

UNA CLASIFICACION DE LOS BANCOS PRIVADOS ARGENTINOS

Tamara Burdóssio y Patricia Botargues, con la
colaboración de Laura D'Amato.

INTRODUCCION

Los bancos pueden ser vistos como firmas multiproducto que requieren, en general, de múltiples plantas (sucursales) para llevar adelante su operatoria. Se observa, sin embargo, que existe una heterogeneidad importante entre bancos en cuanto a su especialización productiva y asociado a ello al número de sucursales con que operan.

Esta diversidad entre firmas plantea la necesidad de encontrar clasificaciones que permitan agrupar a los bancos en conjuntos lo más homogéneos posible respecto de sus características económicas como tipo y diversidad de servicios, número, tamaño y tipo de clientes (familias o firmas, por ejemplo), costos de fondeo, número de sucursales, etc..

De hecho, autores como Arnaudo (1977) y Feldman (1978), que estiman funciones de costos para los bancos argentinos, señalan las limitaciones de la clasificación institucional establecida por el Banco Central. En tanto Arnaudo propone una clasificación alternativa, Feldman plantea que para realizar estudios de costos resultan de mayor utilidad desagregaciones basadas en criterios de homogeneidad de acuerdo al grado de especialización productiva y número de plantas. Muestra, a través de ejemplos numéricos, que bancos privados de Capital Federal y bancos extranjeros, por un lado, y bancos públicos y privados del interior, por el otro, no se diferencian notablemente si se comparan niveles medios para variables como costos totales/total de operatoria, salarios medios, producto por empleado y otros.

El sistema financiero argentino se encuentra actualmente en un proceso de expansión. El nivel de actividad de los bancos ha crecido sostenidamente a partir de 1991. Ello es consecuencia de la remonetización de la economía generada por la estabilidad, que permitió la recomposición del crédito. Esta expansión se opera en un contexto internacional en el que los bancos tienden a ampliar

sus negocios frente al surgimiento de nuevas formas de intermediación financiera (como por ejemplo los fondos de pensión), que compiten con la operatoria tradicional de intermediación bancaria. Por otra parte la remonetización se ha operado en un contexto regulatorio distinto al prevaleciente hasta la Convertibilidad: tasas de interés no reguladas, dolarización de los activos monetarios, condiciones favorables al crecimiento del mercado de capitales local.

El hecho de que el negocio bancario se encuentre en un proceso de ampliación no implica que la tendencia sea necesariamente hacia una mayor diversificación para el conjunto de los bancos. Se puede pensar que es más probable que los bancos grandes deriven economías de la diversificación de sus actividades, en tanto para los de menor tamaño lo más eficiente sea una mayor especialización. El contar con una clasificación de bancos que contenga grupos homogéneos en cuanto a su tecnología puede resultar de utilidad para encarar trabajos de investigación dirigidos al análisis microeconómico de las firmas bancarias (estudios de costos, rentabilidad, etc.). Si el análisis no toma

en cuenta la heterogeneidad antes mencionada las conclusiones a las que se arribará no serán tan precisas como si se trabajara por separado con cada uno de los grupos homogéneos de la población.

La clasificación que se propone en este trabajo se plantea como una primera agrupación posible de los bancos privados que sabemos no agota la posibilidad de posteriores clasificaciones, principalmente entre los bancos que aquí se clasifican como minoristas y que, si bien resultan homogéneos respecto de las variables utilizadas para clasificar, presentan diferencias de tamaño y de grado de diversificación importantes en el interior del grupo.

Se consideraron en este análisis sólo los bancos privados.

Las variables que se utilizan para clasificar a los bancos buscan en principio captar características básicas que permiten diferenciar las operatorias mayorista y minorista.

Los bancos minoristas, si bien pueden ofrecer una amplia gama de servicios, centran su actividad en la operatoria tradicional de intermediación. Ofrecen servicios de liquidez y crédito a clientes de distintos tipos. Una característica tecnológica importante de este tipo de operatoria es que requiere de un número importante de sucursales para su funcionamiento. El número de casas bancarias es en este caso un indicador de escala de operación.

