



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

APRENDER DEMOGRAFÍA I
CON "R-PROJECT"

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A:

PAULA DE LA LAMA ZUBIRÁN

TUTORA:

M. en D. MARIA TERESA VELÁZQUEZ URIBE

2008



FACULTAD DE CIENCIAS
UNAM



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.

Hoja de Datos del Jurado

<p>1. Datos del alumno Apellido paterno Apellido materno Nombre(s) Teléfono Universidad Nacional Autónoma de México Facultad de Ciencias Carrera Número de cuenta</p>	<p>1. Datos del alumno De la Lama Zubirán Paula 56 17 85 49 Universidad Nacional Autónoma de México Facultad de Ciencias Actuaría 0971237099</p>
<p>2. Datos del tutor Grado Nombre(s) Apellido paterno Apellido materno</p>	<p>2. Datos del tutor M. en D. María Teresa Velázquez Uribe</p>
<p>3. Datos del sinodal 1 Grado Nombre(s) Apellido paterno Apellido materno</p>	<p>3. Datos del sinodal 1 Dra. Ruth Selene Fuentes García</p>
<p>4. Datos del sinodal 2 Grado Nombre(s) Apellido paterno Apellido materno</p>	<p>4. Datos del sinodal 2 M. en P. Laura Elena Gloria Hernández</p>
<p>5. Datos del sinodal 3 Grado Nombre(s) Apellido paterno Apellido materno</p>	<p>5. Datos del sinodal 3 Dra. Ivonne Rosa Szasz Pianta</p>
<p>6. Datos del sinodal 4 Grado Nombre(s) Apellido paterno Apellido materno</p>	<p>6. Datos del sinodal 4 L. en C. C. Mauricio Morgado Castillo</p>
<p>7. Datos del trabajo escrito. Título Número de páginas Año</p>	<p>7. Datos del trabajo escrito Aprender Demografía I con "R-Project" 159 p 2008</p>

Índice

Introducción	1
1. Marco conceptual y fuentes de información	5
1.1. Live CD.....	5
1.2. Ubuntu.....	6
1.3. El programa “R-Project”.....	7
1.3.1. Estadística con “R-Project”.....	8
1.4. Fuentes de información.....	8
1.4.1. Censos.....	9
1.4.2. Estadísticas vitales.....	10
1.4.3. Tipos de errores en las fuentes.....	10
1.4.3.1. Error de cobertura.....	10
1.4.3.2. Error de respuesta.....	11
1.4.3.3. Error de codificación.....	11
1.4.4. Información Utilizada.....	11
2. Como funciona "R-Project"	13
2.1. Cómo empieza R.....	13
2.2. Creación, listado y eliminación de objetos en memoria.....	14
2.3. La ayuda en línea.....	15
2.4. Manejando Datos con R.....	16
2.4.1. Objetos.....	16
2.4.2. Leyendo datos desde un archivo.....	18
2.4.3. Guardando datos.....	20
2.4.4. Generación de datos.....	21
2.4.4.1. Secuencias regulares.....	21
2.4.4.2. Secuencias aleatorias.....	23
2.4.5. Manipulación de objetos.....	23
2.4.5.1. Creación de objetos.....	23
2.4.5.2. Operadores.....	27
2.4.5.3. Cómo acceder los valores de un objeto: el sistema de indexación.....	27
2.4.5.4. Funciones aritméticas simples.....	28
2.4.5.5. Cálculos con Matrices.....	30
2.5. Las gráficas en R.....	32
2.5.1. Manejo de gráficos.....	32
2.5.1.1. Abriendo múltiples dispositivos gráficos.....	32
2.5.2. Funciones gráficas.....	34
2.5.3. Comandos de graficación de bajo nivel.....	35
2.5.4. Parámetros gráficos.....	36
2.6. Análisis estadísticos con R.....	37
2.7. Programación práctica con R.....	39
2.7.1. Bucles y Vectorización.....	39
2.7.2. Escribir un programa en R.....	41
3. Aprender Demografía con R-Project	43
3.1 Evaluación de la información.....	43
3.1.1. Las Tasas de crecimiento intercensal.....	44
3.1.1.1. La Tasa aritmética.....	45
3.1.1.2. Tasa geométrica.....	46
3.1.1.3. La Tasa exponencial.....	47

3.1.2. La Pirámide de edades individuales.....	48
3.1.3. Procedimientos para calificar los datos de la población censada.....	53
3.1.3.1. Índice de Whipple.....	53
3.1.3.2. Índice de Myers.....	55
3.1.3.3. Índice de Naciones Unidas.....	59
3.1.3.3.1. El Índice de regularidad por sexo.....	59
3.1.3.3.2. Coeficiente por edad y por sexo.....	61
3.1.4. Conclusiones.....	63
3.2. Corrección de la información.....	65
3.2.1. Distribución de la población no especificada.....	65
3.2.2. Método agrupamiento (demográfico).....	67
3.2.3. Método analítico.....	67
3.2.3.1. Promedio Móvil Simple (PMS).....	68
3.2.3.2. Promedios Móviles Ponderados (1/16).....	69
3.2.4. Población corregida a la mitad del año.....	70
3.2.5. Índice de masculinidad.....	72
3.2.6. El Diagrama de Lexis.....	73
3.2.6.1. Factores de separación.....	76
3.2.7. Pirámides de edad.....	81
3.2.8. Conclusiones.....	83
3.3. Componentes demográficos.....	85
3.3.1. Mortalidad.....	85
3.3.1.1. Tasa bruta de mortalidad.....	86
3.3.1.2. Tasas específicas de mortalidad.....	87
3.3.1.3. Mortalidad infantil.....	88
3.3.1.3.1. Mortalidad neonatal y postnatal.....	89
3.3.1.4. Tabla de mortalidad.....	91
3.3.1.5. Conclusiones.....	99
3.3.2. Fecundidad.....	100
3.3.2.1. Tasa bruta de natalidad.....	101
3.3.2.2. Tasa de fecundidad general.....	102
3.3.2.3. Tasas específicas de fecundidad.....	103
3.3.2.3.1. Edad cúspide.....	105
3.3.2.3.2. Grado de concentración.....	106
3.3.2.4. Factores que afectan las medidas.....	110
3.3.2.4.1. Factores extrínsecos.....	110
3.3.2.4.2. Factores intrínsecos.....	113
3.3.2.5. Medidas resumen de la fecundidad.....	115
3.3.2.5.1. Tasa global de fecundidad.....	115
3.3.2.5.2. Tasa bruta de reproducción.....	116
3.3.2.5.5. Descendencia media final.....	116
3.3.2.5.3. Tasa neta de reproducción.....	117
3.3.2.5.4. Edad media de la fecundidad.....	121
3.3.2.6. Conclusiones.....	122
3.3.3. Migración.....	123
3.3.3.1. Crecimiento natural y crecimiento social.....	123
3.3.3.2. Migración según lugar de nacimiento (migración absoluta).....	125
3.3.3.3. Migración según lugar de residencia anterior.....	127
3.3.3.3.1. Migración según lugar de residencia anterior a una fecha determinada...127	

3.3.3.4. Métodos indirectos para medir la migración.....	129
3.3.3.4.1. Método prospectivo.....	130
3.3.3.4.2. Método retrospectivo.....	133
3.3.3.4.3. Método promedio.....	134
3.3.3.5. Conclusiones.....	136
Conclusiones generales.....	137
Recomendaciones.....	138
Glosario.....	139
Apéndice de archivos.....	142
Bibliografía.....	158

Introducción

Al iniciar mi desarrollo profesional fuera de las aulas universitarias me encontré con que carecía de formación en el uso de las nuevas tecnologías que se estaban desarrollando para el análisis de datos. Sin embargo, gracias a que fui aceptada para realizar mi Servicio Social en la Dirección General de Servicios de Cómputo Académico (DGSCA) de la UNAM me fue posible conocer el programa R-Project, utilizado para el análisis estadístico de una gran cantidad de datos.

Durante el proceso de conocer las cualidades y las ventajas de este programa surgió la idea de compartirlo con los estudiantes de la Facultad de Ciencias de la carrera de Actuaría. Una manera de que estos alumnos conocieran el programa y lo manejaran fue que recurrieran a él, como una herramienta práctica para resolver los dilemas que se presentan en muchos de los problemas actuariales que se aprenden en la licenciatura.

El primer paso fue crear un ambiente de trabajo propicio para que los estudiantes aprendieran la parte teórica de la asignatura y después fueran capaces de insertar los datos en el programa y de esta manera se calculó la parte práctica. Esta manera de encarar el problema corre el riesgo de que el alumno aprenda solamente a utilizar el ambiente y no a manejar los datos ni el programa R-Project por lo que la idea no acababa de estar completa.

Fue entonces cuando se pensó en enseñar a utilizar el programa al mismo tiempo que se aprendía una asignatura. La asignatura elegida fue Demografía I, que se imparte en la licenciatura de Actuaría de la Facultad de Ciencias de la UNAM y que está a cargo de la profesora María Teresa Velázquez Uribe. Las razones de recurrir a esta asignatura fueron que requería de una gran cantidad de datos demográficos y numerosos cálculos que se resuelven en la actualidad mediante el conocido método de “arrastrar el lápiz” utilizando la hoja de Excel. Gracias a la estrategia de enseñar simultáneamente ambos contenidos esperamos incorporar nuevos sistemas que permitan a los egresados de esta Universidad contar con ventajas competitivas que los harán destacar por encima de otras Universidades. Así fue como se inició este proyecto y cuyos objetivos son:

Objetivos

1. Proponer una nueva forma de impartir la asignatura de Demografía I, mediante el uso del programa computacional R-Project.
2. Que los actuarios conozcan las potencialidades que ofrece el uso de las

herramientas computacionales para el estudio de los problemas que aborda la carrera de Actuaría.

Método

La forma con la que se planeó darles a los estudiantes de Demografía un acceso al programa R-Project y a la parte teórica de la asignatura fue a través de crear un live-CD (CD que contiene un ambiente total de trabajo) que para su uso sólo requiere una computadora con acceso a un CD. El alumno, mediante el cd, podrá conocer el programa computacional y manejarlo al mismo tiempo que trabaja su clase de Demografía I.

El live-Cd tiene una plataforma basada en Linux en el cual se diseñaron cuatro ventanas operativas. La primera es un acceso directo a R-Project que es un programa que analiza estadísticamente una gran cantidad de datos. La segunda es una parte de la página de Internet que guarda ciertas instrucciones que facilitan el trabajo con el programa. La tercera ventana contiene los manuales que sirven para conocer los comandos de R-Project. Por último se tiene un acceso a las instrucciones creadas para estudiar el programa computacional junto con la materia de Demografía I.

Este sistema de aprendizaje se inició en enero de 2006 en la subdirección de Prospección e Innovación Tecnológica a cargo del Actuario Mario García Burgos en la Coordinación General de Servicios Educativos en Red (SER-UNAM) y esta adscrito a la DGSCA. Se nos proporcionó una computadora con procesador Intel Pentium IV a 3.0 GHZ con 256MB en RAM. Con la ayuda de Héctor E. Reyes Castro jefe del departamento de Innovación en Implementación y Mauricio Morgado Castillo. Así se pudo conocer las posibilidades de R-Project.

Amablemente la profesora María Teresa Velázquez Uribe se mostró interesada en aplicar nuevas técnicas a su asignatura y de ahí nació este proyecto.

Resumen por capítulos

El trabajo se divide en tres capítulos. El primero hace la presentación del trabajo, es decir, explica lo que es el live-CD y como funciona. Luego describe al programa R-Project y su capacidad para el análisis estadístico de datos. Además muestra las fuentes que se tomaron para hacer el análisis del caso del Estado de Nuevo León, que sirve de ejemplo y muestra los cálculos para determinar los errores o distorsiones que pueden surgir en estos datos debido a fallas en el trabajo de campo.

El segundo capítulo expone brevemente como funciona R para enseñar al usuario las funciones básicas para utilizar el programa. Estas funciones van desde la creación de objetos, su manejo, cómo leer y exportar datos, pues generalmente se maneja una gran cantidad de información, para vectores, matrices, funciones estadísticas y gráficas. Con estos conceptos el alumno se puede adentrar en el trabajo demográfico de los futuros capítulos.

En el tercer capítulo se hace la aplicación de R-Project a demografía con la importación de la información al programa, se considera en la primera parte la evaluación de la calidad de los datos de las edades que se obtienen de las fuentes de información del Instituto Nacional de Estadística Geografía e Informática (INEGI). Se calculan las tasas de crecimiento intercensal para medir como crece la población de un censo a otro y al mismo tiempo se empieza a conocer algunos comandos del programa. Se crean las pirámides de edad. Para la elaboración de estos existen manuales de R-Project, que se incluyen en el disco, ya que se espera que el alumno los utilice para su aprendizaje. Después se estudian los diferentes métodos de corrección.

En la segunda parte se muestra el método de corrección de datos. Los análisis que se hacen se basan en los métodos de agrupamiento, analíticos y gráficos. También se explica el Diagrama de Lexis y los factores de separación que éste utiliza. Además se construyen nuevas pirámides con la función antes creada y se utilizan funciones anidadas para que el alumno entienda las bondades del programa. Después de corregir los datos se pueden volver a evaluar para asegurarse que los resultados de los análisis futuros serán correctos.

En el tercer apartado se habla de los componentes demográficos, que son los sucesos que disminuyen o aumentan la población. Estos se dividen en tres: mortalidad, fecundidad y migración. De la misma manera este capítulo se divide en tres secciones.

La sección de mortalidad maneja tasas para medir la mortalidad en ciertas zonas, en la práctica de la asignatura se usó como ejemplo del Estado de Nuevo León. Se analiza la mortalidad infantil, neonatal y postnatal aunque éstas se midan en el diagrama de Lexis que se estudia anteriormente. Además se construye la tabla de mortalidad que se utiliza tanto en estudios sobre la población como en una compañía de seguros de vida.

En la parte de fecundidad se calculan diferentes tasas para saber a que edades las mujeres tienen hijos, cuántos y cuántas hijas para conocer cuanto va a crecer, por

nacimientos, la población que se esta estudiando.

En migración se aclara la manera de clasificar a un migrante, ya que estos hacen que crezca o disminuya la población. En el caso de los estados se miden por las preguntas del censo (método directo) o por la ecuación compensadora y relaciones de sobrevivencia (métodos indirectos).

En esta seccion se aplica el conocimiento del programa con nuevas gráficas y con nuevos comandos además de que se enseña como exportar los datos para poder entregar un informe.

Por último se presentan las conclusiones y recomendaciones de uso de R-Project y sus aplicaciones.

1. Marco conceptual y fuentes de información

Este trabajo pretende enseñar al usuario con especial interés a los alumnos de la carrera de Actuaría de la Facultad de Ciencias la utilización de el programa “R-Project”. Esto para que él tenga más herramientas de cómputo al momento de terminar la carrera y pueda cumplir con los altos estándares de competencia que se requieren hoy en día en el ámbito laboral.

La presentación de este trabajo es por medio de un live-Cd para facilitar al usuario el programa R-Project, la información básica de éste (manuales), la teoría de demografía y la librería en R-Project con las funciones de la materia. Porque aunque es software libre (ver Glosario) no siempre se tiene acceso a todo esto.

La forma de aprender a manejar este programa se hizo por medio de los temas de demografía. El usuario conocerá las funciones necesarias para el seguimiento del curso de Demografía I. A la vez se creó una librería que contiene las funciones derivadas del problema demográfico. La librería se creó en la plataforma de nuestro programa llamado “R-Project”.

Los ejemplos de los programas y cálculos demográficos se hicieron por medio de información que proveen los censos, de los años 1990 y 2000, además, de las fuentes estadísticas vitales del Estado de Nuevo León. Los temas del nuestro proyecto consintieron en replicar las clases prácticas de la asignatura de Demografía I, en la Facultad de Ciencias de la UNAM. Que imparte la profesora Maria Teresa Velázquez.

1.1. LiveCD

Un Live-CD. Es un ambiente de trabajo (sistema operativo) que se puede ejecutar directamente desde un CD, en vez de tener que hacerlo desde el disco duro.

El CD utiliza la memoria ram (random access memory), como si se tratase del disco duro. Sus ventajas son:

- No necesita del disco duro, ni ninguna instalación por lo que no se pierde tiempo en esto. Optimizando el tiempo de clase.

- Puedes utilizar cualquier equipo, sólo varían por la velocidad con que va a funcionar el programa según la memoria de cada computadora, y no importa que tiene instalado pues no usa el software del equipo.

- Tiene acceso a los recursos del equipo con que se está trabajando.

Sus desventajas son:

- No guarda cambios, por lo que el trabajo realizado desaparece cuando apagas la computadora, y el tamaño del sistema que se usará, no podrá ser mayor que el tamaño de la memoria ram. Aunque esto tiene diferentes arreglos.

(<http://linux.pricoinsa.es> en octubre de 2006).

El live-CD que se utiliza está hecho en GNU/Linux (Ver glosario) elaborado por Ubuntu.

1.2. Ubuntu

Ubuntu es un sistema operativo completo, construido en una plataforma con base Linux, disponible para toda la comunidad. La filosofía ubuntu dice que el software deberá estar siempre disponible sin costo alguno y los usuarios siempre tendrán la libertad de adaptar y alterar el software de acuerdo a sus necesidades particulares (<http://www.ubuntu-es.org> en Octubre de 2006).

Ubuntu usa lo mejor en infraestructura de traducciones y accesibilidad que la comunidad del Software Libre es capaz de ofrecer, para hacer que sea utilizable por el mayor número de personas. Además invita a sus usuarios a mejorar el sistema y a compartirlo (<https://help.ubuntu.com> en octubre de 2006) . Así es como siempre se está actualizando.

El nombre Ubuntu viene de una ideología étnica Sud-Africana donde

"Una persona con ubuntu está abierta y disponible para otros, aclarando a otros, no se siente amenazada por la habilidad y bondad de otros, tiene una seguridad propia que viene de saber que él o ella proviene de un gran todo y es disminuida cuando otros son humillados o disminuidos, cuando otros son torturados u oprimidos."

–Arzobispo Desmond Tutu (<https://help.ubuntu.com> en octubre de 2006)

Por esto se utilizó ubuntu en este trabajo además de la facilidad con que se pueden

hacer modificaciones al live-CD.

1.3. El programa “R-Project”

R-Project es un programa para análisis estadístico que incluye gráficos, creado por la "fundación R" la cual esta situada en Viena, Austria y se hospeda en la Universidad tecnológica de Viena. Ésta es una asociación registrada bajo las leyes de austríacas y es soportado por académicos de todo el mundo. Su lenguaje de programación es considerado como un dialecto del lenguaje “S” creado por los laboratorios AT & Bell. Por lo tanto en “S” también se puede desarrollar un análisis estadístico similar, pero exige que se construya paso a paso y sin las ventajas que tiene “R-Project” (http://cran.r-project.org/doc/contrib/rdebuts_es.pdf en marzo de 2006).

“R-Project” se distribuye gratuitamente bajo los términos de la GNU general Public Licence (Para mayor información: <http://www.gnu.org>); Su desarrollo y distribución son llevados a cabo por varios estadísticos conocidos como el Grupo Nuclear de Desarrollo de R-Project (http://cran.r-project.org/doc/contrib/rdebuts_es.pdf en marzo de 2006).

Los archivos necesarios, para instalar R-Project, se distribuyen desde el sitio de Internet Comprehensive R Archive Network (CRAN (<http://cran.r-project.org> en marzo 2006)) junto con las instrucciones de instalación.

R-Project posee muchas funciones para el análisis estadístico y los gráficos; estos últimos pueden ser visualizados de manera inmediata en su propia ventana y ser guardados en varios formatos (jpg, bmp, pdf, etc.). Los resultados del análisis estadístico se muestran en pantalla y algunos resultados intermedios se pueden guardar, para hacer un análisis parcial, sin necesidad de mostrar sus resultados inmediatamente y extraer solo los resultados que interesen.

El lenguaje de R-Project tiene una sintaxis muy simple e intuitiva. Además permite programar bucles (nuevas funciones) para analizar conjuntos sucesivos de datos; combinar diferentes funciones o realizar otras pruebas estadísticas, así como ayuda en línea. La disponibilidad de los programas escritos para S se encuentra disponible en la red.

R-Project maneja un lenguaje “Orientado a Objetos, que significa que las variables,

datos, funciones, resultados, etc. se guardan en la memoria activa del computador en forma de objetos con un nombre específico. El usuario puede modificar o manipular estos objetos con operadores (aritméticos, lógicos y comparativos) y funciones (que a su vez son objetos)." (http://cran.r-project.org/doc/contrib/rdebuts_es.pdf en marzo de 2006). Además los resultados mismos son objetos, que pueden ser considerados como datos y analizarlos como tales.

R-Project es interpretado y no compilado. Lo que significa que los comandos escritos en el teclado son ejecutables directamente hasta que encuentra un comando no valido.

1.3.1. Estadística con R-Project

El entorno de R-Project se formó con un conjunto de programas de datos, cálculos y gráficos. Algunas de sus características son el almacenamiento y manipulación de datos; operadores para cálculos en vectores y matrices; una variedad de herramientas para el análisis de datos; gráficas de estos; y un lenguaje de programación que incluye: condicionales, ciclos, funciones recursivas y la posibilidad de importar y exportar datos (<http://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf> en marzo de 2006).

Todo esto logra que R-Project sea un vehículo para el desarrollo de nuevos métodos de análisis interactivo de datos, ya que es muy dinámico y posee muchas técnicas estadísticas, algunas incluidas en el entorno base y otras como bibliotecas (packages).

Así R-Project logra manejar una amplia variedad estadística desde las clásicas hasta las últimas metodologías. Además de técnicas para graficar incluye símbolos matemáticos y formulas que dan facilidad al momento de usarse.

1.4. Fuentes de información

Se utilizaron las siguientes fuentes de información:

1.4.1 Censos

El censo de población obtiene información sobre las características económicas y sociales de cada persona y de cada hogar en un territorio determinado, por ejemplo un país, en una determinada fecha (Departamento de Asuntos Económicos y Sociales, División de estadística. “Manual sobre la recolección de datos de fecundidad y mortalidad”, Naciones Unidas, Nueva York, 2005). Además deben cumplir los criterios básicos de universalidad y simultaneidad.

Los censos tienen como objetivo reunir, elaborar y publicar datos demográficos (información sistemática sobre las entidades de algún agregado estadístico), es decir, generan las estadísticas referentes de los aspectos más relevantes de la población total y sus viviendas, así como de las unidades económicas del país de tal manera que pueda compararse, estudiarse y presentarse esta información con gran detalle geográfico, como: municipio, localidad, Área Geoestadística Básica (AGEB) y manzana. Además tienen un gran nivel de desagregación conceptual.

(http://www.inegi.gob.mx/est/contenidos/espanol/proyectos/metadatos/censos/pcn_10.asp?c=1434 en marzo 2006).

Los censos demográficos se usan en general para conocer el número de personas, las características de los hogares y las familias, características económicas, el nivel de instrucción, la ocupación y la situación migratoria, que luego sirven para calcular valiosos indicadores socio-económicos, políticos y de la población en sí. (Departamento de Asuntos Económicos y Sociales, División de estadística. “Manual sobre la recolección de datos de fecundidad y mortalidad”, Naciones Unidas, Nueva York, 2005).

Los censos de población deben considerarse como una fuente potencial de datos sobre fecundidad y mortalidad. Las preguntas especiales para capturar esa información son mínimas, pero los datos producidos pueden ser de gran valor. (Departamento de Asuntos Económicos y Sociales, División de estadística. “Manual sobre la recolección de datos de fecundidad y mortalidad”, Naciones Unidas, Nueva York, 2005.)

Existen dos tipos de censos de población: los de hecho (por lugar de precedencia) y los de derecho (por lugar de residencia habitual).

En México el encargado de efectuar los censos, por lo general cada 10 años, es el

Instituto Nacional de Estadística, Geografía e Informática (INEGI)

1.4.2. Estadísticas vitales

El objetivo de las Estadísticas vitales es generar las estadísticas (Ver Glosario) sobre nacimientos, defunciones, muertes fetales, matrimonios y divorcios, que contribuyan al conocimiento de las características demográficas de la población mexicana.

El Subprograma de Estadísticas vitales de México, lo produjo el Instituto Nacional de Estadística, Geografía e Informática (INEGI). Se elabora a partir de los Sistemas del Registro Civil y de otras fuentes de información, como son las Agencias del Ministerio Público, las cuales reportan las muertes accidentales y violentas; y los Juzgados de lo Familiar, Civiles y Mixtos que reportan los divorcios judiciales.

Las estadísticas vitales cuentan con un amplio desglose geográfico (nacional, entidad federativa, municipio, localidad y tamaño de localidad); ofrecen información con corte anual, desde 1893; muestran aspectos fundamentales de la dinámica poblacional, a partir de la caracterización de los fenómenos de la fecundidad, nupcialidad y mortalidad. (http://www.inegi.gob.mx/est/contenidos/espanol/proyectos/metadatos/continuas/sev_31.asp?c=4301 en marzo de 2006)

Las estadísticas vitales son importantes para el estudio, planeación y evaluación de acciones en materia de bienestar de la población.

(<http://www.inegi.gob.mx/est/contenidos/espanol/proyectos/metadatos/continuas/sev31.asp> en marzo de 2006)

1.4.3. Tipos de errores en las fuentes

Existen varios tipos de errores que pueden afectar los resultados de los análisis demográficos. (Leguina, Joaquín. Fundamentos de la demografía .España. Siglo XXI, 1981 pág 124)

1.4.3.1. Errores de cobertura

1. Por omisiones, tales errores pueden surgir, porque la encuesta se ha diseñado con

un marco incompleto causando que la encuesta no llegue a todos los individuos deseados.

2. Por dobles cuentas, que surge debido al solapamiento en ciertas unidades

1.4.3.2. Errores de respuesta

Estos errores, son muy comunes y se deben a que el encuestado no responde o responde incorrectamente, ya sea involuntariamente por ignorancia o voluntariamente al mentir.

1.4.3.3. Errores de codificación

Son los que se producen por una equivocación al paso del cuestionario a los datos numéricos y de éstos a las fichas estadísticas.

1.4.4. Información utilizada

Se utilizaron los Censos de Población y Vivienda y las Estadísticas Vitales elaboradas por el Instituto Nacional de Estadística Geografía e Informática (INEGI).

En la investigación se utilizan los datos del Estado de Nuevo León de los Censos:

- Población total por sexo y edad de 1990 y 2000.
- Población por sexo según lugar de nacimiento para el 2000
- Población por sexo según lugar de nacimiento para el 1990
- Población de cinco años y más según lugar de residencia hace cinco años (1985 y 1995).
- Población de 12 años y más por estado conyugal del 2000
- Población de 12 años y más por estado conyugal del 1990

De las Estadísticas Vitales se utilizó:

- Defunciones por quinquenios del año 1999, 2000, y 2001.
- Defunciones por quinquenios del año 1989, 1990, y 1991.
- Defunciones de menores de un año de los años de estudio.
- Nacimientos registrados por edad de la madre de los años 1999, 2000, 2001.
- Nacimientos registrados por edad de la madre de los años 1989, 1990, 1991.

2. Como funciona "R- project"

Una vez que ha instalado R en su computadora, el programa se inicia llamando al archivo ejecutable correspondiente. Entonces el cursor aparece, con el símbolo '>', e indica que R está listo para empezar a operar.

2.1. Cómo empieza R

R es un lenguaje interpretado y no compilado (como C, C++, Fortran, Pascal, ...), lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin necesidad de construir ejecutables, esta ventaja además se complementa con que la sintaxis de R es muy simple e intuitiva. Por ejemplo para que una función sea ejecutada en R debe estar siempre acompañada de paréntesis, inclusive en el caso que no haya nada dentro de los mismos, por ejemplo `ls()`. Si se escribe el nombre de la función sin paréntesis, R mostrará el código de la función.

Todas las acciones en R se realizan con objetos que son guardados en la memoria activa del ordenador, sin usar archivos temporales. La lectura y escritura de archivos sólo se realiza para la entrada y salida de datos. Los resultados se pueden visualizar directamente en la pantalla, guardar en un objeto o escribir directamente en un disco (esto es particularmente útil cuando se recurre a gráficos). Esto se debe a que los resultados mismos son objetos, mismos que pueden ser considerados datos y analizados como tales. Los Archivos que contengan datos pueden ser leídos directamente desde el disco local o en un servidor remoto a través de la red.

Las funciones disponibles en R están guardadas en una librería localizada en el directorio R HOME/library (R HOME es el directorio donde R está instalado). Este directorio contiene *paquetes* de funciones, las cuales a su vez están estructuradas en directorios. El paquete denominado base constituye el núcleo de R y contiene las funciones básicas del lenguaje para leer y manipular datos, funciones y gráficos. Cada paquete contiene un directorio cuyo archivo tiene el mismo nombre del paquete (por ejemplo, para el paquete base, existe el archivo R HOME/library/base/R/base). Este archivo está en formato ASCII así como todos los demás.

El comando para visualizar el contenido de un objeto, por ejemplo, "n" que contiene el

valor 10 entonces:

```
> n  
[1] 10
```

El dígito [1] indica que la visualización del objeto comienza con el primer elemento de n (Este comando constituye un uso implícito de la función `print`).

El nombre de un objeto debe comenzar con una letra (A-Z and a-z) y puede incluir letras, dígitos (0-9), y puntos (.). R discrimina entre letras mayúsculas y minúsculas para el nombre de un objeto, de tal manera que `x` y `X` se refiere a objetos diferentes.

2.2. Creación, listado y eliminación de objetos en memoria

Un objeto puede ser creado a partir del operador “asignar”, el cual se denota con el símbolo igual (=) y también con “->” ó “<-” (“->” se forma de dos símbolos menos (-) y mayor que (>)) dependiendo de la dirección en que asigna el objeto:

```
> n <- 15  
> n  
[1] 15  
> 5 -> n  
> n  
[1] 5
```

Si el objeto ya existe, su valor anterior es borrado después de la nueva asignación. El valor asignado puede ser el resultado de una operación y/o de una función:

```
> n <- 10 + 2  
> n  
[1] 12
```

Una expresión a la cual no se le asigna un objeto será ejecutada pero no se guardará en memoria:

```
> (10 + 2) * 5
[1] 60
```

La función `ls` lista los objetos que hay en la memoria. Sólo se muestran los nombres de los mismos.

```
> name <- "Carmen"; n1 <- 10; n2 <- 100; m <- 0.5
> ls()
[1] "m" "n1" "n2" "name"
```

Si se quiere listar solo aquellos objetos que contengan un caracter en particular, se puede usar la opción `pattern` (que se puede abreviar como `pat`):

```
> ls(pat = "m")
[1] "m" "name"
```

Para restringir la lista a aquellos objetos que comienzan con este caracter:

```
> ls(pat = "^m")
[1] "m"
```

Para borrar objetos en memoria, se utiliza la función `rm()`: `rm(x)` elimina el objeto `x`. Para borrar selectivamente algunos objetos se usa: `rm(list=ls(pat="^m"))`.

2.3. La ayuda en línea

La ayuda en línea de R "?" proporciona información útil sobre cómo utilizar las funciones. La ayuda se encuentra disponible directamente para una función dada. Por ejemplo:

```
> ?ls
```

El comando `help(ls)` o `help("ls")` (para acceder a la ayuda con caracteres no convencionales).

Al llamar la ayuda, se obtiene información general sobre la función en la primera línea, tal como el nombre del paquete donde se encuentra dicha función u operador. Después viene el título de la función, seguido de secciones con información específica acerca de la misma.

```
> ?(función)
```

Description: descripción breve.

