



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

fertilityRates. PAQUETE DESARROLLADO EN R PARA
ESTIMAR TASAS DE FECUNDIDAD CON DATOS DE LAS
HISTORIAS DE EMBARAZOS.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaria

PRESENTA:

Yazmin Berenice González Mayorga

DIRECTORES:

Dra. María Marta Mier y Terán y Rocha

Dr. Victor Manuel García Guerrero



Ciudad Universitaria, Ciudad de México, 2021



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*“Hay una temporada para todo,
un tiempo para cada actividad bajo el cielo.
Un tiempo para nacer y un tiempo para morir.
Un tiempo para sembrar y un tiempo para cosechar.
Un tiempo para matar y un tiempo para sanar.
Un tiempo para derribar y un tiempo para construir.
Un tiempo para llorar y un tiempo para reír.
Un tiempo para entristecerse y un tiempo para bailar.
Un tiempo para esparcir piedras y un tiempo para juntar piedras.
Un tiempo para abrazarse y un tiempo para apartarse.
Un tiempo para buscar y un tiempo para dejar de buscar.
Un tiempo para guardar y un tiempo para botar.
Un tiempo para rasgar y un tiempo para remendar.
Un tiempo para callar y un tiempo para hablar.
Un tiempo para amar y un tiempo para odiar.
Un tiempo para la guerra y un tiempo para la paz.”
(Eclesiastés 3:1-8 NTV)*

Agradecimientos

A Cristo Jesús, Dios mismo, que por su sacrificio, amor y misericordia hoy puedo vivir, ser libre y experimentar su poder y gracia.

A mi mamita hermosa, que por su amor, apoyo y educación fue posible que pudiera lograr esta meta.

A mis padres, por su cuidado, guía y ejemplo de trabajo y esfuerzo.

A Lalito, por su compañía y apoyo, por luchar por mi y animarme.

¡Los amo!

Muchas son las personas que han influido de manera distinta en mi vida dentro y fuera de las aulas, sería impracticable nombrar a todos y agradecerles, pero aquí van algunos de ellos por orden de aparición en mi vida.

A Patt, por escucharme, aconsejarme y apoyarme cuando lo necesitaba.

A mis tíos, por su cariño, apoyo, guía y ayuda.

A Dani, Alan, Karla y Luz, por su hermosa amistad, cariño y por realmente estar.

A Itza, mi querida cómplice de aventuras y aprendizaje, por su compañía y apoyo en más de una ocasión para encontrar la perspectiva que necesito.

A Pamic, Lily, Pam, Migue, Paty, Fani y Lizzy, por haber hecho de la universidad una etapa muy divertida y por la oportunidad de seguir aprendiendo de ustedes.

A la Dra. Marta y al Dr. Victor, por su guía, corrección, tiempo y confianza.

A mis sinodales Mtro. José Salvador Zamora, Dra. Laura Elena Gloria y Dra. Ana Escoto por sus observaciones y comentarios.

A Yaneth, Prs. Marce y Raquel, Prs. Carlos y Naye, Rick y Paty, por sus oraciones y consejos.

¡Les agradezco mucho!

Índice general

Introducción	1
1. Conceptos y medidas para el estudio de la fecundidad	3
1.1. Las tasas en Demografía	3
1.2. El diagrama de Lexis	5
1.3. Concepto de fecundidad	7
1.4. Tasas de fecundidad específicas por edad	8
1.5. La tasa global de fecundidad. Una cohorte sintética.	9
1.6. Fuentes de datos para el estudio de la fecundidad	10
2. Las encuestas ENADID de 1992, 1997, 2006, 2009, 2014 y 2018	11
2.1. Una breve historia sobre las encuestas de fecundidad	11
2.2. Diseño conceptual	14
2.3. Referencias metodológicas	16
2.4. Diseño estadístico	18
2.4.1. Esquemas de muestreo	19
2.4.2. Tamaños de muestra	19
2.4.3. Estimadores puntuales	21
2.4.4. Estimadores de intervalo	23
3. Estimación de la fecundidad a partir de la información de la historia de embarazos de las encuestas	27
3.1. Estimación para el periodo comprendido entre enero y el mes del levantamiento de la encuesta	29
3.2. Estimación para periodos anuales anteriores al año del levantamiento de la encuesta	34
3.3. Estimación para periodos trianuales que se encuentren dentro de los cuatro años anteriores a la encuesta	36
3.4. Estimación para periodos anuales de los catorce años anteriores al año de la encuesta	39
4. Desarrollo de un paquete en R	45
4.1. Uso de paquetes y la escritura de programas	45

ÍNDICE GENERAL

4.2. Estilo de programación	46
4.3. R y su orientación a objetos	48
4.3.1. S3	49
4.3.2. S4	51
4.3.3. Reference Classes	52
4.3.4. Organización de un paquete R	53
5. Uso del paquete <i>fertilityRates</i>	55
5.1. Instalación del paquete	55
5.2. Estructura de los datos	56
5.3. Implementación del paquete	60
5.3.1. <i>frts_intvw</i> . Estimaciones en el año del levantamiento	61
5.3.2. <i>frts_yrly</i> . Estimaciones anuales en los cuatro años anteriores al levantamiento	62
5.3.3. <i>frts_3yrs</i> . Estimaciones trianuales dentro de los cuatro años anteriores al levantamiento	62
5.3.4. <i>frts_14yrs</i> . Estimaciones anuales de los catorce años anteriores al levantamiento	63
6. Estimaciones de la fecundidad con las encuestas ENADID	65
Conclusiones y consideraciones	75
A. Edad de las mujeres en la encuesta	77
B. Programación en R	81
C. Código fuente del paquete <i>fertilityRates</i>	99
Bibliografía	117

Introducción

Al estudiar una población humana desde un enfoque demográfico, podemos observar elementos característicos, como su tamaño, distribución o composición y los procesos que la determinan. Con el análisis de estos elementos, es posible diseñar y planificar políticas y programas sociales, económicos y de población de mejor manera. Este tipo de estudios conllevan el uso de datos cuantitativos que se obtienen regularmente de los censos, de las Estadísticas Vitales y de encuestas nacionales sociodemográficas. Además de contar con los datos, es importante darle un procesamiento adecuado y así obtener la información pertinente al tema de interés.

En México, el Instituto Nacional de Estadística Geografía (INEGI) es el principal proveedor de información sobre el territorio, los recursos, la economía y la población, con la cual es posible conocer las características de nuestro país y ayudar a la toma de decisiones (INEGI 2018). Algunos de los quehaceres del Instituto, son llevar a cabo los censos nacionales, las encuestas por muestreo y generar las estadísticas vitales a partir de los registros administrativos. Todo esto desde principios de la década de los ochenta, cuando fue creado.

En el ámbito del comportamiento reproductivo de la población, los datos que proporciona el Instituto constituyen un pilar fundamental para el conocimiento de los niveles y tendencias de la fecundidad en nuestro país, que son elementos indispensables para comprender la dinámica demográfica (CONAPO 2005). Las fuentes de información que brinda el INEGI sobre los nacimientos que ocurren en el país son: los Censos de Población y Vivienda, con la pregunta de la fecha de nacimiento del último hijo nacido vivo, las Estadísticas Vitales de Natalidad con los nacimientos registrados y las encuestas por muestreo, particularmente la Encuesta Nacional de la Dinámica Demográfica (ENADID), con la historia de embarazos de mujeres en edad fértil.

Este trabajo tiene como finalidad desarrollar el paquete *fertilityRates* escrito en el lenguaje de programación **R** que mediante el uso de los datos que proporcionan las historias de embarazos¹ de las encuestas, realice estimaciones puntuales y de intervalo

¹El paquete está diseñado para usarse con datos de las ENADID, pero también podría usarse

de las tasas de fecundidad específicas por edades individuales y por grupos quinquenales de edad, y de las tasas globales de fecundidad, para periodos anuales o trianuales que se encuentren dentro de los cuatro años anteriores a la fecha de levantamiento de la encuesta. Además, que se puedan obtener estimaciones puntuales de las medidas de fecundidad antes mencionadas, para el periodo de catorce años anteriores a la encuesta. Por estas razones, en este trabajo se abarcan temas que van desde conceptos y metodologías demográficas propias de la fecundidad, cuestiones puntuales de muestreo y estructuras de programación y el desarrollo de paquetes en R.

El paquete *fertilityRates* puede ser útil para las personas o instituciones que se encargan de medir y analizar la fecundidad, de manera que su empleo agilizaría su trabajo. Otra motivación por la que se llevó a cabo este trabajo es para que sirva como una guía y ejemplo de desarrollo para paquetes en dicho lenguaje. Actualmente existe una herramienta similar para Encuestas Demográficas y de Salud (DHS), pero ésta tiene su desarrollo en STATA.

El contenido del trabajo inicia con un capítulo que expone de manera detallada conceptos demográficos básicos y presenta indicadores del nivel de la fecundidad empleados en los siguientes capítulos, así como las fuentes de datos para su obtención. En el siguiente capítulo, se brinda un panorama general del muestreo empleado en las encuestas ENADID y los estimadores utilizados de acuerdo al tipo de muestreo.

El tercer capítulo consiste en la metodología para obtener estimaciones de las tasas de fecundidad a partir de las historias de embarazos. En el capítulo siguiente, se explican los elementos necesarios para la creación de un paquete en R.

En el capítulo quinto se expone el uso del paquete con datos de las diferentes encuestas ENADID: cómo descargarlo e instalarlo, cómo estructurar la base de datos para poder obtener las estimaciones, cómo estimar las tasas de fecundidad en los distintos periodos de observación. Se dedica el último capítulo a una breve comparación de los datos estimados con el paquete y las cifras oficiales.

Capítulo 1

Conceptos y medidas para el estudio de la fecundidad

En este capítulo se presentan algunos de los conceptos y medidas que apoyan el estudio y el análisis de la fecundidad. Todos estos elementos son de utilidad para familiarizar al lector con el objetivo de este trabajo, y está organizado de la siguiente manera.

En la sección 1.1 se expone de manera somera el objeto de estudio de la Demografía y los principales índices empleados, en particular las tasas demográficas así como la forma de calcularlas. En la sección 1.2 se introduce el concepto y el uso del diagrama de Lexis, y se presentan los elementos básicos para realizar un análisis. Posteriormente, en la sección 1.3 se expone brevemente el objetivo del estudio de la fecundidad, se define el término fecundidad y se mencionan algunos ejemplos de factores que la afectan. Después, en las secciones 1.4 y 1.5 se presentan los índices de mayor uso para estudiar la fecundidad, así como la manera de calcularlos. Finalmente, en la sección 1.6 se exponen las principales fuentes de datos para el estudio de la fecundidad.

1.1. Las tasas en Demografía

Para hacer un análisis de la población de interés, los demógrafos se valen de diferentes técnicas y metodologías que proporcionan características y estructuras de ésta, y así poder conocer su comportamiento. De acuerdo a [Preston et al. \(2001\)](#), la población es un conjunto de personas vivas en un punto específico en el tiempo y que cumplen ciertos criterios de inclusión

Una medida importante en demografía es la tasa, ya que mantiene una relación de magnitud (en cantidad y tiempo) entre el numerador y el denominador, brindando un panorama de comparabilidad temporal en la población de estudio.

1. CONCEPTOS Y MEDIDAS PARA EL ESTUDIO DE LA FECUNDIDAD

Las tasas demográficas hacen referencia a un fenómeno particular, la mortalidad, la fecundidad, la emigración y la inmigración. Preston et al. (2001, p. 6) señalan que el numerador de las tasas consiste en el número de eventos ocurridos, y el denominador corresponde a los años persona de exposición al riesgo de ocurrencia del evento, ambos para un periodo de tiempo determinado y una población específica, dando lugar a tasas demográficas de dimensión anual.

Así, la tasa demográfica del grupo G en el periodo de tiempo definido t es:

$$Tasa_G = \frac{\sum_{i \in G} N_i}{\sum_{i \in G} T_i} \quad (1.1)$$

donde N_i es el número total de ocurrencias en el individuo i en t y T_i son los años-persona que el individuo i estuvo expuesto al riesgo de ocurrencia durante t (Preston et al. 2001, p. 1).

Una forma de comprender la obtención de las tasas demográficas es a través de las líneas de vida, que son representaciones individuales mediante segmentos que van del nacimiento a la muerte de cada individuo, generalmente, y que a lo largo del segmento se pueden registrar eventos de interés.

Para ejemplificar lo anterior, consideremos la figura 1.1, en la que se muestra el periodo de observación t' que es de dos años naturales y la población de estudio G' que consta de 6 individuos.

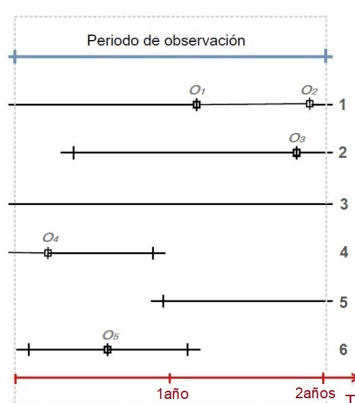


Figura 1.1: Representación de las líneas de vida para seis individuos en estudio.

El individuo 1 tuvo dos eventos O_1 y O_2 , y un tiempo de exposición al riesgo de ocurrencia T_1 de 2 años, no así el individuo 2, que tuvo un evento O_3 y un tiempo

de exposición T_2 de 1.64 años. La observación 3 participó sólo en el denominador T_3 durante 2 años, ya que durante el tiempo de referencia no tuvo evento alguno. El siguiente individuo tuvo la ocurrencia de un evento O_4 y un tiempo de exposición T_4 de 0.88 años. Finalmente, los últimos dos individuos (5 y 6) participaron con un tiempo de exposición T_5 y T_6 de 1.06 años y únicamente el último tuvo un evento O_5 .

Entonces, la tasa demográfica de G' en el periodo t' de acuerdo a la ecuación 1.1 es:

$$\begin{aligned} Tasa_{G'} &= \frac{\sum_{i=1}^6 N_i}{\sum_{i=1}^6 T_i} \\ &= \frac{5}{2 + 1.64 + 2 + 0.88 + 1.06 + 1.06} \\ &= 0.578 \end{aligned}$$

La tasa que se obtuvo se le conoce como *tasa de periodo*, ya que t es un periodo de tiempo del calendario. Existe otro tipo de tasa llamada *tasa de cohorte*, que corresponde a los individuos que pertenecen a la misma cohorte de nacimiento (o generación).

Otras tasas usadas con frecuencia son las *tasas específicas por edad*, que brindan una perspectiva del periodo de manera bruta, ya que eliminan el efecto de la estructura por edad de la población.

1.2. El diagrama de Lexis

El diagrama de Lexis es un instrumento gráfico bidimensional de análisis temporal que permite detallar y profundizar temas estadísticos y demográficos, en donde generalmente intervienen dos duraciones: la edad y el tiempo calendario. En el eje horizontal o de las abscisas se representa el tiempo del calendario (periodos) y en el eje vertical o de las ordenadas se representa la duración que es el tiempo transcurrido a partir de un evento inicial, generalmente la edad que inicia con el evento del nacimiento (ver figura 1.2).

1. CONCEPTOS Y MEDIDAS PARA EL ESTUDIO DE LA FECUNDIDAD

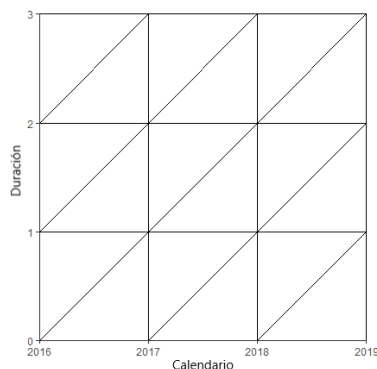


Figura 1.2: Diagrama de Lexis.

Dentro de la perspectiva de análisis que hace uso del diagrama de Lexis se encuentran la de carácter *longitudinal* y la de carácter *transversal*, es decir, la que estudia la evolución de los eventos de la población por cohortes en el tiempo (estudios generacionales) y la que estudia a los eventos de la población en un periodo de tiempo, respectivamente (ver figura 1.3).

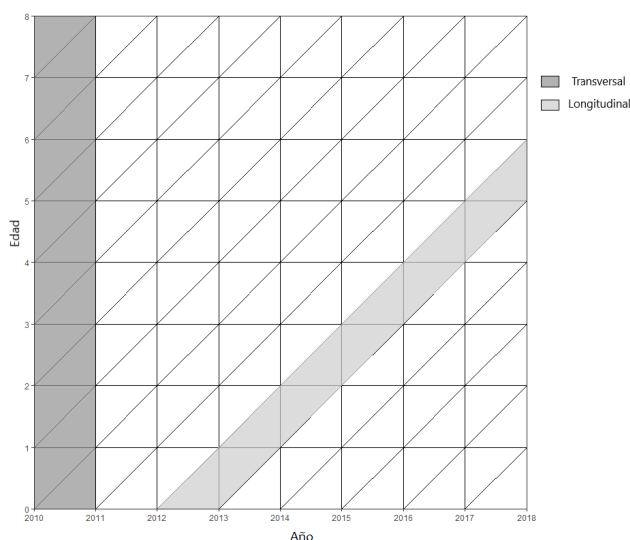


Figura 1.3: Diagrama de Lexis que muestra la cohorte de personas nacidas en el año de 2012 y el periodo de un año calendario (2010).

Un elemento de mucha utilidad para análisis longitudinales son las *líneas de vida*, presentadas en la sección anterior, pero que en el contexto del diagrama son representadas por un segmento que forma un ángulo de 45° con el eje de las abscisas y que inicia en el momento del suceso de origen (en el tiempo calendario).

Por ejemplo, en la figura 1.4 se observa la línea de vida de una persona durante sus dos primeros años de vida, cuyo suceso de origen fue el nacimiento con fecha del 30 de junio de 2015 y culmina con el cumpleaños número 2, el 30 de junio de 2017.

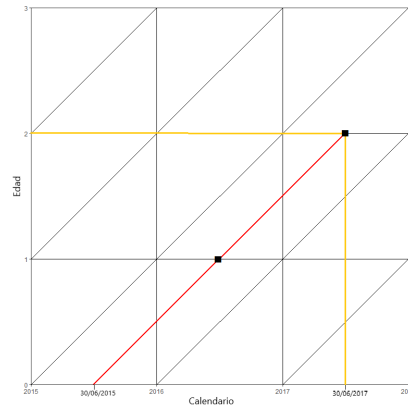


Figura 1.4: Diagrama de Lexis que muestra la línea de vida de una persona desde su nacimiento hasta que cumple uno y dos años de vida (intersecciones).

En algunos tipos de análisis, como el de supervivencia y el análisis por cohortes, es de suma utilidad hacer uso del diagrama de Lexis.

1.3. Concepto de fecundidad

“La fecundidad se refiere al proceso de incremento, en el que los miembros de una población producen nacidos vivos, es decir, nuevos miembros de la población” (Preston et al. 2001, p. 94). Además, la fecundidad es uno de los tres fenómenos principales que estudia la demografía, ya que junto con la mortalidad y la migración, determinan el monto y la dinámica poblacional pasada, presente y futura.

En el estudio de la fecundidad se investiga y analiza, desde diferentes perspectivas, la experiencia reproductiva de una población. Thomas (2018) menciona que en muchos estudios se considera el número de nacimientos que ocurren en un año específico, y suele hacerse una distinción de acuerdo a los atributos de las madres tales como raza, grupo étnico, estado civil, paridad, entre otros. Al considerar estas distinciones, el análisis de la fecundidad se vuelve más específico y permite caracterizar o describir de mejor manera a la fecundidad de la población en estudio.

El Population Reference Bureau plantea que la fecundidad

... se refiere al número de nacidos vivos que tienen las mujeres (...) y está de-

1. CONCEPTOS Y MEDIDAS PARA EL ESTUDIO DE LA FECUNDIDAD

terminada directamente por una serie de factores que, a su vez, se ven afectados por muchos otros factores: sociales, culturales, ambientales, económicos y de salud. (PRB 2011, p. 9)

Algunos de los factores involucrados en la producción de nacimientos vivos en una población que mencionan Preston et al. (2001, pp. 92, 93) son: la infecundidad temporal, la esterilidad primaria y secundaria, la actividad sexual y los comportamientos ante la anticoncepción. Además, una concepción puede no terminar en un nacimiento vivo. Lo anterior evidencia qué tan complicado puede llegar a ser el análisis de la fecundidad.

Es necesario mencionar que la fecundidad puede confundirse con fertilidad, que es la posibilidad fisiológica de reproducción, además, en el habla inglesa los términos son llamados a la inversa, *fertility* y *fecundity*, respectivamente.

1.4. Tasas de fecundidad específicas por edad

Como se mencionó anteriormente, las tasas son un instrumento muy utilizado por los demógrafos para analizar cambios en las poblaciones. En el caso del análisis de la fecundidad, las tasas específicas por edad son las de mayor uso, ya que brindan una comparación detallada del comportamiento de la fecundidad por edad. Al respecto, Preston et al. (2001, p. 21) mencionan que en la mayoría de las poblaciones, los eventos demográficos presentan una relación de ocurrencia diferente en las distintas edades.

Asimismo, las tasas de fecundidad específicas por edad pueden obtenerse controlando por periodo de tiempo, sexo y el orden en la descendencia del recién nacido, y características de la madre como son la escolaridad, el tipo de localidad de residencia, el estado civil.

El cálculo de las tasas es mediante la siguiente expresión:

$${}_n f_x(t) = \frac{\text{Nacimientos ocurridos en el periodo } t \text{ para las mujeres de edad } x \text{ a } x+n}{\text{Años-persona vividos en el periodo } t \text{ por las mujeres de edad } x \text{ a } x+n}$$

El rango de edad de la mujer que generalmente se considera va de los 15 a los 49 años, por ser consideradas edades reproductivas. Además, cuando $n = 1$, se obtienen tasas de fecundidad específicas por edad¹ simple, y cuando $n = 5$, se obtienen las tasas de fecundidad específicas por grupos quinquenales de edad. Estas últimas suelen ser las más comunes, ya que presentan de una manera suavizada los niveles de la fecundidad

¹La notación para las tasas por edad no considera el 1, es decir f_x

para el periodo de estudio, sin perder la estructura por edad.

La forma de interpretar las tasas específicas es como el número promedio o esperado de hijos nacidos vivos que tiene una mujer en cada edad o grupo de edad en un periodo o año determinado.

1.5. La tasa global de fecundidad. Una cohorte sintética.

Una medida de resumen que tiene relevancia a la hora de estudiar la fecundidad en una población para un periodo específico, es la tasa global de fecundidad, conocida como TGF por sus iniciales.

Esta tasa se obtiene de la suma de las tasas de fecundidad específicas por edad de acuerdo a la siguiente expresión:

$$TGF = \sum_{x=15}^{49} f_x \quad (1.2)$$

O bien, si las tasas de fecundidad específicas son por grupos quinquenales de edad, se obtiene con la expresión:

$$TGF = 5 * \sum_{i=3}^9 {}_5f_{5*i} \quad (1.3)$$

La tasa global de fecundidad (TGF) es un índice de gran utilidad, ya que brinda el número de hijos que tendrían las mujeres al final de su vida reproductiva si a lo largo de ésta tuvieran los patrones reproductivos observados en el periodo de estudio. Se supone que el comportamiento reproductivo de las mujeres de una cohorte ficticia es mostrado en la tasa de fecundidad específica por edad del periodo (Thomas 2018). Al mismo tiempo, la TGF resume en un número la fecundidad de todas las mujeres en edades reproductivas en el periodo (PRB 2011).

El valor de la TGF debe ser de al menos 2.1 nacimientos para mantener el reemplazo de la población.

Es por lo anterior, que para proyectar la descendencia final de una cohorte específica se hace uso de la TGF como medida hipotética, ya que en caso de querer saber con precisión cuántos hijos tendrá una cohorte de mujeres durante su vida reproductiva, sería necesario observarla a partir de los 15 años y seguirla hasta los 49 años.

1.6. Fuentes de datos para el estudio de la fecundidad

Tradicionalmente las fuentes de datos para producir estadísticas de la fecundidad son los censos de población, las estadísticas vitales de natalidad y las encuestas sociodemográficas.

Un censo de población suministra la enumeración y las características generales de los individuos en una nación para un momento específico (Thomas 2018). Entre las características que generalmente se proporcionan está la composición por sexo y edad de la población, datos sobre la fecundidad acumulada de las mujeres e información de variables socioeconómicas que pueden relacionarse con la fecundidad. En algunos censos recientes, mediante una pregunta sobre la fecha de nacimiento del último hijo nacido vivo, es posible obtener los nacimientos de los doce meses anteriores al censo y, con ellos, se puede calcular la fecundidad reciente y relacionarla con factores que influyen en el fenómeno.

Las estadísticas vitales de natalidad son registros captados por el Registro Civil que tiene cobertura universal, son procesadas y difundidas por el INEGI, brindan información de los nacimientos (personas nacidas vivas) que tuvieron un registro. Las estadísticas permiten conocer la frecuencia con que ocurren los nacimientos para caracterizar el fenómeno de la fecundidad. Sin embargo, se debe tener en cuenta que los nacimientos registrados tienen deficiencias de registro tardío, subregistro y registro múltiple de los eventos (Potter 1977). Aunado a esto, si el objetivo es obtener medidas de fecundidad mediante tasas, se requiere hacer uso de la población registrada o proyectada para determinar los denominadores.

Una encuesta por muestreo está diseñada para que las personas encuestadas sean representativas de la población examinada, lo que permite hacer inferencias para la población total con base en una muestra. Este tipo de encuestas a nivel nacional puede proporcionar información similar a la de un censo, e incluso datos más detallados o que no es posible recolectar en un evento censal, de manera que permite análisis más complejos de la fecundidad (Shryock and Siegel 2004).

Es importante mencionar que algunas de las encuestas por muestreo de carácter sociodemográfico y que contienen módulos enfocados a la fecundidad, suelen captar las historias de embarazos que son datos relevantes para el propósito que tiene este trabajo. Las encuestas descritas en el siguiente capítulo, son nuestra fuente de datos para obtener los índices mencionados en este capítulo.

Capítulo 2

Las encuestas ENADID de 1992, 1997, 2006, 2009, 2014 y 2018

En este capítulo se busca familiarizar al lector con las Encuestas Nacionales de la Dinámica Demográfica (ENADID) en el contexto histórico (sección 2.1), en la información captada y las poblaciones de estudio (sección 2.2), en las referencias técnicas y de estrategias (sección 2.3), y en cuestiones propias del muestreo (sección 2.4) necesarias para comprender adecuadamente la obtención de estimadores mediante una muestra.

2.1. Una breve historia sobre las encuestas de fecundidad

México cuenta con una gran cantidad de información estadística demográfica, producida a partir de la captación de datos en los registros administrativos, en los cuestionarios censales (censos de población) y en los cuestionarios de encuestas nacionales sociodemográficas. Lo que permite generar conocimiento del monto poblacional y sus principales características, de los niveles y tendencias de los componentes de la dinámica demográfica como son la fecundidad, la mortalidad y la migración, además de explorar los factores socioeconómicos y geográficos que determinan o inciden de forma relevante en el desarrollo general del país. El conocimiento obtenido es un elemento esencial para evaluar, revisar y modificar políticas públicas y gestión de programas, especialmente los referentes a la población, y con ello tener un adecuado sustento para la toma de decisiones (INEGI 2019a).

Para atender de forma específica, confiable, oportuna y con menor costo¹ a la demanda de datos que estadísticamente reflejen las situaciones sociodemográficas de interés, se ha recurrido al levantamiento de encuestas por muestreo, ya que éstas constituyen un mecanismo útil para la recopilación de información de la población (Gloria 2001).

¹Menor costo en comparación a los censos.

Algunas de las primeras encuestas en hogares que se han realizado en el territorio nacional sobre los componentes demográficos de la fecundidad y la mortalidad son:

- Las encuestas pertenecientes al Programa de Encuestas Comparativas (PECFAL) en áreas urbanas y rurales de América Latina en los años 1964 y 1969, respectivamente, a cargo del CELADE, las cuales permitieron la estimación de los niveles de fecundidad, conocer con más detalle las diferencias de la fecundidad entre grupos y evaluar los mecanismos que propician diferencias en los niveles de fecundidad (CELADE 1982).
- La Encuesta Mexicana de Fecundidad (EMF) de 1976, fue la primera en brindar información demográfica (nupcialidad, fecundidad y mortalidad) con cobertura geográfica a nivel nacional. Además, la encuesta formó parte del programa internacional WFS (World Fertility Survey) y tuvo una gran relevancia internacional. El proyecto estuvo a cargo de la Dirección General de Estadística, con el apoyo financiero del Fondo de Naciones Unidas para Actividades de Población, la cooperación del Instituto de Investigaciones Sociales de la UNAM, la asesoría del COLMEX, el Consejo Nacional de Población (CONAPO) y el Sistema de Información para la Planificación Económica y Social (Ordorica and Potter 1981).
- La Encuesta Nacional Demográfica (END) de 1982, realizada por el CONAPO, en coordinación con la Dirección General de Estadística y la Red Móvil de Promotores del DIF, aborda temas específicos de nupcialidad, salud materno-infantil, mortalidad y fecundidad (Mojarro and Núñez 1988).
- La Encuesta Nacional de Fecundidad y Salud (ENFES), cuyo operativo de campo se llevó a cabo en 1987 y con cobertura geográfica nacional, se realizó como parte del Programa Mundial de Encuestas Demográficas y de Salud (DHS), con la finalidad de proporcionar información esencial acerca de la fecundidad, familia, mortalidad infantil y salud materno-infantil. La llevó a cabo la Dirección General de Planificación Familiar de la Secretaría de Salud, con el respaldo financiero y técnico del Instituto para el Desarrollo de los Recursos y el Consejo Nacional de Ciencia y Tecnología (CONACyT), con la participación de la Dirección General de Estadística (Secretaría de Salud/DHS 1990).

En la actualidad se cuenta con otras encuestas que brindan información más completa y abundante de interés demográfico. Lo anterior ha sido derivado de los avances de infraestructura en tecnología de información y comunicación, al interés creciente por estudiar la dinámica demográfica, y a la responsabilidad de dependencias como INEGI, CONAPO, el Instituto Nacional de Salud Pública (INSP) y la Secretaría de Salud en generarla, procesarla y dar a conocer resultados de lo recabado.

