

La edad y distribución de clases del fondo bibliográfico del Centro de Documentación e Información (CDI) del Ministerio de Economía de la Nación (junio 2001)

LOZANO, Marta
mlozan@mecon.gov.ar

OBJETO DE ESTUDIO

En este trabajo se miden las variables edad y clase temática del fondo bibliográfico del Centro de Documentación e Información del Ministerio de Economía (en adelante CDI). La variable edad es estudiada en términos de tendencias centrales y de variabilidad, mientras que a la variable clase (debido a su característica de ser un dato nominal) se le determinarán las siguientes mediciones: moda, frecuencias absolutas, acumuladas y relativas y por tanto la distribución porcentual de éstas últimas.

Como fuente de información se tomó la base de datos bibliográfica principal y una muestra de la colección a partir del estante. Creemos que la información obtenida permitirá evaluar (en función de las dos variables) en qué medida es representativa la base de datos con respecto a la población. Cuan cierto puede ser de que se trate de una muestra sesgada.

Por otra parte, la variable clase es estudiada también en los préstamos activos utilizando como fuente de información la base de datos del sistema de préstamo para luego poder confrontarla con los datos obtenidos de la muestra y de la base de datos bibliográfica, creemos que esto permitirá inferir algunas ideas con respecto al tema de la subutilización y de la sobreutilización de las clases.

BASE DE DATOS BIBLIOGRÁFICA

Procedimientos para extracción de datos

Para las variables (edad y clase) se tomaron de la base de datos bibliográfica la información de los siguientes campos:

- Fecha normalizada (v44) de edición (se utiliza la norma ISO de fecha): los primeros cuatro dígitos.
- Signatura topográfica (v3 en adelante ST) a partir de la octava posición (ya que se encuentra antecedida por el código ISO de país, un guión, el código de biblioteca asignado por el CAICYT y dos puntos: AR-AHA:) tres dígitos. La ST cuenta entre sus elementos con la clase de acuerdo a la CDU. Por otra parte, en este campo se indica si existe más de un ejemplar asentando su ST como un subcampo. Por ej.: AR-AHA:345.1 Arg. M1223d^bAR-AHA:BTT 21 4 2

En el momento en que se extrajeron los datos para el presente estudio la base de datos contaba con cerca de 70.000 registros. En esta base de datos no sólo se registran monografías sino que también analíticas de publicaciones periódicas y de monografía por esta razón se hizo una primera extracción por el nivel de registro bibliográfico (cabe acotar que la estructura de la base responde a las Hojas de Descripción Bibliográfica de la Comisión Económica para América Latina (CEPAL)). Es decir que se extrajo el nivel de registro Monográfico (sólo y con sus variantes: m, mc, mcs y ms) dando como resultado 43.454 registros. Del total de estos registros se extrajeron otros más, a saber:

√ Los correspondientes a libros que se encuentran en el subsuelo (la mayoría son anteriores a 1980) y que su ST carece de clase porque sólo se indica: Sub y un número secuencial.

√ Los correspondientes a documentos de proyectos gubernamentales financiados por organismos internacionales que se encuentran guardados en CAJAS, por tanto en la ST no aparece una clase sino que la denominación caja seguida por un número secuencial.

√ Los pertenecientes a Bibliotecas Técnicas (anexo del CDI que administra colecciones sobre transporte, obras públicas y recursos hídricos pertenecientes originalmente al Ex -Ministerio de Obras y Servicios Públicos) por dos razones: 1) estas colecciones se encuentran ubicadas en otro edificio y 2) la ST no responde a la CDU (ver más arriba el ejemplo de signatura topográfica con subcampo).

√ Los de documentos especiales, a saber: CD-roms, Diskettes, Archivos electrónicos que se encuentran ubicados en una unidad de red (LAN), Microfichas, Papers: artículos de publicaciones periódicas extranjeras solicitados a la British Library, Tesis y Videocintas. Todos ellos fueron descartados porque en su ST no aparece la clase.

√ Los de documentos de organismos tales como la CONADE, ExSecretaría de Planeamiento, CFI, SIGEP ya que en su ST no consta la clase.

√ Los correspondientes a otras bibliotecas (por ejemplo de la CEPAL. Oficina en Buenos Aires, del Instituto Torcuato Di Tella, etc.)

