Contribuţii Botanice, *XXXVIII, (2)*, **2003**
Grădina Botanică “Alexandru Borza”
Cluj-Napoca

L'INTENSITÉ ET LA NATURE DES RELATIONS ENTRE TROIS VARIABLES DANS UNE ÉTUDE PHYTOSOCIOLOGIQUE, EN EMPLOYANT LE COEFFICIENT DE CORRÉLATION MULTIPLE

Valeriu ALEXIU, Monica NEBLEA
Universitatea din Piteşti, Facultatea de Biologie, str. Târgu din Vale, nr. 1, **RO-110040 Piteşti**

Abstract: ***The intensity and nature of relations between three variables in a phytosociological study, using the coefficient of multiple correlation.*** The paper tries to demonstrate if the three variables (the altitude, the number of diploidic species and the number of poliploidic species) in the plants association (Telekio - Alnetum incanae), in four places (Iezer, Piatra Craiului, Râiosu, Răstoliţa), are or they are not connected, and the degree of that connection. We used the Galton's Method, that used the separated variables, and calculates their square mean. If the high values of one variable has trended to be associated with high values of other variable, then the square of that two variables has the same sign. Contrary, if a variable increasing is connected with a decreasing of other variable, then the multiplication of the squares will be negative. The method may be used for three and more variables.

Introduction

On peut analyser une variable à la fois, mais cette attitude est dépourvue d’attraits pour le biologiste, qui, par sa formation, réfléchit en termes de relations à l’intérieur des organismes plutôt que de mesures prises isolément. En vérité, il n’y a peut-être pas de source plus puissante d’incompréhension entre statistician et biologiste que cette tendance du premier à penser dans le langage univariate, et du second dans le langage multivariate, sans qu’ils se rendent jamais compte de leurs differences. Dans ce travaille, nous allons nous mettre à considérer les variables dans leurs relations mutuelles. Les exemples sont tirés de divers travaux de biologie, de physiologie végétale, de phytosociologie et de médicine (Pearce S.C., 1978, Drăgulescu C. et Sîrbu I., 1997, Stancu R., 1998, Săhleanu V., 1957), mais la méthode décrite dans les pages suivantes impliquent, pour la première fois, l’emploi simultané de plusieurs variables phytosociologiques (l'altitude, le nombre d'espèces dyploïdes et le nombre d'espèces polyploïdes). Nous avons souligné que la corrélation entre deux variables peut être imparfaite parce qu’elles sont liées de deux ou plusieurs manières contradictoires; cela signifie qu’un acroissement de l’une entraîne chez l’autre une variation importante ou minime, positive ou négative, selon la cause qui est à l’origine de la variation. La situation n’est pas modifiée si leur association résulte de l’existence d’une cause commune; il est encore possible que la forme de leur association varie selon la nature du stimulus qui agit sur les variables mesurées.

Matériaux et Méthodes

On cherche à connaître si trois variables (l'altitude, le nombre d'espèces dyploïdes et le nombre d'espèces polyploïdes) d'une associations végétale (*Telekio speciosae - Alnetum incanae*), en quatre composantes (Iezer-Păpuşa, Piatra Craiului, Râiosu, Răstoliţa), sont liées ou non et, aussi, on cherche à connaître, leur degré de liason. La corrélation traduit le degré d’association entre des variables et non la nature de cette liaison. En fait, la calcul d’un coefficient de corrélation suppose que deux variables sont soumises à des tendances constantes l’une par rapport à l’autre et cette hypothèse soustend aussi le calcule des coefficients de corrélation partielle; quand deux variables sont ajustées pour une troisième, l’ajustement est fait en supposant qu’il existe une relation linéaire entre chaque paire de variables. Il n’en demeure pas moins que, bien qu'on ait nécessairement supposé l'existence de ces relations, notre objectif était d'étudier **l'intensité** de la liaison. Dans le cas de la régression c'est la **nature** de cette relation que l'on prend en considération. Une variation de *x* est associée à une variation proportionnelle de *y*, dans le même sens ou dans l'autre. Si l'on trace les deux droites de régression, celle de *y* sur *x* et celle de *x* sur *y*, nous voyons qu'elles ne sont pas identiques. Si *x* et *y* étaient strictement associées, c'est-à-dire, si $r = \pm 1$, en sorte qu'aucun ne soit affecté par un autre facteur que l'autre, on n'obtiendrait qu'une seule droite correspondant à la loi qui les associe (dans le cas où la relation peut se traduire par une droite). On doit supposer que la droite représentant la relation naturelle réelle passe quelque part entre les deux droites de régression; sa position précise dépend de l'effet plus ou moins grand de l'erreur sur l'une ou l'autre variables. Quand les deux droites de régression sont à angle droit, c'est-à-dire que $r = 0$, la relation réelle, si elle existe, est complètement masquée par l'erreur et toute estimée en vaut une autre.

