Written test, date: 10. December 2001

Course no. : 02409

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 30 questions for the 11 problems. The answers to the 30 questions must be written into the table below.

Problem	1	1	2	2	2	2	2	2	2	3
Question	1.1	1.2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	3.1
Answer										

Problem	3	3	3	4	4	4	4	5	5	6
Question	3.2	3.3	3.4	4.1	4.2	4.3	4.4	5.1	5.2	6.1
Answer										

Problem	6	6	7	7	8	9	10	11	11	11
Question	6.2	6.3	7.1	7.2	8.1	9.1	10.1	11.1	11.2	11.3
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") 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 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. The last page is page 16; please check that it is there.

Problem 1.

A regression problem with 20 observations has 1 dependent variable and 3 independent variables X_1, X_2, X_3. The following table shows the result of all possible subsets regression. All models are considered significant. It is noted that all models include an intercept.

Variable(s) in model	R^2
X_1	0.48
X_2	0.52
X_3	0.46
X_1, X_2	0.62
X_1, X_3	0.65
X_2, X_3	0.61
X_1, X_2, X_3	0.69

Question 1.1.

An analysis of the same data using forward selection will result in the following sequence of models:

1 □ (X_2), (X_1, X_3), (X_1, X_2, X_3)

2 □ (X_2), (X_1, X_2), (X_1, X_2, X_3)

3 □ (X_1), (X_1, X_3), (X_1, X_2, X_3)

4 □ (X_3), (X_1, X_3), (X_1, X_2, X_3)

5 □ (X_3), (X_2, X_3), (X_1, X_2, X_3)

6 □ Don't know.

The problem continues on the next page

Question 1.2.

We now turn to an analysis using backward elimination. When going from a model with 2 independent variables (e.g. X_1, X_2) to 1 independent variable (e.g. X_1) the (usual) relevant test distribution is:

1 □ F(1,16)

2 □ F(1,17)

3 □ F(1,18)

4 □ F(1,19)

5 □ F(1,20)

6 □ Don't know.

Problem 2.

Enclosure A with SAS-program and SAS-output belongs to this problem. #### indicates that information has been concealed *(Danish: skjult el. fjernet)*.

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

The data are observations from the Landsat 4 Thematic Mapper. Data are from Eastern Greenland, from the north-western part of an island called Ymer Ø.

Each observation consists of values from 3 spectral bands named "tm1" (intensity of blue reflected light), "tm3" (intensity of red reflected light), and "tm7" (intensity of near infrared reflected light). Furthermore, a variable "rock" shows the result of a manual classification of each observation by a trained geologist. Here we will be considering 3 types of rock named "Bed 5", "Bed 8" and "Bed 15" respectively.

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

Question 2.1.

The usual test statistic for the (alternative) hypothesis that at least one of the rock types has a mean which differs from the others is U-distributed with the following degrees of freedom:

1 □ $(4, 2, 1008)$

2 □ $(3, 3, 1006)$

3 □ $(2, 2, 1008)$

4 □ $(3, 2, 1006)$

5 □ $(2, 3, 1006)$

6 □ Don't know.

The problem continues on the next page

Question 2.2.

The pooled estimate of the dispersion (covariance) matrix has the following degrees of freedom:

1 □ 1005

2 □ 1006

3 □ 1007

4 □ 1008

5 □ 1009

6 □ Don't know.

Question 2.3.

An observation with unknown rock-classification has the following values: (tm1,tm3,tm7)=(67,42,29). The value of the discriminant score for rock type "Bed 5" is estimated at:

1 □ $(-52.8 + 98.7 + 99.0) \cdot 67 + (1.8 - 1.3 - 2.1) \cdot 42 + (-0.1 + 0.1 + 1.0) \cdot 29$

2 □ 9.17

3 □ $-52.8 + 1.78 \cdot 67 - 0.09 \cdot 42 - 0.82 \cdot 29$

4 □ $-52.8 + 98.7 + 99.0$

5 □ $1.8 \cdot 67 - 1.3 \cdot 42 - 2.1 \cdot 29$

6 □ Don't know.

Question 2.4.

The prior probability used for rock type "Bed 5" in the linear discriminant analysis is:

1 □ 0.49

2 □ < 0.0001

3 □ 0.037

4 □ $\frac{1}{3}$

5 □ 9.17

6 □ Don't know.

The problem continues on the next page

Question 2.5.