Los bancos mayoristas tienen clientes de mayor tamaño, principalmente firmas o bancos, y no familias. Su principal fuente de fondos prestables no proviene de la captación de depósitos, ya que ésta no es una actividad principal para ellos. En el contexto de apertura financiera que existe hoy en Argentina, los bancos mayoristas operan en mercados competitivos, en el sentido que enfrentan una tasa dada (la tasa de interés internacional más alguna prima de riesgo) ya que es probable que sus clientes puedan acceder a fuentes externas de financiamiento. Deben por lo tanto fondearse con recursos menos costosos que los depósitos. Este menor costo de fondeo parece reflejarse en menores tasas activas.

INTRODUCCION ESTADISTICA

En este trabajo se introducen técnicas estadísticas de agrupamiento ("clustering analysis") para separar a la población de bancos privados en mayoristas y minoristas.

Las técnicas de agrupamiento del análisis multivariado ("cluster analysis") intentan agrupar objetos en base a un conjunto de datos referidos a los mismos. El propósito del "clustering" es ubicar a los objetos o individuos de una población que se supone heterogénea dentro de un grupo o "cluster" sugerido por los datos, no conocido a priori. Esto se

realiza de forma tal que los objetos resulten "similares" (o parecidos) dentro del mismo grupo, por lo cual éste resultará homogéneo, y distintos entre diferentes grupos. Esencialmente, los "clusters" deben exhibir las propiedades de aislamiento externo y cohesión interna. El aislamiento externo requiere que las entidades en un grupo deben estar separadas de las entidades en otros grupos por áreas vacías. La cohesión interna requiere que las entidades dentro del mismo grupo deben ser similares entre sí (Milligan, 1980).

Entre las múltiples aplicaciones que tienen estas técnicas estadísticas se puede mencionar a la antropología, sociología, investigación de mercados, medicina, economía, etc. (Seber, 1984). El desarrollo principal de estos métodos se dio a partir de la década del '60 con el advenimiento de la computación.

PLANTEO DEL PROBLEMA

La definición que se utilizó hasta el presente del grado de mayoreo y minoreo de un banco se basa en el tamaño medio de préstamos (es decir préstamos totales / número de cuentas activas, PRESROM). Se consideraba que un banco era mayorista si su tamaño medio de préstamos superaba cierto valor (Botargues y Burdisso, 1994).

Se observó que la clasificación de los bancos privados a partir de la distribución de la variable tamaño medio de préstamos no alcanzaba para caracterizar adecuadamente el grado de mayoreo y minoreo. El criterio empleado para identificar a ambos grupos dependía de un valor crítico que se establecía en la distribución del tamaño medio de préstamos. La clasificación así realizada evidentemente puede ser criticada por arbitraria. La técnica estadística de agrupamiento utilizada en este trabajo posee la ventaja de no basarse en la fijación arbitraria de valores críticos para las variables (Berlage y Terweduwe, 1988).

Otro intento para realizar la clasificación se basó no sólo en el tamaño medio de préstamos de los bancos sino también en el número de casas. Al respecto Penas y Streb (1994) estratificaron el segmento de bancos privados mediante técnicas estadísticas multivariadas utilizando la transformación logarítmica de ambas variables. Los dos grupos que obtuvieron resultaron similares a los determinados a partir de la distribución de la variable tamaño medio de préstamos.

La aplicación de las técnicas de agrupamiento empleando sólo la variable tamaño medio de préstamos da indicios de la existencia de dos o cuatro grupos en la población. La clasificación en dos grupos tiene un valor crítico excesivamente bajo por lo cual el grupo que contiene a los bancos que se supone son mayoristas resulta ser más numeroso que el que contiene a los minoristas, cuando se sabe que esto no es así. Algo similar ocurre con la clasificación en cuatro grupos.

