

STATIS DUAL: SOFTWARE Y ANÁLISIS DE DATOS REALES

WILLIAM CASTILLO ELIZONDO*– JORGE GONZÁLEZ VARELA†

Recibido: 17 abril 1998

Resumen

En este artículo se presenta una implementación computacional con MATHEMATICA del método Stasis dual y se usa este método para analizar el fenómeno del crecimiento de una variedad de caa de azúcar.

Palabras-clave: análisis multivariado de datos, MATHEMATICA, tablas múltiples, trayectoria, análisis de crecimiento, caa de azúcar.

Abstract

We present a computational implementation in MATHAMATICA of the dual Stasis method and we use the method for the growth analysis of some kind of sugar cane.

Keywords: multivariate data analysis, MATHEMATICA, multiple tables, trajectories, growth analysis, sugar cane.

AMS Subject Classification: 62-07, 62H99, 62P10.

1 Introducción.

El objetivo principal de este artículo es presentar una implementación computacional del método Stasis dual y una aplicación con datos reales. Como se sabe, el Stasis es un método para el estudio simultáneo de K tablas de datos Z_1, \dots, Z_K de individuos por variables continuas, donde los individuos permanecen fijos. Mientras que el Stasis dual se aplica al caso en que los individuos pueden cambiar de una tabla a otra. En este caso cada tabla de

*CIMPA-PIMAD, Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica. Fax: +(506) 207 4397. E-Mail: wcastill@cariari.ucr.ac.cr

†Misma dirección. E-mail: jgonzale@cariari.ucr.ac.cr

¹Centro de Investigaciones en Matemáticas Puras y Aplicadas (CIMPA).
Escuela de Matemática, Universidad de Costa Rica.

datos Z_k es de tamaño $n_k \times p$ y los tripletes (Z_k, M, D_k) se llaman estudios, donde M es una matriz $p \times p$ definida positiva y D_k es la matriz diagonal de los pesos de los individuos. La matriz diagonal D_k es la matriz de pesos de los estudios; generalmente $D_k = \text{diag}(\frac{1}{K})$. En la sección 3 se analizan datos de esta naturaleza donde los individuos son las plantas de caa de azúcar que son cortadas cada mes para medirles ciertas características. El estudio evolutivo de entidades financieras (los individuos) donde unas nacen y otras desaparecen conforme pasa el tiempo, responde a un modelo de datos que puede ser analizado con el método Statis dual.

El Statis dual sigue la misma estrategia que Statis, la cual consiste de tres etapas que vamos a presentar esquemáticamente.

1. **Interestructura:** Esta etapa tiene el objetivo de estudiar la diferenciación global entre tablas de datos. Para eso se utiliza la distancia inducida por el producto escalar de Hilbert-Schmidt entre los operadores $R_k M$, donde se supone que R_k es la matriz de covarianzas o correlaciones de las columnas de Z_k , M es una métrica euclídea en el espacio de individuos, con frecuencia igual a la identidad.

La matriz S de productos escalares de Hilbert-Schmidt es por definición $S_{ij} = \text{tr}(R_i M R_j M) = \sum_s (R_i M)(s)(R_j M)^t(s)$, donde $(R_i M)(s)$ es la fila s de la matriz $R_i M$.

Se sabe que las filas de la matriz $H = U \Delta_\lambda^{\frac{1}{2}}$ forman una imagen euclídea exacta de los operadores $R_k M$, donde U es la matriz cuyas columnas son los vectores propios de SD_k , D_k ortonormados. Δ_λ es la matriz diagonal de los valores propios correspondientes mayores que cero (ver [7]). Tomando los dos primeros vectores propios se obtiene una imagen euclídea bidimensional de los operadores, la cual es óptima entre todas las de rango 2 y cuya representación (usando un referencial ortonormado) da los mapas de la interestructura. En la sección siguiente se indica el cálculo de dos tipos de mapas: un mapa normalizado para observar los cambios en la estructura correlacional de las matrices Z_k y que se obtiene usando la matriz S normalizada. El segundo se construye con la matriz S centrada (y no normalizada) que permite observar la variación aproximada de las distancias entre los operadores.