Which pair of rock types are best separated based on the information in the 3 spectral bands (tm1, tm3, tm7):

1 □ "Bed 5" and "Bed 8"

2 □ "Bed 5" and "Bed 15"

3 □ "Bed 8" and "Bed 15"

4 □ "Bed 5", "Bed 8", and "Bed 15" are equally well seperated.

5 □ Cannot be answered with the information given.

6 □ Don't know.

Question 2.6.

The total number of misclassified observations is:

1 □ 13

2 □ 1009

3 □ 3

4 □ 117

5 □ 16

6 □ Don't know.

Question 2.7.

In this question we **only** consider observations from rock = "Bed 5".

Suppose we want to test the hypothesis that the correlation coefficient between variables "tm1" and "tm7" is 0.5. In that case the usual (approximative) test statistic is estimated at:

1 □ $(0.65 - 0.55)\sqrt{490}$

2 □ $(1.0 - 0.65)\sqrt{490}$

3 □ $\frac{0.5}{1-0.5^2}\sqrt{491}$

4 □ $\frac{0.57}{1-0.57^2}\sqrt{491}$

5 □ 0.57

6 □ Don't know.

Problem 3.

Enclosure 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.

The data are observations from an incinerator *(Danish: et forbrændingsanlæg)*. Each observation contains corresponding values of content of plastics, paper, garbage, water, and the energy content. Furthermore, a weighted average of the combustible products is included. (Calculated as: Combustible = 0.1·Garbage +0.2·Paper +0.7·Plastics)

In this problem we will mainly consider the first 4 variables and the output from `proc princomp`.

Question 3.1

The usual test statistic for the hypothesis: $H_0 : \lambda_1 \geq \lambda_2 \geq \lambda_3 = \lambda_4$ is approximately distributed as:

1 ☐ $\chi^2(1)$

2 ☐ $\chi^2(2)$

3 ☐ $\chi^2(3)$

4 ☐ $\chi^2(28)$

5 ☐ $\chi^2(29)$

6 ☐ Don't know.

Question 3.2

The total variance of the first 2 principal components is:

1 ☐ 34.5

2 ☐ 2

3 ☐ 0.79

4 ☐ 16.4

5 ☐ 11.4

6 ☐ Don't know.

The problem continues on the next page

Question 3.3

The variance of the reconstructed variable "plastics" based on the first 2 principle components is:

1 □ 4.96

2 □ $22.93 + 11.55$

3 □ $(-0.0057)^2 + (-0.3013)^2$

4 □ 2.23

5 □ $22.93 \cdot (-0.0057)^2 + 11.54 \cdot (-0.3013)^2$

6 □ Don't know.

Question 3.4.

Which one of the following interpretations is most sensible:

1 □ The first principle component is mainly a contrast between water content and plastics content.

2 □ A large positive value of principal component 2 implies a large garbage content.

3 □ A high garbage content and a low paper content gives a high value of principal component 1. Principal component 2 mainly describes water content.

4 □ Principal component 1 is mainly a contrast between water content and paper content. Principal component 2 mainly describes water content.

5 □ Principal component 1 is mainly a contrast between garbage and paper content. A large positive value of principal component 2 implies a large plastics content.

6 □ Don't know.

Problem 4.

Enclosure 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.

The data are observations from an incinerator *(Danish: et forbrændingsanlæg)*. Each observation contains corresponding values of content of plastics, paper, garbage, water, and the energy content. Furthermore, a weighted average of the combustible products is included. (Calculated as: Combustible = 0.1·Garbage +0.2·Paper +0.7·Plastics)

In this problem we will mainly consider output from `proc reg`.

The problem continues on the next page

Question 4.1.

The regression coefficient for the variable water is estimated at:

1 ☐ 2242

2 ☐ 30.41

3 ☐ 3.12

4 ☐ -37.36

5 ☐ 1.77

6 ☐ Don't know.

Question 4.2.

The sum of the squared differences between the individual observations of energy content and their average is:

1 ☐ 924.6

2 ☐ 664746

3 ☐ 24964

4 ☐ 689710

5 ☐ 30.41

6 ☐ Don't know.

Question 4.3.

The variance of (the estimate of) the first residual is estimated at:

1 ☐ $924.6(1 - 0.2346)$

2 ☐ 1.29

3 ☐ 1.06

4 ☐ 28.2

5 ☐ 1-0.59

6 ☐ Don't know.

The problem continues on the next page

Question 4.4.

