

9  
2ej



# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Facultad de Ciencias

## *DISTRIBUCION DEL INGRESO. UN ENFOQUE REGIONAL.*

T E S I S  
Que para obtener el título de  
A C T U A R I A  
p r e s e n t a  
LUCIA CAAMAÑO SIERRA



México, D. F.

1993

TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México



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

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

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

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

## CONTENIDO

	Página
<b>Introducción.</b>	
<b>Capítulo I ANTECEDENTES.</b>	
I.1 Principales indicadores económicos del crecimiento y desarrollo del país .....	1
I.1.1 El Porfiriato .....	1
I.1.2 Período de 1910 a 1934 .....	3
I.1.3 El Cardenismo 1934 - 1940 .....	5
I.1.4 Período de 1940 a 1970 .....	5
I.1.5 Período de 1970 a la fecha .....	6
I.1.5.1 Panorama general de la distribución del ingreso en México (1963, 1968, 1975 y 1977) .....	7
I.1.5.2 Gobierno de Luis Echeverría ...	15
I.1.5.3 Gobierno de José López Portillo	16
I.1.5.4 Gobierno de Miguel de la Madrid	18
I.1.5.5 Panorama general de la distribución del ingreso en México (1984 y 1989) .....	19
I.1.5.6 Gobierno de Carlos Salinas ....	23
I.2 Distrito Federal .....	26
<b>Capítulo II METODOLOGIA ESTADISTICA.</b>	
II.1 Importancia de la pronosticación y sus métodos .....	28
II.2 Análisis de regresión con datos de series de tiempo .....	33
II.3 Modelo de regresión bivariable lineal .....	33
II.3.1 Estimación de parámetros de regresión	37
II.3.2 Varianza de la regresión en la muestra .....	40
II.3.3 Inferencias acerca de los coeficientes de regresión de la población ....	41
II.4 Predicción y pronosticación .....	43
II.4.1 Predicción .....	44
II.4.2 Pronosticación .....	45
II.5 Análisis de correlación .....	49
II.5.1 Coeficiente de correlación de la	

población .....	49
II.5.2 Coeficiente de correlación de la muestra .....	50
II.6 Coeficiente de determinación y análisis de varianza en regresión lineal .....	53
II.7 Prueba F sobre $\beta$ .....	57
II.7.1 Cuadro AV en regresión lineal simple .....	58
II.8 Observaciones finales .....	60
<b>Capítulo III APLICACION DEL MODELO.</b>	
III.1 Crecimiento y desarrollo económico .....	63
III.2 Métodos cuantitativos para establecer estratos económicos y demográficos .....	66
III.2.1 Método y algoritmo de distribución por prorrateo .....	67
III.2.2 Curva de Lorenz .....	68
III.2.3 Coeficiente de Gini .....	69
III.3 Cuadros de aplicación .....	70
III.3.1 Población económicamente activa por grupos de ingreso mensual: 1970, 1980, 1990 .....	70
III.3.2 Población económicamente activa por grupos de ingreso mensual y frecuencias (base 1970): 1970, 1980 y 1990 .....	71
III.3.3 Cálculo del coeficiente de Gini en distribución de ingresos: 1970, 1980 y 1990 .....	73
III.3.4 Curvas de Lorenz .....	74
III.3.5 Pronóstico de los porcentajes de población económicamente activa por grupos de ingreso mensual ..	77
III.3.6 Regresión .....	78
III.3.7 Comentarios .....	88
III.4 Aplicación de prueba F sobre $\beta$ .....	90
III.4.1 Cuadros de prueba F .....	91
<b>Conclusiones .....</b>	<b>93</b>
<b>Apéndice .....</b>	<b>94</b>
<b>Referencias .....</b>	<b>95</b>

## INTRODUCCION

En pocas naciones del mundo los ricos son tan ricos y los pobres tan pobres como en México. Los contrastes de riqueza y miseria son realmente violentos. Frente a un pequeño grupo de mexicanos y extranjeros privilegiados que tienen todo y son en rigor los dueños de gran parte de la riqueza nacional, hay todavía una masa enorme de hombres y mujeres que nada tienen y que sólo trabajan para malcomer. Riqueza y miseria es un rasgo distintivo del paisaje social de México: ciudad y campo, colonias residenciales y barriadas proletarias, grandes y modernas avenidas y oscuras y sucias callejuelas, residencias impresionantes y vecindades deprimentes, negociantes prósperos y campesinos miserables.

Uno de los mayores problemas de nuestro país, derivado del modelo de crecimiento histórico seguido hasta hoy, lo constituye la concentración del ingreso en pocas manos. Esto genera una serie de consecuencias negativas para la población, entre las que destacan:

- .Bajos niveles de vida en general.
- .Poca capacidad adquisitiva (bajos ingresos).
- .Desempleo y Subempleo.
- .Analfabetismo y niveles educativos muy bajos.
- .Desnutrición y problemas de alimentación.
- .Insalubridad y malas condiciones higiénicas.
- .Niveles de inversión insuficientes.
- .Poca capacidad de ahorro interno.

Quizá uno de los problemas más graves derivados de la distribución del ingreso sea la desnutrición y en algunos casos el hambre. Ello a su vez

provoca una serie de consecuencias no sólo para los individuos que la padecen sino para el país en su conjunto, ya que en última instancia frena el desarrollo socioeconómico.

Otra consecuencia importante del problema es el aspecto educativo, el cual presenta problemas derivados de la mala distribución del ingreso y obliga a muchos menores de edad a trabajar para percibir ingresos que complementen los de su familia, abandonando los estudios en los diferentes niveles.

Carlos Duayhe ha estudiado el problema educativo y afirma que la consecuencia de la mala distribución del ingreso en buena parte es el carácter piramidal o selectivo de la educación. Normalmente el que mayores ingresos tiene puede aspirar a alcanzar más altos niveles educativos.

Otro indicador importante acerca de la mala distribución del ingreso en el territorio nacional lo constituyen los servicios de salud que se encuentran inadecuadamente distribuidos en las diferentes poblaciones de la República.

El problema de la distribución del ingreso se ve agravado por el creciente número de desempleados y subempleados y por la población joven (12-18 años) que representa una carga para la población económicamente activa que realmente trabaja.

El estudio de la distribución del ingreso es necesario para comprender el problema de la marginalidad que padecen un crecido número de mexicanos y, al mismo tiempo, para buscar la solución al problema, ya que la mala distribución del ingreso tiende a incrementarse afectando más a las familias que menos ingresos tienen. Concluyéndose que el empobrecimiento de grandes masas es una realidad en nuestro país y, que "sectores medios" y sobre todo "altos" de la población concentran más por ingresos.

El presente trabajo pretende medir el comportamiento estadístico de la distribución del ingreso, tomando en consideración los estratos socioeconómicos y la distribución del mismo para años posteriores.

## ANTECEDENTES

Al alborecer el siglo XIX, Alejandro von Humboldt vino a América; y después de recorrer vastas regiones del continente, al conocer lo que entonces era oficialmente la Nueva España, escribió: "México es el país de la desigualdad". Desde entonces han transcurrido casi dos siglos, lapso en el que México logró su emancipación política, perdió más de la mitad de su territorio, sufrió agresión extranjera, fué víctima de una larga dictadura, vivió una cruenta revolución social que dejó en los campos de batalla alrededor de un millón de seres humanos, fue golpeado por la depresión mundial de 1929; padeció un crecimiento concentrador de la riqueza, deterioro del poder adquisitivo del peso, devaluaciones, crisis económicas; fue objeto de varias inversiones extranjeras de capitales y un pseudo crecimiento económico. En esos dos siglos cambiaron mucho las cosas y quizá poco queda de lo que Humboldt consideró como característico de la Nueva España. Más a pesar de que la Nación no es lo mismo que hace cien o doscientos años, aún podría afirmarse que México es el país de la desigualdad.

### PRINCIPALES INDICADORES ECONOMICOS DEL CRECIMIENTO Y DESARROLLO DEL PAIS.

#### El porfiriato.

A nuestro país se le considera capitalista aproximadamente desde el último cuarto del siglo pasado, que corresponde a la época porfirista (1876-1910), debido a que varios fenómenos que ocurrieron durante este período, permitieron el avance y desarrollo de la acumulación de capital, tanto nacional como

extranjero.

A partir de 1880 se inició el proceso de expansión económica con la construcción de ferrocarriles financiados por inversiones extranjeras que también facilitaron el acondicionamiento de los puertos, la explotación de las minas, operaciones de establecimientos bancarios, transportes y grandes casas de comercio. La construcción de ferrocarriles, las inversiones extranjeras en la industria y la política proteccionista del gobierno dieron como resultado el crecimiento de numerosas ramas industriales del país.

Dependiente de una explotación creciente de los recursos naturales, por parte de los inversionistas extranjeros, quienes contaron con mano de obra barata, capital y tecnología, aumentaron la producción destinada a los mercados de ultramar, dando como resultado que el crecimiento económico durante el porfiriato fuera sostenido y superior al crecimiento de la población. Cabe mencionar algunos datos importantes:

.Entre 1900 y 1910, el producto interno bruto (PIB) creció 3.3% en promedio anual, en tanto que la población tuvo tasas de crecimiento anuales de 1.1% en el mismo periodo.

.La producción per cápita aumentó a un ritmo de 2.2% anual en el mismo decenio.

.La producción minera y petrolera era la actividad más dinámica, ya que durante los mismos años creció anualmente 7.2% en promedio.

.La producción manufacturera tuvo un ritmo de expansión de 3.6% anual, en tanto que la agricultura sólo lo hizo en 2.2%.

Estos elementos nos permiten corroborar que durante la época del porfiriato hubo un crecimiento económico capitalista, debido a que el crecimiento de las utilidades permitió a la industria asimilar el progreso tecnológico, mediante la adopción de equipo más moderno y fuentes más costeables de fuerza motriz, lo que trajo consigo una mayor productividad. Por



otra parte, las inversiones extranjeras y el creciente aumento de la infraestructura de los ferrocarriles, energía eléctrica y demás servicios, favorecieron la formación del mercado interno y permitieron a las fábricas operar en mayores escalas y en situaciones de costo más ventajosas, (lo cual reforzado por la baja tasa de salarios), facilitaron márgenes de utilidad que permitieron la formación de capitales, que para finales de la época porfirista habían llegado a sumar 3400 millones de pesos. Cabe mencionar que un tercio de esta suma se encontraba concentrada en ferrocarriles, 24% en minas y metalurgia, 15% en bonos de la deuda pública y 7% en servicios públicos, entre el 3 y 6% variaba la proporción invertida en bienes raíces, banca, comercio, industrias y petróleo.

Sin embargo, el desarrollo socioeconómico no se logró, ya que los beneficios de crecimiento económico no se distribuyeron entre todos los miembros de la sociedad mexicana, lo cual provocó problemas y conflictos debidos a la inestabilidad interna tanto económica como política y militar, que condujo al movimiento armado de 1910.

En efecto, el crecimiento económico porfirista propició la prosperidad de ciertas capas de la sociedad, pero la mayor parte de ésta quedó fuera del proceso de desarrollo, derivado del crecimiento de la población, el poco aumento de la ocupación productiva y la imposibilidad del sistema para alcanzar nuevamente una dinámica de crecimiento, originando el que el descontento se agudizara.

#### **Periodo de 1910 a 1934.**

Los movimientos armados iniciados en 1910 se prolongaron durante varios años, participando en ellos un elevado número de personas, lo que ocasionó una fuerte disminución de la actividad económica, debido principalmente a la mortalidad de la población y a la destrucción de bienes de capital, así como

al abandono de algunas actividades económicas.

Los efectos más negativos de la baja en la actividad económica, de 1910 a 1915, se manifiestan en la minería, especialmente en la disminución de la producción de oro, plata y plomo.

La producción manufacturera disminuyó 0.9% anual en promedio durante el lapso 1910-1921, en tanto que la agricultura mantuvo sus mismos niveles de producción entre 1910 y 1925.

Después de 1920, la actividad económica se recupera debido sobre todo al establecimiento de la paz y a la estabilidad política de los mandatos presidenciales de Obregón (1921-1924) y Calles (1924-1928).

Por lo que respecta al costo del movimiento armado, éste fue alto tanto en bienes de producción como en mano de obra y fue financiado por las actividades dedicadas a la exportación, como la minería y el petróleo.

La recuperación se empezó a dar en la década de 1920 y todavía no llegaba por completo cuando la economía fue golpeada por la depresión mundial de 1929. Así mismo, las actividades económicas disminuyeron de 1929 a 1932: el PIB bajó 5.6% y las actividades que tuvieron decrementos fueron la minería, las manufacturas y los transportes; el petróleo siguió descendiendo en forma drástica. En los primeros años de la década de 1930 la economía se empezó a recuperar, aunque a diferentes ritmos según la actividad; la agricultura se recuperó rápidamente, pero la minería y el petróleo crecieron a ritmos menores.

Aunque en el periodo 1910-1934 el crecimiento económico fue bajo, permitió el avance socioeconómico del país, lo que se manifestó en una mayor generación de energía eléctrica y un incremento de instalaciones telefónicas, así como adelantos técnicos en varias ramas de la producción. Estos hechos contribuyeron a elevar un poco el nivel de vida de cierta parte de la población, principalmente la que habitaba las ciudades, es por eso que en la

década de 1930 empieza la emigración rural en forma masiva buscando mejores niveles de vida.

#### **El cardenismo 1934-1940.**

Las acciones emprendidas por el presidente Cárdenas permiten sentar las bases del desarrollo capitalista del país. Destaca el espíritu nacionalista del gobierno y el fuerte impulso a la acumulación de capital, así como la creación de infraestructura.

De acuerdo con el Sistema de Cuentas Nacionales de la Secretaría de Programación y Presupuesto, el promedio anual de crecimiento del PIB a precios constantes fue de 4.48%, aunque con ritmos variables. En la década que corre de 1930 a 1940, el crecimiento promedio anual de la población fue de 1.7%, por lo que el crecimiento económico de 4.48% anual en el mismo período fue satisfactorio.

Tomando en cuenta los primeros cuatro años del sexenio cardenista de 1934 a 1938, el ritmo de crecimiento anual de algunas ramas fue: agricultura 3.6%, ganadería 2.7%, minerales metálicos 5.8% y petróleo 1.7%.

De acuerdo con las características socioeconómicas del periodo, se considera que hubo avance en el grado de desarrollo de la sociedad mexicana, lo que se tradujo en mejores niveles de vida para más mexicanos. Algunas evidencias de este mejoramiento se manifestaron en una mayor y mejor distribución de la producción y del ingreso, reparto de tierras, estímulo a los ejidatarios y pequeños empresarios, etc.

#### **Período de 1940 a 1970.**

Este período que abarca los sexenios de los presidentes: Manuel Avila Camacho (1940-1946), Miguel Alemán Valdés(1946-1952), Adolfo Ruiz Cortines (1952-1958), Adolfo López Mateos (1958-1964) y Gustavo Díaz Ordaz (1964-1970),

se caracteriza por un crecimiento económico sin desarrollo, o bien, crecimiento sin distribución del ingreso.

De acuerdo a las Estadísticas Históricas que presentan el producto interno bruto de 1941 a 1970, se puede afirmar que:

.El crecimiento promedio anual de la economía durante esos 30 años fue de 6.3% medido por el PIB a precios constantes de 1970.

.Los sexenios donde mayor crecimiento de la economía hubo fue en los de López Mateos (6.7% anual) y Díaz Ordaz (6.85 anual) conocidos como de desarrollo estabilizador.

.El crecimiento más bajo de este período fue en el sexenio de Miguel Alemán: 5.7% anual en promedio.

.El crecimiento de la economía en el sexenio de Avila Camacho fue de 6.1% anual y de 6.3% con Ruiz Cortinez.

El producto corriente per cápita pasó de 396 pesos en 1940 a 8546 en 1970, lo que representó un crecimiento de 20.5 veces; sin embargo, durante este período hubo una fuerte concentración del ingreso, por lo que los beneficios del crecimiento económico fueron sólo para una pequeña capa de la población, en tanto que las mayorías veían agravarse sus problemas y no mejoraban sus niveles de vida. Por ello, al período también se le conoce como de crecimiento concentrador del ingreso.

#### **Período de 1970 a la fecha.**

En el período de 1940-1970, la economía mexicana es impulsada por el Estado, teniendo tasas importantes de crecimiento -superiores al 6%- , en tanto que la población creció a tasas de 3% anual. Este crecimiento económico ayudó a la acumulación privada de capital y propició la formación y desarrollo del capital financiero, pero no atendió los problemas sociales que afectaban y aún siguen afectando a buena parte de la población nacional.

Ante esta situación, ya en 1963 se aplicaban encuestas sobre ingresos y gastos a la población para conocer de manera más exacta sus condiciones de vida. La información recopilada en 4 encuestas efectuadas en distintas épocas: Encuesta Sobre Ingresos y Gastos Familiares en México, 1963 y la Distribución del Ingreso en México, Encuesta Sobre Ingresos y Gastos de las Familias, 1968, ambas realizadas por el Banco de México, S.A. ; la Encuesta de Ingresos y Gastos Familiares, 1975, elaborada por el Centro Nacional de Información y Estadística del Trabajo (CENIET), de la Secretaría del Trabajo y Previsión Social y la Encuesta Nacional de Ingresos y Gastos de los Hogares, 1977, elaborada por la Coordinación General de los Servicios Nacionales, muestra los contrastes que caracterizan a la población en cuanto a niveles de bienestar, desde el punto de vista de su ingreso y gasto familiar.

*Panorama general de la distribución del ingreso en México.* De acuerdo con las encuestas correspondientes a los años 1963 y 1977, el 50% más pobre de la población recibió el 15% y el 16.7%, en esos respectivos años, del ingreso de toda la sociedad. De este grupo, el 20% menos favorecido aún, recibió el 3.5% y 3.3% del ingreso total; por otro lado el 20% más rico de los habitantes, obtuvo proporciones del ingreso que en 1963 fueron de 59.6% y en 1977 del 55.1%. Finalmente, de este 20% que contó con el mayor poder adquisitivo, el 5% de sus integrantes absorbió el 29.8% del ingreso total en el primero de esos años y del 25.5% en el segundo.

Conforme a la encuesta de 1963, el 18.4% de las familias tenían percepciones inferiores al salario mínimo mensual, cuyo promedio nacional era de \$373.20; esta proporción aumentó según la encuesta de 1977 hasta el 41.4% de los hogares, los cuales abarcaban al 37.2% de la población. Por otra parte, se advierte que el 2.6% de los hogares obtenían en ese último año, ingresos promedio superiores hasta en 11.7 veces el salario mínimo mensual, que

entonces ascendía a \$2 381.

Respecto a la distribución del ingreso entre el medio urbano y el rural, en las encuestas de 1963 y 1968, las familias del medio rural (que comprenden a las que vivían en localidades de 2500 habitantes o menos), recibieron en ambos años un poco más del 25% del ingreso, no obstante que representaban aproximadamente el 45% del total; en cambio, las que vivían en el sector urbano captaron cerca del 75% del ingreso y a él correspondían alrededor del 55% de las familias.

En cuanto a la distribución del ingreso por regiones geográficas, los datos disponibles no permiten una exposición amplia; sin embargo, se puede comparar la participación del Distrito Federal en el ingreso total, según las encuestas de 1963 y 1968.

En 1963 estaba concentrado en el Distrito Federal el 12.5% de las familias del país, las cuales percibían el 25.4% del ingreso total; en 1968, esos porcentajes eran de 13.9% y 31.9%, respectivamente. Y así, el ingreso promedio familiar en esta entidad fue, según la primera encuesta, 2.5 veces, y de acuerdo a la segunda 3.1 veces superior al prevaleciente en el medio rural y más del doble, en ambas ocasiones, que el ingreso promedio familiar del país.

Para 1977 en el área metropolitana de la ciudad de México estaba ubicado el 20.8% de los hogares del país y su participación en el ingreso ascendió al 35.2% del total nacional.

*Fuentes de ingreso familiar.* En relación a las fuentes de ingresos de familias, éstos se pueden agrupar básicamente en los provenientes del trabajo y los del capital.

Para 1963 el origen del ingreso promedio familiar fue el siguiente: el 62.7% provenía de sueldos y salarios; el 14.4% de ingresos obtenidos por campesinos y ejidatarios no jornaleros que explotaban sus propios predios; el 13.1% eran las ganancias de propietarios de empresas no agrícolas; el 8.1%

eran ingresos de capital y derivados de inversiones, y el 17% correspondía a otros ingresos.

Cabe señalar, que esta estructura es diferente para cada nivel de ingreso. En el caso de la población de escasos recursos económicos -aquellos que percibían hasta \$1500 mensuales-, la mayor parte de estos provenían básicamente de sueldos y salarios y de la venta de productos agrícolas; situación explicable, ya que en estos niveles se encontraban ubicados obreros, peones, jornaleros y pequeños agricultores, en cambio, en los sectores de altos ingresos, los provenientes del trabajo eran en términos relativos menores; y los ingresos por concepto de ganancias de empresas propias no agrícolas y de capital e inversiones adquieren mayor importancia.

