

EXAM IN MULTIVARIATE METHOD September 30 2020

Time: 6 hours

The exam is for individual solving. It is an open-book exam, but you are not allowed to use the help of other students, friends, family, or similar. In case you need clarification, the teacher is available at Zoom. Time for zoom meetings are 9:00-9:30 and 12:00-12:30.

Join Zoom Meeting from

<https://stockholmuniiversity.zoom.us/j/9255236581>

Meeting ID: 925 523 6581

The exam consists of five questions. To score maximum points on a question solutions need to be clear, detailed and well-motivated.

Question. 1 (2+6+2+3+3=16 Points)

For a data set with observations on two variables x_1 and x_2 the sample correlation matrix was found to be

$$S = \begin{bmatrix} 60.41 & 4.55 \\ 4.55 & 1.30 \end{bmatrix}$$

- Find the correlation matrix (R).
- Using matrix R, construct two principal components that are orthogonal to each other.
- What proportion of variance is accounted by these principal components?
- Compute the loadings of the variables.
- Using the constructed principal components in part (b) show that $\text{var}(\text{PC1}) = \lambda_1$ and $\text{var}(\text{PC2}) = \lambda_2$, where λ_1 and λ_2 are eigenvalues.

Question. 2 (2+2+2+2+3+2.5+2.5=16 Points)

The sample correlation matrix given below arises from the scores of 210 boys in six school subjects: (1) French, (2) English, (3) History, (4) Arithmetic, (5) Algebra, and (6) Geometry.

$$R = \begin{matrix} & \begin{matrix} \text{French} \\ \text{English} \\ \text{History} \\ \text{Arithmetic} \\ \text{Algebra} \\ \text{Geometry} \end{matrix} & \begin{bmatrix} 1 & & & & & \\ 0.45 & 1 & & & & \\ 0.41 & 0.36 & 1 & & & \\ 0.30 & 0.34 & 0.17 & 1 & & \\ 0.29 & 0.32 & 0.22 & 0.60 & 1 & \\ 0.24 & 0.35 & 0.19 & 0.48 & 0.47 & 1 \end{bmatrix} \end{matrix}$$

Based on correlation matrix R the eigenvalues and eigenvectors are given below

	1	2	3	4	5	6
eigenvalues	2.7287	----	0.6153	0.6028	0.5225	0.4019

Eigenvectors						
	V1	V2	V3	V4	V5	V6
French	0.3902	0.4440	0.1974	-0.5319	-0.5694	0.0672
English	0.4178	0.2859	0.6368	0.2142	0.5372	0.0612
History	0.3219	0.5685	-0.6643	0.3142	0.1243	-0.1333
Arithmetic	0.4461	-0.3923	-0.0882	-0.3062	0.1704	-0.7187
Algebra	0.4460	-0.3627	-0.3063	-0.2539	0.2335	0.6758
Geometry	0.4143	-0.3354	0.1127	0.6437	-0.5368	0.0250

The following estimated pattern loadings are extracted by the principal component factoring procedure:

Variable	F1	F2
French	0.65	----
English	----	0.30
History	----	-----
Arithmetic	----	-0.42
Algebra	0.74	----
Geometry	----	----

Based on these reported results obtain:

- The 2nd missing eigenvalue.
- The values of estimated missing pattern loading.
- The specific variances.
- The proportion of variance explained by each factor.
- The estimated/reproduced correlation matrix.
- The residual correlation matrix.
- RMSR.

Question. 3 (8+2+6=16 Points)

For the following data

Observation	Y_1	Y_2	Brand
1	4	5	A
2	6	2	B
3	5	8	A
4	4	12	B
5	6	10	A
6	8	5	B
7	10	4	B
8	5	6	B
9	14	5	A

- Compute the \mathbf{SSCP}_b , \mathbf{SSCP}_w and \mathbf{SSCP}_t matrices.
- Compute the statistical distance between observations 7 and 9.
- Suppose $n_1=5$ and $n_2=6$ are observations in group-1 and group-2, respectively and

$$\text{Within-group covariance matrix for group-I} = S_1 = \begin{bmatrix} 9.70 & -2.45 \\ -2.45 & 3.70 \end{bmatrix}$$

$$\text{Within-group covariance matrix for group-II} = S_2 = \begin{bmatrix} 11 & -4.0 \\ -4.0 & 5.4 \end{bmatrix}$$

$$\bar{X}_1 = \begin{bmatrix} 4.1 \\ 5.8 \end{bmatrix} \text{ and } \bar{X}_2 = \begin{bmatrix} 8 \\ 5 \end{bmatrix}$$

Calculate Fisher's linear discriminant function for this data set.

Question. 4 (4+4+4+4=16 Points)

Observations on two variables were made for five subjects according to the following table.

Subject	Variable-1	Variable-2
1	4	4
2	8	5
3	5	3
4	2	6
5	6	8

- a) Construct a similarity matrix containing squared Euclidean distances
- b) Use the similarity matrix in part (a) and perform a cluster analysis with the following method
 - I. Centroid method.
 - II. Nearest neighbor method.
 - III. Average linkage method.

Question. 5 (2+3+3+4+4=16 Points)

Suppose we have data of 462 persons on following variables

Y=Response variable, 1= Person has coronary heart disease and 0= Person not having coronary heart disease

X1= Person age in years

X2= Person family history of heart disease, a factor with levels Absent=0 and Present=1

Model-1: A logistic regression model was fitted to the data using X1 and X2 as explanatory variables and the output is given below

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.7585	0.4371	-8.600	< 2e-16 ***
x1	0.0597	0.0088	6.787	1.14e-11 ***
x2(Present)	0.9339	0.2163	4.318	1.58e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 596.11 on 461 degrees of freedom

Residual deviance: 506.66 on 459 degrees of freedom

AIC: 512.66

Model-2: Another logistic regression model was fitted to the data using X1 as explanatory variables and the output is given below

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.5217	0.4160	-8.465	< 2e-16 ***
age	0.0641	0.0085	7.513	5.76e-14***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 596.11 on 461 degrees of freedom

Residual deviance: 525.56 on 460 degrees of freedom

AIC: 529.56

- a) Interpret the parameter $\hat{\beta}_2$ in model-1.
- b) What are the odds of an individual having coronary heart disease if he is 20 years old and has heart disease in his family history?
- c) What is the classification of an individual who is 40 years old and do not have heart disease in his family history?
- d) An individual has heart disease in his family history. What is the minimum age that an individual should have before it is classified as a coronary heart disease patient?
- e) Formulate the null and alternative hypothesis and perform testing of hypothesis using deviance statistic to compare the fitted model-1 with the model-2. Use $\alpha = 0.05$.