

18947.06
(046670)

C2



NACIONES UNIDAS
Fondo de Población de las Naciones Unidas
Programa Global de Formación en Población y Desarrollo

Centro Latinoamericano de Demografía

**FACTORES ASOCIADOS AL DESPOBLAMIENTO
RURAL Y EL USO DE VARIABLES
SINTOMÁTICAS PARA
ESTIMACIONES DE
POBLACIÓN EN
CUBA
JULIO CÁRDENAS**

ENSAYO DE INVESTIGACIÓN FINAL



CURSO DE POSTGRADO EN POBLACION Y DESARROLLO 1995

Santiago de Chile

CELADE - SISTEMA DOCPAL
DOCUMENTACION
SOBRE POBLACION EN
AMERICA LATINA

SÓLO PARA USO INTERNO

Este documento corresponde a un trabajo realizado por el participante.

No ha sido sometido a revisión editorial y los juicios en él contenidos son de responsabilidad exclusiva de su autor.

NACIONES UNIDAS

Fondo de Población de las Naciones Unidas

Programa Global de Formación en Población y Desarrollo

Centro Latinoamericano de Demografía

**FACTORES ASOCIADOS AL DESPOBLAMIENTO
RURAL Y EL USO DE VARIABLES
SINTOMÁTICAS PARA
ESTIMACIONES DE
POBLACIÓN EN
CUBA
*JULIO CÁRDENAS***

ENSAYO DE INVESTIGACIÓN FINAL



CURSO DE POSTGRADO EN POBLACION Y DESARROLLO 1995

Santiago de Chile

**Curso de Postgrado en Población
y Desarrollo**

T R A B A J O F I N A L

**FACTORES ASOCIADOS AL DESPOBLAMIENTO RURAL
DE VILLA CLARA Y EL USO DE VARIABLES
SINTOMATICAS PARA ESTIMACIONES DE POBLACION
EN CUBA.**

Autor: Julio Roberto Cárdenas Pérez

Asesor: Martines Dirven

**CELADE
1995**

INDICE

INDICE

Introducción.	1
Desarrollo.	4
Capítulo # 1. Características Generales del país y la provincia.	4
Planteamiento del problema.	5
Población urbana.	6
Población rural.	6
Lugar habitado urbano.	6
Lugar habitado rural.	7
Propósitos y Objetivos.	8
Objetivo General.	9
Objetivos Específicos.	9
Consideraciones metodológicas.	9
Capítulo # 2. Efectos de la migración rural-urbana y los cambios conceptuales en el despoblamiento de las zonas rurales de Villa Clara.	11
Capítulo # 3. Métodos y modelos utilizados.	20
Información básica.	20
Métodos y modelos.	21

Distribución por prorrateo.	21
Ventajas:	22
Desventajas:	22
Distribución proporcional.	22
Ventajas	23
Desventajas	23
Correlación de razón.	24
Ventajas:	25
Desventajas:	26
Correlación de diferencia.	26
Capítulo # 4. Resultados Estadísticos	28
Conclusiones.	47
Recomendaciones sobre las estimaciones	49
Bibliografía	50
Anexos	
Apéndice (Documento independiente)	

INTRODUCCION

TITULO

Factores asociados al despoblamiento rural de Villa Clara y el uso de variables sintomáticas para estimaciones de población en Cuba.

Introducción.

Cuba y la Provincia de Villa Clara, han tenido en los últimos años un importante despoblamiento de las zonas rurales. El despoblamiento se ha considerado como una consecuencia de la migración rural-urbana, especialmente de la población joven, sin embargo no toda la disminución de la población rural ha dependido de esta migración, sino también de los cambios de definición de la población rural y de la población urbana.

En relación al último aspecto debe decirse que en el censo de 1980, los cambios introducidos en las definiciones originan que un importante número de localidades rurales alcancen la categoría de urbana, pues reúnen las condiciones que las caracterizan como tal, sin haberse producido movimientos migratorios, al mismo tiempo que se modifica el número total de la población rural, la cual, como es de suponer disminuye más aceleradamente.

Es importante mencionar que el Estado cubano, ha implementado un conjunto de políticas, encaminadas a la búsqueda del equilibrio entre la ciudad y el campo, ya que el país y en especial, la región central (Villa Clara), son eminentemente agrícolas. En tal sentido, se han puesto en práctica programas y políticas dirigidas a la urbanización de las zonas agrícolas, a través de la instalación y dotación de establecimientos educativos y médicas, salas de videos, estímulos materiales que incluye reordenamiento de la política salarial, etc, con el fin de mejorar las condiciones de vida de la familia campesina y con el fin de hacer atractivas a las áreas rurales.

Dada la situación económica que vive el país desde finales de la década de los 80', se ha considerado importante analizar hasta qué punto la disminución de la población rural es producto de un cambio del lugar de residencia o se debe simplemente al cambio de concepto mencionado anteriormente.

El presente trabajo tiene como objetivos centrales analizar la incidencia de los cambios conceptuales, respecto a las definiciones de lugar habitado urbano y rural

sobre el despoblamiento de las zonas rurales de la provincia de Villa Clara, y proponer una metodología de estimación de población basada en el uso de las variables sintomáticas¹ para áreas pequeñas.

El trabajo se inicia con un primer capítulo que pretende hacer una breve caracterización general del problema a tratar. El sentido es mostrar, en un contexto muy general, la problemática y algunas líneas, también muy generales de como se ha abordado el mismo.

En un segundo capítulo se analizan los efectos de la migración rural-urbana y la incidencia de los cambios en los conceptos de ambos censos en la disminución de la población rural tanto para todo el país como para la provincia de Villa Clara.

Seguidamente, en el capítulo # 3, se trata de mostrar los métodos y modelos utilizados en las estimaciones de población a través del uso de las variables sintomáticas. En particular se describen las ventajas y desventajas de los métodos siguientes; Distribución por Prorratio, Distribución Proporcional, Correlación de Razón y Correlación de Diferencias. (Esta metodología se espera sea un instrumento importante para la verificación y evaluación periódica de la evolución de la población de las áreas rurales).

Los datos utilizados para el desarrollo de esta metodología, fueron tomados de la investigación desarrollada por Guiomar Bay publicados en mayo de 1994 debido a la imposibilidad de obtener información precisa sobre Cuba a ese nivel de desglose.

Los resultados obtenidos en las estimaciones de población mediante el uso de las variables sintomáticas incluyendo un método alternativo en el cual sólo con una función matemática y las mismas variables sintomáticas es posible realizar estas

(1)...Las variables sintomáticas están relacionadas con los cambios en el tamaño de la población, tales como: nacimientos, defunciones, permisos para conducir o viviendas contruidas, impuestos recolectados, registro de automóviles o licencias de conducir, registro electoral, matrícula escolar (por edad y nivel), valores de los depósitos bancarios, tasas de ocupación, mano de obra agrícola, afiliados a Seguridad Social, consultas médicas, superficie sembrada o plantada, consumo de energía eléctrica, etc.

Los nacimientos y las defunciones son considerados acá como variables sintomáticas, o sea, como indicadores del tamaño de la población, y no como componentes de la dinámica poblacional.(Guiomar Bay mayo de 1994)

estimaciones se muestran en el último de los capítulos de este trabajo.

Finalmente, se arriban a conclusiones y recomendaciones para desarrollar otros trabajos vinculados con el tema.

CAPITULO # 1

Desarrollo.

Capítulo # 1.

Características Generales del país y la provincia.

En Cuba, la historia antes de 1959 habla de un poblamiento empujado hacia la montaña por falta de posibilidades de tierra o trabajo en el llano. Posteriormente se produce un éxodo considerable de la población. Dicho éxodo fue un tanto contrarrestado por las políticas diferenciadas aplicadas en las zonas de montaña. Sin embargo, ya a mediados de los años 70 comienza hacer crisis la montaña y otras zonas rurales del país, motivados por dos factores: estancamiento socioeconómico de las zonas y desarrollo acelerado de las cabeceras.

Posteriormente se inició el proceso de despoblamiento rural, originado por la migración -la cual ocurre generalmente en la población joven- de la provincia. Ello unido al descenso de la fecundidad, han dado como resultado el envejecimiento de la población que tiene nocivas consecuencias socioeconómicas que arrojan un cuadro difícil, a mediano plazo, para algunos municipios de la provincia de Villa Clara.

Esta provincia, por su extensión, ocupa el sexto lugar entre las provincias cubanas. Se encuentra situada en la región central del país, limitada con Sancti Spíritus por el este y el sudeste, con Cienfuegos al sudoeste, con Matanzas al oeste y con el Océano Atlántico por el norte. Sus primeros pobladores, antes de los colonizadores españoles fueron aborígenes dedicados a la caza, la pesca y la recolección.

El 3 de mayo de 1514 se establece la estancia que daría origen al núcleo inicial de la población de San Juan de los Remedios, primera villa fundada en su territorio. Un siglo después se funda la Ciudad de Santa Clara, hoy cabecera de provincia, el 15 de Julio de 1689 (Censo de Población y Vivienda, 1982, pág. II).

Con el triunfo de la Revolución surgen perspectivas reales de desarrollo para la economía de la región. Villa Clara cuenta con tierras muy fértiles, dedicadas en gran parte al cultivo de la caña de azúcar. Cuenta además con un gran número de centrales azucarero, ocupando el segundo lugar nacional en su producción. Los centros de producción industrial más importantes están ubicados fundamentalmente en la ciudad

de Santa Clara, los cuales son abastecedores de gran parte de la demanda nacional de piezas de repuesto y equipos para la industria azucarera.

El sistema educacional, regido por el Ministerio de Educación, comprende todos los tipos de enseñanza, incluida la enseñanza superior. La salud pública ha obtenido significativos logros contando con excelentes instalaciones, equipadas con moderna tecnología y con un personal altamente calificado.

Los planes de desarrollo prospectivo de la provincia contemplan un incremento del desarrollo agro-industrial en el que la producción azucarera seguirá teniendo un papel principal. Sin embargo dadas las condiciones económicas actuales del país, muchos de estos planes se encuentra totalmente paralizados o medianamente ejecutados.

Por otra parte, en cuanto a la distribución espacial de la población, durante los últimos decenios, la provincia de Villa Clara ha sostenido un acelerado crecimiento urbano, en el cual ha tenido una alta incidencia la migración rural-urbana, fundamentalmente por factores de índole económico; infraestructura económica y social, niveles culturales y educacionales alcanzados, etc, que propician la atracción hacia los centros urbanos de la población rural.

Planteamiento del problema.

El crecimiento de la población urbana en Cuba se relaciona de manera importante con el proceso de reclasificación de localidades rurales que pasan a la categoría de urbanas pero también con el surgimiento de nuevas localidades.

Desde la década del 80' se observa que los criterios que prevalecen para definir la población urbana, tienden a aumentar su proporción a cuenta de una disminución de la población rural, pues muchas poblaciones rurales se convierten en urbanas.

Originalmente, en la categorización de población urbana y rural, se tenía en cuenta no sólo los volúmenes de población, sino la actividad predominante en la localidad y en los últimos tiempos estas categorías han tenido cambios significativos de manera que esta definición cedió a la que considera el criterio demográfico (tamaño de la población) unido a otros requisitos relativos a la presencia de infraestructura y servicio. Por ejemplo:

En el censo de 1970 se consideraba que:

Población urbana.

Es la población residente en los lugares habitados de 2000 habitantes y más, así como en los de menos de 2 000 o de 500 habitantes que contaban con cuatro de las características siguientes:

Alumbrado público, calles pavimentadas, acueducto, red de alcantarillado o cloacas, servicios médicos asistencial y centro educacional.

No obstante, mediante el registro previo de lugares habitados se definió, excepcionalmente, como población urbana algunos pueblos construidos por la Revolución, que no tenían la población planteada y contaban con las condiciones exigidas.

Población rural.

Es la residente en los lugares habitados por menos de 500 habitantes o las de aquellos de 500 o menos de 2 000 habitantes, que presentan menos de cuatro de las características mencionadas.²

En el censo de 1980 se consideró que:

Lugar habitado urbano.

A los fines censales se consideró como lugar habitado urbano:

A.- Todos los lugares habitados con una población residente de 2 000 o más habitantes.

B.- Todos los lugares habitados con una población residente de 500 a menos de 2 000 habitantes, que contaran con alumbrado público y 3 o más características de las 5 que se relacionan a continuación:

² CEE, (Comité Estatal de Estadística)(1970), Censo de Población y Vivienda 1970, La Habana.

- 1.- Acueductos.
- 2.- Calles pavimentadas.
- 3.- Red de alcantarillado o cloacas.
- 4.- Servicio médico asistencial.
- 5.- Centro educacional.

C.- Todos los lugares habitados con una población de 200 a menos de 500 habitantes que contaran con las 6 características siguientes: alumbrado público, acueducto, calles pavimentadas, red de alcantarillado o cloacas, servicio médico asistencial y centro educacional.

Lugar habitado rural.

Se consideró como rural a los lugares habitados del país que contaban con menos de 200 habitantes.

También dentro de esta categoría se encontraban aquellos lugares habitados que contaran con 200 o menos de 2000 habitantes que no reunieran las características establecidas para ser urbanos, de acuerdo a la definición de lugar habitado urbano.³

El censo de 1990 no se llevó a efecto debido a problemas económicos. Sin embargo, en las consideraciones metodológicas se plantean las mismas definiciones que en el censo de 1980. Estas categorías son las que se mantienen vigentes en la actualidad.

Según nuestro criterio, es obvio (con la generalización de ciertos servicios básicos y con el creciente desarrollo de la red vial) que alcanzar los parámetros para considerarse "urbano" es cada vez más fácil y no significa que la actividad socioespacial tenga características netamente urbanas.

Algunos trabajos realizados en torno a la problemática del despoblamiento en las zonas rurales señalan como una de sus causas principales, los resultados obtenidos en la aplicación durante varios años de las políticas de urbanización del campo, que tienen como objetivo atenuar las diferencias entre el campo y la ciudad y que se

³ CEE, Censo de Población y Vivienda, 1982, Vol. 5, pág. XXXV.

expresó en los lineamientos Económicos y Sociales aprobados en los Congresos del Partido Comunista de Cuba (PCC)⁴.

En el marco de las consideraciones anteriores, cabe decir que el interés del presente trabajo es estudiar como los diferentes conceptos utilizados en los censos de 1970 y 1982, relativos a las definiciones "lugar habitado urbano" y "lugar habitado rural" han incidido en la distribución espacial de la población, dando lugar, en ciertos casos, a interpretarse como despoblamiento de las áreas rurales, cuando en realidad no ha existido movilidad espacial de la población, sino un cambio de definición, lo cual ha ido en detrimento del número total de población rural de la provincia.

Por último, es necesario decir que:

- * las estrategias de desarrollo vigentes en el territorio de Villa Clara;
- * la no realización del Censo de 1992, debido a las condiciones económicas imperantes;
- * las metodologías para determinar el número de habitantes basadas en tendencias pasadas, presentan limitaciones, a menos que se puedan evaluar periódicamente;
- * las limitaciones de los distintos programas o software de proyecciones de población.

Constituyen un conjunto de factores que favorecen el análisis y aplicación de métodos basados en indicadores del tamaño de la población, denominadas variables sintomáticas, a través de modelos de regresión. Esta es una alternativa útil para determinar el número de habitantes en áreas pequeñas y evaluar sus resultados, así como también, evaluar las políticas encaminadas al aumento de la población rural a través de métodos comparativos. En este caso se estudiará como variable dependiente la población a estimar y como variables independientes, por ejemplo, los nacimientos, las defunciones, consumo eléctrico, matrícula escolar, y cantidad de vehículos motorizados registrados.

Propósitos y Objetivos.

⁴ De cierta, forma esta política también incentivó a la concentración de la población en el área urbana.

Se propone analizar sobre los posibles efectos sobre el despoblamiento de las zonas rurales a partir de los cambios de conceptos intercensales (70 - 81) respecto a los lugares habitados urbanos y rurales y proponer una alternativa útil, desde el punto de vista metodológico, para determinar el número de habitantes en áreas pequeñas a partir de las variables sintomáticas.

Objetivo General:

- * Analizar las causas del despoblamiento de las zonas rurales.
- * Proponer una metodología de estimación de población basada en el uso de las variables sintomáticas para áreas pequeñas.

Objetivos Específicos.

