Written test, date: 5. January 2001 *Course no.* : 04241

Course name: Multivariate Statistics "Statistik 2".

Aids allowed: All usual ones

"Weighting": The questions are given equal weight.

This exam is answered by:

____________________ ____________________ ________

(name) (signature) (study no.)

There is a total of 29 questions for the 10 problems. The answers to the 29 questions must be written into the table below.

Problem	1	1	1	1	1	1	2	2	3	4
Question	1.1	1.2	1.3	1.4	1.5	1.6	2.1	2.2	3.1	4.1
Answer										

Problem	4	5	5	5	5	5	5	6	6	6
Question	4.2	5.1	5.2	5.3	5.4	5.5	5.6	6.1	6.2	6.3
Answer										

Problem	6	6	6	7	8	8	8	9	10	XX
Question	6.4	6.5	6.6	7.1	8.1	8.2	8.3	9.1	10.1	XX
Answer										XX XX

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") gives 0 points. The total number of points, needed for a satisfactorily answered exam is determined at the final evaluation of the exam.

Remember to write your name, signature and table 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. The last page is page 18; please check that it is there.

Problem 1.

Supplement A with SAS-program and SAS-output belongs to this problem.

The alternative answers to the questions can contain rounded values from the SAS-output.

A laboratory uses mercury thermometers. It is known that the thermometers are produced with a random deviation from the correct temperature. The random deviation is constant throughout the lifetime of the thermometer. The laboratory has 9 thermometers marked with the numbers 1 to 9. In order to calibrate the thermometers they are put into the same water bath. The water bath has the same temperature throughout the calibration experiment. The thermometers are read-off 5 times each.

First, the data are analysed using an analysis of variance model as shown below:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Here: $\sum_i \alpha_i = 0$, and each of the $\varepsilon_{ij} \in N(0, \sigma^2)$ are assumed independent.

Question 1.1.

What is the best estimate of μ?

1 □ 50.20

2 □ 8.82

3 □ 1.10

4 □ 8

5 □ 180.25

6 □ Don't know.

The problem continues on the next page

Question 1.2.

How many degrees of freedom does the estimated residual variance have?

1 ☐ 32

2 ☐ 36

3 ☐ 8

4 ☐ 44

5 ☐ 45

6 ☐ Don't know.

Question 1.3.

What is the estimate of the standard deviation corresponding to the read-off error?

1 ☐ 1.102

2 ☐ 0.078

3 ☐ 0.975

4 ☐ 0.0061

5 ☐ 0.220

6 ☐ Don't know.

Question 1.4.

The usual test statistic for checking if there is a difference between thermometers is calculated as:

1 ☐ 0.976^2

2 ☐ $\frac{8.82/8}{0.220/36}$

3 ☐ $\frac{1.102}{0.078}$

4 ☐ $\frac{180.25}{50.20}$

5 ☐ $\frac{0.0061}{0.078}$

6 ☐ Don't know.

The problem continues on the next page

Question 1.5.

The estimated deviation $\hat{\alpha}_1$ for thermometer 1 is:

1 ☐ 0.090-0.078

2 ☐ 50.66

3 ☐ 50.66-50.20

4 ☐ 50.61-50.66

5 ☐ 50.66/0.0901

6 ☐ Don't know.

Question 1.6.

Now consider the 9 thermometers as a random selection from a large population. We can characterise the population of thermometers by a variance corresponding to the production variance.

This is usually estimated by:

1 ☐ 8.817/8

2 ☐ 0.220/36

3 ☐ $\frac{1}{9}(0.0901^2 + 0.0261^2 + 0.1003^2 + 0.0305^2 + 0.969^2 + 0.1313^2 + 0.0510^2 + 0.0524^2 + 0.0570^2)$

4 ☐ $(8.817 - 0.220)/8$

5 ☐ $(1.102 - 0.006)/5$

6 ☐ Don't know.

Problem 2.

You are informed that:

$$\underline{X}_i = \begin{pmatrix} X_{1i} \\ X_{2i} \\ X_{3i} \\ X_{4i} \end{pmatrix} \in N\left(\underline{\mu}_X, \underline{\underline{\Sigma}}\right) \quad , \qquad i = 1, \ldots, n_X$$

$$\underline{Y}_j = \begin{pmatrix} Y_{1j} \\ Y_{2j} \\ Y_{3j} \\ Y_{4j} \end{pmatrix} \in N\left(\underline{\mu}_Y, \underline{\underline{\Sigma}}\right) \quad , \qquad j = 1, \ldots, n_Y$$

$$\underline{Z}_k = \begin{pmatrix} Z_{1k} \\ Z_{2k} \\ Z_{3k} \\ Z_{4k} \end{pmatrix} \in N\left(\underline{\mu}_Z, \underline{\underline{\Sigma}}\right) \quad , \qquad k = 1, \ldots, n_Z$$

All observations can be assumed independent of each other.

