

TP 3 Bioestadística

Irisarri - Landa

TP 3: Estandarización de Tasas

Alumnos: Malena Irisarri, Román Landa

UNIVERSIDAD NACIONAL
DE ROSARIO
FACULTAD DE CIENCIAS
ECONÓMICAS Y ESTADÍSTICA



UNR

Rosario, Argentina

29 de Mayo de 2025

Introducción

Estandarización de tasas

La estandarización de tasas es una herramienta fundamental en epidemiología y demografía que permite comparar indicadores de salud entre poblaciones con estructuras demográficas diferentes. Este método ajusta las tasas crudas (por ejemplo, de mortalidad o morbilidad) para eliminar el efecto de, por ejemplo, la edad o el sexo, lo que facilita la identificación de patrones y desigualdades reales entre grupos o regiones.

En este trabajo, se aplicará la estandarización de tasas para analizar las muertes por **infarto agudo de miocardio** durante el **año 2023** en las provincias de Argentina, utilizando como referencia la población estándar del país. A través de este proceso, se calcularán tasas estandarizadas por edad y las razones de mortalidad estándar (RME), lo que permitirá comparar el riesgo de muerte entre provincias de manera precisa.

Infarto agudo de miocardio

El infarto agudo de miocardio, comúnmente conocido como ataque al corazón, es una emergencia médica que ocurre cuando el flujo sanguíneo al corazón se bloquea, lo que priva al corazón de oxígeno y puede causar la muerte de las células cardíacas. Los síntomas incluyen rigidez o dolor en el pecho, el cuello, la espalda o los brazos, así como fatiga, mareos, ritmo cardíaco anormal y ansiedad.

Aunque cualquier persona puede sufrir un Infarto de Miocardio, no todas tienen el mismo riesgo. Las personas con problemas de corazón o que ya han sufrido un evento cardiovascular, las de edad avanzada y las que presentan factores de riesgo (hipertensión arterial, tabaquismo, obesidad, diabetes, elevación del colesterol malo (LDL), descenso del colesterol bueno (HDL)) tienen mayor riesgo. Los infartos no son hereditarios, pero personas con algún familiar de primer grado (padre, madre o hermano) que ya hayan padecido un infarto, o con una enfermedad hereditaria como hipercolesterolemia o diabetes tiene más posibilidades de padecerlo.

Dado que la edad es uno de los principales factores de riesgo, es fundamental tenerla en cuenta al comparar tasas de mortalidad por infarto entre distintas provincias. Por eso, en este análisis se realizará una estandarización por edad, lo cual permitirá una comparación más adecuada entre poblaciones con estructuras etarias diferentes.

Datos

Para poder estandarizar las tasas de mortalidad, se requiere contar con el número de defunciones por infarto agudo de miocardio desagregado por provincia y grupo de edad. Esta información correspondiente al año 2023 fue obtenida de la página de la Dirección de Estadísticas e Información en Salud (DEIS).

Además, se necesitaba la población dividida por grupos de edad para cada provincia, así como para el total del país (utilizado como población estándar). Estos datos fueron descargados del sitio del Instituto Nacional de Estadística y Censos (INDEC), en la sección de Proyecciones y Estimaciones de Población.

