

PRODUCCIÓN CIENTÍFICA EN MÉTODOS ESTADÍSTICOS PARA EL ANÁLISIS DE DATOS CATEGÓRICOS, ENTRE 1997 Y 2002, UTILIZANDO LA BASE CIS

JORGE-EDUARDO ORTIZ(*) Y CAMPO-ELIAS PARDO(*)

jeortizp@unal.edu.co y cepardot@unal.edu.co

(*) Universidad Nacional de Colombia. Departamento de Estadística

Mots clefs :

Analyse de données catégoriques, scientometrics , production scientifique statistique, analyse multivariée de données

Keywords:

Categorical data analysis, scientometrics, statistical scientific production, multivariate data analysis

Palabras clave:

Análisis de datos categóricos, cienciometría, producción científica en estadística, análisis multivariado de datos.

Résumé

On fait une analyse de la production scientifique dans le domaine des méthodes statistiques pour des données catégoriques (DC) entre 1997 et 2002, en utilisant les articles indexés dans le “core” du Current Index to Statistics (CIS: http://www.statindex.org/CIS/index_htm). Les sujets de chaque article sont identifiés dans les mots clefs. On a extrait 1734 articles parmi les 33348 indexés. La méthodologie employée est la suivante: traitement initial, construction de tables de contingence (TC), description par l’analyse des correspondances y classification automatique. On identifie les auteurs, les revues et les sujets les plus importants, des relations auteur – revue, auteur – sujet et l’évolution pendant la période étudiée. La production d’articles sur DC correspond à moins de 19% des publications dans les revues où on publie le plus sur ce sujet et à peu près au 5.4% du total d’articles indexés.

Resumen

Se realiza un estudio de la producción científica en el área de los métodos estadísticos para el análisis de datos categóricos (DC), entre 1997 y 2002, utilizando los artículos indexados en el “core” de la base Current Index to Statistics (CIS: http://www.statindex.org/CIS/index_htm). Las temáticas de cada artículo se identifican a través de las palabras clave. Sobre el tema de estudio se extraen 1734 artículos de un total de 33348 indexados, durante el período de estudio. La metodología utilizada es: tratamiento previo, construcción de tablas de contingencia (TC), descripción mediante análisis de correspondencias y clasificación automática. Se identifican autores, revistas y temas más importantes y sus relaciones así como su evolución durante los años de estudio. La producción de artículos sobre DC cubre menos del 19% en las revistas donde más se publica sobre este tema y cerca del 5.4% del total de artículos indexados.

1 Introducción

El presente trabajo tiene el propósito de realizar un estudio sobre la producción científica en métodos estadísticos para el análisis de datos categóricos a partir del Current Index to Statistics (CIS: www.statindex.org/CIS/index_html). Éste es un “índice bibliográfico” de las publicaciones en estadística y campos relacionados, con referencias completas de 162 revistas científicas (“core”) y complementado con artículos seleccionados de otras fuentes relacionadas, como memorias, libros y otras revistas.

Se utilizan los artículos de las revistas del “core” durante el período comprendido entre 1997 y 2002. Se toma la información sobre año, revista, autor, título y palabras clave. No se incluyen los resúmenes por su ausencia en un gran número de artículos. Con la anterior información es posible identificar, bajo los parámetros de CIS, los autores que más publican, los temas que son objeto de más publicaciones, las revistas que publican más artículos, la evolución del número de artículos publicados por los autores de mayor producción sobre cada tema. Además, se puede obtener un mapa de conocimiento al cruzar los autores con los temas y estudiar la evolución de los mapas de conocimiento (autores-temas).