Usage: para una función proporciona el nombre de la misma con todos sus argumentos y los posibles valores por defecto. Para un operador describe su uso típico.

Arguments: describe en detalle cada uno de los argumentos.

Details: descripción detallada de la función o del operador.

Value: si tiene valor da el tipo de objeto.

See Also: otras páginas de ayuda con funciones u operadores similares.

Examples: algunos ejemplos que generalmente pueden ser ejecutados sin abrir la ayuda con la función `examples()`.

La función `apropos` encuentra todas aquellas funciones cuyo nombre contiene la palabra dada como argumento para los *paquetes* cargados en la memoria:

```
> apropos(help)
[1] "help" "help.search" "help.start"
[4] "link.html.help"
```

2.4. Manejo de datos con R

2.4.1. Objetos

Los objetos en R tienen nombre, contenido y atributos que especifican el tipo de datos representados por el objeto. La acción de una función sobre un objeto depende de los atributos de este último.

Todo objeto tiene dos atributos intrínsecos: tipo y longitud. Existen cuatro tipos principales: numérico, carácter, complejo, y lógico. Existen otros tipos, pero no representan datos como tales, por ejemplo, funciones y expresiones. La longitud es el número de elementos que tiene el objeto. Para observar el tipo y la longitud de un objeto se pueden usar las funciones `mode` y `length`, respectivamente:

```

> x <- 1
> mode(x)
[1] "numeric"
> length(x)
[1] 1

```

Cuando un dato no está disponible se despliega "NA" (del inglés "not available"). Los datos numéricos que son muy grandes se pueden expresar en notación exponencial:

```

> N <- 2.1e23
> N
[1] 2.1e+23

```

R representa $\pm\infty$ como Inf y -Inf, o valores que no son numéricos como "NaN" (del inglés "not a number").

```

> x <- 5/0
> x
[1] Inf
> x - x
[1] NaN

```

Variables que necesitan ser representadas como caracteres se delimitan con comillas. La siguiente tabla resume los tipos de objetos y los datos que representan.

objeto	tipos	se combinan ?
vector	numérico, caracter, complejo o lógico	No
factor	numérico o caracter	No
arreglo	numérico, caracter, complejo o lógico	No
matriz	numérico, caracter, complejo o lógico	No
data.frame	numérico, caracter, complejo o lógico	Si
ts	numérico, caracter, complejo o lógico	Si
lista	numérico, caracter, complejo, lógico función, expresión, . . .	Si

La descripción de cada objeto es la siguiente:

- Un vector tiene la misma conotación que en matemáticas.
- Un factor es una variable categórica.
- Un arreglo es una tabla de dimensión k .
- Una matriz es un caso particular de un arreglo donde $k = 2$.
- Un `data.frame` (marco o base de datos) es una tabla compuesta de uno o más vectores y/o factores de la misma longitud pero que pueden ser de diferentes tipos.
- Un `'ts'` es una serie temporal y como tal contiene atributos adicionales como frecuencia y fechas.
- Una lista puede contener cualquier tipo de objeto incluyendo otras listas.

2.4.2. Leyendo datos desde un archivo

R utiliza el directorio de trabajo para leer y escribir archivos. Para saber cual es este directorio puede utilizar el comando `getwd()` (get working directory) Para cambiar el directorio de trabajo, se utiliza la función `setwd()`; por ejemplo, `setwd("C:/data")`. Es necesario proporcionar la dirección completa del archivo si este no se encuentra en el directorio de trabajo.

R puede leer datos guardados como archivos de texto (ASCII) con las siguientes funciones: `read.table` (con sus variantes), `scan` y `read.fwf`. R también puede leer archivos en otros formatos (Excel, SAS, SPSS, ...), y acceder a bases de datos.

La función `read.table` crea un marco de datos ('data frame') y constituye la manera más usual de leer datos en forma tabular. Por ejemplo:

```
> tareas <- read.table("data.dat")
```

Crearé un marco de datos denominado `tareas`, y cada variable recibirá por defecto el nombre `V1`, `V2`, ... y puede ser accesada individualmente escribiendo `tareas$V1`, `tareas$V2`, ... , o escribiendo `tareas["V1"]`, `tareas["V2"]`, ... , o, también escribiendo `tareas[,1]`, `tareas[,2]`, ... (la diferencia es que `tareas$V1` y `tareas[, 1]` son vectores mientras que `tareas["V1"]` es un marco de datos).

Un ejemplo de la función `read.table` es:

```
read.table(file, header = FALSE, sep = "", quote = "\"'", dec =
".", row.names, col.names, as.is = FALSE, na.strings = "NA",
colClasses = NA, nrows = -1, skip = 0, check.names = TRUE, fill =
!blank.lines.skip, strip.white = FALSE, blank.lines.skip = TRUE,
comment.char = "#")
```

file es el nombre del archivo, posiblemente con su dirección si se encuentra en un directorio diferente al de trabajo (el símbolo '\ ' no es permitido y debe reemplazarse con '/', inclusive en Windows), o una dirección remota al archivo tipo URL (http://...).

header es una variable lógica (FALSE (falso) o TRUE (verdadero)) que indica si el archivo contiene el nombre de las variables en la primera fila o línea.

sep es el separador de campo usado en el archivo; por ejemplo sep="\t" si es una tabulación.

quote son los caracteres usados para citar las variables en modo caracter.

dec es el caracter usado para representar el punto decimal.

row.names es un vector con los nombres de las líneas de tipo caracter o numérico (por defecto: 1, 2,3, ...).

col.names es un vector con los nombres de las variables (por defecto: V1, V2,...).

as.is controla la conversión de variables tipo caracter a factores (si es FALSE) o las mantiene como caracteres (TRUE); **as.is** puede ser un vector lógico o numérico que especifique las variables que se deben mantener como caracteres.

na.strings es el valor con el que se codifican datos ausentes (convertido a NA).

colClasses es un vector de caracteres para las columnas.

nrows es el número máximo de líneas a leer (se ignoran valores negativos).

skip es el número de líneas ignoradas antes de leer los datos.

check.names si es TRUE, chequea que el nombre de las variables sea válido para R

fill si es TRUE y todas las filas no tienen el mismo número de variables, agrega "blancos".

strip.white (condicional a sep) si es TRUE, borra espacios extra antes y después de variables tipo caracter.

blank.lines.skip si es TRUE, ignora líneas en "blanco".

comment.char es un caracter que define comentarios en el archivo de datos; líneas que comiencen con este caracter son ignoradas en la lectura (para desactivar este argumento utilice comment.char = "").

Las variantes de **read.table** son útiles ya que vienen con diferentes opciones por defecto:

```
read.csv(file, header = TRUE, sep = ",", quote="\"", dec=".",
```

```

fill = TRUE, ...)
  read.csv2(file, header = TRUE, sep = ";", quote="\\"", dec=",",
fill = TRUE, ...)
  read.delim(file, header = TRUE, sep = "\t", quote="\\"", dec=".",
fill = TRUE, ...)
  read.delim2(file, header = TRUE, sep = "\t", quote="\\"", dec=",",
fill = TRUE, ...)

```

2.4.3. Guardando datos

La función `write.table` guarda el contenido de un objeto en un archivo.

```

write.table(x, file = "", append = FALSE, quote = TRUE, sep = " ",
eol = "\n", na = "NA", dec = ".", row.names = TRUE, col.names =
TRUE, col.name = TRUE, qmethod = c("escape", "double"))

```

x es el nombre del objeto a exportar.

file es el nombre del archivo (por defecto, el objeto se muestra en la pantalla).

append, si es TRUE anexa los datos al archivo sin borrar datos ya existentes en el mismo.

quote, lógico o numérico: si es TRUE variables de tipo caracter y factores se escriben entre; si es un vector numérico, éste indica el número de las variables a ser mostradas entre (en ambos casos los nombres de las variables se escriben entre pero no si `quote = FALSE`).

sep es el separador de campo utilizado en el archivo.

eol es el caracter que indica el final de línea ("`\n`" es "retorno").

na es el caracter a usarse para datos faltantes.

dec es el caracter usado para el punto decimal.

row.names es una opción lógica que indica si los nombres de las líneas se escriben en el archivo.

col.names sirve para identificar los nombres de las columnas.

qmethod, si es `quote=TRUE`, especifica la manera como se debe tratar las comillas dobles "en variables tipo caracter: si es "escape"(o "e", por defecto) cada "es reemplazada por \.

Una manera sencilla de escribir los contenidos de un objeto en un archivo es utilizando

el comando `write(x, file="data.txt")`, donde `x` es el nombre del objeto (que puede ser un vector, una matrix, o un arreglo). Esta función tiene dos opciones: `nc` o `ncol` que define el número de columnas en el archivo.

2.4.4. Generación de datos

2.4.4.1. Secuencias regulares

Una secuencia regular de números enteros, por ejemplo de 1 hasta 30, se puede generar con:

```
> x <- 1:30
```

El vector resultante `x` tiene 30 elementos. El operador ":" tiene prioridad sobre otros operadores aritméticos en una expresión:

```
> 1:10-1
[1] 0 1 2 3 4 5 6 7 8 9
> 1:(10-1)
[1] 1 2 3 4 5 6 7 8 9
```

La función `seq` puede generar secuencias de números reales:

```
> seq(1, 5, 0.5)
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
```

También se pueden escribir los valores directamente usando la función `c`:

```
> c(1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5)
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
```

La función `rep` crea un vector con elementos idénticos:

```
> rep(1, 25)
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

La función `sequence` crea una serie de secuencias de números enteros donde cada secuencia termina en el número o números especificado(s) como argumento(s):

```
> sequence(4:5)
[1] 1 2 3 4 1 2 3 4 5
> sequence(c(10,5))
[1] 1 2 3 4 5 6 7 8 9 10 1 2 3 4 5
```

La función `g1` (generador de niveles) genera series regulares de factores. La función tiene la forma `g1(k, n)` donde `k` es el número de niveles (o clases), y `n` es el número de réplicas en cada nivel.

```
> g1(3, 5)
[1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
```

Otras funciones que contiene R-Project son:

Distribución/función	función
Gaussian (normal)	<code>rnorm(n, mean=0, sd=1)</code>
exponencial	<code>rexp(n, rate=1)</code>
gamma	<code>rgamma(n, shape, scale=1)</code>
Poisson	<code>rpois(n, lambda)</code>
Weibull	<code>rweibull(n, shape, scale=1)</code>
Cauchy	<code>rcauchy(n, location=0, scale=1)</code>
beta	<code>rbeta(n, shape1, shape2)</code>
'Student' (t)	<code>rt(n, df)</code>
Fisher-Snedecor (F)	<code>rf(n, df1, df2)</code>
Pearson (X^2)	<code>rchisq(n, df)</code>
binomial	<code>rbinom(n, size, prob)</code>
geométrica	<code>rgeom(n, prob)</code>
hypergeométrica	<code>rhyper(nn, m, n, k)</code>
logística	<code>rlogis(n, location=0, scale=1)</code>
lognormal	<code>rlnorm(n, meanlog=0, sdlog=1)</code>
binomial negativa	<code>rnbinom(n, size, prob)</code>
uniforme	<code>runif(n, min=0, max=1)</code>
Estadístico de Wilcoxon's	<code>rwilcox(nn, m, n), rsignrank(nn, n)</code>

2.4.4.2. Secuencias aleatorias

R tiene la capacidad de generar datos aleatorios para un gran número de funciones y distribuciones, por ejemplo, las funciones `rfunc(n, p1, p2, ...)`, donde `r` es la función aleatoria; `func` indica la distribución; `n` es el número de datos generados; y `p1`, `p2`, `p3`, ... son los parámetros de la distribución. Si se reemplaza la letra `r` por alguna de las letras `d`, `p` ó `q` se obtiene: la densidad de probabilidad (`dfunc(x, ...)`), la densidad de probabilidad acumulada (`pfunc(x, ...)`), y el valor del cuartil (`qfunc(p, ...)`), con $0 < p < 1$).

2.4.5. Manipulación de objetos

2.4.5.1. Creación de objetos

Se pueden crear objetos usando el operador de asignación(`->`); También es posible generar un objeto especificando su clase, tipo, longitud, etc. Con esto se puede crear un objeto "vacío" y modificar de manera sucesiva sus elementos. Así mismo pueden crearse nuevos objetos a partir de objetos ya existentes. Como por ejemplo:

Vector. La función `vector`, que tiene dos argumentos `mode` y `length`, crea un vector cuyos elementos pueden ser de tipo numérico, lógico o carácter dependiendo del argumento especificado en `mode`. Las funciones tienen exactamente el mismo efecto y tienen un solo argumento (la longitud del vector): `numeric()`, `logical()`, y `character()`.

Factor. Un factor incluye no solo los valores correspondientes a una variable categórica. La función `factor` tiene las siguientes opciones:

```
factor(x, levels = sort(unique(x), na.last = TRUE), labels =  
levels, exclude = NA, ordered = is.ordered(x))
```

`levels` especifica los niveles del factor (por defecto los valores únicos de `x`). `labels` define los nombres de los niveles. `exclude` especifica los valores de `x` que deben excluirse de los niveles. `ordered` es un argumento lógico que especifica si los niveles del factor están ordenados. Recuerde que `x` es de tipo numérico o carácter. Ejemplos:

```

> factor(1:3)
[1] 1 2 3
> factor(1:3, labels=c("A", "B", "C"))
[1] A B C
Levels: A B C

```

Matriz. Una matriz es realmente un vector con un atributo adicional (dim) el cual a su vez es un vector numérico de longitud 2, que define el número de filas y columnas de la matriz. Se crea con la función `matrix`:

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

La opción `byrow` indica si los valores en `data` deben llenar las columnas sucesivamente (por defecto) o las filas (si `TRUE`). La opción `dimnames` permite asignar nombres a las filas y columnas.

```

> matrix(data=5, nr=2, nc=2)
[,1] [,2]
[1,] 5 5
[2,] 5 5

> x <- 1:15
> x
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> dim(x)
NULL
> dim(x) <- c(5, 3)
> x
[,1] [,2] [,3]
[1,] 1 6 11
[2,] 2 7 12
[3,] 3 8 13
[4,] 4 9 14
[5,] 5 10 15

```

Marco de datos. Un marco de datos (`'data.frame'`) se crea de manera implícita con la función `read.table`; también es posible hacerlo con la función `data.frame`. Los vectores incluidos como argumentos deben ser de la misma longitud, o si uno de ellos es más corto que los otros, es "reciclado" un cierto número de veces:

```

> x <- 1:4; n <- 10; M <- c(10, 35); y <- 2:4
> data.frame(x, n)
  x n
1 1 10
2 2 10
3 3 10
4 4 10

```

Si se incluye un factor en un marco de datos, el factor debe ser de la misma longitud que el vector (o vectores). Es posible dar nombres a las filas con la opción `row.names`.

Lista. Una lista se crea de manera similar a un marco de datos con la función `list`. No existe ninguna limitación en el tipo de objetos que se pueden incluir.

```

> L1 <- list(x, y)
> L1
[[1]]
[1] 1 2 3 4
[[2]]
[1] 2 3 4

```

Serie de tiempo. La función `ts` crea un objeto de clase "ts" (serie de tiempo) a partir de un vector (serie de tiempo única) o una matriz (serie multivariada). Las opciones que caracterizan un objeto de este tipo son:

```

ts(data = NA, start = 1, end = numeric(0), frequency = 1, deltat
= 1, ts.eps = getOption("ts.eps"), class, names)

```

data es un vector o una matriz.

start es el tiempo de la primera observación ya sea un número o un vector con dos enteros.

end es el tiempo de la última observación especificada de la misma manera que **start**.

frequency es el número de observaciones por unidad de tiempo.

deltat es la fracción del período de muestreo entre observaciones sucesivas.

ts.eps es la tolerancia para la comparación de series. Las frecuencias se consideran

iguales si su diferencia es menor que `ts.eps`

class es la clase que se debe asignar al objeto; por defecto es "ts" para una serie univariada, y `c("mts", "ts")` para una serie multivariada

names se utiliza para una serie multivariada, es un vector de tipo caracter con los nombres de las series individuales; por defecto los nombres de las columnas de data, o Serie 1, Serie 2, ...

Algunos ejemplos de series de tiempo creadas con `ts()`:

```
> ts(1:10, start = 1959)
Time Series:
Start = 1959
End = 1968
Frequency = 1
[1] 1 2 3 4 5 6 7 8 9 10
> ts(1:47, frequency = 12, start = c(1959, 2))
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1959 1 2 3 4 5 6 7 8 9 10 11
1960 12 13 14 15 16 17 18 19 20 21 22 23
1961 24 25 26 27 28 29 30 31 32 33 34 35
1962 36 37 38 39 40 41 42 43 44 45 46 47
```

Expresión. Los objetos de clase "expresión" juegan un papel fundamental en R. Una expresión es una serie de caracteres que hace sentido para R. Todos los comandos válidos son expresiones. Cuando se escribe un comando directamente en el teclado, éste es evaluado por R y ejecutado si es válido. En muchos casos, es bueno construir una expresión sin evaluarla: esto es lo que la función `expression` hace. Es posible evaluar la expresión posteriormente con `eval()`.

```
> x <- 3; y <- 2.5; z <- 1
> expl <- expression(x / (y + exp(z)))
> expl
expression(x/(y + exp(z)))
> eval(expl)
[1] 0.5749019
```

Nota: Existen funciones que convierten los objetos de una clases a otra (`as.matrix`,

`as.data.frame`, `as.ts`, `as.expression`,...). Estas funciones pueden afectar atributos diferentes al tipo durante la conversión.

2.4.5.2. Operadores

Existen tres tipos de operadores en R: Aritméticos, comparativos y lógicos como se observa en la siguiente tabla:

Operadores					
Aritméticos		Comparativos		Lógicos	
+	adición	<	menor que	! x	NO lógico
-	substracción	>	mayor que	x & y	Y lógico
*	multiplicación	<=	menor o igual que	x && y	id.
/	división	>=	mayor o igual que	x y	O lógico
^	potencia	==	igual	x y	id.
%%	módulo	!=	diferente	x or (x, y)	O exclusivo
%%/%	división de enteros				

Los operadores aritméticos y comparativos actúan sobre dos elementos ($x + y$, $a < b$). Los operadores lógicos pueden actuar sobre uno (!) o dos objetos de tipo lógico, y pueden devolver uno o varios valores lógicos.

Es necesario usar el operador “AND” para especificar una desigualdad del tipo $0 < x < 1$ la cual puede ser codificada como $0 < x \& x < 1$. La expresión $0 < x < 1$ es válida, pero no devuelve el resultado esperado debido a que ambos operadores son del mismo tipo y se ejecutan sucesivamente de izquierda a derecha. La comparación $0 < x$ se realiza primero y el valor retornado es comparado con 1 (TRUE o FALSE < 1) en este caso, el valor lógico es implícitamente forzado a numérico (1 o 0 < 1).

2.4.5.3. Cómo acceder los valores de un objeto: el sistema de indexación

Se puede acceder a los elementos del objeto selectivamente, por ejemplo: para acceder al tercer elemento de un vector x , se escribe $x[3]$. Si x es una matriz o un marco de datos el valor de la i -ésima fila y la j -ésima columna se accede con $x[i, j]$. Por ejemplo para cambiar todos los valores de la tercera columna.

```
> x[, 3] <- 10.2
```

El no especificar la fila significa que el valor 10.2 se repetirá en toda la columna. El sistema de indexación se puede generalizar para matrices con más de dos dimensiones (por ejemplo una matriz tridimensional: $x[i, j, k]$, $x[, , 3]$, ...).

La indexación también se puede usar para suprimir una o más filas o columnas. Por ejemplo, $x[-1,]$ suprime la primera fila. Es posible acceder a los valores de un elemento usando como índice una expresión comparativa.

El sistema de indexación utiliza los valores lógicos devueltos por los operadores comparativos.

```
> x <- 1:40
> s <- c(FALSE, TRUE)
> x[s]
[1] 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
```

En el caso anterior al aplicar la condición en s al vector x el primer elemento no se selecciona (FALSE), el segundo sí (TRUE), el tercero no (FALSE), etc. Para listas, un ejemplo es: `lista[[3]]`. Ésta accede al tercer elemento de lista. Si este tercer objeto es un vector, entonces, sus valores pueden ser modificados con `lista[[3]][i]`, si es una matriz en tres dimensiones con `lista[[3]][i, j, k]`, y así sucesivamente.

2.4.5.4. Funciones aritméticas simples

Existen muchas funciones en R para manipular datos. La más sencilla es "c" que concatena los objetos listados entre paréntesis. Por ejemplo:

```
> c(1:5, seq(10, 11, 0.2))
[1] 1.0 2.0 3.0 4.0 5.0 10.0 10.2 10.4 10.6 10.8 11.0
```

Los vectores pueden ser manipulados con expresiones aritméticas clásicas:

```
> x <- 1:4
> y <- rep(1, 4)
> z <- x + y
> z
```

```
[1] 2 3 4 5
> x <- 1:3
> y <- 1:2
> z <- x + y
```

Warning message: longer object length is not a multiple of shorter object length in: x+y

```
> z
[1] 2 4 4
```

Nota: cuando R ha devuelto un mensaje preventivo y no un mensaje de error la operación fue realizada. Si queremos agregar (o multiplicar) un mismo valor a todos los elementos de un vector se procede de la siguiente manera:

```
> x <- 1:4
> a <- 10
> z <- a * x
> z
[1] 10 20 30 40
```

Algunas de las funciones aritmeticas se detallan en la siguiente tabla:

sum(x)	suma de los elementos de x
prod(x)	producto de los elementos de x
max(x)	valor máximo en el objeto x
min(x)	valor mínimo en el objeto x
which.max(x)	devuelve el índice del elemento máximo de x
which.min(x)	devuelve el índice del elemento mínimo de x
range(x)	rango de x o c(min(x), max(x))
length(x)	número de elementos en x
mean(x)	promedio de los elementos de x
median(x)	mediana de los elementos de x
var(x) o cov(x)	varianza de los elementos de x (n-1)
cor(x)	matriz de correlación
var(x,y) o cov(x,y)	covarianza entre x y y, o entre las columnas de x y y
cor(x,y)	correlación lineal entre x y y

Estas funciones devuelven un solo valor (o un vector de longitud 1), a excepción de `range()` que da por resultado un vector de longitud 2, y `var()`, `cov()`, y `cor()` que pueden ser matrices.

Las siguientes funciones pueden resultar en vectores más complejos:

<code>round(x,n)</code>	redondea los elementos de <code>x</code> a <code>n</code> cifras decimales
<code>rev(x)</code>	invierte el orden de los elementos en <code>x</code>
<code>sort(x)</code>	ordena los elementos de <code>x</code> en orden ascendente
<code>rank(x)</code>	alinea los elementos de <code>x</code>
<code>log(x, base)</code>	calcula el logaritmo de <code>x</code> en base " <code>base</code> "
<code>scale(x)</code>	si <code>x</code> es una matriz, centra y reduce los datos
<code>pmin(x,y,...)</code>	un vector en el que el <code>i</code> -ésimo elemento es el mínimo de <code>x[i]</code> , <code>y[i]</code> , . .
<code>pmax(x,y,...)</code>	igual que el anterior pero para el máximo
<code>cumsum(x)</code>	un vector en el que el <code>i</code> -ésimo elemento es la suma desde <code>x[1]</code> a <code>x[i]</code>
<code>cumprod(x)</code>	igual que el anterior pero para el producto
<code>cummin(x)</code>	igual que el anterior pero para el mínimo
<code>cummax(x)</code>	igual que el anterior pero para el máximo
<code>match(x,y)</code>	devuelve un vector de la longitud que <code>x</code> con los elementos y NA
<code>which(x == a)</code>	devuelve un vector de los índices de <code>x</code> si la operación es (TRUE)
<code>choose(n,k)</code>	calcula el número de combinaciones de <code>k</code> eventos en <code>n</code> repeticiones
<code>na.omit(x)</code>	elimina las observaciones con datos ausentes (NA)
<code>na.fail(x)</code>	devuelve un mensaje de error si <code>x</code> contiene por lo menos un NA
<code>unique(x)</code>	si <code>x</code> un vector, devuelve otro eliminando elementos duplicados
<code>table(x)</code>	devuelve una tabla con el número de diferentes valores de <code>x</code>
<code>subset(x, ...)</code>	devuelve una selección de <code>x</code> con respecto al criterio
<code>sample(x, size)</code>	remuestrea al azar y sin reemplazo

2.4.5.5. Cálculos con matrices

R posee habilidades para hacer operaciones con matrices. Las funciones `rbind()` y `cbind()` unen matrices. `rbind` une filas y `cbind` une columnas

```
> m1 <- matrix(1, nr = 2, nc = 2)
```

```

> m2 <- matrix(2, nr = 2, nc = 2)
> rbind(m1, m2)
[,1] [,2]
[1,] 1 1
[2,] 1 1
[3,] 2 2
[4,] 2 2
> cbind(m1, m2)
[,1] [,2] [,3] [,4]
[1,] 1 1 2 2
[2,] 1 1 2 2

```

El operador para el producto de dos matrices es '%*%'. Por ejemplo, considerando las dos matrices m1 y m2:

```

> rbind(m1, m2) %*% cbind(m1, m2)
[,1] [,2] [,3] [,4]
[1,] 2 2 4 4
[2,] 2 2 4 4
[3,] 4 4 8 8
[4,] 4 4 8 8
> cbind(m1, m2) %*% rbind(m1, m2)
[,1] [,2]
[1,] 10 10
[2,] 10 10

```

Para transponer una matriz se recurre a la función `t`. Esta función también se utiliza con marcos de datos. La función `diag` se puede usar para extraer o modificar la diagonal de una matriz o para construir una matriz diagonal:

```

> diag(m1)
[1] 1 1
> diag(rbind(m1, m2) %*% cbind(m1, m2))
[1] 2 2 8 8
> diag(m1) <- 10
> m1
[,1] [,2]
[1,] 10 1

```

```

[2,] 1 10
> v <- c(10, 20, 30)
> diag(v)
[,1] [,2] [,3]
[1,] 10 0 0
[2,] 0 20 0
[3,] 0 0 30
> diag(2.1, nr = 3, nc = 5)
[,1] [,2] [,3] [,4] [,5]
[1,] 2.1 0.0 0.0 0 0
[2,] 0.0 2.1 0.0 0 0
[3,] 0.0 0.0 2.1 0 0

```

La función `solve()` permite invertir una matriz. `qr()` la descompone. `eigen()` calcula sus valores y vectores propios. `svd()` hace descomposición singular.

2.5. Las gráficas en R

R ofrece una gran variedad de gráficos. Algunos de ellos se pueden ver con el comando `demo(graphics)`. Pero cada función gráfica en R tiene un enorme número de opciones lo que permite una gran flexibilidad en la producción de gráficos. El resultado de una función gráfica no se asigna a un objeto sino que es enviado a un dispositivo gráfico exclusivo (ventana o archivo).

Existen dos tipos de modificadores para las gráficas: *los modificadores o comandos de alto nivel*, que crean una gráfica y *las funciones de graficación de bajo nivel* que agregan elementos a una gráfica ya existente.

2.5.1. Manejo de gráficos

2.5.1.1. Abriendo múltiples dispositivos gráficos

Al ejecutarse una función gráfica en R se abre una ventana para mostrar el gráfico. El tipo de dispositivos gráficos disponibles depende del sistema operativo. En ambos, Windows y Unix/Linux, se puede abrir una nueva ventana gráfica con el comando `x11()`.

La lista de dispositivos gráficos disponibles se obtiene con el comando `?device` y la función `dev.list()` Muestra una lista con los dispositivos abiertos:

```
> x11(); x11(); pdf()  
> dev.list()  
x11 x11 pdf  
2 3 4
```

Para saber cual es el dispositivo activo se usa la función:

```
> dev.cur()  
pdf  
4
```

y para cambiar el dispositivo activo:

```
> dev.set(3)  
x11  
3
```

La función `dev.off()` cierra el dispositivo. R muestra el número del nuevo dispositivo activo:

```
> dev.off(2)  
x11  
3  
> dev.off()  
pdf  
4
```

El menú "History" se puede seleccionar cuando la ventana gráfica está activa y permite grabar todas las gráficas durante una sesión.

2.5.2. Funciones gráficas

La tabla a continuación resume algunas de las funciones gráficas en R.

<code>plot(x)</code>	gráfica los valores de x (en el eje y) ordenados en el eje x
<code>plot(x, y)</code>	gráfico bivariado de x (en el eje x) y y (en el eje y)
<code>sunflowerplot(x,y)</code>	igual a <code>plot()</code> pero los puntos se dibujan como flores
<code>piechart(x)</code>	gráfico circular tipo 'pie'
<code>boxplot(x)</code>	gráfico tipo 'box-and-whiskers'
<code>stripplot(x)</code>	gráfico de los valores de x en una línea
<code>coplot(x~y z)</code>	gráfico bivariado de x y y para cada valor o intervalo de z
<code>interaction.plot(f1, f2, y)</code>	gráfica el promedio de y con respecto a $f1$ y $f2$
<code>matplot(x,y)</code>	gráfica bivariada por columnas de x vs. y
<code>dotplot(x)</code>	gráficos apilados fila por fila y columna por columna
<code>fourfoldplot(x)</code>	visualiza la asociación entre dos variables dicotómicas
<code>assocplot(x)</code>	muestra desviaciones de independencia de filas y columnas
<code>mosaicplot(x)</code>	los residuales de una regresión log-lineal
<code>pairs(x)</code>	las posibles gráficas bivariadas entre las columnas de x
<code>plot.ts(x)</code>	gráfica x con respecto al tiempo
<code>ts.plot(x)</code>	igual a la anterior con diferentes fechas
<code>hist(x)</code>	histograma de las frecuencias de x
<code>barplot(x)</code>	histograma de los valores de x
<code>qqnorm(x)</code>	cuartiles de x con respecto a una distribución normal
<code>contour(x, y, z)</code>	gráfico de contornos
<code>qqplot(x, y)</code>	cuartiles de y con respecto a los cuartiles de x
<code>filled.contour(x, y, z)</code>	igual al anterior, pero las áreas están coloreadas
<code>image(x, y, z)</code>	igual al anterior pero se grafican los datos crudos
<code>persp(x, y, z)</code>	igual al anterior pero en perspectiva
<code>stars(x)</code>	dibuja una gráfica con segmentos o una estrella
<code>symbols(x, y,...)</code>	dibuja, en las coordenadas dadas por x y y
<code>termplot(mod.obj)</code>	gráfica de los efectos de un modelo de regresión

Las opciones y argumentos para cada una de estas funciones se pueden encontrar con la "ayuda" incorporada en R. Las principales (con sus valores por defecto) son:

add=FALSE Si es TRUE superpone el gráfico en el ya existente (si existe)

axes=TRUE Si es FALSE no dibuja los ejes ni la caja del gráfico

type="p" Especifica el tipo de gráfico; "p": puntos, "l": líneas, "b": puntos conectados por líneas, "o": igual al anterior, pero las líneas están sobre los puntos, "h": líneas verticales, "s": escaleras, los datos se representan como la parte superior de las líneas verticales, "S": igual al anterior pero los datos se representan como la parte inferior de las líneas verticales

xlim=, ylim= Especifica los límites inferiores y superiores de los ejes; por ejemplo con `xlim=c(1, 10)` o `xlim=range(x)`

xlab=, ylab= Títulos en los ejes; deben ser variables de tipo carácter

main= Título principal; debe ser de tipo carácter

sub= Sub-título (escrito en una letra más pequeña)

2.5.3. Comandos de graficación de bajo nivel

R posee un conjunto de funciones gráficas que afectan una gráfica ya existente, llamado comandos de graficación de bajo nivel. Estos son los principales:

<code>points(x, y)</code>	agrega puntos (se puede usar la opción <code>type=</code>)
<code>lines(x, y)</code>	igual a la anterior pero con líneas
<code>text(x, y, labels, ...)</code>	agrega texto dado por <code>labels</code> en las coordenadas <code>(x,y)</code>
<code>mtext(text, side=3, line=0, ...)</code>	agrega texto dado por <code>text</code> en el margen dado por <code>side</code>
<code>segments(x0,y0, x1, y1)</code>	dibuja una línea del punto <code>(x0,y0)</code> al punto <code>(x1,y1)</code>
<code>arrows(x0, y0,x1, y1)</code>	igual al anterior pero con flechas
<code>abline(a,b)</code>	dibuja una línea con pendiente <code>b</code> e intercepto <code>a</code>
<code>abline(h=y)</code>	dibuja una línea horizontal en la ordenada <code>y</code>
<code>abline(v=x)</code>	dibuja una línea vertical en la abscisa <code>x</code>
<code>abline(lm.obj)</code>	dibuja la línea de regresión dada por <code>lm.obj</code>
<code>rect(x1, y1,x2, y2)</code>	dibuja un rectángulo con las coordenadas dadas
<code>polygon(x, y)</code>	dibuja un polígono uniendo los puntos dados por <code>x</code> y <code>y</code>
<code>legend(x, y, legend)</code>	agrega la leyenda en el punto <code>(x,y)</code>
<code>title()</code>	agrega un título y opcionalmente un sub-título
<code>axis(side, vect)</code>	agrega un eje según <code>side</code> , <code>vect</code> dice donde dibujarlo
<code>rug(x)</code>	dibuja los datos <code>x</code> en el eje <code>x</code> como pequeñas líneas
<code>locator(n,type="n", ...)</code>	devuelve las coordenadas <code>(x,y)</code> después de <code>n</code> clicks
<code>identify(x,...)</code>	imprime en el valor de <code>x</code> (leyenda especificada)

Nota: la posibilidad de agregar expresiones matemáticas en una gráfica se hace con `text(x, y, expression(...))`, donde la función `expression` transforma su argumento en una ecuación matemática. Por ejemplo:

```
> text(x, y, expression(p == over(1, 1+e^-(beta*x+alpha))))
```

Se verá en la gráfica como la siguiente ecuación en el punto de coordenadas (x,y):

$$p = \frac{1}{1 + e^{-(\beta x + \alpha)}}$$

Para incluir una variable en una expresión, se utilizan las funciones `substitute` y `as.expression`, por ejemplo, para incluir el valor de R^2 (calculado anteriormente y guardado en un objeto `Rsquared`), se escribe:

```
> text(x, y, as.expression(substitute(R^2==r, list(r=Rsquared))))
```

Se verá en la gráfica en el punto con coordenadas (x,y) la expresión:

$$R^2 = .9856298$$

2.5.4. Parámetros gráficos

La presentación de gráficos se puede mejorar con parámetros gráficos adicionales. La función `par` cambia de manera permanente parámetros gráficos

La lista completa de parámetros gráficos se puede ver con `?par`; los más usados son:

adj	controla la justificación del texto
bg	especifica el color del fondo
bty	controla el tipo de caja que se dibuja alrededor del gráfico
cex	un valor que controla el tamaño del texto y símbolos
col	controla el color de los símbolos
font	un valor que controla el estilo del texto
las	un valor que controla la orientación de los caracteres en los ejes
lty	un valor o caracter que controla el tipo de las líneas
lwd	un número que controla la anchura de las líneas
mar	un vector que controla el espacio entre los ejes y el borde de la gráfica
mfcpl	un vector c(nr,nc) que divide la gráfica como una matriz en columnas
mfrow	igual al anterior, pero las gráficas se dibujan en las filas
pch	controla el tipo de símbolo
ps	un valor que controla el tamaño (en puntos) de textos y símbolos
pty	un caracter que especifica el tipo de región a graficar
tck	un valor para los marcadores de eje como según altura y anchura
tcl	un valor para los marcadores de eje como según altura de una línea
xaxt	xaxt="n" el eje x se coloca pero no se muestra
yaxt	si yaxt="n" el eje y se coloca pero no se muestra

2.6. Análisis estadísticos con R

Al igual que los gráficos, es imposible ahondar en detalles acerca de todas las posibilidades ofrecidas por R para realizar análisis estadísticos. Una visión muy general de las características de R para realizar estos análisis están localizadas en el paquete base, pero la gran mayoría de los métodos estadísticos disponibles en R están distribuidos como paquetes (*packages*). Algunos de éstos vienen instalados junto con *base*, otros están dentro del grupo *recommended*, dado que son los métodos comunmente utilizados en estadística, y finalmente mucho otros son adicionales y están dentro del grupo *contributed* y debe ser instalados por el usuario.

La siguiente tabla muestra los paquetes que se distribuyen en la "base". Con excepción de *ctest* que se carga en memoria cuando R comienza, cada uno de estos puede ser utilizado después de haber sido cargado:

```
> library(eda)
```

La lista de las funciones disponibles en un paquete se puede ver escribiendo:

Paquete	Descripción
ctest	pruebas clásicas ('Student', Wilcoxon, Pearson, Kolmogorov-Smirnov,...)
eda	ajuste lineal robusto y ajuste de medianas
lqs	regresión resistente y estimación de covarianza
methods	definición de métodos y clases para objetos en R y de programación
modreg	regresión moderna (alisamiento y regresión local)
mva	análisis multivariado
nls	regresión no-lineal
splines	representaciones polinómicas
stepfun	funciones de distribución empíricas
tcltk	funciones para hacer interfaz desde R a elementos gráficos Tcl/Tk
tools	herramientas para desarrollo y administración de paquetes
ts	análisis de series temporales

Para ver una lista completa y una descripción de los paquetes "*contributed*" visite la página web del CRAN.

Los paquetes recomendados se describen brevemente en la siguiente tabla.

Paquete	Descripción
boot	métodos de remuestreo y "bootstrapping"
class	métodos de clasificación
cluster	métodos de agregación
foreign	funciones para leer datos en diferentes formatos
KernSmooth	métodos para estimación de densidad
MASS	librerías de "Modern Applied Statistics with S-PLUS"
mgcv	modelos aditivos generalizados
nlme	modelos lineales y non-lineales con efectos mixos
nnet	redes neuronales y modelos multinomiales log-lineales
rpart	particionamiento recursivo
spatial	análisis espaciales ("kriging", covarianza espacial, . . .)
survival	análisis de sobrevivencia

El procedimiento para instalar un paquete depende del sistema operativo usado y de la manera como se instaló R: ya sea desde el código fuente o por medio de archivos binarios pre-compilados.

Existen varias funciones para manejar paquetes tales como `installed.packages()`, `CRAN.packages()` o `download.packages()`.

```
> update.packages( )
```

La función anterior verifica las versiones de los paquete instalados en el sistema y los compara con los disponibles en CRAN . De esta manera, el usuario puede actualizar sus paquetes con las versiones más recientes.

2.7. Programación práctica con R

Una vez que se ha descrito cómo funciona "R", es recomendable regresar al uso del lenguaje de programación para que el usuario diseñe sus propias funciones, que en la practica resultan muy simples.

2.7.1. Bucles y Vectorización

Una ventaja de R comparado con otros programas estadísticos con “menús y botones” es la posibilidad de programar de una manera sencilla una serie de análisis que se puedan ejecutar de manera sucesiva, que hace posible programar sin muchos conocimientos o experiencia previa en esta área.

Bucles

R posee estructuras de control que no son muy diferentes a las de un lenguaje de alto nivel como "C". Si tenemos un vector "x", (a,b,b,c) y para se quiere que para cada elemento de x con valor igual a "b", se le asigne el valor cero (0) y uno (1) para el resto, entonces, se crea un vector "y" de la misma longitud de x y se realiza la siguiente función:

```
y <- numeric(length(x))  
for (i in 1:length(x)) if (x[i] == b) y[i] <- 0 else y[i] <- 1
```

Se pueden usar corchetes para ejecutar varias instrucciones:

```
for (i in 1:length(x)) {  
y[i] <- 0  
...  
}
```

```
if (x[i] == b) {  
y[i] <- 0  
...  
}
```

Otra posibilidad es ejecutar una instrucción siempre y cuando se cumpla una cierta condición es:

```
while (mifun > minimum) {  
...  
}
```

Vectorización.

La vectorización hace los bucles implícitos en las expresiones. Por ejemplo la suma de dos vectores:

```
> z <- x + y
```

Esta suma se hubiera podido escribir como un bucle:

```
> z <- numeric(length(x))  
> for (i in 1:length(z)) z[i] <- x[i] + y[i]
```

Pero solo funciona si "x" y "y" son de la misma longitud.

Las ejecuciones condicionales (if ... else) se pueden evitar con el uso de indexación

lógica; volviendo al ejemplo anterior:

```
> y[x == b] <- 0  
> y[x != b] <- 1
```

2.7.2. Escribir un programa en R

Como se recordará los programas en R se escriben en un archivo que se guarda en formato ASCII con extensión "R". Esto permite al usuario escribir sus propias funciones, que tendrán las mismas propiedades que las del programa. Esto permite un uso flexible, eficiente y racional de R.

Un ejemplo donde se leen unos datos y se dibuja una gráfica de los mismos, quedaría:

```
mifun <- function(S, F)  
{  
  data <- read.table(F)  
  plot(data$V1, data$V2, type="l")  
  title(S)  
}
```

Para que esta función pueda ser ejecutada, es necesario cargarla en la memoria. Las líneas de la función se pueden escribir directamente desde el teclado, como cualquier otro comando, o ser copiada y pegada a un editor de texto. Si el usuario desea que su función sea cargada cada vez que comienza R, se puede guardar en un archivo especial llamado "espacio de trabajo" (del inglés "workspace") .RData cargará en memoria automáticamente si se encuentra en el directorio de trabajo de R.

Una vez que la función es cargada se puede ejecutar con un solo comando como por ejemplo: `mifun("swallow", "Swal.dat")`.

3. Aprender Demografía con R-Project

Con base en la información de población del Estado de Nuevo León se hace un estudio demográfico para los años de 1990 y 2000. Para esto se aplica el programa R-Project, se hace una evaluación y corrección de la información y un análisis de los componentes demográficos: mortalidad, fecundidad y migración.

3.1. Evaluación de la información

Antes de proceder a un análisis de los datos de población, estos últimos deben ser evaluados para determinar su calidad. Esto se debe a que cuando se levanta un censo es común que las edades declaradas por los censados se vean afectadas por los diferentes tipos de errores.

Para hacer la evaluación y la corrección de la información se utilizan los datos de la población total y su distribución según sexo del Estado de Nuevo León del año de estudio en este caso 1990 y 2000.

En este capítulo se determina si los datos son confiables, aproximados, malos o muy malos. con la ayuda de tres índices: Índice de Whipple, Índice de Myers, e Índice de las Naciones Unidas.

En los ejemplos que a continuación se presentan, los datos se han guardado en archivos `csv` utilizando un formato de dos columnas cuyos títulos son “Hombres” y “Mujeres”. El primer número que se presenta en las columnas representa el total de cada población; los quinquenios vienen intercalados con las edades individuales y el último período es de 100 *años y más*, al final se presentan las personas que no especificaron su edad.

Los archivos de nuestro ejemplo se guardan en el disco duro con el nombre de:

- NL_90.csv para los datos de Nuevo León del año 1990, es decir, el año z
- NL_00.csv para los datos de Nuevo León del año 2000, es decir, el año z+10.

El comando utilizado en R-Project:

```

>#importar datos: población censo 1990(población anterior)
>dpant=read.csv("C:/NL_90.csv",header=TRUE,sep=" ",quote="\ ",
dec=".",fill = TRUE)
>#importar datos: población censo 2000(población actual)
>dpact=read.csv("C:/NL_00.csv",header=TRUE,sep=" ",quote="\ ",
dec=".",fill = TRUE)

```

Para extraer los datos de las edades individuales por sexo creamos las siguientes funciones:

```

>#Población de hombres por edades individuales del censo de 1990
>dpantH=dpant[ (c(0:nrow(dpant)) %% 6) != 0, 1]
>#Población de hombres por edades individuales del censo de 2000
>dpactH=dpact[ (c(0:nrow(dpact)) %% 6) != 0, 1]
>#Población de mujeres por edades individuales del censo de 1990
>dpantM=dpant[ (c(0:nrow(dpant)) %% 6) != 0, 2]
>#Población de mujeres por edades individuales del censo de 2000
>dpactM=dpact[ (c(0:nrow(dpant)) %% 6) != 0, 2]

```

Se puede notar que las poblaciones se manejan de igual modo pero cambian debido al cambio de género y año.

A continuación se presentan tres diferentes tasa de crecimiento para analizar el crecimiento de la población entre 1990 y 2000.

3.1.1. Las Tasas de crecimiento intercensal

Estas relaciones miden la frecuencia con que crece la población entre un censo y otro. Existen varias tasas (Aritmética, geométrica y exponencial) a escoger. La tasa más adecuada depende del tipo de trabajo que se va a realizar, para esta selección es necesario tomar en cuenta los supuestos de cada tasa y calcular el tiempo determinado entre los censos.

Para determinar el tiempo transcurrido, en años, entre los censos se calcula h

$$h = \frac{(fecha_censo_actual) - (fecha_censo_anterior)}{365}$$

Donde:

fecha_censo_actual = La fecha en que se realizó el censo más reciente en día, mes y año.

fecha_censo_anterior = La fecha en que se realizó el censo anterior en día, mes y año.

Por ejemplo:

$$h = \frac{14/feb/2000 - 12/marzo/1990}{365}$$

El comando utilizado para calcular "h" en R-Project es el siguiente:

```
># Para obtener h (períodos) según las fechas de los censos.  
>h=(mdy.date(2,14,2000)-mdy.date(3,12,1990))/365
```

Para el período entre los censos de 1990 y el 2000 se obtiene el siguiente resultado:

```
>h  
[1] 9.934247
```

Advertencia: Si R-Project no acepta la función `mdy.date` debes recurrir a las librerías que vienen en Packages →load package →date.

3.1.1.1. La Tasa aritmética

El supuesto para usar esta tasa es considerar que la población se incrementa en una constante "r" en cada período.

$$pob_actual = (pob_anterior) * (1 + rh)$$

Donde:

pob_actual = El total de la población más reciente

$pob_anterior$ = El total de la población anterior

Por ejemplo:

$$pob_2000 = pob_1990(1 + rh)$$

Los comandos utilizado en R-Project son los siguientes:

```
># La tasa de crecimiento Aritmética para hombres
>rH=((dpact[1,1]/dpant[1,1])-1)/h
># La tasa de crecimiento Aritmética para mujeres
>rM=((dpact[1,2]/dpant[1,2])-1)/h
```

Se obtiene:

```
>rH
[1] 0.02383492
>rM
[1] 0.02394361
```

El resultado nos informa que la población masculina creció a una tasa aritmética de .02383492 y la femenina a una tasa de 0.02394361

3.1.1.2. La Tasa geométrica

El supuesto que se utiliza para esta tasa es que la población crece en un año en una proporción “ r ’ ” de la proporción inicial.

$$pob_actual = (pob_anterior) * (1 + r')^h$$

Donde:

pob_actual = El total de la población más reciente

$pob_anterior$ = El total de la población anterior

Por ejemplo:

$$pob_{2000} = pob_{1990}(1 + r')^h$$

El comando utilizado en R-Project es el siguiente:

```
># La tasa de crecimiento Geométrica para hombres  
>rpH=( (dpact[1,1]/dpant[1,1])^(1/h) )-1  
># La tasa de crecimiento Geométrica para mujeres  
>rpM=( (dpact[1,2]/dpant[1,2])^(1/h) )-1
```

Se obtiene:

```
>rpH  
[1] 0.02162239  
  
>rpM  
[1] 0.02171214
```

El resultado nos informa que la población masculina creció a una tasa geométrica de 0.02162239 y la femenina a una tasa de 0.2171214

3.1.1.3. La Tasa exponencial

El supuesto que se considera para usar esta tasa es que la población crece cada instante en una proporción “ r ” de la proporción inicial.

$$pob_{actual} = (pob_{anterior}) * (e^{r^i h})$$

Donde:

pob_actual =El total de la población más reciente
pob_anterior =El total de la población anterior

Por ejemplo:

$$pob_{2000} = (e^{r^*h})pob_{1990}$$

El comando utilizado en R-Project:

```
># Las tasa de crecimiento exponencial para hombres  
>rppH=( (log(dpact[1,1]/dpant[1,1]))/h)  
  
># La tasa de crecimiento exponencial para mujeres  
>rppM=( (log(dpact[1,2]/dpant[1,2]))/h)
```

Se obtiene:

```
>rppH  
[1] 0.02139194  
>rppM  
[1] 0.02147979
```

El resultado nos informa que la población masculina creció a una tasa exponencial de 0.02139194 y la femenina a una tasa de 0.02147979

3.1.2. La Pirámide de edades individuales

Las "pirámides de edad" son una representación gráfica en forma de histograma de la distribución de la población, que representa la población por edad de cada sexo respecto a la población total.

Estas gráficas miden en las ordenadas (x) las edades y en las abscisas (y) la cantidad de personas que existen por edad en un momento dado. En las pirámides también puede distinguirse el comportamiento de cada sexo; ya que las mujeres se representan del lado derecho y los hombres del lado izquierdo de la gráfica.

El perfil general que da una pirámide de edad puede responder o no a la estructura real de la población en un determinado momento. Los detalles de estas desviaciones se aprecian con mayor claridad cuando la gráfica muestra las edades año por año. Por ejemplo, la gente cuando responde a una entrevista tiende a redondear su edad; las mujeres suelen reducirse la edad a partir de determinado momento y los hombres manipulan su edad cuando son aptos para el servicio militar.

Para detectar la magnitud de estas desviaciones en las pirámides de edad se recurre al uso de datos de la población "corregida a mitad de año".

El comando utilizado para elaborar una pirámide poblacional en R-Project es el siguiente:

```
>#función para crear las piramides de población
>piramide<-function(hombres,mujeres,amplitud,escalax,
edadmax,region){
  max1<-max(c(hombres,mujeres))
  n<-length(hombres)

  #Utilizar el factor de escala 'escalax' para el eje x:
  min.x<-(max1%/%escalax+1)*escalax
  max.x<-(max1%/%escalax+1)*escalax

  #Abrir una ventana para el gráfico
  plot(0,0,type="n",xaxt='n',yaxt='n',ylim=c(0,edadmax+5),
xlim=c(min.x,max.x),xlab="",ylab="")

  #dibujar ejes y etiquetarlos
  #permitirá que los datos de los hombres queden a la izquierda.
  ejex1<-seq(0,max1,by=escalax)
  ejex2<-ejex1[order(-ejex1)]
  ejex<-c(ejex2,ejex1)
  axis(1,at=ejex,labels=as.character(abs(ejex)),
cex.axis=0.8,las=2)
  ejey<-c(seq(0,edadmax,by=amplitud))
  axis(2,at=ejey,labels=as.character(ejey),cex.axis=0.8,las=2)

  for(i in 1:n){
    x1<-0
```

```

x2<-mujeres[i]
x3<- hombres[i]

#la escala de edad según la amplitud de los intervalos:
y1<-(i-1)*amplitud
y2<-y1+amplitud
#para dibujar la distribución de las mujeres de rojo
rect(x1,y1,x2,y2,col='red')
#para dibujar la distribución de los hombres de azul
rect(x1,y1,x3,y2,col='blue')
}

#Ubicar leyendas y título:
x.l1<-max1/16-1.5*escalax
x.l2<-max1/16+escalax
title(main=paste("PirámidePoblacional",sep="\n",region),
ylab="edad")
legend(x.l1,edadmax+5,"Hombres",bty="n",xjust=1)
legend(x.l2,edadmax+5,"Mujeres",bty="n")
}

```

Las cinco funciones arriba señaladas se describen a continuación:

1. `plot(0,0,type="n",xaxt='n',yaxt='n',ylim=c(0,edadmax+5),xlim=c(min.x,max.x),xlab=" ",ylab=" ")`

`plot` crea una gráfica con los siguientes parámetros:

Parámetros	Descripción
0	Los datos de x
0	Los datos de y
type="n"	Tipo del gráfico ninguno
xaxt='n'	El eje x no dibujado
yaxt='n'	El eje y no dibujado
ylim =c(0,edadmax+5)	El mínimo y máximo de "y"
xlim =c(min.x,max.x)	El mínimo y máximo de "x"
xlab=" "	Nombre del eje x = ninguno
ylab=" "	Nombre del eje y = ninguno

Por lo tanto esta función abre una ventana para el gráfico.

```
2. axis(1, at = ejex, labels = as.character(abs(ejex)), cex.axis = 0.8, las = 2)
```

`axis` es una función que agrega un eje a la gráfica.

Parámetros	Descripción
1	El eje se agrega abajo
at=ejex	Marcas en el eje x
labels=as.character(abs(ejex))	Etiquetas del eje
cex.axis=0.8	Espacio de las anotaciones
las=2	Orientación horizontal de la etiqueta

```
3. rect(x1, y1, x2, y2, col='red')
```

`rect` dibuja en la pirámide la secuencia de rectángulos de las coordenadas dadas, Estas se dan en orden "x" izquierdo, "y" inferior, "x" derecho, "y" inferior.

Parámetros	Descripción
col='red'	color rojo para rellenar el rectángulo

Por lo tanto dibuja de rojo a la distribución, en este caso de mujeres

```
4.title(main=paste("PirámidePoblacional", sep="\n", region), ylab="edad")
```

`title` hace las anotaciones que requiere la gráfica.

Parámetros	Descripción
main	El título principal
ylab="edad"	Etiqueta del eje y

```
5. legend(x.l1, edadmax+5, "mujeres", bty="n", xjust=1)
```

`legend` agrega textos a la gráfica.

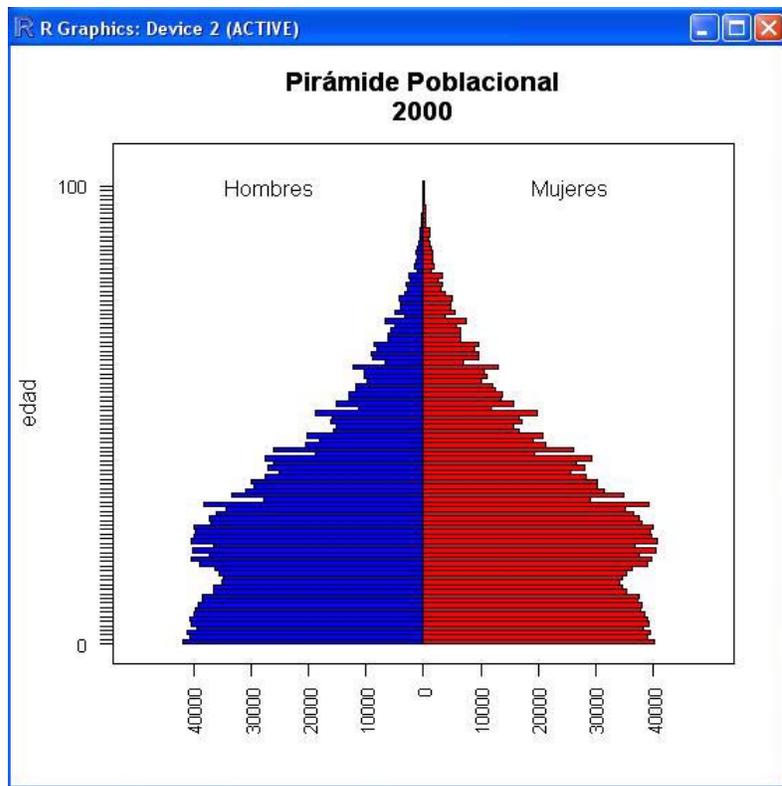
Parámetros	Descripción
x.l1	Posicion de la leyenda para x
edadmax+5	Número en el eje
"hombres"	Título de la leyenda
bty="n"	Ningún dibujo alrededor de la leyenda
xjust=1	Leyenda a la derecha justificada

Por lo tanto en este caso se escribe "hombres" en la gráfica.

La función que dibuja una pirámide poblacional en este programa es de la forma:

```
>#Pirámide de la poblacion corregida a mitad del año actual
(2000), todas las edades
>max(max(dpactMm[1:101]),max(dpactHm[1:101]))
>Pdactm<-piramide(dpactHm[1:101],dpactMm[1:101],1,10000,100,
"2000")
```

Se obtiene el siguiente gráfico:



Nota: La función se emplea la población "dpactHm" que será vista posteriormente.

3.1.3. Procedimientos para calificar los datos de la población censada

Existen tres tipos de indicadores (Whipple, Myers y Naciones Unidas) que permiten evaluar la calidad de los censos. Todos ellos parten de la idea de que la calidad de los datos pueden cuantificarse y de esta manera obtener criterios válidos para tomar decisiones en torno a ellos.

3.1.3.1. El Índice de Whipple

Este indicador utiliza los datos de la edad corregida a mitad de año a partir de los 23 años y hasta los 62, divididos por sexo. Mide también las preferencias por el dígito "0" y el "5". Este método supone que cada una de las edades individuales está distribuida de manera uniforme y se comporta de manera lineal por los grupos de edad. Así, toma las poblaciones de edades $5x$ y las multiplica por 5, pues se supone que se aproximarán a la sumas de edades $(5x - 2) + (5x - 1) + 5x + (5x + 1) + (5x + 2)$ y luego las compara con toda la población, de acuerdo a la siguiente fórmula:

$$I_w = \left(\frac{5 * \sum_{5x}^{12} P_{5x}}{\sum_{23}^{62} P_x} \right) * 100$$

Para evaluar a la población con este método las reglas de interpretación son las siguientes:

- $I_w \in [100, 105]$ los datos son *precisos*.
- $I_w \in (105, 110]$ los datos son *relativamente precisos*.
- $I_w \in (110, 125]$ los datos son *aproximados*.
- $I_w \in (125, 175]$ los datos son *malos*.
- $I_w \in (175, \text{inf})$ los datos son *muy malos*.

El comando utilizado en R-Project es el siguiente:

```
>#índice de Whipple para hombres del censo de 1990
>IWH_90=( ( 5*( sum(dpantHm[ 26 ] , dpantHm[ 31 ] , dpantHm[ 36 ] , dpantHm[ 41 ] ,
  dpantHm[ 46 ] , dpantHm[ 51 ] , dpantHm[ 56 ] , dpantHm[ 61 ] ) ) / ( sum(dpantHm
[ c( 24 : 63 ) ] ) ) * 100 ) )
```

```
>#índice de Whipple para mujeres del censo de 1990
>IWM_90=( (5*(sum(dpantMm[26],dpantMm[31],dpantMm[36],dpantMm[41],
dpantMm[46],dpantMm[51],dpantMm[56],dpantMm[61])))/(sum(dpantMm
[c(24:63)])))*100))
```

```
>#índice de Whipple para hombres del censo del 2000
>IWH_00=( (5*(sum(dpactHm[26],dpactHm[31],dpactHm[36],dpactHm[41],
dpactHm[46],dpactHm[51],dpactHm[56],dpactHm[61])))/(sum(dpactHm
[c(24:63)])))*100))
```

```
>#índice de Whipple para mujeres del censo del 2000
>IWM_00=( (5*(sum(dpactMm[26],dpactMm[31],dpactMm[36],dpactMm[41],
dpactMm[46],dpactMm[51],dpactMm[56],dpactMm[61])))/(sum(dpactMm
[c(24:63)])))*100))
```

Los resultados de nuestro ejemplo (Nuevo León) son los siguientes:

```
>IWH_90
[1] 110.6205
```

Como el dato obtenido se sitúa entre (110, 125] se concluye que los datos son "*aproximados*" para la población masculina de Nuevo León, del censo de 1990.

```
>IWH_00
[1] 109.7092
```

Como el dato obtenido se sitúa entre (105, 110] se concluye que los datos son "*relativamente precisos*" para la población masculina de Nuevo León, del censo del año 2000.

```
>IWM_90
[1] 112.5956
```

Como el dato obtenido se sitúa entre (110, 125] se concluye que los datos son

"aproximados" para la población femenina de Nuevo León, del censo del año 1990.

```
>IWM_00  
[1] 110.0303
```

Como el dato obtenido se sitúa entre (110, 125] se concluye que los datos son "aproximados" para la población femenina de Nuevo León, del censo del año 2000.

3.1.3.2. El Índice de Myers

Este índice utiliza los datos de la población corregida a mitad del año, dividido por sexos, de los 10 a los 80 años. Debido a las irregularidades que se presentan en las primeras edades, se consideran los valores para diez años y más, y para veinte años y más. Este indicador establece una región de aceptación o rechazo para cada una de las edades. La región de aceptación no debe variar más del 10% de la población original. Toda desviación de ese 10% significa que la edad medida debe rechazarse.

Así se elabora el índice de Myers $M = \sum_{j \geq 0} |M_j|$

Donde

$$M_j = \frac{a_j P_j + a'_j P'_j}{\sum_{j \geq 0} (a_j P_j + a'_j P'_j)} = 0,10$$

Si tomamos las edades hasta un límite de 100 años, puede escribirse como i el dígito de las decenas y j el de las unidades.

Entonces:

$P_j = \sum_{i \geq 1} P_{10i+j}$ = Número de personas que han declarado edades terminadas en el dígito

j dentro de la población de 10 años y más.

$P'_j = \sum_{i \geq 1} P_{10i+j'}$ = Número de personas que han declarado edades terminadas en el

dígito j dentro de la población de 20 años y más.

a_j, a'_j = Son los coeficientes creados por Myers para igualar las sumas de las

poblaciones verdaderas con las declaradas y que cumplan con las desviación del 10% para cada j .

Es decir, que el índice de Myers es la suma, en valores absolutos, de los diez índices individuales de cada dígito. Para facilitar la construcción de estos índices individuales se realiza la siguiente tabla:

j	A	B	C	D	E	F	G	G
Dígito j	P_j	P'_j	$P_j a_j$	$P'_j a'_j$	$C + D$	$\left(\frac{E}{\sum E_j} \right) * 100$	$F - 10$	G

La columna j tienen los dígitos del 0 al 9.

La columna A tienen las poblaciones de 10 y +.

La columna B tienen las poblaciones de 20 y +.

La columna C tiene los P_j y se multiplican por a_j que es uno de los índices de Myers.

La columna D tiene los P'_j y se multiplican por a'_j que es el otro de los índices.

La columna $G = M_j$ y por lo tanto si sumas toda la columna de |G| obtendrás el Índice de Myers.

Para evaluar los datos Myers da una regla de interpretación que dice que:

- Si el índice de Myers pertenece al intervalo [0, 5] los datos son *confiables*.
- Si el índice de Myers pertenece al intervalo (5, 15] los datos son *aproximados*.
- Si el índice de Myers pertenece al intervalo (15, 30] los datos son *malos*.
- Si el índice de Myers pertenece al intervalo (30, inf] los datos son *muy malos*.

Se puede calcular en R-Project cada columna de la tabla anterior. Un ejemplo: para hombres del censo de 1990 sería como sigue:

>#Se crea una función para cada fila de la columna "A" de hombres del censo de 1990.

