

## ANEJO 3

### DESCRIPCIÓN DETALLADA DEL PROCEDIMIENTO DE ANÁLISIS DE SUSCEPTIBILIDAD A LOS DESLIZAMIENTOS SUPERFICIALES MEDIANTE EL ANÁLISIS DISCRIMINANTE

#### APLICACIÓN A LA POBLA DE LILLET

El procedimiento aquí descrito es una parte del procedimiento global para obtener un mapa de susceptibilidad a los deslizamientos superficiales. En este Anejo se describe paso a paso, una vez obtenidas las variables, el ajuste de éstas para obtener la muestra que se utilizará en el paquete estadístico SPSS para realizar el análisis discriminante y obtener el mapa de susceptibilidad. Con objeto de facilitar el seguimiento y la comprensión del procedimiento empleado, presentamos en cada apartado una descripción detallada de todos los pasos, ilustrándolo con ejemplos de **La Población de Lillet**.

Aunque no se hayan descrito explícitamente los pasos previos del procedimiento, creemos importante aquí hacer una breve referencia a ellos para situar al lector en el orden cronológico que conlleva un análisis de este tipo. Los pasos son:

1. Selección de la zona de estudio. Debe contener deslizamientos superficiales y disponer de la información necesaria para realizar el trabajo como por ejemplo, mapas topográficos a escala adecuada, fotos aéreas tomadas poco después de producirse los deslizamientos etc.
2. Selección de las variables a utilizar (teniendo en cuenta el tipo de análisis, el tipo de deslizamientos, los antecedentes bibliográficos sobre el tema, disponibilidad de variables digitales, etc.)
3. La obtención o generación (en el caso de que no exista o no sea de la resolución adecuada) de un Modelo Digital de Elevaciones de la zona. La selección de la zona de trabajo debe tener en cuenta la existencia y disponibilidad de un buen modelo con una resolución acorde con el tipo de deslizamientos y análisis de susceptibilidad.
4. Se procurará que las variables primarias no derivadas del MDE sean el menor número posible con objeto de reducir el coste de recogida de la información. Los pasos a seguir normalmente serán interpretación de fotos aéreas, visitas de campo y digitalización de la información. La teledetección puede ser una fuente de información, al igual que la existencia de una base cartográfica digital disponible en el mercado, lo que reducirá el tiempo de recogida de datos.
5. Generación de las variables derivadas del MDE a partir de archivos AML (se pueden consultar en el anejo 1).

Comentados los pasos previos del procedimiento se enumeran aquí los pasos que serán descritos en este apartado.

6. Rasterización de las variables vectoriales (ARC/INFO)

7. Ajuste /categorización de las variables (ARC/INFO)
8. Obtención de la muestra a analizar (ARC/INFO)
9. Test de normalidad de las variables (SPSS)
10. Transformación de las variables y comprobación con el test de normalidad (SPSS)
11. Dependencia entre variables / Análisis Factorial (SPSS)
12. Contraste entre las poblaciones de la muestra. Test T-TEST de comparación de medias y varianzas (SPSS)
13. Contraste entre poblaciones. Test ONEWAY de varianzas múltiples (SPSS)
14. Selección de las variables explicativas
15. Análisis discriminante (SPSS)
16. Obtención de la mejor función discriminante (SPSS)
17. Obtención de la malla regular Discriminante (ARC/INFO)
18. Obtención del Mapa de Susceptibilidad (ARC/INFO)
19. Obtención del Índice de Fiabilidad y validación(ARC/INFO y EXCEL)

### **RASTERIZACIÓN DE LAS VARIABLES VECTORIALES (ARC/INFO)**

La confección del mapa de susceptibilidad mediante el análisis discriminante requiere de una muestra del territorio. Uno de los métodos más utilizados para el muestreo es el uso de una malla regular de parcelación del terreno que facilita el análisis y la representación cartográfica de unidades geohomogéneas (Baeza, 1994). Cada unidad del terreno queda caracterizada por diversas variables que permiten diferenciarla de las unidades adyacentes. El uso de una malla regular simplifica la adquisición de datos, y como se ha observado en la definición previa de las variables, permite obtener variables cuantitativas y descriptivas del relieve derivadas de la variable de elevación (MDE).

El análisis que se presenta utiliza el formato raster. Las variables primarias obtenidas manualmente e introducidas en formato vectorial se deben rasterizar. Esto se realiza en ARC/INFO desde el subprograma ARC que controla la extensión de la zona a rasterizar (que será la del MDE). Los comandos usados son POLYGRID y POINTGRID, según se trate de variables con topología polígono o punto respectivamente (ESRI, Arc commands, 1994). Para rasterizar utilizamos un ítem o campo numérico de la tabla de atributos asociada a la variable vectorial, siendo el valor de este ítem el que tomará la malla regular resultante. En el caso que nos ocupa se ha transformado a formato raster las variables MOV, VEGET, FMS y GROSOR. Al transformarlas a malla regular se les debe dar un nombre distinto del de la variable en formato vectorial.

