

ESTADÍSTICA Y HERRAMIENTAS COMPUTACIONALES PARA LA ANALÍTICA DE DATOS

PERIODO: 21 de Junio a 14 de Julio de 2022

HORARIO: Martes y Jueves, 6:00 a 8:00 pm

NOMBRE PROFESOR MAGISTRAL: Manuel Alejandro Segura Delgado

CORREO ELECTRÓNICO: ma.segura10@uniandes.edu.co

HORARIO Y LUGAR DE ATENCIÓN: Virtual sincrónico

REPOSITORIO DEL CURSO: https://github.com/asegura4488/Estadistica_Herramientas

II Contenido por semanas

Semana 1

■ Temas:

1. Presentación del curso. Introducción a la sintaxis básica de **Python**. Uso de **Python** en la nube. Lectura y escritura de archivos. Extracción de información de interés usando el paquete numérico **numpy**.
2. Graficación de problemas asociados a física y finanzas en dos y tres dimensiones usando el paquete **matplotlib**. Ejemplos diversos y su presentación.

Semana 2

■ Temas:

1. Muestreo aleatorio simple, estratificado y sistemático de una población. Definición de variable aleatoria, interpretación frecuentista de probabilidad y funciones de densidad de variable discreta. Definición del objeto histograma.
2. Ejemplo usando **pandas** para el manejo eficiente de datos teniendo en cuenta: pre-procesamiento, re-formato de tablas y normalización. Uso de **SQL** en **Python**.

Semana 3

■ Temas:

1. Definición del valor esperado. Distribuciones: uniforme, binomial, multinomial y de Poisson. Ejemplos de conteo en física experimental (radiación cósmica de fondo) y media móvil pesada (pronóstico optimizado de ventas) usando **Scipy** y **Pandas**.
2. Distribución de probabilidad en varias variables, distribución conjunta, independencia de sucesos y teorema de Bayes. Definición de varianza y covarianza.

Semana 4

■ Temas:

1. Funciones de densidad de variable continua como: distribuciones exponencial, gaussiana, log-normal, chi-cuadrada. Uso del paquete **scipy** para el cálculo de percentiles y de probabilidades de estas distribuciones.
2. Estudio de la propagación del COVID-19 en Bogotá usando **Pandas**. Uso de cortes, tablas de contingencia y presentación de resultados con gráficos de barras y diagrama circular.

Semana 5

- Temas:

1. Estimación de parámetros. Método de máxima verosimilitud, estimación puntual y en intervalos de confianza.
2. Propagación de errores y su presentación. Ajuste de datos usando el método de mínimos cuadrados matriciales.

Semana 6

- Temas:

1. Estimación de parámetros. Método de re-muestreo (bootstrapping). Teorema del límite central y de cambio de variable.
2. Estimación de parámetros. Cadenas de Markov, método de metropolis-hastings. Modelamiento de efectos sistemáticos en las estimaciones. Ejemplo de conteo en física de partículas y de regresión de datos.

Semana 7

- Temas:

1. Pruebas de hipótesis. Definición de una hipótesis y del proceso de inferencia, error tipo-I y error tipo II. Significancia estadística de una observación, definición del p-valor (p-value) y su correcta interpretación.
2. Pruebas normales y chi-cuadrada, definición de valores críticos y toma de decisión. Diversos ejemplos de aplicación.

Semana 8

- Temas:

1. Pruebas de hipótesis: para la diferencia de medias, proporción, diferencia de proporciones, estimación de la varianza poblacional.
2. Cálculo de valores críticos para modelos de conteo multicanal. Teorema de Neyman-Pearson y su uso en la comparación de modelos.

III Criterios de evaluación

Cada semana habrá un ejercicio teórico y/o práctico para entregar en grupos de máximo 2 personas, el cual tiene el siguiente porcentaje de calificación:

- Entrega de código fuente que incluye las funciones, algoritmos y procedimientos que se solicitan en el enunciado: 30 %.
- El código corre sin errores dentro del intérprete de **Python**: 30 %.
- El código produce la respuesta correcta: 40 %.

Todas las entregas son **en los grupos de trabajo**. El trabajo no incluye la colaboración de otras personas no inscritas en el curso. **Las únicas fuentes autorizadas** para reutilización de código son: el repositorio del curso y los libros de la bibliografía principal.

IV Bibliografía

- *Introduction to statistical with python*. Thomas Haslwanter. Springer 2016.
- *Introduction to probability and statistics for engineers and scientists*. Sheldon M. Ross. Third Edition, Elsevier academic press.

- *Data Analysis: A Bayesian tutorial*. D. S. Sivia, J. Skilling. Second edition, Oxford Science Publications.
- *A survey of Computational Physics - Enlarged Python Book* . R. H. Landau, M. J. Páez, C. C. Bordeianu. WILEY. 2012. <https://psrc.aapt.org/items/detail.cfm?ID=11578>
- *Elements of Scientific Computing* Tveito A., Langtangen H.P., Nielsen B.F., Cai X. Spinger. 2010.