On utilise la méthode de Galton qui consiste à prendre les variables séparément et à calculer leurs écarts à la moyenne. Si les valeurs élevées de l'une ont tendance à être associées à des valeurs élevées de l'autre et s'il en est de même pour les valeurs faibles, les écarts des deux variables sont en général de même signe et la plupart des produits d'écarts sont positifs. Par contre, si une augmentation de l'une est associée à une diminution de l'autre, la plupart des produits d'écarts sont négatifs.

On arriverait à la même conclusion si, le rôle mutuel des variables étant inversé, on s'attachait à la variation de x, et à la mesure dans laquelle on peut l'expliquer par la variation de y et w. Le procédé ne se limite pas au cas de trois variables. Lorsqu'il y en a quatre, on peut calculer la corrélation partielle entre x et y, en éliminant les effets de v et w.

Iezer-Păpuşa			**Piatra Craiului**		
w (alt.)	***x*** (espèces dyploïdes)	***y*** (espèces polyploïdes)	***w*** (alt.)	***x*** (espèces dyploïdes)	***y*** (espèces polyploïdes)
650	20	6	800	16	9
660	15	7	800	13	11
660	14	7	830	23	11
700	20	7	900	16	12
700	17	14	900	16	12
750	8	15	900	11	12
770	8	15	900	22	12
810	6	16	900	16	11
810	5	20	910	21	11

Răstolița			**Râiosu**		
w (alt.)	***x*** (espèces dyploïdes)	***y*** (espèces polyploïdes)	***w*** (alt.)	***x*** (espèces dyploïdes)	***y*** (espèces polyploïdes)
487	26	13	600	25	11
500	16	15	650	16	11
500	20	15	650	16	12
510	17	15	650	20	14
510	26	19	660	17	14
550	20	17	680	12	15
550	17	17	770	21	16
550	16	20	810	15	17
550	18	22	900	17	19

Dans notre cas, on décide de transformer toutes les mesures en logarithmes parce-que la représentation graphique de log (altitude) en fonction de log (les espèces dyploïdes et les espèces polyploïdes) se traduisait en général par une relation linéaire. La base du "coefficient de corrélation" de Galton est une somme de produits d'écarts. On peut la calculer en faisant d'abord le produit des paires de données, qui donne une somme de produits Total, puis en faisant le produit des totaux et le divisant par le nombre de paires de données, ce qui donne un Terme

Correctif. On voit qu'une somme de produits, à l'encontre d'une somme de carrés, peut être négative.

Résultats et Discussions

Une association entre deux variables n'est pas une preuve de causalité, il arrive quelque fois qu'elle résulte d'une causalité commune, les deux variables étant affectées par une troisième. Appelons-les *w, x* et *y*, appelons leurs somme de carrés des écarts C_{ww}, C_{xx} et C_{yy} et les trois somme de produits des écarts: C_{xy}, C_{wy} et C_{wx}. Le coefficient de corrélation entre *x* et *y* peut varié entre -1 et +1. On peut le calculé par l'equation suivante:

$$r_{xy} = \frac{C_{xy}}{\sqrt{C_{xx}C_{yy}}}$$

Si on élimine ensuite l'effet de *w* à la fois sur *x* et *y*, leur coefficient de corrélation devient:

$$\frac{C_{xy}C_{ww} - C_{wx}C_{wy}}{\sqrt{(C_{ww}C_{yy} - C_{wy}^{2})(C_{ww}C_{xx} - C_{ww}C_{xx} - C_{wx}^{2})}} = \frac{r_{xy} - r_{wx}r_{wy}}{\sqrt{(1-r_{wx}^{2})(1-r_{wy}^{2})}} = r_{xy.w}$$

On le désigne par $r_{xy.w}$: c'est un "coefficient de corrélation partielle".

