METODOLOGÍA DE ANÁLISIS EN OPINIÓN PÚBLICA

Universidad de Buenos Aires
Facultad de Ciencias Sociales
Carrera de Ciencia Política

Profesor Adjunto: Tomás A. Olego

Modalidad de la Asignatura

Materia Electiva

Período de Cursada

- Inicio: Lunes 5 de agosto del 2019
- Fin: Sábado 23 de Noviembre del 2019
- Cantidad de Semanas: 16 semanas
- Cantidad de Clases: 29 clases (considerando los feriados)

Carga Horaria Total

96 Horas-Reloj. La cursada se estipula dos veces por semana con clases de tres horas diarias.

Presentación y Fundamentación

La presente asignatura versa sobre los métodos y técnicas de investigación propias de la opinión pública. El propósito de la materia es introducir al estudiante un conjunto de herramientas metodológicas que le permitan articular los contenidos teóricos propios de la orientación de opinión pública junto con los requerimientos empíricos del analista y/o investigador. Para ello la materia cuenta con un fuerte contenido teórico estadístico y una práctica orientada a la implementación computacional en Python de los métodos en problemas de la opinión pública.

La asignatura se estructura en catorce unidades. La primer unidad está destinada a permitirle al alumno instalar Python y aprender a realizar informes automáticos para consultoría. La segunda unidad repasa lo aprendido en Metodología I orientado a la O.P. La tercera unidad se centra en la manipulación de datos con bases de fuentes secundarias, sin necesidad de usar ofimática y preparando al alumno para atacar grandes volúmenes de datos. La cuarta unidad dota al estudiante de herramientas de visualización de datos. La quinta unidad introduce la lógica de Bases de Datos, lo que le permitirá estructurar futuros análisis. La unidad seis introduce al alumno a la programación y se le demuestra su utilidad a la hora de encarar el análisis de una encuesta. La séptima unidad propone sentar las bases de la geocomputación lo que le habilita a realizar trabajos de supervisión de campo y muestreo, al igual que espacializar los resultados obtenidos en las encuestas. La siguiente unidad le presenta los fundamentos de la estadística y la novena le enseña los rudimentos del muestreo. La décima unidad, central a la materia, introduce al alumno a la estadística en el siglo XXI, simplificando los cálculos necesarios para realizar las estimaciones estadísticas mediante el uso de métodos de Monte Carlo. La unidad once versa sobre el análisis de variables continuas, con énfasis en ANOVA. La doce busca dotar al alumno de herramienta para analizar variables categóricas y se centra en la tarea de perfilar electores y predecir resultados. La anteúltima unidad se encarga de aplicar lo aprendido de modelos categóricos para corregir el sesgo en las encuestas. Finalmente, la última unidad recupera el valor de los grupos focales como método complementario y presenta las nuevas técnicas de análisis y los desafíos se que presentan a la opinión pública tradicional.

Cabe destacar la importancia de la materia en la formación del alumno: el siglo XXI trae consigo un volumen inaudito de datos tanto estructurados como no estructurados que han de cambiar la forma tradicional en que se desempeña el analista de opinión pública. Es necesario que el alumno cuente con las herramientas computacionales y estadísticas para poder hacer frente a las nuevas demandas de la disciplina. La presente asignatura pretende por tanto dotar al estudiante de sus primeras armas sin descuidar la conexión entre teoría y praxis. La introducción a problemas de la ciencia de datos desde la opinión pública le permitirá continuar su estudio en problemáticas de Machine Learning, Big Data y/o fundamentos matemáticos de la estadística en futuros cursos.

Objetivos Generales

La presente materia tiene por objetivos generales:

- Permitir al alumno articular sus conocimientos teóricos con la investigación empírica.
- Dotar al alumno de herramientas metodológicas para el análisis de la opinión pública.
- Problematizar las potencialidades y limitaciones de la estadística para el análisis de la realidad social.

Objetivos Específicos

La presente materia tiene por objetivos específicos:

- Permitir al alumno formular una pregunta de investigación en el campo de la opinión pública y a su vez definir los métodos de análisis y sus estrategias de recolección de datos.
- Formar al alumno en el uso de técnicas estadísticas y computacionales aplicadas al análisis de la opinión pública desde un punto de vista de la ciencia de datos.
- Otorgar al alumno conocimiento técnico que le permita iniciarse en el ámbito de la consultoría y/o investigación de la opinión pública.

Metodología de Enseñanza

Clases Teóricas y Prácticas. En las clases teóricas se introducirá al alumno los métodos y las técnicas de análisis. En las clases prácticas se analizaran datos reales de opinión pública en Python.

