

Estadística

Josep Gibergans Bàguena
Àngel J. Gil Estallo
Carles Rovira Escofet

XP08/75057/02301



Universitat Oberta
de Catalunya

www.uoc.edu

Josep Gibergans Bàguena

Doctor en Ciencias Físicas por la Universidad de Barcelona en el año 2001. Profesor titular de escuela universitaria del Departamento de Matemática Aplicada III de la Universidad Politècnica de Catalunya. Desde 1993 su actividad docente se centra en la Escuela Universitaria de Ingeniería Técnica Industrial de Barcelona en materias de matemáticas y estadística en los estudios de dicho centro. Es consultor de la Universitat Oberta de Catalunya desde 1998.

Àngel J. Gil Estallo

Doctor en Ciencias Matemáticas por la Universidad de Barcelona desde el año 1996. Profesor titular de escuela universitaria de la Universidad Pompeu Fabra desde 1991. Su actividad docente se centra en temas de matemáticas, estadística e informática en los estudios de Economía de dicha Universidad. Consultor de la Universitat Oberta de Catalunya desde 1998.

Carles Rovira Escofet

Doctor en Matemáticas por la Universidad de Barcelona. Profesor titular en el Departamento de Estadística de la Universidad de Barcelona. Su investigación se sitúa en el campo de los procesos estocásticos, principalmente en el estudio de ecuaciones diferenciales estocásticas.

Tercera edición: febrero 2009
© Josep Gibergans Bàguena, Àngel J. Gil Estallo, Carles Rovira Escofet
Todos los derechos reservados
© de esta edición, FUOC, 2009
Av. Tibidabo, 45-47, 08035 Barcelona
Realización editorial: Eureka Media, SL
Depósito legal: B-2.292-2009

Ninguna parte de esta publicación, incluido el diseño general y de la cubierta, puede ser copiada, reproducida, almacenada o transmitido de ninguna manera ni por ningún medio, tanto eléctrico como químico, mecánico, óptico, de grabación, de fotocopia, o por otros métodos, sin la autorización previa por escrito de los titulares del copyright.

Introducción

Es fácil apreciar, simplemente examinando nuestro entorno más inmediato, la creciente importancia de la multitud de datos estadísticos que nos rodean: estadísticas sobre precios, sobre viviendas, sobre ordenadores... También en la mayoría de los puestos de trabajo el análisis de datos está tomando una importancia destacable, por ejemplo, en términos de control de calidad, de *data mining*, etc. Por otra parte, en algunos temas más técnicos, como por ejemplo el procesamiento de señales y las comunicaciones, son básicos los llamados modelos probabilísticos.

A todo ello se unen las múltiples facilidades de cálculo estadístico que se nos ofrecen hoy en día (paquetes estadísticos, páginas web, etc.), que exigen, para hacer un uso correcto de las mismas, un buen conocimiento de las técnicas estadísticas, de sus posibilidades y sus limitaciones.

La estadística se preocupa de estudiar situaciones en las que aparece el azar o, dicho de otro modo, situaciones en las que hacemos unas observaciones en las que hay elementos fuera del control del observador. La formulación de una teoría matemática que nos ayude a interpretar de manera coherente y científica este tipo de situaciones es una herramienta fundamental en cualquier investigación o estudio.

Todos hemos oído afirmaciones como “Hay mentiras y estadísticas” o “Con la estadística se puede demostrar cualquier cosa”. En este curso veremos que esto no es cierto. Quizá se puede decir que con la estadística no somos capaces de demostrar nada, pero comprobaremos que es la herramienta fundamental que nos proporciona todos los indicios de cómo es la realidad.

Podéis encontrar constantemente las aplicaciones de la estadística: cuando leéis un diario, cuando os llegan unos datos o unos gráficos al trabajo, cuando en la televisión hablan de índices de audiencia, etc.

Podemos recordar lo que dijo el escritor H.G. Wells:

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.”

Objetivos

La asignatura *Estadística* tiene como objetivo general introducir las técnicas básicas estadísticas para poder establecer un puente entre los modelos matemáticos y los fenómenos reales, de manera que seamos capaces de determinar qué técnica estadística se puede aplicar en cada situación, cómo se tiene que aplicar y hasta qué punto podemos estar seguros de las conclusiones a las que se llegan.

Para lograr este objetivo, se ha distribuido la asignatura en cuatro partes. Tres partes tratan la descripción y aplicación de diferentes técnicas estadísticas: el análisis de datos, la inferencia estadística y los modelos de regresión. La cuarta parte, esencial para entender la metodología propia de la estadística, está dedicada a la teoría de la probabilidad, para así poder dar sentido a las técnicas introducidas y entender sus limitaciones.