Con el objetivo de identificar a los grupos de bancos existentes en la población se definieron dos nuevas variables. Por un lado el cociente depósitos / activo (DT_ACT), que se incluyó en base a la siguiente hipótesis: la banca mayorista otorga créditos a empresas en un mercado más competitivo y de mayores plazos que la minorista. Luego el fondeo de estas operaciones no se basará

en depósitos dado que los mismos presentan una mayor rotación y probablemente tienen un mayor costo explícito que el resto de sus pasivos. Por ello debería observarse para los bancos mayoristas un valor relativamente bajo de la variable DT_ACT. Por otro lado, el grueso de los ingresos de este tipo de bancos no debería provenir de la operatoria "tradicional" del sistema financiero, entendiendo por ella la relacionada con el manejo de créditos y obligaciones, sino más bien de operatorias intensivas en comisiones como el manejo de carteras de títulos y acciones, colocación de obligaciones negociables, etc. Este hecho debería reflejarse en la estructura de los ingresos por servicios en un mayor peso de las comisiones antes mencionadas respecto de las comisiones por obligaciones, créditos, cajas de seguridad y

garantías, más asociadas a la operatoria tradicional de intermediación. Por este motivo se construyó una variable ISOT_IS*, que mide la participación de los ingresos por servicios ligados a la operatoria tradicional respecto del total de los ingresos por servicios, esperándose valores relativamente pequeños de esta variable para los bancos mayoristas.

Se estudiaron las dos nuevas variables construidas para los meses de agosto, octubre y diciembre de 1992, febrero, abril, junio y diciembre de 1993. Ambas variables tienen distribuciones bastante asimétricas en las que sería posible observar dos modas, lo cual podría deberse a la existencia de dos distribuciones diferentes superpuestas. Se estudió además la distribución de la variable PRESROM y no sólo es muy asimétrica sino que además los valores del último cuartil se encuentran muy aislados entre sí, separados por intervalos vacíos.

Se esperaba que la variable ISOT_IS fuera estable en el tiempo ya que las comisiones son un monto fijo por cuenta bancaria, y el número de cuentas no varía significativamente en el corto plazo. Se comprobó esta suposición realizando tests no paramétricos de igualdad de medias, tomando de a dos los diferentes meses considerados. No ocurrió lo mismo con la variable de fondeo DT_ACT. Por este motivo la clasificación se realizó para cada mes y también utilizando las variables promediadas.

(*)ISOT_IS: Ingresos por servicios por operatoria tradicional/Ingresos por servicios

Ingresos por servicios = Ingresos por servicios por operatoria tradicional + ingresos por servicios por operatoria no tradicional.

Ingresos por servicios por operatoria tradicional = Comisiones vinculadas con obligaciones + comisiones vinculadas con créditos + comisiones por garantías otorgadas + alquiler de cajas de seguridad.

Ingresos por servicios por operatoria no tradicional = Comisiones vinculadas con valores mobiliarios + comisiones por operaciones de exterior y cambio + otros en moneda argentina + otros en moneda extranjera.

La correlación de Spearman entre las dos variables construidas es aproximadamente 0.5, y debe destacarse que la correlación entre cada una de ellas con el tamaño medio de préstamos es superior a 0.7.

Se intentó incluir la variable tamaño medio de préstamos para realizar el agrupamiento conjuntamente con las nuevas variables definidas. Para ello fue necesario estandarizar PRES PROM, ya que su rango de variación es de 0 a 5 millones de pesos, mientras que DT_ACT e ISOT_IS varían de 0 a 1. La estandarización se realizó dividiendo a las tres variables mencionadas por sus correspondientes rangos intercuartílicos. No se utilizaron sus desvíos standard debido a la asimetría mencionada en las distribuciones de las mismas. Pero mediante la utilización conjunta de PRES PROM, DT_ACT e ISOT_IS no fue posible realizar ningún tipo de agrupamiento, ya que no se obtienen evidencias de la existencia de diferentes grupos en la población respecto de las tres variables (ver Anexo Estadístico, sección Número de "Clusters"). Si bien utilizando PRES PROM con cada una de las otras dos variables hay indicios de la existencia de dos grupos de bancos, en ambos casos se obtiene que el "cluster" más numeroso es el que contiene a los bancos que se suponen mayoristas.