En resumen y a continuación se detallan los extraídos a partir de los 43.454 registros:

Libros en el Subsuelo	6461	14,86%
Proyectos en Cajas	4400	10,12%
De Bibliotecas Técnicas	2934	6,73%
De materiales especiales	1816	4,17%
De organismos varios	567	1,24%
De Otras Bibliotecas	175	0,39%
TOTAL	16353	37,5%

El resultado de esta depuración fueron 27.101 registros (62,37% sobre la primera selección)

Se optó por dejar el material de nivel MS es decir monografías de series (o también llamadas series monográficas, por sus características de publicación de aparición regular que lleva un mismo título y un número correlativo y secuencial y al mismo tiempo cada ejemplar contiene una única obra intelectual); además, la ST de este material está constituida por clases. También se decidió conservar los registros del material cuyo autor institucional es el Banco Mundial si bien su signatura no contiene (en su mayoría) clase lo hemos incluido por el volumen que representa en la colección (cabe acotar que el CDI es Biblioteca Depositaria desde 1984 de ese organismo internacional) y se lo considerará sólo como material del Banco Mundial (desde ahora BM). Además se dejó el material de referencia y tipo folio.

Las frecuencias tanto de las fechas como las de las clases fueron calculadas mediante el programa CDS/ISIS al invertir los campos, por ej.: para la fecha se tomaron los cuatro primeros dígitos del campo 44 (v44.4) y con la clase se tomaron las tres primeras posiciones después de la séptima para cada subcampo (v3^a*7.3, v3^b*7.3, v3^c*7.3) ya que las primeras siete posiciones están ocupadas por la codificación AR-AHA como se explicó más arriba.

Desde MicroIis y la pantalla a modo de ventana se fueron copiando las frecuencias (los posting) y sus valores correspondientes desde el diccionario y luego pegando a la planilla Excel.

Análisis de los datos

a. Edad de la colección

Como se dijo anteriormente la población resultante fue de 27.101 registros de los cuales se descartaron 204 (0,75%) por no poseer fecha de edición y 125 (0,46%) por pertenecer al 2001. Este último criterio de eliminación fue aplicado debido a que la muestra (ver más adelante) no contaba con ítems pertenecientes al 2001 y con el fin de aplicar tanto para la base de datos como para la muestra la misma base de cálculo para obtener las edades. Finalmente, el resultado fue 26.772 registros.

Las medidas de centralización y de variabilidad resultantes son (para observar la forma de los cálculos realizados, ver archivo [calculos.xls la primera hoja "Base Datos Bgraf.-Edad Colecc"](#)):

Media: **13 años 8 meses 8 horas 55,68 minutos** (1988)

Mediana: **10 años** (1991)

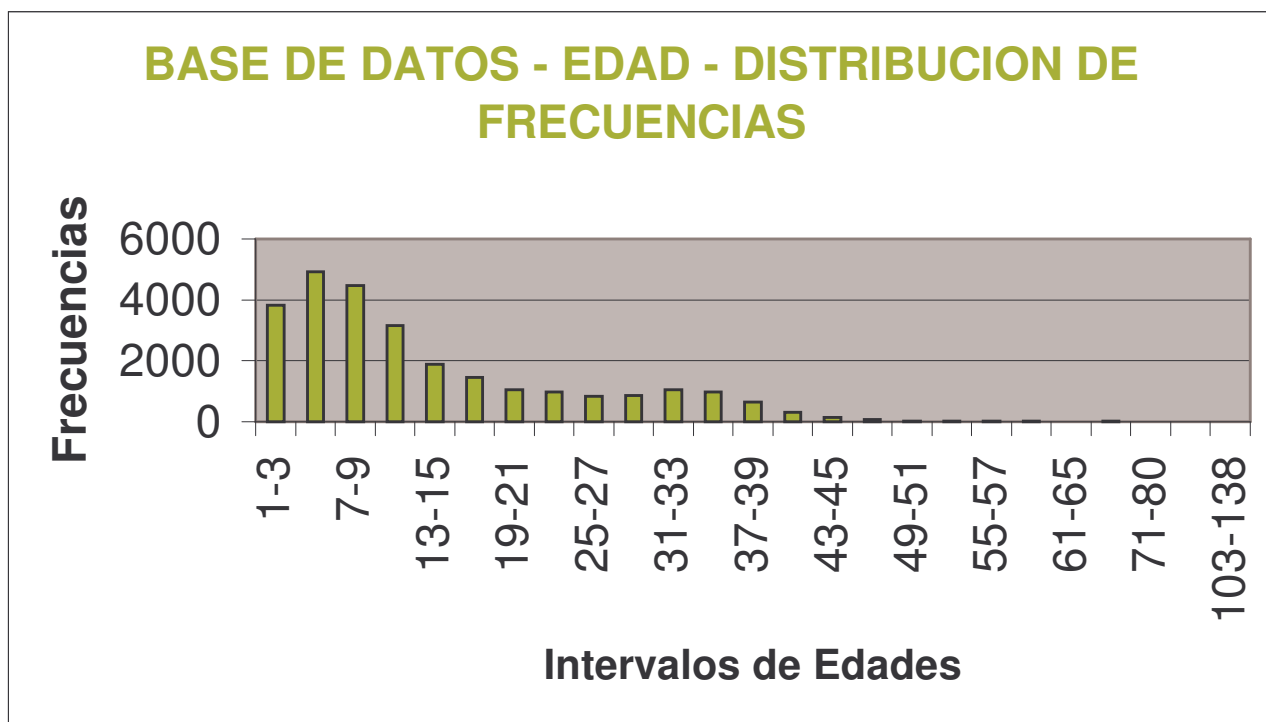
Moda: **5 años** (1996)

