Written examination, date: 11. December 2017
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 7 problems. The answers to the 30 questions must be written into the table below.

Problem	1	1	1	1	2	2	2	2	2	3
Question	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.5	3.1
Answer										

Problem	3	3	4	4	4	5	5	5	5	5
Question	3.2	3.3	4.1	4.2	4.3	5.1	5.2	5.3	5.4	5.5
Answer										

Problem	5	5	6	6	6	7	7	7	7	7
Question	5.6	5.7	6.1	6.2	6.3	7.1	7.2	7.3	7.4	7.5
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.

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. We are interested in the commuting patterns across the Danish municipalities. For each of 8 distance intervals and 8 job descriptors, the proportion of citizens in the municipality with that job description and commuting distance was recorded. Data have been pulled from https://www.statistikbanken.dk.

Municipality name is the SAS-variable: city

Distances [km]	SAS-Variable	Job-descriptions	SAS-name
0	0	Independent contractors	independent
0-5	2.5	Spouces working together	spouces
5-10	5	Top leaders	topleaders
10-20	10	High salary	salaryHigh
20-30	20	Medium salary	salaryMedium
30-40	30	Low salary	salaryLow
40-50	40	Other salary	salaryOther
50+	50	Unknown salary	salaryUnknown

We start by investigating whether the proportion of commuters are the same with respect to both municipality and distance.

Question 1.1.

The usual test-statistic for no city (municipality) effect is:

1 ☐ 1.57639735

2 ☐ 3.60827056

3 ☐ 1.99721058

4 ☐ 0.07372765

5 ☐ 2.33

6 ☐ Don’t know.

 The problem continues on the next page

Question 1.2.

The usual test-statistic for no city (municipality) effect has – under the null-hypothesis – the following distribution:

1 ☐ $\mathrm{U}(8, 98, 686)$

2 ☐ $\mathrm{U}(8, 7, 686)$

3 ☐ $\mathrm{U}(8, 98, 693)$

4 ☐ $\mathrm{U}(8, 7, 787)$

5 ☐ $\mathrm{U}(8, 44.5, 338.5)$

6 ☐ Don't know.

Question 1.3.

The usual test-statistic for no distance effect, is distributed $U(p, q, r)$ under the null-hypothesis. The F-approximation is:

1 ☐ Exact since $p = 2$

2 ☐ Exact since $p - q = 1$

3 ☐ Approximate since r is larger than 2

4 ☐ Approximate since p is larger than 2

5 ☐ Approximate since p and q are larger than 2

6 ☐ Don't know.

Question 1.4.

The usual univariate test-statistic for city effect on the variable 'salaryLow' is:

1 ☐ 0.02159

2 ☐ 0.321032

3 ☐ $\frac{1.36216/98}{0.321032/686}$

4 ☐ $\frac{0.02159/98}{0.321032/686}$

5 ☐ $\frac{\det(Error\ SSCP\ Matrix)}{\det(Error\ SSCP\ Matrix\ +\ Type\ III\ SSCP\ Matrix\ for\ city)}$

6 ☐ Don't know.

The problem continues on the next page

Problem 2.

Enclosure B with SAS program and SAS output belongs to this problem. We still consider the data from problem 1. We now perform a factor analysis on the 'Job'-variables with 3 factors and rotate using the VARIMAX-principle. Further, we produce the score-plots for the distance 50 km+.

Question 2.1.

The 3 factors together describe the following fraction of the variance in the data:

1 ☐ 0.72429919

2 ☐ 0.7242992

3 ☐ (4.6570564 + 2.0695673 + 0.7242992) / 7.450923

4 ☐ 0.0905

5 ☐ 0.5821 + 0.2587 + 0.0905

6 ☐ Don't know.

Question 2.2.

The usual test statistic for the last 3 eigenvalues being equal is:

1 ☐ −2978.1

2 ☐ 126.2

3 ☐ 2961.4

4 ☐ 3027.8

5 ☐ 4512.1

6 ☐ Don't know.

Question 2.3.

The variable with the highest uniqueness is:

1 ☐ spouces

2 ☐ topleaders

3 ☐ salaryHigh

4 ☐ salaryMedium

5 ☐ salaryLow

6 ☐ Don't know.

The problem continues on the next page

Question 2.4.

What fraction of the total variance is explained by the rotated factor 1:

1 ☐ $2.7503238/8$

2 ☐ $4.6570564/8$

3 ☐ $2.7503238\ / (2.7503238 + 2.6559237 + 2.0446753)$

4 ☐ $4.6570564 / (4.6570564 + 2.0695673 + 0.7242992)$

5 ☐ 2.7503238

6 ☐ Don’t know.

Question 2.5.

We consider the score plots and look at the scores giving:

A: *The proportion of high- and medium-salaried as well as of topleaders*
B: *The proportion of independent and of spouces*

When comparing these scores, we will list them from highest to lowest, i.e A(x1,x2,x3), means that city x1 has the highest A and city x3 the lowest A.

Which of the following combinations is true:

1 ☐ A(Ikast-Brande, Albertslund, Gribskov), B(Ikast-Brande, Albertslund, Gribskov)

2 ☐ A(Ikast-Brande, Gribskov, Albertslund), B(Ikast-Brande, Gribskov, Albertslund)

3 ☐ A(Ikast-Brande, Albertslund, Gribskov), B(Ikast-Brande, Gribskov, Albertslund)

4 ☐ A(Gribskov, Albertslund, Ikast-Brande,), B(Albertslund, Gribskov, Ikast-Brande)

5 ☐ A(Gribskov, Albertslund, Ikast-Brande,), B(Gribskov, Albertslund, Ikast-Brande)

6 ☐ Don’t know.

The problem continues on the next page

Problem 3.

Enclosure C with SAS program and SAS output belongs to this problem. *Vino Verde* is a type of wine produced in Portugal. We investigate the whether we can determine the quality of the wine – as assessed by a panel – based on the physical/chemical characteristics of the wine. The data is from:
https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Variables and SAS-name:
Quality from 1 to 10 where 10 is best: quality
The fixed acidity: fixedacidity
The volatile acidity: volatileacidity
Citric Acid: citricacid
The residual sugar content: residualsugar
Cloride content: chlorides
Free sulfur dioxide: freesulfurdioxide
Total Sulfur Dioxide: totalsulfurdioxide
The density of the wine: density
The pH: pH
The sulphate content: sulphates
The alcohol content: alcohol

We are only interested in really poor or really good wine. The lowest classed wine in the dataset is quality 3 and the best is quality 8. Further, we only consider a subset of the variables: ‘alcohol’, ‘pH’ and ‘sulphates’ and perform a discriminant analysis.

