

Estatística

*Enxeñaría Técnica en Informática de Sistemas. 1º curso.
Troncal 2º Cuadrimestre.*

Curso 2005-2006

PROFESORES ENCARGADOS DA DOCENCIA

Manuel Febrero Bande (coordinador)

Beatriz Pateiro López

Pedro Galeano San Miguel

DESCRITORES DA MATERIA NO PLANO DE ESTUDIOS

Estatística descriptiva. Probabilidades. Inferencia estatística. Métodos estadísticos aplicados.

CRÉDITOS

4.5 teóricos + 3 prácticos

DEPARTAMENTO

Departamento de Estatística e Investigación Operativa

OBXECTIVOS

Introducción ós fundamentos e modelos básicos dos métodos estadísticos.
Análise exploratorio e inferencial de datos.

PROGRAMA TEÓRICO

1. Estadística Descriptiva

1.1.Descripción estadística unidimensional

- Introducción a la Estadística Descriptiva. Conceptos básicos. Tipos de datos. Variables estadísticas: cualitativas y cuantitativas, discretas y continuas. Frecuencias. Tablas estadísticas. Representaciones gráficas. Medidas de centralización. Medidas de posición. Medidas de dispersión. Medidas de forma. Otras medidas características. Transformaciones en los datos y su efecto en el análisis descriptivo.

1.2.Descripción estadística conjunta de varias variables

- Variable estadística bidimensional: Definición y clasificación. Distribución conjunta de frecuencias. Representación tabular. Distribuciones de frecuencias marginales y condicionadas. Representaciones gráficas. Momentos. Covarianza y Correlación. Dependencia lineal. Rectas de Regresión.
- Generalización al caso n-dimensional. Distribuciones de frecuencias conjunta, marginales y condicionadas. Vector de medias y matriz de varianzas-covarianzas. Varianza generalizada. Interpretación

2. Modelos de distribución de Probabilidad

2.1.Espacio Probabilístico

- Introducción histórica. Evolución del Cálculo de Probabilidades. Experimentos aleatorios. Sucesos y espacio muestral. Operaciones con sucesos. Frecuencia y probabilidad: Definición empírica o de Richard von Mises de probabilidad. Definición axiomática de Probabilidad: Axiomática de Kolmogorov. Propiedades derivadas de los axiomas. Asignación de probabilidades en la práctica. Aplicaciones de la combinatoria a los modelos probabilísticos. Problemas clásicos de probabilidad geométrica

2.2. Probabilidad condicionada

- Probabilidad condicionada: Definición y concepto. Independencia de sucesos. Probabilidad de intersección de sucesos. Teorema del producto. Teorema de las probabilidades totales. Teorema de Bayes

2.3. Variables aleatorias unidimensionales

- Concepto de variable aleatoria unidimensional. Ejemplos. Función de distribución de una variable aleatoria unidimensional: Definición y propiedades. Variables aleatorias discretas. Definición y ejemplos. Función masa de probabilidad y función de distribución. Variables aleatorias continuas. Definición y ejemplos. Función de densidad y función de distribución. Transformaciones de variables aleatorias. El teorema de cambio de variable.

2.4. Medidas características de una variable aleatoria unidimensional

- Esperanza matemática de una variable aleatoria unidimensional: concepto, propiedades e interpretación. Varianza y desviación típica de una variable aleatoria unidimensional: Definición y propiedades. Momentos de una función de distribución. Momentos respecto al origen. Momentos respecto a la media. Relaciones entre los momentos. Función generatriz de momentos. Cálculo de momentos. Algunas desigualdades entre momentos de interés: Markov, Tchebychev, Lyapunov. Principales características de las distribuciones unidimensionales: Medidas de posición y centralización, Medidas de dispersión, Medidas de concentración, Medidas de asimetría o deformación, Medidas de apuntamiento, curtosis o concentración central

2.5. Modelos unidimensionales discretos de distribuciones de probabilidad

- Distribución degenerada. Distribución uniforme en n puntos. El proceso de Bernoulli. Distribuciones asociadas: Bernoulli, Geométrica, Binomial, Binomial Negativa, Distribución hipergeométrica: La distribución binomial como límite de la distribución hipergeométrica. El proceso de Poisson: Distribución de Poisson. La distribución de Poisson como límite de la distribución binomial.