- * Analizar los criterios censales sobre las definiciones de población rural-urbana y su incidencia sobre el despoblamiento rural de la provincia de Villa Clara.
- * Evaluar los diferentes métodos existentes para la determinación del número de habitantes en áreas pequeñas así como sus resultados mediante el uso de variables sintomáticas.
- * Proponer recomendaciones en relación al uso de los métodos de estimaciones de población, a partir de las condiciones de la condiciones de la provincia de Villa Clara.

Consideraciones metodológicas.

El trabajo ha sido desarrollado en dos partes:

En la primera de ellas se analiza las diferentes definiciones, políticas y programas que han dado lugar a una disminución de la población rural.

En una segunda etapa se estudian las posibilidades de las diferentes metodologías y métodos para la estimación del número de habitantes mediante el uso de variables sintomáticas.

Para este fin se utilizaron las técnicas estadísticas con especial énfasis de los métodos de regresión simple y múltiple, y en los software profesionales actualizados (SPSS/PC Ver 4.1, STEPWISE⁵ y LOTUS 1-2-3).

Dentro de las consideraciones especiales a tomar en cuenta y que se describen en el trabajo están los conceptos siguientes:

Para el **SPSS/PC**:

.- Altamente significativo: Cuando el valor de significación de la prueba ejecutada está por debajo de 0.0000.

.- Significativo: Cuando el valor de significación de la prueba aplicada está por debajo de 0.05.

.- No significativo: Cuando el valor de significación de la prueba es superior a 0.05.

Para el **STEPWISE**:

.- Altamente significativo: Cuando la probabilidad de significación de la prueba tiende a valores cercanos de 0.0000.

.- Significativo: Cuando la probabilidad de significación de la prueba es inferior de 0.95.

.- No significativo: Cuando el valor de significación de la prueba es superior de 0.95.

⁵ Programa desarrollado por el Instituto de Informática de la Universidad Central de Las Villas, Santa Clara, Villa Clara, Cuba

CAPITULO # 2

Capítulo # 2.

Efectos de la migración rural-urbana y los cambios conceptuales en el despoblamiento de las zonas rurales de Villa Clara.

En Cuba, el régimen económico capitalista imperante hasta 1958, impuso su sello distintivo a la evolución de estos fenómenos. La tenencia y las formas de explotación de la tierra, junto a las contradicciones de clase inherentes a tal formación económico-social, y las particularidades de la agricultura prevalecientes, condicionaron el poblamiento rural del país y su propia movilidad. El acelerado crecimiento de la capital imponía también marcadas diferencias entre ellas y el resto de las regiones del país, relevando una vez más la contradicción de carácter antagónico entre el campo y la ciudad.

Entre los rasgos típicos que adquiría entonces el poblamiento rural, estaban la fuerte dispersión con formas diversas según la actividad económica predominante, acompañada de una movilidad territorial que favorecía a los grandes centros urbanos a causa del desempleo imperante en las zonas rurales y el deterioro en que estaba sumido el medio rural.

El poblamiento urbano se concentraba en la ciudad capital, en una proporción tal, que en 1953 alcanzó la cifra del 35% de la población urbana residiendo en ésta. En la ciudad de La Habana, y en menor medida en otras ciudades como Santiago de Cuba y Santa Clara, se concentraba casi toda la actividad industrial y comercial del país, y las principales funciones administrativas y de servicios, circunstancias éstas que frenaban el desarrollo de las llamadas ciudades medianas y pequeñas, que fungían como simples poblados rurales.

La agudización paulatina de las contradicciones entre el medio rural y el urbano debido al propio deterioro del sistema capitalista, impulsaría las migraciones internas, básicamente el éxodo rural, precisamente en dirección de la ciudad capital, lo que explicaba la rápida disminución que experimentaba la población rural del país, que descende de 62 por ciento en 1919 a 46 por ciento en 1953.

El triunfo de la Revolución y la posterior recuperación por la sociedad cubana de los principales medios de producción, constituyeron las premisas básicas para el reordenamiento del territorio y la combinación más armónica de las fuerzas

productivas con arreglo a un plan único de desarrollo económico y social, cuyo objetivo fundamental ha sido la eliminación de las diferencias territoriales comentadas. El país ha hecho un gran esfuerzo por disminuir las diferencias entre el campo y la ciudad con respecto a los servicios y por elevar el nivel de vida de sus habitantes rurales, ofreciéndoles a los jóvenes espacios de participación social.

La revitalización de las ciudades medianas y pequeñas, que van asumiendo nuevas funciones administrativas, especialmente desde 1976; la creación de novedosas e importantes planes de desarrollo agropecuario e industrial a escala regional, y la construcción de más de 300 nuevos poblados rurales o comunidades y otros fortalecidos con más de 1400 cooperativas que agrupan a más de 250 mil habitantes, antes integrantes de la población dispersa rural, modifican en fin la importancia de los movimientos migratorios locales. Tales hechos, explican la peculiaridad de que la creciente mecanización de la agricultura y la formación de unidades de explotación cada vez más grandes, hayan servido de estímulo para evitar un éxodo mayor del campo a la ciudad. Este aspecto y significación han sido reconocidos aún a nivel internacional.

Junto a esto, se suma que a mediados de los años 80, para enfrentar la escasez de mano de obra agrícola, se formularon objetivos explícitos en materias de distribución territorial de la población, a través de ofertas de viviendas, salarios diferenciados y la utilización de otros estímulos, los que también, han ejercido influencia en la desaceleración de la migración rural-urbana.

La unidad de acción entre el Partido Comunista de Cuba (PCC), las organizaciones estudiantiles a todos los niveles, la Unión de Jóvenes Comunistas (UJC)⁶, la Central de Trabajadores de Cuba y otras organizaciones políticas y de masas han logrado que muchos jóvenes urbanos se desplacen por períodos cortos (15 días) o medianos (dos años) para trabajar en la agricultura, atraídos fundamentalmente por la solución a los problemas de vivienda y otras necesidades básicas. Sin embargo, para la mayoría de estos jóvenes, retornar a la ciudad parece seguir siendo su objetivo principal.

En lo que migración rural-urbana respecta, la Encuesta Demográfica Nacional (EDN) levantada en 1979, permitió conocer la evolución experimentada por ésta desde 1959. La migración rural-urbana, que durante el decenio 1959-1969, representaba el

⁶ Organización partidista juvenil.

40% de los movimientos migratorios registrados, había descendido al 37 por ciento en el quinquenio 1975-1979, sin embargo esta diferencia aún no es tan significativa si tomamos en cuenta los esfuerzos realizados hasta esa fecha. Por otra parte se experimentaron incrementos de los traslados entre áreas rurales.

Entre 1970 y 1981, el 84 por ciento del aumento de la población urbana ocurrido en el país (corresponde a unos 1 274 200 efectivos) se produjo en los lugares habitados urbanos que figuran en ambos momentos censales, por efecto combinado del crecimiento natural y las migraciones.

La ubicación territorial concebida sobre la base de subsistemas urbanos con asentamientos de distintas categorías, tamaños y funciones, contribuyó a la descentralización urbana y a desviar la migración rural hacia pequeños y medianos centros urbanos. Respondiendo al interés de satisfacción de las necesidades de servicios de la población rural, surgieron entonces las comunidades construidas por la Revolución que contribuyeron a la introducción de patrones de vida urbanos en el medio rural, al tiempo que favorecían el proceso de concentración de la población rural dispersa.

En sentido contrario aumenta la magnitud de población residente en los núcleos rurales menores de 1999 habitantes. Los núcleos con menos de 200 habitantes, en número de 16006 en 1970, disminuían su cifra a 11420 en 1981, mientras, las localidades rurales de 200 a 1999 habitantes aumentaban de 1811 a 2570 entre uno y otro censo. Esta variación diferencial, traía aparejada cambios de igual tendencia en la población residente en dichos aglomerados. La población dispersa que habitaba en localidades rurales de menos de 200 habitantes, redujo sus efectivos de 2 655 miles en 1970 a 1 098 miles en 1981, es de suponer que la diferencia migró a otros lugares, mientras que la población rural residente de localidades de más de 200 habitantes se incrementó de 724 miles a 1 057 miles en 1981.

El proceso de cooperativización que últimamente ha tenido lugar en el país, ha sido también un importante factor en la concentración de la población rural, especialmente en localidades de tamaño superior y en otras de carácter urbano previamente establecidas, en la mayoría de los casos, y en ellas, mediante la autoconstrucción y con ayuda estatal auspiciada por ministerios afines, se ofrecen soluciones constructivas de viviendas y oferta de servicios a la población agrícola.

El proceso de concentración de la población rural y las inversiones productivas que le acompañan, como estrategia de una política que persigue disminuir la brecha entre el campo y la ciudad, han repercutido favorablemente en los niveles de calidad de la vida del medio rural disperso, el cual resultaba ser el sector territorial más deprimido de la sociedad cubana.

El cuadro # 1, ofrece un resumen de los cambios ocurridos en el país tomando en cuenta las definiciones metodológicas de ambos censos realizados.

Cuadro # 1: Cambios ocurridos en los lugares habitados urbanos según los censos de 1970 y 1981.

CAMBIOS	Población Urbana			
	1970	1981	Difer.	%
Lugares habitados urbanos que figuran en ambos censos (408)	5172.9	6447.2	1274.3	83.3
Lugares rurales en 1970, pero urbanos en 1981 (134)	-	204.9	204.9	12.5
Lugares habitados urbanos en 1970 pero rurales en 1981 (23)	15.0	-	(15.0)	-
Nuevos lugares habitados urbanos en 1981 (Comunidades) (42)	-	59.9	59.9	3.9
Total	5187.9	6712.0	1524.1	100.0

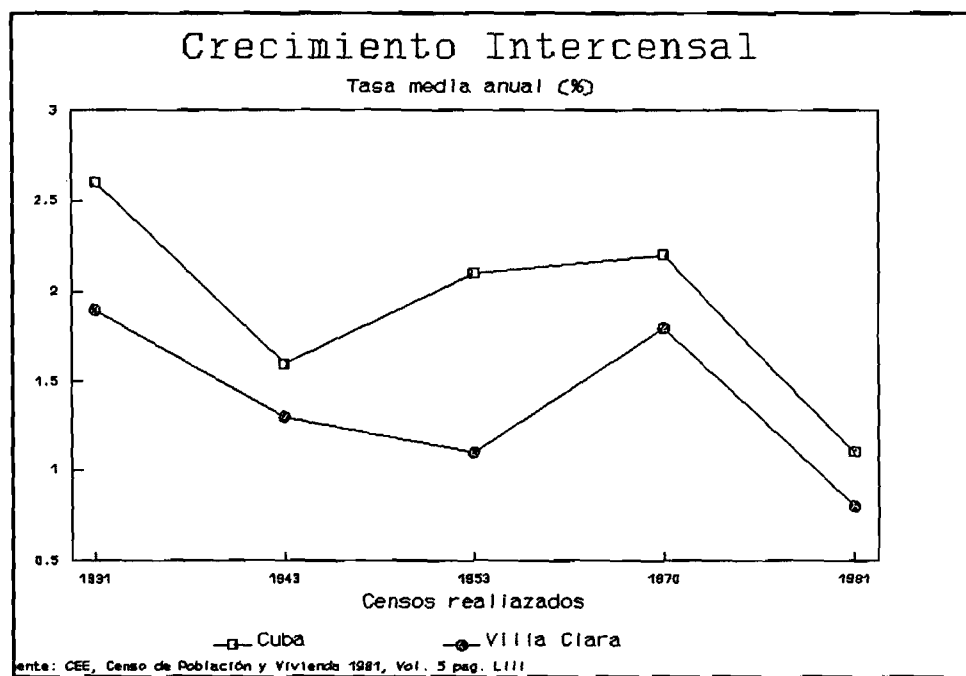
Fuente: Comité Estatal de Estadísticas. Censo de Población y Viviendas, 1981, Vol 16, Cuba, pág LXVII.

Con referencia a la provincia de Villa Clara, la misma contaba con 700 mil habitantes en momento del censo de 1970. En 1981 contaba con la cifra de 765823

habitantes, creciendo a una tasa media anual desde 1970 de 1.8 por ciento mientras el país crecía a un promedio medio anual del 2.2 por ciento ⁷.

El aumento de la población señalado, entre los dos últimos censos del país, se origina, en el crecimiento natural (110 mil personas) y los saldos migratorios hacia el exterior y a otras provincias. (aproximadamente 40 mil personas).

En sentido general, el aumento total de la población urbana de la provincia (147.9 miles de personas) en el período intercensal se genera⁸, en parte por el incremento natural de la población de las áreas urbanas, en una considerable porporción por las migraciones del campo a la ciudad y por el factor ya apuntado de cambios de clasificación y creación de nuevas comunidades urbanas.



La densidad⁹ de población de la provincia en 1970 fue de 80.4 hab/Km² y en

⁷ CEE, Censo de Población y Vivienda, 1982, Cuadro # 1, pág. LIII.

⁸ Según indican las estadísticas de población y la Encuesta Demográfica Nacional de 1979.

⁹ Entiendase por densidad de población el índice que relaciona el volumen de la población con respecto al territorio que ocupa.

1981 alcanzó la cifra de 88.4 hab/Km² similar a la media nacional (87.5 hab/Km²)¹⁰. Como puede apreciarse la densidad aumentó en el período intercensal en 7.6 hab/Km², es decir, un crecimiento similar a otras provincias pero inferior a la media nacional.

La provincia de Villa Clara ocupaba el sexto lugar¹¹ en el país por su porcentaje de población urbana (68.7%), lo que la sitúa entre las provincias más urbanizadas¹² a nivel nacional.

En el cuadro # 2 se presenta la población urbana y rural de Cuba y la provincia de Villa Clara en el período intercensal de 1970 a 1981. En este tiempo (11 años), la población urbana creció en 147.9 habitantes, lo que representa el 28.1 por ciento del total urbano en la provincia en el año 1981. A nivel nacional, en el mismo lapso la población urbana se incrementó en 1514.5 mil personas, representando el 22.5 por ciento de la población urbana total del país en 1981.

Cuadro # 2. Población Urbana y Rural de Cuba y Villa Clara.

Censos	Población Urbana		Población Rural	
	Cuba	Villa Clara	Cuba	Villa Clara
	Cifras Absolutas (en miles)			
1970	5187.3	378.1	3381.3	321.9
1981	6702.3	526.0	3005.3	239.8
	Cifras Relativas (en por cientos)			
1970	60.5	54.0	39.5	46.0
1981	69.0	68.7	31.0	31.3

Fuente: Comité Estatal de Estadística, Censo de Población y Vivienda, 1981, Vol. 5, pág. LV.

¹⁰ Censo de Población y Vivienda 1981, Vol. 5, "Características generales", Cuadro # 2, pág. LIV.

¹¹ El país cuenta con 14 provincias y un municipio especial (Isla de la Juventud)

¹² Entiendase por grado de urbanización a la proporción de habitantes que residen en las áreas urbanas en relación con la población total.

En términos relativos, el grado de urbanización a nivel nacional creció de 60.5 en 1970 a 69.0 por ciento en 1981 (8.5 puntos) mientras que la provincia de 54.0 por ciento pasó a 68.7. El crecimiento de la población urbana en la provincia fue superior a la media nacional en aproximadamente 6.2 puntos.

La situación de la población rural cambió tanto en términos absolutos como relativos, ya que disminuye a nivel de país y a nivel de provincia.

A nivel de país la población rural disminuye en 376 mil efectivos mientras que en la provincia la disminución es del orden de los 82 mil habitantes. Desde el punto de vista porcentual, en el país la población rural pasa de 39.5 por ciento en 1970 a 31.0 en 1981, es decir, disminuye en 8.5 puntos respecto a 1970 . La provincia por su parte, de 46.0 en 1970 pasó a 31.3 por ciento en 1981 para una disminución de 14.7 puntos, lo que quiere decir que disminuye la población rural en la misma proporción en que aumenta la población urbana con respecto al país en general.

En 1981 la provincia de Villa Clara mostró en cifras absolutas la mayor disminución de población rural del país.

Con el fin de profundizar en las variaciones ocurridas en la población según zonas urbanas y rurales, en el cuadro # 3 se muestran los cambios ocurridos en los lugares habitados urbanos de Villa Clara en el período intercensal.

Cuadro # 3: Cambios en la población de los lugares habitados urbanos según los censos de 1970 y 1981.(Villa Clara).

