

Héritabilité des caractères discrets Etude bibliographique critique ⁽¹⁾

J. RAZUNGLES

*Station d'Amélioration génétique des Animaux,
Centre de Recherches de Toulouse, I.N.R.A.,
31320 Castanet-Tolosan*

Résumé

Nous avons tenté de regrouper les différentes méthodes d'estimation de l'héritabilité des caractères discrets en deux catégories :

- celles supposant l'existence d'une variable aléatoire sous-jacente,
- celles utilisant l'analyse de variance sur données brutes ou transformées.

D'autre part nous avons recherché, quand ils n'existaient pas déjà, les variances des différents estimateurs et montré leur fluctuation lorsque le plan d'accouplement variait.

Introduction

Un certain nombre de caractères dont l'on doit analyser la variation en génétique appliquée s'expriment par des variables tout ou rien ou discrètes. L'analyse de cette variation pose alors un certain nombre de problèmes, déjà envisagés par quelques auteurs, dont nous allons faire une synthèse dans le présent mémoire.

Lorsqu'il s'agit d'un caractère à variation continue et à déterminisme polygénique, l'héritabilité a été définie, selon LUSH, comme le rapport de la variance génétique additive (au sens étroit) ou de sa variance génotypique (au sens large) à la variance phénotypique. Ce paramètre revêt une importance particulière en génétique animale appliquée, et sa connaissance permet notamment de discuter la meilleure méthode de sélection à utiliser et d'estimer le progrès génétique attendu. Cependant nous avons montré (RAZUNGLES, 1974) qu'il était parfois souhaitable de discuter aussi bien l'espérance du progrès génétique que son intervalle de confiance. Inversement, à partir d'expériences de sélection, on a pu estimer des valeurs réalisées de l'héritabilité et HILL (1972) a discuté la précision de telles estimations. Plusieurs méthodes, basées sur l'expression et le calcul de la covariance entre phénotypes d'animaux apparentés, ont été proposées pour estimer les

⁽¹⁾ Cet article a fait l'objet d'une communication lors du séminaire sur « L'estimation de la valeur génétique des reproducteurs » organisé à Toulouse les 6 et 7 octobre 1976 par le Département de Génétique Animale de l'I.N.R.A. (France).

paramètres génétiques et notamment l'héritabilité de caractères à variation continue et de déterminisme polygénique. Ces méthodes font appel à l'utilisation soit de l'analyse de la variance dans un modèle à effets aléatoires (voir par exemple la revue de SEARLE, 1968), soit de la régression linéaire. Cela permet d'obtenir des covariances entre phénotypes d'individus apparentés, à partir desquelles on estime les héritabilités. Nous considérons le cas de populations panmictiques et non consanguines. Ces méthodes mathématiques d'estimation de variances et covariances font appel aux hypothèses classiques : variables continues, homocédasticité des variances, linéarité des liaisons.

Un grand nombre de caractères correspondent à des variables qui ne remplissent pas ces hypothèses : il s'agit des variables discrètes et des variables tout ou rien. Pour ces variables se posent des problèmes méthodologiques particuliers pour l'estimation des héritabilités, plus généralement des paramètres génétiques des populations et ensuite des index de sélection. On trouve dans la bibliographie peu d'études méthodologiques relatives à ces cas. L'objectif du présent travail est d'en faire une synthèse critique, en nous limitant aux estimations de l'héritabilité. Les méthodes peuvent être classées par grands groupes selon leurs bases théoriques (nous les étudierons selon ce principe). Elles s'appliquent soit aux caractères discrets, soit aux caractères tout ou rien. Dans certains cas nous avons pu trouver une méthode donnant la variance de l'héritabilité estimée. Enfin nous donnerons une illustration numérique.

Les deux grandes parties selon lesquelles les différentes méthodes d'estimation de l'héritabilité dans le cas de caractères discrets peuvent se classer, sont :

1. Celles qui supposent l'existence de variables aléatoires sous-jacentes aux variables discrètes (cf. fig. 1 et 2).
2. Celles qui utilisent l'analyse de variance soit sur les données brutes, soit sur les données transformées.

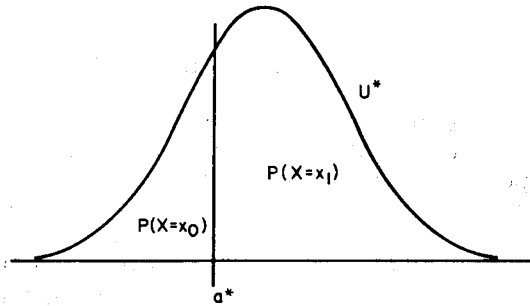


FIG. 1. — Variable normale sous-jacente à une variable tout ou rien.
 FIG. 1. — Normal variable underlying to a hit or miss variable.

I. — Méthodes supposant l'existence de variables sous-jacentes

a. — Méthode de TALLIS (1962b)

Nous étudierons tout d'abord le cas de caractères tout ou rien, c'est-à-dire de caractères dont les valeurs peuvent se regrouper en deux classes. Cela peut être le cas de la viabilité, pour laquelle les valeurs sont 0 ou 1 selon que l'individu est

vivant ou mort. Mais on peut également envisager le classement plus ou moins arbitraire de données en deux parties. Par exemple le nombre d'agneaux nés d'une brebis Romanov peut être, si l'on exclut le cas 0, égal à 1, 2, 3, 4, 5, 6. On peut diviser ces valeurs en deux groupes 1, 2 et 3, 4, 5, 6 ou 1, 2, 3 et 4, 5, 6. Le grand nombre de combinaisons montre que, dans ce dernier cas, les techniques ci-après ne seront employées que pour une estimation grossière.

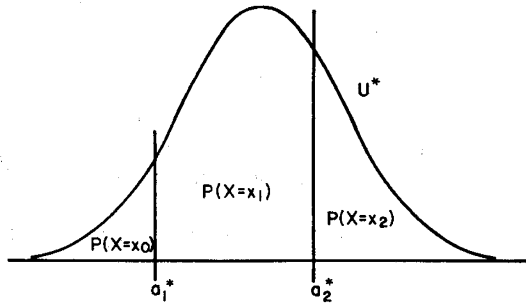


FIG. 2. — Variable normale sous-jacente à une variable discrète ayant trois valeurs.
 FIG. 2. — Normal variable underlying to a discrete variable with three values.

Le calcul de l'héritabilité h^2 est basé sur celui de la corrélation ρ parent-enfant $\left(\frac{1}{2} h^2\right)$. La même méthode pourrait être utilisée pour établir la corrélation entre deux mesures successives d'un même caractère, c'est-à-dire la répétabilité. Les données seront disposées en table de contingence (tableau 1).

TABLEAU I
 Table de contingence 2×2
 Table of 2×2 contingency.

		X	
		x_0	x_1
Y	y_0	n_{00}	n_{10}
	y_1	n_{01}	n_{11}

Sommairement le fondement théorique de la méthode est le suivant : Soit $\varphi(u, v, \rho)$ la densité conjointe de deux variables aléatoires U et V normales et $\Phi(x, y)$ leur fonction de répartition conjointe. On notera également $\varphi(u)$ et $\Phi(x)$ la densité et la fonction de répartition d'une variable normale U.

