

EDUARDO BERRIATUA SANSEBASTIAN
Dr. Ingeniero Industrial. Ldo. en Ciencias Económicas
Instituto Territorial de Vizcaya
Servicio Social de Higiene y Seguridad del Trabajo.

ANTONIO BAIGORRI MATAMALA
Licenciado en Ciencias Económicas
Instituto Territorial de Vizcaya
Servicio Social de Higiene y Seguridad del Trabajo.

JOSE ANTONIO LIBANO ZUMALACARREGUI
Dr. en Medicina
Instituto Territorial de Vizcaya
Servicio Social de Higiene y Seguridad del Trabajo.

Análisis estadístico del saturnismo.

El objeto de este trabajo es señalar una serie de procedimientos estadísticos de análisis de datos utilizables para el diagnóstico y esclarecimiento de enfermedades profesionales. Aunque estos procedimientos están concebidos para una aplicación general, hay que hacer constar que la dificultad en algunos casos o el coste en otros para la obtención de análisis clínicos y ambientales, hace que el número de datos tratados no sea muy elevado y, por tanto, las conclusiones que se obtengan deban aceptarse con cierta reserva; no obstante, las técnicas aplicadas se revelan como un instrumento de gran eficacia en la sistematización del diagnóstico de este tipo de enfermedades.

Este diagnóstico, por parte del médico, está basado en la mayoría de los casos en los resultados de unas exploraciones y de unos análisis que permiten el grado en que se padece una determinada enfermedad.

El caso concreto al que nos referimos es el de la enfermedad profesional de saturnismo.

Desde el punto de vista médico, el diagnóstico de esta enfermedad está basado, principalmente, en la

determinación, para cada paciente, de las cantidades de ALA, ALA-D⁽¹⁾, Plomo en Sangre y Linfocitos que, por unidad de medida, posee su organismo.

Los objetivos de este trabajo pueden dividirse básicamente en dos partes, lo que nos lleva a emplear distintos datos para cada una de ellas.

A: La primera pretende, a grandes rasgos, alcanzar los siguientes fines:

I.— Determinar si la clasificación médica en sanos y enfermos es coherente, es decir, si todos los individuos clasificados como sanos presentan unos resultados en los análisis de ALA, ALA-D, Pb/s y LINF, que pueden considerarse homogéneos y, lo mismo, para los clasificados como enfermos.

II.— Obtener, en orden jerárquico, cuál de los análisis de ALA, ALA-D, Pb, LINF, juegan un papel más relevante en el diagnóstico y cuáles pueden considerarse irrelevantes y, por tanto, suprimibles en lo sucesivo.

(1) ALA \equiv α δ aminolevulinico; ALA-D \equiv ALA deshidrasa

TABLA 1

ENFERMOS

ALA	Pb/s	ALA-D	LINF	ALA	Pb/s	ALA-D	LINF
1,25	71	98	1800	2,25	98	48	2923
1,04	68	93	2232	2,03	73	75	2820
1,53	70	66	3132	1,60	86	49	2790
2,59	87	64	3379	2,58	89	73	2257
1,31	74	76	3680	1,59	62	88	2952
1,20	61	108	2346	1,30	67	55	3731
2,22	78	112	2581	4,24	71	66	3192
1,36	73	68	2232	0,85	71	85	2923
1,18	66	52	3589	2,21	56	66	2820
2,52	86	99	2117	1,43	64	61	3128
1,02	71	50	2997	0,92	66	52	2912
2,15	81	100	2378	3,87	73	70	2730
3,50	81	54	2916	1	73	73	2923
2,16	85	69	3139	1,38	61	97	2412
1,54	67	99	1841	2,18	77	64	3204
1,06	68	92	4472	0,36	77	63	2904
1,22	81	61	5040	2,56	89	41	2548
1,13	94	52	2847	1,48	73	98	1575
1,35	75	56	2660	2,48	59	116	3362
3,83	95	33	3000	1,13	87	102	3902
1,63	84	76	2349	1,38	71	81	3069
1,38	86	68	2583	3,38	62	42	3102
1	79	108	2926	4,28	112	85	2465
1,11	67	81	5104	0,81	89	59	3010
1,77	71	47	2584	0,97	64	77	3232
2,13	77	50	3040	0,76	70	58	3237
2,33	79	68	3400	1,35	89	95	3298
2,59	70	82	2142	2,35	75	60	3102
1,07	71	72	4554	2	77	46	3162
1,87	63	63	3920	3,28	106	30	3030
1,50	63	46	3663	2,31	73	70	4142
2,69	64	41	3360	1,77	68	88	2673
1,91	83	50	2697	3,32	101	38	3360
2,10	87	60	2790	3,25	81	63	3024
4,93	107	57	2257	1,23	72	92	2331
1,56	85	61	2952	1,82	87	70	2448
2,99	91	55	3731	1,04	80	78	3145
1,65	81	71	3192				

