

Métodos Estadísticos para la Computación

Profesores: Rodrigo Morón, Juan José Villalba, Santiago Cárdenas,, Yolanda Cabrera

Departamento de Matemática Aplicada
E.T.S Ingeniería Informática

Titulaciones de:

- Grado en Ingeniería Informática
- Grado en Ingeniería del Software
- Grado en Ingeniería de Computadores

Universidad de Málaga

email: ruyman@uma.es / jvsenciales@uma.es / scardenas@uma.es / / yolandacc@uma.es

Contenido

- 1 Estadística y Ciencia de Datos
 - Estadística
 - Ciencia de Datos
 - Historia de la Estadística
- 2 Descripción de los temas
 - Estadística Descriptiva
 - Cálculo de Probabilidades
 - Estudio estadísticos e Inferencia Estadística.
- 3 Estructura de la asignatura
 - Evaluación
 - Bibliografía

La Estadística

- El objetivo de la Estadística es el de organizar, describir y sacar consecuencias de aquellos fenómenos que no son deterministas. Es decir, que tienen una componente aleatoria.
- La realidad es, que ya sea por desconocimiento o por una gran complejidad, la mayoría de sucesos que observamos en nuestro mundo debemos modelizarlos como sucesos no deterministas.

La Ciencia de datos (Data Science)

- En los últimos años la capacidad de almacenar información y de memoria para procesarla ha hecho que tengamos más datos que nunca.
- Además, en los últimos años se ha desarrollado muchísimo la capacidad de procesamiento en paralelo, por lo que tenemos más capacidad que nunca para procesarlos y analizarlos.
- Todo esto está haciendo que el análisis y el procesado de datos se haya convertido en el área más candente y que más crece dentro del mundo de la informática. El análisis de datos está transformando nuestro mundo.
- Esto ha llevado a acuñar un nuevo término:

Ciencia de Datos (Data science)

Partes de la Ciencia de datos

La Ciencia de datos esta formada principalmente por:

- Estadística.
- Otras técnicas de aprendizaje avanzadas (Machine Learning)
- Técnicas de procesado y limpieza de datos.
- Técnicas y algoritmos de optimización.

Los Macrodatos (Big Data)

- Hay grandes oportunidades para las empresas cuando analizamos cantidades muy grandes de datos, del orden de Terabytes.
- A esto lo llamamos «Big Data» (en español, Macrodatos).
- No ha de confundirse Ciencia de Datos con Big Data.
- La mayor parte de la Ciencia de Datos que hacemos no es sobre Big Data, sino que es sobre conjuntos de datos del orden de Gigabytes.
- La diferencia fundamental está en que los datos quepan en la memoria de un ordenador o no quepan y deban ser tratados en un cluster de ordenadores.

Partes de la Estadística

La Estadística actual esta formada principalmente por:

- Estadística Descriptiva.
- Cálculo de Probabilidades.
- Inferencia estadística.
- Análisis multivariante.

Comienzos de la estadística descriptiva

- Tiene su origen en los censos de la antigüedad (3050 ac Egipto, 2238 ac China)
- La palabra Estadística viene del alemán: Statistik y significa «Ciencia del Estado».
- Se desarrolló como ciencia con el Despotismo Ilustrado.

Comienzos del Cálculo de Probabilidades

- Los juegos de azar tienen un origen remoto (paleolítico).
- Se acepta que la teoría matemática de la probabilidad fue iniciada por Blaise Pascal (1623-1662) y Pierre Fermat (1601-1665) cuando analizaron problemas relacionados con los juegos de dados propuestos por el caballero de Meré.
- Algunos precursores fueron Girolamo Cardano (1501-1576) y Galileo Galilei (1564-1642).
- Siglo XVIII: Leyes grandes números, tablas de mortalidad. Distribuciones. Errores. Nacimiento de las matemáticas sociales.
- Siglo XIX: Teorema Central del Límite. Método Mínimos Cuadrados.
- Siglo XX: Axiomática de Kolmogoroff.