### **AJUSTE / PONDERACIÓN DE LAS VARIABLES (ARC/INFO)**

En este apartado se describe el ajuste o la ponderación de algunas variables para poder utilizarlas en el análisis. Se entiende aquí por **ajuste** a los pequeños cambios o ajustes realizados en algunas variables a fin de evitar que algunas celdas queden sin valor (NODATA) lo que anularía el registro de la celda, o la transformación de alguna variable de tipo circular como la ORIENT a una variable lineal (creación de variables compuestas). La **ponderación** hace referencia a la división en rangos de una variable

cuantitativa continua convirtiéndola en discreta, la reducción del número de rangos de una variable cualitativa o la ponderación de las variables primarias (cualitativas) a partir de unos criterios subjetivos y/o expertos sobre el peso de cada categoría en relación a la estabilidad.

### Ajuste de variables

El subprograma GRID de ARC/INFO permite visualizar el histograma de cualquier malla regular, utilidad que se usa para observar la distribución de las variables y transformarlas. Ello no implica obviar este paso en el SPSS, sino que sirve para ajustar algunas variables para que evite posibles errores o registros nulos una vez estemos trabajando ya en el SPSS. Las variables que precisan de este tipo de ajuste son las que no observan una distribución normal como, por ejemplo, ACUENCA, LONG, AFS, y LFS. El procedimiento de ajuste (realizado en el subprograma GRID) para estas variables es:

$$\text{ACUENCA1} = \text{ACUENCA} + 1,1$$

$$\text{LONG1} = \text{LONG} + 1,1$$

$$\text{AFS1} = \text{AFS} + 1,1$$

$$\text{LFS1} = \text{LFS} + 1,1$$

De estas 4 variables, AFS1 y LFS1 se han generado a partir de una máscara, donde las celdas no seleccionadas (que corresponden a ausencia de formación superficial) son celdas sin valor (NODATA) (ESRI, Cell-based Modelling with GRID, 1994). Esto significa que, según la zona de estudio, un mayor o menor número de celdas no se podrían utilizar en la mayoría de los análisis estadísticos por ser valores perdidos para el SPSS. Para evitar este problema estas dos variables han sido ajustadas de manera parecida a la expresión anterior, dando valor 1.1 a todas las celdas sin valor (NODATA). Esto se puede realizar con las expresiones:

$$\text{AFS2} = \text{CON} (\text{ISNULL} (\text{AFS1}), 1.1, \text{AFS1})$$

$$\text{LFS2} = \text{CON} (\text{ISNULL} (\text{LFS1}), 1.1, \text{LFS1})$$

Estas expresiones se pueden añadir al final del archivo fegrid1.aml. De esta manera ACUENCA1, LONG1, AFS2 y LFS2 son las variables que nos llevaremos al paquete estadístico SPSS. Allí las transformaremos a log10 o log normal, o la que se crea más conveniente. Si alguna de ellas llega a formar parte de la función discriminante final, hay que tener en cuenta que deberá ser generada como malla regular en ARC/INFO en un paso posterior.

La variable ORIENT, la orientación de la ladera, al tener valores de 0 a 360 se ha ajustado transformándola de 0 a 180 para evitar tener que trabajar con estadística circular. Esto anula la información referida al Este y al Oeste, puesto que las dos orientaciones toman el valor de 90 en la variable transformada, describiendo bien las orientaciones Norte (a las que se asigna el valor = 0) y Sur (celdas a las que se asigna el valor 180), debiendo este ajuste ser valorado con el conocimiento de la zona para decidir su aplicación en un área. La otra opción es convertir esta variable en cualitativa (categórica) si se cree conveniente. El primer caso es el que ha sido utilizado mediante el archivo AML *ori180.aml* (véase Anejo 1). La nueva variable ajustada será la que se utilice en el programa SPSS.

Otro retoque que puede realizarse ajusta la variable RUGOS reduciendo el efecto “borde” que sufre esta variable por la forma como se genera. Las expresiones pueden ser:

```
RUG = SELECT (RUGOS, 'VALUE > 0.7')
RUGOS1 = CON (ISNULL (RUG), 0.9, RUG)
```

El valor 0.9 es la media que puede ser cambiada según se crea conveniente. La variable RUGOS1 es la que se puede llevar al programa SPSS, aunque la experiencia demuestra que debido a los valores normalizados de esta variable, su distribución en un histograma y su poca significación en la estabilidad, se puede llevar al SPSS indistintamente la variable RUGOS o la ajustada.