Debe mencionarse que en la literatura se recomienda evitar transformaciones de escala de las variables. Por un lado, éstas pueden atemperar las diferencias entre los grupos con respecto a las variables que mejor los discriminan. Por otro lado, la estandarización se realiza respecto de la población total, cuando lo correcto sería estandarizar las variables dentro de cada grupo, ya que los grupos son diferentes entre sí, pero esto por supuesto es imposible (Seber, 1984; Cormack, 1971). Por lo tanto se continuó el análisis solamente con las dos nuevas variables DT_ACT e ISOT_IS, las cuales tienen el mismo rango de variación.

No hay motivos para esperar que los grupos a identificar sean compactos en cuanto a forma, tamaño o dispersión. En el diagrama de dispersión adjunto (Gráfico 1) se puede observar que no hay

presentes "clusters" compactos en los datos. Por esta razón se usó un método de clasificación no paramétrico ("two-stage density linkage", ver Anexo Estadístico), (SAS/STAT, 1988).

Para precisar el número de grupos existentes en la población de bancos privados se utilizó un criterio sugerido en el producto estadístico SAS (SAS/STAT, 1988) en combinación con un criterio empleado por Wong y Lane (1983) (ver Anexo Estadístico). Se determinó que el número de grupos es dos. Esto indica que en base a las nuevas variables empleadas se puede hablar de dos grupos con características muy diferentes, a los que se llamará grupo de bancos mayoristas y grupo de bancos minoristas.

La clasificación de los bancos privados se realizó para cada uno de los meses de agosto, octubre y diciembre de 1992, y febrero, abril, junio y diciembre de 1993. Se comprobó que los "clusters" mantenían su composición prácticamente inalterada a través del tiempo.

En base a este resultado se promediaron los datos correspondientes a agosto, octubre y diciembre de 1992, y febrero, abril y junio de 1993, construyéndose las variables promedio DT_ACT e ISOT_IS. Los grupos obtenidos con estas variables fueron concordantes con los observados a través del tiempo.

Cabe destacar que la distribución de las variables utilizadas dentro de cada uno de los grupos determinados no difieren significativamente de la normal.

En la Tabla 1 se muestran las correlaciones de Spearman entre las variables DT_ACT e ISOT_IS para la población total y dentro de cada uno de los grupos para los meses considerados. Se observa que mientras el valor es cercano a 0.5 para todos los bancos, dentro de los grupos las correlaciones son cercanas a 0. Tomando en cuenta que la distribución de las variables dentro de los grupos se puede considerar normal, esto en conjunto con las correlaciones cercanas a 0, confirmaría que ambos grupos conforman

"clusters" según la definición citada por Cormack (1971) de "clusters" como conjuntos de entidades dentro de los cuales las variables o atributos son independientes.

TABLA 1.

Correlaciones de Spearman entre las variables DT_ACT e ISOT_IS con sus correspondientes niveles de significación indicados entre paréntesis

AGO-92	OCT-92	DIC-92	FEB-93	ABR-93	JUN-93	DIC-93	PROM9293*
--------	--------	--------	--------	--------	--------	--------	-----------

Total de Bancos Privados

0.51	0.47	0.44	0.46	0.50	0.52	0.59	0.52
(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Bancos Minoristas

0.15	0.02	0.11	0.09	0.02	0.14	0.23	0.14
(0.149)	(0.849)	(0.267)	(0.357)	(0.844)	(0.161)	(0.026)	(0.160)

Bancos Mayoristas

-0.07	-0.09	-0.34	0.09	0.31	0.19	0.21	0.01
(0.730)	(0.652)	(0.099)	(0.688)	(0.095)	(0.347)	(0.295)	(0.948)

* Se refiere a los promedios de las variables DT_ACT e ISOT_IS para agosto, octubre y diciembre de 1992, y febrero, abril y junio de 1993.

Se debe mencionar que una característica que presentan los bancos clasificados como mayoristas es que tienen una o dos casas. Además se observó que este mismo grupo de bancos tiene valores altos de la variable PRESPROM.