Si a et b sont les seuils des variables aléatoires U et V sous-jacentes aux variables discrètes X et Y, on aura les relations :

$$\begin{aligned} \Phi(a) &= P_0 & 1 - \Phi(a) &= P_1 \\ \Phi(b) &= P_0 & 1 - \Phi(b) &= P_1 \end{aligned}$$

et

$$\Phi(a, b) = P_{00}$$

$$\Phi(b) - \Phi(a, b) = P_{10}$$

$$\Phi(a) - \Phi(a, b) = P_{10}$$

$$1 - \Phi(a) - \Phi(b) + \Phi(a, b) = P_{11}$$

avec $P_{ij} = \Pr(X = x_i, Y = y_j)$ $i = 0, 1; j = 0, 1$

x_0 et x_1 étant les valeurs prises par X , y_0 et y_1 celles prises par Y .

Par conséquent $P_{i.} = \Pr(X = x_i)$ et $P_{.j} = \Pr(Y = y_j)$

La vraisemblance est ici déterminée par

$$L = C P_{00}^{n_{00}} P_{10}^{n_{10}} P_{01}^{n_{01}} P_{11}^{n_{11}}$$

où C est une constante ne dépendant pas des paramètres à estimer.

Il suffit maintenant de dériver $\ln L = \log L$ par rapport aux paramètres à estimer, a , b et ρ , et de procéder par itérations successives après détermination de la matrice d'information.

Les premières estimations peuvent être prises égales à :

$$\text{pour } a : a^{(1)} = \Phi^{-1} \left(\frac{n_{00} + n_{01}}{n} \right)$$

$$\text{avec } n = \sum_i \sum_j n_{ij}$$

$$\text{pour } b : b^{(1)} = \Phi^{-1} \left(\frac{n_{00} + n_{10}}{n} \right)$$

pour ρ on peut prendre $\rho^{(1)} = 0$ ou tout autre valeur.

Le propos n'est pas ici de donner une technique explicite de calcul et le lecteur pourra utilement se référer à TALLIS (1962b) ou RAZUNGLES (1974).

Cette méthode se généralise sans aucune difficulté au cas de variables discrètes prenant plus de deux valeurs telles, par exemple, la prolificité ou une notation à l'aptitude au vêlage. Dans le cas où la variable prend trois valeurs comme par exemple la prolificité d'une brebis Berrichonne du Cher, on pourra établir une table de contingence telle que le tableau 2, et écrire les relations avec les mêmes notations que pour le cas tout ou rien.

TABLEAU 2

Table de contingence 3×3

Table of 3×3 contingency

Y \ X	x_0	x_1	x_2
y_0	n_{00}	n_{10}	n_{20}
y_1	n_{01}	n_{11}	n_{21}
y_2	n_{02}	n_{12}	n_{22}

$$\Pr(X = x_0) = \Phi(a_1) = P_0.$$

$$\Pr(X = x_1) = \Phi(a_2) - \Phi(a_1) = P_1.$$

$$\Pr(X = x_2) = 1 - \Phi(a_2) = P_2.$$

$$\begin{aligned} \Pr(Y = y_0) &= \Phi(b_1) = P_{.0} \\ \Pr(Y = y_1) &= \Phi(b_2) - \Phi(b_1) = P_{.1} \\ \Pr(Y = y_2) &= 1 - \Phi(b_2) = P_{.2} \end{aligned}$$

a_1, a_2, b_1, b_2 sont les seuils tels que ceux de la figure 2. La vraisemblance est $L' = C' \prod_{i,j} P_{ij}^{n_{ij}}$; on en dérive le logarithme et on procède par itérations successives en prenant pour premières valeurs :

$$\begin{array}{l|l} a_1^{(1)} = \Phi^{-1} \left(\frac{n_{00} + n_{01} + n_{02}}{n} \right) & b_1^{(1)} = \Phi^{-1} \left(\frac{n_{00} + n_{10} + n_{20}}{n} \right) \\ a_2^{(1)} = \Phi^{-1} \left(1 - \frac{n_{20} + n_{21} + n_{22}}{n} \right) & b_2^{(1)} = \Phi^{-1} \left(1 - \frac{n_{02} + n_{12} + n_{22}}{n} \right) \end{array}$$

et $\rho'' = 0$

La technique détaillée du cas pour laquelle la variable discrète prend trois valeurs est décrite dans TALLIS (1962b) et dans RAZUNGLES (1974).

Cette méthode est à rapprocher de l'étude des tables de contingence lorsque les marges sont connues (KENDALL, 1948).

b. — Méthode de PLACKETT (1965)

La méthode de PLACKETT (1965) telle qu'elle est décrite dans TURNER (1969) apparaît comme originale et semble se distinguer de toute autre méthode. Un examen approfondi montre qu'en fait elle peut se classer parmi les méthodes qui supposent une variable continue sous-jacente. Nous ne montrerons que succinctement cette liaison en utilisant les mêmes notations que dans le paragraphe précédent.

Étant données deux variables aléatoires U et V de fonctions de répartition $\Phi(u)$ et $\Phi(v)$ normales réduites, on définit H(u, v, ψ) de telle manière que

$$\psi = \frac{H(1 - 2\Phi + H)}{(\Phi - H)^2} \tag{1}$$

H étant la distribution conjointe de U et V.

FRÉCHET (1971) a montré que H(u, v) doit satisfaire à

$$\max(\Phi(u) + \Phi(v) - 1, 0) \leq H(u, v) \leq \min(\Phi(u), \Phi(v))$$

On peut montrer que dans cet intervalle, pour ψ et Φ données, (1) a une solution et une seule.

Parmi les quelques solutions évidentes de (1), on peut remarquer que quand $u = 0$ c'est-à-dire $\Phi = \frac{1}{2}$

$$H = \frac{1}{2} \psi^{\frac{1}{2}} \left/ \left(1 + \psi^{\frac{1}{2}} \right) \right. \tag{2}$$

D'autre part si M(x, y, ρ) est la distribution conjointe de deux variables aléatoires normales de coefficient de corrélation ρ , $M(0,0; \rho) = (\arccos(-\rho))/2\pi$.

Si ce terme est égal à $H(0,0; \psi)$ on obtiendra à partir de (2) :

$$\rho = -\cos\left(\pi\psi^{\frac{1}{2}} \left/ \left(1 + \psi^{\frac{1}{2}}\right)\right.\right) \quad (3)$$

Considérons la table de contingence illustrée dans le tableau 1.

L'estimateur $\hat{\psi} = \frac{n_{00}n_{11}}{n_{01}n_{10}}$ de ψ est asymptotiquement normal de moyenne ψ et de variance estimée par :

$$\hat{\psi}^2 \left(\frac{1}{n_{00}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{11}} \right)$$

ρ sera estimée par (3) dans laquelle on aura remplacé ψ par $\hat{\psi}$.