La metodología utilizada es ya clásica en este tipo de análisis y existen programas especialmente adaptados para el efecto, el Tetralogie (<http://atlas.irit.fr/TETRALOGIE/tetrajeu.htm>) es uno de ellos (Dousset, Sosson y Vassard - 2001) y se utiliza aquí para el tratamiento previo de la información y para la construcción de las tablas de contingencia (TC). Las TC se migran al SPAD (CISIA – CERESTA, 2000) donde se realizan análisis factorial de correspondencias (AFC) y clasificación automática (Lebart, Salem y Berry, 1997). Para la clasificación se utiliza un procedimiento mixto (CM) a partir de las coordenadas factoriales del AFC (Lebart, Morineau y Piron, 1995).

En la sección 2 se presenta el procedimiento seguido para la selección del corpus y la sección 3 describe el conjunto de revistas que publican artículos sobre DC. La sección 4 se ocupa de los autores y su relación con las revistas. La sección 5 se dedica a los temas, sus relaciones con autores y revistas, mientras que la sección 6 describe la evolución de los temas. Se presentan luego las conclusiones más importantes del estudio.

2 Selección del corpus

De los 33348 registros del “core” entre 1997 y 2002, 1793 contienen, en los campos de título o palabras clave, los segmentos que se muestran en la tabla 1, correspondientes a temas relacionados con datos categóricos (DC). De manera global, en las revistas que conforman el “core” de CIS, cerca del 5.4% de artículos indexados tratan sobre DC, objeto de este trabajo.

3 Las revistas

La tabla 2 presenta el número de artículos sobre DC por revista y por año, junto con el número total en el período, el número total de artículos y la proporción de artículos sobre DC con respecto al total por revista (CT). El porcentaje de artículos sobre DC en relación con la cantidad total publicada en cada revista es inferior a 19%, lo que significa que en las revistas donde se publican trabajos sobre DC, menos de la quinta parte corresponde a estos temas.

De acuerdo con la cantidad total (CT) y la de artículos sobre datos categóricos (DC), las revistas indexadas en CIS se separan en cuatro grupos: seis revistas con más de 150 artículos en promedio por año, tres de las cuales (Biometrics, Statistics in Medicine y Journal of the American Statistical Association) contienen entre 20 y 26 artículos sobre DC y las otras tres (Journal of Statistical Planning and Inference, Communications in Statistics: Theory and Methods y Statistics & Probability Letters) contienen entre 7.8 y 11.5 artículos en promedio por año sobre DC. Dos revistas más (Computational

Statistics & Data Analysis y Biometrika) publicaron 10.8 y 11.7 artículos promedio por año en el tema y las demás publicaron en promedio menos de 8 artículos sobre DC.

Tabla 1. Términos para la selección de artículos sobre datos categóricos

1	BINARY DATA	16	MULTINOMIAL MODEL
2	BINOMIAL DATA	17	NON*METRIC MULTIDIMENSIONAL SCALING
3	CATEGORICAL	18	OPTIMAL SCALING
4	COLLAPSING TABLE	19	POISSON DATA
5	CONTINGENCY TABLE	20	PREFERENCE ANALYSIS
6	CORRESPONDENCE ANALYSIS	21	PRINQUAL
7	DECISION TREES	22	PROBIT
8	DUAL SCALING	23	QUALITATIVE
9	FREQUENCY DATA	24	QUANTIFICATION METHOD
10	GENERALI* LINEAR MODEL	25	RANK DATA
11	HOMOGENEITY ANALYSIS	26	RANK STATISTIC
12	LOG*LINEAR MODEL	27	RANK TEST
13	LOGISTIC REGRESSION	28	STRATIFIED ANALYSIS
14	LOGIT	29	TEXT MINING
15	MULTINOMIAL DATA	30	TEXTUAL DATA

Se compararon las cantidades de artículos promedio por año en DC publicados entre 1997 y 1999 con las de 2000 a 2002. El promedio de 312 artículos por año antes de 2000 en las 45 revistas de mayor producción bajó a 286. Se ve en la tabla 2 que, exceptuando el año 2000, la publicación de artículos indexados en CIS sobre DC decreció en los últimos cinco años. Sin embargo, es posible que la disminución observada en los artículos indexados del año 2002 sea atribuible a información desactualizada en los archivos disponibles para el análisis.