### **Ponderación de variables**

La ponderación puede comportar el establecimiento de rangos de una variable tanto cuantitativa como cualitativa (por ejemplo, una variable primaria que contenga 20 categorías de información, para el análisis estadístico éstas pueden reducirse a 6 rangos atendiendo a su significancia estadística en cuanto a la susceptibilidad) o la ponderación de las categorías de las variables, tanto cualitativas como cuantitativas, para su inclusión en el análisis estadístico. En este caso la ponderación sólo se ha realizado en las variables primarias que estaban en formato vectorial, creándose un nuevo ítem en la tabla de atributos de las variables con la ponderación numérica deseada. Esto se ha realizado mediante ARC/INFO en el subprograma TABLES con funciones como SELECT, ADDITEM, ITEMS, RESELECT, CALCULATE, NSELECT, UPDATE, etc. (ESRI, Managing Tabular Data, 1994). Con este nuevo ítem se rasteriza la variable vectorial dentro de ARC dando la extensión del MDE con el comando POLYGRID.

Finalmente, la mayor parte de estos ajustes y ponderaciones realizados en las variables comentadas se pueden realizar indistintamente tanto en ARC/INFO como en el programa estadístico SPSS. Realizarlas en ARC/INFO supone una mayor velocidad de cálculo, la ventaja de tener ya la malla regular generada (en el caso de formar parte la variable en la función discriminante final) y la desventaja del espacio de disco ocupado por la malla regular que luego no tiene porque estar en la función final. La experiencia nos hace inclinarnos por la opción de realizar estas tareas en ARC/INFO.

### **OBTENCIÓN DE LA MUESTRA A ANALIZAR (ARC/INFO)**

Para realizar el análisis estadístico en el SPSS necesitamos extraer una muestra de las zonas de estudio, dado que no es necesario llevar la información de toda el área al programa SPSS (en caso de zonas grandes el programa no cogería todos los registros). También hay que tener en cuenta que el número de deslizamientos, y por tanto de celdas con roturas, es muy pequeño en relación con el número de celdas sin ellas o estables obtenidas a partir de la rasterización de la variable MOV. Por ejemplo, en La Poblada de Lillet existen 179225 celdas de las que sólo 270 son celdas donde se ha producido un deslizamiento y el área del valle del Deba de 70441 celdas sólo 150 incluyen deslizamientos.

El análisis discriminante trabaja bien cuando las dos poblaciones (celdas con deslizamientos y sin deslizamientos) tienen un tamaño similar en la muestra extraída

(Baeza, 1994) de manera que ninguna de las dos poblaciones tenga una mayor influencia en los resultados finales. Es obvio entonces que hay que extraer una muestra aleatoria de las celdas clasificadas como estables y que el número de esta muestra debe ser muy similar al de celdas con roturas a analizar.

El proceso se realiza corriendo el subprograma GRID. El objetivo es crear una muestra con número similar de celdas con y sin roturas utilizando todas las variables generadas, mediante la función SAMPLE. Las expresiones se pueden recoger en un archivo aml o en varios, aunque aquí se ha hecho de forma manual. El proceso es el siguiente:

1. Para crear la muestra aleatoria que incluya la mitad de celdas con roturas

Creamos máscara con las celdas con roturas

```
ine = select (mov, 'value = 1')
setmask ine
setcell mdel
setwindows mde
```

Ceamos malla regular aleatoria

```
randine = rand ()
```

De forma interactiva seleccionamos un número de celdas que sean aproximadamente la mitad de las que presentan roturas. Es un proceso iterativo de generación de mallas hasta dar con el número de celdas deseado.

```
Ran = select (randine, 'value < 0.005')
Int = int (ran)
List int.vat
```

Si al listar la VAT (tabla con los valores de la malla regular, para el caso de mallas con valores enteros), el número de celdas obtenido es el deseado nos quedaremos con la malla regular INT borrando la malla regular RAN. Si no es la cantidad de celdas que buscamos, por exceso o por defecto, borramos las mallas regulares generadas RAN y INT, creándolas de nuevo variando la cantidad 'value < .....' por otra más o menos cercana hasta dar con la cantidad aproximada deseada.

Una vez que se dispone de una malla regular INT con la muestra de celdas con roturas con un valor de 0 y el resto sin valor (NODATA), generamos una nueva malla regular dando valor 1 a estas celdas y al resto valor 0 y borramos las mallas regulares intermedias (RAN, INT, INT1). Previamente salimos del estatus máscara con el comando SETMASK OFF.

```
Setmask off
Ines = con (isnull (int + 1), 0, (int + 1))
Kill ran all
Kill int all
```

2. Para crear muestra aleatoria de celdas estables con un número similar al de las celdas sin roturas. El proceso es el mismo variando nombres de las mallas regulares y algún argumento.