Con la finalidad de proporcionar información sobre el perfil sociodemográfico del país que permitiera hacer estimaciones de calidad y con igual profundidad para cada

componente de la dinámica demográfica, así como incorporar preguntas clave para evaluar el Sistema Nacional de Estadísticas Vitales, profundizar en variables captadas en el XI Censo General de Población y fortalecer el Sistema Nacional de Información¹, el INEGI en el año de 1992, levantó la Encuesta Nacional de la Dinámica Demográfica (ENADID). Además, la ENADID en comparación² a las encuestas antes mencionadas, generó información a nivel de entidad federativa para indicadores demográficos básicos (INEGI 1992).

A finales del año 1997, el INEGI con la propósito de actualizar e incrementar los datos brindados por la ENADID de 1992, levantó la segunda edición de la ENADID. Se tomó como base la encuesta previa e incorporó temas de salud marterno-infantil, historia de uniones, anticoncepción y preferencias reproductivas, para así seguir consolidando el Sistema Nacional de Información Estadística (INEGI 1999).

La tercera edición de la ENADID fue realizada en 2006 y coordinada por el INSP, como parte de un proyecto de cooperación interinstitucional entre la Dirección General de Información en Salud y el Centro Nacional de Equidad de Género y Salud Reproductiva de la Secretaría de Salud, el CONAPO y el INEGI. En esta encuesta se tomó como punto de partida la ENADID 1997 y las encuestas anteriores para tener conceptos homogéneos y que existiera comparabilidad entre éstas, asimismo se sumó al Sistema Nacional de Información Estadística y Geográfica³ (INSP 2008b).

La siguiente encuesta de la dinámica demográfica, elaborada por parte del INEGI y en colaboración con el CONAPO, bajo el lineamiento de la Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG)⁴, se levantó en 2009, continuando con la serie de encuestas nacionales en hogares. Ésta agregó variables para conocer

¹El Sistema Nacional de Información se creó a finales de los años 70's, bajo la premisa que se sustentarían los procedimientos de toma de decisiones que el entonces presidente José López Portillo, tomaría durante su mandato, y teniendo en su quehacer el suministrar de forma oportuna, confiable y suficiente la información generada (INEGI 2009).

²La comparación es en relación a temas que sí son abordados en las encuestas mencionadas, pero que no tienen el mismo desglose geográfico que la ENADID 1992.

³El antes Sistema Nacional de Información, cambió a Sistema Nacional de Información Estadística y Geográfica el 4 de octubre de 2005, publicado en la Gaceta parlamentaria, y cuyos datos ya serían considerados como oficiales (INEGI 2009). Para el 7 de abril de 2006 se publicó en el Diario Oficial de la Federación (DOF) el decreto por el que se declaran reformados los artículos 26, fracción XXIX-D de la Constitución Política de los Estados Unidos Mexicanos, incorporando que el Estado contará con un Sistema Nacional de Información Estadística y Geográfica (SNIEG) y que la responsabilidad de normarlo y coordinarlo será de un organismo con autonomía técnica y de gestión, personalidad jurídica y patrimonio propios.

⁴Con la creación de la LSNIEG publicada por el DOF el 16 de abril de 2008; se estableció en el Capítulo IV, Sección I, Art. 21, que el Subsistema Nacional de Información Demográfica y Social debería generar un conjunto de indicadores clave, que atenderían como mínimo los temas siguientes: población y dinámica demográfica, salud, educación, empleo, vivienda, distribución de ingreso y pobreza.

niveles y tendencias de la mortalidad infantil, y actualizó la información de fecundidad, mortalidad, migración y factores que afectan cada fenómeno (INEGI 2010).

En 2014, siguiendo con la producción de información sociodemográfica derivada de las encuestas, se levantó la ENADID, a cargo del INEGI, de conformidad con los Artículos 21 y 22 de la LSNIEG y con determinación de que fuera Información de Interés Nacional¹. Se incorporaron, además, nuevos temas como discapacidad, pertenencia étnica, alfabetismo y se ampliaron otros como anticoncepción y salud (INEGI 2015a).

La versión más reciente, levantada en 2018, da continuidad a los temas que se tratan en otras ediciones y bajo el marco de la LSNIEG, proporcionando datos que son guía y sustento para planes y decisiones de carácter nacional y objetivos internacionales (INEGI 2019a). De forma general, se puede decir que las encuestas ENADID mantienen cierta comparabilidad conceptual.

2.2. Diseño conceptual

Las encuestas en hogares ENADID brindan un panorama amplio en temas de carácter social, económico y demográfico, en mayor o menor medida según la versión de la encuesta y de acuerdo a los requerimientos temporales de cada una. Sin embargo, debido al interés en contar con información equiparable, se ha buscado realizar las mejores prácticas a través del tiempo, incorporando las sugerencias de organismos internacionales, principalmente Naciones Unidas a través del programa de encuestas DHS (INEGI 2016; INEGI2019a). En ese sentido, a partir de 2009 se incluyeron recomendaciones internacionales para la definición del perfil del personal de campo (INEGI 2010) y para los últimos dos levantamientos de la encuesta se buscó la comparabilidad internacional, al tomar en cuenta las recomendaciones respecto a la información por captar, las definiciones y clasificaciones, así como de procesos técnicos y metodológicos (INEGI 2016; INEGI2019a). A continuación se presentará un resumen comparativo de las seis encuestas acerca de la información captada y las poblaciones de estudio que aborda cada una.

Método de recolección

La manera en como se captó la información en todas las encuestas fue mediante una entrevista directa, aunque el medio de captación es un *cuestionario para el hogar* con un *módulo para la mujer*. Esto es diferente para las primeras dos encuestas (1992 y 1997), ya que en éstas sólo se presenta un cuestionario con secciones aplicables a las diferentes poblaciones objetivo entre quienes se encuentran las mujeres en edades reproductivas.

¹Acuerdo que se dió el 10 de noviembre de 2014, con base en la reforma de los Artículos 77, fracción II y 78 de la LSNIEG publicada en el DOF.

Cobertura temática

Los temas que aborda cada ENADID son:

	ENADID					
	1992	1997	2006	2009	2014	2018
Características de la vivienda	✓	✓	✓	✓	✓	✓
Residentes, hogares y lista de personas	✓	✓	✓	✓	✓	✓
Características de las personas	✓	✓	✓	✓	✓	✓
Migración interna e internacional	✓	✓	sólo internacional	sólo internacional	sólo internacional	sólo internacional
Fecundidad e historia de embarazos	✓	✓	✓	✓	✓	✓
Certificado y registro, de nacimientos y defunciones	sólo registro	sólo registro	sólo registro	-	✓	✓
Preferencias reproductivas	-	✓	✓	✓	✓	✓
Anticoncepción	✓	✓	✓	✓	✓	✓
Salud materno infantil (último embarazo)	-	✓	✓	✓	✓	✓
Nupcialidad	-	✓	-	✓	✓	✓

La elaboración de cada encuesta tomó como punto de partida las versiones previas, aunque la profundidad con la que se tocan los temas varía en cada encuesta.

Unidades de análisis

Para atender a los requerimientos conceptuales que cada ENADID plantea se definieron las poblaciones objetivo, y poder así garantizar que la información recolectada para cada tema sea la más adecuada. Las poblaciones objetivos son:

2. LAS ENCUESTAS ENADID DE 1992, 1997, 2006, 2009, 2014 Y 2018

	ENADID					
	1992	1997	2006	2009	2014	2018
Las viviendas del territorio nacional habitadas	✓	✓	✓	-	✓	✓
Los hogares de nacionales y extranjeros de las viviendas	✓	✓	✓	-	✓	✓
Residentes habituales de las viviendas seleccionadas	✓	✓	✓	✓	✓	✓
Mujeres de 15 a 54 años de edad, residentes habituales	✓	✓	✓	✓	✓	✓
Migrantes internacionales en los últimos cinco años, que en el momento de su última partida eran residentes habituales	✓	✓	✓	-	✓	✓
Residentes habituales que fallecieron en los últimos cinco años	✓	✓	-	-	-	-
Hijos nacidos vivos e hijos nacidos vivos actualmente fallecidos	-	-	✓	-	-	-

Las unidades de análisis son muy importantes para el diseño muestral, ya que son las que hacen posible la realización de estimaciones de las características de interés de la población en estudio. Por esto mismo, son seleccionadas con criterios probabilísticos.

2.3. Referencias metodológicas

Informante adecuado

Una de las condiciones para que se dé de manera conveniente la aplicación de la encuesta a la vivienda seleccionada, es su aplicación a informantes adecuados para cada módulo. En el caso de las seis encuestas, fue necesario ubicar a una persona de 15 años o más de edad, residente de la vivienda que conociera las características de ésta. Además, es importante que la persona conociera la información de los integrantes de cada hogar dentro de la vivienda. Para el módulo de la mujer, la informante adecuada es cada una de las mujeres de 15 a 54 años, residentes habituales del hogar.

Desglose geográfico

Cada encuesta fue diseñada para generar información a nivel nacional primeramente, y a diferentes niveles de desagregación según la característica y el periodo de referencia.

En el caso de la ENADID 1992 es posible obtener la paridez media por grupos quinquenales de edad de la mujer y la tasa global de fecundidad del periodo 1987-1991, entre otros, para regiones, cuatro tamaños de localidad¹ y cada entidad federativa. La encuesta de 1997 permite obtener información representativa para localidades de menos de 15 000 habitantes y de 15 000 y más habitantes, y por entidad federativa para características generales de la población.

El nivel de desagregación para la ENADID 2006 es de cuatro tamaños de localidad² únicamente y en el caso de la encuesta de 2009 la cobertura es estatal y por tamaño de localidad³. Para las encuestas de 2014 y 2018 los niveles son estatal y por tamaños de localidad, sin embargo, se hace una distinción entre las variables captadas en el cuestionario del hogar y las captadas en el módulo de la mujer, con cuatro y dos tamaños de localidades, respectivamente.

Esta información es relevante, ya que en caso de querer obtener estimaciones para regiones geográficas específicas se tiene que verificar que la encuesta esté diseñada para dar resultados a ese nivel, con la consideración de las variables involucradas.

Tomando en cuenta el dominio de estudio de este trabajo, las estimaciones generadas por el paquete son para nivel nacional, es decir, ocupando todos los casos de interés captados por cada encuesta.

¹Cada ENADID delimita las localidades de acuerdo al marco muestral utilizado.

²En la documentación de la ENADID 2006 no especifica qué características se consideran en esta desagregación.

³En la documentación de la ENADID 2009 no se especifica qué características se consideran para los niveles de desagregación.

Periodos de levantamiento

Los operativos de levantamiento para la respectiva *ENADID* se desarrollaron en los períodos:

- del 31 de agosto al 30 de noviembre de *1992*
- del 8 de septiembre al 15 de diciembre de *1997*
- del 3 de abril al 22 de mayo del *2006*
- del 18 de mayo al 10 de julio de *2009*
- del 4 de agosto al 26 de septiembre de *2014*
- del 13 de agosto al 5 de octubre de *2018*

Estos datos son relevantes para el propósito de este trabajo, ya que en la obtención de las tasas del año de la encuesta se requiere previamente la fecha de la entrevista (mes, año).

Marcos muestrales

El marco muestral es un elemento esencial a la hora de elaborar un muestreo probabilístico, ya que lista los elementos poblacionales de los que se quiere hacer inferencias.

Para el levantamiento de las encuestas de 1992 y 1997 se elaboraron específicamente marcos de muestreo constituidos por listas de viviendas que conformaron la Unidad Primaria de Muestreo¹ (UPM). En las *ENADID* 2006 y 2009 se usó el Marco Nacional de Viviendas 2002 del INEGI, construido a partir de la información cartográfica y demográfica que se obtuvo del XII Censo General de Población y Vivienda de 2000. Finalmente, en las *ENADID* 2014 y 2018 se considera el Marco Nacional de Viviendas de 2012 del INEGI, construido a partir de información cartográfica y demográfica obtenida del Censo de Población y Vivienda de 2010.

2.4. Diseño estadístico

Esta sección se exponen los datos y aspectos de mayor relevancia en los diseños muestrales² de cada encuesta *ENADID*.

¹Las UPM esán constituidas por agrupaciones de viviendas con características diferenciadas, dependiendo del ámbito al que pertenecen: rural o urbano.

²Las encuestas *ENADID* se pueden considerar Encuestas Complejas (E.C.) con base en las características propuestas por [Wolter \(2007\)](#), entre las que destacan el grado de complejidad del diseño de muestreo, el grado de complejidad de los estimadores de la encuesta y el tamaño de

2.4.1. Esquemas de muestreo

El diseño que se emplea en cada encuesta es:

- Para la ENADID 1992, estratificado¹ y bietápico².
- Para la ENADID 1997, estratificado y bietápico.
- Para la ENADID 2006, estratificado, trietápico³ y por conglomerados⁴.
- Para la ENADID 2009, estratificado, bietápico y por conglomerados.
- Para la ENADID 20014, estratificado, bietápico y por conglomerados.
- Para la ENADID 2018, estratificado, bietápico y por conglomerados.

La vivienda es la última etapa de selección en todas las encuestas, ya que todos los hogares al interior de la misma, así como las otras unidades de estudio, son seleccionadas con certeza.

Es necesario mencionar que el muestreo empleado en la selección de muestras varía dependiendo de la zona o la ENADID y determina los casos para el cálculo de las probabilidades de selección de las viviendas.

2.4.2. Tamaños de muestra

Debido a que las encuestas ENADID tienen propósitos múltiples, calcular el tamaño de muestra de manera que abarque cada caso de estudio, puede resultar complejo. Es por eso que para cada encuesta se seleccionaron indicadores clave para calcular el tamaño de muestra con base en ellos. Estos indicadores se eligieron de manera que se tenga un tamaño de muestra que atienda de manera simultánea al resto de los indicadores y regularmente son de baja frecuencia entre la población objetivo.

ésta. Esto es relevante, ya que la manera de hacer inferencias para las características de interés con base en su diseño muestral respectivo requiere de análisis complejos, sin embargo, al ser consideradas E.C. es posible hacer uso de métodos o técnicas de aproximación “más simples” para estimar los parámetros de interés.

¹Este tipo de muestreo regularmente se emplea en caso de que la variable que es de interés estimar, toma valores similares en diferentes subpoblaciones (estratos) homogéneas de la población total (Crochran 1977).

²Se refiere a que la selección de la muestra final se realiza en dos etapas.

³Se refiere a que la selección de la muestra final se realiza en tres etapas.

⁴El muestreo por conglomerados divide a la población total en un número determinado de subdivisiones (conglomerados) relativamente pequeñas y se seleccionan solo algunas de ellas para incluirlas en la muestra. Este tipo de muestreo es empleado cuando no se tienen un marco muestral, o bien, la encuesta es de un tamaño considerable (Särndal et al. 1992), como es el caso de las ENADID

Tamaños de muestra para proporciones

La fórmula empleada para su cálculo es:

$$n = \frac{z^2}{r^2} \cdot \frac{q}{p} \cdot \frac{DEFF}{(1 - TNR)PMV^*}$$

Donde:

n = el tamaño de la muestra.

p = la estimación de la proporción de interés.

$q = 1 - p$.

r = error relativo máximo aceptable.

z = valor asentado en las tablas estadísticas de la distribución normal estándar para una confianza prefijada.

$DEFF$ = efecto del diseño definido como el cociente de la varianza en la estimación del diseño utilizado entre la varianza obtenida, con un muestreo aleatorio simple para un mismo tamaño de muestra.

TNR = tasa de no respuesta máxima esperada.

PMV = promedio de años de exposición de las mujeres de 15 a 54 años de edad por vivienda. * Este promedio no es incluido en algunos los cálculos.

Los datos para sustituir en la fórmula son obtenidos mediante la información proporcionada por el marco muestral de cada encuesta y por el criterio de el encargado de realizar el muestreo.

Las proporciones o índices de interés considerados para estimar los tamaños de muestra de manera que se cubran los objetivos de cada ENADID son:

- La tasa de mortalidad infantil para la ENADID de 1992.
- A nivel nacional se tomó la proporción de la población migrante internacional del quinquenio 1990-1995 y a nivel entidad federativa se tomó el valor mínimo en la proporción de mujeres de 15 a 49 años embarazadas y unidas para la ENADID de 1997.
- La proporción de la población con migración internacional para la ENADID de 2006.

- La tasa de fecundidad general para la ENADID de 2014 y 2018.

Para la ENADID de 2009 no hay referencia del índice que se utilizó.

Regularmente se escogen índices de baja frecuencia entre las poblaciones objetivo para garantizar que los demás índices de interés queden cubiertos por el tamaño de muestra calculado.

Los tamaños de muestra obtenidos en cada ENADID fueron:



- 64 mil viviendas a nivel nacional (2 mil en cada entidad federativa) en 1992.
- 80 mil viviendas a nivel nacional (2 mil 500 en cada entidad federativa). en 1997.
- 42 mil viviendas a nivel nacional (1 mil 800 en el Distrito Federal, Jalisco, Estado de México y Veracruz, 1 mil 600 en Guanajuato, Nuevo León y Puebla, y 1 mil 200 en el resto de entidades federativas) en 2006.
- 101 mil viviendas a nivel nacional (entre 3 mil y 3 mil 500 en cada entidad federativa, y 4 mil en el Distrito Federal) en 2009.
- 101 mil 389 viviendas a nivel nacional (3 mil en cada entidad federativa y 4 mil en el Distrito Federal) en 2014.
- 119 800 viviendas particulares a nivel nacional (entre 3 mil 500 y 4 mil en cada entidad federativa) en 2018.

2.4.3. Estimadores puntuales

Al hablar de encuestas probabilísticas por muestreo, las características o parámetros pertenecientes a la población muestrada son tratados como estimadores de la población,

ya sea del total o de un subconjunto de ésta. Para propósitos de este trabajo, los estimadores de interés derivados¹ de las encuestas ENADID son razones y suma de razones (tasas y suma de tasas) para el total de la población (total nacional).

Estimador de tasas específicas de fecundidad por edad

En la sección 1.4 se muestra la fórmula para obtener las tasas específicas de fecundidad por edad, que en términos generales, es el cociente de dos características de la población. La forma de considerar estas tasas como estimadores de razón bajo el esquema de muestreo de las ENADID es la siguiente:

El estimador del total de $Y =$ nacimientos ocurridos en el periodo t para las mujeres de edad x a $x + n$, se define como:

$$\hat{Y} = \sum_{h \in E} \sum_{i \in U} \left(\sum_{s \in V} F_{his}^{UA} \sum_{l \in M} Y_{hisl}^{UA} \right) + \sum_{h \in E} \sum_{i \in U} \left(\sum_{s \in V} F_{his}^{CU} \sum_{l \in M} Y_{hisl}^{CU} \right) + \sum_{h \in E} \sum_{i \in U} \left(\sum_{s \in V} F_{his}^{RR} \sum_{l \in M} Y_{hisl}^{RR} \right)$$

Donde:

E = Estratos.

U = Unidades Primarias de Muestreo.

V = Viviendas.

M = Mujeres.

F_{his}^{UA} = factor de expansión final de la s -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, en el dominio² urbano alto.

Y_{hisl}^{UA} = valor observado de la característica de interés Y en la l -ésima mujer, en la s -ésima vivienda, en la i -ésima UPM, en el h -ésimo estrato, en el dominio urbano alto.

F_{his}^{CU} = factor de expansión final de la s -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, en el dominio complemento urbano.

¹Son derivados debido a que el numerador y denominador son características resultantes (no directas) de la información contenida en las historias de embarazos. Esta información es revisada en el siguiente capítulo.

²En el caso de las encuestas de 1992 y 1997 los dominios cambian.

Y_{hisl}^{CU} = valor observado de la característica de interés Y en la l -ésima mujer, en la s -ésima vivienda, en la i -ésima UPM, en el h -ésimo estrato, en el dominio complemento urbano.

F_{his}^R = factor de expansión final de la s -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, del dominio rural.

Y_{hisl}^R = valor observado de la característica de interés Y en la l -ésima mujer, en la s -ésima vivienda, en la i -ésima UPM, en el h -ésimo estrato, del dominio rural.

El estimador del total de X = años-persona vividos en el periodo t por las mujeres de edad x a $x + n$, se define de forma análoga a \hat{Y} , por lo que el estimador de razón o de tasa específica por edad es:

$${}_n\hat{f}_x = \hat{R} = \frac{\hat{Y}}{\hat{X}}$$

Los periodos de tiempo t en este trabajo son especificados en el capítulo 3, por edad individual y grupos quinquenales de edad, de los 15 a los 49 años de edad ($n = 1, 5$).

Estimador de la TGF

El cálculo de la TGF visto en la sección 1.5 es mediante la suma de las tasas específicas de fecundidad por edad o por grupos quinquenales de edad. Obtener esta tasa bajo el esquema de muestreo, no es más que la suma de estimadores de razón de las tasas específicas como se muestra en las siguientes expresiones para edades individuales y grupos quinquenales de edad, respectivamente.

$$\widehat{TGF} = \sum_{x=15}^{49} \hat{f}_x \quad (2.1)$$

O bien, con la expresión:

$$\widehat{TGF} = 5 * \sum_{i=3}^9 {}_5\hat{f}_{5*i} \quad (2.2)$$

2.4.4. Estimadores de intervalo

Los intervalos de confianza o estimadores de intervalo proporcionan, con alta confianza, un rango de valores en el que se encuentra el valor poblacional (o parámetro) de la característica en estudio. Además, de cierta manera ayudan a evaluar los errores de muestreo de las estimaciones, al contener de manera intrínseca la estimación de la

varianza de la estadística en cuestión. Esto es importante, ya que por ser una muestra de la población, existen ciertos grados de imprecisión debidos al azar.

Estimador de tasas específicas de fecundidad

Como se mencionó, es necesario obtener la varianza del estimador para conocer el intervalo de confianza. El hecho que el tipo de muestreo sea complejo, hace que el cálculo de la varianza pueda resultar difícil en términos de costo y tiempo, debido a la complejidad de su estimador.

Es por eso que Hansen et al. (1953) proponen el *método de conglomerados últimos* con la finalidad hacer más simple la obtención de la varianza en este tipo de encuestas. Este método es utilizado por Lumley (2010) en su paquete *survey* de R, fundamental para el desarrollo del paquete *fertilityRates*.

Además de lo anterior, las tasas o estimadores de razón son estadísticas no lineales, por lo que una opción para aproximar el valor de la varianza es por el método de linealización también conocido por el método de series de Taylor. Los supuestos que debe cumplir para aplicar este método es que $\hat{\theta}$ estimador sea consistente para θ parámetro y debidamente diferenciable (Skinner et al. 1989).

Bajo ambas metodologías, el estimador de la varianza de las tasas es:

$$\hat{V}(n\hat{f}_x) = \frac{1}{\hat{Y}} \sum_{h \in E} \frac{n_h}{n_h - 1} \sum_{i \in U} \left[\left(\hat{X}_{hi} - \frac{1}{n_h} \hat{X}_h \right) - n\hat{f}_x \left(\hat{Y}_{hi} - \frac{1}{n_h} \hat{Y}_h \right) \right]^2$$

Donde:

\hat{X}_{hi} = total ponderado de la variable de estudio X en la *i-ésima* UPM, en el *h-ésimo* estrato.

\hat{X}_h = total ponderado de la variable de estudio X en el *h-ésimo* estrato.

\hat{Y}_{hi} = total ponderado de la variable de estudio Y en la *i-ésima* UPM, en el *h-ésimo* estrato.

\hat{Y}_h = total ponderado de la variable de estudio Y en el *h-ésimo* estrato.

n_h = número de UPM en el *h-ésimo* estrato.

Finalmente, los intervalos de confianza al $100(1 - \alpha) \%$ se construyen de la siguiente forma:

$$I_{1-\alpha} = \left(n\hat{f}_x - Z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(n\hat{f}_x)}, n\hat{f}_x + Z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(n\hat{f}_x)} \right)$$

Donde:

$Z_{1-\frac{\alpha}{2}}$ = cuantil $1 - \frac{\alpha}{2}$ de una t de Student con grados de libertad $n - 1$, n el tamaño de la muestra.

Estimador de la TGF

Para estimar el intervalo de confianza de la TGF se requiere el uso de una técnica de remuestreo llamada Jackknife (JK3) para estimar la varianza de la tasa, mediante la “eliminación” de un conglomerado¹ a la vez y obtener el estimador de la TGF tantas veces como conglomerados se tengan. Lo anterior, debido a que el método de Taylor no puede ser empleado por lo complejo del indicador.

Se obtiene la varianza de la \widehat{TGF} a través de la siguiente fórmula (Särndal et al. 1992):

$$\hat{V}(\widehat{TGF}) = \frac{1}{A(A-1)} \sum_{a=1}^A \left(\widehat{TGF}_{(a)} - \widehat{TGF} \right)^2$$

Donde:

\widehat{TGF} es el estimador de la TGF señalado en la ecuación 2.1.

$\widehat{TGF}_{(a)}$ es la estimación calculada a partir de la muestra reducida a $A - 1$ conglomerados.

A es el número total de conglomerados.

De igual manera que para las tasas específicas por edad, el intervalo de confianza para la \widehat{TGF} al $100(1 - \alpha)\%$ se construye como:

$$I_{1-\alpha} = \left(\widehat{TGF} - Z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\widehat{TGF})}, \widehat{TGF} + Z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\widehat{TGF})} \right)$$

Es importante señalar que en el caso que el diseño muestral de la encuesta no involucre un esquema por conglomerados, como es el caso de las encuestas de 1992 y 1997, se particionará la muestra en grupos aleatorios para así poder aplicar el método de

¹El conglomerado conforme al diseño de la encuesta.

remuestreo.

En este capítulo pudimos conocer a detalle las características de los datos que utilizaremos para el desarrollo e implementación del paquete *fertilityRates*. Tener conocimiento de las características conceptuales, metodológicas y técnicas de las encuestas, permite obtener estimaciones de las cuales se deriven inferencias válidas, incluso, tener este contexto sirve de referencia para poder adaptar el paquete a otras encuestas con características similares.

El siguiente capítulo presenta la metodología para obtener los estimadores vistos previamente, que son aplicados a diferentes periodos. Además, los estimadores que son de nuestro interés, no se obtienen de manera directa con los datos de las encuestas, sino que es necesario realizar cierto procesamiento a los datos.

Capítulo 3

Estimación de la fecundidad a partir de la información de la historia de embarazos de las encuestas

Hasta hace 45 años, la medición de la fecundidad en México se hacía únicamente con la información de los nacimientos reportados en las estadísticas vitales y la fecundidad acumulada en los censos de población (CONAPO 2005). A partir de entonces, se empezó a hacer uso de los datos de las historias retrospectivas de embarazos de las mujeres en edad fértil recogidos en las encuestas. Las historias de embarazos consisten en la recopilación de todas las fechas de los nacimientos y los términos de embarazos que terminan en pérdidas y abortos de la mujer a lo largo de su vida, hasta el momento de la encuesta. Esta información permite la estimación de tasas de fecundidad y, con ellas, conocer estructuras por edad y tendencias en el nivel de la fecundidad en el tiempo.

Son muchas las formas de abordar los datos de fecundidad recabados en las historias de embarazos. En este trabajo, es de interés la estimación de la fecundidad de forma directa, a través de los datos que proporcionan las encuestas para diferentes periodos retrospectivos. La metodología empleada y que se presenta en este capítulo, se basa en la propuesta por Moultrie (2013).

Como se mencionó en el primer capítulo, para estudiar la fecundidad con frecuencia se recurre al cálculo de tasas. Para obtener el numerador, es necesaria la información de cada nacimiento en conjunto con la de su madre, y para tener el denominador, la información de cada mujer de 15 a 54 años. La información¹ para realizar esto es:

¹Regularmente esta información es dada en dos archivos diferentes (mujer y embarazos), la estructura de los datos que requiere el paquete son detallados en la sección 5.2.

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

- Conjunto de datos de la mujer
 - Mes y año de nacimiento
 - Edad en la entrevista
 - Los pesos muestrales para ponderar los datos
- Conjunto de datos del niño
 - Mes y año de nacimiento
 - Los pesos muestrales para ponderar los datos
- Conjunto de datos generales
 - Mes y año de la entrevista

Para fines prácticos, es necesario considerar la siguiente notación:

- m_n^M : mes de nacimiento de la mujer/madre
- t_n^M : año de nacimiento de la mujer/madre
- x_e : edad de la mujer/madre en la entrevista
- m_n^N : mes de nacimiento del niño
- t_n^N : año de nacimiento del niño
- m_e : mes de la entrevista
- t_e : año de la entrevista

La variable *Edad en la entrevista* es un requisito extra al procedimiento que se sigue debido a las diferencias que se generaban en las tasas estimadas con la edad de la mujer o madre a un tiempo t y las obtenidas con las fechas de nacimiento, con respecto a las cifras oficiales. La justificación de esta decisión se encuentra detallada en el apéndice [A](#).

Como parte de las recomendaciones de [Moultrie \(2013\)](#) y por sugerencia de expertos ([Schoumaker 2013](#)), para las estimaciones de las tasas, los periodos de referencia que se desarrollarán en el capítulo son de enero al mes del levantamiento de la encuesta (sección [3.1](#)), anuales (años calendario anteriores al de la encuesta) (sección [3.2](#)), trianuales (tres años anteriores al de la encuesta en la sección [3.3](#)) y anuales para los catorce años anteriores al año del levantamiento de la encuesta (sección [3.4](#)).

3.1. Estimación para el periodo comprendido entre enero y el mes del levantamiento de la encuesta

Las tasas cuyo periodo de referencia es de enero al mes de la entrevista (periodo menor de un año) se obtienen de acuerdo al siguiente procedimiento:

1. Es necesario ubicar a los nacimientos ocurridos de enero al mes de la entrevista, considerando que éstos deben ser distribuidos conforme a la edad de la madre al nacimiento.