C A M B I O S	Población Urbana			
	1970	1981	Difer.	%
Lugares habitados urbanos que figuran en ambos censos (44)	377.6	496.3	118.7	80.3
Lugares rurales en 1970, pero urbanos en 1981 (21)	-	25.5	25.5	17.2
Lugares habitados urbanos en 1970 pero rurales en 1981 (1)	0.5	-	(0.5)	
Nuevos lugares habitados urbanos en 1981(Comunidades) (3)	-	4.2	4.2	2.8
Total	378.1	526.0	147.9	100.0

Fuente: Comité Estatal de Estadística, Censo de Población y Vivienda, 1981, Vol. 5, pág. LVI.

Como se observa, el aumento de la población urbana (el 80.3 por ciento, 118.7 mil personas), en el período intercensal, se efectuó en los lugares habitados urbanos que figuran en ambos censos, y sólo el 20.0 por ciento restante se debe a lugares rurales que cambiaron su clasificación a nuevos centros urbanos surgidos en 1981.

Al comparar los resultados provinciales con los existentes a nivel nacional, se puede apreciar, que el incremento de la población urbana de la provincia en lugares habitados urbanos que figuran en ambos censos, es similar (80.3 por ciento) al crecimiento medio nacional (83.6 por ciento). Sin embargo, debido al cambio de lugares rurales a urbanos, el incremento de la población urbana, en términos relativos, es superior (17.2 por ciento) al ocurrido a nivel nacional en aproximadamente 5

puntos. Por último, el crecimiento de la población urbana a nivel de país debido a nuevos lugares habitados urbanos o comunidades fue del orden del 3.9 por ciento y en la provincia de 2.8 por ciento. El 7.1 por ciento de las nuevas comunidades fueron creadas en esta provincia.

Todos estos factores, no se presentan separadamente en municipios específicos, sino que la acción conjunta de todos ellos influye en el crecimiento urbano de la provincia, aunque a veces, predomina en un municipio un factor más que otro.

CAPITULO # 3

Capítulo # 3.

METODOS Y MODELOS UTILIZADOS.

Información básica.

Los datos primarios tomados para el desarrollo de este trabajo fueron los mismos que sirvieron de base para la investigación "EL USO DE VARIABLES SINTOMATICAS EN LA ACTUALIZACION DE POBLACION DE AREAS MENORES" realizada en el Centro Latinoamericano de Demografía por la Consultora del CELADE, Guiomar Bay, publicado en mayo de 1994.

La matriz de datos originales de Chile estaba formada por las siguientes variables: Región, Número de Comuna, Nombre de la Comuna, Población censo 1970, Nacimientos registrados en 1970, Defunciones registradas 1970, Población censo 1982, Nacimientos registrados en 1982, Defunciones registradas 1982, Matrícula en Educación Básica registradas en 1982, Población censo 1992, Nacimientos registrados en 1992, Defunciones registradas 1992, Matrícula en Educación Básica registradas en 1992, Población proyectada para 1992 según el "Método de relación de Cohorte", Condición Urbano/Rural observada en el censo de 1992, Automóviles motorizados registrados en 1982 y Automóviles motorizados registrados en 1992.

Para este trabajo se utilizaron los siguientes datos: Región, Número de Comuna, Población del Censo 1970, Nacimientos registrados en 1970, Defunciones registradas 1970, Población censo 1982, Nacimientos registrados en 1982, Defunciones registradas 1982, Matrícula en Educación Básica registradas en 1982, Automóviles motorizados registrados en 1982, Población censo 1992, Nacimientos registrados en 1992, Defunciones registradas 1992, Matrícula en Educación Básica registradas en 1992, Población proyectada para 1992 según el "Método de Relación de Cohorte" y Automóviles registrados en 1992.

En este conjunto de información existían comunas (aproximadamente 15) cuyos valores de población eran demasiado pequeños, en este caso, se decidió no trabajar con los mismos por considerarlos atípicos.

Todas aquellas comunas cuyos valores de población, en los tres censos, estuviera por debajo de 2000, no se tomaron en consideración para los cálculos. Si alguna comuna tenía valores de población por encima de 2000 en cualquiera de los censos, se mantuvo en el juego de datos con sus valores originales.

Métodos y modelos. ¹³

Los métodos y modelos a estudiar son los que a continuación se describen.

Los modelos se referirán a tres momentos específicos. Ellos son:

- 1.- momento 0: corresponde al año del primer censo;
- 2.- momento t: corresponde al año del segundo censo; y,
- 3.- momento (t + n): corresponde al año de la estimación.

Distribución por prorateo.

Este es el más sencillo de los métodos considerados en este trabajo. Se basa en el supuesto de que la distribución de la población por área es idéntica a la distribución de la variable sintomática. Para su aplicación se necesita solamente la distribución proporcional de la variable sintomática, para el año que se pretende estimar la población del área, y la estimación de la población total de las áreas para dicho año.

Así se tiene que la población estimada del área i en el momento t + n está dada por:

$$P_{(i,t+n)} = \frac{S_{(i,t+n)}}{S_{(.,t+n)}} * P_{(.,t+n)} \quad (1)$$

donde:

$P_{(i,t+n)}$ = población del área i en el momento t + n;

(3)... Metodología descrita en el Trabajo "El uso de variables sintomaticas en la actualización de población de áreas menores"

$P_{(.,t+n)}$ = estimación de la población total (suma de las i áreas) en el momento $t+n$;

$S_{(i,t+n)}$ = valor de la variable sintomática del área i en el momento $t+n$;

$S_{(.,t+n)}$ = valor de la variable sintomática para el total de las áreas en el momento $t+n$.

Ventajas:

a.- Necesita información para un único momento (el de la estimación) por lo que puede disponerse de un mayor número de variables sintomáticas a ser analizadas (no se necesita una serie histórica de la variable elegida);

b.- No es necesario compatibilizar las áreas geográficas en el tiempo; y,

c.- No hay cambios de definición o forma de recolección de las variables en el tiempo.

Desventajas:

a.- El supuesto en que se basa el método no es muy confiable, ya que, es difícil encontrar en la práctica una variable sintomática cuya distribución sea igual o muy similar a la de la población;

b.- Las estimaciones están afectadas por la calidad diferencial -por áreas- de la variable sintomática.

A pesar de sus desventajas, si hay disponibilidad de un conjunto de variables sintomáticas para el período a estimar, se podría entonces obtener varias estimaciones para la población, cuya media aritmética -de las estimaciones- se aproxime más a la realidad.

Distribución proporcional.

Este método supone que la población varía en igual proporción que la variable sintomática. Difiere del anterior en cuanto relaciona cambios de la variable sintomática y de la población entre dos momentos (en general entre el último censo y el año a

estimar).

Para la aplicación del método se necesita la información referente a la variable sintomática, por área, en dos momentos (censo y año a estimar), la población por área en el momento inicial (censo) y una estimación de la población del total de las áreas para el año a estimar.

La población estimada del área i en el momento $t+n$ está dada por:

$$P_{(i,t+n)} = P_{(i,t)} * \frac{S_{(i,t+n)}}{S_{(i,t)}} * F_a \quad (2)$$

$$F_a = \frac{P_{(i,t+n)}}{\sum [P_{(i,t)} * \frac{S_{(i,t+n)}}{S_{(i,t)}}]} \quad (3)$$

donde:

$P_{(i,t+n)}$; $P_{(i,t)}$; $S_{(i,t+n)}$ definidos anteriormente;

$P_{(i,t)}$ = población del área i en el momento t ;

$S_{(i,t)}$ = variable sintomática en el momento t , y;

F_a = factor de ajuste a la fórmula para que la suma de las $P_{(i,t+n)}$ sea igual a $P_{(i,t+n)}$.

Ventajas:

a.- Es sencillo, dado que se basa en una única variable sintomática;

b.- La incorporación de los cambios de la variable sintomática permite llevar a un mayor número de estimaciones; o sea, se podrían usar variables cuyas distribución no se asemeje a la de la población, siempre que su variación sea un buen indicador de los cambios de tamaño de la población.

Desventajas:

- a.- Necesita mayor disponibilidad de información que el anterior.
- b.- Se necesita compatibilización geográfica en los dos momentos, tanto a nivel de población como de la variable sintomática.
- c.- Las estimaciones están afectadas por cambios en la calidad de la información de la variable sintomática, si esta no es constante en el tiempo.

También en este modelo se puede calcular un conjunto de estimaciones y calcular la media de éstas, como una estimación probablemente más adecuada.

Correlación de razón. (Crosetti y Schmitt, 1954).

Este método se basa en el supuesto de que la evolución de la población está correlacionada con la variación de un conjunto de variables sintomáticas; la correlación se estima por medio de un modelo de regresión.

El modelo puede ser descrito de la siguiente forma:

$$Y_{(i,t)} = \beta_{(0)} + \beta_{(1)} * X1_{(i,t)} + + \beta_{(m)} * Xn_{(i,t)} + e_{(i)} \quad (4)$$

donde:

$$Y_{(i,t)} = \frac{\frac{P_{(i,t)}}{P_{(.,t)}}}{\frac{P_{(i,0)}}{P_{(.,0)}}} \quad (5)$$

$$X_{(i,t)} = \frac{\frac{S_{(i,t)}}{S_{(i,0)}}}{\frac{S_{(.,t)}}{S_{(.,0)}}} \quad (6)$$

$Y_{(i,t)}$ = razón entre la proporción de población del área i en el momento 0 y el momento t ;

$X_{j(i,0)}$, ($j = 1, \dots, n$) = razón entre la proporción de la variable j del área i en el momento 0 y el momento t ;

$e_{(i)}$ = error en la estimación según el modelo.

La población del área i en el momento $t+n$, se encuentra estimando el modelo de la ecuación (4), basado en el período $(0,t)$, y con la posterior estimación de $Y_{(i,t+n)}$, con base al período $t,t+n$, o sea:

$$P_{(i,t+n)} = \hat{Y}_{(i,t+n)} * \frac{P_{(i,t)}}{P_{(.,t)}} * P_{(.,t+n)} \quad (7)$$

donde:

$Y_{(i,t+n)}$ = razón de proporción de población del área i , estimada, para el período $t,t+n$, por el modelo de regresión.

Ventajas:

- a.- La principal ventaja es que se basa en una o más variables cuya evolución explica la mayor parte de la variación de la población;
- b.- Las estimaciones no están afectadas por errores de calidad en la información, siempre y cuando éstas se mantengan en el tiempo o cambien en forma similar en todas las áreas consideradas;
- c.- La estimación del modelo es independiente de las estimaciones realizadas, o sea, las áreas utilizadas para estimar el modelo no necesariamente deben ser las mismas áreas que se pretende estimar.

Desventajas:.

- a.- Necesidad de disponibilidad de una mayor cantidad de información;
- b.- Necesidad de compatibilizar la información en los dos momentos utilizados para estimar el modelo;
- c.- Necesidad de compatibilizar la información en los dos momentos necesarios para hacer la estimación.

Correlación de diferencia. (O'Hare, 1976).

Este método es similar al anterior, se basa en el mismo supuesto general, y la diferencia consiste en la forma en que se calculan las variaciones; las razones son sustituidas por diferencias, así se tiene:

$$W_{(t,t)} = \beta_{(0)} + \beta_{(1)} * Z1_{(t,t)} + + \beta_{(n)} * Zn_{(t,t)} + \epsilon_{(t)} \quad (8)$$

donde:

$$W_{(t,t)} = \frac{P_{(t,t)}}{P_{(.,t)}} - \frac{P_{(t,0)}}{P_{(.,0)}} \quad (9)$$

$$Z_{j(i,t)} = \frac{S_{j(i,t)}}{S_{(.,t)}} - \frac{S_{j(i,0)}}{S_{(.,0)}} \quad (10)$$

$W_{(i,0)}$ = diferencia entre la proporción de población del área i entre el momento 0 y el momento t;

$Z_{j(i,t)}$ = diferencia entre la proporción de la variable j del área i entre el momento 0 y el momento t; y,

$e_{(i)}$ = error de la estimación según el modelo.

La población del área i en el momento t + n, estará dada por:

$$P_{(i,t+n)} = \left[\hat{W}_{(i,t+n)} + \frac{P_{(i,t)}}{P_{(.,t)}} \right] * P_{(.,t+n)} \quad (11)$$

donde:

$W_{(i,t+n)}$ = diferencia de proporciones de población del área i estimada, para el período t,t+n, por el modelo de regresión.

Este método presenta las mismas ventajas y desventajas del método anterior. Sin embargo, O'Hare (1976) sugiere calcular las variaciones de proporción medio de diferencias, pues así se tendría una mayor intercorrelación entre las variables (W y Zj) y, además, las estimaciones estarían menos afectadas por cambios temporales de las variables.

Otra característica importante de los métodos de correlación, es que se pueden construir modelos distintos para subgrupos de áreas. Esto es especialmente

importante para aquellos casos en que se considera que un modelo único no representa adecuadamente a todas las divisiones geográficas, para las cuales se desea estimar la población. Se podría generar modelos de regresión para grupos homogéneos de áreas, por ejemplo, según grandes regiones, según predominancia urbana o rural, según la actividad económica principal, etc.

CAPITULO # 4

Capítulo # 4.

Resultados obtenidos a través de los Métodos de Estimaciones de Población con variables sintomáticas.

A través de los métodos de estimaciones de población de variables sintomáticas se estimó la población del año 1992 con las variables sintomáticas Nacimientos y Defunciones ocurridas en cada una de las comunas consideradas. Además se estimó la población haciendo uso de una función obtenida mediante la técnica de regresión múltiple por el procedimiento Stepwise (paso a paso), con la participación de cuatro variables sintomáticas, ellas son: número de nacimientos ocurridos, número de defunciones ocurridas, matrícula en educación básica registrada y número de automóviles. Los resultados de esta función serán detallados posteriormente.

A modo de mostrar la efectividad de cada uno de los procedimientos empleados se presentan los siguientes resultados:

En el cuadro # 8 se muestra una comparación de las proyecciones obtenidas para el año 1992 mediante el uso del Análisis de Varianza Unifactorial (Ver Apéndice). Como se puede apreciar, no se detectan diferencias significativas entre los resultados finales de cada una de las proyecciones al ser el valor de significación de F_{prob} mayor a 0.05 lo que indica que la población media nacional en cada una de las estimaciones es igual, confirmándose de esta forma que al hacer uso de la variable nacimiento como de la variable defunciones se llega a resultados semejantes con el método de Distribución Proporcional. Por otra parte, existe un dato de suma importancia que le da valor, de cierta manera, a la prueba realizada y es el hecho que entre los requisitos de esta técnica estadística está el cumplimiento de la homogeneidad de varianza, en tal caso, como en los demás que se verán más adelante, este supuesto o principio del análisis de varianza se cumple y la prueba tiene validez necesaria.

La utilización de esta técnica permite tener una idea de como se comportan los valores medios de cada una de estas estimaciones, e incluso conocer por ejemplo que el 95% de las comunas analizadas tienen tamaños de población que fluctúan entre 27163 y 66443 habitantes en el caso de las estimaciones realizadas con la variable nacimientos, y entre 26226 y 67379 habitantes para las estimaciones realizadas con la variable defunciones.

En el mismo cuadro # 8 se muestran las estimaciones realizadas por los métodos llamados Correlación de Razón y Correlación de diferencias. Estos dos últimos métodos exigen el cálculo de una ecuación de regresión para luego hacer las estimaciones. En este caso se utilizaron las variables sintomáticas nacimientos y defunciones de los censos 1970 y 1982 como base en las estimaciones de los parámetros de dichas ecuaciones. Para los cálculos de las mismas se utilizó el procedimiento paso a paso, el cual consiste en ir analizando las variables que se incorporan a la función en dependencia de su relación con la variables dependiente a través del coeficiente de correlación múltiple.

Para las estimaciones de población de 1992 mediante el método Correlación de Razón fue necesario calcular dos ecuaciones de regresión debido a las altas diferencias entre los volúmenes de población de cada comuna, existían algunas con valores de 2000 y otras con valores de aproximadamente 2 millones de habitantes.