Iezer

alt (w)	D (x)	P (y)	$w-\overline{w}$	$x-\overline{x}$	$y-\overline{y}$	C_{ww}	C_{xx}	C_{yy}	C_{wx}	C_{wy}	C_{xy}
2,813	1,301	0,778	-0,04494	0,252360	-0,25706	**0,0020**	**0,06369**	**0,0661**	**-0,01134**	**0,01155**	**-0,0649**
2,820	1,176	0,845	-0,03831	0,127421	-0,19011	**0,00147**	**0,01624**	**0,0361**	**-0,00488**	**0,00728**	**-0,0242**
2,820	1,146	0,845	-0,03831	0,097458	-0,19011	**0,00147**	**0,00950**	**0,0361**	**-0,00373**	**0,00728**	**-0,0185**
2,845	1,301	0,845	-0,01276	0,252360	-0,19011	**0,00016**	**0,063686**	**0,0361**	**-0,00322**	**0,00243**	**-0,0480**
2,845	1,230	1,146	-0,01276	0,181779	0,110916	**0,00016**	**0,03304**	**0,0123**	**-0,00232**	**-0,0014**	**0,0202**
2,875	0,903	1,176	0,017204	-0,14558	0,140879	**0,00030**	**0,02119**	**0,0198**	**-0,0025**	**0,00242**	**-0,0205**
2,886	0,903	1,176	0,028633	-0,14558	0,140879	**0,00082**	**0,02119**	**0,0198**	**-0,00417**	**0,00403**	**-0,0205**
2,908	0,778	1,204	0,050627	-0,27052	0,168908	**0,00256**	**0,07318**	**0,0285**	**-0,0137**	**0,00855**	**-0,0457**
2,908	0,698	1,301	0,050627	-0,34961	0,265818	**0,00256**	**0,12229**	**0,0707**	**-0,0177**	**0,01346**	**-0,0930**
25,769	9,438	9,316				**0,01152**	**0,42401**	**0,3257**	**-0,06357**	**0,05560**	**-0,3151**
2,863	1,048	1,035									

Piatra Craiului

alt (w)	D (x)	P (y)	$w-\overline{w}$	$x-\overline{x}$	$y-\overline{y}$	C_{ww}	C_{xx}	C_{yy}	C_{wx}	C_{wy}	C_{xy}
2,903	1,204	0,954	-0,03641	-0,01790	-0,09426	**0,001326**	**0,00032**	**0,0089**	**0,000652**	**0,003432**	**0,0017**
2,903	1,113	1,041	-0,03641	-0,10808	-0,00711	**0,001326**	**0,011681**	**0,0001**	**0,003935**	**0,000259**	**0,0008**
2,919	1,361	1,041	-0,02042	0,139707	-0,00711	**0,000417**	**0,019518**	**0,0001**	**-0,00285**	**0,000145**	**-0,001**
2,954	1,204	1,079	0,014741	-0,01790	0,030677	**0,000217**	**0,00032**	**0,0009**	**-0,00026**	**0,000452**	**-0,0006**
2,954	1,204	1,079	0,014741	-0,01790	0,030677	**0,000217**	**0,00032**	**0,0009**	**-0,00026**	**0,000452**	**-0,0006**
2,954	1,041	1,079	0,014741	-0,18063	0,030677	**0,000217**	**0,032626**	**0,0009**	**-0,00266**	**0,000452**	**-0,0055**
2,954	1,342	1,079	0,014741	0,120402	0,030677	**0,000217**	**0,014497**	**0,0009**	**0,001775**	**0,000452**	**0,0037**
2,954	1,204	1,041	0,014741	-0,01790	-0,00711	**0,000217**	**0,00032**	**0,0001**	**-0,00026**	**-0,00010**	**0,0001**
2,959	1,322	1,041	0,01954	0,100198	-0,00711	**0,000382**	**0,01004**	**0,0001**	**0,001958**	**-0,00014**	**-0,0007**
26,455	10,99	9,436				**0,004537**	**0,089643**	**0,0129**	**0,002012**	**0,005401**	**-0,0021**
2,940	1,222	1,048									

Răstolița

alt (w)	D (x)	P (y)	w-$\overline{w}$	x-$\overline{x}$	y-$\overline{y}$	C_{ww}	C_{xx}	C_{yy}	C_{wx}	C_{wy}	C_{xy}
2,688	1,414	1,113	-0,03048	0,130927	-0,11109	**0,000929**	**0,017142**	**0,0123**	**-0,00399**	**0,003386**	**-0,0145**
2,699	1,204	1,176	-0,01904	-0,07993	-0,04894	**0,000362**	**0,006388**	**0,0024**	**0,001522**	**0,000932**	**0,0039**
2,699	1,301	1,176	-0,01904	0,016984	-0,04894	**0,000362**	**0,000288**	**0,0024**	**-0,00032**	**0,000932**	**-0,0008**
2,708	1,230	1,176	-0,01044	-0,05359	-0,04894	**0,000109**	**0,002873**	**0,0024**	**0,000559**	**0,000511**	**0,0026**
2,708	1,414	1,278	-0,01044	0,130927	0,053718	**0,000109**	**0,017142**	**0,0029**	**-0,00137**	**-0,00056**	**0,0070**
2,740	1,301	1,230	0,022356	0,016983	0,005413	**0,000500**	**0,000288**	**0,0003**	**0,000380**	**0,000121**	**0,0001**
2,740	1,232	1,230	0,022356	-0,05359	0,005413	**0,000500**	**0,002873**	**0,0003**	**-0,00120**	**0,000121**	**-0,0003**
2,740	1,204	1,301	0,022356	-0,07993	0,075994	**0,000500**	**0,006388**	**0,0058**	**-0,00179**	**0,001699**	**-0,0061**
2,740	1,255	1,342	0,022356	-0,02877	0,117387	**0,000500**	**0,000828**	**0,0138**	**-0,00064**	**0,002624**	**-0,0034**
24,487	11,56	11,025				**0,003871**	**0,05421**	**0,0420**	**-0,00685**	**0,009765**	**-0,0115**
2,721	1,284	1,225									

