

**Planificación de la Asignatura:** Tópicos Especiales en Áreas Complementarias: Herramientas Informáticas para Análisis de Datos

**Fecha:** 23/10/2024 13:02

**Código:** OP003-5

**Carrera:** Licenciatura en Bioinformática

**Departamento Académico:** Informática

**Docente a cargo:**

**Correo del docente a cargo:** gmerino@ingenieria.uner.edu.ar

**Régimen de Dictado:** Cuatrimestral 1º Cuatrimestre

**Carga Horaria Semanal:** 5 horas semanales

**Carga Horaria Total:** 70 horas

---

**Contenidos Mínimos:**

Entorno R. Tipos de datos y estructuras propias de R. Uso de funciones y paquetes estadísticos R. Uso de herramientas informáticas para generar gráficas y analíticas de estadística descriptiva. Ajuste e interpretación de modelos lineales generales y mixtos. Herramientas para el análisis de datos multivariados y la aplicación de técnicas no paramétricas.

---

**Correlativas Regulares para cursar:**

Tercer año

**Correlativas Aprobadas para cursar:**

Segundo año

**Correlativas Aprobadas para promocionar o rendir el examen final:**

Segundo año

**Objetivo General:**

Que el alumno logre:

1. Utilizar el programa estadístico R como herramienta para el análisis de datos representativos de fenómenos biológicos.
2. Aplicar los conceptos y las estrategias básicas para el análisis estadístico de datos biológicos mediante herramientas informáticas.
3. Utilizar los conceptos, métodos y estrategias de análisis a problemas reales del ámbito biomédico, biotecnológico y multi-ómico.
4. Interpretar los resultados de la aplicación de modelos estadísticos en el contexto de problemas reales con el objetivo de elegir la mejor solución.
5. Fortalecer el uso del marco conceptual-teórico en la capacidad del alumno para plantear modelos o herramientas estadísticas según el problema bajo estudio.
6. Reforzar la capacidad para adquirir conocimientos en forma autónoma así como también para trabajar en forma grupal.

**Objetivos Particulares:**

Que el alumno logre:

1. Adquirir habilidades en el uso del software estadístico R.
2. Conocer y aprovechar todas las ventajas que ofrece R y la gran diversidad de paquetes libremente disponibles.
3. Utilizar las herramientas previamente estudiadas de análisis estadístico en R.
4. Aprender métodos estadísticos avanzados para el análisis de datos bioinformáticos.
5. Analizar datos biológicos de diversa índole.
6. Adquirir habilidades críticas para la evaluación de los modelos y los resultados estadísticos.
7. Interpretar correctamente los resultados obtenidos por distintas estrategias y en el contexto del problema bajo estudio.
8. Diseñar sus propias estrategias de análisis estadístico para grandes bases de datos biológicos.
9. Utilizar con conocimiento y juicio crítico las herramientas computacionales disponibles.
10. Implementar sus propias funciones en R.
11. Reforzar y adquirir lenguaje estadístico útil para la descripción de las tareas de análisis y la comunicación de los resultados obtenidos.

**Programa Analítico:**

- Unidad Temática 1: Introducción a R. Introducción a R y RStudio. Instalación de paquetes específicos. Tipos de datos y estructuras propias de R. Manipulación de datos y uso de funciones. Espacio de trabajo en RStudio y uso de la ayuda de R.
- Unidad Temática 2: Herramientas del análisis exploratorio. Estadística descriptiva en R. Obtención e implementación de gráficos exploratorios y confirmatorios en R. La librería ggplot2.
- Unidad Temática 3: Uso de modelos lineales. Definición de los modelos lineales en R. Regresión lineal y Análisis de la Varianza con el paquete stats. Implementación del método de mínimos cuadrados. Análisis e interpretación de los coeficientes. Gráficas básicas de diagnóstico en R.
- Unidad Temática 4: Modelos lineales avanzados. Teoría de los modelos lineales avanzados. Ajuste de modelos lineales avanzados. Técnicas de diagnóstico en R.
- Unidad Temática 5: Análisis Multivariado en R. Discriminante Lineal. Análisis de componentes principales. Análisis en conglomerados. Árboles de decisión.
- Unidad Temática 6: Estadística No Paramétrica en R. Estimación de la función de distribución acumulada. Análisis de correlación e independencia en R. Uso e implementación de pruebas no paramétricas.

**Listado de Actividades de Formación Práctica:**

Los trabajos prácticos (TP) que se desarrollarán a lo largo de la asignatura se corresponden con sus seis unidades temáticas:

TP1: Introducción a R.

TP2: Estadística descriptiva en R.

TP3: Modelos lineales en R.

TP4: Modelos lineales avanzados.

