

PROGRAMA DE CURSO
Análisis Exploratorio de Datos
XS0124

Grupo y Profesor:	Horario de clases:	Aula:	Atención a estudiantes:
G-01 Natalia Díaz Ramírez	L: 07-08:50 / M: 07-08:50	015 CE	K: 17:30 – 19:30
G-03 Eugenia Gallardo Allen	L: 13-14:50 / M: 13-14:50	015 CE	L: 15:00 – 17:00
G-04 Gilbert Brenes Camacho	K: 07-08:50 / V: 07-08:50	015 CE	K: 14:00 – 16:00
G-05 Shirley Rojas Salazar	K: 09-10:50 / V: 09-10:50	015 CE	K, V: 11:00 – 12:00
G-06 Karen Corrales Bolívar	L: 07-08:50 / M: 07-08:50	Sala NIDES ME*	L: 17:00 – 19:00
Requisitos: Ninguno	Correquisitos: Ninguno	Ciclo: I-2025	Créditos: 4
Modalidad curso: Presencial		Plataforma: Mediación Virtual	
Coordinadora de Cátedra: Shirley Rojas Salazar		Correo: shirleyelena.rojas@ucr.ac.cr	

*La sala informatizada de NIDES se encuentra en el primer piso de la Facultad de Medicina

Descripción

Este curso constituye el primer acercamiento con la carrera de Bachillerato en Estadística, de modo que pretende brindar a las personas estudiantes los conceptos básicos de la disciplina a través de actividades prácticas que fomenten el razonamiento estadístico. Se discute el papel fundamental de la Estadística en los procesos de investigación científica y en la toma de decisiones en general. También se abordan las diversas formas de recolección de datos, las diferentes maneras de visualización de información estadística, así como conceptos relacionados con la aleatoriedad, la variabilidad, inferencia estadística y técnicas para realizar análisis exploratorio de datos. Para trabajar los elementos prácticos del curso se utilizan casos de situaciones reales y conjuntos de datos con la ayuda de software como Excel, paquetes estadísticos y el lenguaje de programación R. Además, a lo largo del curso se realizan actividades para desarrollar habilidades relacionadas con la investigación científica y la comunicación de resultados de forma oral y escrita.

Objetivo general

Aplicar conceptos básicos y técnicas estadísticas para el análisis exploratorio de los datos mediante visualizaciones y medidas resumen, para la toma de decisiones eficiente en diferentes áreas de conocimiento, tanto a nivel científico como técnico, y para empezar a desarrollar el razonamiento estadístico.

Objetivos específicos

Al finalizar el curso el/la estudiante estará en la capacidad de:

1. Conocer el papel que juega la Estadística en las diferentes etapas de una investigación científica y su interrelación con otras disciplinas para tomar una decisión de forma eficiente.
2. Conocer los diversos métodos de recolección para garantizar una buena calidad en los datos.
3. Utilizar conceptos estadísticos, medidas de tendencia central y variabilidad, y diversas formas de visualización en un análisis exploratorio para describir y mostrar la tendencia de los datos.
4. Presentar resultados estadísticos de manera oral y escrita tanto en artículos académicos como en informes técnicos.

Habilidades y conocimientos (perfil de salida)

Habilidades	Conocimientos
HE04 - Aprender de forma autónoma metodología estadística	CE06 - Conocimientos básicos en la exploración y actualización de metodologías estadísticas CE07 - Conocimientos básicos en técnicas de estudio CE08 - Conocimiento de estrategias de aprendizaje autodidacta

HT01 - Capturar, visualizar, procesar y analizar datos estructurados y no estructurados	CT03 - Conocimientos avanzados en herramientas de recolección de datos (ej: CSPro, DMC, Survey to Go, limesurvey) CT04 - Conocimientos avanzados en manejo de herramientas de construcción de gráficos en capas (ej: ggplot2, plotly, matplotlib, seaborn)
HT03 - Crear y mantener archivos de datos y usar sistemas de bases de datos	CT14 - Conocimientos intermedios en limpieza y revisión de inconsistencias en archivos de datos
HI05 - Contextualizar el problema de investigación y los resultados al campo de aplicación HI06 - Aprender elementos del campo de aplicación de forma autónoma HI07 - Comprender artículos científicos tanto de estadística como de disciplinas sustantivas a las cuales aplica la estadística	CI07 - Conocimiento de estrategias de aprendizaje autodidacta CI08 - Conocimientos intermedios de técnicas de lectura (en español e inglés)
HI08 - Evaluar la calidad de los datos y resultados	CI09 - Conocimientos avanzados en evaluación de fuentes de información CI10 - Conocimientos avanzados en técnicas para mejorar la calidad de los datos
HI09 - Identificar fuentes bibliográficas confiables	CI12 - Conocimientos intermedios de citación CI13 - Conocimientos básicos de búsqueda de fuentes bibliográficas (SIBDI)
HC01 - Expresar y transmitir conocimientos técnicos mediante lenguaje adaptado al público meta HC02 - Comunicarse con profesionales de otros campos para entender sus necesidades de información	CC01 - Conocimientos básicos en técnicas para la comunicación oral efectiva
HC04 - Comunicar asertivamente mediante la manipulación creativa de datos, pero manteniendo su integridad	CC04 - Conocimientos avanzados en construcción de cuadros y gráficos CC05 - Conocimientos básicos para construcción de infografías, mapas conceptuales y otras técnicas emergentes
HC05 - Comunicar adecuadamente de forma escrita diferentes elementos del quehacer estadístico	CC06 - Conocimientos intermedios de técnicas de escritura de artículos e informes técnicos

