Written examination, date: 9th of December 2019
Course name: Multivariate Statistics
Course number: 02409
Aids allowed: All
Exam duration: 4 hours
Weighting: The questions are given equal weight

This exam is answered by:

_______________________________ _______________________________ _____________
(name) (signature) (study no.)

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

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

Problem	2	3	3	3	3	3	3	3	3	4
Question	2.3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	4.1
Answer										

Problem	4	4	5	5	5	6	6	6	6	6
Question	4.2	4.3	5.1	5.2	5.3	6.1	6.2	6.3	6.4	6.5
Answer										

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

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

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

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

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

Problem 1.

You are encouraged to use statistical software in this problem.

We consider the following model

$$\begin{bmatrix} Y_1 & Z_1 & V_1 & W_1 \\ Y_2 & Z_2 & V_2 & W_2 \\ Y_3 & Z_3 & V_3 & W_3 \\ Y_4 & Z_4 & V_4 & W_4 \\ Y_5 & Z_5 & V_5 & W_5 \end{bmatrix} = \begin{bmatrix} 1 & -2 & -4 \\ 1 & -1 & -1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} \alpha_y & \alpha_z & \alpha_v & \alpha_w \\ \beta_y & \beta_z & \beta_v & \beta_w \\ \gamma_y & \gamma_z & \gamma_v & \gamma_w \end{bmatrix} + \begin{bmatrix} \delta_1 & \varepsilon_1 & \epsilon_1 & \vartheta_1 \\ \delta_2 & \varepsilon_2 & \epsilon_2 & \vartheta_2 \\ \delta_3 & \varepsilon_3 & \epsilon_3 & \vartheta_3 \\ \delta_4 & \varepsilon_4 & \epsilon_4 & \vartheta_4 \\ \delta_5 & \varepsilon_5 & \epsilon_5 & \vartheta_5 \end{bmatrix}$$

Where the error terms $[\delta_i \quad \varepsilon_i \quad \epsilon_i \quad \vartheta_i], for\ i = 1, \dots, 5$ are independent and normally distributed $N_4(\mathbf{0}, \mathbf{\Sigma})$, and where $\mathbf{\Sigma}$ is the unknown dispersion matrix.

We have obtained the following observations

$$\begin{bmatrix} 1 & 8 & 2 & 9 \\ 0 & 9 & 4 & 6 \\ 2 & 4 & 4 & 2 \\ 1 & 5 & 9 & 5 \\ 1 & 2 & 8 & 7 \end{bmatrix}$$

We can further calculate $(\boldsymbol{x}^T \boldsymbol{x})^{-1} = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 2.125 & -1.125 \\ 0 & -1.125 & 0.625 \end{bmatrix}$

Question 1.1.

The maximum likelihood estimate for the parameters $[\alpha_y \quad \beta_y \quad \gamma_y]$ are:

1 ☐ $[5.6 \quad -2.5 \quad 0.5]$

2 ☐ $[1 \quad 1 \quad -0.5]$

3 ☐ $[1 \quad 1 \quad 0.5]$

4 ☐ $[5.4 \quad 3.5 \quad -1]$

5 ☐ $[1 \quad 1.5 \quad -0.5]$

6 ☐ Don’t know.

Question 1.2.

The covariance between the maximum likelihood estimates for α_y and β_y is:

1 ☐ 0

2 ☐ -1.125

3 ☐ 1

4 ☐ -0.5

5 ☐ 2.125

6 ☐ Don't know.

Question 1.3.

The estimated covariance between the maximum likelihood estimates for β_y and γ_y is:

1 ☐ 0

2 ☐ -1.125

3 ☐ 1.59375

4 ☐ 0.46875

5 ☐ -0.84375

6 ☐ Don't know.

Question 1.4.

The estimated variance of the maximum likelihood estimate for α_y is:

1 ☐ 0

2 ☐ -1.125

3 ☐ 0.15

4 ☐ 0.46875

5 ☐ 0.38730

6 ☐ Don't know.

Question 1.5.

The observation with the lowest leverage is:

1 ☐ 1

2 ☐ 2

3 ☐ 3

4 ☐ 4

5 ☐ 5

6 ☐ Don't know.

Question 1.6.

The dependent variable with the lowest Mean Squared Error is:

1 ☐ β

2 ☐ Y

3 ☐ Z

4 ☐ V

5 ☐ W

6 ☐ Don't know.

We now test whether $[\beta_y \quad \beta_z \quad \beta_v \quad \beta_w]$ are all equal to 0 simultaneously with the following model

$$\mathrm{H}_0\text{: } \boldsymbol{A}\begin{bmatrix}\alpha_y & \alpha_z & \alpha_v & \alpha_w\\ \beta_y & \beta_z & \beta_v & \beta_w\\ \gamma_y & \gamma_z & \gamma_v & \gamma_w\end{bmatrix}\boldsymbol{B}^T = \boldsymbol{C} \quad \text{vs.} \quad \mathrm{H}_1\text{: } \boldsymbol{A}\begin{bmatrix}\alpha_y & \alpha_z & \alpha_v & \alpha_w\\ \beta_y & \beta_z & \beta_v & \beta_w\\ \gamma_y & \gamma_z & \gamma_v & \gamma_w\end{bmatrix}\boldsymbol{B}^T \neq \boldsymbol{C}$$

Question 1.7.

In the above model **A** is equal to:

1 ☐ $[0 \quad 1 \quad 0]$

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

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

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

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

6 ☐ Don't know.

Question 1.8.

The usual test-statistic for the above model has – under the null-hypothesis – the following distribution:

1 ☐ $U(1,1,3)$

2 ☐ $U(4,3,4)$

3 ☐ $U(2,3,2)$

4 ☐ $U(4,1,2)$

5 ☐ $U(1,2,2)$

6 ☐ Don't know.

Problem 2.

Enclosure A with SAS program and SAS output belongs to this problem. We consider data for the 98 municipalities (kommuner) in Denmark. We have the rates (pr. 1000 capita) of different types of library use, e.g. book loan, music loan, etc. Further, we consider the educational levels as the fraction of the population with that educational level, e.g. a H1 at 0.25 means that 25 % of the population in a given municipality has primary school has the highest education. (Source http://www.statistikbanken.dk)

We shall now investigate the relations between the use of libraries and the educational level by means of a Canonical Correlation Analysis.

We consider the following variables for library use

SAS-name	Meaning
U1	Books
U2	Serial publications
U3	Audio books
U4	Music
U5	Live images (movies)
U6	Multi media material
U7	Other material

And for educational level

SAS-name	Meaning
H1	Primary school
H2	High school
H3	Vocational school
H4	Short further educations
H5	Medium further education
H6	Bachelor level
H7	Master level
H8	Ph.D. level

Question 2.1.

The first canonical correlation describes the following fraction of the variation between V1 and W1:

1 ☐ 0.7648

2 ☐ 1.4090

3 ☐ 0.1222

4 ☐ 0.4823

5 ☐ 0.5849

6 ☐ Don’t know.

Question 2.2.

How much of the variance in U1 is explained by V1:

1 ☐ 0.7532

2 ☐ 0.5673

3 ☐ 0.5760

4 ☐ 0.7211

5 ☐ 0.00067

6 ☐ Don't know.

Question 2.3.

The first canonical variate W1 can be interpreted as:

1 ☐ Mainly primary school, but an overall average

2 ☐ An average of primary and vocational school

3 ☐ A contrast between Ph.D. level and short further education

4 ☐ A contrast between primary and vocational school against all other educations.

5 ☐ A contrast between primary, high school and vocational school against the rest

6 ☐ Don't know.

Problem 3.