SANOS

ALA	Pb/s	ALA-D	LINF	ALA	Pb/s	ALA-D	LINF
0,46	43	262	4256	0,50	40	153	2700
0,22	28	208	4140	0,64	28	282	3266
0,24	20	309	4080	0,19	30	247	2508
0,55	32	169	3264	0,18	22	127	3640
0,39	53	151	2862	0,14	26	250	4416
0,29	27	167	2695	0,33	22	317	2888
0,38	34	162	2905	0,33	35	235	2415
0,63	41	194	3520	0,25	36	225	2117
0,54	36	155	3168	0,19	19	244	2790
0,46	31	190	3040	0,27	32	164	3255
0,23	32	190	4223	0,65	42	207	2997
0,33	35	154	3185	0,65	38	168	2816
0,48	37	216	3030	0,36	21	149	4508
0,18	36	151	3811	0,26	40	115	3608
0,30	41	187	4386	0,41	61	140	2842
0,18	40	142	5848	0,24	47	211	3956
0,23	42	161	2295	0,59	39	152	3402
0,26	39	223	2470	0,33	29	198	2640
0,30	47	150	2449	0,32	44	186	3741
0,46	44	174	3895	0,42	26	254	3150
0,31	44	146	2508	0,33	32	316	3078
0,18	37	225	2440	0,32	27	553	4224
0,33	42	285	2560	0,67	40	162	3375
0,33	48	146	2684	0,38	46	226	3440
0,50	49	145	2652	0,31	35	283	3080
0,36	48	240	2964	0,31	45	151	5117
0,59	33	198	4750	0,50	38	188	3388
0,23	31	238	3525	0,32	31	162	4128
0,37	26	184	3081	0,33	38	225	3379
0,28	16	193	2975	0,34	32	143	3080
0,25	19	219	2262	0,30	33	155	3074
0,29	46	231	3016	0,21	33	175	2652
0,36	28	138	4726	0,45	52	187	3591
0,36	35	326	2808	0,23	28	217	2205
0,10	37	217	3120	0,43	35	182	4033
0,25	38	150	3564	0,31	32	311	3440
0,27	32	274	3115	0,34	26	225	2720
0,51	25	240	3060				

III.- Clasificación de futuros pacientes en sanos o enfermos mediante reglas de decisión basadas, únicamente, en aquellos análisis que hayan resultado relevantes en el apartado anterior.

B: La segunda parte del trabajo tiene por objeto determinar cuáles son las causas o factores que hacen que un paciente presente unos determinados resultados de ALA, ALA-D, Pb/s y LINF, y si pueden interpretarse mediante alguna variable conocida.

Fijados los objetivos, tanto para el apartado A como para el B, la no similitud de éstos nos ha obligado a emplear dos tipos de datos iniciales, empleando para el apartado A los datos de la tabla 1 y para el B los de la tabla 2, que expondremos a continuación.

Tomando en consideración el apartado A, los datos en que se fundamenta el análisis provienen de 150 individuos, de los cuales 75 fueron clasificados médicamente como sanos⁽¹⁾ y 75 como enfermos y a los que se les midieron las cantidades de ALA, ALA-D, Pb/s y LINF.