Contenidos

1. La Estadística y su relación con la investigación científica:
 - a) Breve historia y definición de Estadística. Ejemplos de aplicaciones de diseños experimentales, encuestas, censos en diferentes áreas del conocimiento.
 - b) La Estadística y la lógica de investigación científica.
 - c) Esquema general de una investigación.
 - d) Diseños de investigación científica: estudios transversales, longitudinales, observacionales, experimentales.
 - e) Tipos de investigaciones: exploratoria, descriptiva y explicativa.
 - f) Conceptos básicos: problema, objetivos, unidad estadística, características y observación, atributos y variables, tipos de variables y niveles de medición, población, censo, muestra, muestreo probabilístico (selección de una muestra simple al azar con y sin reemplazo), marco muestral, error de muestreo, sesgos, y muestreo no probabilístico (intencional y por conveniencia), parámetros y estimadores.
2. Obtención de información estadística y recolección de datos:
 - a) Datos existentes y no existentes.
 - b) Fuentes primarias, secundarias y necesidad de evaluar las fuentes de datos.
 - c) Técnicas para la recolección de datos no existentes: El cuestionario y otros instrumentos para la recolección de datos. Conceptos de validez y confiabilidad de los instrumentos de medición.
 - d) Codificación y digitación, empleo de paquetes computacionales para la captura de información.
 - e) Estructura y formatos de datos.
3. Medidas para resumir datos:
 - a) Resumen de información cuantitativa en distribuciones de frecuencias.
 - b) Cálculo e interpretación de medidas de tendencia central (moda, mediana, media aritmética). Media aritmética simple y ponderada, propiedades de la media aritmética. Relación de las medidas de tendencia y simetría de la distribución.
 - c) Media geométrica y armónica, cálculo y usos.
 - d) Cálculo e interpretación de medidas de posición (cuantiles).

- e) Cálculo e interpretación de medidas de variabilidad (recorrido, rango intercuartil, desviación media, varianza, desviación estándar y coeficiente de variación).
 - f) Cálculo e interpretación de razones, proporciones, porcentajes, tasas, índices, modelos de crecimiento, escalas y otros indicadores.
 - g) Análisis bivariado: asociación y correlación entre dos variables.
 - Asociación entre variables nominales
 - Correlación entre variables ordinales (Spearman)
 - Correlación entre variables métricas (Pearson), cálculo e interpretación de coeficientes.
 - h) Correlación y causalidad
 - i) Exploración inicial de datos y limpieza de archivos de datos utilizando medidas resumen.
4. Presentación y visualización de datos estadísticos:
- a) Formato para la publicación de datos estadísticos.
 - b) Cuadros estadísticos: formato e interpretación. Adecuada interpretación de cuadros univariados, bivariados y trivariados.
 - c) Gráficos estadísticos: su construcción e interpretación. barras y columnas – simples, compuestas, comparativas, 100%-, circular, aritméticos, áreas y semilogarítmico.
 - d) Representación gráfica de distribuciones de frecuencias (histogramas y densidades).
 - e) Diagrama de caja.
 - f) Diagrama de dispersión.
 - g) Gráficos y cuadros para publicaciones académicas o reportes (presentación oral y escrita).
 - h) Otras formas de presentación de información estadísticas: infografía
 - i) Exploración inicial de datos y limpieza de archivos de datos utilizando visualizaciones.
 - j) Estrategias para la presentación oral y escrita de análisis estadístico.

Metodología

El curso es presencial. Se imparten clases teóricas y prácticas en el laboratorio. En las prácticas de laboratorio se usa mayoritariamente el lenguaje de programación R. Para la introducción de cada contenido se van a desarrollar actividades que fomenten el razonamiento estadístico. Se utilizan problemas sobre situaciones reales y cotidianas para que los y las estudiantes propongan soluciones y comprendan la utilidad de la estadística. Además, se realiza un trabajo práctico en el cual se desarrolla un análisis exploratorio de datos donde se aplica los conceptos, técnicas y presentación de la información estadística aprendidas durante el curso. El trabajo práctico cuenta con un componente escrito y una exposición oral por parte del estudiantado.

