

EL ANÁLISIS ESTADÍSTICO EN LA ECOLOGÍA VEGETAL

Linnaeus de Valtimore

1. Introducción

Los trabajos que presentamos formaron parte de la investigación que tutelamos a dos alumnos de 2º de Bachillerato Internacional en el I.E.S. Maestro Matías Bravo (Valdemoro, Madrid).

En estos trabajos hemos aplicado métodos estadísticos al estudio de comunidades vegetales. En realidad planteamos dos partes claramente diferenciadas una en la que se emplea la inferencia estadística para tomar decisiones acerca de las poblaciones estudiadas y otra en la que empleamos métodos multivariantes para clasificar la vegetación según las especies presentes.

Es frecuente pensar en las ciencias sistemáticas como meramente descriptivas y ancladas en su imagen romántica decimonónica. La actualidad es muy diferente. Los métodos moleculares y el tratamiento estadístico de la información han permitido desarrollar métodos de estudio tremendamente complejos que resuelven preguntas de investigación antes inabordables. El desarrollo de aplicaciones informáticas cada vez más potentes han acercado enormemente la utilización de métodos estadísticos que de otra manera requerirían tediosos y complejos cálculos que quedarían restringidos prácticamente al ámbito de los matemáticos.

Hemos utilizado en nuestra experiencia el paquete Statgraphics Plus 5.0 que resulta suficientemente asequible como para que alumnos de bachillerato sean capaces de comprender las hipótesis subyacentes al análisis o inferencia que van a realizar e interpretar razonablemente los resultados. Todo ello salvando el escollo infranqueable que supondría para ellos comprender y realizar todo el cálculo formal que ello implicaría.

Independientemente del interés de la investigación realizada y del trabajo experimental que realizaron, los alumnos obtuvieron un beneficio neto: han podido afrontar en situación muy ventajosa una asignatura de Bioestadística en el primer curso de grado.

2. La vegetación de los cerros vesíferos de la laguna de San Juan (Chinchón)

OBJETIVO

La Laguna de San Juan está situada en la margen izquierda del río Tajuña, muy próxima a su desembocadura en el Jarama. La vegetación predominante en los cerros de la Laguna es el espartal. Los cerros tienen una pendiente suave en la zona más baja que progresivamente se verticaliza en las zonas culminantes (Fig.1).

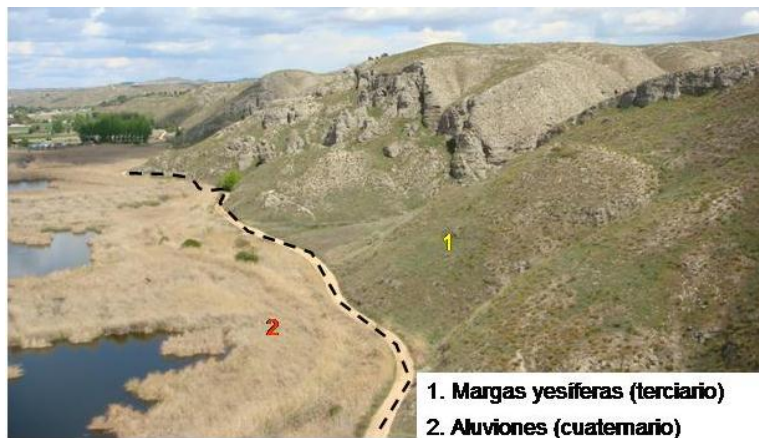


Figura 1. Cerros yesíferos junto a la laguna de San Juan. Las pendientes son suaves en la base y próximas a la vertical en las zonas más altas.

La estructura topográfica de los cerros y la naturaleza soluble y permeable de las rocas nos hacen pensar que pueden desarrollarse gradientes apreciables en diferentes factores abióticos (humedad, pH, concentración de sales...) que repercutirán en una distribución heterogénea de las especies vegetales formando facies diferentes de una misma comunidad.