Râiosu

alt (w)	D (x)	P (y)	w-$\overline{w}$	x-$\overline{x}$	y-$\overline{y}$	C_{ww}	C_{xx}	C_{yy}	C_{wx}	C_{wy}	C_{xy}
2,778	1,397	1,041	-0,06831	0,159540	-0,1079	**0,004666**	**0,02545**	**0,0116**	**-0,0109**	**0,00737**	**-0,01721**
2,813	1,204	1,041	-0,03355	-0,03428	-0,1079	**0,001126**	**0,00117**	**0,0116**	**0,0012**	**0,0036**	**0,003699**
2,813	1,204	1,079	-0,03355	-0,03428	-0,07011	**0,001126**	**0,00117**	**0,0049**	**0,0012**	**0,00235**	**0,002403**
2,813	1,301	1,146	-0,03355	0,062630	-0,00316	**0,001126**	**0,00392**	**0,0009**	**-0,0021**	**0,00010**	**-0,0002**
2,820	1,230	1,146	-0,02692	-0,00795	-0,00316	**0,000725**	**0,00070**	**0,0009**	**0,0002**	**0,00008**	**0,000023**
2,833	1,079	1,176	-0,01395	-0,15922	0,026798	**0,000195**	**0,02535**	**0,0007**	**0,0022**	**-0,0004**	**-0,00427**
2,886	1,322	1,204	0,040028	0,083819	0,054827	**0,001602**	**0,00702**	**0,0030**	**0,0035**	**0,00219**	**0,004596**
2,908	1,176	1,230	0,062023	-0,06231	0,081156	**0,003847**	**0,00388**	**0,0066**	**-0,0039**	**0,00503**	**-0,00506**
2,954	1,230	1,278	0,10778	-0,00795	0,129461	**0,011617**	**0,01674**	**0,0168**	**-0,0009**	**0,01395**	**-0,00103**
25,618	11,15	10,343				**0,026028**	**0,06811**	**0,0553**	**-0,0097**	**0,03434**	**-0,01704**
2,846	1,238	1,149									

Nous calculons les sommes de carrés et de produits sous forme de la somme de quatre composantes, une pour chaque localisation, avec chacune 8 degrés de liberté.

Un test F montrera si la corrélation peut ou non être attribuée au hasard.

Un avantage de la somme de carés des écarts et des degrés de liberté est qu'on peut tous deux les subdiviser en parts. Ainsi, dans le Massif Iezer-Păpuşa, pour altitude (w) il y avait 9 données, avec un total de 25,72. Il y avait donc 8 degrés de liberté et la somme des carrés des écarts était:

$$2{,}81^2 + 2{,}82^2 + 2{,}82^2 + 2{,}84^2 + 2{,}85^2 + 2{,}88^2 + 2{,}89^2 + 2{,}91^2 + 2{,}91^2 - \frac{25{,}72^2}{9} = 65{,}34$$

De la même façon, pour les espèces dyploïdes (x) on disposait, aussi, de 9 données avec la somme des carrés des écarts de 8,75, avec 8 degrés de liberté. Le dernière groupe, des espèces polyploïdes (y), a donné une somme des carrés des écarts de 8,54, avec 8 degré de liberté.

Ces chiffres représentent la variation dans les donnèes. L'altitude, la première variable, a fourni pour 9 données une moyenne de 2,86. Une fois éliminée la variation, cette variable contribue pour $9*(2{,}86)^2$ que l'on peut encore écrire $\frac{25{,}72^2}{9}$. De même, la parte des autres variables est $\frac{9{,}44^2}{9}$ et $\frac{9{,}32^2}{9}$, la somme de ces trois quantités représentant la variation des moyennes des variables par raport à zéro. Si l'on considère à présent la variation de la moyenne générale, il y a 27 données avec un total de 44,48 de sorte que le terme qui la représente est $\frac{44{,}48^2}{27}$ et la variation des moyennes des donnés par rapport à la moyenne générale est :

$$\frac{25{,}72^2}{9}+\frac{9{,}44^2}{9}+\frac{9{,}32^2}{9}-\frac{44{,}48^2}{27}=19{,}786$$ avec 2 degré de liberté car il y a 3 variables.