One now wants to test the following hypothesis: $\underline{\mu}_X = \underline{\mu}_Y = \underline{\mu}_Z$

The following auxillary variables are defined:

$$\overline{\underline{A}} = \frac{1}{n_X + n_Y + n_Z} \left(\sum_{i=1}^{n_X} \underline{X}_i + \sum_{j=1}^{n_Y} \underline{Y}_j + \sum_{k=1}^{n_Z} \underline{Z}_k \right)$$

$$\overline{\underline{A}}_X = \frac{1}{n_X} \sum_{i=1}^{n_X} \underline{X}_i \quad ; \quad \overline{\underline{A}}_Y = \frac{1}{n_Y} \sum_{j=1}^{n_Y} \underline{Y}_j \quad ; \quad \overline{\underline{A}}_Z = \frac{1}{n_Z} \sum_{k=1}^{n_Z} \underline{Z}_k$$

$$\underline{\underline{U}}_0 = \sum_{i=1}^{n_X} (\underline{X}_i - \overline{\underline{A}})(\underline{X}_i - \overline{\underline{A}})' + \sum_{j=1}^{n_Y} (\underline{Y}_j - \overline{\underline{A}})(\underline{Y}_j - \overline{\underline{A}})' + \sum_{k=1}^{n_Z} (\underline{Z}_k - \overline{\underline{A}})(\underline{Z}_k - \overline{\underline{A}})'$$

$$\underline{\underline{U}}_1 = \sum_{i=1}^{n_X} (\underline{X}_i - \overline{\underline{A}}_X)(\underline{X}_i - \overline{\underline{A}}_X)' + \sum_{j=1}^{n_Y} (\underline{Y}_j - \overline{\underline{A}}_Y)(\underline{Y}_j - \overline{\underline{A}}_Y)' + \sum_{k=1}^{n_Z} (\underline{Z}_k - \overline{\underline{A}}_Z)(\underline{Z}_k - \overline{\underline{A}}_Z)'$$

$$\underline{\underline{U}}_2 = n_X (\overline{\underline{A}}_X - \overline{\underline{A}})(\overline{\underline{A}}_X - \overline{\underline{A}})' + n_Y (\overline{\underline{A}}_Y - \overline{\underline{A}})(\overline{\underline{A}}_Y - \overline{\underline{A}})' + n_Z (\overline{\underline{A}}_Z - \overline{\underline{A}})(\overline{\underline{A}}_Z - \overline{\underline{A}})'$$

The problem continues on the next page

Question 2.1.

The usual test statistic, which should be compared to a suitable percentile in an U-distribution, is:

1 ☐ $\dfrac{\det(\underline{\underline{U}}_0)}{\det(\underline{\underline{U}}_1)}$

2 ☐ $\dfrac{\det(\underline{\underline{U}}_0)}{\det(\underline{\underline{U}}_2)}$

3 ☐ $\dfrac{\det(\underline{\underline{U}}_1)}{\det(\underline{\underline{U}}_0)}$

4 ☐ $\dfrac{\det(\underline{\underline{U}}_1)}{\det(\underline{\underline{U}}_2)}$

5 ☐ $\dfrac{\det(\underline{\underline{U}}_2)}{\det(\underline{\underline{U}}_0)}$

6 ☐ Don't know.

Question 2.2.

The U-distribution has the following degrees of freedom:

1 ☐ $(4, 2, n_X + n_Y + n_Z - 3)$

2 ☐ $(3, 3, n_X + n_Y + n_Z - 4)$

3 ☐ $(2, 4, n_X + n_Y + n_Z - 1)$

4 ☐ $(4, 2, n_X + n_Y + n_Z - 2)$

5 ☐ $(3, 2, n_X + n_Y + n_Z - 3)$

6 ☐ Don't know.

Problem 3.

You are informed that:

$$\underline{X} \in N_p\left(\underline{0}, \underline{\underline{I}}\right)$$

Question 3.1

$\underline{a}$ is a p-dimensional constant column-vector. $\underline{\underline{B}}$ is a $p \times p$ matrix of constants.

Which one of the following statements is true for $p > 1$?

1 ☐ $\underline{Y} = \underline{a}\underline{\underline{B}} + \underline{X} \in N\left(\underline{a}, \underline{\underline{BB'}}\right)$

2 ☐ $\underline{Y} = \underline{\underline{B}}\underline{a} + \underline{X} \in N\left(\underline{a}, \underline{\underline{BB'}}\right)$

3 ☐ $\underline{Y} = \underline{\underline{B}} + \underline{a}'\underline{X} \in N\left(\underline{a}, \underline{\underline{BB'}}\right)$

4 ☐ $\underline{Y} = \underline{\underline{B}} + \underline{X}\underline{a}' \in N\left(\underline{a}, \underline{\underline{BB'}}\right)$

5 ☐ $\underline{Y} = \underline{a} + \underline{\underline{B}}\underline{X} \in N\left(\underline{a}, \underline{\underline{BB'}}\right)$

6 ☐ Don't know.

Problem 4.

For a general linear model you are informed that:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} \quad ; \quad \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} \in N(\underline{0}, \sigma^2 \underline{\underline{I}})$$

Question 4.1

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

1 ☐ 1

2 ☐ 0

3 ☐ -1

4 ☐ $\frac{1}{2}$

5 ☐ $-\frac{1}{2}$

6 ☐ Don't know.

Question 4.2

The maximum likelihood estimate of α is:

1 ☐ $\hat{\alpha} = \frac{Y_1+Y_2}{2}$

2 ☐ $\hat{\alpha} = \frac{Y_1-Y_2}{2}$

3 ☐ $\hat{\alpha} = \frac{Y_1-Y_2+Y_3-Y_4}{2}$

4 ☐ $\hat{\alpha} = \frac{Y_1+Y_2+Y_3+Y_4}{4}$

5 ☐ Cannot be calculated from the information given.

6 ☐ Don't know.

Problem 5.

Supplement B with SAS-program and SAS-output belongs to this problem.

The alternative answers to the questions can contain rounded values from the SAS-output.

In the following we will use part of a large data set on the quality of education in "Primary school" from: "The Daily Telegraph", Thursday, 7 December 2000.

The data to be analysed here come from 3 regions (=REGION) in England. Each region has a number of schools (=SCHOOL). At each school the 11-year old pupils' knowledge of English, science and mathematics was tested. The result of the test was then compared to the expectation of the knowledge of 11-year olds in these subjects. The variables ENGLISH, SCIENCE, and MATHS show the percentage of 11-year old pupils at the school, who met the expectation for their age.

