**Reporte Ejecutivo**

Introducción

El objetivo de nuestra investigación es analizar los datos obtenidos por la Secretaría de Salud durante el periodo de 2003 a 2014 sobre los casos de violencia intrafamiliar reportados en toda la República Mexicana. Hablaremos de la calidad de los datos y la transformación que les tuvimos que hacer para poder llevar a cabo un análisis exploratorio. Finalmente haremos comentarios y observaciones con la finalidad de que el Gobierno pueda tomar medidas de mejora.

Definición del problema

La violencia intrafamiliar es un tipo de abuso que puede incluir violencia física, violencia sexual, amenazas de violencia o abuso emocional. Su origen es debatido por muchos aunque se cree que las principales causas son falta de recursos económicos, el abuso del alcohol y la drogadicción, asociados comúnmente con la pobreza.

El problema en el cual nos centraremos es obtener las épocas en que se debe prestar mayor atención al fenómeno, cuál es la tendencia del fenómeno, cuáles son las instituciones que atienden un mayor número de casos, qué entidades, qué grupos de edad y qué sexo son más vulnerables al fenómeno para mejorar el apoyo que se brinda a las víctimas y las acciones preventivas que se toman en cada estado.

Obtención y descripción de los datos

Los datos en los que nos basamos para nuestra investigación se obtuvieron de la Dirección General de Epidemiología de la Secretaría de Salud durante el periodo de 2003 a 2014

La información está dividida en

* Casos reportados por fuentes: muestra el número total de víctimas por entidad federativa, reportado por las principales instituciones de salud.
* Casos por grupo de Edad: Presenta el número total de víctimas por entidad federativa, reportado para los grupos de edad establecidos.
* Casos de ocurrencia por mes: Presenta el número total de víctimas, por entidad federativa, reportado por mes.

Asimismo, para cada uno de estos tres grandes grupos la información se divide en tres categorías según género: Femenino, Masculino y ambos.

Los archivos se descargaron manualmente y se clasificaron en tres carpetas llamadas Fuente (Fue), Grupo (Gru) y Mes (Mes). El nombre que se les asignó fue de tal manera que su identificación fuese sencilla, por ejemplo “FueGen2009” representa al archivo en la carpeta fuente, genero general y año 2009.

La información recopilada determina que se realice un análisis de niveles y que se excluya el análisis de tasas. No se podrán eliminar ciertos sesgos ya que no se hay información disponible que ayude en la corrección de estos.

Calidad de los datos

La información fue obtenida en formato pdf y transformada mediante el programa "Tabula" a formato csv para facilitar el análisis. El análisis de los datos se realizó en "RStudio" y el código se puede encontrar en el reporte técnico rmarkdown.

Las tablas descargadas tenían múltiples errores:

* Columnas desfasadas
* Nombres y orden de columnas distinto
* Múltiples valores faltantes
* Valores faltantes diferentes (N.A. y S/R)
* Títulos que ocupaban filas enteras

Al realizar la prueba de cifras de control, notamos que algunos de los valores totales en las tablas de género General no son iguales a la suma de los totales de las tablas de géneros Masculino y Femenino.

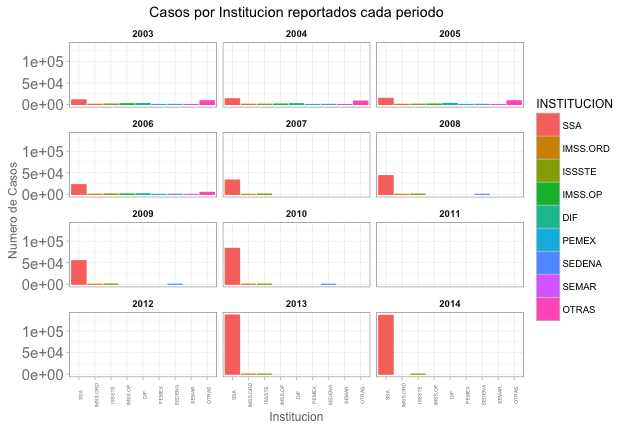
Transformación de los datos

Se realizó la lectura de los 108 archivos para los años incluidos en el periodo 2003-2014. Existen 3 tipos de bases de datos: el primer tipo es casos de violencia intrafamiliar reportada por institución y por año, el segundo tipo es casos de violencia intrafamiliar reportados por mes y por año, y, por último, los casos de violencia intrafamiliar reportados en diferentes grupos de edad por año. Cada tipo se encuentra subdivida por casos reportados por género (femenino, masculino y general).

