Theor Appl Genet (1985) 70:383–389

Theoretical studies on the necessary number of components in mixtures

1. Number of components and yield stability *

M. Hühn

Institut für Pflanzenbau und Pflanzenzüchtung, Universität Kiel, Olshausenstrasse 40–60, D-2300 Kiel, Federal Republic of Germany

Received November 26, 1984; Accepted December 18, 1984
Communicated by P. M. A. Tigerstedt

Summary. Theoretical studies on the optimal numbers of components in mixtures (for example multiclonal varieties or mixtures of lines) have been performed according to phenotypic yield stability (measured by the parameter 'variance'). For each component i, $i = 1, 2, \ldots, n$, a parameter u_i with $0 \leqq u_i \leqq 1$ has been introduced reflecting the different survival and yielding ability of the components. For the stochastic analysis the mean of each u_i is denoted by $\overline{u_i}$ and its variance by σ_i^2. For the character 'total yield' the phenotypic variance V can be explicitly expressed dependent on 1) the number n of components in the mixture, 2) the mean $\overline{\sigma^2}$ of the σ_i^2, 3) the variance of the σ_i^2, 4) the ratio $\overline{\sigma^2}/\lambda^2$ and 5) the ratio σ_u^2/λ^2 where λ denotes the mean of the $\overline{u_i}$ and σ_u^2 is the variance of the $\overline{u_i}$. According to the dependence of the phenotypic stability on these factors some conclusions can be easily derived from this V-formula. Furthermore, two different approaches for a calculation of necessary or optimal numbers of components using the phenotypic variance V are discussed: *A*. Determination of 'optimal' numbers in the sense that a continued increase of the number of components brings about no further significant effect according to stability. *B*. A reduction of b % of the number of components but nevertheless an unchanged stability can be realized by an increase of the mean λ of the $\overline{u_i}$ by 1% (with $\overline{\sigma^2}$ and σ_u^2 assumed to be unchanged). Numerical results on n (from *A*) and l (from *B*) are given. Computing the coefficient of variation v for the character 'total yield' and solving for the number n of components one obtains an explicit expression for n dependent on v and the factors 2.–5. mentioned above. In the special case of equal variances, $\sigma_i^2 = \sigma_0^2$ for each i, the number n depends on v, $x = (\sigma_0/\lambda)^2$ and $y = (\sigma_u/\lambda)^2$. Detailed numerical results for $n = n(v, x, y)$ are given. For $x \leqq 1$ and $y \leqq 1$ one obtains $n = 9$, 20 and 79 for $v = 0.30$, 0.20 and 0.10, respectively while for $x \leqq 1$ and arbitrary y-values the results are $n = 11$, 24 and 95.

Key words: Mixtures – Number of components – Phenotypic yield stability – Stability parameter: variance

Introduction and problem

The central problem of this paper "evaluation of optimal or necessary numbers of components in mixtures" is well-known and has been intensively discussed in many experimental investigations using very different crops (Clay and Allard 1969; Schutz and Brim 1971; Marshall and Allard 1974; Luthra and Rao 1979; Pfahler and Linskens 1979). Some theoretical studies on this topic are present in the literature (Marshall and Brown 1973; Kampmeijer and Zadoks 1977; Trenbath 1977; Østergaard 1983) but they are mainly concerned with special aspects and problems – development of epidemics, analysis of interactive effects in mixtures, etc. No general theoretical approach with regard to an evaluation of optimal numbers of components in mixtures has been worked out.

Recent activities and requests from forestry, especially from forest tree breeding, gave rise to research interest in this theoretical field: in the last years essential improvements of methods of vegetative propagation have been achieved, for

* This publication is an extended version of a lecture given at the 1984-EUCARPIA meeting (Section Biometrics in Plant Breeding) in Stuttgart-Hohenheim (Federal Republic of Germany)

example, by using cuttings and tissue culture techniques. Breeders of very different tree species succeeded in producing clones relevant for practical forestry. Referring to the negative experiences with genetically uniform varieties in agricultural crop science, most forest tree breeders propose the development of multiclonal varieties for maintaining some genetic diversity in the stands. Multiclonal varieties are mixtures of different clones artificially created with definite proportions. At present time this problem is of particular importance for the fast growing tree species like poplars and aspen.

Coming from this background the models and concepts of our following theoretical investigations have been formulated according to this field of applications. But, nevertheless, these studies and results are of an extended validity for the field of agricultural crop science and plant breeding too: 1) including multilines, that means mixtures of isolines that differ by single, major genes for reaction to a pathogen and 2) including mixtures of an arbitrary number of pure lines which are more different among each other than isolines.