In the following the assumptions for performing a linear discriminant analysis, where one compares regions, are assumed to be fulfilled.

Question 5.1.

In determining the discriminant scores one assumes the following prior probability for region 1:

1 □ 0.3333

2 □ 0.4795

3 □ 0.5

4 □ -58.22

5 □ 100%

6 □ Don't know.

Question 5.2.

The estimate of the variance-covariance matrix used for calculating the discriminant scores is based on the following number of degrees of freedom:

1 □ 34

2 □ $(34 + 22 + 14)/3$

3 □ 73

4 □ 72

5 □ 70

6 □ Don't know.

The problem continues on the next page

Question 5.3.

The generalised variance of the variance-covariance matrix can be found as:

1 ☐ $127.82 + 115.95 + 69.83$

2 ☐ $5.07 + 5.13 + 1.80$

3 ☐ $127.69 + 116.20 + 69.10$

4 ☐ $e^{12.33}$

5 ☐ 3

6 ☐ Don't know.

Question 5.4.

An hypothesis on no difference between the 3 regions is best tested by:

1 ☐ Hotelling's T^2 in the two-sample case.

2 ☐ one-way multi-dimensional analysis of variance.

3 ☐ two-way multi-dimensional analysis of variance.

4 ☐ principal component analysis.

5 ☐ cannot be tested.

6 ☐ Don't know.

Question 5.5.

Using the estimated discriminant scores, the number of *mis*-classified schools in region 1 is found to be:

1 ☐ 18

2 ☐ 13+7

3 ☐ 73-18

4 ☐ 9+8

5 ☐ 7+4

6 ☐ Don't know.

The problem continues on the next page

Question 5.6.

We now consider region 3 only. A test for 0 (zero) correlation between "ENGLISH" and "SCIENCE" is:

1 □ significant at level 0.2, but not at level 0.1

2 □ significant at level 0.1, but not at level 0.05

3 □ significant at level 0.05, but not at level 0.01

4 □ significant at level 0.01, but not at level 0.005

5 □ significant at level 0.005, but not at level 0.001

6 □ Don't know.

Problem 6.

Supplement C with SAS-program and SAS-output belongs to this problem. "####" indicates that the corresponding information is deleted.

The alternative answers to the questions can contain rounded values from the SAS-output.

In the following we will use part of a large data set on the quality of education in "Primary school" from: "The Daily Telegraph", Thursday, 7 December 2000. Data correspond to REGION=1 in the previous problem.

At a number of schools in a region in England the 11-year old pupils' knowledge of English, science and mathematics was tested. The result of the test was then compared to the expectation for the knowledge of 11-year olds in these subjects. The variables ENGLISH, SCIENCE, and MATHS show the percentage of 11-year old pupils at the school, who met the expectation for their age.

In order to analyse the correlation between the variables a factor analysis was performed.

The problem continues on the next page

Question 6.1.

The usual test for equality of the 2 smallest eigenvalues in the correlation matrix has the following approximate null-hypothesis distribution:

1 ☐ $\chi^2(1)$

2 ☐ $\chi^2(2)$

3 ☐ $\chi^2(3)$

4 ☐ $\chi^2(4)$

5 ☐ $\chi^2(5)$

6 ☐ Don't know.

Question 6.2.

The principal factor solution corresponding to factor 1 is found in the analysis as:

1 ☐ $\begin{pmatrix} 0.54 \\ 0.61 \\ 0.58 \end{pmatrix}$

2 ☐ $\begin{pmatrix} 0.54 \\ 0.82 \end{pmatrix}$

3 ☐ $\begin{pmatrix} 0.80 \\ 0.59 \end{pmatrix}$

4 ☐ $\begin{pmatrix} 0.80 \\ 0.91 \\ 0.87 \end{pmatrix}$

5 ☐ 2.24

6 ☐ Don't know.

The problem continues on the next page

Question 6.3.

Uniqueness for the variable ENGLISH before varimax rotation is:

1 □ 0.9896

2 □ 2.7525

3 □ 0.8033+0.5869

4 □ 0.0104

5 □ 0.0351

6 □ Don't know.

Question 6.4.

Uniqueness for the variable ENGLISH after varimax rotation is:

1 □ 0.0104

2 □ 0.8026

3 □ 0.2946

4 □ 1.6248

5 □ 3-2.7525

6 □ Don't know.

Question 6.5.

Which one of the following interpretations is most sensible:

1 □ Un-rotated factor 1 describes the schools' quality of teaching in mathematics in particular. Un-rotated factor 2 describes the schools' quality of teaching in English.

2 □ Un-rotated factor 1 describes the general quality of the schools' teaching. Rotated factor 2 describes the schools' quality of teaching in English in particular.

3 □ Un-rotated factor 2 describes the schools' quality of teaching in mathematics. Rotated factor 2 describes the schools' quality of teaching in English.

4 □ Un-rotated factor 1 describes the schools' quality of teaching in English in particular. Rotated factor 1 describes the general quality of the schools' teaching.

5 □ Rotated factor 1 describes the schools' quality of teaching in English in particular. Rotated factor 2 describes the general quality of the schools' teaching.

6 □ Don't know.

The problem continues on the next page

Question 6.6.

The performed varimax rotation corresponds to:

1 ☐ a -44.2° rotation

2 ☐ a 30.9° rotation

3 ☐ a reflection in the 2'nd axis.

4 ☐ a reflection in a line with an angle of -30.9° to the 1'st axis.

5 ☐ a -36.6° rotation

6 ☐ Don't know.

Problem 7.

Consider the stochastic variables Y, X_1, and X_2. You are informed that $\mathrm{Corr}(Y, X_1) \neq 0$ and $\mathrm{Corr}(Y, X_2) \neq 0$.