La Estadística moderna

- Galton (1822-1911) Introdujo el concepto de línea de regresión.
- Pearson (1857-1936) Tablas estadísticas para los contrastes.
- Gosset (1876-1937) (Student) Método de Montecarlo. Contrastes muestras pequeñas.
- Fisher (1890-1962) Análisis discriminante. Diseños experimentales.

La estadística es la base de las decisiones

La Estadística es la base de
las decisiones objetivas
la base de la Ciencia,
y la base de la Democracia.

No somos buenos haciendo apreciaciones...



El sesgo de confirmación

1620 «Novum Organum» de Francis Bacon: El Método Científico.

El entendimiento humano, una vez que ha adoptado una opinión [...] dibuja todo lo demás para apoyar y mostrar conformidad con ella. Y pese a haber un gran número de ejemplos, y de peso, que muestran lo contrario, los ignora o desprecia, prescinde de ellos o los rechaza.

Tema 1: Estadística descriptiva de una variable

El objetivo principal es representar y obtener conclusiones de un conjunto de datos, analizando cada variable de forma aislada.

También, hablaremos de datos extraños (outliers).

Los métodos pueden clasificarse en:

- Agrupar datos en tablas de frecuencias.
- Representaciones gráficas.
- Obtención de estadísticos:
 - De tendencia central (media, mediana,...)
 - De dispersión (varianza, rango,...)
 - De forma (sesgo, curtosis...)

Tema 2: Machine learning y Regresión lineal

Mostramos relaciones básicas entre dos variables y definimos el concepto de Machine Learning. El contenido principal es la técnica más importante de Machine Learning: La regresión lineal.

Los métodos pueden clasificarse en:

- Agrupar datos en tablas de frecuencias bidimensionales.
- Obtención de estadísticos bidimensionales.
- Técnicas de regresión lineal.

Tema 3: Series temporales

Series temporales y series de números índice.

- Definición de series temporales.
- Cálculo de las componentes de una serie temporal.
- Predicción.

Tema 4: Probabilidad

- Concepto de experimento aleatorio.
- Sucesos. Axiomática.
- Probabilidad.
- Probabilidad condicionada.
- El Teorema de Bayes.

Tema 5: Variables aleatorias y distribuciones.

- Concepto de variable aleatoria. Tipos.
- Principales distribuciones discretas: Uniforme, Binomial, Poisson, . . .
- Principales distribuciones continuas: Normal, T-Student, Exponencial, . . .

Tema 6: Estudio estadísticos e Inferencia Estadística.

- Estudios estadísticos.
- Muestra, tipos de muestreo, estimador,
- Estimadores eficientes de la media, proporción, varianza, . . .
- Estimación por intervalos.
- Contraste de hipótesis paramétricos.
- Contraste de hipótesis no paramétricos.

Evaluación Junio

Se realizarán actividades de evaluación continua que constarán de 3 pruebas de conocimiento y un trabajo práctico.

Calificación= Actividades (máx. 7)+ Ex. Final(máx 3).

Evaluación Septiembre y Diciembre

Calificación= Examen Final (máx 10).

Bibliografía

- A. M. Montiel, F. Rius, F. J. Barón. **Elementos básicos de estadística económica y empresarial**. Ed. Prentice Hall.
- C. M. Cuadras, B. Echevarría, J. Mateo, P. Sánchez. **Fundamentos de estadística. Aplicación a las ciencias humanas**. Ed. Promociones Publicaciones Universitarias.
- D. Peña Sánchez de Rivera. **Estadística. Modelos y métodos**. Ed. Alianza Universidad
- J. A. Viedma Castaño. **Exposición intuitiva y problemas resueltos de metodos estadísticos**. Ed. del Castillo
- J. Lóbez Urquia, A. Casa Aruta. **Estadística intermedia**. Ed. Vicens Vives.
- Spiegel. **Estadística**. Ed. Schwam-McGraw-Hil
- V. Quesada, A. Isidoro, L. A. López. **Curso y ejercicios de Estadística**. Ed. Alhambra Universidad.