

SPAD.N (INTEGRADO) NORMAS DE UTILIZACIÓN

por

Juanito Etxeberría

Dpto. de Pedagogía del Lenguaje y Métodos de Investigación
Universidad del País Vasco

RESUMEN

El artículo presenta el paquete estadístico para PCs del SPAD.N INTEGRADO. Es un avance de la versión 2 del SPAD.N, ya que permite mayor manejabilidad en el análisis de correspondencias múltiples. Lo interesante para el investigador es que el autor ofrece un ejemplo de uso del paquete sobre un fichero de datos en forma ASCII en el que se combinan variables cuantitativas con cualitativas. Se explicita la realización de los análisis a través de cuatro fases: de carga del fichero, parámetros de análisis, ejecución y gráficas. Útil y muy didáctico.

ABSTRACT

This contribution presents the statistical package of the Integrated SPAD.N, as an advance of the Version 2. It permits a better manafeability than the first resources in the multiple correspondences analysis. The most interesting thing for the researcher is the proposal of an example of how the package works. The data records in ASCII codes are introduced combining quantitative and qualitative variables. The author details the four steps in work: files charge, parametres for analysis, execution and graphics. Very useful and didactic tool.

INTRODUCCIÓN

Durante mucho tiempo el Análisis de Correspondencias ha tenido una larga lista de espera de posibles usuarios, debido básicamente al hecho de que el paquete estadístico que nos permitía realizar los cálculos pertinentes para su obtención, el

SPAD, tenía unas normas de escritura bastante complejas que a la larga desanimaban a muchos usuarios potenciales.

Sin embargo en la versión para PCs del año 1991, SPAD.N (versión 2) ha introducido como novedad el SPAD.N INTEGRADO.

Esta versión echa por tierra todos los reproches que se hacían a este programa en función de su poca «manejabilidad» ya que ha conseguido facilitar enormemente las normas/pasos necesarios para ejecutar los análisis.

En lo que sigue supondré que el lector/usuario del paquete está familiarizado con los fundamentos teóricos del análisis de correspondencias múltiples, que es el análisis más específico de los realizables mediante el paquete que nos ocupa, y en el que me centraré básicamente en este artículo, a pesar de que las normas en general son válidas para cualquiera de los módulos que conforman este paquete integrado.

Iremos analizando los pasos a dar siguiendo el ejemplo que paso a exponer.

EJEMPLO

Supongamos que tenemos el siguiente fichero de datos introducidos en formato ASCII en el que tenemos mezcladas tanto variables cuantitativas como cualitativas. Los dígitos que ocupan las cuatro primeras columnas los hemos empleado para identificar los elementos. A continuación tenemos 15 variables cualitativas y dos cuantitativas que pueden estar introducidas en un orden cualquiera.

FICHERO DE DATOS: FELIX.DAT

```
0001 4 6 5 2 2 1 6 2 6 2 5 2 3 2 1 37 21
0002 4 1 2 3 1 1 1 6 6 1 1 2 1 1 1 999999 11
0003 1 3 9 3 4 1 2 1 6 1 6 2 2 1 1 16 11
.....
0135 1 1 4 2 1 1 6 6 0 0 0 0 0 8 999999 28
0136 1 3 4 8 9 3 6 1 6 1 6 2 2 1 8 16 18
0137 1 1 1 2 1 1 1 3 3 3 3 3 2 2 8 17 28
0138 2 3 4 2 6 1 6 6 6 5 5 2 2 1 8 25 18
```

NOTAS AL FICHERO DE DATOS:

1. Existen diversas formas de introducir los datos. He optado por la de formato fijo, por ser de muy fácil uso, así como muy utilizada también en ficheros que luego van a ser analizados por otros paquetes, p.ej.: SPSS/PC(+). En cada columna (grupo de columnas) introducimos los valores correspondientes a una variable. Estas las separaremos mediante una columna en blanco. Introduciremos únicamente valores numéricos.