2. **El Compromiso:** Consiste en calcular una configuración llamada compromiso que es representativa de las K configuraciones y cuyo papel es definir un escenario para la representación de las trayectorias de los individuos y las variables. La construcción de dichas configuraciones requiere la diagonalización del operador CM con $C = \sum_{k=1}^K \beta_k \frac{R_k}{\|R_k\|}$ y $\beta = (\beta_1, \dots, \beta_K)$ es un primer vector propio de SD_k tal que $\sum_{k=1}^K \beta_k = 1$.
3. **Intraestructura:** Es la etapa en la cual analizamos la trayectoria de las variables asociadas a cada tabla de datos, a través de los K estudios. También se consideran las trayectorias de los individuos en los períodos donde permanecen fijos. Lo anterior se completa con mapas de las variables observadas en el período entero, llamadas supervariables.

Sea V la matriz cuyas columnas son los vectores propios M ortonormados de CM , entonces la matriz $\beta_k R_k (V \Delta_\lambda^{-\frac{1}{2}})$ contiene en la fila j -ésima las coordenadas de la variable suplementaria x_k^j donde Δ_λ es la matriz diagonal de los valores propios de CM , mayores que cero. Por otra parte, el vector de coordenadas del individuo $x_i(k)$ en los ejes del compromiso es $x_i(k)MV$ donde $x_i(k)$ es la fila i -ésima de la matriz Z_k centrada y reducida. A partir de esta información se construyen los mapas bidimensionales de las trayectorias de las variables, el mapa de las supervariables y cuando corresponde, las trayectorias de los individuos.

2 Implementación del Statis dual con *Mathematica*

Según nuestra información al momento de escribir este artículo existen muy pocas implementaciones computacionales del Statis y Statis dual. En su artículo Lavit, Ch et al. ([8]) mencionan un software que será implementado en SPAD (un paquete especializado en Análisis de Datos) y presentan su programa ACT para Statis y Statis dual. Por otra parte el sistema PIMAD contiene un módulo de implementación de Statis (ver [5]).

A nivel de las necesidades de investigación en Análisis de Tablas Múltiples y otros tópicos del Análisis de Datos, la programación con *Mathematica* es una opción muy adecuada porque facilita el cálculo en espacios euclídeos. El procedimiento ACPPlot (y Conteos como auxiliar) desarrollado por Arce, C. (ver [1]) es una herramienta muy eficiente para la construcción de gráficos (mapas bidimensionales). Puede decirse que todo investigador en Análisis de Datos y de otras áreas que posean alguna experiencia en la manipulación de paquetes, puede sacar provecho de los recursos que hemos mencionado. En esa perspectiva presentamos una secuencia de instrucciones (con breves comentarios) que permiten hacer los cálculos involucrados en el método Statis dual.

El cálculo matricial en *Mathematica* así como la definición de funciones y otros procedimientos siguen muy de cerca las definiciones matemáticas usuales. Esta característica facilita mucho la programación, por ejemplo, de los métodos de análisis de datos cuyos fundamentos teóricos se desarrollan en el marco de los espacios euclídeos, como es el caso del método Statis y Statis dual. Así entonces, por regla general se prefiere escribir en forma de operaciones matriciales las distintas fórmulas propias del Statis-dual y luego traducirlas al lenguaje de *Mathematica*. Dichas fórmulas aparecen en ([7]). En [3] el lector hallará un buen material para adquirir los fundamentos sobre álgebra de matrices con *Mathematica*, y en [4] se encuentran los fundamentos generales para aprender *Mathematica*¹.

2.1 Lectura de datos, definición de algunas variables y funciones

Los datos que se analizan son K tripleteos (Z_k, M, D_k) donde Z_k son las tablas de datos de tamaño $n_k \times p$, n_k es el número de individuos asociados a la ocasión k -ésima, p es el número de variables. Estas tablas de datos son archivos tipo texto los cuales, suponiendo

¹Estos materiales se encuentran disponibles a través de los autores.

que se hallan en `c:\winmath22`, se leerían desde *Mathematica*, con la instrucción²

```
Z1=ReadList[‘nombre.prn’,Real,RecordLists -> True];
```

donde `nombre.prn` es el nombre del archivo que contiene la tabla de datos Z_1 y Z_1 es el nombre de variable que se le da a esta tabla dentro de *Mathematica*.

