

OBTENCIÓN Y ANÁLISIS DE DATOS
Posgrados en Ciencias Ambientales – Facultad de Ciencias – UDELAR
Docente a cargo: Dr. Walter Norbis

Carga horaria: 120

Créditos: 10

Unidad Análisis e Integración

Días y horarios: Martes a Jueves de 08:00 a 10:00, primer clase a las 08:30

Fechas: del 15 de agosto al 15 de diciembre del 2023

OBJETIVOS:

Proporcionar los conocimientos sobre fundamentos teóricos, propiedades, uso e interpretación del análisis estadístico de datos, que pueden ser aplicables a su trabajo de investigación.

CONTENIDO:

MODULO 1

TEMA 1. Niveles de clasificación y ordenación. Atributos y variables continuas, discontinuas y discretas. Ordenación y clasificación. Niveles de información: agrupación, clasificación, adición, sustracción (variables discretas y continuas).

TEMA 2. Estadística descriptiva. Distribuciones de frecuencias (continuas y discontinuas). Discretización de variables continuas (intervalos de clase). Curvas de frecuencia acumulada. Medidas de posición: moda, mediana, media (aritmética, geométrica, armónica) y cuantiles. Principales propiedades. Medidas de dispersión: rango, desviación media, varianza, desvío estándar, coeficiente de variación. Asimetría y kurtosis. Propiedades. Análisis exploratorio de datos: objetivos. Box and Whisker Plot. Interpretación. Otros métodos.

TEMA 3. Experimentos aleatorios y espacio de muestreo: concepto. Variables aleatorias. Definición. Funciones de probabilidad y distribución. Esperanza y varianza. Funciones: Binomial, Poisson y Normal. Funciones de densidad generadas por la normal: Chi-cuadrado, Student y F (Snedecor).

MODULO 2

TEMA 4. Teoría, Técnicas y diseño de muestreo. Obtención de datos: objetivos e hipótesis. Tipos de datos. Variabilidad espacial y temporal. Población o universo. Censo. Muestra. Conveniencia y ventajas del muestreo. Muestreo probabilístico: aleatorio, estratificado (tipos), sistemático. Otros diseños. Ventajas y desventajas. Factores que afectan el

muestreo. Tamaño de muestra y costo. Diseño óptimo. Métodos de re - muestreo: “Jackknife y Bootstrap”.

MODULO 3

TEMA 5. Estimación y test de hipótesis. Teorema central del límite. Parámetros y estimadores: estimador de la media y varianza. Propiedades. Estimador del cociente de varianzas. Hipótesis nula. Errores tipo I y II. Hipótesis alternativa. Potencia de un test.

TEMA 6. Supuestos básicos en estadísticas paramétricas. Aleatoriedad, Independencia, normalidad, homocedasticidad. Test. Transformaciones. Violación de los supuestos y supuestos básicos en estadística no-paramétrica. Ventajas y desventajas. Test de permutaciones y Monte Carlo.

TEMA 7. Análisis de muestras y test de hipótesis. Modelos lineal general y generalizado. Comparación de dos muestras. Planteos básicos y supuestos. Test de Student. Comparación de varias muestras: análisis de varianza (ANVA). Supuestos. ANVA factorial y anidado. Test “a priori” y “a posteriori”. Diseños de experimentos.

TEMA 8. Análisis no – paramétrico: Test Binomial; Chi – cuadrado; Mac Nemar; Test de los signos; Wilcoxon; Fisher; Mann – Whitney; Kolmogorov – Smirnov; Kruskal – Wallis; Friedman; Correlación no - paramétrica: Spearman y Kendall. Otros test.

TEMA 9. Relación de dos variables. Correlación. Regresión lineal. Regresión funcional. Ajuste y validación del modelo. Análisis de varianza del modelo. Análisis de covarianza. Relaciones no – lineales: Regresión potencial y exponencial. Regresión logística.

