Centro de Estudios Económicos
www.colmex.mx
El Colegio de México, A.C.

Serie documentos de trabajo

MODERNIZACIÓN Y EL MANTENIMIENTO DE LA BIODIVERSIDAD GENÉTICA EN EL CULTIVO DE MAÍZ EN MÉXICO

Antonio Yunez Naude y Fernando Barceinas

DOCUMENTO DE TRABAJO

Núm. VII - 1995

Modernización y el mantenimiento de la biodiversidad genética en el cultivo del maíz en México*

Antonio Yúnez-Naude, El Colegio de México
Fernando Barceinas P., UAM-Azcapotzalco
Agosto, 1995

* Agradecemos el apoyo económico de la Fundación Konrad Adenauer, las sugerencias metodológicas del Prof. J. E. Taylor, así como a nuestros auxiliares de investigación, Anabell Luna V. y Margarito Rodriguez H.

Uno de los componentes del estudio de la sustentabilidad agrícola de países que, como México, cuentan con riqueza genética en un contexto de heterogeneidad, es el análisis de las diferencias productivas entre las unidades que componen al sector, así como de las razones que las explican. En el caso del estudio del mantenimiento de la biodiversidad genética del maíz esto exige, entre otras labores, determinar la tendencia de algunas unidades productoras a recurrir más que otras al uso de semillas genéticamente ricas.

Para poner en práctica nuestra propuesta, no sólo es necesario contar con información detallada sobre las condiciones en las que obtienen maíz las distintas unidades productivas del agro mexicano, sino con una metodología que permita establecer, de una manera rigurosa, qué variables están relacionadas al uso de cierto tipo de semillas y su peso en la explicación del mismo. Una forma de hacerlo es mediante los análisis "factorial" y "de conglomerados".

En el presente trabajo presentamos los resultados que hemos obtenido al aplicar tales enfoques. Con ello pretendemos profundizar nuestro estudio sobre la relación entre el tipo de semilla usada en la producción del maíz con algunas características de los predios que lo producen, tales como el grado de capitalización y de vinculación con los mercados, el acceso al agua y el tipo de propiedad. Nuestro propósito es, pues, incluir los análisis factorial y de conglomerados en nuestras indagaciones sobre los determinantes del tipo de semillas usadas en la producción de maíz en México (véase a Yúnez-Naude, A., J. E. Taylor y F. Barceinas: 1994).

El artículo está compuesto por cinco secciones. En la primera exponemos brevemente las características de los datos y de los métodos utilizados. En la segunda describimos la manera en que seleccionamos las variables relevantes a nuestro estudio. La tercera contiene los resultados del análisis factorial y la cuarta los correspondientes al de conglomerados. El artículo concluye con la discusión sobre qué tanto el estudio apoya nuestra hipótesis de trabajo; es decir si, en efecto, existe una tendencia en el subconjunto privado, capitalizado y vinculado al mercado de los productores de maíz a usar semillas mejoradas (SM en adelante) y si, por el contrario, el subconjunto de productores de tipo familiar tiene mayor vocación al uso de semillas autóctonas o criollas (SC en lo que sigue). De ser esto así, tendríamos argumentos adicionales a los del artículo citado para proponer que la pretendida modernización del agro mexicano podría estar en conflicto con el mantenimiento de la biodiversidad genética del maíz.

I. La información y el método

La base de datos que usamos proviene de una encuesta nacional muy detallada sobre producción, costos, productividad y tipo de

semillas usadas por las unidades agrícolas que cultivaron maíz durante el ciclo primavera-verano de 1991, elaborada por BANRURAL, FIRA y la SARH.[1] La información proporcionada por ésta fuente es muy amplia: consta de 2,448 cuestionarios aplicados a predios maiceros en todo México y contiene más de 20 variables para cada uno de ellos. Asimismo, distingue a los productores según régimen de propiedad (ejidal o privado) y acceso al recurso pluvial (riego y temporal, en el cuadro 1 del Anexo Estadístico presentamos un listado de las principales variables de la Encuesta).

Una manera de incluir en el estudio la riqueza de la información proporcionada por la Encuesta y, al mismo tiempo, eliminar las variables con poco o ningún peso explicativo para nuestros objetivos, es usando el análisis factorial (AF en adelante).

El enfoque comúnmente usado para establecer las relaciones entre variables relevantes a un estudio específico es la aplicación de los métodos econométricos. Como se sabe, éste consiste en seleccionar variables dependientes e independientes --proponiendo formas funcionales que las relacionan-- y, posteriormente, estimar los coeficientes de las ecuaciones propuestas, realizar una serie de pruebas estadísticas sobre la pertinencia de los cálculos hechos e interpretar los resultados desde un punto de vista económico.

No obstante la fuerza que ésta metodología tiene para el análisis cuantitativo, es común que enfrente dos problemas: el primero tiene que ver con el hecho de que su uso requiere especificar a priori la ecuación (es decir, recurriendo a la teoría económica determinar de ante mano las variables explicativas adecuadamente y definir una forma funcional); el segundo se debe a que, en muchas ocasiones, se tiene una gran cantidad de variables que, potencialmente son explicativas, pero no se cuenta con una regla de discriminación.

La aplicación del análisis factorial (AF en adelante) es una manera de superar tales limitantes: mediante su uso se establece la relación entre un conjunto de variables observables o datos que son explicadas en términos de un número reducido de variables no observadas, denominadas factores. Además de eliminar las que, en un principio y a juicio del investigador se consideraban irrelevantes, éste método tiene la ventaja de no excluir a priori en el análisis cuantitativo a ninguna de ellas, así como a ningún tipo de relación entre las mismas.

[1] SARH, FIRA y BANRURAL, "Encuesta nacional de costos, coeficientes técnicos y producción agrícola", México, 1992 (en adelante nos referimos a ésta base de datos como Encuesta). Desafortunadamente, la encuesta sólo divide en dos el tipo de semillas usadas en la producción nacional de maíz: mejoradas y criollas y/o autóctonas. No obstante y debido a que el segundo tipo son las ricas en biodiversidad, la distinción nos permite hacer una aproximación al tema de nuestro interés. En Yúnez-Naude, et.al., op.cit., nota 1 y cuadros 1 a 5, presentamos algunos datos obtenidos de la Encuesta que describen el uso de semillas en la producción de maíz por tipo de predio.

Una vez determinadas las variables significativas que arroja el AF, procedimos a aplicar el enfoque conocido como análisis de conglomerados, mediante el cual se hacen agrupaciones a partir de las variables relevantes arrojadas por el AF, de acuerdo a sus características comunes. El procedimiento permite profundizar el estudio sobre qué tipo de unidades productoras de maíz son las más proclives al uso de SC y, en consecuencia, al mantenimiento de la biodiversidad genética del cultivo.[2]

II. Selección de las variables de la Encuesta

Como dijimos, la base de datos que usamos contiene una gran cantidad de información sobre la estructura de producción y costos de cada una de las unidades encuestadas, productoras de maíz en México y representativas de la situación nacional (en adelante nos referiremos a ellas como unidades productivas o UP).

Por sí mismo, el AF sirve para eliminar variables irrelevantes. Sin embargo, las características de nuestra fuente de información y los objetivos específicos del estudio, hicieron necesaria una selección previa. Además de eliminar las correspondientes a la ubicación por estado federativo de las unidades encuestadas,[3] eliminamos aquellas variables de la Encuesta que contaban con gran cantidad de datos iguales a cero, puesto que ellos contribuyen en poco al estudio de la variación de los datos; es decir, al AF.

El segundo criterio de selección, previo al cálculo de los factores, consistió en la estimación de la matriz de correlaciones entre todas las variables y en la revisión de sus coeficientes. Puesto que una de las metas del AF es obtener factores que ayuden a explicar éstas correlaciones, las variables deben estar relacionadas entre sí. Con ésta base eliminamos aquellas cuyo coeficiente de correlación fuera menor al valor absoluto de .25, tomando en cuenta que el valor de tales coeficientes fluctúa entre 1 y -1.

Con el fin de hacer una selección más precisa de las variables a las que le aplicaríamos el AF, a la base de datos seleccionada con el procedimiento anterior le aplicamos pruebas de adecuación muestral, como el índice de Kaiser-Meyer-Olkin y las medidas de adecuación muestral para cada una de las variables (ver Anexo Metodológico).

[2] En el Anexo Metodológico presentamos formalmente los detalles de los análisis factorial y de conglomerados.

[3] No obstante, en el análisis de conglomerados tomamos en cuenta la ubicación geográfica de las UP (Cf. infra., sección IV).

III. Análisis Factorial

Como hemos dicho, nuestro interés al aplicar el AF a los datos seleccionados de la Encuesta es estudiar si hay relación alguna entre el grado de modernidad de los predios y el tipo de semilla que usan. Con ésta base, el último criterio de selección, previo a la aplicación del AF, fue considerar, dentro de las variables relevantes arrojadas por las pruebas expuestas en la sección anterior, aquellas que de alguna forma u otra capturaran el grado de modernización o capitalización de los predios productores de maíz. Para tal fin tuvimos que adecuar la información original. Esto debido a que, como los datos se presentan en términos absolutos, es posible que el valor más elevado de una variable para un predio respecto a otro --digamos, el valor de las ventas-- se deba a que el primero sea mayor que el segundo y no a que sea más capitalista. Por tal motivo, las variables seleccionadas para llevar a cabo el AF se definieron en términos relativos.

Antes de presentar el listado de tales variables, conviene recordar que el mismo AF sirve para desechar algunas de ellas. Asimismo, es necesario aclarar que lo ideal del AF es obtener un sólo factor que conjugue las variables significativas al estudio.

III.1 Resultados del primer ejercicio

Las variables seleccionadas en nuestra primera aproximación al estudio de la relación existente entre modernización y tipo de semillas usadas de los predios maiceros son:

Rendimiento físico (Kg. por Ha.) ..... REND = Q/SUPSEM
Proporción del costo de la mano de obra alquilada sobre la total (sobre la suma del costo de la mano de obra alquilada y la propia o familiar)..... RELMO = CMOA/(CMOA+CMOP)
Proporción de las ventas sobre la producción física ..... PROVEN = VENTA/Q
Valor de los insumos sobre la superficie sembrada..... INSSUP = INS/SUPSEM
Proporción del costo de la maquinaria sobre el costo total ..... RELKC = CMQ/(CMQ+CMO)[4]

[4] En un principio pensamos en variables tales como la relación entre el costo de la maquinaria y de la mano de obra (CMQ/CMO, ver cuadro 1). Sin embargo, si por alguna razón el dato sobre CMO de alguno de los predios encuestados no estaba presente en la información de la encuesta, la relación tendría un cero en el denominador y el dato se perdería. La búsqueda de cómo definir una variable que nos aproximara al conocimiento de la relación capital-trabajo --es decir, a un indicador del grado de capitalización--, finalizó al definirla como sigue: RELKC = CMQ/(CMQ+CMO). La variable estaría midiendo la proporción que el costo de la maquinaria representa en el costo total, entendido el último como la suma de los gastos en maquinaria y en mano de obra. Por construcción tal variable estará acotada al intervalo [0,1], siendo 1 cuando el CMO es cero y 0 cuando el

A éstas 5 variables agregamos una "dummy", de tal forma que capturase el tipo de semilla (TS) usada (le asignamos el valor 1 si la UP utiliza SM y 0 si utiliza SC).

