

# Práctica Tidyverse con EPH - Consignas

Nombre del Docente

2025-03-27

## Práctica con Tidyverse: EPH - INDEC

### Instrucciones

Este documento contiene 10 ejercicios para practicar las funciones principales de Tidyverse utilizando datos de la Encuesta Permanente de Hogares (EPH) del INDEC.

Deberás implementar soluciones utilizando las siguientes funciones de Tidyverse: - `filter()` - `select()` - `mutate()` - `arrange()` - `group_by()` - `summarize()` - `ungroup()`

### Preparación del entorno

Primero, carga las librerías necesarias:

### Carga de datos

Carga los datos de la EPH. Modifica las rutas de acuerdo a donde tengas guardados los archivos.

### Ejercicio 1: Selección y filtrado básico

**Consigna:** Selecciona las variables CODUSU, NRO\_HOGAR, COMPONENTE, CH04 (sexo), CH06 (edad), NIVEL\_ED (nivel educativo), CH15 (lugar de nacimiento), ESTADO (condición de actividad), P21 (ingreso de la ocupación principal) y PONDERA (factor de expansión). Luego filtra para considerar solo personas de 25 años o más.

### Ejercicio 2: Creación de variables categóricas

**Consigna:** A partir del conjunto de datos obtenido en el ejercicio anterior, crea las siguientes variables nuevas utilizando `mutate()` y `case_when()`:

- `sexo_desc`: Descripción del sexo (“Varón” o “Mujer”) a partir de CH04
- `nivel_educativo`: Descripción del nivel educativo (Primaria incompleta, Primaria completa, etc.) a partir de NIVEL\_ED
- `lugar_nacimiento`: Descripción del lugar de nacimiento a partir de CH15
- `condicion_actividad`: Descripción de la condición de actividad (Ocupado, Desocupado, Inactivo) a partir de ESTADO

### Ejercicio 3: Filtrado múltiple

**Consigna:** Filtra los datos para analizar solo a las personas ocupadas (`ESTADO == 1`) que tienen entre 30 y 60 años, y que hayan nacido en otra provincia o en el extranjero.

### Ejercicio 4: Ordenamiento de datos

**Consigna:** Ordena los datos de manera descendente por nivel educativo (`NIVEL_ED`) y luego de manera ascendente por edad (`CH06`). Muestra los primeros 15 registros del resultado.

### Ejercicio 5: Estadísticas por nivel educativo

**Consigna:** Agrupa los datos por nivel educativo y calcula:

- La cantidad de personas (ponderada)
- La edad promedio (ponderada)
- El ingreso promedio de la ocupación principal (ponderado)
- El porcentaje de ocupados Ordena los resultados por ingreso promedio de manera descendente.

### Ejercicio 6: Estadísticas por lugar de nacimiento

**Consigna:** Agrupa los datos por lugar de nacimiento y calcula:

- La cantidad de personas (ponderada)
- El porcentaje de personas con educación superior completa
- El ingreso promedio de la ocupación principal (ponderado)
- La edad promedio (ponderada)

Ordena los resultados por porcentaje de educación superior de manera descendente.

### Ejercicio 7: Medidas de dispersión por nivel educativo

**Consigna:** Para las personas ocupadas, calcula por nivel educativo:

- El desvío estándar del ingreso
- El coeficiente de variación ( $\text{desvío estándar} / \text{media} * 100$ )
- El ingreso mínimo y máximo
- El rango de ingresos ( $\text{máximo} - \text{mínimo}$ )

Ordena los resultados por coeficiente de variación.

### Ejercicio 8: Análisis por sexo y nivel educativo

**Consigna:** Agrupa los datos por sexo y nivel educativo, y calcula:

- La cantidad de personas (ponderada)
- El ingreso promedio (ponderado)
- La tasa de ocupación (porcentaje de ocupados)

Luego calcula la brecha de ingresos entre varones y mujeres para cada nivel educativo.

## Ejercicio 9: Cuartiles de ingreso por nivel educativo

**Consigna:** Para las personas ocupadas, calcula por nivel educativo:

- El primer cuartil de ingresos (P25)
- La mediana de ingresos (P50)
- El tercer cuartil de ingresos (P75)
- El rango intercuartílico ( $IQR = P75 - P25$ )

Ordena los resultados por mediana de ingresos.

## Ejercicio 10: Estadísticas por lugar de nacimiento y sexo

**Consigna:** Agrupa los datos por lugar de nacimiento y sexo, y calcula:

- La cantidad de personas (ponderada)
- El porcentaje de personas con nivel secundario completo o superior
- La edad promedio (ponderada)
- La tasa de actividad (porcentaje de personas económicamente activas)

Finalmente, elimina la agrupación y ordena los resultados por lugar de nacimiento y luego por tasa de actividad de manera descendente.

## Desafío final (Opcional)

**Consigna:** Combina las funciones de Tidyverse para crear un análisis que responda a la siguiente pregunta: ¿Cómo varían los ingresos y la educación según el lugar de nacimiento y qué grupos tienen mayor desigualdad interna?

Para ello deberás:

1. Filtrar solo personas ocupadas entre 25 y 65 años
2. Agrupar por lugar de nacimiento y calcular estadísticas de ingresos (promedio, mediana, desvío, CV)
3. Calcular la proporción de personas con diferentes niveles educativos por lugar de nacimiento
4. Ordenar los resultados según algún criterio relevante