2.6. Modelos unidimensionales continuos de distribuciones de probabilidad

- Distribución uniforme. Distribución exponencial: Relación con la distribución de Poisson. Distribución normal: La distribución normal como límite de la distribución binomial. Distribución gamma. Distribución Weibull. Distribución beta. Distribución de Pareto. Distribución de Cauchy. Distribución doble exponencial. Distribución logarítmico normal

2.7. Variables aleatorias multidimensionales

- Definición de vector aleatorio. Distribución asociada a un vector aleatorio bidimensional. Tipos. Distribución asociada a un vector aleatorio n -dimensional. Distribuciones marginales y condicionadas. Distribuciones truncadas. Independencia de variables aleatorias. Esperanza matemática de un vector aleatorio. Covarianza y correlación. Matriz de varianzas-covarianzas. Matriz de correlaciones. Correlación múltiple y parcial. Momentos. Función generatriz de momentos. La esperanza condicionada. Función de regresión.

2.8. Modelos multidimensionales de distribuciones de probabilidad

- Distribución multinomial. Distribución normal bidimensional. Distribución normal n -dimensional. Introducción a los Teoremas Límite. Sucesiones de variables aleatorias. Convergencia de sucesiones de variables aleatorias. Relaciones entre los distintos tipos de convergencia. Leyes de los Grandes Números: concepto e interpretaciones intuitivas. Teorema Central del Límite: concepto e interpretación. Resultados sobre el Teorema Central del Límite. Ejemplos y aplicaciones.

3. Inferencia Paramétrica

3.1. Introducción a la Inferencia Estadística

- Introducción: Objeto y necesidad de la Inferencia Estadística. Clasificación de los procedimientos de la Inferencia Estadística. Respecto al objetivo del estudio: Muestreo frente a diseño. Respecto al método utilizado: Métodos paramétricos frente a no paramétricos. Respecto a la información considerada: Enfoque clásico frente a Bayesiano. Concepto de población. Distribución teórica. Concepto de muestra aleatoria. Distribución conjunta de la muestra. Definición de estadístico y su distribución en el muestreo.

3.2.Introducción a los métodos de muestreo

- Muestreo aleatorio simple (m.a.s.). Muestreo aleatorio sin reemplazamiento. Muestreo estratificado. Muestreo por conglomerados. Muestreo sistemático.

3.3.Introducción a la Simulación estadística

- Introducción. Números aleatorios y pseudo-aleatorios. Métodos de generación de números pseudo-aleatorios. Método de los cuadrados medios. Métodos congruenciales. Generación de variables aleatorias discretas. Generación de variables aleatorias continuas: método de la transformada inversa. Generación de variables aleatorias continuas: método del rechazo con entorno finito (a,b). Generación de variables aleatorias continuas: método del rechazo con entorno infinito $[0,\infty)$. Métodos de generación de distribuciones univariantes notables. Otros métodos de simulación: El Método de Box-Muller, El Método Polar de Marsaglia.

3.4.Distribución muestral

- Función de distribución muestral o empírica. Características. Momentos muestrales. Comportamiento asintótico. Cuantiles muestrales. Comportamiento asintótico. Distribuciones en el muestreo de poblaciones normales. Distribución de la media muestral. Distribución de la varianza muestral. La distribución χ^2 de Pearson. Estadístico de Student. Distribución t de Student. Distribución de la diferencia de medias muestrales. Distribución del cociente de cuasivarianzas muestrales. Distribución F de Snédecor. Estimación puntual. El concepto de estimador puntual. Comparación de estimadores: Riesgo y error cuadrático medio. Propiedades deseables de los estimadores: Estimadores insesgados o centrados, Estimadores consistentes, Estimadores invariantes (por cambios de origen o escala) , Estadísticos suficientes. Cotas para la varianza de un estimador. Eficiencia de un estimador. Estimación de la varianza de un estimador: Método Jackknife, Método Bootstrap

3.5.Métodos de estimación

- Método de los momentos. Método de máxima verosimilitud. Propiedades asintóticas. Estimación mínimo cuadrática. Estimación Bayesiana: Estimación y riesgo Bayes.

3.6.Intervalos de confianza

- Planteamiento del problema. Conceptos fundamentales: Nivel de confianza, Intervalo de confianza. Método de la cantidad pivotal para la construcción de intervalos de confianza. Intervalos de confianza para los parámetros de distribuciones normales. Intervalo de confianza para la media si la varianza poblacional es conocida. Intervalo de confianza para la media si la varianza poblacional es desconocida. Intervalo de confianza para σ^2 si la media poblacional es conocida. Intervalo de confianza para σ^2 si la media poblacional es desconocida. Intervalo de confianza para la diferencia de medias con varianzas conocidas. Intervalo de confianza para la diferencia de medias con varianzas desconocidas pero iguales. Intervalo de confianza para la diferencia de medias con varianzas desconocidas y desiguales. Intervalo de confianza para la diferencia de medias de dos poblaciones normales no independientes (datos apareados). Intervalo de confianza para el cociente de varianzas poblacionales. Determinación del tamaño muestral.