Enclosure B with SAS program and SAS output belongs to this problem. As in Problem 2, we consider library data, but now for the Capital region only. We want to predict the music loans (pr. 1000 capita) pr. municipality (the U4 variable in Problem 2) based on 5 financial variables (in 1.000 DKK pr. 1.000 capita) for the libraries in each municipality

SAS-name	Meaning
F1	Salary costs
F2	Material costs
F3	Other costs
F4	Income (e.g. late fees)
F5	Netto costs

We consider two models – both with an intercept:

- M1: All variables
- M2: which is the resulting model, after we have reduced the number of explanatory variables by stepwise model selection.

Question 3.1.

The reduction in variance explained when going from model M1 to model M2 is:

1 ☐ 0.0401

2 ☐ 0.0862

3 ☐ 0.0461

4 ☐ 0.6151

5 ☐ 0.5289

6 ☐ Don’t know.

Question 3.2.

If we performed backwards elimination from M1, the first variable to be excluded is:

1 ☐ F1

2 ☐ F2

3 ☐ F3

4 ☐ F4

5 ☐ F5

6 ☐ Don’t know.

Question 3.3.

What is the usual test statstic for M1 vs M2:

1 ☐ $\frac{(137360-124382)/(5-2)}{137360/2}$

2 ☐ $\frac{(137360-124382)/(26-23)}{124382/23}$

3 ☐ $\frac{(5407.83-5283.08)/(26-23)}{39759/5}$

4 ☐ $\frac{137360/3}{124382/5}$

5 ☐ $\frac{(137360-1243\quad)/5}{137360/2}$

6 ☐ Don’t know.

Question 3.4.

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

1 ☐ F(5,26)

2 ☐ F(2,23)

3 ☐ F(5,23)

4 ☐ F(3,23)

5 ☐ F(3,26)

6 ☐ Don’t know.

We now only consider M2

Question 3.5.

The observation with the highest leverage is:

1 ☐ 2

2 ☐ 3

3 ☐ 5

4 ☐ 8

5 ☐ 25

6 ☐ Don't know.

Question 3.6.

The observation with the highest impact on the intercept is:

1 ☐ 2

2 ☐ 3

3 ☐ 5

4 ☐ 8

5 ☐ 25

6 ☐ Don't know.

Question 3.7.

What is the 95% confidence interval for the expected value of observation no. 3 (three):

1 ☐ 245.8493 ± 190.8493

2 ☐ $245.8493 \pm t(26)_{0.975} \times 22.7423$

3 ☐ $55 \pm t(26)_{0.975} \times 22.7423$

4 ☐ $55 \pm t(26)_{0.975} \times 7.1158$

5 ☐ $245.8493 \pm t(26)_{0.975} \times \sqrt{0.5750 \cdot 0.0979}$

6 ☐ Don't know.

Question 3.8.

What is the variance of the model M2, if observation no. 8 (eight) is deleted?

1 ☐ -0.1112^2

2 ☐ 5407.93

3 ☐ 5283.08

4 ☐ 5491.58

5 ☐ 5644.32

6 ☐ Don’t know.

Problem 4.

Enclosure C with SAS program and SAS output belongs to this problem. We now consider a subset of municipalities, that are either high or low crime. There might be a link between crime levels and library use. We will test if we can classify these municipalities as a high or low crime municipality, based on the 7 use of library variables (U1 – U7) described in problem 2. (Source http://www.statistikbanken.dk)

Question 4.1.

The number of misclassifications based on resubstitutions when going from Linear Discriminant Analysis with all variables to Quadratic Discriminant Analysis with all variables is reduced with:

1 ☐ 0

2 ☐ 1

3 ☐ 3

4 ☐ 4

5 ☐ 11

6 ☐ Don't know.

Question 4.2.

When using all variables, the Hotelling's T^2 for the hypothesis of same mean in the two groups is:

1 ☐ 3.62001

2 ☐ 27.6437

3 ☐ 111.22061

4 ☐ 61.17070

5 ☐ 406.5760

6 ☐ Don't know.

Question 4.3.

We now test if U1, U2, U3, U5, and U8 contribute to the discrimination between the groups using Linear Discriminant Analysis. The usual test statistic is given by:

1 ☐ $\frac{21+12-8-1}{8-3} \cdot \frac{61.17070-1.52415}{\frac{(21+1\quad)(21+12-2)}{21 \cdot 12}+1.52415}$

2 ☐ 61.17070

3 ☐ $\frac{21+12-8-1}{8-3} \cdot \frac{3.62001-1.52415}{\frac{(21+12)(21+12\quad\;)}{21 \cdot 12}+1.52415}$

4 ☐ $\frac{21 \cdot 12}{21+12} \cdot (3.62001 - 1.52415)$

5 ☐ $\frac{21 \cdot 12}{21+12} \cdot (61.1707 - 1.52415)$

6 ☐ Don’t know.

Problem 5.

Enclosure D with SAS program and SAS output belongs to this problem. We now consider the library contents among the 98 municipalities in Denmark given in items pr. 1000 capita. We will analyse if there are any patterns or trends by means of a factor analysis. (Source http://www.statistikbanken.dk)

We consider the following variables for library contents

SAS-name	Meaning
B1	Books
B2	Audio books
B3	Music
B4	Live images (movies)
B5	Multi media material
B6	Other material
B7	Electronic resources

Question 5.1.

Considering the screeplot, we should retain how many factors in the model:

1 ☐ 6 factors

2 ☐ 5 factors

3 ☐ 4 factors

4 ☐ 2 factors

5 ☐ 1 factor

6 ☐ Don't know.

Question 5.2.

Irrespective of the answer to Question 5.1., we choose to include 3 factors in the model.
Which of the original variables has the least amount of its variance described by the Factor Analysis:

1 ☐ B1

2 ☐ B2

3 ☐ B5

4 ☐ B6

5 ☐ B7

6 ☐ Don't know.

Question 5.3.

Looking at the score plots, we can conclude:

1 ☐

Læsø and Samsø have a smaller collection of books, Other Material and Electronic Resources than average.
København(vn) has a larger and Gentofte a smaller Music collection.
Odense and Herlev have a smaller Audio Books, Live Images (Movies), Multi Media Material collection than average, while Favrskov has a larger collection.

2 ☐

Læsø and Samsø have a larger collection of books, Other Material and Electronic Resources than average.
København(vn) has a smaller and Gentofte a larger Audio Books, Live Images (Movies), Multi Media Material collection.
Odense and Herlev have a larger Music collection than average, while Favrskov has a smaller collection.

3 ☐

Læsø and Samsø have a larger collection of books, Other Material and Electronic Resources than average.
København(vn) has a smaller and Gentofte a larger Music collection.
Odense and Herlev have a larger Audio Books, Live Images (Movies), Multi Media Material collection than average, while Favrskov has a smaller collection.

4 ☐

Læsø and Samsø have a larger collection of Audio Books, Live Images (Movies), Multi Media Material than average.
København(vn) has a smaller and Gentofte a larger collection of Books, Other Material and Electronic Resources.
Odense and Herlev have a larger Music collection than average, while Favrskov has a smaller collection.

5 ☐

Læsø and Samsø have a larger Music collection than average.
København(vn) has a smaller and Gentofte a larger collection of Audio Books, Live Images (Movies), Multi Media Material.
Odense and Herlev have a larger Books, Other Material and Electronic Resources collection than average, while Favrskov has a smaller collection.

6 ☐ Don’t know.

Problem 6.

We consider a random variable

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

with mean value and dispersion matrix respectively equal to

$$\mu = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

Furthermore we consider the random variables

$$S = X - Y$$
$$T = Y - Z$$

Question 6.1.

The mean value of the two-dimensional random variable $\begin{bmatrix} S \\ T \end{bmatrix}$ is:

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

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

3 ☐ $\begin{bmatrix} 1 + \rho \\ 1 - \varrho \end{bmatrix}$

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

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

6 ☐ Don’t know

Question 6.2.