A continuación se dan las definiciones que proveen a *Mathematica* de la información necesaria para los cálculos posteriores.

```
Z={Z1,Z2,Z3};
K=Length[Z]; (* número de tablas *)
p=Length[Z1[[1]]]; (* número de variables *)
M=IdentityMatrix[p];
Dk=Table[1/K,{i,K}]; (* lista de los pesos de los estudios *)
d=Table[1/Length[Z[[j]]],{j,K},{i,Length[Z[[j]]]}; (* lista de los pesos de los individuos *)
Normaliza[x_,D_]:=x/Sqrt[x.(D x)]; (* esta función normaliza el vector x de acuerdo con la métrica inducida por la matriz D *)
```

2.2 Centraje y reducción de los datos

La operación de centraje de las matrices Z_k consiste en restar a cada entrada de cada columna, su media³. En *Mathematica* esto se escribe así:

```
ZCT = Table[Transpose[Z[[j]]]-Apply[Plus,d[[j]] Z[[j]]],{j,K}];
```

La reducción consiste en dividir cada variable centrada por su desviación estándar.

Esto se logra aplicando a cada fila de la matriz $ZCT[[j]]$ la función `Normaliza`:

```
ZCR = Table[Map[Normaliza[#,d[[j]]]&,ZCT[[j]]],{j,K}];
```

2.3 Matrices de correlación

Nótese que ZCR es la lista de las matrices Z_k transpuestas, centradas y reducidas. Por eso la matriz de correlaciones de las columnas de Z_k es $ZCR[[k]].(d[[k]] Transpose[ZCR[[k]]])$, por lo tanto la lista de matrices de correlaciones es:

```
R=Table[ZCR[[k]].(d[[k]] Transpose[ZCR[[k]]]),{k,K}];
```

2.4 Cálculo de la matriz S de productos escalares entre matrices de correlación

Sea R_k la matriz de correlaciones de las columnas de Z_k y S la matriz de productos escalares. Entonces $S_{kl} = \text{tr}(R_k M R_l M) = \sum_j R_k M(j) (R_l M)^t(j)$ donde $R_k M(j)$ es la fila j -ésima de la matriz $R_k M$. Esto traducido a *Mathematica* es:

²Se distinguirán las expresiones escritas en *Mathematica* porque se enfatizan con el tipo de letra `Typewriter`.

³Normalmente se complementa el estudio multivariado con unas estadísticas univariadas. En este caso las medias y las desviaciones de las variables, las cuales pueden calcularse así:

```
Medias=Table[Apply[Plus,d[[i]] Z[[i]]],{i,K}];
Norma[x_,D_]:= Sqrt[x.(D x)];
desviac=Table[Map[Norma[#, d[[j]]]&,ZCT[[j]]], {j,1,K}];
```

```
RM=Table[R[[k]].M,{k,K}];
S=Table[Sum[RM[[k,j]].Transpose[RM[[s]]][[j]],{j,1,p}],{k,1,K},{s,1,K}];
```

2.5 Coordenadas de la interestructura

Se distinguirán dos tipos de representación de la interestructura: la “no centrada normalizada” y la “centrada no normalizada”. La matriz S normalizada (S_n) se obtiene así:

```
Aux=N[DiagonalMatrix[Table[1/Sqrt[S[[i,i]]],{i,K}]]];
Sn=Aux.S.Aux; (* matriz de productos escalares normalizados *)
```

La matriz de coordenadas de la interestructura no centrada normalizada es $H1 = U\Delta_\lambda^{\frac{1}{2}}$ donde U es la matriz cuyas columnas son los vectores propios de $S_n D_k$, D_k ortonormados y Δ_λ es la matriz diagonal de los valores propios correspondientes. Entonces tenemos:

```
{vlp1,vcp1}=Eigensystem[Transpose[Dk Sn]];
vcpn = Map[Normaliza[#, Dk]&,vcp1];
H1 = Transpose[Map[Sqrt,vlp1] vcpn];
```

Similarmente para el caso de la interestructura ‘centrada no normalizada’, se calcula su matriz de coordenadas H2. Debe recordarse que la matriz S centrada es

$$S_c = (I - 11^t D_k) S (I - D_k 11^t)$$

donde $1 = (1 \dots 1)_{K \times 1}^t$. Por lo tanto:

```
Cent = IdentityMatrix[K]-Table[1/K,{i,K},{i,K}];
Sc = Cent.S.Cent;
{vlp2,vcp2}=Eigensystem[Transpose[Dk Sc]];
vcp2 = Map[Normaliza[#, Dk]&,vcp2];
H2 = Transpose[Sqrt[vlp2] vcp2];
```