Para estudiar dicha variabilidad se seleccionaron 11 cuadrados de 5 metros de lado repartidos aleatoriamente en los cerros. En cada cuadrado se identificaron todas las especies presentes y se asignaron índices de cobertura según la metodología de BRAUN BLANQUET (1964) (Tabla 3). Se pretende realizar una la clasificación numérica de los inventarios teniendo en cuenta sólo la presencia o ausencia de especies.

RESULTADOS

Como medida de la similitud utilizamos el índice de afinidad de Jaccard (CUADRAS, 1991) en el que se relacionan el número de especies exclusivas de cada inventario (A y B), con el número de especies comunes entre los dos (C). De esta manera, con el coeficiente de Jaccard (CJ) podemos establecer una relación matemática sencilla de las similitudes de cada par de inventarios (Tabla 1).

$$C.J = \frac{C}{A + B - C}$$

Para valorar la afinidad hemos realizado la prueba de significación χ^2 (MARGALEF, 1986), que permite comparar la similitud de inventarios por parejas (Tabla 2). El estadístico de contraste es

$$\chi^2 = \frac{n(b \cdot c - a \cdot d)^2}{(a + b)(c + d)(a + c)(b + d)}$$


A partir de la tabla de inventarios generamos una matriz cuadrada con los índices de Jaccard que expresa las afinidades entre las entidades reales que se clasifican. A este análisis matricial en el que se agrupan diferentes grupos o “cluster” se le llama análisis de cluster.

	2	3	4	5	6	7	8	9	10	11
1	0,50	0,18	0,33	0,27	0,19	0,25	0,31	0,30	0,20	0,14
2		0,05	0,33	0,27	0,19	0,25	0,31	0,44	0,20	0,07
3			0,14	0,16	0,13	0,20	0,09	0,23	0,13	0,14
4				0,29	0,19	0,17	0,23	0,31	0,16	0,05
5					0,11	0,20	0,19	0,36	0,25	0,06
6						0,13	0,26	0,11	0,09	0,04
7							0,25	0,37	0,14	0,23
8								0,38	0,28	0,17
9									0,27	0,13
10										0,29

Tabla 1. Valores obtenidos para el coeficiente de Jaccard entre los 11 inventarios estudiados.

Los resultados se muestran en un dendrograma en el que se representa un tipo de análisis cluster jerárquico: el método del vecino más lejano (ABRAIRA & PÉREZ DE VARGAS, 1996). Este método usa el análisis de los grupos midiendo la disimilaridad de los individuos más alejados de cada grupo (Figura 2). La distancia de cada especie con respecto a otra se ha medido con la distancia euclídea al cuadrado.

	2	3	4	5	6	7	8	9	10	11
1	0,89	6,68	2,40	3,44	6,15	2,67	2,76	2,86	5,62	7,47
2		14,70	2,40	3,40	6,1	2,70	2,80	1,30	5,60	11,20
3			11,50	9,30	14,50	2,90	16,10	5,20	13,50	11,50
4				4,20	7,47	4,80	6,49	3,34	9,75	15,39
5					12,06	3,75	7,27	2,36	5,33	13,39
6						6,35	6,38	10,88	16,22	18,28
7							2,18	1,14	5,85	4,55
8								2,03	5,73	9,16
9									4,29	8,57
10										4,96

 diferencia no significativa


 diferencia significativa

Tabla 2. Valores del estadístico χ^2 calculados para realizar la prueba de independencia entre cada par de inventarios.

En el dendrograma (Fig 2) aparecen tres grupos claramente diferenciados. Por una parte, el formado por los inventarios 1, 2, 4 y 5, cuya vegetación coincide con la del espartal típico. Un

segundo grupo formado por el 7, 8 y 9 en el que el esparto pierde protagonismo ante la *Reseda suffruticosa*, por que parece ligado a cotas topográficas escarpadas donde la humedad es menor. Por último, los inventarios 3, 6, 10 y 11 constituyen una zona de plantas de carácter nitrófilo en la base de los cerros.

