



**UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES  
ACATLÁN**

**ESTIMACIÓN DE LOS NACIMIENTOS EN MÉXICO DE  
1990 A 2011; APLICACIÓN DE UN MODELO DE  
REGRESIÓN LINEAL MÚLTIPLE.**

**T E S I S I N A**

**QUE PARA OBTENER EL TÍTULO DE**

**A C T U A R I O**

**P R E S E N T A**

**VIANEY ROCÍO SEPÚLVEDA DEL CASTILLO**

**Asesor: Act. Aszael W. Picazo Sánchez**

Cd. Universitaria, D. F. Diciembre 2015



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**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.



*A mis padres...*

*A mis hermanos...*

*A mi familia...*

*A mis maestros, amigos y compañeros...*

*Por ser un pilar en el cual me apoyo. Por estar cerca de mí compartiendo las experiencias más importantes de mi carrera. Porque gracias a su apoyo, he realizado una de mis mejores metas. Ustedes que sin pedir nada, lo dieron todo. Porque nunca estuve sola. Porque siempre conté con su confianza. Por todo esto, quiero que sientan que el objetivo logrado, también es suyo y que la fuerza que me ayudó a conseguirlo, fue su amor. Para ustedes con cariño y admiración.*

*Gracias!*

## Resumen

---

**E**l Análisis de Regresión Lineal Múltiple nos permite establecer la relación que se produce entre una variable dependiente  $Y$  y un conjunto de variables independientes  $(x_1, x_2, \dots, x_k)$ . El análisis de regresión lineal múltiple, se aproxima a situaciones de análisis real ya que los fenómenos, hechos y procesos sociales, por definición, son complejos y, en consecuencia, deben ser explicados en la medida de lo posible por la serie de variables que, directa e indirectamente, participan en su concreción.

Al aplicar el análisis de regresión múltiple lo más frecuente es que tanto la variable dependiente como las independientes sean variables continuas medidas en escala de intervalo o razón. No obstante, caben otras posibilidades: (1) también podremos aplicar este análisis cuando relacionemos una variable dependiente continua con un conjunto de variables categóricas; (2) o bien, también aplicaremos el análisis de regresión lineal múltiple en el caso de que relacionemos una variable dependiente nominal con un conjunto de variables continuas.

En este trabajo se utiliza el modelo de regresión lineal múltiple para explicar los factores que afectan los nacimientos en México a partir de 1990 y hasta 2011, se utiliza éste periodo ya que es la información más actual que se encuentra publicada en la página de internet del Instituto Nacional de Estadística y Geografía (INEGI) y se cuenta con la información necesaria para poder aplicar dicho modelo.

Como se verá más adelante, el primer capítulo “La situación demográfica en México. Panorama desde las proyecciones de población” trata de sobre el crecimiento de la población para conocer mejor la dinámica demográfica. El segundo capítulo “Aspectos técnicos de un modelo de regresión lineal múltiple” trata del la información teórica necesaria para realizar el modelo, las características de las variables, las propiedades de los parámetros, el método que se utiliza para realizar la regresión, las pruebas de hipótesis para descartar variables y un resumen del paquete estadístico que se utilizó para realizar el modelo de regresión lineal. En el tercer capítulo “Aplicación del modelo de regresión lineal múltiple” se estima la correlación entre la variable dependiente y las independientes, se realizan las pruebas necesarias y, a partir de estos resultados se realiza la estimación del modelo la cual nos conduce a las conclusiones de dicho modelo que es para lo que sirve ésta rama de la estadística.

# Índice Temático

<b>INTRODUCCIÓN.....</b>	<b>1</b>
<b>CAPÍTULO 1. LA SITUACIÓN DEMOGRÁFICA EN MÉXICO. PANORAMA DESDE LAS PROYECCIONES DE POBLACIÓN.....</b>	<b>4</b>
1.1 DINÁMICA DEMOGRÁFICA.....	5
1.2 ESPERANZA DE VIDA.....	6
1.3 BONO DEMOGRÁFICO Y ENVEJECIMIENTO.....	7
1.4 MORTALIDAD.....	8
1.5 FECUNDIDAD.....	9
1.6 TRANSICIONES DE VIDA.....	16
1.7 ATENCIÓN MATERNA.....	17
1.8 CONCLUSIONES.....	18
<b>CAPÍTULO 2. ASPECTOS TÉCNICOS DE UN MODELO DE REGRESIÓN LINEAL MÚLTIPLE .....</b>	<b>20</b>
2.1 ESTIMACIÓN DE LOS PARÁMETROS POR EL MÉTODO DE MÍNIMOS CUADRADOS.....	22
2.2 PROPIEDADES DE LOS ESTIMADORES.....	26
2.2.1 TEOREMA DE GAUSS-MARKOV.....	27
2.3 ESTIMACIÓN DE $\sigma^2$ .....	28
2.4 PRUEBAS DE HIPÓTESIS EN LA REGRESIÓN LINEAL MÚLTIPLE.....	28
2.5 COEFICIENTE DE DETERMINACIÓN $R^2$ Y $R^2$ AJUSTADA.....	30
2.6 CRITERIO DE AKAIKE Y CRITERIO DE SCHWARTZ.....	31
2.7 DETECCIÓN DE LA PRESENCIA DE AUTOCORRELACIÓN.....	31
2.8 INTERVALOS DE CONFIANZA.....	33
2.9 MULTICOLINEALIDAD.....	33
2.10 ESTADÍSTICOS BÁSICOS PARA EL ANÁLISIS DE SERIES.....	34
2.11 EIEWS. PAQUETE ESTADÍSTICO.....	36
<b>CAPÍTULO 3. APLICACIÓN DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE .....</b>	<b>37</b>
3.1 APLICACIÓN DEL MODELO.....	37
TABLA 1. PRUEBAS DE RAÍZ UNITARIA PARA CADA VARIABLE NO ESTACIONARIA Y ESTACIONARIA ..	41
3.2 ESTIMACIÓN DEL MODELO.....	44
3.2.1 NORMALIDAD.....	46
3.2.2 AUTOCORRELACIÓN.....	47
3.2.3 HOMOCEDASTICIDAD.....	48
3.2.4 FORMA FUNCIONAL LINEAL.....	49
3.2.5 ESTABILIDAD DE PARÁMETROS.....	50
<b>CONCLUSIONES.....</b>	<b>51</b>

**REFERENCIAS ..... 52**

FUENTES BIBLIOGRÁFICAS

FUENTES ELECTRÓNICAS

**ANEXOS ..... 53**

A. DATOS ..... 53

B. PRUEBAS DE RAÍZ UNITARIA ..... 54

C. REGRESIONES ..... 59

D. PRUEBA DE NORMALIDAD ..... 63

## Introducción

---

**E**n su origen, la palabra Estadística procede del vocablo “Estado” (la ciencia del Estado) pues era la función principal de los gobiernos establecer registros de población, nacimientos, defunciones, impuestos,... La necesidad de tener datos cifrados sobre la población y sus condiciones materiales de existencia continúa ampliamente a través de los servicios de estadísticas nacionales e internacionales. En particular, los censos comenzaron a suministrar información regular acerca de la población de cada país. Así, los datos estadísticos se referían originalmente a los datos demográficos de una ciudad o estado determinados.

El uso de la estadística se ha extendido más allá de sus orígenes como un servicio al estado. Personas y organizaciones la utilizan para entender datos y tomar decisiones en ciencias naturales, sociales, medicina, negocios y otras áreas.

La estadística asume un papel importante en el campo de estudios y laboral de un egresado de la carrera de actuaría ya que es posible aplicar a distintas áreas del desarrollo del profesionista. La estadística trata del recuento, ordenación, clasificación y análisis de una muestra representativa de datos que permiten explicar correlaciones y dependencias de un fenómeno de ocurrencia ya sea aleatorio o condicional. Es una herramienta transversal a otras disciplinas desde la física, química, economía hasta las ciencias sociales, de la salud e incluso el control de calidad.

La estadística proporciona instrumentos para la toma de decisiones cuando prevalecen condiciones de incertidumbre. Estos instrumentos pueden ser de aplicación y utilidad completamente general en cualquier campo de la ciencia. Son aplicables no solo en el mundo científico, sino también en el de las empresas y de los asuntos cotidianos. El proceso de tomar decisiones en situaciones generales, sobre la base de una información incompleta contenida en datos muestrales, es arriesgado y no puede realizarse con certeza; la probabilidad es una medida de esta incertidumbre. Hay dos incertidumbres con las que nos enfrentamos: 1) La incertidumbre debida a la aleatoriedad, y 2) La incertidumbre debida a nuestra ignorancia del verdadero estado del sistema. La teoría de la estadística (en sentido amplio) puede considerarse que incluye a la teoría de la probabilidad. Incluye también otras cuestiones que no forman parte de la teoría de la probabilidad, propiamente dicha, como las consecuencias del principio de aleatorización, diversos principios de estimación y en general, un principio de toma de decisiones.

El actuario, al encontrarse con un problema experimental determinado, construye un modelo matemático que se ajuste lo mejor posible a la situación experimental; analiza el modelo por métodos matemáticos, y finalmente establece procedimientos para el estudio del problema.

La estadística se divide en dos partes: la estadística descriptiva que registra los datos en tablas, los representa en gráficos y calcula parámetros (medidas de centralización y de dispersión) que



describen el conjunto estudiado y la estadística inferencial que estudia cómo sacar conclusiones generales para toda la población a partir de una muestra y el grado de confiabilidad o significación de los resultados obtenidos.

El empleo del instrumental estadístico no es una simple cuestión de escoger el que mejor se adapte, hay que hacer toda una serie de juicios relativos a la naturaleza y magnitud de los diversos errores que de ellos se obtengan por la discrepancia entre el modelo y el experimento efectivo.

Una técnica estadística utilizada frecuentemente es el análisis de regresión lineal predice los valores de una variable cuantitativa a partir de otra variable explicativa también cuantitativa. En el caso de la regresión lineal múltiple, se aproxima más a situaciones de análisis real puesto que los fenómenos, hechos y procesos sociales, son complejos y en consecuencia deben ser explicados en la medida de lo posible por una serie de variables que directa o indirectamente participan en su correlación.

El término regresión fue utilizado por primera vez en un estudio realizado por Francis Galton sobre variables antropométricas en 1889, al comparar la estatura de padres e hijos resultó que los hijos cuyos padres tenían una estatura superior al valor medio tendían a igualarse a éste; mientras que aquellos cuyos padres eran muy bajos tendían a reducir su diferencia respecto a la estatura media; es decir, “regresaban” al promedio. El término lineal es utilizado para distinguir de las demás técnicas de regresión que emplean modelos basados en cualquier función matemática como por ejemplo cuadráticas, cúbicas, exponenciales, etc.

En el análisis de regresión múltiple, los estadísticos, pruebas y análisis que se aplican para determinar la relación y grado de asociación entre una variable dependiente y sus supuestas variables explicativas, así como la estimación de los parámetros de la ecuación no difieren de los determinados en el análisis de regresión simple. En el análisis de regresión múltiple es necesario calcularlos a medida que vamos introduciendo y/o sacando variables independientes en el modelo.

En dicho análisis la construcción de su correspondiente ecuación se realiza seleccionando las variables una a una. La finalidad perseguida es buscar de entre todas las posibles variables explicativas aquellas que más y mejor expliquen a la variable dependiente sin que ninguna de ellas sea combinación lineal de las restantes. Este procedimiento implica que: (1) en cada paso solo se introduce aquella variable que cumple unos criterios de entrada; (2) una vez introducida, en cada paso se valora si alguna de las variables cumplen criterios de salida; y (3), en cada paso se valora la bondad de ajuste de los datos al modelo de regresión lineal y se calculan los parámetros del modelo verificado en dicho paso. El proceso se inicia sin ninguna variable independiente en la ecuación de regresión y el proceso concluye cuando no queda ninguna variable fuera de la ecuación que satisfaga el criterio de selección (garantiza que las variables seleccionadas son significativas) y/o el criterio de eliminación (garantizar que una variable seleccionada no es redundante).

Uno de los métodos más empleados para definir el modelo (ecuación lineal) de regresión es el método de mínimos cuadrados, el cual fue propuesto por Legendre en 1805 y Gauss en 1809. El término "mínimos cuadrados" proviene de la descripción dada por Legendre "moindres carrés".

Es importante mencionar que la estadística aplicada se ha desarrollado hasta tal punto que puede considerarse que constituye ya un campo de interés especial.

## Capítulo 1. La situación demográfica en México. Panorama desde las proyecciones de población<sup>1</sup>.

---

**L**as proyecciones de población son un instrumento indispensable para la formulación e instrumentación de acciones y programas de la administración pública a escala nacional, estatal y municipal; éstas permiten anticipar necesidades sociales y situaciones que podrían convertirse en problemas, así como canalizar recursos de diverso tipo para su atención.

La política de población tiene como uno de sus soportes técnicos los insumos que le ofrece la prospectiva demográfica, con éstos es posible la formulación de estrategias y líneas de acción sobre mortalidad, fecundidad y salud reproductiva, migración interna e internacional, envejecimiento, bono demográfico, entre otros.

En México se cuenta hoy con un amplio acervo de resultados de diversos ejercicios de proyecciones de población elaborados tanto por organismos internacionales, como por científicos e instituciones estatales abocadas a esta tarea técnica. Desde los escenarios futuros publicados por la Organización de las Naciones Unidas en 1954, hasta las proyecciones de finales de los años noventa del siglo XX y principios de este nuevo siglo, tuvieron entre sus objetivos monitorear el avance en las metas planteadas sobre el crecimiento de la población y conocer mejor la dinámica demográfica. En síntesis, las proyecciones desde el último tercio del siglo XX:

- Alimentan la planeación económica, social y demográfica;
- Establecen metas programáticas;
- Permiten analizar los efectos e interrelaciones de los fenómenos demográficos, conciliando el pasado demográfico reciente;
- Hacen posible explorar posibles escenarios futuros; y
- Permiten monitorear y comparar el avance del país con su propio pasado demográfico y con las tendencias a escala internacional.

---

<sup>1</sup> La información contenida en este capítulo fue obtenida del documento “La situación demográfica de México, 2013” emitido por el Consejo Nacional de Población (CONAPO) en el año 2013.

## 1.1 Dinámica demográfica

*La dinámica demográfica se compone de tres factores: la natalidad, la mortalidad y la migración, tanto interna como internacional. Los tres interactúan para modificar tanto el volumen –cuántos somos- como la estructura –la distribución por edad y sexo- de la población, en estrecha relación con una serie de características como la condición de actividad, la escolaridad, la etnicidad, la distribución en el territorio, la situación conyugal, la religión y otras más.*

Analizar la transición demográfica en México es plantear los desiguales ritmos y evolución de las variables del cambio demográfico en el país. Aún cuando las tendencias generales confirman el descenso en la fecundidad, la mortalidad general y la infantil, así como el aumento en la esperanza de vida de la población en general, el rezago en el bienestar y condiciones de vida de determinados sectores de la población son factores para que dicha transición transcurra de forma incluso más lenta, convirtiéndose, por tanto, en un reto demográfico la democratización de los beneficios de dicha transición.

El Consejo Nacional de Población (CONAPO) estimó que a mediados del año 2013 la población alcanzó 118.4 millones. Durante este mismo año habrá 2.25 millones de nacimientos y alrededor de 673 mil defunciones, lo que implicará en términos absolutos un crecimiento de 1.58 millones de personas, con una tasa de crecimiento anual de 1.13 por ciento. Por su parte, el saldo neto migratorio internacional, que comparado con décadas pasadas es de una magnitud considerablemente menor, continuará en valores negativos, con un número promedio anual de 228 mil personas por año, entre 2010-2015.

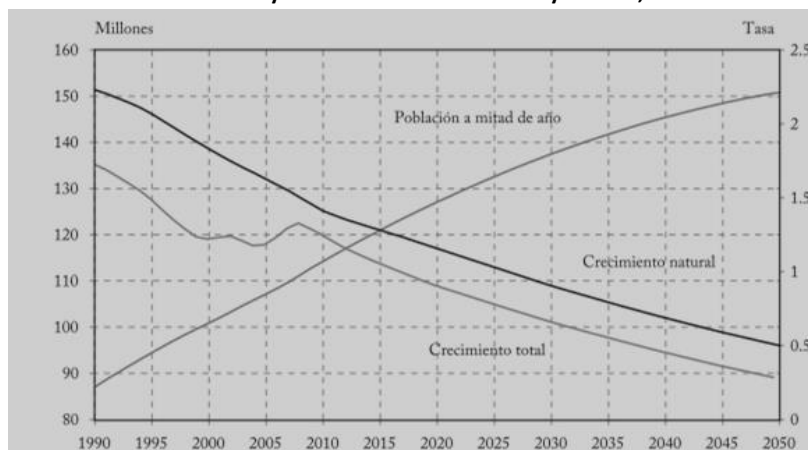
La inercia actual de cambio demográfico se origina en el crecimiento acelerado que tuvo la población en el siglo pasado. Si bien las tasas de crecimiento han tendido a reducirse, su nivel aún se mantiene. Sin embargo, el cambio en las tendencias de la migración internacional durante la primera década del siglo XXI dio un considerable impulso a la dinámica demográfica, dando lugar a un evento en nuestra historia que sólo encuentra un punto de comparación posiblemente con lo ocurrido después de la gran depresión de 1929 y los años subsiguientes.<sup>2</sup>

En el umbral del nuevo siglo, el ritmo de crecimiento de la población era de 1.22 por ciento, y en 2008 éste se incrementó hasta llegar a 1.33 por ciento. Actualmente, la tasa de crecimiento media anual proyectada se sitúa en 1.25 en México.

---

<sup>2</sup> La información estadística de esa época sobre las personas deportadas de los Estados Unidos de América entre 1926-1932, 1933 y 1934 indica que del número total, entre un tercio y la mitad tuvieron como destino México o fueron clasificados de raza mexicana. Véanse los cuadros 97 y 104 de los Statistical Abstract of the United States de 1933 y 1935 del Department of Commerce, Washington D.C.

**Grafica 1. Población y tasas de crecimiento total y natural, 1990-2050.**



Fuente: Estimaciones del CONAPO con base en la conciliación demográfica 1990-2010 y proyecciones de población 2010-2050.

La fecundidad continúa en franco descenso y en el último año del primer cuarto de este siglo estará cada vez más cerca del nivel del reemplazo (que es de 2.1 hijos por mujer). En la última década del siglo pasado ocurrió una disminución considerable en el número promedio de hijos por mujer al pasar de 3.36 a 2.65. Al presente, la tasa global de fecundidad en México es de 2.22 hijos por mujer.