The correlation ρ between Y and $aX_1 + X_2$ is known to be maximal for a certain a. This maximal correlation is denoted ρ_{max}.

Question 7.1

Which one of the following statements is correct (for any a, Y, X_1, and X_2 which fulfill the above mentioned requirements):

1 ☐ $\rho_{\mathrm{max}} = 1$

2 ☐ $\mathrm{Corr}(X_1, X_2) = \rho_{\mathrm{max}}$

3 ☐ ρ^2_{max} is the variance reduction of Y by regression on X_1 and X_2

4 ☐ $\mathrm{Corr}(aX_1, X_2) = \rho_{\mathrm{max}}$

5 ☐ $\rho^2_{\mathrm{max}} = \mathrm{Corr}(aX_1, Y) \cdot \mathrm{Corr}(X_2, Y)$

6 ☐ Don't know.

Problem 8.

$\underline{X}$ and $\underline{Y}$ each contain 2 elements. You are informed that:

$$\mathrm{E}\begin{pmatrix} \underline{X} \\ \underline{Y} \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

$$\mathrm{D}\begin{pmatrix} \underline{X} \\ \underline{Y} \end{pmatrix} = \begin{pmatrix} e & f & g & h \\ f & i & j & k \\ g & j & l & m \\ h & k & m & n \end{pmatrix}$$

Question 8.1

What is $\mathrm{E}(\underline{X} + \underline{Y})$:

1 ☐ $\begin{pmatrix} a+b \\ c+d \end{pmatrix}$

2 ☐ $\begin{pmatrix} ab \\ cd \end{pmatrix}$

3 ☐ $\begin{pmatrix} a+c \\ b+d \end{pmatrix}$

4 ☐ $\begin{pmatrix} ac \\ bd \end{pmatrix}$

5 ☐ $\begin{pmatrix} a+d \\ b+c \end{pmatrix}$

6 ☐ Don't know.

The problem continues on the next page

Question 8.2

What is $\mathrm{D}(\underline{X} + \underline{Y})$:

1 □ $\begin{pmatrix} e+i+f+f & g+k+h+j \\ g+k+j+h & l+n+m+m \end{pmatrix}$

2 □ $\begin{pmatrix} e+n+h+h & f+m+m+f \\ g+k+k+g & i+l+j+j \end{pmatrix}$

3 □ $\begin{pmatrix} e+l+g+g & f+m+h+j \\ f+m+j+h & i+n+k+k \end{pmatrix}$

4 □ $\begin{pmatrix} e+g+h+m & f+h+k+n \\ f+j+g+l & i+k+j+m \end{pmatrix}$

5 □ $\begin{pmatrix} e+l & f+m \\ f+m & i+n \end{pmatrix}$

6 □ Don't know.

Question 8.3

What is $\mathrm{Corr}(X_1, Y_2)$:

1 □ h

2 □ $\frac{h}{\sqrt{e/n}}$

3 □ $\frac{e \cdot n}{h}$

4 □ $\frac{h}{\sqrt{e \cdot n}}$

5 □ $h/e/n$

6 □ Don't know.

Problem 9.

In an experiment we assume that we have independent measurements: $X \in \mathrm{N}(\mu, \sigma^2)$.

Based on n such measurements we estimate the mean ($\overline{x}$) and the variance (s^2).

We are now about to perform a new experiment. Due to better equipment, it is known that the variance on the new measurements will be half that of the previous measurements.

Question 9.1.

What is a $1-\alpha$ prediction-interval for a future measurement with the new equipment?

1 ☐ $\left[\overline{x} - \mathrm{t}(n-1)_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{2}(\frac{1}{n}+1)} \ ; \ \overline{x} + \mathrm{t}(n-1)_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{2}(\frac{1}{n}+1)}\right]$

2 ☐ $\left[\overline{x} - \mathrm{t}(n)_{1-\alpha} s\sqrt{\frac{1}{n}} \ ; \ \overline{x} + \mathrm{t}(n)_{1-\alpha} s\sqrt{\frac{1}{n}}\right]$

3 ☐ $\left[\overline{x} - \mathrm{t}(n-1)_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{n}+\frac{1}{2}} \ ; \ \overline{x} + \mathrm{t}(n-1)_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{n}+\frac{1}{2}}\right]$

4 ☐ $\left[\overline{x} - \mathrm{t}(n)_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{n}+1} \ ; \ \overline{x} + \mathrm{t}(n)_{1-\frac{\alpha}{2}} s\sqrt{\frac{1}{n}+1}\right]$

5 ☐ $\left[\overline{x} - \mathrm{t}(n-1)_{1-\frac{\alpha}{2}} \frac{1}{2} s\sqrt{\frac{1}{n}} \ ; \ \overline{x} + \mathrm{t}(n-1)_{1-\frac{\alpha}{2}} \frac{1}{2} s\sqrt{\frac{1}{n}}\right]$

6 ☐ Don't know.

Problem 10.

A dataset contains correponding values of the independent variable x, and the dependent variable Y .

