

Alberto Hernández Armendáriz

## **Propuesta II curso de R para UPMRIP**

UN Migration

Febrero 14, 2022.

**Objetivo:** aprendizaje de herramientas en R para el análisis estadístico de la migración.

**Duración total:** 40 hrs, divididas en 20 sesiones de 2 horas.

### **Bases de datos:**

Base de datos de devolución de mexicanos desde Estados Unidos – INM UPMRIP

<http://portales.segob.gob.mx/es/PoliticaMigratoria/EvDevMexEUU>

Encuesta sobre migración en la frontera Sur de México (EMIF-Sur)

<https://www.colef.mx/emif/basescuestionarios.html>

Censo de Población y Vivienda 2020 – Hogares y personas, remesas, emigración, inmigración y retorno

<https://www.inegi.org.mx/programas/ccpv/2020/>

Estadísticas de Empleo y Migración INE Guatemala

<https://www.ine.gob.gt/ine/portal-estadistico-1-0/>

### **Semana 1**

#### **Sesión 1**

- Aplicación de pequeño cuestionario para evaluar estado de conocimiento de los participantes con respecto de R
- Presentación de la agenda del curso
- Presentación de bases de datos que se utilizarán durante el curso: Devolución de mexicanos de Estados Unidos UPMRIP, EMIF-Sur y CPyV2020 de INEGI.
- Atender dudas de instalación de R y RStudio. Asegurarse que todos los participantes lo tengan instalado

#### **Sesión 2**

- Diferencias entre registros administrativos, estadística migratoria, censos y encuestas
- Cómo leer bases de datos de distintos formatos (csv, Excel, spss -UPM, dbf) -
- Cómo guardar tablas procesadas en R en archivos csv y Excel
- Buenas prácticas en organización de archivos de tablas de datos (diccionarios)
- Leer y unir columnas pertinentes de tablas de Devolución de Mexicanos (2017,2018 y 2019)
  - Nombrar columnas
  - Observar diferencias

- Creación de índices
- Cómo seleccionar columnas y renglones específicos
- Clases de datos en R (data frame, matrix, vector, character, number)

## **Semana 2**

### **Sesión 1**

- Uso de tablas unidas en la sesión anterior (Mexicanos Devueltos por Estados Unidos, 2017-2019).
- Tipos de datos: categóricos nominales vs continuos
- Análisis de tendencias históricas de mexicanos devueltos por Estados Unidos
- Ggplot: cómo graficar las tendencias a través del tiempo y por distintos grupos
- Distribución de edades de personas repatriadas, ¿cambian de un año a otro? –boxplots

### **Sesión 2**

- Continúa el uso de las tablas unidas de mexicanos devueltos por Estados Unidos 2017, 2018 y 2019
- Reestructuración de tablas: librería tidyverse, gather, spread; cómo “reacomodar” bases de datos para poder analizarlas.
- Recodificar variables en R y su equivalente en SPSS

## **Semana 3**

### **Sesión 1**

- Uso de tablas unidas de mexicanos devueltos por Estados Unidos 2017, 2018, 2019
- Gráficas en R: uso de librería ggplot2
- Gráficas de barras, barras compuestas, scatterplots, boxplots, gráficas de distribuciones estadísticas, etiquetas, títulos, ejes, colores
- Gráficas compuestas

### **Sesión 2**

- Uso de tablas unidas de mexicanos devueltos por Estados Unidos 2017, 2018, 2019
- Análisis exploratorio de datos y estadística descriptiva:
- Media, mediana, varianza y desviación estándar, distribuciones
- Graficar boxplots en R y cómo se utilizan
- Selección de columnas relevantes: de un conjunto grande de variables, ¿cómo seleccionamos las más pertinentes para comenzar el análisis
- Medidas de correlación: tablas de cruces entre variables

## **Semana 4**

## Sesión 1

- Ciclos for y while: introducción a los conceptos y sintaxis
- ¿Cómo funcionan las funciones en R? ejemplos de funciones sencillas
- Funciones para leer muchos archivos de Excel o un archivo de Excel con distintas tabs

## Sesión 2

- Introducción a los mapas en ggplot: cómo agrupar mexicanos devueltos por entidad de origen y su visualización en un mapa *choropleth*

## Semana 5

### Sesión 1

- Análisis de encuestas y estadística inferencial: Uso de la EMIF Sur
- Uso de ponderadores – conceptos y código en R
- Introducción a las muestras: representatividad, potencia estadística, niveles de confianza,
- Énfasis en el cálculo de intervalos de confianza – R, interpretación
- Revisar distintos tipos de estrategias de muestreo

### Sesión 2

- Análisis de encuestas y estadística inferencial: Uso de la EMIF Sur
- Introducción a las pruebas de hipótesis: concepto y noción de significancia estadística
- Pruebas de hipótesis en R: ¿entre los migrantes devueltos, existen diferencias significativas en algunas características?, por ejemplo, ¿existen diferencias significativas en las características de migración (e.g. migración familiar vs individual) entre personas que hablan y no hablan una lengua indígena? ¿existen diferencias significativas en los salarios que perciben hombres y mujeres en el extranjero?