The dispersion matrix for the two-dimensional random variable $\begin{bmatrix} S \\ T \end{bmatrix}$ is:

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

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

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

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

5 ☐ $\begin{bmatrix} \rho & \rho^2 \\ \rho^2 & 2\rho^2 \end{bmatrix}$

6 ☐ Don’t know

Question 6.3.

The covariance between X and S is:

1 ☐ 0

2 ☐ 1

3 ☐ ρ^2

4 ☐ $1+\rho$

5 ☐ $1-\rho$

6 ☐ Don’t know

Question 6.4.

The conditional mean $\mathrm{E}(X|Y)$ is:

1 ☐ $\rho(y-2)+1$

2 ☐ $-\rho(y-1)-1$

3 ☐ $\rho(y-1)+1$

4 ☐ $(y-2)+1$

5 ☐ $-\rho$

6 ☐ Don't know

Question 6.5.

The conditional dispersion matrix $\mathrm{D}\left(\begin{bmatrix} X \\ Z \end{bmatrix} \middle| Y\right)$ is:

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

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

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

4 ☐ $\begin{bmatrix} \rho^2 & 0 \\ 0 & \rho^2 \end{bmatrix}$

5 ☐ $\begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 1-\rho^2 \end{bmatrix}$

6 ☐ Don't know

LAST PAGE:
END OF THE EXAM SET

SAS-PROGRAM

```
proc cancorr data=bibudd;
var U1-U7;
with H1-H8;
run;
```

Some SAS-outputs have been omitted or truncated

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)				Test of H0: The canonical correlations in the current row and all that follow are zero				
					Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	0.7648	0.7188	0.0426	0.5849	1.4090	0.6740	0.4823	0.4823	0.1222	3.7700	56	441.51	<.0001
2	0.6509	0.5898	0.0591	0.4236	0.7350	0.2585	0.2516	0.7338	0.2943	2.7500	42	388.07	<.0001
3	0.5681	0.5318	0.0695	0.3227	0.4765	0.2950	0.1631	0.8969	0.5106	2.0400	30	334.00	0.0014
4	0.3919	0.3138	0.0868	0.1536	0.1815	0.0975	0.0621	0.9590	0.7538	1.2400	20	279.55	0.2182
5	0.2783	0.2110	0.0947	0.0774	0.0839	0.0487	0.0287	0.9877	0.8906	0.8400	12	225.18	0.6092
6	0.1845	0.1636	0.0991	0.0340	0.0352	0.0346	0.0121	0.9998	0.9654	0.5100	6	172.00	0.8004
7	0.0248	-0.1740	0.1025	0.0006	0.0006		0.0002	1.0000	0.9994	0.0300	2	87.00	0.9736

Multivariate Statistics and F Approximations

S=7 M=0 N=39.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
Wilks' Lambda	0.12215361	3.77	56	441.51	<.0001
Pillai's Trace	1.59691375	3.21	56	609	<.0001
Hotelling-Lawley Trace	2.92172975	4.15	56	274.26	<.0001
Roy's Greatest Root	1.40900753	15.32	8	87	<.0001

The CANCORR Procedure

Canonical Correlation Analysis

Raw Canonical Coefficients for the VAR Variables

	V1	V2	V3	V4	V5	V6	V7
U1	-0.00067	-0.00067	-0.00006	0.00052	-0.00065	0.00006	0.00016
U2	0.00670	-0.00655	-0.00257	0.00331	0.01038	0.00025	0.00163
U3	-0.00228	0.00491	0.02591	0.00792	0.00916	-0.01748	0.00135
U4	-0.00112	0.00462	-0.00910	0.00340	0.00436	-0.00446	-0.00351
U5	-0.00150	0.00076	-0.00056	-0.00672	0.00083	0.00657	-0.00408
U6	0.01206	0.01141	-0.00084	0.03235	-0.01847	0.00979	0.00743
U7	-0.01077	0.00545	0.00189	-0.01764	0.02178	0.01646	0.04834

Raw Canonical Coefficients for the WITH Variables

	W1	W2	W3	W4	W5	W6	W7
H1	-12.3599	61.1847	22.0391	-23.9819	21.8006	-5.4010	26.0295
H2	-10.8499	78.9921	25.1711	78.1380	74.0410	58.3459	-14.6767
H3	-2.1986	34.5513	44.4092	-28.3715	72.0741	24.3734	23.8578
H4	-11.9694	40.0035	116.3422	-91.7140	-158.5020	-148.9189	-33.0675
H5	-42.3082	32.9323	-23.7598	-16.9000	95.1033	-54.1121	20.6306
H6	95.5645	-74.9238	-166.8540	-220.3267	-258.4091	-162.5183	128.1785
H7	-33.9872	65.3493	70.1304	-62.2321	124.0281	91.8197	-58.3828
H8	-40.9669	-60.4224	97.8562	195.3959	-255.5424	-307.2344	733.7147

The CANCORR Procedure

Canonical Correlation Analysis

Standardized Canonical Coefficients for the VAR Variables

	V1	V2	V3	V4	V5	V6	V7
U1	-0.7211	-0.7151	-0.0612	0.5540	-0.6998	0.0632	0.1694
U2	0.5388	-0.5267	-0.2069	0.2662	0.8345	0.0201	0.1310
U3	-0.1054	0.2267	1.1966	0.3655	0.4228	-0.8071	0.0625
U4	-0.1115	0.4588	-0.9037	0.3373	0.4326	-0.4426	-0.3488
U5	-0.2745	0.1395	-0.1021	-1.2257	0.1510	1.1981	-0.7437
U6	0.3775	0.3570	-0.0263	1.0124	-0.5778	0.3064	0.2323
U7	-0.2005	0.1015	0.0353	-0.3285	0.4056	0.3065	0.9003

Standardized Canonical Coefficients for the WITH Variables

	W1	W2	W3	W4	W5	W6	W7
H1	-0.3897	1.9289	0.6948	-0.7561	0.6873	-0.1703	0.8206
H2	-0.2074	1.5098	0.4811	1.4935	1.4152	1.1152	-0.2805
H3	-0.0936	1.4712	1.8909	-1.2080	3.0689	1.0378	1.0159
H4	-0.0644	0.2153	0.6262	-0.4936	-0.8531	-0.8015	-0.1780
H5	-0.5363	0.4175	-0.3012	-0.2142	1.2056	-0.6859	0.2615
H6	0.7460	-0.5848	-1.3024	-1.7198	-2.0171	-1.2686	1.0005
H7	-1.2442	2.3923	2.5673	-2.2782	4.5404	3.3613	-2.1373
H8	-0.1832	-0.2702	0.4376	0.8737	-1.1427	-1.3738	3.2808

Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables

	V1	V2	V3	V4	V5	V6	V7
U1	-0.7532	-0.4078	0.0131	0.4316	0.1035	0.2455	-0.0950
U2	0.2246	-0.5347	0.0256	0.3848	0.6136	0.3431	-0.1436
U3	-0.4422	0.2017	0.5650	0.3946	0.4631	0.1152	-0.2472
U4	-0.5616	0.4018	-0.3959	0.3622	0.4362	-0.0245	-0.2108
U5	-0.3860	0.1324	0.2111	0.2110	0.3682	0.6565	-0.4216
U6	0.0958	0.4143	0.1493	0.6380	0.0214	0.6237	-0.0204
U7	-0.4215	0.2987	-0.0629	-0.0146	0.3268	0.2678	0.7419