A pesar de que las encuestas de 1963 y 1968 no son estrictamente comparables, se observa que las fuentes de ingreso de los distintos grupos de familias eran en términos generales, las mismas para ambos años. La mayor parte de los ingresos en las capas sociales de percepciones bajas provenían de la propiedad de pequeños negocios, de sueldos y salarios y de la venta de productos del campo. En los estratos superiores se reducía la participación de las remuneraciones al trabajo en el total de ingreso y aumentaba la proveniente de empresas propias de capital y de inversiones.

En 1975 podía observarse que en el promedio total, los ingresos por sueldos y salarios representaban el 63.6%. En el estrato más bajo de ingreso esa proporción era del 26.9%; ello obedecía a la elevada participación de las percepciones de los trabajadores por cuenta propia. Por otra parte, existía una mayor ponderación de los ingresos por concepto de sueldos y salarios en los estratos intermedios. Cabe destacar, que la proporción de ellos en el ingreso total de las familias varía de acuerdo a la ocupación de los jefes de familias; así, los que trabajaban como personal administrativo obtenían un 88.5% de sus percepciones por concepto de sueldos y salarios, mientras que los

que estaban dedicados a labores agropecuarias sólo recibían el 31.8% por ese concepto y el 39.3% por la venta de productos agrícolas. Lo anterior, también se presentó en el caso de los comerciantes y vendedores, los cuales recibieron sólo 44% de sus ingresos por salarios; el 50.1% provenía de negocios familiares y el restante 5.9% eran otros ingresos.

Para 1977 se mantuvieron como la principal fuente de ingreso los pagos al trabajo, los cuales representaron el 70% del total mensual; les siguieron en orden de importancia los ingresos obtenidos por renta empresarial con el 22.4%, las transferencias con el 5.3% y el restante 2.2% quedó distribuido entre las percepciones por arrendamiento de inmuebles y otros ingresos.

En el período que abarcan las 4 encuestas fue cada vez mayor el número de familias que fundamentalmente obtenían sus ingresos por concepto de sueldos y salarios; además, la participación de éstos en el ingreso total común aumentó básicamente en las familias de percepciones medias.

Los ingresos de campesinos y ejidatarios no jornaleros, originados en sus propias empresas, elevaron su importancia dentro de las familias con percepciones reducidas. El considerable porcentaje que representan los ingresos provenientes de la renta empresarial en los estratos de menores posibilidades económicas se explica por el acentuado crecimiento de vendedores ambulantes, comerciantes en pequeño y prestadores de toda clase de servicios que trabajan por cuenta propia y no perciben ingresos por sueldos y salarios.

La participación de los sueldos y salarios tiende a descender conforme se avanza en la escala de ingresos, mientras que aumenta la de los beneficios generados por operaciones con capitales; este es un patrón que se observa en las cuatro encuestas. Puede afirmarse que la distribución del ingreso en México, a pesar del aumento en la proporción percibida por los grupos de medianos ingresos, mantiene una estructura concentrada que se manifiesta tanto en el plano regional como en las percepciones alcanzadas por los diferentes



estratos poblacionales, situación que no es más que reflejo de las considerables disparidades sociales existentes en el país.

Es común que en los países capitalistas la distribución del ingreso presente agudos desequilibrios; y en las naciones no desarrolladas, tales desigualdades son aún más acentuadas. Ello obedece a que en estos países una gran proporción de la fuerza de trabajo está localizada en actividades cuya productividad es sumamente baja. En resumen; la desigualdad en la distribución del ingreso tiene algunas de sus causas en la heterogénea estructura productiva que priva en estos países y en la inequidad prevaleciente en el reparto del producto entre el factor capital y el factor trabajo.

*Estructura del gasto familiar.* En general, las familias de ingresos reducidos canalizan una gran proporción de los mismos a cubrir sus necesidades básicas y a medida que se dispone de mayores recursos pierde importancia, en términos relativos, la parte del gasto destinada a satisfacer dichas necesidades.

En promedio, durante 1963 las familias dedicaban el 73.3% del gasto a cubrir sus necesidades de alimentación, vestido y habitación. Las familias con menores niveles de ingresos destinaba el 89.2% a esos fines, mientras que en las ubicadas en los estratos superiores el gasto por estos conceptos era de 54.6%. Cabe destacar que el destino de las erogaciones familiares del medio urbano y rural difiere fundamentalmente en lo que respecta al gasto en alimentación, ya que mientras en el área urbana estaba canalizado a éste el 40.5% del ingreso, en las zonas rurales la proporción era de 58.4%.

Para 1968 los gastos de alimentación, vestido y habitación, en promedio representaron el 65.4% del gasto total; en cambio para las familias de altos ingresos esa proporción fue de 46%. En el medio urbano los gastos en alimentos sólo representaban el 37% del ingreso, en tanto que en el rural por este mismo concepto, erogaban el 50.3% del total de las percepciones.

En 1975 se advierte que las erogaciones de las familias mexicanas, principalmente de las de bajos ingresos, destinaban una gran proporción a cubrir necesidades esenciales, lo que limitaba otro tipo de gastos. El 69% de las familias dedicaba a la adquisición de alimentos un porcentaje superior al promedio nacional, para ese año estimado en 40% del gasto total. En ese mismo período, el gasto en alimentación, vestido y habitación, alcanzó en promedio el 71.5% del gasto familiar. En el medio rural la proporción destinada sólo para alimentación alcanzó el 55.5% y en el medio urbano fue de 36.7%.

Lo anterior indica que una considerable parte de la población -principalmente la radicada en el medio rural-, no tenía acceso a una gran cantidad de bienes y servicios y que, en consecuencia, constituía un mercado de dimensiones reducidas. Si bien los desembolsos destinados a la compra de productos de consumo duradero mantuvieron relativamente constante su participación a nivel global, en las familias de mayor volumen de gasto, aquéllos aumentaron de 7.6% en 1963 a 9.2% en 1975.

En lo concerniente a otro aspecto importante del gasto, como es el servicio médico, su participación en el total disminuyó de 3.9% en 1968 a 2.6% en 1975. En términos relativos, el gasto en servicio médico fue mayor en el medio rural que en el urbano, seguramente como consecuencia del hecho de que los servicios de seguridad social estaban más extendidos en las zonas citadas en segundo término. También se observa que el desembolso en servicios médicos tenía mayor peso en las familias de ingresos medios.

Los gastos educativos en las familias más pobres tenían poca significación, pues sólo llegaban al 0.5% del total, mientras que para familias más ricas representaba el 2.7%, según la encuesta de 1975; ello se debía a que los hijos de las primeras asistían a los servicios públicos de educación, en tanto que los de las segundas estudiaban preferentemente en las instituciones privadas de este tipo.

A pesar de que la encuesta de 1977 no es del todo comparable con las anteriores, el patrón de gasto mantuvo las características señaladas. Así, la estructura del gasto de los hogares en promedio nacional presentó los siguientes rasgos: 62.4% fue destinado al renglón de alimentos, vestido y habitación, el 11.7% a transportes y comunicaciones y el 8.2% a gastos en artículos domésticos, mientras que a los renglones de educación y cuidados médicos sólo dedicaron el 2.1% y 3.3% respectivamente. Ello obedeció, en gran medida, a que estos últimos son servicios que el Estado otorga en forma gratuita a amplios grupos sociales.

Los hogares con menor ingreso destinaron en promedio el 79.1% a gastos de alimentación, vestido y habitación, siendo ésta la cifra más significativa del gasto, mientras que en cuidados médicos y educación sólo dedicaron en promedio el 3.1% y el 0.7%, en el orden citado. En los estratos más altos de ingresos, el comportamiento de estas variables fue bastante diferente, ya que los gastos dedicados a transporte y comunicaciones ascendieron al 21.2% que se equiparan a los gastos de alimentación, los cuales representaron el 24%; seguían en orden de importancia los destinados a muebles y aparatos domésticos con el 10.3%, mientras que los renglones de menor cuantía los constituyeron la educación, con el 3.5% y los cuidados médicos, con el 3.8%.

Por lo relativo a los medios urbano y rural, para ese mismo año, el gasto en alimentos, bebidas y tabaco, alcanzó el 44.8% y 55.4%, respectivamente. Al comparar estos porcentajes con los obtenidos en la encuesta de 1975, se advierte que en el medio rural se registró un sustancial incremento en la proporción que tienen dichos gastos en el consumo total.

Las cifras anteriores permiten apreciar el fuerte impacto que ejercieron en la población urbana los acentuados incrementos de precios registrados en estos bienes posteriormente a la devaluación monetaria, así como el drástico descenso de la actividad económica en 1977.

*Aspectos Internacionales.* La comparación sólo fue posible con un número reducido de países seleccionados, debido a la escasez de estadísticas sobre el tema en ese tiempo.

En los países para los que existe información, la distribución de ingreso fue la siguiente: en los Estados Unidos, el 20% de la población de menores percepciones participó del 4.6% del ingreso; en orden ascendente en conjunto, el 50% de los habitantes percibían el 23.4% del ingreso, en tanto que el 20% de las personas de ingresos más elevados recibían el 45.5% y dentro de éstas, el 5% ubicado en la cúspide de la pirámide se adjudicaban el 20% del ingreso total para el año de 1965. En lo que respecta a Argentina, el 5.2% del ingreso fue recibido por el 20% de la población más pobre y 50% del total de habitantes participó con el 20.5% de aquél; por otra parte el 20% de los argentinos de más altas percepciones obtenían el 54.1% del ingreso y sólo el 5% de ellos captaba el 31.2% del mismo.

Las características distributivas en Venezuela eran las siguientes: por un lado, el 20% de la población más pobre sólo percibía el 3% del ingreso, mientras que el 5% de la población más rica se apropiaba del 25.5% del ingreso total. Asimismo, se observaba que en la República de Costa Rica el 20% más pobre de la población captaba sólo el 5.5% del ingreso y el 5% más rico de ésta recibía el 35% de aquél.

Por lo que se refiere a América Latina en su conjunto, el 20% de la población recibía el 3.1% del ingreso, mientras que en promedio el 5% de la población latinoamericana más rica percibía el 33.4% del ingreso de todos esos países.

Al suponer que este esquema distributivo prevalecía en la mayoría de las naciones no desarrolladas, la consecuencia sería que el destino de los gastos en estos países estaba centrado en mayor proporción en cubrir las necesidades elementales de sus habitantes, tales como alimentos, vestido y vivienda. En

este sentido, cabe destacar que en países como Tanzania, las familias destinaban en promedio el 66.3% de su gasto a cubrir necesidades alimentarias, en cambio, en los países desarrollados este porcentaje descendía a 33% en el caso de Francia y a sólo 18% en el caso de Estados Unidos.

Desafortunadamente, la aplicación de este tipo de encuestas es muy costoso, pero permite trazar un plan gubernamental en pro del bienestar social, dando avances en una justa distribución del ingreso, no obstante, se aplicaron en México Las Encuestas Nacionales de Ingresos y Gastos de los Hogares 1984 y 1989, que se tratarán en su momento y cuya información nos da un panorama de lo ocurrido con los ingresos y gastos al pasar de los años y nos permite prospectivamente identificar hacia donde nos dirigimos socioeconómicamente.

Al continuar con nuestro estudio histórico, tenemos que como respuesta a los problemas sociales que afectaban a la población durante el gobierno de Luis Echeverría, se plantea una nueva estrategia económica llamada "desarrollo compartido", cuyos objetivos eran: crear un mayor número de empleos, suficientemente remunerados y socialmente útiles, aumentar el poder adquisitivo de los grupos menos favorecidos, tanto por un imperativo de justicia como para fortalecer nuestro mercado interno y dar así apoyo a las tareas productivas del país; alcanzar óptimo aprovechamiento de nuestros recursos humanos y naturales, al disminuir la concentración económica y obtener un desarrollo regional más equilibrado, al incorporar mano de obra campesina en las actividades manufactureras, al estimular la pequeña y mediana industria, al promover la fabricación de bienes de consumo popular, aumentando la inversión del estado en esta rama, al subordinar la inversión directa de las compañías transnacionales a los imperativos del país, al mejorar la eficiencia de la planta productiva existente, al fomentar la fabricación de

bienes de capital, al impulsar y proteger la inversión nacional y señalarle a la inversión extranjera los campos complementarios en los que podía actuar, al utilizar y crear tecnologías acordes con nuestra realidad y romper los vínculos de sometimiento de la industria mexicana, así como fortalecer a través de ella, nuestra independencia económica; ya que la desigualdad económica que se manifestaba en una desigual distribución del ingreso tanto regionalmente como por actividades económicas y por estratos sociales, representaban un poderoso obstáculo a la propiedad dinámica del crecimiento económico del país, por lo que se hacía urgente una redistribución del ingreso.

En síntesis, durante el periodo 1970-1976 el ritmo de crecimiento de la economía fue menor que en décadas anteriores, a costa de constante y creciente subida de precios. Muchos mexicanos siguieron sin beneficiarse del crecimiento económico, especialmente las capas más bajas de la población. El poder adquisitivo del peso se deterioró en forma paulatina; se incrementó la dependencia del país. Muchos problemas económicos y sociales quedaron sin resolverse y otros más se agudizaron. La tónica no cambió: hubo crecimiento económico sin desarrollo. El saldo del sexenio fue la crisis económica.

El gobierno de José López Portillo (1976-1982) parte de una situación económica bastante deteriorada debido a los problemas ocasionados por la crisis y la recesión iniciada en 1976, además de la devaluación del peso y los convenios firmados con el Fondo Monetario Internacional (FMI).

El propio López Portillo afirmaba que se tenía que "administrar la crisis" y superarla, para lo cual se establecen tres medidas iniciales que van a ser fundamentales en todo el sexenio.

La reforma económica, llamada "Alianza para la Producción", que serviría para refuncionalizar todo el aparato productivo, haciendo que los factores

productivos coadyuvaran a salir de la crisis.

.La reforma administrativa que, aplicada en el sector público, permitiría agilizar todos los trámites administrativos, de manera que no existiera duplicidad de funciones, con el objeto de que coadyudara a la reforma económica a salir de la crisis.

.La reforma política, que complementa a las dos anteriores y pretende legitimar la acción del Estado para darle una mayor base de sustento y apoyo.

Las tasas de crecimiento durante el sexenio 1976-1982, medidas por el incremento del producto interno bruto a precios constantes de 1970 muestran:

.El crecimiento del PIB en el primer año del sexenio, fue muy bajo (3.4%) y similar al crecimiento de la población.

.De los años 1978 a 1981, el ritmo de crecimiento anual de la economía fue muy elevado (8.2%(78), 9.1%(79), 8.3%(80), 8.9%(81)), superior a las tasas históricas que se venían registrando.

.Este impresionante crecimiento del PIB se debió en buena medida al impulso que representaron la explotación y producción petrolera industrial, las cuales tuvieron tasas de crecimiento por arriba de las tasas del PIB.

.En 1982, la economía mexicana se derrumba y se vuelven a manifestar agudos síntomas de una crisis. Por primera vez, desde la década de 1920, no hubo crecimiento de la economía, sino una disminución de las actividades económicas que se manifiestan en una caída del PIB de -0.5.

.El promedio de crecimiento anual de la economía en el sexenio de López Portillo fue de 6%, igual al periodo de Luis Echeverría e inferior a los periodos anteriores.

Una vez más, en el sexenio 1976-1982 hubo un buen crecimiento de la economía (6% anual) comparado con el crecimiento de la población (3.2% anual);

sin embargo, esto se logró a base de expandir todo el sistema económico más allá de sus propias posibilidades de crecimiento.

El desarrollo socioeconómico durante el sexenio siguió ausente; los principales problemas económicos y sociales, lejos de resolverse, se agudizaron, como la elevada inflación, el constante y creciente endeudamiento interno y externo, el desempleo y subempleo, el deterioro salarial, las precarias condiciones de salud y educativas de muchos mexicanos, etc.

En síntesis, durante el periodo 1976-1982 se siguió dando un crecimiento sin desarrollo; es decir, crecimiento económico y concentración del ingreso en pocas manos. El saldo de este sexenio como el del anterior fue la crisis económica.

Cuando Miguel de la Madrid Hurtado asume la Presidencia de la República el 1o de diciembre de 1982, la situación económica del país era grave, por lo cual instrumenta un Programa Inmediato de Reordenación Económica (PIRE) que contiene diez puntos y que pretende resolver los difíciles problemas del país.

A pesar de todas las medidas tomadas por el gobierno de Miguel de la Madrid, los signos de la crisis se siguieron manifestando en 1983, primer año de su sexenio, y aún algunos síntomas críticos fueron persistentes y no se solucionaron en el corto plazo.

Al observar las tasas de crecimiento de la economía en los seis años del sexenio, se desprende que:

.En 1983 hubo una dramática caída del PIB (-4.2%), con lo que, por segundo año consecutivo, la economía mexicana, en lugar de crecer, disminuyó su producción ante el creciente aumento de la población.

.Esta caída repercutió en la economía en los años 1984 y 1985, en que hubo ritmos de crecimiento de 3.5% y 2.5%, muy por abajo de la tasa histórica de crecimiento y apenas arriba del crecimiento de la población, que es



de alrededor de 3%.

.En el año de 1986, se produce una nueva caída del PIB (-3.8%), con lo cual queda de manifiesto que los síntomas de la crisis siguen vigentes en el país.

.El promedio de crecimiento anual de la economía mexicana en el sexenio de Miguel de la Madrid es de apenas 0.13%, cifra muy baja que no alcanza a cubrir las crecientes necesidades de una población que va en aumento, (es decir, fue un sexenio perdido en términos de producción).

Es en 1984, habiendo pasado ya siete años, cuando se vuelve a aplicar una encuesta sobre ingresos y gastos, y dos años más tarde el INEGI lleva a cabo la Encuesta Nacional de Ingresos y Gastos de los Hogares 1989, las dos más recientes cuyos conceptos y definiciones son exactamente comparables.

*Panorama general de la distribución del ingreso en México.* De acuerdo con con las encuestas citadas, el 50% más pobre de la población recibió el 20.87% y 18.76%, respecto a 1984 y 1989, del ingreso de toda la sociedad. De este grupo, el 20% menos favorecido aún, recibió el 6.93% y 5.97% del ingreso total; por otro lado, el 20% más rico de los habitantes, obtuvo proporciones del ingreso que en 1984 fueron de 49.03% y en 1989 del 43.55%. Finalmente, de este 20% que contó con el mayor poder adquisitivo, el 10% de sus integrantes absorbió el 32.43% en el primero de esos años y el 37.93% en el segundo.

Conforme a la encuesta de 1984, el 16.79% de las familias tenían percepciones inferiores al salario mínimo mensual; esta proporción disminuyó según la encuesta de 1989 hasta el 8.5% de los hogares, los cuales abarcaban al 2.2% de la población.

Para fines de selección de la muestra, los hogares se identificaron a través de la vivienda utilizando los conceptos de Alta y Baja Densidad de Población. La zona de alta densidad de población estuvo constituida por

aquellos municipios del país que cumplieron al menos con alguna de las siguientes características según el X Censo General de Población y Vivienda:

- Tener al menos una localidad con 15 000 o más habitantes.
- Que el total de la población sea mayor o igual a 100 000 habitantes.
- Contener a la capital de la entidad.
- Formar parte de alguna de las 12 áreas metropolitanas que se utilizan en la Encuesta Nacional de Empleo Urbano: Ciudad de México, Guadalajara, Monterrey, León, Mérida, Chihuahua, San Luis Potosí, Puebla, Veracruz, Orizaba, Torreón y Tampico.

La Zona de baja densidad estaba formada por aquellos municipios que no pertenecían a ninguno de los grupos anteriores.

Respecto a la distribución del ingreso entre Zonas de Alta y Baja densidad, en las encuestas de 1984 y 1989 las familias del medio rural (que comprenden a las que vivían en zonas de baja densidad) recibieron 23.99% y 20.39% del ingreso respectivamente, no obstante que representaban aproximadamente 37.43% y 35.52% del total para esos años; en cambio los que vivían en el sector urbano captaron cerca del 80% del ingreso y a él correspondían alrededor de 60% de las familias.

*Fuentes de Ingresos en el hogar.* En relación a las fuentes de ingreso en los hogares, éstos se pueden agrupar básicamente en los provenientes del trabajo y los del capital.

Para 1984 el origen del ingreso corriente total promedio por hogar fue el siguiente: El 79.09% provenía del Ingreso Monetario y el 20.91% del Ingreso no Monetario. Dentro del Ingreso Corriente Total Promedio el 49.01% provenía de Remuneraciones al Trabajo; el 20.72% eran las ganancias de los propietarios de empresas; el 5.61% de ingresos obtenidos por transferencias (jubilaciones y pensiones, indemnizaciones, etc.); el 3.052% eran Ingresos de capital y derivados de inversiones; el 0.39% correspondía a otros ingresos y el 0.199% a

### Ingresos netos por cooperativas.

Cabe señalar que esta estructura es diferente para cada nivel de ingreso. En el caso de la población de escasos recursos económicos, la mayor parte de éstos provenía básicamente de la renta empresarial, en segundo término se encontraban las remuneraciones al trabajo, los regalos y donaciones originados en el país. En cambio, en los sectores de altos ingresos, los provenientes del trabajo son los que adquieren mayor importancia y los ingresos por concepto de ganancias de empresas propias son, en términos relativos, menores.

