

UNIVERSIDAD AUTONOMA CHAPINGO
DEPARTAMENTO DE ECONOMÍA AGRÍCOLA
CHAPINGO, MÉXICO

✓ USO DE INSUMOS MEJORADOS EN LA PRODUCCIÓN DE MAÍZ
Y SU RELACIÓN CON FACTORES SOCIOECONÓMICOS
DE LOS PRODUCTORES EN UN SISTEMA DE PRODUCCIÓN

JORGE O. AUED HUERTA

T E S I S

REQUISITO PARCIAL PARA OBTENER
EL GRADO DE MAESTRO EN CIENCIAS
ESPECIALISTA EN ECONOMIA DEL DESARROLLO RURAL

1982

Esta investigación fue realizada bajo la Dirección del Comité de Elaboración de Tesis que abajo se indica. Habiendo sido aprobada por el mismo y aceptada como requisito parcial para la obtención del grado de:

MAESTRO EN CIENCIAS ESPECIALISTA EN
ECONOMIA DEL DESARROLLO RURAL

Chapingo, Méx., Septiembre de 1982.

COMITE DE ELABORACION DE TESIS

DIRECTOR:



M.C. MARCOS PORTILLO VAZQUEZ

ASESOR:



DR. ANGEL MARTINEZ GARZA

AGRADECIMIENTO

Al Ing. Marcos Portillo Vázquez, por su apoyo y dirección durante toda mi carrera y en el desarrollo de este trabajo.

Al Dr. Angel Martínez Garza, por sus valiosas asesorías.

A los profesores José Antonio Avila Dorantes, Pedro A. Villiezca Becerra, Domingo Muñante Pérez, Jorge Gómez Hidalgo y Reymundo Patrón Sánchez, quienes formaron el Sínodo que evaluó este trabajo.

Al Consejo Nacional de Ciencia y Tecnología y a la Organización de Estados Americanos, por el apoyo económico que hizo posible mis estudios de Maestría.

Al Gobierno de la República de Panamá, a través del Ministerio de Desarrollo Agropecuario, y el Instituto para la Formación de Recursos Humanos quienes con su apoyo, permitieron mi superación profesional.

A los profesores del Departamento, quienes por sus enseñanzas y consejos lograron mejorar mi formación profesional.

A la Srita. Alicia Buendía B., por su excelente trabajo mecanográfico.

DEDICATORIA

A la memoria del General Omar Torrijos H.
Incansable defensor de las causas de los campesinos

A mi Esposa Eva

A mis Hijos Jorgito y Ana

A mis Padres Oscar y Alicia

A mis Hermanos Pablo y Mirna

A mis Suegros Federico y Mercedes

A los Campesinos de México y Panamá

INDICE

	Pág.
INTRODUCCION	1
- Importacia del cultivo	3
- Identificación del problema	6
- Objetivos	8
- Hipótesis	9
 I. REVISION DE LITERATURA	 12
- Características generales del área de estudio	 14
- Correlación	16
- Regresión lineal	22
- Regresión múltiple	43
 II. MATERIALES Y METODOS	 48
- Información analizada	48
- Tamaño de la muestra	48
- Obtención de datos de campo	51
- Variables estudiadas	53
- Metodología	55
- Cuantificación de variables	58
- Modelos ensayados	59
 III. DISCUSION DE RESULTADOS	 75
 IV. RESUMEN Y CONCLUSIONES	 95
 V. RECOMENDACIONES	 101
 VI. BIBLIOGRAFIA CONSULTADA	 112
 VII. CUESTIONARIO UTILIZADO	 105

INDICE DE CUADROS

	Pág.
Cuadro No. 1. Observaciones ejemplificadas.	24
Cuadro No. 2. Resumen de observaciones.	28
Cuadro No. 3. Importancia relativa de los diferentes factores socioeco- nómicos según el resultado - del modelo logarítmico.	75
Cuadro No. 4. Modelo definitivo y tabla de análisis de varianza.	78
Cuadro No. 5. Estimación de la influencia relativa de seis factores so- cioeconómicos sobre el uso - de insumos mejorados.	87
Cuadro No. 6. Matriz de correlaciones par- ciales.	89
Cuadro No. 7. Coeficientes de correlación.	90
Cuadro No. 8. Aporte relativo de cada fac- tor al coeficiente de deter- minación (R^2).	91 92
Cuadro No. 9. Procedimiento de regresión por pasos para la obtención del modelo de regresión fi- nal.	93 94

INDICE DE FIGURAS

		Pág.
Figura No. 1.	Coeficiente de correlación <u>ma</u> yor que cero.	19
Figura No. 2.	Coeficiente de correlacion <u>me</u> nor que cero.	19
Figura No. 3.	Coeficientes de correlación - igual a cero.	20
Figura No. 4.	Regresión lineal.	24
Figura No. 5.	Distribución de errores del - ajuste.	34
Figura No. 6.	Recta de regresión.	39
Figura No. 7.	Mejor línea de ajuste.	45
Figura No. 8.	Opción (a)	62
	Opción (b)	62
	Opción (c)	63

INTRODUCCION

Es característico de los países dependientes la existencia de una polarización en la agricultura, en que por un lado se encuentra un sector altamente desarrollado, en donde se usan técnicas de producción moderna, que incorporan el uso de los avances técnico-científico a la producción, tales como mecanización de labores, uso de agroquímicos, semillas mejoradas, altas densidades de población y control biológico. En fin se encuentra en función una relativamente alta composición orgánica de capital en este tipo de agricultura. Esto conduce a que se obtengan altos niveles de productividad de los recursos utilizados, bajos costos unitarios y mejores condiciones de competitividad en el mercado.

Por otro lado se encuentra la presencia de una gran masa de pequeños agricultores, con una producción tradicional y dispersa, en que el nivel de uso de técnicas modernas es mínimo, lo que se traduce en bajos niveles de producción y productividad, relativamente altos costos unitarios y bajo poder competitivo en el mercado.

A pesar de los esfuerzos realizados por los gobiernos de estos países, esta situación persiste e incluso en algunos lugares tiende a hacerse más crítica.

México no escapa a esta problemática dada las condiciones específicas de su estructura económica, encontrándose que esta polarización de la agricultura mexicana cobra relieve, al observar que por un lado se amplían las áreas cultivadas de ciertos cultivos tales como sorgo, soya, cebada, que hacen un uso intensivo de tecnologías modernas, mientras por otro lado son desplazados de estas tierras cultivos básicos de la dieta mexicana como son el maíz y el frijol y se deja su producción en manos de pequeños productores, que utilizan niveles de tecnologías modernas muy bajos y que obtienen pocos excedentes destinados al mercado.

Esta serie de problemas origina que la demanda interna sea mayor que la oferta, obligando con esto a satisfacer esa demanda excedente con importaciones de granos de otros países, sobre todo de los Estados Unidos de Norteamérica, lo que se traduce en una gran fuga de divisas, problemas de balanza de pagos y lo que es peor aún, se refuerzan los mecanismos de dependencia de México frente a este país, al depender en parte de su alimentación básica del suministro foráneo de alimentos.

Este complejo panorama exige un constante estudio, a fin de ir encontrando una posible y efectiva solución a este problema.

Este trabajo se enmarca dentro de esta perspectiva, pretendiendo contribuir parcialmente en la búsqueda de una solución a tan difícil problemática.

Es necesario señalar que una solución integral y definitiva de este problema brevemente enunciado, va más allá de consideraciones meramente técnicas, interviniendo en su solución aspectos de carácter ideológico, político y económico.

Importancia del Cultivo

El cultivo de maíz está muy asociado a la cultura mexicana, este grano ha tenido una destacada importancia desde la época prehispánica y posteriormente a ella, se ha mantenido su importancia, siendo un producto de consumo básico en la dieta del mexicano.

Sin embargo a pesar de su tradicional importancia, dada la estructura y relación económica que ha vivido el país en las últimas décadas, el cultivo ha experimentado un retroceso, y el país ha pasado de exportador a importador del grano en los últimos años.

En los últimos 20 años han ocurrido cambios importantes en la estructura de consumo en México. En las

áreas rurales, durante 1959-1964 a 1979; se registraron - las siguientes variaciones: el consumo promedio de maíz bajó de 407 gramos por día/per-cápita a 324 gramos. En el área urbana el consumo de maíz se mantuvo cerca de los 200 gramos al día por persona.

Por otro lado se calcula que el 10% más pobre - de la población destina algo más del 32% de su gasto total en alimento, al maíz y sus derivados. Esto contrasta con las proporciones en el promedio nacional: poco menos del 10% a maíz y sus derivados.

Dentro de la canasta de consumo actual de la población campesina y de bajos ingresos urbana, el maíz y - frijol ocupan un 36% del gasto en alimentación.

En cuanto a la superficie cosechada de maíz en el país ésta era de 5'558,429 hectáreas en 1960 y pasó a 6'783,184 hectáreas en 1976 y los rendimientos medios na cionales para las mismas fechas pasaron de 975 a 1,181 - Kg/hectárea.

Con relación a la producción total para el mismo período fue de 5'419,782 toneladas en 1960 y de 8'017,294 toneladas en 1976.

La superficie cosechada de maíz en 1960 representó el 49% de la superficie total nacional cosechada y en 1976 ésta fue de 39%. Montañez C. (1979).

Lo anterior demuestra como ha estado disminuyendo la superficie relativa sembrada de maíz en los últimos años y el papel destacada que ha jugado el aumento en los rendimientos unitarios, como forma de hacerle frente a la creciente demanda del grano en el país.

Sin embargo en el período 1960-1970, la superficie y producción de maíz se localizó principalmente en los ejidos y los minifundios, pero los rendimientos unitarios fueron mayores y aumentaron más rápido en los predios privados mayores de 5 hectáreas que en los ejidos y minifundios, utilizando en los predios privados mayores de 5 hectáreas una mayor superficie bajo riego, híbridos, semillas mejoradas y fertilizantes. Montañez C. (1979). Por lo que se ve, la producción de maíz en el país se está concentrando en el sector de los pequeños productores. De ahí que se considere muy importante tratar de elevar los niveles productivos dentro de este sector.

Identificación del Problema

Es conocida la importancia que tiene en el aumento de la disponibilidad de alimentos básicos para la po--blación, el lograr aumentos en los rendimientos unitarios de las parcelas que se destinan a la producción de estos bienes.

Es así como se considera que a través del aumento en el uso de insumos mejorados aplicados en forma racional, se puede contribuir con este propósito. Sin embargo la experiencia indica que muchas veces los pequeños agri--cultores utilizan en forma limitada estos insumos.

En el presente trabajo se pretende conocer las relaciones que guardan los factores socioeconómicos que rodean a los agricultores y la tecnología.

Es conveniente tener cierto nivel de aproxima--ción en el conocimiento de estas relaciones ya que pueden servir de guía en la definición de políticas y programas de desarrollo agrícola, pues a menudo se encuentran pro--gramas que tratan de impulsar el sector agrícola, en los cuales no se toman en consideración este tipo de relacio--nes, propias de los agricultores y que varían según el lugar y características específicas de los productores, en--contrándose en consecuencia muchas veces con frustracio--

nes, desconfianza, fracasos y desgaste de recursos y energías, tanto por parte de los campesinos como de los técnicos y del estado.

El presente trabajo enfoca su atención a este - sector de pequeños agricultores y concretamente a pequeños productores de maíz, con el propósito de conocer con cierto grado de aproximación las relaciones que guardan los - factores socioeconómicos que afectan a los pequeños pro-- ductores y el uso de insumos mejorados, uniéndose en esta forma a los diversos trabajos realizados que tratan de obtener información acerca de cuales son los elementos que influyen en la determinación del uso de tecnológías modernas por parte de pequeños agricultores, con el fin de incrementar la disponibilidad de alimentos para la población.

Este trabajo tiene como propósito aplicar un método relativamente nuevo, para conocer relaciones entre - diferentes variables, en este caso particular, se aplicaría a variables socioeconómicas apoyándose en los avances de técnicas econométricas, con el auxilio del uso de computadoras, con lo cual se puede llegar a conocer en forma relativamente rápida qué elementos influyen en mayor grado, en un momento dado y en un área específica, en el comportamiento de los agricultores frente a la adopción de - nuevas tecnologías.

Se está consciente de que tratándose de aspectos relacionados con la conducta humana, en que intervienen factores tanto de tipo cultural, psicológico, religiosos y económicos, el empleo de este tipo de método en esta clase de análisis, puede resultar restrictivo, sin embargo se consideró que sirve de gran apoyo al investigador en las ciencias sociales y que al complementarse la interpretación de los resultados obtenidos con el apoyo de la economía y la sociología, pueden obtenerse conclusiones satisfactorias.

O b j e t i v o s

Con este estudio se van a explorar algunas de las relaciones que existen entre los factores socioeconómicos y el uso de insumos mejorados entre los pequeños productores de maíz del área de Otumba, México.

Dentro de este objetivo general existen objetivos específicos como son;

1. Probar una metodología que combine técnicas econométricas y análisis socioeconómico en el estudio del comportamiento de la conducta de los pequeños agricultores en cuanto a adopción de tecnologías modernas.

2. Determinar el grado de influencia de algunos factores socioeconómicos en la adopción de tecnologías modernas, entre los productores de maíz del área de Otumba, México.
3. Determinar que combinación de estos factores socioeconómicos influyen con mayor peso en el uso de insumos mejorados entre estos productores de maíz.
4. Generar, a partir de los resultados obtenidos, algunas recomendaciones que puedan ser utilizadas en políticas y programas de desarrollo agrícola en el área.

Hipótesis

El uso de insumos mejorados está relacionado con factores socioeconómicos de los pequeños agricultores, tales como la edad del agricultor, grado de escolaridad, tamaño de la explotación, etc.

Hipótesis de Trabajo

1. Se espera que exista una relación directa entre las cantidades de insumos mejorados utilizados por los agricultores y la asistencia a las demostraciones agrícolas (asistencia técnica).

2. Se considera que a mayor edad del agricultor menor será la cantidad de insumos mejorados que utilice.

3. Se supone que a mayor nivel de escolaridad el agricultor tenderá a usar un número mayor de insumos mejorados, ya que estará más apto para aprender a manejar técnicas más complejas y utilizarlas en forma eficiente.

4. Se espera que al tener otras fuentes de ingreso tenderá a utilizar mayor número de insumos mejorados, ya que dispondrá de mayores recursos monetarios para la adquisición de insumos lo que le permitirá ahorrar fuerza de trabajo en su parcela y dedicarla a otra actividad económica.

5. El tipo de propiedad de la parcela que explota el agricultor afectará el uso de insumos mejorados, ya que si por ejemplo tiene propiedad privada sobre la tierra tendrá más facilidades para conseguir crédito y comprar insumos, dadas las reglas del mercado.

6. Se espera una relación inversa entre el tamaño de la familia y el uso de insumos mejorados, ya que al ser más numerosa la familia habrá mayor disponibilidad de fuerza de trabajo para emplear en la parcela y en consecuencia se utilizará una menor proporción de insumos mejorados.

7. Se espera que exista una relación directa - entre el tamaño de la parcela y el número de insumos mejorados que utilice el agricultor.

8. En la medida en que la superficie sembrada de maíz aumente, se espera se utilice un mayor número de insumos mejorados.

9. Se considera que entre más jornales asalariados utilice el agricultor mayor será el número de insumos mejorados que emplee en su parcela.

I. REVISION DE LITERATURA

Características generales del área de estudio

Considerando de mucha importancia la necesidad de estudiar las relaciones que guardan los factores socioeconómicos y el uso de insumos mejorados entre pequeños - productores de maíz, se pensó en el lugar donde podía llevarse a cabo el estudio, teniendo presente que el área debía tener una alta proporción de agricultores dedicados - al cultivo del maíz y que los mismos en términos generales debían tener como característica que se trataran de pequeños agricultores (con 10 o menos hectáreas por agricultor).

Además se consideró la posibilidad de obtener - una información confiable, accesible y en cantidad suficiente.

Para definir el área de estudio ésta, se seleccionó entre las diferentes zonas homogéneas originadas de una regionalización del área de influencia del Campo - Experimental "Chapingo" perteneciente al Instituto Nacional de Investigaciones Agrícolas, (INIA), en la cual - se encuentran en acción programas tendientes a generar recomendaciones técnicas a los productores de maíz del área.

Por guardar estos requisitos se escogió el área de Otumba, en el Valle de México.

Localización:

Otumba se encuentra ubicada entre los paralelos 19°35' y 19°45' latitud norte y entre los 98°45' y 98°55' longitud oeste. El área está comprendida entre la parte norte del Municipio de Texcoco y el Municipio de San Juan Teotihuacán y desde el extremo occidental del Estado de Tlaxcala hasta el Vaso del Lago de Texcoco.

Medio físico:

En el área predominan los suelos andosoles vítricos, con textura mediana según la nomenclatura geológica.

La precipitación pluvial oscila entre 650 y 700 mm. anuales.

La temperatura media anual es de 16°C, habiéndose se registrado temperaturas máximas de 34°C y mínimas -3.9°C para el período comprendido entre los años 1961 y 1969.

La altura media sobre el nivel del mar es de - 2,260 metros y el clima predominante es clasificado como templado sub-húmedo con lluvias en verano.

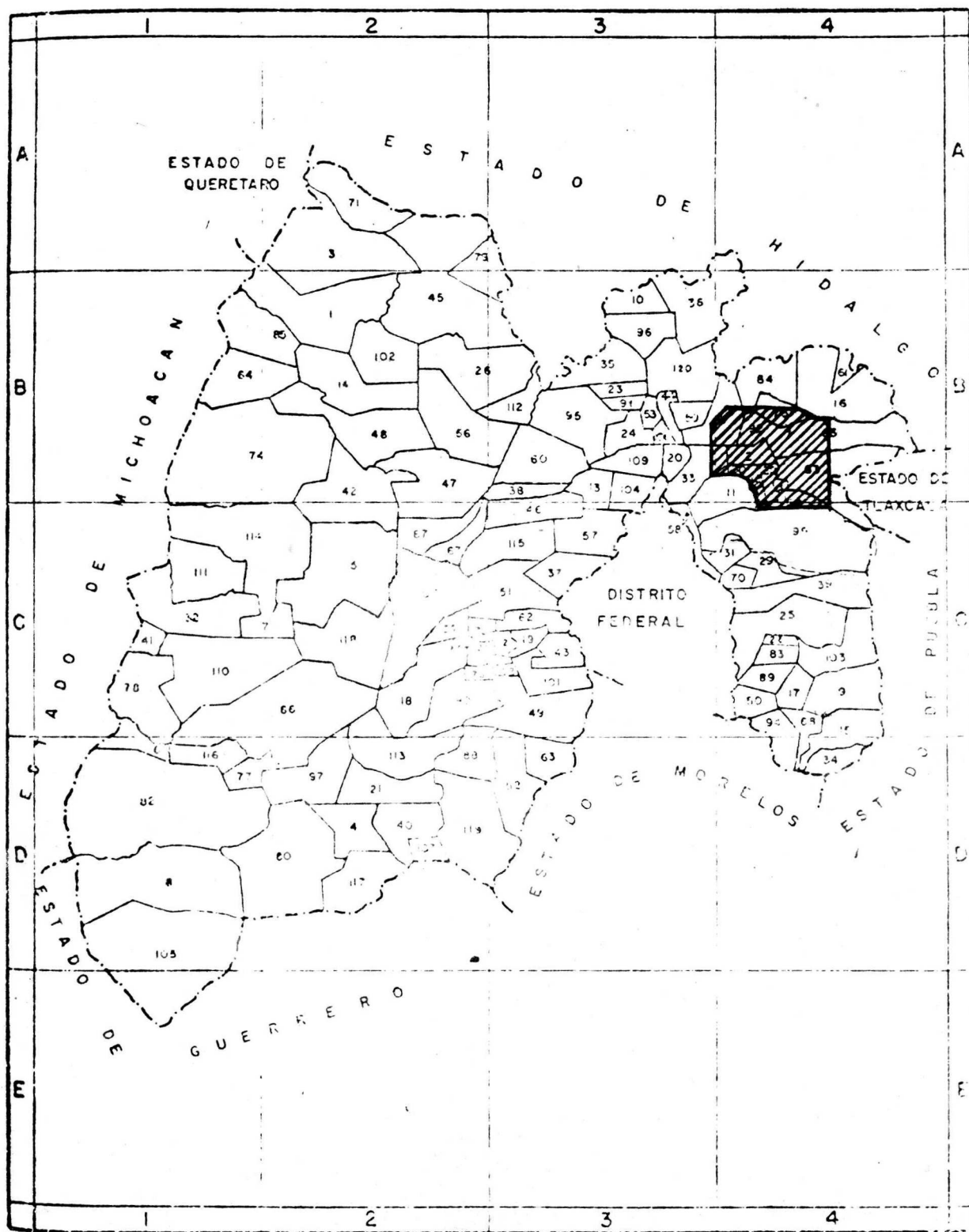
Características socioeconómicas:

El 60% de la población tiene como principal ocupación la agricultura y la ganadería; un 10% se dedica al comercio y transporte, un 7% se dedica a otros servicios, el resto de la población, o sea el 23% se dedican a otras actividades. (Censo de Población de 1970). El 65% del -- área cultivada se dedica al maíz, el 15% a cebada, el 12% frijol y el 8% que resta se siembra con otros cultivos.

El área de estudio comprende los siguientes municipios: Texcoco, Otumba, Papalotla, San Martín, Acolman, Teotihuacán, Tezoyuca, Tecamac, Chiautla, Tepetlaoxtoc y Ecatepec.

Dentro del área de estudio la principal población en cuanto al número de habitantes es San Juan Teotihuacán y de las cercanas al área de estudio están Calpulalpan (Estado de Tlaxcala) y Texcoco (Estado de México). Por tillo M. (1980).

EDO. DE MEXICO



Localización aproximada del
área muestreada

Correlación

El análisis de correlación es una técnica estadística que con mayor frecuencia se aplica a la Economía, a la agricultura y a otras ciencias. Esta técnica tiene dos aspectos: uno es como medida de la bondad de ajuste de una recta de regresión a la distribución de las observaciones, y un segundo uso es como medida de covariabilidad de dos variables X e Y.

"Debe hacerse notar que si bien independencia implica necesariamente una covarianza igual a cero, lo opuesto no es necesariamente cierto. A las variables aleatorias que tienen covarianza igual a cero se les denominan no correlacionadas. O sea que independencia excluye toda clase de relación entre las variables X e Y, mientras que no correlación excluye sólo relaciones lineales entre variables". Matus J. (1979).

Otra forma de analizar lo anterior es a través del coeficiente de correlación (ρ) de X e Y que se define como: "el cociente que resulta de dividir la covarianza entre el producto de las raíces cuadradas de las varianzas de cada una de las variables". Yamane T. (1974).

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{E(X - \mu_X)(Y - \mu_Y)}{\sqrt{E(X - \mu_X)^2} \sqrt{E(Y - \mu_Y)^2}}$$

Con $\sigma_X > 0$, $\sigma_Y > 0$, por lo que si $E\{X - E(X)\} \cdot \{Y - E(Y)\} = \text{Cov}(X, Y) = \sigma_{xy} = 0$, entonces $\rho_{xy} = 0$, es decir, el coeficiente de correlación es cero si la covarianza es igual a cero.

La covarianza es a la vez un indicador del sentido de variación que tiene una variables con respecto a otra; es decir, da información acerca de si al aumentar una variable, la otra también aumenta, o bien varían en direcciones opuestas, y da información acerca "de si ambas variables varían juntas, de la misma manera, o varían de manera muy diversa una de la otra". Yamane T. (1974).

Si la covarianza es positiva, entonces la correlación será positiva, si la covarianza es negativa, la correlación será negativa. Si el coeficiente de correlación es igual a (1), entonces Y es una función lineal de X (y viceversa). Si el coeficiente de correlación es igual a (-1), entonces la función lineal tiene pendiente negativa, es decir:

$$\sigma_{xy} > 0 \Rightarrow \rho_{xy} > 0; \quad \sigma_{xy} < 0 \Rightarrow \rho_{xy} < 0$$

$\rho_{xy} = 1 \Rightarrow Y$ es una función lineal de X y la pendiente es positiva.

$\rho_{xy} = -1 \Rightarrow Y$ es una función lineal de X (y viceversa) y la pendiente es negativa.

En otros términos se dice que cuando la covarianza de $(X, Y) < 0$; X e Y varían en direcciones opuestas; - en cambio si $\text{Cov}(X, Y) > 0$, varían en forma directa cuando la relación entre X e Y es lineal. Cuando la covarianza $= 0$ no implica que X e Y no estén relacionadas, sino - que puede ser que estén relacionadas en forma no lineal. Pero si aún de manera no lineal la covarianza resulta ser cero, se dice que X e Y son variables independientes entre sí.

A manera de ilustración veamos la representación gráfica de:

$$\rho_{xy} \text{ ó } r$$

$\rho_{xy} \text{ ó } r > 0$; implica que la pendiente de la línea es hacia arriba y a la derecha.

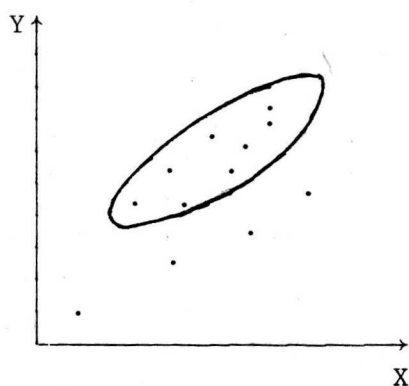


Figura 1.

$r < 0$; la línea de ajuste se inclina hacia abajo y a la derecha.

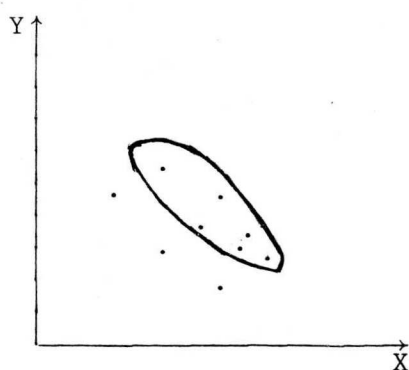


Figura 2.

$r = 0$; la línea es horizontal, es decir que X no sirve para predecir Y como regresión, y por otro lado indica que X e Y son variables no relacionadas en correlación.

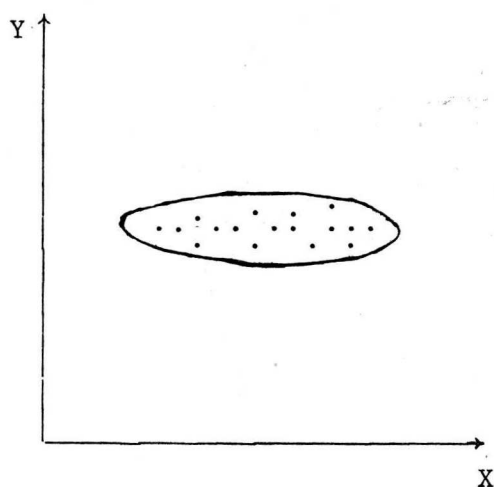


Figura 3.

El coeficiente de correlación (r) es una medida del grado de asociación entre dos variables, mientras que la covarianza no expresa suficientemente el grado de vinculación que existe entre X e Y .

Al dividir la covarianza de X e Y entre el producto de sus respectivas varianzas el coeficiente resultante queda libre de unidades de medidas, ventaja esta sobre la covarianza que es más imprecisa.

Cuando se estima el coeficiente de correlación a partir de una muestra que contiene valores de X e Y , este estimador puede obtenerse a través de la siguiente fórmula:

$$\text{Estimador de } \rho_{xy} = r = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\sqrt{[\Sigma^2 - n\bar{X}^2] [\Sigma Y^2 - n\bar{Y}^2]}}$$

a este estimador del coeficiente de correlación se le conoce como r .

$\therefore r$ = Coeficiente de correlación muestral

X = Valores que toma la variable X

Y = Valores que toma la variable Y

ΣXY = Sumatoria de los productos de los valores de X e Y .

$\bar{X}, \bar{Y}$ = Valores promedios de X e Y respectivamente

n = Número de observaciones.

El coeficiente de correlación de la muestra se puede interpretar desde dos puntos de vista. Uno es como medida de la exactitud del ajuste de la recta de regresión de la muestra a las observaciones, desde este punto de vista r se considerará en sí mismo, como un valor que mide la exactitud de ajuste de la recta de regresión de la muestra.

Otro punto de vista es el de r como un estimador de ρ_{xy} , desde este punto de vista, se puede observar que

cuando $n \rightarrow N$, la $r \rightarrow \rho_{xy}$ y así r es un estimador consistente de ρ_{xy} . Así, cuando n es grande, r gradualmente se convertirá en un estimador insesgado de ρ_{xy} .

"Hay que señalar que el análisis de correlación es técnicamente neutral, y no prueba causa y efecto. Sólo mediante una investigación más completa en la ciencias donde se aplica como por ejemplo la economía, sociología, etc., se puede llegar a alguna conclusión respecto a saber si X es o no la causa de Y ". Yamane T (1974).

Regresión Lineal

Una de las técnicas usadas con mayor frecuencia en investigación económica, en la que se trata de encontrar una relación entre dos o más variables ligadas de un modo causal es el análisis de regresión.

Cuando se trata de encontrar la relación existente entre dos variables se denomina regresión lineal simple, por ejemplo, la producción de arroz y la cantidad de fertilizante utilizado. Cuando se extiende esta técnica al caso general de más de dos variables independientes, como sería la relación entre la producción de café y la cantidad de fertilizante, altura sobre el nivel del mar, distancia entre plantas, cantidad de lluvia caída, etc., se co-

noce como regresión lineal múltiple.

Si se quiere conocer la relación entre dos variables X e Y ; se distinguen dos casos que dependen de como se pretende utilizar la relación: en un caso interesa la relación per sé. Cuando se desea conocer la relación entre altura y peso, o consumo e ingresos, interesa la relación per sé. Sin embargo, si lo que interesa es estimar la longitud de un carrete de alambres, no sería práctico desenrollar el alambre y medir su longitud. Pero podemos establecer la relación entre longitud (Y) y el peso (X) - del alambre de manera que pesando el alambre se puede estimar su longitud. En este segundo caso interesa conocer la longitud (Y), y la relación es, exclusivamente, un procedimiento para estimar la longitud.

A continuación veámos brevemente algunas condiciones en torno al problema de la regresión; para esto tomaremos el caso más simple del análisis de regresión, a través de un ejemplo específico.

Supongamos un conjunto de datos que representan el tamaño (X) y el peso (Y) de un grupo de niños según el cuadro que a continuación presentamos, en el cual aparecen 25 pares de observaciones tal como (50 pulgadas, 40 libras), (50 pulgadas, 41 libras),, (54 pulgadas, 60 libras). Estos 25 pares de observaciones comprenden la población y

y la figura siguiente es la representación de estas observaciones.

CUADRO No. 1

X, PULGADAS						E (Y/X)
50	40	41	42	43	44	42
51	41	43	44	45	46	44
52	41	44	45	48	52	46
53	43	46	47	49	55	48
54	44	46	49	51	60	50

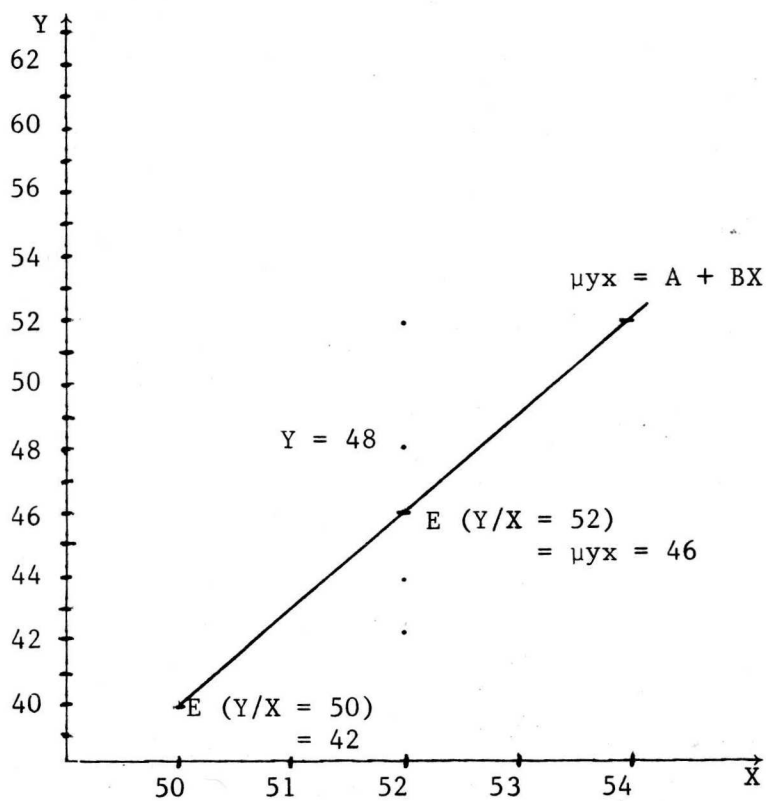


Figura 4.

Dado que los datos se agrupan de acuerdo con de terminadas alturas, tal como 50 pulgadas y 51 pulgadas y cada grupo tiene 5 pares de observaciones. Cada grupo es una subpoblación, y hay 5 subpoblaciones que corresponden a las alturas variables fijas (X). Se dice que tenemos - una colección o familia de subpoblación.

Las características de esta población es que los valores de la variable X se han fijado (variables no alea torias) mientras que los de Y son variables aleatorias.

Se supone una distribución no normal de la variable Y y la media de las Y en cada subpoblación se deno mina valores esperados de Y para las alturas dadas (X) y se escribe E (X/Y). Por ejemplo E (Y/X - 50 pulgadas =
$$\frac{40 + 41 + 42 + 43 + 44}{5} = 42 \text{ lbs.}$$

Los datos se han ajustado para que la media caiga sobre la recta que se ha trazado en la figura anterior. Si vemos la subpoblación que corresponde a X = 52 pulgs. Entonces.

$$E (Y/X = 52) = \frac{41 + 44 + 45 + 48 + 52}{5} = 46$$

La Y = 48 en esta subpoblación se desvía de --
E (Y/X) según Y - E (Y/X = 52) = 48 - 46 = 2 lbs.

Estas desviaciones se representan por ϵ que se denomina término de error o término de perturbación estocástico. Entonces $Y = 48$ se representa como:

$$Y = E(Y/X = 52) + \epsilon$$

que es una expresión general para los valores individuales de Y de la subpoblación $X = 52$. Es decir, cuando $\epsilon = -4$, $Y = E(Y/X = 52) + \epsilon = 46 - 4 = 42$.

Si $\epsilon = -3$, entonces $Y = 43$; si $\epsilon = -1$, $Y = 45$; si $\epsilon = 3$, $Y = 49$; y, cuando $\epsilon = 6$, $Y = 52$.

Los valores individuales de Y en cada una de las subpoblaciones pueden representarse de forma análoga. Es decir,

$$Y = E(Y/X = 50) + \epsilon = 42 + \epsilon$$

$$Y = E(Y/X = 51) + \epsilon = 44 + \epsilon$$

.....

$$Y = E(Y/X = 54) + \epsilon = 50 + \epsilon$$

En general, un Y particular se representa como:

$$Y = E(Y/X) + \epsilon \quad (1)$$

Como el $E(Y/X)$ cae en una recta, como se puede ver en la figura 4 se supone que la relación entre altura

y peso es lineal y expresa $E(Y/X)$ como:

$$E(Y/X) = A + BX$$

Abreviando escribiremos:

$$E(Y/X) = \mu_{yx}$$

Entonces la relación entre altura y peso se supone que es:

$$\mu_{yx} = A + BX \quad (2)$$

y ésta es la línea recta que pasa por los pesos medios - $E(Y/X)$. Los parámetros A y B se determinan por observación como:

$$\mu_{yx} = 42 + 2X$$

Origen: $X = 0$ para 50 pulgs.

En el cuadro No. 2 se resume lo expuesto hasta aquí:

CUADRO No. 2

Altura (Pulgs.)	X	E (Y/X) = μ_{yx}
50	0	42 lbs.
51	1	44
52	2	46
53	3	48
54	4	50

Las ecuaciones (1) y (2) eran:

$$(1) \quad Y = E(Y/X) + \epsilon$$

$$(2) \quad \mu_{yx} = A + BX$$

Combinando estas dos ecuaciones, se obtiene un valor individual de Y por:

$$Y = A + BX + \epsilon \quad (3)$$

Estas tres ecuaciones resumen todos los datos de la población y se denominarán modelo (o modelo de regresión). La ecuación (2) es la denominada curva de regresión de Y sobre X y muestra la relación entre los valores esperados de Y y los valores independientes X. Se llama también simplemente función de regresión. Los coeficientes A y B se denominan coeficientes de regresión de la población.

Para generalizar el modelo, primero en lugar de tener sólo 5 pares de observaciones para una altura fija, dispongamos un gran número de casos, muchos de ellos no podrán observarse pero se sabe que existe. Entonces cada subpoblación resulta muy grande.

Segundo, supóngase que los términos de perturbación ϵ son variables independientes aleatorias (o digamos que las variables aleatorias Y son independientes) y

$$E(\epsilon) = 0 \quad (4)$$

Tercero, supóngase que las desviaciones típicas (varianzas) de todas las subpoblaciones son iguales. La varianza de Y por definición es:

$$\sigma^2 = E [\bar{Y} - E(Y/X)]^2 \quad (5)$$

Se considera que todas las subpoblaciones tienen la misma varianza σ^2 .

Esta varianza se puede expresar en términos de las perturbaciones ϵ . Con (3) resulta:

$$\begin{aligned} \epsilon &= Y - (A + BX) \\ &= Y - e(Y/X) \end{aligned}$$

Sustituyendo esto en (5) tenemos:

$$\begin{aligned}\sigma^2 &= E (\epsilon)^2 \\ &= E [\epsilon - E (\epsilon)]^2 \\ &= \text{Var} (\epsilon)\end{aligned}$$

Esto es:

$$\text{Var} (Y) = \text{Var} (\epsilon) = \sigma^2 \quad (6)$$

Como ϵ representa las perturbaciones estocásticas de Y con respecto a su media $E (Y/X)$, la varianza σ^2 se denomina varianza residual. Detengámonos un momento y pensemos sobre la perturbación ϵ . En el ejemplo de alturas y pesos, se tiene, por ejemplo, una subpoblación para $X = 51$ pulgs. Hay 5 estudiantes que pesan 40, 41, 42, 43 y 44 libras, respectivamente. Encontramos que:

$$E (Y/X \text{ 50 pulgs.}) = 42 \text{ lbs.}$$

Y por ejemplo:

$$\epsilon = Y - E (Y/X) = 40 - 42 = -2 \text{ lbs.}$$

para el primer alumno

¿Qué ha producido esta perturbación de $\epsilon = -2$ libras?

Podríamos pensar en muchas causas. Primero podríamos argumentar teóricamente que si conociéramos todos los factores que afectan el peso, tales como los padres, la alimentación preferida, las condiciones de vida, etc. que son una gran cantidad de factores, y si utilizamos to dos estos factores probablemente podríamos hallar el valor esperado del peso de un estudiante exactamente. Pero como hemos elegido solamente la talla (X) para explicar el valor esperado del peso de un alumno, estamos agrupando - los efectos de todos los factores que afectan el peso y que no son la altura, en el término de error o perturbaciones cuando se agrupan tienen la propiedad de ser estocásticas.

Segundo, al medir el peso (Y) o la talla (X) pue de existir un error en la medición y este también se supo ne incluido en el término de perturbación ϵ .

Una tercera causa de la existencia de este ϵ es que la relación entre la talla (X) y el peso (Y) puede no ser exactamente lineal como hemos supuesto.

Y si esta relación que hemos supuesto no es exac tamente el efecto de la medición de este error de medición también estará incluido en ϵ .

Para el segundo estudiante encontramos que las perturbaciones son:

$$\epsilon = Y - E(Y/X) = 41 - 42 = -1 \text{ lb.}$$

que también se interpreta de igual modo.

Supongamos ahora que la perturbación $\epsilon = -2$ lbs. del primer estudiante y la $\epsilon = -1$ lb. del segundo no están relacionadas; es decir, que son independientes. O bien - podemos decir que los pesos del primer y segundo estudiante son independientes.

Las generalizaciones anteriores pueden ser visualizadas en la figura No. 4. Para una determinada altura (digamos $X = 50$ pulgadas) existen muchos pesos posibles Y . Supongamos que la mayoría de los pesos se distribuyen alrededor de $E(Y/X = 50) = 42$ libras y las restantes se agruparán hacia este punto. La dispersión de los valores individuales Y , alrededor de $E(Y/X = 50) = 42$ tienen una varianza σ^2 . Esto es la dispersión de las perturbaciones ϵ alrededor de $E(Y/X = 50)$ tienen una varianza σ^2 . En la figura se ha considerado que $\sigma = 2$.

Resumiendo tenemos una familia de subpoblaciones de Y (sean pesos) que son las variables aleatorias que corresponden a determinadas X (sean, alturas) que son varia

bles fijas. La distribución de las Y de cada subpoblación se supone distribuída con una media $E(Y/X)$ y una varian-za σ^2 . Entonces la función de regresión de Y sobre X se define como:

$$E(Y/X) = A + BX$$

en la que A y B son los parámetros de regresión de la población (o coeficientes). Los valores individuales de Y se obtienen según:

$$Y = A + BX + \epsilon$$

Los supuestos son:

- a) ϵ es una variable independiente aleatoria y la distribución no está especificada
- b) $E(\epsilon) = 0$ (7)
- c) Varianza (σ) = Var (Y) = σ^2 (8)
- d) X es un conjunto de números fijos.

El supuesto c) de que tiene varianza σ^2 constante se refiere a la homocedasticidad. Las ecuaciones (7) y (8) y estos supuestos proporcionan un modelo de regresión.

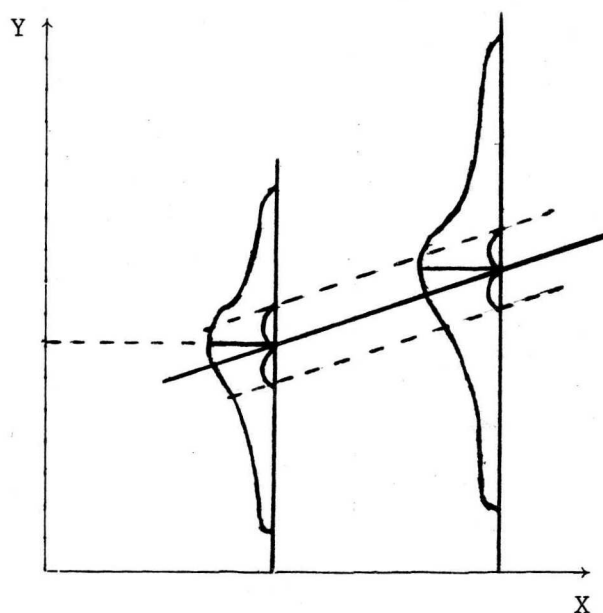


Figura 5.

El modelo de regresión era:

$$Y = E(Y/X) + \epsilon \quad (1)$$

$$\mu_{yx} = A + BX \quad (2)$$

$$Y = A + BX + \epsilon \quad (3)$$

Viendo ahora el problema de la regresión desde un punto de vista estadístico en el que se pone la mayor atención en la estimación de μ_{yx} .

Suponiendo que los valores esperados de Y — es decir, $\mu_{yx} = E(Y/X)$ — caen sobre una línea recta y por

lo tanto se supone una ecuación lineal (2). ¿Cuál es el motivo que justifica la hipótesis de que μ_{yx} cae sobre una línea recta?.

Uno de ellos es que, en la práctica, cuando la amplitud de variación de X (sea, peso o ingreso, etc.) se aproxima a una línea recta. Este procedimiento de aproximación lineal se utiliza siempre en la vida diaria.

Esta aproximación lineal se utiliza por su simplicidad para la interpretación, para el tratamiento matemático y para los empleos prácticos.

Un punto importante es comprender que el modelo expresado por la ecuación (2) es una hipótesis y no un conocimiento a priori.

El problema que se plantea consiste en estimar μ_{yx} empleando la relación entre los valores esperados de Y y X . También se puede decir, que se desea encontrar el valor de $E(Y/X)$ para un determinado X . Según la ecuación (2), es evidente que si se pueden determinar los coeficientes de regresión de la población A y B se puede buscar el valor esperado de Y para valores dados de X . Por ejemplo, si se conoce A y B para la función de regresión de altura y peso, se puede encontrar el peso que le corresponde a una altura dada. En el ejemplo simplificado de la altura

y peso, las A y B se deducen buscando el $E(Y/X)$. Esto fue posible porque sólo teníamos cinco valores de X y muchos (teóricamente una cantidad infinita) valores de Y . ¿Qué debemos hacer?.

Como se puede entender ahora, se toma una muestra de estas observaciones y se estima A y B .

Las estimaciones se representan por a y b y se denominan coeficientes de regresión muestrales. Luego el estimador de $A + BX$ se transforma en $a + bx$.

Dado que $A + B = \mu_{yx}$, $a + bx$ puede llamarse estimador de μ_{yx} .

Por lo tanto vemos que el problema estadístico es seleccionar una muestra y a partir de esta muestra estimar los coeficientes de regresión de la población A y B . Como veremos los estimadores de A y B que hemos denominado a y b resultarán ser formas lineales de los elementos de la muestra. Expliquemos esto utilizando la media de la muestra.

Sabemos que cuando deseamos estimar la media de la población μ seleccionamos una muestra y utilizamos la media de la muestra $\bar{X}$ como un estimador de μ .

Un estimador debe tener varias características, entre ellas hay 3 que son muy importantes:

- a) Debe ser insesgado; esto es que la esperanza del estimador sea igual al parámetro; o sea $E(\bar{X}) = \mu$

En términos de nuestro ejemplo, se desea que $a + bx$ sea un estimador insesgado de μ_{yx} , - lo que significa que $E(a + bx) = A + BX$ (8) y esto será cierto evidentemente cuando:

$$E(a) = A, \quad E(b) = B$$

Cuando la ecuación (8) se cumple, decimos - que $a + bx$ es un estimador lineal insesgado μ_{yx} .

- b) Debe ser consistente; es decir que en el límite en que el tamaño de la muestra se acerca a la población, o sea $\lim_{n \rightarrow N} \bar{X} = \mu$, o bien $\lim_{n \rightarrow N} a + bx = A + BX$.
- c) Debe ser de mínima varianza; o sea que se encuentre lo más cerca posible del parámetro de población que se estima. Y según el teorema del Límite Central que dice que la media muestral, por ser una variable aleatoria se distribuye normalmente con media μ y varianza σ^2/η , es decir $\bar{X}_\eta \sim N(\mu, \sigma^2/\eta)$. A medida que η crece el teorema es más válido, o sea cuando $\eta \rightarrow N$.

En nuestro caso se desea que $a + bx$ sea el estimador de varianza mínima de $A + BX$ (o, de μ_{yx}). Veamos que significa esto.

Supongamos que se elige una muestra de tamaño n , y que por un procedimiento l , se encuentra $a + bx$, que es una estimación de $A + BX$. Se selecciona una segunda muestra y, con el mismo procedimiento l , se encuentra una segunda estimación de $a' + b'x$. Se elige otra muestra y se halla una tercera estimación $a'' + b''x$. Como se ve este proceso - puede repetirse una gran cantidad de veces y se encuentran otras muchas estimaciones - $a + bx$.

Estas estimaciones se obtuvieron por el procedimiento l ; pero se pueden seguir otros - procedimientos diferentes y se encontrará - otro grupo de estimaciones, digamos $c + dx$ que también se agruparán alrededor de $A + BX$.

Si se supone que $a + bx$ es el mejor estima-dor o sea de varianza mínima, lo que se está afirmando es que la dispersión de $a + bx$ alrededor de $A + BX$ será menor que la de cualquier otro tipo de estimador $m + kx$ obtenido por cualquier otro procedimiento.

El encontrar el estimador $a + bx$ de varianza mínima es un problema estadístico. Para poblaciones tipo I (población en que la distribución de Y no se especifica) y tipo II (Y está distribuido normalmente), el procedimien

to de estimación para obtener un estimador de varianza mínima se utiliza el método de los mínimos cuadrados.

Para una población tipo I, el procedimiento de estimación es el Teorema de Gaus-Markov, mientras que para la población de tipo II es el método de máxima verosimilitud.

El método de los mínimos cuadrados minimiza la suma de las desviaciones al cuadrado de los valores observados de Y con respecto a los predichos. Martínes A. (1974).

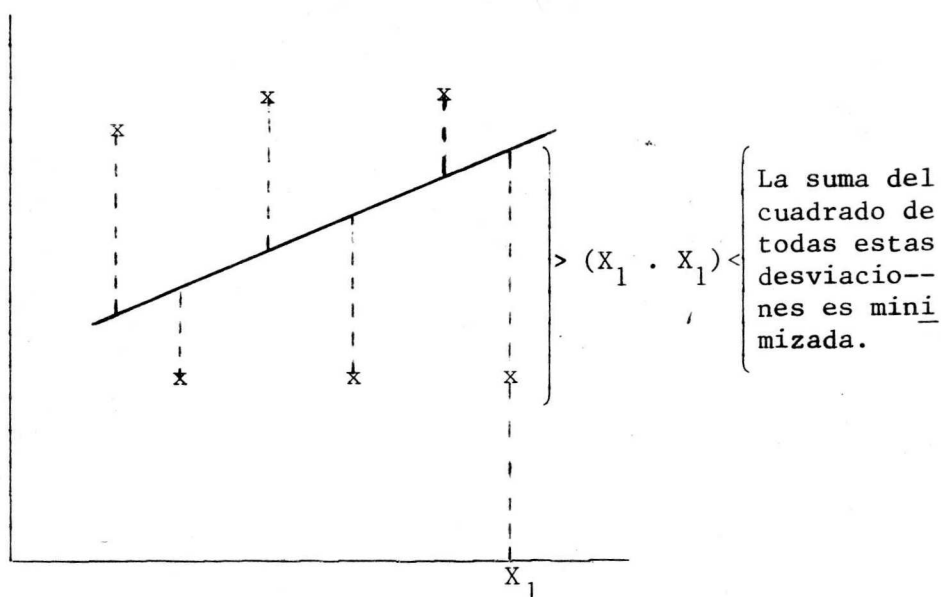


Figura 6.

Para aplicar el método de mínimos cuadrados a un modelo de regresión lineal simple como el siguiente por ejemplo:

$$Y_i = \alpha + \beta X_i + e_i$$

$$E(e_i) = 0, \quad E(e_i^2) = \sigma^2, \quad E(e_i e_{i'}) = 0$$

$$i \neq i', \quad i, i' = 1, 2, \dots, n \quad (9)$$

Se forma Q , la suma de cuadrados de los errores:

$$Q = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$$

Al derivar parcialmente Q con respecto a α y β , e igualando a cero, resulta:

$$\begin{aligned} \frac{\partial Q}{\partial \alpha} &= \sum_{i=1}^n \frac{\partial}{\partial \alpha} (Y_i - \alpha - \beta X_i)^2 \\ &= \sum_{i=1}^n 2 (Y_i - \alpha - \beta X_i) (-1) \quad (10) \\ &= 0, \end{aligned}$$

$$\begin{aligned} \frac{\partial Q}{\partial \beta} &= \sum_{i=1}^n \frac{\partial}{\partial \beta} (Y_i - \alpha - \beta X_i)^2 \\ &= \sum_{i=1}^n 2 (Y_i - \alpha - \beta X_i) (-X_i) \quad (11) \\ &= 0 \end{aligned}$$

A partir de las fórmulas (10) y (11) se obtiene:

$$(-2) \sum_{i=1}^n (Y_i - \alpha - \beta X_i) = 0, \quad (12)$$

$$(-2) \sum_{i=1}^n (Y_i - \alpha - \beta X_i) (X_i) = 0 \quad (13)$$

(12) implica:

$$\sum_{i=1}^n Y_i - \sum_{i=1}^n \alpha - \sum_{i=1}^n \beta X_i = 0, \quad \delta$$

$$n\alpha + \beta \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i \quad (14)$$

A partir de la fórmula (13) se obtiene:

$$\sum_{i=1}^n (X_i Y_i - \alpha X_i - \beta X_i^2) = 0,$$

Por lo tanto;

$$\alpha \sum_{i=1}^n X_i + \beta \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i \quad (15)$$

A las ecuaciones (14) y (15) se les llaman ecuaciones normales y en el sistema de ecuaciones normales cada expresión tiene en su lado izquierdo términos con los parámetros desconocidos como factores y en los extremos de rechos aparecen términos independientes de los parámetros.
Martínez A. (1974).

Si resolvemos este sistema para α , resulta:

$$\alpha = \bar{Y} - \beta \bar{X} \quad (16)$$

$$\alpha = \frac{\sum_i X_i Y_i - \beta \sum_i X_i^2}{n \bar{X}} \quad (17)$$

Donde:

$$\bar{X} = (\sum_{i=1}^n X_i) / n \quad \text{y} \quad \bar{Y} = (\sum_{i=1}^n Y_i) / n$$

Si igualamos las expresiones contenidas en (16) y (17) y resolvemos para β , resulta:

$$\bar{Y} - \beta \bar{X} = \frac{\sum_i X_i Y_i - \beta \sum_i X_i^2}{n \bar{X}}$$

Por consiguiente:

$$\hat{\beta} = \frac{\sum_i X_i Y_i - n \bar{X} \bar{Y}}{\sum_i X_i^2 - n \bar{X}^2} = \frac{\sum_i (X_i - \bar{X}) Y_i}{\sum_i (X_i - \bar{X})^2}, \quad (18)$$

aquí $\hat{\beta}$ es el estimador mínimo cuadrático de β , y si sustituimos β por su estimador en la fórmula (16) queda así:

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X},$$

aquí $\hat{\alpha}$ es el estimador mínimo cuadrático de α . Martínez A. (1974).

Regresión múltiple

La teoría de la regresión simple anteriormente expuesta puede generalizarse a un conjunto de variables aleatorias observables $Y_1, Y_2, \dots, Y_n$ y su valor esperado puede expresarse como una función lineal conocida de p parámetros desconocidos. O sea que:

$$E(Y_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip},$$

$$i = 1, 2, \dots, n, \quad n \geq p \quad (19)$$

Si cada Y_i , $i = 1, 2, \dots, n$, es la suma de su valor esperado dado por (20) más un término aleatorio de error e_i , o en símbolos.

$$Y_i = E(Y_i) + e_i, \quad i = 1, 2, \dots, n, \quad (20)$$

Donde:

$$E(e_i) = 0, \quad E(e_i^2) = \sigma^2, \quad E(e_i e_{i'}) = 0$$

$$i \neq i', \quad i, i' = 1, 2, \dots, n,$$

La fórmula (20) se conoce como el "modelo general de las hipótesis lineales". Martínez A. (1974).

Las n relaciones de la fórmula (20) pueden expresarse en una forma compacta empleando notación matricial. Si escribimos:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1P} \\ X_{21} & X_{22} & \dots & X_{2P} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & & X_{nP} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_P \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}. \quad \text{Martínez A. (1974).}$$

Si r^2 o coeficientes de determinación, o $\hat{\beta}$; es igual a $\frac{[\sum (X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}$;

La suma de cuadrados de la regresión es:

$$\frac{[\sum (X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum (X_i - \bar{X})^2}$$

Por lo tanto se puede decir que r^2 es la proporción de la suma de cuadrados de Y atribuibles a X (más allá de lo cual podemos atribuírselo a $\hat{Y}$ como predictor de Y).

Poniendo esto en forma de diagrama:

	S.C. de la media	Explicados por
S.C.		
Totales de	S.C. de la regresión	Sin explicación, a menos que usemos la información dada por la X.
	S.C. de los residuos	

Por consiguiente:

$$\text{Proporción explicada por la X} = \frac{\text{S.C. Regresión}}{\text{S.C. Regresión} + \text{S.C. Residuo}}$$

Ilustrado gráficamente:

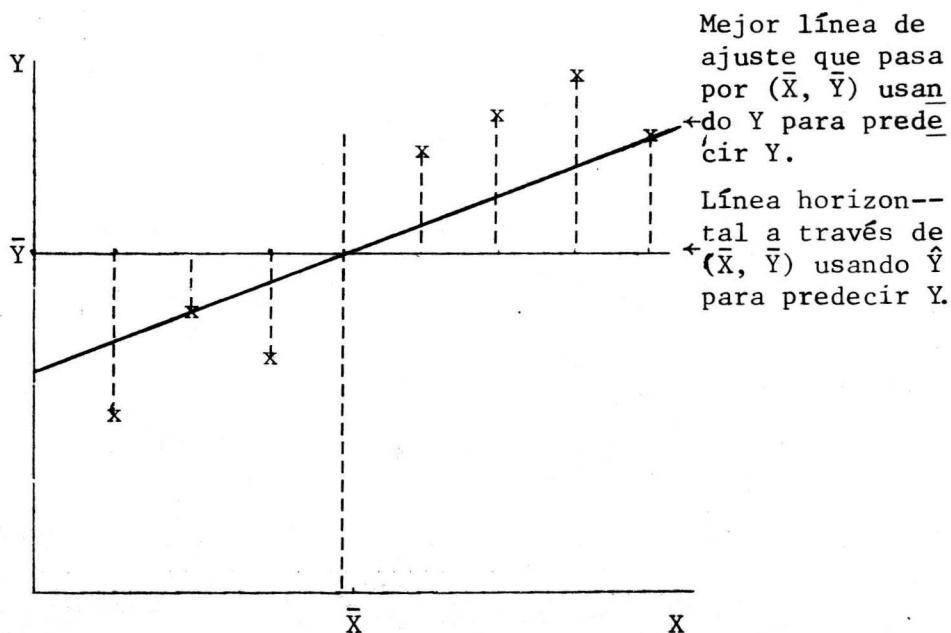


Figura 7.

La suma de las distancias es la S.C. Regresión + S.C. Residuos, o bien la S.C. Totales - S.C. de la media.

O sea;

$$r^2 = \frac{\text{S.C. Regresión}}{\text{S.C. Regresión} + \text{S.C. Residuos}}$$

Todas las S.C. deben ser positivas. R^2 es una medida más útil de la significación de la relación entre X e Y que r.

Según lo anterior los valores que pueden tomar r^2 son: a) para el caso en que la S.C. de los residuos sea = 0; $r^2 = 1$; b) si la S.C. de los residuos aumenta, r^2 tiende a cero; por tanto r^2 debe estar comprendida entre los valores; $0 \leq r^2 \leq 1$.

De ahí que r^2 está cerca de 1 cuando la S.C. de la regresión es mucho más grande que la S.C. de los residuos.

Si $r^2 = 1$ los datos (observaciones) caen exactamente en la línea de regresión y en consecuencia la S.C. de los residuos es = 0. Consecuentemente para r los valores serán;

$$-1 \leq r \leq 1$$

Hay que recordar que r ó r^2 solamente miden la significación de la relación lineal entre X e Y. En este sentido X e Y pueden estar perfectamente relacionados por una función no lineal y sin embargo r^2 puede ser pequeño.

De ahí que sea posible tener r^2 pequeño cuando:

- a) No sea una relación lineal
- b) Se tengan observaciones demasiado dispersas; es decir, X e Y no están relacionadas.

II. MATERIALES Y METODOS

Información analizada

Para realizar este trabajo se usaron los datos de una encuesta levantada en el área de estudio, efectuada por Marcos Portillo en el año de 1977 (7).

Se analizaron 200 cuestionarios, donde aparecen datos socioeconómicos de los agricultores, tecnología utilizada, uso de la tierra, datos de la explotación y mano de obra e instrumentos empleados.

Los agricultores entrevistados se seleccionaron en parte de los asistentes a demostraciones agrícolas del campo experimental "Chapingo" que tenían sus parcelas ubicadas dentro del área de estudio.

Otra parte de los entrevistados se seleccionaron del área geográfica de Otumba, utilizando para esto un mapa de suelos del DETENAL.

Tamaño de la muestra

1. Se utilizó una premuestra formada por 72 - cuestionarios.

2. Para obtener una precisión de 10% con respecto al promedio y una confiabilidad del 95%, se calculó el número de cuestionarios que debían levantarse.

3. Se utilizó el muestreo simple aleatoria cuya fórmula es:

$$N = \frac{No}{1 + (No/N)}$$

Donde:

$$No = \frac{S^2 T^2}{d^2}$$

Al ser "No" muy pequeño en relación con la población total (N), resulta ser (No/N) una fracción despreciable. Por tanto se utilizó la fórmula:

$$N = \frac{S^2 T^2}{d^2}$$

Donde:

T = Valor dado por la tabla "t" de Student, a un 95% de la probabilidad

S² = Varianza estimada

N = Tamaño de la muestra

d = Presición deseada (10% con respecto al va
lor promedio.

Así:

$$s^2 = \frac{\sum_{i=1}^{No} (X_i - \bar{X})^2}{No}$$

X_i = Valor de cada una de las observaciones

$\bar{X}$ = Valor promedio de las observaciones

No = Tamaño de la premuestra.

Por ejemplo para la superficie total en explotaci
ción se calculó así:

$$s^2 = 1.62$$

$$T_{.05} = 1.645$$

$$\bar{X} = 1.65$$

$$d = .165$$

Sustituyendo:

$$N = \frac{s^2 T^2}{d^2} = \frac{(1.62) (1.645)^2}{(.165)^2} = 162$$

De esta manera se calculó para otras variables
el tamaño de muestra que en todos los casos resultó menor
a 162.

Obtención de datos de campo

1. Se dividió el área de estudio en segmentos cuadrados los cuales se numeraron.
2. Se excluyeron aquellos formados por áreas no agrícolas.
3. Aleatoriamente se seleccionaron los segmentos en los cuales debía buscarse una parcela sembrada con maíz.
4. Se localizaba al propietario y se le entrevistaba.

Este procedimiento de muestreo enunciado se denomina "marco-área de segmento abierto", y con este método se obtuvieron 178 agricultores seleccionados.

Con relación a los entrevistados que habían asistido a demostraciones agrícolas, se seleccionaron al azar, con ayuda de la tabla de números aleatorios, aprovechando que la relación estaba numerada, un total de 25 agricultores.

Dadas algunas contradicciones en datos básicos se procedió a eliminar algunos cuestionarios y finalmente

quedaron 200, de los cuales 178 fueron de los seleccionados al azar y 22 de los asistentes a demostraciones agrícolas. Ninguno de los 178 ha asistido a demostraciones agrícolas.

Como se puede apreciar este número de cuestionarios totales es superior al necesario, según lo indican los cálculos para el tamaño de la muestra.

Identificación de las variables

I. Tecnológicas

- Y = Rendimientos de maíz en toneladas por hectárea
- X₁ = Superficie en Ha. cultivada
- X₂ = Clase de suelo (fertilidad, textura, drenaje)
- X₃ = Tipo de tracción (mecánica o animal)
- X₄ = Ton/Ha. de mejoradores del suelo (abonos orgánicos y enmiendas)
- X₅ = Epoca de aplicación de mejoradores del suelo (días antes de la siembra)
- X₆ = Manejo de suelos (rotación de cultivos, conservación de suelos)
- X₇ = Tipo de semilla (mejorada o criolla)
- X₈ = Kgs. de semilla por Ha.
- X₉ = Kgs. de N-P-K por Ha.
- X₁₀ = Epocas de aplicación del fertilizante

- X_{11} = Forma de aplicación del fertilizante (incorporado, banda o voleado)
- X_{12} = Kgs/Ha. de herbicida
- X_{13} = Etapa del cultivo en que se aplicó el herbicida
- X_{14} = Número de aplicaciones de herbicida
- X_{15} = Litros/Ha. de insecticida
- X_{16} = Número de aplicaciones de insecticida
- X_{17} = Kgs/Ha. de fungicida
- X_{18} = Número de aplicaciones de fungicidas
- X_{19} = Suministro de agua (riego, punteado, temporal)
- X_{20} = Número de riegos
- X_{21} = Número de insumos utilizados
- X_{22} = Infraestructura existente
- X_{23} = Cultivos asociados o intercalados
- X_{24} = Siniestros
- X_{25} = Cantidad de mano de obra por Ha. (días-hombre)
- X_{26} = Producción para el mercado
- X_{27} = Producción para autoconsumo
- X_{28} = Producción para otros usos (alimentación animal).

II. Socioeconómicos

- X_{27} = Forma de propiedad (ejidal, privada, alquilada)
- X_{28} = Tamaño en Ha. de la explotación
- X_{29} = Número de miembros de la familia
- X_{30} = Edad del jefe de la familia

- X₃₁ = Año de escolaridad
- X₃₂ = Número de miembros de la familia que trabajan tiempo completo en la explotación
- X₃₃ = Número de miembros de la familia que trabajan tiempo parcial en la explotación
- X₃₄ = Fuentes de ingreso (maíz, otros cultivos, ganadería, comercio, asalariado)
- X₃₅ = Número de asalariados que contrata en la explotación
- X₃₆ = Salario \$/día a nivel predial
- X₃₇ = Número de miembros de la familia que trabajan extrafinca
- X₃₈ = Salario \$/día fuera del sector agrícola
- X₃₉ = Valor de la producción
- X₄₀ = Valor de la producción vendida
- X₄₁ = Valor de la producción para autoconsumo
- X₄₂ = Valor de la producción para otros usos
- X₄₃ = Distancia en Kms. de los centros de mercado
- X₄₄ = Precio de los insumos
- X₄₅ = Precio del maíz pagado al productor
- X₄₆ = Prácticas tecnológicas que podrían modificarse para aumentar los rendimientos de manera más económica
- X₄₇ = Número de visitas técnico-agrícola durante el ciclo de cultivo
- X₄₈ = Número de demostraciones y días de campo asistidos durante el ciclo del cultivo
- X₄₉ = Financiamiento del Banco de Desarrollo Rural (para todas las actividades de cultivo o parte de ellas).

Metodología

En este estudio se utilizaron básicamente matrices de correlaciones parciales y regresión múltiple los -
cuales aplicados en forma complementaria forman la metodología básica de este estudio.

Por otro lado en la interpretación de los resultados se utilizará además de la interpretación econométrica el análisis socioeconómico como complemento a la metodología básica, ya que se trata de interpretar el comportamiento de la conducta humana, como respuesta a factores socioeconómicos.

Matriz de correlaciones parciales

Este método sirvió para conocer las variables -
que debían ser incluidas en el modelo empírico de regresión, a través de los índices de correlación entre las variables independientes y la variable dependiente.

Así como también se vieron los índices de correlación entre las variables independientes para conocer --
aquellas con altos índices de correlación y así evitar problemas de multicolinealidad.

Se utilizaron como variables dependientes el número de insumos mejorados (INMEJ) y el uso de fertilizantes (Fert), con los cuales se quería medir la adopción de tecnologías modernas.

Se utilizó Fertilizantes (Fert) en algunos modelos ya que se observó que uno de los insumos mejorados más utilizados por los agricultores era el fertilizante.

En esta forma se esperaba que el coeficiente de regresión estimado para cada variable diera el aporte con que contribuye cada una de ellas al número de insumos mejorados.

Con el propósito de mejorar el nivel predictivo de los modelos elaborados, se emplearon efectos cuadráticos de los factores socioeconómicos, así como también formas logarítmicas para conocer directamente las elasticidades de cada uno de estos factores.

También utilizamos el modelo de regresión por -pasos (Stepwise), a fin de comparar con el modelo de regresión múltiple los resultados. Este procedimiento utiliza básicamente la prueba de F para cada parámetro estimado, mientras que el modelo general de regresión lineal selecciona las variables más significativas con la prueba "t", y a veces no suelen coincidir las variables seleccionadas con un procedimiento y otro.

El procedimiento de regresión por pasos da a conocer información sobre la aportación que cada una de las variables seleccionadas hace al coeficiente de determinación o R^2 .

Se ensayaron posteriormente varios modelos con las variables más altamente correlacionadas con el número de insumos mejorados ó con el uso de fertilizantes, obtenidos a través de la matriz de correlaciones parciales y con las variables más altamente significativas utilizando la "t" de Student en el modelo de regresión completo (que contenía el conjunto de variables en estudio), se formaron modelos para ser ensayados.

En varios casos las variables obtenidas por un criterio de selección no coincidían nuevamente con las obtenidas con el otro criterio. Se procedió en consecuencia a cuantificar en forma diferente estas variables, a estratificarlas agrupando valores según características prede-terminadas de cada una de ellas; también se generaron nuevas variables relacionando variables entre sí.

Con estas nuevas variables surgidas en la forma antes expuesta, se procedió a probar nuevos modelos reducidos, incluyendo además algunas que no habían sido seleccionadas por las primeras pruebas realizadas.

Cuantificación de variables

En términos estadísticos una variable es una característica que varía aleatoriamente de un individuo a otro de una población. Es decir lo contrario de una característica constante de los individuos de una población. A través de las variables se puede distinguir o describir los elementos componentes de una población, de ahí que las variables (aleatorias) pueden referirse a atributos cualitativos o cuantitativos de los componentes de una población. En el primer caso se tienen variables cualitativas y en el segundo variables cuantitativas.

Las variables cualitativas o conceptuales son comunmente no numéricas, por ejemplo: nacionalidad, color, raza, etc.

Las cuantitativas resultan de una medida numérica: altura, peso, ingreso, número de veces que sale el 3 al lanzar en un dado, etc.

Las variables cuantitativas o numéricas a la vez pueden ser contínuas y discontínuas (discretas).

Las variables contínuas son aquellas que pueden tomar cualquier valor dentro de un rango determinado, o sea que pueden tomar cualquier valor dentro de la recta de números reales (altura, peso, etc.).

Las variables discretas no pueden tener valores intermedios, sólo pueden tomar valores enteros (número de ejidos por Estado).

Cuando se trata de variables cualitativas, en algunos casos es posible cuantificarlas de tal forma que a cada opción que presenten puede asignársele un número y, en algunos casos este número puede tratarse como si fuera un valor que puede tomar la variable a fin de poder procesarlo en un modelo de regresión. La forma de cuantificar estas variables debe estar en concordancia con ciertos supuestos, mismos que se tratan de comprobar con los resultados obtenidos en el modelo de regresión.

En ciertos casos podemos tener variables continuas pero por falta de información no se pueden trabajar así y pueden pasar a trabajarse como variables discontinuas. También por el tipo de información que disponemos de una variable, ésta puede convertirse en variable de tipo conceptual. Cuando se trata de variables conceptuales que pueden cuantificarse en base a sólo dos alternativas, se les conoce a estas como variables dicotómicas.

Ampliando el tratamiento que se le da a las variables dicotómicas, a variables que presentan más de dos opciones podemos procesar algunas variables conceptuales, teniendo presente que los resultados numéricos obtenidos

tengan alguna interpretación lógica y generen información útil para el trabajo en proceso.

El razonamiento teórico que se siguió para procesar variables conceptuales con más de dos opciones fue el siguiente:

En un modelo de la forma $Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + e$, b_i es el producto marginal del insumo X_i ; si X_i es una variable dicotómica y b_i es estadísticamente significativa, se dice que (b_i) es la magnitud en que cambia el producto "Y" cuando se pasa de la opción 1 a la opción 2, en la variable dicotómica, el producto marginal - así calculado es válido sólo para el cambio en X_i que va de 1 a 2 (los únicos valores que tiene la variable independiente).

Este tratamiento ampliándolo a una variable con 3, opciones donde sus valores son cuantificados con 1, 2 y 3, (b_i) sería el producto marginal al pasar de 1 a 2 e - igual al pasar de 2 a 3 que en algunos casos sería difícil de esperar (aunque cabe la posibilidad).

Si se toma como ejemplo el recurso agua con tres opciones 1 = temporal, 2 = humedad y 3 = riego, una hipótesis que se antoja en este caso es, que el paso de humedad a riego (de la opción 2 a la 3) ocasione un incremen-

to en la producción mayor que el que se tiene al pasar de temporal a humedad (de la opción 1 a la opción 2). En este caso el valor de b_i no resuelve el problema planteado - por la hipótesis salvo que se calcule un valor para un término cuadrático de X_i de tal manera que el modelo estimado comprendería los términos siguientes:

$$Y = a + b_i X_i + b'_i X_i^2$$

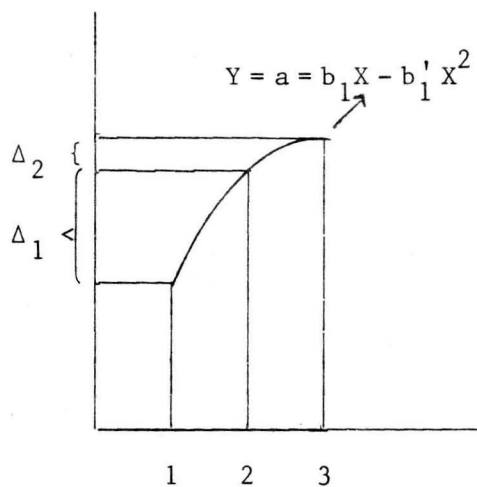
donde i representa una variable del modelo que contiene n variables en estudio.

El incremento en Y que se tiene de pasar de una opción a otra ($\frac{dy}{dx_i}$) sería igual a la expresión $b_i + 2 b'_i X_i$ lo que implica que tendrá valores diferentes, dependiendo de los valores que se tome X_i ; en otras palabras el incremento en el producto (Y) que se tiene al pasar de un valor 1 a un valor 2 en X_i , será diferente al incremento que se tiene al pasar de la opción 2 a la opción 3 en una magnitud dada por la expresión $2 b'_i X_i$ que a su vez viene dada por el término cuadrático de la variable X_i en el modelo de regresión.

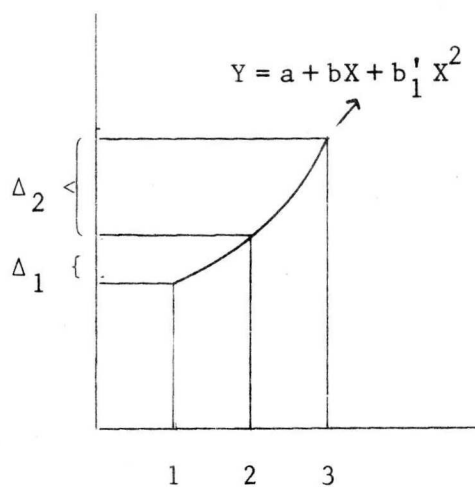
Si el parámetro b'_i es negativo, el producto marginal b_i dado por el modelo lineal se verá disminuido, lo que significará que el paso de la opción 1 a la 2 ocasiona un cambio en el producto mayor que el de pasar de la - opción 2 a la opción 3; si el parámetro b'_i es positivo im

plica que el paso de la opción 1 a la opción 2 es menor - que le que se tiene al pasar de 2 a 3 y por último, si el parámetro b_i^1 tienda a cero, el valor del producto marginal de X_i dado por la expresión $b_i + 2 b_i X_i$ será igual a b_1 que es el parámetro dado por el modelo lineal e implicaría que el incremento que se tiene al pasar de 1 a 2 es igual al que se tiene al pasar de la opción 2 a la opción 3 y por lo tanto podría utilizarse en este último caso un modelo que no incluya el término cuadrático de la variable en cuestión.

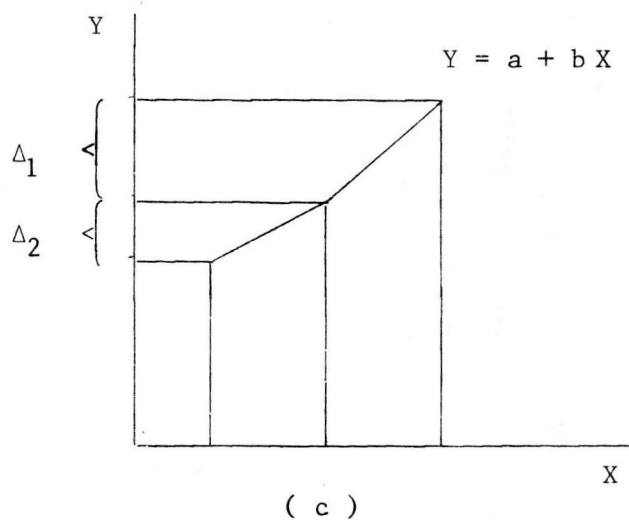
La representación gráfica de la anterior exposición es la siguiente:



(a)



(b)



Nótese las diferencias en incrementos Δ_1 y Δ_2 correspondientes a incrementos en X en las tres diferentes funciones, Δ_1 es mayor que Δ_2 , Δ_1 es menor que Δ_2 y $\Delta_1 = \Delta_2$ en (a), (b) y (c) respectivamente.

Es importantes recalcar que no todas las variables cualitativas pueden representarse de manera numérica y que no siempre es posible trabajarlas como si fueran variables cuantitativas.

Considerando lo anterior, se procedió a cuantificar cada una de las variables en observación, así como también las nuevas variables que se generaron a partir de las originales.

Así para:

$A_I = 1$, si asistió a campos experimentales o charlas de asistencia técnica.

$A_I = 2$; no asistió.

Edad = 1; si A_3 es < 35 años.

Edad = 2; si A_3 está comprendido entre 35 y 50 años.

Edad = 3; si A_3 es > de 50 años.

Se consideró que un agricultor menor de 35 años es joven y está más anuente y capacitado para utilizar insumos mejorados en su parcela.

El agricultor comprendido entre los 35 y 50 años se le consideró como "agricultor maduro" y su disponibilidad a incorporar insumos mejorados en la parcela iba a ser mediana. En cambio el agricultor que tenía mas de 50 años

de edad, considerado como agricultor tradicional tendría una baja disponibilidad a aceptar y utilizar insumos mejorados en su parcela.

Para la variable ESC:

ESC = 1; si A_4 (grado de escolaridad) = 0

ESC = 2; si A_4 es ≤ 4 .

ESC = 3; si A_4 es $>$ que 4, es decir ha cursado mas de 4 grados escolares.

Se pensó que un agricultor que tenía más de cuatro grados de escuela primaria se le consideraba como nivel primaria y tenía una mayor disponibilidad y capacidad para asimilar e incorporar insumos mejorados en su producción.

El agricultor que tenía 4 o menos grados de escuela primaria se le consideraba como alfabetizado y con una capacidad media para incorporar insumos mejorados a su parcela.

El agricultor con "0" grados de escolaridad se le consideró como poco apto para poder incorporar insumos mejorados en su parcela.

Para la variable fuente principal de ingresos

(A_5):

$A_5 = 1$; si la fuente era la parcela

$A_5 = 2$; si es el comercio

$A_5 = 3$; si es el jornal

$A_5 = 4$; si es la ganadería

$A_5 = 5$; si es otra fuente

A partir de A_5 , se generó la variable A_5B .

$A_5B = 1$; si es la agricultura la fuente principal de ingreso.

$A_5B = 2$; si es otra la fuente principal de inreso.

Con relación a la variable tipo de explotación en base al recurso agua (A_{19}), ésta se cuantificó así:

$A_{19} = 2$; si es la temporal

$A_{19} = 3$; si es punteado

$A_{19} = 4$; si es humedad

$A_{19} = 5$; si es riego.

También a partir de A_{19} , se generó una nueva variable, ésta se le denominó "agua" y se cuantificó de la siguiente forma:

Agua = 1; temporal (incluye a humedad y punteado)

Agua = 2; riego

En cuanto a la variable tipo de propiedad (T_p); se cuantificó así:

T_p = 1; si es propiedad privada

T_p = 2; si es propiedad ejidal

T_p = 3; si es rentada, a medias o a tercias

Se generó la variable Jornales Totales (JT), - sumando Jornales Familiares y Jornales Asalariados. También se generó la variable Jornales Familiares por Hectárea (JFH), al dividir Jornales Familiares entre la superficie sembrada de maíz (A_{18}).

La variable Jornales Asalariados por Hectárea - (JAH) se obtuvo al dividir Jornales Asalariados entre A_{18} , igual se procedió para generar la variable Jornales Totales por Hectárea.

También se generó la variable proporción de Jornales Familiares, dividiendo la variable JFH entre JTH y multiplicando por 100.

La variable Proporción Sembrada del Cultivo (PSC) se obtuvo al dividir la superficie sembrada de maíz (A_{18}) entre el tamaño de la explotación (A_{10}).

También se trabajó con la variable Otros Cultivos (OC); si A_{10} es mayor que A_{18} entonces $OC = 2$ si A_{10} es igual a A_{18} entonces $OC = 1$.

Todas las variables que se usaron en este trabajo también fueron generadas en forma logarítmica, para ser probadas en modelos logarítmicos y observar su comportamiento y compararlo con los modelos corridos en forma lineal. Así por ejemplo la variable A_{18} , se generó en forma logarítmica de la siguiente manera:

$$L A_{18} = \text{Log } 10 (A_{18} + 1)$$

Y la variable INMES de la misma forma:

$$L IN = \text{Log } 10 (INMEJ + 1)$$

Más especificaciones sobre esta metodología puede encontrarse en Portillo V. (7).

Modelos ensayados

A continuación se presenta a modo de ilustración algunos ejemplos de modelos probados, en que se utilizaron los procedimientos y variables antes descritas; así como los resultados parciales obtenidos, hasta que se llegó a seleccionar el modelo explicativo que finalmente se escogió.

Se definió el siguiente modelo, al cual se procedió a correr una correlación parcial, una regresión por pasos y una regresión lineal múltiple.

$$INMEJ = f(A_1, A_3, A_4, A_5, A_8, A_9, A_{10}$$

$$A_{18}, A_{19}, Tp, Rend, JF, JA, JT)$$

Los resultados obtenidos fueron los siguientes:

- 1) La matriz de correlaciones parciales reportó a las variables Rend, A_1 , A_{19} , A_{10} , A_{18} - como las más correlacionadas con la variable dependiente (INMEJ).

2) La regresión por pasos seleccionó las variables A_1 , Rend y JT, con una $R^2 = 0.38$.

3) La regresión lineal múltiple seleccionó las variables A_1 , A_5B , Rend, y JF, como las más significativas.

Luego se procedió a correr una regresión por pasos y una regresión lineal múltiple, incluyendo en este modelo nuevas variables generadas a partir de las que no habían entrado en los primeros modelos ensayados, pero que a priori se consideraban importantes, fue así como se probó el siguiente modelo además de los anteriores.

$$\text{INMEJ} = f(A_1, \text{Edad}, \text{Esc}, \text{Agua}, A_5B, A_8, A_9, A_{10}, A_{18}, \text{Tp}, \text{Rend});$$

Los resultados obtenidos fueron los siguientes:

Para la regresión por pasos; entraron las variables A_1 , Agua, A_5B , A_{18} y Rend, con una $R^2 = 0.384$.

En la regresión lineal múltiple, aparecen como las variables más importantes según la prueba de "t": A_1 , Agua, A_5B , A_{18} y Rend.

Como se observa ambos procedimientos seleccionaron las mismas variables como las más importantes. Posteriormente se ensayaron otros modelos incorporando nuevas variables y cuantificando de diferente manera a otras que ya se habían ensayado antes.

Este nuevo modelo fue:

$$\begin{aligned} \text{INMEJ} = f(A_1, \text{Edad}, \text{Esc}, \text{Agua}, A_5B, A_8, A_9, \\ A_{10}, A_{18}, \text{Tp}, \text{Rend}, \text{OC}, \text{PSC}, \text{JF}, \\ \text{JF}, \text{JT}); \end{aligned}$$

A este modelo se le aplicó la regresión por pasos y la regresión lineal múltiple y los resultados obtenidos fueron:

Para la regresión por pasos las variables seleccionadas fueron: A_1 , Rend, JT; y la $R^2 = 0.396$.

La regresión lineal múltiple seleccionó las variables A_1 , Agua, A_5B , Rend, JF.

También se probaron una serie de modelos que tenían como variable dependiente el uso de fertilizante (Fert), ya que se estimó que es uno de los insumos mejorados que utiliza en mayor grado el agricultor. Los diferentes modelos ensayados fueron poco prometedores y solamente la variable Asistencia a Demostraciones Agrícolas (A_1) fue la única que apareció como significativa y en forma sistemática en las diferentes pruebas que se realizaron.

Se prosiguió a usar el Número de Insumos Mejorados (INMEJ) como variable dependiente, para lo cual con los mismos procedimientos técnicos antes señalados se probó el siguiente modelo:

$$\begin{aligned} \text{INMEJ} = f(A_1, \text{Edad, Esc, Agua, } A_5B, A_8, \\ A_9, A_{10}, A_{18}, \text{Tp, Rend, OC, PSC,} \\ \text{JFH, JAH, JTH, PJF}); \end{aligned}$$

En la regresión por pasos se seleccionaron las variable, A_1 , Agua, A_5B , A_{18} y Rend.

En la regresión lineal múltiple se presentaron como más significativas las variables A_1 , Agua, A_5B , A_{18} y Rend.

En esta forma se continuó trabajando y observando que había ciertas variables que sistemáticamente apare

cían en las diferentes pruebas realizadas, por lo que se decidió probar las anteriores variables en modelos logarítmicos, a fin de tratar de encontrar un mejor ajuste de regresión y para eso se procedió a utilizar el siguiente modelo:

$$\begin{aligned} \text{LIN} = f(\text{LA}_1, \text{LAgua}, \text{LRend}, \text{LA}_3, \text{LA}_4, \text{LA}_5\text{B}, \\ \text{LA}_8, \text{LA}_9, \text{LA}_{10}, \text{LA}_{18}, \text{LA}_{19}, \text{LTP}, \\ \text{LEdad}, \text{LEsc}, \text{LPJF}, \text{LJT}, \text{LJA}, \text{LOC}, \\ \text{LPSC}); \end{aligned}$$

donde cada variable está precedida de una L indicando que se trabajó con el logaritmo (base 10) de sus valores.

El procedimiento de regresión por pasos seleccionó las siguientes variables: LA_1 , LAgua, LRend, LA_{19} , - LJT.

La regresión lineal múltiple indica que las variables más significativas son: LA_1 , LAgua, LRend, LA_{19} .

Finalmente, luego de hacer diferentes combinaciones de variables y de usar diferentes modelos, tanto lineales como logarítmicas, se seleccionó el siguiente modelo como el "mejor".

$$\text{INMEJ} = f(\text{A}_1, \text{A}_5\text{B}, \text{Rend}, \text{JF}, \text{A}_{18}, \text{Agua});$$

Este modelo fue el que reportó un mejor ajuste general, y en el que entraron seis variables con un buen nivel de significancia según la prueba de "t" y definió una R^2 de 0.404, la más alta a esos niveles de significancia.

Para tratar de encontrar las elasticidades de cada una de las variables en forma directa y poder determinar el nivel de importancia relativa de ellas, se procedió a ensayar el modelo seleccionado como el "mejor", en forma logarítmica.

$$LIN = f(LA_1, LA_5B, LRe nd, LJF, LAgua);$$

Sin embargo dado que la R^2 en este modelo fue de 0.335, no coincidieron las elasticidades así obtenidas con las calculadas utilizando el modelo lineal, tal como se podrá observar más adelante. Este modelo en forma logarítmica definió el siguiente orden de importancia para las variables.

CUADRO No. 3

IMPORTANCIA RELATIVA DE LOS DIFERENTES FACTORES SOCIOECONOMICOS SEGUN EL RESULTADO DEL MODELO LOGARITMICO

F A C T O R	COEFICIENTE DE ELASTICIDAD
Asistencia a demostraciones (A_1)	1.225
Condición del Recurso Agua (Agua)	0.318
Superficie Sembrada de Maíz (A_{18})	0.285
Rendimiento en Ton/Ha. (Rend)	0.282
Jornales Familiares (JF)	0.080
Fuente de Ingresos (A_5B)	- 0.249

FUENTE: Cálculo directo.

III. DISCUSION DE RESULTADOS

Partiendo de un total de 200 observaciones que fueron utilizadas en el análisis de regresión y luego de haber probado 24 variables y varios modelos, se obtuvo finalmente el modelo explicativo más aceptable, en cuanto a la significancia estadística que presentan los parámetros estimados.

$$Y = -1.034 + 1.105X_1 + 0.412X_2 - 0.256X_3 \\ (0.457) \quad (0.290)^1 \quad (0.204)^2 \quad (0.155)^3 \\ + 0.359X_4 + 0.141X_5 + 0.010X_6 + \hat{\epsilon} \\ (0.087)^4 \quad (0.097)^5 \quad (0.005)^6$$

Donde:

- Y = Número de Insumos Mejorados (INMEJ)
- X₁ = Asistencia a Demostraciones Agrícolas (AI)
- X₂ = Condición del Recurso Agua (Agua)
- X₃ = Fuente de Ingresos (A₅B)
- X₄ = Rendimiento en Ton/Ha. (Rend)
- X₅ = Superficie Sembrada de maíz (A₁₈)
- X₆ = Total de Jornales Familiares Utilizados (JF)
- () = Error Standard de cada parámetro estimado
- $\hat{\epsilon}$ = Término Aleatorio de Error

TABLA DE ANALISIS DE VARIANZA

FUENTE DE VARIACION	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	CUADRADO MEDIO	F CALCULADA
Regresión	6	131.696	21.949	20.48
Error	181	194.021	1.072	
Total <u>co</u> rregido	187	325.718		

F de la Tabla 2.80

Prueba de F 0.0001 (Altamente significativa).

R^2 0.4043

PRUEBA DE "T" PARA LA SIGNIFICANCIA ESTADISTICA
DE CADA PARAMETRO

PARAMETRO	T PARA $H_0 : \hat{\beta}_j = 0$	P R : T
Ordenada al Origen	- 2.26	0.0250
X_1	3.82	0.0002
X_2	2.02	0.0451
X_3	- 1.65	0.1009
X_4	4.10	0.0001
X_5	1.45	0.1498
X_6	2.03	0.0443

CUADRO N.º 4
MODELO DEFINITIVO Y TABLA DE ANALISIS DE VARIANZA

DEPENDENT VARIABLE: INPEJ

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	6	131.69653812	21.94942302	20.48	0.0001	0.404327	86.5089
ERROR	181	154.02154699	1.07194225				
CORRECTED TOTAL	187	325.71808511					
					STD DEV		INPEJ MEAN
					1.03534644		1.19680851

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
1	1	81.01326583	75.53	0.0001	1	15.60359114	14.56	0.0002
AGUA	1	16.29965119	15.21	0.0001	1	4.36289442	4.07	0.0451
58	1	2.46745068	2.30	0.1310	1	2.91488594	2.72	0.1009
ENC	1	15.26627737	14.24	0.0002	1	18.04515834	16.83	0.0001
18	1	12.25395119	11.43	0.0019	1	2.24212255	2.09	0.1496
IF	1	4.39554195	4.10	0.0443	1	4.39554195	4.10	0.0443

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > T	STD ERROR OF ESTIMATE
INTERCEPT	-1.03380656	-2.26	0.0250	0.45744776
1	1.16525036	3.82	0.0002	0.28970139
AGUA	0.41158456	2.02	0.0451	0.20421107
58	-0.26655952	-1.65	0.1009	0.15558335
ENC	0.36622518	4.10	0.0001	0.08755232
18	0.04477332	1.45	0.1496	0.29726750
IF	0.4566650	2.03	0.0443	0.22487237

Verificación. Pruebas de hipótesis:

A) Prueba para la significancia total de la regresión

$$H_0: \hat{\beta}_1 = \hat{\beta}_2 = \dots = \hat{\beta}_6 = 0$$

HA : Cuando menos una ($\hat{\beta}_j$), o una combinación lineal de los ($\hat{\beta}_j$)'s es diferente de cero.

Estadístico:

$$F = \frac{SCE/(K)}{SCE/(n - k - 1)} \sim F(K), (n - k - 1)$$

Regla de decisión:

Si $F_c > F_t$ $\therefore$ se rechaza H_0 .

$$F_c = \frac{131.696/6}{194.021/181} = \frac{21.9493}{1.0719} = 20.48$$

$$F_{\text{tabla}} = [6, (181)] \quad \alpha = 0.01 = 2.80$$

Como la $F_c > F_t$ ($20.48 > 2.80$) se rechaza la hipótesis nula (H_0) de que las varianzas y la suma de los cuadrados son iguales a cero, a un nivel de significancia estadística (α) igual a 0.01 es decir, que el conjunto de variables incluidas en el modelo son significativamente explicativas, por lo que no se rechaza la hipótesis alter

nativa, de que son significativamente distintas de cero - las varianzas y las sumas de los cuadrados.;

La salida del computador (Cuadro No. 4) señala un nivel de significancia igual a 0.0001, lo que indica que el modelo en general tiene buen ajuste, según esta prueba.

B) Prueba para la significancia de cada uno de los parámetros estimados ($\hat{\beta}_j$)

Para la verificación de la significancia estadística de cada parámetro se aplica la prueba de "T" de Student. Hay que tener presente que las pruebas de "T" - son aproximadas, ya que sólo pueden ser exactas si no existiera correlación alguna entre las variables explicativas.

Prueba de relevancia de j -ésima variable explicativa (X_j). La relevancia de la variable (X_j) en el modelo de regresión se determina a través de su coeficiente de regresión ($\hat{\beta}_j$), dado que si las variaciones de (X_j) no "explican" a las variaciones de (Y) entonces ($\hat{\beta}_j$) tendría un valor de cero (en el sentido probabilístico).

Hipótesis:

$$H_0 : \hat{\beta}_j = 0$$

$H_A : \hat{\beta} \neq 0$ (al menos uno de los $\hat{\beta}$ es distinto de cero).

Estadístico de prueba:

$$T_c = \frac{\hat{\beta}_j - 0}{S_{\hat{\beta}_j}} = \frac{\hat{\beta}_j}{S_{\hat{\beta}_j}} \sim t_{(\eta - k - 1)}$$

Regla de decisión:

Si $|t_c| > |t_t| \therefore$ se rechaza H_0 .

1. Variable X_1 (asistencia a demostraciones agrícolas).

Hipótesis:

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 > 0$$

$$t_c = \frac{1.10529236}{.28970139} = 3.82$$

$$t_t = [181 \alpha = (0.005)] = 2.58$$

Como $t_c > t_t$ ($3.82 > 2.58$) se rechaza la hipótesis de que el parámetro $\hat{\beta}_1 = 0$ a un nivel de significancia ($\alpha = 0.005$), es decir que la variable "asistencia a demostraciones agrícolas" es relevante porque sus varia-

ciones "explican" parcialmente las variaciones en el uso de insumos mejorados.

De igual manera, siguiendo este procedimiento - podemos determinar la significancia estadística para cada uno de los parámetros, resultando para:

$$X_2 \quad t_c = 2.02; \quad t_t (\alpha = 0.025) = 1.96$$

$$X_3 \quad t_c = 1.65; \quad t_t (\alpha = 0.05) = 1.64$$

$$X_4 \quad t_c = 4.10; \quad t_t (\alpha = 0.005) = 2.58$$

$$X_5 \quad t_c = 1.45; \quad t_t (\alpha = 0.10) = 1.28$$

$$X_6 \quad t_c = 2.03; \quad t_t (\alpha = 0.025) = 1.96$$

Por lo que observamos podemos concluir que todos los parámetros son significativos al 0.05, exceptuando X_5 que es significativo a nivel de 10%, o sea que la hipótesis nula para X_5 se rechaza a un nivel de significancia - ($\alpha = 0.1$) y a este nivel se comprueba que el cociente $\hat{\beta}_5$, es estadísticamente diferente de cero.

Basándose en el nivel de significancia estadística de los estimadores y en el coeficiente de regresión de cada uno de ellos se pueden deducir las siguientes interpretaciones:

X_1 = Asistencia a demostraciones agrícolas (A_1)

Según el modelo ajustado obtenido, esta variable indica que la asistencia a demostraciones agrícolas, por parte de los productores, hace que estos utilicen un número mayor de insumos mejorados que los agricultores que no asisten a este tipo de eventos, este número es 1.10 veces mayor que los que no asisten.

X_2 = Condición del recurso agua (Agua)

Para esta variable los resultados señalan que al pasar de una opción a otra, de acuerdo al orden establecido, se obtienen resultados crecientes en el número de insumos mejorados que utiliza el productor.

Por lo que agricultores que siembran bajo riego utilizan en promedio 1.236 insumos mejorados más que los agricultores que siembran en condiciones de temporal 0.824 insumos más que los que con riego de auxilio (punta de riego), y 0.412 insumos más que los que siembran en terrenos de humedad.

X_3 = Fuente de ingreso ($A_5 B$)

El modelo estimado indica que los productores - que cuentan con otra fuente de ingresos adicional a la

agricultura utilizan en promedio 0.256 insumos mejorados menos que los agricultores que sólo tienen como fuente de ingresos la agricultura.

X_4 = Rendimiento en Ton/Ha. (Rend)

Este caso puede interpretarse en el sentido de que un aumento de 10% en los rendimientos indican que ha autado en un 3.59% el número de insumos mejorados utilizados por los agricultores y viceversa.

X_5 = Superficie sembrada de maíz (A_{18})

La función estimada para esta variable indica - que agricultores con una mayor superficie sembrada de maíz, utilizan un número mayor de insumos mejorados que aquellos que dedican menos superficie al maíz.

Al respecto se estima que por cada aumento en - promedio del 10% en la superficie dedicada al cultivo de maíz, hay un aumento de 1.41% en el número de insumos me-
jorados.

X_6 = Total de jornales familiares utilizados (JF)

En este caso se puede inferir que un aumento en un 10% en el número total de jornales familiares utilizados, conlleva un aumento de 0.10% en el número total de insumos mejorados empleados por los productores.

La corroboración empírica a que fue sometido el modelo estimado (conjunto de pruebas estadísticas de hipótesis: prueba de "F", pruebas de "t" para los parámetros estimados, y el análisis de la matriz de correlaciones parciales entre las variables) nos dan bases para no rechazar la ecuación de regresión ajustada. La prueba de la bondad del ajuste o prueba de hipótesis sobre la importancia estadística del conjunto de variables explicativas, significativa a nivel de $\alpha = 0.0001$, lo que nos permite concluir que el conjunto de variables incluidas en el modelo son altamente significativas.

En relación a las pruebas de hipótesis sobre la significación estadística de cada variable exógena, tratadas en forma individual, (prueba de "T", permitió comprobar la relevancia de cada una en el modelo; dado que los coeficientes de regresión son significativamente distintos de cero en sentido probabilístico.

En cuanto al coeficiente de determinación ($R^2 = 0.4043$), nos indica que la variación en el número de insu

mos mejorados, utilizados por los agricultores está "explicado" por la variación en las variables independientes contenidas en el modelo en un 40%. Esto implica que el número de insumos mejorados utilizados por los productores - es función de un número mayor de factores que los contenidos en el modelo completo, o que hay un factor que no fue considerado en este análisis y que puede tener mucha influencia sobre la variable dependiente.

Sin embargo es aceptado estadísticamente que seis variables estén explicando este porcentaje, que tiene un número relativamente grande de observaciones para cada variable (188).

Hay que agregar que en este resultado ha podido influir los posibles errores de medición de algunas variables y otras causas aleatorias que contribuyeron a incrementar el valor de la suma de cuadrados del error.

A partir de este modelo se estimó el peso o influencia relativa que cada factor ejerce sobre la variable dependiente (Y), que en este caso es el "número de insumos mejorados", empleados por los productores de la muestra en estudio. Utilizando para esto la fórmula para el cálculo de elasticidades, que nos señala en que porcentaje cambia la variable dependiente ante un cambio de 1% en cada una de las variables independientes; así: para X_1 ;

si X_1 cambia en 1% la variable Y cambia en la relación dada por la siguiente expresión:

$$\frac{dy}{dx_1} \cdot \frac{\bar{X}_1}{\bar{Y}}$$

Sustituyendo datos de la muestra se tiene:

$$1.105 \cdot \frac{1.115}{1.199} = 1.105 \times 0.930 = 1.028$$

De la misma manera y calculando el anterior indicador para las demás variables se obtuvo el siguiente cuadro de resultados.

CUADRO No. 5

ESTIMACION DE LA INFLUENCIA RELATIVA DE SEIS FACTORES SOCIOECONOMICOS SOBRE EL USO DE INSUMOS MEJORADOS

FACTOR ORDEN DE IMPORTANCIA	COEFICIENTE DE ELASTICIDAD
X_1 (A_1)	1.028
X_2 (Agua)	0.423
X_4 (Rend)	0.406
X_6 (J F)	0.189
X_5 (A_{18})	0.152
X_3 ($A_5 B$)	-0.330

De acuerdo a los resultados del cuadro anterior, el orden de importancia de los factores analizados señala que la Asistencia a las Demostraciones Agrícolas es el factor que reporta un mayor peso relativo, en la determina--ción del número de insumos mejorados que utiliza el productor, con un índice de 1.028, es decir que ante un cambio de 1% de X_1 Y cambia en 1.028%, así en orden descendente le sigue el factor Condición del Recurso Agua (X_2), el cual con un cambio de 1%, hace que el Número de Insumos Mejorados (Y) cambie en un 0.423%.

El tercer lugar lo ocupa la variable "Rendimientos en Toneladas por Hectárea" (X_4), la cual hace que con un cambio de 1% el número de insumos mejorados varíe en - un 0.406%, y así sucesivamente en orden de importancia descendente quedan ubicadas las variables "Total de Jornales Familiares" (X_6), con un coeficiente de 0.189, la varia--ble "Superficie Sembrada de Maíz" (X_5) reporta un coeficiente (0.152) y por último la variable "Fuente de Ingresos" (X_3), con un coeficiente de (-0.330).

Las variables independientes que resultaron más correlacionadas con la variable dependiente (INMEJ), según la Matriz de Correlaciones Parciales se pueden apreciar - en el siguiente cuadro:

MATRIZ DE CORRELACIONES PARCIALES/PRUEBA ESTADISTICA DE SIGNIFICANCIA/Y NUMERO DE OBSERVACIONES

	REND	IP	A19	A18	A17	A16	A15	A14	A13	A12	A11	A10	A9	A8	A7	A6	A5	A4	A3	A2	A1	INMEJ
INMEJ	1																					
A1	0.49447	1																				
A3	0.49447	0.49447	1																			
A4	0.49447	0.49447	0.49447	1																		
A5B	0.49447	0.49447	0.49447	0.49447	1																	
A8	0.49447	0.49447	0.49447	0.49447	0.49447	1																
A9	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	1															
A10	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	1														
A18	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	1													
A19	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	1												
IP	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	1											
REND	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	0.49447	1										

CUADRO No. 7

VARIABLE INDEPENDIENTES	COEFICIENTE DE CORRELACION
Asistencia a Demostraciones Agrícolas (A ₁)	0.494
Rendimiento en Ton/Ha. (Rend)	0.494
Condición del Recurso Agua (Agua)	0.357
Jornales Totales (JT)	0.276
Superficie Sembrada de Maíz (A ₁₈)	0.218
Jornales Asalarados (JA))	0.198
Jornales Familiares (JF)	0.191
Fuente de Ingresos (A ₅ B)	-0.131

De acuerdo al anterior cuadro las variables "Asistencia a Demostraciones Agrícolas" (A₁) y "Rendimientos en Toneladas por Hectárea" (Rend), son las variables más correlacionadas con la variable dependiente "Número de Insumos Mejorados" (INMEJ).

En tercer lugar se ubica la variable "Condición del Recurso Agua" (Agua) y así sucesivamente en orden descendente se ubican las siguientes variables: "jornales - Totales" (JT), "Superficie Sembrada de Maíz" (A_{18}), "Jornales Asalariados" (JA), "Jornales Familiares" (JF) y -- "Fuente de Ingresos" (A_5).

Utilizando el procedimiento electrónico de Regresión por Pasos, contenido en el paquete SAS, conocido como "Stepwise", el cual proporciona de manera directa información acerca de la aportación que hace cada una de las variables al coeficiente de determinación o R^2 , se ha elaborado el siguiente cuadro:

CUADRO No. 8

APORTE RELATIVO DE CADA FACTOR AL COEFICIENTE DE DETERMINACION (R^2)

F A C T O R	APORTE AL COEFICIENTE DE DETERMINACION (R^2)
Rendimiento en Ton/Ha. (Rend)	0.254
Asistencia a Demostraciones - Agrícolas (A_1)	0.068
Jornales Familiares (JF)	0.049

.....##

.

F A C T O R	APORTE AL COEFICIENTE DE DETERMINACION (R ²)
Condición del Recurso Agua (Agua)	0.016
Fuente de Ingreso (A ₅ B)	0.010
Superficie Sembra da de Maíz (A ₁₈)	0.007

Al comparar los diferentes procedimientos utilizados para seleccionar los factores que tienen una mayor influencia relativa en la determinación del uso de insumos mejorados entre los productores de maíz de la muestra considerada, se puede observar que la mayoría de las variables escogidas, aparecen seleccionadas sistemáticamente con los diferentes criterios empleados, aunque el orden de importancia definido no coincide exactamente, sin embargo destacan las variables "Asistencia a Demostraciones Agrícolas" y "Rendimientos en Toneladas por Hectárea", como los factores que definen una mayor importancia.

Todo lo anterior indica que existe un nivel -- elevado de consistencia y confiabilidad en la selección de las variables consideradas como las más importantes.

C U A D R O N o . 9

PROCEDIMIENTO DE REGRESION POR PASOS PARA LA OBTENCION DEL MODELO DE REGRESION FINAL (STEP 6)

WARNING: 3 OBSERVATIONS DELETED DUE TO MISSING VALUES.

STEP 1	VARIABLE RECD ENTERED	R SQUARE = 0.25401014	C(P) = 42.67215393		
		DF	SUM OF SQUARES	MEAN SQUARE	F
	REGRESSION	1	82.73862674	82.73862674	63.34
	ERROR	186	242.97545836	1.30634117	0.0001
	TOTAL	187	325.71808511		
	B VALLE	STD ERROR	TYPE II SS	F	PROB>F
	INTERCEPT	0.43076477			
	REND	0.57832369	0.07266832	32.73862674	63.34
					0.0001
STEP 2	VARIABLE AI ENTERED	R SQUARE = 0.32219262	C(P) = 23.95710579		
		DF	SUM OF SQUARES	MEAN SQUARE	F
	REGRESSION	2	104.94396272	52.47198101	43.97
	ERROR	185	220.77412309	1.19337364	0.0001
	TOTAL	187	325.71808511		
	B VALLE	STD ERROR	TYPE II SS	F	PROB>F
	INTERCEPT	-0.74272637			
	AI	1.29099802	0.20528512	22.20533528	18.61
	REND	0.37555873	0.00386648	23.93069619	20.05
					0.0001
STEP 3	VARIABLE JF ENTERED	R SQUARE = 0.37147080	C(P) = 10.98354235		
		DF	SUM OF SQUARES	MEAN SQUARE	F
	REGRESSION	3	120.99475728	40.33158576	36.25
	ERROR	184	204.72332782	1.11262678	0.0001
	TOTAL	187	325.71808511		
	B VALLE	STD ERROR	TYPE II SS	F	PROB>F
	INTERCEPT	-1.06717563			
	AI	1.24773439	0.20528698	20.70981632	18.61
	REND	0.40266719	0.03129339	27.29804246	24.53
	JF	0.01488779	0.00391974	16.05079527	14.43
					0.0002

.....##

STEP 4 VARIABLE AGUA ENTERED

R SQUARE = 0.38730796 C(P) = 8.17129496

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	4	126.15320857	31.53830214	28.92	0.0001
ERROR	183	199.56487653	1.09051845		
TOTAL	187	325.71808511			
	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-1.47915249				
AI	1.21367222	0.28674725	19.53607562	17.91	0.0001
REND	0.32562956	0.08120678	15.53060142	14.29	0.0002
JF	0.01475756	0.01288106	15.76745333	14.46	0.0002
AGUA	0.44545232	0.27433165	5.15845129	4.73	0.0309

STEP 5 VARIABLE ASE ENTERED

R SQUARE = 0.39744313 C(P) = 7.09164491

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	5	129.45441557	25.89088311	24.01	0.0001
ERROR	182	196.26366954	1.07837181		
TOTAL	187	325.71808511			
	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-0.56737618				
AI	1.11693409	0.29045672	15.94633010	14.79	0.0002
ASE	-0.27235832	0.15566412	3.37120622	3.06	0.0819
REND	0.35232646	0.08120678	17.41120358	16.15	0.0001
JF	0.01416173	0.00387438	14.40777049	13.36	0.0003
AGUA	0.44256166	0.20369284	5.09375449	4.73	0.0309

STEP 6 VARIABLE A18 ENTERED

R SQUARE = 0.47432676 C(P) = 7.00000000

	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
REGRESSION	6	131.69653812	21.94942302	20.48	0.0001
ERROR	181	194.02154699	1.07194225		
TOTAL	187	325.71808511			
	B VALUE	STD ERROR	TYPE II SS	F	PROB>F
INTERCEPT	-1.03380658				
AI	1.10525236	0.28970139	15.60359114	14.56	0.0002
ASE	-0.25655552	0.15558335	2.91488574	2.72	0.1009
REND	0.35922558	0.08120678	18.04515334	16.83	0.0001
JF	0.00986690	0.00487237	4.39594195	4.10	0.0443
A18	0.14067332	0.09726750	2.24212255	2.09	0.1498
AGUA	0.41198456	0.20421107	4.36289442	4.07	0.0451

NO OTHER VARIABLES MET THE 0.1500 SIGNIFICANCE LEVEL FOR ENTRY INTO THE MODEL.

IV. RESUMEN Y CONCLUSIONES

En los países dependientes es común encontrar un sector altamente desarrollado, en donde se emplean técnicas de producción moderna que incorporan el uso de los avances técnico-científico a la producción, y polarizadamente a este sector, una masa pequeña de agricultores con una producción tradicional y dispersa en el que el nivel de uso de técnicas modernas es mínimo, lo que se traduce en bajos niveles de producción y productividad, altos costos unitarios y bajo poder competitivo en el mercado.

México no escapa a esta problemática, la investigación enmarcada dentro de esta perspectiva, pretende contribuir parcialmente en la búsqueda de una solución a tan difícil problemática.

El cultivo del maíz es un producto de consumo básico en la dieta del mexicano, sin embargo a pesar de su tradicional importancia, dada la estructura y relación económica que ha vivido el país en las últimas décadas el cultivo ha experimentado un retroceso, y el país ha pasado de exportador a importador del grano en los últimos años.

El presente trabajo enfoca su atención a pequeños productores de maíz, con el propósito de conocer con

cierto grado de aproximación las relaciones que guardan los factores socioeconómicos que afectan a los pequeños productores, y el uso de insumos mejorados con el fin de incrementar la disponibilidad de alimentos para la población.

Hipóticamente se plantea, que el uso de estos insumos mejorados está relacionada con factores socioeconómicos de los pequeños agricultores, tales como la edad del agricultor, grado de escolaridad y tamaño de la explotación.

La información analizada procede de una encuesta realizada en el Valle de México conformada por 200 cuestionarios, donde aparecen datos socioeconómicos de los -- agricultores, tecnología utilizada, uso de la tierra, datos de la explotación, mano de obra e instrumentos empleados.

Para alcanzar los objetivos deseados, se utilizaron en este estudio matrices de correlación parciales, así como el método conocido como regresión múltiple los cuales aplicados en forma complementaria formaron la metodología básica del estudio.

Una vez cuantificadas las variables, se ensayaron diferentes modelos, tanto lineales como logarítmicos

seleccionando el siguiente modelo como el "mejor".

$$Y = \frac{-1.034}{(0.457)} + \frac{1.105X_1}{(0.290)^1} + \frac{0.412X_2}{(0.204)^2} - \frac{0.256X_3}{(0.155)^3} \\ + \frac{0.359X_4}{(0.087)^4} + \frac{0.141X_5}{(0.097)^5} + \frac{0.010X_6}{(0.005)^6} + \hat{\varepsilon}$$

Donde:

Y = Número de Insumos Mejorados (INMEJ).

X_1 = Asistencia a Demostraciones Agrícolas (A_1).

Según el modelo ajustado obtenido, esta variable indica que la asistencia a demostraciones agrícolas, por parte de los productores, hace que éstos utilicen un número mayor de insumos mejorados que los agricultores que no asisten a este tipo de eventos.

X_2 = Condición del recurso agua (Agua).

Para esta variable los resultados señalan que al pasar de una opción a otra, de acuerdo al orden establecido, se obtienen resultados crecientes en el número de insumos mejorados que utiliza el productor.

X_3 = Fuente de Ingresos ($A_5 B$).

El modelo estimado indica que los productores que cuentan con otra fuente de ingresos adicional a la agricultura utili-

zan en promedio 0.256 insumos mejorados menos que los agricultores que sólo tienen como fuente de ingresos la agricultura.

X_4 = Rendimientos en Ton./Ha. (Rend).

Este caso puede interpretarse en el sentido de que un aumento de 10% en los rendimientos hace que aumenten en un 3.59% el número de insumos mejorados utilizados por los agricultores y viceversa.

X_5 = Superficie sembrada de maíz (A_{18}).

La función estimada para esta variable - indica que agricultores con una mayor superficie sembrada de maíz, utilizan un número mayor de insumos mejorados que aquellos que indican menos superficie al maíz.

X_6 = Total de jornales familiares utilizados (JF).

En este caso se puede inferir que un aumento en un 10% en el número total de jornales familiares utilizados, conlleva un aumento de 0.10% en el número total de insumos mejorados empleados por los productores.

Este sería el modelo explicativo más aceptable, en cuanto a la significancia estadística que presentan los parámetros estimados.

Conclusiones

Como resultado más notable de esta investigación es el hecho de que puede evaluarse la influencia de factores socioeconómicos (que muchos de ellos son variables no numéricas), mediante técnicas cualitativas, siempre y cuando se utilice una representación cuantitativa adecuada en cada uno de los factores estudiados.

Las conclusiones específicas que se derivan de este trabajo son las siguientes:

- a) De los factores analizados, el que explica en mayor medida el nivel de insumos mejorados utilizando como indicador el "coeficiente de elasticidad" es la Asistencia a Demostraciones Agrícolas.
- b) De los demás factores, el orden de importancia relativa considerados bajo el mismo indicador son: Condición del Recurso Agua, - Rendimientos en Toneladas por Hectárea, Total de Jornales Familiares Utilizados, Superficie Cosechada de Maíz y Fuente de Ingre--sos.

- c) Los factores anteriormente mencionados determinan en un 40% a la variable "Adopción de Tecnología" representada por el "Número de Insumos Mejorados".

V. RECOMENDACIONES

En base a los resultados obtenidos, se sugiere la utilización del método aplicado en este trabajo en futuras investigaciones, donde se trate de evaluar el comportamiento de variables socioeconómicas, teniendo presente la utilización de una adecuada representación cuantitativa de las variables no numéricas empleadas. Además hay que tener presente el hecho de que este método es un apoyo en la investigación que necesita complementarse con -- otros elementos de las ciencias sociales, de tal forma que se tenga una visión integral del problema en estudio.

Se considera conveniente la incorporación de otras variables tales como cambios en los precios relativos de los insumos y productos agrícolas, ponderación que el agricultor le asigna al riesgo, etc., de tal forma que el análisis sea dinámico y referido a un período determinado de tiempo.

En cuanto a los resultados específicos derivados de este trabajo se estima:

1. Recomendar a las autoridades agrícolas del área de estudio, realizar una reasignación de recursos, de tal forma que se amplíen los programas de asistencia técnica

ca al agricultor, ya que como se observó en el estudio, - esta variable tiene una amplia influencia en la determinación del uso de insumos mejorados.

2. Se recomienda mejorar la disponibilidad del recurso agua a los productores, ya que este factor tiene una influencia significativa en la utilización de otros - insumos mejorados dentro de la población de agricultores incluídos en el área de estudio.

Teniendo presente las condiciones bajo las cuales se realizó esta investigación y los posibles errores de medición de algunas variables, se considera que este - trabajo de bases para futuras investigaciones, en cuanto al método utilizado.

También se estima que dada la rapidez con que - se pueden obtener algunos resultados al aplicar este método, puede contribuir con cierta eficacia a instrumentar - lagunas políticas agrícolas, tendientes a elevar el nivel de productividad de los pequeños agricultores.

VI. BIBLIOGRAFIA CONSULTADA

1. AVILA DORANTES, JOSE ANTONIO. Evaluación de Factores que actúan como Restricciones en la Aplicación de las Recomendaciones del Plan Puebla. Tesis de Maestría, Colegio de Postgraduados, Chapingo, México, 1973.
2. BLANCO, LUIS MARTIN. Función de la Información y Algunos Factores Sico-Sociales con el Uso de Insumos Agrícolas. Tesis de Maestría, Colegio de Postgraduados, Chapingo, México, 1973.
3. DIAZ HERNANDEZ, MOISES. La Toma de Decisiones con el Empleo de Nueva Tecnología y su Efecto Sobre los Rendimientos Unitarios de Maíz de Temporal. Un Estudio de Algunos Factores Sociosicológicos y Físicos que intervienen en el Proceso, en una Comunidad del Plan - Puebla. Tesis de Maestría, Colegio de Postgraduados, Escuela Nacional de Agricultura, Chapingo, México, 1971.
4. FREEBAIRN, D. K. Metodología de la Investigación en Economía Agrícola, Folleto Técnico No. 2, Chapingo, México.
5. MARTIN DEL CAMPO, ANTONIO. La Capacidad del Agricultor para Adoptar Tecnología: Un Caso de - Estudio con Productores de Maíz de Temporal en el Area del Plan Puebla. Tesis de Maestría, Colegio de Postgraduados, Escuela Nacional de Agricultura, Chapingo, México, 1972.
6. MOSCARDI, EDGARDO. Riesgo y Transferencia de Tecnología: Estudio para el Caso del Plan Puebla. Tesis de Maestría, Colegio de Postgraduados, Escuela Nacional de Agricultura, Chapingo, México, 1972.

7. PORTILLO VAZQUEZ, MARCOS. Aspectos Socioeconómicos de la Producción de Maíz en el Area de Otumba, México. Tesis de Maestría en Economía, Colegio de Postgraduados, Chapingo, México, 1980.
8. TELLO DIAGO, RODRIGO. Algunos Factores Determinantes de la Adopción de Tecnología Agrícola. Tesis de Maestría, Colegio de Postgraduados, Escuela Nacional de Agricultura, Chapingo, México, 1974.
9. VILLA ISSA, LUIS A. Adopción de Tecnología Nueva en Zonas de Temporal. El efecto del Factor Incertidumbre. Tesis de Maestría, Colegio de Postgraduados, Escuela Nacional de Agricultura, Chapingo, México, 1974.
10. VILLA ISSA, MANUEL R. El Mercado de Trabajo y la Adopción de Tecnología Nueva de Producción Agrícola: El Caso del Plan Puebla, Colegio de Postgraduados, Centro de Economía Agrícola, Chapingo, México, 1976.
11. MARTINEZ GARZA, ANGEL. Teoría de la Regresión y sus Aplicaciones Econométricas. Notas de Clase, Centro de Estadística y Cálculo, Colegio de Postgraduados, Chapingo, México, 1974.
12. MATUS GARDEA, JAIME A. Notas para Acompañar el Curso de Econometría, Mimeografiado, Chapingo, México, 1979.
13. MONTAÑEZ, CARLOS Y ABURTO, HORACIO. Maíz Política Institucional y Crisis Agrícola: CIDER, Editorial Nueva Imagen; primera edición, México, 1979.
14. WINKELMAN, DONAL. La Adopción de la Nueva Tecnología de Maíz en el Plan Puebla, CIMMYT, México, 1976.
15. YAMANE, TARO. Estadística. Editorial Harla, tercera edición, México, 1974.

VII. CUESTIONARIO UTILIZADO

NUMERO DE CUESTIONARIO ____

SISTEMAS DE PRODUCCION DE MAIZ EN EL AREA
DE INFLUENCIA DEL CAMPO CHAPINGO

1. Datos del agricultor

- 1.1 Nombre _____
- 1.2 Dirección _____
- 1.3 Localidad _____
- 1.4 Municipio _____
- 1.5 Estado _____
- 1.6 Edad _____
- 1.7 ¿Hasta qué año escolar estudió? _____
- 1.8 Fuentes de ingreso _____
- 1.8.1 Parcela _____
- 1.8.2 Comercio _____
- 1.8.3 Jornales _____
- 1.8.4 Ganadería _____
- 1.8.5 Otras _____
- 1.9 ¿Cuántas personas que trabajan fuera de la finca con
tribuyen con usted al sostenimiento de la casa?

- 1.10 ¿Cuántas personas dependen de usted económica-
mente? _____

2. Datos de la explotación

- 2.1 Superficie total de la finca _____ has.

2.2 Tipo de propiedad

Privada _____ ha.

Ejidal _____ ha.

Rentada _____ ha.

A medias _____ ha.

A tercias _____ ha.

2.3 Tipo de suelo predominante (en la terminología del agricultor) _____

2.4 Relieve

Plano _____

Pendiente suave _____

Pendiente fuerte _____

Escarpado _____

3. Uso actual de la tierra

Cultivo	Superficie		Fecha siembra	Cultivo del ciclo anterior*	Tipo, Asociación cultivo actual**	Superficie cosechada	Producción total
	Riego	Temporal(1)					
1. Maíz							
2. Frijol							
3.							
4.							
5.							

(1) Indicar si es punteado

* Indicar si fué con fertilizante (1) o sin fertilizante (2)

* Abono orgánico

** Indicar si es sólo (s), asociado (a) o intercalado (i)

** Abono químico

4. Tecnología. (Se refiere únicamente a la parcela sembrada de maíz)

4.1 Siembra

4.1.1 ¿A qué distancia entre surcos siembra? --

4.1.2 ¿A qué distancia entre matas siembra?

4.1.3 ¿Cuántas semillas por mata deja caer?

4.2 Fertilización

4.2.1 Superficie fertilizada _____

Nombre del fertilizante	Número de bultos en c/aplicación	Peso de cada -- bulto	Labor en la que se efectuó la aplicación	Método de aplicación

4.2.2 En caso de que no fertilice, ¿cuáles son las causas? _____

4.3 Semilla utilizada

Tipo de semilla	Cantidad total en Kg.	Superficie sembrada	Precio por Kg.	Donde la obtiene
1.				
2.				
3.				

4.3.1 ¿Por qué utiliza esos tipos de semilla? -

4.4 Combate de malezas

4.4.1 ¿Cuáles son las principales malezas que atacan su cultivo de maíz? _____

4.4.2 ¿En qué etapa del cultivo causan más daños? _____

4.4.3 ¿En qué etapa del cultivo las combate? _____

4.4.4 ¿Cómo combate esas malezas? _____

a). Métodos culturales _____

b). Métodos químicos

Nombre de la maleza principal	Nombre del herbicida utilizado	Dosis	No. de Aplicaciones	Etapas en que se hizo la aplic.

En caso de que no utilice herbicidas, menciones las causas -

c). Otros métodos

4.5 Control de plagas

4.5.1 ¿Cuáles son las principales plagas que atacan su cultivo de maíz?

4.5.2 ¿En qué etapa del cultivo son más perjudiciales?

4.5.3 ¿En qué etapa del cultivo las combate?

4.6.3 ¿Cómo combate esas enfermedades?

a) Métodos culturales _____

b) Métodos químicos

Nombre de la enfermedad	Nombre del producto	Dosis	No. de aplicaciones	Etapa en - que se hizo la aplicación

En caso de que no las utilice, mencione las causas _____

4.7 Rotación de cultivos

4.7.1 ¿Practica alguna rotación de cultivos?

Si _____ No _____

4.7.2 En caso afirmativo, ¿Cuál es la rotación-típica que practica?

4.5.4 ¿Cómo combate esas plagas?

a) Métodos culturales o manuales _____

¿Controla eficientemente las malas -
hierbas con estos métodos? _____

b) Métodos químicos

Nombre de la plaga princi pal	Nombre del insectici- da	Dosis	No. de Apli caciones	Etapa en que hizo la apli cación

En caso de no utilizar insecticidas, mencione las causas _____

c) Otros métodos _____

4.6 Control de enfermedades

4.6.1 ¿Cuáles son las principales enfermedades-
que atacan su cultivo de maíz? _____

4.6.2 ¿En qué etapa del cultivo? _____

5. Mano de obra utilizada en las diferentes labores

Actividades	Fecha en que se efectuó	Jornales utilizados			Valor del jornal	Instrumentos utilizados
		Familiares	Asalariados	Totales		
1. Barbecho						
2. Cruza						
3. Rastreo						
4. Nivelación						
5. Surcado						
6. Siembra						
7. Escarda						
8. 1a. labor						
9. 2a. labor						
10. Deshierbes						
11. Riego(s) - Anotar cuántos						
12. Abonos						
13. Insecticidas						
14. Corte						
15. Recolección						
16. Otros						