Correlations Between the WITH Variables and Their Canonical Variables

	W1	W2	W3	W4	W5	W6	W7
H1	0.8011	0.2581	0.1093	0.1871	0.2245	-0.0680	0.1485
H2	-0.5932	0.4694	-0.3295	-0.0183	-0.3781	0.1419	-0.1080
H3	0.8678	-0.0786	0.2586	0.0179	0.3341	-0.2469	0.0214
H4	-0.4394	0.1912	0.3919	-0.1344	-0.3453	-0.4744	-0.4129
H5	-0.7879	0.0105	-0.3570	-0.1090	0.1755	-0.4051	-0.1169
H6	-0.5541	0.2595	-0.4377	-0.3858	-0.3719	0.2241	0.0428
H7	-0.8623	-0.0016	-0.1228	-0.2225	-0.2708	0.3133	0.0311
H8	-0.8735	-0.0707	-0.0324	-0.0798	-0.2877	0.2479	0.2460

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables

	W1	W2	W3	W4	W5	W6	W7
U1	-0.5760	-0.2654	0.0074	0.1691	0.0288	0.0453	-0.0024
U2	0.1718	-0.3480	0.0145	0.1508	0.1708	0.0633	-0.0036
U3	-0.3382	0.1313	0.3209	0.1546	0.1289	0.0213	-0.0061
U4	-0.4295	0.2615	-0.2249	0.1419	0.1214	-0.0045	-0.0052
U5	-0.2952	0.0862	0.1199	0.0827	0.1025	0.1211	-0.0105
U6	0.0732	0.2696	0.0848	0.2500	0.0059	0.1151	-0.0005
U7	-0.3223	0.1944	-0.0357	-0.0057	0.0909	0.0494	0.0184

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables

	V1	V2	V3	V4	V5	V6	V7
H1	0.6126	0.1680	0.0621	0.0733	0.0625	-0.0126	0.0037
H2	-0.4536	0.3055	-0.1872	-0.0072	-0.1052	0.0262	-0.0027
H3	0.6636	-0.0512	0.1469	0.0070	0.0930	-0.0456	0.0005
H4	-0.3360	0.1245	0.2227	-0.0527	-0.0961	-0.0875	-0.0102
H5	-0.6026	0.0068	-0.2028	-0.0427	0.0488	-0.0747	-0.0029
H6	-0.4238	0.1689	-0.2487	-0.1512	-0.1035	0.0414	0.0011
H7	-0.6594	-0.0011	-0.0698	-0.0872	-0.0754	0.0578	0.0008
H8	-0.6681	-0.0460	-0.0184	-0.0313	-0.0800	0.0457	0.0061

SAS-PROGRAM

```
title "Model M1";
proc reg data=bibliotek2;
model U4 = F1-F5;
run;

title "Model M2";
proc reg data=bibliotek2;
model U4 = F1-F5 / selection=stepwise influence p;
run;
```

Some SAS-outputs have been omitted or truncated

Model M1

The REG Procedure
Model: MODEL1
Dependent Variable: U4

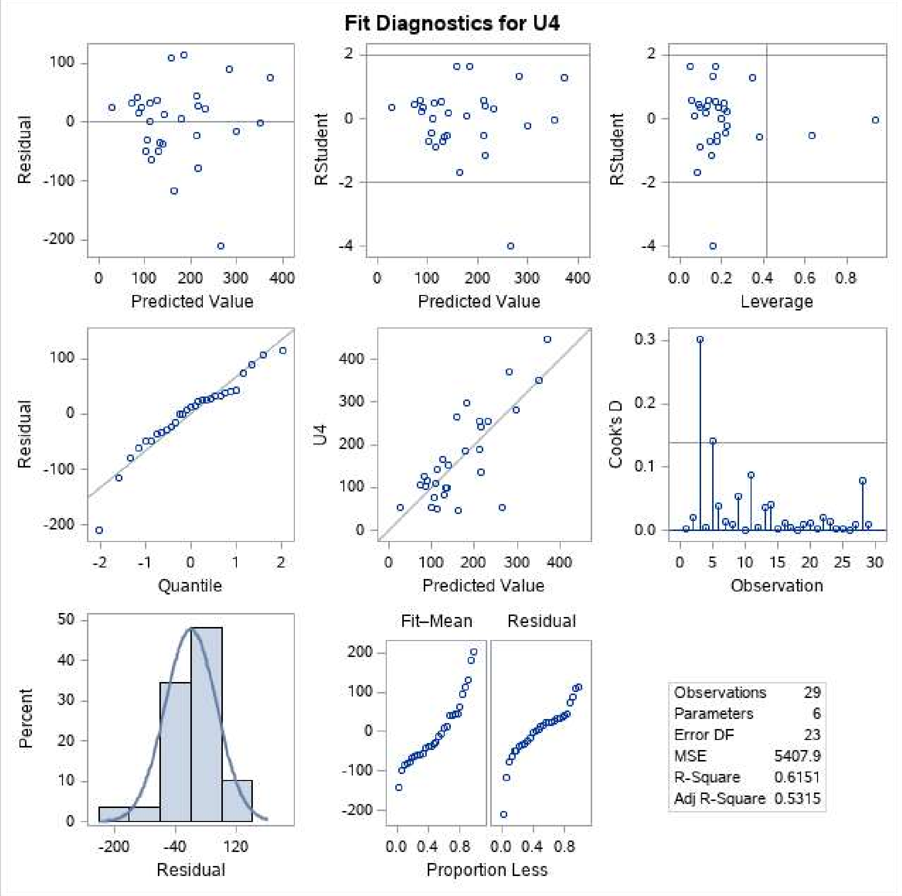
Number of Observations Read	29
Number of Observations Used	29

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	198793	39759	7.35	0.0003
Error	23	124382	5407.93108		
Corrected Total	28	323175			

Root MSE	73.53864	R-Square	0.6151
Dependent Mean	169.58621	Adj R-Sq	0.5315
Coeff Var	43.36357		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-41.74991	72.35605	-0.58	0.5695
F1	1	-22.42111	25.45826	-0.88	0.3876
F2	1	-24.16698	25.64101	-0.94	0.3557
F3	1	-23.23197	25.60095	-0.91	0.3736
F4	1	23.84917	25.64438	0.93	0.3620
F5	1	23.14720	25.53541	0.91	0.3741

Model M1
The REG Procedure
Model: MODEL1
Dependent Variable: U4



Model M2
The REG Procedure
Model: MODEL1
Dependent Variable: U4

Number of Observations Read	29
Number of Observations Used	29

Stepwise Selection: Step 1
Variable F1 Entered: R-Square = 0.5289 and C(p) = 3.1545

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	170918	170918	30.31	<.0001
Error	27	152257	5639.16555		
Corrected Total	28	323175			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-88.85878	48.97157	18566	3.29	0.0807
F1	0.68950	0.12524	170918	30.31	<.0001

Bounds on condition number: 1, 1
Stepwise Selection: Step 2
Variable F4 Entered: R-Square = 0.5750 and C(p) = 2.3998

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	185815	92907	17.59	<.0001
Error	26	137360	5283.08351		
Corrected Total	28	323175			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-81.42341	47.60658	15454	2.93	0.0991
F1	0.62516	0.12713	127747	24.18	<.0001
F4	0.51571	0.30711	14897	2.82	0.1051

Bounds on condition number: 1.0999, 4.3996
All variables left in the model are significant at the 0.1500 level.
No other variable met the 0.1500 significance level for entry into the model.