L'utilisation de variables sous-jacentes a été envisagée par FALCONER (1965) pour mesurer la transmissibilité de certaines maladies en génétique humaine. Cette méthode a été corrigée par EDWARDS (1969) puis SMITH (1970) en a donné une nouvelle démonstration. Nous ne décrivons pas cette méthode ici car les hypothèses qui sont émises, notamment l'existence d'une population témoin dont la variance phénotypique est la même que celle de la population présentant le caractère envisagé, sont peu respectées en génétique animale.

II. — Méthodes utilisant l'analyse de variance

a) LUSH *et al.* (1948) ont utilisé simplement l'analyse de la variance, en considérant un plan d'expérience hiérarchique et une variable aléatoire qui ne prend que les valeurs 0 et 1. Si l'on étudie les taux de mortalité, tout individu qui meurt se voit attribuer la valeur 0 et tout survivant la valeur 1. L'héritabilité peut être estimée à partir de la corrélation entre demi-frères ou sœurs. Les sommes des carrés sont prises égales à des termes tels que $\frac{a(n-a)}{n}$ où a est le nombre d'individus présentant le caractère considéré, par exemple, les survivants, et n le nombre total d'individus considérés.

Dans le cas d'une hiérarchie père-mère-descendant, on établit le tableau 3; l'héritabilité est alors égale

$$\text{soit à } \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_d^2 + \sigma_e^2} \quad \text{soit à } \frac{2(\sigma_d^2 + \sigma_s^2)}{\sigma_s^2 + \sigma_d^2 + \sigma_e^2}$$

Dans le cas d'une hiérarchie père-descendants, on utilise une analyse de variance à un facteur. Le processus est strictement analogue au précédent et l'analyse s'établit comme l'indique le tableau 4. S'il s'agit de familles de demi-frères ou sœurs de père, on peut prendre comme valeur de l'héritabilité

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2}$$

et comme pour le cas précédent un estimateur de h^2 est obtenu en remplaçant les composantes de la variance par leurs estimateurs habituels tirés du tableau 4.

b) Une autre méthode d'estimation a été élaborée par ROBERTSON et LERNER (1949) puis étudiée et discutée par DEMPSTER et LERNER (1950), RENDEL (1956), TURNER et YOUNG (1969), VAN VLECK (1972), et utilisée par RAZUNGLES et al. (1975).

TABLEAU 3

Tableau d'analyse de variance dans un modèle hiérarchique non équilibré à 2 facteurs
 Analysis of variance in a hierarchical non balanced model with two factors

Source	d.d.l	SS	E (MS)
Entre familles de père	$s - 1$	SS_1	$\sigma_e^2 + k_1\sigma_d^2 + k_2\sigma_s^2$
Entre familles de mère intra fam. de père	$\sum_{i=1}^s (N_i - 1)$	SS_2	$\sigma_e^2 + k_3\sigma_d^2$
Résidus (entre pleines sœurs) . .	$N - \sum_{i=1}^s N_i$	SS_3	σ_e^2
Total	$N - 1$	SS	

$N = \sum_{ij} n_{ij}$ $N_i = \sum_j n_{ij}$ n_{ij} = nombre de descendants de la $j^{ème}$ mère accouplée au $i^{ème}$ père.

$$k_1 = \frac{\sum_i (\sum_j n_{ij}^2 / N_i) - \sum_{ij} n_{ij}^2 / N}{s - 1} \quad s = \text{nombre de pères}$$

$$k_2 = \frac{N - \sum_i N_i^2 / N}{s - 1}$$

$$k_3 = \frac{N - \sum_i (\sum_j n_{ij}^2 / N_i)}{\sum_i (N_i - 1)}$$

$$SS = \frac{a(n - a)}{n} \quad a \text{ étant le nombre total de survivants et } n \text{ le nombre total d'individus}$$

$$SS_1 = SS - SS'$$

SS' est la somme de s termes tels que $\frac{a(n - a)}{n}$, chacun calculé séparément pour chaque famille de père

$$SS_2 = SS' - SS''$$

SS'' est la somme de $\sum_{i=1}^s N_i$ termes tels que $\frac{a(n - a)}{n}$, chacun calculé séparément pour chaque famille de mère.

La définition considérée, dans le cas des caractères tout ou rien, est la suivante : l'héritabilité est le rapport de l'espérance du progrès génétique et de la différentielle de sélection phénotypique. Soient G_1, \dots, G_n les valeurs phénotypiques de n individus à une génération donnée, de moyenne $E(G)$ et de variance σ_G^2 . Le phénotype P_i du $i^{ème}$ individu est supposé être de la forme

$$P_i = G_i + E_i$$

E_i incluant tous les facteurs qui déterminent la valeur phénotypique à l'exception de G_i . P_i ne peut prendre que les valeurs 0 ou 1

$$E(E_i) = 0 \quad E(G_i.E_i) = 0$$

TABLEAU 4

Tableau d'analyse de variance dans un modèle hiérarchique non équilibré à 1 facteur
Analysis of variance in a hierarchical non balanced one factor model

Source	d.d.l	SS	E (MS)
Inter pères	$s - 1$	SS_s	$\sigma_e^2 + k\sigma_g^2$
Résidus.	$\sum_{i=1}^s (n_i - 1)$	SS_e	σ_e^2
Total.	$\sum_{i=1}^s n_i - 1$	SS	

n_i étant le nombre de descendants par mâle.

$$k = \frac{1}{s - 1} \left(\sum n_i - \frac{\sum n_i^2}{\sum n_i} \right)$$

Le génotype moyen des individus présentant le caractère considéré est :

$$\bar{\Gamma} = \frac{\sum_{i=1}^n G_i P_i}{\sum_{i=1}^n P_i} = \frac{\sum_{i=1}^n G_i (G_i + E_i)}{\sum_{i=1}^n (G_i + E_i)}$$

alors $E(\bar{\Gamma}) = \frac{E(\sum_{i=1}^n G_i^2)}{E(\sum_{i=1}^n G_i)}$

puisque asymptotiquement l'espérance d'un quotient est égale au quotient des espérances.

$$\begin{aligned} \text{Donc } E(\bar{\Gamma}) - E(G) &= \frac{\sum_{i=1}^n E(G_i^2)}{\sum_{i=1}^n E(G_i)} - E(G) \\ &= \frac{n \{ \sigma_G^2 + E^2(G) \}}{nE(G)} - E(G) \\ &= \frac{\sigma_G^2 G}{E(G)} \end{aligned}$$

Or, dans ce cas, la différentielle de sélection phénotypique est $1 - E(P) = 1 - E(G)$ donc $h^2 = \frac{\sigma_G^2}{E(G) (1 - E(G))}$

La variance génotypique peut maintenant être estimée à partir de l'analyse de variance des variables tout ou rien et cela nous conduit à l'expression bien connue :

$$h^2 = \frac{\frac{\sum a_i^2}{n_i} - \frac{(\sum a_i)^2}{\sum n_i}}{\frac{\sum a_i}{\sum n_i} \left(1 - \frac{\sum a_i}{\sum n_i}\right)} - (s - 1) \quad (4)$$

a_i étant le nombre d'individus présentant le caractère considéré, n_i étant le nombre d'individus considérés dans la $i^{\text{ème}}$ classe (par exemple le nombre de descendants du $i^{\text{ème}}$ mâle)

$$n_0 = \sum n_i - \frac{\sum n_i^2}{\sum n_i} - (s - 1)$$

s étant le nombre de classes

r est le coefficient de parenté entre deux individus de la même classe.