Si se comparan las encuestas de 1984 y 1989, se observa que las fuentes de ingreso corriente total promedio por hogar eran, en términos generales, las mismas para ambos años, con porcentajes del 77.44% proveniente del ingreso monetario y 22.55% del ingreso no monetario. En el ingreso monetario destacan los incrementos provenientes de la renta empresarial y transferencias y la disminución del proveniente de remuneraciones al trabajo.

En cambio, se observa que las fuentes de ingreso de los distintos tipos de familias tuvieron una estructura muy diferente en 1989 comparada con la de 1984. En 1989 podía observarse que en el promedio total, los ingresos por sueldos y salarios representaban el 46.35%. En el estrato más bajo de ingreso esa proporción era del 14.86%; ello obedecía a la elevada participación de las percepciones de los trabajadores por cuenta propia, es decir, el sector informal de la economía. Por otra parte, existía una mayor ponderación de los ingresos por concepto de sueldos y salarios en los estratos intermedios.

Puede afirmarse que la distribución del ingreso en México mantiene una estructura concentrada que se manifiesta tanto en el plano regional como en las percepciones alcanzadas por los diferentes estratos poblacionales como reflejo de las considerables disparidades sociales existentes en el país. Esto ya se observaba con las otras encuestas, pero lo alarmante después de estudiar

las de 1984 y 1989, es que ahora los ricos son más ricos en una proporción poblacional menor y que ha disminuido la población perteneciente a los grupos de medianos ingresos ya que ha pasado a formar parte de la población de bajos recursos.

*Estructura del Gasto del Hogar.* Las observaciones en este aspecto serán las relativas a los rasgos más significativos que distinguen al gasto familiar, en relación al nivel de ingresos.

En promedio, durante 1984 los hogares dedicaban el 51.022% del gasto a cubrir sus necesidades de alimentación, vestido y habitación. Las familias con menores niveles de ingresos destinaba el 61.93% a esos fines, mientras que en las ubicadas en los estratos superiores el gasto por estos conceptos era de 40.94%. Cabe destacar, que el destino de las erogaciones familiares del medio de alta densidad y baja densidad difieren fundamentalmente en lo que respecta al gasto en alimentación, ya que mientras que en la zona de alta densidad estaba canalizado a éste el 35.3% del ingreso, en las zonas de baja densidad la proporción era de 44.7%.

Para 1989 los gastos de alimentación, vestido y habitación, en promedio representaron 43.60% del gasto total; en cambio para las familias de altos ingresos esa proporción fue de 29.97%. En las zonas de alta densidad los gastos en alimentos sólo representaban 30.02% del ingreso, en tanto que en la zona de baja densidad por este mismo concepto, erogaban el 41.07%.

En 1989 se advierte que las erogaciones de los hogares mexicanos, principalmente de los de bajos ingresos, detinaban una gran proporción a cubrir necesidades esenciales, lo que limitaba otro tipo de gastos. El 78.46% de los hogares dedicaba a la adquisición de alimentos un porcentaje superior al promedio nacional para ese año, estimado en 29.78% del gasto total y este porcentaje se ha ido incrementando encuesta con encuesta.

En lo concerniente a otro aspecto importante del gasto, como lo es el

servicio médico, su participación en el total disminuyó de 2.7% en 1989 a 2.6% en 1989. En términos relativos, el gasto en servicio médico fue mayor en las zonas de baja densidad que en las de alta densidad, como consecuencia de que los servicios de seguridad social estaban más extendidos en las zonas citadas en segundo término. También se observa que el desembolso en servicios médicos tenía mayor peso en las familias de ingresos medios.

Los gastos educativos en las familias más pobres tenían poca significación, pues sólo llegaban al 1.16 en 1984 y es alarmante que esta proporción sólo aumentara en 0.19% para 1989 pues no indica que el costo de la educación haya disminuído, si no, que ya pocos integrantes de dichos hogares asistieron a recibir educación; mientras que para las familias más ricas representaban 6.18% en 1984 para aumentar a 9.96% en 1989 dedicado a las Instituciones privadas de este tipo.

Cabe destacar, que después del renglón de ingresos destinados a alimentos, se encuentra el de transporte y comunicaciones para los hogares de mayor ingreso durante las 2 encuestas citadas, mientras que para los hogares de bajos ingresos, en 1984, los renglones de mayor peso eran alimentos y muebles, conservándose para 1989 el primer renglón de alimentos y transformándose el segundo en transporte y comunicaciones, que se explica por la disminución de subsidios dedicados a éste último.

Por el ritmo de la economía mexicana se pensaba que difícilmente se alcanzarían tasas importantes de crecimiento en el sexenio de Salinas de Gortari, pero si recurrimos a las palabras pronunciadas por el Presidente de México ante el Congreso de la Unión en sus Informes de gobierno podemos extraer lo siguiente:

.El crecimiento económico dado en el primer semestre de 1991 fue casi de 5%, el más alto para ese período en los últimos 10 años. La inflación

acumulada en los primeros nueve meses fue la menor en 16 años.

.Logró abatir la inflación, que en 1987 superó 150%, gracias al establecimiento del Pacto para la Estabilidad y el Crecimiento Económico, que permitió la reducción del deslizamiento del tipo de cambio de 80 a 40 centavos promedio por día durante aproximadamente un año, ahora el deslizamiento es de 6 diezmilésimos de nuevos pesos.

.Entre enero y agosto de 1991 la tasa promedio de desempleo abierto en las principales ciudades del país fue de 2.6% frente a 3.6% registrada en 1988.

.El saldo real del ahorro financiero aumentó 13% entre septiembre de 1990 y 1991.

.Se creó el Programa Nacional de Solidaridad para atender servicios básicos, a lo que se incorporó más tarde el reclamo de fuentes de empleo.

.En 1992 la inflación de todo el año fue inferior a la que hace un lustro teníamos en un solo mes.

.En el primer semestre de 1992, el Producto registró una expansión de 2.8%

.La actividad económica reflejó, en el 1er semestre de 1992, un crecimiento anual de 9% real en inversión, de 10% en las exportaciones manufactureras y de 36% en las importaciones de Bienes de capital.

.La producción nacional sigue creciendo por cuarto año consecutivo a una tasa superior a la de la población.

.Con objeto de fortalecer el ahorro nacional y dar pasos adicionales a las prestaciones a los obreros de nuestro país se instauró el Sistema de Ahorro para el Retiro (SAR).

.El Estado ha incrementado en 70% real los recursos para la superación de rezagos y carencias urgentes en gran mayoría de los municipios del país con "Solidaridad".

Aún así los problemas sociales y económicos siguen vigentes en el México de hoy.

Sin perdernos del estudio del período de 1970 a la fecha, lo que sobresale es la inversión de capitales.

La inversión es un componente muy importante del ingreso nacional ya que permite la acumulación y consecuentemente el crecimiento económico; para el análisis de la inversión se ha dividido ésta en: pública, privada y extranjera, donde la inversión privada y pública conforman la inversión Bruta.

A lo largo de 25 años (1952-1977), la inversión bruta tuvo un crecimiento acelerado de 27.5 veces, esto quiere decir que en promedio se incrementó 1.1 veces anualmente. A través de este tiempo la inversión pública también ha variado su participación respecto al total. En 1952 la inversión pública representaba el 36.1% del total en tanto que la privada participaba con el 63.9%. Para 1963 la proporción había variado y la inversión pública representaba el 45.4% del total, y la privada fue de 54.6%. Para 1977 la primera aumentó hasta 55.1%, en tanto que la segunda se redujo a 44.9%. El capital y con él la capacidad productiva de la economía nacional, ha crecido en los últimos años. Pero, lejos de que México ostente la fisonomía propia de un país con un alto grado de desarrollo está atrás de las naciones industriales -en términos relativos cada vez más rezagado- ¿Qué fue entonces lo que sucedió con las grandes inversiones que se hicieron a nivel nacional?, ¿Quizá se hizo uso particular de estas grandes sumas para engrosar cuentas bancarias en el extranjero o canalizar la inversión en negocios personales?. Este tema no es fácil de tratar y menos con la información que está a disposición pública, además que desvía el objetivo del presente trabajo, así que continuando con nuestro punto de estudio...

Respecto a la distribución del capital en la República Mexicana se observa

una gran concentración en el D.F. que en 1960 absorbía ya el 25.7% de dicho capital, en 1965 participaba con el 31.3%, en tanto que el Estado de México lo hacía con el 16.7%, Nuevo León con el 11.9% y Veracruz con el 7.2%. Tomada en conjunto el área del Valle de México, concurre con el 48% de todo el capital industrial privado, y las entidades antes mencionadas, junto con Coahuila, Jalisco, Chihuahua y Puebla concentran el 81.4% del total, mientras que las 24 restantes apenas participan con el 19.6%. Por lo menos en 10 de ellas las inversiones industriales son realmente insignificantes, lo que pone de relieve un desarrollo pobre y sumamente desigual del proceso industrial.

#### **DISTRITO FEDERAL.**

Los apartados anteriores confirman que, el Distrito Federal, con una superficie que no llega a la milésima parte del territorio nacional, constituye el centro cultural, administrativo, político, industrial y financiero más importante del país; las gigantescas dimensiones alcanzadas por la zona metropolitana con su correlativa concentración de recursos productivos y, en consecuencia, de actividades económicas de elevada productividad, dan lugar al desequilibrio regional que manifiesta el sistema económico nacional, al que se califica de tener un desarrollo macrocefálico.

A tal concentración contribuyen factores de muy diversa índole que van desde acontecimientos históricos hasta la política de industrialización del país. En efecto, la ciudad de México nació sobre la antigua Tenochtitlán, por decisión del conquistador que así materializaba el triunfo sobre una cultura, convirtiéndose en la capital de la Nueva España. Esta decisión le dió a la ciudad las características de centro administrativo y desde entonces se convirtió en la mayor concentración urbana de la nueva colonia.

El mayor desarrollo relativo de la ciudad de México, con respecto al resto del país, le permitió ser el principal beneficiario de la política económica,



dando lugar a un proceso de crecimiento que aún no cesa. En efecto, los atractivos que presenta la ciudad de México para la inversión privada son mayores que las desventajas, por lo que los intentos de descentralización han enfrentado obstáculos.

Es por la importancia económica que tiene esta entidad federativa que la he elegido para hacer el estudio de la distribución del ingreso, ya que aquí se encuentra concentrada tanto población urbana como rural, producto de la migración; se llevan a cabo actividades de todo tipo y es políticamente hablando el lugar donde se toman las decisiones que afectan al resto de la República.

Cuadro 1. PRODUCTO INTERNO BRUTO TOTAL Y POR HABITANTE

Presidentes	Año	Producto Total (Millones de pesos)		Población (Miles de habitantes)	Producto por habitante (pesos)		Crecimiento (%)	
		Corrientes	De 1970		Corrientes	De 1970		
	1800	---	11212	---	---	1800		
	1843	---	10752	---	---	1434	-4.10	
	1890	---	10020	---	---	1254	-8.67	
	1877	---	15862	---	---	1567	58.57	
Porfirio Díaz (1854-1877)	1865	852	30637	12832	87	2441	66.51	
	1869	879	31791	12821	66	2478	3.06	
	1867	1176	33623	13013	90	2606	6.71	
	1866	1199	35662	13208	86	2716	5.77	
	1869	1185	34146	13406	86	2547	-4.64	
	1900	1319	36194	13607	86	2509	0.78	
	1901	1774	37371	13756	126	2718	8.59	
	1902	1672	34708	13804	120	2496	-7.13	
	1903	1859	38560	14055	132	2745	11.20	
	1904	1836	39271	14208	129	2764	1.29	
	1905	2273	43352	14360	158	3018	10.96	
	1906	2217	42962	14519	152	2862	-1.13	
1907	2346	45377	14678	156	3091	5.87		
1908	2408	45306	14836	162	3053	-0.15		
1909	2943	49639	14997	178	3199	2.84		
1910	3100	47054	15180	204	3103	0.89		
Alvaro Obregón (1920-1924)	1921	5455	50958	14335	360	3533	7.96	
	1922	5360	51830	14586	315	3558	2.33	
	1923	5014	50620	14861	338	3622	3.44	
Plutarco Elías Calles (1924-1928)	1924	4633	52753	15039	308	3507	-1.62	
	1925	5236	56024	15282	342	3696	6.20	
	1926	5469	56365	15526	352	3624	6.00	
	1927	4987	56773	15778	318	3399	-4.40	
Emilio Portes Gil (1928-1930)	1928	5018	57125	16032	312	3563	0.62	
	1929	4863	54915	16290	298	3301	-3.87	
Pascual Ortiz Rubio (1930-1932)	1930	4668	51473	16553	282	3109	-6.27	
	1931	4218	53179	16840	250	3137	3.31	
Abelardo L. Rodríguez (1932-1934)	1932	3206	45247	17132	187	2941	-4.82	
	1933	2782	50359	17429	216	2989	1.30	
Lázaro Cárdenas (1934-1940)	1934	4151	53357	17721	234	3021	6.73	
	1935	4540	57742	18038	251	3163	7.41	
	1936	5348	62361	18350	291	3396	8.00	
	1937	6000	64425	18668	364	3451	3.31	
	1938	7281	65469	18981	383	3447	1.82	
	1939	7785	68969	19309	402	3570	5.36	
	1940	8249	69441	19654	419	3556	1.26	
	Manuel Ávila Camacho (1940-1946)	1941	8232	76753	19783	437	3600	6.73
		1942	10681	81059	20751	514	3608	5.61
		1943	13025	84061	21323	611	3642	3.70
1944		16801	86623	21910	659	4149	6.18	
1945	20598	93779	22514	913	4185	3.14		
1946	27800	99642	23134	1207	4320	6.57		
Miguel Alemán Valdés (1946-1952)	1947	31023	103364	23771	1303	4346	3.44	
	1948	33101	107864	24438	1350	4436	4.12	
	1949	36412	113544	25006	1450	4598	5.48	
	1950	42183	124779	25791	1634	4636	6.89	
	1951	54375	134429	26585	2045	5000	7.73	
1952	63963	139775	27403	2225	5100	5.98		

Presidentes	Año	Producto (en millones de pesos)		Población (en miles de HAB/1991)	Producto per habitante (pesos)		Crecimiento (%)
		Comentas	De 1970		Comentas	De 1970	
Adolfo Ruiz Cortines (1952-1954)	1953	80064	140158	28240	2147	4802	0.27
	1954	73606	154188	28115	2538	5295	10.00
	1955	90053	187270	30011	3000	5573	8.50
	1956	102600	178706	30605	3329	5778	8.84
	1957	118206	182243	31867	3707	6028	7.58
Adolfo López Mateos (1958-1964)	1958	1331377	255467	32688	3987	8190	5.32
	1959	140772	208323	33880	4155	8154	0.89
	1960	156703	225448	34980	4284	8443	8.12
	1961	173208	236562	36188	4787	8537	4.93
	1962	186781	247815	37427	4880	8615	4.87
Gustavo Díaz Ordaz (1964-1970)	1963	207952	267398	38708	5372	9007	7.90
	1964	245501	298862	40033	6132	7480	11.80
	1965	267420	318000	41404	6458	7980	8.48
	1966	287198	340074	42821	6840	7841	8.00
	1967	325025	361387	44287	7338	8156	6.27
Luis Echeverría Álvarez (1970-1976)	1968	358858	380799	45863	7806	8531	8.14
	1969	387796	415512	47371	8308	8770	0.32
	1970	444271	444271	48225	9212	9212	0.02
	1971	480011	482804	52884	8265	8751	4.17
	1972	584727	520386	54661	10331	8185	8.48
José López Portillo (1976-1982)	1973	680681	544307	56481	12233	8638	8.41
	1974	889707	577568	58320	15427	9803	0.11
	1975	1100550	606978	60153	18287	10140	5.81
	1976	1370965	635831	61979	22118	10258	4.24
	1977	1848263	657722	63813	28878	10307	3.44
Miguel de la Madrid H. (1982-1988)	1978	2337388	711862	65638	35588	10843	8.25
	1979	3087328	777183	67516	45432	11510	0.15
	1980	4270480	841855	69655	61385	12088	8.32
	1981	5874386	868765	71205	82383	12743	7.65
	1982	814089	903839	72568	12008	12388	-0.54
Carlos Salinas de Gortari (1988-1994)	1983	17141864	887647	74833	22980	11472	-1.79
	1984	28748888	885928	78952	37887	11805	-0.19
	1985	47422549	848708	77628	48028	10884	-4.28
	1986	76334300	814702	79903	60738	10238	-0.86
	1987	129814485	828388	81183	2410143	10182	1.43
Carlos Salinas de Gortari (1988-1994)	1988	388258523	890005	81175	4795303.6	11826	16.17
	1989	503667765	960608	81187	6037198.2	12204	3.21
	1990	678823486	1034807	81211	8359994.2	12742	4.44
	1991	852783201	1072855	81223	10468425	13208	3.85
	1992	114821277.4	604712.2	81235	1413445.0	7444	-43.62
	1993	118018679.4	611430.42	81247	1433382.3	7326	1.11
	1994	118415481.3	618148.83	81259	1437372.2	7807	1.10
	1995	120214083.3	624888.88	81271	1437675.8	7888	1.08
	1996	122011685.2	631585.08	81283	1501072.8	7770	1.08
	1997	123808287.2	638333.27	81295	1522953.1	7852	1.08
1998	129006886.1	645021.43	81307	1548447.2	7833	1.05	
1999	127404481.1	651738.7	81319	1586724.8	8015	1.04	
2000	128202093.1	658457.81	81331	1588285.8	8088	1.03	

NOTA: Para 1957 se refiere al Ingreso Nacional. Originalmente los datos eran en dólares estadounidenses de 1954.

Los datos a partir de 1992 fueron ajustados con ayuda de migración y para la población se usó tasa de crecimiento geométrico (ver Apéndice).

FUENTE: INEGI "Estadísticas Históricas de México", Tomo I. Segunda Edición, Aguascalientes, Ato. INEGI Junio de 1994, p.p. 317,318.

## METODOLOGIA ESTADISTICA

### IMPORTANCIA DE LA PRONOSTICACION Y SUS METODOS.

Las naciones deben planear para el futuro, para poder sobrevivir y crecer. Aun las compras a plazos hechas por una familia suponen planes para el futuro. Planes de ventas, producción, inversión, distribución, etc., son hechas diariamente por empresas comerciales para hacer frente a la competencia corriente y mantener un crecimiento continuo. El gobierno federal debe planear ingresos y gastos futuros, no solo para realizar sus funciones rutinarias eficientemente, sino también para influir en la actividad comercial agregada, de modo que el progreso económico de la nación no pueda ser reducido por inflación o deflación.

Una acción comercial o económica emprendida hoy se basa en un plan de ayer y en las expectativas de mañana. Los planes para el futuro no pueden hacerse sin pronosticar hechos y las relaciones que tendrán. La pronosticación no solo puede hacerse para una línea determinada de actividad independientemente; el pronóstico de un tipo de hecho también puede hacerse sobre la base de otros pronósticos.

Los ejecutivos de empresas y los funcionarios de gobierno parece que han reconocido la importancia de la pronosticación como base de decisiones y acciones racionales referentes al futuro. Pero la pronosticación sigue siendo más un arte que una ciencia, ya que el valor de un pronóstico no es meramente su precisión, sino el hecho de que para hacerlo se requiere una consideración equilibrada de factores que influyen en acontecimientos futuros, acertada o errónea.

La dificultad básica -que el pasado nunca puede ser una perfecta guía para el futuro- nos advierte que la pronosticación no debe considerarse como una aplicación rutinaria de ciertas técnicas o ideas teóricas a una lista de variables no cambiantes. La buena pronosticación requiere una mezcla experta de la teoría económica, conocimientos estadísticos importantes y perfecta familiaridad con los datos estadísticos relevantes. Debe utilizar información cuantitativa y cualitativa. El pronosticador debe poseer la capacidad para distinguir entre nuevos hechos que son importantes y los que no lo son. Debe ser competente para juzgar en qué condiciones puede confiarse en las relaciones pasadas y en cuáles no. Debe ser capaz de apreciar los efectos de fuerzas socioeconómicas y políticas no mesurables en las actividades comerciales. En otras palabras, la pronosticación es, y probablemente seguirá siendo, más un arte que una ciencia.

Se han ideado durante las últimas décadas numerosas técnicas de pronosticación con distintos grados de complejidad. La mayor parte de ellas se agrupan en una de tres amplias categorías: el método ingenuo, el método barométrico y el método analítico.

