

Escuela Superior Politécnica del Litoral

Instituto de Ciencias Matemáticas

*"Análisis estadístico de la producción
arrocera en el Ecuador"*

TESIS DE GRADO

Previa a la obtención del título de:

*INGENIERA EN ESTADÍSTICA
INFORMATICA*

Presentada por:

Delia E. Montalvo Roca

Guayaquil - Ecuador

Año

2000

AGRADECIMIENTO

A Dios; a mis padres por haber estado siempre a mi lado apoyándome en todo momento y por ser las personas que me animaban a seguir adelante. Al Ing. Gaudencio Zurita, a David, Flor y José por su apoyo incondicional.

DEDICATORIA

TRIBUNAL DE GRADUACION

Ing. Félix Remírez
Director del ICM

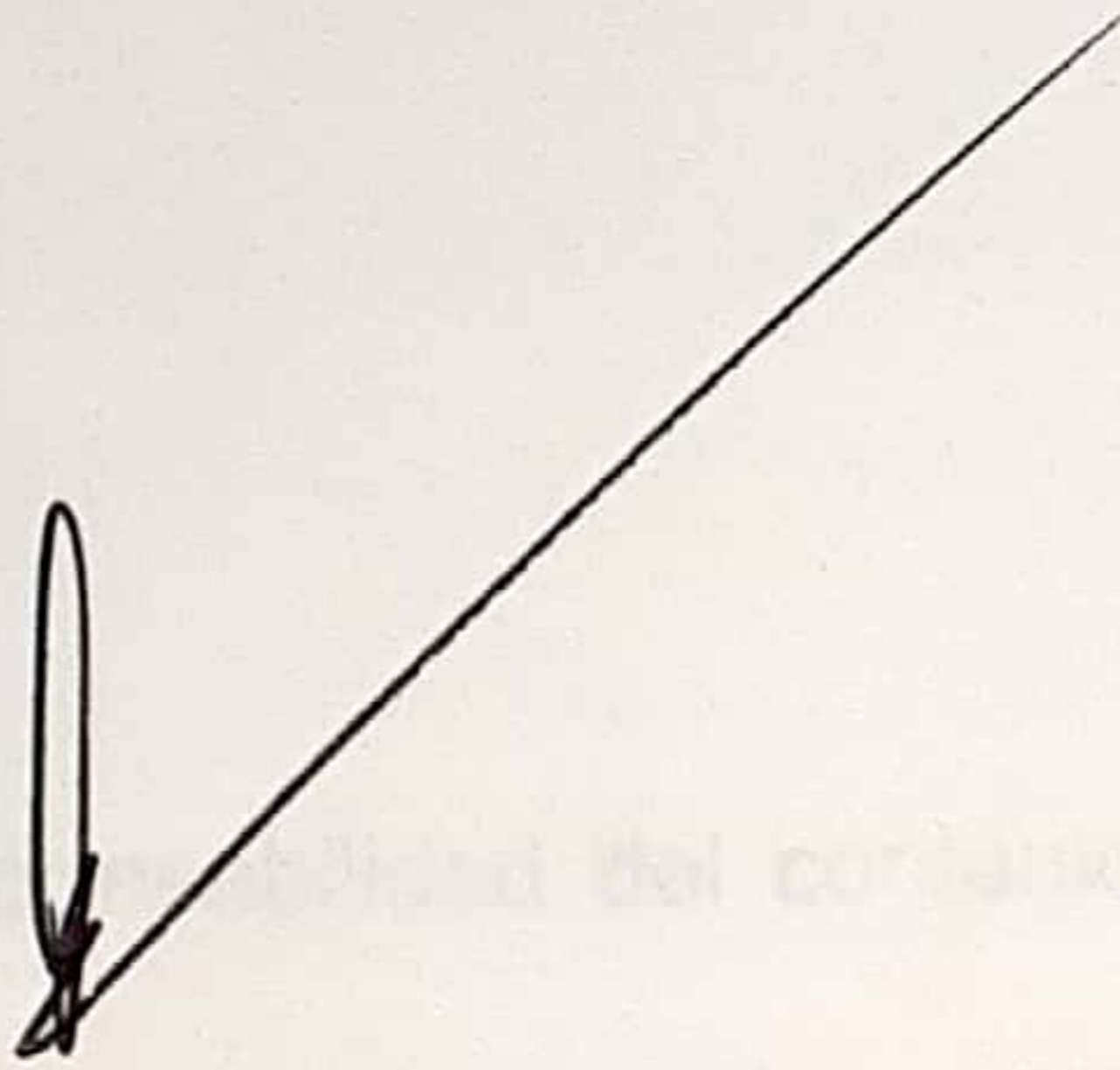
Ing. Jorge Hernández
Vocal

Ing. Sr. Gaudencio Zúñiga

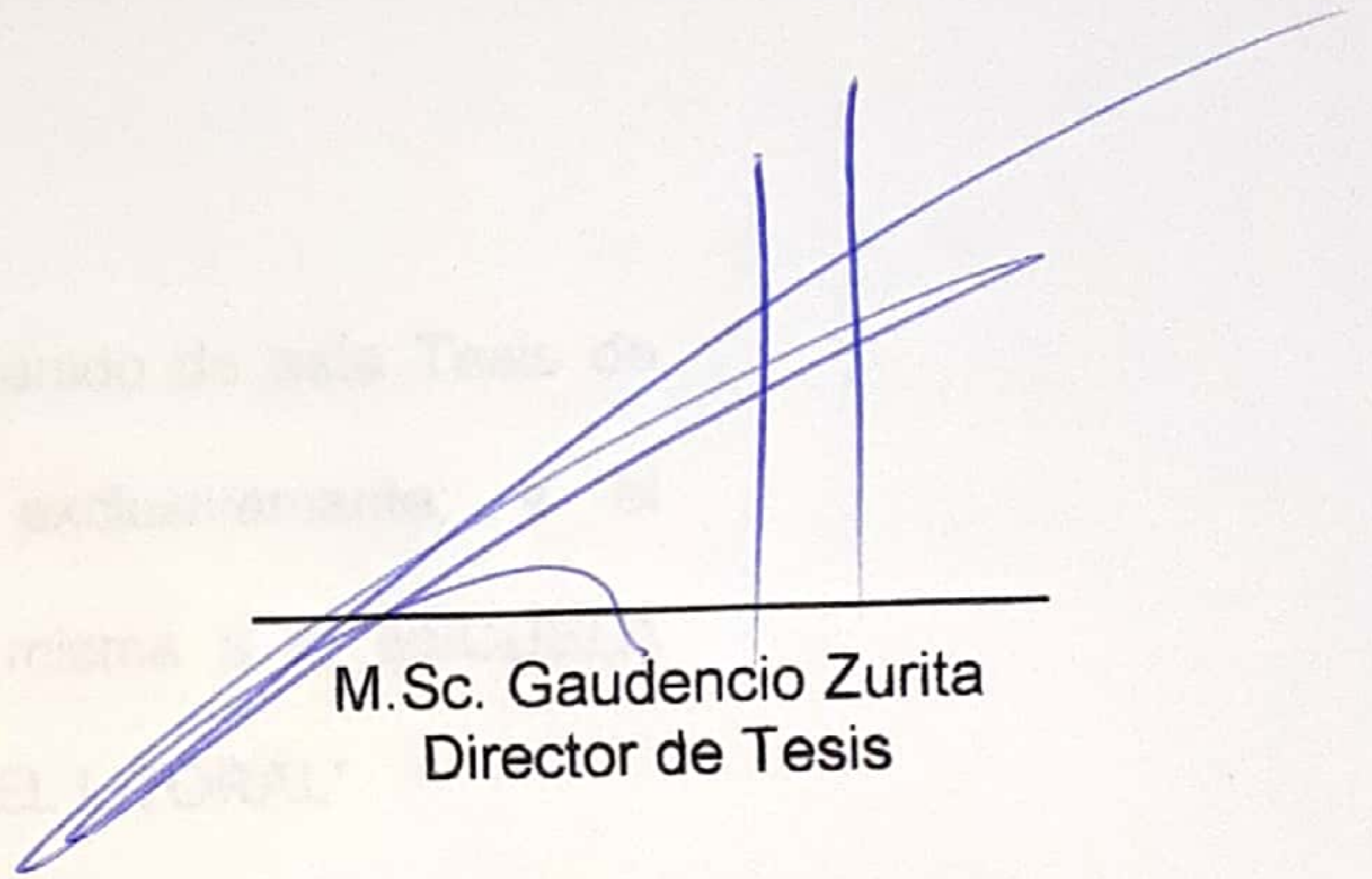
A Dios; a la Virgen María, a mis padres,
y a Kevin, que desde el momento de su
existencia fue la fuerza que me ayudó a
luchar por mis ideales.

Ing. Luis Sánchez
Vocal

TRIBUNAL DE GRADUACIÓN



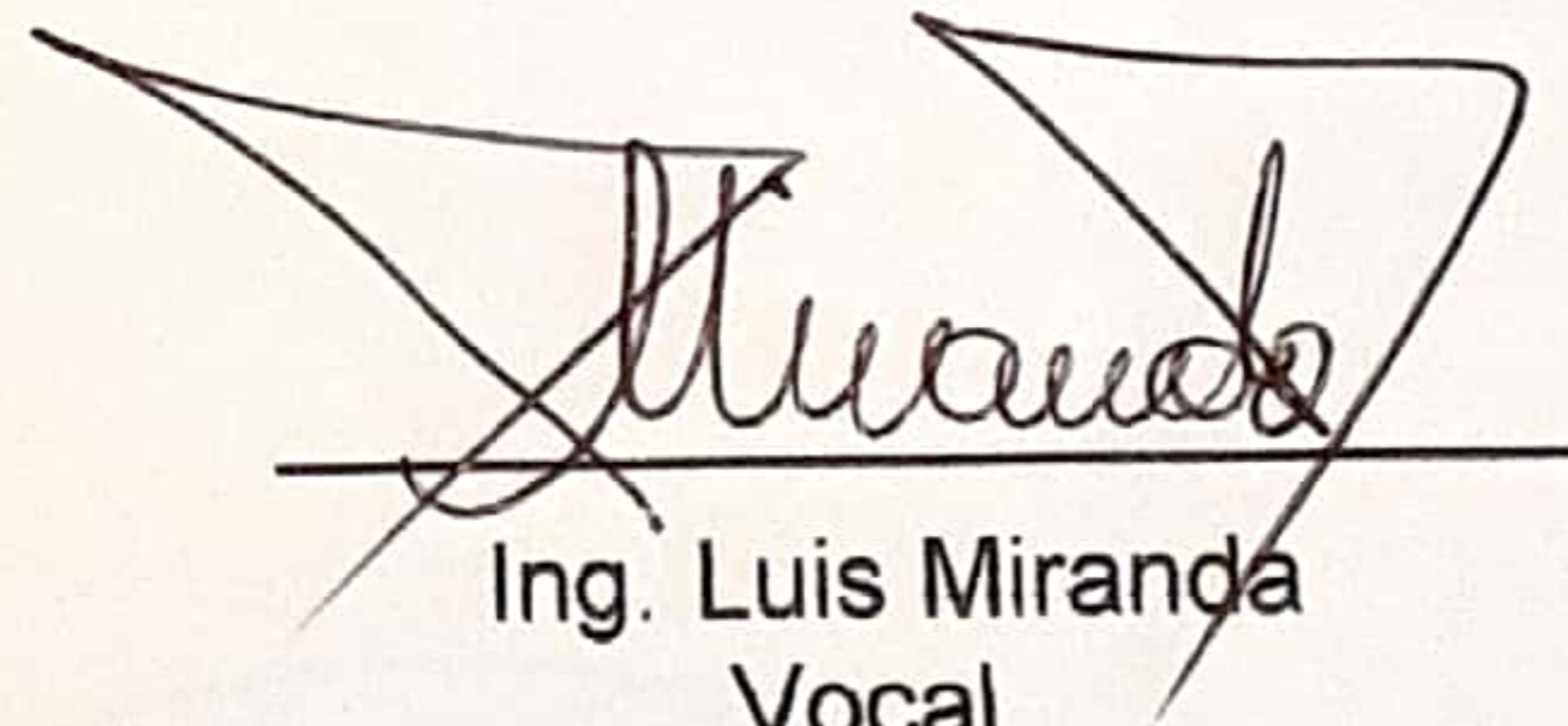
Ing. Félix Ramírez
Director del ICM



M.Sc. Gaudencio Zurita
Director de Tesis



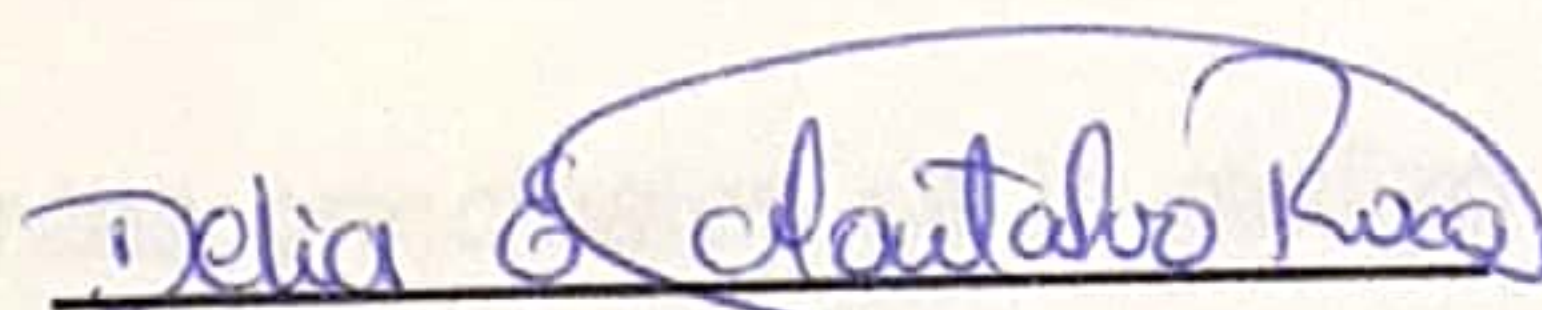
Ing. Jorge Fernández
Vocal



Ing. Luis Miranda
Vocal

DECLARACION EXPRESA

"La responsabilidad del contenido de esta Tesis de Grado, me corresponden exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL"



Delia E. Montalvo Roca

RESUMEN

El arroz más que una costumbre, es una necesidad en todos los hogares a lo largo y ancho del país; ya que constituye un alto porcentaje de la dieta de los ecuatorianos. En este trabajo se ha realizado un estudio estadístico de la producción de este cereal; tomando para ello un esquema de 20 variables; a las cuales se les ha aplicado dos técnicas estadísticas: series temporales y componentes principales.

A las variables: exportaciones del arroz, ipc, y salario mínimo vital se les aplicó series temporales, obteniendo las predicciones de dichas variables, hasta el año 2015. El conjunto de variables restantes fue modelado a través de regresión.

El análisis de componentes principales se lo realizó agrupando todas las 20 variables; de las cuales se obtuvieron 3 variables no observables, a las cuales se las rotuló de la siguiente manera; proceso de producción, industrialización y comercialización del arroz, y clima, de acuerdo con las variables que cada una de ellas contiene.

ÍNDICE GENERAL

	Pág.
RESUMEN	I
ÍNDICE GENERAL	II
ÍNDICE DE FIGURAS	V
ÍNDICE DE TABLAS	VII
I. HISTORIA Y FITOTÉCNIA DEL ARROZ	1
1.1 Introducción	1
1.2 Historia y actualidad del arroz	2
1.3 Información agronómica.....	4
1.3.1 Características generales del cultivo	4
1.4 Clasificación del arroz	5
1.5 Fitotécnia del arroz	7
1.5.1 Introducción	7
1.5.2 Caracteres morfológicos	7

1.5.3	Caracteres fisiológicos	8
1.6	Factores importantes que influyen en el cultivo del arroz	8
1.7	El arroz en el Ecuador	13
II. DETERMINACIÓN DE VARIABLES		16
2.1	Introducción	16
2.2	Descripción de las variables de estudio	17
III. ANÁLISIS DE LAS SERIES TEMPORALES RELACIONADAS CON LA PRODUCCIÓN ARROCERA EN EL PAÍS		23
3.1	Introducción	23
3.2	Aplicación de modelos de series temporales	23
3.3	Análisis de regresión	32
IV. ANÁLISIS MULTIVARIADO		81
4.1	Introducción	81
4.2	Análisis de componentes principales de las variables de estudio.	90
V. CONCLUSIONES Y RECOMENDACIONES		104
APENDICES		

GLOSARIO

BIBLIOGRAFÍA

ÍNDICE DE FIGURAS

	Pág.
Figura 1.1	Planos del arroyo
Figura 2.1	Gráfico de las series de las exportaciones del arroyo
Figura 2.2	Gráfico de las autocorrelaciones de las exportaciones del arroyo
Figura 2.3	Gráfico de las autocorrelaciones parciales de las exportaciones del arroyo
Figura 2.4	Gráfico de la predicción de las exportaciones del arroyo
Figura 3.1	Gráfico de la serie del IPC
Figura 3.2	Gráfico de las autocorrelaciones del IPC
Figura 3.3	Gráfico de las autocorrelaciones parciales del IPC
Figura 3.4	Gráfico de las predicciones para la serie del IPC
Figura 3.5	Gráfico de la serie de desempleo
Figura 3.6	Gráfico de las autocorrelaciones del desempleo
Figura 3.7	Gráfico de las autocorrelaciones parciales del desempleo
Figura 3.8	Gráfico de las predicciones para la serie del desempleo
Figura 3.9	Gráfico de la serie de crecimiento
Figura 3.10	Gráfico de las autocorrelaciones del crecimiento
Figura 3.11	Gráfico de las autocorrelaciones parciales del crecimiento
Figura 3.12	Gráfico de las predicciones para la serie del crecimiento
Figura 3.13	Gráfico de la serie de inflación
Figura 3.14	Gráfico de las autocorrelaciones de la inflación
Figura 3.15	Gráfico de las autocorrelaciones parciales de la inflación
Figura 3.16	Gráfico de las predicciones para la serie de la inflación
Figura 3.17	Gráfico de la serie de PIB
Figura 3.18	Gráfico de las autocorrelaciones del PIB
Figura 3.19	Gráfico de las autocorrelaciones parciales del PIB
Figura 3.20	Gráfico de las predicciones para la serie del PIB
Figura 3.21	Gráfico de la serie de precios
Figura 3.22	Gráfico de las autocorrelaciones de los precios
Figura 3.23	Gráfico de las autocorrelaciones parciales de los precios
Figura 3.24	Gráfico de las predicciones para la serie de los precios
Figura 3.25	Gráfico de la serie de tasas de interés
Figura 3.26	Gráfico de las autocorrelaciones de las tasas de interés
Figura 3.27	Gráfico de las autocorrelaciones parciales de las tasas de interés
Figura 3.28	Gráfico de las predicciones para la serie de las tasas de interés
Figura 4.1	Primeros componentes de la serie del PIB
Figura 4.2	Segundos componentes de la serie del PIB
Figura 4.3	Terceros componentes de la serie del PIB

ÍNDICE DE FIGURAS

		Pág.
Figura 1.1	Planta del arroz	8
Figura 3.1	Gráfico de las series de las exportaciones del arroz	44
Figura 3.2	Gráfico de las autocorrelaciones de las exportaciones del arroz	45
Figura 3.3	Gráfico de las autocorrelaciones parciales de las exportaciones del arroz	45
Figura 3.4	Gráfico de la predicción de las exportaciones del arroz	47
Figura 3.5	Gráfico de la serie del IPC	49
Figura 3.6	Gráfico de las autocorrelaciones del IPC	50
Figura 3.7	Gráfico de las autocorrelaciones parciales del IPC	50
Figura 3.8	Gráfico de las predicciones para la inflación	52
Figura 3.9	Gráfico de las predicciones para la inflación	54
Figura 3.10	Gráfico de la serie de deuda externa en el Ecuador	56
Figura 3.11	Gráfico de autocorrelación de deuda externa	57
Figura 3.12	Gráfico de autocorrelaciones parciales de deuda externa	57
Figura 3.13	Gráfico de regresión Dólar & Deuda Externa	59
Figura 3.14	Gráfico de la serie del salario mínimo vital	60
Figura 3.15	Gráfico de autocorrelación del salario	62
Figura 3.16	Gráfico de autocorrelaciones parciales del salario	62
Figura 3.17	Gráfico de las predicciones del salario mínimo vital	63
Figura 3.18	Gráfico de la serie de remuneraciones	65
Figura 3.19	Gráfico de regresión de remuneraciones	66
Figura 3.20	Gráfico de la serie de producción del arroz en invierno	70
Figura 3.21	Gráfico de regresión de la producción en invierno	71
Figura 3.22	Gráfico de la serie de producción del arroz en verano	73
Figura 3.23	Gráfico de regresión de la producción en verano	74
Figura 3.24	Gráfico de la serie de producción total del arroz	76
Figura 3.25	Gráfico de regresión de la producción total del arroz	77
Figura 3.26	Gráfico de la serie del rendimiento total de la producción del arroz	81
Figura 4.1	Primero componente vs. Componente reducida	104
Figura 4.2	Segunda componente vs. Segunda componente reducida	105
Figura 4.3	Tercera componente principal vs. Tercera componente reducida	106

Figura A.1	Histograma de frecuencias de las exportaciones del arroz.....	Apéndice A
Figura A.2	Histograma de frecuencias del índice de precios al consumidor	Apéndice A
Figura A.3	Histograma de frecuencias de la deuda externa.....	Apéndice A
Figura A.4	Histograma de frecuencias del SMV	Apéndice A
Figura A.5	Histograma de frecuencias de las remuneraciones.....	Apéndice A
Figura A.6	Histograma de frecuencias del precio del dólar.....	Apéndice A
Figura A.7	Histograma de frecuencias de la superficie sembrada en invierno	Apéndice A
Figura A.8	Histograma de frecuencias de la superficie sembrada en verano.. ..	Apéndice A
Figura A.9	Histograma de frecuencias de la superficie sembrada total.....	Apéndice A
Figura A.10	Histograma de frecuencias de la superficie cosechada en invierno	Apéndice A
Figura A.11	Histograma de frecuencias de la superficie cosechada en verano	Apéndice A
Figura A.12	Histograma de frecuencias de la superficie cosechada total.....	Apéndice A
Figura A.13	Histograma de frecuencias de la producción del arroz en invierno.....	Apéndice A
Figura A.14	Histograma de frecuencias de la producción del arroz en verano.....	Apéndice A
Figura A.15	Histograma de frecuencias de la producción total del arroz.....	Apéndice A
Figura A.16	Histograma de frecuencias del rendimiento del arroz en invierno.....	Apéndice A
Figura A.17	Histograma de frecuencias del rendimiento del arroz en verano	Apéndice A
Figura A.18	Histograma de frecuencias del rendimiento total del arroz.....	Apéndice A

ÍNDICE DE TABLAS

		Pág.
Tabla 2.1	Importaciones y exportaciones del arroz (1990 – 1998)	17
Tabla 3.1	Matriz de correlación de las variables económicas	42
Tabla 3.2	Predicción y límites de confianza para las exportaciones del arroz del 2000 - 2018	45
Tabla 3.3	Predicción y límites de confianza para el IPC; modelo ARMA.....	51
Tabla 3.4	Predicción y límites de confianza para el IPC; modelo ARIMA ...	53
Tabla 3.5	Tabla ANOVA del modelo de regresión de la deuda externa	58
Tabla 3.6	Predicción y límites de confianza para el SMV; modelo ARMA ...	61
Tabla 3.7	Tabla ANOVA del modelo de regresión del rendimiento	67
Tabla 3.8	Tabla ANOVA del modelo de regresión de la producción del arroz en invierno.....	69
Tabla 3.9	Tabla ANOVA del modelo de regresión de la producción del arroz en verano	72
Tabla 3.10	Tabla ANOVA del modelo de regresión de la producción total del arroz (primer modelo)	78
Tabla 3.11	Tabla ANOVA del modelo de regresión de la producción total del arroz (segundo modelo)	79
Tabla 3.12	Tabla ANOVA del rendimiento total	82
Tabla 4.1	Matriz de correlaciones de las 20 variables	94
Tabla 4.2	Matriz de covarianzas entre las variables y las componentes principales.....	96
Tabla 4.3	Matriz de correlaciones entre las variables y las componentes principales (datos estandarizados)	98
Tabla 4.4	Matriz de covarianzas entre las dos primeras componentes y las variables con varimax.....	101

CAPITULO I

HISTORIA Y FITOTECNIA DEL ARROZ

1.1 INTRODUCCIÓN

En este trabajo se realiza un análisis de la producción del arroz en nuestro país. Se efectúa un estudio de los factores que influyen directa e indirectamente en una buena producción de este grano, mediante un análisis de series temporales y un análisis multivariado de dichos factores. En el análisis de series temporales, se calculan las proyecciones correspondientes y se estudian los resultados positivos o negativos de los mismos.

Todo lo anteriormente mencionado se irá desarrollando a lo largo de este trabajo en cuatro capítulos, los cuales fueron detallados en el sumario correspondiente. Al término del análisis de series temporales y análisis estadístico multivariado se presentará un conjunto de conclusiones y recomendaciones.

1.2 HISTORIA Y ACTUALIDAD DEL ARROZ

El arroz se cultiva desde tiempos inmemoriales. Su lugar de origen es Asia, que continúa siendo el principal productor, pero en la actualidad se lo cultiva en todos los continentes, y en muchísimos países el arroz es el alimento básico de la dieta de su población. Algunos investigadores han realizado análisis interesantes del origen y la historia del arroz, la mayoría de ellos coinciden en que el arroz es originario específicamente de China. Se permiten concluir, además, que por el hallazgo de glumas de arroz en una excavación en Tailandia, el arroz se remonta a los 3000 años A.C.; lo cual parece estar confirmado por las más antiguas tradiciones y escritos.

Sin embargo, un antiguo manuscrito japonés señala que la siembra del arroz constituía una ceremonia religiosa importante 5000 años A.C. En algunos manuscritos primitivos de la India se mencionan ciertas variedades de arroz que se utilizaban como ofrendas religiosas. Similares testimonios de la antigüedad y la importancia del arroz se encuentran en la literatura primitiva de Tailandia, Birmania, Malasia e Indonesia.

Por otra parte, se ha generalizado la opinión de que el arroz no tuvo su origen en un solo país, sino en una vasta zona, como lo demuestran las

variedades silvestres de *Oryza sativa* L., que se han encontrado en la India y en el suroeste de Asia. Si bien, resulta difícil determinar el verdadero lugar de origen del arroz cultivado, no cabe duda de que su cultivo se remonta a épocas muy remotas.

Entonces, su lugar de origen concreto se desconoce, pero ciertos estudios efectuados permiten suponer que su iniciación como cultivo tuvo lugar en el sur y este de Asia, desde donde fue llevado hacia el norte. Luego parecería haberse extendido hacia el Archipiélago Malayo y desde allí hacia Indonesia. Esto naturalmente no está totalmente demostrado, y es probable que también haya sido llevado hacia otros lugares al mismo tiempo.

En virtud de las numerosas citas históricas y de los datos proporcionados por la literatura asiática, se puede afirmar que el arroz se ha cultivado desde los primeros años de vida racional del género humano. También se puede asegurar con bastante certeza que su cultivo se extendió en primer lugar hacia Asia y Oceanía y luego se fue desplazando hacia Africa y Europa.

En épocas más recientes, el arroz ha merecido la atención de los especialistas en botánica, genética, riego, mecánica, fitopatología, etc. En la mayoría de los países donde se cultiva, se efectúan estudios a fin de

obtener las variedades adecuadas al ambiente, para lo cual se analizan sus tierras, fuentes de agua, condiciones climáticas, fertilizaciones y, entre muchos otros, los problemas socioeconómicos afines. En la actualidad América ocupa el segundo lugar de producción en el cultivo de arroz (después de Asia) y su historia es netamente contemporánea; a pesar de que los primeros cultivos aparentemente pertenecen a la época colonial.

1.3 INFORMACIÓN AGRONÓMICA

1.3.1 CARACTERÍSTICAS GENERALES DEL CULTIVO

Nombre Científico: Oryza sativa L.

Ciclo vegetativo:

- Precoces: hasta 120 días
- Tardías: más de 140 días
- Tempranas: de 120 a 140 días

Variedades:

- Recomendadas por el INIAP

- INIAP 11 - INIAP 12

- INIAP 14

- INIAP 415

Condiciones ecológicas: - Lluvia: 1.200 a 2000 milímetros durante ciclo de cultivo

- Luz: por lo menos 1.000 horas de sol.

- Temperatura promedio: 25 °C.

- Textura de suelo: preferentemente pesados
con buen drenaje
- Ph. : preferiblemente neutro

El INIAP, Instituto Nacional Autónomo de Investigaciones Agropecuarias, fue creado en 1959; su objetivo principal es el de investigar, desarrollar, adaptar y aplicar el conocimiento científico y tecnológico para incrementar la competitividad de los sectores agropecuario y agroindustrial, en un marco de sostenibilidad de los recursos naturales y del ambiente.

El INIAP tiene estructurada técnicamente sus actividades de investigación agropecuaria mediante Programas y Departamentos. Dentro de los programas de este organismo se encuentra el del Arroz; el cual se encarga de mezclar variedades criollas (netas del país) con variedades extranjeras, con el fin de mejorar la calidad de este producto; de ahí obtienen las variedades que fueron anotadas anteriormente.

1.4 CLASIFICACIÓN DEL ARROZ

Las diferentes variedades de arroz que se siembran en el Ecuador, determinan que en el proceso de producción el arroz se clasifique de

acuerdo a la longitud y apariencia del arroz pilado, en los siguientes grados o tipos:

- a) Arroz de grano extra largo: corresponde a las variedades de arroz cuyos granos libres de cáscara, midan más de 7 milímetros de longitud.
- b) Arroz de grano largo: comprende las variedades de arroz, cuyos granos libres de cáscara, midan de 6 a 7 milímetros de longitud. Este tipo de arroz tiene una subclasificación:
 - Arroz de grano largo de primera: comprende todos los arroces cuya relación largo-ancho, sea igual o superior a tres.
 - Arroz de grano largo de segunda: comprende todos los arroces cuya relación largo-ancho sea igual o superior a 2.6 y su apariencia sea opaca.
- c) Arroz en grano mediano: esta clase comprende las variedades de arroz, cuyos granos libres de cáscara, midan de 5 a 6 milímetros de longitud.
- d) Arroz de grano corto: son aquellas variedades de arroz, cuyos granos libres de cáscara, midan menos de 5 milímetros de longitud.
- e) Arroz mezclado: será el arroz que tenga más del 20% (m/m) de clases contrastantes.

1.5 FITOTECNIA DEL ARROZ

1.5.1 INTRODUCCIÓN

El término fitotecnia hace referencia a conseguir los medios y formas que permitan asegurar una producción adecuada de los vegetales a las exigencias ambientales. La fitotecnia del arroz, no sólo trata de obtener mayores rendimientos, sino también de adaptar el producto a las exigencias del consumo, ya que éstas, varían según los distintos mercados.

Además, por medio de la fitotecnia se intenta lograr plantas que puedan ofrecer resistencia a sus diversos enemigos (hongos, virus, insectos, malezas, etc.) y desarrollarse en los distintos climas, altitudes, latitudes, suelos, etc., que den cosechas normales y en épocas convenientes, y provistas de tallos y panojas que soporten los vientos y los cambios de temperatura.

1.5.2 CARACTERES MORFOLÓGICOS

Los caracteres morfológicos se refieren a la descripción de las partes de la planta: tallo, hojas, clorofila, panícula, glumas, espiguillas, pistilo, grano, endosperma, desgrane, etc. (Véase figura No.1)

Planta de arroz con cinco hijos

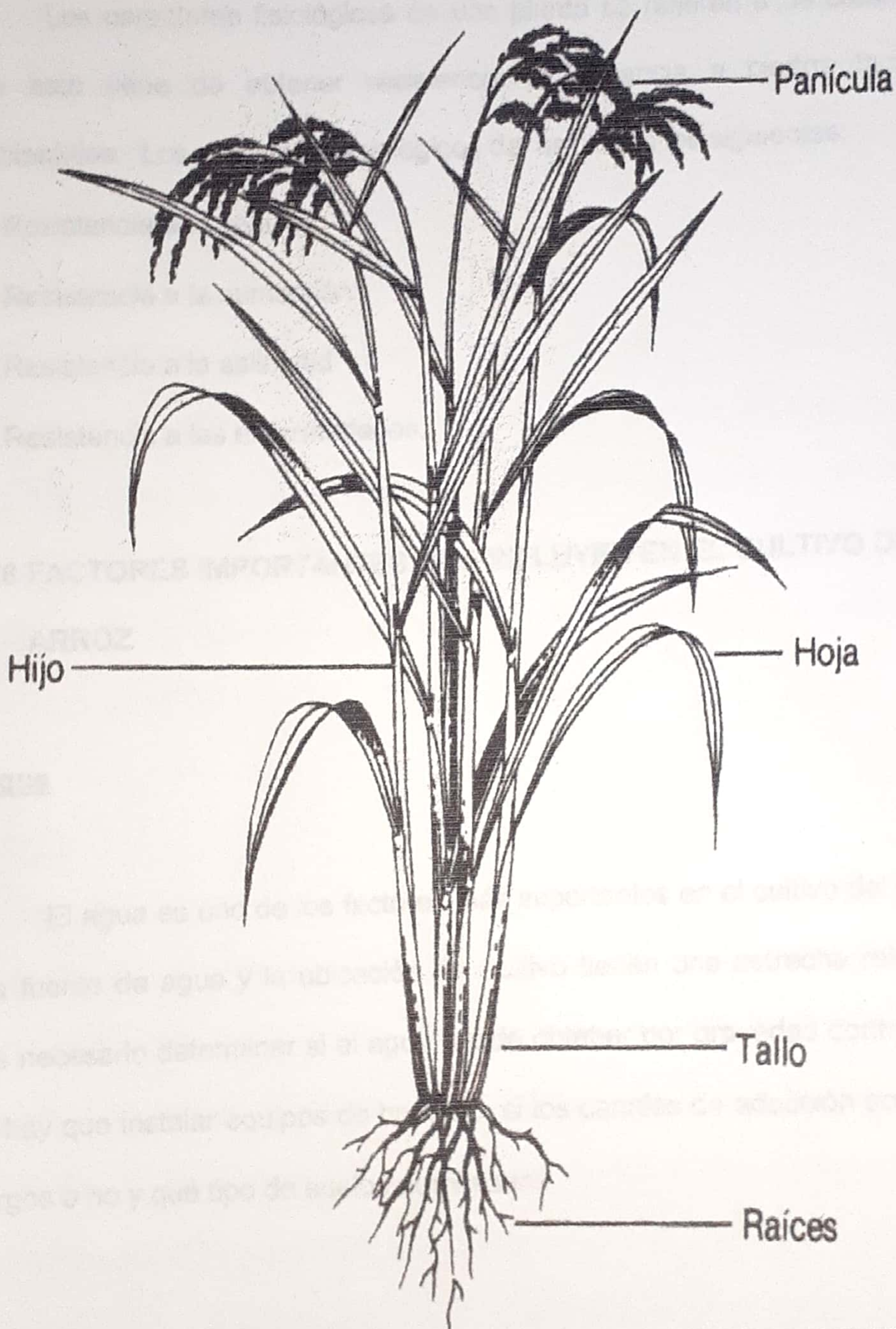


Figura 1.1

1.5.3 CARACTERES FISIOLÓGICOS

Los caracteres fisiológicos de una planta se refieren a la capacidad que esta tiene de obtener resistencia o tolerancia a ciertos factores ambientales. Los caracteres fisiológicos del arroz son los siguientes:

- * Resistencia a la sequía
- * Resistencia a la sumersión
- * Resistencia a la salinidad
- * Resistencia a las enfermedades.

1.6 FACTORES IMPORTANTES QUE INFLUYEN EN EL CULTIVO DEL ARROZ

Agua

El agua es uno de los factores más importantes en el cultivo del arroz. La fuente de agua y la ubicación del cultivo tienen una estrecha relación. Es necesario determinar si el agua puede obtener por gravedad controlada, si hay que instalar equipos de bombeo, si los canales de aducción son muy largos o no y qué tipo de suelos atraviesan.

Sobre la cantidad de agua necesaria para un cultivo, ya sea de arroz u otros, influyen las condiciones del clima, el tipo de suelo que debe ser regado, la extensión del riego, la sistematización del campo de cultivo, los drenajes existentes, la presencia o ausencia de una napa de agua y su eventual profundidad.

Los factores más importantes que gobiernan el uso del agua son: infiltración, evapotranspiración, riego preliminar, drenaje después de una siembra hecha en el agua, condiciones de humedad en el momento del riego, topografía del terreno, intervalos verticales de los bordes, altura deseada de la inundación, drenaje para la aplicación de tratamientos químicos; y, pérdida de agua que den las estructuras.

Suelo

El suelo en sí mismo no determina las posibilidades de supervivencia de la planta del arroz. Esta puede crecer tanto en suelos medios como en pesados, o sea que es relativamente indiferente a las propiedades físicas del suelo. Pero su crecimiento depende del clima, del agua y de los elementos químicos accesibles a la planta.

Drenaje

El drenaje es la evacuación de las aguas en exceso que hay en el suelo de un cultivo; está estrechamente relacionado con la mayoría de las fases del cultivo del arroz, entre ellas laboreo de la tierra, siembra, lucha contra las malezas, defensa contra las plagas y enfermedades, cosecha, etcétera. El apropiado drenaje del cultivo requiere que se consideren los siguientes aspectos: recursos adecuados de agua de riego, el retiro oportuno del exceso de agua de lluvia o de riego, el acceso de la maquinaria y de los implementos agrícolas al campo; y, los impedimentos mínimos para las operaciones agrícolas.

Debido al drenaje se producen los siguientes efectos:

- 1) Aumenta la permeabilidad del suelo.
- 2) Aumenta la porosidad
- 3) Se agrietan las partes profundas de la capa del suelo
- 4) Ocurren modificaciones favorables en los minerales de la arcilla
- 5) Aparecen organismos vivos que trabajan activamente en el suelo.

Cosecha

Cada una de las operaciones que implica el cultivo del arroz tiene su importancia, y cada una de ellas puede ser decisiva en cuanto al resultado final que se espera obtener. Para obtener un buen producto es necesario que todas las operaciones de cultivo hayan sido correctamente procesadas, debe empezarse por elegir el momento propicio para cosechar; este punto presenta ciertas dificultades.

