Badji Mokhtar University, Annaba **Exam in Data Analysis**
Faculty of Economic, Commercial and Management Sciences
Department of Financial and Accounting Science
Duration: 1:30h

Exercise 01 (matrix algebra)

Let the three matrices be:

$$A = \begin{pmatrix} 4 & 2 \\ 6 & 3 \end{pmatrix} \quad , \quad B = \begin{pmatrix} -2 & -3 \\ -7 & 0 \end{pmatrix} \quad , \quad C = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$$

1- Compute: $X = A + 3B - 2C$
2- If X is invertible, Calculate X^{-1}.

Exercise 02 (The Principal Components analysis)

The following table represents two car brands and the ratings given by four individuals to these brands (ratings range from 0 to 10, where a score of 10 indicates a very good rating for the brand, and 0 the opposite).

Individual	BMW	Citroën
A	9	4
B	4	8
C	8	6
D	5	7

1. Calculate the statistics of the above data variables.
2. Calculate the covariance matrix.
3. Calculate the correlation matrix, and then comment on it.
4. Calculate the variances (eigenvalues) of the principal components for this problem statement.
5. What is the percentage of variance explained by each principal component?
6. Find the normalized eigenvector corresponding to the first principal component.

Exercise 03 (Correspondence Analysis)

A survey was conducted among cinema-goers about a movie they watched, and they were asked to specify their age groups.

Using the Correspondence Analysis by SPSS get the following results.

Age Group	Excellent	Good	Moderate	Weak
16–24	41	48	49	69
25–34	22	14	45	148
35–44	28	12	65	170
45–54	29	12	57	159
55–64	26	6	26	122
65–74	23	5	21	106
75+	14	1	7	40

Row Profiles

Age	Evaluation				
	Excellent	Good	Moderate	Weak	Active Margin
16-24	.198	.232	.237	.333	1.000
25-34	.096	.061	.197	.646	1.000
35-44	.102	.044	.236	.618	1.000
45-54	.113	.047	.222	.619	1.000
55-64	.144	.033	.144	.678	1.000
65-74	.148	.032	.135	.684	1.000
75+	.226	.016	.113	.645	1.000
Mass	.134	.072	.198	.596	

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	.305	.093			.859	.859	.033	-.019-
2	.115	.013			.123	.982	.026	
3	.044	.002			.018	1.000		
Total		.108	147.309	.000[a]	1.000	1.000		

a. 18 degrees of freedom

Overview Row Points[a]

Age	Mass	Score in Dimension		Inertia	Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
		1	2		1	2	1	2	Total
16-24	.152	-1.294-	.091	.078	.834	.011	.998	.002	1.000
25-34	.168	.158	-.201-	.003	.014	.059	.457	.281	.739
35-44	.201	.194	-.350-	.005	.025	.215	.423	.522	.945
45-54	.188	.175	-.221-	.003	.019	.080	.586	.353	.939
55-64	.132	.332	.332	.006	.048	.126	.714	.271	.985
65-74	.114	.346	.398	.006	.045	.156	.654	.327	.981
75+	.045	.326	.947	.007	.016	.354	.221	.706	.927
Active Total	1.000			.108	1.000	1.000			

a. Symmetrical normalization

1. What is the nature or classification of these values?
2. What are the key purposes of applying Correspondence Analysis (CA) to this dataset?
3. From the row profiles table, find the average row vector (L_m), and the theoretical diagonal matrix for the average row.
4. Compute the distance between: (25-34) and L_m, (+75) and L_m.
5. Extract from the table of Overview Row Points the inertia values of row, and deduce the percentage contribution of each age group to the construction of the first axis.
6. Is there a correlation between the rows and columns? Justify your answer.