En la Tabla 2. se muestran algunas medidas descriptivas de las variables promedio DT_ACT e ISOT-IS para el total de los bancos privados, para el grupo de bancos que aquí se llaman mayoristas y para el grupo de bancos que se llaman minoristas. Se observa que los parámetros de posición (media, mediana) difieren mucho entre los grupos, quitando significado a la media (mediana) de la población total. Asimismo los parámetros de dispersión (desvío standard, rango intercuartílico) son menores dentro de cada grupo que para el total.

TABLA 2.

Medidas Descriptivas

Total de Bancos Privados

	DT_ACT*	ISOT_IS*
Media	0.55	0.57
Desvío Standard	0.21	0.33
Mediana	0.61	0.63
Rango Inter cuartílico	0.26	0.54

Bancos Mayoristas

	DT_ACT*	ISOT_IS*
Media	0.24	0.11
Desvío Standard	0.15	0.11
Mediana	0.25	0.09
Rango Inter cuartílico	0.24	0.15

Bancos Minoristas

	DT_ACT*	ISOT_IS*
Media	0.66	0.72
Desvío Standard	0.11	0.21
Mediana	0.68	0.76
Rango Inter cuartílico	0.14	0.31

Promedio de los datos correspondientes a agosto, octubre y diciembre de 1992 y febrero, abril y junio de 1993.

A modo de comparación entre los dos grupos obtenidos mediante la metodología utilizada, se consideró un indicador de tasa activa, la tasa activa implícita, que se calcula como el cociente entre los intereses por préstamos y/o ajustes devengados en el período y la semisuma de saldos del período y del período inmediato anterior. Si bien este indicador tiene limitaciones por estar calculado como una tasa media que agrupa préstamos con distintos plazos y distintos riesgos, y quizá no es un buen indicador de la tasa de mercado, sí puede servir para hacer comparaciones entre entidades.

Se calcularon tasas medias por grupo, y mediante tests de diferencias de medias se evaluó si las mismas son distintas entre sí. En la Tabla 3 se presentan las tasas medias por grupo con sus correspondientes errores standard, las desviaciones standard por grupo, los valores del estadístico F para testear igualdad de varianzas en ambos grupos con sus correspondientes niveles de significación, y los valores de los estadísticos T para testear igualdad de medias bajo las hipótesis de igualdad de varianzas y de varianzas distintas. Con respecto al test F de igualdad de varianzas, se observa que con un nivel de significación 0.05 se rechaza esta hipótesis en todos los meses considerados. Los tests de igualdad de medias en ambos grupos confirman que los indicadores de tasa implícita de los bancos mayoristas y de los bancos minoristas son significativamente diferentes en todos los meses analizados. Aún más, a partir de esto se concluye también que la tasa media de los mayoristas es significativamente menor que la de los minoristas.

TABLA 3.

Tasa Activa Implícita en pesos
Tests de Diferencias de Medias

	AGO-92	OCT-92	DIC-92	FEB-93	ABR-93	JUN-93
Media Min.	3.60 (0.11)	3.56 (0.09)	4.03 (0.08)	3.11 (0.17)	3.00 (0.08)	2.84 (0.08)
Media May.	1.76 (0.13)	1.75 (0.11)	3.27 (0.23)	1.42 (0.12)	1.30 (0.24)	1.26 (0.23)
St.D. Min.	1.05	0.93	0.79	1.69	0.81	0.84
St.D. May.	0.74	0.62	1.31	0.69	1.37	1.31
Igualdad de Varianzas F	2.01 (0.031)	2.23 (0.014)	2.75 (0.000)	5.97 (0.000)	2.86 (0.000)	2.45 (0.001)
Sup.Varianzas Distintas T	10.74 (0.000)	12.34 (0.000)	3.14 (0.003)	7.91 (0.000)	6.60 (0.000)	6.40 (0.000)
Sup. Ig. de Varianzas T	9.04 (0.000)	10.14 (0.000)	3.97 (0.000)	5.47 (0.000)	8.43 (0.000)	7.90 (0.000)

CONCLUSIONES

Este trabajo se propuso encontrar una clasificación de los bancos privados en base a su especialización productiva usando técnicas estadísticas de agrupamiento. Para ello se utilizaron dos variables: un indicador de fondeo (proporción de los depósitos respecto del activo) y un indicador que mide el peso de los ingresos asociados a la operatoria tradicional en los ingresos por servicios.