Question 3.1.

The value of Hotellings T^2 statistic for comparing quality 3 and quality 8 when using 3 variables is:

1 ☐ 32.48466

2 ☐ $\frac{0.454545}{0.545455} \cdot 32.48466$

3 ☐ 177.1891

4 ☐ 63.1567

5 ☐ 194.9080

6 ☐ Don’t know.

 The problem continues on the next page

Question 3.2.

We consider the Linear Discriminant Function for classifying between poor and good wine using 3 variables. We classify as poor if the function is larger than zero and good if it is less than zero. If we assume that the prior probability for getting a poor wine is twice that of a good wine, the discriminant function becomes:

1 ☐ $[pH\ sulphates\ alcohol]\begin{bmatrix}-43.29\\28.93\\9.43\end{bmatrix} - 20.48 > 0$

2 ☐ $[pH\ sulphates\ alcohol]\begin{bmatrix}-43.29\\28.93\\9.43\end{bmatrix} - 21.18 > 0$

3 ☐ $[pH\ sulphates\ alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} + 20.48 > 0$

4 ☐ $[pH\ sulphates\ alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} + 0.6931 > 0$

5 ☐ $[pH\ sulphates\ alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} - 21.18 > 0$

6 ☐ Don't know.

Question 3.3.

Based on the theory that students only really care about alcohol content when judging the quality of a wine, we investigate whether 'pH' and 'sulphates' contains additional information when classifying based on 'alcohol'. The F-statistic corresponding to this is:

1 ☐ $\frac{(10+12-3-1)}{2}\frac{(32.48466-4.70685)}{(10+12)(10+12-2)/(10\cdot 12)+4.7068}$

2 ☐ 5.685803

3 ☐ $32.48466 - 4.70685$

4 ☐ $\frac{(10+12-3-1)(32.48466-4.70685)}{(10+12)(10+12-2)/(10\cdot 12)+4.70685}$

5 ☐ $\frac{(10+12-3-1)(32.48466-4.70685)}{(10+12)(10+12-2)/(10\cdot 12)+32.48466}$

6 ☐ Don't know.

The problem continues on the next page

Problem 4.

We consider the random normal variable

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ X_1 \\ X_2 \end{bmatrix}, \quad D(\boldsymbol{Z}) = \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix},$$

You may use the following information:

$$\begin{aligned} \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} - x^2\boldsymbol{\Sigma}_{yy} &= \tfrac{\rho^2}{1-\rho^2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - x^2\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \\ &= \tfrac{2\rho^2}{1+\rho}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - x^2\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \\ &= \tfrac{1}{1+\rho}\begin{bmatrix} 2\rho^2 - x^2(1+\rho) & 2\rho^2 - x^2\rho(1+\rho) \\ 2\rho^2 - x^2\rho(1+\rho) & 2\rho^2 - x^2(1+\rho) \end{bmatrix} \end{aligned}$$

$$a^2 - b^2 = (a+b)(a-b)$$

For a dispersion matrix of type: $\boldsymbol{\rho}_n = \begin{bmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \rho & \cdots & 1 \end{bmatrix} = (1-\rho)\boldsymbol{I}_n + \rho\boldsymbol{1}_n\boldsymbol{1}_n'$

the determinant is given by: $\det(\boldsymbol{\rho}_n) = (1-\rho)^{n-1}[1 + (n-1)\rho]$

Question 4.1.

For $D(\boldsymbol{Z})$ to be positive-definite ρ must be in the interval:

1 ☐ $0 < \rho$

2 ☐ $-1 < \rho < \frac{1}{3}$

3 ☐ $-\frac{1}{3} < \rho < 1$

4 ☐ $-3 < \rho < 1$

5 ☐ $-1 < \rho < 1$

6 ☐ Don’t know.

The problem continues on the next page

Question 4.2.

The first squared canonical correlation between ***Y*** and ***X*** is given by:

1 ☐ $\frac{4\rho^2}{(1+\rho)^2}$

2 ☐ ρ

3 ☐ ρ^2

4 ☐ 0

5 ☐ $\frac{4\rho}{1+\rho}$

6 ☐ Don't know.

Question 4.3.

The first canonical variable V_1 with variance 1 for the ***Y*** variables is given by:

1 ☐ $V_1 = Y_1 - Y_2$

2 ☐ $V_1 = \frac{1}{\sqrt{2(1+\rho)}} Y_1 + \frac{1}{\sqrt{2(1+\rho)}} Y_2$

3 ☐ $V_1 = Y_1 + Y_2$

4 ☐ $V_1 = \frac{1}{\sqrt{2(1+\rho)}} Y_1 - \frac{1}{\sqrt{2(1+\rho)}} Y_2$

5 ☐ $V_1 = \frac{1}{\sqrt{2}} Y_1 + \frac{1}{\sqrt{2}} Y_2$

6 ☐ Don't know.

The problem continues on the next page

Problem 5.

We investigate the number of cyclones around Antarctica. The data are from "*D.A. Howarth (1983), "An Analysis of the Variability of Cyclones Around Antarctica and Their Relation to Sea-Ice Extent", Annals of the Association of American Geographers, Vol. 73, pp 519-537.*"
You are encouraged to use statistical software in this problem.

We have:

Observation	Latitude band: X_1	Season: X_2	Cyclone count: Y
1	1	1	370
2	1	2	452
3	1	3	273
4	1	4	422
5	2	1	526
6	2	2	624
7	2	3	513
8	2	4	1059
9	3	1	980
10	3	2	1200
11	3	3	995
12	3	4	1751

We now propose the following model M1:

$$Y = \mu + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 (X_1 \cdot X_2) + \epsilon$$

Where μ is the intercept and ϵ is the error term.