Se dice que un método de pronosticación es ingenuo si carece de una rígida base teórica. Como método de pronosticación, varía desde un simple lanzamiento de una moneda para decidir sobre un movimiento hacia arriba o hacia abajo hasta la proyección mecánica de una serie de tiempo en el futuro.

Un procedimiento ingenuo frecuentemente usado es el método sencillo, pero no necesariamente inútil, de suponer que las cosas no cambiarán. En realidad, para muchas decisiones a corto plazo es exactamente lo que se hace. Por supuesto, este tipo de pronóstico resulta más dudoso a medida que se alarga el período de pronóstico. El método ingenuo, en general, y el análisis de series de tiempo, en particular, supone que el futuro es cierta clase de extensión del pasado.

El método barométrico implica que patrones históricos pasados tienden a repetirse en el futuro y acaricia la idea de que éstos pueden ser predichos partiendo de ciertos acontecimientos del presente. Así, el comportamiento estadístico pasado que parece asociarse regularmente con fluctuaciones en series particulares o condiciones comerciales generales es descubierto y usado como base para la pronosticación. Se buscan series presagiadoras para proporcionar una lectura anticipada de lo que se espera que siga en las series que han de ser pronosticadas. La búsqueda de series presagiadoras no se basa solo en los indicios realmente relevados por datos históricos. También se emplean consideraciones teóricas sobre los indicios y algunas en varias series. La comprobación cruzada entre empirismo y teoría en la búsqueda de series principales es la mejor esperanza para establecer la tesis de que el futuro no representa una ruptura con el pasado, sino que los cambios son determinados en gran parte por las condiciones presentes. El presagio o las series principales escogidas sirven de barómetros de los cambios futuros en series específicas o en las condiciones comerciales generales. Por tanto, el método barométrico supone el uso de uno o varios indicadores estadísticos que proporcionarían una indicación de la dirección que seguirán series específicas de tiempo o toda la economía.

El método analítico de pronosticación supone análisis detallados de fuerzas causales que operan corrientemente sobre la variable que ha de predecirse. Relaciones de causa y efecto reveladas por el análisis de datos corrientes se usan para juzgar el curso futuro de las fuerzas causales y sus efectos en el comportamiento futuro de la variable que se predice. Las técnicas analíticas pueden ser no matemáticas, tales como el método ingenuo de enumeración de factores o modelos econométricos. En cualquier caso, este método requiere cuidadosos análisis de las fuerzas que actúan, para establecer útiles relaciones y su cuantificación. El análisis de las fuerzas

corrientes actuantes no excluye en modo alguno el empleo de relaciones estadísticas pasadas para hacer el pronóstico. En realidad, a menudo se usan patrones históricos para estimar las componentes contenidas en los modelos matemáticos. Aunque el método analítico puede ser usado para pronosticar la actividad de una empresa o una industria, su mayor progreso en años recientes se ha debido a su uso en la pronosticación del marco de las cuentas del ingreso nacional.

Algunas de las técnicas de pronosticación más frecuentes son:

1. Métodos ingenuos:

- a) El método de ritmo económico, o método de pronosticación de una serie por sí misma.
- b) El método de suavización exponencial.

2. Métodos barométricos:

- a) Indicadores estadísticos;
- b) Índice de difusión.

3. Métodos analíticos:

- a) Análisis de regresión;
- b) Modelos econométricos.

El método que nos interesa en el presente estudio, es aquel que revele relaciones de causa y efecto, para juzgar el curso futuro de las fuerzas causales y sus efectos en el comportamiento futuro de la variable que se predice, empleando relaciones estadísticas pasadas para hacer el pronóstico. No estamos interesados en un método de pronosticación que carezca de una rígida base teórica, ni que sea útil para decisiones a corto plazo, ya que quisiéramos adelantarnos un poco en el tiempo para encontrar soluciones que puestas en práctica hoy lleguen a cambiar el futuro. Tampoco nos interesa saber la dirección que seguirán las series específicas de tiempo sin conocer

las causas y efectos, pues esto es en lo que realmente estamos interesados, ya que nos permitiría construir planes de acción. Esto nos lleva al uso de un método analítico.

Cada una de las técnicas de pronosticación tiene diferentes utilidades, utilidades que con el tiempo se han hecho más específicas con la ayuda de la experiencia en su aplicación, ejemplo de esto es el uso del método de ritmo económico para la programación de producción y presupuestos, o la suavización exponencial para la pronosticación de inventarios y ventas. Dentro de las técnicas de pronosticación más frecuentemente usadas que emplean el método analítico, es Análisis de Regresión con datos de series de tiempo la que llena nuestros requisitos y que trataremos en el apartado siguiente.



## ANÁLISIS DE REGRESIÓN CON DATOS DE SERIES DE TIEMPO.

Para pronosticar, el análisis de regresión suele aplicarse a datos de series de tiempo.

El principal problema al analizar datos bivariantes o multivariantes es descubrir y medir la asociación o covariación entre las variables, es decir, determinar cómo varían juntas las variables. A tal procedimiento se le ha llamado *estimación por asociación*.

Hay dos aspectos relacionados, pero distintos, del estudio de la asociación entre variables. El primero, *análisis de regresión*, trata de establecer la "naturaleza de la relación" entre variables, es decir, estudiar la relación funcional entre las variables y, por tanto, proporcionar un mecanismo de predicción, o pronóstico. El segundo, *análisis de correlación*, tiene como objetivo determinar el "grado de la relación" entre variables.

### MODELO DE REGRESIÓN BIVARIANTE LINEAL.

En el modelo de regresión bivariable lineal, una variable  $Y$  dependiente, o "explicada", se relaciona con una variable  $X$  independiente, o "explicativa", por la siguiente expresión:

$$y_i = \alpha + \beta x_i + u_i \quad (1)$$

donde  $\alpha$  y  $\beta$  son los parámetros de regresión desconocidos llamados *coeficientes de regresión de población*, y  $u_i$  es el "trastorno" al azar o *residual*.

Dos observaciones deben hacerse. Primera, designar las variables como dependientes o independientes aquí se refiere al significado matemático funcional de dependencia. Lo que se pretende al designar por  $Y$  la variable dependiente y por  $X$  la variable independiente es considerar  $Y$  como una función de  $X$ , en el sentido del álgebra ordinaria.

Segunda, la relación de dependencia lineal simple definida por (1) consta de dos partes: la parte sistemática identificada por  $\alpha + \beta x_i$  y la parte estocástica identificada por  $u_i$ . Esta descomposición nos recuerda que (1) es un modelo probabilista, en vez de determinista.

La naturaleza estocástica del modelo de regresión implica que el valor de  $Y$  nunca puede ser predicho exactamente como en el caso determinista. La incertidumbre relativa a  $Y$  es atribuible a la presencia de  $u_i$ , que siendo una variable aleatoria, imparte aleatoriedad a  $Y$ . Por ejemplo, no puede esperarse que robles de la misma edad ( $x_i$ ) tengan la misma altura ( $y_i$ ), debido a la influencia de fuerzas "casuales". Además de esta interpretación del término casual como una aleatoriedad inherente a la conducta tienen mérito otros dos puntos de vista. A veces, surge  $u_i$ , por la exclusión de otras variables explicativas importantes y relevantes en el modelo. Esto conduce al análisis de regresión múltiple. En ocasiones, el error de medición en  $Y$  es la causa de  $u_i$ . En una aplicación particular del análisis de regresión, cualquiera de estas razones podría ser una interpretación razonable de  $u_i$ , o cualquier par de estas razones, o las tres razones juntas.

Como una observación, podría preguntarse como se maneja el error de medición en  $X$ , ahora que el error de medición en  $Y$  ya se ha mencionado. La respuesta es que (1) no permite error de medición en  $X$ . Pero hay otros modelos que sí lo permiten. A pesar de esta limitación en (1), sigue siendo un modelo muy útil.

Cualquiera que sea la forma en que se interprete  $u_i$ , la completa

especificación del modelo de regresión incluye no solo la forma de la ecuación de regresión, sino también una expresión de cómo son determinados los valores de la variable independiente y una especificación de la distribución de  $u_i$  por probabilidades. La especificación completa de lo que se llama *modelo clásico de regresión lineal simple* la hace el siguiente conjunto de supuestos:

1. La variable independiente  $X$  es *fija*. El término "fijo" está en contraste directo con la noción de "estocástico". La expresión "valores fijos de  $X$ " significa que  $X$  tiene valores que son fijados (es decir, escogidos o predeterminados) por el investigador. El supuesto *independiente-variable-fijo* implica que para cada valor fijo de  $X$ ,  $\alpha_i$ , hay una distribución de valores  $Y$  por probabilidades, llamada *subpoblación* de  $Y$ .

2. El término "error"  $u_i$  asociado con cada valor de  $X$ ,  $\alpha_i$ , es una variable aleatoria cuya distribución de probabilidades se supone que es normal con  $E(u_i)=0$ . Este supuesto implica que, en promedio, la parte sistemática de  $\psi_i$  en (1) es  $\alpha + \beta\alpha_i$ . En realidad, la esperanza condicional de  $\psi_i$  dada  $\alpha_i$  es simplemente

$$\begin{aligned} E(\psi_i|\alpha_i) &= \mu_{yx} = E(\alpha + \beta\alpha_i + u_i) \\ &= \alpha + \beta E(\alpha_i) + 0 \\ &= \alpha + \beta\alpha_i \end{aligned} \tag{2}$$

El resultado nuevamente obtenido se llama *ecuación de regresión de población de  $Y$  sobre  $X$* , que nos da el valor medio de  $Y$  dado un valor fijo de  $X$ , y de ahí la notación  $\mu_{yx}$ . En esta expresión,  $\alpha$  es el valor medio de  $Y$  cuando  $X=0$ ;  $\beta$  mide el cambio en el valor medio de  $Y$  por cambio unitario en el valor de  $X$ . El gráfico de (2) se conoce por *línea de regresión de población*, donde,  $\alpha$  es la ordenada en el origen  $Y$  y  $\beta$  es la pendiente de la línea de regresión.

3. La varianza condicional de Y dada X se llama *varianza de la regresión*, representada por  $\sigma^2_{yx}$ . Se supone que esta medida es constante, cualquiera que sea el valor de X, y es igual a la varianza de  $u_i$ , es decir,  $\sigma_u^2$ . Esto es fácil de comprobar usando (1), y para cualquier valor de X, tenemos

$$\begin{aligned}
 V(\psi_i) &= E(\psi_i - E(\psi_i))^2 \\
 &= E(\alpha + \beta x_i + u_i - E(\alpha + \beta x_i + u_i))^2 \\
 &= E(\alpha + \beta x_i + u_i - \alpha - \beta x_i)^2 \\
 &= E(u_i^2) \\
 &= \sigma_u^2 = \sigma^2
 \end{aligned}
 \tag{3}$$

4.  $u_i$  es estadísticamente independiente de  $x_i$ , como podría esperarse, porque cada valor de  $u_i$  es una muestra al azar simple de tamaño uno y de una población normal con media cero y desviación estándar  $\sigma^2$ .

Con los supuestos anteriores, pueden derivarse estimadores para los parámetros de regresión desconocidos y pueden hacerse inferencias con estos estimadores.

## ESTIMACION DE PARAMETROS DE REGRESION.

Suponiendo que se ha extraído una muestra simple al azar de una población bivariable de interés, en la que el investigador ha determinado previamente cada valor de  $X$ , dando observaciones  $(x_i, y_i)$  para  $i=1, 2, \dots, n$ ; entonces, el modelo de regresión lineal de la muestra puede darse como

$$y_i = a + bx_i + e_i \quad (4)$$

y la ecuación de regresión de la muestra de  $Y$  sobre  $X$  como

$$\hat{y}_i = a + bx_i \quad (5)$$

En estas dos ecuaciones,  $y_i$  y  $\hat{y}_i$  designa el valor individual de  $Y$  y el valor calculado de  $Y$ , respectivamente. De (4) tenemos

$$e_i = y_i - a - bx_i,$$

donde  $e_i$  son los "residuales" o estimadores calculados de  $u_i$ , con  $\sum_1^n e_i = 0$ .

En (5),  $a$  es la ordenada en el origen  $Y$  de la muestra, el valor medio de  $Y$  dado que  $X = 0$ , y  $b$  es la pendiente de regresión de la muestra, el cambio en el valor medio de  $Y$  por cambio unitario en el valor de  $X$ . Cabe decir que  $a$  y  $b$  son estimadores de  $\alpha$  y  $\beta$ , respectivamente.

Debido al supuesto de linealidad del modelo,  $(a + bx_i)$  se llama *estimador lineal* de  $(\alpha + \beta x_i)$ . El estimador lineal que buscamos aquí es el MEIL -el mejor estimador insesgado lineal-. El método usado para obtener MEIL se llama método de mínimos cuadrados. Hay amplia motivación y justificación para el uso de este método. En primer lugar, cuando las  $Y$  son variables aleatorias con varianzas iguales y cuando las  $X$  son fijas, entonces, por el Teorema Gauss-Markov, el MEIL de  $\alpha$  y  $\beta$  puede obtenerse por el método de mínimos

cuadrados. Segundo, cuando las  $Y$  están normalmente distribuidas, los estimadores obtenidos por el método de mínimos cuadrados serán idénticos a los obtenidos por el método de probabilidad máxima. Finalmente, el método de mínimos cuadrados tiene la naturaleza de ajuste de curvas.

Dada la dependencia lineal entre  $Y$  y  $X$  y los  $n$  pares de observaciones  $(\psi_i, \alpha_i)$ , este método produce estimadores paramétricos  $a$  y  $b$  en tal forma que

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (\psi_i - \hat{\psi}_i)^2 = \sum_{i=1}^n (\psi_i - (a + b\alpha_i))^2 = \text{RSS}$$

es un mínimo. (Obsérvese que  $\hat{\psi}_i = a + b\alpha_i$ ). La razón del criterio de mínimos cuadrados es que escoge valores para  $a$  y  $b$  que minimizan la suma de cuadrados de las diferencias entre los valores realmente observados  $\psi_i$  y los valores estimados  $\hat{\psi}_i$ . Los estimadores de mínimos cuadrados, EMC, de  $a$  y  $b$  pueden obtenerse resolviendo el siguiente conjunto de "ecuaciones normales":

$$(a) \sum_i \psi_i = na + b \sum_i \alpha_i$$

$$(b) \sum_i \psi_i \alpha_i = a \sum_i \alpha_i + b \sum_i \alpha_i^2$$

dichas ecuaciones normales son el resultado de obtener las derivadas parciales de RSS con respecto a  $a$  y  $b$  e igualarlas a cero.

Ahora, si dividimos (a) por  $n$  y reagrupamos términos, obtenemos

$$a = \bar{\psi} - b\bar{\alpha}. \quad (6)$$

A continuación, si multiplicamos (a) por  $\sum \psi_i$  y (b) por  $n$ , restando luego el primer resultado del segundo, tenemos

$$b = \frac{n \sum \psi_i a_i - \sum \psi_i \sum a_i}{n \sum a_i^2 - (\sum a_i)^2} \quad (7a)$$

Por razones de comodidad en el cálculo y la notación, varias medidas en análisis de regresión pueden calcularse sobre las desviaciones de las medias de las muestras en vez de datos brutos. En términos de desviaciones de las medias de las muestras, las sumas de cuadrados y los productos cruzados pueden definirse como sigue:

$$m_{11} = \sum (\psi_i - \bar{\psi})^2 = \sum \psi_i^2 - n(\bar{\psi})^2 \quad (8)$$

$$m_{22} = \sum (a_i - \bar{a})^2 = \sum a_i^2 - n(\bar{a})^2 \quad (9)$$

$$m_{12} = \sum (\psi_i - \bar{\psi})(a_i - \bar{a}) = \sum \psi_i a_i - n(\bar{\psi})(\bar{a}) \quad (10)$$

Con estas definiciones, por ejemplo (7a), puede escribirse compactamente

$$b = \frac{m_{12}}{m_{22}} \quad (7b)$$

## VARIANZA DE LA REGRESION EN LA MUESTRA

La ecuación de regresión de la muestra se llama *ecuación predictiva* porque su función principal es predecir  $\mu_{xy}$  o  $\psi_i$  asociadas con un valor dado de  $X$ . Pero, ¿cuán conveniente es como un medio de predicción? La respuesta evidentemente reside en parte en la variabilidad de  $\psi_i$  obtenida de  $\hat{\psi}_i$  asociada con  $\alpha_i$ . Una forma de comprender tal dispersión es trazar la línea de regresión en el diagrama de dispersión y unir esta línea con los puntos por líneas verticales. Para trazar la línea de regresión, necesitamos representar gráficamente solo dos valores  $\hat{\psi}_i$  con, quizá, un tercero como comprobación. La medida numérica de las desviaciones es el estimador insesgado de la varianza de la regresión de población, o el término transtorno, que se define como

$$\begin{aligned} \sigma^2 &= \frac{\sum (\psi_i - \hat{\psi}_i)^2}{n - 2} \\ &= \frac{\sum (\psi_i - a - b\alpha_i)^2}{n - 2} \\ &= \frac{\sum \psi_i^2 - a \sum \psi_i - b \sum \psi_i \alpha_i}{n - 2} \\ &= \frac{m_{11} - \left( \frac{m_{12}^2}{m_{22}} \right)}{n - 2} \end{aligned} \quad (11)$$

Aquí  $(n - 2)$  grados de libertad son usados como denominador. La pérdida de dos grados de libertad corresponde al número de coeficientes de regresión. La raíz cuadrada del estimador  $\hat{\sigma}^2$  es la *desviación estandar de la regresión para*



la muestra, que a veces se conoce erróneamente como *error estándar de estimación*.

Se puede observar que  $\hat{\sigma}^2$  o  $\hat{\sigma}$  es una medida de grado de ajuste -el grado de la relación lineal entre Y y X-. Cuando menor es una u otra, tanto más precisa es la ecuación de regresión de la muestra como un instrumento de predicción. Debe tenerse en cuenta que muestras mayores ayudarán a localizar la línea de regresión más precisamente, pero no reducirán mucho la desviación estándar de la regresión.

#### INFERENCIAS ACERCA DE LOS COEFICIENTES DE REGRESION DE LA POBLACION.

Una vez obtenida la ecuación de regresión de la muestra y deducido que la ecuación de regresión puede ser útil sobre la base de la desviación estándar de la regresión de la muestra, se puede llegar a la conclusión de que ahora la ecuación de regresión de la muestra puede ser usada fácilmente como un instrumento de predicción. Pero en realidad, aun si la ecuación de regresión de la muestra es idéntica a la ecuación de regresión de la población, la predicción todavía puede contener cierto error porque la relación entre Y y X en la población bivariable puede no ser perfecta. La magnitud de este error se mide por  $\sigma^2$ . Además, al predecir con la ecuación de regresión de la muestra hay otro error debido a las variaciones casuales en el muestreo. Es decir, para muestreo repetido con el mismo tamaño de la muestra, el estimador lineal de  $(\alpha + \beta x)$  tendería a variar de una muestra a otra. Los errores de muestreo de a y b son evaluados en términos de sus respectivas distribuciones por muestreo.

Las distribuciones de a y b por muestreo, que son EMC, son muy fáciles de

deducir. Primero, puesto que  $a$  y  $b$  son insesgadas, debemos tener  $E(a)=\alpha$  y  $E(b)=\beta$ . Segundo, puesto que  $a$  y  $b$  son combinaciones lineales de variables normales independientes,  $Y$ , deben estar normalmente distribuidas. Finalmente, debido a las propiedades MEIL, los estimadores insesgados para las varianzas de  $a$  y  $b$  deben ser definidos como sigue:

$$\hat{\sigma}_a^2 = \frac{\sigma^2 \sum \bar{x}_i^2}{n \sum (\bar{x}_i - \bar{x})^2} = \hat{\sigma}^2 \left( \frac{\sum \bar{x}_i^2}{nm22} \right); \quad (12)$$

$$\hat{\sigma}_b^2 = \frac{\sigma^2}{\sum (\bar{x}_i - \bar{x})^2} = \hat{\sigma}^2 \left( \frac{1}{m22} \right) \quad (13)$$

Las estadísticas de prueba para las hipótesis acerca de  $\alpha$  y  $\beta$  son las razones  $z$  ordinarias:

$$z = \frac{a - \alpha_0}{\hat{\sigma}_a} \quad (14a)$$

y

$$z = \frac{b - \beta_0}{\hat{\sigma}_b} \quad (15a)$$

Cuando  $n \leq 30$  y cuando  $\sigma$  es desconocida,  $a$  y  $b$  están distribuidas aproximadamente como  $t$  de Student con  $\delta = n - 2$ . La estadística de prueba apropiada es ahora

$$t(\delta) = \frac{a - \alpha_0}{\hat{\sigma}_a} \quad (14b)$$

y

$$t(\delta) = \frac{b - \beta_0}{\hat{\sigma}_b} \quad (15b)$$

respectivamente.