Para las TC construidas con Tetralogie se utilizan 1734 registros, que quedan luego de eliminar los correspondientes a comentario de un artículo anterior, réplica a un comentario, corrección a un artículo y revisión de un libro.

4 Los autores

En CIS los autores no se encuentran normalizados. Se aplica entonces un tratamiento textual, para el cual Tetralogie está especialmente adaptado. Se utiliza una tabla de sinónimos con el apellido y el primer nombre del autor o la inicial, dependiendo de la información disponible. La tabla 3 muestra el número anual y total de artículos por autor, para los de más alta producción en el tema.

La figura 1 muestra los perfiles de las 10 primeras revistas según los 10 primeros autores. Se aprecia, por una parte, la presencia de publicaciones de estos autores en varias revistas, siendo lo más común su participación en cuatro revistas. Por otra parte, con la excepción de Biometrika y Communications in Statistics: Theory and Methods, se encuentra que las revistas más grandes tienen mayor diversidad de autores y además, mayor cantidad de artículos sobre DC por autor. La anterior relación parece indicar la inclinación de los investigadores a publicar más frecuentemente en las revistas más grandes.

Tabla 2. Número de artículos sobre análisis de datos categóricos por revista por año y total

No	CIS code	TÍTULO	1997	1998	1999	2000	2001	2002	Cat.	Total	Cat/ Tot
1	Biomtrcs	Biometrics	20	29	30	36	23	18	156	944	16.5
2	StatMed	Statistics in Medicine	27	25	20	31	18	14	135	1529	8.8
3	JASA	Journal of the American Statistical Association (JASA)	35	20	25	14	17	13	124	1210	10.2