La planta de arroz puede considerarse madura, mientras aún está en pie, cuando sus granos tienen un grado de humedad de un 25.5 ± 2.5 %. Asimismo, se puede establecer que si el arroz es cosechado con un contenido de humedad entre 21 ± 5 %, su calidad molinera puede ser superior.

Pérdidas post cosecha

En el trayecto (desde el sitio de cultivo y el lugar donde temporalmente será almacenado el grano) hasta el embodegado pueden ocasionarse pérdidas considerables de arroz, dependiendo del tipo de transportación. En el proceso de pilado la pérdida más grande se da por el

cambio de humedad de 25% al momento de la cosecha, al 13% que es lo ideal al inicio de la pilada. Se considera también importante el grado de impurezas, que debe llegar solo al 3% al momento del proceso.

De doscientas libras por saco de arroz en cáscara resulta:
 125 libras de arroz blanco (pilado),
 15 libras de arroz partido (arrocillo),
 12 libras de polvillo, que es el resultado de la fragmentación y pulida en la piladora; y, las 48 libras sobrantes son las correspondientes a la cáscara del grano de arroz.

En un cultivo semitecnificado (cultivos con controles fitosanitarios, recurso hídrico controlado y fertilización adecuada), la producción alcanza aproximadamente 16.000 libras de arroz en cáscara por hectárea; es decir 80 sacos de 200 libras por hectárea.

1.7 EL ARROZ EN EL ECUADOR

El arroz es un producto de amplio consumo en el Ecuador; se cultiva en el Litoral, principalmente en las provincias del Guayas y Los Ríos; sólo el 10% del arroz producido en el invierno, corresponde a otras provincias: Cañar, Manabí y El Oro. Nuestro país, gracias a su situación geográfica,

goza de zonas arroceras con un amplio rango en la distribución de los factores climáticos que varía desde el húmedo hasta el seco, precipitaciones máximas de 1500 ± 1000 mm por año con humedad relativa generalmente alta. Estas zonas son fértiles y su mayor obstáculo es la inadecuada disponibilidad de agua, factor que en extensas zonas es extremo llegando a provocar inundaciones y en otras zonas es mínima pues están sujetas a formas irregulares de lluvias.

Existen zonas que reciben el abastecimiento hídrico del complejo hidroeléctrico Daule – Peripa, en donde miles de hectáreas sembradas con arroz en las zonas de Palestina, Santa Lucía, Daule, Nobol, Plan América, Salitre, Samborondón son beneficiadas con este recurso. El agua es un recurso que influye decisivamente sobre las condiciones en que se desarrolla el cultivo del arroz, de allí que lo relacionado con su disponibilidad, forma de permanencia en el suelo y manejo, son variables que sirven de base para diferenciar las áreas arroceras.

El Ecuador, a pesar de poseer grandes extensiones de terreno apto y condiciones climáticas favorables para el cultivo del arroz, no tiene un rendimiento de producción, de la calidad absolutamente deseada; esto se debe al uso de variedades criollas, susceptibles a enfermedades, empleo de prácticas culturales inadecuadas, ataque de plagas y enfermedades,

excesos de agua en las zonas inundables o provisión escasa de lluvias en las zonas de secas, es decir, falta de desarrollo de una adecuada infraestructura para el control hídrico.

El arroz más que una costumbre, es una necesidad en todos los hogares a lo largo y ancho del país; ya que constituye un alto porcentaje de la dieta de los ecuatorianos. El país requiere de 30.000 TM de arroz blanco para abastecer mensualmente a la población. El arroz blanco es un alimento básico en la dieta del pueblo ecuatoriano tanto urbano como rural, el cual a pesar de no ser un cultivo originario en el país, en las últimas décadas ha incrementado su importancia dentro de los productos de consumo de primera necesidad. Tanto en Ecuador, como en el resto de países de América Latina, el arroz no es la fuente principal de alimentación; pero, es uno de los más importantes en el suministro de calorías y proteínas de origen no animal.

CAPITULO II

DETERMINACION DE VARIABLES

2.1 Introducción

En el presente capítulo, se presentan las variables de estudio; las mismas fueron determinadas debido a su directa influencia en la producción del arroz en el Ecuador. Estas variables se las ha clasificado en tres grupos, de acuerdo al tipo de influencia que tienen en la producción de esta planta, como sigue:

- a) Variables económicas: importaciones y exportaciones del arroz, índice de precio al consumidor (IPC), deuda externa, dólar (tipo de cambio), salario mínimo vital y remuneraciones.
- b) Variable climatológica: clima
- c) Variables de producción: superficie sembrada, superficie cosechada, producción y rendimiento por hectárea.

2.2 Descripción de las variables de estudio

Importaciones y Exportaciones

Importaciones: Compra de algún bien o servicio que se realiza en un mercado extranjero

Exportaciones: Venta de bienes o servicios que realiza un país a un mercado extranjero

En los años anteriores al comercio internacional del arroz en nuestro país, se vio orientado hacia la importación del producto para cubrir el déficit de la producción nacional, la cual varió considerablemente debido a los factores climatológicos y económicos.

Año	Importaciones (Tm)	Exportaciones (Tm)
1990	19219	0
1991	0	50
1992	640	12372
1993	1101	36
1994	251	31700
1995	1051	26512
1996	255	92728
1997	2844	108519
1998	99737	54154

Fuente: Boletín Mensual - Febrero 1999; Banco Central del Ecuador

Como se puede observar en la tabla 2.1, el país en el año 1991 no realizó importación de arroz, mientras que en 1990 no exportó arroz. Esto se debe a que la producción en ese año (1990) sirvió solamente para abastecer las necesidades de consumo interno; mientras que en 1991 no hubo necesidad de importar arroz, puesto que la producción alcanzó tanto para consumo interno del país como para exportar, aunque no en grandes cantidades (50 TM). Estas variables están medidas toneladas métricas; ya que lo que nos interesa es la cantidad de arroz que se importó o exportó.

Indice de Precios al Consumidor

La variación del índice de precios al consumidor es la medida de la inflación; la misma que se aprecia en el aumento persistente del nivel general de precios de los bienes y servicios de una economía con la consecuente pérdida del valor adquisitivo del ciudadano.

Se escogió el índice de precios al consumidor como variable destinada a estudio, debido a la relación con la inflación y la gran influencia que ésta tiene en el consumo de alimentos de primera necesidad, entre los cuales se encuentra el arroz; puesto que al haber mayor inflación los precios aumentan y esto repercute tanto en la demanda como en la oferta de dicho producto.

Deuda Externa

La deuda externa es un conjunto de obligaciones de un país contraídas con entidades (empresas, bancos privados, organismos internacionales, etc.) residentes en el exterior, que se derivan de las operaciones de crédito internacional.

Si bien es cierto que la deuda externa no tiene una influencia directa en la producción del arroz; no se la puede pasar por alto, ya que es un factor importante de la economía de nuestro país, y esta sí es importante en este análisis. Esta variable está medida en dólares.

Salario Mínimo Vital (SMV) y Remuneraciones

El salario mínimo vital es la cantidad mínima establecida por las leyes de un país que se le debe pagar a un empleado. Se considera el salario mínimo vital como variable de importancia en este estudio, porque de él depende la capacidad adquisitiva de las personas de nuestro país, y por ende el consumo y demanda de este producto.

La remuneración es la cantidad de dinero que recibe una persona por los servicios que presta a través de un contrato de trabajo. Es una aproximación de la cantidad real de lo que las personas de nuestro país

“ganan” u obtienen como sueldo; es decir, la remuneración es el salario más las bonificaciones, compensaciones y los conocidos como “sueldos décimos” (décimo tercero, cuarto y quinto); es por esa razón que se la incluye dentro de este estudio. Estas variables también están medidas en dólares.

Precio del Dólar

El precio del dólar es una variable de suma importancia, si bien es cierto que es muy difícil de modelar, servirá como una variable explicativa; ya que de ella dependen muchas otras variables que van a ser estudiadas; es decir, otras variables como por ejemplo las remuneraciones dependen del comportamiento que el dólar tenga en un determinado período; además tiene influencia en los precios de los insumos para el cultivo del arroz, ya que para ciertas variedades del arroz que se cultiva en nuestro país es hecha con insumos exportados.

Cabe recalcar que con la dolarización implantada en el país a partir del 10 de enero del presente año, el precio del dólar se fijó en una cantidad determinada (S/. 25.000); con lo cual se afectarán todas las variables aquí descritas, ya que estas en su mayoría dependen o interdependen de la variación del precio del dólar.

Clima

Esta variable no será analizada de manera individual, sino en forma conjunta con el resto de las variables; para medir su influencia en las mismas, especialmente en las variables de producción. Esta variable se mide en porcentaje de humedad. El clima del territorio ecuatoriano es muy variado, gracias a sus peculiares características geográficas. En la zona costera, que es la de mayor producción de arroz, el clima es cálido, con temperaturas que van de 25 a 31 grados centígrados. Durante el año hay dos estaciones: una lluviosa y una seca. La primera es caliente y húmeda por un período, aproximadamente de seis meses (de Diciembre a Mayo). La estación seca es el resto del año; las mismas que son beneficiosas para los cultivos dependiendo del tipo de cosecha que se haga.

Superficie Sembrada y Superficie Cosechada

La superficie sembrada comprende la cantidad de suelo en hectáreas ocupada por semillas o plantas destinadas a la producción agrícola en el período de referencia. Mientras que la superficie cosechada se refiere a la cantidad de suelo en hectáreas que luego de haber sido sembrada o plantada ha sido cosechada, generando productos agrícolas dentro del período de referencia. Esta superficie puede ser igual o menos a la superficie sembrada.

Estas dos variables se escogieron porque se estima una fuerte y directa relación en la producción del arroz; ya que de ellas va a depender la cantidad de arroz de se produzca en un determinado período.

Producción y Rendimiento

La producción agrícola es la cantidad total de producto primario obtenido de un cultivo; medido en toneladas métricas. Esta variable es la de mayor importancia, puesto que el objetivo de este estudio es el analizar las variaciones de la producción del arroz. El rendimiento por hectárea es la cantidad de producto (expresada en toneladas métricas) dividida entre el número de hectáreas que se produjeron.

Estas últimas cuatro variables son las más relacionadas con la producción del arroz; como se podrá comprobar en capítulos posteriores.

CAPITULO III

ANÁLISIS DE SERIES TEMPORALES RELACIONADAS CON LA PRODUCCIÓN ARROCERA EN EL PAÍS

3.1 Introducción

En este capítulo se realizará un análisis de series temporales con el fin de obtener proyecciones de cada una de las variables establecidas en el capítulo anterior. Para ello primero se hará mención a ciertos conceptos básicos de las series temporales.

3.2 Aplicación de modelos de series temporales.

Definición 3.2.1: Proceso Estocástico

Un *proceso estocástico* es una familia de variables aleatorias $\{X_t\}$, donde t es un punto en un espacio T llamado *espacio parametral*; y donde para cada $t \in T$, X_t es un punto en un espacio S llamado *espacio de estados*.

Definición 3.2.2: Serie temporal

Una *serie temporal* es un proceso estocástico; una sucesión o conjunto de observaciones generadas secuencialmente en el tiempo. Si la sucesión es continua, la serie de tiempo es llamada continua; mientras que si la sucesión es discreta, la serie de tiempo será denominada discreta.

Un modelo es una estimación de la realidad. En series temporales se utilizan ciertos modelos con el fin de aproximar o estimar procesos; los modelos pueden ser: estacionarios y no estacionarios.

Definición 3.2.3: Proceso estacionario

Un proceso estocástico X_t es estacionario si está basado en la suposición de que el proceso está en un particular estado de *equilibrio estadístico*; es decir:

$$\text{Distribución } (X_t, \dots, X_k) \equiv \text{Distribución } (X_{t+h}, \dots, X_{k+h})$$

La media de un proceso estacionario μ está dada por:

$$\mu = E[X_t] = \int_{-\infty}^{\infty} xp(x)dx;$$

y su varianza se define como:

$$\sigma_x^2 = E[(x_t - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx;$$

donde $p(x)$ es la distribución de probabilidad de X_t .

Definición 3.2.4: Proceso débilmente estacionario

Se dice que un proceso estocástico es débilmente estacionario si:

- Es de segundo orden : $E[X_t^2] < \infty$
- $E[X_t] \equiv \mu$, para todo $t \in T$.
- $\text{Cov}(X_t, X_{t+h}) \equiv g(h)$; para todo $t, k \in T$

Definición 3.2.5: Ruido blanco.

Un proceso $\{X_t\}$, se dice un ruido blanco si:

- $E[X_t] \equiv 0$, $\forall t \in T$.
- $E[X_t^2] = \text{Var}[X_t] = \sigma^2 < \infty$

$$\text{c) } \text{Cov}(X_t, X_{t+k}) = \begin{cases} 0 & \text{si } k \neq 0 \\ \sigma^2 & \text{si } k = 0 \end{cases}$$

Definición 3.2.6: Operadores utilizados

- B** es el operador llamado *operador de retroceso o cambio hacia atrás*; que se define como: $BX_t \equiv X_{t-1}$; en general:

$$B^s(X_t) = X_{t-s};$$

- ∇ es el operador de diferenciación; y puede como:

$$\nabla X_t \equiv X_t - X_{t-1} \equiv (1 - B)X_t$$

Definición 3.2.7: Autocovarianza y Autocorrelación

La covarianza entre X_t y su valor X_{t+k} , separados por k intervalos de tiempo, que bajo el supuesto de ser estacionario, debería ser el mismo para todo t , es llamada la *autocovarianza* en un intervalo de longitud k y está definida por:

$$\gamma_k = \text{COV}[X_t, X_{t+k}] = E[(X_t - \mu)(X_{t+k} - \mu)]$$

siendo: $\mu = E[X_t] = E[X_{t+k}]$, $k = 0, 1, 2, \dots$

Similarmente, la *autocorrelación* en un intervalo de longitud k es:

$$\rho_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sqrt{E[(X_t - \mu)^2]E[(X_{t+k} - \mu)^2]}}$$

$$= \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma_x^2}$$

entonces, para un proceso estacionario, la varianza $\sigma^2 = \gamma_0$ es la misma tanto para el tiempo $t+k$ como para el tiempo t . De manera que la autocorrelación en un intervalo de longitud k , esto es, la correlación entre X_t y X_{t+k} , es:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

lo que implica que $\rho_0 = 1$. Se puede demostrar que: $-1 \leq \rho_k \leq 1$.

El coeficiente de autocorrelación ρ_k mide la correlación entre dos variables X_t y X_{t+k} un tiempo k más tarde.

3.2.1 Procesos lineales estacionarios

Definición 3.2.1.1: Proceso lineal

Un proceso $\{X_t\}$, se dice lineal si puede expresarse de la siguiente manera:

$$X_t = \sum_{i=0}^{\infty} Y_i \mu_{t-i}$$

donde: Y_i es un proceso cualquiera; y

μ_t es un ruido blanco

Definición 3.2.1.2: Procesos autoregresivos (modelos AR)

$\{X_t\}$ se dice que es un proceso autoregresivo de orden p o simplemente, un proceso AR(p) si:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \mu_t$$

También puede ser representado como :

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p) X_t = \phi(B) X_t = \mu_t$$

donde:

μ_t es un ruido blanco.

Función de autocorrelación parcial

Antes de definir la función de autocorrelación parcial, es necesario deducir las *ecuaciones de Yule – Walker*.

◆ Ecuaciones de Yule-Walker

Si se sustituye $k = 1, 2, \dots, p$ en la ecuación de la función de autocorrelación:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad k > 0 \quad (3.2.1)$$

se obtiene un conjunto de ecuaciones lineales para $\phi_1, \phi_2, \dots, \phi_p$ en términos de $\rho_1, \rho_2, \dots, \rho_p$, esto es:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned} \quad j = 1, 2, \dots, k \quad (3.2.2)$$

Estas ecuaciones son las usualmente llamadas *ecuaciones de Yule – Walker*.

♦ Función de autocorrelación parcial

Inicialmente, no se conoce el orden del proceso autoregresivo que se adapte a la serie observada. La función de autocorrelación parcial es un recurso que explota el hecho de que, a pesar que un proceso AR(p) tiene una función de autocorrelación que puede tomar infinitos valores; esta puede describirse en términos de p funciones diferentes de cero de las autocorrelaciones. Se denota por ϕ_{kj} , al j -ésimo coeficiente en una representación autoregresiva de orden k , de manera que ϕ_{kk} es el último coeficiente. De la ecuación (3.2.1), el ϕ_{kj} satisface el conjunto de ecuaciones:

$$\rho_j = \phi_{k1}\rho_{j-1} + \dots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \quad j = 1, 2, \dots, k \quad (3.2.2)$$

esto conduce a las ecuaciones de Yule – Walker, las mismas que en términos matriciales pueden escribirse como:

$$\begin{bmatrix} 1 & \rho_1 & \dots & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \dots & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \dots \\ \dots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \dots \\ \rho_k \end{bmatrix}$$

En general, la cantidad ϕ_{kk} observada como una función en un intervalo de longitud k , es llamada **función de autocorrelación parcial**. Para un proceso autoregresivo de orden p , la función de autocorrelación

parcial ϕ_{kk} deberá ser *diferente de cero para k menor o igual que p y cero para k mayor que p .*

Definición 3.2.1.4: Procesos media móvil (modelos MA)

Considerando el caso en que solo se define un parámetro q de los pesos ϕ diferente de cero; se dice que $\{X_t\}$ es un proceso *media móvil de orden q* , o simplemente un MA(q), si el proceso puede ser representado de la siguiente manera:

$$X_t = \mu_t - \theta_1\mu_{t-1} - \theta_2\mu_{t-2} - \dots - \theta_q\mu_{t-q};$$

o también : $X_t = \theta(B)\mu_t;$

donde $\theta(B) = 1 - \theta_1B - \theta_2B^2 - \dots - \theta_qB^q.$

Definición 3.2.1.6: Procesos ARMA

Este modelo combina a los procesos autoregresivos y media móvil dentro de un mismo proceso y está representado por:

$$\phi(B)X_t = \theta(B)\mu_t$$

esto es llamado un proceso autoregresivo = media móvil (p,q), o simplemente un proceso ARMA(p,q).

Definición 3.2.1.7: Procesos ARIMA

Algunas series tienen un comportamiento no estacionario homogéneo; es decir, no varían alrededor de una media fija o no tienen una varianza constante durante el proceso. En particular, el nivel general de las fluctuaciones ocurridas puede ser diferente en diferentes tiempos; sin embargo, utilizando la diferenciación se las puede tratar de manera similar a las series anteriores. La diferenciación se la realiza por medio del operador autoregresivo $\phi(B)$, en el que una o más de las raíces de la ecuación $\phi(B) \equiv 0$ cae dentro del círculo unitario. En particular, si hay d raíces unitarias, el operador $\phi(B)$ puede escribirse como:

$$\phi(B)X_t = \phi(B) (1-B)^d$$

donde $\phi(B)$ es el operador estacionario. De manera que un modelo puede representar un comportamiento homogéneo no estacionario es de la forma:

$$\phi(B)w_t = \theta(B) \mu_t \quad (3.2.3)$$

donde $w_t = \nabla^d X_t$ (3.2.4)

El comportamiento homogéneo no estacionario puede ser algunas veces representado por un modelo que "diferencie" d veces el proceso para hacerlo estacionario. El proceso definido por las ecuaciones (3.2.3) y (3.2.4) provee un poderoso modelo para describir series de tiempo

estacionarias y no estacionarias y es llamado un **proceso autoregresivo integrado media móvil (ARIMA) de orden (p,d,q)** . El proceso es definido por:

$$w_t \equiv \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \mu_t - \theta_1 \mu_{t-1} - \dots - \theta_q \mu_{t-q};$$

donde : μ_t es un ruido blanco; y $w_t = \nabla^d X_t$.

3.3. Análisis de regresión

En muchas situaciones, es necesario describir la relación existente entre dos o más variables. La idea principal de la regresión es explicar una variable cuantitativa Y , en términos de valores de una o más variables también cuantitativas X_1, X_2, \dots, X_{p-1} .

Formalmente, si tenemos la distribución conjunta de dos variables aleatorias X y Y y se sabe que X toma el valor de x , el problema básico de **regresión bivariada** es determinar la media condicional $\mu_{Y/x}$; es decir, el valor "promedio" de Y para el valor de X dado. En problemas en los cuales intervienen más de dos variables; es decir, en **regresión múltiple**, nos interesan correspondientes cantidades como $\mu_{Z/x,y}$, la media del valor de Z para valores de X y Y dados.

Si $f(x,y)$ es el valor de la densidad conjunta de dos variables aleatorias X y Y en (x,y) entonces el problema de la regresión bivariada es,

simplemente, determinar la densidad condicional de Y dado $X = x$ y después la evaluación de la integral:

$$\mu_{Y/x} = E(Y/x) = \int_{-\infty}^{\infty} y \cdot w(y/x) dy$$

donde: $w(y/x) = \frac{f(x,y)}{g(x)}$; y $g(x)$ es la densidad marginal de X

La ecuación resultante recibe el nombre de *ecuación de regresión de Y en X* :

$$\mu_{X/y} = E(X/y) = \int_{-\infty}^{\infty} x \cdot f(x/y) dx$$

Ejemplo 3.1: Dadas las variables aleatorias X y Y que tienen la densidad conjunta

$$f(x,y) = \begin{cases} xe^{-x(1+y)} & \text{para } x > 0 \text{ y } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

obtenga la ecuación de regresión de Y en X .

Solución:

Al integrar y se puede obtener que la densidad marginal de X está dada por:

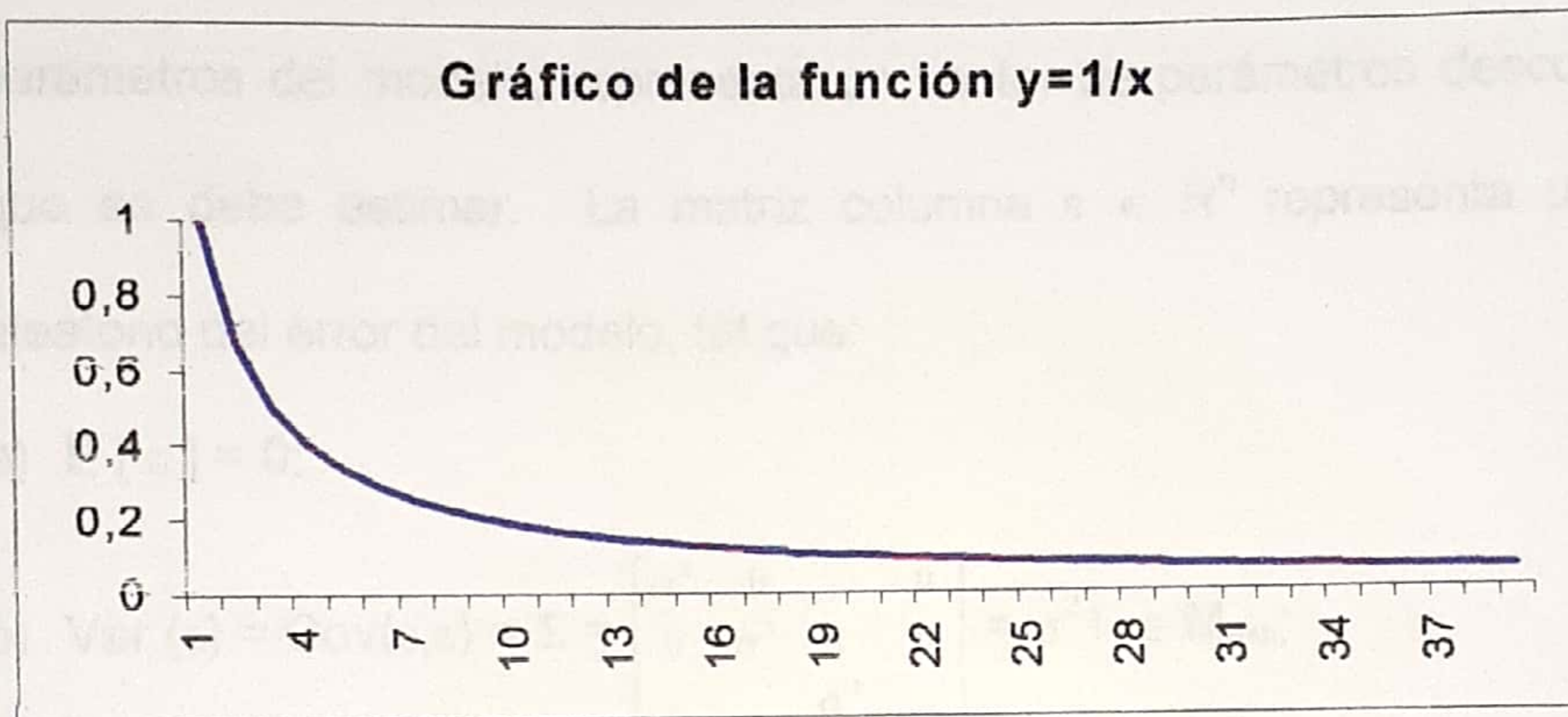
$$g(x) = \begin{cases} e^{-x} & \text{para } x > 0 \\ \text{en cualquier otra parte} \end{cases}$$

y por consiguiente, la densidad condicional de Y dado $X = x$ está dada por:

$$w(y/x) = \frac{f(x, y)}{g(x)} = \frac{x e^{-x(1+y)}}{e^{-x}} = x e^{-xy}$$

para $y > 0$ y $w(y/x) > 0$ en cualquier otra parte. Luego evaluamos la ecuación de regresión; dándonos como resultado:

$$\mu_{Y/X} = \int_0^{\infty} y \cdot x \cdot e^{-xy} dy = \frac{1}{x}$$



3.3.1 Regresión Lineal

Si $E[Y/X=x] = \beta_0 + \beta x$ se tiene un modelo de regresión lineal simple. Los modelos de regresión lineal son aquellos en los que se supone que Y está compuesta por una media, que depende en una manera continua de las X y de un error aleatorio ϵ , que explica el error de medida y el efecto de otras variables que no se especifican en el modelo. El modelo general en regresión lineal es:

$$Y = X\beta + \varepsilon$$

La matriz columna $Y \in \mathbb{R}^n$ representa un vector de las observaciones de la variable a ser explicada. La matriz $X \in M_{n \times p}$ denominada *matriz de diseño*, representa una matriz con los valores de las variables de explicación; y donde n es el número de observaciones y p representa el número de parámetros a estimar. La matriz columna $\beta \in \mathbb{R}^p$, donde p es el número de parámetros del modelo; representa un vector de parámetros desconocidos, que se debe estimar. La matriz columna $\varepsilon \in \mathbb{R}^n$ representa un vector aleatorio del error del modelo, tal que:

a) $E[\varepsilon] = 0$;

b) $\text{Var}(\varepsilon) = \text{Cov}(\varepsilon, \varepsilon) = \Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdot & 0 \\ 0 & \sigma^2 & \cdot & 0 \\ \cdot & \cdot & \sigma^2 & \cdot \\ 0 & 0 & \cdot & \sigma^2 \end{bmatrix} = \sigma^2 I \in M_{n \times n}$;

Σ es una matriz real simétrica, y por tanto diagonalizable ortogonalmente; y,

c) $\text{Cov}[\varepsilon_i, \varepsilon_j] = 0$, para todo $i \neq j$

3.3.2 Regresión Lineal Múltiple

Aunque en muchos casos una variable puede predecirse con bastante precisión en términos de otra, parece razonable que las predicciones deban mejorar si adicionalmente se considera información relevante. Este es el caso de la regresión lineal múltiple; es decir, si existen dos o más variables

de explicación, se tiene un modelo de regresión múltiple. El modelo de regresión lineal múltiple es el siguiente:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1}; i = 1, \dots, n \quad (3.3.2.1)$$

donde p es el número de parámetros a estimar; si por ejemplo se tienen dos variables de explicación X_1, X_2 , el modelo sería:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i;$$

donde el término $\beta_3 x_{i1} x_{i2}$ contiene la interacción entre X_1 y X_2 ; y la matriz de diseño sería:

$$\begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} \end{bmatrix}$$

Se puede demostrar que:

$$E[y_i] = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1}$$

Uno de los métodos más utilizados en la estimación de parámetros de un modelo de regresión lineal, es el método de mínimos cuadrados.

Supongamos el modelo (3.3.2.1), donde:

- y_i es el valor observado,
- $\varepsilon_i \sim N(0, \sigma^2)$
- $\text{cov}(\varepsilon_i, \varepsilon_j) = 0; \quad i \neq j$

El valor estimado de y_i se denota por \hat{y}_i . Como no se conocen los parámetros, entonces se emplearán las estimaciones de los parámetros para obtener el valor de \hat{y}_i . Esto es

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{p-1} x_{i,p-1} = b_0 + b_1 x_{i1} + \dots + b_{p-1} x_{i,p-1}$$

La diferencia entre el valor observado y_i y el valor estimado \hat{y}_i , es igual al estimador del error $e_i = \hat{\varepsilon}_i$ correspondiente a la i -ésima observación, es decir

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

Ahora, se requiere de una función diferenciable, de tal manera que el error total de la estimación, sea mínimo. Si minimizamos la suma del error, obtendríamos estimaciones con un error pequeño en valor relativo, pero de magnitud grande. Si minimizamos la suma de los valores absolutos del error, entonces la función resultante no será diferenciable.

Se define la suma cuadrática del error como la suma de los estimadores del error elevados al cuadrado. Se denota por Q , y es una medida del error total de la estimación.

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_{p-1} x_{i,p-1})^2$$

El método de mínimos cuadrados consiste en establecer el valor de los β_i que minimicen la suma cuadrática del error. Esto se hace derivando parcialmente la suma cuadrática del error con respecto a cada estimador, e igualando cada derivada a cero. Luego

$$\frac{\partial Q}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_{p-1} x_{i,p-1}) = 0$$

$$\frac{\partial Q}{\partial \hat{\beta}_j} = -2 \sum_{i=1}^n x_{ij} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_{p-1} x_{i,p-1}) = 0, \quad j = 1, \dots, p-1$$

Este es un conjunto de p ecuaciones con p incógnitas. La resolución de este sistema de ecuaciones, daría como resultado la estimación de los parámetros β_i por el método de mínimos cuadrados.

3.3.2.1 Regresión lineal múltiple (notación matricial)

Para expresar las ecuaciones normales en notación matricial se definirán las tres matrices siguientes:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p-1} \\ 1 & x_{21} & & x_{2p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & & x_{np-1} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

Y además:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

La matriz \mathbf{X} es la *matriz de diseño* del modelo lineal general, previamente definido, mientras que la matriz $\boldsymbol{\beta}$ es el *vector de parámetros* del modelo, $\boldsymbol{\varepsilon}$ es el *vector de residuos* o *vector de error* del modelo; y, \mathbf{Y} es la matriz con los valores observados de y .

Mediante el uso de estas matrices, se puede anotar un importante resultado:

Las estimaciones de mínimos cuadrados de los coeficientes de regresión múltiple están dadas por:

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

donde \mathbf{X}^T es la transpuesta de \mathbf{X} y $(\mathbf{X}^T \mathbf{X})^{-1}$ es la inversa de $\mathbf{X}^T \mathbf{X}$.

Ejemplo 3.2:

Los siguientes datos son ficticios; muestran el número de hectáreas de superficie sembrada, el número de hectáreas de superficie cosechada y los precios a los que se vendió en 1999 el quintal de arroz:

Número de hectáreas Superficie sembrada	Número de hectáreas de Superficie cosechada	Precio
x_1	x_2	y
3	2	78800
2	1	74300
4	3	83800
2	1	74200
3	2	79700
2	2	74900
5	3	88400
4	2	82900

Encuentre una ecuación lineal que nos permita predecir el precio de la venta promedio de un quintal de arroz dado en términos del número de hectáreas de superficie sembrada y el número de hectáreas de superficie cosechada.

Solución:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 2 \\ 1 & 2 & 1 \\ 1 & 4 & 3 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \\ 1 & 5 & 3 \\ 1 & 4 & 2 \end{pmatrix} \quad \mathbf{X}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 4 & 2 & 3 & 2 & 5 & 4 \\ 2 & 1 & 3 & 1 & 2 & 2 & 3 & 2 \end{pmatrix} \quad \mathbf{X}^T \mathbf{X} = \begin{pmatrix} 8 & 25 & 16 \\ 25 & 87 & 55 \\ 16 & 55 & 36 \end{pmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{84} \begin{pmatrix} 107 & -20 & -17 \\ -20 & 32 & -40 \\ -17 & -40 & 71 \end{pmatrix} \quad \mathbf{X}^T \mathbf{Y} = \begin{pmatrix} 637000 \\ 2031100 \\ 1297700 \end{pmatrix}$$

Y por último:

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \frac{1}{84} \begin{pmatrix} 107 & -20 & -17 \\ -20 & 32 & -40 \\ -17 & -40 & 71 \end{pmatrix} \begin{pmatrix} 637000 \\ 2031100 \\ 1297700 \end{pmatrix}$$

$$= \frac{1}{84} \begin{pmatrix} 5476100 \\ 374200 \\ 63700 \end{pmatrix}$$

$$= \begin{pmatrix} 65191.7 \\ 4133.3 \\ 758.3 \end{pmatrix}$$

De manera que la ecuación sería:

$$\hat{Y}_i = 65191.7 + 4133.3 x_{i1} + 758.3 x_{i2} ; \quad i = 1, 2, \dots, 8$$

	IPC	Deuda pública	Salario	Salario mínimo	Dolar tipo de cambio		
IPC	0.712	0.152	1				
Deuda pública	0.774	0.546	0.634	1			
Salario	0.005	-0.31	-0.214	-0.215	1		
Salario mínimo	0.392	0.616	0.193	0.675	-0.432	1	
Dolar	0.38	0.762	0.293	0.248	-0.471	0.808	1

Tabla 2.4

3.4 Análisis de las variables de estudio mediante series temporales

A continuación se presentan las series, los modelos y el análisis en general (mediante series temporales) de las diferentes variables de estudio, definidas en el capítulo anterior. Analizaremos el primer grupo de variables:

- ◆ Importaciones y exportaciones del arroz
- ◆ IPC (Inflación)
- ◆ Deuda externa
- ◆ Salario mínimo vital
- ◆ Remuneraciones
- ◆ Dólar (tipo de cambio)

Primero se calculará la matriz de correlación entre estas variables, y se analizará cada serie para decidir si se la modela por regresión o por series temporales. La matriz de correlación de estas variables es:

	Matriz de Correlación						
	Importaciones del arroz	Exportaciones del arroz	IPC	Deuda Externa	Salarios	Remun.	Dólar
Importaciones	1						
Exportaciones	0.049	1					
IPC	0.265	0.069	1				
Deuda Externa	0.256	0.546	0.639	1			
Salarios	0.008	-0.31	-0.214	-0.215	1		
Remuneraciones	0.392	0.816	0.193	0.675	-0.432	1	
Dólar	0.38	0.767	0.291	0.748	-0.471	0.985	1

Tabla 3.1

Para saber si modelamos estas series mediante regresión o series temporales; primero se analizan los valores de las series y la matriz de correlación, para ver con que variables están altamente correlacionadas y así poder establecer un modelo que se ajusta a la realidad. Las variables que tienen correlaciones más altas son las más propicias a modelarse mediante análisis de regresión; por ejemplo la variable remuneración con el dólar. A continuación se procederá a analizar cada una de las variables:

Exportaciones del arroz

En la tabla 3.1 podemos observar que esta variable tiene una alta correlación con las variables: remuneración, el tipo de cambio del dólar y la deuda externa. La serie de las exportaciones del arroz está modelada a través de series de tiempo; cuyo gráfico es el que se muestra en la figura 3.1. Para saber los parámetros del modelo o de los modelos que mejor ajusten a esta serie, primero se calculan y grafican las autocorrelaciones y autocorrelaciones parciales. El gráfico que muestra la Figura 3.3 es el de correlaciones parciales de las exportaciones del arroz; de aquí se puede hallar el valor del parámetro p para el modelo ARMA (p,q).