```
AiH_90<-function(idx_init){
  imAH_90=dphantHm[(c(0 : 69) %% 10) == idx_init]
  imA1H_90=sum(imAH_90[c(2:7)])
  return (imA1H_90)
}
```

># Se crea una función para cada fila de la columna "B" de hombres del censo de 1990

```
BiH_90<-function(idx_init){  
  imBH_90=dpantHm[(c(0 : 79) %% 10) == idx_init]  
  imBlH_90=sum(imBH_90[c(3:8)])  
  return (imBlH_90)  
}
```

># Columna A de hombres del censo de 1990

```
colAH_90=c(AiH_90(0),AiH_90(1),AiH_90(2),AiH_90(3),AiH_90(4),  
AiH_90(5),AiH_90(6),AiH_90(7),AiH_90(8),AiH_90(9))
```

># Columna B de hombres del censo de 1990

```
colBH_90=c(BiH_90(0),BiH_90(1),BiH_90(2),BiH_90(3),BiH_90(4),  
BiH_90(5),BiH_90(6),BiH_90(7),BiH_90(8),BiH_90(9))
```

># Columna C de hombres del censo de 1990

```
colCH_90=c(colAH_90*((0:9)+1))
```

># Columna D de hombres del censo de 1990

```
coldH_90=c(colBH_90*(9-(0:9)))
```

># Columna E de hombres del censo de 1990

```
coleH_90=c(colCH_90+coldH_90)
```

># Columna F de hombres del censo de 1990

```
colFH_90=c(coleH_90/sum(coleH_90)*100)
```

># Columna G de hombres del censo de 1990

```
colGH_90=c(abs(colFH_90-10))
```

>#indice de Myers para hombres del censo de 1990

```
IMH_90=sum(colGH_90)
```

Los resultados de aplicar el índice de Myers sobre el Estado de Nuevo León para la población masculina del censo de 1990 fue el siguiente:

```
>IMH_90  
[1] 5.535664
```

Podemos concluir que debido a que el resultado se ubica entre las calificaciones 5 y 15, los datos pueden calificarse como "aproximados".

De igual forma, el usuario puede efectuar cálculos para las otras poblaciones y posteriormente evaluar los datos. Lo que debe hacer es cambiar los nombres de las variables que se crean, como por ejemplo AiH_90 por AiH_00, para cambiar de censo; y AiH_90 por AiM_90 para cambiar de género. Además, dentro de los comandos se utilizan diferentes variables para diferenciar a cada población por ejemplo: dpantHm que son los datos de la población masculina a mitad del año de censo de 1990; cambia por dpactMm para los datos de la población femenina a mitad de año del censo del 2000.

Entonces tenemos:

```
>#Índice de Myers para hombres del censo de 2000  
>IMH_00  
[1] 4.929964
```

Se concluye que los datos de Nuevo León del censo del año 2000 de hombres son *confiables*.

```
>#Índice de Myers para mujeres del censo de 1990  
>IMM_90  
[1] 5.916904
```

Se concluye que los datos de Nuevo León del censo del año 1990 de mujeres son *aproximados*.

```
>#Índice de Myers para mujeres del censo de 2000  
>IMM_00  
[1] 4.997174
```

Se concluye que los datos de Nuevo León del censo del año 2000 de mujeres son *confiables*.

3.1.3.3. El Índice de las Naciones Unidas (NU)

Este indicador recurre a los datos de la población corregida a mitad del año, en quinquenios; parte de los 0-4 años hasta los 84-89.

Para extraer los quinquenios de nuestra población inicial en R-Project se recurre al siguiente comando:

```
>#datos de la poblacion 2000 (actual) de hombres en quinquenios  
>dpactHQ=dpact[ (c(0:nrow(dpact)) %% 6 ) == 0, 1]
```

```
>#datos de la poblacion 2000 (actual) de mujeres en quinquenios  
>dpactMQ=dpact[ (c(0:nrow(dpact)) %% 6 ) == 0, 2]
```

```
>#datos de la poblacion 1990 (anterior) de hombres en quinquenios  
>dpantHQ=dpant[ (c(0:nrow(dpant)) %% 6 ) == 0, 1]
```

```
>#datos de la poblacion 1990 (anterior) de mujeres en quinquenios  
>dpantMQ=dpant[ (c(0:nrow(dpant)) %% 6 ) == 0, 2]
```

El índice de las Naciones Unidas nos permite conocer la calidad de los datos por edad y sexo y se genera a partir de tres índices:

- Índice de regularidad por sexo
- El coeficiente por edad de hombres
- El coeficiente por edad de mujeres

3.1.3.3.1. El Índice de regularidad por sexo

Para la construcción de este índice se elabora la siguiente tabla:

Grupos	Hombres	Mujeres	Índice de masculinidad	Incremento del Imasc
Quinquenios <i>i</i>			$\frac{\text{hombres}}{\text{mujeres}}$	$\text{Imasc}_{i-1} - \text{Imasc}_i$

Los "grupos" sirven para colocar los quinquenios desde 0-4 hasta 80-84.

La población de "hombres" es corregida a mitad de año, por quinquenios.

La población de "mujeres" es corregida a mitad de año, por quinquenios.

El "Índice de masculinidad" es la proporción de hombres con respecto a mujeres.

El "Incremento del índice de masculinidad" (Ímasc) hace la diferencia de éste con la edad siguiente.

La suma de todos los Incrementos del índice de masculinidad en valor absoluto da como resultado el índice de regularidad por sexo.

El comando utilizado en R-Project para el primer índice del año 1990 es el siguiente:

```
>#índice de masculinidad del censo de 1990
>im_90=(dpantHmQ[c(1:17)] / dpantMmQ[c(1:17)])

>#índice de regularidad por sexo del censo de 1990
>irs_90=abs(sum(im_90[2:17] - im_90[1:16]))
```

Por lo tanto se obtiene:

```
>irs_90
[1] 0.2375066
```

El comando utilizado en R-Project para el primer índice del año 2000 es el siguiente:

```
>#índice de masculinidad del censo de 2000
>im_00=(dpactHmQ[c(1:17)] / dpactMmQ[c(1:17)])

>#índice de regularidad por sexo del censo de 2000
>irs_00=abs(sum(im_00[2:17] - im_00[1:16]))
```

Por lo tanto se obtiene:

```
>irs_00
[1] 0.2601636
```

3.1.3.3.2. Coeficiente por edad y por sexo

Este coeficiente asume que los grupos de edad decrecen linealmente.

La unidad u a que se recurre para cada grupo representa la población de cualquier grupo multiplicada por dos y el resultado se divide por la suma de los grupos anterior y posterior. En otras palabras:

$$u = \frac{2P_u}{(P_{u-1} + P_{u+1})}$$

Para construir los coeficientes se requiere de una tabla como la que sigue:

Grupos	Población	Coeficiente por edad	incremento del coeficiente
Quinquenios		u	$1 - u$

Donde:

Los "grupos" de edad sirven para separar los quinquenios desde 1-4 hasta 85-89.

La "población" contiene a la población corregida a mitad del año de hombres o mujeres.

El "coeficiente por edad" de cada quinquenio es igual a u , y empieza del quinquenio 5-9 hasta 80-84.

El "Incremento del coeficiente" por edad es la diferencia con 1.

Para obtener el coeficiente se suman todos los incrementos y se le aplica de valor absoluto. Esto se realiza para hombres y para mujeres.

Los comandos en R-Project para el segundo y tercer índice son los siguientes:

```
>#El coeficiente por edad de hombres del censo de 1990
>ceH_90=abs(sum(1-(2*(dpantHmQ[c(2:17)]/(dpantHmQ[c(1:16)]+dpantHmQ
[c(3:18)]))))))
>#El coeficiente por edad de hombres del censo de 2000
>ceH_00=abs(sum(1-(2*(dpactHmQ[c(2:17)]/(dpactHmQ[c(1:16)]+dpactHmQ
[c(3:18)]))))))
```

```

>#El coeficiente por edad de mujeres del censo de 1990
>ceM_90=abs(sum(1-(2*(dpantMmQ[c(2:17)]/(dpantMmQ[c(1:16)]+dpantMmQ
[c(3:18)])))))
>#El coeficiente por edad de mujeres del censo de 2000
>ceM_00=abs(sum(1-(2*(dpactMmQ[c(2:17)]/(dpactMmQ[c(1:16)]+dpactMmQ
[c(3:18)])))))

```

Los cálculos de estos índices dan como resultado:

```

>ceH_90
[1] 0.2132681

```

```

>ceH_00
[1] 0.3084347

```

```

>ceM_90
[1] 0.157306

```

```

>ceM_00
[1] 0.2537157

```

Por lo tanto, el índice de las Naciones Unidas (I_{NU}) es el porcentaje de la media de las diferencias en valor absoluto de los tres índices presentados y se resuelve de la siguiente forma:

$$I_{NU} = \left(3 \left(\frac{IRS + CEH + CEM}{16} \right) \right) * 100$$

Donde:

IRS es el índice de regularidad por sexo

CEH es valor absoluto de la suma de las diferencias del coeficiente por edad de hombres

CEM es valor absoluto de la suma de las diferencias del coeficiente por edad de mujeres

El denominador (16) se utiliza en vez de (18) ya que representa a los quinquenios, a los cuales se les puede calcular la u .

Para la interpretación del Índice de las Naciones Unidas se utiliza la siguiente regla:

- Si $I_{NU} < 20$ los datos son *satisfactorios*.
- Si $20 \leq I_{NU} \leq 40$ los datos son *intermedios*.
- Si $I_{NU} > 40$ los datos son *deficientes*.

Los comandos en R-Project para calcular el índice para cada censo son los siguientes:

```
>#Índice de las Naciones del censo de 1990
>INU_90=(3*((irs_90+ceH_90+ceM_90)/(18-2)))*100

>#Índice de las Naciones del censo de 2000
>INU_00=(3*((irs_00+ceH_00+ceM_00)/(18-2)))*100
```

Así los índices a interpretar son:

```
>INU_90
[1] 11.40151
```

Para el índice de las Naciones Unidas los datos del censo del año 1990 para el Estado de Nuevo León son "*satisfactorios*".

```
>INU_00
[1] 15.41839
```

Para el índice de las Naciones Unidas los datos del censo del año 2000 para el Estado de Nuevo León son "*satisfactorios*".

3.1.4. Conclusiones

Aquí el alumno empieza a conocer el programa R-Project como la inserción de datos, la utilización de nuevas bibliotecas como "date", las entradas de los vectores, el uso de cada dato, además de utilizar la función para crear pirámides de edad. Esto con el motivo de que el alumno pueda practicar lo que encuentre en los manuales y al mismo tiempo conocer la

forma de evaluación de datos de una población para demografía.

El uso de los indicadores de Whipple, Myers y Naciones Unidas para la misma población permiten validar la información proporcionada por los censos de población, sin embargo, cabe destacar que cada uno de ellos llega a diferentes resultados, pero aceptables, para el caso del estado de Nuevo León. Esto se debe a que cada indicador analiza características diferentes de la población. El investigador que recurre a ellos debe tomar en cuenta las circunstancias, suposiciones y los objetivos que persigue para tomar el indicador que mejor se adecuó al problema que desea resolver.

Se recomienda que antes de seleccionar uno de los indicadores se tomen en cuenta los supuestos para evitar seleccionarlo por conveniencia.

3.2. Corrección de la información

Cuando se estudian los censos de población y se encuentran que los datos se clasifican como inexactos, el investigador puede efectuar procesos de "corrección" que permiten corregir estas desviaciones y por ende utilizarlos para posteriormente hacer inferencias válidas.

Supóngase que se encuentran con unos datos de población por edades que fueron señalados por alguno de nuestro índices como "muy inexactos", en el presente capítulo procederemos a realizar la corrección de esos datos.

El primer paso de este proceso consiste en obtener una *distribución de los datos no especificados*. Una vez que se realizó este proceso se pueden aplicar los métodos de corrección, que son tres:

- Método de agrupamiento. Consiste en agrupar a los individuos según las edades por quinquenios o por grandes grupos de edad (0-14, 15-64, 65y+).
- Método analítico. Se busca ajustar los datos a diferentes funciones matemáticas, ya sean polinomios o rectas para encontrar su mejor distribución.
- Método gráfico. Ayuda a suavizar los datos por medio de trazos libres en las pirámides de edad o por medio del método de Lexis.

3.2.1. Distribución de la población no especificada

Lo que se hace es distribuir a la población "no especificada" (*NE*) en cada edad, en función del porcentaje que cada edad represente sobre el total de la población, sin incluir la población no especificada.

Primero se calcula el porcentaje de los "no especificados" de las poblaciones, tanto de hombres como de mujeres, en cada censo. El cálculo se realiza a la población "corregida a mitad del año". La regla de interpretación dice que si $\%NE < 3\%$ entonces los datos son correctos.

Después, suponiendo que la distribución es uniforme a cada quinquenio, se le aplica la siguiente fórmula para obtener la población ajustada, es decir, P_i^{\wedge}

$$P_i^{\wedge} = P_i \left(1 + \frac{\%NE}{100} \right)$$

Donde

P_i = Es la población a mitad de año de cada quinquenio.

$\%NE$ = Es el porcentaje de los no especificados.

El comando utilizado en R-Project: para calcular el porcentaje de los "no especificados" en cada una de nuestras poblaciones es:

```
>#El porcentaje de los No Especificados hombres del censo (1990)
>NEH_90=(dpantHm[102]/(dpant[1,1]*((1+rpH)^hm90))
-dpantHm[102]))*100

>#El porcentaje de los No Especificados mujeres del censo (1990)
>NEM_90=(dpantMm[102]/(dpant[1,2]*((1+rpM)^hm90))
-dpantMm[102]))*100

>#El porcentaje de los No Especificados hombres del censo (2000)
>NEH_00=(dpactHm[102]/(dpact[1,1]*((1+rpH)^hm00))
-dpactHm[102]))*100

>#El porcentaje de los No Especificados mujeres del censo (2000)
>NEM_00=(dpactMm[102]/(dpact[1,2]*((1+rpM)^hm00))
-dpactMm[102]))*100
```

Los resultados en porcentaje para el estado de Nuevo León fueron los siguientes:

```
> NEH_90
[1] 0.5345196
```

Para hombres del censo de 1990

```
> NEM_90
[1] 0.542166
```

Para mujeres del censo de 1990

```
> NEH_00  
[1] 1.229382
```

Para hombres del censo de 2000

```
> NEM_00  
[1] 1.175636
```

Para mujeres del censo de 2000

Dado que los resultados de todos los análisis fueron menores al 3% podemos concluir que los datos de estas poblaciones son correctos y no necesita que se distribuyan a los "no especificados".

Una vez realizado el procedimientos de "distribución de los no especificados" se procede a aplicar los métodos de corrección.

3.2.2. Método de agrupamiento

Este método agrupa a las edades en quinquenios. El primer quinquenio va de 0-4 años y siguen hasta 80-84, 85 y + ó 90-99, 100 y + años.

Estos datos ya se tienen agrupados en las variables: `dpantHQ`, `dpantMQ`, `dpactHQ` y `dpactMQ`.

3.2.3. Método analítico

Existen dos procedimientos analíticos para corregir nuestro datos de manera matemática, el Promedio Móvil Simple y el Promedios Moviles ponderado, mejor conocido como Método de 1/16avo.

3.2.3.1. Promedio Móvil Simple (PMS)

Este método utiliza la población a "mitad de año". Sirve para corregir los quinquenios de 5 a 9 años y de 75 a 79 años, por sexo. Se resuelve cuando se recurre a los quinquenios 0-4, 5-9, 10-14 para corregir el primero y 70-74, 75-79, 80-84 para el segundo.

Para corregir estos quinquenios recurrimos a la siguiente fórmula:

$${}_5P^{\wedge}_{(x)} = \frac{{}_5P_{(x-5)} + {}_5P_{(x)} + {}_5P_{(x+5)}}{3}$$

Donde:

x = Toma los valores de 5 a 75 años.

P^{\wedge} = Es la población corregida en el quinquenio correspondiente.

$P_{()}$ = Es la población corregida a mitad de año del quinquenio marcado por los paréntesis.

El comando utilizado en R-Project para corregir de 5 a 9 años es el siguiente:

```
>#corrige de 5-9 años de los hombres del censo anterior (1990)  
>pms1H_90=sum(dpantHmQ[1],dpantHmQ[2],dpantHmQ[3])/3
```

Se obtiene:

```
> pms1H_90  
[1] 178875.8
```

Para el estado de Nuevo León, la fórmula informa que la población del quinquenio de hombres del censo de 1990 es de 178 876 individuos.

El comando es igual, salvo que se cambia las poblaciones correspondientes para calcular los quinquenios de 5 a 9 años de las otras poblaciones (mujeres del censo anterior (1990), y ambos sexos para el censo actual (2000)).

El siguiente proceso utilizado en R-Project corrige el quinquenio de 75 a 79 años.

```
>#corrige de 75-79 años en hombres del censo de 1990  
>pms2H_90=sum(dpantHmQ[15],dpantHmQ[16],dpantHmQ[17])/3
```

El resultado para el estado de Nuevo León es el siguiente:

```
> pms2H_90  
[1] 10347.83
```

El resultado indica que hay 10348 hombres para el quinquenio 75-79 años, del censo de 1990

3.2.3.2. Promedios Móviles Ponderados (1/16)

Para corregir los quinquenios de 10-14 hasta 70-74 se utiliza un Promedio Móvil Ponderado ("1/16"). Este método supone que la población distribuye con un polinomio de 4º grado. La ecuación demanda que se den "pesos" a cada grupo de edad, de acuerdo a la cercanía con el quinquenio a corregir. Así el grupo que se va a corregir se le da un peso de 10, y los contiguos de 4 y los siguientes de -1. Y se calcula el promedio.

La formula para el quinquenio con poblaciones de edad x hasta $x + 4$ separados por sexo y por año es:

$${}_5P_x^\wedge = \frac{-5 P_{(x-10)} + 4{}_5P_{(x-5)} + 10{}_5P_{(x)} + 4{}_5P_{(x+5)} -5 P_{(x+10)}}{16}$$

Donde

x = Es la primera edad del quinquenio a corregir.

$P_{()}$ = Es la población corregida a mitad de año del quinquenio marcado por los paréntesis.

El comando utilizado en R-Project para hacer el cálculo es el siguiente:

```
>#Promedio móvil ponderado hombres censo anterior(1990)
>pmpH_90=c((( -dpantHmQ[1:13])+(4*dpantHmQ[2:14])+(10*dpantHmQ[3:15])+(4*dpantHmQ[4:16])-(dpantHmQ[5:17])))/16)
```

Por lo que para el estado de Nuevo León en la población masculina del censo de 1990 se obtiene el siguiente vector que es la población corregida en quinquenios de 10 a 74 años por el método de 1/16.

```
> pmpH_90
[1] 189928.82 192469.09 169535.21 135340.15 110160.09
[6] 89603.81 72446.30 60752.09 49432.85 37944.34
[11] 29474.72 21896.27 14719.28
```

El mismo procedimiento se puede realizar para las tres poblaciones restantes.

3.2.4. Población corregida a la mitad del año

Para este trabajo se supone que la población crece en un año en una proporción r' de la proporción inicial, es decir, se recurre a la tasa de incremento intercensal antes calculada (véase el capítulo de Evaluación) para a su vez calcular el crecimiento geométrico de la población a mitad del año.

Posteriormente se calcula el tiempo h , definido como el período de tiempo que hay entre la fecha en que se levante el censo y lo que sería el 30 de junio del año censal.

Por ejemplo: Para el censo que se realizó el 12 de marzo de 1990 la fecha a mitad de ese año sería 30 de junio de 1990. Por lo tanto h es:

$$h = \frac{30/06/90 - 12/03/90}{365}$$

Advertencia: En R-project: se necesita una librería auxiliar para poder utilizar el formato de fechas.

El comando utilizado en R-Project para calcular h es:

```
># para obtener h para mitad de año del censo anterior (1990)
>hm90=(mdy.date(6,30,1990)-mdy.date(3,12,1990))/365
```

El resultado de este cálculo es el siguiente:

```
> hm90
[1] 0.3013699
```

Una vez calculado h y recurriendo a r' se puede estimar la población para cada una de las edades mediante la siguiente formula:

$$P_x^- = P_x(1 + r')^h$$

Donde:

P_x^- =Es la población a mitad de año para cada edad x .

P_x = Es la población original del censo.

El comando en R-Project para calcular la población corregida a mitad de año es:

```
>#Poblacion corregida a mitad del año del censo anterior (1990)
de hombres.
>dpantHm=(dpantH)*(((1+rpH)^hm90))
```

Para agrupar la población en quinquenios se necesita ubicarlos a mitad de año, por lo que les aplicamos la formula de P_x^- donde el último número es la suma de 85 y más años y el anterior es de 100 y más.

El comando utilizado en R-Project para hombres del año del censo anterior (1990)

```
>#Poblacion corregida a mitad del año del censo anterior (1990)
de hombres, en quinqueneos
>rangoy=c(0:19)
```

```

dpantHmQ=c(0:21)*0
dpantHmQ[21]=sum(c(dpantHm[97:101]))
for(k in rangoy)
{
  dpantHmQ[k+1]=sum( dpantHm[ ((k*5)+1):((k*5)+5) ] )
}
dpantHmQ[22]=sum( dpantHmQ[18:21])

```

De manera similar se calculan los quinquenio para las mujeres de ese censo y del censo actual (2000)

3.2.5. Índice de masculinidad

El índice de masculinidad (Im^t) indica cuántos hombres hay por cada 100 mujeres. Para calcular este índice se utiliza la población a mitad de año.

$$Im^t = \frac{H^t}{M^t} * 100$$

Donde:

H^t = Es el total de la población masculina en el año t .

M^t =Es el total de la población femenina en el año t .

El comando utilizado en R-Project para clacular el índice de masculinidad es el siguiente:

```

>#para el censo anterior (1990)
>Im_90=(sum(dpantHm[1:102]))/(sum(dpantMm[1:102]))*100

>#para el censo actual (2000)
>Im_00=(sum(dpactHm[1:102]))/(sum(dpactMm[1:102]))*100

```

Los resultados para el estado de Nuevo León son los siguientes:

```
> Im_90  
[1] 99.13572
```

```
> Im_00  
[1] 99.0486
```

Los resultados indican que tanto en el censo de 1990 como en el de 2000 el número de hombres por cada 100 mujeres es de 99 aproximadamente.

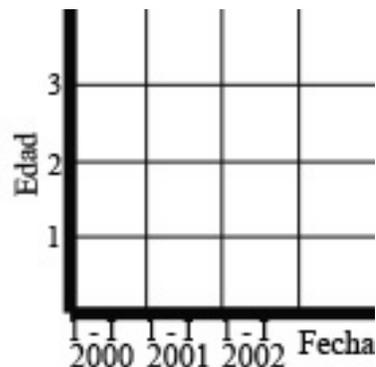
3.2.6. El Diagrama de Lexis

El diagrama de Lexis se utiliza para visualizar el comportamiento de vida de la población, ya sea por generaciones o edades o por años.

Se inicia con una sencilla y correcta delimitación del tiempo. Éste se representa mediante los parámetros:

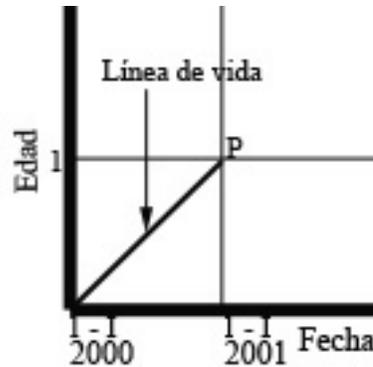
- Tiempo calendario, es decir, Fechas y,
- Duraciones o edades

El tiempo es representado por líneas rectas o semirrectas donde cada punto de la gráfica representa un instante. El diagrama de Lexis recurre a dos dimensiones para formar un cuadrante que se representa mediante un plano cartesiano. En las abscisas (X) se representan las fechas y en las ordenadas (Y) las edades.



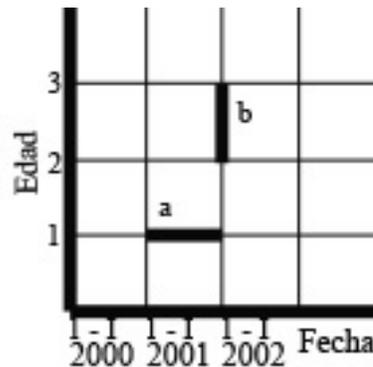
El desarrollo de una persona se dibuja con una recta en un ángulo de 45°, respecto al eje de las fechas (X). Inicia en el origen (que representa su nacimiento). A esta recta se le

llama "línea de vida". Por ejemplo, si una persona nace el 1º de enero del año 2000 tiene 0 años, exactamente. Conforme el tiempo transcurre el punto que representa a este niño se desplaza sobre la bisectriz del ángulo recto. Así al año el punto (o el niño) estará en P.

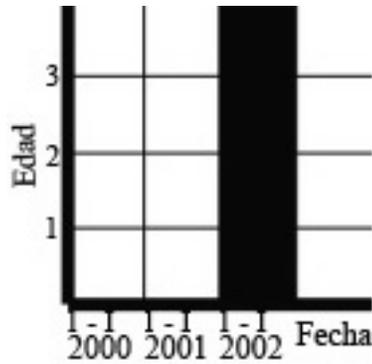


En el diagrama de Lexis a las líneas horizontales (Y) de cada edad se le conoce como "edad exacta" y coincide con el cumpleaños de la persona estudiada. Las líneas verticales (X) de cada año se les conoce como "edad cumplida" y se representa de la siguiente manera:

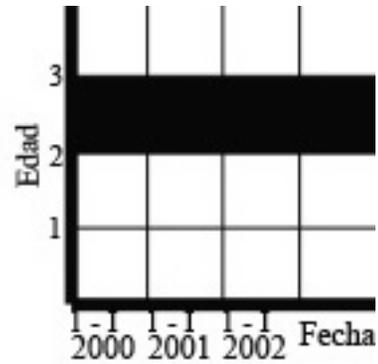
- a) Edad exacta un (1) año en el 2001
- b) Edad cumplida 2 a finales del 2001



En el diagrama de Lexis también se pueden observar las "líneas de momentos", que como su nombre lo dice, estudian en un momento dado a todas las edades de la población. También se pueden registrar las "líneas de coetaneos" que miden el crecimiento de la población por años.



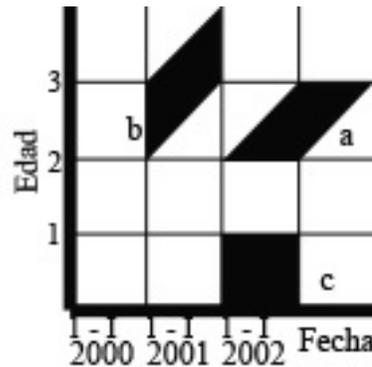
Línea de momentos



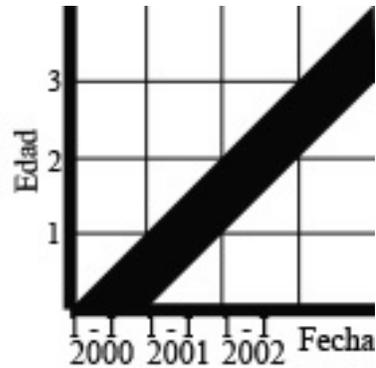
Línea de coetanio

Este diagrama permite visualizar otros fenómenos demográficos, por ejemplo:

- a) Los nacidos durante 2000, entre las edades exactas 2 y 3 años.
- b) Los nacidos en 1998, durante el año 2001.
- c) Personas de 0 años de edad cumplida, durante el año 2002.



Otra forma de agrupar a la población son las cohortes. Representan aquellas personas que comparten el mismo evento origen (nacimiento) en común y se les llama "generación". Otro tipo de evento origen es el matrimonio y se les da el nombre de "promoción".



Las magnitudes demográficas se dividen en dos clases:

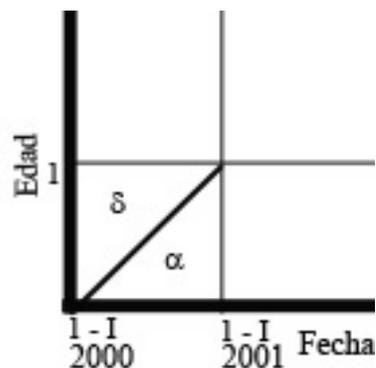
- Efectivos o stocks, que representan al instante
- Flujo, que son períodos de tiempo

Estos cálculos permiten conocer el número de personas vivas en cada momento a partir de las defunciones ocurridas en un lapso de tiempo determinado. Estas medidas no remiten al cálculo de defunciones.

3.2.6.1. Factores de separación

Para calcular las defunciones de los menores de un año y de uno a cuatro años se separan estas defunciones por cohorte o generación. Para esto existen los factores de separación (${}_n f_x$). Estos factores de separación se basan en la hipótesis de la distribución uniforme o lineal de las muertes, para cada uno de los intervalos.

Por ejemplo: En el diagrama de Lexis las defunciones de los menores a un año en el año 2000 se pueden dividir en las defunciones α (D_α) que son de las personas nacidas en el 2000 y las defunciones δ (D_δ) que son las personas de la generación anterior (1999).



Para calcular las defunciones ocurridas a los menores a un año en el año 2000 se tendría:

$$D_T = D_a + D_\delta$$

Donde:

D_T = Defunciones totales del año correspondiente.

Se utiliza como factor de separación a:

$${}_n f_x = \frac{D_\delta}{D_T}$$

En 1964 se hicieron tablas regionales para el cálculo de los factores de separación por Coale y Demeny. Para este trabajo se utiliza el modelo oeste (alta mortalidad adulta y alta mortalidad infantil) y los factores son:

	hombres	mujeres
$f_{<1}$	$0.0425 + 2.875 * q_{<1}$	$0.05 + 3 * q_{<1}$
f_{1-4}	$(1.653 - 3.013 * q_{<1})/4$	$(1.524 - 1.627 * q_{<1})/4$

Para llevar a cabo los cálculos tenemos que tener la información por años de los nacimientos y las defunciones de menores a 1 año y de 1-4 años, tanto de hombres como de mujeres.

Para trabajar en R-Project el documento se guarda como archivo `csv` y tiene 7 columnas:

En la columna 1º se inscribe el título "años" y los años van de 1984 hasta 2001.

La columna 2º lleva el título de "nacimientos H" pues son los datos referidos a los hombres.

En la columna 3º se lee el título "def<1 H" y se refiere a las defunciones de menores a un año del género masculino.

En la columna 4^o que lleva por título "def1-4 H" se incluyen las defunciones de 1 a 4 años de hombres.

La columna 5^a lleva el título "nacimientos M" y se refiere a los datos de los mujeres.

En la 6^o columna se inscribe el título "def<1 M" que son las defunciones de mujeres menores de un año.

La 7^a columna lleva el título "def1-4 M". Son las defunciones de 1 a 4 años de mujeres.

Por lo que para importar los datos y manejarlos en el programa R-Project volvemos a utilizar la función "read.csv" de la siguiente forma:

```
># importar datos para lexis
>lexis_NL=read.csv("C:/lexi_LN.csv",header=TRUE,sep=" ",quote="\\"
dec=".",fill = TRUE)
```

Una vez que el programa R-Project lee los datos, se procede a dar un nombre a cada columna. Por ejemplo, en el caso de hombres para el estado de Nuevo León se escribe:

```
># años trabajados
>años=lexis_NL[c(1:12),1]

># nacimientos de hombres
>nach=lexis_NL[c(1:12),2]

># defunciones de menores a 1 año de hombres
>def_men1H=lexis_NL[c(1:12),3]

># defunciones de 1 a 4 años de hombres
>def_1a4H=lexis_NL[c(1:12),4]
```

El siguiente paso consiste en calcular las tasas de mortalidad infantil de menores a 1 año, para luego calcular los factores de separación (muertos y vivos).

