

Seuils de signification des coefficients de corrélation génotypique, phénotypique et environnementale. Etude du cas d'un test clonal

G. NEPVEU

*Station de Recherches sur la Qualité des Bois
Centre national de Recherches forestières, I.N.R.A.,
Champenoux, 54280 Seichamps*

Résumé

Cet article est une application très simple de recherches développées par Tallis et Scheinberg. Il rappelle l'expression de la variance d'échantillonnage de l'estimation de coefficients de corrélation (génétique, génotypique, due à l'environnement, phénotypique) obtenus en analyse de variance-covariance.

Il particularise au cas d'un test clonal et montre que l'on commet une erreur parfois importante en jugeant la signification des corrélations génotypique et phénotypique uniquement à l'aide des caractéristiques du dispositif : nombre de clones, nombre d'individus par clone.

Introduction

Cet article ne constitue qu'une application très simple de recherches développées par Tallis (1959) et Scheinberg (1966).

Son but est de mettre en évidence le danger de certaines approximations assez couramment acceptées en France notamment dans des tests concernant la signification des corrélations génotypique et phénotypique.

Il est vrai que le calcul des coefficients de corrélation génotypique et phénotypique sert surtout aujourd'hui à fournir une matrice pour l'établissement d'index de sélection faisant intervenir un nombre parfois élevé de caractères ; on pourrait dans ce cadre discuter de l'importance réelle de la signification des coefficients de corrélation.

Il reste cependant des cas où l'on sélectionne sur un ou deux caractères et où l'on désire connaître les conséquences de l'opération sur d'autres paramètres ; il est alors important de tester avec une certaine rigueur les liaisons entre caractères.

Dans un premier chapitre, nous rappellerons quelques résultats permettant, dans un cas très général, d'estimer la variance d'échantillonnage de l'estimation de coefficients de corrélation obtenus en analyse de variance-covariance ; nous en donnerons un intervalle de confiance sous l'hypothèse de nullité.

En second lieu, nous particulariserons au cas d'un test clonal, donnerons les expressions des variances d'échantillonnage des coefficients de corrélations génoty-

pique, phénotypique et due au milieu, et vérifierons si les tests de signification adoptés de façon courante pour ces coefficients sont satisfaisants.

Dans un troisième chapitre, nous étudierons un exemple.

1. — Variances d'échantillonnage de l'estimation de coefficients de corrélation obtenus en analyse de variance-covariance

Nous nous placerons dans le cas le plus simple qui puisse se poser au généticien : celui d'une analyse de variance-covariance à un facteur, les effets étudiés pouvant être des effets « clone », « famille de demi-frères », ou même « famille de plein-frères » que nous désignerons sous le vocable de « groupe ».

Nous choisirons un test équilibré et appellerons n le nombre d'individus par groupe et p le nombre de groupes.

Nous attribuerons les indices 1 et 2 aux deux caractères que l'on désire corrélérer.

1.1. — Modèle

Les individus étudiés sont considérés comme un échantillon au hasard d'une population ; ils permettent d'estimer des paramètres de la population pour les caractères étudiés.

Pour chaque caractère, la valeur observable d'un individu y_{ij} est une variable aléatoire qu'on écrit :

$$y_{ij} = \mu + g_i + e_{ij}$$

avec μ : moyenne générale du caractère dans la population de référence

g_i : variable aléatoire représentant l'effet du groupe i ($i = 1, \dots, p$)

e_{ij} : variable aléatoire représentant la déviation intragroupe ($j = 1, \dots, n$).

Remarque : On insistera auprès du lecteur sur le fait que pour considérer g_i comme une variable aléatoire, il faut que les groupes forment un échantillon au hasard de la population de référence.

1.2. — Tableau d'analyse de variance-covariance

TABLEAU 1

Analysis of variance and covariance

Source de variation	ddl	Analyse de variance (caractère q , $q = 1$ et 2)		Analyse de covariance (caractères 1 et 2)	
		carré moyen	espérance des carrés moyens	produit moyen	espérance des produits moyens
Entre groupes	$p - 1$	V_{qq}	$\sigma_{e(q)}^2 + n\sigma_g^2(q)$	V_{12}	$\text{cov}_{e(1,2)} + n \text{cov}_{g(1,2)}$
Individu dans groupe ...	$p(n - 1)$	v_{qq}	$\sigma_{e(q)}^2$	v_{12}	$\text{cov}_{e(1,2)}$
Total	$pn - 1$				

avec $\sigma_{g(q)}^2$: variance entre groupes dans la population du caractère q (q = 1 et 2)
 $\sigma_{e(q)}^2$: variance intragroupe dans la population du caractère q
 $\text{cov}_{g(1,2)}$: covariance entre groupes dans la population des caractères 1 et 2.
 $\text{cov}_{e(1,2)}$: covariance intragroupe dans la population des caractères 1 et 2.

1.3. — Rappel sur les moments des carrés moyens

Les carrés et coproduits moyens V_{qq} , V_{12} , v_{qq} , v_{12} sont des variables aléatoires dont les espérances figurent dans le tableau d'analyse de variance-covariance.

Les résultats de Fisher (1928) permettent de définir les variances (covariances) de (entre) ces variables aléatoires à partir de leurs espérances.

Elles s'écrivent de façon générale :

$$\text{var}(V_{qr}) = \frac{E(V_{qq}) \times E(V_{rr}) + [E(V_{qr})]^2}{\text{ddl affectés à } V_{qr}}$$

$$\text{cov}(V_{qr}, V_{st}) = \frac{E(V_{qs}) \times E(V_{rt}) + E(V_{qt}) \times E(V_{rs})}{\text{ddl affectés à } V_{qr}}$$

On peut donner une estimation de ces variances et covariances (Tallis, 1959, op. cit.) :

$$\widehat{\text{var}}(V_{qr}) = \frac{V_{qq} \times V_{rr} + (V_{qr})^2}{\text{ddl affectés à } V_{qr}}$$

$$\widehat{\text{cov}}(V_{qr}, V_{st}) = \frac{V_{qs} \times V_{rt} + V_{qt} \times V_{rs}}{\text{ddl affectés à } V_{qr}}.$$

1.4. — Interprétation génétique des termes $\sigma_{g(q)}^2$, $\sigma_{e(q)}^2$ (respectivement $\text{cov}_{g(1,2)}$ et $\text{cov}_{e(1,2)}$) dans différents cas

Nous nous contenterons de quelques rappels sans revenir sur les conditions nécessaires à ces interprétations (panmixie...).

Nous appellerons :

- $\sigma_{A(q)}^2$ variance des effets génétiques additifs
- $\sigma_{D(q)}^2$ variance des effets génétiques non additifs
- $\sigma_G^2(q)$ variance génotypique (= $\sigma_{A(q)}^2 + \sigma_{D(q)}^2$)
- $\sigma_E^2(q)$ variance des effets dus à l'environnement
- $\sigma_P^2(q)$ variance phénotypique (= $\sigma_{A(q)}^2 + \sigma_{D(q)}^2 + \sigma_E^2(q)$).

TABLEAU 2

Interprétation génétique des termes $\sigma_{g(q)}^2$, $\sigma_{e(q)}^2$ (respectivement $\text{cov}_{g(1,2)}$ et $\text{cov}_{e(1,2)}$) dans différents cas
 Genetic interpretation of $\sigma_{g(q)}^2$, $\sigma_{e(q)}^2$ ($\text{cov}_{g(1,2)}$ and $\text{cov}_{e(1,2)}$ respectively) in some cases

	Clones	Familles de demi-frères	Familles de plein frères non apparentées
$\sigma_{g(q)}^2$	$\sigma_G^2(q)$	$\frac{1}{4} \sigma_{A(q)}^2$	$\frac{1}{2} \sigma_{A(q)}^2 + \frac{1}{4} \sigma_{D(q)}^2$
$\sigma_{e(q)}^2$	$\sigma_E^2(q)$	$\frac{3}{4} \sigma_{A(q)}^2 + \sigma_{D(q)}^2 + \sigma_E^2(q)$	$\frac{1}{2} \sigma_{A(q)}^2 + \frac{3}{4} \sigma_{D(q)}^2 + \sigma_E^2(q)$
$\sigma_{g(q)}^2 + \sigma_{e(q)}^2$	$\sigma_P^2(q)$	$\sigma_P^2(q)$	$\sigma_P^2(q)$

On voit qu'il est possible d'obtenir des estimations de $\sigma_{A(q)}^2, \sigma_{G(q)}^2, \sigma_{E(q)}^2$ ou $\sigma_{P(q)}^2$ (respectivement $\text{cov}_{A(1,2)}, \text{cov}_{G(1,2)}, \text{cov}_{E(1,2)}, \text{cov}_{P(1,2)}$) à l'aide d'une combinaison linéaire de V_{qq} et v_{qq} (respectivement V_{12} et v_{12}) de la forme $\lambda V_{qq} + \mu v_{qq}$ (respectivement $\lambda V_{12} + \mu v_{12}$).

Nous désignerons par le terme général θ_{qq} (respectivement θ_{12}) les termes $\sigma_{A(q)}^2, \sigma_{G(q)}^2, \sigma_{E(q)}^2, \sigma_{P(q)}^2$, (respectivement $\text{cov}_{A(1,2)}, \text{cov}_{G(1,2)}, \text{cov}_{E(1,2)}, \text{cov}_{P(1,2)}$).

On a $\hat{\theta}_{qq} = \lambda V_{qq} + \mu v_{qq}$ qui est une variable aléatoire d'espérance θ_{qq} et de variance :

$$\lambda^2 \text{var } V_{qq} + \mu^2 \text{var } v_{qq} + 2 \lambda \mu \text{cov } (V_{qq}, v_{qq}) .$$

Or V_q et v_q sont supposés indépendants par le modèle d'analyse de variance, d'où :

$$\text{var } \hat{\theta}_{qq} = \lambda^2 \text{var } V_{qq} + \mu^2 \text{var } v_{qq} .$$

De même, l'estimation de la covariance $\hat{\theta}_{12} = \lambda V_{12} + \mu v_{12}$ a pour espérance θ_{12} et pour variance $\lambda^2 \text{var } V_{12} + \mu^2 \text{var } v_{12}$.

On montrerait de même que

$$\begin{aligned} \text{cov } (\hat{\theta}_{qq}, \hat{\theta}_{12}) &= \lambda^2 \text{cov } (V_{qq}, V_{12}) + \mu^2 \text{cov } (v_{qq}, v_{12}) \\ \text{cov } (\hat{\theta}_{11}, \hat{\theta}_{22}) &= \lambda^2 \text{cov } (V_{11}, V_{22}) + \mu^2 \text{cov } (v_{11}, v_{22}) . \end{aligned}$$

1.5. — Valeurs prises par λ et μ dans différents cas ; rappels sur l'interprétation des θ_{qq} obtenus (respectivement θ_{12}).

TABLEAU 3

Valeurs prises par λ et μ dans différents cas ; rappels sur l'interprétation des θ_{qq} obtenus (respectivement θ_{12})
 Values of λ and μ in some cases ; interpretation of θ_{qq} calculated (θ_{12} respectively)

	Test clonal	Test de descendance	
		demi-frères	plein frères
λ	$\frac{1}{n}$	$\frac{4}{n}$	$\frac{2}{n}$
μ	$-\frac{1}{n}$	$-\frac{4}{n}$	$-\frac{2}{n}$
Interprétation	$\hat{\sigma}_G^2$	$\hat{\sigma}_A^2$	$\hat{\sigma}_A^2 + \frac{\hat{\sigma}_D^2}{2}$
λ	0	$-\frac{3}{n}$	$-\frac{1}{n}$
μ	1	$\frac{n+3}{n}$	$\frac{n+1}{n}$
Interprétation	$\hat{\sigma}_E^2$	$\hat{\sigma}_D^2 + \hat{\sigma}_E^2$	$\frac{\hat{\sigma}_D^2}{2} + \hat{\sigma}_E^2$
λ	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$
θ	$\frac{n-1}{n}$	$\frac{n-1}{n}$	$\frac{n-1}{n}$
Interprétation	$\hat{\sigma}_P^2$	$\hat{\sigma}_P^2$	$\hat{\sigma}_P^2$

1.6. — Définition des coefficients de corrélation

De même qu'on définit des variances (et covariances) génétique, génotypique, due à l'environnement, phénotypique, on peut définir entre deux caractères 1 et 2, des coefficients de corrélation génétique, génotypique, due à l'environnement, phénotypique de la forme

$$r_{\theta(1,2)} = \frac{\theta_{12}}{\sqrt{\theta_{11} \times \theta_{22}}}.$$

On va s'attacher en 1.7 à définir la variance d'échantillonnage de l'estimation de $r_{\theta(1,2)}$ que l'on peut obtenir par l'analyse de variance-covariance.

1.7. — Calcul de $\text{var}(\hat{r}_{\theta(1,2)})$

On a :

$$\hat{r}_{\theta(1,2)} = \frac{\hat{\theta}_{12}}{\sqrt{\hat{\theta}_{11} \times \hat{\theta}_{22}}}$$

$\hat{r}_{\theta(1,2)}$ est une fonction qui dépend de trois variables aléatoires $\hat{\theta}_{12}$, $\hat{\theta}_{11}$ et $\hat{\theta}_{22}$ qui ont respectivement pour espérances θ_{12} , θ_{11} et θ_{22} .

Kendall et Stuart (1963, p. 232) donnent une expression générale approchée de la variance d'une fonction g qui dépend de variables aléatoires x_1, x_2, \dots, x_k de moyennes m_1, m_2, \dots, m_k :

$$\text{var } g = \sum_{i,j=1}^k \frac{\partial g}{\partial m_i} \times \frac{\partial g}{\partial m_j} \text{cov}(x_i, x_j).$$

Appliqué au cas qui nous intéresse :

$$\begin{aligned} \text{var } \hat{r}_{\theta} &= \frac{2 \partial r_{\theta}}{\partial \theta_{12}} \times \frac{\partial r_{\theta}}{\partial \theta_{11}} \times \text{cov}(\hat{\theta}_{12}, \hat{\theta}_{11}) + \frac{2 \partial r_{\theta}}{\partial \theta_{12}} \times \frac{\partial r_{\theta}}{\partial \theta_{22}} \times \text{cov}(\hat{\theta}_{12}, \hat{\theta}_{22}) \\ &+ \frac{2 \partial r_{\theta}}{\partial \theta_{11}} \times \frac{\partial r_{\theta}}{\partial \theta_{22}} \times \text{cov}(\hat{\theta}_{11}, \hat{\theta}_{22}) + \left(\frac{\partial r_{\theta}}{\partial \theta_{12}}\right)^2 \times \text{var}(\hat{\theta}_{12}) + \left(\frac{\partial r_{\theta}}{\partial \theta_{11}}\right)^2 \times \text{var}(\hat{\theta}_{11}) \\ &+ \left(\frac{\partial r_{\theta}}{\partial \theta_{22}}\right)^2 \times \text{var}(\hat{\theta}_{22}) \\ \text{var } \hat{r}_{\theta} &= \frac{\theta_{12} \times \text{cov}(\hat{\theta}_{12}, \hat{\theta}_{11})}{\theta_{11}^2 \times \theta_{22}} - \frac{\theta_{12} \times \text{cov}(\hat{\theta}_{12}, \theta_{11})}{\theta_{22}^2 \times \theta_{11}} + \frac{\theta_{12}^2 \times \text{cov}(\hat{\theta}_{11}, \hat{\theta}_{12})}{2 \theta_{11}^2 \times \theta_{22}^2} \\ &+ \frac{\text{var}(\hat{\theta}_{12})}{\theta_{11} \times \theta_{22}} + \frac{\theta_{12}^2 \times \text{var}(\hat{\theta}_{11})}{4 \theta_{11}^3 \times \theta_{22}} + \frac{\theta_{12}^2 \times \text{var}(\hat{\theta}_{22})}{4 \theta_{22}^3 \times \theta_{11}}. \end{aligned} \quad (1)$$

On aura noté au paragraphe 14 de ce chapitre que les termes $\text{cov}(\hat{\theta}_{12}, \hat{\theta}_{11})$, $\text{cov}(\hat{\theta}_{12}, \hat{\theta}_{22})$... s'expriment en fonction des variances et covariances des termes $V_{12}, V_{qq}, v_{qq}, v_{12}$ ($q = 1, 2$).

1.8. — Cas de l'hypothèse de nullité de r_0

C'est le cas qui répond souvent à la préoccupation du généticien : une corrélation diffère-t-elle significativement de zéro ?

On est dans le cas où $\theta_{12} = 0$, c'est-à-dire qu'à partir de (1) du paragraphe 1.7, on obtient :

$$\text{var } \hat{r}_0 = \frac{\text{var } (\hat{\theta}_{12})}{\theta_{11} \times \theta_{22}} = \frac{\lambda^2 \text{var } (V_{12}) + \mu^2 \text{var } (v_{12})}{\theta_{11} \times \theta_{22}}.$$

D'après les résultats de 1.3, nous avons :

$$\text{var } (V_{12}) = \frac{E(V_{11}) \times E(V_{22}) + E(V_{12})^2}{(p-1)}$$

en se souvenant que :

$$E(V_{qq}) = \sigma_{e(q)}^2 + n \sigma_{g(q)}^2 \quad (q = 1, 2)$$

et que :

$$E(V_{12}) = \text{cov}_{e(1,2)} + n \text{cov}_g(1,2)$$

Il vient :

$$\text{var } (V_{12}) = \frac{(\sigma_{e(1)}^2 + n \sigma_{g(1)}^2) (\sigma_{e(2)}^2 + n \sigma_{g(2)}^2) + (\text{cov}_{e(1,2)} + n \text{cov}_{g(1,2)})^2}{(p-1)}.$$

De même :

$$\text{var } (v_{12}) = \frac{\sigma_{e(1)}^2 \times \sigma_{e(2)}^2 + (\text{cov}_{e(1,2)})^2}{p(n-1)}$$

d'où l'expression de $\text{var } (\hat{r}_0)$

$$\begin{aligned} \text{var } (\hat{r}_0) = \lambda^2 \frac{(\sigma_{e(1)}^2 + n \sigma_{g(1)}^2) (\sigma_{e(2)}^2 + n \sigma_{g(2)}^2) + (\text{cov}_{e(1,2)} + n \text{cov}_{g(1,2)})^2}{(p-1)} + \\ + \mu^2 \frac{\sigma_{e(1)}^2 \times \sigma_{e(2)}^2 + (\text{cov}_{e(1,2)})^2}{p(n-1)} \quad (2) \end{aligned}$$

2. — Cas d'un test clonal

Ainsi qu'il a été indiqué au paragraphe 1.4, on se souvient que :

$$\left. \begin{aligned} \sigma_{g(q)}^2 &= \sigma_{G(q)}^2 \\ \sigma_{e(q)}^2 &= \sigma_{E(q)}^2 \end{aligned} \right\} (q = 1, 2) \quad \begin{aligned} \text{cov}_{g(1,2)} &= \text{cov}_{G(1,2)} \\ \text{cov}_{e(1,2)} &= \text{cov}_{E(1,2)} \end{aligned}$$

On a :

$$\sigma_{E(q)}^2 + \sigma_{G(q)}^2 = \sigma_{P(q)}^2 : \text{variance phénotypique du caractère } q \quad (q = 1, 2)$$

$$\text{cov}_{E(1,2)} + \text{cov}_{G(1,2)} = \text{cov}_{P(1,2)} : \text{covariance phénotypique des caractères 1 et 2}$$

$$\frac{\sigma_{G(q)}^2}{\sigma_{P(q)}^2} = h^2 \text{ est l'héritabilité au sens large du caractère } q$$

$$r_{G(1,2)} = \frac{\text{COV}_{G(1,2)}}{\sqrt{\sigma_{G(1)}^2 \times \sigma_{G(2)}^2}} : \text{corrélation génotypique entre les caractères 1 et 2}$$

$$r_{E(1,2)} = \frac{\text{COV}_{E(1,2)}}{\sqrt{\sigma_{E(1)}^2 \times \sigma_{E(2)}^2}} : \text{corrélation due à l'environnement entre les caractères 1 et 2}$$

$$r_{P(1,2)} = \frac{\text{COV}_{P(1,2)}}{\sqrt{\sigma_{P(1)}^2 \times \sigma_{P(2)}^2}} : \text{corrélation phénotypique entre les caractères 1 et 2.}$$

On s'attachera dans ce cas particulier à donner des expressions de $\widehat{\text{var}} r_G$, $\widehat{\text{var}} r_E$, $\widehat{\text{var}} r_P$. On en déterminera un seuil de signification à 5 p. 100 qui permettra de tester l'hypothèse de nullité de r_0 si la valeur estimée dépasse les bornes de l'intervalle ainsi défini. On fera remarquer au lecteur que nous nous sommes placés pour ce faire dans le cas de la normalité de la représentation de \widehat{r}_0 . G. H. Brown (1969) a effectué une étude par simulation de la distribution du coefficient de corrélation génétique et a conclu à la normalité approximative de sa distribution pour des héritabilités assez fortes (de l'ordre de 0,4) avec un nombre élevé d'échantillons ($> 1\,000$), mais ses calculs s'appliquent à des estimations de r_G par régression parents-descendants. En ce qui concerne les estimations obtenues en analyse de variance-covariance, les travaux de ce genre, qui seraient très lourds, font défaut. On ne peut faire que des parallèles avec les corrélations simples, qui deviennent asymptotiquement normales lorsque le nombre d'observations est grand, et remarquer que la détermination des seuils de r_G faite plus loin, en supposant vérifiée la normalité, sera d'autant plus entachée d'erreur que le nombre de clones et les héritabilités sont plus faibles.

2.1. — Test fréquemment utilisé en France par les améliorateurs

Afin de tester les hypothèses r_G , r_E ou $r_P = 0$, on les assimile souvent à des coefficients de corrélation dont les nombres de degrés de liberté sont les suivants :

- pour r_G : nombre de degrés de liberté de l'effet clone : $p - 1$,
- pour r_E : nombre de degrés de liberté de l'effet individu dans clone : $p(n - 1)$,
- pour r_P : on prend, tantôt le nombre de degrés de liberté de l'erreur ($p(n - 1)$), tantôt le nombre total d'individus moins un, $(pn - 1)$. En vérité, comme la plupart du temps le nombre de clones est faible devant le nombre total d'individus, les tests pour r_P sont assez voisins dans les deux éventualités.

On rejette l'hypothèse d'absence de liaison entre les deux caractères (aux niveaux génotypique, environnemental ou phénotypique) à 5 (ou 1 p. 100) si les coefficients r_0 estimés dépassent les valeurs seuils indiquées dans les tables classiques en fonction du nombre de degrés de liberté de r .

On voit qu'avec cette méthode, il n'est pas tenu compte, par exemple, des héritabilités des caractères : ainsi, rien ne nous empêche d'annoncer un r_G non nul quand bien même l'effet clone sur l'un ou l'autre (ou les deux) caractère n'est pas significatif. A la limite, on peut trouver, avec $\widehat{\sigma}_{G(1)}^2 \simeq 0$, un \widehat{r}_G infini ou, à tout le moins, largement supérieur à 1 en valeur absolue : r_G sera déclaré significatif.

Remarque : Nous nous sommes placés volontairement ici dans le cas d'un test clonal. Ces tests sont également utilisés dans le cas d'études faisant intervenir des

fratries ou demi-fratries : voir Arbez et Millier (1972), Nepveu *et al.* (1978), Polge et Illy (1968), Teissier du Cros (1976).

2.2. — Calcul de $\text{var}(\hat{r}_G)$ sous l'hypothèse de nullité de r_G

Comme il a été indiqué au paragraphe 1.5, nous avons :

$$\lambda = \frac{1}{n}, \quad \mu = -\frac{1}{n}.$$

En outre θ_{11} et θ_{22} représentent ici respectivement $\sigma_{G(1)}^2$ et $\sigma_{G(2)}^2$.

En se souvenant que $\text{cov}_G = 0$ sous l'hypothèse de nullité de r_G , l'expression (2) de 1.8 s'écrit :

$$\begin{aligned} \text{var}(\hat{r}_G) &= \frac{1}{n^2} \left(\frac{(\sigma_{E(1)}^2 + n\sigma_{G(1)}^2)(\sigma_{E(2)}^2 + n\sigma_{G(2)}^2) + (\text{COV}_{E(1,2)})^2}{(p-1) \times \sigma_{G(1)}^2 \times \sigma_{G(2)}^2} + \right. \\ &\quad \left. + \frac{\sigma_{E(1)}^2 \times \sigma_{E(2)}^2 + (\text{COV}_{E(1,2)})^2}{p(n-1) \times \sigma_{G(1)}^2 \times \sigma_{G(2)}^2} \right) \\ \text{var}(\hat{r}_G) &= \frac{1}{n^2} \left[\frac{1}{(p-1)} \left(\left(\frac{1}{h_1^2} + (n-1) \right) \left(\frac{1}{h_2^2} + (n-1) \right) + r_E^2 \left(\frac{1-h_1^2}{h_1^2} \right) \left(\frac{1-h_2^2}{h_2^2} \right) \right) + \right. \\ &\quad \left. + \frac{1}{p(n-1)} \left(\left(\frac{1-h_1^2}{h_1^2} \right) \left(\frac{1-h_2^2}{h_2^2} \right) + r_E^2 \left(\frac{1-h_1^2}{h_1^2} \right) \left(\frac{1-h_2^2}{h_2^2} \right) \right) \right] \end{aligned}$$

En prenant

$$A = \sqrt{\frac{1-h_1^2}{h_1^2}} \quad \text{et} \quad B = \sqrt{\frac{1-h_2^2}{h_2^2}}$$

afin d'alléger la formule, il vient :

$$\text{var}(\hat{r}_G) = \frac{1}{n^2} \left[\frac{(A^2 + n)(B^2 + n) + r_E^2 A^2 B^2}{(p-1)} + \frac{A^2 B^2 (1 + r_E^2)}{p(n-1)} \right]$$

En se plaçant dans le cas de la normalité de la représentation de \hat{r}_G , on peut en déterminer un seuil de signification à 5 p. 100 qui est en valeur absolue :

$$\frac{1,96}{n} \sqrt{\frac{(A^2 + n)(B^2 + n) + r_E^2 A^2 B^2}{(p-1)} + \frac{A^2 B^2 (1 + r_E^2)}{p(n-1)}}. \quad (1)$$

On conclura à la non-nullité de r_G si la valeur estimée dépasse les bornes de l'intervalle ainsi défini.

On voit que le seuil de signification dépend certes du nombre de clones (p), mais également des héritabilités des deux caractères, du coefficient de corrélation due à l'environnement et du nombre d'individus par clone. On remarque que l'approximation consistant à ne tenir compte que de p pour déterminer un seuil de signification revient à faire dans l'expression (1) :

$$r_E = 0 \quad \text{et} \quad A = B = 0 \quad (\text{soit } h_1^2 = h_2^2 = 1).$$

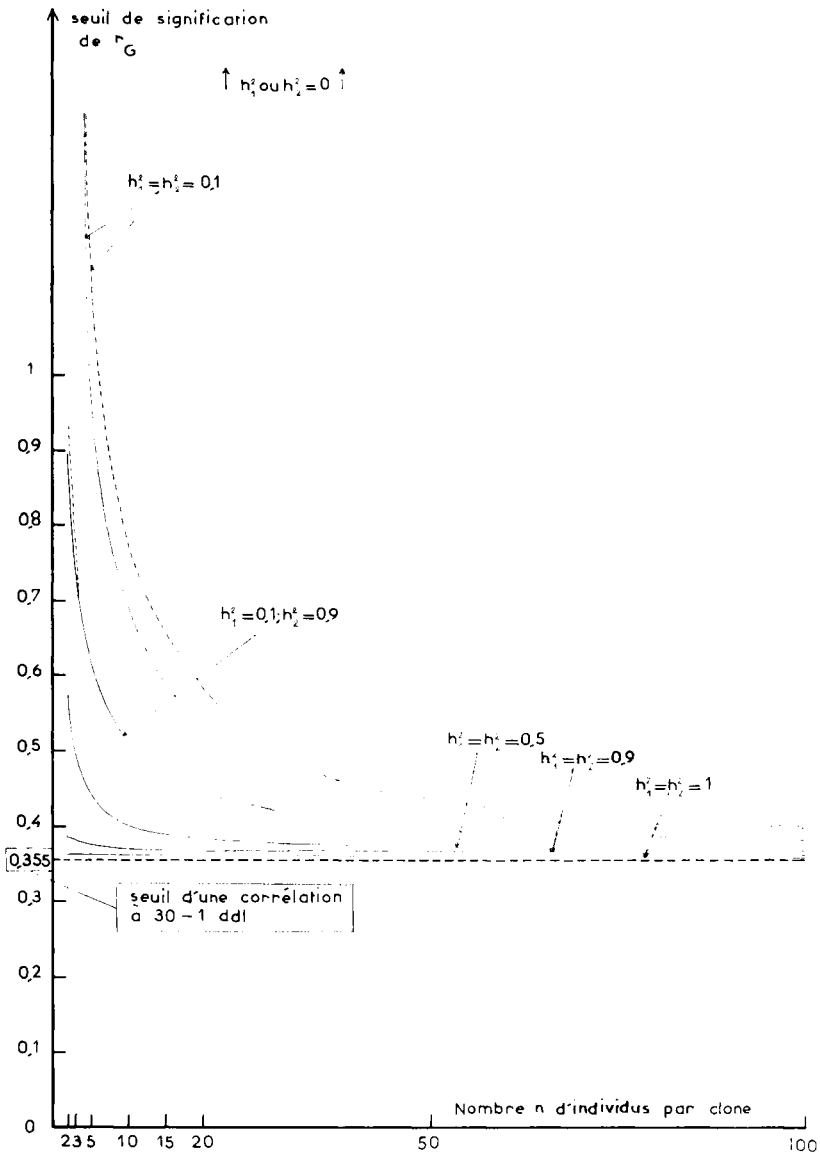


FIG. 1. — Seuil de signification contre zéro à 5 p. 100 du coefficient de corrélation génotypique r_G en fonction des héritabilités au sens large des caractères, du coefficient de corrélation due au milieu r_E et du nombre n d'individus par clone.

p , nombre de clones, égal à 30 ; — $r_E = 0$; - - - - $|r_E| = 1$.

Significance level against zero (at 5 p. 100 level) of genotypic correlation coefficient r_G in terms of broad sense heritabilities of two characters, environmental correlation coefficient r_E and size n of each clone.

p , number of clones equal to 30 ; — $r_E = 0$; - - - - $|r_E| = 1$.

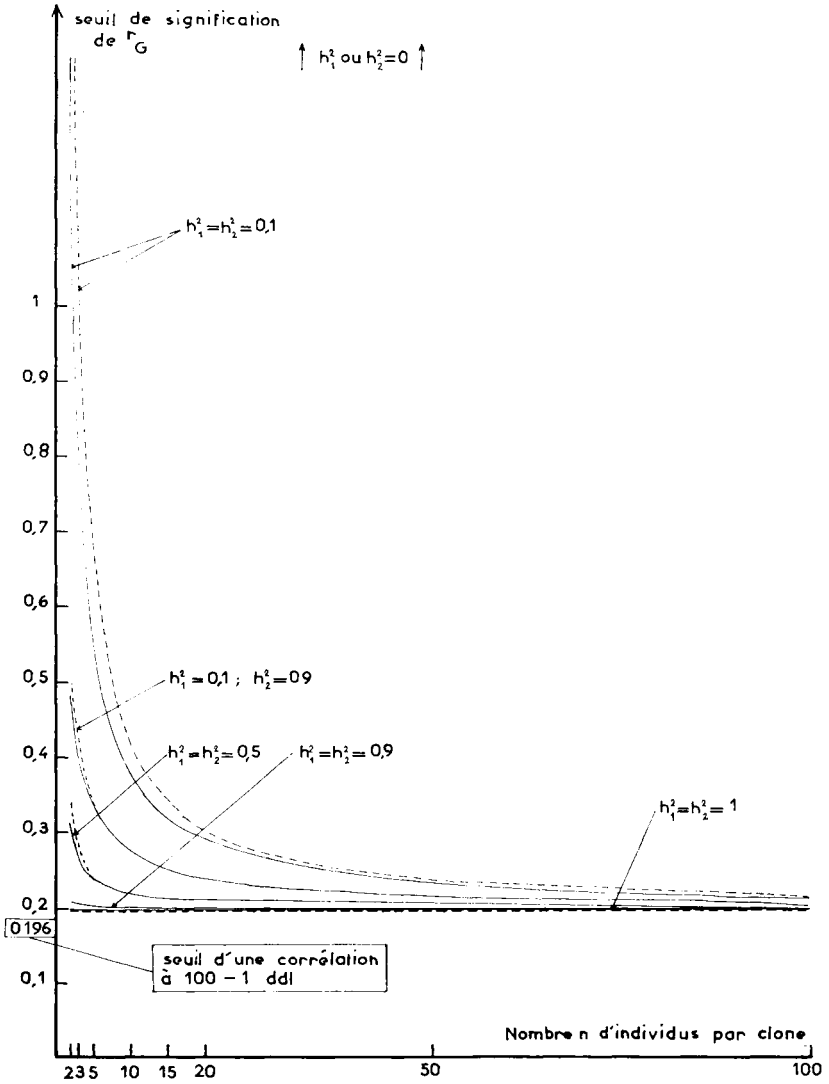


FIG. 2. — Seuil de signification contre zéro à 5 p. 100 du coefficient de corrélation génotypique r_G en fonction des héritabilités au sens large des caractères, du coefficient de corrélation due au milieu r_E et du nombre n d'individus par clone.

p , nombre de clones, égal à 100 ; — $r_E = 0$; - - - - $|r_E| = 1$.

Significance level against zero (at 5 p. 100 level) of genotypic correlation coefficient r_G in terms of broad sense heritabilities of two characters, environmental correlation coefficient r_E and size n of each clone.

p , number of clones equal to 100 ; — $r_E = 0$; - - - - $|r_E| = 1$.

En effet (1) s'écrit $\frac{1,96}{\sqrt{p-1}}$, valeur voisine du seuil de signification d'un coefficient de corrélation à $p-1$ degrés de liberté lorsque $p > 30$. En d'autres termes, l'approximation n'est licite que lorsque seuls des effets génétiques sont responsables de la variabilité des deux caractères. Il ne sert alors à rien de multiplier le nombre d'individus par clone pour estimer r_G puisque, les héritabilités étant égales à 1, tous les individus d'un même clone sont rigoureusement identiques.

Les figures 1 et 2 comparent, respectivement pour $p = 30$ et $p = 100$ les seuils utilisés en prenant pour r_G la signification contre zéro d'un coefficient de corrélation à $p-1$ degrés de liberté et les valeurs-seuils telles que nous les avons définies plus haut.

Le trait horizontal correspond au seuil de signification à 5 p. 100 avec $p-1$ degrés de liberté. Les courbes en traits pleins correspondent à $r_E = 0$, celles en traits pointillés à $|r_E| = 1$.

Nous voyons que l'approximation faite ordinairement est particulièrement dangereuse lorsque :

- les héritabilités sont faibles,
- le nombre d'individus par clone est faible.

Les valeurs des seuils sont peu sensibles aux variations de r_E .

Exemple : Prenons l'exemple de l'estimation d'une corrélation génotypique avec 100 clones (avec $r_E = 0$ pour fixer les idées). La méthode approximative va rejeter l'hypothèse de nullité de r_G à 5 p. 100, quels que soient n , h_1^2 et h_2^2 (et r_E) si r_G dépasse 0,195.

Il faut au contraire ne déclarer r_G significativement différent de zéro que s'il dépasse les valeurs suivantes :

TABLEAU 4

Seuil de signification contre zéro au niveau 5 p. 100 de r_G avec $p = 100$ et $r_E = 0$
Significance level against zero (at 5 p. 100 level) of r_G with $p = 100$ and $r_E = 0$

Valeurs des héritabilités	Nombre d'individus par clone				
	2	5	10	20	50
$h_1^2 = h_2^2 = 0,1$	1,398	0,579	0,379	0,286	0,232
$h_1^2 = 0,1$; $h_2^2 = 0,5$	0,638	0,366	0,285	0,243	0,217
$h_1^2 = h_2^2 = 0,5$	0,312	0,237	0,217	0,207	0,201
$h_1^2 = 0,1$; $h_2^2 = 0,9$	0,485	0,334	0,273	0,238	0,215
$h_1^2 = 0,5$; $h_2^2 = 0,9$	0,250	0,219	0,208	0,203	0,199
$h_1^2 = h_2^2 = 0,9$	0,209	0,202	0,199	0,198	0,198

En prenant 30 clones (et toujours $r_E = 0$), on aura les seuils suivants (alors que la méthode approximative donne une valeur limite de 0,355) :

TABLEAU 5

Seuil de signification contre zéro au niveau 5 p. 100 de r_G avec $p = 30$ et $r_E = 0$
 Significance level against zero (at 5 p. 100 level) of r_G with $p = 30$ and $r_E = 0$

Valeurs des héritabilités	Nombre d'individus par clone				
	2	5	10	20	50
$h_1^2 = h_2^2 = 0,1$	2,569	1,069	0,700	0,529	0,429
$h_1^2 = 0,1 ; h_2^2 = 0,5$	1,175	0,676	0,527	0,449	0,400
$h_1^2 = h_2^2 = 0,5$	0,574	0,438	0,401	0,382	0,371
$h_1^2 = 0,1 ; h_2^2 = 0,9$	0,895	0,616	0,505	0,440	0,396
$h_1^2 = 0,5 ; h_2^2 = 0,9$	0,462	0,404	0,384	0,374	0,368
$h_1^2 = 0,9 ; h_2^2 = 0,9$	0,385	0,372	0,368	0,367	0,365

2.3. — Cas de r_E sous l'hypothèse nulle

r_E étant estimé par la corrélation intraclone obtenue dans l'analyse de variance-covariance, on devrait trouver une variance d'échantillonnage égale à celle d'un coefficient de corrélation à $p(n - 1)$ degrés de liberté qui est le nombre de degrés de liberté de la variance intraclone.

Dans la formule (2) du paragraphe 1.8, θ_{11} et θ_{22} représentent cette fois $\sigma_{E(1)}^2$ et $\sigma_{E(2)}^2$.

D'après le paragraphe 1.5, on a $\lambda = 0$ et $\mu = 1$.

Il vient :

$$\text{var}(\hat{r}_E) = \frac{\sigma_{E(1)}^2 \times \sigma_{E(2)}^2 + \text{COV}_{E(1,2)}}{p(n - 1) \times \sigma_{E(1)}^2 \times \sigma_{E(2)}^2}$$

Sous l'hypothèse de nullité de r_E , on a $\text{COV}_{E(1,2)} = 0$, d'où

$$\text{var}(\hat{r}_E) = \frac{1}{p(n - 1)}$$

On reconnaît bien là la variance d'échantillonnage d'un coefficient de corrélation à $p(n - 1)$ degrés de liberté sous l'hypothèse de nullité.

2.4. — Cas de r_P sous l'hypothèse nulle

Cette fois nous avons

$$\lambda = \frac{1}{n}$$

$$\mu = \frac{n - 1}{n}$$

$$\theta_{11} = \sigma_{P(1)}^2$$

$$\theta_{22} = \sigma_{P(2)}^2$$

$$\text{COV}_{P(1,2)} = 0 \text{ (hypothèse de nullité de } r_P \text{).}$$

En reprenant (2) de 1.8, il vient :

$$\begin{aligned} \text{var}(\hat{r}_P) &= \frac{1}{n^2} \left[\frac{(\sigma_{E(1)}^2 + n\sigma_{G(1)}^2)(\sigma_{E(2)}^2 + n\sigma_{G(2)}^2) + (\text{COV}_{E(1,2)} + n\text{COV}_{G(1,2)})^2}{(p-1) \times \sigma_{P(1)}^2 \times \sigma_{P(2)}^2} + \right. \\ &\quad \left. + \frac{(n-1)}{p} \times \frac{\sigma_{E(1)}^2 \times \sigma_{E(2)}^2 + (\text{COV}_{E(1,2)})^2}{\sigma_{P(1)}^2 \times \sigma_{P(2)}^2} \right] \\ \text{var}(\hat{r}_P) &= \frac{1}{n^2} \left[\frac{(1 + (n-1)h_1^2)(1 + (n-1)h_2^2) + (n-1)^2 r_G^2 h_1^2 h_2^2}{(p-1)} + \right. \\ &\quad \left. + \frac{(n-1)}{p} \times ((1-h_1^2)(1-h_2^2) + r_G^2 h_1^2 h_2^2) \right]. \end{aligned}$$

En admettant la normalité de la représentation de \hat{r}_P , le seuil de signification à 5 p. 100 de r_P vaut :

$$\frac{1,96}{n} \sqrt{\frac{(1 + (n-1)h_1^2)(1 + (n-1)h_2^2) + (n-1)^2 r_G^2 h_1^2 h_2^2}{(p-1)} + \frac{n-1}{p} \times ((1-h_1^2)(1-h_2^2) + r_G^2 h_1^2 h_2^2)}.$$

On pourra rejeter l'hypothèse de nullité de r_P si la valeur estimée dépasse les bornes de l'intervalle ainsi défini.

Les figures 3 et 4 comparent pour quelques valeurs de h_1^2 , h_2^2 et r_G les seuils obtenus en prenant pour r_P la signification contre zéro d'un coefficient de corrélation à $np - 1$ degrés de liberté et les valeurs-seuils telles que nous les avons définies plus haut.

— La figure 3 correspond à une population totale de deux cents individus. Le trait horizontal représente le seuil de signification à 5 p. 100 d'un coefficient de corrélation à 199 degrés de liberté (0,139) ; les courbes en traits pleins correspondent à $r_G = 0$; celles en pointillés à $|r_G| = 1$.

— La figure 4 correspond à une population de mille individus. Le trait horizontal représente le seuil de signification à 5 p. 100 d'un coefficient de corrélation à 999 degrés de liberté (0,062). Les courbes en traits pleins et en pointillés correspondent respectivement à $r_G = 0$ et $|r_G| = 1$.

Au nombre total d'individus donné, on voit que la méthode qui consiste à prendre $np - 1$ degrés de liberté pour r_P nous conduit à déclarer r_P non nul de façon erronée ; le risque encouru est d'autant plus important que r_G et les héritabilités sont plus grandes, surtout lorsqu'on prend beaucoup d'individus par clone. Ceci est assez normal : lorsque le contrôle génétique est fort, c'est la variabilité clonale qui représente l'essentiel de la variance phénotypique. On a donc intérêt à avoir un grand nombre de clones pour estimer r_P .

3. — Un exemple

Nepveu, Keller et Tessier du Cros (1978) ont étudié vingt-huit clones de *Populus euramericana* pour quelques caractéristiques de vigueur et de qualité du bois.

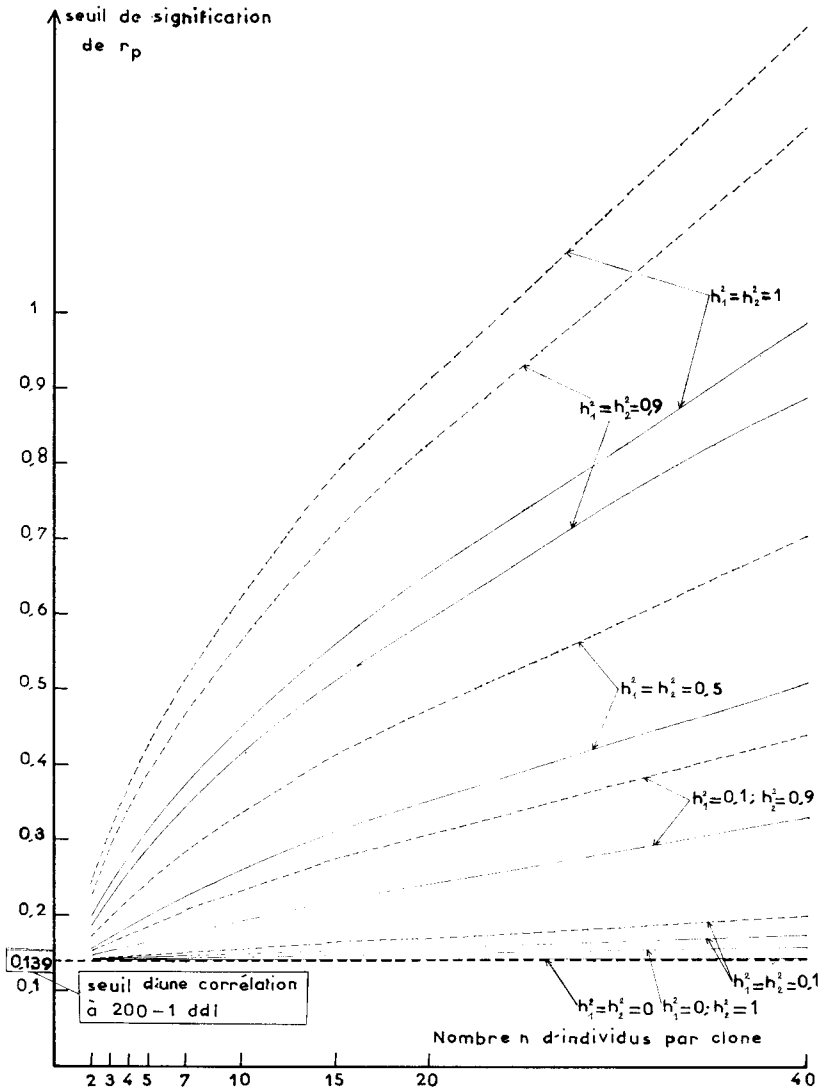


FIG. 3. — Seuil de signification contre zéro à 5 p. 100 du coefficient de corrélation phénotypique r_P en fonction des héritabilités au sens large des caractères, du coefficient de corrélation génotypique r_G et du nombre n d'individus par clone.

n_p , nombre total d'individus fixé, égal à 200 ; — $r_G = 0$; - - - - $|r_G| = 1$.

Significance level against zero (at 5 p. 100 level) of phenotypic correlation coefficient r_P in terms of broad sense heritabilities of two characters, genotypic correlation coefficient r_G and size n of each clone.

n_p , total number of individuals equal to 200 ; — $r_G = 0$; - - - - $|r_G| = 1$.

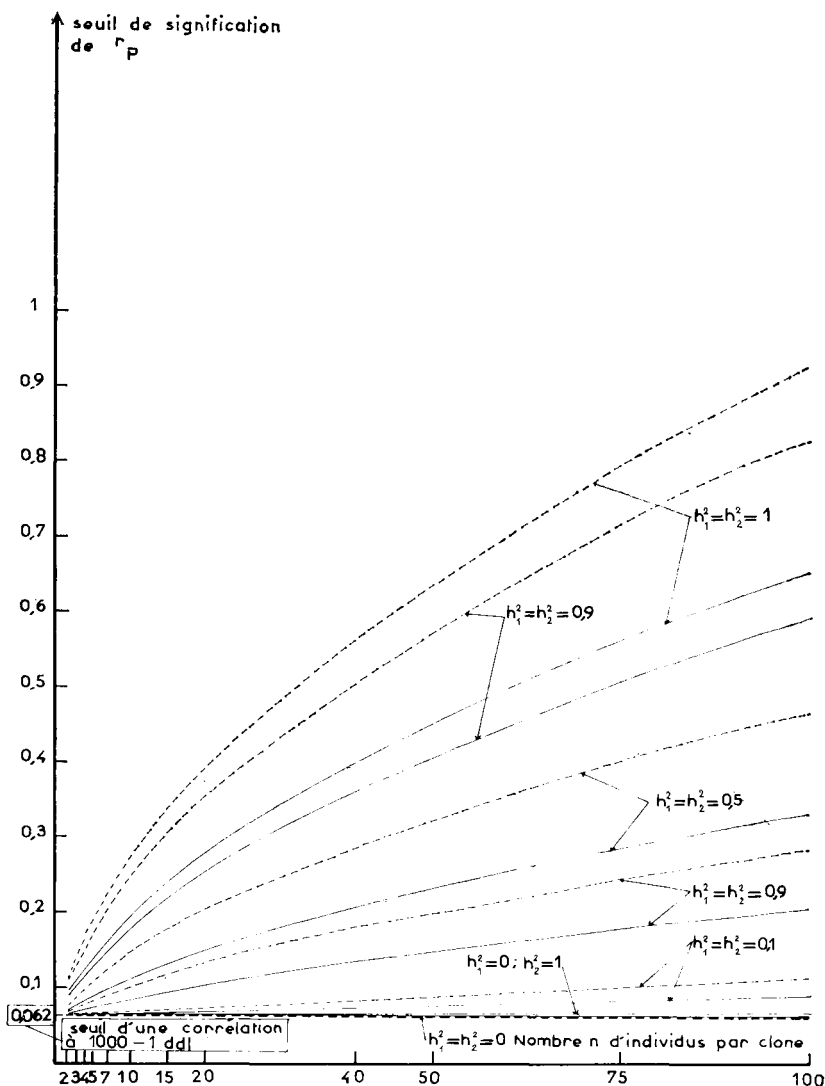


FIG. 4. — Seuil de signification contre zéro à 5 p. 100 du coefficient de corrélation phénotypique r_p en fonction des héritabilités au sens large des caractères, du coefficient de corrélation génotypique r_G et du nombre n d'individus par clone.

n_p , nombre total d'individus fixé, égal à 1 000 ; — $r_G = 0$; - - - - $|r_G| = 1$.

Significance level against zero (at 5 p. 100 level) of phenotypic correlation coefficient r_p in terms of broad sense heritabilities of two characters, genotypic correlation coefficient r_G and size n of each clone.

n_p , total number of individuals equal to 1 000 ; — $r_G = 0$; - - - - $|r_G| = 1$.

Le nombre d'individus par clone était égal à 2. Les estimations des héritabilités au sens large sont les suivantes :

— infradensité du bois, côté cœur	0,755
— infradensité du bois, côté écorce	0,711
— accroissement en hauteur, état juvénile	0,838
— accroissement en circonférence, état juvénile	0,824
— couple de torsion	0,792
— retrait tangentiel, côté écorce	0,434
— longueur de fibres, côté écorce	0,470
— accroissement en hauteur, état adulte	0,801
— accroissement en circonférence, état adulte	0,901

Par la méthode approximative, r_G , r_E et r_P sont déclarées non nuls à 5 p. 100 s'ils dépassent les valeurs-seuils 0,367 (r_G), 0,361 (r_E) et 0,261 (r_P) qui correspondent respectivement à un nombre de degrés de liberté de $p - 1 = 27$, $p(n - 1) = 28$, $pn - 1 = 55$.

En réalité, les valeurs-seuils obtenues pour r_G et r_P , en remplaçant h_1 , h_2 , r_E et r_G par leurs estimations, sont, suivant les couples de caractères considérés, les suivantes :

TABLEAU 6

Seuils de signification de r_G et r_P dans quelques cas
Significance level for r_G and r_P in some cases

	Valeur-seuil r_G	Valeur-seuil r_P
Infradensité cœur × infradensité écorce	0,452	0,355
Infradensité cœur × accroissement en hauteur, état juvénile	0,428	0,351
Infradensité cœur × accroissement en circonférence, état juvénile	0,431	0,341
Couple de torsion × infradensité écorce	0,445	0,360
Couple de torsion × retrait tangentiel écorce	0,530	0,320
Couple de torsion × longueur de fibres écorce	0,512	0,322
Couple de torsion × accroissement en hauteur, état adulte	0,428	0,342
Couple de torsion × accroissement en circonférence, état adulte	0,414	0,351
Infradensité écorce × retrait tangentiel écorce	0,551	0,315
Infradensité écorce × longueur de fibres écorce	0,533	0,325
Infradensité écorce × accroissement en hauteur, état adulte	0,426	0,334
Infradensité écorce × accroissement en circonférence, état adulte ..	0,426	0,344

Remarque : Pour calculer $\text{var } \hat{r}_G$ et $\text{var } \hat{r}_P$, on a remplacé h_1^2 , h_2^2 , r_E et r_P par leurs estimations fournies dans les analyses de variance-covariance des données observées. On doit normalement utiliser les vraies valeurs de ces coefficients que l'on ignore la plupart du temps. On sait en revanche construire un intervalle de confiance exact de h_1^2 , h_2^2 (Bogyo et Becker, 1963) et de r_E (Lamotte, 1962).

On est alors capable, la variance d'échantillonnage de r_G variant de façon simple avec h_1^2 , h_2^2 et r_E , de déterminer une valeur minima et une valeur maxima du seuil de signification de r_G en remplaçant, dans la formule donnant ($\text{var } \hat{r}_G$), les paramètres h_1^2 , h_2^2 et r_E par les bornes de leurs intervalles de confiance qui minimisent ou maximisent la valeur-seuil de r_G .

Les figures 1 et 2 montrent qu'une borne supérieure du seuil de signification de r_G est obtenue avec les bornes inférieures de h_1^2 et h_2^2 ainsi qu'avec la borne supérieure — en valeur absolue — de r_E . La borne inférieure du seuil est obtenue avec les bornes supérieures de h_1^2 et h_2^2 et avec la borne inférieure en valeur absolue de r_E .

On fera cependant deux remarques pour mettre en garde le lecteur :

— le test clonal est particulièrement simple puisque r_E , estimé par la corrélation intracclone, a un intervalle de confiance bien connu (ce n'est pas le cas de r_G et r_P). Dans le cas de fratries ou semi-fratries, on ne connaît pas d'intervalle de confiance exact de r_E . L'extrapolation à d'autres cas que le test clonal n'est donc pas correcte.

— on a défini un intervalle de confiance *très large* pour le seuil de signification de r_G . En effet, si les intervalles de h_1^2 , h_2^2 et r_E sont bien définis à 5 p. 100 (pour fixer les idées), les liaisons entre h_1^2 , h_2^2 et r_E font que la probabilité :

(bornes inférieures de h_1^2 , h_2^2 , r_E) < (h_1^2 , h_2^2 , r_E) < (bornes supérieures de h_1^2 , h_2^2 , r_E) est supérieure à 0,95.

A titre d'exemple, et pour fixer les idées, prenons le cas de la liaison infradensité du bois, côté cœur-infradensité du bois, côté écorce. On avait :

$$r_E = -0,170, \quad h_1^2 = 0,755, \quad h_2^2 = 0,711.$$

L'intervalle de confiance de r_E à 5 p. 100 est d'après Lamotte (1962, *op. cit.*) :

$$-0,445 < r_E < +0,295.$$

Les intervalles de confiance de h_1^2 et h_2^2 sont d'après Bogyo et Becker (1963, *op. cit.*) :

$$0,540 < h_1^2 < 0,877, \quad 0,469 < h_2^2 < 0,853.$$

L'expression de la limite supérieure du seuil de signification de r_G est obtenue avec $r_E = 0,445$; $h_1^2 = 0,540$; $h_2^2 = 0,469$. Elle vaut 0,602.

La limite inférieure est donnée en faisant $r_E = 0$; $h_1^2 = 0,877$ et $h_2^2 = 0,853$. Elle est égale à 0,408.

Ces deux valeurs : 0,408 et 0,602 encadrent bien la valeur 0,452 du seuil de signification précédemment trouvé en remplaçant h_1^2 , h_2^2 et r_E par leurs estimations.

On peut donc dire que le seuil de signification de r_G est compris entre 0,408 et 0,602 (avec une probabilité supérieure à 0,95). Cet intervalle n'encadre pas, on le remarquera, la valeur-seuil couramment utilisée en France qui vaut ici 0,367 (seuil de signification contre 0 à 5 p. 100 d'un coefficient de corrélation à $p-1 = 27$ degrés de liberté).

4. — Conclusion

Dans le cas d'un test clonal, la détermination des seuils de signification des coefficients de corrélation génotypique et phénotypique à l'aide des seules caractéristiques du dispositif (nombre de clones, nombre d'individus par clone) conduit toujours à rejeter abusivement l'hypothèse de nullité de r_G ou r_P .

— Pour la corrélation génotypique, l'erreur commise croît, pour un nombre de clones donné, à mesure que les héritabilités et le nombre d'individus par clone dimi-

nient. Lorsque tous ces paramètres sont fixés, l'influence de r_E , corrélation due à l'environnement, est faible.

— Pour la corrélation phénotypique, en se plaçant à nombre total d'individus fixé, l'erreur reste faible pour le cas d'un contrôle génotypique minime quel que soit le nombre d'individus par clone. En revanche, elle est importante lorsque les héritabilités sont fortes et le nombre de clones faible. Lorsque tous ces paramètres sont fixés, l'erreur croît assez sensiblement avec $|r_G|$.

Ces considérations pourraient être étendues, à l'aide des formules générales rappelées dans l'article, aux cas de tests faisant intervenir des fratries ou semi-fratries.

On rappellera au lecteur les approximations faites, notamment en ce qui concerne le caractère normal des représentations de \hat{r}_G et \hat{r}_P .

Le problème des intervalles de confiance de paramètres génétiques mériterait d'autres développements, en particulier en ce qui concerne les estimations fondées sur l'analyse de plans de croisement de modèles variés.

Reçu pour publication en avril 1979.

Summary

Significance level for the genotypic, phenotypic and environmental correlation coefficients. Clonal test assumption

This paper gives a very simple application of research developed by Tallis and Scheinberg.

It recalls the expression of the sampling variance of correlation coefficient estimation (genetic, genotypic, environmental and phenotypic) calculated from analyses of variance and covariance.

It mentions in detail the case of a clonal test and shows that a serious error may occur when the significance level of the genotypic and phenotypic correlations is appreciated by means of clonal test characteristics only : number of clones, size of each clone.

Références bibliographiques

- ARBEZ M., MILLIER C., 1972. Variabilité, héritabilité et corrélations entre caractères chez de jeunes pins laricio de Calabre (*Pinus nigra* Arn, ssp. laricio, var Calabrica). Conséquences et problèmes des indices de sélection. *IUFRO Genetics, SABRAO Joint Symposia*, Tokyo.
- BOGYO T. P., BECKER W. A., 1963. Exact confidence intervals for genetic heritability estimated from paternal half-sib correlations. *Biometrics*, **19**, 494-496.
- BROWN G. H., 1969. An empirical study of the distribution of the sample genetic correlation coefficient. *Biometrics*, **25**, n° 1.
- FISHER R. A. 1928. Moments and products moments of sampling distributions. *Proc. Lond. math. Soc.*, **30**, 199-238.
- KENDALL M. G., STUART A., 1963. The advanced theory of statistics. Vol. 1, 2nd Ed. Griffin, London, Hafner publishing Company, New York.
- LAMOTTE M., 1962. Initiation aux méthodes statistiques en biologie, Masson et Cie, 2^e édition.
- NEPVEU G., KELLER R., TEISSIER du CROS E., 1978. Sélection juvénile pour la qualité du bois chez certains peupliers noirs. *Ann. Sci. forest.*, **35** (1), 69-92.
- POLGE H., ILLY G., 1968. Héritabilité de la densité du bois et corrélations avec la croissance étudiées à l'aide de tests non destructifs sur plants de Pins maritimes de quatre ans. *Silvae Genetica*, **17**, 5-6, 173-181.
- SCHEINBERG E., 1966. The sampling variance of the correlation coefficients estimated in genetic experiments. *Biometrics*, **22**, 187-191.
- TALLIS G. M., 1959. Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance. *Aust. Jour. of Stat.* **1**, 35-43.
- TEISSIER du CROS E., 1976. Premiers enseignements d'un test de descendance de *Picea Abies* Karts des Vosges (peuplement du Kertoff). *Ann. Sci. forest.*, **32** (3), 143-155.