Describiremos a continuación los objetivos más concretos que hay que alcanzar en cada una de las partes mencionadas:

1. Estadística descriptiva, introducción al análisis de datos y muestreo (módulos 1 y 2). En esta parte básicamente se introducen los conceptos que más adelante permitirán hacer un trabajo estadístico: definición de la variable de interés, recogida y organización de datos y cálculo de parámetros que resuman el comportamiento de esta variable. También se insiste en la representación gráfica de los datos como primer paso en su estudio.
2. Probabilidad y variables aleatorias (módulos 3, 4 y 5). En esta parte se describe el modelo matemático de los fenómenos aleatorios que “soporta” la teoría estadística y que permite construir los modelos estadísticos de una manera rigurosa: la teoría de probabilidades. En ésta se estudia con detenimiento el concepto de variable aleatoria y se presentan las más habituales, entre las cuales destacan por su importancia la distribución normal y sus propiedades. También se presenta la distribución t y se trata la aproximación de la binomial a la normal como paso previo para presentar el teorema del límite central. Se introduce asimismo el concepto de control de calidad y sus aplicaciones.
3. Inferencia y modelización estadística (módulos 6, 7, 8 y 9). En esta parte se presenta la teoría que permite sacar conclusiones sobre una población a partir de una muestra (un subconjunto de los individuos de la población) y precisar con qué márgenes de confianza son válidas estas afirmaciones. Como aplicación de las distribuciones muestrales y del teorema del límite central se definen los intervalos de confianza para la media y la proporción

de una población, y se insiste en la interpretación del resultado que se obtiene al calcular uno de estos intervalos. Finalmente, se introducen las distribuciones necesarias para llevar a cabo la inferencia estadística (concretamente, la F de Snedecor y la χ^2 cuadrado), y se estudian básicamente los contrastes de hipótesis sobre medias y varianzas para una o dos poblaciones. Estos contrastes permiten, dentro de las limitaciones del análisis estadístico, decidir entre dos hipótesis mutuamente excluyentes referidas a los posibles valores de la media, la proporción o la varianza de los valores de un conjunto de datos.

4. Regresión lineal múltiple y análisis de la varianza (módulos 10, 11 y 12). En esta parte se desarrolla la metodología que permite explicar una variable como combinación lineal de otras, lo que da lugar a los llamados modelos de regresión (ya sea simple, en los que se utiliza una única variable explicativa, o múltiple, en la que tenemos más variables explicativas). También se tiene que hacer inferencia sobre alguno de los parámetros del modelo y entender el concepto de bondad del ajuste. Finalmente, se introduce la comparación de medias entre varias poblaciones mediante el llamado ANOVA.

Como veis, son unos objetivos ambiciosos, que nos darán una rica panorámica de la práctica y la metodología de la estadística.

Contenidos

Módulo 1

Estadística descriptiva

Àngel J. Gil Estallo

1. Tipos de datos y su representación gráfica
2. Medidas de centro y propiedades
3. Medidas de dispersión

Módulo 2

Muestreo

Àngel J. Gil Estallo

1. Muestreo

Módulo 3

Probabilidad

Àngel J. Gil Estallo

1. Introducción a la probabilidad
2. Combinatoria y técnicas de recuento
3. Probabilidad
4. El teorema de Bayes

Módulo 4

Variables aleatorias

Carles Rovira Escofet

1. Introducción a las variables aleatorias. Variables aleatorias discretas
2. Esperanza y varianza
3. Algunas distribuciones discretas
4. Variables aleatorias continuas
5. Algunas leyes continuas. La ley normal
6. Procesos estocásticos

Módulo 5

Teorema del límite central

Carles Rovira Escofet

1. La distribución de la media muestral
2. El teorema del límite central

Módulo 6

Intervalos de confianza

Àngel J. Gil Estallo

1. Introducción a los intervalos de confianza. El caso de la media aritmética
2. Intervalo de confianza para la proporción

Módulo 7

Contraste de hipótesis

Carles Rovira Escofet

1. Introducción al contraste de hipótesis
2. Contrastes sobre la media y sobre la proporción

Módulo 8

Contraste de dos muestras

Josep Gibergans Bàguena

1. Contrastes sobre la diferencia de medias
2. Contrastes sobre la diferencia de proporciones

Módulo 9

Contraste de varianzas

Josep Gibergans Bàguena

1. Contraste de la varianza
2. Comparación de varianzas

Módulo 10

Regresión lineal simple

Josep Gibergans Bàguena

1. El modelo de regresión simple
2. La calidad del ajuste
3. Inferencia en la regresión

Módulo 11

Regresión lineal múltiple

Josep Gibergans Bàguena

1. El modelo de regresión múltiple
2. La calidad del ajuste
3. Inferencia en la regresión lineal múltiple

Módulo 12

El análisis de la varianza (ANOVA)

Josep Gibergans Bàguena

1. El análisis de la varianza (ANOVA)