TABLA 2

Ind	Pb/aire*	Pb/sangre	ALA	ALA-D
1	0,12	80	2,23	63,13
2	0,08	98	0,59	180,59
3	0,44	71	0,83	102,87
4	0,00	67	0,34	124,05
5	2,5	84	0,35	110
6	0,67	80	1,61	76,52
7	0,63	63	0,78	79,53
8	0,13	83	0,82	70,93
9	0,95	87	2,28	100,55
10	1,52	108	2,43	58,35
11	0,5	53	0,31	23,93
12	0,8	80	1,5	25,56
13	0,55	67	0,20	187
14	0,26	80	2,04	66,52
15	1,43	65	0,85	84,24
16	0,08	77	0,75	107,20
17	0,54	62	0,34	172,36
18	0,26	60	0,67	111,5
19	0,24	39	0,67	156
20	0,22	43	0,57	52
21	2	21	0,30	64,33
22	0	42	0,38	49
23	0,10	24	0,58	74
24	0,15	44	0,45	74
25	0	30	0,14	39,95

*: Pb/aire = mgr/mm³

(1) Sanos son trabajadores sin alteraciones clinicoanalíticas, aunque su vida fabril se desarrolle en medios contaminados por el Pb. En personas que no tuvieran riesgo de saturnismo, no se plantea el problema de diagnóstico diferencial.

Esta clasificación se hizo mediante exploración y examen de los siguientes parámetros analíticos: ALA, ALA-D, Pb/s, Fe, Porfirinas, Hematocrito y Linfocitos.

Las unidades de medida utilizadas, han sido las siguientes:

ALA: mg/100 mililitros

ALA-D: unidades/hematie

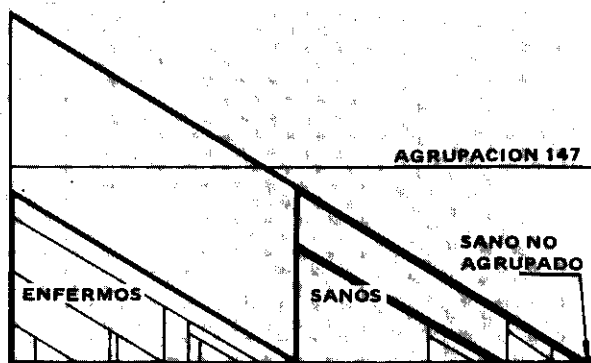
Pb/s: microgramo/100 mililitros

LINF: unidades/mm.

A.I. Para determinar si la clasificación médica expuesta en la tabla 1 es coherente, se han sometido los 150 datos considerados conjuntamente a un análisis de agrupación, de tal forma que se van formando grupos con aquellos individuos (puntos en su representación geométrica), que presentan una menor distancia entre sí, teniendo en cuenta que esta diferencia considera las correlaciones existentes entre las variables.

Este procedimiento empieza suponiendo que cada individuo es un grupo (en nuestro caso 150), y finaliza cuando se ha formado un único grupo, como vamos obteniendo todos los pasos intermedios, según el caso concreto que analicemos, pararemos en una determinada etapa el proceso de agrupación.

En el gráfico que se expone a continuación, pueden seguirse las sucesivas agrupaciones.



Si se detiene el proceso en la agrupación 147, observamos por una parte un bloque formado por los 75 individuos enfermos, un segundo grupo formado por 74 individuos sanos y un individuo aislado, el 110 que, médicamente, había sido clasificado como sano pero que no ha sido incluido en ninguno de los bloques anteriores.

Según estos resultados puede afirmarse que la clasificación médica es coherente. Salvo ese individuo anteriormente citado y que representa una proporción muy pequeña respecto al total. Existe pues una clara diferencia en los resultados de las medidas de ALA, ALA-D, Pb/s y LINF, entre los individuos sanos y enfermos.

A.II: Importancia de las variables o análisis para el diagnóstico.

Cuando a una persona se le diagnostica la

enfermedad de saturnismo, se hace en base a los resultados de ALA, ALA-D, Pb y LINF que de él se obtienen. Ahora bien, todos los análisis que intervienen en la clasificación no tienen la misma importancia. Este es el motivo que nos lleva a señalar en cuáles de ellos debe fijar el médico su atención para emitir su decisión final.