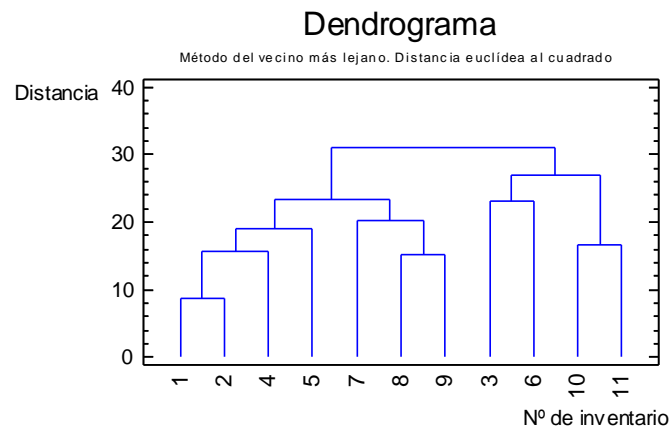


Figura 2. Dendrograma en el que se mide la distancia entre los grupos de inventarios.

El método de agrupamiento de cluster nos ha permitido establecer una clasificación de los inventarios estudiados y plantear hipótesis acerca de sus diferencias. Estos resultados abren nuevas expectativas de investigación. Las hipótesis planteadas deben verificarse y para ello se puede hacer un diseño experimental para comprobar si las diferencias realmente se deben a la humedad edáfica o a la concentración de nitratos en el suelo.

El método tradicional de estudio propio de la fitosociología ya citado anteriormente permite una primera aproximación al estudio de los inventarios. Posiblemente es suficientemente eficaz para comparar inventarios pertenecientes a diferentes comunidades pero la mera observación de la Tabla 1 revela las dificultades para encontrar pautas lógicas en la agrupación de los inventarios levantados. Prácticamente la única información relevante que podríamos obtener es que algunas especies son dominantes en la comunidad y aparecen en todos o casi todos los inventarios.

La utilización del Análisis de Cluster se muestra en este ejemplo como una herramienta tremendamente práctica para el estudio de la vegetación. En realidad simplifica incluso la forma de muestrear ya que no es preciso asignar índices a las diferentes especies del inventario puesto que basta consignar su presencia en cada uno de ellos.

Somos conscientes de la complejidad matemática que encierra este tipo de análisis que excede no sólo los conocimientos de los alumnos de bachillerato sino también de muchos alumnos universitarios de ciencias. El estudio que hemos llevado a cabo no implica que nuestros alumnos de bachillerato comprendan la base matemática del análisis pero sí ha servido para que lo conozcan y comprendan la enorme utilidad que puede tener.