Realizando la prueba de parámetros respectiva para ver que modelo ajusta mejor a esta serie, se escogieron dos modelos:

- ARMA (1,0) y

Ecuador: Exportación anual de arroz a partir del 1970

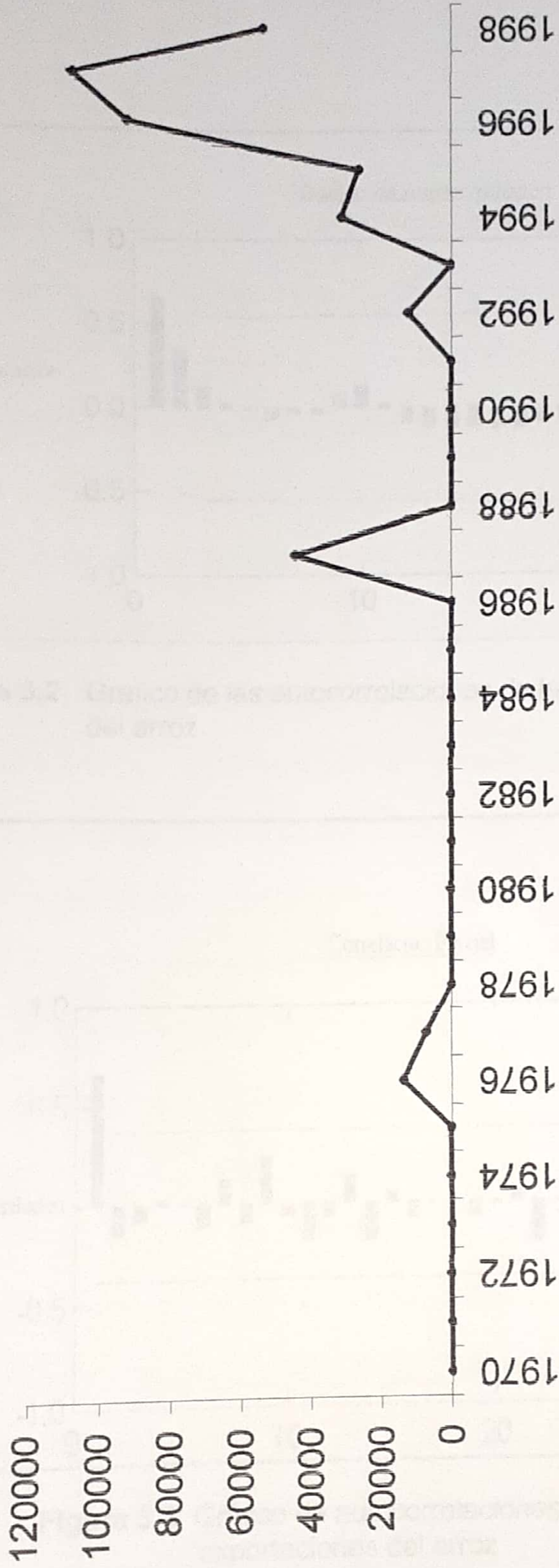


Figura 3.1 Gráfico de la serie de las exportaciones del arroz

Fuente: Banco Central del Ecuador, Boletín anual de 1975, 1980, 1985, 1990, 1995, 1997
Ver datos en el Apéndice B

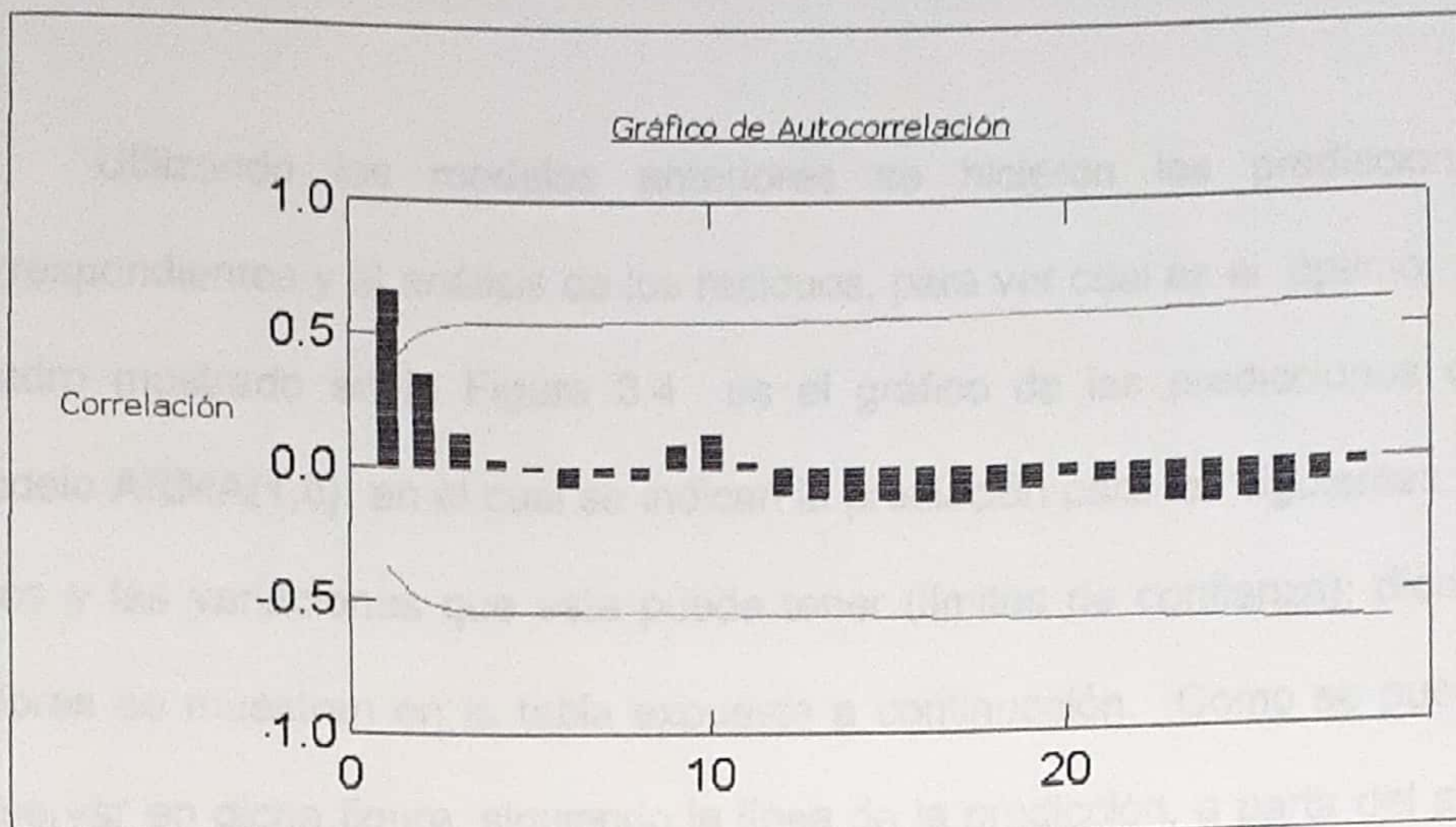


Figura 3.2 Gráfico de las autocorrelaciones de las exportaciones del arroz

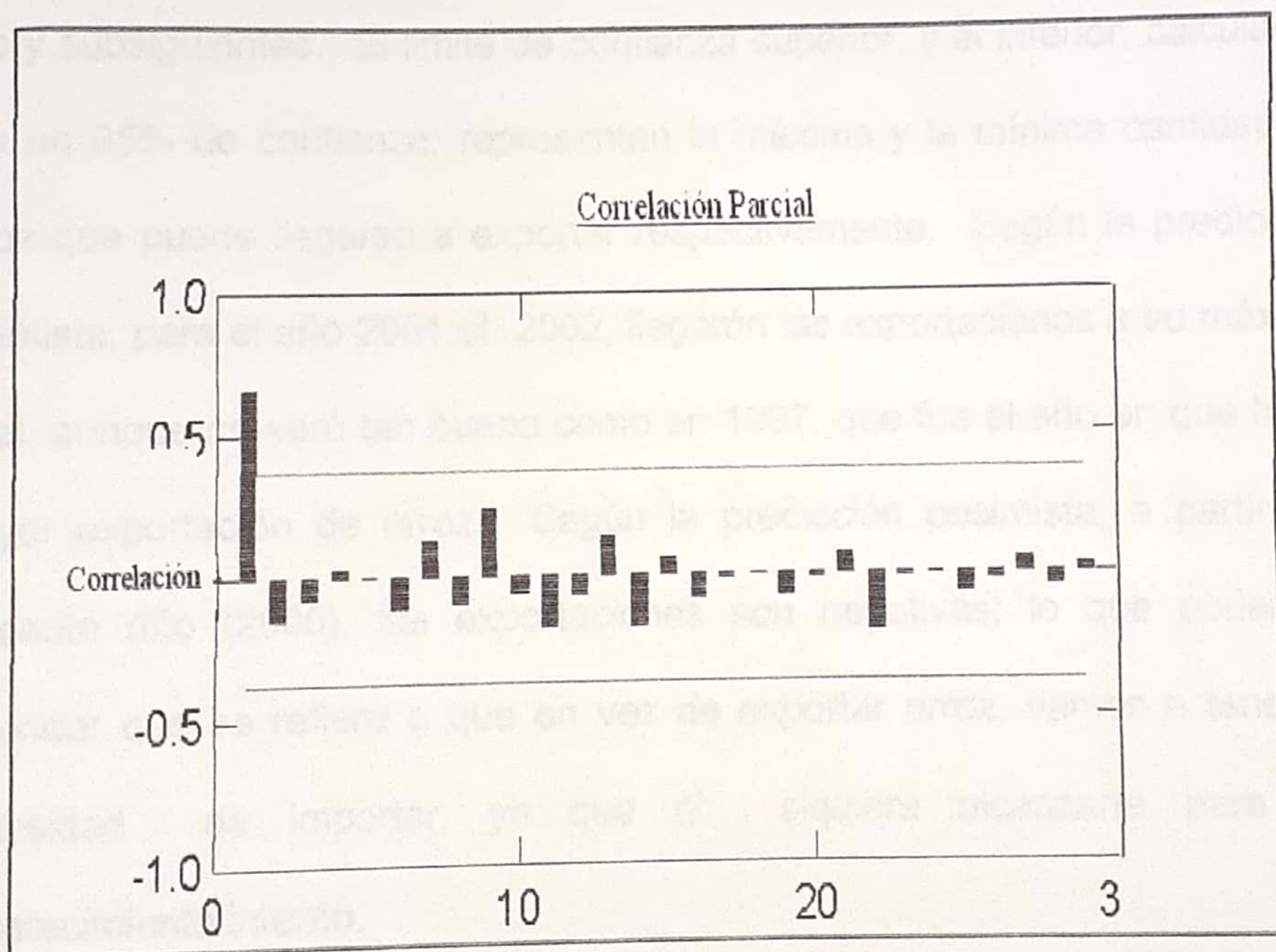


Figura 3.3 Gráfico de autocorrelaciones parciales de las exportaciones del arroz

- ARMA (0,1)

Utilizando los modelos anteriores se hicieron las predicciones correspondientes y el análisis de los residuos, para ver cual es el óptimo. El cuadro mostrado en la Figura 3.4 es el gráfico de las predicciones del modelo **ARMA(1,0)**; en el cual se indican la predicción para los siguientes 20 años y las variaciones que esta puede tener (límites de confianza); dichos valores se muestran en la tabla expuesta a continuación. Como se puede observar en dicha figura, siguiendo la línea de la predicción, a partir del año de 1997, las exportaciones del arroz empezaron a disminuir, y según este modelo de predicción continuará con la misma tendencia para el presente año y subsiguientes. El límite de confianza superior, y el inferior, calculados con un 95% de confianza; representan la máxima y la mínima cantidad de arroz que puede llegarse a exportar respectivamente. Según la predicción optimista, para el año 2001 al 2002, llegarán las exportaciones a su máximo nivel, aunque no será tan buena como en 1997, que fue el año en que hubo mayor exportación de arroz. Según la predicción pesimista, a partir del presente año (2000), las exportaciones son negativas; lo que podemos entender que se refiere a que en vez de exportar arroz, vamos a tener la necesidad de importar, ya que ni siquiera alcanzaría para un abastecimiento interno.

Proyección de la exportación del arroz del 2000 – 2018
Modelo ARMA (1,0)

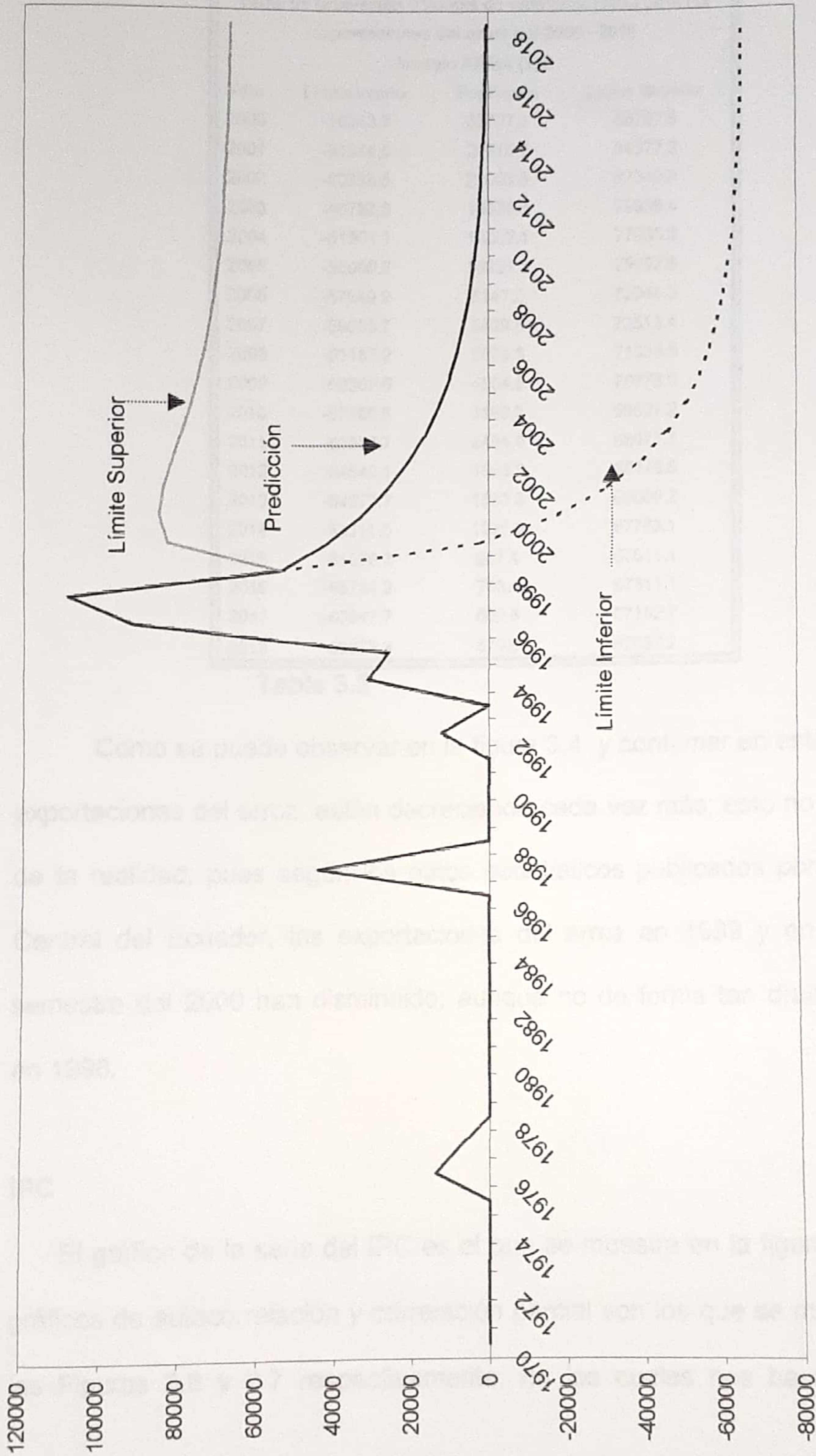


Figura 3.4 Gráfico de la predicción de las exportaciones del arroz

Tabla de predicción y límites de confianza (95%) para las exportaciones del arroz del 2000 - 2018			
Modelo ARMA (1,0)			
Año	Límite Inferior	Predicción	Límite Superior
2000	-18343.7	33727.1	85797.9
2001	-31344.0	26616.7	84577.3
2002	-40338.5	21005.3	82349.0
2003	-46782.6	16576.9	79936.4
2004	-51501.1	13082.1	77665.3
2005	-55009.6	10324.1	75657.8
2006	-57649.2	8147.5	73944.3
2007	-59653.7	6429.9	72513.4
2008	-61187.2	5074.3	71335.8
2009	-62367.6	4004.5	70376.6
2010	-63280.6	3160.3	69601.2
2011	-63989.7	2494.0	68977.7
2012	-64542.1	1968.2	68478.6
2013	-64973.7	1553.3	68080.2
2014	-65311.5	1225.8	67763.1
2015	-65576.3	967.4	67511.1
2016	-65784.3	763.4	67311.1
2017	-65947.7	602.5	67152.7
2018	-66076.3	475.5	67027.2

Tabla 3.2

Como se puede observar en la figura 3.4 y confirmar en esta tabla las exportaciones del arroz, están decreciendo cada vez más; esto no está lejos de la realidad; pues según los datos estadísticos publicados por el Banco Central del Ecuador, las exportaciones del arroz en 1999 y en el primer semestre del 2000 han disminuido; aunque no de forma tan drástica como en 1998.

IPC

El gráfico de la serie del IPC es el que se muestra en la figura 3.5. Los gráficos de autocorrelación y correlación parcial son los que se muestran en las Figuras 3.6 y 3.7 respectivamente; en los cuales nos basamos para

Ecuador: IPC

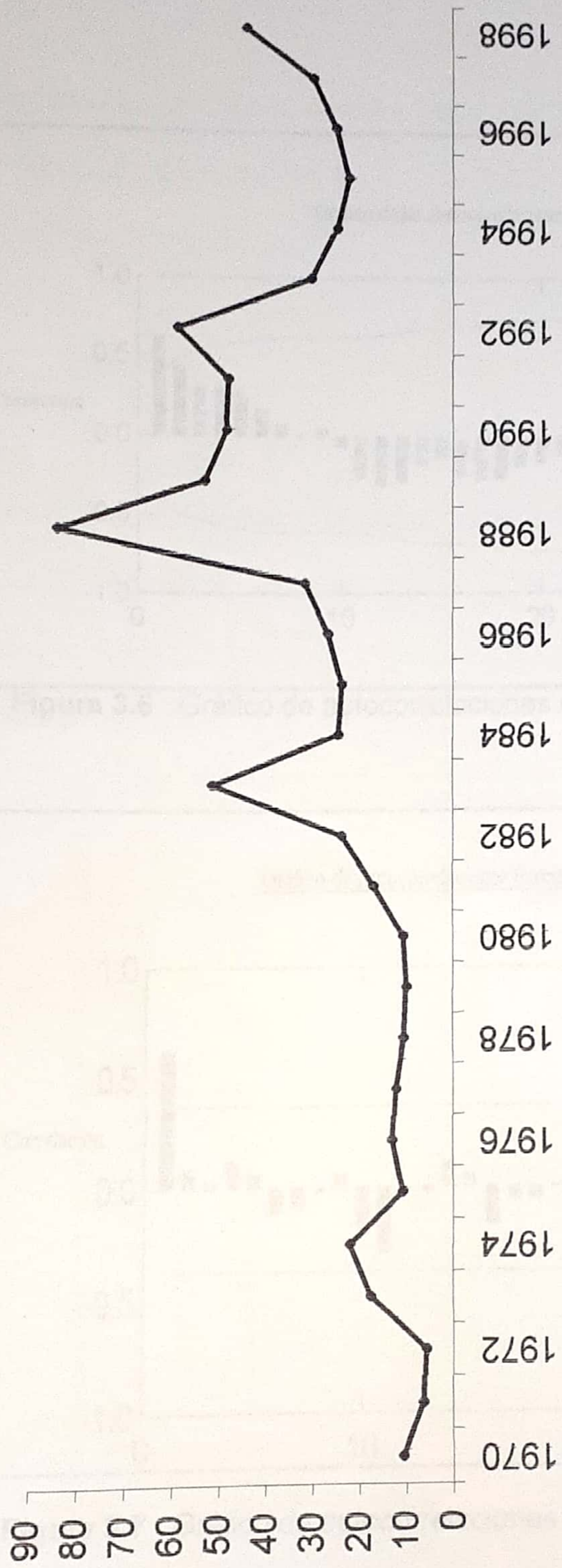


Figura 3.5 Gráfico de la serie del IPC

Fuente: INEC, Boletín anual de 1970 – 1998
Ver datos en el Apéndice B

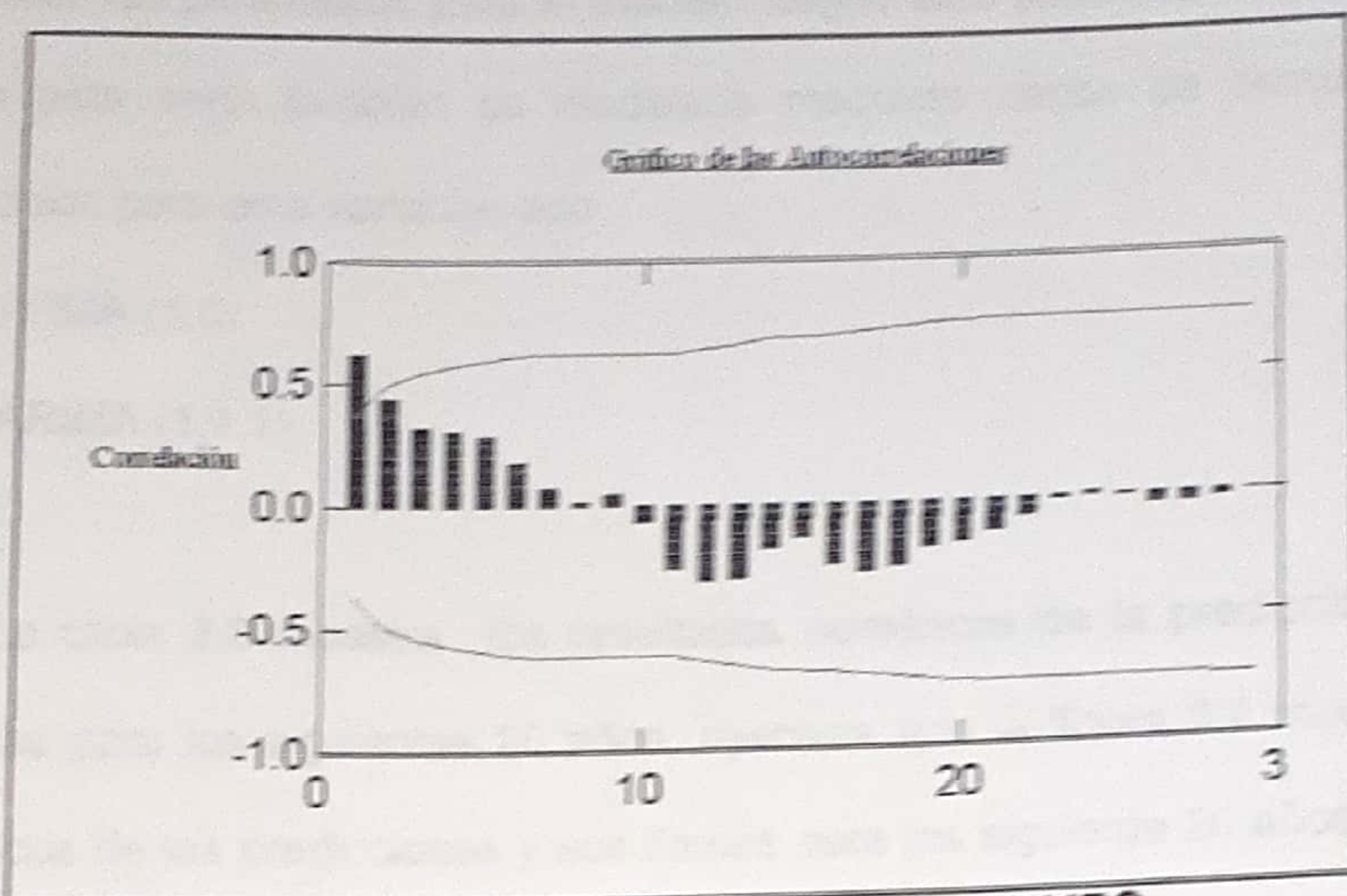


Figura 3.6 Gráfico de autocorrelaciones del IPC

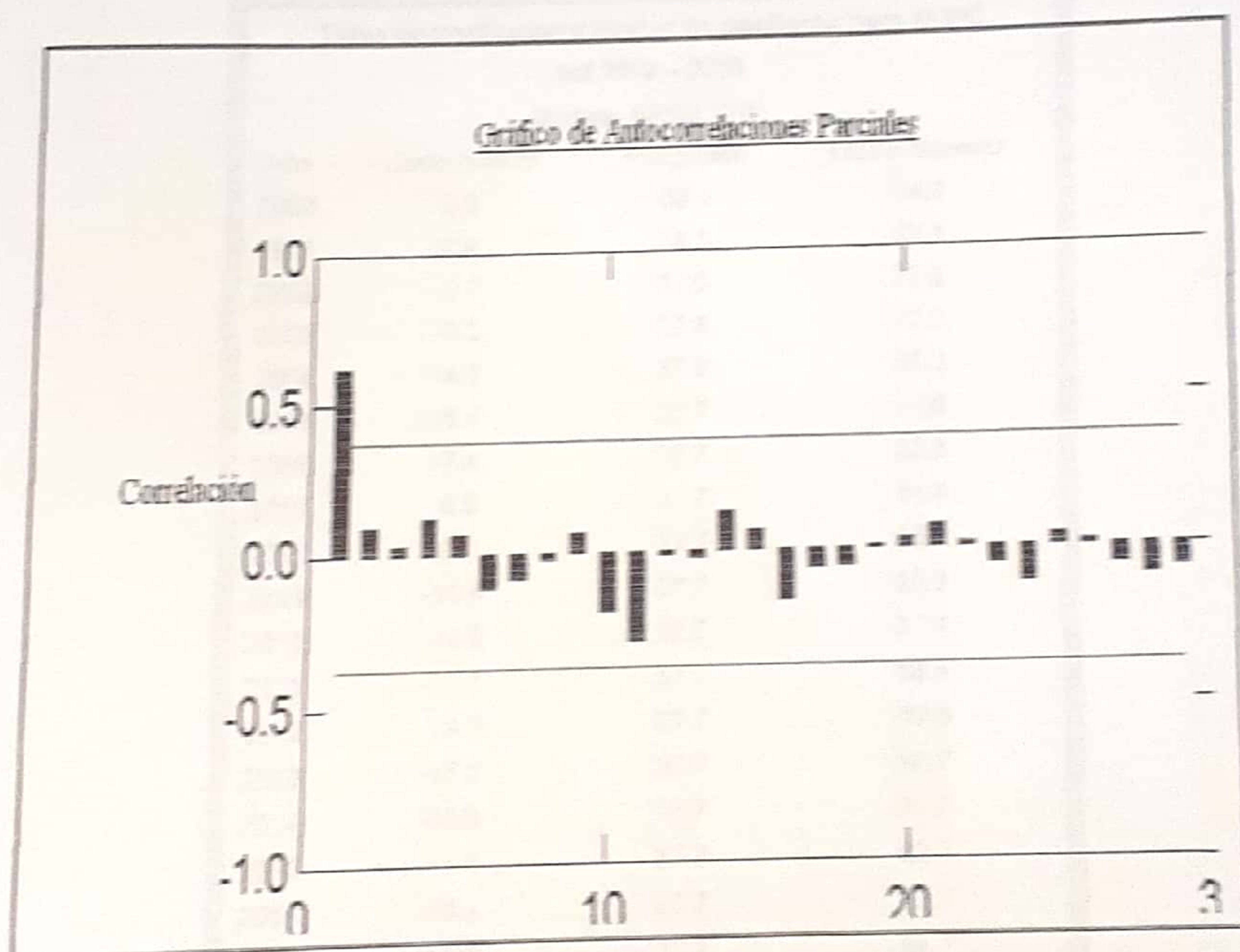


Figura 3.7 Gráfico de autocorrelaciones parciales del IPC

estimar los parámetros para el modelo; según esto podemos darnos cuenta que esta serie también es modelada mediante series de tiempo. Los modelos para esta variables son:

- ARMA (1,0)
- ARIMA (1,1,1)

La tabla 3.3 muestra los resultados numéricos de la predicción y sus límites para los siguientes 20 años; mientras que la figura 3.8 muestra los gráficos de las predicciones y sus límites para los siguiente 20 años, con un 95% confianza.

Tabla de predicción y límites de confianza para el IPC del 2000 – 2018 Modelo ARMA (1,0)			
Año	Límite Inferior	Predicción	Límite Superior
2000	3.8	39.0	74.2
2001	0.4	38.3	76.1
2002	-1.7	37.9	77.6
2003	-3.3	37.8	79.0
2004	-4.8	37.8	80.3
2005	-6.1	37.7	81.6
2006	-7.4	37.7	82.8
2007	-8.6	37.7	84.0
2008	-9.7	37.7	85.2
2009	-10.9	37.7	86.3
2010	-12.0	37.7	87.4
2011	-13.1	37.7	88.5
2012	-14.2	37.7	89.6
2013	-15.2	37.7	90.7
2014	-16.3	37.7	91.7
2015	-17.3	37.7	92.7
2016	-18.3	37.7	93.7
2017	-19.2	37.7	94.7
2018	-20.2	37.7	95.6
2019	-21.1	37.7	96.6
2020	-22.1	37.7	97.5

Tabla 3-3

Predicción de la inflación

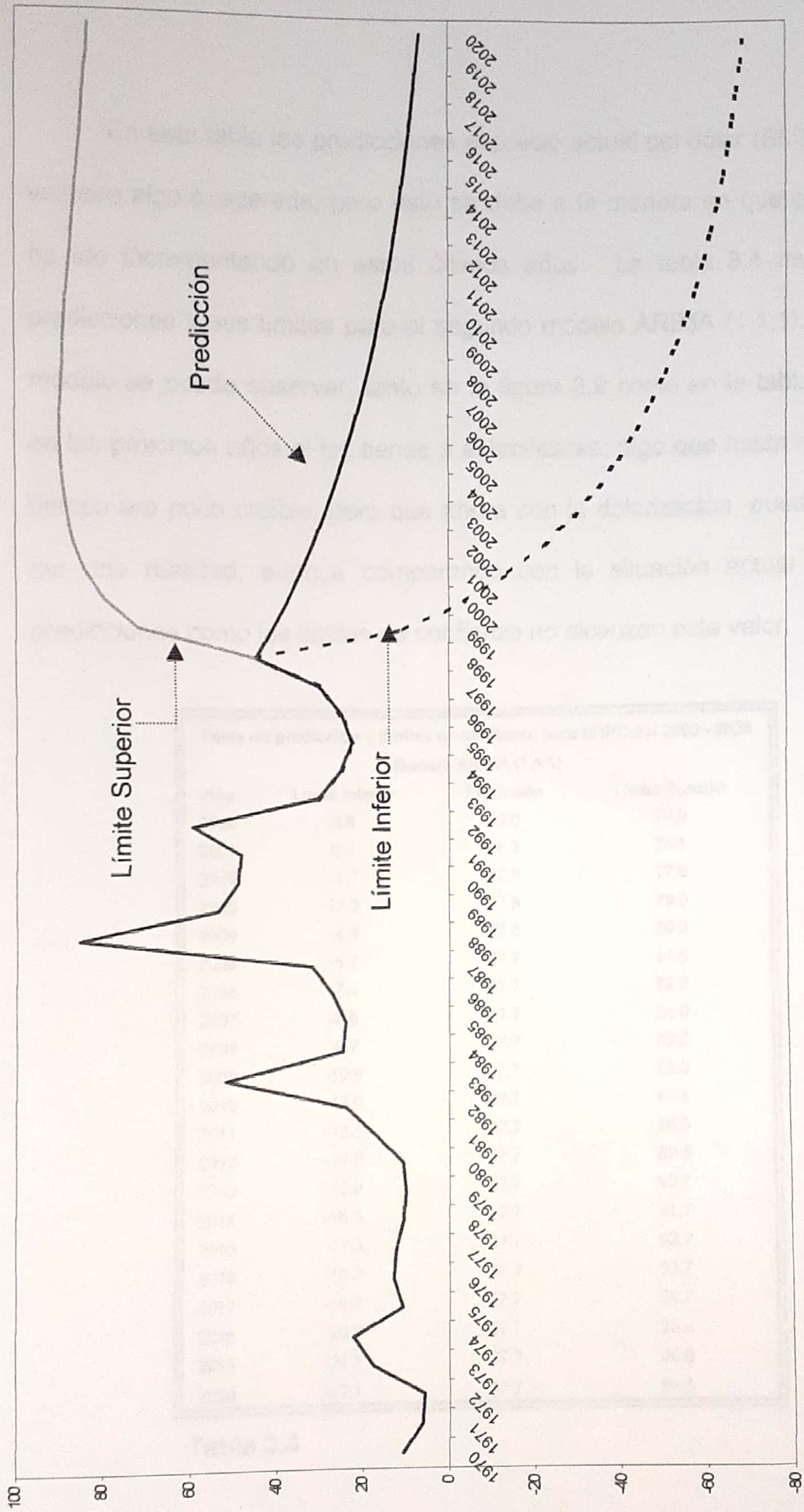


Figura 3.8 Gráfico de las predicciones para la inflación

En esta tabla las predicciones al precio actual del dólar (S/. 25000) tal vez sea algo exagerada; pero esto se debe a la manera en que el dólar se ha ido incrementando en estos últimos años. La tabla 3.4 muestra las predicciones y sus límites para el segundo modelo ARIMA (1,1,1). En este modelo se puede observar, tanto en la figura 3.9 como en la tabla 3.4, que en los próximos años el ipc tiende a estabilizarse; algo que hasta hace poco tiempo era poco creíble, pero que ahora con la dolarización puede llegar a ser una realidad; aunque comparando con la situación actual tanto las predicciones como los límites de confianza no alcanzan este valor.

Tabla de predicción y límites de confianza para el IPC del 2000 - 2020			
Modelo ARIMA (1,1,1)			
Año	Límite Inferior	Predicción	Límite Superior
2000	3.8	39.0	74.2
2001	0.4	38.3	76.1
2002	-1.7	37.9	77.6
2003	-3.3	37.8	79.0
2004	-4.8	37.8	80.3
2005	-6.1	37.7	81.6
2006	-7.4	37.7	82.8
2007	-8.6	37.7	84.0
2008	-9.7	37.7	85.2
2009	-10.9	37.7	86.3
2010	-12.0	37.7	87.4
2011	-13.1	37.7	88.5
2012	-14.2	37.7	89.6
2013	-15.2	37.7	90.7
2014	-16.3	37.7	91.7
2015	-17.3	37.7	92.7
2016	-18.3	37.7	93.7
2017	-19.2	37.7	94.7
2018	-20.2	37.7	95.6
2019	-21.1	37.7	96.6
2020	-22.1	37.7	97.5

Tabla 3.4

Predicción de la inflación según el modelo ARIMA (1,1,1)

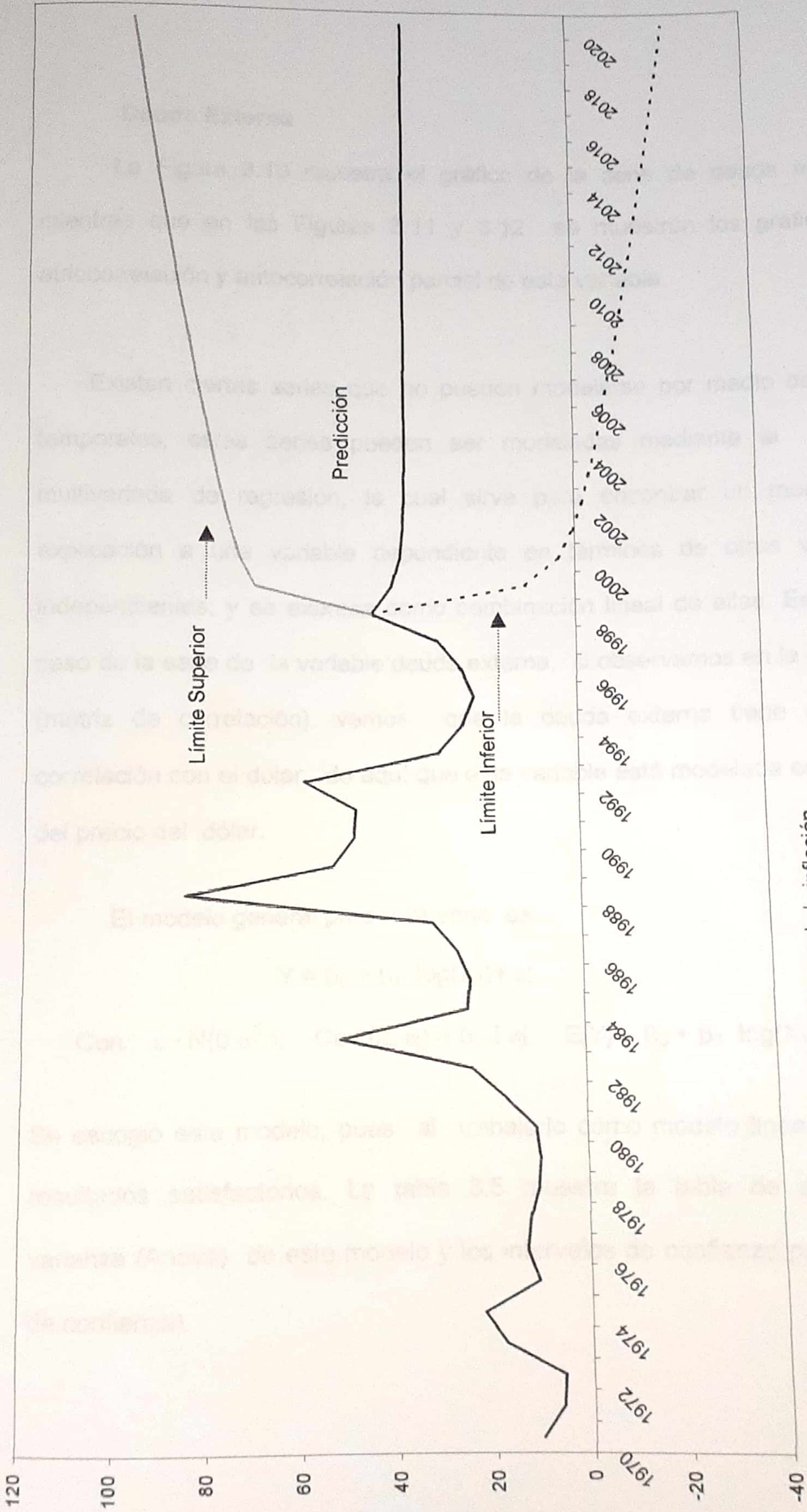


Gráfico de las predicciones de la inflación

Figura 3.9

Deuda Externa

La Figura 3.10 muestra el gráfico de la serie de deuda externa; mientras que en las Figuras 3.11 y 3.12 se muestran los gráficos de autocorrelación y autocorrelación parcial de esta variable.

Existen ciertas series que no pueden modelarse por medio de series temporales; estas series pueden ser modeladas mediante la técnica multivariada de regresión; la cual sirve para encontrar un modelo de explicación a una variable dependiente en términos de otras variables independientes; y se expresa como combinación lineal de ellas. Este es el caso de la serie de la variable deuda externa; si observamos en la tabla 3.1 (matriz de correlación), vemos que la deuda externa tiene una alta correlación con el dólar, de aquí que esta variable está modelada en función del precio del dólar.

El modelo general para esta serie es:

$$Y = \beta_0 + \beta_1 \log(X_1) + \varepsilon_i$$

$$\text{Con } \varepsilon_i \sim N(0, \sigma^2); \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j; \quad E[Y] = \beta_0 + \beta_1 \log(X_1)$$

Se escogió este modelo, pues al trabajarlo como modelo lineal, no daba resultados satisfactorios. La tabla 3.5 muestra la tabla de análisis de varianza (Anova) de este modelo y los intervalos de confianza para β (95% de confianza).