The correlation between the parameter estimates of "Combustible" and "Water" is estimated at:

1 □ $\frac{1.71}{\sqrt{14.9 \cdot 3.12}}$

2 □ $3.86 \cdot 1.77$

3 □ 1.71^2

4 □ $\frac{-37.4}{41.1}$

5 □ Cannot be computed with the information given.

6 □ Don't know.

Problem 5.

Consider the following experiment with 6 observations:

Yield	Enzyme	Temperature
Y_{11}	1	t_1
Y_{12}	1	t_2
Y_{13}	1	t_3
Y_{21}	2	t_1
Y_{22}	2	t_2
Y_{23}	2	t_3

The yield Y of the process is assumed to depend on which type of enzyme (type 1 or type 2) was used and on the process temperature t. The yield differs by a constant amount for the 2 enzymes. The increase in yield per unit temperature is the same for the 2 enzymes.

Question 5.1

The described model for the experiment has:

1 □ 1 residual degree of freedom

2 □ 2 residual degrees of freedom

3 □ 3 residual degrees of freedom

4 □ 4 residual degrees of freedom

5 □ 5 residual degrees of freedom

6 □ Don't know.

The problem continues on the next page

Question 5.2

Which 2 of the following models both give a correct description of the experiment:

(A) $E(Y_{ij}) = \mu_i + \beta t_j$, where $i = 1, 2,\ j = 1, 2, 3.$

(B) $E(Y_{ij}) = \mu + \beta_i t_j$, where $i = 1, 2,\ j = 1, 2, 3.$

(C) $E(Y_{ij}) = \mu + \alpha_i + \beta t_j$, where $\alpha_1 + \alpha_2 = 0,\ i = 1, 2,\ j = 1, 2, 3.$

(D) $E(Y_{ij}) = \mu + \beta t_j$, where $i = 1, 2,\ j = 1, 2, 3.$

where: μ, μ_1, μ_2, α_1, α_2, β, β_1, β_2 are parameters in the relevant models.

1 ☐ models (A) and (B)

2 ☐ models (A) and (C)

3 ☐ models (B) and (C)

4 ☐ models (B) and (D)

5 ☐ models (C) and (D)

6 ☐ Don't know.

Problem 6.

Assume:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

We now consider the transformed variables: $Z_1 = X_1 + X_2$ and $Z_2 = X_1 - X_2$.

Let:

$$D\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

The problem continues on the next page

Question 6.1.

σ_{11} is equal to:

1 ☐ 0

2 ☐ $\sigma^2(1-\rho^2)$

3 ☐ $2\sigma^2(1-\rho)$

4 ☐ $2\sigma^2(1+\rho)$

5 ☐ $2\sigma^2$

6 ☐ Don't know.

Question 6.2.

σ_{22} is equal to:

1 ☐ 0

2 ☐ $\sigma^2(1-\rho^2)$

3 ☐ $2\sigma^2(1-\rho)$

4 ☐ $2\sigma^2(1+\rho)$

5 ☐ $2\sigma^2$

6 ☐ Don't know.

Question 6.3.

σ_{12} is equal to:

1 ☐ 0

2 ☐ $\sigma^2(1-\rho^2)$

3 ☐ $2\sigma^2(1-\rho)$

4 ☐ $2\sigma^2(1+\rho)$

5 ☐ $2\sigma^2$

6 ☐ Don't know.

Problem 7.

Consider the following one-sided analysis of variance:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

The residuals are assumed normally distributed.

Question 7.1

In this question the observations are mutually independent. However, the first and second observations are assumed to have twice the variance of the third observation.

The maximum likelihood estimate of the parameters μ_1 and μ_2 is:

1 ☐ $\hat{\mu}_1 = (Y_1 + Y_2)$ and $\hat{\mu}_2 = Y_3$

2 ☐ $\hat{\mu}_1 = \frac{1}{2}(Y_1 + Y_2)$ and $\hat{\mu}_2 = Y_3$

3 ☐ $\hat{\mu}_1 = \frac{1}{3}(Y_1 + Y_2)$ and $\hat{\mu}_2 = \frac{2}{3}Y_3$

4 ☐ $\hat{\mu}_1 = (Y_1 + Y_2)$ and $\hat{\mu}_2 = 2Y_3$

5 ☐ Cannot be estimated

6 ☐ Don't know.