Se encontró que respecto de estas variables existen dos grupos de bancos privados bien diferenciados, a los que se denominó grupo de bancos mayoristas y grupo de bancos minoristas.

Se encontró además que la tasa activa implícita es significativamente menor para los bancos llamados mayoristas que para los minoristas.

Si bien ambos grupos son homogéneos respecto de las variables utilizadas aquí, se cree que dentro del grupo de bancos minoristas persisten diferencias de tamaño y grado de diversificación que sugieren la necesidad de continuar investigando sobre este grupo de entidades.

ANEXO ESTADISTICO

METODOLOGIA

"Clustering" es el proceso de agrupar objetos similares utilizando datos referidos a los mismos. Es parte del proceso científico general de buscar patrones en los datos y luego tratar de construir leyes que expliquen estos patrones. Los objetivos del "cluster analysis" son variados e incluyen actividades muy diferentes, siendo ejemplos notables la clasificación de plantas y animales (taxonomía), y la clasificación de enfermedades (Seber, 1984).

Los "clusters" o grupos pueden ser descriptos como regiones continuas de puntos de alta densidad separadas éstas de otras regiones de puntos de baja densidad. Los "clusters" descriptos de esta manera son referidos como "clusters naturales" y una inspección visual en dos o tres dimensiones podría ayudar a detectar los grupos naturales. Una ventaja de considerar los "clusters" de este modo es que no se restringe tan rígidamente la forma de los mismos como lo hacen otras definiciones propuestas.

El problema de las técnicas de "clustering" o agrupamiento es que la mayoría de las mismas encuentra "clusters" de determinada forma. Sin embargo, no hay razón a priori para creer que los grupos presentes en los datos tengan una forma particular, y el uso de una técnica de "clustering" no adecuada podría imponer una estructura particular sobre los datos más que hallar la verdadera estructura subyacente en los mismos (Everitt, 1980). La mayoría de las técnicas de agrupamiento no tienen dificultad en encontrar "clusters" compactos o esféricos pero sí en hallar los verdaderos "clusters naturales" presentes en los datos.

Con el fin de elegir la técnica de "clustering" apropiada es importante tener en cuenta que la mayoría de éstas tienden a encontrar "clusters" con determinadas características relativas al tamaño, forma y dispersión. Por ejemplo, los métodos basados en cuadrados mínimos como el "k-means" y el método de Ward tienden a encontrar "clusters" con el mismo número de elementos en cada grupo. El método "average-linkage" tiene tendencia a determinar "clusters" de igual varianza.

La mayoría de los métodos de agrupamiento son incapaces de detectar "clusters" de forma irregular o muy elongados. Los métodos con menor sesgo son aquéllos basados en estimadores de la densidad no paramétricos tales como el "density-linkage" y sus variantes. Por lo tanto, si no se tiene conocimiento a priori del tipo de grupos presentes en los datos se aconseja incluir en el análisis al menos un método no paramétrico (SAS/STAT, 1988).

Todos los métodos de agrupamiento que hasta aquí se han citado excepto el "k-means" son los llamados jerárquicos aglomerativos (Seber, 1984; SAS/STAT, 1988). Estos métodos comienzan con tantos grupos como individuos se tienen. Los dos individuos más cercanos se fusionan en un "cluster" que reemplaza a los "clusters" unitarios. Esta técnica de fusionar los dos "clusters" más cercanos se repite hasta la obtención de un único "cluster" que contiene a todos los individuos de la población. La

fusión de los grupos más cercanos se hace en base a una medida de distancia. Las distintas formas de definir esta distancia da origen a los diferentes métodos de agrupamiento.