Serie de la Deuda Externa (1970 - 1998)

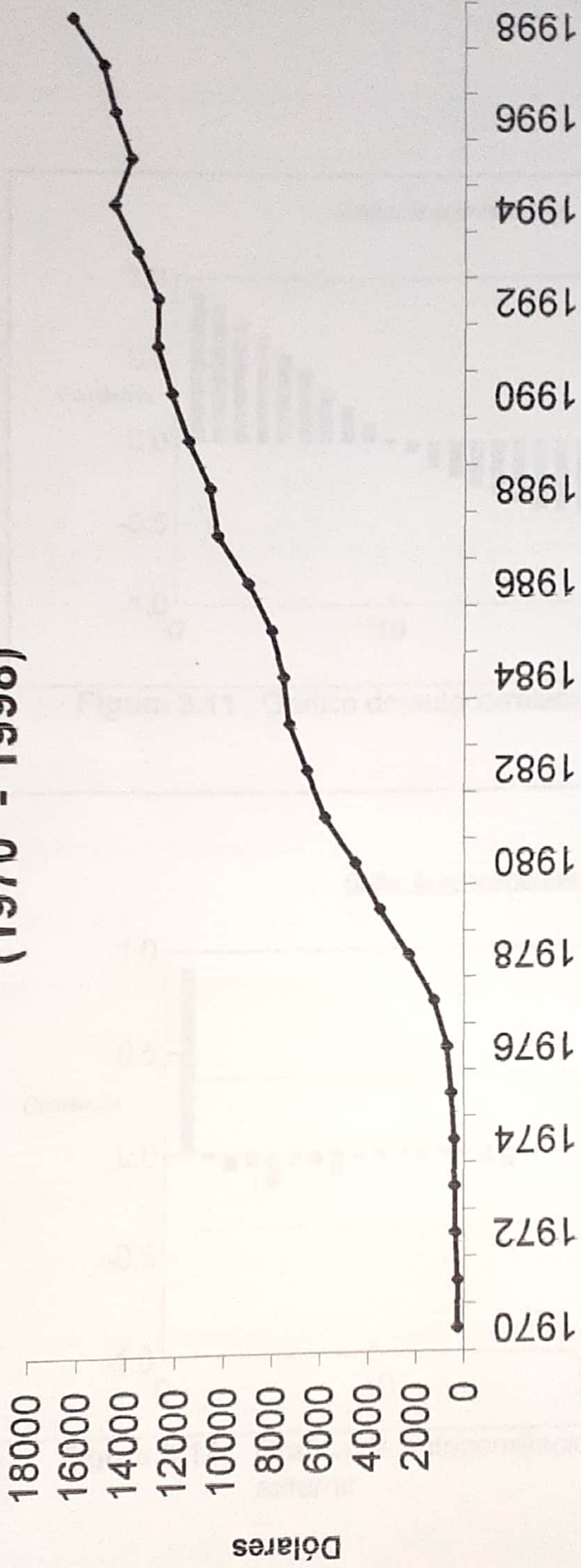


Figura 3.10 Gráfico de la serie de Deuda Externa en el Ecuador

Fuente: Banco Central del Ecuador, Boletín anual 1975 - 1998
Ver datos en el Apéndice B

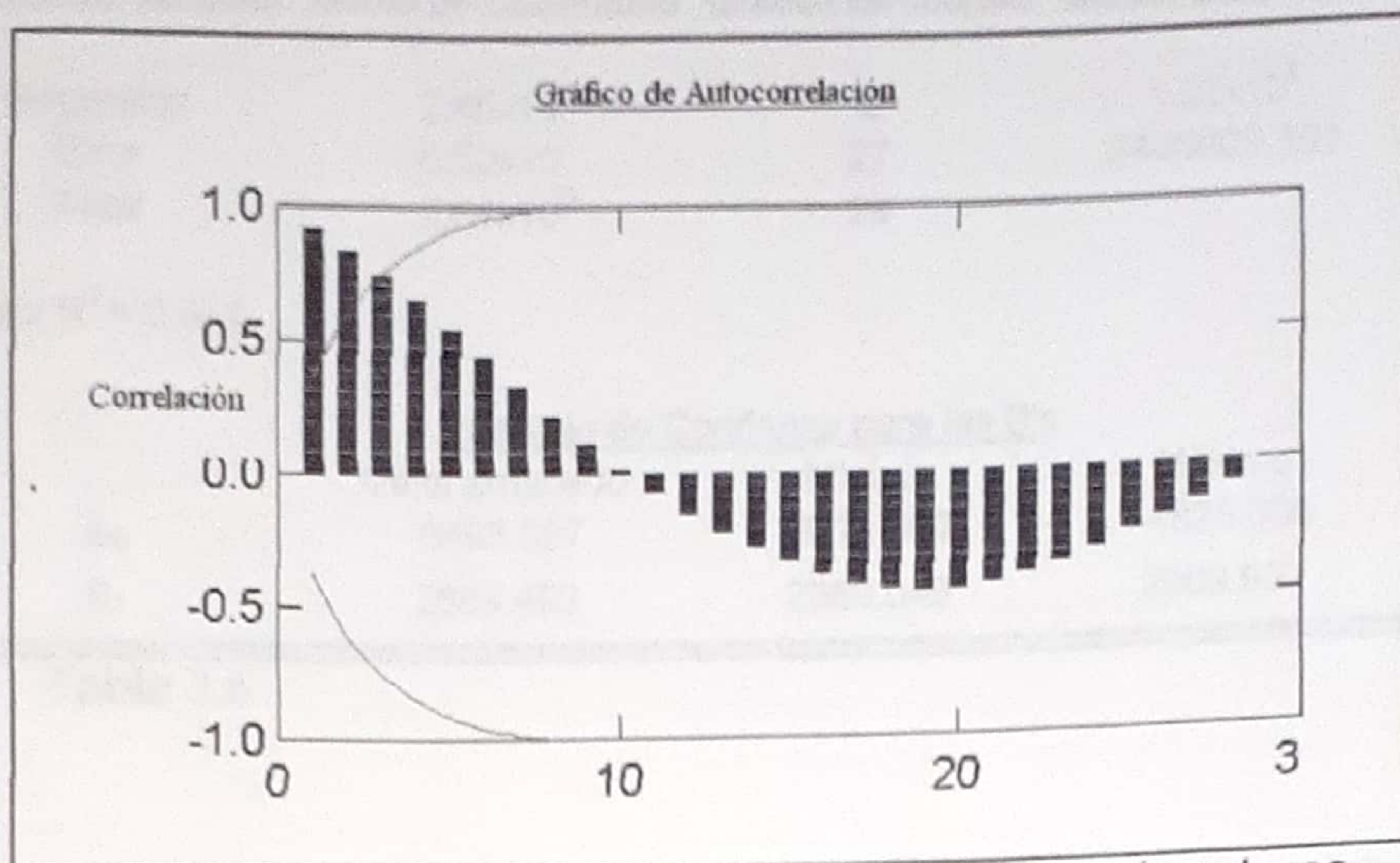


Figura 3.11 Gráfico de autocorrelación de deuda externa

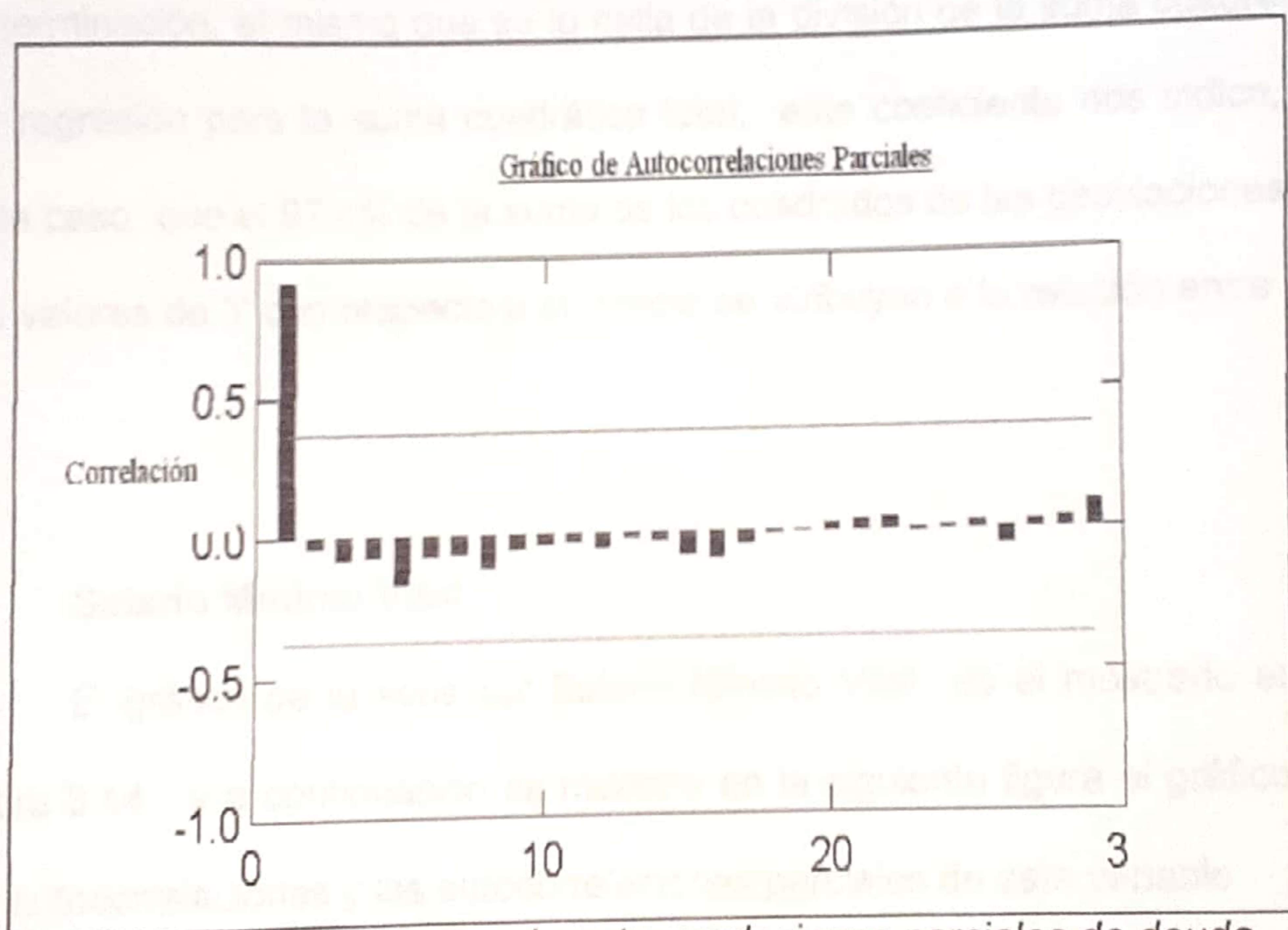


Figura 3.12 Gráfico de autocorrelaciones parciales de deuda externa

<u>Tabla Anova</u>				
Fuente de variación	Suma de Cuadrados	Grados de libertad	Media Cuadrática	Valor F
Regresión	2.45×10^9	2	1.22×10^9	5.08×10^2
Error	6.50×10^7	27	2408609.191	
Total	2.51×10^9	29		
Valor $R^2 = 0.974$				
<u>Intervalo de Confianza para las B's</u>				
	Valor estimado	Mínimo	Máximo	
B_0	-6498.597	-8171.607	-4825.586	
B_1	2689.493	2389.049	2989.937	

Tabla 3.5

En esta tabla se puede observar el valor R^2 que es el coeficiente de determinación, el mismo que se lo halla de la división de la suma cuadrática de regresión para la suma cuadrática total; este coeficiente nos indica, en este caso, que el 97.4% de la suma de los cuadrados de las desviaciones de los valores de Y con respecto a su media se atribuyen a la relación entre Y y X.

Salario Mínimo Vital

El gráfico de la serie del Salario Mínimo Vital es el mostrado en la figura 3.14 y a continuación se muestra en la siguiente figura el gráfico de las autocorrelaciones y las autocorrelaciones parciales de esta variable.

Realizando el análisis de la serie y apoyándonos en estos gráficos, podemos tomar como modelo un ARMA, cuyos parámetros serían (1,0). De

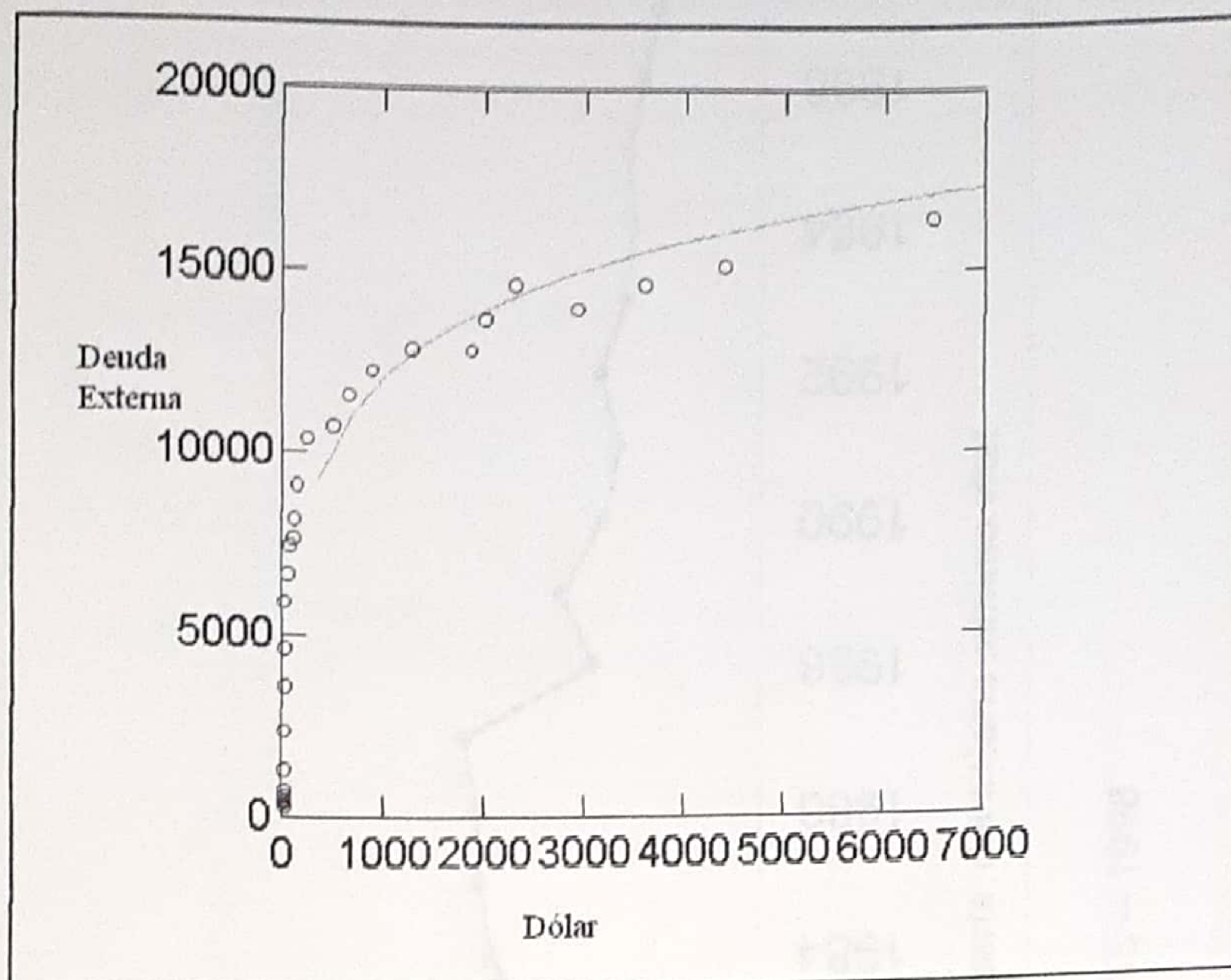


Figura 3.13 Gráfico de regresión, modelo:
 $Y = \beta_0 + \beta_1 \log (X_1) + \varepsilon$

Ecuador: Salario mínimo vital en dólares desde 1970 - 1998

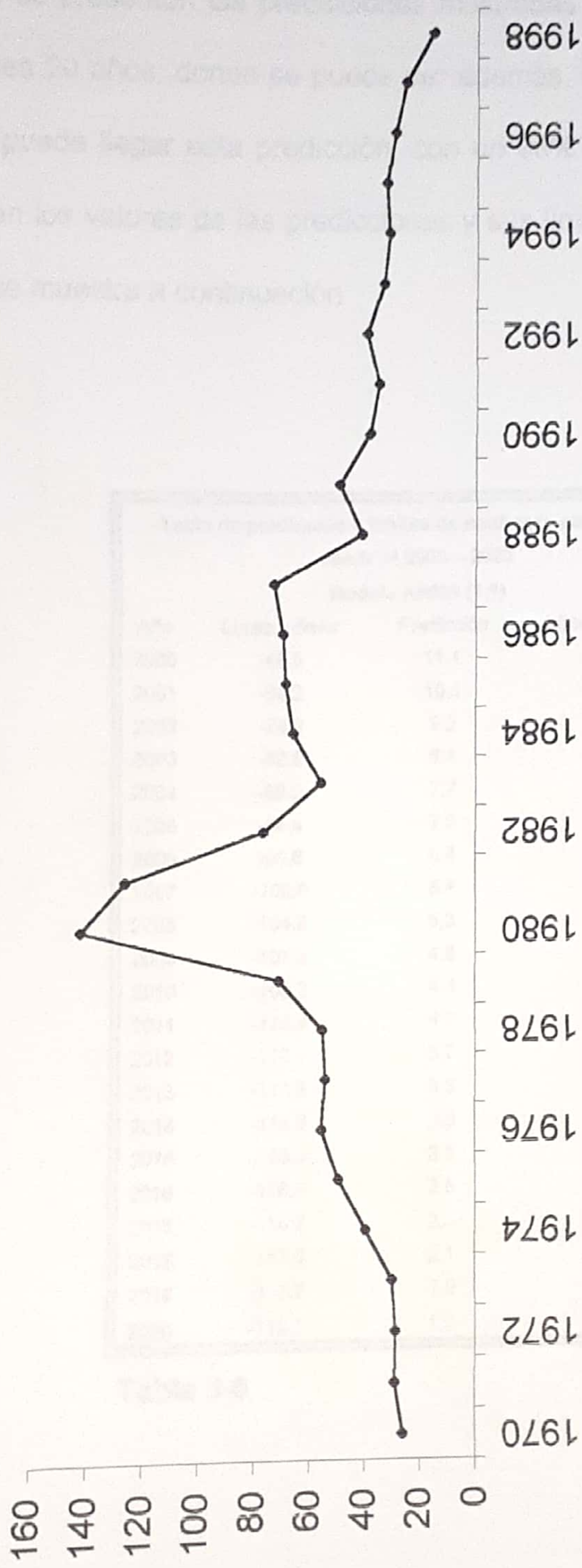


Figura 3.14 Gráfico de la serie del salario mínimo vital

Fuente: Banco Central del Ecuador, Boletín anual 1975 - 1998
Ver datos en el Apéndice B

ahí que se presenten las predicciones mostradas en la figura 3.17 para los siguientes 20 años; donde se puede ver además, el límite máximo y mínimo al cual puede llegar esta predicción, con un 95% de confianza. Además se muestran los valores de las predicciones y sus límites en la tabla 3.6 que es la que se muestra a continuación.

Tabla de predicción y límites de confianza para el SMV desde el 2000 – 2020 Modelo ARMA (1,1)			
Año	Límite Inferior	Predicción	Límite Superior
2000	-49.5	11.1	71.6
2001	-64.2	10.1	84.4
2002	-74.8	9.2	93.2
2003	-82.9	8.4	99.7
2004	-89.3	7.7	104.6
2005	-94.4	7.0	108.4
2006	-98.6	6.4	111.3
2007	-102.0	5.8	113.6
2008	-104.8	5.3	115.4
2009	-107.2	4.8	116.9
2010	-109.2	4.4	118.0
2011	-110.9	4.0	118.9
2012	-112.3	3.7	119.6
2013	-113.5	3.3	120.1
2014	-114.5	3.0	120.5
2015	-115.3	2.8	120.9
2016	-116.1	2.5	121.1
2017	-116.7	2.3	121.3
2018	-117.2	2.1	121.4
2019	-117.7	1.9	121.5
2020	-118.1	1.7	121.6

Tabla 3-6

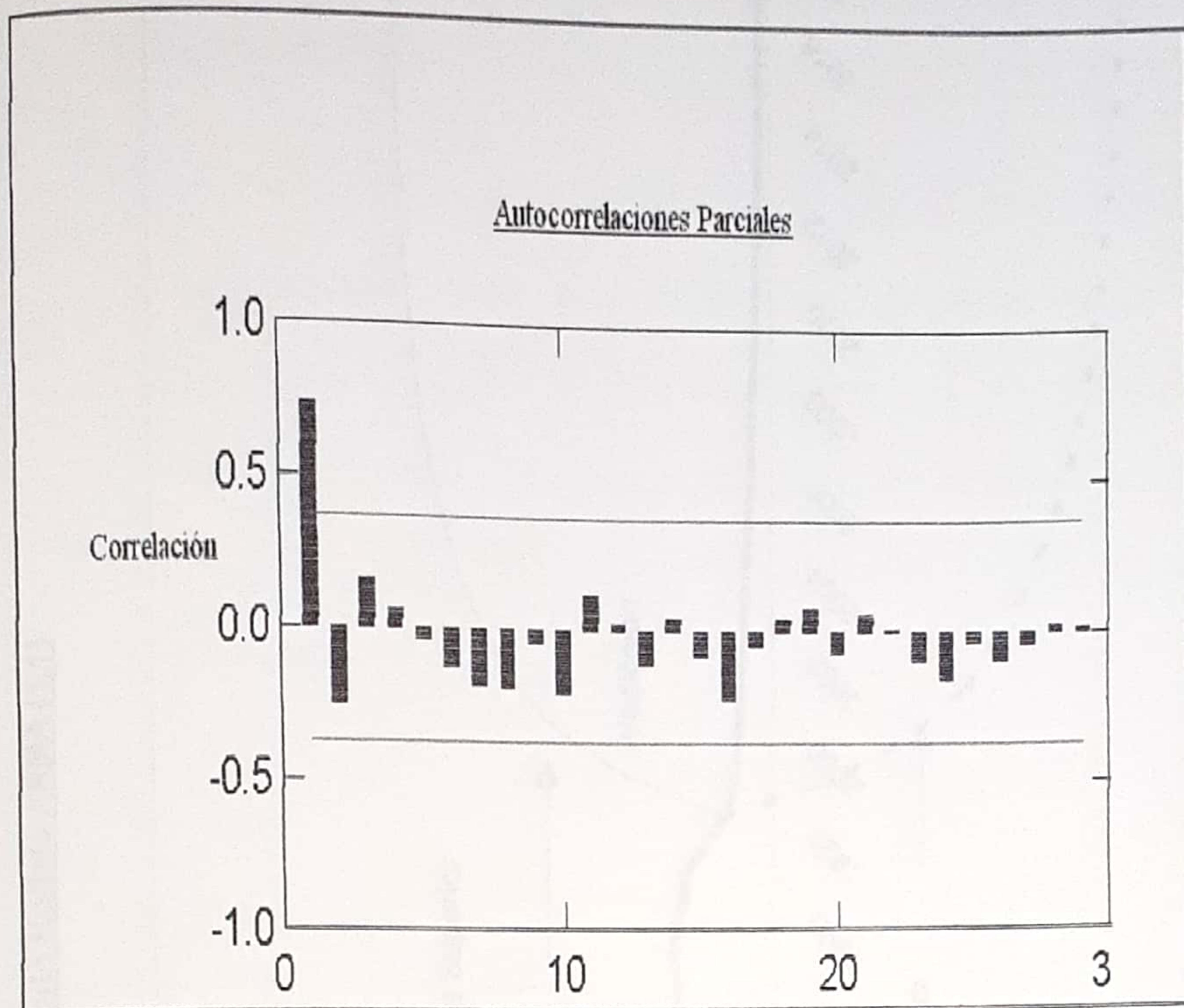


Figura 3.15 Gráfico de autocorrelación del salario

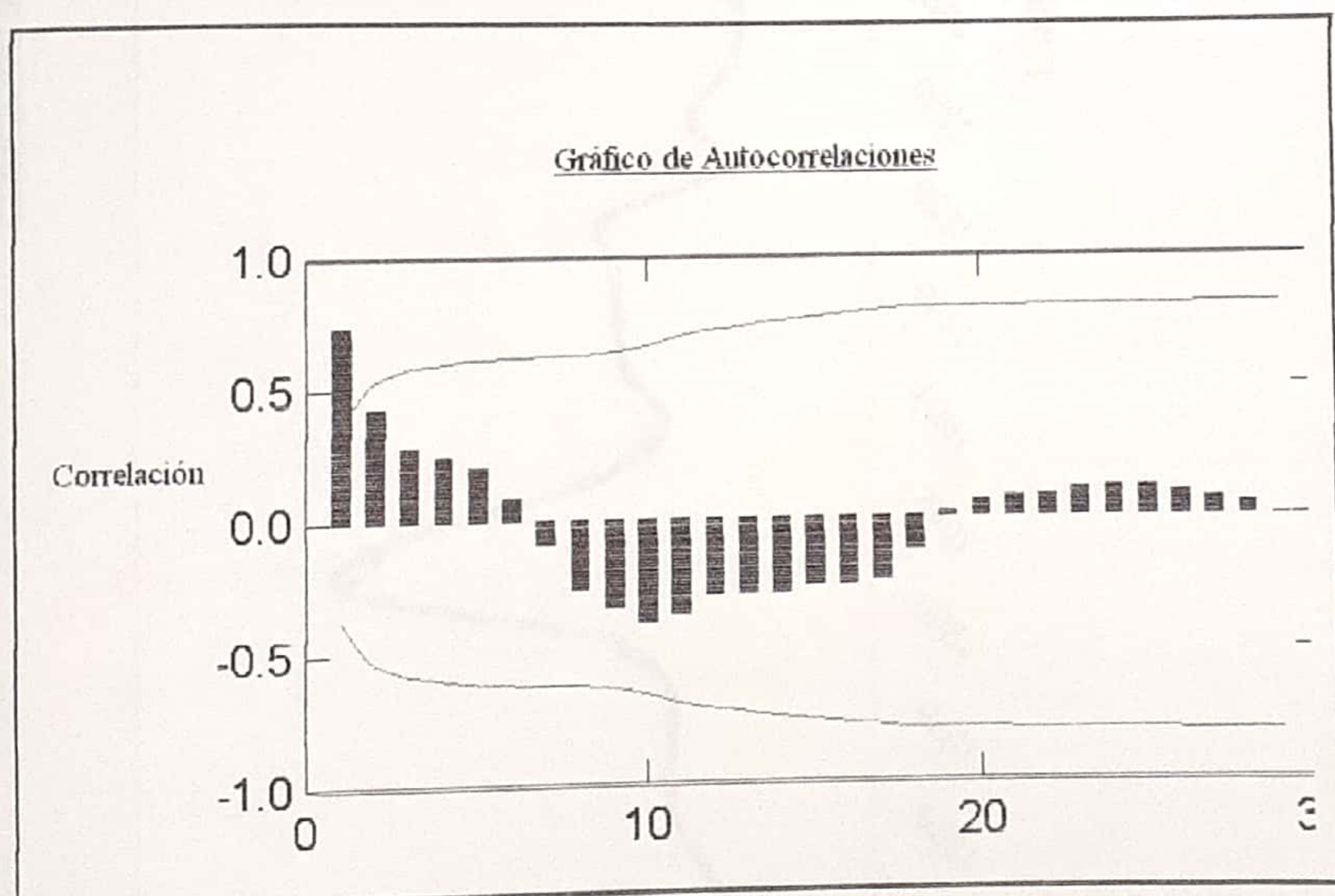


Figura 3.16 Gráfico de autocorrelaciones parciales del salario

Predicción del Salario Modelo ARMA (1,1)

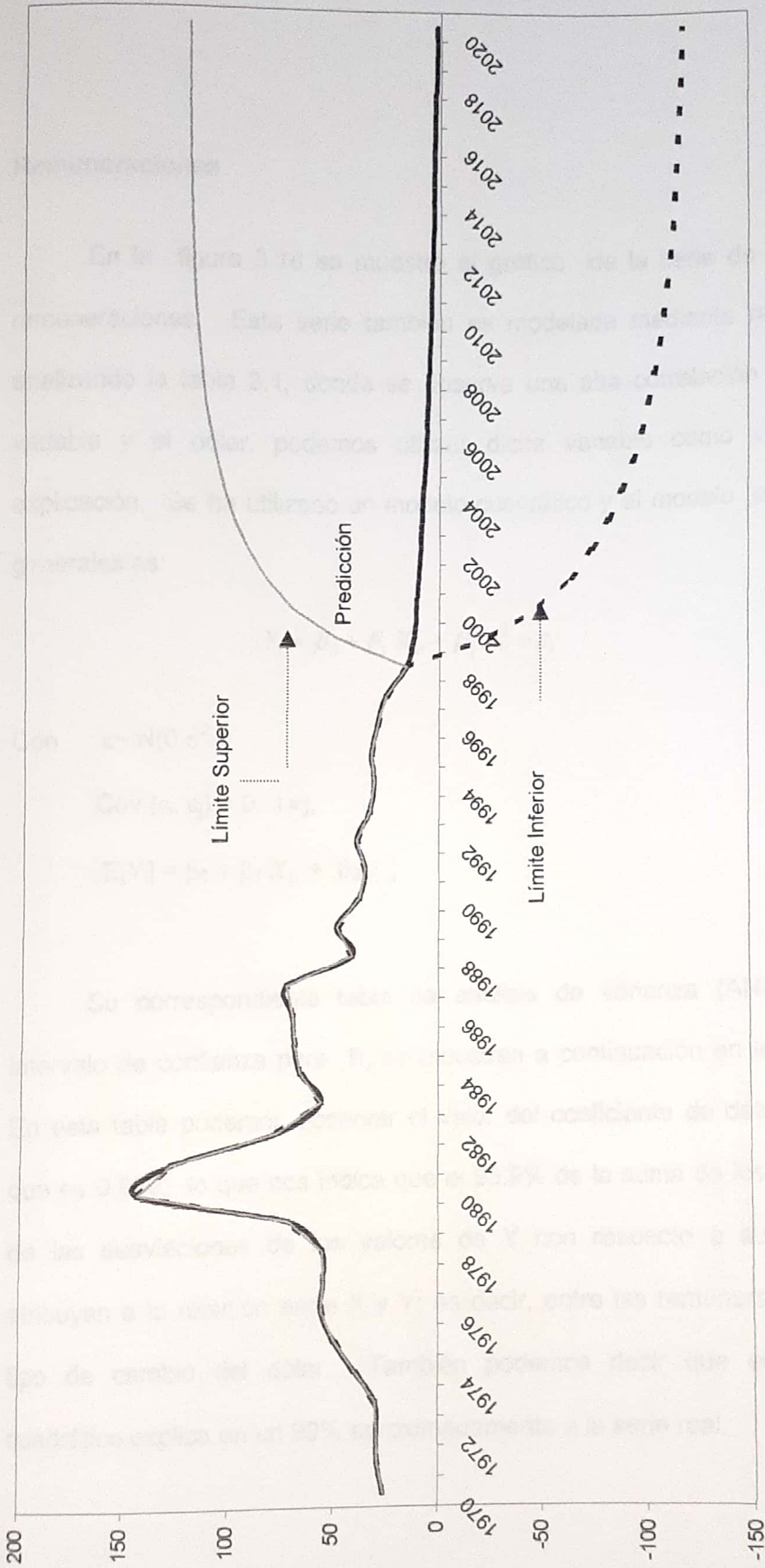


Figura 3.17 Gráfico de las predicciones del salario mínimo vital

Remuneraciones

En la figura 3.18 se muestra el gráfico de la serie de la variable remuneraciones. Esta serie también es modelada mediante regresión; y analizando la tabla 3.1, donde se observa una alta correlación entre esta variable y el dólar, podemos utilizar dicha variable como variable de explicación. Se ha utilizado un modelo cuadrático y el modelo en términos generales es:

$$Y_i = \beta_0 + \beta_1 X_{it} + \beta_2 X_{it}^2 + \varepsilon_i$$

Con $\varepsilon \sim N(0, \sigma^2)$;

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j;$$

$$E[Y_i] = \beta_0 + \beta_1 X_{it} + \beta_2 X_{it}^2$$

Su correspondiente tabla de análisis de varianza (ANOVA) y el intervalo de confianza para B, se muestran a continuación en la tabla 3.7. En esta tabla podemos observar el valor del coeficiente de determinación, que es 0.989; lo que nos indica que el 98.9% de la suma de los cuadrados de las desviaciones de los valores de Y con respecto a su media se atribuyen a la relación entre X y Y; es decir, entre las remuneraciones y el tipo de cambio del dólar. También podemos decir que este modelo cuadrático explica en un 99% aproximadamente a la serie real.

Ecuador: Remuneraciones desde el año 1970 - 1998

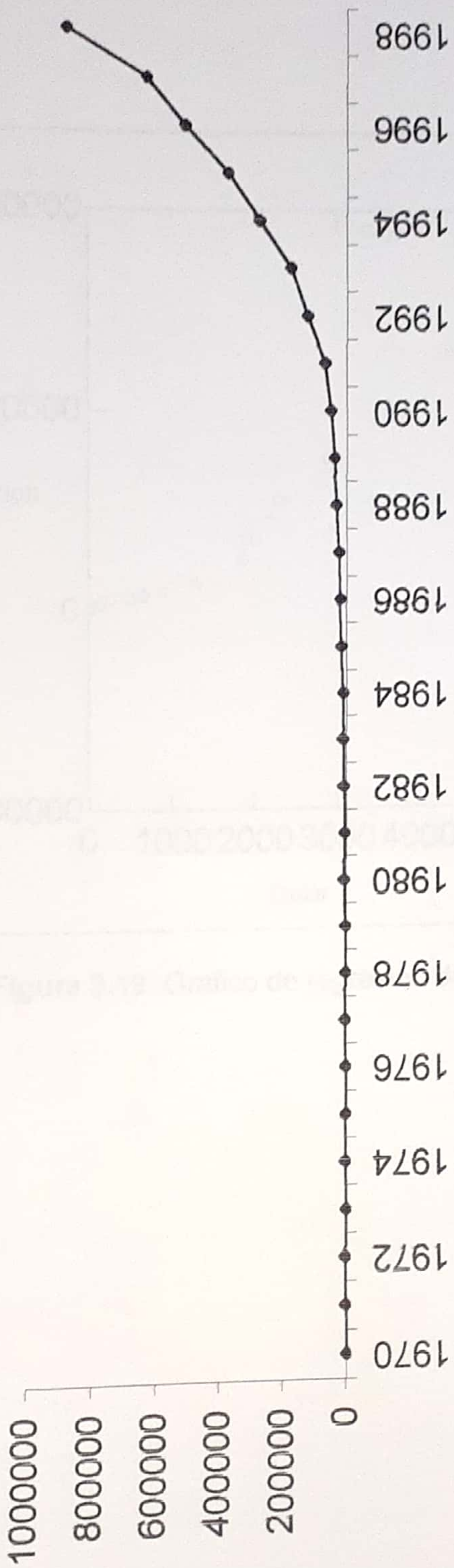


Figura 3.18 Gráfico de la serie de remuneraciones

Fuente: Banco Central del Ecuador, Boletín anual de 1970 - 1998
Ver datos en el Apéndice B

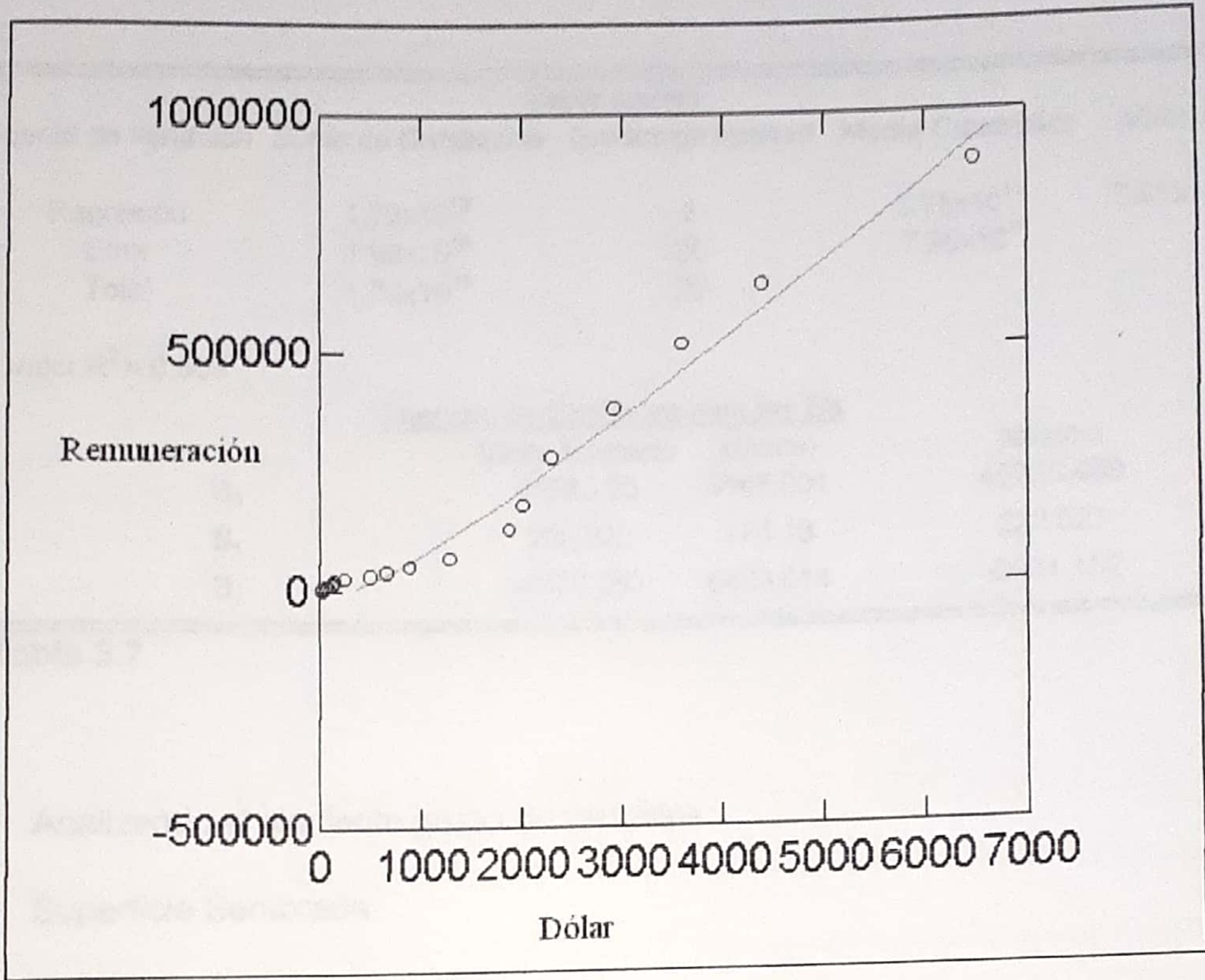


Figura 3.19 Grafico de regresión de remuneraciones

<u>Tabla Anova</u>				
Fuente de variación	Suma de Cuadrados	Grados de libertad	Media Cuadrática	Valor F
Regresión	1.72×10^{12}	3	5.75×10^{11}	7.91×10^2
Error	1.89×10^{10}	26	7.26×10^8	
Total	1.74×10^{12}	29		

Valor $R^2 = 0.989$

<u>Intervalo de Confianza para las B's</u>			
	Valor estimado	Mínimo	Máximo
B ₀	25785.26	5845.031	45725.488
B ₁	198.325	174.13	222.521
B ₂	-4670.083	6409.014	-2931.152

Tabla 3.7

Analizando el siguiente grupo de variables:

- ◆ Superficie Sembrada
- ◆ Superficie Cosechada
- ◆ Producción
- ◆ Rendimiento

Estas cuatro variables están además subdivididas en:

- ◆ invierno, verano y total

Primero se calculará la matriz de correlaciones para analizar la dependencia entre cada una de estas variables:

Matriz de Correlación

Matriz de correlación de las variables de producción del arroz												
	ssinv	ssver	sstot	scinv	scver	sctot	pinv	pver	ptot	rinv	rver	rtot
Ssinv	1											
Ssver	0.462	1										
Sstot	0.957	0.7	1									
Scinv	0.98	0.404	0.922	1								
Scver	0.507	0.961	0.723	0.449	1							
Sctot	0.958	0.647	0.983	0.954	0.696	1						
Pinv	0.965	0.421	0.916	0.977	0.5	0.952	1					
Pver	0.662	0.833	0.806	0.598	0.921	0.789	0.682	1				
Ptot	0.941	0.582	0.948	0.929	0.669	0.971	0.973	0.832	1			
Rinv	0.649	0.353	0.638	0.625	0.49	0.667	0.755	0.661	0.781	1		
Rver	0.459	-0.14	0.324	0.433	-0.011	0.345	0.513	0.369	0.505	0.46	1	
Rtot	0.642	0.252	0.6	0.607	0.415	0.627	0.747	0.696	0.786	0.915	0.759	1

Ssinv: superficie sembrada en invierno	Ssver: superficie sembrada en verano	Sstot: superficie sembrada total
Scinv: superficie cosechada en invierno	Scver: superficie cosechada en verano	Sctot: superficie cosechada total
Pinv: producción en invierno	Pver: producción en verano	Ptot: producción total
Rinv: rendimiento en invierno	Rver: rendimiento en verano	Rtot: rendimiento total

En este cuadro podemos observar que las correlaciones son en su mayoría bastante altas; esto se debe a que algunas de estas variables dependen de otras; por ejemplo, las variables totales, generalmente son la suma de las variables de invierno y verano; por esta razón para el análisis de estas variables sólo se ha tomado en cuenta las variables de producción y rendimiento, ya que estas variables pueden ser explicadas por las otras variables de este grupo, como la superficie cosechada y sembrada. A continuación el análisis por variable.