The problem continues on the next page

Question 7.2

In this question the second and third observations are assumed to be dependent such that:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \in N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{3} & -\frac{2}{3} \\ 0 & -\frac{2}{3} & \frac{4}{3} \end{pmatrix}\right)$$

The following may be used as a hint to help answer the question:

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

The maximum likelihood estimate of the parameters μ_1 and μ_2 is:

1 ☐ $\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \begin{pmatrix} Y_1 + Y_2 + \frac{1}{2}Y_3 \\ \frac{1}{2}Y_2 + Y_3 \end{pmatrix}$

2 ☐ $\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \begin{pmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}^{-1} \begin{pmatrix} Y_1 + Y_2 + \frac{1}{2}Y_3 \\ \frac{1}{2}Y_2 + Y_3 \end{pmatrix}$

3 ☐ $\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \begin{pmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} Y_1 + Y_2 \\ Y_2 + Y_3 \end{pmatrix}$

4 ☐ $\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{3} & -\frac{2}{3} \\ 0 & -\frac{2}{3} & \frac{4}{3} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$

5 ☐ Cannot be estimated

6 ☐ Don't know.

Problem 8.

In this problem we consider example 6.4 in the lecture notes.

Question 8.1

We want to test the simultaneous hypothesis:

$$\theta_{11} = 0.3 \quad \wedge \quad \theta_{12} = 0.4 \quad \wedge \quad \theta_{21} = 0.8 \quad \wedge \quad \theta_{22} = 0.2$$

To accomplish this we can use the following matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$:

1 ☐ $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\boldsymbol{C} = \begin{bmatrix} 0.3 & 0.4 \\ 0.8 & 0.2 \end{bmatrix}$

2 ☐ $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{C} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

3 ☐ $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{C} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

4 ☐ $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\boldsymbol{C} = \begin{bmatrix} 0.3 & 0.4 \\ 0.8 & 0.2 \end{bmatrix}$

5 ☐ $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{C} = \begin{bmatrix} 0.3 & 0.4 \\ 0.8 & 0.2 \end{bmatrix}$

6 ☐ Don't know.

Problem 9.

Consider

$$\boldsymbol{X} \in N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known and $\boldsymbol{\Sigma}$ has full rank.

Question 9.1.

$P\left\{(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) > 10\right\}$ is

1 ☐ between 0.25 and 0.5

2 ☐ between 0.1 and 0.25

3 ☐ between 0.05 and 0.1

4 ☐ between 0.025 and 0.05

5 ☐ between 0.01 and 0.025

6 ☐ Don't know.

Problem 10.

Consider the following (incomplete) analysis of variance table:

Source of variation	Sum of squares	Degrees of freedom
Model - hypothesis		4
Observations - model	20	
Observations - hypothesis	24	9

Question 10.1.

The usual test statistic for the hypothesis is:

1 ☐ $\frac{4}{9}$

2 ☐ $\frac{4}{5}$

3 ☐ $\frac{24}{9}$

4 ☐ $\frac{1}{4}$

5 ☐ $\frac{1}{5}$

6 ☐ Don't know.

Problem 11.

Assume:

$\boldsymbol{X}_{1i} \in N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $i = 1, \ldots, n_1$

$\boldsymbol{X}_{2i} \in N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, $i = 1, \ldots, n_2$

$\hat{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{X}}_1 = \sum_{i=1}^{n_1} \boldsymbol{X}_{1i}$

$\hat{\boldsymbol{\mu}}_2 = \overline{\boldsymbol{X}}_2 = \sum_{i=1}^{n_2} \boldsymbol{X}_{2i}$

$\boldsymbol{W}_1 = \sum_{i=1}^{n_1} (\boldsymbol{X}_{1i} - \overline{\boldsymbol{X}}_1)(\boldsymbol{X}_{1i} - \overline{\boldsymbol{X}}_1)'$

$\boldsymbol{W}_2 = \sum_{i=1}^{n_2} (\boldsymbol{X}_{2i} - \overline{\boldsymbol{X}}_2)(\boldsymbol{X}_{2i} - \overline{\boldsymbol{X}}_2)'$

The problem continues on the next page

Question 11.1.

In this question we **only** consider data $\boldsymbol{X}_{1i}$.

A test for the hypothesis $\boldsymbol{\mu}_1 = \mathbf{0}$ has the usual test statistic:

1 ☐ $\frac{n_1-p}{p} n_1 \overline{\boldsymbol{X}}_1' \boldsymbol{W}_1^{-1} \overline{\boldsymbol{X}}_1$

2 ☐ $\frac{n_1-p}{p} n_1 \overline{\boldsymbol{X}}_1' \boldsymbol{\Sigma}^{-1} \overline{\boldsymbol{X}}_1$

3 ☐ $\frac{n_1-p}{p} n_1 \boldsymbol{\mu}_1' \boldsymbol{W}_1^{-1} \boldsymbol{\mu}_1$

4 ☐ $\frac{n_1-p}{p} n_1 \boldsymbol{\mu}_1' \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1$

5 ☐ $\frac{n_1-p}{p} n_1 \overline{\boldsymbol{X}}_1' (\mu_1 \mu_1')^{-1} \overline{\boldsymbol{X}}_1$

6 ☐ Don't know.

