



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**ESTADÍSTICA PARAMÉTRICA Y NO  
PARAMÉTRICA: ANÁLISIS DE LA  
INDEPENDENCIA Y DEL SUPUESTO DE  
NORMALIDAD EN LAS ENCUESTAS NACIONALES  
A HOGARES RURALES EN MÉXICO 2002 Y 2007**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**LICENCIADO EN ACTUARIA**

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

**ARTURO ISAÍN CISNEROS YESCAS**



**ASESOR: M.EN.C JOSÉ ANTONIO FLORES  
DÍAZ**

**2012**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

1. Datos del alumno  
Cisneros  
Yescas  
Arturo Isaín  
22324103  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
305263801
2. Datos del asesor  
M. en. C  
José Antonio  
Flores  
Díaz
3. Datos del sinodal 1  
Dra.  
María del Pilar  
Alonso  
Reyes
4. Datos del sinodal 2  
M. en I. O.  
María Isabel  
Escalante  
Membrillo
5. Datos del sinodal 3  
Act.  
Francisco  
Sánchez  
Villareal
6. Datos del sinodal 4  
Act.  
Jaime  
Vázquez  
Alamilla
7. Datos del trabajo escrito  
Estadística paramétrica y no paramétrica: Análisis de la independencia y del supuesto de normalidad en las Encuestas Nacionales a Hogares Rurales en México 2002 y 2007.  
92p.  
2012

## ÍNDICE

<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>CAPÍTULO 1</b>	
<b>LOS ESTUDIOS LONGITUDINALES .....</b>	<b>4</b>
Las encuestas aplicadas en distintas áreas en específico en el sector rural.....	6
Encuestas agrícolas .....	6
Encuestas sociodemográficas enfocadas al campo .....	8
Encuestas de empleo y desempleo enfocadas al campo.....	9
Las encuestas longitudinales .....	11
Los diseños tipo panel.....	12
Clasificación de los diseños tipo panel.....	13
Aspectos fundamentales de un diseño tipo panel .....	15
Ventajas de los diseños tipo panel .....	16
Desventajas de los diseños tipo panel .....	17
<b>CAPÍTULO 2</b>	
<b>LA ENCUESTA NACIONAL A HOGARES RURARLES EN MÉXICO ( ENHRUM I Y LA ENHRUM II).....</b>	<b>19</b>
La ENHRUM I.....	20
Marco conceptual.....	21
Cobertura .....	22
Secciones.....	23
La ENHRUM II.....	24
Cobertura .....	24
Secciones.....	25
Descripción de las secciones del cuestionario .....	25
La ENRHUM I y la ENHRUM II como encuestas tipo panel.....	28
Diferencia entre la ENHRUM I y la ENHRUM II .....	29
Cantidad de hogares sobre el tiempo.....	30
Cantidad de individuos sobre el tiempo.....	30

### **CAPÍTULO 3**

<b>ANÁLISIS ESTADÍSTICO.....</b>	<b>32</b>
Tratamiento de la información .....	34
La prueba de Kolmogorov Smirnov .....	37
Análisis del supuesto de normalidad para la característica de la:	
Migración a los Estados Unidos .....	38
Migración femenina a los Estados Unidos .....	41
Migración masculina a los Estados Unidos.....	43
Migración dentro territorio mexicano .....	45
Migración femenina dentro del territorio mexicano .....	47
Migración masculina dentro del territorio mexicano .....	49

### **CAPÍTULO 4**

<b>APLICACIÓN DE LAS PRUEBAS ESTADISTISTCAS .....</b>	<b>52</b>
La prueba de Suma de Rangos de Wilcoxon .....	52
Aplicación de la prueba de Wilcoxon para la característica de la:	
Migración a los Estados Unidos .....	54
Migración femenina a los Estados Unidos .....	55
Migración masculina a los Estados Unidos.....	56
Migración dentro del territorio mexicano I .....	57
Migración femenina dentro del territorio mexicano .....	58
Migración masculina dentro del territorio mexicano .....	59
La prueba T para muestras apareadas .....	60
Aplicación de la prueba de T para la característica de la:	
Migración a los Estados Unidos.....	62
Migración femenina a los Estados Unidos .....	62
Migración masculina a los Estados Unidos.....	63
Migración dentro del territorio mexicano .....	64
Migración femenina dentro del territorio mexicano .....	64
Migración masculina dentro del territorio mexicano .....	65

### **CAPÍTULO 5**

<b>ANÁLISIS DE LA INDEPENDENCIA DE LAS MUESTRAS .....</b>	<b>66</b>
Prueba de Man Whitney Wilcoxon .....	67
Aplicación de la prueba de Mann Whitney para la característica de la:	
Migración a los Estados Unidos.....	69
Migración femenina a los Estados Unidos .....	70
Migración masculina a los Estados Unidos.....	71

Migración dentro del territorio mexicano .....	72
Migración femenina dentro del territorio mexicano .....	73
Migración masculina dentro del territorio mexicano .....	73

**CONCLUSIONES**

Conclusiones.....	75
-------------------	----

**ANEXO**

Anexo I .....	78
---------------	----

**BIBLIOGRAFÍA**

Bibliografía .....	90
--------------------	----

## AGRADECIMIENTOS

Con el término de este proyecto culmina una de las maravillosas etapas de mi vida, la cual me llena de satisfacción y orgullo, sin embargo hay demasiadas personas a las que tengo que agradecer y no hay forma de pagar todo ese esfuerzo que junto conmigo hicieron y me apoyaron a lo largo de estos diecinueve años de estudio.

Antes que nada quiero agradecerle a dos seres maravillosos que a lo largo de mi vida me han dado muchas enseñanzas y que sin su apoyo no hubiera logrado muchas de las cosas que me he propuesto. Le agradezco a mis papas que gracias a su apoyo, consejos, impulsándome en cada proyecto que emprendía. Quiero agradecerle a mi papá por sus grandes enseñanzas, grandes platicas y consejos, a mi mamá que siempre estuvo presente en todo momento y más que nada en mi época estudiantil, no hay palabras para describir lo que ellos hicieron por mi y su apoyo que me brindaron durante este tiempo, gracias mamá y papá por brindarme y enseñarme los valores más importantes de esta vida. Padres muchas gracias por enseñarme a nunca rendirme y siempre dar todo de mi por mas insignificante que fuera la situación además de enseñarme que todo tiene su recompensa.

Agradezco a mis dos hermanos un gran apoyo en esta vida, siempre dándome consejos. Mi hermano Gerardo una gran persona que siempre trato de guiarme por el mejor camino y nunca me negó la ayuda cuando mas la necesite. Mi hermano Edgar una persona con grandes virtudes que sin su apoyo en diversas ocasiones no hubiera podido salir a flote. Agradezco a ellos dos, personas dignas de admirar que siempre me impulsaron a ser mejor y nunca rendirme.

Agradezco a CONACYT (Consejo nacional de la ciencia y la tecnología) por el auspicio que me otorgo por la realización de este proyecto, sin ellos no hubiera podido ser posible la culminación de este trabajo.

Agradezco al COLMEX en especial al equipo de PRECESAM (Programa de Estudios del Cambio Económico y la Sustentabilidad del Agro Mexicano) por brindarme la oportunidad de participar con ellos y poder desarrollar este proyecto juntos, por su comprensión y entendimiento ante momentos y situaciones difíciles y por las charlas con respecto a este trabajo.

Agradezco a mi asesor José Antonio Flores Díaz por guiarme a lo largo de este proyecto y por esas pláticas de preocupación con respecto al futuro de dicho trabajo, por su ayuda en los momentos más complicados y por brindarme ese apoyo y esa confianza necesaria.

Agradezco a todos mis amigos de la facultad por su apoyo a lo largo de la carrera, por los trabajos en equipo, los momentos de angustia y los momentos de diversión, y por supuesto a mis amigos de toda la vida que sin ellos no hubiera logrado muchas cosas Fernanda, Liliana, Gabriela Michel Roberto y Hugo, .Del mismo modo quiero agradecerle a una persona en especial que siempre estuvo presente durante este proyecto, que escucho mis preocupaciones, mis temores que nunca dejó que me diera por vencido por más difícil que pareciera, por momentos de charla donde todo se centraba en mi tesis y esas preocupaciones.

## INTRODUCCIÓN

La investigación social hoy en día es un instrumento fundamental para entender o analizar los múltiples fenómenos que ocurren en el seno de las poblaciones de cualquier parte del mundo.

Para realizar la investigación social, es necesario recurrir a múltiples herramientas, se requieren labores demográficas, económicas, políticas, entre otras muchas. Sin embargo un elemento fundamental para las teorías antes señaladas y por ende para la investigación social, es el análisis de la información o la estadística, la cual permite corroborar ciertas características posibles o condiciones, que aunque son inherentes a los datos cuantitativos o cualitativos, éstos están íntimamente ligados o vinculados a los fenómenos que se desarrollan en el núcleo de las poblaciones.

En México la Encuesta Nacional de Hogares Rurales de México (ENHRUM) es un proyecto enfocado al sector rural, el cual es realizado conjuntamente por el Colegio de México y la Universidad de Davis en California. Existen tres etapas de esta encuesta las cuales son ENHRUM I, ENHRUM II y la ENHRUM II.5.

El análisis que aquí se presenta está relacionado con el fenómeno de la migración rural, que es muy complejo y que requiere para su comprensión un estudio a fondo, el cual aquí se bosqueja o ejemplifica, debido a que son múltiples las razones que la originan y han originado el abandono del campo.

En todo caso se plantea una primera fase, que desde el punto de vista estadístico resulta interesante pues se consideran revisar si las muestras o conjuntos de datos ya recolectados tienen ciertas cualidades que posteriormente permitan construir o elaborar hipótesis así como suposiciones más completas sobre el fenómeno en cuestión, el cual se insiste no se desarrollará aquí.

En este trabajo se utilizó solamente la ENHRUM I y ENHRUM II con las cuales se realizó el análisis del supuesto de normalidad para aplicar estadística paramétrica. Asimismo se estudió la independencia de las muestras entre sí cuando se presenta una encuesta con el diseño tipo panel, por ello en este trabajo se

propone averiguar que si la normalidad y la independencia de las muestras, son consideraciones o supuestos que se pueden contemplar cuando se aplican ciertas exploraciones estadísticas.

Sólo se consideraron variables de la sección 2 de sociodemografía, donde se incluyeron todos los integrantes del hogar y en específico los que migraron.

Se abordó la estadística paramétrica donde se explicó en qué consiste y los supuestos con los que deben de cumplir los datos. Asimismo, se dio a conocer qué es la estadística no paramétrica y algunas de sus ventajas y desventajas.

Se consideraron cuatro pruebas, tres de ellas no paramétricas y una paramétrica. De las no paramétricas se utilizaron la de **Kolmogorov Smirnov** para verificar la existencia de normalidad de los datos. La segunda que se aplicó es la **Suma de rangos de Wilcoxon** para comparar la mediana de dos muestras dependientes. La tercera es la prueba de **Mann Whitney** que funciona igual que la antes, señalada pero para muestras independientes. Finalmente por el lado de la estadística paramétrica se utilizó la **Prueba T** la cual es equivalente a las dos últimas.

Se planteó este trabajo como un ejercicio estadístico donde se vio cuál es el significado de que no se cumplan ciertos supuestos en los datos. Del mismo modo se analizaron cuál es el efecto que tiene el planteamiento de estos casos y cómo resolverlo ¿será que la estadística paramétrica con la violación de un supuesto sea una opción para aplicar o realmente se tendrá que acudir a la estadística no paramétrica?

Asimismo se planteó qué diferencia hay entre utilizar pruebas no paramétricas del mismo tipo donde lo único que las hace distintas es la independencia de las muestras. Es claro que cada prueba está hecha con sus supuestos y sus criterios específicos, pero lo que se pretendió realizar con este trabajo es ir más allá de esas restricciones para ver qué tanto difieren los resultados.

El contenido de esta tesis es el siguiente. En el primer capítulo se describirán los estudios longitudinales explicando sus características ventajas y desventajas. Las

encuestas enfocadas en diversas áreas donde se habla de cuáles son las consideraciones que se deben de asumir para cada tipo de encuesta así como las características de éstas y los tipos de diseños longitudinales que existen, señalando las particularidades más importantes de éstos y en específico del diseño tipo panel, el cual es el que más interesa para este estudio.

En el segundo capítulo se abordó la ENHRUM etapa I y II donde se describe cada una de las dos encuestas, tomando en cuenta la cobertura y el contenido de ellas, así como los cambios que hubo del 2002 al 2007 y la formación del panel.

En el tercer capítulo se reporta el análisis estadístico que se aplicó así como la explicación y la diferencia que existe entre la estadística paramétrica y no paramétrica, del mismo modo se da a conocer cuáles son las variables que se tomaron para comprender el estatus migratorio y poder realizar el análisis.

En el capítulo cuatro se abordan dos pruebas, una paramétrica (prueba T) y otra no paramétrica (Rangos de Wilcoxon). Además, se analizan los resultados arrojados por cada una de ellas, tomando en cuenta que el supuesto de normalidad no se cumple.

El capítulo cinco versó sobre la independencia de las muestras, donde se da una explicación de por qué en un panel no se puede hablar de muestras independientes. Asimismo, se aplicó la prueba no paramétrica para muestras independientes este hecho fue para verificar si realmente modifica los resultados o se sigue con la misma tendencia.

Finalmente, en el último apartado se concluye con los resultados y las implicaciones que se desprendieron de ellos.

# CAPÍTULO I

## LOS ESTUDIOS LONGITUDINALES

Los estudios longitudinales tienen el objetivo de mostrar no sólo los perfiles individuales de las personas, sino determinar si hay cambios en ellos a lo largo del tiempo y si éstos son significativos entre los distintos elementos de la muestra.

Esta clase de trabajos están compuestos por dos tipos de estudios: los prospectivos y los retrospectivos. Los primeros se caracterizan por recolectar la información de acuerdo a los criterios del investigador y con los fines establecidos de la investigación. Por el contrario, en los segundos los datos se obtuvieron con anterioridad a su planeación, estableciéndose con fines ajenos al trabajo de investigación que se pretende realizar.

Ahora, si ambos se realizan en varias ocasiones, se puede decir que es una situación de medidas repetidas donde las observaciones son seleccionadas a través del tiempo. Así, los sujetos son medidos en diferentes ocasiones con el propósito de conseguir la curva continua del cambio sobre el tiempo, esto sirve para tener comparaciones de las variables medidas en más de un par de ocasiones y con ello analizar las características de un proceso de cambio. En este caso se debe observar la evolución a través de una serie de escenarios diferentes<sup>1</sup>.

Algunas de las dificultades de este tipo de estudios son:

- El análisis suele ser más complejo debido a la dependencia que se da entre las medidas repetidas de la misma unidad observacional.
- El investigador no puede controlar las circunstancias bajo las que obtiene las medidas repetidas, de modo que los datos son incompletos.

Ahora bien, ante cualquier estudio de carácter longitudinal es posible plantear cuestiones básicas que tienen, sin duda, una gran importancia desde el punto de

---

<sup>1</sup>Cfr: Jaume Arnau et al, *Estudios Longitudinales Modelos de Diseño y Análisis*. P. 33

vista aplicado, el cambio de la muestra a través del tiempo, es decir, cual es la forma del cambio intraindividual en función del tiempo, así como las diferencias interindividuales en los procesos de cambio y si existe alguna predicción o explicación de las diferencias en relación a los cambios observados.

Teniendo en cuenta estos factores, ahora se puede tratar el tema de las encuestas longitudinales, pero, antes de pasar a ello, hay que definir qué es una encuesta y cuáles son sus tipos:

Una *encuesta* es un método por el cual se obtiene información de una muestra o de una población de individuos, la primera es sólo una parte de la población y su tamaño depende de las metas estadísticas y de los recursos disponibles. La segunda se aplica en situaciones especiales que no provoquen el uso de recursos financieros en exceso o demasiado tiempo. Las encuestas proveen una fuente importante de conocimiento básico. Economistas, psicólogos, profesionales de la salud y sociólogos llevan a cabo encuestas para estudiar fenómenos diversos tales como los patrones de ingreso y gastos en los hogares, las raíces del prejuicio étnico o racial, las implicaciones de los problemas de salud en la vida de las personas, etcétera, lo cual, resulta útil para comparar diversos aspectos tanto sociales como económicos o demográficos, ente otros<sup>2</sup>.

El objetivo de una encuesta no es intentar describir la situación particular de un individuo, sino obtener un perfil estadístico de la población.

Las encuestas pueden ser clasificadas de varias formas:

- Por el tamaño de la muestra.
- Por el tipo de población en la que se concentran.
- Por el método bajo el que se lleva a cabo.
- Por el área en la cual se aplican.

Este trabajo se enfocará en esta última categoría, ya que es la que permite analizar la encuesta seleccionada en el.

---

<sup>2</sup> Ferber Robert, et al, *What is A Survey?*

## **LAS ENCUESTAS APLICADAS EN DISTINTAS ÁREAS EN ESPECÍFICO EN EL SECTOR RURAL**

Una encuesta es compleja cuando se tiene que enfocar a distintos campos, es decir, cuando debe de considerar a una población específica, ésta puede encontrarse en la agricultura, la industria, el transporte, o en diversos lugares. Esta labor resulta compleja porque la información se tiene que recolectar considerando las características de lo que se quiere estudiar.

A continuación se señalarán algunos ejemplos de estos tipos de encuestas, sus enfoques, así como las ventajas y desventajas que tienen, poniendo énfasis especial en el sector rural que es el que interesa, donde se destacan distintos aspectos como lo agrícola, lo laboral y lo sociodemográfico, ya que de ahí partirá el análisis posterior en esta investigación.

### ***ENCUESTAS AGRÍCOLAS***

Se puede obtener una idea de la estructura agrícola de un país subdividiendo su área total de acuerdo con la utilización de la tierra. Los métodos de recolección de datos sobre esta área se concentran en dos tipos:

- Se pide la información a los agricultores del país acerca de la tierra que trabajan.
- Se visita cada pedazo de tierra del país para determinar su utilización.

En el primer caso se emplea el método de la entrevista, donde el encuestador cuestiona al agricultor sobre ciertas características. La entrevista se realiza, de preferencia, en el terreno para que se pueda verificar la información en caso necesario. La ventaja de recolectar datos para cada predio permite tabular la información por unidad administrativa, grandes o pequeñas. Otra ventaja es que los datos pueden someterse a clasificaciones cruzadas a cualquier nivel, permitiendo el estudio de las relaciones existentes entre las variables.

Sus desventajas son que:

- Los datos no representan la imagen de toda la tierra de un país.

- Los datos sobre áreas recolectadas por el método de la entrevista pueden estar considerablemente errados, ya que es posible que no se revelen algunas parcelas del predio.
- Es necesario hacer este tipo de encuestas con frecuencia, ya que la situación agrícola de un país cambia cada año, provocando que los agricultores puedan empezar a dedicar una mayor área a cierto cultivo o abandonar algún otro cultivo, es por eso que resulta importante la recolección de estadísticas cada año.
- Puede darse el caso de que haya poca cooperación de parte del entrevistado (el agricultor), quien muchas veces ignora la superficie real del área que trabaja y muy a menudo por temor a los impuestos o por otra causa, se resiste a dar la información correcta de toda la tierra que trabaja.
- Los predios varían mucho respecto al área total y en relación con las áreas en las que se realizan diversos cultivos.
- El resultado es que los errores muestrales tienden a ser demasiado grandes cuando el tamaño de la muestra es razonablemente pequeño.