## 1.2 Esperanza de vida

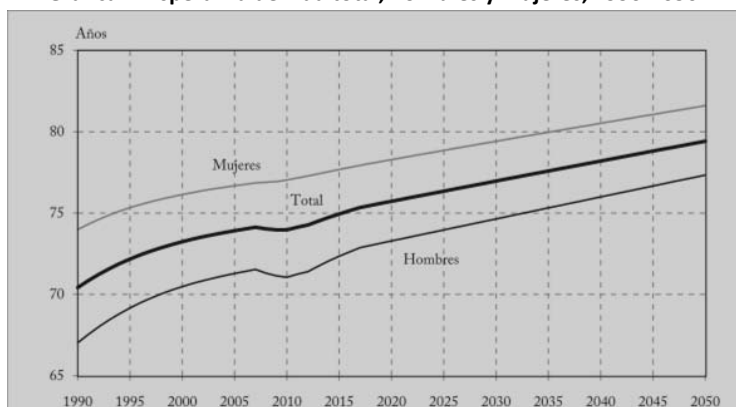
*La esperanza de vida al nacer indica la cantidad de años que viviría un recién nacido si los patrones de mortalidad vigentes al momento de su nacimiento no cambian a lo largo de la vida del infante.*

Durante la década de 1990 al año 2000, la esperanza de vida al nacimiento para el total de la población pasó de 70.4 a 73.2; al año 2007, ya situados en la primera década del siglo XXI, la esperanza de vida total alcanzó 74.1, lo que constituye un avance de casi cuatro años más de vida para la población en su conjunto en un lapso de 15 años, lo cual refleja la contribución de los avances logrados durante el siglo XX en el desarrollo económico y social del país. Por circunstancias coyunturales, en los años posteriores a 2007 se registró un descenso ligero en la esperanza de vida, como resultado de una sobremortalidad en algunos grupos de población. La esperanza de vida actualmente se sitúa en un valor de 74.5 años (Véase gráfica 2).

El diferencial por sexo de la esperanza de vida implica ya un número mayor de mujeres en edades avanzadas respecto al volumen de hombres, situación que contribuye al reto de constituírnos en una población con una estructura y composición demográfica heterogéneas y con tendencia hacia el envejecimiento. La tasa de crecimiento media anual de la población de 60 años o más es

actualmente de 3.4 por ciento, la mayor comparada con el grupo de 0 a 14 años o con el de 15 a 59 años de edad.

**Grafica 2. Esperanza de vida total, hombres y mujeres, 1990-2050.**



Fuente: Estimaciones del CONAPO con base en la conciliación demográfica 1990-2010 y proyecciones de población 2010-2050.

### 1.3 Bono demográfico y envejecimiento

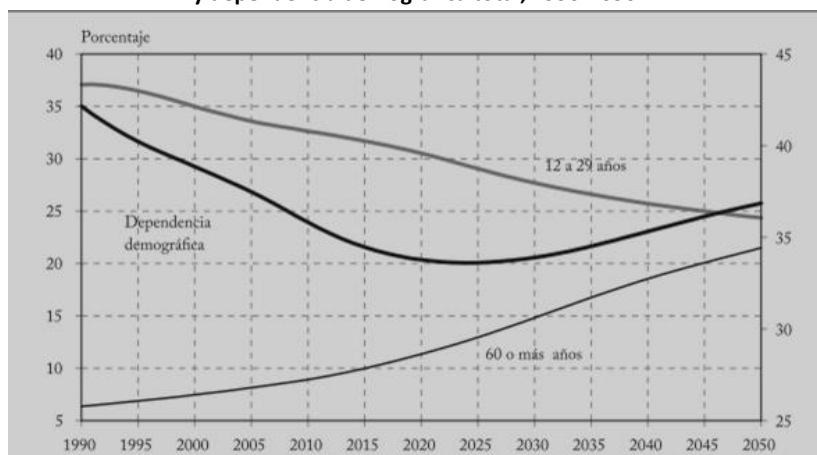
*El Consejo Nacional de Población (CONAPO) define al bono demográfico como el fenómeno que se da dentro del proceso de transición demográfica en el que la población en edad de trabajar (económicamente activa) es mayor que la dependiente (niños y adultos mayores), y por tanto, el potencial productivo de la economía es superior.*

Usualmente, se acude a indicadores de las relaciones de dependencia infantil y de edades avanzadas como descriptores de la situación global que guarda una sociedad, un país, respecto al potencial que representa este bono, en tanto que una ventana de oportunidad para aprovechar el potencial para el desarrollo de esta estructura centrada en edades productivas. Los cambios en la razón de dependencia demográfica para las edades avanzadas, que se derivan de las proyecciones de población, pueden indicarse como sigue: De 1990 a 2013 la razón de dependencia demográfica para la población de 65 años o más respecto a la población total pasó de 4.3 a 6.5 adultos en dichas edades por cada 100 habitantes. En las próximas décadas el CONAPO prevé que esta razón siga aumentando.

En 2013, en México, la población de 15 a 64 años de edad es de 77.04 millones de personas, 65.1 por ciento de la población total del país. Para 2020 el CONAPO proyecta que la población en esas edades será de 84.17 millones, representando dos terceras partes (66.2%) de la población total, de la cual más de la mitad estará constituida por mujeres (43.67 millones).

En la gráfica 3 se muestra información de proyecciones sobre la población joven (12 a 29 años de edad), la cual representa cerca de un tercio de la población total del país (32.1%), mientras que la población de 60 años o más actualmente constituye sólo el 9.5 por ciento. Valores que implican que por cada 3.4 jóvenes hay una persona adulta mayor. Hacia el año 2030 esta última relación será de dos a uno, y se prevé que hacia el 2050 la brecha en la estructura demográfica se cierre, en consecuencia, los porcentajes de ambos grupos serán similares y su relación será prácticamente de uno a uno (1.1).

**Gráfica 3. Proporción de población de 12 a 29 años y 60 años o más y dependencia demográfica total, 1990-2050.**



Fuente: Estimaciones del CONAPO con base en la conciliación demográfica 1990-2010 y proyecciones de población 2010-2050.

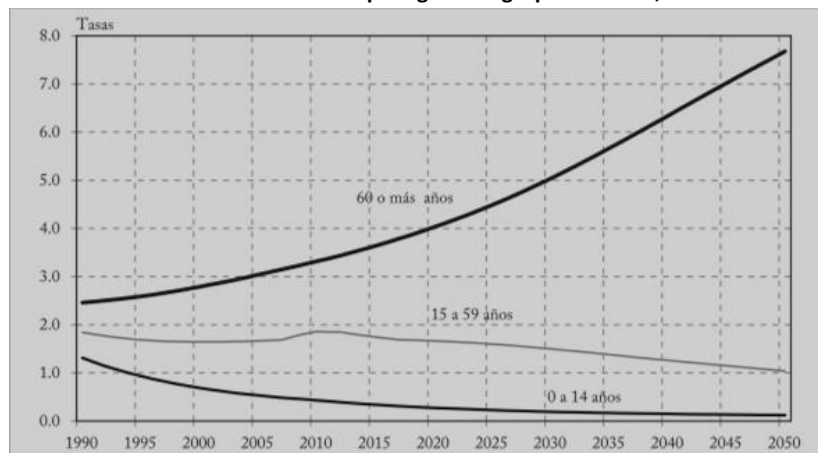
#### 1.4 Mortalidad

*Esta variable proporciona el número medio anual de muertes durante un año por cada 1000 habitantes, también conocida como tasa bruta de mortalidad. La tasa de mortalidad, a pesar de ser sólo un indicador aproximado de la situación de mortalidad en un país, indica con precisión el impacto actual de mortalidad en el crecimiento de la población. Este indicador es significativamente afectado por la distribución por edades. La mayoría de los países eventualmente mostrarán un aumento en la tasa de mortalidad general, a pesar del continuo descenso de la mortalidad en todas las edades, a medida que una disminución en la tasa de fecundidad resulta en un envejecimiento de la población.*

Por otro lado, el siglo XXI comienza con un aumento en la mortalidad general de la población al compararlo con la situación demográfica del pasado siglo, sin embargo, destaca que ésta se encuentra mucho más relacionada con las causas de muerte de una población que progresivamente envejece.

La estructura y composición demográfica de la mortalidad que se observa en la gráfica 4 corrobora que conforme la estructura demográfica del país envejece, la mortalidad se sobrepone en edades avanzadas. No obstante, una mirada de mayor rigor permite dar cuenta de una situación inusual que se presentó ya avanzado el segundo quinquenio de este nuevo siglo: se trata de los fallecimientos de hombres en edades predominantemente laborales, anclados a la situación de violencia vivida en el país en esos años. La información aritmética de esta situación muestra entonces un viraje importante en la dinámica de este componente demográfico, que se articula a procesos mucho más complejos vinculados a los distintos ámbitos de la realidad social.

**Gráfica 4. Tasas de mortalidad para grandes grupos de edad, 1990-2050.**



Fuente: Estimaciones del CONAPO con base en la conciliación demográfica 1990-2010 y proyecciones de población 2010-2050.

### 1.5 Fecundidad

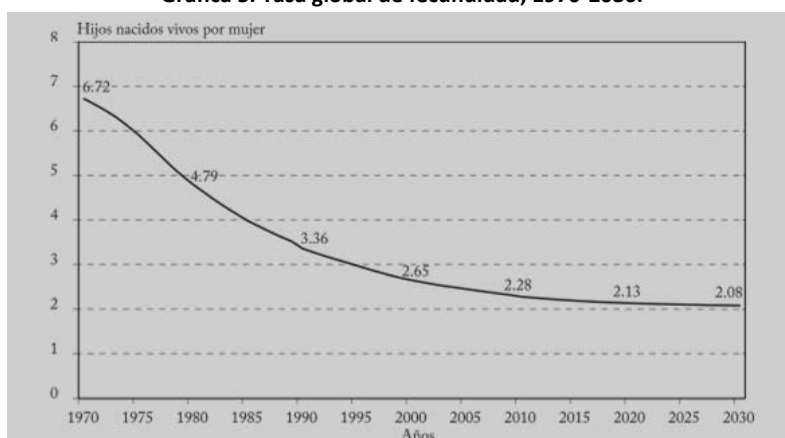
*Esta variable da el número promedio de hijos que nacerían por mujer si todas las mujeres vivieran hasta el final de sus años fértiles y dieran a luz de acuerdo a la tasa de fecundidad promedio para cada edad. La tasa total de fecundidad es una medida más directa del nivel de fecundidad que la tasa bruta de natalidad, ya que se refiere a los nacimientos por mujer. Este indicador muestra el potencial de los cambios demográficos en el país. Un promedio mayor a dos hijos por mujer se considera la tasa de sustitución para una población, dando lugar a una relativa estabilidad en términos de cifras totales. Promedios por encima de dos hijos por mujer indican poblaciones en aumento y cuya edad media está disminuyendo. Tasas más elevadas también puede indicar dificultades para las familias, en algunas situaciones, para alimentar y educar a sus hijos y para las mujeres que desean entrar a la fuerza de trabajo. Promedios por debajo de dos hijos por mujer indican una disminución del tamaño de la población y una edad media cada vez más elevada.*



A finales de los años setenta comienza a disminuir la fecundidad en México, debido principalmente a la instrumentación de acciones en materia de planificación familiar. En los primeros veinte años (entre 1970 y 1990) se apreció una acelerada reducción de la Tasa Global de Fecundidad (TGF)<sup>3</sup>, por ejemplo, de 1970 a 1980 descendió en casi dos hijos y de 1980 a 1990, en poco más de un hijo (Véase gráfica 5). Así, en 1990, el número de hijos que tuvieron las mujeres fue la mitad de los que tenían en 1970; asimismo, el descenso fue continuo pero a un ritmo menos acelerado, es decir, tuvieron que pasar veinte años (1990 a 2010) para reducir la TGF en alrededor de un hijo y se espera que en las próximas dos décadas (2010 a 2030) se logre estar alrededor del nivel de reemplazo generacional (2.1 hijos por mujer).

Como se ha mencionado, la reducción de la fecundidad ha sido favorecida por la implementación de programas gubernamentales de planificación familiar que promovieron el uso de métodos anticonceptivos; en un principio, sólo las mujeres residentes de zonas urbanas tuvieron acceso a estos beneficios, pero con el paso del tiempo se ha buscado que la cobertura se extienda al resto de las mujeres del país.

**Gráfica 5. Tasa global de fecundidad, 1970-2030.**



Fuente: Estimaciones del CONAPO, retrospectiva con base en el INEGI, XII Censo General de Población y Vivienda 2000; la Conciliación demográfica 1990-2010 y proyecciones de población 2010-2050.

El uso de métodos anticonceptivos entre mujeres unidas en edad fértil a nivel nacional se ha incrementado, en 1976 fue de 30.2%, en 1987, de 52.7, en 1997, de 68.5, y en 2009, de 72.5%. Estos datos muestran que la mayor cobertura del uso de métodos anticonceptivos sucedió en las dos primeras décadas<sup>4</sup>, y dicho indicador todavía continúa en ascenso, lo que reafirma la influencia que han tenido las políticas públicas en la disminución de la fecundidad (Moreno y Singh, 1996). Sin embargo, al interior del país la reducción de la fecundidad no ha sido homogénea, pues en 2010 había entidades federativas con alta TGF, similar a la estimada a nivel nacional en el año

<sup>3</sup> La Tasa Global de Fecundidad (TGF) es el número de hijos que nacerían por mujer si la mujer o mujeres tuvieran sus hijos en todos sus años reproductivos de acuerdo con las tasas de fecundidad específicas por edad de la población y épocas del estudio.

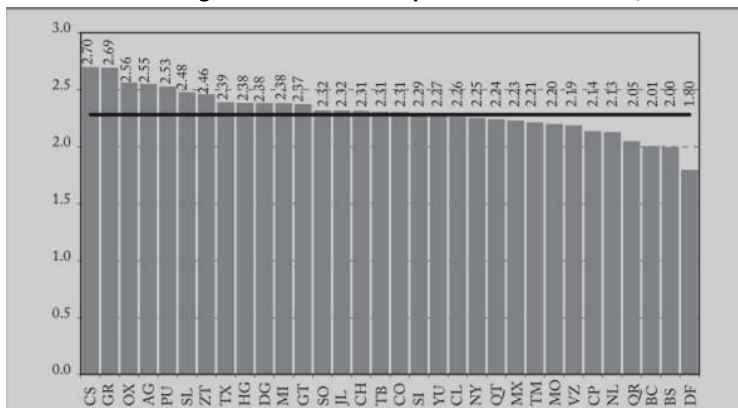
<sup>4</sup> A partir del año de implementación de la Ley General de Población en 1974 la consideramos como década inicial.

2000, es decir, las entidades presentan una fecundidad que se tenía diez años atrás y, por otro lado, se observó que hubo entidades que en 2010 ya habían alcanzado una fecundidad incluso por debajo del reemplazo generacional, situación que el país en su conjunto experimentará hasta el 2030.

Las entidades que en 2010 tuvieron la TGF más alta fueron Chiapas, Guerrero y Oaxaca (Véase gráfica 6), que por lo general se caracterizan por tener grados de marginación muy alto o alto, tienen un mayor porcentaje de población rural, que habla lengua indígena o que vive en localidades de difícil acceso; además, su prevalencia anticonceptiva fue baja. En contraste, el Distrito Federal y estados como Baja California Sur y Baja California registraron las tasas de fecundidad más bajas, y cuentan con todos los servicios necesarios para que la población tenga información y acceso a la planificación familiar.

Al analizar la fecundidad de acuerdo al calendario de las mujeres<sup>5</sup>, se observó que, entre 1990 y 2010, las de 20 años de edad en adelante redujeron su descendencia, sobre todo, las mujeres que se encontraban en los últimos grupos de edad reproductiva (Véase gráfica 7).

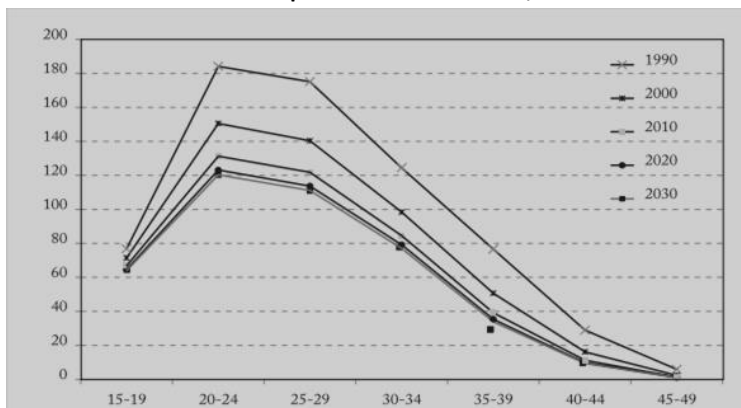
**Gráfica 6. Tasa global de fecundidad por entidad federativa, 2010.**



Fuente: Estimaciones del CONAPO con base en la Conciliación demográfica 1990-2010.

<sup>5</sup> Por calendario se entiende la distribución por edad del fenómeno durante el ciclo de vida, distribución que podrá ser más o menos precoz, más o menos tardía.

**Gráfica 7. Tasas Específicas de Fecundidad, 1990-2030.**



Fuente: Estimaciones del CONAPO con base en la Conciliación demográfica 1990-2010.

De 1990 a 2010, el patrón de fecundidad temprano se conserva, es decir, las mujeres entre 20 y 24 años fueron las que más contribuyeron a la fecundidad al presentar el mayor número de nacidos vivos.

De acuerdo con lo anterior, en 2030 se espera que el grupo de mujeres de 20 años en adelante reduzca su fecundidad en alrededor de 13 por ciento en promedio, que el mayor número de nacimientos ocurra en mujeres entre 20 y 24 años, además de que su número de nacimientos sea más cercano al de las mujeres de entre 25 y 29 años, por lo que el patrón de fecundidad tenderá a ser dilatado.

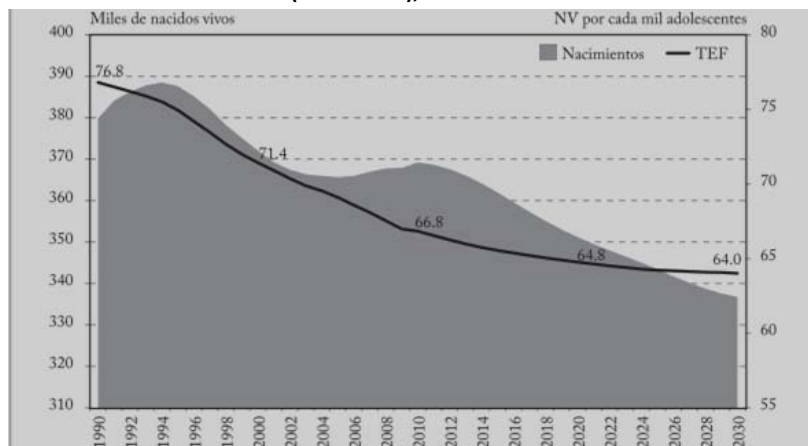
Mientras la fecundidad disminuye en las mujeres adultas, en las adolescentes (15 a 19 años) los cambios son más ligeros, puesto que entre 1990 y 2000 la reducción fue de 7.0 por ciento, de 2000 a 2010, de 6.4 por ciento, y se espera que de 2010 a 2030 el descenso sea de 4.2 por ciento. La contribución de las adolescentes a la fecundidad entre 1990 y 2010 ha sido constante, alrededor de 16 por ciento de los nacimientos ha sido de mujeres en este grupo de edad, y se prevé que continuará con esa tendencia hasta el año 2030 (Véase gráfica 8).

