

EST-46114 Métodos Multivariados
Primavera 2019
Examen parcial 1
Fecha: 05-Mar-2019

Alumno: _____
Clave única: _____
Calificación: _____

Lean cuidadosamente las preguntas. Sus respuestas deben estar acompañadas con procedimientos y justificaciones.
¡Éxito!

1. (50 puntos) Argumente respecto a la veracidad de los siguientes enunciados:

- El análisis de componentes principales poblacionales (PPCA) es equivalente al análisis de componentes principales muestrales (SPCA).
- Los resultados de aplicar PCA sobre estas dos matrices de covarianzas muestrales,

$$\Sigma_1 = \begin{bmatrix} 1 & 0.1 & 0.3 \\ 0.1 & 2 & 0.2 \\ 0.3 & 0.2 & 3 \end{bmatrix}, \text{ y } \Sigma_2 = \begin{bmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 3 & 0.1 \\ 0.3 & 0.1 & 2 \end{bmatrix},$$

respectivamente, son equivalentes.

- Los resultados del modelo de factores (FA) son invariantes a la forma en cómo se modela la media de todas dimensiones originales en un conjunto de datos.
 - Realizar predicciones de $\mathbf{X} = (X_1, \dots, X_p)$, no observado, con base en datos $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, donde $\mathbf{x}_n = (x_{1n}, \dots, x_{pn})$ para $n = 1, \dots, N$, empleando PCA reduce incertidumbre respecto a realizar la predicción empleando un modelo gaussiano p -dimensional.
 - El resultado en el que después de aplicar PCA a un conjunto de datos se obtiene que la estructura de los datos originales y la de los componentes principales resultante es la misma, es evidencia exclusiva que los datos originales son ortogonales.
 - La matriz de cargas $\mathbf{\Lambda}^{(k)}$ del modelo FA con k factores, aplicado a un vector p -dimensional de mediciones, incluye toda la información acerca de la estructura de correlación de las p mediciones.
 - Si el número de componentes retenidos aumenta, por ejemplo, de m a $(m+1)$, las primeras m componentes del SPCA no habrán de cambiar. Este es también el caso en el modelo de análisis de factores (FA).
2. (20 puntos) Después de aplicar SPCA a un conjunto de pacientes con base 8 mediciones químicas sanguíneas (**rblood**-globulos rojos, **plate**-coagulacion, **wblood**-globulos blancos, **neut**-neutrofenia para combatir infecciones especificas, **lymph**-sistema linfatico/circulacion, **bilir**-bilirrubina, **sodium**-sodio/presion arterial, **potass**-potasio/activacion muscular), se obtienen los siguientes resultados:

a) Eigenvectores

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
rblood					0.943	0.329		
plate	-0.999							
wblood		-0.192		-0.981				
neut						0.758	-0.650	
lymph						-0.649	-0.760	
bilir		0.961	0.195	-0.191				
sodium		0.193	-0.979					
potass					0.329			

b) Eigenvalores

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
1.67101	1.23758	1.11771	0.88227	0.78839	0.69917	0.66002	0.31996

Elabora un análisis y diagnóstico interpretativo acerca de estos resultados.

3. (20 puntos) Continuando con los mismos datos de la pregunta anterior, se ajustan dos modelos de factores con $k = 1$ y $k' = 2$, respectivamente, arrojando los siguientes resultados:

rblood	plate	wblood	neut	lymph	bilir	sodium	potass
0.986	0.852	0.713	0.256	0.005	0.940	0.996	0.982

Loadings:

	Factor1
rblood	0.120
plate	0.384
wblood	0.536
neut	0.862
lymph	-0.997
bilir	-0.244
sodium	
potass	-0.133

BIC = 1200.77

y

rblood	plate	wblood	neut	lymph	bilir	sodium	potass
0.743	0.638	0.487	0.007	0.181	0.604	0.878	0.966

Loadings:

	Factor1	Factor2
rblood		0.506
plate	0.322	0.508
wblood	0.456	0.552
neut	0.994	
lymph	-0.879	-0.214
bilir	-0.139	-0.614
sodium		-0.346
potass	-0.137	-0.122

BIC = 247.35

Elabora también un análisis y diagnóstico interpretativo acerca de estos resultados.

4. (10 puntos) Responde al menos dos las siguientes tres preguntas (recuerda incluir el desarrollo):
- En SPCA, ¿cómo podemos recuperar la matriz de covarianzas muestral a partir de los *eigenectores* y *eigenvalores* muestrales?
 - Si $\mathbf{X} = (X_1, \dots, X_p)$ denota un vector p -dimensional en escala original sobre el cual se aplica el modelo FA con k factores fijos ($k < p$). Describe el efecto que tendrá sobre los factores el reescalar cada columna de \mathbf{X} por c_1, \dots, c_p valores fijos positivos.
 - Desarrolla la expresión para la matriz de covarianzas en el modelo FA si se considera que los factores puedan estar correlacionados.

Si respondes los incisos de esta pregunta, la tercera correcta sumará 5 puntos a la acumulación de puntos del examen.