Ahora, si se considera el problema de la estimación de áreas para diversos cultivos por métodos de muestreo, los cuales se ajustan siempre a las condiciones existentes de cada lugar. En la mayoría, se realiza una lista de localidades y se estratifica de acuerdo a la información disponible. Para cada estrato se toma una muestra de localidades con probabilidad proporcional a alguna medida del tamaño de la muestra.

Otro tipo de muestreo es el realizado por áreas administrativas, en éste se hace una lista de las familias de la localidad y se toma una muestra de ella. Se miden las parcelas trabajadas por la familia de la muestra y se determina el área para cada cultivo. En realidad no existe ningún método de estimación de áreas que sea el mejor, este depende mucho de las condiciones locales.

El problema de la estimación de la producción anual de cultivos alimenticios y otros productos agrícolas tiene gran importancia para los países en desarrollo que les interesa elevar sus niveles de vida. Asimismo un método de estimación de la

producción total de tales cultivos consiste en el área total para ellos y la tasa de rendimiento.

El método de estimación de la tasa de rendimiento consiste en seleccionar una muestra de parcelas o campos al azar y emplearlos en las observaciones. Ahora, con respecto a la medición física del rendimiento en el momento de la cosecha donde las unidades muestrales son terrenos de una dimensión prescrita situados y marcados en la parcela.

Las encuestas por muestreo de rendimiento agrícola se realizan generalmente cultivo por cultivo. La razón es que algunos de ellos maduran en épocas diferentes, por lo que la muestra de un cultivo en particular sólo puede tomarse cuando la cosecha está madura, esto quiere decir que sólo hay un periodo muy breve anterior a la cosecha efectiva en que el campo esta listo para el muestreo.

#### ***ENCUESTAS SOCIODEMOGRÁFICAS ENFOCADAS AL CAMPO***

En la mayoría de las encuestas es necesario tener información sociodemográfica donde se utiliza al hogar como unidad muestral, éste puede constar de un miembro o de varios, pero, siempre conviene hacer la definición más flexible e incluir información sobre todas las personas pertenecientes a él. Las formas de clasificación de los miembros del hogar al momento del levantamiento de la encuesta son las siguientes:

- Residentes presentes en el momento de la encuesta,
- Residentes pero ausentes en el momento de la encuesta,
- Residentes temporales.

La mayoría de las tabulaciones demográficas se subdividen por edad y sexo. Se define la edad como el intervalo existente entre la fecha de nacimiento y la de la encuesta. La experiencia demuestra que los informes sobre dicha característica están sujetos a errores considerables porque muchas personas mienten acerca de ella.

Ahora, la composición de una familia se determina a partir del estado civil de cada miembro y de la relación con el jefe de la casa. La información sobre el estado civil es vital para los estudios de estructura de la población, niveles de fecundidad, tendencias y es definido en cuatro categorías:

- Solteros
- Casados
- Viudos
- Divorciados

### ***ENCUESTAS DE EMPLEO Y DESEMPLEO ENFOCADAS AL CAMPO***

Las encuestas de empleo y desempleo sirven para medir o definir el status de la fuerza de trabajo de un país. Estos datos tienen una función muy útil como indicadores económicos fundamentales, pero la recolección de ellos plantea problemas muy difíciles.

Lo primero que se debe de considerar es definir a la población objetivo a la que se referirán las estadísticas de empleo. Uno de los propósitos de esta encuesta es la determinación del porcentaje de la población que se encuentra en la fuerza de trabajo, es decir, de las personas económicamente activas.

La distribución de la fuerza de trabajo por tipo de actividad económica se relaciona estrechamente con la organización económica del país y arroja luz sobre la calidad de la población trabajadora.

En México son muy diferentes las condiciones de las áreas rurales y de las urbanas. En el campo no hay un mercado de trabajo organizado y no está bien definido. Predominan las personas ocupadas por cuenta propia y el empleo tiene un carácter muy estacional e irregular. No se aplica de la misma manera el concepto para las áreas urbanas, pues, una persona que trabaja en su propio pedazo de tierra durante la estación agrícola no tiene trabajo fuera de la estación.

Sin embargo, a pesar de no estar laborando no puede llamársele desempleada, pues, es más apropiado decir que está disponible para trabajar.

Además, debe de advertirse que en el caso de los niños que laboran en el campo no se contempla ninguna edad mínima, y del mismo modo, tienen que considerarse a aquellos que trabajan por cuenta propia, por lo que el concepto de trabajo debe de interpretarse de manera más laxa en las encuestas agrícolas.

Sumado a ello, debe de hacerse hincapié en el hecho de que las técnicas de la medición de la actividad económica, el empleo y el subempleo, no son enteramente satisfactorias. Se necesita mucha investigación para encontrar los métodos adecuados para un conjunto dado de condiciones. Dado el carácter estacional del empleo en los países en desarrollo, es importante que la referencia de tiempo de la encuesta sea muy específica.

Una de las formas de evaluación del nivel económico de los individuos o de las familias de una población consiste en descubrir como emplea la comunidad su ingreso o, más específicamente, el patrón de gastos de las familias con diversas características sociales y económicas. Una de las maneras de conocerlo es a través del “gasto familiar en consumo”, el cual, se refiere a todos los gastos en bienes destinados para el consumo y los servicios.

Sin embargo, también hay métodos alternativos para la cuantificación del gasto, por ejemplo, cuando se considera el valor de los bienes y servicios consumidos por la familia en un periodo específico de referencia.

El segundo método es la medición del gasto en consumo que se basa en el valor total de todos los bienes y servicios recibidos durante el periodo, independientemente del momento en el que se haga el pago total o de que se consuman enteramente los bienes durante dicho intervalo de tiempo .

Ahora, una de las desventajas de las encuestas de empleo y desempleo consiste en que la recolección de datos puede estar sujeta a grandes errores de respuesta y, la correcta clasificación de los datos depende considerablemente de la

capacidad de los encuestados para recordar los renglones de gastos y colocarlos correctamente en el tiempo.

Aparte de las fallas de memoria de los encuestados, las diferencias de las prácticas y los enfoques de entrevistadores pueden generar sesgos y variaciones apreciables en la recolección de la información.

### **LAS ENCUESTAS LONGITUDINALES<sup>3</sup>.**

Una encuesta longitudinal es una colección de datos de una muestra de elementos en varias ocasiones en el tiempo. Las cuales pueden tratar de diversos temas así como tener diferentes propósitos. Tienen un mayor alcance que las encuestas transversales, pues no se aplican una sola vez en el tiempo.

Las encuestas longitudinales tienden a variar en muchas dimensiones, principalmente en los aspectos metodológicos, los cuales incluyen características como el diseño, así como los intervalos entre la recolección de datos y el modo en el que éstos son recolectados, es por eso que existen varias encuestas longitudinales con distintos objetivos. Algunas de ellas son:

- Encuestas de negocios: Comúnmente se basan en indicadores económicos y sólo contienen un rango limitado de información. Ésta es recolectada en intervalos de frecuencia muy seguidos. Una de las razones fundamentales por las que este tipo de encuestas se realizan como panel –del cual se hablará más adelante– es porque pueden mejorar la precisión de las estimaciones.
- Encuestas de graduados: Sirven principalmente para mantener un seguimiento en las universidades sobre los estudiantes graduados y así conocer la actividad que desarrollaron después del término de sus estudios.
- Encuestas a hogares tipo panel: Se han llevado a cabo en distintos países, estas encuestas suelen ser multietápicas y cumplen con el propósito

---

<sup>3</sup>Cfr: Peter Lynn *Methodology of Longitudinal Surveys*, 2009 Pp. 1-11

general de recolectar datos del comportamiento, actitudes y circunstancias de un específico rango de datos y problemáticas sociales y económicas. Su principal objetivo es probar que una fuente de datos puede ser utilizada por una amplia gama de usuarios. La estructura suele ser compleja ya que envuelve un cierto número de entrevistas con cada persona del hogar y puede que adicionalmente se agreguen cuestionarios de complementación.

- Cohortes de nacimiento: Consiste en analizar y dar un seguimiento a personas nacidas en un lugar específico en un tiempo muy particular. La muestra es tomada sobre un periodo, por ejemplo un año.
- Estudios epidemiológicos: Pretenden hacer un seguimiento de cierto grupo con respecto a los cambios en su salud, considerando a los que tengan cierta enfermedad o alguna característica en especial.
- Estudios de medidas repetidas: Son estudios que se hacen periódicamente para medir la misma variable de la misma forma o en cada unidad de la muestra. La mayoría de las encuestas de tipo longitudinal incluyen el estudio de medidas repetidas.

#### **LOS DISEÑOS TIPO PANEL**

Cuando el objetivo de una investigación es analizar cierto tipo de población, pero en vez de hacer una única medición a cada entrevistado se realizan varias, se dice que se está en presencia de diseños longitudinales tipo panel. Este planteamiento selecciona la muestra al principio de la investigación y lleva a cabo una recolección de datos sobre el mismo grupo de sujetos en diferentes momentos en el tiempo, con el objetivo de obtener y reportar los cambios individuales de las personas seleccionadas. Asumiendo que a través de este proceso es posible conocer como determinadas circunstancias que generan cambios en los comportamientos y en las actitudes de los individuos, lo que permite conocer con gran precisión las causas de las modificaciones en determinadas conductas,

opiniones e interpretaciones. Algunas características de las encuestas tipo panel son las siguientes:<sup>4</sup>

- Recolecta información longitudinal, obtenida a partir de encuestas que se realizan periódicamente a un mismo grupo de personas.
- Es una herramienta que permite la realización de estudios sociales y físicos que no pueden ser inferidos mediante un estudio transversal.
- Permite la medición y el análisis de las variables micro-económicas de una sociedad y puede controlar los efectos de las que no se ven reflejadas en cierto periodo porque no dependen del tiempo.

Las principales diferencias respecto al resto de diseños es la periodicidad de las mediciones y la posibilidad de realizar un seguimiento continuo de los fenómenos objeto de estudio, lo que proporciona un conocimiento más profundo y detallado del comportamiento de los sujetos analizados tales como:

- Describir las pautas de estabilidad y cambio,
- Mostrar el orden temporal de los fenómenos,
- Estudio de transiciones, llegar a establecer análisis de trayectorias vitales, ocupacionales, etc,
- Por último, conviene destacar su utilización para exponer la incidencia de nuevos eventos entre dos o más fechas.

#### **CLASIFICACIÓN DE LOS DISEÑOS TIPO PANEL<sup>5</sup>**

Las encuestas que recolectan información sobre la misma unidad en múltiples ocasiones pueden variar de acuerdo a la naturaleza de la información, de la población que está siendo estudiada y de los objetivos a los que se quiere llegar

---

<sup>4</sup>Cfr: Vidal Díaz de Rada, *Tipos de encuestas considerando la dimensión temporal*, Pp. 131-145.

<sup>5</sup>Cfr: Peter Lynn, *Methodology of Longitudinal Surveys*, 2009, pp 21-24.

con dicha investigación. Con esto, se tienen distintos tipos de encuestas tipo panel a partir de su diseño. A continuación se explica cada uno de los paneles.

*Fijo:* Este tipo de panel intenta recoger los datos de las encuestas de las mismas unidades en múltiples ocasiones. Después de la selección de la muestra no hay cambios en ella en ninguna ocasión. La única pérdida teórica de elementos en este tipo de diseño son los fallecimientos de los entrevistados.

*Repetido:* Este tipo de diseño envuelve una serie de encuestas tipo panel, las cuales pueden o no traslaparse en el tiempo. Cada una de ellas es diseñada para representar una población equivalente, según el objeto de la investigación, teniendo para cada una las mismas características de la población aplicados en diferentes puntos en el tiempo.

*Rotativo:* En este caso, el panel está predeterminado por las proporciones de las unidades muestrales que son remplazadas cada vez que se realiza el trabajo de campo. Cada unidad cambiará en la muestra por la misma ocasión de aplicaciones. Normalmente, cada unidad permanece en la muestra durante el mismo número de años.

*Dividido:* Éste envuelve una muestra combinada de corte transversal y un panel cada vez que se realiza el trabajo de campo, es decir, implica una muestra tipo panel en cada ocasión adicionándole un suplemento de muestra de corte transversal en cada ocasión.

Por supuesto que los tipos de diseño señalados anteriormente no son todos los que posiblemente podrían existir, un claro ejemplo es que un panel repetido puede o no incluir muestras de nuevos individuos. Comúnmente se acostumbra usar un panel fijo, es decir, se selecciona una sola muestra y se le da seguimiento para obtener información detallada sobre cambios en las características.

Un panel repetido y los paneles rotativos son de gran utilidad para evaluar cambios en estimaciones longitudinales ya que representan poblaciones comparables, cada uno tratado con una misma metodología. En muchos aspectos los diseños son asintóticamente equivalentes, la diferencia inicia cuando alguno de

ellos comienza a ser más fundamental que el otro para la investigación. Esto es útil para estudios de transición y estabilidad en varios campos, tal es el caso de fenómenos que cambian lentamente o tienen efectos a largo plazo. Cuando el cambio es rápido y a mediano plazo se puede plantear paneles cortos o rotativos.

Un panel dividido puede ser usado si el cambio es gradualmente considerable y es necesario ser observado sólo sobre periodos cortos.

Un aspecto importante es el de la posibilidad de que la ejecución de la encuesta ofrezca estimaciones transversales y longitudinales. En principio, un panel fijo más los nacimientos pueden ofrecer esta posibilidad, pero en muchas situaciones, realizar este trabajo es muy complejo y costoso.

Un segundo aspecto que se debe tomar en cuenta es la unidad principal de análisis. En buena parte del estudio longitudinal, la unidad de observación es algún tipo de suceso, antes que una persona, un hogar o un negocio. Por ello la muestra debe ser representativa de todos esos sucesos durante un período específico. Esto sólo se puede lograr si la muestra es representativa de todos los individuos que han vivido dichos eventos, dado que los datos son proporcionados por ellos. Con esto se subraya la importancia de incluir en el marco de la muestra a todas las personas que puedan haber experimentado dichos sucesos, incluso si la posibilidad de haber vivido la situación hubiera sido de sólo un breve período de tiempo. Dicho de otro modo, es preferible incluir siempre en la muestra a todos los nacimientos en la población de individuos en riesgo.

#### ***ASPECTOS FUNDAMENTALES DE UN DISEÑO***

En cualquier encuesta es necesario definir cuál será la población objetivo sobre la que se harán inferencias y el marco de la población de las unidades que serán muestreadas, comúnmente en muchos casos no es totalmente congruente y en ciertos casos es apropiado el sistema de ponderación para ajustar la muestra, sin embargo dicha población muy difícil de definir ya que en este tipo de encuestas está en constante cambio. Así que es necesario utilizar algunos métodos para

definir una población fija sobre la cual está basado un diseño, los cuales son los siguientes:

Una población estática sobre la base de la población en el momento de seleccionar la muestra del primer levantamiento: No hay nacimientos, el único cambio son las muertes y el cambio de localización. Una desventaja que resulta es que la precisión se reduce con el tamaño de la muestra y la ventaja es que la población es fija en todos los levantamientos, así que la información será constante lo cual hace que la ponderación y el análisis de los subsecuentes levantamientos sea relativamente sencillo.

Una encuesta definida de tipo transversal: Con esta aproximación, tanto nacimientos como muertes durante el estudio son excluidos de la población en estudio que tiene este diseño, es necesario adicionar individuos para poder calibrar la muestra y el objetivo, del mismo modo sólo se obtienen estimaciones imparciales de la población objetivo si no hay relación entre la probabilidad de estar en la intersección y la variable de estudio.

La población es tomada como la unión de varias encuestas transversales en cada levantamiento: Esta aproximación parece ser la más utilizada para una población dinámica pero requiere de muestras dinámicas para introducir nacimientos y evitar parcialidad.

#### ***VENTAJAS DE LOS DISEÑOS TIPO PANEL***

- Es el único diseño que permite medir los cambios a nivel individual, bruto y agregado.
- Conocimiento de la dinámica del cambio, detectando las circunstancias en las que se producen determinados acontecimientos.
- Permite conocer la reacción a una variable experimental.
- Agregando nuevas preguntas en las siguientes mediciones, es posible lograr gran información.
- Reducción de errores al utilizar verificaciones y controles en las respuestas.

- Realizar varias entrevistas con los encuestados permite una cierta familiaridad que se traduce en una mejora en la calidad —y cantidad— de la información recogida.
- Dan la posibilidad de construir medidas derivadas de datos recolectados en distintos puntos del tiempo, donde los datos no podrían ser recolectados de una manera retrospectiva en un punto del tiempo.

### ***DESVENTAJAS DE LOS DISEÑOS TIPO PANEL***

Muchas veces puede ocurrir que en una encuesta las respuestas dadas por una persona que ya participó en ella y que previamente había contestado de una manera, modifique su respuesta al volvérselo a entrevistar. En este caso, las contestaciones pueden estar condicionadas por la experiencia previa de los que formaron parte de la misma encuesta. Así, se puede decir que el comportamiento de la muestra se encuentra modulada por la experiencia de la primera entrevista, pudiendo afectar la manera en la que se responde el cuestionario en la segunda entrevista.

Además, otra desventaja es que en ocasiones los entrevistados por razones propias o ajenas a ellos, deciden ya no participar en el estudio, provocando con ello un desgaste del panel, que quiere decir que la muestra se reduce por la creciente ausencia de entrevistados. Es por eso que es de gran importancia considerar el desgaste porque puede producir distintas estimaciones y cambios en las mediciones. En general, las desventajas que otorgan este tipo de diseños son las siguientes:

- Respecto a las relacionadas con el muestreo, la primera crítica señala la dificultad para seleccionar personas que colaboren durante un largo período de tiempo.
- Un efecto directo de la desventaja anterior es la falta de representatividad de las muestras, debido a que las personas que participan en los paneles no suelen representar a una selección aleatoria del universo poblacional.

- Pudiera decirse que existe un sesgo de representatividad de la muestra debido a la autoselección de los entrevistados.
- El desgaste de la muestra debido a las molestias causadas por las sucesivas mediciones, que se traducen en abandonos a lo largo de la investigación. El cual puede tener causas naturales (fallecimiento del panelista), aunque, en su mayor parte, se produce por otros motivos, como pérdida de interés en el proyecto, falta de respuesta, dificultad de localización por cambios de domicilio.

# CAPÍTULO II

## LA ENCUESTA NACIONAL A HOGARES RURALES EN MEXICO ENHRUM I Y LA ENHRUM II

El sector agropecuario ha sufrido grandes transformaciones en los últimos años debido a los procesos de urbanización, globalización, migración y el abandono que se ha hecho de éste. Por otro lado, los avances tecnológicos, las mejoras en la productividad, así como en la genética de los productos para aumentar su variedad, han generado un nuevo esquema de organización para la producción comercial y nuevos métodos de inserción en el mercado mundial e, incluso, han apoyado el surgimiento de esquemas de desarrollo rural diferentes. Por lo anterior es que el objetivo de la ENHRUM es obtener información representativa en el plano nacional sobre la economía y sociedad rural de México para tener mayor información con respecto a dicho sector y así poder hacer un análisis con estos datos.