Para ello, primero se obtiene la edad de la madre al nacimiento del niño. Hay dos formas de obtenerla, una es únicamente con las fechas de nacimiento tanto del niño como de la madre¹, y la otra contempla, además, la edad de la madre en la entrevista y el año de la entrevista. La forma en cómo se realiza este cálculo en el paquete, es mediante la segunda opción², como se muestra a continuación:

$$x_{n\ n} = \begin{cases} x_{1\cdot 1\cdot e} - (t_e - t_n^N) + 1 & \text{si } m_n^N > m_n^M, \\ x_{1\cdot 1\cdot e} - (t_e - t_n^N) & \text{si } m_n^N < m_n^M, \\ x_{1\cdot 1\cdot e} - (t_e - t_n^N) + X & \text{e.o.c.} \end{cases} \quad (3.1)$$

donde $X \sim Ber(0.5)$ y $x_{1\cdot 1\cdot e}$ es la edad de la madre al 1º de enero del año de la entrevista, obtenida de la siguiente forma:

$$x_{1\cdot 1\cdot e} = \begin{cases} x_e & \text{si } m_n^M \geq m_e, \\ x_e - 1 & \text{e.o.c.} \end{cases} \quad (3.2)$$

El uso de la v.a. (variable aleatoria) Bernoulli en el caso que el niño y la madre tengan el mismo mes de nacimiento, ayuda a determinar aleatoriamente el día de nacimiento del niño, es decir, si el niño nació antes ($X = 0$) o después ($X = 1$) del día que cumple años su madre.

Con la edad de la madre al nacimiento de cada uno de los niños, se debe distinguir cuando el nacimiento ocurrió de enero al mes de la entrevista, considerando

¹Esta es la propuesta presentada por [Moultrie \(2013\)](#).

²La obtención de tasas con las otras dos fuentes de datos (censos de población y estadísticas vitales) se hace regularmente considerando la edad declarada, pero al hacerlo como en la primera propuesta, se generaban discrepancias en cuanto a la estructura por edad en las tasas.

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

que cada nacimiento representa cierto número de nacimientos de la población, conforme al peso muestral de la madre. Finalmente, se agrupan los nacimientos de acuerdo a la edad individual de la madre o los grupos quinquenales de edad, y se obtienen así los numeradores de las tasas de fecundidad específicas por edad individual o por grupos quinquenales de edad.

Por ejemplo, se tienen tres nacimientos ocurridos: uno el mes de noviembre del año 2017, el segundo el mes de marzo del año 2018 y el tercero el mes de agosto del año 2018, con pesos muestrales de 120, 90 y 64, respectivamente; y la edad de la madre al nacimiento es de 16 años en los tres casos. Considerando que la encuesta se levantó en el mes de septiembre del año 2018, los únicos nacimientos que cuentan para el numerador de las tasas referentes al periodo enero-septiembre son el segundo y el tercero. El valor del numerador para la tasa específica de fecundidad de edad 16 sería de $90 + 64 = 154$, es decir, 154 nacimientos ocurridos durante enero-septiembre del 2018 de madres de edad 16. Si la tasa fuera para el grupo de 15 a 19 años, faltaría incluir a los nacimientos de las mujeres de 15, 17, 18 y 19 años. En la figura 3.1 se puede apreciar el ejemplo anterior gráficamente.

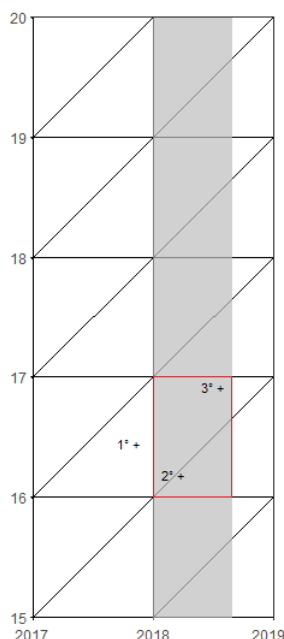


Figura 3.1: Diagrama de Lexis con el periodo de referencia enero-septiembre de 2018 para las edades de 15 a 19, considerando los nacimientos del ejemplo anterior.

2. El otro elemento para el cálculo de las tasas específicas son los denominadores, con los tiempos de exposición de las mujeres al riesgo de quedar embarazadas en el periodo de referencia antes mencionado. En los denominadores se consideran a

las todas las mujeres, ya sea que hayan sido madres o no.

Los denominadores se calculan en función de la edad de la mujer, y es importante mencionar que una mujer puede aportar tiempo de exposición en dos edades consecutivas, según la ubicación de su línea de vida en el periodo de referencia.

Antes del cálculo del tiempo de exposición, se requiere la edad de la mujer al 1º de enero del año de la entrevista ($x_{1.1.e}$), calculada como en la expresión 3.2. Una vez calculada esta edad $x_{1.1.e}$, se observa si la entrevista tuvo lugar antes o después del cumpleaños de la mujer en ese año. Entonces:

- Si $m_n^M > m_e$, la mujer estará expuesta al riesgo de quedar embarazada a la edad de $x_{1.1.e}$, por un tiempo de:

$$E(x_{1.1.e}, t_e) = \frac{m_n^M - 0.5}{12} \text{ años, y} \quad (3.3)$$

a la edad de $x_{1.1.e} + 1$, por un tiempo de:

$$E(x_{1.1.e} + 1, t_e) = \frac{m_e - m_n^M - 0.5}{12} \text{ años} \quad (3.4)$$

- En caso contrario, cuando $m_n^M < m_e$, la mujer estará expuesta al riesgo de quedar embarazada a la edad de $x_{1.1.e}$, por un tiempo de:

$$E(x_{1.1.e}, t_e) = \frac{m_e - 1}{12} \text{ años, y} \quad (3.5)$$

a la edad de $x_{1.1.e} + 1$, por un tiempo de:

$$E(x_{1.1.e} + 1, t_e) = 0 \text{ años} \quad (3.6)$$

Al igual que con los nacimientos, el tiempo de exposición medido en años que aporta cada mujer debe ser ponderado conforme a su peso en la muestra, para que, finalmente, se agrupen los tiempos de exposición por edad individual o grupos quinquenales de edad, y obtener los denominadores de las tasas de fecundidad específicas por edad individual o por grupos quinquenales de edad, respectivamente.

Un forma de ejemplificar cómo son considerados los tiempos de exposición en los denominadores para cada edad, es analizando a tres mujeres que cumplieron: la primera 15 años el mes de mayo del 2016, la segunda el mes de octubre del 2016

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

y la tercera el mes de enero del 2017, con pesos muestrales de 3186, 1812 y 972, respectivamente.

Se toma nuevamente septiembre como el mes de la entrevista y la edad de la mujer los 16 años para la obtención del denominador. Se tendría que la primera mujer que cumple años en mayo, aporta desde el 1^o de enero del año de la encuesta hasta antes de su cumpleaños 17, un tiempo de $\frac{5-0.5}{12} = 0.375$ años *para el denominador de edad 16* y desde su cumpleaños 17 hasta septiembre del 2018, un tiempo de $\frac{9-5-0.5}{12} = 0.2916$ *para el denominador de edad 17* (ecuaciones 3.3 y 3.4). La segunda mujer cumple años en octubre, aporta desde el 1^o de enero del año de la encuesta hasta antes de su cumpleaños 17 (en este caso, hasta la fecha de la entrevista), un tiempo de $\frac{9-1}{12} = 0.666$ años *para el denominador de edad 16* y un tiempo de 0 años *para el denominador de edad 17* (durante el periodo de referencia siempre tuvo 16 años), (ecuaciones 3.5 y 3.6). Por último, la mujer que cumple años en enero, aporta desde el 1^o de enero del año de la encuesta hasta antes de su cumpleaños 16, un tiempo de $\frac{1-0.5}{12} = 0.0416$ años *para el denominador de edad 15* y desde su cumpleaños 16 hasta septiembre del 2018, un tiempo de $\frac{9-1-0.5}{12} = 0.6255$ *para el denominador de edad 16*, (ecuaciones 3.3 y 3.4).

La figura 3.2 muestra las líneas de vida de las mujeres en el tiempo, el periodo del año antes de la entrevista es representado por la franja gris y con una línea roja, la división de mujeres con fecha de cumpleaños antes y después de la entrevista, respectivamente.

3.1 Estimación para el periodo comprendido entre enero y el mes del levantamiento de la encuesta

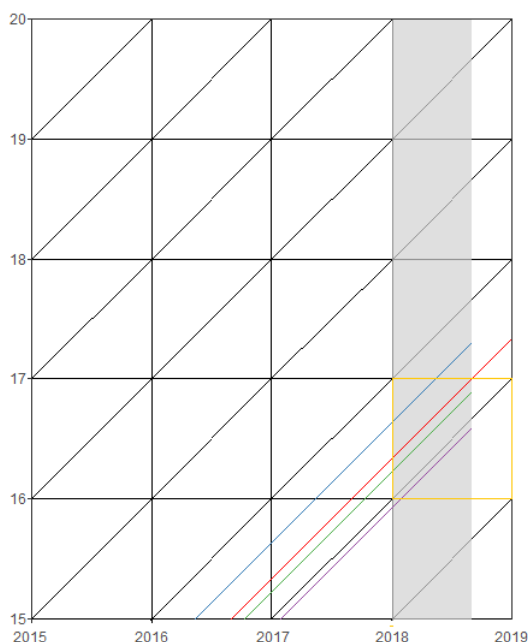


Figura 3.2: Diagrama de Lexis con el periodo de referencia enero-septiembre de 2018 para las edades de 15 a 19 y con las líneas de vida de mujeres.

Al aplicar los pesos muestrales a los tiempos de exposición de edad 16, tenemos $0.375 * 3186 + 0.666 * 1812 + 0.6255 * 972 = 3009.5$, es decir, 3009.5 años-persona de exposición al riesgo de embarazo durante enero-septiembre del 2018 de mujeres de edad 16.

- Al tener calculados los nacimientos conforme a la edad de la madre al nacimiento y la exposición al riesgo de todas las mujeres según su edad para el periodo de referencia, las tasas de fecundidad específicas por edad individual, para edad x en el año t_e están dadas por

$$f_x(t_e) = \frac{N(x, t_e)}{E(x, t_e)}, \text{ para } x = 15, 16, \dots, 49 \quad (3.7)$$

Si consideramos los ejemplos de los puntos 1. y 2., la tasa específica de fecundidad de edad 16 para el periodo enero-septiembre del 2018 sería:

$$\begin{aligned} f_{16}(2018) &= \frac{154}{3009.5} \\ &= 0.0511 \end{aligned}$$

Las tasas de fecundidad específicas por edad para grupos quinquenales, se obtie-

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

nen sumando los nacimientos de acuerdo a cada grupo de edad y dividiendo por la suma del tiempo de exposición de ese grupo de edad, es decir,

$${}_5f_{5*i}(t_e) = \frac{\sum_{j=5*i}^{5*i+4} N(j, t_e)}{\sum_{j=5*i}^{5*i+4} E(j, t_e)}, \text{ para } i = 3, 4, \dots, 9 \quad (3.8)$$

3.2. Estimación para periodos anuales anteriores al año del levantamiento de la encuesta

La estimación de tasas de fecundidad específicas por edad para periodos anuales anteriores, suelen calcularse para los cuatro años calendario anteriores al de la encuesta, es decir, $t_s < t_l < t_e$, $l = 1, 2, 3, 4$, ya que para esos años se tiene la información completa de las mujeres de edades 15 a 49 años, esto, al considerar que la encuesta se aplica a mujeres de 15 a 54 años. En la figura 3.3 se muestra la delimitación de la información con la que se cuenta, que es justamente el triángulo inferior delimitado por la línea de vida roja, y con amarillo se delimitan los cuatro años calendario anteriores al levantamiento para los que se cuenta con información completa.

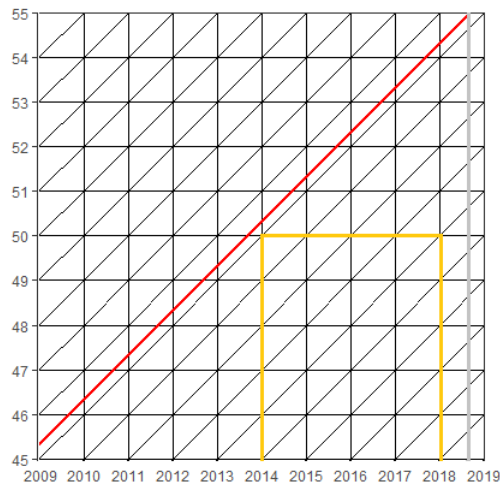


Figura 3.3: Diagrama de Lexis con la información de las historias de embarazos en la encuesta de 2018 para las edades de 45 a 54 años.

Es importante hacer mención que metodológicamente se puede regresar en el tiempo los años deseados, pero entre más distantes sean los años, se carecerá de más informa-

ción para estimar las tasas en las edades que se encuentren hacia el final de la vida reproductiva. Por cada año que se retroceda en el tiempo, habrá una edad más para la que falte la información.

Para obtener las tasas de fecundidad para periodos anuales es un poco más sencillo que para periodos menores a un año, debido a que se tiene el año calendario completo y los casos especiales se reducen. La metodología es la siguiente:

1. Al igual que en la sección anterior, se requieren los nacimientos ocurridos en el año calendario de estudio, distribuidos por edad de la madre al nacimiento.

Para obtener el numerador, se debe calcular la edad de la madre al nacimiento del niño con la expresión 3.1 antes presentada, para cada nacimiento que ocurrió en el año calendario de referencia (t_l , $l = 1, 2, 3, 4$). Cada nacimiento cuenta el número de veces conforme al peso muestral de la madre.

2. El tiempo de exposición al riesgo de embarazo que aporta una mujer en un periodo anual está dividido en dos edades consecutivas, de manera similar a la de periodos menores a un año, con la diferencia de que aquí no interesa el mes de la entrevista.

Para saber el tiempo de exposición que aporta cada mujer en las dos edades, es necesario calcular la edad de la mujer al 1^o de enero del año de estudio, con la expresión:

$$x_{1:l} = x_{1:e} - (t_e - t_l), \quad t_5 < t_l < t_e, \quad l = 1, 2, 3, 4 \quad (3.9)$$

donde t_l es el año de referencia y $x_{1:e}$ se obtiene como en 3.2.

La edad de la mujer en el año t_l es $x_{1:l}$ desde el inicio del año hasta antes de su cumpleaños y $x_{1:l} + 1$ desde su cumpleaños hasta el final del año, por lo que el tiempo de exposición que se aporta a cada edad está dado por:

$$E(x_{1:l}, t_l) = \frac{m_n^M - 0.5}{12} \text{ años, y} \quad (3.10)$$

$$E(x_{1:l} + 1, t_l) = 1 - \frac{m_n^M - 0.5}{12} \text{ años, } l = 1, 2, 3, 4 \quad (3.11)$$

Finalmente, es necesario ponderar cada valor obtenido con el peso muestral respectivo a cada mujer y sumar de acuerdo a las edades de las mujeres, es decir de 15 a 49 años.

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

3. De acuerdo a los puntos anteriores, las tasas de fecundidad específicas por edad y por grupos quinquenales de edad para periodos anuales anteriores al levantamiento de la encuesta están dadas por:

$$f_x(t_l) = \frac{N(x, t_l)}{E(x, t_l)}, \text{ para } x = 15, 16, \dots, 49, \quad l = 1, 2, 3, 4 \quad (3.12)$$

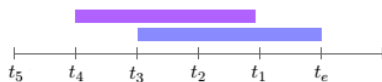
$${}_5f_{{}_5i}(t_l) = \frac{\sum_{j=5*i}^{5*i+4} N(j, t_l)}{\sum_{j=5*i}^{5*i+4} E(j, t_l)}, \text{ para } i = 3, 4, \dots, 9, \quad l = 1, 2, 3, 4 \quad (3.13)$$

3.3. Estimación para periodos trianuales que se encuentren dentro de los cuatro años anteriores a la encuesta

Una de las sugerencias de [Moultrie \(2013\)](#) para obtener estimaciones más confiables es respecto a los periodos de observación. El autor menciona que éstos deben ser mayores a un año calendario pero menores o iguales a tres años calendario, ya que de lo contrario se tendrían estimaciones erráticas o se aplanaría la tendencia de la fecundidad.

Las estimaciones para tres años calendario que se encuentren dentro de los cuatro años anteriores a la encuesta siguen la misma lógica de los cuatro años señalada con periodos anuales en la sección 3.2. El proceso para obtener las tasas se desarrolla a continuación:

1. Los numeradores para las tasas de periodos trianuales son los nacimientos ocurridos del año t_l al $t_l - 2$, donde $t_5 < t_l$ y $t_l - 2 < t_e$, $l = 3, 4$, es decir,



para los años calendario de t_4 a t_2 o de t_3 a t_1 , asociados y distribuidos por la edad de la madre al nacimiento, tomando en cuenta el peso muestral de la madre. La edad de la madre al nacimiento se calcula con base en la expresión 3.1.

2. Como el periodo de observación es más largo (tres años), el tiempo de exposición de la mujer al riesgo de quedar embarazada es mayor. Por ejemplo, en la figura

3.3 Estimación para periodos trianuales que se encuentren dentro de los cuatro años anteriores a la encuesta

3.4 se muestra el periodo de observación de 2015 a 2017 y la línea de vida de una mujer cuya fecha de nacimiento es el 9 de septiembre de 1998, esta mujer aporta tiempo de exposición al riesgo de embarazo para las edades 16, 17, 18 y 19, de manera fraccionaria para las edades 16 y 19 y de manera completa para las edades 17 y 18.

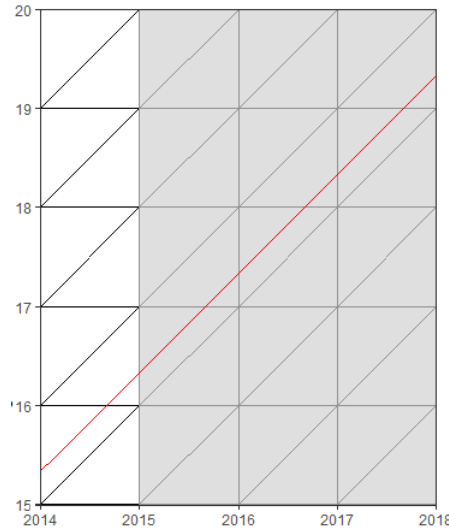


Figura 3.4: Diagrama de Lexis con la línea de vida de una mujer con fecha de nacimiento del 9 de septiembre de 1998 en el periodo de observación de 2015 a 2017.

Para calcular el tiempo de exposición es necesario contar con la edad de la mujer al inicio del primer año de referencia. Si es t_p el primer año de referencia, t_q el segundo y t_r el tercero, donde, $t_p < t_q < t_r$, la edad al 1º de enero del año t_p está dada por:

$$x_{1,t_p} = x_{1,t_e} - (t_e - t_p), \quad t_p < t_e < t_r \quad (3.14)$$

donde x_{1,t_e} es obtenida como 3.2.

Teniendo las edades al 1º de enero del año t_p (x_{1,t_p}), los tiempos de exposición están dados por:

$$E(x_{1,t_p}, t_p) = \frac{m_n^M - 0.5}{12} \text{ años}, \quad (3.15)$$

$$E(x_{1,t_p} + 1, t_p) = 1 - \frac{m_n^M - 0.5}{12} \text{ años}, \quad (3.16)$$

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

$$E(x_{1.1.p} + 1, t_q) = \frac{m_n^M - 0.5}{12} \text{ años,} \quad (3.17)$$

$$E(x_{1.1.p} + 2, t_q) = 1 - \frac{m_n^M - 0.5}{12} \text{ años,} \quad (3.18)$$

$$E(x_{1.1.p} + 2, t_r) = \frac{m_n^M - 0.5}{12} \text{ años y} \quad (3.19)$$

$$E(x_{1.1.p} + 3, t_r) = 1 - \frac{m_n^M - 0.5}{12} \text{ años} \quad (3.20)$$

Sin embargo, es claro que las edades de las ecuaciones 3.16 y 3.17, y 3.18 y 3.19 suman la unidad, por lo que al agruparlos para obtener los denominadores, los tiempos de exposición para los tres años estarían dados por:

$$E(\cdot, t_p, t_q, t_r) = \begin{cases} \frac{m_n^M - 0.5}{12} & \text{para } x_{1.1.p}, \\ 1 & \text{para } x_{1.1.p} + 1 \text{ o } x_{1.1.p} + 2, \\ 1 - \frac{m_n^M - 0.5}{12} & \text{para } x_{1.1.p} + 3 \end{cases} \quad (3.21)$$

3. Por último, las tasas de fecundidad específicas por edad individual y por grupos quinquenales de edad para periodos trianuales anteriores al levantamiento de la encuesta están dadas por:

$$f_x(t_p, t_q, t_r) = \frac{N(x, t_p, t_q, t_r)}{E(x, t_p, t_q, t_r)}, \text{ para } x = 15, 16, \dots, 49, t_5 < t_p < t_q < t_r < t_e \quad (3.22)$$

$${}_5f_5^i(t_p, t_q, t_r) = \frac{\sum_{j=5^*i}^{5^*i+4} N(j, t_p, t_q, t_r)}{\sum_{j=5^*i}^{5^*i+4} E(j, t_p, t_q, t_r)}, \text{ para } i = 3, 4, \dots, 9, t_5 < t_p < t_q < t_r < t_e \quad (3.23)$$

3.4. Estimación para periodos anuales de los catorce años anteriores al año de la encuesta

La metodología de cálculo para tasas específicas anuales en esta sección es similar a la que se presenta en la sección 3.2. Sin embargo, la diferencia entre ambas secciones es la estimación de las tasas de las últimas edades, ya que como se explicó con el diagrama de Lexis 3.3, entre más distantes sean las estimaciones con respecto al año de la entrevista, son menos las tasas que se pueden obtener con la información de la encuesta. Es necesario mencionar que, las tasas de las últimas edades tiene valores bajos generalmente y éstos afectan poco el valor de la TGF.

En esta sección, el periodo de retrospección para las estimaciones es de 14 años. Si se considera la aplicación de la encuesta a mujeres de 15 a 54 años y la obtención de tasas para edades de 15 a 49 años, se tendría información hasta los 39 años de edad para el año más distante, y faltaría información para calcular 55 tasas por edad individual o 15 tasas por grupos quinquenales de edad pertenecientes a las últimas edades. El siguiente diagrama de Lexis 3.5 ejemplifica la distribución de dicha información para las edades a partir de los 38 años.

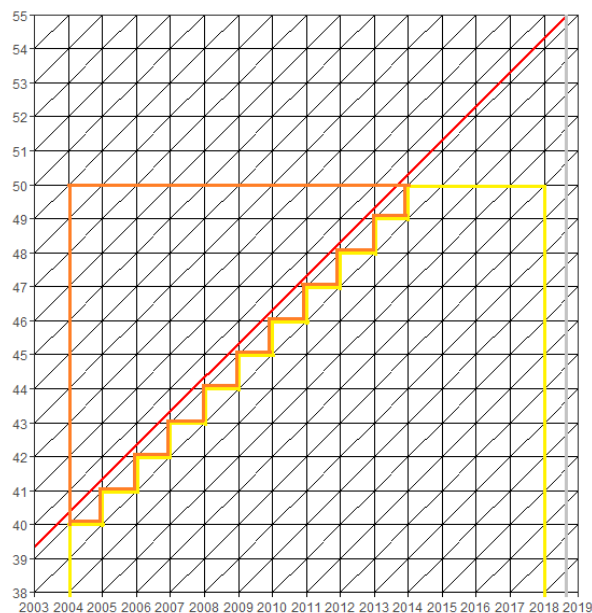


Figura 3.5: Diagrama de Lexis de la información para las edades de 38 a 49 años.

Es necesario mencionar que el triángulo inferior de la diagonal roja marca toda la información que se tiene. El área delimitada de amarillo muestra los datos que son

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

ocupados para la obtención de tasas de esta sección, y el área delimitada de naranja muestra los datos que se necesitarían para tener las estimaciones completas¹.

Dicho lo anterior, el enfoque principal de esta sección es la metodología para obtener las tasas que no pueden ser obtenidas de manera directa (área naranja). Una propuesta de estimación es mediante la simulación de las tasas por grupos quinquenales de edad, y con estas tasas completas, obtener las tasas por edad individual faltantes a través de desagregación.

Como primer paso se requiere tener las tasas por edad y por grupos quinquenales de edad que sí se pueden calcular mediante la metodología de la sección 3.2 para cada uno de los 14 años anteriores a la encuesta. Es importante decir que si existe información faltante en alguna tasa agrupada, la tasa debe ser estimada, tal y como se muestra en el siguiente diagrama 3.6.

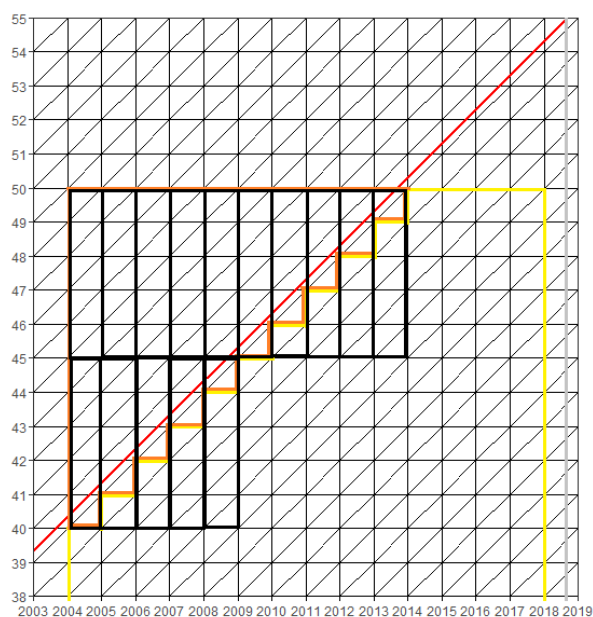


Figura 3.6: Diagrama de Lexis que muestra la información existente y faltante de las últimas edades reproductivas para los 15 años anteriores al levantamiento de la encuesta.

Debido a la poca información, es difícil emplear un modelo matemático sofisticado para resolver el problema de los datos faltantes. No obstante, se propone simular las tasas a manera de reproducir artificialmente los valores, con la experiencia de los propios grupos de edad en años más recientes.

Para obtener las 5 primeras tasas pertenecientes al grupo de 40 a 44 años, primero se calcula la media muestral de las 9 tasas existentes del mismo grupo, es decir:

¹Estimaciones que van de los 15 a los 49 años para los 14 años anteriores.

$$\hat{\mu}_a = \frac{\sum_{i=6}^{14} {}_5f_{40}(t_i)}{9} \quad \text{donde } t_i \text{ es el } i\text{-ésimo año de los catorce.}$$

Después, se obtiene la varianza muestral de las mismas, es decir:

$$\hat{\sigma}_a^2 = \frac{\sum_{i=6}^{14} ({}_5f_{40}(t_i) - \hat{\mu}_a)^2}{8}$$

Finalmente, se simulan 5 observaciones de una variable aleatoria $X \sim N(\hat{\mu}_a, \hat{\sigma}_a^2)$ y se toma el valor absoluto de éstas para cada ${}_5f_{40}(t_i)$ faltante, $i = 1, 2, \dots, 5$. En el programa R, la simulación se hace con la función `rnorm()` con sus respectivos estimadores de μ y σ^2 .

En el caso de las tasas pertenecientes al grupo 45 a 49 años, de manera análoga pero con menos información, se calcula la media muestral con la siguiente fórmula:

$$\hat{\mu}_b = \frac{\sum_{i=11}^{14} {}_5f_{45}(t_i)}{4}$$

también se calcula la varianza muestral con la fórmula:

$$\hat{\sigma}_b^2 = \frac{\sum_{i=11}^{14} ({}_5f_{45}(t_i) - \hat{\mu}_b)^2}{3}$$

En este caso, se simulan 10 observaciones de una variable aleatoria $X \sim N(\hat{\mu}_b, \hat{\sigma}_b^2)$ y se considera su valor absoluto para cada ${}_5f_{45}(t_i)$ faltante, $i = 1, 2, \dots, 10$.

Con las tasas específicas de grupos de edades quinquenales completas, se obtienen las tasas específicas por edad faltantes mediante la desagregación de la estructura de la fecundidad que propone y desarrolla [Chackiel \(1979\)](#) mediante del doble logaritmo de la función Gompertz.

La exposición del método es de manera general para un año, pero en la aplicación de este trabajo se emplea para los 10 años más distantes de la encuesta, que son los que requieren de la estimación de al menos una tasa por edad.

Como base de este método, se emplea la función Gompertz que representa la fecundidad acumulada, la expresión es la siguiente:

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

$$F(x) = (TGF) * a^{b^x} \quad (3.24)$$

donde:

- x es la edad, con $x = 15, 16, \dots, 50$.
- $F(x)$ es la fecundidad acumulada hasta la edad x , y mediante sumas de tasas quinquenales la fórmula es:

$$F(x) = 5 * \sum_{i=3}^{\frac{x}{5}-1} {}_5f_{5*i}, \quad x = 20, 25, \dots, 50$$

con $F(50) = TGF$.

- TGF^1 es la tasa global de fecundidad del año.
- a y b son parámetros que están en el intervalo $[0, 1]$.

Ahora bien, faltaría determinar a y b . No obstante, una manera de conocer un estimador de $F(x)$ sin contar directamente con a y b es linealizando la función 3.24.

Entonces linealizando 3.24 tenemos que:

$$\frac{F(x)}{TGF} = a^{b^x}$$

aplicando \ln :

$$\ln\left(\frac{F(x)}{TGF}\right) = b^x * \ln(a)$$

multiplicamos por -1 y aplicamos \ln :

$$\ln\left(-\ln\left(\frac{F(x)}{TGF}\right)\right) = x * \ln(b) + \ln(-\ln(a))$$

si hacemos:

- $V(x) = \ln\left(-\ln\left(\frac{F(x)}{TGF}\right)\right)$
- $m = \ln(b)$
- $c = \ln(-\ln(a))$

¹La TGF para cada caso se obtiene con la expresión 1.3.

tendríamos la ecuación de una recta de la forma $V(x) = m * x + c$.