4	Biomtrka	Biometrika	14	11	13	7	17	8	70	478	14.6
5	JStPIInf	Journal of Statistical Planning and Inference	11	12	7	12	13	14	69	1319	5.2
6	CommStA	Communications in Statistics: Theory and Methods	15	13	8	13	12	6	67	1047	6.4
7	CmpStDA	Computational Statistics & Data Analysis	12	11	9	8	12	13	65	545	11.9
8	BiomtrcJ	Biometrical Journal	12	11	6	9	4	6	48	439	10.9
9	StPrLet	Statistics & Probability Letters	4	9	10	6	9	9	47	1459	3.2
10	JEconmtx	Journal of Econometrics	8	8	5	8	6	4	39	550	7.1
11	JABES	Journal of Agricultural, Biological, and Environmental S.	5	11	8	4	4	3	35	196	17.9
12	ApplStat	Journal of the Royal Statistical Society, Series C: Applied	3	9	8	8	4	3	35	220	15.9
13	AmerStat	American Statistician, The	3	3	5	5	6	10	32	590	5.4
14	AnlsStat	Annals of Statistics, The	4	3	6	11	4	3	31	539	5.8
15	CandJSt	Canadian Journal of Statistics, The	6	2	6	9	4	3	30	316	9.5
16	CommStB	Communications in Statistics: Simulation and Comp.	9	6	2	10	2	0	29	388	7.5
17	CmpStQ	Computational Statistics	11	4	3	1	4	3	26	204	12.7
18	JCmpGrSt	Journal of Computational and Graphical Statistics	3	2	9	3	5	3	25	265	9.4
19	JRSS-B	Journal of the Royal Statistical Society, Series B: Stat.M.	5	5	5	4	3	3	25	355	7.0
20	Ststcian	Journal of the Royal Statistical Society, Series D: The St.	7	0	6	7	3	2	25	290	8.6
21	JAppStat	Journal of Applied Statistics	4	2	7	5	3	2	23	480	4.8
22	StSinica	Statistica Sinica	2	3	1	6	4	5	21	444	4.7
23	SocMethR	Sociological Methods and Research	3	3	4	5	3	3	21	112	18.8
24	JStCmpSm	Journal of Statistical Computation and Simulation	3	7	4	4	2	NA	20	274	7.3
25	Psymtrka	Psychometrika	2	1	4	6	2	4	19	176	10.8
26	JRSS-A	Journal of the Royal Statistical Society, Series A: Statist.	0	0	5	4	6	3	18	352	5.1
27	JMultiAn	Journal of Multivariate Analysis	1	5	3	2	0	7	18	401	4.5
28	AusNZJSt	Australian & New Zealand Journal of Statistics	NA	4	3	2	4	5	18	218	8.3
29	JBioPhSt	Journal of Biopharmaceutical Statistics	3	4	5	3	2	1	18	224	8.0
30	RStApp	Revue de Statistique Appliquée	4	3	1	6	3	NA	17	102	16.7
31	JBES	Journal of Business & Economic Statistics	2	3	4	5	2	1	17	283	6.0
32	JKoreaSt	Journal of the Korean Statistical Society	3	6	3	0	2	2	16	201	8.0
33	EnvrMtrc	Environmetrics	2	3	6	1	2	1	15	322	4.7
34	StatComp	Statistics and Computing	2	3	3	2	3	2	15	196	7.7
35	AnInStMa	Annals of the Institute of Statistical Mathematics	3	1	4	1	3	3	15	301	5.0
36	JNnparSt	Journal of Nonparametric Statistics	2	3	5	3	2	NA	15	171	8.8
37	Technmcs	Technometrics	0	9	1	0	3	1	14	563	2.5
38	BJMaStPs	British Journal of Mathematical and Statistical Psych.	1	2	5	1	1	2	12	142	8.5
39	KorApS	Korean Journal of Applied Statistics, The	3	2	1	0	6	NA	12	157	7.6
40	JOffStat	Journal of Official Statistics	0	6	1	0	3	1	11	198	5.6
41	Chance	Chance	2	2	1	3	2	1	11	308	3.6
42	SankhyaB	Sankhya, Series B	0	3	2	3	0	3	11	149	7.4
43	JEdBehSt	Journal of Educational and Behavioral Statistics	2	0	2	3	1	3	11	136	8.1
44	PrInStMa	Proceedings of the Institute of Statistical Mathematics	1	2	2	4	1	0	10	105	9.5
45	StMedRes	Statistical Methods in Medical Research	2	3	2	2	0	1	10	137	7.3
46	Biostat	Biostatistics (Oxford)	NA	NA	NA	4	6	NA	10	58	17.2
		Otras	39	54	37	77	46	29	282	10077	2.8
		Total	300	328	307	358	282	218	1793	29170	6.1

Tabla 3. Número de artículos relacionados con datos categóricos, por autor y año

o.	Abrv.Autor	4.1 Autor	1997	1998	1999	2000	2001	2002	Total
1	MOLENBE	Molenberghs, Geert	4	6	8	0	6	1	25
2	LIPSITZ	Lipsitz, Stuart	0	6	10	1	5	1	23
3	AGRESTI	Agresti, Alan	3	2	4	5	6	1	21
4	CARROLL	Carroll, Raymond	5	3	1	2	5	1	17
5	IBRAHIM	Ibrahim, Joseph	1	2	6	4	3	1	17
6	CHEN,_M	Chen, Ming-Hui	0	2	6	4	1	2	15