The maximum likelihood estimated parameters to model M1 are:

Variable	DF	Parameter Estimate
$\hat{\mu}$	1	1000.66667
$\hat{\beta}_1$	1	-339.75000
$\hat{\beta}_2$	1	124.87500
$\hat{\beta}_3$	1	-531.91667
$\hat{\beta}_4$	1	87.58333
$\hat{\beta}_5$	1	106.55000

We also propose the reduced model M2:

$$Y = \mu + \gamma_1 x_1 + \gamma_2 x_2 + \epsilon$$

Where μ is the intercept and ϵ is the error term.

 The problem continues on the next page

Question 5.1.

The fraction of variance explained by model M1 is:

1 ☐ 0.91

2 ☐ $\frac{177.29}{763.75}$

3 ☐ 0.84

4 ☐ 0.0043

5 ☐ 0.79

6 ☐ Don't know.

Question 5.2.

The first variable to remove from model M1 when performing backwards elimination using the F-value is:

1 ☐ $(X_1 \cdot X_2)$

2 ☐ X_2^2

3 ☐ X_2

4 ☐ X_1^2

5 ☐ X_1

6 ☐ Don't know.

Question 5.3.

Which of the variables exhibit the highest degree of multicolliniarity, as measured by the *Tolerance* and *Variance Inflation*:

1 ☐ $(X_1 \cdot X_2)$

2 ☐ X_2^2

3 ☐ X_2

4 ☐ X_1^2

5 ☐ X_1

6 ☐ Don't know.

The problem continues on the next page

Question 5.4.

The observation that is most influential as measured by Cooks D is:

1 ☐ 1

2 ☐ 4

3 ☐ 8

4 ☐ 11

5 ☐ 12

6 ☐ Don't know.

Question 5.5.

The observation that – when deleted – will lead to the largest change in the intercept is:

1 ☐ 1

2 ☐ 4

3 ☐ 8

4 ☐ 11

5 ☐ 12

6 ☐ Don't know.

Question 5.6.

The usual F-statistic for the hypothesis M2 vs M1 is:

1 ☐ 12.17

2 ☐ 17.20

3 ☐ $\frac{(188581-435744)/(5-2)}{188581/6}$

4 ☐ $\frac{(1912595-1665432)/(5-2)}{188581/6}$

5 ☐ $\frac{(1912595-1665432)/6}{188581/6}$

6 ☐ Don't know.

The problem continues on the next page

Question 5.7.

In the model M2 the usual F-test for all parameters except the intercept being equal to zero is:

1 ☐ Significant at the 0.0005 but not 0.0001 level.

2 ☐ Significant at the 0.001 but not 0.0005 level.

3 ☐ Significant at the 0.005 but not 0.001 level.

4 ☐ Significant at the 0.01 but not 0.005 level.

5 ☐ Significant at the 0.05 but not 0.01 level.

6 ☐ Don’t know.

The problem continues on the next page

Problem 6.

Enclosure C with SAS program and SAS output belongs to this problem. We consider the data from problem 3. We now want to investigate the variables related to acidity. The dependent variables are:

The fixed acidity: Y1
The volatile acidity: Y2
Citric Acid: Y3
The pH: Y4

And the independent variables are:

The residual sugar content: X1
Cloride content: X2
Free sulfur dioxide: X3
Total Sulfur Dioxide: X4
The density of the wine: X5
The sulphate content: X6
The alcohol content: X7

We test the influence of sulfur *X3* and *X4* by the hypothesis that all parameters related to these are zero, against all alternatives, using Theorem 5.9 in the lecture notes.

Question 6.1.

The C matrix is given by

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

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

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

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

5 ☐ $[0]$

6 ☐ Don’t know.

The problem continues on the next page

Question 6.2.

The A matrix is given by

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

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

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

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

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

6 ☐ Don’t know.

Question 6.3.

The usual test statistic for the hypothesis that the parameters are zero, follow this distribution under the null hypothesis:

1 ☐ $\mathrm{U}(4, 2, 1572)$

2 ☐ $\mathrm{U}(7, 4, 1565)$

3 ☐ $\mathrm{U}(4, 2, 1768)$

4 ☐ $\mathrm{U}(4, 2, 1565)$

5 ☐ $\mathrm{U}(2, 4, 1565)$

6 ☐ Don’t know.

The problem continues on the next page

Problem 7.

We consider the two bivariate normal random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$.
Assume $E(\boldsymbol{X}) = \boldsymbol{\mu}_x$, $E(\boldsymbol{Y}) = \boldsymbol{\mu_y}$

$$D\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix} = D\begin{pmatrix}\boldsymbol{X_1}\\ \boldsymbol{X_2}\\ \boldsymbol{Y_1}\\ \boldsymbol{Y_2}\end{pmatrix} = \begin{bmatrix}1 & \rho & \rho & \rho\\ \rho & 1 & \rho & \rho\\ \rho & \rho & 1 & \rho\\ \rho & \rho & \rho & 1\end{bmatrix}$$

Furthermore, consider the constant matrices $\boldsymbol{A}$, $\boldsymbol{B}$ with appropriate dimensions

Question 7.1.

$E(\boldsymbol{AX} - \boldsymbol{BY} + \boldsymbol{AX})$ equals

1 ☐ $2\boldsymbol{A\mu_x} - \boldsymbol{B\mu_y}$

2 ☐ $\boldsymbol{A\mu_x} - \boldsymbol{B\mu_y}$

3 ☐ $2\boldsymbol{A\mu_x} + \boldsymbol{B\mu_y}$

4 ☐ $\boldsymbol{A\mu_x A} - \boldsymbol{B\mu_y}$

5 ☐ $\boldsymbol{A\mu_x} - \boldsymbol{B\mu_y} + \boldsymbol{A}$

6 ☐ Don’t know.

Question 7.2.