Número de inventario	1	2	3	4	5	6	7	8	9	10	11
Superficie (m ²)	25	25	25	25	25	25	25	25	25	25	25
Pendiente (°)	9	10	15	35	35	25	30	50	85	40	40
Exposición	N	N	W	W	NW	W	W	NW	N	N	N
Fecha	16-4	16-4	17-4	17-4	21-4	21-4	02-5	02-5	16-5	16-5	16-5
<i>Stipa tenacissima</i> L.	2	3	3	2	4	2	4	2	3	3	4
<i>Thapsia villosa</i> L.	+	+	4	2	2	1	2	1	2	1	.
<i>Asphodelus albus</i> L.	2	3	.	1	3	.	.	2	2	2	.
<i>Stipa lagascae</i> Roemer & Schultz	4	4	.	2	.	2	.	2	.	.	.
<i>Reseda suffruticosa</i> Loeffl. ex Koelp.	.	.	.	+	.	.	4	2	3	.	1
Bryophyta sp.	.	.	3	2	3	.	.	.	4	.	.
<i>Grypsophila struthium</i> L. subsp. struthium	+	.	3	.	.	.	2	.	.	1	1
<i>Sedum album</i> L.	2	1	4	.
<i>Helianthemum squamatum</i> (L.) Dum.-Cours.	.	.	1	3	3
<i>Lepidium subulatum</i> L.	2	.	1	2
<i>Erodium cicutarium</i> L'Hérit.	.	+	2	1	.	.
<i>Bromus madritensis</i> L.	3	.	5	.	.	.
<i>Centaurea hyssopifolia</i> Vahl.	5	3
<i>Asparagus acutifolius</i> L.	2	.	3	.	.	.
<i>Poa bulbosa</i> L.	3	1
<i>Leontodon taraxacoides</i> (Vill.) Mèrat.	.	.	1	.	.	2
<i>Mathiola fruticulosa</i> (Loeffl. ex L.) Maire subsp. fruticulosa	+	1	.
<i>Herniaria fruticosa</i> L.	+	+
<i>Helianthemum angustatum</i> Pomel	.	.	4
<i>Bromus tectorum</i> L.	4
<i>Medicago intertexta</i> (L.) Mill.	4	.	.	.
<i>Bupleurum frutescens</i> (Loeffl. ex L.) Maire subsp. fruticulosa	.	.	.	3
Bryophyta sp.	.	.	3
<i>Diplotaxis muralis</i> (L.) DC.	.	.	2
<i>Lygeum spartum</i> L.	.	.	.	2
<i>Helianthemum salicifolium</i> (L.) Mill.	.	.	.	2
<i>Alysum alyssoides</i> (L.) L.	2
<i>Reseda phyteuma</i> L.	2
<i>Melilotus spicatus</i> (Sm.) Breistr.	2	.	.	.
<i>Lithodora fruticosa</i> (L.) Griseb.	1
<i>Helianthemum ledifolium</i> (L.) Mill.	1
<i>Cynosurus eclinatus</i> L.	.	1
<i>Galium hispanicum</i> L.	1
<i>Iberis saxatilis</i> L.	.	.	1
<i>Ferula communis</i> L.	1	.
<i>Avena barbata</i> Pott. Ex Link	1

Además: *Papaver rhoeas* L. + en 6; *Bromus rubens* L. + en 6; *Sanguisorba verrucosa* (Link ex G.Don) Ces. + en 5; *Medicago polymorpha* L. + en 4; *Sedum caespitosum* (Cav.) DC. + en 3; *Eryngium campestre* L. + en 3; *Bellis perennis* L. + en 3; *Tragopogon porrifolius* L. + en 6; *Gladiolus illyricus* Koch + en 11; *Reseda stricta* Pers. subsp. *stricta* + en 11.

Tabla 3. Inventarios mostrando las especies recolectadas en cada una de ellos. Los números representan el valor del índice de abundancia de Braun Blanquet.

3. Metabolismo y transpiración en ciertas especies del parque de las Bolitas del Airón (Valdemoro)

OBJETIVO

¿Responden del mismo modo las plantas a situaciones ambientales de estrés? Intuitivamente podemos pensar que no, pero por otra parte aquellas especies que conviven en un mismo medio deberían mostrar comportamientos similares. Hemos intentado responder a esta respuesta seleccionando ciertas especies y midiendo de alguna manera la intensidad de su actividad. Para ello hemos utilizado ocho especies que crecen en el parque Bolitas del Airón en Valdemoro (Madrid). El interés de este lugar es que se encuentra en una cubeta en la que el nivel freático queda muy próximo y forma suelos muy húmedos. Esto da lugar a la formación de ecotonos (límites entre ecosistemas) muy notorios que separan los suelos higromorfos de los áridos suelos yesíferos del entorno (Fig.3).



Figura 3. Ecotono entre los suelos áridos y los suelos higromorfos en el parque Bolitas del Airón. Nótese el diferente aspecto de la vegetación en ambos medios.

Para evaluar la actividad de los vegetales hemos recurrido a dos indicadores: la tasa de transpiración y la tasa metabólica. La tasa de transpiración mide la cantidad de agua perdida a través de los estomas por los vegetales. En situaciones de estrés hídrico las plantas tienden a cerrar los estomas para evitar pérdidas excesivas de agua aunque esto conlleve no poder obtener CO₂ para realizar la fotosíntesis.