Maintenant, la variation totale du système est celle des 27 donnèes par raport à la moyenne générale. La somme des carés des écarts est: Somme de toutes les données élevées au carré - $\frac{44{,}48^2}{27}$ = 102,42 avec 26 degré de liberté. Il apparaît maintenant que les composantes s'additionnent correctement comme le montre le tableau suivant:

Source de la variation	dl	Somme de carrés
Altitude (w)	8	65,34
Les espèces dyploïdes (x)	8	8,75
Les espèces polyploïdes (y)	8	8,54
Entre variétés	2	19,79
Total	26	102,42

Dans cette étude, on n'avait pas consideré d'autre variation que celle entre les variables: c'est-à/dire que l'erreur est constituée par toute la variation dans les variables et qu'on peut obtenir en cumulant les composantes de la variation dans chacune des trois variables prises isolèment. Cela donne l'analyse de variance de tableau suivant:

Source	dl	Somme de carrés
Variables	2	19,79
Erreur	24	82,63
Total	26	102,42

En employant le test F (Fisher-Snedecor) dans l'analyse de la variance, le carré moyen "variable" était $\frac{1}{2}(19{,}79)=9{,}86$. On doit chercher les valeurs de F dans des tables (Snedecor, 1968, Ceapoiu, 1968, Militaru C., Dumitrescu A., 2001). Leur lecture montre qu'avec un numérateur ayant 2 degrés de liberté et un dénominateur en ayant 26, le F dépasse la valeure de 5,53 par pur hasard seulement une fois sur cent (P=0,01); on dit alors qu'il est très significativ, ce que l'on indique par deux astérisques.

De la même façon on peut faire l'analyse de la variance de ces trois variables dans les suivantes composantes: Piatra Craiului, Răstolița, Râiosu:

Source	Variables		Total		F
	dl	Somme de carrés	dl	Somme de carrés	
Iezer-Păpuşa	2	19,79	26	102,42	$9{,}86^{**}$
Piatra Craiului		19,67		109,52	$9{,}83^{**}$
Răstolița		12,87		97,15	$6{,}43^{**}$
Râiosu		16,42		104,06	$8{,}21^{**}$

Nous obtiendrons le même résultat final, mais cela demand plus de travail, c'est pourquoi on ne procède ainsi que dans le cas où ils pourra être nécessaire de refaire le calcule par la suite de façon plus détaillée.

Iezer-Păpuşa

C_{ww} = 0,011524	C_{xx}= 0,424007	C_{yy} = 0,32569
C_{xy} = -0,31511	C_{wy}= 0,055598	C_{wx} = -0,06357

Piatra Craiului

C_{ww} = 0,004537	C_{xx} = 0,089643	C_{yy} = 0,310801
C_{xy} = -0,00207	C_{wy} = 0,005401	C_{wx} = 0,002012

Răstoliţa

C_{ww} = 0,003871	C_{xx} = 0,05421	C_{yy} = 0,04202
C_{xy} = -0,01146	C_{wy} = 0,009765	C_{wx} = -0,006850

Râiosu

C_{ww} = 0,026028	C_{xx} = 0,068111	C_{yy} = 0,05529
C_{xy} = -0,01704	C_{wy} = 0,034342	C_{wx} = -0,00963

On additionne les quatre composantes pour répondre à notre besoin présent:

C_{ww}	0,045959214	C_{xx}	0,635971205	C_{yy}	0,435866
C_{xy}	-0,34567614	C_{wy}	0,105106372	C_{wx}	-0,07804

On peut faire l'analyse de la variance. Si on désigne par C_{xx} et C_{yy} les sommes des carrés des écarts respectives de x et de y, et par C_{xy} la somme des produits d'écarts, le coefficient de corrélation, que l'on a l'habitude d'appeler $\boldsymbol{r_{xy}}$, est alors:

$$r_{xy} = \frac{C_{xy}}{\sqrt{C_{xx}C_{yy}}} = \frac{-0,34567614}{\sqrt{0,635971205 * 0,435866}} = -0,65655975$$

Le limite de **r** sont ± 1. Il est égal à +1 quand les variations d'une variable sont associées à des variations de l'autre exactement proportionnelles et de même sens. Si elles sont proportionelles, mais en sens inverse, **r = -1**. Si une variation de l'une n'est pas associée à une variation de l'autre, r = 0.

Dans la même manière, on calcule le coefficient de corrélation $\boldsymbol{r_{wy}}$ et $\boldsymbol{r_{wx}}$**:**

$$r_{wy} = \frac{C_{wy}}{\sqrt{C_{ww}C_{yy}}} = 0,742619472$$

$$r_{wx} = \frac{C_{wx}}{\sqrt{C_{ww}C_{xx}}} = -0,45644785$$

Deux décimales suffisent en général pour un coefficient: nous en avons pris davantage car nous aurons à les utiliser par la suite dans des calculs.

