

42
20



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

UNA APLICACION DEL ANALISIS DE REGRESION
A LA PROBLEMATICA AGRARIA.

T E S I S
Que para obtener el Título de
A C T U A R I O
p r e s e n t a
J. Nicomedes Olgún Ramírez



TESIS CON
FALLA DE ORIGEN



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

UNA APLICACION DEL ANALISIS DE REGRESION

A LA PROBLEMÁTICA AGRARIA

INDICE

	Pags.
Introducción.....	1
Capítulo I	"El Problema".....2
I.1	Región Seleccionada.....2
I.2	Fuentes de Información.....7
I.3	Determinación de Variables.....8
Capítulo II	"Conceptos Básicos del Análisis de Regresión Lineal".....13
II.1	El Modelo.....13
II.2	La Estimación de Parámetros.....14
II.3	Los coeficientes de Correlación y de Determinación.....21
II.4	Pruebas de hipótesis.....26
II.5	Procedimiento.....31
Capítulo III	"Resultados Obtenidos".....37
III.1	Presentación e Interpretación de Resultados.....37
III.2	Hipotesis que se desprenden.....89
Conclusiones y Recomendaciones.....	91
Bibliografía.....	95

INTRODUCCION

El propósito del presente trabajo es la exposición de los principales elementos del análisis de regresión y su aplicación, mediante un modelo que nos permita explicar la problemática agraria, al analizar los distintos aspectos que lo caracterizan en un plan de interrelación.

Es importante señalar que el problema de la tierra es tan antiguo como los orígenes de la Nación y que no puede ser tratado separadamente de la problemática general del país, por lo que al realizar este estudio, se está consciente de la necesidad de enfocarlo a la luz de criterios que se derivan de diversas disciplinas, aplicando la técnica seleccionada en conjunción con alguna teoría causal, que permita evaluar las consecuencias lógicas de un modelo estructural de eslabonamiento causa-efecto afirmado a priori.

En la primera parte del trabajo, se hace una breve descripción de la región seleccionada y la determinación de las variables, con el fin de ilustrar la aplicación práctica de la teoría.

El capítulo II lo constituye la exposición detallada del modelo de regresión lineal, con la descripción de la teoría, la presentación de cálculos y demostraciones necesarias en su aplicación, profundizando en ciertas secciones con el fin de dejar clara su aplicación en este caso particular.

El diseño, interpretación de los resultados obtenidos e hipótesis que se desprenden, se integran en un tercer capítulo; incorporando como parte final de esta presentación las conclusiones y recomendaciones que se derivan de los resultados obtenidos, así como del desarrollo de las diversas actividades realizadas.

La necesidad de estudiar en su conjunto las diferentes causas que determinan la problemática que afronta una parte importante de la población rural, como es el ejido y la comunidad agraria; que no puede ni debe ser calificada partiendo de uno de los aspectos que le singularizan (el régimen de tenencia de la tierra), porque entraña varios aspectos de naturaleza diversa, tales como: El aprovechamiento, arrendamiento y acaparamiento parcelario, el cacicazgo, las deficiencias técnicas de explotación de la tierra, la insuficiente inversión y hasta la desviación del crédito, la insolvencia económica de los acreditados, la ausencia de sentido gregario o de asociación para realizar propósitos comunes, la impericia para almacenar, transportar y comercializar sus productos, la incapacidad para transformarlos, etc; dió origen al presente estudio, seleccionando el Estado de Hidalgo como la región a ser analizada.

I.1 REGION SELECCIONADA

El 15 de enero de 1869 se decreta la creación del Estado de Hidalgo como estado libre y soberano; con una superficie de 20,987 km², que podría considerarse repartida en tres regiones conocidas como "La Sierra", "La Huasteca" y "El Valle", localizándose en esta última los valles fértiles de Tulancingo, el Valle del Mezquital que es una zona de gran aridez que abarca varios municipios y los llanos de Apan que se extienden hasta el Estado de México y Tlaxcala.

La población económicamente activa se concentra principalmente en el sector primario (63.2%); el sector industrial absorbe el 18.5% y el de servicio el 18.3%.

Por su topografía accidentada y con un nivel bajo en el

aprovechamiento de sus recursos, la economía del Estado está basada principalmente en las actividades comprendidas dentro del sector primario, con bajas posibilidades productivas e imposibilitado a proporcionar empleo y retener la fuerza de trabajo que se localiza en su medio rural.

La actividad que predomina es la agricultura, aún cuando el Estado presenta características climatológicas y de suelos poco propicias para su desarrollo, ya que una parte del estado es zona árida o semiárida, otra parte tiene elevaciones o problemas de erosión. Dentro de éstas se encuentra el área laborable con aproximadamente el 28% de su extensión; un 20% de ella es superficie de riego, localizada en la región del Valle, con altos rendimientos que no obstante la pequeña superficie que ocupa genera casi tres cuartas partes del producto agrícola estatal; el resto la constituye el área de temporal.

La actividad ganadera tiene altas perspectivas de desarrollo, sin embargo ha sido explotada mínimamente y sin la incorporación de técnicas modernas ni apoyo crediticio.

La actividad forestal apenas se inicia. La importancia de los recursos maderables, justifican cualquier acción encaminada a tomar en cuenta esta actividad como parte esencial de la economía estatal y como una fuente de empleo tan necesaria y urgente.

La minería desempeña un papel importante por los yacimientos que existen en explotación y sus reservas estimadas.

La actividad industrial, a pesar de contribuir en forma importante a la corriente de bienes y servicios producidos, tiene un crecimiento desequilibrado, ya que sólo beneficia a un reducido número de poblaciones.

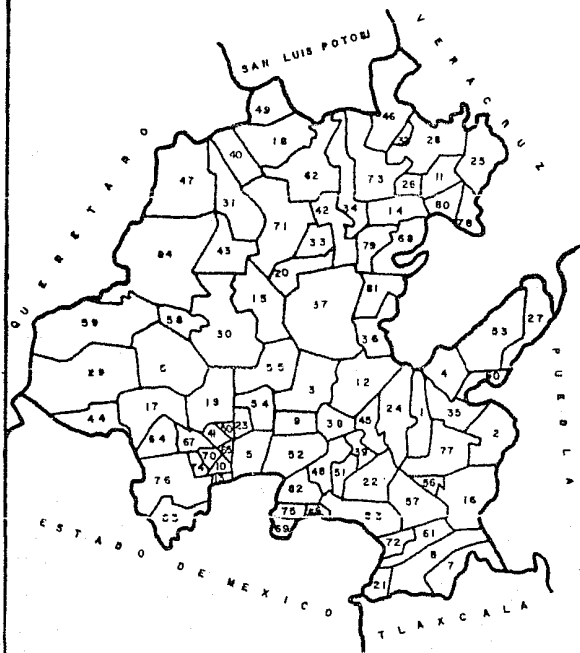
La actividad turística está poco desarrollada a pesar de contar con grandes atractivos.

Respecto a la tenencia de la tierra, en la actualidad se registran como tierras ejidales y comunales un 49% de la superficie del Estado, el resto es propiedad privada, en su mayoría gran propiedad, con una pequeña fracción como pequeños y medianos predios.

En el aspecto ejidal, la dotación por ejidatario fluctua entre 0.5 y 60 hectáreas; sin embargo, el área cultivable varía entre un cuarto y una hectárea, debido a que comprende agostadero y monte alto principalmente.

Debo hacer notar que dentro del universo de trabajo, la unidad de análisis es el municipio y que debido a la naturaleza del estudio se realiza en aquellos en que existe la forma de tenencia de la tierra ejidal y/o comunal. De los 84 municipios que integran la región, en 4 no se obtuvo información (Calnali, Juárez Hidalgo, Nicolás Flores y Progreso). Por otra parte, en adelante no se hará distinción entre ejido o comunidad y ejidatarios o comuneros, siendo el término a utilizar ejido o ejidatario según sea el caso. A continuación se presenta un mapa con la división municipal del Estado de Hidalgo.

H I D A L G O



HIDALGO

- | | |
|----------------------------|------------------------------|
| 1. - Acatlán | 43. - Nicolás Flores |
| 2. - Acaxochitlán | 44. - Nopala |
| 3. - Actopan | 45. - Omitlán de Juárez |
| 4. - Agua Blanca Iturbide | 46. - Orizatlán |
| 5. - Ajacuba | 47. - Pacula |
| 6. - Alfajayucan | 48. - Pachuca |
| 7. - Almoloya | 49. - Pisaflores |
| 8. - Apan | 50. - Progreso |
| 9. - Arenal, El | 51. - Reforma, La |
| 10. - Atitalaquia | 52. - San Agustín Tlaxiaca |
| 11. - Atlapexco | 53. - San Bartolo Tutotepec |
| 12. - Atotonilco El Grande | 54. - San Salvador |
| 13. - Atotonilco Tula | 55. - Santiago |
| 14. - Calnali | 56. - Santiago Tulantepec |
| 15. - Cardonal | 57. - Singuilucan |
| 16. - Cuautepec | 58. - Taxquillo |
| 17. - Chapantongo | 59. - Tecozautla |
| 18. - Chapulhuacan | 60. - Tenango de Doria |
| 19. - Chilcuautla | 61. - Tepeapulco |
| 20. - Eloxochitlán | 62. - Tepehuacán de Guerrero |
| 21. - Emiliano Zapata | 63. - Tepeji del Río |
| 22. - Epazoyucan | 64. - Tepetitlán |
| 23. - Francisco I Madero | 65. - Tetepango |
| 24. - Huasca | 66. - Tezontepec |
| 25. - Huautla | 67. - Tezontepec de Aldama |
| 26. - Huazalingo | 68. - Tianguistengo |
| 27. - Huehuetla | 69. - Tizayuca |
| 28. - Huejutla | 70. - Tlahuelilpan |
| 29. - Huichapan | 71. - Tlahuilltepa |
| 30. - Ixmiquilpan | 72. - Tlanalapan |
| 31. - Jacala | 73. - Tlanchinol |
| 32. - Jalteacán | 74. - Tlaxcoapan |
| 33. - Juárez Hidalgo | 75. - Tolcayuca |
| 34. - Lolotla | 76. - Tula de Allende |
| 35. - Metepec | 77. - Tulancingo |
| 36. - Metzquititlán | 78. - Xochiatipán |
| 37. - Metztlitlán | 79. - Xochicoatlán |
| 38. - Mineral del Chico | 80. - Yahualica |
| 39. - Mineral del Monte | 81. - Zacualtipán |
| 40. - Misión, La | 82. - Zapotlán de Juárez |
| 41. - Mizquiahuala | 83. - Zempoala |
| 42. - Molango | 84. - Zimapán |

I.2 FUENTES DE INFORMACION

Para el desarrollo del presente estudio se utilizó la información que se consideró más confiable; aún cuando se pudiera trabajar con indicadores más elaborados, se decidió por la recuperación de las fuentes de información existentes en la Secretaría de la Reforma Agraria en el año de 1978, mismas a las que tuve acceso por encontrarme en esa época desempeñando funciones de coordinación de la Dirección General de Programación y Evaluación con la Dirección General de Servicios Electrónicos, clasificando aquella que cuantitativamente y cualitativamente describía mejor a la región seleccionada y en especial al sector ejidal, complementando en su caso con la información del Censo General de Población 1970. Los archivos utilizados fueron:

- 1).- Historial Agrario.- Relación cronológica de los Mandatos Presidenciales publicados en el Diario Oficial de la Federación, que nos muestra la verdad jurídica en cuanto a tenencia de la tierra.
- 2).- Ejecución de Resoluciones Presidenciales.- Integrado a partir del historial, concentrando las acciones agrarias ejecutadas en cada uno de los núcleos agrarios.
- 3).- Documentos Básicos.- Relación actualizada de ejidos con documentos básicos completos; esto es, por cada acción agraria llevada a cabo en el ejido.
- 4).- Padrón Nacional de Ejidatarios.- Relación de ejidatarios en pleno goce de sus derechos.
- 5).- Inventario de Recursos en Ejidos y Comunidades 1974 .- Encuesta realizada en 1974 por la Secretaría de la Reforma Agraria a través de su Dirección General de

Planeación, por medio de un cuestionario aplicado a cada uno de los ejidos del país, que por su contenido estaba encaminado a determinar sus recursos y el grado de aprovechamiento de los mismos. Al término de la encuesta y después del procesamiento de los datos captados, se obtuvo información rica en detalle sobre una cobertura del 74% de la región seleccionada en este estudio.

I.3 DETERMINACION DE VARIABLES

Un estudio estadístico involucra a un grupo de individuos con algunos atributos comunes y ubicados en el tiempo y el espacio; las mediciones efectuadas son designadas como variables estadísticas, de tal manera que tendremos un conjunto de valores de las observaciones que podrán ser resumidas en una matriz de datos de la forma:

$$X = \begin{bmatrix} X_{n1} & \dots & X_{np} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{bmatrix} = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ X_n \end{bmatrix}$$

donde X denota el vector de variables aleatorias o su matriz de valores observados.

A continuación se dan a conocer las variables que fueron determinadas en base a las fuentes de información descritas en el punto anterior. Conciente de sus limitaciones, considero que son suficientes para efectuar el análisis que se pretende, cuya descripción teórica se presenta en el Capítulo siguiente; dichas variables intentan representar solas o en grupo los diversos eventos que en base a las encuestas y concentraciones de información se pudieron medir.

No.	IDENTIFICACION	DESCRIPCION	OBSERVACIONES
E S T R U C T U R A D E L A T I E R R A			
1	EJCEDO	PROPORCION DE EJIDOS DEL MUNICIPIO RESPECTO AL TOTAL DEL ESTADO	SUPERFICIE EJIDAL ES EL NUMERO DE HECTAREAS QUE HASTA EL MOMENTO SE HAN DOTADO AL EJIDO SIN IMPORTAR LOS TRAMITES AGRARIOS QUE REQUIRIRIA PARA HACER SU CONSTITUCION LEGAL
2	SUPTOT	SUPERFICIE EJIDAL RESPECTO A LA MUNICIPAL	
3	SUPLAB	SUPERFICIE DE LABOR EJIDAL RESPECTO A LA SUPERFICIE DE LABOR MUNICIPAL	
4	SUPRIE	SUPERFICIE DE RIEGO EJIDAL RESPECTO A LA SUPERFICIE DE LABOR EJIDAL	
5	EJIRIE	NUMERO DE EJIDOS CON RIEGO RESPECTO AL TOTAL DE EJIDOS	
6	DOTEJI	DOTACION PROMEDIO POR EJIDATARIO	
P O B L A C I O N A L			
7	DENPOB	DENSIDAD DE POBLACION EJIDAL	EJIDATARIO ES LA PERSONA A QUIEN LA SECRETARIA DE LA REFORMA AGRARIA LE HA OTORGADO PODER LEGAL SOBRE UNA PORCION DE TIERRA, APEGANDOSE A LOS LINEAMIENTOS QUE ESTABLECE LA LEY PARA LA CONSERVACION DE LA MISMA SON AQUELLAS PERSONAS QUE NO ALCANZARON TIERRA EN EL MOMENTO DE LA DOTACION, PERO QUE GUARDAN UN LUGAR PRIORITARIO PARA NUEVAS ACCIONES DOTATORIAS SE CONSIDERO PARA SU CALCULO A LA POBLACION DE 10 AÑOS O MAS QUE SAREN LEER Y ESCRIBIR
8	EJIPEA	EJIDATARIOS CON RESPECTO A LA PEA (POBLACION ECONOMICAMENTE ACTIVA) MUNICIPAL	
9	DERSAL	CAMPESINOS CON DERECHOS A SALVO RESPECTO AL TOTAL DE EJIDATARIOS	
10	POBALF	PORCENTAJE DE POBLACION ALFABETA EN EL EJIDO	
11	POBIND	PORCENTAJE DE POBLACION INDIGENA EN EL EJIDO	

No.	IDENTIFICACION	DESCRIPCION	OBSERVACIONES
T R A N S I T A C I O N A G R A R I A			
12	RESEJE	RESOLUCIONES PRESIDENCIALES EJECUTADAS RESPECTO A LAS PUBLICADAS	SE TRATA DE MEDIR CON ESTAS VARIABLES EL GRADO DE AVANCE EN LAS GESTIONES AGRARIAS PARA LA CONSTITUCION LEGAL DE LOS EJIDOS
13	DOCBAS	EJIDOS CON DOCUMENTACION BASICA COMPLETA RESPECTO AL TOTAL DE EJIDOS	
14	DEREGU	PORCENTAJE DE EJIDATARIOS CON DERECHOS AGRARIOS REGULARIZADOS	
P R O B L E M A S A G R A R I O S			
15	PROINS	PORCENTAJE DE EJIDOS CON DESAPROVECHAMIENTO DE AREAS LABORALES POR PROBLEMAS INSTITUCIONALES	SON LAS DIFICULTADES QUE ATANEN A UN EJIDO PARA SU POSESION LEGAL COMO PROBLEMAS INSTITUCIONALES A: A) FALTA DE CREDITO B) AUSENTISMO DE EJIDATARIOS C) PROBLEMAS AGRARIOS (VARIABLES 17 A 20) Y COMO FENOMENOS NATURALES A: A) EROSIONES B) INUNDACIONES C) FALTA DE AGUA
16	PRONAT	PORCENTAJE DE EJIDOS CON DESAPROVECHAMIENTOS DE AREAS LABORALES POR FENOMENOS NATURALES	
17	DESLIN	PORCENTAJE DE EJIDOS CON PROBLEMAS DE DESLINDES Y/O CONFLICTOS POR LINDEROS	
18	CENSUC	PORCENTAJE DE EJIDOS CON PROBLEMAS DE DEPURACION CENSAL Y/O SUCESION DE DERECHOS AGRARIOS	
19	EJEPRE	PORCENTAJE DE EJIDOS CON PROBLEMAS DE EJECUCIONES PRESIDENCIALES	
20	DESPOJ	PORCENTAJE DE EJIDOS CON DESPOJO DE TIERRAS	

No.	IDENTIFICACION	DESCRIPCION	OBSERVACIONES
O R G A N I Z A C I O N			
21	ORCHED	PORCENTAJE DE EJIDOS ORGANIZADOS PARA OBTENER CREDITOS	SE REFIERE AL APOYO PRESTADO POR INSTITUCIONES PUBLICAS O PRIVADAS AL NUCLEO AGRI-CO, CON PERSONAL CALIFICADO, EN EL DESARROLLO DE LAS DISTINTAS ACTIVIDADES
22	CREDOF	PORCENTAJE DE EJIDOS CON CREDITO DE BANCOS OFICIALES	
23	CREPRI	PORCENTAJE DE EJIDOS CON CREDITO DE PARTICULARES U ORGANISMOS PRIVADOS	
24	ASITEC	PORCENTAJE DE EJIDOS CON ASISTENCIA TECNICA	
25	ORPROD	PORCENTAJE DE EJIDOS ORGANIZADOS PARA LA PRODUCCION	
26	ORINSU	PORCENTAJE DE EJIDOS ORGANIZADOS PARA LA COMPRA DE INSUMOS	
27	ORVENT	PORCENTAJE DE EJIDOS ORGANIZADOS PARA LA VENTA DE SU PRODUCCION	
28	AFIPOL	PORCENTAJE DE EJIDOS AFILIADOS A ALGUNA ORGANIZACION POLITICA	
E C O N O M I C A S			
29	AGRIEJ	VALOR PROMEDIO DE LA PRODUCCION AGRICOLA EJIDAL	ES EL MONTO A PRECIOS CORRIENTES DE LA PRODUCCION GENERADA POR LA ACTIVIDAD, SIN IMPORTAR SI DICHS PRODUCTOS SON DESTINADOS AL CONSUMO INTERNO DEL EJIDO O A SU COMERCIALIZACION CON EL EXTERIOR
30	GANAEJ	VALOR PROMEDIO DEL INVENTARIO GANADERO EJIDAL	
31	FRUTEJ	VALOR PROMEDIO DE LA PRODUCCION FRUTICOLA Y AGAVES	

No.	IDENTIFICACION	DESCRIPCION	OBSERVACIONES
E C O N O M I C A S			
32	MACRIC	PORCENTAJE DE EJIDOS QUE USAN MAQUINARIA AGRICOLA	SE CONSIDERAN LAS UNIDADES DE TIPO ELECTRICO O MECANICO COMO SON TRACTORES, SECADORAS, TRILLADORAS, DESGRANADORAS, PICADORAS, Y EMPACADORAS
33	INSREJ	PORCENTAJE DE EJIDOS QUE USAN INSUMOS MEJORADOS	
34	PEAPRI	PROPORCION DE LA PEA EN EL SECTOR PRIMARIO	
35	PEASEC	PROPORCION DE LA PEA EN EL SECTOR SECUNDARIO	
S E R V I C I O S			
36	EDACEJ	PORCENTAJE DE EJIDOS CON ESCUELA PRIMARIA	SE DETECTA SI CUENTA CON PERSONAL DESTINADO A LA ENSEÑANZA DEL NIVEL PRIMARIO Y Y SI ES INTEGRAL LA INSTITUCION, SIN CONSIDERAR EL EDIFICIO O LAS CONDICIONES DE LOS LOCALES EN QUE SE IMPARTEN LAS CLASES.
37	AGULJZ	PORCENTAJE DE EJIDOS CON AGUA ENTUBADA Y/O LUZ ELECTRICA	SEÑALA SI GOZA DE LOS BENEFICIOS FEDERALES OTORGADOS PARA LA DISTRIBUCION DE ESTOS SERVICIOS EN SU ZONA URBANA
38	VIACOH	PORCENTAJE DE EJIDOS CON CARRETERA PAVIMENTADA Y/O CAMINO DE TERRACERIA	
39	MECOM	PORCENTAJE DE EJIDOS CON MEDIOS DE COMUNICACION	SE REFIERE A SERVICIOS FEDERALES O PARTICULARES, POR MENOS DE LOS CUALES SE GUARDA UNA ESTRECHA RELACION QUE INTERCAMBIA NOTICIAS Y FLUJO DE PERSONAS ENTRE EL EJIDO Y OTRAS ZONAS, COMO SON:
40	VIVSER	PORCENTAJE DE EJIDOS CON VIVIENDAS CON SERVICIOS	A) LINEA DE CAMIONES B) FERROCARRIL C) TRANSPORTE FLUVIAL D) CORREO E) TELEGRAFO F) TELEFONO

CAPITULO II. CONCEPTOS BASICOS DEL ANALISIS DE REGRESION LINEAL.

II.1 EL MODELO

El análisis de regresión es una técnica estadística a través de la cual se puede analizar la relación entre una variable dependiente o criterio y una o varias variables independientes o predictivas. Los usos más importantes de esta técnica, como instrumento descriptivo, son:

- a) Encontrar la mejor ecuación de predicción y evaluar correctamente esa predicción.
- b) Controlar la confusión de los factores que contribuyen a la especificación de una variable en particular o un conjunto de variables.
- c) Encontrar relaciones estructurales y generar explicaciones de las relaciones multivariadas aparentemente complejas.

Es decir, mediante el análisis de regresión podrá obtenerse una ecuación lineal de predicción que indica cómo deben ponderarse los valores de las variables independientes para obtener una explicación de la variable dependiente. La función matemática más simple es la expresada por el modelo:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + c_i \quad i=1,2,\dots,n \quad (1)$$

el cual es llamado modelo lineal múltiple, porque es expresado como una combinación lineal de parámetros.

Las n ecuaciones pueden formularse en notación matricial como sigue:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & \dots & X_{1k} \\ 1 & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

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

\underline{Y} es el vector n-dimensional cuyas componentes son las Y_i .
 X es una matriz de elementos fijos y conocidos, que consiste de las variables independientes, de orden $(n \times k)$ y rango k menor que n ; donde n es el número de observaciones. Observe que X es tal que $X'X$ es invertible (ó no singular); es decir que:

$$|X'X| \neq 0,$$

por lo que existe una matriz B única, tal que

$$BX'X = X'XB = I$$

en donde B es la inversa de $X'X$ y se denota por

$$B = (X'X)^{-1}$$

$\underline{\beta}$ es el vector k-dimensional de parámetros desconocidos.

$\underline{\epsilon}$ es un vector aleatorio (vector de errores) de orden $(n \times 1)$, tal que :

$$E(\underline{\epsilon}) = 0$$

dado que puede tomar tanto valores positivos como negativos y los errores más próximos a cero son los más frecuentes,

$$\text{Var}(\underline{\epsilon}) = \sigma^2 I_n$$

los errores tienen varianza constante y finita, siendo la covarianza

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0 \quad (i \neq j) \quad i, j = 1, 2, \dots, n$$

$$\text{con } E(\underline{Y}) = X \underline{\beta}$$

II.2. ESTIMACION DE PARAMETROS

La estrategia del análisis de regresión se basa en tomar los valores de los estimadores, en tal forma que la suma de los cuadrados de los residuales sea mínima. Criterio de mínimos

cuadrados); en donde el residual está definido como la distancia vertical de la observación Y a la línea estimada, o sea $(Y_i - \hat{Y}_i)$.

Sea $\hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]'$ un vector columna de las estimaciones β ; esto implica que podemos escribir:

$$Y = X \hat{\beta} + e \quad (3)$$

en donde e representa al vector columna de los n residuos $(Y - X\hat{\beta})$ y

$$\begin{aligned} \sum_{i=1}^n e_i^2 &= e'e & (4) \\ &= (Y - X \hat{\beta})' (Y - X \hat{\beta}) \\ &= Y'Y - 2 \hat{\beta}' X' Y + \hat{\beta}' X' X \hat{\beta} \end{aligned}$$

dado que $\hat{\beta}' X' Y$ es un escalar, por lo tanto es igual a su traspuesta $Y' X \hat{\beta}$.

Los valores de $\hat{\beta}$ que hacen mínima la expresión, son aquellos con los que la derivada parcial respecto a $\hat{\beta}$ es igual a cero

$$\frac{d}{d \hat{\beta}} e'e = 0 - 2 X' Y + 2 X' X \hat{\beta} = 0$$

de donde $(X' X) \hat{\beta} = X' Y$

como $r(X) = k$

entonces $r(X' X) = k$

implica que $(X' X)^{-1}$ existe.

por lo que el vector de estimadores (de mínimos cuadrados) está dado por:

$$\hat{\beta} = (X' X)^{-1} X' Y \quad (5)$$

para obtener la media y la matriz de covarianzas de $\hat{\beta}$, tenemos:

$$\begin{aligned} \hat{\beta} &= (X' X)^{-1} X' (X \beta + \varepsilon) \\ &= \beta + (X' X)^{-1} X' \varepsilon \end{aligned}$$

puesto que $I = (X' X)^{-1} (X' X)$

esto expresa a $\hat{\beta}$ como una función lineal de la verdadera pero desconocida β y de las perturbaciones ε ,

$$E(\hat{\beta}) = E(\beta) + E[(X'X)^{-1}X'\epsilon] \\ = \beta + (X'X)^{-1}X'EC\epsilon$$

puesto que los valores de X permanecen fijos y $E(\epsilon) = 0$

$$E(\hat{\beta}) = \beta$$

lo cual demuestra que los estimadores minimocuadráticos son insesgados. Por lo tanto como $E(\hat{\beta}_i) = \beta_i$, para $i=1, 2, \dots, k$; podemos deducir que $E(\hat{\beta}_i - \beta_i)^2$ es la varianza de $\hat{\beta}_i$ y $E[(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)]$ es la covarianza de $\hat{\beta}_i$ y $\hat{\beta}_j$; es decir: $E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] =$

$$\begin{bmatrix} E(\hat{\beta}_1 - \beta_1)^2 & E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) & \dots & E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k) \\ E(\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1) & E(\hat{\beta}_2 - \beta_2)^2 & \dots & E(\hat{\beta}_2 - \beta_2)(\hat{\beta}_k - \beta_k) \\ \vdots & \vdots & \ddots & \vdots \\ E(\hat{\beta}_k - \beta_k)(\hat{\beta}_1 - \beta_1) & E(\hat{\beta}_k - \beta_k)(\hat{\beta}_2 - \beta_2) & \dots & E(\hat{\beta}_k - \beta_k)^2 \end{bmatrix}$$

es una matriz simétrica que contiene las varianzas en la diagonal y las covarianzas en los lugares restantes, que se denomina matriz de varianzas y covarianzas de $\hat{\beta}$ y se identifica por $\text{var}(\hat{\beta})$. La expresión de esta matriz, puede obtenerse como:

$$\text{var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \quad \text{como } (\hat{\beta} - \beta) = (X'X)^{-1}X'\epsilon \\ = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\ = (X'X)^{-1}X'E(\epsilon\epsilon')X(X'X)^{-1} \\ = (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}$$

en virtud de que $\epsilon\epsilon'$ es una matriz simétrica de orden n (ϵ es un vector columna $n \times 1$ y ϵ' un vector fila), al tomar esperanzas matemáticas en cada uno de los elementos de la matriz tenemos en la diagonal principal las varianzas $E(\epsilon_i^2) = \sigma^2$, para toda i , los términos restantes son covarianzas $\text{Cov}(\epsilon_i, \epsilon_j) = 0$, para $i \neq j$, es decir los valores de ϵ no están correlacionados por parejas. Como σ^2 es un escalar se puede colocar antes o después de una matriz según convenga e $I_n = (X'X)^{-1}(X'X) = (X'X)(X'X)^{-1}$ puede suprimirse, tenemos que:

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

por lo que la varianza de cualquier coeficiente $\hat{\beta}_i$ se puede obtener tomando el i -ésimo término de la diagonal principal

de $(X' X)^{-1}$ y multiplicarlo por σ^2 que es la varianza de ϵ .

Con objeto de determinar las propiedades de los parámetros, recordemos el teorema de Gauss-Markov.

Teorema de Gauss-Markov

Los mejores estimadores lineales insesgados, en el sentido de varianza mínima, de los componentes del vector de parámetros β , son aquellos que minimizan la suma de cuadrados de los errores de predicción (residuales).

Demostración

Para ver que los estimadores minimocuadráticos son los de mínima varianza, definamos por ejemplo:

$\tilde{\beta} = ZY$ en donde Z es una matriz $K \times n$

$$\begin{aligned} E(\tilde{\beta}) &= E [Z (X\beta + \epsilon)] \\ &= Z X \beta \\ &= C \beta \end{aligned}$$

siempre y cuando $ZX = C$, en donde C es una matriz $r \times k$ de constantes conocidas, con $r \leq k$; $c\beta$ representa un conjunto de funciones lineales de los parámetros β ;

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= E [(C\tilde{\beta} - C\beta)(C\tilde{\beta} - C\beta)'] \\ &= E (Z \epsilon \epsilon' Z') \end{aligned}$$

ya que

$$\tilde{\beta} = ZY = ZX\beta + Z\epsilon = C\beta + Z\epsilon$$

por lo tanto

$$\text{Var}(\tilde{\beta}) = \sigma^2 ZZ'$$

para minimizar las varianzas de nuestros estimadores lineales insesgados, deberemos escoger los elementos de Z tales que $ZX = C$ y que los elementos de la diagonal principal de ZZ' sean los más pequeños posibles. Consideremos la identidad

$$\begin{aligned} ZZ' &\cong [CCX'X)^{-1}X'] [CCX'X)^{-1}X']' + \\ &\quad [Z - CCX'X)^{-1}X'] [Z - CCX'X)^{-1}X']' \end{aligned}$$

que se puede establecer multiplicando los términos del

segundo miembro y utilizando $ZX = C$. Los dos términos del segundo miembro son de la forma de una matriz multiplicada por su transpuesta, por lo tanto, los elementos diagonales de cada producto son necesariamente no negativos; por otra parte, considerando el primer producto, vemos que los elementos de C y X son constantes conocidas, por consiguiente los términos diagonales de este producto son constantes dadas. Estas constantes representan los valores posibles mínimos de los estimadores en $\tilde{\beta} = ZY$ y la única manera de evitar que nuestros estimadores tengan varianzas mayores que éstas es establecer:

$$Z = C(CX'X)^{-1}X'$$

con lo que el último término del segundo miembro de ZY , se compone de una matriz nula multiplicada por su transpuesta, lo que nos da una matriz nula; sustituyendo en nuestra función original, los óptimos estimadores lineales insesgados de $C\beta$, están dados por

$$\tilde{\beta} = C(CX'X)^{-1}X'Y$$

que es lo mismo si estimamos $C\beta$ reemplazando β por su estimador minimocuadrático definido en (5). Por lo tanto, los estimadores minimocuadráticos son los óptimos lineales insesgados; si $C = I_k$ tenemos que $C\beta = \beta$ y el estimador lineal insesgado óptimo de β es $(X'X)^{-1}X'Y$, que es el minimocuadrático obtenido en (5).

Ahora bien, podemos obtener un estimador de σ^2 a partir de los residuos

$$\begin{aligned} e &= Y - X\hat{\beta} \\ &= X\beta + \varepsilon - X[(CX'X)^{-1}X'(X\beta + \varepsilon)] \\ &= \varepsilon - X(CX'X)^{-1}X'\varepsilon \\ &= [I - X(CX'X)^{-1}X']\varepsilon \end{aligned}$$

por lo que los residuos observados son una función lineal de las perturbaciones desconocidas. Si definimos

$A = I - X(CX'X)^{-1}X'$; se puede ver que A es una matriz simétrica idempotente, ya que $A' = A$ y

$$A^2 = [I - X(CX'X)^{-1}X'] [I - X(CX'X)^{-1}X']$$

$$\begin{aligned}
 A^2 &= I - 2 X (X' X)^{-1} X' + X (X' X)^{-1} X' X (X' X)^{-1} X' \\
 &= I - X (X' X)^{-1} X' \\
 &= A
 \end{aligned}$$

de tal forma que

$$\begin{aligned}
 e'e &= e' A' A e \\
 &= e' A e \\
 &= e' [I - X(X'X)^{-1}X']e
 \end{aligned}$$

por lo que

$$\begin{aligned}
 E(e'e) &= E\{e' [I - X(X'X)^{-1}X'] e\} \\
 &= \sigma^2 t_r [I - X(X'X)^{-1}X'] \\
 &= \sigma^2 [t_r(I) - t_r(X(X'X)^{-1}X')] \\
 &= \sigma^2 [n - t_r(X(X'X)^{-1}X'X)] \\
 &= (n - k) \sigma^2
 \end{aligned}$$

puesto que

$$\begin{aligned}
 E(e'Ae) &= E\{t_r(e'Ae)\} \\
 &= t_r\{E(Aee')\} \\
 &= t_r\{CA\} E(ee') \\
 &= t_r\{CA\sigma^2 I_n\} \\
 &= \sigma^2 t_r\{CA\}
 \end{aligned}$$

$$t_r(A+B) = t_r(A) + t_r(B)$$

$$t_r(ABC) = t_r(BCA) = t_r(CAB)$$

$$(X'X) \text{ es de orden } k, (X'X)^{-1}(X'X) = I_k$$

luego

$$S^2 = \frac{e'e}{n - k}$$

de (4) tenemos:

$$\begin{aligned}
 S^2 &= \frac{Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'Y}{n - k} \\
 &= \frac{Y'Y - \hat{\beta}'X'Y}{n - k}
 \end{aligned}$$

S^2 es un estimador insesgado de σ^2

Si ahora utilizamos el supuesto $\varepsilon \sim N(0, \sigma^2 I)$, la función de verosimilitud de β y σ^2 es:

$$L = (2\pi)^{-n/2} \sigma^{-n} \exp[-e'e/2\sigma^2]$$

$$L = \frac{1}{(2\pi \sigma^2)^{n/2}} \exp \left[- \frac{(Y - X\beta)' (Y - X\beta)}{2 \sigma^2} \right]$$

maximizar esta expresión respecto a β , equivale a minimizar la suma de cuadrados $(Y - X\beta)' (Y - X\beta)$. Es decir, el estimador maximoverosímil de β es el minimocuadrático ya obtenido $\hat{\beta} = (X'X)^{-1}X'Y$, que puede expresarse como una función lineal de β y ϵ , $\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon$, lo que significa que tiene una distribución normal multivariante $\hat{\beta} \sim N[\beta, \sigma^2(X'X)^{-1}]$.

Por otra parte tenemos que $(n - k) s^2 / \sigma^2 \sim \chi^2_{(n-k)}$; puesto que $e'e = e'Ac = e'[I - X(X'X)^{-1}X']\epsilon$ en donde A es una matriz simétrica idempotente y que su traza y su rango es $n-k$, es posible hallar una matriz ortogonal P tal que $P'AP = E_{n-k}$ en donde E_{n-k} es matriz diagonal con $n - k$ unidades y k ceros en su diagonal principal; P puede utilizarse para definir una transformación del vector ϵ al vector V , de tal forma que:

$$\epsilon = PV \quad \text{o} \quad V = P'\epsilon \quad \text{ya que } P^{-1} = P'$$

con lo que

$$e'e = \epsilon'Ac$$

$$= V'P'APV$$

$$= V'E_{n-k}V$$

$$= V_1^2 + V_2^2 + \dots + V_{n-k}^2$$

que expresa la suma de cuadrados de los residuos como la suma de cuadrados de $n-k$ variables aleatorias normales independientes con media cero y varianza σ^2 .

Finalmente, si consideramos la matriz simétrica $E[e(\hat{\beta} - \beta)']$ y sustituimos por los valores obtenidos anteriormente, tenemos que:

$$E[e(\hat{\beta} - \beta)'] = E[(I - X(X'X)^{-1}X')\epsilon\epsilon'X(X'X)^{-1}]$$

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

$$= 0$$

dado que e y $\hat{\beta}$ son funciones lineales de variables aleatorias normales, tienen también una distribución normal y como su covarianza es cero, se puede concluir que se distribuyen independientemente.

II.3 LOS COEFICIENTES DE CORRELACION Y DE DETERMINACION

Con el análisis de regresión, también podemos obtener estadísticos que nos permiten conocer el grado de relación lineal que existe entre las variables, así como la variación en la variable dependiente explicada por la influencia lineal conjunta de las variables independientes. Un medio adecuado para medir el grado de determinación conjunta de las relaciones entre los hechos y expresarlas en fórmulas breves, es el coeficiente de correlación.

Un coeficiente de correlación nos proporciona simultáneamente tres datos principales, en un sólo valor:

- 1) La existencia o no existencia de una relación lineal entre las variables estudiadas
- 2) La dirección de esta relación, si es que existe y
- 3) El grado de esta relación.

El coeficiente de correlación momento-producto de Pearson (r), es un coeficiente de correlación paramétrico que nos indica el grado de correlación lineal entre dos variables; es decir, hasta que punto una variación en una se corresponde a una variación en la otra; su expresión es:

$$r = \frac{n \sum X Y - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad (6)$$

6

$$r = \frac{\sum x_i y_i}{n S_x S_y} \quad (7)$$

$$x_i = X_i - \bar{X} \quad y_i = Y_i - \bar{Y}$$

$$S_x = \sqrt{\sum x_i^2 / n} \quad S_y = \sqrt{\sum y_i^2 / n}$$

En la práctica, raramente se dan correlaciones perfectas y los valores del coeficiente oscilan entre +1 y -1, pasando por el cero que es la no-correlación.

Si el análisis de regresión es aplicado conjuntamente con el análisis de correlación, obtenemos una relación entre el coeficiente de correlación y la pendiente estimada (coeficiente de regresión) \hat{b} en el modelo $Y_i = a + bx_i + \varepsilon_i$

$$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2} = r \frac{S_y}{S_x}$$

$$r = \frac{\sum x_i y_i}{n S_x S_y} = \hat{b} \frac{S_x}{S_y} \quad (8)$$

si $\hat{b} = 0$, r también es cero

por la definición de línea mínimocuadrática:

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \quad i = 1, 2, 3, \dots, n$$

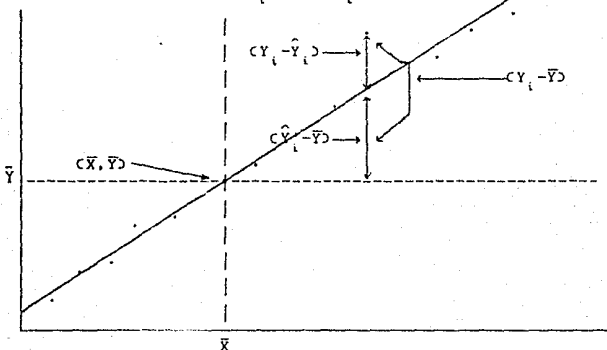
desviación Total = desv. explicada + desv. no explicada

lo cual implica

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

variación total = var. explicada + var. no-explicada

Recta ajustada $\hat{Y}_i = \hat{a} + \hat{b}x_i$ \longrightarrow



en términos de desviaciones respecto a las medias tenemos:

$$\begin{aligned} \sum y_i^2 &= \sum (\hat{y}_i + e_i)^2 \\ &= \sum \hat{y}_i^2 + \sum e_i^2 \\ &= \hat{b}^2 \sum x_i^2 + \sum e_i^2 \end{aligned}$$

puesto que

$$\begin{aligned} \sum \hat{y}_i e_i &= \hat{b} \sum x_i e_i \\ &= \hat{b} \sum x_i (y_i - \hat{b}x_i) = 0 \end{aligned}$$

$$\frac{\hat{\Sigma}y_i^2}{\Sigma y_i^2} = \frac{\Sigma(\hat{b} x_i)^2}{\Sigma y_i^2} \quad \hat{\Sigma}y_i^2 = \hat{b}^2 \Sigma x_i^2$$

de donde se deduce que

$$r^2 = \hat{b}^2 \frac{\Sigma x_i^2}{\Sigma y_i^2} \quad (9)$$

es decir, el cuadrado del coeficiente de correlación, llamado coeficiente de determinación, es igual a la proporción de la varianza de Y explicada por la influencia lineal de X.

sustituyendo, tenemos que:

$$r^2 = 1 - \frac{\Sigma e_i^2}{\Sigma y_i^2} \quad (10)$$

así los límites de r son ± 1

Se pueden diseñar otros coeficientes de correlación para medir la contribución de cada variable a la variación de la variable dependiente, como es el de:

Correlación Parcial $r_{Y_1.2} = \frac{r_{Y_1} - r_{Y_2} r_{12}}{\sqrt{1 - r_{Y_2}^2} \sqrt{1 - r_{12}^2}}$

que es una correlación simple entre dos residuales, el de Y con X_1 , donde se eliminan los efectos de X_2 . Es decir, describe la relación entre dos variables controlando los efectos de una o más variables adicionales; esto es la correlación entre dos variables, eliminando de estas la información contenida en las demás variables incluidas en el modelo.

Para ejemplificar lo anterior, sea $r_{12.2}$:

eliminaremos de Y y X_2 toda la influencia lineal de X_2 , de la forma siguiente:

tomemos la regresión lineal de Y en X_2

$$Y_i = a_{12} + b_{12} X_{2i} + e_i$$

los residuos inexplicados en Y_i , serán obtenidos por

$$\begin{aligned}\mu_i &= Y_i - \hat{a}_{13} - \hat{b}_{13} X_{3i} \\ &= y_i - \hat{b}_{13} x_{3i}\end{aligned}$$

similarmente los residuos no explicados en X_2 , después de remover la influencia lineal de X_3 son

$$\begin{aligned}\nu_i &= X_{2i} - \hat{a}_{23} - \hat{b}_{23} X_{3i} \\ &= x_{2i} - \hat{b}_{23} x_{3i}\end{aligned}$$

finalmente, el coeficiente de correlación parcial entre Y con X_2 considerando a X_3 en el modelo, está definido por el coeficiente de correlación simple entre μ_i y ν_i

$$\begin{aligned}r_{\mu\nu} &= \frac{\sum \mu_i \nu_i}{\sqrt{\sum \mu_i^2} \sqrt{\sum \nu_i^2}} \\ &= \frac{\sum \mu_i \nu_i}{n S_\mu S_\nu} \\ &= r_{12.3}\end{aligned}$$

En general, el cuadrado del coeficiente de correlación parcial, representa el incremento proporcional en la variación explicada debido a X_i , (X_i variable independiente, con $i = 1, \dots, k$) expresado como una proporción de la variación inexplicada por las variables independientes ya incluidas en la ecuación.

Como una definición general, tenemos que el coeficiente de un orden dado puede expresarse en términos del siguiente de orden más bajo, por ejemplo:

$$\begin{aligned}b_{12.3} &= \frac{r_{12} - r_{13} r_{23}}{1 - r_{23}^2} \frac{S_1}{S_2} \\ b_{13.2} &= \frac{r_{13} - r_{12} r_{23}}{1 - r_{23}^2} \frac{S_1}{S_3}\end{aligned}$$

En el caso del coeficiente de correlación múltiple $R_{1.23}$ tendríamos que:

$$R_{1.23}^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2} \quad \text{según (10)}$$

$$= 1 - \frac{S_{1.23}^2}{S_1^2}$$

en donde $S_{1.23}^2 = 1/n \sum (Y_i - \hat{Y}_i)^2$

$$R_{1.23}^2 = \frac{b_{12.3} \sum y_i x_{2i} + b_{13.2} \sum y_i x_{3i}}{\sum y_i^2}$$

$$= \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}$$

La otra forma de encontrar el coeficiente de determinación R^2 , es medir el coeficiente de correlación simple entre Y_i con los valores de regresión \hat{Y}_i , de tal manera que:

$$r_{Y\hat{Y}}^2 = \frac{(\sum y_i \hat{y}_i)^2}{(\sum y_i^2)(\sum \hat{y}_i^2)}$$

como $\sum y_i \hat{y}_i = \sum (y_i + e_i) \hat{y}_i = \sum y_i \hat{y}_i$ para $\sum y_i e_i = 0$

$$r_{Y\hat{Y}}^2 = \frac{(\sum y_i \hat{y}_i)^2}{(\sum y_i^2)(\sum \hat{y}_i^2)}$$

simplificando tenemos

$$r_{Y\hat{Y}}^2 = \frac{\sum y_i \hat{y}_i^2}{\sum y_i^2} \quad (11)$$

resultado general que funciona para k variables explicativas. Es decir, para el caso general podemos definir al coeficiente de determinación en términos de la variación del residual sobre el plano de regresión:

$$\sum y_i \hat{y}_i^2 = \sum Y_i^2 - \frac{1}{n} (\sum Y_i)^2$$

$$= Y'Y - \frac{1}{n} (\sum Y_i)^2$$

$$\begin{aligned}\hat{\Sigma y_i^2} &= \Sigma y_i^2 - \Sigma e_i^2 \\ &= Y'Y - e'e - \frac{1}{n} (\Sigma Y_i)^2 \\ &= \hat{\beta}'X'Y - \frac{1}{n} (\Sigma Y_i)^2\end{aligned}$$

Aplicando (4)

por lo tanto:

$$R^2_{1, 23 \dots k} = \frac{\hat{\beta}'X'Y - \frac{1}{n} (\Sigma Y_i)^2}{Y'Y - \frac{1}{n} (\Sigma Y_i)^2} \quad (12)$$

Si los resultados se dan en términos de desviaciones, tenemos:

$$R^2 = \frac{\hat{\beta}'X'Y}{Y'Y} \quad (13)$$

siendo la suma de cuadrados, debido a la influencia lineal de las variables explicativas, igual a:

$$\hat{\beta}'X'Y = \hat{\beta}_2 \Sigma x_{2i} y_i + \dots + \hat{\beta}_k \Sigma x_{ki} y_i$$

II.4 PRUEBAS DE HIPOTESIS

Para las pruebas de hipótesis comunes en un modelo lineal, se supone que x tiene distribución normal, con independencia entre sus valores x_i , con media cero y varianza constante y finita igual a σ^2 . Es importante recalcar que estas condiciones son lo suficientemente generales para que el modelo pueda ser aplicado a un gran número de fenómenos, con "cierta seguridad" de que las condiciones se cumplen por lo menos con una aproximación satisfactoria; es decir, lo suficiente para que los resultados de los análisis aplicados al modelo, conduzcan a conclusiones útiles en la práctica. En cualquier caso, existen procedimientos estadísticos que permiten verificar si los supuestos son válidos; básicamente mediante la axaminación de residuales, con el objeto de aprovechar cualquier cosa de interés que ellos contengan o de determinar como alterar el modelo, para la obtención de mejores resultados. El análisis de residuales

es usualmente realizado de dos maneras, gráficamente o en base a estadísticas; el método gráfico es muy fácil de llevar a cabo y muy revelador cuando los supuestos son violados; entre las técnicas estadísticas utilizadas podemos señalar la prueba de Durbin-Watson, la prueba de rachas y la prueba de bondad de ajuste.

Una hipótesis en el modelo $Y = X\beta + \epsilon$, consistirá en la suposición de que uno o más elementos del vector de parámetros β toman valores específicos o bien que cierto grupo son iguales o cumplen una determinada relación matemática. Por ejemplo:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_0 : \beta_1 - \beta_3 = \beta_5 = \beta^*$$

$$H_0 : 3\beta_1 + 8\beta_7 - 5\beta_4 - 6\beta_2 = 0$$

De una manera muy general, la hipótesis consiste en establecer restricciones sobre los parámetros; estas restricciones se pueden representar por $C\beta = \gamma$, donde C es una matriz $q \times k$, de rango ($Cq \leq K$) y γ es un vector $q \times 1$.

Un recurso importante, dentro de esta etapa de análisis, es el Análisis de Varianza, que divide la suma total de cuadrados en diferentes partes independientes, cada parte atribuible a una fuente, causa o factor conocido. Si el experimento ha sido diseñado en forma apropiada, será posible analizar la varianza total en una cantidad de componentes independientes, cada uno de los cuales puede identificarse con una de las fuentes de variación. Los pasos necesarios para la realización de la prueba pueden ser resumidos en la llamada Tabla de Análisis de Varianza (ANOVA) y el procedimiento es el siguiente:

- A) Con el modelo completo obtener los estimadores que minimizan la suma de cuadrados de los errores (SC). De
- (B) tenemos que el vector de estimadores que minimizan

$e'e$ está dado por $\hat{\beta} = (X' X)^{-1} X' Y$

B) Con los estimadores del punto anterior, obtener el valor de la SC de los errores (SC_{omc}).

De (4) y (5), tenemos

$$e'e = Y' Y - \hat{\beta}' X' Y = Y' Y - Y' X (X' X)^{-1} X' Y \quad (14)$$

$$= Y' [I - X (X' X)^{-1} X'] Y$$

C) Establecer la hipótesis en el modelo generando el llamado modelo reducido, modelo bajo la hipótesis; obtener en éste los estimadores que minimizan la SC y con ellos calcular la SC del error en el modelo reducido (SC_{omr}).

Si se establece la hipótesis $C\beta = \gamma$ se genera el modelo reducido. Usando multiplicadores de Lagrange se tendrá:

$$Q^* = (Y - X\hat{\beta}^*)' (Y - X\hat{\beta}^*) - 2\lambda(C\hat{\beta}^* - \gamma)$$

en donde λ es un vector $1 \times q$ de multiplicadores de Lagrange.

$$\frac{dQ^*}{d\hat{\beta}^*} = -2 Y' Y + 2 X' X \hat{\beta}^* - 2C' \lambda$$

$$-2X' Y + 2X' X \hat{\beta}^* - 2C' \lambda = 0$$

$$X' X \hat{\beta}^* = X' Y + C' \lambda$$

$$\hat{\beta}^* = (X' X)^{-1} X' Y + (X' X)^{-1} C' \lambda$$

$$\hat{\beta}^* = \hat{\beta} + (X' X)^{-1} C' \lambda \quad \text{al aplicar (5)}$$

$$\frac{dQ^*}{d\lambda} = C \hat{\beta}^* - \gamma \quad \text{de donde} \quad C \hat{\beta}^* = \gamma$$

sustituyendo, tenemos que

$$-\lambda = [C (X' X)^{-1} C']^{-1} C (X' X)^{-1} X' Y - [C (X' X)^{-1} C']^{-1} \gamma$$

$$\hat{\beta}^* = \hat{\beta} + (X' X)^{-1} C' [C (X' X)^{-1} C']^{-1} (\gamma - C\hat{\beta})$$

Si los estimadores $\hat{\beta}$ en el modelo completo cumplen la hipótesis $C\hat{\beta} = \gamma$, entonces $\hat{\beta}^* = \hat{\beta}$; es decir, habrá una concordancia máxima entre la hipótesis y los datos.

Para encontrar la suma de cuadrados del error en el modelo reducido, se tiene:

$$SC_{\text{omr}} = e^* e^* = Y' Y - \hat{\beta}^* X' X \hat{\beta}^* + C' \gamma - C \hat{\beta}^* \cdot [C (X' X)^{-1} C']^{-1} (\gamma - C\hat{\beta})$$

D) Si denotamos.

$$\begin{aligned} SC_{H_0} &= \text{"Suma de cuadrados debida a la hipótesis"} \\ &= SC_{\text{emr}} - SC_{\text{emc}} \quad \text{para este caso será} \end{aligned}$$

$$SC_{H_0: C\beta=\gamma} = (Y - C\hat{\beta})' [C(X'X)^{-1}C']^{-1} (Y - C\hat{\beta}) \quad (15)$$

Dicha diferencia es una medida de la discrepancia que los datos presentan con la hipótesis. El calificativo de "grande" o "pequeña" dependerá básicamente de la escala de medida usada.

E) Para estandarizar estas medidas de variabilidad (SC), se obtienen los cuadrados medios (CM).

$$CM_{\text{emc}} = \frac{SC_{\text{emc}}}{(n - k)} \quad (16)$$

donde n es el tamaño de la muestra, k es el número de parámetros independientes en el modelo completo y (n - k) se conoce como grados de libertad.

CM_{emc} resulta ser un estimador insesgado de σ^2 , denotado frecuentemente por S^2 .

A la suma de cuadrados debida a la hipótesis, se le asocia como grados de libertad las restricciones de la hipótesis impuestas en los parámetros, en este caso k que es el rango de C.

$$CM_{H_0} = \frac{SC_{H_0}}{k} \quad (17)$$

F) Finalmente se comparan las magnitudes relativas de CM_{H_0} y CM_{emc} mediante la llamada razón F de Snedecor.

$$F_{\text{cal}} = \frac{CM_{H_0}}{CM_{\text{emc}}} = \frac{(Y - C\hat{\beta})' [C(X'X)^{-1}C']^{-1} (Y - C\hat{\beta})}{Y' [I - X(X'X)^{-1}X'] Y} \frac{(n-k)}{k} \quad (18)$$

Si la hipótesis es $\beta = 0$ ($\beta_1 = \beta_2 = \dots = \beta_k = 0$), entonces $C = I$, $\gamma = 0$ y:

$$F_{\text{cal}} = \frac{\hat{\beta}' X' X \hat{\beta}}{Y' [I - X(X'X)^{-1}X'] Y} \frac{(n - k)}{k}$$

Si esta F calculada resulta mayor que la F de tablas, con k y (n - k) grados de libertad y un nivel de significancia α ,

se puede afirmar que existe una relación lineal entre la variable dependiente y las variables independientes, cuyo grado de explicación estará dado por el coeficiente de determinación R^2 .

Tabla de Analisis de Varianza para $H_0: \beta = 0$

FUENTE DE VARIACION	SC	GRADOS DE LIBERTAD	CM	F
X_1, X_2, \dots, X_k	$\hat{\beta}' X' X \hat{\beta}$	k	$\frac{\hat{\beta}' X' X \hat{\beta}}{k}$	$\frac{\hat{\beta}' X' X \hat{\beta} / k}{Y' Y - \hat{\beta}' X' Y / (n-k)}$
RESIDUOS	$Y' Y - \hat{\beta}' X' Y$	(n-k)	$\frac{Y' Y - \hat{\beta}' X' Y}{(n-k)}$	
TOTAL	$Y' Y$	n		

De igual forma el procedimiento puede abarcar cualquier coeficiente o subgrupo de coeficientes. Para verificar la hipótesis de que $\beta_k = 0$, es decir X_k no ejerce influencia lineal sobre Y, calculamos:

$$F_{\text{cal}} = \frac{\hat{\beta}_k^2}{[\Sigma e^2 / (n-k)] a_{kk}}$$

$\hat{\beta}_k^2$ mide el incremento en la suma de cuadrados explicada debido a la adición de X_k a la lista de variables predictivas.

a_{kk} es el k-ésimo elemento de la diagonal principal de $(X'X)^{-1}$.

Finalmente comparamos el valor resultante de esta F-parcial con un punto preseleccionado en la distribución F con (1, n-k) grados de libertad.

Las pruebas de correlación, regresión y varianza, son todas ellas formas equivalentes para verificar o probar la hipótesis básica de no-relación lineal entre las X y Y. Según (13) tenemos que:

$$\hat{\beta}' X' Y = Y' Y \cdot R^2$$

$$e'e = Y' Y (1 - R^2) \quad \text{que dan origen a}$$

$$F_{\text{cal}} = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}$$

esta relación de la varianza explicada a la varianza no

explicada, nos sirve para efectuar la prueba de la significación global de la regresión. Si la relación F calculada excede el valor tabulado de F a un nivel especificado de significación y grados de libertad, se acepta la hipótesis de que los parámetros de la regresión no son todos iguales a cero y que R^2 es significativamente diferente de cero.

II.5. PROCEDIMIENTO.

Como se comenta en el Capítulo 1, la información utilizada fue aquella que describía mejor a la región seleccionada para el estudio y una vez establecido el número de variables se procedió a su codificación, integrando una matriz de 80x40; 40 características de cada uno de los 80 municipios considerados, medidas en una escala de intervalo.

El contar con la matriz de datos y con el modelo descrito en éste Capítulo, como herramienta de análisis para describir el comportamiento conjunto de dichas variables, permite multitud de posibilidades con las cuales intentar explicar la situación económica, social y política que caracteriza el sector ejidal en el Estado de Hidalgo ; sin embargo, en el presente trabajo la exposición se limita a cuatro aplicaciones de Análisis de Regresión Múltiple.

Para realizar el análisis descrito en las secciones anteriores utilicé el SPSS (Statistical Package for the Social Sciences : paquete estadístico para las ciencias sociales) versión 6, constituido por módulos que reciben el nombre de subprogramas. El subprograma "Regression" efectúa análisis de regresión y a solicitud del usuario calcula los estadísticos que permiten analizar que tan correcto es el ajuste y que tan válidos son los valores calculados ; entre éstos se encuentran los valores de F para efectuar pruebas de hipótesis sobre los coeficientes, el estadístico de Durbin-Watson que permite determinar si hay o no autocorrelación serial de primer orden, así como también

calcula y grafica los residuales.

La inclusión de variables independientes a la ecuación de regresión, fue por el procedimiento de selección Stepwise con un sólo nivel de inclusión para toda la lista de variables. El procedimiento consiste en ir incluyendo variables de una en una en cada paso hasta que la ecuación de regresión es satisfactoria ; el orden de la inclusión es determinado utilizando el coeficiente de correlación parcial, como una medida de la importancia de la variable predictiva en potencia ; es decir :

1) Seleccionamos la X_i (X_1) más correlacionada con Y , obteniendo $\hat{Y} = f(X_1)$, verificamos su significancia; si la prueba es no significativa el proceso se concluye, en caso contrario se retiene como variable predictiva y continuamos con el paso siguiente.

2) Examinamos los coeficientes de correlación parcial de todas las predictivas que no han sido incluidas a la ecuación y seleccionamos X_j que tenga el mayor coeficiente en valor absoluto. El valor parcial de F para las j variables incluidas en la ecuación es examinado; el menor de estos valores de F es comparado con una F porcentual apropiado y la correspondiente variable predictiva es retenida o rechazada, dependiendo si la prueba es significativa o no significativa. Esta prueba del último predictor útil en la ecuación es efectuada en cada paso del procedimiento Stepwise.

Un predictor puede ser el mejor candidato de entrada en un paso del procedimiento y en uno posterior ser removido del modelo, dada su correlación con las variables actuales en la regresión, mediante la evaluación de los valores parciales de F (el menor). Lo anterior establece un juicio sobre la contribución de los últimos valores en la regresión en ese paso, tratada como si ella fuera la variable más recientemente introducida ; si la prueba da como resultado una contribución no significativa, ella es removida del modelo, continuando el proceso con la examinación de la

mejor de las variables fuera del modelo, si no puede tener lugar en la ecuación el proceso se concluye.

El subprograma "Regression" permite especificar tres criterios estadísticos para regresiones Stepwise, con la finalidad de separar un subconjunto de variables predictivas evaluadas, que producirán una ecuación de predicción óptima con tan pocos términos como sea posible. Los criterios estadísticos son :

1) El número máximo de variables independientes a entrar en una ecuación de regresión, siempre y cuando cumpla los otros criterios. En este trabajo, dicho número se representa mediante n y se consideró $n = 20$.

2) El valor mínimo de una estadística F parcial para aceptar la inclusión o exclusión de una variable en el modelo. En este trabajo se denota mediante F y se consideró $F = 3.08$, que corresponde a un valor en tablas de $\alpha = 0.05$ y (1,70) grados de libertad.

3) La tolerancia de una variable considerada para inclusión, es la proporción de la varianza de la variable no explicada por las variables independientes ya incluidas en la ecuación y su rango posible es de 0 a 1 ; una tolerancia de cero indicará que la variable es una combinación lineal perfecta de las ya incluidas ; una tolerancia de uno indicará que dicha variable es no correlacionada con las otras variables independientes ya incluidas en la ecuación. En este trabajo se identificó con la letra T , estableciendo $T = 0.05$ como restricción.

El procedimiento básico mediante este subprograma consiste en:

1) De la matriz de correlación, elaborada en sus primeros cálculos, selecciona la variable cuyo coeficiente de correlación es más significativo y proporciona la primera tabla de resultados con R^2 , $\hat{\sigma}_e$, $\hat{\beta}$ y realiza el análisis de varianza, que en este primer paso la prueba para $\hat{\beta}_1$ es igual que para la regresión, de donde la razón F sirve para

comprobar si la variable independiente tiene una relación significativa con la dependiente si

$$F = \frac{r^2 / k}{(1-r^2)/(n-k)} > F_{(ab)}$$

es decir cae en la zona crítica, se

acepta la hipótesis alternativa, o sea, el coeficiente de correlación es significativamente diferente de cero y, por lo tanto, hay una relación significativa entre las variables.

2) En base a la variable incluida, calcula los coeficientes de correlación parcial, así como los valores de F y T. Las variables a incluirse en los siguientes pasos, son analizadas considerando los criterios estadísticos establecidos; calculando después de cada inclusión una nueva tabla con los valores correspondientes $\hat{\beta}$, R^2 , T y F), para ser evaluados en el siguiente paso.

3) Los valores de R^2 , σ_e y $\hat{\beta}$ van variando en cada paso conforme la ecuación de predicción contiene más variables; en el caso de $\hat{\beta}_1$, su magnitud es diferente una vez que X_1 es confundida con los efectos de las otras variables predictivas incluidas a la ecuación en pasos posteriores, debido a la falta de una total independencia lineal entre tales variables (esto en estadística recibe el nombre de multicolinealidad). Por otra parte, en el caso de la regresión múltiple, el subprograma proporciona dos valores de F; uno corresponde a la variable que se incluye y el otro a la regresión total hasta ese punto.

4) Cuando ya no hay ninguna variable independiente que incluir, porque se haya llegado a la n deseada o las restantes no cumplen con el valor de F y T establecido, el subprograma proporciona un cuadro resumen de resultados.

5) Finalmente, para la última regresión imprime una tabla con los valores de Y , \hat{Y} y e , grafica \hat{Y}_i con el correspondiente valor e_i y proporciona el valor de la prueba de Durbin-Watson, útiles en análisis posteriores.

El examen de residuales es una herramienta auxiliar para mejorar el análisis de los datos a los cuales se aplica el

Análisis de Regresión; al analizar los residuales obtenidos de una primera aplicación, se busca determinar si es necesario alterar el modelo para una segunda aplicación, a partir de la cual se puedan obtener mejores resultados. Si el modelo es correcto, los residuales deben exhibir tendencias que tiendan a confirmar las suposiciones hechas, o al menos no deben exhibir una negación de tales suposiciones. Una forma de analizar los residuales para verificar los supuestos en el modelo es gráficamente. Al graficar contra \hat{Y}_i , dado que e_i y \hat{Y}_i no son correlacionados, se forma una banda. (En este caso el subprograma la proporciona en forma vertical, de -2 a +2 desviaciones estandar), si es uniforme nos indica no anomalía y por lo tanto su análisis de mínimos cuadrados no debe ser invalidado; el obtener valores muy grandes fuera de esta banda, nos indica que deberán recibir un tratamiento especial dependiendo de la causa que se considere más probable de su valor y de cual sea el objetivo del análisis estadístico.

Con el propósito de abundar en este tipo de análisis, presento en cada aplicación, gráficas adicionales que permiten validar los supuestos del modelo.

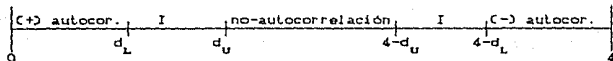
El estadístico de Durbin-Watson que nos sirve para determinar si existe autocorrelación; se calcula con la siguiente fórmula:

$$d = \frac{\sum_{t=1}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

donde e_t = residual en la observación t

e_{t-1} = residual en la observación t-1

el valor calculado de d oscila entre 0 y 4; sin autocorrelación cuando d está próximo a 2, es decir, se comprueba la naturaleza aleatoria de los datos



6) Es importante señalar, que el subprograma "Regression", realiza el procedimiento con variables estandarizadas, con lo que el término independiente del modelo descrito en (2) desaparece, con los mismos supuestos establecidos en el modelo y con el único cambio que se produce en la fórmula de R^2 , que se convierte en

$$R^2 = \frac{\hat{\beta}' X' Y}{Y' Y}$$

CAPITULO III. RESULTADOS OBTENIDOS

En base a lo señalado en el "Procedimiento", sección II.5, la exposición del trabajo se limita a cuatro aplicaciones del análisis de regresión múltiple, efectuadas con un mismo tipo de diseño del subprograma "Regression" del SPSS y eligiendo como variables dependientes (una para cada caso) las que se consideraron más representativas para tratar de explicar el "Problema"

Por otra parte, en la presentación de los resultados obtenidos del "Modelo", sólo se realizan con mayor profundidad en la primera aplicación por la similitud con los tres casos restantes.

III.1. Presentación e interpretación de resultados.

Variable dependiente : X_{15} -PROINS- Evidos con desaprovechamiento de áreas laborables por problemas institucionales.

Cuadro 1
Variables incluidas en la ecuación

Variable No.	Ident.	$\hat{\beta}_i$	$\sqrt{\text{Var}(\hat{\beta}_i)}$	T	F	α
16	Pronat	.41147	.08032	0	26.24	0
11	Pobind	.27867	.08857	.69	9.90	.002
17	Deslin	.11510	.05536	.68	4.32	0.41
33	Insmej	-.22003	.06053	.58	15.60	0
27	Orvent	.41093	.14874	.81	7.63	.007
2	Suptot	.12622	.05724	.30	4.86	.031

T = Índice de tolerancia con el que se incluyó X_i

F = Valor parcial de F que determinó la aceptación de X_i

α = Nivel de significancia descriptivo.

Un coeficiente de regresión parcial β_i representa el cambio esperado en Y con un cambio de unidad en X_i , cuando las demás variables son controladas; los efectos combinados son aditivos, es decir, al cambiar una unidad en cada una de las variables predictivas, el cambio esperado en Y podrá ser la suma de las β_i . En este caso en particular,

con los valores registrados en el Cuadro 1, tenemos que en el cálculo de los valores de "Proins" las variables con mayor peso son "Pronat" (Ejidos con desaprovechamiento de áreas laborables por problemas naturales, como son erosiones, inundaciones y/o falta de agua) y "Orvent" (Ejidos organizados para la venta de su producción), observándose que el efecto de "Insmej" (ejidos que usan insumos mejorados) en dicho cálculo es negativo.

Con los valores $\hat{\beta}_i$ asociados a las 6 variables predictivas seleccionadas, podemos representar nuestra ecuación de regresión:

$$\hat{Y}_i = \hat{\beta}_{10} X_{i10} + \hat{\beta}_{11} X_{i11} + \hat{\beta}_{17} X_{i17} + \hat{\beta}_{33} X_{i33} + \hat{\beta}_{27} X_{i27} + \hat{\beta}_2 X_{i2}$$

los valores \hat{Y} son calculados sustituyendo en la ecuación los valores correspondientes para cada observación (ver Cuadro 6). Podemos representar las n ecuaciones por :

$$Y = X\beta + \varepsilon = X\hat{\beta} + e$$

en donde

$$E(\varepsilon) = 0$$

$$\text{Var}(\varepsilon) = \sigma^2 I$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

$$\text{para } i \neq j$$

$$E(Y) = X\beta$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\text{Var}(\hat{\beta}_i) = E[(\hat{\beta}_i - \beta_i)^2]$$

$$\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = E[(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)]$$

la matriz simétrica de varianzas y covarianza de las $\hat{\beta}$ $E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ contiene en su diagonal las varianzas y en sus lugares restantes covarianzas (ver cuadro 2) ; la raíz cuadrada del i-ésimo término de la diagonal nos da el error estandar de $\hat{\beta}_i$ estimada, es decir, la variabilidad en su valor.

Cuadro 2
Matriz de Varianza / Covarianza

Suptot	.00328					
Pobind	-.00114	.00784				
Pronat	-.00084	-.00272	.00645			
Deslin	-.00154	-.00011	-.00058	.00306		
Orvent	.00110	.00037	-.00261	-.00067	.02212	
Insmej	-.00142	.00073	-.00025	-.00066	.00226	.00365
	Suptot	Pobind	Pronat	Deslin	Orvent	Insmej

Además del interés por conocer como se determinan los valores de "Proins" a partir de las variables independientes incluidas en la ecuación de predicción, es el medir el porcentaje de explicación en su variación en función a dichas variables, estos valores se presentan en el cuadro siguiente:

Cuadro 3
Cuadro de Resultados

Paso	Variabiles	R ²	r ² _{15(i,j--k)}	F-total (α=0)	$\hat{\sigma}_x$
1	16	.27600	.27600	96.62	20.11
2	16,11	.38423	.10823	62.94	18.67
3	16,11,17	.42461	.04038	48.13	18.16
4	16,11,17,33	.47915	.05055	39.27	17.46
5	16,11,17,33,27	.51366	.03651	34.54	16.92
6	16,11,17,33,27,2	.54365	.02999	31.17	16.00

La exactitud total de la ecuación de predicción, proporción de la variación explicada por las variables incluidas en la ecuación, es reflejada por R² (coeficiente de determinación), es decir:

en el primer paso $R^2_{15,16} = \hat{\beta}'X'Y / Y'Y = 27.6\%$

en el paso 6 $R^2_{15,16,11,17,33,27,2} = 54.365\%$

que nos indica que el 54.4% de la variación de "Proins" es explicada por las 6 variables operando conjuntamente.

El cuadrado del cambio producido por la inclusión de cada

variable se registra en la columna de r^2 . Con este coeficiente podemos visualizar si una variable contribuye sustancialmente a la variación explicada por las variables ya incluidas o si su contribución es más bien trivial.

La exactitud de predicción en unidades absolutas está reflejada por el error estándar de estimación y los valores resultantes se registran en la columna de $\hat{\sigma}_e$; en el caso del paso 8 (última ecuación de predicción obtenida), nos indica que las estimaciones de Y_i se desviarán de los puntajes reales por un factor proporcional a 16.50 unidades sobre su escala de medición. Esto depende de la confianza con que se desea realizar la estimación.

$$S^2 = \frac{e'e}{n-k} = \frac{Y'Y - \beta'X'Y}{n-k} = (16.50)^2$$

donde S^2 es un estimador insesgado de σ^2

Por otra parte, las pruebas efectuadas para las β se desarrollaron en base al método conocido como "Método de Regresión Estandar", en donde cada variable es tratada como si hubiera sido aumentada a la ecuación de regresión en un paso separado después de que las otras variables predictivas han sido incluidas. Los valores F obtenidos, se presentan en el cuadro 1; si les comparamos con la F de tablas con 1 y 74 grados de libertad, tenemos que los 8 coeficientes de regresión son significativos a un nivel del 0.05. Del Capítulo anterior tenemos que para verificar la hipótesis de que $\beta_k = 0$, es decir X_k no ejerce influencia lineal sobre Y , podemos calcular:

$$F_{\text{cal}} = \frac{\hat{\beta}_k^2}{(\sum_i e_i^2 / (n-k)) a_{kk}}$$

en donde a_{kk} es el k -ésimo elemento de la diagonal principal de $(X'X)^{-1}$

como $\text{Var}(\hat{\beta}_k) = a_{kk} \sigma^2$, de los resultados registrados en los Cuadros 2 y 3 podemos calcular manualmente:

$$F_{\text{cal}} = \frac{\hat{\beta}_{10}^2}{(e'e / 74)_{10 \ 10}} = \frac{(0.41147)^2}{(16.50)^2 [0.00845 / (16.50)^2]} = 26.24$$

$$F_{\text{cal}} = \frac{\hat{\beta}_{11}^2}{(16.50)^2 [0.00784 / (16.50)^2]} = \frac{(0.27867)^2}{0.00784} = 9.80$$

La hipótesis nula $H_0: R=0$, es equivalente a la hipótesis de que todos los k coeficientes de regresión son iguales a cero:

$$H_0: \beta_{10} = \beta_{11} = \beta_{17} = \beta_{33} = \beta_{27} = \beta_2 = 0$$

si la hipótesis es $\beta = \gamma$ en donde $C = I$ y $\gamma = 0$; (18) se convierte en:

$$F_{\text{cal}} = \frac{\hat{\beta}'X'X\hat{\beta}}{Y'[I - X(X'X)^{-1}X']Y} \frac{n-k}{k}$$

En el siguiente cuadro se presenta el análisis de varianza del R^2 total obtenido:

Cuadro 4
Análisis de Varianza (última ecuación)

F.V.	G.L.	S.C.	C.M.	F	α
Regresión	6	50894.59	8482.43	31.17	0
Residual	74	20140.41	272.17		

el valor en tablas para F con (6,74) grados de libertad y un 99% ($\alpha=0.01$) esta entre 2.21 y 2.32. Nos indica que la probabilidad de obtener $F_{\text{tab}} \geq 31.17$ es practicamente cero; de donde se puede afirmar que es improbable que la muestra se haya obtenido de una población en que $R = 0$ y por lo tanto, hay una relación significativa entre las variables.

Si un subconjunto de variables explican una proporción sustancial de variación en una variable dependiente, no podrá esperarse que con la adición a la ecuación de predicción de otra variable se logre incrementar a R^2 mucho más, pero si se podrá observar si su contribución, cuando se use el conjunto completo, es no importante. En el Cuadro 5, se presentan los resultados obtenidos

correspondientes a dichas variables, que no contienen los criterios estadísticos establecidos, pero que pueden ser incluidas a la ecuación mediante la eliminación o disminución de alguno de tales criterios.

El cuadrado del coeficiente de correlación parcial $r_{15i.j---k}^2$ representa el incremento proporcional en la variación explicada si X_i se incluyera a la ecuación en un siguiente paso, expresado como una proporción de la variación no explicada. Es decir, en un siguiente paso pueden ser incluidas tres variables más (8-Ejipea, 12-Peseje y 32-Magric) con α 5.10, incrementando en 6.2% la variación explicada por el primer subconjunto.

De los resultados obtenidos tenemos que Proins--Ejidos con desaprovechamiento de áreas laborables por problemas institucionales, como son la falta de crédito, ausentismo de ejidatarios, deslindes, conflictos por linderos, depuración censal, sucesión de derechos agrarios, ejecuciones presidenciales y/o despojo de tierras, es explicada principalmente por Pronat--Ejidos con desaprovechamiento de áreas laborables por problemas naturales y Pobind--Ejidos con población indígena; que operando conjuntamente las 8 variables, el porcentaje de explicación en su variación es de 60.6%. Sin embargo, dicho porcentaje se considera "pequeño", debiendo existir otras variables u otro tipo de relación que nos de un mayor grado de explicación.

Cuadro 5
Variables no incluidas en la ecuación

VARIABLE		r _{15i. j---k}	T	F	a
No.	IDENT.				
1	EJCEDO	.07982	.31	.47	.498
3	SUPLAB	.03678	.22	.10	.754
4	SUPRIE	-.05948	.49	.26	.612
5	EJIRIE	-.08807	.29	.57	.452
6	DOEJI	.00339	.85	.00	.977
7	DENPOB	.08446	.38	.52	.471
8	EJIPEA	-.22216	.20	3.79	.068
9	DERSAL	-.07051	.71	.36	.548
10	POBALF	-.15136	.11	1.71	.198
12	RESEJE	.20938	.80	3.35	.071
13	DOCBAS	-.18738	.21	2.66	.107
14	DEREGU	-.13503	.13	1.36	.248
18	CENSUC	-.07967	.16	.47	.497
19	EJEPRE	-.04180	.30	.13	.722
20	DESPOJ	-.16779	.44	2.11	.150
21	ORCRED	-.04707	.34	.16	.688
22	CREDOF	-.05729	.54	.24	.623
23	CREPRI	.04038	.69	.12	.731
24	ASI TEC	.03745	.24	.10	.750
25	ORPROD	-.09751	.36	.67	.414
26	ORINSU	.02112	.26	.03	.837
28	AFIPOL	-.06351	.10	.30	.588
29	AGRIEJ	.02179	.46	.03	.853
30	GANAEJ	-.01190	.29	.01	.919
31	FRUTEJ	.04722	.92	.16	.687
32	MAGRIC	-.20507	.27	3.20	.078
34	PEAPRI	-.12123	.20	1.09	.300
35	PEASEC	.04779	.34	.17	.684
36	EDUCEJ	.16801	.23	2.07	.155
37	AGULUZ	-.03490	.17	.09	.766
38	VIACOM	.01596	.11	.19	.892
39	MEDCOM	.00475	.27	.00	.968
40	VI VSER	-.02096	.16	.03	.858

Validación de supuestos distribucionales

1) La gráfica del Cuadro 8, e_i contra \hat{Y}_i , nos presenta a los residuales dentro de una banda uniforme que nos indica una relación libre de anomalías.

2) Los residuales graficados en un histograma y en el papel de probabilidad normal (ver Fig. 1 y 2), no dan evidencia de que se este violando el supuesto de normalidad.

3) En la Fig. 3 podemos observar que los residuales no tienden a alinearse, es decir no están correlacionados. Respecto al estadístico de Durbin-Watson, como $d = 2.24016 > d_u = 1.62$, nos indica que no hay evidencia de autocorrelación de primer orden con 1% de nivel de significación; sin embargo, con 5% nos ubica en la zona de indecisión ($d_u = 1.77$).

4) Al igual que en la gráfica de residuales contra \hat{Y}_i del Cuadro 8, al graficar contra las variables predictivas podemos observar que e_i se distribuye dentro de una banda horizontal de -2 a $+2$ desviaciones estandar (ver Fig. 4.1 a 4.8); sin embargo, su dispersión indica anomalías que requieren corregirse, como es el caso de:

Fig. 4.2 y 4.3. - Indica que la varianza de los residuales depende del valor de la variable predictiva graficada en el eje horizontal; es decir, los datos dan evidencia de que se esta violando el supuesto de varianza constante, lo que puede conducir a estimaciones sesgadas e ineficientes de los errores estandar. Una posibilidad de superar este problema es mediante la transformación de Y.

Fig. 4.5 y 4.6. - Indica una relación lineal entre los residuales y la variable sobre el eje X debido a errores de cálculo. La posible solución a este problema sería:

- a) Los errores de medición en la variable dependiente se incorporen en el término de error o
- b) Reemplazar la variable predictiva sujeta a errores de medición.

CUADRO 6

MULTIPLE REGRESSION

OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-2SD	0.0	+2SD
1.	0	8.373873	-8.373873			
2.	10.00000	4.090087	5.909913			
3.	0	22.35115	-22.35115			
4.	14.00000	21.88509	-0.885088			
5.	50.00000	27.07173	22.92827			
6.	35.00000	37.69868	-2.698678			
7.	0	-5.467843	5.467843			
8.	7.000000	2.145068	4.854932			
9.	0	19.49494	-19.49494			
10.	0	-4.282557	4.282557			
11.	60.00000	36.90097	23.09903			
12.	10.00000	9.479012	.5209876			
13.	0	2.138017	-2.138017			
14.	58.00000	44.07440	13.92560			
15.	0	10.72553	-10.72553			
16.	0	10.07535	-10.07535			
17.	44.00000	31.17139	12.82862			
18.	0	29.41169	-29.41169			
19.	0	27.01540	-27.01540			
20.	0	-2.529349	2.529349			
21.	10.00000	15.19136	-5.191359			
22.	25.00000	13.79709	11.20291			
23.	21.00000	13.13385	7.866147			
24.	17.00000	19.79800	-22.79800			
25.	75.00000	46.22905	28.77095			
26.	0	14.87606	-14.87606			
27.	91.00000	69.58599	21.41401			
28.	23.00000	27.87778	-4.877779			
29.	38.00000	16.53269	21.46731			
30.	25.00000	23.79517	1.204828			
31.	0	14.90065	-14.90065			
32.	50.00000	28.13865	21.86135			
33.	0	11.36130	-11.36130			
34.	33.00000	12.24882	20.75118			
35.	0	7.587958	-7.587958			
36.	20.00000	11.62891	8.371094			
37.	50.00000	16.97652	33.02348			
38.	0	-3081781	-3081781			
39.	50.00000	54.21680	-4.216800			
40.	29.00000	27.54889	1.451115			
41.	13.00000	21.48057	-8.480549			
42.	0	11.73130	-11.73130			
43.	77.00000	55.14818	21.85182			
44.	0	41.72904	-41.72904			
45.	13.00000	14.48010	-1.480098			
46.	0	-9862682	-9862682			
47.	0	17.03226	-17.03224			
48.	33.00000	32.53201	-4779914			
49.	0	17.17950	-17.17950			
50.	22.00000	30.65448	-8.654478			
51.	22.00000	39.56636	-17.56636			
52.	0	-8.649318	8.649318			

45

CUADRO 6

MULTIPLE REGRESSION

OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-RSD	0.0	*2SD
1.	0	8.373873	-8.373873			
2.	10.00000	4.090887	5.909113			
3.	0	22.35115	-22.35115			
4.	14.00000	21.88509	-8.885088			
5.	50.00000	27.07173	22.92827			
6.	35.00000	37.69868	-2.698678			
7.	0	-6.467843	6.467843			
8.	7.000000	2.145068	4.854932			
9.	0	19.49494	-19.49494			
10.	0	-4.282557	4.282557			
11.	60.00000	36.99097	23.00903			
12.	10.00000	9.479012	.5209876			
13.	0	2.138017	-2.138017			
14.	58.00000	44.07440	13.92560			
15.	0	10.72553	-10.72553			
16.	0	10.07535	-10.07535			
17.	44.00000	31.17138	12.82862			
18.	0	29.61169	-29.61169			
19.	0	27.01540	-27.01540			
20.	0	-2.529349	2.529349			
21.	10.00000	15.19136	-5.191359			
22.	25.00000	12.79709	11.20291			
23.	21.00000	13.13385	7.866147			
24.	17.00000	19.79800	-22.79800			
25.	75.00000	46.22905	28.77095			
26.	0	14.87606	-14.87606			
27.	91.00000	69.58590	21.41410			
28.	23.00000	27.87778	-4.877779			
29.	38.00000	16.53269	21.46731			
30.	25.00000	23.79517	1.204828			
31.	0	14.90065	-14.90065			
32.	50.00000	28.13865	21.86135			
33.	0	11.36130	-11.36130			
34.	33.00000	12.24862	20.75138			
35.	0	7.587958	-7.587958			
36.	20.00000	11.62891	8.371094			
37.	50.00000	16.97652	33.02348			
38.	0	-30.1781	-30.1781			
39.	50.00000	54.21680	-4.216800			
40.	29.00000	27.54885	1.451155			
41.	13.00000	21.48057	-8.480565			
42.	0	11.7730	-11.773030			
43.	77.00000	55.15838	21.84162			
44.	0	41.72904	-41.72904			
45.	13.00000	14.8210	-1.820222			
46.	0	-9802682	-9802682			
47.	0	17.03226	-17.03226			
48.	33.00000	32.43201	.4679916			
49.	0	17.17950	-17.17950			
50.	22.00000	30.65448	-8.654478			
51.	22.00000	39.56616	-17.56616			
52.	0	-8.649318	8.649318			

CUADRO 6

continuación

..... MULTIPLE REGRESSION

OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-2SD	0.0	+2SD
53.	0	7.019895	-7.019895			
54.	33.00000	14.47635	18.52365			
55.	24.00000	10.04257	5.957431			
56.	20.00000	17.2459	2.975413			
57.	0	2.678781	-2.678781			
58.	58.00000	35.21953	22.78047			
59.	10.00000	-6.694927	16.69493			
60.	0	10.61545	-10.61545			
61.	0	15.32323	-15.32323			
62.	17.00000	31.92790	-14.92790			
63.	0	9.912525	-9.912525			
64.	0	21.50945	-21.50945			
65.	0	-1.967428	1.967428			
66.	0	-5.823194	5.823194			
67.	0	1.204845	-1.204845			
68.	0	16.21919	-16.21919			
69.	99.00000	43.63054	35.36946			
70.	0	-12.06633	12.06633			
71.	50.00000	23.41385	26.58615			
72.	12.00000	10.71582	1.284180			
73.	0	-2.008598	2.008598			
74.	7.000000	46.52600	-39.52600			
75.	0	10.43282	-10.43282			
76.	43.00000	41.29838	1.701620			
77.	25.00000	9.162026	15.83797			
78.	0	17.73385	-17.73385			
79.	15.00000	16.34692	-1.346924			
80.	29.00000	27.10031	1.899690			

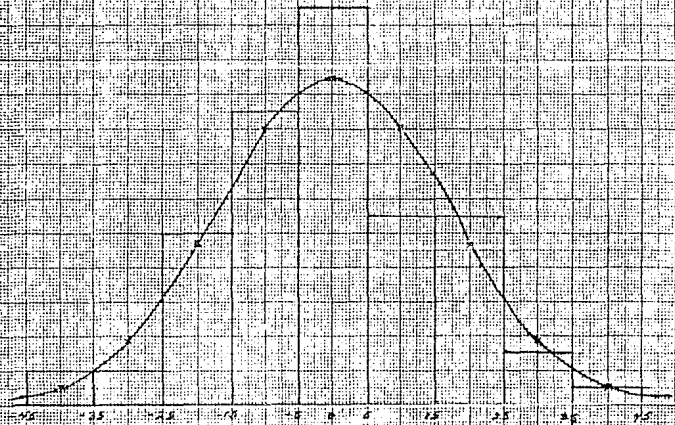
46

NOTE - (*) INDICATES ESTIMATE CALCULATED WITH MEANS SUBSTITUTED
R INDICATES POINT OUT OF RANGE OF PLOT

NUMBER OF CASES PLOTTED	80.
NUMBER OF 2 S.D. OUTLIERS	4. OR 5.00 PERCENT OF THE TOTAL
VON NEUMANN RATIO	2.21881
DURBIN-WATSON TEST	2.24014
NUMBER OF POSITIVE RESIDUALS	39.
NUMBER OF NEGATIVE RESIDUALS	41.
NUMBER OF RUNS OF SIGNS	45.
EXPECTED NUMBER OF RUNS OF SIGNS	41.
EXPECTED S.D. OF RUN DISTRIBUTION	4.44893
UNIT NORMAL DEViate-	
Z = (EXPECTED-OBSERVED)/S.D.	1.69447
PROBABILITY OF OBTAINING .4E. ABS(Z)	.04509

GRAFICA DE RESIDUALES

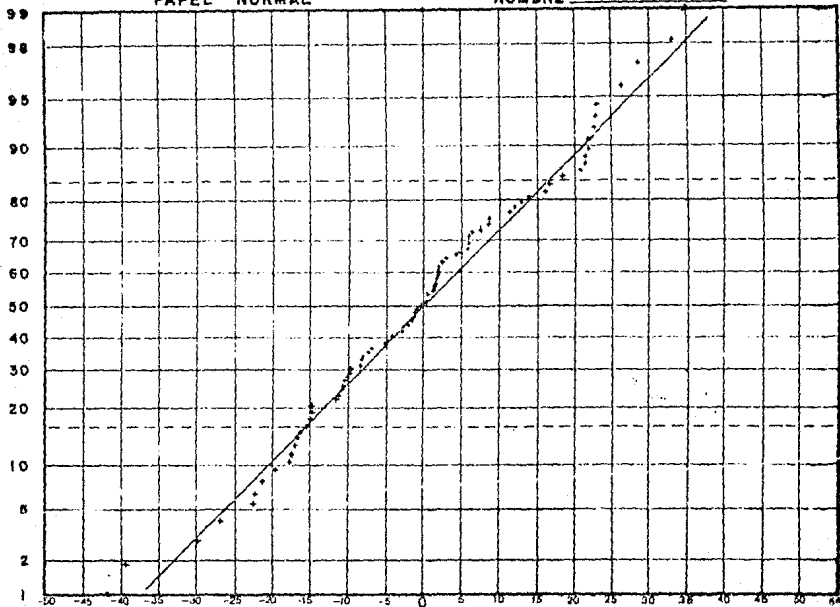
FIG. 1



LABORATORIO DE ESTADISTICA

PAPEL NORMAL

NOMBRE FIG. 2

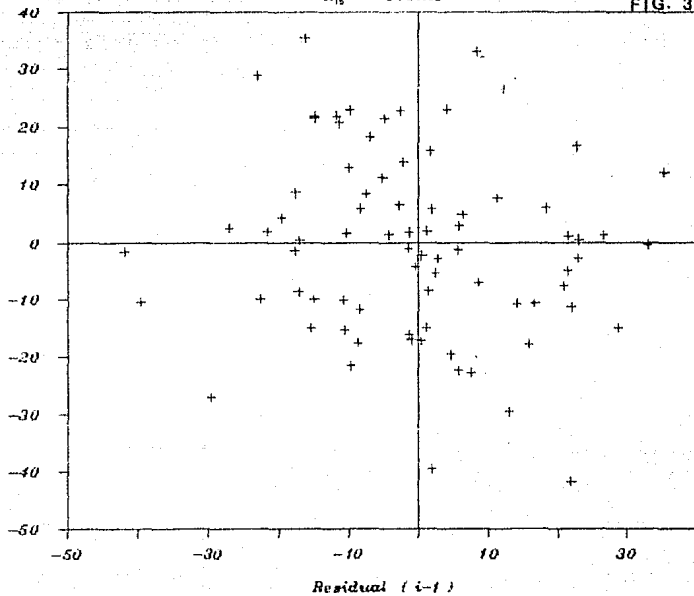


GRAFICA DE RESIDUALES

X_{15} -- Proins

FIG. 3

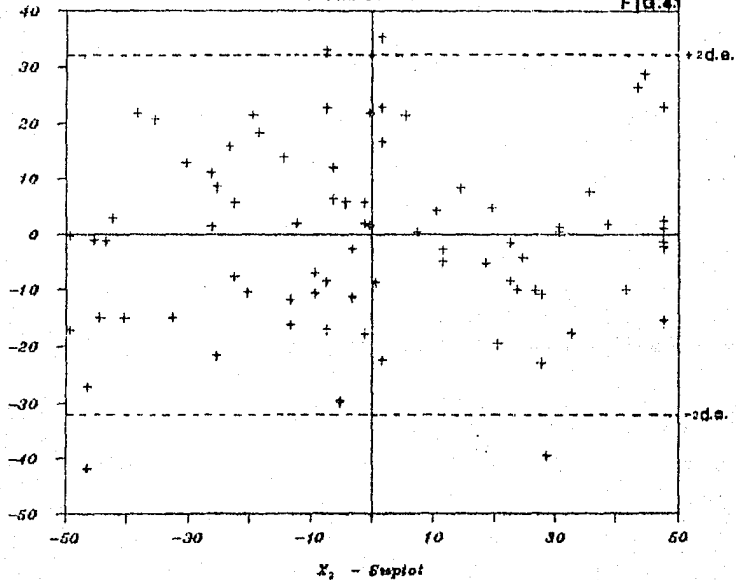
Residual (i)



GRAFICA

RESIDUAL/PREDICTIVA

FIG. 41

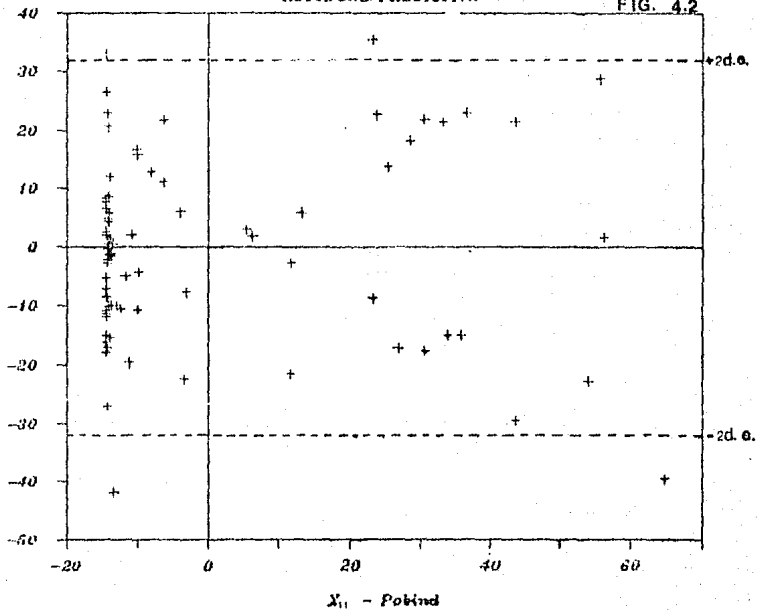


50
RESIDUAL

GRAFICA

RESIDUAL/PREDICTIVA

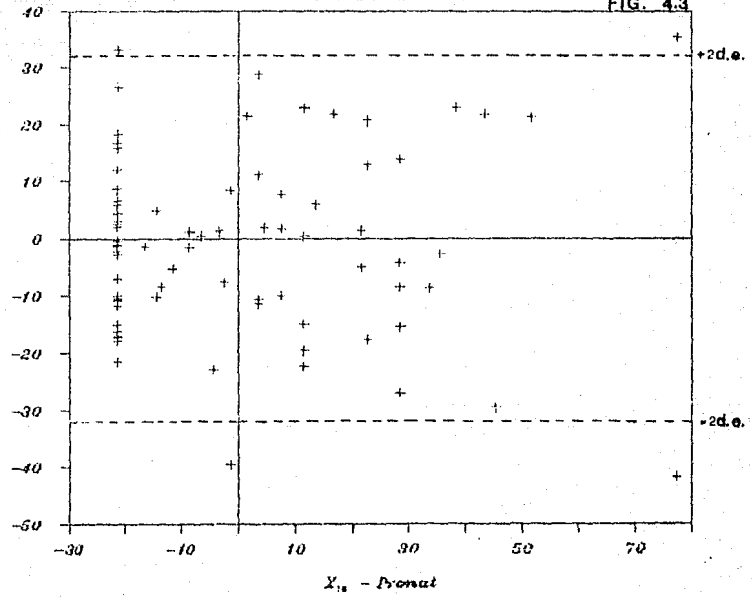
FIG. 4.2



GRAFICA

RESIDUAL/PREDICTIVA

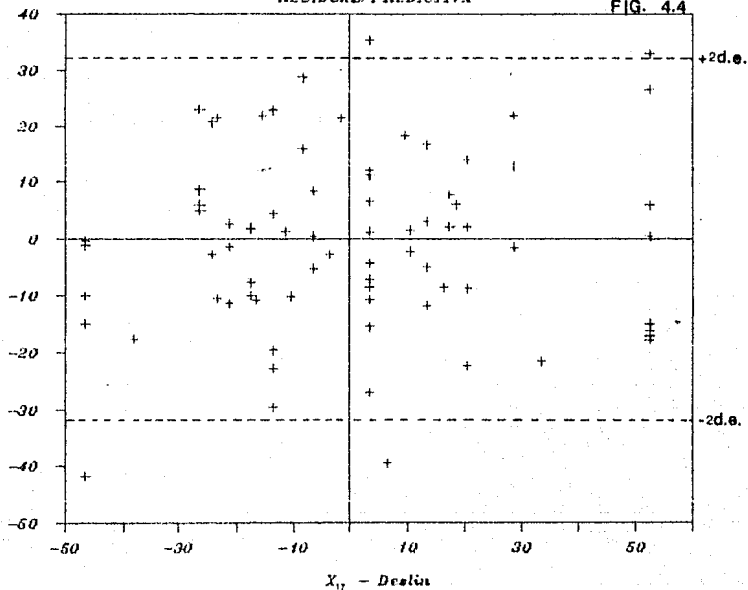
FIG. 4.3



GRAFICA

RESIDUAL/PREDICTIVA

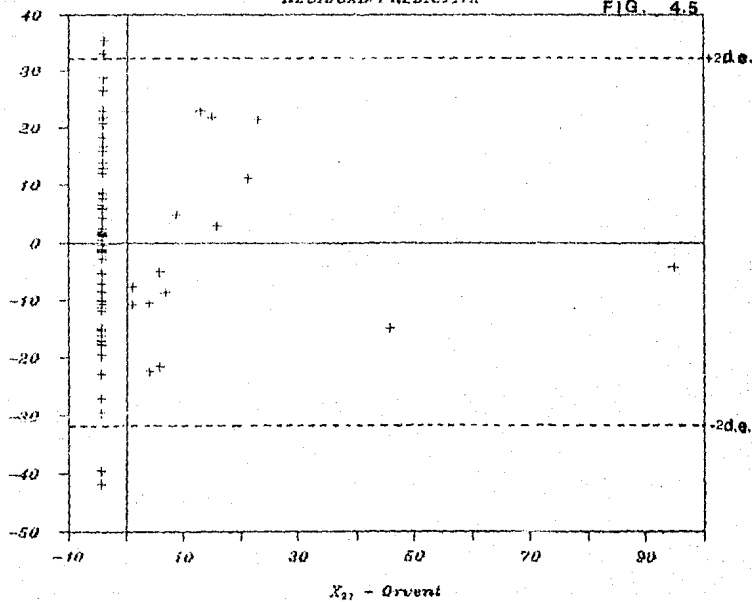
FIG. 4.4



GRAFICA

RESIDUAL/PREDICTIVA

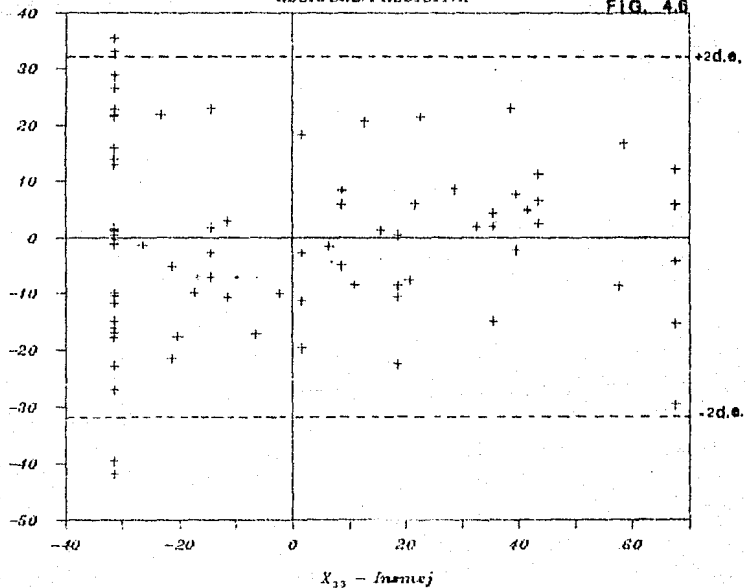
FIG. 4.5

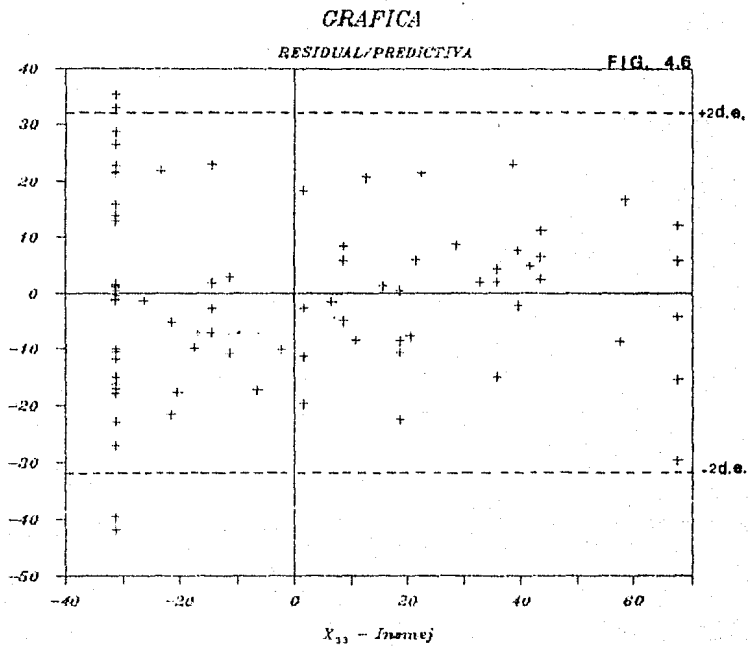


GRAFICA

RESIDUAL/PREDICTIVA

FIG. 4.6





Variable dependiente : X_{21} = Orcred--Ejidos organizados para obtener crédito .

Cuadro 7
Variables incluidas en la ecuación

Variable		$\hat{\beta}_i$	Var($\hat{\beta}_i$)	T	F	α
No.	Ident.					
22	Credof	.98426	.08903	0	122.23	0
23	Crepri	.82729	.12075	.95	46.94	0
27	Orvent	.31116	.08024	.75	15.04	0
32	Magric	.08608	.02423	.46	12.62	.001

T = Índice de tolerancia con el que se incluyó X_i

F = Valor parcial de F que determinó la aceptación de X_i

α = Nivel de significancia descriptivo.

donde:

$$\hat{Y}_i = \hat{\beta}_{22} X_{i22} + \hat{\beta}_{23} X_{i23} + \hat{\beta}_{27} X_{i27} + \hat{\beta}_{32} X_{i32}$$

Es decir, la organización para obtener crédito en un ejido es una característica de aquellos que reciben crédito de bancos oficiales, particulares u organismos privados, están organizados para la venta de su producción y utilizan maquinaria en su actividad agrícola.

Sustituyendo en la ecuación, obtenemos los valores estimados de Y, que se presentan en el cuadro 10.

Cuadro 8
Cuadro de resultados

Paso	Variable	R ²	$r_{21(i,j--k)}$	F-total $\alpha = 0$	$\hat{\sigma}_e$
1	22	.52746	.52746	191.94	13.85
2	22,23	.75169	.22423	215.54	10.11
3	22,23,27	.79532	.04364	177.57	9.23
4	22,23,27,32	.82448	.02915	156.43	8.61

Como se observa en el Cuadro 8, el porcentaje de explicación en la variación de Orcred determinado por las 4 variables incluidas en la ecuación de predicción es significativo y el

incremento en R^2 por la inclusión, en un siguiente paso, de otra variable predictiva no sería sustancial. Por otra parte, las pruebas efectuadas nos indican que los coeficientes estimados son significativos para un α prácticamente igual a cero.

Cuadro 9
Análisis de Varianza (última ecuación)

F.V.	G.L.	S.C.	C.M.	F	α
Regresión	4	46359.95	11589.49	156.43	0
Residual	76	5530.04	74.08		

Es decir, si ejercen influencia lineal sobre Y ; que si cambiáramos una unidad en cada una de ellas, el cambio esperado podrá ser la sumatoria de las β_i .

Validación de supuestos distribucionales

- 1) La gráfica del Cuadro 10 indica una relación libre de anomalías.
- 2) Los residuales en el histograma de la Fig. 5 se comportan como las observaciones de una distribución normal con media cero; es decir, no dan evidencia de que se este violando el supuesto de normalidad.
- 3) En la Fig. 6 (papel normal) existe evidencia que indica falta de normalidad.
- 4) Al observar la Fig. 7, el supuesto de no correlación no queda muy claro que se cumpla, al presentar una evidencia de agrupamiento entre los residuales negativos. Sin embargo, el estadístico de Durbin-Watson nos indica que no existe autocorrelación en los valores observados.
- 5) Al analizar las Fig. 8.1 a 8.4, se observa una evidencia de que se esta violando el supuesto de varianza constante, combinado con agrupamientos de residuales en un determinado valor de la variable predictiva, posiblemente por errores de cálculo.

CUADRO 10

MULTIPLE REGRESSION

OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-2SD	0.0	+2SD
1.	33.00000	34.32387	-1.323869			
2.	0	0	0			
3.	33.00000	39.74190	-6.741895			
4.	14.00000	18.98515	-4.985150			
5.	17.00000	27.87582	-10.87582			
6.	9.000000	9.433129	-6.331287			
7.	33.00000	37.07892	-4.078916			
8.	27.00000	32.25258	-5.252579			
9.	0	5.767543	-5.767543			
10.	0	8.522190	-8.522190			
11.	0	0	0			
12.	35.00000	42.01322	-7.013215			
13.	0	4.111874	-4.111874			
14.	0	0	0			
15.	5.000000	13.79416	-8.794160			
16.	29.00000	20.17331	8.826694			
17.	13.00000	12.79544	0.2045562			
18.	0	8.522190	-8.522190			
19.	0	0	0			
20.	50.00000	53.81102	-3.811019			
21.	20.00000	26.20749	-8.207488			
22.	50.00000	45.51446	-15.51446			
23.	14.00000	8.497592	5.302408			
24.	0	0	0			
25.	0	0	0			
26.	0	0	0			
27.	27.00000	17.25973	9.740274			
28.	13.00000	7.183478	5.816522			
29.	7.000000	10.84966	-3.849660			
30.	0	0	0			
31.	0	0	0			
32.	0	0	0			
33.	17.00000	18.10790	-1.107900			
34.	22.00000	21.45303	0.3461720			
35.	0	5.257361	-5.257361			
36.	0	1.721655	-1.721655			
37.	0	0	0			
38.	0	0	0			
39.	99.00000	88.54036	10.45964			
40.	0	0	0			
41.	38.00000	16.06659	21.93341			
42.	40.00000	42.81391	-2.813905			
43.	8.000000	9.221209	-1.221209			
44.	0	0	0			
45.	25.00000	41.99984	-16.99984			
46.	0	0	0			
47.	0	4.304136	-4.304136			
48.	50.00000	21.20877	28.79123			
49.	0	0	0			
50.	33.00000	32.73794	0.2620403			
51.	11.00000	11.42949	-0.4294931			
52.	20.00000	26.57192	-6.571916			

85

R

R

CUADRO 10

continuación

OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-RSD	n.d	*250
53.	50.00000	39.82561	10.37439		1	
54.	11.00000	10.82691	-17.30860		1	
55.	18.00000	7.455079	10.54492		1	
56.	0	6.223217	-6.223217		1	
57.	11.00000	.6856618	10.31134		1	
58.	0	1.244643	-1.244643		1	
59.	20.00000	20.54613	-.5461254		1	
60.	0	2.152068	-2.152068		1	
61.	0	8.522190	-8.522190		1	
62.	33.00000	40.81273	-7.812735		1	
63.	0	0	0		1	
64.	0	3.111608	-3.111608		1	
65.	47.00000	41.00293	25.99707		1	R
66.	50.00000	49.88660	-1133968		1	
67.	0	0	0		1	
68.	33.00000	8.522190	24.47781		1	R
69.	0	0	0		1	
70.	50.00000	29.20440	20.79560		1	
71.	0	8.522190	-8.522190		1	R
72.	12.00000	5.595377	6.404623		1	
73.	45.00000	33.68455	11.31545		1	
74.	0	0	0		1	
75.	0	2.489287	-2.489287		1	
76.	0	0	0		1	
77.	0	0	0		1	
78.	50.00000	57.73544	-7.735435		1	
79.	20.00000	8.177859	11.82214		1	
80.	0	.2582482	-.2582482		1	

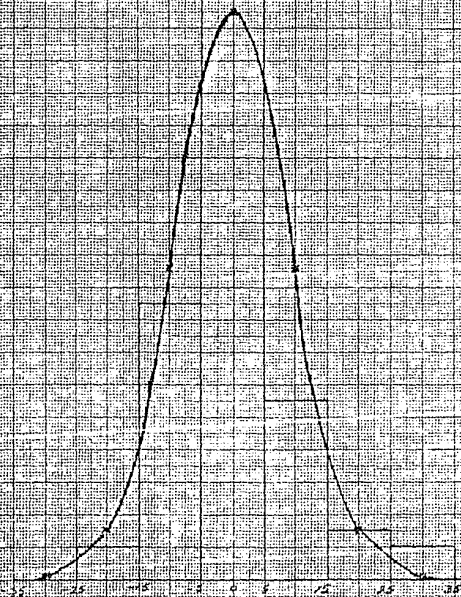
59

NOTE - (+) INDICATES ESTIMATE CALCULATED WITH MEANS SUBSTITUTED
R INDICATES POINT OUT OF RANGE OF PLOT

NUMBER OF CASES PLOTTED 80.
 NUMBER OF 2 S.D. OUTLIERS 5. OR 6.25 PERCENT OF THE TOTAL
 VON NEUMANN RATIO 2.15117 DURBIN-WATSON TEST 2.12428
 NUMBER OF POSITIVE RESIDUALS 43.
 NUMBER OF NEGATIVE RESIDUALS 37.
 NUMBER OF RUNS OF SIGNS 22.
 EXPECTED NUMBER OF RUNS OF SIGNS 41.
 EXPECTED S.D. OF RUN DISTRIBUTION 4.41842
 UNIT NORMAL DEViate-
 Z=(EXPECTED-OBSERVED)/S.D. --28856
 PROBABILITY OF OBTAINING .GR. ABS(Z) .38646

GRAFICA DE RESIDUALES

FIG. 5

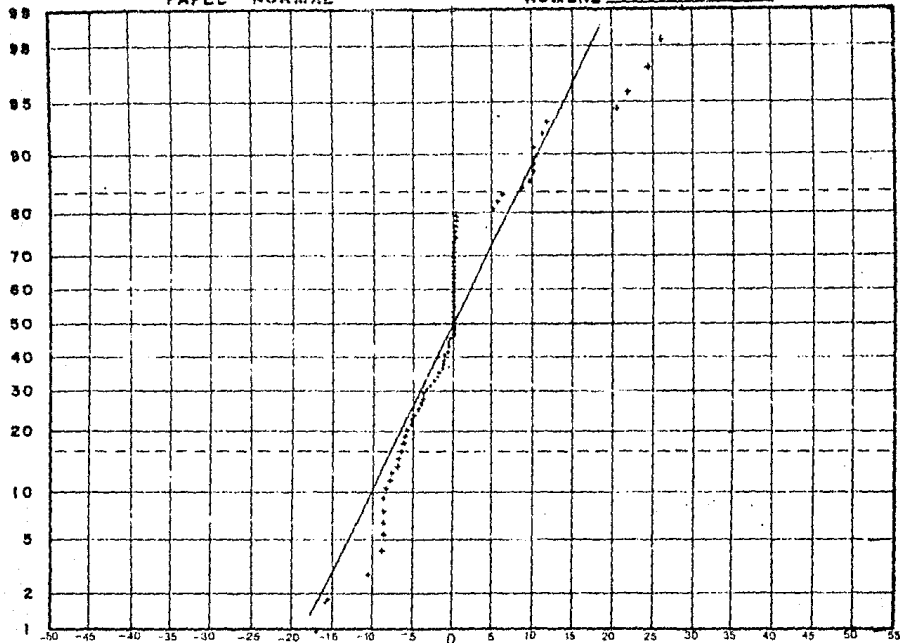


LABORATORIO DE ESTADISTICA

PAPEL NORMAL

NOMBRE _____ FIG. 6

61

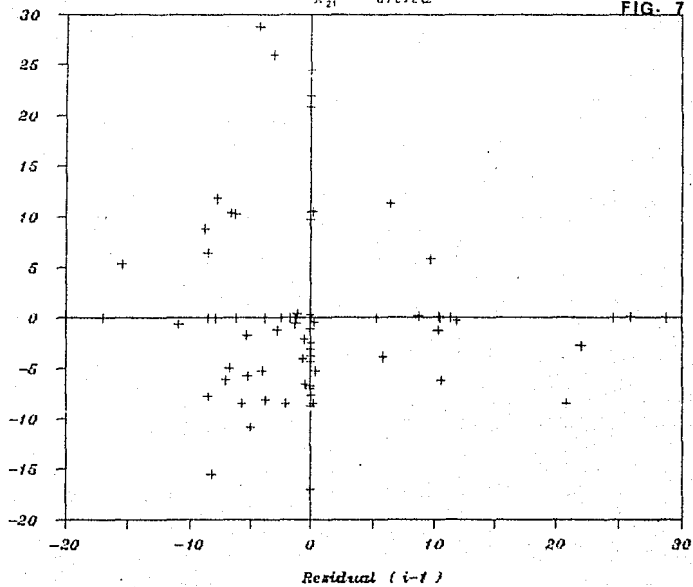


GRAFICA DE RESIDUALES

X_{21} — Overred

FIG. 7

62
Residual (i)



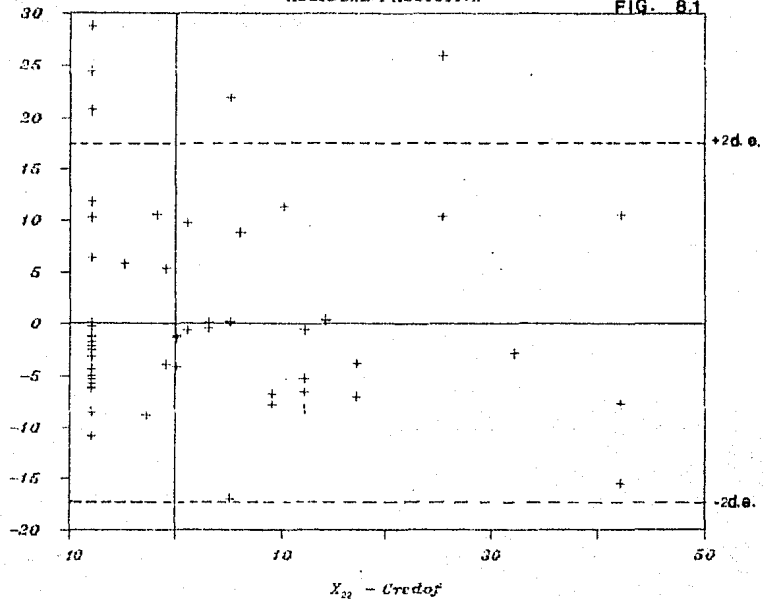
GRAFICA

RESIDUAL/PREDICTIVA

FIG. 8.1

88

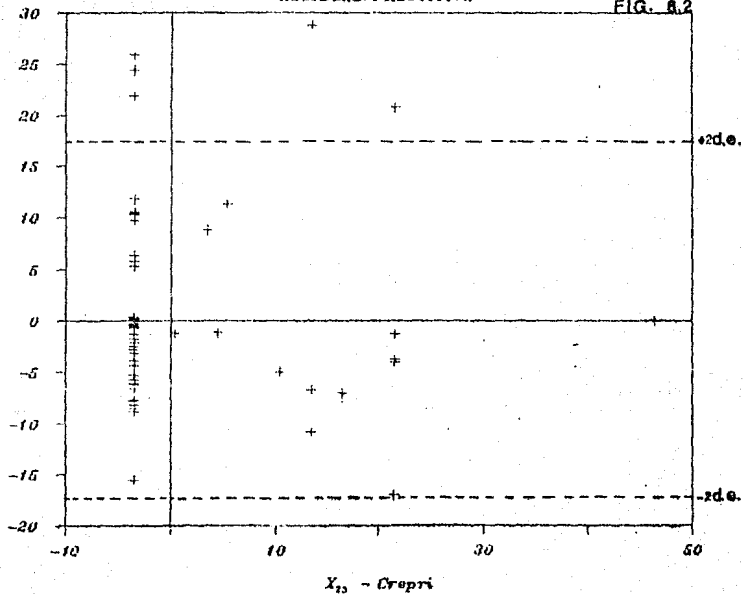
RESIDUAL



GRAFICA

RESIDUAL/PREDICTIVA

FIG. 8.2



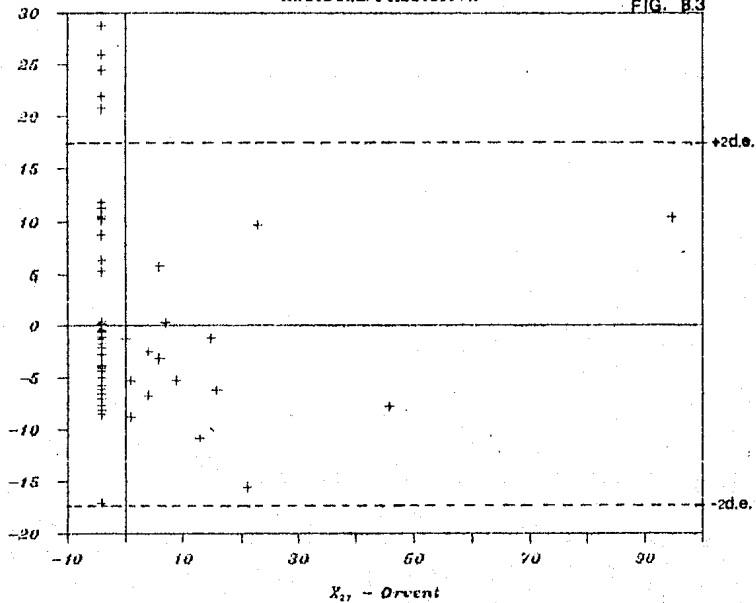
GRAFICA

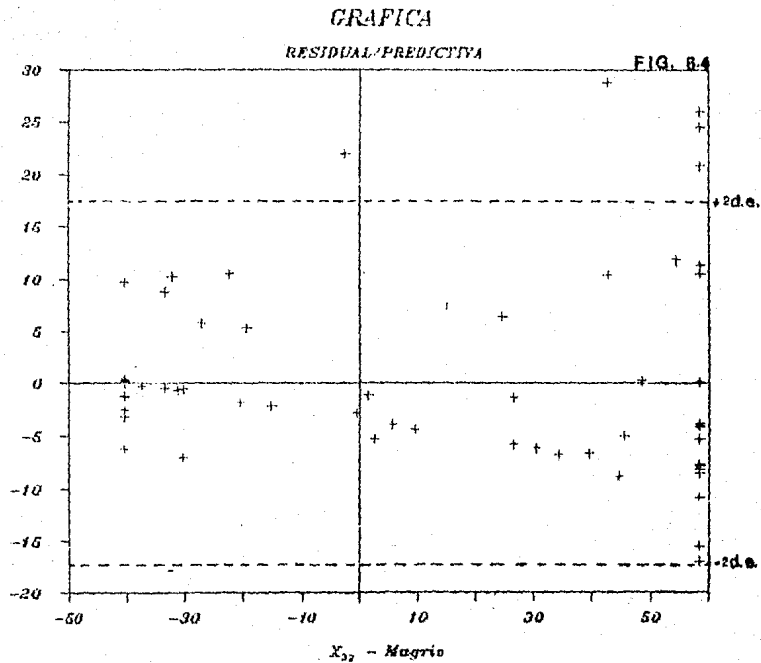
RESIDUAL/PREDICTIVA

FIG. 8.3

59

RESIDUAL





Variable dependiente: X_{24} -Asitec -Ejidos con asistencia técnica.

Cuadro 11
Variables en la ecuación

Variable No.	Ident..	$\hat{\beta}_i$	$\sqrt{\hat{\text{Var}}(\hat{\beta}_i)}$	T	F	α
27	Orvent	.79864	.07824	0	103.66	0
5	Ejirie	.11983	.02229	.89	28.89	0
23	Crepri	.29888	.08957	.83	11.13	.001
30	Ganaej	-.00740	.00179	.44	17.06	0
21	Orcred	.09941	.04602	.32	4.67	.034
25	Orprod	-.09398	.04499	.41	4.36	.040

T = Índice de tolerancia con el que se incluyó X_i

F = Valor parcial de F que determinó la aceptación de X_i

α = Nivel de significancia descriptivo.

Nuestra ecuación de predicción es :

$$\hat{Y}_i = \hat{\beta}_{27} X_{i27} + \hat{\beta}_5 X_{i5} + \hat{\beta}_{23} X_{i23} + \hat{\beta}_{30} X_{i30} + \hat{\beta}_{21} X_{i21} + \hat{\beta}_{25} X_{i25}$$

sustituyendo en la ecuación obtenemos los valores estimados de Y, que se presentan en el cuadro 14.

Las distintas medidas de relación de Asitec con el conjunto de variables predictivas incluidas en la ecuación, se presentan en el siguiente cuadro.

Cuadro 12
Cuadro de resultados

Paso	Variabes	R^2	$r^2_{24(i, j---k)}$	F-total $\alpha=0$	$\hat{\sigma}_e$
1	27	.66284	.66284	186.18	7.97
2	27,5	.74150	.07866	131.75	7.02
3	27,5,23	.78757	.02606	99.31	6.70
4	27,5,23,30	.80422	.03665	90.83	6.19
5	27,5,23,30,21	.81795	.01373	78.25	6.01
6	27,5,23,30,21,25,	.82808	.01014	68.86	5.88

es decir, el 82.81% de la variación en "Ejidos con asistencia técnica" es explicada por las 6 variables operando conjuntamente, siendo la variable "Ejidos organizados para la venta de su producción" la de mayor contribución en dicha explicación (66.3). Las pruebas

efectuadas nos indican que los valores obtenidos son significativos; presentandose en el Cuadro 13 el desarrollo de la última prueba.

Cuadro 13
Análisis de Varianza (última ecuación)

F.V.	G.L.	S.C.	C.M.	F	α
Regresión	8	14280.25	2380.04	68.86	0
Residual	74	2537.75	34.56		

De los datos registrados en los Cuadros 11 y 12, se observa que la asistencia técnica en un ejido de la región analizada, se presenta cuando a su vez existen diversas formas de organización : para obtener crédito, para la producción y para la venta de sus productos (ésta última es la de mayor peso) y son precisamente aquellos ejidos que cuentan con riego, que reciben crédito de organismos privados y desarrollan a su vez la actividad ganadera ; es decir, la asistencia técnica se presenta en un ejido cuando las condiciones le son favorables a la institución que la da o cuando forma parte de un paquete crediticio y no por necesidades propias del núcleo de población.

Validación de supuestos distribucionales

Por su similitud con las aplicaciones anteriores, señalaré únicamente que la dispersión que se observa en las Fig. 12.1 a 12.8 indican que existe evidencia de que se esta violando el supuesto de varianza constante y que a la vez existe error en los cálculos. La posible solución sería una transformación de Y.

CUADRO 14

MULTIPLE REGRESSION

OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-2SD	+2SD
1.	8.000000	12.93254	-4.932543		
2.	0	.8740803E-01	-.8740803E-01		
3.	17.000000	14.62174	2.370041		
4.	0	5.326663	-5.326663		
5.	0	17.34656	-17.34656		
6.	0	3.324033	-3.324032		
7.	0	7.496499	-7.496499		
8.	13.000000	10.03206	2.967941		
9.	0	-1.285278	1.285278		
10.	0	-1.228230	1.228230		
11.	0	-2.279133	2.279133		
12.	5.000000	5.814082	-.8140825		
13.	0	-5.778752	5.778752		
14.	0	-4.494122	4.494122		
15.	0	5.104388	-5.104388		
16.	0	4.447443	-4.447443		
17.	6.000000	.8115201	5.388480		
18.	0	2.359612	-2.359612		
19.	0	2.176868	-2.176868		
20.	0	2.660062	-2.660062		
21.	0	1.865910	-1.865910		
22.	25.000000	29.48278	-4.482786		
23.	0	8.889862	-8.889862		
24.	0	-5.793125	5.793125		
25.	0	-2.293835	2.293835		
26.	0	-4.373085	4.373085		
27.	27.000000	16.57297	10.42703		
28.	6.000000	10.16419	-4.164185		
29.	7.000000	5.127302	1.872698		
30.	0	3.739538	-3.739538		
31.	0	-4.373085	4.373085		
32.	0	-2.006058	2.006058		
33.	8.000000	1.459511	6.540489		
34.	0	-8.140077	8.140077		
35.	0	8.823478	-8.823478		
36.	0	-2.686007	2.686007		
37.	0	-2.293835	2.293835		
38.	0	-4.373085	4.373085		
39.	99.000000	90.23923	8.760769		
40.	0	-2.958375	2.958375		
41.	13.000000	7.769898	5.230102		
42.	0	2.488926	-2.488926		
43.	4.000000	14.49018	-10.49018		
44.	0	-2737803	2737803		
45.	0	4.222761	-4.222761		
46.	0	-4.373085	4.373085		
47.	0	-.3699734E-01	.3699734E-01		
48.	0	3.364923	-3.364923		
49.	0	.5426659	-.5426659		
50.	22.000000	14.09734	7.902657		
51.	0	-2.011595	2.011595		
52.	26.000000	5.746659	14.25334		

89

CUADRO 14

continuación

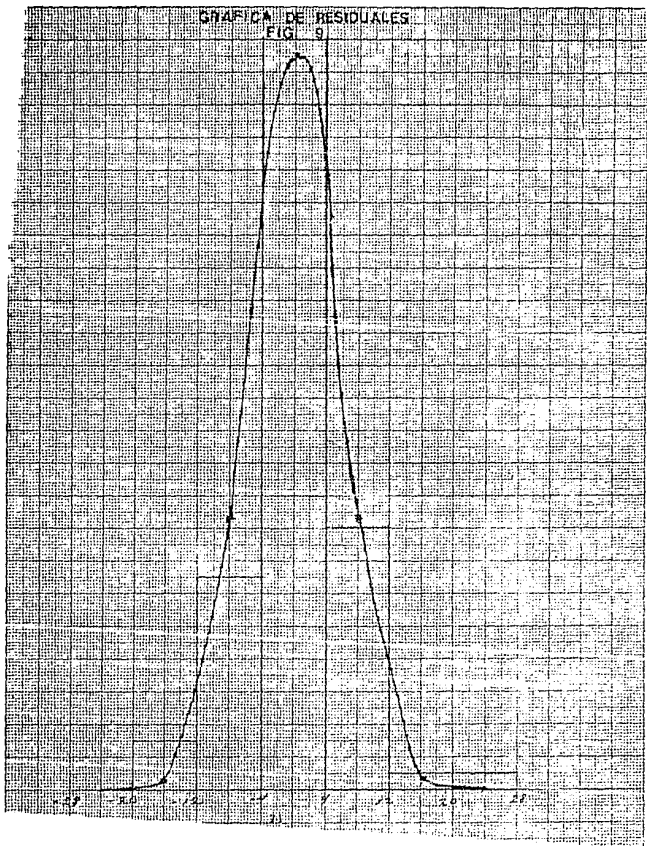
OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-2SD	0.0	+2SD
53.	0	3.053811	-3.053811			
54.	11.00000	10.95338	.4661654E-01			
55.	6.000000	4.263959	1.836041			
56.	0	12.44757	-12.44757			
57.	0	-.9487929	.9487929			
58.	0	-.9376109E-01	.9376109E-01			
59.	10.00000	5.652512	4.347488			
60.	0	1.441822	-1.441822			
61.	0	2.769067	-2.769067			
62.	33.00000	40.83550	-7.835504			
63.	0	-4.373085	4.373085			
64.	0	.6479958	-.6479958			
65.	0	3.148441	-3.148441			
66.	58.00000	27.40397	22.59603			
67.	0	0	0			
68.	0	-.5821415	.5821415			
69.	0	-2.575015	2.575015			
70.	0	12.17744	-12.17744			
71.	0	-8.302202	8.302202			
72.	6.000000	5.546349	.4536513			
73.	0	5.941862	-5.941862			
74.	0	-1.840825	1.840825			
75.	0	1.461950	-1.461950			
76.	0	0	0			
77.	0	-.2466872	.2466872			
78.	0	4.760940	-4.760940			
79.	0	-3.998108	3.998108			
80.	0	1.532773	-1.532773			

70

NOTE - (*) INDICATES ESTIMATE CALCULATED WITH MEANS SUBSTITUTED
R INDICATES POINT OUT OF RANGE OF PLOT

NUMBER OF CASES PLOTTED 80.
 NUMBER OF 2 S.D. OUTLIERS 5. OR 6.25 PERCENT OF THE TOTAL
 VON NEUMANN RATIO 2.09348 DURBIN-WATSON TEST 2.06731
 NUMBER OF POSITIVE RESIDUALS 43.
 NUMBER OF NEGATIVE RESIDUALS 37.
 NUMBER OF RUNS OF SIGNS 39.
 EXPECTED NUMBER OF RUNS OF SIGNS 41.
 EXPECTED S.D. OF RUN DISTRIBUTION 4.41842
 UNIT NORMAL DEViate -
 $Z = (\text{EXPECTED} - \text{OBSERVED}) / \text{S.D.}$ -2.8856
 PROBABILITY OF OBTAINING .GE. ABS(Z) .36646

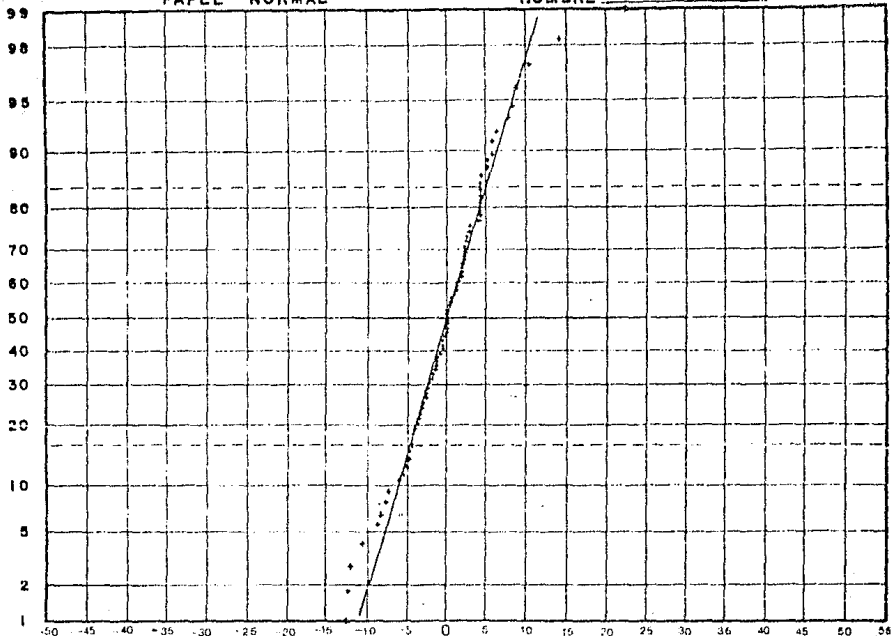
GRAFICA DE RESIDUALES
FIG. 9



LABORATORIO DE ESTADISTICA

PAPEL NORMAL

NOMBRE _____ FIG. 10



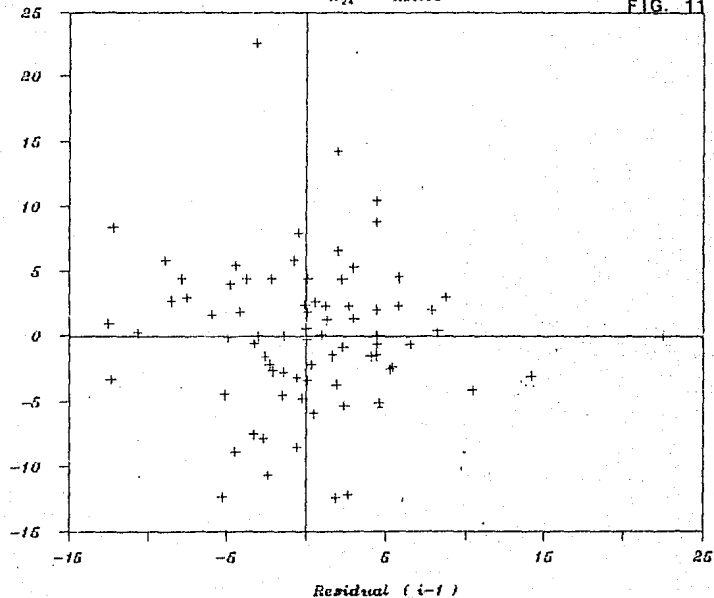
GRAFICA DE RESIDUALES

X_{24} — Aceite

FIG. 11

Residual (i)

73

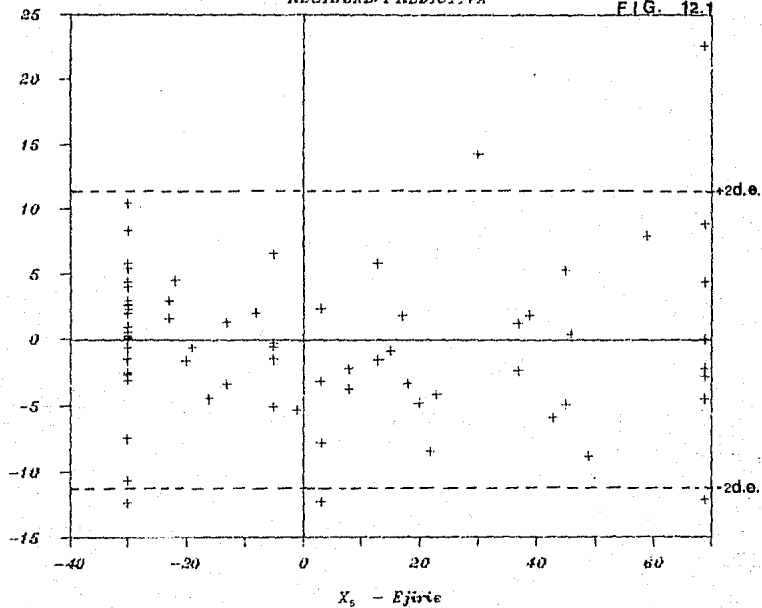


GRAFICA

RESIDUAL/PREDICTIVA

FIG. 12.1

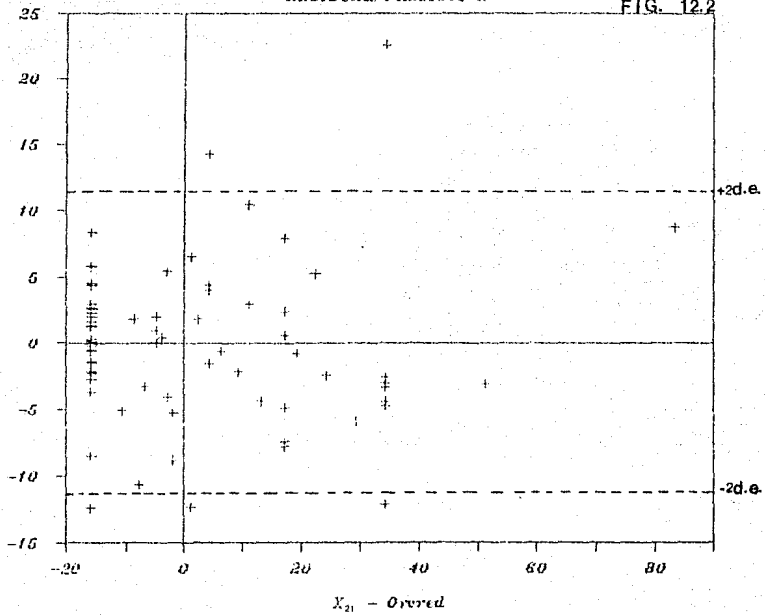
74
RESIDUAL



GRAFICA

RESIDUAL/PREDICTIVA

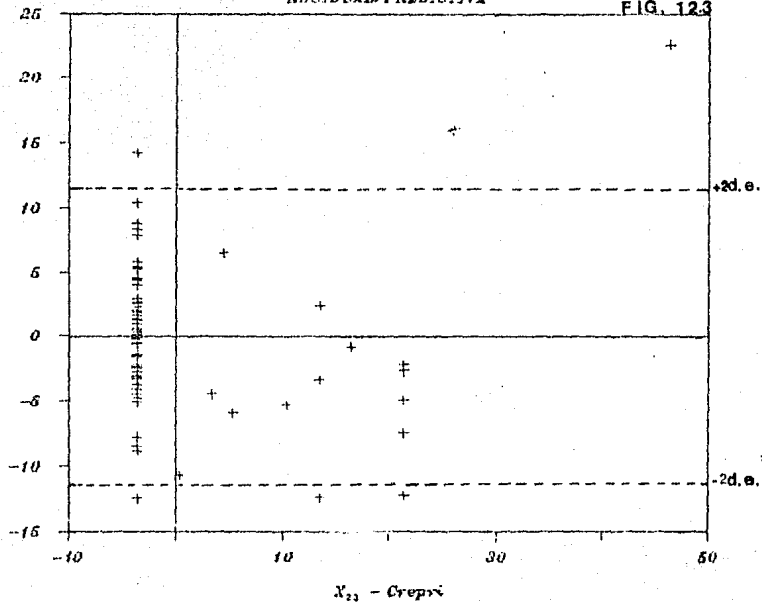
FIG. 12.2



GRAFICA

RESIDUAL/PREDICTIVA

FIG. 123

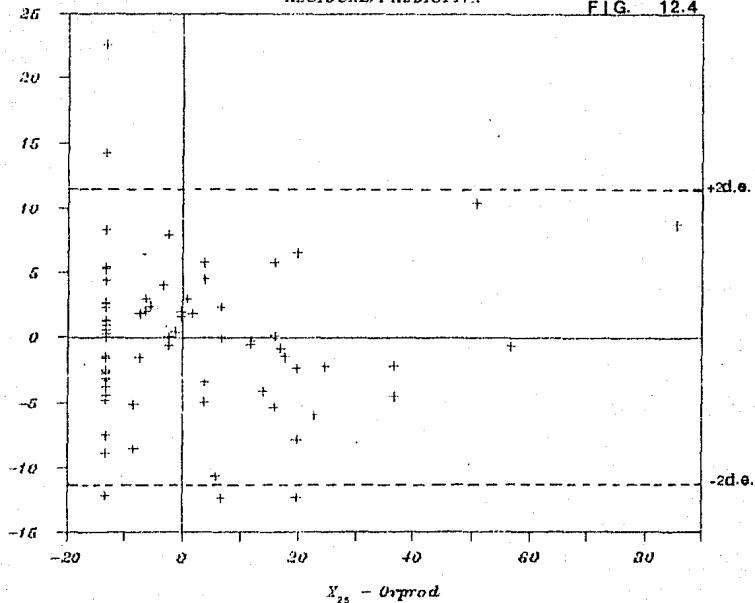


GRAFICA

RESIDUAL/PREDICTIVA

FIG. 12.4

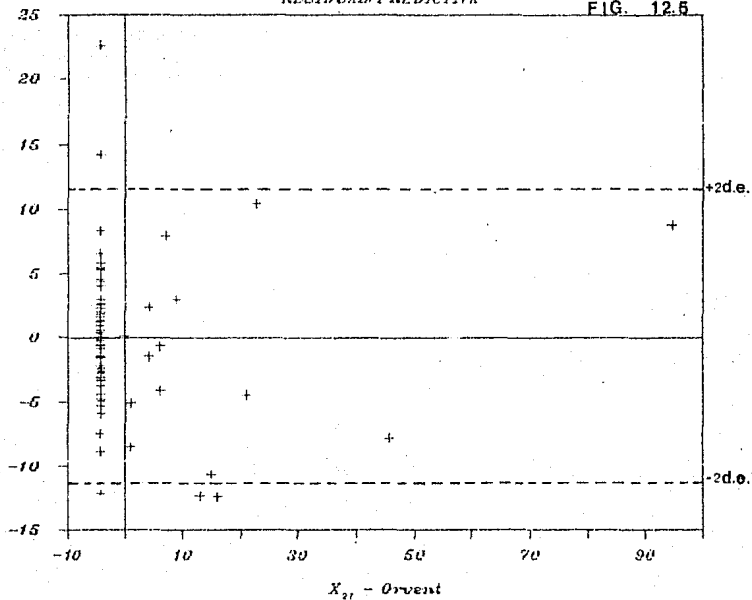
11
RESIDUAL



GRAFICA

RESIDUAL/PREDICTIVA

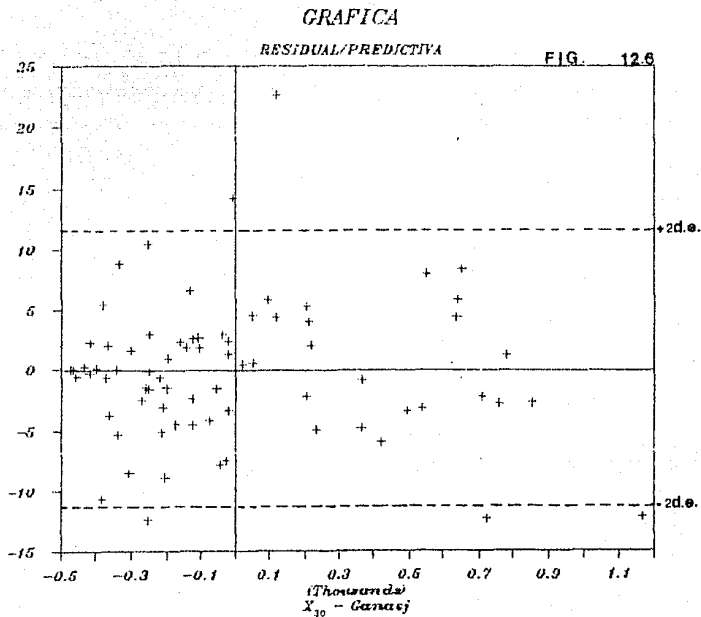
FIG. 12.5



ESTÁ TESIS NO DEBE
SALIR DE LA BIBLIOTECA

79

RESIDUAL



Variable dependiente : X_{2d} = Orinsu = Ejidos organizados para la compra de insumos.

Cuadro 15
Variables incluidas en la ecuación

Variable No.	Ident.	$\hat{\beta}_i$	$\sqrt{\text{Var}(\hat{\beta}_i)}$	T	F	α
27	Orvent	.82533	.07010	0	138.61	0
5	Ejirie	.10116	.02148	.89	22.23	0

T = Índice de tolerancia con el que se incluyó X_i

F = Valor parcial de F que determinó la aceptación de X_i

α = Nivel de significancia descriptivo.

La ecuación de regresión de X_{2d} respecto a X_{27} y X_5 es:

$$\hat{Y}_i = \hat{\beta}_{27} X_{27} + \hat{\beta}_5 X_{15}$$

Esta sencilla ecuación de predicción nos proporciona una idea más clara respecto de los fenómenos que se dan en el sector ejidal. El nivel de organización para la compra de insumos, está en función al nivel de organización para la venta de su producción, así como el contar con riego para desarrollar la actividad agrícola. Es decir, la organización para la compra de insumos, es un grado superior en la organización para la producción, que se da principalmente cuando existe un alto grado de tecnificación en el desarrollo de sus actividades productivas, como es el caso de contar con riego.

Las distintas medidas de relación calculadas, la última prueba realizada, así como los valores y gráfica de los residuales, se presentan en los cuadros siguientes.

A través de los valores de la variable dependiente registrados en el Cuadro 18, podemos calcular el coeficiente de determinación:

$$R_{\hat{Y}Y}^2 = \Sigma \hat{Y}_i^2 / \Sigma Y_i^2 = 0.68$$

De las pruebas efectuadas, se puede afirmar que si existe

una relación entre tales variables.

Cuadro 18
Cuadro de resultados

paso	variables	R ²	r ² _{20(1.1)}	F-total α=0	$\hat{\sigma}_e$
1	27	.58574	.58574	157.81	9.24
2	27,9	.67760	.09187	111.22	8.20

Cuadro 17
Análisis de Varianza (última ecuación)

F.V.	G.L.	S.C.	C.M.	F	α
Regresión	2	14957.92	7478.96	111.22	0
Residual	78	5245.08	67.24		

Validación de supuestos distribucionales

- 1) La Fig. 14 nos presenta cierta evidencia que indica falta de normalidad.
- 2) Los datos de la Fig 18.1 y 18.2 dan evidencia de que se está violando el supuesto de varianza constante.
- 3) El estadístico de Durbin-Watson nos indica que no hay autocorrelación.

Finalmente, para las cuatro aplicaciones, se observa que la variable dependiente fue medida en porcentaje, existiendo demasiados valores iguales a cero; por otra parte las gráficas indican básicamente un problema de varianza no constante. Por lo anterior, lo recomendable sería una transformación de Y que nos permita estabilizar dicha varianza; esta transformación podría ser una función de tipo logarítmica.

CUADRO 18

MULTIPLE REGRESSION

OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-2SD	0.0	+2SD
1.	8.000000	7.586666	.413334			
2.	0	3.036666	-3.036666			
3.	17.000000	9.940000	7.059197			
4.	14.000000	2.933511	11.066489			
5.	0	17.34881	-17.34881			
6.	4.000000	4.855466	-.855466			
7.	0	0	0			
8.	20.000000	11.43743	8.562573			
9.	0	1.719644	-1.719644			
10.	0	6.777422	-6.777422			
11.	20.000000	0	20.000000			
12.	0	4.552000	-4.552000			
13.	0	4.349688	-4.349688			
14.	0	.8092444	-.8092444			
15.	5.000000	6.655557	-1.655557			
16.	7.000000	1.416178	5.583822			
17.	0	0	0			
18.	0	6.777422	-6.777422			
19.	0	10.01440	-10.01440			
20.	0	0	0			
21.	0	1.011555	-1.011555			
22.	50.000000	30.64774	19.35226			
23.	21.000000	7.991288	13.00871			
24.	0	0	0			
25.	0	0	0			
26.	0	0	0			
27.	9.000000	22.28401	-13.28401			
28.	3.000000	13.61488	-10.61488			
29.	0	2.172722	-4.979733			
30.	0	3.843911	-3.843911			
31.	0	0	0			
32.	0	0	0			
33.	8.000000	2.528889	5.471111			
34.	0	1.112711	-1.112711			
35.	5.000000	9.386757	-4.386757			
36.	0	0	0			
37.	0	0	0			
38.	0	0	0			
39.	99.000000	91.72244	7.277562			
40.	0	0	0			
41.	25.000000	7.586666	17.41333			
42.	0	0	0			
43.	25.000000	15.68134	7.318699			
44.	0	0	0			
45.	25.000000	3.843911	21.15609			
46.	0	0	0			
47.	0	0	0			
48.	17.000000	1.719644	15.28036			
49.	25.000000	2.828889	22.17111			
50.	11.000000	18.08151	-7.08151			
51.	0	2.225422	-2.225422			
52.	0	4.069333	-4.069333			

B2

CUADRO 18

continuación

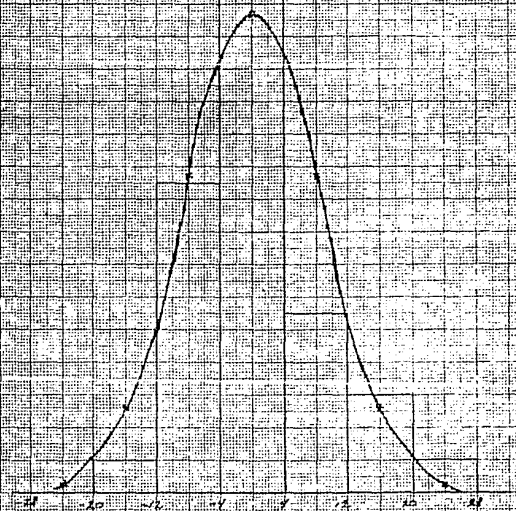
OBSERVATION	Y VALUE	Y ESTIMATE	RESIDUAL	-2SD	0.0	+2SD
53.	0	0	0			
54.	0	10.01440	-10.01440			
55.	6.000000	4.754311	1.245689			
56.	20.00000	16.50667	3.493325			
57.	0	0	0			
58.	4.000000	3.301335	.6986651			
59.	30.00000	10.01440	19.98560			
60.	13.00000	2.528889	10.47111			
61.	0	10.01440	-10.01440			
62.	33.00000	44.60482	-11.60482			
63.	0	0	0			
64.	0	2.253337	-0.253337			
65.	0	3.330133	-3.330133			
66.	0	10.01440	-10.01440			
67.	0	0	0			
68.	0	0	0			
69.	0	0	0			
70.	0	10.01440	-10.01440			
71.	0	0	0			
72.	12.00000	7.607822	4.312178			
73.	9.000000	7.384355	1.615645			
74.	0	-7.080888	-7.080888			
75.	0	6.602670	-6.602670			
76.	0	0	0			
77.	13.00000	2.528889	10.47111			
78.	0	5.057177	-5.057177			
79.	5.000000	0	5.000000			
80.	0	4.349668	-4.349668			

NOTE - (*) INDICATES ESTIMATE CALCULATED WITH MEANS SUBSTITUTED
 R INDICATES POINT OUT OF RANGE OF PLOT

NUMBER OF CASES PLOTTED	80.		
NUMBER OF 2 S.D. OUTLIERS	7.	OR	8.75 PERCENT OF THE TOTAL
VON NEUMANN RATIO	1.76375	DURBIN-WATSON TEST	1.76146
NUMBER OF POSITIVE RESIDUALS	46.		
NUMBER OF NEGATIVE RESIDUALS	32.		
NUMBER OF RUNS OF SIGNS	32.		
EXPECTED NUMBER OF RUNS OF SIGNS	39.		
EXPECTED S.D. OF RUN DISTRIBUTION	4.26371		
UNIT NORMAL DEViate			
Z (EXPECTED-OBSERVED)/S.D.	-1.61831		
PROBABILITY OF OBTAINING .GE. ABS(Z)	.05280		

GRAFICA DE RESIDUALES

FIG. 13

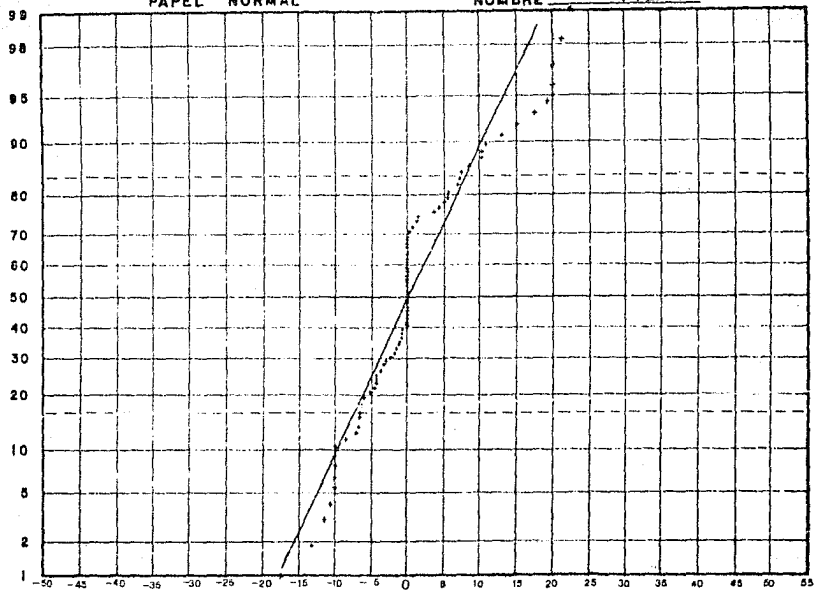


12

LABORATORIO DE ESTADISTICA

PAPEL NORMAL

NOMBRE _____ FIG. 14



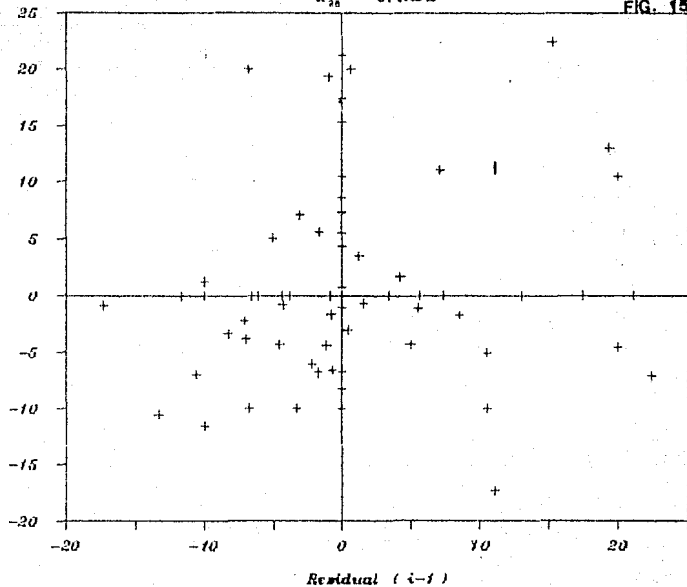
GRAFICA DE RESIDUALES

X_{20} -- Orinau

FIG. 15

98

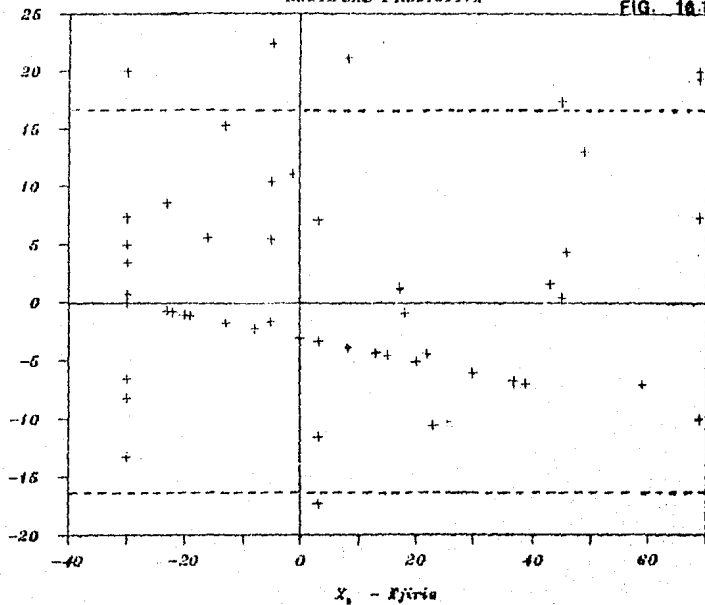
Residual (i)



GRAFICA

RESIDUAL/PREDICTIVA

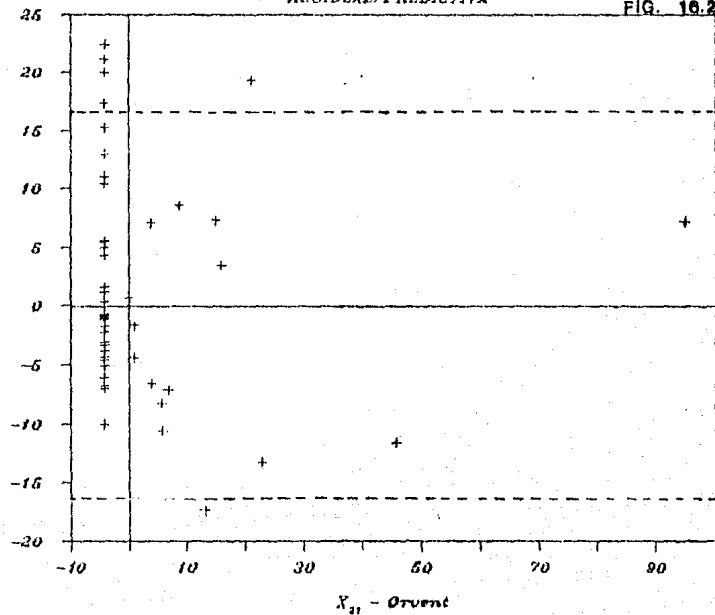
FIG. 10.1



GRAFICA

RESIDUAL/PREDICTIVA

FIG. 10.2



III.2. -Hipótesis que se desprenden.

Como mencionaba al principio de este estudio, el sector ejidal se caracteriza por diversos factores que determinan o condicionan su situación en el ámbito de la problemática general del país. Al efectuar el análisis estadístico, sobre un grupo de variables que de alguna manera miden características de tipo económico, social o político en dicho sector, nos permitió determinar el valor de una variable partiendo del conocimiento de una o varias variables relacionadas con aquella; en este caso en particular la relación supuesta fue a través de una función de regresión lineal, calculando distintas medidas de relación y efectuando las pruebas que nos determinan su aceptación.

De los resultados presentados en la sección anterior y específicamente para la región analizada, se desprenden las hipótesis siguientes:

a) De la primera aplicación del modelo de regresión, en donde la variable dependiente Proins mide las distintas instancias que debe superar un núcleo agrario para la posesión legal y garantizada de sus tierras, nos indica que la problemática agraria en el sector ejidal, no se determina únicamente por la garantía documentada en la tenencia de la tierra o por la existencia de problemas agrarios específicos, en virtud de que existe una estrecha relación con la presencia de grupos indígenas y con los niveles de organización del grupo campesino.

Cuando un ejido cuenta con todos sus trámites agrarios concluidos y documentados, se presenta en el grupo la seguridad en la tenencia, se estimula la producción, se amplía el aprovechamiento de sus recursos, se incentiva la organización y se canalizan más fácilmente los recursos externos y servicios hacia el núcleo agrario, así como también se simplifica el proceso de comercialización. Lo

anterior, nos determina los distintos grados de problemática agraria existentes en el sector ejidal.

b) De las tres aplicaciones restantes, se desprende que la organización de los grupos campesinos, como factor necesario para utilizar y administrar eficientemente los recursos naturales, humanos y de capital de que dispone, queda explicada por la presencia de indicadores que miden diversas formas de organización (para la producción, comercialización u obtención de crédito), con la influencia sustancial del indicador de sistemas de riego y utilización de maquinaria especializada en el desarrollo de su actividad agrícola.

Conclusiones tentativas.

Las fuentes de información, mencionadas en el Capítulo I, presentaron criterios distintos de organización y actualización, por lo que la primera dificultad que se tiene al querer realizar un estudio como el presente, es el no contar con una infraestructura de información básica y apoyos técnicos, que permitan la captación y el análisis de la información dentro de los márgenes de profundidad, tiempo y confiabilidad deseados.

La relación observada entre las variables de Problemas Agrarios con las de Organización, nos indican el fenómeno conocido como "situación legal y organización de los grupos campesinos" ; en dicha relación la situación legal es un factor determinante para el buen logro de la organización económica ejidal, siendo el caso de que en los ejidos donde se presentan problemas sobre la tenencia de la tierra no es fácil iniciar un proceso organizativo, debido a que los conflictos internos provocan cierta apatía en los ejidatarios, que consideran en primer término la legalización en la tenencia de su parcela.

Del grupo de variables Poblacionales se identifica un aspecto importante en las relaciones resultantes, como es la existencia de grupos indígenas, que según los Censos de Población representan aproximadamente el 17% de la población total del Estado, de los cuales el 39% son menonlimgues, lo que dificulta aun más la tarea de su integración al desarrollo regional ; en general estos grupos indígenas viven en comunidades cercanas casi siempre a un centro rector, su organización económica está basada principalmente en una agricultura de subsistencia o en la elaboración de diversas artesanías que venden en los centros cercanos; por otra parte a su alrededor proliferan los intermediarios de sus productos.

Finalmente, de los resultados que el análisis estadístico arroja, se puede concluir lo siguiente: Los factores limitantes del desarrollo regional son:

A) Físico.- La conformación topográfica del terreno y la calidad de los suelos que lo constituyen, presentan un alto índice de erosión y precipitación, que trae como consecuencia bajos rendimientos productivos o el desaprovechamiento de grandes extensiones.

B) Legal.- Referido principalmente a la tenencia de la tierra como una limitante compleja. Ejidatarios y pequeños propietarios que en su mayoría tienen menos de 5 hectáreas, objetivamente la situación de unos y otros es muy parecida :

- * Poca tierra que cuando es buena está arrendada.
- * Falta de titulación en la tenencia.
- * Carencia de capital de trabajo.
- * Dependencia de la usura.
- * Falta de efectividad de la instrumentación legal, que provoca enfrentamientos entre ellos ; como por ejemplo, Ejidatarios sin tierra presionan sobre reducidas extensiones de pequeños propietarios o estos sobre confusos linderos ejidales; en la mayoría de las veces son conflictos manipulados por grupos regionales que persiguen beneficios particulares.

C) Institucional.- Se da como una consecuencia de las dos limitantes anteriores, si la estructura de la tierra que le ha sido dotada no asegura "rendimientos aceptables " y/o no ha resuelto el factor legal, no pueden ser sujetos de crédito y la asistencia técnica no existe :

- * Los campesinos no son poseedores legales de su tierra, lo que "dificulta " el movimiento de las instituciones encargadas de promover el desarrollo regional.
- * No ofrecen las garantías financieras que se piden.
- * El crédito institucional no alcanza para todos los solicitantes.
- * Son campesinos marginados, es decir casi inexistentes.

DD Humano.- Incidentalmente, esta situación produce un sentimiento de inseguridad en los campesinos a quienes no se les garantiza la posesión legal de su tierra, por lo que, consecuentemente no quieren invertir en ellas más capital o esfuerzo que el necesario para que les dé precariamente para subsistir

Recomendaciones.

Como señalaba en la sección 1.2, en los años 70 colaboré en la Dirección General de Programación y Evaluación de la Secretaría de la Reforma Agraria. Dicha dependencia, para cumplir con las necesidades de planeación del Sector, inició en 1974 un sistema de encuestas en los ejidos del país, con diversas limitantes en el desarrollo de las actividades correspondientes, como son el carecer de infraestructura básica (datos previos de identificación, organización y control del trabajo de campo y disponibilidad de personal capacitado), cartografía y sobre todo de diversos apoyos técnicos, como por ejemplo un sistema probado de proceso electrónico de la información. Por otra parte, cabe destacar que soy originario del Estado de Hidalgo, teniendo la oportunidad de conocer la situación económica, social y política que prevalece en dicha región; por lo que, en el estudio que aquí presento, me permito asentar las siguientes recomendaciones:

A) Se debe estructurar y fundamentar un banco de datos, elaborando en principio una metodología sustentada en bases firmes teóricas y prácticas, adaptadas a las condiciones particulares del sector, tanto en lo relativo a infraestructura técnica por el lado del diseño de captación y manejo de información, como en cuanto al desarrollo cultural de quienes proporcionan dicha información. Responsabilizando tanto a los generadores del banco de datos como a los usuarios de la información, de tal forma que mientras los primeros deben usar conceptos homogéneos, los

segundos deben plantear sus demandas con mayor precisión.

B) Integrar y unificar los programas de organización y capacitación con propósitos de producción, aumento de recursos crediticios, adquisición de insumos y maquinaria, comercialización y difusión de las nuevas tecnologías para un mejor aprovechamiento de los recursos con que cuenta el ejido ; como una importante posibilidad para afrontar la complejidad de la problemática regional. Desarrollando una tecnología agropecuaria y forestal acorde con las características de la región, para subir a un plano de primera importancia actividades económicas accidentadas y en el mejor de los casos complementarias ; incrementando el desarrollo integral del ejido, dotando de infraestructura social a la población; proponiendo medidas concretas para mejorar la capacitación y capacidad de gestión de los grupos campesinos.

C) Intensificar las actividades encaminadas a regularizar el régimen legal de la tenencia de la tierra. Definiendo correctamente los límites de propiedad ; otorgando seguridad jurídica a sus distintas modalidades ; manteniendo vigentes los derechos individuales y de los grupos campesinos, mediante la expedición de certificados de derechos agrarios, integración de la documentación básica, deslindes, depuración censal, etc..

Bibliografía

- 1.- Draper N. R. y Smith H.
Applied Regression Analysis
Wiley international
2a. Edición, U.S.A., 1981
- 2.- García Uribe José
Recorriendo el Estado de Hidalgo
Gráficos Olimpo, México 1979
- 3.- Hool Paul G.
Introducción a la Estadística Matemática
Ed. Ariel, S. A., Barcelona, 1976
- 4.- Johnston J.
Métodos Econométricos
Ed. Vicens-Vives, 1972
- 5.- Nie Norman, et al.
Statistical Package for the Social Sciences (SPSS)
Mc Graw - Hill, New York, 1975
- 6.- Padua Jorge
Paquete Estadístico para las Ciencias Sociales (SPSS)
Oferta y condiciones para su utilización e interpretación
de resultados
El Colegio de México, Centro de Estudios Sociológicos
Cuaderno no. 12, México 1975
- 7.- Restrepo Fernández Ivan, et'al
Los problemas de la Organización Campesina
Editorial Campesina. México, 1976
- 8.- Varman Arturo
Los Campesinos Hijos Predilectos del Regimen
Nuestro Tiempo, México 1973