



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

FACULTAD DE CIENCIAS

LA PIÑA EN LA CUENCA DEL PAPALOAPAN  
(1990-2000): UN ESTUDIO ESTADÍSTICO  
A TRAVES DE LA REGRESIÓN LINEAL PARA  
IMPULSAR SU DESARROLLO

T E S I S  
QUE PARA OBTENER EL TÍTULO DE:  
A C T U A R I O  
P R E S E N T A :  
ADRIANA ALVAREZ DURÁN

DIRECTOR DE TESIS: ACT. HUMBERTO SANTILLANA LOYO



MÉXICO, D.F.

FACULTAD DE CIENCIAS  
UNAM



FACULTAD DE CIENCIAS  
SECCIÓN ESCOLAR

2004



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

ESTA TESIS NO SALE  
DE LA BIBLIOTECA



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: ALVAREZ DURAN  
ADRIANA  
FECHA: 14 ENERO 2004  
FIRMA: FCO LOPEZ H02  
P.A. FCOLOPEZ1402

**DRA. MARÍA DE LOURDES ESTEVA PERALTA**  
Jefa de la División de Estudios Profesionales de la  
Facultad de Ciencias  
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:  
"LA PIÑA EN LA CUENCA DEL PAPALOAPAN (1990-2000): UN ESTUDIO ESTADISTICO  
A TRAVES DE LA REGRESION LINEAL PARA IMPULSAR SU DESARROLLO".

realizado por ADRIANA ALVAREZ DURAN

con número de cuenta 09455855-1 , quien cubrió los créditos de la carrera de: ACTUARIA

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario ACT. HUMBERTO SANTILLANA LOYO

Propietario MAT. MARGARITA ELVIRA CHAVEZ CANO

Propietario ACT. FRANCISCO SANCHEZ VILLARREAL

Suplente M. EN C. JOSE ANTONIO FLORES DIAZ

Suplente M. EN AP. MARIA DEL PILAR ALONSO REYES

Consejo Departamental de

MATEMATICAS



M. EN C. JOSE ANTONIO FLORES DIAZ

FACULTAD DE CIENCIAS

CONSEJO DEPARTAMENTAL

DE  
MATEMATICAS

## *A Dios*

Por la vida, amor y compañía

## *A mis padres*

Florencio  
María Guadalupe  
Por su amor, apoyo, paciencia y motivación

## *A mis tías*

Basilisa y Aurora  
Por su comprensión y complicidad

## *A mi esposo*

Por su amor, comprensión, complicidad y compañía

## *A mi hija*

Por su amor, ternura, compañía e ilusión

## *A mis amigos*

Ivonne, Juan José, Fernando, Luis Felipe y Francisco  
Por su respeto y amistad

## *A mis otros amigos*

Norma, Blanca, Edalid, Gloria,  
David, Ranferi, Pablo, Juan y Alejandro  
Por su amistad y consejos

# CONTENIDO

PRESENTACIÓN .....	4
INTRODUCCIÓN .....	7
1. ANTEDECENTES .....	9
1.1 Origen de la piña ( <i>ananas comosus</i> ) .....	9
1.2 Historia de la piña en México .....	10
1.3 Descripción botánica de la piña .....	10
1.4 Principales variedades de piña .....	13
1.5 Variedades cultivadas en México .....	16
2. TECNOLOGÍA Y PRODUCTIVIDAD DE LA PIÑA .....	17
2.1 Requerimientos climáticos para la producción de piña .....	17
2.2 Plagas y enfermedades de la piña .....	19
2.3 Proceso productivo .....	22
2.3.1 Preparación del terreno .....	22
2.3.2 Siembra .....	22
2.3.2.1 Producción de hijuelos .....	24
2.3.2.2 Corte y selección .....	24
2.3.2.3 Desinfección .....	24
2.3.3 Densidades de plantación y normas del mercado .....	25
2.3.3.1 Sistemas de plantación .....	26
2.3.3.2 Método de plantación .....	27
2.3.3.3 Fertilización .....	28
2.3.3.4 Control de malezas .....	29
2.4 La producción de piña en el mundo en el periodo 1990 – 2000 .	30
2.4.1 Producción .....	31
2.4.2 Comercialización de la piña en el mundo .....	33
2.4.2.1 Principales países exportadores e	
importadores del mundo .....	33
2.4.2.2 La importancia de la comercialización ....	39
2.5 La producción de piña en México en el periodo 1990 – 2000 ....	41
2.5.1 Análisis de Indicadores en producción de piña por	
entidad y a nivel nacional .....	44
2.5.1.1 Superficie sembrada y cosechada .....	44
2.5.1.2 Producción .....	49
2.5.1.3 Rendimiento .....	51

2.5.2	Comercialización interna y externa .....	53
2.5.2.1	Exportaciones e Importaciones de piña durante el periodo 1990 – 2000 .....	56
2.5.2.2	Comportamiento de los precios y rentabilidad del producto .....	59
2.5.2.3	Importancia económica del producto .....	62
2.5.3	Normas de calidad de la piña .....	63
2.5.3.1	Tipos de defectos y tolerancias .....	67
<b>3.</b>	<b>EL MODELO DE REGRESIÓN LINEAL</b> .....	<b>69</b>
3.1	Conceptos fundamentales en el análisis de regresión .....	70
3.1.1	Origen del término regresión .....	70
3.1.2	Interpretación del análisis de regresión .....	71
3.1.3	Las relaciones causa – efecto .....	71
3.1.4	Usos de la regresión .....	72
3.2	El modelo de regresión lineal simple .....	73
3.2.1	Diagrama de dispersión .....	75
3.2.2	Estimación de los parámetros por mínimos cuadrados.....	77
3.2.2.1	Estimación de $\beta_0$ y $\beta_1$ .....	77
3.2.3	Propiedades de los estimadores mínimos cuadrados y el modelo de regresión ajustado .....	80
3.2.4	Estimación de $\sigma^2$ .....	89
3.2.5	El Teorema de Gauss – Markov .....	85
3.3	Estimación de intervalos en regresión lineal simple .....	86
3.3.1	Intervalos de confianza para $\beta_0, \beta_1$ y $\sigma^2$ .....	86
3.3.2	Tabla de análisis de varianza para el modelo de regresión lineal simple .....	87
3.3.3	Pruebas individuales o pruebas de $t$ .....	90
3.3.4	Coefficiente de determinación .....	91
3.3.5	Coefficiente de correlación .....	92
3.4	El modelo de regresión lineal múltiple .....	96
3.4.1	El modelo y descripción de los datos .....	97
3.4.2	Propiedades de los mínimos cuadrados .....	99
3.4.3	Valores estimados y el error estándar .....	101
3.4.4	Coefficiente de correlación múltiple .....	102
3.4.5	Prueba de hipótesis en el modelo lineal .....	104
3.4.6	Tabla de análisis de varianza para un modelo con $k$ variables .....	106
3.4.7	El proceso de modelación .....	108
3.4.8	Supuestos sobre las variables explicatorias .....	110
3.5	Riesgos y limitaciones asociados con el análisis de regresión ....	111
3.5.1	Riesgos en el uso de la regresión .....	111
3.5.2	Medidas del ajuste del modelo .....	115
3.5.3	Análisis de residuos .....	116

3.5.4 Detección y tratamiento de puntos aberrantes o fuera de rango .....	118
4. METODOLOGÍA Y RESULTADOS .....	121
4.1 Diseño del modelo .....	121
4.2 Origen de los datos .....	122
4.3 Métodos de la regresión lineal múltiple .....	122
4.4 Interpretación de resultados .....	123
5. CONCLUSIONES .....	127
5.1 Predicciones del consumo de piña en México .....	127
5.2 Conclusiones .....	129
5.3 Recomendaciones .....	130

## ANEXOS

Anexo A	
Indicadores internacionales en la producción, consumo y comercialización de la piña .....	A
Anexo B	
Indicadores de producción y comercialización de la piña en México durante el periodo 1990 – 2000 .....	B
Anexo C	
Datos empleados en el análisis del modelo .....	C
Anexo D	
Fundamentos teóricos para la función de la demanda .....	D
Anexo E	
Construcción del índice estacional de precios de la piña mediana de la central de abastos del Distrito Federal .....	E

## BIBLIOGRAFÍA

## PRESENTACIÓN

El presente trabajo tiene como objetivo el analizar el comportamiento del consumo de piña en México en función de diversas variables, como su precio, cantidad producida, tamaño e ingreso de la población y precios de productos relacionados; cuantificar la influencia de cada una de las variables mencionadas anteriormente; proporcionar algunas recomendaciones a los productores, a partir del análisis de las variables influyentes y realizar predicciones sobre el comportamiento del consumo de la piña empleando el modelo econométrico resultante del análisis.

La producción de piña en México presenta épocas de sobreoferta y falta de producto. A diferencia de otras regiones piñeras del mundo, en donde la mayoría de la superficie cultivada con este frutal pertenece y es administrada por unos cuantos productores (por ejemplo Hawaii, en donde tres empresas transnacionales controlan el total de las 9 mil hectáreas); en México participan en esta actividad no menos de 2500 productores, cada uno de manera independiente al resto, programando sus plantaciones de acuerdo con sus necesidades, conceptos técnicos y recursos con que cuente en su momento.

La falta de producto se debe principalmente a que su precio fluctúa de tal manera que a los productores no les es redituable su producción, o bien simplemente debido a una baja producción o rendimiento por la ciclicidad del cultivo.

Cuando se presenta falta de producto, el problema se resuelve importando de otros países productores. Sin embargo cuando se presenta una sobreoferta la situación se vuelve un tanto complicada pues son miles de familias afectadas ante una falta de mercado. Pero, ¿Qué es lo que origina una sobreoferta del producto?, Las razones pueden ser muchas y entre ellas puede encontrarse la sobreproducción o la sobreimportación que desplaza el producto nacional (en productos enlatados principalmente), además de otros factores económicos.

Pare remediar la falta de mercado o la sobreoferta es necesario conocer los factores que influyen o determinan el consumo de piña en México, pues de esta manera se podrá ver la forma de generar un incremento en tal consumo manipulando uno o varios factores influyentes, como puede ser el precio mismo de la piña, el ingreso de la población, los niveles de producción, etc.; a fin de prevenir o solucionar en cierta medida situaciones como las que actualmente enfrentan los piñeros de México (falta de mercado en la época de abundancia).

Así pues, dado que resulta importante conocer la serie de factores que determinan o influyen en el consumo de piña en México, se cree necesaria la búsqueda o el establecimiento estructural de un modelo econométrico que explique en la mayor medida posible la variedad que se presenta en el consumo de la piña en México, con el fin de tomar decisiones en la manipulación de variables influyentes de tal forma que permita elevar o disminuir el consumo del producto, según se desee, para beneficio de los productores y del país mismo al prevenir o remediar en cierta forma situaciones como la actual.

Las hipótesis planteadas para el análisis estadístico del consumo de piña en México son:

- ✓ El consumo de piña en México guarda una relación inversa o negativa con su precio, pues se espera que al incrementarse su precio, el consumo disminuirá (ley de la demanda).
- ✓ El consumo de piña en México guarda una relación directa o positiva con el ingreso y tamaño de la población, ya que al incrementarse o disminuirse cada una de estas variables, el comportamiento del consumo de piña será de igual manera (bien normal).
- ✓ El consumo de piña guarda una relación directa o positiva con el precio de los productos competitivos pues se espera al incrementarse el precio de estos últimos, el consumo de piña se incremente, ya que la gente tenderá a optar por el consumo de piña en vez de los productos competitivos (elasticidad cruzada).

El presente trabajo consta de 5 capítulos. En el primer capítulo, *Antecedentes*, se describe el origen de la piña, así como su taxonomía, historia, genética, descripción botánica y las variedades existentes y cultivadas en México.

El segundo capítulo denominado *Tecnología y Productividad*, contiene los principales requerimientos de clima y suelo para la producción de la piña, control de plagas y enfermedades. Así mismo, una descripción del proceso productivo como es: preparación del terreno, siembra, densidades y métodos de plantación; se realiza un análisis estadístico de la producción de piña en la última década (1990-2000) a nivel mundial y nacional; no omitiendo su comercialización e importancia de la misma.

Por otra parte y con base a los objetivos planteados, en el tercer capítulo se desarrolla *El Modelo de Regresión Lineal Simple, El Modelo de Regresión Lineal Múltiple*, explicando los elementos necesarios de los modelos citados; y se hace ver los riesgos en el uso de la regresión así como el análisis de residuos y la detección de puntos aberrantes.

En el cuarto capítulo, se presenta la metodología, en la cual se emplea un modelo econométrico incluyendo a las variables más determinantes en el consumo de la piña en México. Así mismo, la aplicación de métodos de la regresión lineal múltiple e interpretación de los resultados.

Por último, en el capítulo 5, se presentan las predicciones del consumo de piña en México, conclusiones y recomendaciones en el proceso de producción y comercialización de la piña; seguido de los anexos, tres de los cuales corresponden a la información de indicadores nacionales e internacionales en la producción, consumo y comercialización de la piña en la última década; así también, los datos empleados en el modelo utilizado y, dos más presentan el índice nacional del precio de la piña y los fundamentos de la teoría de la demanda.

## INTRODUCCIÓN

México ha logrado consolidarse en el mercado internacional de frutas tropicales como el principal exportador de papaya, mango y aguacate. Esta posición es sumamente significativa, si se considera que el comercio internacional representa en términos de volumen, según datos de la FAO, sólo el 3% de la producción de estas frutas, ya que éstas se consumen fundamentalmente en el mercado interno de los países productores.

La piña constituye el caso de excepción entre las frutas tropicales. La falta de competitividad no sólo está afectando el acceso del producto a los mercados internacionales, sino también al mercado interno. Costa Rica y Honduras por un lado y Tailandia, Filipinas y Taiwán por el otro, han desarrollado mejores tecnologías y penetrado en el mercado de Estados Unidos, el principal socio comercial del país.

La producción de piña en México representa una fuente generadora de ingresos para aproximadamente 10,000 familias. Sin embargo el cultivo nacional se ha visto afectado por el atraso tecnológico y los altos costos. Los volúmenes de una producción que se concentra en cinco estados y pocos productores y el desarrollo de la agroindustria vinculada a ella, han estado estrechamente relacionados con los ciclos económicos del país.

La piña que se produce en México tiene como destino final tres usos; el mercado en fresco que absorbe el 70% de la producción nacional; la agroindustria en la elaboración de almíbares y jugo de piña que consume entre el 23 y 25% del volumen total que se canaliza al mercado en fresco; sin embargo, cabe señalar que no todo es consumido en la Ciudad de México y la zona conurbada, sino que parte de estos volúmenes son redistribuidos a otras centrales del país.

La Central de Abasto de Monterrey es la que sigue en importancia con alrededor de 20%, Guadalajara con 10%, mientras que otro 15% se distribuye entre diversas plazas de segundo orden de importancia como Puebla, Chihuahua, Tamaulipas, Yucatán y Michoacán. El porcentaje restante se distribuye en plazas menores de la República y cadenas de autoservicio.

En lo que respecta a las exportaciones, éstas tienen como destino principal los países de Estados Unidos absorbiendo el 96.73% y Canadá con el 3.07% del total. En tanto que sólo el 0.2% tiene otro destino (Países de Europa).

Sobre el comportamiento de dichas exportaciones en los últimos años se puede mencionar que éstas han disminuido, pues en la década más reciente (1990-2000), las exportaciones fueron en promedio de 12,261 toneladas por año, en tanto que en la década anterior (1980-1989) éstas fueron de 20,545 toneladas, lo que significa una reducción del 40.3%.

En este trabajo se hace un estudio sobre el comportamiento del consumo de la piña en el mercado interno, realizando un análisis cualitativo y cuantitativo sobre los diversos factores que intervienen en dicho consumo.

# 1. ANTEDECENTES

La piña (*Ananas comosus*), es considerada como una de las frutas tropicales más finas no sólo de México, sino del mundo. Su agradable sabor y aroma, así como su contenido en vitaminas A y C, la hacen altamente demandada en los diversos mercados. Originaria de América, la piña se ha distribuido a diversas partes del mundo, de forma que hoy en día, la participación de países asiáticos como Tailandia, Filipinas y Taiwán es fundamental en la producción mundial de dicha fruta.

La piña pertenece a la familia de las Bromeliáceas, cuya especie es la *Ananas Comosus*, siendo la variedad Cayena lisa la que ofrece grandes perspectivas desde el punto de vista económico en el incremento de las plantaciones en forma comercial con bases técnicas.

## 1.1 ORIGEN DE LA PIÑA (*Ananas Comosus*)

La historia de la piña puede dividirse en dos grandes periodos: antes y después del descubrimiento de América; en el primer caso, esta historia se ha reconstruido a través de las fuentes documentales de aquellos que vivieron en el nuevo mundo, hasta antes de la llegada de los españoles. Se considera de acuerdo a especialistas, que las culturas precolombinas habían desarrollado mecanismos para domesticar esta fruta. Hoy en día, existe el consenso de que *el punto de origen de la piña fue Brasil*.

El segundo periodo se comienza a escribir a partir del descubrimiento a lo que los europeos llamaron el nuevo mundo. Se dice que el primer contacto de la piña con

la población se ubica en el año de 1493 la cual coincide con el segundo viaje a Colón. En dicho viaje Colón y sus marineros desembarcaron en una Isla de las Antillas menores, ahí encontraron una villa poblada por grupos indígenas locales, en donde precisamente hallaron la piña (Irving 1884). Así, mientras para los europeos la piña resultaba ser una fruta extraña y exótica, para los habitantes de la América tropical era un alimento no sólo importante, sino también común en la dieta.

## 1.2 HISTORIA DE LA PIÑA EN MÉXICO.

El primer registro que en México se tiene de esta fruta, fue señalado por Geronims Benzoni que vivió en México durante el periodo de 1541-1555 y quien hizo una descripción de la planta de piña en su Historia del Nuevo Mundo de 1578, en cuyo trabajo hizo especial referencia a México. Otro registro que se tiene de esta fruta, fue el realizado por Francisco Hernández en su *Rerum Medicarun Hispaniae Thesaurus* (Roma 1651), en donde señala que encontró piña (lo que llamó *pineae indiaca*) en las regiones cálidas de México y Haití, proporcionado incluso, un dibujo que la piña pudo haber sido conocida por las principales culturas del mundo antiguo de Asia y Europa, a pesar de que casi todas (con excepción de una) las variedades de la familia de la piña se consideran nativas de América.

## 1.3 DESCRIPCIÓN BOTÁNICA DE LA PIÑA.

La piña es una planta monocotiledónea que pertenece a la familia de las Bromeliáceas, de las cuales existen cerca de 50 géneros y alrededor de 200 especies. La mayoría son xerófitas o epífitas, haciéndolas altamente eficientes al uso del agua.

Todos los tipos cultivados de piña pertenecen al género de Ananas y en particular a la especie comestible *Comosus*, dentro de esta especie existen cinco grupos: Reyna, Española, Pernambunco, Mordelona y Cayena.

La Ananas es una planta herbácea perenne, después de la recolección del fruto, las yemas axilares del tallo prosiguen su desarrollo y forman una nueva planta semejante a la primera, que da un segundo fruto o retoño, generalmente de tamaño inferior al tamaño original (Claude 1968).

**Las raíces** llamadas primarias, que tienen por origen el embrión de la semilla y por tanto existen sólo sobre las semillas, desaparecen para dar lugar a las siguientes; las raíces adventicias, típicas de numerosas monocotiledóneas, que nacen del tejido muy vascularizado que separa el cilindro central de la corteza. El conjunto del sistema radical de la planta adulta es muy superficial, pero su importancia depende esencialmente de las características físicas del suelo, estructura, aireación y humedad. Su longitud puede llegar a los 2 metros cuando el medio le es favorable. Se extiende principalmente por la capa de los 15 cm del suelo, se encuentran algunas a los 30 cm de profundidad y muy excepcionalmente a 60 cm o más (Laude, 1968).

**El tallo** tiene la típica forma de una masa de 25 cm a 30 cm de largo por 2.5 a 3.5 cm en su base y de 5.5-6.5 cm por debajo del meristemo terminal. Entre las dos partes esenciales del tallo (corteza y cilindro central) se encuentra un tejido vascular muy delgado, producido por el meristemo, típico de los tallos de las Bromeliáceas. Compuesto esencialmente de xilema y de una pequeña cantidad de floema, separados por un típico de célula especial, este tejido vascular no es continuo: está perforado por aberturas que dan pasos a los haces vasculares que sirven a las hojas. A su nivel nacen las raíces adventicias, que se encuentran en toda la longitud del tallo llegando hasta menos de 1 cm del meristemo terminal (Py Laude 1968).

**Las hojas** están dispuestas en roseta con las más jóvenes al centro y las más antiguas en el exterior siguiendo la filotaxia 5/13, la forma de las hojas es variable, según su posición en la planta, pues su edad ha llevado a clasificarlas en diferentes categorías (Py Laude, 1968).

Desde el punto de vista del manejo del cultivo, las más importantes del conjunto de hojas son las "D", éstas son las más jóvenes en términos de fisiología, en general corresponden a las hojas más altas de la planta a partir del nivel del suelo; son también las más largas después de los 8 a los 12 meses de edad de la planta; se han usado como indicadores del nivel nutricional de la planta para evaluar el efecto del ambiente sobre su estado hídrico y de desarrollo. Su peso está altamente relacionado con el peso de la planta entera. (INIFAP, 1998).

**El péndulo:** la primera manifestación visible de un cambio en el meristemo terminal, que normalmente produce las hojas, es un engrosamiento después de un corto periodo durante la cual se había estrechado esa manifestación: corresponde al comienzo de la diferenciación del péndulo, en cuyo extremo apical se desarrolla la inflorescencia que dará origen al fruto. (Py Laude, 1998).

**Flor:** Da un fruto individual llamado baya; es hermafrodita de tipo trímera, es decir, con 3 sépalos y 6 estambres situados en dos verticilos y un pistilo con ovario ínfero. Las flores están dispuestas en espiral alrededor de un eje o corazón que es la prolongación del pedúnculo. El número de flores por espiral varía mucho; se considera un total de 100 a 200 flores en los 8 espirales que conforman el fruto compuesto. Por polinización cruzada con otras variedades o individuos "fuera de tipo", lo cual es comercialmente indeseable.

**Fruto:** Se forma por partenogamia natural, es decir, sin la fecundación de óvulo y por lo tanto sin la formación de semilla; después de la antesis, todas las piezas florales contribuyen a formar el fruto partenocárpico, excepto el pistilo, los estambres y los pétalos, los cuales se marchitan. Botánicamente el fruto es una sorosis; constituido por un eje carnoso o corazón, del cual parten las flores que son concrecentes (se fusionan entre si) durante el desarrollo del fruto. Las brácteas y los carpelos se unen al eje para constituir el conjunto comestible. En la parte exterior se localizan las cavidades de los ovarios protegidos por el cáliz que forma la cáscara, en la parte superior del fruto se localiza la corona, la cual se desarrolla mientras dura la

formación del fruto después entra en estado de letargo y sólo reanuda su desarrollo cuando se separa del fruto y se establece algún medio de cultivo.

La planta de piña presenta una reproducción asexual, dentro de las mismas variedades sus flores son estériles y por tanto no producen semilla. La autoincompatibilidad disminuye a medida que las variedades se alejan más genéticamente. Es altamente heterocigótica (INIFAP, 1998)

#### 1.4 PRINCIPALES VARIEDADES DE PIÑA.

Las principales variedades de piña que se cultivan en los países productores son las siguientes:

***Española Roja:*** Se caracteriza por poseer hojas largas y estrechas de coloración verde oscuro y bandas rojizas. Sus bordes están cubiertos de espinas presentando menor número en el tipo "pinareña" que en la "camagüeyana". Su fruto tiene forma redonda y puede alcanzar hasta 2 Kg. de peso. La coloración externa, cuando está completamente madura puede ser amarillo-rojizo.

La pulpa es blanca, algo fibrosa y ligeramente ácida. Produce numerosos vástagos basales y de 1 a 2 yemas axilares, es poco sensible a las enfermedades y su fruto se adapta bien a la exportación en fresco.

La diferencia que existe entre los dos tipos de Española Roja (pinareña y camagüeyana) radica en que la camagüeyana presenta espinas alrededor de los bordes, mientras que en la pinareña se encuentran en la base y extremos apicales. En los ensayos realizados en Cuba no se han encontrado diferencias notables entre ambos tipos en cuanto a las características físicas y químicas del fruto, sin embargo el tipo pinareña facilita considerablemente el trabajo de los obreros ya que posee pocas espinas lo cual la convierte en una variedad que se presta a la exportación, pues presenta cualidades importantes como la dureza de la fruta que aminora los efectos de los golpes recibidos en la travesía, a demás de ser resistente a enfermedades.

**Cayena Lisa:** Es la variedad más extendida de forma comercial en los principales países productores de piña en el mundo. Las hojas no tienen espinas, excepto en su extremidad, son de color verde oscuro con manchas rojizas. Son anchas y algo cortas. El fruto es cilíndrico, de color naranja a rojo en su madurez y ligeramente abultado en el centro.

La pulpa es de color amarillo pálido hasta amarillo oro, tiene alto contenido en azúcar y ácido, su sabor es agradable, lo que la hace muy apreciada en el mundo.

El peso promedio del fruto es elevado, aproximadamente 2 kilogramos, aunque sobrepasa ampliamente este peso cuando se planta a menor densidad. El fruto de esta variedad se adapta al igual que la Española Roja, a la exportación. Sin embargo es más sensible a las enfermedades.

**Piña Blanca:** Es un fruto apreciado para el consumo debido a que presenta un sabor dulce y un alto contenido de jugo.

Las hojas son largas, espinosas y de color verde claro. El fruto tiene forma piramidal, con color amarillo intenso en su madurez, frecuentemente presenta una coloración verde en la cáscara aún cuando hayan alcanzado su madurez fisiológica. La pulpa posee poca fibra, sabor dulce y abundante jugo. Esta variedad produce numerosos vástagos basales.

No se recomienda para uso industrial ni para la exportación en fresco. El primer caso por su pobre resistencia al embarque dado que presenta pulpa frágil y el segundo por la baja calidad de las rodajas. La planta es menos resistente que la Española Roja a la incidencia de plagas y enfermedades, especialmente a la pudrición del corazón producido por un hongo (*Phytophthora parasitica*).

**Cayena de Martinica:** Esta variedad presenta los hijos basales (bulbillos) en cantidades variables (1-10), sus claveles son poco numerosos y son omitidos tardíamente, esto trae como consecuencia que se vea afectada la segunda cosecha.

**Cayena de Guinea:** Esta variedad no presenta hijos basales (bulbillos), sin embargo sus hijos axilares (claveles) son numerosos, generalmente no presenta problemas con la segunda cosecha, sin embargo, presenta una tendencia marcada hacia

fasciación de los frutos. En algunos casos se encuentra hasta un 50% de deformaciones del fruto, la corona o ambos a la vez, lo cual no ocurre con la Cayena de Martinico. También se aprecian diferencias en cuanto al vigor, siendo ésta superior en la Cayena de Guinea.

**Cayenas 25 y 32-33:** Son selecciones de la Cayena Guinea y presentan algunas ventajas respecto a la originaria.

- Mayor rendimiento por unidad de superficie.
- Mayor resistencia a los insectos y las enfermedades.
- Sus frutos presentan una coloración más uniforme para un mismo grado de madurez.

**Cayena de Cuba:** El color de las hojas es verde morado por el haz y verde claro por el envés. El fruto es ligeramente cónico, de color morado en estado verde y amarillo en su madurez y de menor tamaño que el de las demás Cayenas. Los hijos basales (bulbillos) son sumamente numerosos y muy próximos a la base del fruto. La emisión de claveles es tardía en la mayoría de los casos.

**Baron de Rothschild:** Ésta se diferencia de la Cayena Lisa por las hojas espinosas, su fruto es más ácido, de color más oscuro y de sabor superior. Se cultiva en Guinea para la exportación en fresco, principalmente por su resistencia a la manipulación, presenta numerosos hijos basales.

**Cabezona:** Esta variedad se caracteriza por ser una planta de crecimiento vigoroso, hojas anchas, carnosas, con espinas en los bordes. El fruto es grande, de forma cónica, llegando a pesar 4.5 Kg. o más; la masa es de color amarillo pálido de sabor agradable. La planta no se comporta de la misma forma en las diferentes regiones, variando extensiblemente su desarrollo y ciclo vegetativo. Su cultivo no está definido.

## **1.5 VARIEDADES CULTIVADAS EN MÉXICO.**

Entre las variedades que se producen en México se encuentran la Española Roja, Cabezona, Sugar Loaf, Esmeralda y la Cayena Lisa. Ésta última variedad produce frutos grandes, cilíndricos de sabor excelente, es la variedad más importante y de mayor cultivo a nivel mundial y nacional por su calidad y sus óptimas condiciones para presentarse enlatada.

La variedad Cayena Lisa fue introducida en México debido a los embarques provenientes de la colonia francesa y con destino a los Estados Unidos, las cuales pasaban por el puerto de Salina Cruz, lugar donde fue reconocida esta variedad y difundida a las zonas productoras del país.

## 2. TECNOLOGÍA Y PRODUCTIVIDAD DE LA PIÑA

### 2.1 REQUERIMIENTOS CLIMÁTICOS PARA LA PRODUCCIÓN DE PIÑA.

**Altitud:** La altitud sobre la cual la piña tiene mayor perspectiva de éxito es de 0 a 800 metros sobre el nivel del mar, aunque se le ha visto crecer de manera normal a una elevación de 1500 metros. A mayor altitud sus hojas se acortan y se hacen más angostas; el pedúnculo es el más largo en relación con el tamaño de la planta; los frutos son pequeños con los "ojos" salientes y puntiagudos; la pulpa es de color amarillo pálido y el sabor es pobre y altamente ácido. Alturas mayores de mil metros retrasan el ciclo del cultivo por lo menos en un mes.

**Temperaturas:** Las temperaturas óptimas para el desarrollo del cultivo durante el día y la noche son de 30° y 20°, respectivamente; sin embargo, existen grandes diferencias en el promedio anual de temperaturas en los diferentes países donde se cultiva la piña. El promedio más bajo es de 17° C en África del Sur y el más alto de 27° C en Vietnam. Las raíces y las hojas crecen mejor a 32° C y 29° C, respectivamente y su crecimiento cesa prácticamente por debajo de los 20° C y por arriba de los 36° C. Altas temperaturas nocturnas suelen ser perjudiciales. Debe existir una diferencia de por lo menos 4° C entre el día y la noche. Las regiones productoras de México tienen temperaturas promedio de 24° C.

La temperatura también influye en la calidad de la fruta: en los frutos que se desarrollan en regiones donde prevalecen temperaturas a niveles de radiación bajos,

el contenido de ácido se incrementa y los sólidos solubles se disminuyen; en contraste, los frutos cosechados en Cuba, Filipinas, Haití y México, contiene menor cantidad de ácido y mayores valores de azúcar.

**Radiación solar:** Cuando la luminosidad es escasa las hojas se alargan, pierden anchura, permanecen erguidas y presentan un color verde oscuro. Cuando es muy fuerte el follaje se pone amarillo o rojizo. Si existe una luminosidad adecuada, pero la temperatura es baja, el fruto obtiene una coloración rojiza en la cáscara. El cultivo requiere una luminosidad de 1200 a 1500 luxes por año.

**Precipitación:** La piña puede desarrollarse en gran rango de precipitación; existen plantaciones en la Isla Molokay, en Hawai, donde el promedio de lluvias es de 565 milímetros, en contraste con la Isla Guadalupe y el Norte de Costa Rica en donde llueven de 3,500 milímetros al año o más. En México en el Bajo Papaloapan la precipitación anual varía desde 1,100 milímetros hasta los 1,900, en donde el 85% cae de junio a noviembre.

En Tabasco llueve alrededor de 2,000 milímetros por año, con escasas lluvias que comprenden el periodo de enero a mayo. Un óptimo para la piña es de 1,500 a 1,800 milímetros bien distribuidos durante el año. De lo contrario hay que suplir con riesgos de auxilio o permanente.

Los efectos negativos de la sequía son variables, de acuerdo con la edad y estado vegetativo de la planta; se presenta en la siembra, se retarda la recuperación y el establecimiento del retoño; cuando ocurre cuatro o cinco meses después de la plantación, su efecto es mínimo, aunque el ciclo de la planta se prolonga; en cambio cuando se presenta en la etapa de floración y fructificación las consecuencias son graves, ya que el tamaño y calidad de los frutos disminuyen con relación a la severidad de ésta.

Por otra parte las plantas que sufren exceso de agua durante la época de lluvias, son más sensibles a la sequía, debido al deficiente desarrollo del sistema radical.

Cuando hay demasiada agua en el suelo, las raíces empiezan a asfixiarse, lo cual se traduce en un atraso en el crecimiento vegetativo; la planta reacciona emitiendo nuevas raíces, pero si el exceso se prolonga la planta cesa en su crecimiento y toma un color rojizo. Por ello es necesario drenar perfectamente bien los excesos de agua del terreno, ya que es preferible que la planta sufra por un poco de agua y no por lo contrario.