La primera de las ecuaciones calculadas fue para todas aquellas comunas cuyos valores de población estaban entre 2000 a 10000 y la segunda para las mayores de 10000. Los modelos de las funciones tienen la forma siguiente:

$$Y_{est} = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \epsilon_i$$

cuyos parámetros son los siguientes:

Cuadro # 4. Parámetros y supuestos del modelo de regresión.
(Método Correlación de Razón)

PARAMETROS (Mayores 10000)	VALORES
Constante	6.21066E-01
Coefficiente β_1	2.863711E-01
Coefficiente β_2	5.243294E-01
F de Ratio	114.54
F probabilidad	1.000
Correlación R^2	0.5505
Prob($H < H_0$) ¹⁴	0.9986
Sqrt(n) * D > D* ¹⁵	0.82 < 0.89

Fuente: Cálculos propios. Software Stepwise¹⁶

Llama la atención el hecho de que la función no cumple con los requisitos establecidos por el análisis de regresión en cuanto a la homogeneidad de varianza y esto puede estar dado por las diferencias de tamaños de población entre las distintas comunas de la base de datos (Ver Apéndice), sin embargo los residuales siguen una distribución normal lo que de cierta forma permite realizar las estimaciones.

Los valores estimados por la función no provocaron diferencias significativas con respecto a la población real, así como tampoco con respecto a las demás

¹⁴ Prueba de homogeneidad de varianza, (Ver Apéndice)

¹⁵ Prueba de normalidad de los residuos, (Ver Apéndice)

¹⁶ Desarrollado en el Instituto de Informática de la Universidad Central de Las Villas, Villa Clara, Cuba.

estimaciones.

La ecuación calculada para las comunas menores de 10000 y mayores de 2000, es una ecuación confiable, sin embargo, es importante destacar que en la misma la variable nacimiento no es muy consistente, su incorporación a la función no aporta mucho en la estimación (Ver Anexo), su participación en la función no provoca un alza del valor del coeficiente de correlación así como tampoco una disminución en la varianza residual S^2 . A pesar de ello la ecuación cumple con todos los supuestos descritos en el Apéndice.

Cuadro # 5. Parámetros y supuestos del modelo de regresión.
(Método Correlación de Razón)

PARAMETROS 2000 a 10000	VALORES
Constante	8.014603E-01
Coefficiente β_1	-5.68646E-03
Coefficiente β_2	3.62210E-02
F de Ratio	64.21
F probabilidad	1.000
Correlación R^2	0.62
Prob($H < H_{obs}$)	0.3529
Sqrt(n)*D > D*	0.32 < 2.50

Fuente: Cálculos propios. Software Stepwise.

En las estimaciones de población por el método de Correlación de Diferencia también fue necesario calcular dos ecuaciones de regresión debido a las razones antes mencionadas y cuyos resultados son:

Cuadro # 6. Parámetros y supuestos del modelo de regresión.
(Método Correlación de Diferencia)

PARAMETROS Mayores 10000	VALORES
Constante	-3.928205E-05
Coefficiente β_1	-2.456516E-01
Coefficiente β_2	6.26606E-01
F de Ratio	350.85
F probabilidad	1.000
Correlación R^2	0.7985
Prob($H < H_{obs}$)	0.9999
Sqrt(n) * D > D*	1.04 > 0.89

Fuente: Cálculos propios. Software Stepwise.

Es importante observar que en esta ecuación los valores de correlación de las variables con la función son ligeramente altos en el sentido de que el coeficiente de correlación múltiple de 0.7885 indica que cerca del 80% de las estimaciones de población son expresados por las variables nacimiento y defunciones. EL supuesto de homogeneidad no es cumplido pues en este caso **Prob($H < H_{obs}$) = 0.9999**, al igual que el supuesto de normalidad. Sin embargo, las estimaciones no provocan diferencias significativas al 5% respecto al resto de las estimaciones realizadas por los demás métodos y la población real .

Los mejores resultados se observan cuando se calcula la función para las comunas menores de 10000, pero solo en cuanto a las hipótesis de los supuesto de la regresión ya que los niveles de relación de las variables con la función son muy bajos ($R^2 = 0.22$), o sea, las variables sólo son capaces de expresar hasta el 22% aproximadamente del tamaño de la población.

Cuadro # 7. Paramétros y supuestos del modelo de regresión.
(Método Correlación de Diferencia)

PARAMETROS 2000 a 10000	VALORES
Constante	-9.008416E-05
Coficiente β_1	4.031438E-02
Coficiente β_2	2.70319E-01
F de Ratio	11.81
F probabilidad	0.9999
Correlación R^2	0.2279
Prob(H < Hobs)	0.6312
Sqrt(n) * D > D*	1.59 < 2.50

Fuente: Cálculos propios. Software Stepwise.

Cuadro # 8. Comparación respecto a población censal 1992.

Año 1992	Resultados ONEWAY		
	Media	Desv.	Intervalo 95%
Estimación D.P. ¹⁷ Nacimiento	46803	168445	27163;66443
Estimación D.P. Defunciones	46803	176479	26226;67379
Método Correlación de Razón.	44810	169846	25006;64613
Método Correlación de Diferencia.	46125	183244	24760;67490
Población Real	46803	173813	26537;67069

¹⁷ D.P. "Método de distribución proporcional".

La validación estadística de esta prueba se demuestra a través del cálculo de la prueba F del análisis de varianza el cual es la relación que existe entre las varianzas dentro de los grupos y entre los grupos, su significación estadística se detalla en el cuadro siguiente:

Estadístico Asociado	Valor	Significación
F. Fischer	0.0070	0.9999
Cochrans C.	0.2207	0.4320

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile. Procesados en SPSS/PC ver. 4.1.

A pesar de cumplirse el principio de homogeneidad de varianza (Cochrans sig. > 0.05) se empleó la técnicas alternativa llamada Análisis de Varianza de clasificación simple Kruskal-Wallis (Ver Apéndice). El objetivo del uso de la misma surge por las dudas acerca de la normalidad de las variables originales, un pre-requisito del análisis de varianza unifactorial, además esta técnica permite llegar a otro tipo de información relacionadas con las estimaciones.

Como puede apreciarse en el Cuadro # 9, la comparación de conjunto que se realiza no muestra realmente que existan diferencias entre las estimaciones obtenidas mediante estos métodos y la población real del censo 92 (sig $X^2 > 0.05$) (Ver Apéndice, Prueba K-W).

Con el uso de esta prueba se espera que todos los métodos tengan los mismos valores medios de rango. Sin embargo si se observan las medias de rango de cada una de las estimaciones es comprensible la semejanza. Por ejemplo, el valor medio de los rangos asignados a cada uno de los valores de la población real es de 717.58 puntos. Los promedios de los rangos de las estimaciones obtenidas oscilan alrededor de este valor, siendo el más cercano el de las estimaciones del método de Correlación de Razón, o sea, que los valores estimados de población de estas comunas son muy cercanos a los que realmente tienen en el censo de 1992. Los que más se alejan de los reales son los estimados por el método de Distribución Proporcional con el uso de

la variable número de nacimientos registrados con 727.72 puntos. De cualquier manera, esto indica que en el orden aritmético que realiza la prueba algunos valores están por encima de los reales, al menos en una cantidad mayor. En sentido contrario están las estimaciones realizadas por el método de Correlación de Diferencia, donde una buena parte de las estimaciones (aproximadamente en 100 comunas) son inferiores a la población real.

Cuadro # 9. Análisis de las estimaciones obtenidas para 1992. (No paramétricos).

Conceptos	Resultados KRUSKALL-WALLIS	
	Casos	Mean Rank
Estimación D.P. Nacimiento	285	727.72
Estimación D.P. Defunciones	285	709.58
Método de Correlación de Razón	285	711.77
Método de Correlación de Diferencia.	285	698.35
Población Real 1992	285	717.58

Estadístico Asociado	Valor	Significación
Chis-square	0.7832	0.9407

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile. Procesados en SPSS/PC ver. 4.1.

El cuadro # 10 muestra los mismos resultado pero en forma más detallada, buscando las diferencias en pares de métodos, mediante la técnica Mann-Whitney (Ver Apéndice). Los resultados de la prueba indican que no existen diferencias relevantes entre las estimaciones de población de las comunas y su población censada.

Cuadro # 10. Comparación pareada entre las estimaciones. (No paramétrica).

Comparaciones	M A N N W H I T N E Y		
	Mean Rank	U	Significación
Población Real by D.P. Nacimiento	283.29 287.71	39981.5	0.7482
Población Real by D.P. Defunciones	287.14 283.86	40144.0	0.8116
Población Real by Método Correlación de Razón	286.61 284.39	40295.0	0.8717
Población Real by Método Correlación de Diferencia	289.54 281.46	39461.0	0.5581
Correlación de Diferencia by D.P. Nacimientos	279.81 291.19	38992.0	0.4098
Correlación de Diferencia by D.P. Defunciones	283.20 287.80	39957.5	0.7390
Correlación de Diferencia by Correlación de Razón	282.88 288.12	39865.5	0.7040
Correlación de Razón by D.P. Nacimientos	282.26 288.74	39689.0	0.6385
Correlación de Razón by D.P. Defunciones	286.00 285.00	40469.5	0.9420
D.P. Nacimientos by D.P. Defunciones	289.08 281.92	39593.0	0.6041

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile.
Procesados en SPSS/PC ver. 4.1.

En muchos, casos la diferencia entre los puntajes¹⁸ es mínima

¹⁸ Ver Apéndice las pruebas no paramétricas de K-W y M-W.

(aproximadamente

4 a 5 puntos) y en otros la diferencia de puntajes es cero. Este resultado es muy confiable pues, mientras más cercanos son los puntajes cuando comparamos con la población real esto indica que los valores de las estimaciones son extremadamente cercanos de los valores reales. Por otra parte, cuando comparamos los métodos entre sí, y las diferencias de puntajes es mínima, indica que se llegan a resultados similares por un método y otro, por ejemplo, los resultados entre las estimaciones del método de Correlación de Razón y el de Distribución Proporcional con la variable Defunciones son prácticamente los mismos, entre ellos solo hay un punto de diferencia.

En el trabajo se parte de hecho de que existen dos puntos de comparación, uno es la Población censada en 1992 y el otro las proyecciones realizadas en CELADE (Centro Latinoamericano de Demografía), mediante el Método de Relación de Cohorte.

El cuadro # 11 muestra las comparaciones realizadas con respecto a estas proyecciones y así como en el caso anterior, tampoco observamos diferencias relevantes entre los métodos empleados y las proyecciones de CELADE. En este caso también se tomaron en cuenta el cumplimiento de los supuestos o pre-requisitos de la prueba.

Cuadro # 11. Comparación respecto a Método de Relación de Cohorte.

Año 1992	Resultados ONEWAY		
	Media	Desv.	Intervalo 95%
Estimación D.P. Nacimiento	46803	168445	27163;66443
Estimación D.P. Defunciones	46803	176479	26226;67379
Método Correlación de Razón.	44810	169846	25006;64613
Método Correlación de Diferencia.	46125	183244	24760;67490
Método de Relación de Cohorte	47679	180443	26640;68718

Estadístico Asociado	Valor	Significación
F. Fischer	0.0105	0.9998
Cochrans C.	0.2173	0.6260

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile.
Procesados en SPSS/PC ver. 4.1.

El uso de la prueba no paramétrica para el análisis de conjunto muestran también que se llegan a los mismos resultados mediante el uso de los métodos de estimación a través de las variables sintomáticas que mediante el Método de Relación de Cohorte.

Cuadro # 12. Análisis de las estimaciones obtenidas para 1992. (No parámetros).

Conceptos	Resultados KRUSKALL-WALLIS	
	Casos	Mean Rank
Estimación D.P. Nacimiento	285	729.20
Estimación D.P. Defunciones	285	711.01
Método de Correlación de Razón	285	713.50
Método de Correlación de Diferencia.	285	700.09
Método de Relación de Cohorte	285	711.21

Estadístico Asociado	Valor	Significación
Chis-square	0.7346	0.9470

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile.
Procesados en SPSS/PC ver. 4.1.

En cuanto a la comparación individual de cada uno de los métodos con respecto al de Relación de Cohorte, tampoco se observan diferencias entre ellos. Las diferencias de puntajes son muy pequeñas, y en algunos casos están en el orden de las décimas como es caso de las estimaciones del método de Distribución Proporcional con la variable defunciones, lo que indica que se llegan a resultados semejantes tanto por un método que por el otro.

Cuadro # 13. Comparación pareada entre las estimaciones. (No paramétrica).

Comparaciones	MANN WHITNEY		
	Mean Rank	U	Significación
Método Relación de Cohorte by D.P. Nacimiento	281.81 289.19	39559.5	0.5922
Método Relación de Cohorte by D.P. Defunciones	285.71 285.29	40551.5	0.9752
Método Relación de Cohorte by Método Correlación de Razón	284.88 286.12	40437.0	0.9289
Método Relación de Cohorte by Método Correlación de Diferencia	287.81 283.19	39955.5	0.7382

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile.
Procesados en SPSS/PC ver. 4.1.

Junto a estos métodos de estimaciones de población fue calculada, como anteriormente se comentó, una función de regresión tomando como base para los cálculos de los parámetros las estadísticas vitales y la población censada de 1982. Posteriormente se procedió a estimar la población de 1992, sustituyendo en la función calculada, por los valores de las estadísticas vitales del propio año.

Para calcular la ecuación de regresión antes mencionada se utilizó el mismo procedimiento y software anteriormente descrito (Ver Apéndice, Stepwise).

Entre los resultados iniciales de la función¹⁹ está, en primer lugar, la matriz de correlación de las variables empleadas, como puede observarse en el cuadro # 14.

Cuadro # 14. Matriz de correlación de las variables .

Variables	Nacimientos	Defuncione	Matrícula Escolar	Número de Automóviles	Población 1982
Nacimientos	1.0000	0.9992	0.9994	0.9917	0.9999
Defunciones	0.9992	1.0000	0.9990	0.9902	0.9993
Matrícula Escolar	0.9994	0.9990	1.0000	0.9904	0.9995
Número de Automóviles	0.9917	0.9902	0.9904	1.0000	0.9919
Población 1982	0.9999	0.9993	0.9995	0.9919	1.0000

Fuente: Cálculos propios. Software Stepwise.

Se muestra que, los niveles de correlación entre las variables es sumamente alto (valores por debajo de la diagonal principal), todos, sin excepción, están por encima del 99%, sin embargo, el orden de importancia de las variables en la introducción en la función difiere un poco de los resultados preliminares.

La primera variable en introducirse en la ecuación fue los nacimientos ocurridos seguidos en orden ascendente por la matrícula escolar, las defunciones y por último el número de automóviles. De esta forma se conforma una ecuación con los siguientes coeficientes de regresión:

¹⁹ El modelo es el mismo utilizado para las anteriores funciones, ecuación número 4 del Capítulo 3.

$$\hat{Y} = -1.28E+02 + 2.8E+02 * Nac + 2.06E+01 * Def + 1.05 * ME + 4.970E-01 * Auto$$

La validación de los supuestos se muestran en el cuadro # 15.

Cuadro # 15. Supuestos del modelo de regresión.

PARAMETROS	VALORES
F de Ratio	555477.8
F probabilidad	1.000
Correlación R ²	0.9998
Varianza Residual S ²	2.753E + 06
Prob(H < Hobs)	0.7458
Sqrt(n) * D > D*	1.23 > 0.89

Fuente: Cálculos propios. Software Stepwise.

En primer lugar se logra que exista linealidad entre los parametros de la ecuación dado que el valor de **Fprob = 1.000** por tanto al menos un β_j ²⁰ es distinto de 0, el nivel de correlación de las variables con la función es del orden del 99%, o sea, las variables son capaces de describir el comportamiento de la población de las comunas en ese por ciento. Sin embargo este resultado no es del todo confiable, ya que el coeficiente de correlación puede estar influido por el número de variables que participan, y la varianza residual es en extremo alta, está en el orden de los millones y esto puede ser peligroso cuando se pretende hacer estimaciones a largo plazo ya que los resultados pueden ser muy distintos a los esperados. Por último no cumple con el requisito de normalidad de sus residuos.

²⁰ Recordemos que con la prueba F se testa la hipótesis siguiente:

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

H1 : Al menos un β_j es distinto de cero.

A pesar de las limitaciones de esta función, con ella se realizaron estimaciones de la población de 1992 y se compararon con el resto de las estimaciones, la población real y las proyecciones realizadas por el método de Relación de Cohorte. Los resultados de las comparaciones se muestran en el cuadro #16.

Cuadro # 16. Comparación respecto a Método Directo.

