Exam solution

> **Solution** n°1

1. The matrix representing the linear transformation T is:

   $T = \begin{pmatrix} 4x + y \\ 2x + 3y \end{pmatrix}$. ***(1pts)***

2. The eigenvalues of T:

To find the eigenvalues λ, we solve the characteristic equation:

$$\det(A - \lambda I) = 0.$$

Where I is the identity matrix, and λ represents the eigenvalue.

First, compute $A - \lambda I$:

$$A - \lambda I = \begin{pmatrix} 4 - \lambda & 1 \\ 2 & 3 - \lambda \end{pmatrix}.$$

Now, find the determinant of $A - \lambda I$:

$$\det(A - \lambda I) = (4 - \lambda)(3 - \lambda) - 2 \cdot 1 = 0.$$

Expanding the determinant:

$$(4 - \lambda)(3 - \lambda) - 2 = 12 - 7\lambda + \lambda^2 - 2 = \lambda^2 - 7\lambda + 10 = 0.$$

Solve the quadratic equation:

$$\lambda^2 - 7\lambda + 10 = 0.$$

The roots of this quadratic equation are:

$$\lambda = \frac{7 \pm \sqrt{49 - 40}}{2} = \frac{7 \pm \sqrt{9}}{2} = \frac{7 \pm 3}{2}.$$

Thus, the eigenvalues are:

$$\lambda_1 = \frac{7 + 3}{2} = 5, \quad \lambda_2 = \frac{7 - 3}{2} = 2.$$ ***(1.5pts)***

Find the eigenvectors associated with each eigenvalue: ***(1.5pts)***

For each eigenvalue λ, we solve the equation $(A - \lambda I)\mathbf{v} = 0$ to find the corresponding eigenvector $\mathbf{v}$.

1. Eigenvalue $\lambda_1 = 5$:

We need to solve:

$$(A - 5I)\mathbf{v} = 0.$$

First, compute $A - 5I$:

$$A - 5I = \begin{pmatrix} 4-5 & 1 \\ 2 & 3-5 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}.$$

Now, solve $\begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$:

$$-x + y = 0 \quad \text{and} \quad 2x - 2y = 0$$

Both equations give the same condition: $x = y$.

Thus, the eigenvector corresponding to $\lambda_1 = 5$ is:

$$\mathbf{v_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Eigenvalue $\lambda_2 = 2$:

We need to solve:

$$(A - 2I)\mathbf{v} = 0.$$

First, compute $A - 2I$:

$$A - 2I = \begin{pmatrix} 4-2 & 1 \\ 2 & 3-2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}.$$

Now, solve $\begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$:

$$2x + y = 0 \quad \text{and} \quad 2x + y = 0.$$

Both equations give the same condition: $y = -2x$.

Thus, the eigenvector corresponding to $\lambda_2 = 2$ is:

$$\mathbf{v_2} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

4. Calculate the image of the vector $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ under T:

We compute $T(v) = A \cdot v$, where $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$:

$$T(v) = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \cdot 1 + 1 \cdot 0 \\ 2 \cdot 1 + 3 \cdot 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

Thus, the image of v under T is:

$T(v) = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$. ***(1pts)***

> **Solution** n°2

Exercice p. 4

1. Nature or Classification of Values: The values represent a contingency table showing the joint distribution of two ***categorical variables***: gender (rows) and profession (columns), which are ***qualitatives***. ***(0.5pt)***

2. Purpose of Correspondence Analysis (CA): ***(1pt)***

- Visualization and identify the datasets relationship (association between gender and profession).
- Reduce dimensionality while preserving most of the variance (inertia).

3. Theoretical Frequencies (Independence Assumption): ***(1 Pts)***

$$E_{ij} = \frac{R_i \cdot C_j}{N}$$

	Profession 01	Profession 02	Σ
Men	19.44	15.56	35
Women	30.56	24.44	55
Σ	50	40	90

4. χ-Square Statistic: ***(1 Pts)***

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 0.0594.$$

Degrees of Freedom: $df = 1$,

Critical Value: 3.841.

Since $\chi^2 = 0.0594 < 3.841$, we ***fail to reject the null hypothesis***. There is ***no significant association*** between gender and profession. ***(0.5 Pts)***

5. Inertia:

$$\text{Inertia} = \frac{\chi^2}{N} = \frac{0.0594}{90} = 0.00066$$ ***(0.5 Pts)***

Conclusion: The association contributes very little, or even nothing, to the variance. ***(0.5 Pts)***

6. Row-Profile Matrix L: ***(1 Pts)***

$$L = \begin{bmatrix} 0.571 & 0.429 \\ 0.545 & 0.455 \end{bmatrix}$$

7. Column-Profile Matrix C: (1 Pts)

$$C = \begin{bmatrix} 0.4 & 0.375 \\ 0.6 & 0.625 \end{bmatrix}.$$

Comment : ***(0.5 Pts)***

The row profiles show the proportion of individuals within each profession for each gender, while the column profiles show the proportion of individuals within each gender for each profession. (Both profiles help identify the distribution patterns and potential associations between gender and profession.)

8. Diagonal Matrix D_r: ***(0.5 Pts)***

$$D_r = \text{diag}\,(0.389, 0.611)$$

$$D_r = \begin{pmatrix} 0.389 & 0 \\ 0 & 0.611 \end{pmatrix}.$$

- ***Bonus answer***

Total Inertia for Row Profiles: ***(+2 Pts)***

$$\text{Total Inertia} = \text{trace}(\Sigma_r) = 0.014 + 0.039 = 0.053$$

A low inertia value indicates minimal differences in row profiles.