**Suelos:** La planta de piña puede crecer en diversos tipos de suelo. Sin embargo y por tener un sistema radical poco profundo, los óptimos para su desarrollo son los arenosos o migajón arenosos (70% arena, 20% limo y 10% arcilla) ricos en materia orgánica con PH de 4.5 a 5.5. En los trópicos se cultiva en suelos rojizos o café-rojizos lateríticos, comunes en estas regiones.

La permeabilidad del suelo es uno de los factores que limitan o favorecen el desarrollo; los suelos arcillosos se saturan de agua durante la temporada de lluvias, se disminuye su aireación y en consecuencia las raíces se asfixian; mientras que en la estación seca se agrietan y las raíces se rompen. Los suelos limosos pueden ser perjudiciales al cultivo.

El contenido nutrimental de los suelos es importante pero no determinante. Los suelos ricos generalmente son pesados y con un PH demasiado elevados, en tanto que los ácidos son de textura gruesa y poco fértiles. Siendo el PH y drenaje los dos factores más importantes del suelo que hay que tomar en cuenta para elegir el terreno donde se establecerá el cultivo.

## **2.2 PLAGAS Y ENFERMEDADES DE LA PIÑA.**

Las principales plagas que atacan al cultivo de la piña son:

**Acaro o Araña Roja (*Dolichotetranychus floridanus* Banks).** La sintomatología que presenta es que las hojas de la planta se marchitan y se tornan rosáceas, salvo el ápice que se inclina, el resto permanece rígido, característica que distingue a este daño del causado por la marchites roja.

Cuando las condiciones de sequía se prolongan es frecuente que ésta plaga invada las diferentes partes del fruto. El control es a base de Disyston 10 G.

**Nematodos:** se considera el segundo problema parasitológico más importante en el mundo; en México se encuentra muy generalizado, pueden causar la pérdida del 15 al 45% del pudrimiento de la piña.

Los géneros predominantes en la región productora de piña son *Pratylenchus*, *Helicotylenchus*, *Radaphulus*, *Tylenchus*.

Los nematodos sobreviven en estado latente a los ambientes favorables como frío y periodos largos de sequía; sin embargo se mantienen en constante actividad cuando se localizan en suelos ligeros, con contenidos de humedad del 40 al 60% de su capacidad de campo.

La sintomatología se presenta en las raíces, ya que éstas al ser atacadas presentan necrosis, deformaciones y pudriciones, lo que ocasiona la disminución en el ritmo de crecimiento y consecuentemente las hojas rojizas y pérdida de turgencia de los bordes y puntos de las hojas, flacidez del pedúnculo y frutos pequeños de mala calidad y sin valor comercial.

Los daños pueden reducirse mediante: la preparación paulatina del terreno, la rotación con otros cultivos y tratamientos químicos preventivos, como aplicación antes del último rastreo, siempre que exista la humedad suficiente, los nematicidas granulados Nemaicur 10% o Macab 10 o 15% en dosis de 100 Kg. por hectárea o nematicidas líquidos como el Mocab gel a razón de 8 litros por hectárea. Para aplicaciones durante en transcurso del ciclo además del Mocab gel 68% se puede aplicar Vydate 24% o Nemaicur a razón de 5 a 8 litros por hectárea. El intervalo entre aplicaciones incluida la de preplantación, es de 2 a 3 meses, deben de suspenderse de 50 a 70 días antes del tratamiento de inducción floral.

Otras plagas importantes son el Barrenador del fruto, el comejen, la gallina ciega, los sinfilidos, el picudo negro, las ratas y las aves.

Por otra parte las enfermedades de la piña son relativamente pocas en comparación con aquellos que atacan a otros frutales. Algunas de ellas son:

Enfermedades fungosas que son provocadas por diversas especies del género *Phytophthora* como *P. Palmivora*, *P. Cinnamoni*, *P. Parasitica* y *P. Meaddi*. Su daño principalmente es la pudrición del corazón de las raíces.

**Complejo piojo harinoso-Marchites Roja-Hormiga.** La incidencia del virus que provoca esta enfermedad se asocia a la presencia del piojo harinoso. Los daños provocan la pérdida del 10 al 15% de la producción.

La enfermedad se presenta con mayor frecuencia e intensidad durante la floración, ya que en esta etapa la planta se debilita al utilizar las reservas para la formación y desarrollo de la inflorescencia. El fruto de las plantas afectadas tiene un crecimiento raquítico, ácido y sin ningún valor comercial.

A diferencia de la marchites provocada por sequía o un ataque por nematodos, la marchites roja muestra sus síntomas con mayor violencia y siempre en manchones más o menos circulares que crecen a medida que las plantas vecinas se van infectando.

Para evitar que se propague se sugiere el control cultural: Sanidad del material vegetal, destrucción de acahuales, residuos de cosecha y rotación de cultivos, control legal y control químico, con aplicaciones de insecticidas granulados con el Disyston 10 después de los 15 a 20 de su aplicación, continuándose las aplicaciones con intervalos de 2.5 a 3 meses, y suspendiéndose una vez hincada la inducción floral o con aplicaciones líquidas durante los meses secos con productos como el Vydate a razón de 2.5 litros por Ha y Folidol a razón de 2.5 Lt por Ha.

El Marmoleo o mancha café del fruto. Provocado por la bacteria *Erwinia ananas* la cual se presenta con mayor frecuencia en el periodo de abril a junio que es cuando se presentan cambios bruscos de temperatura y humedad ambiental provocado por las lluvias. Los síntomas son que en el exterior del fruto se observan áreas de color verde opaco y al partirlos muestran en su pulpa un tejido de color café oscuro. Otras enfermedades de menor importancia son la pudrición del cogollo y la raíz.

## 2.3 PROCESO PRODUCTIVO.

### 2.3.1 Preparación del terreno.

Una buena cama de siembra se logra con las siguientes labores: chapeo, quema e incorporación de residuos de cosecha, barbecho, rastreo, nivelación y drenaje.

**a) Chapeo :** Este es necesario para destruir los residuos del cultivo anterior. Para que éstas puedan quemarse o incorporarse con mayor facilidad, debe procurarse que la trituración sea lo más fina posible.

**b) Incorporación de residuos de cosecha:** Los residuos de la cosecha anterior pueden utilizarse para elevar el contenido de nutrimentos y materia orgánica de los suelos. Para ello se requiere incorporarlos mediante un barbecho profundo, por lo menos cinco meses antes de efectuarse la plantación.

**c) Barbecho:** Su propósito es romper y aflojar la capa arable del suelo, incorporar los residuos de la cosecha anterior, destruir algunas plagas del suelo, así como mejorar la aeración y la penetración del agua. Debe ser profundo, tanto como lo permita el terreno y la maquinaria disponible. Se efectúa por lo menos antes de la plantación.

**d) Rastreo:** Consiste en romper y desmoronar los terrenos que quedan después de barbecho. Para que el suelo quede mullido se sugieren por lo menos dos pasos de rastra en forma cruzada, a 20 cm de profundidad. Puede dejarse un tiempo razonable entre rastreos, para que la maleza germine y se destruya por lo menos una generación de ésta con el rastreo posterior.

Una variante para preparar el suelo consiste en dar dos o tres pasos de rastra pesada y posteriormente, un barbecho con un arado de reja. Esto deja listo el terreno para la plantación, sustituye a la preparación convencional con arado de discos y permite una mejor cama de siembra en terrenos con alto contenido de arena.

Algunos productores realizan prácticas de mejoramiento del suelo como es el control de la acidez, la cual se realiza con la finalidad de neutralizar los residuos ácidos de los fertilizantes utilizados durante el ciclo y reintegrarle al terreno las cantidades de calcio y magnesio que se pierden por la erosión, lixiviación y la cosecha de fruto y vástagos, se recomienda aplicar de ½ a 2 toneladas de cal agrícola por hectárea, según el grado de acidez del terreno. Es indispensable distribuirla uniformemente en el terreno, ya sea manualmente o con maquinaria, inmediatamente después de aplicarla se incorpora mediante un paso de rastra.

### 2.3.2 Siembra.

#### Material de propagación.

La propagación de la piña es asexual y para su establecimiento se utilizan los brotes vegetativos que la misma planta emite en forma natural, los cuales son:

- **Corona:** Se localiza en la parte superior del fruto y es de hecho el material apical de la planta, este material sólo es disponible durante el periodo de actividad de la industria.
- **Gallo:** Se desarrolla a partir de yemas axilares del pedúnculo del fruto. Se produce en promedio 2 por planta, aunque en la cosecha de los meses de mayo a julio se incrementa en 5 debido a que la diferenciación floral de la planta madre ocurre en forma natural o inducida durante los meses de noviembre, diciembre y enero.
- **Clavo:** Son los vástagos que se originan de las yemas axilares del tallo, es el tipo más abundante, se produce en promedio 4 brotes por planta.

Los tres tipos difieren en su forma y en la longitud de su ciclo, en condiciones normales, la corona requiere en promedio 19 meses para fructificar, mientras que el clavo y el gallo requieren de 16 a 17 meses respectivamente.

### **2.3.2.1 Producción de hijuelos.**

Para producir un buen material de siembra es necesario un manejo especial a los acahuales después de cosechar la fruta. Primeramente se arrancan y se sacan de la plantación todas las plantas madres previamente marcadas; los hijuelos grandes se corren y se sacan del piñal para uniformizar la siguiente camada de vástagos; las plantas madre se podan con machete para permitir mayor iluminación de los tallos y así aumentar la emisión de nuevos retoños. Estas actividades deben complementarse con la aplicación de fertilizantes nitrogenados.

### **2.3.2.2 Corte y selección.**

Los hijuelos deben cortarse cuando alcanzan el peso adecuado, según la fecha de plantación programada; así se detiene su crecimiento. Se colocan sobre la planta madre con la base expuesta al sol para que cicatrice la herida causada al momento del corte. En el material de corona, es necesario remover todos los restos de pulpa adherida, ya que propician el desarrollo de hongos en la base del retoño. Se coloca sobre una superficie libre de encharcamiento, con la base expuesta al sol para evitar el ataque de pudrición negra.

### **2.3.2.3 Desinfección.**

La práctica más utilizada por los productores consiste en sumergir el material vegetativo en una solución insecticida, utilizando para ello tinas metálicas y cajas de malla, que permitan una mayor eficiencia. Se requiere 50 ml de producto comercial (prefiriéndose los de contacto como Paration metílico o diazinon al 50 o 60% de concentración) por litro de agua; a la solución se le puede agregar 2.5 gramos de Alliete 80% por litro de agua, con el fin de prevenir el ataque de hongos al cogollo y raíz de los hijuelos o coronas.

Las plantas se sumergen completamente durante 10 segundos y después se vacía en una charola para recuperar los excedentes de solución.

### 2.3.3 Densidades de plantación y normas del mercado.

La densidad de plantación depende del desarrollo de la cosecha. El aumento en el número de plantas por hectárea disminuye el peso medio del fruto; sin embargo el tonelaje total se incrementa.

Para optar por una u otra densidad se debe considerar la cantidad de lluvia, pendiente y tipo de suelo. La siguiente tabla presenta algunos rendimientos de acuerdo a diferentes densidades de plantación.

**Tabla 1.** Rendimiento potencial de la piña en toneladas de fruta fresca y su porcentaje obtenido por tipo de mercado según la densidad de población.

Densidad de plantación por Ha.	Porcentaje de Frutos por tipo de Mercado			Peso promedio del fruto (Kg.)	Rendimiento potencial (Ton/Ha)
	Fresco Nacional	Industrial	Fresco de exportación		
	(> de 2.5 Kg)	(De 1.8 a 2.5 Kg)	(De 1.3 a 1.8 Kg)		
30,000	72	28	0	2.8	84
35,000	68	27	5	2.6	91
40,000	55	35	10	2.4	96
45,000	46	40	14	2.3	104
50,000	40	42	18	2.2	110
55,000	25	45	30	2.1	118
60,000	18	47	35	2.1	126
65,000	12	47	41	2.0	130
70,000	5	45	50	1.9	133
80,000	4	20	76	1.6	128

Fuente: INIFAP. Tecnología para la producción de piña en México, 1998.

A continuación se mencionan las normas de los principales mercados a los cuales se destina la producción:

**Mercado fresco nacional.** El fruto destinado a este mercado debe ser del tipo A, de 2.5 kilogramos en adelante. La mayor cantidad de esta fruta se comercializa en las grandes ciudades por los vendedores de rebanadas de cocteles de frutas, por lo que: su pulpa debe ser firme, sin huecos, jugosa, de color amarillo brillante y madurez uniforme; la forma del fruto debe ser lo más cilíndrica posible, para incrementar su rendimiento en rodajas; su apariencia externa debe ser de colores uniformes y brillantes.

**Mercado industrial.** La industria enlatadora, cuyo principal producto son las rebanadas en almíbar, requiere frutos del tipo B, de 1.8 a 2.5 kilogramos. Además de las características antes mencionadas, deben tener una relación de sólidos solubles y ácido cítrico, un diámetro uniforme de la médula o corazón que no exceda los 3 centímetros y; la corona debe representar menos del 15 por ciento del peso total del fruto.

**Mercado fresco de exportación.** Requiere frutos entre 1.3 a 2.2 kilogramos; los más comunes son de 1.5 a 1.8 kilogramos, clasificados como tipo C. Los mercados internacionales varían en sus requerimientos; sin embargo, de manera general solicitan frutos con corona simple, recta y que no sobrepase del 10 por ciento del peso total del fruto.

**Frutos de rezaga.** Consiste en frutos "lacrados" por el barrenador, malformados, dañados o menores de 1.3 kilogramos. Normalmente se comercializan para la obtención de jugo y mermelada.

#### **2.3.3.1 Sistemas de plantación.**

Las plantaciones pueden establecerse mediante dos sistemas: hilera sencilla o hilera doble. El primer sistema se utiliza cuando la densidad de población es menos de 30 mil plantas por hectárea y el segundo cuando la densidad de población es mayor.

Este último sistema reduce el problema de malezas en la calle angosta, facilita el paso por la calle para las diferentes labores; además, en caso de dejar la plantación para una segunda cosecha (acahual) se reduce el porcentaje de frutos ladeados, lo cual evita su deformación y caída, ya que en las calles angostas las plantas se sostienen entre sí.

Cuando se utiliza el sistema de hilera sencilla la amplitud de la calle ancha es de 80 cm y una distancia de 41 cm entre planta, lo que da una densidad de 30,000 plantas por hectárea y cuando se planta a hilera doble la amplitud de la calle ancha oscila entre los 60 y 90 cm, en la angosta de 40 a 55 cm y la distancia entre plantas es de 25 a 45 cm, para dar una densidad de población de 30 mil a 80 mil plantas por hectárea.

#### **2.3.3.2 Método de plantación.**

La plantación de los vástagos y las coronas es completamente manual. Las plantas se colocan en hoyos del tamaño adecuado, o en un pequeño surco denominado "raya".

a) *Ah oyado*: Para este sistema se tiene tres alternativas; el espeque, cuyo funcionamiento es más eficiente bajo buenas condiciones de preparación del suelo y humedad; el cava hoyo que permite ahoyar bajo condiciones de poca humedad y terrenos muy arenosos; y la espátula "cuchara" o "polín" que permite al operador hacer el hoyo y plantar el vástago al mismo tiempo.

b) *Rayado* : Para realizarlo se utilizan los cinceles de los subsoleos, los cuales se entierran a la profundidad requerida (25 a 30 cm). Sus ventajas son la rapidez de la operación y el buen asentamiento de los vástagos, ya que se entierra a presión y se apisonan posteriormente.

**Tabla 2:** *Diferentes tipos de material vegetativo que se utilizan en la siembra.*

<b>Tipo de material</b>	<b>Peso del material (gr.)</b>	<b>Tamaño en (cm.)</b>	<b>Mes de plantación</b>
Corona	250	16	Todo el año
	350	20	Todo el año
Gallo	300	31	julio
	400	37	agosto
	600	41	septiembre
	800	45	octubre
	1,000	50	noviembre-enero
Clavo	300	42	julio, agosto
	400	45	septiembre
	600	55	octubre
	800	60	noviembre
	1,000	65	diciembre, enero

Fuente: INIFAP. Tecnología para la producción de piña en México, 1998.

### **2.3.3.3 Fertilización.**

Requiere para su desarrollo una gran cantidad de nutrientes. Una hectárea de piña extrae alrededor de 350 kilogramos de nitrógeno, 20 de fósforo, 450 de potasio, 116 de azufre, 42 de magnesio y 100 de calcio.

Según densidad de plantas utilizadas, se requiere durante el ciclo del cultivo, de 12 a 18 gramos de potasio y de 2 a 4 gramos de magnesio. Estos nutrimentos se distribuyen en cuatro o cinco aplicaciones sólidas o sus equivalentes 12 o 15 aplicaciones líquidas antes de que la planta inicie su floración.

El sulfato y nitrato de amonio son las fuentes de nitrógeno más adecuadas y en el caso del potasio la mejor fuente es el sulfato de potasio y para el magnesio el sulfato de magnesio.

El intervalo entre las aplicaciones de fertilizantes sólidos es de dos a tres meses y para las fertilizaciones foliares el intervalo es de 20 a 30 días, estimándose

que 3.5 de ellas equivalen a una de sólido; y si se hace con una aspersora de alto volumen se puede hacer cada 15 días.

En la época seca, que abarca generalmente de marzo a mayo, la planta puede presentar algunas deficiencias de elementos, las cuales se corrigen con fertilizantes disueltos en el agua con el uso de bomba de mochila o aspersoras de alto volumen, en el primer caso, por cada 100 litros de agua se disuelve de tres a cinco kilogramos de urea + 1 a 2 kilogramos de 18-46-00 + 3 a 5 kilogramos de cloruro de potasio. Se aplican 50 mililitros de esta solución en las axilas de las hojas. Cuando se realiza con equipo de aspersión de alto volumen no debe exceder del 5%.

Durante el desarrollo vegetativo del cultivo se deben programar al menos tres aplicaciones de micro nutrientes. Las dos primeras aplicaciones deben llevarse a cabo a las 4 a 6 meses después de la plantación; para ello se disuelven 100 gramos de sulfato de zinc y 50 gramos de sulfato de hierro por cada 100 litros de agua, la tercer aplicación se efectúa a los 8 meses de edad y consiste en la aplicación de 100 gramos de Borax por cada 100 litros de agua.

Otra opción práctica es aplicar micro nutrientes en presentaciones comerciales, entre los que destaca Fertipíña, el cual se aplica a razón de 7 kilogramos por hectárea a los 4.6 y 8 meses después de la siembra, junto con los elementos mayores en las aspersiones programadas. Este producto contiene las cantidades de fierro, zinc, boro, magnesio y cobre necesario.

#### **2.3.3.4 Control de malezas.**

Las principales especies de malezas que se presentan son: pelo de conejo *Cynodon dactylon* (L); bejuco de tuza *Ipomoea indica* Burm.; matalí *Phaeosphaerion biocarpum* Bent.; coquillos *Cyperus spp*; zacate privilegio *Panicum maxcimum* Jacq.; hierba peluda *Ageratum tomentosum* L. y zacate grama *Digitaria sanguinalis* L Scop.

Para controlar adecuadamente la maleza se puede combinar la preparación oportuna del suelo (2 meses antes de plantar); el uso de tarpala (control mecánico) y la aplicación de herbicidas.

Cuando el terreno se encuentra fuertemente infectado con malezas de difícil control (pelo de conejo, bejuco de tuza y/o coquillo) se sugiere aplicar de 3 a 4 litros de glifosato por hectárea antes de preparar el terreno.

El control químico de las malezas se efectúa generalmente con dos aplicaciones de herbicidas: la primera después del establecimiento y la segunda al inicio de la temporada de lluvias del siguiente año. En ambos casos se sugiere aplicar 4 kilogramos de karmex más un kilogramo de Aviar o de 4 a 6 kilogramos de Gesapax Combi 80 GDA.

El volumen normal de agua para aplicar herbicida es por lo menos de 800 litros por hectárea.

## **2.4 LA PRODUCCIÓN DE PIÑA EN EL MUNDO EN EL PERIODO 1990 – 2000.**

La piña (*ananas comosus*), es considerada como una de las frutas tropicales más finas no sólo de México, sino del mundo entero. Su agradable sabor y aroma, así como su contenido en vitaminas A y C, además de otras propiedades, la hacen altamente demandada en los diversos mercados internacionales, sobre todo en Estados Unidos y Europa donde la participación de países asiáticos como Tailandia, Filipinas o Taiwan cobra gran relevancia.

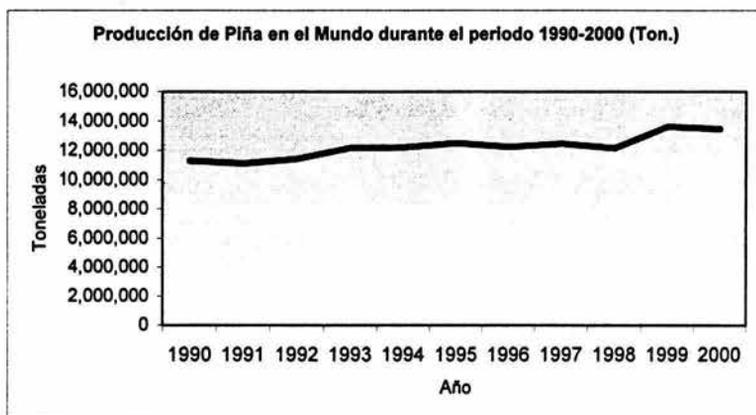
Durante el periodo 1990-2000 los principales indicadores tanto en el proceso de producción como en la comercialización del producto sufren algunas fluctuaciones que se traducen al final como incrementos o reducciones. A continuación se describe el comportamiento de algunos indicadores en la producción y comercialización mundial de la piña.

### 2.4.1 Producción.

El cultivo de piña en el mundo se concentra en países cuyo clima es principalmente de tipo tropical, en este sentido, países como **Tailandia, Brasil, Filipinas, India y Nigeria** tienen una importante participación en la producción mundial de este fruto pues sus condiciones climáticas les permite destacar como países productores. Durante el periodo 1990-2000, en estos cinco países se concentró un poco más del 54% de la producción en todo el mundo. Tailandia se ubica como el principal productor de la fruta al aportar aproximadamente un 17.50% de la producción mundial durante el periodo antes mencionado, seguido por Filipinas y Brasil con un 11.30% y 10.26%, respectivamente. De los cinco principales países productores es Nigeria el que menor aporta a la producción mundial con tan solo un 6.70%.

Para el mismo periodo, tanto el volumen de producción como la superficie sembrada no experimentaron cambios drásticos. En el primer indicador sólo se tuvo un ligero incremento de aproximadamente 1.91% en promedio anual al pasar de 11 millones 291 mil 972 toneladas en 1990 a 13 millones 448 mil 930 toneladas en el año 2000. Por otro lado, la superficie sembrada sólo experimenta un incremento de 2.10% en promedio anual, pasando de 622 mil seis hectáreas en 1990 a 752 mil 405 hectáreas en el año 2000. Sin embargo, el rendimiento sufre una ligera reducción de 0.1539% en promedio anual al pasar de 18.15 ton/ha. en 1990 a 17.87 ton/ha para el año 2000, obteniéndose un rendimiento promedio de 18 toneladas por hectárea durante el periodo en todo el mundo. De los principales países productores, Filipinas y Brasil ocupan el primer lugar con un rendimiento promedio de 28 ton/ha, en tanto que Nigeria es el país con menor rendimiento alcanzando sólo una razón de 8 toneladas por hectárea, muy por debajo del rendimiento promedio a nivel mundial.

**Gráfico 1. Comportamiento del volumen de producción de piña a nivel mundial durante el periodo 1990-2000.**



Fuente: Base de Datos de la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO).

Describiendo brevemente el comportamiento de los principales países productores en los indicadores de producción y superficie sembrada, se tiene que durante este periodo Filipinas es quien más incrementa su producción al experimentar un aumento promedio anual de 3.19% pasando de 1,155,748 toneladas en 1990 a 1,524,567 toneladas en el año 2000, a este país le siguen la India y Tailandia con incrementos promedios anuales de 2.49% y 2.26%, respectivamente; en tanto que Brasil y Nigeria experimentan incrementos de aproximadamente 1.71% y 1.55% en promedio anual, respectivamente.

Por otro lado, en la superficie sembrada es Brasil quien mayor incremento tiene al pasar de 33,167 Ha., en 1990 a 55,749 Ha., en el año 2000 traduciéndose en un incremento medio anual de 6.81%, seguido por la India con un 4.66% de incremento promedio anual. A estos países le siguen Tailandia con un 3.06%, Nigeria con 1.50% y Filipinas una reducción de aproximadamente 2.71% en promedio anual pasando de 59,619 hectáreas en 1990 a 43 mil 449 Ha., para el año 2000.

Nótese que a pesar de que Filipinas sufre una reducción en su superficie sembrada, durante este periodo logra incrementar su producción debido a los altos

rendimientos obtenidos como consecuencia de la implementación de mejores técnicas de cultivo.

Finalmente, como dato adicional se puede comentar que de los 5 principales países productores Nigeria es quien destina mayor superficie para la siembra teniendo un promedio de 105 mil hectáreas durante el periodo tratado. Sin embargo, debido a sus bajos rendimientos es quien menor aportación tiene en la producción mundial; en contraste con Brasil quien a pesar de ser el país que menor fracción de superficie destina a la siembra (45,715 hectáreas), sus altos rendimientos le permiten ocupar el tercer lugar en producción del fruto a nivel mundial.

#### **2.4.2 Comercialización de la piña en el mundo.**

El comercio mundial de piña ha registrado una tendencia hacia el crecimiento en los últimos años a tal grado que se ha encontrado por arriba del crecimiento en la producción, lo cual indica la buena aceptación del producto en la dieta de los consumidores en gran parte del mundo. Por ejemplo, para el periodo 1990-2000 las exportaciones mundiales experimentan una serie de fluctuaciones moderadas tanto a la alza como a la baja traduciéndose en un incremento considerable de 8.10% en promedio anual al pasar de 574 mil 102 toneladas en 1990 a 1 millón 39 mil 209 toneladas para el año 2000. La red mundial de comercio de la piña lo componen básicamente países de los continentes Americano, Asiático y Europeo.

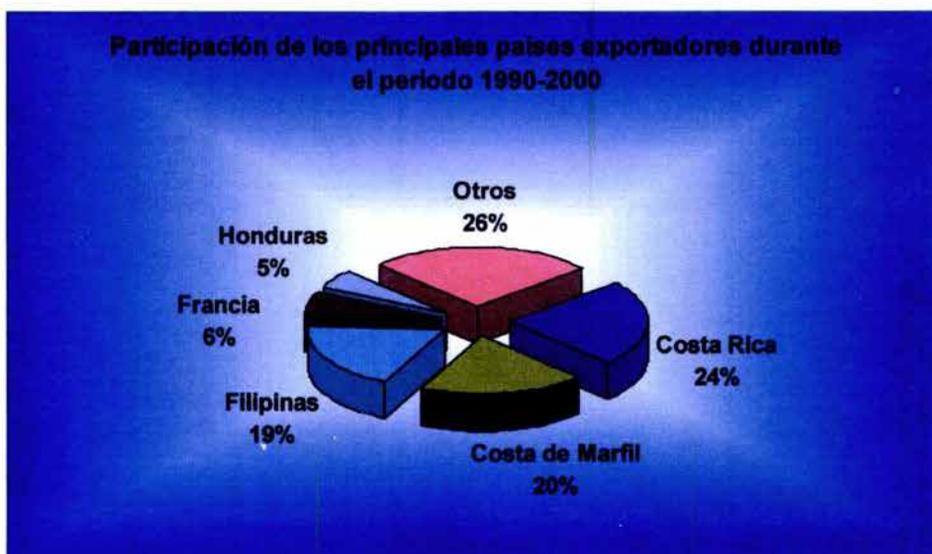
##### **2.4.2.1 Principales países exportadores e importadores del mundo.**

Dado que entre los principales países productores de piña se encuentran Tailandia, Brasil y Filipinas, podría esperarse que dichas federaciones figuraran en la lista de los principales países exportadores del producto. Sin embargo, contrario a las expectativas, de dichos países sólo figura Filipinas ocupando el tercer lugar en importancia, por debajo de **Costa Rica** quien ocupa el primer lugar al contribuir con aproximadamente el 24% de lo exportado a nivel mundial durante el periodo 1990-2000. A lo largo del mismo cuando este país incrementaba sus niveles de

exportación hasta en un 23.63% en promedio anual al pasar de 95 mil 880 toneladas en 1990 a 322 mil 453 toneladas para el año 2000 exportando un promedio de 186 mil 595 toneladas anuales debido a que sus bajos niveles de consumo en su mercado interno le permite exportar más del 80% de su producción. Sus exportaciones promedio le generan ingresos de 72 millones 528 mil dólares en promedio anual.

En segundo lugar se encuentra **Costa de Marfil** con exportaciones promedio de 155 mil 187 toneladas anuales a lo largo del periodo. Este país presenta un crecimiento en este rubro de 3.80% en promedio anual al pasar de 138 mil 683 toneladas a 191 mil 398 toneladas de 1990 al año 2000. Al igual que Costa Rica, Costa de Marfil tiene un consumo interno bastante limitado, de tal modo que durante el período en descripción exporta un total de 1 millón 707 mil 53 toneladas habiendo producido 2 millones 459 mil 143 toneladas, es decir, el 69.42% de su producción total.

**Gráfico 2.** Principales países exportadores de piña en el mundo y su contribución durante el periodo 1990-2000.



Fuente: Base de Datos de la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO).

**Filipinas** ocupa el tercer lugar como país exportador de piña en el mundo al contribuir con aproximadamente el 19% de lo exportado en el mundo durante el periodo antes mencionado percibiendo ingresos de 24 millones 96 mil dólares en promedio anual. En dicho país se experimenta una serie de fluctuaciones en las exportaciones obteniéndose como saldo final una ligera reducción de 0.74% en promedio anual al pasar de 146 mil 323 toneladas en 1990 a 135 mil 484 toneladas en el año 2000. En promedio, Filipinas exporta 146 mil 778 toneladas anuales a lo largo del periodo, es decir, aproximadamente 40 mil toneladas menos que Costa Rica.

Los siguientes dos principales países exportadores son **Francia y Honduras** que en conjunto participaron con el 11% de lo exportado durante el periodo en cuestión. Pese a no ser un productor de piña Francia destaca como uno de los principales exportadores al destinar la comercialización externa en un promedio de 50 mil 169 toneladas anuales durante el periodo aportando aproximadamente el 6% de las exportaciones mundiales debido a que gran parte de sus compras, al ser también un principal importador, son canalizados a otros países, principalmente de la Unión Europea, sirviendo de esta forma como intermediario en el comercio del producto.

Finalmente, Honduras participa durante el mismo lapso con el 5% de lo comercializado en el exterior, al exportar un promedio de 41 mil 700 toneladas anuales permitiéndole captar ingresos promedios de 13 millones 490 mil dólares.

En resumen, se puede decir que durante el periodo 1990-2000 los países de Costa Rica, Costa de Marfil, Filipinas, Francia y Honduras, aportan el 74% de lo comercializado en el mundo, debido principalmente a que en algunos de estos países se presenta una reducida demanda interna del producto.

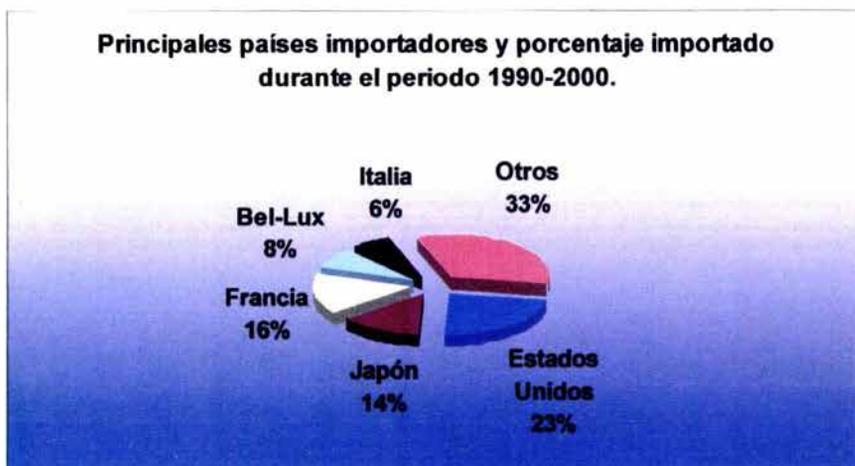
Por otro lado, la lista de los principales países **importadores** es encabezada por los **Estados Unidos de América** al experimentar un incremento considerable en este rubro de 17.99% en promedio anual pasando de 113 mil 885 toneladas en 1990 a 318 mil 837 toneladas para el año 2000 lo que produce que a lo largo del lapso

este país adquiriera el 22.72% de las importaciones mundiales pagando un costo total de 884 millones 704 mil dólares.

El incremento en la demanda de piña en el mercado estadounidense se debe a que lo largo del periodo, la producción de este país no satisface ni el 18% de la demanda interna pues en promedio se producen sólo 342 mil 214 toneladas anuales, en tanto que la demanda interna se encuentra en un promedio de 1 millón 930 mil 783 toneladas (Ver datos de anexo A). Cabe hacer mención que gran parte del producto demandado por éste país es empleada para los requerimiento de la industria de conservas, que una vez procesada es comercializado tanto interna como externamente.

**Francia** es el segundo país que mayor cantidad de Piña importa absorbiendo el 15.42% del total mundial durante el periodo mencionado. Dicho país a experimentado una serie de variaciones moderadas tanto de incrementos como de reducciones en sus importaciones que sin embargo se traducen en un incremento de 8.49% en promedio anual pasando de 80 mil 166 toneladas a 148 mil 239 toneladas en 1990 al año 2000. En promedio, Francia importa 119 mil 40 toneladas anuales pagando un promedio de 36 millones 679 mil dólares. De acuerdo a los reporte de la FAO éste país se abastece del producto del mercado internacional tanto para satisfacer su suministro interno como su mercado externo dado que no figura como país productor.

**Gráfico 3.** Principales importadores de Piña y sus importaciones durante el periodo 1990-2000.



Fuente: Base de datos de la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO), con información actualizada hasta el año 2002.

Otro de los principales países importadores de piña es **Japón**, quien ocupa el tercer lugar al adquirir el 14.17% del producto comercializado en el mundo durante el periodo en mención. Este país importa un promedio anual 109 mil toneladas durante el lapso y experimenta una serie de reducciones en sus niveles de importación a partir de 1995 que se traducen en un saldo final negativo de 2.19% en promedio anual al pasar de 128 mil 250 toneladas en 1990 a 100 mil 92 toneladas para el año 2000 como resultado de la disminución de la demanda del producto en su mercado interno. Sus importaciones tienen un valor promedio anual de 49 millones 632 mil dólares, es decir casi 13 millones de dólares más que las de Francia, a pesar de que este último importa alrededor de 10 mil toneladas más que Japón.

Finalmente, entre **Bélgica-Luxemburgo** e **Italia** importan aproximadamente el 14.65% de lo comercializado a nivel mundial. Los primeros participan absorbiendo el 8.23% y durante el periodo experimentaron un incremento de aproximadamente 2.65% en promedio anual debido a que tuvieron incrementos medios anuales tanto

en su demanda interna como en sus exportaciones de alrededor de 0.99% y 8.39%, respectivamente. Estos países importaron un promedio de 63 mil 513 toneladas durante cada año con un valor medio de 29 millones 44 mil dólares. Como dato interesante, se señala que comentar que al igual que Francia, Bélgica y Luxemburgo fungen también como países intermediarios en la comercialización de la piña ya que sus exportaciones son prácticamente el doble de la cantidad suministrada en el país, teniendo promedios anuales de 61 mil 392 toneladas y 29 mil 947 toneladas, respectivamente.

Por último, **Italia** importa el 6.42% del total mundial y durante el periodo experimenta un incremento en sus importaciones al pasar de 40 mil 499 toneladas en 1990 a 67 mil 677 en el año 2000, lo que da como resultado una tasa de cambio positiva de aproximadamente 6.71% en promedio anual. Este país adquiere durante el periodo un total de 49 mil 545 toneladas en promedio anual con un costo aproximado de 37 millones 947 mil dólares.

Las cifras mundiales de los indicadores exportaciones e importaciones, arrojan saldos positivos al obtenerse incrementos de 8.10% y 7.53% en promedio anual, respectivamente. Por otro lado, el valor de las exportaciones se incrementaron hasta en un 131.11% durante el periodo pasando de 179 millones 566 mil dólares en 1990 a 414 millones 995 mil dólares para el año 2000, en tanto que el valor de las importaciones se incrementan hasta en un 65.47% al pasar de 376 millones 535 mil dólares a 623 millones 43 mil dólares durante el mismo lapso.

Se analiza la participación de México en la comercialización de la piña en el mundo, se podrá ver que está prácticamente insignificante pues las importaciones sólo representan el 0.009% de lo comercializado a nivel mundial en tanto que las exportaciones sólo significan el 1.65%. Por otra parte, la producción representó el 2.95% de lo producido en todo el mundo durante el periodo 1990-2000.

El comportamiento de los principales indicadores tanto en la producción como en la comercialización en el mundo así como los países participantes han sido descritos de manera general en este apartado empleando los datos que se presentan en el anexo A.

#### **2.4.2.2 La importancia de la comercialización.**

En los términos más escuetos, la exportación consiste en obtener beneficios mediante la venta de productos (o servicios) en mercados exteriores. Significa encontrar clientes a los pequeños y medianos productores donde éstos pueden satisfacer mejor sus necesidades que con los actuales proveedores.

Significa atender a esos clientes con tanto éxito que los pequeños y medianos productores puedan crecer y prosperar, aumentando a la vez el empleo directo e interno. Distinguiremos entre “vender” y “comercializar”. La “venta” se refiere a un producto que ya se tiene. La “comercialización” se aplica a un producto que el mercado necesita. Se dejará a un lado el término “vender” y se utilizará “comercializar”.

Aunque la exportación ofrece numerosas ventajas a las empresas, muchas de éstas no han aprovechado las increíbles oportunidades que existen en el mercado mundial. La reestructuración masiva de fronteras políticas, la apertura y globalización de nuevos mercados de consumo, los convenios históricos y la fundación de la Organización Mundial del Comercio, han creado oportunidades sin precedentes para que las empresas exporten.

Muchos países en desarrollo han llegado a convertirse en serios rivales para economías consolidadas, debido a los enlaces que permiten los sistemas mundiales de comunicación y al mejorado acceso a la información impresa, electrónica y televisiva. Nunca antes ha existido un momento tan oportuno para las empresas. Estas deben aprovechar los cambios que se están produciendo en los mercados para exportar porque la exportación permite:

**Incrementar las ventas y utilidades.** Si la empresa rinde bien a nivel nacional, es probable que la expansión a mercados extranjeros mejore su rentabilidad. Sin embargo, analizando la estructura de exportación son pocos los grupos y productos locales que se exportan, aunque todos ellos podrían hacerlo.

**Incrementar la cuota en el mercado global.** Al exportar, la empresa sabrá como trabajan sus competidores, cuales son sus estrategias y lo que han hecho para conseguir entrar en los principales mercados extranjeros.

**Disminuir la dependencia para con los mercados nacionales:** Al exportar, la empresa incrementará su base de mercado y reducirá la competencia interna en el país.

**Estabilizar las fluctuaciones de mercado:** Desaparece la dependencia por parte de las empresas hacia los cambios económicos, gustos del consumidor y fluctuaciones estacionales, dentro de la economía nacional.

**Sacar provecho de la capacidad excedentaria de producción:** Al exportar, la capacidad y duración de los períodos de producción pueden aumentar, disminuyendo por tanto, los costos medios por unidad e incrementando las economías de escala.

**Ser más competitivo:** La exportación aumenta las ventajas competitivas de las empresas y sus respectivos países. Al tiempo que la empresa se beneficiará del contacto con nuevas tecnologías, métodos y procesos, mejorará también la balanza comercial del país.

**Crear empleos en el país:** La exportación aumenta la necesidad de mano de obra y así el empleo en el país.

**Disponer de excelentes “expertos en exportación” a bajo costo o gratuitamente:** Muchas empresas deciden no exportar por miedo a lo desconocido. En la actualidad, existen organizaciones promotoras del comercio por todo el país, para ayudar a las empresas que, a pesar de ser fuertes a nivel nacional, aun no se han lanzado a los mercados de exportación. Estas organizaciones ayudan a las empresas en cada paso del proceso de exportación.

## **Ventajas y riesgos que conlleva la exportación.**

Entre las **ventajas** directas para la empresa exportadora, se encuentran:

- La oportunidad de ampliar su participación en el mercado.
- El incremento en la producción, si no está operando con toda su capacidad en el mercado nacional.
- La disminución de la dependencia en las ventas nacionales o la compensación en caso del estancamiento del mercado nacional.
- Disolución de la competencia interna al extenderse a mercados extranjeros menos competitivos.
- El seguimiento de los exportadores líderes nacionales en los mercados extranjeros reduciendo los costos de estudio de dichos mercados.

La adaptación de los propios productos a las necesidades del mercado, debido a la competencia vigente en los mercados internacionales, lo cual desemboca en un mejoramiento de su nivel de pericia tecnológica.

Los **riesgos** de la exportación son parecidos a los del mercado nacional. Entre los riesgos potenciales que conlleva extenderse a nuevos mercados están:

- No alcanzar las ventas proyectadas.
- Una competencia mayor a la prevista.
- Impagos o demoras por parte de los clientes.

## **2.5 LA PRODUCCIÓN DE PIÑA EN MÉXICO EN EL PERIODO 1990 - 2000.**

A continuación se describirá el comportamiento de la producción y la comercialización de la piña en México empleando la información numérica presentada en el anexo B.

La producción de piña en nuestro país está caracterizada por una marcada concentración, no solo en algunos estados sino además, en una región específica. De acuerdo a cifras del periodo mas reciente (1990-2000) las entidades que concentraron el 99% tanto de la superficie cosechada y sembrada así como de la

producción fueron, en orden de importancia, los siguientes: Veracruz, Oaxaca, Tabasco, Nayarit y Jalisco.

En los dos primeros estados se ubica la zona productora de piña por excelencia. Esta zona es conocida como la zona del Bajo Papaloapan o cuenca del Papaloapan, y la forman municipios la Isla, Rodríguez Clara, Azueta, Medellín, Chacaltianguis, Alvarado y Martínez de la Torre. Además los de Loma Bonita y Tuxtepec, que se ubican en el estado de Oaxaca.

Se considera que esta región durante el periodo de 1900-1998, contribuyó con un 78% de las superficies cosechadas en el ámbito nacional, así como el 86% de la producción total nacional, de ahí su importancia. Otra región también importante la constituye el municipio de Huimanguillo en Tabasco, pues en el mismo periodo participó con el 11.19% de la superficie cosechada y 9.48% de la producción nacional. Los municipios productores con menor participación son Compostela, Valle de Banderas y Ruíz, en el estado de Nayarit; Tomatlán en Jalisco y Carrillo Puerto en el estado de Quintana Roo.

A diferencia de lo que sucede en otras naciones productoras de piña, donde la superficie está concentrada en unas cuantas manos, en México participan en esta actividad no menos de 2,500 productores de los cuales se estima que aproximadamente 1,400 pertenecen a la zona del Bajo Papaloapan. Cada uno trabaja de manera independiente empleando sus propios programas de plantación, conceptos técnicos y recursos económicos. La superficie por productor varía al grado de que existen predios de 0.5 hasta 200 hectáreas. Se estima que por lo menos 300 productores cuentan con plantaciones que superan las 25 hectáreas.

Así se puede clasificar a los productores, según su superficie de producción como sigue:

- **Pequeño productor diversificado.** Se caracteriza por cultivar de una a tres hectáreas de piña, siendo frecuente que también se dedique a la ganadería de doble propósito y/o a otros cultivos como maíz y chile. Generalmente utiliza bajas densidades de plantación y un paquete tecnológico con menor uso de fertilizantes y de insumos para el control de plagas y enfermedades. Por esta razón, sus rendimientos son bajo y comúnmente destinados al mercado nacional.
- **Productor mediano.** Cultivo de 10 a 50 hectáreas. Puede estar especializado en piña o articular este cultivo con la actividad ganadera o con el cultivo de sandía. Aplica un paquete tecnológico más intensivo en el empleo de insumos e incluye la aplicación de algunos fitorreguladores a la planta y al fruto. Por este obtiene mayores rendimientos y un costo unitario menor; además realiza una producción escalonada, a través del control de las fechas de plantación y de la inducción floral, con lo que se previene de las fluctuaciones de los precios. A este tipo de productor pertenece la mayoría de los productores de piña de la región.
- **Productores grandes.** Cultiva más de 50 hectáreas. Además de utilizar el paquete tecnológico intensivo, ha mecanizado el manejo del cultivo aplicando, por ejemplo, la aspersión de agroquímicos con maquinaria. Sus costos de producción por hectárea son mayores, pero ha reducido de manera importante sus costos unitarios debido a los altos rendimientos obtenidos. Es frecuente que realice una comercialización más directa, dándose casos de productores que se han integrado al empaque en fresco y están vinculados a compañías agroexportadoras. Es importante señalar que por el proceso de concentración de la producción, estos tipos de productores tienden a controlar los mayores volúmenes.
- **Productores intermediarios.** Los productores de éste tipo, a la vez que cultivan grandes extensiones de tierra (más de 50 hectáreas) tienen vínculos

con bodegueros o bien son propietarios de bodegas, fungiendo como intermediarios al captar la producción de otros productores para cumplir con sus compromisos de abasto.

## **2.5.1 Análisis de Indicadores en producción de piña por entidad y a nivel nacional.**

### **2.5.1.1 Superficie sembrada y cosechada**

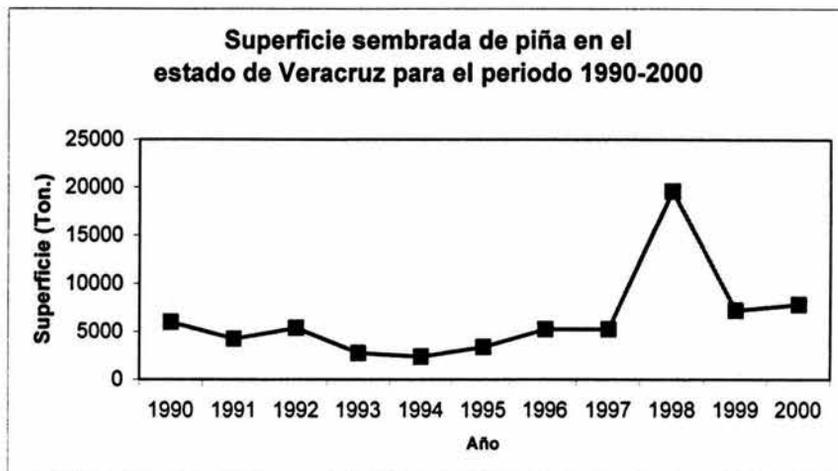
Para este periodo, la superficie sembrada experimentó una serie de incrementos y decrementos en cada uno de los estados productores de nuestro país, así, en el estado de Veracruz se produjo una disminución en la superficie cultivada para 1991 al pasar de 6,000 has. a 4,240 has. Dicha disminución obedece a la sobreoferta que se produjo durante 1989 y 1990 lo cual genera un impacto directo en los precios al provocar que durante los primeros 4 años de la década de los 90' el precio de dicho producto estuviera por debajo de los \$1000 pesos/tonelada, esto produce una serie de fluctuaciones durante los primeros cuatro años experimentándose tanto incrementos como reducciones en este rubro. El incremento más notable se produce en 1998 pues casi se cuadruplica en este año pasando de 5,226 has de 1997 a 19,562 para el año antes mencionado. Durante el periodo, el estado de Veracruz destina un promedio de 6288.18 hectáreas anuales para siembra de piña, que son en su totalidad de temporal.

Como balance general, al final de este periodo el estado de Veracruz experimentó un ligero incremento en la superficie sembrada pues se estima que se produce un aumento de 3.018% en promedio anual.

En casi todos los años que abarca este periodo, la superficie sembrada fue cosechada en su totalidad a excepción de 1998 en el que sólo se cosechó el 33.28% del total de la superficie sembrada debido a la sequía que se presentó durante ese año, además de las fluctuaciones desfavorables en el precio y en las exportaciones. Sin embargo, en términos de promedio en dicho estado se cosecharon durante cada

año un total de 4,922.73 hectáreas habiendo sembrado, como se mencionó anteriormente, un promedio de 6,288.18 hectáreas, es decir, sólo se cosecha en cada año del periodo el 78.29% de la superficie sembrada.

**Gráfico 4:** Comportamiento de la superficie sembrada de piña en el estado de Veracruz para el periodo 1990-2000.



Fuente: Centro de Estadística Agropecuaria (CEA).

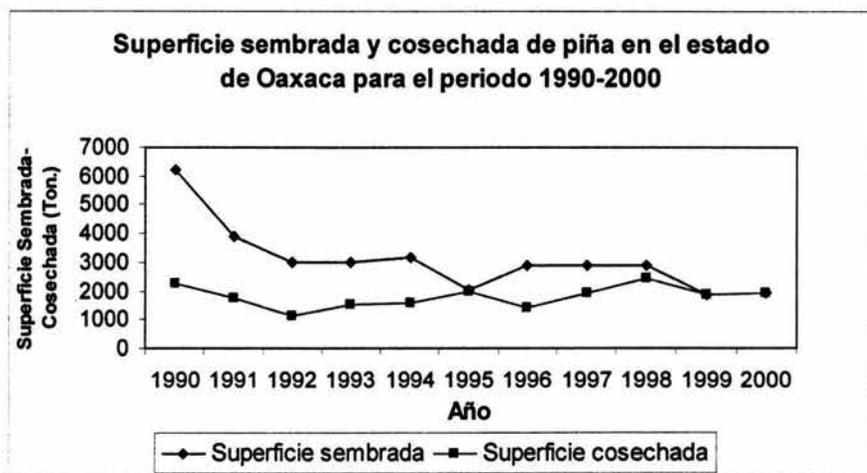
En el estado de Oaxaca se experimentaron fluctuaciones quizá más moderadas que en el estado de Veracruz, el cambio más notable se experimenta en 1991 al reducirse dicha superficie en un 37.58% con respecto al año anterior y a partir de ese año se producen más decrementos que incrementos que se traducen en una reducción de alrededor de 6.9% en promedio anual para este periodo.

Para este estado, la superficie cosechada tuvo un comportamiento distinto al de la superficie sembrada pues de 1990 a 1998 la superficie cosechada es menor que la sembrada y sólo en dos últimos años se cosecha el total de la superficie sembrada. Por lo que en términos de media se puede decir que a lo largo del lapso comentado, Oaxaca destina 3,068.36 hectáreas para la siembra del fruto durante cada año, de los cuales sólo se cosecha el 58.36%.

La razón por la cual se presenta dicha situación no sólo en este estado, sino a nivel nacional en general, se le atribuye al hecho de que durante este periodo se presentan una serie de casos desfavorables para los productores tales como la desarticulación de distintas entidades estatales como CONAFRUT, BANRURAL, FERTIMEX y el Fondo de Fomento de Apoyo a la Agricultura, perdiéndose con ello apoyos en asistencia técnica y subsidios en fertilizantes y plaguicidas para los productores. Aún más, durante el mismo periodo se da inicio al proceso de apertura económica a la competencia internacional.

Así, toda esta serie de acontecimientos provoca que durante este periodo se tenga un marcado contraste en cuanto a la superficie sembrada y cosechada tanto en el estado de Oaxaca como a nivel nacional.

**Grafico 5:** Comportamiento de la superficie sembrada y cosechada de piña en el estado de Oaxaca para el periodo 1990-2000.



Fuente: Centro de Estadística Agropecuaria (CEA)

En el resto de los principales estados productores se experimentaron incrementos en la superficie sembrada pues a pesar de que en cada estado se tuvo reducción en algunos años de este periodo, al final se obtuvieron incrementos de alrededor de 2.27%, 8.02% y 2.2% en promedio anual para los estados de Tabasco, Nayarit y Jalisco, respectivamente. Así mismo, dichos estados destinan durante cada año un promedio de 1,379.67, 703.82 y 63.55 hectáreas, respectivamente.

En lo que respecta a la superficie cosechada, a diferencia de Veracruz y Oaxaca, en los estados de **Tabasco, Nayarit y Jalisco**, no existe una marcada diferencia en la superficie sembrada y la cosechada pues en casi todo el periodo se cosecha entre el 80 y el 100% de la superficie sembrada, quizá el contraste más marcado para el estado de Jalisco se presenta durante 1990, 1991 y 1996 pues en los dos primeros años se cosecha apenas alrededor del 50% de la superficie sembrada, mientras que para el último año sólo se cosecha el 16.85%. Esta misma situación se presenta en el estado de Tabasco ya que durante 1992 sólo se cosecha la mitad de la superficie que se sembró.

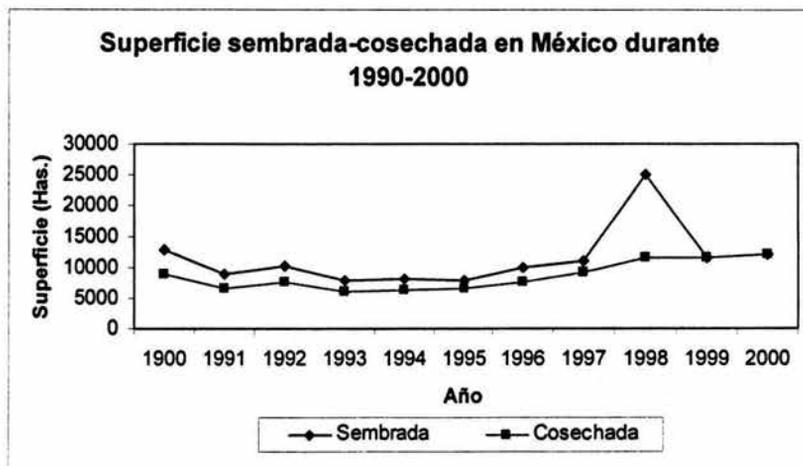
Resumiendo lo anterior en términos de promedio, se puede decir que durante cada uno de los años del periodo en descripción en el estado de Tabasco se cosechó el 85.77% de la superficie sembrada; en tanto que en el estado de Nayarit se cosecha el 97.46%, muy por encima de Jalisco en donde sólo se cosecha el 74.96% de la superficie sembrada. Nótese que a lo largo del periodo, Nayarit es quien en promedio cosecha el mayor porcentaje de su superficie sembrada, en tanto que el estado de Oaxaca es quien lo hace en menor proporción; esto se debe a que en el segundo, el total de la superficie sembrada es de temporal por lo que depende mucho de las variaciones climáticas durante el proceso de producción, en tanto que en el primer estado una parte de la superficie sembrada cuenta con sistema de riego.

Finalmente, a nivel nacional se experimentó una disminución de aproximadamente 0.63% en promedio anual en la superficie sembrada pasando de 12 mil 871 hectáreas a 12 mil 58 hectáreas de 1990 al año 2000, los cambios más

notables en incrementos y reducciones se presentan durante 1991 cuando se reduce un 31.89% de la superficie sembrada en el año anterior. Para 1998 se presenta el incremento más notable al sembrarse en este año un 132.15% más que en el año anterior, impulsado por el alza de precios experimentado al pasar de \$ 1015.78 a \$ 2118.23 por tonelada. En promedio, México destina durante cada año del periodo un total de 11 mil 369.64 hectáreas, muy por debajo de la superficie destinada por los principales países productores en donde quien menor superficie siembra es Brasil con 45 mil 715 hectáreas durante cada año del lapso en descripción, es decir, aproximadamente 75.13% menos.

Durante el mismo periodo, la superficie cosechada siempre estuvo por debajo de la sembrada, aunque en los últimos dos años se cosechó casi el total de la superficie sembrada ya que en 1999 sólo se dejaron de cosechar 42 hectáreas, en tanto que para el año 2000 tan sólo son 48 has las que no se cosechan. La diferencia más notable se presenta en 1998 pues en este año únicamente se cosechó el 45.78% de la superficie sembrada. La explicación a este acontecimiento se adjudica al hecho de que durante ese año nuestro país es perjudicado por una intensa sequía que acaba con las cosechas pues la mayor parte de la superficie sembrada es de temporal.

**Gráfica 6:** Comportamiento de la superficie sembrada y cosechada de piña a nivel nacional durante el periodo 1990-2000.



### **2.5.1.2 Producción.**

Para el periodo de 1990-2000 la producción de piña en el estado de **Veracruz** sufre un ligero incremento de aproximadamente 0.53% en promedio anual.

A lo largo de este periodo se experimentan una serie de fluctuaciones que tienden más a la alza que a la baja motivadas por diversos factores como incrementos en la superficie sembrada y cosechada, el alza y baja de precios, condiciones climáticas favorables etc. La pequeña caída de que sufre la producción se presenta en los primeros años del periodo. Para 1991 la producción pasa de 334,800 toneladas producidas en 1990, a 204 mil 73 toneladas (el 60.95% de lo que en el año anterior se había producido), debido a que en el mismo año se da una reducción en la superficie sembrada como consecuencia de la sobreoferta generada en 1989. Para 1992 la producción vuelve a sufrir una ligera caída y pasa de 204,073 toneladas a 169,636 toneladas. A partir de 1994 la producción experimenta sólo incrementos y el cambio más notable se presenta en 1996 cuando se produce un 71.90% más que en el año anterior.

En el estado de **Oaxaca** se presenta una situación un tanto diferente a la encontrada en el estado de Veracruz ya que en esta entidad la producción se reduce en aproximadamente un 0.79% en promedio anual. Al igual que en el estado de Veracruz, la producción se reduce durante 1991 en un 23.27% con respecto al año anterior, y en 1992 en un 37.05% con respecto a 1991 y a partir de 1993 la producción se incrementa de manera moderada hasta sufrir nuevamente una ligera reducción de 6.28% en el año 2000.

Para el resto de los principales estados productores Nayarit es quien experimenta mayores fluctuaciones, es decir, no conserva una tendencia a lo largo del periodo pues tiene tanto incrementos como disminuciones en la producción que se traduce en una reducción de 2.65% en promedio anual a lo largo del periodo. El cambio más notable se da en 1993 cuando sólo se produce el 22.80% de lo producido en 1992. Para 1994 nuevamente se incrementa la producción al pasar de 2,604 a 10 mil 590 toneladas y a partir de ese año se presenta un periodo de

estabilidad hasta 1999 cuando nuevamente se reduce la producción en un 44.69% con respecto a lo producido en el año anterior.

**Gráfico 7: Producción de Piña en México (toneladas), por entidad y a nivel nacional durante el periodo 1990-2000.**



Fuente: Centro de Estadística Agropecuaria (CEA).

**Tabasco** experimenta un caso diferente al estado de Nayarit ya que aquí se presenta una tendencia de crecimiento en la producción durante el lapso 1992-1995 como consecuencia al ligero incremento que se vio en los precios. Sin embargo, la producción sufre una caída drástica en 1996 al pasar de 70,500 a 40 toneladas debido a la reducción de la superficie cosechada en dicha entidad. Sin embargo, esto no es factor suficiente para frenar la tendencia en el incremento de la producción ya que se obtiene como saldo final un aumento de 8.73% en promedio anual para este periodo.

**Jalisco** es el estado que mayor incremento tiene en la producción durante este periodo, ya que en dicha entidad se logra un aumento de hasta un 678.75%.

Finalmente, a nivel nacional se presenta una tendencia a la baja en la producción durante los primeros 3 años del periodo debido a la sobreoferta que se había generado a finales de la década anterior, sin embargo a partir de 1994, motivada por la mejora relativa en los precios, la producción inicia una tendencia al crecimiento que se traduce en un aumento de 1.49% en promedio anual durante este periodo. Así mismo, México presenta producciones de 360 mil 748 toneladas en promedio anual.

Veracruz es quien mayor aporte tiene a la producción nacional durante el periodo pues se estima que en todo el lapso participa con el 61.3% de la producción nacional aportando un total de 2 millones 433 mil 31 toneladas de los 3 millones 968 mil 228 producidas en el país; en tanto que Oaxaca lo hace con un 24.6%, Tabasco con 9.8% y entre Nayarit, Jalisco y otros estados aportan el 4.2% restante.

#### **2.5.1.3 Rendimiento.**

El rendimiento se puede obtener dividiendo la magnitud de la producción entre el número de hectáreas cosechadas en cada año, la dimensión obtenida al efectuar dicha operación dá una idea del posible impacto causado por la implementación de un nuevo esquema tecnológico en la segunda mitad de este periodo, según especialistas. Sin embargo sería inmaduro atribuir la alteración del rendimiento en este periodo a dicho esquema pues se debe considerar que alrededor del 97% del área cultivada de piña en México pertenece a superficies de temporal, por lo que se debe atribuir parte de estas alteraciones al factor clima.

Las alteraciones que se presentan en este indicador durante el periodo tratado se traducen en una reducción a nivel nacional de aproximadamente 1.66% en promedio anual ya que pasa de 51.97 ton./ha., en 1990 a 43.35 ton./ha. para el año 2000 en contraste al periodo anterior donde pasa de 42.20 en 1980 a 52.10 ton./ha en 1989 para tener un incremento promedio anual de 2.34%. Como dato adicional se tiene que para este periodo el rendimiento promedio es de 42.18 toneladas por hectárea en tanto que para el periodo anterior (1980-1990) el rendimiento promedio

había sido de 45.8 ton./ha., lo cual indica que el rendimiento se redujo en términos de promedio por década en un 7.90% para este periodo.

Además, debe notarse que el rendimiento promedio de nuestro país es sumamente superior al rendimiento promedio a nivel mundial el cual es de 18 toneladas por hectárea durante el mismo lapso, incluso es también superior hasta en un 33.61% al de los países con mayor rendimiento dentro de los principales productores los cuales son Brasil y Filipinas con rendimientos promedios de 28 toneladas por hectárea durante el mismo lapso.

Si se analiza la situación que guarda cada uno de los principales estados productores se tiene que para el caso de **Veracruz**, se presenta una reducción en el rendimiento de aproximadamente 1.91% en promedio anual al pasar de 55.8 ton./ha en 1990 a 45.13 para el año 2000.

En el estado de **Oaxaca**, contrario a lo que ocurre en Veracruz, experimenta un incremento en su rendimiento de aproximadamente 0.80% en promedio anual al pasar de 48.8 a 52.69 ton./ha., durante el mismo lapso.

Por otro lado, los estados de Tabasco, Nayarit y Jalisco presentan tanto incrementos como reducciones en el rendimiento de aproximadamente de -2.59%, -6.01% y 159.36% en promedio anual, respectivamente. De estos estados, Nayarit es quien presenta menor rendimiento dado que su rendimiento promedio anual para este periodo es de 14.29 toneladas por hectárea, en tanto que para Tabasco y Jalisco llegan a ser de 35.99 y 46.96 ton./ha., respectivamente.

## **2.5.2 Comercialización interna y externa.**

Como se mencionó anteriormente, la piña que se produce en México tiene como destino final tres usos: el mercado en fresco que absorbe el 70% de la producción nacional; la industria que absorbe entre el 23 y 25% de la producción y las exportaciones en fresco con el porcentaje restante.

Cada uno de estos destinos o consumos define las características de la fruta que requiere a fin de obtener los mejores precios o bien como en el caso de la industria los mayores rendimientos. Así, el mercado nacional exige una fruta de clasificación A, es decir, por arriba de los 2.5 kilogramos, en tanto que la industria requiere el producto tipo B, entre 1.8 y 2.5 kg. y finalmente el mercado de exportación tiene entre sus requerimientos frutas tipo C con pesos de entre 1.5 y 1.8 kilogramos. Durante mucho tiempo se consideró que la fruta destinada al mercado de exportación era en realidad la producción de desecho debido a que no cubría los requerimientos tanto del mercado nacional en fresco como el de la industrial. En la actualidad esta visión ha cambiado a tal grado que en la región productora (La región del Bajo Papaloapan) se está creando en algunos productores una cultura de exportación, al producir especialmente los tamaños y calidad demanda en este mercado.

Los destinos de la piña en México están influenciados por la temporalidad de la época de cosecha. En los meses de septiembre-noviembre, lapso en el que hay escasez del producto, se destina principalmente a la venta en fresco en el mercado nacional; en tanto que durante el lapso diciembre-marzo, que es el periodo en el que la mayoría de los productores de la región del Bajo Papaloapan obtienen su cosecha, la industria se convierte en un importante receptor de la abundante oferta.

Las exportaciones también se hacen durante la etapa alta de producción, por lo que durante la primera mitad del año (enero-junio) se realizan aproximadamente el 67% de las exportaciones nacionales.

Para la comercialización nacional en fresco las centrales de abasto son el mecanismo básico de distribución y de ellas destacan las centrales del Distrito Federal adquiriendo volúmenes de entre el 30 y 35%, la central de Abasto de Monterrey con 20% y la de Guadalajara con 10%, en tanto que el porcentaje restante es absorbida por otras plazas como las de Puebla, Chihuahua, etc.

La producción de piña además del problema de estacionalidad, enfrenta también el de la comercialización, ya que la cadena de mercado se hace grande por la intervención de diversos agentes, a tal grado que se considera que los productores reciben la cuarta parte del precio pagado por el consumidor final.

Los canales de comercialización y sus diversas modalidades que se distinguen son esencialmente las siguientes:

- ✓ **Compra – Venta en báscula.** En este pueden encontrar dos modalidades.
  - a) **Con participación de intermediario.** Este es el caso más frecuente. El productor se vincula con los compradores de piña a través de un intercambio cuyo papel es contactar oferentes y demandantes, recibiendo a cambio una comisión por su intervención. Puestos en contacto, ambas partes pactan los volúmenes, precios y calidades de la fruta en transacción, siendo los compradores los que se encarguen de cubrir los costos de las labores de cosecha y transporte, que incluyen las actividades de selección, corte, estiba y transporte a la zona consumidora. El intermediario tiene una comisión de aproximadamente el 10% del precio al que se vende el producto. En el caso de los acopiadores o compradores éstos pueden ser regionales o foráneos, los primeros trabajan más con la industria local, mientras que los segundos se encargan de abastecer la oferta de los comerciantes mayoristas en diversas ciudades e industrias ubicadas fuera de la región productora.

b) **Sin participación de intermediario.** En esta modalidad, los productores buscan prescindir de los intermediarios, de manera que se vincula directamente el productor con los compradores, los cuales siguen siendo los mismos.

- ✓ **Compra directa a productores.** Este canal de comercialización es el menos frecuente en la región del Bajo Papaloapan. Se caracteriza porque el demandante compra directamente al productor quien costea los gastos de corte, estiba, báscula y guía, mientras que el comprador cubre los gastos de flete, para el traslado de la fruta desde el lugar de compra hasta la plaza de venta. Este tipo de comercialización, permite garantizar a los productores la colocación de importantes volúmenes, sobre todo en la época de temporada alta, que es cuando se presentan los problemas de comercialización. A los compradores o demandantes les garantiza obtener volúmenes en las calidades y tiempos requeridos, sobre todo cuando la cosecha alcanza su punto mínimo, además de que los costos son comúnmente menores.
- ✓ **Envíos de grandes productores a bodegas de diversas centrales de abasto.** Es el canal menos representativo aunque no se dispone de información sobre los volúmenes comercializados de este modo. Se distingue porque ofrece la ventaja a los grandes productores de controlar e integrar tanto a la fase de producción como la de comercio, lo que necesariamente incide en los costos y en los márgenes de ganancia.

Es importante aclarar que los canales de comercialización de piña que se distinguen en la zona del Bajo Papaloapan, están vinculados directamente con las características y tamaño de los productores, las cuales se mencionaron anteriormente.

### **2.5.2.1 Exportaciones e Importaciones de piña durante el periodo 1990 – 2000.**

Las exportaciones de piña fresca en México durante el periodo 1990-2000 experimentaron un incremento de aproximadamente 18.11% en promedio anual al pasar de 8 mil 683 toneladas durante al año de 1990 a 24 mil 409 toneladas para el año 2000. En promedio se exportaron alrededor de 13 mil 76 toneladas anuales a lo largo del lapso habiéndose reducido dicha cantidad hasta en un 36.35% si se compara con el periodo 1980-1989 donde en promedio se exportaron 20 mil 545 toneladas anuales.

Las razones por las que se presenta dicha reducción se debió principalmente a dos situaciones por un lado, el comportamiento de los volúmenes de producción representan un factor decisivo puesto que el incremento de la producción permite tener volúmenes extra para comercializar en el exterior tomando en cuenta que durante el periodo mencionado el consumo aparente de piña, representó alrededor del 95% de la producción.

El otro factor que influyó fue el reacomodo de la producción mundial apareciendo en la escena países asiáticos como Filipinas, Tailandia y Taiwán, a demás de Costa Rica y Honduras en el continente Americano. Esto impacta en el mercado mundial dando como consecuencia que de las importaciones de fruta fresca o congelada que hizo Estados Unidos durante el periodo 1990-1998, México sólo contribuyó con el 5.86% en tanto que Costa Rica y Honduras lo hicieron con el 62.9% y 17.82%, respectivamente.

El mercado Norteamericano se convierte en el caso de las exportaciones mexicanas en un referente necesario, ya que la mayor parte de las exportaciones de piña del periodo 1997-2000 tuvieron como destino principal Estados Unidos, seguido en menor medida por Canadá y otros Países.

Por otro lado, las importaciones de piña fresca durante el mismo periodo se puede decir que fueron prácticamente insignificantes pues durante todo el lapso sólo se importaron un total de 769 toneladas con un valor acumulado de 1 millón 360 mil dólares.

Las exportaciones de piña conservada o preparada tuvieron un comportamiento similar al de las frescas o secas. Durante la primera mitad de los noventa, muestran un comportamiento a la baja debido a los niveles de producción, en tanto que a partir de 1995 inician un proceso de recuperación teniéndose como saldo final que durante el periodo en descripción la menor cantidad de piña preparada o conservada exportada fue de 852 toneladas, en tanto que la mayor llegó a ser de 4,376 toneladas. En promedio a lo largo del periodo se exportaron alrededor de 2,875.8 toneladas anuales con un valor aproximado de 2,060.82 miles de dólares.

Las importaciones de esta misma modalidad de piña se ubicaron en un promedio 3,577.18 toneladas anuales, teniendo como cantidad mínima 367 toneladas y un máximo de 19 mil 15 toneladas durante el lapso en descripción. Al final del periodo, se importaron un total de 39 mil 349 toneladas con un costo acumulado de 31 millones 459 mil dólares. Si se comparan los volúmenes de importación y exportación de piñas preparadas se obtendrá que al final del periodo se importaron 7,715 toneladas más de lo que se exportó, lo cual implica que hay una demanda de producción procesada no satisfecha por la industria mexicana debido a diversos factores como:

- El desfase tan profundo que existe entre industria y producción que hace que la primera funcione sólo en épocas de producción alta, cuando hay sobreproducción o bien cuando el precio de la fruta fresca es muy bajo.
- La agroindustria de piña funciona con bajos niveles de diversificación en las líneas de producción, así como tecnología obsoleta.
- Esto ha generado que se importe volúmenes de otros países, los que tienen ventajas en costos de producción y financiamiento.

En las exportaciones de piña procesada a diferencia de la fresca o seca, se puede observar una diversificación de los países de destino. Por ejemplo, se considera que durante el periodo de 1997-2000, del total de los volúmenes comercializados en el extranjero, el 58.82% se destina a EE.UU., mientras que Argentina el 18.08, Chile 10.14% y Canadá 9.88%, el porcentaje restante se exporta a diversos países.

Para las importaciones, también se observa una diversificación de países de origen. En este caso EE.UU. es el principal abastecedor de nuestro país con 53.83% del total de los volúmenes adquiridos en el extranjero. Le siguen Tailandia con 32.85%, China 7.78%, Indonesia 2.57% y Filipinas con 1.15% mientras que el porcentaje restante corresponde a diversos países. Estos datos señalan por lo tanto el papel tan activo que tienen los países asiáticos en las importaciones de piña procesada, ya que si sumamos su participación, abastecen a nuestro mercado con casi el 45%.

**Gráfico 8:** Comportamiento de las exportaciones e importaciones de piña conservada o procesada durante el periodo 1990-2000.



Fuente: Base de Datos de la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO), con información actualizada hasta el año 2002.

### **2.5.2.2 Comportamiento de los precios y rentabilidad del producto.**

Se sabe que el precio de cualquier producto es el principal factor que interviene para que se establezca el punto de equilibrio en la cantidad demandada y ofrecida, es decir, el punto en el cual tanto demandantes como oferentes establecen volúmenes de producto y precio de tal manera que no existe una sobreoferta o una demanda insatisfecha. Las fluctuaciones en el precio de cualquier producto están determinadas, de acuerdo a la teoría microeconómica, principalmente por los niveles de producción y la cantidad demandada que a su vez es determinada por el precio del producto y la existencia de sustitutos.

Así, dado que el comportamiento del precio de cualquier producto juega un papel muy importante en el consumo o comercialización de la misma, en este apartado se hace una breve descripción sobre el comportamiento del precio medio rural de la piña en los últimos años debido a que este precio quien determina el comportamiento de los precios al mayoreo en las diversas centrales de abasto del país y con ello el nivel de consumo del mismo.

El precio medio rural de la piña ha tenido en los últimos 10 años una tendencia alcista, es decir los precios se han incrementado aunque de una manera no muy notoria en algunos años del periodo 1990-2000. Los grandes volúmenes de producción de los años 1989 y 1990, generaron un proceso de sobreoferta que tuvo implicaciones directas en el precio, de modo que durante los años 1990-1994, este no fue superior a los \$1000 pesos por tonelada. Sin embargo, a partir de 1995, como producto de una reducción en la producción, el precio inicia un periodo de recuperación a tal grado que para 1998 se duplica al pasar de \$1,015.78/tonelada (precio del año anterior) a \$2,118.23/tonelada, es decir, prácticamente el doble.

Una posible explicación a este suceso es que para este año se registra la menor cantidad de superficie cosechada a pesar de que la cantidad sembrada se había duplicado con respecto al año anterior. Este último hecho implica que las

inversiones de los productores se acrecentaran en gran medida durante el mismo año sin incrementarse la producción, por lo que la única manera de al menos recuperar lo invertido fue el incremento del precio de la fruta.

En los siguientes años, el precio medio rural no varía en gran proporción con respecto al establecido en 1998 ya que para los años 1999 y 2000 éstos son de \$2166.18/tonelada y \$2142.21/tonelada, respectivamente.

Por otro lado, el comportamiento del precio al mayoreo en las principales centrales de abasto permite apreciar claramente la estacionalidad que tiene el producto al registrarse las mayores cotizaciones con la etapa de menor producción, en tanto que durante los primeros meses se registran las menores cotizaciones debido a la abundancia que se tiene del producto.

Describiendo el comportamiento de los precios en las principales centrales de abastos tenemos que en la central de abasto de Iztapalapa en el Distrito Federal, por ejemplo, se ha presentado en los últimos cuatro años (1998-2001) poca variación en los precios al encontrarse durante este lapso un mínimo de 6.1 pesos por cada pieza de piña con tamaño mediana y un máximo de 6.88 pesos/pieza; en tanto que en el Mercado de Abasto de Guadalajara los precios mantuvieron un mínimo de \$2.75/Kg y un máximo de \$3.77/Kg durante el mismo lapso. En el mismo intervalo, en la central de abasto de Guadalupe, Nuevo León se presenta un costo mínimo de \$2.67/Kg y un máximo de \$3.65/Kg.

En términos generales, el comportamiento de los precios al mayoreo de la piña en la central de abasto del Distrito Federal durante el periodo 1993-2001 se resume diciendo que ésta tuvo un ligero incremento de apenas 0.43% en promedio anual al pasar de \$6.1/pieza en 1993 a \$6.31/pieza para el año 2001. El precio promedio por cada pieza de tamaño mediana a lo largo del periodo en descripción fue de aproximadamente \$5.24.

En lo que respecta a las centrales de abasto de Guadalajara y Nuevo León, éstas tuvieron tasas de cambio aproximadamente -1.04 y 3% en promedio anual respectivamente, además de que los precios promedio durante el lapso fueron de \$2.94/Kg y \$3.26/Kg, respectivamente. Nótese que el comportamiento de la variable precio en éstas dos centrales es casi similar a excepción de 1996, año en el que el precio de Venta en la Central de Abasto de Guadalupe en Nuevo León duplica prácticamente al precio dado en el Mercado de Abasto de Guadalajara.

Si habla de los costos de producción de la piña, se estará deduciendo aunque de manera bosquejada la rentabilidad de este producto. De acuerdo a la información proporcionada por los productores de la región del bajo Papaloapan, el costo de producción de una hectárea para el año 2000 ascendió a \$40543.4, los cuales se distribuyen en diversas actividades o insumos siendo tres las que absorben el 60% del costo de producción total.

La primera se refiere a la siembra del producto, a la que corresponde el 17.54%, de esta actividad el mayor gasto se vincula con el material vegetativo. La segunda esta relacionada con la fertilización con un 25.91% siendo el insumo de mayor gasto el NPK (Nitrógeno, Fósforo y Potasio). Por otro lado, el tercer gasto se refiere a las labores de cosecha las cuales absorben alrededor del 16.16%; y finalmente el porcentaje restante se distribuye en otras actividades como: la preparación del terreno, control de maleza, control de plagas, etc.

Así para establecer un perfil de la rentabilidad de esta fruta, se considera como costo de producción la cantidad de \$40,000 por tonelada. Ahora, si considera que el rendimiento promedio para el mismo año en el estado de Veracruz fue de cuando menos 45 toneladas, se estaría hablando de un costo de producción de \$889 por tonelada.

Los precios del producto en las temporadas de alta, media y baja producción han variado en máximos de 0.60, 1.60 y 3.0 pesos por kilogramo, por lo que en promedio se estaría hablando de un precio de 1.73 pesos por kilogramo, lo que nos

arrojaría una cantidad de \$1733.33/tonelada. Por lo tanto, a este precio necesitaríamos alrededor de 23 toneladas para cubrir los costos de producción, obteniendo así una utilidad de 22 toneladas por hectárea y concluir con esto que el producto es muy rentable.

Por otro lado, los precios mínimos para las mismas temporadas se ubicaron en 0.20, 0.80 y 1.90, arrojando un promedio final de 0.97 pesos por kilogramo. A este precio necesitaríamos alrededor de 41.4 toneladas para cubrir los costos de producción, por lo que nuestra utilidad sería únicamente de 3.6 toneladas.

En resumen, la rentabilidad de la producción de piña en México depende mucho de la estacionalidad que presenta, pues en la temporada de baja producción o de escasez el producto resulta ser muy rentable; sin embargo en la temporada de alta producción puede no ser rentable.

### **2.5.2.3 Importancia económica del producto.**

En México, la producción de piña tiene una larga tradición convirtiéndose no sólo en una importante actividad económica para algunas regiones de país, sino también, de vital importancia social por la gran cantidad de jornales y empleos que se generan tanto en la parte productiva como en la comercialización e industrialización de ésta fruta.

La importancia económica que tiene éste producto en nuestro país consiste básicamente en que representa una fuente generadora de ingresos para cerca de 10,000 familias. Su producción se dedica básicamente a satisfacer las necesidades del mercado interno, sin embargo, la exportación en fresco representa la alternativa de mayor crecimiento potencial y con ello una mayor captación de divisas. Las exportaciones en esta modalidad de piña generaron durante el intervalo 1998-2000 un total acumulado de 21 millones 290 mil dólares.

Si hablamos en términos del valor que representan las producciones tenemos por ejemplo que para el año 2000, con una producción de alrededor de 500 mil toneladas y un precio medio rural de \$2000/tonelada, se tendría un valor de 1 billón de pesos, lo cual representaría alrededor del 0.015% del PIB Nacional en términos reales, dado que éste último fue de aproximadamente 6 mil billones 413 mil millones 46 mil 159 pesos.

### 2.5.3 Normas de calidad de la piña.

La piña para la industria es demandada fundamentalmente por las empresas procesadoras de la región en su conjunto y por otras plantas procesadoras de piña ubicadas en diferentes centros de consumo. De estas destacan por su capacidad instalada y su volumen de producción:

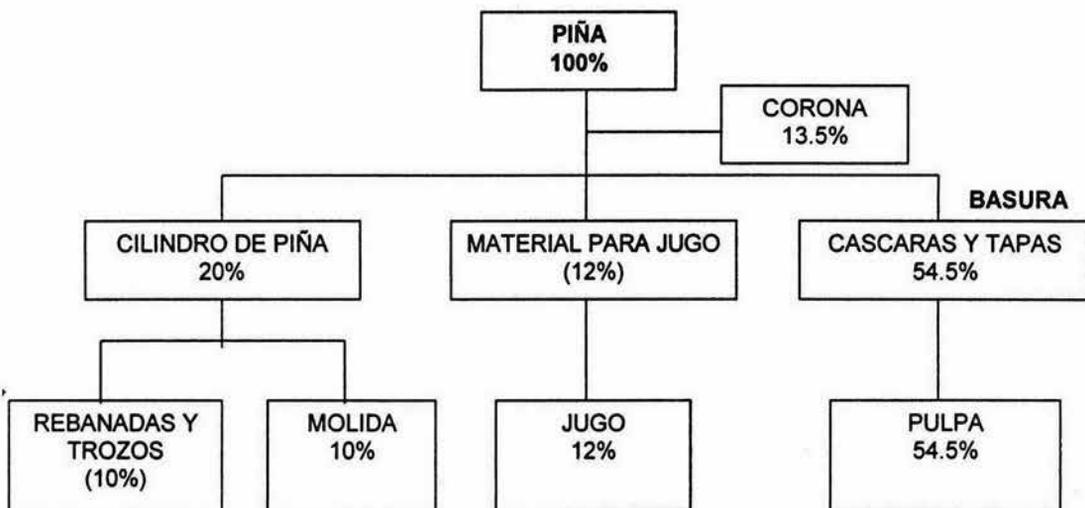
No.	EMPRESA
1	Conservas La Torre, S.A. de C.V.
2	Clemente Jacques y Cía., S.A. de C.V.
3	Jugos del Valle, S.A. de C.V.
4	Kraft Foods de México, S.A. de C.V.
5	Productos Loma Bonita, S.A. de C.V.
6	Herdez, S.A. de C.V.
7	Tropifrut, S.A. de C.V.
8	Jugomex, S.A. de C.V.
9	Jugos del Centro, S.A. de C.V.
10	Agroindustrias Loma Bonita, S.A. de C.V.
11	Conservas Pegaso, S.A. de C.V.
12	Empacadora del Trópico, S.A. de C.V.

La industria cuyo principal producto son las rebanadas, chunks, tidbits, molida y juegos requiere frutos de 1.4 a 1.8 y de 1.8 a 2.3 kg., con un tamaño de 4" hasta 5"  $\frac{1}{2}$  de diámetro, dependiendo el tipo de rebanada a elaborar de 2 o 2 $\frac{1}{2}$ , como un mínimo de grado brix de 10 y la madurez de la piña de hoja pinton, pintcona,  $\frac{1}{2}$  cara de madura y  $\frac{3}{4}$  de madura .

Los principales productos que se obtienen por el proceso de la piña se muestran en la siguiente tabla:

<b>DESCRIPCIÓN DE LOS PRODUCTOS</b>	
<b>Rebanadas.</b>	Es un trozo completo cortado transversalmente del eje del cilindro de la piña cuando ha sido pelada y eliminado el corazón.  Número de rebanadas por bote del número 2 $\frac{1}{2}$ 8 con peso de 80 gramos por rebanada de piña fresca.  Número de piezas por bote del número 10,50 y bote número 2 , 10 rebanadas del número 2, peso 43 gramos por rebanada de piña fresca.
<b>Chunks.</b>	Son trozos obtenidos al cortar en 8 partes una rebanada de 2 $\frac{1}{2}$ con un grosor especial de $\frac{3}{4}$ " , largo 1" y un grosor de $\frac{1}{2}$ " , y con un número más o menos de 170 piezas por bote del número 10.
<b>Tidbits.</b>	Son pequeños triángulos obtenidos al cortar en 16 partes una rebanada, con un largo de 1" y un grosor de $\frac{1}{2}$ " , y con un número mas o menos de 470 piezas por bote del número 10.
<b>Mermeladas.</b>	
<b>SUBPRODUCTOS</b>	
<b>Cáscara de piña y corona</b>	Elaboración de vinagre y tepache para la propagación de cultivo.

## APROVECHAMIENTO REAL DE LA PIÑA



**NOTA:** La rebajada de piña del 2½ pesa 80 Grs. cada una, el bote se lleva 8 y 24 botes por caja. En total se necesitan 15.36 Kgs. por caja de producto terminado.

La rebajada de piña del 2 pesa 43 Grs. cada una, el bote se lleva 10 y 24 botes por caja. En total se necesitan 10.32 Kgs. por caja de producto terminado.

La rebajada de piña del 2 pesa 43 Grs. cada una, el bote se lleva 50 y 6 botes por caja. En total se necesitan 12.90 Kgs. por caja de producto terminado.

La norma oficial mexicana para la piña, NOM-FF-28-1982, establece las características de calidad que debe cumplir la piña (Ananas comosus) en estado fresco, de la variedad Cayena lisa, destinada al consumo humano directo.

La norma define a la piña como la fruta cuyo color va del verde oscuro al amarillo, de sabor y olor característico, perteneciente a la familia de las Bromeliáceas del género Ananas y especie Comosus. La clasifica en tres grados de calidad, en orden descendente: México Extra, México No. 1 y México No. 2. El producto no

clasificado de acuerdo con los grados enunciados previamente se designa como "No clasificado" término que denota, exclusivamente, que no se ha dado algún grado de calidad al lote.

En este documento se estipula que la piña debe cumplir con las siguientes especificaciones sensoriales: ser fresca, limpia, sana, entera y bien desarrollada; tener forma, sabor y olor característico; estar exenta de humedad exterior y prácticamente libre de descomposición o pudrición; libre de defectos de origen mecánico, meteorológico, microbiológico o genético-fisiológico, y tener pedúnculo con longitud máxima de un centímetro sin infecciones y un máximo de tres bacterias.

La norma indica, además, las especificaciones de madurez de la piña. El grado de madurez se determina por el contenido de sólidos solubles totales y acidez titulable en punto sazón; los sólidos solubles no deben ser menores de 9% y la acidez titulable no debe de ser mayor de 0.9%. La piña, de acuerdo a tales especificaciones, alcanza su punto sazón, cuando adquiere un tono verde oscuro y las bayas se tornan planas y bien formadas. El fruto se clasifica en función de su madurez en punto sazón (sazona), de "ojo" "pintona",  $\frac{1}{4}$  de madurez,  $\frac{1}{2}$  madurez,  $\frac{3}{4}$  de madurez y madura.

En cuando a las correspondientes especificaciones de defectos y de presentación, la norma establece los siguientes criterios: la calidad México Extra debe estar libre de cualquier defecto y dentro de las tolerancias establecidas. Asimismo las piñas de esta calidad deben ser envasadas siguiendo una rigurosa selección, dejando cada envase perfectamente presentado, su aspecto global debe ser uniforme en cuanto a color y tamaño aunque en estos casos se acepta una tolerancia de 5% por lote.

Por su parte, las piñas de calidad México No. 1 pueden presentar como máximo un defecto menor por fruta y las de la calidad México No. 2 un defecto mayor. En ambos casos los defectos deben estar dentro de las tolerancias establecidas para estas categorías. Las piñas de las calidades mencionadas, pueden

presentar variaciones en cuanto a homogeneidad de color y tamaño, y estar dentro de las tolerancias establecidas para estos aspectos, en el caso de la calidad México No.1 son del 10% por lote y de 15% para la México No. 2

### 2.5.3.1 Tipos de defectos y tolerancias.

De acuerdo a la norma, en la piña de todos los grados de calidad se permiten las siguientes tolerancias de defectos.

TIPOS DE DEFECTOS	PUNTO DE EMBARQUE	PUNTO DE ARRIBO
CITRICOS	4%	5%
MAYORES	6%	7%
MENORES	10%	12%
ACUMULATIVO	10%	12%
PUDRICCION	0.5%	1%

En las tolerancias de color, tamaño y defecto, el porcentaje permitido se da para el lote, el que no corresponda a la designación declarada se evalúa por conteo.

De acuerdo a la norma, los residuos tóxicos están sujetos a las tolerancias establecidas por la Secretaría de Agricultura, Ganadería y Desarrollo Rural y la Secretaría de Salud, incluyendo aquellos que corresponden a los residuos de plaguicidas, productos mejoradores de las apariencias y otros.

Se establece asimismo, lineamientos relativos a marcado, etiquetado, envase y embalaje, que cada envase debe llevar en el exterior en etiqueta o en impresión permanente, con caracteres legibles e indelebles, redactados en español o en otro idioma cuando el producto sea para exportación y el importador lo requiera, que tenga como mínimo los datos siguientes: piña en estado fresco; identificación simbólica del producto o envasador; nombre y dirección del productor, distribuidor o exportador y, cuando se requiera, el del importador; zona regional de producción y la leyenda "Producto de México"; fecha de envasado; designación del producto y contenido neto en gramos o kilogramos.

En cuanto a envasado y presentación, la norma estipula que la fruta no debe sobresalir del nivel superior del envase, mismo que deben reunir la calidad y resistencia que garanticen el estibado y transportación al lugar de consumo. Señala la norma que los envases pueden ser de cartón corrugado rígido u otro material aceptable y contiene, de las dimensiones que se adapten a las necesidades de transporte nacional e internacional, y reunir las condiciones de higiene, ventilación resistencia a la humedad y temperatura que garanticen una adecuada conservación de la fruta y su manejo.

### 3. EL MODELO DE REGRESIÓN LINEAL

A sido inquietud del ser humano encontrar el por qué de la ocurrencia de la infinidad de fenómenos que suceden a su alrededor. Con el desarrollo de las matemáticas fue posible el análisis de una gran cantidad de estos fenómenos, encontrándose las formas y tipos de relaciones que existen en los diversos factores que intervienen en la ocurrencia de un fenómeno determinado.

Así, con la ayuda de las matemáticas fue posible encontrar la relación entre la velocidad, distancia y tiempo, esto mediante el modelo  $V = d/t$ ; entre la masa de un cuerpo, su energía y la velocidad de la luz por medio del modelo  $e = mc^2$ ; etc.

Sin embargo se han encontrado fenómenos cuyas relaciones entre sus factores no son exactas, y no es posible encontrar modelo alguno que exprese tal relación. Como ejemplo de éstos fenómenos podemos citar; los gastos de consumo, de una familia, relacionados con sus ingresos, número de elementos, religión, etc.; la estatura del ser humano relacionada con su peso, edad, sexo, etc. A éste tipo de fenómenos se les llama fenómenos aleatorios, aunque se sabe que se presentan, no se sabe con seguridad cual es su resultado. Por ejemplo sabemos que para una persona de cierta edad le corresponderá una estatura, pero no podemos asegurar que estatura, esto se debe a que en la estatura intervienen otros factores, además de la edad, que no se toman en cuenta.

Las matemáticas no fueron suficientes para tratar estos problemas. Fue hasta que surgió el concepto de probabilidad y su desarrollo que fue posible establecer modelos para las relaciones entre factores que intervienen en un fenómeno aleatorio. La idea principal para la construcción de éstos modelos es aceptar que existe una parte del fenómeno que si se puede explicar con exactitud y otra parte que es aleatoria, por ejemplo: podemos decir que para cierta edad se debe tener una estatura promedio mas una cantidad aleatoria que puede ser positiva o negativa, esta cantidad aleatoria tiene su origen en los otros factores que no se toman en cuenta para determinar la estatura. Esta parte o cantidad aleatoria siempre está presente y cuando el fenómeno se puede observar un gran número de veces bajo idénticas condiciones, presenta cierta regularidad. Esta regularidad se estudia por medio de la Estadística, a través de la teoría de probabilidad.

Es así como por medio de las matemáticas y de la probabilidad fue y es posible crear modelos para fenómenos aleatorios, a estos modelos se les conoce como Modelos Estadísticos o Modelos Estocásticos.

Los Modelos Estadísticos más sencillos son los Modelos Lineales, que a su vez se dividen en: Modelos de Regresión y Modelos de Análisis de Varianza.

### **3.1 CONCEPTOS FUNDAMENTALES EN EL ANALISIS DE REGRESIÓN.**

#### **3.1.1 Origen del término regresión.**

La palabra "regresión" se usó por primera vez en este contexto por Francis Galton (1822-1911) en sus estudios biológicos sobre la herencia. En ellos se notó que las características promedio de la siguiente generación de un grupo en particular tendían a moverse en la dirección de las características promedio de la población general, más que hacia las de la generación previa de ese grupo. Esta tendencia fue referida como una regresión hacia la media de la población.

De manera básica, la regresión tiene dos significados, uno surge de la distribución conjunta de probabilidad de dos variables aleatorias; el otro es empírico y nace de la necesidad de ajustar alguna función a un conjunto de datos.

### **3.1.2 Interpretación del análisis de regresión.**

El objetivo primordial del análisis de regresión es estimar el valor de una variable aleatoria (*la variable dependiente*) dado que el valor de una variable asociada (*la variable independiente*) es conocido. La variable dependiente también se llama *variable de respuesta*, mientras que la variable independiente también se llama variable de *predicción*. La ecuación de regresión es la fórmula algebraica por la cual se determina el valor estimado de la variable dependiente, o de predicción. Las aplicaciones de regresión son numerosas y ocurren en casi todos los campos, incluyendo la ingeniería, las ciencias físicas, las ciencias biológicas, las ciencias sociales, economía, entre otras.

### **3.1.3 Las relaciones causa – efecto.**

El modelo de regresión no implica una relación causa – efecto entre las variables. Si bien una relación empírica fuerte puede existir entre dos o más variables, esto no puede considerarse como una evidencia de que las variables predictoras y las variables respuestas estén relacionadas como una relación causa – efecto. Para establecer causalidad, la relación entre las variables regresoras o predictoras y las variables respuesta deben tener un fenómeno físico y lógico. Por ejemplo, la relación puede ser sugerida por consideraciones teóricas. El análisis de regresión puede ayudar en confirmar una relación causa – efecto, pero no puede ser la base única de tal reclamo.

### 3.1.4 Usos de la regresión.

Los modelos de regresión se usan para varios propósitos, incluyendo los siguientes:

1. La descripción de datos.
2. La estimación de parámetros.
3. Predicción y estimación.
4. Control.

Los profesionistas y los científicos frecuentemente usan las ecuaciones para resumir o describir un conjunto de datos. El análisis de regresión es útil en el desarrollo de tales ecuaciones. Por ejemplo, podemos recolectar una cantidad de datos donde estén involucradas el volumen de entrega y el tiempo de entrega, y un modelo de regresión probablemente sería mucho más conveniente que el uso de una simple tabulación de esos datos.

A veces, problemas de estimación del parámetro pueden ser resueltos por métodos de regresión. Por ejemplo, suponer que un circuito eléctrico contiene una resistencia desconocida de  $R$  ohms. Las diferentes corrientes conocidas que pasan mediante el circuito y los voltajes correspondientes medidos, pueden ser la base para estimar tal resistencia. El diagrama de dispersión indica la relación de ese voltaje y la corriente a través del origen, pasando por una línea recta, la pendiente es  $R$  (porque el voltaje  $E$  y la corriente  $I$  están relacionadas por la ley de Ohm:  $E=IR$ ). El análisis de regresión puede usarse para ajustar este modelo a los datos, y producir una estimación de la resistencia desconocida.

Muchas aplicaciones de regresión involucran estimación de la variable respuesta. Por ejemplo, se puede predecir el tiempo de entrega para un número específico de casos de bebidas que deberá ser entregado. Estas estimaciones, pueden ser útiles en la planificación de actividades de entrega tal como la programación y evaluación de la productividad en las operaciones de entrega.

Los modelos de regresión pueden usarse con propósitos de control. Por ejemplo, un ingeniero químico podría usar el análisis de regresión para desarrollar un modelo relacionando la fortaleza tensora de papel con la concentración en la pulpa de la madera. Esta ecuación podría entonces ser usada para controlar en valores convenientes para variar el nivel de concentración en la pulpa. Cuando una ecuación de regresión se usa para los propósitos de control, es importante que las variables estén relacionadas en una manera causal. Note que una relación causa – efecto no puede ser necesaria si la ecuación sólo es usada para el pronóstico. En este caso es sólo necesario si la relación que existe entre los datos originales que son usados para construir la ecuación de regresión son todavía válidos. Por ejemplo, el consumo diario de electricidad durante un mes determinado en un lugar dado, puede ser un buen estimador para la temperatura diaria máxima en ese mes. Sin embargo, cualquier intento de reducir la temperatura máxima del consumo de electricidad se condena claramente al fracaso.

El análisis de regresión es una técnica estadística que se utiliza para investigar y modelar la relación entre variables.

### 3.2 EL MODELO DE REGRESIÓN LINEAL SIMPLE.

Se considera que el modelo de regresión lineal simple, es un modelo con una sola variable regresora  $X$  que tiene una relación con una variable de respuesta  $Y$  que es una línea recta. Este modelo de regresión lineal simple es:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (3.1)$$

donde la ordenada al origen  $\beta_0$  y la pendiente  $\beta_1$  son constantes desconocidas y el  $\varepsilon$  es un componente aleatorio o de error. Los errores son distribuidos con media cero y varianza desconocida  $\sigma^2$ . Además comúnmente se asume que los errores son no correlacionados. Esto significa que el valor de un error no depende del valor de cualquier otro error.

Es conveniente considerar que la variable predictora  $X$  es controlada por el analista y medida con un error insignificante, mientras que la variable respuesta  $Y$  es una variable aleatoria. Es decir, hay una distribución de probabilidad para  $Y$  en cada valor posible para  $X$ . La media de esta distribución es:

$$E(Y/x) = \beta_0 + \beta_1 x \quad (3.2)$$

y la varianza es:

$$V(Y/x) = V(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2 \quad (3.3)$$

Entonces la media de  $Y$  es una función lineal de valor de  $X$  aunque la varianza de  $Y$  no depende del valor de  $X$ . Además, debido a que los errores son no correlacionados, las respuestas son también no correlacionadas. Los parámetros son comúnmente llamados coeficientes de regresión.

La pendiente  $\beta_1$  es el cambio en la media de la distribución de  $y$  producida por una unidad cambiada en  $X$ . Si el intervalo de datos sobre  $X$  incluye  $x = 0$ , entonces el intercepto  $\beta_0$  es la media de la distribución de la respuesta  $Y$  cuando  $x = 0$ . Si el intervalo de valores de  $X$  no incluye cero, entonces el intercepto  $\beta_0$  no tiene interpretación práctica.

Dado que la ecuación (3.1) involucra una sola variable de regresión, ésta se conoce como un modelo de regresión lineal simple. En general, la variable respuesta puede ser  $K$  dimensionada con varios componentes  $x_1, x_2, \dots, x_k$  esto es:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (3.4)$$

Esta ecuación se conoce como un modelo de regresión lineal múltiple, ya que en este modelo está involucrada más de una variable regresora. El adjetivo lineal es empleado para indicar que el modelo es lineal es los parámetros  $\beta_0, \beta_1, \dots, \beta_k$  y no porque  $Y$  sea una función lineal de  $X$ .

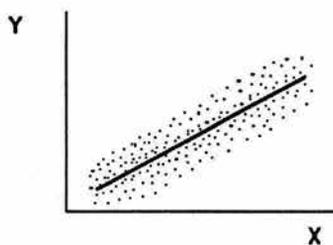
### 3.2.1 Diagrama de dispersión.

Un diagrama de dispersión es una gráfica en la que cada punto trazado representa un par de valores observados de las variables independiente y dependiente. El valor de la variable independiente  $X$  se identifica respecto del eje horizontal, mientras que el valor de la variable dependiente  $Y$  se identifica respecto del eje vertical.

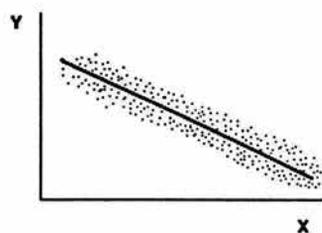
La forma de la relación representada por el diagrama de dispersión puede ser curvilínea más que lineal. Aunque el análisis de regresión de relaciones curvilíneas escapa al interés de este trabajo. En el caso de relaciones no lineales, un enfoque frecuente consiste en determinar un método de transformación de valores de una o ambas variables a fin de que la relación de los valores transformados sea lineal. Así, el análisis de regresión lineal puede aplicarse a los valores transformados, y los valores de la variable dependiente pueden retransformarse a la escala de medición original.

Si el diagrama de dispersión indica en general una relación lineal, se ajusta una línea recta a los datos. La ubicación precisa de esta línea es determinada por el método de mínimos cuadrados, (que desarrollaremos en la siguiente sección). Tal como se ilustra en el Ejemplo 1, una línea de regresión con pendiente positiva indica una relación directa entre las variables, una pendiente negativa indica una relación inversa entre las variables y una pendiente de cero indica que las variables no tienen relación entre sí. Además, el grado de dispersión vertical de los puntos trazados respecto de la línea de regresión indica el grado de relación entre las dos variables.

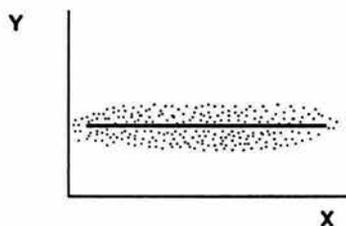
**Ejemplo 1.** La figura incluye varios diagramas de dispersión y sus líneas de regresión asociadas en demostración de varios tipos de relaciones entre las variables.



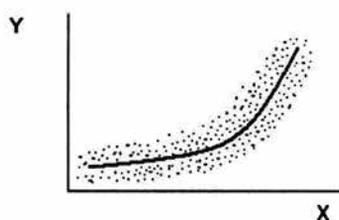
a) Relación lineal directa



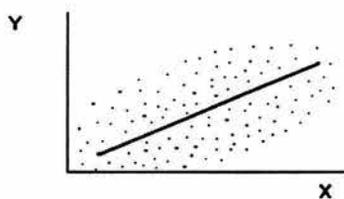
b) Relación lineal inversa



c) Sin relación



d) Relación curvilínea directa



e) Relación lineal directa con menor grado de relación que en a)



f) Relación lineal directa con mayor grado de relación que en a)

### 3.2.2 Estimación de los parámetros por mínimos cuadrados.

Los parámetros  $\beta_0$  y  $\beta_1$  son desconocidos y deben estimarse usando datos de una muestra. Suponga que se tiene  $n$  pares de datos,  $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ . Estos datos de una muestra pueden resultar o bien un diseño experimental controlado específicamente para coleccionar los datos o bien de registros históricos existentes.

#### 3.2.2.1 Estimación de $\beta_0$ y $\beta_1$ .

El método de mínimos cuadrados se utiliza para estimar  $\beta_0$  y  $\beta_1$ . Es decir, se estimarán estos parámetros, tales que la diferencia de la suma de cuadrados entre las observaciones  $y_i$  y la línea recta sea mínima. De la ecuación (3.1) se puede escribir lo siguiente:

$$y = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3.5)$$

El criterio de mínimos cuadrados establece minimizar:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (3.6)$$

Los estimadores mínimos cuadrados de  $\beta_0$  y  $\beta_1$ , es decir  $\hat{\beta}_0$  y  $\hat{\beta}_1$  deben satisfacer

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (3.7)$$

$$\left. \frac{\partial S}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Al simplificar estas dos ecuaciones se obtiene que:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (3.8)$$

$$\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i$$

Las ecuaciones (3.8) son llamadas ecuaciones normales de mínimos cuadrados. La solución a las ecuaciones normales es:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3.9)$$

y

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \quad (3.10)$$

donde

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (3.11)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

son los promedios de  $y_i$  y de  $x_i$ , respectivamente. Por lo tanto,  $\beta_0$  y  $\beta_1$  en (3.9) y (3.10) son los estimadores mínimos cuadrados de la ordenada del origen y la pendiente. El modelo de regresión lineal simple ajustado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (3.12)$$

La ecuación (3.12) da una estimación puntual de la media de  $Y$  para un valor particular de  $X$ . El denominador de (3.10) es la suma de cuadrados de  $x$ , corregida, y el numerador es la suma de los productos cruzados de  $x$ , y  $y$ , también corregida; así, se pueden escribir estas cantidades en una notación más compacta como

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.13)$$

y

$$S_{xy} = \sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n} = \sum_{i=1}^n y_i (x_i - \bar{x}) \quad (3.14)$$

Con lo que se obtiene una manera más conveniente de escribir (3.10), que es:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (3.15)$$

La diferencia entre el valor observado y el valor estimado correspondiente es un residuo:

$$e_i = y_i - \hat{y} = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \quad i = 1, 2, \dots, n \quad (3.16)$$

Después de obtener los estimadores de mínimos cuadrados, un número de preguntas interesantes deben considerarse:

1. ¿Qué tan bien, esta ecuación estima a los datos?
2. ¿Es probable que el modelo sea útil como un estimador?
3. ¿Qué tanto se cumplen los supuestos básicos, tales como varianza constante y errores no correlacionados?

Todos estos puntos deben investigarse antes de que el modelo sea ajustado finalmente. Como se ha mencionado anteriormente, los residuos juegan un papel importante en la evaluación del ajuste del modelo. Los residuos pueden verse como

realizaciones de los errores del modelo 4. Así para verificar los supuestos de la varianza constante y los errores no correlacionados nos debemos preguntar si los residuos se presentan de una muestra aleatoria y provienen de una distribución con estas propiedades.

### 3.2.3 Propiedades de los estimadores mínimos cuadrados y el modelo de regresión ajustado.

Los estimadores mínimos cuadrados de  $\beta_0$  y  $\beta_1$  tienen diferentes propiedades estadísticas importantes. Primero, notar que de las ecuaciones (3.9) y (3.10) los  $\hat{\beta}_0$  y los  $\hat{\beta}_1$  son las combinaciones lineales de las observaciones  $y_i$ . Por ejemplo,

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^n c_i y_i \quad (3.17)$$

donde  $c_i = (x_i - \bar{x})/S_{xx}$  para  $i = 1, 2, \dots, n$ .

Se tiene lo siguiente para  $\hat{\beta}_1$

$$\begin{aligned} E(\hat{\beta}_1) &= E\left(\sum_{i=1}^n c_i y_i\right) \\ &= \sum_{i=1}^n c_i E(y_i) \\ &= \sum_{i=1}^n c_i (\beta_0 + \beta_1 x_i) \\ &= \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i x_i \end{aligned} \quad (3.18)$$

además se puede mostrar directamente que

$$\sum_{i=1}^n c_i = 0 \quad \text{y} \quad \sum_{i=1}^n c_i x_i = 1 \quad (3.19)$$

Por lo tanto

$$E(\hat{\beta}_1) = \beta_1 \quad (3.20)$$

es decir, la  $E(y_i) = \beta_0 + \beta_1 X_i$ ,  $\hat{\beta}_1$ , es un estimador insesgado de  $\beta_1$ . Similarmente podemos mostrar que  $\hat{\beta}_0$  es un estimador insesgado de  $\beta_0$ ,

De la expresión (3.9) el estimador  $\hat{\beta}_0$  puede ser expresado como una función lineal de  $y_i$ , es decir,

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_0 &= \bar{y} - \sum_{i=1}^n c_i y_i \bar{x} \\ \hat{\beta}_0 &= \frac{\sum_{i=1}^n y_i}{n} - \sum_{i=1}^n c_i \bar{x} y_i \\ \hat{\beta}_0 &= \left( \sum_{i=1}^n \frac{1}{n} - \sum_{i=1}^n c_i \bar{x} \right) y_i \\ \hat{\beta}_0 &= \sum_{i=1}^n \left( \frac{1}{n} - \sum_{i=1}^n c_i \bar{x} \right) y_i \\ \hat{\beta}_0 &= \sum_{i=1}^n \left( \frac{1}{n} - \bar{x} c_i \right) y_i \end{aligned} \quad (3.21)$$

Se demostrará que  $E(\hat{\beta}_0) = \beta_0$

$$\text{Sea } \hat{\beta}_0 = \sum_{i=1}^n \left( \frac{1}{n} - \bar{x} c_i \right) y_i$$

$$\hat{\beta}_0 = \sum_{i=1}^n \left( \frac{1}{n} - \bar{x} c_i \right) (\beta_0 + \beta_1 x_i + \varepsilon)$$

$$\hat{\beta}_0 = \beta_0 - \beta_0 \bar{x} \sum_{i=1}^n c_i + \beta_1 \bar{x} - \beta_1 \bar{x} \sum_{i=1}^n c_i x_i + \sum_{i=1}^n \left( \frac{1}{n} - \bar{x} c_i \right) \varepsilon$$

$$\hat{\beta}_0 = \beta_0 + \beta_0 \bar{x} - \beta_0 \bar{x} + \sum_{i=1}^n \left( \frac{1}{n} - \bar{x}c_i \right) \varepsilon \quad \text{usando (3.19)}$$

$$\hat{\beta}_0 = \beta_0 + \sum_{i=1}^n \left( \frac{1}{n} - \bar{x}c_i \right) \varepsilon$$

$$E(\hat{\beta}_0) = E \left( \beta_0 + \sum_{i=1}^n \left( \frac{1}{n} - \bar{x}c_i \right) \varepsilon \right)$$

$$E(\hat{\beta}_0) = \beta_0 + \sum_{i=1}^n \left( \frac{1}{n} - \bar{x}c_i \right) E(\varepsilon)$$

$$E(\hat{\beta}_0) = \beta_0 \quad (3.22)$$

De las expresiones (3.15) y (3.21) tenemos que:

$$\hat{\beta}_1 = \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) y_i$$

$$\hat{\beta}_0 = \sum_{i=1}^n \left( \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) y_i$$

y se obtiene:

$$Var(\hat{\beta}_1) = \sum_{i=1}^n \left( \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 Var(y_i)$$

$$Var(\hat{\beta}_1) = \sigma^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.23)$$

$$\begin{aligned}
 \text{Var}(\hat{\beta}_0) &= \sum_{i=1}^n \left( \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 \text{Var}(y_i) \\
 \text{Var}(\hat{\beta}_0) &= \sigma^2 \sum_{i=1}^n \left( \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 \\
 \text{Var}(\hat{\beta}_0) &= \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \tag{3.24}
 \end{aligned}$$

Frecuentemente se dice que los estimadores mínimos cuadrados son los mejores estimadores insesgados lineales, donde la palabra "mejor" implica la varianza mínima. Hay otras propiedades útiles de los estimadores mínimos cuadrados:

1. La suma de los residuos en cualquier modelo de regresión que contiene un intercepto  $\hat{\beta}_0$  es siempre cero; es decir,

$$\sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0 \tag{3.25}$$

2. La suma de los valores observados  $y_i$  es igual a la suma de los valores estimados  $\hat{y}_i$ , o bien

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i \tag{3.26}$$

3. La línea de regresión de los mínimos cuadrados siempre pasa por el centro de los datos (punto  $(\bar{y}, \bar{x})$ ).

4. La suma de los residuos ponderados para el valor correspondiente de la variable regresora es siempre igual a cero; es decir,

$$\sum_{i=1}^n x_i e_i = 0 \quad (3.27)$$

5. La suma de los residuos ponderados para el valor estimado correspondiente de la variable regresora es siempre igual a cero; es decir,

$$\sum_{i=1}^n \hat{y}_i e_i = 0 \quad (3.28)$$

### 3.2.4 Estimación de $\sigma^2$ .

Además para estimar  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , se requiere un estimador de  $\sigma^2$  para probar hipótesis y construir un intervalo estimado pertinente al modelo de regresión.

Lo ideal sería que este estimador no dependiera del ajuste del modelo estimado. Esto sólo es posible cuando existen varias observaciones en  $Y$  y por lo menos un valor en  $X$  o cuando en la información anterior,  $\sigma^2$  es conocida. Cuando esta aproximación no puede ser utilizada, la estimación de  $\sigma^2$  se obtiene de la suma de cuadrados del error,

$$\begin{aligned} SS_E &= \sum_{i=1}^n e_i^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned} \quad (3.29)$$

Una fórmula conveniente para calcular  $SS_E$  puede ser encontrada por la sustitución de  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$  y simplificación en (3.29), como se muestra a continuación

$$SS_E = \sum_{i=1}^n y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy} \quad (3.30)$$

pero

$$\sum_{i=1}^n y_i^2 - n\bar{y}^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \equiv S_{yy} \quad (3.31)$$

es sólo la suma de cuadrados corregida de las observaciones, así pues

$$SS_E = S_{yy} - \hat{\beta}_1 S_{xy} \quad (3.32)$$

La suma de cuadrados del error tiene  $n-2$  grados de libertad, debido a que los dos grados de libertad están asociados con la estimación de  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , involucrados en la obtención de  $\hat{y}_i$ . Ahora, el valor esperado de  $SS_E$  es  $E(SS_E) = (n-2)\sigma^2$ , así entonces un estimador insesgado de  $\sigma^2$  es

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} = MS_E \quad (3.33)$$

La cantidad  $MS_E$  es llamada la media del cuadrado del error o media del cuadrado del residuo. La raíz cuadrada de  $\hat{\sigma}^2$  es llamada algunas veces como el error estándar de regresión, y tiene las mismas unidades como la variable de respuesta  $Y$ . Porque  $\hat{\sigma}^2$  depende de la suma de cuadrados del error y cualquier violación de los supuestos del modelo puede dañar seriamente la utilidad de  $\hat{\sigma}^2$  como un estimador de  $\sigma^2$ .

### 3.2.5 El Teorema de Gauss – Markov.

Un resultado importante relacionado con la calidad de los estimadores mínimos cuadrados de  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , es el teorema de Gauss-Markov, el cual afirma que el modelo de regresión (3.1) con los supuestos  $E(\varepsilon) = 0$ ,  $V(\varepsilon) = \sigma^2$ , y los errores no correlacionados, los estimadores mínimos cuadrados son estimados y tienen una varianza mínima cuando son comparados con otros estimadores insesgados que son combinaciones lineales de los  $y_i$ .

### 3.3 ESTIMACION DE INTERVALOS EN REGRESION LINEAL SIMPLE.

En esta sección se considera la estimación por intervalo de confianza de los parámetros del modelo de regresión. También se discute la estimación por intervalo de la media de la variable respuesta  $E(Y)$  para valores dados de  $X$ .

#### 3.3.1 Intervalos de confianza para $\beta_0, \beta_1$ y $\sigma^2$ .

Además de la estimación puntual de  $\beta_0, \beta_1$  y  $\sigma^2$ , se pueden obtener estimaciones por intervalo de confianza de estos parámetros. El tamaño de estos intervalos es una medida de la calidad total de regresión lineal. Si los errores son distribuidos normal e independientemente, entonces la distribución normal de ambos es

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{MS_E / S_{XX}}} \quad \text{y} \quad \frac{\hat{\beta}_0 - \beta_0}{\sqrt{MS_E \left(1/n + \bar{x}^2 / S_{xx}\right)}} \quad (3.34)$$

con  $n-2$  grados de libertad. Así, un intervalo de confianza porcentual  $100(1-\alpha)$  para  $\beta_1$  está dada por

$$\hat{\beta}_1 - t_{\alpha/2, n-2} \sqrt{MS_E \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} \sqrt{MS_E \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \quad (3.35)$$

y para  $\beta_0$  su intervalo de confianza porcentual  $100(1-\alpha)$  correspondiente es el siguiente

$$\hat{\beta}_0 - t_{\alpha/2, n-2} \sqrt{MS_E \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2, n-2} \sqrt{MS_E \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \quad (3.36)$$

Estos intervalos de confianza tienen la interpretación usual de frecuencia. Es decir, si se toman muestras repetidas del mismo tamaño en los mismos niveles de  $X$

y se construye por ejemplo, 95 intervalos de confianza en la pendiente para cada muestra, entonces 95 por ciento de esos intervalos contendrán el verdadero valor de  $\beta_1$ .

La cantidad

$$se(\hat{\beta}_1) = \sqrt{\frac{MS_E}{S_{xx}}} \quad (3.37)$$

en (3.35) se llama error estándar de la pendiente  $\hat{\beta}_1$ . Es una medida de cómo se ha estimado la pendiente. Similarmente

$$se(\hat{\beta}_0) = \sqrt{MS_E \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \quad (3.38)$$

en (3.36) es el error estándar del intercepto  $\hat{\beta}_0$ . Los programas de regresión por computadora comúnmente informan los errores estándares de los coeficientes de regresión. Si los errores son distribuidos normal e independientemente, la distribución muestral de

$$(n-2)MS_E / \sigma^2$$

es Chi-cuadrada con n-2 grados de libertad. Así entonces

$$P \left\{ x^2_{1-\alpha/2, n-2} \leq \frac{(n-2)MS_E}{\sigma^2} \leq x^2_{\alpha/2, n-2} \right\} = 1 - \alpha$$

y consecuentemente un  $100(1-\alpha)$  por ciento del intervalo de confianza para  $\sigma^2$  es

$$\frac{(n-2)MS_E}{x^2_{\alpha/2, n-2}} \leq \sigma^2 \leq \frac{(n-2)MS_E}{x^2_{1-\alpha/2, n-2}} \quad (3.39)$$

### 3.3.2 Tabla de análisis de varianza para el modelo de regresión lineal simple.

Se sabe que con el modelo de regresión lineal lo que se busca es explicar la variabilidad que existe en  $Y$  en base a  $X$ , por lo que una vez estimados los parámetros de modelo de regresión  $\beta_0$  y  $\beta_1$ , lo que nos interesa conocer ahora, es que magnitud de la varibilidad total ha sido explicado por el modelo, o bien que tan

bueno es el modelo para explicar la variabilidad existente en  $Y$ . En primera instancia podemos plantear un juego de hipótesis con el fin de juzgar si el modelo planteado es adecuado para explicar la variabilidad en  $Y$  o no. Dicho juego de hipótesis se plantea de la siguiente manera:

$$H_0 : \beta_1 = 0 \text{ (El modelo no sirve)} \quad \text{vs} \quad H_a : \beta_1 \neq 0$$

En seguida, se averiguar si se acepta o rechaza la hipótesis nula mediante un Análisis de Varianza. El análisis de varianza se realiza mediante la tabla del Análisis de Varianza (ANOVA) la cual se obtiene de la suma de cuadrados de los errores (residuales) a través de la siguiente igualdad:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \bar{y})^2 - \beta_1 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (3.40)$$

Suma de cuadrados = Suma de cuadrados del - Suma de cuadrados de la  
 regresión  
 del error total corregido

Empleando la igualdad (3.40) se construye la tabla de ANOVA obteniéndose lo siguiente.

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F calculada
Regresión	$P-1$	$\beta_1 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$	$\frac{SCR}{P-1}$	$\frac{CMR}{CME}$
Error	$n-P$	$SCT_c - SCE$	$\frac{SCE}{n-P}$	
Total Corregido	$n-1$	$\sum_{i=1}^n (y_i - \bar{y})^2$		

De la tabla de ANOVA podemos comentar lo siguiente:

1. Los grados de libertad que aparecen en la tabla tienen que ver con el número de observaciones independientes en los cuales las sumas de cuadrados se basa. Así, la Suma de Cuadrados del Total corregido ( $SCT_c$ ) tiene  $n-1$  grados de libertad ( $gl$ ) porque se pierde un  $gl$  en el cálculo de la media muestral ( $\bar{Y}$ ). La suma de cuadrados de la regresión ( $SCR$ ) tiene  $p-1$   $gl$  porque es el resultado de restar los  $gl$  del error de los  $gl$  del total corregido. Finalmente la suma de cuadrados del error ( $SCE$ ) tiene  $n-p$  grados de libertad porque los  $gl$  se definen como el número de observaciones menos los parámetros estimados.
2. Los cuadrados medios se obtienen al dividir las Sumas de Cuadrados de cada fuente de variación entre sus respectivos grados de libertad. El cuadrado medio del error ( $CME$ ) funge como un estimador insesgado de la varianza poblacional.

Para concluir si se rechaza o se acepta la hipótesis nula se emplea la siguiente regla de decisión.

*Regla de decisión:* Rechazar la hipótesis nula si  $F_c > F_{tab}$  donde  $F_{tab}$  se obtiene de una tabla de  $F$  con  $p-1$  grados de libertad en el numerador y  $n-p$  en el denominador, y algún  $\alpha$  para la confiabilidad.

Otra posible manera para deducir la aceptación o rechazo de la hipótesis nula podría ser observando directamente el valor obtenido en la  $F_c$ ; esto es, si dicho valor resultara ser mayor que 1, se estaría rechazando la hipótesis nula pues ello indicaría que se está explicando más del 50% de la variabilidad en  $Y$  con nuestro modelo.

### 3.3.3 Pruebas individuales o pruebas de $t$ .

Mediante la tabla del análisis de varianza (ANOVA) se realizan pruebas de hipótesis que involucran a todo el modelo, es decir, la prueba únicamente sirve para determinar si el modelo planteado es válido o no para explicar la variabilidad existente en  $Y$ . Sin embargo en muchas ocasiones la prueba de ANOVA indica que el modelo como tal es válido para explicar a  $Y$  pero en realidad sólo una de las variables independientes que componen el modelo (para un modelo de regresión lineal múltiple) es la que explica en gran medida o de manera significativa a dicha variable. Para detectar esto es necesario realizar pruebas individuales sobre los parámetros que sirven como coeficientes de las variables independientes, estas pruebas son conocidas como: *pruebas de significancia para los coeficientes de regresión o prueba de  $t$* .

En términos generales, una prueba de significancia es un procedimiento mediante el cual se utilizan los resultados de la muestra para verificar la veracidad o falsedad de una hipótesis. La idea consiste en utilizar un estadístico de prueba (estimador) y la distribución muestral de dicho estadístico bajo la hipótesis nula.

Bajo el supuesto de normalidad el estadístico  $t = \frac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)}$  se distribuye como

una  $t$  con  $n-2$  grados de libertad.

Aquí:

$se(\hat{\beta}_i)$  = el error estándar del estimador del parámetro  $\beta_i$ .

Para el caso del modelo de regresión lineal simple,  $var(\hat{\beta}_0) = \frac{S^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$  y

$var(\hat{\beta}_1) = \frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$  por lo que el error estándar de cada uno de estos parámetros

es simplemente la raíz cuadrada de las dos últimas cantidades.

Así, para realizar una prueba de significancia para el parámetro  $\beta_1$  que es el que concierne al caso simple, se plantean un juego de hipótesis de la siguiente manera:

$$H_0: \beta_1 = 0 \quad \text{vs} \quad H_a: \beta_1 \neq 0$$

La regla de decisión será: Rechazar la hipótesis nula cuando el estadístico  $t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$ , evaluada bajo la hipótesis nula, sea mayor al valor tabular  $t$  con  $n-2$  gl.

Nótese que si no rechaza la hipótesis nula, implica que la variable  $X$  no ayuda a explicar la variabilidad existente en  $Y$ .

### 3.3.4 Coeficiente de determinación.

La cantidad

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \quad (3.41)$$

es llamado el coeficiente de determinación. El  $S_{yy}$  es una medida de la variabilidad en  $Y$  sin considerar el efecto de la variable regresora  $X$  y  $SS_E$  es una medida de la variabilidad en  $Y$  que pertenece a un  $X$  que ha sido considerado, es llamado frecuentemente como la proporción de la variación explicada por la variable regresora  $X$ . Debido a que  $0 \leq SS_E \leq S_{yy}$ ,  $R^2$  se encuentra dentro del intervalo 0 y 1, es decir  $0 \leq R^2 \leq 1$ . Los valores de  $R^2$  que se aproximan a 1 implican que la mayoría de la variabilidad es explicada por el modelo de regresión.

El estadístico  $R^2$  debería usarse con cuidado, siempre es posible hacer  $R^2$  grande agregando suficientes términos al modelo. Por ejemplo, si no hay puntos repetidos indica (más de un  $Y$  en el mismo valor  $X$ ), un polinomio de  $n-1$  grados y dará un ajuste "perfecto" ( $R^2 = 1$ ) para  $n$  datos.

Cuando hay puntos repetidos,  $R^2$  nunca puede ser exactamente 1 porque el modelo no puede explicarse la variabilidad verdadera al error "puro".

Aunque  $R^2$  incremente, si se agrega una variable regresora al modelo, esto no significa necesariamente que el modelo nuevo sea superior al anterior. A menos que la suma de cuadrados del error en el modelo nuevo sea reducida por una cantidad igual a la media del cuadrado del error original, el modelo nuevo tendrá una media del cuadrado del error más grande que el anterior debido a la pérdida de un grado de libertad para el error. Así el modelo nuevo realmente será mejor que el anterior. La magnitud de  $R^2$  también depende del rango de variabilidad en la variable regresora. Hahn (1973) observa que el valor esperado de  $R^2$  de una línea recta de regresión es aproximadamente

$$E(R^2) = \frac{\hat{\beta}_1^2 S_{xx}}{\hat{\beta}^2 S_{xx} + \sigma^2} \quad (3.42)$$

Claramente el valor esperado de  $R^2$  incrementará (o disminuirá) como  $S_{xx}$  (una medida de dispersión de las  $X$ ) aumente (o disminuya). Así un valor grande de  $R^2$  puede resultar simplemente porque  $X$  ha variado sobre un rango grande.

Existen otros conceptos sobre  $R^2$ , en general,  $R^2$  no mide la magnitud de la pendiente de la línea de regresión, un valor grande de  $R^2$  no implica una pendiente. Además  $R^2$  no mide el ajuste del modelo lineal,  $R^2$  frecuentemente será grande si  $Y$  y  $X$  están relacionados no linealmente. Recuerde que aunque  $R^2$  sea grande, esto no implica necesariamente que el modelo de regresión sea un buen predictor.

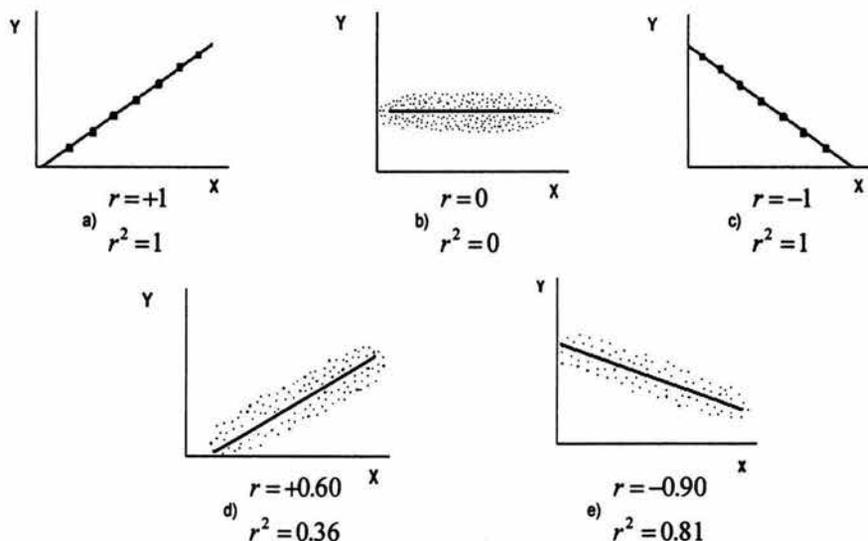
### 3.3.5 Coeficiente de correlación.

Aunque el coeficiente de determinación  $R^2$  es relativamente fácil de interpretar, no se presta para realizar una prueba estadística. En cambio, la raíz cuadrada del coeficiente de determinación, llamada coeficiente de correlación  $R$ , sí se presta para realizar una prueba estadística, porque puede servir para definir una

estadística de prueba con distribución  $t$  cuando la correlación poblacional  $\rho$  equivale a 0. El valor del coeficiente de correlación puede ir de -1.00 a +1.00. El signo aritmético asociado con el coeficiente de correlación, siempre igual al asociado con  $\beta_1$  en la ecuación de regresión, indica la dirección de la relación entre  $X$  y  $Y$  (positivo = directa; negativo = inversa).

En síntesis, el signo del coeficiente de correlación indica la dirección entre las variables  $X$  y  $Y$ , mientras que el valor absoluto del coeficiente indica el grado de relación. El valor cuadrado del coeficiente de correlación es el coeficiente de determinación, el cual indica la proporción de la varianza de  $Y$  explicada por el conocimiento de  $X$  (y viceversa).

**Ejemplo 2.** En la figura se ilustra la apariencia general de los diagramas de dispersión asociados con diversos valores de correlación.



Para el análisis de regresión se asume que  $X$  es una variable de medida controlable con un error insignificante y que  $Y$  es una variable aleatoria. Muchas de las aplicaciones del análisis de regresión involucran situaciones donde ambas variables son aleatorias y los niveles de  $X$  no pueden controlarse. En estas situaciones comúnmente se asume que las observaciones  $(y_i, x_i)$ ,  $i = 1, 2, \dots, n$  son variables aleatorias conjuntamente distribuidas. Por ejemplo, supóngase que se desea desarrollar un modelo de regresión que relacione ventas de bebidas con la temperatura máxima diaria. Claramente no se puede controlar la temperatura máxima diaria. Se seleccionarían aleatoriamente  $n$  días y se observaría una temperatura máxima  $(x_i)$  y un nivel de ventas  $(y_i)$  para cada día. Así  $(y_i, x_i)$  son variables conjuntamente distribuidas. En tales modelos las interferencias que resultan son condicionales de los  $X$ 's observados.

Se asume que la distribución conjunta de  $Y$  y  $X$  es la distribución normal bivariada. Esto es,

$$f(y, x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp\left\{-\frac{1}{2(1-p^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x-\mu_2}{\sigma_2}\right)^2 - 2p\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)\right]\right\} \quad (3.43)$$

donde  $\mu_1$  y  $\sigma_1^2$  son la media y la varianza de  $Y$ ,  $\mu_2$  y  $\sigma_2^2$  son la media y la varianza de  $X$ , y

$$f(y/x) = \frac{1}{\sqrt{2\pi}\sigma_{12}} \exp\left[-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1x}{\sigma_{12}}\right)^2\right] \quad (3.44)$$

es el coeficiente de correlación entre  $Y$  y  $X$ . El término  $\sigma_{12}$  es la covarianza de  $Y$  y  $X$ .

La distribución condicional de  $Y$  para un valor dado de  $X$  es

$$f(y/x) = \frac{1}{\sqrt{2\pi}\sigma_{12}} \exp\left[-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1x}{\sigma_{12}}\right)^2\right] \quad (3.45)$$

donde

$$\beta_0 = \mu_1 - \mu_2 \rho \frac{\sigma_1}{\sigma_2} \quad (3.46a)$$

$$\beta_1 = \frac{\sigma_1}{\sigma_2} \rho \quad (3.46b)$$

y

$$\sigma_{12}^2 = \sigma_1^2 (1 - \rho^2) \quad (3.46c)$$

Esto es, la distribución condicional de  $Y$  dado  $X$  es normal con media

$$E(y/x) = \beta_0 + \beta_1 x \quad (3.47)$$

y varianza  $\sigma_{12}^2$ . Note que la media de la distribución condicional de  $Y$  dado  $X$ , es un modelo de regresión lineal, además existe una relación entre el coeficiente de correlación  $\rho$  y la pendiente  $\beta_1$ . De (3.46b) vemos que si  $\rho = 0$ , entonces  $\beta_1 = 0$ , lo cual implica que hay regresión no lineal de  $Y$  sobre  $X$ . Esto es, el conocimiento de  $X$  no nos ayuda a predecir  $Y$ .

Es posible hacer inferencias sobre el coeficiente de correlación  $\rho$  en este modelo. El estimador de  $\rho$  es el coeficiente de correlación de muestreo

$$r = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}} = \frac{S_{xy}}{[S_{xx} S_{yy}]^{1/2}} \quad (3.48)$$

note que

$$\hat{\beta}_1 = \left( \frac{S_{yy}}{S_{xx}} \right)^{1/2} r \quad (3.49)$$

así la pendiente  $\hat{\beta}_1$  es simplemente el coeficiente de correlación muestral  $r$  multiplicado por un factor escalar que es la raíz cuadrada de un "spread" de las  $Y$ 's dividido por el "spread" de las  $X$ 's. Así  $\hat{\beta}_1$  y  $r$  se relacionan estrechamente, aunque ellos proporcionan información diferente. El coeficiente de correlación de muestreo  $r$  es una medida de asociación entre  $X$  y  $Y$ , mientras que  $\hat{\beta}_1$  mide el cambio estimado para una unidad de cambio en  $X$ . En el caso de una variable controlable  $X$ ,  $r$  no tiene ningún significado porque la magnitud de  $r$  depende de los valores de  $X$ . También se puede escribir como (3.49),

$$r^2 = \hat{\beta}_1^2 \frac{S_{xx}}{S_{yy}} = \frac{\hat{\beta}_1 S_{xy}}{S_{yy}} = \frac{SS_R}{S_{yy}} = R^2 \quad (3.50)$$

La cual se había mencionado en un principio como el coeficiente de determinación. Es decir, el coeficiente de determinación  $R^2$  es simplemente el cuadrado del coeficiente de correlación entre  $X$  y  $Y$ .

Mientras que la regresión y la correlación se relacionan estrechamente, la regresión es una herramienta más poderosa en muchas situaciones. La correlación es simplemente una medida de asociación y es de poco uso para la predicción. Sin embargo, los métodos de regresión son útiles en el desarrollo cuantitativo de relaciones entre variables, los cuales pueden ser usados en predicciones.

### 3.4 EL MODELO DE REGRESIÓN LINEAL MÚLTIPLE.

El modelo general de regresión múltiple se presenta a continuación. Los resultados son esencialmente extensiones de los resultados dados por el modelo de regresión simple. La presentación sirve como una revisión de los resultados estándares.

### 3.4.1 El modelo y descripción de los datos.

Los datos consisten de  $n$  observaciones sobre una variable dependiente o de respuesta  $Y$  y  $p$  variables independientes (explicatorias)  $X_1, X_2, \dots, X_p$ . Las observaciones se representan comúnmente como se indica a continuación:

Número de Observación	$y$	$x_1$	$x_2$	$x_3$	...	$x_p$
1	$y_1$	$x_{11}$	$x_{21}$	$x_{31}$		$x_{p1}$
2	$y_2$	$x_{12}$	$x_{22}$	$x_{32}$		$x_{p2}$
3	$y_3$	$x_{13}$	$x_{23}$	$x_{33}$		$x_{p3}$
.	.					.
.	.					.
.	.					.
$n$	$y_n$	$x_{1n}$	$x_{2n}$	$x_{3n}$		$x_{pn}$

La relación entre  $Y$  y  $X_1, X_2, \dots, X_p$  se formula como un modelo lineal

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + e_i \quad (3.51)$$

donde  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  son constantes referidas a los coeficientes del modelo de regresión parcial (o simplemente como los coeficientes de regresión) y  $e_i$  es una distribución aleatoria. Se asume que para cualquier conjunto de valores fijos de  $X_1, X_2, \dots, X_p$  que caen dentro del rango de los datos, la ecuación lineal (3.51) satisface una aproximación aceptable a la relación verdadera entre  $Y$  y las  $X$ 's. En otras palabras,  $Y$  es aproximadamente una función lineal de  $X$  y  $e_i$  mide la distancia en esa aproximación para la observación  $Y$ . En particular  $e_i$  contiene información no sistemática para la determinante  $Y$  que no ha sido capturada en  $X$ .

Se asume que los valores de  $e_i$  son cantidades aleatorias, independientemente distribuidas con media cero o varianza constantes  $\sigma_2$ .

El coeficiente de regresión  $\beta_i$  puede interpretarse como el incremento en  $Y$  que corresponde a una unidad incrementada en  $X_i$  cuando todas las variables permanecen constantes. Claramente, esta interpretación se mantiene independiente de los valores de las  $X$ 's.

Los  $\beta$ 's son estimadores para minimizar la suma de cuadrados residuales, lo que se conoce como el método de mínimos cuadrados. A continuación se ve que:

$$S(\beta_0, \beta_1, \beta_2, \dots, \beta_p) = \sum_{i=1}^n e_i^2 \quad (3.52)$$

Aplicando los cálculos anteriores, puede mostrarse que los cuadrados mínimos estimados  $b_0, b_1, b_2, \dots, b_p$  los cuales minimizan a  $S(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ , son obtenidos por la solución del siguiente sistema de ecuaciones:

$$S_{11}b_1 + S_{12}b_2 + S_{13}b_3 + \dots + S_{1p}b_p = S_{y1}$$

$$S_{21}b_1 + S_{22}b_2 + S_{23}b_3 + \dots + S_{2p}b_p = S_{y2}$$

$$S_{p1}b_1 + S_{p2}b_2 + S_{p3}b_3 + \dots + S_{pp}b_p = S_{yp}$$

(3.53)

donde

$$S_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \quad i, j = 1, 2, \dots, p$$

$$S_{yi} = \sum_{k=1}^n (y_k - \bar{y})(x_{ik} - \bar{x}_i) \quad i, j = 1, 2, \dots, p$$

$$\bar{x}_i = \frac{\sum_{k=1}^n x_{ik}}{n}, \quad \bar{y} = \frac{\sum_{k=1}^n y_k}{n}$$

(3.54)

y

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - b_3 \bar{x}_3 - \dots - b_p \bar{x}_p \quad (3.55)$$

Las fórmulas que se encuentran arriba en el sistema de ecuaciones se llaman ecuaciones normales,  $b_0$  es referido comúnmente como el intercepto, y  $b_i$  como el estimador del coeficiente de regresión de  $X_i$ .

Se asume que el sistema de ecuaciones es soluble y tiene una solución única. No se dirá nada más sobre el proceso real de la solución de las ecuaciones normales. Se asumen que existe software de computadora que proporciona soluciones numéricamente precisas.

Usando los coeficientes estimados de regresión se define un valor estimado o predicho, en el renglón siguiente se ve representado este valor:

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + e_i \quad (3.56)$$

y un residuo observado, está dado por:

$$e_i = y_i - \hat{y}_i \quad (3.57)$$

para cada observación. Los  $e_i$ 's se usan como se mostró anteriormente para evaluar la exactitud del modelo.

### 3.4.2 Propiedades de los mínimos cuadrados.

1)  $b_i$  es una estimación insesgada de  $\beta_i$  que tiene una varianza de  $\sigma^2 c_{ii}$  donde  $c_{ii}$  es el  $i$ -ésimo elemento de la diagonal de la matriz inversa de productos de la suma corregida de cuadrados. La covarianza entre  $b_i$  y  $b_j$  es  $\sigma^2 c_{ij}$  donde  $c_{ij}$  es el elemento en la  $i$ -ésima fila,  $j$ -ésima columna de la matriz inversa de productos de la suma corregida de cuadrados. Para todos los estimadores insesgados que son lineales en las observaciones, la estimación de mínimos cuadrados tiene la varianza más pequeña.

2) Una estimación insesgada de  $\sigma^2$  está dada por:

$$s^2 = \frac{SSE}{n-p-1} \quad (3.58)$$

donde

$$SSE = \sum_{i=1}^n (y_i - b_0 - b_1 x_{1i} + b_2 x_{2i} + \dots + b_p x_{pi})^2 \quad (3.59)$$

Con el supuesto adicional de que los  $e_i$ 's se distribuyen normalmente.

3)  $b_i$  es normalmente distribuido con media  $\beta_i$  y varianza  $\sigma^2 c_{ii}$

4)  $W = SSE/\sigma^2$  tiene una distribución Chi-cuadrada  $\chi^2$  con  $(n-p-1)$  grados de libertad.

El conjunto de  $b_i$  y  $W$  son independientemente distribuidos para cada uno.

5) El vector  $b = (b_1, b_2, \dots, b_p)$  tiene una distribución normal bivariada con vector media  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  y matriz de varianzas y covarianzas  $\sigma^2(c_{ij})$ .

Los resultados presentados arriba permiten probar diversas hipótesis sobre  $\beta$  y construir intervalos de confianza.

El estadístico para probar  $H_0(\beta_i = \beta_i^0)$  donde  $\beta_i^0$  es una constante elegida por el investigador, es el siguiente.

$$t = \frac{b_i - \beta_i^0}{s \sqrt{c_{ii}}} \quad (3.60)$$

que tiene una distribución  $t$ -student con  $(n-p-1)$  grados de libertad.

La prueba es efectuada para comparar el valor observado con el valor crítico  $t$  apropiado. La prueba usual es para  $\beta_i^0 = 0$ , en la cual  $t$  reduce a la proporción de  $b_i$  a su error estándar. Los límites de confianza para  $\beta_i$  con un coeficiente de confianza  $\alpha$  están dados por  $b_i \pm t(n-p-1)$ ,  $(\alpha/2)s\sqrt{c_{ii}}$ , donde  $t(n-p-1, \alpha)$  es el  $(1-\alpha)$  por ciento de la distribución  $t$  con  $(n-p-1, \alpha)$  grados de libertad.

### 3.4.3 Valores estimados y el error estándar.

La ecuación de regresión múltiple ajustada puede usarse para predecir el valor de la variable dependiente correspondiente a una observación  $x_0 = (x_{10}, x_{20}, x_{30}, \dots, x_{p0})$ . El valor estimado  $\hat{y}_0$  está dado por

$$\hat{y}_0 = b_0 + b_1 x_{10} + b_2 x_{20} + \dots + b_p x_{p0} \quad (3.61)$$

y su varianza es

$$Var(\hat{y}_0) = \sigma^2 \left( 1 + \frac{1}{n} \right) + \sum_{i=1}^n (x_{i0} - \bar{x}_i)^2 V(b_i) + 2 \sum_i \sum_{i < j} (x_{i0} - \bar{x}_i)(x_{j0} - \bar{x}_j) Cov(b_i, b_j) \quad (3.62)$$

donde

$$\begin{aligned} V(b_i) &= \text{varianza de } b_i, \\ Cov(b_i, b_j) &= \text{Covarianza entre } b_i \text{ y } b_j \end{aligned} \quad (3.63)$$

Los límites de confianza para  $\hat{y}_0$  con un coeficiente de confianza  $\alpha$ , se denotan como sigue:

$$\hat{y}_0 \pm t(n-p-1, \alpha/2) s.e.(\hat{y}_0)$$

donde  $s.e.(\hat{y}_0)$  es el error estándar de  $\hat{y}_0$  y es obtenido tomando la raíz cuadrada de (3.62) después de sustituir  $\sigma^2$  por  $s^2$ , donde  $s^2$  es la varianza de los residuos y está dada por (3.58).

Como ya se ha mencionado, en relación con la regresión simple, en lugar de predecir la variable respuesta  $y$  correspondiente a una observación  $x_0 = (x_{10}, x_{20}, x_{30}, \dots, x_{p0})$ , se puede desear estimar la media de la respuesta correspondiente a esa observación. Se denotará la media de la respuesta en  $(x_{10}, x_{20}, x_{30}, \dots, x_{p0})$  por  $\mu_0$  y su estimador por  $\hat{\mu}_0$ . Entonces se tiene que:

$$\hat{\mu}_0 = b_0 + b_1 x_{10} + b_2 x_{20} + \dots + b_p x_{p0} \quad (3.64)$$

pero la varianza de la media de respuesta estimada para esta observación es:

$$Var(\hat{\mu}_0) = \frac{\sigma^2}{n} + \sum_{i=1}^n (x_{i0} - \bar{x}_i)^2 V(b_i) + 2 \sum_i \sum_{i < j} (x_{i0} - \bar{x}_i)(x_{j0} - \bar{x}_j) Cov(b_i, b_j) \quad (3.65)$$

donde  $V(b_i)$  y  $Cov(b_i, b_j)$  tienen el mismo significado de antes. El error estándar de la media de respuesta que corresponde a la observación  $(x_{10}, x_{20}, x_{30}, \dots, x_{p0})$  es obtenido sacando la raíz cuadrada de (3.65) y puede calcularse sustituyendo  $\sigma^2$  con su estimador  $s^2$ .

### 3.4.4 Coeficiente de correlación múltiple.

Después de ajustar el modelo lineal a los datos, una evaluación es realizar la exactitud del ajuste. La medida que se utiliza con más frecuencia, es el coeficiente de correlación múltiple  $R$ , o más frecuentemente el cuadrado del coeficiente múltiple de correlación  $R^2$ . Hay varias maneras equivalentes en que  $R^2$  pueden definirse y ser interpretado. Se define a  $R^2$  como

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (3.66)$$

e interpretarlo como la proporción de variabilidad total, que es explicada por la ecuación de regresión.  $R^2$  se encuentra entre el intervalo cero y uno. Cuando el modelo se ajusta bien a los datos, es claro ver que el valor de  $R^2$  se aproxima a la unidad. Con un ajuste bueno, los valores observados y predichos estará cerca uno del otro, y  $\sum (y_i - \hat{y}_i)^2$  será mínima. Entonces  $R^2$  estará cerca de la unidad. Por otra parte, si no hay relación lineal entre la variable independiente y las variables dependientes, el modelo lineal da un ajuste pobre, y el mejor predicho para una observación  $y_i$ , será  $\bar{y}$ ; es decir, en la ausencia de cualquier relación, la mejor estimación es la media muestral, para este caso la media muestral minimiza la desviación de la suma de cuadrados. Así la ausencia de cualquier relación lineal entre  $Y$  y las  $X$ 's,  $R^2$  estará cerca de cero. Por lo tanto el valor de  $R^2$  se usa como

una medida resumida para ajustar el modelo lineal a un determinado grupo de datos. Como se mencionó anteriormente, un valor grande de  $R^2$  no significa necesariamente que el modelo se ajuste bien a los datos. Es necesario un análisis más detallado para asegurar que el modelo describa a los datos adecuadamente.

Una cantidad relativa a  $R^2$ , conocida como R-cuadrada ajustada ( $R_a^2$ ) es también usada para juzgar la bondad de ajuste.

$R_a^2$ ; y se define como

$$R_a^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n - p - 1)}{\sum (y_i - \bar{y})^2 / (n - 1)} \quad (3.67)$$

$$= 1 - \frac{s^2}{s_y^2}$$

donde  $s_y^2 = \sum (y_i - \bar{y})^2 / (n - 1)$  y  $s^2$  como se definió en (3.58). De las ecuaciones (3.66) y (3.67) se ve que:

$$R_a^2 = 1 - \frac{n - 1}{n - p - 1} (1 - R^2) \quad (3.68)$$

$R_a^2$  se utiliza algunas veces para comparar modelos que tienen diferente número de variables explicatorias. Para comparar la bondad de ajuste de modelos con diferentes variables explicatorias,  $R_a^2$  trata de ajustar las diferentes variables en los modelos diferentes. A diferencia de  $R^2$ ,  $R_a^2$  no puede ser interpretado como la proporción de la variabilidad total de  $Y$  explicado por la ecuación de regresión.

La mayoría de los paquetes computacionales de regresión proveen valores para  $R^2$ ,  $R_a^2$ .

### 3.4.5 Prueba de hipótesis en el modelo lineal.

Además de existir hipótesis para los  $\beta$ 's, existen también varias hipótesis que son consideradas en relación al análisis de los modelos lineales. Las hipótesis que son investigadas con más frecuencia son.

- 1) Todos los coeficientes de regresión son cero. Esto implica que no hay relación lineal entre la variable dependiente y el conjunto de variables independientes.
- 2) Un subconjunto de los coeficientes de regresión son cero.
- 3) Un subconjunto de los coeficientes de regresión son iguales a cada uno.

Las diferentes hipótesis sobre los coeficientes de regresión pueden ser probadas del mismo modo por un método unificado. Para describir las pruebas individuales, se mostrará el método general.

El modelo dado en (3.9) es referido como el modelo completo (FM). La hipótesis nula a ser probada específica valores para algunos coeficientes de regresión. Cuando estos valores se sustituyen en el modelo completo, el modelo resultante se llama: modelo reducido (RM).

El número de parámetros distintos a ser estimados en el modelo reducido es menor que el número de parámetros a ser estimados en el modelo completo. La prueba para cualquier hipótesis involucra una comparación de la bondad de ajuste que se obtiene cuando es usado el modelo completo con los datos, para la bondad de ajuste que resulta de usar el modelo reducido especificado por la hipótesis nula. Si el modelo reducido proporciona un buen ajuste como el modelo completo, entonces la hipótesis nula que define el modelo reducido (por especificar algunos valores de  $\beta_i$ ), no es rechazada.

Se sabe que  $\hat{y}_i$  y  $\hat{y}_i^*$  son los valores predichos o estimados de  $y_i$  para el modelo completo y el modelo reducido, respectivamente. La falta de ajuste en los

datos asociada con el modelo completo es  $\sum (y_i - \hat{y}_i)^2$ . Se denota por SSE (FM), la suma de cuadrados debido al error asociada con el modelo completo,

$$SSE(FM) = \sum (y_i - \hat{y}_i)^2 \quad (3.69)$$

La falta de ajuste en los datos asociada con el modelo reducido es  $\sum (y_i - \hat{y}_i^*)^2$ . Esta cantidad es denotada por SSE (RM), que es la suma de cuadrados debido al error asociada con el modelo reducido,

$$SSE(MR) = \sum (y_i - \hat{y}_i^*)^2 \quad (3.70)$$

En el modelo completo existen  $(p+1)$  parámetros  $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$  y supóngase que para el modelo reducido hay  $k$  parámetros distintos. Para ver si el modelo reducido se ajusta, es necesario comparar  $SSE(RM) - SSE(FM)$  con  $SSE(FM)$  y en lugar de comparar estas dos cantidades directamente, se tiene que:

$$\frac{SSE(RM) - SSE(FM)}{p+1-k} \quad \text{y} \quad \frac{SSE(FM)}{n-p-1} \quad (3.71)$$

Los divisores se introducen para compensar el número diferente de parámetros involucrados en los dos modelos, así como también para asegurar que la prueba estadística resultante tenga una distribución estadística estándar. La relación

$$F = \frac{[SSE(RM) - SSE(FM)](p+1-k)}{SSE(FM)(n-p-1)} \quad (3.72)$$

tiene distribución  $F$  con  $(p+1-k)$  y  $(n-p-1)$  grados de libertad. Si el valor de  $F$  observado es grande, en comparación al valor tabulado de  $F$  con  $(p+1-k)$  y  $(n-p-1)$  grados de libertad con un nivel de confianza alfa determinado, el resultado es significativo, es decir; el modelo reducido no es el adecuado y la hipótesis nula con valores de  $\beta$ 's en el modelo completo, es rechazada.

La teoría distribucional, intervalos de confianza, y las pruebas de hipótesis planteadas anteriormente son válidas únicamente si se mantienen las suposiciones estándares para los términos residuales del modelo. Se asumen que los términos son independientes y normalmente distribuidos con media cero y varianza constante. Cuando estos supuestos no se cumplen pueden ocurrir serios errores cuando se

realice una aplicación de los errores estándares. Como en el caso de la regresión simple, el análisis gráfico de los residuos observados puede usarse para la verificación de los supuestos. Hay más gráficas que pueden ser examinadas en problemas de regresión múltiple que en el caso de regresión simple.

El análisis gráfico de los residuos pueden indicar violaciones serias en uno o más de los supuestos estándares. Estas violaciones podrían invalidar los procedimientos formales de inferencia estadística descritos arriba. Una situación de mayor importancia que se debe mencionar, es que el análisis de residuos puede conducir sugerencias de estructura o indicar información en los datos que pueden ser extraños si los resúmenes estadísticos ya han sido evaluados. Estas sugerencias pueden conducir un mejor entendimiento y posiblemente un mejor proceso del modelo estudiado. La realización cuidadosa de un análisis gráfico de residuos puede resultar frecuentemente la parte más importante del análisis de regresión.

### 3.4.6 Tabla de análisis de varianza para un modelo con k variables.

Para el caso de un modelo de regresión lineal de k variables, la forma que adquiere la tabla del análisis de varianza es la siguiente:

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F calculada
Regresión	$P-1$	$\hat{\beta}' X'Y - n\bar{Y}^2$	$\frac{SCR}{P-1}$	$\frac{CMR}{CME}$
Error	$n-P$	$Y'Y - \hat{\beta}' X'Y$	$\frac{SCE}{n-P}$	
Total Corregido	$n-1$	$Y'Y - n\bar{Y}^2$		

Dicha tabla puede ser empleada, al igual que el caso anterior (modelo de regresión lineal simple), para realizar una prueba de hipótesis con el fin de verificar si el modelo es adecuado o no para explicar a  $Y$ .

En lo que respecta a las pruebas individuales o pruebas de  $t$ , el procedimiento a seguir es exactamente la misma que en el caso de un modelo de regresión lineal simple, sólo que para este caso obtenemos la varianza de cada una de los estimadores  $\beta_i$  a partir de su matriz de varianzas y covarianzas, la cual es calculada mediante la siguiente expresión:

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

Es evidente que para el cálculo de esta expresión se debe obtener una estimación de dicha matriz sustituyendo  $\sigma^2$  por la varianza estimada  $S^2$  (CME).

Nótese que los elementos diagonales de esta matriz representan las varianzas de los parámetros estimados, en tanto que los elementos fuera de ella nos representan estimaciones de las covarianzas entre tales parámetros.

Para el caso del coeficiente de determinación  $R^2$ , debemos aclarar que la expresión que se empleó para su cálculo en el modelo de regresión lineal simple, se mantiene para el caso de un modelo con  $k$  variables. Por otro lado, para realizar predicciones de un valor esperado, digamos  $y_0$ , dado un valor determinado para cada variable independiente se emplea la expresión:  $y_0 = x_0\hat{\beta}$ .

Donde:  $x_0$  = vector que contiene un valor determinado para cada variable independiente.

### 3.4.7 El proceso de modelación.

En general, la modelación estadística puede verse como un proceso iterativo que en su fase terminal permite obtener el modelo que mejor logra explicar el fenómeno bajo estudio.

Este proceso puede sufrir algunas modificaciones, que en su mayoría son mínimas, dependiendo del modelo que se desea ajustar. Para el caso específico de regresión lineal múltiple se resume este proceso; como primer paso dentro del proceso de modelación se encuentra, como es natural, un análisis exploratorio y descriptivo de los datos, el cual no ayuda a conocer el comportamiento de cada una de las variables y las relaciones existentes entre ellas.

En base a estos conocimientos es posible postular un modelo tentativo (fase dos) que sea acorde con los resultados obtenidos. En la fase 3, ajuste del modelo, debemos obtener una estimación de los parámetros presentes en el modelo.

La evaluación del modelo ajustado es la siguiente fase dentro del proceso, en el cual se debe contrastar la hipótesis de que las variables  $X$  contribuyen significativamente de manera conjunta para explicar  $Y$ , contra la alternativa de que no lo hacen. Con esta prueba de hipótesis se comprueba si el modelo se ajusta adecuadamente a los datos o simplemente si el modelo es adecuado. Otro indicador que debemos observar en este punto es el coeficiente de determinación  $R^2$  que valora la adecuación global del modelo ajustado.

Revisión y diagnóstico es uno de los principales puntos dentro del proceso de modelación. En esta fase debemos revisar los residuales (cumplimiento de supuestos), además de valorar el posible efecto adicional no considerado en el modelo para las gráficas de residuos contra cada una de las variables  $X$  por separado.

Nótese que en el proceso de revisión y diagnóstico puede llevar a replantear el modelo. En este punto, el proceso se hace más interactivo, agregando términos al modelo ajustado, ajustando, verificando residuos y violaciones graves de los supuestos, eliminando puntos aberrantes o fuera de rango, volviendo a ajustar, etc. El cada paso debe evaluarse la adecuación del modelo a través de la  $R^2$  ajustada y el p-value asociado a la prueba de hipótesis respectiva, así como observando los indicadores de la significancia para cada parámetro en el modelo. Estas fases, en las que se define un ciclo, terminan cuando el modelo que tenemos es el más "aceptable", en términos de la mayor  $R^2$  ajustada, el menor  $\sigma^2 = CM_E$ , sin problemas de multicolinealidad o influencia de observaciones y sin problemas graves de violación a los supuestos del modelo.

Un aspecto adicional que se debe evaluar en la regresión múltiple es el efecto de la multicolinealidad, que es más bien un problema de grado; esto quiere decir que la multicolinealidad está siempre presente, sin embargo su efecto empieza a ser nocivo sólo si afecta el ajuste del modelo de manera sensible. Algunas técnicas para detectar y resolver problemas de multicolinealidad se pueden ver en Goldstein y Dillon (1984).

Una vez que se ha obtenido un modelo aceptable en los términos establecidos viene una fase final que consiste en elegir el modelo más simple. Esto es así porque usualmente no todas las  $X$  en el modelo tienen una contribución significativa. Para tal efecto, existen varios métodos con los cuales podemos obtener el modelo más simple y sin perder información esencial. Los métodos más usuales son: eliminación hacia atrás (backward), selección hacia delante (forward) y regresión paso a paso (stepwise). Más detalles sobre estos métodos están disponibles en Montgomery y Peck (1992).

Al obtener el modelo más simple, es posible hacer una interpretación más sencilla de él y llegar a una conclusión sobre el fenómeno que está modelando (fase última).

### 3.4.8 Supuestos sobre las variables explicatorias.

Hay dos supuestos en lo que concierne a las variables explicatorias que aún no se han discutido. Estos supuestos son más sutiles que los supuestos de mínimos cuadrados estándares mencionado anteriormente, en lo que concierne a la forma de la ecuación y a las propiedades probabilísticas de los términos de residuos. Los supuestos son:

1. Las variables explicativas no son estocásticas, es decir, los valores de las  $X$ 's se fijan o se seleccionan por adelantado, y
2. Las  $X$ 's se miden sin el error.

Estos supuestos no pueden ser válidos, así que no juegan un papel importante en el análisis de regresión, ni influyen en la interpretación de los resultados de regresión.

El primer supuesto se cumple únicamente cuando el experimentador puede colocar los valores de  $X$  variables en los niveles predeterminados. Es claro ver, que bajo situaciones observacionales, éste supuesto no se cumple. Los resultados teóricos que se han presentado continuarán presentes, pero su interpretación tiene que ser modificada. Cuando las  $X$  no son un conjunto fijo de constantes, todas las inferencias son condicionadas sobre los datos observados. Se debe notar que este aspecto condicional de la inferencia es consistente con el método de análisis de datos ya presentado. El objetivo principal consiste en extraer la cantidad máxima de información de los datos disponibles.

El segundo supuesto, de que las variables independientes se miden sin el error, difícilmente se cumple. Los errores afectan a la varianza del residuo, al coeficiente de correlación múltiple y a la estimación individual de los coeficientes de regresión. La magnitud exacta de los efectos dependerá de varios factores, los más importantes son, la desviación estándar de los errores medios y la estructura de la correlación entre los errores. El efecto de los errores medios está considerado para

aumentar la varianza del residuo y reducir la magnitud del coeficiente observado de la correlación múltiple. Los efectos de los errores medios sobre coeficientes individuales de regresión son más difíciles de evaluar. La estimación del coeficiente de regresión para una variable, se afecta no solamente por sus errores propios de medida, sino también por errores medios de otras variables incluidas en la ecuación.

La corrección para errores medios sobre los coeficientes estimados de regresión, aún en el caso más simple donde todos los errores medios son no correlacionados, requiere de un conocimiento de relación entre las varianzas de errores medios para las diferentes variables y la varianza del error aleatorio. Estas cantidades son raramente conocidas (particularmente en las ciencias sociales, donde este problema es muy agudo), nunca se puede eliminar completamente el efecto de errores medios de los coeficientes estimados de regresión. Si los errores medios no son grandes comparados con los errores aleatorios, el efecto de errores medios es insignificantes. Para interpretar los coeficientes en tal análisis, este punto debería recordarse. Aunque hay algún problema en la estimación de los coeficientes de regresión cuando las variables están en el error, la ecuación de regresión puede ser usada aún para la predicción. Sin embargo, la presencia de errores en las  $X$ 's variables disminuye el ajuste de predicción. Para una discusión más extensiva de este problema, el lector debe consultar al Fuller (1987).

### **3.5 RIESGOS Y LIMITACIONES ASOCIADOS CON EL ANÁLISIS DE REGRESIÓN.**

#### **3.5.1 Riesgos en el uso de la regresión.**

En casi todas las aplicaciones de regresión, la ecuación de regresión es sólo una aproximación a la relación que existe entre las variables. Por ejemplo, en la Figura 1, se ha ilustrado una situación donde una relación relativamente compleja

entre la  $X$  y la  $Y$  puede ser bien aproximada por una ecuación de regresión lineal. A veces una función más aproximadamente compleja es necesaria.

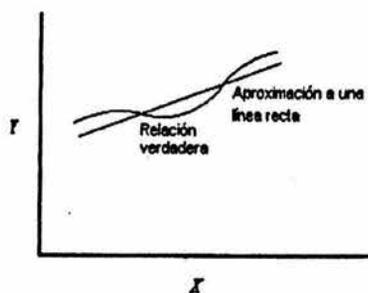


Figura 1. Aproximación de regresión lineal de una relación compleja.

Generalmente las ecuaciones de regresión son válidas sólo bajo la región o dominio de las variables regresoras contenidas en los datos observados. Por ejemplo, considere la Figura 2, y supóngase que los datos sobre el eje  $X$  y  $Y$  fueron colectados dentro del intervalo  $x_1 \leq x \leq x_2$ . Sobre este intervalo la ecuación de regresión lineal mostrada en la Figura 2 es una buena aproximación de la verdadera relación.

Sin embargo, supóngase que esta ecuación se usó para predecir valores de  $Y$  para valores de la variable regresora en la región  $x_2 \leq x \leq x_3$ . Claramente el modelo de regresión lineal es inútil para este rango  $X$ . Finalmente es importante recordar que el análisis de regresión es la parte de una aproximación de un enfoque analítico para la solución del problema. Es decir, la ecuación de regresión por sí misma no puede ser el objetivo primario del estudio.

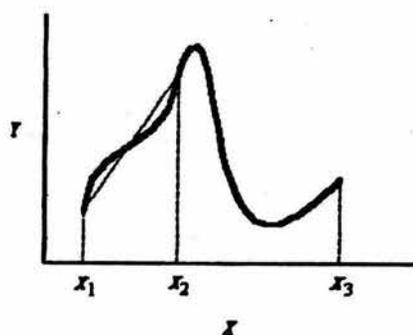


Figura 2. El riesgo de extrapolación en regresión.

Un aspecto esencial del análisis de regresión es la colección de datos. Dado que las conclusiones del análisis son condiciones sobre los datos; un buen esfuerzo de colección de datos puede tener muchos beneficios, incluyendo un análisis simplificado y un modelo generalmente más aplicable. Los datos usados en un análisis de regresión deben ser representativos del sistema estudiado. Sin datos representativos el modelo de regresión y las conclusiones obtenidas de éste, podrán tener un gran error. Por lo tanto deberá tenerse cuidado en la colección precisa de datos. Muchas de las técnicas usadas en el análisis de regresión pueden ser seriamente distorsionadas por datos registrados inadecuadamente. La edición preliminar de los datos para el análisis de regresión es conducido frecuentemente para la identificación de estos errores.

El análisis de regresión se usa ampliamente y, desafortunadamente, con frecuencia su empleo no es el adecuado. Hay varios abusos comunes de regresión que deben mencionarse.

1) Los modelos de regresión son proyectados como las ecuaciones de interpolación bajo el rango de la (s) variable (s) regresora (s) para ajustar el modelo. No pueden ser válidos para la extrapolación fuera de este rango. Ver figura 2.

2) La disposición de los valores de  $X$  juegan un papel importante en los cuadrados mínimos ajustados. Mientras que todos los puntos tienen un peso igual en la determinación de la altura de la línea, la pendiente es fuertemente influida por los valores remotos de  $X$ .

Por ejemplo, considere los datos de la figura 3. La pendiente de los mínimos cuadrados dependen fuertemente de cada uno de los puntos A y B o de ambos. Además los datos restantes darían una estimación muy diferente de la pendiente si A y B se borrarán. Tales situaciones como estas requieren de una acción correctiva, como el análisis adicional y eliminación posible de los puntos inusuales, estimación de los parámetros del modelo con alguna técnica que sea menos influida por estos puntos que los mínimos cuadrados, o reestructuración del modelo, posiblemente por introducción de regresores adicionales.

Una situación diferente se ilustra en el Figura 4, donde una de las 18 observaciones es muy remota en el espacio de las  $X$ . En este ejemplo la pendiente es determinada principalmente por el punto extremo. Si este punto es borrado, el estimador de la pendiente es probablemente  $\alpha$

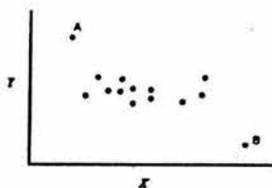


Figura 3. Dos observaciones influyentes.

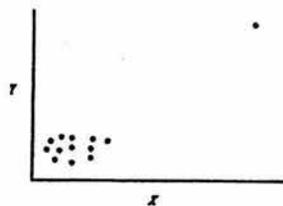


Figura 4. Un punto remoto en el espacio de  $X$ .

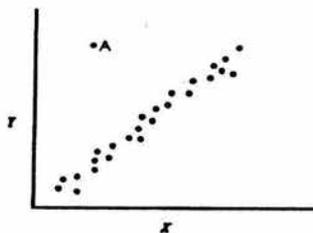


Figura 5. Un valor fuera de rango.

Debido a que en el intervalo existen dos grupos de puntos, se tiene realmente dos únicas unidades distintas de información con las cuales se ajusta el modelo. Así efectivamente hay menos distancia entre los 16 grados de libertad para el error.

Situaciones como estas ocurren muy frecuentemente en la práctica. En general, deberíamos ser conscientes que en algunos grupos de datos, uno de ellos (o un grupo pequeño de puntos) puede controlar las propiedades del modelo.

3) Los valores aberrantes o fuera de rango pueden perturbar seriamente los mínimos cuadrados estimados.

Por ejemplo, considere los datos en la Figura 5. La observación A parece ser un punto fuera de rango o un valor erróneo porque cae lejos de la línea que ajusta al resto de los datos. Si este punto es realmente un punto aberrante, entonces la estimación del intercepto puede ser incorrecto y el cuadrado de la media del residuo puede ser una estimación excesiva de  $\sigma^2$ . Por otra parte, los datos pueden no ser valores erróneos y pueden ser una evidencia útil en lo que respecta al proceso de investigación.

4) Como se ha mencionado, un análisis de regresión indica una relación fuerte entre dos variables, esto no implica que las variables estén relacionadas en ningún sentido causal. Nuestras expectativas de descubrir relaciones causa y efecto de regresión deberán ser modestas.

5) En algunas explicaciones de regresión el valor de la variable regresora  $X$  requerida para predecir la variable  $Y$ , es desconocida.

### 3.5.2 Medidas del ajuste del modelo.

Los supuestos importantes que se han hecho en el estudio de análisis de regresión son los siguientes:

1. La relación entre  $X$  y  $Y$  es lineal, o por lo menos se aproxima a una línea recta.
2. El término de error  $\varepsilon$  tiene media cero.

3. El término de error  $\varepsilon$  tiene varianza constante  $\sigma^2$
4. Los errores son no correlacionados.
5. Los errores se distribuyen normalmente.

Los supuestos 4 y 5 implican que los errores son variables aleatorias independientes. El supuesto 5 se requiere para pruebas de hipótesis e intervalos de estimación.

Siempre se considerará la validez de estos supuestos para conducir y examinar la suficiencia del modelo. Los tipos de modelos inadecuados discutidos aquí tienen consecuencias potencialmente serias. Las violaciones de los supuestos pueden proporcionar un modelo inestable en el sentido de que un muestreo diferente podría conducir a un modelo totalmente diferente con conclusiones opuestas. Comúnmente no se puede detectar violaciones de supuestos por pruebas del resumen estadístico estándar, tales como el estadístico  $t$ ,  $F$  o  $R^2$ . Estas son propiedades del modelo general, las cuales no aseguran suficiencia en el modelo.

### 3.5.3 Análisis de residuos.

#### Definición de Residuos

Se han definido a los residuos como

$$\begin{aligned}e_i &= y_i - \hat{y}_i \\ i &= 1, \dots, n\end{aligned}\tag{3.73}$$

donde  $y_i$  es una observación y  $\hat{y}_i$  es el valor estimado correspondiente. Un residual puede observarse como la desviación entre los datos y el ajuste, es una medida de la variabilidad no explicada por el modelo de regresión. Es conveniente pensar que los residuales son valores observados de los errores. Así cualquier violación de los supuestos sobre los errores es mostrado en la ecuación anterior de los residuos. El análisis es un método efectivo para descubrir diversos tipos de modelos deficientes.

Los residuos tienen diversas propiedades importantes. Los residuos tienen media cero y su varianza es aproximada a un promedio

$$\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-2} = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{SS_E}{n-2} = MS_E \quad (3.74)$$

Los residuos no son independientes, sin embargo los  $n$  residuos tienen únicamente  $n-2$  grados de libertad asociados a ellos. Esta no independencia de los residuos tiene poco efecto sobre su uso para investigar el ajuste del modelo mientras que  $n$  no sea pequeña.

Algunas veces es útil trabajar con los residuos estandarizados

$$d_i = \frac{e_i}{\sqrt{MS_E}} \quad i = 1, \dots, n \quad (3.75)$$

Los residuos estandarizados tienen media cero y una varianza aproximada a la unidad. La ecuación anterior reduce los residuos dividiéndolos por su desviación estándar promedio. En algunos grupos de datos de residuos en regresión pueden tener desviaciones estándares que difieren mucho.

En el modelo de regresión lineal simple

$$\begin{aligned} V(e_i) &= V(y_i - \hat{y}_i) \\ &= V(y_i) + V(\hat{y}_i) - 2Cov(y_i, \hat{y}_i) \\ &= \sigma^2 + \sigma^2 \left[ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right] - 2Cov(y_i, \hat{y}_i) \end{aligned} \quad (3.76)$$

Ahora, se puede mostrar que

$$\begin{aligned} \text{Cov}(y_i, \hat{y}_i) &= \text{Cov}\left[y_i, \bar{y} + \frac{S_{xy}}{S_{xx}}(x_i - \bar{x})\right] \\ &= \sigma^2 \left[ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right] \end{aligned} \quad (3.77)$$

En consecuencia, la varianza del  $i$ -ésimo residuo es

$$V(e_i) = \sigma^2 \left[ 1 - \left( \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \right] \quad (3.78)$$

Los residuos estudentizados están definidos como

$$r_i = \frac{e_i}{\sqrt{MS_E \left[ 1 - \left( \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \right]}}, i = 1, \dots, n \quad (3.79)$$

Note que en (3.79) el residuo  $e_i$  de mínimos cuadrados ordinarios ha sido dividido por su error estándar exacto, más que por un valor promedio como en (3.75). Los residuos estudentizados son extremadamente útiles en los análisis de regresión.

En grupos de datos pequeños los residuos estudentizados son frecuentemente una reducción más apropiada que los residuos estandarizados, porque la diferencia entre los dos métodos de reducción de los mínimos cuadrados residuales.

### 3.5.4 Detección y tratamiento de puntos aberrantes o fuera de rango.

Un punto aberrante o fuera de rango es una observación extrema. Los residuos que son considerablemente más grandes en valor absoluto que los otros, cuando se dice que hay tres o cuatro desviaciones estándares de la media, son puntos aberrantes potenciales. Los puntos aberrantes son puntos de datos que no son típicos del resto de los datos. Dependiendo de su ubicación en el espacio  $X$ , los puntos aberrantes pueden tener moderados o severos efectos en el modelo de

regresión (p. ej., ver Figuras 3, 4 y 5). Las gráficas de residuos y gráficas de probabilidad normal, son útiles para identificar puntos aberrantes. La examinación de residuos es una manera óptima para identificar puntos aberrantes potenciales. Un método general para el tratamiento óptimo de los puntos aberrantes se encuentra en Barnett y Lewis [1978] y Myers [1990].

Los puntos aberrantes deberán ser cuidadosamente investigados para ver si existe una razón justificada para su comportamiento. Algunas veces estos puntos son valores erróneos, que ocurren como resultados inesperados, pero que a su vez son sucesos explicables.

Los ejemplos incluyen medidas o análisis defectuosos, edición incorrecta de datos, y fracaso de un instrumento medidor. Si éste es el caso, entonces los puntos aberrantes deberían corregirse (si es posible) o borrarse de la base de datos. Claramente, desechar valores erróneos es lo deseable porque la extracción de mínimos cuadrados de la ecuación estimada minimiza la suma de cuadrados de residuos. Sin embargo, se ha enfatizado que habría una fuerte evidencia no estadística de que el punto aberrante o fuera de rango es un valor erróneo antes de que sea desechado. A veces se encuentra que el punto aberrante es algo inesperado, pero perfectamente una observación plausible. Borrar estos puntos para mejorar el ajuste de la ecuación puede ser peligroso, ya que puede dar falsas predicciones en la estimación. Ocasionalmente se ha encontrado que el punto aberrante es más importante que el resto de los datos porque puede controlar muchas propiedades claves del modelo. Los puntos aberrantes pueden indicar también insuficiencias en el modelo, tales como fracasos para ajustar los datos en una región dada del espacio  $X$ . Si el punto fuera de rango es un punto de respuesta particularmente deseable (bajo costo, alto rendimiento, etc.), el conocimiento de los valores regresivos cuando esa respuesta es observada, puede ser sumamente valioso. Los análisis de seguimiento e identificación de puntos aberrantes frecuentemente resultan un proceso importante o un nuevo conocimiento en lo que concierne a factores cuyo efecto se desconoce.

Las diversas pruebas estadísticas han sido propuestas para detectar y rechazar puntos aberrantes o fuera de rango. Por ejemplo, ver Anscombe [1960], Anscombe y Tukey [1963], Elienberg [1976], y Rosner [1975], Stefansky [1971, 1972] que han propuesto una prueba aproximada para identificar puntos aberrantes basados en los residuos máximos estandarizados.

$$|e_i| / \sqrt{\sum_{i=1}^n e_i^2} \quad (3.80)$$

que es particularmente fácil de aplicar. Los ejemplos de esta prueba y las otras referencias conexas están en Cook y Prescott [1981], Daniel [1976], y Williams [1973].

Mientras estas pruebas pueden ser útiles para identificar puntos aberrantes, éstos no debería ser interpretados para indicar que los puntos descubiertos deben rechazarse inmediatamente. Como se ha notado, estos puntos pueden ser pistas importantes que contienen información valiosa.

El efecto de puntos aberrantes sobre el modelo de regresión puede ser fácilmente verificado para eliminar esos puntos y volver a estimar la ecuación de regresión. Se ha encontrado que los valores de los coeficientes de regresión o el resumen estadístico tales como los estadísticos  $t$ ,  $F$  o  $R^2$ , y el cuadrado medio del residuo puede ser muy sensible a los valores fuera de rango. Las situaciones en las que un porcentaje relativamente pequeño de los datos tienen un impacto importante sobre el modelo que no puede ser aceptable por el usuario de la ecuación de regresión. Generalmente se asume que una ecuación de regresión es válida si no es demasiado sensible a pocas observaciones. Se desearía que la relación se fijara en todas las observaciones y no únicamente en pocos datos.

## 4. METODOLOGÍA Y RESULTADOS

### 4.1 DISEÑO DEL MODELO.

Para lograr los objetivos propuestos y probar las hipótesis planteadas en el presente trabajo se empleó un modelo econométrico incluyendo a las variables más determinantes en el consumo de la piña en México, considerando los fundamentos básicos de la teoría de la demanda.

Esto es, se empleó un modelo econométrico de la forma:

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_8 X_{8i} + \varepsilon_i$$

Con:

$\hat{Y}_i$  = Cantidad consumida de piña en el año i en Toneladas (Estimación).

$X_1$  = Precio medio rural de la piña en el año i (\$/Ton.)

$X_2$  = Cantidad producida de piña en año i (\$/Ton.)

$X_3$  = Tamaño de la población en el año i.

$X_4$  = Ingreso de la población en el año i, obtenida de dividir el PIB nacional en términos reales entre la población total (\$/Habitante)

$X_5$  = Cantidad exportada en el año i (Ton.)

$X_6$  = Precio medio rural del melón (\$/Ton.)

$X_7$  = Precio medio rural de la sandía (\$/Ton.)

$X_8$  = Precio medio rural de la manzana (\$/Ton.)

## 4.2 ORIGEN DE LOS DATOS.

El modelo econométrico fue evaluado estadísticamente empleando una serie de datos correspondientes a cada una de las variables durante el periodo 1980-2000 (Ver anexo C). Nótese que en este caso las variables precio del melón, precio de la sandía precio de la manzana fungen como productos relacionados.

La recopilación de la información referente a cada una de las variables incluidas en el modelo inicial fue obtenida de diversas fuentes tales como: anuarios estadísticos de INEGI, Bases de datos de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), bases de datos de la SAGARPA, bases de datos del Banco de México (BANXICO), e información proporcionada por la Secretaría General de Fomentos a la Agricultura.

## 4.3 METODOS DE LA REGRESION LINEAL MÚLTIPLE.

Finalmente se evaluó dicho modelo empleando regresión lineal múltiple mediante los métodos Stepwise, Backward y Forward con el fin de asegurarnos que el modelo obtenido fuese el mejor pues dichos métodos nos permiten obtener únicamente modelos con variables significantes.

**Bacward:** Este método inicia considerando todas las variables que se incluyen en el modelo y en cada uno los siguientes pasos va eliminando a las variables menos significantes. Se detiene hasta tener en el modelo solo variables significantes.

**Fordward:** Este método incluye en cada uno de los pasos a la variable de mayor significancia (la que explica en mayor proporción a la variable dependiente) y se detiene en el momento que no existen variables significativas.

**Stepwise:** Es una combinación de los dos métodos anteriores, es decir, agregando la variable más significativa y en el siguiente paso realiza pruebas de significancia, se detiene cuando no existen ya variables significativas.

#### 4.4 INTERPRETACION DE RESULTADOS.

Empleando los métodos que se mencionaron en el apartado anterior se obtuvo que en cada uno de ellos sólo las variables producción ( $X_2$ ) y exportaciones ( $X_3$ ) resultaron ser significativa con un nivel  $\alpha = 0.05$ . Cada uno de los métodos proponía modelos diferentes con niveles de significancias también diferentes, por lo que se realizaron pruebas individuales en cada uno de los modelos propuestos anteriormente.

El modelo encontrado hasta aquí explicaba el 99.19% de la variabilidad presentada en el consumo de piña en México, sin embargo a fin de evitar problemas de colinealidad y dado que la variable exportación ( $X_3$ ) no explica una proporción considerable de la variabilidad en cuestión (Consumo de piña en México) se optó por eliminar dicha variable para simplificar el análisis estadístico.

Así mismo, el modelo obtenido es:

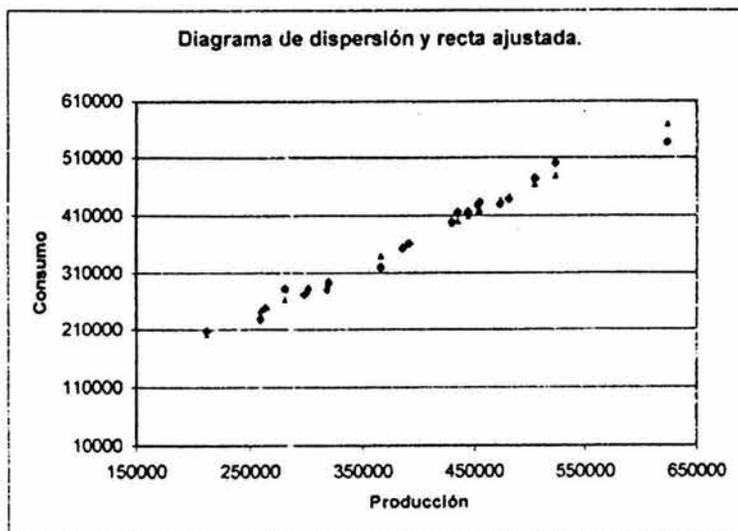
$$\text{Consumo} = f(\text{Producción}) \text{ o bien } \text{Consumo} = \beta_0 + \beta_1 + \varepsilon$$

Ahora haciendo:  $Y = \text{Consumo}$  y  $X = \text{Producción}$ , y dado que los parámetros de regresión serán estimaciones por estar trabajando con una muestra, tenemos que el modelo será:  $Y_i = \beta_0 + \beta_1 X_i + \varepsilon'_i$  y al evaluarla se tiene que  $\beta_0 = 11289$  y  $\beta_1 = 0.89522b$  por lo que el modelo de regresión que servirá para estimar el consumo de piña en México es:

$$Y'_i = 11289 + 0.89522X_i + \varepsilon'_i$$

Es decir, independientemente de los niveles de producción de la piña en México se prevé que se consumirán o habrá una demanda de alrededor de 11,289 toneladas anuales (los cuales podrán ser obtenidos a partir de las importaciones) y por cada tonelada que se produzca el consumo se incrementará en 0.895 toneladas.

**Gráfico 9:** Diagrama de dispersión. Nótese que los datos efectivamente tienen una tendencia lineal.



Para verificar la validez del modelo se plantea un juego de hipótesis de la siguiente manera:

$$H_0 : \beta_1 = 0 \text{ (El modelo no sirve)} \text{ vs } H_a : \beta_1 \neq 0$$

Ahora, para definir la aceptación o rechazo del modelo se empleó la tabla de análisis de varianza.

Los resultados obtenidos fueron:

Tabla de análisis de varianza:					
FV	GL	Suma de Cuadrados	Cuadrados Medios	F calculada	F tabular
Modelo	1	1.744984E11	1.744984E11	925.50	4.38
Error	19	3582362610	188545401		
Total	20	1.780808E11			
Corregido					

De esta tabla se puede ver que el modelo propuesto es válido debido a que se rechaza la hipótesis nula a favor de la alterna pues el valor de  $F_{calculado}$  es mayor que el valor de  $F_{tabular}$ , esto es:  $F_c = 925.50 > F_{tab} = 4.38$ . Todo esto con una confiabilidad del 95%.

Además el modelo explica de acuerdo al coeficiente de determinación  $R^2$  (Ver anexo G) una proporción de 97.99% de la variabilidad total que se presenta en la variable consumo. Por lo tanto, debe esperarse que los valores observados y estimados sean muy parecidos.

Para evaluar la efectividad del modelo se eliminaron las dos últimas observaciones de la variable dependiente (Consumo) para posteriormente ser estimadas junto con sus respectivos intervalos de confianza y obtener lo siguiente:

Valor eliminado	Valor estimado	Límite inferior	Límite superior
474428	457839	429054	486624
501104	473572	444455	502689

De dichas estimaciones se pudo apreciar que los intervalos de confianza contienen a las observaciones previamente eliminadas.

Para reforzar lo anterior, se elaboró también una gráfica de los valores observados y estimados en la variable consumo obteniéndose lo siguiente:

**Gráfico 12:** Estimaciones del consumo de piña en México empleando el modelo de regresión obtenido.



En ésta gráfica se aprecia la validez del modelo, pues no existe una marcada diferencia entre los valores observados y estimados, por lo tanto se concluye que el modelo obtenido es el adecuado.

## 5. CONCLUSIONES

### 5.1 PREDICCIONES DEL CONSUMO DE PIÑA EN MÉXICO.

Empleando el modelo de regresión obtenido para explicar la variabilidad que se presenta en el consumo de piña en México, se realiza una serie de predicciones sobre el comportamiento de éste a partir de la producción.

Así tenemos por ejemplo que el volumen de la producción para el año 2001 fue 535 mil toneladas de acuerdo a las cifras reportadas por la FAO, ahora dado que se desconoce el nivel de consumo para el mismo año podemos estimarla empleado el modelo y con ello obtener que el volumen consumido durante ese año fue de aproximadamente 490 mil 233 toneladas, aunque ésta cantidad puede variar entre 459 mil 515 y 520 mil 950 toneladas, de acuerdo a los intervalos de confianza estimadas a un nivel  $\alpha = 0.05$ .

De la misma manera, si quisiéramos estimar la cantidad de piña consumida durante el año 2002 sólo se necesitaría el volumen de la producción, la cual se desconoce por el momento. Esto puede resolverse usando series de tiempo o un modelo de regresión que nos emite la producción (Oferta), la última opción nos resulta la menos viable pues no se dispone de la información de algunas variables que la función de la oferta considera, en tanto que mediante serie de tiempo sólo necesitamos el comportamiento de la variable producción en los últimos años (al menos 30 años).

Así empleando el comportamiento anual de la variable producción correspondiente al periodo 1961-2001 se estima, mediante métodos de series de tiempo, que la producción para el año 2002 será de aproximadamente 546 mil 832 toneladas por lo que se espera un consumo de 500 mil 825 toneladas aunque ésta puede variar entre 469 mil 890 y 531 mil 760 toneladas de acuerdo a los intervalos con 95% de confiabilidad obtenidos a partir del modelo.

El modelo de series de tiempo que ajusta a los datos de producción es:

$$X_t + 1.0392X_{t-1} + 0.6248X_{t-2} - 0.1568X_{t-3} - 0.2568X_{t-4} = 0.0935Z_{t-1} - 0.3295Z_{t-2} - 0.7638Z_{t-3} + Z_t$$

con  $Z_t \approx WN(0.348177E+10)$

Para llegar a este modelo se diferenciaron 2 veces los datos a distancia 1 debido a que estos no se comportaban como un proceso estacionario (condición necesario en el ajuste de series de tiempo), posteriormente se graficaron las funciones de autocorrelación habitual (obtenida a partir de  $Z_t$ ) y parcial a fin de determinar de manera preeliminar el orden del modelo. Así el parámetro  $d$  tiene dimensión 2, porque se diferenciaron dos veces los datos a distancia 1; el parámetro  $p$  tiene dimensión 4 de acuerdo al gráfico de autocorrelación parcial y finalmente el parámetro  $q$  es de dimensión 3 de acuerdo al gráfico de autocorrelación tomando en cuenta que los parámetros de los modelos ARIMA<sup>1</sup> en series de tiempo son  $(p,d,q)$  donde  $p$  representa el orden del modelo autorregresivo,  $d$  el número de diferenciaciones a distancia  $k$  y  $q$  el orden del promedio móvil.

---

<sup>1</sup> Mayores detalles sobre los modelos de ARIMA y algunos modelos de series de tiempo se pueden ver en Brockwell & Davis 1996 cap 6.

## 5.2 CONCLUSIONES.

Después de haber realizado el análisis estadístico para obtener el modelo de regresión que describa el comportamiento del consumo de piña en México, se concluye lo siguiente:

- El consumo de Piña en México está determinada en mayor proporción por los volúmenes de producción y no por los factores que se establecen en la teoría de la demanda.
- Contrario a las hipótesis planteadas al inicio de este trabajo, las variaciones en el precio, el ingreso, el tamaño de la población y el precio de los productos relacionados, no han sido factores determinantes o influyentes en el comportamiento del consumo de piña en México durante el periodo analizado. (1990-2000).
- El principal problema que enfrentan los productores de piña en México no es la falta de mercado o la competitividad internacional pues en los últimos años el mercado interno ha absorbido alrededor del 95% de lo producido; así, el problema consiste más bien en la estacionalidad del producto ya que ocasiona la variación exagerada en los precios pagados a los productores ocasionado con ello problemas sociales y políticos como las vividas recientemente en el año 2001 (plantones de productores en demanda de mejores precios para su producto).
- El volumen de las importaciones recientes de piña procesada que se han hecho en México ha incrementado debido a que la capacidad instalada en la industria está siendo subutilizada al funcionar sólo en épocas de alta producción. Por ello surge la necesidad de activarla de manera inmediata pues durante el periodo analizado las importaciones de piña procesada rebasan por mucho a las exportaciones.
- México se encuentra a la altura de cualquier productor de piña en el mundo si se comparan rendimientos, ya que éstos están por encima de los promedios internacionales. Así, México debería ser competitivo en la comercialización

internacional de este fruto, lo cual no sucede debido a que la mayoría de los países productores son subsidiados por organismos gubernamentales o por empresas de frutas (Dole y Del Monte en Costa Rica y Honduras) en tanto que los productores de México carecen de subsidios.

### **5.3 RECOMENDACIONES.**

Algunas alternativas de solución a los problemas encontrados en el proceso de producción y comercialización de la piña son:

- ✓ A fin de evitar la caída de los precios en las épocas de abundancia del producto, es necesario regular la producción aplicando las técnicas de carburación, que es el control de la floración a través de reguladores de crecimiento (mediante carburo de calcio); sólo que dicha programación debe hacerse de una manera más sistematizada de cómo se ha venido haciendo en la Región del Bajo Papaloapan evitando realizar ésta carburación de manera independiente, sino organizando a los productores para establecer programas que nos lleven a evitar sobreproducciones en los periodos que éstos ocurren.
- ✓ Por otro lado, a fin de reducir el crecimiento de las importaciones de piña procesada en México, es necesario fortalecer la industria nacional a través de la integración de los productores e industriales mediante el establecimiento de compromisos de tiempos, cantidades, calidades y precios que representen ventajas para ambos y con ello abastecer a la industria no sólo en la época de sobreproducción sino de manera permanente.
- ✓ Finalmente, otra posible alternativa que surge para frenar la marcada variación de los precios de la piña en México debido a la estacionalidad del producto es mediante el establecimiento de un rango de precios que abarquen un mínimo y un máximo conveniente para demandantes y productores. Para ello es necesario que órganos gubernamentales como SAGARPA, la Secretaría General de Fomento a la Agricultura, etc., adquieran un papel de mayor protagonismo.

# ANEXOS

# ANEXO A

**INDICADORES INTERNACIONALES EN LA PRODUCCION, CONSUMO Y  
COMERCIALIZACIÓN DE LA PIÑA**

**Principales países productores de piña en el mundo durante 1990-2000 (Ton.)**

Pais/Año	1990	1991	1992	1993	1994	1995	1996
Tailandia	1865290	1930691	2180000	2589000	2370000	2087707	1986700
Brasil	1103897	1190307	1238991	1251873	1484327	1426361	1145981
Filipinas	1155748	1117156	1135151	1287398	1334960	1442820	1542240
India	881000	768000	859000	1000000	1010000	1060000	1200000
Nigeria	763000	800000	800000	800000	800000	800000	800000
Otros	5523037	5297263	5199405	5231274	5205287	5659765	5551585
Mundial	11,291,972	11,103,417	11,412,547	12,159,545	12,204,574	12,476,653	12,226,506

Pais/Año	1997	1998	1999	2000	TCM	Promedio	Participación (%)
Tailandia	2083390	1786234	2371791	2287000	2.26	2,139,800	17.50
Brasil	1073263	1113219	1477030	1292800	1.71	1,254,368	10.26
Filipinas	1638000	1488700	1530033	1524567	3.19	1,381,525	11.30
India	1250000	1280000	1006000	1100000	2.49	1,037,636	8.48
Nigeria	830000	857000	881000	881000	1.55	819,273	6.70
Otros	5573440	5624404	6339638	6363563	1.52	5,597,151	45.77
Mundial	12,448,093	12,149,557	13,605,492	13,448,930	1.91	12,229,753	100.00

**Superficie sembrada de piña (Has.) en el mundo durante el periodo 1990-2000.**

Pais/Año	1990	1991	1992	1993	1994	1995	1996
Tailandia	74,514	79,693	89,760	99,840	99,360	90,493	83,341
Brasil	33,167	35,904	37,402	39,719	45,218	44,384	45,843
Filipinas	59,619	57,745	60,588	66,925	68,708	68,600	45,046
India	54,587	63,647	57,059	59,436	70,000	80,000	82,000
Nigeria	100,000	100,000	100,000	100,000	100,000	100,000	100,000
Otros	300,119	286,433	283,999	295,345	289,050	306,269	300,124
Mundial	622,006	623,422	628,808	661,265	672,336	689,746	656,354

Pais/Año	1997	1998	1999	2000	TCM	Promedio	Participación (%)
Tailandia	84,698	81,874	97,101	97,280	3.1	88,905	13.23
Brasil	53,567	54,998	56,917	55,749	6.8	45,715	6.80
Filipinas	40,441	37,714	37,432	43,449	-2.7	53,297	7.93
India	82,000	82,000	74,200	80,000	4.7	71,357	10.62
Nigeria	110,000	115,000	115,000	115,000	1.5	105,000	15.62
Otros	305,788	320,222	338,644	360,927	2.0	307,902	45.81
Mundial	676,494	691,808	719,294	752,405	2.1	672,176	100.00

**Rendimiento (hectogramos/Ha.) a nivel mundial durante el periodo 1990-2000.**

Pais/Año	1990	1991	1992	1993	1994	1995	1996
Tailandia	250,327	242,266	242,870	259,315	238,527	230,704	238,382
Brasil	332,830	331,525	331,263	315,182	328,260	321,368	249,980
Filipinas	193,856	193,464	187,356	192,364	194,295	210,324	342,370
India	161,394	120,666	150,546	168,248	144,286	132,500	146,341
Nigeria	76,300	80,000	80,000	80,000	80,000	80,000	80,000
Mundial	181,541	178,104	181,495	183,883	181,525	180,888	186,279

Pais/Año	1997	1998	1999	2000	Promedio	Ton/Ha
Tailandia	245,979	218,169	244,260	235,095	240,535.82	24
Brasil	200,358	202,411	259,506	231,897	282,234.55	28
Filipinas	405,034	394,734	408,750	350,887	279,403.09	28
India	152,439	156,098	135,580	137,500	145,963.45	15
Nigeria	75,455	74,522	76,609	76,609	78,135.91	8
Mundial	184,009	175,620	189,151	178,746	181,931.00	18

**Principales países exportadores de pifia en el mundo durante el periodo 1990-2000 (Ton.)**

Pais/Año	1990	1991	1992	1993	1994	1995	1996
Costa Rica	95,880	100,286	93,491	97,061	160,526	177,604	179,451
Costa de Marfil	138,683	127,760	129,117	128,782	136,634	138,233	173,726
	146,323	167,520	151,946	154,333	161,512	163,524	143,994
Francia	4,302	4,354	3,997	26,094	45,130	44,254	77,181
Honduras	37,700	48,744	49,073	54,339	42,915	44,228	30,636
Otros	151,214	160,959	162,205	217,133	205,591	209,920	234,164
Mundial	574,102	609,623	589,829	677,742	752,308	777,763	839,152

Pais/Año	1997	1998	1999	2000	TCM	Promedio	Participación (%)
Costa Rica	250,100	271,272	304,418	322,453	23.63	186,595	24
Costa de Marfil	180,749	154,339	207,632	191,398	3.80	155,187	20
Filipinas	144,802	117,436	127,682	135,484	-0.74	146,778	19
Francia	85,445	75,560	108,172	77,371	169.80	50,169	6
Honduras	22,949	43,100	43,500	41,511	1.01	41,700	5
Otros	217,540	198,765	262,916	270,992	7.92	208,309	26
Mundial	901,585	860,472	1,054,320	1,039,209	8.10	788,737	100

**Principales países importadores de pifia en el mundo durante el periodo 1990-2000 (Ton.)**

Pais/Año	1990	1991	1992	1993	1994	1995	1996
Estados Unidos	113,885	115,155	123,680	127,134	131,115	124,609	135,255
Japón	128,250	137,786	127,466	120,963	113,527	107,940	96,618
Francia	80,166	86,055	75,967	103,958	117,728	112,090	142,250
Bel-Lux	50,203	45,699	53,068	47,018	46,954	68,094	81,924
Italia	40,499	41,799	50,647	42,294	42,505	39,235	47,820
Otros	182,750	197,734	201,141	221,846	247,229	244,010	272,909
Mundial	595,753	624,228	631,969	663,213	699,058	695,978	776,776

Pais/Año	1997	1998	1999	2000	TCM	Promedio	Participación (%)
Estados Unidos	203,993	252,848	283,090	318,837	18.00	175,418	23
Japón	96,088	84,710	89,866	100,092	-2.20	109,391	14
Francia	142,833	131,940	168,211	148,239	8.49	119,040	15
Bel-Lux	81,708	72,663	87,799	63,513	2.65	63,513	8
Italia	50,755	47,396	74,366	67,677	6.71	49,545	6
Otros	292,219	271,182	328,142	345,975	8.93	255,012	33
Mundial	867,596	860,739	1,031,474	1,044,333	7.53	771,920	100

**Producción de piña y suministro interno en algunos países del mundo durante el periodo 1990-2000 (Ton.)**

		1990	1991	1992	1993	1994	1995
Estados Unidos	Producción	521,630	503,500	498,950	335,660	331,122	313,000
	Suministro Interno	2,143,089	2,213,461	2,218,296	2,122,864	1,751,233	1,810,428
Japón	Producción	31,900	29,300	29,400	26,700	25,000	25,600
	Suministro Interno	213,388	236,167	217,461	215,753	219,756	207,021
Francia	Producción						
	Suministro Interno	150,600	153,085	140,165	124,465	125,535	119,684
Costa Rica	Producción	160,000	170,000	180,000	190,000	240,000	260,000
	Suministro Interno	64,120	69,714	86,509	92,939	68,021	71,271
Costa de Marfil	Producción	232,543	201,571	193,033	202,949	211,686	209,528
	Suministro Interno	93,860	73,811	63,916	74,167	75,052	71,310

		1996	1997	1998	1999	2000	Promedio
Estados Unidos	Producción	315,000	292,860	301,180	319,300	32,150	342214
	Suministro Interno	1,915,280	1,860,039	1,608,684	2,025,121	1,570,121	1930783
Japón	Producción	18,800	14,700	12,800	12,600	13,000	21800
	Suministro Interno	176,023	166,072	146,042	160,131	169,417	193385
Francia	Producción						
	Suministro Interno	114,323	103,977	98,641	101,394	115,415	122480
Costa Rica	Producción	260,000	355,000	400,000	445,000	475,000	285000
	Suministro Interno	69,451	90,946	117,992	130,921	137,190	90825
Costa de Marfil	Producción	250,628	260,556	213,974	257,000	225,675	223558
	Suministro Interno	76,914	79,839	59,656	49,403	34,305	68385

**Valor de las importaciones mundiales de piña (miles de dólares) durante el periodo 1990-2000**

Pais/Año	1990	1991	1992	1993	1994	1995	1996
Estados Unidos	54,870	48,401	52,346	55,301	51,243	50,210	64,029
Japón	57,600	53,072	56,961	52,655	51,759	52,004	44,710
Francia	71,821	65,401	63,181	72,655	77,651	91,300	94,777
Bélgica-Luxemburg	29,879	25,620	30,417	24,628	27,140	58,703	58,503
Italia	40,176	36,246	41,151	28,240	29,298	30,939	35,185
Mundial	376,535	353,788	379,453	370,568	386,816	446,691	480,130

Pais/Año	1997	1998	1999	2000	Promedio
Estados Unidos	102,248	103,259	145,206	157,591	80428
Japón	45,667	37,964	44,007	49,557	49632
Francia	95,146	89,693	89,588	93,524	82249
Bélgica-Luxemburg	59,659	59,039	54,639	57,779	44182
Italia	37,469	37,567	49,327	51,822	37947
Mundial	524,530	504,601	585,293	623,043	457404

**Valor de las exportaciones de piña en el mundo durante el periodo 1990-2000 (miles de dólares)**

Pais/Año	1990	1991	1992	1993	1994	1995	1996
Costa Rica	38,438	38,492	34,741	34,462	56,526	58,621	68,932
Filipinas	22,699	26,961	22,946	23,158	24,378	24,757	24,514
Francia	4,841	4,285	4,217	21,724	37,234	42,789	55,467
Honduras	16,277	14,007	10,477	18,532	9,481	11,750	8,744
Mundial	179,566	184,671	178,352	223,007	262,571	300,105	346,072

Pais/Año	1997	1998	1999	2000	Promedio
Costa Rica	102,848	114,968	128,234	121,550	72,528
Filipinas	27,189	20,841	22,814	24,794	24,096
Francia	59,753	55,494	63,926	53,737	36,679
Honduras	5,645	18,500	19,200	15,774	13,490
Mundial	369,308	377,037	438,355	414,995	297,640

Fuente: Base de datos de la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO), con información actualizada hasta el año 2002.

Nota: Para el suministro de piña en cada uno de los países abarcados en la tabla, éste considera también piñas en lata, en tanto que la producción, las exportaciones e importaciones son de piña en fresco.

# ANEXO B

**INDICADORES DE PRODUCCIÓN Y COMERCIALIZACION DE LA PIÑA EN MEXICO DURANTE EL PERIODO 1990-2000.**

**SUPERFICIE SEMBRADA (Has.)**

Estado/Año	1990	1991	1992	1993	1994	1995	1996
Veracruz	6000	4240	5374	2751	2384	3375	5226
Oaxaca	6200	3870	3000	3000	3150	2041	2891
Tabasco			1100	1350	1519	1500	920
Nayarit	445	444	629	629	698	699	690
Jalisco	50	55	55	95	57	61	89
Otros	176	157	52	45	165	97	146
Nacional	12871	8766	10210	7870	7973	7773	9962

Estado/Año	1997	1998	1999	2000	Promedio	TCM
Veracruz	5226	19562	7221	7811	6288.18	3.018
Oaxaca	2900	2900	1880	1920	3068.36	-6.903
Tabasco	1600	1600	1528	1300	1379.67	2.273
Nayarit	971	964	771	802	703.82	8.022
Jalisco	85	15	76	61	63.55	2.200
Otros	46	97	101	204	116.91	1.591
Nacional	10828	25138	11577	12098	11369.64	-0.632

**SUPERFICIE COSECHADA (Has.)**

Estado/Año	1990	1991	1992	1993	1994	1995	1996
Veracruz	6000	4238	5166	2751	2384	2384	4499
Oaxaca	2250	1750	1105	1500	1600	1956	1436
Tabasco			550	1100	1304	1500	920
Nayarit	436	435	609	561	669	671	676
Jalisco	15	25	45	75	51	61	15
Otros	47	71	21	38	165	48	70
Nacional	8748	6519	7496	6025	6173	6620	7616

Estado/Año	1997	1998	1999	2000	Promedio	% de la sembrada
Veracruz	5186	6510	7221	7811	4922.73	78.29
Oaxaca	1900	2400	1880	1920	1790.64	58.36
Tabasco	920	1528	1528	1300	1183.33	85.77
Nayarit	967	964	755	802	685.91	97.46
Jalisco	85	15	76	61	47.64	74.96
Otros	46	92	75	156	75.36	64.46
Nacional	9104	11509	11535	12050	8490.45	74.68

**PRODUCCIÓN (Ton.)**

Estado/Año	1990	1991	1992	1993	1994	1995	1996
Veracruz	334800	204073	169636	107200	96477	116905	200959
Oaxaca	109800	84252	53040	64500	70400	82120	86160
Tabasco			29150	36000	47072	70500	40
Nayarit	9713	9363	11419	2604	10590	9566	9266
Jalisco	40	339	660	1860	1890	1710	3770
Otros	315	499	242	238	32933	379	1212
Nacional	454668	298526	264147	212402	259362	281180	301407

Estado/Año	1997	1998	1999	2000	Promedio	TCM	Participación (%)
Veracruz	233040	293389	324017	352535	221184.6	0.5297	61.3
Oaxaca	108500	110000	107940	101160	88897.5	-0.7869	24.6
Tabasco	32200	61128	58064	54600	43194.9	10.9134	9.8
Nayarit	13839	12997	7189	7136	9425.6	-2.6531	2.6
Jalisco	2780	375	3160	2755	1758.1	678.75	0.5
Otros	1132	2966	1398	4236	4140.9	124.4762	1.1
Nacional	391491	480855	501768	522422	360748.0	1.4902	100.0

**RENDIMIENTO (Has.)**

Estado/Año	1990	1991	1992	1993	1994	1995	1996
Veracruz	55.80	48.15	32.84	38.97	40.47	49.04	44.67
Oaxaca	48.80	48.14	48.00	43.00	44.00	41.98	60.00
Tabasco			53.00	32.73	36.10	47.00	0.04
Nayarit	22.28	21.52	18.75	4.64	15.83	14.26	13.71
Jalisco	2.67	13.56	14.67	24.80	37.06	28.03	251.33
Otros	6.70	7.03	11.52	6.26	199.59	7.90	17.31
Nacional	51.97	45.79	35.24	35.25	42.02	42.47	39.58

Estado/Año	1997	1998	1999	2000	TCM	Promedio
Veracruz	44.94	45.07	44.87	45.13	-1.91	44.54
Oaxaca	57.11	45.83	57.41	52.69	0.8	49.72
Tabasco	35.00	40.01	38.00	42.00	-2.59	35.99
Nayarit	14.31	13.48	9.52	8.90	-8.01	14.29
Jalisco	32.71	25.00	41.58	45.16	159.36	48.96
Otros	24.61	32.24	18.64	27.15	30.52	32.63
Nacional	43.00	41.78	43.50	43.35	-1.66	42.18

Fuente: Centro de Estadística Agropecuaria (CEA)

**Importaciones y exportaciones de piña fresca en México.**

	1990	1991	1992	1993	1994	1995	1996
Importaciones (Ton.)	35	5	4	23	10	1	49
Valor (miles de dólares)	57	3	10	16	17	3	124
Exportaciones (Ton.)	8683	9817	9768	8184	6558	8438	10198
Valor (miles de dólares)	1001	1145	2276	1968	1864	1633	3362

	1997	1998	1999	2000	Promedio	TCM	Total
Importaciones (Ton.)	103	23	213	303	69.91	76.571	769
Valor (miles de dólares)	203	66	383	478	123.64	73.86	1360
Exportaciones (Ton.)	18337	19827	19612	24409	13075.55	18.111	143831
Valor (miles de dólares)	4596	5961	7032	8297	3557.73	72.887	39135

**Importaciones y exportaciones de piña enlatada en México durante el periodo 1990-2000.**

	1990	1991	1992	1993	1994	1995	1996
Importaciones (Ton.)	367	600	1997	3746	19015	6468	808
Valor (miles de dólares)	346	614	2473	2771	13392	5705	901
Exportaciones (Ton.)	4376	3737	4084	2553	1061	1529	3268
Valor (miles de dólares)	2431	2782	2910	1860	755	1050	2558

	1997	1998	1999	2000	Promedio	TCM	Total
Importaciones (Ton.)	533	648	693	4474	3577.18	1219.07	39349
Valor (miles de dólares)	679	794	636	3148	2859.91		31459
Exportaciones (Ton.)	3841	4148	2185	852	2875.82	-8.053	31634
Valor (miles de dólares)	2917	2779	1806	821	2060.82		22669

Fuente: Base de datos de la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO), con información actualizada hasta el año 2002.

# ANEXO C

**DATOS EMPLEADOS EN EL ANALISIS DEL MODELO.**

<b>Año</b>	<b>Consumo (toneladas)</b>	<b>Producción (Toneladas)</b>	<b>Precio medio rural (\$Ton.)</b>	<b>Exportaciones (toneladas)</b>	<b>Población (miles de habitantes)</b>	<b>Precio del melon (\$/Ton.)</b>	<b>Precio de la sandía (\$/Ton.)</b>	<b>Precio de la manzana (\$/Ton.)</b>	<b>Ingreso per cápita anual en pesos</b>	<b>PIB en términos reales (Base=1993)</b>
1980	536501.00	622729.00	2.08	41873.00	66846.80	4.05	3.07	8.99	56763.07	3794429274
1981	429081.00	473031.00	2.21	30593.00	69186.00	5.02	4.14	10.13	59519.66	4117927389
1982	416095.00	444286.00	6.01	21182.00	70779.00	9.07	6.28	15.04	57877.07	4096480892
1983	397893.00	429624.00	18.00	16704.00	72350.00	19.48	13.84	53.12	54646.31	3953660289
1984	427745.00	453260.00	20.69	16773.00	73908.00	31.39	25.29	40.26	55318.94	4088512293
1985	291593.00	319814.00	26.80	13139.00	75462.00	47.50	35.22	73.66	55365.04	4177956395
1986	280928.00	318368.00	87.53	18255.00	77012.00	80.69	61.90	162.09	52580.36	4049318982
1987	319887.00	366240.00	134.28	20671.00	78559.00	217.87	187.65	283.32	52432.77	4119066044
1988	353042.00	385866.00	198.82	16326.00	80106.00	420.26	346.25	462.84	52080.05	4171924412
1989	415265.00	434822.00	215.57	9929.00	80848.00	595.43	430.27	655.84	53720.60	4343203158
1990	431956.00	454668.00	216.73	8683.00	81249.64	641.19	588.72	841.82	56221.75	4567997294
1991	272303.00	298526.00	273.08	9817.00	84797.00	854.63	678.21	1219.14	56140.28	4760527181
1992	248165.00	264147.00	353.30	9768.00	86381.00	866.48	717.22	889.02	57062.34	4929102324
1993	207219.00	212402.00	545.83	8184.00	87970.00	1021.98	786.41	1072.96	57119.29	5024783882
1994	228580.00	259362.00	805.76	6558.00	89557.00	1041.73	790.81	1346.50	58607.85	5248801721
1995	281682.00	281180.00	1077.38	8438.00	91185.29	1191.89	905.15	1847.98	53982.74	4922431918
1996	281610.00	301406.96	1121.67	10198.00	92710.00	1830.37	1433.63	2580.47	55823.93	5175436431
1997	361420.00	391491.00	1015.78	18337.00	94272.00	1613.72	1389.59	1589.84	58618.69	5526100685
1998	438259.00	480856.00	2118.23	19827.00	95822.00	2119.70	1682.10	3365.09	60500.10	5797240239
1999	474428.00	504339.00	2166.18	19612.00	97356.00	2164.49	1516.78	3542.86	61773.27	6013998388
2000	501104.00	522422.00	2142.21	15444.00	97483.41	2175.80	1721.20	3602.23	64862.11	6413046159

Fuente: Base de datos de la FAO, SAGARPA, INEGI y BANXICO

# ANEXO D

## FUNDAMENTOS TEORICOS PARA LA FUNCION DE LA DEMANDA.

La demanda es el factor decisivo del mercado. Si esto no se toma en cuenta, se provocarían periodos de sobreproducción o déficit en algún producto y se emprendería inversiones antieconómicas.

La demanda constituye un punto clave en los mercados, así como para la formación de precios, por lo cual es necesario conocerla ampliamente.

La demanda del consumidor se *define como las distintas cantidades de un bien que el consumidor puede y está dispuesto a pagar según la variación del precio del bien, permaneciendo constantes todos los demás factores de la demanda.*

La función de demanda del consumidor puede escribirse de dos formas: como un cuadro o tabla de precios y cantidades, y como una gráfica o función algebraica de precios y cantidades (una curva de demanda).

El comportamiento de la curva de demanda indica que a precios altos se demandan menores cantidades de producto, y a precios bajos mayores cantidades. Ello sienta las bases de la "Ley de la Demanda", la cual establece que el precio y la cantidad varían en razón inversa; es decir, la curva de demanda tiene pendiente negativa.

La suma horizontal de las funciones individuales de demanda da origen a la curva de demanda de mercado.

La demanda de mercado se define como las cantidades totales de un producto que los consumidores están en disposición y capacidad de adquirir por periodo de tiempo a los distintos precios alternativos en tanto que otros factores permanezcan constantes.

El supuesto que se encuentra detrás de la teoría de la demanda es que el consumidor desea maximizar su utilidad. Una curva de demanda cualquiera, relaciona las cantidades de equilibrio que se compran a diferentes precios de mercado del bien, sin que cambien los otros factores que influyen en la demanda.

Sin embargo, se debe diferenciar entre un cambio en la cantidad demandada, esto es, entre movimientos a lo largo de la curva de demanda y desplazamiento en la posición de la misma.

## FACTORES QUE PROVOCAN CAMBIOS EN LA DEMANDA DE UN PRODUCTO.

En general, la demanda de un bien no se puede explicar como resultado de un solo factor explicativo. Más bien la demanda debe explicarse en función de un conjunto de factores.

Los factores que provocan desplazamientos en el nivel de la curva pueden agruparse en cuatro principalmente:

- 1) El tamaño de la población y su distribución por edad, área geográfica, etc.
- 2) El ingreso del consumidor y su distribución.
- 3) Los precios y la disponibilidad de otros bienes y servicios.
- 4) Los gustos y preferencias del consumidor.

La función de demanda define simplemente la relación entre el precio y la cantidad comprada por unidad de tiempo mientras permanecen constantes los demás factores y se denota de la siguiente forma:

$$Q_m = D(P_m, P_1, \dots, P_n, P_{ob}, Y, G)$$

En donde para todo periodo  $t$ :

$Q_m$  = Cantidad demandada de un bien.

$P_m$  = Precio del bien considerado.

$P_1, \dots, P_n$  = Precios de los bienes relacionados.

$Y$  = Ingreso

$P_{ob}$  = Población.

$G$  = Gustos y preferencias.

El precio de un bien produce cambios en la cantidad demandada, mientras que la curva de demanda no se mueva, los demás determinantes establecen el nivel de posición de la curva de demanda.

Los cambios a lo largo de la curva de demanda debido a variaciones en el precio pueden presentar tres posibilidades:

- a) La demanda decrece cuando aumenta el precio y aumenta cuando el precio desciende. A esta posibilidad se le llama "*Ley fundamental de la demanda*" e implica que una cantidad de bienes mayores solo puede ser comprada a precios más bajos.
- b) La demanda aumenta cuando se incrementa el precio y disminuye cuando el precio desciende. A este tipo de bienes se les llama "bienes giffen" y constituye la única excepción de la ley de la demanda.
- c) La demanda permanece constante y es insensible a las variaciones del precio.

Otro determinante de la cantidad demandada es el **precio de bienes relacionados** (sustitutos y complementarios), debido a que condicionan el nivel de demanda del bien en cuestión. Existen bienes sustitutos que al aumentar el precio de uno de ellos genera un aumento en el consumo del otro. En los bienes complementarios se da una relación opuesta.

El **ingreso** es también un determinante de la cantidad demandada. Para muchos productos agrícolas el ingreso y la cantidad demandada esta positivamente relacionados. Así, un incremento en el ingreso desplazará la curva de la demanda hacia arriba y a la derecha. Sin embargo para un pequeño grupo de productos puede suceder lo contrario. Por lo tanto, en principio, la interrelación entre el ingreso y la demanda puede oscilar en el rango de positiva hasta negativa, dependiendo de la elasticidad ingreso de la demanda, la cual se define como la relación relativa de la cantidad demandada ante cambios en el ingreso y sugiere que el bien en cuestión es normal. En cambio una elasticidad ingreso mayor que uno, nos indica que el bien es superior o de "lujo". Un bien con elasticidad ingreso negativo será un bien inferior.

**La población** es otro determinante que hace que la cantidad demandada aumente o disminuya proporcionalmente a los aumentos o disminuciones de la misma.

Por su parte, **los cambios en los gastos y preferencias** también contribuyen a los desplazamientos en la demanda de productos, aunque sus efectos a menudo resultan difíciles de separar porque parecen estar asociados a cambios en el ingreso u otras variables.

# ANEXO E

## CONSTRUCCIÓN DEL ÍNDICE ESTACIONAL DE PRECIOS DE LA PIÑA MEDIANA DE LA CENTRAL DE ABASTOS DEL DISTRITO FEDERAL.

La construcción del índice estacional de precios de piña de tamaño mediana presentada en el gráfico 9 nos permitió apreciar con mayor claridad el carácter estacional de la piña en nuestro país. Dicho índice se construyó a partir de una serie de precios reales (Base = 100, junio 2002) de los años 1993-2001.

El procedimiento empleado para tal construcción fue la siguiente:

A partir del gráfico de los datos eliminamos los componente de la serie (Componente de tendencia, cíclico y estacionalidad) mediante una suavización conocida como "el método de promedios móviles", la cual consiste básicamente en la obtención de promedios y centralización de los datos.

Un promedio móvil  $M_t$  de  $k$  puntos en un tiempo  $t$  es formado  $k$  valores secuenciales de  $Y_t$  (variable en observación o descripción). El tiempo  $t$  es tomado para ser un punto medio de esos intervalos de tiempo.

Los promedios móviles son fáciles de calcular cuando  $k$  es impar. Por ejemplo, un promedio móvil centrado para una serie de tiempo  $Y_1, Y_2, \dots, Y_n$  de  $n$  observaciones con  $k = 3$  es definida como la serie de media dadas por:

$$M_2 = \frac{Y_1 + Y_2 + Y_3}{3} \quad M_3 = \frac{Y_2 + Y_3 + Y_4}{3} \quad M_4 = \frac{Y_3 + Y_4 + Y_5}{3} \quad \dots, \quad M_{n-1} = \frac{Y_{n-2} + Y_{n-1} + Y_n}{3}$$

el primer promedio (involucrando tres primeras observaciones) es centrado en  $t = 2$ , el siguiente en  $t = 3$  y el último en  $t = n - 1$ . Mótese que estos promedios móviles son centrados a periodos de tiempos dados en la serie original y que perdemos un punto al inicio y al final de la serie; en general, para  $k$  impar perdemos  $k - 1$  puntos en el proceso de suavización.

Cuando  $k$  es par, como en el caso de datos mensuales o trimestrales, los promedios móviles son centrados entre los periodos dados en la serie original. Por ejemplo, con  $k = 4$ , los primeros tres promedios móviles.

$$M_{2.5} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} \quad M_{4.5} = \frac{Y_3 + Y_4 + Y_5 + Y_6}{4} \quad M_{3.5} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4};$$

debemos centrarlos en  $t = 2.5, 3.5$  y  $4.5$ , para evitar el problema que se tiene en la centralización de los promedios móviles entre los periodos usados en la serie original, los datos son suavizados un poco mas usando pequeños valores de  $k$ , como  $k = 2$ . Con  $k = 2$ , una segunda suavización produce los valores centrados.

$$M_3 = \frac{M_{2.5} + M_{3.5}}{2} \quad M_4 = \frac{M_{3.5} + M_{4.5}}{2} \quad M_5 = \frac{M_{4.5} + M_{5.5}}{2}$$

La doble suavización de la serie ha perdido  $k = 4$  puntos, dos al inicio y dos al final, y es ahora centrada a periodos de tiempo correspondiente a la serie original. Cuando  $k$  es par y la serie es doblemente suavizado como se indicó, la suavización de la serie resultante es referida como un promedio móvil centrado  $k$ -puntos. Teniendo la suavización de la serie centrada o puntos correspondientes de la serie original se facilita la estimación de tendencia, estacionalidad y componente cíclico de la serie.

Para el caso de nuestro índice, se empleó una doble suavización dado que utilizamos una  $k = 12$ , nuestros datos eran mensuales. Una vez llegado hasta esta instancia, se calcula el porcentaje del índice que representa el precio de éste promedio móvil centrado y por último se promedian todos los porcentajes correspondientes a cada mes.

Los datos empleados para la construcción del índice fueron:

AÑO	MES	Precio Nominal de la piña	IPC	Precio real	Suma	Promedio Movil	Promedio movil centrado	Porcentaje del promedio móvil
1993	ENERO	2.23	17.63	12.648894				
	FEBRERO	2.54	19.07	13.319350				
	MARZO	2.74	19.29	14.204251				
	ABRIL	3.20	21.10	15.165877				
	MAYO	2.91	21.87	13.305898				
	JUNIO	2.32	19.83	11.699445		14.279761		
	JULIO	3.60	20.56	17.509728		14.077960	14.1789	123.4618
	AGOSTO	3.06	20.50	14.926829		13.882692	13.9803	106.7651
	SEPTIEMBRE	3.30	20.50	16.097561		13.719330	13.8010	116.6592
	OCTUBRE	3.48	22.41	15.528782		13.525843	13.6226	114.0041
	NOVIEMBRE	3.14	22.86	13.735783		13.482864	13.5044	101.7065
	DICIEMBRE	3.12	23.61	13.214740		13.677353	13.5801	97.3169
1994	ENERO	2.70	26.40	10.227273	171.3571	13.494088	13.5857	75.2878
	FEBRERO	2.53	23.05	10.976139	168.9355	14.029483	13.7618	79.7437
	MARZO	2.51	20.50	12.243902	166.5923	14.383813	14.2066	86.1678
	ABRIL	2.80	21.80	12.844037	164.6320	14.493256	14.4385	88.9545
	MAYO	2.70	21.11	12.790147	162.3101	14.562020	14.5276	88.0245
	JUNIO	2.78	19.81	14.033317	161.7944	14.431366	14.4967	96.7911
	JULIO	3.18	20.77	15.310544	164.1282	14.446307	14.4388	106.0574
	AGOSTO	5.15	24.12	21.351575	161.9291	14.285611	14.3660	148.6538
	SEPTIEMBRE	5.24	25.75	20.349515	168.3538	13.922159	14.1039	144.2744
	OCTUBRE	4.48	26.60	16.842105	172.6058	13.533568	13.7279	122.6892
	NOVIEMBRE	3.93	26.99	14.560948	173.9191	13.384730	13.4591	108.1870
	DICIEMBRE	3.14	26.96	11.646884	174.7442	12.938366	13.1615	88.4752
1995	ENERO	2.79	26.81	10.406565	173.1764	12.154843	12.5466	82.9509
	FEBRERO	2.48	27.41	9.047793	173.3557	11.177967	11.6664	77.5621
	MARZO	2.20	27.91	7.882479	171.4273	10.491408	10.8347	72.7515
	ABRIL	2.55	31.17	8.180943	167.0659	9.846262	10.1688	80.4552
	MAYO	3.77	34.26	11.004086	162.4028	9.132536	9.4894	115.9657
	JUNIO	3.43	39.53	8.676954	160.6168	8.802493	8.9675	96.7511
	JULIO	2.28	38.59	5.908266	155.2604	8.599040	8.7008	67.9114
	AGOSTO	3.79	39.36	9.629065	145.8581	8.443266	8.5212	112.9987
	SEPTIEMBRE	5.64	46.57	12.110801	134.1356	8.433915	8.4386	143.5150
	OCTUBRE	5.25	57.69	9.100364	125.8969	8.386442	8.4102	108.2055
	NOVIEMBRE	3.50	58.37	5.996231	118.1551	8.155484	8.2710	72.4979
	DICIEMBRE	3.98	51.78	7.686365	109.5904	8.004758	8.0801	95.1252

AÑO	MES	Precio Nominal de la piña	IPC	Precio real	Suma	Promedio Movil	Promedio movil centrado	Porcentaje del promedio
1996	ENERO	4.02	50.47	7.965128	105.6299	8.218222	8.1115	98.1885
	FEBRERO	3.74	52.10	7.178503	103.1885	8.069684	8.1440	88.1425
	MARZO	3.68	47.36	7.770270	101.3192	8.298832	8.1843	94.9408
	ABRIL	3.54	46.51	7.611266	101.2070	8.463971	8.3814	90.8203
	MAYO	3.95	47.98	8.232597	100.6373	8.562766	8.5134	96.6966
	JUNIO	3.43	49.94	6.868242	97.8658	8.433683	8.4982	80.8198
	JULIO	3.93	46.40	8.469828	96.0571	8.291031	8.3624	101.2851
	AGOSTO	3.99	50.85	7.846608	98.6187	8.233888	8.2625	94.9743
	SEPTIEMBRE	8.90	59.89	14.860578	96.8362	8.178091	8.2060	181.0944
	OCTUBRE	7.20	64.97	11.082038	99.5860	8.093235	8.1357	136.2075
	NOVIEMBRE	4.35	60.57	7.181773	101.5677	7.916982	8.0051	89.7210
	DICIEMBRE	3.78	61.59	6.137360	102.7532	7.969410	7.9432	77.2621
1997	ENERO	3.54	56.61	6.253312	101.2042	7.925767	7.9476	78.6879
	FEBRERO	3.51	54.06	6.492786	99.4924	7.964345	7.9451	81.7177
	MARZO	3.61	50.84	7.100708	98.8067	7.510596	7.7375	91.7718
	ABRIL	3.52	53.39	6.592995	98.1371	7.164025	7.3373	89.8493
	MAYO	3.32	54.27	6.117560	97.1188	7.006662	7.0853	86.3466
	JUNIO	4.29	57.22	7.497379	95.0038	6.967951	6.9873	107.3021
	JULIO	4.60	57.89	7.946105	95.6329	6.930061	6.9490	114.3425
	AGOSTO	4.95	59.57	8.309552	95.1092	6.934732	6.9324	119.8598
	SEPTIEMBRE	6.38	67.76	9.415584	95.5721	6.876412	6.9056	136.3557
	OCTUBRE	5.02	72.51	6.923183	90.1272	6.951644	6.9140	100.1412
	NOVIEMBRE	3.59	67.82	5.293424	85.9683	7.111393	7.0315	75.2796
	DICIEMBRE	3.44	60.64	5.672823	84.0799	7.151197	7.1313	79.5452
1998	ENERO	3.30	56.91	5.798629	83.6154	7.075310	7.1133	81.5191
	FEBRERO	3.52	53.75	6.548837	83.1607	7.243763	7.1595	91.4641
	MARZO	3.50	54.68	6.400878	83.2168	7.429074	7.3364	87.2471
	ABRIL	4.43	59.10	7.495770	82.5169	7.821245	7.6252	98.3001
	MAYO	4.93	61.36	8.034550	83.4197	8.219158	8.0202	100.1747
	JUNIO	5.62	70.47	7.975025	85.3367	8.453969	8.3366	95.6634
	JULIO	4.82	68.51	7.035469	85.8144	8.577803	8.5159	82.6223
	AGOSTO	7.46	72.21	10.330979	84.9037	8.564374	8.5711	120.5357
	SEPTIEMBRE	9.72	83.51	11.639325	86.9252	8.564575	8.5645	135.8970
	OCTUBRE	10.20	87.71	11.629233	89.1489	8.548537	8.5566	135.9079
	NOVIEMBRE	9.13	90.68	10.068372	93.8549	8.513832	8.5312	118.0241
	DICIEMBRE	7.92	93.28	8.490566	98.6299	8.393630	8.4537	100.4397

AÑO	MES	Precio Nominal de la piña	IPC	Precio real	Suma	Promedio Movil	Promedio movil centrado	Porcentaje del promedio
1999	ENERO	6.90	94.72	7.284628	101.4476	8.348110	8.3709	87.0208
	FEBRERO	6.35	99.41	6.387687	102.9336	8.255876	8.3020	76.9444
	MARZO	5.90	92.14	6.403299	102.7725	8.272356	8.2641	77.4874
	ABRIL	6.60	90.37	7.303309	102.7749	8.181601	8.2270	88.7702
	MAYO	6.79	89.13	7.618086	102.5824	8.087327	8.1345	93.6472
	JUNIO	5.75	88.02	6.532606	102.1660	7.959145	8.0232	81.7387
	JULIO	5.12	78.90	6.489227	100.7236	7.893058	7.9261	82.5113
	AGOSTO	7.74	83.91	9.224169	100.1773	7.875688	7.8844	117.9219
	SEPTIEMBRE	12.41	104.84	11.837085	99.0705	7.900304	7.8880	151.2467
	OCTUBRE	12.41	117.74	10.540173	99.2683	7.935578	7.9179	134.1573
	NOVIEMBRE	9.56	106.97	8.937085	98.1792	7.817423	7.8765	114.3578
	DICIEMBRE	5.84	84.00	6.952381	97.0479	7.922790	7.8701	79.5671
2000	ENERO	5.71	87.96	6.491587	95.5097	8.105507	8.0141	81.6315
	FEBRERO	5.24	84.80	6.179245	94.7167	8.033477	8.0695	77.1699
	MARZO	5.16	77.03	6.698689	94.5083	7.768693	7.9011	85.4454
	ABRIL	5.72	74.03	7.726597	94.8036	7.564039	7.6664	101.6013
	MAYO	4.36	70.32	6.200228	95.2269	7.406766	7.4854	83.5184
	JUNIO	5.27	67.59	7.797011	93.8091	7.428436	7.4176	105.5466
	JULIO	6.83	78.67	8.681836	95.0735	7.433283	7.4309	116.8318
	AGOSTO	7.63	91.27	8.359812	97.2661	7.498033	7.4657	111.9772
	SEPTIEMBRE	9.00	103.93	8.659675	96.4017	7.624982	7.5615	114.5192
	OCTUBRE	7.90	97.72	8.084323	93.2243	7.585998	7.6055	106.2993
	NOVIEMBRE	5.52	78.30	7.049808	90.7685	7.693588	7.6398	92.2821
	DICIEMBRE	5.11	70.85	7.212421	88.8812	7.634685	7.6641	94.1098
2001	ENERO	4.93	75.27	6.549754	89.1412	7.453900	7.5443	86.8138
	FEBRERO	4.42	63.54	6.956248	89.1994	7.681132	7.5675	91.9256
	MARZO	5.05	61.42	8.222077	89.9764	8.022991	7.8521	104.7189
	ABRIL	4.92	67.78	7.258778	91.4998	8.014715	8.0189	90.5260
	MAYO	5.39	71.95	7.491313	91.0320	7.964028	7.9894	93.7616
	JUNIO	5.48	77.29	7.090180	92.3231		3.9820	178.0471
	JULIO	4.25	65.26	6.512412	91.6162			
	AGOSTO	8.04	72.52	11.086597	89.4468			
	SEPTIEMBRE	12.30	96.38	12.761984	92.1736			
	OCTUBRE	8.52	106.70	7.985005	96.2759			
	NOVIEMBRE	6.62	102.77	6.441569	96.1766			
	DICIEMBRE	5.80	93.17	6.225180	95.5683			

# BIBLIOGRAFIA

Johnston J., *Econometric Methods*; 2<sup>nd</sup> Edition; Mc Graw Hill; Tokio 1972.

Canavos C. George; *Probabilidad y Estadística*; México 1988.

Said Infante Gil y Guillermo P. Zárate de Lara; *Métodos estadísticos*; 2<sup>a</sup> . Edición; Trillas; México 2003.

Montgomery, D.C. and Peck, E.A; *Introduction to lineal Regression Analisis*; 2<sup>nd</sup>. Edition; Wiley, New York 1991.

Montgomery, D.C and Hines, W.W.; *Probabilidad y Estadística*; 3<sup>a</sup>. Edición; México 2001.

Hoel Paul G; *Introducción a la Estadística Matemática*; 2<sup>a</sup>. Edición; Ariel; España 1980.

Rebolledo M. A.; *Tecnología para la producción de piña en México*; INIFAP; México 1998.

Damodar N. Gujarati; *Econometría*; 2<sup>a</sup>. Edición; Mc Graw Hill; México 1992.

Parkin Michael; *Microeconomía*; 5<sup>a</sup>. Edición; Addison-Wesley Iberoamericana; México 2001.

William Mendenhall & Roberto J. Beaver; 1992. *A course in business statistics*, 3<sup>nd</sup> Edition. PWS

Sistema Electrónico para la Consulta de Información del Sector Alimentario en México 2002.

Universidad de Barcelona; *Análisis Estadístico con SPSS 11.0 para Windows*. Volumen 1. Estadística Básica; 2<sup>a</sup>. Edición; Mc Graw Hill; España 2002.

[www.inegi.gob.mx](http://www.inegi.gob.mx)

[www.fao.org](http://www.fao.org)

[www.sagarpa.gob.mx](http://www.sagarpa.gob.mx)

[www.banxico.org](http://www.banxico.org)

[www.sniim.gob.mx](http://www.sniim.gob.mx)