Metodo de Estandarización Directa

Para calcular las tasas de mortalidad estandarizadas utilizaremos el método directo de estandarización. Este consiste en calcular las tasas brutas por grupos de edad para cada provincia y multiplicarla por la población estándar, en este caso la de Argentina.

```

# Inicializar un vector para almacenar los nombres de los data frames generados
nombres_dfs_generados <- c()

# Bucle para procesar cada hoja del archivo de proyecciones
for (hoja in nombres_hojas) {
  # print(hoja) # Descomentar para depuración

  # 1. Leer los datos de la hoja actual
  datos_hoja <- read_excel(ruta_archivo,
                           sheet = hoja,
                           range = "F60:H84")

  # Eliminar filas donde TODAS las columnas son NA
  datos_hoja <- datos_hoja[!apply(is.na(datos_hoja), 1, all), ]

  # Verificar y unir 'edades' al data frame actual
  if (nrow(edades) != nrow(datos_hoja)) {
    warning(paste("El número de filas en 'edades' (", nrow(edades), ") no coincide con 'datos_hoja' (",
                  next # Saltar a la siguiente iteración del bucle
  }
  datos_hoja <- cbind(Edad = edades$Edad, datos_hoja) # Asegurar que la columna se llame 'Edad'

  # Extraer los dos primeros dígitos del nombre de la hoja para 'codigo_prov'
  codigo_provincia <- str_extract(hoja, "^\\d{2}")

  # Agregar la columna 'codigo_prov' al data.frame (como carácter)
  datos_hoja$codigo_prov <- codigo_provincia

  # Limpiar el nombre de la hoja para usarlo como nombre de variable en R
  nombre_limpio <- tolower(hoja)
  nombre_limpio <- str_replace_all(nombre_limpio, "^\\d{2}-", "") # Eliminar prefijo numérico
  nombre_limpio <- str_replace_all(nombre_limpio, "ñ", "n") # Reemplazar 'ñ'
  nombre_limpio <- str_replace_all(nombre_limpio, "[áéíóúü]", function(x) { # Quitar tildes
    chartr("áéíóúü", "aeiouu", x)
  })
  nombre_limpio <- str_replace_all(nombre_limpio, "[[:space:]]", "_") # Reemplazar espacios por guiones
  nombre_limpio <- str_replace_all(nombre_limpio, "[^a-z0-9_]", "") # Eliminar caracteres no alfanuméri

  # Eliminar la primera fila (asumiendo que es una fila de totales)
  datos_hoja <- datos_hoja[-1, ]

  # Convertir columnas numéricas a tipo numérico para cálculos
  cols_numericas <- c("Ambos sexos", "Varones", "Mujeres")
  for (col in cols_numericas) {
    if (col %in% colnames(datos_hoja)) {
      datos_hoja[[col]] <- as.numeric(datos_hoja[[col]])
    }
  }

  # Calcular la columna '%poblacion'
  datos_hoja$`%poblacion` <- (
    datos_hoja$`Ambos sexos` / sum(datos_hoja$`Ambos sexos`, na.rm = TRUE)
  ) * 100

```

```

# Lógica para colapsar grupos de edad "80 y más"
grupos_a_colapsar <- c("80-84", "85-89", "90-94", "95-99", "100 y más")
nombre_columna_grupo_edad <- "Edad"

filas_a_colapsar <- datos_hoja %>%
  filter(get(nombre_columna_grupo_edad) %in% grupos_a_colapsar)

filas_no_colapsadas <- datos_hoja %>%
  filter(!get(nombre_columna_grupo_edad) %in% grupos_a_colapsar)

if (nrow(filas_a_colapsar) > 0) {
  fila_colapsada <- filas_a_colapsar %>%
    summarise(
      !!nombre_columna_grupo_edad := "80 y más",
      `Ambos sexos` = sum(`Ambos sexos`, na.rm = TRUE),
      `Varones` = sum(`Varones`, na.rm = TRUE),
      `Mujeres` = sum(`Mujeres`, na.rm = TRUE),
      `%poblacion` = sum(`%poblacion`, na.rm = TRUE),
      codigo_prov = first(codigo_prov) # Tomar el código de provincia
    )
  datos_hoja <- bind_rows(filas_no_colapsadas, fila_colapsada)
}

# Realizar el merge con 'defunciones'
datos_hoja <- datos_hoja %>%
  left_join(defunciones,
    by = c("codigo_prov" = "PROVRES", "Edad" = "GRUPEDAD"))

# Asignar el data frame procesado al entorno global con el nombre limpio
assign(nombre_limpio, datos_hoja, envir = .GlobalEnv)

# Mensaje de confirmación y almacenamiento del nombre
# cat("Creado data frame:", nombre_limpio, "\n")
nombres_dfs_generados <- c(nombres_dfs_generados, nombre_limpio)
}

##### Metodo Directo Estandarizacion #####

tasas_totales <- list()
tasa_estandar_directo <- list()
for (nombre_df_provincia in nombres_dfs_generados) {
  provincia_df <- get(nombre_df_provincia, envir = .GlobalEnv)

  # Calcula la 'tasa_por_edad'
  provincia_df <- provincia_df %>%
    mutate(
      tasa_por_edad = ifelse(
        !is.na(`Ambos sexos`) & `Ambos sexos` != 0,
        cuenta_total / `Ambos sexos`,
        0 # 0 NA
      )
    )
  provincia_df$defunciones_estandar <- provincia_df$tasa_por_edad*total_del_pais$`Ambos sexos`
}

