	10
	100.00	0.00	100.00
2	0	11	11
	0.00	100.00	100.00
Total	10	11	21
	47.62	52.38	100.00
Priors	0.5	0.5	

Error Count Estimates for Piece			
	1	2	Total
Rate	0.0000	0.0000	0.0000
Priors	0.5000	0.5000	

Enclosure C – SAS program

```
Title 'Reflectance Measurements on Second Lard Data Set';
Title2 'The Correlation Matrix';

ods graphics on;

data typelard(type=corr);
set sasuser.lardcorr;
_type_ = 'corr';
run;

proc print data=typelard;
var _name_ slo500 slo580 slo750 slo830 slo910 int500 int580
int750 int830 int910;
run;

Title'Principal Component Analysis on the Lard Data';
proc princomp data=typelard;
var slo500 slo580 slo750 slo830 slo910 int500 int580 int750 int830
int910;
run;

Title 'Factor Analysis on the Lard Data';
proc factor data=typelard rotate=varimax plots=(scree
initloadings(vector) loadings(vector)) ;
var slo500 slo580 slo750 slo830 slo910 int500 int580 int750 int830
int910;
run;

Title 'Canonical Analysis on the Lard Data';
proc cancorr data=sasuser.lardcorr;
var slo500 slo910;
with int500 int910;
run;

ods graphics off;
```

Enclosure C – SAS output

Reflectance Measurements on Second Lard Data Set

The Correlation Matrix

Obs	_NAME_	slo500	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
1	slo500	1.00000	0.54123	0.50443	0.67020	0.62326	0.39537	0.12104	0.26611	-0.17035	-0.21488
2	slo580	0.54123	1.00000	0.96360	0.93989	0.92607	-0.22834	-0.63711	-0.50498	-0.75865	-0.81688
3	slo750	0.50443	0.96360	1.00000	0.96283	0.95643	-0.24279	-0.62413	-0.54306	-0.80164	-0.85952
4	slo830	0.67020	0.93989	0.96283	1.00000	0.99155	-0.00107	-0.42261	-0.30682	-0.65176	-0.71651
5	slo910	0.62326	0.92607	0.95643	0.99155	1.00000	-0.00749	-0.40729	-0.30009	-0.62054	-0.69500
6	int500	0.39537	-0.22834	-0.24279	-0.00107	-0.00749	1.00000	0.85029	0.90585	0.68150	0.65104
7	int580	0.12104	-0.63711	-0.62413	-0.42261	-0.40729	0.85029	1.00000	0.95477	0.91718	0.89811
8	int750	0.26611	-0.50498	-0.54306	-0.30682	-0.30009	0.90585	0.95477	1.00000	0.86818	0.85097
9	int830	-0.17035	-0.75865	-0.80164	-0.65176	-0.62054	0.68150	0.91718	0.86818	1.00000	0.98827
10	int910	-0.21488	-0.81688	-0.85952	-0.71651	-0.69500	0.65104	0.89811	0.85097	0.98827	1.00000

The PRINCOMP Procedure

Observations	10000
Variables	10

Eigenvalues of the Correlation Matrix

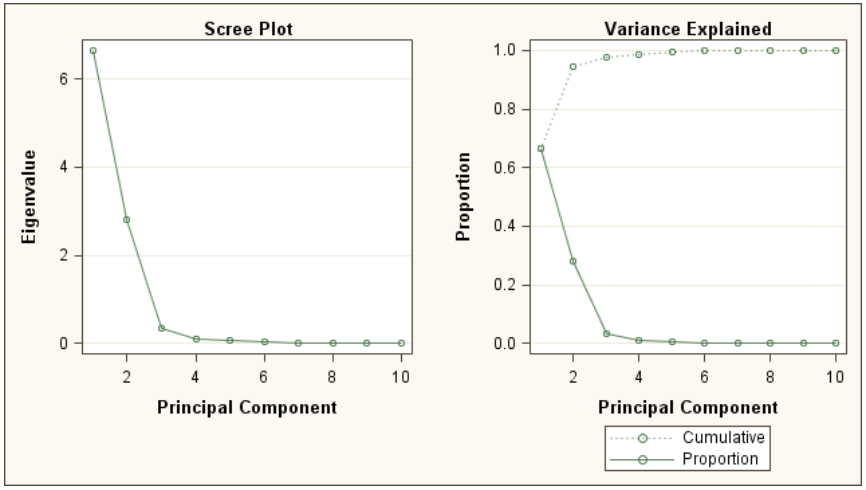
	Eigenvalue	Difference	Proportion	Cumulative
1	6.64039665	3.83052724	0.6640	0.6640
2	2.80986941	2.48138197	0.2810	0.9450
3	0.32848744	0.22425307	0.0328	0.9779
4	0.10423437	0.02544652	0.0104	0.9883
5	0.07878785	0.05190333	0.0079	0.9962
6	0.02688453	0.01945873	0.0027	0.9989
7	0.00742579	0.00576423	0.0007	0.9996
8	0.00166156	0.00006616	0.0002	0.9998
9	0.00159540	0.00093841	0.0002	0.9999
10	0.00065698		0.0001	1.0000