Además, a través de la fecundidad por grupos quinquenales podemos tener $F(x)$, la TGF del año en cuestión y $V(x)$ para ciertas edades como se muestra en la siguiente tabla:

Valores conocidos		Valores obtenidos	
x	$F(x)$	$\frac{F(x)}{TGF}$	$V(x) = \ln \left(-\ln \left(\frac{F(x)}{TGF} \right) \right)$
15	$F(15) := 0$	0	-
20	$F(20)$	✓	$V(20)$
25	$F(25)$	✓	$V(25)$
30	$F(30)$	✓	$V(30)$
35	$F(35)$	✓	$V(35)$
40	$F(40)$	✓	$V(40)$
45	$F(45)$	✓	$V(45)$
50	$F(50) = TGF$	1	-

Con los 6 valores obtenidos, ajustamos 2 puntos por el método de promedios móviles, es decir:

$$A : \left(\frac{20 + 25 + 30}{3}, \frac{V(20) + V(25) + V(30)}{3} \right)$$

$$B : \left(\frac{35 + 40 + 45}{3}, \frac{V(35) + V(40) + V(45)}{3} \right)$$

Al determinar la recta que pasa por A y B , obtendremos el estimador de $V(x)$ y con ello $\hat{F}(x)$. Para esto, podemos hacer uso de la ecuación en la forma:

$$V(x) - a_2 = \left(\frac{a_2 - b_2}{a_1 - b_1} \right) (x - a_1)$$

Entonces, aplicando la ecuación anterior a los puntos dados y despejando, se tiene:

$$\hat{V}(x) = \frac{V(35)+V(40)+V(45)-(V(20)+V(25)+V(30))}{15} (x - 25) + \frac{V(20) + V(25) + V(30)}{3}$$

3. ESTIMACIÓN DE LA FECUNDIDAD A PARTIR DE LA INFORMACIÓN DE LA HISTORIA DE EMBARAZOS DE LAS ENCUESTAS

Una vez obtenida la ecuación $\hat{V}(x)$ y con el estimador de $F(x)$ podemos evaluar para $x = 16, 17, \dots, 49$ y finalmente obtener la fecundidad por edades individuales f_x .

Los estimadores finales para eso son:

$$\hat{F}(x) = TGF * \exp\left(-\exp\left(\hat{V}(x)\right)\right)$$

$$\hat{f}_x = \hat{F}(x + 1) - \hat{F}(x)$$

Se han presentado los estimadores y la metodología que cubre nuestro objetivo en temas demográficos y de muestreo. A continuación se exponen los principales elementos para la adecuada construcción y desarrollo de un paquete en R.

Capítulo 4

Desarrollo de un paquete en R

En la actualidad el lenguaje de programación R se ha convertido en un gran recurso estadístico, debido a que es un software libre en constante desarrollo gracias al aporte de la comunidad científica.

En este capítulo se describe brevemente cómo se hace uso de las funciones contenidas en un paquete y la manera de guardar códigos (4.1). En la sección siguiente (sección 4.2), se muestran algunas convenciones de la programación en R. Además, se explica y ejemplifica cada uno de los tres sistemas de orientación de objetos que maneja R entorno al desarrollo de paquetes en R (4.3). Finalmente, en el anexo [Programación en R](#) se describen algunos elementos necesarios de la programación en R y que son útiles para el desarrollo de paquetes.

4.1. Uso de paquetes y la escritura de programas

Los paquetes en R son herramientas que permiten al usuario obtener de forma rápida y gratuita colecciones de funciones y códigos compartidos por otros usuarios. El principal repositorio de paquetes estables es CRAN. Para descargar un paquete es necesario contar con internet y ejecutar la función `install.packages()` con el nombre del paquete entre comillas. Una vez descargado, cada vez que se deseen ocupar sus funciones y consultar su documentación, es necesario ejecutar la función `library()` con el nombre del paquete como argumento.

Un ejemplo de esto es el paquete `foreign` que tiene funciones para leer y escribir datos con diferentes formatos.

```
> install.packages("foreign")
```

```
Installing package into 'C:/Users/Yaz/Documents/R/win-library/3.4'  
(as 'lib' is unspecified)  
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.4/foreign_0.8-70.zip'  
Content type 'application/zip' length 324176 bytes (316 KB)  
downloaded 316 KB
```

4. DESARROLLO DE UN PAQUETE EN R

```
package 'foreign' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\Yaz\AppData\Local\Temp\Rtmpkv9PiY\downloaded_packages
```

```
> library(foreign)
> fec1<-read.dbf('E97CMU.DBF')
> head(fec1,5)
```

```
  ENT MUN ZONA ESTRATO TAM_LOC   UPM F_VIV HOGAR ...
1  01 006  30     1         4 000001  01    1
2  01 006  30     1         4 000001  01    1
3  01 006  30     1         4 000001  02    1
4  01 006  30     1         4 000001  03    1
5  01 006  30     1         4 000001  04    1 ...
```

Existen paquetes que necesitan de otros para poder funcionar, por lo que es importante primero consultar si es el caso.

Hasta el momento las instrucciones han sido mostradas como se ejecutan en consola, sin embargo, ésta es una práctica no muy recomendable, ya que si se requieren ejecutar en más de una ocasión, es necesario escribirlas nuevamente. Para esta cuestión, se tienen los *script*, que son archivos en donde se guardan los programas y pueden ser ejecutados en cualquier momento. Los nombres de los archivos deben ser significativos y la extensión de estos es **.R* (Wickham 2015).

```
fit_models.R      adecuado
utility_functions.R  adecuado
foo.r             inadecuado
stuff.r           inadecuado
```

4.2. Estilo de programación

De forma general se han mostrado los elementos necesarios para programar en R. Sin embargo, es importante que los programas sean escritos bajo ciertos criterios convencionales o de consistencia, y que permitan a otro desarrollador o lector comprender la estructura y el funcionamiento del programa. Siguiendo el estilo de Wickham (2014) que se basa en [la guía de estilo R de Google](#)¹, se exponen algunas reglas de programación en R para hacer más fácil de leer, compartir y verificar el código.

Nombres de los objetos

Recomendaciones:

¹<https://google.github.io/styleguide/Rguide.xml>

- los nombres de las variables y las funciones deberán estar en minúsculas;
- separar palabras dentro de un nombre con un guión bajo (el punto está reservado para los métodos S3);
- las variables deben ser nombradas con sustantivos;
- las funciones deben ser nombradas con verbos;
- evitar nombres de variables y funciones existentes o que causen confusión.

<code>day_one</code>	adecuado
<code>day_1</code>	adecuado
<code>first_day_of_the_month</code>	inadecuado
<code>DayOne</code>	inadecuado
<code>dayone</code>	inadecuado
<code>djm1</code>	inadecuado
<code>T <- FALSE</code>	inadecuado
<code>c <- 10</code>	inadecuado
<code>mean <- function(x) sum(x)</code>	inadecuado

Espaciado

Recomendaciones:

- poner espacio alrededor de operadores (`=`; `+`; `-`; `/`; `*`; entre otros) y cuando se use `=` para funciones llamadas;
- poner espacio después de la coma, nunca antes;
- no poner espacio después del nombre de una función;
- no poner espacios extra;
- no colocar espacios alrededor del código entre paréntesis o corchetes, a menos que haya una coma.

<code>average <- mean(feet / 12 + inches, na.rm = TRUE)</code>	adecuado
<code>x <- 1:10</code>	adecuado
<code>plot(x, y)</code>	adecuado
<code>if (debug) do(x)</code>	adecuado
<code>diamonds[5,]</code>	adecuado
<code>average<-mean(feet/12+inches,na.rm=TRUE)</code>	inadecuado
<code>x <- 1 : 10</code>	inadecuado
<code>plot (x, y)</code>	inadecuado
<code>x[1,]</code>	inadecuado
<code>x[1 ,]</code>	inadecuado

Llaves

Recomendaciones:

- una llave de apertura nunca debe ir en una línea propia;
- una llave de cierre siempre debe ir en su propia línea, a menos que esté seguida por un else;
- el código lleva sangría dentro de las llaves;
- está bien dejar declaraciones muy cortas en la misma línea (sin uso de llaves).

```
if (y < 0 && debug) {           adecuado
message("Y is negative")
}
if (y == 0) {                  adecuado
log(x)
} else {
y ^ x
}
if (y < 0 && debug) message("Y is negative")  adecuado
if (y < 0 && debug)             inadecuado
message("Y is negative")
if (y == 0) {                  inadecuado
log(x)
}
else {
y ^ x
}
```

Longitud de línea

La recomendación para el largo máximo de una línea de código es de 80 caracteres, ya que así se adapta a una página impresa con una fuente de tamaño razonable.

4.3. R y su orientación a objetos

Como ya se mencionó en el subcapítulo anterior, el principal paradigma de desarrollo de software en R es la orientación a objetos, por lo que los programas son definidos en términos de clases, objetos y métodos. Las *clases* definen las características y el comportamiento de los *objetos*, y describen su estado, que es la especificación de los tipos de datos representados; las clases definen y seleccionan los *métodos*, que son conjuntos de instrucciones/algoritmos.

R cuenta con tres sistemas orientados a objetos S3, S4 y RC, los cuales son diferenciados por cómo se definen las clases y los métodos. De acuerdo con [Wickham \(2014\)](#), el contraste es que, S3 no tiene definición formal de clases e implementa una programación orientada a objetos, llamada función genérica. En ésta, los cálculos se realizan

mediante métodos, para que después, la función genérica decida qué método llamar. Mientras que S4 tiene definición formal de clases y funciones especiales para definir métodos. Las *Reference Classes*, llamadas RC por sus siglas en inglés, son diferentes a S3 y S4, y muy similares a la orientación de objetos de C++, Java o Python, debido a que implementa la orientación a objetos de paso de mensajes, por lo que los métodos pertenecen a las clases, no a las funciones.

Para ejemplificar S3, S4 y RC, a continuación se expondrán ejemplos simples de la implementación.

4.3.1. S3

Clases y objetos

Como ya se mencionó, S3 no tiene una definición formal de clases, por lo que basta con tener el objeto y hacerlo una instancia de clase con la función `class()` o bien, `attr()`, por asignación directa o con una función.

Los nombres de las clases deben ser en minúsculas, y debe evitarse el uso de punto.

Código 4.1: Creación y definición de clase *cuenta*

```
1 account <- list(titular = "Helena", saldo = 98.5)
2 class(account) <- "cuenta" # se establece la clase cuenta (directo)
```

Código 4.2: Función constructora de objetos de clase *cuenta*

```
1 cuenta <- function(a, b) {
2   if(!is.character(a)) stop("Nombre no valido")
3   if(!is.numeric(b)) stop("Saldo no valido")
4   if(b < 0) stop("Saldo no valido")
5
6   v <- list(titular = a, saldo = b)
7   attr(v, "class") <- "cuenta" # se establece la clase cuenta
8   v
9 }
```

Para tener uniformidad en los objetos, es común que el nombre de la función constructora sea el mismo que el de la clase.

Métodos y funciones genéricas

Una vez teniendo el objeto con la clase a la que hará referencia, es importante llamar al método adecuado para éste. S3 reconoce los métodos por su nombre, ya que son compuestos del nombre de la función genérica, un punto y el nombre de la clase (`nombre_generico.nombre_clase()`).

4. DESARROLLO DE UN PAQUETE EN R

Por ejemplo, la función `mean()` obtiene el promedio de los datos que recibe, si estos fueran de la clase `Date`, el método que se estaría reconociendo es el `mean.Date()`, en caso de no existir el método específico para `Date`, se tomaría un método alternativo llamado `mean.default()`. Es justo por esta razón que no se recomienda que los nombres de las clases lleven punto, para evitar la confusión con un método.

Código 4.3: Método para imprimir el objeto *cuenta*

```
1 print.cuenta <- function(x,...){
2   print("El titular de la cuenta es: \n")
3   print(x$titular)
4   cat("y tiene con un saldo de $", x$saldo)
5 }
```

Para S3, los métodos pertenecen a las funciones genéricas (en el ejemplo 4.3 pertenece a la función `print()`) y no al objeto o clase. Por lo que si se desea una función genérica propia, es necesario crearla.

Código 4.4: Creación de la función genérica `deposito()`

```
1 deposito <- function(x) {
2   UseMethod("deposito")
3 }
```

Por sí sola, la función del código 4.4 es incapaz de realizar alguna acción, por lo que es necesario la implementación de métodos.

Código 4.5: Método predeterminado

```
1 deposito.default <- function(x, ...) {
2   print("Funcion para hacer depositos")
3 }
```

Código 4.6: Método para los objetos de la clase *cuenta*

```
1 deposito.cuenta <- function(x, cantidad) {
2   if(!is.numeric(cantidad)) stop("Cantidad no valida")
3   if(cantidad >= 0){
4     x$saldo = x$saldo + cantidad
5     cat(x$titular, ":\n")
6     cat("Tu saldo ahora es de $", x$saldo)
7   } else{
8     stop("Esa cantidad no puede ser depositada")
9   }
10 }
```

4.3.2. S4

Clases

Para definir una clase en S4 es necesario hacer uso de la función `setClass()`, y establecer el nombre y las variables (slots) que tendrá. Existen argumentos opcionales que pueden ser consultados en la ayuda de R.

Código 4.7: Creación y definición de la clase *cuenta*

```
1 setClass("cuenta",  
2   slots = list(titular = "character", saldo = "numeric"))
```

Objetos

Los objetos pueden ser creados con la función `new()` o con una función generadora.

Código 4.8: Creación de objetos de clase *cuenta*

```
1 c9 <- new("cuenta", titular = "Helena", saldo = 98.5)
```

Código 4.9: Función generadora de objetos de clase *cuenta*

```
1 cuenta <- setClass("cuenta",  
2   slots = list(titular = "character", saldo = "numeric"))  
3  
4 cuenta("Helena", 98.5)
```

Para tener acceso a los elementos de un objeto, es necesario hacer uso del `@` o bien de la función `slot()` (`@` es equivalente a `$`, y `slot()` a `[]`).

Métodos y funciones genéricas

De manera similar a S3, los métodos implementados para clases de S4 también pertenecen a funciones genéricas más que a la clase. Para implementar los métodos asociados a una función genérica se usa la función `setMethod()`. Si es necesario escribir funciones genéricas propias, se hace uso de la función `setGeneric()`.

Código 4.10: Creación de la función genérica `retiro()`

```
1 setGeneric("retiro", function(x) standardGeneric("retiro"))
```

`standardGeneric()` en S4 es equivalente a `UseMethod()` en S3.

4. DESARROLLO DE UN PAQUETE EN R

Código 4.11: Método para los objetos de la clase *cuenta*

```
1 setMethod("retiro", "cuenta", function(x, cantidad) {
2   if(!is.numeric(cantidad)) stop("Cantidad no valida")
3   if(cantidad <= saldo){
4     x$saldo = x$saldo - cantidad
5     cat(x$titular, ":\n")
6     cat("Tu saldo ahora es de $", x$saldo)
7   } else{
8     stop("Saldo insuficiente")
9   }
10 }
11 )
```

4.3.3. Reference Classes

En comparación con S3 y S4, al programar bajo RC se tiene como ventaja un ahorro importante de memoria, debido a que los objetos de la clase son mutables, es decir, se modifica el objeto y no una copia de éste, teniendo con ello mayor flexibilidad a cambios.

Clases y objetos

La definición de clases es parecida a S4. La función `setRefClass()` define la clase mediante el nombre y las variables (fields).

Código 4.12: Creación y definición de la clase *cuenta*

```
1 setRefClass("cuenta",
2   fields = list(titular = "character", saldo = "numeric"))
```

De la misma manera que S4, la creación de objetos es mediante la función generadora de la clase.

Código 4.13: Función generadora de objetos de clase *cuenta*

```
1 cuenta <- setRefClass("cuenta",
2   fields = list(titular = "character", saldo = "numeric"))
3 cuenta("Eduardo", 1000)
```

Los campos de un objeto pueden ser mostrados o modificados usando el operador `$`.

Métodos

En Reference Classes, los métodos son de la clase de referencia y no de las funciones genéricas. Además, todas las clases de referencia tienen métodos heredados de la superclase `envRefClass`. Algunos de los métodos son: `copy`, `export`, `field`, `getClass`, `getRefClass`, `import`, `initFields`, `show`, `usingMethods`.

Los métodos propios de la clase de referencia se crean junto con la definición de la clase.

```

1 cuenta <- setRefClass("cuenta",
2   fields = list(titular = "character", saldo = "numeric"),
3   methods = list(
4     deposito = function(x) {
5       saldo <<- saldo + x # <<- operador de asignacion no local
6     },
7     retiro = function(x) {
8       saldo <<- saldo - x
9     }
10  )
11 )
12
13
14 cuenta$deposito(4900)
15 cuenta$retiro(356) # la implementacion de estos metodos se hace uso del operado $

```

4.3.4. Organización de un paquete R

Un paquete R ayuda a agrupar el código, los datos y la documentación, entre otros, y así compartir las nuevas extensiones o implementaciones de manera más eficiente. Para ello, es necesario seguir la estructura adecuada y tener buenas prácticas. Una guía que describe el proceso de creación de paquetes es [Writing R Extensions](#)¹ o bien, existen paquetes como *devtools* y *roxygen2*, en conjunto con RStudio² y Rtools³, que automatizan las tareas de desarrollo comunes con buenas prácticas.

En la actualidad⁴, existen más de 12,000 paquetes puestos a disposición en el repositorio CRAN, con el constante aporte de la comunidad científica, que dota a R de múltiples capacidades. Por lo que es importante que antes de implementar procedimientos, se investigue si alguien más no los ha llevado a cabo, con la finalidad de no duplicar funcionalidades.

A continuación se muestran los componentes esenciales de un paquete.



fertilityRates

Directorio que contiene el paquete. En este caso el paquete es `fertilityRates`.



DESCRIPTION

Fichero que contiene información básica, como la descripción de lo que hace el paquete, lo que necesita para funcionar, contacto para dudas, entre otros.

¹<https://cran.r-project.org/doc/manuals/r-release/R-exts.html>

²RStudio es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R.

³Rtools es una colección de herramientas necesarias para la creación de paquetes en R y es un programa aparte.

⁴Consultado el 31 de mayo de 2018 en <https://cran.r-project.org/web/packages/>

4. DESARROLLO DE UN PAQUETE EN R



Fichero que indica las funciones exportadas y reconocidas por el paquete.



Subdirectorio que contiene los archivos del código R. Esta carpeta contiene lo más importante de un paquete.



Subdirectorio que contiene los ficheros de datos en formato *RData* que son utilizados por los ejemplos de la implementación.



Subdirectorio que contiene los ficheros de documentación necesaria para el funcionamiento de las funciones del paquete.



`fertilityRates.Rproj`, archivo proyecto RStudio, cuya función es facilitar el uso del paquete en ese IDE. Este archivo es sólo auxiliar.

Existen otros componentes que pueden ser incluidos en el paquete, como son vignettes, que son archivos que contienen documentación más completa, o test, que son archivos de prueba para un adecuado comportamiento, entre otros.

Es importante decir que, el nombre de un paquete únicamente puede tener letras, números y puntos, siempre debe comenzar con letra y nunca terminar en punto. (Wickham 2015).

Se sugiere el uso de *devtools* y *roxygen2*, que son herramientas para que el desarrollo de un paquete sea de manera más rápida y fácil, ya que crean de forma automática los ficheros `DESCRIPTION` y `NAMESPACE`, los directorios `R` y `man`, así como la documentación.

Con los elementos principales presentes para la construcción del paquete *fertilityRates*, en el siguiente capítulo se expone el uso y la aplicación de éste.

Capítulo 5

Uso del paquete *fertilityRates*

Debido a la extensión del código fuente del paquete, resulta poco práctico presentarlo en este capítulo, pero se incluye de manera íntegra en el apéndice C. En este capítulo se indica cómo obtener el código e instalarlo en R (sección 5.1), la preparación de los datos para generar las tasas (sección 5.2), y finalmente, cómo hacer uso del paquete *fertilityRates* (sección 5.3).

5.1. Instalación del paquete

Como se mencionó en la sección 4.1, para hacer uso de un paquete en R, es necesario que éste sea descargado e instalado desde un repositorio en línea. En el caso del paquete *fertilityRates* el repositorio donde se puede consultar y descargar es GitHub.

Existen dos formas de hacer uso del código. Una es descargando el paquete en una carpeta ZIP desde el repositorio <https://github.com/YazGonzalez/fertilityRates.git>, como se muestra en la figura 5.1, o bien, al tener previamente instalado y cargado el paquete *devtools*, se ejecuta `devtools::install_github("YazGonzalez/fertilityRates")`.

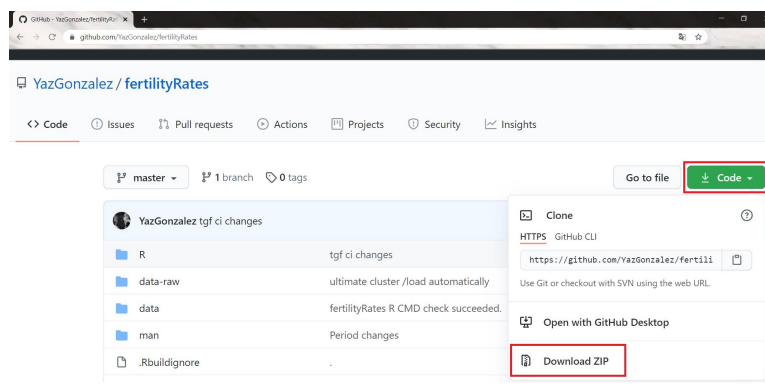


Figura 5.1: Página de GitHub donde está el repositorio *fertilityRates*; los cuadros rojos señalan los pasos para descargar la carpeta.

5.2. Estructura de los datos

Anteriormente se mencionó que este paquete está enfocado a los datos proporcionados en las historias de embarazos de encuestas por muestreo. Generalmente en las encuestas se organiza el conjunto de datos en secciones (entidades), de acuerdo a un modelo entidad-relación, generando diversos archivos de información que tienen un tipo de relación entre sí. En el caso de las encuestas ENADID, la distribución de la información suele estructurarse conforme a los datos de la vivienda, el hogar y la persona, para cada uno de los fenómenos demográficos.

Para el propósito del paquete, es de interés un conjunto de datos en un formato largo, que contenga ciertas *características de las mujeres de 15 a 54 años*, su respectiva *historia de embarazos*, el mes de la entrevista y las variables del diseño muestral. Para esto, es importante conocer las secciones o el nombre de los archivos que contienen la recopilación y la organización de la información y cómo están relacionados entre sí para, mediante operaciones de datos, obtener el conjunto requerido.

Como se mencionó, es necesario identificar la sección que contiene la información de interés. En el caso de las encuestas ENADID, ésta se encuentra en los archivos de la tabla 5.1.

ENADID						
1992	1997	2006	2009	2014	2018	
FECUNDIDAD_1	E97CMU	tbl_mujer	tr_cmu	TMMujer1	TMUJER1	Caract. de la mujer
FECUNDIDAD_2	E97HEM	tbl_fecundidad	tr_fec_hemb	TFec_Hemb	TFECHISEMB	Historia de embarazos
	E97VHO	tbl_hogares				Mes de entrevista

Tabla 5.1: Nombre de los archivos de cada ENADID 2018.

Estas secciones de la información están relacionadas mediante una variable o la agrupación de varias variables, según la encuesta de que se trate. La siguiente figura 5.2 muestra la relación de uno a muchos que tienen TMUJER1 y TFECHISEMB, es decir, un registro de la sección características de la mujer se puede asociar a ninguno, a uno o a varios registros de la sección historia de embarazos. En este caso, la relación es posible por la llave primaria (PK) y la llave foránea (FK), que es la variable LLAVE_MUJ que ambos archivos tienen, y que para TMUJER1 identifica de forma única a cada fila.

Además de esto, la figura 5.2 muestra las variables que son de interés para la obtención de los principales índices de la fecundidad. Dentro de las historias de embarazos se captan los hijos nacidos vivos, mortinatos y abortos, por lo que es importante excluir los dos últimos grupos de nuestros datos, ya que los índices de fecundidad solo involucran información de los hijos nacidos vivos.

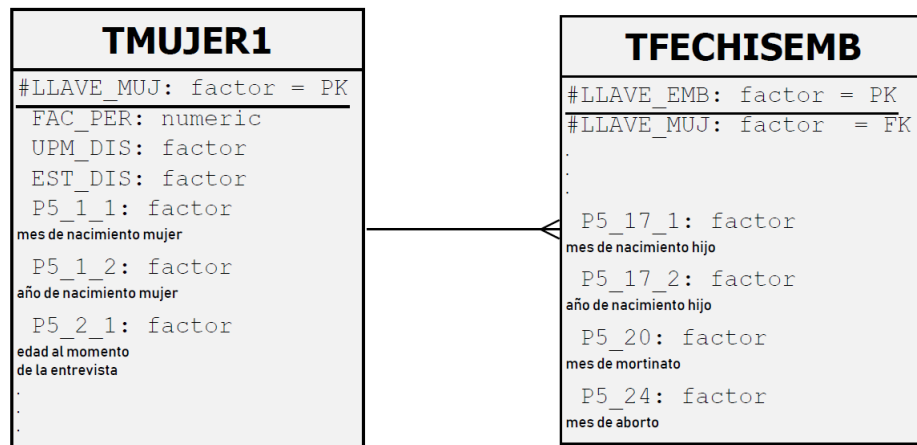


Figura 5.2: Diagrama UML de las secciones de la base de datos ENADID 2018.

Es necesario hacer mención que el tipo de dato *factor* no es el más adecuado para nuestro propósito, ya que impide realizar operaciones aritméticas, lógicas y comparativas importantes en la información, por lo que es necesario convertirlo a numérico. Asimismo, se debe tomar en cuenta que cuando se carece de información (missing) en la variable año de nacimiento del hijo, el nacimiento será excluido; en cambio, cuando falta el mes de nacimiento de la mujer o del hijo, estos faltantes se pueden proratear, ya que generalmente, los meses de nacimiento se distribuyen de manera uniforme.

En figura 5.3 se muestra el cambio en los archivos TMUJER1 y TFECHISEMB (ahora MUJER_18 e HISTORIA_18, respectivamente), con la eliminación de variables y registros, y la conversión de tipo de dato de variables, conforme a lo antes mencionado.

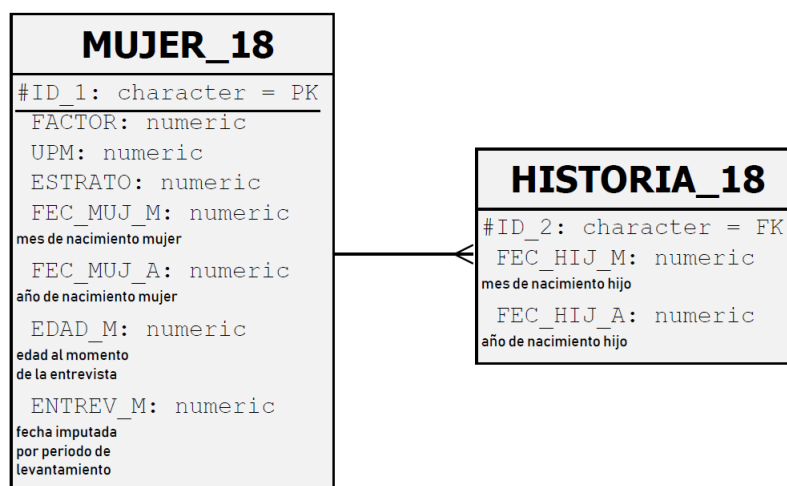


Figura 5.3: Diagrama UML de las secciones transformadas de la base de datos ENADID 2018.

Con la finalidad de tener los datos de la mujer y sus respectivos embarazos (si es el caso) en una sola base de datos, se unen ambas secciones, MUJER_18 e HISTORIA_18, de forma que estén todas las filas y columnas, y en caso de que no exista registro coincidente se toma como NA. Es decir, cada fila corresponde a un hijo nacido vivo con la información de su madre, o bien, a una mujer sin hijos nacidos vivos.

Por ejemplo, en la figura 5.4 se tiene la información de tres mujeres con meses de nacimiento en diciembre, septiembre y diciembre y con años de nacimiento 1971, 1991 y 1997, cuyas edades al momento de la entrevista son 46, 27 y 20, respectivamente. En la figura 5.5 se tiene el mes y el año de nacimiento de tres embarazos diferentes, pero cuya madre para los primeros dos es la mujer de 46 años y para el último es la mujer de 27 años.

ID_1	FACTOR	UPM	ESTRATO	FEC_MUJ_M	FEC_MUJ_A	EDAD_M	ENTREV_M
010000305102	151	1	3	12	1971	46	9
010002602104	116	3	2	9	1991	27	10
010002602105	116	3	2	12	1997	20	8

Figura 5.4: Tabla MUJER_18.

ID_2	FEC_HIJ_M	FEC_HIJ_A
010000305102	2	1997
010000305102	8	2001
010002602104	12	2011

Figura 5.5: Tabla HISTORIA_18.

Después de unir ambas tablas con ayuda de las variables ID_1 e ID_2, de manera que cada fila contenga la información de la madre y el niño nacido vivo, y en caso de que la mujer no haya sido madre aún, aparece NA en las variables FEC_HIJ_M y FEC_HIJ_A. En la figura 5.6 se observa el conjunto de datos que resulta.

ID_1	FACTOR	UPM	ESTRATO	FEC_MUJ_M	FEC_MUJ_A	EDAD_M	ENTREV_M	FEC_HIJ_M	FEC_HIJ_A
010000305102	151	1	3	12	1971	46	9	2	1997
010000305102	151	1	3	12	1971	46	9	8	2001
010002602104	116	3	2	9	1991	27	10	12	2011
010002602105	116	3	2	12	1997	20	8	NA	NA

Figura 5.6: Tabla unión de MUJER_18 e HISTORIA_18.

Esta operación relacional que se empleó para obtener la figura 5.6 se le conoce como *full join* y es representada por la figura 5.7:



Figura 5.7: Representación de la operación full join a través de un diagrama de Venn.

Por último, para propósito del cálculo de los denominadores, es necesario considerar sólo una vez a cada mujer en nuestra base de datos final, por lo que es fundamental la creación de una variable que determine cuáles elementos de ID_1 (identificador/llave) son duplicados. En R, la función *duplicated* del paquete base nos ayuda a este propósito. La figura 5.8 representa el conjunto de datos con la estructura y variables adecuadas para la obtención de los índices por medio del paquete *fertilityRates*.