Question 11.2.

The usual unbiased *(Danish: centrale)* estimate for $\boldsymbol{\Sigma}$ is:

1 ☐ $\frac{1}{n_1-1} \boldsymbol{W}_1 + \frac{1}{n_2-1} \boldsymbol{W}_2$

2 ☐ $\boldsymbol{W}_1 + \boldsymbol{W}_2$

3 ☐ $(n_1 - 1)\boldsymbol{W}_1 + (n_2 - 1)\boldsymbol{W}_2$

4 ☐ $(n_1 + n_2 - 2)(\boldsymbol{W}_1 + \boldsymbol{W}_2)$

5 ☐ $\frac{1}{n_1+n_2-2}(\boldsymbol{W}_1 + \boldsymbol{W}_2)$

6 ☐ Don't know.

Question 11.3.

The usual test statistic for the hypothesis $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ should be compared to a suitable percentile in an:

1 ☐ $\chi^2(n_1 + n_2 - p - 1)$

2 ☐ $\chi^2(2)$

3 ☐ $F(n_1 + n_2, n_1 + n_2 - p - 1)$

4 ☐ $F(1, n_1 + n_2 - 2)$

5 ☐ $F(p, n_1 + n_2 - p - 1)$

6 ☐ Don't know.

```
/* encla.sas   Crtd: 21.10.01 22:26 by BKE. Updt: 11/13/01 23:31 */
/* Purpose: enclosure A for exam in Multivariate Statistics 02409 */
/*          on 10 December 2001. */

title1 'Enclosure A - Landsat data from Ymer OE';

proc print data=stat2.encla;
var rock tm1 tm3 tm7;
title2 'Print of observations (only first 20 and last 20 shown)';
run;

proc discrim data=stat2.encla wcov wcorr pcov pcorr pool=yes;
class rock;
var tm1 tm3 tm7;
title1 'Output from proc discrim';
run;
```



```
                 Enclosure A - Landsat data from Ymer OE                  1
             Print of first 20 and last 20 observations only

                     Obs    rock     tm1    tm3    tm7

                       1    Bed 5     74     43     32
                       2    Bed 5     74     43     32
                       3    Bed 5     71     46     32
                       4    Bed 5     73     46     33
                       5    Bed 5     69     45     32
                       6    Bed 5     75     47     34
                       7    Bed 5     74     43     34
                       8    Bed 5     79     48     37
                       9    Bed 5     70     41     32
                      10    Bed 5     71     44     34
                      11    Bed 5     55     21     21
                      12    Bed 5     66     34     27
                      13    Bed 5     71     39     27
                      14    Bed 5     58     25     25
                      15    Bed 5     70     38     26
                      16    Bed 5     70     40     31
                      17    Bed 5     70     40     31
                      18    Bed 5     60     29     26
                      19    Bed 5     72     43     29
                      20    Bed 5     71     43     30


          * * * observations 21-989 removed from output * * *


                     990    Bed 15    118     66     77
                     991    Bed 15    116     65     78
                     992    Bed 15    111     64     75
                     993    Bed 15    111     65     72
                     994    Bed 15    109     62     69
                     995    Bed 15    111     73     61
                     996    Bed 15    124     70     79
                     997    Bed 15    117     67     80
                     998    Bed 15    105     60     66
                     999    Bed 15    110     65     67
                    1000    Bed 15    118     69     74
                    1001    Bed 15    118     68     75
                    1002    Bed 15    111     64     75
                    1003    Bed 15    115     66     72
                    1004    Bed 15    117     67     80
                    1005    Bed 15    112     67     75
                    1006    Bed 15    104     53     70
                    1007    Bed 15    113     67     63
                    1008    Bed 15    111     60     65
                    1009    Bed 15    118     66     69
```

```
                         Output from proc discrim                          18

                          The DISCRIM Procedure

            Observations    1009          DF Total               1008
            Variables          3          DF Within Classes      1006
            Classes            3          DF Between Classes        2


                         Class Level Information

             Variable                                                Prior
   rock      Name          Frequency       Weight    Proportion    Probability

   Bed 15    Bed_15              389     389.0000      0.385530       0.333333
   Bed 5     Bed_5               493     493.0000      0.488603       0.333333
   Bed 8     Bed_8               127     127.0000      0.125867       0.333333

/><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>></
                         Output from proc discrim                          19

                          The DISCRIM Procedure
                     Within-Class Covariance Matrices

                       rock = Bed 15,     DF = 388

         Variable                 tm1                tm3                tm7

         tm1              108.5275488         78.1582501         70.8226022
         tm3               78.1582501         63.3794291         51.8532319
         tm7               70.8226022         51.8532319         68.0409986

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


                        rock = Bed 5,     DF = 492

         Variable                 tm1                tm3                tm7

         tm1              23.13996768         9.02407691         8.04561421
         tm3               9.02407691        13.86331404         7.12912894
         tm7               8.04561421         7.12912894         8.61168555

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


                        rock = Bed 8,     DF = 126

         Variable                 tm1                tm3                tm7

         tm1              59.30221222        42.37726534        26.82977128
         tm3              42.37726534        57.55905512        29.00593676
         tm7              26.82977128        29.00593676        31.01474816

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



