

# Análisis Estadístico de Datos

Diego Ravnani  
diego.ravnani@iteda.cnea.gov.ar

2021

**Clases teóricas:** Martes y jueves de 10 a 12 (hora Argentina)

**Carga horaria:** 120 horas (60 teóricas / 60 ejercicios)

**Comienzo:** Martes 17 de agosto

**Duración:** 15 semanas

**Evaluación:** Monografía final

## Descripción

El análisis de los datos generados en diversas ramas de la ciencia, la tecnología y otras disciplinas requiere cada vez más el uso intensivo de herramientas estadísticas. Este curso del Instituto Sabato dependiente de la Comisión Nacional de Energía Atómica y de la Universidad Nacional de San Martín, está dirigido a estudiantes de grado, posgrado y profesionales interesados en el análisis de datos. El mismo provee las herramientas estadísticas necesarias para resolver nuevos problemas además de una sólida base conceptual para poder especializarte luego en campos como, por ejemplo, la inteligencia artificial.

## Contenidos

**Probabilidad discreta:** Introducción histórica. Variables aleatorias discretas. Distribuciones de Bernoulli y uniforme. Media, varianza y desviación estándar. Independencia. Probabilidad conjunta, condicional y marginal. Teorema de Bayes. Distribuciones binomial y Poissoniana. Suma y convolución de variables aleatorias.

**Probabilidad continua:** Variables aleatorias continuas. Densidad de probabilidad, probabilidad acumulativa y cuantiles. Distribución uniforme continua. Teorema central del límite. Distribución Gaussiana o normal. Distribución chi-cuadrado. Distribución de Maxwell-Boltzmann en un gas ideal. Histogramas. Memoria y distribución exponencial: decaimiento radioactivo.

**Probabilidad multidimensional:** Variables aleatorias bidimensionales. Probabilidad conjunta, marginal y condicional. Covarianza y correlación. Distribución Gaussiana multidimensional. Combinación lineal de variables Gaussianas. 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. Función de verosimilitud. Verosimilitud Gaussiana, Poissoniana, binomial y binormal. Verosimilitud relativa y función de costo. Propiedades de los estimadores de máxima verosimilitud.

**Estimación de intervalos:** Intervalo de confianza clásico, cinturón de confianza y nivel de confianza. Intervalo de Student para muestras pequeñas. Intervalos aproximados y probabilidad de cobertura. Límites de confianza: cota superior para un flujo de neutrinos. Región de verosimilitud multidimensional.

**Test de hipótesis:** *Test de significancia:* la catadora de té. *Test de hipótesis:* ensayo clínico de la vacuna covid. Tests de hipótesis e intervalos de confianza. *Análisis discriminante lineal:* clasificación de rayos cósmicos. *Bondad de un ajuste de datos:* test chi-cuadrado y valor p.

**Ajuste de datos:** *Combinación de mediciones:* masa del bosón de Higgs. El principio chi-cuadrado y el método de máxima verosimilitud. Ajuste chi-cuadrado lineal y no-lineal. Ajuste de verosimilitud. Estimadores, errores y bondad del ajuste. Ajuste de un histograma. *Ajuste bivariado:* estimación de la eficiencia con pares de detectores.

## Requisitos

Haber cursado materias de análisis vectorial y de álgebra lineal. Conocimientos de algún lenguaje de computación o software de análisis de datos como Python, C++, ROOT, Matlab, R, etc...

## Bibliografía

- [1] Statistical Data Analysis, G. Cowan, Clarendon Press (1998).
- [2] Probability and Statistics in Particle Physics, A. G. Frodesen, O. Skjeggstad, and H. Tofte, Universitetsforlaget (1979).
- [3] P.A. Zyla et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020). (chapters 39-41)
- [4] Statistical Inference, G. Casella & R. Berger, Thomson Press (2006).
- [5] Knight, Keith. Mathematical Statistics. Chapman & Hall/CRC (2000).

- [6] W. T. Eadie, D. Drijard, F. James, M. Roos, and B. Sadoulet. *Statistical Methods in Experimental Physics*. North-Holland (1971).