## Semana 6

### Sesión 1

- Uso del cuestionario ampliado del Censo de Población y Vivienda 2020
- Repaso de muestras: representatividad, potencia, confiabilidad de la muestra del cuestionario ampliado del CPV2020

### Sesión 2

- Uso del cuestionario ampliado del Censo de Población y Vivienda 2010
- Cómo pegar respuestas de distintos niveles en R: **hogares y personas**, creación de índices (llaves)

- Repaso de principales conceptos relacionados al estudio de la migración en el CPV2020: emigración, inmigración, retorno, remesas, movilidad laboral y educativa y cómo se construyen.

## Semana 7

### Sesión 1

- Estimación de municipios expulsores de población (origen de emigrantes) en la última década. Uso de Censos de Población y Vivienda 2010 y 2020.
- ¿cuáles municipios tuvieron crecimiento negativo o inferior a tasas nacionales urbanas y rurales? ¿podría afirmarse que son municipios expulsores de migrantes?
- Repaso de métodos: comparación de índices de los municipios mexicanos con clave oficial de INEGI

### Sesión 2

- Se utilizarán los datos preparados durante la sesión anterior
- Creación de mapa con ggplot y gráficas compuestas con otros parámetros pertinentes
- Repaso de buenas prácticas en cartografía: material de curso de visualizaciones de 2019
- Introducción a conceptos de análisis espacial: ¿se observan patrones en el espacio? ¿hay agrupación o dispersión de municipios expulsores de población? Prueba de I de Morán

## Semana 8

### Sesión 1

- Análisis de Remesas: Uso del Censo de Población y Vivienda 2020
- Estimación del valor de remesas recibidas en 2020
- Cómo se comparan los datos censales a otras fuentes oficiales de estimación de remesas (Banxico y ENIGH) (En este punto, los participantes no trabajarán estas bases de datos mencionadas, sino que serán datos preparados por el consultor para la enseñanza de la comparación de distintas fuentes de datos)
- Se pedirá que los participantes preparen un pequeño reporte de remesas con medidas de estadísticas descriptivas vistas hasta el momento utilizando el CPyV2020

### Sesión 2

- Análisis de Remesas: Uso del Censo de Población y Vivienda 2020
- ¿Qué estados reciben más y menos remesas?
- Creación de mapa *choropleth* de remesas por entidad federativa
- ¿Qué porcentaje representaron las remesas del total del PIB en 2020?
- ¿Qué características sociodemográficas tienen los hogares que recibieron remesas en 2020? (Información preparada previamente por el consultor para su exposición como ejemplo y cálculo en clase)

- Análisis estadístico de datos financieros: introducción al concepto de series de tiempo (sólo como punto teórico para que participantes se familiaricen con el concepto)

## **Semana 9**

### **Sesión 1**

- Construcción de indicadores
- Introducción a los indicadores, metodologías y buenas prácticas  
Bibliografía:  
The Good Indicators Guide: Understanding how to use and choose indicators. NHS UK.  
How to develop good indicators. Compass.  
Good Practice Guidelines for Indicator Development and Reporting. Third World Forum on Statistics, knowledge and policy. OECD 2009.
- Mesa redonda con los participantes: cuál de la información analizada en las semanas anteriores podría ser convertida en indicadores para su uso público o interno

### **Sesión 2**

- Creación de narrativas
- Mesa redonda con los participantes: cuál de la información analizada en las semanas anteriores podría ser de interés para su difusión pública y porqué. Qué herramientas utilizaríamos: visualizaciones, periodismo, medios, datos duros.
- Cómo elegir qué tanta información incluir en nuestra narrativa
- Observar momentos de coyuntura que puedan ayudar a la difusión pública de información de interés

## **Semana 10**

- Se proponen dos sesiones como espacio de repaso en dónde se pueda profundizar en los temas vistos que sean propuestos por los participantes