```

```

# Calcula la 'tasa_total' ponderada por fila
tasa_total_ponderada <- sum(
  (provincia_df$tasa_por_edad * (provincia_df$`%poblacion` / 100)), # %poblacion / 100
  na.rm = TRUE # Importante para que la suma ignore los NA
)
tasa_estandar <- sum((provincia_df$defunciones_estandar), na.rm = TRUE) / sum(total_del_pais$`Ambos s

# Almacena la tasa total ponderada en la lista 'tasas_totales'
tasas_totales[[nombre_df_provincia]] <- tasa_total_ponderada
tasa_estandar_directo[[nombre_df_provincia]] <- tasa_estandar

# Asigna el data frame actualizado con 'tasa_por_edad' de nuevo al entorno global
assign(nombre_df_provincia, provincia_df, envir = .GlobalEnv)
}

```

Metodo de Estandarización Indirecto

Utilizamos el método indirecto para calcular la relación entre el número observado y esperado de casos, es decir las razones de mortalidad estandarizadas (RME).

```

# Recalcular defunciones_totales
defunciones_totales <- defunciones %>%
  group_by(GRUPEDAD) %>%
  summarise(
    cuenta_total = sum(cuenta_total, na.rm = TRUE)
  ) %>%
  ungroup()

defunciones_totales <- defunciones_totales %>%
  left_join(
    total_del_pais %>% select(Edad, `Ambos sexos`),
    by = c("GRUPEDAD" = "Edad")
  ) %>%
  mutate(
    tasa = ifelse(
      !is.na(`Ambos sexos`) & `Ambos sexos` != 0,
      cuenta_total / `Ambos sexos`,
      0 # Si 'Ambos sexos' es NA o 0, la tasa es 0.
    )
  )

rme_indirecta <- list()

for (nombre_df_provincia in nombres_dfs_generados) {
  provincia_df <- get(nombre_df_provincia, envir = .GlobalEnv)

  # --- PASO 1: Joinear la columna 'tasa_nacional'
  provincia_df <- provincia_df %>%
    left_join(
      # Seleccionamos GRUPEDAD, y renombramos 'tasa' a 'tasa_nacional'
      defunciones_totales %>% select(GRUPEDAD, tasa_nacional = tasa),

```

```

    by = c("Edad" = "GRUPEDAD") # unimos por las columnas de grupo de edad
  ) %>%
  mutate(tasa_nacional = replace_na(tasa_nacional, 0))
# --- PASO 2: Calcular 'defunciones_indirectas' usando la nueva columna 'tasa_nacional' ---
provincia_df <- provincia_df %>%
  mutate(
    # Ahora usamos 'tasa_nacional' para el cálculo
    defunciones_indirectas = `Ambos sexos` * tasa_nacional
  )
# --- PASO 3: Calculo de RME ---
total_esperado <- sum(provincia_df$defunciones_indirectas, na.rm = TRUE)
total_observado <- sum(provincia_df$cuenta_total, na.rm = TRUE)
if (total_esperado != 0) {
  rme_indirecta[[nombre_df_provincia]] <- total_observado / total_esperado
} else {
  rme_indirecta[[nombre_df_provincia]] <- NaN
  warning(paste("Total esperado es cero para", nombre_df_provincia, ". RME será NaN/Inf."))
}

assign(nombre_df_provincia, provincia_df, envir = .GlobalEnv)
}

# print("Razones de Mortalidad Estandarizadas Indirectas (RME):")
# print(rme_indirecta)