La hipótesis subyacente es que una UP será más capitalizada --y utilizará SM-- en cuanto las variables consideradas adquieran, dentro de su rango de variación, valores más altos. En otras palabras, esperamos una correlación positiva entre todos los pares de las variables.

Presentamos las estadísticas descriptivas de las variables usadas en el cuadro 2.a. Respecto a las medias, es necesario aclarar que, siendo medias aritméticas simples de la información original, sus valores no corresponden al concepto de valor "representativo" de los datos. Tomando como ejemplo el rendimiento, su media real sería la suma de la producción de las 2,448 UP sobre la suma de sus correspondientes superficies sembradas, y no el dato de 1888.4 Kg. del cuadro (que le está dando la misma ponderación tanto a las UP grandes como a las pequeñas). No obstante, éstas medias se presentan porque aportan información sobre las tendencias; por ejemplo, según los datos, la proporción de ventas de los productores de maíz en México sobre su producción fue durante 1989 y en promedio de .42 (de lo cual se infiere que el autoconsumo de maíz representó el 58% de la producción).

El cálculo de la matriz de correlaciones entre todas las variables --que aporta información de la asociación entre ellas-- arrojó los resultados presentados en el cuadro 2.b (las correlaciones significativas están en negritas, merece la pena resaltar que RELMO no se relaciona significativamente con ninguna variable).

Por su parte, el índice Kaiser-Meyer-Olkin tiene un valor satisfactorio de .78 y algo similar sucede con los resultados de las medidas de adecuación muestral para cada una de las variables (cuadro 2.c).

El paso siguiente fue calcular los factores. Los resultados de éste ejercicio sobre los valores propios y la proporción de variación explicada por ellos está contenida en el cuadro 2.d.

Tomando en cuenta que los factores con valores propios mayores a la unidad son los significativos (ver Anexo Metodológico), tenemos que dos de ellos lo son y que, unidos, explican el 64.1% de la variación existente en las 6 variables originales.

La matriz asociada a éstos dos factores se presenta en el cuadro 2.e. Nótese que ella indica que el factor 1 estaría asociado básicamente con REND, PROVEN, INSSUP, RELKC y TS, mientras que el

---

CMQ es cero.

factor 2 lo estaría únicamente con RELMO.

Conviene aclarar que existe un problema interpretativo surgido del signo negativo de RELKC. Según los resultados contenidos en los cuadros 2.b y 2.e, mayores rendimientos, mayor proporción de ventas sobre la producción y la tendencia a usar SM estarían asociados con bajas proporciones de gasto en maquinaria sobre el gasto total (en cálculos posteriores de factores tratamos de resolver la cuestión, infra. sección III.2).

La proporción de variación explicada a través del AF, esto es, la varianza reproducida para cada una de las variables, proporciona un índice de ajuste del AF para cada variable. De esto da cuenta el cuadro 2.f.

La primera fase del AF proporciona los elementos necesarios para eliminar las variables RELMO y RELKC. La primera de ellas debido a que no se relaciona significativamente con otra variable (cuadro 2.b) y a su baja medida de adecuación muestral (cuadro 2.c), mientras la segunda debido a la difícil interpretación por sus signos negativos (cuadros 2.b y 2.e)

III.2 Resultados definitivos

El siguiente paso del AF consistió pues, en eliminar las variables RELMO y RELKC. Al proceder de ésta forma tratamos de mejorar el ajuste del AF para cada variable, evitamos el problema de interpretación del signo de RELKC y esperamos la formación de un sólo factor significativo.[5]

Con el fin de capturar la influencia que tiene el tipo de propiedad y acceso al agua en el grado de capitalización de los predios maiceros, en éste ejercicio incluimos una variable cualitativa (MODTP), formada de la siguiente manera:

		0	si la UP es de temporal y ejidal
MODTP	=	1	si la UP es de temporal y privada
		2	si la UP es de riego y ejidal
		3	si la UP es de riego y privada

Este último ejercicio también incluye implícitamente la variable MOD, que recoge el tipo de acceso al agua de los predios maiceros (si la UP es irrigada, la variable adquiere el valor de 1, y si es

[5] Esto debido a que, si un factor está asociado fundamentalmente a una o más variables y éstas se eliminan, lo más factible es que el factor también lo sea (es el caso de RELMO).

de temporal, su valor es de 0).[6]

Las estadísticas descriptivas de MODTP son, media = 0.598 y desviación estándar = 0.909 y las de MOD: media = 0.222 y desviación estándar = 0.416.

Con este nuevo grupo de variables la matriz de correlaciones presenta un panorama bastante distinto que la anterior (cuadro 3.a, compárese con el cuadro 2.b).

Si consideramos al valor absoluto de .25 como el mínimo para mostrar un cierto grado de relación entre variables (señalados en el cuadro con negritas), tenemos que, respecto al AF anterior, el número de coeficientes significativos se ha incrementado de 9 a 11. Además, hay que notar que la variable REND se relaciona positiva y significativamente con todas las demás y, lo más interesante para nuestros propósitos, es que lo mismo sucede con TS. Lo último implica que, efectivamente, el uso de SM está asociado con: los predios privados irrigados, mayores rendimientos y mayor proporción de canalización del producto a los mercados.

En cuanto a las medidas de adecuación muestral se tiene que el índice Kaiser-Meyer-Olkin tiene un valor satisfactorio de 0.69. Por su parte, los valores de la adecuación muestral para cada una de las variables tienen niveles aceptables (cuadro 3.b). Destacan los valores de REND, INSSUP y TS, pues son más elevados que los obtenidos en el primer ejercicio del AF (véase cuadro 2.c).

Los resultados sobre el grado de vinculación entre los factores calculados y las variables se encuentran en el cuadro 3.c. Respecto al previo, éste ejercicio mejoró la proporción de variación explicada por el factor 1 (49.1% vrs. 45.9% del primero, ver cuadro 2.d). Sin embargo, debido a que el valor propio del segundo factor es mayor a la unidad, no es posible eliminarlo, de tal forma que nos quedáramos con uno sólo. No obstante, el poder explicativo de éstos dos factores es superior al del primer ejercicio (71.4% vrs. 64.1%).

La matriz de los dos factores significativos (cuadro 3.d) muestra que, no obstante la permanencia de un segundo factor, el primero está asociado significativamente (con valores mayores al valor absoluto de 0.5) y de manera positiva con todas las variables. Esto quiere decir que el primer factor está relacionado positivamente a los rendimientos, a las UP con irrigación (sean privadas o ejidales), a la vinculación de los predios con el mercado del maíz

[6] Podría decirse que hubiera sido suficiente utilizar una variable dummy que sólo hiciera distinción del tipo de propiedad. Sin embargo, al proceder de ésta forma en el AF, la variable no fue significativa, pero lo opuesto resulta cuando la estructura de propiedad se introduce en asociación con el tipo de acceso al agua.

y al uso de SM.[7]

Por último, los índices de ajuste del AF para cada variable (cuadro 3.e) son aceptables.

III.3 Síntesis

A partir de los resultados del AF podemos decir que, dentro de la veintena de variables usadas en la Encuesta, seis de ellas son significativas en nuestro propósito de estudiar la vinculación entre el grado de capitalización o "modernidad" de los predios maiceros contemporáneos y el tipo de semillas que usan como insumos. Las variables son: los rendimientos (volumen de la producción sobre la superficie sembrada), la proporción del volumen producido canalizado al mercado, el tipo de semilla usada (mejorada o criolla), el tipo de propiedad y acceso al agua (riego privado, riego ejidal, temporal privado y temporal ejidal) y el valor de los insumos aplicados por hectárea.

La datos de éstas variables son, precisamente, los que usaremos para completar el estudio, mediante el análisis de conglomerados.

IV. Análisis de conglomerados

Nuestro objetivo al aplicar el análisis de conglomerados (AC en adelante) es el de agrupar a las unidades productoras de maíz de acuerdo a sus características, en base a las variables seleccionadas mediante el AF.

Un problema inicial para llevar a acabo el AC se debió al excesivo número de observaciones (que, una vez eliminadas las variables irrelevantes al estudio, suman alrededor de 20,000); algo que lo hacía poco práctico, tanto desde el punto de vista computacional como interpretativo.[8] Por tal motivo, obtuvimos una muestra aleatoria formada por el 10% de la población original, lo que equivale a 245 cuestionarios o UP. A ésta base de datos fue a la que le aplicamos el AC.

La formación de grupos consistió en 12 etapas; la primera es

[7] Recordar que la interpretación de las variables cualitativas es de ésta forma pues: a TS se le asigna un valor de uno al uso de semilla mejorada y de cero al uso de semilla criolla, mientras que a MODTP se le dán los valores más altos a la irrigación (de 3 a los predios privados y de 2 a los ejidales) y sólo el valor de 1 a los predios privados de temporal y de 0 a los ejidales temporaleros. Así pues, solamente se le asigna peso al uso de semillas mejoradas y mayores pesos a la irrigación.

[8] En una investigación en curso aprovecharemos en su totalidad la riqueza informativa proporcionada por la Encuesta. Esto lo haremos al elaborar los análisis factorial y de conglomerados, así como un estudio econométrico, por entidad federativa.

trivial, pues es aquella en donde cada observación forma un grupo (en nuestro caso serían 245). Además, en las dos etapas siguientes se forman un gran número de grupos, lo que dificulta la interpretación. Por tal motivo, nuestra revisión de las agrupaciones es a partir de la cuarta etapa, en la que se formaron 16 grupos (cuadro 4.a).[9]

Cuarta etapa

Los dieciseis grupos formados en ésta fase (cuadro 4.a) pueden subdividirse claramente en dos: los primeros 8 corresponden a la modalidad temporal (del 4.1 al 4.8) y el resto son UP de riego (4.9 al 4.16). A su vez, dentro de los grupos de temporal, se tiene que los primeros 4 utilizan SC (4.1 al 4.4), mientras los restantes usan SM (4.5 al 4.8).