```
                         Output from proc discrim                          20

                          The DISCRIM Procedure

              Pooled Within-Class Covariance Matrix,     DF = ####

         Variable                 tm1                tm3                tm7

         tm1              60.60201967        39.86558877        34.61050004
         tm3              39.86558877        38.43380712        27.11862173
         tm7              34.61050004        27.11862173        34.33868292

/><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>><<>></
                         Output from proc discrim                          21

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

                              rock = Bed 15

               Variable           tm1           tm3           tm7

               tm1            1.00000       0.94239       0.82417
                                             <.0001        <.0001

               tm3            0.94239       1.00000       0.78962
                               <.0001                      <.0001

               tm7            0.82417       0.78962       1.00000
                               <.0001        <.0001

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


                               rock = Bed 5

               Variable           tm1           tm3           tm7

               tm1            1.00000       0.50383       0.56995
                                             <.0001        <.0001

               tm3            0.50383       1.00000       0.65247
                               <.0001                      <.0001

               tm7            0.56995       0.65247       1.00000
                               <.0001        <.0001

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


                               rock = Bed 8

               Variable           tm1           tm3           tm7

               tm1            1.00000       0.72534       0.62560
                                             <.0001        <.0001

               tm3            0.72534       1.00000       0.68651
                               <.0001                      <.0001

               tm7            0.62560       0.68651       1.00000
                               <.0001        <.0001

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

Output from proc discrim 22

The DISCRIM Procedure

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

Variable	tm1	tm3	tm7
tm1	1.00000	0.82603 <.0001	0.75870 <.0001
tm3	0.82603 <.0001	1.00000	0.74648 <.0001
tm7	0.75870 <.0001	0.74648 <.0001	1.00000

Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	9.17313

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

Output from proc discrim 23

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 rock

From rock	Bed 15	Bed 5	Bed 8
Bed 15	0	41.87313	14.99329
Bed 5	41.87313	0	58.33204
Bed 8	14.99329	58.33204	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 rock

Variable	Bed 15	Bed 5	Bed 8
Constant	−98.97062	−52.78887	−98.68610
tm1	2.11822	1.77772	1.31253
tm3	−1.03239	−0.08501	−0.12105
tm7	0.54367	−0.82218	0.90957

Output from proc discrim 24

The DISCRIM Procedure
Classification Summary for Calibration Data: STAT2.ENCLA
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 rock

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, \sum_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into rock

From rock	Bed 15	Bed 5	Bed 8	Total
Bed 15	386 99.23	0 0.00	3 0.77	389 100.00
Bed 5	0 0.00	493 100.00	0 0.00	493 100.00
Bed 8	13 10.24	0 0.00	114 89.76	127 100.00
Total	399 39.54	493 48.86	117 11.60	1009 100.00
Priors	0.33333	0.33333	0.33333	

Error Count Estimates for rock

	Bed 15	Bed 5	Bed 8	Total
Rate	0.0077	0.0000	0.1024	0.0367
Priors	0.3333	0.3333	0.3333	

```
/* enclb.sas   Crtd: 11/09/01 10:19 by BKE. Updt: 11/13/01 23:40 */
/* Purpose: enclosure B for exam in Multivariate Statistics 02409 */
/*          on 10 December 2001. */

title1 'Enclosure B - Incinerator data';

proc print data=stat2.enclb;
title1 'Print of observations';
run;

proc princomp cov data=stat2.enclb;
var plastics paper garbage water;
title1 'Output from proc princomp';
run;

proc reg data=stat2.enclb;
model energy_content=combustible water/influence covb;
title1 'Output from proc reg';
run;
```