En el 2002 se aplicó la ENHRUM y nuevamente se realizó en el 2007 para poder darle un seguimiento tanto de los integrantes de los hogares, como de las actividades que eran realizadas, además de otros factores que predominaban dentro de estas localidades. Esta encuesta es un proyecto del Programa de Estudios del Cambio Económico y Sustentabilidad del Agro Mexicano (PRECESAM) del Centro de Estudios Económicos de El Colegio de México y del *Rural Economies of the Americas and Pacific Rim* (REAP) de la Universidad de California en Davis.

Lo que se planteó hacer con el diseño de esta encuesta fue un análisis tipo panel, como el descrito previamente, donde para ciertas secciones se hizo una categorización de las variables que tuvieron en común tanto la ENHRUM I y la II. Sin embargo para la primera encuesta se hicieron ciertas modificaciones a las variables con el fin de que pudiera trascender el análisis. A continuación se

describirán las características generales de ambas, para explicar con más detalle cómo es que se hizo el diseño tipo panel para estas dos encuestas.

## **LA ENHRUM I<sup>6</sup>**

La realización de la ENHRUM 2002 se llevó a cabo en dos etapas, la primera fue la encuesta comunitaria (CC) y la segunda etapa consistió en la encuesta de hogares (CH). La CC tuvo un periodo de cobertura de agosto a octubre de 2002, y se aplicó a las autoridades de cada localidad, con el propósito de obtener la siguiente información:

- Las relaciones de cada comunidad con su entorno (comercio, trabajo, migración, etc.).
- Las unidades de medida locales.
- La infraestructura económica y social.
- Las principales actividades económicas.
- El tipo de propiedad de la tierra.
- Acceso y uso de los recursos naturales.
- Los mercados del maíz.

Ya realizada la CC se procedió a aplicar la CH la cual tuvo como periodo de cobertura de enero a marzo de 2003. El principal objetivo de la aplicación de esta encuesta fue el obtener los siguientes datos de cada miembro del hogar:

- Sexo
- Escolaridad
- Historia de trabajo
- Historia migratoria
- Costo e ingresos de sus actividades:
  - Uso de trabajo familiar.
  - Producción para el consumo propio.

---

<sup>6</sup>La información que a continuación se expone fue obtenida de la nota metodológica de la ENHRUM 2002.

- Origen de compras y ventas

### **MARCO CONCEPTUAL**

Teniendo en cuenta que el marco conceptual es una manera organizada de pensar en el cómo y el por qué de la realización de un proyecto, y en cómo se entienden sus actividades, dicha encuesta sigue las definiciones de la Encuesta Nacional de Ingresos y Gastos de los Hogares de 2000 (ENIGH-2000) que tiene como objetivo general proporcionar información sobre la distribución, monto y estructura del ingreso y el gasto de los hogares.

Para la realización de la CH–ya referida– se tomaron los elementos sociodemográficos y los económicos de la ENIGH 2000, los cuales fueron:

- Edad
- Sexo
- Nivel de instrucción
- Relación de parentesco
- Remuneración al trabajo
- Autoconsumo-ingreso
- Gasto

Los elementos que se definieron a partir del objetivo propio de la encuesta fueron:

- Miembros del hogar
- Miembros permanentes del hogar.
- Trabajo local
- Trabajo familiar

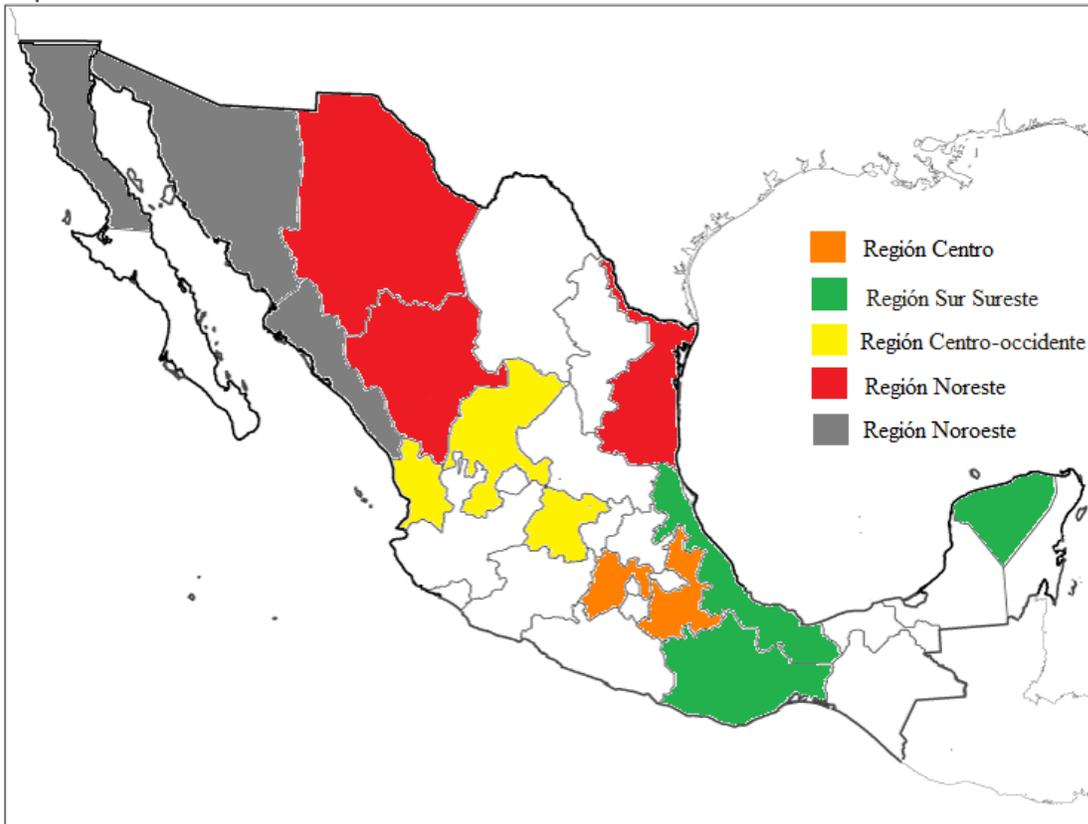
## COBERTURA

Para la ENHRUM 2002 se utilizó un tipo de cobertura temporal que se llevó a cabo en el momento del levantamiento de la encuesta. Además, se tomaron otros periodos de referencia como se muestra en la siguiente tabla:

VARIABLES	PERIODO DE REFERENCIA
De tipo económico(ingreso-gasto)	1° de enero al 31 de diciembre 2002
Sociodemográficas, parentesco, sexo, edad, religión, estado civil, escolaridad.	Momento de la entrevista
La historia de trabajo de miembros del hogar.	1980-2002

Asimismo existió una cobertura geográfica nacional, donde se tomaron en cuenta poblaciones de 500 a 2499 individuos agrupadas en cinco regiones retomadas del Plan Nacional de Desarrollo.

REGIÓN	ESTADO
1.Sur Sureste	Oaxaca
	Veracruz
	Yucatán
2.Centro	Estado de México
	Puebla
3.Centro-occidente	Guanajuato
	Nayarit
	Zacatecas
4.Noroeste	Baja California
	Sinaloa
	Sonora
5.Noreste	Chihuahua
	Durango
	Tamaulipas



**SECCIONES**

El cuestionario de hogares estuvo compuesto por diez secciones las cuales se subdividieron en distintos cuadros para una organización más precisa de la encuesta. De igual manera se consideraron los distintos aspectos en los que podrían variar las respuestas de los encuestados.

<b>SECCIONES DEL CUESTIONARIO</b>	
1	Vivienda
2	Miembros del hogar
3	Características de las parcelas
4	Producción de cultivos
5	Ganadería
6	Bienes y servicios
7	Recursos naturales
8	Otros gastos e ingresos
9	Activo
10	Herencia y activos

## LA ENHRUM II<sup>7</sup>

Como la ENHRUM 2002 tuvo un gran impacto en la determinación de un panorama general del campo mexicano, en el 2007 se decidió aplicar la segunda ronda de esta encuesta tratando de realizar un diseño tipo panel. La ENHRUM 2007 se aplicó en los mismos meses que la primera encuesta, tomando el diseño muestral y los lineamientos de la encuesta de 2002.

Los propósitos básicos en el 2007 fueron:

- 1) Recabar datos sobre la economía rural y sus recursos naturales.
- 2) Ampliar la información obtenida en 2002 y 2007.
- 3) Apoyar y entrenar a jóvenes estudiantes para que analicen con rigor la economía rural.
- 4) Elaborar estudios sobre la demografía, la economía, la sociedad, la migración y usos de los recursos naturales en el medio rural de México y los cambios experimentados durante los últimos 5 años.

La versión 2007, al igual que su edición anterior, contó con los mismos componentes: esto es la CC y la CH. Adicionalmente, se aplicó un test de memoria a los entrevistados.

### COBERTURA

VARIABLES	PERIODO DE REFERENCIA
De tipo económico(ingreso-gasto)	1° de enero al 31 de diciembre 2007
Sociodemográficas, parentesco, sexo, edad, religión, estado civil, escolaridad.	Momento de la entrevista
La historia de trabajo de miembros del hogar.	1980-2007

---

<sup>7</sup>La información que a continuación se expone fue obtenida de la nota metodológica de la ENHRUM 2007.

La cobertura fue nacional y se tomaron en cuenta poblaciones de 500 a 2499 individuos, siguiendo los mismos lineamientos de la ENHRUM 2002.

### **SECCIONES**

El cuestionario de hogares estuvo compuesto por 15 secciones que a su vez se integraron por cuadros referentes a los distintos temas de la sección. A continuación se señalaran las secciones principales del cuestionario aplicado en 2007.

<b>SECCIONES DEL CUESTIONARIO</b>	
1	Hoja de recontacto
2	Miembros del hogar
3	Características de las parcelas
4	Producción de cultivos anuales
5	Permanentes
6	Contabilidad de cultivos anuales y permanentes
7	Ganadería
8	Bienes y servicios
9	Recursos naturales
10	Activos
11	Vivienda
12	Crédito
13	Ahorro e inversiones financieras
14	Otros ingresos y gastos
15	Eventos exógenos

### **DESCRIPCIÓN DE LAS SECCIONES DEL CUESTIONARIO<sup>8</sup>**

El objetivo de cada sección fue recabar información concisa y necesaria para poder cumplir con el objetivo planteado para esta encuesta, desde los aspectos sociodemográficos hasta los económicos, es por ello que a continuación se hace una breve descripción de las secciones utilizadas en la ENHRUM 2007, teniendo

---

<sup>8</sup> En este apartado se utiliza la descripción de las secciones que se encuentran en la nota metodológica de la ENHRUM 2007.

en cuenta que esta misma descripción aplica para las secciones de la encuesta 2002.

1. *Hoja de recontacto*. Contiene información sobre la forma en qué se localizó al hogar seleccionado y encuestado en 2002 y tuvo como objetivo registrar los términos en los que se volvió a contactarlo, así como la dirección y localización de la vivienda del hogar.

2. *Miembros del hogar*. Indagó acerca de las características sociodemográficas, laborales y migratorias de las personas que residen en el hogar y de las que no residen en él.

3. *Características de las parcelas*. Recabó información referente a las características de la tierra o parcelas que el hogar posee y/o trabaja. Los datos se registraron en cuatro niveles:

- a) Registradas en 2002.
- b) Nuevas.
- c) Dadas a otros agricultores en renta, préstamo, a medias o al tercio en 2007.
- d) Recibidas de otros agricultores en renta, préstamo, a medias o al tercio.

4. *Producción de cultivos anuales*. La finalidad fue explorar de manera minuciosa las características del proceso productivo de los cultivos temporales que tuvo el hogar durante los ciclos primavera-verano y otoño-invierno de 2007.

5. *Permanentes*. Como la sección anterior, el propósito de ésta fue averiguar las características del proceso productivo, en este caso, de los cultivos permanentes que tuvo el hogar en 2007.

6. *Contabilidad de cultivos anuales y permanentes*. La finalidad de esta sección fue realizar un inventario con la contabilidad de la producción, autoconsumo y

comercialización de todos los cultivos registrados. Las secciones utilizadas fueron la de 'Cultivos temporales' y 'Cultivos permanentes'.

7. *Ganadería*. Se capturaron los ingresos y gastos que se derivaron de la actividad pecuaria.

8. *Bienes y servicios*. Esta sección acopió información referente a los ingresos y gastos provenientes de los negocios del hogar y también indagó sobre los negocios registrados en 2002.

9. *Recursos naturales*. El objetivo de este apartado fue registrar los cambios entre 2002 y 2007 respecto al aprovechamiento de los recursos naturales por parte del hogar, así como los ingresos y costos.

10. *Activos*. El interés en esta parte del cuestionario radicó en capturar las características de la propiedad, compra y renta de los activos del hogar como son vehículos de carga y transporte, tractores y sus implementos y, bienes inmuebles.

11. *Vivienda*. En la sección se reunieron dos tipos de información: 1) la de propiedad de la vivienda donde habita el hogar y las obras que realizaron para su mejora o ampliación durante 2007; y, 2) de la estructura de la vivienda y el tipo de material que predomina en su construcción y los servicios con los que cuenta.

12. *Crédito*. Capturó detalles de las condiciones de los préstamos activos del hogar en el área de actividades productivas y de consumo.

13. *Ahorros e inversiones financieras*. El propósito de esta sección fue conocer la capacidad de ahorro del hogar entrevistado.

14. *Otros ingresos y gastos*. Se registraron los gastos efectuados en el consumo del hogar y el lugar donde éstos se realizaron. Los cuales incluyeron las compras efectuadas durante el 2007.

15. *Eventos exógenos*. En este apartado se cuantificó la presencia de “shocks” que hayan afectado los intereses del hogar o de alguno de sus miembros; tales como pérdidas de insumos agrícolas, desastres naturales, enfermedades graves, robos e incendios.

## **LA ENHRUM I Y ENHRUM II COMO ENCUESTA TIPO PANEL**

Como se reportó anteriormente el propósito de la ENHRUM I y II fue darle seguimiento a la información para crear datos de corte longitudinal tipo panel. Para generar este diseño se tuvieron que hacer ciertas modificaciones a las secciones, ajustándolas a la ENHRUM 2002, exceptuando las secciones 1, 12, 13 y 15 que no cuentan con el número adecuado de observaciones ya que estas secciones se incorporaron en la encuesta del 2007, así que no se consideraron dentro del panel. Sin embargo, para estos cuatro casos es posible sacar preguntas en común para ambas encuestas.

Como es posible percatarse las ENRHUM tienen un diseño tipo panel fijo, ya que, consideran la misma muestra para ambas rondas donde la única salida fue el movimiento de los hogares.

Para poder desarrollar el panel se siguieron los siguientes pasos

- Se identificaron las variables equivalentes dentro de las dos encuestas y en algunos casos se hizo un ajuste a las variables de la ENHRUM 2002 ya que no eran estrictamente iguales a la segunda.
- Como segundo paso se tomó como referencia las etiquetas del ejercicio en 2007 para poder renombrar las variables de la ENHRUM 2002. Se realizó una recodificación para éstas, donde las opciones de respuesta no coincidían o no eran equivalentes.
- Como paso siguiente se planteó la unión de todas las secciones que fueron posibles, es decir, se creó un código único para cada una de ellas y

tomando en cuenta el número de cuestionario se trataron de unir todas las tablas que comprendieron a la sección correspondiente.

- En seguida, fue necesaria la creación de un código único; por ejemplo en la sección 2, 'Miembros del hogar' se combinó el número de hogar y el número de individuo en todas las tablas que fueron parte de la sección, cabe aclarar que este código sólo se hizo para ciertas secciones.

Para la realización del panel fue necesario verificar que la información y los datos obtenidos en el 2002 coincidieran con los de la segunda ronda que en este caso fue el ejercicio del 2007.

Para poder verificar estos datos se utilizaron las siguientes variables como clave:

- Edad: Para la variable edad se espera que la diferencia entre las dos encuestas este entre 4 y 6 años como máximo.
- Sexo: Para esta variable codificada como 1 para el caso masculino y 2 para el femenino, se espera que la diferencia sea cero.
- Parentesco con jefe del hogar. En este caso la codificación se hizo de acuerdo con la relación que existía con éste. Del mismo modo se esperaba que la diferencia fuera cero.
- Estado civil: De esta variable se esperó que no hubiera grandes modificaciones dentro de la muestra, sin embargo para aquellos casos donde existió algún cambio se verificó que estuvieran en una edad razonable (menores de 60 y mayores de 10).

#### **DIFERENCIA ENTRE LA ENHRUM I Y LA ENHRUM II**

Debido a que la ENHRUM es una encuesta tipo panel, como se planteó en el capítulo previo, una de las grandes desventajas es que al estar la población en constante movimiento, se presentan muchos factores que dificultan el seguimiento de la información, ya que es difícil tener el control de la población o de la muestra. Además, como se observó, dentro de las dos encuestas se hicieron algunos cambios, unos más significativos que otros, trayendo como consecuencia que no se pudiera recolectar la misma información obtenida en el 2002.

El número de secciones del cuestionario en el 2002 fueron 10, mientras que para la del 2007 se anexaron 4 secciones más, las cuales fueron:

- Hoja de recontacto
- Crédito
- Ahorro e inversiones financieras
- Eventos exógenos

#### **CANTIDAD DE HOGARES SOBRE EL TIEMPO**

Al hablar de una encuesta tipo panel es evidente que existe un desgaste en ella. Asimismo, con el tiempo se pierde información debido a los cambios generados dentro de la población muestreada, es por ello que en comparación con la ENHRUM I, en la ENHRUM II existió una reducción de la muestra como se presenta en el siguiente cuadro:

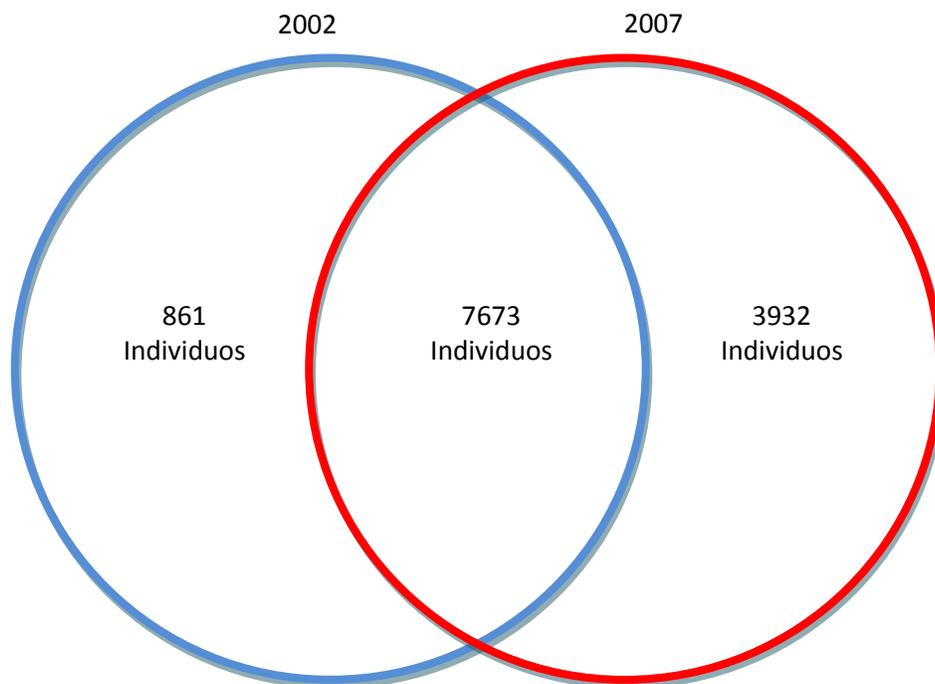
#### **HOGARES ENCUESTADOS**

<b>ENHRUM 2002</b>	<b>ENHRUM 2007</b>
1765	1543

La disminución de hogares encuestados se debió a la movilidad de la población durante el periodo de 5 años. Es decir 222 hogares no se les pudo aplicar la encuesta.