Año 1992	Resultados ONEWAY		
	Media	Desv.	Intervalo 95%
Estimación D.P. Nacimiento	46803	168445	27163;66443
Estimación D.P. Defunciones	46803	176479	26226;67379
Método Correlación de Razón.	44810	169846	25006;64613
Método Correlación de Diferencia.	46125	183244	24760;67490
Método de Relación de Cohorte	47679	180443	26640;68718
Población Real	46803	173813	26537;67069
Método Directo ²¹	42560	156954	24260;60860

Estadístico Asociado	Valor	Significación
F. Fischer	0.0286	0.9999
Cochrans C.	0.1604	0.4230

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile. Procesados en SPSS/PC ver. 4.1.

Una vez practicada la prueba convenienté se observa que no existen diferencias entre los distintos métodos utilizados para las estimaciones de población en relación

²¹ Ecuación de Regresión calculada a partir de las variables nacimiento, defunciones, matrícula escolar y número de automóviles.

con la población real y las proyecciones realizadas con el Método de Relación de Cohorte. Todos los supuestos son evaluados satisfactoriamente.

El cuadro # 17, muestra la comparación de conjunto entre los distintos métodos mediante la prueba alternativa no paramétrica, y en la misma se puede observar como los puntajes de los diferentes métodos son tan cercanos unos de otros. El de menor puntaje es sin dudas los que se obtienen con las estimaciones de la función de regresión, es decir, las estimaciones de población por este método están, -por decirlo de alguna forma- por debajo de los valores reales obtenidos en el censo y las demás estimaciones, pero no podemos olvidar que la misma tiene una varianza residual extremadamente alta y no cumple con el supuesto de normalidad.

Cuadro # 17. Análisis de las estimaciones obtenidas para 1992. (No parámetros).

Conceptos	Resultados KRUSKALL-WALLIS	
	Casos	Mean Rank
Estimación D.P. Nacimiento	285	1028.32
Estimación D.P. Defunciones	285	1002.44
Método de Correlación de Razón	285	1006.18
Método de Correlación de Diferencia.	285	986.76
Método de Relación de Cohorte	285	1002.99
Población Real 1992	285	1014.54
Método Directo	285	944.78

Estadístico Asociado	Valor	Significación
Chis-quare	3.6616	0.7224

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile. Procesados en SPSS/PC ver. 4.1.

Por último, en las comparaciones de a dos, se pone de manifiesto que las

estimaciones llevadas a efecto con esta ecuación están por debajo de las realizadas con cualquier otro método, pero la diferencia de estos puntajes no es lo suficientemente grande como para concluir que el modelo no es utilizable (en ningún caso los valores de significación están por debajo de 0.05). Todo lo contrario, es posible construir una función capaz de realizar estimaciones más certeras de las poblaciones de áreas pequeñas mediante las variables sintomáticas. Este modelo, por ejemplo, se aproxima bastante a esos resultados.

Cuadro # 18. Comparación pareada entre las estimaciones. (No paramétrica).

Comparaciones	MANN WHITNEY		
	Mean Rank	U	Significación
Método Directo by D.P. Nacimientos	273.59 297.41	37219.0	0.0843
Método Directo by D.P. Defunciones	277.42 293.58	38311.0	0.2417
Método Directo by Correlación de Razón	276.70 294.70	38104.5	0.2021
Método Directo by Correlación de Diferencia	279.79 291.21	38984.0	0.4075
Método Directo by Población Real 1992	275.25 295.75	37690.5	0.1372
Método Directo by Método de Relación de Cohorte	277.02 293.98	38197.0	0.2192

Fuente: Cálculos propios, Base de datos "Población de las comunas de Chile. Procesados en SPSS/PC ver. 4.1.

A modo de resumen se presenta el cuadro # 19, en el cual se describen los errores absolutos de cada uno de los métodos utilizados para estimar las poblaciones de las comunas.

Cuadro # 19. Resumen de resultados comparativos.²²

Método utilizado	Error Absoluto		Población estimada vs. Pob. Censada ²³		
	Medio	Desv.	< 10%	Normal	10% >
Distrib. Proporcional					
Var. Nacimientos	2.56	19.3	21.05	51.53	25.6
Var. Defunciones	-0.41	24.3	30.18	44.21	25.6
Correlación de Razón	-2.16	11.3	18.60	70.88	10.5
Correlac. de diferencia	-5.56	18.1	35.09	50.53	14.4
Método Directo	-1.23	16.6	23.8	60.70	15.4

Fuente: Cálculos propios.

Como puede apreciarse, el método de mejores resultados es el de Correlación de Razón, debido a que la desviación estándar de los errores absolutos es la menor de todos y el 70.88 por ciento de las comunas (aproximadamente 200), tienen un valores estimados que oscila entre un 10 por ciento menor y un 10 por ciento mayor de la población censada.

El método directo también ofrece buenos resultados a pesar de las limitaciones de la función, es capaz de estimar hasta un 60.7 por ciento de las comunas, la desviación estándar es la segunda más baja de todas, aunque en algunos casos (15.4 por ciento) realiza estimaciones por debajo del valor real.

²² Los errores absolutos fueron calculados como (Población Estimada - Población Censada) / Población Censada.

²³ Los cálculos fueron realizados de la siguientes forma:

- < 10%, a la población censada se le cálculo el 10% y se le restó; este valor se comparó con el estimado.
- >10%, a la población censada se le cálculo el 10% y se le sumó; este valor se comparó con el estimado.
- Normal, son todas aquellas que no cumplen ninguna de las condiciones anteriores.

CONCLUSIONES

Conclusiones.

* La población urbana de Cuba en el período intercensal (1970-1981) creció en 1524.1 miles de habitantes lo que representa un aumento de 22.7 por ciento con relación al censo de 1970.

* Las políticas y programas desarrolladas desde el triunfo de la revolución hasta el censo de 1981, con vista a disminuir las diferencias entre el campo y la ciudad, unido a su vez con las definiciones de lugar habitado rural, hicieron posible que (134) lugares que en 1970 eran considerados rurales pasaran a la clasificación de lugar habitado urbano lo que trajo consigo un aumento del 12.5 por ciento de la población urbana a nivel de país.

* La creación de (42) nuevas comunidades ayudó en un 3.9 por ciento al aumento de la población urbana a nivel nacional.

* Entre 1970 y 1981, la población urbana de la provincia de Villa Clara creció en 147.9 miles de habitantes, lo que representa el 28.1 por ciento del total urbano del país. Es uno de los mayores por cientos de crecimiento registrados a nivel nacional.

* A nivel nacional disminuye la población rural en 8.5 puntos con respecto al censo de 1970, mientras que Villa Clara lo hace en 14.7 puntos. En cifras absolutas, en el censo de 1981 se registra en la provincia la mayor disminución de población rural del país (82 mil habitantes).

* El 15.6 por ciento de los lugares habitados en 1981 que eran rurales en 1970 y que luego pasaron a urbanos (21), corresponden a la provincia de Villa Clara, de esta forma la población urbana de la provincia se incrementó en un 17.2 por ciento.

* Los nuevos lugares habitados urbanos (Comunidades) generados en la provincia provocaron un aumento de la población urbana del 2.8 por

ciento.

* En general el aumento de la población urbana de la provincia se genera, en parte por el crecimiento natural de la población de las áreas urbanas, en una considerable proporción por las migraciones del campo a la ciudad y por el factor ya apuntado de cambios de clasificación y creación de nuevas comunidades.

* Es plausible el uso de las variables sintomáticas en la estimación de población de áreas menores a través de los métodos descritos en el trabajo.

* Para el caso de Chile (fuente de datos utilizada en el trabajo) los mejores resultados son aportados con el método de Correlación de Razón, debido a que la desviación estándar de los errores absolutos es la menor de todas y el 70.8 por ciento de las comunas (200 aproximadamente) tienen un valor estimado que oscila entre un 10 por ciento menor y un 10 por ciento mayor de la población censada.

* Es posible realizar estimaciones de población de áreas pequeñas mediante la creación de una ecuación de regresión haciendo uso de las variables sintomáticas como el calculado en el presente trabajo.

* No es posible predecir con exactitud cual modelo o método es mejor sin considerar el país o región que se desea estimar, al menos por el momento, pues ello depende de las condiciones geográficas de cada país o región, en el caso de Chile resultó mejor el método antes mencionado pero puede que para Cuba o algunas de sus provincias no resulte ser el mismo.

* El método de Distribución Proporcional no es muy recomendable dada la magnitud de sus desventajas y en particular porque sólo permite hacer uso de una sola variable sintomática a la vez, puede que su mejor estimación sea el promedio de las estimaciones realizadas con cada una de las variables que se utilicen.

* En el caso de los métodos de Correlación de Razón y Correlación de Diferencia, que incluyen ecuaciones de regresión, es importante tomar muy en cuenta los supuestos del análisis de regresión, pues en la medida

que estos se cumplan pueden llegarse a obtener coeficientes de regresión más adecuados y con ello a un mejor modelo de base para las estimaciones.

RECOMENDACIONES

Recomendaciones sobre las estimaciones de población mediante el uso de variables sintomáticas.

* Para el caso de los métodos de Correlación de Razón y Correlación de Diferencia es importante tomar en cuenta el cálculo de la ecuación de regresión. Se plantea, según la metodología, obtener una ecuación de regresión lineal (Fórmula # 4 y 8, capítulo 3), que sirva de base en las estimaciones a realizar, sin embargo no siempre es posible llegar a una ecuación de este tipo que cumpla con los supuestos ya apuntados (Ver Apéndice), no obstante es posible llegar a otros tipos de modelos que no son precisamente lineales en cuanto a sus variables pero pueden lograrse por ejemplo;

$$\sqrt{Y} = \beta_0 + \beta_1 * X_1 + \dots + \beta_2 * X_2 + \epsilon_i$$

donde Y constituye la población a estimar y X_i las variables sintomáticas a utilizar por el modelo.

* No siempre un solo modelo de regresión para distintas áreas geográficas llega a ser el más adecuado, pueden constituirse distintas agrupaciones de áreas pequeñas y para ello existen diferentes métodos estadísticos como por ejemplo el Análisis de Cluster, el cual tiene como principio el de agrupar en grupos a los elementos comunes (localidades, áreas, comunas, etc) a partir de un conjunto de atributos que pueden ser las variables sintomáticas.

* Deben ser exploradas otras vías de estimación de población de áreas pequeñas, como por ejemplo las series de tiempo multivariadas, que incluyen métodos de regresión y que por demás están disponibles en algunos software como por ejemplo en el SPSS/PC a partir de la versión para DOS 4.1 hasta las versiones más avanzadas.

BIBLIOGRAFIA

BIBLIOGRAFIA

BRAVO, Jorge, (1993), Algunas observaciones acerca de los métodos de estimación de población a través de variables sintomáticas. Celade, mimeo.

CELADE (Centro Latinoamericano de Demografía), El uso de variables sintomáticas en la actualización de población de áreas menores. (versión preliminar).

----- (1993), Determinantes socioeconómicos de la migración interna, Santiago de Chile.

----- (1992), Transformaciones agrarias y población. El caso Cubano, Santiago de Chile.

----- (1980), Redistribución espacial de la población en América Latina Santiago de Chile, Serie E, N^{ro} 28.

----- (1976), Cuba, algunos aspectos demográficos de las áreas urbanas y rural de la provincia de Camagüey, San José de Costa Rica.

----- (1975), Migraciones internas. Teoría, método y factores sociológicos, Santiago de Chile, Serie E, N^{ro} 19.

CEPAL (Comisión Económica para América Latina y el Caribe)(1995 a), América Latina y El Caribe dinámica de la población y desarrollo, Santiago de Chile, Cuadernos de la CEPAL, LC/G. 1862-P.

----- (1995 b), Programa de acción regional para las mujeres de América Latina y El Caribe, 1995-2001, Santiago de Chile.

Comité Estatal de Estadística (1986), Población, La Habana, Series Cronológicas.

----- (1984), Estudios y datos sobre la población cubana, La Habana, Publicación 13.

----- (1982), Estudios y datos sobre la población cubana, La Habana, Publicación N^{ro} 11.

----- (1981), Estudios y datos sobre la población cubana, La Habana, Publicación N^{ero}. 10.

----- (1980), Estudios y datos sobre la población cubana, La Habana, Publicación N^{ero}. 9.

Grau Avalo, Ricardo, (1993), Apuntes para un libro, Villa Clara, Cuba.

INSIE (Instituto de Investigaciones Estadísticas) (1988 a), Revista de Estadística, La Habana, N^{ero}. 25.

----- (1988 b), Revista de Estadística, La Habana, N^{ero}. 24.

Naciones Unidas (1994), Boletín de Población de las Naciones Unidas, Nueva York, Serie N, N^{ero}. 36.

Revista Cuba Económica (1991). La Habana, año 1, N^{ero}. 2.

ANEXOS

Anexo # 1. Relación de Variables de los datos de Chile por comunas.

Variables	Definición de las variables
Reg	Región a la cual pertenece la comuna
Cód	Código de la comuna
X3	Población Censo 1970
X4	Nacimientos Registrados en 1970
X5	Defunciones Registradas en 1970
X6	Población Censo 1982
X7	Nacimientos Registrados en 1982
X8	Defunciones Registradas en 1982
X9	Matrícula en Educación Básica Registradas en 1982
X16	Automoviles motorizados registrados en 1982
X10	Población Censo 1992
X11	Nacimientos Registrados en 1992
X12	Defunciones Registradas en 1992
X13	Matrícula en Educación Básica Registradas en 1982
X14	Población Proyectada para 1992 según el "Método de relación de Cohortes"
X15	Condición urbano/rural observada en Censo de 1992
X17	Automoviles motorizados registrados en 1992

Reg	Cód	Año 1970			Año 1982					
		X3	X4	X5	X6	X7	X8	X9	X16	
1	1	92523	3019	580	147013	3479	628	28696	14299	
1	2	991	19	10	731	25	8	104	1204	
1	3	1767	66	27	3356	37	15	315	1232	
1	4	814	6	10	1087	23	4	253	599	
1	5	64977	1523	633	110991	2951	643	18388	14020	
1	6	2697	37	37	2171	42	25	307	167	
1	7	1262	31	16	1488	52	12	349	324	
1	8	716	0	0	1499	54	10	371	594	
1	9	1525	33	17	1796	28	15	417	719	
1	10	7936	192	47	5012	95	30	831	2725	
2	11	22318	652	227	22202	466	164	3982	1677	
2	12	14805	683	137	16022	413	59	3395	891	
2	13	70105	2492	600	100401	2680	418	19946	10685	
2	14	911	26	23	654	9	4	151	199	
2	15	1784	57	28	2578	67	19	549	161	
2	16	138798	3894	1381	195428	4379	1259	34969	15694	
2	17	3255	71	10	4417	69	21	656	1223	
3	18	12969	435	127	13000	286	89	2728	1279	
3	19	23800	963	158	26201	643	59	6200	2475	
3	20	51440	284	54	71612	1758	379	14026	4605	
3	21	3477	70	14	4713	98	22	826	523	
3	22	8130	284	54	7550	183	38	1563	1391	
3	23	36719	1155	389	42775	1024	219	9000	2310	
3	24	7121	152	58	5410	97	20	1185	709	
3	25	4981	163	44	7015	170	40	1303	466	
3	26	5251	116	70	5131	104	27	1120	526	
4	27	72141	2098	582	95118	2205	502	19414	8793	
4	28	5506	345	79	3648	85	17	751	434	
4	29	61423	1639	479	91997	2114	495	16698	5924	
4	30	9002	353	78	9559	219	66	2156	538	
4	31	15655	381	115	19502	378	86	3815	908	
4	32	4320	188	54	3934	119	26	792	202	
4	33	59038	1510	578	72825	1611	436	12690	3185	
4	34	4573	128	53	4795	121	37	1213	1511	
4	35	22823	686	242	26460	609	156	6082	827	
4	36	11990	340	147	12509	224	95	4168	333	
4	37	10596	265	79	9848	258	78	3907	343	
4	38	21822	615	210	25584	656	153	5367	1222	
4	39	18783	585	214	20928	477	115	5111	759	
4	40	10728	325	104	12546	317	65	2682	1494	
4	41	10246	298	132	10703	243	62	2313	335	
5	42	17199	447	174	21728	467	131	3906	1469	
5	43	7660	235	80	8969	199	70	1894	244	
5	44	12421	398	125	14281	373	96	3087	723	
5	45	2897	62	15	3818	86	17	593	1280	
5	46	2595	59	19	3393	73	19	547	942	
5	47	30501	790	308	40743	957	230	8168	3450	
5	48	8474	226	61	11005	217	60	1464	2886	
5	49	7166	180	51	9302	145	48	1522	633	
5	50	4132	132	28	4907	89	23	877	963	