La formule ainsi établie pour l'héritabilité peut-être directement démontrée par l'analyse de la variance des variables tout ou rien sans aucune transformation, et s'apparente ainsi à la méthode précédente (cf. annexe).

Un certain nombre de précautions sont à prendre pour utiliser la formule (4).

α) dans une analyse entre mâles intra-années, n_i doit être en moyenne plus grand que 12,

β) dans une analyse entre femelles intra-mâles et intra-années, n_i doit être en moyenne plus grand que 5, parce que le premier terme du numérateur de (4) est un χ^2 d'hétérogénéité d'une table de contingence $2 \times s$.

Cependant ces deux dernières méthodes ont un important défaut inhérent au caractère particulier des données à traiter : la variance des observations est fortement dépendante de leur moyenne quand celle-ci n'est pas comprise entre 0,3 et 0,7. En effet si \bar{p} est le pourcentage moyen d'individus possédant le caractère considéré, la variance est égale à $\bar{p}(1 - \bar{p})$, et ce produit est alors compris entre 0,21 et 0,25.

A l'extérieur de l'intervalle [0,3, 0,7] il faut transformer les données. Notre propos n'est pas ici d'expliciter les différentes possibilités qui sont offertes, aussi ne ferons nous qu'une description sommaire.

Les deux transformations du même type à signaler tout d'abord sont $\arcsin \sqrt{\bar{p}}$ et $\sqrt{\bar{p} + \frac{3}{8}}$. D'autre part la transformation « probit » est abondamment utilisée et une étude relativement détaillée en est faite dans ROBERTSON et LERNER (1949).

Cependant les estimations faites à l'aide de données transformées n'ont pas la même signification que celles faites à partir de données brutes, et les avantages de l'une deviennent un inconvénient pour l'autre.

En effet les deux premières transformations citées ne sont pas réversibles et l'on ne peut pas passer de l'héritabilité calculée avec les données brutes à celles calculées avec des données transformées contrairement à la méthode des probits.

Mais cette méthode-ci suppose l'existence d'une variable aléatoire normale sous-jacente à la variable tout ou rien. On rejoint ainsi les hypothèses faites dans la première partie de ce mémoire.

TABLEAU 5
 Tableau d'analyse de variance corrigée à 2 facteurs
 Corrected two factor analysis of variance

Sources de variation	d.d.l	SS	MS	E (MS)
Entre classes	M — I	$\Sigma(y_{i..} / n_i) - (y^2 \dots / N)$	M_1	$\sigma_{03}^2 + \sigma_2^2 + k_1 \sigma_1^2 + k_2 \sigma_4^2$
Entre sous-classes intra classes	S — M	$\Sigma \Sigma (y_{ij}^2 / n_{ij}) - \Sigma (y_{i.}^2 / n_i)$	M_2	$\sigma_{12}^2 + \sigma_2^2 + k_3 \sigma_4^2$
Entre sous-sous-classes intra sous-classes	N — S	$\Sigma \Sigma \Sigma y_{ijk}^2 - \Sigma \Sigma (y_{ij}^2 / n_{ij})$	M_3	$\sigma_{01}^2 + \sigma_2^2$
Résidus				0,25

y_{ijk} est le pourcentage considéré (transformé par arc sin $\sqrt{\quad}$)
 n_{ij} le nombre de pourcentages dans le j^{ème} sous-classe de la i^{ème} classe
 n_i le nombre de pourcentages dans la i^{ème} classe soit Σn_{ij}
 N le nombre total de pourcentages dans l'expérience soit Σn_i
 k_1, k_2 et k_3 sont les mêmes que dans le tableau 3.

Si « h^2 » est l'héritabilité calculée sur l'échelle des probits, alors $\frac{h^2 z^2}{\bar{p}(1 - \bar{p})}$ est l'héritabilité calculée sur l'échelle des pourcentages, \bar{p} étant le pourcentage moyen d'individus ayant le caractère considéré et z , l'ordonnée de la courbe normale réduite de fonction de répartition F au point de troncature x tel que $F(x) = 1 - \bar{p}$.

Tout cela nous conduit aux méthodes employant l'analyse de variance non plus directement sur des données dichotomiques mais sur des pourcentages.

La méthode de GABRIEL (1963) qui nous a permis (RAZUNGLES, 1976) d'analyser la fertilité des brebis Romanov, Berrichonnes et croisées s'applique aux effets fixes et ne peut par conséquent être utilisée pour l'estimation de l'héritabilité. Cochran (1943) considère que la variance d'un pourcentage contient une partie binomiale de la forme $\frac{pq}{n}$ et une partie non binomiale σ^2 ne dépendant pas en général de n .

Il suggère de pondérer chaque fraction f par

$$W = \frac{1}{\frac{pq}{n} + \sigma^2} = \frac{n}{pq + n\sigma^2}$$

et étudie la pondération la plus correcte selon que la variation est binomiale ou non, afin de déterminer les tests statistiques les plus puissants, mais son attention ne se porte pas sur l'estimation de composantes de la variance.

BOGYO et BECKER (1965) ont proposé un modèle tenant compte de la variabilité binomiale de la valeur phénotypique du caractère tout ou rien.

$$y_{ijk} = p_{ijk} + b_{ijk} = p + h_i + s_{ij} + d_{ijk} + b_{ijk}$$

$i = 1, 2, \dots, M; \quad j = 1, 2, \dots, S; \quad k = 1, 2, \dots, n_{ij}$

$$\sum_j n_{ij} = n_i; \quad \sum_i \sum_j n_{ij} = N; \quad \sum_i s_i = S$$

y_{ijk} est un pourcentage, par exemple un taux de fertilité, p est la probabilité dans la population qu'une femelle soit fertile, h_i est l'effet de la $i^{\text{ème}}$ année, s_{ij} est l'effet du $j^{\text{ème}}$ mâle la $i^{\text{ème}}$ année, d_{ijk} est l'effet de la $k^{\text{ème}}$ femelle accouplée au $j^{\text{ème}}$ père la $i^{\text{ème}}$ année. b_{ijk} est un effet binomial mesuré en écart à la moyenne p_{ijk} de la sous classe.

Les h_i sont distribuées $N(0, \sigma_h^2)$, les s_{ij} $N(0, \sigma_s^2)$, les d_{ijk} $N(0, \sigma_d^2)$ et les covariances sont nulles.

Si r_{ijk} est le nombre d'individus sur lequel est calculé le pourcentage

$$E(b^2_{ijk}) = \frac{p_{ijk} q_{ijk}}{r_{ijk}}$$

avec $q_{ijk} = 1 - p_{ijk}$

d'où $\text{var } y_{ijk} = \sigma_h^2 + \sigma_s^2 + \sigma_d^2 + (p_{ijk} q_{ijk} / r_{ijk})$.