7	NELDER,	Nelder, John	1	2	3	5	2	1	14
8	BRUNNER	Brunner, Edgar	3	0	1	3	1	2	10
9	WANG,_C	Wang, C.Y.	3	1	0	1	3	1	9
10	TEN_HAV	Ten Have, Thomas R	1	2	2	1	2	1	9
11	PAN,_WE	Pan, Wei	0	0	2	0	4	3	9
12	LEE,_AN	Lee, Andy H.	2	2	2	0	1	1	8
13	AERTS,_	Aerts, Marc	4	2	1	0	1	0	8
14	ZHANG,_	Zhang, Biao	1	0	1	3	1	1	7
15	RYAN,_L	Ryan, Louise M.	1	1	2	1	2	0	7
16	WANG,_S	Wang, Suojin	2	1	1	1	1	1	7
17	GELFAND	Gelfand, Alan E.	2	1	1	1	0	2	7
18	GOETGHE	Goetghebeur, Els	1	0	2	1	3	0	7
19	GASTWIR	Gastwirth, Joseph	0	1	3	2	1	0	7
20	GIBBONS	Gibbons, Robert D.	1	2	1	2	0	0	6
21	FARRELL	Farrell, Patrick J.	3	0	0	2	1	0	6
22	KAUERMA	Kauermann, Göran	1	1	1	2	1	0	6
23	CORDEIR	Cordeiro, Gauss M.	1	2	1	0	2	0	6
24	SUTRADH	Sutradhar, Brajendra	1	0	2	1	2	0	6
25	BERGER,	Berger, Vance	1	2	0	1	1	1	6
26	CHIB,_S	Chib, Siddhartha	1	1	1	1	2	0	6
27	NEWTON,	Newton, Michael	1	2	1	1	0	1	6
28	SIMONOF	Simonoff, Jeffrey S.	1	3	0	0	1	1	6
29	LINDSEY	Lindsey, J. K.	1	1	1	2	1	0	6
30	BEH,_ER	Beh, Eric J.	0	2	2	0	2	0	6
31	COULL,_	Coull, Brent A.	0	2	1	1	1	1	6
32	TOMIZAW	Tomizawa, Sadao	0	3	2	0	1	0	6
33	REBOUSS	Reboussin, Beth A.	0	1	1	1	2	1	6
34	AITKIN,	Aitkin, Murray	0	1	2	1	0	2	6
35	KENWARD	Kenward, Michael	0	0	3	1	2	0	6
36	SHAO,_Q	Shao, Qi-Man	0	0	2	3	1	0	6
37	ALBERT,	Albert, Paul S.	0	0	2	1	1	2	6