$D(\boldsymbol{AX} - \boldsymbol{BY})$ equals

1 ☐ $\boldsymbol{A}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{1}\\ \boldsymbol{1} & \boldsymbol{1}\end{bmatrix}\boldsymbol{B}' + \boldsymbol{B}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{1}\\ \boldsymbol{1} & \boldsymbol{1}\end{bmatrix}\boldsymbol{B}' - \boldsymbol{A}\begin{bmatrix}\boldsymbol{\rho} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{\rho}\end{bmatrix}\boldsymbol{B}' - \boldsymbol{B}\begin{bmatrix}\boldsymbol{\rho} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{\rho}\end{bmatrix}\boldsymbol{A}'$

2 ☐ $\boldsymbol{A}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{A}' + \boldsymbol{B}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{B}'$

3 ☐ $\boldsymbol{A}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{A}' + \boldsymbol{B}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{B}' - \boldsymbol{2A}\begin{bmatrix}\boldsymbol{\rho} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{\rho}\end{bmatrix}\boldsymbol{B}'$

4 ☐ $\boldsymbol{A}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{A}' + \boldsymbol{B}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{B}' + \boldsymbol{A}\begin{bmatrix}\boldsymbol{\rho} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{\rho}\end{bmatrix}\boldsymbol{B}' + \boldsymbol{B}\begin{bmatrix}\boldsymbol{\rho} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{\rho}\end{bmatrix}\boldsymbol{A}'$

5 ☐ $\boldsymbol{A}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{A}' + \boldsymbol{B}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{1}\end{bmatrix}\boldsymbol{B}' - \boldsymbol{A}\begin{bmatrix}\boldsymbol{\rho} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{\rho}\end{bmatrix}\boldsymbol{B}' - \boldsymbol{B}\begin{bmatrix}\boldsymbol{\rho} & \boldsymbol{\rho}\\ \boldsymbol{\rho} & \boldsymbol{\rho}\end{bmatrix}\boldsymbol{A}'$

6 ☐ Don’t know.

The problem continues on the next page

Question 7.3.

The partial correlation $\rho_{x_1 x_2|y_2}$ is:

1 ☐ $\frac{\rho-\rho^2}{1-\rho}$

2 ☐ $\frac{\rho-\rho^2}{1-\rho^2}$

3 ☐ $\frac{\rho}{1-\rho^2}$

4 ☐ ρ^2

5 ☐ $\frac{\rho^2}{1-\rho}$

6 ☐ Don't know.

Question 7.4.

Assume that we have 10 observations of the variables ***X*** and ***Y***. The usual test of $\rho_{x_1 x_2|y_2} = 0$ against all alternatives, has – under the null hypothesis – the following distribution:

1 ☐ F(2,1)

2 ☐ t(8)

3 ☐ t(7)

4 ☐ N(0,1)

5 ☐ t(4)

6 ☐ Don't know.

The problem continues on the next page

Question 7.5.

The multiple correlation $\rho^2_{x_1|,y_1,y_2}$ is (*hint: see problem 4*):

1 ☐ $\frac{1}{1-\rho^2}$

2 ☐ $\frac{1+2\rho^3-2\rho^2}{1-\rho^2}$

3 ☐ $\frac{1+2\rho^3-3\rho^2}{1-\rho^2}$

4 ☐ $\frac{1-3\rho^2}{1-\rho^2}$

5 ☐ $\frac{2\rho^3-3\rho^2}{1-\rho^2}$

6 ☐ Don’t know.

LAST PAGE:
END OF THE EXAM SET

SAS-PROGRAM

```
proc print data=pendling(obs=20);
run;

proc glm data=pendling;
class city distance;
model independent
         spouces
         topleaders
         salaryHigh
         salaryMedium
         salaryLow
         salaryOther
         salaryUnknown = city distance;
manova h=_all_/printe printh;
run;
```

Some SAS-outputs have been omitted

Obs	city	distance	independent	spouces	topleaders	salaryHigh	salaryMedium	salaryLow	salaryOther	salaryUnknown
1	Koebenhavn	0.000000	0.028640	0.000294	0.000291	0.001292	0.000318	0.001186	0.000252	0.003259
2	Frederiksberg	0.000000	0.047590	0.000617	0.000854	0.003511	0.000901	0.002040	0.000427	0.006025
3	Dragoer	0.000000	0.102620	0.001619	0.002914	0.005827	0.001295	0.008741	0.001295	0.028812
4	Taarnby	0.000000	0.024870	0.000682	0.000401	0.000963	0.000321	0.002286	0.000401	0.004252
5	Albertslund	0.000000	0.018740	0.000557	0.000405	0.000557	0.000203	0.001672	0.000405	0.002736
6	Ballerup	0.000000	0.014930	0.000237	0.000237	0.000878	0.000237	0.001876	0.000356	0.002944
7	Broendby	0.000000	0.020050	0.000739	0.000261	0.000478	0.000217	0.001783	0.000174	0.003045
8	Gentofte	0.000000	0.052910	0.000732	0.001867	0.005728	0.001161	0.002473	0.000252	0.013499
9	Gladsaxe	0.000000	0.026140	0.000484	0.000339	0.001282	0.000605	0.001814	0.000290	0.003410
10	Glostrup	0.000000	0.013900	0.000144	0.000337	0.000529	0.000096	0.001155	0.000241	0.003416
11	Herlev	0.000000	0.019980	0.000283	0.000236	0.000661	0.000378	0.001370	0.000331	0.003732
12	Hvidovre	0.000000	0.027290	0.000389	0.000354	0.000742	0.000283	0.002934	0.000636	0.004808
13	Hoeje-Taastrup	0.000000	0.020340	0.000347	0.000520	0.000694	0.000462	0.002832	0.000607	0.005028
14	Ishoej	0.000000	0.034380	0.000236	0.000473	0.000945	0.000354	0.003190	0.000945	0.008389
15	Lyngby-Taarbaek	0.000000	0.034140	0.000577	0.000941	0.002943	0.000759	0.001457	0.000243	0.007040
16	Roedovre	0.000000	0.033740	0.000808	0.001305	0.000621	0.000808	0.003231	0.000497	0.006400
17	Vallensbaek	0.000000	0.051600	0.001290	0.000860	0.001290	0.000860	0.005590	0.001290	0.015481
18	Alleroed	0.000000	0.038770	0.001262	0.001893	0.002664	0.000982	0.003576	0.000771	0.009185
19	Egedal	0.000000	0.083080	0.001988	0.002334	0.004582	0.001816	0.008213	0.001556	0.019193
20	Fredensborg	0.000000	0.086230	0.002448	0.002527	0.004501	0.001500	0.005922	0.000790	0.019662