To provide a simultaneous discussion of these situations – multilines, multiclonal varieties and mixtures of lines – we use the general terms 'mixture' and 'components'.

Higher yields, improved resistance properties and especially an increased phenotypic stability are the main advantages of such a variety structure. Here in this paper we want to restrict our theoretical studies to an analysis of the phenotypic yield stability measured by the parameter 'variance'. There are many other stability parameters in discussion but here we don't enter into this topic.

The aim of this paper is to give some statistical approaches and numerical results concerning the optimal or necessary number of components in mixtures with respect to phenotypic stability.

In these investigations we don't consider successive generations. Only one period from the initial composition of the mixture until the final harvest shall be analysed.

Theoretical investigations and some numerical results

For each component i, $i = 1, 2, \ldots, n$, we introduce a parameter u_i with $0 \leqq u_i \leqq 1$ reflecting the different survival of the components. u_i gives the relative probability of surviving for component i.

For such long-living organisms as forest trees, some special aspects must be considered: evaluation of juvenile-mature correlations, effects of different thinning procedures, etc. For the quantitative analysis, therefore, it is convenient to use two characters and parameters: 'mortality' (natural and artificial (thinning)) or survival on the one side and 'yielding-ability' on the other side.

$$\text{Total yield} = \sum_{i=1}^{n} \left(\frac{f_i \cdot u_i}{\bar{u}} N \right) \cdot L_i \qquad (1)$$

where:

f_i = frequency of component i in the initial composition of the mixture

u_i = survival-parameter of component i

$\bar{u}$ = mean of the u_i's: $\bar{u} = \sum_{i=1}^{n} f_i u_i$

n = number of components

L_i = mean yield of component i (per plant)

N = final number of plants.

For all the following theoretical investigations we presume a proportionality between L_i and u_i: $L_i : L_j = u_i : u_j$ for all i and j (see "discussion"). Without loss of generality the maximal value of the u_i's may be assumed to be one and the corresponding L_i will be denoted by L_{max}. Using this component as a reference component one obtains: $L_i : L_{max} = u_i : 1$ or $L_i = u_i \cdot L_{max}$ (for each i). From (1) it follows:

$$\text{total yield} = N\, L_{max} \cdot \sum_{i=1}^{n} \frac{f_i \cdot u_i^2}{\bar{u}}. \qquad (2)$$

For the special case of equal proportions of the components in the initial composition of the mixture ($f_i = 1/n$ for each i) formula (2) gives for the total yield (expressed in $N\,L_{max}$-units):

$$\text{total yield} = \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n} u_i}. \qquad (3)$$

To enable a stochastic analysis we assume that the u_i, $i = 1, 2, \ldots, n$, are independent random variables with mean $\overline{u_i}$ and variance σ_i^2. The simplifying assumption of a negligible skewness γ_i and kurtosis δ_i of the distribution of u_i (for each i) seems to be justified (see Hühn 1984 and Appendix 1).

The phenotypic yield stability of the mixture can be expressed quantitatively by the variability of the 'total yield' from (3) (for each n and variable u_i, $i = 1, 2, \ldots, n$). Here we use the variance V of the total yield as a stability parameter.

An investigation of V requires a computation of the variance of the ratio of the two random variables $\sum_{i=1}^{n} u_i^2$ and $\sum_{i=1}^{n} u_i$. After some extensive algebraic and statistical calculations one obtains the following approximation (proof, see Appendix 1):

$$V = \frac{\overline{\sigma^2}}{n} \cdot \left[1 + x^2 + y^2 + 2y + 2x \left(y + \frac{V(\sigma_i^2)}{(\overline{\sigma^2})^2} \right) \right] \qquad (4)$$

where:

$\overline{\sigma^2}$ = mean of the σ_i^2,

$V(\sigma_i^2)$ = variance of the σ_i^2,

x = $\overline{\sigma^2}/\lambda^2$,

y = σ_u^2/λ^2,

λ = mean of the $\overline{u_i}$ and

σ_u^2 = variance of the $\overline{u_i}$.

In the special case of equal variances $\sigma_i^2 = \sigma_0^2$ for each i, $i = 1, 2, \ldots, n$, the ratio $V(\sigma_i^2)/(\overline{\sigma^2})^2$ vanishes and the variance V depends on x, y, n and σ_0^2:

$$V = \frac{\sigma_0^2}{n} \cdot (1 + x^2 + y^2 + 2y + 2xy) . \qquad (5)$$

For convenience we proceed from this expression (5) in the following discussion of conclusions. In the general case of unequal variances σ_i^2, $i = 1, 2, \ldots, n$, we only have to introduce the additional term $V(\sigma_i^2)/(\overline{\sigma^2})^2$ and we must use the generalized expression (4).