5. USO DEL PAQUETE *FERTILITYRATES*

ID_1	FACTOR	UPM	ESTRATO	FEC_MUJ_M	FEC_MUJ_A	EDAD_M	ENTREV_M	FEC_HIJ_M	FEC_HIJ_A	MUJER
010000305102	151	1	3	12	1971	46	10	2	1997	TRUE
010000305102	151	1	3	12	1971	46	10	8	2001	FALSE
010002602104	116	3	2	9	1991	27	9	12	2011	TRUE
010002602105	116	3	2	12	1997	20	8	NA	NA	TRUE

Figura 5.8: Conjunto de datos final.

El código para procesar los datos conforme a lo explicado en esta sección se encuentra en el apéndice C.

5.3. Implementación del paquete

El objetivo del paquete *fertilityRates* es la obtención de estimadores de razón, de la suma de las razones y de intervalo, donde las razones son las tasas específicas de fecundidad, la suma de las razones es la tasa global de fecundidad y los intervalos son los intervalos de confianza para las tasas de fecundidad. Esto se logra mediante el diseño muestral presentado en la sección 2.4 y la metodología del capítulo 3 para X y Y como variables del cociente.

La idea general de qué realiza cada función es la siguiente:

Las funciones *frts_intvw*, *frts_yrly* y *frts_3yrs* crean un objeto *frts_intvw*, *frts_yrly* y *frts_3yrs* que contiene un `data.frame` con las variables¹ necesarias para realizar las tasas/cocientes y un `survey.design` de acuerdo a lo especificado por el usuario en el argumento de la función, esto conforme al periodo de referencia de cada clase.

La función *frts_14yrs* crea un objeto *frts_14yrs*, que contiene cuatro matrices (`matrix`) que corresponden a los nacimientos ocurridos y los tiempos de exposición en los catorce años calendario anteriores al año de levantamiento de la encuesta por edad individual y por grupos quinquenales de edad, respectivamente, y a un `numeric` que es el año de levantamiento de la encuesta.

Para obtener las estimaciones, se programaron los métodos `summary` para cada clase (*frts_intvw*, *frts_yrly*, *frts_3yrs* y *frts_14yrs*). Los métodos para las tres primeras clases proporcionan una lista con las tasas de fecundidad específicas por edad individual y por grupos quinquenales de edad, y la tasa global de fecundidad (estimaciones puntuales y de intervalo). El método para *frts_14yrs* proporciona las tasas de fecundidad específicas por edad individual y por grupos quinquenales de edad, y la tasa global de

¹Estas variables son creadas de acuerdo a la metodología de estimación del capítulo 3, entre ellas están la edad de la madre al nacimiento del hijo nacido vivo y los tiempos de exposición que aporta cada mujer, ya sea por edades individuales o por grupos quinquenales.

fecundidad para los catorce años calendario antes mencionados.

Es necesario mencionar que para poder hacer uso de las funciones del paquete ya instalado, se requiere ejecutar previamente `library(fertilityRates)`, como cualquier otra biblioteca en R. Además, el tiempo promedio de ejecución para obtener estimaciones de acuerdo a cada función es 1885.5 segundos, esto con 4 núcleos, 8 subprocesos y 4,00 GHz. de frecuencia máxima en CPU, y 12 GB en RAM. El código fuente de cada una de las funciones presentadas a continuación se encuentra en el apéndice C.

5.3.1. *frts_intvw*. Estimaciones en el año del levantamiento

Con los datos de la ENADID 2018, las tasas de fecundidad específicas por edad individual de enero a septiembre¹ son:

Código 5.1: Función *frts_intvw* con datos de la ENADID 2018

```
1 mg1 <- frts_intvw(m.intvw = ENTREV_M, y.intvw = 2018, m.wmn = FEC_MUJ_M, y.wmn = FEC_MUJ_A,
2   age.wmn = EDAD_M, m.child = FEC_HIJ_M, y.child = FEC_HIJ_A, wmn.dummy = MUJER, id.wmn = ID_1,
3   ids = UPM, strata = ESTRATO, weights = FACTOR, data = enadid_2018)
```

Código 5.2: Estimaciones

```
1 dat1 <- summary(mg1, level = 0.9) # 90% de confianza
2 dat1[["as_fr_s"]] # tasas por edad
```

```
> dat1[["as_fr_s"]]
  as_fr_s  l_ci u_ci
15  0.019 0.013 0.025
16  0.050 0.039 0.060
17  0.067 0.055 0.079
18  0.100 0.083 0.117
19  0.094 0.080 0.109
20  0.103 0.088 0.119
21  0.106 0.089 0.122
22  0.109 0.093 0.124
23  0.091 0.077 0.104
24  0.090 0.076 0.104
25  0.115 0.099 0.131
26  0.097 0.082 0.112
27  0.108 0.092 0.124
28  0.085 0.071 0.100
29  0.084 0.070 0.099
30  0.076 0.063 0.089
31  0.071 0.057 0.085
32  0.071 0.057 0.085
33  0.062 0.047 0.077
34  0.058 0.047 0.070
35  0.049 0.038 0.060
36  0.053 0.038 0.068
37  0.034 0.023 0.045
38  0.036 0.026 0.046
```

¹Se consideró el mes de levantamiento conforme a lo mencionado en 2.3

5. USO DEL PAQUETE *FERTILITYRATES*

```
39 0.020 0.013 0.027
40 0.021 0.012 0.029
41 0.006 0.003 0.009
42 0.004 0.001 0.006
43 0.003 0.000 0.007
44 0.005 0.002 0.008
45 0.003 0.001 0.005
46 0.002 0.000 0.005
47 0.002 0.000 0.004
48 0.000 0.000 0.000
49 0.002 0.001 0.004
```

5.3.2. *frts_yrly*. Estimaciones anuales en los cuatro años anteriores al levantamiento

Con los datos de la ENADID 2014, las tasas de fecundidad específicas por grupos quinquenales de edad para el año 2013 son:

Código 5.3: Función *frts_yrly* con datos de la ENADID 2014

```
1 mg2 <- frts_yrly(m.intvw = ENTREV_M, y.intvw = 2014, y.ref = 2013, m.wmn = FEC_MUJ_M,
2   y.wmn = FEC_MUJ_A, age.wmn = EDAD_M, m.child = FEC_HIJ_M, y.child = FEC_HIJ_A,
3   wmn.dummy = MUJER, id.wmn = ID_1, ids = UPM, strata = ESTRATO, weights = FACTOR,
4   data = enadid_2014)
```

Código 5.4: Estimaciones

```
1 dat2 <- summary(mg2, level = 0.9) # 90% de confianza
2 dat2[["as_fr_g"]] # tasas por grupos quinquenales de edad
```

```
> dat2[["as_fr_g"]]
  as_fr_g l_ci u_ci
15-19    0.080 0.075 0.085
20-24    0.121 0.116 0.127
25-29    0.106 0.100 0.112
30-34    0.075 0.070 0.080
35-39    0.036 0.033 0.040
40-44    0.009 0.007 0.010
45-49    0.000 0.000 0.001
```

5.3.3. *frts_3yrs*. Estimaciones trianuales dentro de los cuatro años anteriores al levantamiento

Con los datos de la ENADID 2009, la tasa global de fecundidad para el trienio de 2006 a 2008 es:

Código 5.5: Función *frts_3yrs* con datos de la ENADID 2009

```
1 mg3 <- frts_3yrs(m.intvw = ENTREV_M, y.intvw = 2009, y.first = 2008, y.second = 2007,
2   y.third = 2006, m.wmn = FEC_MUJ_M, y.wmn = FEC_MUJ_A, age.wmn = EDAD_M, m.child = FEC_HIJ_M,
3   y.child = FEC_HIJ_A, wmn.dummy = MUJER, id.wmn = ID_1, ids = UPM, strata = ESTRATO,
4   weights = FACTOR, data = enadid_2009)
```

Código 5.6: Estimaciones

```
1 dat3 <- summary(mg3, level = 0.9) # 90% de confianza
2 dat3[["t_fr"]] # tasa global de fecundidad
```

```
> dat3[["t_fr"]]
      t_fr
t_fr 2.255
l_ci 2.214
u_ci 2.296
```

5.3.4. *frts_14yrs*. Estimaciones anuales de los catorce años anteriores al levantamiento

Con los datos de la ENADID 2006, las tasas de fecundidad específicas por grupos quinquenales de edad anuales de 1992 a 2005 son:

Código 5.7: Función *frts_14yrs* con datos de la ENADID 2006

```
1 mg4 <- frts_14yrs(m.intvw = ENTREV_M, y.intvw = 2006, m.wmn = FEC_MUJ_M, y.wmn = FEC_MUJ_A,
2   age.wmn = EDAD_M, m.child = FEC_HIJ_M, y.child = FEC_HIJ_A, wmn.dummy = MUJER, id.wmn = ID_1,
3   weights = FACTOR, data = enadid_2006)
```

Código 5.8: Estimaciones

```
1 dat4 <- summary(mg4)
2 dat4[["as_fr_g"]] # tasas por edad
```

```
> dat4[["as_fr_g"]]
      1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005
15-19 0.087 0.093 0.092 0.097 0.086 0.084 0.081 0.082 0.074 0.080 0.077 0.066 0.063 0.063
20-24 0.187 0.170 0.169 0.173 0.162 0.149 0.157 0.159 0.165 0.140 0.141 0.135 0.121 0.131
25-29 0.164 0.151 0.160 0.161 0.136 0.139 0.135 0.138 0.151 0.135 0.134 0.124 0.116 0.124
30-34 0.131 0.115 0.114 0.109 0.108 0.095 0.109 0.104 0.105 0.093 0.089 0.086 0.080 0.091
35-39 0.063 0.065 0.062 0.054 0.060 0.052 0.063 0.050 0.048 0.056 0.042 0.045 0.036 0.043
40-44 0.036 0.033 0.030 0.027 0.024 0.022 0.018 0.013 0.017 0.013 0.013 0.011 0.011 0.014
45-49 0.003 0.003 0.003 0.003 0.003 0.003 0.002 0.002 0.002 0.002 0.002 0.003 0.002 0.003
```

Hacer uso del paquete brinda al usuario una manera ágil y óptima del obtener los principales índices de la fecundidad, ya que como se mostró en el presente capítulo, la manera de implementar cada función del paquete es sencilla y rápida. A continuación se presentan las estimaciones obtenidas con el paquete y con la intención de validarlas, se contrastan con otras cifras.

Estimaciones de la fecundidad con las encuestas ENADID

En este capítulo se presentan algunas de las estimaciones de la fecundidad obtenidas con los datos de las seis encuestas ENADID mediante el paquete *fertilityRates* que se desarrolló en este trabajo. Además, con la intención de validar las estimaciones que se generaron, se hace un comparativo con cifras oficiales.

Las tasas globales de fecundidad

Para revisar de manera general las TGF estimadas con el paquete *fertilityRates* se presenta la figura 6.1. En ella se muestran las TGF posibles de las seis encuestas ENADID, de manera puntual y por intervalo, y para periodos trianuales, anuales y menores a un año, respectivamente.

Dentro de las cualidades que tiene *fertilityRates* se encuentran la facilidad y rapidez para obtener tasas de tres periodos diferentes, con intervalos de confianza según el caso, y para 14 años anteriores a la encuesta. Es por esto que la figura 6.1, además de mostrar el comparativo de manera trianual por encuesta, también proporciona la evolución en el tiempo del nivel de la fecundidad al considerar todas las estimaciones de la TGF realizadas por el paquete. Esto permite tener un mejor contraste de la información y muestra de manera gráfica que las estimaciones son plausibles entre fuentes.

6. ESTIMACIONES DE LA FECUNDIDAD CON LAS ENCUESTAS ENADID

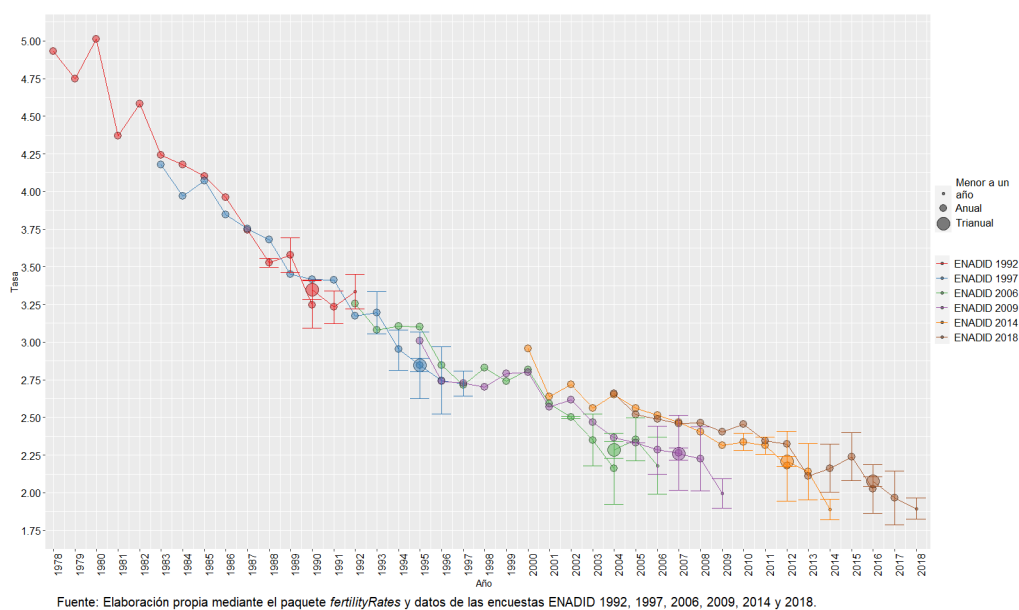


Figura 6.1: TGF y sus intervalos de confianza para diferentes periodos con información de la ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas con el paquete *fertilityRates*.

Es posible notar que algunas de las tasas puntuales para el año de la entrevista (de periodo menor a un año) son más bajas que las del año anterior, e incluso, respecto al mismo año, pero estimada con otra encuesta. Las estimaciones que son diferentes estadísticamente son: de 2006 de la ENADID 2006 respecto a las estimaciones puntuales anuales de las ENADID 2014 y 2018, y de 2014 de la ENADID 2014 respecto a la estimación puntual anual de la ENADID 2018.

Además, para el periodo de 2000 a 2009, se tienen estimaciones de cuatro encuestas, en donde se aprecia que las tasas puntuales de las ENADID 2014 y 2018 tienen valores más altos que las de ENADID 2006 y 2009. Con base en los intervalos de confianza, solo se puede asegurar que las tasas no resumen de manera estadísticamente significativa la fecundidad para el periodo de 2000 a 2006 y para el año 2009, conclusión que no se puede tener para las tasas de 2007 y 2008. Sin embargo, esto es diferente para estimaciones anteriores a 2000 y las estimaciones de las dos encuestas más recientes (ENADID 2014 y 2018), donde las estimaciones tienen gran coincidencia y los estimadores puntuales caen dentro de los intervalos de confianza.

Una manera externa de validar las estimaciones es mediante la tabla 6.1, donde se presentan cifras oficiales (CONAPO (2016) y INEGI (2019d)) y las estimadas con el paquete en R.

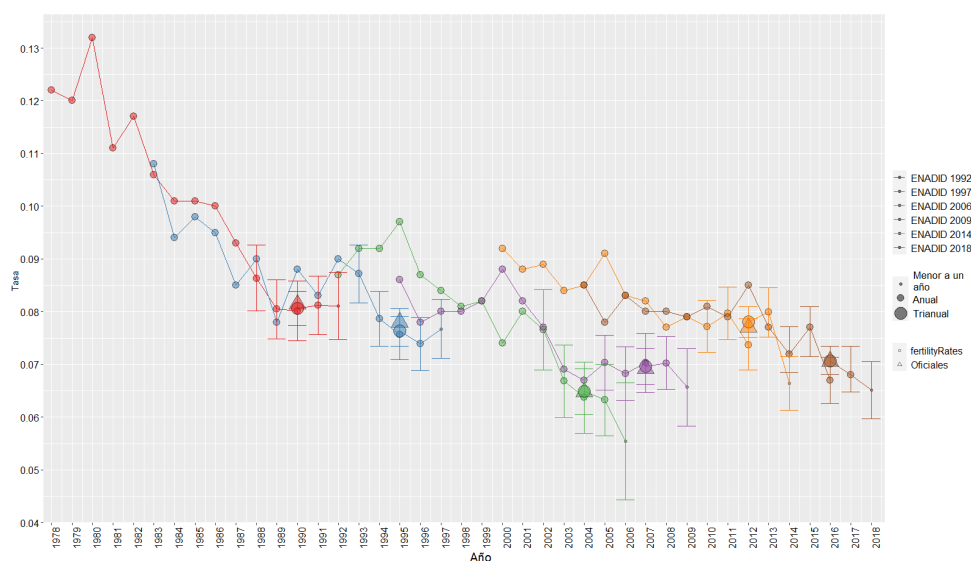
	TGF					
	1990	1995	2004	2007	2012	2016
<i>Oficiales</i>	3.35	2.86	2.28	2.25	2.21	2.07
<i>fertilityRates</i>	3.35	2.85	2.29	2.26	2.21	2.07

Tabla 6.1: Estimaciones trianuales de las TGF con base en la información de la ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI, y con el paquete *fertilityRates*. Se hace la referencia al año central del periodo correspondiente.

Se aprecia de forma clara que no existen diferencias entre tasas y el tener tal semejanza sugiere que los cálculos con el paquete realizan el resumen de la TGF adecuadamente.

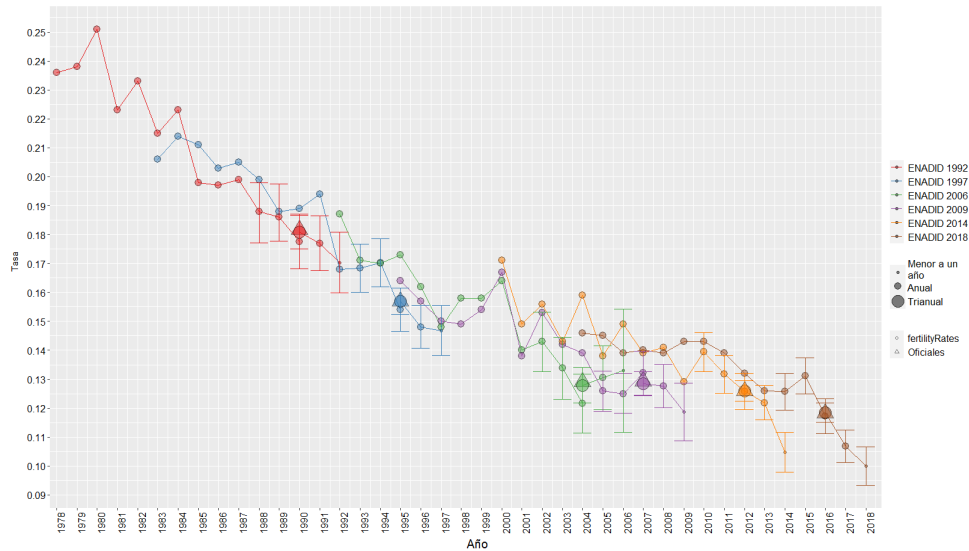
Las tasas de fecundidad específicas por edad

Otro contraste entre las estimaciones realizadas con *fertilityRates* y las estimaciones realizadas por el CONAPO y el INEGI, es observar de manera particular los diferentes grupos quinquenales de edad. Las tasas específicas por edad que se obtienen con el paquete permiten observar, al igual que con las TGF, el cambio en el tiempo. A continuación se muestran las tasas de fecundidad de cada grupo quinquenal de edad para las diferentes encuestas y distintos periodos de observación.



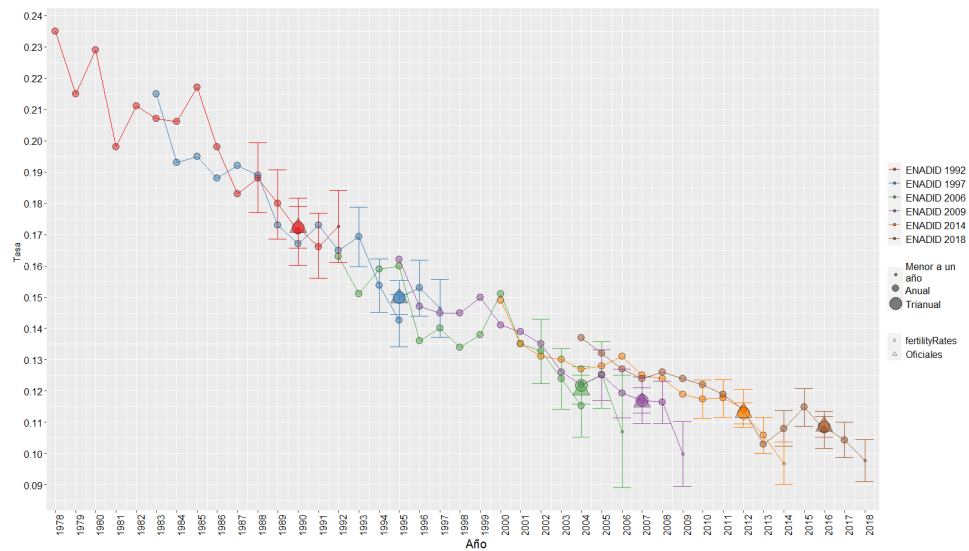
Fuente: Elaboración propia mediante el paquete *fertilityRates* y datos de las encuestas ENADID 1992, 1997, 2006, 2009, 2014 y 2018, y las cifras publicadas por el CONAPO(2016) y el INEGI(2019d).

Figura 6.2: Tasas de fecundidad del grupo quinquenal 15-19 y sus intervalos de confianza para diferentes periodos con información de las ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI y con el paquete *fertilityRates*.



Fuente: Elaboración propia mediante el paquete *fertilityRates* y datos de las encuestas ENADID 1992, 1997, 2006, 2009, 2014 y 2018, y las cifras publicadas por el CONAPO(2016) y el INEGI(2019d).

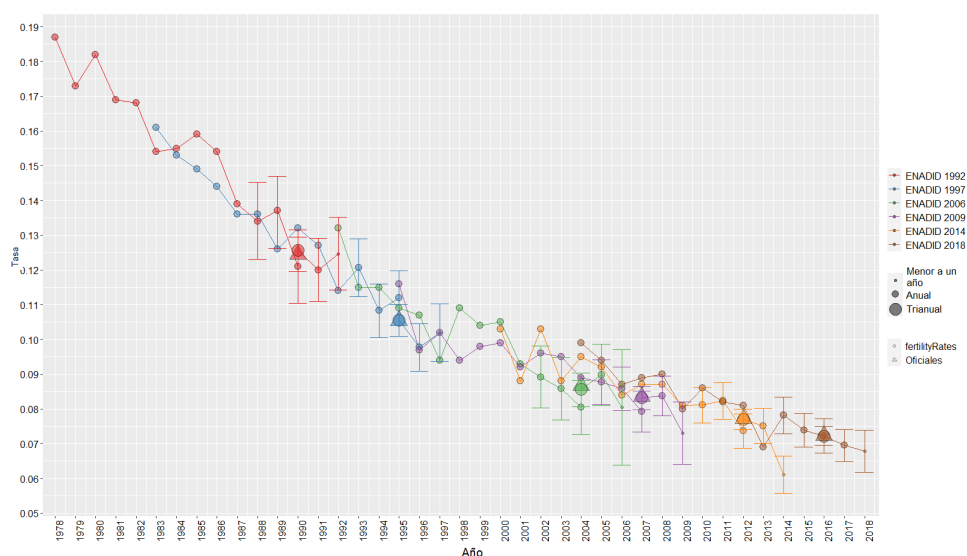
Figura 6.3: Tasas de fecundidad del grupo quinquenal 20-24 y sus intervalos de confianza para diferentes periodos con información de las ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI y con el paquete *fertilityRates*.



Fuente: Elaboración propia mediante el paquete *fertilityRates* y datos de las encuestas ENADID 1992, 1997, 2006, 2009, 2014 y 2018, y las cifras publicadas por el CONAPO(2016) y el INEGI(2019d).

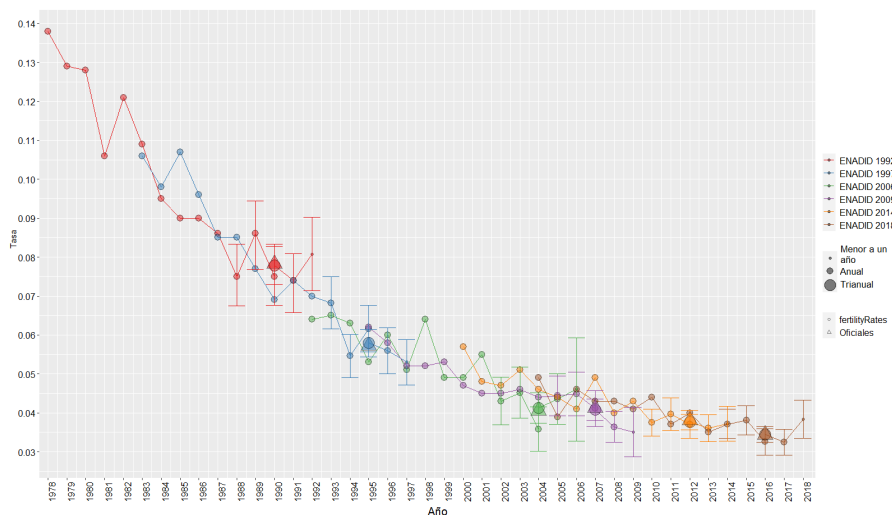
Figura 6.4: Tasas de fecundidad del grupo quinquenal 25-29 y sus intervalos de confianza para diferentes periodos con información de las ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI y con el paquete *fertilityRates*.

6. ESTIMACIONES DE LA FECUNDIDAD CON LAS ENCUESTAS ENADID



Fuente: Elaboración propia mediante el paquete *fertilityRates* y datos de las encuestas ENADID 1992, 1997, 2006, 2009, 2014 y 2018, y las cifras publicadas por el CONAPO(2016) y el INEGI(2019d).

Figura 6.5: Tasas de fecundidad del grupo quinquenal 30-34 y sus intervalos de confianza para diferentes periodos con información de las ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI y con el paquete *fertilityRates*.



Fuente: Elaboración propia mediante el paquete *fertilityRates* y datos de las encuestas ENADID 1992, 1997, 2006, 2009, 2014 y 2018, y las cifras publicadas por el CONAPO(2016) y el INEGI(2019d).

Figura 6.6: Tasas de fecundidad del grupo quinquenal 35-39 y sus intervalos de confianza para diferentes periodos con información de las ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI y con el paquete *fertilityRates*.

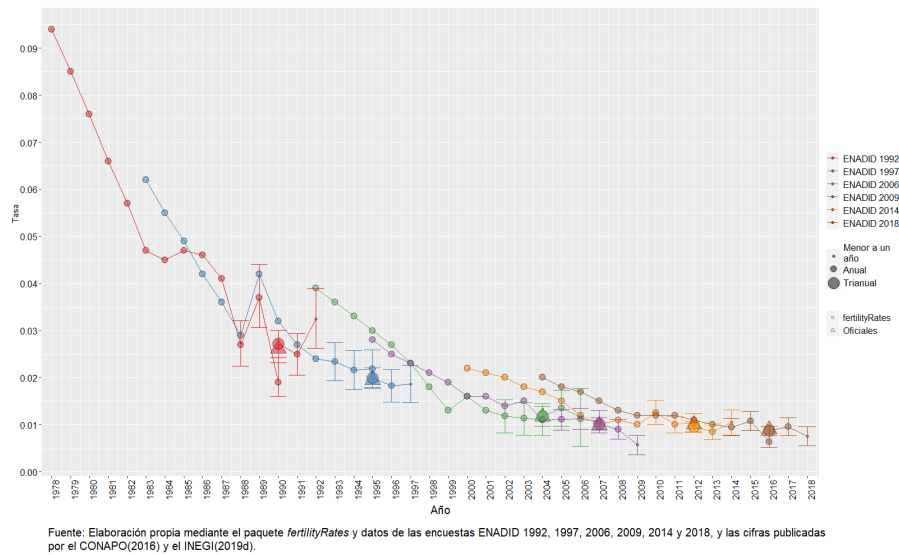


Figura 6.7: Tasas de fecundidad del grupo quinquenal 40-44 y sus intervalos de confianza para diferentes periodos con información de las ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI y con el paquete *fertilityRates*.

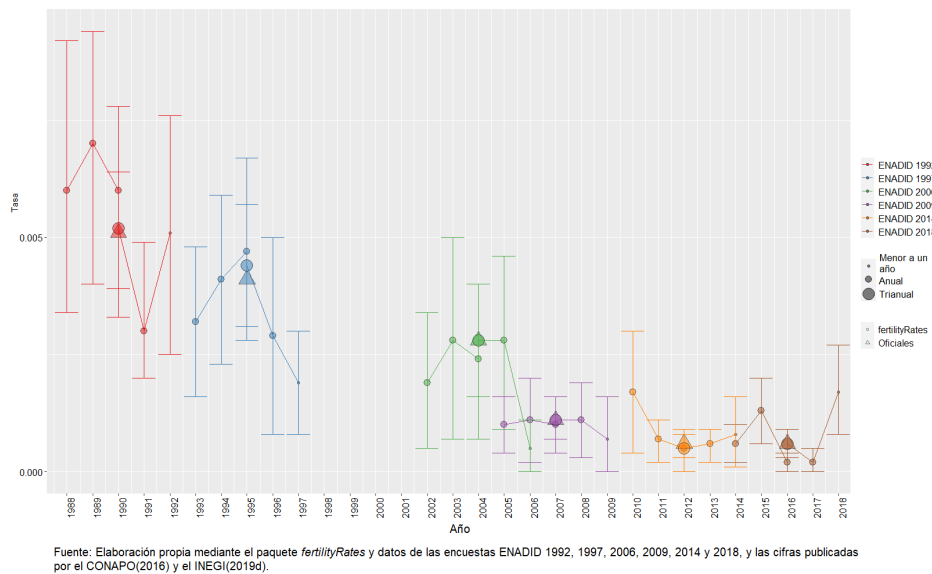


Figura 6.8: Tasas de fecundidad del grupo quinquenal 45-49 y sus intervalos de confianza para diferentes periodos con información de las ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI y con el paquete *fertilityRates*.