Enclosure C – SAS output

The PRINCOMP Procedure

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10
slo500	-.130694	0.487662	-.811231	0.109990	0.033337	0.153203	0.138534	-.148265	-.092048	0.026123
slo580	-.357271	0.189657	0.126734	0.522397	-.446694	0.167612	-.333231	0.414583	0.021404	0.190834
slo750	-.365486	0.179262	0.222800	-.055600	0.184634	0.290129	-.086913	-.470507	0.648618	0.116843
slo830	-.324513	0.319584	0.183316	-.052294	0.116497	-.093938	0.275549	0.309360	0.066098	-.746875
slo910	-.318683	0.313912	0.316282	0.088061	0.374134	-.351926	0.177417	-.110976	-.450071	0.426529
int500	0.206416	0.478488	0.283326	-.464216	-.543971	0.230356	0.085188	-.164085	-.217408	0.047501
int580	0.326552	0.304564	0.048824	-.197396	0.529566	0.346538	-.298041	0.493986	0.051325	0.152453
int750	0.297465	0.375375	-.049354	0.030567	-.099413	-.715831	-.333998	-.036931	0.363050	-.014743
int830	0.366532	0.142771	0.208174	0.558954	0.152600	0.208461	-.207220	-.422963	-.262308	-.363482
int910	0.377829	0.102013	0.118889	0.365509	-.046913	0.050611	0.710090	0.159796	0.336035	0.227297



Enclosure C – SAS output

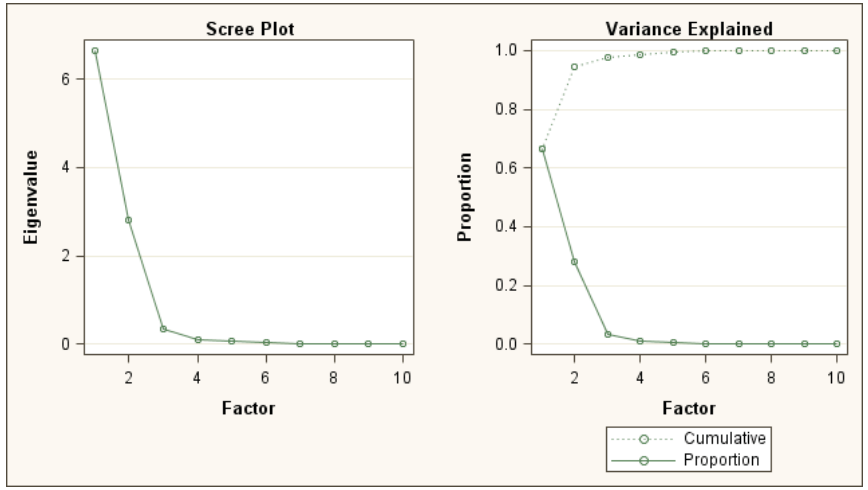
Factor Analysis on the Lard Data

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 10 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	6.64039665	3.83052724	0.6640	0.6640
2	2.80986941	2.48138197	0.2810	0.9450
3	0.32848744	0.22425307	0.0328	0.9779
4	0.10423437	0.02544652	0.0104	0.9883
5	0.07878785	0.05190333	0.0079	0.9962
6	0.02688453	0.01945873	0.0027	0.9989
7	0.00742579	0.00576423	0.0007	0.9996
8	0.00166156	0.00006616	0.0002	0.9998
9	0.00159540	0.00093841	0.0002	0.9999
10	0.00065698		0.0001	1.0000



Enclosure C – SAS output

The FACTOR Procedure

Initial Factor Method: Principal Components

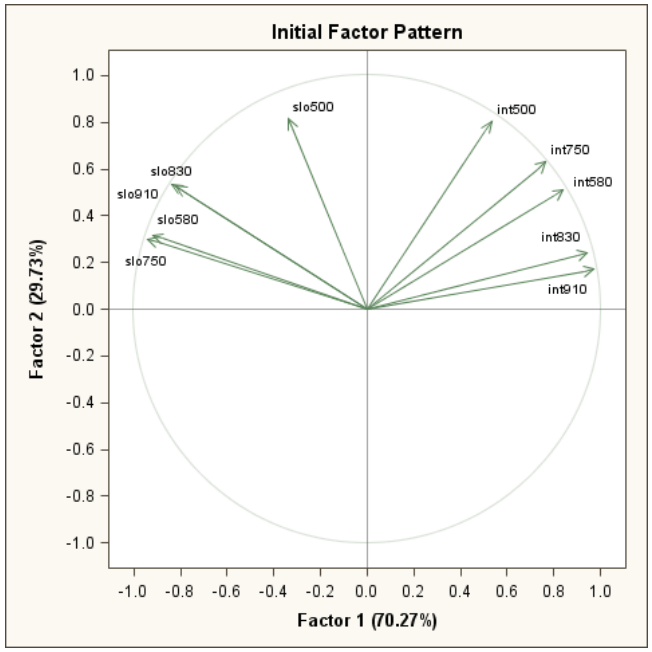
Factor Pattern		
	Factor1	Factor2
slo500	-0.33679	0.81745
slo580	-0.92065	0.31792
slo750	-0.94182	0.30049
slo830	-0.83624	0.53571
slo910	-0.82121	0.52620
int500	0.53191	0.80207
int580	0.84149	0.51053
int750	0.76654	0.62923
int830	0.94451	0.23932
int910	0.97363	0.17100