```

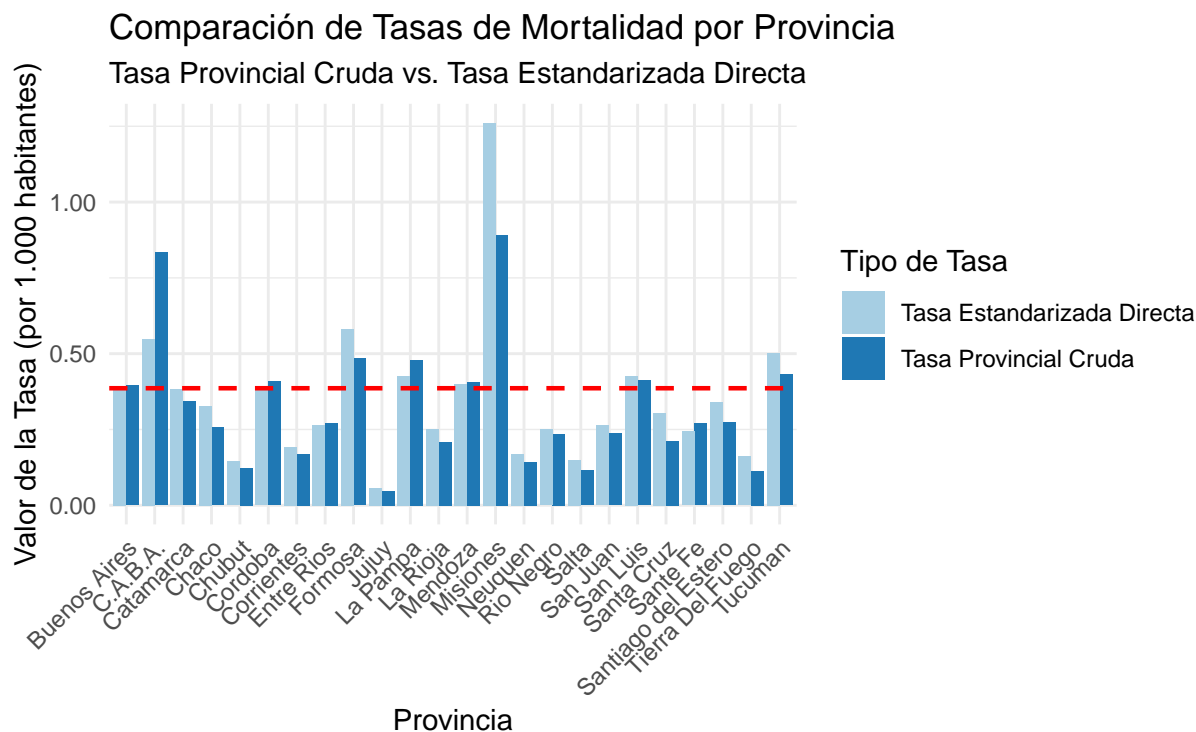
Resultados

Nos parece importante comparar las tasas crudas y las estandarizadas ya que nos indican si la edad de cada provincia está distribuída parecida a la población Argentina. Por ejemplo:

Table 1: Tasas de Mortalidad por Provincia (por 1.000 habitantes)

Provincia	Tasa Cruda	Tasa Estandarizada
Buenos Aires	0.40	0.38
C.A.B.A.	0.83	0.55
Catamarca	0.34	0.38
Chaco	0.26	0.33
Chubut	0.12	0.14
Cordoba	0.41	0.38
Corrientes	0.17	0.19
Entre Rios	0.27	0.26
Formosa	0.48	0.58
Jujuy	0.05	0.06
La Pampa	0.48	0.43
La Rioja	0.21	0.25
Mendoza	0.41	0.40
Misiones	0.89	1.26
Neuquen	0.14	0.17
Rio Negro	0.23	0.25
Salta	0.11	0.15
San Juan	0.24	0.26
San Luis	0.41	0.43
Santa Cruz	0.21	0.30
Sante Fe	0.27	0.24
Santiago del Estero	0.27	0.34
Tierra Del Fuego	0.11	0.16
Tucuman	0.43	0.50

- Si la **Tasa Cruda** $>$ **Tasa Estandarizada**, la provincia tiende a tener una población más envejecida que la población de referencia. Su tasa cruda parece alta, pero al ajustar por edad, su mortalidad es en realidad más baja de lo que la tasa cruda sugiere.
- Si la **Tasa Cruda** $<$ **Tasa Estandarizada**, la provincia tiende a tener una población más joven que la población de referencia. Su tasa cruda parece baja, pero al ajustar por edad, su mortalidad es en realidad más alta de lo que la tasa cruda sugiere.