La técnica que hemos utilizado en este apartado se denomina análisis discriminante y consiste en obtener funciones de las variables consideradas para cada uno de los grupos en estudio en forma progresiva, (en nuestro caso, cuatro variables ALA, ALA-D, Pb, LINF, y dos grupos: sanos y enfermos). En un primer paso, las dos funciones clasificadoras, una para el grupo de "sanos" y otra para el de "enfermos", se hacen depender de una sola variable, la que produce menor error de clasificación (se entiende por error de clasificación, en este método, la proporción de individuos que han quedado mal clasificados). En pasos sucesivos, las funciones clasificadoras se completan con dos, tres... variables y se sigue observando el error de clasificación. El proceso se detiene en el momento en que, al añadir una nueva variable, no disminuye el error de clasificación.

En nuestro caso, la variable con mayor poder discriminador ha resultado ser la cantidad de Plomo en Sangre por unidad de medida o, lo que es lo mismo, la variable que hemos denominado Pb/s (plumbemia).

La segunda variable introducida ha resultado ser ALA-D y, en un tercer paso, la variable ALA.

En el caso analizado, ocurre que una vez se han tomado en cuenta las variables Pb, ALA-D y ALA, la introducción de la nueva variable LINF no disminuye el error de clasificación.

Este tipo de análisis nos ha permitido establecer los siguientes resultados, para los 150 individuos considerados:

1. Las variables que mejor permiten diagnosticar a un paciente la enfermedad de saturnismo son por este orden Pb, ALA-D y ALA.
2. La variable LINF es irrelevante en el análisis o, lo que es lo mismo que para este caso concreto, es indiferente que se tenga en cuenta o no en el diagnóstico.

El uso de este tipo de funciones de decisión resulta adecuado siempre que el error de clasificación final sea menor que un porcentaje prefijado, porcentaje que, como es lógico, depende del tipo de estudio y criterios de los investigadores del caso.

Considerando las tres variables Pb, ALA-D y ALA, que es la situación óptima a la que se puede llegar, resumiremos estos errores en el siguiente cuadro:

	ENFERMOS	SANOS	CLASIFICACION MEDICA
Enfermos	75	0	75
Sanos	1	74	75
Clasificación Estadística	76	74	150

Según se observa en la tabla considerada de los 75 individuos clasificados "a priori" como enfermos, también ahora son clasificados como enfermos, luego el porcentaje de error por clasificar un individuo enfermo al grupo de sanos es 0%.

Por lo que respecto al grupo de sanos, de los 75 existentes, 74 son clasificados como sanos y 1 como enfermo, con lo cual el error de clasificar un individuo enfermo en sano es del 1,3%.

Por último, considerados conjuntamente los 150 individuos, el % de error de clasificación es el 0,75%.

A.III: En este apartado utilizaremos los resultados obtenidos en el anterior para la clasificación de futuros pacientes en alguno de los grupos de sanos o enfermos.

Las funciones dependientes de las tres variables, anteriormente aludidas, son para cada grupo:

PARA EL GRUPO DE ENFERMOS

$$(1) Z_1 = 0,24X_1 + 0,07X_2 + 0,78X_3 - 33,72$$

PARA EL GRUPO DE SANOS

$$(2) Z_2 = -1,21X_1 + 0,11X_2 + 0,47X_3 - 20,30$$

Siendo: $X_1 = ALA$

$X_2 = ALA-D$

$X_3 = Pb/s$

Ante un nuevo individuo nos interesa conocer los valores que éste toma en los análisis: ALA, ALA-D, y Pb/s, que denominaremos X_1 , X_2 , X_3 . Sustituidos estos valores en las funciones (1) y (2), obtenemos los valores particulares Z_1 y Z_2 adoptando la siguiente regla de decisión:

Al individuo en cuestión lo asignaremos al grupo de enfermos si $Z_1 < Z_2$ y al grupo de sanos si $Z_2 < Z_1$.

No cabe duda de que, si en lugar de haber tomado esta muestra de 75 individuos sanos y 75 enfermos hubiésemos considerado el total de estos dos colectivos (en la práctica imposible), los parámetros que aparecen en las funciones (1) y (2) serían los verdaderos y no estimaciones de éstos, como ocurre aquí. Esta nueva causa de error es común a todos los modelos muestrales y, en consecuencia, las predicciones que se realicen mediante estos procedimientos deben considerarla junto con

el error de clasificación anteriormente expuesto.

La segunda parte de este trabajo trata de analizar tres variables Pb, ALA y ALA-D. Hemos visto son relevantes para el diagnóstico del saturnismo, pueden estar determinadas por un número menor de factores o causas internas.