2.6 Calidad de la representación centrada no normalizada

La calidad de la representación de una matriz de correlación será el cociente de la norma de su proyección ortogonal entre la norma de la matriz:

```
<<Conteos.ma;
I2={1,1};
Cal=100 Map[Norma[#,I2]&, Columna[H2,{1,2}]]/Table[Sqrt[Sc[[i,i]]],{i,K}];
```

2.7 Cálculo y diagonalización del operador compromiso

Recordemos que el operador compromiso es un promedio ponderado de las matrices de correlación R_k normalizadas: $C = \sum_{k=1}^K \beta_k \frac{R_k}{\|R_k\|}$ donde $\beta = (\beta_1 \dots \beta_K)$ es un primer vector propio de $S_n D_k$ tal que $\sum_k \beta_k = 1$. El cálculo y diagonalización del compromiso C procede así:

```
beta = vcp1[[1]]/Apply[Plus,vcp1[[1]]];
Comp = beta.(R/Table[Sqrt[S[[i,i]]],{i,K}]); (* aquí Comp=C *)
{vlpc,vpcp} = Eigensystem[Comp.M]; (* los vectores propios vpcp[[j]] son M
ortonormados, siempre que M=I*)
```

2.8 Coordenadas de la intraestructura

Se sabe que la representación de la intraestructura en Statistica dual concierne a las trayectorias de las variables asociadas a cada tabla, a las supervariables y a los individuos en los períodos donde no cambian.

Sea V la matriz cuyas columnas son los vectores propios M ortonormados de CM , entonces la matriz $\beta_k R_k (V \Delta_\lambda^{-\frac{1}{2}})$ contiene en la fila j -ésima las coordenadas de la variable suplementaria x_k^j . Por lo tanto las matrices de coordenadas de las variables suplementarias y de los individuos, se pueden calcular así:

```
coordvar = beta R.Transpose[vpcp/Sqrt[vlpc]]; (* aquí vpcp= $V^t$  *)
```

2.9 Representaciones gráficas

Para lograr una buena presentación de los gráficos es necesario ensayar varias veces hasta que las etiquetas queden ubicadas de manera legible. Por esa razón se aconseja usar una nueva hoja de *Mathematica* para la construcción de los gráficos. Para eso se deben guardar las coordenadas calculadas anteriormente (H1, H2 y coordvar).

Las representaciones bidimensionales de la interestructura, así como de la intraestructura se harán utilizando el procedimiento `Columna` (el cual se encuentra definido en el paquete `Conteos.ma`), y el paquete `ACPPlot.ma` (ver [1] y [2]). Cada vez que se quiere construir un plano principal se deben elegir los dos ejes y definir la lista de coordenadas. Se indicará brevemente cómo se hace esto.

2.9.1 Representación de la interestructura

Las siguientes órdenes producen el primer plano principal de la interestructura “no centrada normalizada”.

```
<<ACPPlot.ma;
<<Conteos.ma;
ACPPlot[Columna[H1,{1,2}], lista de opciones];
```

La representación de la interestructura “centrada no normalizada” se hace igual usando H2 en lugar de H1. En este caso no se debe utilizar la opción `Circulo -> 1`.

2.9.2 Trayectoria de las variables suplementarias

La instrucción,

```
cdvar=Transpose[Table[Columna[coordvar[[k]],{1,2}], {k,K}]];
produce la lista de coordenadas de las trayectorias en los dos primeros ejes. El plano de las trayectorias se obtiene con:
ACPPlot[cdvar,lista de opciones];
```

2.9.3 Representación de las supervariables

La matriz de coordenadas de las supervariables es,

```
supervar = Apply[Plus, cdvar, 1];
```

y su representación usando los dos primeros ejes se consigue con
`ACPPlot[supervar, lista de opciones];`

3 Análisis de la variedad cp 722086

Utilizamos el método Stasis dual para analizar el crecimiento de la caña de azúcar en la variedad cp 722086, para lo cual se realizaron durante once meses desde marzo de 1995 a enero de 1996, mediciones mensuales de cinco variables en 30 matas seleccionadas al azar mes a mes. Las mediciones de las variables implican la destrucción de la mata, por lo que no hay trayectorias de los individuos. Tenemos entonces once tablas de datos de tamaño 30×5 .

Las siguientes son las variables consideradas activas:

- AFV: Area Foliar verde (cm^2).
- Altura: Altura de los tallos (cm).
- PSHF: Peso seco de las hojas verdes (grs).
- PSCO: Peso seco de los cogollos (grs).
- PSTA: Peso seco de los tallos (grs).