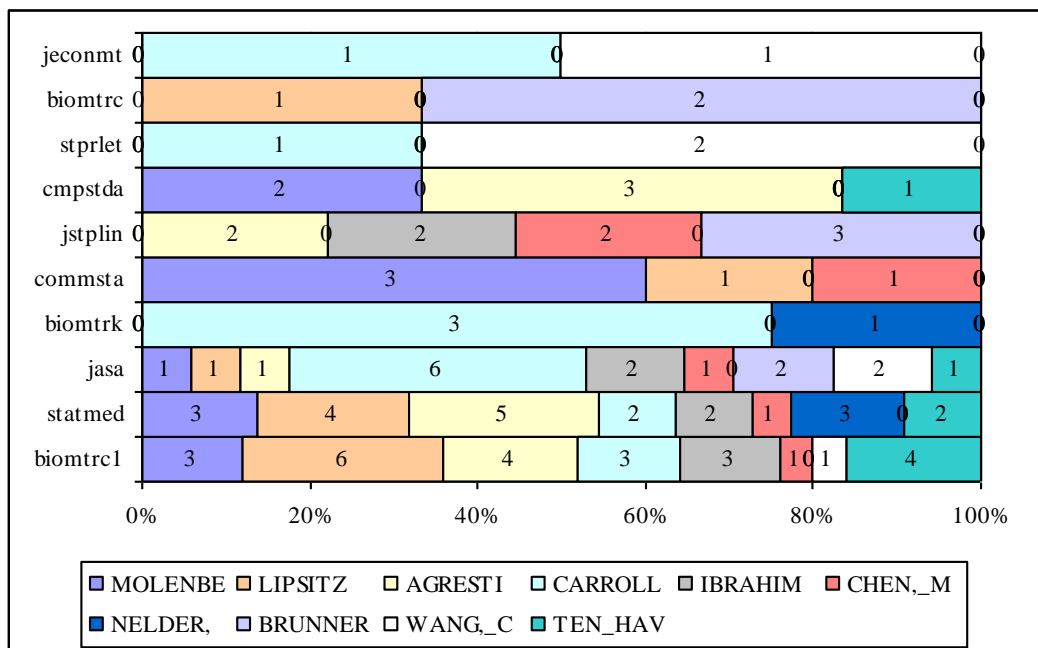


Figura 1. Perfíles revista-autor

5 Los temas

Luego de un tratamiento inicial con Tetralogie, se retienen 94 palabras clave con frecuencias superiores a 6 (véase tabla 4). Se observa una preferencia por los métodos inferenciales y el estudio de modelos, especialmente los modelos lineales generalizados y los métodos para el tratamiento de tablas de contingencia y datos binarios.

La tabla 5 muestra las principales asociaciones entre temas y autores, obtenidas mediante el procedimiento de clasificación mixta (Lebart, Morineau y Piron, 1995), sobre las 7 primeras coordenadas del AFC y con un corte del árbol en 9 clases, tal como se encuentra implementado en el SPAD.

Algunos grupos elaborados, como el que incluye los siguientes temas: exponential family, dose-response, overdispersion, corresponden a la temática de los modelos lineales generalizados. Su conformación en un grupo separado se puede interpretar como la actividad investigativa más específica por parte de los autores en las publicaciones indexadas.

Tabla 4. Palabras clave con frecuencia superior a 6

IDEN	PALABRA CLAVE	FR	DEN	PALABRA CLAVE	FR
logisti	logistic regression	264	asympto	asymptotic relative efficiency	14
general2	generalized linear models	225	logit	logit	14
rank_te	rank tests	156	bayesia	bayesian analysis	13
maximum	maximum likelihood	152	cross-v	cross-validation	13
statist	statistical tests	136	sensiti	sensitivity	13
conting	contingency tables	111	robustn	robustness	13
categor	categorical data	97	smoothi	smoothing	13
log-lin	log-linear models	93	time_se	time series	12
likelih	likelihood	92	censore	censored data	12
markov_	markov chain monte carlo methods	88	hierarc	hierarchical models	12
probit_	probit models	84	bias	bias	12
binary_	binary data	84	metropo	metropolis algorithm	11
regress	regression	74	latent_	latent variable	11
logit_m	logit models	59	missing	missing at random	11
bootstr	bootstrap	56	princip	principal component analysis	11
em_algo	em algorithm	55	margin	marginal model	11
gibbs_s	gibbs sampling	54	case-co	case-control study	10
random_	random effects	53	bayesia1	bayesian inference	10
theory_	theory and methods	53	censori	censoring	10
least_s	least squares	42	predict	prediction	9
missing1	missing data	42	specifi	specificity	9
general1	generalized estimating equations	41	proport	proportional hazard	9
longitu	longitudinal data	40	simulat	simulation	9
odds_ra	odds ratio	39	imputat	imputation	9
classif	classification	35	mixed_m	mixed model	9
clinica	clinical trials	33	power	power	9
surviva	survival analysis	31	efficie	efficiency	9
error	error	30	binomia	binomial distribution	8
goodnes	goodness-of-fit tests	30	correla	correlation	8
exponen	exponential family	29	akaike_	akaike information criterion	8
power_o	power of a test	28	marting	martingale	8
hypothe	hypothesis testing	27	conditi	conditional inference	8
model_s	model selection	26	cox_reg	cox regression	8
mixture	mixture models	26	jackkni	jackknife	8
corresp	correspondence analysis	26	sports	sports	8
general	generalized estimating equation	25	linear_	linear regression	8
repeate	repeated measures	24	consist	consistency	8
nonpara1	nonparametric regression	24	epidemi	epidemiology	7
overdis	overdispersion	22	non-lin	non-linear regression	7
confide	confidence interval	21	chi-bar	chi-bar squared distribution	7
mcm_met	mcm methods	20	multino	multinomial distribution	7
dose-re	dose-response	17	nonpara	nonparametric inference	7
applica	applications and case studies	17	link_fu	link function	7
asympto1	asymptotic theory	17	incompl	incomplete data	7
bayes_f	bayes factor	16	discrim	discriminant analysis	7
poisson	poisson regression	16	aids	aids	7
cluster	clustered data	14	high-fr	high-frequency data	7