Los grupos 1 a 4 (de temporal, que usan SC) son, en su mayoría de tenencia ejidal, sobre todo el 1 y el 4. El primero de ellos contiene la mitad de la muestra (124 casos) con un rendimiento bajo de 1213 kilos por hectárea, con sólo 92,808 pesos de insumos aplicados por hectárea y una proporción de únicamente 28% de su producción destinada a la venta. Los estados que tienen mayor representación en este grupo son: Chihuahua, Oaxaca y Puebla, es decir, sus UP están localizadas en entidades federativa de las tres zonas en las que hemos dividido al país.[10]

El grupo 2 es el de menores rendimientos (la excepción es el grupo 8, formado por una sola observación), con 1048 kilos de maíz producido por hectárea. Además, se trata prácticamente de un grupo formado por unidades de autoconsumo, que, sin embargo, aplican más insumos por hectárea que el grupo 1. En su mayoría, el grupo está formado por unidades ubicadas en el centro y sur del país (Guerrero, Hidalgo y México).

El grupo 3 está constituido sólo por tres UP, que se distinguen de los dos grupos anteriores por un mayor rendimiento (1666 kilos por hectárea), pero, sobretodo, por un altísimo costo en insumos (667,466 pesos por hectárea). La totalidad de las unidades pertenecen al Estado de México.

Las características distintivas del grupo 4 dentro de los de

[9] Presentamos los resultados del resto de las etapas en los cuadros 4.b a 4.g. La primera columna de los cuadros 4 contiene, en la primera parte, las variables consideradas y, en la segunda, los estados. Las columnas que siguen presentan los grupos formados por el AC.

[10] Sin embargo, la representatividad de Chihuahua también es alta en los grupos 5 (de temporal, que usan SC, pero compuesto por UP modernas), así como en los 13, 14 y 15 (de riego y modernos); característica que no comparten los predios maiceros de Oaxaca y Puebla.

temporal que usan SC son básicamente dos: por una parte, es el que tiene el mayor rendimiento dentro de ese subgrupo (3234 kilos por hectárea), lo cual indica el probable uso de las UP que lo componen de excelentes tierras de temporal y, por otra parte, una vocación a la venta de su producción (90%). La hipótesis del tipo de tierras se refuerza por el hecho de que gran parte de las UP de este grupo pertenecen a los estados de Jalisco, Puebla y Veracruz, donde es conocida la existencia de regiones con buenas tierras de temporal.

Los grupos 5 a 8 también son de temporal, pero utilizan SM. Una diferencia respecto a los grupos 1 a 4 es de orden cuantitativo: mientras que éstos reúnen al 65% de las UP, los numerados del 5 al 8 contienen sólo al 9% de las UP.

Por su parte, éstos últimos pueden dividirse en dos: los 7 y 8, que son de autoconsumo y de bajísimos rendimientos, y los 5 y 6, comerciales y con elevada producción por hectárea.

Las UP que forman éstos grupos --es decir, con tierras de temporal, que usan SM--, se encuentran en todo el país. Sin embargo, todas las de autoconsumo y pobres (grupos 7 y 8) están ubicadas exclusivamente en el centro y sur (Chiapas. Guerrero, Jalisco, México, Tlaxcala y Zacatecas). Por su parte, el grupo 6, el más productivo de los temporaleros, usa SM y la mitad de sus UP son privadas.

Podemos sintetizar los hallazgos del AC para la agricultura temporalera productora de maíz diciendo que las UP de autoconsumo, menos productivas y ubicadas en el Centro y Sur del país, son las más proclives al uso de SC. Sin embargo, hay un grupo, el 4, que no obstante usar SC para su producción, es productivo, comercial y cuyas UP se ubican mayoritariamente en el Centro y Sur del país.

Los grupos 9 a 16 contienen las UP de riego. En términos del tipo de semilla que usan, puede hacerse una subdivisión: grupos 9 y 10, que utilizan SC, y 11 a 16, que usan SM.

Los grupos 9 y 10 tienen todo en común, salvo las ventas: mientras que el primero comercializa el 80% de su producción, el segundo sólo introduce al mercado el 2.6%. No hay un patrón claro respecto a la distribución geográfica de las UP en éstos dos grupos.

Por su parte, los grupos 11 y 12 son poco representativos (en ellos sólo hay tres y dos UP respectivamente), y están distribuidos en cinco estados, por lo que la información proporcionada por el AC no es, en éste caso, valiosa.

Los grupos 13 y 15 son parecidos: con altos rendimientos y una producción mercantil; su diferencia estriba en el tipo de propiedad y en el valor de los insumos que usan: la mitad de las UP del grupo 15 son privadas y casi todas las pertenecientes al 13 son ejidales; además, el primero aplica 283,433 pesos de insumos por ha.,

mientras que el segundo lo hace en más del doble. La mayor parte de las UP de éstos dos grupos se encuentra en los estados norteños de Chihuahua y Sinaloa. A los grupos 13 y 15 se le puede agregar el 16, cuya única gran diferencia es el valor de los insumos aplicados.

La principal diferencia entre el grupo 14 y los demás de riego con elevados rendimientos y que usan SM (13, 15 y 16) es que el primero no es comercial (consume prácticamente las 3/4 partes de su producto). Es posible que la localización geográfica de las UP de éste grupo (Coahuila, Chihuahua, Puebla y Querétaro), es decir, en zonas de difusión y acceso a las SM, explique el porqué de su uso, no obstante de ser predios de autoconsumo.

Si, dentro de las UP de riego (grupos 9 a 16) ignoramos a los que cuentan con pocas UP (11, 12 y 14), es posible constatar, en términos generales, que los usuarios de SM tienen características más capitalizadas que los que aplican SC.

Además de lo anterior, las agrupaciones arrojadas por la cuarta etapa del AC indican que, no obstante la gran heterogeneidad en la producción de maíz en México, la mayoría de las UP que lo cultivan son ejidos temporaleros, usan SC y no son modernas, es decir, canalizan su producción al autoconsumo, usan pocos insumos comerciales y tienen bajos rendimientos (grupo 1). Estas UP se ubican, sobre todo, en el centro y sur del país.

Quinta etapa

En ésta fase se forman 14 grupos (cuadro 4.b), de los cuales la mitad está formado por UP de temporal y el resto de riego. Esto como consecuencia de la fusión de los grupos 2 y 3, y 13 y 14 de la cuarta etapa.

Dentro de los grupos de temporal, los primeros tres son usuarios de SC y los restantes cuatro de SM. El nuevo grupo 2 está constituido por UP de bajo rendimiento (1172 kg/ha), baja proporción de venta (14%), a pesar de que aplican un nivel elevado de insumos (434,438 $/ha). Ninguna de las UP del grupo se localiza en el norte del país.

Con ésta etapa del AC se hace evidente que las UP de temporal que usan SM (grupos 4 y 5) tienen más rasgos capitalistas que los predios que aplican SC (grupos 1 y 2).

Por lo que corresponde a los grupos de riego (8 a 14), decíamos que el 12 es el resultado de la unión de los grupos 13 y 14 de la etapa anterior. Este nuevo grupo es representativo puesto que contiene el 25% de las unidades de riego y sus características son de un alto nivel de capitalización con el atenuante de un nivel medio de aplicación de insumos. La mayor parte de sus UP se localizan en los estados del norte (Coahuila, Sinaloa y Chihuahua).

Si ignoramos a los grupos 10 y 11, es decir a los que tienen poca representatividad, es notorio el mayor grado de capitalización de los conjuntos usuarios de SM (12, 13 y 14) respecto a los 8 y 9, que aplican SC. Además, los grupos 12 a 14 tienden a localizarse más en los estados del norte respecto a los otros dos.

Sexta etapa

La sexta etapa forma 11 grupos (cuadro 4.c), de los cuales seis son de temporal y cinco de riego. Dentro de los temporaleros, los tres primeros utilizan SC y otros tres SM. El nuevo grupo 4 resulta de la asociación de los 4 y 5 de la etapa anterior y está formado por las UP temporaleras usuarias de SM, que son las más capitalistas dentro de las maiceras de temporal y están distribuidas en todo el país. Aunque en menor medida, el grupo 3 comparte tales características, con la diferencia de que es usuario de SC. Lo anterior nos lleva a proponer que éstos dos grupos se localizan en las regiones de México con tierras de buen temporal.

Cabe observar que el grupo 1 es bastante homogéneo, puesto que en las etapas descritas no se ha asociado con ningún otro: es el que contiene la mitad de la muestra, con una elevada proporción de UP ejidales, de bajo uso de insumos y una alta propensión al autoconsumo.

Los grupos de riego se reducen de 7 a 5 por medio de la unión de los 8 y 9, y de los 13 y 14 de la etapa anterior. Estos cinco grupos de UP irrigadas están compuestos por tres conjuntos de predios en su gran mayoría ejidales, que producen en condiciones precarias (7 a 9) y dos de predios modernos con propiedad mixta (10 y 11).

El grupo 7 se distingue del 8 y 9 por ser usuario de SC, mientras que el 8 se diferencia del 9 por producir exclusivamente para el autoconsumo. Además, mientras que las UP del grupo 7 se distribuyen en todo el país, las pertenecientes a los otros dos se concentran en un reducido número de estados. No obstante lo anterior, hay que notar que los grupos 8 y 9 son poco representativos, con sólo 5 UP.

Por su parte, la principal diferencia entre los dos grupos de predios más productivos de la muestra --que son irrigados y que usan SM (10 y 11)--, estriba en el valor de los insumos aplicados por ha. (el primero usa mucho menos que el segundo); en la estructura de la propiedad de sus UP (hay más del doble de predios privados en el 11) y en su localización geográfica (la mayoría de las UP del grupo 11 se encuentran en los estados norteños de Chihuahua y Sinaloa). Lo último sugiere que, dentro de las UP más modernas, las más capitalistas se localizan en el norte de México.

Séptima etapa

Esta fase se caracteriza por la formación de ocho grupos (cuadro

4.d), o sea, por la agrupación de tres pares de conjuntos de la etapa anterior. Por lo que respecta a las UP de temporal, se unen los grupos 1 y 2, y 5 y 6 de la sexta etapa, de tal manera que se forman cuatro grupos de temporal, dos de los cuales utilizan SC y el resto SM. En ambos casos, uno de los grupos es claramente más capitalizado que el otro (el grupo 2 lo es con respecto al 1 y el grupo 3 en relación al 4). La ubicación geográfica de ellos se da en todo el país.

Los grupos 5 a 8 están formados por UP de riego (el 6 es la unión entre los 8 y 9 de la sexta etapa), de los cuales el 5 es el único que utiliza SC y el 6 se caracteriza por tener bajos niveles de capitalización, pocas UP y una diseminación geográfica. Como en la etapa anterior, los grupos 7 y 8 son los de mayores niveles de capitalización, sobretodo el último.

Octava etapa

La fase produce 6 grupos (cuadro 4.e): tres de temporal y tres de riego. En el primero se unen los conjuntos 1 y 2 de la etapa previa, contiene el 65% de la muestra e incluye a todos los predios temporaleros que usan SC.