Variance Explained by Each Factor	
Factor1	Factor2
6.6403967	2.8098694

Final Communality Estimates: Total = 9.450266									
slo500	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
0.781651	0.948667	0.977319	0.986273	0.951277	0.926253	0.9687470	0.983507	0.949383	0.977189

The FACTOR Procedure

Initial Factor Method: Principal Components



The FACTOR Procedure

Rotation Method: Varimax

Orthogonal Transformation Matrix		
	1	2
1	-0.72894	0.68458
2	0.68458	0.72894

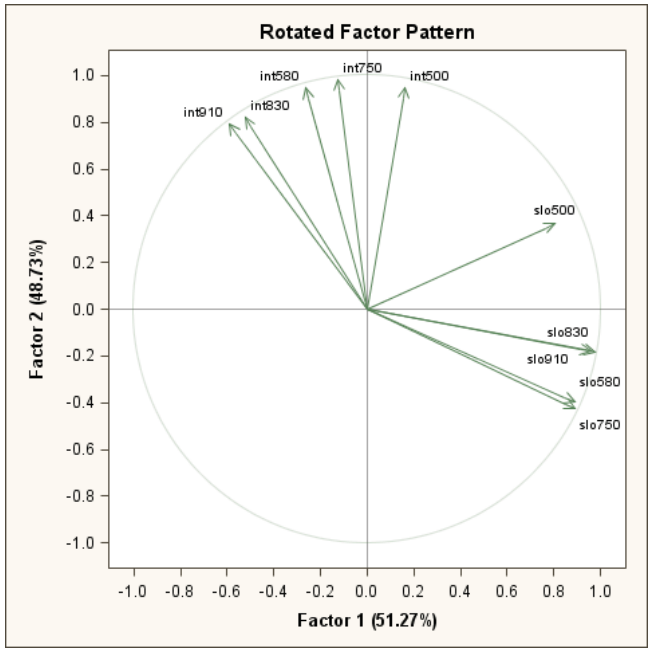
Rotated Factor Pattern		
	Factor1	Factor2
slo500	0.80511	0.36531
slo580	0.88873	-0.39852
slo750	0.89224	-0.42571
slo830	0.97630	-0.18197
slo910	0.95884	-0.17862
int500	0.16135	0.94880
int580	-0.26389	0.94821
int750	-0.12800	0.98342
int830	-0.52466	0.82105
int910	-0.59265	0.79118

Variance Explained by Each Factor	
Factor1	Factor2
4.8452152	4.6050509

Final Communality Estimates: Total = 9.450266									
slo500	slo580	slo750	slo830	slo910	int500	int580	int750	int830	int910
0.781651	0.948667	0.977319	0.986273	0.951277	0.926253	0.968747	0.983507	0.949383	0.977189

The FACTOR Procedure

Rotation Method: Varimax



Canonical Analysis on the Lard Data

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	0.999507	0.999451	0.000329	0.999013	1012.5301	1010.8801	0.9984	0.9984
2	0.789072	.	0.125789	0.622634	1.6500		0.0016	1.0000

	Test of H0: The canonical correlations in the current row and all that follow are zero				
	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	0.00037233	152.47	4	12	<.0001
2	0.37736560	11.55	1	7	0.0115

Raw Canonical Coefficients for the VAR Variables		
	V1	V2
slo500	-0.085547329	-4.977743368
slo910	-1.351109859	2.2629395224

Raw Canonical Coefficients for the WITH Variables		
	W1	W2
int500	-1.081677911	-5.795879442
int910	1.7643410493	2.9191594038

Standardized Canonical Coefficients for the VAR Variables		
	V1	V2
slo500	-0.0326	-1.8961
slo910	-0.9722	1.6283

Standardized Canonical Coefficients for the WITH Variables		
	W1	W2
int500	-0.5220	-2.7969
int910	1.4717	2.4350

Enclosure C – SAS output

The CANCORR Procedure
Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables		
	V1	V2
slo500	-0.8586	-0.5126
slo910	-0.9999	0.0172

Correlations Between the WITH Variables and Their Canonical Variables		
	W1	W2
int500	0.8558	-0.5173
int910	0.9830	-0.1835

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables		
	W1	W2
slo500	-0.8582	-0.4045
slo910	-0.9994	0.0136

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables		
	V1	V2
int500	0.8554	-0.4082
int910	0.9825	-0.1448