```
>#tasa de mortalidad infantil menores a 1 año de hombres
>tmort_inf1H=c(1:11)*0
tmort_inf1H[1]=0
for(i in 2:11){
```

```

tmort_inf1H[i]=
mean(def_men1H[(i-1):(i+1)])/mean(nach[(i-1):(i+1)])
}

>#factor de separación menores a un año de hombres
>fac_sep_m1H=c(tmort_inf1H*2.875+.0425)

>#factor de separación de 1 a 4 años de hombres
>fac_sep_1a4H=c((1.653-3.013*tmort_inf1H)/4)

```

A partir de esta información se crea una función para conocer la población de menores de un año y de 1 a 4 años.

Para ello, creamos una función con dos variables que sirve para calcular la edad cumplida de diferentes poblaciones.

La primera variable puede ser uno (1) si se quiere obtener la población de menores a un año ó dos (2) si se desea obtener la población de 1 a 4 años. La segunda variable se refiere al año para el cual queremos calcular la población.

En R-Project la función se escribe de la siguiente forma:

```

># función de lexis para conocer poblaciones de menores a un año
y de 1 a 4 años.
> lexis=function(m,año){
  i=año-1989
  if(m==1)
    return(pob=nach[i]-((1-fac_sep_m1H[i])*def_men1H[i]))
  else {
    if(m==2)

return(pob=sum(nach[(i-1):(i-4)])-(sum(def_men1H[(i-3):(i-1)]))
      -(sum(def_1a4H[(i-3):i])*(1-(fac_sep_1a4H[i])))
      -((1-fac_sep_m1H[i])*def_men1H[i-4])
      -(fac_sep_m1H[i]*def_men1H[i]))
    else {

```

```
        return(pob="no valido")
    }
}
}
```

Por ejemplo: Se desea conocer del estado de Nuevo León, la población de hombres, menores de un año, del año 1990:

```
> lexis(1,1990)
[1] 40778.51
```

Si lo que se quiere conocer es la población a mitad de año, del año 2000, se saca la población del año 1999 y luego se promedia con la del 2000, de la siguiente manera:

```
> a=lexis(1,2000)
> a
[1] 47272.02
> b=lexis(1,1999)
> b
[1] 46039.74
```

Una vez que se tienen estos números, se crea un vector para sacar el promedio de esos resultados. para ello utilizamos la función **mean**:

```
> x=c(a, b)
> mean(x)
[1] 46655.88
```

Por tanto, el número anterior constituye la población del estado de Nuevo León, a mitad del año 2000, de menores de un año, hombres.

Para sacar la población de 1 a 4 años se utiliza el mismo procedimiento, con los datos correspondientes.

Un ejemplo: Se quiere calcular del estado de Nuevo León, la población de hombres, menores de 1 a 4 años, del año 1990:

```
> lexis(2,2000)
[1] 176842.2
```

La población correspondiente es 176842 hombres de 1 a 4 años en 2000.

3.2.7. Pirámide de edad

En este método recurrimos a la población corregida a mitad de año. Se crean las pirámides por quinquenios y en grandes grupos de edad con la finalidad de suavizar los datos agrupados. Este método también permite hacer trazos libres para suavizar aun más los datos, si fuera necesario.

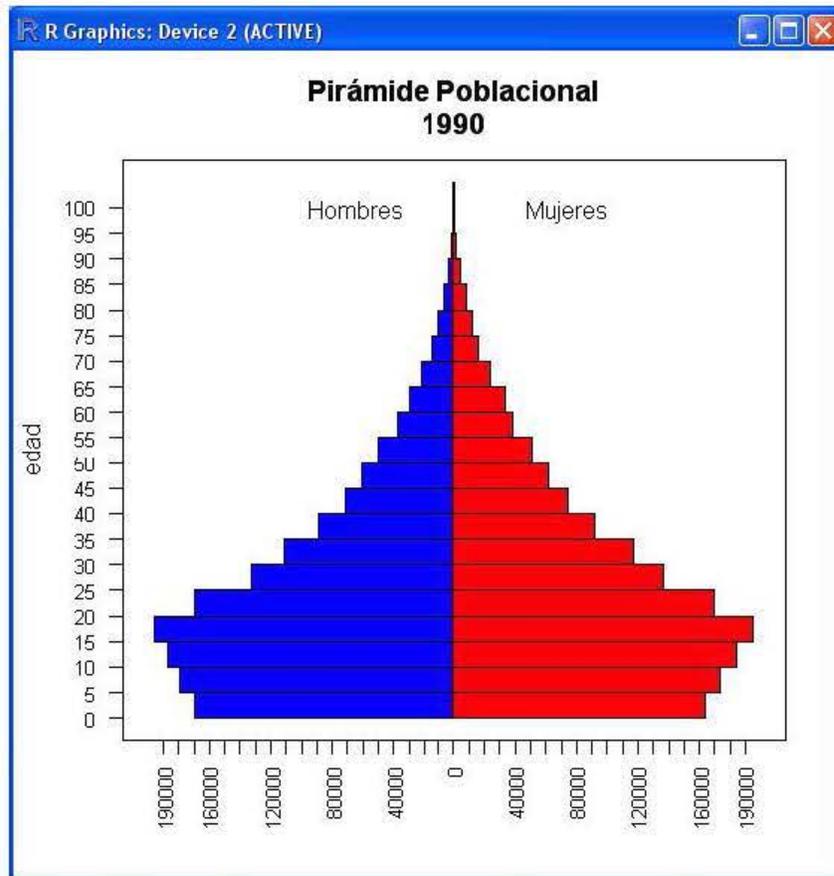
El comando que se utiliza en R-Project para crear estas piramides es el mismo que se utilizó en las piramides de edades individuales; pero se cambian los datos con los que se trabaja, es decir, se utilizan los datos agrupados en quinquenios o se forman los grandes grupos de edad.

Para la pirámide de edad en quinquenios se ocupan las poblaciones que calculamos para la población a mitad del año, es decir, para hombres del censo anterior utilizamos `dpantHmQ`. De la misma manera sería `dpantMmQ` para las mujeres del censo anterior, para los hombres del censo actual (2000) utilizamos `dpactHmQ`; en cambio, para las mujeres del censo actual (2000) recurrimos a la variable `dpactMmQ`.

Por tanto para hombres del censo anterior el comando en R-Project es:

```
>PdpantmQ<-piramide(dpantHmQ[1:21],dpantMmQ[1:21],5,10000,100,"1990")
```

La gráfica que se obtuvo para el estado de Nuevo León se puede observar en la grafica inferior:



Para construir la pirámide de "grandes grupos de edad" en R-project, el primer paso consiste en agrupar a las edades de la siguiente manera: de 0-14 años, 15-64 años y de 65 y más años separándolos por sexo. Esta fórmula se escribe de la siguiente manera:

```
>#para formar la poblacion de grandes grupos de edad del censo anterior (1990) para los hombres
```

```
>dpantHmG<-c(sum(dpantHm[1:15]),sum(dpantHm[16:66]),sum(dpantHm[67:101]))
```

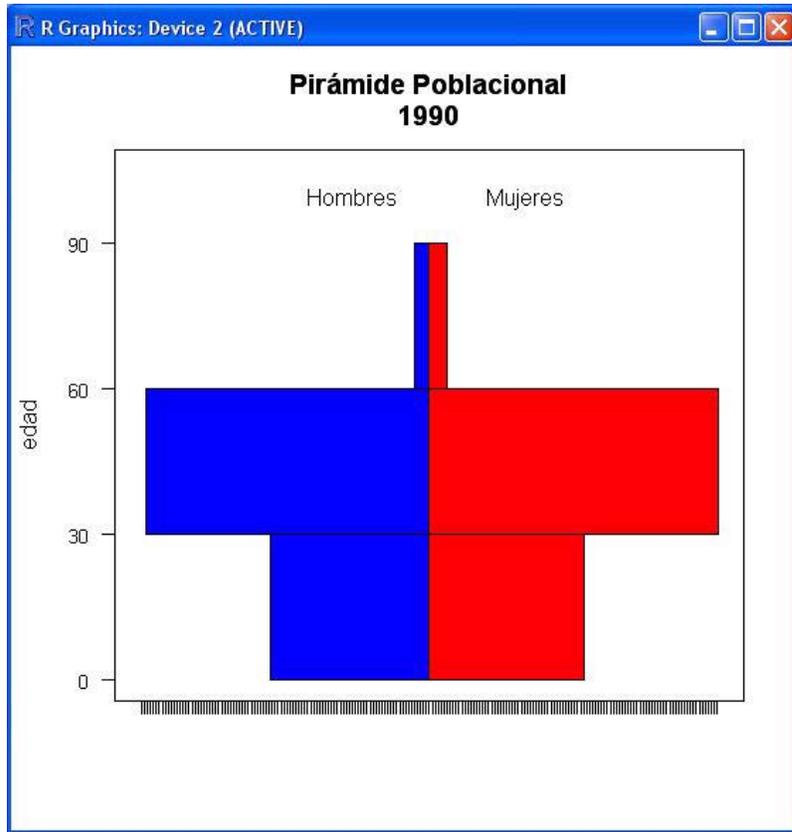
```
>#para formar la poblacion de grandes grupos de edad del censo anterior (1990) para los mujeres
```

```
>dpantMmG<-c(sum(dpantMm[1:15]),sum(dpantMm[16:66]),sum(dpantMm[67:101]))
```

El resultado para el caso del estado de Nuevo León la pirámide de grandes grupos de edad es la siguiente:

```
>PdpantmG<-piramide ( dpantHmG[1:3],dpantMmG[1:3] , 30 , 10000 , 100 , "1990" )
```

La gráfica que se dibuja adopta la siguiente forma:



3.2.8. Conclusiones

En este capítulo en R-Project se hace más sencillo el manejo de los comandos, como por ejemplo en las pirámides de edad, que sólo se utiliza la función anteriormente creada. Además se hacen pequeñas funciones para facilitar los cálculos demográficos y el uso continuo de ellos. Así el alumno puede hacer cualquier función que necesite para su futuro manejo de R-Project en cualquier tema, es decir, el alumno va personalizando su programa.

El capítulo presenta la manera de corregir los datos del estado de Nuevo León ya que como se vio en el capítulo anterior no son muy confiables. En las pirámides de edad puede apreciarse el crecimiento de la población y un ligero comienzo de la transición de la

población joven ya que la mayor cantidad de población ahora no se encuentra en los primeros años de edad. Además de conocer un método apropiado para calcular con precisión las defunciones de menores de 5 años de nuestra población ya que son edades problemáticas debido a la alta mortalidad que presentan y el método de Lexis nos permite ver una evaluación más real de la población.

3.3. Componentes demográficos

Los componentes demográficos son los sucesos que hacen que crezca o disminuya la población:

Componente Demográfico	Sucesos	Signo
<i>Mortalidad</i>	<i>Defunciones</i>	–
<i>Fecundidad</i>	<i>Nacimientos</i>	+
<i>Migración</i>	<i>Inmigración</i>	+
	<i>Emigración</i>	–

3.3.1. Mortalidad

“El concepto de mortalidad se emplea para expresar la acción de la muerte sobre los integrantes de una población” (Wolti, Carlos, 1997, p.75), es la salida de un individuo debido a su fallecimiento. Estos datos se pueden encontrar en las estadísticas vitales del país.

Para continuar con el estudio de mortalidad en R-Project se utilizarán los datos de los archivos guardados en `csv`, acomodados en cinco columnas.

La primera con título "hombres" o "mujeres" contiene los grupos de edad empezando en < 1, 1 - 4,5 - 9,... 80 - 84 hasta 85 y +.

La segunda con título del año anterior al año del censo en cuestión, por ejemplo: para la tabla de mortalidad del censo del 2000 sería "1999". El resto de la columna son los datos de las defunciones según los quinquenios.

La tercera y cuarta columna tienen de título "2000" y "2001", es decir, el año del censo y un año después con sus respectivos datos.

La quinta columna es la población a mitad de año del 2000, año del censo.

Entonces para poder trabajar con los datos tenemos que mandarlos llamar con la

funcion read.csv

El comando es:

```
># importando datos de defunciones de hombres para la tabla de mortalidad de hombres del 2000
>TMH00_NL=read.csv("C:/TMH_00.csv",header=TRUE,sep=" ",
quote="\ " ,dec=".",fill= TRUE)
```

Trabajando así los datos se tienen cuatro archivos: hombres del 2000, mujeres del 2000, hombres del 1990 y mujeres del 1990. Para nuestros ejemplos utilizaremos los datos de hombres del 2000 por lo que conviene nombrar al archivo **TMH00_NL** como **TMH**.

Para este cambio basta con hacer en R-Project

```
>TMH =TMH00_NL
```

3.3.1.1. Tasa bruta de mortalidad

Se utiliza para medir a la mortalidad dentro de una zona determinada, se puede abreviar como "TBM" y es el número de defunciones por cada mil habitantes, en un año.

Entonces la formula es:

$$TBM = \frac{D^z}{P^z} * 1000$$

Donde:

D^z =Es el número total de defunciones ocurridas en el año z de una zona.

P^z =Es la población total de la zona a mitad año.

Se multiplica por 1000 para facilitar su interpretación ya que así se evita resultados con demasiados decimales facilitando su comparación.

Esta interpretación quedaría que: “en el periodo definido, por cada mil personas fallece la cantidad que señala la tasa.” (Welti, Carlos, 1997, pp.75-76).

En R-Project es:

```
>#tasa bruta de mortalidad para hombres del año 2000  
>tbmH=(sum(sum(TMH[2]),sum(TMH[3]),sum(TMH[4]))/sum(TMH[5]))*1000
```

Da como resultado

```
> tbmH  
[1] 12.53566
```

Para las mujeres se calcula de manera similar y da como resultado:

```
> tbmM  
[1] 10.18431
```

Si hacemos lo mismo para el año 1990 la TBM de los hombres se obtiene:15.32117 y de las mujeres es: 12.00896

3.3.1.2. Tasas específicas de mortalidad

Para estudiar el comportamiento de la mortalidad a través de las edades y por quinquenios se utiliza la fórmula:

$${}_n m_x^z = \frac{{}_n D_x^z}{{}_n P_x^z}$$

Donde:

${}_n m_x^z$ = Es la tasa de mortalidad del grupo de edad x a $x + n - 1$ durante el año z

${}_n D_x^z$ = Es el número de defunciones entre la edades x y $x + n - 1$ del grupo de personas

en el año z.

${}_n P_x^z$ = Es la población a mitad del año entre las edades x y $x + n - 1$.

Es común que esta tasa se presente separada por sexo y que se utilice en el numerador el promedio de las muertes de tres años consecutivos, con el fin de suavizar irregularidades o fluctuaciones en la información básica. Como en la tabla de mortalidad.

En R-Project primero creamos una función para calcular la tasa y luego otra para crear la columna que se utilizara en la tabla de mortalidad al calcular las personas que van a fallecer de un periodo a otro.

```
># Tasa de mortalidad especifica por quinquenio de hombres
>tseH=function(i) {
  tse=c(((sum(TMH[i,2], TMH[i, 3], TMH[i,4])))/3)/TMH[i,5])
  return(tse)
}
```

3.3.1.3. Mortalidad infantil

La mortalidad infantil es la que ocurre antes de cumplir un año de vida.

La tasa para medir la mortalidad infantil es el número de defunciones de menores de un año de edad por cada mil nacimientos ocurridos en un período determinado, referido a un año y se calcula:

$$TMI = \frac{D_0^z}{B_0^z} * 1000$$

Donde:

TMI Es la tasa de mortalidad infantil.

D_0^z = Es el total de defunciones de menores de un año ocurrida en el año z.

B_0^z = Es el número de nacidos vivos en el año z.

Para utilizar R-Project se necesitan los datos de mortalidad infantil que se tienen guardados en el archivo mortalidad infantil_00.csv

Entonces se hace:

```
># Importar datos de mortalidad infantil del 2000
>mort_inf=read.csv("C:/mort_inf.csv",header=FALSE,sep=" ",
quote="\",dec=".",fill=TRUE)
```

En este documento se muestra la mortalidad infantil del año 2000 dividida en:

- Total
- Menores de 1 semana
- De 7 a 13 días
- De 14 a 20 días
- Del 21 al 29 día por día
- No especificados
- De 1 a 11 meses
- Nacidos vivos

Para calcular la tasa de mortalidad infantil del año 2000 queda:

```
> TMI=(mort_inf[1,2]/mort_inf[16,2])*1000
```

Da como resultado:

```
> TMI
[1] 13.80140
```

3.3.1.3.1. Mortalidad neonatal y postnatal

La tasa de mortalidad neonatal comprende la mortalidad que ocurre dentro del primer mes de vida (desde el nacimiento hasta antes de cumplir los 28 días). Estas son causadas generalmente por accidentes de parto o por malformaciones congénitas (defunciones endógenas)

$$\frac{D_{-28}^z}{B^z}$$

Para este cálculo se distribuyen los no especificados, por lo que en R-Project se trabaja de la siguiente forma:

```
>#Mortalidad infantil con la distribución de los no especificados
>mort_infNE=c(1:15)
for(j in 1:15){
  mort_infNE[j]=c(mort_inf[j,2]*(1+(mort_inf[14,2]/
(mort_inf[1,2]-mort_inf[14,2])))
}

># tasa de mortalidad neonatal
>neo=c(1:10)
for(j in 1:10){
  neo[j]=c(mort_infNE[j+1]);
}
tneo=(sum(neo)/mort_inf[16,2])*1000
```

Por lo tanto el resultado de la tasa de mortalidad neonatal en el estado de Nuevo León es de:

```
> tneo
[1] 8.494115
```

La tasa mortalidad posneonatal se refiere a la mortalidad que ocurre entre el momento de cumplir 28 días de nacido y el correspondiente al primer cumpleaños. Estas son causadas por enfermedades respiratorias, accidentes alimenticios, etc. (causas externas, defunciones exógenas)

$$\frac{D_{28-365}^z}{B^z}$$

Para calcular la tasa posneonatal se ejecuta:

```
>posneo=c(mort_infNE[12],mort_infNE[13],mort_infNE[15])
```

```
>tposneo=(sum(posneo)/mort_inf[16,2])*1000
```

El resultado para el estado que se trabaja es:

```
> tposneo  
[1] 5.307287
```

3.3.1.4. Tabla de mortalidad

Las tablas de mortalidad son la representación estadística de la mortalidad. El supuesto de la tabla es l_0 que generalmente $l_0 = 100000$ que son los sobrevivientes a edad exacta cero y es llamado el radix de la tabla de mortalidad. Las tablas se diferencian por sexo por las causas biológicas y socioeconómicas sin olvidar que los hombres están más expuestos a accidentes y violencia.

En la tabla de mortalidad usualmente se utilizan los grupos quinquenales de edad pero por la variación de la mortalidad al principio de la vida, se divide al primer grupo quinquenal en dos subgrupos: menores de un año y uno a cuatro años cumplidos. Se termina en el grupo de 85 y +.

Se utiliza la tasa de mortalidad específica con el promedio de tres años consecutivos por las irregularidades en los datos. Este promedio se divide entre la población a mitad del año que se busca. Al incluir estas columnas en la tabla de mortalidad se facilita el cálculo y además se puede incluir otra columna de la tasa de mortalidad específica corregida y sirve para corregirla a mano alzada si lo creemos conveniente.

Por lo tanto la primera parte de la tabla queda:

<i>Quinquenios</i>	<i>Defunciones</i> $t - 1$	<i>Defunciones</i> t	<i>Defunciones</i> $t + 1$	<i>Población</i> <i>mitad_año</i>	${}_5m_x^t$	${}_5\hat{m}_x^t$
--------------------	-------------------------------	---------------------------	-------------------------------	--------------------------------------	-------------	-------------------

Donde:

Quinquenios son los grupos de edad.

Defunciones son según cada grupo de edad el año $t - 1, t, t + 1$.

Poblaciones mitad_año Es el total de vivos para cada grupo de edad a mitad de año.

${}_5m_x^t$ = Es la tasa específica de mortalidad en quinquenios de la edad x a $x + 5 - 1$ que se calcula como:

$${}_n m_x^t = \frac{\left(\frac{1}{3}\right) * ({}_n D_x^{t-1} + {}_n D_x^t + {}_n D_x^{t+1})}{{}_n P_x^t}$$

\hat{m}_x^t = Es la tasa específica de mortalidad ajustada.

En R-Project se creó la columna de ${}_5m_x^t$ con ayuda de una función para que recorra todos los quinquenios con la fórmula de la tasa específica de mortalidad antes calculada.

```
# la columna de tasa de mortalidad específica
>coltseH<-function( ) {
  k=c(1:19)
  for(j in 1:19) {
    k[j]=tseH(j);
  }
  return (k)
}
```

También se puede crear un gráfico por si aún con el promedio utilizado la tasa no crece suavemente corregirla. Para esto se utiliza la gráfica del tipo "línea" para ver la suavidad con la que crece. El gráfico se crea de la siguiente forma:

```
# grafica de tasa mortalidad específica
>GTme=plot(coltseH( ), type="l", xlim=c(1,19), xlab="quinquenios",
ylab="tasa de mortalidad específica")
```

Así se calcula la tasa específica de mortalidad corregida.

En lo que respecta a la tabla de mortalidad está formada por:

Edades	\hat{m}_x^i	${}_nq_x$	l_x	${}_nd_x$	L_x	T_x	e_x
< 1	${}_1\hat{m}_0^i$	${}_1q_0 = 1 - \exp[1 * -{}_1\hat{m}_0^i - 0.008 * ({}_1\hat{m}_0^i)^2]$	$l_0 = 100000$	${}_1d_0 = {}_1q_0 * l_0$	$L_0 = f_1 * l_0 + (1 - f_1) * l_1$	$T_0 = \sum_0^{18} L_x$	$e_0 = \frac{T_0}{l_0}$
1 - 4	${}_4\hat{m}_1^i$	${}_4q_1 = 1 - \exp[4 * -{}_4\hat{m}_1^i - 0.008 * 4^3 * ({}_4\hat{m}_1^i)^2]$	$l_1 = l_0 - {}_1d_0$	${}_4d_1 = {}_4q_1 * l_1$	$L_1 = f_2 * l_1 + (1 - f_2) * l_3$	$T_1 = \sum_1^{18} L_x$	$e_1 = \frac{T_1}{l_1}$
5 - 9	${}_5\hat{m}_5^i$	${}_5q_5 = 5 * {}_5\hat{m}_5^i / [1 + (5/2) * {}_5\hat{m}_5^i]$	$l_2 = l_1 - {}_4d_1$	${}_5d_5 = {}_5q_5 * l_2$	$L_3 = 5/2 * (l_3 + l_4)$	$T_2 = \sum_2^{18} L_x$	$e_2 = \frac{T_2}{l_2}$
...
80 - 84	${}_5\hat{m}_{80}^i$	${}_5q_{80} = 5 * {}_5\hat{m}_{80}^i / [1 + (5/2) * {}_5\hat{m}_{80}^i]$	$l_{17} = l_{16} - {}_5d_{75}$	${}_5d_{80} = {}_5q_{80} * l_{17}$	$L_{17} = 5/2 * (l_{17} + l_{18})$	$T_{17} = \sum_{17}^{18} L_x$	$e_{17} = \frac{T_{17}}{l_{17}}$
85y +	${}_85\hat{m}_{85}^i$	${}_85q_{85} = 1$	$l_{18} = l_{17} - {}_5d_{80}$	${}_85d_{85} = {}_85q_{85} * l_{18}$	$L_{18} = T_{18}$	$T_{18} = e_{18} * l_{18}$	$e_{18} = \frac{1}{\hat{m}_{85}^i}$

Donde:

${}_nq_x$ Es la probabilidad para los sobrevivientes a edad x de fallecer antes de llegar a la edad $x+n-1$.

l_x Representa a los sobrevivientes de edad x que van a llegar con vida a la edad $x+n-1$.

${}_nd_x$ Representa a las personas que van a fallecer entre x y $x+n-1$.

T_x Representa los años - personas vividos entre las edades x y $x+n-1$ y también las personas vivas a edad cumplida x años. Sirve para estimar la serie de las esperanzas de vida a edad x .

f_1 =factor de separación en el caso de hombres $f_1 = 0.0425 + (2.875 * q_0)$, en el caso de mujeres $f_1 = 0.05 + (3 * q_0)$

f_2 =factor de separación en el caso de hombres $f_2 = 1.653 - 3.013 * q_1$, en el caso de mujeres $f_2 = 1.524 - (1.627 * q_1)$

e_x =Es la esperanza de vida

Para R-Project se colocan cada columna como un vector para su fácil manejo. Entonces queda para hombres del 2000:

```
># la columna de la probabilidad de sobrevivientes
>nqx=c(1-exp(1*-tseH(1)-.008*tseH(1)^2),1-exp(4*-tseH(2)-.008*4^3*
tseH(2)^2),5*coltseH()[3:18]/(1+(5/2)*coltseH()[3:18]),1)
```

La columna de sobrevivientes l_x se va creando al mismo tiempo que la columna de las personas que van a fallecer ${}_nd_x$

```
># La columna de sobrevivientes lx
># La columna de las personas que van a fallecer ndx
>ndx=c(1:19)*0
lx=c(1:19)*0
cst=100000
ndx[1]=cst*nqx[1]
lx[1]=cst
for(j in 2:19){
  lx[j]=lx[j-1]-ndx[j-1]
  ndx[j]=lx[j]*nqx[j]
}
```

Recordando que para esta columna las edades de menores a un año y de 1 a 4 años se utilizan factores de separación.

```
># La columna de Lx
>asd=.0425+(2.875*nqx[1])
asc=1.653-(3.013*nqx[2])
Lx=c(asd*lx[1]+(1-asd)*lx[2], asc*lx[2]+(1-asc)*lx[3],
      5/2*(lx[3:18]+lx[4:19]), (1/tseH(19))*lx[19])
```

```
># La columna de Tx
>Tx=(1:19)*0
Tx[19]=(1/tseH(19))*lx[19]
for(i in 1:19){
  Tx[i]=c(sum(Lx[i:19]))
}
```

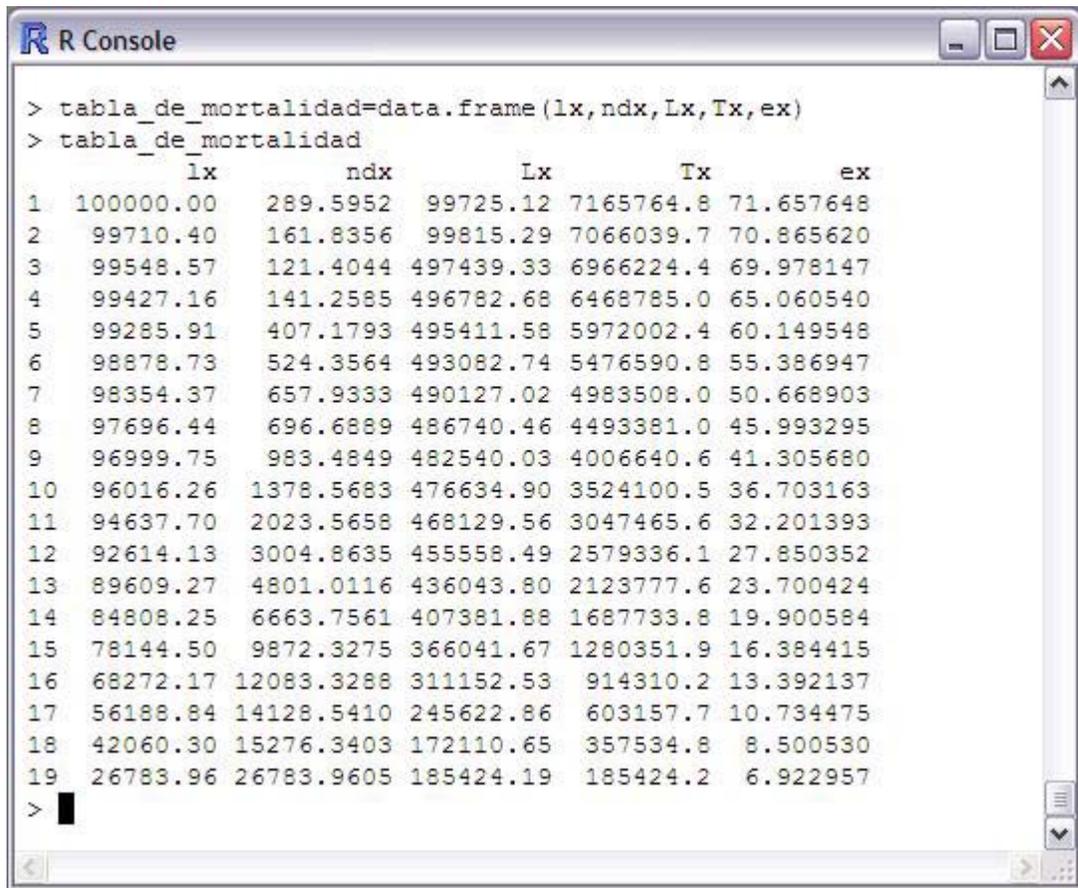
```
># La columna de la esperanza de vida
>ex=c(Tx[1:18]/lx[1:18],1/tseH(19))
```

La Esperanza de vida es un índice muy utilizado que da la duración media de vida a partir de una edad dada, es decir, por ejemplo: e_0 representa los años que en promedio espera vivir una persona al momento de su nacimiento.

Para poder ver la tabla de mortalidad se utiliza una función en R-Project que es:

```
>tabla_de_mortalidad=data.frame(lx,ndx,Lx,Tx,ex)
```

Por lo que se puede ver en la consola de la siguiente manera.

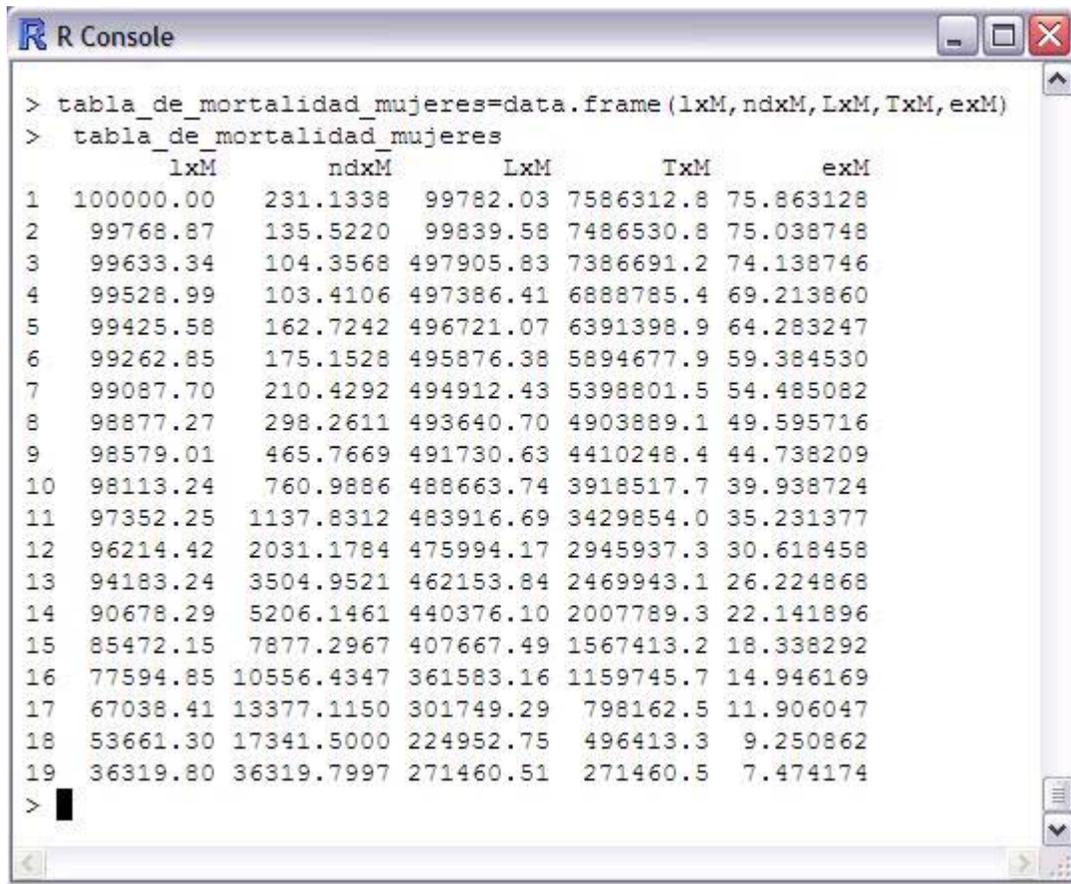


