

Introducción al procesamiento de datos de la Encuesta Permanente de Hogares (EPH) - INDEC

Docente: Pablo Tiscornia

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.

Objetivos del curso

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 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.

Clases y contenido

- 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 y 3: Procesamiento de datos con el paquete **Tidyverse** y **eph**:
 - Descripción del programa “R”. Lógica sintáctica del lenguaje y comandos básicos
 - 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).
 - 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 y 5: Visualización de la información con **ggplot2**
 - 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
 - * Extensiones de ggplot
- Clase 6: Técnica de panel con la EPH
 - Alcances y límites en el uso de la técnica de panel con la EPH.
 - Construcción y análisis de la matriz del flujo entre categorías de la condición de actividad entre dos trimestres.

Librerías a instalar

```
install.packages(c("tidyverse","openxlsx",'ggplot2','ggthemes', 'eph','viridis'))
```