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
city	99	Aabenraa Aalborg Aarhus Albertslund Alleroed Assens Ballerup Billund Bornholm Broendby Broenderslev Christiansoe Dragoer Egedal Esbjerg Faaborg-Midtfyn Fanoe Favrskov Faxe Fredensborg Fredericia Frederiksberg Frederikshavn Frederikssund Furesoe Gentofte Gladsaxe Glostrup Greve Gribskov Guldborgsund Haderslev Halsnaes Hedensted Helsingoer Herlev Herning Hilleroed Hjoerring Hoeje-Taastrup Hoersholm Holbaek Holstebro Horsens Hvidovre Ikast-Brande Ishoej Jammerbugt Kalundborg Kerteminde Koebenhavn Koege Kolding Laesoe Langeland Lejre Lemvig Lolland Lyngby-Taarbaek Mariagerfjord Middelfart Morsoe Naestved Norddjurs Nordfyns Nyborg Odder Odense Odsherred Randers Rebild Ringkoebing-Skj Ringsted Roedovre Roskilde Rudersdal Samsoe Silkeborg Skanderborg Skive Slagelse Soenderborg Solroed Soroe Stevns Struer Svendborg Syddjurs Taarnby Thisted Toender Vallensbaek Varde Vejen Vejle Vesthimmerlands Viborg Vordingborg aeroe
distance	8	0 2.5 5 10 20 30 40 50

Number of Observations Read	792
Number of Observations Used	792

The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	independent	spouces	topleaders	salaryHigh	salaryMedium	salaryLow	salaryOther	salaryUnknown
independent	0.072405	0.002343	0.006323	0.030663	0.012864	0.045904	0.005506	0.016977
spouces	0.002343	0.000144	0.000170	0.000743	0.000361	0.000656	0.000125	0.000418
topleaders	0.006323	0.000170	0.004038	0.020799	0.007197	0.024251	0.002766	0.006113
salaryHigh	0.030663	0.000743	0.020799	0.173111	0.047890	0.148231	0.015724	0.028085
salaryMedium	0.012864	0.000361	0.007197	0.047890	0.019857	0.046981	0.006808	0.008552
salaryLow	0.045904	0.000656	0.024251	0.148231	0.046981	0.321032	0.049171	0.066990
salaryOther	0.005506	0.000125	0.002766	0.015724	0.006808	0.049171	0.016025	0.009173
salaryUnknown	0.016977	0.000418	0.006113	0.028085	0.008552	0.066990	0.009173	0.024238

The GLM Procedure
Multivariate Analysis of Variance

H = Type III SSCP Matrix for city

	independent	spouces	topleaders	salaryHigh	salaryMedium	salaryLow	salaryOther	salaryUnknown
independent	0.01711	0.00059	-0.00212	-0.01228	-0.01076	0.00012	-0.00162	0.00895
spouces	0.00059	0.00003	-0.00008	-0.00063	-0.00038	0.00008	0.00003	0.00035
topleaders	-0.00212	-0.00008	0.00065	0.00218	0.00176	-0.00121	-0.00010	-0.00108
salaryHigh	-0.01228	-0.00063	0.00218	0.03716	0.01171	-0.02145	-0.00742	-0.00927
salaryMedium	-0.01076	-0.00038	0.00176	0.01171	0.01004	-0.00558	-0.00030	-0.00650
salaryLow	0.00012	0.00008	-0.00121	-0.02145	-0.00558	0.02159	0.00498	0.00149
salaryOther	-0.00162	0.00003	-0.00010	-0.00742	-0.00030	0.00498	0.00465	-0.00021
salaryUnknown	0.00895	0.00035	-0.00108	-0.00927	-0.00650	0.00149	-0.00021	0.00627

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall city Effect
H = Type III SSCP Matrix for city
E = Error SSCP Matrix

S=8 M=44.5 N=338.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
Wilks' Lambda	0.07372765	2.68	784	5435.1	<.0001
Pillai's Trace	1.99721058	2.33	784	5488	<.0001
Hotelling-Lawley Trace	3.60827056	3.12	784	4916.4	<.0001
Roy's Greatest Root	1.57639735	11.03	98	686	<.0001

H = Type III SSCP Matrix for distance								
	independent	spouces	topleaders	salaryHigh	salaryMedium	salaryLow	salaryOther	salaryUnknown
independent	0.30942	0.00967	-0.01396	-0.09990	-0.04599	-0.15815	-0.03566	0.02740
spouces	0.00967	0.00030	-0.00047	-0.00334	-0.00151	-0.00557	-0.00126	0.00071
topleaders	-0.01396	-0.00047	0.00421	0.02888	0.01012	0.07105	0.01644	0.01264
salaryHigh	-0.09990	-0.00334	0.02888	0.20854	0.06996	0.51511	0.11987	0.09132
salaryMedium	-0.04599	-0.00151	0.01012	0.06996	0.02537	0.16360	0.03792	0.02551
salaryLow	-0.15815	-0.00557	0.07105	0.51511	0.16360	1.36216	0.31792	0.27245
salaryOther	-0.03566	-0.00126	0.01644	0.11987	0.03792	0.31792	0.07439	0.06381
salaryUnknown	0.02740	0.00071	0.01264	0.09132	0.02551	0.27245	0.06381	0.06812

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall distance Effect H = Type III SSCP Matrix for distance E = Error SSCP Matrix S=7 M=0 N=338.5					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
Wilks' Lambda	0.01356382	79.93	56	3661.8	<.0001
Pillai's Trace	2.06653647	35.87	56	4795	<.0001
Hotelling-Lawley Trace	13.11634859	158.70	56	2441.4	<.0001
Roy's Greatest Root	8.44103511	722.76	8	685	<.0001

SAS PROGRAM

```
proc factor data=pendling
rotate = varimax nfactors=3
         plots=(scree loadings) score outstat = factor_stat_out;
         var      independent
         spouces
         topleaders
         salaryHigh
         salaryMedium
         salaryLow
         salaryOther
         salaryUnknown;
run;

proc score data=pendling score=factor_stat_out out=Fscore;
var independent
         spouces
         topleaders
         salaryHigh
         salaryMedium
         salaryLow
         salaryOther
         salaryUnknown;
run;

data Fscore2;
set Fscore;
if distance NE 50 then delete;
run;

title 'Score Plots';
proc sgscatter data=Fscore2;
  plot Factor1*Factor2 /  datalabel=city  group=distance;
run;
proc sgscatter data=Fscore2;
  plot Factor1*Factor3 /  datalabel=city  group=distance;
run;
proc sgscatter data=Fscore2;
  plot Factor2*Factor3 /  datalabel=city  group=distance;
run;
```