#### **CANTIDAD DE INDIVIDUOS SOBRE EL TIEMPO**

Considerando el panel en su totalidad, el número de individuos dentro de él fueron 12,466 donde 7673 es decir el 61.55% se encuentra en ambas encuestas, 861 individuos (6.91%) sólo en el 2002 y 3932 (31.54%) sólo en el 2007.



A partir de la información anterior se tiene que el 61.55% de la población encuestada fue la que se pudo localizar para realizar la encuesta en las dos rondas. De tal manera que éstas fueron las que conformaron el panel.

# CAPÍTULO III

## ANÁLISIS ESTADÍSTICO

Como se ha señalado, en 2002 y 2007 se realizaron dos encuestas que se trataron de seguir con los lineamientos de un diseño tipo panel. Éste da la posibilidad de realizar una comparación de variables a través del tiempo. Para el caso que ocupa aquí se propone realizar un análisis utilizando tanto la *estadística no paramétrica* así como la *estadística paramétrica*

La estadística paramétrica está basada en parámetros específicos, como la media ( $\mu$ ), la desviación estándar ( $\sigma$ ) o la proporción ( $p$ ) dependiendo de la función de densidad de probabilidad de que se trate. Estos métodos paramétricos usualmente tienen que ajustarse a algunas condiciones completamente estrictas, por ejemplo el requisito de que los datos de la muestra provengan de una población normalmente distribuida. Es por ello que, entre otras cosas, en esta investigación se pretende comprobar este supuesto haciendo una comparación y evaluando si la normalidad es realmente determinante para la diferencia en los resultados.

A diferencia de la anterior, la *estadística no paramétrica* es aquella en la que el modelo no especifica las condiciones de los parámetros de la población de la que se obtuvo la muestra. Además, las suposiciones que se asocian a la estadística no paramétrica comparadas con las paramétricas son débiles y muy pocas.

Las ventajas de utilizar estadística no paramétrica para hacer el análisis son las siguientes:

- Se asume que las observaciones obtenidas no siguen alguna distribución, es decir no se hace una serie de suposiciones severas acerca de los parámetros.
- Hay pruebas estadísticas no paramétricas adecuadas para observaciones hechas en poblaciones diferentes.

- Las pruebas estadísticas no paramétricas son típicamente mucho más fáciles de aplicar que las pruebas paramétricas.

En cuanto a las desventajas de utilizar estadística no paramétrica se encuentran.

- Los métodos no paramétricos tienden a perder información porque datos numéricos exactos son frecuentemente reducidos a una forma cualitativa.
- Las pruebas no paramétricas no son tan eficientes como las pruebas paramétricas, de manera que con una prueba no paramétrica generalmente se necesita evidencia más fuerte (así como una muestra más grande o mayores diferencias) antes de rechazar una hipótesis nula.

Las pruebas que se aplicarán a continuación responden a los dos tipos de estadística señalados. Para la *estadística no paramétrica* se utilizará la comparación de medianas para muestras dependientes con la suma de rangos de Wilcoxon y la comparación de medianas para muestras independientes con la prueba de Mann Whitney , como parte de la *estadística paramétrica*, la prueba t de comparación de medias. En la siguiente sección se abordará profundamente cada una de las pruebas en cuestión.

Parte de lo que se pretende verificar con estos análisis estadísticos es la influencia del supuesto de normalidad en los resultados. Se compararan ambas pruebas y se determinará si existe alguna diferencia cuando se usa cada una de ellas, respectivamente, con el objetivo de comprobar si realmente existe algún cambio en los resultados si es que no se cumple el supuesto antes indicado. De ocurrir esto, se reforzará el hecho de que la violación de un supuesto para esta muestra en específico es realmente importante para el desarrollo de las pruebas paramétricas y, asimismo, ayudará a saber que tanto pueden diferir los resultados utilizando una estadística y la otra.

Asimismo en este proyecto se pretende utilizar dos pruebas de estadística no paramétrica para ver la diferencia que existe cuando las muestras son dependientes y se toman como muestras independientes.

Para verificar el supuesto de normalidad se utilizará la prueba de Kolmogorov Smirnov específicamente en la sección 2 de las encuestas en estudio. Con esto, se comprobará si los datos se distribuyen como una normal y posteriormente se pasará a aplicar las pruebas contempladas previamente.

### **TRATAMIENTO DE LA INFORMACIÓN**

Dada la posición geográfica que tiene México se analizarán principalmente la migración dentro de esta encuesta, del universo de variables contenidos en el panel se decidió tomarlas para realizar dicha comparación y verificar lo que se explicó en hojas anteriores ya que es una de las variables que tiene más relevancia dentro de la investigación.

Como se pudo apreciar en capítulos anteriores, las encuestas están compuestas por diversas secciones donde se trató de hacer un seguimiento tipo panel. La sección que se seleccionó para trabajar fue la 2 que es la sección que se ha verificado con mayor detenimiento y donde se tienen los datos más certeros, del mismo modo las variables seleccionadas que son las de migración, su análisis es un ejercicio interesante ya que por una parte se corroborará que tanto difieren el aplicar los dos tipos de estadísticas y del mismo modo en este lapso de cinco años que tanto ha cambiado la migración. A continuación se describen las variables que se tomaron.

Para realizar el análisis, la base de datos construida incluyó sólo a los mexicanos del sector rural que migran hacia los Estados Unidos y dentro del mismo territorio mexicano en los dos periodos de tiempo en que se realizaron las encuestas. El ejercicio analítico servirá para comparar la migración del 2007 con respecto a la de 2002. Las variables que se tomarán en cuenta son las siguientes:

- Migración a los Estados Unidos
- Migración dentro del territorio mexicano
- Migración masculina a los Estados Unidos
- Migración femenina a los Estados Unidos
- Migración masculina dentro México

- Migración femenina dentro del territorio mexicano

Estas variables fueron tomadas de la sección 2 de las encuestas y corresponde a la parte de sociodemografía. Las preguntas del cuestionario que fueron analizadas son las siguientes:

En la ENHRUM I, sección 2:

- ¿Cuántos meses del 2002 vivió en esta casa?
- ¿Alguien de esta casa (o sus hijos) trabajó en E.U en el 2002? ¿Quién?
- ¿Alguien vivió y trabajo en otras partes de México en el 2002? ¿Quién?

En la ENHRUM II, sección 2:

- ¿Cuántos meses del 2007 vivió en esta casa?
- ¿Es miembro fuera?
- ¿Trabajó en Estados Unidos en 2007?
- ¿Trabajó en otras partes de México en 2007 viviendo fuera de la localidad?

Algunos conceptos que se definen en las encuestas y que son pertinentes esclarecer antes de pasar al análisis son: a) se define al migrante internacional como los miembros del hogar que vivieron o trabajaron en Estados Unidos por más de tres meses. b) Se define al migrante nacional como los miembros del hogar que vivieron o trabajaron en otras partes de México que no fuera su localidad.

Es pertinente aclarar que se tomaron en cuenta solamente las observaciones que estuvieron presentes dentro de las dos rondas, teniendo en claro que en el 2007 los hogares encuestados fueron 1543, esos hogares se tomaron para el 2002, es decir, se excluyeron los 222 hogares que ya no estuvieron presentes en la segunda ronda, por lo que sólo se tomaron en cuenta los hogares que contestaron en los dos levantamientos. Para este primer análisis se medirán los individuos que han migrado por cada hogar aplicando la prueba de suma de rangos de Wilcoxon y la prueba T, pero para empezar se verificará el supuesto de normalidad para esto se procede con las siguientes pruebas realizadas en SPSS.

Es importante considerar que todas las pruebas realizadas se harán a una significancia del 5%.

Para cuestiones de codificación en el paquete estadístico se renombró a las variables de la siguiente forma:

Migración a los Estados Unidos en el 2002= migus\_h2002

Migración a los Estados Unidos en el 2007= migus\_h2007

Migración dentro del territorio mexicano 2002= migmx\_h\_2002

Migración dentro del territorio mexicano 2007= migmx\_h\_2007

Migración masculina a los Estados Unidos 2002= migusmale\_2002

Migración masculina a los Estados Unidos 2007= migusmale\_2007

Migración femenina a los Estados Unidos 2002= migusfem\_2002

Migración femenina a los Estados Unidos 2007= migusfem\_2007

Migración masculina dentro México 2002= migmxmale\_2002

Migración masculina dentro México 2007= migmxmale\_2007

Migración femenina dentro del territorio mexicano 2002= migmxfem\_2002

Migración femenina dentro del territorio mexicano 2007= migmxfem\_2007

## LA PRUEBA DE KOLMOGOROV SMIROV

La prueba se emplea para probar el grado de concordancia entre la distribución de datos empíricos de la muestra y alguna distribución teórica específica.

Desarrollo de la prueba:

- Se desarrolla la distribución acumulativa de la distribución teórica así como la de los datos empíricos.
- Se comparan y se selecciona aquel intervalo de clase que tenga mayor desviación absoluta entre las desviaciones teóricas y observadas.
- Se compara la desviación con los valores críticos de la tabla de Kolmogorov Smirnov.

En este caso se aplicará esta prueba para verificar si los datos se distribuyen como una normal, así que las hipótesis a contrastar son:

$H_0$ : *Los datos se distribuyen como una normal*

*vs*

$H_1$ : *Los datos no siguen una distribución normal*

La estadística de prueba es:

*Si  $D < D_\alpha \rightarrow$  Se acepta  $H_0$*

*Si  $D > D_\alpha \rightarrow$  Se rechaza  $H_0$*

Donde se obtiene que D es la diferencia entre la distribución empírica y la forma en la que se distribuyen los datos, se tiene que:

$$D^+ = \max_{1 \leq i \leq n} \left| \frac{i}{n} - F(x_i) \right| \quad \text{y} \quad D^- = \max_{1 \leq i \leq n} \left| F(x_i) - \frac{i-1}{n} \right|$$

Donde:

$F(x_i)$ : Distribucion Teorica

$n$  = Número total de observaciones

$$D = \max\{|D^+, D^-|\}$$

La D que se usa a una confianza del 95% es:

$$D_{.05} = .0228$$

Del mismo modo se tiene que si:

*Si el valor  $p \geq \alpha \rightarrow$  Se acepta  $H_0$*

*Si el valor  $p < \alpha \rightarrow$  Se rechaza  $H_0$*

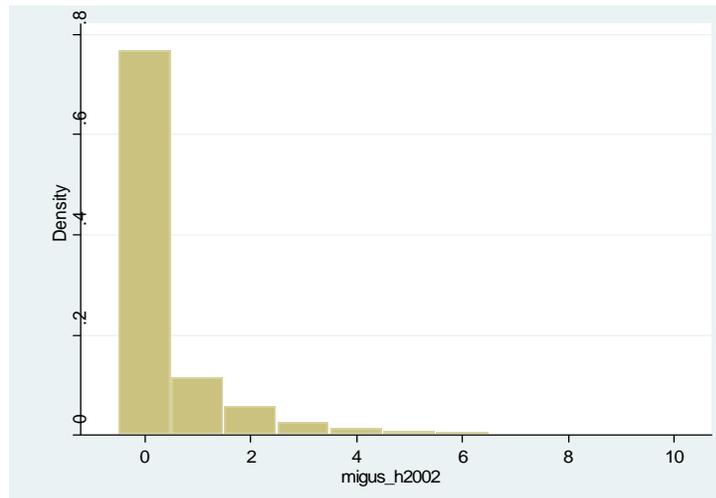
**Tabla 1 Prueba de Kolmogorov Smirnov para la característica de migrantes a los Estados Unidos en 2002**

		migus_h2002
N		1543
Parámetros normales <sup>a,b</sup>	Media	.50
	Desviación típica	1.197
Diferencias más extremas	Absoluta	.428
	Positiva	.428
	Negativa	-.338
Z de Kolmogorov-Smirnov		16.801
Sig. asintót. (bilateral)		.000

En la tabla 1 se observa que dadas las 1543 observaciones la media arroja un valor de .5, es decir, no se puede considerar que exista al menos un miembro por hogar que se fue a los Estados Unidos en el 2002, a una confianza del 95 % se rechaza la hipótesis nula, donde se asevera el hecho que los datos siguen una distribución normal ya que el valor p es altamente significativo. Con respecto a las diferencias se tiene que la diferencia máxima positiva es de .382 y la máxima

negativa es de  $-.314$ , lo cual indica que tanto se alejan de la distribución empírica. Con la estadística de prueba se tiene que la  $D_{.05} = 0.0228$ , esto indica que  $D_{.05} < D$ , por lo tanto se rechaza la hipótesis nula de que los datos se distribuyan como una normal.

Gráfico 1 Distribución de la migración a Estados Unidos en el 2002



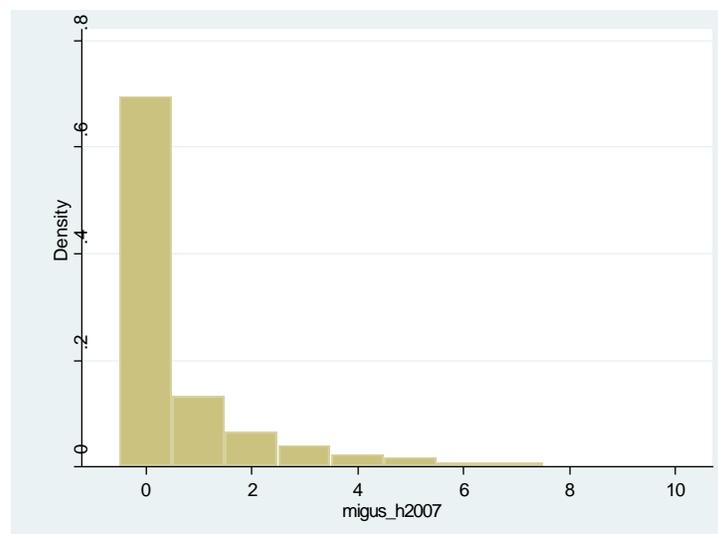
En el Grafico 1 se muestra el histograma de la función de densidad para los migrantes a Estados Unidos, el cual no tiene un comportamiento de una normal.

**Tabla 1.1 Prueba Kolmogorov Smirnov para la característica de migrantes a los Estados Unidos 2007**

		migus_h2007
N		1543
Parámetros normales <sup>a,b</sup>	Media	.75
	Desviación típica	1.545
Diferencias más extremas	Absoluta	.382
	Positiva	.382
	Negativa	-.314
Z de Kolmogorov-Smirnov		15.022
Sig. asintót. (bilateral)		.000

La Tabla 1.1, presentada en la página anterior muestra la prueba de bondad y ajuste de Kolmogorov Smirnov para individuos que migraron a Estados Unidos en el 2007. Se observa que la media para estos datos es mayor que para el 2002, es decir, en promedio hubo un mayor número de migrantes para este año, del mismo modo el valor p es altamente significativo, lo que quiere decir que se rechaza la hipótesis nula y por lo tanto los datos no siguen una distribución normal. Con la estadística de prueba se rechaza la hipótesis nula, dado que las diferencias si son significativas.

Gráfico 2 Distribución de la migración a Estados Unidos en el 2007



En el Gráfico 2 se muestra el histograma de la función de densidad del número de miembros que están en Estados Unidos, claramente se ve que no sigue una distribución normal para los individuos que migraron en el 2007.

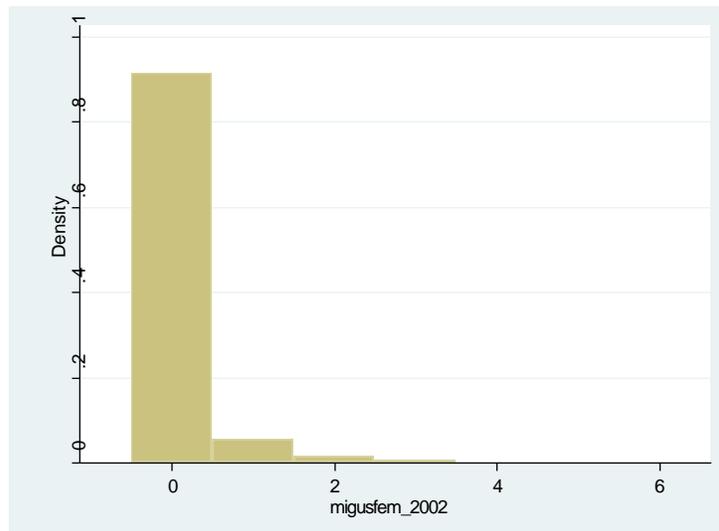
En la tabla 1.2 que se presenta en la siguiente página, se muestra la prueba de normalidad para la población femenina que ha migrado a los Estados Unidos en el 2002, las diferencias máximas negativas como positivas son de .517 y -.398 respectivamente con respecto a la distribución teórica. El valor p es altamente significativo y con una confianza del 95% se rechaza la hipótesis nula, por lo tanto los datos no se distribuyen como una normal. Ocurre lo mismo con la estadística

de prueba, la hipótesis nula es rechazada. Distribución de la migración femenina a Estados Unidos en el 2002

**Tabla 1.2 Prueba de KolmogorovSmirnov para la característica de la migración femenina a los Estados Unidos 2002**

		migusfem_2002
N		1543
Parámetros normales	Media	.13
	Desviación típica	.493
Diferencias más extremas	Absoluta	.517
	Positiva	.517
	Negativa	-.398
Z de Kolmogorov-Smirnov		20.323
Sig. asintót. (bilateral)		.000

**Gráfico 3 Distribución de la migración femenina a Estados Unidos en el 2002**



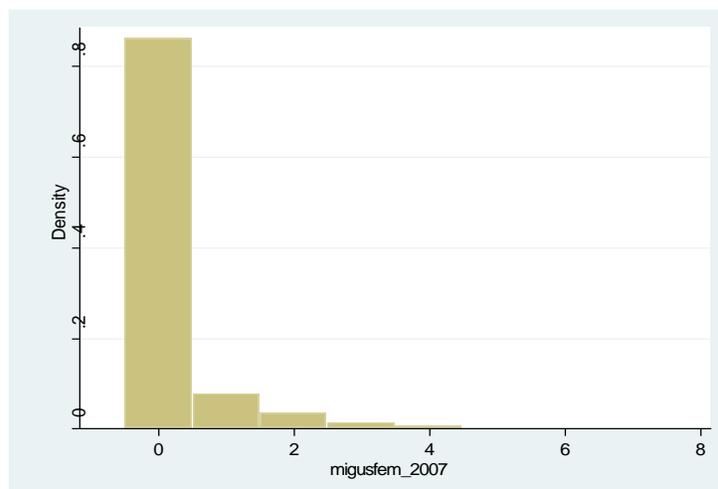
En el Gráfico 3 se ve claramente de que los datos no se comportan como una normal.