Voyons à présent si *y*, le logarithme des espèces polyploïdes, est lié à *w* et *x*, les logarithmes d'altitude et des espèces dyploïdes; c'est ici que le coefficient de corrélation partielle peut nous aider. On a:

$$r_{xy.w} = \frac{r_{xy} - r_{wx}r_{wy}}{\sqrt{(1 - r_{wx}{}^2)(1 - r_{wy}{}^2)}} = -0,38926404$$

D'après cela, il peut sembler que les espèces polyploïdes sont liés à les espèces dyploïdes, surtout parce que les deux sont liés à l'altitude. Toutefois $r_{wy.x}$ est lui-même significative:

$$r_{wy.x} = \frac{r_{wy} - r_{wx}r_{xy}}{\sqrt{(1-r_{wx}^2)(1-r_{xy}^2)}} = 0{,}50189864$$

Ce que montrent ces résultates, c'est que qui veut estimer *y* peut le faire s'il connaît soit *w* soit *x*. S'il connaît l'un, il ne lui sert pas à grand chose de connaître l'autre. Cette fois encore, nous n'avons utilisé les coefficients de corrélation que dans un but descriptif; ils ne traduisent pas nécessairmenet des lois naturelles.

Dans le cas présent, la somme des carrés de *y* est 0,435866, dont on peut attribuer la fraction suivante aux corrélations (corrélation multiple avec *w* et *x* - CM):

$$CM = \frac{C_{wy}(C_{wy}C_{xx} - C_{wx}C_{xy}) + C_{xy}(C_{ww}C_{xy} - C_{wx}C_{wy})}{(C_{ww}C_{xx} - C_{wx}^2)} = 0{,}295906736$$

avec 2 degrés de liberté.

$\frac{C_{wy}^2}{C_{ww}} = 0{,}240372899$ provient de la corrélation avec *w* (1) et

$\frac{C_{xy}^2}{C_{xx}} = 0{,}187888997$ provient de celle avec *x* (2), chacune ayant 1 degré de liberté.

Analyse de la variance des corrélations d'une variable *y* avec deux autres, *w* et *x*:

Source	Degré de liberté	Somme de carrés
Corrélation avec *w*, seul (1)	1	0,240373
Corrélation partielle avec *x*	1	0,055534
Corrélation avec *x*, seul (2)	1	0,187888
Corrélation partielle avec *w*	1	0,108019
Corrélation multiple avec *w* et *x* (CM)	2	0,295907
Erreur	30	0,139959

Dans le cas de deux variables, la part de la variation C_{yy} de *y* qui est attribuable à *x* vaut $C_{yy}r^2$. Nous avons ici trois variables et la parte de la variation de *y* attribuable à *w* et *x* vaut $C_{yy}R^2$, où R s'appelle le "coefficient de corrélation multiple. Ici $R^2 = \frac{CM}{C_{yy}} = 0{,}678894$, le signe étant toujours positif. Il n'y a pas de raison d'attacher de l'importance aux corrélations partielles avec *x*, mais l'on voit que *w* nous fournit de l'information supplémentaire sur *y* par rapport à ce que *x* fournit à lui seul. Il n'y a pas de raison pour que le procédé se limite au cas de trois variables. Lorsqu'il y en a quatre, la corrélation partielle entre *x* et *y*, en éliminant les effets de *v* et *w*, peut s'écrire:

$$r_{xy.vw} = \frac{r_{xy.w} - r_{vx.w}r_{vy.w}}{\sqrt{(1-r_{vx.w}^2)(1-r_{vy.w}^2)}}$$

expression analogue à celle de $r_{xy.w}$ en fonction de r_{xy} et r_{wy}. Toutefois, les formules de sommes de carrés correspondant aux coefficients de corrélation partielles d'ordre plus élevé peuvent être compliquées.

Conclusion

Ce travail ne contient que peu de mathématique; on a essayé d'examiner des situations expérimentales pour les phytosociologistes. Les concepts mathématiques, très précis mais peut utiles, et les concepts biologiques, plus subtils mais parfois trop vagues pour se prêter à une vérification expérimentale, combinés, ces deux types de concepts peut avoir des qualités que ne possédent ni l'un ni l'autre séparement. Mais, il serait très dangereux d'ériger le coefficient de corrélation au rang d'expression d'une loi de la nature. Quand on les inities aux méthodes de corrélation, la plupart des gens ont l'impression d'avoir acquis un outil de valeur immense pour l'étude de cas complexes. Ce n'est pas vrai, elles ont leur utilité, mais dans un nombre limité de circonstences. Il ne faut pas confondre une association avec une relation de cause à effet. Un certain nombre de variables peuvent être en corrélation entre elles, mais l'examen des données ne peut pas révéler celle qui est plus correct.

La méthode décrite l'emploi simultané de plusieurs variables phytosociologiques (l'altitude, le nombre d'espèces dyploïdes et le nombre d'espèces polyploïdes) dans une associations végétale (*Telekio speciosae - Alnetum incanae*), en quatre composantes (Iezer-Păpuşa, Piatra Craiului, Râiosu, Răstoliţa).