L'analyse de variance n'est valable que si $p_{ijk} q_{ijk}$ est constant.

Deux cas peuvent alors se présenter :

$$\alpha) 0,3 < p_{ijk} < 0,7$$

alors $0,21 < \hat{p}_{ijk} q_{ijk} < 0,25$ et l'on peut considérer que le numérateur de $E(b_{ijk}^2)$ est relativement constant. L'analyse de variance peut donc être faite normalement.

$$\beta) \hat{p}_{ijk} < 0,3 \text{ ou } \hat{p}_{ijk} > 0,7$$

Il faut avoir recours à une transformation. BOGYO et BECKER (1965) utilisent arc sin $\sqrt{\hat{p}}$.

Le résultat de l'analyse de variance est résumé dans le tableau 5 avec :

$$\begin{aligned} \sigma_{b1}^2 &= \frac{1}{N-S} \left(\sum_i \sum_j \sum_k \frac{c}{r_{ijk}} - \sum_i \sum_j \frac{1}{n_{ij}} \left(\sum_k \frac{c}{r_{ijk}} \right) \right) \\ \sigma_{b2}^2 &= \frac{1}{N-M} \left(\sum_i \sum_j \frac{1}{n_{ij}} \left(\sum_k \frac{c}{r_{ijk}} \right) - \sum_i \frac{1}{n_i} \left(\sum_j \sum_k \frac{c}{r_{ijk}} \right) \right) \\ \sigma_{b3}^2 &= \frac{1}{M-1} \left(\sum_i \frac{1}{n_i} \left(\sum_j \sum_k \frac{c}{r_{ijk}} \right) - \frac{1}{N} \left(\sum_i \sum_j \sum_k \frac{c}{r_{ijk}} \right) \right) \end{aligned}$$

$c = 0,25$ ou 821 selon que les angles sont exprimés en termes de radians ou de degrés.

En outre BARTLETT (1947) a suggéré de substituer $\frac{1}{2r}$ ou $\frac{1}{4r}$ à 0 et $1 - \frac{1}{2r}$ ou $1 - \frac{1}{4r}$ à 1 avant transformation surtout lorsque r_{ijk} a une valeur faible.

BOGYO et BECKER ne pensent pas que cette correction affecte les espérances des σ_{bi}^2 .

A l'aide de ces résultats on pourra donc prendre pour estimateur de l'héritabilité

$$\begin{aligned} \hat{h}^2 &= \frac{4\hat{\sigma}_s^2}{\hat{\sigma}_d^2 + \hat{\sigma}_a^2 + c} \text{ avec } \hat{\sigma}_d^2 = M_3 = \sigma_{b1}^2 \\ \hat{\sigma}_s^2 &= (M_2 - \hat{\sigma}_d^2 - \sigma_{b2}^2) / k_3 \end{aligned}$$

III. — Variance des estimateurs de l'héritabilité

Dans le cas d'un plan d'accouplement équilibré où tous les s mâles par exemple ont un même nombre n de descendants contrôlés, FISCHER (1941) a montré que :

$$\sigma_t = (1 + (n-1)t)(1-t) \sqrt{\frac{2}{n(n-1)(s-2)}}$$

où t est le coefficient de corrélation intra classe et, puisque l'héritabilité est alors égale à $\frac{t}{r}$, $\sigma_h^2 = \frac{\sigma_t}{r}$, ce qui n'est qu'un estimateur grossier.

En fait, dans le cas précis de l'estimateur établi par ROBERTSON et LERNER, nous pouvons calculer la variance de h^2 d'autre manière :

$$\text{en effet } \text{var } h^2 = \text{var } \frac{\chi^2 - (s-1)}{rn_0} = \frac{1}{r^2 h_0^2} \text{var } \chi^2$$

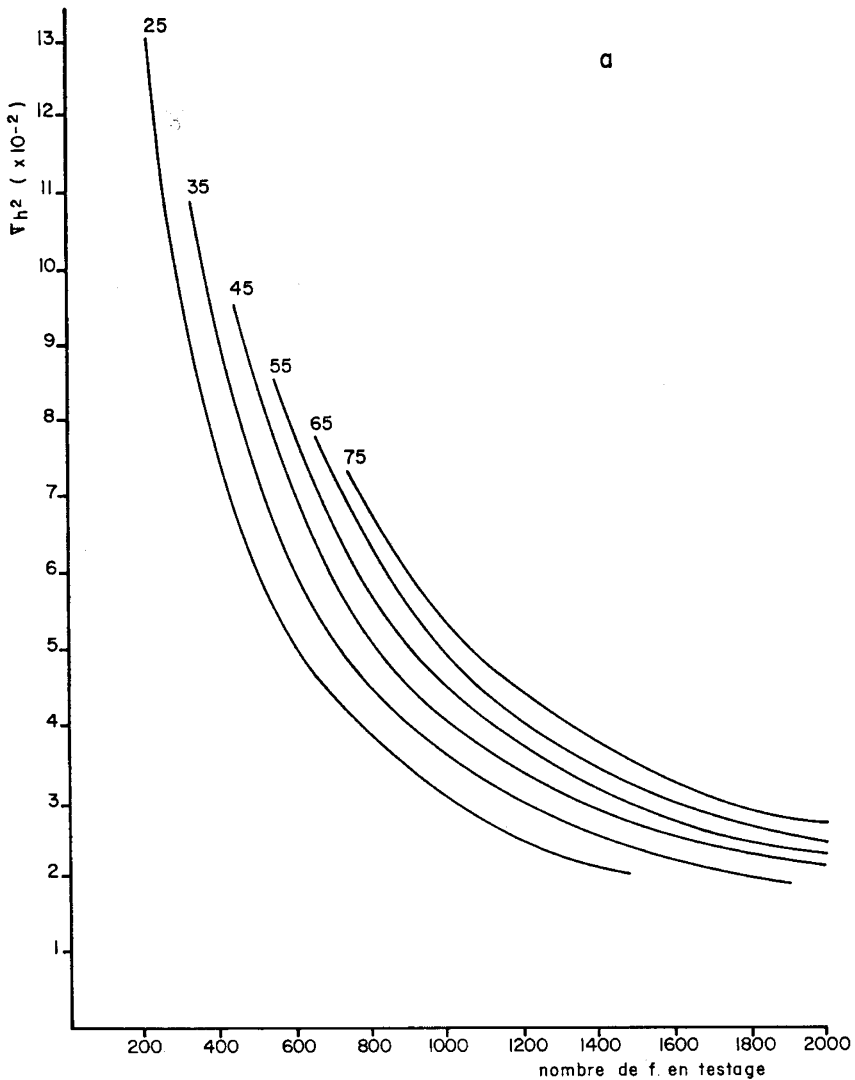
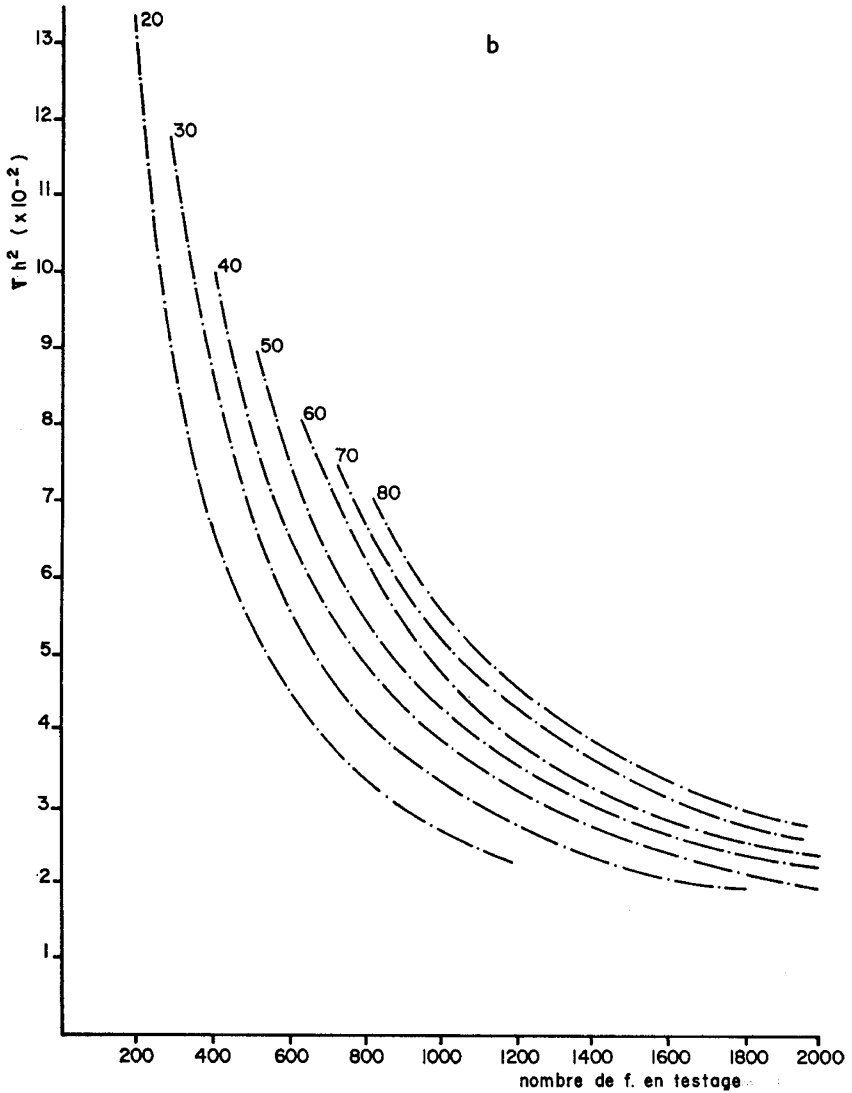