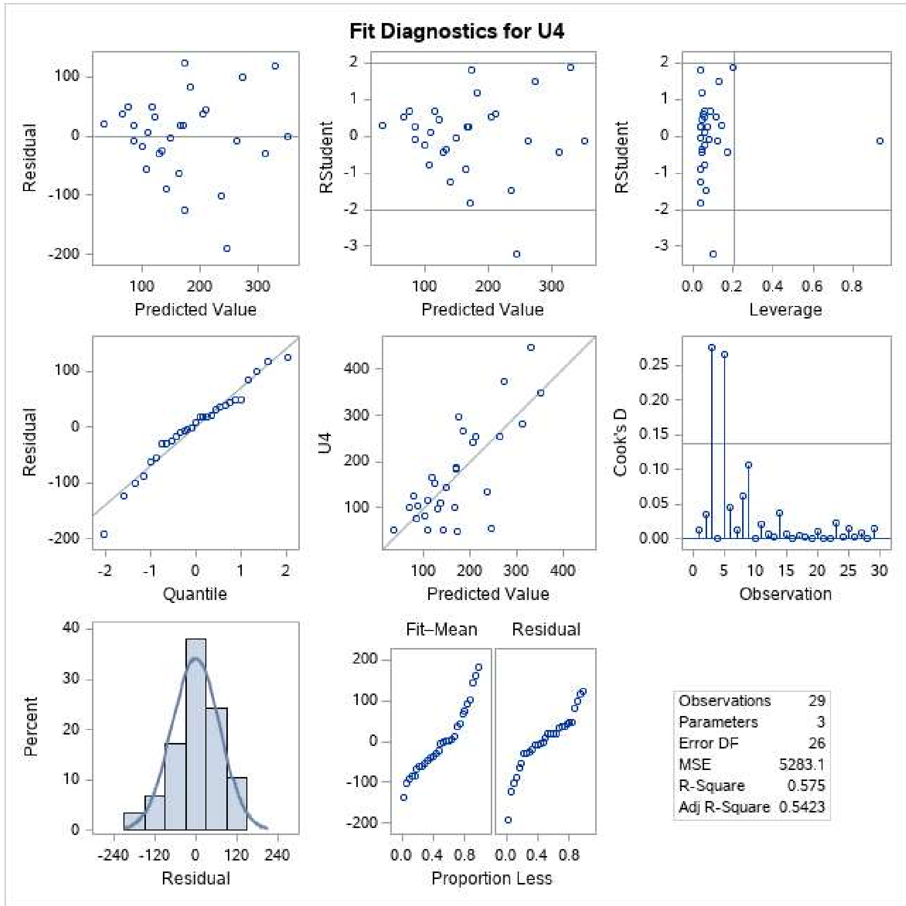
Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	F1		1	0.5289	0.5289	3.1545	30.31	<.0001
2	F4		2	0.0461	0.5750	2.3998	2.82	0.1051

Model M2
The REG Procedure
Model: MODEL1
Dependent Variable: U4

Output Statistics

Obs	Dependent Variable	Predicted Value	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS Intercept	DFBETAS F1	DFBETAS F4
1	103	66.9892	36.0108	0.5196	0.1163	1.2326	0.1885	0.1776	-0.1581	0.0469
2	297	173.7829	123.2171	1.7979	0.0346	0.8090	0.3403	0.0783	0.0182	-0.0009
3	55	245.8493	-190.8493	-3.2262	0.0979	0.4389	-1.0629	0.5743	-0.8387	0.4138
4	254	262.8382	-8.8382	-0.1276	0.1257	1.2840	-0.0484	0.0294	-0.0405	0.0192
5	447	329.3542	117.6458	1.8943	0.1972	0.9370	0.9389	-0.7064	0.8379	-0.1010
6	136	236.8447	-100.8447	-1.4649	0.0635	0.9382	-0.3813	0.0933	-0.1382	-0.1657
7	53	108.1888	-55.1888	-0.7757	0.0565	1.1101	-0.1898	-0.1312	0.0757	0.0641
8	350	352.0847	-2.0847	-0.1112	0.9360	17.5583	-0.4255	0.0080	0.0546	-0.4112
9	372	273.4345	98.5655	1.4890	0.1318	1.0041	0.5801	-0.3719	0.4962	-0.1946
10	185	167.2975	17.7025	0.2439	0.0389	1.1621	0.0491	0.0084	0.0081	-0.0162
11	266	182.8171	83.1829	1.1784	0.0427	0.9992	0.2490	-0.0065	0.0882	-0.0883
12	243	205.0262	37.9738	0.5312	0.0594	1.1563	0.1335	-0.0377	0.0786	-0.0581
13	99	129.0066	-30.0066	-0.4160	0.0464	1.1554	-0.0917	-0.0497	0.0202	0.0338
14	48	172.5641	-124.5641	-1.8210	0.0355	0.8026	-0.3492	-0.0658	-0.0424	0.0508
15	255	211.4022	43.5978	0.6074	0.0484	1.1312	0.1370	-0.0359	0.0734	-0.0230
16	144	148.6520	-4.6520	-0.0640	0.0386	1.1694	-0.0128	-0.0053	0.0011	0.0035
17	53	33.4659	19.5341	0.2850	0.1423	1.2987	0.1161	0.1125	-0.0960	-0.0010
18	110	134.6479	-24.6479	-0.3404	0.0414	1.1573	-0.0708	-0.0457	0.0271	0.0017
19	77	84.3695	-7.3695	-0.1035	0.0765	1.2164	-0.0298	-0.0267	0.0209	0.0005
20	101	164.5000	-63.5000	-0.8882	0.0404	1.0679	-0.1824	-0.0324	-0.0298	0.0694
21	116	108.9383	7.0617	0.0980	0.0555	1.1896	0.0238	0.0192	-0.0137	-0.0006
22	82	100.4049	-18.4049	-0.2567	0.0617	1.1894	-0.0658	-0.0554	0.0410	0.0023
23	52	140.7586	-88.7586	-1.2609	0.0408	0.9747	-0.2602	-0.1644	0.1026	-0.0292
24	105	85.8387	19.1613	0.2690	0.0740	1.2042	0.0760	0.0672	-0.0516	-0.0042
25	283	313.0220	-30.0220	-0.4464	0.1702	1.3236	-0.2021	0.1470	-0.1785	0.0281
26	154	122.5045	31.4955	0.4368	0.0464	1.1531	0.0964	0.0661	-0.0391	-0.0163
27	165	116.8929	48.1071	0.6743	0.0568	1.1298	0.1654	0.1334	-0.1036	0.0330
28	188	169.9855	18.0145	0.2480	0.0371	1.1596	0.0487	0.0085	0.0074	-0.0123
29	125	76.5392	48.4608	0.6909	0.0876	1.1648	0.2141	0.1979	-0.1621	0.0118

Sum of Residuals	0
Sum of Squared Residuals	137360
Predicted Residual SS (PRESS)	165793



SAS-PROGRAM

```
Title "Quadratic Discriminant Analysis";
proc discrim data=krimbib2 pool=no;
class type;
var U1-U8;
run;
Title "Linear Discriminant Analysis";
proc discrim data=krimbib2 pool=yes;
class type;
var U1-U8;
run;
Title "Linear Discriminant Analysis - Reduced Model";
proc discrim data=krimbib2 pool=yes;
class type;
var U4 U6 U7;
run;
```

Some SAS-outputs have been omitted or truncated

Quadratic Discriminant Analysis

The DISCRIM Procedure

Total Sample Size	33	DF Total	32
Variables	8	DF Within Classes	31
Classes	2	DF Between Classes	1

Number of Observations Read	33
Number of Observations Used	33

Class Level Information

type	Variable Name	Frequency	Weight	Proportion	Prior Probability
HighCrim	HighCrim	21	21.0000	0.636364	0.500000
LowCrime	LowCrime	12	12.0000	0.363636	0.500000

Within Covariance Matrix Information

type	Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
HighCrim	8	56.25449
LowCrime	8	53.24209

Quadratic Discriminant Analysis

The DISCRIM Procedure

Generalized Squared Distance to type

From type	HighCrim	LowCrime
HighCrim	56.25449	111.22061
LowCrime	61.17070	53.24209

Quadratic Discriminant Analysis

The DISCRIM Procedure

Classification Summary for Calibration Data: WORK.KRIMBIB2

Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into type

From type	HighCrim	LowCrime	Total
HighCrim	20 95.24	1 4.76	21 100.00
LowCrime	0 0.00	12 100.00	12 100.00
Total	20 60.61	13 39.39	33 100.00
Priors	0.5	0.5	