Print of observations 1

Obs	Plastics	Paper	Garbage	Water	Energy_ content	Combustible
1	18.69	15.65	45.01	58.210	947	20.714
2	19.43	23.51	39.69	46.310	1407	22.272
3	19.24	24.23	43.16	46.630	1452	22.630
4	22.64	22.20	35.76	45.850	1553	23.864
5	16.54	23.56	41.20	55.140	989	20.410
6	21.44	23.65	35.56	54.240	1162	23.294
7	19.53	24.45	40.18	47.200	1466	22.579
8	23.97	19.39	44.11	43.820	1656	25.068
9	21.45	23.84	35.41	51.010	1254	23.324
10	20.34	26.50	34.21	49.060	1336	22.959
11	17.03	23.46	32.45	53.230	1097	19.858
12	21.03	26.99	38.19	51.780	1266	23.938
13	20.49	19.87	41.35	46.690	1401	22.452
14	20.45	23.03	43.59	53.570	1223	23.280
15	18.81	22.62	42.20	52.980	1216	21.911
16	18.28	21.87	41.50	47.444	1334	21.320
17	21.41	20.47	41.20	54.680	1155	23.201
18	25.11	22.59	37.02	48.740	1453	25.797
19	21.04	26.27	38.66	53.220	1278	23.848
20	17.99	28.22	44.18	53.370	1153	22.655
21	18.73	29.39	34.77	51.060	1225	22.466
22	18.49	26.58	37.55	50.660	1237	22.014
23	22.08	24.88	37.07	50.720	1327	24.139
24	14.28	26.27	35.80	48.240	1229	18.830
25	17.74	23.61	37.36	49.920	1205	20.876
26	20.54	26.58	35.40	53.580	1221	23.234
27	18.25	13.77	51.32	51.380	1138	20.661
28	19.09	25.62	39.54	50.130	1295	22.441
29	21.25	20.63	40.72	48.670	1391	23.073
30	21.62	22.71	36.22	48.190	1372	23.298

Output from proc princomp 2

The PRINCOMP Procedure

Observations	30
Variables	4

Simple Statistics

	Plastics	Paper	Garbage	Water
Mean	19.89933333	23.41366667	39.34600000	50.52413333
StD	2.22729608	3.37747047	4.04942107	3.30405635

Covariance Matrix

	Plastics	Paper	Garbage	Water
Plastics	4.96084782	−1.12502506	−0.80857517	−1.90337577
Paper	−1.12502506	11.40730678	−8.62165379	−0.05936947
Garbage	−0.80857517	−8.62165379	16.39781103	0.95401710
Water	−1.90337577	−0.05936947	0.95401710	10.91678840

Total Variance 43.682754028

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	22.9317107	11.3843145	0.5250	0.5250
2	11.5473962	5.7582089	0.2643	0.7893
3	5.7891873	2.3747274	0.1325	0.9218
4	3.4144599		0.0782	1.0000

Eigenvectors

	Prin1	Prin2	Prin3	Prin4
Plastics	−.005669	−.301272	−.583156	0.754409
Paper	−.597537	0.149511	0.595600	0.515613
Garbage	0.798993	0.030402	0.466246	0.378552
Water	0.067293	0.941253	−.296320	0.147340

Output from proc reg 3

The REG Procedure
Model: MODEL1
Dependent Variable: Energy_content

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	664746	332373	359.48	<.0001
Error	27	24964	924.58557		
Corrected Total	29	689710			

Root MSE	30.40700	R−Square	0.9638
Dependent Mean	1281.26667	Adj R−Sq	0.9611
Coeff Var	2.37320		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2242.49136	139.52549	16.07	<.0001
Combustible	1	41.09413	3.86421	10.63	<.0001
Water	1	−37.36370	1.76519	−21.17	<.0001