3.7.Otros métodos de construcción de intervalos de confianza.

- Método general de Neyman. Intervalos de verosimilitud. Intervalos de confianza basados en distribuciones asintóticas. Intervalos de confianza Bayesianos. Contrastes de hipótesis paramétricas. El Lema de Neyman-Pearson. Planteamiento de un problema de contraste de hipótesis. Formulación de hipótesis. Hipótesis nula y alternativa; hipótesis simples y compuestas. Conceptos básicos: región crítica y región de aceptación; nivel de significación; error de tipo I y error de tipo II. Nivel crítico. Función de potencia. Contrastes uniformemente más potentes (UMP). Contrastes de hipótesis simples. El Lema de Neyman-Pearson. Formulación de contrastes de hipótesis unilaterales y bilaterales.

3.8.Métodos de contraste.

- Contrastes de hipótesis clásicos:Contrastes acerca de la media, Contrastes sobre la varianza, Contrastes para una proporción. Contrastes sobre la diferencia de medias de dos poblaciones, Contrastes acerca del cociente de varianzas de dos poblaciones, Contrastes para la diferencia entre proporciones de dos poblaciones. Test de la razón de verosimilitudes. Relación entre los contrastes de hipótesis y los intervalos de confianza.

4. Inferencia No Paramétrica

4.1. Contrastes de bondad de ajuste.

- Efectos de la mala especificación de la distribución. Contrastes basados en la distribución multinomial: Test χ^2 de Pearson. Contrastes basados en la función de distribución empírica: Test de Kolmogorov-Smirnov. Contrastes de normalidad: Diagramas probabilísticos normales. Contraste de Shapiro-Wilks, Contraste de Kolmogorov-Smirnov-Lilliefors, Contrastes basados en la asimetría y la curtosis (Jarque-Bera). Transformaciones de Box-Cox para conseguir normalidad. Estimación por máxima verosimilitud de la transformación.

4.2. Contrastes de posición

- El problema de la localización no paramétrica. Test de los signos. Test de los rangos con signo de Wilcoxon.

4.3. Contrastes de independencia

- Consecuencias de la dependencia de los datos. Test de independencia en muestras bivariantes: Test de la τ de Kendall, Test de Spearman de correlación por rangos entre componentes, Test de la independencia en tablas de contingencia. Contrastes de aleatoriedad en una muestra: Concepto de racha. Test basado en el número total de rachas, Test basado en el coeficiente de correlación por rangos de Spearman, Test basado en la τ de Kendall.

4.4. Contrastes de homogeneidad

- Heterogeneidad y sus consecuencias. Efecto de los datos atípicos. Test de valores atípicos. Contraste de homogeneidad relativos a dos o más muestras: Test del tipo Kolmogorov-Smirnov para comparar dos muestras, Test de Mann-Whitney-Wilcoxon y otros análogos, Otros test de comparación basados en la mediana o en las rachas de la muestra conjunta, Test χ^2 de homogeneidad en tablas de contingencia, Test de Kruskal-Wallis

4.5. Introducción a la estimación no paramétrica de curvas

- Consideraciones sobre la estimación no paramétrica de curvas. Estimación no paramétrica de la función de densidad. Histograma de partición fija y aleatoria. Método núcleo. El estimador de Rosenblatt-Parzen. El problema del parámetro de suavizado. Estimación no paramétrica de la función de distribución. Método núcleo.

5. Modelos de Regresión.

5.1. Introducción a los modelos de regresión.

- Dependencia funcional y dependencia aleatoria. Modelos de diseño fijo. Modelos de diseño aleatorio. Algunos ejemplos reales de distintos modelos de regresión (lineal, polinómico, no lineal). Análisis gráfico. Metodología en el estudio de los modelos de regresión: fases de un estudio.

5.2. Modelo de Regresión Lineal Simple: Formulación del modelo e inferencias respecto a sus parámetros.

- Formulación del modelo de regresión simple: Hipótesis básicas. Estimación de los parámetros: métodos de máxima verosimilitud y de mínimos cuadrados. Varianza residual. Interpretación geométrica de la estimación. Propiedades de los estimadores. Intervalos de confianza y contrastes sobre los parámetros. La hipótesis de horizontalidad (contraste de regresión): Interpretación. El contraste de linealidad: el F test. La tabla ANOVA completa. El coeficiente de correlación lineal y el coeficiente de determinación. Inferencias acerca del coeficiente de correlación lineal.