Some SAS-outputs have been omitted

The FACTOR Procedure

Input Data Type	Raw Data
Number of Records Read	792
Number of Records Used	792
N for Significance Tests	792

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 8 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	4.65705641	2.58748913	0.5821	0.5821
2	2.06956728	1.34526809	0.2587	0.8408
3	0.72429919	0.51707082	0.0905	0.9314
4	0.20722838	0.05358036	0.0259	0.9573
5	0.15364802	0.06830463	0.0192	0.9765
6	0.08534339	0.02056548	0.0107	0.9871
7	0.06477790	0.02669848	0.0081	0.9952
8	0.03807943		0.0048	1.0000

3 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2	Factor3
independent	-0.24655	0.93064	0.17312
spouces	-0.27165	0.91379	0.19040
topleaders	0.90502	0.01777	0.29552
salaryHigh	0.89555	-0.03649	0.30534
salaryMedium	0.85518	-0.17244	0.41343
salaryLow	0.93862	0.14703	-0.26322
salaryOther	0.87277	0.12077	-0.36860
salaryUnknown	0.72618	0.54852	-0.31849

Variance Explained by Each Factor

Factor1	Factor2	Factor3
4.6570564	2.0695673	0.7242992

Final Communality Estimates: Total = 7.450923

independent	spouces	topleaders	salaryHigh	salaryMedium	salaryLow	salaryOther	salaryUnknown
0.95685759	0.94505116	0.90671210	0.89657028	0.93199361	0.97191100	0.91217650	0.92965065

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2	3
1	0.71650	0.67400	-0.17986
2	-0.07248	0.32837	0.94177
3	0.69381	-0.66174	0.28413

Rotated Factor Pattern

	Factor1	Factor2	Factor3
independent	-0.12399	0.02486	0.96998
spouces	-0.12877	-0.00903	0.96353
topleaders	0.85219	0.42027	-0.06208
salaryHigh	0.85615	0.38957	-0.10868
salaryMedium	0.91208	0.24619	-0.19875
salaryLow	0.47923	0.85510	-0.10514
salaryOther	0.36084	0.87182	-0.14797
salaryUnknown	0.25958	0.88032	0.29547

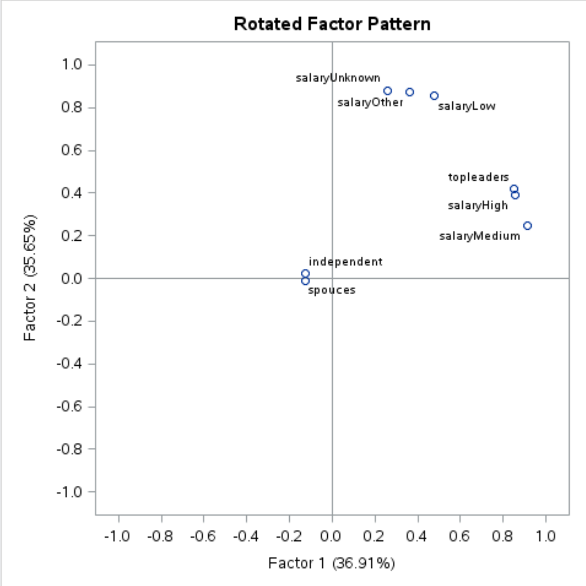
Variance Explained by Each Factor

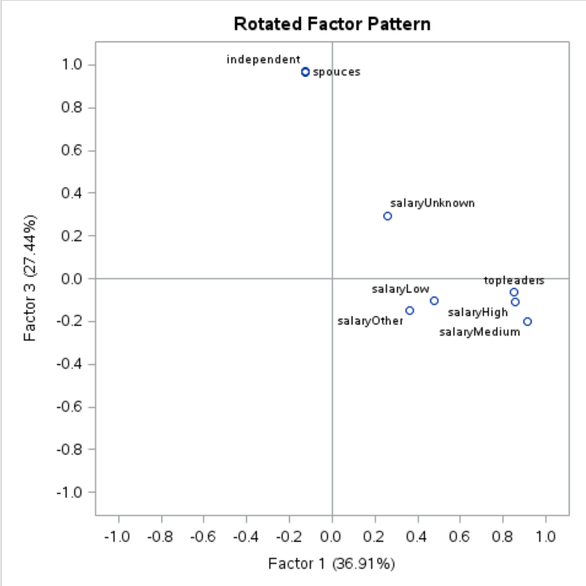
Factor1	Factor2	Factor3
2.7503238	2.6559237	2.0446753

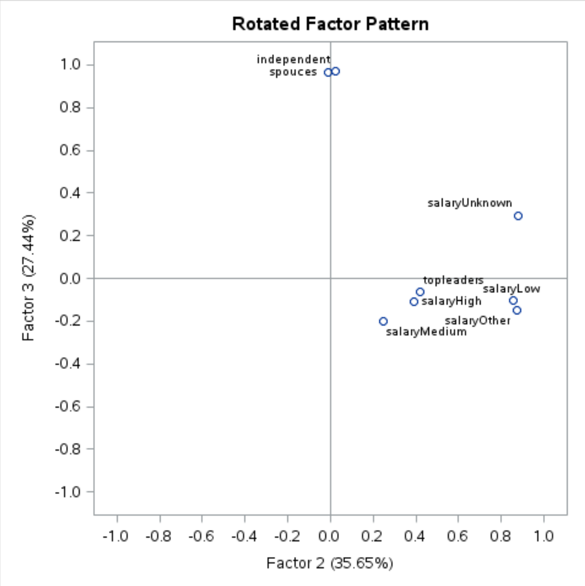
Final Communality Estimates: Total = 7.450923