La importancia de abatir la fecundidad adolescente radica en que la adolescencia constituye una etapa de vida crucial para desarrollar el capital humano de la persona. La postergación de la edad al matrimonio, además del inicio de la paternidad o maternidad, permite a nivel individual aumentar la posibilidad de contar con una mayor preparación y maduración, lo que brindará a este grupo de población ventajas en etapas posteriores de su ciclo de vida.

Los riesgos de un embarazo en la adolescencia están fuertemente asociados con la desigualdad, pobreza e inequidad de género. Asimismo, las probabilidades de que las adolescentes mueran

debido a complicaciones durante la gestación o el parto son dos veces mayores a las de una mujer de entre 20 y 30 años<sup>6</sup>.

**Gráfica 8. Número de nacimientos y Tasa de fecundidad de las adolescentes (15-19 años), 1990-2030.**



Fuente: Estimaciones del CONAPO con base en la conciliación demográfica 1990-2010 y Proyecciones de población 2010-2050.

En 2009 se estimó que 40.6 por ciento de las adolescentes no había planeado o deseado el embarazo<sup>7</sup>, a pesar de que 97.0 por ciento de este grupo de mujeres dijo conocer al menos un método anticonceptivo; 61.5 por ciento de ellas no se protegió en su primera relación sexual, condición experimentada por la mitad de las adolescentes a los 15.4 años o antes.

Por otra parte, la dificultad que tienen las mujeres adolescentes unidas para ejercer sus derechos reproductivos se hace evidente al presentar un bajo uso de métodos anticonceptivos (44.7 por ciento), con una necesidad insatisfecha de métodos anticonceptivos de 24.6 por ciento.

La fecundidad adolescente muestra un comportamiento diferenciado entre las entidades federativas. Según se aprecia en la gráfica 9, las mayores tasas no necesariamente se encuentran en las entidades con las mayores TGF<sup>8</sup>. Los niveles más altos se distinguen en estados de la frontera norte (Coahuila, Chihuahua y Sonora), contrarios a las entidades del centro donde se registra la menor fecundidad en este grupo de mujeres (Distrito Federal<sup>9</sup>, Guanajuato, Morelos).

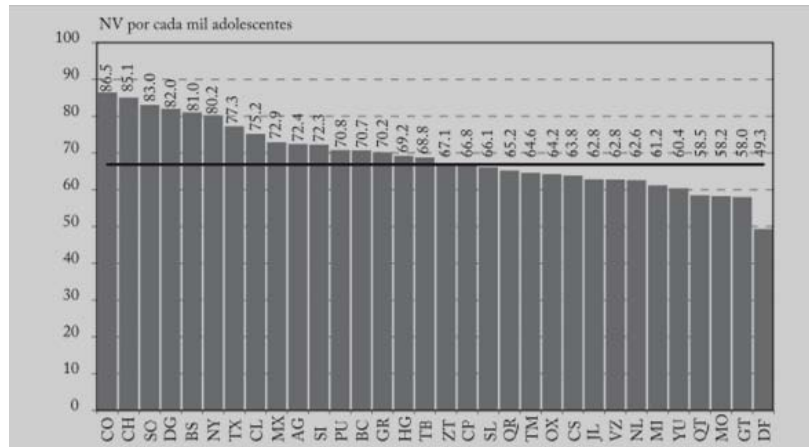
<sup>6</sup> UNFPA México “Salud Sexual y Reproductiva en Adolescentes y Jóvenes”. [http://www.unfpa.org.mx/ssr\\_adolescentes.php](http://www.unfpa.org.mx/ssr_adolescentes.php)

<sup>7</sup> Datos estimados con base en la Encuesta Nacional de la Dinámica Demográfica, 2009.

<sup>8</sup> Las mayores TGF se localizan en estados con grandes rezagos socioeconómicos y mayores retrasos en términos de la transición demográfica.

<sup>9</sup> Es necesario señalar que el 24 de abril de 2007 la Asamblea Legislativa del Distrito Federal aprobó en la Ciudad de México una reforma para despenalizar el aborto realizado en las primeras 12 semanas de gestación, por lo que éste puede ser un factor que contribuya al reporte de una baja tasa de fecundidad en las adolescentes.

**Gráfica 9. Tasa de fecundidad de las adolescentes (15-19 años) por entidad federativa, 2010.**



Fuente: Estimaciones del CONAPO con base en la conciliación demográfica 1990-2010.

La disminución de la fecundidad ha sido el principal determinante en la reducción del crecimiento demográfico registrado en el país durante los últimos cuarenta años. Pero el descenso de la fecundidad no ha ocurrido de manera simultánea, ni con la misma intensidad entre los diferentes ámbitos poblacionales o en los distintos grupos étnicos del país, pues aún continúan presentando altos niveles de fecundidad.

Entre 1999 y 2009, la Tasa Global de Fecundidad<sup>10</sup> (TGF) de las mujeres indígenas disminuyó en cerca de un hijo, al pasar de 4.14 a 3.25 hijos por mujer<sup>11</sup>. Sin embargo, continúa por arriba de la TGF estimada en mujeres no indígenas, que fue de 2.73 y 2.27 hijos, respectivamente. En tanto, la TGF estimada a nivel nacional fue de 2.85 en 1999 y disminuyó a 2.32 en 2009. Dichos datos confirman que la fecundidad de hablantes de lengua indígena sigue siendo mayor con respecto a la estimada a nivel nacional, contrario a lo que sucede con la fecundidad de no hablantes de lengua indígena, la cual es menor.

De acuerdo a la estructura por edad, se observó que las mujeres indígenas mantienen un patrón de fecundidad temprano, es decir, la cúspide se encuentra en el grupo de jóvenes (20 a 24 años). Por su parte, las no indígenas, desde el año 2000, tienden a tener un patrón dilatado, la cúspide se encuentra compartida entre el grupo de jóvenes y el grupo de mujeres entre 25 y 29 años.

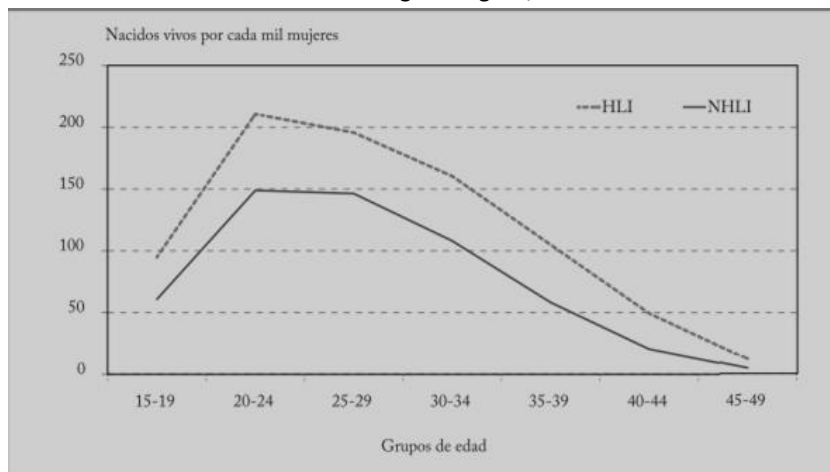
Tanto en la población indígena como en la no indígena descendieron las tasas específicas de fecundidad (TEF), aunque en mayor proporción en los indígenas. Por ejemplo, en las adolescentes hablantes de lengua indígena la tasa bajó 27.8%, al pasar de 94.8 a 68.4 nacidos vivos por cada mil,

<sup>10</sup> La TGF mide el número de hijos que tendría una mujer al final de su vida reproductiva en caso de seguir el patrón que registra la estructura de la fecundidad en el año de referencia.

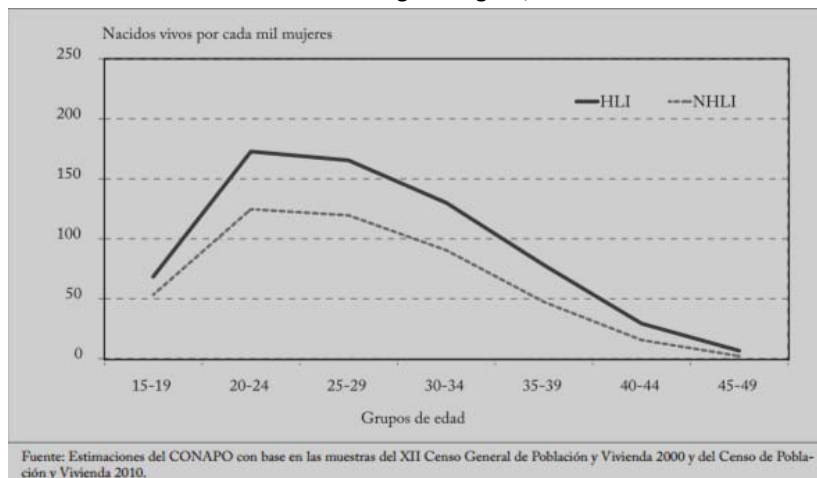
<sup>11</sup> La estimación se realizó a partir del último hijo nacido vivo ocurrido un año previo al levantamiento del censo.

mientras que en las no hablantes de lengua indígena la reducción fue de 12.2%, debido a que la tasa pasó de 60.7 a 53.3 (véanse gráficas a y b). Cabe señalar que la brecha entre la fecundidad de adolescentes indígenas y no indígenas se redujo a la mitad (de 34.1 a 15.1 hijos por cada mil).

**Gráfica A. Tasa específica de fecundidad de mujeres en edad fértil por condición de habla de lengua indígena, 2000.**



**Gráfica B. Tasa específica de fecundidad de mujeres en edad fértil por condición de habla de lengua indígena, 2010.**



Fuente: Estimaciones del CONAPO con base en las muestras del XII Censo General de Población y Vivienda 2000 y del Censo de Población y Vivienda 2010.

Por otra parte, las preferencias e ideales reproductivos influyen en la determinación de la fecundidad y descendencia final de las parejas; asimismo, dan cuenta de las necesidades de la población en materia de planificación familiar y salud reproductiva (CONAPO, 2011). En ese sentido, es importante conocer el número ideal promedio de hijos, sobre todo de las mujeres en edad fértil unidas, quienes tienen mayor probabilidad de embarazarse. A nivel nacional, el número

ideal de hijos promedio pasó de 3.2 en 1997 a 3.0 en 2009, en tanto que en las mujeres indígenas se mantuvo en 3.9; en las no indígenas tampoco se han mostrado cambios, al estimarse en 3.0 hijos.

En cuanto al número promedio ideal de hijos por nivel de escolaridad, éste cambia a 4.6 hijos como ideal en mujeres indígenas que no cuentan con algún nivel de escolaridad; en las no indígenas con la misma característica, a 4.1 hijos; y en ambos grupos disminuye si cuentan con secundaria o más, de manera que en las indígenas el ideal baja a 3.1 hijos y en las no indígenas, a 2.7 hijos.

### *1.6 Transiciones de vida*

La población pasa por diferentes transiciones a lo largo de la vida (casarse por primera vez, tener la primera relación sexual, el nacimiento del primer hijo, entre otras) que pueden ser complejas, ya que dependen de un conjunto de factores estructurales e institucionales que se caracterizan por situaciones históricas particulares, sin embargo, la mayoría de las personas transita por las mismas fases (Loggia 2006).

Las edades medianas a la que ocurren los eventos que constituyen transiciones en la vida de las jóvenes se muestran en el cuadro 1, donde se distingue que no siguen el orden más conservador (unión<sup>12</sup>-sexualidad-reproducción). Por ejemplo, la primera relación sexual sucede antes que la primera unión. En 2009, el inicio de la vida sexual de las indígenas ocurrió a una edad mediana más temprana con respecto a las no indígenas. En tanto que la edad mediana a la primera unión en mujeres indígenas se mantuvo casi igual entre 1997 y 2009.

En 2009 las mujeres no indígenas se unieron a una edad más temprana que en 1997, aunque las indígenas son las que experimentan esta transición siendo más jóvenes, incluso antes de cumplir la mayoría de edad. En cuanto a la llegada del primer hijo, se mostró un ligero incremento en la edad mediana de las mujeres indígenas; en las no indígenas se mantuvo la misma edad.

Finalmente, la edad mediana al primer uso de métodos anticonceptivos en las mujeres indígenas en 1997 ocurría 4.5 años después del nacimiento del primer hijo; en 2009 se redujo a 2.6 años. En el caso de las mujeres no indígenas, en 1997 la edad mediana al primer uso de anticonceptivos ocurrió después de 1.8 años de nacido el primer hijo y en 2009 adoptaron el método a menos de un año del nacimiento.

No obstante, la diferencia entre la edad mediana a la primera relación sexual y al primer uso de anticonceptivos fue de 3.9 años en mujeres indígenas y de 2.2 años en mujeres no indígenas. Esta

---

<sup>12</sup> Se refiere al primer matrimonio o unión consensual.

brecha apunta a que persisten conductas de riesgo en las primeras relaciones sexuales tanto de hablantes como de no hablantes.

**Cuadro 1. Edad mediana de cada una de las transiciones seleccionadas, de mujeres entre 25 y 34 años, según condición de habla de lengua indígena, 1997 y 2009.**

Transiciones de vida	Habla de Lengua indígena		No hablantes de lengua indígena	
	1997	2009	1997	2009
Relación sexual		17.3		18.5
Unión	17.7	17.6	20.2	18.9
Hijo	18.2	18.6	19.7	19.8
Uso de métodos	22.7	21.2	21.5	20.7

Fuente: Estimaciones del CONAPO con base en la Encuesta Nacional de la Dinámica Demográfica de 1997 y 2009.

### 1.7 Atención Materna

Entre las medidas para prevenir riesgos de muerte o enfermedad tanto de la mujer como del niño se encuentra la atención del parto por personal calificado<sup>13</sup>, es decir, por un médico, además de contar con todos los servicios en caso de emergencia.

A nivel nacional, en 1997 se estimó que 84.2% de las mujeres era atendido por un médico, 11.9%, por una partera, 2.2%, ella sola o por otra persona, y 1.7%, por una enfermera o una auxiliar. En 2009 se logró que 94.0% de las mujeres contara con la atención de un médico, lo que implicó la disminución del porcentaje de mujeres atendidas por una partera a 4.0, en tanto que 1.5% contó con una enfermera o auxiliar y disminuyó a 0.6% en las mujeres que estuvieron solas durante el proceso de parto.

En el caso de las indígenas, todavía en 2009 una tercera parte careció de la atención de un médico y una de cada cuatro acudió a una partera; cabe destacar que en ese mismo año casi la mitad de las mujeres indígenas sin escolaridad recurrió a una partera y aumentó el porcentaje de atendidas por un médico conforme ellas tenían un mayor nivel de escolaridad; sólo es hasta que cuentan con secundaria o más que disminuye drásticamente el porcentaje de mujeres que utilizó a una partera (Véase cuadro 2).

<sup>13</sup> A partir de 2008, el Gobierno Federal estableció el *Programa Embarazo Saludable* a nivel nacional, por medio del cual se asegura en todo momento la incorporación al Seguro Popular de todas aquellas mujeres con diagnóstico de embarazo y al momento del parto.



Esta situación contrasta con la atención que recibieron las mujeres no indígenas, ya que en 2009 casi todas contaron con la atención de un médico y sólo las mujeres sin escolaridad presentaron el mayor porcentaje que fue asistido durante el parto por una partera.

**Cuadro 2. Distribución porcentual de las mujeres en edad fértil por grupos de edad y nivel de escolaridad de acuerdo al tipo de agente que la atendió durante el parto, según condición de habla indígena, 1997 y 2009.**

Característica seleccionada	Médico		Enfermera/auxiliar		Partera		Ella sola/otro	
	1997	2009	1997	2009	1997	2009	1997	2009
<b>Hablantes de Lengua indígena</b>	45.6	70.4	3.9	2.3	40.0	25.0	10.5	2.3
Edad de la madre al momento del nacimiento								
12-19	48.1	68.8	3.6	3.5	39.0	26.1	9.4	1.6
20-34	46.2	72.2	4.0	1.5	39.6	24.5	10.2	1.8
35-49	40.8	64.2	3.8	4.4	42.6	26.2	12.8	5.2
Nivel de escolaridad								
Sin escolaridad	27.3	46.4	3.7	4.0	51.2	47.6	17.8	2.0
Primaria incompleta	42.9	64.3	4.3	2.8	42.7	29.9	10.2	3.0
Primaria completa	51.4	72.5	5.1	1.0	36.7	24.4	6.9	2.1
Secundaria y más	80.2	86.2	1.3	1.9	16.1	9.7	2.4	2.1
<b>No hablantes de Lengua indígena</b>	88.0	95.7	1.5	1.4	9.1	2.4	1.4	0.4
Edad de la madre al momento del nacimiento								
12-19	86.0	95.1	1.9	1.7	11.0	2.9	1.2	0.4
20-34	88.9	96.0	1.3	1.4	8.5	2.3	1.3	0.3
35-49	84.7	95.1	1.7	1.2	11.1	2.7	2.5	1.1
Nivel de escolaridad								
Sin escolaridad	63.7	76.9	2.3	4.1	28.1	16.1	6.0	2.9
Primaria incompleta	72.4	87.2	2.5	1.7	21.7	9.3	3.4	1.8
Primaria completa	86.0	93.1	2.1	1.9	10.8	4.2	1.1	0.8
Secundaria y más	96.4	97.8	0.8	1.2	2.4	0.9	0.4	0.1

Fuente: Estimaciones del CONAPO con base en la Encuesta Nacional de la Dinámica Demográfica de 1997 y 2009.

## 1.8 Conclusiones

El diagnóstico presentado en este apartado analiza algunas de las variables determinantes en la calidad de la salud sexual y reproductiva de las mujeres. La actualización en general de los diferentes indicadores demográficos, en particular los de salud sexual y reproductiva, de las mujeres hablantes de lengua indígena es importante porque muestra cambios en la conducta reproductiva y sexual de un grupo de la población tradicionalmente vulnerable.

La población indígena, en especial las mujeres, ha tenido un mayor acceso a la educación que en el pasado, sin embargo, persisten desigualdades con respecto a las mujeres no indígenas.

Entre 2000 y 2010, la participación de las mujeres indígenas en el mercado laboral disminuyó, mientras que aumentó el porcentaje que se dedica a los quehaceres domésticos. Esto implica que debe haber mujeres indígenas que al dejar de formar parte de un trabajo formal, pierden la

protección de sistemas de salud como el IMSS o el ISSSTE, en posible detrimento de acceso a tratamientos médicos, así como a métodos anticonceptivos. Sin embargo, la creación y puesta en marcha del Seguro Popular benefició principalmente a la población indígena, en particular, a todas aquellas mujeres que carecen de un empleo formal.