Error Count Estimates for type

	HighCrim	LowCrime	Total
Rate	0.0476	0.0000	0.0238
Priors	0.5000	0.5000	

Linear Discriminant Analysis
The DISCRIM Procedure

Total Sample Size	33	DF Total	32
Variables	8	DF Within Classes	31
Classes	2	DF Between Classes	1

Number of Observations Read	33
Number of Observations Used	33

Class Level Information					
type	Variable Name	Frequency	Weight	Proportion	Prior Probability
HighCrim	HighCrim	21	21.0000	0.636364	0.500000
LowCrime	LowCrime	12	12.0000	0.363636	0.500000

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
8	58.45721

Linear Discriminant Analysis
The DISCRIM Procedure

Generalized Squared Distance to type		
From type	HighCrim	LowCrime
HighCrim	0	3.62001
LowCrime	3.62001	0

Linear Discriminant Function for type		
Variable	HighCrim	LowCrime
Constant	-8.88572	-8.61921
U1	0.50100	0.06392
U2	0.51592	0.08165
U3	0.57176	0.10774
U4	0.50486	0.05290
U5	0.46841	0.03793
U6	0.57602	0.09638
U7	0.54500	0.05615
U8	-0.49884	-0.06074

Linear Discriminant Analysis
The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.KRIMBIB2
Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into type			
From type	HighCrim	LowCrime	Total
HighCrim	17 80.95	4 19.05	21 100.00
LowCrime	1 8.33	11 91.67	12 100.00
Total	18 54.55	15 45.45	33 100.00
Priors	0.5	0.5	

Error Count Estimates for type			
	HighCrim	LowCrime	Total
Rate	0.1905	0.0833	0.1369
Priors	0.5000	0.5000	

Linear Discriminant Analysis - Reduced Model

The DISCRIM Procedure

Total Sample Size	33	DF Total	32
Variables	3	DF Within Classes	31
Classes	2	DF Between Classes	1

Number of Observations Read	33
Number of Observations Used	33

Class Level Information					
type	Variable Name	Frequency	Weight	Proportion	Prior Probability
HighCrim	HighCrim	21	21.0000	0.636364	0.500000
LowCrime	LowCrime	12	12.0000	0.363636	0.500000

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	21.15909

Linear Discriminant Analysis - Reduced Model

The DISCRIM Procedure

Generalized Squared Distance to type		
From type	HighCrim	LowCrime
HighCrim	0	1.52415
LowCrime	1.52415	0

Linear Discriminant Function for type		
Variable	HighCrim	LowCrime
Constant	-4.07166	-1.41854
U4	0.01093	0.00466
U6	0.06282	0.04511
U7	0.06171	0.02448

Linear Discriminant Analysis - Reduced Model

The DISCRIM Procedure

Classification Summary for Calibration Data: WORK.KRIMBIB2

Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into type			
From type	HighCrim	LowCrime	Total
HighCrim	14 66.67	7 33.33	21 100.00
LowCrime	4 33.33	8 66.67	12 100.00
Total	18 54.55	15 45.45	33 100.00
Priors	0.5	0.5	

Error Count Estimates for type			
	HighCrim	LowCrime	Total
Rate	0.3333	0.3333	0.3333
Priors	0.5000	0.5000	

SAS-PROGRAM

```
proc factor data=bibliotek nfactors=3 rotate=varimax
plots=all outstat=factorOUT score;
var B1 B2 B3 B4 B5 B6 B7;
run;

proc score data=bibliotek score=factorOUT out=fScore;
var B1 B2 B3 B4 B5 B6 B7;
run;

title "Rotated factor Score Plots";
proc sgscatter data=fScore;
plot factor2*factor1 /datalabel=LA;
run;
proc sgscatter data=fScore;
plot factor3*factor1 /datalabel=LA;
run;
proc sgscatter data=fScore;
plot factor3*factor2 /datalabel=LA;
run;
```

Some SAS-outputs have been omitted or truncated

The FACTOR Procedure

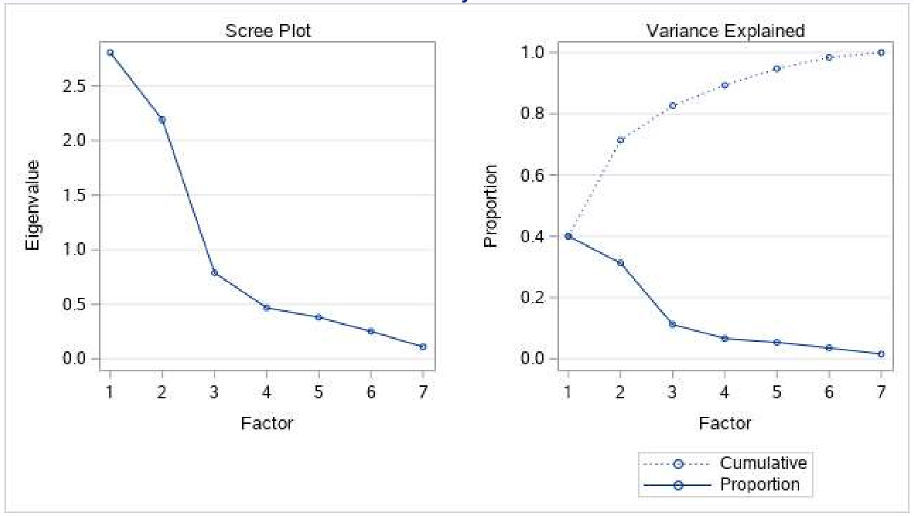
Input Data Type	Raw Data
Number of Records Read	97
Number of Records Used	97
N for Significance Tests	97

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	2.80603350	0.61618379	0.4009	0.4009
2	2.18984970	1.40081357	0.3128	0.7137
3	0.78903613	0.32001120	0.1127	0.8264
4	0.46902493	0.08768645	0.0670	0.8934
5	0.38133849	0.12830640	0.0545	0.9479
6	0.25303208	0.14134692	0.0361	0.9840
7	0.11168517		0.0160	1.0000

3 factors will be retained by the NFACTOR criterion.