Producción Invierno

La variable de producción de invierno se la modela mediante regresión; está explicada en términos de la variable superficie cosechada en

invierno. El modelo que mejor ajusta a esta variable (por mayor porcentaje de explicación) es:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

Con: $\varepsilon \sim N(0, \sigma^2)$;

$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$; y, $E[Y] = \beta_0 + \beta_1 X_1$

Y la correspondiente tabla ANOVA se muestra a continuación, junto con el intervalo de confianza para los estimadores de las constantes.

Tabla Anova				
	Suma de Cuadrados	Grados de libertad	Media Cuadrática	Valor F
Regresión	3.82×10^{12}	2	1.91×10^{12}	1.47×10^3
Error	3.52×10^{10}	27	1.30×10^9	
Total	3.85×10^{12}	29		
Valor $R^2 = 0.991$				
Intervalo de Confianza para las B's				
	Valor estimado	Mínimo	Máximo	
B_0	-44183.83	-74425.269	-13942.398	
B_1	3.793	3.507	4.078	

Tabla 3.8

El gráfico correspondiente al modelo anteriormente señalado se muestra en la figura 3.21, con un coeficiente de determinación R^2 de 0.991; es decir, que este modelo explica a la variable de Producción de invierno en un 99.1%, con un 95% de confianza.

Ecuador: Producción del arroz en invierno

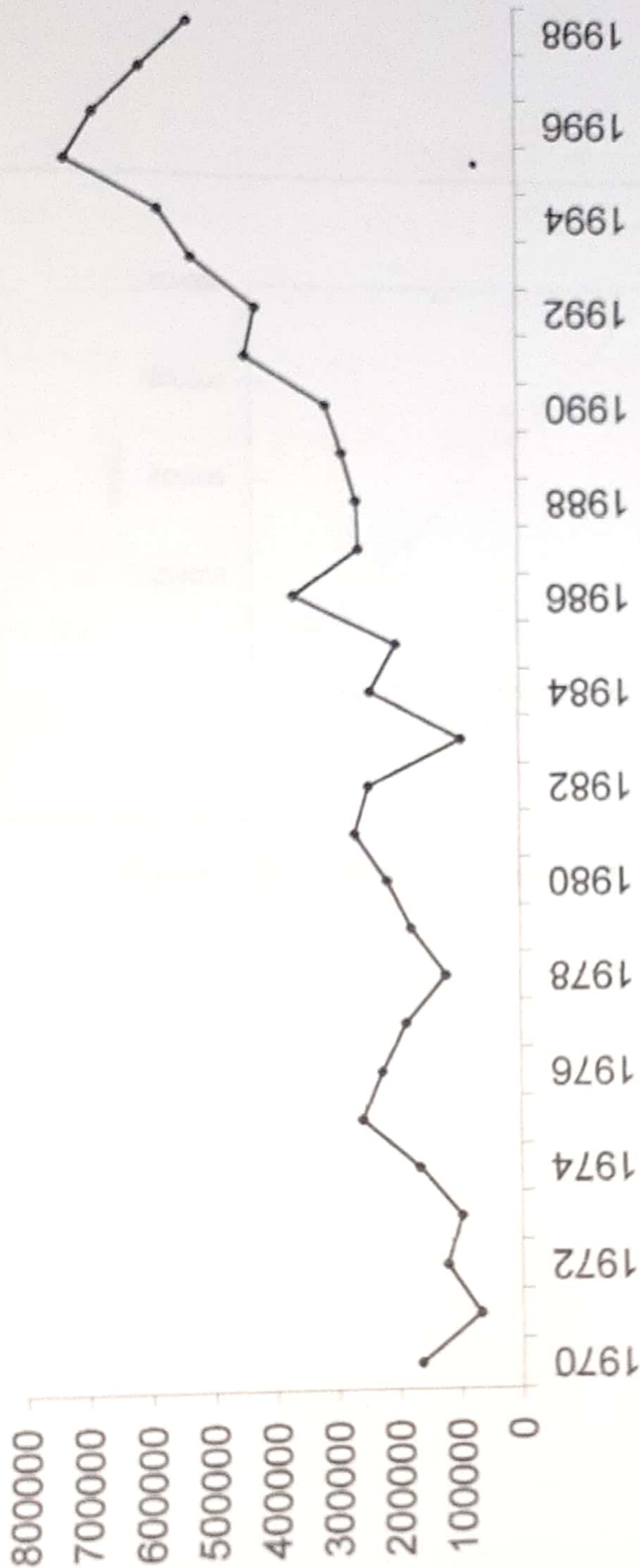


Figura 3.20 Gráfico de la serie de producción del arroz en invierno

Fuente: Ministerio de Agricultura, Superficie cosechada, superficie cosechada, superficie sembrada y producción 1997
Ver datos en el Apéndice B

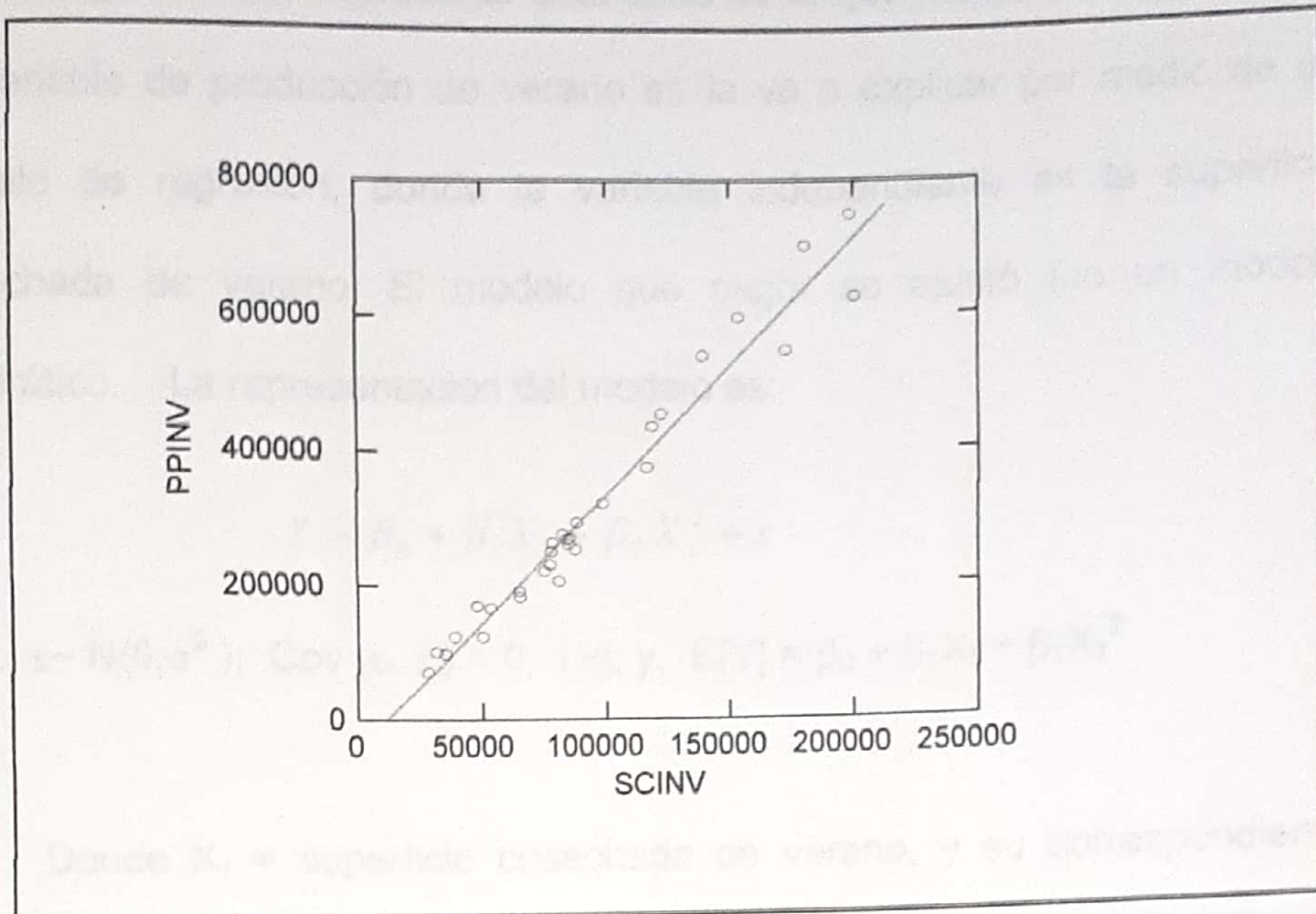


Figura 3.21 Gráfico de regresión, modelo: $Y = \beta_0 + \beta_1 X_1 + \epsilon$

Factor de producción	Unidad de medida	Cantidad utilizada	Unidad de medida	Valor P
Capital	10^6	100	10^6	7.22x10 ⁻⁷
Trabajo	10^3	1000	10^3	2.29x10 ⁻⁸

Estadísticas de regresión para el modelo			
	Valor	Unidad	Significado
R	0.992		0.000000
R Cuadrado	0.984		0.000000
Estadístico F	150.000		0.000000

Tabla 3.3

Producción Verano

El gráfico que representa esta serie es el que muestra la figura 3.22. La variable de producción de verano se la va a explicar por medio de un modelo de regresión; donde la variable independiente es la superficie cosechada de verano. El modelo que mejor se ajustó fue un modelo cuadrático. La representación del modelo es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon$$

Con: $\varepsilon \sim N(0, \sigma^2)$; $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$; y, $E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$

Donde X_1 = superficie cosechada de verano, y su correspondiente tabla anova se muestra en la tabla 3.9. El coeficiente de determinación R^2 es del 0.988, y el gráfico correspondiente a este modelo es el que se muestra en la figura 3.23

Tabla Anova				
Fuente de variación	Suma de Cuadrados	Grados de libertad	Media Cuadrática	Valor F
Regresión	1.14×10^{12}	3	3.80×10^{11}	7.23×10^2
Error	1.37×10^{10}	26	5.26×10^8	
Total	1.16×10^{12}	29		

Valor $R^2 = 0.988$

Intervalo de Confianza para las B's			
	Valor estimado	Mínimo	Máximo
B_0	-30229.4	-279565.5	219106.6
B_1	3.434	-1.04	7.907
B_2	140.027	-1972.891	2252.946

Tabla 3.9

Ecuador: Producción del arroz en verano

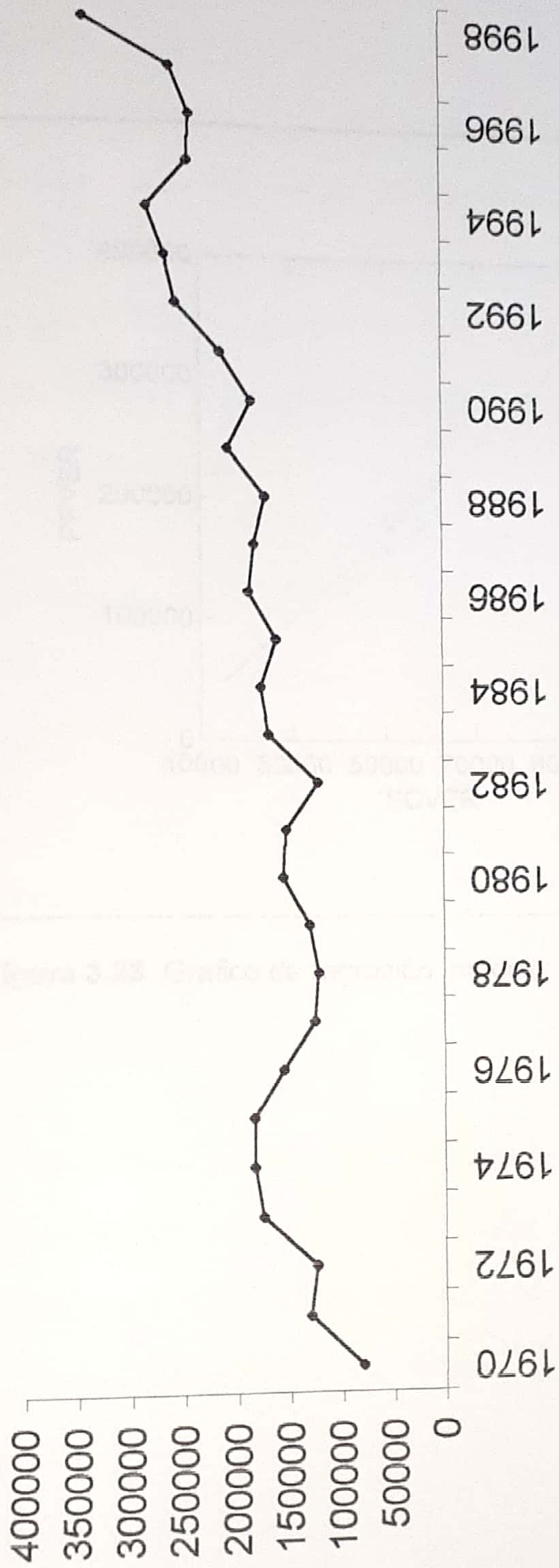


Figura 3.22 Gráfico de la serie de producción del arroz en verano

Fuente: Ministerio de agricultura, Superficie sembrada, superficie cosechada y producción 1997
Ver datos en el Apéndice B

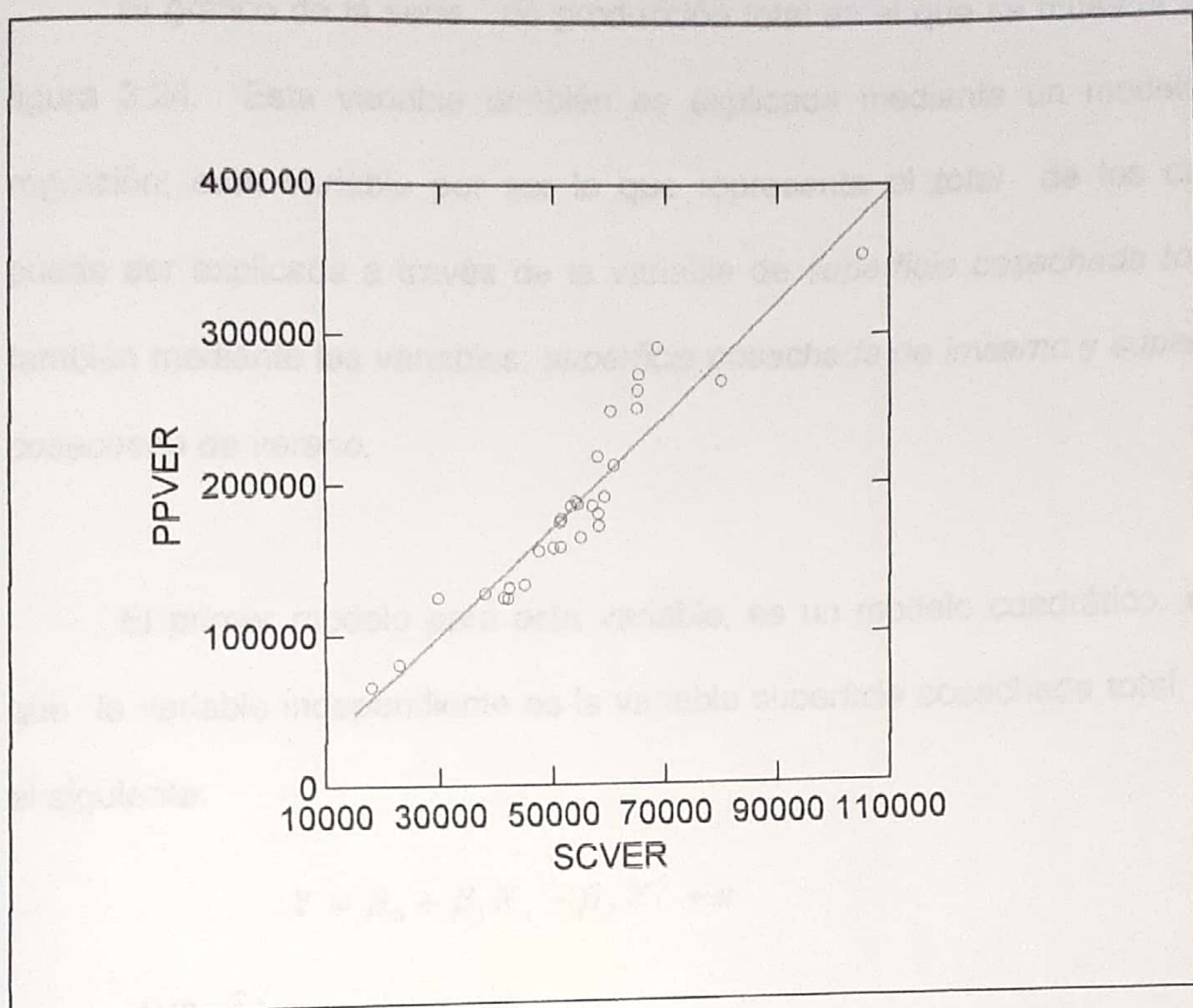


Figura 3.23 Grafico de regresión, modelo: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon$

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$$

Su correspondiente tabla y el intervalo de confianza para los β 's se muestran en la tabla 3.10. El coeficiente de determinación R^2 para la estimación de esta función es de 0.991, y la prueba de parámetros asociados. El gráfico de estimación del modelo es el que se muestra en la figura 3.25.

Producción Total

El gráfico de la serie de producción total es el que se muestra en la figura 3.24. Esta variable también es explicada mediante un modelo de regresión; esta variable por ser la que representa el *total* de los datos, puede ser explicada a través de la variable de *superficie cosechada total* o también mediante las variables: *superficie cosechada de invierno* y *superficie cosechada de verano*.

El primer modelo para esta variable, es un modelo cuadrático; en el que la variable independiente es la variable superficie cosechada total; y es el siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon$$

con $\varepsilon \sim N(0, \sigma^2)$;

$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$; y,

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$$

Su correspondiente tabla y el intervalo de confianza para las B's se muestran en la tabla 3.10. El coeficiente de determinación R^2 para la estimación de esta variable es de 0.991; y la prueba de parámetros aceptada. El gráfico de estimación del modelo es el que se muestra en la figura 3.25.

Ecuador: Producción total del arroz

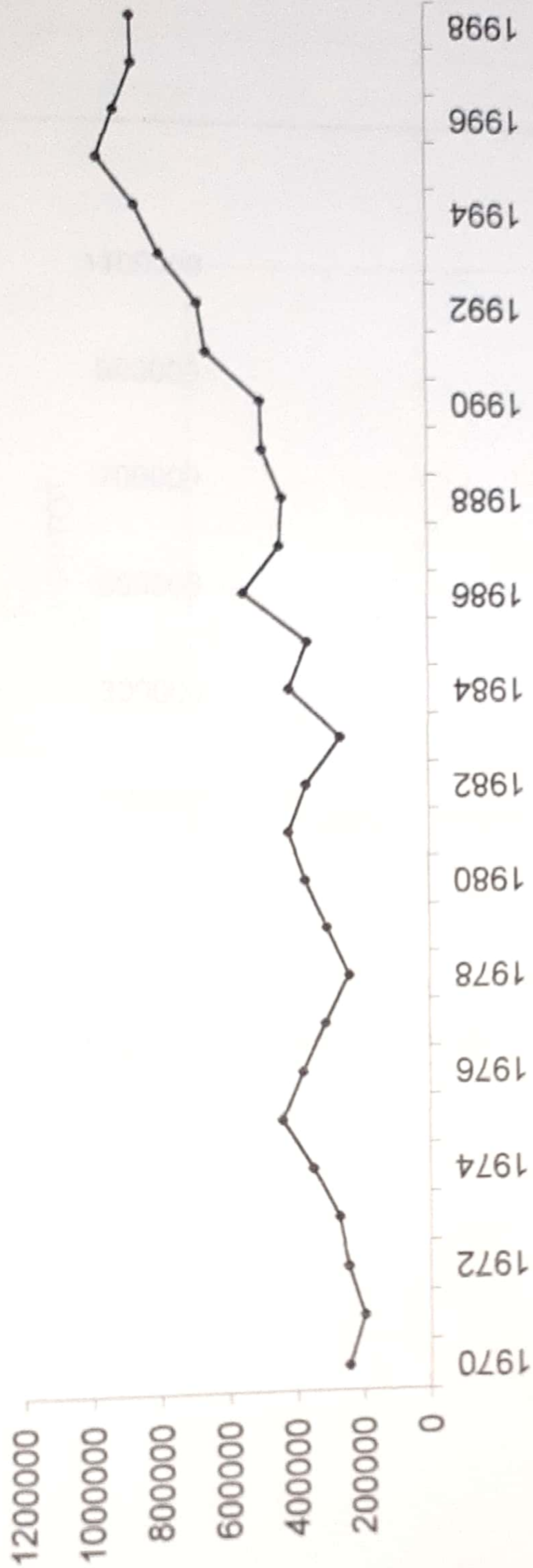


Figura 3.24 Gráfico de la serie de producción total del arroz

Fuente: Ministerio de agricultura, Superficie sembrada, superficie cosechada y producción 1997
Ver datos en el Apéndice B

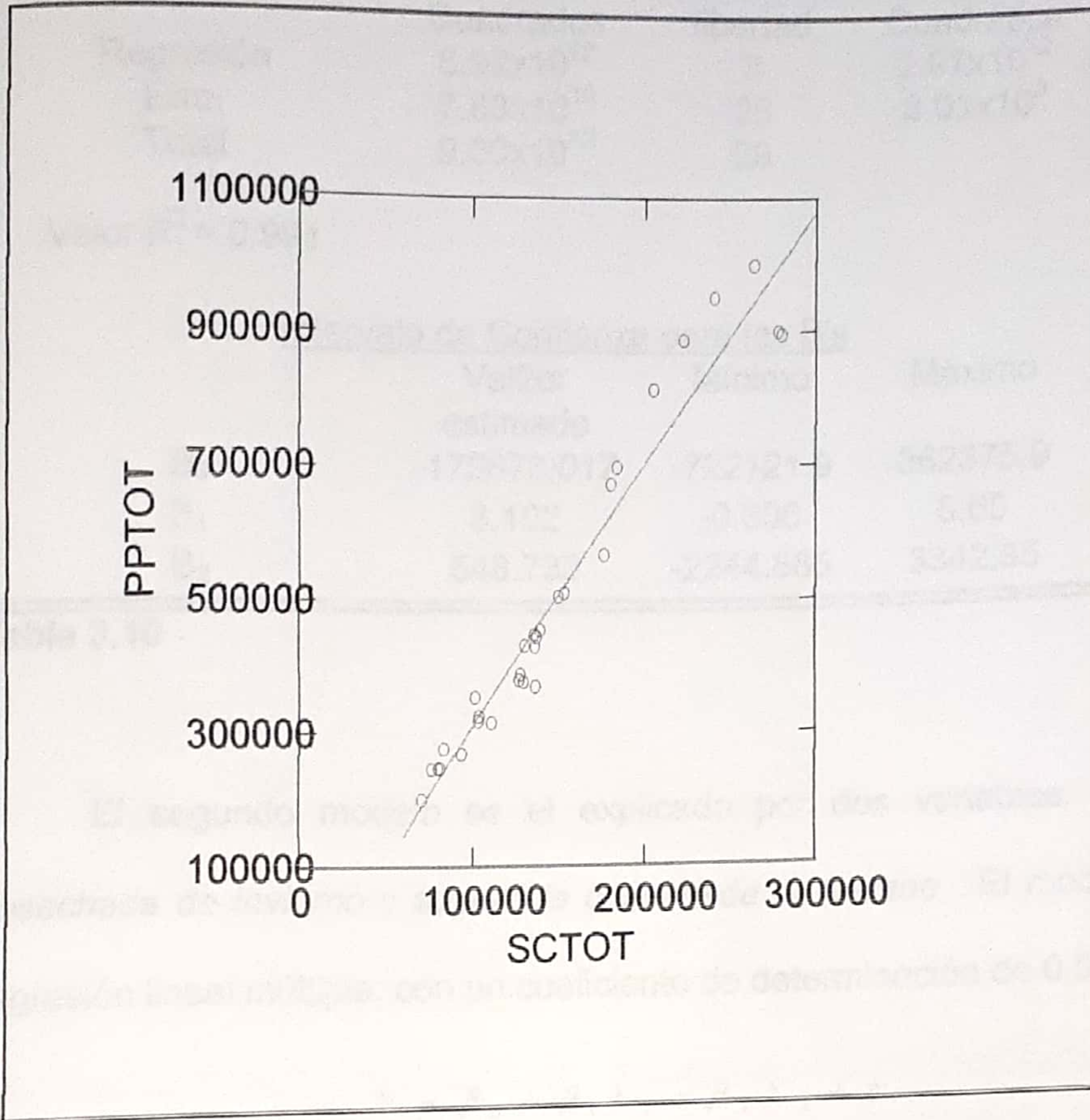


Figura 3.25 Grafico de regresión de la producción total del arroz

Tabla Anova				
Fuente de variación	Suma de Cuadrados	Grados de libertad	Media Cuadrática	Valor F
Regresión	8.92×10^{12}	3	2.97×10^{12}	9.88×10^2
Error	7.83×10^{10}	26	3.01×10^9	
Total	9.00×10^{12}	29		

Valor $R^2 = 0.991$

Intervalo de Confianza para las B's			
	Valor estimado	Mínimo	Máximo
B_0	-179873.017	-722121.9	362375.9
B_1	3.192	-0.306	6.69
B_2	548.733	-2244.885	3342.35

Tabla 3.10

El segundo modelo es el explicado por dos variables: *superficie cosechada de invierno* y *superficie cosechada de verano*. El modelo es de regresión lineal múltiple; con un coeficiente de determinación de 0.992.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

con: $\varepsilon \sim N(0, \sigma^2)$;

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j; \quad y,$$

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

La tabla Anova y el intervalo de confianza para B se muestran a continuación en la Tabla 3.11.

Tabla Anova				
Fuente de variación	Suma de Cuadrados	Grados de libertad	Media Cuadrática	Valor F
Regresión	8.93×10^{12}	3	2.98×10^{12}	1.12×10^3
Error	6.94×10^{10}	26	2.67×10^9	
Total	9.00×10^{12}	29		

Valor $R^2 = 0.992$

Intervalo de Confianza para las B's			
	Valor estimado	Mínimo	Máximo
B_0	-22283.146	-99792.3	55226.027
B_1	4.318	3.732	4.905
B_2	2.171	0.275	4.066

Tabla 3.11

Los valores que toman las variables correspondientes al **Rendimiento**, son el cociente entre la Producción y la Superficie Cosechada correspondiente al término que se requiere analizar. Es por eso que las variables de Rendimiento tienen una correlación alta con la producción y la superficie cosechada; de allí que se estima estas variables mediante modelos de regresión.

Para ilustrar lo expuesto en el párrafo anterior, vamos a realizar el análisis sólo de la variable **Rendimiento Total**, ya que las otras dos, tienen el mismo formato; es decir, que la variable rendimiento en invierno y rendimiento en verano son modeladas por medio de regresión, y al igual que la variable de rendimiento total, las variables independientes en estos

modelos son la superficie cosechada y. producción en invierno y verano, respectivamente. Cabe mencionar que el rendimiento total no es la suma del rendimiento de invierno con el de verano; si no que se lo haya como se explicó anteriormente; a esto se debe la similitud en el análisis de estas variables. Empezaremos con el gráfico de la serie de esta variable; el mismo que se muestra en la figura 3.26

Como era de esperarse, esta serie también se la puede ajustar mediante dos modelos; los cuales tienen el mismo valor de coeficiente de determinación. El primer modelo tiene las dos variables mencionadas anteriormente, y el segundo modelo tiene desglosadas las variables anteriores; este modelo consta de cuatro variables, y estas son: Superficie cosechada en invierno, producción de invierno, Superficie cosechada en verano, producción de verano. Los modelos se muestran a continuación:

Modelo I

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad ; \quad \varepsilon \sim N(0, \sigma^2);$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j; \quad y,$$

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Modelo II

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon; \quad \varepsilon \sim N(0, \sigma^2);$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j; \quad y,$$

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Ecuador: Rendimiento total del arroz

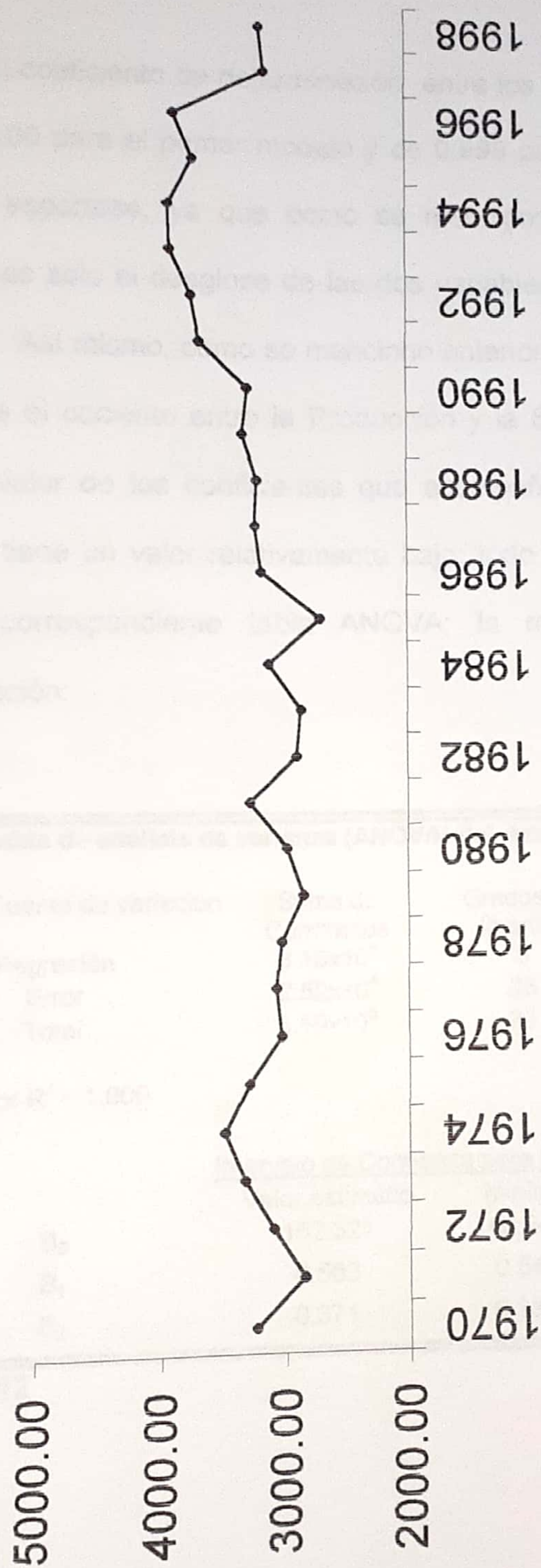


Figura 3.26 Grafico de la serie del rendimiento total de la producción del arroz

Fuente: Ministerio de agricultura, Superficie sembrada, superficie cosechada y producción 1997
Ver datos en el Apéndice B

El coeficiente de determinación entre los dos modelos es muy similar; es de 1.00 para el primer modelo y de 0.999 para el segundo modelo; esto era de esperarse, ya que como se mencionó anteriormente el segundo modelo es solo el desglose de las dos variables independientes del primer modelo. Así mismo, como se mencionó anteriormente, el Rendimiento no es más que el cociente entre la Producción y la Superficie Cosechada; por lo que el valor de los coeficientes que acompañan a cada variable en este modelo tiene un valor relativamente bajo; todo esto se lo podrá comprobar en la correspondiente tabla ANOVA; la misma que se presenta a continuación:

Tabla de análisis de varianza (ANOVA) del rendimiento total				
Fuente de variación	Suma de Cuadrados	Grados de libertad	Media Cuadrática	Valor F
Regresión	3.19×10^8	3	1.06×10^8	1.10×10^5
Error	2.52×10^4	26	9.69×10^2	
Total	3.19×10^8	29		
Valor $R^2 = 1.000$				
	Intervalo de Confianza para las B's			
	Valor estimado	Mínimo	Máximo	
B_0	152.321	40.241	264.4	
B_1	0.583	0.545	0.622	
B_2	0.371	0.332	0.409	

Tabla 3.12

CAPITULO IV

ANÁLISIS MULTIVARIADO

4.1 Introducción

El Análisis Multivariado es un conjunto de técnicas estadísticas que simultáneamente analizan las múltiples medidas existentes sobre un conjunto de objetos de interés,. Las técnicas multivariadas se clasifican en: dependientes e interdependientes.

- **Interdependientes**

Son las técnicas en las cuales una variable o grupo de variables están definidas como entes independientes o dependientes.

Para el estudio de las variables utilizadas se ha escogido la técnica multivariada de Componentes Principales.

El Análisis de Componentes Principales es una técnica multivariada de interdependencia, en la que se estudian p variables observables de interés, que constituyen un vector aleatorio $\mathbf{X}^T = (X_1, X_2, \dots, X_p)$; en el que estas p variables observables, generarán k variables latentes, donde $k < p$, y se pretende contenga tanta información como las p variables originales.

Algebraicamente, las componentes principales son combinaciones lineales de las p variables aleatorias observables X_1, X_2, \dots, X_p . Geométricamente, esas combinaciones lineales representan la selección de un nuevo sistema de coordenadas obtenido por la rotación del sistema original con X_1, X_2, \dots, X_p como los ejes de coordenados.

Las componentes principales dependen únicamente de la matriz de covarianzas Σ (o de la matriz de correlación ρ) de X_1, X_2, \dots, X_p . Cabe mencionar que el desarrollo de componentes principales no supone que la población considerada sea normal p -variada.

Tenemos un vector aleatorio $X^T = [X_1, X_2, \dots, X_p]$, que tiene una matriz de covarianzas Σ , con sus respectivos valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Consideremos las combinaciones lineales:

$$Y_1 = a_1^T X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a_2^T X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

.

.

.

$$Y_p = a_p^T X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

Entonces:

$$\text{Var}(Y_i) = a_i^T \Sigma a_i \quad i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i, Y_k) = a_i^T \Sigma a_k \quad i, k = 1, 2, \dots, p$$

Las **componentes principales** son esas *combinaciones lineales no correlacionadas* Y_1, Y_2, \dots, Y_p , cuyas varianzas son tan grandes como sea posible.