2. **DATOS PERDIDOS:** Para las variables cualitativas utilizaremos el 0 como indicador de dato perdido. Para las variables continuas el procedimiento que se sigue por defecto es asignar a los valores perdidos el número 999999. El SPAD.N (integré) automáticamente lo considera como dato perdido. Esto trae consigo el que caso de que no reservemos seis espacios a cada variable continua, los valores de cada variable ocupen columnas diferentes. Este hecho carece de importancia siempre que los valores de cada elemento se introduzcan en el mismo orden. (Ver manual SPAD.N integré pág. 23 -Manuel de reference: procedure ARDON, pág. 9).

3. **CODIFICACIÓN DE VARIABLES CUALITATIVAS:** Teniendo en cuenta que el valor 0 es empleado para denotar un dato perdido, la codificación de las distintas categorías de una variable cualitativa debe comenzar por el 1 y tomar valores correlativos hasta el último. Por ejemplo una variable que tenga tres categorías debe ser codificada por los valores 1, 2 y 3. Si por error codificáramos dicha variable con los valores 1, 2 y 7 este hecho sería considerado como una variable de 7 categorías en las que ningún sujeto toma los valores 3, 4, 5 y 6 que sin embargo deberán ser definidos en el diccionario de etiquetas.

FICHERO DICCIONARIO: FELIX.DIC

Una vez que tenemos el fichero de datos antes presentado, para poder hacer uso del SPAD.N, necesitaremos crear un fichero DICCIONARIO, que lo deberemos escribir también en código ASCII haciendo uso de uno cualquiera de los procesadores de texto, que admita esta forma de escritura/grabación.

Presento a continuación el fichero de etiquetas, que para una mejor visualización lo presento a cuatro columnas, pero que en su versión para el SPAD.N estará escrito en una única.

El orden de etiquetas, tanto de variables como de valores de las mismas, debe ser el mismo que el utilizado en la introducción de los datos.

4	AITALANA	9	DNORESPE	7	GNONBIZI	7	JLANESPEK
	BAJO		D1MA		G1NE		J1FA
	MEDIO		D2LA		G2HE		J2LA
	ALTO		D3SE		G3HI		J3ER
	OTROS		D4RO		G4ME		J4IN
6	ALANMAIL		D5AS		G5BE		J5CO
	AGOI		D6AS		G6EZ		J6EZ
	AERD		D7EZ		G7OT		J7OT
	AKUA		D8BE	7	HNONESPEK	3	IDIOCASA
	APEO		D9OT		H1FA		CAST
	AKIR	3	ESEMEALA		H2LA		EUSK
	ABES		E1BA		H3ER		DOS
9	BLANESPEK		E2EZ		H4IN	3	LEESEUSK
	BLAG		E3OT		H5CO		MUCH
	BOSP	7	FSEMESPEK		H6EZ		REGU
	BDIR		F1AL		H7OT		POCO
	BGOG		F2LA	7	ILANLEKU	2	SEXO
	BZAI		F3GI		I1NE		MUTI
	BJAB		F4AR		I2HE		NESKA
	BBID		F5ED		I3HI	8	IKASTETX
	BEZD		F6EZ		I4ME		SANT
	BBES		F7OT		I5BE		ETXA
9	CNORBIZI		F8AM		I6EZ		LEKU
	C1SE		F9MN		I7OT		IRUR
	C2SE		F10A				AOIZ
	C3SE		F11B				ELIZ
	C4AN						LEIZA
	C5PA						GOIZ
	C6LA						0 CONTINU1
	C7EZ						0 CONTINU2
	C8GU						
	C9BA						