Puntuación más baja: **1 año** (2000)

Puntuación más alta: **138 años** (1863)

Rango:137
 Desviación Media: 9,09
 Desviación Estándar:11,40
 Varianza: 129,95

Como se puede observar en el siguiente gráfico de frecuencias absolutas los casos se concentran entre los 2 y 25 años de antigüedad (además, se puede ver en [calculos.xls en la segunda hoja "Base Datos Bgraf.-Edad-Gaus -Frec"](#) el mismo gráfico más extendido).



b. Distribución de clases

Primero se unificaron las clases y frecuencias de las monografías con las MS y los documentos de referencia, ya que las MS se encuentran precedidas en la ST por la letra H y los documentos de referencia por la letra R; es decir se unificaron en total 20886 ítems distribuidos entre: 16032 de monografías (58%), 4778 (17%) de MS y 77 (0,3%) de material de Referencia. Los 6581 ítems (24% sobre el total) publicados por el BM se mantuvieron aparte. Luego se determinó la moda ordenando 259 clases por sus frecuencias en forma ascendente, esto dio como resultado que la clase 338 (Producción. Movimiento de la Economía. Situación Económica) es la que cuenta con la mayoría de los ítems (4845). Que de entre las 36 clases que más frecuencias tienen (desde 100 a 4845) 23 pertenecen a la clase 3, es decir un 63,89% sobre el total (ver en [calculos.xls en la tercer hoja "Base Datos Bgraf.-Clase"](#))

MUESTRA

Elaboración de la muestra

A continuación se presenta la lista de ítems, con los datos de ST y fecha de edición (primero están los correspondientes a las monografías seriadas y luego los libros):

datos sin depurar	
ST	año
312 5	1994
323 2	1984
330 16a	1990
331 6a	1994
332 3b	2000
332 57g	1970
332 21a	1992
334 1	1998
338 28a	1987
338 72b	1994
338 73d	1987
350 2g	1996
350 6b	1999
36 13a	1996
380 32l	1996
381 10	1995
380 31d	1985
613 6	1996
62 3	1996
631 20 ^a	1986
631 36	1997
66 4b	1997
711 14b	1982
669 9	1975
bm 16	1988
bm 28 bo 1995	1995
bm 338.973 m 476	1987
325.25 M352h	1969
329 Arg. U46ed	1983
330.84 S916w	1997
336.2 China B853	1998
338.94 Am.Lat. G212	1995
338.984 M5 Anal	1975
382.1 V28	1964
382.4 Grupo de los 8 G758	1988
341.232.3 J41	1991
551.46 Arg. F5re	1984
631 Arg.Sta.Fe A754	1983
654 M866te	1984
658.562 E95a	1995