La primera componente principal es la combinación lineal con máxima varianza; esto es, maximiza la $\text{Var}(Y_1) = a_1^T \Sigma a_1$. Está claro que la

$\text{Var}(Y_1) = a_1^T \Sigma a_1$ puede ser incrementada multiplicando cualquier a_1 por alguna constante. Para eliminar esta indeterminación, es conveniente restringir la atención a vectores de coeficientes de norma unitaria. Entonces definimos:

Primera componente principal $Y_1 =$ la combinación lineal $a_1^T X$ que

maximiza $\text{Var}(a_1^T X)$, sujeto a

$$\|a_1\| = +\sqrt{\langle a_1, a_1 \rangle} = 1$$

Segunda componente principal $Y_2 =$ la combinación lineal $a_2^T X$ que

maximiza $\text{Var}(a_2^T X)$, sujeto a

$$\|a_2\| = +\sqrt{\langle a_2, a_2 \rangle} = 1 \text{ y } \text{Cov}(a_1^T X, a_2^T X) = 0$$

En el i -ésimo paso:

La i -ésima componente principal $Y_i =$ la combinación lineal $a_i^T X$ que

maximiza $\text{Var}(a_i^T X)$, sujeto a

$$\|a_i\| = +\sqrt{\langle a_i, a_i \rangle} = 1 \text{ y } \text{Cov}(a_i^T X, a_k^T X) = 0;$$

para $k < i$.

Bajo estas condiciones, se puede demostrar que:

Si Σ es la matriz de covarianzas correspondiente a un vector aleatorio $X^T = [X_1, X_2, \dots, X_p]$; Σ tiene asociado los pares de valores y vectores propios $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$, donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$; entonces la i -ésima componente principal está dada por:

$$Y_i = e_i^T X = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p, \quad i = 1, 2, \dots, p \quad (4.1)$$

De modo que ahora tenemos:

$$\text{Var}(Y_i) = e_i^T \Sigma e_i = \lambda_i \quad i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i, Y_k) = e_i^T \Sigma e_k = 0$$

Si algunas λ_i son iguales, la elección de los correspondientes vectores de coeficientes e_i y por lo tanto de Y_i ; no son únicas. Recuérdese que

$$\|e_1\| = \|e_2\| = \dots = \|e_p\| = 1, \text{ y que } \langle e_i, e_j \rangle = 0, \text{ para } i \neq j$$

Otro resultado importante, dadas las condiciones anteriores y siendo

$Y_1 = e_1^T X, Y_2 = e_2^T X, \dots, Y_p = e_p^T X$ las componentes principales; entonces:

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i)$$

De aquí se puede decir que:

$$\text{Varianza total de la población} = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p;$$

y consecuentemente, la proporción de varianza total explicada por la k -ésima componente principal es:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}; \quad k = 1, 2, \dots, p$$

Si la mayoría (por ejemplo 80 al 90%) de la varianza total de la población para todas las p , puede ser atribuida a la primera, segunda y tercera componente, entonces estas componentes pueden reemplazar las p variables originales sin mucha pérdida de información.

Componentes principales obtenidas de variables estandarizadas.

Cuando las unidades de medida de las variables son diferentes, se recomienda estandarizar las variables para evitar resultados erróneos que podrían generarse por diferencias demasiado grandes entre las magnitudes de las variables. Las componentes principales pueden ser obtenidas de variables estandarizadas:

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}}$$

$$Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}}$$

$$Z_p = \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}}$$

En notación matricial:

$$Z = (V^{1/2})^{-1} (X - \mu)$$

donde $V^{1/2}$ es la matriz diagonal de desviación estándar, definida por:

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_{22}} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

donde, $E[Z] = 0$ y

$$\text{Cov}(Z) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = \rho$$

Las componentes principales de Z pueden ser obtenidas de los vectores propios de la matriz de correlación ρ de X . La teoría que se aplica para hallar las componentes principales de Σ , es la misma que se aplica cuando se utiliza ρ ; pero los resultados son diferentes ya que los pares de valores y vectores propios (λ_i, e_i) obtenidos de Σ , no son los mismos que los obtenidos a partir de ρ .

Rotación de factores

Una importante instrumento de apoyo en la interpretación de factores es la **rotación**. El término rotación significa específicamente que los ejes de los factores son girados alrededor del origen hasta alguna otra posición alcanzable. El efecto de rotar la matriz de factores es redistribuir la varianza desde los primeros factores hasta los finales para asegurar una manera más simple y teóricamente más fácil de comprender. Existen varios casos de rotación: rotación *ortogonal* y rotación *oblicua*.

Métodos de rotación ortogonal.- El objetivo es de todos los métodos de rotación ortogonal es simplificar las filas y columnas de la matriz de factores para facilitar su interpretación. Existen tres métodos de rotación ortogonal: *quartimax*, *varimax* y *equimax*.

Quartimax.- simplifica las filas de la matriz de factores

Varimax .- en contraste con *quartimax*, el criterio de *varimax*, se centra en simplificar las columnas de la matriz de factores.

Equimax.- este método más bien que concentrado en la simplificación de filas o en la simplificación de columnas, trata de realizar algo de cada una.

Método de rotación oblicua.- La rotación oblicua es similar a la rotación ortogonal; excepto que las rotaciones oblicuas permiten factores correlacionados, en vez de mantener independencia entre los factores rotados.

4.2 Análisis de Componentes Principales de las variables de estudio

Para el análisis de componentes principales se trabajará con todas las variables agrupadas. El análisis se lo realizará en tres partes; primero se trabajará con los datos originales, luego se hará una rotación y por último se estandarizaran los datos para el último análisis y se escogerá la mejor opción.

En la tabla 4.1 se muestra la matriz de correlaciones de las 20 variables. Como podemos observar en esta matriz, existen correlaciones sumamente altas, sobre todo entre las variables de "producción"; esto se debe a que en este grupo unas variables dependen de otras y se pueden expresar en términos de ellas; por ejemplo la variable producción total de arroz tiene una correlación de 1.00 con la variable producción de arroz en invierno, debido a que la mayor cantidad de arroz que se produce anualmente proviene de la producida en el primer ciclo; es decir, en el invierno. También las variables de la superficie cosechada (X_{11} , X_{12} , X_{13}) tienen alta correlación con las variables de la superficie sembrada (X_8 , X_9 , X_{10}), ya que la superficie que se cosecha depende directamente de la que se siembra; en ciertas ocasiones hasta es la misma; pero no siempre, ya que a veces hay pérdidas en los sembrados. También podemos observar que las remuneraciones tienen una correlación negativa con respecto al dólar, lo

Matriz de correlación

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀	
X ₁	1																				
X ₂	0.05	1																			
X ₃	0.27	0.07	1																		
X ₄	0.26	0.55	0.6	1																	
X ₅	0.01	-0.31	-0.2	-0.22	1																
X ₆	0.39	0.82	0.2	0.68	-0.4	1															
X ₇	0.38	0.77	0.3	0.75	-0.5	1.0	1														
X ₈	0.15	0.75	0.3	0.85	-0.3	0.9	0.88	1.00													
X ₉	0.49	0.68	0.4	0.72	-0.2	0.9	0.85	0.77	1												
X ₁₀	0.25	0.77	0.3	0.86	-0.3	0.9	0.92	0.98	0.87	1.00											
X ₁₁	0.09	0.72	0.3	0.85	-0.3	0.8	0.85	0.99	0.74	0.97	1										
X ₁₂	0.54	0.55	0.4	0.75	-0.2	0.8	0.82	0.75	0.95	0.84	0.72	1									
X ₁₃	0.20	0.72	0.3	0.87	-0.3	0.9	0.89	0.98	0.84	0.99	0.99	0.83	1								
X ₁₄	0.02	0.67	0.3	0.84	-0.3	0.8	0.82	0.97	0.67	0.94	0.98	0.68	0.96	1							
X ₁₅	0.35	0.58	0.4	0.82	-0.4	0.8	0.86	0.84	0.85	0.88	0.81	0.93	0.89	0.8	1						
X ₁₆	0.10	0.67	0.3	0.86	-0.3	0.8	0.86	0.97	0.74	0.95	0.97	0.76	0.98	1.0	0.9	1					
X ₁₇	-0.14	0.28	0.3	0.57	-0.2	0.4	0.44	0.61	0.29	0.55	0.61	0.42	0.60	0.7	0.6	0.73	1				
X ₁₈	-0.27	0.27	0.2	0.42	-0.5	0.3	0.40	0.49	0.06	0.40	0.49	0.15	0.43	0.6	0.5	0.58	0.63	1			
X ₁₉	-0.21	0.29	0.3	0.55	-0.4	0.4	0.46	0.61	0.22	0.53	0.60	0.36	0.58	0.7	0.6	0.73	0.94	0.85	1		
X ₂₀	0.08	0.41	-0.1	-0.05	-0.4	0.3	0.30	0.11	0.29	0.16	0.07	0.29	0.13	0.1	0.3	0.15	0.04	0.22	0.1	1	

Tabla 4.1

X₁: Importaciones del arroz
 X₆: remuneraciones
 X₁₁: sup. cosechada en invierno
 X₁₆: producción total

X₂: exportación del arroz
 X₇: dólar
 X₁₂: sup. cosechada en verano
 X₁₇: rendimiento en invierno

X₃: ipc
 X₈: superficie sembrada en invierno
 X₁₃: sup. cosechada total
 X₁₈: rendimiento en verano

X₄: deuda externa
 X₉: superficie sembrada en invierno
 X₁₄: producción en invierno
 X₁₉: rendimiento total

X₅: salarios
 X₁₀: sup. sembrada total
 X₁₅: producción en verano
 X₂₀: clima

cual sonaría absurdo, pues esto representa una relación inversa; es decir, que mientras sube el valor del dólar, el valor de las remuneraciones disminuye; lo que sí disminuiría es el poder adquisitivo de las personas; aunque en la actualidad con la dolarización, ya no hay cambios en el valor del dólar; pero estos resultados son basados en datos anteriores a la dolarización.

Ahora procederemos a realizar el análisis de componentes principales. Trabajando con los datos originales y siguiendo el proceso anteriormente descrito para este análisis, se obtuvieron los siguientes resultados, primero los valores propios correspondientes de dicha matriz; se escogieron los dos primeros, ya que con ellos se logra un gran porcentaje de explicación (98%).

$$\lambda_1 = 1.46 \cdot 10^{11} \quad \lambda_2 = 1.33 \cdot 10^{10}$$

Obtenemos ahora la matriz de covarianzas entre las variables originales y las dos primeras componentes; la misma que se muestra en la siguiente tabla:

MATRIZ DE COVARIANZAS		
	1	2
X ₁	4297.255	-12362.967
X ₂	21211.093	-8737.281
X ₃	5.55	2.524
X ₄	4704.644	981.851
X ₅	-10.685	6.128
X ₆	199148.154	-94910.544
X ₇	1491.972	-536.22
X ₈	51598.367	3928.159
X ₉	14995.75	-6014.557
X ₁₀	66594.117	-2086.398
X ₁₁	47759.714	6612.3
X ₁₂	12097.584	-3148.484
X ₁₃	59857.298	3463.816
X ₁₄	183125.366	44851.879
X ₁₅	54101.707	-605.584
X ₁₆	237227.073	44246.295
X ₁₇	255.039	208.885
X ₁₈	205.758	147.002
X ₁₉	222.096	176.166
X ₂₀	0.518	-0.966

Tabla 4.2: Matriz de covarianzas entre las variables y las componentes principales

Como se puede notar en esta matriz hay valores muy discrepantes; es decir, que los valores tienen mucha diferencia entre sí; esto se debe a que las variables no están medidas en las mismas unidades.

Ahora obtenemos las componentes:

$$Y_1 = 0.011X_1 + 0.055X_2 + 0.012X_4 + 0.521X_6 + 0.004X_7 + 0.135X_8 + 0.039X_9 +$$

$$0.174X_{10}+0.125X_{11}+0.032X_{12}+0.157X_{13}+0.479X_{14}+0.141X_{15}+0.620X_{16}+0.001X_{17}+0.001X_{18}+0.001X_{19}$$

$$Y_2 = -0.107X_1 - 0.076X_2 + 0.009X_4 - 0.822X_6 - 0.005X_7 + 0.034X_8 - 0.052X_9 - 0.018X_{10} + 0.057X_{11} - 0.027X_{12} + 0.030X_{13} + 0.388X_{14} - 0.005X_{15} + 0.383X_{16} + 0.002X_{17} + 0.001X_{18} + 0.002X_{19}$$

Donde Y_1 tiene un porcentaje de explicación de 90.258%, y Y_2 de 8.228%, con lo que tenemos un total de 98.486%. Si observamos cada una de las componentes, en ellas hay coeficientes que no son representativos, por lo cual podríamos reducir las componentes, dejando así los valores significativos para poder darles una denominación. Así tenemos que Y_1 sería:

$$Y_1 = 0.521X_6 + 0.479X_{14} + 0.620X_{16}$$

Es decir; que Y_1 se reduce a estas tres variables que son: remuneración, producción de arroz en invierno y producción total de arroz. De la misma manera Y_2 sería:

$$Y_2 = -0.822X_6$$

En este caso tenemos que sólo una variable es significativa, aunque de manera negativa.

Ahora realizando el análisis de componentes principales con los datos estandarizados, tenemos los valores de los tres primeros valores propios,

que son los que tomaremos para encontrar las componentes, ya que son suficientes para obtener un buen porcentaje de explicación.

$$\lambda_1 = 12.406 \quad \sum_{i=1}^{20} \lambda_i = 20.001$$

$$\lambda_2 = 2.521$$

$$\lambda_3 = 1.634$$

Obtenemos ahora la matriz de correlaciones correspondiente, que representa la correlación existente entre las variables con cada una de las componentes:

Matriz de correlaciones			
	1	2	3
IMPORT	0.237	0.729	0.13
EXPORT	0.738	0.151	-0.36
INFLACIÓN	0.397	0.063	0.43
DEUDA	0.876	0.016	0.354
SALARIO	-0.399	0.195	0.564
REMUNER	0.888	0.287	-0.244
DÓLAR	0.93	0.206	-0.174
SSINV	0.966	-0.035	0.067
SSVER	0.835	0.488	-0.005
SSTOT	0.976	0.107	0.05
SCINV	0.95	-0.072	0.096
SCVER	0.848	0.391	0.063
SCTOT	0.977	0.04	0.093
PPINV	0.945	-0.215	0.09
PPVER	0.937	0.06	0.012
PPTOT	0.976	-0.154	0.073
RENDINV	0.653	-0.577	0.186
RENDVER	0.523	-0.705	-0.197
RENDTOT	0.656	-0.69	0.027
CLIMA	0.237	0.103	-0.808

Tabla 4.3: Matriz de correlación entre las variables y las componentes principales

Continuando con el proceso, tenemos las componentes, cuyos porcentajes de explicación son: Y_1 tiene un porcentaje de varianza explicada de 62.031%; Y_2 de 12.603%; y, Y_3 de 8.169%; lo que da un total de 82.8%.

Las componentes son:

$$Y_1 = 0.067X_1 + 0.209X_2 + 0.113X_3 + 0.249X_4 - 0.249X_5 + 0.252X_6 + 0.264X_7 + 0.274X_8 + 0.237X_9 + 0.277X_{10} + 0.270X_{11} + 0.241X_{12} + 0.277X_{13} + 0.268X_{14} + 0.266X_{15} + 0.277X_{16} + 0.185X_{17} + 0.148X_{18} + 0.186X_{19} + 0.067X_{20}$$

$$Y_2 = 0.459X_1 + 0.095X_2 + 0.040X_3 + 0.010X_4 + 0.123X_5 + 0.181X_6 + 0.130X_7 - 0.022X_8 + 0.307X_9 + 0.067X_{10} - 0.045X_{11} + 0.246X_{12} + 0.025X_{13} - 0.136X_{14} + 0.038X_{15} - 0.097X_{16} - 0.363X_{17} - 0.444X_{18} - 0.435X_{19} + 0.065X_{20}$$

$$Y_3 = 0.101X_1 - 0.282X_2 + 0.336X_3 + 0.277X_4 + 0.441X_5 - 0.191X_6 - 0.136X_7 + 0.052X_8 - 0.004X_9 + 0.039X_{10} + 0.075X_{11} + 0.050X_{12} + 0.073X_{13} + 0.070X_{14} + 0.010X_{15} + 0.057X_{16} + 0.145X_{17} - 0.154X_{18} + 0.021X_{19} - 0.633X_{20}$$

donde: X_1 representa a la importación del arroz;

X_2 representa a la exportación del arroz;

X_3 representa a la ipc;

X_4 representa a la deuda externa;

X_5 representa a los salarios;

X_6 representa a las remuneraciones;

X_7 representa al dólar;

X_8 representa la superficie sembrada de arroz en invierno;

X_9 representa la superficie sembrada de arroz en verano;

X_{10} representa la superficie sembrada de arroz total;

X_{11} representa la superficie cosechada de arroz en invierno;

X_{12} representa la superficie cosechada de arroz en verano;

X_{13} representa la superficie cosechada de arroz total;

X_{14} representa la producción de arroz en invierno;

X_{15} representa la producción de arroz en verano;

X_{16} representa la producción de arroz total;

X_{17} representa el rendimiento del arroz en el invierno;

X_{18} representa el rendimiento del arroz en verano;

X_{19} representa el rendimiento del arroz total; y,

X_{20} representa el porcentaje de humedad relativa (clima).

Ahora para reducir estas componentes analizamos sus valores y tomamos sólo las variables con mayor significancia; de ahí que tengamos:

$$Y_1 = 0.274X_8 + 0.277X_{10} + 0.270X_{11} + 0.277X_{13} + 0.268X_{14} + 0.266X_{15} + 0.277X_{16}$$

$$Y_2 = 0.459X_1 - 0.444X_{18} - 0.435X_{19}$$

$$Y_3 = 0.633X_{20}$$

Ahora haciéndole una rotación varimax a los datos originales obtenemos los siguientes resultados:

$$\lambda_1 = 9.86 \cdot 10^{10}$$

$$\lambda_2 = 6.08 \cdot 10^{10}$$

Y la matriz de covarianzas correspondiente es la que se muestra en la tabla 4.3; y las componentes tienen un porcentaje de explicación total de 98.486%, donde Y_1 tiene una porcentaje de explicación de 60.914% y Y_2 de 37.572%.

Matriz de correlación de varimax		
	1	2
X_1	-3950.4	12478.129
X_2	11773.224	19688.619
X_3	5.958	1.296
X_4	4357.645	2026.979
X_5	-4.898	-11.302
X_6	102835.282	195174.003
X_7	874.983	1322.087
X_8	43701.425	27712.946
X_9	8420.586	13789.167
X_{10}	52122.011	41502.114
X_{11}	42230.438	23265.918
X_{12}	7812.139	9758.84
X_{13}	50042.577	33024.758
X_{14}	173586.386	73582.316
X_{15}	42996.015	32843.631
X_{16}	216582.401	106425.947
X_{17}	329.327	-14.865
X_{18}	252.821	5.254
X_{19}	283.358	-8.347
X_{20}	-0.163	1.083

Tabla 4.4: Matriz de covarianzas entre las dos primeras componentes y las variables con varimax

Como se mencionó anteriormente, una vez obtenido los tres resultados, escogemos el más conveniente, el de los **datos estandarizados**; no tanto por el porcentaje de explicación, sino porque en los datos originales, como se mencionó anteriormente existe mucha discrepancia entre los datos, debido a las unidades en las que están medidas; lo cual no permite con claridad dar buenas conclusiones.

Analizamos ahora estas componentes para poder rotularlas. Si observamos Y_1 no tiene valores mayormente significantes, pero en relación con los otros coeficientes de esta componente son los más relevantes; basándonos en esto, las variables son las de superficie sembrada, superficie cosechada y las tres variables de producción de arroz; es decir, que esta componente comprende todo el proceso de producción; de ahí que se podría denominar: **proceso de producción**. Algo similar ocurre con Y_2 ; y estas variables con: exportaciones del arroz, y rendimientos; esto es bastante lógico, pues las exportaciones del arroz dependen del rendimiento de la producción de dicho producto; por esto podemos nombrar esta componente como: **industrialización y comercialización del arroz**. Y_3 es más sencillo nombrarla, pues la única variable relevante es el **clima**.

A continuación se presenta el gráfico de cada componente original versus la componente reducida. Como podemos observar en la figura 4.1 la serie de la componente reducida es muy similar a la serie de la componente

principal; mientras que en la figura 4.2, la serie de la componente reducida es más precisa, pues no tiene tantas variables, ya que se reduce a dos: exportaciones del arroz y rendimientos de la producción del arroz. En la figura 4.3 observamos que el gráfico de la componente reducida aparentemente es una recta; pero se debe a que los valores de esta componente con la principal, son muy discrepantes, ya que al reducir esta componente nos queda sólo en función de la variable clima; por lo que el gráfico de esta serie debería ser igual al de la serie de la variable clima.

Primera componente vs. Componente reducida

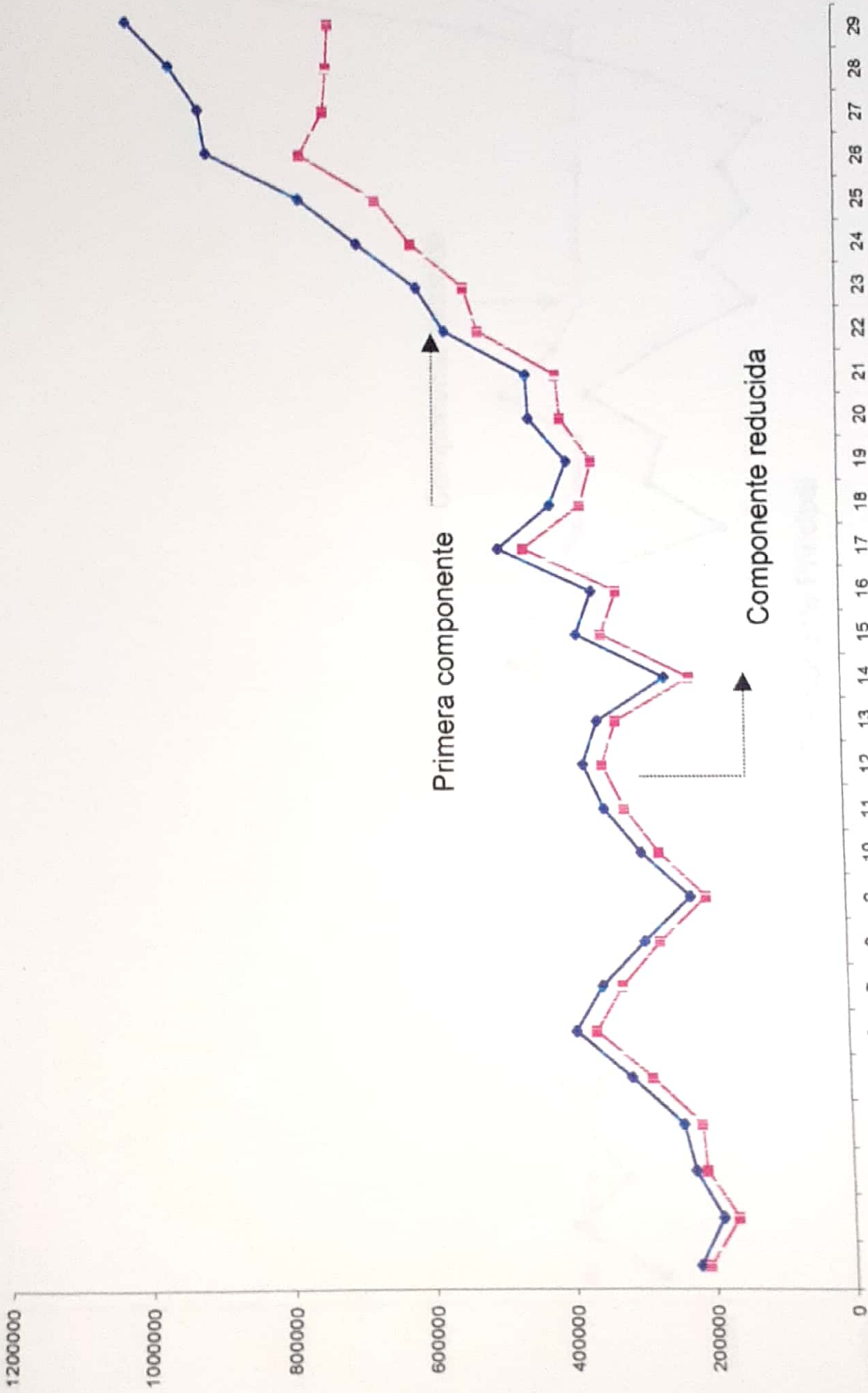


Figura 4.1

Segunda componente vs. Segunda componente reducida

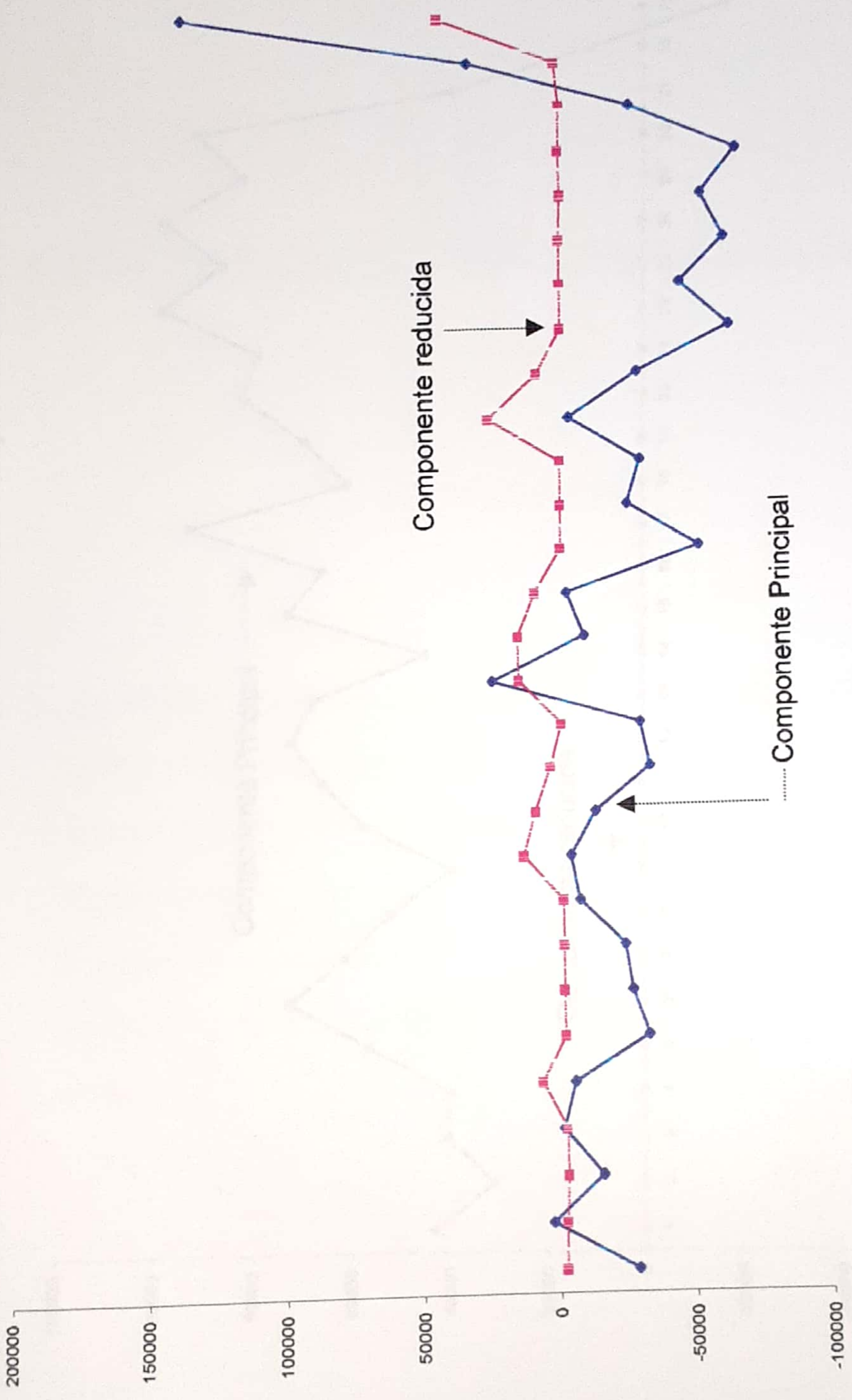


Figura 4.2

Tercera componente principal vs. Tercera componente reducida

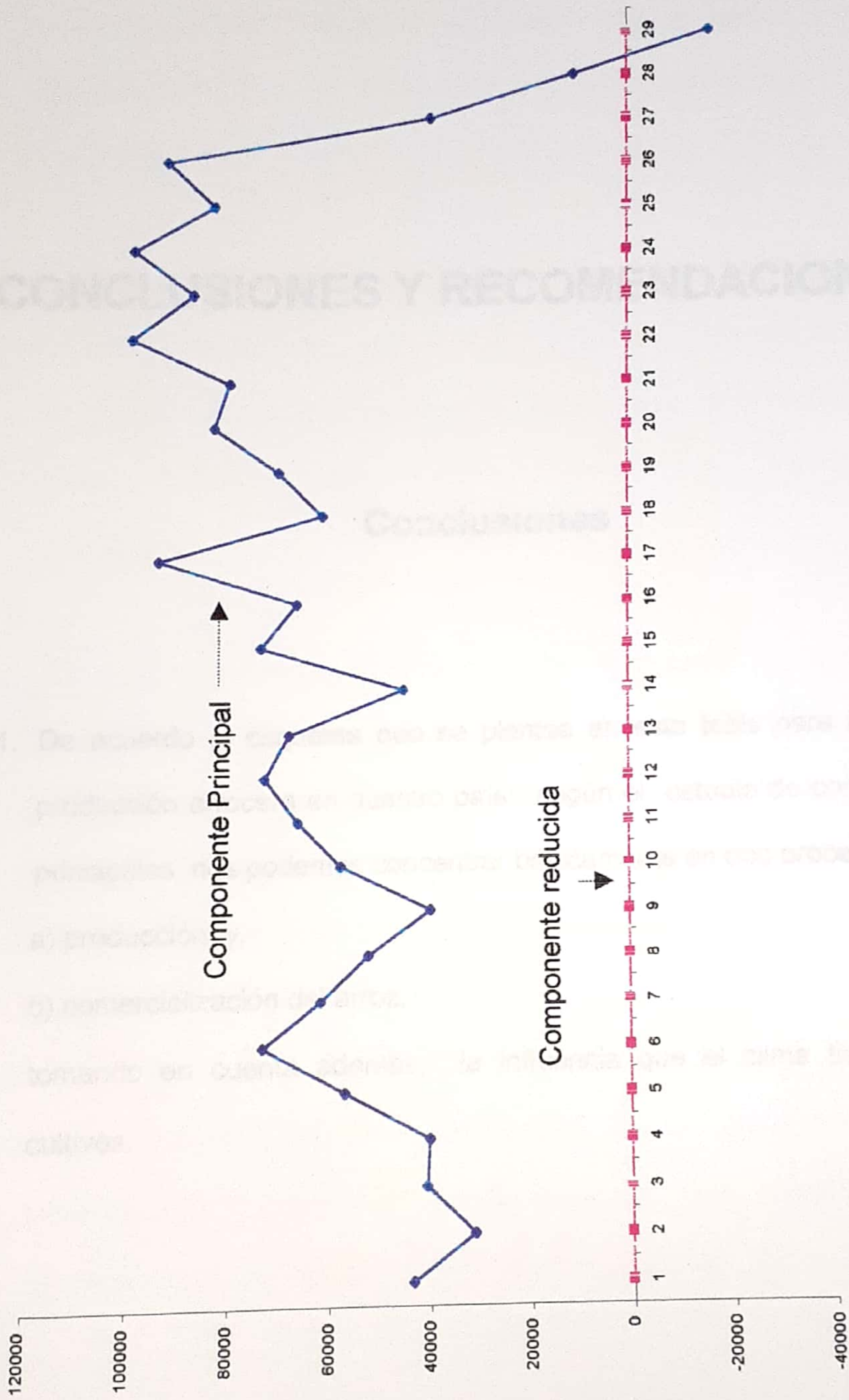


Figura 4.3

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

1. De acuerdo al esquema que se plantea en esta tesis para analizar la producción arrocerá en nuestro país; según el estudio de componentes principales nos podemos concentrar básicamente en dos procesos:
 - a) producción; y,
 - b) comercialización del arroz;tomando en cuenta además, la influencia que el clima tiene en los cultivos.

2. Como era de esperarse; la producción del arroz depende exclusivamente de la cantidad de hectáreas que se siembra y las que se cosechan; es decir, la producción del arroz depende de la superficie sembrada y de la superficie cosechada en un determinado período.
3. De este mismo análisis, podemos concluir que otro factor importante en la producción del arroz, es el clima; medido en porcentaje de humedad, de ahí que en período de invierno exista mayor producción que en el verano, lo cual lo podemos confirmar comparando los datos en el apéndice B.
4. De acuerdo al esquema planteado en el capítulo de componentes principales; las 20 variables iniciales observables, se han reducido a 3 variables no observables; las cuales se las ha rotulado como: proceso de producción, industrialización y comercialización del arroz y clima. Se las ha rotulado así, basándonos en las variables de explicación que contiene cada una.
5. Según el análisis de series temporales, en la proyecciones de las exportaciones del arroz, en los próximos 20 años seguirá la misma tendencia que ha llevado hasta el momento; es decir, disminuirán,

aunque a partir del año 2015 aproximadamente tenderá a estabilizarse alrededor de las 67000 TM.

6. La correlación entre las variables "económicas" (llamadas así debido al estrecho vínculo con la economía del país) y las variables de producción, no es muy alta; excepto en las variables de índice de precios al consumidor y deuda externa que influyen en la cantidad de arroz producido; debido a lo que se refiere a créditos y beneficios de los trabajadores. Recordemos que las variables de producción son: superficie sembrada, superficie cosechada, producción y rendimiento; y las variables económicas son: inflación, importaciones y exportaciones del arroz, deuda externa, salario, precio del dólar y remuneración.
7. Según el análisis univariado, mostrado en el apéndice C; se puede concluir que la mayor cantidad de arroz que se produce es en el período de invierno; de allí que del promedio de producción total del arroz, aproximadamente el 62% corresponda a la producción en invierno y tan sólo el 38% a la producción en verano.
8. También en este análisis se puede observar que la menor cantidad producida de arroz fue de 69181.92 TM. correspondiente al primer ciclo

de cultivo; es decir, en invierno. Esto ocurrió en 1971 (ver figura 3.20); época en la que mayor producción se obtuvo en el verano (ver figura 3.22)

Recomendaciones

9. Además podemos concluir que el rendimiento de producción máximo que ha tenido nuestro país hasta el año 1998 es de 3988.61 has.; que fue en 1994 (ver figura 3.26).
10. Cabe recalcar que para el análisis realizado en esta tesis, no se ha tomado en cuenta otros factores de vital importancia, como son la transferencia tecnológica, el uso de nuevas semillas certificadas que den un mayor rendimiento, entre otros. De aquí que los resultados obtenidos en este trabajo se cumplirían si es que en nuestro país se siguiera llevando la misma tecnología.

Recomendaciones

1. Se recomienda que para un análisis de producción arrocerá, se basen fundamentalmente en dos puntos: el proceso de producción; es decir, desde la superficie que se siembra hasta la industrialización; y segundo, la comercialización de este cereal, tanto interna como externa.
2. De acuerdo con los resultados del análisis de componentes principales, se concluyó que las variables más influyentes en la producción del arroz son las de superficies sembradas y cosechadas, de ahí se desprende lo que es una buena producción, por lo que se recomienda llevar un control adecuado en las siembras, tanto en fertilizaciones, como en control hídrico.

3. Se recomienda manejar adecuadamente las cosechas para el mejor rendimiento del producto agrícola; así se aumentaría su calidad y cantidad, de tal manera que sirva para abastecer no solo a la población ecuatoriana, sino también para exportar.
4. Además se recomienda que no sólo se utilice el arroz como alimento básico; si no también como materia prima para elaborar otros productos que aunque no son de primera necesidad, algunos de ellos como por ejemplo los cereales sirven como suplemento alimenticio.
5. También se recomienda que la mayor cantidad de tierras sembradas se lo haga durante el primer ciclo de cultivo; es decir durante los meses de diciembre a mayo, ya que es en este tiempo en el que el clima es más favorable.
6. Se recomienda a los productores, utilizar semillas certificadas, además de implementar la tecnología adecuada para poder obtener mayores rendimientos en los cultivos.