Il est connu que si les valeurs élevées de l'une (l'altitude - w) ont tendance à être associées à des valeurs élevées de l'autre (les espèces polyploïdes - y) et s'il en est de même pour les valeurs faibles (les espèces dyploïdes - x), les écarts des deux variables sont en général de même signe et la plupart des produits d'écarts sont positifs. Par contre, si une augmentation de l'une (w) est associée à une diminution de l'autre (x), la plupart des produits d'écarts sont négatifs (Fig. 1):

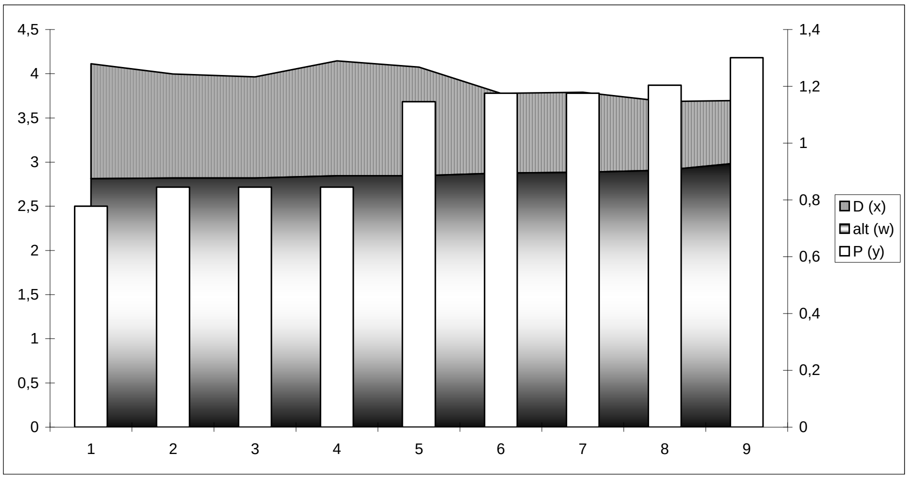

Fig. 1: La liaison entre trois variables – l'altitude, les espèces dyploïdes et les espèces polyploïdes d'associations *Telekio - Alnetum incanae*

BIBLIOGRAPHIE

1. Alexiu, V., 1998, *Vegetaţia Masivului Iezer-Păpuşa. Studiu fitocenologic*. Edit. Cultura, Piteşti.
2. Bailey, N.T.J, 1968, *Statistical methods in Biology*. The English Universities Oress L.T.D., Aylesbury-Oxford.
3. Ceapoiu, N., 1968, *Metode statistice aplicate în experienţele agricole şi biologice*. Edit. Agro-Silvică, Bucureşti.
4. Cox, G.W., 2000, *Laboratory manual of General Ecology*. San Diego State College; WM.C.Brown Company Publishers, Iowa.
5. Cristea, V., 1993 - *Fitosociologie şi Vegetaţia României*. Edit. Universităţii, Cluj-Napoca.
6. Cristea, V., Gafta, D., Pop, I., 2002, Bogăţia specifică şi proporţia endemitelor în fitocenozele lemnoase din judeţul Cluj, *Argessis. Studii şi comunicări*, **IX-X**: 25-34.
7. Davidescu, V., Davidescu D., 1999, *Compendium agrochimic*. Edit. Acad. Române, Bucureşti.
8. Dragomirescu, L., 1995, *Tehnică de calcul pentru studenţii Facultăţii de Biologie*. Edit. Universităţii Bucureşti.

9. Dragomirescu, L., 1999, *Lucrări practice de Biostatistică*. Edit. "Ars Docendi", Bucureşti.
10. Dragomirescu, L., Wanzer-Drane, J., 2001, *Biostatistică pentru începători*. Vol. **I**, Edit. "Ars Docendi", Bucureşti.
11. Drăgulescu, C., Sîrbu, I., 1997, *Practicum de fitocenologie*. Univ. "Lucian Blaga", Sibiu.
12. Lewis, T., Taylor, L.R., 1974, *Introduction to Experimental Ecologi. A student Guide to Fieldwork and Analysis*. Academic Press, London-New York.
13. Mihăilescu, S., 1999, *Flora şi vegetaţia masivului Piatra Craiului*. Edit. "Vergiliu", Bucureşti.
14. Militaru, C., Dumitrescu, A., 2001, *Statistică tehnică*. Edit. Bren, Bucureşti.
15. Oroian, S., 1998, *Flora şi vegetaţia Defileului Mureşului între Deda şi Topliţa*. Edit. SC "Apostrof", Tg. Mureş.
16. Pearce S.C., 1978, *Introduction a la statistique en biologie*. Edition S.E.I., CNRA-Versailles.
17. Săhleanu V., 1957, *Metode matematice în cercetarea medico-biologică*. Edit. Medicală, Bucureşti.
18. Snedecor G.W., 1968, *Metode statistice aplicate în cercetările de agricultură şi biologie*. Edit. Did. şi Ped., Bucureşti.
19. Stancu, D.I., 2002, *Flora şi vegetaţia Munţilor Râiosu şi Buda, Masivul Făgăraş*. Teză de doctorat, Bucureşti.
20. Stancu R., 1998, The analyses of the dependence of some physiological processes and a vegetal production as opposed to the nutrients in soil. *Revue Roumaine de Biologie*, **43**, (1): 65-72.

