**Curso de Introducción al software estadístico R para Ciencias Sociales.**

**Docentes a cargo:** Tiscornia, Pablo y Weksler, Guido Ezequiel.

**Presentación.**

En los últimos años se han difundido muchas herramientas estadísticas novedosas para el análisis de información socioeconómica. En particular el software denominado "R", por tratarse de un software libre, se extiende cada vez más en diferentes disciplinas y recibe el aporte de investigadores e investigadoras en todo el mundo, multiplicando sistemáticamente sus capacidades.

Este programa se destaca, entre varias cosas, por su capacidad de trabajar con grandes volúmenes de información, utilizar múltiples bases de datos en simultáneo, generar reportes, realizar gráficos a nivel de publicación y por su comunidad de usuarios y usuarias que publican sus sintaxis y comparten sus problemas, hecho que potencia la capacidad de consulta y de crecimiento. A su vez, la expresividad del lenguaje permite diseñar funciones específicas que permiten optimizar de forma personalizada el trabajo cotidiano con R.

**Objetivo.**

El presente Taller tiene como objetivo principal introducirse en el aprendizaje del lenguaje de programación “R” aplicado procesamiento de la Encuesta Permanente de Hogares (EPH) - INDEC. Se apunta a brindar a los y las participantes herramientas prácticas para el procesamiento de datos haciendo énfasis en la producción y el análisis de estadísticas socioeconómicas en pos de abrir puertas para realizar investigaciones propias sobre diversas temáticas relacionadas al mercado de trabajo y las condiciones de vida de la población.

La Encuesta Permanente de Hogares será la base de datos de aplicación elegida para el curso, dado que representa un insumo fundamental para realizar estudios sobre el mercado de trabajo y las condiciones de vida de la población. Se hará una introducción a los lineamientos conceptuales principales de la encuesta, en pos de que los y las participantes puedan abordar con datos distintas problemáticas vinculadas al mercado de trabajo y las condiciones de vida de la población.

**Destinatarios.**

* Alumnos/as de carreras vinculadas a las ciencias sociales y al procesamiento de datos.
* Investigadores/as.
* Docentes.
* Trabajadores/as del sector público y privado con tareas vinculadas al procesamiento de datos.
* Público en general interesado.

**Requisitos.**

No se requiere conocimientos previos.

**Cursada.**

Modalidad: presencial

Fecha de inicio: a definir

Fecha de finalización: a definir

Cursada: [día] de [hora] a [hora]

Cantidad de clases: 9

Carga horaria: 27hs .

**Contenidos.**

Las clases del curso serán teórico-prácticas. Cada una estará acompañada por una nota de clase en la cual se plasman conjuntamente las líneas de código que se desarrollarán durante la cursada y su correspondiente explicación respecto a los procedimientos realizados. A su vez, los participantes dispondrán de un *script* con el código de cada clase para ir corriendo secuencialmente los ejercicios en paralelo al docente. Finalmente, se asignará un espacio de 30 minutos, para realizar ejercicios prácticos en clase, en los cuales se dará a los y las participantes una consigna relacionada con lo abordado en el día y a resolver en grupos.

**Clase 1 - Conceptos Principales de EPH:**

* Presentación del curso.
* Presentación de la Encuesta Permanente de Hogares: Lineamientos conceptuales y metodología
* Abordaje del marco teórico y analítico de la EPH y sus aplicaciones prácticas.
* Síntesis del operativo de campo, cobertura y periodicidad de la Encuesta
* Definiciones de las principales variables de interés a abordar en el curso: Condición de actividad, categoría ocupacional, precariedad y pobreza
* Metodología usuaria de las Bases de microdatos. Utilización del Diseño de Registro.

**Clase 2 – Introduciendo a R:**

* Descripción del programa “R”. Lógica sintáctica del lenguaje y comandos básicos
* Presentación de la plataforma RStudio para trabajar en “R”
* Caracteres especiales en “R”
* Operadores lógicos y aritméticos
* Definición de Objetos: Valores, Vectores y DataFrames
* Tipos de variable (numérica, de caracteres, lógicas)
* Lectura y Escritura de Archivos