Creamos máscara con las celdas estables

```
est = select (mov, 'value = 0')
setmask est
setcell mde
setwindows mde
```

Ceamos malla regular aleatoria

```
rand = rand ()
```

De forma interactiva seleccionamos un número de celdas que sean aproximadamente el mismo número de las celdas estables seleccionadas.

```
Ran = select (randine, 'value < 0.005')
Int = int (ran)
List int.vat
```

Ahora tenemos una malla regular INT con la muestra de celdas estables con un valor de 0 y el resto sin valor (NODATA). Generamos una nueva malla regular dando valor 2 a estas celdas y al resto valor 0 y borramos las mallas regulares intermedias.

```
Setmask off
Esta = con (isnull (int + 2), 0, (int + 2))
Kill ran all
Kill int all
```

3. Para crear la muestra final con las celdas estables e inestables.

```
Total = ines + esta
Sample = select (total, 'value > 0')
List sample.vat
Kill total all
```

La malla regular “sample” es la muestra con las celdas estables e inestables a analizar en el SPSS.

4. Creamos el archivo ASCII muestra con todas las variables a analizar en el SPSS.

```
MUESTRA = SAMPLE (SAMPLE, MDE, PEND, PENDS, ORIENT, ACUENCA1,
LONG1, AFS2, LFS2, PENDM, CURVAR, PERFIL, PLA, RUGOS1, INSOL,
GROSOR, VEG)
```

El archivo MUESTRA contiene todos los datos de las variables para las celdas estables e inestables seleccionadas previamente. En este punto se llevan los datos al programa estadístico SPSS.

## INTRODUCCIÓN DE LA MUESTRA EN EL PAQUETE ESTADÍSTICO SPSS

Una vez obtenida la muestra con todas las variables a analizar se introduce el archivo en el programa estadístico SPSS v.9. Esta versión permite entrar el archivo con todos sus datos de forma automática. El programa reconoce el orden de las variables e incluso el nombre en la parte superior. Al entrar en el programa vamos a *Archivo, Leer datos de texto* y damos nombre y localización del archivo que queremos abrir. Se abre la ventana del *Asistente para la importación de Texto* con 6 pasos sucesivos, a los que se llega con el botón inferior NEXT. Se trata de una serie de preguntas que el programa nos hace, aunque él mismo ya se las contesta, para que confirmemos el formato del archivo que él está detectando.

Se debe tener en cuenta que el programa SPSS, como otros programas, permite trabajar con un lenguaje propio, la **Sintaxis** donde se puede tener programado todos los pasos a realizar para que los ejecute de una sola vez. En los siguientes apartados se describen los pasos uno a uno, realizados de forma manual, sin usar una sintaxis.

## TEST DE NORMALIDAD DE LAS VARIABLES (SPSS)

Para verificar la hipótesis de normalidad de las variables se deben realizar dos pasos. **Primero** hay que realizar los **histogramas** de todas las variables donde observaremos las tendencias de sus distribuciones. Ello nos dará una pista sobre las variables que deberemos transformar para normalizar su distribución, en la medida de lo posible. Para obtener los histogramas vamos a *Gráficos, Histograma*, seleccionamos la variable y pedimos que nos dibuje la curva normal. Para cada histograma el gráfico resultante indicará el número de casos utilizados, la media y la desviación estándar. Las variables reconocidas por el programa o definidas como de cadena (cualitativas) no figuran en la lista de variables para realizar el histograma debido a su definición. En el caso de querer realizar su histograma hay que cambiar la definición de la variable de cadena a numérica. Esto último se hace con un clic con el botón derecho del ratón sobre la variable en el Editor de datos, entrando en *Tipo* y cambiando el tipo de variable.

Un procedimiento rápido para realizar todos los histogramas a la vez es en *Analizar, Estadísticos descriptivos, Frecuencias* en donde desactivamos *Mostrar tablas de frecuencias* y en el botón *Gráficos* activamos *Histograma con Curva normal* para todas las variables.

El **segundo** paso a realizar es el **Test de Kolmogorov-Smirnov** para chequear la bondad de ajuste de las variables a una distribución normal. Esto se realiza desde *Analizar, Pruebas no paramétricas, K-S de 1 muestra*. Una vez aquí se entra en la ventana de la prueba K-S donde se puede seleccionar todas las variables y se acepta la Distribución de contraste Normal defecto que presenta la ventana (que es la que nos interesa). El programa nos mostrará una tabla con todas las variables seleccionadas y con la media, la desviación estándar, la Z de Kolmogorov-Smirnov y la Significancia asintótica (bilateral) para cada variable. Las variables con valores altos de la K-S Z y nula significancia deberán ser transformadas para adecuarse a una distribución más normal.