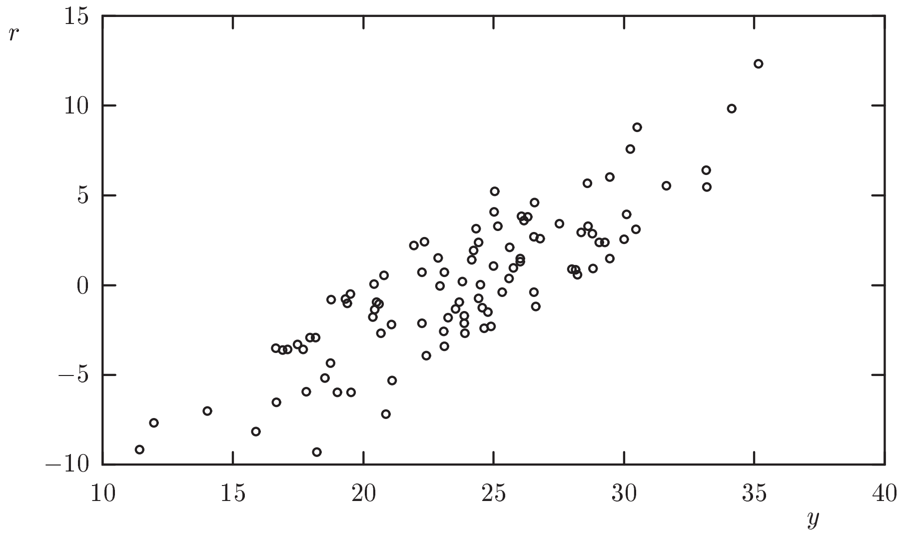
One has analysed the data by means of the folowing linear regression:

$$E(Y_i) = \alpha + \beta x_i$$

The empirical residuals are found as:

$$R_i = Y_i - \hat{Y}_i$$

In order to analyse the residuals one has made the following plot:



Question 10.1.

From the above plot one should suggest:

1 □ including a term of the form γx^2 in the analysis

2 □ omitting the constant term α from the analysis

3 □ performing a weighted analysis

4 □ taking the logarithm of the dependent variable

5 □ plotting the residuals as a function of $\hat{Y}$

6 □ Don't know.

```
/* a.sas   Crtd: 11-29-00 14:07 by BKE. Updt: 12-10-00 22:54 */
/* Purpose: Bilag/Supplement for exam in 04241 on 5 Jan. 2001 */

title1 'Bilag A - termometer data';
title2 'Supplement "Bilag A" - thermometer data';

proc print data=stat2.therm;
title3 'PROC PRINT output';

proc anova data=stat2.therm;
  class thermom;
  model temp=thermom;
  means thermom;
title3 'PROC ANOVA output';

run;
```


Bilag A - termometer data 1
Supplement "Bilag A" - thermometer data
PROC PRINT output

Obs	THERMOM	REP	TEMP
1	1	1	50.71
2	1	2	50.65
3	1	3	50.73
4	1	4	50.51
5	1	5	50.71
6	2	1	50.64
7	2	2	50.62
8	2	3	50.57
9	2	4	50.60
10	2	5	50.60
11	3	1	50.55
12	3	2	50.30
13	3	3	50.38
14	3	4	50.51
15	3	5	50.45
16	4	1	50.08
17	4	2	50.05
18	4	3	50.02
19	4	4	50.07
20	4	5	50.01
21	5	1	50.14
22	5	2	50.16
23	5	3	50.01
24	5	4	50.22
25	5	5	50.00
26	6	1	49.17
27	6	2	48.90
28	6	3	49.06
29	6	4	49.03
30	6	5	49.24
31	7	1	50.27
32	7	2	50.15
33	7	3	50.25
34	7	4	50.26
35	7	5	50.27
36	8	1	50.14
37	8	2	50.26
38	8	3	50.23
39	8	4	50.22
40	8	5	50.15
41	9	1	50.39
42	9	2	50.36
43	9	3	50.50
44	9	4	50.37
45	9	5	50.38

```
                       Bilag A - termometer data                        2
                Supplement "Bilag A" - thermometer data
                           PROC ANOVA output

                          The ANOVA Procedure

                        Class Level Information

                 Class         Levels    Values

                 THERMOM            9    1 2 3 4 5 6 7 8 9


                      Number of observations    45

/><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><</
                       Bilag A - termometer data                        3
                Supplement "Bilag A" - thermometer data
                           PROC ANOVA output

                          The ANOVA Procedure

Dependent Variable: TEMP

                                        Sum of
 Source                      DF        Squares    Mean Square   F Value   Pr > F

 Model                        8     8.81711111     1.10213889    180.25   <.0001

 Error                       36     0.22012000     0.00611444

 Corrected Total             44     9.03723111


              R-Square     Coeff Var      Root MSE     TEMP Mean

              0.975643      0.155774      0.078195      50.19756


 Source                      DF       Anova SS    Mean Square   F Value   Pr > F

 THERMOM                      8     8.81711111     1.10213889    180.25   <.0001

/><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><</
                       Bilag A - termometer data                        4
                Supplement "Bilag A" - thermometer data
                           PROC ANOVA output

                          The ANOVA Procedure

                Level of           -------------TEMP------------
                THERMOM      N             Mean          Std Dev

                1            5       50.6620000       0.09011104
                2            5       50.6060000       0.02607681
                3            5       50.4380000       0.10034939
                4            5       50.0460000       0.03049590
                5            5       50.1060000       0.09685040
                6            5       49.0800000       0.13133926
                7            5       50.2400000       0.05099020
                8            5       50.2000000       0.05244044
                9            5       50.4000000       0.05700877
```

```
/* b.sas   Crtd: 13.11.96 12:53 by BKE. Updt: 12-10-00 22:57 */
/* Purpose: Bilag/Supplement for exam in 04241 on 5 Jan. 2001 */

title1 'Bilag B - undervisningskvalitet data';
title2 'Supplement "Bilag B" - teaching quality data';

proc print data=stat2.quality;
title3 'PROC PRINT output';

proc discrim data=stat2.quality
  pool=yes bcov pcov tcov wcov bcorr pcorr tcorr wcorr simple;
class region;
var english science maths;
title3 'PROC DISCRIM output';

run;
```


Bilag B - undervisningskvalitet data 1
Supplement "Bilag B" - teaching quality data
PROC PRINT output

