



Máster en Metodología de la Investigación en Ciencias Sociales: Innovaciones y Aplicaciones

Departamento de Sociología IV
Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid



PROGRAMA DE LA ASIGNATURA

“ANÁLISIS DE FUENTES Y DE DATOS SECUNDARIOS Y TÉCNICAS DE *DATA MINING*”

Curso 2016-2017

(Primer cuatrimestre)

Profesor: Juan José Torres Serrano



Máster en Metodología de la Investigación en Ciencias Sociales: Innovaciones y Aplicaciones

Departamento de Sociología IV
Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid



OBJETIVOS FORMATIVOS

La asignatura tiene una doble finalidad, por un lado identificar las principales fuentes de datos secundarios en Sociología y aprender a utilizarlas para incorporarlas a sus investigaciones, y por otro lado pretende mostrar las principales herramientas de minería de datos para el análisis de datos secundarios y Big Data y enseñar a manejarlas para poder explotar bases de datos con el fin de encontrar información valiosa que sirva de ayuda a la investigación y a la toma de decisiones de las empresas e instituciones.

COMPETENCIAS A DESARROLLAR

En consonancia con estos objetivos, se pretende que en esta asignatura el alumno desarrolle las siguientes competencias o destrezas:

- CE9 - Capacidad de búsqueda y conocimiento de las principales fuentes de datos secundarios y valoración crítica de las limitaciones y las posibilidades de los datos secundarios.
- CT4 - Capacidad de elaborar y presentar informes finales de investigación a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- CT5 - Capacidad de auto-aprendizaje y de búsqueda autónoma de la información, que permita abordar el estudio de técnicas novedosas o áreas de aplicaciones nuevas, así como avanzar hacia la consecución de una tesis doctoral en los casos que se busque ese objetivo.
- CE4 - Manejo de programas informáticos aplicados a la investigación social, así como de tecnologías disponibles de la información y la comunicación. Capacidad de gestionar datos, elaborar modelos e interpretar los resultados de programas de ordenador.
- CE14 - Capacidad de evaluar la calidad de la investigación cuantitativa y cualitativa.

METODOLOGIA DOCENTE

Las clases serán de contenido teórico-práctico. A lo largo del curso, en cada bloque habrá una parte de carácter teórico basado en clases expositivas que se desarrollará paralelamente con un carácter más aplicado, mostrando las aplicaciones prácticas de lo aprendido en la parte teórica. En clase se contará con un ordenador permanentemente conectado a internet, que permita el acceso a bases de datos y a los programas para analizar estos.



Máster en Metodología de la Investigación en Ciencias Sociales: Innovaciones y Aplicaciones

Departamento de Sociología IV
Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid

En el último bloque se hará especial énfasis en el análisis de los datos y la presentación de los resultados, si bien, la utilización de los programas de análisis de datos secundarios y Data Mining se hará siempre desde el punto de vista del usuario no estadístico; no se incidirá en la fundamentación estadística, sino en la aplicación de los diferentes modelos estadísticos a la investigación y a la interpretación de los datos.

TEMARIO

Bloque I: Introducción a los datos secundarios.

- El proceso de investigación social y el lugar de los datos primarios y secundarios. Aplicaciones de los datos secundarios.
- Tipos de datos secundarios y su utilización como ayuda a nuestras investigaciones.
- Consideraciones metodológicas en el análisis de los datos secundarios. Ventajas e inconveniente de los datos secundarios.

Bloque II: Localización y uso de datos secundarios.

- Análisis primario de datos secundarios. Los microdatos como fuente para nuestras investigaciones. Como recopilar y utilizar los datos primarios producidos en otras investigaciones. Repositorios de microdatos.
- Bases de datos en España y en el extranjero.
- Datos sindicados: Paneles de distribuidores, consumidores y audiencias.
- Otras fuentes de datos privadas.
- Principales organismos nacionales proveedores de datos: INE, CIS, EGM, Organismos oficiales...
- Principales fuentes: Censo, Padrones, encuestas sociales, laborales, políticas, etc.
- Fuentes de datos estadísticas europeas e internacionales (Eurostat, ONU, CIA, etc.). Los problemas de armonización.
- Otras fuentes de datos: datos cualitativos, legislación, repositorios de datos de cuestionarios y preguntas, etc.



Máster en Metodología de la Investigación en Ciencias Sociales: Innovaciones y Aplicaciones

Departamento de Sociología IV
Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid

El objetivo de este bloque es conocer la extensa panoplia de fuentes de datos secundarios y saber seleccionar y utilizar aquellos que sean idóneos para apoyar nuestras investigaciones.

Bloque III: Minería de datos, análisis y utilización de los datos secundarios.