> **Solution** n°3

Exercice p. 5

1. Statistical Analysis of Variables x (Salary) and y (Consumption Expenses):

Means: ***(1pt)***

$$\bar{x} = \frac{\sum x_i}{n} = \frac{35 + 8 + 12 + 25}{4} = 20.$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{10.7 + 2.5 + 3.8 + 7.0}{4} = 6.$$

Variances: ***(1pts)***

The variance formulas are:

$$\text{Var}(x) = \frac{\sum(x_i - \bar{x})^2}{n}, \quad \text{Var}(y) = \frac{\sum(y_i - \bar{y})^2}{n}$$

$$\sum(x_i - \bar{x})^2 = (35-20)^2 + (8-20)^2 + (12-20)^2 + (25-20)^2 = 15^2 + (-12)^2 + (-8)^2 + 5^2 = 225 + 144 + 64 + 25 = 458.$$

For x:

For y:

$$\sum(y_i - \bar{y})^2 = (10.7-6)^2 + (2.5-6)^2 + (3.8-6)^2 + (7-6)^2 = 4.7^2 + (-3.5)^2 + (-2.2)^2 + 1^2 = 22.09 + 12.25 + 4.84 + 1 = 40.18$$

$$\text{Var}(x) = \frac{458}{4} = 114.5$$

$$\text{Var}(y) = \frac{40.18}{4} = 10.045$$

The standard deviations of x and y are: ***(0.5Pts)***

$$\sigma_x = \sqrt{114.5} \approx 10.7, \quad \sigma_y = \sqrt{10.045} \approx 3.17.$$

2- Calcul of covariance matrix:(1.5pts)

Standardization of Data :

$$Z = X - \mu$$

Now we substract the means from each corresponding value in X:

$$Z = X - \mu = \begin{bmatrix} 35-20 & 10.7-6 \\ 8-20 & 2.5-6 \\ 12-20 & 3.8-6 \\ 25-20 & 7-6 \end{bmatrix} \qquad Z = \begin{bmatrix} 15 & 4.7 \\ -12 & -3.5 \\ -8 & -2.2 \\ 5 & 1 \end{bmatrix}$$

$$\Sigma = \frac{1}{4}(Z^T Z) = \begin{bmatrix} \frac{458}{4} & \frac{135.1}{4} \\ \frac{135.1}{4} & \frac{40.18}{4} \end{bmatrix} = \begin{bmatrix} 114.5 & 33.775 \\ 33.775 & 10.045 \end{bmatrix}$$

.

The diagonal off represent $Cov(x, y) = 33.775$

3- Calcul of the correlation matrix :

We know that the correlation matrix is:

$$R = \begin{bmatrix} 1 & \rho(x,y) \\ \rho(x,y) & 1 \end{bmatrix}$$

Where: $\rho(x,y) = \dfrac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_y} = \dfrac{33.775}{\sqrt{114.5}\,\sqrt{10.045}} = 0.9959.$

Thus from above we get **(0.5PTS)**

$$R = \begin{bmatrix} 1 & 0.9959 \\ 0.9959 & 1 \end{bmatrix}$$

Conclusion : indicates that the variables x (salary) and y (consumption expenses) are ***positively correlated***, with a correlation value of **0.9959**, which is very close to 1. This suggests a ***very strong positive*** linear relationship between the two variables. **(0.5PTS)**

4- Eigenvalues and Principal Component Analysis: ***(0.75pts)***

To find the eigenvalues, we solve the characteristic equation:

$$\det\left(\begin{bmatrix} 1-\lambda & 0.9959 \\ 0.9959 & 1-\lambda \end{bmatrix}\right) = 0.$$

The determinant is :

$$\det(R - \lambda I) = (1-\lambda)^2 - 0.9959^2 = (0.0041 - \lambda)(1.9959 - \lambda).$$

Thus:

$$\lambda_1 = 1.995, \quad \lambda_2 = 0.004$$

-Calculate the percentage of variance explained by each principal component (0.5pts)

To calculate the percentage of variance explained by each principal component, we use the following formula:

$$\text{Percentage of variance explained by } \lambda_i = \frac{\lambda_i}{\sum \lambda} \times 100$$

First, calculate the total variance:

$$\sum \lambda = \lambda_1 + \lambda_2 = 1.995 + 0.004 = 1.999$$

Then, calculate the percentage of variance explained by each principal component:

1. **For** λ_1:

Percentage of variance explained by the first principal component is approximately

$$T_{\lambda_1} = \frac{1.995}{1.999} \times 100 \approx 99.80\%$$

2. **For** λ_2 :

$$\text{Percentage of variance explained by } \lambda_2 = \frac{0.004}{1.999} \times 100 \approx 0.20\%$$

6-Find the normalized eigenvector corresponding to the first principal component (0.75pts)

This lead to find the eigenvector corresponding to $\lambda_1 = 1.995$, we solve the equation:

;

This simplifies to:

$$-0.995v_1 + 0.9959v_2 = 0$$

$$0.9959v_1 - 0.995v_2 = 0$$

So, the eigenvector $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is proportional to:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1.0009 \\ 1 \end{bmatrix} v_2$$

4. Normalize the eigenvector.

To normalize the eigenvector, we calculate its magnitude:

$$\| \begin{bmatrix} 1.0009 \\ 1 \end{bmatrix} \| = \sqrt{1.0009^2 + 1^2} \approx \sqrt{1.0018 + 1} = \sqrt{2.0018} \approx 1.414$$

Thus, the normalized eigenvector is:

$$\frac{1}{1.414} \begin{bmatrix} 1.0009 \\ 1 \end{bmatrix} \approx \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$$

This eigenvector indicates the direction of maximum variance in the data and corresponds to the first principal component.