APENDICE A

HISTOGRAMAS DE FRECUENCIA

Histograma de frecuencias de las exportaciones del arroz

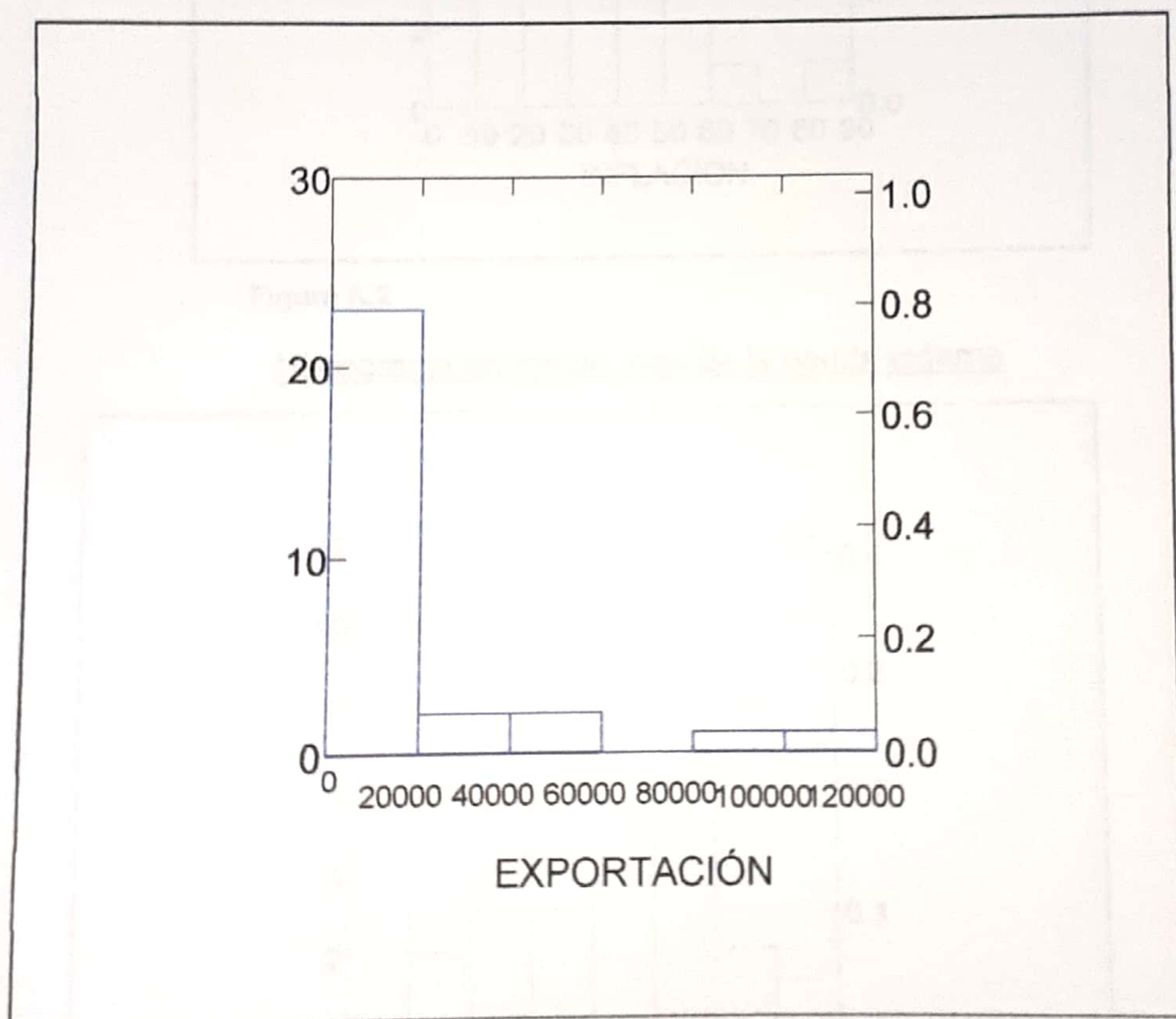


Figura A.1

Histograma de frecuencias del índice de precios al consumidor

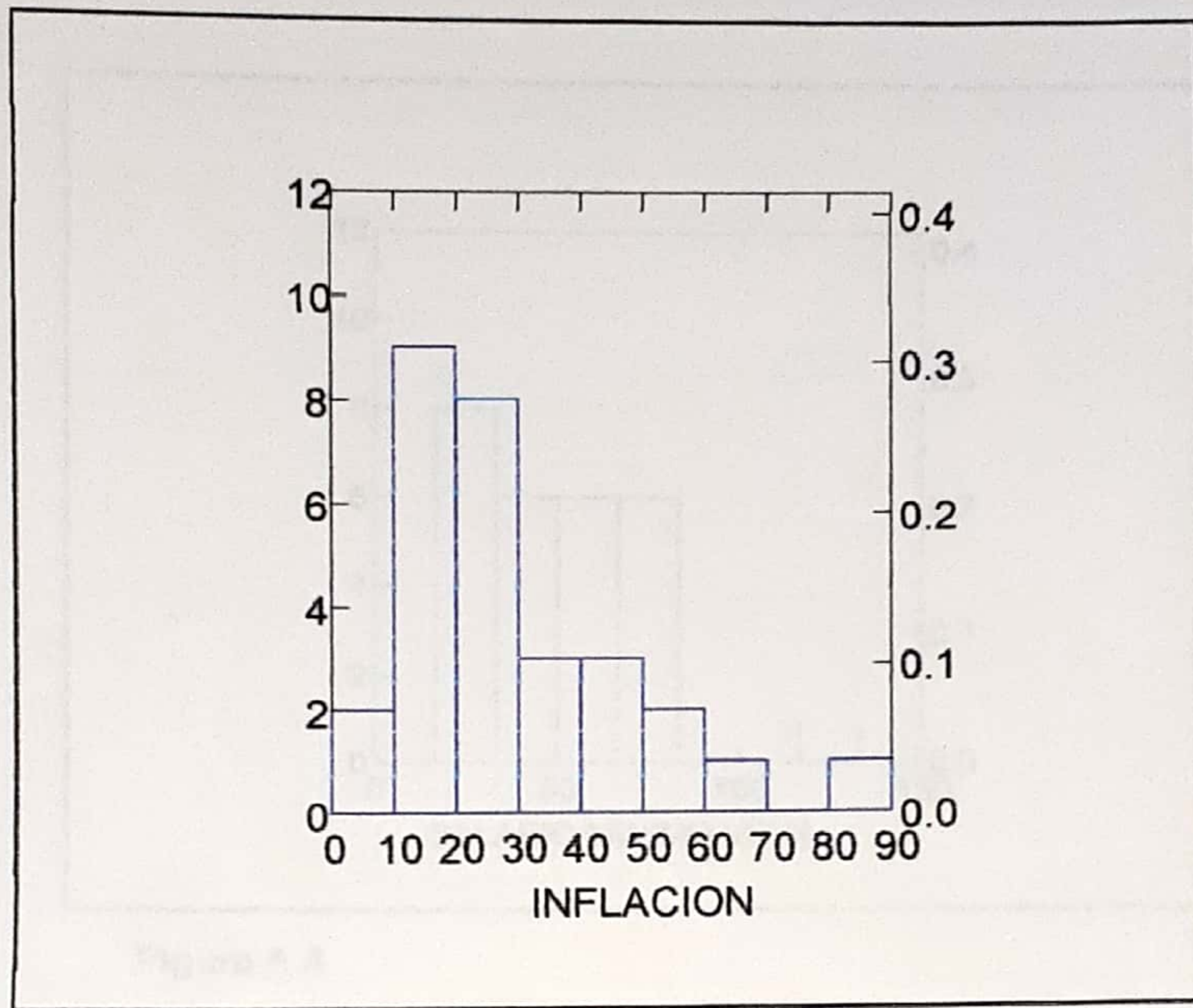


Figura A.2

Histograma de frecuencias de la deuda externa

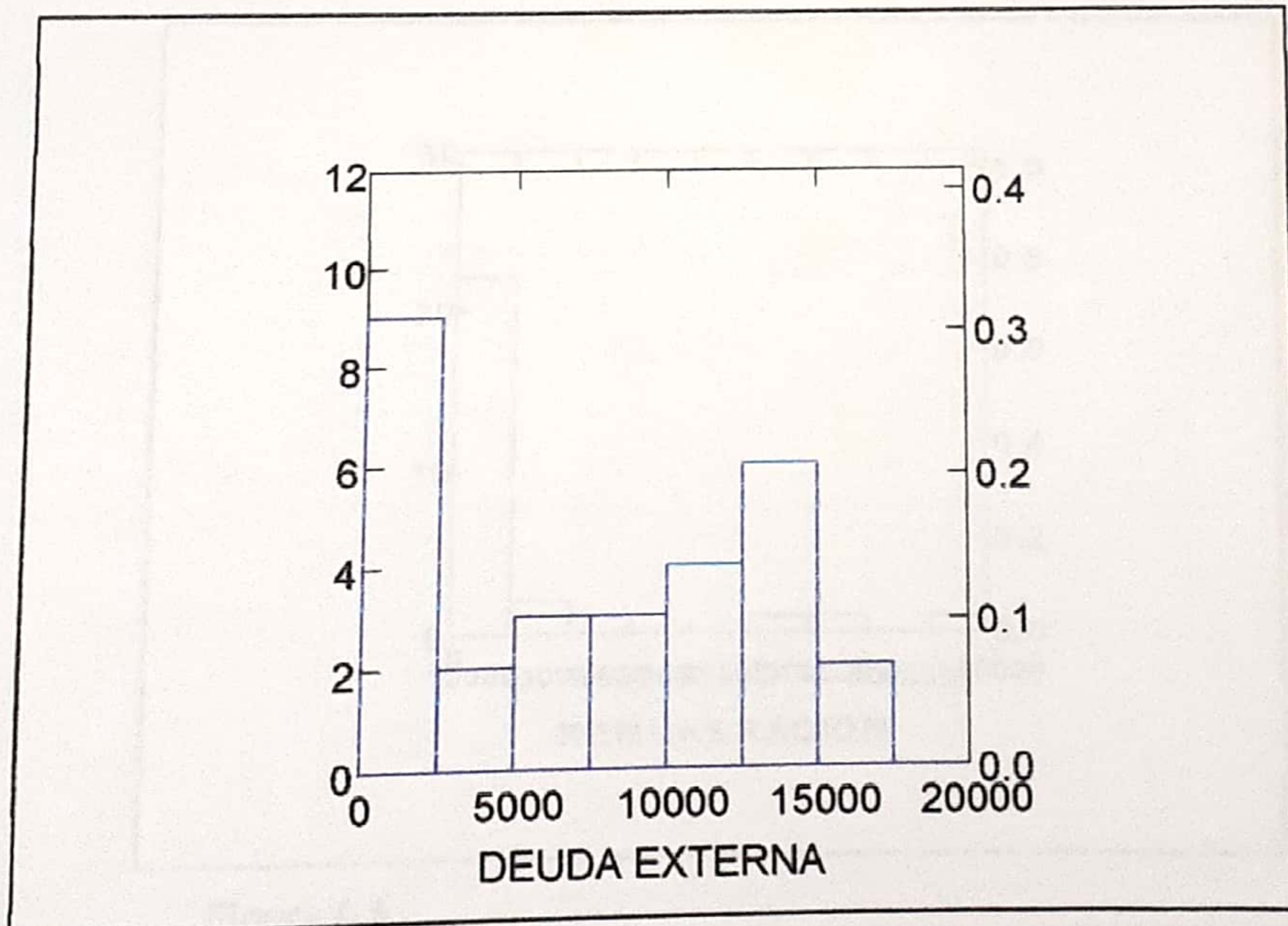


Figura A.3

Histograma de frecuencias del Salario Mínimo Vital

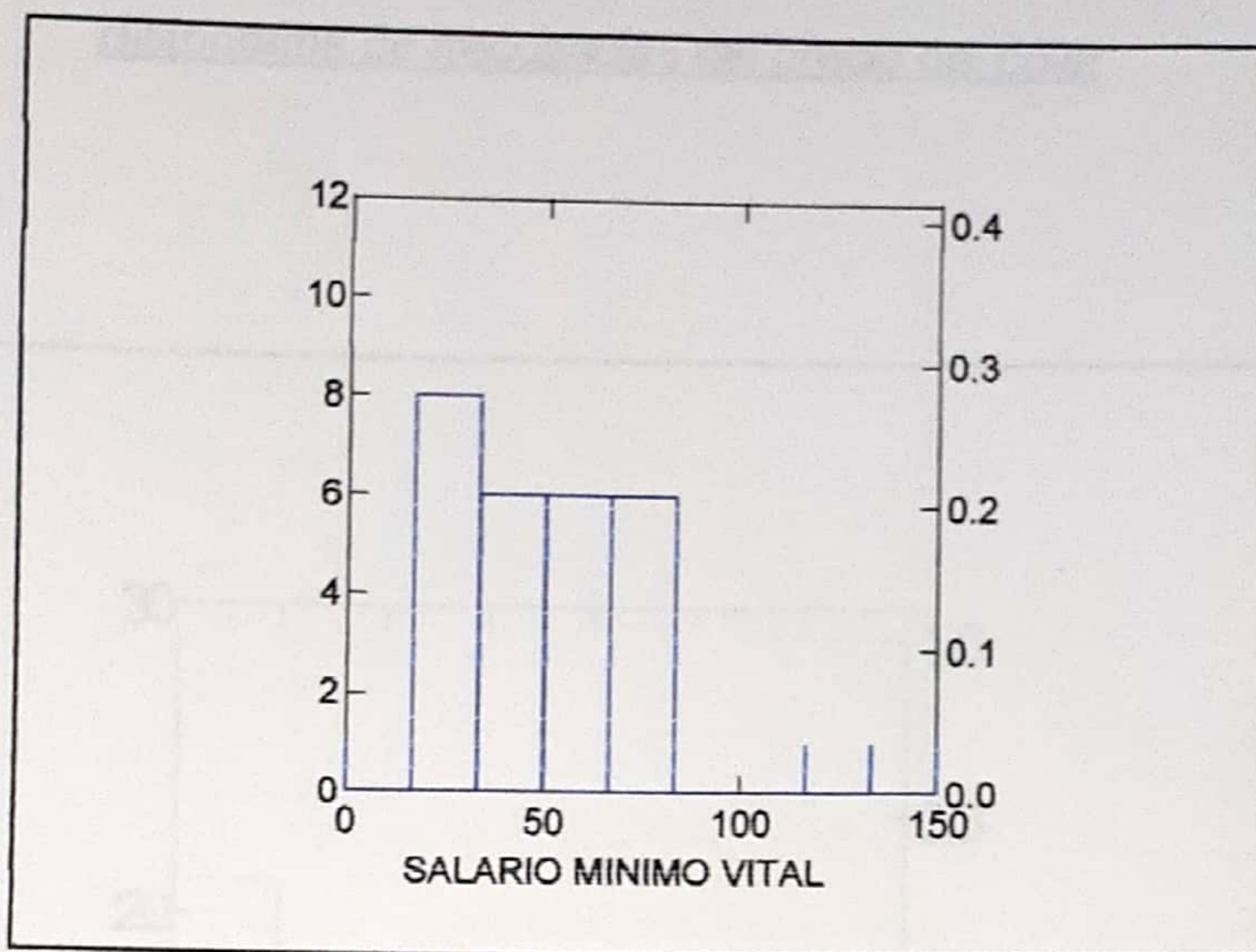


Figura A.4

Histograma de frecuencias de las remuneraciones

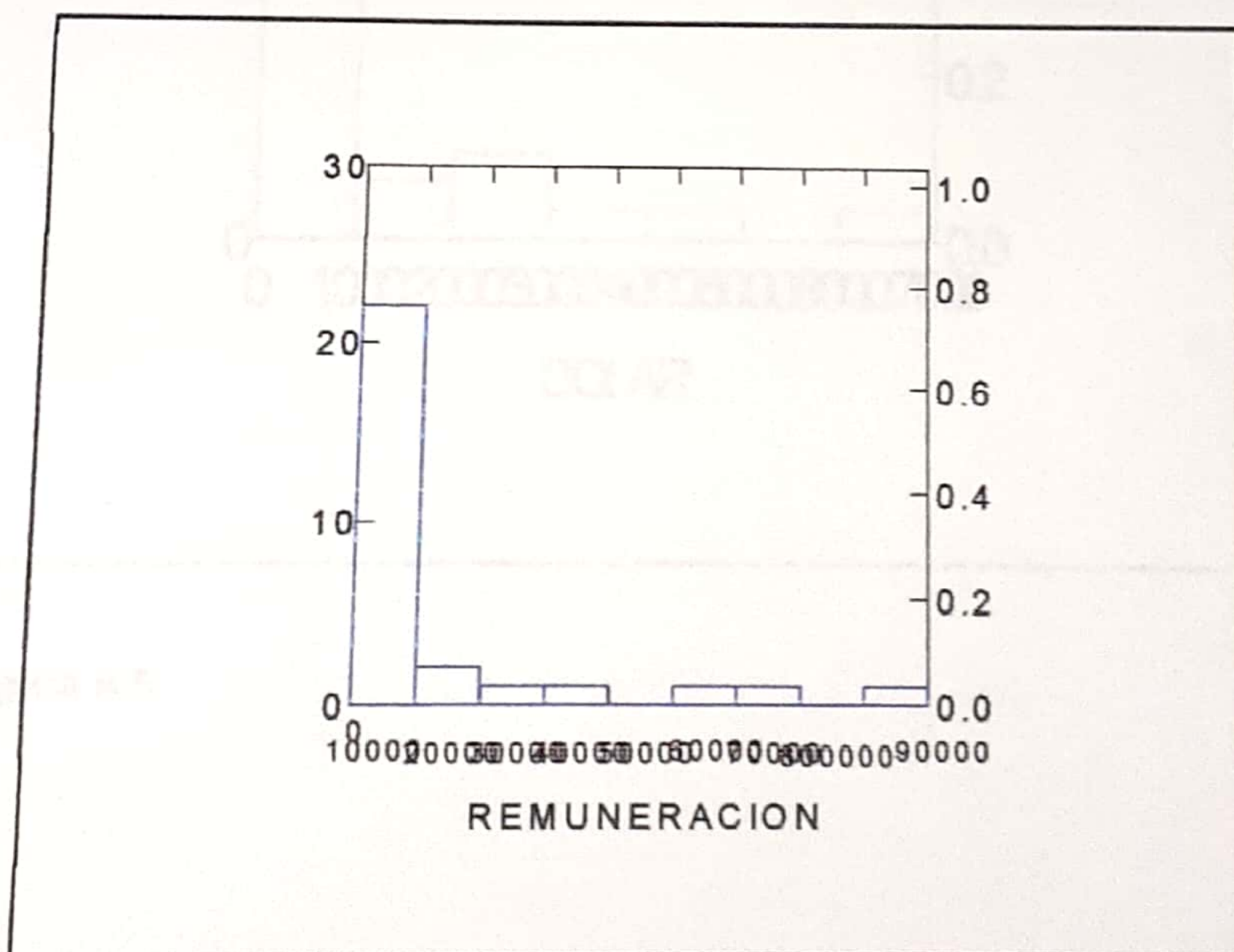


Figura A.5

Histograma de frecuencias del precio del dólar

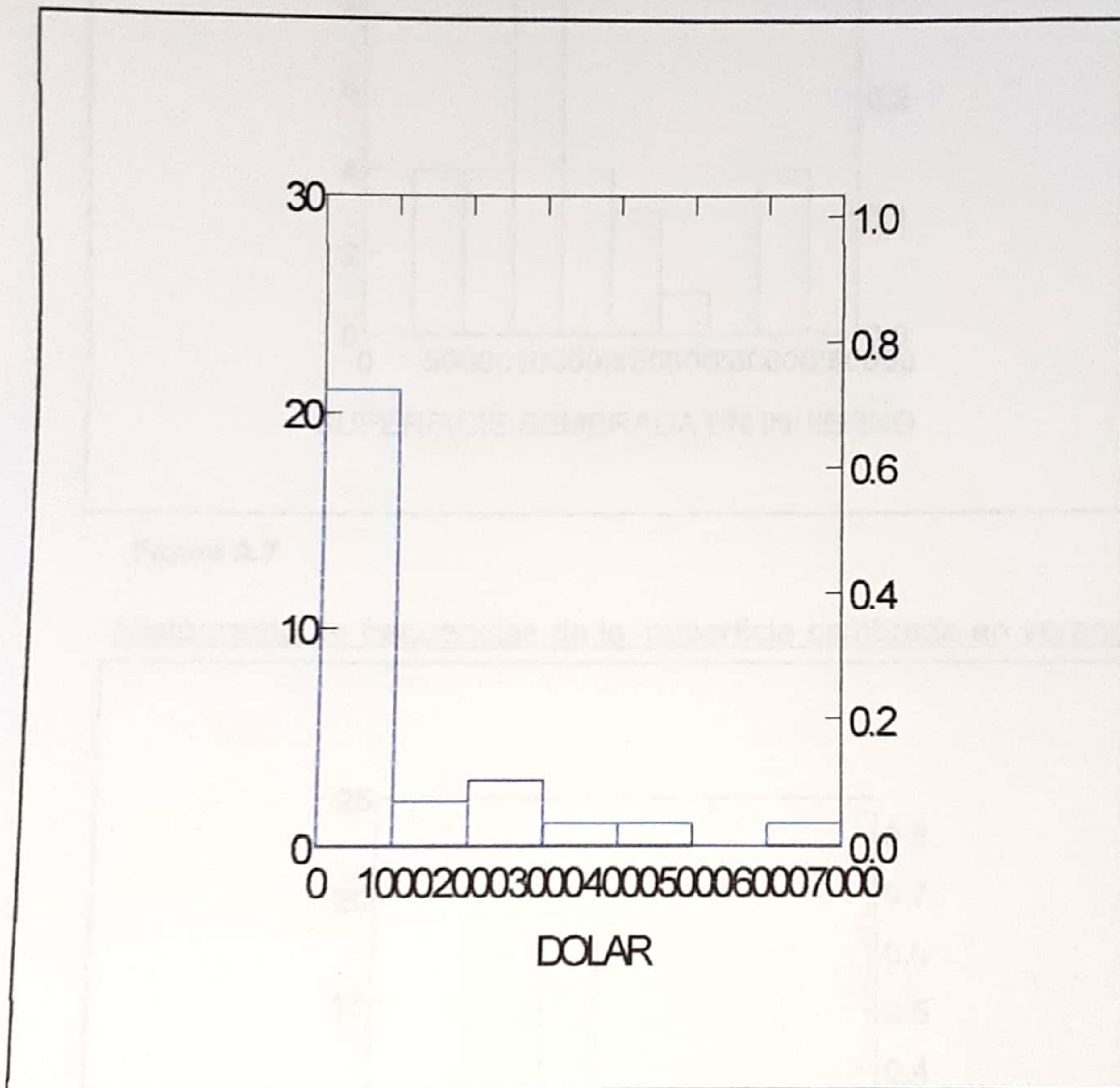


Figura A.6

Histograma de frecuencias de la superficie sembrada en invierno

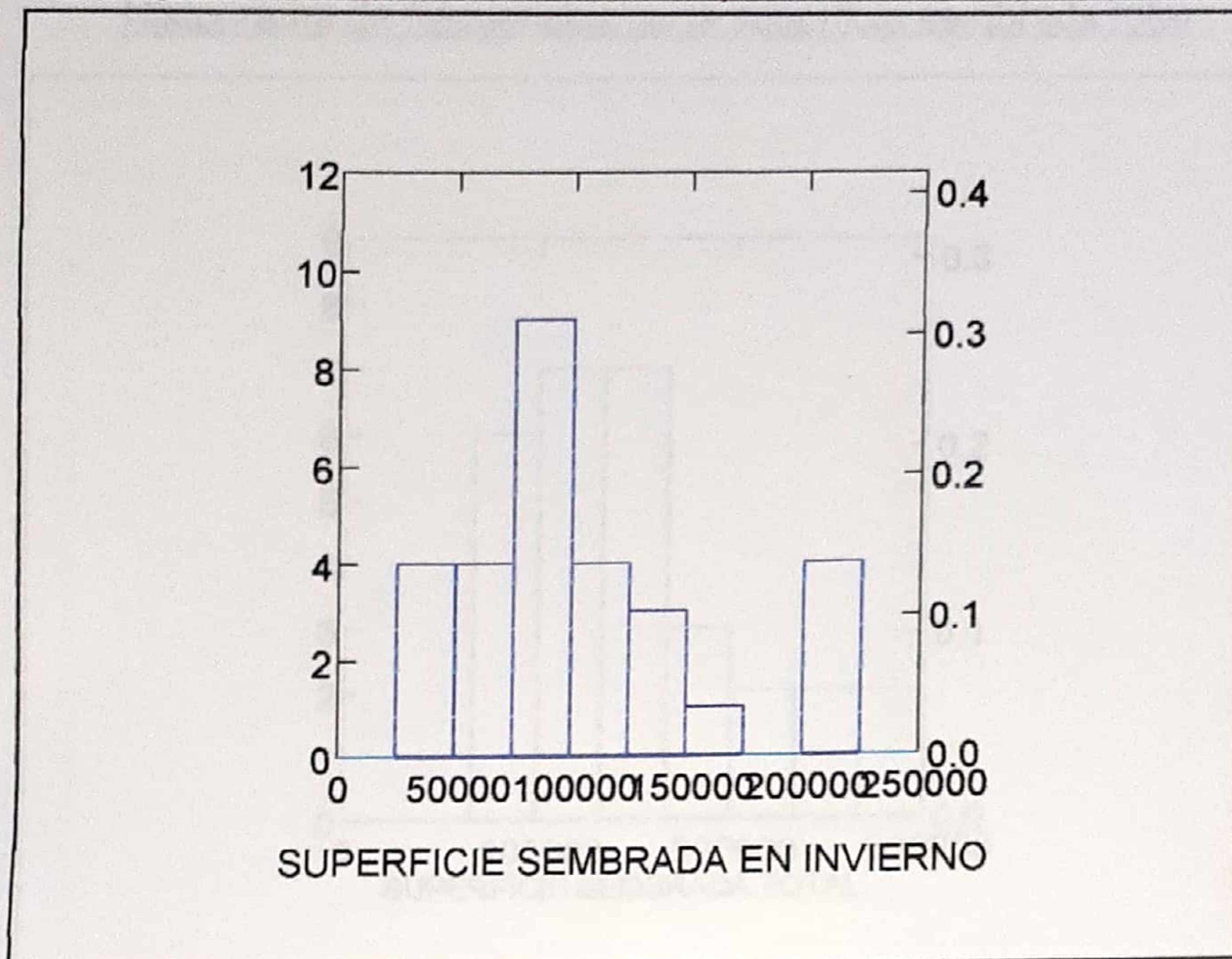


Figura A.7

Histograma de frecuencias de la superficie sembrada en verano

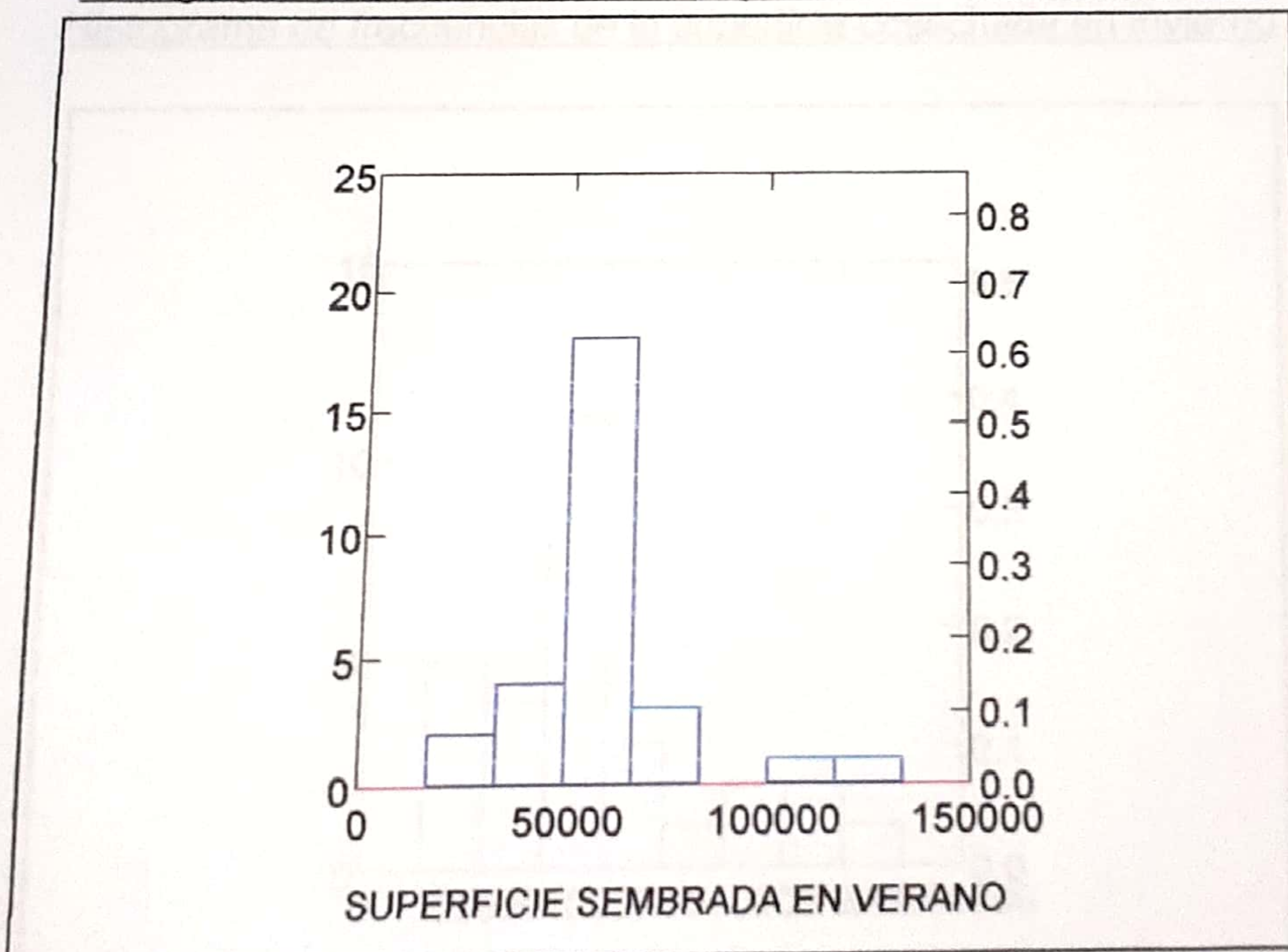


Figura A.8

Histograma de frecuencias de la superficie sembrada total

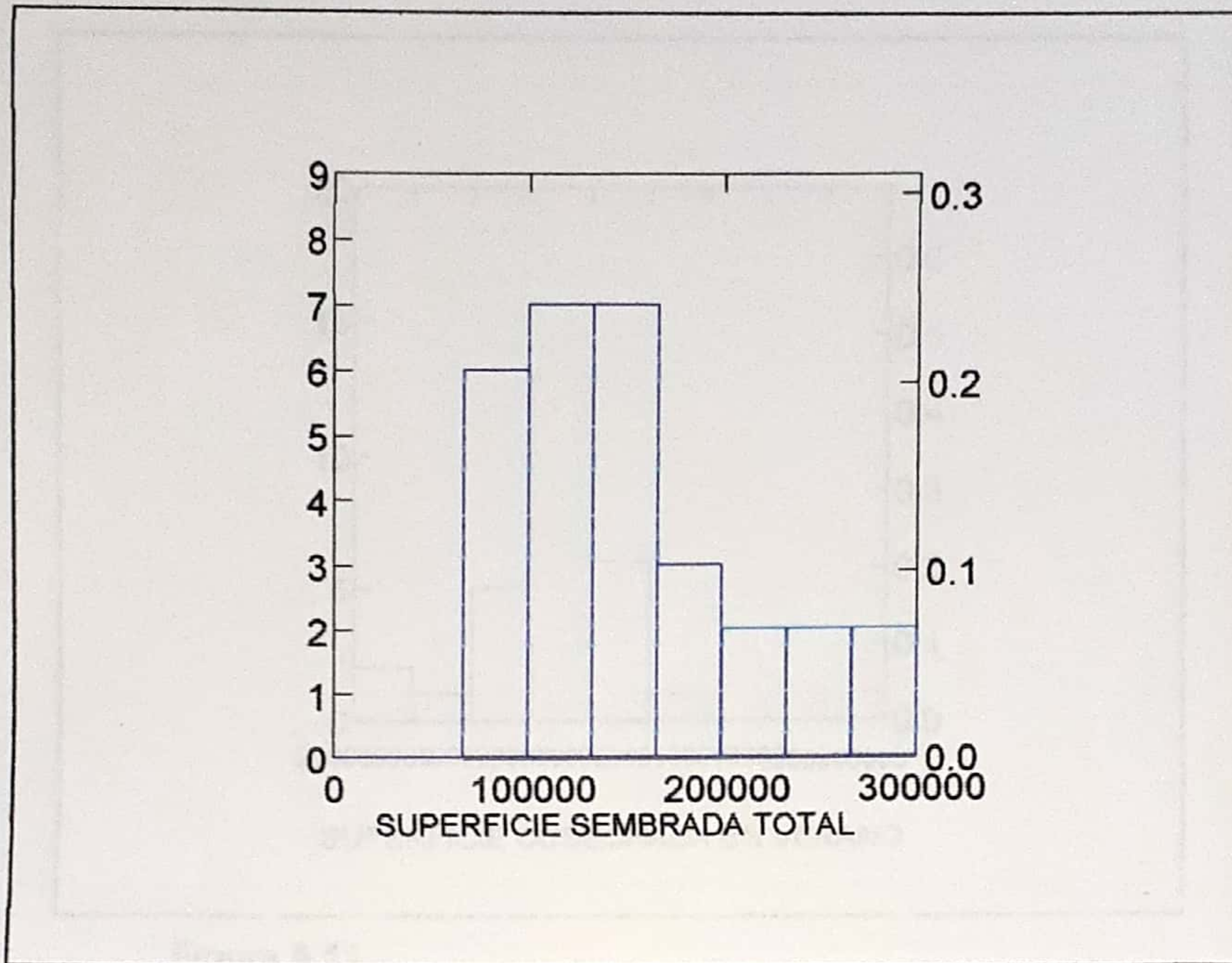


Figura A.9

Histograma de frecuencias de la superficie cosechada en invierno

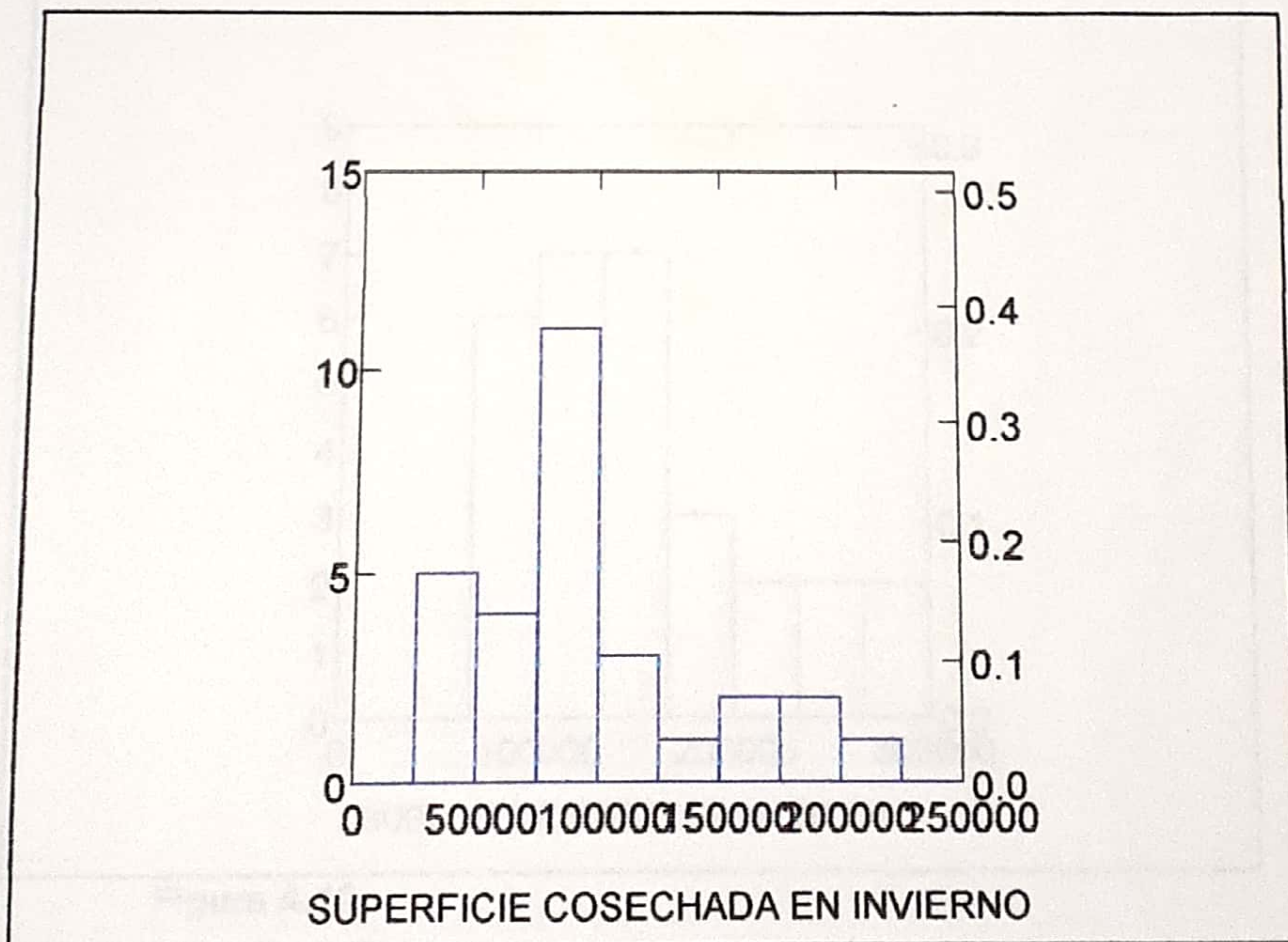


Figura A.10

Histograma de frecuencias de la superficie cosechada en verano

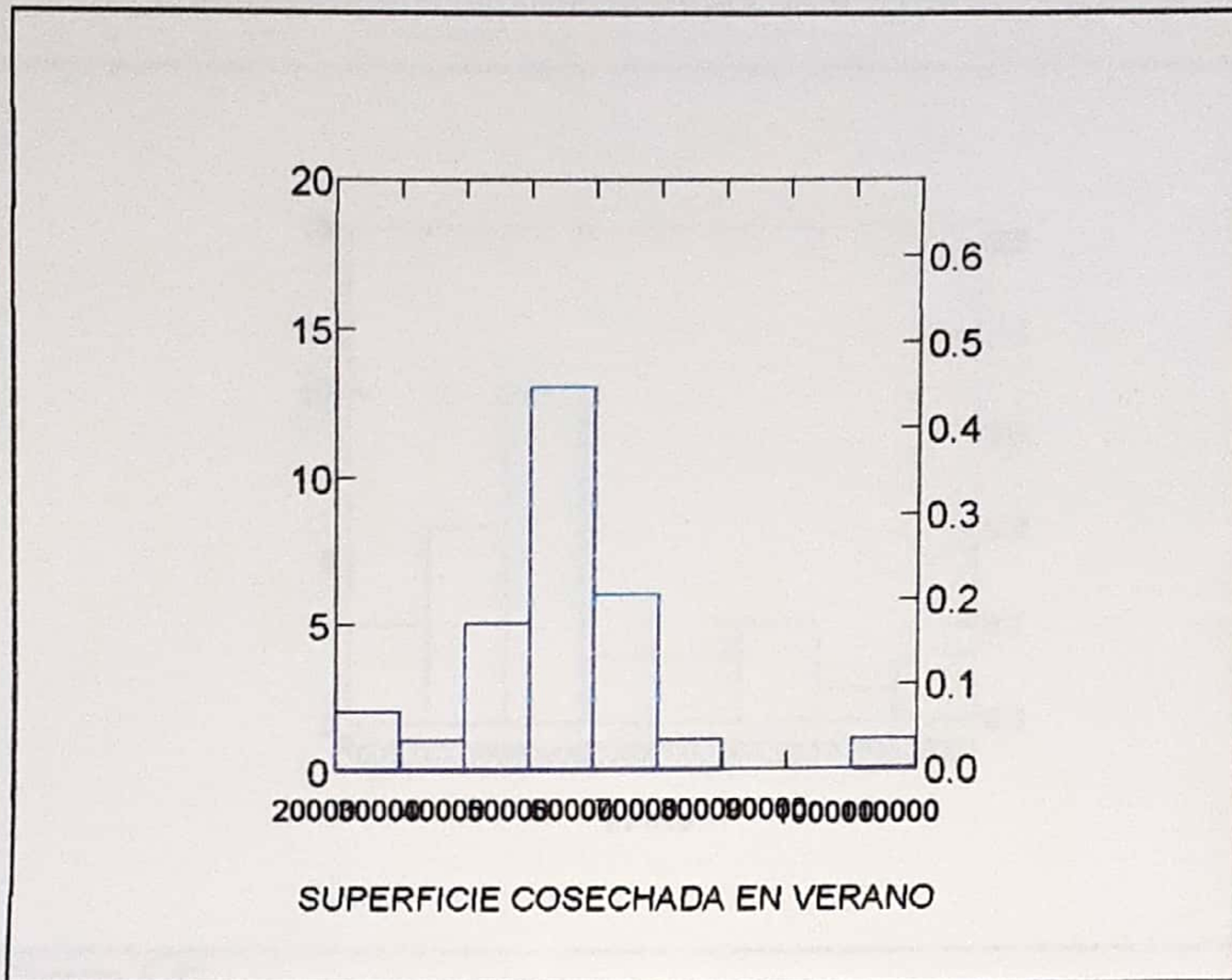


Figura A.11

Histograma de frecuencias de la superficie cosechada total

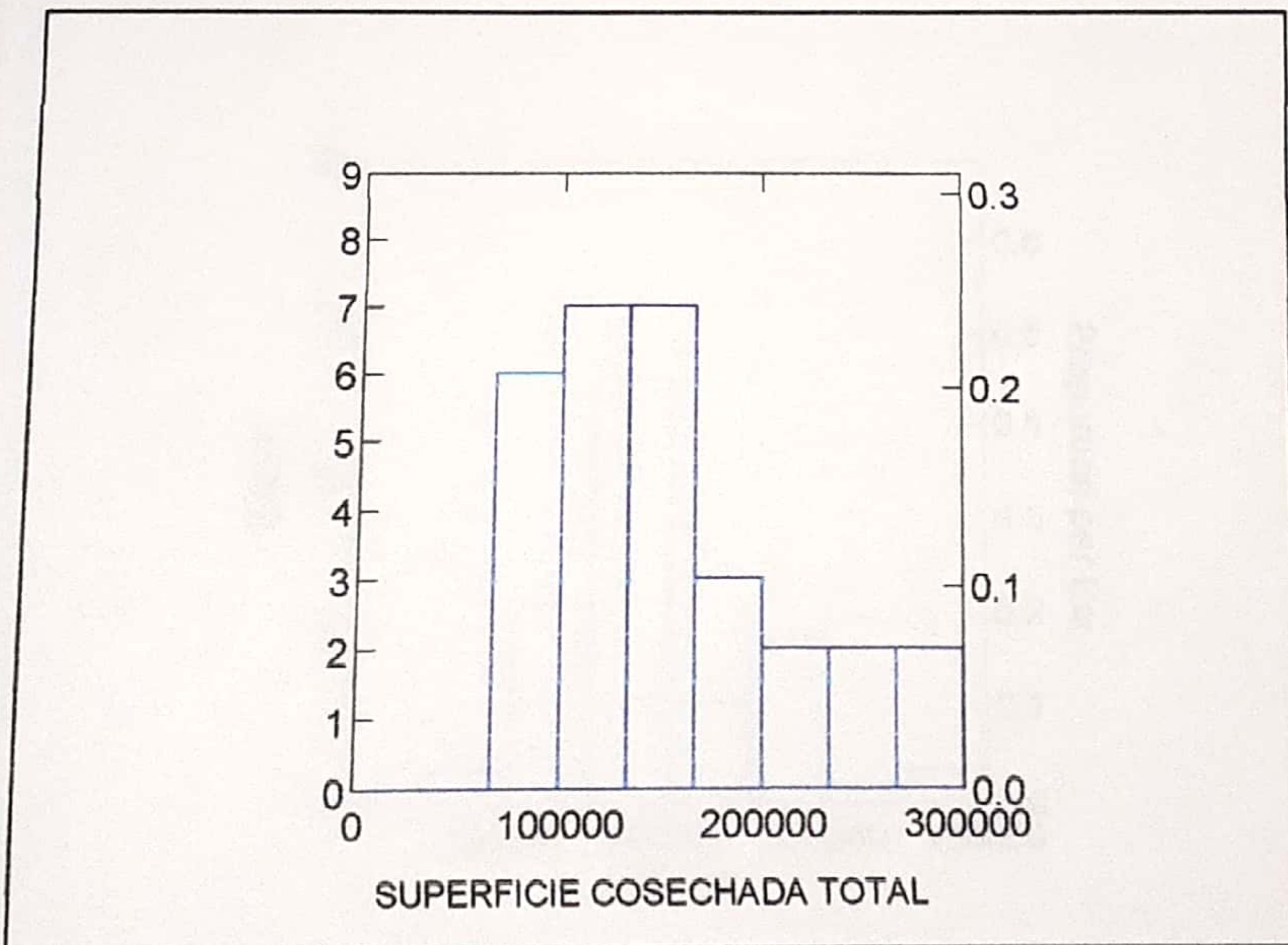


Figura A.12

Histograma de frecuencias de la producción de arroz en invierno

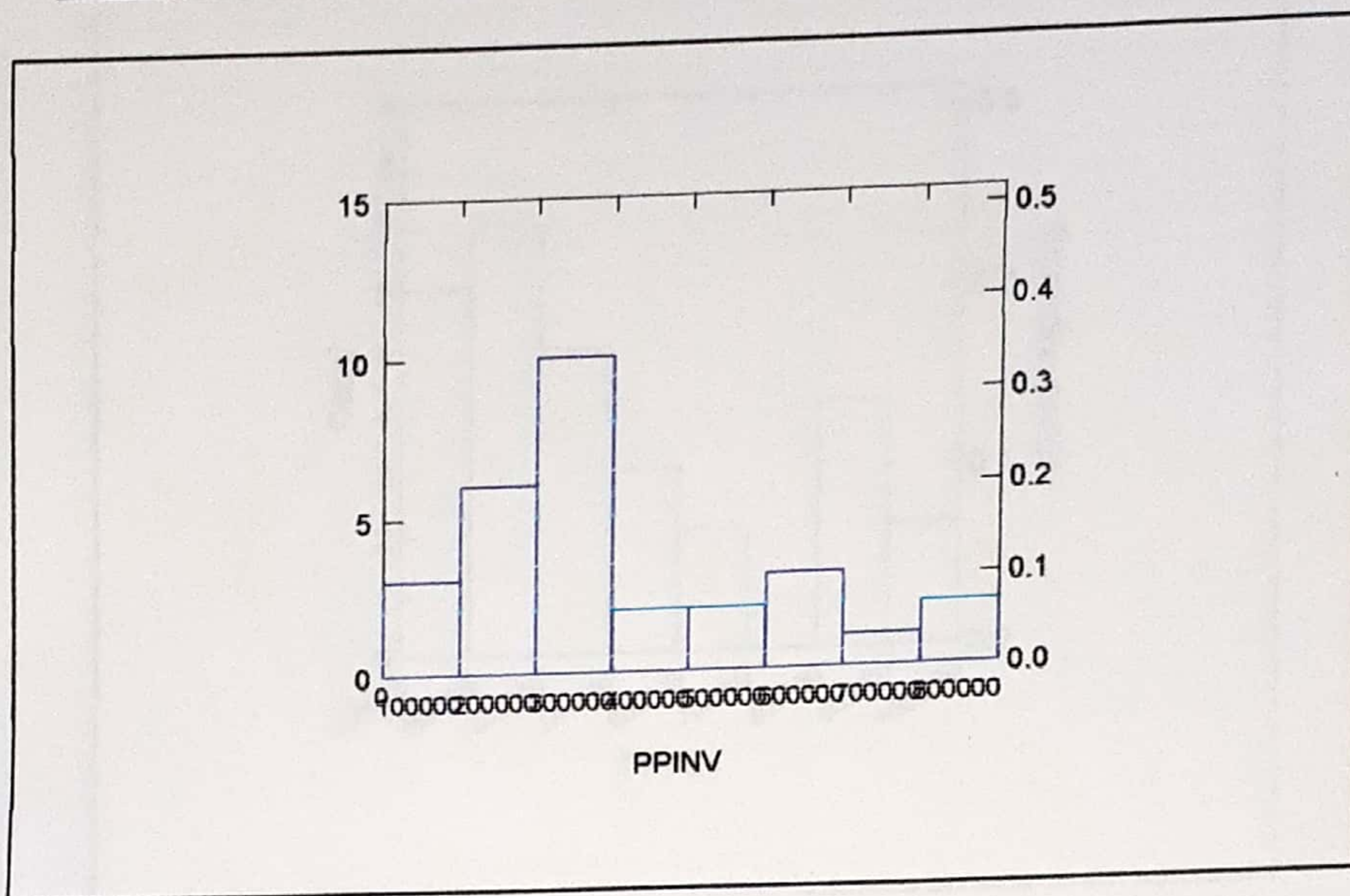


Figura A.13

Histograma de frecuencias de la producción de arroz en verano

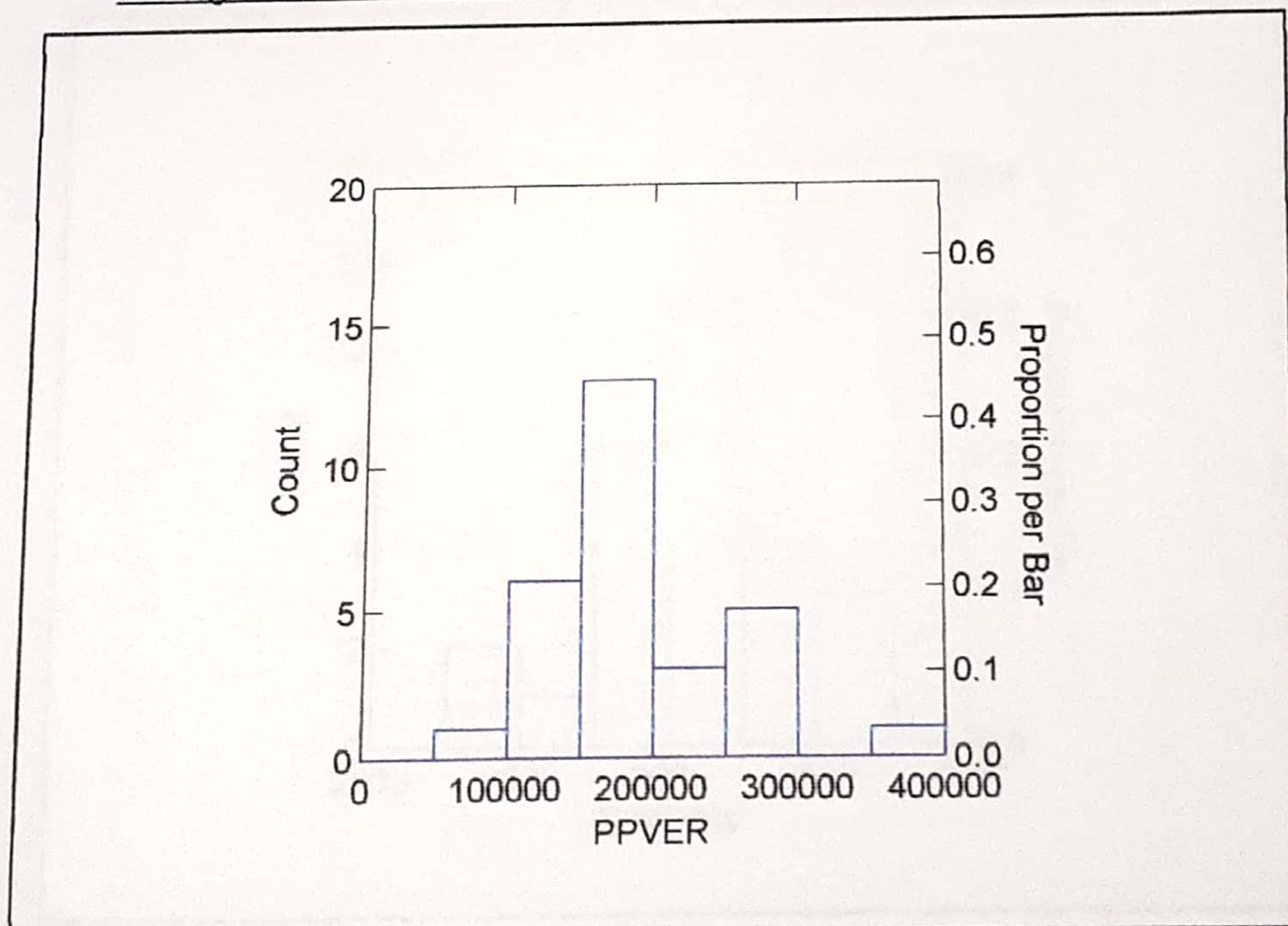


Figura A.14

Histograma de frecuencias de la producción de arroz total

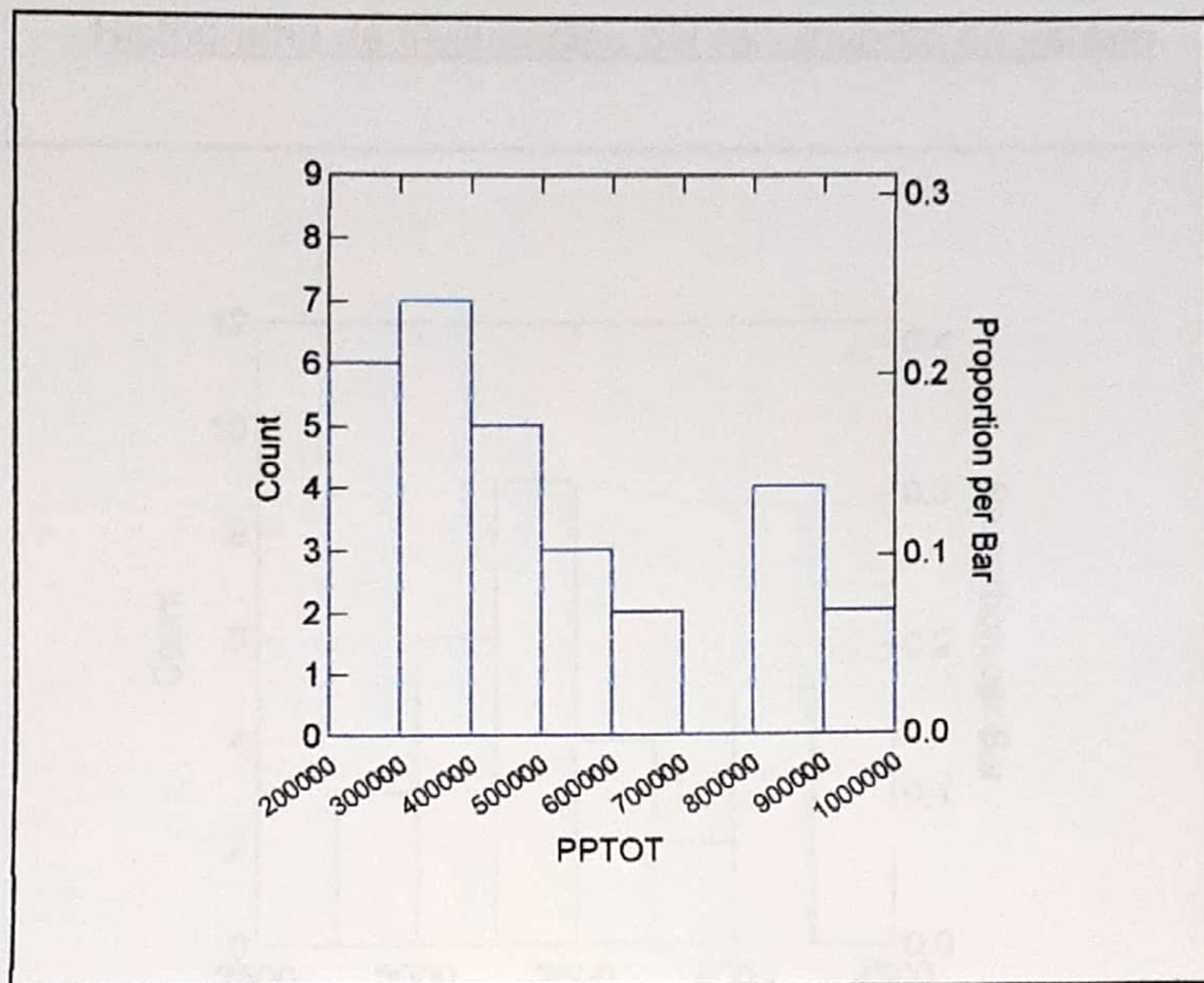


Figura A.15

Histograma de frecuencias del rendimiento en invierno

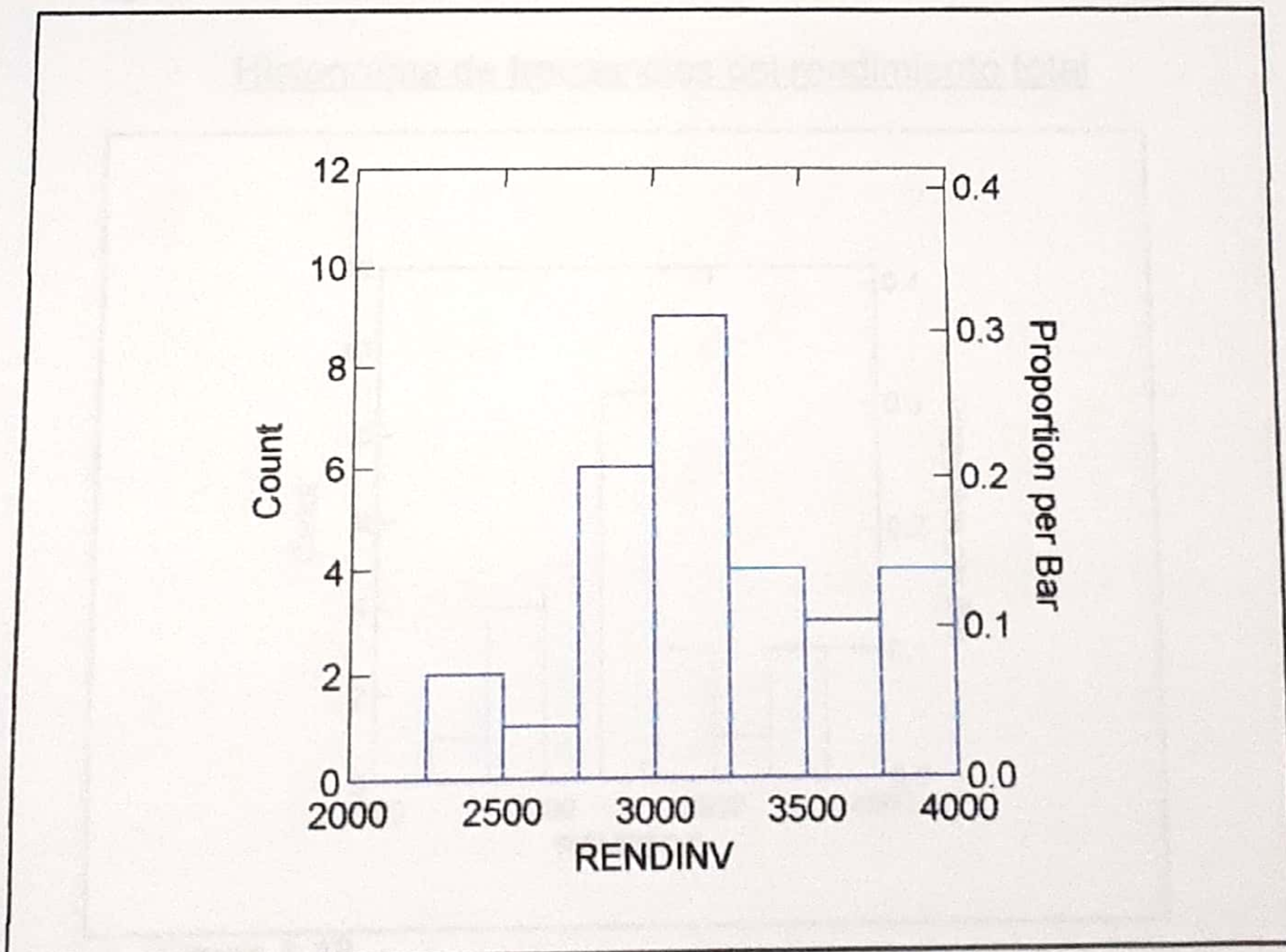


Figura A.16

Histograma de frecuencias del rendimiento en verano

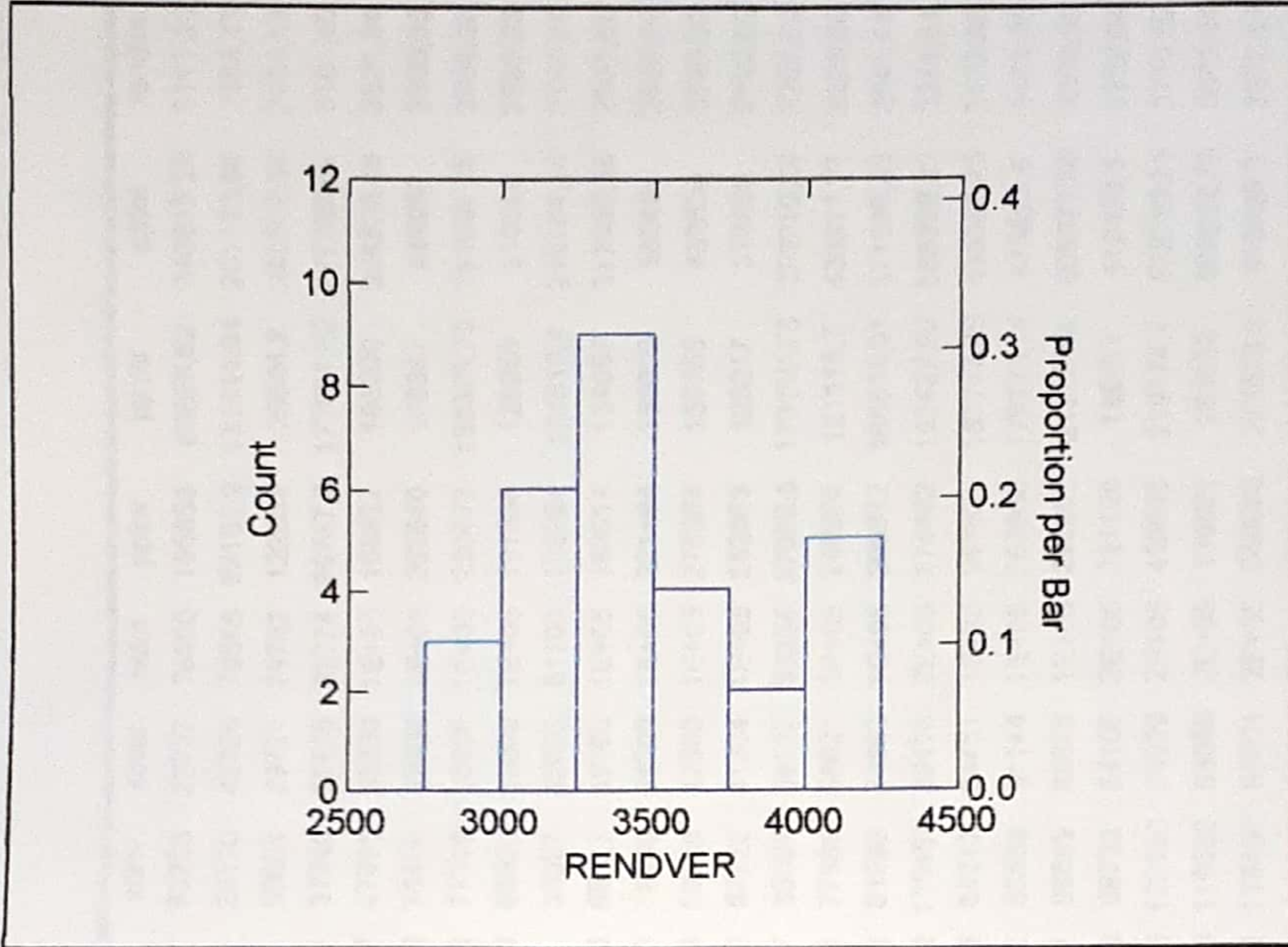


Figura A.17

Histograma de frecuencias del rendimiento total

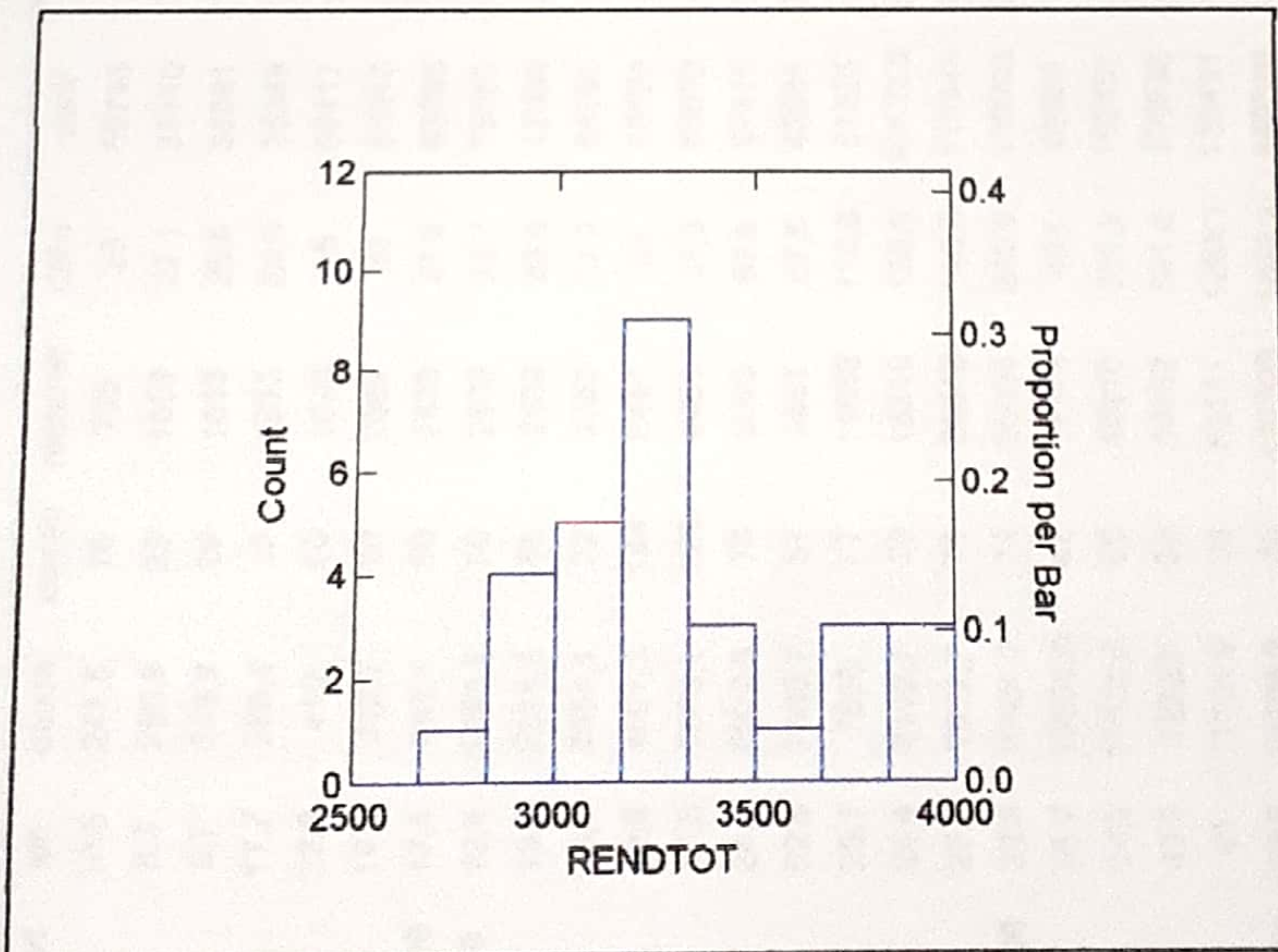


Figura A.18

Matriz de datos

import	export	ipc	deuda	salario	remuner	dólar	ssinv	ssver	sstot	scinv	scver	sctot	ppinv	ppver	pptot	rendinv	rendver	rendtot	clima
1970	1400	50	241,5	26	795	25	59716	22797	82513	53203	22797	76000	165850	80825,62	246675,29	3117,30	3545,45	3245,73	74
1971	0	0	260,8	29	1009	27,1	33410	42326	75736	28190	42326	70516	69181,9	131941,94	201123,86	2454,13	3117,28	2852,17	75
1972	0	0	343,9	29	1013	25,8	55061	32720	87781	50061	29721	79782	122858	125504,8	248362,64	2454,16	4222,77	3113,02	77
1973	70	0	380,4	30	1013	24,8	35349	51425	86774	31349	51425	82774	99747,8	177651,82	277399,59	3181,85	3454,58	3351,29	79
1974	19500	0	410	40	1620	25	68417	56000	124417	47891	53200	1E+05	168627	186200	354826,99	3521,06	3500,00	3509,98	75
1975	0	0	512,7	50	2067	26	86240	60000	146240	78400	57000	1E+05	262640	186347	448987	3350,00	3269,25	3316,00	77
1976	0	14000	693,1	56	2439	27,2	83090	52000	135090	77340	50000	1E+05	230473	158828,75	389301,75	2980,00	3176,58	3057,18	78
1977	0	7300	1263,7	55	2439	26,7	75000	40000	115000	65000	38000	1E+05	191050	128250	319300	2939,23	3375,00	3100,00	74
1978	0	0	2314,2	56	2539	26,8	42000	44000	86000	39037	42063	81100	123294	124810,4	248104,14	3158,38	2967,23	3059,24	73
1979	32000	0	3554,1	72	3352	27,2	68765	47000	115765	65327	45000	1E+05	183014	134067	317080,86	2801,50	2979,27	2874,01	72
1980	22302	0	4601,3	144	6451	28	80100	54800	134900	75200	51300	1E+05	221464	159030	380494	2945,00	3100,00	3007,86	72
1981	10000	0	5868,1	128	6451	33,3	85570	50709	136279	82600	47500	1E+05	275884	156750	432634	3340,00	3300,00	3325,40	72
1982	0	0	6632,8	78	8026	62,9	97416	55084	152500	87722	41304	1E+05	252869	125217	378086	2882,62	3031,60	2930,31	71
1983	35000	0	7380,7	57	9451	87,8	43245	64301,8	107547	35097	58137	93234	97099,5	173110,73	270210,24	2766,59	2977,63	2898,19	80
1984	35991	0	7596	67	11686	118,9	81325	61916	143241	77895	58021	1E+05	249696	181115,7	430811,77	3205,56	3121,55	3169,70	74
1985	21022	0	8110,7	70	15319	125,6	93623,2	66228,6	159852	81099	54883	1E+05	206072	165514,31	371586,43	2541,01	3015,78	2732,63	71
1986	0	0	9062,7	71	20642	145,6	119948	60063,5	180012	116451	59176	2E+05	374462	192427,92	566889,82	3215,61	3251,82	3227,81	73
1987	0	45000	10335,5	74	25094	246,9	109093	54582,3	163675	84347	54451	1E+05	267621	187747,39	455368,75	3172,86	3447,99	3280,79	76
1988	0	0	10668,8	42	31318	499	88960	54071	143031	85559	51144	1E+05	269850	175773,74	445623,4	3153,96	3436,84	3259,79	72
1989	60000	0	11532,6	50	40310	661,2	102455	60876	163331	88599	60876	1E+05	292852	213120,3	505971,89	3305,36	3500,89	3384,99	75
1990	19219	0	12222	39	50832	891,6	106533	54106	160639	98738	54106	2E+05	321100	189371	510470,5	3252,04	3500,00	3339,81	71
1991	0	50	12801,9	36	69111	1284,1	130431	58029	188460	122136	58029	2E+05	453616	219174,1	672789,75	3714,02	3776,98	3734,30	71
1992	640	12372	12795,2	40	127333	1873,5	140369	65069	205438	118582	65069	2E+05	435601	263052	698652,75	3673,41	4042,66	3804,24	76
1993	1101	36	13630,9	34	177375	2014,5	149300	65201	214501	139450	65201	2E+05	540065	273523,3	813588,3	3872,82	4195,08	3975,49	75