## **TRANSFORMACIÓN DE LAS VARIABLES Y COMPROBACIÓN CON EL TEST DE NORMALIDAD (SPSS)**

Las variables con sesgo positivo y negativo marcados se transforman en este paso para reducir el sesgo y tender a una distribución normal. Una vez transformadas se realizan los histogramas de las variables transformadas y el Test K-S. Esto nos permitirá ver si se ha mejorado con la transformación de la variable. Si no se ha obtenido mejora será necesario decidir sobre la utilización de la variable original o la transformada.

### Transformación de variables con sesgo positivo

Las variables van a ser transformadas a un modelo de distribución lognormal de un componente. Ello se realiza asignando a la variable el valor de su logaritmo decimal. En el SPSS esto se realiza desde *Transformar, Calcular* y una vez en la ventana de Calcular hay que definir un nuevo nombre para la Variable de Destino y en la Expresión numérica utilizar la variable a transformar de la lista de variables, la función LG10 de la lista de funciones y los paréntesis que sean necesarios terminando con el botón Aceptar. Las variables transformadas se llamarán LACUENCA, LLONG, LAFS, LLFS (se añade una L de logaritmo para señalar la transformación).

### Transformación de variables con sesgo negativo

Para reducir el sesgo negativo se puede asignar a la variable el valor de su cuadrado o cubo. El proceso es el mismo que para el logaritmo, entrando en Transformar y en Calcular. Aquí transformamos la variable RUGOS1 y podemos realizar las dos transformaciones para decidir cuál es la mejor o si mantenemos la variable original. Las nuevas variables se han denominado RUGOS2 y RUGOS3 respectivamente.

En este punto se realizan los histogramas de las variables transformadas y la prueba K-S y se comparan con los resultados de las mismas variables no transformadas. En el caso de que no exista variación sustancial se utilizará la variable inicial.

A pesar de la hipótesis de normalidad para las variables, no se rechazará ninguna variable del análisis discriminante por no cumplir este requisito previo, debido a la robustez de la técnica discriminante (*Nie et al. 1981*).

## **DEPENDENCIA ENTRE VARIABLES / ANÁLISIS FACTORIAL (SPSS)**

El siguiente paso es comprobar la dependencia entre variables utilizándose para ello el Análisis Factorial de Componentes Principales que permite, por un lado, determinar las correlaciones y por otro, conocer la estructura de la muestra (*Baeza, 1994*).



Para poder trabajar con todas las variables (exceptuando la variable de agrupación de laderas con y sin roturas) es necesario que todas estén definidas como variables numéricas (dentro del SPSS).

Para realizar el Análisis Factorial nos dirigimos a *Analizar, Reducción de Datos, Análisis Factorial* entrando aquí en la ventana del mismo. Por un lado seleccionaremos las variables a analizar y daremos algunas especificaciones antes de Aceptar el análisis. En el botón de Descriptivos pediremos los Coeficientes de la Matriz de Correlaciones y en Rotación especificaremos la rotación Varimax, y que el programa nos muestre la Solución Rotada así como el Gráfico de Saturaciones.

### **CONTRASTE ENTRE POBLACIONES DE LA MUESTRA. TEST T-TEST DE COMPARACIÓN DE MEDIAS Y VARIANZAS (SPSS)**

En este paso como en el siguiente se realizan diversos tests de contraste entre las poblaciones estables e inestables de la variable de agrupación para cada variable de la muestra. Estos tests permiten conocer el poder discriminante de cada variable respecto a la variable de agrupación y seleccionar las que deben entrar en el análisis.

Para realizar el test hay que ir a *Analizar, Comparar medias y Prueba T para muestras independientes*. En la ventana de la Prueba T definiremos la variable de agrupación y sus rangos (los dos valores 1 y 2). Seleccionaremos las variables y aceptaremos. Este análisis muestra dos tablas resultados la primera de las cuales muestra los estadísticos de grupo para cada variable según los casos inestables (1) y estables (2), mostrándose la Media y la Desviación Típica de cada una. Serán estos datos los que nos ayudarán a seleccionar las variables suficientemente discriminantes para las poblaciones estables/inestables. Las variables con medias entre las dos poblaciones y desviaciones estándar entre poblaciones estables/inestables muy similares deberán ser rechazadas para el análisis discriminante. Esto es debido a que la poca diferencia entre grupos implica poca diferencia significativa en la población original de la variable y bajo o nulo poder discriminante de la variable respecto a la estabilidad / inestabilidad del terreno en cada celda. El programa también muestra una tabla con distintos estadísticos (Prueba Levene y Prueba T).