The variance V and, consequently, the phenotypic stability of the mixture also depend on four factors:

1. Number n of components.
2. Variance σ_0^2 of the distribution of the u_i for each component i, $i = 1, 2, \ldots, n$.
3. Mean λ of the $\overline{u_i}$.
4. Variance σ_u^2 of the $\overline{u_i}$.

According to these factors 1 – 4, some conclusions can be easily derived from (5):

Ad 1 For fixed parameters σ_0^2, λ and σ_u^2 the stability of the mixture increases with an increasing number of components.

Ad 2 For fixed parameters n, λ and σ_u^2 the stability of the mixture decreases with an increasing variance σ_0^2.

Ad 3 For fixed parameters σ_0^2, n and σ_u^2 the stability of the mixture increases with an increasing mean λ of the $\overline{u_i}$.

Ad 4 For fixed parameters σ_0^2, n and λ the stability of the mixture decreases with an increasing variability of the $\overline{u_i}$.

All these conclusions concerning the stability of a mixture are of course, very obvious. The main importance of this theoretical approach, however, doesn't refer to these obvious statements but to the explanation of the explicit functional dependence of the phenotypic variance V on the interesting parameters n, x, y and σ_0^2 given by (5). Furthermore, this function $V = V(n, x, y, \sigma_0^2)$ can be used to derive some useful numerical results on the optimal or necessary number of components. Two such approaches shall be mentioned here without going into any theoretical and numerical details:

A. For fixed σ_0^2, λ and σ_u^2 the stability increases with increasing n. Then we can ask for the number n where the stability difference between n and n+1 components is smaller than a certain percentage g of the stability with n components.

This condition gives (using the phenotypic standard deviation):

$$n \geqq [(1-g)^{-2} - 1]^{-1} . \qquad (6)$$

For g = 0.10, 0.05 and 0.01 we obtain n = 5, 10 and 50 respectively. That means: if the improvement in stability caused by the addition of one further component shall be smaller than 5% of the previous stability this requirement may be accomplished by n = 10 components. For g = 10% this number reduces to n = 5 (Hühn 1984).

Numbers n determined in this way may be considered to be 'optimal' in the sense that a further increase of the number of components is of no significant effect according to stability.

B. A reduction of the number n of components in a mixture results in a decrease of the stability (with fixed σ_0^2, λ and σ_u^2). We may ask for the conditions that this stability decrease can be compensated and avoided by changed parameter values of σ_0^2, λ and σ_u^2. In spite of the reduction of the number of components the stability of the mixture would remain unchanged.

There are several possibilities for realization. But here we only mention the following simple approach: assuming unchanged σ_0^2 and σ_u^2 reduction of b % of the number of components may be achieved by an increase of the mean λ of the $\overline{u_i}$ by l %. Such an increase of λ can be possibly realized by a selection of suitable components.

(5) and the condition "equal stability for n and n (1 – b/100) components" give (for unchanged σ_0^2 and σ_u^2):

$$\frac{1 + (x^*)^2 + (y^*)^2 + 2x^*y^* + 2y^*}{1 + x^2 + y^2 + 2xy + 2y} = 1 - \frac{b}{100} \qquad (7)$$

where

$$x^* = \frac{x}{\left(1 + \frac{l}{100}\right)^2} \quad \text{and} \quad y^* = \frac{y}{\left(1 + \frac{l}{100}\right)^2} .$$

Using the abbreviations $K^2 = 1 + x^2 + y^2 + 2xy + 2y$, $d = (1 + l/100)^2$ and $\varepsilon = 1 - b/100$ the solution of (7) for d gives:

$$d = \frac{y \pm \sqrt{y^2 + (x+y)^2 \cdot (\varepsilon K^2 - 1)}}{\varepsilon K^2 - 1} . \qquad (8)$$

For a given percentage b and given numerical values for x and y the percentage l can be calculated from (8). A reduction of b % of the number of components and an increase of the mean λ of the $\overline{u_i}$ by l % are compensating effects.

