Facultad de Matemáticas
Unidad de Postgrado e Investigación
Especialización en Estadística

MISIÓN

Formar profesionales altamente capacitados, desarrollar investigación y realizar actividades de extensión en matemáticas y computación así como en sus diversas aplicaciones.

Taller de análisis exploratorio de datos

Programa

Período: Enero – junio, 2010
Horario: Viernes de 19:30 a 21:00 horas
Profesor: Luis A. Rodríguez Carvajal

Objetivo General

Al terminar el curso el alumno será capaz de:

i) Realizar un análisis exploratorio de datos provenientes de una o más muestras.
ii) Conocer las principales aplicaciones de la función de distribución empírica en la validación de supuestos.
iii) Depurar una base de datos para su posterior análisis estadístico.

Descripción del curso.

Muchos de los métodos o modelos estadísticos se basan fuertemente en supuestos acerca de la población de donde provienen las muestras, tales como normalidad, independencia y homocedasticidad. La verificación de la normalidad y la homocedasticidad se apoya tanto en métodos exploratorios como en pruebas de hipótesis. En cuanto a los métodos exploratorios, las medidas de tendencia central, dispersión, asimetría o sesgo y curtosis son de las más utilizadas y los métodos gráficos más comunes son el histograma, el diagrama de caja y bigotes y el gráfico de probabilidad. En este curso se presentan éstas y otras herramientas; se enfatiza su rol en la verificación de los supuestos mencionados. En donde sea pertinente, se hará uso de un software estadístico.

Contenido

1. Preliminares (dos sesiones)
   Objetivo: El alumno comprenderá la diferencia entre población y muestra, el concepto de variable como un atributo a medir en una población y la independencia estadística de variables aleatorias. También comprenderá los distintos tipos de medición.
   1.1. Conceptos básicos:
      1.1.1. Población, muestra, estadísticos y parámetros.
      1.1.2. Variables reales y aleatorias.
   1.2. Clasificaciones de los datos:
      1.2.1. Cualitativa (nominales y ordinales) y cuantitativa (discretas y continuas).
      1.2.2. Nominal, ordinal, de intervalo y de razón.

2. Gráficas y estadísticas (siete sesiones)
   Objetivo: El alumno utilizará los métodos gráficos y numéricos para analizar características de la muestra de datos, como sesgo, simetría, modas. Determinará de acuerdo a la naturaleza de los datos las medidas de tendencia central y de dispersión apropiadas.
   2.1. Medidas de tendencia central: media, mediana, moda.
   2.2. Medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación.
   2.3. Características de la distribución: asimetría o sesgo, curtosis, cuantiles.
   2.4. Histogramas como aproximación de la función de densidad de la variable.
   2.5. Gráfica de barras y circulares.
   2.6. Gráfica de caja y bigotes.
   2.7. Gráfica de tallo y hojas.
   2.8. Gráficas de tendencia (en el tiempo y en el espacio).
   2.9. Regla empírica para la descripción de datos.
   2.10. Teorema de Tchebyshev.

3. La función de distribución empírica (cuatro sesiones)

Objetivo: El alumno aprenderá a calcular y graficar la función de distribución empírica para una muestra, y la interpretará como una aproximación a la función de distribución poblacional. Comprenderá el concepto de gráfica de probabilidad y lo utilizará como herramienta para validar supuestos distribucionales. Comprenderá e interpretará las gráficas Q-Q.

3.1. La distribución empírica como aproximación de una función de distribución verdadera.
3.2. Teorema de Glivenko-Cantelli
3.3. La gráfica de probabilidad: definición y usos.
  3.3.1. normal
  3.3.2. exponencial
  3.3.3. otros
3.4. Gráficas Q-Q: definición y aplicaciones

4. Exploración de datos de dos o más muestras (tres sesiones)
   Objetivo: El alumno explorará y describirá dos o más muestras utilizando las técnicas gráficas y las medidas estudiadas en las unidades anteriores.
   4.1. Medidas descriptivas y de asociación.
   4.2. Histogramas y cajas y bigotes múltiples.
   4.3. Gráficas de dispersión.

5. Preparación final de bases de datos (cuatro sesiones).
   Objetivo: El alumno detectará posibles inconsistencias en la base de datos y propondrá soluciones encaminadas a depurar la base de datos para su posterior análisis estadístico.
   5.1. Detección de datos atípicos.
   5.2. Tratamiento de datos faltantes.
   5.3. Transformaciones.

Estrategias de enseñanza

Este taller es eminentemente aplicado. Para cada tema, a partir de un ejemplo de preferencia con datos reales, se introducen los conceptos básicos de estadística necesarios al abordar un problema de interés. Se motiva el uso de estadísticos y gráficas para estudiar las características de los datos. Se depuran los datos con miras al análisis estadístico. A lo largo del curso, y donde es oportuno, se utiliza software estadístico.

Criterios de evaluación

Se harán tres exámenes parciales, uno después de las dos primeras unidades; otro después de la unidades 3 y uno más, después de las unidades 4 y 5. La calificación final será el promedio final de los tres exámenes.

Bibliografía

Devore, J. y Peck, R. (1986) *Statistics: the Exploration and Analysis of Data*, West Pub., Nueva York.
Evans, M.J. y Rosenthal, J. (2005) *Probabilidad y Estadística*, Reverté, Barcelona, España.
Gotkin, L.G. y Goldstein, L.S. (1973) *Estadística Descriptiva,* texto programado, Vol 1*,* Limusa, México.
Hartwig, F. y Dearing, B. (1980) *Exploratory Data Análisis*. Editorial SAGE, Beverly Hills, California.
Holguin Quiñones, F. (1988) *Estadística descriptiva Aplicada a las Ciencias Sociales,* Unam, México.
Jarrell, S. (1994) *Basic Statistics*, Wm. C. Brown, Dubuque, Iowa.
Milliken, G. y Johnson, D. (2009) *Analysis of Messy Data Volume I: Designed Experiments*, 2ª edición, Kansas State University, Manhattan, Kansas, USA
Milliken, G. y Johnson, D. (1989) *Analysis of Messy Data Volume II: Nonreplicated Experiments*, Kansas State University, Manhattan, Kansas, USA.
Milliken, G. y Johnson, D. (2001) *Analysis of Messy Data Volume III:Analysis of Covariance*, Kansas State University, Manhattan, Kansas, USA
Newman, I. y Newman, C. (1994) *Conceptual Statistics for Beginners,* University Press of América, Nueva York.
Peck, R.; Casella, G.; Cobb, G.; Hoerl, R.; Nolan, D. Starbuck, R. y Stern, H. (2006) *Statistics: A Guide to the Unkonwn*, Thomson, Belmont, California.
Tanur, J. (1992) *Estadística: Una Guía a lo Desconocido,* Alianza, Madrid.
Tufte, E (2003) *The Visual Display of Quantitative Information*, 2ª edición, Graphics Press, York, Reino Unido.