La otra unión que se hace en ésta etapa es la de los 7 y 8 de la séptima, que significa la agrupación de las UP irrigadas más modernas, que usan SM.

Así pues, en ésta etapa se forman tres grupos de predios maiceros temporaleros: el 1 es el mayoritario, con UP en todo el país, que usa SC, poco moderno y con propiedad mixta (aunque la mayoría de las UP que lo forman son ejidales); no obstante que el 2 también es de propiedad mixta, es mucho más productivo y moderno, usa SC y probablemente sus UP están situadas en buenas tierras de temporal de los Estados de Chiapas, Chihuahua, Guerrero, Jalisco y Michoacán. Por último, el grupo 3, no obstante de usar SM, es muy ineficiente, no vende el maíz que produce y la mayor parte de sus UP se localizan en el sur del país.

Por lo que respecta a los tres grupos de predios irrigados, los 4 y 5 están compuestos por UP improductivas, en su mayoría ejidales y cuya diferencia reside en el monto de insumos aplicados y en el tipo de semilla aplicada (el 4 recurre en mucho menor medida a la aplicación de insumos y usa SC). En contraste, el grupo 6 es de propiedad mixta, el más moderno (no obstante que, al igual que el 5, usa SM). La ubicación geográfica del conjunto se da, sobre todo, en el norte del país.

Novena etapa

La única diferencia de ésta fase respecto a la anterior es la unión de los grupos 2 y 3 (cuadro 4.f). O sea que las UP de temporal ahora se subdividen en 2 grupos: las que usan SC (grupo 1) y las

que aplican SM (grupo 2). Tal unión muestra la importancia de las semillas usadas en la tipificación de las UP temporaleras.

Décima etapa

En ésta fase sólo se unen los grupos 3 y 4 de la previa, es decir, se asocian los predios irrigados poco productivos, independientemente del tipo de semilla que usen (cuadro 4.g). Que el grupo 5 de la novena etapa se mantenga en la décima (siendo ahora el 4), pone en evidencia que las UP más modernas tienden a usar SM.

Undécima etapa

La etapa forma dos grupos, que, en términos generales subdividen a la UP de maíz en México conforme a su acceso al agua, régimen de propiedad, tipo de semillas que usan, grado de modernidad y ubicación geográfica (cuadro 4.h). En efecto, las UP menos modernas (grupo 1 de bajos rendimientos y reducido mercadeo de su producto) son casi todas de temporal y ejidales, más proclives al uso de SC y se encuentran en todo el país. Por su parte, las UP modernas (grupo 2) son irrigadas, con un peso mayor de predios privados, altos rendimientos y una elevada proporción de producción vendida. Además, aunque la gran mayoría de éste tipo de predios se ubican en el norte y centro del país, es en los estados de la región donde se concentran.

V. Conclusiones

Los resultados que arrojan los análisis factorial y de conglomerados no contradicen la hipótesis que hemos estudiado sobre la vinculación entre grado de capitalización o modernidad de los predios maiceros de México y el tipo de semillas que usan (Yúnez-Naude, A., et.al., op.cit.: 1994).

En efecto, nuestro AF indica que hay un factor --el de mayor peso-- que está relacionado positivamente con los rendimientos, con los predios irrigados, con la vinculación de éstos al mercado del maíz, con la cantidad de insumos utilizados y con el uso de SM.

Por su parte, la agrupaciones de predios maiceros elaboradas mediante el AC indican que el tipo de semillas usadas es un componente significativo de las diferencias entre las UP y que la mayor parte de los predios con mayor tendencia al uso de semillas ricas en biodiversidad genética son las que no tienen acceso al riego y las menos modernas, localizadas en su mayoría en la zona centro y sur de México.

Los grupos formados por el AC sirven para elaborar una tipología de productores maiceros. Revirtiendo la presentación previa de las etapas del AC, es decir, iniciando con la quinta, podemos decir lo siguiente.

a) Hay una diferencia fundamental entre predios según su acceso al agua: los irrigados son los más modernos --con mayores rendimientos y una elevada proporción de maíz vendida--, los que aplican más insumos y usan exclusivamente SM (cuadro 4.h).

b) Así mismo, las UP más modernas son las que tienen la tendencia a usar una mayor proporción de SM, sean éstas irrigados o de temporal (cuadro 4.g).

c) Los predios privados --irrigados o no-- son más proclives al uso de SM. Esto se observa al notar que la proporción de éste tipo de UP en cierto grupo es mayor respecto a otro cuando el peso del uso de SM es más alto (cuadros 4.e y 4.f).

d) Unido al acceso al recurso pluvial, un componente fundamental de la estructura de la producción de maíz en México es el tipo de semillas que los predios usan; conclusión que surge al notar que las agrupaciones que se hacen de las etapas 4 a la 6 son, además del tipo de acceso al agua, entre unidades con proporciones similares en el uso de SM (Cuadros 4.a a 4.c).

Todo esto nos lleva a reforzar nuestra hipótesis de que, en el caso de la producción de maíz en México, existe una relación positiva entre desarrollo regional, modernización y propiedad privada de los predios maiceros y el uso de semillas mejoradas. Si tomamos en cuenta que el otro tipo de semillas: las autóctonas y criollas, son las que contribuyen en mayor grado a la conservación de la biodiversidad genética del cultivo, podemos decir que los objetivos gubernamentales declarados de transformar la producción tradicional de cultivos básicos en México --incluyendo las expectativas de modernización que se tienen a raíz de la reciente Reforma Ejidal-- y que no reparan en las cuestiones que nos han ocupada en la presente investigación, estarían en conflicto con el mantenimiento de la riqueza genética del maíz (en ibid. presentamos una reflexión detallada sobre el tema).

No obstante lo anterior, nuestro estudio y, sobre todo, investigaciones microrregionales (Bellon, Mauricio R., y J. Edward Taylor: 1993, Brush, B. Stephen: 1992 y Brush, Stephen, B., et.al.: 1988), indican que, en México, las semillas que los campesinos usan para obtener maíz son aún ricas en biodiversidad. Esto significa que es posible tomar medidas de conservación in situ, para lo cual se requieren diagnósticos más precisos y un mejor conocimiento de las condiciones necesarias para que los campesinos mantengan la riqueza genética del maíz y, al mismo tiempo, logren salir de las precarias condiciones en las que viven.[11] A lo anterior habrá que

[11] La investigación binacional en curso "Conservation of genetic diversity and improvement of crop production in Mexico: A farmer-based approach" tiene, precisamente éstos objetivos. El proyecto se realiza en la UNAM y Universidad de California en Davis y cuenta con el apoyo financiero de la fundación McKnight.

agregar la necesidad de que en los planes y acciones gubernamentales hacia el subsector se tome en cuenta la problemática, pues de todo esto dependerá el logro de un desarrollo alimenticio sustentable.

Cuadro 1
Variables capturadas para cada unidad productiva en la "Encuesta nacional de costos, coeficientes técnicos y producción agrícola para el maíz."

1. Datos generales

MOD = modalidad (privado irrigado y de temporal y ejidal irrigado y de temporal)
TS = tipo de semilla usada (mejorada y criolla autóctona)
EDO = Estado donde se localiza la unidad productiva.

2. Gastos

2.1 Asociados directamente al proceso productivo:

SEMCRIO = gasto en semilla criolla o autóctona
SEMMEJ = gasto en semilla mejorada
INSUMOS = gasto en insumos

2.2 Gastos indirectos:

INTERESES = gasto en intereses
IMPUESTOS = gasto en impuestos
SEGURO = gasto en seguro
SERVICIOS = gasto en servicios (sobretodo pagos por uso de agua)
GASDIV = gastos diversos

3.Características físicas y destino de la producción:

SUPSEM = superficie sembrada
Q = volumen de la producción
VENTA = volumen de venta
AUTOCON = volumen de autoconsumo (para consumo humano, animal y uso de semilla propia)

4. Costos

CPHC = costo ponderado por hectárea
CMO = costo de la mano de obra
CMQ = costo de la maquinaria
CMOP = costo mano de obra propia
CMOA = costo mano de obra alquilada

5. Precios:

PSEMC = precio de la semilla criolla
PSEMM = precio de la semilla mejorada
PFER = precio de los fertilizantes
PHER = precio de los herbicidas
PFUN = precio de los fungicidas
PINSE = precio de los insecticidas

Fuente: SARH, FIRA y BANRURAL, Encuesta

Cuadro 2.a
Estadísticas de las variables

Variable	MEDIA	DESV.EST.
REND	1888.415	1480.399
RELMO	0.730	0.358
PROVEN	0.415	0.424
INSSUP	200971.932	196474.178
RELKC	0.642	0.296
TS	0.209	0.406

Cuadro 2.b.
Matriz de correlaciones

	REND	RELMO	PROVEN	INSSUP	RELKC	TS
REND	1.000					
RELMO	-0.065	1.000				
PROVEN	**0.518**	-0.147	1.000			
INSSUP	**0.493**	0.019	**0.323**	1.000		
RELKC	**-0.437**	0.237	**-0.550**	0.242	1.000	
TS	**0.527**	-0.090	**0.392**	**-0.356**	**-0.430**	1.000

Cuadro 2.c.
Medidas de adecuación muestral

Variable	MSA
REND	0.768
RELMO	0.672
PROVEN	0.785
INSSUP	0.795
RELKC	0.767
TS	0.825

Cuadro 2.d
Resultados del cálculo de factores

Factor	Valor propio	Porcentaje de variación	Porcentaje acumulado
1	2.755	45.9	45.9
2	1.089	18.2	64.1
3	0.710	11.8	75.9
4	0.606	10.1	86.0
5	0.452	7.5	93.6
6	0.385	6.4	100.0

Cuadro 2.e
Matriz de factores

Variable	FACTOR 1	FACTOR 2
REND	**0.811**	0.204
RELMO	-0.225	**0.856**
PROVEN	**0.766**	-0.115
INSSUP	**0.616**	0.441
RELKC	**-0.736**	0.311
TS	**0.732**	0.099

Cuadro 2. f
Varianzas reproducidas

Variable	Comunalidad
REND	0.699
RELMO	0.784
PROVEN	0.600
INSSUP	0.574
RELKC	0.640
TS	0.545

Cuadro 3.a
Matriz de correlaciones

	REND	MODTP	PROVEN	TS	MOD	INSSUP
REND	1.000					
MODTP	**0.420**	1.000				
PROVEN	**0.518**	0.188	1.000			
TS	0.527	0.322	0.392	1.000		
MOD	**0.420**	**0.918**	0.164	0.325	1.000	
INSSUP	**0.493**	0.187	**0.323**	0.356	0.174	1.000

Cuadro 3.b
Medidas de adecuación muestral

Variable	MSA
REND	0.778
MODTP	0.595
PROVEN	0.777
TS	0.863
MOD	0.587
INSSUP	0.803

Cuadro 3.c
Resultados del cálculo de factores

Factor	Valor propio	Porcentaje de variación	Porcentaje acumulado
1	2.945	49.1	49.1
2	1.338	22.3	71.4
3	0.681	11.4	82.7
4	0.582	9.7	92.5
5	0.370	6.2	98.6
6	0.081	1.4	100.0

Cuadro 3. d
Matriz de factores

Variable	FACTOR 1	FACTOR 2
REND	**0.813**	0.260
MODTP	**0.755**	**-0.612**
PROVEN	0.587	0.483
TS	**0.693**	0.257
MOD	**0.748**	**-0.625**
INSSUP	0.570	0.451

Cuadro 3.e
Varianzas reproducidas

Variable	Comunalidad
REND	0.730
MODTP	0.946
PROVEN	0.578
TS	0.547
MOD	0.952
INSSUP	0.529

Cuadro 4.a Resultados de la cuarta etapa del AC.