**Tabla 1.4 Prueba de KolmogorovSmirnov para la característica de la migración femenina a los Estados Unidos 2007**

		migusfem_2007
N		1543
Parámetros normales <sup>a,b</sup>	Media	.24
	Desviación típica	.719
Diferencias más extremas	Absoluta	.491
	Positiva	.491
	Negativa	-.371
Z de Kolmogorov-Smirnov		19.270
Sig. asintót. (bilateral)		.000

En la Tabla 1.4 se observa que para el 2007 hubo un aumento en la media con respecto al 2002, pero este dato sigue indicando que el promedio de migración femenina hacia los Estados Unidos durante este año es bajo, las diferencias máximas son de .491 y de -.371 con respecto a la distribución empírica, se observa que el valor p es altamente significativo por lo tanto los datos no se distribuyen como una normal. Del mismo modo se cumple que  $D_{.05} < D$  por lo tanto se rechaza la hipótesis nula.

**Gráfico 4 Distribución de la migración femenina a Estados Unidos en el 2007**



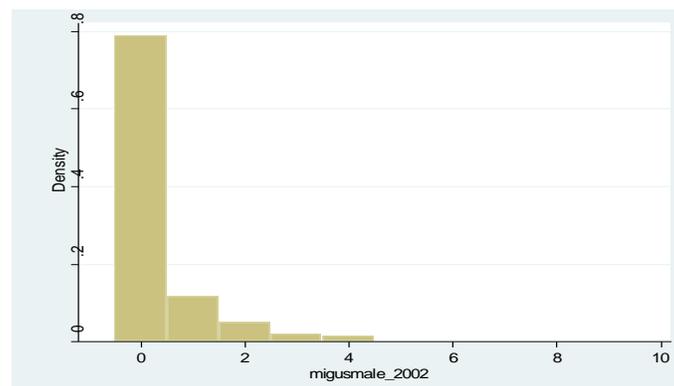
En el Gráfico 4 se muestra la función de densidad para la migración femenina en el 2007, es claro que los datos no se distribuyen como una normal.

**Tabla 1.6 Prueba de Kolmogorov Smirnov para la característica de la migración masculina a los Estados Unidos 2002**

		migusmale_2002
N		1543
Parámetros normales <sup>a,b</sup>	Media	.37
	Desviación típica	.895
Diferencias más extremas	Absoluta	.452
	Positiva	.452
	Negativa	-.339
Z de Kolmogorov-Smirnov		17.745
Sig. asintót. (bilateral)		.000

En la Tabla 1.6 se observa que la media para los hombres que migraron hacia los Estados Unidos es de .37, es decir mayor que la población femenina que ha migrado a este país, a una confianza del 95% se rechaza la hipótesis nula que indica que los datos se comportan como una normal ya que el valor p es altamente significativo. Del mismo modo si se analizan las diferencias con respecto a la función empírica son de .452 y -.339, verificando la estadística de prueba se observa que  $D_{\alpha} < D$  por lo tanto se asevera el hecho de que los datos no se distribuyen como una normal.

**Gráfico 5 Distribución de la migración masculina a Estados Unidos en el 2002**



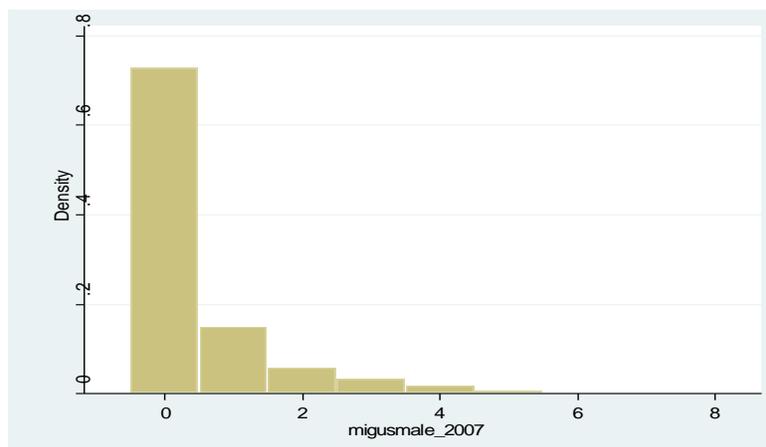
Al verificar el Gráfico 5, se tiene que la función de densidad de los datos para la migración masculina no cumple con una distribución normal.

**Tabla 1.7 Prueba de KolmogorovSmirnov para la característica de la migración masculina a los Estados Unidos 2007**

		migusmale_2007
N		1543
Parámetros normales	Media	.51
	Desviación típica	1.061
Diferencias más extremas	Absoluta	.413
	Positiva	.413
	Negativa	-.314
Z de Kolmogorov-Smirnov		16.215
Sig. asintót. (bilateral)		.000

En la Tabla 1.7 se observa que la media para este año aumentó con respecto a la media del año 2002, el valor p es altamente significativo, es decir, se rechaza la hipótesis nula a una confianza del 95%, si se analiza el estadístico de prueba se tiene que la D es igual a .413, por lo tanto  $D_{.05} < D$ , es decir se rechaza la hipótesis nula, se confirma que los datos no se distribuyen como una normal.

**Gráfico 6 Distribución de la migración masculina a Estados Unidos en el 2007**



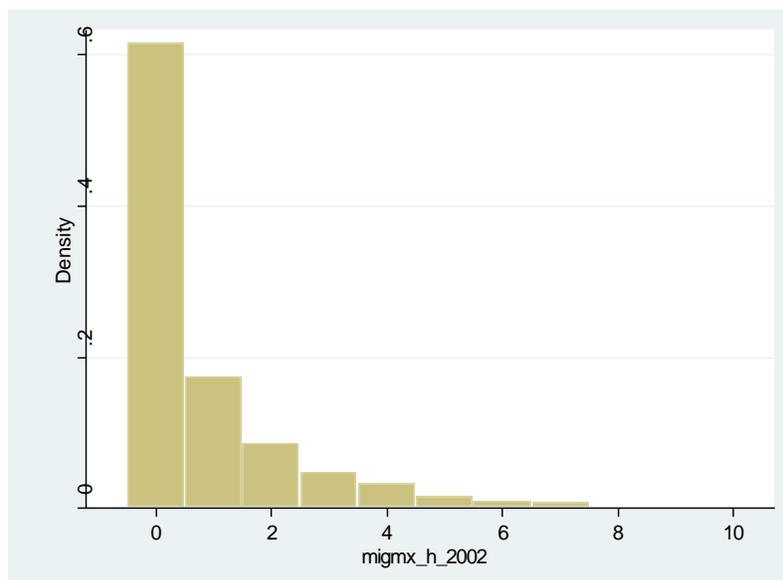
En el Gráfico 6 se muestra que los datos de la migración masculina a Estados Unidos en el 2007 no tiene un comportamiento normal.

**Tabla 1.8 Prueba de KolmogorovSmirnov para la migración dentro del territorio mexicano 2002**

		migmx_h_2002
N		1543
Parámetros normales <sup>a,b</sup>	Media	.90
	Desviación típica	1.582
Diferencias más extremas	Absoluta	.330
	Positiva	.330
	Negativa	-.286
Z de Kolmogorov-Smirnov		12.944
Sig. asintót. (bilateral)		.000

En la Tabla 1.8 se observa que la media es muy cercana a un miembro por familia que migró dentro del territorio nacional, la estadística de prueba indica que se rechaza la hipótesis nula ya que  $D_\alpha < D$ , del mismo modo si se analiza el valor p es evidente que es altamente significativo, es decir que se rechaza el supuesto de normalidad.

**Gráfico 7 Distribución de la migración dentro de México en el 2002**



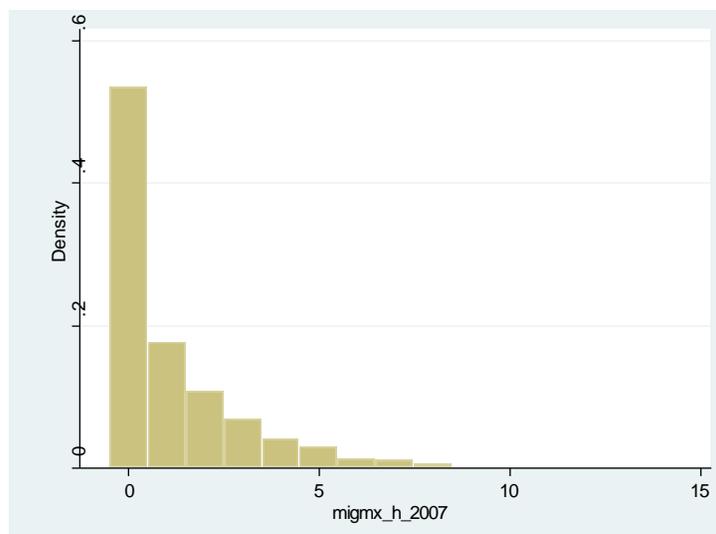
En el Gráfico 7 se tiene la función de densidad para los miembros que migraron dentro del territorio nacional, se ve claramente que no siguen una distribución normal.

**Tabla 1.9 Prueba de Kolmogorov Smirnov para la característica de la migración dentro del territorio mexicano 2007**

		migmx_h_2007
N		1543
Parámetros normales <sup>a,b</sup>	Media	1.20
	Desviación típica	1.836
Diferencias más extremas	Absoluta	.278
	Positiva	.278
	Negativa	-.256
Z de Kolmogorov-Smirnov		10.918
Sig. asintót. (bilateral)		.000

En la tabla 1.9 se observa que para el 2007 el número promedio de individuos que migraron ya supera el 1, es decir hubo un aumento considerable en estos cinco años, con respecto al valor p es altamente significativo, por lo tanto se rechaza la hipótesis nula.

**Gráfico 8 Distribución de la migración dentro de México en el 2007**



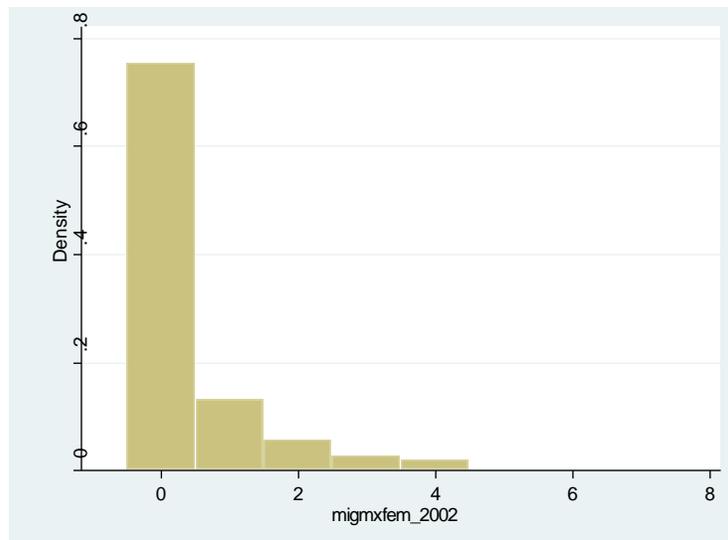
En el Gráfico 8 se tiene la función de densidad de la migración dentro del territorio nacional, claramente los datos no se distribuyen como una normal.

**Tabla 1.10 Prueba de Kolmogorov Smirnov para la característica de la migración femenina dentro del territorio mexicano 2002**

		migmx fem_2002
N		1543
Parámetros normales	Media	.45
	Desviación típica	.965
Diferencias más extremas	Absoluta	.434
	Positiva	.434
	Negativa	-.321
Z de Kolmogorov-Smirnov		17.066
Sig. asintót. (bilateral)		.000

Al analizar la Tabla 1.10, la media para la población femenina es de .45, es decir que el número de mujeres que migraron dentro del territorio mexicano para ese año es bajo, así mismo se asevera el hecho de que los datos estudiados no se distribuyen como una normal ya que el valor p es altamente significativo.

**Gráfico 9 Distribución de la migración femenina dentro de México en el 2002**



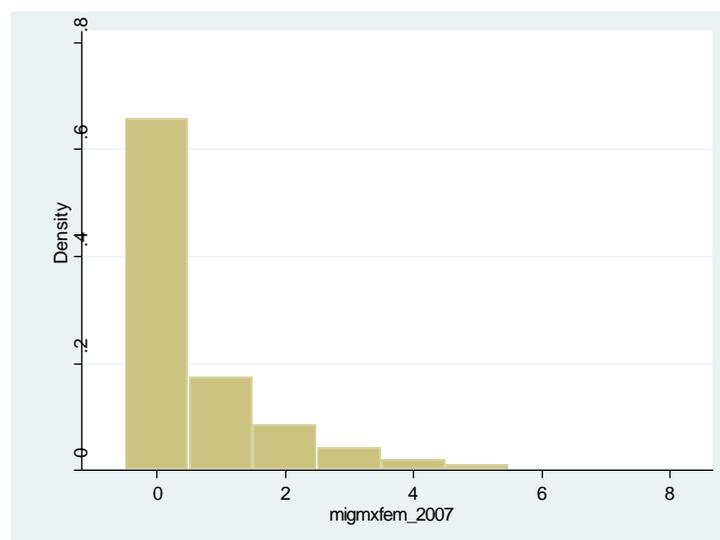
En el Grafico 9, que se muestra en la página anterior, se tiene la función de densidad de las mujeres que migraron dentro del territorio mexicano, se ve claramente que los datos no siguen un comportamiento de una normal.

**Tabla 1.11 Prueba de Kolmogorov Smirnov para la característica de la migración femenina dentro del territorio mexicano 2007**

		migmx fem_2007
N		1543
Parámetros normales <sup>a,b</sup>	Media	.66
	Desviación típica	1.162
Diferencias más extremas	Absoluta	.371
	Positiva	.371
	Negativa	-.286
Z de Kolmogorov-Smirnov		14.588
Sig. asintót. (bilateral)		.000

En la Tabla 1.11 se observa que la media es baja a comparación de la media generada por todos los individuos que migraron dentro del territorio Mexicano, es decir que en promedio el número de mujeres que migró por hogar no llega a uno, analizando el valor  $p$ , es altamente significativo, es decir se rechaza la hipótesis nula a una significancia del 5%, del mismo modo el estadístico de prueba es de .371 por lo tanto es mayor que la  $D_{\alpha}$ . Se rechaza el hecho de que nuestros datos se distribuyan como una normal.

**Gráfico10 Distribución de la migración femenina dentro de México en el 2007**



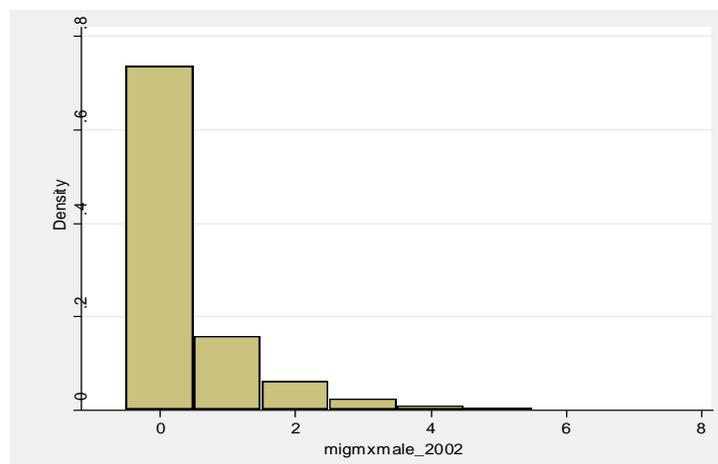
En el Gráfico 10, que se muestra en la página anterior, es evidente que la función de densidad de la migración femenina en el 2007 no sigue una distribución normal.

**Tabla 1.12 Prueba de Kolmogorov Smirnov para la característica de la migración masculina dentro del territorio mexicano 2002**

		migmxmale_2002
N		1543
Parámetros normales <sup>a,b</sup>	Media	.45
	Desviación típica	.909
Diferencias más extremas	Absoluta	.417
	Positiva	.417
	Negativa	-.312
Z de Kolmogorov-Smirnov		16.395
Sig. asintót. (bilateral)		.000

En la Tabla 1.12 se observa que no se cumple el supuesto de normalidad ya el valor p es altamente significativo, por lo tanto los datos de esta muestra no se distribuyen normal. Asimismo si se analiza el estadístico de prueba se tiene una D de .417, mayor a la  $D_{\alpha}$ . Se rechaza la hipótesis nula y se asevera el hecho de que los datos no se distribuyen normal.

Gráfico 11 Distribución de la migración masculina dentro de México en el 2002



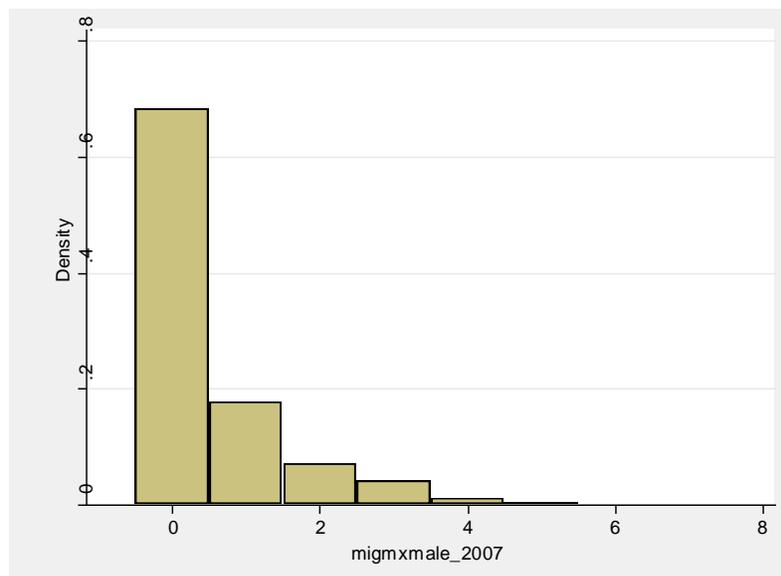
En el caso de la migración masculina, al analizar Gráfico 11 se observa que no siguen una distribución normal.

**Tabla 1.13 Prueba de KolmogorovSmirnov para la característica de la migración masculina dentro del territorio mexicano 2007**

		migmxmale_2007
N		1543
Parámetros normales <sup>a,b</sup>	Media	.55
	Desviación típica	1.007
Diferencias más extremas	Absoluta	.393
	Positiva	.393
	Negativa	-.293
Z de Kolmogorov-Smirnov		15.426
Sig. asintót. (bilateral)		.000

En la tabla 1.13 a un nivel de significancia del 5% se rechaza nuestra hipótesis nula. Analizando la estadística de prueba, indica que la  $D$  es de .393, por lo tanto es mayor que la  $D_{\alpha}$ , es decir, se rechaza el hecho de que los datos se distribuyan como una normal.

**Gráfico 12 Distribución de la migración masculina dentro de México en el 2007**



En el gráfico 12, presentado en la página anterior, muestra que la función de densidad de los hombres que migraron dentro del territorio nacional no se comporta como una distribución normal.

En conclusión se observa que en todas la variables analizadas las cuales son migración a los Estados Unidos tanto masculina como femenina, así como dentro del territorio nacional no siguen con una distribución normal, en este caso corresponde aplicar una prueba no paramétrica pero qué pasa si se viola ese supuesto, se comprobará si realmente la normalidad influye en los resultados, se aplicaran las pruebas correspondientes para aclarar este hecho.