5	51	36510	837	329	44730	1101	282	8217	3913
5	52	10615	297	130	11915	266	106	2019	515
5	53	8137	155	61	9720	204	52	1949	1187
5	54	4234	121	25	4787	141	18	805	513
5	55	14105	468	146	17851	429	114	3357	1014
5	56	8721	275	74	10200	240	52	1885	641
5	57	49937	1311	491	59264	1460	396	10315	3830
5	58	8958	225	82	9780	189	59	1639	1693
5	59	31690	706	260	40282	989	240	7154	2042
5	60	12875	471	124	16272	484	121	3107	1079
5	61	10194	140	47	12695	224	45	2277	833
5	62	23405	574	288	30695	699	233	5697	1641
5	63	9481	242	102	11361	306	72	1754	2011
5	64	254812	5998	2850	273039	6213	2348	46217	22331
5	65	188403	4411	1530	263775	5655	1489	41883	35032
5	66	11647	304	83	14740	337	85	2721	1275
5	67	7770	199	59	8395	165	49	1446	1485
5	68	56998	1346	434	85837	1862	528	14836	6018
5	69	37540	826	248	56690	1143	340	8888	3781
5	70	12304	360	117	14213	290	77	2943	1359
5	71	28291	1516	555	71465	1616	476	13188	7465
5	72	4632	194	80	13287	251	100	2215	2330
5	73	3153	104	37	9002	147	50	1590	2982
5	74	1585	39	10	1936	46	8	496	538
6	75	95030	2659	919	148758	3415	878	29028	12966
6	76	13537	385	186	18949	434	101	3938	1697
6	77	11473	306	108	15302	301	86	3437	1667
6	78	6698	138	42	7987	159	57	1184	905
6	79	28524	689	161	19711	448	100	3590	1617
6	80	5422	122	44	7892	168	27	1312	3860
6	81	11842	584	179	14396	292	78	3296	1154
6	82	27309	689	296	33725	803	234	6752	2504
6	83	8996	280	119	10262	241	83	1873	932
6	84	8036	197	78	8786	200	53	1625	978
6	85	26796	782	285	30850	629	213	6577	1753
6	86	13858	336	143	14906	253	88	3048	1124
6	87	11306	391	114	12348	260	72	2430	769
6	88	11836	250	99	13332	236	83	2547	864
6	89	4936	133	47	5356	132	41	947	377
6	90	8831	273	115	11903	315	74	2256	991
6	91	2585	322	127	14926	228	84	3076	900
6	92	41291	1117	434	47386	1251	336	9716	3997
6	93	21954	461	219	26613	453	152	5597	967
6	94	6483	149	50	7137	139	50	1319	601
6	95	8070	291	106	9401	269	85	1834	588
6	96	12726	314	96	14048	264	64	2837	618
6	97	34356	626	225	39409	704	229	8496	2139
6	98	7350	219	66	6661	118	42	1516	427
6	99	4039	76	34	3915	64	23	845	248
6	100	7282	245	77	8215	160	54	2024	416
6	101	7746	225	70	8844	177	70	2121	463
6	102	5932	173	55	5673	85	47	1214	682
6	103	5788	82	79	5140	86	30	1058	410
6	104	3393	67	42	2687	44	21	388	376

6	105	5502	134	42	5632	77	30	1175	352
6	106	6764	235	79	6522	144	60	1530	713
7	107	63062	1635	771	84698	1933	542	17042	5495
7	108	16615	577	159	19901	432	126	4370	1142
7	109	8720	245	69	10131	230	59	1984	1563
7	110	28815	817	428	32122	793	257	6484	2014
7	111	15164	347	136	16927	257	74	3446	547
7	112	8565	272	87	8193	127	64	2264	1465
7	113	5393	144	61	5295	139	38	1346	3798
7	114	3641	114	34	3740	101	20	761	2025
7	115	7005	162	81	7010	147	46	1445	1307
7	116	105529	2598	1162	142600	3343	924	26147	10505
7	117	10907	427	130	13095	308	105	3344	547
7	118	9579	229	70	11441	216	49	2660	294
7	119	30968	860	260	34246	679	193	7949	1278
7	120	10829	294	65	11619	261	89	3028	827
7	121	3244	95	36	4326	89	34	1840	96
7	122	8977	266	68	7814	152	60	1893	349
7	123	21928	513	252	31597	717	217	6747	2014
7	124	13108	352	131	12623	271	89	2711	350
7	125	55555	1487	696	62857	1757	557	14187	3669
7	126	11916	217	54	13615	223	68	2550	1325
7	127	12935	312	132	16786	360	101	3614	1572
7	128	21884	508	174	23442	469	143	5929	630
7	129	30064	997	431	34926	827	302	7329	2152
7	130	15356	460	92	18415	308	96	4601	432
7	131	14100	470	202	15224	308	120	3402	642
7	132	32588	1068	439	34578	691	317	7792	1697
7	133	38074	1011	404	39518	858	346	7842	2019
7	134	4687	84	40	5297	95	29	907	768
7	135	8269	253	83	8551	239	73	1930	174
8	136	104789	3338	1334	136496	3199	1003	28026	8432
8	137	37819	1502	488	43736	937	345	9548	2240
8	138	10684	176	45	12395	228	70	2749	88
8	139	4311	111	36	3862	64	14	1174	172
8	140	18264	475	198	22338	320	147	5069	625
8	141	7306	297	84	9235	168	69	2001	845
8	142	13542	392	148	15003	286	116	3707	310
8	143	13194	498	154	14076	295	120	3351	387
8	144	13747	339	155	14861	265	121	3392	532
8	145	7594	221	51	8187	168	68	1997	235
8	146	16295	542	181	18612	512	154	4415	761
8	147	13434	350	148	15026	268	143	3239	680
8	148	6815	140	66	7403	141	75	1344	233
8	149	7580	257	85	7172	134	66	2002	296
8	150	13082	476	198	15638	327	158	3401	494
8	151	5033	120	56	5960	92	45	1084	108
8	152	5667	171	57	6079	137	56	1454	120
8	153	9924	307	138	10936	283	132	2404	337
8	154	6374	187	58	7032	135	46	1764	79
8	155	7752	149	48	8005	134	57	2121	120
8	156	88278	3271	1114	115568	2876	853	26715	6128
8	157	15502	546	196	18607	394	154	4304	924
8	158	11750	386	131	11838	233	96	2661	573

8	159	10052	164	61	5266	120	36	1144	1048
8	160	12486	354	149	11438	173	67	2734	349
8	161	14835	304	129	16715	471	128	3483	472
8	162	4843	86	39	4615	69	26	1056	143
8	163	23781	830	306	27973	705	219	5861	1225
8	164	7809	199	74	7106	119	40	1602	370
8	165	17902	530	195	22990	541	173	4443	842
8	166	18651	709	218	22581	520	145	5107	887
8	167	5013	122	53	4432	81	24	878	255
8	168	20616	739	283	21548	467	258	4753	1254
8	169	217819	5588	2036	272679	6226	1748	54399	18037
8	170	150030	3843	1187	207219	4908	1086	36714	13805
8	171	28245	848	285	33360	871	224	6579	2106
8	172	43723	1453	481	47716	1055	378	10078	1496
8	173	8465	367	122	10757	211	96	2398	452
8	174	13501	437	144	15091	376	124	3154	1056
8	175	10395	284	110	11260	234	91	2668	448
8	176	50128	1886	553	48814	1228	324	10889	1731
8	177	58740	2373	696	70371	1691	430	14321	2205
8	178	17191	747	234	22445	560	129	4909	490
8	179	18796	546	192	24674	605	129	5216	828
8	180	21722	814	260	27692	717	161	6308	589
8	181	11631	390	128	14314	319	86	3587	342
8	182	15440	478	200	20533	509	141	4672	846
8	183	9929	191	81	12159	188	70	2645	108
8	184	3387	122	37	7075	166	63	1669	45
9	185	35048	1144	401	39599	942	269	8471	1707
9	186	7467	262	110	8717	186	41	2398	471
9	187	16446	461	173	19422	423	154	4786	718
9	188	9101	277	111	9694	259	75	2474	162
9	189	17315	547	231	18475	519	131	3932	783
9	190	7387	180	97	8863	149	66	2176	218
9	191	28376	788	317	31530	663	240	6862	1459
9	192	19681	474	237	20448	506	176	4194	827
9	193	13446	220	127	11883	271	105	2703	207
9	194	10254	290	96	12634	331	100	2898	251
9	195	8787	202	80	9341	166	67	2136	188
9	196	138300	3443	1472	189994	4736	1284	39899	10540
9	197	24855	740	381	27525	655	222	7249	1108
9	198	5381	142	71	5971	111	44	1370	494
9	199	20565	476	170	20245	358	109	4637	710
9	200	16119	510	160	16834	383	154	4145	523
9	201	6837	164	88	5593	120	39	1446	97
9	202	6464	120	67	6070	138	51	1578	199
9	203	10623	306	132	12589	420	118	2741	1102
9	204	27691	711	254	32973	720	198	6667	2545
9	205	19671	281	125	21486	357	124	4642	865
9	206	18935	434	207	19919	307	144	4322	1046
9	207	12514	391	153	14129	350	129	2821	1154
9	208	18932	439	206	23173	505	149	4428	930
9	209	11546	251	147	12178	305	119	2705	269
9	210	12570	82	31	13179	153	47	2901	222
9	211	12387	225	208	13890	282	147	3423	388
9	212	22774	414	266	24564	547	199	5707	560

9	213	30442	532	396	34051	711	283	7924	778
9	214	12096	133	84	13263	248	92	3459	225
10	215	92055	2290	911	109387	2346	710	20035	6046
10	216	15369	349	190	16960	395	126	3843	793
10	217	12962	337	145	12910	280	123	3227	559
10	218	15364	432	168	17651	370	111	3871	863
10	219	10722	304	126	12261	264	64	2961	578
10	220	5553	137	54	5298	92	42	1178	305
10	221	7230	142	57	7067	114	49	1690	646
10	222	31400	893	470	30192	749	219	7625	880
10	223	32140	803	330	35324	711	227	7138	1586
10	224	15621	425	180	18461	419	133	3896	847
10	225	28428	601	286	31786	479	200	7006	1289
10	226	8431	253	127	9862	276	104	2160	576
10	227	84130	2334	1017	110994	2517	830	22251	7785
10	228	11173	161	90	11308	203	88	2283	788
10	229	9866	163	66	11452	162	39	2755	957
10	230	11416	283	79	10568	162	60	2637	649
10	231	18735	569	239	20545	425	137	4246	867
10	232	15834	372	184	16983	361	118	3312	924
10	233	9522	161	166	8449	142	81	2806	682
10	234	81482	2197	768	103680	2388	717	20235	6024
10	235	20283	683	255	21831	448	151	4439	1589
10	236	4987	115	34	4498	88	32	611	59
10	237	22133	528	232	23965	409	190	4294	285
10	238	11947	322	137	13301	246	111	2552	304
10	239	14564	403	160	16228	226	97	2597	540
10	240	12546	293	133	12638	213	59	2265	385
10	241	12972	378	105	13466	273	94	1938	1006
10	242	12326	269	122	11954	233	61	1948	882
10	243	22899	649	259	26984	480	189	5392	1445
10	244	23931	649	221	29423	579	151	6162	1489
10	245	9422	229	108	9422	237	81	2223	282
10	246	6131	78	49	5848	69	28	1259	368
10	247	3525	43	41	3055	59	29	647	152
10	248	8339	205	111	9055	218	65	2338	169
10	249	4079	88	51	4374	109	44	938	40
10	250	9524	132	94	9594	165	76	1997	256
10	251	4341	169	64	4753	125	41	996	93
10	252	7887	128	85	10206	240	54	2297	172
10	253	4718	97	27	7016	82	24	1286	180
10	254	5624	165	65	6302	134	45	1530	225
10	255	2402	9	8	1809	53	13	416	236
10	256	2515	64	17	1848	38	10	452	196
11	257	23695	733	182	37305	955	205	7132	3159
11	258	980	13	7	931	12	6	426	0
11	259	13525	377	117	13538	319	83	3164	889
11	260	2143	43	6	3293	57	17	656	209
11	261	1103	30	21	1107	23	2	258	0
11	262	3483	248	63	7450	111	46	1675	409
11	263	1794	0	0	2161	53	10	453	104
11	264	226	0	0	284	5	1	54	0
11	265	227	0	0	292	6	0	84	0
12	266	7330	107	15	17383	305	122	3017	2008

12	267	35026	1573	536	103359	2353	635	14501	21847
12	268	2561	16	1	7625	69	37	764	1844
12	269	1673	14	2	2190	15	2	249	259
12	270	708	8	0	1357	5	1	195	0
13	271	2E+06	52712	17209	2462424	58448	14837	413291	158283
13	272	174290	4607	986	282510	6507	1093	42884	8490
13	273	273095	7139	1908	419317	9098	2253	53028	27207
13	274	164107	3854	755	272031	5401	1059	25653	60263
13	275	58698	1350	301	191883	4423	667	30513	21410
13	276	55048	1199	287	80452	1712	409	9229	13718
13	277	121437	1718	1001	115449	2083	838	13126	70724
13	278	18107	568	134	28776	747	168	5455	1927
13	279	12190	329	87	17834	372	85	3992	1684
13	280	9405	241	63	10412	241	65	1488	654
13	281	76694	2382	719	113211	2752	614	22085	6418
13	282	9103	280	62	10405	238	54	1931	1698
13	283	7968	260	44	8659	202	39	1557	1660
13	284	31325	1072	351	41790	989	213	8303	2843
13	285	21924	600	177	28021	632	154	5332	1524
13	286	6263	168	38	8936	196	38	1844	3658
13	287	50084	500	114	64267	1600	416	12704	4946
13	288	6011	128	25	7329	180	41	1182	1528
13	289	11767	286	70	14329	302	91	2931	1639
13	290	3456	109	31	3415	79	22	714	555
13	291	6303	138	52	6368	106	27	1104	695
13	292	22985	724	198	32193	868	188	6344	3972
13	293	37355	1062	252	62167	1368	322	11275	4016
13	294	14405	363	35	17207	308	94	3141	1511
13	295	14064	315	131	18712	427	97	2674	1395

A ñ o 1 9 9 2						
X10	X11	X12	X13	X14	X15	X17
169456	3519	688	28089	199859	1	18207
848	10	0	95	427	0	630
2803	17	3	383	4414	0	991
1012	14	10	185	1436	0	1433
151677	3605	758	23469	156518	1	32938
1972	25	17	262	1490	0	172
1422	37	6	296	1534	0	85
1555	38	9	302	2785	0	155
2512	34	15	392	1888	1	321
6322	115	23	1103	2902	1	1285
24985	534	166	3937	20490	1	1996
13660	340	21	2383	16091	1	1579
121807	3208	426	21192	119005	1	20478
443	3	0	74	442	0	152
2829	46	18	580	3249	1	248
240685	5287	1356	36147	230773	1	24142
6315	129	32	916	5775	1	507
13936	301	71	2260	12537	1	1776
27515	573	65	5032	28059	1	1970
100907	2474	485	16076	83005	1	13383
12061	276	32	2015	6408	1	1393
11724	290	35	1804	7101	1	1584
47248	1179	221	7925	47269	1	5444
5221	131	23	928	3956	1	368
7516	165	23	1159	9654	1	690
4745	71	20	805	4678	0	163
120816	2770	635	20168	120522	1	13085
3498	83	12	612	2338	0	2199
122766	2719	574	16866	119217	1	10389
12246	274	79	1924	9665	1	958
21660	475	109	3220	22445	1	1558
3772	86	19	488	3295	0	170
84982	1886	464	14229	85977	1	7282
5090	101	51	1031	4809	0	738
28374	641	155	5091	29897	0	552
14382	260	90	2433	12534	0	515
8723	203	57	1589	8801	0	247
29007	652	158	5023	29045	1	1765
23126	454	145	3904	22317	1	1179
15805	304	79	2736	14380	1	2358
10140	192	58	1829	11165	0	293
27322	571	148	4224	25469	1	3111
9273	206	66	1591	10822	1	362
17520	421	101	3075	16002	1	1269
4554	92	25	655	4951	1	1730
3896	71	17	548	4368	1	1120
49747	1060	300	8934	53818	1	5253
12153	231	69	1354	13984	0	2475
9860	171	52	1314	12066	0	945
5765	122	17	793	5873	1	1499