En las gráficas anteriores se observa de manera general que las cifras generadas con el paquete en R son semejantes entre encuestas para periodos de tiempo cercanos y con respecto a las cifras oficiales, salvo para el primer y el último grupo de edad quinquenal,

6. ESTIMACIONES DE LA FECUNDIDAD CON LAS ENCUESTAS ENADID

en los que se aprecian diferencias puntuales que se explican a continuación.

Dentro del grupo 15-19, las diferencias entre las estimaciones del paquete son propias de las fuentes, es decir, las ENADID 2006 y 2009 muestran niveles diferentes con respecto a las ENADID 2014 y 2018 para el periodo de 2000 a 2009, sin embargo, esto sugiere deficiencias en la calidad de la información¹ y no en la metodología que se aplicó para calcular las tasas. Por parte del grupo 45-49, no se observa una tendencia clara como en los otros grupos quinquenales edad, esto se debe a la poca información retrospectiva con la que se cuenta; no obstante, las tasas obtenidas con el paquete *fertilityRates* coincide con los niveles de las tasas oficiales.

La siguiente tabla 6.2 muestra las estimaciones trianuales oficiales y las realizadas con *fertilityRates*. En ella se puede apreciar una gran similitud entre los valores de las tasas, y aunque no son puntualmente iguales en algunos casos, la estructura y el nivel de la fecundidad son sustancialmente los mismos. Esto permite afirmar que la metodología programada en el paquete estima adecuadamente los índices de la fecundidad antes presentados.

¹En el capítulo 2 se mostró el tamaño de muestra de las diferentes encuestas, por lo que es posible que las diferencias que se mencionan se deban a un problema de representatividad de la ENADID 2006 y posiblemente de la ENADID 2009.

Grupos de edad	Oficiales						<i>fertilityRates</i>					
	1990	1995	2004	2007	2012	2016	1990	1995	2004	2007	2012	2016
15-19	0.0814	0.0780	0.0648	0.0692	0.077	0.0706	0.0806	0.0763	0.0649	0.0700	0.0780	0.0710
20-24	0.1817	0.1570	0.1290	0.1294	0.1260	0.1182	0.1808	0.1570	0.1279	0.1280	0.1260	0.1180
25-29	0.1722	0.1467	0.1203	0.1164	0.1131	0.1088	0.1723	0.1499	0.1210	0.1170	0.1130	0.1090
30-34	0.1244	0.1055	0.0867	0.0836	0.0772	0.0724	0.1256	0.1055	0.0856	0.0830	0.0770	0.0720
35-39	0.0783	0.0572	0.0406	0.0412	0.0381	0.0344	0.0778	0.0579	0.0413	0.0410	0.0370	0.0340
40-44	0.0262	0.0195	0.0117	0.0098	0.0100	0.0090	0.0272	0.0199	0.0118	0.0100	0.0100	0.0090
45-49	0.0051	0.0041	0.0028	0.0011	0.0006	0.0060	0.0052	0.0044	0.0028	0.0010	0.0010	0.0010

Tabla 6.2: Estimaciones trianuales de las tasas de fecundidad específicas por grupos quinquenales de edad con base en la información de la ENADID 1992, 1997, 2006, 2009, 2014 y 2018 estimadas por el CONAPO y el INEGI, y con el paquete *fertilityRates*. Se hace la referencia al año de en medio del periodo correspondiente.

Conclusiones y consideraciones

En este trabajo se presentó el paquete *fertilityRates* con implementación en **R**. El paquete puede ser una herramienta para analizar, investigar y enseñar aspectos de la fecundidad mediante sus principales índices calculados con datos de las historias de embarazos presentes en las encuestas demográficas por muestreo. Incluso, si el nivel de desagregación de la encuesta lo permite, se pueden obtener índices representativos para segmentos o características específicas.

La facilidad y rapidez con la que se obtiene la información son aspectos destacables del paquete. Debido a esto, los resultados obtenidos a partir de su uso pueden ser de gran utilidad para instituciones locales y federales como el CONAPO o el INEGI, ya que brindan la posibilidad de acelerar su análisis para la toma de decisiones o bien, como apoyo para la elaboración de la Conciliación Demográfica de México, por mencionar algunos usos.

Aunque el paquete fue hecho para realizar las estimaciones con información de las ENADID, éste puede ser fácilmente aplicable a otras encuestas, siempre y cuando se tenga la información necesaria y estructurada de acuerdo a lo presentado.

fertilityRates es de distribución libre y código abierto lo que facilita la participación para mejorar o añadir funcionalidades. Otra ventaja es que se encuentra desarrollado en **R** y al ser un software libre, su uso no representa ningún costo o limitación en comparación al módulo desarrollado por Schoumaker (2013) que genera estimaciones similares, pero su implementación requiere de Stata, que es un software de pago.

Algunas áreas de oportunidad para el paquete serían la estimación de índices para años naturales, y con ello poder obtener tasas a mitad de año, la optimización de subprocesos como es el caso de la varianza para la TGF, y programación de excepciones para que en caso de que el usuario ingrese mal algún o algunos parámetros en las funciones, éste pueda identificar fácilmente los errores en las entradas.

Comentarios finales

En el proceso de elaboración de este trabajo se tomó la decisión de cambiar la manera de calcular la edad de la mujer/madre a un tiempo t (capítulo 3). Esta decisión fue porque la forma propuesta por [Moultrie \(2013\)](#) de utilizar la fecha de nacimiento generaba estimaciones de una fecundidad más tardía que las obtenidas con la edad declarada y en comparación a las cifras oficiales (ver apéndice A). Esto sugiere la necesidad de una revisión especial en la captación de las fechas de nacimiento y las edades declaradas de las mujeres en las encuestas ENADID.

Apéndice A

Edad de las mujeres en la encuesta

Las tasas de fecundidad se obtienen mediante la edad de la mujer en el numerador y en el denominador (1.4). Hay dos formas de calcular la edad de la mujer con la información presente en las encuestas que captan las historias de embarazos, una es con la fecha de nacimiento y la otra con la edad declarada al momento de la entrevista. Este apéndice ayuda a entender la elección del cálculo de la edad de la mujer con base en la edad declarada aplicado en el paquete *fertilityRates*.

Se consideraron los datos de las encuestas ENADID 1992, 2006 y 2014 para contrastar las dos metodologías de edad de la mujer aplicadas al cálculo de tasas específicas por edad y por grupo quinquenal para periodos trianuales anteriores al levantamiento de la encuesta.

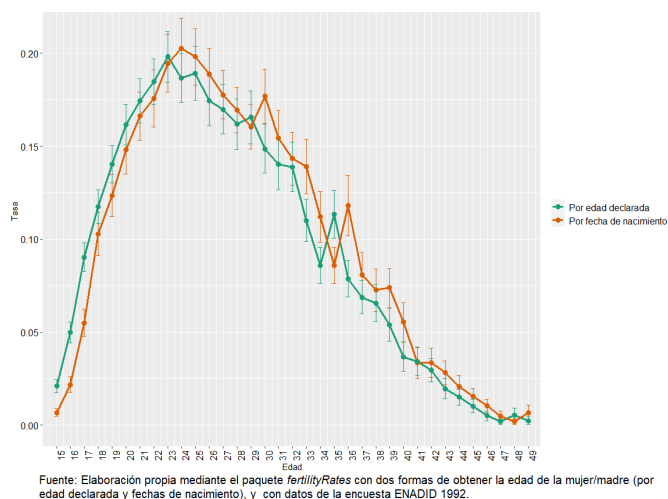


Figura A.1: Tasas específicas de fecundidad trianuales por edad individual y sus intervalos de confianza para el periodo 1989-1991 con información de la encuesta ENADID 1992, obtenidas mediante el paquete *fertilityRates* y dos maneras diferentes de calcular la edad de la mujer/madre.

A. EDAD DE LAS MUJERES EN LA ENCUESTA

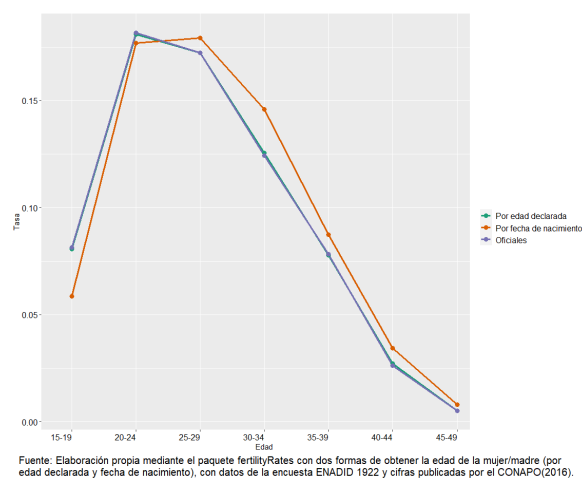


Figura A.2: Tasas específicas de fecundidad trianuales por grupo quinquenal de edad para el periodo 1989-1991 con información de la encuesta ENADID 1992, obtenidas mediante el paquete *fertilityRates* y dos maneras diferentes de calcular la edad de la mujer/madre, y la cifras publicadas por el CONAPO.

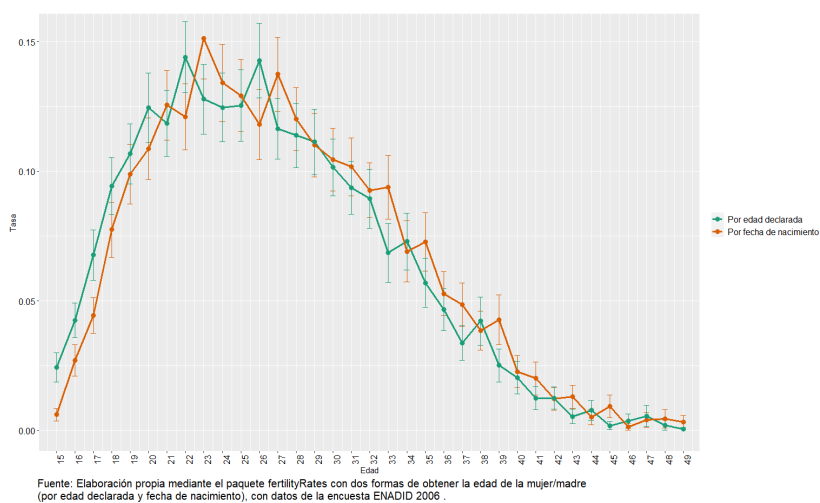
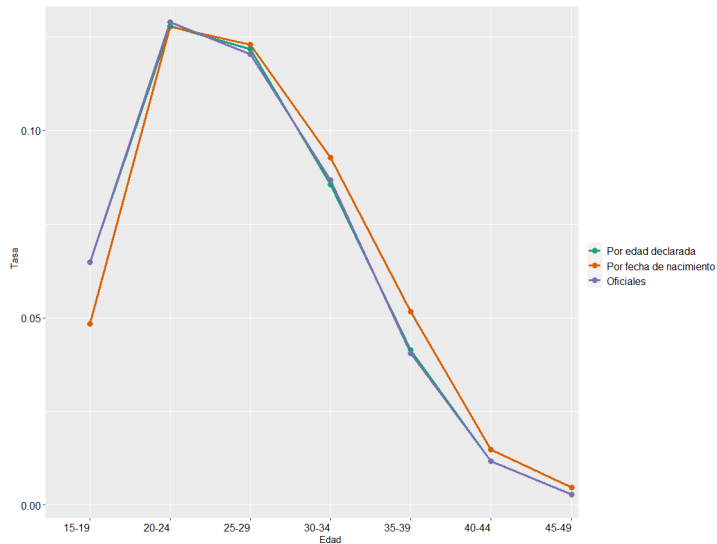
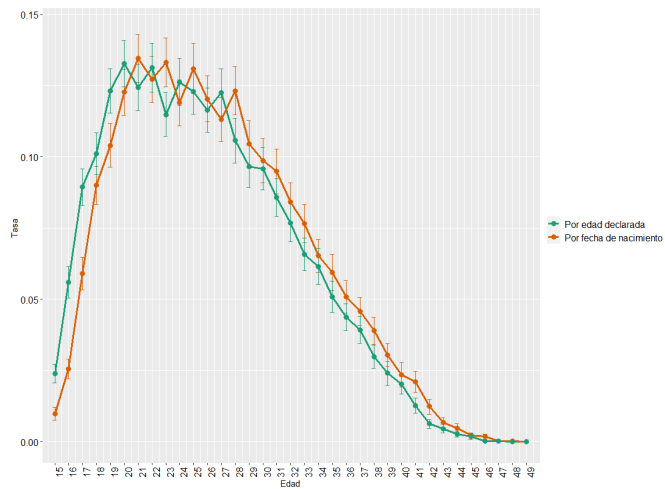


Figura A.3: Tasas específicas de fecundidad trianuales por grupo quinquenal de edad para el periodo 2003-2005 con información de la encuesta ENADID 2006, obtenidas mediante el paquete *fertilityRates* y dos maneras diferentes de calcular la edad de la mujer/madre.



Fuente: Elaboración propia mediante el paquete *fertilityRates* con dos formas de obtener la edad de la mujer/madre (por edad declarada y fecha de nacimiento), con datos de la encuesta ENADID 2006 y cifras publicadas por el CONAPO(2016).

Figura A.4: Tasas específicas de fecundidad trianuales por grupo quinquenal de edad para el periodo 2003-2005 con información de la encuesta ENADID 2006, obtenidas mediante el paquete *fertilityRates* y dos maneras diferentes de calcular la edad de la mujer/madre, y la cifras publicadas por el CONAPO.



Fuente: Elaboración propia mediante el paquete *fertilityRates* con dos formas de obtener la edad de la mujer/madre (por edad declarada y fecha de nacimiento), con datos de la encuesta ENADID 2014.

Figura A.5: Tasas específicas de fecundidad trianuales por grupo quinquenal de edad para el periodo 2011-2013 con información de la encuesta ENADID 2014, obtenidas mediante el paquete *fertilityRates* y dos maneras diferentes de calcular la edad de la mujer/madre.

A. EDAD DE LAS MUJERES EN LA ENCUESTA

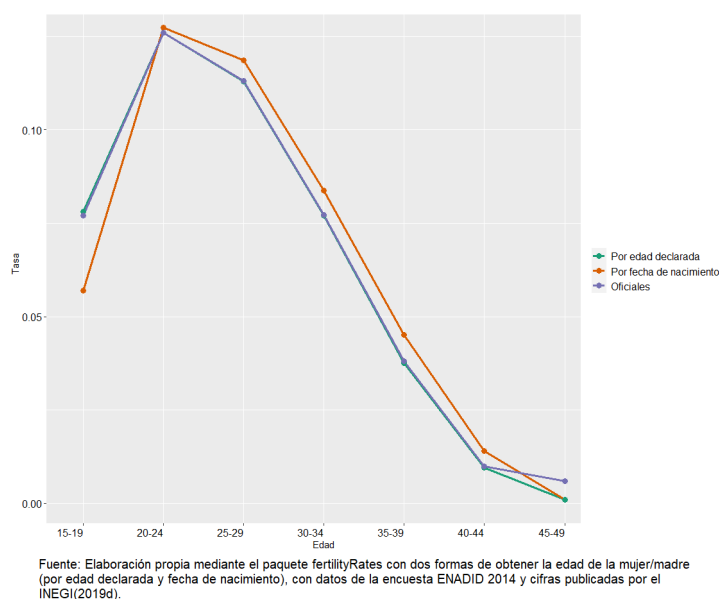


Figura A.6: Tasas específicas de fecundidad trianuales por grupo quinquenal de edad para el periodo 2011-2013 con información de la encuesta ENADID 2014, obtenidas mediante el paquete *fertilityRates* y dos maneras diferentes de calcular la edad de la mujer/madre, y la cifras publicadas por el INEGI.

Como se puede observar en las figuras anteriores, existe una errónea declaración en las fechas de nacimiento de las mujeres, a partir de las cuales se obtienen las edades de las mujeres, o bien, en la declaración de las edades de las mujeres. Estos errores tienen implicaciones en los niveles y tendencias de la fecundidad en diferentes magnitudes y sentidos (CONAPO 2005). En este caso, con la intención de que las estimaciones sean comparables con las cifras oficiales y exista una forma de validar que se hizo un adecuado desarrollo del paquete, independiente de la calidad de los datos para dichas variables, se tomó la decisión de elegir para la implementación el cálculo de la edad de la mujer/madre mediante la edad declarada en el momento de la entrevista .

Apéndice B

Programación en R

R es un lenguaje de programación y *software* libre con licencia de GNU¹, el cual tiene un ambiente enfocado al cómputo estadístico y gráfico, donde es posible implementar técnicas estadísticas y así ampliar la gama de aplicaciones en el procesamiento de datos. Este lenguaje se compila y se ejecuta en una gran variedad de plataformas como Linux, Windows y MacOS. Para descargar R, se debe elegir el sistema operativo con el que se trabaja y un espejo de CRAN².

La creación de R fue llevada a cabo por Ross Ihaka y Robert Gentleman en Nueva Zelanda en 1992, cuyo nombre “R”, hace referencia a las iniciales Ross y Robert. La finalidad de R era ser un lenguaje didáctico para el curso de Introducción a la Estadística de la Universidad de Nueva Zelanda, por lo que acogieron ideas del lenguaje S desarrollado en Laboratorios Bell Laboratories (AT&T) en 1976 (Santana and Farfán 2014, 8). Tuvieron que pasar cerca de 7 años luego de su creación, para tener la version 1.0. de R, siendo un software completo, estable, libre y con un soporte (R-help y R-devel).

Actualmente, R proporciona un manejo y almacenamiento seguro de los datos, cuenta con un conjunto de herramientas para cuestiones matriciales y de análisis de datos, y diferentes estructuras de control, teniendo con ello, flexibilidad y diversidad de instrumentos. Además de esto, R nos permite extender o implementar técnicas estadísticas por medio de paquetes, los cuales son puestos a disposición en el sitio de internet CRAN.

Fundamentos

Existen diversos lenguajes de programación, los cuales siguen diferentes modelos para abordar los problemas, a estos modelos se les conocen como paradigmas de programación. Algunos de los paradigmas de programación más ocupados son:

¹GNU *General Public License*, es la autorización formal de *copyright*, permitiendo la modificación y redistribución del software bajo la misma licencia.

²CRAN *The Comprehensive R Archive Network* es una red de servidores web de todo el mundo donden se almacena el código fuente, los manuales y la documentación de R, y los paquetes aportados. <https://cran.r-project.org/>

- Imperativo
- Declarativo
- Lógico
- Funcional
- Orientado a objetos

Es justamente, la funcionalidad y la orientación a objetos con las que cuenta R, haciendo que éste tenga una sintaxis simple e intuitiva.

El hecho de que R sea orientado a objetos hace que las variables, bases de datos, funciones, entre otros, se almacenen en la memoria de la computadora como *objetos* con un nombre especificado por el usuario, para después ser modificados con *operadores* o *funciones* e incluso poder *escribirlos*¹ (Paradis 2003, 5). Por otra parte, que R sea funcional proporciona la facilidad de usar funciones matemáticas y así mismo implementar diferentes metodologías.

Datos

Los tipos de datos definidos en R, y denominados *básicos* o *atómicos*, son:

- Tipo numérico
 - números reales (`numeric`)
 - números enteros (`integer`)
 - números complejos (`complex`)
- Tipo caracter (`character`)
- Tipo booleano (`logical`)

Es importante mencionar que R indica con `NA` (“Not Available”) como valores faltantes o missing value y con `NaN` (“Not a Number”) a un cálculo que tuvo indeterminación 0/0.

Para manejar y hacer uso de estos tipos de datos, es necesario encapsularlos en un objeto, de tal forma que éste se puede “ver” como un conjunto de datos.

¹Hace referencia a salvar los datos en un fichero.

Operadores

Los operadores en programación nos ayudan a realizar una determinada operación sobre uno o más valores/variables y obtener un determinado resultado.

Operadores aritméticos

Los operadores aritméticos pueden ser implementados a tipos numéricos o tipos booleanos y actúan en al menos dos elementos.

Suma [+]

```
> 1+9  
[1] 10
```

Resta [-]

```
> 2-8  
[1] -6
```

Multiplicación [*]

```
> 3*7  
[1] 21
```

División [/]

```
> 4/6  
[1] 0.6666667
```

Potencia [^]

```
> 5^5  
[1] 3125
```

División entera [%/%]

```
> 6 %/% 4  
[1] 1
```

Operadores lógicos

Los operadores lógicos pueden ser empleados en tipos numéricos o tipos booleanos y devuelven uno o más valores lógicos (TRUE o FALSE).

Negación / NO lógico [!]

```
> !1  
[1] FALSE
```

Éste se puede emplear en uno o varios datos.

Conjunción/ Y lógico [&]

B. PROGRAMACIÓN EN R

```
> 0 & 1
[1] FALSE
```

```
> FALSE & TRUE
[1] FALSE
```

Disyunción / O lógico [|]

```
> 0 | 1
[1] TRUE
```

Para los operadores de conjunción (Y) o disyunción (O), la comparación se puede dar para todos los elementos del objeto (& o |) y con esto obtener valores lógicos para cada comparación, o bien la comparación únicamente para el primer elemento del objeto (&& o ||) y tener un valor lógico.

Operadores comparativos

Los operadores comparativos al igual que los aritméticos, actúan sobre dos o más elementos de tipo numérico o booleano, devolviendo uno o varios valores lógicos.

Menor que [<] / Menor o igual que [<=]

```
> 2 < 90
[1] TRUE
```

Mayor que [>] / Mayor o igual que [>=]

```
> 2 >= 90
[1] FALSE
```

Igual que [==]

```
> 2 == 90
[1] FALSE
```

Diferente que [!=]

```
> 2 != 90
[1] TRUE
```

En caso de querer hacer una comparación de más de dos elementos, por ejemplo una desigualdad doble $0 < 129 \leq 237$, es necesario hacer uso del operador conjunción, es decir:

```
> 0 < 129 & 129 <= 237
[1] TRUE
```

Operadores de asignación

Para poder crear variables u objetos y hacer uso de éstos, es necesario que el usuario los asigne a un nombre específico, tomando en consideración que R distingue entre mayúsculas y minúsculas.

Asignación izquierda [< -]

```
> valor <- 'Hola'
> valor
[1] "Hola"
```

Asignación derecha [- >]

```
> TRUE -> bandera
> bandera
[1] TRUE
```

Función asignar [assign()]

```
> assign("x",1568)
> x
[1] 1568
```

Además de estos operadores, es posible hacer uso del =, sin embargo, es una mala práctica y se debe tratar de evitar.

Objetos

En R, el manejo de los datos es a través de objetos, existe una gran variedad de éstos. Algunos de los más usados son vectores, matrices, factores, listas, data frames y funciones. Cada uno de éstos cuenta con nombre, contenido y atributos que son definidos por el usuario cuando son creados y que nos ayudan a reconocerlos, usarlos, modificarlos y saber los tipos de datos que representan (datos básicos), para así poder dar un tratamiento estadístico adecuado.

En general, la forma de creación de los objetos es como se muestra en la figura B.1.

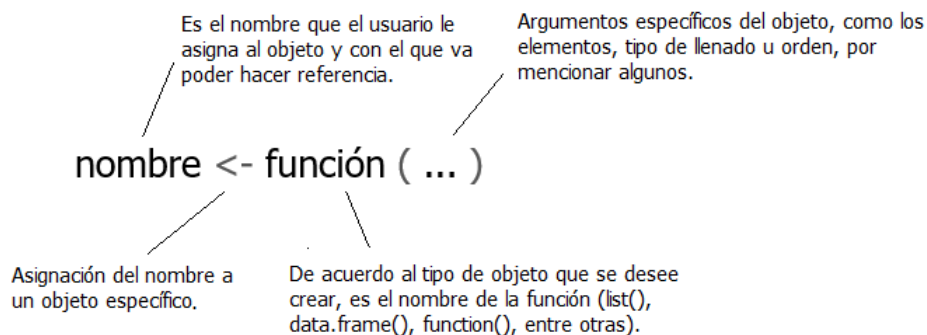


Figura B.1: Creación de un objeto

Vectores

Un vector es una serie de datos del *mismo* tipo, cuya longitud puede ser desde 1, como es el caso de los ejemplos de asignación. Para crear un vector podemos hacer uso de diferentes funciones, algunas de ellas son:

Función concatenar [c()]: Para valores específicos.

```
> val <- c(9, -6, -78)
> val
[1] 9 -6 -78
```

Función asignar [assign()]: Para valores específicos.

```
> assign("nom", c('Gato', 'Perro', 'Pez'))
> print(nom)
[1] "Gato" "Perro" "Pez"

> class(nom)
[1] "character"
```

La función `class()` nos ayuda a saber el tipo de dato que caracteriza a los elementos del vector, en este caso, el vector `nom` está formado por caracteres.

Función vector [vector()]: Vector de ceros o FALSE (según se especifique) y con longitud determinada.

```
> ceros <- vector("integer", 21)
> ceros
[1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

> falsos <- vector("logical", 7)
> falsos
[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

Operador [:]: Permite generar un vector de enteros crecientes o decrecientes indicado por sus extremos.

```
> vec <- 10:5
> vec
[1] 10 9 8 7 6 5
```

Matrices

Las matrices son vectores de vectores, es decir, vectores bidimensionales. Éstas cuenta con renglones y columnas y puede ser creadas con la función `matrix()` o con el atributo de dimensión.

Función matriz [matrix()]

```
matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)
```

Donde el argumento `data` son los valores que contendrá la matriz, `nrow` y `ncol` especifican el número de renglones y columnas respectivamente, `byrow` ayuda a indicar si el llenado se hará por renglón (TRUE) o por columna (FALSE) y `dimnames` permite nombrar las filas y las columnas.

```
> ar <- matrix(data = 80:89, nrow =5, ncol =2)
> ar
      [,1] [,2]
[1,]  80  85
[2,]  81  86
[3,]  82  87
[4,]  83  88
[5,]  84  89

> mat <- matrix(c(9,-8,7,-6,5,-4), 2, 3, byrow=TRUE,
+             dimnames = list(rows = c("row1", "row2"),
+             cols = c("C.1", "C.2", "C.3")))
> mat
      cols
rows  C.1 C.2 C.3
row1   9 -8  7
row2  -6  5 -4
```

También es posible crear una matriz cuyos elementos sean caracteres, sin embargo, todas las entradas de ésta deberán ser caracteres. Es decir, las matrices, al igual que los vectores aceptan un solo tipo de dato.

Función dimensión [`dim()`]

La función `dim()` nos ayuda a saber el número de renglones y columnas de una matriz, sin embargo, es posible crear una matriz a partir de un vector agregándole a este dimensión, es decir:

```
> obj <- rep(2, 12)
> obj
[1] 2 2 2 2 2 2 2 2 2 2 2 2

> dim(obj)
NULL

> dim(obj) <- c(3, 4)
> obj
      [,1] [,2] [,3] [,4]
[1,]  2   2   2   2
[2,]  2   2   2   2
[3,]  2   2   2   2
```

Es importante mencionar que el vector que se le asigna a la dimensión cuenta con dos elementos, donde el primero indica el número de renglones y el segundo el número de columnas. En nuestro ejemplo, indicamos que se querían 3 renglones y 4 columnas.

Factores

Los factores son usados para codificar los elementos de un vector como categorías de variables nominales u ordinales, para darle así un tratamiento estadístico más adecuado. A diferencia de un vector de caracteres, que también puede contener las categorías de una variable categórica, el factor sí reconoce los niveles y orden de los datos.

Función factor [factor()]

```
> sexo <- factor(c('F', 'M', 'M', 'M', 'F'))
> sexo
[1] F M M M F
Levels: F M

> not <- factor(1:3, labels = c("A", "B", "C"))
> not <- ordered(not, levels = c("A", "B", "C") )
> not
[1] A B C
Levels: A < B < C
```

Data frames

Un data frame está compuesto de variables que pueden ser vectores y factores, y éstos deben de tener la misma longitud. Para crear un data frame, es necesario emplear la función `data.frame()`. Sin embargo, también es posible tener uno a través de la lectura de archivos como se verá más adelante en la subsección 1.1.5..

Función data frame [data.frame()]

```
> sexo <- factor(c('F', 'M', 'M', 'M', 'F'))
> not <- factor(1:3, labels=c("A", "B", "C"))
> group <- data.frame(sexo, not)
Error in data.frame(sexo, not) :
  arguments imply differing number of rows: 5, 3
```

En este caso, R nos genera un error debido a que el factor `sexo` tiene longitud 5, mientras que `not` sólo 3.