Comenzaremos escribiendo en la 4ª columna el número de modalidades que tiene la primera variable introducida. Si ésta es una variable continua escribiremos un 0, valor que hace que el SPAD.N la considere como tal. Si como en el caso que nos ocupa, la primera variable tiene 4 categorías, escribimos a continuación, en la misma fila y separado por un espacio el nombre de la variable. En las cuatro filas siguientes escribiremos las etiquetas correspondientes a cada una de las cuatro categorías de esta primera variable. En nuestro caso bajo, medio, alto y otros. Así continuaremos con todas las variables. En el ejemplo aquí analizado podemos observar que el hecho de que el número de categorías sea superior a 9 no nos produce ningún tipo de distorsión en la redacción de las etiquetas. También pode-

mos observar que el hecho de que las etiquetas de los valores sobrepasen las 4 letras, no nos producirá error, sin embargo solo serán utilizadas las cuatro primeras.

En el caso de las variables continuas bastará indicar mediante el 0 el carácter continuo de la misma. En este caso es obvia la no pertinencia de la indicación de las categorías.

ANÁLISIS DE CORRESPONDENCIAS

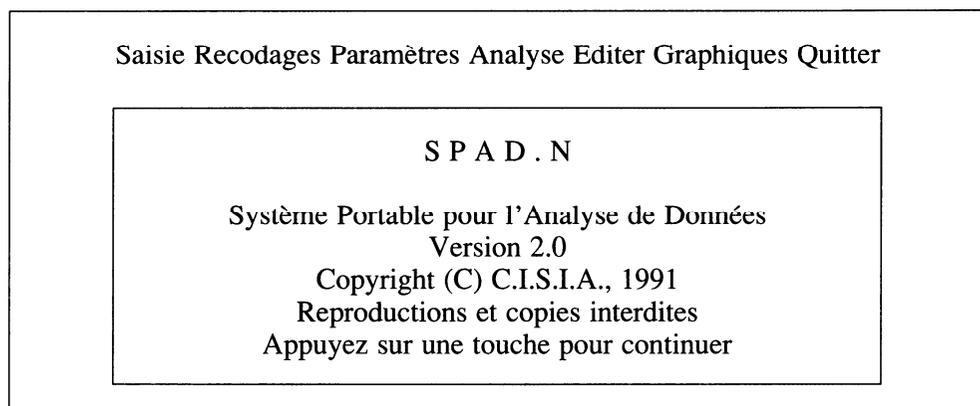
Una vez creados los dos ficheros antes mencionados (datos y diccionario) paso a exponer las etapas que son necesarias para realizar un análisis de correspondencias múltiples, así como obtener las gráficas correspondientes a los ejes factoriales.

El lector se dará cuenta que para poder realizar otros análisis (Componentes principales, correspondencias simples, Análisis discriminantes, regresión...) las etapas son muy similares, y en cualquier caso no revisten ninguna dificultad.

Las distintas etapas que debemos cumplir para llevar a cabo un análisis de correspondencias múltiples y obtener las gráficas correspondientes las podemos a su vez agrupar en cuatro fases:

- 1.- CARGA DEL FICHERO DE DATOS Y DE ETIQUETAS (**Saisie**)
- 2.- PARÁMETROS DEL ANÁLISIS DE CORRESPONDENCIAS (**Paramètres**)
- 3.- EJECUCIÓN (**Lancer**)
- 4.- GRÁFICAS (**Graphiques**)

Una vez hemos cargado el programa SPAD.N, nos aparecerá en pantalla la presentación del mismo:



Una vez pulsada cualquier tecla nos desaparecerá la pantalla de presentación, y sólo quedará en pantalla, la siguiente cabecera.

Saisie Recodages Paramètres Analyse Editer Graphiques Quitter

Para poder acceder a cualquiera de los menus teclearemos la inicial del menu a la vez que la tecla <Alt>, esto es para acceder al menu de graficos, teclearemos <Alt> + G, con lo que se nos desplegará el menú correspondiente.

Pasemos al análisis de cada una de las cuatro fases en que hemos dividido la generación/ejecución del programa.

