**Tipo de actividad**: Taller

**Nombre de la actividad:** Análisis estadístico de datos con R (2021)

**DocenteS ResponsableS:**

* Esp. María Paula DIESER
* Prof. Lorena Verónica CAVERO
* Lic. Laura Beatriz WAGNER
* Lic. Sofía FUNKNER

**Fundamentación:**

Un curso de Probabilidad y Estadística supone un desarrollo teórico de los contenidos, la resolución de ejercicios y situaciones problemáticas, y la utilización de una computadora, ya sea como herramienta para la resolución de tales ejercicios y problemas, para la simulación de experimentos caracterizados por la aleatoriedad, o para el análisis de un conjunto de datos reales. Habitualmente las dos primeras se desarrollan de manera conjunta e interrelacionada pero el uso de la computadora no se implementa adecuadamente por diversas razones (dificultades en el acceso a hardware y software para atender las necesidades de todos los estudiantes, escasa disponibilidad de tiempos académicos para el tratamiento de los temas involucrados, entre otros). En este contexto, los equipos docentes de las asignaturas Probabilidad e Inferencia Estadística1 vienen impulsando, desde 2010, distintas propuestas para los/las estudiantes que cursan estos espacios curriculares correspondientes a los planes de estudio de Profesorado en Matemática y Licenciatura en Matemática. Las mismas han ido evolucionando con el transcurrir de los años, en búsqueda de enfoques innovadores, hasta presentar un diseño modular basado en la resolución de problemas auténticos y el desarrollo de proyectos con datos reales. La propuesta incluye un conjunto de actividades que favorecen la apropiación de los procedimientos básicos para operar el lenguaje de programación R a través del paquete R Commander que, mediante una interfaz gráfica, facilita el uso del programa de base.

Hasta 2018 tales actividades se llevaron a cabo en horarios extracurriculares sujetos a la disponibilidad horaria del equipo docente y los/as estudiantes que deciden participar de la propuesta. Surgió entonces la necesidad de sistematizar la propuesta bajo la modalidad de taller que permita a los estudiantes acreditar los aprendizajes alcanzados. Dicha iniciativa viene desarrollándose desde 2019 en forma ininterrumpida.

**Objetivos:** El propósito general del taller es introducir al/la estudiante en la utilización del programa R para la resolución de problemas estadísticos sencillos.

Así se pretende que, al finalizar el taller, los/las estudiantes sean capaces de:

* Comprender el funcionamiento general del entorno R.
* Utilizar R Commander para facilitar la operación de R.
* Identificar las principales estructuras de datos de R y valorar la utilidad de los data frame para el almacenamiento de datos estadísticos.
* Comprender y utilizar las principales sentencias de programación de R para el análisis estadístico de datos univariados y bivariados.
* Emplear las funcionalidades de R e interpretar las salidas correspondientes para:

a. Realizar análisis descriptivos de un conjunto de datos univariados y bivariados.

b. Producir algunas de las representaciones gráficas más comunes del análisis estadístico de datos univariados y bivariados.

c. Realizar análisis inferenciales clásicos paramétricos y no paramétricos de un conjunto de datos univariados y bivariados.

**Arancel:** No arancelado

**Modalidad:** virtual

El trabajo en el Taller será abordado desde un enfoque teórico – práctico y combinando los encuentros sincrónicos por videoconferencia con las posibilidades que ofrecen los entornos virtuales de enseñanza y aprendizaje y otras tecnologías digitales actuales como mediadoras del proceso de enseñanza y de aprendizaje.

La propuesta didáctica incluirá una serie de materiales y recursos digitales, diseñados ad hoc o reutilizados, que serán distribuidos a través de diferentes medios: (a) visual (textos, imágenes, figuras, gráficos); (b) audiovisual; y (c) hipermedia. Parte de ellos conformarán un cuerpo de materiales obligatorios de carácter único (común a todos los destinatarios). Adicionalmente, se ofrecerán otros materiales y recursos distribuidos bajo una estrategia opcional, pudiendo el estudiante optar por un determinado medio en función de sus habilidades y necesidades.

El cuerpo de materiales y recursos se distribuirá a través de un curso online en el entorno virtual de enseñanza y aprendizaje Moodle de la Facultad de Ciencias Exactas y Naturales. Dicho entorno permitirá el acceso a información codificada, de forma rápida e interactiva, favoreciendo la construcción del conocimiento a través de diversos recursos y actividades, permitiendo además la incorporación de otras estrategias y herramientas novedosas para favorecer la comprensión.

Se emplearán dinámicas que promuevan el trabajo colaborativo y la participación activa de los estudiantes con un fuerte apoyo de medios y materiales para la autogestión del aprendizaje. Para el desarrollo de estas actividades se emplearán herramientas disponibles en Moodle (e.g. foro, taller, tarea, base de datos) en conjunto con otras desagregadas (e.g. herramientas de Google).

Se realizará un seguimiento permanente de las actividades propuestas, atendiendo al carácter formativo de la evaluación e incorporando al error como parte del proceso de aprendizaje (reconstrucción a partir del error). Para los diferentes módulos indicados en el programa, se propondrán una serie de actividades de acreditación que los estudiantes deberán resolver individual o grupalmente (según corresponda). Estas actividades estarán mediadas por tareas de Moodle.

Las tutorías combinarán encuentros sincrónicos y un sistema de tutorías online cuyos objetivos generales y características se detallan a continuación:

− Los encuentros sincrónicos se desarrollarán mediante aplicaciones de videoconferencias (e.g. Zoom). Tendrán el objetivo didáctico de introducir los temas a trabajar en cada unidad, y realizar actividades de taller de resolución de problemas y aplicaciones. En este sentido, se abordarán casos y ejemplos cotidianos vinculados con problemas reales que permitan abordar los conceptos y procedimientos vinculados con los contenidos detallados en el programa del Taller.