```
> df <- data.frame(num = 1:10, et = factor(c(rep('L',3), rep('S',7))))
> df
  num et
1   1  L
2   2  L
3   3  L
4   4  S
5   5  S
6   6  S
7   7  S
8   8  S
9   9  S
10  10 S
```

```

> names(df)
[1] "num" "et"

> names(df) <- c('Numeros', 'Etiqueta')
> df
  Numeros Etiqueta
1        1        L
2        2        L
3        3        L
4        4        S
5        5        S
6        6        S
7        7        S
8        8        S
9        9        S
10       10        S

```

Para nombrar las variables del data frame en caso de que sean vectores/factores no creados con anterioridad, puede nombrarse directamente o bien, con la función `names()` asignándole un vector de caracteres de la misma longitud.

Listas

Las listas son usadas como contenedores de diferentes tipos de objetos, como son datos atómicos, vectores, matrices, data.frames, funciones e incluso otras listas, por mencionar algunos.

Función lista [`list()`]

```

> df <- data.frame(num = 1:10, et = factor(c(rep('L',3), rep('S',7))))
> ar <- matrix(data = 80:89, nrow =5, ncol =2)
> obj <- rep(2, 12)
> lista <- list(df, ar, 2+3i, obj)
> lista
[[1]]
  num et
1    1  L
2    2  L
3    3  L
4    4  S
5    5  S
6    6  S
7    7  S
8    8  S
9    9  S
10  10  S

[[2]]
  [,1] [,2]
[1,]  80  85
[2,]  81  86
[3,]  82  87
[4,]  83  88
[5,]  84  89

[[3]]

```

```
[1] 2+3i
```

```
[[4]]
```

```
[1] 2 2 2 2 2 2 2 2 2 2 2 2
```

Como se observa, basta con poner argumentos a los elementos que contendrá la lista. En este caso se guardó un data frame, una matriz, un número complejo y un vector.

Funciones

Como ya se mencionó, R también nos brinda la posibilidad aplicar diversas metodologías. La implementación de funciones como herramienta para elaborar operaciones, de acuerdo a las necesidades del usuario es una característica importante con la que cuenta. El funcionamiento general de una función se muestra en la figura B.2

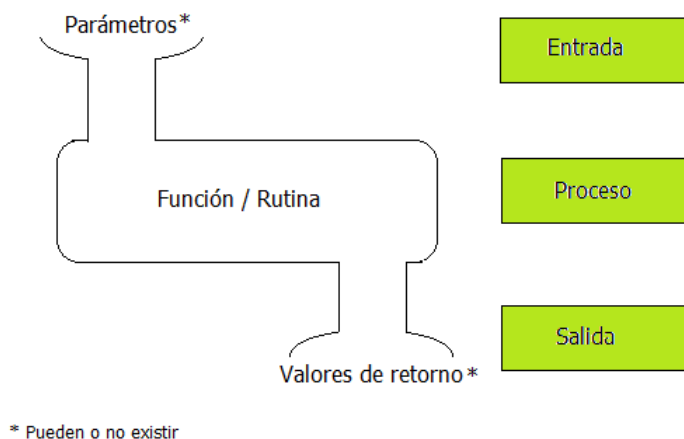


Figura B.2: Diagrama de actividad funcional

Para crear una función es necesario utilizar la función `function()`, la cual puede o no recibir valores/objetos, además de que se le debe asignar un nombre para aplicarla después. Las operaciones o procedimientos que realizará van dentro de las llaves, tal como se muestra en la siguiente sintaxis.

Función función [function()]

```
variable <- function(val1, val2, val3, ...){
  expresion(es)
}
```

Al igual que las demás funciones que son empleadas para crear objetos, los parámetros de éstas son separados por comas. El argumento `...` hace alusión a que la función puede recibir parámetros no especificados.

```

> misc <- function(x, y, ...) {
+   z1 <- 2*x + y
+   z2 <- x + 2*y
+   z3 <- 2*x + 2*y
+   z4 <- x/y
+   return(c(z1, z2, z3, z4))
+ }

> misc(1,1)
[1] 3 3 4 1

```

En el caso anterior, la función `misc` recibe dos valores, x y y , para así evaluar las funciones

- $z_1 = 2x + y$
- $z_2 = x + 2y$
- $z_3 = 2x + 2y$
- $z_4 = \frac{x}{y}$

y dar como valor de salida un vector con los resultados de las evaluaciones.

Conversión

Algunos de los tipos de datos y objetos antes mencionados admiten conversión a otros tipos. La forma de convertirlos es a través de funciones, algunos ejemplos son:

Conversión de real a entero [`as.integer()`]

```

> real <- 83.09875
> class(real)
[1] "numeric"

> entero <- as.integer(real)
> entero
[1] 83

> class(entero)
[1] "integer"

```

Conversión de entero a complejo [`as.complex()`]

```

> complejo <- as.complex(entero)
> complejo
[1] 83+0i

```

Conversión de complejo a caracter [`as.character()`]

```

> caracter <- as.character(complejo)
> caracter
[1] "83+0i"

```

B. PROGRAMACIÓN EN R

Conversión de real a booleano [`as.logical()`]

```
> booleano <- as.logical(real)
> booleano
[1] TRUE
```

Conversión de vector a factor [`as.factor()`]

```
> vector <- 1:6
> vector
[1] 1 2 3 4 5 6

> factor <- as.factor(vector)
> factor
[1] 1 2 3 4 5 6
Levels: 1 2 3 4 5 6
```

Conversión de factor a vector numérico [`as.numeric(as.character())`]

```
> vectorNumerico <- as.numeric(as.character(factor))
> vectorNumerico
[1] 1 2 3 4 5 6
```

En este caso es importante no aplicar directamente la función `as.numeric()`, debido a que el factor maneja etiquetas y aunque éstas sean números, R les da el valor de numérico por orden aparición en los niveles.

Conversión de matriz a data frame [`as.data.frame()`]

```
> matriz <- matrix(data = 80:89, nrow =5, ncol =2)
> matriz
  [,1] [,2]
[1,]  80  85
[2,]  81  86
[3,]  82  87
[4,]  83  88
[5,]  84  89

> dataframe <- as.data.frame(matriz)
> dataframe
  V1 V2
1 80 85
2 81 86
3 82 87
4 83 88
5 84 89
```

Es importante recordar que todos los elementos de una matriz son del mismo tipo, mientras que los data frames, admiten diferentes variables (vectores numéricos, vectores de caracteres, factores, etc.), por lo que debe tenerse cuidado en las conversiones.

Otras funciones de conversión son:

-
- `as.vector()` aplicada a matrices
 - `as.matrix()` aplicada a `data.frame`

Comprobación

A veces es necesario saber o comprobar el tipo de dato u objeto que es nuestra variable, por lo que con las funciones `is.numeric()`, `is.character()`, `is.na()`, `is.integer()`, `is.vector()`, `is.data.frame()` son una herramienta para saberlo. El argumento de estas funciones es el objeto a probar, y el valor que devuelven es un booleano (TRUE o FALSE).

Función `is.na()`

Esta función en especial es de gran utilidad, ya que nos ayuda a saber si algún vector tiene valores faltantes (NA).

```
> vct <- c(12, 14, NA, 16, 18, 20)
> is.na(vct)
[1] FALSE FALSE TRUE FALSE FALSE FALSE
```

El tercer elemento de `vct` es un tipo de dato NA, por lo que al “preguntarle” a `vct` si alguno de sus elementos es NA, el vector de booleanos que regresa muestra la tercera posición como TRUE, indicando la existencia y el lugar en el que se encuentra.

Acceso y selección

Los operadores de acceso y selección de elementos para los distintos objetos son `[]`, `[[[]]` y `$`.

Operador `[]`

Para el caso de vectores y factores, acceder al elemento que ocupa la posición ‘i’, basta con poner el nombre del objeto, seguido del número de posición o bien el nombre de esa posición.

```
> obj
[1] 2 2 2 2 2 2 2 2 2 2 2
> obj[3]
[1] 2
```

En este caso, el vector `obj`, en su posición 3, tiene el elemento 2. Si se requiere más de un elemento, en lugar de poner un número, se pone entre los corchetes un vector con las posiciones deseadas.

Para las matrices y data frames, la dimensión es 2, por lo que para acceder al k-ésimo renglón y a la l-ésima columna, se tienen que indicar los índices separados por una coma.

B. PROGRAMACIÓN EN R

```
> mat
      cols
rows  C.1 C.2 C.3
row1  9 -8  7
row2 -6  5 -4

> mat[1, 2]
[1] -8

> mat['row1', 'C.1']
[1] 9
```

Para tener acceso a sólo columnas o sólo renglones es necesario dejar un espacio, según sea el caso.

```
> mat[ , c(1,3)]
      cols
rows  C.1 C.3
row1  9  7
row2 -6 -4
```

Es posible aplicar operaciones lógicas dentro de los operadores de acceso y selección, y así obtener los elementos que obtiene como valor lógico TRUE.

Operador [[]]

Este operador es especialmente para listas. Dentro de los corchetes se pone el número de posición o el nombre entre comillas del objeto al que se quiere acceder y sólo permite un elemento a la vez.

```
> lista
[[1]]
 [,1] [,2]
[1,]  80  85
[2,]  81  86
[3,]  82  87
[4,]  83  88
[5,]  84  89

[[2]]
[1] 2+3i

[[3]]
[1] 2 2 2 2 2 2 2 2 2 2 2 2

> lista[[2]]
[1] 2+3i
```

Operador \$

El signo de pesos \$ nos ayuda a seleccionar elementos de una lista o un data frame, a partir del nombre.

```

> df
  Numeros Etiqueta
1         1         L
2         2         L
3         3         L
4         4         S
5         5         S
6         6         S
7         7         S
8         8         S
9         9         S
10        10         S

> df\$Etiqueta
[1] L L L S S S S S S S
Levels: L S

```

Estructuras de control

En los lenguajes de programación, las estructuras de control nos permiten controlar el flujo de ejecución de las instrucciones de un programa, a través de condiciones o ejecuciones repetitivas de ciertas sentencias.

If-else

La estructura *if* o *if-else* en combinación, permiten ejecutar una sentencia en función de una condición lógica.

```

if( condición/expresión lógica ){
  sentencia
}

```

```

if( condición/expresión lógica ){
  sentencia
} else {
  sentencia
}

```

```

> x <- runif(1, 0, 100) #número aleatorio entre 0 y 100
> if(x > 50) {
+   y <- TRUE
+ } else {
+   y <- FALSE
+ }
> x; y
[1] 67.57339
[1] TRUE

```

En este ejemplo la condición lógica a cumplir para que *y* sea TRUE, es que *x* sea mayor a 50, en caso de que no se cumpla, *y* es FALSE. La última línea separa dos sentencias con un punto y coma (;), para que así puedan ser ejecutadas en una línea.

For

El bucle *for* es una estructura de repetición que ejecuta un conjunto de comandos mientras iteran a través de los valores en un vector.

```
for( variable in vector ) {  
  ...  
  sentencia(s)  
}  
  
> fibonacci <- c(1, 1, 0, 0, 0, 0, 0, 0)  
> for (i in 3:8) {  
+   fibonacci[i] <- fibonacci[i-1] + fibonacci[i-2]  
+ }  
> fibonacci  
[1] 1 1 2 3 5 8 13 21
```

While

El bucle *while* evalúa una expresión o función, siempre y cuando una condición sea verdadera.

```
while( condición ) {  
  ...  
  sentencia(s)  
}  
  
> x<-0  
> while (x < 10)  
+ {  
+   x<- x+4  
+   print (x)  
+   if (x==8)  
+   {  
+     break  
+   }  
+ }  
[1] 4  
[1] 8
```

La palabra reservada **break**, es usada para detener y salir de un bucle.

Repeat

El bucle **repeat** a diferencia de los anteriores, este se ejecuta sin condición, es decir, infinitamente. La forma de parar el bucle es con **break** o **stop()**.

```
repeat( condición ) {  
  ...  
  sentencia(s)  
}  
  
> sum <- 1  
> repeat{  
+   sum <- sum + 2  
+   print(sum)  
+   if (sum >= 11)  
+     stop ("Los valores deben ser < 12")  
+ }  
[1] 3  
[1] 5  
[1] 7  
[1] 9  
[1] 11
```

Error: Los valores deben ser < 12

La función `stop()`, muestra el mensaje que recibe como argumento y detiene la ejecución en caso de no cumplirse la especificación. Esta función es útil a manera de excepción. La función `print()` nos ayuda a mostrar los valores de las variables.

Lectura y escritura de datos

El enfoque que tiene R hacia el análisis de datos, hace que la importación y exportación sean tareas necesarias. Existen muchas formas de llevarlas a cabo, dependiendo del formato de los archivos.

Antes de poder realizar estas tareas, es necesario que el directorio donde se ubican los archivos para importar o bien donde se desea escribir los nuevos archivos, sea conocido. Para saber el directorio de trabajo que está reconociendo R se utiliza la función `getwd()`. Para cambiar el directorio de trabajo es con la función `setwd()`, recibiendo como argumento la ruta del directorio como una cadena de caracteres, es decir, `setwd('C:/...')`. Además de esto, es importante saber la codificación de los archivos que se quieren leer o bien escribirlo con una codificación deseada, y así tener una adecuada representación de los caracteres.

Lectura

Algunas de las formas más usuales para importar datos son con las siguientes funciones: `read.table()` para archivos con extensión *.txt*, `read.csv()` para *.csv* delimitado por comas (,) y `read.csv2()` para *.csv* delimitado por punto y coma (;). Además de estos formatos de almacenamiento de datos, es posible leer archivos de otras extensiones como *.dbf*, *.dta*, *.spss*, por mencionar algunos, y cuya implementación es referida al paquete `foreign`.

B. PROGRAMACIÓN EN R

```
nombre1 <- read.table("NombreArchivo.txt", header = TRUE)
nombre2 <- read.csv("NombreArchivo.csv")
nombre3 <- read.csv("NombreArchivo.csv", header = FALSE)
```

En los ejemplos se muestra cómo emplear las funciones mencionadas. El objeto que reconoce R es de tipo data frame, donde `header` hace alusión a si tiene o no nombres las variables/columnas. Sin embargo, es necesario consultar la documentación en caso de que se quiera especificar otro tipo de cosas como la codificación de los missing values, nombres de variables en caso de no tener, el caracter para representar el punto decimal, por mencionar algunas.

Una manera de consultar *ayuda* para cualquier función de paquetes instalados es con `help()` y de argumento la función a consultar entre comillas.

Escritura

Dependiendo del tipo de archivo que se quiere generar, es la función que se emplea. Para archivos con extensión **.txt* la función es `write.table()`, para **.csv* `write.csv()` o `write.csv2()` según como se quieran delimitar. Al igual que la lectura, es posible escribir en otros formatos con el apoyo del paquete antes mencionado.

```
write.table(base1, file = "NombreArchivo.txt")
write.csv(base2, file = "NombreArchivo.csv")
write.csv2(base3, file = "NombreArchivo.csv")
```

Tanto `base1`, `base2` y `base3`, son objetos tipo data frame o matriz.

Apéndice C

Código fuente del paquete *fertilityRates*

Estructura de los datos

Código C.1: Código para procesar los datos

```
1
2 # Los archivos TMMujer1.dbf y TFec_Hemb.dbf estan disponibles en:
3 # https://www.inegi.org.mx/programas/enadid/2014/default.html#Microdatos
4
5 library(foreign)
6 library(dplyr)
7
8
9
10 # leer los archivos dbf
11 temporal <- tempfile()
12 download.file("http://www.inegi.org.mx/contenidos/programas/enadid/2014/microdatos/base_datos_
13   enadid14_dbf.zip",temporal)
14 files = unzip(temporal, list=TRUE)$Name
15 unzip(temporal, files=files[grepl("dbf",files)])
16 muj_14 <- data.frame(read.dbf("TMMujer1.dbf"))
17 hist_14 <- data.frame(read.dbf('TFec_Hemb.dbf'))
18
19 # borrar los registros que son abortos o mortinatos
20 hist_14 <- hist_14[which(is.na(hist_14[, 'P5_20'])==T & is.na(hist_14[, "P5_24"])==T), ]
21
22
23 # modificar los tipos de datos
24 muj_14 <- transform(muj_14, ID_1 = LLAVE_MUJ, FACTOR = as.numeric(as.character(FAC_PER)),
25   UPM = as.numeric(as.character(UPM_DIS)), ESTRATO = as.numeric(as.character(EST_
26   DIS)),
27   FEC_MUJ_M = as.numeric(as.character(P5_1_1)), FEC_MUJ_A = as.numeric(as.
28   character(P5_1_2)),
29   EDAD_M = as.numeric(as.character(P5_2)))
30 hist_14 <- transform(hist_14, ID_2 = LLAVE_MUJ, FEC_HIJ_M = as.numeric(as.character(P5_17_1)),
31   FEC_HIJ_A = as.numeric(as.character(P5_17_2)))
32
33 # seleccionar las variables necesarias
34 muj_14 <- muj_14[,c('ID_1', "FACTOR", 'UPM', 'ESTRATO', 'FEC_MUJ_M', 'FEC_MUJ_A', 'EDAD_M')]
35 hist_14 <- hist_14[,c('ID_2', 'FEC_HIJ_M', 'FEC_HIJ_A')]
36
```

C. CÓDIGO FUENTE DEL PAQUETE *FERTILITYRATES*

```
37 # eliminar los registros que no especifican año de nacimiento
38 table(muj_14$FEC_MUJ_A); table(is.na(muj_14$FEC_MUJ_A))
39 table(hist_14$FEC_HIJ_A); table(is.na(hist_14$FEC_HIJ_A)) # 1031 (9999 es valor no especificado)
40
41 hist_14 <- hist_14[which(hist_14$FEC_HIJ_A!="9999"), ] # eliminar valores
42
43
44 # prorratear de manera uniforme los registros que no se especifico mes de nacimiento
45 table(muj_14$FEC_MUJ_M); table(is.na(muj_14$FEC_MUJ_M)) #159 (99 es valor no especificado)
46 table(hist_14$FEC_HIJ_M); table(is.na(hist_14$FEC_HIJ_M)) # 2014 (99 es valor no especificado)
47
48 hist_14$FEC_HIJ_M[which(hist_14$FEC_HIJ_M == '99')] <- round(runif(2014, 1, 12), 0) # uniform
49 muj_14$FEC_MUJ_M[which(muj_14$FEC_MUJ_M=='99')] <- round(runif(159, 1, 12), 0) # uniform
50
51
52 # establecer el mes de entrevista
53 muj_14$ENTREV_M <- as.numeric(09)
54
55
56
57 # union de los datos y revisar duplicados
58 enadid_2014 <- data.frame()
59 enadid_2014 <- full_join(muj_14, hist_14, by = c('ID_1' = 'ID_2')) # where there are not matching
    values, returns NA
60 enadid_2014$MUJER <- !duplicated(enadid_2014$ID_1)
61
62
63 # escribir la base de datos procesada
64 save(enadid_1994, file = 'enadid_1992.RData', compress = "xz")
```

frts_intvw. Función para obtener estimaciones en el año del levantamiento

Código C.2: Función constructora de objetos de clase *frts_intvw*

```
1 frts_intvw<- function(m.intvw, y.intvw, m.wmn,
2                       y.wmn, age.wmn, m.child, y.child,
3                       wmn.dummy, id.wmn, ids,
4                       strata, weights, data){
5   if(is.data.frame(data)){
6     if (requireNamespace("survey", quietly = TRUE)) {
7
8       attach(data)
9       db <- data.frame(m.intvw, y.intvw,
10                      m.wmn, y.wmn,
11                      m.child, y.child,
12                      wmn.dummy, id.wmn)
13
14       db$child.dummy <- ifelse(is.na(y.child)==FALSE, 1, 0)
15
16       intvw.age <- ifelse(m.wmn >= m.intvw, age.wmn, age.wmn-1)
17       age.wmn <- NULL
18
19       age.wmn <- intvw.age
20       age2.wmn <- intvw.age +1
21
```

```

22 db$age.wmn <- age.wmn
23 db$age2.wmn <- age2.wmn
24 db$intvw.age <- intvw.age
25
26 db$expo1 <- 0
27 db$expo2 <- 0
28 db$expo3 <- 0
29
30 for (i in 1:dim(db)[1]) {
31
32   if(db$wmn.dummy[i]==TRUE){
33     if(db$m.intvw[i] > db$m.wmn[i]){
34       db$expo1[i] <- ((db$m.wmn[i]-0.5)/12)
35       db$expo2[i] <- ((db$m.intvw[i]-db$m.wmn[i]-0.5)/12)
36     }else{
37       db$expo3[i] <- ((db$m.intvw[i]-1)/12)
38     }
39   }
40
41 }
42 db$m.intvw <- NULL
43
44
45 auxiliary <- function(age, age.wmn, age2.wmn, exposition1, exposition2, exposition3){
46   expo1 <- rep(0,length(exposition1))
47   aux<-grep(age, age.wmn, value=FALSE)
48   expo1[aux] <- exposition1[aux]
49   aux2<-grep(age, age2.wmn, value=FALSE)
50   expo1[aux2] <- exposition2[aux2]
51   expo2 <- rep(0,length(exposition2))
52   aux3<-grep(age, age.wmn, value=FALSE)
53   expo2[aux3] <- exposition3[aux3]
54   expo <- expo1 + expo2
55   return(expo)
56 }
57
58 nom <- sprintf("exposition_%s", 15:49)
59 data <- data.frame(matrix(ncol =(49- 15)+1, nrow = length(db$age.wmn)))
60 colnames(data) <- nom
61 for(j in 1:((49 - 15)+1)){
62   data[,j]<- auxiliary(j+15-1, db$age.wmn, db$age2.wmn, db$expo1,
63     db$expo2, db$expo3)
64 }
65
66
67 db$expo1 <- NULL
68 db$expo2 <- NULL
69 db$expo3 <- NULL
70
71 db <- cbind(db,data)
72
73 estimate_age <- function(m.child, y.child, m.wmn, intvw.age, y.intvw){
74   age <- NULL
75   for(i in 1:length(m.child)){
76     if(is.na(m.child[i]) == F){
77       if(m.child[i] > m.wmn[i] | m.child[i] < m.wmn[i]){
78         if(m.child[i] > m.wmn[i] ){
79           age = c(age, intvw.age[i]-(y.intvw-y.child[i])+1)
80         }
81         else{
82           age = c(age, intvw.age[i]-(y.intvw-y.child[i]))
83         }
84       }
85     }
86   }
87 }

```

C. CÓDIGO FUENTE DEL PAQUETE *FERTILITYRATES*

```
84     }
85     else{
86         age = c(age, intvw.age[i]-(y.intvw-y.child[i])+stats::rbinom(1,1,0.5))
87     }
88 }
89 else{age = c(age,0)}
90 }
91 return(age)
92 }
93
94
95 db$age.mother <- estimate_age(db$m.child, db$y.child, db$m.wmn, db$intvw.age, y.intvw)
96 db$age.group <- cut(db$age.mother, c(-1, 14, 19, 24, 29, 34, 39, 44, 49, 60))
97 levels(db$age.group) <- c('0_14', '15_19', '20_24', '25_29', '30_34', '35_39', '40_44', '
98     45_49', '50_60')
99
100 db$expo15_19 <- apply(db[,paste('exposition_',15:19, sep = '')],1,sum)
101 db$expo20_24 <- apply(db[,paste('exposition_',20:24, sep = '')],1,sum)
102 db$expo25_29 <- apply(db[,paste('exposition_',25:29, sep = '')],1,sum)
103 db$expo30_34 <- apply(db[,paste('exposition_',30:34, sep = '')],1,sum)
104 db$expo35_39 <- apply(db[,paste('exposition_',35:39, sep = '')],1,sum)
105 db$expo40_44 <- apply(db[,paste('exposition_',40:44, sep = '')],1,sum)
106 db$expo45_49 <- apply(db[,paste('exposition_',45:49, sep = '')],1,sum)
107
108 db$y.wmn <- NULL
109 db$m.child <- NULL
110
111 db$ids <- ids
112 db$strata <- strata
113 db$weights <- weights
114
115 detach(data)
116 attach(db)
117
118
119 ds <- survey::svydesign(id = ~ids,
120     strata = ~strata,
121     weights = ~weights,
122     data = db, nest=TRUE) #JUST NEST
123
124 value <- list(df = db, ds = ds)
125
126 attr(value, 'class') <- 'frts_intvw'
127 value
128
129
130 } else {
131     stop("Library 'survey' must be installed")
132 }
133 } else {
134     stop('There is not some data.frame')
135 }
136 }
137
138 }
```

Código C.3: Método que proporciona los objetos de la clase *frts_intvw*

```
1 summary.frts_intvw <- function(object, level, ...){
```

```

2  if (requireNamespace("survey", quietly = TRUE)) {
3
4  val <- c()
5  i_ci <- c()
6  u_ci <- c()
7  variance <- c()
8
9  for (i in 15:49) {
10   temp <- paste('exposition_',as.character(i),sep = '')
11   rate <- survey::svyratio(~ child.dummy * (y.child == y.intvw & age.mother == i),
12     ~ object$df[,grep(temp, names(object$df), value=TRUE)], object$ds, na.rm=TRUE)
13   ci <- stats::confint(rate, level = level, df=survey::degf(object$ds))
14   val <- c(val, rate$ratio[2])
15   i_ci <- c(i_ci, ci[2])
16   u_ci <- c(u_ci, ci[4])
17   variance <- c(variance,rate$var[2])
18 }
19
20 as_fr <- data.frame(round(val, 3), round(i_ci, 3), round(u_ci, 3), variance)
21 names(as_fr) <- c('as_fr_s','l_ci','u_ci','var')
22 row.names(as_fr) <- 15:49
23
24
25 rate <- apply(as_fr,2,sum)[1]
26 se <- sqrt(apply(as_fr,2,sum)[4])
27 a <- (1 - level)
28 t_fr <- round(c(rate, rate + stats::qt(a/2, df=survey::degf(object$ds)) * se, rate + stats::qt
29   (1-a/2, df=survey::degf(object$ds)) * se), 3)
30 names(t_fr) <- c('t_fr','l_ci','u_ci')
31 t_fr <- as.data.frame(t_fr)
32
33
34 val2 <- c()
35 i_ci2 <- c()
36 u_ci2 <- c()
37
38 for (i in levels(object$df$age.group)[-c(1,9)]) {
39   temp <- paste('expo', i, sep = '')
40   rate <- survey::svyratio(~ child.dummy * (y.child == y.intvw & age.group == i),
41     ~ object$df[,grep(temp, names(object$df), value=TRUE)], object$ds, na.rm=TRUE)
42   ci <- stats::confint(rate, level = level, df=survey::degf(object$ds))
43   val2 <- c(val2, rate$ratio[2])
44   i_ci2 <- c(i_ci2, ci[2])
45   u_ci2 <- c(u_ci2, ci[4])
46
47 }
48
49 ag_fr <- data.frame(round(val2, 3), round(i_ci2, 3), round(u_ci2, 3))
50 names(ag_fr) <- c('as_fr_g','l_ci','u_ci')
51 row.names(ag_fr) <- c('15-19', '20-24', '25-29', '30-34', '35-39', '40-44', '45-49')
52
53 as_fr$var <- NULL
54
55
56
57
58 value <- list(as_fr_s = as_fr, t_fr = t_fr, as_fr_g = ag_fr)
59 value
60 } else {
61   stop("Library 'survey' must be installed")
62 }

```

63 }

frts_yrly. Función para obtener estimaciones anuales en los cuatro años anteriores al levantamiento

Código C.4: Función constructora de objetos de clase *frts_yrly*

```
1
2 frts_yrly<- function(m.intvw, y.intvw, y.ref, m.wmn, y.wmn, age.wmn,
3                     m.child, y.child,
4                     wmn.dummy, id.wmn, ids,
5                     strata, weights, data){
6   if(is.data.frame(data)){
7     if (requireNamespace("survey", quietly = TRUE)) {
8
9       attach(data)
10      db <- data.frame(m.wmn, y.wmn,
11                      m.child, y.child,
12                      wmn.dummy, id.wmn)
13
14      db$child.dummy <- ifelse(is.na(y.child)==FALSE, 1, 0)
15
16
17      intvw.age <- ifelse(m.wmn >= m.intvw, age.wmn, age.wmn-1)
18      age.wmn <- NULL
19
20      age.wmn <- intvw.age - (y.intvw-y.ref)
21      age2.wmn <- age.wmn + 1 #age2.wmn-(y.intvw-y.ref-1)
22
23
24      db$age.wmn <- age.wmn
25      db$age2.wmn <- age2.wmn
26      db$intvw.age <- intvw.age
27
28      db$expo1 <- 0
29      db$expo2 <- 0
30
31
32      for (i in 1:dim(db)[1]) {
33
34        if(db$wmn.dummy[i]==TRUE){
35          db$expo1[i] <- ((db$m.wmn[i]-0.5)/12)
36          db$expo2[i] <- (1-db$expo1[i])
37        }
38
39      }
40
41
42
43      auxiliary <- function(age, age.wmn, age2.wmn, exposition1, exposition2){
44        expo1 <- rep(0,length(exposition1))
45        aux<-grep(age, age.wmn, value=FALSE)
46        expo1[aux] <- exposition1[aux]
47        aux2<-grep(age, age2.wmn, value=FALSE)
48        expo1[aux2] <- exposition2[aux2]
49
```

```

50     return(expo1)
51 }
52
53
54
55 nom <- sprintf("exposition_%s", min(db$age.wmn):max(db$age2.wmn))
56 data <- data.frame(matrix(ncol =(max(db$age2.wmn)- min(db$age.wmn))+1, nrow = length(db$age
57 .wmn)))
58 colnames(data) <- nom
59 for(j in 1:(max(db$age2.wmn)- min(db$age.wmn))+1){
60     data[,j]<- auxiliary(j+min(db$age.wmn)-1,db$age.wmn,db$age2.wmn,db$expo1,db$expo2)
61 }
62
63 db$expo1 <- NULL
64 db$expo2 <- NULL
65
66 db <- cbind(db,data)
67
68 estimate_age <- function(m.child, y.child, m.wmn, intvw.age, y.intvw){
69     age <- NULL
70     for(i in 1:length(m.child)){
71         if(is.na(m.child[i]) == F){
72             if(m.child[i] > m.wmn[i] | m.child[i] < m.wmn[i]){
73                 if(m.child[i] > m.wmn[i] ){
74                     age = c(age, intvw.age[i]-(y.intvw-y.child[i])+1)
75                 }
76                 else{
77                     age = c(age, intvw.age[i]-(y.intvw-y.child[i]))
78                 }
79             }
80             else{
81                 age = c(age, intvw.age[i]-(y.intvw-y.child[i])+stats::rbinom(1,1,0.5))
82             }
83         }
84         else{age = c(age,0)}
85     }
86     return(age)
87 }
88
89
90 db$age.mother <- estimate_age(db$m.child, db$y.child, db$m.wmn, db$intvw.age, y.intvw)
91 db$age.group <- cut(db$age.mother, c(-1, 14, 19, 24, 29, 34, 39, 44, 49, 60))
92 levels(db$age.group) <- c('0_14', '15_19', '20_24', '25_29', '30_34', '35_39', '40_44', '
93 45_49', '50_60')
94
95 db$expo15_19 <- apply(db[,paste('exposition_',15:19, sep = '')],1,sum)
96 db$expo20_24 <- apply(db[,paste('exposition_',20:24, sep = '')],1,sum)
97 db$expo25_29 <- apply(db[,paste('exposition_',25:29, sep = '')],1,sum)
98 db$expo30_34 <- apply(db[,paste('exposition_',30:34, sep = '')],1,sum)
99 db$expo35_39 <- apply(db[,paste('exposition_',35:39, sep = '')],1,sum)
100 db$expo40_44 <- apply(db[,paste('exposition_',40:44, sep = '')],1,sum)
101 db$expo45_49 <- apply(db[,paste('exposition_',45:49, sep = '')],1,sum)
102
103 db$y.wmn <- NULL
104 db$m.child <- NULL
105
106 db$ids <- ids
107 db$strata <- strata
108 db$weights <- weights
109

```

C. CÓDIGO FUENTE DEL PAQUETE *FERTILITYRATES*

