

Universidad de Costa Rica

Sede de Occidente

Departamento de Ciencias Naturales

Sección de Matemática

MA0372 Principios de Estadística Matemática

Licenciado Luis Gerardo Araya Aguilar

Licenciado Jorge Cruz Campos

Lunes 8 de agosto del año 2005.

Entrada:

Ha sido una constante en la formación de profesores de enseñanza media y es tendencia en la educación media la presencia del tema de la estadística y la probabilidad.

Los profesores futuros reciben uno o dos cursos en el campo tendientes hacia la estadística matematizada, cuando estos cursos son ofrecidos por escuelas de matemáticas de las universidades, o tendientes a la estadística inferencial cuando los mismos son ofrecidos por educadores. En los colegios o liceos de educación media, en especial a corrientes extranjeras, se ha insistido en varias ocasiones y este momento no es la excepción, en que debe enseñarse el tema. Sin embargo, la realidad de toda esta situación es que la estadística aprendida por los profesores nunca llega a las aulas de las escuelas pues resulta prácticamente imposible su enseñanza. El formalismo mal entendido sume el tema en un marasmo incomprensible para el propio profesor y más aún para los estudiantes mismos.

Objetivos del curso de estadística matemática.

Además de las hipótesis generales implícitas en toda actividad científica, la Estadística Matemática (EM) hace intervenir un cierto número de hipótesis de comodidad que no se introducen para ser confirmadas o rechazadas, sino para permitir la ejecución de una prueba o el ajuste de un modelo. (Ejemplo la hipótesis de Normalidad en los residuos del modelo lineal). La prueba o el ajuste, sólo son realizables si el conjunto de hipótesis permite una formulación analítica relativamente simple, o permite utilizar tablas estadísticas ya existentes.

Las técnicas de análisis multivariado de datos, en cambio, no están hipotecadas por modelos restrictivos, sólo que no es posible utilizarlas en dimensiones reales sin el concurso de un ordenador. El desarrollo y difusión de la informática, permite actualmente la atención a los métodos de estadística descriptiva multidimensional, cuyos principios teóricos son conocidos desde hace mucho tiempo, pero que ele obstáculo del cálculo los había mantenido alejados. Los campos de observación, ayer juzgados ayer como demasiados amplios, complejos pueden ser ahora estudiados, así

por ejemplo las tablas grandes proporcionadas por los datos de encuestas pueden, en cierta medida, ser trasladados en forma global y sintetizadas.

Serán objetivos de este curso, el conocimiento de los fundamentos teóricos de los métodos diferentes de análisis multivariado de datos y en mayor grado el análisis de componentes principales (ACP), así como el manejo de paquetes integrados que permitan realizar análisis con datos reales, de los cuales hay en abundancia en el país.

El ACP permite obtener un resumen descriptivo y gráfico de una población de n individuos a los cuales se les ha medido p caracteres o variables, de este modo se dispone de una matriz como columnas las p variables, así los análisis se harán sobre los espacios \mathbb{R}^p y \mathbb{R}^n según se trate de individuos o de variables.

Antes de realizar cualquier análisis de datos, es necesario estudiar la información de que se dispone, para ello, se utilizan frecuentemente, histogramas, diagramas de bastones, cálculos de medias, medianas, variancia, desviación corriente, etc.

2

Estos estudios previos nos ayudarán a decidir, por ejemplo, si la matriz a diagonalizar es de covariancia, o la de correlaciones (caso en que las dispersiones de las variables son muy diferentes), o lo que es lo mismo decidir cual es la métrica que debemos usar.

Enfoques diferentes de los métodos de análisis multivariado se encuentran disponibles en la literatura, en particular en la producida en la Escuela Francesa. Dado que este es un curso de Estadística Matemática, y que tiene dos cursos previos de álgebra, el desarrollo teórico del curso debe fundamentarse fuertemente en las ideas del álgebra lineal.

Conceptos tales como espacios vectoriales con producto interno, proyectores, valores y vectores propios, valores extremos de formas estadísticas, son indispensables.

Emplearemos las notas: Principios de Análisis de Datos MA0372, del doctor Trejos Zelaya Javier [1], que han sido producidas por El, para este curso, febrero de 1995, se consultará el texto Tratamiento Estadístico de Datos [2], contiene un desarrollo que se adapta a este esquema, el cual se complementará con Introducción a l'analisi donnes [3]. Se puede consultar Estadística de Taro Yamane [4], en el enfoque tradicional y otra forma de ver el ACP.

Para otros aspectos de bibliografía, ver [1].

En lo referente a Software, se cuenta con el paquete CHADOC y el doctor Javier Trejos, proporciona unas notas manuales en español, el cual contiene los principales métodos de análisis multivariado.

Se revisará y utilizará los materiales producidos por los investigadores de nuestro país.

Contenidos.

1. Entrada a la estadística descriptiva.
2. Entrada a la estadística multidimensional.
3. Análisis de componentes principales.
4. Análisis de correspondencias.
5. Regresión lineal múltiple.
6. Clasificación automática.
7. Análisis de discriminantes.

Evaluación:

1. Tres parciales (70%)
2. Trabajos o tareas (30%)
3. Las fechas de exámenes, trabajos y tareas, se definirán oportunamente, para los meses de setiembre, al inicio, octubre, a mediados y noviembre, casi al final.

Bibliografía:

- [1] Trejos Zelaya Javier. Principios de análisis de datos. UCR. 1995.
- [2] Lebart, Morinesu y Fenelon. Tratamiento estadístico de datos. Maracaibo, Barcelona, México, 1985.
- [3] Cailiez, Pages J. Introducción a l'analyse des donnes, Smash. París, Francia. 1979.
- [4] Yamane, Taro. Estadística. México. 1979.
- [5] Castillo, William, Trejos Javier. Métodos matemáticos aplicados a las ciencias, VII y VIII simposios. Editorial de la Universidad de Costa Rica. 1995.

Observaciones: 7-8-2005.