1994	251	31700	14589,4	32	277139	2298,1	154230	68605	222835	153885	68605	2E+05	596088	291337,15	887425,05	3873,59	4246,59	3988,61	75
1995	1051	26512	13934	33	379063	2914,8	207000	70000	276855	198640	64895	3E+05	748128	251039,25	999167,4	3766,25	3868,39	3791,40	78
1996	255	92728	14586,1	30	513681	3596,2	210000	110200	280000	180418	60200	2E+05	700881	249562,2	950442,82	3884,76	4145,55	3950,01	77
1997	2844	108519	15099,2	26	637590	4393	206840	110200	317040	200220	79936	3E+05	625965	268966	894931,25	3126,39	3364,77	3194,40	77
1998	99737	54154	16399	16	890257	6480	207437	117030	324467	172962	105165	3E+05	546961	352877,4	899838,05	3162,32	3355,46	3235,35	78

APÉNDICE C

ANÁLISIS UNIVARIADO DE LAS VARIABLES DE PRODUCCIÓN Y RENDIMIENTO DEL ARROZ

Producción del arroz en invierno				
Media	Desviación estándar	Varianza	Máximo	Mínimo
313551,98	189288,08	35829978819	748128,15	69181,92

Producción del arroz en verano				
Media	Desviación estándar	Varianza	Máximo	Mínimo
190452,99	60719,32	3686835397	352877,4	80825,62

Producción total del arroz				
Media	Desviación estándar	Varianza	Máximo	Mínimo
504004,98	241601,59	58371326176	999167,4	201123,86

Rendimiento de la producción de arroz en invierno				
Media	Desviación estándar	Varianza	Máximo	Mínimo
3200,41	402,57	162064,64	3884,76	2454,13

Rendimiento de la producción de arroz en verano				
Media	Desviación estándar	Varianza	Máximo	Mínimo
3458,17	400,55	160440,89	4246,59	2967,23

Rendimiento total de la producción de arroz				
Media	Desviación estándar	Varianza	Máximo	Mínimo
3300,33	349,26	121981,93	3988,61	2732,63

GLOSARIO

Arrocillo.- Resultado de la fragmentación del grano en el proceso de pilado.

Cotiledón.- Reservorio de las sustancias nutritivas.

Desgrane.- Separar el grano de la espiga.

Deuda Externa.- Conjunto de obligaciones de un país contraídas con entidades residentes en el exterior, que se derivan de las operaciones de crédito internacional.

Drenaje.- Evacuación de las aguas en exceso que hay en el suelo de un cultivo.

Evapotranspiración.- Suma de un volumen de agua utilizada en el crecimiento vegetativo para la transpiración del tejido de la planta en un área

dada y el volumen evaporado de los suelos adyacentes, por niveles o precipitaciones en determinado tiempo.

Exportación.- Venta de bienes o servicios que realiza un país a un mercado extranjero.

Fisiología.- Capacidad que tiene una planta de obtener resistencia o tolerancia a ciertos factores ambientales.

Fitopatología.- Es el estudio de las enfermedades de las plantas.

Fitotécnia.- Descripción de medios y formas que permiten a los vegetales obtener una mayor producción según las exigencias ambientales.

Gramíneas.- Todas aquellas especies vegetales cuyas hojas son finas y poseen un solo cotiledón.

Importación.- Compra de algún bien o servicio que se realiza en un mercado extranjero

IPC.- Índice de precios al consumidor.- es la medida de la inflación.

Inflación.- Aumento persistente en el nivel de precios de los bienes y servicios de una economía

Morfología.- Descripción de las partes de una planta.

Napa.- Nivel del agua en las capas del suelo

Ph.- Proporción de hidrogeno necesario para el cultivo del arroz.

Polvillo.- Resultado de la pulida del grano del arroz.

Producción.- Cantidad total de producto primario obtenido de un cultivo.

Remuneración.- Cantidad de dinero que recibe una persona por los servicios que presta a través de un contrato de trabajo.

Rendimiento.- Cantidad de producto (expresado en toneladas métricas) dividida entre el número de hectáreas que se produjeron

Salario Mínimo Vital.- Cantidad mínima establecida por las leyes de un país que se le debe pagar a un empleado

Superficie sembrada.- Cantidad de suelo ocupada por semillas o plantas destinadas a la producción agrícola en un periodo de referencia.

Superficie Cosechada.- Cantidad de suelo que luego de haber sido sembrado o plantado ha sido cosechado, generando productos agrícolas dentro del periodo de referencia.

BIBLIOGRAFÍA

1. Vergara Benito S., Manual para el nuevo arrocero, Surco Editores Agrícolas, Quito – Ecuador, 1985
2. Centro de Investigaciones Universidad de especialidades Espíritu Santo, Maximizando la producción del arroz, guayaquil, 1999 – 2000.
3. Topolanski Eugenio, El arroz. Su cultivo y producción, 1985
4. .Box G, Jenkins G., Prinsel G., Time Series Analysis.- Forecasting and Control, Prentice Hall, New Jersey, 1994
5. Mendenhall Willian, Wackerly D., Scheaffer R., Estadística Matemática con Aplicaciones, Segunda edición, Grupo editorial Iberoamérica.

6. Johnson R., Wichern Dean, Applied multivariate Statistical Analysis, Fourth edition, Prentice Hall, New Jersey, 1998
7. Hair, Anderson, Tatham, Black, Multivariate Data Analysis, fifth edition, Prentice Hall, New Jersey, 1998
8. <http://www.sica.gov.ec>
9. Ministerio de Agricultura, Superficie sembrada, superficie cosechada, producción del arroz, Ecuador, 1997
10. Banco Central del Ecuador, Boletín anual 1997.
11. Ministerio de Agricultura, edición a publicar.
12. Manual de Systat 6.0