La tasa metabólica es un indicador de la velocidad de las reacciones metabólicas. Para medirla utilizamos la producción neta de CO₂ que depende de la velocidad a la que se gasta dicho gas en la fotosíntesis y su producción durante la respiración. Como ambos procesos son antagonistas, hemos realizado las mediciones en condiciones de oscuridad, de modo que las variaciones de CO₂ sólo puedan deberse a la respiración.

Para estimar la tasa de transpiración introdujimos los individuos seleccionados en el extremo de un tubo de goma relleno de agua con un sensor de presión en el otro extremo. Se sellaron las juntas con vaselina para asegurar el hermetismo. Cuando la planta transpira agua provoca un ascenso en cadena de la columna de agua en sus tejidos que termina transmitiéndose al sensor. El sensor detecta presiones negativas puesto que la planta succiona agua como lo haría una jeringuilla en un vaso de agua.

Como los muestreos se realizaron durante el mes de julio no había muchas especies en desarrollo debido a la sequía estival, especialmente en los pastos xerófilos. Las plantas seleccionadas fueron:

- | | |
|---------------------------|----------------------------|
| 1. PASTOS XERÓFILOS | 2. PASTOS HIGRÓFILOS |
| 1.1. Retama sphaerocarpa | 2.1. Ulmus minor |
| 1.2. Gypsophila struthium | 2.2. Dittrichia viscosa |
| 1.3. Limonium toletanum | 2.3. Alcea hirsuta |
| 1.4. Salsola vermiculada | 2.4. Eleagnus angustifolia |

Se toma para todo el experimento un modelo de efectos aleatorios, ya que se ha tomado tan sólo una muestra de los diferentes valores que la tasa metabólica puede tomar. Tomamos como hipótesis nula que no habrá una diferencia estadísticamente significativa entre los datos de ensayos de una misma planta ni entre plantas de una misma zona. Sí habrá una diferencia estadísticamente significativa entre plantas de zonas diferentes.

Esta hipótesis parte de la suposición de que las muestras de una misma especie tienen un metabolismo muy similar. Además, se añade otra suposición: las plantas de una misma zona presentarán muchas similitudes en su tasa metabólica porque sus adaptaciones al medio son similares. No se tienen en cuenta otros factores que no sean la disponibilidad de agua. Para poder realizar esta simplificación, se realizó el experimento con plantas procedentes de un mismo parque natural, evitando otros factores ambientales como el clima.

El Análisis de la Varianza permite comparar si los valores de un conjunto de datos son significativamente diferentes a los de otro conjunto. El método comparativo está basado en la varianza global que presentan las muestras. Permite dar la probabilidad de que la media de los datos de una población sea diferente de la media de los datos de otra población. En este experimento, se utiliza el Análisis de la Varianza para determinar si existe una diferencia significativa entre la media de la tasa metabólica de una especie vegetal respecto a otra especie, a partir de una muestra estadística de ambas. El ANOVA de un factor parte de dos suposiciones iniciales:

- 1) Los datos de las muestras se ajustan a una distribución normal.
- 2) Homocedasticidad: las varianzas de cada conjunto de datos no deben diferir de forma significativa.

Bondad de ajuste

El estudio de la normalidad permite establecer si un conjunto de datos se ajusta a la distribución normal estándar. El test de bondad de ajuste se basa en el estudio de los estadísticos DMAS, DMENOS y DN global de Kolmogorov.

Se utiliza el test Chi-Cuadrado, que divide el rango de la variable dependiente (del factor a medir) en intervalos no solapados y compara el número de observaciones en cada intervalo con el número esperado basado según la distribución ajustada. El test de Kolmogorov-Smirnov calcula la distancia máxima entre la distribución acumulada de la variable dependiente y la distribución de probabilidad acumulada de la distribución normal ajustada.

Homocedasticidad

El estudio de la homocedasticidad puede realizarse utilizando el contraste de Cochran, el contraste de Bartlett o el contraste de Hartley. Estos dos últimos presentan la ventaja de que pueden realizarse aunque el tamaño muestral no sea constante.