FIG. 3. — Écart type de l'héritabilité pour différents types d'accouplements :

- a et b) Chaque courbe est tracée pour un même nombre de mâles soumis au testage (nombres inscrits près des courbes).
 c) Chaque courbe est tracée pour un même nombre de femelles soumises au testage.
 d) Regroupements des 3 formules de courbes en a, b, et c.

Standard deviation of heritability for different types of matings.

- a and b) Each curve corresponds to a given number of males in the testing program (see that number near the curves).
 c) Each curve corresponds to a given number of females in the testing program.
 d) Grouping the three formulas of curves in a, b and c.



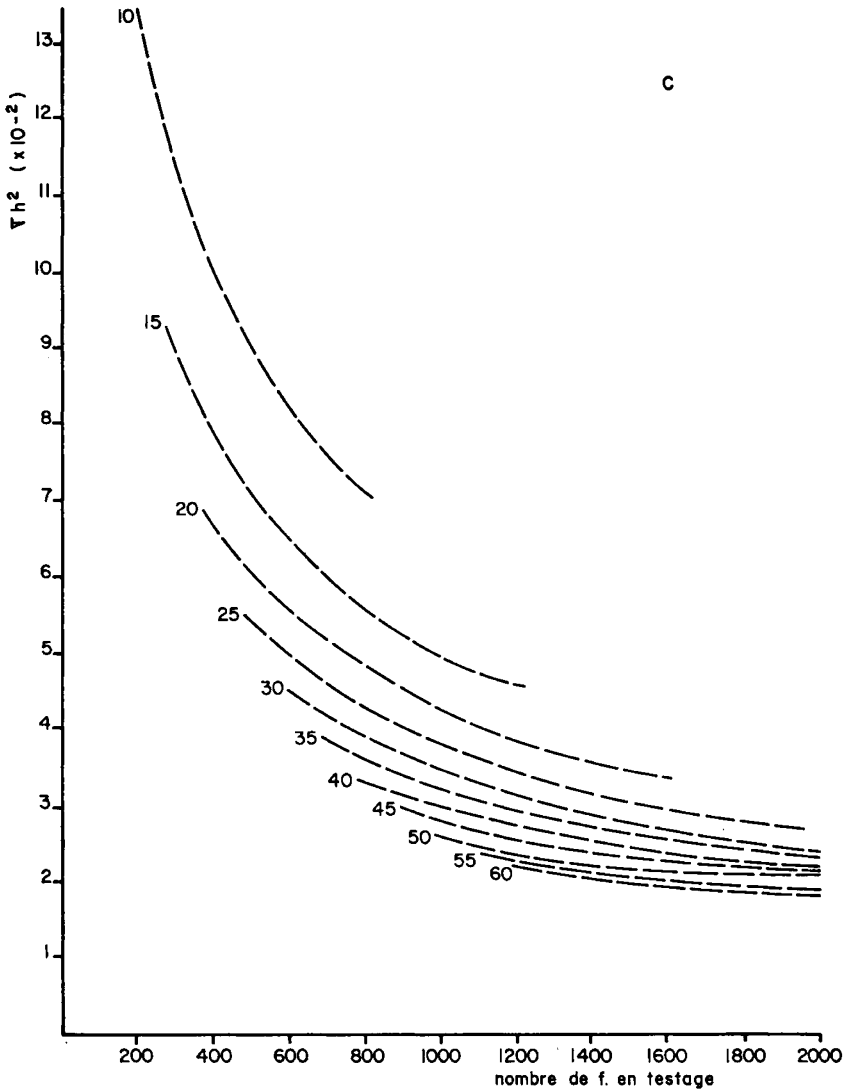
Or COCHRAN (1936) a montré que, dans un cas équilibré,

$$E(\chi^2) = s - 1 + \frac{1}{n} \left(1 - \frac{1}{s} \left(1 - \frac{1}{n} \right) \right) = v$$

$$\text{var } h^2 = \frac{\text{Var } \chi^2}{r^2 n_0^2} = \frac{2E(\chi^2)}{r^2 n_0^2} = \frac{2v}{r^2 n_0^2}$$