# CAPÍTULO IV

## APLICACIÓN DE LAS PRUEBAS ESTADÍSTICAS

### PRUEBA DE SUMA DE RANGOS DE WILCOXON

La prueba de la suma de los Rangos de Wilcoxon, se utiliza para comparar las medianas que tienen dos muestras tomada una antes y la otra después de un evento, con ella se analizará la diferencia de las respuestas para ver que tanto cambia la mediana de la primera muestra con respecto de la segunda y comprobar si son iguales o existe alguna diferencia:, cuando los datos cumplen el supuesto de normalidad, esta es equivalente a la prueba T en la estadística paramétrica.

Las hipótesis a contrastar son las siguientes:

$$H_0: \theta_D = 0 \quad vs. \quad H_1: \theta_D \neq 0$$

Sea  $X_1, \dots, X_n$  la muestra tomada antes del evento  $X_1, \dots, X_n$

Sea  $Y_1, \dots, Y_n$  la muestra tomada después del evento  $Y_1, \dots, Y_n$

Se define  $\theta_D$  como la mediana existente sobre la diferencia de las observaciones  $X_i, Y_i$   $1 \leq i \leq n$ , por lo tanto la hipótesis nula indica que la mediana de la diferencia de los puntajes es igual a cero contra la hipótesis alternativa que dice que la mediana de la diferencia de los puntajes es distinta a cero, lo cual se traduce a que la suma de los rangos positivos ( $\sum R^+$ ) sea igual a la suma de los rangos negativos ( $\sum R^-$ ). Es decir que la mediana de las dos muestras es igual.

También se tiene que si:  $\sum R^+ > \sum R^-$  entonces se tiene la siguiente hipótesis alternativa  $H_1: \theta_D > 0$ , es decir que la mediana de la primera población es mayor a la de la segunda, en este caso la mediana de  $X_1, \dots, X_n$  es mayor que la de  $Y_1, \dots, Y_n$

Otra hipótesis alternativa se presenta cuando se tiene el caso donde:  $\sum R^+ < \sum R^-$  entonces  $H_1: \theta_D < 0$ . Estaría pasando lo contrario a lo descrito en el párrafo anterior.

A continuación se describe la forma en la que se realiza la prueba de Wilcoxon para obtener los rangos y determinar el estadístico de prueba.

- Para cada par igualado, se determina la diferencia entre los dos puntajes.

$$d_i = X_i - Y_i \quad \forall 1 \leq i \leq n$$

- Se calcula los rangos de las diferencias sin tomar en cuenta el signo de las mismas (o sea en valor absoluto). En el caso de haber empate se asigna un rango promedio a todas las diferencias empatadas es decir:

$$\frac{\text{menor rango del grupo o del empate} + \text{mayor rango del grupo del empate}}{2}$$

- Finalmente el estadístico  $w$  de Wilcoxon será la suma de los rangos correspondientes a las diferencias cuyo signo sea menos frecuente.

Cuando  $n$  es mayor que 16, se usa aproximación Normal para hallar el valor  $p$  de la prueba pues, se puede mostrar que el estadístico de Wilcoxon se aproxima a una normal donde:

$$\mu = \frac{n(n+1)}{4}, \sigma^2 = \frac{n(n+1)(2n+1)}{24}$$

Por lo tanto para obtener el estadístico se obtiene de la siguiente forma:

$$z = \frac{w - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1)$$

Para el estadístico de prueba a una confianza del 95% se afirma que:

$$\text{Si } z < z_{0.05} \rightarrow \text{Se acepta } H_0$$

$$\text{Si } z > z_{.05} \rightarrow \text{Se rechaza } H_0$$

Para el análisis se maneja la  $z_{.05} = 1.96$

Así mismo obtenemos que:

$$\text{Si el valor } p \geq \alpha \rightarrow \text{Se acepta } H_0$$

$$\text{Si el valor } p < \alpha \rightarrow \text{Se rechaza } H_0$$

A continuación se presenta la prueba de rangos de Wilcoxon, donde se compara a los individuos que migraron en el 2007 con respecto a los del 2002 y se analiza la mediana para ver si la distribución para cada año referido ha cambiado o se ha mantenido igual.

**Tabla 2.1 Prueba de rangos de Wilcoxon para la característica de migración a Estados Unidos**

		N	Rango promedio	Suma de rangos
migus_h2007 - migus_h2002	Rangos negativos	131 <sup>a</sup>	191.90	25139.50
	<b>Rangos positivos</b>	<b>312<sup>b</sup></b>	<b>234.64</b>	<b>73206.50</b>
	Empates	1100 <sup>c</sup>		
	Total	1543		

a. migus\_h2007 < migus\_h2002

b. migus\_h2007 > migus\_h2002

c. migus\_h2007 = migus\_h2002

**Tabla 2.2 Estadísticos de contraste**

	migus_h2007 - migus_h2002
Z	-9.256 <sup>a</sup>
<b>Sig. asintót. (bilateral)</b>	<b>.000</b>

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

En las tablas 2.1 y 2.2, presentada en la página anterior, se observa que el valor p es altamente significativo, es decir, que la mediana de las dos poblaciones analizadas es distinta, por lo tanto se afirma que la migración en el 2007 y en el 2002 no son iguales. La suma de los rangos positivos es mayor que la de los negativos, así que en el año 2007 hubo un mayor número de individuos que migraron por cada hogar en comparación con el 2002.

**Tabla 2.3 Prueba de Rangos de Wilcoxon para la característica de migración femenina a Estados Unidos**

		N	Rango promedio	Suma de rangos
migufem_2007 - migufem_2002	Rangos negativos	48 <sup>a</sup>	85.08	4084.00
	<b>Rangos positivos</b>	<b>148<sup>b</sup></b>	<b>102.85</b>	<b>15222.00</b>
	Empates	1347 <sup>c</sup>		
	Total	1543		

a. migufem\_2007 < migufem\_2002

b. migufem\_2007 > migufem\_2002

c. migufem\_2007 = migufem\_2002

**Tabla 2.4 Estadísticos de contraste**

	migufem_2007 - migufem_2002
Z	-7.358 <sup>a</sup>
<b>Sig. asintót. (bilateral)</b>	<b>.000</b>

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

Analizando las tablas 2.3 y la 2.4, se rechaza la hipótesis nula ya que el valor p es altamente significativo, se observa que hubo un aumento en la migración por hogar para el año 2007, es decir, que la suma de los rangos positivos es mayor que la de los negativos.

**Tabla 2.5 Prueba de Rangos de Wilcoxon para la característica de migración masculina a Estados Unidos**

		N	Rango promedio	Suma de rangos
migusmale_2007 - migusmale_2002	Rangos negativos	119 <sup>a</sup>	176.33	20983.00
	<b>Rangos positivos</b>	<b>259<sup>b</sup></b>	<b>195.55</b>	<b>50648.00</b>
	Empates	1165 <sup>c</sup>		
	Total	1543		

a. migusmale\_2007 < migusmale\_2002

b. migusmale\_2007 > migusmale\_2002

c. migusmale\_2007 = migusmale\_2002

**Tabla 2.6 Estadísticos de contraste<sup>b</sup>**

	migusmale_2007 - migusmale_2002
<b>*Z</b>	<b>-7.395<sup>a</sup></b>
Sig. asintót. (bilateral)	.000

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

En la prueba mostrada en las tablas 2.5 y 2.6 se observa el mismo caso de la migración femenina, es decir, no se acepta la hipótesis nula, lo cual indica que la mediana sea igual para las dos poblaciones, con la estadística de prueba se rechaza la hipótesis en cuestión, ya que el valor absoluto de \*z que es igual a 7.395, mayor que el valor de  $z_{.05}$  que en el planteamiento de la prueba se hizo mención de dicho valor. Ahora al analizar la suma de rangos se observa que hay un número mayor de rangos positivos, por lo que se acepta la hipótesis alternativa, donde indica que la migración masculina aumentó en el 2007 con respecto al 2002.

En conclusión analizando los tres casos, se observa que la migración a Estados Unidos aumentó para el año 2007, es decir que existieron diversos factores los cuales generaron mayor migración de parte de los miembros del hogar hacia Estados Unidos. Se analizarán las diferencias entre el 2002 y 2007 con respecto a la migración en el territorio mexicano, al igual que lo hecho con Estados Unidos se procederá a hacer el mismo estudio y los comentarios pertinentes para esta parte.

**Tabla 2.7 Prueba de Rangos de Wilcoxon para la característica de migración dentro del territorio mexicano**

		N	Rango promedio	Suma de rangos
migmx_h_2007 - migmx_h_2002	Rangos negativos	268 <sup>a</sup>	335.91	90024.50
	<b>Rangos positivos</b>	<b>463<sup>b</sup></b>	<b>383.42</b>	<b>177521.50</b>
	Empates	812 <sup>c</sup>		
	Total	1543		

a. migmx\_h\_2007 < migmx\_h\_2002

b. migmx\_h\_2007 > migmx\_h\_2002

c. migmx\_h\_2007 = migmx\_h\_2002

**Tabla 2.8 Estadísticos de contraste<sup>b</sup>**

	migmx_h_2007 - migmx_h_2002
<b>*Z</b>	<b>-7.852<sup>a</sup></b>
Sig. asintót. (bilateral)	.000

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

Para este caso analizando las tablas 2.7 y 2.8, el valor  $p$  altamente significativo, por lo que se rechaza la hipótesis nula donde se indica que las medianas para ambas muestras son iguales, del mismo modo si se analiza el valor absoluto de  $*z$  es igual a 7.852, mientras que la  $z_{.05} = 1.96$ , por lo tanto se refuerza el hecho de rechazar la hipótesis nula. Con respecto a la suma de los rangos se tiene que el

número de rangos positivos es de 177521.50 contra 90024.50, por lo que hubo un mayor número de migrantes por hogar para el 2007 que para el 2002.

**Tabla 2.9 Prueba de Rangos de Wilcoxon para la característica de la migración femenina dentro del territorio mexicano**

		N	Rango promedio	Suma de rangos
migmxfer_2007 -	Rangos negativos	158 <sup>a</sup>	238.39	37666.00
migmxfer_2002	<b>Rangos positivos</b>	<b>346<sup>b</sup></b>	<b>258.94</b>	<b>89594.00</b>
	Empates	1039 <sup>c</sup>		
	Total	1543		

a. migmxfer\_2007 < migmxfer\_2002

b. migmxfer\_2007 > migmxfer\_2002

c. migmxfer\_2007 = migmxfer\_2002

**Tabla 2.10 Estadísticos de contraste<sup>b</sup>**

	migmxfer_2007 - migmxfer_2002
<b>Z</b>	<b>-8.277<sup>a</sup></b>
Sig. asintót. (bilateral)	.000

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

En las tablas 2.8 y 2.10 se observan los resultados de la variable que indica el número de mujeres que migraron por hogar, se rechaza la hipótesis nula dado el estadístico de prueba y el valor p, es decir la mediana de la primera muestra es distinta a la de la segunda, del mismo modo al verificar los rangos, se observa que la suma de los rangos positivos es mayor que la de los negativos, por lo tanto se afirma que la mediana del 2007 es mayor que la del 2002.

**Tabla 2.11 Prueba de Rangos de Wilcoxon para la característica de la migración masculina dentro del territorio mexicano**

	N	Rango promedio	Suma de rangos
migmxmale_2007 - migmxmale_2002 Rangos negativos	207 <sup>a</sup>	235.69	48787.50
<b>Rangos positivos</b>	<b>292<sup>b</sup></b>	<b>260.15</b>	<b>75962.50</b>
Empates	1044 <sup>c</sup>		
Total	1543		

a. migmxmale\_2007 < migmxmale\_2002

b. migmxmale\_2007 > migmxmale\_2002

c. migmxmale\_2007 = migmxmale\_2002

**Tabla 2.12 Estadísticos de contraste<sup>b</sup>**

	migmxmale_2002 - migmxmale_2007
<b>Z</b>	<b>-4.431<sup>a</sup></b>
Sig. asintót. (bilateral)	.000

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

Para las tablas 2.11 y 2.12 se observa que el valor p es altamente significativo, por lo tanto se rechaza la hipótesis nula, no hay evidencia para decir que la mediana de los datos es igual. Analizando la suma de los rangos se observa que es mayor el número los positivos, por lo tanto se llega a la conclusión de que en el 2007 hubo mayor número de migrantes por hogar que en el 2002.

En conclusión para esta parte se tiene que con la prueba de rangos de Wilcoxon para todos los casos analizados, la migración aumentó para el año 2007, del mismo modo revisando los rangos y el valor p es claro que la migración en el último año de aplicación es mayor para todas las variables.

## LA PRUEBA T PARA MUESTRAS APAREADAS

La prueba t-Student se utiliza para contrastar hipótesis sobre medias en poblaciones con distribución normal. Esta prueba es apropiada para el contraste de medias de dos muestras dependientes cuando las varianzas de las poblaciones son desconocidas.

La prueba se aplica cuando dos poblaciones pareadas, o una población que ha sido medida dos veces. Dada una muestra aleatoria  $X_1, X_2, \dots, X_n$  con una distribución  $N(\mu_X, \sigma)$ , pareada con otra muestra aleatoria  $Y_1, Y_2, \dots, Y_n$  con una distribución  $N(\mu_Y, \sigma)$ . Las hipótesis a contrastar son las siguientes

Las hipótesis a contrastar es la siguiente:

$$H_0: \mu_X = \mu_Y \quad vs \quad H_1: \mu_X \neq \mu_Y$$

Donde la hipótesis nula indica que las medias de las dos poblaciones son iguales y la hipótesis alternativa indica que las medias de dichas poblaciones son distintas. Del mismo modo al contrastar con una sola cola se tiene la siguiente hipótesis alternativa.

$$H_1: \mu_X > \mu_Y$$

La media de la primera población es más grande que la media de la segunda población, esta hipótesis es evaluada en una prueba de una sola cola y sólo es considerada cuando el valor de la estadística de la prueba de t es positivo y mayor que  $t_{\alpha, k-1}$  que es el valor en la tablas a cierto nivel de significancia  $\alpha$  y con los  $k-1$  grados de libertad.

La hipótesis que se tiene es:

$$H_1: \mu_X < \mu_Y$$

Donde la media de la primera población es menor a la media de la segunda población en pruebas de una sola cola, esto solo se cumple cuando el valor de t

es negativo y el valor absoluto de  $t$  es igual o mayor que  $t_{\alpha, k-1}$  que es el valor en las tablas a cierto nivel de significancia  $\alpha$  y con los  $k-1$  grados de libertad.

Para el cálculo de la estadística de prueba se obtiene de la siguiente forma:

$$t = \frac{\bar{D}}{\bar{s}_D}$$

Donde:

$\bar{D}$  = Diferencia de medias de las dos poblaciones analizadas

$\bar{s}_D$  = Error estandar de la diferencia de medias

$$\bar{s}_D = \frac{S_D}{\sqrt{n}}$$

$n$  = Número de observaciones

Para calcular la desviación estándar de las diferencias:

$$S_D = \sqrt{\frac{\sum d^2}{n-1}}$$

$$\sum d^2 = \sum D^2 - \frac{(\sum D)^2}{n}$$

$$D = X_i - Y_i \quad \forall i = 1 \dots n$$

De tal forma a una confianza del 95% afirmamos que

$$\text{Si } t < t_{.05, k-1} \rightarrow \text{Se acepta } H_0$$

$$\text{Si } t > t_{.05, k-1} \rightarrow \text{Se rechaza } H_0$$

Así mismo obtenemos que:

*Si el valor  $p \geq \alpha \rightarrow$  Se acepta  $H_0$*

*Si el valor  $p < \alpha \rightarrow$  Se rechaza  $H_0$*

**Tabla 3.1 Prueba T para la característica de la migración a Estados Unidos**

	Diferencias relacionadas					t	Gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par migus_h2007 - 1 migus_h2002	.251	1.157	.029	.194	.309	8.541	1542	.000

En la tabla 3.1 se observa que el valor p es insignificante, es decir se rechaza la hipótesis nula donde las medias de las dos poblaciones son iguales, asimismo la estadística de prueba arroja una t positiva, por lo tanto la media del año 2007 es mayor que la del 2002, es decir se concluye que hubo un mayor número de individuos que migraron por hogar en el 2007.

**Tabla 3.2 Prueba T para la característica de la migración femenina a Estados Unidos**

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par migusfem_2007 - 1 migusfem_2002	.110	.612	.016	.080	.141	7.073	1542	.000

Analizando el resultado de la tabla 3.2, mostrada en la página anterior, se observa que el valor p es altamente significativo, por lo que se rechaza la hipótesis nula, es decir que la media de las dos poblaciones, en este caso la de la migración femenina por hogar es diferente en el 2002 y 2007, ahora si se verifica la estadística de prueba se tiene una t positiva, por eso es que la migración de mujeres por hogar fue mayor para el 2007 a comparación del 2002.

**Tabla 3.3 Prueba T para la característica de la migración masculina a Estados Unidos**

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par migusmale_2007 - l migusmale_2002	.141	.804	.020	.101	.181	6.901	1542	.000

Analizando la Tabla 3.3 se tiene que el valor p es altamente significativo, es decir se rechaza la hipótesis nula de que las medias de las dos poblaciones son iguales. Analizando la estadística de prueba se tiene una t positiva, por lo tanto se puede afirmar que la media del 2007 es mayor que la del 2002.

Ahora corresponde analizar la migración dentro del territorio nacional, comparar los datos obtenidos en el 2002 contra los del 2007, es decir se hará un análisis de comparación de medias, violando el supuesto de normalidad y se revisará si los resultados difieren o no.

**Tabla 3.4 Prueba T para la característica de la migración dentro del territorio mexicano**

	Diferencias relacionadas					T	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par 1 migmx_h_2007 - migmx_h_2002	.307	1.583	.040	.228	.386	7.621	1542	.000

Analizando la tabla 3.4, que se presenta en la hoja anterior, se observa que el valor p altamente significativo, es decir, se rechaza la hipótesis nula, donde se especifica que la media de los datos son iguales, al verificar la estadística de prueba se tiene una t positiva, la media de la migración en el 2007 es mayor que en el 2002, esto quiere decir que hubo un mayor número de individuos que migraron por hogar en el año 2007.

**Tabla 3.5 Prueba T para la característica de la migración femenina dentro del territorio mexicano**

	Diferencias relacionadas					T	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par 1 migmx_fem_2007 - migmx_fem_2002	.205	.993	.025	.156	.255	8.129	1542	.000

Analizando la variable migración femenina dentro del territorio nacional, en la tabla 3.5 se observa que el valor p es altamente significativo, es decir se rechaza la hipótesis nula donde se asevera que la media de ambas poblaciones es igual. La

estadística de prueba es positivo por lo tanto la migración femenina por hogar es mayor para el 2007.

**Tabla 3.6 Prueba T para la característica de la migración masculina dentro del territorio mexicano**

	Diferencias relacionadas					T	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par migmxmale_2007 - l migmxmale_2002	.102	.922	.023	.056	.148	4.335	1542	.000

En la tabla 3.6 el valor p es altamente significativo, es decir, se rechaza la hipótesis nula a una confianza del 95%, con la estadística de prueba se afirma que la media de la migración es mayor para el 2007 que para el 2002.

