

The statistical analysis of compositional data

Prof. Dr. Juan José Egozcue

Práctica sobre coordenadas

Se basa en el libro Excel llamado “Prac_coord”

1. Primera hoja: “elections”

Escenario: se han registrado el número de votos en 5 distritos en unas elecciones a las que se presentan 6 partidos. Los que tienen inicial L se suponen de “Izquierda”, y los de inicial R se suponen de “Derecha”. Se trata de analizar la información de estos pocos datos.

Procedimiento:

- Construir en las columnas I-N las frecuencias relativas observadas. ¿Qué pasaría si hubiera algún cero en los datos de votación?
- Determinar las coordenadas (columnas O-S) de las composiciones de voto según la partición secuencial binaria codificada, en las mismas columnas, en el cuadro amarillo. Se trata de construir la matriz clr de la base en el cuadro en blanco titulado *coefficients*. Y con esa matriz obtener las coordenadas. Si la construcción es correcta, al cambiar el código de signos de la partición deben cambiarse adecuadamente las coordenadas. Ayuda: al elaborar la matriz clr de coeficientes clr conviene utilizar la instrucción SI(;;;) o IF(;;;) de excel.
- Construir, a partir de las coordenadas, las inter-distancias entre los distritos en las columnas U-Y. Examinar e interpretar estas inter-distancias. Lo más práctico es calcularlas manualmente (por eso hay pocos datos).
- En las columnas AA—AE construir la matriz de correlaciones de las coordenadas. Analizarlas e interpretarlas. Hay una función Excel CORRELACION ó CORRELATION.
- Obtener las representaciones que se crean necesarias para visualizar los datos (tanto en el diagrama ternario como en coordenadas).

2. Segunda hoja: “Normals2”

Objetivo: simular una muestra normal bi-variada y suponer que esos valores son coordenadas de una composición de S3. Obtener la representación de la muestra en el ternario y en coordenadas.

Procedimiento:

- Introducir valores en la matriz triangular A9:B10 “Cholesky”. Con ello se obtiene una matriz de covarianzas en A13:B14 que corresponde a la simulación que se pretende.
- En las columnas D y E simular una muestra uniforme (0,1) de la extensión deseada (p.e. 50, 100). Utilícese ALEATORIO() ó RANDOM() de Excel (los resultados son volátiles).
- En las columnas F-G, obtener una simulación normal con la media especificada en A6:B6 y la covarianza en A13:B14. Para ello debe multiplicarse la simulación uniforme D-E como columna por la matriz “Cholesky”. A

- continuación debe sumarse la media. Representar esa muestra normal en un gráfico de dispersión de Excel y comprobar que lo anterior es correcto.
- (d) En las columnas I-J construir el código de signos y la matriz Psi de la base que se desee.
 - (e) En las columnas L-N calcular las partes que corresponden a la simulación normal, según la base elegida, sin clausurar. Obtener los valores clausurados en las columnas P-R.
 - (f) Obtener un gráfico en el ternario de la muestra. Si se hace con CoDaPack el gráfico será fijo para la muestra aleatoria vigente en el momento de hacer el gráfico. Si se quiere obtener un gráfico dinámico, puede construirse un gráfico ternario siguiendo las instrucciones del apéndice A de los apuntes del curso. También pueden hallarse ejemplos de este tipo de gráficos en hojas Excel que acompañan a la de esta práctica.