D'autre part dans ce même cas

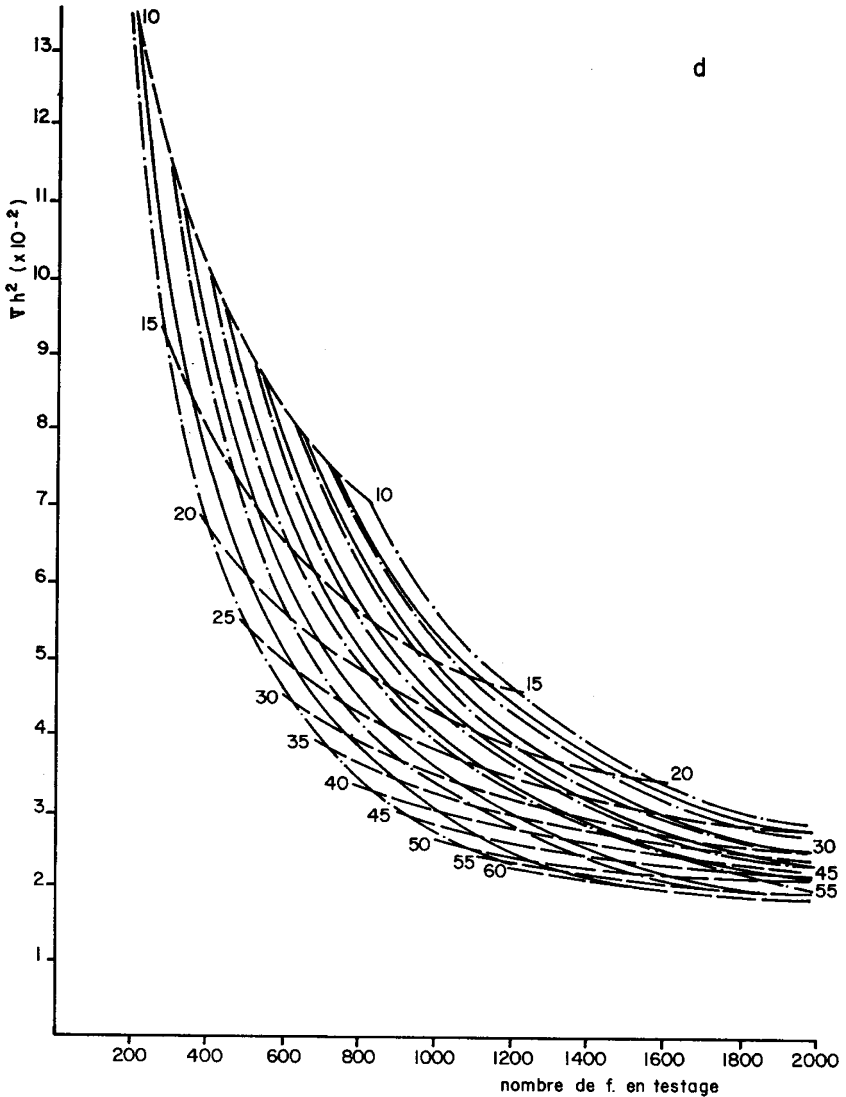
$$n_0 = \sum n_i - \frac{\sum n_i^2}{\sum n_i} = sn - n = n(s - 1)$$



donc $\text{var } h^2 = \frac{2v}{r^2n^2(s-1)^2}$

Une expression plus simple peut être donnée si l'on prend l'espérance approchée du χ^2 égale à $s - 1$, alors

$$\text{var } h^2 = \frac{2(s-1)}{r^2n^2(s-1)^2} = \frac{2}{r^2n^2(s-1)}$$



L'écart type de l'estimateur établi par PLACKETT (1965) est

$$\sigma_{h^2} = \frac{\pi \sin [\pi \psi^{1/2} / (1 + \psi^{1/2})]}{\psi^{1/2} (1 + \psi^{1/2})^2} \left(\psi \left(\frac{1}{n_{00}} + \frac{1}{n_{01}} + \frac{1}{n_{10}} + \frac{1}{n_{11}} \right)^{1/2} \right)$$

et celui de l'estimateur établi par BOGYO et BECKER (*op. cit.*), est approximativement si $\sigma_{b_1}^2 = \sigma_{b_2}^2$

$$\sigma_{h^2} = \frac{4 \sqrt{\frac{2}{k_3^2} \left(\frac{M_2^2}{S-M} + \frac{M_1^2}{N-S} \right)}}{\widehat{\sigma}_S^2 + \widehat{\sigma}_d^2 + 0,25}$$

Afin de déterminer l'importance numérique des écarts-type de l'héritabilité nous avons étudié le cas de ROBERTSON et LERNER (*op. cit.*) pour différents plans d'accouplement (fig. 3).

Les courbes tracées montrent les différents comportements de la précision selon que le nombre de femelles par mâle augmente ou lorsque le nombre de mâles augmente.

Conclusion

Les méthodes que nous avons décrites dans cette étude ont été arbitrairement divisées en deux catégories pratiquement de même importance. Il faut cependant bien remarquer, comme nous l'avons démontré en annexe, que la méthode de ROBERTSON et LERNER (*op. cit.*) n'est qu'une analyse de variance d'une variable tout-ou-rien et que la transformation probit n'est que le reflet de l'existence supposée d'une variable sous-jacente.

Cela montre bien que ces méthodes présentent en fait une réelle unité. Cependant nous tenons à insister sur le fait que la supposition d'une variable sous-jacente affecte le modèle génétique et le modèle statistique, et qu'en outre, il faut, dans toute utilisation de telles techniques, bien préciser si les hypothèses génétiques sont faites sur la variable décrivant le caractère ou sur la variable sous-jacente et cela avant ou après transformation.

Reçu pour publication en février 1977.

Summary

Heritability of discrete characters A critical bibliographic study

We have tried to regroup the different methods of estimating the heritability of discrete characters in two categories :

- these supposing the existence variable an underlying random of.
- these using analysis of variance on either rough or transformed data. On the other hand, we have search, when they did not exist already, after the variances of different estimators and shown their fluctuations when the mating scheme vary.

Références bibliographiques

- ABPLANALP M., 1961. Linear heritability estimates. *Genet. Res. Camb.*, **2**, 439-448.
- BARTLETT M. S., 1947. The use of transformation. *Biometrics*, **3**, 39-52.
- BOGYO T. P., 1964. Coefficients of variation of heritability estimates obtained from variance analysis. *Biometrics*, **20**, 122-129.
- BOGYO T. P., BECKER W. A., 1965. Estimates of heritability from transformed percentages with unequal subclass numbers. *Biometrics*, **21**, 1001-1007.
- COCHRAN W. G., 1936. The distribution for the binomial and Poisson series with small expectations. *Ann. Eugen.*, **7**, 207-217.
- COCHRAN W. G., 1943. Analysis of variance of percentages based on unequal numbers. *J.A.S.A.*, **38**, 287-301.
- DEMPSTER E. R., LERNER I. M., 1950. Heritability of threshold characters. *Genetics*, **35**, 212-236.
- EDWARDS J. H., 1969. Familial predisposition in man. *Br. Med. Bu.*, **25**, 58-63.
- FALCONER D. S., 1965. The inheritance of liability to certain diseases estimated from the incidence among relatives. *Ann. of Human Genetics*, **29**, 51-76.