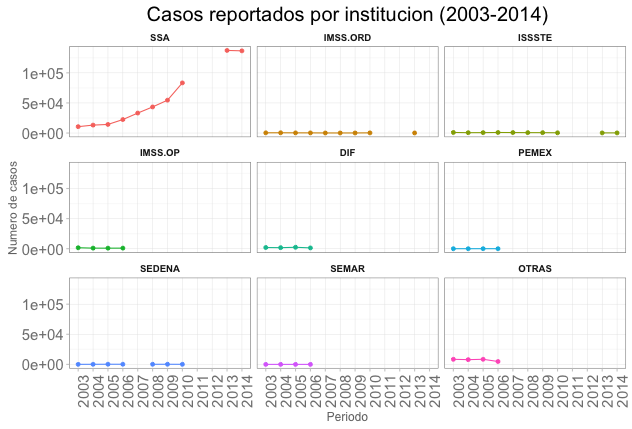
Debido al mal estado de los datos en formato CSV tuvimos que hacer una limpieza manual, eliminamos columnas vacías, las columnas de tasa se cambiaron por columnas de incidencia y, en caso de no existir, se crearon con valores faltantes. Además, eliminamos las filas que solo tenían un título.

Después de la limpieza manual utilizamos R para pasar los nombres de las columnas a mayúsculas, reducirlos a una sola palabra y eliminar los espacios sobrantes. Los valores faltantes fueron sustituidos por “NA”. Juntamos toda la información en tres data frames violencia.intrafamiliar.fuente, violencia.intrafamiliar.edad y violencia.intrafamiliar.mes. Decidimos introducir dos nuevas variables “SEXO” y “PERIODO” para poder clasificar los datos y realizar el análisis.

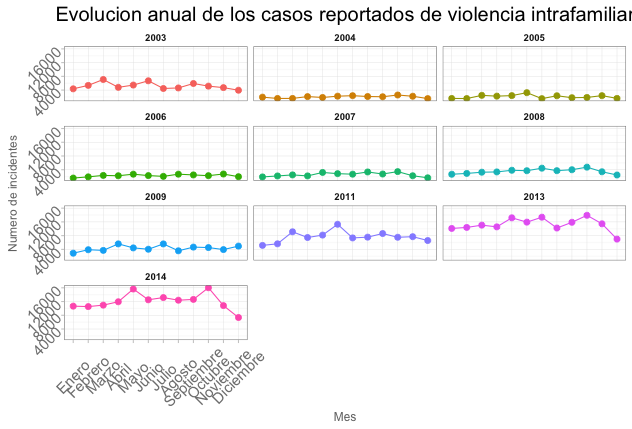
Análisis de exploración

Utilizando la tabla, se elabora la siguiente gráfica, mostrando en ella la tendencia del número de incidentes reportados por institución, durante el periodo de observación.

En esta gráfica podemos ver claramente que en cada periodo, la institución que tuvo mayores casos de violencia intrafamiliar reportados fue la Secretaría de Salud. Sin embargo, notamos que en los años 2003, 2004 y 2005 la cantidad de casos reportados a la Secretaría de Salud es similar a la suma de los casos reportados en “otras” instituciones.

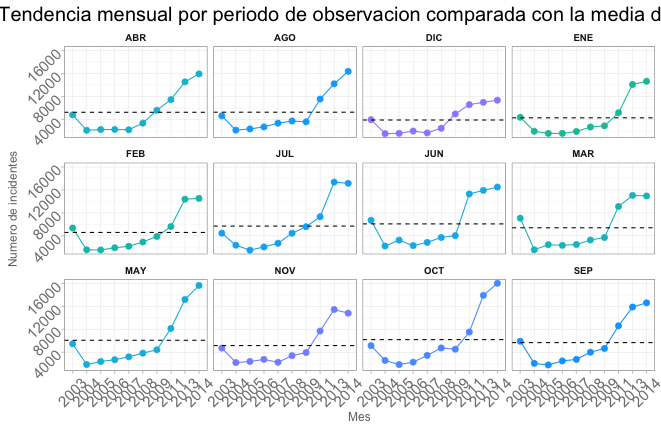


En esta gráfica podemos ver los casos que fueron reportados por institución en cada periodo. Vemos que la Secretaría de Salud empezó en el 2003 con cierto número de casos que con el paso de los años fue aumentando notablemente. La mayoría del resto de las instituciones mantuvo el número de casos reportados en el 2003, aunque con algunos pequeños cambios.