1. CARGA DEL FICHERO DE DATOS Y DE ETIQUETAS (Saisie)

Tras la aparición de la pantalla de presentación, una vez pulsada cualquier tecla nos aparece la pantalla siguiente:

Saisie Recodages Paramètres Analyse Editer Graphiques Quitter

Comenzaremos por cargar el fichero diccionario que lo tenemos archivado bajo el nombre FELIX.DIC. Para ello, elegiremos el menú Saisie, al que accederemos con <Alt> + S + <Intro>. Nos aparece en pantalla el siguiente menú:

Fichiers Libellés Données

Siendo nuestro deseo cargar las etiquetas (libellés), teclearemos <Alt> + L, para elegir a continuación la opción <Importer>.

Ante la pregunta del nombre del diccionario, escribiremos el de nuestro fichero DIC.

Nom du fichier de libellés: FELIX.DIC

A continuación nos aparecerá en pantalla un fichero resumen del diccionario, similar al que sigue:

Fichiers	Libellés	Données	
Nº		Libellé	Type
1	AITALANA		4
2	ALANMAILA		6
3	BLANESPEK		9
4	CNORBIZI		9
5	DNORESPE		9
6	ESEMEALA		3

17	CONTINU2		CONT
FELIX.LAD			

En la columna de Type nos indica el número de categorías en el caso de que la variable sea cualitativa, y las letras CONT en el caso de que la variable sea continua, es decir en el caso que en el fichero diccionario: FELIX.DIC hayamos indicado 0 como número de categorías.

El número de la variable que figura en este fichero es el que utilizaremos luego a la hora de indicar el carácter activo o ilustrativo de cada una de ellas a la hora de calcular los planos factoriales o de proyectar las categorías y/o las variables cuantitativas en los mismos.

En la parte inferior de la pantalla nos aparece el nombre FELIX.LAD que es bajo el que guardará esta información el paquete que nos ocupa.

A continuación cargaremos el fichero de datos, para lo cual en primer lugar le indicaremos el formato elegido. Este punto se reduce a indicarle al paquete la LONGITUD EXACTA que tiene el IDENTIFICADOR en el caso de que existiera.

Para ello, teniendo en pantalla las etiquetas cargadas, teclearemos <Alt> + D, eligiendo la opción: **fOrmat d'importation**

Nos aparecen tres opciones:

1. Tenemos un identificador ocupando las primeras columnas sin marcas.
2. Tenemos un identificador ocupando las primeras columnas con marcas.
3. No tenemos identificadores

Siendo nuestro caso el primero, deberemos teclear el 1, como primer valor, para a continuación escribir un 4, dado que es esta la longitud exacta de los identificadores.

ES IMPORTANTE VALIDAR estos dos parámetros, para lo que teclearemos <F10>.

Cargaremos a continuación los datos, eligiendo las siguientes opciones, Données + Importer, para ello: <Alt> + D y a continuación elijo la opción **Importer**.

Nombre del fichero: **FELIX.DAT**

Una vez cargados los dos ficheros, veremos en la parte inferior de la pantalla: FELIX.LAD FELIX.DAT. Son los nombres de los ficheros con los que posteriormente trabajará el paquete.

ES NECESARIO SALVAR ambos ficheros. Para ello teclearemos <Alt> + F Y elegiremos la opción: **Sauver sous: FELIX**. Se acaba de generar un nuevo fichero: FELIX.DAD, que lo utilizará en sus algoritmos.

Para volver al menú principal, teclearemos <Alt> + F + Retour.