Tabla 5. Principales asociaciones palabras clave-autores

<p>generalized linear models maximum likelihood marginal model em algorithm likelihood logistic regression missing at random longitudinal data odds ratio applications and case studies</p>	<p>Molenberghs, Geert Nelder, John Lipsitz, Stuart</p>	<p>exponential family dose-response overdispersion bias</p>	<p>Cordeiro, Gauss M. Aerts, Marc Gelfand, Alan Ryan, Louise</p>
<p>akaike information criterion cross-validation</p>	<p>Pan, Wei</p>	<p>random effects binary data generalized estimating equation consistency bootstrap survival analysis classification</p>	<p>Albert, Paul S. Farrell, Patrick Reboussin, Beth A. Ten Have, Thomas Zhang, Biao</p>
<p>regression error generalized estimating equations case-control study linear regression robustness nonparametric regression efficiency asymptotic theory</p>	<p>Carroll, Ramond Wang, C.Y. Wang, Suojin</p>	<p>gibbs sampling markov chain monte carlo methods probit models mcm methods logit models mixture models proportional hazard power of a test bayes factor</p>	<p>Chen, Ming-Hui Chib, Siddhatha Agresti, Alan Shao, Qi-Man</p>
<p>Log-linear models</p>	<p>Beh, Eric J. Lindsey, J. K.</p>	<p>repeated measures contingency tables clinical trials categorical data statistical tests theory and methods</p>	<p>Brunner, Edgar Berger, Vance</p>
<p>rank tests</p>	<p>Brunner, Edgar Berger, Vance Gastwirth, Joseph</p>		

Un análisis similar al anterior se realizó con los temas y las revistas. Las principales asociaciones se presentan en la tabla 6.

Tabla 6. Asociaciones más destacadas entre temas y revistas

correspondence analysis logit principal component analysis	Revue de Statistique Appliquée	logit models probit models high-frequency data	Journal of Econometrics Sociological Methods and Research Journal of Business & Economic Statistics
theory and methods applications and case studies consistency	Journal of the American Statistical Association (JASA)	sports	Chance Journal of the Royal Statistical Society, Series D

6 Temas por años

En la figura 2 se aprecia un efecto del nuevo milenio en los temas investigados. El primer eje los contrasta por su comportamiento diferencial con relación al nuevo milenio: en la parte derecha se encuentran los temas que fueron objeto de mayor publicación antes del fin de siglo y que con la llegada del año 2000 su publicación disminuyó. Para los temas ubicados en la parte izquierda, las publicaciones se incrementaron en los años 2000 a 2002.

