Tarea 1
CA-0411
I CICLO 2025

1. Instrucciones

A continuación se muestran las instrucciones de la tarea:

- La solución a cada tarea se debe subir en el aula virtual, no puedn ser enviadas por correo u otro medio.
- Las tareas se pueden hacer en parejas, pero cada persona deberá entregar la solución.
- Todas las tareas tienen el mismo valor en la nota final del curso.
- Las tareas se pueden entregar tarde, pero cada día de atraso tendrá un rebajo de 10 puntos.

2. Preguntas

1. Pruebe que las componentes principales del A.C.P. general de la nube $\mathcal{N}(\mathbf{X}, \mathbf{M}, \mathbf{D})$ tiene las siguientes propiedades:

   a) Son centradas:
   $$\overline{\mathbf{c}^k} = 0$$

   b) $\mathbf{c}^k$ tiene varianza λ_k:

   c) Cada par de ellas tiene correalción cero:

2. Para un A.C.P. general, sean λ_1 el mayor valor propio de $\mathbf{VM}$ y $\mathbf{u}_1$ un vector propio unitario asociado. Probar que $\lambda_1 = I_{\Delta\mathbf{u}_1^{\perp}}(\mathcal{N})$.

3. Sea $\mathbf{u}_k$ un eje principal del A.C.P. de la nube $\mathcal{N}(\mathbf{X}, \mathbf{M}, \mathbf{D})$ con valor λ_k. Si se define $v_k = \mathbf{M}\mathbf{u}_k$ pruebe que entonces $\mathbf{v}_k$ es vector propio de $\mathbf{MV}$ asociado al mismo valor propio λ_k. Se llama a $\mathbf{v}_k$ un *factor principal*. Pruebe que además los factores principales son ortogonales respecto de los productos internos definidos por las métricas $\mathbf{V}$ y $\mathbf{M^{-1}}$ en el dual $(\mathbb{R}^p)^*$.

4. Pruebe que la matriz de datos $\mathbf{X}$ se puede expresar como $\mathbf{X} = \sum_{j=1}^{p} \mathbf{c}^j \mathbf{u}_j^t$ donde $\mathbf{c}^j$ son las componentes principales y $\mathbf{u}_j$ los ejes principales del A.C.P. general de $\mathcal{N}(\mathbf{X}, \mathbf{M}, \mathbf{D})$. Esta expresión es conocida como *fórmula de reconstrucción de los datos*. (Sugerencia: en vista de que los $\mathbf{u}_j$ forman una base ortonormada de $\mathbb{R}^p$, el $i-$ésimo individuo puede expresarse como $x_i = \sum_{j=1}^{p} c_{ij} \mathbf{u}_j$).

5. Sea $\mathbf{X}$ una tabla de datos asociada a p variables cuantitativas centradas y denótese $\mathbf{W} = \mathbf{X}\mathbf{M}\mathbf{X}^t$, con $\mathbf{M}$ la métrica sobre el espacio de individuos. Pruebe que las componentes principales $\mathbf{c}^k$ son vectores propios de $\mathbf{WD}$ asociados los valores propios λ_k (Sugerencia: ver el esquema de dualidad para el A.C.P. general, Figura 2.4 página 34).

6. Sea la nube $\mathcal{N}(\mathbf{X}, \mathbf{M}, \mathbf{D})$ con $\mathbf{W} = \mathbf{X}\mathbf{M}\mathbf{X}^t$, matriz $n \times n$ de entradas w_{ij}. Pruebe que la distancia entre dos individuos $\mathbf{x}_i, \mathbf{x}_j$ se puede escribir

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = ||\mathbf{x}_i - \mathbf{x}_j||^2 = w_{ii} + w_{jj} - 2wij$$

7. Considérese que se ha observado el crecimiento de 20 novillos, midiéndose el peso al nacimiento ($\mathbf{x}$), a los 50 días ($\mathbf{y}$) y a los 100 días ($\mathbf{z}$) obteniéndose la matriz de covarianzas $\mathbf{V}$ (para pesos iguales a los individuos):

$$\begin{pmatrix} 17{,}56 & 31{,}09 & 47{,}69 \\ 31{,}09 & 91{,}29 & 140{,}39 \\ 47{,}69 & 140{,}39 & 267{,}08 \end{pmatrix}$$

   a) Determine la matriz de correlaciones $\mathbf{R}$.

   b) Probar que los vectores $\mathbf{v}_1 = \begin{pmatrix} 0{,}5496 \\ 0{,}5995 \\ 0{,}5818 \end{pmatrix}$ y $\mathbf{v}_2 = \begin{pmatrix} 0{,}8160 \\ -0{,}2360 \\ -0{,}5276 \end{pmatrix}$ son vector propios de la matriz a diagonalizar en el A.C.P. y determine los valores propios a los que están asociados. ¿Cuánto vale el tercer valor propio?

   c) Mostrar que $\mathbf{v}_1$ y $\mathbf{v}_2$ son ortogonales, y que son de norma 1.

   d) Calcule la inercia de la nube de puntos; calcule el porcentaje de inercia explicada por el primer eje factorial, y el porcentaje de inercia explicada por el segundo eje factorial.

   e) Las correlaciones entre las variables originales y las dos primeras componentes principales son:

	$\mathbf{c}^1$	$\mathbf{c}^2$
$\mathbf{x}$	0.751	0.438
$\mathbf{y}$	0.949	0.309
$\mathbf{z}$	0.997	-0.115

Dibuje el círculo de correlaciones y respresente en él a las tres variables originales.