Test de Kruskal-Wallis

Si encontramos en nuestros datos que no se ajustan a la distribución normal o que no podemos suponer la igualdad de varianzas, la prueba de Kruskal-Wallis permite comparar muestras que no se adaptan a la normalidad o que no presentan homocedasticidad, mediante el estudio de las medianas muestrales.

RESULTADOS

Estudio de la tasa de transpiración

Se realizaron mediciones en doce individuos de cada especie. La distribución normal ajustada tiene una media de -0,004885 y una desviación típica de 0,004286. La tabla 4 muestra que el p-valor es menor que 0.01 y podemos rechazar que la tasa de transpiración proceda de una distribución normal con un nivel de confianza de al menos un 99%.

AJUSTE A LA NORMAL
$\chi^2=68,6961$ con 5 g.l. p-Valor = $1,9140 \cdot 10^{-13}$ Estadístico DMAS de Kolmogorov = 0,164417 Estadístico DMENOS de Kolmogorov = 0,201349 Estadístico DN global de Kolmogorov = 0,201349
CONTRASTE PARA IGUALDAD DE VARIANZAS
Contraste C de Cochran: 0,301882 P-valor = 0,0203625 Contraste de Bartlett: 1,4729 P-valor = 0,0000524448 Contraste de Hartley: 15,2568

Tabla 4: Test de bondad de ajuste a la normal y de homocedasticidad para la tasa de transpiración

Respecto a la homocedasticidad, como el menor de los p-valores < 0,05 entonces hay diferencia estadísticamente significativa entre las desviaciones típicas para un nivel de confianza del 95,0%.

El test de Kruskal-Wallis arroja un p-valor de 0,00163191 (menor que 0,05), que indica una diferencia estadísticamente significativa entre las medianas a un nivel de confianza del 95,0%. Para determinar cuáles son las medianas significativamente diferentes entre si, se realiza un gráfico de Caja y Bigotes (Figura 4).

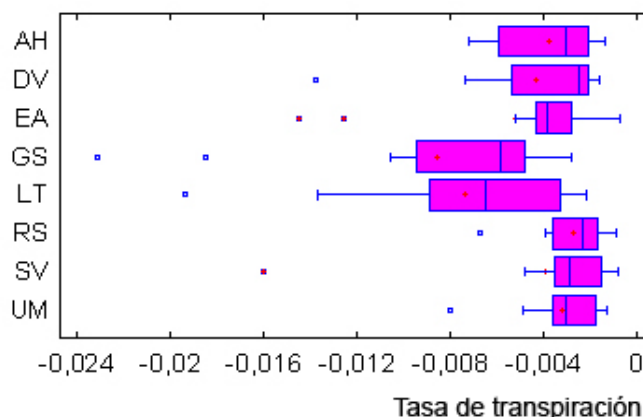


Figura 4: Diagrama de cajas y bigotes para tasa de transpiración

En el gráfico (Fig.4) distinguimos dos grupos: uno con menor presión de succión (que indica mayor tasa de transpiración) formado por las especies *Gypsophila struthium* y *Limonium toletanum*, y otro grupo de mayor tasa de transpiración formada por el resto de especies. Una transpiración más intensa implica que a esas especies les afecta menos la sequía, que pertenecen a la zona de pastos secos. El resto de especies con estas características (*Retama sphaerocarpa* y *Salsola vermiculata*) no presentan diferencias significativas respecto a las plantas de la zona de pastos higrófilos. Su comportamiento nos muestra que, además de la adaptación a la disponibilidad de agua, también influyen otros factores.

Estudio de la tasa respiratoria

La distribución normal ajustada a partir de los resultados tiene una media de 0,269809 y una desviación típica de 0,114509.