MODULO 4

TEMA 10. Relación de tres o más variables. Repaso sobre matrices y determinantes: adición y producto de matrices. Determinantes. Rango de una matriz. Matriz inversa y traspuesta. Valores propios y vectores propios. Propiedades. Correlación y regresión múltiple. Modelos lineales generales (GLM) y modelos aditivos generales (GAM).

TEMA 11. Análisis multidimensional. Análisis numérico de datos múltiples: matriz básica de datos. Datos cualitativos y cuantitativos. Transformaciones en análisis multivariado. Clasificación: índices de similitud y de distancia. Propiedades.

TEMA 12. Análisis Factorial. Matrices de dispersión y de correlación. Análisis de Componentes Principales. Interpretación. Análisis Factorial de Correspondencias. Interpretación. Otros análisis derivados: Análisis de Redundancia y factorial de correspondencias. Análisis multidimensional métrico y no métrico. Ventajas y desventajas. Análisis discriminante y canónico de poblaciones.

APROBACIÓN DEL CURSO

Cada estudiante debe presentar a lo largo del curso y de manera obligatoria, un seminario teórico basado en la separata de un trabajo o un tema previamente seleccionado. Este seminario tendrá una nota de calificación. Al final del curso el estudiante deberá realizar un trabajo final escrito que deberá contener una breve introducción del problema a abordar, los objetivos del trabajo, el planteo de hipótesis (y aquellas que puedan ser estadísticamente verificables), una descripción fundamentada de los métodos a utilizar, como serán o fueron obtenidos los datos y un desarrollo completo del análisis de datos y sus resultados. Plazo de presentación del trabajo: febrero- marzo 2024. Además: Asistencia al 75% de las clases teóricas y 90% de las clases prácticas.

BIBLIOGRAFÍA:

(Durante el curso se brinda una lista de bibliografía y separatas que amplían cada uno de los temas que forman parte de cada módulo).

- Cochran, W. (1986) Métodos de muestreo. CECSA, México. (La versión original del año 1977 es en inglés).
- Conover, W. (1980) Practical nonparametric statistic. Second Edition. John Wiley and Sons.
- Crisci, J. y M.F. Armengol (1983) Introducción a la teoría y práctica de la taxonomía numérica. OEA, Serie de Biología. Monografía número 26.
- Cuadras, C. (1993) Análisis Multivariante. EUNIBAR, Barcelona.
- Daniels, W. (1989) Bioestadística. Métodos aplicados a las Ciencias Biológicas y de la Salud.
- Kent, J. (2012). Vegetation Description and Data Analysis: A Practical Approach 2nd Edición.
- Legendre, P. and L. Legendre (2012) Numerical Ecology. 3rd. Edition. Elsevier. 1006 p.
- Manly, B.F.J., 1994. Multivariate Statistical Methods: A Primer. Second Edition. Chapman & Hall/CRC, London, New York
- Plá, L. (1986) Análisis multivariado: método de componentes principales. OEA, Ser. Matemáticas, Monografía número 27.
- Quinn, G. and M. Keough (2002) Experimental Design and Data Analysis for Biologists. Cambridge University Press
- Santaló, L. (1975) Probabilidad e inferencia estadística. OEA, Serie de Matemática. Monografía número 11.
- Sneath, P. and R. Sokal (1973) Numerical taxonomy. The principles and practice of numerical classification. Freeman, San Francisco.
- Siegel, S. (1956) Non – parametric Statistic for the Behavioral Sciences. Mc Graw – Hill.
- Sokal, R. y F. Rohlf (2012) Biometry: the principles and practice of statistics in biological research. W.H. Freeman and Company. 915 p.
- Steel, R. y J. Torrie (1985) Bioestadística: principios y procedimientos. Editorial MacGraw - Hill.
- Snedecor, G. y W. Cochran (1984) Métodos Estadísticos. CECSA, México.
- Zar, J. (2010) Biostatistical analysis. Prentice - Hall International Editions.