En general, la ecuación de regresión de la muestra debe ser conservada como un instrumento de predicción si se encuentra que  $b$  es significativa; en otro caso, debe ser desechada. Los intervalos de confianza para coeficientes de regresión pueden ser calculados en la forma usual.

## PREDICCIÓN Y PRONOSTICACION

Se ha dicho repetidamente que el objeto final del análisis de regresión es la estimación por asociación. A este fin, se distinguen dos casos: el del valor medio de  $Y$  y el de una observación individual de  $Y$ . Llamaremos al caso de estimar la media de  $Y$ ,  $\mu_{yx}$ , *predicción*, y designaremos por  $\hat{\bar{y}}_0$  al estimador de  $\mu_{yx}$  asociado con una  $x_0$  específica. Llamaremos también arbitrariamente *predictor* a  $\hat{\bar{y}}_0$  y escribiremos

$$\hat{\bar{y}}_0 = a + b x_0 \quad (16a)$$

Llamaremos *pronosticación* a la proyección de un solo valor de  $Y$ ,  $y_0$ , correspondiente a  $x_0$ . Al estimador se le llama ahora *pronosticador*, que será designado por  $\hat{y}_0$ , y puede expresarse como

$$\hat{y}_0 = a + b x_0 \quad (16b)$$

Es importante hacer constar que las fórmulas empleadas para "predicción" y "pronosticación" son idénticas -a la ecuación de regresión de la muestra establecida-; pero los valores que se proyectan son muy diferentes teóricamente.

## Predicción.

"Predicción", tal como se usa el término, al ser un procedimiento para estimar  $\mu_{yx}$ , se relaciona con la precisión de toda la línea de regresión de la muestra. Puesto que sabemos que  $(a + b\alpha_0)$  es un estimador insesgado de  $(\alpha + \beta\alpha_0)$  sabemos que  $E(\hat{y}_0) = \mu_{yx}$ . Además  $(a + b\alpha_0)$ , siendo una combinación lineal de variables aleatorias distribuidas normalmente, debe estar también normalmente distribuida. Finalmente, para determinar el intervalo de confianza para cualquier punto de la línea de regresión de la población, debemos tener el error estándar del predictor. Para obtener esta medida, observamos que el error al determinar el valor de Y condicional a  $\alpha_0$  surge de dos fuentes: variaciones casuales en a y variaciones causales en b. Así dado que  $\alpha$  y  $\beta$  son los parámetros verdaderos, el error en la predicción representado por  $e_p$ , es claramente

$$\begin{aligned}e_p &= \alpha + \beta\alpha_0 - (a + b\alpha_0) \\ &= (\alpha - a) + (\beta - b)\alpha_0.\end{aligned}$$

Elevando al cuadrado y tomando la esperanza de esta expresión, obtenemos la varianza del predictor, representada por  $\sigma_p^2$ , como sigue:

$$\sigma_p^2 = \sigma^2 \left[ \frac{1}{n} + \frac{(\alpha_0 - \bar{\alpha})^2}{\sum_1 (\alpha_1 - \bar{\alpha})^2} \right]$$

La raíz cuadrada de este resultado es el error estándar de  $\hat{y}_0$ . Usando el estimador insesgado para  $\sigma^2$ , obtenemos el estimador (ligeramente sesgado) del error estándar de  $\hat{y}_0$  como

$$\hat{\sigma}_p = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(\alpha_0 - \bar{\alpha})^2}{\sum_1 (\alpha_1 - \bar{\alpha})^2}}$$

$$= \frac{\sigma}{\sqrt{n}} \sqrt{\frac{(\alpha_0 - \bar{\alpha})^2}{m_{22}}} \quad (17)$$

En conclusión, vemos que la razón

$$\frac{\hat{y}_0 - (\alpha + \beta\alpha_0)}{\sigma_p}$$

es aproximadamente  $n(0,1)$  y  $n > 30$  o si  $\sigma$  es desconocida. En otro caso, tiene una distribución  $t$  con  $\delta = n - 2$ . Esto conduce a la construcción de un intervalo de confianza para  $\mu_{y\alpha_0}$  sobre la base del predictor como sigue:

$$\hat{y}_0 \pm t(\delta; 1-\alpha/2) \hat{\sigma}_p \quad (18)$$

Las estimaciones de intervalos para  $\mu_{yx}$  construídas con varios valores de  $\alpha_0$  en un nivel de confianza dado forman lo que se llama la *banda de confianza* para  $\mu_{yx}$ .

#### Pronosticacion.

Muy a menudo podemos interesarnos por estimar un valor individual de  $Y$ ,  $y_0$  asociado con  $\alpha_0$ . Precisamente, dado  $X = \alpha_0$ , nuestra tarea es ahora determinar el valor del pronosticador,  $\hat{y}_0$ , que es una variable aleatoria con sus valores dispersos alrededor del punto de la línea de regresión correspondiente a  $\alpha_0$ . Como tal, nunca podemos conocer su valor antes del experimento. Aquí, la diferencia entre el valor real  $y_0$  y el valor pronosticado  $\hat{y}_0$  se llama *error de pronosticación*, que es una combinación lineal de variables aleatorias normalmente distribuidas, los trastornos residuales. Observemos que

$$y_0 - \hat{y}_0 = (\alpha + \beta\alpha_0 + u_0) - (a + b\alpha_0).$$

Esto indica que el error de pronosticación es también una variable normalmente distribuida cuya distribución se define completamente por su esperanza y su varianza. La esperanza del error de pronosticación puede determinarse fácilmente de su definición anterior:

$$\begin{aligned} E(\hat{y}_0 - y_0) &= E(\alpha + \beta x_0 + u_0 - a - bx_0) \\ &= \alpha + \beta x_0 + E(u_0) - E(a) - E(b)x_0 \\ &= 0 \end{aligned}$$

La varianza del error de pronosticación, representada por  $\sigma_F^2$ , consta de dos partes: 1) el error de predicción, o variación debido a muestreo,  $\sigma_p^2$ , y 2) variación debida a trastornos causales medida por  $\sigma^2$ . Así,

$$\begin{aligned} \sigma_F^2 &= V(e_p) + V(u) = \sigma_p^2 + \sigma^2 \\ &= \sigma^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_1 (x_1 - \bar{x})^2} \right] \end{aligned}$$

Como es usual, la raíz cuadrada de esta medida es el error estándar para el error de pronosticación. Ahora, usando los estimadores insesgados de  $\sigma^2$  y  $\sigma_p^2$ , tenemos el estimador (ligeramente sesgado) del error estándar para el error de pronosticación como

$$\begin{aligned} \hat{\sigma}_F &= \sqrt{\hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_1 (x_1 - \bar{x})^2} \right]} \\ &= \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{m_{22}}} \end{aligned} \quad (19)$$

Como para el caso del predictor, el error de pronosticación está normalmente distribuido y la razón

$$(\hat{y}_0 - y_0) / \hat{\sigma}_F$$

es aproximadamente  $n(0,1)$  cuando  $n > 30$  o si  $\sigma$  es desconocida. Cuando  $n \leq 30$  y  $\sigma$  es desconocido, está distribuido como  $t_{n-2}$ . Este resultado permite escribir

$$P \left\{ -t(\delta; 1-\alpha/2) \leq \frac{\hat{y}_0 - y_0}{\hat{\sigma}_F} \leq +t(\delta; 1-\alpha/2) \right\} = 1 - \alpha$$

que reagrupando términos, resulta

$$\hat{y}_0 - t(\delta; 1-\alpha/2) \hat{\sigma}_F \leq y_0 \leq \hat{y}_0 + t(\delta; 1-\alpha/2) \hat{\sigma}_F$$

Los límites de confianza para  $y_0$  -el valor pronosticado de  $Y$ - son

$$\hat{y}_0 \pm t(\delta; 1-2\alpha/2) \hat{\sigma}_F \quad (20)$$

Es interesante observar que las bandas de confianza para  $\mu_{xy}$  y  $y_0$  siguen la misma forma general que con  $\hat{\sigma}_p^2$  y  $\hat{\sigma}_F^2$  (y, por tanto, los correspondientes límites de confianza de  $\mu_{xy}$  y  $y_0$ ) aumentando cuando  $\alpha_0$  se aleja de la media de  $X$  para la muestra. Es decir, cuando  $\alpha_0$  se aleja de  $\bar{x}$ , nuestra estimación resulta más incierta.

Debe tenerse en cuenta que, en la estimación con el análisis de regresión clásico, generalmente no debe fijarse el valor de  $X$  fuera de la escala de valores  $X$  empleada para establecer la ecuación de regresión de la muestra. Una razón para ello es que la relación entre las dos variables abajo o arriba de la escala de valores  $X$  de la muestra puede que no sea la misma que la

observada dentro de la escala. Otra razón reside en el hecho de que la predicción y la pronosticación se vuelven muy imprecisas cuando  $x_0$  se aleja demasiado de  $\bar{x}$  en virtud de los grandes errores de predicción y pronosticación. Pero debe hacerse constar que en los estudios de economía o pronosticación de negocios, especialmente cuando se emplean datos de series de tiempo en el análisis de regresión, más a menudo que no, nos interesamos por los valores de variables del modelo fuera de la amplitud de la muestra. Esta práctica se justifica muchas veces en el supuesto de que la relación establecida, a juicio del usuario, se producirá más allá de la amplitud de los valores de la muestra.



## ANALISIS DE CORRELACION.

### Coefficiente de correlacion de la poblacion

Hasta ahora, hemos tratado el tipo y el grado de relación que existe entre dos variables. Pero, en algunas situaciones, podemos interesarnos solo por el grado de relación entre dos variables.

La medida del grado de relación entre dos variables se llama *coeficiente de correlación*, representado universalmente por  $\rho$ . Los supuestos que constituyen un modelo de población por correlación lineal bivariable, para el cual se calcula o estima  $\rho$ , se describen brevemente a continuación:

1. Y y X son variables aleatorias. Como tales, no deben ser designadas como dependiente e independiente; cualquier designación dará el mismo resultado.

2. La población bivariable es normal. Una población normal bivariable es, entre otras cosas aquella en la que Y y X están normalmente distribuidas. La esperanza y la varianza de X son  $E(X) = \mu_x$  y  $V(X) = \sigma_x^2$ , respectivamente, y la esperanza y la varianza de Y son  $E(Y) = \mu_y$  y  $V(Y) = \sigma_y^2$ , respectivamente.

3. La relación entre Y y X es, en cierto sentido, lineal. Este supuesto implica que todas las medias de Y asociadas con valores X,  $\mu_{y|x}$ , caen sobre una línea recta, que es la línea de regresión de Y sobre X:  $\mu_{y|x} = \alpha + \beta x_i$ . Igualmente todas las medias de X asociadas con valores de Y,  $\mu_{x|y}$ , caen sobre una línea recta, que es la línea de regresión de X sobre Y:  $\mu_{x|y} = \alpha' + \beta' y_i$ . Además, las líneas de regresión de población en las dos ecuaciones son iguales si, y solo si, la relación entre Y y X es perfecta -es decir,  $\rho = \pm 1$ . En otro caso, con Y dependiente, las ordenadas en el origen y las pendientes diferirán de la ecuación de regresión con X dependiente.

Según los supuestos anteriores, el coeficiente de correlación de la población se define como la covarianza entre Y y X, así:

$$\rho = \frac{\text{cov}(Y, X)}{\sigma_y \sigma_x} = \frac{E\{(Y - \mu_y)(X - \mu_x)\}}{\sqrt{E\{Y - \mu_y\}^2} \sqrt{E\{X - \mu_x\}^2}} \quad (21)$$

Pueden observarse algunas cosas acerca de la definición anterior. Primera, es una ecuación que contiene los cinco parámetros de la población normal bivariable:  $\mu_x$ ,  $\sigma_x$ ,  $\mu_y$ ,  $\sigma_y$  y  $\rho$ . El último, como se mencionó antes, es el coeficiente de correlación para la población normal bivariable. Segundo,  $\rho$  es simétrico con respecto a Y y X; es decir, intercambiar X y Y no cambia  $\rho$ . En otras palabras,  $\rho$  es un número puro, porque se define como la razón de la covarianza entre Y y X al producto de sus respectivas desviaciones estándares. Como tal, cuando  $\text{cov}(Y, X) = 0$ ,  $\rho$  sería 0, indicando que no hay relación entre las dos variables. Cuando hay covariabilidad perfecta entre Y y X, y Y y X varían en la misma dirección,  $\rho = 1$ . Análogamente, cuando hay covariabilidad perfecta, pero Y y X varían en sentidos opuestos,  $\rho = -1$ . Cuando existe cierto grado de covariabilidad entre Y y X, tenemos  $-1 < \rho < 0$  ó  $0 < \rho < 1$ .

#### Coefficiente de correlación de la muestra.

Cuando se extrae una muestra de n pares de valores, cada valor Y es una observación al azar de la población Y, y cada valor de X es una observación al azar de la población X; pero las dos no son necesariamente independientes. Además, cuando se cumple el supuesto de una población normal bivariable, el estimador de probabilidad máxima de  $\rho$ , representado generalmente por r, se obtiene por la siguiente expresión

$$r = \hat{\rho} = \frac{\sum_1 (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\left[ \sum_1 (y_i - \bar{y})^2 \right] \left[ \sum_1 (x_i - \bar{x})^2 \right]}}$$

$$= \frac{m_{12}}{\sqrt{m_{11}m_{22}}}$$
(22)

que puede variar, como  $\rho$ , de  $-1$  a  $0$  a  $+1$ . También como para  $\rho$ , cuando  $r = -1$  o  $+1$ , se dice que  $Y$  y  $X$  están perfectamente correlacionadas negativa o positivamente. Cuando  $r = 0$ , se dice que las dos variables son independientes o no correlacionadas; hablando en sentido riguroso,  $X$  y  $Y$  son independientes sí, y solo sí,  $\rho = 0$ , en vez de  $r = 0$ , y es bueno tener presente esto. También, sí a  $X$  y  $Y$  se les permite tener cualquier distribución conjunta, su independencia implica que su varianza es cero, pero lo inverso no es cierto, porque las variables dependientes pueden tener covarianza cero. Sin embargo, sí  $X$  y  $Y$  se restringen a la distribución normal bivariable, entonces su independencia significa que su covarianza es cero, y por consiguiente, su coeficiente de correlación (población) es cero; y lo inverso es cierto, porque las variables normales bivariadas dependientes no pueden tener covarianza cero. Otros valores de  $r$  son interpretados en función del coeficiente de determinación que se tratará después.

Puede señalarse que  $r$  es un estimador sesgado de  $\rho$ , excepto que si  $\rho = 0$ , entonces  $r$  es insesgado, pero se usa generalmente debido a su relativa sencillez y porque el sesgo es despreciable para muestras grandes.

Para hacer inferencias de  $r$ , debemos empezar con algunas observaciones acerca de su distribución de muestreo. La distribución de  $r$  es simétrica cuando  $\rho = 0$  y es asimétrica cuando  $\rho \neq 0$ . Para una población bivariable normal, la distribución de  $r$  se aproxima a una distribución normal cuando  $n$  se

aproxima al infinito. Cuando  $\rho = 0$ , hay una transformación por la cual los valores transformados de  $r$  son distribuidos como  $t_{n-2}$ :

$$t(\delta) = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (23)$$

Puesto que esta transformación solo se aplica en el caso de que  $\rho = 0$ , no puede ser empleada para estimación del intervalo. Pero puede ser usada como la estadística de prueba para la hipótesis nula de que  $\rho = 0$  contra una alternativa apropiada.

Hay que observar que las hipótesis nulas  $\rho = 0$  y  $\beta = 0$  son equivalentes, porque  $\beta = \rho(\sigma_y/\sigma_x)$ . Como resultado, las pruebas  $t$  en ambos casos deben dar el mismo valor.

Con la prueba  $t$  recién introducida, no podemos enunciar explícitamente la relación entre las dos variables cuando la hipótesis nula es rechazada. Si deseamos probar una hipótesis de que  $\rho$  tiene cierto valor distinto de cero, o si queremos construir un intervalo de confianza para  $\rho$ , podemos emplear lo que se llama transformación  $z$ . Es decir, podemos hacer una transformación de la distribución de  $r$  en una distribución aproximadamente normal como sigue:

$$z_r = \frac{1}{2} \ln \frac{1+r}{1-r} \quad (24)$$

donde  $z_r$  está aproximada y normalmente distribuida con  $E(z_r) = z_p$ , y el error estándar estimado es

$$\hat{\sigma}_z = \frac{1}{\sqrt{n-3}} \quad (25)$$

Para probar la hipótesis acerca de  $\rho$  por  $r$ , tenemos ahora la estadística de prueba

$$z = \frac{z_r - z_p}{\hat{\sigma}_z} \quad (26)$$

que es aproximadamente  $n(0,1)$ .

Los límites de confianza pueden calcularse para  $z_p$ , como sigue:

$$P(z_r - z(1-\alpha/2\sigma_z) < z_p < z_r + z(1-\alpha/2\sigma_z)) = 1 - \alpha \quad (27)$$

Puede señalarse, antes de concluir esta sección, que la significación de una diferencia entre dos coeficientes de correlación de muestra obtenidos de dos muestras independientes puede probarse por la estadística

$$z = \frac{z_{r1} - z_{r2}}{\sqrt{\sigma_{z1}^2 + \sigma_{z2}^2}} \quad (28)$$

donde

$$\sigma_{z1}^2 = \frac{1}{n1-3} \quad \text{y} \quad \sigma_{z2}^2 = \frac{2}{n2-3}$$

#### COEFICIENTE DE DETERMINACION Y ANALISIS DE VARIANZA EN REGRESION LINEAL

Se ha mencionado antes que  $\sigma^2$ , como una medida de aproximación del ajuste, no tiene un significado obvio y exacto, y que los valores de  $r$ , distintos de  $-1$ ,  $+1$  y  $0$ , no pueden ser interpretados fácilmente. Parece conveniente entonces tener una medida estándar, y libre de unidad, de aproximación del ajuste, y una medida por medio de la cual puedan interpretarse fácilmente los valores de  $r$ . La medida que puede servir para ambos fines es una estadística llamada *coeficiente de determinación*. Esta medida puede derivarse por el método de análisis de varianza: descomponiendo la varianza de  $Y$  en porciones significativas y distintivas, como sigue

$$(\psi_i - \bar{y}) = (\hat{\psi}_i - \bar{y}) + (\psi_i - \hat{\psi}_i)$$

distancia total a $\bar{y}$ ; "error total"	distancia de la línea de regresión a $y_i$ ; "error explicado"	distancia de una observación individual a la regresión; "error residual"
---	--	--

Esta identidad se aplica a una sola observación. Pero necesitamos una medida sumaria para todas las observaciones de muestra. Por tanto elevamos al cuadrado ambos miembros de esta identidad, sumamos todas las observaciones de la muestra y obtenemos

$$\begin{aligned}
 \sum_i (\psi_i - \bar{y})^2 &= \sum_i \left[ (\hat{\psi}_i - \bar{y}) + (\psi_i - \hat{\psi}_i) \right]^2 \\
 &= \sum_i \left[ (\hat{\psi}_i - \bar{y}) + e_i \right]^2 \\
 &= \sum_i (\hat{\psi}_i - \bar{y})^2 + 2 \sum_i (\hat{\psi}_i - \bar{y}) e_i + \sum_i e_i^2
 \end{aligned}$$

Ahora, puede evaluarse el segundo término del segundo miembro de este resultado como sigue:

$$\begin{aligned}
 2 \sum_i (\hat{\psi}_i - \bar{y}) e_i &= 2 \sum_i (a + b x_i - \bar{y}) e_i \\
 &= 2 \left( a \sum_i e_i + b \sum_i x_i e_i + \bar{y} \sum_i e_i \right)
 \end{aligned}$$

Pero, puesto que  $\sum_i e_i = 0$ , y como  $x_i$  es estadísticamente independiente de  $e_i$ , deducimos que

$$2 \sum_i (\hat{\psi}_i - \bar{y}) e_i = 0.$$

Por tanto tenemos

$$\sum_i (\psi_i - \bar{y})^2 = \sum_i (\hat{\psi}_i - \bar{y})^2 + \sum_i (\psi_i - \hat{\psi}_i)^2$$

suma de cuadrados total (SCT)	suma de cuadrados debida a regresión (SCR)	suma de cuadrados por error (SCE)
-------------------------------	--	-----------------------------------

La expresión anterior indica que la variabilidad total en  $Y$ , medida por la *suma de cuadrados total*,

$$SCT = \sum_i (y_i - \bar{y})^2 = m11 \quad (29)$$

puede descomponerse en dos partes. Una describe los valores ajustados de  $Y$ , que representan aquella porción de  $SCT$  que ha sido explicada por regresión de  $Y$  sobre  $X$ . A esta parte se le llama *suma de cuadrados debida a regresión*  $SCR$ . la otra parte describe la variación de los residuales, llamada *suma de cuadrados por error*,  $SCE$ , que representa aquella parte de  $SCT$  que no ha sido explicada por la ayuda de  $X$ . Alternativamente, se dice que  $SCR$  representa el efecto estimado de  $X$  en la variación en  $Y$  y  $SCE$  representa el efecto estimado del trastorno al azar.