La línea punteada roja representa la Tasa de Mortalidad Nacional de Argentina: 0.39 por 1.000 habitantes.

Particularmente llama la atención en el gráfico 1 es Misiones, ya que tiene una tasa estandarizada muy por arriba de su tasa cruda, ambas muy elevadas en comparación con otras provincias y la tasa de mortalidad de Argentina. Esto es una señal de alerta ya dicha provincia tiene una estructura de edad más joven que la población de referencia (o bien, presenta una alta proporción de población en edades de baja mortalidad), pero aún así presenta una mortalidad muy elevada. Esto sugiere que el riesgo de muerte allí es alto y no se debe al envejecimiento de la población.

Provincias con Tasas Estandarizadas por encima de la Nacional

Además de Misiones, CABA, Formosa, La Pampa, Mendoza, San Luis y Tucumán muestran sus tasas estandarizadas por encima de la línea nacional. Esto significa que, si estas provincias tuvieran la misma estructura de edad que el país en su conjunto, su mortalidad sería más alta que el promedio. Esto nos indica de que existen factores de riesgo de mortalidad inherentes a la población de estas provincias que no se explican por la edad, y que necesitan atención en salud pública.

Provincias con Tasas Estandarizadas por debajo de la Nacional

Chaco, Chubut, Corrientes, Entre Ríos, Jujuy, La Rioja, Neuquén, Rio Negro, Salta, San Juan, Santa Cruz, Santa Fe, Santiago del Estero y Tierra del Fuego tienen tasas estandarizadas por debajo de la línea nacional. Esto indica que, controlando el efecto de la estructura de edad de su población, el riesgo de morir en estas provincias es menor que el promedio del país. Es una señal de una situación de mortalidad relativamente más favorable.

En el mapa también vemos que la provincia con mayor tasa es Misiones, ya que aparece en color rojo.