Evaluación

1. Tres exámenes parciales. Se evalúan tanto aspectos teóricos como prácticos.
 - I Parcial: Historia y conceptos básicos, obtención de información estadística (15%)
 - II Parcial: Medidas resumen de datos (20%)
 - III Parcial: Examen práctico de visualización de datos (15%)
2. Resolución de asignaciones (casos prácticos, laboratorios evaluados, quices, pruebas cortas, otras asignaciones): se asignan ejercicios que contengan datos de situaciones reales o que hayan sido recolectados en el contexto del aula y su solución se puede realizar utilizando Excel, paquetes estadísticos o el lenguaje de programación R, además de actividades dirigidas a fortalecer la comunicación oral y escrita. (25%)
3. Trabajo práctico: se aplican todos los conocimientos adquiridos en el curso, se realiza un análisis exploratorio con datos recolectados por los estudiantes, el cual se entrega mediante un reporte escrito básico y se hace una presentación oral. Se realizarán entregas de avances. (25%)

Fechas de evaluaciones para cada grupo:

Evaluación	G1	G3	G4	G5	G6
I Parcial	M 9 abril	M 9 abril	K 8 abril	K 8 abril	M 9 abril
II Parcial	M 28 mayo	M 28 mayo	V 30 mayo	V 30 mayo	M 28 mayo
Trabajo	L 23 junio / M 25 junio	L 23 junio / M 25 junio	K 24 junio / V 27 junio	K 24 junio / V 27 junio	L 23 junio / M 25 junio
III Parcial	M 2 julio	M 2 julio	V 4 julio	V 4 julio	M 2 julio
Ampliación			V 11 julio 8am		

Las fechas de reposición de las evaluaciones se comunicarán posteriormente a las personas con una solicitud de reposición aprobada. La solicitud debe cumplir con lo establecido en el **Artículo 24** del Reglamento de Régimen Académico Estudiantil.

Cronograma

Semana	Contenido (tema)	Otros detalles
1	T1	
2	T1	
3	T2	
4	T2	
5	T2	Parcial 1 (T1 y T2)
*	-	Semana Santa
6	T3	Semana U
7	-	Actividad evaluada* # 1
8	T3	
9	T3	
10	T3	Actividad evaluada* # 2
11	-	Parcial 2 (T3)
12	T4	
13	T4	Actividad evaluada* # 3
14	T4	
15	-	Entrega de trabajo y exposiciones
16	-	Parcial 3 / Actividad evaluada* #4

*Las fechas de las actividades evaluadas pueden cambiar dependiendo del progreso en la cobertura de los contenidos. El/la docente les comunicará con anticipación si hay cambios en las fechas de dichas actividades.

Bibliografía

- Aguilar, E. y Zamora, A. (2020). Introducción a la Estadística Descriptiva con R. Editorial Universidad Nacional.
<https://www.euna.una.ac.cr/index.php/EUNA/catalog/view/255/27/24>
- Agresti et al. (2018) The Art and Science of Learning from Data. Editorial Pearson.
- Anderson, M. (1992). The history of women and the history of statistics. Journal of Women's History, 4(1), 14-36. Johns Hopkins University Press.
- Anderson, M. (1992). The history of women and the history of statistics. Journal of Women's History, 4(1), 14-36. Johns Hopkins University Press.
- Gómez Barrantes, M. (2016). Elementos de Estadística Descriptiva. Tercera edición. Editorial EUNED. Costa Rica.

SIGNATURA SIBDI: 310 G633eL5

- Gómez Villegas, M., & de Mora Charles, M. S. (2018). Historia de la probabilidad y de la estadística. UNED - Universidad Nacional de Educación a Distancia.
<https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/lc/sibdi/titulos/48925>
- Hernández Rodríguez, Óscar (2016), Estadística Elemental para Ciencias Sociales. Editorial de la Universidad de Costa Rica, cuarta Edición.
SIGNATURA SIBDI: 300.21 H557e4
- Ross, S. (2010). A first course on probability. New Jersey: Prentice Hall.
- Santos del Cerro, J. (Dir.). (2017). Historia de la probabilidad y la estadística (VIII). Delta Publicaciones.
<https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/lc/sibdi/titulos/169702>
- Otero, H. (2018). Socio-history of statistics on Latin America: a review. Histoire & Mesure, 33(2), 13-32.

Reglamentación

- La reglamentación sobre sus deberes y derechos como estudiante se encuentra en el **Reglamento de Régimen Académico Estudiantil**
https://www.cu.ucr.ac.cr/normativ/regimen_academico_estudiantil.pdf
- La reglamentación y sanciones ante fraudes en las evaluaciones o comportamientos anómalos por parte de los y las estudiantes, la pueden encontrar en **Reglamento de Orden y Disciplina de los Estudiantes de la Universidad de Costa Rica** https://www.cu.ucr.ac.cr/normativ/orden_y_disciplina.pdf
- Con el fin de garantizar un espacio libre de violencia y sexismo en el desarrollo de este curso, les recomiendo que revisen **el Reglamento de la Universidad de Costa Rica contra el Hostigamiento Sexual**
https://www.cu.ucr.ac.cr/normativ/hostigamiento_sexual.pdf