Obs	REGION	SCHOOL	ENGLISH	SCIENCE	MATHS
1	1	1	96	88	96
2	1	2	87	90	95
3	1	3	90	85	95
4	1	4	84	80	98
5	1	5	74	88	98
6	1	6	87	78	95
7	1	7	79	85	92
8	1	8	85	80	88
9	1	9	72	86	93
10	1	10	88	71	92
11	1	11	79	77	93
12	1	12	79	76	94
13	1	13	71	73	94
14	1	14	70	75	94
15	1	15	73	79	85
16	1	16	75	71	91
17	1	17	67	71	91
18	1	18	73	63	90
19	1	19	64	67	96
20	1	20	73	62	90
21	1	21	71	67	87
22	1	22	60	64	93
23	1	23	87	50	80
24	1	24	57	66	93
25	1	25	61	69	85
26	1	26	63	66	85
27	1	27	66	61	85
28	1	28	70	63	75
29	1	29	67	59	81
30	1	30	54	67	85
31	1	31	61	57	84
32	1	32	70	55	74
33	1	33	58	56	69
34	1	34	57	52	74
35	1	35	57	57	64
36	2	1	94	89	95
37	2	2	94	86	94
38	2	3	94	77	97
39	2	4	88	88	89
40	2	5	89	86	89
41	2	6	91	77	91
42	2	7	88	77	89
43	2	8	84	82	89
44	2	9	79	82	91
45	2	10	78	79	93
46	2	11	72	79	91
47	2	12	82	74	85
48	2	13	77	65	86
49	2	14	76	71	82
50	2	15	78	69	78
51	2	16	70	68	88
52	2	17	66	63	87
53	2	18	63	65	87
54	2	19	70	62	73
55	2	20	61	60	71
56	2	21	59	62	68
57	2	22	47	56	83
58	2	23	53	53	78
59	3	1	87	82	96

Bilag B - undervisningskvalitet data 2
Supplement "Bilag B" - teaching quality data
PROC PRINT output

Obs	REGION	SCHOOL	ENGLISH	SCIENCE	MATHS
60	3	2	77	85	100
61	3	3	84	80	96
62	3	4	79	79	93
63	3	5	83	73	90
64	3	6	75	70	92
65	3	7	80	68	82
66	3	8	75	64	83
67	3	9	79	60	83
68	3	10	74	57	82
69	3	11	72	55	82
70	3	12	68	63	76
71	3	13	63	54	86
72	3	15	66	62	71
73	3	16	62	59	69

/><><><><><><><><><><><><><><><><><><><><><><><><><><><><></

Bilag B - undervisningskvalitet data 3
Supplement "Bilag B" - teaching quality data
PROC DISCRIM output

The DISCRIM Procedure

Observations	73	DF Total	72
Variables	3	DF Within Classes	70
Classes	3	DF Between Classes	2

Class Level Information

REGION	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	_1	35	35.0000	0.479452	0.333333
2	_2	23	23.0000	0.315068	0.333333
3	_3	15	15.0000	0.205479	0.333333

Bilag B - undervisningskvalitet data 4
Supplement "Bilag B" - teaching quality data
PROC DISCRIM output

The DISCRIM Procedure
Within-Class Covariance Matrices

REGION = 1, DF = 34

Variable	ENGLISH	SCIENCE	MATHS
ENGLISH	121.5378151	73.9831933	46.9957983
SCIENCE	73.9831933	122.4571429	68.7554622
MATHS	46.9957983	68.7554622	70.9697479

REGION = 2, DF = 22

Variable	ENGLISH	SCIENCE	MATHS
ENGLISH	182.3596838	125.6798419	70.3122530
SCIENCE	125.6798419	112.3399209	59.4743083
MATHS	70.3122530	59.4743083	59.2411067

REGION = 3, DF = 14

Variable	ENGLISH	SCIENCE	MATHS
ENGLISH	57.3523810	56.1000000	50.5285714
SCIENCE	56.1000000	105.8285714	73.3285714
MATHS	50.5285714	73.3285714	83.6857143

```
                 Bilag B - undervisningskvalitet data                     5
             Supplement "Bilag B" - teaching quality data
                         PROC DISCRIM output

                        The DISCRIM Procedure

          Pooled Within-Class Covariance Matrix,     DF = 70

  Variable            ENGLISH            SCIENCE              MATHS

  ENGLISH         127.8161727         86.6540728         55.0303815
  SCIENCE          86.6540728        115.9517303         66.7531500
  MATHS            55.0303815         66.7531500         69.8267968


             Between-Class Covariance Matrix,     DF = 2

  Variable            ENGLISH            SCIENCE              MATHS

  ENGLISH         5.072628265        1.833183017       -2.797124139
  SCIENCE         1.833183017        5.133230921        0.057816774
  MATHS          -2.797124139        0.057816774        1.797823698


             Total-Sample Covariance Matrix,     DF = 72

  Variable            ENGLISH            SCIENCE              MATHS

  ENGLISH         127.6944444         85.4861111         51.6111111
  SCIENCE          85.4861111        116.2005327         64.9379756
  MATHS            51.6111111         64.9379756         69.1023592
```



```
                 Bilag B - undervisningskvalitet data                     6
             Supplement "Bilag B" - teaching quality data
                         PROC DISCRIM output

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

                             REGION = 1

        Variable       ENGLISH        SCIENCE          MATHS

        ENGLISH        1.00000        0.60644        0.50602
                                       0.0001         0.0019

        SCIENCE        0.60644        1.00000        0.73753
                        0.0001                        <.0001

        MATHS          0.50602        0.73753        1.00000
                        0.0019         <.0001

--------------------------------------------------------------------------------


                             REGION = 2

        Variable       ENGLISH        SCIENCE          MATHS

        ENGLISH        1.00000        0.87808        0.67648
                                       <.0001         0.0004

        SCIENCE        0.87808        1.00000        0.72904
                        <.0001                        <.0001

        MATHS          0.67648        0.72904        1.00000
                        0.0004         <.0001

--------------------------------------------------------------------------------


                             REGION = 3

        Variable       ENGLISH        SCIENCE          MATHS

        ENGLISH        1.00000        0.72009        0.72935
                                       0.0025         0.0020

        SCIENCE        0.72009        1.00000        0.77919
                        0.0025                        0.0006

        MATHS          0.72935        0.77919        1.00000
                        0.0020         0.0006

--------------------------------------------------------------------------------
```