INTENSITATEA ŞI NATURA RELAŢIILOR ÎNTRE TREI VARIABILE ÎNTR-UN STUDIU FITOSOCIOLOGIC, FOLOSIND COEFICIENTUL DE CORELAŢIE MULTIPLĂ

(Rezumat)

Dacă vrem să ajungem, în biologie, la cunoaşterea legilor vieţii, trebuie nu numai să observăm şi să constatăm fenomenele vitale, ci şi să precizăm prin cifre relaţiile de intensitate în care se găsesc unele în raport cu celelalte. Pentru aceasta este nevoie ca datele supuse calculului să constituie rezultatele îndeajuns de analizate, astfel încât să dobândim siguranţa că am cunoscut pe deplin condiţiile fenomenelor între care vrem să stabilim o ecuaţie. În lucrarea de faţă, am încercat să examinăm situaţii experimentale privind relaţiile ce exprimă legătura între mărimile cunoscute şi cele căutate, necunoscute. Am încercat să redăm o tehnică de lucru adecvată, convinşi fiind că există şi alte metode prin care să se poată ajunge la aceleaşi rezultate. Lucrarea se adresează celor dornici să utilizeze tehnicile statistice în lucrările lor, privind măsurarea variaţiei sau analiza multivariată.

Am căutat să demonstrăm matematic relaţiile dintre trei variabile, într-un studiu fitocenologic, folosind o asociaţie vegetală, identificată în patru staţiuni distincte. Variabilele luate în calcul sunt: altitudinea, numărul speciilor diploide, numărul speciilor poliploide (puteam, de asemenea, să folosim orice alte variabile între care să existe o anumită relaţie: raportul dintre speciile unei asociaţii în funcţie de trei parametri ecologici - U,T,R - sau raportul dintre speciile unei asociaţii în corelaţie cu ponderea anumitor bioforme, sau cu ponderea anumitor geoelemente, sau raportul dintre altitudine şi constanţa sau abundenţa-dominanţa etc.). Cunoscută fiind relaţia dintre altitudine şi numărul speciilor diploide şi poliploide, am căutat să demonstrăm, prin metode matematice, veridicitatea acestei legi biologice, alegând relevee în care să nu existe specii de plante care reprezintă excepţii de la această regulă. Demonstaţia folosită încearcă să reprezinte o metodă de lucru în abordarea altor corelaţii dintre diverşi parametri.

Am folosit metoda lui Galton, cu variabile distincte, calculând deviaţia de la medie, pătratul deviaţiilor şi produsul deviaţiilor. Coeficientul de corelaţie se încadrează între -1 şi +1. Dacă valorile ridicate ale unei variabile au tendinţa de a fi asociate cu valorile ridicate ale altei variabile, amândouă spre +1 sau spre -1, pătratul celor două variabile este de acelaşi semn şi pozitiv. Din contră, dacă o creştere a valorilor unei variabile este asociată cu o diminuare a valorilor altei variabile, produsul pătratelor va fi negativ. Când între cele două variabile nu există nici o relaţie, coeficientul de corelaţie va fi zero. Un test F ne va arăta dacă este reală sau nu corelaţia. Metoda se poate aplica nu numai la trei variabile. Când sunt patru variabile, se poate calcula, pe rând, corelaţia parţială între x şi y, eliminând efectele lui v şi w. Pentru o mai mare uşurinţă în calcul, valorile numerice ale tuturor variabilelor au fost transformate în valori logaritmice, pentru că reprezentarea grafică a logaritmului altitudine în funcţie de logaritmul numărului de specii diploide şi poliploide se traduce în general printr-o relaţie liniară.

Conceptele matematice, foarte precise dar puţin folosite, şi conceptele biologice, mai subtile dar, de multe ori, prea vagi pentru a se preta la o verificare experimentală, combinate, aceste două tipuri de concepte pot avea calităţi pe care nu le au separate. Dar este greşit să absolutizăm coeficientul de corelaţie la rangul expresiei unor legi ale naturii. Metoda matematică vine să elucideze relaţiile cauzale de care depinde eventuala corelaţie dintre diferiţi parametri, să testeze măsura în care aceste relaţii apropie variabile mai mult sau mai puţin asociate.