The stability of the mixture will be unchanged. For a wide range of parameter values relevant for practical applications some numerical results shall be mentioned: For a reduction of b = 5% of the number of components an increase l of the mean λ of 2 – 4% would be required. For b = 10% we obtain 3 – 9%, for b = 20% an increase to 7 – 20%, for b = 30% we have 11 – 40% and finally for b = 50% these percentages increase up to 21 – 50%. Very low x-values together with

Table 1. Necessary numbers of components in mixtures for different x- and y-values and some numerical values for the coefficient of variation v

x \ y	v = 0.30									
	0.2	0.6	1.0	1.4	1.8	2.2	2.6	3.0	3.4	3.8
0.2	2	2	3	3	3	3	3	3	3	3
0.6	5	6	6	6	6	7	7	7	7	7
1.0	7	8	9	10	10	10	11	11	11	11
1.4	10	11	12	13	14	14	14	15	15	15
1.8	12	14	15	16	17	18	18	18	19	19
2.2	16	18	19	20	21	21	22	22	22	23
2.6	19	21	22	23	24	25	26	26	26	27
3.0	22	24	26	27	28	29	29	30	30	31
3.4	26	28	29	31	32	32	33	34	34	34
3.8	30	32	33	34	35	36	37	37	38	38

x \ y	v = 0.20									
	0.2	0.6	1.0	1.4	1.8	2.2	2.6	3.0	3.4	3.8
0.2	4	5	5	5	5	5	5	5	5	5
0.6	10	12	13	13	14	14	14	15	15	15
1.0	15	18	20	21	22	23	23	24	24	24
1.4	21	25	27	29	30	31	32	32	33	33
1.8	27	31	34	36	38	39	40	41	41	42
2.2	35	39	42	44	46	47	48	49	50	51
2.6	42	46	49	52	54	55	57	58	59	59
3.0	50	54	58	60	62	64	65	66	67	68
3.4	58	62	66	68	70	72	74	75	76	77
3.8	67	71	74	77	79	81	82	84	85	86

x \ y	v = 0.10									
	0.2	0.6	1.0	1.4	1.8	2.2	2.6	3.0	3.4	3.8
0.2	16	18	19	19	20	20	20	20	20	20
0.6	38	46	50	52	54	56	56	57	58	58
1.0	60	71	79	83	87	89	91	93	94	94
1.4	83	97	107	113	118	122	125	127	129	130
1.8	108	124	135	144	150	155	159	161	163	166
2.2	137	154	166	174	182	188	192	195	198	201
2.6	167	184	196	206	214	220	226	229	233	236
3.0	198	215	229	239	247	254	259	264	267	271
3.4	231	248	261	272	280	287	294	299	302	306
3.8	265	281	294	305	315	321	328	333	337	342

very low y-values are the only exceptions of these summarized numerical results. In these extreme situations enlarged l-percentages are necessary (for explanations and further numerical results, see Hühn 1984).

In the preceding theoretical investigations the 'variance' has been used as a stability parameter. In several aspects, however, the statistic 'coefficient of variation' shows some advantages compared to the common 'variance' – especially if variability comparisons are intended. The main argument for the preference of the 'coefficient of variation' depends on the fact that we want to characterize yield stability independent from the yield level.

Table 2. Necessary numbers of components in mixtures for different intervals of x and y and for different values for the coefficient of variation v

Intervals for x and y	No. n of components		
	v = 0.30	v = 0.20	v = 0.10
$0 < x \leqq 0.5,\ 0 < y \leqq 0.5$	5	10	38
$0 < x \leqq 0.5,\ 0.5 < y \leqq 1$	5	11	42
$0 < x \leqq 0.5,\ 1 < y \leqq 2$	6	12	46
$0 < x \leqq 0.5,\ 2 < y \leqq 3$	6	12	48
$0.5 < x \leqq 1,\ 0 < y \leqq 1$	9	20	79
$0.5 < x \leqq 1,\ 1 < y \leqq 2$	10	22	88
$0.5 < x \leqq 1,\ 2 < y \leqq 3$	11	24	93
$1 < x \leqq 2,\ 0 < y \leqq 1$	17	38	150
$1 < x \leqq 2,\ 1 < y \leqq 2$	19	42	168
$1 < x \leqq 2,\ 2 < y \leqq 3$	20	45	178
$2 < x \leqq 3,\ 0 < y \leqq 1$	26	58	229
$2 < x \leqq 3,\ 1 < y \leqq 2$	28	63	251
$2 < x \leqq 3,\ 2 < y \leqq 3$	30	66	264

After some simple calculations the expectation $\mathfrak{E}$ of the total yield (expressed in $N\,L_{max}$-units, see formula (3)) can be approximated by

$$\mathfrak{E}\,(\text{total yield}) = \frac{1 + x + y}{\sqrt{x}} \sqrt{\overline{\sigma^2}} \tag{9}$$

(proof, see Appendix 2).