Además,  $SCR$  puede desarrollarse como sigue:

$$\begin{aligned} SCR &= \sum_i (\hat{y}_i - \bar{y})^2 \\ &= \sum_i (a + bx_i - \bar{y})^2 \\ &= \sum_i \left[ (\hat{y}_i - b\bar{x}) + bx_i - \bar{y} \right]^2 \\ &= \sum_i \left[ b(x_i - \bar{x}) \right]^2 \\ &= b^2 m22 \end{aligned} \quad (30)$$

Obviamente

$$SCE = \sum_i (y_i - \hat{y}_i)^2 = SCT - SCR \quad (31)$$

Ahora, si dividimos cada término de la identidad

$$SCT = SCR + SCE$$

por  $SCT$ , obtenemos

$$1 = \frac{SCR}{SCT} + \frac{SCE}{SCT}$$

De esto se define el *coeficiente de determinación de la muestra*, representado por  $r^2$ , como sigue:

$$\begin{aligned} r^2 &= 1 - \frac{SCE}{SCT} \\ &= \frac{SCR}{SCT} && (32) \\ &= \frac{\text{suma de cuadrados explicada}}{\text{suma de cuadrados total}} \\ &= \frac{\text{error explicado}}{\text{error total}} \end{aligned}$$

De la definición de  $r^2$  vemos que el valor de  $r^2$  varía de 0 a 1 porque  $SCR \leq SCT$ . Si  $SCE = 0$ , entonces  $SCR = SCT$ ; por tanto,  $r^2 = 1$ . Obsérvese que cuando  $r^2=1$ , o  $\hat{y}_i = \hat{y}_i$ , o  $e_i = 0$ , lo que indica que todos los puntos del diagrama de dispersión se encuentran en la línea de regresión de la muestra. Si  $SCR = 0$ , entonces  $SCE = SCT$ ; por tanto  $r^2 = 0$ . Una condición necesaria para que  $r^2 = 0$  es que la pendiente de regresión  $b$  sea cero. Pero esto puede suceder por diferentes razones. La línea de regresión de la muestra podría ser horizontal:

a) si las observaciones se dispersan alrededor de  $\bar{y}$  en una forma al azar, b) si las observaciones se dispersan alrededor de una curva tal que la línea mejor ajustada es una línea recta horizontal y c) si todas las observaciones tienen el mismo valor, cualquiera que sea el valor de  $X$ . El caso (a) es obvio. El caso (b) refleja un ajuste muy malo debido a la alta relación no lineal entre las dos variables. En el caso (c), con todos los valores de  $Y$  iguales, o con  $\hat{y}_i = \bar{y}$ , no hay variación en  $Y$  que deba explicarse; por consiguiente, la cuestión de descomposición de la variación total,  $SCT$ , es irrelevante. Alternativamente, para este caso, la dificultad es que las subpoblaciones de  $Y$



no tienen ninguna variación, porque  $SCT = 0$ . Puesto que  $SCT = 0$ ,  $SCR = 0$  y  $SCE = 0$ . Como consecuencia, el numerador y el denominador de (32) son cero, y por tanto  $r^2$  es indeterminada.

Puede observarse ahora que  $r^2$  puede ser interpretada desde tres puntos de vista. Primero, puede considerarse como una medida del grado de mejora debido a la línea de regresión. Segundo, puede ser interpretada como una medida del grado de ajuste de la línea de regresión a los puntos del diagrama más grueso. Finalmente, puede considerarse como el grado de linealidad de la dispersión de los puntos.

Es importante observar que tres de las diferentes interpretaciones de  $r^2$  son simplemente diferentes aspectos del mismo resultado. Además,  $r^2$  como una medida que muestra la mejora en términos de reducir el error total no debe considerarse como una medida de covariabilidad de las dos variables si la variable independiente  $X$  asume valores fijos. Pero, si  $X$  y  $Y$  son variables aleatorias,  $r^2$  también puede ser interpretada como una medida del grado de covariabilidad. Finalmente, puede señalarse que la raíz cuadrada de  $r^2$  es el coeficiente de correlación de la muestra.

$$\sqrt{r^2} = \pm r, \quad (33)$$

donde el signo de  $r$  es igual que el de la pendiente de regresión  $b$ .

### PRUEBA F SOBRE $\beta$

El análisis del procedimiento de varianza en el análisis de regresión lineal puede emplearse para probar la significación de  $b$  mediante una razón  $F$ . Resumiendo los resultados de la sección anterior en un cuadro AV, cuadro 1. Se puede observar que para  $SCR$ ,  $\delta = k$ , donde  $k$  es el número de variables

independientes o el número de coeficientes  $b$ . Al calcular SCE se pierden dos grados de libertad: uno es representado por  $b$  y el otro por  $a$ . Así, para SCE,  $\delta = n - k - 1$ . También se puede observar que  $CME = \hat{\sigma}^2$ . Sin embargo, debido a errores de redondeo, CME difiere ligeramente de  $\hat{\sigma}^2$  calculada con (11). El número de grados de libertad asociados con SCT es la suma de los grados de libertad asociados con SCR y SCE.

Cuadro 11.1. Cuadro AV en regresión lineal simple

Fuente de variación	TM	$\delta$	CM
Regresión	SCR	$k$	$SCR / k$
Residual	SCE	$n - k - 1$	$SCE / n - k - 1$
Total	SCT	$n - 1$	$SCT / n - 1$

Consideremos ahora una prueba alternativa sobre la pendiente de población. Para esto, observemos que si la hipótesis nula  $\beta = 0$  es cierta, las variaciones en  $Y$  de una observación a otra no serán afectadas por las variaciones en  $X$ , pero deben ser explicadas por trastorno al azar solamente. Esto significa que SCR se aleja de cero solo debido al error de muestreo. Así, si hay una relación significativa entre  $Y$  y  $X$ , la razón  $SCR/SCE$  sería diferente de cero solo debido a variaciones en el muestreo. Esta afirmación puede aplicarse igualmente a  $CMF/CME$  que, siendo una razón de dos variables chi cuadrado, tiene una distribución  $F$  con  $\delta_1 = 1$  y  $\delta_2 = n - 2$  en el caso de regresión lineal simple. Usando esta  $F$  como estadística de prueba para  $\beta = 0$  contra  $\rho \neq 0$ , se definiría la región de rechazo como

$$\frac{CMF}{CME} > F_{1, n-2; 1-\alpha} \quad (34)$$

Cabe destacar que la prueba  $F$  (34) es equivalente a una prueba  $t$  bilateral alrededor de  $\beta$  en que ambas pruebas dan la misma respuesta para el mismo nivel de significación y el mismo conjunto de datos de muestra. La diferencia es que la prueba  $f$  puede ser generalizada para abarcar más de una variable explicativa, mientras que la prueba  $t$  solo puede usarse para un coeficiente de regresión cada vez.

## OBSERVACIONES FINALES

En este capítulo se ha tratado, en forma bastante detallada, modelos simples de regresión y correlación lineal presentando la medida de coeficiente de determinación como el vínculo entre los dos. Puesto que estos dos modelos se elaboran bajo diferentes supuestos, proporcionan diferentes tipos de información, y no siempre está claro qué medida debe ser usada en problemas determinados. Puede ser útil entonces comparar de nuevo los dos modelos.

Primero, se verifica el mismo tipo de relación para el análisis de regresión y el análisis de correlación, como lo indica el hecho de que  $r$  adopta el mismo signo que  $b$ . Igualmente, si el valor de  $b$  es significativo en un nivel de significancia dado,  $r$  es también significativo en dicho nivel.

Segundo,  $r$ , por medio de  $r^2$ , puede considerarse como una medida del grado de ajuste de la línea de regresión. En general, cuanto mayor es el valor de  $r^2$ , tanto mejor es el ajuste y más útil es la ecuación de regresión como un instrumento de predicción. (Obrévese que cuanto mayor es el valor de  $r^2$ , tanto más se acerca  $r$  a  $+1$  o  $-1$ ).

Dadas las dos primeras observaciones, se debe comprender que el objetivo principal del análisis de regresión es establecer dependencia estadística entre las variables dependientes e independientes, de modo que la primera pueda predecirse por la segunda; el coeficiente de correlación es una medida del grado de ajuste de la línea de regresión o una medida del grado de covariabilidad entre  $Y$  y  $X$ , o ambas cosas a la vez.

Tercero, un valor dado del coeficiente de correlación es compatible con un número infinito de líneas de regresión. Por ejemplo, en el caso de correlación perfecta, cuando todos los puntos son desplazados hacia arriba o hacia abajo igual cantidad, el valor del coeficiente de correlación seguirá siendo 1, pero

resultará una diferente ordenada en el origen Y. Análogamente, si todos los puntos giran en un punto dado de modo que todos caen precisamente sobre una línea recta diferente, la pendiente, y quizá la ordenada en el origen Y, cambiará, mientras que el coeficiente de correlación asumirá aún el valor de +1 o -1 (o resultará indeterminado para una línea recta vertical u horizontal).

Por tanto, teóricamente, la regresión es un método direccional, pero la correlación no lo es. La primera debe ser usada si una variable es claramente dependiente de la otra. La segunda es apropiada cuando ninguna de las dos variables puede considerarse como una consecuencia de la otra.

Prácticamente, el objeto de la investigación es quizá la consideración más importante al escoger entre técnicas de regresión y de correlación. En otras palabras, la elección depende de si deseamos tener una ecuación predictiva o simplemente queremos determinar el grado de asociación. En situaciones prácticas, a menudo nos interesamos más por averiguar la naturaleza de la relación para el objetivo de predicción. A este respecto, la razón de nuestro interés por el grado de relación debe ser solamente escoger la variable independiente más útil, entre varias variables independientes posibles, para el análisis de regresión con la variable dependiente que ha sido escogida.

El análisis de asociación, aunque es un instrumento muy útil, a menudo ofrece muchas oportunidades para la mala interpretación. La interpretación más injustificada surge de la confusión entre asociación y causalidad, debido principalmente a la designación convencional de variables dependiente e independiente.

Muchas personas consideran que la variable independiente es la causa, y la variable dependiente es el efecto en una situación dada, pero esto no es forzoso que sea sí.

El vínculo entre asociación y causalidad puede expresarse así: la

presencia de asociación no implica causalidad; pero la existencia de causalidad siempre implica asociación. La evidencia estadística solo puede establecer la presencia o ausencia de asociación entre variables. Si existe o no causalidad depende puramente del razonamiento.

## APLICACION DEL MODELO

Antes de describir y analizar el proceso de crecimiento y desarrollo del país, veamos que significan ambos conceptos.

Crecimiento económico significa el incremento de las actividades económicas, lo que es un fenómeno objetivo, por lo que puede ser observable y también medible. En el crecimiento económico se manifiesta la expansión de las fuerzas productivas, es decir, de la fuerza de trabajo, del capital, de la producción, de las ventas y del comercio.

El crecimiento es un concepto que permite entender en términos cuantitativos el desenvolvimiento de la economía de una nación; se trata de una medición que se hace por periodos específicos con el objeto de hacer comparaciones; puede ser anual, por decenios o sexenal, como en nuestro país.

Las manifestaciones o formas de medir el crecimiento económico y social de un país son los incrementos en las principales variables económicas tales como:

- .La producción total del país, medida por el producto interno bruto (PIB).
- .La producción sectorial, o sea, la producción de los sectores agropecuario, industrial y de servicios.
- .La producción por ramas económicas seleccionadas: producción agrícola, ganadera, pesquera, petrolera, minera, metalúrgica, energética y otras.
- .Los ingresos generados en el país, medidos por el ingreso nacional (IN).
- .Los ingresos per cápita (IPC), que son el promedio resultante de dividir

el ingreso nacional entre la población del país.

.La productividad de la economía, medida en forma general, por sectores o por ramas de actividad.

.El nivel de empleo.

.El nivel de ahorro e inversión, ya sea pública, privada o externa.

.La infraestructura: construcción de caminos, carreteras, escuelas, obras de irrigación, entre otras.

.El volúmen de ventas, etc.

Desarrollo económico es el proceso mediante el cual los países pasan de un estado atrasado de su economía a un estado avanzado de la misma. Este nivel alcanzado en el desarrollo representa mejores niveles de vida para la población en su conjunto; trae como consecuencia que los niveles de vida se vayan mejorando día con día, lo que representa cambios cualitativos y cuantitativos. Dos expresiones fundamentales del desarrollo económico son : aumento de la producción y productividad per cápita en las diferentes ramas económicas y aumento del ingreso real per cápita.

Las manifestaciones y formas de medir el desarrollo económico y social de un país son más difíciles de precisar, dado que implican cambios cuantitativos y cualitativos; entre dichas manifestaciones destacan:

.Aumento de la producción per cápita, medido en general en toda la economía, por sectores o por ramas.

.Aumento de la productividad per cápita, que es el resultado de dividir la producción total de cada trabajador entre el tiempo empleado.

.Aumento de los salarios nominales y reales.

.Disminución del analfabetismo.

.Mejoramiento en la cantidad y calidad de alimentos ingeridos por persona.



**.Incremento del número de viviendas y mejoras en la condición de las ya existentes.**

**.Disminución de enfermedades infecciosas.**

**.Aumento en la esperanza de vida.**

Para que haya desarrollo se requiere que el crecimiento económico sea superior al crecimiento de la población y que se mejoren los niveles de vida de la población en su conjunto, para lo cual es necesario distribuir el ingreso en forma más equitativa entre los distintos miembros de la sociedad y no destruir el ambiente físico.

## **METODOS CUANTITATIVOS PARA ESTABLECER ESTRATOS ECONOMICOS Y DEMOGRAFICOS.**

Es conveniente describir qué es Condición de Actividad, tomando en cuenta la clasificación de la población dentro de este rubro, antes de operar con los datos contenidos en las Características Económicas de un Censo. Condición de Actividad es la situación que distingue a las personas de 12 años y más de acuerdo con el desempeño o no de una actividad económica en la semana de referencia (intervalo de tiempo fijado para cada censo), este criterio permite clasificar a la población en: población económicamente activa y población económicamente inactiva; en el caso de las personas que simultáneamente realizaron una actividad económica y una no económica, se da prioridad a la actividad económica.

La Población económicamente activa es el total de personas de 12 años y más que en la semana de referencia al censo se encontraban ocupadas o desocupadas. La Población Ocupada es el total de personas de 12 años y más que realizaron cualquier actividad económica en la semana de referencia, a cambio de sueldo, salario, jornal u otro tipo de pago en dinero o especie; incluye además, a las personas que tenían trabajo pero no trabajaron en la semana de referencia por alguna causa temporal, (vacaciones, licencia, enfermedad, mal tiempo, huelga o estaban en espera de iniciar o continuar con las labores agrícolas); incluye también a las personas que ayudaron en el predio, fábrica, tienda o taller de algún familiar sin recibir sueldo o salario de ninguna especie; y a los aprendices o ayudantes que trabajaron sin remuneración; la población ocupada a su vez puede ser clasificada por Rama de Actividad (actividad a la que se dedica la empresa donde labora el trabajador), Ocupación principal (actividad que realiza el trabajador) y

Posición en el trabajo (la Relación que establece el trabajador con la empresa). La Población Desocupada es el total de personas de 12 años y más que en la semana de referencia no tenían trabajo, pero lo buscaban activamente; incluye tanto a los buscadores de trabajo que ya habían trabajado, como a los que buscaron trabajo por primera vez.

La Población Económicamente Inactiva es el total de personas de 12 años y más que en la semana de referencia no realizaron ninguna actividad económica, ni buscaron trabajo. La población económicamente inactiva se clasifica en :

.Estudiantes

.Personas dedicadas a los quehaceres de su hogar

.Jubilados o Pensionados

.Incapacitados permanentemente para trabajar

.Otro tipo de inactivos

Dentro de la información acerca de las características económicas se encuentran los grupos de ingreso, que es una clasificación de los ingresos que la población económicamente activa declaró recibir, expresados en rangos de salarios mínimos mensuales vigentes en la semana de referencia. Dentro de estos cuadros aparece la clase de NO ESPECIFICADO que es aquella que no pertenece a ningún rango especificado, para no perder dicha información se trata de distribuir proporcionalmente la cantidad entre los rangos establecidos, esto es lo que se llama Método de Distribución por Prorrateo, que se describe a continuación.

#### **Método de Distribución por prorrateo.**

El método de distribución por prorrateo consiste en distribuir proporcionalmente una cantidad entre varias clases. Este método se aplica generalmente en situaciones en la que dicha cantidad no se quiere manipular

dentro de un análisis debido a sus características particulares. Este método es aplicable a la clase de datos no especificados dentro del Censo de población.

#### Algoritmo.

1. Se tienen las cantidades dentro de las clases especificadas  $\alpha_1, \alpha_2, \dots, \alpha_n$  y la cantidad no especificada  $\alpha^*$ . Sea  $T = \sum_{i=1}^n \alpha_i$

2. Hacer  $\alpha_{i1} = \alpha^* (\alpha_i / T)$  para  $i=1, \dots, n$

3. Redondear las  $\alpha_{i1}$  siguiendo el siguiente criterio:

Si la parte decimal de  $\alpha_{i1}$  es mayor que 0.5,

$$\alpha_{i1} = \alpha_i + \alpha_{i1} - (\text{parte decimal}) + 1$$

Si no,

$$\alpha_{i1} = \alpha_i + \alpha_{i1} - (\text{parte decimal})$$

Nota: Al ir realizando este redondeo siempre se tendrá en cuenta que

$$\sum_{i=1}^n \alpha_{i1} = T + \alpha^*$$

En la sección de ingreso mensual se hace necesario el introducir algunos conceptos técnicos básicos, como lo son la curva de Lorenz y el coeficiente de Gini.

#### Curva de Lorenz.

La curva de Lorenz es la representación gráfica de frecuencias acumuladas y puede ser usada para comparar lo que sería una distribución perfectamente equitativa con la distribución del ingreso de una cierta población. Se dice que el ingreso está distribuido de una manera equitativa si para cualquier  $\alpha$  porcentaje que se tome de la población, ésta tendrá siempre el mismo ingreso  $I(\alpha)$ ; la traducción geométrica de este hecho es una recta que pasa por el origen y con pendiente igual al ingreso más alto dentro del rango que se este

manejando. La curva de Lorenz se construye con las frecuencias acumuladas obtenidas de dividir el rango de cada clase entre el ingreso mayor. En un cuadro de distribución del ingreso jamás aparece un ingreso límite, sino (- +) pero éste puede ser obtenido al sumar el doble del penúltimo rango al extremo izquierdo del último intervalo.

A los porcentajes de población se les agrupa en  $P_i$ , a las frecuencias acumuladas de éstos en  $p_i$ ; al resultado de dividir el rango de cada clase entre el ingreso mayor en  $Q_i$  y a las frecuencias acumuladas de este último resultado en  $q_i$ . Teniendo estas columnas se grafica  $q_i$  contra el extremo derecho de cada intervalo de ingreso para obtener la curva de Lorenz. El resto de datos nos son útiles para calcular el coeficiente de Gini.

#### Coefficiente de Gini.

El coeficiente de Gini es una medida que nos indica que tan mal se distribuye cierta población con respecto a la ideal distribución de ésta. Este coeficiente se expresa en términos porcentuales y se define como:

$$\text{COEFICIENTE DE GINI} = \text{AREA A} / \text{AREA (A + B)}$$

donde:

.Area A es el área entre la gráfica de la línea de perfecta distribución y la curva de Lorenz.

.Area B es el área entre la curva de Lorenz y el eje de las abscisas, donde  $0\% < \alpha < 100\%$

Una manera de obtener el coeficiente de Gini (G) con los datos con que contamos es usar el siguiente resultado:

$$G = \sum_{i=1}^n (p_{i-1} \cdot q_i) - (p_i \cdot q_{i-1})$$

donde  $0 \leq G \leq 1$ .

**Cuadro III.1. POBLACION ECONOMICAMENTE ACTIVA POR GRUPOS DE INGRESO MENSUAL**

1970		Salario=27.93	
Gpos. de Ing.	PEA	Prorrateo	Total
0 - 100	77782	2332.233	80114.23
200 - 499	266921	8003.419	274924.4
500 - 999	674915	20236.8	695151.8
1000-1499	493842	14807.47	508649.5
1500-2499	329495	9879.652	339374.7
2500-4999	206218	6183.287	212401.3
5000-9999	81036	2429.802	83465.8
10000 - +	35830	1074.335	36904.33
Suma	2166039		
No especificado	64947	64947	
Totales	2230986		2230986

Fuente: INEGI. "D.F., IX Censo General de Poblacion y vivienda, 1970". Mexico. Cuadro 28.