Comentando los resultados arrojados es evidente que con la prueba T, la media no es la misma para ninguno de los seis casos, es decir para el 2007 hubo un aumento de la migración, al comparar los resultados con la prueba de rangos de Wilcoxon es evidente que no hay ningún cambio para todos los casos analizados por lo que se puede considerar que el supuesto de normalidad en este caso no fue determinante para modificar dichos resultados.

## CAPÍTULO IV

### ANÁLISIS DE LA INDEPENDENCIA DE LAS MUESTRAS

A lo largo de este proyecto se ha señalado que esta encuesta está bajo un diseño de medidas repetidas, es decir, un diseño tipo panel por lo que se debe de tener cuidado con su análisis y con los supuestos que se hacen sobre las observaciones, es por ello que en este caso y para los demás las muestras no pueden ser independientes, es decir se tiene que asumir dos muestras dependientes. Se siguió con el mismo diseño muestral y el mismo número de hogares para las dos rondas, del mismo modo para una encuesta tipo panel se tiene que el cálculo de la varianza del estimador de la diferencia está dado por la siguiente expresión:

$$V(Y_{t2}-Y_{t1}) = V(Y_{t2}) + V(Y_{t1}) - 2COV(Y_{t1}, Y_{t2})^9$$

Donde el número de migrantes en el año 2002 está definido por  $Y_{t1}$  y en el año 2007 por  $Y_{t2}$ .

De la expresión anterior se deduce que si el valor de la covarianza es positivo, entonces la varianza estimada para muestra panel será menor que la proveniente de dos mediciones independientes.

Si se utilizan muestras independientes el estimador de la diferencia sería:

$$V(Y_{t2}-Y_{t1}) = V(Y_{t2}) + V(Y_{t1})^{10}$$

Pero qué pasaría si se asume que existe independencia en los datos, lo que se pretende es analizar si este hecho es determinante para tener resultados distintos.

---

<sup>9</sup>Cfr: Medina, Fernando, "Las encuestas panel" pp. 139

<sup>10</sup>Idem

Se aplicará la U de Mann-Whitney como prueba no paramétrica para muestras independientes, a continuación se explicará las características de esta prueba.

### PRUEBA DE MANN WHITNEY

La prueba de Mann Whitney es empleada con datos de carácter ordinal donde se envuelve un diseño de dos muestras independientes, lo que se compara con esta prueba son las medianas de las dos poblaciones, las pruebas de hipótesis que se contrastan son las siguientes:

Sea  $X_1, \dots, X_n$  y  $Y_1, \dots, Y_n$  dos muestras independientes

Se define a  $\theta_1$  como la mediana de la población  $X_i (i = 1 \dots n)$  y  $\theta_2$  como la mediana de la población  $Y_i (i = 1 \dots n)$

$$H_0: \theta_1 = \theta_2 \quad vs \quad H_1: \theta_1 \neq \theta_2$$

La hipótesis nula indica que la mediana de las dos poblaciones comparadas es igual, contra la hipótesis alternativa que dice que la mediana de las dos poblaciones a comparar es distinta, lo cual se traduce a que la suma de los rangos positivos sea diferente a la suma de los rangos negativos y por lo tanto las medianas de las dos poblaciones a comparar sean distintas.

También se tiene que si:  $\sum R_1 > \sum R_2$  entonces  $H_1: \theta_1 > \theta_2$  esta es la hipótesis alternativa para pruebas de una cola. Es decir que la mediana de la primera población es mayor que la de la segunda.

Del mismo modo para el caso donde:  $\sum R_1 < \sum R_2$  entonces  $H_1: \theta_1 < \theta_2$ , por lo que se concluye que la mediana de la segunda población es mayor que la de la primera.

Para la obtener los rangos se sigue de esta manera:

- Se determina los valores  $n_1$  y  $n_2$  donde  $n_1$  es el número de observaciones de la primera muestra y  $n_2$  el de la segunda muestra.

- Se ordenan juntos los puntajes de ambos grupos, asignando el rango 1 al puntaje que se sea algebraicamente más bajo. Se hace la suma los rangos de cada grupo por separado.

$$\sum R_1 = \text{Suma de rangos de la primera poblacion}$$

$$\sum R_2 = \text{Suma de rangos de la segunda poblacion}$$

El estadístico U se obtiene de la siguiente manera:

$$U = n_1 n_2 + \frac{n_x(n_x + 1)}{2} - \sum R_x$$

Donde:

$n_1, n_2 =$  Número de casos de cada grupo

$$\sum R_x = \text{Suma de rangos del grupo correspondiente}$$

$n_x =$  Número de casos del grupo correspondiente

Con muestras demasiados grandes se tienen la siguiente fórmula para el estadístico de prueba:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

De tal forma que a una confianza del 95% afirmamos que:

$Si Z < Z_{.05} \rightarrow$  Se acepta  $H_0$

$Si Z > Z_{.05} \rightarrow$  Se rechaza  $H_0$

Así mismo obtenemos que:

*Si el valor  $p \geq \alpha \rightarrow$  Se acepta  $H_0$*

*Si el valor  $p < \alpha \rightarrow$  Se rechaza  $H_0$*

**Tabla 4.1 Prueba de ManWhitney para la característica de migración a Estados Unidos**

	Survey	N	Rango promedio	Suma de rangos
migus	2002	1543	1484.50	2290577.50
	2007	1543	1602.50	2472663.50
	Total	3086		

**Tabla 4.2 Estadísticos de contraste**

	Migus
U de Mann-Whitney	1099381.500
W de Wilcoxon	2290577.500
Z	-4.721
Sig. asintót. (bilateral)	.000

En la tabla 4.1 y 4.2 se observa que el valor p es altamente significativo por lo que se rechaza la hipótesis nula, ahora si se toma en cuenta la suma de los rangos, se observa que para la segunda encuesta la suma es mayor, es decir, la mediana de los datos de la encuesta 2007 es mayor que para 2002, esto se traduce que en el 2007 hubo mayor número de migrantes por hogar que en el 2002 hacia los Estados Unidos.

**Tabla 4.3 Prueba de ManWhitney para la característica de migración femenina a Estados Unidos**

	survey	N	Rango promedio	Suma de rangos
migusfem	2002	1543	1500.56	2315359.00
	2007	1543	1586.44	2447882.00
	Total	3086		

**Tabla 4.4 Estadísticos de contraste**

	Migusfem
U de Mann-Whitney	1124163.000
W de Wilcoxon	2315359.000
Z	-4.903
Sig. asintót. (bilateral)	.000

Analizando el caso de la migración femenina hacia los Estados Unidos en la tablas 4.3 y 4.4 se observa que la hipótesis nula se rechaza ya que el valor p es altamente significativo, ahora comparando la suma de los rangos se tiene que en 2007 es mayor la suma de ellos, por lo que se concluye que la migración femenina por hogar para el 2007 aumentó, es decir, hubo mayor número de mujeres por hogar que se fueron a los Estados Unidos en este último año que en el 2002.

**Tabla 4.5 Prueba de MannWhitney para la característica de migración masculina a Estados Unidos**

	Survey	N	Rango promedio	Suma de rangos
migusmas	2002	1543	1493.12	2303884.00
	2007	1543	1593.88	2459357.00
	Total	3086		

**Tabla 4.6 Estadísticos de contraste**

	Migusmas
U de Mann-Whitney	1112688.000
W de Wilcoxon	2303884.000
Z	-4.196
Sig. asintót. (bilateral)	.000

Para el caso de la migración masculina hacia los Estados Unidos en las tablas 4.5 y 4.6 se observa que el valor p es altamente significativo, es decir, se rechaza la hipótesis nula, la cual afirma que la mediana de la migración en el 2002 sea igual a la del 2007, ahora al verificar la suma de los rangos se observa que para el año 2007 es mayor, por lo que se llega a la conclusión que en este último año hubo una mayor migración masculina hacia los Estados Unidos que en el 2002.

**Tabla 4.7 Prueba de Mann Whitney para la característica de migración dentro del territorio mexicano**

survey	N	Rango promedio	Suma de rangos
migmx 2002	1543	1468.98	2266636.50
2007	1543	1618.02	<b>2496604.50</b>
Total	3086		

**Tabla 4.8 Estadísticos de contraste**

	migmx
U de Mann-Whitney	1075440.500
W de Wilcoxon	2266636.500
Z	-5.182
Sig. asintót. (bilateral)	.000

En las tablas 4.7 y 4.8 se verifican los resultados de la variable migración hacia los Estados Unidos, se observa que el valor p es altamente significativo, es decir, se rechaza la hipótesis nula donde se asevera que las medianas para los dos años son iguales, ahora si se analizala suma de los rangos, se tiene que para el 2007 es mayor que para el 2002, por lo que se llega al resultado que en el 2007 hubo mayor número de migrantes por hogar dentro del territorio mexicano.

**Tabla 4.9 Prueba de Mann Whitney para la característica de migración femenina dentro del territorio mexicano**

	Survey	N	Rango promedio	Suma de rangos
Migmx fem	2002	1543	1466.88	2263392.00
	2007	1543	1620.12	2499849.00
	Total	3086		

**Tabla 4.10 Estadísticos de contraste**

	migmx fem
U de Mann-Whitney	1072196.000
W de Wilcoxon	2263392.000
Z	-5.956
Sig. asintót. (bilateral)	.000

Analizando la migración femenina, presentada en la pagina anterior la información, dentro del territorio nacional, en las tablas 4.9 y 4.10 se puede ver que sigue el mismo comportamiento que en casos anteriores, donde el valor p es altamente significativo, es decir, que se rechaza la hipótesis nula, ahora si se verifica la suma de los rangos se observa que es mayor para el 2007, por lo que se llega a la conclusión de que en este año la mediana de los datos es mayor, es decir que hubo mayor migración de mujeres por hogar en el 2007.

**Tabla 4.11 Prueba de Mann Whitney para la característica migración masculina dentro del territorio nacional**

	Survey	N	Rango promedio	Suma de rangos
Migmx masc	2002	1543	1506.57	2324633.00
	2007	1543	1580.43	2438608.00
	Total	3086		

**Tabla 4.12 Estadísticos de  
contraste**

	migmxmasc
U de Mann-Whitney	1133437.000
W de Wilcoxon	2324633.000
Z	-2.877
Sig. asintót. (bilateral)	.004

Del mismo modo que en el análisis anterior en la tablas 4.11 y 4.12 es claro que la hipótesis nula se rechaza, ya que el p valor es muy significativo, por lo tanto al verificar la suma de los rangos se tiene que es mayor para 2007 por lo que se concluye que la mediana es mayor, es decir, hubo mayor numero de migrantes masculinos por hogar en el 2007.

Hasta aquí concluye el análisis de las variables seleccionadas con las distintas pruebas que se decidieron aplicar, del mismo modo se observa los cambios en la migración durante este periodo, es decir se concluye que para todos los casos en el 2007 aumentó.

## CONCLUSIONES

- Al observar los resultados es evidente que el supuesto de normalidad, que se comprobó con la prueba de Kolmogorov Smirnov, de las variables elegidas de la ENHRUM I y ENHRUM II no es fuerte, ya que al realizar los estudios con la prueba de Wilcoxon y la prueba T, el resultado condujo a la misma conclusión. Es decir, que para las seis variables seleccionadas no se modificó el resultado en ningún aspecto a pesar de que éstas no cumplieron con el supuesto señalado.
- Se afirma que los resultados obtenidos no muestran modificaciones si se utiliza estadística paramétrica o estadística no paramétrica, encontrándose, para el caso específico, que la migración para el año 2007 aumentó en comparación con 2002 en las seis variables estudiadas.
- Ante esta situación, donde los datos no cumplen con algún supuesto, se muestra que es posible abordar de distintas formas el problema, se señala que se puede proceder a utilizar estadística no paramétrica o hacer una transformación de los datos. En el caso de esta tesis se optó por utilizar estadística paramétrica, resaltándose que es una opción viable aunque se viole algún supuesto. Lo anterior es evidencia de que no hay que ser tan rigurosos y determinantes ante el incumplimiento de algún supuesto estadístico, pues, al trabajar con datos reales es muy probable encontrarse con problemas de este tipo. Si bien, al trabajar con estadística paramétrica y trabajar con alguna violación de algún supuesto se debe de tomar con más cautela los resultados pero no descartarlos de inmediato.
- Esta investigación fue un ejercicio para verificar la rigurosidad de las pruebas, ya que, a lo largo de la formación como actuario al aplicar pruebas se supone que éstas se pueden aplicar si y solo si cumplen con ciertos supuestos, excluyendo de antemano las demás opciones. Sin embargo este trabajo muestra que no necesariamente se debe de cumplir con todos los requisitos para aplicar pruebas estadísticas de este tipo.

- Del mismo modo al analizar la independencia (la cual no se cumple por tener observaciones tipo panel, ya que la encuesta del 2002 no puede ser independiente a la del 2007), se decidió aplicar la prueba de Mann Whitney, la cual es usada para muestras independientes. Como resultado de esta aplicación, no se presentó algún cambio en la tendencia ni en los resultados de la migración hacia los Estados Unidos y dentro del territorio nacional, de igual forma que con la suma de rangos de Wilcoxon es decir los resultados obtenidos para el análisis de la mediana de la migración en el 2002 y 2007 fueron diferentes para los dos años. Del mismo modo para la prueba de comparación de medias, se observó el mismo resultado que en las pruebas anteriores, por lo que se concluye que la influencia del supuesto de normalidad, así como el de independencia de las muestras no fue tan fuerte en este caso, ya que se llegaron a los mismos resultados con tres pruebas distintas.
- En suma, esta tesis permite al investigador darse cuenta, percatarse que no es necesario ser tan riguroso en el momento de los análisis ni de los supuestos para esta encuesta. Para este caso se puede afirmar que con las variables tomadas y las pruebas utilizadas es posible seguir distintos caminos para poder realizar un análisis, es decir no ser tan rigurosos con los supuestos, ya que, lo más probable, dada la evidencia presentada, es que se llegue al mismo resultado. Obviamente es necesario tomar con cautela dichos resultados cuando los supuestos no se cumplen, haciendo las pruebas pertinentes.
- Ahora bien, se resalta que trabajar con datos tipos panel genera cierto tipo de inconvenientes. Como se vio a través de este proyecto, el desgaste del panel es un factor que complica el tratamiento de los datos y el análisis, pues, se van perdiendo observaciones y el propósito de hacer una comparación va careciendo de sentido. Como se observó en el caso de estudio, lo que se tuvo que hacer fue mantener las observaciones que estaban en la segunda encuesta para poder realizar la comparación con

mayor sentido, dejando de lado las observaciones que se perdieron por el desgaste.

- Es importante tomar en cuenta que este proyecto se ha llevado a cabo en varias etapas y el hecho de juntar dos encuestas donde se aplicó el diseño panel fue muy complicado, pues, hubo ciertas modificaciones dentro del cuestionario para las dos rondas, lo cual provocó cierta dificultad para trabajar con ciertas variables y es por eso que se hizo la elección de las que sólo se presentan en el estudio.
- Del mismo modo, con este trabajo se entiende lo complicado que es llevar una encuesta, desde su levantamiento, captura, revisión y congruencia de los datos, pues, a lo largo de este proceso es inevitable perder información. Asimismo, como se planteó al principio de esta tesis, no se puede asegurar que la información dada por el encuestado sea la correcta, es por eso que se vuelve complejo y aún más, con este tipo de diseños panel, ya que hay que darle seguimiento a la información a través del tiempo.
- Asimismo, la realización del panel es una tarea compleja ya que se tienen que recodificar las variables para que sea congruente la información. Esta labor resulta complicada ya que, de los cuestionarios de ambas encuestas se tienen que ir eligiendo variables y códigos concretos. Además, para el año 2010 se volvió a aplicar esta encuesta con el propósito de verificar que tanto había afectado la crisis económica a México. Desafortunadamente no se pudo tomar como referencia este levantamiento para este proyecto, ya que la base de datos estaba incompleta y aun faltaba su validación. A continuación se presenta un cuadro resumen donde se maneja el resultado de cada prueba, teniendo en cuenta las 6 variables analizadas en su totalidad.

## CUADRO DE CONCLUSIONE

<b>PRUEBA</b>	<b>P-VALOR</b>	<b>ACEPTACION O RECHAZO DE LA HIPOTESIS NULA</b>	<b>SE CONCLUYE QUE:</b>
<b>Kolmogorov-Smirnov</b>	Altamente significativo	Se rechaza la hipótesis nula	Los datos no presentan normalidad
<b>Suma de Rangos de Wilcoxon</b>	Altamente significativo	Se rechaza la hipótesis nula	En el 2007 hubo mayor numero de migrantes por hogar
<b>Prueba T</b>	Altamente significativo	Se rechaza la hipótesis nula	En el 2007 hubo mayor numero de migrantes por hogar
<b>Pueba de Mann Whitney</b>	Altamente significativo	Se rechaza la hipótesis nula	En el 2007 hubo mayor numero de migrantes por hogar

## **ANEXO I**

### **DISEÑO MUESTRAL PARA UNA ENCUESTA TIPO PANEL Y EN ESPECÍFICO PARA LA ENHRUM**

Existen algunos aspectos del diseño de una encuesta longitudinal y en específico de las encuestas tipo panel que son definidos particularmente para éstas, un poco diferente con respecto a las encuestas transversales, ya que la información debe de ser obtenida de una manera más cuidadosa y con un diseño más adecuado, para eso se debe de tomar en cuenta algunos aspectos para una encuesta de este estilo, ya que el tipo de muestreo debe de ser cuidadoso y específico.

#### **TAMAÑO DE LA MUESTRA**

Como primera consideración se debe de tener en cuenta la definición de la población ya que la recolección de datos toma lugar en varios puntos del tiempo para cada unidad muestreada. Por lo que los datos son definidos por ellos mismos a través de los distintos periodos. Es necesario un plan para el tamaño de la muestra requerida para estudios longitudinales, ya que si tomamos en cuenta el desgaste de la encuesta y otros factores por los cuales hay una modificación de la muestra.

#### **CONGLOMERACIÓN**

La conglomeración muchas veces son decisiones de la naturaleza así que la medida de una muestra puede depender de los diferentes criterios para los cuales se quiera aplicar el análisis longitudinal. Específicamente la relación existente entre la agrupación y la colección de datos puede cambiar sobre los levantamientos de cada encuesta debido a la movilidad y al desgaste de la muestra, el impacto de conglomerar sobre la varianza de las estimaciones puede también cambiar, si tomamos en cuenta que en el diseño las etapas requieren un conocimiento del numero y tiempo de los levantamientos.

La identificación del nivel óptimo de conglomeración de una muestra es aún más complicado con paneles de largo plazo especialmente si la duración de la vida del panel es desconocida al tiempo de plantear el diseño, de ahí para paneles de larga duración es no recomendable hacer un muestreo por conglomerados para que la precisión en años posteriores pueda transformar el esfuerzo adicional que se hizo en los años.

## **ESTRATIFICACIÓN**

Normalmente para muestreo estratificado debería estar relativamente definido como homogéneo, en términos de las variables clave en una encuesta longitudinal, los objetivos principales por inferencia pueden ser medidos de manera individual o como una variación en conjunto, ambos general y por dominios por los cuales la estratificación transversal estándar puede ser un poco de ayuda. Si se considera una estratificación se toma una muestra inicial para recopilar información sobre la base de esta muestra se asignan las unidades de la muestra a los “L” estratos que se desean hacer.