Análisis de los datos

a. Edad de la colección

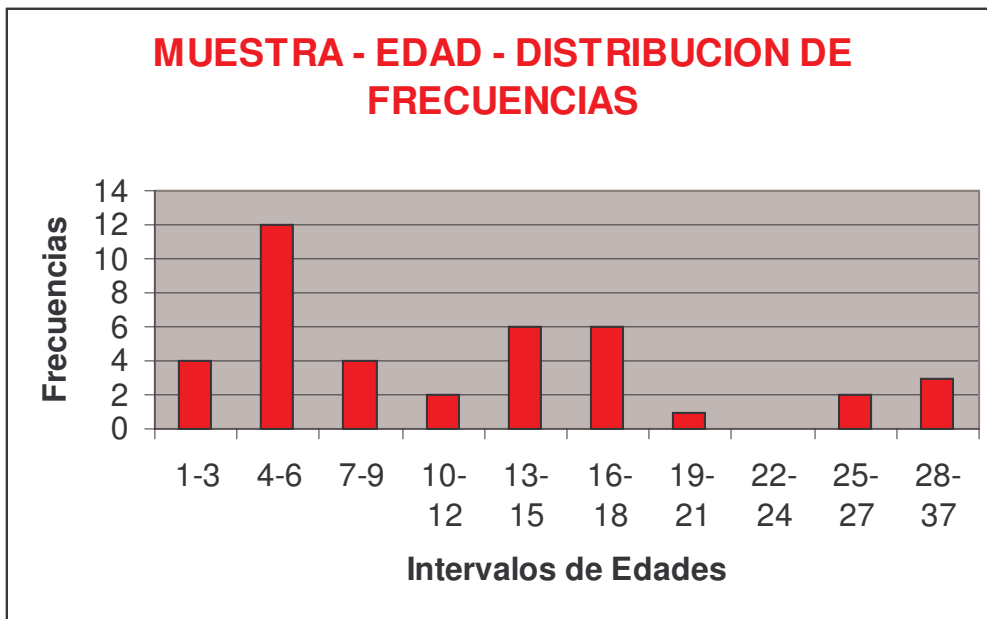
Las medidas de centralización y de variabilidad resultantes son (para observar la forma de los cálculos realizados, ver archivo [calculos.xls cuarta hoja "Muestra-Edad de la Colect"](#)):

Media: **12 años 2 meses 23 días y 1 hora** (1989)

Mediana: **10 años** (1991)
 Moda: **5 años** (1996)
 Puntuación más baja: **1 año** (2000)
 Puntuación más alta: **37 años** (1964)

Rango: **36**
 Desviación Media: **7,12**
 Desviación Estándar: **8,76**
 Varianza: **76,72**

Como se puede observar en el siguiente gráfico de frecuencias absolutas los casos se concentran entre los 3 y 21 años de antigüedad (también puede ser visto en [calculos.xls en la quinta hoja "Muestra-Edad-Gauss-Frec"](#)):

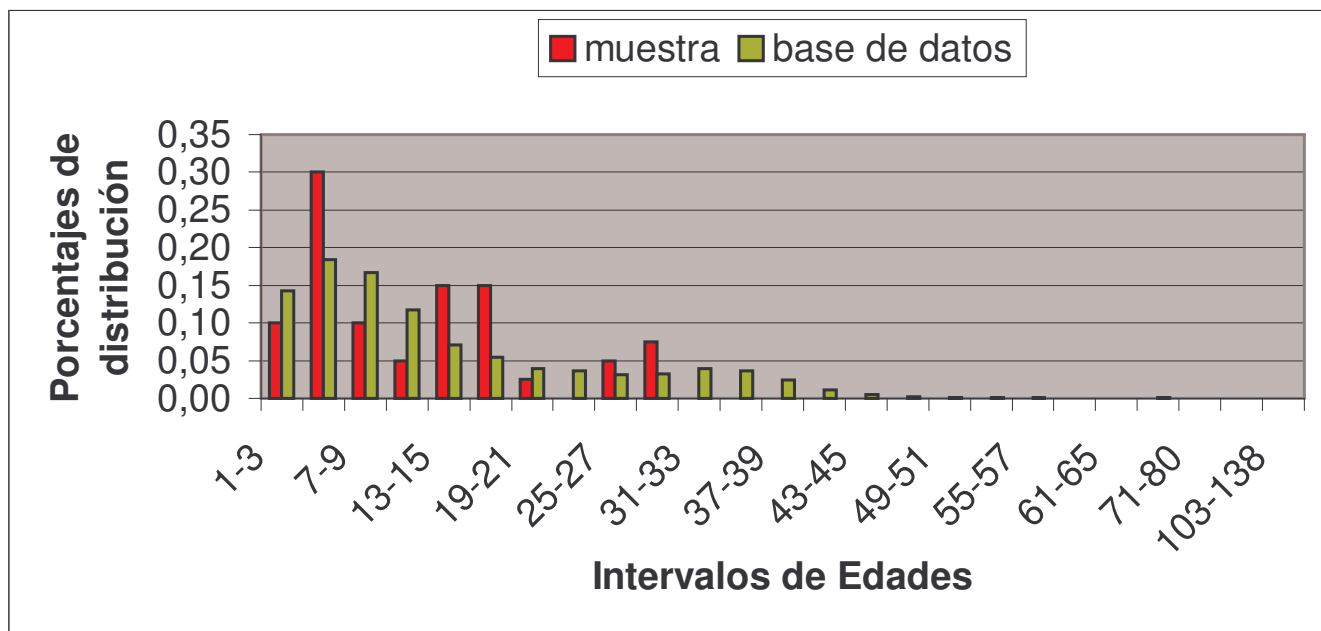


COMPARACIONES SOBRE LA VARIABLE EDAD

Llegado a este punto es necesario establecer algunas similitudes entre los resultados obtenidos de la base de datos y de la muestra. A continuación presentamos un cuadro a fin de observar en cuales datos hay semejanzas y en cuales diferencias:

MEDIDAS		BASE DE DATOS			MUESTRA				
Media		13			12			dif. 1	
Mediana		10			10			=	
Moda		5			5			=	
Puntuación más alta		138			37				
Puntuación más baja		1			1				
Rango		137			36				
Desviación Media		9,09			7,12			dif.1,97	
Varianza		129,95			76,72			dif.53,23	
Desviación Estandar		11,40			8,76			dif.2,64	
Media - + 1DS	Items	%	2,3-25,1	21049	79%	3,5-21	33	83%	dif. 4%
Media - + 2DS			-9,1-36,5	25479	95,17%	-5,3-29,8	37	92,50%	dif. 2,67%
Media - + 3DS			-20,5-47,9	26629	99,47%	-14-38,5	40	100%	dif. 0,53%

Por otra parte, se observa similitudes en la distribución de los porcentajes de frecuencias relativas como se ve en el gráfico siguiente (también puede ser visto más extendido en : [calculos.xls en la quinta hoja "Muestra-Edad-Gauss-Frec"](#)):



b. Distribución de clases

Se ordenaron las clases según la CDU llevando sus respectivas frecuencias. Cabe acotar que de los 40 ítems tomados al azar 24 se corresponden a Hemeroteca (es decir que se tratan de MS) esto es el 60%, 13 pertenecen al fondo bibliográfico: 32,50% y 3 son del BM es decir 7,50% sobre el total. Se determinaron las frecuencias absolutas, frecuencias acumuladas, frecuencias relativas y representación porcentual de éstas últimas. Además, se determinó la clase con más cantidad de ítems ordenando las frecuencias en forma descendente dando como resultado la clase 338 como la que más ítems posee.

PRESTAMOS

Procedimientos para la extracción de datos

Se tomó desde la base de datos en Access del Sistema de Préstamo la tabla de 1034 préstamos activos. Esta tabla fue exportada a Excel y depurada. Se quitaron los préstamos correspondientes a documentos ubicados en el subsuelo, Bibliotecas Técnicas, pertenecientes a otras instituciones o cuya ST fuera incompleta o incorrecta. De esta depuración resultaron un total de 718 préstamos, los datos tomados en definitiva fueron el autor y la ST.

Análisis de los datos

a. Distribución de clases

El tratamiento de los datos a fin de estudiar la distribución de las clases fue semejante al realizado con los de la Base de Datos y la muestra. Primero se ordenaron (de acuerdo a la CDU) las clase

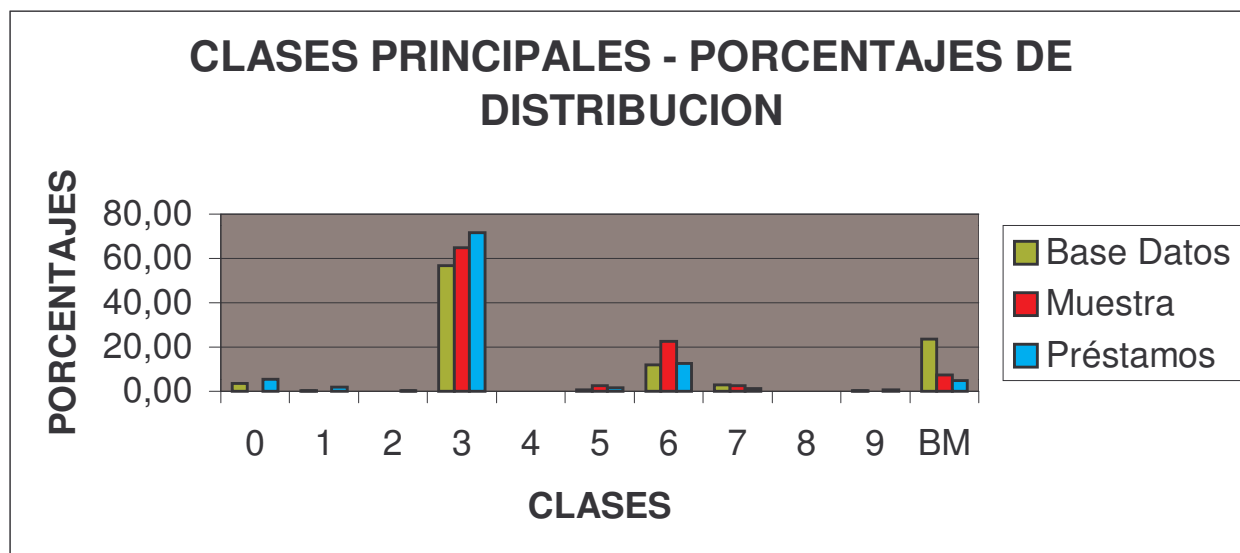
obtenidas de las ST y se fusionaron las pertenecientes al fondo bibliográfico (530 ítems 73,82%) con las de hemeroteca (137 ítems 19,08%) y las de referencia (17 ítems 2,37%) dejando apartadas las pertenecientes a la colección BM (34 ítems 4,74%). Una vez esto se contabilizaron las frecuencias de cada clase pudiendo así constar las frecuencias absolutas, también se determinaron las acumuladas y también las relativas. Se ordenaron las frecuencias absolutas en forma descendente pudiendo de esta forma establecer que la moda es la clase 338 con 105 ítems (esto se puede ver en [calculos.xls en la séptima hoja "Préstamo-Clase"](#)).