En lo que respecta al número ideal de hijos, las mujeres indígenas presentaron en 2009 un ideal casi igual al que reportaban en 1997, en ambos casos es más alto que el estimado para las no indígenas. Otra situación que coloca a las mujeres indígenas en una posición de mayor vulnerabilidad es el hecho de que tuvieron su primera relación sexual a una edad más temprana que las no indígenas.

Cabe agregar que las indígenas no sólo reportan haber iniciado su vida sexual más jóvenes, sino que comenzaron a usar métodos anticonceptivos casi dos años después de haber tenido a su primer hijo, en tanto que las no indígenas comienzan a protegerse casi en el mismo año de haber tenido a su primer hijo. Las mujeres indígenas presentaron mayores rezagos en el uso de métodos anticonceptivos, lo que se refleja en una necesidad insatisfecha más elevada que la mostrada por las no indígenas, en especial en las adolescentes.

El análisis por condiciones sociodemográficas sugiere que elevar el nivel de educación formal de las mujeres indígenas implica que ellas tengan mayor acceso a la información, así como mayor capacidad para ejercer sus derechos. Sin embargo, los datos también indican que la falta de acciones de programas de salud reproductiva y de estrategias de planificación familiar comienzan a afectar de manera negativa el uso de métodos anticonceptivos en mujeres con escolaridad de secundaria y más.

Finalmente, podemos concluir que las revisiones médicas a las que acuden las mujeres embarazadas indígenas apenas sobrepasa lo establecido por la Norma Mexicana. El aumento en el número de revisiones prenatales pudiera estar relacionado con el incremento de embarazos atendidos por un médico, en especial de las mujeres con mayores niveles educativos. A pesar de este incremento, la importancia de las parteras sigue siendo notoria al atender alrededor de uno de cada cuatro partos de mujeres indígenas.

Si bien es cierto que la mayoría de los indicadores presentados mejoran para las mujeres indígenas entre 1997 y 2009, se sigue observando que, en este país, ser hablante de lengua indígena representa una condición de vulnerabilidad, situación que debe abatirse mediante el refuerzo de políticas y programas públicos, sin dejar de lado el respeto a su identidad, sus tradiciones y sus costumbres.

## Capítulo 2. Aspectos Técnicos de un Modelo de Regresión Lineal Múltiple

---

**E**l análisis de regresión es una de las técnicas de uso más frecuente para analizar datos multifactoriales. Su atractivo y utilidad general son el resultado del proceso conceptualmente lógico de usar una ecuación para expresar la relación entre una variable de interés (la respuesta) y un conjunto de variables predictivas relacionadas. El análisis de regresión tiene también interés teórico, por las elegantes matemáticas que se usan y por una teoría estadística bien desarrollada. Para usar bien la regresión se requiere apreciar tanto la teoría como los problemas prácticos que se suelen presentar cuando se emplea una técnica con datos del mundo real. Un modelo de regresión donde intervenga más de una variable regresora se llama Modelo de Regresión Lineal Múltiple. En éste capítulo se explica el ajuste y análisis de los modelos.

En la regresión lineal múltiple se utiliza más de una variable explicativa; esto ofrece la ventaja de utilizar más información en la construcción del modelo y, consecuentemente, realizar estimaciones más precisas.

En estadística la regresión lineal o ajuste lineal es un método matemático que modela la relación entre una variable dependiente  $Y$ , las variables independientes o explicativas  $X_i$  y un término aleatorio  $U$ . Este modelo puede ser expresado como:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + U_i \quad \forall i = 1 \dots k$$

$Y_i$ : es la variable dependiente, explicada o regresando.

$X_1, X_2, \dots, X_k$ : son variables independientes, explicativas o regresores.

$\beta_0, \beta_1, \dots, \beta_k$ : son parámetros o coeficientes de regresión, miden la influencia que las variables explicativas tienen sobre el regresando.

Este modelo describe un hiperplano en el espacio de  $k$  dimensiones de las variable regresoras  $X_j$ . El parámetro  $\beta_j$  representa el cambio esperado en la respuesta  $Y$  por cambio unitario en  $X_j$ . Cuando todas las demás variables regresoras  $X_j$  ( $i \neq j$ ) se mantienen constantes. Por esta razón, a los parámetros  $\beta_j$   $j = 0, 1, \dots, k$  se les llama con frecuencia coeficientes de regresión parcial.

En la práctica, se debe elegir cuidadosamente qué variables se van a considerar como explicativas.

Algunos criterios que deben cumplir serán los siguientes:

- Tener sentido numérico.
- No deberá de haber variables repetidas o redundantes.
- Las variables introducidas en el modelo deberán de tener cierta justificación teórica.

- La relación entre variables explicativas entre el modelo y los casos deben ser como mínimo de 1 a 10.
- La relación de las variables explicativas con la variable dependiente debe ser lineal, es decir, proporcional.

Para realizar un análisis de regresión lineal múltiple se hacen las siguientes consideraciones sobre los datos:

- ✓ Normalidad: La distribución de la perturbación (error) aleatoria tiene distribución normal

$$U \sim N(0, \sigma^2)$$

- ✓ Autocorrelación: Es la correlación existente entre los miembros de una serie de observaciones

$$d = \sum_{t=2}^n \frac{(e_t - e_{t-1})^2}{e_t^2}$$

- ✓ Homocedasticidad: Todas las perturbaciones (errores) tienen la misma varianza, es decir, la varianza de los errores es constante

$$V(U_i) = \sigma^2$$

- ✓ Linealidad: Los valores de la variable dependiente están generados por el siguiente modelo lineal:

$$Y = X * \beta + U$$

- ✓ Forma Funcional Lineal: Esta prueba verifica si las combinaciones no lineales de los valores ajustados ayudan a explicar la variable de respuesta  $\hat{y} = E\{y | x\} = \beta x$ . El test de Ramsey prueba si  $(\beta x)^2, (\beta x)^3, \dots, (\beta x)^k$  tiene algún poder en explicar  $Y$ . Esto es ejecutando la siguiente estimación de una regresión lineal  $y = \alpha x + \gamma_1 \hat{y}^2 + \dots + \gamma_{k-1} \hat{y}^k + \varepsilon$ .

- ✓ Estabilidad de parámetros: En general el hecho de cambio de parámetros lleva implícito que existe otra variable independiente que condiciona el modelo y los parámetros cambian de acuerdo a esa variable.

$$b = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon$$

- ✓ Independencia: Las perturbaciones (errores) aleatorias son independientes entre sí:

$$E(U_i * U_j) = 0 \quad \forall i \neq j$$

- ✓ Las variables explicativas  $X_k$  se obtienen sin errores de medida.

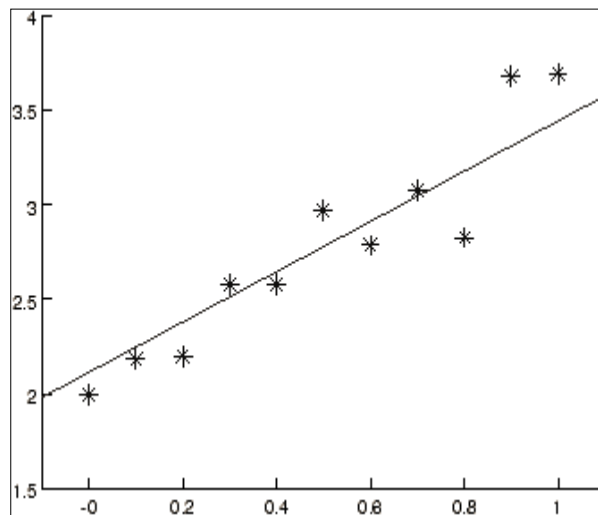
Si admitimos que los datos presentan estas hipótesis se les llama estacionarios, entonces el teorema de Gauss-Markov establece que el método de estimación de mínimos cuadrados va a

producir estimadores óptimos, en el sentido que los parámetros estimados van a estar centrados y van a ser de mínima varianza.

## 2.1. Estimación de los parámetros por el Método de Mínimos Cuadrados

El método consiste en someter al sistema a diferentes condiciones, fijando para ello los distintos valores de las variables independientes  $X_k$ , y registrando en cada caso el correspondiente valor para la variable dependiente  $Y_i$ . De este modo se dispone de una serie de puntos  $(x_1, y_1), \dots, (x_k, y_k)$ , que representados gráficamente, deberían caer sobre una línea recta. Sin embargo, los errores experimentales siempre presentes hacen que no se hallen perfectamente alineados.

Gráfica de método de mínimos cuadrados



El método de mínimos cuadrados determina los valores de los parámetros  $\beta$  y  $U$  de la recta que mejor se ajusta a los datos experimentales.

Los coeficientes son elegidos de forma que la suma de cuadrados entre los valores observados y los pronosticados sea mínima, es decir, que se va minimizando la varianza residual la cual descompone la variabilidad de la variable dependiente  $Y$  en dos componentes o fuentes de variabilidad: una componente va a representar la variabilidad explicada por el modelo de regresión y la otra componente va a representar la variabilidad no explicada por el modelo y, por tanto, atribuida a factores aleatorios.

Se considera la variabilidad de la variable dependiente como:

$$n * \sigma^2 = \sum (y_i - \bar{Y})^2$$

Es decir, la variabilidad de  $Y$  es la suma cuadrática de los valores que toma la variable respecto a la media de la variable.

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

Es decir,

$$VT = VE + VNE$$

Donde:

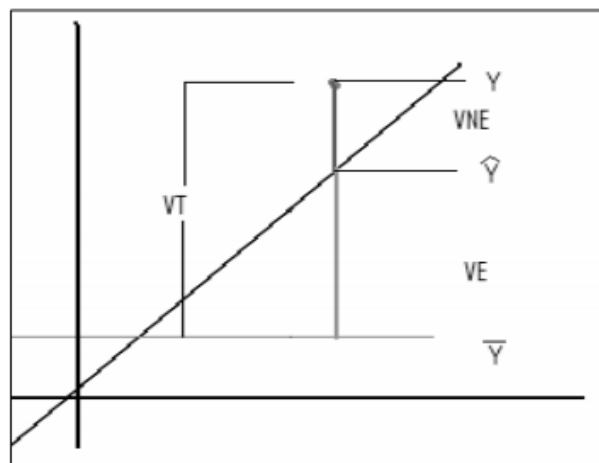
VT es la variabilidad total

VE es la variabilidad explicada

VNE es la variabilidad no explicada

Gráficamente se observa la relación

**Grafica de variabilidad**



Se puede aplicar el método de mínimos cuadrados para estimar los coeficientes de regresión de la ecuación.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + U_i \quad \forall i = 1 \dots n \quad (1)$$

Supongamos que se dispone de  $n > k$  observaciones, y sea  $Y_i$  la  $i$ -ésima respuesta observada, y  $X_{ij}$  la  $i$ -ésima observación o nivel del regresor  $X_j$ . Los datos aparecerán como en la siguiente tabla.

Suponemos que el término del error  $U$  del modelo tiene  $E(U) = 0$ ,  $Var(U) = \sigma^2$  y que los errores no están correlacionados.

**Datos para la regresión lineal múltiple**

Observación	Respuesta	Regresores			
$i$	$y$	$X_1$	$X_2$	...	$X_k$
1	$y_1$	$X_{11}$	$X_{12}$	...	$X_{1k}$
2	$y_2$	$X_{21}$	$X_{22}$	...	$X_{2k}$
⋮	⋮	⋮	⋮		⋮
$n$	$y_n$	$X_{n1}$	$X_{n2}$	...	$X_{nk}$

Desde un punto de vista estadístico, un requisito implícito para que funcione el método de mínimos cuadrados es que los errores de cada medida estén distribuidos de forma aleatoria. El teorema de Gauss-Márkov prueba que los estimadores mínimos cuadráticos carecen de sesgo y que el muestreo de datos no tiene que ajustarse. La técnica de mínimos cuadrados se usa comúnmente en el ajuste de curvas.

Cuando las  $X$  son variables aleatorias sólo es necesario que las observaciones con cada regresor sean independientes, y que la distribución no dependa de los coeficientes de regresión (las  $\beta$ ) o de  $\sigma^2$ . Cuando se prueban hipótesis o se establecen intervalos de confianza, se debe suponer que la distribución condicional de  $Y$  dadas  $X_1, X_2, \dots, X_k$  es normal, con promedio  $\beta_1 + \beta_2 X_2 + \dots + \beta_k X_k$  y varianza  $\sigma^2$ .

Se puede escribir en la siguiente forma el modelo muestral de regresión que corresponde a la ecuación (1):

$$\begin{aligned}
 Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + U_i \\
 &= \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + U_i, \quad i = 1, 2, \dots, n
 \end{aligned}$$

La función de mínimos cuadrados es:

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n U_i^2$$

$$= \sum_{i=1}^n \left( Y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{ij} \right)^2$$

Se debe minimizar la función  $S$  respecto a  $\beta_0, \beta_1, \dots, \beta_k$ . Los estimadores de  $\beta_0, \beta_1, \dots, \beta_k$  por mínimos cuadrados deben satisfacer

$$\frac{\delta S}{\delta \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left( Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij} \right) = 0$$

y

$$\frac{\delta S}{\delta \beta_j} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left( Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij} \right) X_{ij} = 0 \quad j = 1, 2, \dots, k$$

Al simplificar la ecuación anterior, se obtienen las ecuaciones normales de mínimos cuadrados

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_{i1} + \hat{\beta}_2 \sum_{i=1}^n X_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{ik} &= \sum_{i=1}^n Y_i \\ \hat{\beta}_0 \sum_{i=1}^n X_{i1} + \hat{\beta}_1 \sum_{i=1}^n X_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n X_{i1}X_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{i1}X_{ik} &= \sum_{i=1}^n X_{i1}Y_i \\ \vdots & \vdots \\ \hat{\beta}_0 \sum_{i=1}^n X_{ik} + \hat{\beta}_1 \sum_{i=1}^n X_{ik}^2 + \hat{\beta}_2 \sum_{i=1}^n X_{ik}X_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{ik}^2 &= \sum_{i=1}^n X_{ik}Y_i \end{aligned}$$

Nótese que hay  $p=k+1$  ecuaciones normales, una para cada uno de los coeficientes desconocidos de la regresión. La solución de las ecuaciones normales serán los estimadores por mínimos cuadrados  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ .

A veces ayuda tener una interpretación geométrica intuitiva de mínimos cuadrados. Se puede imaginar que el vector de las observaciones  $Y' = [Y_1, Y_2, \dots, Y_n]$  define un vector del origen al punto A de la siguiente figura. Nótese que  $Y_1, Y_2, \dots, Y_n$  forman las coordenadas de un espacio muestral  $n$  dimensional. En dicha figura, el espacio muestral es tridimensional.



## 2.2 Propiedades de los estimadores de mínimos cuadrados.

Las propiedades estadísticas del estimador de mínimos cuadrados  $\hat{\beta}$  se demuestra con facilidad.

Examinemos primero el sesgo:

$$\begin{aligned} E(\hat{\beta}) &= E[(X'X)^{-1}X'Y] = E[(X'X)^{-1}X'(X\beta + U)] \\ &= E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'U] = \beta \end{aligned}$$

porque  $E(U) = 0$  y  $(X'X)^{-1}X'X = 1$ , Entonces  $\hat{\beta}$  es un estimador insesgado de  $\beta$ .

La propiedad de la varianza de  $\hat{\beta}$  se expresa con la matriz de covarianza

$$Cov(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\}$$

que es una matriz simétrica de  $p \times p$ , cuyo  $j$ -ésimo elemento diagonal es la varianza de  $\hat{\beta}_j$  y cuyo  $(ij)$ -ésimo elemento fuera de la diagonal es la covarianza entre  $\hat{\beta}_i$  y  $\hat{\beta}_j$ . La covarianza de la matriz de  $\hat{\beta}$  es

$$Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Por consiguiente, si hacemos  $C = (X'X)^{-1}$ , la varianza de  $\hat{\beta}_j$  es  $\sigma^2 C_{jj}$ , y la covarianza entre  $\hat{\beta}_i$  y  $\hat{\beta}_j$  es  $\sigma^2 C_{ij}$ .

### 2.2.1 Teorema de Gauss-Markov

El teorema de Gauss-Markov establece que el estimador de mínimos cuadrados ordinarios (OLS) de  $\beta$ , que es  $\hat{\beta} = (X'X)^{-1}X'y$ , es el mejor estimador lineal insesgado. En este caso, quiere decir que  $\hat{\beta}$  tiene la mínima varianza, en cierto sentido significativo, entre la clase de todos los estimadores insesgados que son combinaciones lineales de los datos. Un problema es que  $\hat{\beta}$  es un vector, y en consecuencia, su varianza es en realidad una matriz. Por ello se tratará de demostrar que  $\hat{\beta}$  minimiza la varianza para cualquier combinación lineal de los coeficientes estimados,  $f'\hat{\beta}$ . Se observa que

$$\begin{aligned} Var(f'\hat{\beta}) &= f'Var(\hat{\beta})f \\ &= f'[\sigma^2(X'X)^{-1}]f \\ &= \sigma^2 f'(X'X)^{-1}f \end{aligned}$$

que es un escalar, Sea  $\hat{\beta}$  otro estimador insesgado de  $\beta$  que sea una combinación lineal de los datos. De lo que se trata es, entonces, demostrar que  $Var(\hat{\beta}) \geq \sigma^2 \mathbf{1}'(X'X)^{-1}\mathbf{1}$  con al menos una  $\mathbf{1}$  tal que  $Var(\hat{\beta}) > \sigma^2 \mathbf{1}'(X'X)^{-1}\mathbf{1}$ .