54591	1215	243	9236	52396	1	4729
12806	247	108	1589	12263	0	640
11575	218	50	1633	11243	0	1162
5900	138	40	851	5398	0	768
20276	430	114	3470	21540	1	1866
11295	218	69	1767	11508	1	1063
67007	1271	433	11089	66233	1	5889
10771	222	56	1425	10065	1	2090
45776	932	288	7173	48231	1	3303
18669	374	97	2730	19399	1	1696
13938	277	82	2339	15005	0	1440
34962	686	223	5754	38637	1	3100
12603	258	97	1763	12985	1	1471
283328	5751	2056	40063	300038	1	28025
304203	5667	1830	44364	316682	1	36109
17796	325	91	2618	17589	1	1528
10661	192	54	1384	9824	1	1015
104203	1998	681	16811	113949	1	7449
71672	1383	408	9424	75253	1	4388
16590	344	112	2385	15553	1	1620
84376	1656	507	13664	83764	1	8182
16419	279	103	2190	17796	1	1726
12065	207	48	1730	12997	1	2896
2764	70	10	546	2167	1	432
187324	4049	942	30091	205364	1	21116
22453	507	135	3602	23083	1	2295
18138	369	101	2849	18127	1	1467
9600	188	51	1303	8549	0	943
24152	443	133	3013	13702	1	2871
11332	241	41	1347	10761	1	3796
19432	435	87	2859	15416	1	1841
43617	992	272	6506	36773	1	3746
12252	256	64	1598	10669	0	1165
10782	246	46	1676	8746	0	1048
35167	678	219	5485	31616	0	3163
16594	286	96	2301	14401	0	835
12843	298	94	1680	11672	0	1126
15205	317	92	2202	13398	0	960
5823	119	39	897	5284	0	571
14578	302	86	2355	14142	1	1049
17738	351	106	2585	14442	0	2577
56368	1367	370	9761	49671	1	6434
30665	575	172	4643	30182	0	2212
7799	27	15	1119	7508	0	1154
14414	308	71	2006	10476	1	1398
14101	253	89	2225	14818	0	748
40122	700	219	6370	42850	0	4083
5944	98	39	973	5548	0	700
3773	59	32	566	3457	0	175
9144	148	63	1420	8594	0	687
10510	218	62	1565	9511	1	607
5423	75	25	798	5261	0	276
5466	88	33	786	4405	0	607
2779	49	19	348	2091	0	961

6209	112	43	861	5357	0	1082
6622	107	42	1118	6063	0	321
104113	2254	560	16864	111623	1	7552
24090	433	123	3447	23526	0	4120
11490	204	56	1713	11748	0	2753
35674	684	245	5480	35655	1	2712
16894	291	82	2605	18925	0	1891
9298	182	69	1483	8279	0	2798
6345	130	54	978	5363	0	1269
4931	92	22	712	3892	0	11399
7822	156	52	1072	6936	0	2094
171503	3749	994	27050	186796	1	16330
14194	299	85	2263	15261	0	663
13038	255	60	1894	13367	0	639
36414	702	239	5789	37154	0	2141
13769	266	48	1849	12086	0	535
4554	102	26	726	5868	0	234
7854	133	55	1174	6757	0	1067
40340	1001	220	6583	45499	1	3397
12285	223	80	1965	12232	0	577
77316	1669	454	13195	72415	1	5573
15500	256	59	1931	15847	0	1562
16950	335	104	2867	21957	0	936
28018	579	168	4561	25627	0	1001
38067	893	280	5991	41001	1	4265
19703	408	94	3249	22076	0	584
15150	230	113	2623	16571	0	863
35587	768	309	5674	36748	1	2254
40279	753	356	6442	41136	1	3045
5471	104	43	843	6181	0	1193
9492	215	55	1574	9299	0	306
166225	3384	1073	25021	164491	1	14580
48129	1010	378	7651	46456	1	3609
13156	175	74	2043	13328	0	152
3803	77	27	810	3203	0	82
22585	403	124	3609	25595	0	645
8932	193	66	1509	10775	0	566
16499	310	102	2927	15580	0	365
14161	254	98	2628	13838	0	567
15290	341	120	2440	14929	1	830
8413	153	48	1608	8182	0	354
19713	419	119	3374	19746	0	1461
14562	239	114	2330	15313	0	509
6404	81	37	1022	7479	0	217
5970	109	46	1164	6344	0	317
16630	351	123	2702	17331	1	941
5637	73	33	803	6593	0	84
6257	98	39	1067	6087	0	111
10971	230	79	1892	11134	1	546
6417	106	39	1120	7064	0	108
9495	142	62	1510	7553	0	399
140535	3254	789	23527	138313	1	10682
21705	460	136	3534	20433	1	1319
12020	258	77	2066	10387	0	570

4062	90	28	742	2475	0	544
10492	186	80	1790	8997	0	363
17257	384	134	2926	17464	0	522
4379	65	31	772	3828	0	171
29934	649	198	4864	30428	1	1736
8347	171	55	1221	5488	0	411
25994	621	166	4122	27555	1	1564
24350	560	106	4032	25245	1	1639
4375	81	22	620	3320	1	105
20460	409	159	3507	20606	1	1241
331027	6605	1777	52219	314953	1	37071
248543	5501	1070	34836	254542	1	15647
40359	798	197	5996	35429	1	2120
49284	1026	304	8050	49988	1	1695
10437	206	79	1773	12668	0	786
16156	347	92	2815	15814	1	3640
11957	244	79	1960	11288	0	638
50256	1112	296	8451	43607	1	2354
83426	1866	419	12580	77561	1	2884
24748	632	125	4357	26314	1	832
29657	674	144	5068	29760	1	1862
33631	742	167	6128	32103	1	1607
16870	335	103	3345	16240	1	483
29323	579	135	5483	24943	1	1398
6736	139	47	1200	13707	0	97
8736	168	65	2150	13577	0	322
46226	1052	269	7910	44412	1	3443
9197	180	53	1729	10320	1	491
22767	456	150	4125	23168	1	1090
9099	180	52	1860	10459	0	218
18135	378	137	3088	20095	1	1143
8842	179	68	1636	10750	0	294
32979	685	220	5845	34321	1	2613
20622	385	131	3625	21006	1	1330
12258	221	102	1874	10666	0	412
13917	325	97	2756	16018	0	539
8995	154	70	1547	10313	0	355
243561	5200	1398	39230	255186	1	23255
28725	583	192	5182	29620	1	1682
5886	122	39	970	6373	0	485
20887	338	128	3825	19728	0	1049
18339	316	127	3059	17011	0	891
5313	101	43	1014	4509	0	158
5978	103	45	1277	5791	0	191
14356	370	89	2281	14735	0	1258
35867	824	245	6093	37759	1	3019
22997	373	141	3825	23146	0	1285
20026	332	163	3318	20178	0	1297
14652	332	127	2157	15329	1	915
23643	438	152	3822	27878	1	1127
12061	223	65	2146	12931	0	261
15028	238	113	3065	13667	0	387
14432	245	108	2990	15403	0	142
25500	474	161	4660	26173	0	975

36878	592	273	7050	36872	0	1167
14076	259	94	2881	14684	0	341
122168	2591	738	17311	124518	1	10030
17952	392	116	3222	18680	0	1170
13757	290	109	2416	12499	1	953
18564	397	107	3198	19947	0	718
14048	311	74	2618	14093	0	651
5765	103	35	777	4788	1	36
7176	154	37	1296	6821	0	542
30162	669	217	5917	28364	0	1293
38740	776	226	6471	38005	1	2894
18152	378	136	2993	21735	1	1168
32981	613	228	5369	34862	0	1947
10460	152	64	1762	11300	0	433
127769	2878	807	20383	138488	1	12860
11178	201	62	1771	10917	0	1086
11207	167	46	1910	12804	0	1343
11051	193	54	1721	9537	0	794
20176	381	142	3180	21380	1	1405
16026	246	112	2459	17492	0	894
9778	116	63	2027	7187	0	303
129970	2903	757	22325	126982	1	13575
26529	558	156	4931	22304	1	2769
4361	96	30	992	3936	0	0
27027	565	173	6056	24373	0	474
17115	351	106	3018	15766	0	471
17054	276	92	3798	16756	0	663
13013	297	100	3193	11952	0	688
14386	312	87	2730	13218	0	1631
13107	254	74	2352	10965	0	1047
29931	658	216	4606	30519	1	3502
37516	816	208	5796	34318	1	2406
8188	142	58	1349	8788	0	152
7763	200	57	1174	5317	0	345
3021	47	26	484	2462	0	89
9088	165	67	1673	9346	0	164
4248	81	43	769	4651	0	76
10627	173	86	1563	9086	0	392
4952	107	36	789	5052	0	71
15055	331	68	2403	13469	0	553
7256	142	28	1154	9876	0	281
8104	157	36	1254	6611	0	96
1735	28	8	295	1107	1	95
1653	39	9	264	1077	0	92
43297	997	187	7744	52259	1	5500
1168	19	3	154	872	0	0
19090	421	102	3197	12241	1	1074
5353	98	20	775	5050	1	125
1283	23	8	230	986	0	0
6529	130	37	1152	7844	0	341
2996	69	12	515	2405	1	75
337	5	0	61	364	0	0
448	3	2	80	373	0	0
17757	299	122	2765	18974	1	2863

116511	2260	646	17331	131525	1	22029
6356	67	23	730	10910	1	829
1629	5	1	151	2812	0	262
1945	26	3	292	1598	1	28
2755211	58420	15496	401023	3E+06	1	215295
403742	9180	1610	52890	432885	1	17538
473064	9542	2592	56314	547977	1	52086
337500	5855	1450	40420	328009	1	96689
328881	6927	1194	56783	416307	1	26033
92410	1588	463	13107	97737	1	23811
111182	1769	1085	19655	83843	1	66361
52769	1156	211	7739	46500	1	3935
25033	622	161	3515	24519	1	2169
12838	280	56	1835	10809	0	948
254673	6047	970	33047	209161	1	11316
11646	238	73	1822	13021	1	1783
11368	263	56	1203	10554	0	3957
52792	1278	255	8194	52087	1	4673
37529	827	179	5432	34175	0	3318
11843	258	43	1739	12076	0	4181
80255	1798	467	12861	77941	1	6682
8735	175	44	1299	8555	0	935
19053	428	108	3123	17187	1	2138
4013	77	28	729	3256	0	137
6746	120	42	998	6979	0	825
44908	1120	192	8077	39550	1	4743
79520	1670	314	11103	93999	1	5012
20344	415	98	2914	18658	1	15921
21882	468	104	3483	22205	1	1457

APENDICE

Método de Regresión.¹

Mirado desde un punto de vista más general, el análisis de un proceso conduce en muchos casos a la concepción del mismo bajo el principio de la caja negra. Esto significa que dicho proceso se representa en forma de esquema como:



donde:

\vec{x} - el vector de variables de entrada (que incluye las variables controladas).

\vec{y} - es el vector de variables de salida.

y el rectángulo o caja negra es el proceso cuya esencia interna se desconoce exactamente y se desea investigar. En estos casos la investigación se formula como el problema de encontrar la función que relaciona las variables de entrada y salida a partir de conjuntos de valores experimentales de esas variables.

El objetivo del análisis de regresión y de los paquetes de programas que la realizan - con mayor o menor grado de generalidad- es precisamente determinar para cada componente y del vector Y , la función F que la relaciona con las componentes x_1, x_2, \dots, x_i del vector X , esto es:

$$Y = F (X_1 , X_2 , \dots , X_i) \quad (1)$$

Por lo tanto es necesario tener resultados experimentales donde para cada conjunto de variables de entrada se haya medido el valor de la variable de salida objeto de análisis.

El caso más simple, frecuente en la práctica, es aquel en que la función F es lineal, esto

¹ Tomado de los apuntes para un libro de texto, Autor: Dr. Ricardo Grau Avalo.

es, se buscan coeficientes β_j , tales que:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + + \beta_i * X_i + \epsilon$$

de donde: ϵ representa el error de estimación.

No siempre este modelo es el más adecuado y por ello son importantes también los casos en que, por ejemplo, la función F es cuadrática, más generalmente un polinomio, o incluso, una expresión más compleja en que aparezcan funciones trascendentes. Los problemas de regresión no lineal pueden ser reducidos a problemas de regresión lineal siempre y cuando la igualdad (1) pueda ser reducida a una cierta dependencia lineal entre funciones de las mismas variables mencionadas, o más precisamente, si (1) es equivalente a la determinación de los coeficientes β_j de una expresión como :

$$f_0 = \beta_0 + \beta_1 * f_1 + \beta_2 * f_2 + + \beta_m * f_m$$

con las funciones $f_i(x_1, x_2, \dots, x_i, y)$ arbitrarias pero conocidas.

De esta forma, si se hace una nueva denominación de las variables:

$$Y = f_0(x_1, \dots, x_i, y)$$

$$X_1 = f_1(x_1, \dots, x_i, y)$$

.....

$$X_m = f_m(x_1, \dots, x_i, y)$$

el problema se reduce al Análisis de Regresión Lineal Múltiple

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + + \beta_m * X_m + \epsilon$$

Por tanto, a partir del juego de datos experimentales se debe determinar según un nuevo juego de datos de la forma:

$$\begin{array}{cccc}
 X_{1,1} & X_{1,2} & \dots & X_{1,m} & Y_1 \\
 X_{2,1} & X_{2,2} & \dots & X_{2,m} & Y_2 \\
 & & & & \dots \\
 X_{n,1} & X_{n,2} & \dots & X_{n,m} & Y_n
 \end{array}$$

y encontrar los coeficientes β_j que cumplan con un cierto criterio de optimalidad de los errores ϵ_j .

Para su análisis teórico se resume en la forma matricial siguiente:

$$\vec{Y} = \vec{X} \beta + \epsilon \quad (2)$$

Cálculo de la ecuación de regresión.

La teoría del Análisis de regresión se formula en muchos textos de estadísticas e incluso de otras especialidades, pero frecuentemente se trata de una forma demasiado simplificada. En lo que sigue se destacan los aspectos que en general han sido poco tratados. Se trata de utilizar notaciones que faciliten la descripción posterior del programa utilizado para los cálculos de las ecuaciones de regresión usadas en las distintas metodologías llamado **STEPWISE**, elaborado en el Centro de Estudios de Informática de la Universidad Central de Las Villas, para realizar el Análisis de Regresión Lineal Múltiple.

Suponga que se desea obtener de la forma descrita en la expresión (2) donde:

\vec{Y} - es el vector de observaciones ($n * 1$).

\vec{X} - es la matriz de puntos muestrales ($n * p$).

$\vec{\beta}$ - es el vector de parámetros a estimar ($p * 1$).

$\vec{\epsilon}$ - es el vector de errores aleatorios independientes ($n * 1$).

con valor esperado o nulo [$E(\epsilon)=0$] y varianza constante [$V(\epsilon)=I\sigma^2$]. Es importante, destacar que el análisis de regresión lineal múltiple parte de las hipótesis de independencia de las observaciones Y_i , normalidad de los errores y homogeneidad de sus varianzas.

El criterio de optimalidad es minimizar la suma de los cuadrados de los errores lo cual conduce a las ecuaciones normales:

$$(\vec{X}'\vec{X})\vec{\beta} = \vec{X}'\vec{Y}$$

En el caso de que la matriz $X'X$ sea inversible existe el vector solución β y tiene las siguientes propiedades:

- 1.- Minimiza la suma de los cuadrados de los errores independientemente de su distribución.
- 2.- Los elementos de β son funciones lineales de las observaciones Y_1, Y_2, \dots, Y_n que dan estimados insesgados de los elementos de β con mínima varianza, independientemente de las propiedades de distribución de los errores.
- 3.- Si los errores son independientes y se distribuye normalmente con media 0 y varianza σ^2 ($\epsilon_i \sim N(0, \sigma^2)$), entonces el vector β es el estimador de máxima verosimilitud de β .