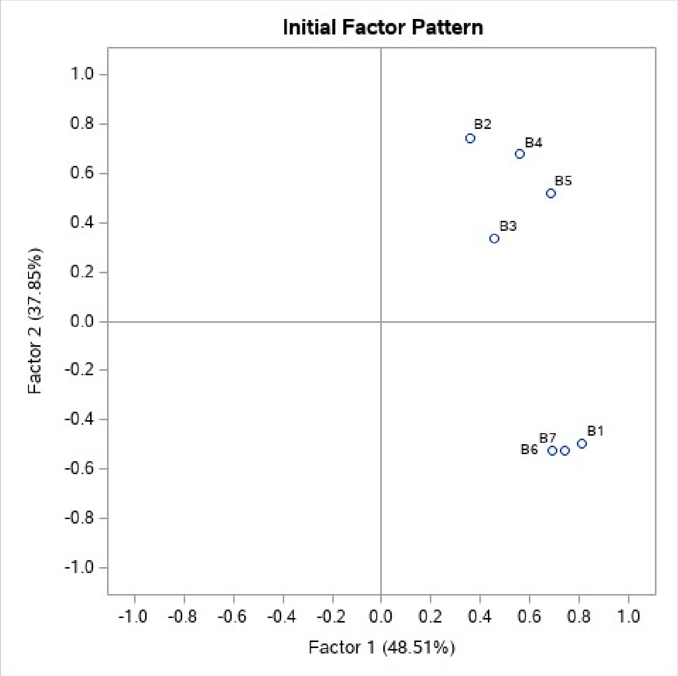


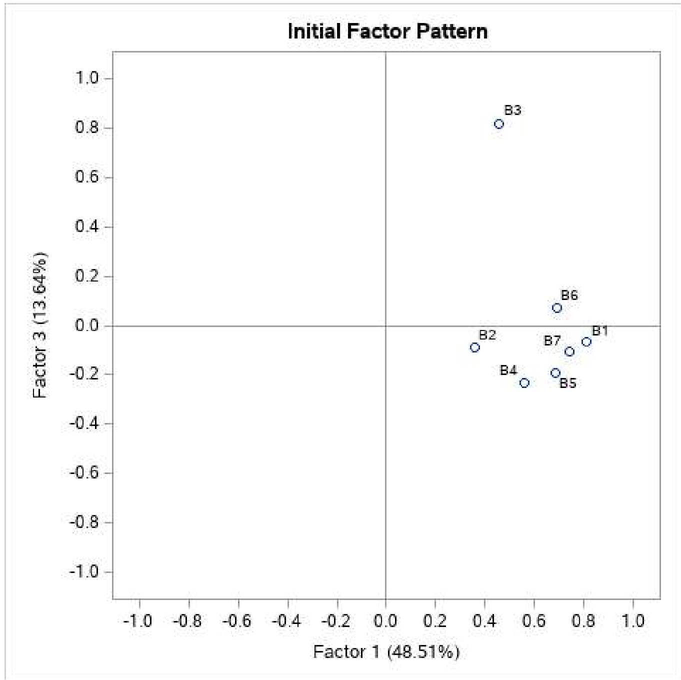
Factor Pattern			
	Factor1	Factor2	Factor3
B1	0.81346	-0.49574	-0.06424
B2	0.35836	0.74375	-0.08765
B3	0.45721	0.33596	0.81800
B4	0.56123	0.67773	-0.23340
B5	0.68265	0.51811	-0.19331
B6	0.69030	-0.52392	0.07276
B7	0.74118	-0.52519	-0.10470

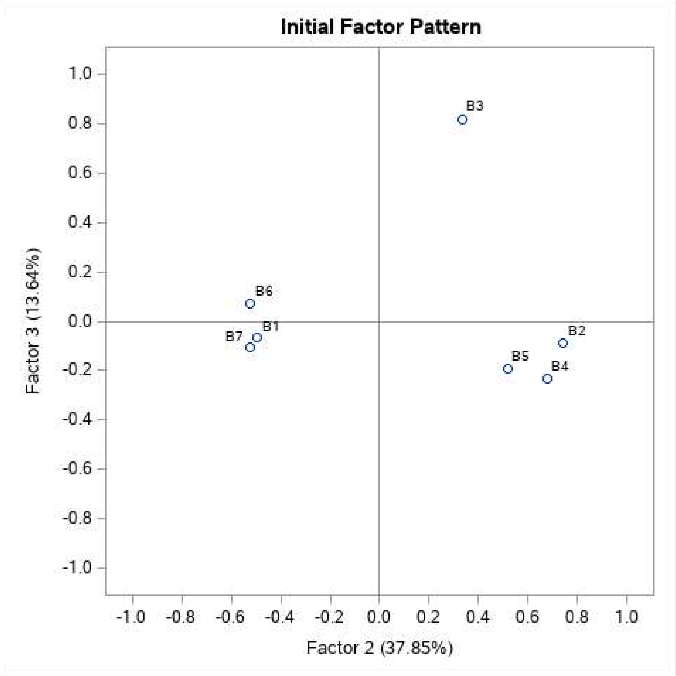
Variance Explained by Each Factor		
Factor1	Factor2	Factor3
2.8060335	2.1898497	0.7890361

Final Communality Estimates: Total = 5.784919						
B1	B2	B3	B4	B5	B6	B7
0.91159983	0.68926825	0.99103354	0.82877011	0.77181742	0.75630004	0.83613014

The FACTOR Procedure
Initial Factor Method: Principal Components







The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2	3
1	0.78908	0.55643	0.26027
2	-0.60986	0.76040	0.22328
3	-0.07367	-0.33492	0.93936

Rotated Factor Pattern

	Factor1	Factor2	Factor3
B1	0.94895	0.09719	0.04069
B2	-0.16436	0.79431	0.17701
B3	0.09562	0.23591	0.96241
B4	0.04673	0.90580	0.07815
B5	0.23693	0.83856	0.11177
B6	0.85886	-0.03865	0.13104
B7	0.91285	0.04813	-0.02271

Variance Explained by Each Factor

Factor1	Factor2	Factor3
2.5659076	2.2235022	0.9955096

Final Communality Estimates: Total = 5.784919

B1	B2	B3	B4	B5	B6	B7
0.91159983	0.68926825	0.99103354	0.82877011	0.77181742	0.75630004	0.83613014

The FACTOR Procedure
Rotation Method: Varimax
Scoring Coefficients Estimated by Regression

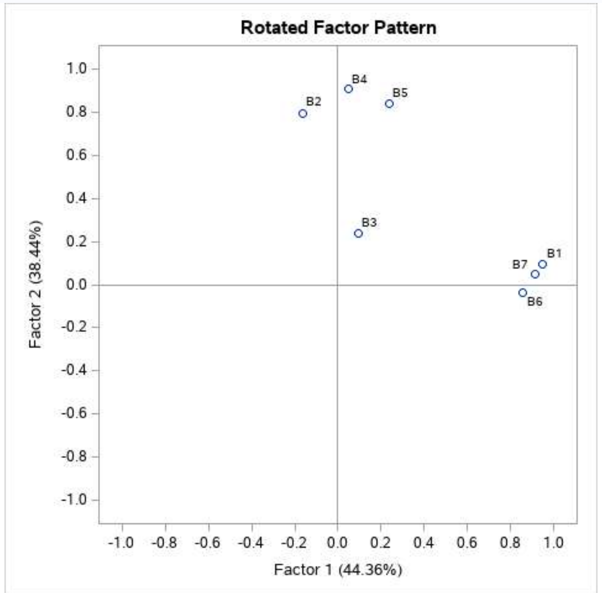
Squared Multiple Correlations of the Variables with Each Factor

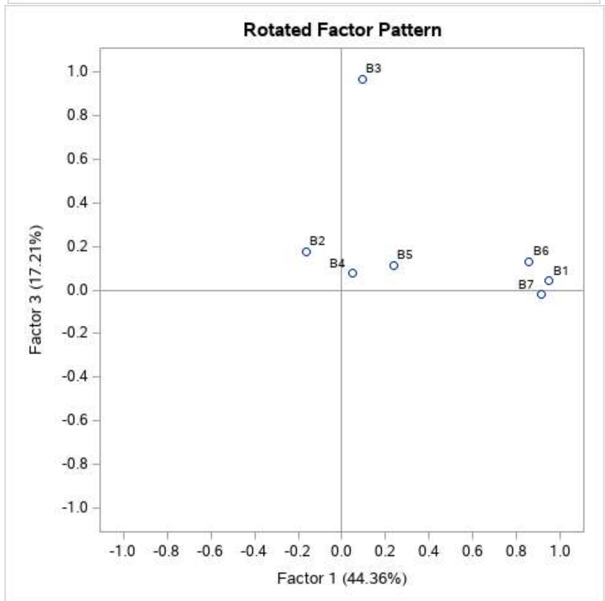
Factor1	Factor2	Factor3
1.0000000	1.0000000	1.0000000