**Cuadro III.2. POBLACION ECONOMICAMENTE ACTIVA POR GRUPOS DE INGRESO MENSUAL**

1980		Salario=140.69	
Gpos. de Ing.	PEA	Prorrateo	Total
0	263649	30643.02	294292
1- 435	20651	2400.195	23051.2
436-590	10803	1255.596	12058.6
591-800	25821	3001.087	28822.09
801-1080	37745	4386.973	42131.97
1081-1480	49251	5724.276	54975.28
1481-1970	72985	8482.798	81467.0
1971-2670	135851	15789.5	151640.5
2671-3610	148325	17239.31	165564.3
3611-4890	481076	55913.82	536989.8
4891-6610	561386	65247.98	626634
6611-8950	387030	44983.18	432013.2
8951-12110	327584	38073.97	365658
12111-16390	161890	18815.92	180705.9
16391-22170	121369	14106.3	135475.3
22171-30000	85908	9984.794	95892.79
30001 - +	76336	8872.273	85208.27
Suma	2967660		
No especificado	344921	344921	
Totales	3312581		3312581

Fuente: INEGI. "D.F., X Censo General de Poblacion y vivienda ,1980". Mexico. Características Económicas, cuadro 13.

**Cuadro III.3. POBLACION ECONOMICAMENTE ACTIVA POR GRUPOS DE INGRESO MENSUAL**

1990		Salario=9138.89	
Gpos. de Ing.	PEA	Prorrateso	Total
0	31231	977.2535	32208.25
1-137070	113677	3557.083	117234.1
137070-274140	446221	13962.76	460183.8
274141-548310	1199572	37535.97	1237108
548311-822480	455570	14255.3	469825.3
822480-1370820	325132	10173.75	335305.8
1370820-2741640	196796	6157.971	202954
2741641 - +	103221	3229.903	106450.9
Suma	2871420		
No especificado	89850	89850	
Totales	2961270		2961270

Fuente: INEGI. "D.F., XI Censo General de Poblacion y vivienda Resultados Definitivos, 1990". Mexico. Cuadro 36.

**Cuadro III.4. POBLACION ECONOMICAMENTE ACTIVA DE 1970 POR GRUPOS DE INGRESO MENSUAL Y FRECUENCIAS (BASE 1970)**

1970		Base 1970=100			
Gpos. de Ing \$	PEA	Pi	pi	Qi	qi
0 - 199	80114	3.59%	0.03591	1.00%	0.009951
200 - 499	274924	12.32%	0.15914	1.50%	0.024902
500 - 999	695152	31.16%	0.470729	2.50%	0.049855
1000-1499	508650	22.80%	0.698722	2.50%	0.074807
1500-2499	339375	15.21%	0.850841	5.00%	0.124762
2500-4999	212401	9.52%	0.946046	12.50%	0.249725
5000-8999	83466	3.74%	0.983458	25.00%	0.4997
10000-18998	36904	1.65%	1	49.99%	0.99965
TOTALES	2230986	1		0.99965	

**Cuadro III.5. POBLACION ECONOMICAMENTE ACTIVA DE 1980 POR GRUPOS DE INGRESO MENSUAL Y FRECUENCIAS (BASE 1970)**

1980	Base 1970=100/1980=508*				
Gpos. de Ing \$	PEA	PI	pi	QI	qi
0-86	317343	9.58%	0.095799	0.96%	0.009567
87-116	12059	0.36%	0.09944	0.32%	0.012793
117-157	28822	0.87%	0.10814	0.44%	0.017243
158-213	42132	1.27%	0.120859	0.61%	0.023362
214-287	54975	1.66%	0.137455	0.81%	0.031483
288-388	81468	2.46%	0.162049	1.11%	0.042608
389-526	151641	4.58%	0.207826	1.52%	0.057848
627-711	165564	5.00%	0.257806	2.05%	0.078318
712-963	536990	16.21%	0.419912	2.79%	0.106241
964-1301	626634	18.92%	0.60908	3.75%	0.143731
1302-1782	432013	13.04%	0.739496	5.12%	0.194905
1783-2384	365658	11.04%	0.849881	6.91%	0.263989
2385-3226	180706	5.46%	0.904432	9.36%	0.357548
3227-4384	135475	4.09%	0.945329	12.65%	0.484036
4385-5906	95893	2.89%	0.974277	17.14%	0.655468
5907-8989	85208	2.57%	1	34.29%	0.998331
<b>TOTALES</b>	<b>3312581</b>	<b>1</b>		<b>0.998331</b>	

\*Fuente: Banco de Mexico. "Informe Anual 1991". Mexico.

**Cuadro III.6. POBLACION ECONOMICAMENTE ACTIVA DE 1990 POR GRUPOS DE INGRESO MENSUAL Y FRECUENCIAS (BASE 1970)**

1990	Base 1970=100/1990=65608.7*				
Gpos. de Ing \$	PEA	PI	pi	QI	qi
0-209	149442	5.05%	0.050466	2.50%	0.025006
210-418	460184	15.54%	0.205866	2.49%	0.049892
419-836	1237108	41.78%	0.623629	4.99%	0.099785
837-1254	469825	15.87%	0.782286	4.99%	0.149677
1255-2089	335306	11.32%	0.895516	9.98%	0.249462
2090-4179	202954	6.85%	0.964052	24.99%	0.499402
4180-8358	106451	3.59%	1	49.99%	0.999282
<b>TOTALES</b>	<b>2961270</b>	<b>1</b>		<b>0.999282</b>	

\*Fuente: Banco de Mexico. "Informe Anual 1991". Mexico.



**Cuadro III.7. CALCULO DEL COEFICIENTE DE GINI EN DISTRIBUCION DE INGRESOS 1970**

Gpos. de Ing. \$	pi	qi	Sumandos del Coeficiente de Gini
0-199	0.03591	0.00951	-0.00069
200-499	0.15914	0.024902	-0.00379
500-999	0.470729	0.049655	0.000379
1000-1499	0.698722	0.074807	0.023525
1500-2499	0.850841	0.124762	0.094445
2500-4999	0.946046	0.249725	0.227145
5000-8999	0.983458	0.4997	0.483414
10000-19998	1	0.99965	0
Coeficiente de Gini			0.824431

**Cuadro III.8. CALCULO DEL COEFICIENTE DE GINI EN DISTRIBUCION DE INGRESOS 1980**

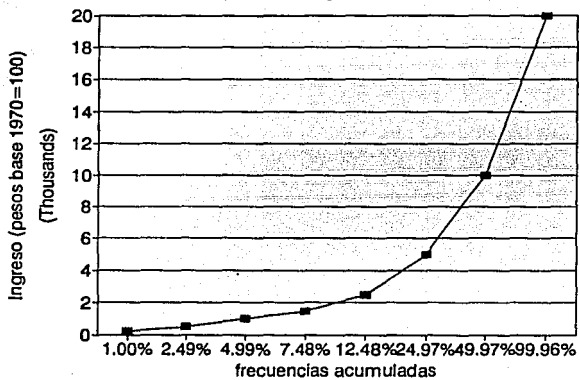
Gpos. de Ing. \$	pi	qi	Sumandos del Coeficiente de Gini
0-88	0.095799	0.009567	0.000274
87-116	0.09944	0.012793	0.000331
117-157	0.10814	0.017243	0.000442
158-213	0.120859	0.023362	0.000594
214-287	0.137455	0.031483	0.000755
288-388	0.162049	0.042608	0.000519
389-526	0.207826	0.057840	0.001363
527-711	0.257806	0.078318	-0.0055
712-963	0.419912	0.106241	-0.00435
964-1301	0.609908	0.143731	0.012424
1302-1762	0.739496	0.194905	0.029573
1763-2384	0.849881	0.263989	0.065113
2385-3229	0.904432	0.357548	0.099777
3227-4364	0.945329	0.484036	0.148048
4365-5906	0.974277	0.655468	0.317184
6907-8989	1	0.998331	0
Coeficiente de Gini			0.667764

**Cuadro III.9. CALCULO DEL COEFICIENTE DE GINI EN DISTRIBUCION DE INGRESOS 1990**

Gpos. de Ing. \$	pi	qi	Sumandos del Coeficiente de Gini
0-209	0.050466	0.025006	-0.00263
210-418	0.205866	0.049892	-0.01057
419-836	0.623629	0.099785	0.015283
837-1254	0.782286	0.149677	0.061112
1255-2099	0.895516	0.249462	0.206728
2099-4179	0.964052	0.499402	0.463958
4180-8358	1	0.999282	0
Coeficiente de Gini			0.73388

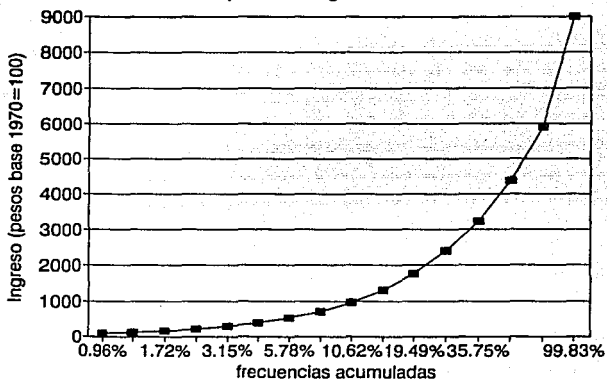
# CURVA DE LORENZ

## Grupos de Ingreso 1970



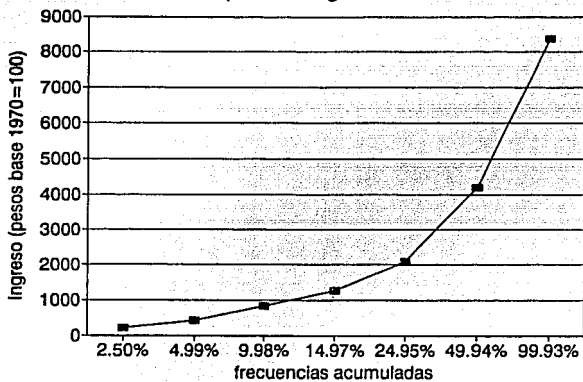
# CURVA DE LORENZ

## Grupos de Ingreso 1980



# CURVA DE LORENZ

Grupos de Ingreso 1990



**Cuadro III.10. PRONOSTICO DE LOS PORCENTAJES DE POBLACION  
ECONOMICAMENTE ACTIVA POR GRUPOS DE INGRESO MENSUAL**

Veces el	1	2	3	4	5	6	7	8	
Sal. Mínim	0-0.5	0.51-1	1.01-2	2.01-3	3.01-5	5.01-7	7.01-10	10.01-+	TOTALES
1970	15.91	31.16	22.80	15.21	8.18	3.07	2.01	1.66	100.00
1980	16.20	25.79	31.96	11.04	9.54	2.89	2.27	0.31	100.00
1990	5.05	15.54	41.78	15.87	11.32	3.73	3.12	3.59	100.00
I. 1991	6.41	15.57	42.62	14.40	11.41	3.59	3.08	2.91	100.00
II. 1992	5.87	14.79	43.57	14.44	11.56	3.63	3.13	3.01	100.00
III. 1993	5.33	14.01	44.52	14.47	11.72	3.66	3.19	3.11	100.00
IV. 1994	4.78	13.23	45.47	14.50	11.88	3.69	3.24	3.20	100.00
V. 1995	4.24	12.45	46.41	14.53	12.03	3.72	3.30	3.30	100.00
VI. 1996	3.70	11.67	47.36	14.57	12.19	3.76	3.35	3.40	100.00
VII. 1997	3.16	10.89	48.31	14.60	12.35	3.79	3.41	3.49	100.00
VIII. 1998	2.61	10.11	49.26	14.63	12.51	3.82	3.47	3.59	100.00
IX. 1999	2.07	9.32	50.21	14.67	12.66	3.86	3.52	3.69	100.00
X. 2000	1.53	8.54	51.16	14.70	12.82	3.89	3.58	3.78	100.00

I.1. Regression Output:		I.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	4.67036	Std Err of Y Est	1.992252
R Squared	0.729986	R Squared	0.96849
No. of Observations	3	No. of Observations	3
Degrees of Freedom	1	Degrees of Freedom	1
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.330244	Std Err of Coef.	0.140873

I.3. Regression Output:		I.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.269444	Std Err of Y Est	3.674235
R Squared	0.999597	R Squared	0.015877
No. of Observations	3	No. of Observations	3
Degrees of Freedom	1	Degrees of Freedom	1
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.019053	Std Err of Coef.	0.259808

I.5. Regression Output:		I.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.171464	Std Err of Y Est	0.416413
R Squared	0.994072	R Squared	0.556748
No. of Observations	3	No. of Observations	3
Degrees of Freedom	1	Degrees of Freedom	1
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.012124	Std Err of Coef.	0.029445

I.7. Regression Output:		I.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.240866	Std Err of Y Est	1.89019
R Squared	0.91393	R Squared	0.34266
No. of Observations	3	No. of Observations	3
Degrees of Freedom	1	Degrees of Freedom	1
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.017032	Std Err of Coef.	0.133657

II.1. Regression Output:		II.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	3.302444	Std Err of Y Est	1.408735
R Squared	0.79717	R Squared	0.97811
No. of Observations	4	No. of Observations	4
Degrees of Freedom	2	Degrees of Freedom	2
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.193676	Std Err of Coef.	0.082617

II.3. Regression Output:		II.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.190526	Std Err of Y Est	2.598076
R Squared	0.999723	R Squared	0.022916
No. of Observations	4	No. of Observations	4
Degrees of Freedom	2	Degrees of Freedom	2
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.011174	Std Err of Coef.	0.152367

II.5. Regression Output:		II.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.121244	Std Err of Y Est	0.294449
R Squared	0.995914	R Squared	0.646142
No. of Observations	4	No. of Observations	4
Degrees of Freedom	2	Degrees of Freedom	2
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.00711	Std Err of Coef.	0.017268

II.7. Regression Output:		II.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.170318	Std Err of Y Est	1.336566
R Squared	0.93916	R Squared	0.431112
No. of Observations	4	No. of Observations	4
Degrees of Freedom	2	Degrees of Freedom	2
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.009989	Std Err of Coef.	0.078385

III.1. Regression Output:		III.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	2.696434	Std Err of Y Est	1.150227
R Squared	0.829221	R Squared	0.982207
No. of Observations	5	No. of Observations	5
Degrees of Freedom	3	Degrees of Freedom	3
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.142273	Std Err of Coef.	0.06069

III.3. Regression Output:		III.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.155563	Std Err of Y Est	2.12132
R Squared	0.999776	R Squared	0.02816
No. of Observations	5	No. of Observations	5
Degrees of Freedom	3	Degrees of Freedom	3
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.008208	Std Err of Coef.	0.111928

III.5. Regression Output:		III.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.098995	Std Err of Y Est	0.240416
R Squared	0.99669	R Squared	0.692863
No. of Observations	5	No. of Observations	5
Degrees of Freedom	3	Degrees of Freedom	3
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.005223	Std Err of Coef.	0.012685

III.7. Regression Output:		III.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.139064	Std Err of Y Est	1.091301
R Squared	0.950176	R Squared	0.483531
No. of Observations	5	No. of Observations	5
Degrees of Freedom	3	Degrees of Freedom	3
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.007337	Std Err of Coef.	0.057581



IV.1 Regression Output:		IV.2 Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	2.33518	Std Err of Y Est	0.996126
R Squared	0.849632	R Squared	0.984671
No. of Observations	6	No. of Observations	6
Degrees of Freedom	4	Degrees of Freedom	4
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.114217	Std Err of Coef.	0.048722

IV.3 Regression Output:		IV.4 Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.134722	Std Err of Y Est	1.837117
R Squared	0.999807	R Squared	0.032619
No. of Observations	6	No. of Observations	6
Degrees of Freedom	4	Degrees of Freedom	4
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.006589	Std Err of Coef.	0.089856

IV.5 Regression Output:		IV.6 Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.085732	Std Err of Y Est	0.208207
R Squared	0.997155	R Squared	0.72415
No. of Observations	6	No. of Observations	6
Degrees of Freedom	4	Degrees of Freedom	4
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.004193	Std Err of Coef.	0.010184

IV.7 Regression Output:		IV.8 Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.120433	Std Err of Y Est	0.945095
R Squared	0.956883	R Squared	0.521413
No. of Observations	6	No. of Observations	6
Degrees of Freedom	4	Degrees of Freedom	4
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.005891	Std Err of Coef.	0.046226

V.1. Regression Output:		V.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	2.088649	Std Err of Y Est	0.890962
R Squared	0.864716	R Squared	0.988426
No. of Observations	7	No. of Observations	7
Degrees of Freedom	5	Degrees of Freedom	5
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.096051	Std Err of Coef.	0.040973

V.3. Regression Output:		V.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.120499	Std Err of Y Est	1.643168
R Squared	0.99983	R Squared	0.036742
No. of Observations	7	No. of Observations	7
Degrees of Freedom	5	Degrees of Freedom	5
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.005541	Std Err of Coef.	0.075564

V.5. Regression Output:		V.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.076681	Std Err of Y Est	0.186226
R Squared	0.997484	R Squared	0.74809
No. of Observations	7	No. of Observations	7
Degrees of Freedom	5	Degrees of Freedom	5
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.003526	Std Err of Coef.	0.008564

V.7. Regression Output:		V.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.107719	Std Err of Y Est	0.845318
R Squared	0.961693	R Squared	0.552064
No. of Observations	7	No. of Observations	7
Degrees of Freedom	5	Degrees of Freedom	5
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.004954	Std Err of Coef.	0.038874

VI.1. Regression Output:		VI.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	1.906667	Std Err of Y Est	0.813333
R Squared	0.876879	R Squared	0.9878
No. of Observations	8	No. of Observations	8
Degrees of Freedom	6	Degrees of Freedom	6
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.083066	Std Err of Coef.	0.035434

VI.3. Regression Output:		VI.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.11	Std Err of Y Est	1.5
R Squared	0.999847	R Squared	0.040769
No. of Observations	8	No. of Observations	8
Degrees of Freedom	6	Degrees of Freedom	6
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.004792	Std Err of Coef.	0.065349

VI.5. Regression Output:		VI.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.07	Std Err of Y Est	0.17
R Squared	0.997741	R Squared	0.767923
No. of Observations	8	No. of Observations	8
Degrees of Freedom	6	Degrees of Freedom	6
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.00305	Std Err of Coef.	0.007406

VI.7. Regression Output:		VI.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.098333	Std Err of Y Est	0.771667
R Squared	0.965485	R Squared	0.578638
No. of Observations	8	No. of Observations	8
Degrees of Freedom	6	Degrees of Freedom	6
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.004284	Std Err of Coef.	0.033618

VII.1 Regression Output:		VII.2 Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	1.76523	Std Err of Y Est	0.753
R Squared	0.887225	R Squared	0.988943
No. of Observations	9	No. of Observations	9
Degrees of Freedom	7	Degrees of Freedom	7
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.073171	Std Err of Coef.	0.031213

VII.3 Regression Output:		VII.4 Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.10184	Std Err of Y Est	1.38873
R Squared	0.999862	R Squared	0.044843
No. of Observations	9	No. of Observations	9
Degrees of Freedom	7	Degrees of Freedom	7
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.004221	Std Err of Coef.	0.057565

VII.5 Regression Output:		VII.6 Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.064807	Std Err of Y Est	0.157389
R Squared	0.997955	R Squared	0.785183
No. of Observations	9	No. of Observations	9
Degrees of Freedom	7	Degrees of Freedom	7
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.002686	Std Err of Coef.	0.006524

VII.7 Regression Output:		VII.8 Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.091039	Std Err of Y Est	0.714425
R Squared	0.968652	R Squared	0.802691
No. of Observations	9	No. of Observations	9
Degrees of Freedom	7	Degrees of Freedom	7
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.003774	Std Err of Coef.	0.029614

VIII.1. Regression Output:		VIII.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	1.651222	Std Err of Y Est	0.704367
R Squared	0.896328	R Squared	0.989929
No. of Observations	10	No. of Observations	10
Degrees of Freedom	8	Degrees of Freedom	8
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.065291	Std Err of Coef.	0.027851

VIII.3. Regression Output:		VIII.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.095263	Std Err of Y Est	1.299038
R Squared	0.999874	R Squared	0.049063
No. of Observations	10	No. of Observations	10
Degrees of Freedom	8	Degrees of Freedom	8
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.003767	Std Err of Coef.	0.051365

VIII.5. Regression Output:		VIII.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.060622	Std Err of Y Est	0.147224
R Squared	0.998139	R Squared	0.800672
No. of Observations	10	No. of Observations	10
Degrees of Freedom	8	Degrees of Freedom	8
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.002397	Std Err of Coef.	0.005821

VIII.7. Regression Output:		VIII.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.085159	Std Err of Y Est	0.668293
R Squared	0.971394	R Squared	0.625056
No. of Observations	10	No. of Observations	10
Degrees of Freedom	8	Degrees of Freedom	8
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.003367	Std Err of Coef.	0.026424

IX.1. Regression Output:		IX.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	1.556787	Std Err of Y Est	0.664084
R Squared	0.904508	R Squared	0.990799
No. of Observations	11	No. of Observations	11
Degrees of Freedom	9	Degrees of Freedom	9
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.05881	Std Err of Coef.	0.025087