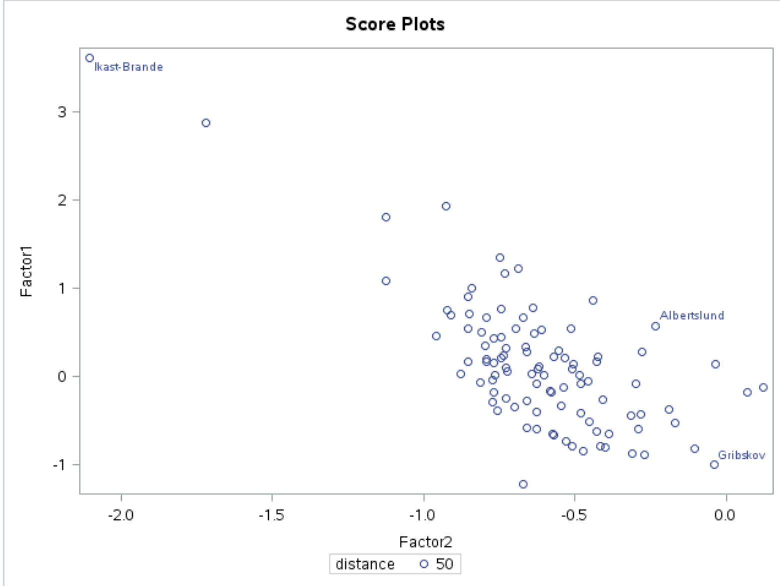
independent	spouces	topleaders	salaryHigh	salaryMedium	salaryLow	salaryOther	salaryUnknown
0.95685759	0.94505116	0.90671210	0.89657028	0.93199361	0.97191100	0.91217650	0.92965065

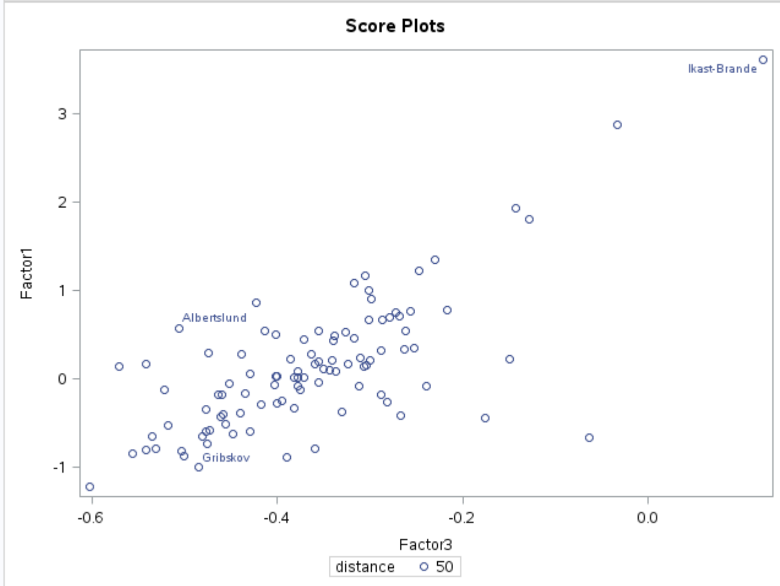
The FACTOR Procedure
Rotation Method: Varimax

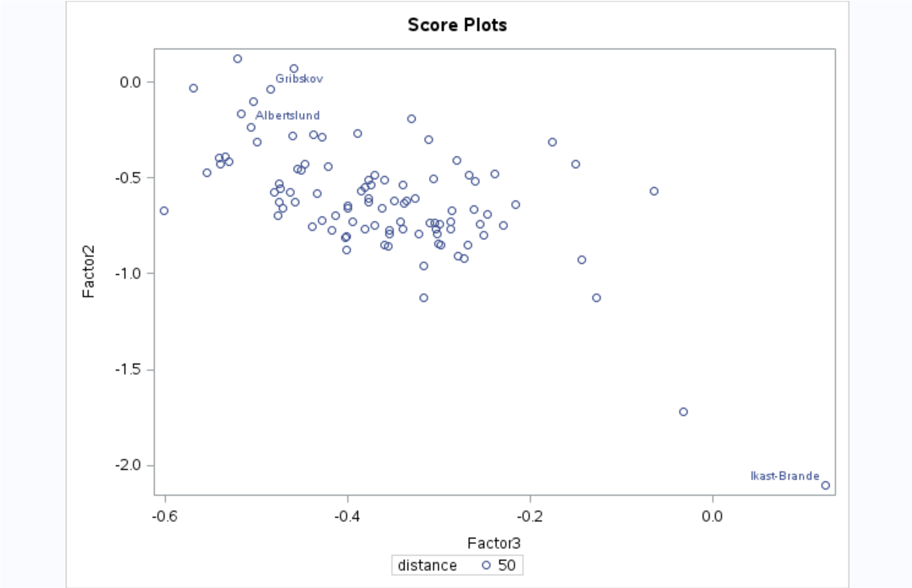





Rotated Factor Pattern
independent
spouces
salaryUnknown
topleaders
salaryLow
salaryHigh
salaryOther
salaryMedium
Factor 3 (27.44%)
Factor 2 (35.65%)


Score Plots
Ikast-Brande
Albertslund
Gribskov
Factor1
Factor2
distance 50


Score Plots
Ikast-Brande
Albertslund
Gribskov
Factor1
Factor3
distance 50


Score Plots
Gribskov
Albertslund
Ikast-Brande
0.0
-0.5
-1.0
-1.5
-2.0
Factor2
-0.6
-0.4
-0.2
0.0
Factor3
distance
50

SAS PROGRAM

```
title 'Full data-set - first twenty observations';
proc print data=exam.wine2(obs=20);
run;

title 'Summary statistics of full dataset';
proc means data=exam.wine2;
run;

* Reduce dataset to only quality 3 and 8;
data wine;
set exam.wine2;
if quality ne 3 and quality ne 8 then delete;
run;

title 'Classification with 3 variables';
proc discrim data=wine pool=test;
class quality;
var pH sulphates alcohol;
run;

title 'Classification with 1 variable';
proc discrim data=wine pool=yes;
class quality;
var alcohol;
run;
```


