Clase 5: Técnicas Multivariadas

Justo Andrés Manrique Urbina

21 de septiembre de 2019

1. Componentes principales

La información de una variable univariada se encuentra en su desviación estándar. Similarmente, la información de variables multivariada se encuentra en la matriz de varianza y covarianza Σ. Por propiedad, tenemos que:

$$\sum_{i=1}^{p} \sigma_{ii}^2 = \sum_{i=1}^{p} \lambda_i.$$

Lo que deseamos explicar es la varianza total, la cual es la traza de Σ.Ello se puede explicar a través de los autovalores y los componentes principales. Para hallar los componentes principales, se debe realizrar:

- Construir matriz de covarianza de $X_1 \dots X_p$
- Hallar autovalores y autovectores: $\lambda_i \to e_i$
- $Y_i = e_i^{'} X$, $var(Y_i) = \lambda_i$

Para hallar la correlación entre el componente principal y una variable específica, se utiliza la siguiente fórmula:

$$p_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_{ii}}}{\sqrt{\sigma_{kk}}}.$$

2. Regresión y componentes principales

Sea $Y = X\beta + \varepsilon$, $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2 I$. Si hay multicolinealidad, la varianza de β aumenta considerablemente. Esto se debe a que si existe colinealidad, los autovalores se vuelven 0. Al aplicar inversa y aplicar decomposición espectral, la varianza de β tiene la siguiente expresión:

$$\sigma^2 I(\sum_{i=1}^{p} \frac{1}{\lambda_i} \mu_k \mu_k^T).$$

3. Análisis Factorial

El análisis de componentes principales busca explicar la varianza total entre el menor número de variables. En el análisis factorial buscan explicar una estructura latente a los datos. A través de este análisis se identifican dimensiones que representan esquemas conceptuales de análisis.

El modelo factorial se define como:

$$X_j - \mu_j = l_{j1}F_1 + l_{j2}F_2 + \ldots + l_{jm}F_m + \varepsilon_j.$$

Cada $F.$ se define como una variable latente (no observable). Estas se llaman **factores comunes** y los ε_j se lllaman **factores no comunes**. También se puede hacer un **análisis factorial confirmatorio**, que es un caso de la ecuación estructural.

3.1. Ecuaciones estructurales

En las ecuaciones estructurales existen dos tipos de variables, **latentes y medibles**.

La varianza de X_i, bajo el modelo sería igual a

$$\Sigma = LL^T + \Psi.$$

En dónde LL^T se define como la comunalidad.