ANÁLISIS COMPARATIVO DE LA DISTRIBUCIÓN DE CLASES

Finalmente, analizamos las clases en forma general y la clase tres (dado su alto porcentaje de representación en las distintas fuentes) en forma desagregada.

Así, todas las clases desagregadas fueron resumidas en las diez (10) principales clases de la CDU. Esto se aplicó para las tres fuentes: base de datos, muestra y préstamos. Para todos ellos se estableció: las frecuencias absolutas, relativas y porcentajes de éstas últimas. Se elaboró la siguiente tabla y su correspondiente gráfica donde se muestra la distribución porcentual de cada clase y para cada fuente:

PORCENTAJES DE DISTRIBUCION				
CLASES PRINCIPALES		Base Datos	Muestra	Préstamos
Generalidades	0	3,47	0	5,43
Filosofía	1	0,20	0	1,81
Religión	2	0,05	0	0,28
Ciencias Sociales	3	56,47	65	71,59
Sin ocupar	4	0,01	0	0,00
Ciencias Puras	5	0,58	2,5	1,67
Ciencias Aplicadas	6	11,97	22,5	12,53
Arte	7	2,87	2,5	1,39
Linguística	8	0,01	0	0,00
Geograf. Historia	9	0,40	0	0,56
Banco Mundial	BM	23,96	7,5	4,74
Total		100	100	100



La gráfica muestra claramente cuál es el núcleo temático principal (3) y cuales son los otros que lo secundan (6, BM, 7 y 5). Esto se observa con mayor detalle en la siguiente tabla donde se rankean las clases por su porcentaje de representación y en forma decreciente:

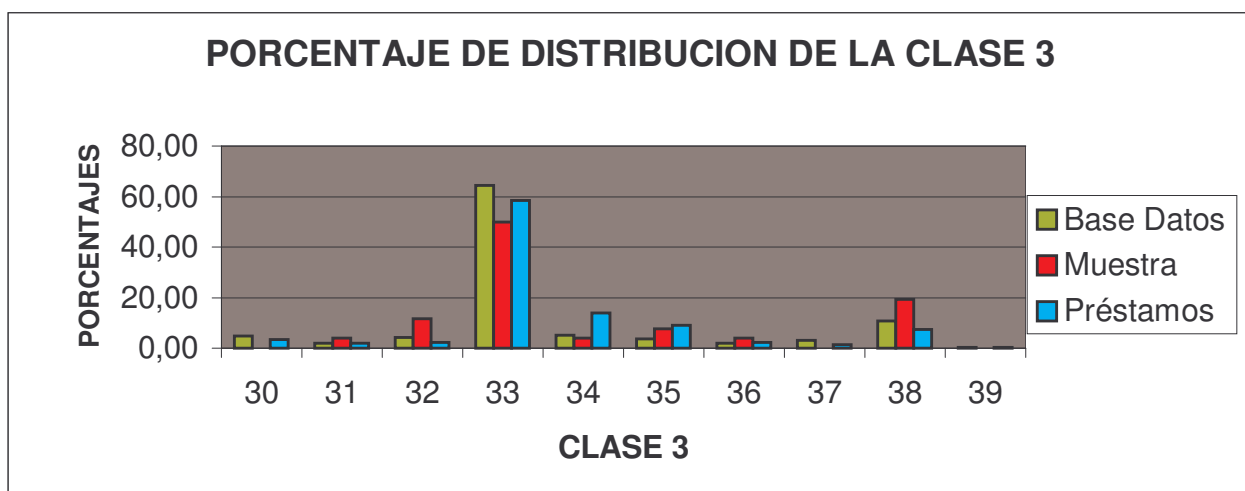
RANKING DE DISTRIBUCION DE CLASE					
	Base Datos		Muestra		Préstamos
3	56,77	3	65	3	71,59
bm	23,55	6	22,5	6	12,53
6	12,04	bm	7,5	0	5,43
0	3,49	5	2,5	bm	4,74
7	2,88	7	2,5	1	1,81
5	0,58	0	0	5	1,67
9	0,40	1	0	7	1,39
1	0,20	2	0	9	0,56
2	0,05	4	0	2	0,28
8	0,01	8	0	4	0,00
4	0,01	9	0	8	0,00