Primero, se observa que se puede escribir cualquier otro estimador de  $\beta$  que sea una combinación lineal de los datos en la forma

$$\tilde{\beta} = [(X'X)^{-1}X' + B]y + b_0$$

Siendo B una matriz de  $p \times n$ , y  $b_0$  es un vector de  $p \times 1$  de constantes, que ajusta en forma adecuada al estimador OLS para formar el estimado alternativo. A continuación se observa que si el modelo es correcto, entonces

$$\begin{aligned} E(\tilde{\beta}) &= E([(X'X)^{-1}X' + B]y + b_0) \\ &= [(X'X)^{-1}X' + B]E(y) + b_0 \\ &= [(X'X)^{-1}X' + B]X\beta + b_0 \\ &= (X'X)^{-1}X'X\beta + BX\beta + b_0 \\ &= \beta + BX\beta + b_0 \end{aligned}$$

Por consiguiente  $\tilde{\beta}$  es insesgado si, y sólo si tanto  $b_0 = 0$  como  $BX = 0$ . La varianza de  $\tilde{\beta}$  es

$$\begin{aligned} Var(\tilde{\beta}) &= Var([(X'X)^{-1}X' + B]y) \\ &= [(X'X)^{-1}X' + B]Var(y)[(X'X)^{-1}X' + B]' \\ &= [(X'X)^{-1}X' + B] \sigma^2 I [(X'X)^{-1}X' + B]' \\ &= \sigma^2 [(X'X)^{-1}X' + B][X(X'X)^{-1} + B'] \\ &= \sigma^2 [(X'X)^{-1} + BB'] \end{aligned}$$

porque  $BX = 0$ , lo cual a su vez implica que  $(BX)' = X'B' = 0$ . Como resultado,

$$\begin{aligned} Var(\mathbf{1}'\tilde{\beta}) &= \mathbf{1}'Var(\tilde{\beta})\mathbf{1} \\ &= \mathbf{1}'(\sigma^2[(X'X)^{-1} + BB'])\mathbf{1} \\ &= \sigma^2 \mathbf{1}'(X'X)^{-1}\mathbf{1} + \sigma^2 \mathbf{1}'BB'\mathbf{1} \\ &= Var(\mathbf{1}'\tilde{\beta}) + \sigma^2 \mathbf{1}'BB'\mathbf{1} \end{aligned}$$

Primero se observa que  $BB'$  es al menos una matriz positiva semidefinida, y por consiguiente  $\sigma^2 \mathbf{1}'BB'\mathbf{1} \geq 0$ . A continuación, se ve que se puede definir a  $\mathbf{1}^* = B'\mathbf{1}$ . Por consiguiente,

$$\mathbf{1}'BB'\mathbf{1} = \mathbf{1}^* \mathbf{1}^* = \sum_{i=1}^p \mathbf{1}_i^{*2}$$

que debe ser estrictamente mayor que 0 para cierta  $\int \neq 0$ , a menos que  $B = 0$ . Por lo anterior, el estimado OLS de  $\beta$  es el mejor estimador lineal insesgado.

### 2.3 Estimación de $\sigma^2$

Se puede desarrollar un estimador de  $\sigma^2$  a partir de la suma de los cuadrados de residuales:

$$SS_{Res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 = e'e$$

Se sustituye  $e = Y - X\hat{\beta}$  y se obtiene:

$$\begin{aligned} SS_{Res} &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\ &= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

Como  $X'X\hat{\beta} = X'Y$ , la última ecuación se transforma en:

$$SS_{Res} = Y'Y - \hat{\beta}'X'Y$$

### 2.4 Pruebas de Hipótesis en la Regresión Lineal Múltiple

Una vez estimados los parámetros del modelo, surgen de inmediato dos preguntas:

1. ¿Cuál es la adecuación general del modelo?
2. ¿Cuáles regresores específicos parecen importantes?

Hay varios procedimientos de prueba de hipótesis que demuestran su utilidad para contestar estas preguntas. Las pruebas formales requieren que los errores aleatorios sean independientes y tengan una distribución normal con promedio  $E(U_i) = 0$  y una varianza  $Var(U_i) = \sigma^2$ .

La prueba de la significancia de la regresión es para determinar si hay una relación lineal entre la respuesta  $Y$  y cualquiera de las variables regresoras  $X_1, X_2, \dots, X_k$  (Véase Anexo C. Regresiones). Este procedimiento suele considerarse como una prueba general o global de la adecuación del modelo. Las hipótesis pertinentes son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_a: \beta_j \neq 0 \text{ al menos para una } j.$$

El rechazo de la hipótesis nula implica que al menos uno de los regresores  $X_1, X_2, \dots, X_k$  contribuye al modelo en forma significativa.

Una vez determinado que al menos uno de los regresores es importante, la pregunta lógica es ¿Cuál(es) sirve(n) de ellos? Si se agrega una variable a un modelo de regresión, la suma de cuadrados de la regresión aumenta, y la suma de cuadrados residuales disminuye. Se debe decidir si el aumento de la suma de cuadrados de la regresión es suficiente para garantizar el uso del regresor adicional en el modelo. La adición de un regresor también aumenta la varianza del valor ajustado  $\hat{y}$ , por lo que se debe tener cuidado de incluir sólo regresores que tengan valor para explicar la respuesta. Además, si se agrega un regresor no importante se puede aumentar el cuadrado medio de residuales, y con eso se disminuye la utilidad del modelo.

Las hipótesis para probar la significancia de cualquier coeficiente individual de regresión, como por ejemplo  $\beta_j$ , son:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Si no se rechaza  $H_0: \beta_j = 0$ , quiere decir que se puede eliminar el regresor  $X_j$  del modelo. El estadístico de prueba para esta hipótesis es:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

donde  $C_{jj}$  es el elemento diagonal de  $(X'X)^{-1}$  que corresponde a  $\hat{\beta}_j$ . Se rechaza la hipótesis nula  $H_0: \beta_j = 0$  si  $|t_0| > t_{\alpha/2, n-k-1}$ . Nótese que ésta es en realidad una prueba parcial o marginal, porque el coeficiente de regresión  $\hat{\beta}_j$  depende de todas las demás variables regresoras  $X_i$  ( $i \neq j$ ), que hay en el modelo. Así que se trata de una prueba de contribución de  $X_j$  dados los demás regresores del modelo.

## 2.5 Coeficiente de Determinación $R^2$ y $R^2$ ajustada

El coeficiente de determinación mide la proporción de la variabilidad de la variable dependiente explicada por las variables independientes que en ese momento han sido admitidas en el modelo. A partir del resumen de los modelos generados paso a paso podemos calcular el incremento de  $R^2$ , siendo éste una estimación de la importancia relativa que tiene la variable que acabamos de introducir en el paso correspondiente para predecir la variable dependiente. El coeficiente de determinación ajustado  $\bar{R}_k^2$  no queda influenciado por el número de variables que introducimos.

Una medida de la adecuación de un modelo de regresión es el coeficiente de determinación múltiple,  $R^2$ . Sea  $R_k^2$  el coeficiente de determinación múltiple para un modelo de regresión con subconjunto de  $k$  términos, esto es,  $k - 1$  regresores y un término  $\beta_1$  de ordenada al origen. Este coeficiente se calcula como sigue:

$$R_k^2 = \frac{SS_R(k)}{SS_T} = 1 - \frac{SS_{Res}(k)}{SS_T}$$

en donde  $SS_R(k)$  y  $SS_{Res}(k)$  representan la suma de cuadrados de la regresión y la suma de cuadrados residuales, respectivamente, para un modelo de subconjunto de  $k$  términos. Nótese que hay  $\binom{K}{k-1}$  valores de  $R_k^2$  para cada valor de  $k$ , uno para cada posible modelo de subconjunto de tamaño  $k$ , ahora bien,  $R_k^2$  aumenta al aumentar  $k$ , y es máximo cuando  $k = K + 1$ ; por consiguiente, el analista usa este criterio agregando regresores al modelo hasta el punto en el que una variable adicional ya no es útil porque sólo produce un pequeño aumento de  $R_k^2$ .

Para evitar las dificultades de interpretar a  $R^2$ , algunos analistas prefieren usar la estadística  $R^2$  ajustada, definida para una ecuación de  $k$  términos como sigue:

$$\bar{R}_k^2 = 1 - \left( \frac{n-1}{n-k} \right) (1 - R_k^2)$$

La estadística  $\bar{R}_k^2$  no necesariamente aumenta cuando se introducen más regresores al modelo.

## 2.6 Criterio de Akaike y Criterio de Schwartz

El Criterio de Akaike al igual que el Criterio de Schwartz son una medida que compara modelos con distinta cantidad de variables explicativas

$$AK = -\frac{(2 * \ln(\hat{L}))}{T} + \frac{(2 * k)}{T}$$
$$SC = -\frac{(2 * \ln(\hat{L}))}{T} + \frac{(k * \ln(T))}{T}$$

donde  $k$  es el número de parámetros del modelo y  $T$  el número de períodos.

Siendo mejor el modelo que menor AIC presente, una vez realizadas las correlaciones pertinentes para tener en cuenta las posibles diferencias en la variable dependiente de los modelos. Sin embargo, se elegirá como mejor modelo el que presente mayor  $\ln(L)$  (de nuevo realizando las correcciones pertinentes ante diferencias en la medida de la variable dependiente).

Para detectar la autocorrelación suelen ser útiles las gráficas residuales. La presentación más adecuada es la de los residuales en función del tiempo. Si hay autocorrelación positiva, los residuales de igual signo se presentarán en grupo, que indica que no hay los suficientes cambios de signo en la secuencia de los residuales. Por otra parte, si hay autocorrelación negativa, los residuales cambiarán de signo con demasiada rapidez.

## 2.7 Detección de la presencia de autocorrelación

Para detectar la autocorrelación suelen ser útiles las gráficas de residuales. La presentación más adecuada es la de los residuales en función del tiempo. Si hay autocorrelación positiva, los residuales de igual signo se presentarán en grupos, que indica que no hay los suficientes cambios de signos en la secuencia de los residuales. Por otra parte, si hay autocorrelación negativa, los residuales cambiarán de signo con demasiada rapidez.

Para detectar la presencia de autocorrelación se pueden aplicar diversas pruebas estadísticas. Una prueba de raíz unitaria (Véase Anexo B. Pruebas de Raíz Unitaria) es un AR(1), es decir, proceso auto regresivo de orden 1 es de la forma:

$$y_t = \alpha + \beta y_{t-1} + U_t$$

$$U_t \sim N(0, \sigma^2)$$

Si  $\beta = 1$  vamos a llamarlo prueba de raíz unitaria (caminata aleatoria)

$$AR(1): y_t = \alpha + \beta y_{t-1} + U_t$$

$$y_t - y_{t-1} = \alpha + \beta y_{t-1} + U_t - y_{t-1}$$

$$y'_t: \Delta y_t = \alpha + (\beta - 1)y_{t-1} + U_t$$

$$y'_t = \alpha + \varphi y_{t-1} + U_t \quad (2)$$

La ecuación (2) se llama de Dickey Fuller con intercepto ( $\alpha$ )

$$y_t = \alpha + \beta y_{t-1} + \gamma t + U_t$$

$$y_t - y_{t-1} = \alpha + \beta y_{t-1} - y_{t-1} + \gamma t + U_t$$

$$\Delta y_t = \alpha + (\beta - 1)y_{t-1} + \gamma t + U_t$$

$$y'_t = \alpha + \varphi y_{t-1} + \gamma t + U_t \quad (3)$$

La ecuación (3) es de Dickey Fuller con tendencia e intercepto

$$y_t = \beta y_{t-1} + U_t$$

$$y_t - y_{t-1} = \beta y_{t-1} - y_{t-1}$$

$$\Delta y_t = (\beta - 1)y_{t-1} + U_t \quad (4)$$

La ecuación (4) es de Dickey Fuller sin tendencia ni intercepto.

Si se agregan rasgos de  $y_t$  a las especificaciones, entonces (5) es D-F con intercepto, (6) es D-F con tendencia e intercepto y (7) sin tendencia ni intercepto

$$I: y'_t = \alpha + \varphi y_{t-1} + \varphi_1 y'_{t-1} + \dots + \varphi_n y'_{t-k} + U_t \quad (5)$$

$$T e I: y'_t = \alpha + \varphi y_{t-1} + \gamma t + \varphi'_1 y'_{t-1} + \dots + \varphi'_k y'_{t-k} + U_t \quad (6)$$

$$N: y'_t = \varphi y_{t-1} + \varphi_1 y'_{t-1} + \dots + \varphi_k y'_{t-k} + U_t \quad (7)$$

En la revisión de supuestos  $\varphi_0$  es una prueba que hace una regresión auxiliar entre el residuo y el momento actual y el residuo en momentos anteriores.

## 2.8 Intervalos de confianza

Para construir estimados de intervalos de confianza de los coeficientes de regresión  $\beta_j$ , se continuará suponiendo que los errores  $U_i$  están distribuidos normal e independientemente, con promedio cero y varianza  $\sigma^2$ . En consecuencia, las observaciones  $y_i$  están distribuidas en forma normal e independiente, con media  $\beta_0 + \sum_{j=1}^k \beta_j X_{ij}$ , y varianza  $\sigma^2$ . Como el estimador  $\hat{\beta}$  por mínimos cuadrados es una combinación lineal de las observaciones, también está distribuido normalmente, con vector medio  $\beta$  y matriz de covarianza  $\sigma^2(X'X)^{-1}$ . Esto implica que la distribución marginal de cualquier coeficiente de regresión  $\hat{\beta}_j$  es normal, con media  $\beta_j$  y varianza  $\sigma^2 C_{jj}$ , donde  $C_{jj}$  es el j-ésimo elemento diagonal de la matriz  $(X'X)^{-1}$ . En consecuencia, cada uno de los estadísticos

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}, \quad j = 0, 1, \dots, k \quad (2)$$

se distribuye como  $t$ , con  $n-p$  grados de libertad.

De acuerdo con el resultado de la ecuación (2), se puede definir un intervalo de confianza de **100(1- $\alpha$ )** por ciento para el coeficiente de regresión  $\beta_j$ ,  $j = 0, 1, \dots, k$ , como sigue

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}}.$$

## 2.9 Multicolinealidad

Los modelos de regresión se usan en una gran cantidad de aplicaciones. Un problema serio que puede influir mucho sobre la utilidad de un modelo de regresión es la multicolinealidad, o dependencia casi lineal entre las variables de regresión.

La multicolinealidad implica una dependencia casi lineal entre los regresores. La presencia de dependencias casi lineales puede influir en forma dramática sobre la capacidad de estimar coeficientes de regresión.

El uso y la interpretación de un modelo de regresión lineal múltiple dependen, con frecuencia, en forma explícita o implícita, de los estimados de los coeficientes individuales de regresión. Entre los ejemplos de las inferencias que se hacen a menudo están:

- Identificación de los efectos relativos de las variables regresoras.



- Predicción y/o estimación.
- Selección de un conjunto adecuado de variables para el modelo.

Si no hay relación lineal entre los regresores, se dice que son ortogonales. Cuando los regresores son ortogonales se pueden hacer con relativa facilidad inferencias como las anteriores, desafortunadamente, en la mayor parte de las aplicaciones de regresión, los regresores no son ortogonales. A veces no es grave la falta de ortogonalidad, sin embargo, en algunos casos los regresores tienen una regresión lineal casi perfecta, y en esos casos, las inferencias basadas en el modelo de regresión pueden ser engañosas o erróneas. Cuando hay dependencias casi lineales entre los regresores, se dice que existe el problema de multicolinealidad.

$$Y = X\beta + U$$

donde  $y$  es un vector de  $n * 1$  de respuestas,  $X$  es una matriz de  $n * t$  de las variables regresoras,  $\beta$  es un vector de  $t * 1$  de las constantes desconocidas y  $U$  es un vector de  $n * 1$  de los errores aleatorios, siendo  $U_i \sim NID(0, \sigma^2)$ .

### *2.10 Estadísticos básicos para el análisis de las series*

Obtener los estadísticos descriptivos básicos de cualquier serie de datos es muy simple en E-Views. Así, una de las opciones de vista de cualquier serie individual es, precisamente, la de histograma y estadísticos. Si ejecutamos esta opción podremos observar en una misma pantalla el histograma de frecuencias de la serie y algunos estadísticos básicos.

- El histograma de frecuencias representa gráficamente la distribución de frecuencias de los valores de la serie. Esta representación tiene en general poco sentido para series temporales en las que, habitualmente, no suelen repetirse valores en distintos períodos a lo largo de la muestra considerada. Por ello, EViews genera unos intervalos de igual extensión y representa con barras el número de ocasiones en las que la serie toma valores comprendidos en cada uno de estos intervalos. Los estadísticos descriptivos ofrecidos a la derecha del histograma son conceptualmente muy sencillos y fáciles de interpretar. Aparecen en primer lugar dos medidas de tendencia central de la serie:
  - La media de la serie, calculada como promedio aritmético.
  - La mediana de la serie; es decir, aquel valor que separa los valores de la serie en dos conjuntos de igual densidad de frecuencias. En el caso de una serie temporal representada por intervalos, se toma como mediana la marca de clase del intervalo mediano, convenientemente desplazada según la mayor o menor carga a izquierdas y a derechas del mismo.

- El valor máximo y mínimo de la serie.
- La desviación típica de la serie (raíz de la varianza de la serie).
- La simetría de la serie con respecto a su media es un cálculo sencillo (que utiliza el momento de orden tres con respecto a la media así como el cubo de la desviación típica de la serie). Esta expresión toma el valor cero en el caso de una distribución perfectamente simétrica (con igual densidad de frecuencias a izquierda y derecha de la media). Valores positivos indican asimetría a derechas respecto a la normal y valores negativos asimetría a izquierdas.

$$S = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})^3}{n}}{\sigma^3}$$

- La curtosis de una serie indica si su distribución de frecuencias es más aplanada o más apuntada que una distribución normal, es decir, si alrededor de la media se concentran más o menos valores que en el caso de una normal y por tanto sus “colas” son más o menos estrechas. La curtosis de una serie puede calcularse utilizando el momento de orden cuatro con respecto a la media y la potencia cuarta de la desviación típica de la serie. Para una distribución normal el valor de este coeficiente de curtosis es 3 (distribución mesocúrtica). Valores superiores a 3 indican un apuntamiento mayor que el de una distribución normal (distribución leptocúrtica) y valores inferiores a 3, un apuntamiento menor (distribución platocúrtica).

$$K = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})^4}{n}}{\sigma^4}$$

- El test o contraste de Jarque-Bera (Véase Anexo C. Regresiones), analiza la relación entre los coeficientes de simetría y curtosis de los residuos de la ecuación y los correspondientes a los de una distribución normal, de forma tal que si estas relaciones son suficientemente diferentes se rechazaría la hipótesis nula de normalidad de los residuos. La idea del cálculo se apoya en comparar simultáneamente los valores obtenidos para los coeficientes de simetría y curtosis con los de referencia para una normal. La ventaja del cálculo es que su resultado permite contrastar la hipótesis nula de que la serie se distribuye como una Normal ya que esta expresión (JB) se distribuye como una  $\chi^2$  con dos grados de libertad.

$$JB = \frac{n-k}{6} \left[ S^2 + \frac{1}{4}(k-3)^2 \right]$$

El término “k” minúscula se refiere al número de variables exógenas utilizadas, cuando este test se usa para contrastar la normalidad de los residuos de una determinada regresión y toma el valor “0” cuando contrastamos la normalidad de una serie individual. El valor de la probabilidad ofrecido por E-views, se entiende como el nivel de significación asociado al rechazo de la nula: valores pequeños para esa probabilidad (inferiores a 0,05 ó 0,1) indicarían, por tanto, ausencia de normalidad en la distribución de valores de la variable analizada.

Como en otras ocasiones el valor del contraste viene acompañado con el correspondiente nivel de probabilidad asociado al rechazo de la hipótesis nula siendo cierta, de forma tal que si dicho valor de probabilidad fuera inferior al 10%, rechazaríamos la hipótesis nula, con el 90% de confianza, y deberíamos admitir la no normalidad del residuo.

