

UNIVERSIDAD DE COSTA RICA  
SEDE DE OCCIDENTE  
CIUDAD CARLOS MONGE ALFARO SAN RAMON  
DEPARTAMENTO DE CIENCIAS NATURALES  
SECCION DE MATEMATICA

Prof. Luis Gerardo Araya Aguilar. 27 de febrero de 1995  
Programa de Principios de Estadística Matemática MA0372

Entrada:

Ha sido siempre una constante en la formación de profesores de enseñanza media y es una fuerte tendencia en la escuela secundaria el tener siempre presente el tema de las probabilidades y la estadística.

Los profesores futuros reciben, usualmente, uno o dos cursos en el campo con una marcada tendencia hacia la estadística matematizada, cuando estos cursos son ofrecidos por Escuelas de Matemáticas de las Universidades, o con una tendencia a la estadística inferencial cuando los mismos son ofrecidos por educadores. En la escuela secundaria, en especial debido 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 todo 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 asume al tema en un marasmo incomprensible para el propio maestro y más aún para el estudiante.

Objetivos del curso estadística

Además de las hipótesis generales implícitas en toda actividad científica, la Estadística 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 el obstáculo del cálculo los había mantenido alejados. Los campos de observación, ayer juzgados como demasiado amplios, complejos pueden ser ahora estudiados, así por ejemplo las tablas grandes proporcionadas por los datos de encuestas pueden, en cierta medida, ser tratados en forma global y sintetizadas.

Serán objetivos de este curso, el conocimiento de los fundamentos teóricos de los diferentes métodos 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 real de tamaño  $n \times p$ , que tiene como filas los  $n$  individuos y como columnas las  $p$  variables, así los análisis se harán sobre los espacios  $R^n$  y  $R^p$  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. Estos estudios previos nos ayudarán a decidir, por ejemplo, si la matriz a diagonalizar es la de covarianzas 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.

Diferentes enfoques de los Métodos de Análisis Multivariado se encuentran disponibles en la literatura, en particular en la producida por la escuela Francesa. Dado que este es un curso de Estadística 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, vectores y valores propios, valores extremos de formas cuadráticas, son indispensables.

Emplearemos las notas: PRINCIPIOS DE ANALISIS DE DATOS MA0372, del doctor Javier Trejos Zelaya [1], que han sido producidas por El para este curso, febrero de 1995, se consultará el texto TRATAMIENTO ESTADISTICO DE DATOS [2], contiene un desarrollo que se adapta a este esquema, el cual se complementará con INTRODUCCION A L'ANALISE DONNES [3]. Se puede consultar ESTADISTICA 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, proporcionó unas notas manuales en español, el cual contiene los principales métodos de análisis multivariado.

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

Contenidos:

1. Entrada a la Estadística Descriptiva.
2. Entrada a la Estadística Multidimensional.
3. Análisis en Componentes Principales.

4. Análisis de Correspondencias
5. Regresión lineal múltiple.
6. Clasificación automática.
7. Análisis discriminante.

Evaluación:

1. Tres parciales (78%)
2. Trabajos o tareas (22%)
3. Las fechas de exámenes, trabajos y tareas, se definirán.

Bibliografía:

- [1] Trejos Javier, PRINCIPIOS DE ANALISIS DE DATOS, UCR, 1995.
- [2] Lebart, Morineau, Fénelon. TRATAMIENTO ESTADISTICO DE DATOS, Marcombo, Barcelona-México, 1985.
- [3] Caillez, Pages J., INTRODUCTION A L'ANALYSE DES DONNEES, SMASH, París, Francia, 1976.
- [4] Yamane, Taro, ESTADISTICA, Harla, México, 1979.

Observaciones: