

MEMORIAS

V FORO NACIONAL DE ESTADISTICA

XALAPA VERACRUZ

10 AL 13 DE SEPTIEMBRE DE 1990

ORGANIZADORES:

FACULTAD DE ESTADISTICA E INFORMATICA
UNIVERSIDAD VERACRUZANA
ASOCIACION MEXICANA DE ESTADISTICA

AUSPICIADAS POR CONACYT

MEMORIAS

V FORO NACIONAL DE ESTADISTICA

XALAPA VERACRUZ

10 AL 13 DE SEPTIEMBRE DE 1990

ORGANIZADORES:

FACULTAD DE ESTADISTICA E INFORMATICA
UNIVERSIDAD VERACRUZANA
ASOCIACION MEXICANA DE ESTADISTICA

AUSPICIADAS POR CONACYT

V FORO NACIONAL DE ESTADISTICA

XALAPA VERACRUZ
Del 10 al 13 de Septiembre de 1990.

UNIVERSIDAD VERACRUZANA
FACULTAD DE ESTADISTICA E INFORMATICA

ASOCIACION MEXICANA DE ESTADISTICA

INSTITUCIONES PARTICIPANTES

Universidad Veracruzana
Colegio de Postgraduados (Centro de Estadística y Cálculo)
Universidad Nacional Autónoma de México (IIMAS)

COMITE ORGANIZADOR LOCAL

(U. V.)

Victor M. Méndez (Coordinador)
Fernando U. González
Mario Miguel Ojeda
Gisela Vargas
J. Aurora Montano
Pablo E. Suardíaz
Juan Muñoz Navarro
Ma. Cristina Ortiz
José Luis Colorado
Claudio R. Castro

COMITE ORGANIZADOR NACIONAL

Enrique de Alba - ITAM
Alfredo Bustos - IIMAS
Said Infante - CEC
José Luis Fara - INEGI
Guillermo Baz - IIMAS
Manuel Mendoza - FC-UNAM
Felix González - CEC
Eduardo Sojo - INEGI
Alberto Castillo - CEC
Ignacio Méndez - IIMAS
Guillermo Zárata - CEC

PATROCINADO POR:

Consejo Nacional de Ciencia y Tecnología
Universidad Veracruzana

M E M O R I A S

V FORO NACIONAL DE ESTADISTICA

COMITE EDITORIAL

MANUEL MENDOZA (UNAM)

EDUARDO SOJO (INEGI)

FEDERICO O'REILLY (UNAM)

ALBERTO CASTILLO (CP)

MARIO MIGUEL OJEDA (U.V.)
(COORDINADOR)

PRESENTACION

La colección de artículos en esta memoria está constituida por algunos artículos presentados como contribuciones libres en el V FORO NACIONAL DE ESTADISTICA, celebrado del 10 al 13 de septiembre de 1990 en la Facultad de Estadística e Informática de la Universidad Veracruzana. El proceso de selección de los trabajos fue realizado por los miembros del Comité Editorial, quienes a su juicio y de manera independiente se auxiliaron de revisores calificados, miembros de la comunidad estadística mexicana, para emitir las opiniones definitivas sobre 16 artículos presentados a consideración. Se buscó que cada artículo fuese calificado por al menos dos personas independientemente. Debido a que la selección de los trabajos se realizó en una sola ronda, se tuvo que admitir un criterio de mayor flexibilidad que la usual en la revisión de trabajos técnicos para una revista especializada, cosa que el lector podrá verificar en la lectura de los artículos.

El Comité Editorial tuvo que resolver el dilema de la forma de garantizar la aparición oportuna de las memorias, y es en ese sentido y considerando las restricciones de tiempo y recursos para la tarea editorial, que se justifica la decisión de que el proceso de selección fuera realizado en una sola ronda.

Los trabajos se presentan ordenados alfabéticamente por el apellido paterno del primer autor, solo por un criterio de facilidad de organización del material.

Se espera que esta memoria constituya el esfuerzo continuado de una labor editorial que ha pretendido dejar constancia de los trabajos y contribuciones de algunos sectores de la comunidad estadística mexicana en el marco de los foros de Estadística.

Xalapa-Equez., Ver. agosto de 1991

EL COMITE EDITORIAL

I N D I C E

Evaluación de la exposición ocupacional a plaguicidas organofosforados y carbámicos en el municipio de Villa Guerrero, Edo. de México.	
Arroyo-López P. E., Amaya-Chávez y Sánchez-Mesa.	1
Aspectos metodológicos del modelo de regresión logística.	
Colorado-Hernández J.L., Ojeda M.M. y Jácome-Avila N.	13
Aplicaciones ecológicas del modelo logístico.	
Cruz-Kuri L. y Sosa-Galindo I.	27
Análisis estadístico de algunas características morfológicas de la gaviota parda.	
Mendoza-Blanco J.R. y Villalba-Cohen R.	39
Conceptos paralelos en las áreas de diseños experimentales y muestreo.	
Ojeda M.M. y Cortina-Carmona C.H.	55
Análisis de intervención de series de tiempo univariadas.	
Rodríguez-Muñoz J.E.	68
Uso de técnicas de cúmulos en un análisis inicial de secuencias de aminoácidos.	
Suardíaz-Solé P. y Vargas-Mestizo G.	82
Componentes principales: análisis de datos sobre cobertura de atención primaria a la salud.	
Valencia G.J. y Serrano O.V.	90
Directorio de participantes	107

EVALUACION DE LA EXPOSICION OCUPACIONAL A PLAGUICIDAS
ORGANOFOSFORADOS Y CARBAMICOS EN EL MUNICIPIO DE
RO, ESTADO DE MEXICO

Arroyo López, Pilar Esther
ITESM, Campus Toluca. A.P. 1101
Amaya Chávez, Araceli, y Sánchez
Meza, Juan Carlos. Facultad de
Química de la U.A.E.M. A.P. 50.
Toluca, edo. de México.

RESUMEN: Tres métodos de determinación de colinesterasa en sangre fueron probados con el propósito de seleccionar el más adecuado para realizar el diagnóstico temprano de intoxicación por plaguicidas. Los métodos fueron evaluados sobre tres grupos de individuos: sanos, expuestos a plaguicidas pero asintomáticos y expuestos a plaguicidas y con síntomas de intoxicación. Los individuos seleccionados fueron de diferentes edades y sexo. Los métodos fueron comparados estadísticamente empleando análisis de covarianza multivariado. El mejor método fue seleccionado en base a su sensibilidad y especificidad.

PALABRAS CLAVE: Análisis de covarianza multivariado, Tabla de contingencia, Sensibilidad, Especificidad, Validación de métodos.

1. INTRODUCCION

El uso de plaguicidas se ha incrementado en los últimos años en toda América Latina. En México, el Comité Nacional Calificador de Insumos estimó que en 1988 se usaron en la agricultura más de 53,000 ton. de plaguicidas, aún cuando el uso de estos productos tiene un beneficio inmediato, los riesgos a corto y largo plazo que sobre el ambiente y las poblaciones humanas tiene su uso indiscriminado, no ha sido debidamente valorado.

En México, se ha estimado que de la población económicamente activa (24 millones), aproximadamente 5 millones tiene contacto regular con plaguicidas (SARH, 1988 y SPP, 1987). Entre los grupos de plaguicidas más empleados están los organofosforados y

carbamatos, que aunque menos persistentes en el ambiente, son altamente tóxicos al hombre (Albert, 1986). El efecto tóxico principal es la inhibición de la acetilcolinesterasa, enzima que degrada la acetilcolina, compuesto que participa como neurotransmisor del sistema nervioso. Las consecuencias de la inhibición pueden abarcar desde debilidad muscular, hormigueo, espasmos musculares y náusea, hasta depresión respiratoria y muerte. Los síntomas de intoxicación no se muestran hasta que ocurra al menos un 50% de inhibición enzimática. La determinación de la actividad de la acetilcolinesterasa representa la prueba de diagnóstico más adecuada y de bajo costo para la detección de exposición a plaguicidas. De los métodos existentes para la evaluación, se tienen determinaciones en eritrocitos y en plasma. Dadas las diferencias entre los métodos disponibles, se decidió comparar los resultados obtenidos aplicando cada uno de ellos con el fin de determinar el más apropiado para emplearse en la vigilancia de poblaciones expuestas a plaguicidas.

2. SELECCION DE LAS MUESTRAS EN EL ESTUDIO

El estudio incluyó a los habitantes de los municipios de Villa Guerrero y Tenancingo, municipios aledaños en donde la principal actividad es la floricultura. Para la elaboración del marco de muestreo, se efectuaron tres visitas a la zona de estudio y se empleó la información disponible en la Secretaría de Salud y las empresas floricultoras del área. Los grupos incluidos en el estudio fueron los siguientes:

Grupo Control: Los individuos del grupo control se eligieron al azar de entre los estudiantes y trabajadores de la Escuela Normal de Tenancingo. De la encuesta efectuada en este lugar, se eligieron individuos que no habían tenido exposición a plaguicidas en por lo menos los últimos tres años y clínicamente sanos. Se excluyeron mujeres embarazadas y personas con antecedentes de enfermedad hepática. En este grupo se tuvieron 60 individuos (40 mujeres y 20 hombres), con edades entre 15 y 24 años.

Grupo Expuesto B: Los individuos en este grupo se eligieron al azar de la lista de trabajadores de las empresas floricultoras de la zona de Villa Guerrero. Los individuos en este grupo fueron 35

*Testigo

(28 mujeres y 11 hombres), con edades entre 15 y 27 años, los cuales habían estado expuestos a plaguicidas en por lo menos el último año.

Grupo Expuesto C: Los individuos de este grupo fueron aquellos registrados en la lista de la Jurisdicción de Santa Ana, Tenancingo, dependiente de la Secretaría de Salud. Se incluyeron 45 personas, todos hombres, con edades entre 15 y 62 años, los cuales tenían antecedentes clínicos de intoxicación por plaguicidas organofosforados y carbamatos.

3. COMPARACION ENTRE METODOS

En una primera etapa del estudio, se procedió a efectuar una comparación entre los métodos denominados: Magnotti, Pharmachim y Michel, considerando los grupos Control y Expuesto B. La variable de respuesta considerada fué el vector de actividad colinesterásica, el cual contiene las siguientes entradas:

X_1 = concentración de colinesterasa en eritrocitos, determinada por el método de Magnotti (UI/ml).

X_2 = concentración de pseudocolinesterasa en plasma, determinada por el método de Pharmachim (UI/ml).

X_3 = concentración de pseudocolinesterasa en plasma, determinada por el método de Michel (Aph/hr, transformada luego a UI/ml).

Se condujo un Análisis de Covarianza Multivariado (MANOVA) sobre el vector de respuesta, empleando para ello el paquete estadístico SYSTAT versión para computadora personal IBM compatible (SYSTAT, 1984), el modelo propuesto fué el siguiente:

$Y = \text{constante} + \text{edad} + \text{sexo} + \text{nivel de exposición}$
de acuerdo a la revisión bibliográfica, no se espera la presencia de interacciones entre los factores involucrados.

Se probaron las hipótesis de no diferencia entre sexos, niveles de exposición y edades respecto a los niveles de colinesterasa de los participantes en el estudio, declarándose diferencias en la respuesta observada para los factores sexo y grado de exposición, ver el Anexo I. Los análisis univariados incluidos en el listado permitieron concluir que el método de

Magnotti fué susceptible a las diferencias entre sexo; en tanto que el método Michel en la prueba univariada no es sensible al grado de exposición de los individuos ($\alpha=0.05$)

Se construyeron entonces intervalos de confianza simultáneos para la diferencia entre la actividad enzimática media para el grupo control y el grupo expuesto B. Los intervalos tienen la forma (Morrison, 1984):

$$(\bar{X}_{1i} - \bar{X}_{2i}) \pm c \sqrt{(1/n_1 + 1/n_2) e_{ii,p}} \quad \text{para } i=1,2,3$$

en donde e_{ii} es el i -ésimo elemento de la diagonal de la matriz de error E , X_{1i} y X_{2i} denotan las medias del grupo control y expuesto respectivamente, para el i -ésimo método y $c^2 = 3/99 [F_{.05}(3,97)]$ las medias no fueron corregidas por la covariable edad, ya que la hipótesis nula para este efecto no fué rechazada.

Los intervalos construidos son los siguientes:

Método Magnotti:	$0.0832 < \mu_{11} - \mu_{21} < 0.4008$
Método Pharmachim:	$0.3448 < \mu_{12} - \mu_{22} < 0.8106$
Método Michel:	$-0.0895 < \mu_{13} - \mu_{23} < 0.1773$

Un 25% de inhibición en la actividad enzimática respecto a los valores medios normales, de acuerdo a la bibliografía, es ya indicio de intoxicación por plaguicidas. Dado que en los tres métodos no se detectó un número considerable de individuos con un 25% de inhibición enzimática, que permitiera evaluar la calidad de los métodos respecto a reconocer los casos de exposición verdaderos, se procedió a incluir un tercer grupo en el estudio, el cual se indentificó como grupo expuesto C.

Un Análisis de Covarianza Multivariado se efectuó sobre los vectores de respuesta del grupo Control y Expuesto C, el modelo propuesto fué:

$$Y = \text{constante} + \text{edad} + \text{grado de exposición}$$

el factor sexo ya no fué considerado dado que todos los individuos en el grupo C fueron del sexo masculino. Se probaron las hipótesis de no diferencia entre edades y grados de exposición, ver Anexo II, declarándose únicamente diferencias significantes respecto al grado de exposición a plaguicidas. Los intervalos de confianza

simultáneos para los grupos control y expuesto C son
 construidos y tienen la siguiente forma general:

$$(\bar{X}_{1i} - \bar{X}_{2i}) \pm c \sqrt{(1/n_1 + 1/n_2) e_{ii,p}} \quad \text{para } i=1,2,3$$

en donde ahora $c^2 = 3/95$ [F₀₅(3,95)].

Se obtuvieron los siguientes intervalos de confianza:

Método Magnotti: $0.3256 < \mu_{11} - \mu_{21} < 0.5037$

Método Pharmachim: $0.7577 < \mu_{12} - \mu_{22} < 1.1837$

Método Michel: $0.5279 < \mu_{13} - \mu_{23} < 0.7179$

Como se observa de los límites de los intervalos, el método Magnotti aún cuando detectó diferencias, éstas no excedieron el 25% de inhibición enzimática, en tanto que un mayor grado de inhibición era esperado dados los datos clínicos de los integrantes del grupo expuesto C. Los métodos Michel y Pharmachim mostraron mayor grado de diferenciación entre los grupos.

4. SENSIBILIDAD Y ESPECIFICIDAD DE LOS METODOS

La validez de un método dependerá de dos factores: sensibilidad y especificidad. La sensibilidad refiere a la capacidad que tiene un procedimiento para detectar a los casos verdaderos positivos, en tanto la especificidad refiere a la capacidad del método de identificar correctamente a los casos verdaderos negativos. Para medir estas características de validez, se construyeron tablas de contingencia 2x2, con la siguiente estructura (Mausner & Bahn, 1977):

Datos de encuesta

		Expuesto	no expuesto
Método	Expuesto	(+,+)=a	(+,-)=b
	no expuesto	(-,+)=c	(-,-)=d

en donde para clasificar a un individuo como "expuesto" en base a la actividad enzimática determinada por algún método, ha de registrarse una reducción en el nivel de colinesterasa mayor del 25% del nivel normal medio.

En base a la tabla se define

$$\text{Sensibilidad} = \frac{a}{a + c} \times 100 \quad \text{Especificidad} = \frac{b}{b + d} \times 100$$

Además, empleando la prueba de McNemar (Bennett, & Underwood, 1970), se probó la hipótesis:

H_0 = número de casos positivos en base al método es igual al número de casos positivos de acuerdo a la encuesta (verdaderos) contra la alternativa

H_1 = número de casos positivos en base al método difiere del número de casos positivos verdaderos.

el estadístico de prueba $X^2 = \frac{(b - c)^2}{b + c} \sim \chi^2 (1)$

empleando $\alpha = 0.01$, la región de rechazo serían ($X^2 > 6.635$). En el Anexo III se muestran los resultados obtenidos para cada método. En base a estos resultados, se define que el método Pharmachim es el más apropiado para diagnosticar intoxicación por plaguicidas organofosforados y carbamatos, ya que tiene los porcentajes más altos en sensibilidad y especificidad, además de no rechazarse la hipótesis propuesta en la prueba de McNemar.

5. VALIDACION DE LOS METODOS

En la última etapa de trabajo, se efectuará la validación de los métodos respecto a las condiciones de operación en que se trabajará generalmente para el control de individuos expuestos a plaguicidas organofosforados y carbamatos. A la fecha se ha efectuado únicamente la validación del método de Magnotti.

El término validación se refiere a la capacidad de un método analítico para cuantificar apropiadamente el producto deseado para ello se evalúan la exactitud, precisión, linealidad, límite de detección y especificidad del método (Eugene, 1987). Para la validación del método se trabajó con cantidades conocidas de colinesterasa eritrocítica humana y con muestras de individuos sin exposición a plaguicidas.

Para probar la linealidad del método, se prepararon una serie de diluciones a 0.00, 0.025, 0.05, 1, 2, ..., 7 U.I./ml de enzima realizándose 5 determinaciones a cada concentración excepto para 0.00, donde se hicieron sólo dos. Con estos datos se ajustó la

siguiente recta de regresión :

$$Y = -0.40828 + 12.984 X \quad \text{con } r^2 = .8381$$

SCFA = 2.1026 con 9 gl. , SCEP = 0.6404 con 36gl, lo que permite declarar falta de ajuste en el modelo propuesto ($F=13.13$) y pobre desempeño del método en este rango de concentraciones.

Dado que la variación entre las determinaciones efectuadas a 0.025 y 0.05 fué notable respecto a las otras concentraciones ($CV > 10\%$), se efectuó otro análisis de regresión para concentraciones de 1.0 a 7.0 U.I., a intervalos de una unidad y también por quintuplicado. La recta de regresión ajustada fué:

$$Y_i = 0.036086 + 7.5857 \times 10^{-2} X_i \quad \text{con } r^2 = .9928$$

SCFA = 0.0383 con 5 gl. , SCEP = 0.1074 con 28gl, lo que permite concluir linealidad del método siempre que se opere en el rango de 1.0 a 7.0 U.I, dado que las observaciones efectuadas sobre los grupos en estudio no reportaron valores menores a 1.0 U.I. el método se consideró apropiado, estableciéndose como límite de cuantificación 1.0 U.I.

El intervalo de confianza para la pendiente de la recta es:

$$0.0684 < \beta_1 < 0.0833$$

dado que no se incluye el valor "1", para emplear el método habrá de construirse la curva estandar correspondiente.

La exactitud se evaluó tomando las concentraciones 2, 4 y 6 U.I y expresando resultados como porcentaje de recuperación respecto al estandar añadido, se probó entonces la hipótesis:

$$H_0: \mu = 100 \% \quad \text{vs.} \quad H_1: \mu \neq 100 \%$$

el estadístico de prueba $t^{14} = 0.405676$, no significativo a $\alpha = .05$. por tanto el método se declaró exacto. Para evaluar repetibilidad, se calculó el coeficiente de variación, $CV = 5.07\%$, el cual no excede el 8% que es la recomendación vigente para métodos biológicos dada por el colegio de Químicos Farmacobiólogos de México.

Para evaluar precisión (reproducibilidad), se efectuaron determinaciones a dos concentraciones, por dos analistas y en dos días diferentes. Los datos se analizaron mediante un MANOVA para dos factores, considerando como entradas del vector-respuesta:

X_1 = determinación con 4 U.I y X_2 = determinación con 6 U.I
ningún efecto fué declarado significativo en este análisis.

6. CONCLUSIONES Y RECOMENDACIONES

La metodología estadística resulta vital para analizar los resultados experimentales en Toxicología y Farmacia.

Los tres métodos evaluados: Magnotti, Pharmachim y Michel, detectan diferencias en la actividad colinesterásica cuando se ha tenido exposición a plaguicidas organofosforados y carbamatos.

El método Magnotti no detecta inhibición enzimática mayor al 25% aun para grupos expuestos a plaguicidas con síntomas de intoxicación, por tanto resulta poco apropiado en diagnóstico.

El método Pharmachim es el de más alta especificidad y sensibilidad, siendo de bajo costo y fácil uso, desafortunadamente es un método completamente de importación.

La validación del método Magnotti en base a concentraciones puras de enzima acetilcolinesterasa permite declarar al método apropiado, pero dadas las observaciones biológicas efectuadas en el estudio, su desempeño parece decrecer en sistemas vivos, requiere de mayor estudio en este aspecto si se desea implementar.

Se requiere completar las validaciones de los otros dos métodos a fin de tener información adicional sobre el comportamiento de los tres métodos en condiciones de laboratorio y en vivo.

7. BIBLIOGRAFIA

- 1.- Albert, L.A. (1986). "Repercusiones del uso de plaguicidas sobre ambiente y salud." *Boletín ECO/OPS*, 4.
- 2.- Bennett, B.M. and Underwood, R.E. (1970). "On McNemar test for the 2x2 tabla and its power function." *Biometrics*, 26, 339-313.
- 3.- Eugene, L.I., et. al.(1987). "General Method Validation Guidelines for Pharmaceutical Samples." *Journal of Chromatographic Science*, 25.
- 4- Mausner, J.S. and Bahn, D.K. *Epidemiologia*. Ed. Iiteramericana.

México, 1977.

5.- Morrison, D.F. *Multivariate Statistical Methods*. 2nd. ed. McGraw-Hill, New York, 1984.

6.- SAHR. *Estimacion de insumos para fitosanidad en Mexico*. SAHR. México, 1988.

7.- SPP. Dirección general de Estadística y Censo. México, SPP-1987.

8.- SYSTAT. *User's manual*. Evanston, 1984. 182-188, y 220-224.

ANEXO I

MODELO $Y = \text{constante} + \text{sexo} + \text{edad} + \text{grado de exposición}$

Número de casos procesados: 99

Medias de las variables dependientes

$$X_1 = 5.026 \quad X_2 = 3.154 \quad X_3 = 1.411$$

Estimación de efectos $B = (X'X)^{-1}X'Y$

		X(1)	X(2)	X(3)
Constante		4.490	2.955	2.012
Edad		0.040	0.010	-0.032
Sexo	1	-0.229	0.100	-0.006
Expo	1	0.485	0.264	0.052

Hipotesis para el efecto SEXO

Pruebas F-univariadas

Variable	F(1,95)	Probabilidad
X(1)	6.353	0.013
X(2)	2.083	0.152
X(3)	0.021	0.886

Lambda de Wilks = 0.915 $F_{\text{aprox}}(3,93) = 2.878 \quad \alpha = 0.04$

Traza de Pillai = 0.085 $F_{\text{aprox}}(3,93) = 2.878 \quad \alpha = 0.04$

Traza de Hotelling-Lawley = 0.093 $F_{\text{aprox}}(3,93) = 2.878 \quad \alpha = 0.04$

Hipotesis para el efecto EDAD

Variable	F(1,95)	Probabilidad
X(1)	1.737	0.191
X(2)	0.179	0.673
X(3)	2.205	0.094

Lambda de Wilks = 0.98 $F_{\text{aprox}}(3,93) = 2.411 \quad \alpha = 0.079$

Traza de Pillai = 0.06 $F_{\text{aprox}}(3,93) = 2.411 \quad \alpha = 0.079$

Traza de Hotelling-Lawley = 0.057 $F_{\text{aprox}}(3,93) = 2.411 \quad \alpha = 0.079$

Hipotesis para efecto GRADO DE EXPOSICION

Variable	F(1,95)	Probabilidad
X(1)	34.433	0.000
X(2)	17.584	0.000
X(3)	6.136	0.014

Lambda de Wilks = 0.65 $F_{\text{aprox}}(3,93) = 16.725 \quad \alpha = 0.000$

Traza de Pillai = 0.35 $F_{\text{aprox}}(3,93) = 16.725 \quad \alpha = 0.000$

Traza de Hotelling-Lawley = 0.54 $F_{\text{aprox}}(3,93) = 16.725 \quad \alpha = 0.000$

ANEXO II

Modelo $Y = \text{constante} + \text{edad} + \text{grado de exposición}$

Número de casos procesados: 102

Medias de las variables dependientes

$$X_1 = 4.690 \quad X_2 = 2.535 \quad X_3 = 1.136$$

Estimación de efectos $B = (X'X)^{-1}X'Y$

		X(1)	X(2)	X(3)
Constante		4.234	2.425	1.015
Expo	1	0.147	0.385	0.313
Edad		0.017	0.004	0.003

Hipotesis para efecto EDAD

Variable	F(1,98)	Probabilidad
X(1)	3.671	0.068
X(2)	0.197	0.658
X(3)	0.794	0.375
Lambda de Wilks = 0.959	$F_{\text{aprox}}(3,96) = 1.363$	$\alpha = 0.259$
Traza de Pillai = 0.041	$F_{\text{aprox}}(3,96) = 1.363$	$\alpha = 0.259$
Traza de Hotelling-Lawley = 0.043	$F_{\text{aprox}}(3,96) = 1.363$	$\alpha = 0.259$

Hipotesis para efecto GRADO DE EXPOSICION

Variable	F(1,98)	Probabilidad
X(1)	365.964	0.000
X(2)	148.432	0.000
X(3)	134.385	0.000
Lambda de Wilks = 0.158	$F_{\text{aprox}}(3,96) = 170.682$	$\alpha = 0.000$
Traza de Pillai = 0.842	$F_{\text{aprox}}(3,96) = 170.682$	$\alpha = 0.000$
Traza de Hotelling-Lawley = 5.334	$F_{\text{aprox}}(3,96) = 170.682$	$\alpha = 0.000$

ANEXO III

Metodo de Magnotti

		Exposición	
		+	-
Método	+	0	1
	-	45	59

Sensibilidad = 0%
Especificidad = 1.6%
 $X^2 = 42.1 **$

Metodo de Pharmachim

		Exposición	
		+	-
Método	+	30	4
	-	15	56

Sensibilidad = 66%
Especificidad = 6%
 $X^2 = 6.37 \text{ ns}$

Metodo de Michel

		Exposición	
		+	-
Método	+	33	0
	-	12	60

Sensibilidad = 73.3%
Especificidad = 0%
 $X^2 = 12 *$

ASPECTOS METODOLOGICOS DEL MODELO DE REGRESION LOGISTICA

José Luis Colorado Hdz.
M. Miguel Ojeda Rmz.
Nancy Jácome A.

Laboratorio de Investigación y Asesoría Estadística.
Facultad de Estadística e Informática, U.V.
Av. Xalapa s/n, esq. con Avila Camacho.
C. P. 91000

RESUMEN. - Se presenta de manera general y con la intención de dar un panorama informativo, los aspectos metodológicos básicos del Modelo de Regresión Logística.

Palabras Claves. - Modelo de regresión logístico; metodología; significancia del modelo; bondad de ajuste.

1.- INTRODUCCION.

El análisis de datos mediante el modelo de regresión Logit, se utiliza con gran frecuencia para analizar la posible relación entre una variable dicotómica y un conjunto de variables explicativas. En el contexto de los modelos de regresión, resulta ser una alternativa interesante debido a que se incorpora de manera directa el análisis sobre variable de tipo categórico.

La metodología del modelo de regresión logística se ha usado en diversas áreas, inicialmente en el análisis de Bioensayos, [Berkson, J., (1944)], como una alternativa al análisis Probit, [Finney (1964)], y su uso se ha extendido a diversas áreas, en la sección final de este trabajo se mencionan algunas aplicaciones importantes, particularmente dentro del área de la epidemiología, en donde actualmente tiene un gran auge. Una de las áreas de más reciente aplicación, lo constituye el análisis de encuestas, en donde se han desarrollado conceptos interesantes como por ejemplo

el modelaje multiniveles, para el análisis de datos, [Prosser, B., *et al*, (1990)], así como desarrollos teóricos, también interesantes, como el trabajo de Roberts, Rao y Kumar, [Roberts, G., *et al*, (1987)].

En este trabajo presentamos una serie de aspectos metodológicos importantes, los que permitirán una rápida comprensión del modelo de regresión logit, pensando principalmente en aquellas personas que requieran de realizar aplicaciones.

2.- EL MODELO.

Supongamos que estamos interesados en la relación de un conjunto de variables aleatorias independientes:

$$\underline{X}^t = (X_1, X_2, \dots, X_p)$$

como variables de predicción sobre una variable dicotómica Y , tal que:

$$P[Y=y] = \begin{cases} p & ; y=1 \\ 1-p & ; y=0 \end{cases}$$

El modelo de partida es el siguiente:

$$[1] \quad \pi(\underline{X}) = P[Y=1/\underline{X}] = \frac{\text{EXP}(\beta_0 + \beta^t \underline{X})}{1 + \text{EXP}(\beta_0 + \beta^t \underline{X})}$$

$$\text{donde } \beta^t = (\beta_1, \beta_2, \dots, \beta_p)$$

el cual, luego de algunas operaciones algebraicas puede ser expresadas como:

$$[2] \quad \log \left[\frac{\pi(\underline{X})}{1 - \pi(\underline{X})} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

A este modelo le llamaremos el modelo de regresión logística, y en particular al miembro izquierdo le llamaremos la transformación "logit".

$$\text{logit}(\underline{X}) = \log \left[\frac{\pi(\underline{X})}{1 - \pi(\underline{X})} \right]$$

Es claro que la expresión en (2), representa una relación lineal entre el logit de \underline{X} , y las variables independientes. Por otro lado es interesante notar que hasta ahora no se ha planteado ninguna restricción para el modelo, lo cual nos permite pensar en la posibilidad de que por ejemplo, las variables de predicción puedan ser de cualquier nivel de medición, incluyendo nominales, lo cual resulta sumamente útil en la práctica.

3.- LA ESTIMACION.

El problema de la estimación de los parámetro del modelo, ha sido resuelto en la práctica, por lo regular mediante el método de máxima verosimilitud, de hecho la mayor parte del software que hemos usado utiliza técnicas iterativas para la obtención de dichos estimadores. Sin embargo se pueden utilizar algunos otros métodos de estimación con resultados semejantes, como el de mínima Chi cuadrada, [Berkson, J, (1955); Smith, K.C., (1984)], el método GSK, [Grizzle, J.E., *et al*, (1969)], y existen también la opción de las técnicas de remuestreo, tales como el método Jackknife, y el Bootstrap, [Efron, B., (1982)], los cuales ofrecen opciones interesantes de investigación con apoyo de las computadoras.

En la actualidad la mayoría de los paquetes estadísticos computacionales ofrecen rutinas para el proceso que implica el ajuste y evaluación de modelos logísticos, entre ellos podemos mencionar al BMDP, SYSTAT, SOLO, GLIM, GENSTAT, GAUSS, entre otros.

A continuación y de forma somera, presentamos las ideas principales en la obtención de los estimadores de máxima verosimilitud.

4.- LOS ESTIMADORES DE MAXIMA VEROSIMILITUD.

Supongamos n observaciones independientes del vector de observaciones X con sus correspondientes observaciones en Y , según el modelo planteado será necesario estimar los valores de $\underline{\beta}^t = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$. Mediante el método de máxima verosimilitud se obtendrán $(p + 1)$ ecuaciones de verosimilitud a partir de las cuales, mediante derivación de los logaritmos de las funciones de verosimilitud con respecto a los $(p + 1)$ coeficientes, se obtendrán las expresiones para $\hat{\underline{\beta}}^t = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$; el vector de estimadores.

Siguiendo este planteamiento tenemos:

$$L(\underline{\beta}) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

$$\ln [L(\underline{\beta})] = \sum_{i=1}^n \{y_i \ln [\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\}$$

posteriormente se obtienen las derivadas de $\ln [L(\underline{\beta})]$ con respecto de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$, y luego se iguala con cero, obteniéndose:

$$\sum_{i=1}^n [y_i - \pi(\underline{x}_i)] = 0$$

$$\sum_{i=1}^n x_{ij} [y_i - \pi(\underline{x}_i)] = 0; \quad \text{para } j=1,2,\dots,p$$

Las cuales deberán de resolverse para β mediante técnicas numéricas computacionales.¹

Las estimaciones de las varianzas y errores estandar de los $\hat{\beta}_i$'s para $i=0,1,\dots,p$; se obtienen a partir de la matriz de segundas derivadas parciales de la función logaritmo de la verosimilitud:

$$\frac{\delta^2 \text{Log} [L(\beta)]}{\delta \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi(\underline{x}_i) (1 - \pi(\underline{x}_i))$$

$$\frac{\delta \text{Log} [L(\beta)]}{\delta \beta_j \delta \beta_u} = - \sum_{i=1}^n x_{ij} x_{iu} \pi(\underline{x}_i) (1 - \pi(\underline{x}_i))$$

para $j, u = 0, 1, 2, \dots, p$.

Con los resultados de estas ecuaciones se obtiene una matriz de orden $(p + 1) * (p + 1)$, cuya inversa nos proporciona la matriz de varianzas y covarianzas $\hat{\beta}$, denotada por $\Sigma(\beta)$.

¹Para detalles sobre las técnicas que se usan para resolver este problema, se puede revisar Mc Cullagh and Nelder (1983).

$$\Sigma(\beta) = \begin{bmatrix} \sigma^2_{(\beta_0)} & \sigma_{(\beta_0, \beta_1)} & \cdot & \cdot & \sigma_{(\beta_0, \beta_p)} \\ \sigma_{(\beta_1, \beta_0)} & \sigma^2_{(\beta_1)} & \cdot & \cdot & \sigma_{(\beta_1, \beta_p)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{(\beta_p, \beta_0)} & \sigma_{(\beta_p, \beta_1)} & \cdot & \cdot & \sigma^2_{(\beta_p)} \end{bmatrix}$$

Los estimadores de los elementos de $\Sigma(\beta)$ se obtienen evaluando $\Sigma(\beta)$ en $\hat{\beta}$.²

Por lo regular lo que se usará en las aplicaciones son los errores standar de los estimadores, cuyas estimaciones denotamos por:

$$\hat{\sigma}_{(\beta_j)} = (\sigma^2_{(\beta_j)})^{1/2}$$

5.- LA SIGNIFICANCIA DEL MODELO.

Para probar la significancia del modelo estimado, recurrimos a probar la hipótesis de que los coeficientes β_i ; $i = 1, 2, \dots, p$ son iguales a cero; o sea que no tienen efecto en la predicción de $\pi(\underline{x}_i)$. Para ésto se utiliza una prueba de razón de verosimilitudes, La cual proporciona la siguiente estadística:

² Se entiende que dichas elementos quedaran expresados en función de los $\hat{\beta}_i^s$; $i = 1, 2, \dots, p$.

$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n_0} \right)^{n_1} \left(\frac{n_0}{n_1} \right)^{n_0}}{\prod_{i=1}^n \hat{\pi}(x_i)^{y_i} (1 - \hat{\pi}(x_i))^{(1-y_i)}} \right]$$

$$G = 2 \left\{ \sum_{i=1}^n [y_i \ln(\hat{\pi}(x_i)) + (1-y_i) \ln(1 - \hat{\pi}(x_i))] \right. \\ \left. - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right\}$$

con $n_1 = \sum_{i=1}^n y_i$ y $n_0 = \sum_{i=1}^n (1-y_i)$

la cual tiene distribución $\chi^2_{(p)}$ para cuando n tiende a ∞ .

Existen dos pruebas alternativas la de Wald y la de Puntajes (Score Test). La primera en su versión univariada nos permite evaluar la significancia de cada una de las variables independientes por separado, en base al estadístico:

$$W_j = \frac{\hat{\beta}_j}{SE_{\hat{\beta}_j}}$$

el cual, bajo la hipótesis de nulidad, $\beta_j=0$, tiene distribución Normal Estandar. Las versiones multivariadas de la prueba de Wald, y la de puntajes, requieren de cálculos matriciales, y no es frecuente encontrar los resultados en la mayoría de los paquetes computacionales. sin embargo es altamente probable que si se dispone de software especializado, como GLIM, GENSTAT, o GAUSS, se puedan programar de manera relativamente fácil. Para los aspectos teóricos de estas pruebas, Lemshow y Hosmer (1990), recomiendan

los libros de Cox y Hinkley (1974), así como el de Dobson (1983).

6.- LA EVALUACION DE LA BONDAD DEL AJUSTE.

Para evaluar la bondad del ajuste del modelo se han utilizado tradicionalmente dos pruebas, una basada en la estadística de Pearson, y otra basada en la Deviance. Las dos estadísticas tienen distribución Chi Cuadrada bajo la hipótesis de nulidad (buen ajuste).

Dichas pruebas están basadas en las diferencias entre los valores observados y los valores estimados, usualmente llamados residuales.

Supongamos, como se ha planteado anteriormente, que tenemos p variables independientes en el modelo. Usaremos la siguiente terminología.

J = El número de diferentes combinaciones de valores de las p variables independientes, de tal manera que si algunas observaciones coinciden en sus valores, entonces $J < n$.

m_j = el número de observaciones tales que $\underline{x} = \underline{x}_j$, para $j=1,2,\dots,J$. Y entonces $\sum m_j = n$.

Y_j = Es el número de respuestas $y=1$ para el j -ésimo grupo de observaciones tales que $\underline{x} = \underline{x}_j$. Y entonces $\sum Y_j = n_1$ es el número de resultados $y=1$.

Para la prueba basada en la estadística de Pearson, los residuales se definen como:

$$r[y_j, \pi(x_j)] = \frac{y_j - m_j \pi(x_j)}{\sqrt{m_j \pi(x_j) [1 - \pi(x_j)]}}$$

y la estadística de prueba será:

$$\chi^2_{\text{cal}} = \sum_{j=1}^J (r[y_j, \pi(x_j)])^2$$

Para el caso de la Deviance, el residual se define como:

$$d[y_j, \pi(x_j)] = 2 \left\{ y_j \ln \left[\frac{y_j}{m_j \pi(x_j)} \right] + (m_j - y_j) \ln \left[\frac{m_j - y_j}{m_j (1 - \pi(x_j))} \right] \right\}^{1/2}$$

donde el signo inicial es el de $(y_j - m_j \pi(x_j))$

Y la estadística de prueba será:

$$D_{\text{cal}} = \sum_{j=1}^J (d[y_j, \pi(x_j)])^2$$

Tanto la estadística de Pearson como la de Deviance, tienen distribución Chi cuadrada con $(J - (p+1))$ grados de libertad, bajo la hipótesis nula. (Hosmer D.W., Jr., Lemeshow S. (1989)), para el caso en que el número J de estratos es considerablemente menor que n.

Existe otra prueba de reciente definición, [Lemeshow and Hosmer (1982)], también basada en la estadística de Pearson, la cual resulta ser una buena opción para cuando $J \approx n$; caso en el cual las estadísticas mencionadas anteriormente no tienen una buena aproximación a la Chi cuadrada. Para otras pruebas se puede consultar Tsiatis, A.A. (1980) y LeCessie, S.-Howelingen, J.C. (1989).

Para explicar la forma de obtener el estadístico para esta prueba, pensemos una tabla de 2 por n, en donde las filas

corresponden a los resultados 1,0 de la variable Y, y las n columnas corresponden a las n probabilidades estimadas para cada una de las n observaciones.

Se proponen dos estrategias para colapsar la tabla en 10 grupos, de tal manera que se tengan solo J columnas. una en base a los percentiles y otra en base a valores fijos de las probabilidades estimadas. La manera de obtener los valores esperados para las filas será sumando las probabilidades estimadas por cada grupo, para el primer renglón, y para el segundo sumando [1-(la probabilidad estimada)], también para cada uno de los 10 grupos. Con estos resultados se realiza una prueba de Chi cuadrada sobre la nueva tabla de 2 por 10. El estadístico resultante es:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n_k \cdot \bar{\pi}(x_k))^2}{n_k \cdot \bar{\pi}(x_k) (1 - \bar{\pi}(x_k))}$$

donde:

n_k = número de observaciones tales que $x=x_k$

$O_k = \sum_{j=1}^{n_k} y_j$ = número de respuestas $y=1$ en los n_k patrones de covariables.

$\bar{\pi}_k = \sum_{j=1}^{n_k} [\hat{\pi}_j(x_k)] / n_k$ = el promedio de las probabilidades estimadas en el k-ésimo grupo.

7.- ALGUNAS AREAS ADICIONALES DE APLICACION DEL MODELO LOGIT.

El modelo logit es sin duda uno de los modelos más populares en el análisis de respuestas dicotómicas, de hecho los modelos

loglineales para esta situación pueden ser vistos como un caso particular de estos modelos [Haberman (1979)]. Quiza una de las áreas donde el modelo logit ha encontrado mayores aplicaciones es en el área de investigación médica y epidemiológica. Por ejemplo, si $y=1$ cuando la enfermedad está presente y cero en otro caso, y \underline{x} es un vector de p variables explicatorias asociadas, (generalmente los factores de riesgo de la enfermedad), y estamos interesados en modelar $E[y/\underline{x}] = P(y=1/\underline{x}) = P(\underline{x})$, el modelo logístico describirá el riesgo de prevalencia de la enfermedad como una función de los factores de riesgo, de la siguiente manera

$$\text{logit}[P(\underline{x})] = \log \left[\frac{P(\underline{x})}{(1-P(\underline{x}))} \right] = \alpha + \beta^t \underline{x}$$

Generalmente en estudios epidemiológicos se tienen datos sin agrupar, es decir no como en el caso de los bioensayos donde se fija un punto en R^1 y se ensaya repetidamente, n_i veces, para observar $P(x_i)$. Un enfoque bastante aceptado es agrupar los datos y graficar las proporciones para tener una idea de la razonabilidad al ajustar el modelo logit.

El modelo logit puede ser adaptado facilmente para situaciones de estudios observacionales. Por ejemplo se puede usar la versión condicional para estudios de casos y controles apareados [Breslow y Day (1980), Holford y otros (1989), Breslow y Zhao (1988)], para estudios retrospectivos y prospectivos [Prentice (1976)], para diseños seccionales cruzados (cross-sectional), y en general para estudios observacionales [Davies y Pickles (1985) y Davies (1989)]. En este sentido existe una basta literatura sobre el tema, la cual considera desde aspectos tradicionales de análisis, problemas computacionales, y generalizaciones del modelo logit que permite incorporar cada vez mas situaciones de estudio. El libro de Hosmer y Lemeshow (1989), la excelente monografía de McCullagh y Nelder (1983), así como el trabajo de Frome (1986), son un adecuado punto de partida para el estudio de esta importante familia de modelos.

BIBLIOGRAFIA

- Berkson, J. (1955). Maximum Likelihood and minimum χ^2 estimates of the logistic function. Journal of the American Statistical Association 50, 130-162.
- Berkson J. (1944). Application of the logistic function to bio-assay. Journal of the American Statistical Association 39, 357-365.
- Binder D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. Int. Stat. Review 51; p.p. 279-292.
- Binder D.A., Gratton M.A., Hidioglow M.A., Kumar S. and Rao J.N.K. (1984). Analysis of Categorical Data From Surveys With Complex designs: Some Canadian Experiences. Survey Methodology; 10, p.p. 141-56.
- Breslow N.E. and Day N.E. (1980). Statistical Methods in cancer research vol 1. Int. Agency for Research on Cancer, Scientific. Publication # 32.
- Breslow N.E. and Zhao L.P. (1988). Logistic Regression for stratified case control studies. Biometrics 44; p.p. 891-899.
- Cox, D.R., and Hinkley, D.V. (1974). Theoretical Statistics. Chapman Hall, London.
- Dobson, A. (1983). An Introduction to statistical Modelling. Chapman Hall, New York.
- Davies R.B. (1989). Robustness in modelling Dynamics of choice. Haver et.al. (eds); Urban Dinamics and Spatial Choice Behaviours, p.p. 59-80. Kluwer Academic Publishers.
- Davies R.B. and Pickles A.R. (1985). Longitudinal versus cross-sectional methods for behavioural research; a first-round Rnockout. Environment an Planing A. 17; p.p. 1315-1329.
- Efron B. , Gong G. (1982) A Leisurely Look at the BootStrap, the Jackknife and cross-validation. Technical Report Num. 75. Stanford University, Division of Biostatistic.
- Finney, D.J., (1964), Statistical Method in Biological Assay. Griffin, London.
- Frome E.L. (1986). Regression methods for Binomial y Poisson Distributed Data. Multiple Regression Analysis: Applications in the Health Sciences; Herbert and Myers (eds). The American Institute of Physics.

- Grizzle, J. E. , Starmer, C. F. , and Koch, G. G. (1909). Analysis of Categorical Data By Linear Models. Biometrics, 25: 489-504.
- Haberman S.J (1979). Analysis of Qualitative data. vol 2. Academic Press, New York.
- Holford T.R., Bracken M.B. and Eskenazi B. (1989). Loglinear models for the analysis of Matched cohort studies. Am. Jour. of Epidemiology, vol. 130, # 6; p.p. 1247-1253.
- Hosmer D.W. Jr., Lemeshow S. (1989). Applied Logistic Regression. John Wiley & Sons.
- LeCessie, S. and van Houwelingen, J.C. (1989). Goodness of Fit Tests for Binary Regression Models Based on Smoothing Methods. Technical Report 7, Dept. of Medical Statistics, University of Leiden, The Netherlands.
- Lemeshow, S., and Hosmer, D.W. (1982). The use of goodness-of-fit statistics in the development of regression models.
- Mc Cullagh, P. , and Nelder . J. A. (1983). Generalized Linear Models. Chapman Hall, Londen.
- Neuhaus J.M., Kalbfleisch J.D. and Hauck (1990) A comparison of clusters specific and population-averaged approaches for analyzing correlated binary data. Tech. Rep. # 3, Group in Biostatistics, U. of California, Sn. Francisco.
- Prosser B. Rasbash J. and Goldstein M. (1990) ML3 - Software for the tree-level Analysis. Institute of Education University London.
- Prentice R. (1976). Use of the logistic model in Retrospective Studies. Biometrics 32: p.p. 599-606.
- Qu Y, Beck G.J. and Williams (1990). Polya-Eggenberger Distribution: Parameter Estimation and Hypothesis test. Biom. J. 32,2 229-242.
- Qu Y. Williams G.W., Beck G.J. and Goormastic (1987). A Generalized Model of Logistic Regression for clustered Data. Commun. Statist-Theory-Meth. 16(12), p.p. 3447-3476.
- Rosner B. (1984). Multivariate Methods in Ophthalmologig with application to other paired data situations. Biometrics 40: p.p. 1025-35.
- Roberts G. , Rao J. N. K. , Kumar S. (1987). Logistic Regression analysis of sample survey data. Biometrika (1987), 74. 1. pp. 1-12.

Smith K.C., Savin N.E., Robertson J.L. (1984). A Montecarlo Comparison of Maximum Likelihood and Minimum Chi-Square Sampling Distribution in Logit Analysis. Biometrics 40, 471-482 June 1984.

Tsiatis, A.A., (1980). A note on a Goodness-of-Fit Test for the Logistic Regression Model. Biometrika 67, 250-251.

APLICACIONES ECOLOGICAS DEL MODELO LOGISTICO

L. Cruz-Kuri

I. Sosa-Galindo

INSTITUTO DE CIENCIAS BASICAS, U. V.
APDO. POSTAL 177, XALAPA, VER.

RESUMEN : En este trabajo se presenta la derivación clásica del modelo logístico, via la solución de una ecuación diferencial en el contexto determinista de crecimiento de poblaciones; asimismo se discute el enfoque probabilista y se hace una aplicación de ajuste a un conjunto de datos generados por un experimento de búsqueda de cebos en una especie de hormigas. Congruente con esto último, se plantea el problema acerca de cual distribución espacial puede dar lugar al modelo logístico; problema que se está buscando resolver mediante simulación por computadora, con lo cual se justificaria la utilización de dicho modelo para todos los problemas similares, según la interpretación de Holling [ver referencia al final]. En al menos un caso, se hace ver como pueden obtenerse resultados exactos.

PALABRAS CLAVE: Modelo logístico, procesos estocásticos, captura, simulación por computadora, ajuste.

1 INTRODUCCION

Motivado por una serie de experimentos sobre captura de cebos por colonias de hormigas [D. Nestel, 1989] en las que se tomaron datos sobre los tiempos de captura, se trató de ajustar los resultados mediante el modelo logístico. Existen antecedentes, e.g., en los trabajos de C.S. Holling en los que se consideran poblaciones de insectos y se asume un esquema de búsqueda uniforme por parte de los depredadores, donde las "presas" son los insectos o sus huevecillos a capturar, los cuales se supone se encuentran distribuidos de manera regular sobre el habitat. Surge entonces la pregunta natural si dicho esquema genera como modelo estocástico el de la distribución logística y, en caso negativo, bajo que condiciones de distribución de presas y de esquema de distribución probabilista de búsqueda por parte de los depredadores puede surgir el modelo logístico. En el presente trabajo se dan pasos

hacia la solución de dicho problema por medio de simulaciones de procesos de captura en el plano ejecutadas con una computadora digital; asimismo, se hace ver como se puede resolver de manera exacta el problema cuando el esquema de búsqueda se particulariza al de la distribución uniforme sobre alguna región finita del plano, permitiendo establecer comparaciones entre los resultados aproximados mediante simulación y los resultados exactos de nuestro análisis. Las situaciones en que se incluyen distribuciones espaciales de búsqueda distintas de la uniforme, son necesariamente más complejas y requieren la consideración adicional de las distribuciones de ubicación de las presas.

2. DERIVACION DEL MODELO LOGISTICO

En esta sección se presenta la derivación clásica del modelo logístico, vía la solución de una ecuación diferencial en el contexto determinista de crecimiento de poblaciones, cuando éste depende de la densidad.

Sea $N(t)$ el tamaño de la población en el instante t . Se hace la suposición de que el medio ambiente donde se desarrolla la población tiene recursos finitos (con lo cual el tamaño de la población queda acotado); asimismo se hace la suposición de que la tasa de crecimiento por individuo es una función del tamaño de la población. Es decir,

$$dN(t)/dt = Nf(N)$$

Para N grande, df/dN debe ser negativa, o sea que, entre más grande sea la población, mayor será el efecto inhibitorio para un crecimiento adicional. La manera más simple de hacer que se cumpla esto es suponer que $f(N)$ es una función lineal de N .

$$f(N) = r - sN \quad r, s > 0 \quad ; \quad N > 0.$$

Esto conduce a

$$dN/dt = N(r - sN)$$

que es la conocida ecuación logística de Verhulst-Pearl. Las cantidades r y s que aparecen en dicha ecuación tienen las siguientes interpretaciones: r es la tasa intrínseca de crecimiento natural y $s = r/K$, donde K es el tamaño máximo alcanzable para la población. Mediante procedimientos directos de integración, se

puede resolver facilmente la ecuación logística. Una forma de solución es la que se presenta a continuación.

$$N(t) = K/[1 + \exp [-r(t-t_0)]]$$

donde se tiene como condición inicial $N(t_0) = N_0$. Por supuesto, r y s están relacionados por

$$s = \exp [-r t_0]/C$$

$$C = N_0/(r - N_0 s)$$

Además, cuando $t \rightarrow \infty$, $N(t) \rightarrow K = r/s$. Así que, K representa el nivel de saturación de la población (el cual no puede ser excedido en virtud de las limitantes del medio ambiente).

A continuación se presenta como ilustración el caso del modelo de crecimiento logístico de una población con los valores paramétricos y condiciones iniciales que se indican.

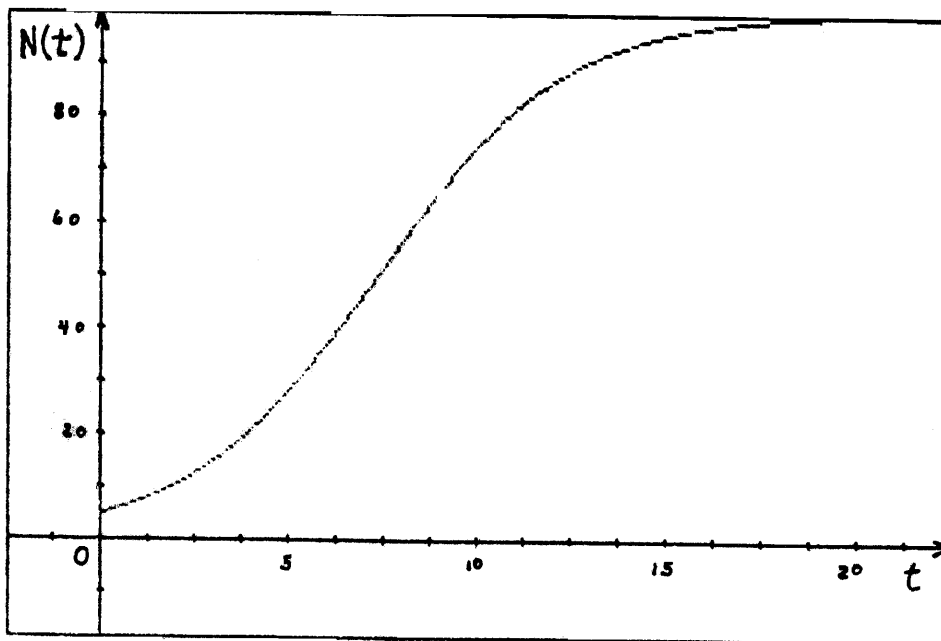


FIG. 1. Curva logística con $r = 0.5$, $K = 100$, $N_0 = 5$, $t_0 = 5.89$.

3. PROCESOS ESTOCÁSTICOS DE CAPTURA

En una serie de experimentos de captura de "cebos" por colonias de hormigas, D. Nestel (comunicación personal a uno de los autores, 1989) colectó datos sobre los tiempos de captura de cada uno de los cebos. Las hormigas en su organización altamente estructurada tienen "exploradoras" que buscan las "presas". Una vez encontrada una, lo transmiten a la colonia, quien se encarga de enviar hormigas

especializadas para la explotación del "recurso". Las hormigas exploradoras continúan con la búsqueda de la siguiente "presa". Las gráficas de los tiempos promedio tienen un aspecto parecido al de la figura 1, motivo por lo cual parece natural el tratar de ajustar un modelo logístico en el que, por supuesto, se hace una estimación de los parámetros. Sin embargo, C.S. Holling (1961) había considerado un modelo en que los depredadores mediante cierto esquema de búsqueda capturan a sus presas. Holling sugirió una simulación analógica de "captura" de discos de cartón distribuidos de manera regular sobre una mesa y donde los "depredadores" eran personas con los ojos vendados intentando encontrar cada disco. En ese experimento, cada vez que un disco era "capturado" por una persona se quitaba de la mesa de tal manera que la captura de discos adicionales representaba un esfuerzo cada vez mayor en virtud de la disminución de densidad de "presas".

Motivados por las situaciones anteriores, decidimos avocarnos a la modelación de tales experimentos partiendo de suposiciones relativas, primero, al esquema de la distribución espacial de las presas y, segundo, al esquema de búsqueda de las presas por parte de los depredadores. Mas específicamente, se busca incorporar al modelo características tales como

A. Esquema de distribución espacial de las presas

- A1. Patrón regular
- A2. Patrón aleatorio

y

B. Esquema de búsqueda de los depredadores

- B1. Distribución uniforme
- B2. Otra distribución probabilista (e.g. normal bivariada truncada).

Para facilitar la discusión se supondrá que el habitat donde se distribuyen las presas es una región finita del plano, la cual tomaremos de forma rectangular para el presente artículo, y se representaran las "presas" como círculos o discos de un radio r dado, dispersas según cierto esquema de los considerados en A. Entonces, el proceso de búsqueda es como sigue: se selecciona aleatoriamente un punto del plano (rectángulo) de acuerdo a cierta distribución bivariada; si el punto seleccionado pertenece a algún disco, se dirá que ha ocurrido una "captura" y, en tal caso, se

remueve para la siguiente etapa: en la segunda etapa, se hace la selección nuevamente de un punto del plano con la misma distribución y se remueve un disco en caso de caer el punto dentro de tal. El proceso continua de esta manera hasta lograr la eliminación o "captura" de todos los discos, anotandose los intentos requeridos en cada caso para la eliminación de cada disco. En simbolos, sean

A = area del habitat (rectangulo)

r = radio de la "presa" (disco)

N = numero de presas a capturar

T_k = tiempo requerido para capturar el k-ésimo disco

X_k = tiempo acumulado hasta la captura de los k primeros discos ($k = 1, 2, 3, \dots, N$)

Se sigue que

$$X_k = T_1 + \dots + T_k$$

donde los sumandos tienen distribuciones que dependen de los esquemas de distribución de los discos, de la distribución de búsqueda y, por supuesto, de A, r y N.

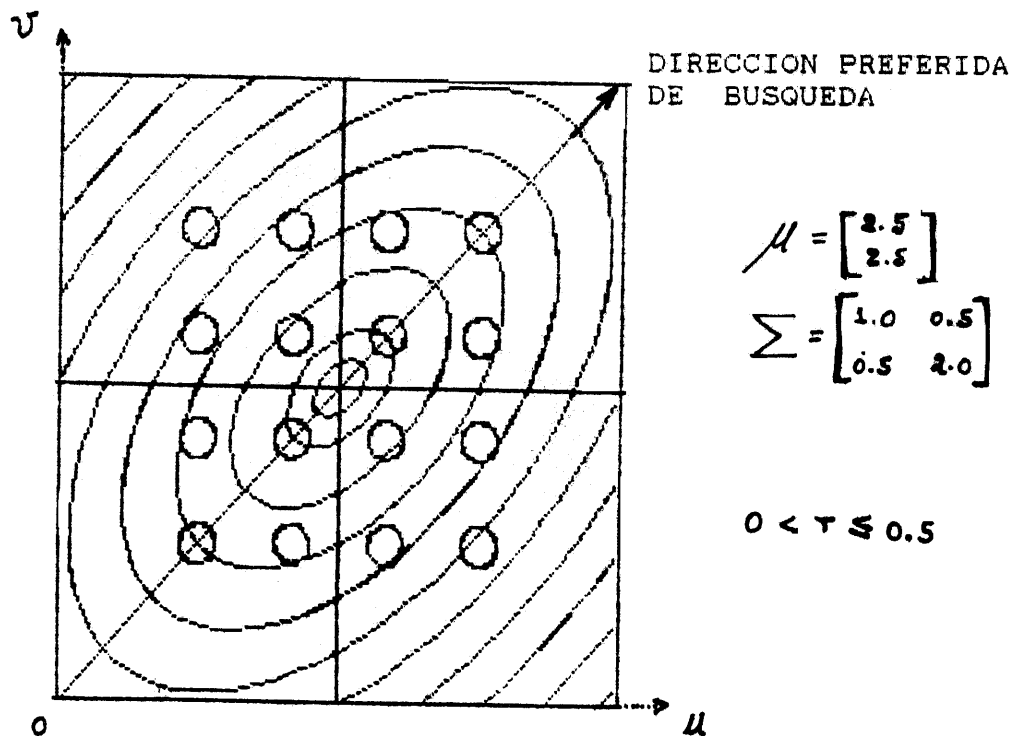


FIG. 2. Configuración 4 x 4 de 16 discos con esquema de búsqueda según la distribución normal bivariada truncada. Las elipses concéntricas corresponden a las curvas de nivel de la densidad.

El problema en general es: encontrar la distribución probabilista de X_k . Asimismo, calcular los momentos de segundo orden de X_k y X_l ($k, l = 1, 2, 3, \dots, N$).

↳ Bajo que condiciones se tiene que los puntos $(k, E(X_k))$ pertenecen a una misma curva logística para $k = 1, 2, \dots, N$?

Segun se vera en las secciones que restan, tales problemas quedan resueltos de manera exacta o aproximada dependiendo de las suposiciones acerca de los esquemas de distribución de los discos y de búsqueda de éstos.

4. ANALISIS EXACTO DEL CASO (A1.B1)

Con la notación de la sección anterior, sea

$$p_0(r) = \pi r^2 / A$$

$$p_k(r) = (N-k+1) p_0(r)$$

Es claro que T_k es una variable aleatoria con distribución geométrica especificada por el parametro $p_k(r)$. Además, como el esquema de búsqueda es uniforme, no importa el patron de distribución de los discos, aunque lo suponemos regular para propósitos de hacer concreta la simulación en la siguiente sección. También es claro que las T 's son independientes.

Mediante el cálculo de la función generadora de probabilidades de T_k , por ejemplo, y de la independencia de las T 's se obtiene

$$E(X_k) = \sum_{j=1}^k (1/p_j(r))$$

$$\text{VAR}(X_k) = \sum_{j=1}^k (q_j(r)/[p_j(r)]^2)$$

donde $q_j(r) = 1 - p_j(r)$. De manera más explícita,

$$E(X_k) = [(1/N) + (1/(N-1)) \dots (1/(N-k+1))] / p_0(r)$$

$$\text{VAR}(X_k) = [(1-Np_0)/N^2 + \dots + (1-(N-k+1)p_0)/(N-k+1)^2] / p_0^2.$$

Una vez más, aprovechando la independencia de las T 's, es posible obtener de manera directa el coeficiente de correlación entre X_i y X_j . Mas precisamente,

$$\text{CORR}(X_i, X_j) = [\text{VAR}(X_i) / \text{VAR}(X_j)]^{1/2}$$

siempre que $1 \leq i < j \leq N$. Los resultados se pueden hacer

explicitos mediante la formula anteriormente obtenida para las variancias.

Resumiendo, se han encontrado de manera exacta los momentos hasta el segundo orden del proceso estocastico $\{X_k : k = 1, 2, \dots, N\}$.

5 SIMULACION DE LAS TRAYECTORIAS

Bajo las suposiciones de distribución no uniforme para la búsqueda de los discos, el proceso estocastico de captura adquiere una complejidad que hace muy difícil un análisis exacto como el realizado en la sección 4, motivo por lo cual se recurre a un análisis aproximado mediante el metodo de simulacion por computadora. Para tales propositos, los autores del presente trabajo, con la ayuda de un especialista en ingenieria de sistemas computacionales [A. J. Garcia Banda, del Instituto de Ciencias Basicas de la Universidad Veracruzana], y el auxilio del paquete de computo estadístico BMDP, elaboraron programas de computo en FORTRAN que permiten la simulacion de las trayectorias del proceso de captura. Asimismo, con la aplicacion del BMDP se realizaron los analisis estadísticos de las trayectorias simuladas

El proceso de simulación parte de una rutina confiable de generacion de numeros pseudo aleatorios con distribución uniforme sobre el intervalo $[0,1]$. Para el caso de la distribución normal bivariada, tambien se hizo uso de una subrutina de simulacion. Las dificultades técnicas encontradas y los algoritmos utilizados se discuten en otro lugar. Aquí unicamente se hace mención que el programa de simulacion tiene suficiente flexibilidad como para incluir distintos patrones de distribución de los discos sobre el plano y de manejar mediante instrucciones condicionales distintas distribuciones espaciales de búsqueda. El programa incorpora una capacidad de captura jerarquizada de los discos, aun cuando aparezcan traslapes entre estos, sin llevar a ambigüedades. Una vez realizada una serie de simulaciones, quedan registradas las trayectorias del proceso de captura en un archivo de la computadora, el cual se analiza estadísticamente con los programas usuales del BMDP, tales como el P2D de estadísticas básicas y distribuciones de

frecuencias y el P3D de cálculo de correlaciones y T-cuadrada de Hotelling para propósitos de comparación de medias vectoriales.

En lo que sigue se presentan los resultados de dos simulaciones del proceso de captura de 9 discos, cada uno de radio 0.2 unidades, distribuidos regularmente dentro de un cuadrado de lado igual a 4 unidades, teniendo los discos sus ubicaciones de tal manera que la separación horizontal entre los centros de dos discos consecutivos es igual a una unidad y lo mismo que la separación vertical. El proceso de búsqueda en este ejemplo se toma igual al de la distribución uniforme sobre el cuadrado con la finalidad de poder estudiar la precisión de las simulaciones ya que, en este caso, se pudieron obtener los resultados exactos, tal y como se menciona en la sección 4.

TABLA 1. Tiempos promedio de captura y desviaciones estandar basado en dos simulaciones del proceso de captura de nueve discos de radio 0.2 y búsqueda uniforme.

Tiempo de Captura	M e d i a s		Desviaciones Estandar	
	Simulación 1	Simulación 2	Simulación 1	Simulación 2
X_1	12.29	11.11	13.48	9.93
X_2	26.87	24.95	18.36	16.97
X_3	42.95	36.68	21.78	20.31
X_4	60.69	59.23	27.24	29.41
X_5	84.98	79.72	32.76	32.46
X_6	113.88	112.87	45.29	44.29
X_7	165.20	159.42	72.14	69.12
X_8	237.26	240.75	101.58	106.22
X_9	430.78	445.05	296.73	237.27

TANDA 1 : $n_1 = 116$ trayectorias simuladas

TANDA 2 : $n_2 = 111$ trayectorias simuladas

(NOTA Cada tanda de simulaciones requirió aproximadamente de 50,000 pares de números pseudo aleatorios)

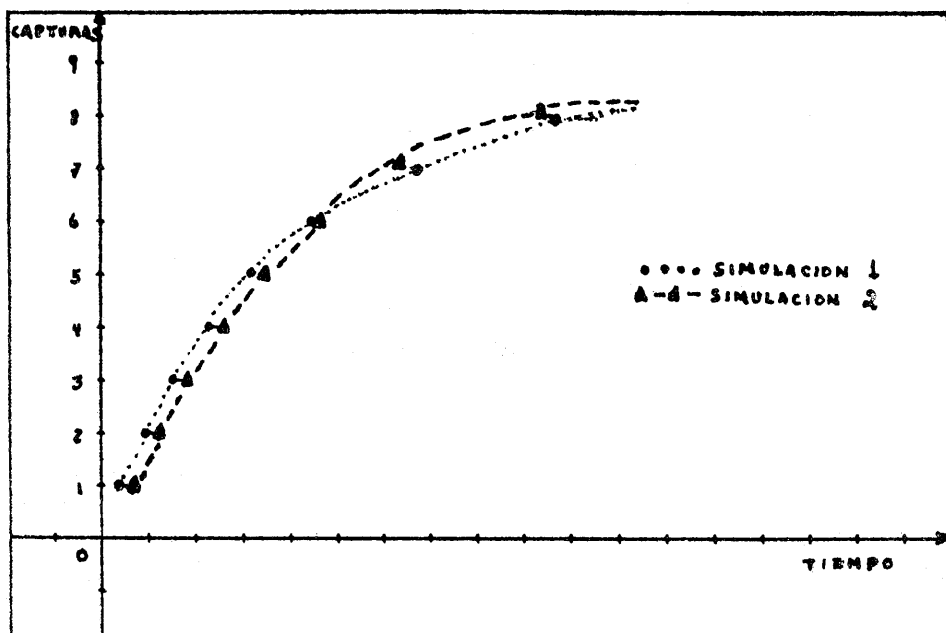


FIG.3. Tiempos promedio de captura para 9 discos ($r = 0.2$)

TABLA 2. Valores extremos para dos tandas de simulaciones del proceso estocástico de captura de 9 discos ($r = 0.2$)

Tiempos de Captura	S I M U L A C I O N 1		S I M U L A C I O N 2	
	Minimo	Maximo	Minimo	Maximo
X_1	1	78	1	45
X_2	2	97	3	123
X_3	8	118	5	124
X_4	9	135	15	168
X_5	24	184	23	182
X_6	38	309	44	227
X_7	46	522	52	366
X_8	52	712	76	561
X_9	54	1936	127	1241

$n_1 = 116$ trayectorias

$n_2 = 111$ trayectorias

TABLA 3. Matriz de correlaciones entre los tiempos de captura.
Segunda tanda de simulaciones (r = 0.2)

	X1	X2	X3	X4	X5	X6	X7	X8	X9
X1	1.0000								
X2	0.4844	1.0000							
X3	0.3078	0.7628	1.0000						
X4	0.1444	0.4741	0.7033	1.0000					
X5	0.0910	0.3917	0.6133	0.8386	1.0000				
X6	0.0776	0.2847	0.4830	0.6011	0.7447	1.0000			
X7	0.1405	0.1522	0.3238	0.4975	0.6200	0.7778	1.0000		
X8	0.0076	0.1545	0.2826	0.3646	0.4612	0.5456	0.7010	1.0000	
X9	0.1707	0.0203	0.0469	0.0919	0.0917	0.0365	0.2578	0.4281	1.0000

n = 111 trayectorias

TABLA 4. Instrucciones en BMDP para ejecución del programa P3D que analiza vectorialmente dos tandas de simulaciones del proceso de captura de 9 discos.

```

$ PROBLEM          TITLE IS "TRAYECTORIAS PROCESO ESTOCASTICO".

$ INPUT            VARIABLES ARE 11.
                   FILE IS 'UNION_AB.DAT'.
                   FORMAT IS '(F7.0,8F6.0,F9.0,F1.0)'.

$ VARIABLES        NAMES ARE FOR J = 1 TO 11 % TIEMPO (J, %.
                   USE = TIEMPO1 TO TIEMPO 9, TIEMPO11.
                   GROUPING IS TIEMPO11.

$ GROUP            CODES(1) ARE 1,2.
                   NAMES(1) GRUPO1 TO GRUPO2.

$ TEST            VARIABLES ARE TIEMPO1 TO TIEMPO9.
                   HOTELLING.
                   CORRELATION.
                   NONPARAMETRIC.

$ PRINT            LINE = 72.
                   PAGE = 0.

$ END

```

6. COMPARACION DE LAS SIMULACIONES CON LOS RESULTADOS EXACTOS

Presentamos a continuación la comparación entre algunos de los resultados obtenidos mediante simulación y los correspondientes del análisis exacto de la sección 4 para el caso (A1,B1) aplicado a la configuración particular 3 x 3 sobre el cuadrado [0,4] x [0,4] cuando el radio de cada disco es $r = 0.2$.

En la notación de la sección 4, se tiene

$$N = 9$$

$$A = 4 \times 4 = 16$$

$$p_0(r) = \Pi(0.2)^2/16 = 0.007854$$

$$p_1(r) = 9p_0(r) = 0.0707$$

con lo cual

$$E(X_1) = E(T_1) = 1/p_1(r) = 14.1471$$

$$\text{VAR}(X_1) = \text{VAR}(T_1) = q_1(r)/(p_1(r))^2 = 185.9935$$

$$\text{DE}(X_1) = 13.6379.$$

Análogamente,

$$p_2(r) = 8p_0(r) = 0.062832$$

$$E(T_2) = 1/p_2(r) = 15.9153$$

$$E(X_2) = E(T_1) + E(T_2) = 30.0626$$

$$\text{VAR}(T_2) = p_2(r)/(q_2(r))^2 = 237.3875$$

$$\text{DE}(X_2) = [\text{VAR}(T_1) + \text{VAR}(T_2)]^{1/2} = 20.5762$$

$$\text{CORR}(X_1, X_2) = \text{DE}(X_1)/\text{DE}(X_2) = 0.6628.$$

El valor esperado exacto, por ejemplo, para X_n se obtiene mediante la relación

$$E(X_n) = [(1/9) + (1/8) + (1/7) + \dots + (1/2)]/p_0(r) = 232.8709.$$

En tanto que el valor estimado en la primera tanda de 116 simulaciones es de 237.26 y, para la segunda tanda de 111 simulaciones, es de 240.75.

La correlación estimada entre X_1 y X_2 es de 0.6825 para la primera tanda, las desviaciones estándar son de 13.4785 y 18.3642, respectivamente, en congruencia con la tabla 1. Para la segunda tanda se tienen una correlación de = 0.4844 entre X_1 y X_2 y desviaciones estándar de 9.9347, 16.9689 siendo menos satisfactoria la concordancia con los valores exactos.

anterior hace pensar en la necesidad de aumentar la cantidad de pares de numeros pseudo aleatorios generados (que como se ha mencionado fue de aproximadamente 50,000 pares en cada tanda), asi como tener mayor seguridad de que el algoritmo de generaci3n pasa las pruebas usuales de aleatoriedad. Todos los programas de computo fueron ejecutados en un equipo DIGITAL VAX-11/730.

RECONOCIMIENTOS

Los autores agradecen la valiosa colaboraci3n del Ingeniero en Sistemas Computacionales A. JAIME GARCIA BANDA del Instituto de Ciencias B3sicas de la U. V. por la elaboraci3n de los programas de c3mputo para las simulaciones del proceso de captura de discos.

BIBLIOGRAFIA

- D. Nestel (1989) Comunicaci3n personal sobre sus experimentos de captura de cebos por colonias de hormigas. Instituto de Ecologia, Xalapa, Ver.
- B M D P Statistical Vol.1 & Vol.2. University of California Software Manual (1988) Press. Berkeley.
- E.C. Pielou (1977) Mathematical Ecology. John Wiley and Sons. New York.
- C.S. Holling (1961) Principles of Insect Predation. Annual Review of Entomology 6 : 163-182.
- D. Rogers (1971) Random Search and Population Models. Journal Anim. Ecol. 40 : 369-382.

ANÁLISIS ESTADÍSTICO DE ALGUNAS CARACTERÍSTICAS MORFOLÓGICAS DE LA GAVIOTA PARDA *Larus heermanni*.

José R. Mendoza Blanco y Renata Villalba Cohen
Depto. de Matemáticas, Fac. de Ciencias. UNAM.
Circuito Exterior, Cd. Universitaria. 04510 México D. F.

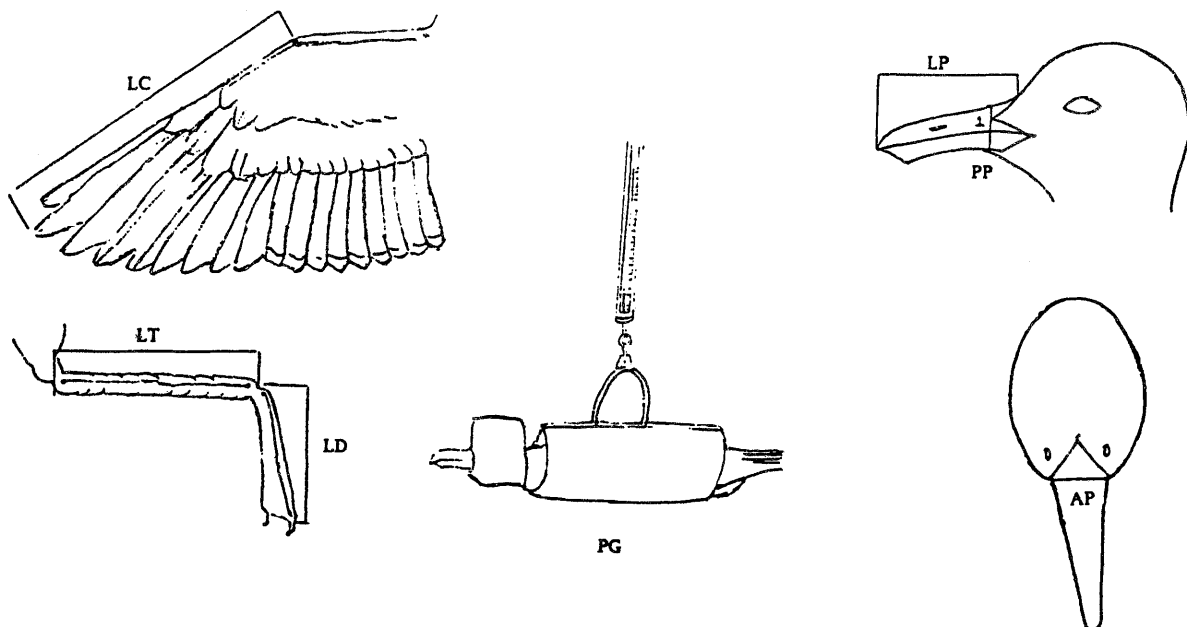
Resumen

En este trabajo se analiza un conjunto de datos correspondiente a registros de variables relacionadas con la morfología de gaviotas pardas *Larus Heermanni*. El objetivo es construir una regla de clasificación por sexo de las gaviotas, basada en las variables registradas. La presente investigación fué presentada al Laboratorio de Estadística de la Facultad de Ciencias de la UNAM por la Biol. Sachiko Chiriguchi. Los datos fueron colectados en Isla Rasa, Baja California Norte, México. Este trabajo fué elaborado utilizando una computadora personal IBM PS/50z, que IBM de México cedió en calidad de préstamo al Laboratorio de Estadística de la Facultad de Ciencias de la UNAM.

PALABRAS CLAVE: Análisis Multivariado, Clasificación, Análisis Discriminante.

1. Introducción

Con objeto de clasificar por sexo a las gaviotas, en este estudio se definieron las siguientes características morfológicas, las cuales se indican en la siguiente figura:



en donde las variables representan:

- LC Longitud de la cuerda en milímetros.
- LP Longitud del pico en milímetros.
- AP Ancho del pico en milímetros.
- PP Profundidad del pico en milímetros.
- LT Longitud del tarso en milímetros.
- LD Longitud del dedo en milímetros.
- PG Peso de la gaviota en gramos.

En este estudio se registraron las variables de interés en 181 gaviotas adultas, las cuales corresponden a gaviotas provenientes de 91 nidos. En 53 de estas 181 gaviotas, provenientes de 27 nidos, el sexo registrado al momento de hacer la captura fué corroborado mediante observación de la cópula, resultando que 26 gaviotas corresponden a hembras y 27 a machos. El motivo por el que se dispone de registros de un macho más que el número de hembras es que en una de las parejas la hembra no pudo ser capturada. Al comparar la clasificación preliminar del investigador con la clasificación obtenida mediante observación de la cópula, se obtuvo que en 51 casos la clasificación inicial coincidió con la obtenida por registro de la cópula; es decir, en sólo una pareja de gaviotas el registro del sexo al momento de la captura fué incorrecto. Los 128 registros restantes corresponden a gaviotas provenientes de 64 nidos, en las cuales no se tiene corroborado el sexo, pero se dispone de la clasificación preliminar del investigador.

Adicionalmente se cuenta con registros de las variables morfológicas de 220 pollos de gaviota. En 30 de estos individuos se registró el sexo mediante disección resultando 17 machos y 13 hembras, en 98 pollos no se cuenta con registro de sexo y en 92 pollos se dispone de un registro de sexo obtenido a simple vista por experiencia del investigador.

En base a la información recolectada, los objetivos de este estudio son:

1. Construir una regla de clasificación por sexo de las gaviotas adultas con las variables morfológicas.
2. A partir de la relación encontrada en el punto anterior, determinar el sexo de las gaviotas adultas en las que no se corroboró este registro por observación de la cópula y comparar este resultado con la clasificación preliminar del investigador.
3. Construir una regla de clasificación por sexo de los pollos con las variables morfológicas.

4. A partir de la relación obtenida en el punto anterior, determinar el posible sexo de los pollos para los que no se tiene este registro o éste no fué corroborado.

2. Planteamiento Estadístico.

La solución a los objetivos uno y tres descritos en la introducción puede plantearse estadísticamente construyendo reglas de clasificación mediante técnicas de Análisis Discriminante [1], [5]. Ante el problema de clasificación de gaviotas adultas, se puede proceder al menos de dos formas. La primera consiste en determinar el posible sexo de cada gaviota en forma individual, en base a las variables morfológicas. En este caso los dos grupos a discriminar son hembras y machos y las variables discriminantes corresponden a los registros morfológicos. En el proceso de clasificación de gaviotas en forma individual es posible combinar la opinión a priori del investigador sobre el sexo de la gaviota (plasmada en una distribución de probabilidades) con la evidencia que proporcionan los registros. Puede resultar que al clasificar dos individuos provenientes de un mismo nido en forma separada, se asignen a un mismo sexo; esto es, se clasifiquen como dos machos o dos hembras, situación que puede ocurrir aunque resulta muy poco probable. Para evitar este posible inconveniente, la segunda forma de proceder consiste en determinar el posible sexo de cada pareja en forma conjunta; es decir, dado que dos individuos provienen de un mismo nido, interesa asignarlos a uno de los siguientes grupos: (macho, hembra), (hembra, macho), (hembra, hembra) o (macho, macho). Evidentemente en el caso de los pollos de gaviota sólo resulta apropiada la clasificación individual. Utilizando los dos procedimientos de asignación anteriores se dió solución a los objetivos dos y cuatro con base en las reglas de asignación. La información de gaviotas adultas y de pollos se analiza por separado.

Dado que las variables se registran en una escala continua y resulta razonable suponer comportamientos unimodales y simétricos, se recurrió al empleo de la distribución normal para modelar la variabilidad de estas características. De esta manera y con objeto de utilizar reglas de discriminación de máxima verosimilitud, se recurrió al supuesto de que la variación conjunta de las variables morfológicas puede describirse mediante el modelo Normal Multivariado de dimensión siete; esto es, si se definen los vectores

$$X_{ij}^t = [LC_{ij}, LP_{ij}, AP_{ij}, PP_{ij}, LT_{ij}, LD_{ij}, PG_{ij}] \quad i=1,2; \quad j=1,2,\dots,n_1,$$

donde $i=1$ representa el grupo de machos e $i=2$ representa el grupo de hembras, se tiene que las suposiciones distribucionales son

$$X_{ij} \sim N_7(\mu_i, \Sigma_i) \quad i=1,2; \quad j=1,2,\dots,n_i$$

con μ_i y Σ_i el vector de medias y la matriz de covarianzas del grupo i respectivamente, n_1 el número de machos, n_2 el número de hembras y $n=n_1+n_2$ el número total de observaciones.

El Método de contraste de hipótesis utilizado en todas las pruebas en este trabajo es el de cociente de verosimilitudes generalizado [1], [5]. En todos los casos se reporta el nivel de significancia descriptivo (nsd). Las pruebas de hipótesis descritas en esta sección se basan en las siguientes funciones de los datos:

$$\begin{aligned} \bar{X}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad i=1,2, & \bar{X} &= \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{n_i} X_{ij}, \\ V_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^t \quad i=1,2, & T_x &= \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{n_i} (X_{ij} - \bar{X})(X_{ij} - \bar{X})^t, \\ W_x &= \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^t \quad y & B_x &= \frac{1}{n} \sum_{i=1}^2 n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^t. \end{aligned}$$

Para el análisis discriminante correspondiente a las gaviotas adultas se tiene $n_1=27$ y $n_2=26$ y para construir la regla de clasificación por sexo para los pollos de gaviota se tiene $n_1=17$ y $n_2=13$. En estos dos casos el procedimiento seguido en el análisis es el siguiente:

Con objeto de evaluar si puede producirse una regla de clasificación de utilidad, se contrastó el siguiente juego de hipótesis (prueba de igualdad de poblaciones):

$$\begin{aligned} H_a &: \mu_1 = \mu_2 \text{ y } \Sigma_1 = \Sigma_2. \\ \text{vs.} \\ H_a &: \text{no } H_a. \end{aligned}$$

En caso de no ser rechazada la hipótesis H_a , se tendría evidencia a favor de que no existen diferencias significativas en las poblaciones, por lo que en este caso la construcción de una regla de clasificación no brindaría utilidad práctica. La estadística de prueba asociada a H_a está definida por [2]:

$$P_a = n \ln |T_x| - \sum_{i=1}^2 n_i \ln |V_i|$$

y asintóticamente sigue una distribución χ^2 con 35 grados de libertad. La hipótesis H_a se rechaza a un nivel de significancia α si P_a resulta mayor que el cuantil de orden $1-\alpha$ de la distribución χ^2_{35} . En caso de rechazarse la hipótesis H_a se contaría con evidencia a favor de que existen diferencias en el comportamiento de las variables, por lo que se procedería, con objeto de simplificar el análisis, a contrastar el juego de hipótesis:

$$\begin{aligned} & H_b : \Sigma_1 = \Sigma_2, \\ \text{vs.} & H_{\bar{b}} : \text{no } H_b, \end{aligned} \tag{1}$$

donde si esta última hipótesis nula no es rechazada, se efectuaría el análisis suponiendo que

$$X_{ij} \sim N_j(\mu_j, \Sigma) \quad i=1,2; \quad j=1,2,\dots,n_1,$$

y en caso de ser rechazada H_b , la regla de clasificación se construiría suponiendo que

$$X_{ij} \sim N_j(\mu_j, \Sigma_j) \quad i=1,2; \quad j=1,2,\dots,n_1.$$

El interés por contrastar el juego de hipótesis (1) está asociado a simplificar el procedimiento de clasificación, ya que si las matrices de covarianzas de los dos grupos resultan estadísticamente iguales, la regla de discriminación se basa en la simple evaluación de combinaciones lineales de las variables involucradas. La estadística de prueba asociada a H_b está definida por [2]:

$$P_b = n \ln |W_x| - \sum_{i=1}^2 n_i \ln |V_i|$$

que tiene distribución asintótica χ^2 con 28 grados de libertad. La hipótesis H_b se rechaza a un nivel de significancia α si P_b resulta mayor que el cuantil de orden $1-\alpha$ de la distribución χ_{28}^2 .

Como siguiente paso en la construcción de la regla de clasificación y con objeto de sólo incluir en ésta información discriminante de utilidad práctica, se efectuó una selección de variables dentro del conjunto de mediciones morfológicas. Esta selección de variables se realizó considerando el poder discriminatorio de éstas, para lo cual se utilizó como criterio de selección minimizar la estadística de prueba de igualdad de medias obtenida vía cociente de verosimilitudes generalizado. Este procedimiento se conoce en la literatura como procedimiento de selección basado en la distribución Λ de Wilks [1], [3], y [4].

Una vez realizada la selección de variables, resulta conveniente definir los vectores

$$Y_{ij} \quad i=1,2; \quad j=1,2,\dots,n_1, \tag{2}$$

cuyas entradas corresponden a las variables en las que se desea basar la clasificación de gaviotas. Las reglas de asignación por sexo, considerando los vectores definidos en (2) se basan en las matrices

$$W_y = \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)(Y_{ij} - \bar{Y}_i)^t, \quad B_y = \frac{1}{n} \sum_{i=1}^2 n_i (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})^t$$

y

$$S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)(Y_{ij} - \bar{Y}_i)^t; \quad i=1,2,$$

de donde

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad i=1,2, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{n_i} Y_{ij}.$$

Con objeto de simplificar la notación, la densidad de una variable Y con distribución normal multivariada de media μ y matriz de covarianzas Σ se denotará por

$$N(Y; \mu, \Sigma) = |2\pi\Sigma|^{-\frac{k}{2}} \exp\left\{-\frac{1}{2}(Y-\mu)^t \Sigma^{-1}(Y-\mu)\right\}.$$

Los procedimientos de clasificación de gaviotas en forma individual y por parejas se describen a continuación.

2.1. Clasificación de Gaviotas en Forma Individual.

En este caso, las poblaciones a discriminar corresponden a machos ($i=1$) y hembras ($i=2$). Sea Π_i el evento de que un individuo con vector de atributos Y , seleccionado en forma aleatoria de la población total de gaviotas, pertenezca a la i -ésima población. Utilizando el teorema de Bayes se obtiene que

$$P(\Pi_i|Y) = \frac{P(Y|\Pi_i)P(\Pi_i)}{\sum_{i=1}^2 P(Y|\Pi_i)P(\Pi_i)} \quad i=1,2,$$

donde $P(\Pi_i)$ representa la probabilidad a priori del investigador sobre Π_i y $P(Y|\Pi_i)$ es la densidad de Y dado Π_i . Dado que $P(Y|\Pi_i)$ es una densidad Normal con parámetros desconocidos, se utiliza el principio de invarianza para construir $\hat{P}(\Pi_i|Y)$, el estimador máximo verosímil de $P(\Pi_i|Y)$.

La regla de clasificación consiste en asignar la gaviota con vector de atributos Y a la población para la cual $P(\Pi_i|Y)$ es mayor. En el caso en que las matrices de covarianzas de las dos poblaciones sean distintas se tiene que $P(Y|\Pi_i) = N(Y; \mu_i, \Sigma_i) \quad i=1,2$. En este caso, el procedimiento de asignación consiste en clasificar al individuo en la población para la cual

$$D_i(Y) = -\frac{1}{2} \ln |S_i| - \frac{1}{2} (Y - \bar{Y}_i)^t S_i^{-1} (Y - \bar{Y}_i) + \ln P(\Pi_i) \quad i=1,2 \quad (3)$$

es mayor y los estimadores de máxima verosimilitud de $P(\Pi_i|Y)$ están dados por

$$\hat{P}(\Pi_i|Y) = \frac{\exp\{D_i(Y)\}}{\sum_{i=1}^2 \exp\{D_i(Y)\}} \quad i=1,2.$$

En el caso en que $\Sigma_1 = \Sigma_2 = \Sigma$, se tiene que $P(Y|\Pi_i) = N(Y; \mu_i, \Sigma)$ $i=1,2$. El procedimiento de asignación consiste en clasificar al individuo en la población para la cual

$$C_i(Y) = \bar{Y}_i^t W_y^{-1} Y - \frac{1}{2} \bar{Y}_i^t W_y^{-1} \bar{Y}_i + \ln P(\Pi_i) \quad (4)$$

resulta mayor. Las estimaciones máximo verosímiles de $P(\Pi_i|Y)$ están dadas por

$$\hat{P}(\Pi_i|Y) = \frac{\exp\{C_i(Y)\}}{\sum_{i=1}^2 \exp\{C_i(Y)\}} \quad i=1,2.$$

2.2. Clasificación de Gaviotas por Parejas.

En este caso el objetivo es clasificar en forma conjunta a una pareja de gaviotas. Dada una pareja, sea $Y^t = (Y_1^t, Y_2^t)$ donde Y_1 y Y_2 son los vectores de atributos de los individuos sin importar a qué gaviota de la pareja se le denote por Y_i .

Dado un vector Y extraído en forma aleatoria de la población total de parejas, existen cuatro posibles clasificaciones para los individuos de la pareja:

Π_1 : El primer individuo es macho y el segundo es hembra.

Π_2 : El primer individuo es hembra y el segundo es macho.

Π_3 : Los dos individuos son hembras.

Π_4 : Los dos individuos son machos.

Utilizando el teorema de Bayes se obtiene que

$$P(\Pi_i|Y) = \frac{P(Y|\Pi_i)P(\Pi_i)}{\sum_{i=1}^4 P(Y|\Pi_i)P(\Pi_i)} \quad i=1,2,3,4,$$

donde de nuevo, $P(\Pi_i)$ representa la probabilidad a priori del investigador sobre la posible clasificación de la pareja y $P(Y|\Pi_i)$ denota la densidad del vector Y cuando éste proviene de la población Π_i .

Considerando que $P(\Pi_i|Y)$ depende de parámetros desconocidos, se utiliza el principio de invarianza para construir $\hat{P}(\Pi_i|Y)$, el estimador máximo verosímil de $P(\Pi_i|Y)$. La regla de clasificación consiste en asignar la pareja de gaviotas con vector Y a la población i para la cual $\hat{P}(\Pi_i|Y)$ resulta mayor ($i=1,2,3,4$).

En el caso en que las matrices de covarianzas de machos y hembras se consideren diferentes, se tiene que

$$P(Y|\Pi_1) = N(Y_1; \mu_1, \Sigma_1) N(Y_2; \mu_2, \Sigma_2),$$

$$P(Y|\Pi_2) = N(Y_1; \mu_2, \Sigma_2) N(Y_2; \mu_1, \Sigma_1),$$

$$P(Y|\Pi_3) = N(Y_1; \mu_2, \Sigma_2) N(Y_2; \mu_2, \Sigma_2)$$

y

$$P(Y|\Pi_4) = N(Y_1; \mu_1, \Sigma_1) N(Y_2; \mu_1, \Sigma_1).$$

Definiendo

$$D_1(Y) = -\frac{1}{2} \ln |S_1 S_2| - \frac{1}{2} (Y_1 - \bar{Y}_1)^t S_1^{-1} (Y_1 - \bar{Y}_1) - \frac{1}{2} (Y_2 - \bar{Y}_2)^t S_2^{-1} (Y_2 - \bar{Y}_2) + \ln P(\Pi_1),$$

$$D_2(Y) = -\frac{1}{2} \ln |S_1 S_2| - \frac{1}{2} (Y_1 - \bar{Y}_2)^t S_2^{-1} (Y_1 - \bar{Y}_2) - \frac{1}{2} (Y_2 - \bar{Y}_1)^t S_1^{-1} (Y_2 - \bar{Y}_1) + \ln P(\Pi_2),$$

(5)

$$D_3(Y) = -\ln |S_2| - \frac{1}{2} (Y_1 - \bar{Y}_2)^t S_2^{-1} (Y_1 - \bar{Y}_2) - \frac{1}{2} (Y_2 - \bar{Y}_2)^t S_2^{-1} (Y_2 - \bar{Y}_2) + \ln P(\Pi_3)$$

y

$$D_4(Y) = -\ln |S_1| - \frac{1}{2} (Y_1 - \bar{Y}_1)^t S_1^{-1} (Y_1 - \bar{Y}_1) - \frac{1}{2} (Y_2 - \bar{Y}_1)^t S_1^{-1} (Y_2 - \bar{Y}_1) + \ln P(\Pi_4),$$

la regla de asignación consiste en clasificar la pareja a la población i para la cual $D_i(Y)$ resulta mayor ($i=1,2,3,4$).

Los estimadores máximos verosímiles de $P(\Pi_i|Y)$ se obtienen como:

$$\hat{P}(\Pi_i|Y) = \frac{\exp\{D_i(Y)\}}{\sum_{i=1}^4 \exp\{D_i(Y)\}} \quad i=1,2,3,4.$$

En el caso en que se considere una matriz de covarianzas común Σ para hembras y machos, se tiene que:

$$P(Y|\Pi_1) = N(Y_1; \mu_1, \Sigma) N(Y_2; \mu_2, \Sigma),$$

$$P(Y|\Pi_2) = N(Y_1; \mu_2, \Sigma) N(Y_2; \mu_1, \Sigma),$$

$$P(Y|\Pi_3) = N(Y_1; \mu_2, \Sigma) N(Y_2; \mu_2, \Sigma)$$

y

$$P(Y|\Pi_4) = N(Y_1; \mu_1, \Sigma) N(Y_2; \mu_1, \Sigma),$$

de donde se deduce que una pareja con vectores de atributos Y se asigna a la población i en la cual $C_i(Y)$ resulta mayor, con

$$C_1(Y) = \bar{Y}_1^t W_y^{-1} Y_1 + \bar{Y}_2^t W_y^{-1} Y_2 - \frac{1}{2} (\bar{Y}_1^t W_y^{-1} \bar{Y}_1 + \bar{Y}_2^t W_y^{-1} \bar{Y}_2) + \ln P(\Pi_1),$$

$$C_2(Y) = \bar{Y}_1^t W_y^{-1} Y_2 + \bar{Y}_2^t W_y^{-1} Y_1 - \frac{1}{2} (\bar{Y}_1^t W_y^{-1} \bar{Y}_1 + \bar{Y}_2^t W_y^{-1} \bar{Y}_2) + \ln P(\Pi_2),$$

$$C_3(Y) = \bar{Y}_2^t W_y^{-1} (Y_1 + Y_2) - \bar{Y}_2^t W_y^{-1} \bar{Y}_2 + \ln P(\Pi_3)$$

y

$$C_4(Y) = \bar{Y}_1^t W_y^{-1} (Y_1 + Y_2) - \bar{Y}_1^t W_y^{-1} \bar{Y}_1 + \ln P(\Pi_4).$$

Los estimadores máximo verosímiles de $P(\Pi_i|Y)$ se obtienen como

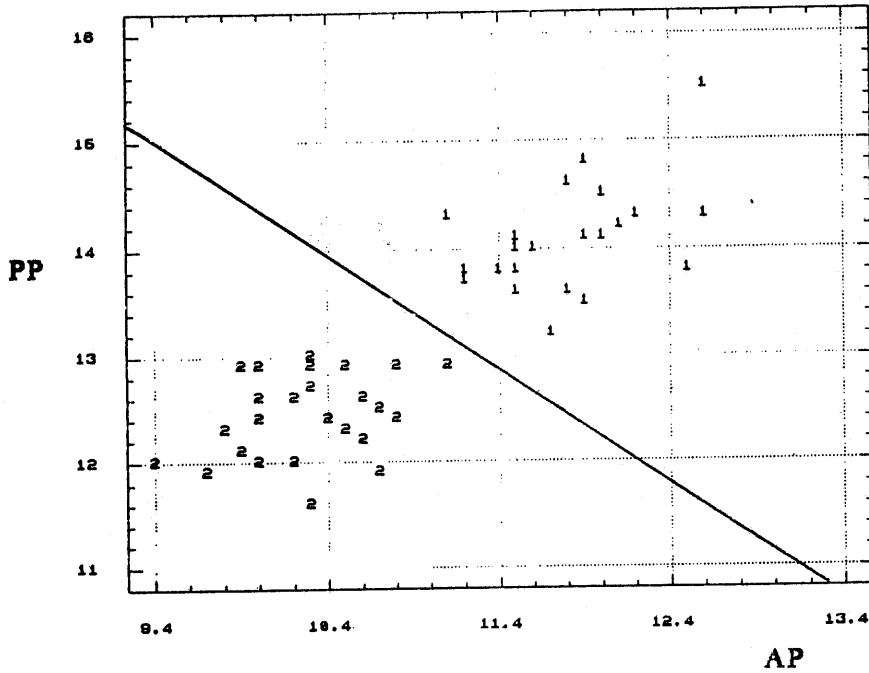
$$\hat{P}(\Pi_i|Y) = \frac{\exp\{C_i(Y)\}}{\sum_{i=1}^4 \exp\{C_i(Y)\}} \quad i=1,2,3,4.$$

3. Análisis Estadístico para Gaviotas Adultas.

La estadística de prueba de igualdad de poblaciones resultó 150.93 con distribución aproximada χ^2 con 35 grados de libertad y nsd menor que 0.0005. Este hecho indica que tiene sentido construir una regla de clasificación por sexo de los individuos en base a las variables consideradas. La estadística de prueba de igualdad de matrices de covarianzas es 34.78 con distribución aproximada χ^2 con 28 grados de libertad y nsd 0.18, lo cual muestra una fuerte evidencia a favor de que puede establecerse una regla de clasificación lineal.

Como resultado de la selección de variables, se obtuvo que las variables significativas en la discriminación son LP, AP, PP y LD. La estadística de prueba de igualdad de medias en las dos poblaciones, considerando las cuatro variables seleccionadas, fué 93.65 y sigue una distribución $F_{(4,48)}$ con un nsd menor que 0.0005, lo cual indica que existen fuertes diferencias entre las dos poblaciones. Utilizando estas variables, pueden construirse las funciones de clasificación individual y por pareja y puede verificarse que en ambos casos, considerando distribuciones a priori uniformes, los porcentajes de clasificación global correcta son del 100%.

De la gráfica 1 se observa que las variables AP y PP producen una separación de las poblaciones que puede ser utilizada para clasificar a las gaviotas en forma individual; esto es, la gráfica sugiere que una regla de clasificación basada en sólo estas dos variables puede brindar resultados satisfactorios. La estadística de prueba de igualdad de medias por sexo en estas dos variables es 137.63 y tiene distribución asociada $F_{(2,50)}$ con nsd menor de 0.0005, por lo que se infiere que existen diferencias significativas por sexo en estas dos variables.



GRAFICA 1. Datos de gaviotas adultas con sexo verificado y la recta de asignación.

Al construir el procedimiento de asignación en forma individual basado en las variables AP y PP (utilizando una distribución a priori uniforme) se obtuvo, como era evidente, que el porcentaje de clasificación global correcta es del 100%. De esta manera resulta preferible para clasificar gaviotas utilizar sólo dos variables en comparación a las cuatro que se obtuvieron mediante el procedimiento de selección. Definiendo el vector $Y^t=(AP,PP)$, las funciones de clasificación en forma individual, utilizando una distribución a priori uniforme, corresponden a

$$C_1(Y)=53.012(AP)+61.847(PP)-747.800$$

para el grupo de machos y

$$C_2(Y)=45.656(AP)+55.155(PP)-578.037$$

para el grupo de hembras. Utilizando estas funciones, un individuo se asigna a la población que tiene asociado el mayor puntaje de clasificación. Para una mayor simplicidad en la clasificación de los individuos, puede definirse la recta de asignación $A(Y)=C_1(Y)-C_2(Y)$, la cual en este caso resulta

$$A(Y)=7.536(AP)+6.692(PP)-169.763. \quad (7)$$

Al evaluar esta función se tiene que si $A(Y) \geq 0$ el individuo se asigna a la población de machos y si $A(Y) < 0$ a la población de hembras. La recta de asignación (7) se exhibe en la gráfica 1, donde las observaciones que aparecen por encima de la recta se clasifican como machos y como hembras en caso contrario.

Definiendo $Y_i^t = (AP_i, PP_i)$ $i=1,2$, las funciones de clasificación por pareja, considerando una distribución a priori uniforme, están dadas por

$$C_1(Y) = 53.012(AP_1) + 61.847(PP_1) + 45.656(AP_2) + 55.155(PP_2) - 1325.144,$$

$$C_2(Y) = 45.656(AP_1) + 55.155(PP_1) + 53.012(AP_2) + 61.847(PP_2) - 1325.144,$$

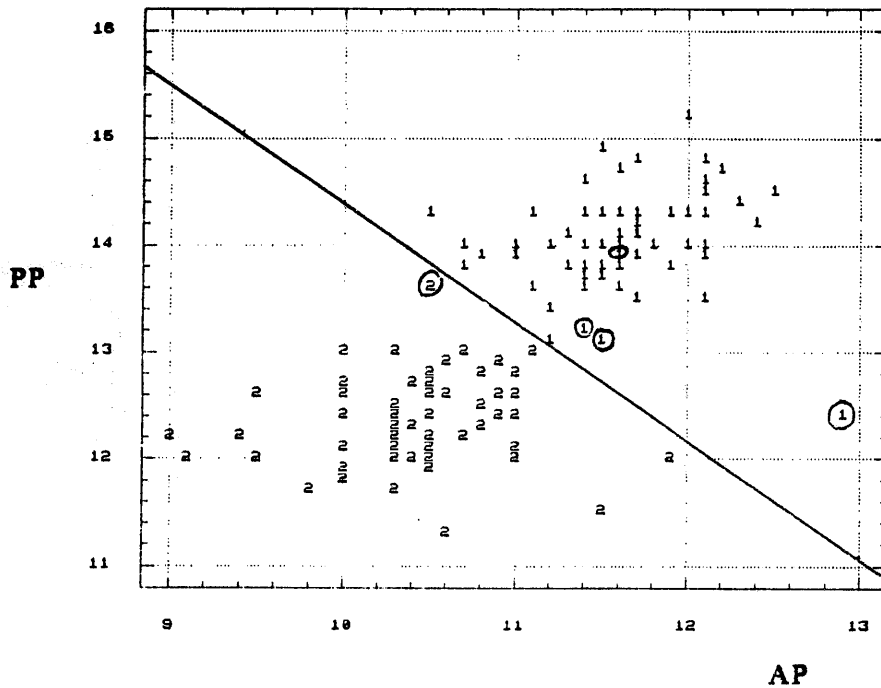
$$C_3(Y) = 45.656(AP_1 + AP_2) + 55.155(PP_1 + PP_2) - 1155.381$$

y

$$C_4(Y) = 53.012(AP_1 + AP_2) + 61.847(PP_1 + PP_2) - 1494.907.$$

Una pareja de gaviotas se clasifica en la i -ésima población si $C_i(Y) \geq C_j(Y)$, $j=1,2,3,4$. Al clasificar a las 26 parejas de gaviotas en las que se verificó el registro del sexo, utilizando las funciones anteriores, se obtuvieron los mismos resultados que en la clasificación individual.

Para responder al objetivo dos se clasificaron en forma individual las 128 gaviotas en las que no se corroboró el registro del sexo, utilizando la recta de asignación (7). De las 64 gaviotas clasificadas inicialmente como machos por el investigador, en sólo un caso la asignación basada en la recta de asignación resultó diferente. De las 64 gaviotas clasificadas inicialmente como hembras por el investigador, en 4 casos esta clasificación difiere de la obtenida mediante la recta de asignación. Las diferencias en clasificación corresponden a que en una de las parejas clasificada inicialmente por el investigador como macho y hembra, la clasificación mediante la recta de asignación los clasificó como hembra y macho. Adicionalmente en tres individuos clasificados inicialmente como hembras por el investigador, la clasificación mediante la recta de asignación resultó en tres machos; esto es, se obtuvo que en tres nidos, las dos gaviotas resultaron machos. En la gráfica 2 se exhibe la asignación basada en la recta (7), donde las observaciones clasificadas como machos se señalan por '1' y las clasificadas como hembras por '2'. Los datos en los que la clasificación mediante la recta de asignación difiere de la del investigador se encuentran marcados con un círculo.



GRAFICA 2. Datos de gaviotas adultas sin verificación del sexo y la recta de asignación.

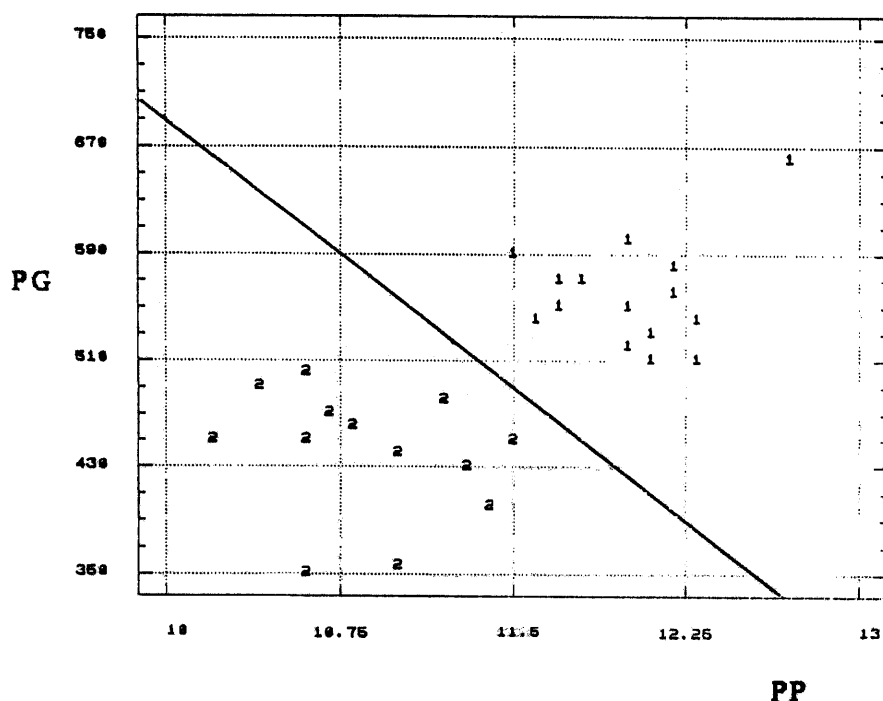
La clasificación en parejas de estos individuos, utilizando una distribución a priori uniforme, produjo los mismos resultados que la clasificación en forma individual.

4. Análisis Estadístico para Pollos de Gaviota.

La prueba de hipótesis de igualdad de poblaciones arrojó una estadística de prueba de 101.77 con distribución aproximada χ^2 con 35 grados de libertad y nsd menor que 0.0005, lo cual indica que existen diferencias significativas entre las poblaciones por lo que tiene sentido construir una regla de clasificación. En la prueba de igualdad de matrices de covarianzas, la estadística de prueba resultó ser 39.75 con distribución aproximada χ^2 con 28 grados de libertad y nsd de 0.073, por lo que los datos sugieren construir una regla de clasificación lineal.

Como resultado de la selección de variables basada en la distribución Λ de Wilks, se obtuvo que el vector de variables seleccionadas es $Y^t=(PP,PG)$. En la gráfica 3 se presenta la distribución de las observaciones en las variables PP y PG, observándose que las poblaciones de machos y hembras pueden separarse perfectamente por una recta, de donde se sigue que puede construirse un procedimiento de clasificación lineal basado en estas dos variables que clasifique a estas gaviotas sin error. La estadística de prueba de igualdad de medias, considerando estas dos variables, es 64.54 con distribución $F_{(2,27)}$

y nsd menor que 0.005, lo que significa que existen fuertes diferencias entre los vectores de medias de las dos poblaciones.



GRAFICA 3. Datos de pollos con sexo verificado y la recta de asignación.

Las funciones de clasificación en este caso están dadas por

$$C_1(Y) = 103.694(PP) + 0.357(PG) - 721.616$$

y

$$C_2(Y) = 93.958(PP) + 0.283(PG) - 573.691.$$

Utilizando una distribución inicial uniforme sobre el sexo, la recta de asignación en las dos poblaciones, dada por $A(Y) = C_1(Y) - C_2(Y)$, resulta en este caso

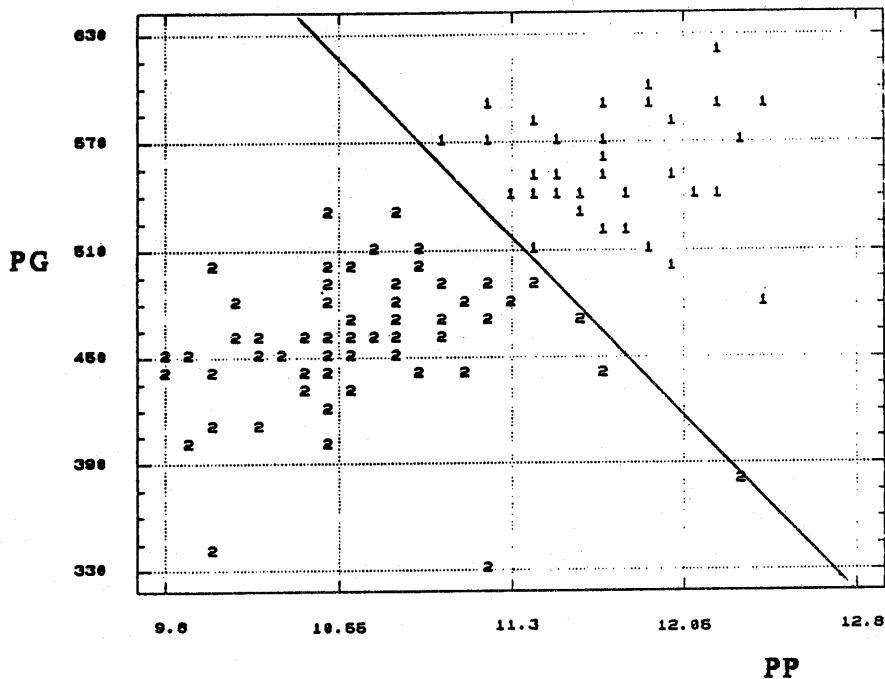
$$A(Y) = 9.736(PP) + 0.074(PG) - 147.925 \quad (8)$$

y un individuo se clasifica como macho si $A(Y) \geq 0$ y como hembra en caso contrario.

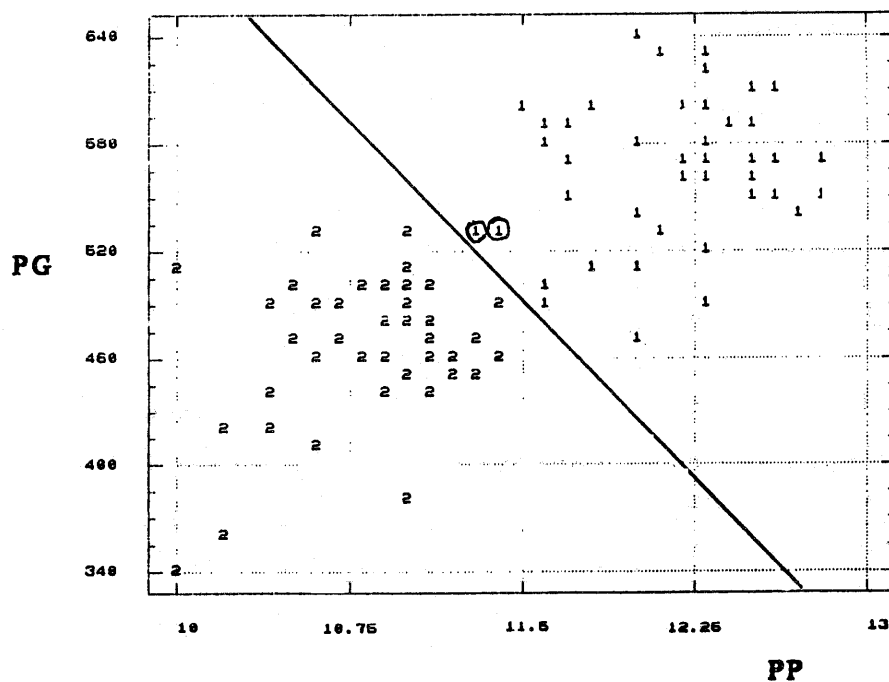
Como resultado de la aplicación de recta de asignación (8) a los datos de pollos verificados se obtuvo un porcentaje de clasificación global correcta del 100%. La recta de asignación (8) se presenta en la gráfica 3, donde si un punto se localiza por encima de la recta se clasifica como macho y en caso contrario como hembra.

Utilizando la recta de asignación (8), los resultados de la clasificación de los 98 pollos no sexados se presentan en la gráfica 4 y en la gráfica 5 se exhibe la clasificación de los 92

pollos en los que no se verificó el registro del sexo. Los casos en los que la clasificación mediante la recta de asignación difiere de la clasificación inicial del investigador se encuentran marcados con un círculo.



GRAFICA 4. Datos de pollos sin registro del sexo-y la recta de asignación.



GRAFICA 5. Datos de pollos sin verificación del sexo y la recta de asignación.

Se observa de la gráfica 5 que en sólo dos casos clasificados como hembras por el investigador, la clasificación basada en la recta de asignación resultó diferente.

Conclusiones

En el caso de gaviotas adultas existen varias características morfológicas que presentan diferencias significativas por sexo. Considerando que uno de los propósitos de este trabajo es construir una regla de asignación por sexo para gaviotas adultas, resulta conveniente construir reglas de asignación que utilizando la menor cantidad de información posible brinden resultados satisfactorios. En este sentido se encontró que resulta suficiente considerar las variables ancho del pico (AP) y profundidad del pico (PP) para establecer una regla de diferenciación de las gaviotas adultas por sexo. El procedimiento de clasificación basado en la recta de asignación, utilizando una distribución a priori uniforme, produjo resultados altamente satisfactorios, ya que la clasificación de las gaviotas adultas verificadas basada en esta recta coincidió con la clasificación verificada mediante observación de la cópula y las probabilidades con que se clasificaron las gaviotas a los grupos en que fueron asignados son en general muy próximas a uno. Dado que en este caso las dos variables discriminantes corresponden a mediciones del pico se sigue que el tamaño de éste constituye una fuente de diferenciación sexual en gaviotas adultas.

En relación a la clasificación por parejas de las gaviotas en que se verificó el registro del sexo (considerando una distribución inicial uniforme sobre las poblaciones), los resultados obtenidos coincidieron con los correspondientes a la clasificación en forma individual. Este resultado se explica por una parte por el hecho de que las probabilidades de pertenencia a la población a que fueron asignados son en general muy cercanas a uno y por otra, por el hecho de haber considerado distribuciones iniciales uniformes sobre el posible sexo de las gaviotas.

La clasificación de las gaviotas adultas no verificadas basada en la recta de asignación, utilizando una distribución a priori uniforme, produjo resultados que podrían considerarse confiables, ya que por un lado la aplicación de la regla a los adultos verificados no condujo a clasificaciones incorrectas y por otro de la gráfica 2 se observa que existen dos grupos en los datos, mismos que la regla asignó como correspondientes a sexos distintos. De nuevo, la clasificación por parejas produjo los mismos resultados, al utilizar una distribución a priori uniforme.

Para los pollos de gaviotas existen, al igual que en el caso de gaviotas adultas, varios atributos morfológicos que presentan diferencias por sexo. En este caso se encontró que en la diferenciación por sexo de los pollos de gaviota resulta suficiente considerar las variables profundidad del pico (PP) y peso de la gaviota (PG). La calidad de la regla de discriminación resultó altamente satisfactoria, ya que al aplicar la regla a los pollos verificados no se obtuvo ningún error de clasificación; es decir, la clasificación por sexo de los pollos verificados basada en la recta de asignación coincidió con la clasificación verificada mediante disección de los mismos.

En relación a la clasificación de los pollos no sexados y los no verificados, puede decirse que de nuevo la asignación puede juzgarse bastante confiable, ya que por una parte la regla de clasificación se consideró de buena calidad y de las gráficas 5 y 6 se observa en ambos casos que la separación de los dos grupos de observaciones resulta razonable. De hecho, la clasificación del investigador para los 92 pollos con registro preliminar del sexo resultó prácticamente igual a la obtenida mediante la recta de asignación, dado que en sólo dos casos las clasificaciones no coincidieron.

Como último punto cabe señalar que aunque en este trabajo se utilizaron distribuciones iniciales uniformes sobre el posible sexo de las gaviotas, la técnica mostrada permite utilizar cualquier distribución inicial, como se muestra en las expresiones (3) a la (6). En este sentido se podría considerar, por ejemplo, el hecho de que dos gaviotas de un mismo nido tengan más posibilidad de tener sexo diferente que de ser de un mismo sexo.

BIBLIOGRAFIA

1. Mardia K.V., Kent J.T. and Bibby J.M. (1979). *Multivariate Analysis*. Academic Press, London.
2. Mendoza J.R. (1987). *La Distribución Normal Multivariada y su Relación con Otras Distribuciones*. Tesis de Licenciatura en Actuaría, Facultad de Ciencias, UNAM.. México.
3. Nie N. H., Hull C. H., Jenkins J. G., Steinbrenner K. and Bent D. H. (1975). *Statistical Package for the Social Sciences, Second Edition*. McGraw-Hill, New York.
4. Norusis M. J. (1979). *SPSS: Statistical Algorithms*. SPSS Inc., Illinois.
5. Seber G.A.F. (1984). *Multivariate Observations*. Wiley, New York.

CONCEPTOS PARALELOS EN LAS AREAS DE DISEÑOS
EXPERIMENTALES Y MUESTREO

M. M. Ojeda
C. H. Cortina-Carmona
LINA E, Fac. de Estadística U. V.
Av. Xalapa esq. A. Camacho
Xalapa Ver.

RESUMEN: En éste artículo se presenta una revisión de una serie de trabajos que identifican y discuten los conceptos paralelos en las áreas de diseños experimentales y muestreo. Se hace especial énfasis en lo que a la tendencia unificadora de inferencia en éstas áreas se refiere.

PALABRAS CLAVE: Muestreo, Diseños Experimentales, Inferencia Analítica, Inferencia Causal, Conceptos Paralelos.

1. INTRODUCCION

Dos de las principales áreas de la Estadística son el Muestreo Estadístico (ME) y los Diseños Experimentales (DE), ambas consideran procedimientos de obtención y de análisis de datos. El ME se desarrolló siguiendo la línea teórica planteada por Neyman (1934) mientras los DE se desarrollaron siguiendo los lineamientos de los trabajos de Fisher (1935) y colaboradores. En el área de ME Godambe (1955) planteó los cimientos de una teoría unificada que incorporara al tratamiento de los problemas típicos, los conceptos generales de la Inferencia Estadística, como verisimilitud, mínima varianza, etc. Esto y el advenimiento del enfoque superpoblacional (Smith (1984)), incorporando la posibilidad de realizar inferencia analítica, han incrementado el interés sobre el estudio comparativo de conceptos en ME y DE. Existen trabajos recientes que presentan magníficas revisiones de los paralelos existentes, por ejemplo Tanur y Fienberg (1983, 1985, 1987, 1988), Rao (1985), Smith y Sugden (1988).

En éste trabajo se presentará una revisión de la definición de conceptos centrales en DE y ME, estableciendo los paralelos. Se pretende divulgar los principales avances de investigación en ésta

interesante dirección, que incluye aspectos teóricos y filosóficos, y que plantean grandes posibilidades de tener una mejor visión de la práctica de la Estadística en éstas áreas.

El trabajo está estructurado de la siguiente forma: en la sección 2 se presenta una revisión de conceptos paralelos siguiendo una propuesta de Rao (1985), incorporada a una clasificación de Smith y Sugden (1988), lo que se aprovecha para introducir el tema; la tercera sección se dedica a revisar los paralelos identificados más recientemente, y una cuarta sección plantea la revisión de las posibilidades identificadas para unificar criterios de inferencia en éstas áreas.

2. MUESTREO Y EXPERIMENTOS: CONCEPTOS PARALELOS Y DIVERGENCIAS

Consideremos la situación en la que contamos con un conjunto de N unidades $U = \{U_1, \dots, U_N\}$ las cuales denominaremos nuestra población finita. Cada una de éstas unidades puede recibir o recibió un tratamiento, que definimos como un estímulo o conjunto de estímulos que impactan la respuesta u observable, incluso cuando no se tenga control sobre la asignación. Esta definición de tratamiento nos permite considerar, variables como sexo, edad, nivel educativo, etc., para incluirlos como "tratamientos en un estudio".

En la práctica ésta situación descrita se presenta en múltiples investigaciones, y el investigador puede manipular la aleatoriedad en dos maneras: (1) al seleccionar las unidades que conformarán la muestra, un subconjunto $s \subset U$ de tamaño n , y (2) al asignar los tratamientos a las unidades en la muestra. La tabla 1, adaptada de Smith y Sugden (1988) pag. 167, presenta una clasificación general de los estudios en las que éstas dos maneras de manejar la aleatoriedad se combinan.

		Asignación de tratamientos	
		Control	No control
Selección de la muestra	Control	Experimentos dentro de muestreo o muestreo dentro de experimentos	Muestreo analítico
	No control	Experimento	Estudio observacional no controlado

TABLA 1. Clasificación de estudios con tratamientos.

En el área de muestreo se estudian fundamentalmente procesos de inferencia descriptiva, es decir cuando interesa hacer inferencia para totales, varianzas, proporciones, razones, etc., pensando que el objetivo es la población finita. En éste enfoque, denominado clásico [Smith (1984)], los tratamientos no se introducen con el objetivo de estudiar relaciones causales [Smith y Sugden (1988)], Holland (1986), Rubin (1978)] y por tal motivo no se puede hablar de inferencias en analogía a lo que es posible hacer en diseños experimentales, donde la población finita es sólo un conjunto de unidades de referencia y la población objetivo es una población "aparente" o un modelo que describe la relación causal.

La inferencia descriptiva basada en el muestreo por lo general, se basa en la distribución originada por las aplicaciones repetidas de los mecanismos de selección de muestra. Cuando el muestreo es aleatorio y las probabilidades de selección conocidas obtenemos la distribución por la aleatorización, la cual es más apropiada para decisiones acerca de la selección del diseño antes de los datos, y sus limitaciones en inferencia condicional después de los datos e inferencia predictiva, fué discutida por Smith (1976, 1983), Royall (1976). Mientras que la inferencia de experimentos regularmente se basa en un modelo estadístico, por lo general un modelo lineal, aunque algunas veces son empleadas las pruebas de aleatorización en lugar de los procedimientos basados

en el modelo. La idea debe ser que la inferencia experimental se fundamenta en el modelo y la inferencia en muestreo clásico en la aleatorización.

El desarrollo de las dos áreas, como dos cuerpos de conocimientos distintos tiene varias explicaciones. La primera es que las dos ideas fundamentales, una en cada área, surgen como dos concepciones a dos problemas aparentemente distintos. El desarrollo posterior, liderado por grandes estadísticos, incluso con rivalidades intelectuales declaradas (recuérdese la polémica de Fisher y Neyman), planteó una tradición de poca comunicación a pesar de que la práctica lo requiriera, según lo señalan Stephan (1948) y recientemente Fienberg y Tanur (1985), retomando un discurso del primero en el I.S.I., en el cual decía:

"...Los desarrollos en agricultura e ingeniería tienen efectos directos e indirectos en las prácticas de inspección de muestreo, ya que proveen los principios del diseño y contribuyen al incremento de la Estadística Matemática aplicada. Sin embargo hay muchos problemas prácticos y obstáculos que retrasan la extensión inmediata de los métodos desarrollados para pruebas de campo e inspección en la fabricación a gran escala. Uno de los obstáculos fue la relativa falta de comunicación entre los Estadísticos en los diferentes tipos de investigación y aplicación" (p. 30).

Sin embargo a pesar de las diferencias centrales en la práctica del muestreo y el diseño de experimentos se han identificado una serie de conceptos, que se han denominado conceptos similares [Rao (1985)] o conceptos paralelos [Fienberg y Tanur (1985)]. La figura 1 adaptada de Rao (1985), presenta un resumen de éstas analogías.

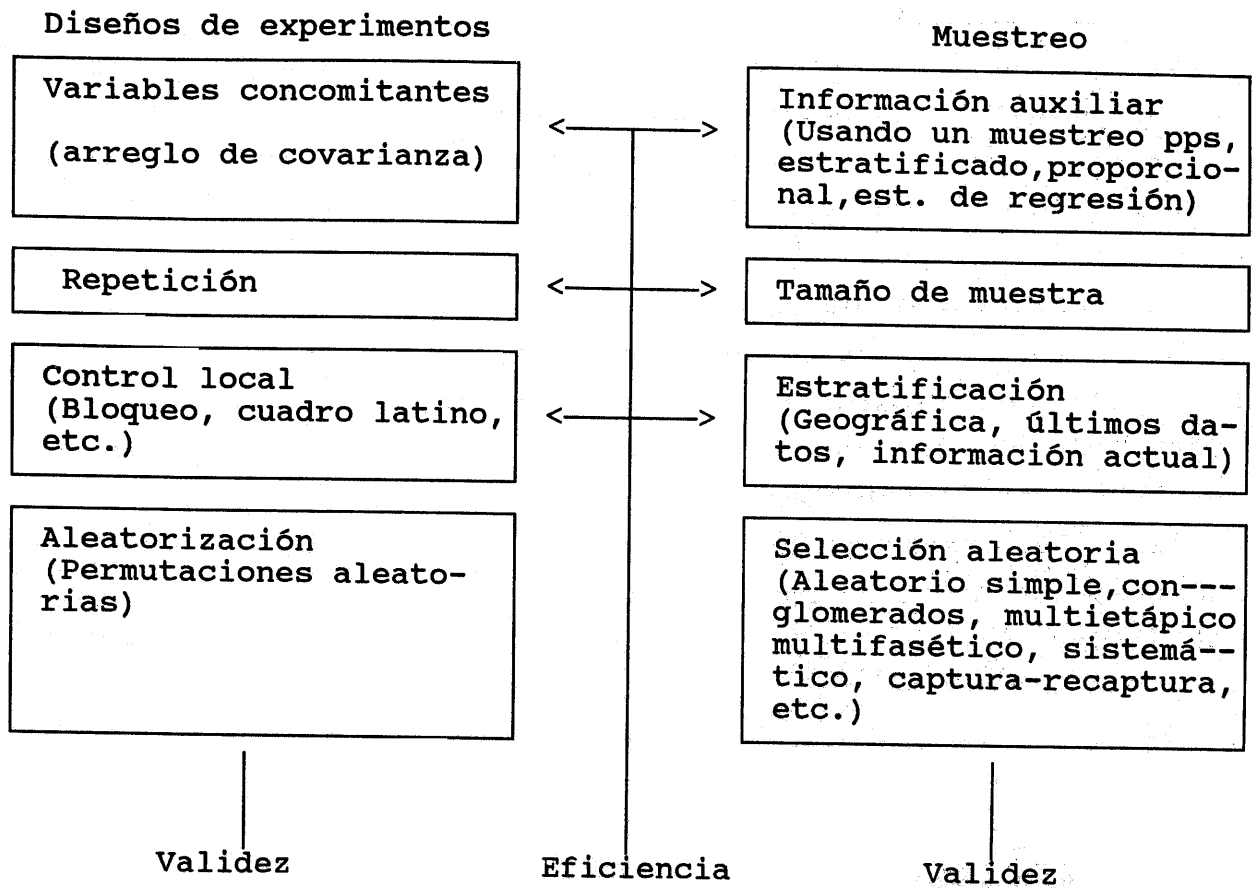


FIGURA 1. Principio de diseño de experimentos y muestreo (conceptos semejantes).

Los conceptos y estructuras paralelas son mostrados muy claramente, en los experimentos simples de dos tratamientos o dos grupos y en su estructura paralela, el muestreo aleatorio simple.

Considere una población finita U y una función de selección de muestra $A = \{A_1, \dots, A_n\}$ donde $A_i = 1$ si $i \in T_1$ y $A_i = 0$ si $i \in T_2$

Mediante esta función podemos formalizar el esquema de elección de una muestra. En un experimento de dos tratamientos, la función de selección de muestra A , indica a cuales miembros de la población se les asigna el tratamiento 1 (T_1) y a cuales el tratamiento 2 (T_2). En el caso de muestreo asigna a T_1 los miembros de la población que son seleccionados para la muestra y asigna en T_2 los que no fueron seleccionados.

En el contexto de muestreo, la selección aleatoria nos permite hacer la generalización para el resto de la población, y en el contexto experimental, a través de la aleatorización "retenemos todas las constantes" y así podemos atribuir cualquier efecto a la diferencia de tratamientos.

La utilización de grupos homogéneos es común, en diseños experimentales son utilizados para minimizar el error experimental por medio del recurso de bloqueo [Cochran y Cox (1957)] y cada repetición es realizada sobre un grupo homogéneo. En paralelo, en muestreo son utilizados para minimizar el error de muestreo mediante el recurso de estratificación [Cochran (1977)], con muestras obtenidas de cada grupo homogéneo en que fué dividida la población.

En el diseño experimental la herramienta de cuadro latino y grecolatino [Fisher (1935), Cochran y Cox (1957)] es útil para minimizar el error experimental cuando se controlan dos o más fuentes de variación: en paralelo, cuando en muestreo hay dos o más dimensiones de estratificación seleccionar una muestra de cada uno de los cruces de clasificación es muy difícil, la aplicación de cuadro latino o grecolatino produce un método de selección de estratos llamado muestreo de lattice [(Cochran (1977))].

El diseño de parcelas divididas [Cochran y Cox (1957)] es repetido en las técnicas de muestreo análogas, muestreo por conglomerados [Hansen, et. al. (1953b), Cochran (1977)].

En la fase de análisis, el análisis de covarianza [Cochran y Cox (1957)] es una investigación experimental que ajusta estimaciones de efectos de tratamientos a influencias ambientales en la misma manera que la post-estratificación y la estimación de regresión [Cochran (1977)] son usados para ajustar los resultados de estimación basada en muestreo. La estructura de análisis de varianza es usada por ambas áreas como una manera de resumir información asociada con muchos de éstos métodos de control

básicos.

3. CONCEPTOS PARALELOS IDENTIFICADOS MAS RECIENTEMENTE

En los últimos años de 1950 y 1960, los investigadores empezaron a investigar una nueva clase de paralelos entre experimentos y muestreo, que saca partido de las estructuras paralelas de los diseños en innovadoras maneras. El trabajo de Chakrabarti (1963), referido en Tanur y Fienberg (1987), relata la idea del apoyo de un plan de muestreo para la estructura de un diseño de bloques incompletos balanceados (BIBD). El apoyo de un plan de muestreo consiste de todas aquellas muestras que tienen probabilidad positiva de selección. El objetivo de la investigación es desarrollar planes de muestreo equivalentes a un muestreo aleatorio simple (SRS) con respecto a la probabilidad de inclusión de primer y segundo orden para cada unidad y par de unidades de la población, pero las cuales tienen un apoyo más pequeño que un SRS. Si el plan de muestreo tiene probabilidad de inclusión de primer orden equivalente al de el SRS entonces se obtendrán muestras con medias que son estimadores insesgados de la media poblacional. Pero si además, el plan de muestreo tiene probabilidad de inclusión de segundo orden equivalente al de el SRS, la media de la muestra tendrá varianza igual a la de la media de la muestra obtenida por SRS. La motivación para éste objetivo proviene de: (1) La idea de aleatorización restringida que elimina las posibles muestras que no son "representativas" sin sacrificar las ganancias logradas por la aleatorización, (2) el deseo de reducir el costo de la colección de datos a través de un cambio implícito entre estratificación y agrupamiento.

Recordamos la definición de un BIBD con $bk=rt$ unidades experimentales:

- (a) cada uno de los b bloques contiene k unidades.
- (b) cada uno de los t tratamientos ocurre el mismo número de

veces, r , en todos.

- (c) todos los pares de tratamientos ocurren a un mismo tiempo, en el mismo bloque, el mismo número de veces, λ .

Después de identificar la semejanza entre pares de tratamientos en un bloque y pares de unidades en una muestra, la estructura paralela se presenta en la tabla 2:

B I B D	PLAN DE MUESTREO
Bloques Tratamientos Tamaño del bloque, k No. de tratamientos, t No. de bloques, b	Muestras Unidades Tamaño de la muestra, n No. de unidades en la población Tamaño de apoyo para el plan de muestreo.

TABLA 2. Paralelos entre BIBD y un plan de muestreo.

Una manera para restringir el número de muestras en el apoyo de un plan de muestreo es comenzar con alguna forma de estratificación, y la selección de las unidades dentro de los estratos es de acuerdo a alguna variable de estratificación adicional, y no es obligatorio usar una estratificación más estricta. Goodman y Kish (1950), referidos en Rao (1985), sugieren un mecanismo llamado selección controlada, la cual es especialmente útil en el muestreo multietápico. Aquí la selección controlada se usa para seleccionar un número pequeño de unidades de muestreo en la primera etapa, y agrupación típica en la etapa final. Morris (1979), también referido en Rao (1985), propone el Modelo de Selección Finita (FSM) para asignación de tratamientos en un experimento por "selección balanceada" de sujetos de la población finita disponible para la asignación, donde el balance es realizado con respecto a un gran número de covariables.

Otros ejemplos de paralelos nuevos que surgen dentro de la literatura, están estrechamente enlazados con los conceptos de diseño óptimo. Chen and Li (1983, 1985) proponen clases de criterios óptimos para obtener la teoría del diseño experimental

óptimo para mostrar el enlace entre el criterio minimax y las estrategias tradicionales de muestreo sugeridas por Rao (1962). Finalmente, Iachan y Jones (1984) proponen una clase general de rotación de diseños de muestreo contando con modelos ANOVA e ideas relatadas sobre la investigación óptima para BIBD. Todas estas contribuciones son referidas y acotadas en los trabajos de Tanur y Fienberg (1985, 1987).

4. INFERENCIA ANALITICA: CRITERIOS DE UNIFICACION DE INFERENCIA EN MUESTREO Y DISEÑO DE EXPERIMENTOS.

Recientemente el advenimiento y popularización del enfoque superpoblacional considerando el problema de inferencia para datos de encuestas [Ojeda (1990)] ha propiciado que se discuta la necesidad de unificar los criterios en las dos áreas. Esta necesidad se sigue del hecho de que los modelos de regresión y de análisis de varianza han sido los que se han incorporado al análisis de datos de encuestas. Fundamentalmente, en el caso de encuestas bajo diseños complejos (selección en varias etapas) se ha hecho necesaria la consideración de modelos de efectos mixtos [Pfefferman y Smith (1985), Aitkin y Longford (1986) y Tate y otros en (1989)]. Aunado a esto se debe considerar la fuerza unificadora del enfoque predictivo para la inferencia en la población finita, el cual permite incorporar el modelo superpoblacional para resolver el problema de estimación de funciones de los valores de la variable respuesta en la población finita. [Smith (1984), Rodrigues y otros (1985), Tam (1988) y Brewer y otros (1988)]. Finalmente, debemos también señalar las ideas unificadoras que han surgido como modelos generales de inferencia, destacando la inferencia causal [Rubin (1978), Holland (1986a), (1986b), Rosenbaum (1984)].

Estudios recientes, como el de Smith y Sugden (1988), muestran que el enfoque superpoblacional combinado con la teoría de la inferencia causal, considerando el manejo de la aleatoriedad en los niveles que se definen en la sección 2, permiten unificar el

análisis inferencial en experimentos y muestras obtenidas según diseños específicos. Sin embargo dichas tendencias no están libres de discusión y los "clásicos" en ambas áreas defienden la idea de que los problemas que se atacan son distintos y por lo tanto las soluciones son en general distintas.

Para el Estadístico que se dedica fundamentalmente a las aplicaciones, la última postura aún sigue resultando la más razonable, sobre todo por la tradición de "prácticas diferentes", de enseñanza de métodos por separado, y fundamentalmente porque los trabajos unificadores son aún poco difundidos y poco discutidos en los salones de clase y seminarios dedicados a éstos temas.

Agradecemos las observaciones y sugerencias de dos arbitros anónimos, las cuales nos ayudaron a mejorar la redacción del artículo en algunas partes.

BIBLIOGRAFIA

Aitkin M., and Longford N. (1986). Statistical modelling issues in school effectiveness studies. Jour. of the Royal Statistical Society A, Vol. 149, 1-43.

Brewer, K.R.W., Hanif M. and Tam S.M. (1988). How nearly can Model-Based prediction and Design-based estimation be reconciled?. Jour. Am. Stat. Ass. 83, 401. p.p. 128-132

Cochran, W.C. (1977). Sampling Techniques. Third Edition. New York: Wiley. ((1982) Técnicas de muestreo; CECSA, México).

Cochran, W.C. and Cox M. (1957). Experimental Designs. Second Edition. New York: Wiley. ((1986) Diseños experimentales; Trillas, México).

Fienberg, S.E. and Tanur, J.M. (1983). Reuniting the twain:

remarking on the parallels between sample surveys and randomized experiments. Paper presented at International Statistical Institute Meetings, Madrid, Spain.

Fienberg, S.E. and Tanur, J.M. (1985). A long and honorable tradition: Intertwining concepts and constructs in experimental design and sample surveys. Bull. Internat. Statist. Inst., Book II, 10.1-1-10.1-18.

Fienberg, S.E. and Tanur, J.M. (1985). The design and analysis of longitudinal surveys: controversies and issues of cost and continuity. Unpublished manuscript.

Fienberg, S.E. and Tanur, J.M. (1987). Experimental and sampling structures: Parallels diverging and meeting. Internat. Statist. Rev., 55, 75-96.

Fienberg, S.E. and Tanur, J.M. (1988). From the inside out and the outside in: combining experimental and sampling structures. The Canadian Journal of Statistics. Vol. 16, No. 2, pages 135-151.

Fisher, R.A. (1935). The Design of Experiments. Oliver & Boyd, Edinburgh.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. Journal of the Royal Statistical Society (B), 17: 269-278.

Hansen, M.H. Hurwitz, W.N. and Madow, W.G. (1953b). Sample Survey Methods and Theory, Vol. II. New York: Wiley.

Holland, P.W. (1986). Statistics and causal inference. J. Am. Statist. Assoc. 77, 237-250.

Holland, P.W. (1986a). "Which comes first, cause or effect?" New York Statistician 38: 1-6.

- Hollad, P.W. "Statistics and causal inference." J. of the Amer. Stat. Assn. 81: 945-970.
- Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society (A), 109: 558-606.
- Ojeda M.M. (1990). Sobre la problemática del análisis de datos de encuestas; en las memorias del IV FORO DE ESTADISTICA, Monterrey Nuevo León, México. (en prensa).
- Pfefferman, D. and Smith, T.M.F. (1985). Regression Models for grouped populations in cross-section Surveys; Int. Stat. Review 53, 37-59.
- kao, C.R. (1985). Evolution of data collection censuses, sample surveys, and design of experiments; I.S.I., Congress.
- Rodrigues J., Bolfarine, H. and Rogatko (1985). A General Theory of Prediction in Finite Populations; Int. Stat. Review 53, 3, p.p. 239-254.
- Rosenbaum, P.R. (1984a). "From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment." J. of the Amer. Stat. Assn. 79: 41-48.
- Royall, R.M. (1976). Current advances in sampling theory: implications for human observational studies. Am. J. Epidemiol. 104, 463-473.
- Rubin, D.B. (1978). Bayesian inference for causal effects: the role of randomisation. Am. Statistician 6, 34-58.
- Smith, T.M.F. (1976). The foundations of survey sampling: a review J.R. Statist. Soc. A 139, 183-204.

Smith, T.M.F. (1983). On the validity of inferences from non-random samples. J.R. Statist. Soc. A 146, 394-403.

Smith, T.M.F. (1984). Present position and potential developments, some personal views, sample surveys. J.R. Statist. Soc. A 147, 208-221.

Smith, T.M.F. and Sugden, R.A. (1988). Sampling and Assignment Mechanisms in Experiments, Survey and Observational Studies. International Statistical Review, 56, 2 pp. 165-180.

Stephan, F.F. (1948). History of the uses of modern sampling procedures. J. Am. Statist. Assoc. 43, 12-39.

Tam, S.M. (1988). Asymptotically design-unbiased predictors in survey sampling; Biometrika 75, 1 p.p. 175-7.

ANALISIS DE INTERVENCION DE SERIES DE TIEMPO UNIVARIADAS

Mat. José Elías Rodríguez Muñoz
Centro de Investigación en Matemáticas
Apartado Postal 402
Guanajuto, Gto
C. P. 36000

RESUMEN: El análisis de series de tiempo basado en los modelos ARIMA tradicionales no consideran los efectos causados por fenómenos exógenos a su comportamiento histórico. Box y Tiao proponen el Análisis de Intervención como una ampliación de la metodología basada en los modelos ARIMA. En el presente artículo se revisa la metodología y se ejemplifica con una aplicación.

PALABRAS CLAVE: serie de tiempo, modelos ARIMA, modelo Dinámico de Intervención.

1. INTRODUCCION

Una serie de tiempo puede verse, informalmente, como las observaciones de una variable ordenadas respecto al tiempo.

El ANÁLISIS DE SERIES DE TIEMPO tiene gran utilidad en el mundo actual. Por ejemplo, en el campo de la Economía, se está familiarizado con series como la del índice de precios al consumidor, del producto interno bruto, etc. En Medicina, un encefalograma es de interés para analizar la actividad cerebral de un individuo. En una industria el registro periódico del promedio de una característica del producto servirá para detectar o prevenir anomalías en el proceso de manufactura.

El fin último en el análisis de series de tiempo es el de PRONOSTICAR, explicar el fenómeno en estudio o para efectos de control. En el presente artículo el objetivo es el pronóstico a corto plazo de las series analizadas. La experiencia ha mostrado que los modelos ARIMA son una herramienta útil para este objetivo.

Ahora, formalmente una serie de tiempo puede verse como la realización de un proceso estocástico, esto es, como las observaciones de un conjunto de variables aleatorias $\{Y(t); t \in T\}$, donde T es un conjunto índice y $Y(t)$ es la variable aleatoria correspondiente al

elemento t de T . En particular, si consideramos una serie de tiempo discreta como las observaciones de un proceso estocástico en los instantes t_1, t_2, \dots, t_N , las variables aleatorias respectivas se denotarán por $\{Y(t_1), Y(t_2), \dots, Y(t_N)\}$. En adelante se identificará $Y(t_1), Y(t_2), \dots, Y(t_N)$ con y_1, y_2, \dots, y_N , ya que se considera que las observaciones del proceso se hicieron a intervalos de tiempo regulares.

Así pues, dos supuestos fundamentales con los que se trabajará en la construcción de los modelos ARIMA son: que las series son ERGÓDICAS y ESTACIONARIAS (al menos de segundo orden).

Ahora bien, las series con las que se trabaja en ocasiones se ven afectadas por fenómenos exógenos, alterando su comportamiento histórico. Por ejemplo, la serie económica del índice de precios al consumidor puede ser afectada por devaluaciones de la moneda nacional frente a una divisa extranjera; la cantidad de plomo medido diariamente en la atmósfera de una gran urbe puede ser modificada por un cambio en el tipo de gasolina usada por los automóviles, la producción mensual de una empresa puede verse alterada por cambios tecnológicos hechos en la misma o fenómenos naturales pueden afectar a series de diferentes tipos.

Este tipo de situaciones no son previstos en la modelación ARIMA tradicional, por lo cual resulta conveniente ampliar el análisis y los modelos, de modo que se incluyan los casos arriba mencionados.

El ANÁLISIS DE INTERVENCIÓN presentado por primera vez de manera explícita por Box y Tiao (1975), permite cubrir los efectos causados por fenómenos exógenos al comportamiento histórico de la serie, el cual, junto con los modelos ARIMA, lo convierte en una herramienta poderosa para analizar este tipo de situaciones.

La sección 2 hace una breve presentación de los modelos ARIMA. El Análisis de Intervención es el tema central del presente artículo, por lo cual, la parte teórica, basada principalmente en el artículo de Box y Tiao (1975) y las notas de Guerrero (1983), se muestra en la sección 3. Una aplicación práctica de esta metodología a series

de tiempo de interés para TELMEX se muestra en la sección 4. Finalmente, la sección 5 presenta las conclusiones.

2. MODELOS ARIMA

Se supondrá que la serie $\{y_t\}_{t=1}^N$ sigue un Modelo Autorregresivo y de Promedios Móviles ARMA(p,q)

$$\varphi(B)y_t = \theta(B)a_t \quad 2.1$$

donde

- i) y_t proviene de la serie original,
- ii) a_t proviene de una serie de variables aleatorias de ruido blanco,
- iii) B es el operador de retraso ($By_t = y_{t-1}$) y
- iv) $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ y $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$

Si el nivel de la serie sigue una TENDENCIA POLINOMIAL ADAPTIVA DE GRADO d , será necesario aplicar a la serie original el operador diferencia d veces

$$\nabla^d y_t = (1-B)^d y_t \quad 2.2$$

De esta forma la expresión (2.1) se convierte en un Modelo Autorregresivo e Integrado de Promedios Móviles ARIMA(p,d,q)

$$\phi(B)(1-B)^d y_t = \theta(B)a_t \quad 2.3$$

También, si la serie tiene un efecto ESTACIONAL de periodo s , entonces el modelo (2.3) se convierte en

$$\phi_1(B^s)\phi_2(B)(1-B^s)^D(1-B)^d y_t = \theta_1(B^s)\theta_2(B)a_t \quad 2.4$$

que da origen a los Modelos ARIMA multiplicativos estacionales, ARIMAS(p,d,q)x(P,D,Q)_s.

La referencia clásica para esta familia de modelos es Box y Jenkins (1976).

3. ANALISIS DE INTERVENCION: TEORIA

A menudo, las series con las que uno trabaja son INFLUENCIADAS POR FENÓMENOS AJENOS A SU COMPORTAMIENTO HISTÓRICO, que pueden ser cambios de políticas sociales o económicas, cambios provocados por la implantación de una ley o cambios provocados por la naturaleza.

En particular, como se expone en Guerrero (1983), pensemos que una serie de tiempo estacionaria está constituida por dos partes, una determinística que refleja el nivel de la serie y otra estocástica, que da origen a las fluctuaciones alrededor de dicho nivel.

Así, podemos distinguir tres tipos de efectos:

- i) Aquellos que se dejan sentir como una elevación o caída momentánea del nivel y desaparece sin influir sobre el comportamiento posterior de la serie.
- ii) Los que ejercen una influencia sostenida sobre el nivel de la serie, pero dejan intacta la estructura básica de su parte estocástica.
- iii) Los efectos que, independientemente de influir o no sobre la parte determinística, sí alteran la estructura de la parte estocástica.

Las intervenciones que causan los efectos (i) y (ii) son los que, de manera ortodoxa, pueden estudiarse mediante el Análisis de Intervención, mientras que las intervenciones causantes de los efectos del tipo (iii) requieren otro tipo de análisis que considere un cambio estructural en el modelo de la serie.

Ahora bien, se considera que para la serie $\{y_t\}$, el modelo ARIMA(p,d,q)

$$\phi(B)\nabla^d y_t = \theta(B)a_t \quad 3.1$$

o bien

$$y_t = \frac{\theta(B)}{\phi(B)\nabla^d} a_t \equiv N_t \quad 3.2$$

no es suficiente para explicar totalmente el comportamiento de dicha serie, ya que hace falta incluir en el modelo los efectos de la intervención a la variable observada.

Un modelo que contempla dichos efectos fue propuesto en Box y Tiao (1975) como

$$y_t = N_t + \sum_{\mathfrak{S}} \varepsilon_t(T) \quad 3.3$$

donde y_t viene de la serie original $\{y_t\}$, la variable N_t representa a y_t sin los efectos de intervención, es decir N_t está dada por el modelo ARIMA(p,d,q) (3.2) y $\varepsilon_t(T)$ es una función que representa los efectos al tiempo t de la intervención en el momento $T \in \mathfrak{S}$, siendo \mathfrak{S} el conjunto de fechas donde se dieron las intervenciones. Esto es

$$\sum_{\mathfrak{S}} \varepsilon_t(T) \quad 3.4$$

representa el conjunto de efectos de las intervenciones al tiempo t .

En general, la función (3.4) se llama la **Función Dinámica de Intervención** y se representa por la ecuación en diferencias

$$\delta_T(B)\nabla^{d_T} \varepsilon_t(T) = \omega_T(B)P_t(T) \quad 3.5$$

donde

$$\begin{aligned} \nabla^{d_T} &= (1-B)^{d_T} \\ \delta_T(B) &= 1 - \delta_{T,1}B - \dots - \delta_{T,r_T}B^{r_T} \\ \omega_T(T) &= \omega_{T,0} + \omega_{T,1}B + \dots + \omega_{T,s_T}B^{s_T} \end{aligned} \quad 3.6$$

ya que las ecuaciones en diferencias proporcionan una manera adecuada para especificar modelos dinámicos, como se muestra en Guerrero (1983). Aquí el operador B opera sobre el índice t, y $P_t(T)$ es la función de pulso definida por

$$P_t(T) = \begin{cases} 1 & \text{si } t=T \\ 0 & \text{si } t \neq T \end{cases} \quad 3.7$$

donde T denota el momento en el que ocurrió la intervención. En otras palabras $P_t(T)$ nos indica la ocurrencia ($t=T$) o la no ocurrencia ($t \neq T$) de la intervención.

Además, se supone que

$$\varepsilon_t(T) = 0 \quad \text{para } t < T \quad 3.8$$

lo que indica que antes que ocurriera la intervención no existían efectos atribuibles a ella.

Reescribiendo la expresión (3.5) * $\omega_{T,0}$

$$\varepsilon_t(T) = \frac{\omega_T(B)}{\delta_T(B)\nabla^{d_T}} P_t(T) \quad 3.9$$

que es lo que se conoce como la función dinámica general de intervención.

Con lo anterior, la expresión (3.3) se convierte en el Modelo General de intervención que queda como

$$y_t = \sum_{\mathfrak{S}} \left[\frac{\omega_T(B)}{\delta_T(B)\nabla^{d_T}} P_t(T) \right] + \frac{\Theta(B)}{\Phi(B)\nabla^d} a_t \quad 3.10$$

4. ANALISIS DE INTERVENCION: APLICACION

En esta sección se ilustrará la metodología del Análisis de Intervención con series de tiempo de interés para Teléfonos de México (TELMEX). Se analizará la serie LDNT^{1,2} (miles de llamadas de larga distancia nacional, LADA).

Como aplicación del modelo, se presentan los pronósticos de 6 períodos, tanto en forma puntual como por intervalos.

4.1 Larga Distancia Nacional Total

La gráfica de la serie LDNT se muestra en la figura 4.1 y abarca de enero de 1975 a diciembre de 1989. Si se observa su gráfica, se nota su rápido crecimiento con el tiempo, así como un incremento en la variabilidad con respecto de su media; para estabilizarla se buscó una transformación potencia del tipo propuesto por Guerrero (1983), encontrándose que la transformación logaritmo natural era conveniente para estabilizarla. Con lo anterior, el modelo ARIMA (estacional, ARIMAS(2,1,0)x(0,1,1)₁₂) que se propone para la serie es

$$(1-\phi_1 B-\phi_2 B^2)\nabla\nabla_{12}LN(LDNT_t)=(1-\theta_{12}B^{12})a_t \quad 4.1$$

Los resultados obtenidos para el modelo (4.1) se muestran en el cuadro 4.1, donde el cuadro se compone de:

- (1) El período de observaciones de la serie y el número de observaciones de que consta.
- (2) Los valores estimados de cada uno de los parámetros.
- (3) Los intervalos del 95% de confianza para cada uno de los parámetros.
- (4) Aquellas correlaciones entre parámetros estimados que sean mayores que 0.5 en valor absoluto.
- (5) La media aritmética de los residuales.

¹La serie analizada fue proporcionada por la Gerencia de Estudios Económicos de TELMEX.

²La serie se presenta en Rodríguez, J. E. (1990).

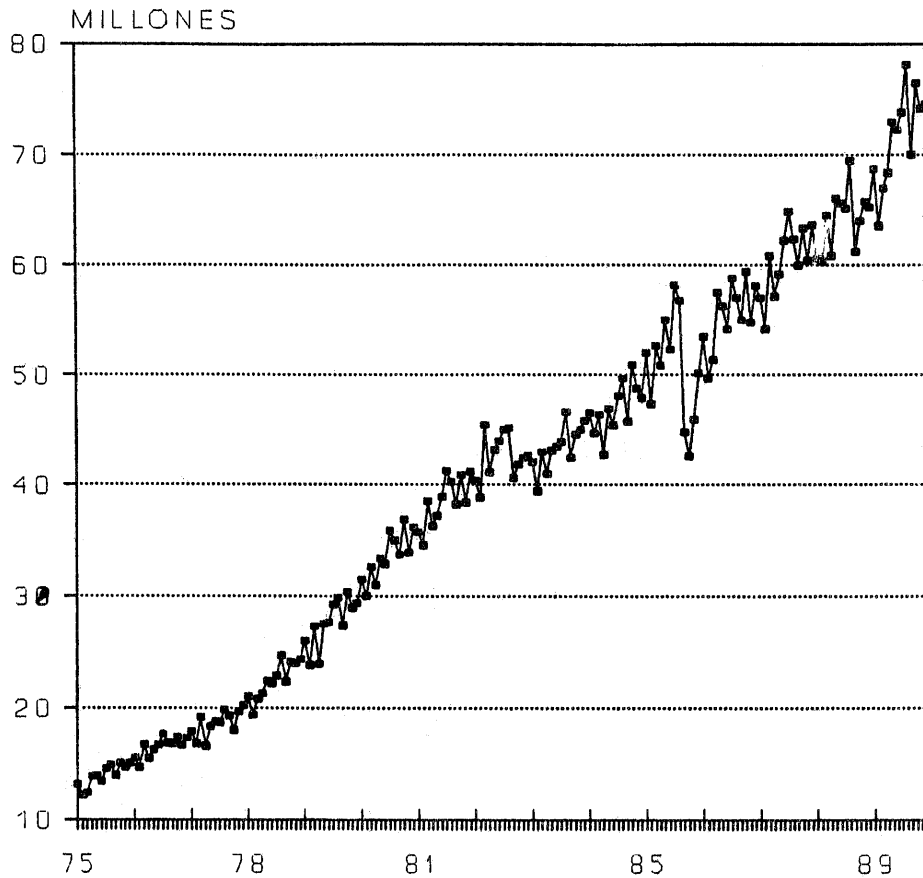


Figura 4.1 LDNT

- (6) La desviación estándar residual.
- (7) El estadístico Q' de Ljung y Box (1978) con sus correspondientes grados de libertad (g.l.).
- (8) Las autocorrelaciones de los residuales que sean significativamente diferentes de cero.
- (9) Los residuales que estén fuera del intervalo $(-3s, 3s)$.

La gráfica de residuales se muestra en la figura 4.2. Como se ve del cuadro 4.1 y de la figura 4.2 existen dos residuales "grandes" que corresponden a septiembre y octubre de 1985. Esta situación también era posible preverla observando la serie original, pues precisamente en septiembre de 1985 la serie parece ser alterada en su comportamiento. La causa de esto coincide con la caída de una de las torres de TELMEX en México, D. F. a causa del temblor de septiembre de 1985.

cuadro 4.1

Serie: LDNT _t		Transformación: LN(LDNT _t)		
Modelo: $(1-\phi_1B-\phi_2B^2)\nabla\nabla_{12}LN(LDNT_t)=(1-\theta_{12}B^{12})a_t$				
(1) Período y número de observaciones	(2) Parámetros estimados	(3) Intervalos del 95% de confianza		(4) Correlación entre parámetros >0.5 ó <-0.5
01,1975-12,1989 N=180	$\phi_1=-0.5381$ $\phi_2=-0.3499$ $\theta_{12}=0.9227$	-0.6767, -0.3995 -0.4881, -0.2117 0.8728, 0.9656		
ANÁLISIS DE RESIDUALES				
(5) m(\hat{a})	(6) s	(7) Q', g.l.	(8) r _k (\hat{a})	(9) Residuales grandes
-0.0040	0.0471	36, 20	r ₉ =0.17 r ₁₄ =0.19	$\hat{a}_{129}=-3.5s$ $\hat{a}_{130}=-4.5s$

Con la observación anterior se propone un modelo de intervención que incluya los efectos de esta causa.

Como el temblor se dió ya avanzado el mes de septiembre (último tercio aproximadamente), el número de conferencias de larga distancia de este mes no se ve tan afectado como el mes de octubre; donde se acentúa la caída en el número de llamadas. Después de octubre la serie recupera su tendencia al crecer con el tiempo.

Con esto, y observando los residuales de la figura 4.2, se proponen las siguientes funciones de intervención:

a) $(\omega_0 + \omega_1 B) P_t(T)$

para abarcar el efecto de caída en el número de llamadas de los meses de septiembre y octubre. Los términos ω_0 y ω_1 representan la caída momentánea de la serie en los meses de septiembre y octubre, respectivamente.

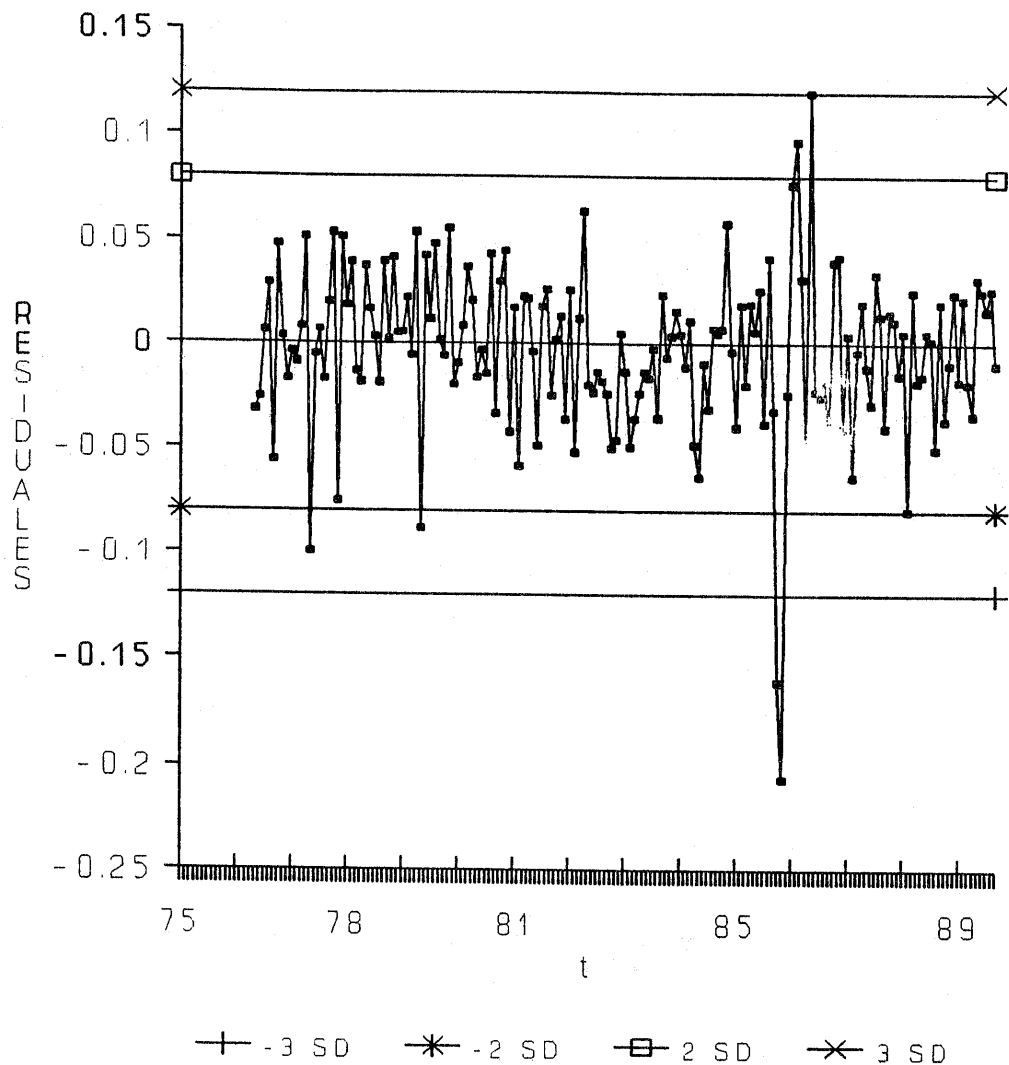


Figura 4.2 RESIDUALES DEL MODELO (4.1)

b)
$$\frac{\omega_2 B^2}{1 - \delta_1 B} P_t(T)$$

para abarcar la posible recuperación, con una tasa mensual δ_1 , a su nuevo nivel a partir del mes de noviembre. El término ω_2 representa el punto desde donde empezará esta recuperación.

c)
$$\frac{\omega_3 B^2}{1 - B} P_t(T)$$

para abarcar el posible cambio de nivel de la serie a partir del mes de noviembre. El término ω_3 representa el nuevo nivel de la serie.

Los tres posibles efectos de intervención se dieron en la misma fecha (T=septiembre de 1985) y tienen la misma causa que explica el efecto observado. Por tanto, la función dinámica de intervención queda como

$$\varepsilon_t(T) = \left(\omega_0 + \omega_1 B + \frac{\omega_2 B^2}{1 - \delta_1 B} + \frac{\omega_3 B^2}{1 - B} \right) P_t(T) \quad 4.2$$

donde, nuevamente, ω_0 representa la caída momentánea de la serie en el mes de septiembre, ω_1 representa la caída momentánea en el mes de octubre (se espera mayor a la del mes de septiembre), ω_2 representa el punto desde donde empezará la recuperación de la serie con una tasa δ_1 al nuevo nivel de la serie ω_3 desde el mes de noviembre.

De esta forma el modelo de intervención para la serie queda como

$$LN(LDNT_t) = N_t + \varepsilon_t(T) \quad 4.3$$

con

$$N_t = \frac{1 - \theta_{12} B^{12}}{(1 - \phi_1 B - \phi_2 B^2) \nabla \nabla_{12}} a_t \quad 4.4$$

y $\varepsilon_t(T)$ como en (4.2).

La estimación de los parámetros del modelo (4.3) se muestran en el cuadro 4.2.

Obsérvese del cuadro 4.2 que ya no existen residuales grandes (lo que se esperaba del Análisis de Intervención).

Finalmente; y como uno de los objetivos del análisis, utilicemos el modelo estimado (4.3) para pronosticar el primer semestre de 1990. Como se aplicó la transformación logaritmo natural a la serie, tendremos que corregir por sesgo los pronósticos, según lo sugiere Guerrero (1983). El cuadro 4.3 muestra tales pronósticos en miles de llamadas.

cuadro 4.2

Serie: LDNT _t		Transformación: LN(LDNT _t)		
Modelo:				
$LN(LDNT_t) = \frac{1 - \theta_{12} B^{12}}{(1 - \phi_1 B - \phi_2 B^2) \nabla \nabla_{12}} a_t + \left(\omega_0 + \omega_1 B + \frac{\omega_2 B^2}{1 - \delta_1 B} + \frac{\omega_3 B^2}{1 - B} \right) P_t(T)$				
INTERVENCIONES: T=SEPTIEMBRE DE 1985				
Período y número de observaciones	Parámetros estimados	Intervalos del 95% de confianza		Correlación entre parámetros >0.5 ó <-0.5
01, 1975-12, 1989 N=180	$\phi_1 = -0.799$ $\phi_2 = -0.5257$ $\theta_{12} = 0.9046$ $\omega_0 = -0.1503$ $\omega_1 = -0.3153$ $\omega_2 = -0.0491$ $\omega_3 = -0.1220$ $\delta_1 = -0.7204$	-0.6767, -0.3995 -0.4881, -0.2117 0.8728, 0.9656 -0.2107, -0.0899 -0.3749, -0.2557 -0.0965, -0.0016 -0.1806, -0.0634 -1.0585, -0.3823		$r(\phi_1, \phi_2) = 0.52$ $r(\omega_2, \delta_1) = -0.67$ $r(\omega_0, \omega_3) = 0.51$
ANÁLISIS DE RESIDUALES				
m(â)	s	Q', g.l.	r _k (â)	Residuales grandes
-0.0026	0.0333	37, 20	$r_9 = 0.18$ $r_{14} = 0.23$	

cuadro 4.3

Fecha	Pronósticos (miles)	Intervalos al 50% de confianza	Intervalos al 95% de confianza
ene-1990	77933.08	76198.82, 79706.80	73003.26, 83195.80
feb-1990	73206.42	71544.97, 74906.43	68485.59, 78252.64
mar-1990	80351.25	78444.48, 82304.36	74938.73, 86154.69
abr-1990	77028.72	74897.86, 79220.20	71002.90, 83565.93
may-1990	81376.40	79053.20, 83767.87	74812.25, 88516.49
jun-1990	81665.92	79223.14, 84184.01	74772.96, 89194.30

5. CONCLUSIONES

Al final de la presente aplicación se tuvo acceso a los datos reales de la serie LDNT para los meses de enero, febrero y marzo de 1990, presentados en el cuadro 5.1. En cada celda del cuadro se presentan: el pronóstico (condicionado a observaciones hasta diciembre de 1989), la observación y el error relativo, en orden descendente.

cuadro 5.1

	P E R I O D O		
	ENE-1990	FEB-1990	MAR-1990
SERIE LDNT	77933.08	73206.42	80351.25
	78369.00	70355.40	78001.22
	(0.6%)	(4%)	(3%)

Con el cuadro 5.1 (o comparando los datos observados con sus respectivos intervalos de confianza) se puede ver la "precisión" de las predicciones de la serie, lo que nos da confianza en los modelos obtenidos para fines de pronóstico.

No se quiere terminar el presente artículo sin hacer la siguiente observación. Las referencias fundamentales fueron Box y Jenkins (1976) para la parte de modelación ARIMA, Box y Tiao (1975) para la de Análisis de Intervención y Guerrero (1983) que incluye ambas partes, que podríamos considerarlos pioneros en la aplicación de esta metodología en el extranjero y en el país, respectivamente. Sin embargo, es de llamar la atención la diferencia que hay entre la gran bibliografía existente sobre los modelos ARIMA y la escasa sobre Análisis de Intervención, que se presentó sólo 5 años después de la primera edición del libro de Box y Jenkins. Es posible que esto se deba a la poca disponibilidad de programas de cómputo estadístico que incluya esta metodología (SAS, BMDP, SCA), o bien, a que los usuarios de técnicas de pronósticos, que generalmente no son estadísticos, se inclinen por métodos menos sofisticados, pero aún en revistas especializadas son pocos los artículos escritos

sobre el tema.

Por último, el fin para el cual fue creado el Análisis de Intervención de series de tiempo es para incluir los efectos causados por fenómenos exógenos al comportamiento de la serie, siempre y cuando se conozca explícitamente estas causas, y no simplemente para reducir errores estimados que se consideran "grandes".

REFERENCIAS

Box, G. E. P. y Jenkins, G. M. (1976). *Time Series Analysis: forecasting and control*. Revised ed. Holden-Day.

Box, G. E. P. y Tiao, G. C. (1975). *Intervención Analysis with Applications to Economic and Environmental problems*. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION. 70, p. 70.

Guerrero, B. M. (1983). *Libro no publicado*. Disponible en fotocopias. México.

Ljung, G. M. y Box, G. E. P. (1978). *On a Measure of lack of Fit in Time Series Models*. BIOMETRIKA, 65, p. 297.

Rodríguez, J. E. (1990). *Análisis de Intervención de Series de Tiempo Univariadas*. Tesis de Licenciatura. México.

USO DE TECNICAS DE CUMULOS EN UN ANALISIS INICIAL DE SECUENCIAS DE AMINOACIDOS

Pablo E. Suardíaz
Gisela Vargas
LINAIE, Fac. de Estadística; U.V.
Av. Xalapa esq. Av. Avila Camacho
Xalapa, Veracruz

RESUMEN: Se presenta la aplicación de metodología estadística multivariada en una investigación del área de biología molecular teórica, desarrollada por los investigadores del Centro de Investigaciones Biológicas (CIB) de la Universidad Veracruzana.

Se analiza estadísticamente la variabilidad de aminoácidos, utilizando los alineamientos de cerca de 1000 inmunoglobulinas presentados en el atlas de Kabat (1987), con respecto a los agrupamientos planteados por Jiménez (1984). Se aplican técnicas de cúmulos para analizar los agrupamientos propuestos en la investigación. Se exponen puntos a ser desarrollados y conclusiones generales.

PALABRAS CLAVE: Técnicas de cúmulos, inmunoglobulinas, aplicación.

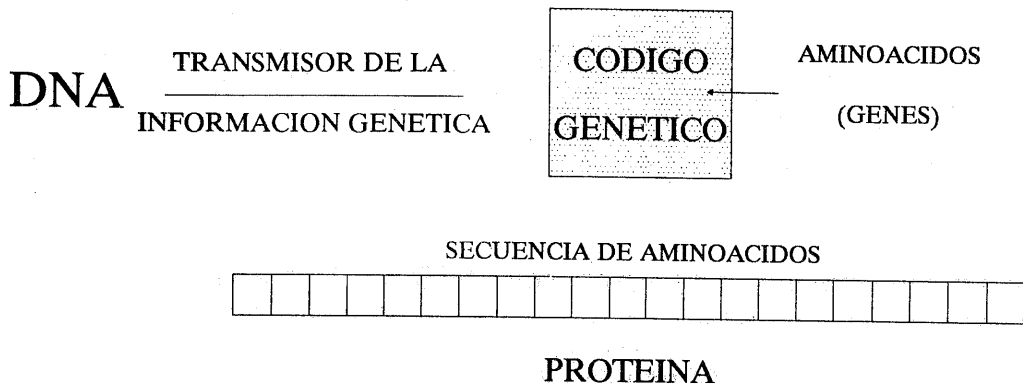
1. INTRODUCCION

"Desde la primera aparición de un mínimo vital, la posición especial asignada a los organismos vivientes por su origen y propósito consta de mensajes codificados que especifican las estructuras moleculares y se transmiten como instrucciones de generación en generación" [Jiménez, 1988].

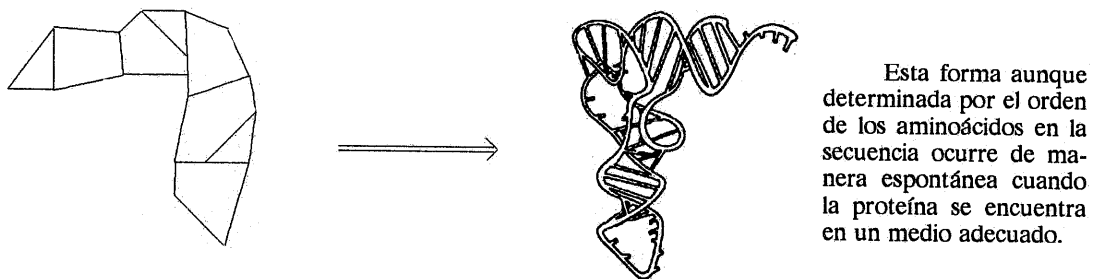
La conformación de este código fue descubierta y descifrada en nuestros tiempos por la biología molecular.

La información genética que es producida por el DNA se expresa al traducirse por medio de dicho código genético a un "alfabeto" de 20 "letras" o aminoácidos (mínima

información genética existente (-GEN-) y constituyen los elementos de la secuencia o estructura primaria de las proteínas.



Para que esta secuencia adquiera un significado o función biológica se despliega en el espacio en una conformación específica.



Ohno (- -) plantea una secuencia tipo como la secuencia primaria común en todo organismo (conformada por 16 aminoácidos), la cual se ve "transformada" por el ruido evolutivo, creándose así una diversificación de secuencias. Posteriormente algunos autores agregan a esta teoría la clasificación de los aminoácidos en grupos que al ser compatibles dentro de grupos de acuerdo a ciertas propiedades pueden sustituirse unos a otros dentro de alguna posición particular sin alterar funcionalmente la secuencia. Jiménez (1984) presenta una jerarquía de categorización de los aminoácidos.

Para delimitar el efecto del ruido dentro de la variación, en un estudio posterior se incorporará el concepto de excepción, el cual permita caracterizar a aquellos casos que se

presenten con una frecuencia no significativa dentro de cada casilla en el alineamiento, ya que estos podrían ser consecuencia del ruido evolutivo o de causas externas al fenómeno bajo estudio.

2. METODOLOGIA

2.1 Bases de datos.

La base de datos original se encuentra constituida por las frecuencias de aparición de los aminoácidos en cada una de las posiciones de cerca de 1000 inmunoglobulinas alineadas (concensos por posición para cada uno de los 20 aminoácidos). De esta manera, se cuenta con una matriz de 20 columnas (correspondientes a los aminoácidos) y 119 filas (posiciones de las inmunoglobulinas), la cual tiene en la entrada (i,j) la frecuencia n_{ij} del j-ésimo aminoácido en la i-ésima posición del alineamiento.

	(1)	(2)	(3)	(4)	. . .	(20)
	W	I	Y	F	. . .	D
1	0	2	0	431	. . .	1
2	0	0	213	1	. . .	1
3	9	0	1	510	. . .	5
.
.
.
119	0	91	124	308	. . .	9

Figura 1. Base de datos IGVL AMINOACIDOS.

La base de datos IGVL AMINOACIDOS fue creada en el paquete DBASE III PLUS.

2.2 Análisis de cúmulos.

Se utilizaron técnicas del análisis de cúmulos para explorar la existencia de algún patrón de agrupación en las frecuencias de aparición de los aminoácidos considerando sus propiedades de polaridad y volumen, que se encuentran en la base de datos IGVL AMINOACIDOS, teniendo como variables a 20 aminoácidos y como individuos a cada posición.

Los métodos utilizados fueron los jerárquicos tradicionales: el algoritmo de conexión simple, de conexión completa, centroide y promedio, para la formación de grupos por aminoácidos, y la medida de similitud fue el coeficiente de correlación de Pearson, medida seleccionada por considerarla más representativa, dado que el análisis de la estructura para este problema nos confiere a la formación de grupos por variable.

Este tipo de algoritmos permiten dar una estructura de clasificación sin requerir de información *a priori* y hacer comparaciones explorando cada dendrograma resultante, pudiendo identificar una agrupación consistente a partir de la detección de ligamientos de alta regularidad [Everitt, 1980].

3. RESULTADOS Y CONCLUSIONES

A partir del análisis de cúmulos se obtienen cuatro dendrogramas, uno por cada algoritmo distinto aplicado.

A continuación se presenta el dendrograma hipotético [Jiménez, 1984] y uno de los dendrogramas resultantes del análisis.

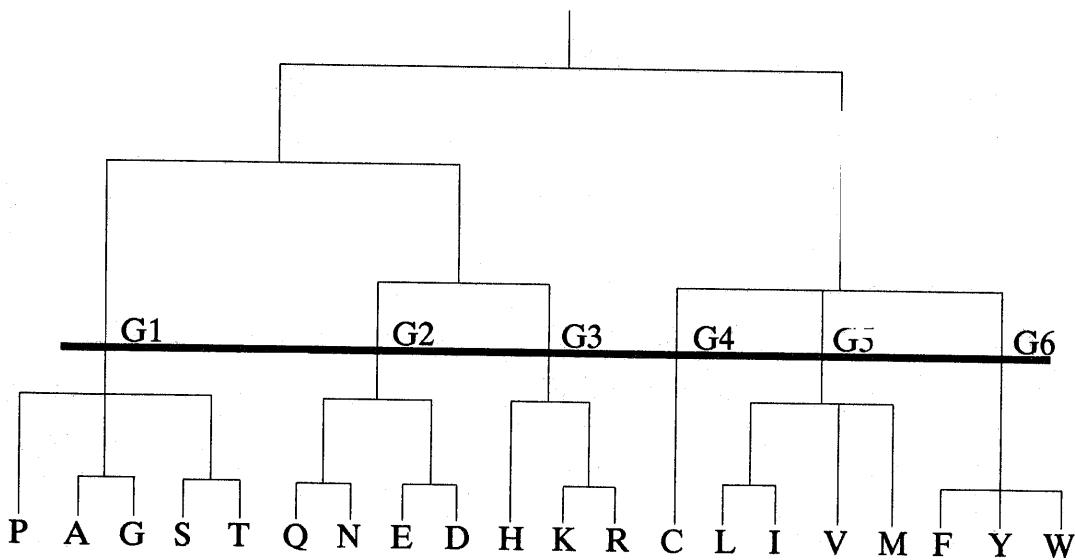


Figura 2. Arbol de agrupación hipotético.

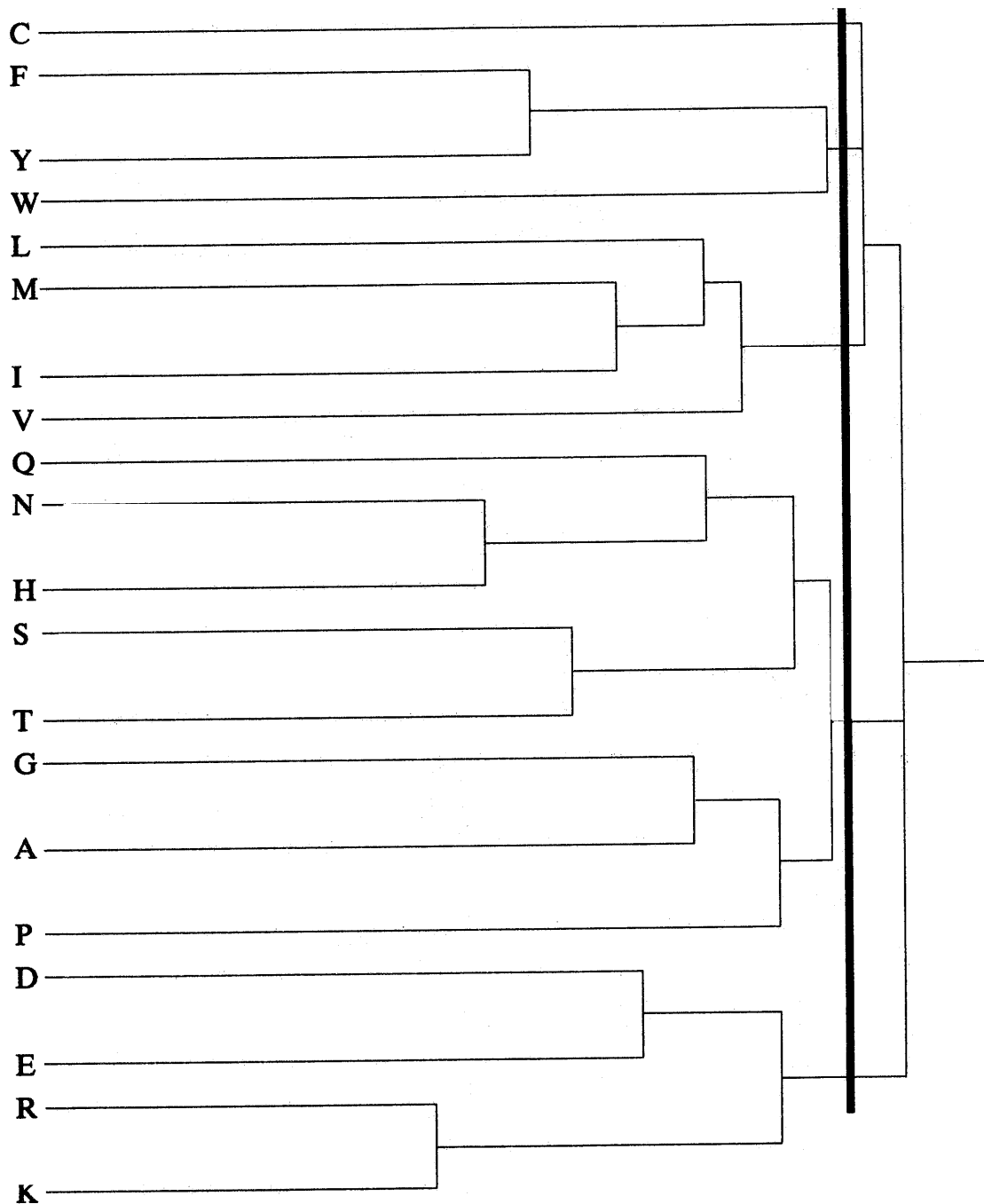


Figura 3. Dendrograma del algoritmo promedio.

Los ligamientos propuestos por Jiménez (1984), Miyata *et al.* (1979), Ratner y Bat-chinsky (1976) y Dayoff *et al.* (1972) se resumen a una clasificación de seis grupos, siendo estos: G1(PAGST), G2(QNED), G3(HKR), G4(C), G5(NLIV) y G6(FYW), ver figura 2.

De la exploración del árbol del algoritmo promedio en un nivel alto marcado por un corte en la gráfica de la figura 3 se obtienen las siguientes características de estructura en cinco grupos:

- El aminoácido C aparece como un grupo desligado de los demás y corresponde al grupo G4 del hipotético, esta característica se presenta en todos los árboles.
- F, Y y W forman un grupo al mismo nivel que equivale al grupo G6.
- L, M, I y V al mismo nivel que los anteriores forman el grupo G5.
- Los aminoácidos D, E, R y K forman un grupo con elementos correspondientes a G2 y G3 excepto H.
- El grupo Q, N, H, S, T, G, A y P está formado por el grupo G1 y elementos de G2 y G3.

El objetivo del análisis de cúmulos aquí realizado no reside en encontrar una consistencia estricta de categorías, sino en probar la existencia de regularidades en los patrones de sustituibilidad.

En el dendrograma presentado a pesar de no encontrarse la estructura hipotética exacta se logran identificar grupos formados por aminoácidos que coinciden con los hipotéticos aunque no en su totalidad, de manera similar ocurre con los dendrogramas restantes para otros grupos. Esto nos motiva a aplicar otro tipo de técnicas para el análisis de grupos como pueden ser las técnicas no jerárquicas con selección del número de grupos y comparación de clasificaciones, dado que se está analizando una información muy compleja con "regiones hipervariables" con alto índice de aparición de excepciones que ameritan un tratamiento especial. Este análisis se tiene contemplado en una segunda fase de la investigación.

Expectativas.

A partir de la discusión del análisis de cúmulos se plantean los siguientes puntos para ser desarrollados en un segundo estudio:

- Caracterizar los valores a considerarse como excepciones y realizar el mismo análisis con propósitos comparativos.
- Llevar a cabo una nueva clasificación sumando las frecuencias de los aminoácidos correspondientes a cada uno de los seis grupos hipotéticos con el propósito de verificar los niveles altos del árbol.
- Adicionalmente se plantea categorizar la variabilidad en seis casos en que puede comportarse el perfil de frecuencias de variabilidad en cada posición de las inmunoglobulinas. Este punto queda pendiente de análisis para corroborar el sistema de sustituibilidad de aminoácidos.
- Realizar el mismo análisis para la base de datos de inmunoglobulinas del dominio variable pesado (IGVH) contenida en el atlas de Kabat.
- Hacer comparación de clasificaciones mediante una matriz de valores cofenéticos de los árboles obtenidos a partir de los distintos métodos utilizados, a través del índice de correlación de Pearson.

BIBLIOGRAFIA

Everitt B. S. (1980)

Cluster Analysis, 2a ed., London, Heineman Education Books, Ltd.

Jiménez-Montaña M. A. (1984)

Bull. Math. Biol., 46, pp. 641-60.

Jiménez Montaña M. A. (1988)

Sobre la lingüística estructural y el lenguaje genético;

La Ciencia y el Hombre, 2, pp. 61-70.

Jiménez Montaña M. A. Vargas E. Ojeda M. M. Vargas G. y Suardías P. (1990)
Análisis estadístico de la variabilidad de los aminoácidos en los dominios variables de las inmunoglobulinas; por aparecer en La Ciencia y el Hombre.

Rohlf F.J. (1974)

Methods of comparing classifications

Annual Rev. of Ecology and Systematics, vol. 5, pp. 101-113.

Vargas G. (1988)

Técnicas de análisis de cúmulos: Algunos problemas estadísticos;

Tesis de Licenciatura. Universidad Veracruzana.

AGRADECIMIENTOS

Agradecemos el apoyo brindado para la realización de este artículo a los investigadores del CIB, Dr. Miguel Angel Jiménez Montaña y Lic. Enrique Vargas; así como la colaboración en la conducción de la asesoría estadística, al M. C. Miguel Ojeda y a Leticia Maldonado estudiante del octavo semestre de la carrera de Estadística.

**COMPONENTES PRINCIPALES.
ANÁLISIS DE DATOS SOBRE COBERTURA
DE ATENCIÓN PRIMARIA A LA SALUD.**

Gustavo J. Valencia^{1,2} y Olga V. Serrano¹.

RESUMEN.

En este trabajo utilizando el procedimiento estadístico conocido como Componentes Principales se construyen tres nuevos indicadores de la cobertura de la Atención Primaria a la Salud. Estos indicadores permiten una descripción simple de la cobertura de los servicios básicos de salud para cada una de las entidades federativas de la República Mexicana.

PALABRAS CLAVE.

ESTADÍSTICA MULTIVARIADA, COMPONENTES PRINCIPALES, SALUD, APS.

1.- INTRODUCCION.

La Organización Mundial de La Salud (OMS) propuso en 1978 la meta "Salud para todos en el año 2000". Desde ese momento, se vuelve muy importante la obtención de indicadores confiables, robustos y con calidad sobre la cobertura de Atención Primaria a la Salud (APS). En 1981 la OMS propuso trece indicadores de cobertura de APS y en 1985 propone como un indicador compuesto (IC) al promedio aritmético (PROM) de los trece indicadores.

Ahora construimos nuevos indicadores, mediante Componentes Principales (CP), basándonos en los trece indicadores anteriores.

2.- ANTECEDENTES.

Existe literatura referente al análisis de Componentes Principales desde 1900. Las aplicaciones del análisis de CP se vieron impedidas en gran medida por la cantidad y dificultad de los cálculos involucrados. Es hasta la segunda mitad del siglo,

1.- Centro de Investigaciones sobre Enfermedades Infecciosas, Instituto Nacional de Salud Pública, México.

2.- Facultad de Ciencias, UNAM.

que con la disponibilidad de computadoras electrónicas se tuvo acceso generalizado a este procedimiento.

3.- ANALISIS DE COMPONENTES PRINCIPALES.

La idea básica del análisis de CP es reducir el conjunto de variables originales correlacionadas, en otro conjunto de variables ortogonales de menor dimensión, de tal manera que se retenga la mayor información esencial. Las nuevas variables son combinaciones lineales de las variables originales.

De manera general, en la práctica sólo se consideran las primeras combinaciones lineales que son las que explican la mayor variabilidad de los datos y así, se descartan las "sobrantes" y se reduce el número de variables a considerar. Esta reducción de variables es de gran utilidad práctica, ya que facilita el manejo matemático y la presentación gráfica de los datos.

Una limitante de importancia del procedimiento de CP es la dificultad de interpretar las nuevas variables. La interpretación de las componentes principales está influida por la creatividad y el conocimiento del investigador.

4.- OBJETIVOS.

La pregunta de interés es ¿pueden los trece indicadores de APS correspondientes a México, en un periodo de tiempo dado, combinarse para obtener un indicador de cobertura único o un número pequeño de indicadores?.

Para responder esta pregunta se ha planteado la siguiente meta: obtener las componentes principales o indicadores compuestos de cobertura de APS.

5.- VARIABLES.

Se han planteado los siguientes 13 indicadores para México (Kumate y Barnard, 1988):

1. Cuidado prenatal: proporción de embarazadas que tuvieron por lo menos una consulta prenatal.

2. Atención del parto: proporción de nacimientos atendidos por personal preparado.
3. Puericultura: proporción de lactantes que tuvieron por lo menos un contacto con una instalación de salud.
4. Vigilancia de crecimiento: proporción de preescolares que han sido pesados por lo menos 2 veces antes de 5 años.
5. Vacunación con BCG: proporción de menores de un año vacunados con BCG.
6. Vacunación con DPT: proporción de menores de un año que recibieron 3 dosis de DPT.
7. Diagnóstico y tratamiento de enfermedades endémicas: proporción del número estimado de casos que se diagnostican y tratan. Se determinó como enfermedades endémicas a las enteritis y a otras enfermedades diarreicas.
8. Atención de la enfermedad terminal: proporción de muertes certificadas por médicos.
9. Instalaciones. La población con acceso razonable a instalaciones médicas.
10. Consultas. El número de consultas médicas por habitantes: se consideró apropiado que fueran 4 consultas/año/habitante=100 por ciento.
11. Hospitalizaciones. El número de egresos hospitalarios. Se determinó como: 8 egresos hospitalarios/año/100 habitantes= 100 por ciento.
12. Agua potable. La proporción de la población con acceso adecuado al agua potable.
13. Facilidades sanitarias. La proporción de la población que dispone de facilidades sanitarias. Se determinó considerar sólo el acceso a drenaje.

Los datos empleados corresponden al año de 1985.

6.- METODO.

Se utilizó el Análisis de CP sobre los trece indicadores de cobertura anteriormente descritos.

El procedimiento de CP producirá trece IC (las componentes principales). Se podrá reducir el número dependiendo de la estructura de correlación existente entre los indicadores originales y de los datos disponibles.

Se consideró apropiado estandarizar los indicadores originales. Esto se hizo por considerar a cada indicador con igual importancia. Por esta razón, se realizó el análisis de CP sobre la matriz de correlaciones.

No se consideró el realizar un contraste estadístico de hipótesis respecto al número de indicadores a utilizarse por dos razones: primero la inexistencia de procedimiento Bayesianos para realizar este contraste; segundo, que los datos no son una muestra, sino que constituyen un censo. En este caso, los procedimientos clásicos para contraste de hipótesis resultan inapropiados. Entonces para elegir el número de componentes resultantes, consideramos otros criterios (ver Jolliffe, 1986, 92-114, y Hair y col., 1987, 246-7). Consideramos como criterios adecuados por sus características los siguientes: el criterio de la raíz principal y el criterio del porcentaje de variación. A continuación se describen someramente estos criterios.

En el criterio de la raíz principal sólo se consideran significativos los valores propios mayores que uno. Esto tiene base al considerar CP realizados con datos estandarizados y por lo tanto, con matrices de correlaciones.

El criterio del porcentaje de variación consiste en declarar como diferentes de cero, a tantas raíces características como sea necesario para que las nuevas variables expliquen un porcentaje de la variación original considerado como satisfactorio. Nosotros decidimos que explicar al menos 80% de variación observada era aceptable.

Para poder interpretar los nuevos indicadores compuestos es necesario analizar los vectores característicos y determinar cuales de sus componentes son suficientemente grandes, para influir de manera significativa en la composición del vector.

En este caso consideramos apropiado el criterio planteado por Hair y col. (1987), respecto a considerar sólo componentes con valor absoluto mayor a 0.30.

7.- RESULTADOS.

Se obtuvieron los valores y los vectores propios correspondientes. En la tabla 1 se presentan los valores propios. Se decidió que el número apropiado de nuevos indicadores compuestos era tres. Esto se hizo considerando que los tres

primeros valores principales resultaban mayores que uno, que la disminución de tamaño del tercer al cuarto valor principal es de 50% aproximadamente y que el porcentaje de variación explicada por estos tres componentes es de 82.2%.

En la tabla 2, se presentan los vectores principales correspondientes a estos tres valores principales.

Con base en que sólo las componentes de valor absoluto mayores que 0.30 intervienen significativamente en la conformación del vector, se procedió a la interpretación e identificación de los vectores principales.

El primer IC de cobertura de APS (IC1), resulta estar formado por los siguientes indicadores:

Cuidado prenatal, Atención del parto, Puericultura, Vigilancia de crecimiento, Consultas y Hospitalizaciones.

Puede interpretarse a este indicador como la Demanda de Atención Médica Institucional (explica un 59.4% de la variación original).

El segundo IC de la APS (IC2), resulta estar formado por:

Diagnóstico y tratamiento de enfermedades endémicas e Instalaciones.

Considerando que el diagnóstico y tratamiento de enfermedades endémicas se da sólo en instalaciones médicas, IC2 puede interpretarse como el Acceso a Instalaciones para Diagnóstico y Tratamiento de Enfermedades Endémicas (explica el 13.042% de la variación original).

El Tercer IC resulta estar constituido principalmente por dos grupos de indicadores con signos opuestos, resultando de esta manera un contraste entre estos dos grupos. Los indicadores que constituyen el primer grupo son los siguientes: Vacunación con BCG y Vacunación con DPT. Los indicadores del segundo grupo son: Atención de la enfermedad terminal, Agua potable y Facilidades sanitarias.

Este tercer indicador IC3, que explica el 9.82% de la variación original, puede pensarse como la Comparación entre

**TABLA 1.- VALORES PROPIOS DE LA MATRIZ DE CORRELACIONES.
DATOS DE 1985.**

	%	%ACUMULADO
7.72193	59.399	59.399
1.69542	13.042	72.441
1.27593	9.815	82.256
0.60673	4.667	86.923
0.425565	3.274	90.197
0.331519	2.550	92.747
0.267184	2.055	94.802
0.189608	1.458	96.260
0.187195	1.440	97.700
0.139909	1.076	98.776
0.081291	0.625	99.401
0.050366	0.387	99.788
0.027352	0.210	100.000

**TABLA 2.- VECTORES PROPIOS DE LA MATRIZ DE CORRELACIONES.
DATOS DE 1985.**

1	2	3
0.30368	0.04642	-0.24861
0.33015	-0.00431	-0.02323
0.31483	0.13996	-0.27177
0.33977	-0.04592	-0.09543
0.27998	-0.13598	-0.32918
0.25604	-0.01401	-0.39500
0.08271	0.68632	0.18495
0.26525	-0.21261	0.30550
0.17633	0.57325	0.23129
0.31490	0.06000	-0.01161
0.32313	-0.01066	0.11603
0.25999	-0.27564	0.39641
0.24660	-0.18134	0.48839

Prevención por Inmunización y Acceso a Recursos de Higiene y Salud.

En la tabla 3, se presentan las correlaciones entre los tres nuevos IC, el indicador PROM y los indicadores originales. Con base en esta tabla puede observarse que PROM e IC1 se correlacionan linealmente de manera perfecta (considerando dos cifras decimales).

Esta correlación muestra que la cobertura de APS puede representarse adecuadamente mediante IC1 y que será equivalente utilizar a PROM o a IC1 para realizar cualquier inferencia o descripción de cobertura de APS. Además, los indicadores simples más importantes de cobertura son los que conforman el indicador IC1. Esto es relevante, pensando en la explicación del fenómeno y en la posibilidad de impulsar intervenciones que puedan aumentar la cobertura.

El segundo indicador IC2, sólo se correlaciona con los indicadores correspondientes a Diagnóstico y tratamiento de enfermedades endémicas e Instalaciones, proporcionando respectivamente explicaciones de 79.21% y 56.25% de la variación de estos indicadores.

El tercer indicador IC3, no se relaciona linealmente y de manera significativa con algún indicador simple. Este indicador proporciona una información no explícita en los datos. Muestra que una de las componentes importantes de los datos de cobertura es el balance entre las acciones de prevención por inmunización (vacunaciones) y el acceso a los recursos de higiene y salud (el agua potable y el drenaje).

En la figura 1, se representan a los indicadores simples y a PROM utilizando las correlaciones de IC1 e IC2 con cada uno de ellos. Las claves para la identificación de los indicadores están en la tabla 4.

Esta representación muestra agrupaciones de indicadores. Por una parte están los indicadores 7 y 9, que son los que forman a IC2. En la parte inferior derecha se distingue a PROM (letra E), las variables 1, 2, 3, 4, A y B que son las que forman a IC1 y

TABLA 3.- CORRELACIONES ENTRE INDICADORES COMPUESTOS E INDICADORES SIMPLES.

		IC1	IC2	IC3
1	PRENAT	0.84	0.06	-0.28
2	PARTO	0.92	-0.01	-0.03
3	LACTAN	0.87	0.18	-0.31
4	PREESC	0.94	-0.06	-0.11
5	BCG	0.78	-0.18	-0.37
6	DPT	0.71	-0.02	-0.45
7	DIARREAS	0.23	0.89	0.21
8	TERMINAL	0.74	-0.28	0.35
9	INSTAL	0.49	0.75	0.26
10	CONS	0.88	0.08	-0.01
11	HOSP	0.90	-0.01	0.13
12	AGUAP	0.72	-0.36	0.45
13	DRENAJE	0.69	-0.24	0.55
	PROM	1.00	0.06	0.00

TABLA 4.- IDENTIFICADORES DE LOS INDICADORES SIMPLES Y EL INDICADOR COMPUESTO PROM.

1	Cuidado prenatal.
2	Atención del parto.
3	Puericultura.
4	Vigilancia de crecimiento
5	Vacunación con BCG.
6	Vacunación con DPT.
7	Diag. y trat. de enfermedades endémicas.
8	Atención de la enfermedad terminal.
9	Instalaciones.
A	Consultas.
B	Hospitalizaciones.
C	Agua potable.
D	Facilidades sanitarias.
E	Indicador compuesto PROM.

por último a las variables 5, 6, 8, D y C que son las que forman a IC3.

8.- DISCUSION.

Utilizando las dos primeras CP se presentan en la figura 2 las 32 entidades federativas. Las agrupaciones corresponden a las regiones descritas anteriormente. Con base en IC1 e IC2, pueden explicarse las diferencias entre entidades federativas, resultando que:

- La región 1 esta formada por entidades con valores de IC1 altos (los más grandes) y valores de IC2 también grandes.
- La región 2, esta formada por entidades con valores (en general) más pequeños que los observados respecto a IC1 en la región 1. Los valores correspondientes a IC2, son en general más altos que los de la región 1. Sin embargo, al presentar valores bajos de IC1 y valores más altos de IC2, puede inferirse que la prevalencia de la enfermedad endémica (enfermedades diarreicas) debe ser mayor en estas entidades que en las consideradas en la región 1.
- en la región 3 se tienen los valores más bajos de IC1 y valores de IC2 comparables a los de la región 2. Recordando la discusión presentada en el párrafo anterior, resulta aquí también una prevalencia alta de la enfermedad endémica. Esta prevalencia debe ser la mayor de las seis regiones.

Puede observarse que Guerrero es el estado con menor cobertura de IC1. Esto puede deberse a varios factores: uno, la baja cobertura de servicios de salud en general y otra el posible uso de medicina alternativa, específicamente medicina indígena.

- La región 4 presenta valores de demanda intermedios, entre los observados en la región 3 y la región 2. Los valores correspondientes a IC2, son más bajos que los

FIGURA 1.- REPRESENTACION DE LOS INDICADORES SIMPLES Y PROM
UTILIZANDO LAS CORRELACIONES RESPECTO A IC1 Y IC2

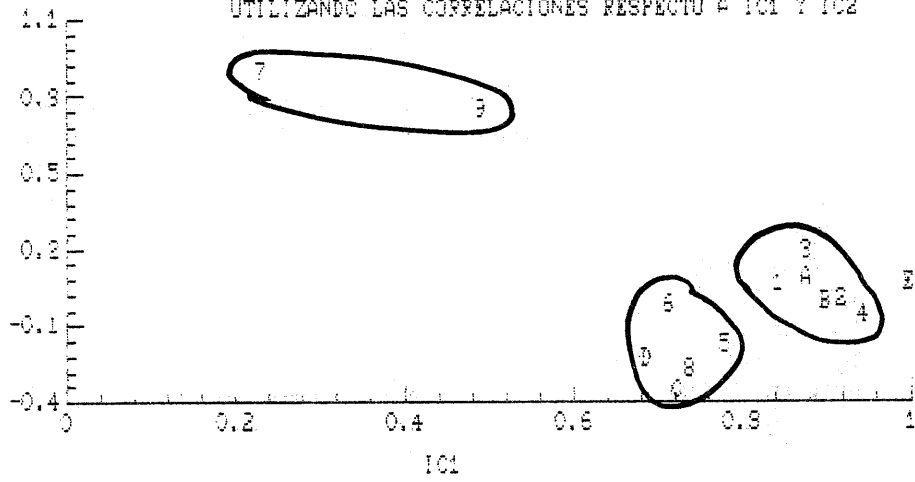


FIGURA 2.- REPRESENTACION DE LAS 32 ENTIDADES FEDERATIVAS
UTILIZANDO LOS DOS PRIMEROS INDICADORES COMPUESTOS.

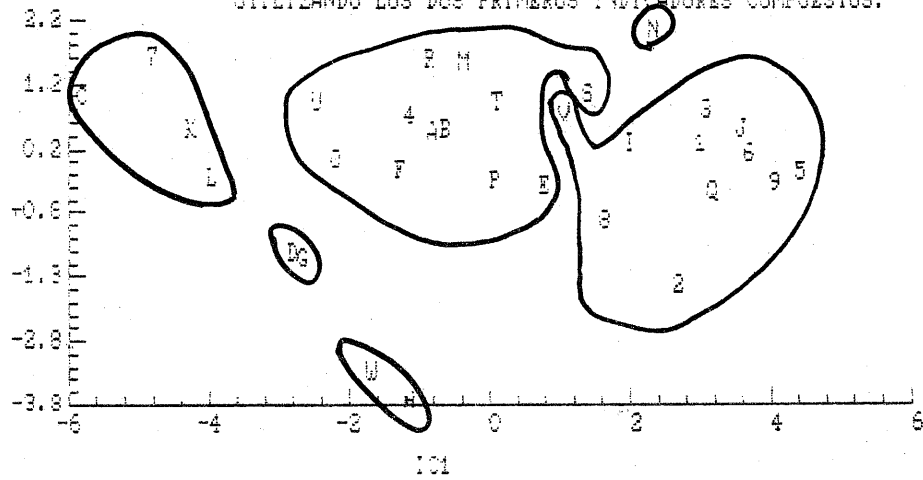


TABLA 5.- IDENTIFICADORES DE LAS ENTIDADES FEDERATIVAS.

1	AGUASCALIENTES	2	BAJA CALIFORNIA
3	BAJA CALIFORNIA SUR	4	CAMPECHE
5	COAHUILA	6	COLIMA
7	CHIAPAS	8	CHIHUAHUA
9	DISTRITO FEDERAL	A	DURANGO
B	GUANAJUATO	C	GUERRERO
D	HIDALGO	E	JALISCO
F	MEXICO	G	MICHOACAN
H	MORELOS	I	NAYARIT
J	NUEVO LEON	K	OAXACA
L	PUEBLA	M	QUERETARO
N	QUINTANA ROO	O	SAN LUIS POTOSI
P	SINALOA	Q	SONORA
R	TABASCO	S	TAMAULIPAS
T	TLAXCALA	U	VERACRUZ
V	YUCATAN	W	ZACATECAS

observados en la región 2, lo que indica una falta de acceso a las instalaciones para el diagnóstico y tratamiento de la enfermedad endémica.

- La región 5 presenta valores de IC1 comparables a los bajos de la región 2. Sin embargo, los valores de IC2 son los más bajos de todos. Morelos presenta el valor más pequeño. Esto indica que las entidades de la región presentan los niveles más bajos de acceso al diagnóstico y tratamiento de la enfermedad endémica.

En este caso y por presentar valores de IC1 comparables a los observados entre algunas de las entidades de la región 2, podría pensarse que esta región está formada por entidades donde la población no acude a los centros de salud para diagnóstico y tratamiento de la enfermedad endémica.

Posiblemente, sectores importantes de la población tratan esta enfermedad con medicina alternativa.

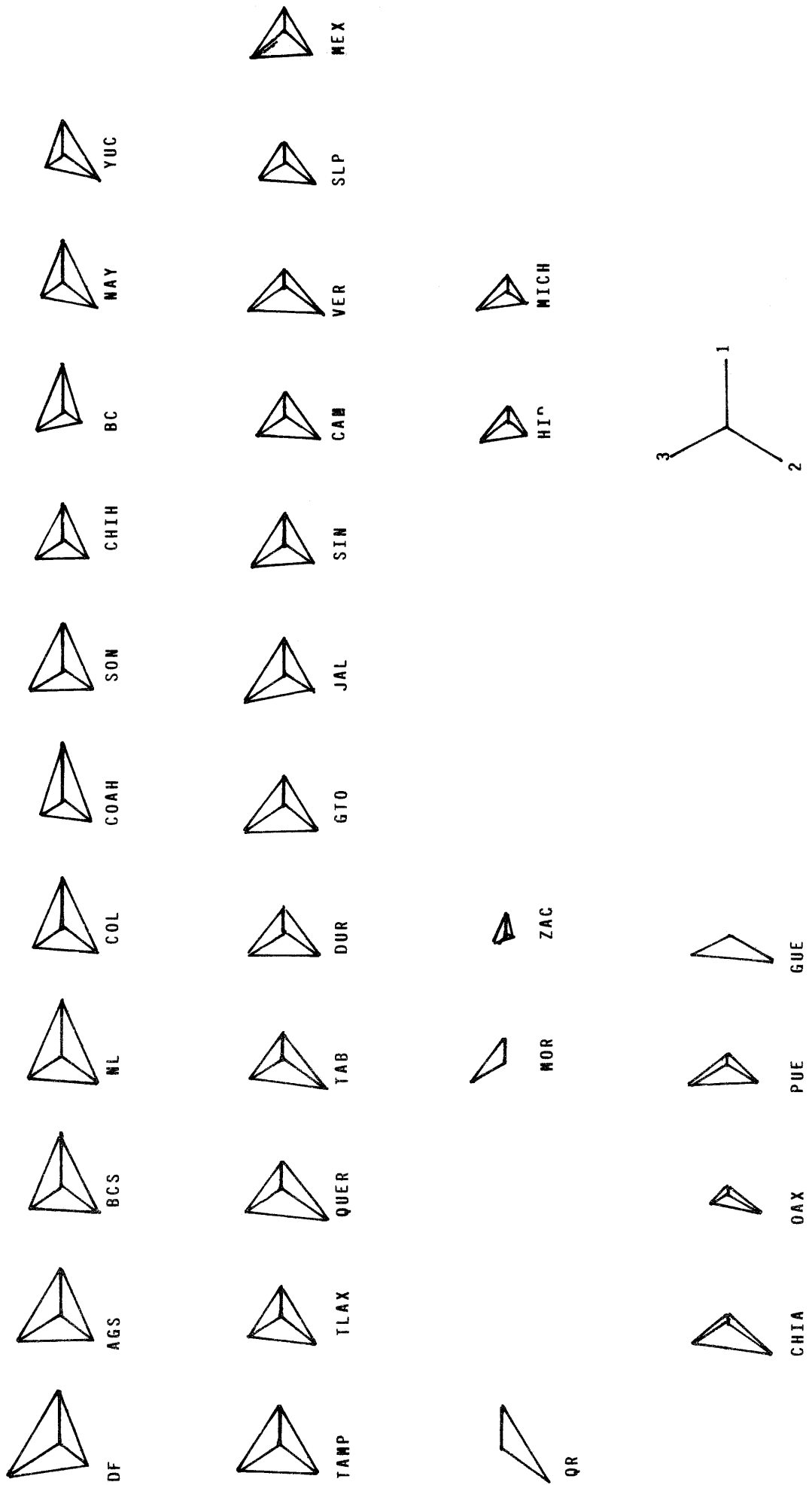
- el grupo 6, formado por Quintana Roo, presenta valores de demanda comparables con los observados en el grupo 1 y valores de IC2, más altos que los del grupo 1. Los valores de IC1 son tan grandes, que esta entidad se diferencia significativamente de las entidades del grupo 1.

En la figura 3, se presentan las gráficas estrella, que en este caso tienen forma triangular. Con base en los tres indicadores compuestos obtenidos se construyen las figuras correspondientes a las 32 entidades federativas.

Puede observarse en estas representaciones, que las entidades ubicadas en los dos renglones superiores difieren del resto de las entidades. Este hecho ilustra la polarización de la cobertura de los servicios de salud.

Además, debe observarse que las representaciones ubicadas en los dos renglones inferiores difieren entre sí, pudiéndose agrupar en cuatro regiones: una formada por Quintana Roo, otra

FIGURA 3.- FIGURAS TRIANGULARES CON BASE EN LOS INDICADORES COMPUESTOS DE COBERTURA DE APS.



por Morelos y Zacatecas, un tercer grupo por Hidalgo y Michoacán y el cuarto formado por Chiapas, Guerrero, Oaxaca y Puebla.

Las agrupaciones que podrían formarse entre las entidades de los dos primeros renglones de la figura 3, no resultan evidentes. Utilizando la regionalización obtenida en Valencia (1989) mediante Análisis Cluster respecto cobertura de APS, es posible agrupar a las entidades de estos dos primeros renglones.

Los grupos que se obtuvieron son los siguientes:

REGION 1: Aguascalientes, Baja California, Baja California Sur, Coahuila, Colima, Chihuahua, Distrito Federal, Nayarit, Nuevo León, Sonora y Yucatán.

REGION 2: Campeche, Durango, Guanajuato, Jalisco, México, Queretaro, San Luis Potosí, Sinaloa, Tabasco, Tamaulipas, Tlaxcala y Veracruz.

REGION 3: Chiapas, Guerrero, Oaxaca y Puebla.

REGION 4: Hidalgo y Michoacán.

REGION 5: Morelos y Zacatecas.

REGION 6: Quintana Roo.

Debe observarse en la figura 3 que Quintana Roo presenta los valores más bajos de IC3. Esto muestra que en esta entidad hay un desbalance entre la prevención por medio de inmunizaciones y el acceso a los recursos de higiene y salud. Quintana Roo, merece por las características presentadas una mayor atención.

9.- COMENTARIOS Y CONCLUSIONES.

El análisis de Componentes Principales ha probado ser desde sus primeras aplicaciones uno de los instrumentos estadísticos multivariados de mayor utilidad para el análisis de datos.

En ocasiones se utiliza previamente a la aplicación de otros procedimientos estadísticos, buscando en estos casos, simplificar el manejo y posiblemente la interpretación de la información.

Se presentaron tres nuevos indicadores de APS. Estos indicadores permiten una descripción simple del estado de

cobertura que guarda cada una de las entidades federativas de la República Mexicana.

La aplicación de CP sobre los datos de cobertura de servicios de salud ha permitido construir IC que permiten una explicación y comprensión más clara y sencilla del problema.

La utilidad de los IC de la cobertura de APS, consiste en representar la información proporcionada por los trece indicadores mediante tres indicadores, sin gran pérdida de información. Estos indicadores explican más del 82% de la variación presentada por los trece indicadores simples.

Los tres IC de la cobertura de APS son:

- Demanda de Atención Médica Institucional,
- Acceso a Instalaciones para Diagnóstico y Tratamiento de Enfermedades Endémicas y
- Comparación entre Prevención por Inmunización y Acceso a Recursos de Higiene y Salud.

El indicador PROM (OMS, 1985) y La Demanda de Atención Médica Institucional (IC1), se correlacionan linealmente de manera perfecta. Esta correlación muestra que la cobertura de la APS se representa adecuadamente mediante IC1 y que será equivalente utilizar a PROM o a IC1 para realizar cualquier inferencia o descripción de la cobertura de APS.

10.- BIBLIOGRAFIA.

- Hair, J.F., Anderson, R.E. y Tatham, R.L. (1987): Multivariate Data Analysis. With Readings. New York: Macmillan Publishing Co.
- Jolliffe, I.T. (1986): Principal Components Analysis. New York: Springer-Verlang.
- Kumate, J. y Barnard, A. (1988): La Atención de Cobertura de Los Servicios de Salud. II. Cobertura por Servicios de Atención Médica Primaria en México. Gaceta Médica de México, V. 124, 166-175.

- OMS (1978): Atención Primaria de Salud. Informe de La Comisión Internacional sobre Atención Primaria de Salud. Alma-Ata, URSS, 6-12 sep. 1978. Ginebra: OMS.
- OMS (1981): Development of Indicators for Monitoring Progress Towards Health for ALL by the Year 2000. Geneve: OMS HFA Series No. 4.
- OMS (1985): Development of Composite Measures of Coverage With Primary Health Care. Geneve: OMS Documento 854R, CMA/gd.
- Valencia, G. (1989): Regionalización de La Atención Primaria a La Salud en México. Análisis Estadístico. Presentado en el XXII Congreso Nacional de Matemáticas. Puebla, Pue. Noviembre 1989.

DIRECTORIO DE PARTICIPANTES

ABARCA ARENAS LUIS GERARDO

CLAVE 2 2 1

ALCOLEA HERRERA GABRIEL

PLANEAC. Y CAPACITACION, INEGI
HEROES DE NACUZARI SUR 2301
CD. INDUSTRIAL
20290 AGS., AGS.
TEL. 81280

HEROES DE NACUZARI SUR 1905-308
JARDINES DE LA ASUNCION
AGS. AGS.

CLAVE 2 1 2

ALMANZA PEREZ JOSE LUIS

UDIEEI, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 53398

JUAN SOTO 9
91500 COATEPEC, VERACRUZ

CLAVE 1 2 2

ARANDA ORDAZ FRANCISCO J.

IIMAS, UNAM
EDIFICIO IIMAS
CD. UNIVERSITARIA
A.P. 20-726
1000 MEXICO, D.F.
TEL. 5505536

HURHURTAN 87
H. DE PADIERNA
14200 MEXICO, D.F.
TEL. 6526134

CLAVE 2 1 1

ARROYO TORAL HEDILBERTO

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

XICOTENCATL 14
CENTRO
91000 XALAPA, VERACRUZ
TEL. 58573

CLAVE 1 2 2

BAEZ BARRADAS MA. DEL ROSARIO

UDIEEI, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 53398

ALLENDE 95
CENTRO
91000 XALAPA, VERACRUZ
TEL. 70747

CLAVE 1 2 2

BECERRIL CHAVEZ M. SAMUEL

UDIEEI, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

ALTAMIRANO 145
91000 XALAPA, VERACRUZ
TEL. 79251

CLAVE 1 2 2

BEHAR GUTIERREZ ROBERTO

UNIV. DEL VALLE CALI COLOMBIA
APARTADO AEREO 25360
CALI COLOMBIA
TEL. 398462
CALLE 12-B 59-31
CALI COLOMBIA
TEL. 301679

CLAVE 2 2 1

BENAVIDES ZAVALA ALEJANDRO

IIMAS, UNAM
MEXICO, D.F.

CFE 14, CASA 96
ELECTRA
34060 TLALNEPANTLA, MEXICO
TEL. 3610837

CLAVE 1 2 2

BORREGO A. SALVADOR A.

SABA CONSULTORES
1a. AVENIDA 112-1
CUMBRES
64610 MONTERREY, NUEVO LEON
TEL. 466288

EL DORIAN 4950
VILLA MITRAS
MONTERREY, NUEVO LEON
TEL. 703380

CLAVE 2 1 1

BRAVO SEGOVIA JESUS

IIMAS, UNAM
EDIFICIO IIMAS
CD. UNIVERSITARIA
A.P. 20726
1000 MEXICO, D.F.
TEL. 5505586

CAÑAMO 22
PALMAS
A.P. 18
1110 MEXICO, D.F.
TEL. 2717428

CLAVE 2 2 2

BURGUETE H. JOSE FRANCISCO

CEC, COLEGIO DE POST.
CHAPINGO, MEXICO
TEL. 42200 EXT. 5762

OROSCO 27
U. E. ZAPATA
CHAPINGO, MEXICO

CLAVE 2 1 1

BUSTOS Y DE LA TIJERA VICTOR A.

IIMAS, UNAM

VIADUCTO TLAPAN INT. 4
14620 MEXICO, D.F.
TEL. 5735749

CLAVE 2 1 1

CAMACHO CASTILLO OSVALDO

CEC, COLEGIO DE POST.
TEL. (91-595) 42200 EXT 5104

CANAL MARTINEZ MARGARITA E.
5 DE FEBRERO 68
CENTRO
91000 XALAPA, VERACRUZ
TEL. 86174

CLAVE 1 2 2

CARRILLO GAMBOA OLIVIA

SABA CONSULTORES
1a. AVENIDA 112-1
CUMBRES
64610 MONTERREY, NUEVO LEON
TEL. 466288

COCOTERO 5249
VALLE VERDE
64360 MONTERREY, NUEVO LEON
TEL. 703332

CLAVE 2 2 1

CASTILLO MORALES ALBERTO

CEC., COLEGIO DE POST.
CHAPINGO, MEXICO

JUAREZ 230-4
14000 MEXICO, D.F.
TEL. 5738236

CLAVE 2 1 1

CASTRO GONZALEZ ROBERTO G

FACULTAD DE BIOLOGIA, U.v.
ZONA UNIVERSITARIA
91000 XALAPA, VERACRUZ
TEL. 79202

RIO JAMAPA 26
CUAHUTEMOC
91060 XALAPA, VERACRUZ

CLAVE 2 2 2

CASTRO LOPEZ CLAUDIO R.

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

CASUORINA A 502
JARDINES DE XALAPA
91000 XALAPA, VERACRUZ

CLAVE 2 1 2

COLORADO HERNANDEZ JOSE LUIS

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
CENTRO
91000 XALAPA, VERACRUZ
TEL. 50263

RETORNO HIGO 18
INFONAVIT SUMIDERO
XALAPA, VERACRUZ

CLAVE 2 2 1

CORDOVA CONTRERAS ARTURO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ

AV. LAS AMERICAS 157
91030 XALAPA, VERACRUZ

CLAVE 1 2 2

CORTAZAR MARTINEZ ERNESTO

FACPYA, UNIV. AUTONOMA DE N. L.
CD. UNIVERSITARIA
MONTERREY, NUEVO LEON
TEL. 768500

VIVEROS 2572
ALTA VISTA INVERNADERO
MONTERREY, NUEVO LEON
TEL. 590045

CLAVE 2 2 2

CORTES A. ALMA ROSA

UAM-A
CALZADA DEL HUESO 1100
VILLA QUIETUD
MEXICO, D.F.

NICOLAS LEON 22
JARDIN BALBUENA
15900 MEXICO, D.F.
TEL. 5529391

CLAVE 2 2 1

CORTINA CARMONA AGUSTIN

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

JUAN DE LA BARRERA 22
CENTRO
91270 PEROTE, VERACRUZ
TEL. 50874

CLAVE 1 2 2

CORTINA CARMONA CLARA H.

FAC. DE ESTADISTICA, U.V.

JUAN DE LA BARRERA NORTE 22
CENTRO
91270 PEROTE, VERACRUZ
TEL. 50874

CLAVE 1 2 1

COVARRUBIAS AGUIRRE PATRICIA

AEROBIAS DE MEXICO
REFORMA 445
TORRE B TERCER PISO 6500
MEXICO, D.F.
TEL. 5254321

AV. DE LA TEJA A 102
VILLA COAPA
14390 MEXICO, D.F.
TEL. 5943878

CLAVE 2 1 2

CRUZ KURI LUIS

INSTITUTO DE CIENCIAS B., U.V.

CLAVE 2 1 1

CRUZ MAVIL MIGUEL ANGEL

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

CORREGIDORA 8
91000 XALAPA, VERACRUZ
TEL. 84714

CLAVE 2 2 2

CUEVAS CUEVAS CRISTINA

CIA. PAPELERA MALDONADO
AV. DE LA JUVENTUD 280
SAN NICOLAS, N.L.
TEL. 769179

LEOPOLDO KIEL 40
RAFAEL LUCIO
91110 XALAPA, VERACRUZ
TEL. 41786

CLAVE 1 2 2

CASANOVA DEL ANGEL FRANCISCO

IPN
ACUEDUCTO DE GUADALUPE 125
ACUEDUCTO DE GUADALUPE
7270 MEXICO, D.F.
TEL. 3912779
MEXICO, D.F.

CLAVE 2 1 1

DE LA VARA SALAZAR ROMAN

CLAVE 1 1 1

DEL CALLEJO ANGELES ENRIQUE

CONSULTOR PRIVADO
5 DE FEBRERO 68
CENTRO
91000 XALAPA, VERACRUZ

CLAVE 1 1 1

DIAZ AGUILAR BLANCA LUZ

CEC, COLEGIO DE POST.
56230 TEXCOCO, MEXICO
TEL. 42200 EXT. 5715

CIRCUITO PRYWER 221
TULIPANES
56230 TEXCOCO, MEXICO
TEL. 44803

CLAVE 2 1 1

DIAZ AVALOS CARLOS

CLAVE 1 2 2

DIAZ CAMACHO JULIAN FELIPE

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

Z. MERCADO 21
MISION SAN CARLOS
91170 XALAPA, VERACRUZ
TEL. 53435

CLAVE 2 2 2

DIAZ FRANCES MURGUIA ELOISA

IIMAS, UNAM
UXMAL 39
NARVARTE
3020 MEXICO, D.F.
TEL. 5196138

CLAVE 1 1 1

DIAZ GARCIA JOSE ANTONIO

UACAN
BUENA VISTA
SAL. COAHUILA
TEL. 43100

BOSQUES DE ARANJUEZ 2225
OLMO
25280 SALTILLO, COAHUILA
TEL. 54919

CLAVE 2 2 2

DIAZ OLIVEROS SALVADOR

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

BASURTO 27
CENTRO
91000 XALAPA, VERACRUZ
TEL. 82883

CLAVE 1 2 2

DIAZ TORRES CONSUELO

UAM-IZT.
AV. MICHOACANA Y PURISIMA
VICENTINA
MEXICO, D.F.
TEL. 6860322

CALLE 1501 64
UNIDAD ARAGON
A.P. 14
7920 MEXICO, D.F.
TEL. 7942134

CLAVE 2 2 2

DIAZ VIVEROS CAMERINA

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

ANTONIO PEREZ RIVERA 41
MIRADOR
XALAPA, VERACRUZ
TEL. 53822

CLAVE 1 2 2

DOMINGUEZ DOMINGUEZ LAURA EUGENIA

CLAVE 1 2 2

DUCOING WATTY ADRIANA M.

IIMAS, UNAM
NUNKINI 110
JARDINES DEL AJUSCO
14200 MEXICO, D.F.

CLAVE 2 1 1

ELIAS VILLA JOSE

GOB. DEL EDO.
SEC. DES. ECON.
AVILA CAMACHO 191
CENTRO
91000 XALAPA, VERACRUZ
TEL. 87099 EXT. 117

ANDADOR PITAYA 1/
INFONAVIT SUMIDERO
91000 XALAPA, VERACRUZ
TEL. 71993

CLAVE 2 2 2

ESCALANTE SALAS HORACIO

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

FIDENCIO OCANA 68-1
FCO. FERRER GUARDIA
XALAPA, VERACRUZ

CLAVE 1 2 2

ESLAVA GOMEZ GUILLERMINA

IIMAS, UNAM

GUADALUPE VICTORIA 4
SANTO TOMAS AJUSCO TLAPAN
14710 MEXICO, D.F.
TEL. 8460227

CLAVE 2 1 1

FABILA GARCIA MA, PATRICIA

IPN
AV. INSTITUTO POLITECNICO NAL,
MEXICO, D.F.

ANTONIO ANCONA 55
CONSTITUCCION
A.P. 9260
1917 MEXICO, D.F.
TEL. 6913402

CLAVE 1 2 2

FELIX MEDINA MARTIN HUMBERTO

ESC. DE CIENCIAS FISICO-MAT., U. A. S
CD. UNIVERSITARIA
CULIACAN, SINALOA

BENITO JUAREZ 1080 OTE.
LAS VEGAS
CULIACAN, SINALOA
TEL. 34683

CLAVE 1 2 2

FERNANDEZ ORTEGA LUIS RAUL

SABA CONSULTORES
1a. AVENIDA 112-1
CUMBRES
MONTERREY, NUEVO LEON
TEL. 466288

PROL. ALVARADO 5436
BRISAS
MONTERREY, NUEVO LEON

CLAVE 1 2 2

FLORES RAMIREZ FELICITAS

FACULTAD DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO

CEMPOALA 33
MODELO

91040 XALAPA, VERACRUZ

CLAVE 1 2 2

GARCIA BANDA AGUSTIN JAIME

INSTITUTO DE CIENCIAS B., U.V.
XALAPA, VERACRUZ

CLAVE 2 2 1

GARCIA PACHECO JOSE DE JESUS

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESO. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

EMILIANO ZAPATA 22
GUADALUPE RODRIGUEZ
90500 XALAPA, VERACRUZ
TEL. 72438

CLAVE 1 2 2

GARCIA ROMERO JAIME S.

COOR. DE INVESTIGACION C., UNAM
CIUDAD UNIVERSITARIA
COYOACAN
A.P. 20
45100 MEXICO, D.F.
TEL. 5505825

SAN RAFAEL 27
PUEBLO NUEVO (BAJO)
DEL. MAGDALENA CONTRERAS
A.P. 20
10640 MEXICO, D.F.
TEL. 6427312

CLAVE 2 2 2

GARCIA SOTELO JUAN CARLOS

INEGI AGS.
HEROES DE NACUZARI 2301
CD.INDUSTRIAL
489 AGS. AGS.
TEL. 83395

LA TINAJA 114
OJO CALIENTE
489 AGS. AGS.

CLAVE 2 2 2

GARCIA VALDEZ ROBERTO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

PERAL 27
INFONAVIT SUMIDERO
91150 XALAPA, VERACRUZ
TEL. 45890

CLAVE 1 2 2

GARCIA ZAMORA MARCELO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

JAZMIN 2
BENITO JUAREZ
XALAPA, VERACRUZ
TEL. 78245

CLAVE 1 2 2

GASCON VARELA MAURO ANTONIO

INEGI AGS.
HEROES DE NACUZARI 2301
CD. INDUSTRIAL
489 AGS. AGS.

TEL. 83395

MONTORO 229
OJO CALIENTE
489 AGS. AGS.

CLAVE 1 2 2

GOMEZ HERNANDEZ CELSO

UNAM
CD. UNIVERSITARIA
MEXICO, D.F.

VICENTE GUERRERO 1
PRADERAS SAN MATEO
MEXICO, D.F.

CLAVE 2 2 2

GONZALEZ COLORADO FERNANDO ULISES

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

CLAVE 2 2 2

GONZALEZ DOMINGUEZ JESUS

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

AV. XALAPA 164-BIS ALTOS A
XALAPA, VERACRUZ

CLAVE 1 2 2

GUTIERREZ ALBA JUSTINO

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

GUANAJUATO 13
PROGRESO
91130 XALAPA, VERACRUZ
TEL. 53514

CLAVE 1 2 2

GUTIERREZ PEÑA EDUARDO

IIMAS, UNAM
MACEDONIA
00018 MEXICO, D.F.

LOMAS ESTRELLA
09890 MEXICO, D.F.
TEL. 6563571

CLAVE 2 2 1

GUTIERREZ PULIDO HUMBERTO

UNIV. DE GUANAJUATO

CLAVE 1 1 1

GUZMAN TOLEDO BEATRIZ

DIAG. SAN ANTONIO 1935
NARVARTE
A.P. 12
3020 MEXICO, D.F.
TEL. 5388634

CLAVE 1 2 2

HERNANDEZ GONZALEZ SERGIO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AV. CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

ISMAEL CRISTEN 26-B
RAFAEL LUCIO
91110 XALAPA, VERACRUZ
TEL. 57595

CLAVE 2 2 2

HERNANDEZ HERNANDEZ MATILDE

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

NUEVO LEON 204

PROGRESO
91130 XALAPA, VERACRUZ
TEL. 41725

CLAVE 1 2 1

HERNANDEZ OSORIO CRESCENCIO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESO. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

PRINCIPAL 39-D
BALCONES DE XALAPA
XALAPA, VERACRUZ

CLAVE 1 2 1

HERNANDEZ RIVERA PROSPERO

ENRIQUE ETA MERCADO 95
JUAN DE LA LUZ ENRIOUEZ
91160 XALAPA, VERACRUZ
TEL. 59827

CLAVE 1 1 2

IÑIGUEZ VALADEZ MA. ANTONIA

ESC. ECONOMIA, UNIV. DE NAYARIT
TEPIC NAYARIT

CLAVE 2 2 2

64160 MONTERREY, NUEVO LEON
TEL. 466288

CD. LIMON 121
MITRAS NORTE
MONTERREY, NUEVO LEON

CLAVE 1 1 1

ISLAS CAMARGO ALEJANDRO

UAM-I
AV. MICHOACAN Y PURISIMA
VICENTINA
MEXICO, D.F.
TEL. 6860322

RICARTE 427
LINDAVISTA
07300 MEXICO, D.F.

CLAVE 2 2 2

LANDOIS PALENCIA LUIS LEON

CEC, COLEGIO DE POST.

5 DE DICIEMBRE 27
LA PAZ SAN SEBASTIAN
56170 TEXCOCO, MEXICO
TEL. 46630

CLAVE 2 1 1

JACOME AVILA NANCY

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

27 DE SEPTIEMBRE 2
ELECTRICISTAS
91060 XALAPA, VERACRUZ
TEL. 79619

CLAVE 2 2 1

LARA PEREZ SOTO CLAUDIA

IIMAS, UNAM
SUR 114
A.P. 59
MEXICO, D.F.

COVE 1120
MEXICO, D.F.
TEL. 5159777

CLAVE 1 2 2

JUAREZ CERRILLO SERGIO FCO.

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

CLAVE 1 2 1

LOPEZ ESTRADA JESUS

FAC. DE CIENCIAS, UNAM
CIRCUITO EXTERIOR
CD. UNIVERSITARIA
04510 MEXICO, D.F.
TEL. 5505215 EXT. 3909

CALLE CENTRAL EDIFICIO 7-33
ALIANZA POPULAR REV.
04800 MEXICO, D.F.
TEL. 6774517

CLAVE 2 2 1

LANDA REYES PIEDAD

SABA CONSULTORES
1a. AVENIDA 112-1
CUMBRES

LOPEZ ORTIZ MIRNA ACELA

SCT
CARRETERA XALAPA-VERACRUZ
ANIMAS
XALAPA, VERACRUZ
TEL. 85677 EXT. 53

AV. ADOLFO RUIZ CORTINEZ 74
BELISARIO DOMINGUEZ
91050 XALAPA, VERACRUZ
TEL. 83220

CLAVE 2 2 2

LOPEZ RIVERA ENRIOUE

FAC. CONT. PUB. Y ADMON., UANL
CD. UNIVERSITARIA
S. NICOLAS DE G., NUEVO LEON
TEL. 768500

AMACUZAC 214
RESIDENCIAL ANCIHUAC
66450 S. NICOLAS DE G., NUEVO LEON
TEL. 766955

CLAVE 2 2 2

LUNA RAMIREZ ENRIOUE

FAC. MATEMATICAS, UNIV. DE GTO.
CALLEJON JALISCO
A.P. 402
36000 GTO., GTO.
TEL. 27175

GUALDRA 18
36010 GTO., GTO.
TEL. 21500

CLAVE 1 2 1

MAHE BACA MOISES

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ

TEL. 50263

AV. AMERICAS 157
91030 XALAPA, VERACRUZ
TEL. 84820

CLAVE 1 2 1

MALDONADO MORALES LETICIA

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

ANDADOR EUCALIPTO 17
UNIDAD FERROCARRILERA
91120 XALAPA, VERACRUZ
TEL. 41840

CLAVE 1 2 2

MARQUEZ AMARO MIGUEL ANGEL

UDIEEI, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

AV. AMERICAS 243-BIS
CARDEL
91030 XALAPA, VERACRUZ
TEL. 81058

CLAVE 1 2 2

MARTINEZ ALCANTARA ANTONIO

CEC, COLEGIO DE POST.

RETORNO 4 DE SUR 12C 57
AGRICOLA ORIENTAL
8500 MEXICO, D.F.
TEL. 7586371

CLAVE 2 1 1

MARTINEZ GARZA ANGEL

CEC, COLEGIO DE POST.
MONTECILLO
CHAPINGO, MEXICO
TEL. 42200 AREA 595

CONSTITUCION LOCAL 510
FRACC. SAN MATEO
56110 TEXCOCO, MEXICO
TEL. 40946

CLAVE 2 1 1

MARTINEZ RAMOS ALEJO A.

ALDAMA 12
CENTRO
91000 XALAPA, VERACRUZ

CLAVE 1 2 2

MARTINEZ RAMOS NARDA A.

ALDAMA 12
CENTRO
91000 XALAPA, VERACRUZ

CLAVE 1 2 2

MAVIL PENSADO JAVIER

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

16 DE SEPTIEMBRE 65-A
MIGUEL HIDALGO
XALAPA, VERACRUZ
TEL. 58681

CLAVE 1 2 2

MENDEZ RAMIREZ IGNACIO

IIMAS, UNAM
CIRCUITO INT. CD. UNIV.
01000 MEXICO, D.F.
TEL. 5500047

MESINA 19
LOMAS ESTRELLA
09890 IXTALAPA, D.F.
TEL. 6564171

CLAVE 2 1 1

MENDEZ SANCHEZ VICTOR M.

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

DELFINO VICTORIA 16
XALAPA, VERACRUZ
TEL 56761

CLAVE 2 1 2

MENDOZA DOMINGUEZ OLIVIA

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 53398

5 DE MAYO 57
91500 COATEPEC, VERACRUZ
TEL. 61136

CLAVE 1 2 2

MENDOZA RAMIREZ MANUEL

FAC. DE CIENCIAS, UNAM
CIRCUITO EXTERIOR
CD. UNIVERSITARIA
04510 MEXICO, D.F.
TEL. 5505215 EXT. 3909

DON MANUELITO 55-B 108
OLIV. DE LOS PADRES
01780 MEXICO, D.F.
TEL. 5952315

CLAVE 2 2 1

MENESES A. DE SESMA BEATRIZ

CLAVE 2 2 2

MIRANDA ABURTO BENJAMIN

CLAVE 1 2 1

MOLET GEORGINA

FAC. DE CIENCIAS, UNAM
CD. UNIVERSITARIA
MEXICO, D.F.
TEL. 5505215

FRATERNIDAD 79 NOGAL A-2
SAN ANGEL
01090 MEXICO, D.F.
TEL. 5485560

CLAVE 1 2 1

MONTANO RIVAS JULIA AURORA

FACULTAD DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
CENTRO
91000 XALAPA, VERACRUZ
TEL. 50263

INDEPENDENCIA 79
ESTANZUELA, VERACRUZ

CLAVE 2 2 1

MONTES DE OCA GARCIA FRANCISCO

CLAVE 2 2 1

MORA ACOSTA MARISSA

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

CARLOS MIGUEL PALACIOS 30
91000 XALAPA, VERACRUZ
TEL. 77252

CLAVE 1 2 2

MORALES RIVAS ANA LUISA

SABA CONSULTORES
1a. AVENIDA 112-1
CUMBRES
64610 MONTERREY, NUEVO LEON

CD. LIMON 121
MITRAS-NORTE
MONTERREY, NUEVO LEON

CLAVE 1 2 2

MORALES RODRIGUEZ RENE

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

DOM. CON. EL AGUAJE
MNPIO. DOS RIOS, VERACRUZ

CLAVE 1 2 2

MORALES SANTIAGO JULIAN

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

RAFAEL RAMIREZ 49
UNIV. VERACRUZANA
91000 XALAPA, VERACRUZ
TEL. 76634

CLAVE 1 2 2

MUÑOZ NAVARRO JUAN

UDIEEI, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 53398

CIRCUITO AGUILA 114.
FRACC. LAS ANIMAS
9100 XALAPA, VERACRUZ
TEL. 85950

CLAVE 2 2 2

NAKAMURA SAVOY MIGUEL

CEIM
A.P. 402
36000 GTO., GTO.
TEL. 4732715

LOMAS CHAPARROSA 3
MARFIL
36250 GTO., GTO.

CLAVE 2 1 1

NARANJO CORTES ADA ISABEL

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA VERACRUZ
TEL. 50263

CARLOS M. PALACIOS 32
VENUSTIANO CARRANZA
XALAPA, VERACRUZ

CLAVE 1 2 2

NITA PEREZ JOSE ARTURO

DPTO. SALUD PUBLICA, U.V.
ERNESTO ORTIZ MEDINA 3
91000 XALAPA, VERACRUZ

INSURGENTES 106-3
91000 XALAPA, VERACRUZ

CLAVE 2 2 2

OCHOA C. OCTAVIO A.

UNIDAD CENTRAL, U.V.
LOMAS DEL ESTADIO
91000 XALAPA, VERACRUZ
TEL. 85219

PRIVADA DE ACUEDUCTO 10
UNIDAD MAGISTERIAL
XALAPA, VERACRUZ
TEL. 43040

CLAVE 2 2 2

OCHOA ZUMAYA ESTEBAN

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

MARTIREZ DE XALAPA 54
CARRILLO PUERTO
A.P. 910
91081 XALAPA, VERACRUZ
TEL. 80743

CLAVE 1 2 2

OJEDA RAMIREZ MARIO MIGUEL

LINAE, U.V.
AV.XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

EDIFICIO EBANO "B" 401
JARDINES DE XALAPA
91120 XALAPA, VERACRUZ

CLAVE 2 1 1

OLIVERA TORO MAYA ALEJANDRO M.

UACPYP/CCH, UNAM
CIRCUITO INT. CD. UNIV.
MEXICO, D.F.

UXMAL 718
LETRAN VALLE 3650
MEXICO, D.F.
TEL. 6049908

CLAVE 1 2 2

O'REILLY FEDERICO

IIMAS, UNAM

CERRO DE LA MANO 60
ROMERO TERREROS
COYOACAN
04320 MEXICO, D.F.
TEL. 5546063

CLAVE 2 1 1

ORTEGA ROMERO ALFREDO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 53398

OAXACA
PROGRESO MACUILTEPELT
XALAPA, VERACRUZ

CLAVE 1 2 2

ORTIZ ANDRADE IRMA

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

AV. MARTIRES 28 AGTO 133
FERRER GUARDIA
XALAPA, VERACRUZ
TEL. 43719

CLAVE 1 2 2

ORTIZ LEON MARIA CRISTINA

DEPTO. DE SALUD PUBLICA, U.V.
ERNESTO ORTIZ MEDINA
CENTRO
91000 XALAPA, VERACRUZ
TEL. 57404

ZARAGOZA 17
CENTRO
COATEPEC, VERACRUZ

CLAVE 2 2 1

ORTIZ RODRIGUEZ JOSE

ADALBERTO LARA HERNANDEZ 53
AGUACATAL
91130 XALAPA, VERACRUZ
TEL. 42128

CLAVE 2 2 2

ORTIZ SEDANO ADOLFO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

13 DE SEPTIEMBRE 66
91090 XALAPA, VERACRUZ
TEL. 79600

CLAVE 2 2 2

OSEGUERA CERVANTES ISAIAS

LINAE, U.V.
AV. XALAPA ESQ. A. CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

LA PAZ 55
CARRILLO PUERTO
91080 XALAPA, VERACRUZ

CLAVE 2 1 2

OSPINA VARON BEATRIZ

IIMAS, UNAM

PETEN 652 DEPTO. 3
VERTIZ NARVARTE
03600 MEXICO, D.F.
TEL. 6047650

CLAVE 1 2 2

PADRON CORRAL HECTOR

UNIVERSIDAD DE NUEVO LEON
CD. UNIVERSITARIA
MONTERREY, NUEVO LEON
TEL. 7768500

CERBA 4763

RESIDENCIAL DE CEDROS
MONTERREY, NUEVO LEON
TEL. 768500

CLAVE 2 2 2

PADRON HERNANDEZ ARNOLDO

DPTO. DE SUELOS

CLAVE 2 1 2

PARRA LOPEZ SAUL

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

ANDADOR 3 10
UNIDAD POMONA
91040 XALAPA, VERACRUZ
TEL. 84476

CLAVE 1 2 2

PEREZ ABREU CARRION RAFAEL

TELEFONOS DE MEXICO S.A. DE C.V.

CLAVE 1 1 1

PEREZ-ABREU VICTOR

CEIM
ARCO DE GUADALUPE 420
MARFIL
36000 GTO., GTO.

CLAVE 2 1 1

PEREZ FLORES ARMANDO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

CAYETANO RODRIGUEZ B. 61
CENTRO
91000 XALAPA, VERACRUZ
TEL. 82664

CLAVE 2 2 2

PEREZ PORTUGAL ANDRES

INSTITUTO DE CIENCIAS B., U.V.

ENRIQUE GUERRA 3-A
JOSE CARDEL
91030 XALAPA, VERACRUZ
TEL. 71455

CLAVE 1 2 1

PUCHETA HERNANDEZ JOSE LUIS

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
RAFAEL HERNANDEZ OCHOA 1
MARIA ENRIQUETA
91500 COATEPEC, VERACRUZ

CLAVE 1 2 2

PUERTAS TELLEZ LETICIA

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

20 DE NOVIEMBRE OTE. 285-9
CENTRO
91000 XALAPA, VERACRUZ

CLAVE 1 2 2

PULIDO RIOS RICARDO

SABA CONSULTORES
1a. AVENIDA 112-1
CUMBRES
64610 MONTERREY, NUEVO LEON
TEL. 466288

JIMENEZ 306 OTE.
GPE., NUEVO LEON
TEL. 374858

CLAVE 2 2 1

QUIROZ MENDOZA FERNANDO

FACULTAD DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ

CLAVE 1 2 2

RAMIREZ PEREZ FILEMON

U. A. CH.
CHAPINGO, EDO. DE MEXICO
TEL. 42200 EXT. 5345

DEPTO. DE FITOTECNIA
56230 TEXCOCO, EDO. DE MEXICO

CLAVE 2 2 2

RAMIREZ VALVERDE BENITO

CEICADAR, COLEGIO DE POST.
CARR. FED. MEXICO-PUEBLA KM. 125.5
A.P. I-12
PUEBLA, PUE.
TEL. 480088

2a PRIV. PLAN DE GPE. 4506 DPTO. 5
SAN BALTAZAR
PUEBLA, PUE.
TEL. 375591

CLAVE 2 1 1

RAMIREZ VALVERDE GUSTAVO

CEC, COLEGIO DE POST.

CLAVE 2 1 1

REYES GONZALEZ JOSE ALBERTO

CLAVE 1 - 1

RINCON SOLIS LUIS ANTONIO

IIMAS, UNAM
EDIF. 5 DPTO. 302
MEXICO, D.F.

GUELATAO DE JUAREZ II
09221 MEXICO, D.F.
TEL. 7637032

CLAVE 1 2 2

RIVERA RENDON MA. DEL REFUGIO

16 DE SEPTIEMBRE 11
LA CRUZ
08310 MEXICO, D.F.
TEL. 6572906

CLAVE 1 2 2

RODRIGUEZ CARVAJAL LUIS A.

UNIVERSIDAD AUT. DE YUCATAN
MERIDA, YUCATAN
YUCATAN 60 491-A X 57
CENTRO
97000 YUCATAN
TEL. 248000

CLAVE 2 2 2

RODRIGUEZ MUÑOZ JOSE ELIAS

CEIM
TEPETAPA 53
36000 GTO., GTO.
TEL. 27155

CLAVE 2 1 1

ROJAS FIGUEROA MIGUEL ANGEL

CLAVE 2 2 2

ROMERO ANGULO M. ARACELI

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

VICTORIA 131
CENTRO
91000 XALAPA, VERACRUZ
TEL. 82736

CLAVE 1 2 2

ROMERO MELGAREJO JUAN

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

FRANCISCO LAGOS CHAZARO 131
EL MIRADOR
91170 XALAPA, VERACRUZ
TEL. 42164

CLAVE 2 1 2

ROMERO RAMIREZ JESUS

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

EDIFICIO TUXPAN 501
FOVISSTE
XALAPA, VERACRUZ

CLAVE 2 1 2

ROMERO VADILLO ELEONORA

CLAVE 2 2 2

RUEDA DIAZ DEL CAMPO RAUL

IIMAS, UNAM
CD. UNIVERSITARIA
ALVARO OBREGON
A.P. 20276
01000 MEXICO, D.F.
TEL. 5505586

AMATLAN 159
CONDESA
06170 MEXICO, D.F.
TEL. 2116605

CLAVE 2 2 1

RUIZ CAMPOS SALVADOR

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

AV. XALAPA 232
PROGRESO
91130 XALAPA, VERACRUZ
TEL. 55648

CLAVE 1 2 1

RUIZ RAMIREZ JUAN

INIFAP, SARH
CARRETERA XALAPA-VERACRUZ KM. 4
XALAPA, VERACRUZ
TEL. 88212

INSURGENTES 142
CENTRO
91000 XALAPA, VERACRUZ
TEL. 81600

CLAVE 2 1 2

RUIZ VELASCO ACOSTA SILVIA

IIMAS. UNAM
3 RETORNO EPSILON 47 INT. 316
04310 MEXICO, D.F.

CLAVE 2 1 1

SAID INFANTE GIL

CEC, COLEGIO DE POST
CHAPINGO, MEXICO
TEL. 45999
LOTE 9
CIRCUITO PRYWER
SAN LUIS HUEXOTLA, MEXICO
TEL. 42909

CLAVE 2 1 1

SANDOVAL RODRIGUEZ ROGELIO

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

JESUS CARRANZA 11
CENTRO
91000 XALAPA, VERACRUZ
TEL. 77636

CLAVE 1 2 2

SANTIZO RINCON JOSE ANTONIO

CEC, COLEGIO DE POST.
A.P. 5504

DIVULGACION DEP.6 16
COL. GILBERTO PALACIO
56230 CHAPINGO, MEXICO

CLAVE 2 1 2

SAUCEDO ROCHA LUIS ROBERTO

SABA CONSULTORES
1a.AVENIDA 112-1
CUMBRES
64610 MONTERREY, NUEVO LEON
TEL. 466288

ARTEAGA 602
CONST. DE QUERETARO VI SECTOR
S.NICOLAS DE G., NUEVO LEON

CLAVE 1 2 2

SESMA Y MUÑOZ BISMARCK

CLAVE 2 2 2

SOLANA SANSORES RAFAEL

CINVESTAV
2699 22-A 85 RINCONADA DE CHUBURNA
97200 MERIDA, YUCATAN

CLAVE 2 1 1

SOLIS A. CARLOS GILBERTO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

DR. LUCIO 127
CENTRO
91000 XALAPA, VERACRUZ
TEL. 73417

CLAVE 1 2 2

SOSA GALINDO ISMAEL

INSTITUTO DE CIENCIAS B., U.V.
XALAPA, VERACRUZ

CLAVE 2 1 1

SOTO FERNANDEZ GUADALUPE

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

MARTIRES 28 DE AGOSTO 93
FERRER GUARDIA
91100 XALAPA, VERACRUZ
TEL. 79223

CLAVE 1 2 2

SOTRES RAMOS DAVID

CEC, COLEGIO DE POST.
56230 CHAPINGO, MEXICO
TEL. 6520823

SIERRA APANECA 43
JARDINES EN LA MONTAÑA
14210 MEXICO, D.F.
TEL. 6520823

CLAVE 2 1 1

SUARDIAZ BRIZZIO ENRIQUE E.

FAC. DE ECONOMIA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
CENTRO
91000 XALAPA, VERACRUZ
TEL. 50263

LUIS J. JIMENEZ 1-A
DEL MAESTRO
91030 XALAPA, VERACRUZ
TEL. 54539

CLAVE 2 1 2

SUARDIAZ SOLE PABLO E.

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

LUIS J. JIMENEZ 1-A
DEL MAESTRO
91030 XALAPA, VERACRUZ
TEL. 54539

CLAVE 2 1 1

TORRES HERNANDEZ MA. TERESA

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

BENITO FENTANES 3
AGUACATAL
XALAPA, VERACRUZ
TEL. 50006

CLAVE 1 2 2

TORREZ MEJIA PEDRO

FAC. DE ESTADISTICA, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
CENTRO
91000 XALAPA, VERACRUZ
TEL. 50263

CALIFORNIA 32
BELISARIO DOMINGUEZ
91030 XALAPA, VERACRUZ
TEL. 84235

CLAVE 2 2 2

VALERO PACHECO ELIZABETH

CINVESTAV
CARRETERA ANTIGUA PROGRESO
UNIDAD M
MERIDA, YUCATAN
TEL. 259831

CLAVE 1 2 2

VALLEJO GONZALEZ NOE M.

IIMAS, UNAM
CIRCUITO ESC. CD. UNIVERSITARIA
01000 MEXICO, D.F.
TEL. 5505215

DOCTOR GARCIA DIEGO 189
DOCTORES VERACRUZANOS

CLAVE 2 1 1

VAQUERA HUERTA HUMBERTO

CLAVE 2 1 1

VARGAS CHANES DELFINO

CIMAT
CALLEJON JALISCO
MINERAL DE VALENCIANA
A.P. 402
36000 GTO., GTO.
TEL. 27155

HDA. PANUCO 4
MARFIL
A.P. 402
36000 GTO., GTO.
TEL. 27155

CLAVE 2 1 1

VARGAS HERNANDEZ MATEO

U. A. DE CHAPINGO
CARRETERA MEXICO TEXCOCO
TEXCOCO, MEXICO
TEL. 42200

DALIA 100
LA CONCEPCION
56170 TEXCOCO, MEXICO

CLAVE 2 1 2

VARGAS MESTIZO GISELA

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
91000 XALAPA, VERACRUZ
TEL. 50263

AV. AMERICAS 400
2 DE ABRIL
91030 XALAPA, VERACRUZ

CLAVE 2 1 1

VEGA VALDIVIA DIXIA DANIA

CEC, COLEGIO DE POST.
CHAPINGO, MEXICO

CLAVE 1 2 2

VELASCO GONZALEZ CRUZ

SOCIOLOGIA RURAL, UACH
JUAREZ SUR 315
EL CARMEN
56130 TEXCOCO, MEXICO

CLAVE 2 1 1

VELAZQUEZ CAMACHO MICAELA

FAC. ODONTOLOGIA, UNAM
CD. UNIVERSITARIA
COYOACAN, DEL. BENITO JUAREZ

A.P. 20
45100 MEXICO, D.F.

TLACOQUEMECATL 70
DEL VALLE
A.P. 12
03100 MEXICO, D.F.
TEL. 5759316

CLAVE 1 2 2

VIESCA GONZALEZ F. CARLOS

PARASIT. AGR., U. A. CHAPINGO
CARRETERA MEX.-VER. KM. 38.5
56230 CHAPINGO, MEXICO
TEL. 42200 EXT. 5356

TOPOGRAFOS 241-A
MISIONES DE SNT.ESPERANZA
TOLUCA, MEXICO

CLAVE 2 1 2

VILLALBA COHEN RENATA

FAC. DE CIENCIAS, UNAM
CIUDAD UNIVERSITARIA
04510 MEXICO, D.F.
TEL. 5505215

AMATLAN 159 DEP. 3
COL. CONDESA
06170 MEXICO, D.F.
TEL. 5534795

CLAVE 2 2 1

ZARATE DE LARA GUILLERMO

CEC, COLEGIO DE POST.
CHAPINGO, MEXICO

CLAVE 2 1 1

ZARATE DE LARA OSCAR E.

CEC, COLEGIO DE POST.
56230 TEXCOCO, MEXICO
TEL. 42200 EXT. 5715

CIRCUITO PRYWER 221
TULIPANES
56230 TEXCOCO, MEXICO
TEL. 44803

CLAVE 2 1 1

ZAVALA ARREOLA OMAR

LINAE, U.V.
AV. XALAPA ESQ. AVILA CAMACHO
CENTRO
91000 XALAPA, VERACRUZ
TEL. 50263

JUSTINO SARMIENTO 170
HIDALGO
XALAPA, VERACRUZ

CLAVE 1 2 2