8. En una encuesta de opinión las respuestas a la pregunta "Se debe capacitar a los empleados desempleados" se distribuyeron por sexo y modalidad de respuesta de acuerdo con la tabla de contigenia siguiente:

Sexo	Respuestas				
	1	2	3	4	5
Masculino	17	27	99	279	232
Femenino	21	41	125	352	267

Cuadro 1: Respuesta según sexo

donde : 1 = en total desacuerdo, 2 = no tan de acuerdo, 3 = podría estar de acuerdo, 4 = bastante de acerdo y 5 totalmente de acuerdo.

a) Calcule los centros de gravedad de la nubes de perfiles-fila y perfiles columna Se encuentra que el centro de gravedad de los perfiles-fila es igual a

b) Calcule la tabla de frecuencias bajo la hipótesis de independencia de las variables "sexo" y "Se debe capacitar a los desempleados".

9. Calcule las correspondientes tablas de perfiles-fila y de perfiles-columna de la tabla de contigencia del ejercicio anterior. Haga un gráfico de los perfiles fila. ¿Se puede afirmar que hay dependencia de las respuestas de los entrevistados a la pregunta "Se debe capacitar a los desempleados", con respecto al sexo?

10. Sean $\mathbf{x}$ y $\mathbf{y}$ dos variables cualitativas con tabla de contingencia asociada $\mathbf{X} = (x_{ij})_{p\times q}$.

    Probar que las siguientes proposiciones son equivalentes:

    a) Las variables $\mathbf{x}$ y $\mathbf{y}$ son independientes.

    b) Los perfiles-fila son iguales a su centro de gravedad $\mathbf{g}_x$.

    c) Los perfiles-columna son iguales a su centro de gravedad $\mathbf{g}_y$.

11. Probar que $I(\mathcal{N}_x) = I(\mathcal{N}_y) = \frac{\chi^2(\mathbf{x},\mathbf{y})}{\mathbf{x}_{..}}$

12. Probar la propiedad de equivalencia distribucional. Es decir, si se tiene la igualdad de perfiles-fila $\mathbf{pf}_i = \mathbf{pf}_{i'}$ entonces

$$d^2_{\chi^2}(\mathbf{pc}_j, \mathbf{pc}_{j'}) = d^2_{\chi^2}(\widetilde{\mathbf{pc}}_j, \widetilde{\mathbf{pc}}_{j'}), \qquad \text{para todo } j, j'$$

donde la matriz $\tilde{X}$ es definida a partir de la tabla de contingencia $\mathbf{X}$ por sustitución de las filas i e i' por la fila que resulta de la suma de ambas: $(x_{i1} + x_{i'1}, \ldots, x_{iq} + x_{i'q})^t$, y $\widetilde{\mathbf{pc}}_j$ es el perfil-columna de la modalidad j de la variable $\mathbf{x}$, calculado con la tabla $\tilde{\mathbf{X}}$.

13. Probar que

    a) $\mathrm{coord}_{u_h}(\mathbf{pf}_i - \mathbf{g}_x) = \mathrm{coord}_{u_h}(\mathbf{pf}_i)$.

    b) $\mathrm{coord}_{v_h}(\mathbf{pc}_j - \mathbf{g}_y) = \mathrm{coord}_{v_h}(\mathbf{pf}_j)$.

14. Probar que, en el ACP de la nube $(\mathbf{X}_{pf}, \mathbf{D}_y^{-1}, \mathbf{D}_x)$, (resp. de la nube $(\mathbf{X}_{pc}^t, \mathbf{D}_x^{-1}, \mathbf{D}_y)$), las componentes principales son $\sqrt{\lambda_h}\mathbf{D}_x^{-1}\mathbf{v}_h$ (resp. $\sqrt{\lambda_h}\mathbf{D}_y^{-1}\mathbf{u}_h$).

15. Probar que los vectores propios asociados a valores propios no nulos en Análisis de Correspondencias, permanecen invariables si se efectúa la operación de centraje de los perfiles. Sugerencia: sea $\hat{\mathbf{X}}_{pf}$ la matriz de perfiles-fila centrados. Utilice la identidad matricial $\hat{V}_x = \hat{\mathbf{X}}_{pf}^t \mathbf{D}_x \hat{\mathbf{X}}_{pf}$ $\hat{\mathbf{V}}_x = \mathbf{V}_x - \mathbf{X}_{pf}^t \mathbf{D}_x \mathbf{1}_p \mathbf{1}_p^t \mathbf{D}_x \mathbf{X}_{pf} \mathbf{D}_y^{-1}$ para deducir que $\hat{\mathbf{V}}_x \mathbf{u}_h = \lambda_h \mathbf{u}_h$ donde $\mathbf{u}_1, ..., \mathbf{u}_r$ son los vectores propios del ACP de $(\mathbf{X}_{pf}, \mathbf{D}_y^{-1}, \mathbf{D}_x)$ asociados a $1 > \lambda_1 \geq \cdots \geq \lambda_r$.

16. Use el Excel que se encuentra en mediación virual y use la hoja relacionada con Profesión-FMI para realizar un análisis factorial de correspondencias. En donde se analice la relación entre la profesión y la respuesta a la frase .El FMI ayuda a resolver la crisis". Donde el nivel se ubica en orden ascendente es decir 1 es muy en desacuerdo y 5 es muy de acuerdo. Debe analizar el porcentaje de inercia conservada, se debe analizar los primeros dos planos factoriales haciendo el gráfico y interpretándolo y determinar cuál perfil fila y cuál primer perfil columna tiene el mayor grado de aporte al primer eje factorial. Puede usar paquetes de Python como puede ser *prince*, *sklearn* o cualquier otro.