Covariance of Estimates

Variable	Intercept	Combustible	Water
Intercept	19467.362005	−422.9756113	−195.9413675
Combustible	−422.9756113	14.932090307	1.7081690743
Water	−195.9413675	1.7081690743	3.115887334

```
                          Output from proc reg                          4

                           The REG Procedure
                             Model: MODEL1
                   Dependent Variable: Energy_content

                           Output Statistics

                                         Hat Diag         Cov
       Obs      Residual     RStudent           H       Ratio       DFFITS

         1       28.2258       1.0636      0.2346      1.2877       0.5889
         2      −20.4269      −0.7009      0.0987      1.1746      −0.2319
         3       21.8178       0.7432      0.0834      1.1470       0.2241
         4       42.9640       1.5370      0.1122      0.9718       0.5465
         5      −31.9881      −1.1424      0.1424      1.1274      −0.4656
         6      −11.1309      −0.3795      0.0991      1.2227      −0.1259
         7       59.2109       2.1508      0.0702      0.7369       0.5909
         8       20.6383       0.7650      0.2250      1.3517       0.4122
         9      −41.0485      −1.4064      0.0453      0.9414      −0.3063
        10      −16.9084      −0.5606      0.0411      1.1265      −0.1160
        11       27.3312       0.9728      0.1479      1.1806       0.4053
        12      −25.5103      −0.8690      0.0764      1.1127      −0.2499
        13      −19.6256      −0.6675      0.0844      1.1623      −0.2026
        14       25.4107       0.8680      0.0815      1.1191       0.2586
        15       52.6240       1.8589      0.0544      0.8145       0.4460
        16      −11.9348      −0.4081      0.1036      1.2256      −0.1387
        17        2.1309       0.0728      0.1085      1.2554       0.0254
        18      −28.4899      −1.0449      0.1932      1.2270      −0.5114
        19       43.9919       1.5637      0.0981      0.9483       0.5158
        20      −26.3782      −0.8923      0.0620      1.0906      −0.2293
        21      −32.9216      −1.1063      0.0342      1.0101      −0.2083
        22      −17.2925      −0.5725      0.0377      1.1208      −0.1133
        23      −12.3757      −0.4168      0.0756      1.1874      −0.1191
        24       15.1311       0.5898      0.3054      1.5492       0.3911
        25      −30.1765      −1.0381      0.0834      1.0816      −0.3131
        26       25.6747       0.8766      0.0802      1.1156       0.2588
        27      −33.7903      −1.1712      0.0873      1.0516      −0.3622
        28        3.3576       0.1103      0.0342      1.1579       0.0208
        29       18.8351       0.6270      0.0458      1.1220       0.1373
        30      −27.3457      −0.9222      0.0543      1.0752      −0.2210

                           Output Statistics

                    −−−−−−−−−−−−−−−DFBETAS−−−−−−−−−−−−−−
                Obs    Intercept    Combustible        Water

                  1      −0.2006        −0.1473       0.4715
                  2      −0.1693         0.0710       0.1871
                  3       0.1388        −0.0357      −0.1734
                  4       0.1498         0.1622      −0.3742
                  5      −0.0088         0.2522      −0.2467
                  6       0.0955        −0.0596      −0.0958
                  7       0.3515        −0.0987      −0.4281
                  8       0.0570         0.1937      −0.2685
                  9       0.1334        −0.1523      −0.0762
                 10      −0.0193        −0.0178       0.0411
                 11       0.1585        −0.3186       0.0754
                 12       0.1713        −0.1764      −0.1060
                 13      −0.1352         0.0473       0.1574
                 14      −0.1870         0.1245       0.1813
                 15      −0.0818        −0.0862       0.2339
                 16      −0.1172         0.0865       0.0939
                 17      −0.0193         0.0111       0.0202
                 18       0.2727        −0.4503       0.0002
```



```
                          Output from proc reg                          5

                           The REG Procedure
                             Model: MODEL1
                   Dependent Variable: Energy_content

                           Output Statistics

                    −−−−−−−−−−−−−−−DFBETAS−−−−−−−−−−−−−−
                Obs    Intercept    Combustible        Water

                 19      −0.4066         0.3368       0.3259
                 20       0.1243        −0.0508      −0.1554
                 21       0.0106         0.0028      −0.0321
                 22      −0.0316         0.0384       0.0053
                 23       0.0696        −0.0889      −0.0269
                 24       0.3420        −0.3578      −0.1776
                 25      −0.2187         0.2398       0.0957
                 26      −0.1850         0.1204       0.1821
                 27      −0.1909         0.2785       0.0127
                 28       0.0040        −0.0022      −0.0029
                 29       0.0260         0.0256      −0.0583
                 30      −0.0381        −0.0583       0.1058


             Sum of Residuals                           0
             Sum of Squared Residuals               24964
             Predicted Residual SS (PRESS)          30574
```