De ocurrir esto, se intenta identificar estos factores con algún fenómeno conocido.

La técnica utilizada para el estudio de estas causas internas es el Análisis Factorial y, en este caso, el método de "factores principales".

Definimos la "comunalidad" como el tanto por ciento de la varianza de todas las variables que explican los factores comunes.

Hemos encontrado que con dos factores F_1 y F_2 pueden explicarse el 99,7% de la comunalidad de las tres variables Pb, ALA y ALA-D, resultando sus expresiones:

$$Pb = 0,854 F_1 + 0,243 F_2$$

$$ALA = 0,069 F_1 + 0,601 F_2$$

$$ALA-D = 0,813 F_1 - 0,307 F_2$$

El método permite, además, jerarquizar la importancia de cada factor en la explicación de las variables, resultando ser F_1 el factor más importante (explica el 73% de la comunalidad, frente a un 26,7 de explicación del factor F_2).

Esta reducción de tres variables a dos factores, se ha estudiado en 25 individuos (de los 150 iniciales) para los que, además de datos analíticos, se dispone de datos sobre las condiciones ambientales en que se desarrolla su trabajo.

En la tabla 2 se reflejan estos datos.

Como novedad aparece en la columna primera una nueva variable, Pb/aire, que corresponde a una medición hecha en el puesto de trabajo de cada individuo.

En la tabla 3 se muestra la puntuación de cada individuo en los factores F_1 y F_2 .

TABLA 3

Ind	Factor 1	Factor 2	Ind	Factor 1	Factor 2
1	1,413	-1,450	14	1,228	-0,683
2	0,535	1,089	15	-0,019	0,209
3	0,091	0,501	16	0,180	0,506
4	-0,368	0,837	17	-0,406	-0,338
5	-0,034	2,025	18	-0,239	-0,151
6	0,841	0,593	19	-0,705	-0,727
7	-0,117	0,229	20	-0,709	-0,278
8	0,389	0,290	21	-1,414	-0,458
9	1,542	-0,391	22	-0,843	-0,587
10	2,151	-0,368	23	-1,111	-0,996
11	-0,666	-0,023	24	-0,756	-0,460
12	0,778	0,416	25	-1,304	-0,505
13	-0,456	0,719			

La última fase del estudio trata de la identificación de los factores F_1 y F_2 identificación que, de ser realizada, supondría poder conocer las causas inmediatas determinantes del saturnismo.

De las dos principales vías de penetración del plomo consideradas en medicina preventiva laboral, la dérmica y la respiratoria, sólo se dispone de datos cuantificables para la segunda (datos de la tabla 2).

El coeficiente de correlación entre el Pb/aire y el Factor F_1 , resulta ser 0,36, lo que no permite asegurar esta identificación ni rechazarla.

La ambigüedad de este último resultado puede provenir de las condiciones del experimento, ya que los datos disponibles son pocos (25 solamente) y los puestos de trabajo (desguace de barcos), por sus características, hacen difícil que las observaciones hechas sobre concentraciones de plomo en aire sean realmente representativas de las condiciones de trabajo.



Las técnicas empleadas pueden consultarse en:

HARMAN (1967)
Modern Factor Analysis.
Chicago Press.

ESCUADERO, L. (1975)
Nuevos avances en las técnicas de Análisis Cluster.
UAM. - IBM.

ANDERSON, T.W. (1973)
Introduction to multivariate statistical Analysis.

MINISTERIO DE TRABAJO. I.N.P. (1966)
Disposiciones Generales sobre Prevención Médica de las Enfermedades Profesionales
Serie Legislativa.

TOLA, S. et AL. (1976)
Occupational lead exposure in Finland.
Scand. J. Work environ and health 2 (2). 115.

ZIELHUIS, R.L. (1974)
Biological Quality Guide for Inorganic Lead.
Int. Arch. Ar Beitsmed 32: 103.

ARROYO, M. MARCOS DOMINGUEZ, E. OTERO, J. MARTIN SERRANO, J. (1978)
La Plumbemia como parámetro indicador en la exposición al plomo. II. Estudio comparativo de diferentes poblaciones.
Rev. Medicina y Seguridad del Trabajo, Núm. 101: pág. 82.