Al ser una materia con clases prácticas, la ejercitación es central para la adquisición del conocimiento por lo cual se espera que el alumno destine cinco horas semanales para el estudio y resolución de los ejercicios.

Evaluación

La modalidad de evaluación constará de tres instancias. Las primeras dos instancias tomarán la forma de exámenes domiciliarios. La tercera instancia consistirá en una trabajo práctico integrador en forma de una exposición oral del alumno. Para promocionar la materia el alumno deberá obtener un promedio igual o mayor a siete (7) y no haber desaprobado ninguna de las instancias.

Programa por Unidades Temáticas

Introducción a Python

Introducción a la programación. Componentes de una Computadora. El Sistema Operativo. Directorios, carpetas y archivos. Instalación de Python. Anaconda y entornos virtuales. Jupyter Notebook.

Aplicación: Redacción de informes de opinión pública con Jupyter-Notebooks.

Repaso

Matriz de datos. Descripción univariada y bivariada. Medidas de tendencia central. Medidas de dispersión. Medidas resúmenes. Tablas de Contingencia. Primeros pasos con Python

Aplicación: Lectura de informes de Opinión Pública.

Manipulación de Datos con Pandas

Librería Pandas. Introducción a los objetos de Pandas. Tipo de Datos. Creación de Tablas. Lectura de datos: CSV y JSON. Indexación y selección. Operaciones con datos-columnas en Pandas. Datos Faltantes. Combinando tablas de datos: concat y append. Orden, filtrado, agregación, agrupamiento, filtrado en grupos (having) y lambdas. Funciones Ventana. Tablas Pivote.

Aplicación: Análisis del Censo 2010 - INDEC y los resultados electorales a Jefe de Gobierno en CABA de la Dirección Nacional Electoral (DINE).

Visualización Estadística de Datos

Matplotlib-Seaborn. Gráficos de barras. Gráfico de barras apiladas. Histogramas. Diagramas de Caja. Diagramas de dispersión.

Aplicación: Confección de un reporte sobre el resultado electoral de CABA 2017 (datos DINE) para diseñar una estrategia de campaña.

Introducción a la Lógica de Bases de Datos

Lógica de Base de Datos: las tablas y sus relaciones. Claves primarias y foráneas. Tipos de relaciones: uno-a-uno, varios-a-uno, uno-a-varios, varios-a-varios. Union de tablas: inner, outer, right, left join. Aplicación de la lógica de bases de datos a Pandas.

Aplicación: Estructura del Censo 2010. Construcción de un Redatam Casero a partir de los micro-datos. Extracción de parámetros poblacionales para el diseño de una encuesta/muestreo.

Introducción a la Programación en Python

Una segunda mirada a la sintaxis básica. Tipos de variables. Listas, tuplas y diccionarios. Strings- Operaciones. Estructuras de decisión. Loops. Funciones. Clases y Objetos.

Aplicación: Codificación y detección de errores en una encuesta.

Introducción a la GeoComputación en GeoPandas

Objetos Espaciales. Shapes y Geojsons. Proyecciones. Relaciones Espaciales. Base de datos Espaciales. Visualización y generación de Mapas.

Aplicación: Mapeo de los radios censales y los circuitos electorales.

Introducción a la Probabilidad y la Estadística

Espacio muestral y definición de probabilidad. Axiomas. Probabilidad conjunta y distribución marginal. Probabilidad condicional e Independencia. Teorema de Bayes. Variables aleatoria. Esperanza y varianza. Dis-

tribución de probabilidad discreta. Función de densidad y función de distribución acumulada. Distribución Empírica. Distribución Normal, Poisson, Bernoulli y Binomial, Categorial y Multinomial, Gamma. Propiedades. Parámetros y su estimación. Convergencia. Ley de los grandes números y Teorema central del límite. Aplicaciones en Python.

Aplicación: Análisis de los resultados de una encuesta.

Muestreo y Encuestas

Censos y muestras. Extrapolación y estimadores. Errores sistemáticos y aleatorios. Tamaño de muestra, variabilidad y convergencia. Generación de números aleatorios. Esquemas de muestreo: muestreo aleatorio simple, sistemático, estratificado, por conglomerados, con probabilidad proporcional al tamaño, polietápico. Muestreo no probabilísticos: muestreo por cuotas. Toma de muestras en Python.

Aplicación: Muestreo de radios censales de CABA para realizar una encuesta coincidental.

Aplicaciones de la Estadística Computacional

Introducción a Bootstrap y a los test de permutaciones. Aplicaciones para la estimación de intervalos de confianza. Aplicaciones para los Test de hipótesis. Error de tipo I y tipo II, nivel de confianza y significación estadística. Aplicaciones en Python.