### **CONTRASTE ENTRE POBLACIONES DE LA MUESTRA. TEST ONE-WAY DE VARIANZAS MÚLTIPLES (SPSS)**

En este paso se realiza el Test ONE-WAY para varianzas múltiples, para complementar con el Test-T la selección de las variables que van a ser utilizadas en el análisis discriminante.

Para realizar el Test hay que ir a *Analizar, Comparación de medias y ANOVA de un factor* donde se presenta la ventana del Test. En ella únicamente debemos rellenar el Factor con la variable de agrupación y la ventana de Dependientes con las variables que seleccionemos de la lista de variables originales. Utilizaremos las opciones defecto de los tres botones inferiores de la ventana del test ANOVA. El test ofrece como resultado una tabla con los resultados dados por el valor F y por la Probabilidad F (o

significación). Estos dos valores para cada variable serán los que utilizaremos para seleccionar las variables finales que entrarán en el análisis discriminante. Cuánto mayor es el valor de F mejor discrimina una variable.

### **SELECCIÓN DE LAS VARIABLES EXPLICATIVAS (SPSS)**

Las variables que finalmente entrarán en el análisis discriminante se seleccionan a partir de los resultados del Análisis Factorial, el Test-T y el Test ONE-WAY. Como se ha comentado previamente el hecho que una variable no cumpla la hipótesis de normalidad no implica la exclusión de la misma para realizar el análisis discriminante. Ello permite entrar en el análisis las variables como RUGOS con un sesgo muy negativo y variables bimodales como ACUENCA o PENDM, entre otras.

A partir de la matriz de correlaciones obtenemos aquellos grupos de variables que están correlacionados (correlación superior al 0.9 entre dos variables). Sean dos, tres, cuatro o más variables las que estén correlacionadas en un grupo sólo una se escogerá para el análisis discriminante. Para decidir qué variable es la mejor para entrar en el análisis discriminante utilizaremos los resultados del Test ONE-WAY. De las variables correlacionadas, en el test ONE-WAY seleccionaremos la que tenga un F de Fisher mayor con una significancia menor. Con ello reduciremos redundancia de información en el discriminante, a la vez que nos aseguramos que la variable entrada tiene más poder discriminante que sus variables correlacionadas. Hay que tener en cuenta que en el caso de variables correlacionadas con un valor discriminante alto no se rechazará ninguna de las variables, y se realizará el análisis por separado con cada una de ellas para comparar su comportamiento con las otras variables, y valorar qué variable da mejores resultados en el análisis discriminante.

Para aquellas variables no correlacionadas con ninguna otra los resultados de los tests T y ONE-WAY serán útiles para valorar su introducción en el análisis discriminante final. El Test-T nos ayudará a rechazar aquellas variables con medias y varianzas iguales para las dos poblaciones a discriminar (son las variables con baja o nula significación estadística según la T de Student). El test ONE-WAY con su estadístico F de Fisher también nos ayudará a valorar la selección de la variable según se ha comentado previamente.

### **ANÁLISIS DISCRIMINANTE (SPSS)**

Una vez seleccionadas las variables se realiza el análisis discriminante, con todas ellas o combinándolas. Para ejecutarlo debemos ir a *Analizar, Clasificar y Discriminante* donde se abrirá la ventana del análisis discriminante. En ella definiremos la variable de agrupación que es la variable que define los casos estables e inestables y seleccionaremos de la lista de variables las variables independientes (las seleccionadas previamente o una combinación de las mismas). Especificaremos *Usar método de inclusión por pasos* dado que este método permite controlar mejor cómo van entrando las variables en la función discriminante y limita la entrada a las variables con una significación muy baja. También se puede usar el método con todas las variables juntas,

método que no limita la entrada a la función de ninguna variable obteniendo en la función coeficientes muy bajos para aquellas variables con muy poca significación para la discriminación entre celdas con y sin roturas.

En el botón *Estadísticos* especificaremos los Coeficientes de la Función No Tipificados que serán los que utilizaremos en ARC/INFO para generar la malla regular discriminante. En el botón *Método* definiremos los valores F de Entrada y Salida para las variables (los defectos son 3.84 para la entrada y 2.71 para la salida). Utilizaremos estos valores como valores de partida para irlos reduciendo y observar cuántas variables entran en la función para los nuevos valores. Finalmente en el botón *Clasificar* se debe especificar Todos los grupos iguales y pediremos que muestre la Tabla Resumen y los Gráficos de Grupos Separados. El Gráfico para los Grupos Combinados se puede especificar pero el programa no lo realizará mostrando una nota de advertencia en la cabecera de los resultados del análisis discriminante.