5.3. Modelo de Regresión Lineal Simple: diagnóstico, validación y predicción.

- Diagnóstico y validación del modelo: Análisis de los residuos, Gráficos de residuos, Observaciones con influencia en regresión simple. Transformaciones en regresión simple: Las transformaciones Box-Cox, Estimación de la transformación: método gráfico y por máxima verosimilitud, Transformaciones para conseguir homocedasticidad, Consecuencias de las transformaciones. Estimación de la media condicionada. Predicción de una nueva observación. Introducción al tratamiento del modelo lineal con regresor aleatorio. Observaciones con influencia en regresión simple.

5.4. Modelo de Regresión Lineal General: planteamiento y estimación.

- Formulación del Modelo Lineal General. Hipótesis básicas. Estimación de los parámetros: métodos de máxima verosimilitud y de mínimos cuadrados. Interpretación geométrica. Estimación por etapas. Distribución muestral de los estimadores. Teorema de Gauss-Markov. Varianza residual: propiedades. Estimación de la varianza.

5.5. Modelo de Regresión Lineal General: intervalos de confianza, contrastes y predicción.

- Intervalos de confianza y contrastes para los coeficientes del modelo. Regiones de confianza y contrastes para conjuntos de coeficientes. Interpretación de los contrastes. Intervalos de confianza para la varianza. El coeficiente de determinación y el coeficiente de determinación corregido. Correlación parcial. Relación con la regresión por etapas. Predicción e intervalos de confianza para el valor medio. Predicción e intervalos de confianza para una nueva observación.

5.6. Modelo de Regresión Lineal General: diagnosis y validación.

- Problemas que pueden surgir al construir un modelo de regresión. Multicolinealidad. Estudio de los residuos. Análisis gráfico. Errores de especificación. La hipótesis de normalidad. Robustez del modelo. Observaciones influyentes. Heterocedasticidad. Autocorrelación. Los contrastes de Durbin-Watson y de Ljung-Box. Estimador de mínimos cuadrados generalizados: Aplicación para corregir heterocedasticidad y/o dependencia. Construcción de modelos: selección de variables. Validación del modelo. Análisis de modelos de regresión lineal múltiple mediante un paquete estadístico.

5.7. Generalizaciones de los Modelos de Regresión Lineal.

- Regresión polinómica. Formulación del modelo de Regresión Polinómica. Elección del grado del polinomio. Una introducción al estudio de Superficies de Respuesta. Regresión con variables regresoras cualitativas. Variables ficticias. Inclusión de interacciones entre atributos. El modelo de Análisis de la Covarianza. Modelos de regresión y de diseño experimental: El Modelo Lineal General. Modelos de regresión con respuesta cualitativa. Formulación del problema. El modelo logístico ("logit"). El modelo probit.

PROGRAMA PRÁCTICAS

6. Análisis Exploratorio de Datos.

6.1. Métodos para resumir la información contenida en la muestra.

- Métodos numéricos basados en medidas de localización: mediana, media aritmética, cuadrática, geométrica y armónica, medias recortadas y truncadas, moda y cuantiles. Métodos numéricos basados en medidas de dispersión: varianza y desviación típica, desviación absoluta (versiones truncadas y recortadas), MEDA, coeficiente de variación, recorrido intercuartílico. Medidas de asimetría y curtosis. Métodos gráficos basados en medidas de centralización y dispersión: diagrama de cajas. Aplicación a la identificación de datos atípicos. Métodos basados en la función de distribución empírica y cuantil empírica. Métodos gráficos basados en la estimación de la densidad: histogramas de partición fija y aleatoria. Diagramas de tallo y hojas. Métodos gráficos para resumir datos categóricos: diagramas de tarta, diagramas de barras (caso unidimensional), tratamiento gráfico de tablas de contingencia (caso bidimensional).

6.2. Gráficos dinámicos para análisis de datos

- Representaciones gráficas de datos bidimensionales. Tratamiento visual de grandes muestras: binned plots. Representaciones gráficas de datos multidimensionales: gráficos de Draftman, gráficos de superficies de nivel. Manipulaciones gráficas de datos bidimensionales: etiquetado (localización e identificación), eliminación y linkado de datos o grupos de datos de interés mediante técnicas de cepillado (brushing). Reescalado y utilización de factores de visión (zooms).