**Clase 3 – Trabajando con bases de datos y Mercado de Trabajo:**

* Limpieza de Base de datos: Renombrar y recodificar variables, tratamiento de valores faltantes *(missing values/ NA´s)*
* Seleccionar variables, ordenar y agrupar la base de datos para realizar cálculos
* Creación de nuevas variables
* Aplicar filtros sobre la base de datos
* Construir medidas de resumen de la información
* Tratamiento de variables numéricas (edad, ingresos, horas de trabajo, cantidad de hijos / componentes del hogar, entre otras).
* Ejercicios prácticos para aplicar lo expuesto: Replicar Informe Técnico de Mercado de Trabajo (EPH-INDEC)
* Cálculo de tasas básicas del mercado de trabajo (tasa de actividad, empleo, desempleo, entre otras)
* Cálculo de tasas para distintos subconjuntos poblacionales (por aglomerado, sexo, grupos de edad)

**Clase 4 - Variables de ingresos y gráficos:**

* Procesamiento de indicadores agregados sobre las variables de ingresos (Ingresos laborales, no laborales, de ocupación principal, total familiar)
* Gráficos básicos de R (función “plot”): Comandos para la visualización ágil de la información
* Gráficos elaborados en R (función “ggplot”):
  + Gráficos de línea, barras, Boxplots y distribuciones de densidad
  + Parámetros de los gráficos: Leyendas, ejes, títulos, notas, colores
  + Gráficos con múltiples cruces de variables.

**Clase 5: Cálculo de tasas de Pobreza e Indigencia.**

* Definición de la Canasta Básica Alimentaria y Canasta Básica Total
* Metodología del cálculo de pobreza por línea (para personas y hogares)
* Ejercicio de estimación de tasas de Pobreza e Indigencia
* Ejercicio de estimación de tasas de Pobreza e Indigencia para subgrupos poblacionales (Género, Edad, Regiones)

**Clase 6: Pool de Datos en Panel. Gráficos y Matrices de Transición.**

* Metodología de trabajos en panel con EPH: (Esquema de rotación de la EPH, Variables para la identificación de los individuos en distintos períodos, Consistencias)
* Proceso para la construcción de paneles en R.
* Cálculo de Frecuencias de transiciones de estados (Categorías Ocupacionales, Situaciones de Pobreza/Indigencia)
* Gráficos de Transición de estados

**Clase 7: Documentación en R. Generación de reportes/informes.**

* Manejo de las extensiones del software “Rmarkdown” y “RNotebook” para elaborar documentos de trabajo, presentaciones interactivas e informes:
  + Opciones para mostrar u ocultar código en los reportes
  + Definición de tamaño, títulos y formato con el cual se despliegan los gráficos y tablas en el informe
  + Caracteres especiales para incluir múltiples recursos en el texto del informe: Links a páginas web, notas al pie, enumeraciones, cambios en el formato de letra (tamaño, negrita, cursiva)
  + Código embebido en el texto para automatización de reportes

**Clase 8: Introducción a R Intermedio:**

* Acercamiento a técnicas más sofisticadas en R, útiles para automatizar el procesamiento periódico de la información:
  + Estructuras de código condicionales
* Loops
* Creación de funciones a medida del usuario
* Herramientas para continuar el aprendizaje en R: Foros reconocidos de usuarios de R. Comunidades donde se comparten conocimientos, experiencias, consultas. Comandos para acceder a los documentos de ayuda.

**Clase 9: Trabajos prácticos: R en concreto / examen**

* Se presentará a los/as alumnos/as problemas concretos vinculados a la EPH y en relación a las experiencias que se fueron volcando a lo largo de la cursada. Se deberán abordar mediante el uso de R, aplicando lo aprendido en los módulos anteriores. En conjunto se expondrán los desafíos que emergen en el momento y la evaluación de las herramientas adecuadas para su abordaje y resolución. Entre los temas giran aquellos vinculados a la distribución personal del ingreso/Construcción de percentiles de ingreso; el mercado laboral a través de la herramienta Panel de datos; la documentación en R (Estilo de Notas de Clase).
* Espacio para consultas puntuales sobre los temas vistos durante el curso y presentación del trabajo final a entregar.

**Evaluación.**

Se debe cumplir un 75% de asistencia mínima y la aprobación de un trabajo práctico domiciliario. El mismo consta en realizar un breve informe sobre un problema concreto, debiendo los alumnos entregar un Script que incluya el código para el procesamiento de los datos y gráficos que integren el mismo. A aquellos/as que cumplan el requisito de la asistencia pero no la entrega/aprobación del trabajo práctico, les será entregado un certificado de asistencia al curso.