A grandes rasgos el programa divide los resultados en tres apartados. El primer apartado es un resumen de los datos introducidos y número de casos. Se muestran dos tablas, una con el Resumen del procesado para el análisis (clases no ponderadas, N y porcentaje para cada clase) y la tabla Estadísticos del Grupo con el número de casos ponderados de cada variable para cada grupo (estables / inestables) y con el total de los casos. El segundo apartado es el Análisis propiamente dicho que a su vez consta de dos subapartados: Estadísticos por pasos y Resumen de las Funciones Canónicas Discriminantes. En Estadísticos por pasos se muestran 4 tablas: la tabla de Variables introducidas/eliminadas con la variable introducida en cada paso y su estadístico entre otros, la tabla Variables en el Análisis con las variables que forman cada paso, su tolerancia, el F que eliminar y la Lambda de Wilks. La tercera tabla es Variables No Incluidas en el Análisis, siendo la misma tabla que la anterior pero sólo con las no introducidas en cada paso y por último la tabla Lambda de Wilks para cada paso. En el subapartado de Resumen de las Funciones Canónicas Discriminantes se muestran 6 tablas pequeñas. Se trata de la tabla de Autovalores de las funciones, la de Lambda de Wilks, la de los Coeficientes Estandarizados de las Funciones Discriminantes Canónicas con los coeficientes de las variables de la función final, la tabla Matriz de Estructura con las variables ordenadas por el tamaño de la correlación con la función discriminante, siendo las dos últimas tablas la de los Coeficientes de las funciones Canónicas Discriminantes (los no tipificados) y la de los Centroides para cada Grupo o población (con y sin roturas). El Tercer y último apartado describe los Estadísticos de Clasificación con 3 tablas y 2 gráficos. Primero se muestra la Tabla de Probabilidades previas para los grupos (especificadas previamente) observándose seguidamente los dos gráficos con los Histogramas de los Grupos por separado. Se muestra luego la Tabla Resumen de la Clasificación con el porcentaje de aciertos para cada grupo (población) y el porcentaje total y, finalmente, una tabla Resumen del Proceso de Clasificación con el número de casos excluidos y usados para la clasificación.

De todos estos datos hay que destacar la tabla de Variables en el Análisis, Los Coeficientes Estandarizados y los no Tipificados, así como los histogramas para cada grupo, los Centroides y la tabla resumen de la Clasificación. Con la observación de estos datos podemos ir modificando la combinación de variables inicial del análisis por otra combinación, variando también la F de entrada y de salida (dentro de los rangos permitidos por el programa).

En general, con este procedimiento los valores discriminantes grandes están asociados a laderas con roturas, mientras que valores pequeños se asocian a laderas estables. Por ello, valores elevados de las variables con un coeficiente estandarizado positivo de la función discriminante contribuyen a incrementar el valor de la función, teniendo un efecto negativo en la estabilidad de la ladera. Variables con signo negativo y valores elevados tienden a favorecer la estabilidad.

### SELECCIÓN DE LA MEJOR FUNCIÓN DISCRIMINANTE (SPSS)

En este punto se evalúan los resultados con las distintas combinaciones de variables, hechas con el análisis discriminante, para decidir cuál será la función más adecuada para generar el mapa de susceptibilidad. Los criterios que rigen la selección de la función discriminante final se han comentado en el Capítulo 3 de la presente memoria (mayor acierto de celdas inestables combinado con el menor número de variables en la función). La selección se realiza con la ayuda de la tabla creada con todos los análisis de las combinaciones hechas (véase Tabla 3.7).

Una vez escogida la función deberemos tener presentes los coeficientes no tipificados para generar la malla regular discriminante en el paso siguiente.

### OBTENCIÓN DE LA MALLA REGULAR DISCRIMINANTE (ARC/INFO)

Una vez escogida la función final retornamos a ARC/INFO dentro del subprograma GRID para generar la malla regular discriminante. Para ello necesitaremos los coeficientes No Tipificados para cada variable de la función final. Una vez definida la extensión del área de trabajo en GRID (el MDE) generaremos la malla regular discriminante con una expresión como la siguiente:

$$DIS = (C1 * VAR1) + (C2 * VAR2) + \dots + (CN * VARN)$$

Donde C1... CN son los coeficientes no tipificados de cada variable y VAR1...VARN son las variables respectivas de estos coeficientes.