```
> tabla_de_mortalidad=data.frame(lx,ndx,Lx,Tx,ex)
> tabla_de_mortalidad
```

	lx	ndx	Lx	Tx	ex
1	100000.00	289.5952	99725.12	7165764.8	71.657648
2	99710.40	161.8356	99815.29	7066039.7	70.865620
3	99548.57	121.4044	497439.33	6966224.4	69.978147
4	99427.16	141.2585	496782.68	6468785.0	65.060540
5	99285.91	407.1793	495411.58	5972002.4	60.149548
6	98878.73	524.3564	493082.74	5476590.8	55.386947
7	98354.37	657.9333	490127.02	4983508.0	50.668903
8	97696.44	696.6889	486740.46	4493381.0	45.993295
9	96999.75	983.4849	482540.03	4006640.6	41.305680
10	96016.26	1378.5683	476634.90	3524100.5	36.703163
11	94637.70	2023.5658	468129.56	3047465.6	32.201393
12	92614.13	3004.8635	455558.49	2579336.1	27.850352
13	89609.27	4801.0116	436043.80	2123777.6	23.700424
14	84808.25	6663.7561	407381.88	1687733.8	19.900584
15	78144.50	9872.3275	366041.67	1280351.9	16.384415
16	68272.17	12083.3288	311152.53	914310.2	13.392137
17	56188.84	14128.5410	245622.86	603157.7	10.734475
18	42060.30	15276.3403	172110.65	357534.8	8.500530
19	26783.96	26783.9605	185424.19	185424.2	6.922957

```
>
```

Para mujeres se hace el mismo procedimiento pero con la población femenina y da como resultado:



```
> tabla_de_mortalidad_mujeres=data.frame(lxM,ndxM,LxM,TxM,exM)
> tabla_de_mortalidad_mujeres
```

	lxM	ndxM	LxM	TxM	exM
1	100000.00	231.1338	99782.03	7586312.8	75.863128
2	99768.87	135.5220	99839.58	7486530.8	75.038748
3	99633.34	104.3568	497905.83	7386691.2	74.138746
4	99528.99	103.4106	497386.41	6888785.4	69.213860
5	99425.58	162.7242	496721.07	6391398.9	64.283247
6	99262.85	175.1528	495876.38	5894677.9	59.384530
7	99087.70	210.4292	494912.43	5398801.5	54.485082
8	98877.27	298.2611	493640.70	4903889.1	49.595716
9	98579.01	465.7669	491730.63	4410248.4	44.738209
10	98113.24	760.9886	488663.74	3918517.7	39.938724
11	97352.25	1137.8312	483916.69	3429854.0	35.231377
12	96214.42	2031.1784	475994.17	2945937.3	30.618458
13	94183.24	3504.9521	462153.84	2469943.1	26.224868
14	90678.29	5206.1461	440376.10	2007789.3	22.141896
15	85472.15	7877.2967	407667.49	1567413.2	18.338292
16	77594.85	10556.4347	361583.16	1159745.7	14.946169
17	67038.41	13377.1150	301749.29	798162.5	11.906047
18	53661.30	17341.5000	224952.75	496413.3	9.250862
19	36319.80	36319.7997	271460.51	271460.5	7.474174

```
>
```

En el caso de 1990 las tablas de mortalidad para hombres y mujeres quedan:

```
R Console
> Tabla_mortalidad_90=data.frame(lxH90,ndxH90,LxH90,TxH90,exH90)
> Tabla_mortalidad_90
  lxH90    ndxH90    LxH90    TxH90    exH90
1 100000.00 2468.7868 97811.37 6755431.6 67.554316
2  97531.21  439.2410 97812.08 6657620.2 68.261431
3  97091.97 196.4309 484968.78 6559808.1 67.562827
4  96895.54 219.3426 483929.35 6074839.3 62.694725
5  96676.20 492.0614 482150.84 5590910.0 57.831297
6  96184.14 710.9765 479143.25 5108759.2 53.114363
7  95473.16 868.3109 475195.03 4629615.9 48.491281
8  94604.85 986.5012 470558.00 4154420.9 43.913403
9  93618.35 1200.7315 465089.91 3683862.9 39.349796
10 92417.62 1698.7081 457841.32 3218773.0 34.828565
11 90718.91 2449.3213 447471.24 2760931.7 30.433916
12 88269.59 3679.2496 432149.82 2313460.4 26.209032
13 84590.34 5592.7820 408969.74 1881310.6 22.240254
14 78997.56 7751.2330 375609.70 1472340.9 18.637803
15 71246.32 10221.3293 330678.29 1096731.2 15.393513
16 61024.99 12285.6606 274410.82 766052.9 12.553100
17 48739.33 13800.4240 209195.61 491642.1 10.087172
18 34938.91 13954.1940 139809.06 282446.5 8.084009
19 20984.72 20984.7152 142637.39 142637.4 6.797204
>
```

```
R Console
> Tabla_mortalidad_mujeres_90=data.frame(lxM90,ndxM90,LxM90,TxM90,exM90)
> Tabla_mortalidad_mujeres_90
  lxM90    ndxM90    LxM90    TxM90    exM90
1 100000.00 1984.5802 98232.8 7246833.2 72.468332
2  98015.42  425.7324 98235.5 7148600.4 72.933426
3  97589.69 148.0722 487578.3 7050364.9 72.244979
4  97441.62 143.5118 486849.3 6562786.6 67.350963
5  97298.10 185.5578 486026.6 6075937.3 62.446616
6  97112.55 240.4268 484961.7 5589910.7 57.561159
7  96872.12 340.1821 483510.1 5104949.0 52.697816
8  96531.94 433.1757 481576.7 4621438.9 47.874715
9  96098.76 624.8998 478931.6 4139862.2 43.079246
10 95473.86 943.6645 475010.1 3660930.6 38.344847
11 94530.20 1446.8177 469033.9 3185920.5 33.702675
12 93083.38 2414.2854 459381.2 2716886.5 29.187666
13 90669.09 3791.8730 443865.8 2257505.3 24.898290
14 86877.22 5762.2592 419980.5 1813639.6 20.875893
15 81114.96 8497.9792 384329.9 1393659.1 17.181283
16 72616.98 10940.2973 335734.2 1009329.2 13.899355
17 61676.68 14497.0181 272140.9 673595.1 10.921389
18 47179.67 16596.7956 194406.3 401454.2 8.509051
19 30582.87 30582.8711 207047.9 207047.9 6.770060
>
```

Para poder exportar las tablas creadas a documentos `csv` con la finalidad de poder manejarla después. Se utiliza el siguiente comando:

```
># exportar los datos de la tabla de mortalidad para hombres del
2000
> write.table(tabla_de_mortalidad, file = "mortalidad.csv",
sep=" , ", col.names=NA, qmethod="double" )
```

Se utiliza `write.table` pues el argumento `tabla_mortalidad` es una tabla, es decir, un `data.frame`. Se guarda en un archivo `csv` que puede usarse en Excel y otras aplicaciones para elaborar un informe de manera sencilla.

Para poder leerla en R-Project entonces utilizas la instrucción:

```
># leer el archivo CSV que fue creado
>read.table("mortalidad.csv", header=TRUE, sep=" , ",
row.names = 1)
```

3.3.1.5. Conclusiones

La tabla de mortalidad es muy útil para demografía sobre todo por el indicador de la esperanza de vida que permite comparar la mortalidad de diferentes poblaciones o la misma población a lo largo del tiempo. En el caso de Nuevo León se puede ver que para las personas mayores a 85 años tienen una esperanza de vida de casi 7 años más. La tabla de mortalidad también es utilizada para el cálculo de primas de seguros de vida por lo que es muy importante saber como se construye. Yo decidí que para R-Project era mejor manejarla por cada vector que se va necesitado para no perder de vista cada parte de la Tabla. Además se muestra una forma de exportar datos, en este caso de una tabla, a otros archivos para que el alumno conozca la forma de exportar sus resultados y pueda elaborar el informe final que se pide en el curso.

3.3.2. Fecundidad

La fecundidad se define como la capacidad de una mujer, hombre, pareja de producir un nacimiento, un nacido vivo. Un fenómeno relacionado con la fecundidad es la natalidad pues son los nacimientos que tiene una población en cierto periodo de tiempo.

Por otro lado, para no confundir, la fertilidad se refiere a la capacidad biológica de una mujer, hombre, pareja de engendrar un hijo. Una mujer puede ser fértil y decidir no tener hijos, es decir, tener fecundidad nula.

Para el estudio de la fecundidad se necesitan los nacimientos de un área y momento determinados. Esta información se encuentra en las estadísticas vitales. También se necesitan las personas que estuvieron en riesgo y se consiguen de los datos de población de los censos.

A pesar de que la concepción es compartida por ambos sexos el análisis toma a la mujer en edad fértil, entre 15 y 50 años exactos (15 a 49 años cumplidos), como la población que está expuesta al riesgo, por su papel preponderante.

Para trabajar con los datos se llama a la función `read.csv` con un documento que tenga para el censo del año z:

- En la primera fila los quinquenios desde 15 - 19 hasta 45 - 49 y en el lugar del título puedes ponerle Nacimientos
- En las tres siguientes se colocan los nacimientos de los años z-1, z, z+1. Los títulos son los años correspondientes.
- En la cuarta columna se colocan el punto medio de los quinquenios con el título de "punto medio".

Así utilizamos la función como sigue:

```
>#importar los datos
>fec_NL= read.csv("C:/fecundidad 2000.csv"
,header=TRUE,sep=" ",quote="\ ",dec=".",fill = TRUE)
```

Como el ejemplo es del estado de Nuevo León el archivo es llamado `fec_NL` pero para efectos prácticos a nuestra variable la vamos a llamar `fec`.

3.3.2.1. Tasa bruta de natalidad

Se refiere al número de nacimientos que ocurren en una población por cada mil habitantes durante un periodo dado. Y su formula es:

$$b^z = \frac{B^z}{P^z} * 1000$$

Donde:

B^z = El número total de nacimientos ocurridos el año z .

P^z = Es la población total a mitad del año z .

Para obtener la población total a mitad del año se hace primero una función que sume las poblaciones a mitad de año de hombres y mujeres.

```
># suma de las poblaciones a mitad de año.  
>dpactTm=sum(sum(dpactHm[1:102]),sum(dpactMm[1:102]))
```

Luego se sigue la fórmula de la tasa bruta de natalidad.

```
>#tasa bruta de natalidad del año actual (2000)  
>tbn=(sum(fec[3])/dpactTm)*1000
```

Por lo que queda:

```
> tbn  
[1] 22.67617
```

Para el año de 1990 es de:

```
> tbn_90
[1] 25.91333
```

Esto quiere decir que por cada mil habitantes ocurren 23 nacimientos en el estado de Nuevo León en el 2000 y 26 en el año de 1990

También se puede calcular la tasa bruta de natalidad para cada género.

```
>#tasa bruta de natalidad femenina del 2000
>tbnF=(sum(fec[3])/sum(dpactMm[1:102]))*1000

>#tasa bruta de natalidad masculina del 2000
>tbnM=(sum(fec[3])/sum(dpactHm[1:102]))*1000

>#tasa bruta de natalidad femenina de 1990
>tbnF_90=(sum(fec[3])/sum(dpantMm[1:102]))*1000

>#tasa bruta de natalidad masculina de 1990
>tbnM_90=(sum(fec[3])/sum(dpantHm[1:102]))*1000
```

Los resultados son:

```
> tbnF
[1] 45.1366
> tbnM
[1] 45.57016
> tbnF_90
[1] 55.96172
> tbnM_90
```

3.3.2.2. Tasa de fecundidad general

La tasa de fecundidad general (TFG) se interpreta como la cantidad de nacimientos por cada mil mujeres en edad fértil ocurridos durante cierto lapso de tiempo, usualmente un año calendario. Su formula es:

$$TFG = \frac{B^z}{P_{f15-49}^z} * 1000$$

Donde:

B^z = El total de nacimientos ocurridos el año z .

P_{f15-49}^z =La población femenina en edad fértil a mitad del año z .

Por lo que la función es igual a la anterior pero con la población femenina a mitad del año. Entonces obtenemos con el siguiente código:

```
># El total de la población femenina en edad fértil a mitad del
año del 2000
>dpactfm=sum(dpactMmQ[4:10])
```

Entonces queda:

```
>#tasa de fecundidad general del año actual (2000)
>TFG=(sum(fec[3]))/(3*dpactfm)*1000
```

Esto da:

```
> TFG
[1]27.09212
```

```
> TFG_90
[1] 31.84249
```

Significa que por cada 1000 mujeres en edad fértil en Nuevo León se dan 27 y 32 nacimientos según los años respectivos.

3.3.2.3. Tasas específicas de fecundidad

Esta tasa indica cómo se distribuye la fecundidad a lo largo del periodo fértil. Por la

dificultad en los datos se presentan en quinquenios. Así si existen nacimientos en mujeres menores de 15 años se suman al grupo de las de 15-19 años. Igualmente si existen nacimientos en mayores de 50 años exactos se le suman al grupo de 45-49 años. Las tasas se calculan como siguen:

$${}_5f_x^z = \frac{{}_5B_x^z}{P_{f5-x}^z}$$

Donde:

${}_5f_x^z$ = La tasa de fecundidad específica de la edad x a $x+5-1$. al año z .
 ${}_5B_x^z$ = Los nacimientos de mujeres de edad x a $x+5-1$ en el año z .
 P_{f5-x}^z = La población femenina de edad x a $x+5-1$ a mitad del año z .

Para reflejar más los cambios en las variaciones anuales se calcula con los datos de los tres años consecutivos como sigue:

$${}_5f_x^z = \frac{(Nac_{z-1} + Nac_z + Nac_{z+1})/3}{P_{f5-x}^z}$$

Donde:

Nac_i = Son los nacimientos de mujeres de edad x a $x+5-1$ del año i .
 P_{f5-x}^z = La población femenina de edad x a $x+5-1$ a mitad del año z .

Por lo que queda:

```
># tasa especifica de fecundidad del año actual (2000)
>tef=c(1:7)*0
for(i in 1:7){
tef[i]=c((sum(fec[i,2],fec[i,3],fec[i,4]))/3)/dpactMmQ[i+3])
}

> tef
[1]0.06474197 0.1254197 0.139965822 0.0977966281 0.041025271
0.0077516889 0.0006685056
```

tef es un vector con siete elementos (por cada quinquenio manejado) donde se puede ver como se distribuye la fecundidad en el estado de Nuevo Leon en el año 2000.

3.3.2.3.1. Edad cúspide

La edad cúspide representa el grupo de edad donde se produce la mayor fecundidad, se encuentra donde la tasa específica de fecundidad es más alta.

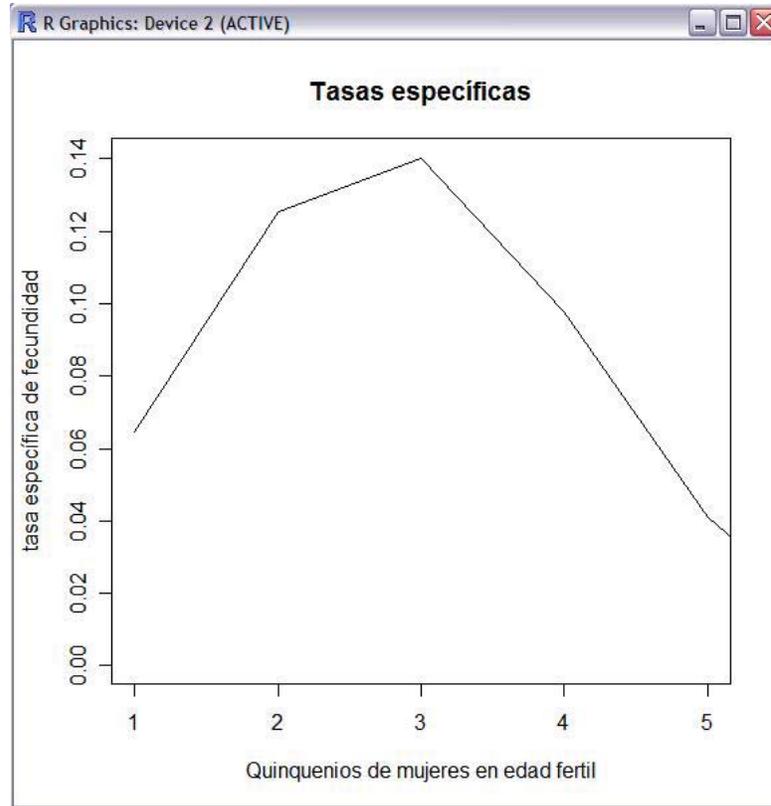
A la edad cúspide se le divide según el grupo donde se encuentre:

- En el grupo de 20-24 años se le llama "Cúspide temprana".
- En el grupo de 25-29 años se le llama "Cúspide tardía".
- Cuando es en los grupos de 20-24 y 25-29 años se dice "Cúspide dilatada".

Para su mejor vista es recomendable crear una gráfica, para poder notar si la edad cuspide es dilatada, Esta se manda llamar de la siguiente forma:

```
># grafica de Tasa especifica de fecundidad
>Gtef=plot(tef,type="l",xlim=range(1,5),xlab="Quinquenios de
mujeres en edad fértil",ylab="tasa específica de
fecundidad",main="Tasas específicas")
```

Con esto se abre una ventana para el gráfico y queda:



Así es más fácil ver que en el tercer quinquenio es la edad cuspide, es decir, que entre las edades de 25 a 29 años es cuando las mujeres de Nuevo León tienen más hijos.

3.3.2.3.2. Grado de concentración

Para calcular el grado de concentración de la fecundidad las edades se dividen en tres grupos y se calculan las tasas específicas de fecundidad (se suman las tasas específicas de fecundidad de los grupos correspondientes) y se dividen entre la suma total de las mismas.

	${}_n f_x^z$
15 – 19	${}_5 f_{15}$
20 – 34	${}_{15} f_{20}$
35 – 49	${}_{15} f_{35}$

Donde:

${}_5 f_x^z$ = La tasa de fecundidad específica de la edad x a $n+5-1$. al año z .

Por lo que queda en R-Project:

```
># nuevas tasas específicas de fecundidad
>ntef=c(tef[1], sum(tef[2:4]), sum(tef[5:7]))
```

```
> ntef
[1] 0.06474198 0.36318216 0.04944547
```

Estas son las tasas con los nuevos grupos de edad

La fórmula del grado de concentración para cada nuevo grupo de edad es:

$$\frac{{}_5 f_x^z}{\sum {}_5 f_{45}^z}$$

Por lo que se necesita la suma de las tasas específicas que es:

```
>#suma de las tasas específicas  
>stef=sum(ntef)
```

Da como resultado:

```
> stef  
[1] 0.4773696
```

Para saber el grado de concentración se hace:

```
># Grado de concentración del año actual (2000)  
>gc=c(ntef[1]/stef, (ntef[2])/stef, (ntef[3])/stef)
```

Por lo que el grado de concentración para Nuevo León en el año 2000 con los nuevos grupos de edad es:

```
> gc  
[1] 0.1356223 0.7607987 0.1035790
```

Luego se clasifican según si son países de alta fecundidad o de baja fecundidad.

	<i>países alta fecundidad</i>	<i>países baja fecundidad</i>
15 – 19	> 10%	< 10%
20 – 34	< 75%	> 75%
35 – 49	> 15%	< 15%

Así se puede ver que si son de alta o baja concentración según cada grupo

En R-Project se hacer con una función que se escribe así:

```
># Función que dice de cada grupo que tipo de concentración tiene
>GC_fec=function(i){
  if(i==1)
  {
    if(gc[i]>.1)
      return (gca="Alta")
    else
      return (gca="baja")
  }
else
  {
    if(i==2)
    {
      if(gc[i]<.75)
        return (gcb="Alta")
      else
        return (gcb="baja")
    }
    else
    {
      if(i==3){
        if(gc[i]>.15)
          return (gcc="Alta")
        else
          return (gcc="baja")
        }
      else{
        return(gcd="no valido")
        }
      }
    }
  }
}
```

Los resultados por cada grupo son:

```
> GC_fec(1)
[1] "Alta"
> GC_fec(2)
[1] "baja"
> GC_fec(3)
[1] "baja"
```

Ahora si se quiere saber cual es la concentración en total se hace otra función de la siguiente manera:

```
># Función que dice en total que grado de concentración tiene
>GCT_fec=function(){
if( GC_fec(1)=="Alta"&GC_fec(2)=="Alta"&GC_fec(3)=="Alta" )
  return(GTC="alta" )
else{
  if( GC_fec(1)=="Alta"&GC_fec(2)=="Alta"&GC_fec(3)=="baja" )
    return(GTC="Alta" )
  else{
    if( GC_fec(1)=="Alta"&GC_fec(2)=="baja"&GC_fec(3)=="Alta" )
      return(GTC="Alta" )
    else{
      if( GC_fec(1)=="baja"&GC_fec(2)=="Alta"&GC_fec(3)=="Alta" )
        return(GTC="Alta" )
      else
        return(GTC="Baja" )
    }
  }
}
}
```

Así se sabe que grado de concentración tiene el Estado. Si se ejecuta la función el resultado para Nuevo León es: "Baja" en el 2000. Al hacer el mismo procedimiento para el año 1990 también se muestra "Baja".

3.3.2.4. Factores que afectan las medidas de la fecundidad

Los Factores que afectan las medidas se pueden dividir en extrínsecos (afectan sólo a las tasas) e intrínsecos (afectan a las tasas y a la fecundidad).

3.3.2.4.1. Factores extrínsecos:

Son aquellos que cuando ocurren los cambios afectan los niveles de las tasas pero no los niveles de la fecundidad

- Porcentaje de mujeres de 15-49 años respecto a la población total femenina.

$$\frac{P_{f_{15-49}}^z}{P_f^z}$$

En el programa queda:

```
># porcentaje de mujeres en edad fértil  
>pm=dpactfm/sum(dpactMmQ)
```

Da el resultado de 0.5584949 que quiere decir que el 55% de las mujeres está en edad fértil en el 2000. Para 1990 es de 0.5405317 que significa el 54%.

- Estructura por edad de las mujeres en edad fértil.

$$\frac{P_{f_{5-x}}^z}{P_{f_{15-49}}^z}$$

En R-Project se hace un ciclo así:

```
># la estructura por edad  
>ee=c(1:7)*0  
for(i in 1:7){  
ee[i]=c(dpactMmQ[i+3]/dpactfm)  
}
```

Da como resultado un vector donde cada componente se refiere a los diferentes grupos de edad para mujeres en edad fértil.

```
> ee  
[1] 0.17455930 0.18289987 0.17375254 0.15296655 0.12857772  
0.10678866 0.08045538
```

Para su mejor visualización se puede hacer una gráfica de pastel como sigue:

```
>#Gráfica de la estructura por edad de las mujeres en edad fértil  
>Gee=pie(ee,ee,col=rainbow(8),main="Estructura por edad de mujeres  
en edad fértil")
```

La gráfica se dibuja de la siguiente manera:



- Estructura de las tasas de fecundidad por edad.

$$\frac{{}_5f_x}{\sum {}_5f_{45}}$$

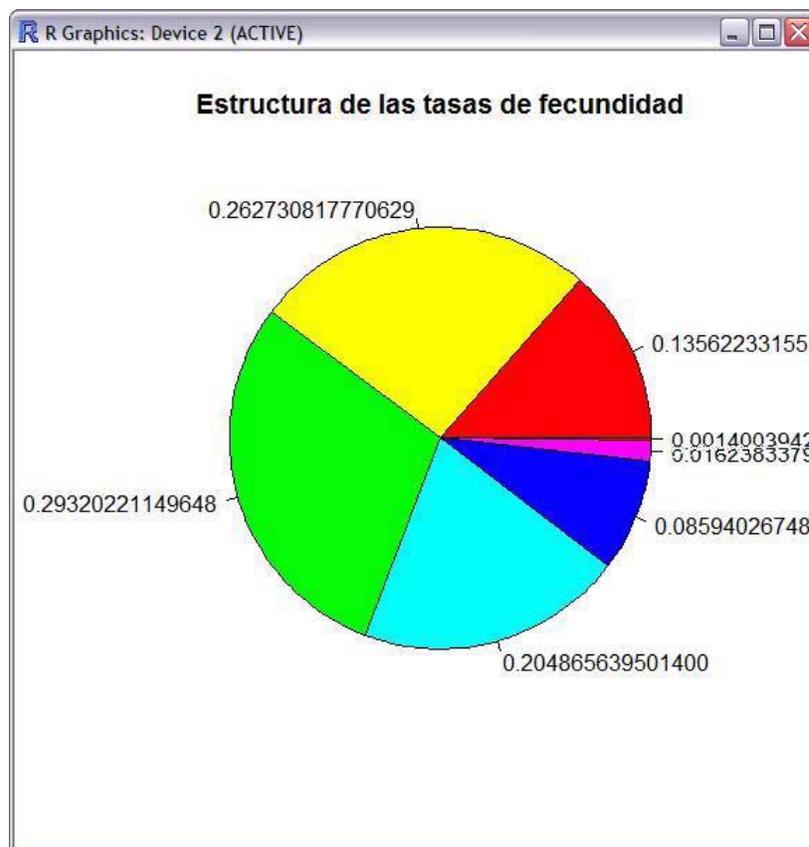
En R-Project se escribe:

```
># Estructura de las tasas de fecundidad
> etef=c(1:7)
> for(j in 1:7){
+ etef[j]=c(tef[j]/stef)
+ }
```

Para verla se usa otra gráfica de pastel

```
>Getef=pie(etef,etef,col=rainbow(6),main="Estructura de las tasas
de fecundidad")
```

Por lo que la gráfica queda:



3.3.2.4.2. Factores intrínsecos:

Son aquellos que cuando ocurren los cambios de dicho factor afectan las tasas de fecundidad solamente.

Para poder observar estos hechos se necesitan los valores de la población de mujeres por estado conyugal en Nuevo León por lo que importamos los datos como se ha hecho anteriormente

```
># Importar datos según estado conyugal
>fec_conyugal=read.csv("C:/conyugal.csv",header=TRUE,sep="," ,vquote
= TRUE)
```

- Proporción de mujeres que se casan.

$$\frac{P_{casadas}^z}{P_f^z}$$

En este cálculo se toman en cuenta las mujeres que se casan, es decir, las que se encuentran en unión libre, casadas por el civil y religiosamente, casadas por el civil, y casadas religiosamente contra el total de la población femenina .

En R-Project es:

```
>#proporcion de mujeres que se casan
>pMc=sum(fec_conyugal[3,2],fec_conyugal[4,2],fec_conyugal[5,2],
fec_conyugal[6,2])/fec_conyugal[1,2]
```

Da como resultado:

```
> pMc
[1] 0.5645347
```

```
> pMc_90
[1] 0.530318
```

Esto quiere decir que se casan el 56% de las mujeres en edad fértil en el 2000 y 53% en el años 1900.

- Estructura de mujeres según el estado conyugal considerando si son casadas ó unidas.

$$\frac{P_{f_{5-x}}^z}{P_{f_{15-49}}^z}$$

En R-Project se hace:

```
>#Estructura según estado conyugal  
>eec=c(1:8)  
for(j in 1:8){  
  eec[j]=c(fec_conyugal[j+1,2]/fec_conyugal[1,2])  
}
```

La gráfica queda:

```
>Geec=pie(eec,eec,col=rainbow(7),main="Estructura de mujeres  
segun el estado conyugal")
```



3.3.2.5. Medidas resumen de la fecundidad

3.3.2.5.1. Tasa global de fecundidad

La tasa global de fecundidad (TGF) se interpreta como el número de hijos, que en promedio, que tendría cada mujer desde el inicio hasta el final de su período fértil. Con el supuesto de no mortalidad. Su cálculo en quinquenios es:

$$TGF^z = 5 * \sum_{x=15}^{45} {}_n f_x^z$$

Donde:

${}_n f_x^z$ =La tasa específica de fecundidad de x a $x + n - 1$ años de edad en el año z .

Se multiplica por 5 ya que se refleja la situación media de una mujer en todo el tramo quinquenal.

Programandolo sería muy fácil pues ya se tiene la suma de las tasas específicas de fecundidad de los grupos formados para el grado de concentración entonces queda:

```
>#TGF tasa global de fecundidad del año actual (2000)  
>TGF=5*stef
```

Da como resultado:

```
> TGF  
[1] 2.386848
```

```
> TGF_90  
[1] 2.761428
```

Por lo que el número de hijos en promedio que tienen las mujeres del estado de Nuevo León es dos en el año 2000 y tres en el 1990.

3.3.2.5.2. Tasa bruta de reproducción

La tasa bruta de reproducción (TBR) se interpreta como el número de hijas, en promedio, que tendrá una mujer en su periodo reproductivo (15-50 años exactos). Con el supuesto de no mortalidad. Su cálculo en quinquenios es:

$$TRB^z = k * 5 * \sum_{x=15}^{45} n f_x^z$$

Donde:

K = 100/205 es la proporción teórica de nacimientos femeninos sobre el total de nacimientos. Basado en un promedio de 105 nacimientos masculinos por cada 100 femeninos.