TP5: Herramientas para el análisis multivariado.

TP6: Estadística no paramétrica en R.

Adicionalmente, se trabajará con dos guías de problemas integradores (GPI) entre los cuales cada grupo de alumnos deberá seleccionar un problema, que deberá resolver utilizando el software estadístico y las herramientas vistas en las unidades abarcadas por cada uno de ellos. La GPI1 abarca los contenidos de las primeras tres unidades mientras que la GPI2, los de las últimas unidades.



**Metodología de Evaluación Durante el cursado:**

La asignatura cuenta con dos instancias distintas de evaluación: a) Resolución de problemas integradores y b) Trabajo Final.

a) Resolución de problemas integradores: las evaluaciones de los informes de resolución de los problemas integradores planteados en las guías de problemas integradores (GPIs) estarán destinadas a acreditar el aprendizaje, la comprensión y la capacidad de relacionar los conceptos desarrollados en las clases e incluidos en la guía de trabajos prácticos (GTP). La evaluación se realizará a partir del informe escrito presentado por cada grupo de alumnos (dos a tres miembros), el cual deberá, al menos, contener las siguientes secciones:

- Introducción: presentación del problema a resolver
- Métodos: descripción de la estrategia y herramientas de análisis utilizada, justificando la elección.
- Resultados: conteniendo los resultados obtenidos en forma de tablas y/o gráficos, incluyendo, cuando fuese necesario, fragmentos del código utilizado en el análisis y su correspondiente salida.
- Discusión: debatiendo acerca de los resultados obtenidos y/o métodos utilizados.

Cada informe será evaluado en una escala del 0 al 100. La evaluación considerará tanto la elección e implementación de la estrategia de análisis como la calidad de la presentación del informe. El examen se considerará aprobado cuando se haya conseguido un puntaje de 60 o superior.

b) Trabajo final: La evaluación del trabajo final consiste de dos partes. Por un lado, se evaluará el informe escrito presentado por el grupo de alumnos y por otro, la defensa oral de dicho informe. Durante el cursado, los alumnos deberán realizar una búsqueda bibliográfica con el objetivo de encontrar un trabajo científico, que sea de su interés, donde se aplique alguna de las metodologías estadísticas estudiadas en la asignatura. La temática y el alcance del trabajo deberán acordarse con el docente responsable antes de comenzar el trabajo. En base a la publicación seleccionada, deberán replicar el trabajo o parte de él que involucra los conceptos de interés. Una vez hecho esto, deberán presentar un informe, con la misma estructura de los informes de resolución de los problemas integradores. Posteriormente, deberán defender su trabajo delante de toda la clase, con una presentación de 15 minutos. Dado que la defensa es grupal, los alumnos deberán organizarse previamente de modo de asegurar la participación de todos los miembros del equipo en la defensa. Tanto el docente como los alumnos participarán luego en una instancia de preguntas/discusión del trabajo (5 minutos), la cual también será evaluada. El trabajo final será evaluado en una escala del 0 al 100 en forma individual. El 60% de los puntos se asignarán en función del informe escrito, por lo que todos los alumnos de un mismo grupo tendrán el mismo puntaje. El 40% restante será de

carácter individual y se obtendrá en función de la participación y calidad de la defensa del informe (30%) y en función de la participación del alumno en el debate de los otros trabajos (10%).

**Metodología de Evaluación en Exámenes Finales:**

Evaluación del alumno regular: La evaluación del alumno regular será de carácter individual y se basará en los dos trabajos integradores desarrollados durante el cursado de la asignatura y en el trabajo final. Cada alumno deberá presentar impreso los informes de los dos trabajos integradores y del trabajo final corregidos acorde a las sugerencias que el docente haya hecho en las evaluaciones correspondientes. Luego, los alumnos deberán defender oralmente sus trabajos, respondiendo las preguntas que el docente considere pertinentes respecto de la metodología utilizada. Para aprobar la asignatura el alumno deberá lograr al menos 60 puntos sobre un total de 100.

Evaluación del alumno libre: Todo aquel alumno que cuente con las materias correlativas aprobadas, podrá rendir de forma libre la asignatura. Para ello, deberá comunicarse de forma fehaciente a la cátedra con al menos 15 días de anticipación su intención de rendir con esta modalidad. La cátedra preparará un examen integrador de todas las unidades, cuya resolución deberá realizarse en computadora, utilizando el programa estadístico R. Los resultados obtenidos deberán organizarse en un informe que deberá entregarse, junto con el código R utilizado para el procesamiento de los datos, en el correspondiente turno de examen. En dicha instancia también deberá defender el trabajo en forma oral ante el tribunal evaluador, con una presentación de 15 minutos, seguida de una instancia de preguntas de 5 minutos. Adicionalmente, se le presentará un conjunto de ejercicios que el alumno deberá resolver en esa instancia para demostrar sus habilidades tanto respecto de los contenidos como del uso del software estadístico. El contenido y presentación del informe, la metodología utilizada, la defensa oral y la resolución de los ejercicios presentados durante la mesa de exámen serán conjuntamente evaluados para obtener una calificación global en una escala del 0 al 100. Para aprobar la asignatura el alumno deberá lograr al menos 60 puntos.

