# **Entrenamiento en el procesamiento y análisis de encuestas de hogares usando R**

Módulo 0: introducción a R/Rstudio y a la inferencia basada en el diseño de muestreo

* R y Rstudio
* Creación de un proyecto y sus componentes
* Carga de bases de datos
* Consultas básicas con tidyverse
* Principio de representatividad
* Distribución de los estimadores de muestreo
* Estimación puntual y varianza estimada

Módulo 1: análisis con variables continuas

* Distribución del ingreso
* Estimación en subgrupos poblacionales
* Totales, medias e intervalos de confianza
* Histogramas y diagramas de caja
* Desviación estándar, percentiles y razones
* Medidas de desigualdad: Gini y curva de Lorenz
* Pruebas de hipótesis en muestras complejas: diferencia de medias y diferencias en subgrupos
* Contrastes de hipótesis
* Procesamiento en bloque (varios países)

Módulo 2: análisis con variables discretas

* Estimación de tamaños y proporciones
* Estimación en subgrupos poblacionales
* Tablas de contingencia
* Pruebas de independencia
* Razón de Odds
* Contrastes de hipótesis
* Procesamiento en bloque (varios países)

Módulo 3: análisis gráfico con ggplot2 para encuestas complejas

* Definición del *layout* gráfico
* Histogramas, densidades, y diagramas de caja
* Diagramas de dispersión
* Diagramas de barras
* Mapas con tmap y ggplot2

Módulo 4: modelos de regresión

* Modelo nulo y modelos saturados
* Modelos poblacionales, modelos muestrales y modelos ponderados
* Medidas de bondad del ajuste
* ¿Ponderar o no ponderar?
* Diagnóstico del modelo: residuales estandarizados, pruebas de normalidad, homoscedasticidad, distancia de Cook
* Datos influyentes
* Predicción fuera de la muestra

Módulo 5: modelos loglineales y logísticos

* Tablas de contingencia
* Pruebas de independencia
* Modelos loglineales
* Modelos logísticos
* Estimación de los parámetros de regresión y estimación en subgrupos poblacionales

Módulo 6: otros modelos lineales generalizados

* Máxima pseudo-verosimilitud
* Regresión gamma
* Regresión multinomial

Módulo 7: modelos multinivel

* Pendientes paralelas y efectos aleatorios
* Modelo nulo
* Modelos con intercepto aleatorio
* Modelos con intercepto y pendiente aleatoria
* Modelo logit multinivel
* Predicciones

Módulo 8: datos faltantes e imputación

* Algunos casos de ausencia de respuesta
* Imputaciones por media condicional y no condicional
* Imputación hot-deck y cold-deck
* Imputación por regresión
* Imputación con el vecino más cercano
* Imputación por regresión
* Imputación con el vecino más cercano y modelos de regresión
* Imputación múltiple: Bootstrap
* Estimación puntual y estimación de varianzas con valores plausibles