```
>#TBR -tasa bruta de reproducción del año actual (2000)
>TBR=TGF*100/205
```

Con esto se obtiene:

```
> TBR
[1] 1.164316

> TBR_90
[1] 1.347038
```

Quiere decir que el número de hijas en promedio que tienen las mujeres de Nuevo Leon es de una para los dos años.

3.3.2.5.3. Descendencia media final

La descendencia media final es una medida de la intensidad de la fecundidad y representa el número de hijos que en promedio tuvieron las mujeres de una cohorte al final

del periodo fértil. Y se calcula:

$$dmf = 5 * \sum_5 f_x^z$$

Donde:

${}_5f_x^z$ =Son las tasas de fecundidad específicas de los quinquenios.
Se multiplica por cinco por los quinquenios.

Queda así:

```
>#dmf descendencia media final  
>dmf=5*sum( tef )
```

El vector queda:

```
> dmf  
[1] 2.386848  
  
> dmf_90  
[1] 2.761428
```

Por lo que en promedio las mujeres de Nuevo León tienen en cada cohorte al final del periodo reproductivo dos hijos ahora ya que antes tenían 3.

3.3.2.5.4. Tasa neta de reproducción

La tasa neta de reproducción (TNR) se refiere a los nacimientos femeninos, pero en este caso se toma en cuenta la mortalidad de las mujeres. Es la capacidad de auto reemplazarse.

Por lo tanto

$$TNR^z = 5 * k * \sum_{x=15}^{45} {}_n f_x^z * {}_{x+2.5} P_0(f)$$

Donde:

${}_{x+2.5} P_0(f)$ = La probabilidad de sobrevivencia femenina entre el nacimiento y la edad $x + 2.5$ años, que es el punto medio de los diferentes grupos de edad.

Entonces:

$${}_{x+2.5} P_0(f) = \frac{l_{x+2.5}(f)}{{}_5 l_0} = \frac{{}_5 L_x(f)}{{}_5 l_0}$$

${}_5 L_x(f)$ = El tiempo vivido entre las edades x a $x+5-1$ en la tabla de mortalidad femenina para el año o periodo en estudio.

Las l_x para mujeres que se necesitan se toman de la tabla de mortalidad para mujeres que se calculo anterior mente.

Para obtener la ${}_{x+2.5} P_0(f)$ se hace:

```
>#P'(x) Probabilidad de sobrevivencia hasta la edad x
>P=c(1:7)*0
P[7]=lxM[11]/lxM[1]
for(i in 1:6){
P[i]= c(((sum(lxM[i+4], lxM[i+5]))/2)/lxM[1])
}
```

Se tiene:

```
> P
[1] 0.9934421 0.9917528 0.9898249 0.9872814 0.9834613 0.9773275
0.9735225
```

Una forma fácil para calcular la TNR sería con las siguientes tablas:

La primera para calcular las tasas específicas de fecundidad conforme los tres años consecutivos.

<i>edades</i>	Nac^{z-1}	Nac^z	Nac^{z+1}	P_{5fx}	${}_5f_x^z$
---------------	-------------	---------	-------------	-----------	-------------

Donde:

edades Son los quinquenios desde 15-19 hasta 45-49

Nac^i = Los nacimientos para el año i .

P_{nfx} = La población femenina de x a $x + 5 - 1$ años de edad a mitad del año z .

${}_n f_x^z$ = La tasa de fecundidad específica de x a $x + 5 - 1$ años de edad para el año z .
Calculada con el promedio de los nacimientos de los tres años entre la población correspondiente.

La segunda tabla para calcular la TNR es:

<i>edades</i>	${}_5f_x^z$	$k * {}_5f_x^z$	$5 * k * {}_5f_x^z$	${}_5L_x$	${}_{x+2.5}P_0 = \frac{{}_5L_x}{5 * l_0}$	$5 * k * {}_5f_x^z * {}_{x+2.5}P_0$
---------------	-------------	-----------------	---------------------	-----------	---	-------------------------------------

Donde:

edades Son los quinquenios desde 15-19 hasta 45-49

${}_5f_x^z$ Viene de la tabla anterior.

$k * {}_5f_x^z$ = La tasa de fecundidad específica sólo para los nacimientos femeninos de x a $x + 5 - 1$ años de edad para el año z .

$5 * k * {}_5f_x^z$ = Es el número medio de los nacimientos femeninos entre las edades x y $x + 5 - 1$.

${}_5L_x$ = Es L_x de la tabla de mortalidad femenina para el año correspondiente.

${}_{x+2.5}P_0$ = Es la probabilidad de sobrevivencia femenina entre el nacimiento y la edad $x + 2.5$

$5 * k * {}_n f_x^z * P_0$ = Es el número esperado de nacimientos femeninos en cada grupo de edad que sobrevivirán a edad de la madre al momento del parto.

Por lo tanto la tasa neta de reproducción es la suma de toda la columna de $5 * k * {}_n f_x^z * P_0$.

En R-Project queda:

```
>#TNR-tasa neta de reemplazo  
>TNR=DFN*100/205
```

Con el siguiente resultado:

```
> TNR  
[1] 1.152124  
  
> TNR_90  
[1] 1.302039
```

Como la TNR > 1 y la TGF >2.1 Esto quiere decir que sí existe la capacidad de reemplazo del estado de Nuevo León.

Después se calcula la descendencia final neta (DFN) que es el número de hijas e hijos que tiene una mujer en edad fértil considerando mortalidad. Se calcula como la TNR pero sin considerar el índice de feminidad al nacimiento k .

```
>#DFN  
>DFNi=c(1:7)*0  
for(i in 1:7){  
DFNi[i]=c(P[i]*tef[i])  
}  
DFN=sum(DFNi)*5  
  
> DFN  
[1] 2.361853  
  
> DFN_90  
[1] 2.669179
```

3.3.2.5.5. Edad media de la fecundidad

La interpretación de este índice es analítica pues entrega un valor que no existe pero se refiere a la edad promedio que se situarían todos los nacimientos de cada mujer. Su fórmula es:

$$m = \frac{\sum x_i * {}_5f_x^z}{\sum {}_5f_x^z}$$

Donde:

x_i = El punto medio del intervalo de la edad

${}_5f_x^z$ = Son las tasas de fecundidad específicas de los quinquenios.

La función en R-project es:

```
#Edad media de la fecundidad
mi=c(1:7)
for(i in 1:7){
mi[i]=fec[i,5]*tef[i]
}
```

Por quinquenios se tiene que:

```
> mi
[1] 1.13298461 2.82194336 3.84906010 3.17839041 1.53844766
0.32944678 0.03175402
```

Para el promedio se hace:

```
m=(sum(mi))/(stef)
```

Da como resultado:

```
> m  
[1] 26.98544
```

```
> m_90  
[1] 26.97419
```

Donde **m** es la edad media de la fecundidad y dice que la edad promedio donde se realizan los nacimientos es a los 27 años.

3.3.2.6. Conclusiones

El tema de fecundidad es uno de los más relevantes para la asignatura de demografía, dado que permite conocer los índices de crecimiento de la población. Una variable fundamental para determinar la calidad de vida de la población. Conocer la fecundidad en cada región y sobre todo en cada país permite incidir sobre las políticas públicas. Por ejemplo, cuando una población empieza a cambiar la estructura de la pirámide de población con una disminución en las edades jóvenes, esto significa que la fecundidad está disminuyendo, lo que conlleva al ensanchamiento de los grupos en edades más avanzadas. En México se observa un envejecimiento poblacional que puede afectar el programa de pensiones. De este modo se detectan los problemas que pueden surgir años después, previniéndolos antes de que ocurran.

En R se utiliza la función “**plot**” para conocer otro tipo de gráficas que son muy importantes pues ilustran muy bien algunos fenómenos que presentan datos en "x", "y". En este capítulo también se crean funciones para enseñar como se programa en R y como se acoplan a cualquier tipo de trabajo. Además de que lo hace de una manera muy sencilla.

3.3.3. Migración

La migración se reconoce como componente de cambio de la población y está relacionado con el movimiento, o desplazamiento de un lugar de origen a otro de destino donde implique un traslado de residencia de los individuos de la población.

Para que un desplazamiento se considere migratorio tienen que haber una residencia fija, es decir que no se toman en cuenta los traslados de poca duración y además tiene que ser de alguna delimitación administrativa o geográfica, la más importante es la migración internacional, aunque la migración interna de cada país es importante por los volúmenes de gente que involucra, Por lo que cada investigador debe dejar en claro que es lo que desea estudiar.

En la migración se mide a los migrantes que es “toda persona que traslada su lugar de residencia habitual de una división geográfica o administrativa a otra” (Wolti, Carlos 1997). Un migrante es un emigrante respecto de su lugar de residencia original y un inmigrante respecto de su lugar de residencia actual.

3.3.3.1. Crecimiento natural y crecimiento social

Debemos recordar que la población cambia según el crecimiento natural que es la diferencia entre la tasa bruta de natalidad y la tasa bruta de mortalidad. Además cambia según el crecimiento social que varía por la migración de la población, es decir, la diferencia que existe entre la inmigración y la emigración.

La ecuación compensadora es una forma de estimar el cambio total de una población, o para estimar cualquier componente desconocido del movimiento de la población a partir de otros componentes conocidos. Queda de la siguiente forma:

$$r' = b - m + i - e$$

Donde:

r' = Es la tasa de crecimiento geométrico con que crece la población que se escribió anteriormente como r_p .

b = Es la tasa bruta de natalidad, que se maneja como tb_n .

m = Es la tasa bruta de mortalidad, se uso como tbm .

i = Es el número de Inmigrantes en el área geográfica.

e = Es el número de emigrantes del área.

Por lo tanto el crecimiento natural es:

$$CN = b - m$$

Entonces en R-Project queda:

```
>#Crecimiento natural para hombres del 1990
>CNH_90=tb_nM_90-tbmH_90
>#Crecimiento natural para mujeres del 1990
>CNM_90=tb_nF_90-tbmM_90

>#Crecimiento natural para hombres del 2000
>CNF=tb_nM-tbmH
>#Crecimiento natural para mujeres del 2000
>CNM=tb_nF-tbmM
```

Los resultados son:

```
> CNH_90
[1] 41.12844

> CNM_90
[1] 43.95276

> CNF
[1] 30.24899

> CNM
[1] 33.12764
```

El crecimiento social queda dado de la siguiente manera:

$$CS = i - e$$

Para medir la migración también se utilizan los métodos directos que vienen de las preguntas que puedes encontrar en los censos según el lugar de nacimiento, al lugar de residencia anterior o al lugar de residencia en una fecha fija anterior. Para después determinar si el individuo es un migrante o un no migrante.

3.3.3.2. Migración según lugar de nacimiento (migración absoluta)

Esta pregunta permite estudiar la migración interna y la internacional. La internacional según el país de nacimiento, si es diferente al lugar donde se realizó el censo es inmigrante para este país y emigrante para el país de nacimiento. La migración interna depende del lugar de nacimiento y su residencia habitual; si coinciden el individuo es "no migrante" y si no coinciden es migrante volviéndose emigrante para el lugar de nacimiento e inmigrante para la residencia habitual.

Un problema de esta pregunta es que se desconoce el momento en que ocurrió la migración evitando conocer el motivo y las circunstancias de la migración; otro problema que causa confusión entre el lugar habitual de la madre y el lugar del parto; otro error de esta pregunta es que no mide la migración de retorno pues si el individuo se trasladó de su residencia habitual y regresa antes del censo, esto no se nota en la declaración.

Para observar más fácilmente de que estados son los inmigrantes de la población del estado de Nuevo León se puede hacer una gráfica circular con los diez estados con mayor número de estos inmigrantes al estado.

Un ejemplo son los datos que se van a trabajar son los diez estados con el número más alto de personas que nacieron fuera de Nuevo León y que ahora residen ahí. Se puede tomar en cuenta a los que nacieron en otro país ya que también son inmigrantes para el estado. Los datos se guardan en un archivo "csv" como antes se ha hecho.

En R-Project primero se tienen que introducir los datos por lo que usamos la función `read.csv`

```
># importar datos de migración por lugar de nacimiento del 2000
>Mn_00=read.csv("C:/migracion_nacimiento.csv",header=TRUE,sep="," ,
quote="\\" , dec="." , fill=TRUE)
```

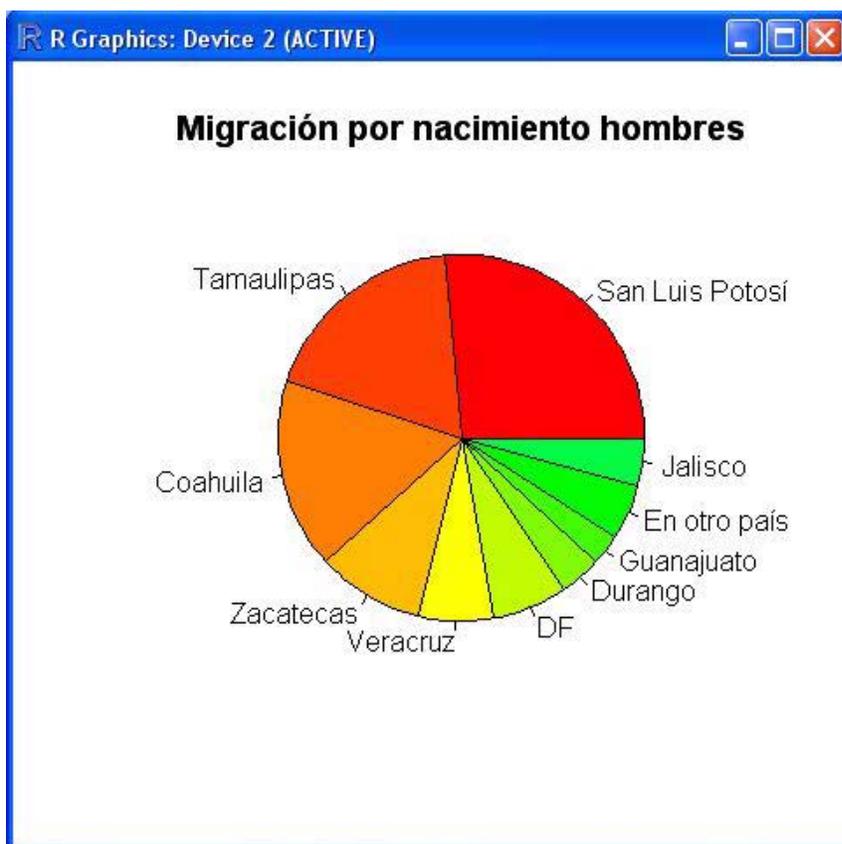
Los archivos tienen los títulos de "estados", "hombres", "mujeres" y "pob total" (población total) para cada columna.

Ahora con la función "pie" se forman las gráficas circulares (pastel) por ejemplo:

```
># Gráfica circular de migración por nacimiento de hombres del
censo del 2000
>G_Mn_00H=pie(Mn_00[1:10,2],Mn_00[1:10,1],col=rainbow(24),main=
"Migración por nacimiento hombres")
```

Donde los argumentos son los datos a graficar, los nombres de los estados correspondientes, los colores y el título del gráfico.

La pantalla de gráficos se dibuja de la siguiente manera:



Se ver a los inmigrantes que nacieron en otro país u otro estado y ahora residen en Nuevo León

3.3.3.3. Migración según lugar de residencia anterior

A veces se hace la pregunta para saber el lugar de residencia anterior de cada individuo, esto permite identificar a las personas como migrantes (si existe residencia anterior) o no migrantes (si no existe residencia anterior). De estos últimos se dividen en emigrantes para el lugar de la residencia anterior e inmigrantes para el lugar actual.

Una ventaja es que sí se captan a los migrantes de retorno. El problema de esta pregunta es que sólo mide el último traslado por lo que no se puede estimar la frecuencia de la migración. Además de que no siempre se conoce la ubicación de la residencia anterior, más las confusiones entre las palabras de residencia y domicilio ya que si el individuo se muda de una vivienda a otra en el mismo lugar de residencia no se notará su calidad de migrante.

3.3.3.3.1. Migración según lugar de residencia anterior a una fecha determinada

Al contarse con la residencia anterior a “n” años sirve, al compararla con la residencia al momento del censo, para conocer áreas de atracción o rechazo de la población y así se puede estudiar algunas hipótesis sobre factores que afectaron a la migración.

Pero los problemas vienen con la extensión del intervalo de la fecha pues no puede ser muy largo (omites movimientos) ni muy corto (capta pocos movimientos moviéndose una muestra poco representativa y puede captar desplazamientos ocasionales). En el caso de los censos por convención se hace la pregunta a los 5 años ya que los censos son cada 10.

Además existen combinaciones que permiten obtener más información de estas preguntas y no debemos olvidar que “los niños no adoptan decisiones migratorias por sí mismos” (Welti, Carlos 1997).

Para estudiar como está la migración en el estado de Nuevo León de los dos censos que estamos trabajando se utilizan las tablas de "Migración por lugar de residencia en

enero de 1985" y "Migración por lugar de residencia en enero de 1995". Estas tablas las ordenas de forma descendente según la cantidad de personas que vivían en otro estado, luego se toman los primeros diez estados y se hacen unas gráficas (circulares) para poder observar de cuales estados principalmente son los inmigrantes de Nuevo León.

Para hacer estas gráficas en R-Project primero se introducen los datos:

```
># importando datos de migración por lugar de residencia en enero
de 1985
>Ra_85=read.csv("C:/residencia_1985.csv",header=TRUE,sep=" ",
quote="\\"",dec=".",fill = TRUE)

># importando datos de migración por lugar de residencia en enero
de 1995
>Ra_95=read.csv("C:/residencia_1985.csv",header=TRUE,sep=" ",
quote="\\"",dec=".",fill = TRUE)
```

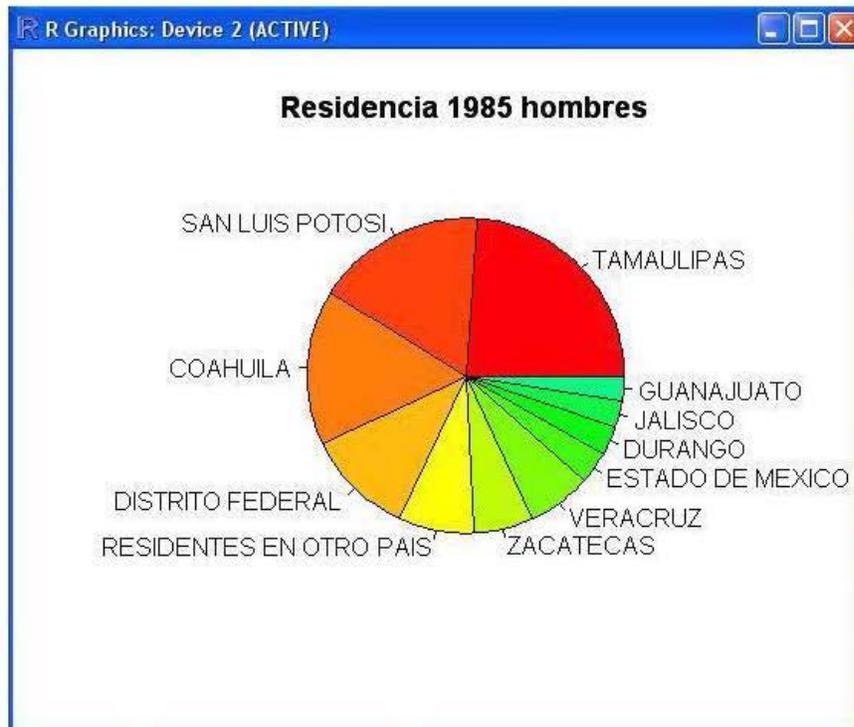
Los archivos tienen los títulos de "Estados", "Hombres" y "Mujeres" para cada columna. Además en los datos no excluimos a los que vivieron en otro país.

Luego con la función "pie" se forman las gráficas circulares (pastel) por ejemplo:

```
># Grafico de residencia en 1985 para hombres
> G_Ra85H=pie( Ra_85[1:11,2],Ra_85[1:11,1], col=rainbow(24),
main="Residencia 1985 hombres" )
```

Donde los argumentos son los datos a graficar, los nombres de los estados correspondientes, los colores y el título de la gráfica.

Por lo tanto en la pantalla de gráficos se dibuja de la siguiente manera:



Esta gráfica muestra a los inmigrantes (hombres) de Nuevo León que hace cinco años residían en otro estado o en otro país para el censo de 1990.

Para el censo del 2000 el comando en R-Project para hacer la gráfica circular de los principales estados donde residían los ahora inmigrantes del sexo femenino es:

```
># Gráfico de residencia en 1985 para mujeres
>G_Ra85M=pie(Ra_95[1:11,3],Ra_85[1:11,1],main="Residencia 1995
mujeres")
```

Los datos de migración generalmente están incompletos, pero se puede obtener información gracias a los métodos indirectos.

3.3.3.4. Métodos indirectos para medir la migración

Cuando no se dispone de los datos de los censos para aplicar la ecuación compensadora se puede utilizar otros métodos que utilizan la tabla de mortalidad.

3.3.3.4.1. Método prospectivo

Se dispone para los tiempos t y $t+10$ la distribución por edades de la población, entonces la migración neta (el efecto neto de la inmigración y la emigración sobre la población estudiada) durante los 10 años para cada grupo de edad x es:

$$M_x^p = Pob_{t+10} - \frac{Pob_t}{S_x} < 0 \Rightarrow \text{Emigrante}$$
$$M_x^p = Pob_{t+10} - \frac{Pob_t}{S_x} > 0 \Rightarrow \text{Inmigrante}$$

Donde:

$$S_x = \frac{L_x}{L_{x+10}}$$

L_x es el promedio de L_x del año 1990 y del año 2000 para el caso que se trabaja y los grupos de edad se trabajan por decenios.

En el programa R-Project primero se formaron los grupos de población de decenios de la siguiente manera:

```
>#población en decenios del año 2000 para hombres
>pobactHD=c(1:9)*0
pobactHD[1]=sum(TMh[1,5],TMh[2,5],TMh[3,5])
for(i in 2:9){
pobactHD[i]=sum(TMh[i*2,5],TMh[2*i+1,5])
}
```

Después se hacen los decenios de L_x que se trabajaron en las tablas de mortalidad.

```
>#Lx de la tabla de mortalidad en decenios
>LxD=c(1:9)*0
LxD[1]=sum(Lx[1],Lx[2],Lx[3])
for(i in 2:9){
LxD[i]=sum(Lx[i*2],Lx[2*i+1])
}
```

Lo que sigue es hacer un promedio de las L_x que se calcularon:

```
>LxH90_00=c((LxD+LxD90)/2)
```

Además se tiene el caso para mujeres que será:

```
>LxM90_00=c((LxMD+LxMD90)/2)
```

Ahora se calcula la S_x como sigue:

```
>Sx para el caso de los hombres
>SxH=c(1:8)*0
SxH[8]=LxH90_00[9]/(LxH90_00[8]+LxH90_00[9])
for(i in 1:7){
SxH[i]=LxH90_00[i+1]/LxH90_00[i]
}
```

Para calcular la fórmula para saber si son emigrantes o inmigrantes se hace:

```
>fórmula de migracion método prospectivo
>MigHP=c(1:10)*0
MigHP[1]=0
MigHP[9]=pobactHD[9]-((pobantHD[8]+pobantHD[9])*SxH[8])
for(i in 2:8){
MigHP[i]=pobactHD[i]-((pobantHD[i-1])*SxH[i-1])
}
MigHP[10]=sum(MigHP[1:9])
```

Se puede hacer una función para que el programa diga que tipo de migrantes son:

```
>Función que dice emigrante o inmigrante
>Migracion=function(Mig){
if(Mig<0)
return("emigrantes")
else
return("inmigrantes")
}
```

Para que funcione sobre todo el vector **MigHP**.se hace otra función

```
>vector con información de emigrantes o inmigrantes
>MigrantesHP=c(1:10)*0
MigrantesHP[1]="nulo"
for(i in 2:10){
MigrantesHP[i]=Migracion(MigHP[i])
}
```

Por último se puede calcular la tasa de migración para hombres de los años 1990 y 2000 con el método prospectivo de la forma siguiente:

```
>tmigHP=1000*MigHP[10]/sum(pobactHD)
```

Que da como resultado:

```
> tmigHP
[1] -56.56387
```

En el caso de las mujeres es:

```
> tmigMP
[1] -51.06878
```

Si a estas variable se les aplica la función de **Migracion** dan:

```
> Migracion(tmigHP)
[1] "emigrantes"
```

```
> Migracion(tmigMP)
[1] "emigrantes"
```

Por lo tanto hubo emigración

3.3.3.4.2. Método retrospectivo

Para este método se calcula de manera similar al anterior

$$M_x^R = Pob_t - \frac{Pob_{t+10}}{S_x} < 0 \Rightarrow Emigrante$$

$$M_x^R = Pob_t - \frac{Pob_{t+10}}{S_x} > 0 \Rightarrow Inmigrante$$

Donde de igual forma:

$$S_x = \frac{L_x}{L_{x+10}}$$

En R-Project cambia la fórmula de migración para hacerla por el método retrospectivo.

```
>MigHR=c(1:9)*0
for(i in 1:8){
MigHR[i]=pobantHD[i]-(pobactHD[i]*SxH[i])
}
MigHR[9]=sum(MigHR[1:8])
```

Después se utilizala función de **Migracion** y se hace otra función de **Migrantes** para sacar el vector.

```
>MigrantesHR=c(1:9)*0
for(i in 1:9){
MigrantesHR[i]=Migracion(MigHR[i])
}
```

Para terminar se vuelve a calcular la tasa de migración para hombres de los años 1990 y 2000 del estado de Nuevo León que queda:

```
># tasa de migracion para hombres método retrospectivo
>tmigHR=1000*MigHR[9]/sum(pobantHD)
```

El resultado para hombres y para mujeres respectivamente es:

```
> tmigHR
[1] -461.2437
```

```
> tmigMR
[1] -463.6797
```

Al volver a aplicar la función de Migración dice:

```
> Migracion(tmigHR)
[1] "emigrantes"
```

```
> Migracion(tmigMR)
[1] "emigrantes"
```

Por lo tanto hubo emigración

3.3.3.4.3. Método promedio

El método promedio se utiliza para mejorar el resultado de los métodos prospectivo y retrospectivo por lo que da la siguiente fórmula:

$$M_x^M = M_x^P + M_x^R$$

En el programa queda:

```
>método promedio para hombres
>MigHM=c(1:10)*0
MigHM[10]=MigHP[10]
for(i in 1:9){
MigHM[i]=MigHP[i]+MigHR[i]
}
```

Se puede volver a usar las funciones de Migración y de Migrantes para al final calcular la tasa de migración por el método promedio que es:

```
>#Tasa de migración hombres prospectivo
>tmigHM=1000*2*MigHM[10]/(sum(pobantHD)+sum(pobactHD))
```

El resultado es:

```
> tmigHM
[1] -65.21413
```

La función dice que:

```
> Migracion(tmigHM)
[1] "emigrantes"
```

Por lo tanto hubo emigración

En el caso de las mujeres la tasa de migración por el método medio es:

```
>#Tasa de migración mujeres prospectivo
>tmigMM=1000*2*MigMM[10]/(sum(pobantMD)+sum(pobactMD))
```

Su cálculo es:

```
> tmigMM
[1] -58.80741
```

Por lo que se puede hacer:

```
> Migracion(tmigMM)
[1] "emigrantes"
```

Se concluye que hubo emigración

Para facilitar mejor la comprensión de los resultados se puede hacer la siguiente tabla de las tasas de emigración:

Método	Mujeres	Hombres
Prospectivo	-51.06878	-56.56387
Retrospectivo	-463.6797	-461.2437
Promedio	-58.80741	-65.21413

3.3.3.5. Conclusiones

La migración es un componente importante del cambio poblacional, porque destaca el desplazamiento de la población entre zonas geográficas. Cualquiera piensa, cuando se habla de este tema, que en México los emigrantes son aquellos que se van a Estados Unidos, ya que causan un gran conflicto con aquél país, pero no se debe olvidar la migración interna, pues también es un factor relevante que afecta a la población. Sin olvidar que los inmigrantes vienen principalmente de Centroamérica. Así que el estudio de la migración en nuestro país se puede abordar desde distintas ópticas.

En este capítulo R-Project se utiliza para un nuevo tipo de gráfica, llamado "pastel" que permite visualizar de otra manera los problemas que de esta manera se exponen.

Conclusiones generales

La presente tesis genera una nueva forma de impartir asignaturas que requieren numerosos cálculos estadísticos. Une la asignatura de “teoría de Demografía I” a un programa de computación con el objeto de facilitar el aprendizaje de los conceptos básicos de dicha materia.

Este modelo creemos que facilita el aprendizaje de la materia de demografía porque simplifica los cálculos de las fórmulas estadísticas típicas de la materia para la evaluación y corrección de datos demográficos. Se estima que es superior a un programa como el “Excel” porque se pueden manejar mayor cantidad de datos, además posee funciones más fáciles de usar para calcular pirámides de población, y crear funciones que el usuario considere necesaria para sus propias necesidades.

Por otra parte, el programa de computación escogido, es relativamente más fácil de usar que los programas “compilados”, porque no necesita conocer a profundidad un lenguaje de programación. Otra ventaja del programa utilizado es que trabaja línea por línea y no como los programas que compilan todo el archivo antes de ejecutarse.

Este programa llamado “R-Projet” además tiene como ventaja que puede recurrir a una gran cantidad de funciones estadísticas que permiten al estudiante de demografía, actuaría o ciencias sociales utilizarlo para otras muchas actividades profesionales, como por ejemplo, las clases de Probabilidad, Estadística, Regresión y cualquier otra donde se necesite realizar un análisis cuantitativo de datos y hacer los cálculos necesarios.

Esta herramienta de trabajo, invita a los estudiantes de Actuaría, inicialmente, a practicar en todo momento las nociones de programación que se imparten en los primeros semestres y aplicarlos constantemente. De esta manera estará mucho mejor preparado cuando se enfrente al mundo laboral. Además de que apoya la formación profesional de cada estudiante sin importar a que área se va a enfocar.

Otra ventaja de usar R-Projet en la clase de Demografía es que le permite a la profesora o profesor que imparte la asignatura de Demografía I agilizar y modernizar sus métodos pedagógicos, sin perder de vista que los alumnos aprenden el por qué de cada cálculo y cómo se realiza éste.

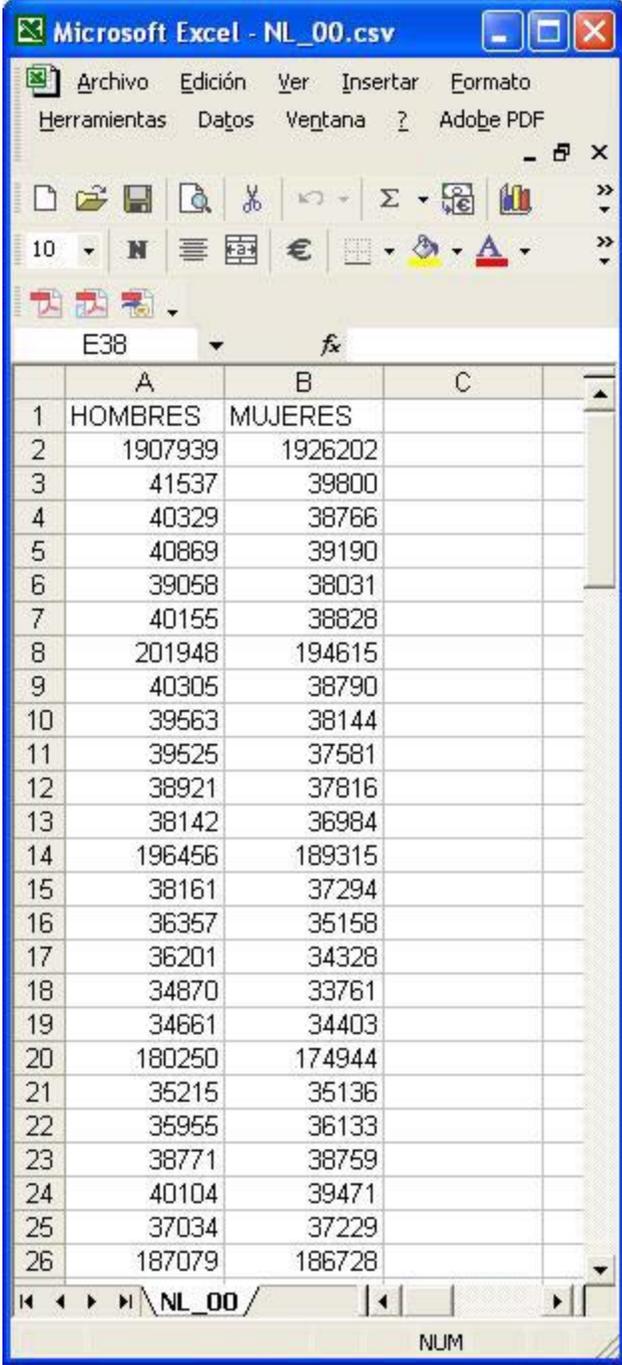
Recomendaciones

Una segunda fase de esta investigación requiere que este modelo de aprendizaje basado en dos conocimientos (programación y demografía) sea puesto a prueba para determinar en qué medida funciona mejor que el método actual. Se sugiere hacerle un experimento pedagógico que contemple la existencia de un grupo experimental y otro grupo denominado control. El primero aplicaría los resultados de esta investigación y el segundo aplicaría un aprendizaje tradicional.

Los resultados se analizarían mediante técnicas estadísticas para determinar si el conocimiento adquirido por el grupo experimental es significativamente superior al del grupo control. Con ello, se estaría impulsando una evaluación de la educación de manera objetiva, sistemática, verificable y apoyaría los cambios pertinentes para seguir a la vanguardia en la carrera de actuaría.

Apéndice de archivos

Los datos de población actual, es decir, los datos de cuantos hombres y mujeres se encuentran en el Estado de Nuevo León del año 2000 divididos por edades y quinquenios. Estos datos se pueden ingresar en Excel y guardarlos como CSV.

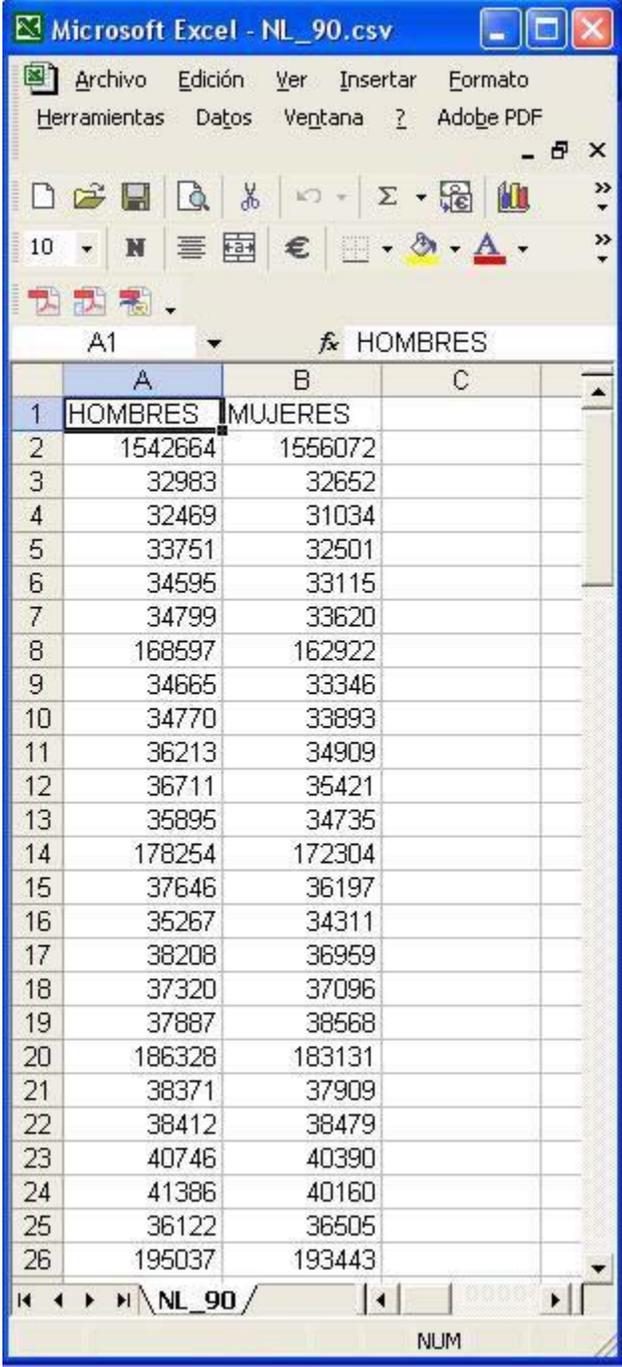


The screenshot shows a Microsoft Excel window titled "Microsoft Excel - NL_00.csv". The menu bar includes "Archivo", "Edición", "Ver", "Insertar", "Formato", "Herramientas", "Datos", "Ventana", and "Adobe PDF". The toolbar contains various icons for file operations and editing. The active cell is E38. The spreadsheet displays the following data:

	A	B	C
1	HOMBRES	MUJERES	
2	1907939	1926202	
3	41537	39800	
4	40329	38766	
5	40869	39190	
6	39058	38031	
7	40155	38828	
8	201948	194615	
9	40305	38790	
10	39563	38144	
11	39525	37581	
12	38921	37816	
13	38142	36984	
14	196456	189315	
15	38161	37294	
16	36357	35158	
17	36201	34328	
18	34870	33761	
19	34661	34403	
20	180250	174944	
21	35215	35136	
22	35955	36133	
23	38771	38759	
24	40104	39471	
25	37034	37229	
26	187079	186728	

The status bar at the bottom shows the file path "\NL_00/" and the word "NUM".

Los datos de población anterior, es decir, los datos de cuantos hombres y mujeres se encuentran en el Estado de Nuevo León del año 1990 divididos por edades y quinquenios. Estos datos se pueden ingresar en Excel y guardarlos como CSV.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - NL_90.csv". The menu bar includes "Archivo", "Edición", "Ver", "Insertar", "Formato", "Herramientas", "Datos", "Ventana", and "Adobe PDF". The toolbar contains various icons for file operations, editing, and formatting. The active cell is A1, containing the text "HOMBRES". The spreadsheet displays population data for men and women in 1990, organized by age groups and quinquenniums. The data is presented in a table with columns A, B, and C, and rows 1 through 26. The first row (row 1) has "HOMBRES" in column A and "MUJERES" in column B. The subsequent rows (rows 2-26) contain numerical values for men and women. The status bar at the bottom shows the file path "\NL_90/" and the data type "NUM".

	A	B	C
1	HOMBRES	MUJERES	
2	1542664	1556072	
3	32983	32652	
4	32469	31034	
5	33751	32501	
6	34595	33115	
7	34799	33620	
8	168597	162922	
9	34665	33346	
10	34770	33893	
11	36213	34909	
12	36711	35421	
13	35895	34735	
14	178254	172304	
15	37646	36197	
16	35267	34311	
17	38208	36959	
18	37320	37096	
19	37887	38568	
20	186328	183131	
21	38371	37909	
22	38412	38479	
23	40746	40390	
24	41386	40160	
25	36122	36505	
26	195037	193443	

Los datos para la mortalidad infantil del año 2000

	A	B
1		2000
2	Total	38,621
3	Menores de 1 semana	17,533
4	De 7 a 13 días	3,158
5	De 14 a 20 días	1,738
6	21 días	179
7	22 días	169
8	23 días	180
9	24 días	132
10	25 días	139
11	26 días	142
12	27 días	152
13	28 días	155
14	29 días	129
15	Edad en días no especificada	402
16	De 1 a 11 meses	14,413
17	nacidos vivos	2,798,339
18		

Datos para la tabla de mortalidad de hombres del año 2000

Microsoft Excel - TMH_00.csv

Adobe PDF

Arial 10

A1 hombres

	A	B	C	D	E
1	hombres	1999	2000	2001	Población
2	< 1	564	613	580	201948
3	1-4	90	73	85	203576.029
4	5-9	41	60	44	198039.755
5	10-14	51	48	56	181703.108
6	15-19	159	149	157	188587.161
7	20-24	213	209	207	197161.733
8	25-29	254	244	246	184747.454
9	30-34	229	213	246	160223.331
10	35-39	274	290	266	135744.57
11	40-44	325	322	316	110984.561
12	45-49	363	339	378	83282.0192
13	50-54	478	466	470	71457.4563
14	55-59	587	616	591	54312.3442
15	60-64	714	748	761	45300.2733
16	65-69	878	828	856	31664.2237
17	70-74	836	896	923	22789.2492
18	75-79	882	902	921	15675.3583
19	80-84	642	681	735	7728.80848
20	85 o más	1085	1069	1086	7476.79307
21					

TMH_00

Listo NUM

Los datos de la tabla de mortalidad de mujeres del año 2000

Microsoft Excel - TMM_00.csv

Adobe PDF

Arial 10

A1 mujeres.

	A	B	C	D	E
1	mujeres.	1999	2000	2001	Población
2	< 1	454	447	450	194615
3	1-4	58	76	66	196190.382
4	5-9	33	42	45	190847.479
5	10-14	40	39	31	176360.148
6	15-19	68	66	51	188239.538
7	20-24	70	73	66	197233.76
8	25-29	75	74	90	187369.552
9	30-34	103	114	82	164954.563
10	35-39	135	132	127	138654.376
11	40-44	177	172	189	115157.701
12	45-49	200	200	212	86760.6763
13	50-54	312	312	326	74208.8872
14	55-59	425	430	426	56303.1062
15	60-64	547	607	551	48074.0279
16	65-69	649	687	678	34742.9814
17	70-74	797	761	738	26214.4988
18	75-79	764	796	823	17917.8781
19	80-84	742	732	825	9940.82345
20	85 o más	1415	1471	1465	10840.0441
21					

TMM_00

Listo NUM

Datos para la tabla de mortalidad de hombres del año 1990

Microsoft Excel - TMH_90.csv

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana

Adobe PDF

I26 fx

	A	B	C	D	E
1	hombres	1989	1990	1991	Población
2	< 1	875	818	796	33196.3246
3	01--04	156	155	151	136491.112
4	05--09	66	62	90	179406.896
5	10--14	78	78	99	187533.116
6	15-19	167	206	228	196298.443
7	20-24	208	268	279	169603.9
8	25-29	229	230	268	132620.238
9	30-34	196	244	262	111617.269
10	35-39	212	217	263	89346.1517
11	40-44	271	262	267	71872.8658
12	45-49	306	351	347	61140.9006
13	50-54	385	434	456	49918.7856
14	55-59	501	513	512	37196.0272
15	60-64	570	618	653	29737.095
16	65-69	662	690	719	22333.5185
17	70-74	585	617	675	13974.8042
18	75-79	711	707	689	10646.4155
19	80-84	630	624	669	6422.27047
20	85 o más	775	770	796	5304.08484
21					

TMH_90

Dibujo Autoformas

Listo

Datos para la tabla de mortalidad de mujeres del año 1990

Microsoft Excel - TMM_90.csv

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana

Adobe PDF

H24

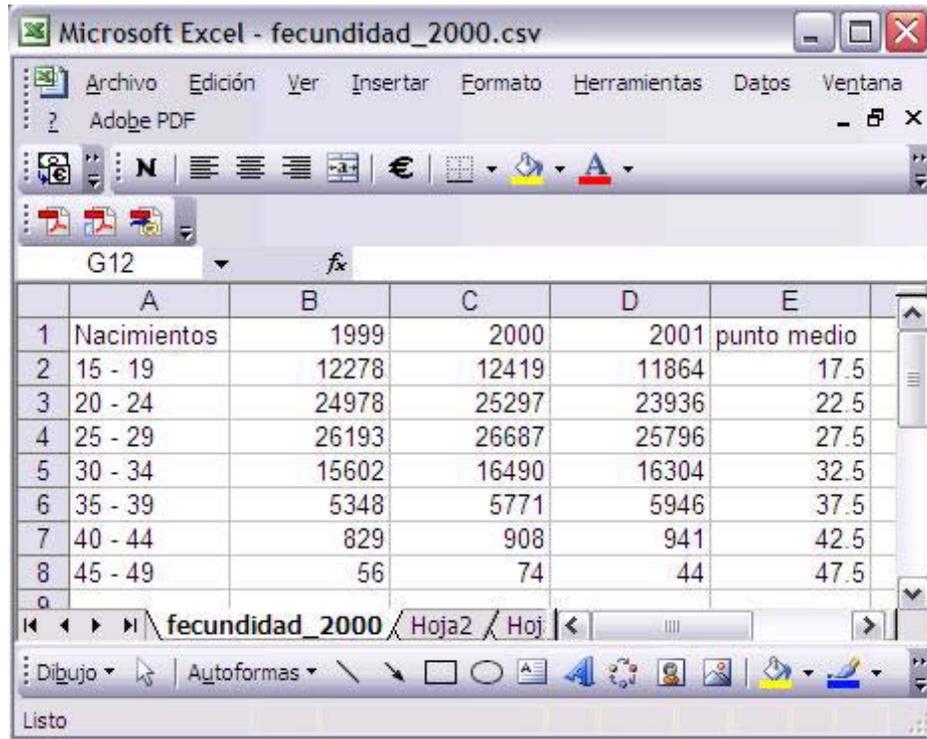
	A	B	C	D	E
1	mujeres	1989	1990	1991	Población
2	< 1	667	692	617	32864.0538
3	01--04	119	150	159	131116.02
4	05--09	60	49	49	173423.004
5	10--14	49	60	54	184320.318
6	15-19	62	89	72	194699.288
7	20-24	75	84	94	170107.611
8	25-29	98	101	89	136447.421
9	30-34	90	117	108	116732.209
10	35-39	84	134	143	92225.078
11	40-44	145	146	152	74330.6148
12	45-49	205	192	172	61486.7403
13	50-54	274	281	249	50994.0366
14	55-59	299	351	334	38394.7403
15	60-64	455	468	438	33065.3526
16	65-69	538	508	550	24060.2476
17	70-74	499	513	562	16100.8902
18	75-79	712	667	630	12571.1145
19	80-84	672	736	655	8054.97434
20	85 o más	1016	1011	1116	7092.76574
21					

TMM_90

Dibujo Autoformas

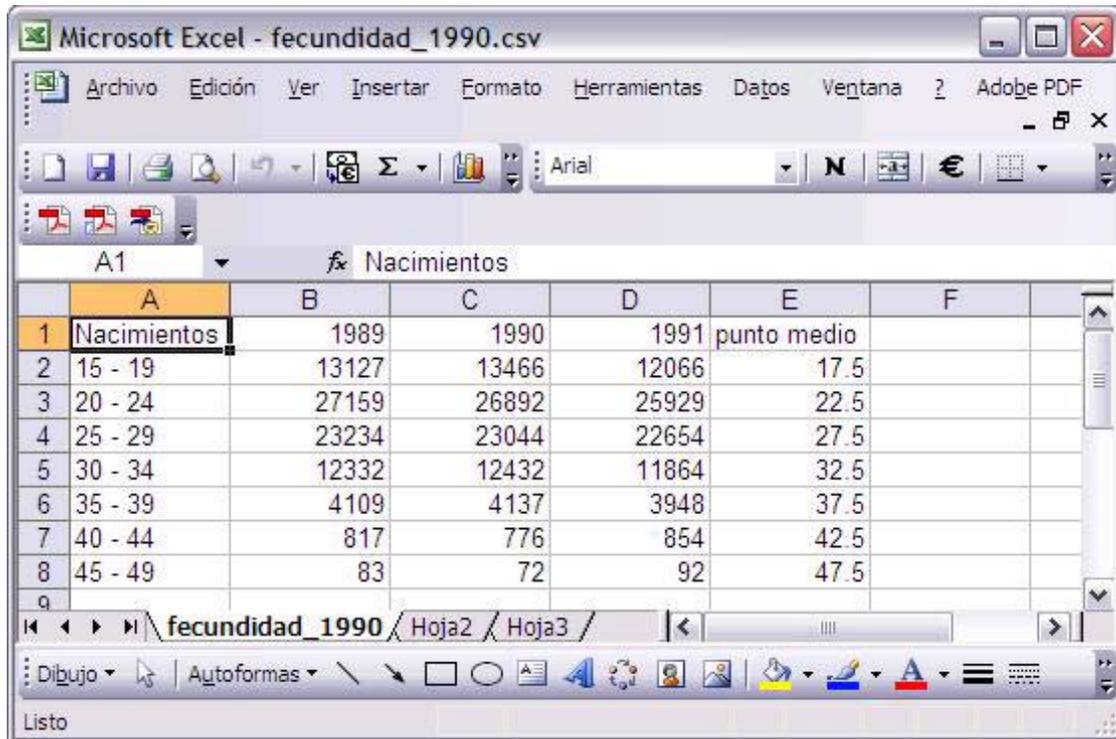
Listo

Los datos de la sección de fecundidad para el análisis del año 2000. Se calcula el punto medio de los quinquenios para facilitar su manejo.



	A	B	C	D	E
1	Nacimientos	1999	2000	2001	punto medio
2	15 - 19	12278	12419	11864	17.5
3	20 - 24	24978	25297	23936	22.5
4	25 - 29	26193	26687	25796	27.5
5	30 - 34	15602	16490	16304	32.5
6	35 - 39	5348	5771	5946	37.5
7	40 - 44	829	908	941	42.5
8	45 - 49	56	74	44	47.5

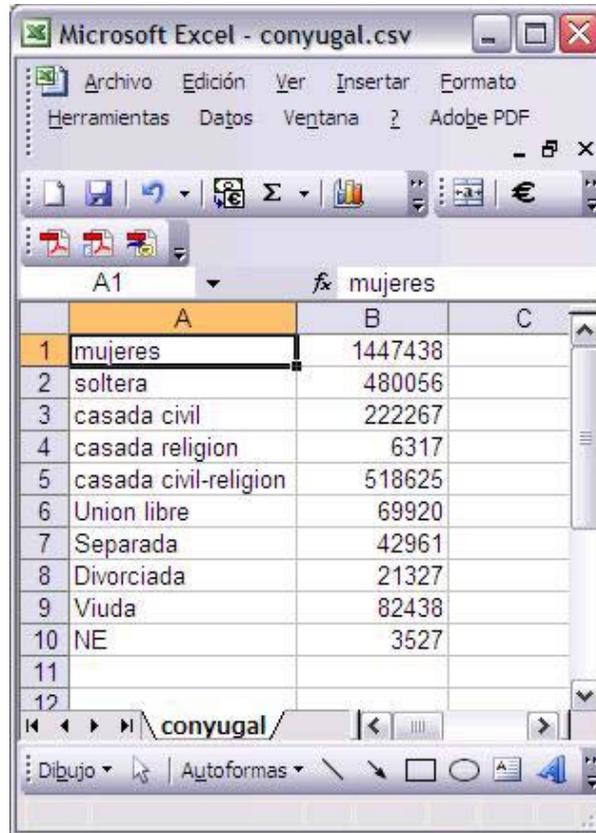
Los datos de la sección de fecundidad para el análisis del año 1990. Se calcula el punto medio de los quinquenios para facilitar su manejo.



The screenshot shows a Microsoft Excel spreadsheet titled "Microsoft Excel - fecundidad_1990.csv". The spreadsheet contains a table with the following data:

	A	B	C	D	E	F
1	Nacimientos	1989	1990	1991	punto medio	
2	15 - 19	13127	13466	12066	17.5	
3	20 - 24	27159	26892	25929	22.5	
4	25 - 29	23234	23044	22654	27.5	
5	30 - 34	12332	12432	11864	32.5	
6	35 - 39	4109	4137	3948	37.5	
7	40 - 44	817	776	854	42.5	
8	45 - 49	83	72	92	47.5	

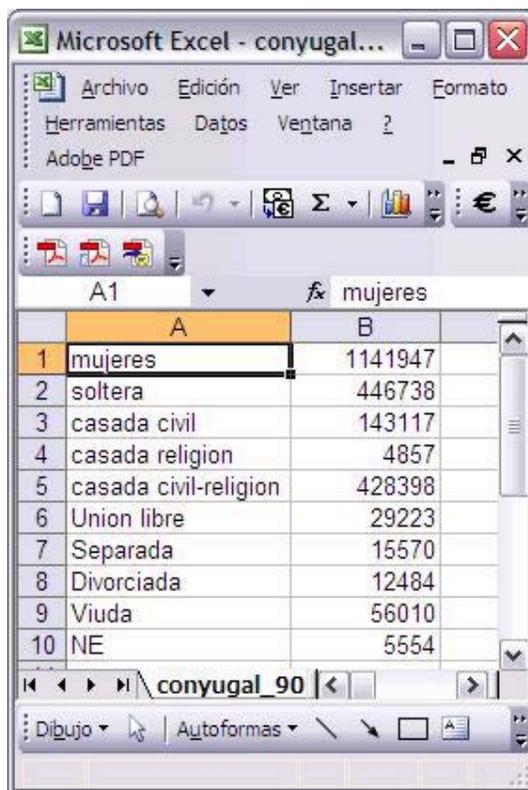
Datos del estado conyugal de las mujeres del Estado de Nuevo León del año 2000.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - conyugal.csv". The spreadsheet contains a table with two columns: marital status (A) and the number of women (B). The data is as follows:

	A	B	C
1	mujeres	1447438	
2	soltera	480056	
3	casada civil	222267	
4	casada religion	6317	
5	casada civil-religion	518625	
6	Union libre	69920	
7	Separada	42961	
8	Divorciada	21327	
9	Viuda	82438	
10	NE	3527	
11			
12			

Datos del estado conyugal de las mujeres del Estado de Nuevo León del año 1990.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - conyugal...". The menu bar includes "Archivo", "Edición", "Ver", "Insertar", "Formato", "Herramientas", "Datos", and "Ventana". The toolbar contains various icons for file operations and data analysis. The active sheet is named "conyugal_90". The data is presented in a table with two columns: "A" (marital status) and "B" (count). The data is as follows:

	A	B
1	mujeres	1141947
2	soltera	446738
3	casada civil	143117
4	casada religion	4857
5	casada civil-religion	428398
6	Union libre	29223
7	Separada	15570
8	Divorciada	12484
9	Viuda	56010
10	NE	5554

Migración por lugar de nacimiento de las diez entidades con mayor relevancia para el año 2000. En las columnas se tienen los estados, la población de hombres, mujeres y la población total.

Microsoft Excel - nacimiento_2000

Archivo Edición Ver Insertar Formato Herramientas Datos

Ventana ? Adobe PDF

F14

	A	B	C	D
1	Estados	hombres	mujeres	pob total
2	San Luis Potosí	101660	116309	217969
3	Tamaulipas	71251	73955	145206
4	Coahuila	64453	69411	133864
5	Zacatecas	36543	40122	76665
6	Veracruz	25659	23134	48793
7	DF	25066	21677	46743
8	Durango	14536	16305	30841
9	Guanajuato	10353	10537	20890
10	En otro país	18883	9721	9162
11	Jalisco	16366	7848	8518
12				

\\nacimiento_2000/

Dibujo Autoformas

Listo

Migración por lugar de nacimiento de las diez entidades con mayor relevancia para el año 1990.

Microsoft Excel - nacimientos_1990.csv

Archivo Edición Ver Insertar Formato Herramientas Datos
Ventana Adobe PDF

Arial N €

A1 Estados

	A	B	C	D
1	Estados	hombres	mujeres	pob total
2	San Luis Potosi	93215	101342	194557
3	Coahuila	61989	65965	127954
4	Tamaulipas	61612	63705	125317
5	Zacatecas	36625	39143	75768
6	Distrito Federal	16708	14559	31267
7	Durango	14127	15728	29855
8	Guanajuato	10497	10398	20895
9	Veracruz	10664	9888	20552
10	Jalisco	7319	7955	15274
11				

\\ nacimientos_1990 / Hoja2

Dibujo Autoformas

Listo

Migración por lugar de Residencia del año 1985 de las diez entidades con mayor relevancia. En las columnas se tienen los estados, la población de hombres, mujeres.

	A	B	C
1	Estados	Hombres	mujeres
2	TAMAUlipAS	12152	11842
3	SAN LUIS POTOSI	8803	10988
4	COAHUILA	8140	7935
5	DISTRITO FEDERAL	5570	5351
6	RESIDENTES EN OTRO PAIS	3936	3032
7	ZACATECAS	3043	3578
8	VERACRUZ	3348	3042
9	ESTADO DE MEXICO	1736	1648
10	DURANGO	1481	1584
11	JALISCO	1451	1550
12	GUANAJUATO	1268	1253
13			
14			

Migración por lugar de Residencia del año 1995 de las diez entidades con mayor relevancia. En las columnas se tienen los estados, la población de hombres, mujeres.



Microsoft Excel - residencia_1995

Archivo Edición Ver Insertar Formato Herramientas
Datos Ventana ? Adobe PDF

G15 fx

	A	B	C
1	Estados	hombres	mujeres
2	TAMAULIPAS	11237	11096
3	SAN LUIS POTOSÍ	8797	11911
4	VERACRUZ-LLAVE	9912	8400
5	COAHUILA DE ZARAGOZA	7081	6958
6	EN OTRO PAÍS	6805	4914
7	DISTRITO FEDERAL	4617	4072
8	MÉXICO	3295	2848
9	ZACATECAS	1791	2211
10	JALISCO	1962	1841
11	CHIHUAHUA	1434	1366
12	HIDALGO	1130	1602
13			

residencia_1995

Dibujo Autoformas

Listo

Glosario

ASCII es el "American Standard Code for Information Interchange". Es de facto el estándar del World Wide Web para el código utilizado por computadoras para representar todas las letras (mayúsculas, minúsculas, letras latinas, números, signos de puntuación, etc.).

Linux es un sistema operativo y un núcleo. Es uno de los paradigmas del desarrollo de software libre (y de código abierto), donde el código fuente está disponible públicamente y cualquier persona puede libremente usarlo, modificarlo y/o redistribuirlo.

RAM son las siglas de: Sigla de Random Access Memory (Memoria de acceso aleatorio). La RAM se usa para mantener los programas mientras se están ejecutando, y los datos mientras los procesa. La RAM es volátil, lo que significa que la información escrita en la RAM desaparecerá cuando se apague la alimentación de energía del ordenador.

Ubuntu es un sistema operativo basado en el núcleo Linux y en algunas herramientas del Proyecto GNU. La estructura técnica del sistema está basada en el Proyecto Debian, pero el ideario está inspirado en los principios de la corriente ubuntu, un movimiento humanista encabezado por el obispo Desmond Tutu, Premio Nobel de la Paz 1984. Económicamente el proyecto se sostiene con aportaciones de la empresa Canonical del sudafricano Mark Shuttleworth.

Glosario demográfico

Crecimiento absoluto natural es el saldo de los nacimientos menos las defunciones en un año determinado.

Crecimiento absoluto social es el saldo de los inmigrantes menos los emigrantes en un año determinado.

Crecimiento absoluto total es el balance del crecimiento absoluto natural y el

crecimiento absoluto social.

Datos pueden definirse como la información sistemática sobre las entidades de algún agregado estadístico.

Demografía es la ciencia que tiene como objetivo el estudio de las poblaciones humanas y trata de su dimensión, estructura, evolución y características generales, todo desde un punto de vista cuantitativo. Por tanto la Demografía estudia estadísticamente la estructura y la dinámica de las poblaciones humanas y las leyes que rigen estos fenómenos.

Dinámica de la población se entiende por dinámica de la población, el cambio en las variables demográficas básicas (fecundidad, mortalidad y migración) y sus consecuencias sobre la magnitud y estructura de ella.

Distribución de nacimientos según edad de la madre se refiere a la distribución porcentual del total de nacimientos vivos en un año determinado, clasificados de acuerdo con la edad de la madre al tener los hijos.

Edad reproductiva (período de reproducción) en la mujer, es la edad en que es capaz de procrear y que, según se ha supuesto para fines estadísticos en la mayoría de los países, está comprendida entre los 15 y 49 años de edad.

Espaciamiento de los nacimientos este concepto hace referencia a la acción voluntaria de las parejas en relación con el retardo o el espaciamiento de los nacimientos de sus hijos. En demografía también se lo denomina intervalo intergenésico.

Estadística es una función calculada a partir de los datos observados. Los recuentos de nacimientos, defunciones y número de personas, así como las tasas y medidas de resumen de todo tipo pueden considerarse resultados estadísticos.

Estructura de la población es la composición por edad y sexo de la población que tiene importantes repercusiones sobre los fenómenos demográficos y el análisis demográfico.

Fecundidad de reemplazo es el nivel de la fecundidad en que cada mujer de una cohorte, procrea en promedio, una hija a lo largo de su vida reproductiva. Equivale aproximadamente a una tasa global de fecundidad de 2.1 hijos por mujer.

Fenómenos demográficos es para el estudio de la estructura y dinámica de la población, la Demografía se concentra en la investigación de los principales factores que influyen sobre ambos aspectos: La mortalidad, la Fecundidad y la Migración.

Fertilidad e infertilidad son los términos de fertilidad e infertilidad, se usan para designar respectivamente la capacidad y la incapacidad fisiológica de una mujer o un hombre para procrear.

Mujeres en edad fértil. En demografía comprende al grupo de mujeres de los 15 a 49 años de edad; generalmente se abrevia como MEF.

Mujeres unidas. En los estudios demográficos se hace referencia al estado civil de las mujeres en edad fértil, clasificándolas en diferentes grupos. Así, se habla frecuentemente de las mujeres unidas o actualmente unidas, para referirse a las casadas más las que viven en unión libre o consensual en un momento dado.

Población es un agregado cuyos miembros cambian con el tiempo, como resultado de su ingreso o egreso del mismo. Debe ser entendido como un conjunto de personas que se agrupan en cierto ámbito geográfico y está propenso a continuos cambios.

Tendencia es el comportamiento de la población o de un fenómeno demográfico en un tiempo determinado.

Bibliografía

- De la Lama García, Alfredo, (2005) *Estrategias para elaborar investigaciones científicas*, México. Trillas.
- Departamento de asuntos económicos y sociales, División de estadística, (2005) *Manual sobre la recolección de datos de fecundidad y mortalidad*, Naciones Unidas, Nueva York.
- Leguina, Joaquín, (1981) *Fundamentos de la demografía*, España. Siglo XXI.
- Mina Valdés, Alejandro, (2001) *Curso básico de demografía*, México publicaciones del departamento de matemáticas de la facultad de ciencias, Séptima edición.
- Mina Valdés, Alejandro, (2000) *Elaboración y utilidad de la tabla abreviada de mortalidad*, Cuarta Edición, México, publicaciones del departamento de matemáticas de la facultad de ciencias. Cuarta Edición.
- San Roman Muñoz, Roman, (2003) *Análisis Demográfico interactivo*, México. UNAM tesis de licenciatura de Actuaría en la Facultad de Ciencias.
- Welti, Carlos, *Demografía I México*, (1997) PROLAP-IISUNAM, pp. 17-158.
- Sistemas Nacionales Estadístico y de Información Geográfica, INEGI, (1991) Estadísticas demográficas *Cuaderno de población No.3*.
- The R Project por Statistical Computing, en el Departamento de estadística y matemáticas de la universidad de Viena Wirtschaftsuniversität Wien, en septiembre de 2006, < <http://www.r-project.org/index.html> >.
- The R Project por Statistical Computing, en el Departamento de estadística y matemáticas de la universidad de Viena Wirtschaftsuniversität Wien, en marzo de

2006, <<http://cran.r-project.org/>>.

- Distribuciones GNU/Linux en Pricoinsa, PRINUX, en octubre de 2006 <<http://linux.pricoinsa.es/>>

- Ubuntu-es, xuzo <http://www.ubuntu-es.org/> en Octubre de 2006

- Sistemas Nacionales Estadístico y de Información Geográfica, INEGI, en septiembre de 2006, <<http://www.inegi.gob.mx>>.

- Ubuntu, documentation team, <https://help.ubuntu.com/> en octubre de 2006

Otros sitios visitados:

- <http://www.gun.org>

- es.wikipedia.org/wiki/Ubuntu_Linux

- www.linux-es.org/Faq/Files/Html/FAQ_Linux_V2.0.2-145.html

- www.es.wikipedia.org/wiki/Linux

- www.red.es/glosario/glosario.html