3.1 La Interestructura

Usamos la métrica de Hilbert-Schmidt con $M = I_{11}$ y $D_k = \frac{1}{11}I_{11}$ para obtener una imagen euclídea plana de las once matrices de correlaciones. Primeramente con la matriz de productos internos Sn normalizada (Figura 1) y luego con la matriz Sc centrada (Figura 2). En el círculo (Figura 1) podemos observar aproximadamente las relaciones entre las matrices de correlaciones a través del RV -coeficiente definido como: $RV(R_i, R_j) = \frac{\langle R_i, R_j \rangle}{\|R_i\| \|R_j\|}$.

Por otro lado, en la Figura 2 tenemos una representación aproximada de las distancias entre las matrices de correlaciones. Así entonces, la proximidad de dos puntos bien representados observada en el círculo y en la Figura 2, significa que las matrices de correlación correspondientes son aproximadamente iguales.

Para ayudarnos en la lectura (Figura 2) construimos una tabla con las normas y la calidad de la representación de cada matriz definida como el cociente de la norma de su proyección ortogonal entre la norma de la matriz.

<i>Mes</i>	<i>ma</i>	<i>ab</i>	<i>my</i>	<i>jn</i>	<i>jl</i>	<i>ag</i>	<i>se</i>	<i>oc</i>	<i>no</i>	<i>di</i>	<i>en</i>
<i>100Cal</i>	96	21	99	96	89	49	94	77	68	95	98
<i>Nor</i>	4.1	3.1	4.2	3.9	3.6	3.1	3.3	3.1	2.8	2.7	3.0

De ambos gráficos podemos observar lo siguiente.

1. Los meses de enero, setiembre y mayo son los puntos más alejados, cuestión que muestra cambios importantes (relativamente) en algunas de las variables.

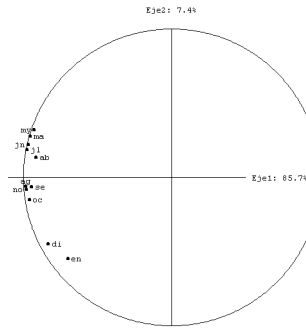


Figura 1: Interestructura normalizada

2. Abril y noviembre están muy próximos, pero es debido a una mala representación del mes de abril (21%). El RV entre ambos es 0.91, pero hay diferencia entre sus normas.
3. {Marzo, junio, julio}, {Agosto, octubre, noviembre} son grupos estables, en el sentido que la estructura de correlaciones se mantiene aproximadamente igual.

3.2 El Compromiso

Para poder encontrar cuáles son las variables que están incidiendo en las observaciones anteriores procedemos a estudiar sus trayectorias a través del compromiso, matriz de correlación que recoge lo que es común a las 11 matrices de correlaciones. La siguiente tabla muestra los pesos que cada mes tiene en el compromiso: $C = \sum_{k=1}^K \beta_k \frac{R_k}{\|R_k\|}$

R_i	ma	ab	my	jn	jl	ag	se	oc	no	di	en
$100\beta_i$	9.39	9.02	9.14	9.54	9.61	9.70	9.32	9.47	9.68	8.20	6.91

3.2.1 Trayectorias de las variables

La diagonalización del compromiso nos permite definir un plano en el cual representamos las trayectorias de las cinco variables. Cada trayectoria está compuesta de 11 puntos, uno

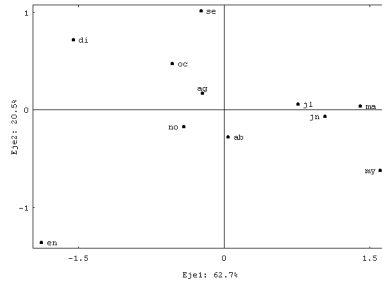


Figura 2: Interestructura centrada

por cada mes. La distancia entre dos puntos es una medida del cambio en la correlación de esa variable con las demás en el período comprendido en esos dos meses.

De las relaciones que siguen concluimos que los cambios más grandes entre puntos de una trayectoria son los responsables de la distancia en las respectivas matrices de correlaciones([7]).

