

Análisis Estadístico de Datos

Instituto Sabato (CNEA/UNSAM)

Primer cuatrimestre 2024

Profesor: Diego Ravignani, ITeDA (CNEA/CONICET/UNSAM)

Ayudante: Carmina Pérez Bertolli, ITeDA (CNEA/CONICET/UNSAM)

Clases teóricas: Martes y jueves de 14 a 16 h (Hora Argentina)

Clases prácticas: Jueves de 16 a 17 h

Carga horaria: 128 horas (64 teóricas / 64 ejercicios)

Modalidad: A distancia con clases sincrónicas

Comienzo: Martes 5 de marzo de 2024

Duración: 16 semanas

Evaluación: 6 trabajos prácticos y 1 examen final

Programa

Probabilidad discreta: Introducción histórica. Variable aleatoria discreta. Distribuciones de Bernoulli y uniforme. Media, varianza y desviación estándar. Independencia. Probabilidad conjunta, condicional y marginal. Teorema de Bayes. Distribución binomial. Ley de los eventos raros, distribución de Poisson. Suma y convolución de variables aleatorias. Simulaciones Monte Carlo.

Probabilidad continua: Variable aleatoria continua. Densidad de probabilidad, probabilidad acumulativa y cuantiles. Distribución uniforme continua. Suma de variables continuas. Teorema central del límite. Distribución normal o Gaussiana. Correspondencia entre las distribuciones binomial, Poissoniana y normal. Distribución de Cauchy. Cambio de variables. Distribución chi-cuadrado. Histogramas. Proceso sin memoria y distribución exponencial: Decaimiento radioactivo.

Probabilidad multidimensional: Variable aleatoria bidimensional. Probabilidad conjunta, marginal y condicional. Covarianza y correlación. Distribución normal multivariada. Elipse de covarianza. Combinación lineal de variables normales. Cambio de variables. Propagación de la varianza.

Estimación de parámetros Estadística y probabilidad. Muestreo de datos. Estadísticas y estimadores de parámetros. Consistencia, sesgo y eficiencia. Estimadores de la media y la varianza de una distribución normal. Función de verosimilitud. Estimador de máxima verosimilitud. Verosimilitud de una muestra de datos. Verosimilitud normal univariada y multivariada, Poisson y binomial. Cociente de verosimilitudes y función de costo o pérdida. Propiedades de los estimadores de máxima verosimilitud.

Estimación de intervalos: Intervalo de confianza clásico, cinturón de confianza y nivel de confianza. Intervalo de confianza de la media de una distribución normal. Intervalo de Student para muestras pequeñas. Intervalos aproximados y probabilidad de cobertura. Límites. Región de confianza normal bivariada: parámetros de un ajuste lineal. Región de confianza con la función de verosimilitud multivariada: aproximación parabólica y método gráfico.

Test de hipótesis: *Introducción histórica:* La catadora de té. *Test de hipótesis Z:* Aditivo para pinturas. Test de significancia para histogramas. *Test t de Student:* Efecto hipnótico de la hiosciamina. *ANOVA:* Edad de los delincuentes de guante blanco. *Clasificación:* Cáncer de mama. Hipótesis nula y alternativa. Estadístico de prueba. Significancia, nivel de confianza y poder. Región crítica. Errores del primer y segundo tipo. Tests centrados y de cola inferior y superior. p-valor. Test de hipótesis y de significancia.

Ajuste de datos: *Combinación de resultados:* Masa del bosón de Higgs. Muestreo y ajuste de datos. Variable independiente y dependiente. Modelos lineales y no-lineales. El principio de cuadrados mínimos y el método de máxima verosimilitud. Cuadrados mínimos lineal y no-lineal. Ajuste de datos con la función de verosimilitud. Estimadores, errores y bondad del ajuste. Regiones de confianza. Matrices de covarianza y de correlación. Ajuste de un histograma. Evaluación numérica de la performance de un ajuste.

Requisitos

Conocimientos de análisis matemático, álgebra lineal y lenguaje de programación Python.

Bibliografía

- [1] Probabilidad y estadística para ingeniería y ciencias, J. Devore, Paraninfo (2002).
- [2] All of Statistics, L. Wasserman, Springer (2004), doi:10.1007/978-0-387-21736-9.
- [3] Statistical inference, G. Casella and R. Berger, Thomson Press (2006).
- [4] Statistical methods for data analysis in particle physics, L. Lista, Springer Cham (2017), doi:10.1007/978-3-319-62840-0.
- [5] Statistical data analysis, G. Cowan, Clarendon Press (1998).
- [6] Statistics in criminal justice, D. Weisburd and C. Britt, Springer (2014).