Aplicación: Cálculo del intervalo de confianza de los resultados de una encuesta.

Análisis de Variables Continuas

Regresión lineal. Regresión lineal múltiple. Modelo lineal. Interpretación. Supuestos. Inferencia. Validación y bondad de ajuste. Contrastes. Anova de un factor. Anova de dos factores. Predicción. Selección de modelos. Aplicaciones en Python.

Aplicación: Análisis de un experimento por encuestas para determinar la estrategia política de un candidato

Análisis de Variables Categóricas

Modelos log-lineales. Regresión Logística. Interpretación. Supuestos. Inferencia. Validación y bondad de ajuste. Predicción. Selección de modelos. Lógica del Modelo Lineal Generalizado (GLM). Aplicaciones en Python.

Aplicación: Perfilado del Votante.

Corrección de Sesgos

No respuesta y sesgos. Imputación. Pesos y ponderadores, calibración y post-estratificación. Aplicaciones en Python.

Aplicación: Como calibrar mediante GLM una encuesta.

Viejos y Nuevos Métodos para el Análisis de la Opinión Pública

Límites de la Investigación Cuantitativa. Grupos focales: Teoría y práctica. Análisis de redes sociales. Comunidades y difusión. Muestreo por bola de nieve. Introducción al procesamiento del lenguaje natural. Análisis del Sentimiento. Big Data, ciencia de datos y el futuro de la opinión pública.

Cronograma de Clases

Unidad	**Clase**	**Práctico**
Introducción a Python	1	Redacción de Informes
Repaso	2	Lectura de Informes
Manipulación de Datos con Pandas	3, 4, 5, 6	Análisis de Datos Secundarios
Visualización de Datos	7	Estrategia de Campaña
Lógica de Bases de Datos	8, 9	Extracción de Parámetros Poblacionales
Programación en Python	10, 11, 12, 13	Codificación y Detección de Errores en Encuestas
GeoComputación en GeoPandas	14	Mapeo de Radios y Circuitos
Probabilidad y Estadística	15, 16, 17, 18, 19	Análisis de los Resultados de una Encuesta
Muestreo y Encuestas	20	Muestreo de Radios Censales para Encuestas
Estadística Computacional	21, 22	Cálculo de la Incertidumbre de una Estimación
Análisis de Variables Continuas	23, 24, 25	Estrategia Política y Experimento por Encuestas
Análisis de Variables Categóricas	26	Perfilado del Votante
Corrección de Sesgos	27	Calibración de una Encuesta
Viejos y Nuevos Métodos	28 y 29	Ejemplos de Aplicación

Bibliografía

La bibliografía de la materia será autocontenida en los apuntes de clase [11]. Los mismos se basan en [10], [3] y [9] para los aspectos computacionales de la materia. Para la parte estadística, el curso seguirá por un lado a [1] y [5], y por el otro a [4] y a [7]. Los dos primeros poseen un nivel de abstracción adecuado para un primer curso de metodología cuantitativa y los segundos se destacan por su enfoque computacional.

Referencias

[1] A. Agresti. *Métodos Estadísticos para Ciencias Sociales*. Pearson, 2018.

[2] A. Agresti and C.A. Franklin. *Statistics: The Art and Science of Learning from Data*. Pearson Education, 2012.

[3] D.Y. Chen. *Pandas for Everyone: Python Data Analysis*. Addison-Wesley data and analytics series. Addison-Wesley, 2018.

[4] D.M. Diez, C.D. Barr, and M. Çetinkaya-Rundel. *Introductory Statistics with Randomization and Simulation*. CreateSpace Independent Publishing Platform, 2014.

[5] D.M. Diez, C.D. Barr, and M. Çetinkaya-Rundel. *OpenIntro Statistics*. OpenIntro, Incorporated, 2015.

[6] M.G. Ferrando. *Socioestadística: Introducción a la Estadística en Sociología*. Ciencias sociales. Alianza Editorial, 2004.

[7] P.I. Good. *Resampling Methods: A Practical Guide to Data Analysis*. Birkhäuser Boston, 2013.

[8] S.L. Lohr and O.A.P. Velasco. *Muestreo: diseño y análisis*. Thomson: Matemáticas. Thomson, 2000.

[9] E. Matthes. *Python Crash Course: A Hands-On, Project-Based Introduction to Programming*. No Starch Press, 2015.

[10] W. McKinney. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, 2017.

[11] T. Olego. *Metodología computacional y estadística de la opinión publica en Python*. https://gitlab.com/olegotomas/met4op/apuntes. Creative Commons 3.0, 2019.

[12] M. Pilgrim. *Inmersión en Python 3*. http://diveintopython3.org/. Creative Commons 3.0, 2009.