Bilag B - undervisningskvalitet data 7
Supplement "Bilag B" - teaching quality data
PROC DISCRIM output

The DISCRIM Procedure

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

Variable	ENGLISH	SCIENCE	MATHS
ENGLISH	1.00000	0.71180 <.0001	0.58250 <.0001
SCIENCE	0.71180 <.0001	1.00000	0.74186 <.0001
MATHS	0.58250 <.0001	0.74186 <.0001	1.00000

Between-Class Correlation Coefficients / Pr > |r|

Variable	ENGLISH	SCIENCE	MATHS
ENGLISH	1.00000	0.35925 0.7661	-0.92624 0.2461
SCIENCE	0.35925 0.7661	1.00000	0.01903 0.9879
MATHS	-0.92624 0.2461	0.01903 0.9879	1.00000

Total-Sample Correlation Coefficients / Pr > |r|

Variable	ENGLISH	SCIENCE	MATHS
ENGLISH	1.00000	0.70179 <.0001	0.54943 <.0001
SCIENCE	0.70179 <.0001	1.00000	0.72468 <.0001
MATHS	0.54943 <.0001	0.72468 <.0001	1.00000

Bilag B - undervisningskvalitet data 8
Supplement "Bilag B" - teaching quality data
PROC DISCRIM output

The DISCRIM Procedure
Simple Statistics

Total-Sample

Variable	N	Sum	Mean	Variance	Standard Deviation
ENGLISH	73	5402	74.00000	127.69444	11.3002
SCIENCE	73	5135	70.34247	116.20053	10.7796
MATHS	73	6329	86.69863	69.10236	8.3128

REGION = 1

Variable	N	Sum	Mean	Variance	Standard Deviation
ENGLISH	35	2525	72.14286	121.53782	11.0244
SCIENCE	35	2454	70.11429	122.45714	11.0660
MATHS	35	3074	87.82857	70.96975	8.4244

REGION = 2

Variable	N	Sum	Mean	Variance	Standard Deviation
ENGLISH	23	1753	76.21739	182.35968	13.5041
SCIENCE	23	1670	72.60870	112.33992	10.5991
MATHS	23	1974	85.82609	59.24111	7.6968

REGION = 3

Variable	N	Sum	Mean	Variance	Standard Deviation
ENGLISH	15	1124	74.93333	57.35238	7.5731
SCIENCE	15	1011	67.40000	105.82857	10.2873
MATHS	15	1281	85.40000	83.68571	9.1480

Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	12.33048

Bilag B - undervisningskvalitet data 9
Supplement "Bilag B" - teaching quality data
PROC DISCRIM output

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

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

Generalized Squared Distance to REGION

From REGION	1	2	3
1	0	0.55390	0.48298
2	0.55390	0	0.52165
3	0.48298	0.52165	0

Linear Discriminant Function

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

Linear Discriminant Function for REGION

Variable	1	2	3
Constant	-58.22599	-55.34677	-56.47142
ENGLISH	0.19237	0.23696	0.27211
SCIENCE	-0.39123	-0.33579	-0.45083
MATHS	1.48021	1.36339	1.43957

Bilag B - undervisningskvalitet data 10
Supplement "Bilag B" - teaching quality data
PROC DISCRIM output

The DISCRIM Procedure
Classification Summary for Calibration Data: STAT2.QUALITY
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

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

Posterior Probability of Membership in Each REGION

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, \text{SUM}_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into REGION

From REGION	1	2	3	Total
1	18	9	8	35
	51.43	25.71	22.86	100.00
2	7	13	3	23
	30.43	56.52	13.04	100.00
3	4	4	7	15
	26.67	26.67	46.67	100.00
Total	29	26	18	73
	39.73	35.62	24.66	100.00
Priors	0.33333	0.33333	0.33333	

Error Count Estimates for REGION

	1	2	3	Total
Rate	0.4857	0.4348	0.5333	0.4846
Priors	0.3333	0.3333	0.3333	

```
/* c.sas   Crtd: 12-09-00 13:22 by BKE. Updt: 12-10-00 22:57 */
/* Purpose: Bilag/Supplement for exam in 04241 on 5 Jan. 2001 */

title1 'Bilag C - undervisningskvalitet data (kun region 1)';
title2 'Supplement "Bilag C" - teaching quality data (only region 1)';

proc print data=stat2.quality1;
title3 'PROC PRINT output';

proc factor data=stat2.quality1 eigenvectors nfactors=2 residuals
  preplot nplot=2 plot rotate=varimax;
var english science maths;
title3 'PROC FACTOR output';

run;
```


Bilag C - undervisningskvalitet data (kun region 1) 1
Supplement "Bilag C" - teaching quality data (only region 1)
PROC PRINT output