Los estadísticos de normalidad de una serie son comúnmente utilizados en el contexto de la validación de un modelo básico de regresión, ya que la normalidad de las variables (en especial de la perturbación aleatoria) es una de las hipótesis básicas sobre la que se apoyan las buenas propiedades de los procedimientos de estimación e inferencia.

### *2.11 EViews. Paquete estadístico.*

Para este modelo utilizaremos EViews que es un paquete estadístico para Microsoft Windows, usado principalmente para análisis econométrico. Ha sido desarrollado por Quantitative Micro Software (QMS). La primera versión, 1.0 salió al mercado en marzo de 1994, reemplazando al MicroTSP. La versión más actualizada de EViews es la 7.2 que fue lanzada en noviembre de 2011.

EViews combina la tecnología de hoja de cálculo con tareas tradicionales encontradas en software estadístico tradicional, empleando una interfaz de usuario gráfica. Estas características se combinan con un poderoso lenguaje de programación.

EViews puede ser empleado para análisis estadístico general, pero es especialmente útil para realizar análisis econométrico, como modelos de corte transversal, datos en panel y estimación y predicción con modelos de series de tiempo. Entre los tipos de archivo con los que es compatible destacan el Excel, SPSS, SAS, Stata, RATS, y TSP.

## Capítulo 3. Aplicación del Modelo de Regresión Lineal Múltiple

---

### 3.1 Aplicación del Modelo

**E**l objetivo principal de éste trabajo es encontrar si existe una correlación entre los nacimientos que se registran en México y las tasas de fecundidad, mortalidad, natalidad, nupcialidad, los matrimonios y la población del mismo.

Para este modelo se utilizan las variables explicativas:

- Tasa de fecundidad, indica la relación entre el número total de nacimientos registrados y la población femenina en edad fértil (dada en porcentaje) en México de 1990 a 2011.
- Tasa de mortalidad, es la relación que hay entre las defunciones registradas en todo el país y la población total (dada en porcentaje) en México de 1990 a 2011.
- Tasa de natalidad, es la relación entre el número total de nacimientos registrados y la población total (dada en porcentaje) en México de 1990 a 2011.
- Tasa de nupcialidad, indica la relación entre los matrimonios registrados en todo el país y la población total (dada en porcentaje) en México de 1990 a 2011.
- Matrimonios en México de 1990 a 2011.
- Población Total en México de 1990 a 2011.

Las variables descritas explicarán los nacimientos ocurridos en México de forma anual en el periodo de 1990 a 2011 a través de un modelo de regresión lineal múltiple.

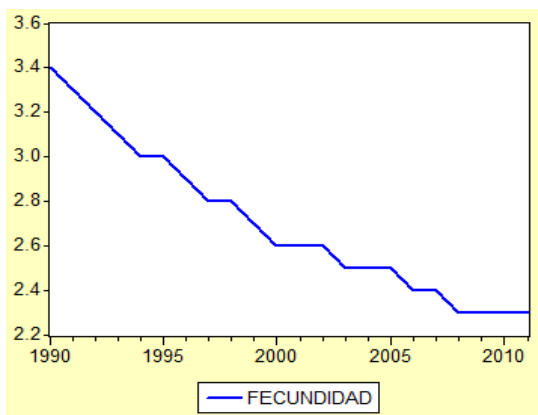
Los datos fueron extraídos de la página de internet del Instituto Nacional de Estadística y Geografía (INEGI) en junio de 2013.

Debido a que los datos están registrados como una serie de tiempo, se tiene que hacer el cambio para que los datos estén de forma estacionaria (caminata aleatoria) y así poder realizar la regresión.

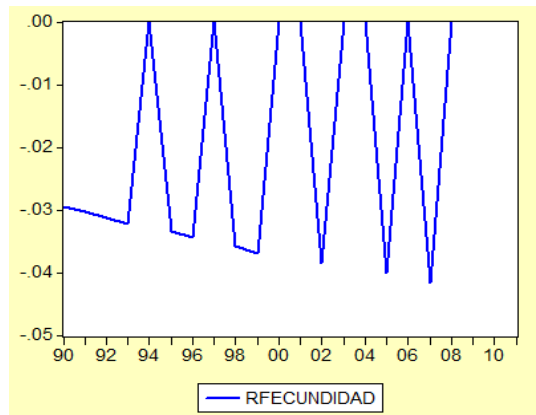
A continuación se muestran las gráficas de cada una de las variables independientes tanto en serie de tiempo como de forma estacionaria.

Una prueba que tenemos que hacer a los datos que vamos a utilizar para la regresión es un gráfico para poder identificar los componentes de una serie de tiempo: tendencia, estacionalidad, ciclicidad e irregularidad. Para verificar los gráficos siguientes, se hace una prueba en E-Views con los datos diferenciados, es decir, de forma estacionaria para saber si los gráficos tienen algún cambio y poder visualizarlo con mayor facilidad, es decir:  $\frac{X_t - X_{t-1}}{X_{t-1}}$ .

**Gráfica 1. Fecundidad. Serie de Tiempo.**<sup>14</sup>

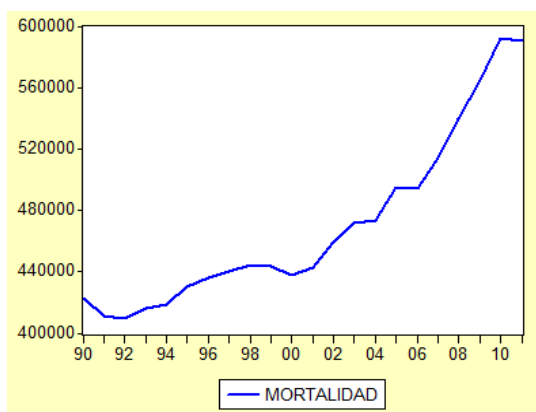


**Gráfica 2. Fecundidad. Estacionaria.**

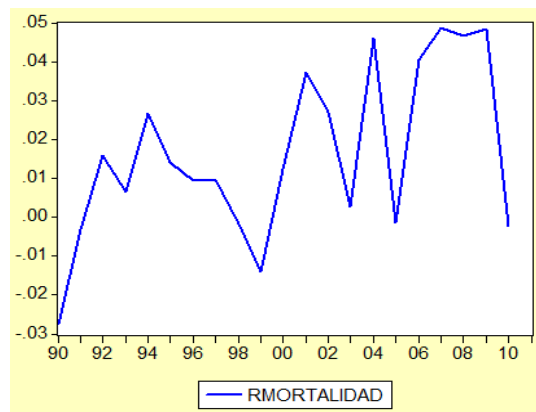


La grafica 1 muestra como ha descendido la tasa de fecundidad en México en un lapso de 22 años, durante el período 1990 a 2011. La gráfica 2 muestra la tasa de fecundidad pero de forma estacionaria para poder realizar la regresión.

**Gráfica 3. Mortalidad. Serie de Tiempo.**



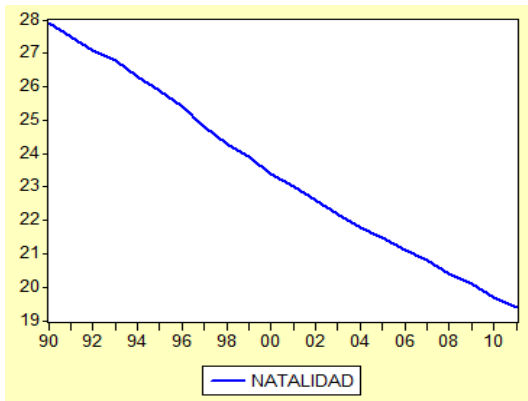
**Gráfica 4. Mortalidad Estacionaria.**



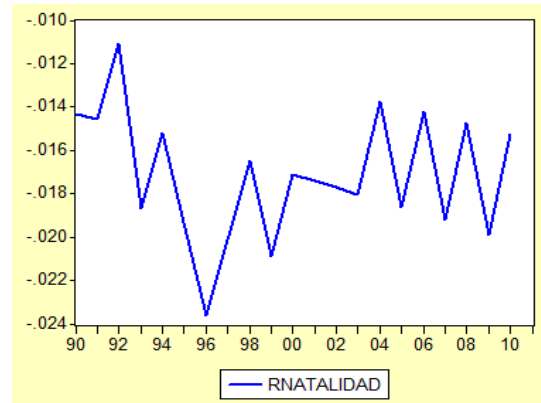
En la gráfica 3 se puede ver que la tasa de mortalidad en general ha tenido un aumento durante el período de 1990 a 2012. La gráfica 4 muestra la tasa de mortalidad de manera estacionaria para poder realizar la regresión.

<sup>14</sup> Gráficas 1-14 realizadas con datos de la página de internet del INEGI 2013. Elaboración propia con el programa EViews.

**Gráfica 5. Natalidad. Serie de Tiempo.**

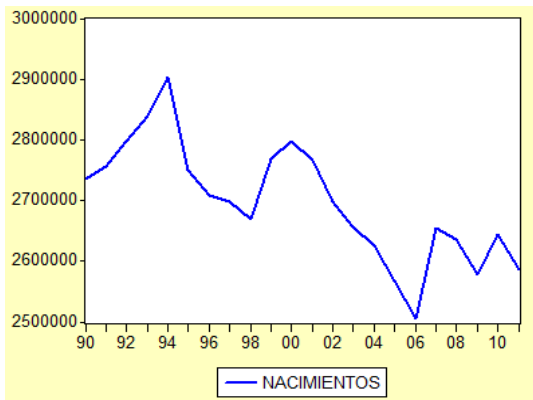


**Gráfica 6. Natalidad Estacionaria.**

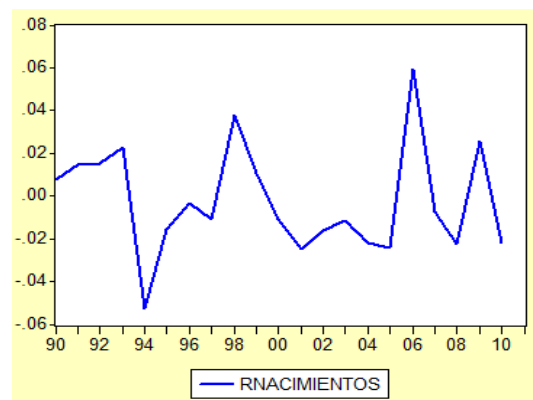


Aquí se muestra el descenso en la tasa de natalidad en un periodo de 22 años. Lo mismo que en la gráfica 6 de manera estacionaria para poder realizar la regresión.

**Gráfica 8. Nacimientos. Serie de tiempo.**

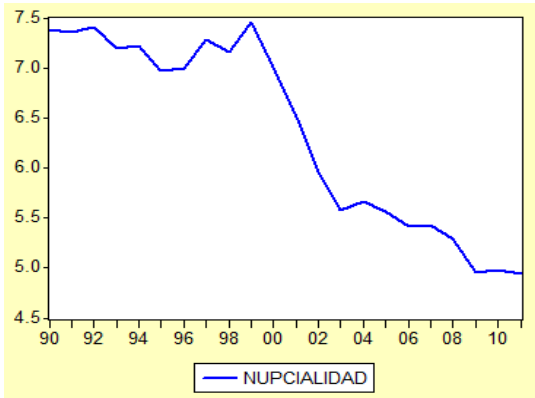


**Gráfica 9. Nacimientos Estacionaria.**

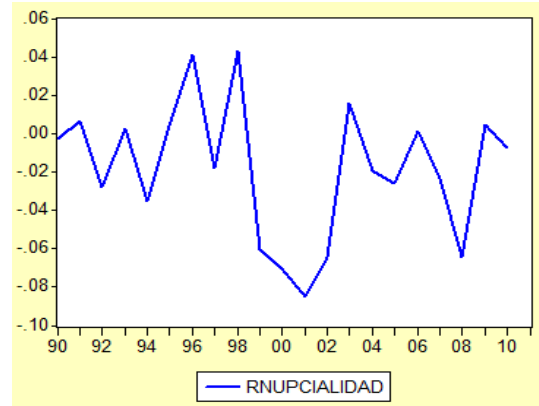


En la gráfica 8 se puede observar un descenso en los nacimientos registrados en México de 1990 a 2011. La gráfica 12 de manera estacionaria para poder realizar la regresión.

**Gráfica 9. Nupcialidad. Serie de tiempo.**

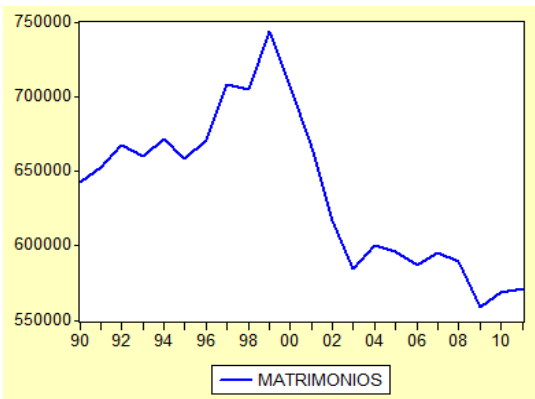


**Gráfica 10. Nupcialidad Estacionaria.**

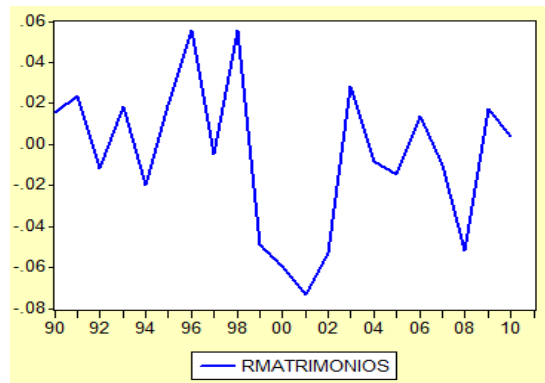


En la gráfica 9 observamos el período de 1990 a 2011; donde los primeros 11 años la tasa de nupcialidad se ha mantenido con el mismo comportamiento; pero después tiene un descenso muy importante. La gráfica 10 se muestra de forma estacionaria para poder realizar la regresión.

**Gráfica 11. Matrimonios. Serie de tiempo.**

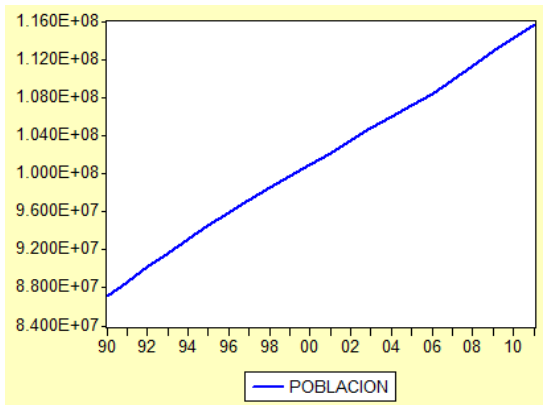


**Gráfica 12. Matrimonios Estacionaria.**

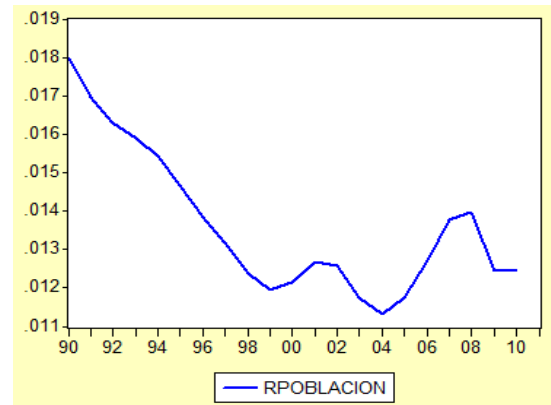


En la gráfica 11 se muestran los matrimonios registrados en México de 1990 a 2011. De igual manera, se observa que los primeros 11 años tiene un crecimiento y de manera posterior un descenso importante. La gráfica 12 se muestra de manera estacionaria para poder realizar la regresión.

Gráfica 13. Población.



Gráfica 14. Población Estacionaria.



Gráfica 13 que muestra claramente el aumento de la población total en México de 1990 a 2011. Gráfica 14 de manera estacionaria para poder realizar la regresión.

Otra forma para saber si la serie es no estacionaria es a través de pruebas de raíz unitaria (Véase Anexo B. Pruebas de Raíz Unitaria), dicha prueba es la que nos va a indicar cuales son los datos que vamos a utilizar para realizar la regresión.

En la tabla siguiente se muestran los resultados de las pruebas de raíz unitaria para cada una de las variables tanto no estacionaria (serie de tiempo) como estacionarias (caminata aleatoria).

Tabla 1. Pruebas de Raíz Unitaria para cada variable no estacionaria y estacionaria

Variable	Intercepto	Tendencia e Intercepto	Ninguno
Nacimientos	-1.436992(0.5447)	-2.954909(0.1669)	-0.523388(0.4778)
rnacimientos	-4.553593(0.0020)	-4.440987(0.0112)	-4.625794(0.0001)
Matrimonios	-0.616925(0.8467)	-1.888019(0.6249)	-0.693689(0.4042)
rmatrimonios	-3.503950(0.0189)	-3.493231(0.0675)	-3.516432(0.0013)
Población	-0.817973(0.7931)	-1.986899(0.5745)	34.96914(0.9999)
rpoblación	-2.818156(0.0735)	-1.649841(0.7353)	-2.193900(0.0304)
Fecundidad	-2.935616(0.0581)	-1.707839(0.7115)	-5.311174(0.0000)
rfecundidad	-5.319676(0.0004)	-6.406006(0.0002)	-2.995445(0.0048)
Mortalidad	2.223534(0.9998)	-1.061599(0.9120)	3.648945(0.9997)
rmortalidad	-3.657651(0.0137)	-3.999754(0.0264)	-2.379727(0.0202)
Nupcialidad	-0.056657(0.9424)	-1.884261(0.6268)	-2.251756(0.0266)
rnupcialidad	-3.562185(0.0168)	-3.525046(0.0637)	-2.980199(0.0050)
natalidad	-2.126694(0.2369)	0.091662(0.9948)	-25.51592(0.0001)
rnatalidad	-4.288087(0.0036)	-4.121636(0.0209)	-0.470721(0.04986)

Datos calculados en el programa EViews a partir de datos extraídos de la página de internet del INEGI. Elaboración propia.