6.3. Métodos de suavización de datos

- Suavización del histograma unidimensional: el histograma móvil y el estimador tipo núcleo (estimador de Parzen-Rosenblatt). Influencia y elección del parámetro de suavizado. Suavización de diagramas acumulativos unidimensionales. Suavización tipo núcleo. Influencia y elección del parámetro de suavizado. Suavización de datos

bidimensionales: el regresograma móvil, el estimador tipo núcleo, suavizadores mínimo cuadráticos locales. Influencia y elección del parámetro de suavizado. Suavización de datos multidimensionales: el desastre de la dimensionalidad. Introducción a los suavizadores semiparamétricos.

6.4.Introducción a las técnicas de reducción de la dimensión.

- Técnicas de tipo global: técnicas basadas en la matriz de covarianzas muestral o en la matriz de correlaciones muestral. Análisis de componentes principales. Análisis factorial. Técnicas de tipo local: introducción a la suavización con modelos aditivos. Técnicas de búsqueda de la proyección (Projection Pursuit)

6.5.Análisis exploratorio de datos temporales

- Representaciones gráficas de datos temporales (time plots). Detección visual de varianza inestable, tendencia y estacionalidad. Transformación Box-Cox para estabilizar la varianza de una serie temporal. Determinación de la transformación. Suavización de diagramas temporales: suavizadores polinómicos y suavizadores de medias móviles (filtro lineal) como mecanismos de eliminación de posibles líneas de tendencia (estabilización de la media). Diferenciación ordinaria y estacional de diagramas temporales para transformarlos en datos estacionarios. El análisis exploratorio mediante las funciones de autocorrelación simple (fas) y autocorrelación parcial (fap): el correlograma. Introducción al análisis exploratorio mediante series sinusoidales: el periodograma.Estructura de computador

DESENVOLVEMENTO DO TEMARIO

O programa de teoría baséase en clases semanais, nas que se inclúen as explicacións dos temas e os seminarios de problemas. O programa de prácticas artúculase habitualmente en sesións semanais de dúas horas e estará orientado a practicar do xeito máis directo posible os temas explicados nas clases teóricas.

CRITERIOS E SISTEMAS DE AVALIACIÓN

A avaliación consistirá nunha proba separada en dúas partes: parte teórica e parte de problemas). A proba da parte teórica(30% da nota) consistirá nun cuestionario tipo test. A parte de problemas consistirá na resolución de varios problemas vencellados os contidos da materia.

Para aproba-la materia deberán superarse conxuntamente as dúas partes así como ter completados alomenos o 75% das clases practicas.

BIBLIOGRAFÍA

- D. MONTGOMERY, G. RUNGER (1998), “Probabilidad y Estadística Aplicadas a la Ingeniería”, Mc Graw Hill.
- L. COCKING, W. SMITH (2001), “Á Estadística ¡en caricaturas!”, Publicado pola SGAPEIO.
- J. MARTÍN PLIEGO, L. RUÍZ-MAYA (1995), “Estadística I: Probabilidad”, Editorial AC.
- J. FREUND, I. MILLER, M. MILLER (2000), “Estadística Matemática con Aplicaciones”, Prentice Hall.
- J. DEVORE (2001), “Probabilidad y Estadística para Ingeniería y Ciencias”, Thomson.
- I. UÑA, V. PERUCHA, J. MARTÍN MORENO (2003), “Lecciones de Cálculo de Probabilidades”, Thomson.
- I. UÑA, V. PERUCHA (2003), “Lecciones de Estadística Descriptiva”, Thomson.
- R. CAO et al (2001), “Introducción a la Estadística y sus aplicaciones”. Pirámide
- D. PEÑA (1991), “Estadística; Modelos y Métodos: 1.Fundamentos”, Alianza Universal Textos.
- D. PEÑA (1992), “Estadística; Modelos y Métodos: 2.Modelos lineales y series temporales”, Alianza Universal Textos.
- D. PEÑA (2001), “Fundamentos de Estadística”, Ciencias Sociales Alianza Editorial.
- D. PEÑA (2002), “Regresión y diseño de experimentos”, Ciencias Sociales Alianza Editorial.
- R. WALPOLE, R. MYERS, S. MYERS (1998), “Probabilidad y Estadística para Ingenieros”, Prentice Hall.
- R. VÉLEZ, A. GARCÍA (1994), “Principios de Inferencia Estadística”, UNED.R.A.