Método "Density-linkage"

Su nombre proviene de la utilización de estimadores de la densidad de probabilidad no paramétricos para construir la medida de distancia. Utilizando la técnica del "kth-nearest-neighbor" desarrollada por Wong y Lane (1983), la densidad se estima de la siguiente forma: dado un individuo x y un valor k se define $r_k(x)$ como la distancia del punto x a la k -ésima observación más cercana. Se considera entonces la esfera con centro en x y radio $r_k(x)$. Luego la densidad estimada en x , llamada $f(x)$, es la proporción de individuos que están dentro de la esfera dividido por el volumen total de la esfera. En base a esta densidad estimada $f(x)$ se define una medida de disimilaridad d^* de la siguiente forma:

$$d^*(x, y) = \frac{1}{2} \left(\frac{1}{f(x)} + \frac{1}{f(y)} \right) \text{ si } d(x, y) \leq \max(r_k(x), r_k(y))$$

$$= \infty \quad \text{a fuera}$$

a partir de la cual se agruparán los individuos. Estos se fusionan hasta que todos ellos son asignados a algún "cluster" y luego se comienza a fusionar los "clusters". Para ello se define distancia entre dos "clusters" C_1 y C_2 como la mínima distancia d^* que existe entre los individuos de C_1 y los individuos de C_2 .

El número de "clusters" resultantes de este método depende directamente del valor k que se elige para estimar la densidad (la elección de k se detalla en la sección Número de "Clusters").

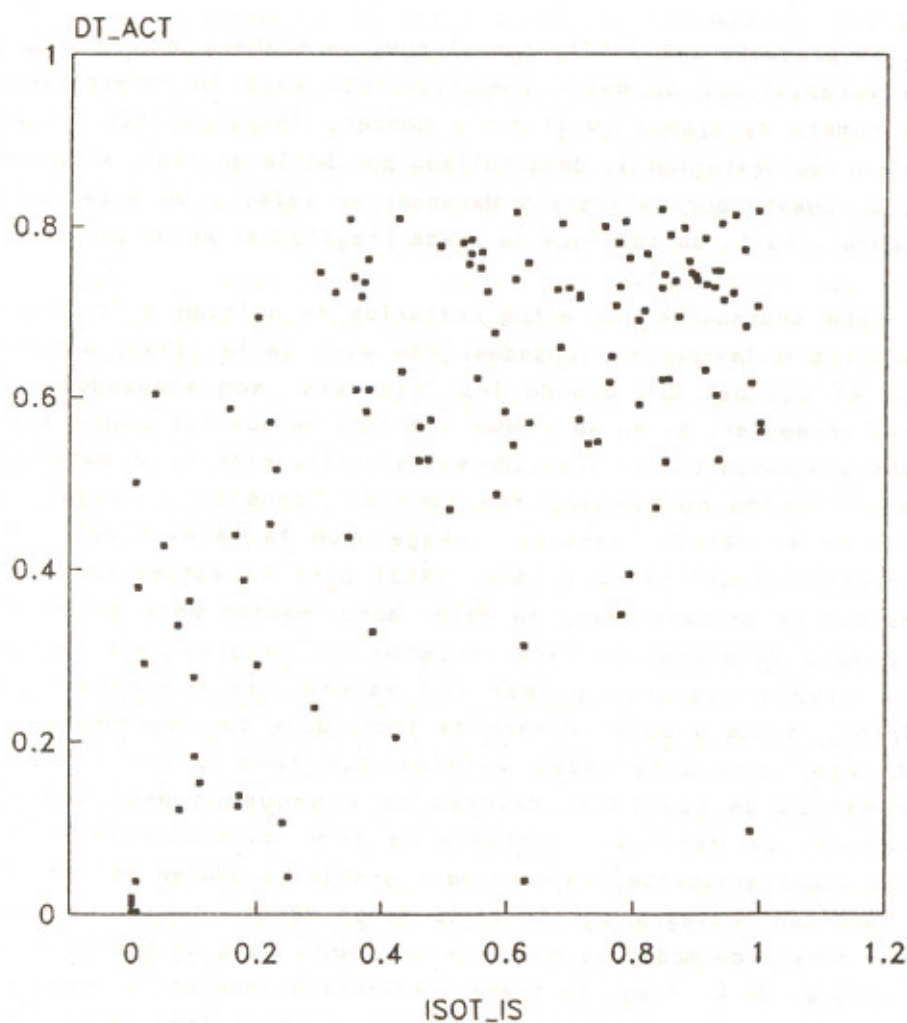
Esta forma de fusionar o agrupar primero los individuos y recién después fusionar los "clusters" es una variante del método "density-linkage" llamada "two-stage density-linkage".