Las variables deben existir en ARC/INFO en forma de mallas regulares, situación que corresponde a la mayor parte de las variables, excepto a aquellas que se hayan transformado en el programa SPSS y no estén generadas como malla regular (como los logaritmos de las variables ACUENCA, LONG, AFS, LFS, etc.). En estos casos se utiliza la función LOG10 de GRID para generar las variables transformadas como se muestra a continuación:

$$LACUENCA = LOG10 (ACUENCA1)$$

Una vez generadas las variables transformadas necesarias se puede generar la malla regular discriminante con los coeficientes no estandarizados usando la siguiente expresión (para el caso de la Población de Lillet) en el subprograma GRID:

$$DIS = (0.181 * PEND) + (- 0.257 * PLA) + (- 0.038 * PENDM) + (- 0.078 * VEGET) + (0.357 * LONG) + (0.761 * GROSOR) + (-0.002 * MDE) - 1.927$$

**OBTENCIÓN DEL MAPA DE SUSCEPTIBILIDAD (ARC/INFO)**

Dependiendo de las variables entradas en la función y de los coeficientes no tipificados, los valores de la malla regular discriminante DIS pueden oscilar de  $-6.5$  a  $6.5$ . Lo usual, sin embargo, son valores entre  $-4$  y  $4$ . Para poder categorizar la malla regular DIS necesitamos conocer sus valores máximo y mínimo así como su distribución en un histograma. Con el comando DESCRIBE obtenemos los valores máximo y mínimo, la media y la desviación estándar. Con el comando HISTOGRAM creamos un histograma de la malla regular DIS para observar los rasgos más destacados de su distribución.

El mapa de susceptibilidad generado aquí se realiza con **rangos equidistantes de los valores de la malla regular DIS**. Esto conlleva que las categorías medias abarcan la mayor parte de las celdas de la zona estudiada debido a la poca separación entre poblaciones que hace la función discriminante. El mapa de susceptibilidad se realiza íntegramente en GRID utilizando la función SLICE con el argumento EQINTERVAL para calcular rangos equidistantes de la malla regular discriminante. El comando es:

```
SUS = SLICE (DIS, EQINTERVAL, 7)
```

De esta forma el programa crea una malla regular con 7 valores (del 1 al 7) siendo el 1 menos susceptible y el 7 susceptibilidad máxima. Comandos como GRIDSHADES permiten visualizar el resultado (ESRI, Arcplot Commands, 1994).

**OBTENCIÓN DEL ÍNDICE DE FIABILIDAD / VALIDACIÓN DE LA FUNCIÓN (ARC/INFO Y EXCEL)**

En este último apartado del procedimiento se describe un índice indicativo del grado de fiabilidad de la función (Baeza, 1994) y como aplicarlo a la zona de estudio para validar la función.

Para calcular el índice necesitamos listar la malla regular SUS con el comando LIST. Esto permite ver el número de celdas para cada categoría del mapa de susceptibilidad. Pero el número de celdas listado contiene tanto las celdas utilizadas en la muestra con la que se ha obtenido la función discriminante, como el resto de celdas. Para poder validar la función y usar el índice de fiabilidad debemos generar una malla regular nueva que sea igual a SUS pero sin las celdas de la muestra utilizada. Esto se realiza usando la malla regular SAMPLE que es la malla regular con la muestra aleatoria de las celdas con y sin roturas, usándola para construir una máscara inversa con el comando ISNULL de GRID, donde los valores NODATA serán los de las celdas de la muestra utilizada. Las expresiones para generar esta nueva malla regular son las siguientes:

```
GRID = ISNULL (SAMPLE)  
GRID1 = SELECT (GRID, 'VALUE = 1')
```

La malla regular GRID1 la utilizaremos como máscara para generar la malla regular de susceptibilidad del resto de celdas (que llamaremos SUSCEPT). El paso

siguiente es obtener el número de celdas inestables que caen en cada categoría de susceptibilidad, tarea que se realiza generando una malla regular intermedia (que luego borraremos) sumando SUSCEP con la malla regular de deslizamientos MOV (estando todavía en el estatus de la máscara de GRID1). Listando esta nueva malla regular y **comparando** los resultados con la malla regular SUSCEP se obtiene el número de celdas con roturas para cada categoría del mapa de susceptibilidad. Estos pasos se realizan con las expresiones siguientes:

```
SETMASK GRID1
SUSCEP = SUS
LIST SUSCEP.VAT
GRID2 = MOV + SUSCEP
LIST GRID2.VAT
```

Una vez se tienen los datos se puede salir del estatus de máscara con el comando SETMASK OFF de GRID y se pueden borrar las mallas regulares intermedias generadas (GRID, GRID1, GRID2). Con estos datos calculamos el Índice de Fiabilidad de la función en el **programa EXCEL** o en alguna hoja de cálculo a partir de la expresión:

$$\text{Índice de Fiabilidad} = (n / N) / \Sigma (n / N) * 100$$

Siendo n el número de deslizamientos y N el número de celdas de una categoría o nivel de susceptibilidad determinado y  $\Sigma (n / N)$  es el sumatorio de todos los niveles de susceptibilidad, todo ello multiplicado por 100. Valores altos de este índice implican niveles de susceptibilidad con gran densidad de deslizamientos en el territorio.