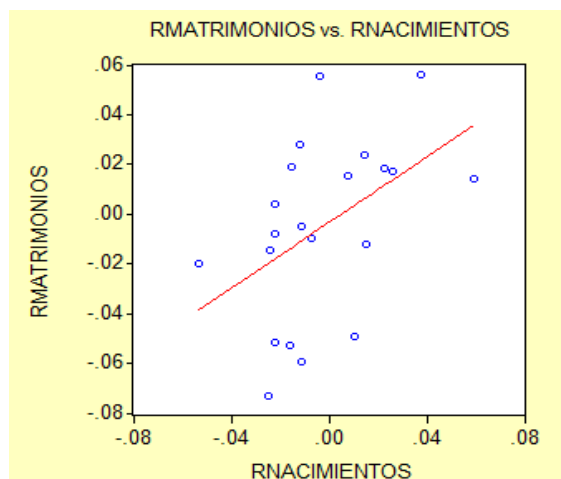


Los datos mostrados en la tabla 1 sirven para seleccionar las variables a utilizar en la regresión, en este caso, se toman los resultados de probabilidad menor a 0.1, es decir, tienen una probabilidad significativa del 90% de confianza de no rechazar la hipótesis nula de la variable, es decir, si la probabilidad es mayor a 0.1, se descarta la variable en cuestión y se continúa con el proceso hasta que solo queden las variables significativas para el modelo de regresión. En este caso, sólo queda la variable que se explicará (rnacimientos) y 4 variables explicativas (rmatrimonios, rfecundidad, rmortalidad y rnpucialidad). (Véase Anexo B. Pruebas de Raíz Unitaria).

Una vez realizadas las pruebas y con la información antes obtenida, se inicia el análisis para explicar los nacimientos a partir de los matrimonios, de la tasa de fecundidad, tasa de mortalidad y la tasa de nupcialidad ocurridos en México de 1990 a 2011.

En seguida se muestran los diagramas de dispersión de cada una de las variables estacionarias que nos ayudarán a explicar el problema propuesto.

**Diagrama de dispersión 1.**<sup>15</sup>



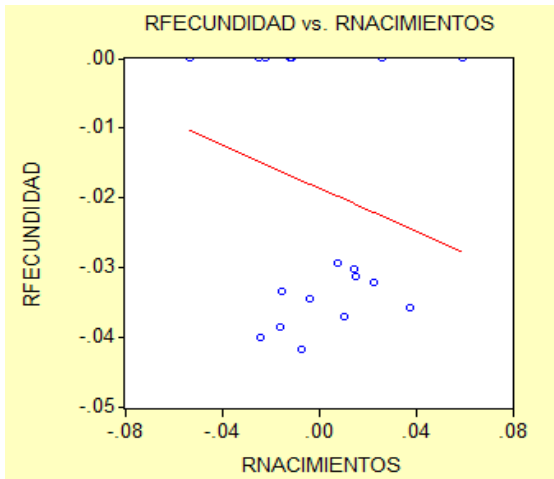
En este diagrama se puede observar que existe una relación positiva entre Rnacimientos y Rmatrimonios. El coeficiente de correlación entre estas dos variables es 0.4664, es decir, hay un 46.64% de relación positiva entre Rnacimientos y Rmatrimonios.

**Matriz de Coeficiente de Correlación.**

	RNACIEMENTOS	RMATRIMONIOS
RNACIEMENTOS	1.000000	0.466403
RMATRIMONIOS	0.466403	1.000000

<sup>15</sup> Diagramas de dispersión 1-4 realizados con el programa EViews a partir de los datos obtenidos en la tabla 1. Elaboración propia.

**Diagrama de dispersión 2.**

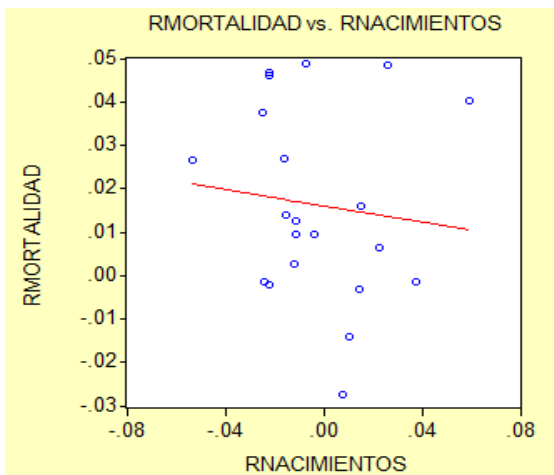


En este diagrama se puede observar que existe una relación negativa entre Rnacimientos y Rfecundidad. El coeficiente de correlación entre estas dos variables es -0.2178, es decir, hay un 21.78% de relación negativa entre Rnacimientos y Rfecundidad.

Matriz de Coeficiente de Correlación.

	RNACIMIENTOS	RFECUNDIDAD
RNACIMIENTOS	1.000000	-0.217812
RFECUNDIDAD	-0.217812	1.000000

**Diagrama de dispersión 3.**

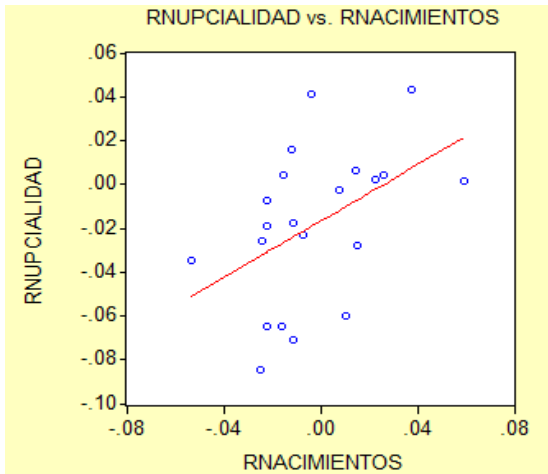


En el diagrama se puede observar que existe una relación negativa entre Rnacimientos y Rmortalidad. El coeficiente de correlación entre estas dos variables es -0.1090, es decir, hay un 10.90% de relación negativa entre Rnacimientos y Rmortalidad.

Matriz de Coeficiente de Correlación.

	RNACIMIENTOS	RMORTALIDAD
RNACIMIENTOS	1.000000	-0.109024
RMORTALIDAD	-0.109024	1.000000

Diagrama de dispersión 4.



Este diagrama indica que existe una relación positiva entre Rnacimiento y Rnupcialidad. El coeficiente de correlación entre estas dos variables es 0.4640, es decir, hay un 46.40% de relación positiva entre Rnacimiento y Rnupcialidad.

Matriz de Coeficiente de Correlación.

	RNACIEMENTOS	RNUPCIALIDAD
RNACIEMENTOS	1.000000	0.464032
RNUPCIALIDAD	0.464032	1.000000

### 3.2 Estimación del modelo

Como ya se vio anteriormente, la fórmula de regresión lineal múltiple es:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} + U_i \quad \forall i = 1 \dots n$$

Primero hacemos la regresión de una variable con la que se explicará para saber si alguna de ellas vuelve a rnacimientos no explicativa, si esto pasa, se descarta dicha variable y continuamos con las demás (Véase Anexo C. Regresiones)

Por lo que con los cálculos efectuados en el programa EViews, llegamos a lo siguiente:

$$Rnacimientos = 0.8887 + (0.0331)Rmatrimonios$$

Dependent Variable: RNACIMIENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 20:45				
Sample (adjusted): 1990 2010				
Included observations: 21 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000725	0.005108	-0.141903	0.8887
RMATRIMONIOS	0.329167	0.143223	2.298287	0.0331
R-squared	0.217531	Mean dependent var		-0.002355
Adjusted R-squared	0.176349	S.D. dependent var		0.025544
S.E. of regression	0.023183	Akaike info criterion		-4.600421
Sum squared resid	0.010211	Schwarz criterion		-4.500942
Log likelihood	50.30442	F-statistic		5.282123
Durbin-Watson stat	1.782626	Prob(F-statistic)		0.033068

Tenemos que la variable dependiente Rnacimientos fue estimada a partir de 7 variables explicativas con el método de mínimos cuadrados con un tamaño muestral real de 21 observaciones. (Véase Anexo C. Regresiones).

Ahora bien, el coeficiente asociado a la variable Rmatrimonios presenta un signo positivo (0.329) con una probabilidad de 3.31% de rechazar la hipótesis nula siendo cierta, es decir, tenemos un 96.69% de nivel de confianza de no rechazar la variable Rmatrimonios.

De acuerdo con el coeficiente de determinación ( $R^2$ ) el modelo explica el 21.75% de la variación de Rnacimientos, mientras que el coeficiente de determinación ajustado ( $\bar{R}^2$ ) explica el 17.63% de la variación de la variable endógena, esto es  $\bar{R}^2$  hace una penalización a  $R^2$  cada vez que aumenta el modelo.

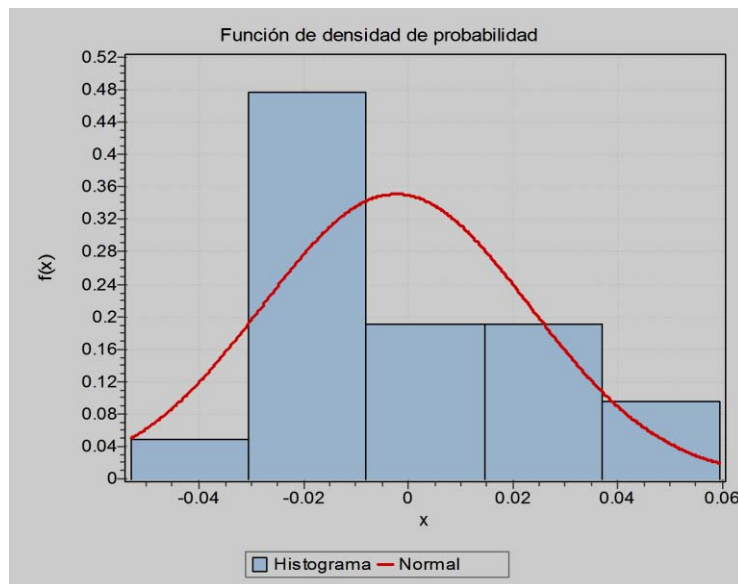
En el modelo que presentamos, la variable de Rmatrimonios presenta un estadístico t con un valor cercano a 2, por lo que podríamos dudar de su significatividad. Comprobando su probabilidad asociada, veríamos que con dicho valor t podríamos rechazar la hipótesis nula de no significatividad siendo esta cierta en un 3.31% de las ocasiones, lo que supondría que estamos trabajando con un 96.69% de confianza.

A continuación se hace un análisis de los supuestos que debe cumplir el modelo de regresión.

### 3.2.1 Normalidad

El modelo presentado cumple con el supuesto de normalidad ya que tenemos que la desviación estándar es de 0.022596, esto quiere decir que es la media de la distancia entre los datos, el coeficiente de curtosis es de 3.250225 que es la proporción de la varianza que se explica por la combinación de datos extremos respecto a la media en contraposición con datos poco alejados de la misma y el estadístico Jarque-Bera muestra que la probabilidad es muy elevada, 0.789966, por lo que debemos asumir la hipótesis nula y afirmar que la distribución de residuos es normal. (Véase Anexo D. Prueba de Normalidad).

**Gráfica 15. Normalidad.**



En la gráfica se puede observar que para el modelo de regresión se cumple el supuesto de normalidad para  $\alpha = 0.2$ ,  $\alpha = 0.1$ ,  $\alpha = 0.5$  y  $\alpha = 0.01$ . (Véase Anexo D. Prueba de Normalidad)

### 3.2.2 Auto correlación

Como puede verse en el estadístico Durbin-Watson es muy cercano a 2, por lo tanto estamos en presencia de no autocorrelación positiva.

Dependent Variable: RNACIMIENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 22:55				
Sample (adjusted): 3 22				
Included observations: 20 after adjustments				
Convergence achieved after 6 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000891	0.004551	-0.195734	0.8471
RMATRIMONIOS	0.327968	0.139581	2.349664	0.0311
AR(1)	-0.188415	0.244625	-0.770218	0.4517
R-squared	0.212339	Mean dependent var		-0.001235
Adjusted R-squared	0.119673	S.D. dependent var		0.025673
S.E. of regression	0.024088	Akaike info criterion		-4.476719
Sum squared resid	0.009864	Schwarz criterion		-4.327359
Log likelihood	47.76719	F-statistic		2.291444
Durbin-Watson stat	2.006758	Prob(F-statistic)		0.131487
Inverted AR Roots	-0.19			

### 3.2.3 Homocedasticidad

Hablamos de homocedasticidad si el error cometido por el modelo, con dos o más muestras, tiene siempre la misma varianza. El no cumplimiento de esta propiedad puede conllevar a que las conclusiones que se extraigan del modelo sean falsas.

Con los resultados obtenidos por el programa, observamos que no hay indicios de auto correlación ya que los resultados del test muestran que el valor de la  $R^2$  en la regresión instrumental es bastante aceptable para un nivel de significación del 51%, es decir, del 49.02%. Por otro lado, aparecen coeficientes estadísticamente significativos al 99.63 para la variable de “rmatrimonios al cuadrado”.

White Heteroskedasticity Test:				
F-statistic	0.617057	Probability	0.550558	
Obs*R-squared	1.347418	Probability	0.509814	
Test Equation: Dependent Variable: RESID^2 Method: Least Squares Date: 10/07/15 Time: 20:54 Sample: 1990 2010 Included observations: 21				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000630	0.000220	2.862966	0.0103
RMATRIMONIOS	0.001035	0.005199	0.199122	0.8444
RMATRIMONIOS^2	-0.108977	0.120354	-0.905463	0.3772
R-squared	0.064163	Mean dependent var	0.000486	
Adjusted R-squared	-0.039819	S.D. dependent var	0.000747	
S.E. of regression	0.000762	Akaike info criterion	-11.38924	
Sum squared resid	1.05E-05	Schwarz criterion	-11.24002	
Log likelihood	122.5870	F-statistic	0.617057	
Durbin-Watson stat	2.354878	Prob(F-statistic)	0.550558	

### 3.2.4 Forma Funcional Lineal

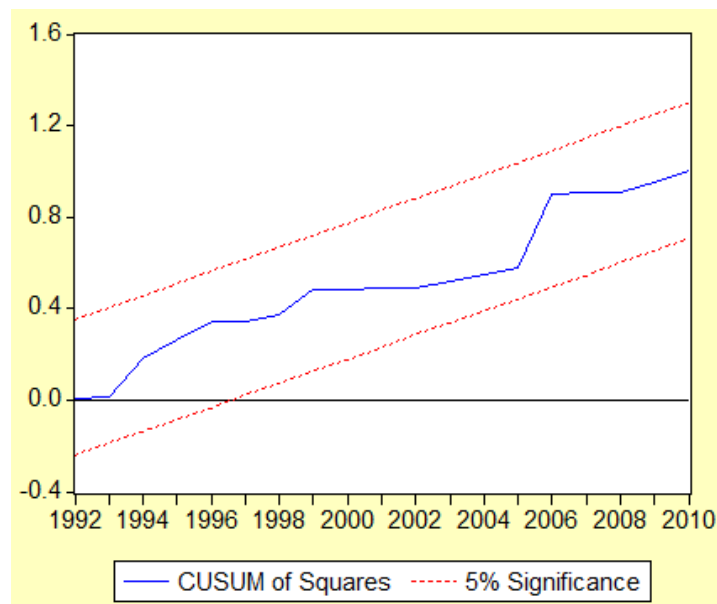
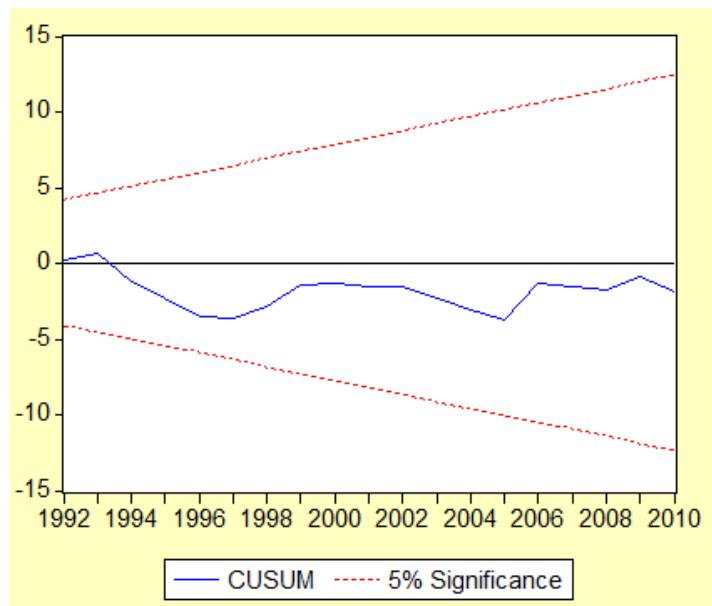
Decimos que al tener F y  $\chi^2$  probabilidades mayores a 5%, debemos rechazar la no linealidad en el modelo, y concluir que sí hay linealidad.

Ramsey RESET Test:				
F-statistic	0.238264	Probability	0.631356	
Log likelihood ratio	0.276151	Probability	0.599235	
Test Equation:				
Dependent Variable: RNACIMIENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 20:57				
Sample: 1990 2010				
Included observations: 21				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002889	0.006844	-0.422089	0.6780
RMATRIMONIOS	0.370595	0.169035	2.192415	0.0417
FITTED^2	16.83288	34.48490	0.488123	0.6314
R-squared	0.227754	Mean dependent var	-0.002355	
Adjusted R-squared	0.141948	S.D. dependent var	0.025544	
S.E. of regression	0.023662	Akaike info criterion	-4.518333	
Sum squared resid	0.010078	Schwarz criterion	-4.369115	
Log likelihood	50.44249	F-statistic	2.654310	
Durbin-Watson stat	1.783373	Prob(F-statistic)	0.097680	



### 3.2.5 Estabilidad de parámetros

Como se puede observar en ambos gráficos, se muestra estabilidad en las muestras con un grado de significancia del 5%.



## Conclusiones

---

Al utilizar como antecedente para esta investigación el documento de la CONAPO “La Situación Demográfica de México, 2013” y con base en las proyecciones de población que muestra el mismo, es importante mencionar que dada la situación de baja calidad en la educación de éste país y como consecuencia la escasez de información sobre planificación familiar, en México, hasta 2013 sigue habiendo un incremento en la población a nivel general.

Como se pudo observar, en las zonas de mayor presencia según condición de habla de lengua indígena es en donde se deben hacer mayores campañas para informar a la población sobre educación sexual y reproductiva sin interferir en la medida de lo posible en los usos y costumbres de los mismos.

El resultado presentado en este trabajo comprende algunos de los indicadores demográficos que se utilizaron como variables para explicar los nacimientos en México ocurridos de 1990 a 2011 fueron (1) tasa de fecundidad la cual nos indica la relación total de nacimientos registrados y la población femenina en edad fértil, (2) tasa de mortalidad que es la relación entre las defunciones registradas en todo el país y la población total, (3) tasa de natalidad que es la relación del número total de nacimientos registrados y la población total, (4) tasa de nupcialidad que indica la relación entre los matrimonios registrados en todo el país y la población total, (5) matrimonios registrados y (6) población total.

Teniendo en cuenta estos indicadores y los resultados arrojados por el modelo, podemos concluir con un 97% de confianza que el factor más importante para que haya nacimientos en México son los matrimonios pues aunque la tasa de fecundidad y la tasa de natalidad han descendido de manera importante, la población en general ha aumentado.