Volveremos a tener en pantalla el menú principal:

Saisie Recodages Paramètres Analyse Editer Graphiques Quitter

2. PARÁMETROS DEL ANÁLISIS DE CORRESPONDENCIAS

Una vez posicionados en el menú principal, elegimos la opción Paramètres y dentro de ésta el Themascopie, donde optaremos por la opción Multiples Correspondences: CORMU. La secuencia es la que sigue:

<Alt> + P; <Alt> + T; Multiples Correspondences CORMU + <Intro>

En este momento nos aparece una pantalla similar a la que sigue:

Fichiers Options Statistiques Thémascope Decisions

ANALYSE DES CORRESPONDENCES

Titre de l'analyse: Analyse des correspondences multiples

Nombre d'individus: 138

Selection des variables Exemple 3, 7, 10—15, 20

NOMInales ACT :

NOMInales II.I. :

CONTInues ILL :

Numéro de la variable poids des individus : UNIF

Nombre d'axes de coordonnées à calculer : 5

Petit graphique(1)ou grand graphique(2): 1 Valeur de Zoom: 2.3

Liste des axes à caractériser : 1—2

Nom (facultatif) du fichier texte des coordonnées: .GUS

OPTION DE CLASSIFICATION

Dentro de la OPTION DE CLASSIFICATION, nos permite realizar, bien una clasificación jerárquica, bien una mixta, que no la analizaré en este momento.

Me centraré en el Análisis de Correspondencias.

Supongamos que estoy interesado en trabajar con las variables AITALANA, ALANMAIL, DNORESPE, ESEMEALA, FSEMESPEK, GNONBIZI, HNONESPEK, ILANLEKU, JLANESPEK, LEESEUSK Y SEXO: Estas variables son las que consideraré como activas, esto es son las variables que utilizaré para calcular los ejes factoriales. Supongamos también que estoy interesado en proyectar sobre los planos factoriales obtenidos con las variables anteriores las variables cualitativas IDIOCASA e IKASTETX, y las dos cuantitativas, esto es y según el orden en el que están incluidas las variables en nuestro fichero de datos, orden que es el que ha memorizado el paquete cuando ha leído el fichero diccionario: FELIX.DIC, y que reproduzco a continuación, el análisis de correspondencias que pretendo realizar a cabo es con las siguientes variables:

NOMINALES ACTIVAS: 1, 2, 5—11, 13, 14

NOMINALES ILUSTRATIVAS: 12, 15

CONTINUAS ILUSTRATIVAS: 16, 17

Como podemos observar en la lista que presento a continuación ha habido variables que no han entrado a formar parte del análisis, en ninguna de las opciones posibles.

1.- AITALANA	7.- FSEMESPEK	13.- LEESEUSK
2.- ALANMAIL	8.- GNONBIZI	14.- SEXO
3.- BLANESPEK	9.- HNONESPEK	15.- IKASTETX
4.- CNORBIZI	10.- ILANLEKU	16.- CONTINU1
5.- DNORESPE	11.- JLANESPEK	17.- CONTINU2
6.- ESEMEALA	12.- IDIOCASA	

IMPORTANTE: A diferencia de otros paquetes de programas, nótese que para denotar a las variables 5 hasta la 11, HEMOS UTILIZADO dos guiones, esto es: 5—11.

En la mayor parte de los casos, en el resto de las opciones que se han de completar en la pantalla que nos ocupa, las que se que se plantean por defecto funcionan de forma satisfactoria en gran número de casos.

Una forma válida puede ser la siguiente: Peso de los individuos, el mismo para todos los casos; UNIF. Número de ejes de coordenadas a calcular, 5 ó 6 son números que muchas veces pueden resultar convenientes. Cualquiera de los dos tipos de gráficas (pequeñas y grandes) pueden resultar adecuadas. La lista de ejes a caracterizar puede ampliarse, en función sobre todo de las gráficas que nos interese obtener.

ES IMPORTANTE completar el nombre del fichero texto de coordenadas, que lo presenta como «**facultatif**», y que sin embargo es imprescindible en el caso de que queramos obtener las gráficas. En nuestro caso: **FELIX.GUS**.