- Minería de datos: conceptos, técnicas y programas informáticos.
- El análisis estadístico de los datos secundarios con herramientas informáticas: IBM Spss Modeler, Gandía Barbwin, Excel, IBM Spss Statistics y otras herramientas.
- los procesos de minería de datos: Obtención de datos, procesado (selección, limpieza, exploración, modificación), modelización e interpretación.
- La modelización con minería de datos, aplicaciones para la toma de decisiones:
 - Modelos de clasificación: Árboles de decisión, redes neuronales
 - Modelos estadísticos: Regresión, Análisis factorial, Discriminante.
 - Modelos de segmentación: Cluster, Kohonen.
- Los distintos usos del análisis secundario y la minería de datos (indicadores sociales, investigación de mercados y márketing, informes sociales).
- La presentación de los resultados, elaboración tablás, gráficos e informes.

El objetivo de este bloque es adquirir un conocimiento sobre las diferentes herramientas de análisis de datos secundarios y minería de datos que nos permita elegir y aplicar aquella o aquellas que sean más adecuadas para obtener la mejor información para nuestra investigación. También aprenderemos la forma adecuada de presentar los datos.



BIBLIOGRAFÍA BÁSICA

Bloque I: Introducción a los datos secundarios.

- Blumer, M. (1980): "Why Don't Sociologists make more use of official Statistics?", *Sociology* 14, pp. 505-523.
- Desrosières, A. (2004): "Clasificar y codificar", en *La Política de los Grandes Números*. Barcelona: Editorial Melusina, pp. 259-303.

Bloque II: Localización y uso de datos secundarios.

- Cea, M^a.A. (1996) "El uso de fuentes documentales y estadísticas", en Cea D'Ancona, *Metodología cuantitativa: estrategias y técnicas de investigación social*. Madrid: Síntesis. pp. 220-237.
- Corti, L.; Witzel, A. y Bishop, L. (2005): "On the Potentials and Problems of Secondary Analysis. An Introduction to the FQS Special Issue on Secondary Analysis of Qualitative Data", en *Forum: Qualitative Social Research (FQS)*. Volume 6, No. 1, Art. 49 – Enero 2005.
- González Rodríguez, B. (2000) "Nuevas perspectivas en la explotación y aprovechamiento de los datos secundarios", en M. García Ferrando, J. Ibáñez y F. Alvira (comp.) *El análisis de la realidad social*. Madrid: Alianza. pp. 299-341. 3^a edición (revisada).

Bloque III: Minería de datos, análisis y utilización de los datos secundarios.

- Pérez López, C. (2014) *Técnicas de Minería de datos e inteligencia de negocios: SPSS Modeler*. Madrid: Garceta editorial.
- Pérez López, C. y Santín González, C. (2007): *Minería de datos: técnicas y herramientas*. Madrid: Thomson Editores.
- Manual de usuario IBM SPSS Modeler
<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/UsersGuide.pdf>
- Manual Gandia Barbwin
<http://download.tesigandia.com/manuales/gbw7.pdf>



Máster en Metodología de la Investigación en Ciencias Sociales: Innovaciones y Aplicaciones

Departamento de Sociología IV
Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid

BIBLIOGRAFÍA COMPLEMENTARIA

Aunque se proporcionará bibliografía específica en función del tema escogido para el trabajo de investigación final, se recomienda la siguiente bibliografía de carácter genérico:

- Cordero Valdavida, M. (1998) *Bancos de datos*. Madrid: CIS. (Cuadernos Metodológicos núm. 24).
- Hakim, C. (1982): *Secondary Analysis in Social Research*. Londres: Allen and Unwin.
- Hakim, C. (1983): "Research based on administrative records", en *Sociological Review* 31, (3), pp. 489-519.
- Heaton, J. (2004): *Reworking Qualitative Data*. London: Sage.
- Kiecolt, K y Nathan, L. (1985): *Secondary Analysis of Survey Data*. Longres: Sage. Quantitative Applications in the Social Sciences Series núm. 53.
- Selltiz, C. y otros (1980): *Métodos de Investigación en las Relaciones Sociales*. Madrid: Ediciones Rialp (pp. 507-541)
- Steward, D. W. (1984): *Secondary Research: Information, Sources and Methods*. Beverly Hills: Sage.

METODO DE EVALUACION

Se adoptará un modelo de evaluación continua de las actividades que desarrolle el estudiante dentro y fuera de la clase. La asistencia a clase es obligatoria.

En todo caso, para ser evaluado y superar la asignatura (que tiene un carácter presencial), se requiere la asistencia al 80% de las sesiones, como mínimo. La calificación final se compondrá y ponderará del modo siguiente:

Mediante examen (25%), un trabajo de curso en equipo (50%), practicas (15%) y participación en clase (10%).

INFORMACION DE CONTACTO

Juan José Torres

Departamento de Sociología IV Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid Campus de Somosaguas, s/n 28223 -
Pozuelo de Alarcón Madrid (España)

Tlf: (+34) 91 394-2973

Correo: juanjo@cps.ucm.es

<http://www.ucm.es/info/mmccss>

Horario de tutorías: Miércoles de 15:30 a 18:30 hs. Despacho 2401A