El año 2014 tuvo un número elevado de incidentes, en particular durante los meses de mayo y octubre. Otro año que presentó un número notable de incidentes fue el 2013, en particular durante los meses de mayo, julio y octubre. Los años durante los cuales no se presentaron grandes cambios son 2004, 2005, 2006, 2007, 2008 y 2009.

Tendencia mensual por periodo de observación comparada con la media



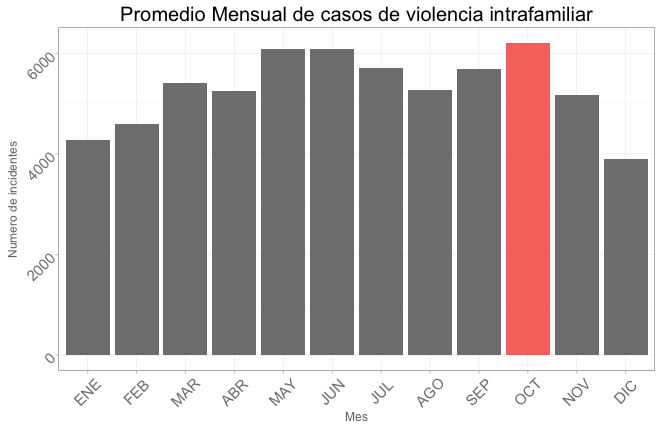
En esta gráfica podemos ver que existe una tendencia a la alza durante el periodo de observación.

A los largo de los años, la diferencia entre el número de casos registrados para hombres y para mujeres es poca, sin embargo, se registra que existe un mayor número de casos de violencia intrafamiliar hacia mujeres.

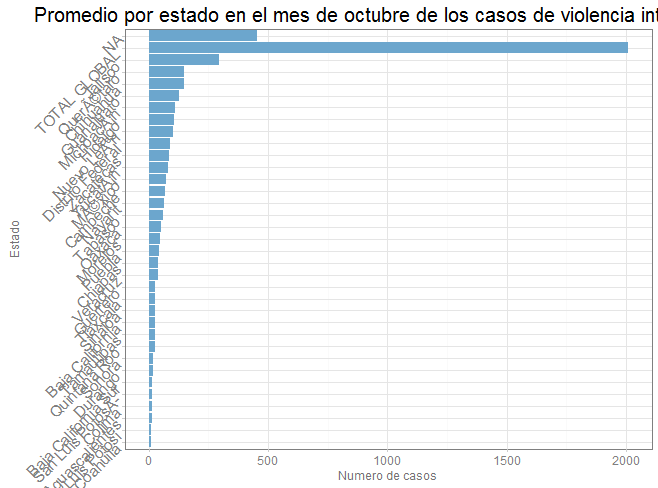
También notamos que el grupo de edad que se encuentra más propenso a sufrir violencia intrafamiliar es el de los 25 a los 44 años.

Por un lado, los grupos de edades que presentan un mayor número de casos de violencia intrafamiliar hacia hombres son aquellos menores a los 19 años, pero por otro lado, se registra un mayor número de casos de violencia intrafamiliar hacia mujeres en los grupos de edad que van de los 19 a los 60 años.

Por último, utilizando la tabla generada que contiene el promedio de incidentes mostramos los promedios mensuales de casos registrados.



Por último, para observar el fenómeno de violencia intrafamiliar en el mes de octubre se generó la siguiente tabla, misma que contiene el número promedio de violencia intrafamiliar por entidad durante el periodo de observación.



Conclusión

Desafortunadamente, los resultados de los análisis no son completamente confiables, debido al mal estado de las bases de datos. Es imperativo que el gobierno mejore sus métodos de recopilación de datos.

En general, las gráficas sugieren que el número de reportes de violencia intrafamiliar ha aumentado durante los años estudiados, pero se mantiene constante en cada año. El mayor número de reportes ocurre durante el mes de enero en cada año y se realizan en la Secretaria de Salud.

Como sugerencia para un análisis futuro se sugiera una homogenización de la información contenida en los archivos, pues como hicimos notar anteriormente existen datos que se reportan solo para ciertos grupos o años, así como definir los grupos de edad de manera uniforme, es decir, involucrando cada uno de estos la misma cantidad de años.

Fuentes

* http://www.violenciaenlafamilia.conapo.gob.mx/en/Violencia\_Familiar/Violencia\_en\_la\_Familia
* <http://www.epidemiologia.salud.gob.mx/anuario/html/anuarios.html>