- FISHER R. A., 1941. *Statistical methods for research workers*. 8^e ed. xiii + 399 pp. Edinburgh : Oliver and Boyd.
- GABRIEL, 1963. Analysis of variance of proportions with unequal frequencies. *J.A.S.A.*, **58**, 1133-1157.
- KEMPTHORNE O., TANDON O. B., 1953. The estimation of heritability by regression of offspring on parent. *Biometrics*, **9**, 890-100.
- KENDALL M. G., STUART A., 1948. *The advanced theory of statistics*, vol. 1, 4 ed., 457, 354-356.
- LUSH J. L., LAMOREUX W. F., HAZEL L. N., 1948. The heritability of resistance to death in the fowl. *Poultry science*, **27**, 375-388.
- PLACKETT R. L., 1965. A class of bivariate distributions. *J. Amer. Stat. Assoc.* **60**, 516-531.
- RAZUNGLES J. L., 1974. *Estimation de la corrélation à partir de tables de contingence*. Document à diffusion restreinte.
- RAZUNGLES J. L., 1976. Analyse de la fertilité des brebis *Romanov*, *Berrichonnes* et croisées. Mise en évidence de l'hétérosis. *Bull. tech. Dep. Génét. anim. (I.N.R.A.)*, n° 24, 79-86
- RAZUNGLES J. L., RICORDEAU G., TCHAMITCHIAN L., PRUD'HOM M., 1975. — Variations génétiques de la fertilité des brebis *Merinos d'Arles*. *Ann. génét. Sel anim.*, **7** (4), 409-419.
- RENDEL J., 1956. Heritability of multiple birth in sheep. *J. anim. Sci.*, **15**, 193-201.
- ROBERTSON A., LERNER JI. M., 1949. The heritability of all-or-none trait : *Genetics*, **34**, 395-411.
- SMITH C., 1970. Heritability of liability and concordance in monozygous twins. *Ann. Hum. Genet., Lond.*, **34**, 85-91.
- TALLIS G. M., 1962a. The use of a generalized multinomial distribution in the estimation of correlation in discrete data. *J. Royal Stat. Soc. B.*, **324**, 530-534.
- TALLIS G. M., 1962b. The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, **18**, 343-353.
- TALLIS G. M., KLOSTERMAN E. W., 1968. Efficient estimates of heritability from paternal half-sib correlations. *J. anim. Sci.*, **18**, 622-628.
- TURNER-NEWTON Helen, YOUNG SS. Y., 1969. *Quantitative genetics in sheep breeding*, 2 ed. XIII + 332 pp., Cornell University Press, New York.

Annexe

Soit une variable aléatoire y_{ij} représentant le caractère considéré du $j^{\text{ème}}$ descendant du $i^{\text{ème}}$ mâle

i varie de 1 à s .

j varie de 1 à n_i .

Ce qui signifie que l'on contrôle n_i individus de la famille du $i^{\text{ème}}$ mâle.

y_{ij} est une variable « tout ou rien » ou variable de comptage puisqu'elle ne prend que les valeurs 0 ou 1, 0 pour les individus ne présentant pas le caractère et 1 pour les autres.

Nous nous proposons donc de faire l'analyse de la variance d'une telle variable.

On pose le modèle habituel à un facteur :

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

μ étant une constante

s_i étant l'effet dû au mâle numéro i

e_{ij} étant la variable d'erreur.

L'analyse de la variance se résume dans le tableau 4.

On sait que $h^2 = \frac{\sigma_G^2}{\sigma_P^2} = \frac{\sigma_S^2/r}{\sigma_P^2}$ dans le cas présent, σ_G^2 et σ_P^2 étant respectivement les variances génotypiques et phénotypiques du caractère considéré.

Calcul des sommes de carrés

Une première remarque importante est à faire tout d'abord : $\sum_{i=1}^{n_i} y_{ij} = a_i$ est le nombre d'animaux possédant le caractère considéré.

Alors $y_{i.} = \frac{\sum_j y_{ij}}{n_i} = \frac{a_i}{n_i}$, c'est-à-dire la proportion d'animaux présentant le caractère pour la famille du $i^{\text{ème}}$ mâle, et $y_{..} = \frac{\sum a_i}{N}$.

$$\text{Soit } N = \sum_{i=1}^s n_i$$

$$\text{Alors } \sum n_i (y_{i.} - y_{..})^2 = \sum n_i y_{i.}^2 - N y_{..}^2 = \sum \frac{a_i^2}{n_i} - \frac{(\sum a_i)^2}{N}$$

$$\text{Par conséquent : } \sum \frac{a_i^2}{n_i} - \frac{(\sum a_i)^2}{N} = (\hat{\sigma}_e^2 + k \hat{\sigma}_s^2) (s - 1)$$

D'autre part, soit $p = \text{Pr}(y_{ij} = 1)$.

Alors $E(y_{ij}^2) = E(y_{ij}) = p$ et par conséquent :

$$\sigma_p^2 = \text{var } y_{ij} = E(y_{ij})^2 - (E(y_{ij}))^2 = p - p^2 = p(1 - p)$$

Or il est évident que $\hat{p} = \frac{\sum a_i}{\sum n_i}$ c'est-à-dire la fertilité moyenne de l'échantillon considéré.

$$\text{Donc on peut prendre } \sigma_p^2 = \hat{p}(1 - \hat{p}) = \frac{\sum a_i}{\sum n_i} \left(1 - \frac{\sum a_i}{\sum n_i} \right)$$

Or $\hat{\sigma}_s^2 = \hat{\sigma}_p^2 - \hat{\sigma}_e^2$ donc

$$\hat{\sigma}_s^2 = \frac{M_1 - \hat{\sigma}_e^2}{k} = \frac{M_1 - \hat{\sigma}_e^2 + \hat{\sigma}_s^2}{k} = \frac{M_1 - \hat{\sigma}_p^2}{k} + \frac{\hat{\sigma}_s^2}{k}$$

$$\text{d'où } \hat{\sigma}_s^2 = \frac{M_1 - \hat{\sigma}_p^2}{k} \left(\frac{k}{k - 1} \right) = \frac{M_1 - \hat{\sigma}_p^2}{k - 1}$$

$$\text{Alors } \hat{h}^2 = \frac{\hat{\sigma}_s^2 / r}{\hat{\sigma}_p^2} = \frac{\left(\frac{M_1}{\hat{\sigma}_p^2} - 1 \right)}{r(k - 1)} = \frac{\frac{\sum a_i^2}{n_i} - \frac{(\sum a_i)^2}{\sum n_i}}{\sum n_i \left(1 - \frac{\sum a_i}{\sum n_i} \right)} - (s - 1)$$

c.q.f.d.

NB : Le « chapeau » au-dessus d'un paramètre signifie que l'on considère une estimation de ce paramètre.