Después de obtener el vector β e independientemente de la distribución de los errores:

- 4.- Los valores estimados se obtienen por la matriz $Y_i = X_i * \beta$.
- 5.- Los errores o residuos están dados por $e_i = Y_i - \hat{Y}_i$.
- 6.- La varianza del vector de parámetros estimados se encuentra por:

$$\vec{V}(\vec{Y}) = (\vec{X}' \vec{X})^{-1} \sigma^2$$

7.- La varianza de la estimación en un punto X_0 ($1 \times p$) es:

$$\vec{V}(Y) = X_0 V(\beta) X_0' = X_0 (\vec{X}' \vec{X})^{-1} X_0' \sigma^2$$

8.- Al utilizar el estimado s^2 por σ^2 se puede encontrar un intervalo de confianza $1 - \alpha$ para Y_0 en X_0 como:

$$\hat{Y}_0 \pm t(n-p, 1 - \frac{\alpha}{2}) s \sqrt{X_0 (\vec{X}' \vec{X})^{-1} X_0'}$$

9.- Se cumple la igualdad:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 + \sum_{k=1}^n (\hat{Y}_k - \bar{Y})^2$$

*Total corregido
por la media*
Error
Regresión

Sobre esta base puede construirse la Tabla # 1 del Análisis de Varianza mediante el cual, si se supone que no existe falta de ajuste y el modelo es adecuado, tomando s_2 como estimación de σ_2 , puede probarse la hipótesis:

$$H_0 = \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

contra:

$$H_1 = \text{Existe } i \text{ tal que } \beta_i = 0 \text{ para } i = 1, 2, \dots, \text{ ó } p - 1.$$

Para esto, si $\text{Prob}(F_r, p-1, n-p) > 1 - \alpha$ puede considerarse el modelo como significativo, es decir, rechazar H_0 .

<i>ORIGEN</i>	<i>SUMA DE CUADRADOS</i>	<i>G. L.</i>	<i>CUADRADO MEDIO</i>	<i>RAZON</i>
<i>REGRESION</i>	$\sum_{k=1}^n (\hat{Y}_k - \bar{Y})^2$	$p - 1$	$\frac{S. C. R.}{p - 1}$	$F r = \frac{C.M.R.}{C.M.E.}$
<i>ERROR</i>	$\sum_{j=1}^n (Y_j - \hat{Y}_j)^2$	$n - p$	$s_2 = \frac{S.C.E.}{n - p}$	
<i>TOTAL</i>	$\sum_{i=1}^n (Y_i - \bar{Y})^2$	$n - 1$		

Como un dato de interés generalmente se calcula:

$$R^2 = 1 - \frac{ERROR}{TOTAL}$$

denominado coeficiente de correlación múltiple. Aunque por sí solo no permite probar la significación del modelo (pues no incluye el tamaño de la muestra), debe tenerse en cuenta que:

$$F r = \frac{R^2}{1 - R^2} * \frac{n - p}{p - 1}$$

tiene distribución de Fisher $F(p-1, n-p)$ por lo cual si se puede utilizar para probar la significación de la ecuación de regresión encontrada...

11.- Se cumple que:

$$\vec{\beta} \sim N \left(\vec{\beta}, (\mathbf{X}^t \mathbf{X})^{-1} \sigma^2 \right)$$

12.- Se obtiene una región de confianza $1 - \alpha$ para todos los coeficientes β_i por la ecuación:

$$(\vec{\beta} - \vec{B})^t \mathbf{X}^t \mathbf{X} (\vec{\beta} - \vec{B}) \leq p * s^2 * F(p, n - p, 1 - \alpha)$$

y para el error estándar estimado de b_i es igual a la raíz cuadrada del i -ésimo término de la diagonal de la matriz $(\mathbf{X}^t \mathbf{X})^{-1} s^2$.

Generalmente las dos regiones no son equivalentes por lo que siempre que sea posible debe tomarse la primera.

Examen de los residuos.

El examen de los residuos es necesario y útil, no solo porque permite comprobar la validez de los supuestos hechos en el Análisis de Regresión, sino también porque, en el caso de fallar algunos de ellos, da indicaciones para lograr su cumplimiento: cambiando la forma del modelo, transformando las variables, rechazando observaciones o utilizando pesos (Mínimos Cuadrados Ponderados).

Una vez que los residuos han sido calculados se presentan varias alternativas para su examen:

- 1.- Analizar la relación de los residuos y las variables.
- 2.- Analizar los residuos en conjunto para rechazar observaciones o probar si su distribución es normal.
- 3.- Graficar los residuos según el orden de ejecución de los experimentos para detectar si es necesario incluir el tiempo como variable independiente.

El método utilizado para el examen de los residuos fue el número dos de los mencionados anteriormente.

Prueba de normalidad.

Según la cantidad de observaciones n se aplica una de las tres pruebas siguientes:

- 1.- Prueba W si $(3 \leq n \leq 50)$.
- 2.- Prueba W^2 si $(50 \leq n \leq 100)$.
- 3.- Prueba de Kolmogorov-Smirnov si $(n \geq 100)$.

Tadas exigen que los errores sean colocados en orden creciente tal que $e_1 \leq e_2 \leq \dots \leq e_n$, y la de Kolmogorov-Smirnov requiere además de la definición de distribución de probabilidad empírica siguiente:

$$F_n(x) = \begin{cases} 0 & \text{para } x < e_1 \\ \frac{k}{n} & \text{para } e_k \leq x < e_{k+1} \\ 1 & \text{para } x \geq e_n \end{cases}$$

Observaciones extremas.

Probablemente la razón más importante para calcular los residuos sea detectar las observaciones extremas que tienen residuos muy grandes en comparación con las otras y deben ser tratadas especialmente.

La mayoría de los procedimientos para examinar residuos son sensibles a la presencia de observaciones extremas. Las pruebas numéricas para detectar la no adecuación del modelo y la varianza no constante a veces reaccionan a las observaciones extremas, de ahí que usualmente sea inseguro aplicar un procedimiento numérico a los residuos antes de haber eliminado las observaciones extremas (un resultado positivo podría indicar una observación extrema o la presencia de lo que debía detectar el procedimiento numérico).

Utilizar una regla fija para rechazar observaciones extremas da protección contra errores groseros, pero puede ocasionar que, aún si todas las observaciones fueran buenas y cumplieran los supuestos iniciales, se rechazara algunas de ellas, lo cual incrementaría la varianza de los parámetros estimados. Podría considerarse el porcentaje de incremento de esta varianza como el precio que se paga por la regla de rechazo, la cual es una garantía contra observaciones malas. Un tipo posible de regla a usar es:

"Rechazar la observación con el residuo de mayor magnitud y mayor que C_n , donde C es una constante dada. Si esta observación se rechaza, recalcular todos los residuos y s y aplicar

la regla nuevamente hasta que no haya más rechazos".

Prueba de homogeneidad de varianzas.

El método consiste en calcular un coeficiente de regresión lineal de los e_i^2 contra Y_i dividido por s^2 :

$$h = \frac{\sum_{i=1}^n e_i * (\hat{Y}_i - \bar{Y}_i)}{s^2 * H}$$

donde H, en general, es un poco más pequeña que la suma de los cuadrados del total, pero sin mucho error puede sustituido por ésta.

Para probar la significación de la desviación de h de cero se utiliza que:

$$V(h) = \frac{2(n-p)}{(n-p+2)H}$$

Si h es significativamente diferente de cero se puede elegir

la transformación potencia Y^p con el estimado:

$$p = 1 - 0.05 * h * Y$$

cuando $p = 0$. Si $p = 0$ se puede considerar que la transformación a efectuar es $\ln(Y)$.

Búsqueda de la mejor ecuación de regresión.

Generalmente se desea obtener un modelo que se ajuste a los datos experimentales con un número mínimo de variables por lo cual se han desarrollado diversos métodos que tienden a lograr este objetivo. Entre los métodos están:

- . Todas las regresiones posibles.
- . Eliminación hacia atrás (**BACKWARD**).
- . Selección hacia adelante (**FORDWARD**).
- . Paso a paso (**STEPWISE**).

Este método (**STEPWISE**) es una modificación al método anterior (**FORDWARD**), ha

sido utilizado para el cálculo de las distintas ecuaciones de regresión buscadas. Se adiciona un chequeo para determinar si alguna variable incluida en el modelo se hace no significativa.

Del máximo R^2 (R^2 - CHANGE).

En forma similar el Stepwise va incluyendo variable pero elige para incluir a la que logre un mayor aumento del coeficiente de correlación múltiple.

Los dos últimos métodos están considerados en la literatura consultada como los mejores. Sin embargo, las facilidades de interacción que brindan las micros y minicomputadoras pueden llevar a confeccionar programas de tal forma que el usuario elija el método a utilizar. En particular si el programa trabaja en modo automático sigue el método Stepwise, pero si el usuario lo desea puede decidir qué variables entran o salen de la ecuación en cualquier momento, de tal forma que pueden utilizar uno de los métodos anteriores o una mezcla de ellos.

El análisis de varianza unifactorial paramétrico. Hipótesis necesarias.

Son muchos los textos que de una manera u otra tratan el tema del Análisis de Varianza Unifactorial, por tanto, solo se hará una breve descripción de los supuestos utilizados y principios tomados en consideración para el desarrollo del trabajo.

El análisis de varianza paramétrico utilizado descansa en tres prerequisites fundamentales:

1^{ro}. - La independencia de las variables. El cumplimiento de esta hipótesis se logra normalmente con la independencia de los grupos en el diseño de investigación.

2^{do}. - La normalidad de las variables. Como en el caso de las pruebas verticales con el test de Student, esta hipótesis es usualmente verificada bien por investigaciones anteriores o a partir de un test aplicado a la misma muestra. En este caso debe probarse que en cada uno de los grupos $i = 1, 2, 3, \dots, p$, la variable objeto de medición tiene una distribución normal.

En algunos textos simples de estadística se pretende "**argumentar**" una reducción de la importancia de esta hipótesis con una explicación (**FALSA**) como esta:

"En última instancia la formulación matemática del análisis de varianza se basa no tanto en la distribución normal de las variables originales, sino en sus medias, y esto siempre se tiene si los volúmenes de las muestras son suficientemente grandes".

En realidad, la fundamentación matemática del test se basa en la distribución de **Fischer-Snedecor** para el estadígrafo F. Ella se hereda de las distribuciones X^2 de las sumas de

cuadrados correspondientes y estas a su vez de la normalidad de las variables. La normalidad de las medias -sin la normalidad de las variables originales- no basta para garantizar la normalidad de los términos que se elevan al cuadrado en esta sumas.

Si se incumple el pre-requisito de normalidad de las variables es siempre preferible utilizar alternativas no paramétricas.

3^{er}o. - La tercera hipótesis está relacionada con la homogeneidad de las varianzas. Este pre-requisito debe ser también verificado, bien por las investigaciones anteriores, o bien a partir de los datos de la muestra utilizando test especiales para esto. Lamentablemente los paquetes estadísticos no siempre incluyen tests de homogeneidad de las varianzas para velar por el cumplimiento de esta hipótesis. Existen varios test para probar esta hipótesis. Los más conocidos son:

.- El test C de Cochran que se basa en la razón de la máxima de las dispersiones dentro de grupos entre la suma de todas las dispersiones. Es un test aproximado para muestras de volumen grande, pero tiene la ventaja de que es poco sensible a las hipótesis de normalidad.

.- El test F de Bartlett-Box tiene un fundamento similar y es exacto (menos exigente para rechazar la normalidad que el test de Cochran) pero es muy sensible a la hipótesis de normalidad de las variables originales.

.- El test de max-min dispersión de Hartley que se basa en la relación entre la máxima y la mínima dispersión dentro de grupos.

Cuando estamos seguros de normalidad de las variables originales es preferible utilizar el test de F de Bartlett-Box. En condiciones de alguna duda sobre la normalidad de las variables, es preferible el test C de Cochran.

Si la hipótesis de homogeneidad de varianza falla, debemos tener al menos volúmenes similares de las muestras, (lo que se suele llamar experimento equilibrado) porque la distribución de Fischer Snedecor para F -y por tanto el análisis de varianza- es menos sensible a la violación de esta hipótesis cuando el diseño es equilibrado. Si no se puede tener volúmenes similares de muestra y falla esta hipótesis es necesario utilizar alternativas no paramétricas.

En términos generales constituye un craso error para un investigador desatender las hipótesis o pre-requisitos del análisis de varianza y aplicarlo indiscriminadamente. Los resultados obtenidos pueden ser absolutamente contrarios a los pretendidos o verdaderos.

Pruebas no paramétrica utilizadas en el trabajo.

La prueba KRUSKAL-WALLIS.(Análisis de Varianza de una clasificación por rangos.

"El análisis de varianza de una clasificación por rangos de Kruskal-Wallis es una prueba extremadamente útil para decidir si k muestras independientes son iguales o si aproximadamente siguen la misma distribución de probabilidad, cualquiera que sea esta. Los valores de las muestras generalmente difieren un poco y la cuestión radica en que las diferencias entre las muestras signifiquen diferencias genuinas o simplemente variaciones aleatorias."

"La técnica, en definitiva, examina la hipótesis de nulidad que supone que las k muestras poseen la misma distribución de probabilidad, con respecto a los promedios. La prueba supone que la variable en estudio tiene como base una distribución continua. Requiere, por lo menos, una medida ordinal."

Fundamento y Método.

Al calcular la prueba Kruskal-Wallis, cada una de las observaciones es reemplazada por rangos. Esto es, todos los puntajes de la k muestras combinadas se ordenan en una sola serie. El puntaje más pequeño es reemplazado por el rango 1, el siguiente en tamaño por el rango 2 y el más grande por el rango N . N es el número total de observaciones independientes en las k muestras.

Cuando se ha hecho esto, se encuentra la suma de los rangos de cada muestra (columna). La prueba determina si la desigualdad entre la suma de los rangos es tan grande que probablemente no proceden de la misma población o sus distribuciones de probabilidad son diferentes entre sí.

Puede demostrarse que si las k muestras efectivamente proceden de la misma población o de poblaciones idénticas, es decir, la hipótesis de nulidad es verdadera, entonces la H está distribuida como chi-cuadrada con $df=k-1$, siempre que los tamaños de las diferentes k muestras no sean demasiado pequeñas.

La test en cuestión es el siguiente:

$$H = \frac{12}{N(N + 1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N + 1)$$

donde:

k = número de muestras.

n_j = número de casos en la muestra de orden j .

N = sumatoria de n_j , el número de casos de todas las muestras combinadas.

R_j = suma de rangos en la muestra de orden j .

$\Sigma_{j=1}^k$ = indica sumar las k muestras (columnas).

está distribuida aproximadamente como chi-cuadrada con $df = k-1$, para tamaños muestrales (n_j) suficientemente grandes."²

La prueba U de Mann-Whitney.³

Cuando se ha logrado por lo menos, una medida ordinal, la prueba U de Mann-Whitney puede usarse para probar si dos grupos independientes han sido tomados de la misma población. Es una de las pruebas no paramétricas más poderosas y constituye la alternativa más útil ante la prueba t cuando el investigador desea evitar la suposiciones que ésta exige o si la medición en la investigación es más vaga que la escala de intervalo.

Supongamos que tenemos muestras de dos poblaciones A y B. La hipótesis de nulidad supondrá que A y B tienen la misma distribución.

Método.

Sea n_1 el número de casos del más pequeño de los dos grupos independientes, y n_2 el número de casos del más grande. Para aplicar la prueba U, se empieza por combinar las observaciones o puntajes de ambos grupos, para clasificarlos de menor a mayor. Aquí se clasifica el tamaño algebraico, es decir, las clasificaciones más bajas se asignan a los números negativos más grandes, si los hay.

En seguida se estudia uno de los grupos, digamos el grupo con n_1 casos. El valor de U (la estadística usada en esta prueba) es dado en la clasificación por el número de veces que un puntaje del grupo con n_2 casos precede a un puntaje del grupo con n_1 casos.

La hipótesis de nulidad a probar consiste en que ambas poblaciones siguen la misma distribución de probabilidad.

Para $n_2 > 20$ ha sido demostrado (Mann-Whitney 1947) que cuando n_1 y n_2 aumentan de tamaño, la distribución muestral de U se acerca rápidamente a la distribución normal, con

$$\text{Media} = \mu_u = \frac{n_1 * n_2}{2}$$

y

² Tomado de libro "Análisis experimental no paramétrico", pág 215.

³ Ident (1) , pág 145.

$$\text{desviación estándar} = \sqrt{\frac{(n_1)(n_2)(n_1+n_2+1)}{12}}$$

Es decir, cuando $n_2 > 20$ podemos determinar la significación de un valor observado de U por medio de:

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

z está prácticamente distribuida en forma normal con media cero y varianza de uno. La probabilidad asociada con la ocurrencia conforme a la hipótesis nula de valores extremos como un z observada puede determinarse, y se encuentra disponible en el software SPSS/PC.