	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10	4.11	4.12	4.13	4.14	4.15	4.16
No. de observaciones	124	12	3	22	11	4	6	1	14	20	3	2	12	4	4	3
MOD (% temporal)	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0
MODTP *																
0 (%)	87.1	58.3	66.7	90.9	100.0	50.0	100.0	100.0								
1 (%)	12.9	41.7	33.3	9.1		50.0										
2 (%)									85.7	100.0	100.0	100.0	83.3	100.0	50.0	100.0
3 (%)									14.3				16.7		50.0	
REND (ton./ha.)	1213.3	1048.5	1666.7	3234.2	2947.0	5254.5	1499.7	1000.0	2052.3	1793.7	1200.8	1330.8	5001.0	4175.4	6745.5	3847.8
TS (% criolla)	100	100	100	100	0	0	0	0	100	100	0	0	0	0	0	0
PROVEN (%)	0.283	0.086	0.373	0.901	0.909	0.950	0.000	0.000	0.796	0.027	0.000	1.000	0.983	0.264	0.992	1.000
INSSUP **	92808.9	376180.6	667466.7	305077.6	293941.0	496212.5	261066.7	770333.0	144993.8	113550.0	274500.0	294958.5	283433.3	335350.0	615471.3	774286.6
Distribución de las UP por Estado (%)																
Coahuila (n)	1.6								14.3	10.0		50.0		25.0		
Chihuahua (n)	11.3			4.6	18.2					10.0	33.3		16.7	25.0	75.0	
Sinaloa (n)	0.8								14.3			50.0	33.3		25.0	66.7
Sonora (n)													16.7			
Distrito Federal (c)	4.8								7.1							
Guanajuato (c)	8.9	8.3								5.0			8.3			
Hidalgo (c)	4.0	25.0							21.4	30.0						
Jalisco (c)				13.6	9.1	50.0	16.7									
México (c)	4.0	25.0	100.0	9.1			16.7		14.3	5.0	33.3					33.3
Michoacán (c)	4.8	8.3		9.1	18.2	50.0			7.1	5.0			16.7			
Morelos (c)	0.8															
Puebla (c)	12.9	8.3		31.8					14.3	10.0				25.0		
Queretaro (c)	1.6													25.0		
San Luis Potosí (c)	3.2								7.1							
Tlaxcala (c)	3.2			4.6			16.7			10.0						
Veracruz (c)	6.5			13.6	9.1											
Zacatecas (c)	4.0						16.7				33.3					
Campeche (s)	0.8															
Chiapas (s)	5.6			4.6	27.3		16.7									
Guerrero (s)	4.0	25.0		4.6	9.1		16.7	100.0								
Oaxaca (s)	11.3			4.6						15.0			8.3			
Yucatán (s)	4.8				9.1											

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privada.

** Valor de insumos aplicados por hectárea cosechada.

(n) Norte del país

(c) Centrodel país

(s) Sur del país

Cuadro 4.b Resultados de la quinta etapa del AC.

	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	5.10	5.11	5.12	5.13	5.14
No. de observaciones	124	15	22	11	4	6	1	14	20	3	2	16	4	3
MOD (% temporal)	100	100	100	100	100	100	100	0	0	0	0	0	0	0
MODTP *														
0 (%)	87.1	60.0	90.9	100.0	50.0	100.0	100.0							
1 (%)	12.9	40.0	9.1		50.0									
2 (%)								85.7	100.0	100.0	100.0	87.5	50.0	100.0
3 (%)								14.3				12.5	50.0	
REND (ton./ha.)	1213.3	1172.1	3234.2	2947.0	5254.5	1499.7	1000.0	2052.3	1793.7	1200.8	1330.8	4794.6	6745.5	3847.8
TS (% criolla)	100	100	100	0	0	0	0	100	100	0	0	0	0	0
PROVEN (%)	0.283	0.143	0.901	0.909	0.950	0.000	0.000	0.796	0.027	0.000	1.000	0.804	0.992	1.000
INSSUP **	92808.9	434437.8	305077.6	293941.0	496212.5	261066.7	770333.0	144993.8	113550.0	274500.0	294958.5	296412.5	615471.3	774286.6
Distribución de las UP por Estado (%)														
Coahuila (n)	1.6							14.3	10.0		50.0	6.3		
Chihuahua (n)	11.3		4.6	18.2					10.0	33.3		18.8	75.0	
Sinaloa (n)	0.8							14.3			50.0	25.0	25.0	66.7
Sonora (n)												12.5		
Distrito Federal (c)	4.8							7.1						
Guanajuato (c)	8.9	6.7							5.0			6.3		
Hidalgo (c)	4.0	20.0						21.4	30.0					
Jalisco (c)			13.6	9.1	50.0	16.7								
México (c)	4.0	40.0	9.1			16.7		14.3	5.0	33.3				33.3
Michoacán (c)	4.8	6.7	9.1	18.2	50.0			7.1	5.0			12.5		
Morelos (c)	0.8													
Puebla (c)	12.9	6.7	31.8					14.3	10.0			6.3		
Queretaro (c)	1.6											6.3		
San Luis Potosí (c)	3.2							7.1						
Tlaxcala (c)	3.2		4.6			16.7			10.0					
Veracruz (c)	6.5		13.6	9.1										
Zacatecas (c)	0.8													
Campeche (s)	5.6		4.6	27.3		16.7								
Chiapas (s)	4.0	20.0	4.6	9.1		16.7	100.0							
Guerrero (s)	11.3		4.6						15.0			6.3		
Oaxaca (s)	4.8			9.1										
Yucatán (s)	4.0					16.7				33.3				

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privada.

** Valor de insumos aplicados por hectárea cosechada.

(n) Norte del país

(c) Centrodel país

(s) Sur del país

Cuadro 4.c Resultados de la sexta etapa del AC.

	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10	6.11
No. de observacione	124	15	22	15	6	1	34	3	2	16	7
MOD (% temporal)	100	100	100	100	100	100	0	0	0	0	0
MODTP *											
0 (%)	87.1	60.0	90.9	86.7	100.0	100.0					
1 (%)	12.9	40.0	9.1	13.1							
2 (%)							94.1	100.0	100.0	87.5	71.4
3 (%)							5.9			12.5	28.6
REND (ton./ha.)	1213.3	1172.1	3234.2	3562.3	1499.7	1000.0	1900.2	1200.8	1330.8	4794.6	5503.6
TS (% criolla)	100	100	100	0	0	0	100	0	0	0	0
PROVEN (%)	0.283	0.143	0.901	0.920	0.000	0.000	0.343	0.000	1.000	0.804	0.997
INSSUP **	92808.9	434437.8	305077.6	347880.1	261066.7	770333.0	126497.4	274500.0	294958.5	296412.5	683535.0
Distribución de las UP por Estado (%)											
Coahuila (n)	1.6						11.8		50.0	6.3	
Chihuahua (n)	11.3		4.6	13.3			5.9	33.3		18.8	42.9
Sinaloa (n)	0.8						5.9		50.0	25.0	42.9
Sonora (n)										12.5	
Distrito Federal (c)	4.8						2.9				
Guanajuato (c)	8.9	6.7					2.9			6.3	
Hidalgo (c)	4.0	20.0					26.5				
Jalisco (c)			13.6	20.0	16.7						
México (c)	4.0	40.0	9.1		16.7		8.8	33.3			14.3
Michoacán (c)	4.8	6.7	9.1	26.7			5.9			12.5	
Morelos (c)	0.8										
Puebla (c)	12.9	6.7	31.8				11.8			6.3	
Queretaro (c)	1.6									6.3	
San Luis Potosí (c)	3.2						2.9				
Tlaxcala (c)	3.2		4.6		16.7		5.9				
Veracruz (c)	6.5		13.6	6.7							
Zacatecas (c)	4.0				16.7			33.3			
Campeche (s)	0.8										
Chiapas (s)	5.6		4.6	20.0	16.7						
Guerrero (s)	4.0	20.0	4.6	6.7	16.7	100.0					
Oaxaca (s)	11.3		4.6				8.8			6.3	
Yucatán (s)	4.8			6.7							

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privad

** Valor de insumos aplicados por hectárea cosechada.

(n) Norte del país

(c) Centrodel país

(s) Sur del país

Cuadro 4.d Resultados de la séptima etapa del AC.

	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8
No. de observaciones	139	22	15	7	34	5	16	7
MOD (% temporal)	100	100	100	100	0	0	0	0
MODTP *								
0 (%)	83.0	90.9	86.7	100.0				
1 (%)	17.1	9.1	13.3					
2 (%)					94.1	100.0	87.5	71.4
3 (%)					5.9		12.5	28.6
REND (ton./ha.)	1208.8	3234.0	3562.0	1428.0	1900.0	1253.0	4795.0	5504.0
TS (% criolla)	100	100	0	0	100	0	0	0
PROVEN (%)	0.268	0.900	0.920	0.000	0.343	0.401	0.804	0.997
INSSUP **	129675.3	305078.0	347880.0	371114.0	126497.0	282683.0	296413.0	683535.0
Distribución de las UP por Estado (%)								
Coahuila (n)	1.4				11.8	20.0	6.3	
Chihuahua (n)	10.1	4.5	13.3		5.9	20.0	18.8	42.9
Sinaloa (n)	0.7				5.9	20.0	25.0	42.9
Sonora (n)							12.5	
Distrito Federal (c)	4.3				2.9			
Guanajuato (c)	8.6				2.9		6.3	
Hidalgo (c)	5.8				26.5			
Jalisco (c)	0.7	13.6	20.0	14.3				
México (c)	7.9	9.1		14.3	8.8	20.0		14.3
Michoacán (c)	5.0	9.1	26.7		5.9		12.5	
Morelos (c)	0.7							
Puebla (c)	12.2	31.8			11.8		6.3	
Queretaro (c)	1.4						6.3	
San Luis Potosí (c)	2.9				2.9			
Tlaxcala (c)	2.9	4.5		14.3	5.9			
Veracruz (c)	5.8	13.6	6.7					
Zacatecas (c)	3.6			14.3		20.0		
Campeche (s)	0.7							
Chiapas (s)	5.0	4.5	20.0	14.3				
Guerrero (s)	5.8	4.5	6.7	28.6				
Oaxaca (s)	10.1	4.5			8.8		6.3	
Yucatán (s)	4.3		6.7					

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privada.