AJUSTE A LA NORMAL
$\chi^2 = 8,36358$ con 5 g.l. p-Valor = 0,137304 Estadístico DMAS de Kolmogorov = 0,0858163 Estadístico DMENOS de Kolmogorov = 0,0649009 Estadístico DN global de Kolmogorov = 0,0828163
CONTRASTE PARA IGUALDAD DE VARIANZAS
Contraste C de Cochran: 0,2972 P-valor = 0,0295142 Contraste de Bartlett: 1,18787 P-valor = 0,0657921 Contraste de Hartley: 6,34804

Tabla 5: Test de bondad de ajuste y homocedasticidad para tasa metabólica

La tabla 5 muestra cómo el p-valor es mayor que 0.10, luego no podemos rechazar que la variable tasa metabólica proceda de una distribución normal con un nivel de confianza de al menos un 90%. En este caso se cumple la primera suposición del ANOVA.

La distribución normal de la variable no es suficiente para asumir la validez del análisis de la varianza que es muy sensible a la heterocedasticidad (PÉREZ DE VARGAS & MARTÍNEZ CALVO, 2000:88)

Dado que los p-valores son inferiores a 0,05, hay diferencia estadísticamente significativa entre las desviaciones típicas para un nivel de confianza del 95,0%. Esto infringe una de las asunciones importantes del ANOVA.

El p-valor obtenido en el test de Kruskal-Wallis es $9,0947 \cdot 10^{-5}$ (inferior a 0,05) por lo que hay diferencia estadísticamente significativa entre las medianas a un nivel de confianza del 95,0%.

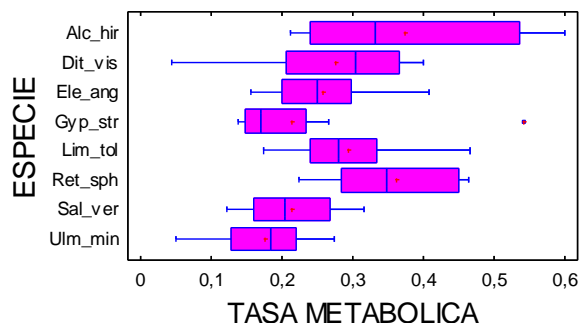


Figura 5: Gráfico de cajas y bigotes de la tasa metabólica

En esta ocasión, son las especies *Ulmus minor*, *Salsola vermiculata* y *Gypsophila struthium* las que parecen tener una tasa metabólica significativamente más baja que el resto. El test de Kruskal-Wallis ha permitido formar otro grupo compuesto por *Dittrichia viscosa*, *Retama sphaerocarpa* y *Alcea hirsuta*, especies con mayor tasa respiratoria (indicador de mayor actividad metabólica). Las especies *Eleagnus angustifolia* y *Limonium toletanum* podrían organizarse en un grupo intermedio entre las plantas con tasa metabólica alta y las especies con tasa metabólica baja.

4. Evaluación

La utilización de las actuales aplicaciones estadísticas en el ámbito de la ecología vegetal está suponiendo una auténtica revolución metodológica en el estudio de los ecosistemas. Sin embargo la simplicidad de uso de estos paquetes informáticos debe tomarse con cautela. Las aplicaciones permite obviar las dificultades de cálculo que los test estadísticos conllevan. Sin embargo hay que conocer muy bien todas las hipótesis y suposiciones que subyacen a cada modelo para no llegar a conclusiones erróneas.

La introducción de este tipo de estudios estadísticos en pequeñas investigaciones llevadas a cabo por alumnos de Bachillerato les sirve de magnífica puerta de entrada en el método científico. Se acostumbran a expresar los resultados con rigor y descubren el tremendo potencial que pone la estadística a sus pies.

5. Referencias bibliográficas

ABRAIRA, V & PÉREZ DE VARGAS, A. (1996) Métodos multivariantes en bioestadística. Editorial Centro de Estudios Ramón Areces. Madrid. 452 pags.

BRAUN-BLANQUET, J. (1964). Fitosociología. Editorial H. Blume.

CUADRAS, C.M. (1991) Métodos estadísticos de análisis multivariante. Ed. Promociones y Publicaciones Universitarias, S.A. Barcelona. 644 pags.

MARGALEF, R. (1986) Ecología. Ed. Omega. Barcelona. 951 pags.