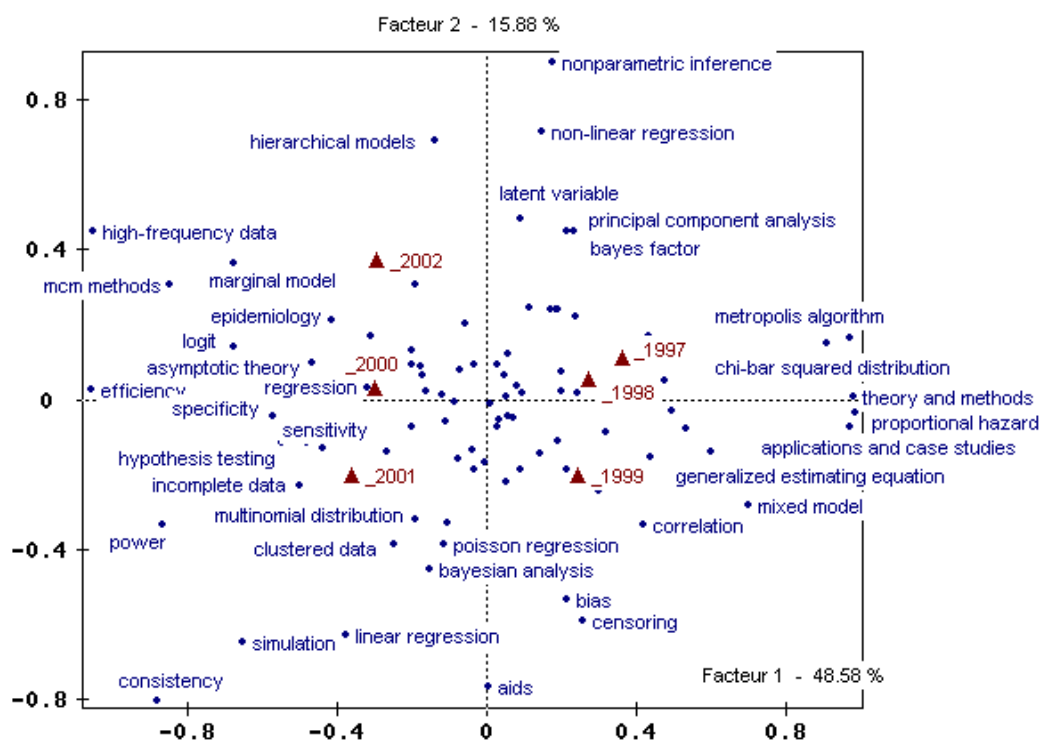


Figura 2. Plano principal del AC entre palabras clave y año de publicación

El segundo eje destaca otra característica diferente de los temas publicados: en la parte superior se encuentran las palabras clave de poca frecuencia en general pero cuyo uso disminuyó durante los años 1999 y 2001 aumentando de nuevo levemente en el último año; este es el caso de *nonparametric infe-*

rence, de *non-linear regression* y *hierarchical models*. En la parte inferior se observa el fenómeno contrario: las palabras clave como *aids*, *consistency*, *simulation* y *linear regression* tuvieron un ligero incremento durante los años intermedios.

7 Conclusiones

Los resultados más importantes del estudio están en los comentarios de cada una de las secciones, sin embargo, en términos más generales se presentan las siguientes conclusiones:

- La producción indexada en *CIS* sobre *DC* disminuyó en los años 2000-2002 en relación con la de 1997-1999
- La producción de artículos sobre *DC* cubre menos del 19% en las revistas donde más se publica sobre este tema
- En general, cerca del 5.4% de los artículos publicados e indexados en *CIS* corresponden a *DC*.
- Los autores más productivos en *DC* tuvieron entre 20 y 25 artículos publicados en revistas del *core* de *CIS* entre 1997 y 2002.
- Las revistas más grandes tienen mayor participación de los autores más productivos y estos se inclinan a publicar más frecuentemente en las revistas más grandes.
- Se observa una preferencia por el estudio de modelos, especialmente los lineales generalizados y de tablas de contingencia.

Nota de agradecimiento. Los autores agradecen el apoyo del “*Management Committee*” de *CIS* y del profesor Alan Zaslavsky, por autorizar el uso de la base y poner a disposición los registros correspondientes a las publicaciones entre 1997 y 2002, indexadas en la base de datos *CIS*.

8 Bibliografía

[1] CISIA – CERESTA *Spad Versión 4.5. Logiciel d'analyse des données*. Montreuil, Cedex, 2000.

[2] LEBART, L., MORINEAU, A., PIRON, M. *Statistique exploratoire multidimensionnelle*. Dunod, Paris, 1995.

[3] LEBART, L., SALEM A., BERRY L. *Exploring textual data*. Kluwer Academic Publishers, 1997.

[4] DOUSSET, B., SOSSON, D., VASSARD M. *Portail pour la navigation en ligne dans les analyses stratégiques*. En : *IRIT : Actes I. Veille stratégique scientifique et technologique*. 2001