Si R_1, R_2 son dos matrices de correlaciones la distancia al cuadrado entre ellas es:

(1) $\|R_1 - R_2\|^2 = \sum_{i=1}^5 \sum_{j=1}^5 [\text{corr}(X_1^i, X_1^j) - \text{corr}(X_2^i, X_2^j)]^2$, donde $\text{corr}(X_k^i, X_k^j)$ es el coeficiente de correlación lineal entre las variables X_k^i y X_k^j .

(2) $\|R_1\|^2 = \sum_{i,j}^5 \text{corr}^2(X_1^i, X_1^j)$

Una aproximación óptima de orden dos para el compromiso (74.6%) nos muestra que los ejes quedan determinados de la siguiente manera:

3.2.2 Primer eje:(56.4%)

Este eje está definido fundamentalmente por la variable Peso seco de los cogollos (PSCO) y en un menor grado por las variables Peso seco de las hojas verdes (PSHF) y peso seco de los tallos (PSTA). Nos referiremos a él como Eje de los gramos.

3.2.3 Segundo Eje (18.2%)

Este eje está definido por la variable Altura de los tallos (Altura) y en menor grado por Area foliar verde (AFV). Lo denominamos Eje de los centímetros. Teniendo en cuenta esta caracterización del plano, en lo que sigue comentamos las trayectorias de las cinco variables. Los comentarios están en concordancia con la variación observada de los promedios mensuales de las variables (ver sección 4).

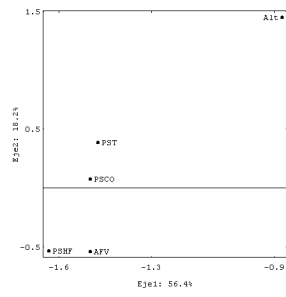


Figura 3: Supervariables

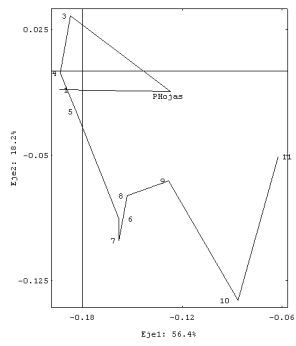


Figura 4: Peso seco hojas verdes

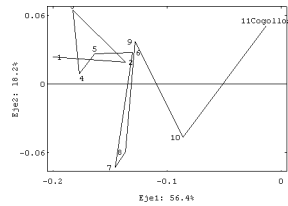


Figura 5: Altura y A. Foliar

3.2.4 Peso seco de las hojas verdes (PSHF):

En el período de setiembre (7) a enero (11) hay cambios en la dirección del primer eje, expresados en una correlación decreciente con el peso seco de los cogollos: 0.75, 0.64, 0.45, 0.35, 0.47 y análogamente con el peso seco de los tallo 0.65, 0.52, 0.25, -0.009 , -0.034 . Esto se traduce en que las tres variables: PSHF, PSTA, PSCO crecen de la misma manera hasta el mes de setiembre (7), a partir del cual PSHF empieza a decrecer, al contrario de las otras dos. En el segundo eje hay una variación de julio (5) a agosto (6) que se muestra por un cambio en la correlación con la altura de los cogollos de -0.01 a -0.1 (Figura 4).

3.2.5 Area foliar verde

Su trayectoria es totalmente análoga al peso seco de las hojas verdes. (ver Figura 4) Correlación entre el área foliar verde y el peso seco, decreciente con el peso seco de los cogollos desde setiembre (7) a enero (11) 0.68, 0.44, 0.37, 0.49, -0.51 y en julio (5), agosto(6) con la altura es 0.3 y 0.11.

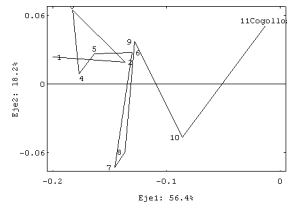


Figura 6: Peso seco: cogollos

3.2.6 Altura

Los cambios más importantes ocurren de setiembre (7) a octubre (8) con correlaciones muy bajas en las variables: AFV, PSHF, PSCO, PSTA que son las siguientes: octubre (8) 0.22, 0.11, -0.19 , -0.028 y setiembre (7) -0.23 , -0.1 , -0.1 , 0.27. Esta variable crece hasta setiembre, a partir de aquí sus cambios son mínimos.

3.2.7 Peso seco de los cogollos

Cambios en el segundo eje en octubre (8) y noviembre (9), la correlación con la altura pasa de -0.19 a 0.4.