** Valor de insumos aplicados por hectárea cosechada.

Cuadro 4.e Resultados de la octava etapa del AC.

	8.1	8.2	8.3	8.4	8.5	8.6
No. de observaciones	161	15	7	34	5	23
MOD (% temporal)	0	100	100	0	0	0
MODTP *						
0 (%)	85.1	86.7	100.0			
1 (%)	14.9	13.3				
2 (%)				94.1	100.0	82.6
3 (%)				5.9		17.4
REND (ton./ha.)	1485.6	3562.3	1428.3	1900.2	1252.8	5010.4
TS (% criolla)	100	0	0	100	0	0
PROVEN (%)	0.355	0.920	0.000	0.343	0.401	0.862
INSSUP **	153643.3	347880.1	371114.3	126497.4	282683.4	414232.4
Distribución de las UP por Estado (%)						
Coahuila (n)	1.2			11.8	20.0	4.3
Chihuahua (n)	9.3	13.3		5.9	20.0	26.1
Sinaloa (n)	0.6			5.9	20.0	30.4
Sonora (n)						8.7
Distrito Federal (c)	3.7			2.9		
Guanajuato (c)	7.5			2.9		4.3
Hidalgo (c)	5.0			26.5		
Jalisco (c)	2.5	20.0	14.3			
México (c)	8.1		14.3	8.8	20.0	4.3
Michoacán (c)	5.6	26.7		5.9		8.7
Morelos (c)	0.6					
Puebla (c)	14.9			11.8		4.3
Queretaro (c)	1.2					4.3
San Luis Potosí (c)	2.5			2.9		
Tlaxcala (c)	3.1		14.3	5.9		
Veracruz (c)	6.8					
Zacatecas (c)	3.1	6.7	14.3		20.0	
Campeche (s)	0.6					
Chiapas (s)	5.0	20.0	14.3			
Guerrero (s)	5.6	6.7	28.6			
Oaxaca (s)	9.3			8.8		4.3
Yucatán (s)	3.7	6.7				

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privada.

** Valor de insumos aplicados por hectárea cosechada.

(n) Norte del país

(c) Centrodel país

(s) Sur del país

Cuadro 4.f Resultados de la novena etapa del AC.

	9.1	9.2	9.3	9.4	9.5
No. de observaciones	161	22	34	5	23
MOD (% temporal)	100	100	0	0	0
MODTP *					
0 (%)	85.1	90.9			
1 (%)	14.9	9.1			
2 (%)			94.1	100.0	82.6
3 (%)			5.9		17.4
REND (ton./ha.)	1485.6	2883.3	1900.2	1252.8	5010.4
TS (% criolla)	100	0	100	0	0
PROVEN (%)	0.355	0.627	0.343	0.401	0.862
INSSUP **	153643.3	355272.8	126497.4	282683.4	414232.4
Distribución de las UP por Estado (%)					
Coahuila (n)	1.2		11.8	20.0	4.3
Chihuahua (n)	9.3	9.1	5.9	20.0	26.1
Sinaloa (n)	0.6		5.9	20.0	30.4
Sonora (n)					8.7
Distrito Federal (c)	3.7		2.9		
Guanajuato (c)	7.5		2.9		4.3
Hidalgo (c)	5.0		26.5		
Jalisco (c)	2.5	18.2			
México (c)	8.1	4.5	8.8	20.0	4.3
Michoacán (c)	5.6	18.2	5.9		8.7
Morelos (c)	0.6				
Puebla (c)	14.9		11.8		4.3
Queretaro (c)	1.2				4.3
San Luis Potosí (c)	2.5		2.9		
Tlaxcala (c)	3.1	4.5	5.9		
Veracruz (c)	6.8	4.5			
Zacatecas (c)	3.1	4.5		20.0	
Campeche (s)	0.6				
Chiapas (s)	5.0	18.2			
Guerrero (s)	5.6	13.6			
Oaxaca (s)	9.3		8.8		4.3
Yucatán (s)	3.7	4.5			

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privada.

** Valor de insumos aplicados por hectárea cosechada.

(n) Norte del país

(c) Centrodel país

(s) Sur del país

Cuadro 4.g Resultados de la décima etapa del AC.

	10.1	10.2	10.3	10.4
No. de observaciones	161	22	39	23
MOD (% temporal)	100	100	0	0
MODTP *				
0 (%)	85.1	90.9		
1 (%)	14.9	9.1		
2 (%)			100.9	82.6
3 (%)			5.1	17.4
REND (ton./ha.)	1485.6	2883.3	1817.2	5010.4
TS (% criolla)	100	0	87	0
PROVEN (%)	0.355	0.627	0.351	0.862
INSSUP **	153643.3	355272.8	5714330	414232.4
Distribución de las UP por Estado (%)				
Coahuila (n)	1.2		12.8	4.3
Chihuahua (n)	9.3	9.1	7.7	26.1
Sinaloa (n)	0.6		7.7	30.4
Sonora (n)				8.7
Distrito Federal (c)	3.7		2.6	
Guanajuato (c)	7.5		2.6	4.3
Hidalgo (c)	5.0		23.1	
Jalisco (c)	2.5	18.2		
México (c)	8.1	4.5	10.3	4.3
Michoacán (c)	5.6	18.2	5.1	8.7
Morelos (c)	0.6			
Puebla (c)	14.9		10.3	4.3
Queretaro (c)	1.2			4.3
San Luis Potosí (c)	2.5		2.6	
Tlaxcala (c)	3.1	4.5	5.1	
Veracruz (c)	6.8	4.5		
Zacatecas (c)	3.1	4.5	2.6	
Campeche (s)	0.6			
Chiapas (s)	5.0	18.2		
Guerrero (s)	5.6	13.6		
Oaxaca (s)	9.3		7.7	4.3
Yucatán (s)	3.7	4.5		

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privada.

** Valor de insumos aplicados por hectárea cosechada.

(n) Norte del país

(c) Centrodel país

(s) Sur del país

Cuadro 4.h Resultados de la cuarta etapa del AC.

	11.1	11.2
No. de observaciones	222	23
MOD (% temporal)	82	0
MODTP *		
0 (%)	70.7	
1 (%)	11.7	
2 (%)	16.7	82.6
3 (%)	0.9	17.4
REND (ton./ha.)	1682.4	5010.4
TS (% criolla)	88	0
PROVEN (%)	0.381	0.862
INSSUP **	172373.4	414232.4
Distribución de las UP por Estado (%)		
Coahuila (n)	3.2	4.3
Chihuahua (n)	9.0	26.1
Sinaloa (n)	1.8	30.4
Sonora (n)		8.7
Distrito Federal (c)	3.2	
Guanajuato (c)	5.9	4.3
Hidalgo (c)	7.7	
Jalisco (c)	3.6	
México (c)	8.1	4.3
Michoacán (c)	6.8	8.7
Morelos (c)	0.5	
Puebla (c)	12.6	4.3
Queretaro (c)	0.9	4.3
San Luis Potosí (c)	2.3	
Tlaxcala (c)	3.6	
Veracruz (c)	5.4	
Zacatecas (c)	3.2	
Campeche (s)	0.5	
Chiapas (s)	5.4	
Guerrero (s)	5.4	
Oaxaca (s)	8.1	4.3
Yucatán (s)	3.2	

* 0 % unidades de temporal ejidal, 1 % unidades de temporal privada, 2 % unidades de irrigación ejidal, 3 % unidades de irrigación privada.

** Valor de insumos aplicados por hectárea cosechada.

(n) Norte del país

(c) Centrodel país

ANEXO METODOLOGICO

En éste anexo presentamos los principales rasgos formales de las metodologías usadas en el artículo, es decir, la de los análisis factorial y de conglomerados. Lo iniciamos con una exposición del método estadístico de "componentes principales", ya que el análisis factorial es su generalización.

Componentes principales (CP)

El método se usa para describir la variación de un conjunto de datos multivariados en términos de un conjunto de variables no correlacionadas, cada una de las cuales es una combinación lineal particular de las variables originales. Las nuevas variables son derivadas en orden de importancia decreciente, de tal forma que, por ejemplo, la primera componente principal contabiliza la mayor variación de los datos originales. El objetivo básico de éste tipo de análisis es ver si algunas de las primeras CP dan cuenta de la mayoría de la variación de los datos originales. Si esto es así, entonces esas variables pueden ser usadas para resumir los datos con una pérdida mínima de información, dando lugar a una reducción de la dimensionalidad de los datos. El análisis de CP es entonces de gran utilidad cuando se cuenta con una gran cantidad de variables, como es el caso de nuestra investigación empírica.

Formalmente, el problema que plantea la teoría de los CP es la representación de un vector m-dimensional en un espacio de dimensión menor. La idea básica consiste en encontrar alguna función lineal de los elementos del vector m-dimensional que capture la mayor parte de la variabilidad original.

Sea pues x un vector m-dimensional con media μ y varianza M. Se trata entonces de encontrar una combinación lineal de los elementos de x que tengan máxima varianza. Si

$$l' = a'x$$

es la combinación lineal que se busca, entonces

$$\mathrm{Var}(l') = a'M\,a$$

Sin embargo, el problema así planteado no tiene solución, porque si se elige

$$a^* = c\,a$$

para un valor fijo de a se puede hacer la varianza l' arbitrariamente elevada tomando valores grandes de c. Este problema de escala se resuelve a traves de la normalización convencional

$$a'a = 1$$

lo que significa que el vector resultante de la combinación lineal sea de tamaño unitario.