Nombre de la variable poids des individus : **UNIF**
Nombre d'axes de coordonnées à calculer : **5**
Petit graphique(1) ou grand graphique(2): **1** Valeur de Zoom: **2.3**
Liste des axes à caractériser : **1-2**
Nom (facultatif) du fichier texte des coordonnées: **FELIX.GUS**

Por último al igual que en el caso de los ficheros diccionario y de datos hay que **VALIDAR** los parámetros anteriores tecleando **<F10>**.

A continuación, tecleando **<Alt> + F + Retour** volveremos al menú principal, para pasar a ejecutar el programa que en términos de SPAD.N equivale a LANCER.

3. EJECUCIÓN (Lancer)

Saisie Recodages Paramètres Analyse Editer Graphiques Quitter

Nos encontramos posicionados en el menú principal, y para lanzar el programa los pasos a dar son estos: Elegir la opción Analyse, y dentro de esta la Opción: Lancer. Para ello teclearé:

<Alt> + A + Lancer + <Intro>

El análisis se ejecutará. Obtendremos como fichero de resultados: **FELIX.LST**.

Para poder visualizar los resultados, podemos hacerlo mediante el comando Editer que aparece en el menú principal. La secuencia que deberá introducir es: **<Alt> + E + <Resultats>**. Esta instrucción hará que nos aparezca en pantalla la última página del fichero de resultados.

Para salir del editor: **<Alt> + X**

4. GRÁFICAS (Graphiques)

Esta versión integrada del SPAD.N también contiene un módulo gráfico que nos permite obtener las gráficas de los planos factoriales. El proceso que deberemos seguir para su obtención es el siguiente.

Una vez efectuado el Análisis de Correspondencias, recordemos que en los parámetros hemos indicado que se salvara un fichero texto de coordenadas bajo el nombre FELIX.GUS. Es el que nos permitirá obtener las gráficas.

Iniciaremos el proceso cargando el módulo gráfico tecleando <Alt> + G, con lo que obtendremos una pantalla equivalente a la siguiente:

Module graphique de SPAD.N

Plans factoriels
Regressions, Analyses de variance
Graphiques (x,y)
Choisir une imprimante
Quitter le programme

Graphiques de plans factoriels issus des procedures....

Una vez elegida la opción de <Plans factoriels> + <Intro>, nos presenta el menú principal de este módulo que es el que sigue:

Fichiers Axes Données Dessiner Options Config

En primer lugar elegimos la opción Fichiers y en este caso Charger, para cargar el fichero: FELIX.GUS; la secuencia es:

<Alt> + F + <Charger> + <Intro> + FELIX.GUS + <Intro>

A continuación elegimos la Opción de Datos <Données>, para elegir los elementos que nos interesa que sean representados. Las modalidades que se activen aparecerán con un asterisco. Para activar/desactivar la modalidad me iré posicionando sobre cada elemento, y aquellos sobre los que pulse la tecla <Intro> quedarán activados la primera ocasión (con un asterisco), o desactivados la segunda vez, (desaparecerá el asterisco).

Pulsando la tecla <Esc> volveré a la pantalla precedente.

Por último elegiré el menú <Dessiner>, con la secuencia:

<Alt> + D + <Intro>

El gráfico aparecerá en pantalla.

Para poder imprimir el gráfico la forma más sencilla es la del «Pantallazo», esto es pulsando la tecla **ImprPant** o su versión inglesa **PrtSc**.

Para salir del módulo gráfico, entramos en la opción de Fichiers, para una vez elegida la alternativa <Quitter>, activarla mediante la tecla <Intro>, con lo que volveremos al menú principal del que saldremos al sistema mediante la secuencia: <Alt> + Q.

BIBLIOGRAFÍA:

LEBART, L., MORINEAU, A., LAMBERT, T., PLEUVERT, P. (1991). *SPAD. N, Version 2.0. Système portable pou l'Analyse des Données*. Paris: CISIA.

LEBART, L., MORINEAU, A., LAMBERT, T., PLEUVERT, P. (1991). *SPAD. N, Intégré. Version PC*. Paris: CISIA.