3.2.8 Peso seco de los tallos

Movimiento en la dirección del segundo eje de octubre (8) a diciembre (10), con la variable Altura tenemos correlaciones de : $-0,028$, 0.25, 0.25. y con el Area foliar verde de: 0.29, 0.21, 0.005.

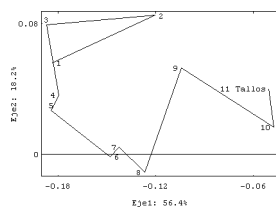
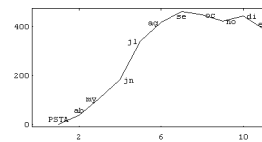
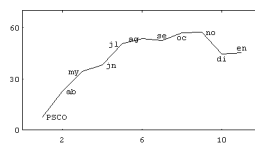
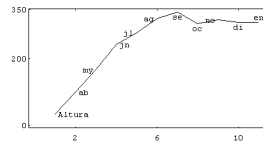
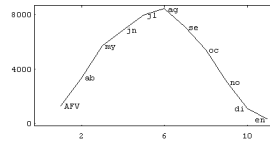
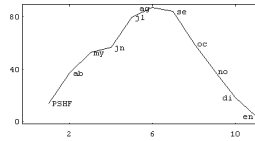


Figura 7: Peso seco: tallos

4 Promedios mensuales de las variables

Con el propósito de que el lector pueda corroborar con los datos los comentarios que sobre las trayectorias hemos hecho, se incluyen los gráficos de los promedios mensuales de las variables.



5 Conclusión

La implementación del Statis dual y su aplicación con datos reales ha permitido mostrar el uso del método y su eficacia en el estudio conjunto de tablas de datos de individuos por variables cuando los individuos cambian. Por otra parte, aun cuando algunos resultados obtenidos sobre el crecimiento de la variedad de caa de azúcar CP722086 eran predecibles para el biólogo experto, el método Statis dual aporta una base científica para explicar el fenómeno y corroborar hipótesis sobre el crecimiento. Por ejemplo, la estabilidad correlacional observada en algunos subperíodos y las variaciones significativas en otros, eran esperadas por el biólogo analista y fueron corroboradas por el método. Además de lo anterior el método tiene la flexibilidad de poder representar trayectorias de los individuos que permanecen durante períodos, lo cual enriquece el análisis. Desde luego este no es el caso de los datos estudiados pues los individuos (las matas de caa de azúcar) cambian en el experimento de un mes al siguiente.

Agradecimientos: Agradecemos a los profesores Walter Marín y Fermín Subirós de la Escuela de Biología de la Universidad de Costa Rica por facilitar los datos y por sus acertadas observaciones sobre los resultados de los análisis. Finalmente agradecemos a nuestro colega Carlos Arce su valiosa ayuda en materia de programación con *Mathematica*.

Referencias

- [1] Arce, C. (1998) “ACPPlot: gráficos del ACP con Mathematica”, *Revista de Matemática: Teoría y Aplicaciones* 5(2): 133–147.
- [2] Arce, C. (1998) *Frecuencias y Tablas de Contingencia con Mathematica*. Escuela de Matemática, Universidad de Costa Rica, San José.
- [3] Arce, C. (1998) *Introducción a Mathematica con Algebra Lineal*. Escuela de Matemática, Universidad de Costa Rica, San José.
- [4] Gray, J.W. (1994) *Mastering Mathematica: Programming Methods and Applications*. Academic Press, Inc., USA.
- [5] González, J.; Rodríguez, O. (1995) “Algoritmo e implementación del método Statis”, *IX Simposio Métodos Matemáticos Aplicados a las Ciencias*, J. Trejos (ed.), UCR-ITCR, Turrialba.
- [6] Lavit, Ch. (1988) *Analyse Conjointe de Tableaux Quantitatifs. Méthode+Programmes*. Masson, Paris.
- [7] Castillo, W. y González, J. (1997) “Análisis de tablas múltiples con individuos cambiantes y variables fijas”, *Revista de Matemática: Teoría y Aplicaciones* 4(1): 51–60.
- [8] Lavit, C.; Escoufier, Y.; Sabatier, R.; Traissac, P. (1994) “The ACT (Statis method)”, *Computational Statistics & Data Analysis, North-Holland*, 18: 97–119.
- [9] Saporta, G.; Lavallard, F. (1996) *L'Analyse des Données Evolutives: Méthodes et Applications*. Editions Technip, París.