Este problema de maximización restringida es equivalente a la maximización del Lagrangiano

$$L = a'Ma + \lambda(1-a'a)$$

que, de acuerdo a las condiciones de primer orden, tiene una solución dada por

$$a'Ma = \lambda$$

por lo tanto λ debe ser uno de los valores característicos de M. Por el planteamiento del problema se debe elegir el valor característico mayor, digamos λ_1 , que tiene asociado un vector característico normalizado a_1. De esta manera se define el primer CP de los elementos de x como

$$\zeta = a_1 x$$

donde a_1 cumple con las siguientes condiciones

$$(\lambda_1 I-M)a_1 = 0 \qquad \text{y} \qquad a_1'a_1 = 1$$

y λ_1 es el vector característico mayor de

$$|\lambda I-M| = 0$$

donde, además, se tiene que

$$E(\zeta_1) = a_1'\mu \qquad \text{y} \qquad Var(\zeta_1) = \lambda_1$$

Si ahora se requiere otra combinación lineal de los elementos de x que tenga una varianza máxima, pero que además no esté correlacionado con ζ_1, se plantearía el mismo tipo de problema anterior, pero con la condición de que,

$$Cov(\zeta_1,\zeta) = a_1'Ma = 0 \qquad \text{donde} \qquad \zeta = a'x$$

siendo a el vector de la segunda combinación lineal. Nuevamente este problema es equivalente a la maximización del Lagrangiano

$$L = a'Ma + \lambda(1-a'Ma) + \mu_1 a_1'Ma$$

Es posible deducir la equivalencia de las condiciones de primer orden del problema planteado con la forma siguiente (véase Dhrymes, P.J., 1973, p. 55)

$$(\lambda I-M)a = 0$$

Por lo tanto, para definir la segunda combinación lineal se debe elegir el segundo mayor vector característico de M, digamos λ_2, que tendrá como vector característico a a_2. La segunda combinación lineal buscada está dada por

$$\zeta_2 = a_2'x$$

y tiene las siguientes propiedades

$$E(\zeta_2) = a_2'\mu \qquad Var(\zeta_2) = \lambda_2 \qquad Cov(\zeta_1, \zeta_2) = 0$$

Las variables ζ_1 y ζ_2 son llamadas, respectivamente, la primera y la segunda CP del vector x. De esta manera se puede proceder para derivar los CP que restan (ver <u>infra</u>. "Análisis Factorial").

Análisis factorial (AF)

En muchas ocasiones es imposible medir directamente los conceptos que son de nuestro interés (por ejemplo el grado de modernización de los agentes productivos). Sin embargo, si se pudiera hacer acopio de información de variables que pueden servir de indicadores del concepto o los conceptos en cuestión, podría utilizarse el AF, cuyo objetivo consiste en establecer la relación entre un conjunto de variables observables que son explicadas en términos de un número pequeño de variables no observadas, denominadas factores. El modelo básico del AF es esencialmente el mismo que el de la regresión múltiple, excepto que en el primero las variables observadas son regresionadas en factores no observables.

En este sentido, el AF no es más que una generalización del análisis de CP, esto debido a que los "factores" no son más que los CP del vector de variables observadas x.

El AF se compone de 4 etapas.

1. Primero, se calcula la matriz de correlaciones de todas las variables. Las que no parecen estar relacionadas con otras pueden identificarse en esta matriz y en sus estadísticos asociados. Con este procedimiento se puede saber si es pertinente aplicar el modelo de factores a los datos con los que se pretende hacer el estudio empírico o si, para efectuarlo, es necesario eliminar algunas variables.

2. En la segunda etapa se extraen factores, esto es, se determina el método y el número de factores necesarios. Con ella también es posible conocer qué tan correctamente el modelo de AF se ajusta a los datos.

3. La tercera etapa se refiere a la construcción de la matriz de factores.

4. En la última etapa el coeficiente de cada factor se calcula para cada caso.

La primera etapa del AF sirve para constatar que cada una de las variables tenga al menos un coeficiente de correlación "significativo" con el resto de las variables de la base de datos. Como es conocido, el coeficiente de correlación varía de -1 a 1, siendo el valor de cero el caso en donde no existe relación alguna entre las variables. No hay criterio único para determinar si la correlación entre variables es significativa o no lo es (en nuestro análisis tomamos valores mayores al valor absoluto de .25 como referencia).

Para elaborar la segunda etapa utilizamos básicamente el método de CP. Como mencionamos, el primer CP es la combinación que toma en consideración el mayor monto de varianza de los datos; el segundo captura el monto que sigue en cuanto a varianza; los componentes sucesivos explican progresivamente menores montos de varianza, estando todos, nuevamente, no correlacionados entre si.

Es posible obtener tantos CP como variables. Sin embargo, en la práctica se consideran sólo aquellos que tienen un valor característico mayor a 1. La justificación de esto radica en el hecho de que, en principio, todas las variables aportan una unidad a la varianza total y, además, el valor característico de los CP representa la varianza explicada por tal componente. Por lo tanto, un CP con valor característico menor que 1 no debe ser considerado debido a que explica menos que una variable en su forma original. Así entonces, la suma de la varianza explicada de los CP con valores característicos mayores que 1 --es decir de los componentes significativos-- y representa el grado de explicación del modelo.

La matriz de factores (tercera etapa del AF) representa los coeficientes de los factores significativos en términos de las variables originales. La matriz, además de proporcionar los índices estandarizados de las variables, determina con qué variables y en qué sentido están relacionados los factores con ellas. Lo último permite caracterizar al factor, esto es, determinar con qué variables está asociado.

En general el modelo para la variable estandarizada i-ésima se puede escribir de la siguiente manera

$$X_i = A_{i1}F_1 + A_{i2}F_2 + \ldots + A_{ik} + U_i$$

donde las F's son los factores comunes (o sea, los CP ζ's), la U es el factor único y las A's son las constantes usadas para combinar los k factores, donde obviamente $k < m$ (ver <u>supra</u>, "Componentes Principales"). Se supone, además, que los factores únicos no están

correlacionados entre sí, ni tampoco con los factores comunes.

Dado un vector x de dimensión m se tendrían, en principio, m factores. Sin embargo, de acuerdo a la teoría de los CP, las variables son standarizadas de forma tal que la varianza total explicada por el vector x es igual a m y, por otra parte, cada factor o CP tendrá una varianza igual a su correspondiente vector característico. Por lo tanto, la regla para determinar el número k de factores a utilizar es simple: sólo aquellos factores que tengan un valor característico mayor a 1 serán considerados.

Bajo estas suposiciones del modelo de AF, la varianza σ_i^2 de la variable x_i está dada por

$$\sigma_i^2 = \sum_{j=1}^{k} A_{ij} + \Psi_i$$

donde Ψ_i es la varianza de u_i. De acuerdo al modelo anterior se tiene que la varianza de una variable observada puede ser dividida en dos partes. La primera

$$h_i^2 = \sum_{j=1}^{k} A_{ij}$$

es conocida como "comunalidad" de las variables y representa las varianza compartida con las otras variables a través de los factores comunes. La segunda parte, Ψ_i, es denominada la varianza *única* o *específica* y se refiere a la variabilidad en x_i que no es compartida con otras variables.

La expresión general (no vectorial) para estimar el j-ésimo factor F es

$$F_j = \sum_{i=1}^{m} W_{ji} X_i = W_{j1} X_1 + W_{j2} X_2 + \ldots + W_{jm} X_m$$

donde las W's son conocidas como los coeficientes de los factores y m es el número de variables (ésto forma la cuarta etapa del AF).

Medidas de adecuación muestral. El AF es un método que no necesariamente arroja buenos resultados: es posible que los datos de la información original guarden entre sí poca relación y que, por lo tanto, no sea posible reducirla a un mínimo de factores.

Sin embargo, existen algunas medidas estadísticas para comprobar hasta qué grado el AF es adecuado. La primera está basada en el

siguiente argumento. Si las variables comparten factores comunes, los coeficientes de correlación parcial entre pares de variables debería ser pequeña cuando los efectos lineales de otras variables son eliminados. Las correlaciones parciales son entonces estimaciones de las correlaciones entre los factores únicos y deberían ser cercanas a cero si el AF es factible. El negativo de la matriz de coeficientes de correlación parcial es llamada matriz anti-imagen de correlación. Si la proporción de grandes coeficientes en esta matriz es alta, se debería reconsiderar la utilización del AF. Por ejemplo, si tenemos que las variables standarizadas de X_1 y X_2 son

$$X_1 = A_{11}F_1 + A_{12}F_2 + A_{13}F_{3....} + A_{1k}F_k + U_1$$

$$X_2 = A_{21}F_1 + A_{21}F_2 + A_{23}F_{3....} + A_{2k}F_k + U_2$$

donde F_i corresponde al i-ésimo factor y U_i es el factor único de la variable i. Definamos a $q_{i,j}$ como el coeficiente de correlación entre los factores únicos. Entonces, si en el ejemplo X_1 y X_2 comparten factores comunes, $q_{1,2}$, el coeficiente de correlación, debería ser pequeño. Por lo tanto, si en la matriz la proporción de coeficientes grandes es alta, debería de reconsiderarse el uso del modelo del AF.

Otra medida de adecuación de la muestra lo es el índice de Kaiser-Meyer-Olkin (KMO), que sirve para comparar las magnitudes de los coeficientes de correlaciones observados con los de correlación parcial. El índice es calculado de la manera siguiente:

$$KMO = \frac{\sum\sum_{i \neq j} r_{i,j}^2}{\sum\sum_{i \neq j} r_{i,j}^2 + \sum\sum_{i \neq j} q_{i,j}^2}$$

Si la suma de los cuadrados de las correlaciones parciales $q_{i,j}$ entre todos los pares de variables es pequeña cuando se compara con la suma al cuadrado de los coeficientes de las correlaciones observados $r_{i,j}$, entonces el índice será cercano a 1. En consecuencia, valores del índice cercanos a cero estarían indicando que el AF no procede.

También es posible calcular una medida de adecuación muestral (MSA) para cada variable en forma individual. En lugar de incluir todos los pares de variables en las sumatorias, se considerarían únicamente los coeficientes que involucran a tal variable. Para la i-ésima variable la medida de adecuación muestral es:

$$MSA_i = \frac{\sum_{i \neq j} r_{i,j}^2}{\sum_{i \neq j} r_{i,j}^2 + \sum_{i \neq j} q_{i,j}^2}$$

Esta medida se localiza en la diagonal de la matriz de correlación de anti-imagen y se necesitan valores razonablemente altos para la realización de un buen AF. Con este indicador se puede considerar la posibilidad de eliminar variables con pequeños valores de MSA.

De acuerdo a lo anterior, la "comunalidad" de una variable puede ser vista como el coeficiente de correlación múltiple entre ésta y el resto de variables. Indica, por lo tanto, el grado de asociación lineal entre las variables. En este sentido el cálculo de la comunalidad también sirve, junto a la MSA_i, como una medida de adecuación muestral para las variables consideradas individualmente.

Matriz de correlación reproducida. Uno de los supuestos básicos del AF es que las correlaciones observadas entre variables es debida al hecho de compartir factores comunes. En consecuencia, las correlaciones estimadas entre los factores y las variables pueden ser usadas para estimar las correlaciones entre las variables. En general, si los factores son ortogonales, el coeficiente de correlación estimado para las variables i y j es:

$$r_{ij} = \sum_{f=1}^{k} r_{fi} r_{fj} = r_{1i} r_{1j} \quad r_{2i} r_{2j} \quad \ldots . r_{ki} r_{kj}$$

donde k es el número de factores comunes y r_{fi} es la correlación entre el f-ésimo factor y la i-ésima variable. La diferencia entre las correlaciones observadas y las estimadas se denomina residual.