− Las tutorías asincrónicas estarán mediadas por mensajería y foros de comunicación. Su objetivo será acompañar a los estudiantes en el proceso de aprendizaje, ofreciendo las ayudas y orientaciones necesarias para el abordaje de las diferentes temáticas y la realización de las actividades propuestas.

Particularmente, se incluirá un Foro General de Consultas que será coordinado por el equipo docente en el que los estudiantes pueden, no solo comunicar sus dudas, sino además participar respondiendo a las consultas de sus compañeros. En el caso particular de las actividades colaborativas que impliquen el uso de herramientas de Google (particularmente documentos compartidos de Google), los docentes mantendrán un rol de observadores del proceso interviniendo, en caso de ser necesario, para encauzar el trabajo hacia el logro de los objetivos propuestos.

**Programa:**

El programa del curso se estructura en 6 unidades distribuidas en 3 módulos. Los módulos no son consecutivos y el estudiante puede realizar diferentes itinerarios de aprendizaje como se muestra en la Figura 1.



Figura 1. Itinerario de aprendizaje del taller.

**Módulo 1 Introducción a R y R Commander**

1. Primeros pasos en R. Instalación y administración de R. Cálculos y objetos en R. Estructuras de datos. Funciones predefinidas. Entrada y salida de datos.
2. El entorno R Commander. Adquisición y sesión de R Commander. Creación, importación, exportación y manipulación de datos.

**Módulo 2 Análisis exploratorio de datos**

3. Análisis exploratorio de datos univariados. Estadísticos resumen. Tablas de distribución de frecuencias. Gráficos: sectores, barras, tallo y hojas, histogramas, diagramas de puntos, caja y brazos.

4. Análisis exploratorio de datos bivariados. Factores como elemento de agrupamiento para análisis exploratorio de datos cuantitativos. Análisis de tablas de contingencia. Gráficos de dispersión.

**Módulo 3 Elementos de inferencia clásica**

5. Inferencia clásica. Inferencias sobre una población: TH e IC para medias, varianzas y proporciones. Inferencias sobre dos poblaciones: TH e IC para diferencias de medias, cocientes de varianzas y diferencias de proporciones.

6. Pruebas chi cuadrado. Pruebas de bondad de ajuste, independencia, y homogeneidad.

**Bibliografía:**

Arriaza Gómez, A. J. y otros. (2008). Estadística Básica con R y R–Commander. Servicio de Publicaciones de la Universidad de Cádiz. Cádiz.

Braun, W. J. y Murdoch, D. J. (2007). A first course in statistical programming with R. Cambridge University Press. New York.

Cavero, L. V., Dieser, M. P. y Martín, M. C. (2013) Introducción al lenguaje y ambiente R para el análisis estadístico de datos. Facultad de Ciencias Exactas y Naturales - UNLPam.

Paradis Emmanuel. (2002). R para principiantes. Institut des Sciences de l’Évolution, Universit Montpellier II. Francia. Disponible en: <http://cran.r-project.org/doc/contrib/rdebuts_es.pdf>

R Development Core Team. (2000). Introducción a R. Disponible en: http://www.r-project.org/ R Development Core Team. (2007). R: A Language and Environment for Statistical Computing. Disponible en: <http://www.r-project.org/>

R Development Core Team. Sitio oficial de R. (2010) http://www.r-project.org/

**Carga horaria:** El taller tendrá una carga horaria total de 40 horas reloj, distribuidas en 6 encuentros sincrónicos por videoconferencia de 2 horas reloj cada uno y las restantes horas destinadas a actividades bajo tutoría online.

Las fechas de inicio y finalización de cada uno de los módulos del Taller (Figura 1) se detallan en la Tabla 1 y pueden estar sujetas a modificación. La distribución de encuentros para cada módulo y horarios se detallan en la Tabla 2.



Figura 1. Itinerario de aprendizaje del taller.

|  |  |  |
| --- | --- | --- |
| **MÓDULO** | **INICIO** | **FINALIZACIÓN** |
| 1. Introducción a R y R Commander
 | 16 de agosto | 27 de agosto |
| 1. Análisis exploratorio de datos
 | 30 de agosto | 17 de septiembre |
| 1. Elementos de inferencia clásica
 | 11 de octubre | 12 de noviembre |

Tabla 1. Fechas tentativas de inicio y finalización de cada módulo del Taller.

|  |  |
| --- | --- |
| **MÓDULO** | **ENCUENTROS SINCRÓNICOS** |
| **Fecha** | **Horario** |
| 1. Introducción a R y R Commander
 | 24 de agosto | 15:00 a 17:00 |
| 1. Análisis exploratorio de datos
 | 31 de agosto | 15:00 a 17:00 |
| 7 de septiembre | 15:00 a 17:00 |
| 1. Elementos de inferencia clásica
 | 12 de octubre | 15:00 a 17:00 |
| 2 de noviembre | 15:00 a 17:00 |
| 9 de noviembre | 15:00 a 17:00 |

Tabla 2. Fechas y horarios de encuentros sincrónicos del Taller.

La fecha de entrega del trabajo final es el 26 de noviembre.

**Destinatarios:** Estudiantes de Profesorado en Matemática o Licenciatura en Matemática matriculados en (o que hayan regularizado) Inferencia Estadística o asignaturas equivalentes.

**Cupo:** 20 estudiantes.

**Requisitos de aprobación:** Se otorgará **certificado de aprobación** a aquellos/as estudiantes que completen las actividades finales propuestas para cada uno de los tres módulos.