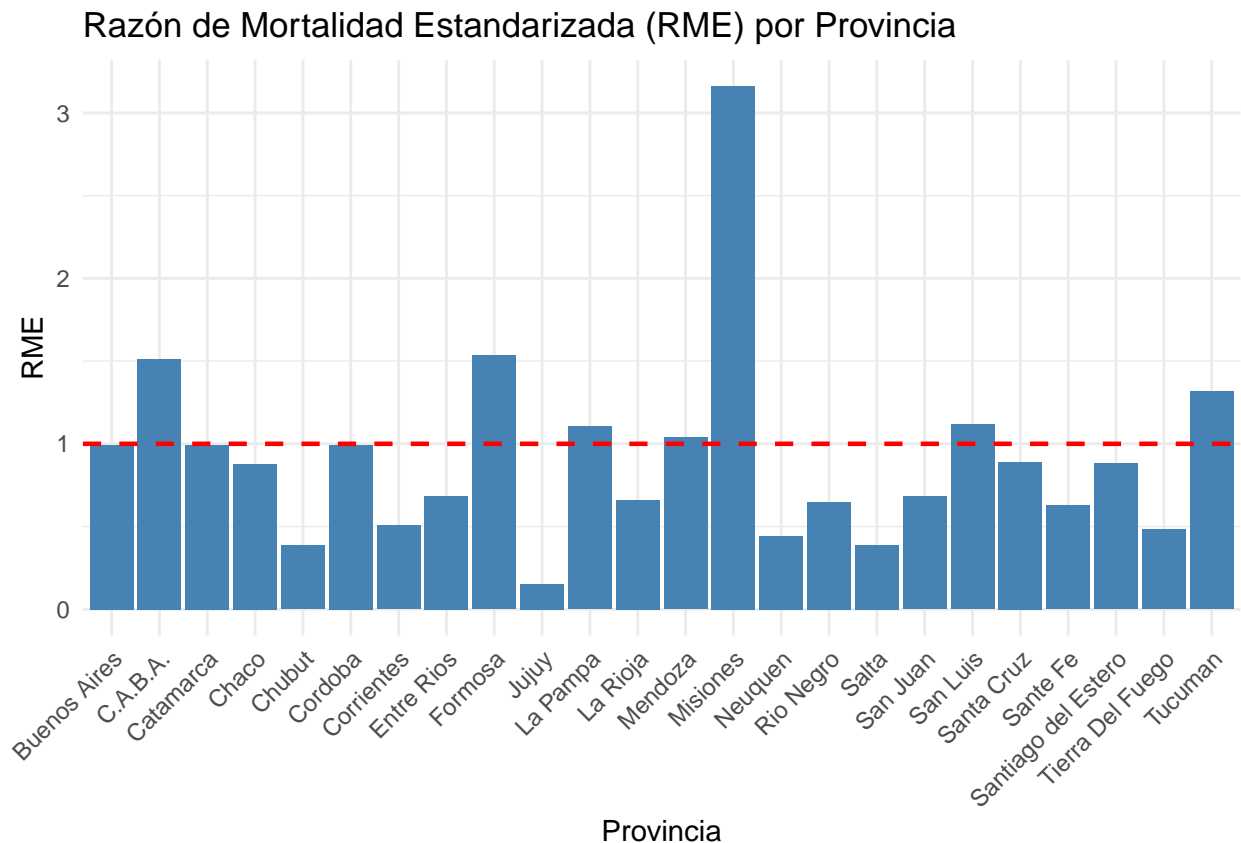
Figure 1: Tasas de Mortalidad Estandarizada por Provincia en Argentina

Razones de Mortalidad Estandarizadas

Una vez calculadas las razones de mortalidad estandarizadas para cada provincia por el método indirecto (RME), podemos interpretarlas comparándolas con el valor de 1, que actúa como nuestro punto de referencia nacional.

Table 2: Razones de Mortalidad Estandarizadas por Provincia

Provincia	RME
Buenos Aires	0.9938174
C.A.B.A.	1.5126843
Catamarca	0.9915571
Chaco	0.8772738
Chubut	0.3848355
Cordoba	0.9931523
Corrientes	0.5071060
Entre Rios	0.6836319
Formosa	1.5334733
Jujuy	0.1515395
La Pampa	1.1036673
La Rioja	0.6605521
Mendoza	1.0386637
Misiones	3.1635893
Neuquen	0.4434360
Rio Negro	0.6472000
Salta	0.3894801
San Juan	0.6815738
San Luis	1.1162325
Santa Cruz	0.8910036
Sante Fe	0.6282376
Santiago del Estero	0.8811364
Tierra Del Fuego	0.4833651
Tucuman	1.3173715



Vemos que CABA, Formosa, La Pampa, Mendoza, Misiones, San Luis y Tucumán tiene una RME mayor que 1, lo que indica que la mortalidad observada en esa provincia es superior a la que se esperaría si su población por edad hubiera experimentado las tasas de mortalidad nacionales. En otras palabras, mueren más personas de las que se prevería basándose en la estructura de edad de la provincia y las tasas de mortalidad del país.

Por el contrario, para el resto de las provincias, la mortalidad observada es inferior a la que se esperaría con las tasas de mortalidad nacionales aplicadas a su estructura de edad. Esto sugiere que mueren menos personas de las que se prevería.

Este análisis nos permite identificar facilmente qué provincias tienen una mortalidad “peor” o “mejor” de lo esperado, considerando únicamente el efecto de la distribución por edad de su población y comparándolas con un estándar nacional.

Por último, en la figura 2, observamos el mapa según RME, donde nuevamente se destacan Misiones y Jujuy como “peor” y “mejor” respectivamente.



Figure 2: Razón de Mortalidad Estandarizada (RME) por Provincia en Argentina