Un problema común a todas las técnicas de agrupamiento es la determinación del número de "clusters" o grupos en la población. La decisión no es sencilla y existen una treintena de tests estadísticos para estimar el número de grupos (Milligan y Cooper, 1985).

El producto SAS / STAT con el cual se trabaja provee tres de los estadísticos de mejor comportamiento para la determinación del número de grupos (Milligan y Cooper, 1985): el CCC ("cubic clustering criterion"), desarrollado por Sarle en 1983; el pseudo F^2 , propuesto por Calinski y Harabasz en 1974; y un estadístico de Duda y Hart de 1973 que se puede transformar en un pseudo t^2 .

Debe destacarse que estos criterios se aplican a "clusters" compactos o levemente elongados. Más aún, Sarle (1983) desaconseja el uso del CCC cuando los "clusters" son elongados o de forma irregular. Si no se tienen indicios de que los grupos buscados son compactos lo indicado es la utilización de un método de clasificación no paramétrico como el "density linkage". Al utilizar el método "density linkage" con la técnica del "kth-nearest-neighbor" (Wong y Lane, 1983) para la estimación de la densidad de probabilidad, la mejor aproximación para determinar el número de grupos se hace variando los valores de k . Wong y Lane (1983) sugieren probar con valores de k alrededor de $2 \cdot \ln(N)$, donde N es el número de individuos considerados en el análisis. Los diferentes valores que toma k son llamados parámetros de suavidad. Valores de k pequeños producen estimadores de densidad "dentados" y como consecuencia muchas modas. Contrariamente, valores de k grandes producen estimadores de densidad suaves y por lo tanto pocas modas. A raíz de esto, si el número de modas se mantiene constante para un amplio rango de valores de k , luego se tiene evidencia de que en la población hay por lo menos esa cantidad de modas. Luego a cada valor modal le corresponde un subgrupo de la población llamado "cluster modal".

GRAFICO 1
Diagrama de dispersión



REFERENCIAS

- * Arnaudo, A.A.A.: "Economías de Operación y Economías de Escala en el Sistema Bancario Comercial Argentino. 1960 - 1972", Ensayos Económicos, No. 2, junio 1977.
- * Berlage, L. y Terweduwe, D.: "The Classification of Countries by Cluster and by Factor Analysis", World Development, Vol. 16, No. 12, 1988.
- * Botargues, P.B. y Burdisso, T.: "Concentración del Sistema Financiero", Indicadores del Sistema Financiero, B.C.R.A., Febrero - Abril 1994.
- * Cormack, R.M.: "A Review of Classification", J.R.Stat.Soc.A, 134, 321-367, 1971.
- * Everitt, B.: Cluster Analysis. Second Edition. John Wiley & Sons, 1980.
- * Feldman, E.: "Costos Bancarios: Estimaciones mediante Análisis de Corte Transversal y Series de Tiempo", Series de Estudios Económicos, C.E.M.y B., B.C.R.A., No. 30, 1978.
- * Milligan, G.W.: "An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms", Psychometrika, Vol. 45, No. 3, Sept. 1980.
- * Milligan, G.W. y Cooper, M.C.: "An Examination of Procedures for Determining the Number of Clusters in a Data Set", Psychometrika, Vol. 50, No. 2, Junio 1985.
- * Penas, F. y Streb, J. : "Criterios de Estratificación", mimeo. B.C.R.A., 1994.

* Sarle, W.S.: "Cubic Clustering Criterion", SAS Technical Report A-108, Cary, NC: SAS Institute Inc., 1983.

* SAS/STAT User's Guide, Release 6.03 Edition, 1988.

* Seber, G.A.F.: Multivariate Observations. John Wiley & Sons, 1984.

* Wong, M.A. and Lane, T.: "A Kth Nearest Neighbor Clustering Procedure", J.R.Stat.Soc.B, 45, 362-368, 1983.