**Condiciones de Regularidad :**

En términos generales, la condición de promoción de esta asignatura implica que el alumno ha alcanzado un nivel de conocimientos, destrezas y habilidades suficientes, asociadas al objetivo de la asignatura. El alumno regular será aquel cuyos conocimientos adquiridos son mínimos, de modo que le permitirán preparar, sin mayores dificultades, el examen final de esta asignatura. Teniendo en cuenta esto, al final del cursado, un alumno al final del cursado de la asignatura puede lograr cuatro condiciones (Regular, Promocionado, Libre por evaluación y Libre por inasistencia) que se describen a continuación:

**a) Regular.**

Para acceder a la condición de regularidad, el alumno deberá cumplir la totalidad de los siguientes requisitos:

- Asistir al menos al 75% de las clases.
- Aprobar la evaluación de los problemas integradores. Logrará esta condición cuando, junto con su grupo de trabajo, haya presentado los dos informes correspondiente, habiendo alcanzado una calificación igual o superior a 60 puntos en ambos informes. En caso de haber desaprobado uno de los dos informes, el grupo podrá acceder a una instancia de recuperatorio donde deberán reevaluar/rehacer el informe siguiendo las correcciones del docente encargado.
- Aprobar la implementación y presentación del trabajo final. Logrará esta condición cuando, junto con su grupo de trabajo, haya presentado el informe correspondiente, habiendo alcanzado una calificación igual o superior a 60 puntos.

**b) Promoción.**

Accederá a la condición de promovido aquel alumno que además de haber alcanzado los requisitos exigidos para la regularidad, haya cumplido con las siguientes condiciones:

Obtener un promedio de 80 puntos en las evaluaciones de problemas integradores.

Obtener una calificación del trabajo final superior a 80 puntos.

**c) Libre por evaluación.**

Se considera libre por evaluación a aquel alumno que:

- acredite asistencia al menos al 75% de las clases.
- obtuvo calificación menor a 60 puntos en alguno de los informes de resolución de los problemas integradores, habiendo utilizado las instancias de recuperatorio, o en la implementación y presentación del trabajo final.

d) Libre por inasistencia.

Se considera libre por inasistencia a aquel alumno que:

- No acredita un mínimo de asistencia a clases del 75%.
- No presentó alguno de los informes de resolución de los problemas integradores, incluso en las instancias de recuperatorio, o del informe del trabajo final, o no participó de la instancia de defensa del trabajo final.



**Bibliografía Principal:**

Cabe destacar que no existe un único libro que cubra por completo los contenidos que involucra la asignatura. Es por ello que se torna necesario consultar a varios títulos que cubren aspectos específicos para cada unidad temática. El contenido teórico debe completarse con aplicaciones y ejemplos que pueden ser tomados tanto de libros como de artículos científicos.

EVERITT, Brian; HOTHORN, Torsten. An introduction to applied multivariate analysis with R. Springer Science & Business Media, 2011.

FISCHETTI, Tony. Data Analysis with R. Packt Publishing Ltd, 2015.

HOTHORN, Torsten; EVERITT, Brian S. A handbook of statistical analyses using R. Chapman and Hall/CRC, 2009.

UGARTE, Maria Dolores; MILITINO, Ana F.; ARNHOLT, Alan T. Probability and Statistics with R. CRC Press, 2008.

VENABLES, W. N.; SMITH, D. M. An Introduction to R: Notes on R, A Programming Environment for Data Analysis and Graphics, v. 3.5. 0. Network Theory, Bristol, UK, 2018.

WEHRENS, Ron. Chemometrics with R: multivariate data analysis in the natural sciences and life sciences. Springer Science & Business Media, 2011.

WEST, Brady T.; WELCH, Kathleen B.; GALECKI, Andrzej T. Linear mixed models: a practical guide using statistical software. Chapman and Hall/CRC, 2014.

ZELTERMAN, Daniel. Applied multivariate statistics with R. Cham: Springer, 2015.

**Bibliografía Complementaria:**

BERRIDGE, Damon Mark; CROUCHLEY, Robert. Multivariate generalized linear mixed models using R. CRC Press, 2011.

FARAWAY, Julian J. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. CRC press, 2016.