IX.3. Regression Output:		IX.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.089815	Std Err of Y Est	1.224745
R Squared	0.999885	R Squared	0.053501
No. of Observations	11	No. of Observations	11
Degrees of Freedom	9	Degrees of Freedom	9
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.003393	Std Err of Coef.	0.046267

IX.5. Regression Output:		IX.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.057155	Std Err of Y Est	0.138804
R Squared	0.998301	R Squared	0.814841
No. of Observations	11	No. of Observations	11
Degrees of Freedom	9	Degrees of Freedom	9
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.002159	Std Err of Coef.	0.005244

IX.7. Regression Output:		IX.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.080289	Std Err of Y Est	0.630063
R Squared	0.973824	R Squared	0.646191
No. of Observations	11	No. of Observations	11
Degrees of Freedom	9	Degrees of Freedom	9
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.003033	Std Err of Coef.	0.023802

X.1. Regression Output:		X.2. Regression Output:	
Constant	1087.527	Constant	1570.543
Std Err of Y Est	1.476898	Std Err of Y Est	0.630005
R Squared	0.911955	R Squared	0.991579
No. of Observations	12	No. of Observations	12
Degrees of Freedom	10	Degrees of Freedom	10
X Coefficient(s)	-0.543	X Coefficient(s)	-0.781
Std Err of Coef.	0.053354	Std Err of Coef.	0.022759

X.3. Regression Output:		X.4. Regression Output:	
Constant	-1846.84	Constant	-51.3
Std Err of Y Est	0.085206	Std Err of Y Est	1.161895
R Squared	0.999895	R Squared	0.058213
No. of Observations	12	No. of Observations	12
Degrees of Freedom	10	Degrees of Freedom	10
X Coefficient(s)	0.949	X Coefficient(s)	0.033
Std Err of Coef.	0.003078	Std Err of Coef.	0.041974

X.5. Regression Output:		X.6. Regression Output:	
Constant	-301.18	Constant	-62.11
Std Err of Y Est	0.054222	Std Err of Y Est	0.131681
R Squared	0.998446	R Squared	0.82795
No. of Observations	12	No. of Observations	12
Degrees of Freedom	10	Degrees of Freedom	10
X Coefficient(s)	0.157	X Coefficient(s)	0.033
Std Err of Coef.	0.001959	Std Err of Coef.	0.004757

X.7. Regression Output:		X.8. Regression Output:	
Constant	-107.423	Constant	-189.217
Std Err of Y Est	0.076169	Std Err of Y Est	0.59773
R Squared	0.976009	R Squared	0.666352
No. of Observations	12	No. of Observations	12
Degrees of Freedom	10	Degrees of Freedom	10
X Coefficient(s)	0.0555	X Coefficient(s)	0.0965
Std Err of Coef.	0.002752	Std Err of Coef.	0.021593

## Comentarios acerca de los cuadros.

Existen en la actualidad paquetes de software útiles para el manejo de datos numéricos, con los que también se puede realizar regresión, herramienta que sirvió para hacer pronósticos. La ayuda de este tipo aminora el trabajo, no con esto lo suple, ya que es necesario estudiar los resultados y ver su grado de eficiencia, así mismo analizar la relación de variables para poder aplicarlas en un algoritmo como éste.

La implantación del modelo implica el análisis de los resultados de las tablas de regresión, respecto de nuestra metodología estadística:

$$\text{Constant} = a$$

$$\text{Std Err of Y Est} = \sigma$$

$$\text{R Squared} = r^2$$

$$\text{No of Observations} = n$$

$$\text{Degrees of Freedom} = n - 2$$

$$\text{X Coefficient(s)} = b$$

$$\text{Std Err of Coef} = \sigma_b$$

Con base en la teoría estadística especificada en el capítulo II se pueden comentar los datos de la Regresión, no sin antes hacer referencia a los cuadros IV.7., IV.8. y IV.9. que fueron el origen del cuadro Resumen de Pronóstico. El Coeficiente de Gini en el cuadro IV.7. indica que para 1970 el 82.44% de la población se encontraba mal distribuida respecto del ingreso, ya para 1980 esta situación había mejorado con 66.77% de la población en este rubro (cuadro IV.8.). El cuadro IV.9. refleja que para 1990 la situación no mejoró, habiendo un incremento del 6.61%.

Los tres primeros renglones del cuadro Pronóstico de los porcentajes de población económicamente activa por grupos de ingreso mensual, fueron



elaborados en base a la información ya existente y con ayuda del Coeficiente de Gini obtenido para cada década. Partiendo de esos renglones se construyó el cuarto; el quinto fue el resultado de aplicar regresión a los cuatro primeros renglones y así sucesivamente hasta llegar al renglón número 13 o del año 2000. Para calcular la celda (I,1) fue necesario el cuadro I.1. y así respectivamente para cada celda; si se observan los cuadros de regresión para 1991, resaltan como buen pronóstico los porcentajes de población económicamente activa para los grupos de ingreso 2, 3, 5 y 7, ya que  $r^2$  es muy cercana a 1, lo que indica que más del 90% de la variación total en porcentaje se ha explicado por las variaciones en fecha para los rangos de ingreso [0.51-1], [1.01-2], [3.01-5] y [7.01-10], mientras que el peor pronóstico se ha hecho para el grupo de ingreso 4 o [2.01-3], ya que menos del 2% de la variación total en porcentaje se ha explicado por las variaciones en fecha. Esta perspectiva se repite para la regresión de II hasta X, con la salvedad de que el pronóstico es cada vez mejor, esto se explica por la cantidad de datos agregados para cada uno de los cálculos.

Respecto de la tabla de pronóstico, se podría obtener una visión optimista basada únicamente en los extremos: gente con escasos recursos (rango 1) que desciende drásticamente en 14.38%, partiendo desde 1970 hasta el año 2000, y gente con altos recursos (rango 8) que aumenta considerablemente; pero si observamos el resto de los rangos vemos que las cosas se mantienen al mismo nivel poblacional, incluso empeoran, como en el caso del rango 6 y 3 respectivamente, además, por el párrafo anterior podemos notar que los pronósticos más optimistas no son muy buenos, inclusive son más negativos, así que no podemos afirmar que la distribución del ingreso será más equitativa en el futuro, sino, que de continuar como vamos llegaremos a empeorar al respecto.

### Prueba F sobre $\beta$ .

Otra manera de medir que tan buenos o malos estuvieron los pronósticos es con la tabla ANOVA, pero para eso necesitamos calcular SCT o lo que es lo mismo  $m_{11}$ , para así poder obtener SCR. Ya que contamos con  $\sigma$  y al elevarla al cuadrado obtenemos SCE.

Si observamos las pruebas  $f$  obtenidas y las comparamos con las  $F_{(1, n-2)}^{\alpha}$  correspondientes verificamos los resultados anteriores, cuyo parámetro era  $r^2$ .

Para el cuadro de regresión I no podemos rechazar la hipótesis nula para los grupos de ingreso 1, 2, 4, 6, 7 y 8 ni con  $\alpha = 0.10$ , para el cuadro II la situación es la misma en los grupos 4 y 8, pero a partir del cuadro III y hasta el X rechazamos la hipótesis nula  $\beta = 0$  para todos los grupos de ingreso. La  $F$  que sirve para rechazar la hipótesis nula en el cuadro III y todos los grupos de ingreso es aquella con  $\alpha = 0.10$ , para el IV  $\alpha = 0.025$ , para el V  $\alpha = 0.01$  y del VI en adelante con  $\alpha = 0.005$  ya podemos rechazar  $\beta = 0$  para todos los grupos de ingreso. Esto comprueba aquello de que el pronóstico se mejora al agregar datos para el cálculo del mismo.

Por otro lado, se veía que para ciertos grupos de ingreso, más del 90% de la variación total en porcentaje de población se explicaba por las variaciones en fecha, dicho fenómeno se verifica también con la prueba  $f$ , ya que si ordenamos para cada cuadro las  $F$  respecto a los grupos de ingreso que más rechazan la hipótesis nula, obtenemos el mismo resultado en los diez :grupo de ingreso 3, 5, 2, 7, 1, 6, 8 y 4. Lo que significa que nuestro pronóstico optimista acerca de los más pobres y los más ricos no es muy confiable, pues los grupos 1 y 8 se encuentran dentro de los mal pronosticados y además reitera que las cosas seguirán iguales para los grupos intermedios. Lo que nos da por conclusión general que si la situación sigue como hasta ahora, la distribución del ingreso no se hará equitativa para años posteriores

			SCT - SCR	F(1,1)									
I.1.	541.0706	450.28853	58.0698	2.7035155									
I.2.		1877.5613	1751.8	121.9922	30.73574								
I.3.			3206.85	3106.6572	180.1202	2480.0945							
I.4.	F=39.9	G=90		605.0826	591.3648	0.2178	0.0161333						
I.5.	F=181	G=95			286.0664	281.1072	4.9298	167.68027					
I.6.	F=848	G=97.5				31.8899	31.2967	0.2178	1.2560554				
I.7.	F=4050	G=99					18.9274	18.25333	0.61805	10.6185			
I.8.	F=16200	G=99.5						15.7356	10.304533	1.84245	0.5212834		

			SCT - SCR	F(1,2)									
II.1.	582.20572	474.66611	96.63348	17.720942									
II.2.		2120.0589	1938.7436	179.33069	180.72833								
II.3.			5103.7202	4841.3068	261.68904	14428.563							
II.4.	F=8.53	G=90		812.52901	796.71238	7.0660267	2.0938153						
II.5.	F=18.5	G=95			418.18605	408.98995	7.1813967	977.06078					
II.6.	F=38.5	G=97.5				44.599549	44.109522	0.4033267	9.3039619				
II.7.	F=98.5	G=99					28.396355	27.442755	0.824591	63.74658			
II.8.	F=199	G=99.5						24.236053	17.9557	4.403945	5.0312629		

			SCT - SCR	F(1,3)									
III.1.	616.67045	486.94842	120.45127	49.659624									
III.2.		2338.8424	2115.7753	221.74404	502.8125								
III.3.			7001.3958	6677.8313	325.54428	40106.795							
III.4.	F=5.54	G=90		1020.9271	1007.0359	9.3911688	6.2607782						
III.5.	F=10.1	G=95			549.91214	541.02882	8.6735208	27.16.3839					
III.6.	F=17.4	G=97.5				57.747425	57.182856	0.5087606	26.302879				
III.7.	F=34.1	G=99					38.209955	37.045513	1.1451036	177.63744			
III.8.	F=55.8	G=99.5						33.304182	26.386405	5.726638	14.426025		

			SCT - SCR	F(1,4)									
IV.1.	643.05448	499.99533	139.60608	102.40556									
IV.2.		2535.1318	2276.1991	257.9405	1039.8031								
IV.3.			8983.1631	8606.6323	376.50567	82978.456							
IV.4.	F=4.54	G=90		1230.2791	1216.3239	10.580202	12.539499						
IV.5.	F=7.71	G=95			687.29399	676.9813	10.325332	56.19.2283					
IV.6.	F=12.2	G=97.5				71.135706	70.507104	0.585252	54.002491				
IV.7.	F=21.2	G=99					48.374362	47.028801	1.331057	367.08265			
IV.8.	F=31.3	G=99.5						42.96281	35.487473	6.572133	29.431717		

			SCT - SCR	F(1,5)									
V.1.	667.94751	506.71379	156.87127	179.79707									
V.2.		2710.1471	2417.7536	291.59967	1836.7016								
V.3.			11050.32	10624.392	425.9137	145664.5							
V.4.	F=4.06	G=90		1440.5871	1426.5721	11.314941	20.953595						
V.5.	F=6.61	G=95			828.38067	818.89601	11.678976	9931.1018					
V.6.	F=10.0	G=97.5				84.76657	84.078229	0.6536614	94.241844				
V.7.	F=16.3	G=99					58.895735	57.3812	1.5029315	647.62922			
V.8.	F=22.8	G=99.5						53.230562	45.254381	7.2619173	50.811572		

			SCT - SCR	F(1,8)																
VI.1.	685 93925	508 77842	173 52548	286 39465																
VI.2.		2865 1081	2539 7636	324 68078	2944 9007															
VI.3.			13204 672	12730 096	474 56465					235321.31										
VI.4.	F=3.78	G=90		1651 8533	1637.7795	11 823767				31.530045										
VI.5.	F=5.99	G=95			973 22209	960 20575				13 011442	15932.378									
VI.6.	F=8.81	G=975			98 642195	97 895028	0 7182669	149 12115												
VI.7.	F=13.7	G=99				69 780236	68 099313	1 6712539				1037.032								
VI.8.	F=18.6	G=995							64 126063	55 646854	7 8837389	79 437214								

			SCT - SCR	F(1,7)																
VII.1.	699 81939	508 205	190 29835	427 49427																
VII.2.		3001 2348	2642 2684	358 39936	4424 6091															
VII.3.			15448 021	14923 799	524 21201				353807											
VII.4.	F=3.59	G=90		1864 0799	1849 9481	12 205227	44 300452													
VII.5.	F=5.59	G=95			1121 867	1107 4918	14 370918	23751.53												
VII.6.	F=8.07	G=975				112 76478	111 95756	0 7824266	221 10093											
VII.7.	F=12.2	G=99						81 034024	78 183302	1 8424341	1556 092									
VII.8.	F=16.2	G=995							75 667936	66 67539	8 4821438	116 3298								

			SCT - SCR	F(1,8)																
VIII.1.	709 57762	496 17993	207 87115	609 33391																
VIII.2.		3119 747	2725 6489	383 80399	6348 7453															
VIII.3.			17782 167	17208 07	578 08792			507848.1												
VIII.4.	F=3.48	G=90		2077 2691	2063 0726	12 509024	59 302042													
VIII.5.	F=5.32	G=95			1274 3648	1258 5699	15 791225	34375 457												
VIII.6.	F=7.57	G=975					127 13644	126 26652	0 8482494	313 07936										
VIII.7.	F=11.3	G=99						92 663261	90 635117	2 0208925	2229 3087									
VIII.8.	F=14.7	G=995							87 874008	78 345876	9 0823297	162 69211								

			SCT - SCR	F(1,9)																
IX.1.	718 40364	487 98764	275 99742	839 24294																
IX.2.		3221 8648	2790 4784	430 94437	8784 6353															
IX.3.			20208 912	19577 763	631 14022	704 164.7														
IX.4.	F=3.36	G=90		2291 4231	2277 16	12 763092	78 578552													
IX.5.	F=5.12	G=95			1430 7848	1413 4632	17 29836	47658 747												
IX.6.	F=7.21	G=975					141 75942	140 82292	0 9172253	428 48166										
IX.7.	F=10.6	G=99						104 67411	102 45787	2 2099858	3085 4725									
IX.8.	F=13.6	G=995							100 7653	90 667137	9 7011848	219 83738								

			SCT - SCR	F(1,10)																
X.1.	720 68717	472 94685	245 55909	1125 7844																
X.2.		3308 806	2837 4563	470 95478	11805 63															
X.3.			22730 056	22039 898	690 15088	950621 01														
X.4.	F=3.29	G=90		2506 544	2492 2095	12 984448	96 181083													
X.5.	F=4.96	G=95			1591 1164	1572 1997	18 913756	64332 504												
X.6.	F=8.94	G=975					156 63568	155 62862	0 9905062	571 22621										
X.7.	F=10.0	G=99						117 07272	114 65448	2 4124506	4158 213									
X.8.	F=12.8	G=995							114 35804	103 64971	10 351047	289 7167								

## CONCLUSIONES

Ahora lo importante para el desarrollo del país es buscar la implantación de un modelo que al mismo tiempo que permita el crecimiento económico, se de desarrollo social, distribuyendo más equitativamente el ingreso entre todos los miembros de la sociedad mexicana. Es decir, un modelo redistribuidor del ingreso, para lo cual sería necesaria una verdadera y real reforma económica, cuya característica principal fuera atacar los problemas de fondo, los estructurales y buscando su impacto en el rubro productivo para que paulatinamente se resuelvan los problemas y se alcance la meta del desarrollo económico.

Lo anterior, obliga a definir esquemas de política económica, acordes con los tiempos de modernidad que el país debe enfrentar, derivado de su incorporación al TLC y al GATT, en el que la sociedad en su conjunto se beneficia con dicha incorporación.

Es impostergable el proceso de reducción de la brecha entre la miseria y la riqueza, lo cual evitará la inestabilidad político-social, elemento fundamental para que un país se desarrolle en armonía.

## APENDICE

Razon es la relación de tamaño entre 2 números cualesquiera a/b, se dividen en tres categorías: a) aquellas en que intervienen los elementos de un universo (mujeres/hombres, universo = población), b) aquellas en las que se considera una parte del universo con el todo (población masculina/población total), c) aquellas que involucran 2 universos (densidad de población = población/Km<sup>2</sup>).

Razon de cambio es el cambio que hay en la población del tiempo 1 al 2 en relación a la población que había en el tiempo 1  $(p_2 - p_1)/p_1$ .

Una Tasa debe expresar el crecimiento como un cambio relativo en el tamaño por año. Para traducir la razón de cambio a tasa de crecimiento, se divide entre el intervalo de tiempo que ha transcurrido  $(p_{t+h} - p_t)/(p_t \cdot h)$

Crecimiento aritmético, su hipótesis es que la población crece una constante poblacional año con año

tasa de crecimiento aritmético =  $r'' = [(p_{t+h}/p_t) - 1]/h$

Crecimiento geométrico, su hipótesis es que la población crece una proporción de la población inicial

tasa de crecimiento geométrico =  $r' = (p_{t+h}/p_t)^{1/h} - 1$

Crecimiento exponencial, su hipótesis es que la población crece de manera continua en cada instante

tasa de crecimiento exponencial =  $r = [\ln(p_{t+h}/p_t)]/h$

$$r'' > r' > r$$

¿Cuál es la tasa más adecuada para pronosticar el crecimiento poblacional?

La hipótesis de la tasa de crecimiento poblacional se acerca más a la realidad, ya que el crecimiento real de la población es continuo y en ese sentido la  $r$  es la tasa más adecuada, pero, los datos o la información para calcular dicho crecimiento está dada de manera discreta, entonces es mejor usar  $r'$ .

## REFERENCIAS

- Aguilar M. Alonso y Carmona Fernando. "Los grandes problemas nacionales - México Riqueza y Miseria". Editorial nuestro tiempo, 1972.
- Banamex. "Examen de la Situación económica de México 1925-1976".
- Banco de México. Dirección de Investigación Económica. "Índices de precios". Noviembre de 1992.
- Banco de México, S.A. "Informe Anual 1991". México.
- Box, George E.P. and Jenkins, G. Wilym M. "Time Series Analysis forecasting and control". San Francisco, California. Holden-Day. 1976.
- Coordinación General de los Servicios Nacionales de Estadística, Geografía e Informática de la Secretaría de Programación y Presupuesto. "Encuesta Nacional de Ingresos y Gastos de los Hogares 1977".
- Chou Ya-Lun. "Análisis estadístico". Segunda edición. México, D.F. Editorial Interamericana. 1977.
- INEGI. "Anuario Estadístico del Distrito Federal". Aguascalientes, Ags. Edición 1991.
- INEGI. "Cuaderno de Información Oportuna". México 1992.
- INEGI. "Distrito Federal, IX Censo General de Población y Vivienda, 1970".
- INEGI. "Distrito Federal, X Censo General de Población y Vivienda, 1980".
- INEGI. "Distrito Federal, Resultados definitivos, XI Censo General de Población y Vivienda, 1990".
- INEGI. "Encuesta Nacional de Ingreso-Gasto de los Hogares, Cuarto Trimestre 1984".
- INEGI. "Encuesta Nacional de Ingreso-Gasto de los Hogares, Tercer Trimestre 1989".
- INEGI. "Estadísticas Históricas de México". Tomos I y II. México, 1990.

López Rosado, Diego G. "Problemas Económicos de México". Dirección General de Publicaciones Editorial Tierra Firme UNAM. 1984.

Méndez Morales, J. Silvestre. "Problemas Económicos de México". Segunda Edición. Mc Graw-Hill. 1991.

Méndez Morales, J. Silvestre y Ballesteros, Nicolás. "Problemas y Política Económicos de México", Vol. II. Mc Graw-Hill. 1991.

Salinas de Gortari, Carlos. "Tercer Informe de Gobierno". En *Comercio Exterior*. México, D.F.. Mensual. 7 de noviembre de 1991.

Salinas de Gortari, Carlos. "Cuarto Informe de Gobierno". En *Excelstor*. Cuarte Parte de la Sección A. México, D.F.. Diario. Lunes 2 de noviembre de 1992.

Subdirección General de estudios administrativos de la Sria. de la Presidencia. "México a través de los informes presidenciales". Tomo V, Vol 5 La Administración Pública. México, 1976.

"Gabinetes". En *Enciclopedia de Mexico, tomo V*. México, D.F.. Enciclopedia de México S.A., 1971. p.p. 1-49.