```
110 detach(data)
111 attach(db)
112
113
114 ds <- survey::svydesign(id = ~ids,
115                       strata = ~strata,
116                       weights = ~weights,
117                       data = db, nest=TRUE)
118
119 value <- list(df = db, ds = ds, year = y.ref)
120
121
122 attr(value, 'class') <- 'frts_yrly'
123 value
124
125
126
127 } else {
128   stop("Library 'survey' must be installed")
129 }
130 } else {
131   stop('There is not some data.frame')
132
133 }
134
135 }
```

Código C.5: Método que proporciona los objetos de la clase *frts_yrly*

```
1
2 summary.frts_yrly <- function(object, level, ...){
3   if (requireNamespace("survey", quietly = TRUE)) {
4
5
6     val <- c()
7     i_ci <- c()
8     u_ci <- c()
9     variance <- c()
10
11    for (i in 15:49) {
12      temp <- paste('exposition_',as.character(i),sep = '')
13      rate <- survey::svyratio(~ child.dummy * (y.child == object$year & age.mother == i),
14                             ~ object$df[,grep(temp, names(object$df), value=TRUE)], object$ds, na.rm=TRUE)
15      ci <- stats::confint(rate, level = level, df=survey::degf(object$ds))
16      val <- c(val, rate$ratio[2])
17      i_ci <- c(i_ci, ci[2])
18      u_ci <- c(u_ci, ci[4])
19      variance <- c(variance,rate$var[2])
20
21    }
22    as_fr <- data.frame(round(val, 3), round(i_ci, 3), round(u_ci, 3), variance)
23    names(as_fr) <- c('as_fr_s','l_ci','u_ci','var')
24    row.names(as_fr) <- 15:49
25
26
27    rate <- apply(as_fr,2,sum)[1]
28    se <- sqrt(apply(as_fr,2,sum)[4])
29    a <- (1 - level)
30    t_fr <- round(c(rate, rate + stats::qt(a/2, df=survey::degf(object$ds)) * se, rate + stats::qt
      (1-a/2, df=survey::degf(object$ds)) * se), 3)
```

```

31 names(t_fr) <- c('t_fr', 'l_ci', 'u_ci')
32 t_fr <- as.data.frame(t_fr)
33
34
35 val2 <- c()
36 i_ci2 <- c()
37 u_ci2 <- c()
38
39
40 for (i in levels(object$df$age.group)[-c(1,9)]) {
41   temp <- paste('expo', i, sep = '')
42   rate <- survey::svyratio(~ child.dummy * (y.child == object$year & age.group == i),
43     ~ object$df[,grep(temp, names(object$df), value=TRUE)], object$ds, na.rm=TRUE)
44   ci <- stats::confint(rate, level = level, df=survey::degf(object$ds))
45   val2 <- c(val2, rate$ratio[2])
46   i_ci2 <- c(i_ci2, ci[2])
47   u_ci2 <- c(u_ci2, ci[4])
48
49
50 }
51 ag_fr <- data.frame(round(val2, 3), round(i_ci2, 3), round(u_ci2, 3))
52 names(ag_fr) <- c('as_fr_g', 'l_ci', 'u_ci')
53 row.names(ag_fr) <- c('15-19', '20-24', '25-29', '30-34', '35-39', '40-44', '45-49')
54
55
56 as_fr$var <- NULL
57
58
59
60
61 value <- list(as_fr_s = as_fr, t_fr = t_fr, as_fr_g = ag_fr)
62 value
63 } else {
64   stop("Library 'survey' must be installed")
65 }
66
67 }

```

frts_3yrs. Función para obtener estimaciones trianuales dentro de los cuatro años anteriores al levantamiento

Código C.6: Función constructora de objetos de clase *frts_3yrs*

```

1
2 frts_3yrs<- function(m.intvw, y.intvw, y.first, y.second, y.third, m.wmn,
3   y.wmn, age.wmn, m.child, y.child,
4   wmn.dummy, id.wmn, ids,
5   strata, weights, data){
6
7   if(is.data.frame(data)){
8     if (requireNamespace("survey", quietly = TRUE)) {
9
10      attach(data)
11      db <- data.frame(m.wmn, y.wmn,
12        m.child, y.child,
13        wmn.dummy, id.wmn)

```

C. CÓDIGO FUENTE DEL PAQUETE *FERTILITYRATES*

```
14
15 db$child.dummy <- ifelse(is.na(y.child)==FALSE, 1, 0)
16
17
18 intvw.age <- ifelse(m.wmn >= m.intvw, age.wmn, age.wmn-1)
19 age.wmn <- NULL
20
21
22
23 age.wmn <- intvw.age -(y.intvw - y.first)
24 age2.wmn <- age.wmn + 1
25
26 age3.wmn <- intvw.age -(y.intvw - y.second)
27 age4.wmn <- age3.wmn + 1
28
29 age5.wmn <- intvw.age -(y.intvw - y.third)
30 age6.wmn <- age5.wmn + 1
31
32 db$age.wmn <- age.wmn
33 db$age2.wmn <- age2.wmn
34 db$age3.wmn <- age3.wmn
35 db$age4.wmn <- age4.wmn
36 db$age5.wmn <- age5.wmn
37 db$age6.wmn <- age6.wmn
38 db$intvw.age <- intvw.age
39
40
41 db$expo1 <- 0
42 db$expo2 <- 0
43
44
45 for (i in 1:dim(db)[1]) {
46
47   if(db$wmn.dummy[i]==TRUE){
48     db$expo1[i] <- ((db$m.wmn[i]-0.5)/12)
49     db$expo2[i] <- (1-db$expo1[i])
50   }
51
52 }
53
54
55
56 auxiliary <- function(age, age.wmn, age2.wmn, exposition1, exposition2){
57   expo1 <- rep(0,length(exposition1))
58   aux<-grep(age, age.wmn, value=FALSE)
59   expo1[aux] <- exposition1[aux]
60   aux2<-grep(age, age2.wmn, value=FALSE)
61   expo1[aux2] <- exposition2[aux2]
62
63   return(expo1)
64 }
65
66
67
68 nom <- sprintf("exposition_%s", min(db$age.wmn):max(db$age2.wmn))
69 data <- data.frame(matrix(ncol =(max(db$age2.wmn)- min(db$age.wmn))+1, nrow = length(db$age
70   .wmn)))
71 colnames(data) <- nom
72 for(j in 1:(max(db$age2.wmn)- min(db$age.wmn)+1)){
73   data[,j]<- auxiliary(j+min(db$age.wmn)-1,db$age.wmn,db$age2.wmn,db$expo1,db$expo2)
74 }
```

```

75 temp <- rep(0, nrow(data))
76 data2 <- cbind(data[, -1], temp)
77 data3 <- cbind(data2[, -1], temp)
78 data <- data + data2 + data3
79
80
81 db$expo1 <- NULL
82 db$expo2 <- NULL
83
84
85 db <- cbind(db, data)
86
87 estimate_age <- function(m.child, y.child, m.wmn, intvw.age, y.intvw){
88   age <- NULL
89   for(i in 1:length(m.child)){
90     if(is.na(m.child[i]) == F){
91       if(m.child[i] > m.wmn[i] | m.child[i] < m.wmn[i]){
92         if(m.child[i] > m.wmn[i] ){
93           age = c(age, intvw.age[i]-(y.intvw-y.child[i])+1)
94         }
95       } else{
96         age = c(age, intvw.age[i]-(y.intvw-y.child[i]))
97       }
98     }
99     else{
100      age = c(age, intvw.age[i]-(y.intvw-y.child[i])+stats::rbinom(1,1,0.5))
101    }
102  }
103  else{age = c(age,0)}
104 }
105 return(age)
106 }
107
108
109 db$age.mother <- estimate_age(db$m.child, db$y.child, db$m.wmn, db$intvw.age, y.intvw)
110 db$age.group <- cut(db$age.mother, c(-1, 14, 19, 24, 29, 34, 39, 44, 49, 60))
111 levels(db$age.group) <- c('0_14', '15_19', '20_24', '25_29', '30_34', '35_39', '40_44', '
112 45_49', '50_60')
113
114 db$expo15_19 <- apply(db[,paste('exposition_', 15:19, sep = ' ')], 1, sum)
115 db$expo20_24 <- apply(db[,paste('exposition_', 20:24, sep = ' ')], 1, sum)
116 db$expo25_29 <- apply(db[,paste('exposition_', 25:29, sep = ' ')], 1, sum)
117 db$expo30_34 <- apply(db[,paste('exposition_', 30:34, sep = ' ')], 1, sum)
118 db$expo35_39 <- apply(db[,paste('exposition_', 35:39, sep = ' ')], 1, sum)
119 db$expo40_44 <- apply(db[,paste('exposition_', 40:44, sep = ' ')], 1, sum)
120 db$expo45_49 <- apply(db[,paste('exposition_', 45:49, sep = ' ')], 1, sum)
121
122 db$y.wmn <- NULL
123 db$m.child <- NULL
124
125 db$ids <- ids
126 db$strata <- strata
127 db$weights <- weights
128
129 detach(data)
130 attach(db)
131
132
133 ds <- survey::svydesign(id = ~ids,
134   strata = ~strata,
135   weights = ~weights,

```

C. CÓDIGO FUENTE DEL PAQUETE *FERTILITYRATES*

```
136         data = db, nest=TRUE)
137
138
139     period <- c(y.first, y.second, y.third)
140
141     value <- list(df = db, ds = ds, period = period)
142
143     attr(value, 'class') <- 'frts_3yrs'
144     value
145
146
147
148   } else {
149     stop("Library 'survey' must be installed")
150   }
151 } else {
152   stop('There is not some data.frame')
153 }
154 }
155 }
156 }
```

Código C.7: Método que proporciona los objetos de la clase *frts_3yrs*

```
1
2 summary.frts_3yrs <- function(object, level, ...){
3   if (requireNamespace("survey", quietly = TRUE)) {
4
5
6     val <- c()
7     i_ci <- c()
8     u_ci <- c()
9     variance <- c()
10
11   for (i in 15:49) {
12     temp <- paste('exposition_',as.character(i),sep = '')
13     rate <- survey::svyratio(~ child.dummy * ((y.child == object$period[1] | y.child == object$
14       period[2] | y.child == object$period[3])
15       & age.mother == i),
16       ~ object$df[,grep(temp, names(object$df), value=TRUE)], object$ds, na.rm=TRUE)
17     ci <- stats::confint(rate, level = level, df=survey::degf(object$ds))
18     val <- c(val, rate$ratio[2])
19     i_ci <- c(i_ci, ci[2])
20     u_ci <- c(u_ci, ci[4])
21     variance <- c(variance,rate$var[2])
22   }
23   as_fr <- data.frame(round(val, 3), round(i_ci, 3), round(u_ci, 3), variance)
24   names(as_fr) <- c('as_fr_s', 'l_ci', 'u_ci', 'var')
25   row.names(as_fr) <- 15:49
26
27
28   rate <- apply(as_fr,2,sum)[1]
29   se <- sqrt(apply(as_fr,2,sum)[4])
30   a <- (1 - level)
31   t_fr <- round(c(rate, rate + stats::qt(a/2, df=survey::degf(object$ds)) * se, rate + stats::qt
32     (1-a/2, df=survey::degf(object$ds)) * se), 3)
33   names(t_fr) <- c('t_fr', 'l_ci', 'u_ci')
34   t_fr <- as.data.frame(t_fr)
```

```

35
36 val2 <- c()
37 i_ci2 <- c()
38 u_ci2 <- c()
39
40
41 for (i in levels(object$df$age.group)[-c(1,9)]) {
42   temp <- paste('expo', i, sep = '')
43   rate <- survey::svyratio(~ child.dummy * ((y.child == object$period[1] | y.child == object$
44     period[2] | y.child == object$period[3])
45     & age.group == i),
46     ~ object$df[,grep(temp, names(object$df), value=TRUE)], object$ds, na.rm=TRUE)
47   ci <- stats::confint(rate, level = level, df=survey::degf(object$ds))
48   val2 <- c(val2, rate$ratio[2])
49   i_ci2 <- c(i_ci2, ci[2])
50   u_ci2 <- c(u_ci2, ci[4])
51
52 }
53 ag_fr <- data.frame(round(val2, 3), round(i_ci2, 3), round(u_ci2, 3))
54 names(ag_fr) <- c('as_fr_g', 'l_ci', 'u_ci')
55 row.names(ag_fr) <- c('15-19', '20-24', '25-29', '30-34', '35-39', '40-44', '45-49')
56
57
58 as_fr$var <- NULL
59
60
61 value <- list(as_fr_s = as_fr, t_fr = t_fr, as_fr_g = ag_fr)
62 value
63 } else {
64   stop("Library 'survey' must be installed")
65 }
66
67 }

```

frts_15yrs. Función para obtener estimaciones anuales de los quince años anteriores al levantamiento

Código C.8: Función constructora de objetos de clase *frts_15yrs*

```

1
2 frts_15yrs<- function(m.intvw, y.intvw, m.wmn,
3   y.wmn, age.wmn, m.child, y.child,
4   wmn.dummy, id.wmn,
5   weights, data){
6   if(is.data.frame(data)){
7
8     attach(data)
9     db <- data.frame(m.wmn, y.wmn,
10      m.child, y.child,
11      wmn.dummy, id.wmn, weights)
12
13     db$child.dummy <- ifelse(is.na(y.child)==FALSE, 1, 0)
14
15     intvw.age <- ifelse(m.wmn >= m.intvw, age.wmn, age.wmn-1)
16     age.wmn <- NULL

```

C. CÓDIGO FUENTE DEL PAQUETE *FERTILITYRATES*

```
17
18
19 age.wmn <- intvw.age - 1
20 age2.wmn <- age.wmn + 1
21
22
23 db$age.wmn <- age.wmn
24 db$age2.wmn <- age2.wmn
25 db$intvw.age <- intvw.age
26 db$expo1 <- 0
27 db$expo2 <- 0
28
29 for (i in 1:dim(db)[1]) {
30
31   if(db$wmn.dummy[i]==TRUE){
32     db$expo1[i] <- ((db$m.wmn[i]-0.5)/12)
33     db$expo2[i] <- (1-db$expo1[i])
34   }
35
36 }
37
38 db$expo1 <- db$expo1 * db$weights
39 db$expo2 <- db$expo2 * db$weights
40
41
42 auxiliary <- function(age, age.wmn, age2.wmn, exposition1, exposition2){
43   aux <- grep(age, age.wmn, value=FALSE)
44   s1 <- sum(exposition1[aux])
45   aux2 <- grep(age, age2.wmn, value=FALSE)
46   s2 <- sum(exposition2[aux2])
47   val <- s1+s2
48   return(val)
49 }
50
51 exposition <- matrix(rep(0,560), nrow = 40, byrow = T )
52 temp <- (y.intvw - 14):(y.intvw - 1)
53 colnames(exposition) <- as.character(temp)
54 temp2 <- 15:54
55 rownames(exposition) <- as.character(temp2)
56
57
58 for (i in c(15:54)) {
59   exposition[i-14,14] <- auxiliary(i, db$age.wmn, db$age2.wmn, db$expo1, db$expo2)
60 }
61
62 for(i in 40:1){
63   for(j in 14:1){
64     exposition[i-1,j-1] <- exposition[i,j]
65   }
66 }
67
68 db$expo1 <- NULL
69 db$expo2 <- NULL
70
71 gqe<-as.matrix(rbind(apply(exposition[c(1:5),c(1:14)],2,sum),
72                           apply(exposition[c(6:10),c(1:14)],2,sum),
73                           apply(exposition[c(11:15),c(1:14)],2,sum),
74                           apply(exposition[c(16:20),c(1:14)],2,sum),
75                           apply(exposition[c(21:25),c(1:14)],2,sum),
76                           apply(exposition[c(26:30),c(1:14)],2,sum),
77                           apply(exposition[c(31:35),c(1:14)],2,sum)))
78
```

```

79
80 estimate_age <- function(m.child, y.child, m.wmn, intvw.age, y.intvw){
81   age <- NULL
82   for(i in 1:length(m.child)){
83     if(is.na(m.child[i]) == F){
84       if(m.child[i] > m.wmn[i] | m.child[i] < m.wmn[i]){
85         if(m.child[i] > m.wmn[i] ){
86           age = c(age, intvw.age[i]-(y.intvw-y.child[i])+1)
87         }
88         else{
89           age = c(age, intvw.age[i]-(y.intvw-y.child[i]))
90         }
91       }
92     }
93     else{
94       age = c(age, intvw.age[i]-(y.intvw-y.child[i])+stats::rbinom(1,1,0.5))
95     }
96   }
97   else{age = c(age,0)}
98 }
99 return(age)
100 }
101
102 db$age.mother <- estimate_age(db$m.child, db$y.child, db$m.wmn, db$intvw.age, y.intvw)
103
104 auxiliary_2 <- function(date, age, data){
105   aux <- grep(date, data[, 'y.child'], value=FALSE)
106   aux2 <- data[aux, c('weights', 'age.mother')]
107   aux3<-grep(age,aux2[,2],value=FALSE)
108   aux4<-aux2[aux3,1:2]
109   aux5<-rep(aux4[,2],aux4[,1])
110   return(length(aux5))
111 }
112
113
114 birth <- matrix(rep(0,560),nrow = 40, byrow = T )
115 temp <- (y.intvw - 14):(y.intvw - 1)
116 colnames(birth) <- as.character(temp)
117 temp2 <- 15:54
118 rownames(birth) <- as.character(temp2)
119
120
121 for (j in (y.intvw - 14):(y.intvw - 1) ) {
122
123   for (i in c(15:54)) {
124
125     birth[i-14, j-(y.intvw - 15)]<- auxiliary_2(j, i, db)
126   }
127 }
128
129 gqb<-as.matrix(rbind(apply(birth[c(1:5),c(1:14)],2,sum),
130   apply(birth[c(6:10),c(1:14)],2,sum),
131   apply(birth[c(11:15),c(1:14)],2,sum),
132   apply(birth[c(16:20),c(1:14)],2,sum),
133   apply(birth[c(21:25),c(1:14)],2,sum),
134   apply(birth[c(26:30),c(1:14)],2,sum),
135   apply(birth[c(31:35),c(1:14)],2,sum)))
136
137
138 value <- list(expo = exposition, birth = birth, gqe = gqe, gqb = gqb, y.intvw = y.intvw)
139
140 attr(value, 'class') <- 'frts_15yrs'

```

C. CÓDIGO FUENTE DEL PAQUETE *FERTILITYRATES*

```
141     value
142
143
144
145
146 } else {
147   stop('There is not some data.frame')
148
149 }
150 }
```

Código C.9: Método que proporciona los objetos de la clase *frts_15yrs*

```
1
2 summary.frts_15yrs <- function(object, ...){
3
4   rgp <- object$gqb/object$gqe
5
6   m1 <- c(mean(rgp[7, 13:14]), mean(rgp[7, 12:13]), mean(rgp[7, 11:12]))
7   m2 <- c(mean(rgp[6, 13:14]), mean(rgp[6, 12:13]), mean(rgp[6, 11:12]), mean(rgp[6, 10:11 ]),
8     mean(rgp[6, 9:10]), mean(rgp[6, 8:9]), mean(rgp[6, 7:8]), mean(rgp[6, 6:7]))
9
10  est1 <- c(rgp[7, 11],rep(stats::sd(m1),10))
11  est1 <- rev(cumsum(est1))
12  est2 <- c(rgp[6, 6],rep(stats::sd(m2),5))
13  est2 <- rev(cumsum(est2))
14
15  rgp[7, 1:11] <- est1
16  rgp[6, 1:6] <- est2
17  row.names(rgp) <- c('15-19', '20-24', '25-29', '30-34', '35-39', '40-44', '45-49')
18
19
20  #X=15,20, 25,...,45,50
21  #X=50 TGF
22  FX <- function(x, year, db){
23    if(x==15){
24      fx=NA
25    }
26    else{
27      rtsg <- db[, (dimnames(db)[[2]]==as.character(year))]
28      fx <- 5*sum(rtsg[1:((x/5)-3)])
29    }
30    return(fx)
31  }
32
33
34  TGFQ <- function(year,db){
35    rtsg <- db[, (dimnames(db)[[2]]==as.character(year))]
36    tgf <- 5*sum(rtsg)
37    return(tgf)
38  }
39
40
41
42  VX1 <- function(fx, tgf){
43    vx1 <- log(-log( fx/ tgf ))
44    return(vx1)
45  }
46
```

```

47
48 VX2 <- function(y, year, db){
49   m <- ((VX1(FX(35, year, db), TGFQ(year, db)) + VX1(FX(40, year, db), TGFQ(year, db)) + VX1(FX
      (45, year, db), TGFQ(year, db))- VX1(FX(20, year, db),TGFQ(year, db))- VX1(FX(25, year,
      db), TGFQ(year, db))-VX1(FX(30, year, db), TGFQ(year, db)))/3)/15
50   vx2 <- m*(y-25)+(VX1(FX(20, year, db),TGFQ(year, db)) + VX1(FX(25, year, db), TGFQ(year, db))
      + VX1(FX(30, year, db), TGFQ(year, db)))/3
51   return(vx2)
52 }
53
54
55 auxFX <- function(vx, year, db){
56   fx1 <- exp(-exp(vx)) * TGFQ(year, db)
57   return(fx1)
58 }
59
60
61 disaggregate <- function(auxfx1,auxfx2){
62   f1x1 <- auxfx2-auxfx1 #la tasa corresponde al anio de auxfx1
63   return(f1x1)
64 }
65
66 rts_dis <- matrix(0, nrow = 35, ncol = 14)
67
68 for(j in (object$y.intvw-14):(object$y.intvw-1)){
69   for(i in 15:49){
70     rts_dis[i-14,j-(object$y.intvw-15)] <- disaggregate(auxFX(VX2(i, j, rgp), j, rgp), auxFX(
      VX2(i+1, j, rgp), j, rgp))
71   }
72 }
73
74 as_frs <- object$birth/object$expo
75 as_frs <- as_frs[1:35, ]
76
77 for(i in 1:10){
78   if(i==1){
79     as_frs[25+i, 1]= NA }
80   else{
81     as_frs[i+25, c(1:i)]= NA
82   }
83 }
84
85
86 for(i in 1:14){
87   as_frs[(which(is.na(as_frs[,i])==T)),i] <- rts_dis[(which(is.na(as_frs[,i])==T)),i]
88 }
89
90 t_frs <- apply(as_frs, 2, sum)
91 t_frs <- as.data.frame(t_frs)
92
93
94 value <- list(as_fr_s = round(as_frs, 3), t_frs = round(t_frs, 3), as_fr_g = round(rgp,3))
95 value
96
97 }

```

Bibliografía

- CELADE, N. C. (1982), *El contexto social de cambio de la fecundidad en América Latina Rural: aspectos metodológicos y resultados empíricos*, Sede de la CEPAL en Santiago (Estudios e Investigaciones).
URL: <https://repositorio.cepal.org/handle/11362/32438>
- Chackiel, J. (1979), 'Estructura de la fecundidad por edades: ajuste y proyección mediante la función Gompertz linealizada', *Notas de Población* **20**(9), 9–18.
- CONAPO (2005), *La fecundidad en México. Niveles y tendencias recientes*, 1 edn, Consejo Nacional de Población (CONAPO), México, D. F.
- CONAPO (2016), *Situación de la Salud Sexual y Reproductiva*, Consejo Nacional de Población (CONAPO), México, D. F.
URL: www.cedes.org
- Cox, P. R. (1976), *Demography*, fifth edit edn, Cambridge University Press.
- Crochran, W. C. (1977), *Técnicas de Muestreo*, Compañía Editorial Continental, México.
- DataMentor (n.d.), 'R OOP (Objects and Classes Introduction)'.
URL: <https://www.datamentor.io/r-programming/object-class-introduction>
- DHS (1997), *An Analysis of Sample Designs and Sampling Errors of the Demographic and Health Surveys*, The Demographic and Health Surveys (DHS), U.S.A., Beltsville.
- Efron, B. (1993), *An Introduction to the Bootstrap*, Chapman & Hall/CRC, U.S.A.
- Feres, J. C., Mancero, X. & Pizarro, R. (2001), 'Estudios Estadísticos Y Prospectivos', *Estudios estadísticos y prospectivos* **4**(66), 52.
URL: <http://www.cepal.org/publicaciones/xml/8/14038/lc2024e.pdf>
- Freund, J. E. & Simon, G. A. (1994), *Estadística elemental*, 8 edn, Prentice Hall Hispanoamericana, México, Edo.Méx.
- Gloria, L. E. (2001), Niveles y tendencias de fecundidad y mortalidad infantil en México a partir de encuestas demográficas, 1950-1996, Maestría, Facultad Latinoamericana de Ciencias Sociales – Sede México.

BIBLIOGRAFÍA

- Gutierrez, A. (2015), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*, 2 edn, ISBN:9789586316088, Bogotá, D.C.
- Hansen, M., Hurwitz, W. & Madow, W. (1953), *Sample Survey Methods and Theory. Volume I. Methods and Applications.*, Vol. 1, John Wiley & Sons, INC.
- INEGI (1992), ENADID 1992, Características metodológicas.
- INEGI (1999), ENADID 1997, Características metodológicas.
- INEGI (2003), ENADID 1997, Síntesis metodológica.
- INEGI (2009), *Colección Memoria*.
URL: https://unstats.un.org/unsd/wsd/docs/Mexico_wsd_125_anos_DGE.pdf
- INEGI (2010), ENADID 2009, Informe operativo.
- INEGI (2013), ENADID 2009, Reemplazo de microdatos, Technical report.
- INEGI (2015a), ENADID 2014, Diseño muestral.
- INEGI (2015b), ENADID 2014, Informe operativo.
- INEGI (2016), ENADID 2014, Marco de referencia.
- INEGI (2018), 'Quiénes somos'.
URL: https://www.inegi.org.mx/inegi/quienes_somos.html
- INEGI (2019a), ENADID 2018, Diseño conceptual.
- INEGI (2019b), ENADID 2018, Diseño muestral.
- INEGI (2019c), ENADID 2018, Informe operativo y de procesamiento.
- INEGI (2019d), ENADID 2018, Principales resultados, Technical Report 1.
- INSP (2008a), ENADID 2006, Diseño muestral.
- INSP (2008b), ENADID 2006, Informe final.
- Lohr, S. (2000), *Muestreo: diseño y análisis*, International Thomson Editores, Arizona State University.
- Lumley, T. (2010), *Complex Surveys: A Guide to Analysis Using R*, John Wiley & Sons, INC., Hoboken, New Jersey.
- Mina, A. (2006), *Curso Básico de Demografía*, Vínculos matemáticos, Servicios Editoriales de la Facultad de Ciencias, UNAM, México.
- Mojarro, O. & Núñez, L. (1988), 'Mortalidad infantil en México: tendencias y factores determinantes.', *Salud Pública de México* **30**(3), 329–345.

- Moultrie, T. (2013), Direct estimation of fertility from survey data containing birth histories, *in* 'Tools for Demographic Estimation', 1 edn, International Union for the Scientific Study of Population (IUSSP), pp. 110–117.
URL: http://getinthepicture.org/sites/default/files/resources/Tools_for_Demographic_Estimation.pdf#page=121
- ONU (2007), 'Encuestas de Hogares en los países en Desarrollo y en Transición', (96).
- Ordorica, M. & Potter, J. E. (1981), 'Scientific Reports', *Evaluation of the Mexican Fertility Survey 1976-77* **21**(17).
URL: https://wfs.dhsprogram.com/WFS-SR/ISI-WFS_SR-21-Ordorica_1981-Evaluation_of_the_Mexican_Fertility_Survey_1976-77.pdf
- Paradis, E. (2003), R para Principiantes (Jorge A. Ahumada, trad.).
URL: <https://cran.r-project.org/doc/contrib/rdebut.es.pdf>
- Partida, V. (2017a), *Conciliación demográfica de México 1950-2015*, México, D. F.
URL: <http://www.cepal.org/publicaciones/xml/8/14038/lc2024e.pdf>
- Partida, V. (2017b), 'Lo que cuenta la Conciliación Demográfica en México, 1950-2015', *Coyuntura Demográfica* pp. 15–21.
- Potter, J. E. (1977), 'Problems in using birth-history analysis to estimate trends in fertility*', *Population Studies* **31**(2), 335–364.
- PRB (2011), *Population Reference Bureau's Population Handbook*, Vol. 36.
URL: <https://www.prb.org/wp-content/uploads/2011/09/prb-population-handbook-2011-1.pdf>
- Preston, S. H., Heuveline, P. & Guillot, M. (2001), *Demography: measuring and modeling population processes*, 1 edn, Blackwell Publishers Ltd, Gran Bretaña.
- Santana, J. S. & Farfán, E. M. (2014), *El arte de programar en R: un lenguaje para la estadística*, 1 edn, Instituto Mexicano de Tecnología del Agua, Jiutepec, Morelos.
URL: https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf
- Särndal, C.-E., Swensoon, B. & Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer, New York.
- Schoumaker, B. (2004), 'Une approche personnes-périodes pour l'analyse des histoires génésiques', *Population* **59**(5), 783.
- Schoumaker, B. (2013), 'A Stata module for computing fertility rates and TFRs from birth histories: tfr2', *Demographic Research* **28**(May), 1093–1144.
URL: <http://www.demographic-research.org/volumes/vol28/38/>
- Secretaría de Salud/DHS (1990), México: ENFES1987: informe resumido.

BIBLIOGRAFÍA

Shryock, H. S. & Siegel, J. S. (2004), *The methods and materials of demography*, second edn, Elsevier Academic Press.

Skinner, C. J., Holt, D. & Smith, T. (1989), *Analysis of complex surveys*, John Wiley & Sons., England.

Thomas, R. K. (2018), *Concepts, Methods and Practical Applications in Applied Demography*, Springer Nature, USA.

Wickham, H. (2014), *Advanced R*, CRC Press.

URL: <http://adv-r.had.co.nz/>

Wickham, H. (2015), *R Packages*, 1 edn, O'Reilly Media.

URL: <http://r-pkgs.had.co.nz/>

Wolter, K. M. (2007), *Introduction to Variance Estimation*, segunda edn, Springer.