El tratamiento de este punto puede ser visto en calculos.xls en la octava hoja “Comparación-Clases”

La clase 3 fue estudiada en forma desagregada pero agrupándola hasta las 2 primeras posiciones, es decir que se tomaron en cuenta: el 30, 31, 32, 33, 34, 35, 36, 37, 38 y 39. De la misma forma que las clases principales la clase 3 fue analizada en términos de frecuencias absolutas, relativas y porcentajes de distribución.

En la siguiente tabla y gráfica se detallan los porcentajes de distribución de la clase 3.

PORCENTAJES DE DISTRIBUCION DE LA CLASE 3			
CLASE 3 AGRUPADA HASTA 2 PRIMEROS DIGITOS			
	Base Datos	Muestra	Préstamos
Sociología	304,93	0	3,31
Estadística	311,88	3,85	1,95
Política	324,35	11,54	2,14
Economía	3364,27	50,00	58,56
Derecho. Jursiprudencia	345,00	3,85	13,81
Administración Pública. Derecho Administrativo. Arte y Ciencia Militar.	353,64	7,69	8,95
Previsión y Asistencia Social Seguros. Uniones para fines Sociales.	362,01	3,85	2,14
Educación. Enseñanza. Formación. Tiempo Libre.	373,02	0,00	1,56
Comercio. Tráfico	3810,68	19,23	7,39
Etnología. Folklore. Usos y costumbres. Vida Social	390,23	0	0,19
	100	100	100



También, para poder establecer con cierta certeza cuales son las clases utilizadas, subutilizadas y sobreutilizadas elaboramos la siguiente tabla donde se rankea la clase 3 de acuerdo a los porcentajes de representación en las fuentes trabajadas.

RANKING DE DISTRIBUCION DE LA CLASE 3		
Base Datos	Muestra	Préstamos
33 64,27	33 50,00	33 58,56
38 10,68	38 19,23	34 13,81
34 5,00	32 11,54	35 8,95
30 4,93	35 7,69	38 7,39
32 4,35	31 3,85	30 3,31
35 3,64	34 3,85	32 2,14
37 3,02	36 3,85	36 2,14
36 2,01	30 0	31 1,95
31 1,88	37 0,00	37 1,56
39 0,23	39 0	39 0,19

Los detalles de cálculo de este tema pueden ser vistos en [calculos.xls en la octava hoja "Distribución Clase 3"](#)

INTERPRETACIÓN DESCRIPTIVA DE LOS RESULTADOS

Del análisis y estudio de los datos surgen algunas interpretaciones como las que siguen.

Con respecto a la edad de la colección se puede afirmar lo siguiente:

que las medidas de centralización y dispersión tomadas de la Base de Datos y de la muestra son muy aproximadas: con una diferencia en la media de un año, siendo la moda y la mediana iguales, que la diferencia en la Desviación Estandar es de 2,64 años,

que las áreas de concentración alrededor de la media tanto en una como en la otra son muy similares:

Base de Datos: entre 2 y 25 años (21049 casos un 79% sobre el total)

Muestra: entre 3 y 21 años (33 casos un 83% sobre el total).

