**Universidad del Valle - Escuela de Estadística**

**Curso: Estadística Aplicada 3**

**Profesor: Andrés Felipe Ochoa Muñoz**

**Laboratorio 2: ACP con datos faltantes**

En este laboratorio realice un ACP con datos faltantes usando las librerías del software R: ade4, plsdepot, FactoMineR o missMDA. La idea del laboratorio es realizar un ACP de una matriz con datos completos y comparar los resultados de un ACP en presencia de datos faltantes. Para trabajar un ACP con datos faltantes utilice el algoritmo NIPALS y el algoritmo EM-PCA.

Para comparar un ACP con datos completos y un ACP con datos faltantes, generé una matriz con el 5%, 10%, 15% y 20% de porcentaje de datos faltantes. Para la comparación tenga en cuenta los siguientes indicadores:

* Poder descriptivo en el eje 1 y 2
* Valores propios, Componentes principales t, Vectores propios u (comparación datos completos vs datos faltantes)
* Correlación entre la componente 1 con datos completos y la componente con datos faltantes
* La estructura factorial en la nube de individuos y en la nube de variables.
* La ortogonalidad en las componentes y la ortonormalidad en los vectores propios.
* Seleccione una matriz con el 20% de NAs y compare la imputación obtenida por NIPALS y EM-PCA.

Las base de datos a trabajar son las mismas del laboratorio 1. Es posible cambiarse a las siguientes:

**Caso. Base de datos milk. (data(milk) en FactomineR)**

**Caso. Base de datos senso. (data(wine) en FactomineR)**

**Otros aspectos para tener en cuenta:**

* Trabaje con el RV de scouffier para realizar las comparaciones.
* Construir indicadores en presencia de datos faltantes.
* Generar no solo una matriz con 5%, 10%, 15% y 20%. Generar m=1000 matrices con dichos porcentajes.
* Estudie el ACP iterativo regularizado.
* ¿Si aumenta el porcentaje de datos faltantes, que pasa con el porcentaje de varianza explicado y que pasa con la correlación de las componentes en datos completos vs datos faltantes?

**Algunas links de consulta:**

* [**https://www.researchgate.net/publication/237414741\_Gestion\_des\_donnees\_manquantes\_en\_Analyse\_en\_Composantes\_Principales**](https://www.researchgate.net/publication/237414741_Gestion_des_donnees_manquantes_en_Analyse_en_Composantes_Principales)
* [**https://www.tdx.cat/bitstream/handle/10803/284659/TVMGR1de1%20.pdf**](https://www.tdx.cat/bitstream/handle/10803/284659/TVMGR1de1%20.pdf)
* [**http://juliejosse.com/wp-content/uploads/2017/10/MissingValuesDauphine.pdf**](http://juliejosse.com/wp-content/uploads/2017/10/MissingValuesDauphine.pdf)
* [**https://ciencias.medellin.unal.edu.co/eventos/seminario-institucional/images/presentaciones/pres\_victor1.pdf**](https://ciencias.medellin.unal.edu.co/eventos/seminario-institucional/images/presentaciones/pres_victor1.pdf)
* [**https://hal.archives-ouvertes.fr/hal-00763227/document**](https://hal.archives-ouvertes.fr/hal-00763227/document)