Full data-set – first twenty observations

Obs	Fixed acidity	Volatile acidity	citric acid	Resi dual sugar	chlor ides	Free sulfur di oxide	Total sulfur di oxide	density	pH	Sul phates	alcohol	Quality
1	11.6	0.580	0.66	2.20	0.074	10	47	1.00800	3.25	0.57	9.00	3
2	10.4	0.610	0.49	2.10	0.200	5	16	0.99940	3.16	0.63	8.40	3
3	7.4	1.185	0.00	4.25	0.097	5	14	0.99660	3.63	0.54	10.70	3
4	10.4	0.440	0.42	1.50	0.145	34	48	0.99832	3.38	0.86	9.90	3
5	8.3	1.020	0.02	3.40	0.084	6	11	0.99892	3.48	0.49	11.00	3
6	7.6	1.580	0.00	2.10	0.137	5	9	0.99476	3.50	0.40	10.90	3
7	6.8	0.815	0.00	1.20	0.267	16	29	0.99471	3.32	0.51	9.80	3
8	7.3	0.980	0.05	2.10	0.061	20	49	0.99705	3.31	0.55	9.70	3
9	7.1	0.875	0.05	5.70	0.082	3	14	0.99808	3.40	0.52	10.20	3
10	6.7	0.760	0.02	1.80	0.078	6	12	0.99600	3.55	0.63	9.95	3
11	7.4	0.590	0.08	4.40	0.086	6	29	0.99740	3.38	0.50	9.00	4
12	5.7	1.130	0.09	1.50	0.172	7	19	0.99400	3.50	0.48	9.80	4
13	8.8	0.610	0.30	2.80	0.088	17	46	0.99760	3.26	0.51	9.30	4
14	4.6	0.520	0.15	2.10	0.054	8	65	0.99340	3.90	0.56	13.10	4
15	8.3	0.675	0.26	2.10	0.084	11	43	0.99760	3.31	0.53	9.20	4
16	8.3	0.625	0.20	1.50	0.080	27	119	0.99720	3.16	1.12	9.10	4
17	5.0	1.020	0.04	1.40	0.045	41	85	0.99380	3.75	0.48	10.50	4
18	9.2	0.520	1.00	3.40	0.610	32	69	0.99960	2.74	2.00	9.40	4
19	7.6	0.680	0.02	1.30	0.072	9	20	0.99650	3.17	1.08	9.20	4
20	7.3	0.550	0.03	1.60	0.072	17	42	0.99560	3.37	0.48	9.00	4

Summary statistics of full dataset
The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
fixedacidity	1572	8.2883588	1.6918340	4.6000000	15.9000000
volatileacidity	1572	0.5276431	0.1791441	0.1200000	1.5800000
citricacid	1572	0.2700763	0.1950462	0	1.0000000
residualsugar	1572	2.5023855	1.2735243	0.9000000	15.5000000
chlorides	1572	0.0872500	0.0470103	0.0120000	0.6110000
freesulfurdioxide	1572	15.7837150	10.3346445	1.0000000	72.0000000
totalsulfurdioxide	1572	46.3244275	32.8802395	6.0000000	289.0000000
density	1572	0.9969575	0.0030224	0.9900700	1.0320000
pH	1572	3.3128435	0.1533583	2.7400000	4.0100000
sulphates	1572	0.6568193	0.1700827	0.3300000	2.0000000
alcohol	1572	10.4226781	1.0590473	8.4000000	14.9000000
quality	1572	5.6272265	0.7960576	3.0000000	8.0000000

Classification with 3 variables

The DISCRIM Procedure

Total Sample Size	22	DF Total	21
Variables	3	DF Within Classes	20
Classes	2	DF Between Classes	1

Number of Observations Read	22
Number of Observations Used	22

Class Level Information					
quality	Variable Name	Frequency	Weight	Proportion	Prior Probability
3	_3	10	10.0000	0.454545	0.500000
8	_8	12	12.0000	0.545455	0.500000

Within Covariance Matrix Information		
quality	Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	3	-10.03796
8	3	-8.39821
Pooled	3	-8.79576

Classification with 3 variables

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
5.685803	6	0.4593

Since the Chi-Square value is not significant at the 0.1 level, a pooled covariance matrix will be used in the discriminant function.
Reference: Morrison, D.F. (1976) Multivariate Statistical Methods p252.

Classification with 3 variables

The DISCRIM Procedure

Generalized Squared Distance to quality		
From quality	3	8
3	0	32.48466
8	32.48466	0

Linear Discriminant Function for quality		
Variable	3	8
Constant	-176.84275	-156.35796
pH	130.26761	86.97796

Linear Discriminant Function for quality		
Variable	3	8
sulphates	33.05130	61.98355
alcohol	-10.82904	-1.39863

Classification with 3 variables

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

Number of Observations and Percent Classified into quality			
From quality	3	8	Total
3	10 100.00	0 0.00	10 100.00
8	0 0.00	12 100.00	12 100.00
Total	10 45.45	12 54.55	22 100.00
Priors	0.5	0.5	

Error Count Estimates for quality			
	3	8	Total
Rate	0.0000	0.0000	0.0000
Priors	0.5000	0.5000	

Classification with 1 variable

The DISCRIM Procedure

Total Sample Size	22	DF Total	21
Variables	1	DF Within Classes	20
Classes	2	DF Between Classes	1

Number of Observations Read	22
Number of Observations Used	22

Class Level Information					
quality	Variable Name	Frequency	Weight	Proportion	Prior Probability
3	_3	10	10.0000	0.454545	0.500000
8	_8	12	12.0000	0.545455	0.500000

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
1	0.05349

Classification with 1 variable

The DISCRIM Procedure

Generalized Squared Distance to quality		
From quality	3	8
3	0	4.70685
8	4.70685	0

Linear Discriminant Function for quality		
Variable	3	8
Constant	-46.97020	-70.35130
alcohol	9.43650	11.54878

Classification with 1 variable

The DISCRIM Procedure

Classification Summary for Calibration Data: WORK.WINE

Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into quality			
From quality	3	8	Total
3	10 100.00	0 0.00	10 100.00
8	3 25.00	9 75.00	12 100.00
Total	13 59.09	9 40.91	22 100.00
Priors	0.5	0.5	

Error Count Estimates for quality			
	3	8	Total
Rate	0.0000	0.2500	0.1250
Priors	0.5000	0.5000	