También, el estudio de las clases bajo sus dos formas (las diez principales y la tres desglosada) arrojó similitudes (entre la base de datos y la muestra) que se observan en los extremos de los ranking de distribución de clases no así en el resto de la distribución, a saber.

el primer lugar en el porcentaje de la clase tres (con un 56,77% y 65% respectivamente),

el quinto lugar en el porcentaje de la clase siete (con un 2,88% y 2,5% respectivamente),

el décimo lugar en el porcentaje de la clase ocho (con un 0,01% y 0% respectivamente),

el primer lugar en el porcentaje de la clase treinta y tres (con 64,27% y 50% respectivamente),

el segundo lugar en el porcentaje de la clase treinta y ocho (con 10,68% y 19,23%

respectivamente),

el décimo lugar en el porcentaje de la clase treinta y nueve (con 0,23% y 0% respectivamente)

Lo dicho anteriormente nos lleva a plantear el siguiente razonamiento: siendo que fueron estudiadas dos variables en términos de medidas de centralización y dispersión en una y de distribución de frecuencias en la otra a partir de una muestra (tomada al azar), y que esas medidas y distribución de frecuencias constituyen datos estadísticos que representan al conjunto de la población y viendo que esos datos son semejantes a los obtenidos a partir del estudio de la base de datos se concluye que la base de datos es representativa del fondo bibliográfico-hemerográfico. No obstante ello, hacemos saber de la presencia de algunos sesgos que conviven desde sus orígenes con la base de datos, producto (en su mayoría) de criterios de automatización de las colecciones:

En lo atinente a las clases:

La desmedida presencia de la colección del Banco Mundial en la base de datos. Esto se debe a lo que dijimos al principio: al ser (el CDI) biblioteca depositaria de este organismo internacional y tal vez al recibir anualmente por parte de este organismo algunos "controles" (encuestas, visitas, etc.) se generó una especie de exigencia interna al CDI de tener todo lo que se pudiera de la colección en la Base de Datos.

De acuerdo a la muestra la clase 6 le sigue en cantidad de ítems a la clase 3 mientras que en la Base de Datos se encuentra en tercer lugar después del BM. Con la clase 5 sucede lo mismo

El 0 no se encuentra en la muestra como lo está notablemente en la base de datos. Esto tiene una explicación y es que existe desde hace años, dentro del CDI, una política de registrar todo lo atinente a ciencia y tecnología.

En lo atinente a la edad:

La automatización del Centro de Documentación comenzó a principios de 1988 y la política de carga de registros bibliográficos era primordialmente (y lo sigue siendo) todo lo que ingresaba nuevo, secundariamente lo de esa época (década del '80) y en último lugar por tema dejando de lado la fecha de edición. Aproximadamente, a partir de 1996 es cuando comienza a haber (en el contexto de políticas de automatización dentro del CDI) una tendencia más o menos firme con respecto a la reconversión retrospectiva de catálogos, intensificándose entre 1997 y 1999, esto se ve reflejado en el 14,86% sobre las 43.454 monografías registradas en la base de datos.

El dato de la moda (338: Producción. Movimiento de la economía. Situación económica) o, si se quiere, la clase que más frecuencias tuvo fue contundente por la coincidencia en las tres fuentes: 4845 ítems en la base de datos, 5 en la muestra y 105 en los préstamos.

Desde el punto de vista del préstamo se observa (con respecto a la base de datos) sobreutilizada la colección del Banco Mundial

En cambio, las más sobreutilizadas fueron:

Como resumen y a modo de conclusión afirmamos que: es indiscutible la especialización de esta colección en el área de las ciencias sociales, especialmente en las clases 33 y 338. Que en orden de importancia le siguen la colección del BM, el 6 y el 7. Que por el promedio de edad la colección es relativamente joven (12 años). Nada mal tratándose de las ciencias sociales, aunque quedaría pendiente para otro estudio confrontar este dato con índices de obsolescencia mas o menos estandares (si es que existen) en las diferentes áreas del conocimiento humano.

BIBLIOGRAFÍA CONSULTADA

Busha, Charles H. y Harter, Stephen P. **Métodos de investigación en bibliotecología: técnicas e interpretación.** México, UNAM, 1990. 408 p.

CEPAL **Sistema de información bibliográfica para el uso de Micro-CDS/ISIS.** Santiago de Chile, Cepal, 1991. (Manual de Procedimientos, 2) 80 p.

Hernández Sampieri y otros **Metodología de la investigación.** México, McGraw Hill, 1994. 505 p.

Madrid. Consejo Superior de Investigaciones Científicas. Instituto Nacional de Racionalización y Normalización **CDU: Clasificación decimal universal.** 3ra. ed. abrev. Madrid, IRANOR, 1975. 225 p.

Universidad Nacional de Mar del Plata. Facultad de Humanidades. Departamento de Bibliotecología **Seminario C: elementos de bibliometría.** Mar del Plata, 1999?