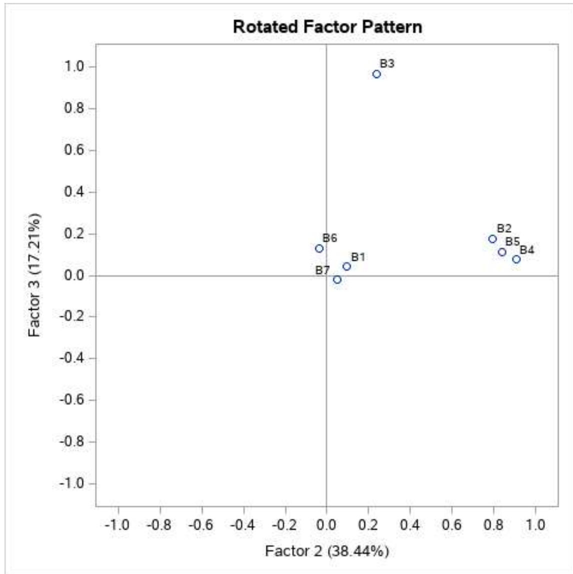
Standardized Scoring Coefficients

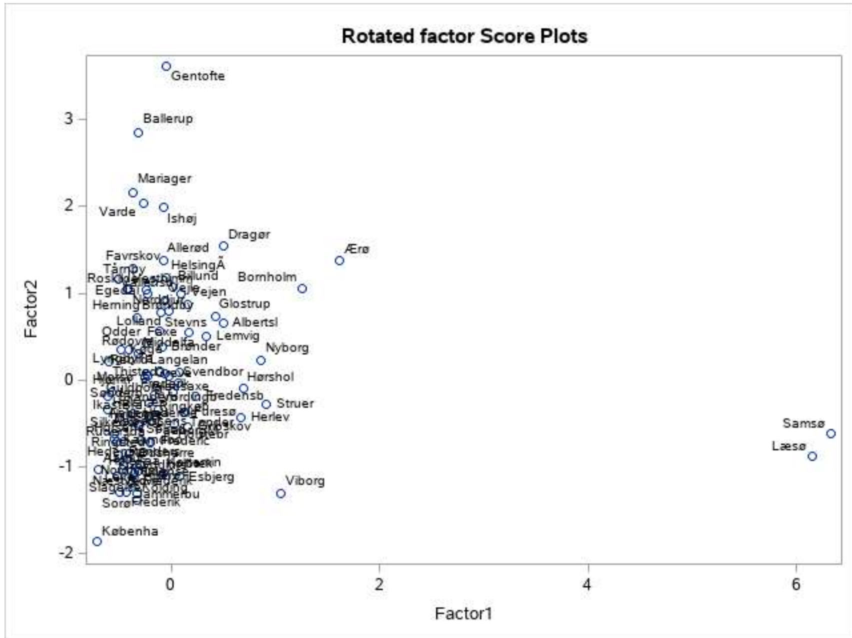
	Factor1	Factor2	Factor3
B1	0.37281	0.01644	-0.05157
B2	-0.09817	0.36652	0.00473
B3	-0.04137	-0.13989	1.05051
B4	-0.00913	0.44570	-0.15671
B5	0.06572	0.39733	-0.11400
B6	0.33323	-0.07592	0.09723
B7	0.36446	0.00905	-0.10945

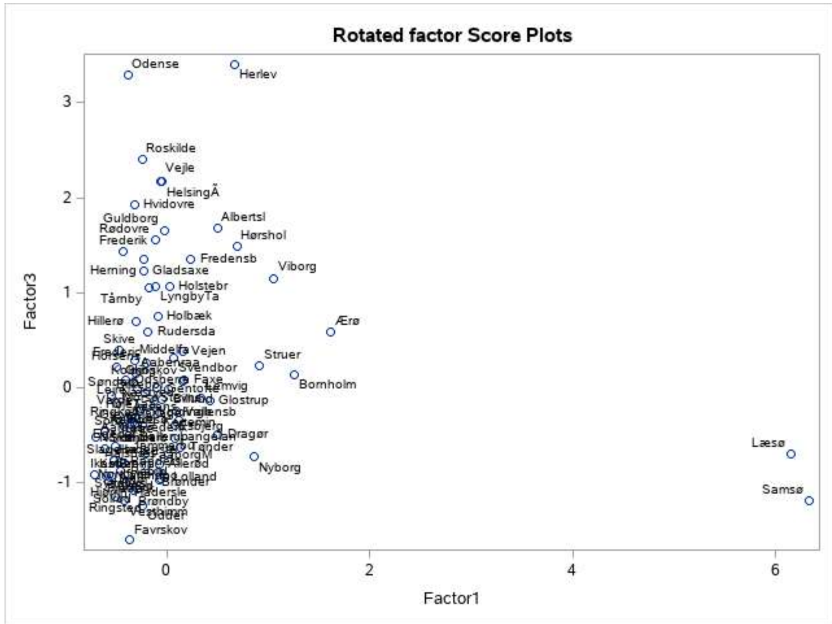
The FACTOR Procedure
Rotation Method: Varimax

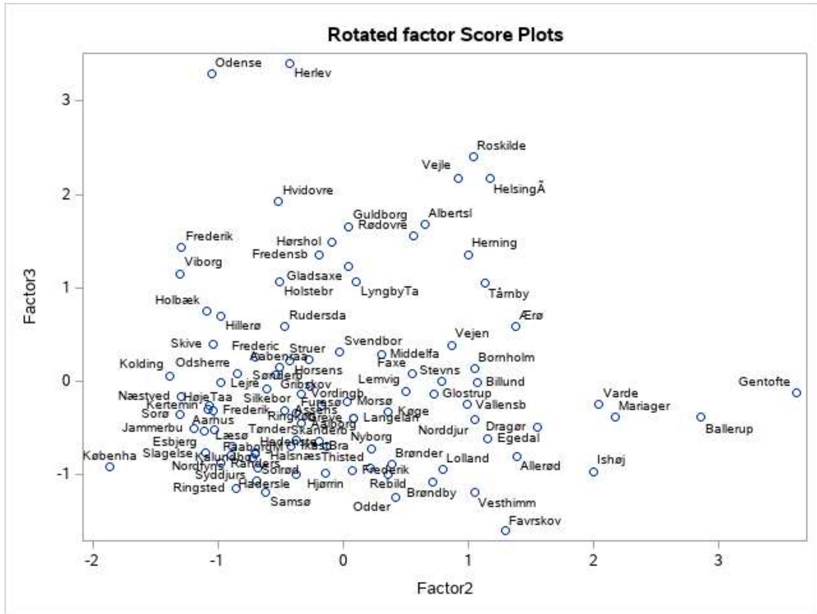









Rotated factor Score Plots
Factor3
Factor1
Odense
Herlev
Roskilde
Vejle
HelsingÃ
Hvidovre
Guldborg
Rødovre
Frederik
Albertsl
Hørshol
Fredensb
Viborg
Herning
Gladsaxe
Holstebr
Tårnby
LyngbyTa
Holbæk
Hillerø
Rudersda
Ærø
Skive
Struer
Svendbor
Bornholm
Glostrup
Dragør
Nyborg
Læsø
Samsø
Favrskov


Rotated factor Score Plots
Factor3
Factor2
Odense
Herlev
Roskilde
Vejle
HelsingÃ
Hvidovre
Guldborg
Albertsl
Rødovre
Frederik
Hørshol
Herning
Fredensb
Viborg
Gladsaxe
Holstebr
LyngbyTa
Tårnby
Holbæk
Hillerø
Rudersda
Ærø
Vejen
Skive
Svendbor
Middelfa
Bornholm
Kolding
Odsherre
Stevns
Lemvig
Billund
Gentofte
Varde
Mariager
Ballerup
Næstved
Glostrup
Vallensb
Køge
Langelan
Norddjur
Dragør
Egedal
Nyborg
Esbjerg
Jammerbu
Sorø
Aarhus
København
Slagelse
Lolland
Allerød
Ishøj
Ringsted
Hjørrin
Rebild
Brøndby
Vesthimm
Samsø
Odder
Favrskov