Toda la información anterior se encuentra contenida en la denominada matriz de correlación reproducida, la cual contiene en el triángulo inferior izquierdo los coeficientes de correlación estimados y en el superior derecho los residuales entre las correlaciones observadas y las estimadas. Cabe mencionar que las magnitudes de los residuos indican qué tan bien el modelo reproduce las correlaciones observadas.

En nuestro análisis empírico las magnitudes de los residuos fueron sustancialmente pequeñas, por lo que puede afirmarse que nuestro modelo reprodujo correctamente las correlaciones observadas.

Análisis de conglomerados (AC)

En estadística la búsqueda de grupos de objetos relativamente homogéneos se conoce como AC. En ésta técnica los conceptos de distancia y similitud son básicos. La distancia es una medida de qué tan lejos dos objetos se encuentran; por el contrario, la similitud mide su cercanía. En el AC estos conceptos son especialmente importantes puesto que los objetos son agrupados sobre la base de su cercanía. En nuestro caso se consideran, por ejemplo, las características de una unidad productiva (UP) y se supondrá "cerca" de otra si ambas son similares.

Para ver cómo la distancia es calculada, considérese el siguiente cuadro que contiene información hipotética sobre los valores de empleo y producción de dos UP.

	empleo	producción
UP 1	144	43
UP 2	157	48

entre las dos existe una diferencia de 13 en empleo y de 5 en producción. Esta combinación resulta en un índice simple o medida de distancia de varias formas. La más común de ellas es la denominada *distancia euclidiana*, la cual es la suma de la diferencia de los cuadrados de todas las variables. En este ejemplo la distancia euclidiana es $13^2 + 5^2 = 194$.

Esta medida tiene dos desventajas: depende de las unidades de medida de las variables y de la escala de las mismas. Esto se resuelve expresando a todas ellas en forma estandarizada, es decir, variables con media 0 y desviación estándar 1.

Supongamos que el total de la muestra de UP tiene las siguientes estadísticas descriptivas

	media	desviación estándar
empleo	132.40	30.25
producción	49.65	14.38

Por lo tanto, los valores estandarizados para las dos UP de nuestro ejemplo serían las siguientes:

	empleo	producción
UP 1	0.38	-0.46
UP 2	0.81	-0.11

siendo ahora la distancia euclidiana de 0.307.

El siguiente paso consiste en formar un conglomerado. Así como existen varias formas para medir la distancia entre objetos, hay varios métodos para combinar objetos en un conglomerado. El más

común es el llamado jerárquico, que, a su vez, tiene dos formas: el de aglomeración, en donde los conglomerados son formados agrupando casos en conglomerados más y más grandes hasta que todos los casos formen parte de un sólo conglomerado, y el de división, que comienza con todos los casos en un conglomerado y se va dividiendo hasta que existan tantos conglomerados como casos (en el estudio usamos el primero de ellos).

El criterio para decidir qué casos deben ser combinados en cada etapa está basado en una matriz de distancia o similitud entre pares de casos. El método más conocido es el denominado *vínculo simple*, también conocido como "vecino más cercano". Los dos primeros casos combinados son aquellos que tienen la distancia más pequeña entre sí. La distancia entre el nuevo conglomerado y los casos individuales es calculada como la distancia mínima entre un caso individual y un caso del conglomerado. Las distancias entre los casos que no han sido asignados no cambian. En cada etapa, la distancia entre dos conglomerados es la que hay entre sus puntos más cercanos.

REFERENCIAS

Bellon, Mauricio R., y J. Edward Taylor, "`Folk' soil taxonomy and the partial adoption of new seed varieties", *Economic Development and Cultural Change*, Vol. 41, No. 4 1993, pp. 763-86.

Brian, S. Everitt, y Graham Dunn, *Applied multivariate data analysis*, Edward Arnold, London, 1991.

Brush, B. Stephen. "Reconsidering the Green Revolution: Diversity and Stability in Craadle Areas of Crop Domestication", *Human Ecology*, Vol. 20, No. 2. June, 1992, pp. 145-67.

Brush, Stephen, B., Mauricio Bellon Corrales y Ella Schmidt, "Agricultural development and maize diversion in Mexico", *Human Ecology*, Vol. 16, No. 3, 1988, pp. 307-28.

Dhrymes, P. J., *Econometrics: Statistical foundations and applications*, Harper International Edition, Univ. of Pennsylvania, 1973, Cap. 1.

SARH, FIRA y BANRURAL, "Encuesta nacional de costos, coeficientes técnicos y producción agrícola", México, 1992.

Yúnez-Naude, A., E. Taylor y F. Barceinas, "Reflexiones sobre la biodiversidad genética de las semillas; problemas de análisis y el caso del maíz en México", en Yúnez-Naude, A. (compilador), *Medio ambiente: problemas y soluciones*, El Colegio de México, 1994, pp. 63-98.

SERIE DOCUMENTOS DE TRABAJO

The following working papers from recent year are still available upon request from:

Rocío Contreras,
Centro de Documentación, Centro de Estudios Económicos, El Colegio de México A.C., Camino al Ajusco # 20 C.P. 01000 México, D.F.

90/I Ize, Alain. "Trade liberalization, stabilization, and growth: some notes on the mexican experience."

90/II Sandoval Musi, Alfredo. "Construction of new monetary aggregates: the case of Mexico."

90/III Fernández, Oscar. "Algunas notas sobre los modelos de Kalecki del ciclo económico."

90/IV Sobarzo, Horacio E. "A consolidated social accounting matrix for input-output analysis."

90/V Urzúa, Carlos M. "El déficit del sector público y la política fiscal en México, 1980 - 1989."

90/VI Romero, José. "Desarrollos recientes en la teoría económica de la unión aduanera."

90/VII García Rocha, Adalberto. "Note on mexican economic development and income distribution."

90/VIII García Rocha, Adalberto. "Distributive effects of financial policies in Mexico."

90/IX Mercado, Alfonso and Taeko Taniura "The mexican automotive export growth: favorable factors, obstacles and policy requirements."

91/I Urzúa, Carlos M. "Resuelve: a Gauss program to solve applied equilibrium and disequilibrium models."

91/II Sobarzo, Horacio E. "A general equilibrium analysis of the gains from trade for the mexican economy of a North American free trade agreement."

91/III Young, Leslie and José Romero. "A dynamic dual model of the North American free trade agreement."

91/IV Yúnez-Naude, Antonio. "Hacia un tratado de libre comercio norteamericano; efectos en los sectores agropecuarios y alimenticios de México."

91/V Esquivel, Hernández Gerardo. "Comercio intraindustrial México-Estados Unidos."

91/VI Márquez, Colín Graciela. "Concentración y estrategias de crecimiento industrial."

92/I Twomey, J. Michael. "Macroeconomic effects of trade liberalization in Canada and Mexico."

92/II Twomey, J. Michael. "Multinational corporations in North America: Free trade intersections."

92/III Izaguirre Navarro, Felipe A. "Un estudio empírico sobre solvencia del sector público: El caso de México."

92/IV Gollás, Manuel y Oscar Fernández. "El subempleo sectorial en México."

92/V Calderón Madrid, Angel. "The dynamics of real exchange rate and financial assets of privately financed current account deficits"

92/VI Esquivel Hernández, Gerardo. "Política comercial bajo competencia imperfecta: Ejercicio de simulación para la industria cervecera mexicana."

93/I Fernández, Jorge. "Debt and incentives in a dynamic context."

93/II Fernández, Jorge. "Voluntary debt reduction under asymmetric information."

93/III Castañeda, Alejandro. "Capital accumulation games."

93/IV Castañeda, Alejandro. "Market structure and innovation a survey of patent races."

93/V Sempere, Jaime. "Limits to the third theorem of welfare economics."

93/VI Sempere, Jaime. "Potential gains from market integration with individual non-convexities."

93/VII Castañeda, Alejandro. "Dynamic price competition in inflationary environments with fixed costs of adjustment."

93/VIII Sempere, Jaime. "On the limits to income redistribution with poll subsidies and commodity taxation."

93/IX Sempere, Jaime. "Potential gains from integration of incomplete markets."

93/X Urzúa, Carlos M. "Tax reform and macroeconomic policy in Mexico."

93/XI Calderón, Angel. "A stock-flow dynamic analysis of the response of current account deficits and GDP to fiscal shocks."

93/XII Calderón, Angel. "Ahorro privado y riqueza financiera neta de los particulares y de las empresas en México."

93/XIII Calderón, Angel. "Política fiscal en México."

93/XIV Calderón, Angel. "Long-run effects of fiscal policy on the real levels of exchange rate and GDP."

93/XV Castañeda, Alejandro. "On the invariance of market innovation to the number of firms. The role of the timing of innovation."

93/XVI Romero, José y Antonio Yúnez. "Cambios en la política de subsidios: sus efectos sobre el sector agropecuario."

94/I Székely, Miguel. "Cambios en la pobreza y la desigualdad en México durante el proceso de ajuste y estabilización".

94/II Calderón, Angel. "Fiscal policy, private savings and current account deficits in Mexico".

94/III Sobarzo, Horacio. "Interactions between trade and tax reform in Mexico: Some general equilibrium results".

94/IV Urzúa, Carlos. "An appraisal of recent tax reforms in Mexico". (Corrected and enlarged version of DT. Núm. X-1993)

94/V Sáez, Raúl E. and Carlos Urzúa. "Privatization and fiscal reform in Eastern Europe: Some lessons from Latin America".

94/VI Feliz, Raúl. "Terms of trade and labour supply: A revision of the Laursen-Metzler effect".

94/VII Feliz, Raúl and John H. Welch. "Cointegration and tests of a classical model of inflation in Argentina, Bolivia, Brazil, Mexico, and Peru".

94/VIII Sempere, Jaime. "Condiciones para obtener ganancias potenciales de liberalización de comercio".

94/IX Sempere, Jaime y Horacio Sobarzo. "La descentralización fiscal en México: Algunas propuestas".

94/X Sempere, Jaime. "Are potential gains from economic integration possible with migration?".

94/XI Gollás, Manuel. "México 1994. Una economía sin inflación, sin igualdad y sin crecimiento".

95/I Schettino, Macario. "Crecimiento económico y distribución del ingreso".

95/II Schettino, Macario. "A function for the Lorenz curve".

95/III Székely P., Miguel. "Economic Liberalization, Poverty and Income Distribution in Mexico".

EL COLEGIO DE MEXICO

3 905 0567940 S