



### **Denominación del Taller:**

*“Métodos Multivariantes: Análisis de Componentes Principales y de Cluster. Técnicas para la reducción de dimensiones y la clasificación espacial de datos”*

**Fecha del Dictado:** 07 de noviembre de 2014 (de 15:00 a 19:30 hs)

**Profesor Responsable:** *Dra. Blanca Elizabeth Ponce*

Profesora y Licenciada en Geografía. Egresada de la Facultad de Humanidades. Doctora en Demografía. Egresada de la Escuela de Graduados de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba. Actualmente se desempeña como Auxiliar docente en la cátedra Geografía y Sociedad del Profesorado y Licenciatura en Geografía y es becaria del CONICET, realizando trabajos de investigación vinculados con la Demografía y Determinantes Socioeconómicos de la Salud en el IIGHI-CONICET-UNNE.

**Profesor Responsable:** *Da Silva Cristian Javier*

Prof. en Geografía. Egresado de la Facultad de Humanidades. Actualmente cursa los estudios de Maestría en Teledetección y Sistemas de Información Geográfica en la Universidad del Centro de la Provincia de Buenos Aires (UNCPBA) y el Doctorado en Geografía en la Universidad Nacional del Nordeste (UNNE). Es becario doctoral de CONICET y trabaja en temas relacionados con la Planificación de Transportes público.

### **Fundamentación**

En estudios exploratorios en los que se tiene una gran cantidad de datos, los analistas suelen utilizar muchas variables y observaciones. El análisis subsiguiente es muy complejo, ya que las variables están muy relacionadas entre sí, y el número de observaciones es tan grande que en ocasiones, es difícil utilizar técnicas descriptivas para reducir la dimensión de los datos. Las técnicas de análisis multivariante son muy adecuadas para poder comprender la estructura de estas grandes masas de datos.

En las últimas décadas, el análisis multivariante ha experimentado un importante crecimiento debido básicamente, al desarrollo de nuevas herramientas informáticas. Cualquier paquete estadístico tiene hoy día, implementadas las más importantes herramientas de análisis multivariante. En efecto, el programa estadísticos SPSS es perfectamente adecuados en este contexto.

Dentro de este marco, el siguiente taller tiene por objetivo principal presentar dos herramientas metodológicas: Componentes Principales y Cluster. Las mismas podrían ser utilizadas en Geografía Humana, a la hora de reducir el número de variables seleccionadas y aproximarnos a una regionalización a partir de la clasificación espacial de dichas variables.

El *análisis de componentes principales* permite conocer la interdependencia presentada por un cierto número de variables, susceptibles de ser sintetizadas en un conjunto de factores comunes que subyacen tras ella. El número de factores que se extraen es inferior al número de variables

analizadas, sin embargo, dichos factores son suficientes para resumir la mayor parte de la información contenida en las variables originales. Como consecuencia, los factores o nuevas variables podrán ser utilizados en sustitución de éstas, lo que explica que a menudo se haga referencia a este análisis como técnica de reducción de datos (Antelo y Alonso, 2003).

Por su parte, el análisis de cluster es un método que puede usarse para encontrar qué entidades (sean éstas sujetos u objetos), de un conjunto determinado son similares entre sí. Esta metodología proporciona clasificaciones a partir de datos inicialmente no clasificados, tratando de encontrar grupos en los datos (Picón Prado y otros, 2007).

### **Objetivos General del Taller**

- Presentar dos herramientas metodológicas del Análisis Multivariado (Componentes Principales y Cluster) e identificar sus posibles aplicaciones en Geografía Humana.

### **Objetivos Específicos del Taller**

- Identificar en cada una de las técnicas las ventajas y desventajas que presentan para el análisis geográfico.
- Orientar en selección de variables de carácter demográficas, sociales y económicas, de acuerdo con la unidad espacial y problemática analizada, para la construcción de nuevas variables síntesis.
- Alcanzar una regionalización en el espacio seleccionado con las variables analizadas.
- Interpretar correctamente los resultados obtenidos.

### **Contenidos**

El concepto de Región. Construcción y selección de variables demográficas, sociales y económicas, a partir de la información censal.

Análisis de Componentes Principales: correlación entre las variables; reducción de dimensiones; cantidad de componentes a retener; caracterización. Análisis de Cluster: obtención de los conglomerados a través de procedimientos jerárquicos y no jerárquicos; ventajas y desventajas de ambos procedimientos; caracterización. Representación Cartográfica.

### **Fuentes de Información**

Las principales fuentes de información serán los Censos Nacionales de Población, Hogar y Vivienda correspondientes a los años 2001 y 2010 (Redatam + SP).

### **Actividades:**

En una primera instancia se indagará sobre los intereses de los presentes en el taller, con el fin de seleccionar la unidad espacial y la problemática a analizar, como así también el conjunto de variables con las cuales se trabajarán.

Posteriormente, se aplicará sobre este conjunto de variables, un análisis de componentes principales por un lado y, un análisis de cluster por otro.

Finalmente, los resultados serán representados cartográficamente.

### **Criterios de Evaluación**

Para cumplir con los requisitos de aprobación del taller, se solicitará dos tipos de análisis de los resultados antes alcanzados:

**Componentes Principales:**

1. Análisis de la correlación entre las variables.
2. Identificación del número de factores (variables) que arroja el programa.
3. Reconocimiento de las antiguas variables que integran los nuevos factores (variables).
4. Denominación de los nuevos factores (variables).

**Clusters:**

1. Análisis espacial de los resultados e identificación de los patrones de distribución.
2. Reconocimiento del número de clusters (o posibles regiones).
3. Identificación de las variables que tienen mayor y menor peso en la clasificación de las variables.
4. Denominación de las posibles regiones.

La modalidad de trabajo será grupal. Cada grupo expondrá sus resultados de manera oral ante los demás grupos.