Si bien es cierto que la mayoría de los indicadores presentados muestran un descenso importante en su comportamiento, podemos observar que en general las mujeres esperan a estar casadas o en unión para tener hijos, por lo que se deben seguir realizando políticas y programas públicos para mantener o disminuir los nacimientos pues estamos en una época en la que se deben tener hijos por elección y no por costumbre o falta de información de una comunidad o sociedad a la que se pertenezca.

## Referencias

---

### Fuentes Bibliográficas

- Montgomery, D., Peck, E. y G. Vining., (2011). *Introducción al Análisis de Regresión Lineal*. Tercera edición, quinta reimpression, México, Grupo Editorial Patria.
- Devore, Jay L., (1998). *Probabilidad y Estadística para Ingeniería y Ciencias*. International Thomson Editores. México. ISBN-10: 9706864571.
- Walpole, Ronald E.; Raymond H.; Myers, Sharon L.,(2012). *Probabilidad y Estadística para Ingenieros*. Prentice-Hall Hispanoamericana, S.A. México 2012. ISBN-10: 9701702646.
- Canavos, George C., (1986). *Probabilidad y Estadística. Aplicaciones y Métodos*. McGraw-Hill. México. ISBN-10: 9684518560.
- Abdi, H., (2003). Least-squares.» *M. Lewis-Beck, A. Bryman, T. Futing (Eds): Encyclopedia for research methods for the social sciences. Thousand Oaks (CA): Sage. pp. 792-795.*

### Fuentes Electrónicas

- CONAPO: [En línea]. México, disponible en [http://www.conapo.gob.mx/work/models/CONAPO/Resource/1727/1/images/La\\_Situacion\\_Demografica\\_de\\_Mexico\\_2013\\_COMPLETO.pdf](http://www.conapo.gob.mx/work/models/CONAPO/Resource/1727/1/images/La_Situacion_Demografica_de_Mexico_2013_COMPLETO.pdf) [Accesado el día 17 de febrero de 2015].
- INEGI: [En línea]. México, disponible en <http://www3.inegi.org.mx/Sistemas/temasV2/Default.aspx?s=est&c=17484>. [Accesado el día 22 de junio de 2013].
- Análisis de Regresión Lineal Múltiple: [En línea]. <http://rua.ua.es/dspace/bitstream/10045/8143/1/Regresion%20MUTIPLE.pdf> [Accesado el día 2 de Noviembre de 2014]
- Guía de manejo del programa E-Views: [En línea]. [http://www.uam.es/personal\\_pdi/economicas/anadelsur/pdf/eviews\\_II.PDF](http://www.uam.es/personal_pdi/economicas/anadelsur/pdf/eviews_II.PDF)
- Eviews es una marca registrada de QMS (Quantitative Micro Software). Toda la información sobre licencias, modalidades de adquisición, actualizaciones, etc. Puede encontrarse en [www.Eviews.com](http://www.Eviews.com)
- EasyFit permite ajustar automáticamente las distribuciones a los datos de la muestra y seleccionar el mejor modelo. Puede encontrarse en <http://www.mathwave.com/es/home.html>

## Anexos

### A. Datos

obs	FECUNDID...	MATRIMO...	MORTALID...	NACIMIEN...	NATALIDAD	NUPCIALI...	POBLACION
1990	3.400000	642201.0	422803.0	2735312.	27.90000	7.376123	87064847
1991	3.300000	652172.0	411131.0	2756447.	27.50000	7.358288	88630941
1992	3.200000	667598.0	409814.0	2797397.	27.10000	7.406844	90132585
1993	3.100000	659567.0	416335.0	2839686.	26.80000	7.200462	91600655
1994	3.000000	671640.0	419074.0	2904389.	26.30000	7.217644	93055300
1995	3.000000	658114.0	430278.0	2750444.	25.90000	6.964882	94490336
1996	2.900000	670523.0	436321.0	2707718.	25.40000	6.993600	95876664
1997	2.800000	707840.0	440437.0	2698425.	24.80000	7.281960	97204604
1998	2.800000	704456.0	444665.0	2668428.	24.30000	7.152896	98485424
1999	2.700000	743856.0	443950.0	2769089.	23.90000	7.460489	99706067
2000	2.600000	707422.0	437667.0	2798339.	23.40000	7.011411	1.01E+08
2001	2.600000	665434.0	443127.0	2767610.	23.00000	6.516050	1.02E+08
2002	2.600000	616654.0	459687.0	2699084.	22.60000	5.962737	1.03E+08
2003	2.500000	584142.0	472140.0	2655894.	22.20000	5.578138	1.05E+08
2004	2.500000	600563.0	473417.0	2625056.	21.80000	5.668278	1.06E+08
2005	2.500000	595713.0	495240.0	2567906.	21.50000	5.559565	1.07E+08
2006	2.400000	586978.0	494471.0	2505939.	21.10000	5.414485	1.08E+08
2007	2.400000	595209.0	514420.0	2655083.	20.80000	5.421470	1.10E+08
2008	2.300000	589352.0	539530.0	2636110.	20.40000	5.295213	1.11E+08
2009	2.300000	558913.0	564673.0	2577214.	20.10000	4.952593	1.13E+08
2010	2.300000	568632.0	592018.0	2643908.	19.70000	4.975369	1.14E+08
2011	2.300000	570954.0	590693.0	2586287.	19.40000	4.938265	1.16E+08

Los datos fueron extraídos de la página de internet del Instituto Nacional de Estadística y Geografía (INEGI) en junio de 2013.

La siguiente tabla muestra los datos de forma diferenciada para poder realizar la regresión.

obs	RFECUNDI...	RMATRIM...	RMORTALI...	RNACIMIE...	RNATALIDAD	RNUPCIAL...	RPOBLACI...
1990	-0.029412	0.015526	-0.027606	0.007727	-0.014337	-0.002418	0.017988
1991	-0.030303	0.023653	-0.003203	0.014856	-0.014545	0.006599	0.016943
1992	-0.031250	-0.012030	0.015912	0.015117	-0.011070	-0.027864	0.016288
1993	-0.032258	0.018304	0.006579	0.022785	-0.018657	0.002386	0.015880
1994	0.000000	-0.020139	0.026735	-0.053004	-0.015209	-0.035020	0.015421
1995	-0.033333	0.018855	0.014044	-0.015534	-0.019305	0.004123	0.014672
1996	-0.034483	0.055654	0.009433	-0.003432	-0.023622	0.041232	0.013851
1997	0.000000	-0.004781	0.009600	-0.011116	-0.020161	-0.017724	0.013177
1998	-0.035714	0.055930	-0.001608	0.037723	-0.016461	0.043003	0.012394
1999	-0.037037	-0.048980	-0.014152	0.010563	-0.020921	-0.060194	0.011933
2000	0.000000	-0.059354	0.012475	-0.010981	-0.017094	-0.070651	0.012156
2001	0.000000	-0.073306	0.037371	-0.024760	-0.017391	-0.084915	0.012687
2002	-0.038462	-0.052723	0.027090	-0.016002	-0.017699	-0.064500	0.012589
2003	0.000000	0.028111	0.002705	-0.011611	-0.018018	0.016160	0.011762
2004	0.000000	-0.008076	0.046097	-0.021771	-0.013761	-0.019179	0.011321
2005	-0.040000	-0.014663	-0.001553	-0.024131	-0.018605	-0.026095	0.011739
2006	0.000000	0.014023	0.040344	0.059516	-0.014218	0.001290	0.012716
2007	-0.041667	-0.009840	0.048812	-0.007146	-0.019231	-0.023288	0.013769
2008	0.000000	-0.051648	0.046602	-0.022342	-0.014706	-0.064704	0.013959
2009	0.000000	0.017389	0.048426	0.025878	-0.019900	0.004599	0.012432
2010	0.000000	0.004083	-0.002238	-0.021794	-0.015228	-0.007458	0.012492
2011	NA	NA	NA	NA	NA	NA	NA

Elaboración propia con el programa E-Views.

Para la obtención de estos datos se utilizó la fórmula  $\frac{X_t - X_{t-1}}{X_t}$ .

## B. Pruebas de Raíz Unitaria

Se realizan las pruebas de raíz unitaria para cada una de las variables estacionarias tanto de la que se va a explicar como de las que la explican, se muestran los cuadros de las pruebas que son significativas para la regresión, esto lo determina la probabilidad cuando sea menor a 0.1 en cada una de ellas.

- ADF Rnacimientos Intercepto

Null Hypothesis: RNACIMIENTOS has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.553593	0.0020

- ADF Rnacimientos Tendencia e Intercepto

Null Hypothesis: RNACIMIENTOS has a unit root		
Exogenous: Constant, Linear Trend		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.440987	0.0112

- ADF Rnacimientos Ninguno

Null Hypothesis: RNACIMIENTOS has a unit root		
Exogenous: None		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.625794	0.0001

- ADF Rmatrimonios Intercepto

Null Hypothesis: RMATRIMONIOS has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.503950	0.0189

- ADF Rmatrimonios Tendencia e Intercepto

Null Hypothesis: RMATRIMONIOS has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.493231	0.0675

- ADF Rmatrimonios Ninguno

Null Hypothesis: RMATRIMONIOS has a unit root Exogenous: None Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.516432	0.0013

- ADF Rfecundidad Intercepto

Null Hypothesis: RFECUNDIDAD has a unit root Exogenous: Constant Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.319676	0.0004

- ADF Rfecundidad Tendencia e Intercepto

Null Hypothesis: RFECUNDIDAD has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.406006	0.0002

- ADF Rfecundidad Ninguno

Null Hypothesis: RFECUNDIDAD has a unit root Exogenous: None Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.995445	0.0048

- ADF Rmortalidad Intercepto

Null Hypothesis: RMORTALIDAD has a unit root Exogenous: Constant Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.657651	0.0137

- ADF Rmortalidad Tendencia e Intercepto

Null Hypothesis: RMORTALIDAD has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.999754	0.0264

- ADF Rmortalidad Ninguno

Null Hypothesis: RMORTALIDAD has a unit root Exogenous: None Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.379727	0.0202

- ADF Rnupcialidad Intercepto

Null Hypothesis: RNUPCIALIDAD has a unit root Exogenous: Constant Lag Length: 0 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.562185	0.0168

- ADF Rnupcialidad Tendencia e Intercepto

Null Hypothesis: RNUPCIALIDAD has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.525046	0.0637

- ADF Rnupcialidad Ninguno

Null Hypothesis: RNUPCIALIDAD has a unit root		
Exogenous: None		
Lag Length: 0 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.980199	0.0050

A continuación se muestra una tabla con cada una de las variables y de los valores arrojados por el programa E-Views los cuales nos indican cuales se toman en cuenta y cuales se descartan.

Variable	Intercepto	Tendencia Intercepto	Ninguno	t-Statistic	Prob.
Nacimientos				-1.436992	0.5447
				-2.954909	0.1669
				-0.523388	0.4778
RNacimientos				-4.553593	0.0020
				-4.440987	0.0112
				-4.625794	0.0001
Matrimonios				-0.616925	0.8467
				-1.888019	0.6249
				-0.693689	0.4042
RMatrimonios				-3.503950	0.0189
				-3.493231	0.0675
				-3.516432	0.0013
Población				-0.817973	0.7931
				-1.986899	0.5745
				34.96914	0.9999
RPoblación				-2.818156	0.0735
				-1.649841	0.7353
				-2.193900	0.0304
Fecundidad				-2.935616	0.0581
				-1.707839	0.7115
				-5.311174	0.0000
RFecundidad				-5.319676	0.0004
				-6.466006	0.0002
				-2.995445	0.0048
Mortalidad				2.223534	0.9998
				-1.061599	0.9120
				3.648945	0.9997
RMortalidad				-3.657651	0.0137
				-3.999754	0.0264
				-2.379727	0.0202



Variable	Intercepto	Tendencia Intercepto	Ninguno	t-Statistic	Prob.
Nupcialidad				-0.056657	0.9424
				-1.884261	0.6268
				-2.251756	0.0266
RNupcialidad				-3.562185	0.0168
				-3.525046	0.0637
				-2.980199	0.0050
Natalidad				-2.126694	0.2369
				0.091662	0.9948
				-25.51592	0.0001
RNatalidad				-4.288287	0.0036
				-4.121636	0.0209
				-0.470721	0.4986

En la tabla anterior, se muestra que las únicas variables explicativas que se toman en cuenta para la regresión son RMatrimonios, RFecundidad, RMortalidad y RNupcialidad al cumplirse el supuesto de que la probabilidad sea menor a 0.1.

### C. Regresiones

A partir de las pruebas de raíz unitarias realizadas en el anexo anterior, se hace la combinación lineal de cada una de ellas para saber si son significativas para la regresión.

#### Regresiones de una variable explicativa.

- rnacimientos – rmatrimonios

Dependent Variable: RNACIMIENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 20:30				
Sample (adjusted): 1990 2010				
Included observations: 21 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000725	0.005108	-0.141903	0.8887
RMATRIMONIOS	0.329167	0.143223	2.298287	0.0331
R-squared	0.217531	Mean dependent var		-0.002355
Adjusted R-squared	0.176349	S.D. dependent var		0.025544
S.E. of regression	0.023183	Akaike info criterion		-4.600421
Sum squared resid	0.010211	Schwarz criterion		-4.500942
Log likelihood	50.30442	F-statistic		5.282123
Durbin-Watson stat	1.782626	Prob(F-statistic)		0.033068

Se observa que la variable explicativa RMatrimonios es significativa pues su probabilidad es menor a 0.1 por lo que sí se toma en cuenta para la regresión.

Los valores resaltados son los que se utilizan en la fórmula de regresión.

- nacimientos – rfecundidad

Dependent Variable: RNACIMENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 20:33				
Sample (adjusted): 1990 2010				
Included observations: 21 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.007978	0.008035	-0.992877	0.3332
RFECUNDIDAD	-0.307574	0.316182	-0.972774	0.3429
R-squared	0.047442	Mean dependent var	-0.002355	
Adjusted R-squared	-0.002693	S.D. dependent var	0.025544	
S.E. of regression	0.025579	Akaike info criterion	-4.403724	
Sum squared resid	0.012431	Schwarz criterion	-4.304245	
Log likelihood	48.23910	F-statistic	0.946290	
Durbin-Watson stat	2.295082	Prob(F-statistic)	0.342888	

Se observa que la variable explicativa Rfecundidad no es significativa pues su probabilidad es mayor a 0.1 por lo que no se toma en cuenta para la regresión.

- nacimientos - rmortalidad

Dependent Variable: RNACIMENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 20:35				
Sample (adjusted): 1990 2010				
Included observations: 21 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000302	0.007125	-0.042403	0.9666
RMORTALIDAD	-0.126117	0.263803	-0.478073	0.6381
R-squared	0.011886	Mean dependent var	-0.002355	
Adjusted R-squared	-0.040120	S.D. dependent var	0.025544	
S.E. of regression	0.026052	Akaike info criterion	-4.367077	
Sum squared resid	0.012895	Schwarz criterion	-4.267598	
Log likelihood	47.85431	F-statistic	0.228554	
Durbin-Watson stat	2.201332	Prob(F-statistic)	0.638053	

Se observa que la variable explicativa Rmortalidad no es significativa pues su probabilidad es mayor a 0.1 por lo que no se toma en cuenta para la regresión.

- Rnacimientos – rnpucialidad

Dependent Variable: RNACIMIENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 20:37				
Sample (adjusted): 1990 2010				
Included observations: 21 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003786	0.005736	0.660016	0.5172
RNUPCIALIDAD	0.335284	0.146836	2.283388	0.0341
R-squared	0.215325	Mean dependent var		-0.002355
Adjusted R-squared	0.174027	S.D. dependent var		0.025544
S.E. of regression	0.023215	Akaike info criterion		-4.597606
Sum squared resid	0.010240	Schwarz criterion		-4.498127
Log likelihood	50.27486	F-statistic		5.213861
Durbin-Watson stat	1.775538	Prob(F-statistic)		0.034092

Se observa que la variable explicativa Rnupcialidad es significativa pues su probabilidad es menor a 0.1 por lo que sí se toma en cuenta para la regresión.

A partir de las regresiones que realizamos, se observa que la siguiente regresión que debemos hacer es sobre las dos variables que si son significativas y que no afectan de manera independiente a la que se va a explicar

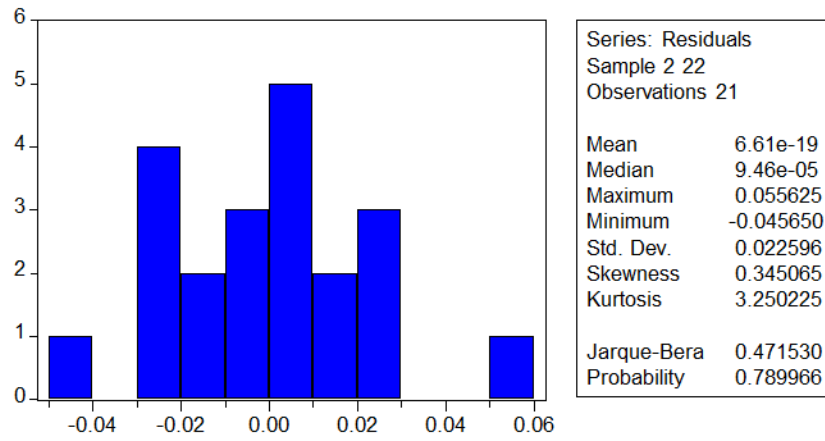
### Regresiones de dos variables explicativas.

Dependent Variable: RNACIEMENTOS				
Method: Least Squares				
Date: 10/07/15 Time: 20:40				
Sample (adjusted): 1990 2010				
Included observations: 21 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.007123	0.039677	-0.179519	0.8595
RNUPCIALIDAD	-0.474485	2.916741	-0.162677	0.8726
RMATRIMONIOS	0.792009	2.848969	0.277999	0.7842
R-squared	0.218680	Mean dependent var		-0.002355
Adjusted R-squared	0.131867	S.D. dependent var		0.025544
S.E. of regression	0.023801	Akaike info criterion		-4.506652
Sum squared resid	0.010196	Schwarz criterion		-4.357434
Log likelihood	50.31984	F-statistic		2.518969
Durbin-Watson stat	1.793256	Prob(F-statistic)		0.108508

Aquí se puede observar que la variable explicativa Rnupcialidad no es significativa, pues la probabilidad es mayor a 0.1 y por tanto afecta a la variable explicativa RMatrimonios y la vuelve también no significativa por lo que se descarta de la regresión.

Por lo que, para este trabajo, se toman en cuenta sólo la variable que se explicará (rnacimientos) y la explicativa (rmatrimonios).

#### D. Prueba de Normalidad



Elaboración propia usando el programa E-Views  
con los datos de la regresión.

Como se puede observar, la regresión cumple con el supuesto de normalidad.