

**1. Título del Curso*****ANÁLISIS ESTADÍSTICO MULTIVARIADO*****2. Unidades de Créditos Académicos (UCAs) que otorga:**

4 UCAS (60 hs)

**3. Número de inscriptos admisibles o cupo:**

20 alumnos

**4. Docente coordinador:**

Dr. (Ing. Agr.) Agustín ALESSO

**5. Docentes del curso:**

Dr. Ing. Agr. Agustín ALESSO

**6. Destinatarios:**

Estudiantes de posgrado de carreras afines a la Agronomía, Biología, Ciencias Ambientales con formación básica en estadística (modelos lineales).

**7. Justificación:**

La estadística es la ciencia del aprendizaje a partir de los datos. Una vez definido el problema de investigación, el investigador se enfrenta al desafío de diseñar experimentos o muestreos que permitan recolectar datos pertinentes al problema y, mediante la aplicación de métodos y técnicas estadísticas, generar nuevo conocimiento en un contexto de incertidumbre. En estos experimentos o muestreos frecuentemente las unidades experimentales o individuos son sometidos a diversas determinaciones o mediciones resultando en bases de datos complejas. El análisis multivariado permite analizar conjunto de datos multidimensionales para extraer sintetizar o extraer información sobre la relación entre las variables consideradas.



## 8. Objetivos

- Entender la naturaleza de los conjuntos de datos multivariados y la necesidad de enfoques analíticos integradores.
- Introducir a los alumnos en las técnicas de análisis multivariado, reconociendo sus ventajas, desventajas, aplicaciones y limitaciones.
- Facilitar la introducción de los alumnos al procesamiento y análisis estadístico de datos utilizando el lenguaje estadístico R

## 9. Programa:

- **Unidad 1.** Introducción. Revisión de los conceptos claves del pensamiento estadístico. Estadística descriptiva: gráficos y medidas de resumen. Inferencia estadística: estimación por intervalo de confianza y contraste de hipótesis. Concepto del valor p. Errores. Introducción al lenguaje R.
- **Unidad 2.** Introducción a los métodos multivariados: ventajas, desventajas, aplicaciones, limitaciones. Repaso elementos de álgebra matricial. Descomposición de valores y vectores característicos. Distancias. Distribución normal multivariada.
- **Unidad 3.** Componentes principales. Matriz de varianza y covarianza. Matriz de correlación. Reducción de dimensionalidad. Extracción de componentes principales. Proporción de varianza explicada. Interpretación de loadings. Aplicaciones.
- **Unidad 4.** Análisis factorial. Matriz de Comunalidad. Modelo subyacente y variables latentes (factores). Rotaciones. Selección del número de factores. Aplicaciones.
- **Unidad 5.** Análisis discriminante. Estimación de función discriminante lineal y cuadrática. Criterios de clasificación: criterio de fisher, probabilidades posteriores. Matriz de confusión. Métodos de validación: resubstitución, validación independiente, validación cruzada.
- **Unidad 6.** Análisis de clusters. Distancias multivariadas. Métodos aglomerativos jerárquicos y no jerárquicos. Métodos divisivos. Número de clusters: criterios.
- **Unidad 7.** Intervalos de confianza y pruebas de hipótesis. Prueba Z y t para una y dos muestras independientes. Análisis multivariado de la Varianza. Criterios Hotelling (T2), Wilk (likelihood ratio test), Pillai (Trace), Roy (Maximum Root Criterion).



- **Unidad 8.** Correlaciones canónicas. Maximización de correlación lineal entre dos conjuntos de variables

#### 10. Actividades Prácticas:

Las actividades prácticas se realizarán en gabinete donde los alumnos deberán resolver ejercicios propuestos relacionados con los distintos diseños experimentales mediante el uso del programa R e interpretar los resultados obtenidos. Los ejercicios propuestos y conjuntos de datos, así como también material adicional y código R será brindado a través de la plataforma Entorno Virtual de la UNL.

#### 11. Cronograma de dictado y duración del curso:

La duración total del curso será de 60 hs la cual incluye la elaboración del informe final. Se dictará en tres módulos: uno virtual y dos presenciales de 2 días de duración cada uno. El siguiente cronograma es una estimación sujeta a cambios en función del progreso de los alumnos.

Modulo	Día	Unidad
1	-	1
2	1	2 y 3
	2	3 y 4
3	1	5 y 6
	2	7 y 8

#### 12. Número de horas teóricas:

30 horas

#### 13. Número de horas prácticas, seminarios y realización de trabajo final integrador:

30 horas

#### 14. Sistema de Evaluación:

Al finalizar el curso los alumnos deberán, en base a unas consignas generales, elaborar un trabajo final que consistirá en el análisis e interpretación de resultados a partir de un set de datos reales. La



presentación de los resultados deberá ir acompañada con el código utilizado y salidas obtenidas.

## 15. Referencias Bibliográficas

1. Box, George E. P., J. Stuart Hunter, and William G. Hunter. 2005. *Statistics for Experimenters: Design, Innovation, and Discovery* , 2nd Edition. 2nded. Wiley-Interscience.
2. Crawley, M.J. 2007. *The R book*. Wiley. Chichester, England; Hoboken, N.J.
3. Ott, L.; Longnecker, M. 2010d. 6.a ed. *An introduction to statistical methods and data analysis*. Brooks/Cole Cengage Learning. Belmont, CA. 1296 pp.
4. R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
5. Sokal, Robert R., and F. James Rohlf. 2011. *Biometry*. 4th edition. New York: W. H. Freeman.
6. Wasserstein, R.L. y Lazar, N.A. 2016. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician* 70(2): 129-133.
7. Johnson, D. E. 1998. *Applied multivariate methods for data analysis*. Duxbury Press.
8. Johnson, R.A. & D.W. Wichern. 2007. *Applied multivariate statistical analysis*. 7<sup>th</sup> Ed. Prentice Hall, NJ.
9. Searle, S.R. 1982. *Matrix algebra useful for statistics*. Ed. Wiley, NY.
10. Everitt, B., and T. Hothorn. 2011. *An Introduction to Applied Multivariate Analysis with R*. Use R!