Clase 6: Técnicas Multivariadas

Justo Andrés Manrique Urbina

28 de septiembre de 2019

1. Diferencia entre métodos

Los componentes principales son la combinación lineal de los autovectores con los datos originales cuyo objetivo es formar nuevas variables para evaluar Σ. Sin embargo, el objetivo del análisis factorial es hallar **estructuras** en grupos de variables. Estas variables se definen como los *constructos* que son determinados por los factores.

2. Extracción de factores

2.1. Componentes principales

Recordemos que la varianza puede descomponerse en:

$$\Sigma = P\Lambda P^T.$$

$$= P(\Lambda)^{0,5}(\Lambda)^{0,5}P^T.$$

$$P\Lambda^{0,5}(P\Lambda^{0,5})^T.$$

Recordemos que los componentes principales utilizan todos los datos, por lo que uno puede elegir cuántos componentes principales desean capturar. Si $m \leq p$, entonces los factores se pueden obtener de la siguiente forma:

$$\Sigma =_{aprox} (P\Lambda^{0,5})(P\Lambda^{0,5})^T + \Phi.$$

2.2. Definiciones

$$\Sigma : \text{Matriz de variana y covarianza de Pearson.}$$

$$R : \text{Matriz de correlación de Pearson.}$$

Si X_i son ordinales, $i = 1, 2, \ldots, p$ se debe usar la correlación de Spearman. Para variables nominales, se debe utilizar el análisis de correspondencia.

3. Aproximación a los factores F

$$X_i = LF + \varepsilon.$$

Aplicar regresión, para ello estimar L. Usando el método de mínimos cuadrados ponderados pues $Var(\varepsilon) = \sigma^2 ought^o$

4. Análisis Discriminante

El anàlisis discriminante es un aprendizaje supervisado de clasificación. Para que este método funcione, todas las variables $(X_1, X_2, \ldots, X_p)$ deben ser continuas. El *target* es una variable nominal. Suposiciones del modelo:

- La matriz de varianza y covarianza para todos los grupos deben ser iguales.
- Las variables numéricas tienen una distribución normal multivariada.

Considerar que estas suposiciones son para el análisis discriminante lineal.

$$X \in \Omega, \Omega = \mathbb{R}_1 U \mathbb{R}_2.$$

$$X \in \mathbb{R}_1 \iff X \in \text{Primera población } \Pi_1.$$

$$X \in \mathbb{R}_2 \iff X \in \text{Primera población } \Pi_2.$$