Ponderar, consiste en darle a cada unidad muestral de una encuesta un valor numérico (Peso), representando la contribución que hará la unidad a las estimaciones basadas en los datos de la encuesta. Los pesos han sido diseñados para que la muestra pueda ser representativa de la población de estudio. El peso de cada unidad de respuesta particular se puede interpretar como el número relativo de unidades de población que representa. El cálculo y la aplicación de los pesos es parte del proceso de inferencia estadística, por la cual se pueden sacar conclusiones sobre una población de interés sobre la base de las informaciones de una muestra obtenida de la población.

Existen tres razones principales para realizar la ponderación:

- Corregir las diferencias de las tasas de cobertura
- Para elegir las probabilidades y probabilidades de respuesta (y falta de respuesta)

- Para corregir los efectos de las varianzas de muestreo aleatorio

El objetivo básico de la ponderación es asegurar que los grupos de las unidades de muestreo estén representados en la misma proporción en la que aparecen dentro de la población. Estos factores sirven para cualquier encuesta. Los rasgos de la ponderación que son distintivos en el caso de las encuestas longitudinales tienen que ver principalmente con la definición de la población de estudio y con el tratamiento de los patrones de falta de respuesta complejos. En el caso de algunas encuestas longitudinales, pueden ser necesarios también algunos métodos especiales para estimar las probabilidades de selección.

Muchas encuestas longitudinales proporcionan dos tipos de pesos a los usuarios de datos: pesos transversales pesos longitudinales. Esta distinción refleja una diferencia en la población que se ha de representar, lo cual, a su vez, está relacionado con diferentes objetivos de estimación.

Un problema común de las encuestas longitudinales es encontrar una buena manera de estimar la distribución de la población, entre las variables relevantes, para cada una de las poblaciones del estudio. Las estimaciones transversales de la población se pueden conseguir fácilmente de fuentes externas, pero casi nunca es el caso de las poblaciones longitudinales. Normalmente, un enfoque basado en un modelo se usa para estimar las características de las poblaciones longitudinales, tomando la población transversal del primer ciclo como punto de partida y después utilizando estimaciones basadas en muestras de los nacimientos y muertes para dar forma a los cambios de la estructura de la población. Esto puede ser dudoso, dado que es probable que las estimaciones que se basan en muestras estén sujetas a errores de falta de respuesta, que es una de las fuentes de error cuyos errores se corrijan bajo las ponderaciones.

Tomando en cuenta estos factores anteriores plantearemos el diseño muestral de la ENHRUM:

El diseño muestral de la ENHRUM se presentó de la siguiente forma:

- Probabilístico
- Estratificado
- Multietápico
- Conglomerados

Para la estratificación se dividió al país en 5 regiones las cuales son centro, sur-este, norte, occidente, sur, posteriormente se paso a realizar cuatro etapas de selección en las que se conglomeró a sus unidades correspondientes, se pudo obtener la muestra que formaría la encuesta, es decir el número de hogares deseados.

Sabiendo que el marco de muestreo constituye el marco de referencia o sitio donde supuestamente se pueden localizar a todas las unidades de la población para dicha encuesta se tomo en cuenta:

- El censo de población y vivienda del 2000 para localidades entre 500 y 2499 habitantes
- Cartografía fue otorgada por el INEGI
- El equipo especializado de la ENHRUM realizó el listado de viviendas de casa localidad, previamente a su levantamiento

Para el tamaño de la muestra se considero la variable ingreso corriente definida en el ENIGH 2000 de la siguiente forma:

Ingreso Corriente Total: Percepciones en efectivo y/o en especie que recibieron los miembros del hogar durante el período de referencia cambio de la venta de su fuerza de trabajo a una empresa, institución o patrón, así como el ingreso en efectivo y/o en especie de un negocio agropecuario o bien el ingreso en efectivo de un negocio no agropecuario o los rendimientos derivados de cooperativas de producción, propiedad de algún miembro del hogar, se incluyeron los alquileres,

intereses, dividendos y regalías derivados de la posesión de activos físicos o no físicos, las transferencias recibidas que no constituyeron un pago por su trabajo y otros ingresos corrientes percibido.

Tomando en consideración el tamaño de muestra utilizamos la siguiente expresión de muestreo aleatorio simple:

$$n_0 = \frac{s^2 z^2}{r^2 \bar{x}^2}$$

Donde:

$s^2 =$  *varianza poblacional*

$z =$  *valor de las tablas de la distribución normal a una confianza del 90%*

$\bar{x} =$  *Media poblacional*

$r =$  *Error máximo esperado*

Ahora si se pondera  $n_0$  con el Promedio de Hogares por Vivienda (PHV) se tiene que:

$$n_1 = \frac{n_0}{PVH} \rightarrow n_1 = \frac{\frac{s^2 z^2}{r^2 \bar{x}^2}}{PVH} \rightarrow n_1 = \frac{s^2 z^2}{r^2 \bar{x}^2 PVH}$$

Como se está hablando de un diseño diferente al de muestreo aleatorio simple se tiene que tomar en cuenta el efecto de diseño que es el cociente entre la varianza del diseño y la varianza del muestreo aleatorio simple por lo que:

$$n_2 = \frac{s^2 z^2}{r^2 \bar{x}^2 PVH} * DEFF$$

La tasa de no respuesta se obtiene bajo la siguiente expresión:

$$n = \frac{s^2 z^2}{r^2 \bar{x}^2 PVH(1 - tnr)} * DEFF$$

Se consideró:

- 90% de confianza
- Efecto de diseño de 3.9

Ahora considerando los datos tomados por la ENIGH 2002:

- Varianza poblacional 91902770.54
- Ingreso promedio corriente por hogar 6364.87
- Error relativo máximo esperado 15%
- Promedio de hogares por vivienda 1.02
- Tasa de no respuesta 15%
- Muestra de 1227 que se ajusta a 1600 a nivel nacional

La selección de la muestra se realizó de manera independiente en cada región y el procedimiento de selección es como a continuación se describe:

1. Se seleccionaron en cada una de las regiones  $k$  entidades con probabilidad proporcional al tamaño, empleando muestreo sistemático<sup>2</sup> con arranque aleatorio.
2. En cada entidad seleccionada se seleccionaron  $s$  localidades con probabilidad proporcional al tamaño, empleando muestreo sistemático.
3. De cada localidad seleccionada se seleccionaron  $V_{rei}$  viviendas con probabilidad igual empleando muestreo sistemático con arranque aleatorio.

4. En cada vivienda seleccionada se seleccionó un hogar con igual probabilidad.

Para el punto 1, 2 y 3 se utilizó el muestreo sistemático aleatorio de arranque se entiende como la numeración de todos los elementos de la población y de ahí solo se selecciona un solo elemento. Es decir se parte de un número aleatorio  $X$ , que es el número elegido al azar y de ahí se van seleccionando los demás elementos de la muestra de acuerdo a un número determinado  $i$ , así el primer elemento de la muestra ocupara el lugar  $X$ , el segundo  $X + i$ , el tercero  $X + 2i$  y así hasta llegar al elemento  $X + 2(n - 1)i$ .

De esta manera se obtienen cuatro unidades de muestreo en cuatro etapas que se presentan de la siguiente forma:

- a) Unidades de muestreo de primera etapa (UPM) son las entidades federativas.
- b) Unidades de muestreo de segunda etapa (USM) son las localidades.
- c) Unidades de muestreo de tercera etapa (UTM) son las viviendas.
- d) Unidades de muestreo de cuarta etapa (UCM) son los hogares

Para la selección de la Tercera Unidad de Muestreo (UTM) que en este caso es la vivienda se partió de lo siguiente:

La probabilidad de seleccionar una vivienda en la  $i$ -ésima localidad, de la  $e$ -ésima entidad, en la  $r$ -ésima región es:

$$P[V_{rei}] = \frac{K m_{re}}{m_r} \frac{s m_{rei}}{m_{re}} \frac{V_{rei}}{m_{rei}} = \frac{K s m_{rei} V_{rei}}{m_r m_{rei}}$$

$V_{rei}$  = Es la vivienda seleccionada en la  $i$ -ésima localidad, de la  $e$ -ésima entidad, en la  $r$ -ésima región

$k$  = Son las entidades seleccionadas por región.

$s$  = Son las localidades seleccionadas por entidad

$m_r$  = Es el total de viviendas en la  $r$ -ésima región, según el XII Censo de Población y vivienda 2000.

$m_{re}$  = Es el total de viviendas en la  $e$ -ésima entidad seleccionada, de la  $r$ -ésima región, según el XII Censo de Población y Vivienda 2000.

$m_{rei}$  = Es el total de viviendas en la  $i$ -ésima localidad seleccionada, de la  $e$ -ésima entidad seleccionada, en la  $r$ -ésima región, según el XII Censo de Población y Vivienda 2000.

$m_{rei}$  = Es el total de viviendas en la  $i$ -ésima localidad seleccionada, de la  $e$ -ésima entidad, en la  $r$ -ésima región, al momento de realizar la entrevista.

Para la selección de la Cuarta Unidad de Muestreo que en este caso son los hogares se procedió de la siguiente forma:

Se multiplica la probabilidad de seleccionar una vivienda en la  $i$ -ésima localidad, de la  $e$ -ésima entidad, en la  $r$ -ésima región por el inverso del número de hogares en las viviendas seleccionadas

$$P[V_{reij}] = P[V_{rei}] \frac{1}{H} = \frac{K m_{re}}{m_r} \frac{s m_{rei}}{m_{re}} \frac{V_{rei}}{m_{rei}} \frac{1}{H} = \frac{K s m_{rei} V_{rei}}{H m_r m_{rei}}$$

#### **AJUSTE DE LOS FACTORES DE EXPANSIÓN**

El factor de expansión es el inverso de la probabilidad de inclusión o la probabilidad de selección. Este factor determina a cuántos elementos de la población está representando nuestra unidad muestreada.

Para esta encuesta el factor de expansión para vivienda está determinado por:

$$F_{rei} = \frac{m_r m_{rei}}{K s m_{rei} V_{rei}}$$

Donde:

$F_{rei}$  = Factor de expansión de la la i-esima localidad, de la e-ésima entidad, en la r-ésima región

Del mismo modo el factor de expansión para los hogares esta determinado por:

$$F_{reij} = \frac{Hm_r m_{rei}}{Ksm_{rei} V_{rei}}$$

$F_{reij}$  = Factor de expansión de la la i-esima localidad, de la e-ésima entidad, en la r-ésima región en la j-ésima vivienda

El ajuste de los factores de expansión o pesos muestrales se realizan para tener una mayor precisión en el momento de proyectar, es decir se toma en cuenta un parámetro ya conocido y se crea una nueva expresión recreando el factor de expansión, en el caso de ENHRUM se hicieron dos tipos de ajuste para, el primero fue el ajuste por no respuesta y el segundo fue el ajuste por proyección.

#### **AJUSTE POR NO RESPUESTA**

Para realizar el ajuste de factor de expansión por no respuesta se utilizó el cociente de las viviendas seleccionada entre las viviendas encuestadas, dicho ajuste se hizo a nivel de localidad seleccionada mediante la siguiente expresión:

$$F_{reij}^* = F_{reij} * \frac{V_{rei}}{V_{rei}^*}$$

$F_{reij}^*$  = Factor de expansión corregido por no respuesta para los hogares de la j-esima vivienda de la i-esima localidad, de la e-esima entidad, en la r-esima región en la j-esima vivienda.

$F_{reij}$  = Factor de expansión de la la i-ésima localidad, de la e-ésima entidad, en la r-ésima región en la j-ésima vivienda

$V_{rei}$  = Número de viviendas habitadas seleccionadas en la i-ésima localidad, de la e-ésima entidad, en la r-ésima región.

$V_{rei}^*$  = Número de viviendas habitadas seleccionadas con respuesta en la i-ésima localidad, de la e-ésima entidad, en la r-ésima región.

### **AJUSTE DEL FACTOR DE PROYECCIÓN**

Ajustado nuestro Factor de expansión por no respuesta se procede a realizar un segundo ajuste por proyección para poder asegurar la población total determinada por la proyección de población generada por el INEGI referida al punto medio del levantamiento.

$$F_{reij}^{**} = F_{reij}^* * \frac{Y_r}{\hat{Y}_E}$$

$F_{reij}^{**}$  = Factor final corregido por proyección.

$F_{reij}^*$  = Factor de expansión corregido por no Respuesta.

$Y_r$  = Población en la r-ésima región, según la proyección.

$\hat{Y}_E$  = Población en la r-ésima región, a la que expande la encuesta.

Ahora para el cálculo de los estimadores que se emplearon dos tipos, los totales y los de razón:

#### *Estimadores*

Los estimadores que se pueden obtener son de totales y de razón.

El estimador del total de la característica X es:

$$\hat{X} = \sum_{r=1}^5 \sum_{e=1}^k \sum_{i=1}^s \sum_{j=1}^{V_{rei}} F_{reij}^{**} * \hat{X}_{reijh}$$

Para esta expresión se está sumando primero sobre el número de viviendas habitadas seleccionadas en la  $i$ -ésima localidad, de la  $e$ -ésima entidad, en la  $r$ -ésima región. Luego sobre las localidades, las  $k$  entidades y la última suma es sobre los 5 estratos.

La estimación de proporciones y promedios se realiza por medio de los estimadores de razón, los cuales son de la siguiente forma:

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}$$

Donde  $Y$  se estima de forma similar a  $X$

$\hat{X}_{reijh}$  = Es el valor observado de la característica  $X$ , en el  $h$ -ésimo hogar, de la  $j$ -ésima vivienda, de la  $i$ -ésima localidad, de la  $e$ -ésima entidad, en la  $r$ -ésima región.

$\hat{X}$  = Estimador total de la característica  $X$

$\hat{Y}$  = Estimador total de la característica  $Y$

### **ESTIMADORES DE LA PRECISIÓN**

Para la evaluación de los errores de muestreo de las principales estimaciones se usó el método de Conglomerados Últimos, basado en que la mayor contribución a la varianza de un estimador, en un diseño polietápico es la que se presenta entre las unidades primarias de muestreo (UPM). El término "Conglomerados Últimos" se utiliza para denotar el total de unidades en muestra de una unidad primaria de muestreo.

Las expresiones de la varianza, el error estándar, el coeficiente de variación y el efecto de diseño para los estimadores del total y de razón son las siguientes:

Estimadores totales:

$$\hat{V}(\widehat{X}) = \sum_{r=1}^5 \frac{k}{k-1} \sum_{e=1}^k (\hat{X}_{re} - \hat{\bar{X}}_r)^2$$

$$E.S = \sqrt{\hat{V}(\widehat{X})}$$

$$C.V = \frac{\sqrt{\hat{V}(\widehat{X})}}{\hat{X}}$$

$$DEFF = \frac{\hat{V}(\widehat{X})}{\hat{V}(\widehat{X})_{MAS}}$$

Estimadores para las razones:

$$\hat{V}(\widehat{R}) = \frac{1}{Y^2} \sum_{r=1}^5 \frac{k}{k-1} \sum_{e=1}^k \left( \hat{X}_{re} - \hat{\bar{X}}_r \frac{1}{K} \right) - \hat{R} \left( \hat{Y}_{re} - \hat{Y}_r \frac{1}{k} \right)$$

$$E.S = \sqrt{\hat{V}(\widehat{R})}$$

$$C.V = \frac{\sqrt{\hat{V}(\widehat{R})}}{\hat{R}}$$

$$DEFF = \frac{\hat{V}(\widehat{R})}{\hat{V}(\widehat{R})_{MAS}}$$

Donde:

$\hat{X}_{re}$  = Es la estimación del valor total de la variable X en la e-ésima entidad, de la r-ésima región.

$\hat{\bar{X}}_r$  = Estimación del promedio de la variable X en la r-ésima región.

$\hat{Y}_{re}$  = Estimación del valor total de la variable Y, en la e-ésima entidad, de la r-ésima región.

$\hat{V}(\widehat{X})_{MAS}$  = Estimador de la varianza del muestreo aleatorio simple.

$\hat{V}(\widehat{R})_{MAS}$  = Estimador de la varianza del estimador de razón empleando muestreo aleatorio simple.

## BIBLIOGRAFÍA

Arnau, Jaume et al. (2008) *Estudios longitudinales modelos de diseño y análisis*, en Revista Escritos de Psicología 2-1, pp. 32-41

Cochran, William G (1976). *Técnicas de muestreo*, 2ª Edición, Compañía Editorial Continental, México, pp. 507.

Chávez, Angeles.(2007) Reporte de la encuesta nacional a hogares rurales de México 2002: una semblanza. Tesis, Facultad de Ciencias, UNAM. pp77

Díaz de Rada, V. (2007). *Tipos de encuestas considerando la dimensión temporal*, en Papers-Revista de Sociología nº 86, pp. 131-145

Dyer Leal, G et al. (2002) *Guía del Cuestionario a Hogares*, documentos de apoyo en e el levantamiento de la Encuesta a Hogares de la Encuesta Nacional a Hogares Rurales de México. Programa de Estudios del Cambio Económico y la Sustentabilidad del Agro Mexicano, El Colegio de México. pp. 23.

Dyer Leal, G et al. (2002) *Manual de la Encuesta Comunitaria*, documentos de apoyo en e el levantamiento de la Encuesta a Hogares de la Encuesta Nacional a Hogares Rurales de México. Programa de Estudios del Cambio Económico y la Sustentabilidad del Agro Mexicano, El Colegio de México. pp23.

Escalante, Roberto et al. (2008) *Situación actual del sector agropecuario en México: perspectivas y reto*, en Economía informa, núm 350 enero-febrero. pp. 21

Ferber Rober et al. (1982) *What is a survey*, Subcommittee of the Section on Survey Research Methods American Statistical Association Washington, D.C. pp.18

Kish, Leslie (1965). *Survey Sampling*, John Wiley and Sons, Estados Unidos, pp. 642.

Lynn, Peter (2009), *Methodology of longitudinal surveys*, John Wiley and Sons, United Kingdom, pp. 396.

Medina, Fernando, *Las encuestas panel*, en Programa para el mejoramiento de las encuestas y la medición de las condiciones de vida en América Latina y el Caribe. CEPAL. pp. 139-145

Méndez, Ignacio. (2011). *El protocolo de investigación* 2ª Edición, Trillas, México, pp 210.

PRECESAM (Programa de Estudios del Cambio Económico y la Sustentabilidad del Agro Mexicano) (2002), Encuesta Nacional a Hogares Rurales de México, reporte metodológico.

PRECESAM (Programa de Estudios del Cambio Económico y la Sustentabilidad del Agro Mexicano) (2010), Informe metodológico y de procedimientos de la Encuesta Nacional a Hogares Rurales de México 2007.

Sheski, David , J. (2007). *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall, Estados Unidos, pp.1772.

Siegel, Sidney(1988), *Estadística no paramétrica aplicada a las ciencias de la conducta*, 11 Edición, Editorial Trillas, México, pp. 343.

Steven G. Heeringa et al, *Applied surveys data analysis*, Taylor and Francis group, Estados Unidos, pp. 467

Bases de datos:

PRECESAM,2002, Encuesta Nacional a Hogares Rurales en México I (ENHRUM I)

PRECESAM, 2007, Encuesta Nacional a Hogares Rurales en México (ENHRUM II)