Obs	REGION	SCHOOL	ENGLISH	SCIENCE	MATHS
1	1	1	96	88	96
2	1	2	87	90	95
3	1	3	90	85	95
4	1	4	84	80	98
5	1	5	74	88	98
6	1	6	87	78	95
7	1	7	79	85	92
8	1	8	85	80	88
9	1	9	72	86	93
10	1	10	88	71	92
11	1	11	79	77	93
12	1	12	79	76	94
13	1	13	71	73	94
14	1	14	70	75	94
15	1	15	73	79	85
16	1	16	75	71	91
17	1	17	67	71	91
18	1	18	73	63	90
19	1	19	64	67	96
20	1	20	73	62	90
21	1	21	71	67	87
22	1	22	60	64	93
23	1	23	87	50	80
24	1	24	57	66	93
25	1	25	61	69	85
26	1	26	63	66	85
27	1	27	66	61	85
28	1	28	70	63	75
29	1	29	67	59	81
30	1	30	54	67	85
31	1	31	61	57	84
32	1	32	70	55	74
33	1	33	58	56	69
34	1	34	57	52	74
35	1	35	57	57	64

```
           Bilag C - undervisningskvalitet data (kun region 1)            2
       Supplement "Bilag C" - teaching quality data (only region 1)
                            PROC FACTOR output

                           The FACTOR Procedure
                 Initial Factor Method: Principal Components

                     Prior Communality Estimates: ONE


       Eigenvalues of the Correlation Matrix: Total = 3  Average = 1

               Eigenvalue    Difference    Proportion    Cumulative

           1    2.23827138    1.72406243        0.7461        0.7461
           2    0.51420894    0.26668926        0.1714        0.9175
           3    0.24751968                      0.0825        1.0000

           2 factors will be retained by the NFACTOR criterion.


                              Eigenvectors

                                       1               2

                 ENGLISH         0.53690         0.81840
                 SCIENCE         0.61004        -0.20889
                 MATHS           0.58275        -0.53534


                             Factor Pattern

                                 Factor1         Factor2

                 ENGLISH         0.80325         0.58686
                 SCIENCE         0.91266        -0.14979
                 MATHS           0.87184        -0.38389


                      Variance Explained by Each Factor

                            Factor1         Factor2

                          2.2382714       0.5142089


               Final Communality Estimates: Total = 2.752480

                   ENGLISH         SCIENCE           MATHS

                0.98961314      0.85539307      0.90747410


          Residual Correlations With Uniqueness on the Diagonal

                          ENGLISH         SCIENCE           MATHS

         ENGLISH          0.01039        -0.03876         0.03100
         SCIENCE         -0.03876         0.14461        -0.11567
         MATHS            0.03100        -0.11567         0.09253
```



```
           Bilag C - undervisningskvalitet data (kun region 1)            3
       Supplement "Bilag C" - teaching quality data (only region 1)
                            PROC FACTOR output

                           The FACTOR Procedure
                 Initial Factor Method: Principal Components

       Root Mean Square Off-Diagonal Residuals: Overall = 0.07267038

                   ENGLISH         SCIENCE           MATHS

                0.03509320      0.08626093      0.08467862


                Partial Correlations Controlling Factors

                          ENGLISH         SCIENCE           MATHS

         ENGLISH                1              -1               1
         SCIENCE               -1               1              -1
         MATHS                  1              -1               1


       Root Mean Square Off-Diagonal Partials: Overall = 1.00000000

                   ENGLISH         SCIENCE           MATHS

                 1.0000000       1.0000000       1.0000000
```

```
            Bilag C - undervisningskvalitet data (kun region 1)             4
        Supplement "Bilag C" - teaching quality data (only region 1)
                          PROC FACTOR output

                         The FACTOR Procedure
                 Initial Factor Method: Principal Components

Plot of Factor Pattern for Factor1 and Factor2

                               Factor1
                                  1

                              B  .9
                       C
                                 .8                A

                                 .7

                                 .6

                                 .5

                                 .4

                                 .3

                                 .2

                                 .1                              F
                                                                 a
                                                                 c
        -1 -.9-.8-.7-.6-.5-.4-.3-.2-.1  0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0t
                                                                 o
                                -.1                              r
                                                                 2
                                -.2

                                -.3

                                -.4

                                -.5

                                -.6

                                -.7

                                -.8

                                -.9

                                 -1

                  ENGLISH=A     SCIENCE=B     MATHS=C
```



```
            Bilag C - undervisningskvalitet data (kun region 1)             5
        Supplement "Bilag C" - teaching quality data (only region 1)
                          PROC FACTOR output

                         The FACTOR Procedure
                       Rotation Method: Varimax

                   Orthogonal Transformation Matrix

                                     1               2

                       1       0.80259         0.59653
                       2      -0.59653         0.80259


                       Rotated Factor Pattern

                              Factor1         Factor2

               ENGLISH        0.29460         0.95017
               SCIENCE        0.82185         0.42421
               MATHS          0.92873         0.21197


                   Variance Explained by Each Factor

                        Factor1         Factor2

                      1.6247685       1.1277118


           Final Communality Estimates: Total = ########

                 ENGLISH         SCIENCE           MATHS

               ##########      ##########      ##########
```

```
              Bilag C - undervisningskvalitet data (kun region 1)              6
          Supplement "Bilag C" - teaching quality data (only region 1)
                               PROC FACTOR output

                              The FACTOR Procedure
                            Rotation Method: Varimax

Plot of Factor Pattern for Factor1 and Factor2

                                     Factor1
                                        1
                                              C
                                       .9

                                       .8           B

                                       .7

                                       .6

                                       .5

                                       .4

                                       .3                               A

                                       .2
                                                                        F
                                       .1                               a
                                                                        c
        -1 -.9-.8-.7-.6-.5-.4-.3-.2-.1  0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0t
                                                                        o
                                      -.1                               r
                                                                        2
                                      -.2

                                      -.3

                                      -.4

                                      -.5

                                      -.6

                                      -.7

                                      -.8

                                      -.9

                                       -1

                     ENGLISH=A     SCIENCE=B     MATHS=C
```