We denote the coefficient of variation for the character 'total yield' by v. Combining (9) and (4) and solving for the number n of components gives:

$$n = \frac{x\left[1 + x^2 + y^2 + 2y + 2x\left(y + \frac{V(\sigma_i^2)}{(\overline{\sigma^2})^2}\right)\right]}{v^2\,(1 + x + y)^2}. \tag{10}$$

For the special case of equal variances $\sigma_i^2 = \sigma_0^2$ for each i, i = 1, 2, ..., n, formula (10) reduces to:

$$n = \frac{x\,[1 + x^2 + y^2 + 2y + 2xy]}{v^2\,(1 + x + y)^2}. \tag{11}$$

The following numerical calculations proceed from (11). The estimates for the number n of components obtained by this approach are therefore lower bounds for the 'true' number of components. With unequal variances σ_i^2, i = 1, 2, ..., n, the term $V(\sigma_i^2)/(\overline{\sigma^2})^2$ must be considered and the numerator in (11) increases leading to an increased number n. Numerical calculations have been performed for the intervals $0 < x \leqq 4$ and $0 < y \leqq 4$ and the following values for the coefficient of variation: v = 10%, 20% and 30%. These xy-intervals are extremely widened to include all possible numerical x- and y-values. Most realistic intervals for x and y are of course $0 < x \leqq 1$ and $0 < y \leqq 1$. Extensive numerical calculations and results for n dependent on x, y and v are presented in Hühn (1984).

Some of these results are given in Table 1. For $x \leqq 1$ and $y \leqq 1$ one obtains $n = 9$, 20 and 79 for $v = 0.30$, 0.20 and 0.10 respectively while for $x \leqq 1$ and $y \leqq 3$ the results are $n = 11$, 24 and 93. In Table 2 further results have been summarized for some interesting intervals for x and y. These numbers from Table 2 are the largest values of $n = n(x, y, v)$ for the given xy-regions. For practical applications the most interesting situations will be "low x, low y, larger v". For example: For $x \leqq 1$, $y \leqq 2$ and $v = 0.30$ one obtains $n = 10$.

The necessary number of components increases with 1) increasing x, 2) increasing y and 3) decreasing v.

Discussion

The preceding theoretical investigations are based on a simple model originally developed for studies with forest trees. Hence two parameters have been introduced to provide a quantitative analysis: A 'survival-parameter' describing shifts of the composition of the population and a 'yielding ability-parameter' characterizing the different yielding potential of the components independent on the selective effects above mentioned.

We don't consider successive generations. Only one period from the initial composition of the mixture until the final harvest shall be analysed.

This general 'two-parameter-approach' (survival and yielding ability) has been simplified in our theoretical studies by assuming a proportionality between these two characters: With regard to long-living organisms like forest trees this assumption seems to be justified. Many natural selection processes and the different artificial thinning procedures too are in accordance with this simplifying assumption.

For annual agricultural crops one might critically discuss it's validity and suitableness. Many examples have been cited that selection may eliminate the best-yielding types in mixtures (for example: Harlan and Martini 1938; Suneson and Wiebe 1942; Suneson 1949; Jennings and Herrera 1968). But on the other hand, many experimental investigations are known from the literature which confirm that the lower yielding components are often maintained in the mixtures with lower frequencies, and vice versa (for example: Suneson 1956; Suneson and Ramage 1962; Workman and Allard 1964; Allard and Adams 1969; Murphy, Helsel, Elliott, Thro and Frey 1982). Therefore, the model used in the previous theoretical investigations and consequently the derived numerical results on necessary numbers of components are of interest and relevance for agricultural crops too.

For the same crop a different characterization and interpretation may be necessary according to the kind of the character just studied: utilization of vegetative or generative parts of the plant.

Furthermore, in our opinion the simplifying assumptions:

a) Uncorrelated variables u_i and u_j for $i \neq j$: $\mathrm{Cov}(u_i, u_j) = 0$ for $i \neq j$.

b) No skewness and no kurtosis of the distribution of u_i (for each i): $\gamma_i = \delta_i = 0$ for each i.

c) No correlation between the variance σ_i^2 and the mean $\overline{u_i}$ and $\overline{u}_i^2$ respectively: $\mathrm{Cov}(\overline{u_i}, \sigma_i^2) = \mathrm{Cov}(\overline{u}_i^2, \sigma_i^2) = 0$.

d) Equal proportions of the components in the mixture: $f_i = 1/n$ for each i.

are no serious restriction and they don't affect the numerical results too much (Hühn 1984).

The effect of another simplyfying assumption (equal variances σ_i^2 of all the components: $\sigma_i^2 = \sigma_0^2$ for each i, $i = 1, 2, \ldots, n$) has been explained completely by the previous investigations: For equal variances one obtains expressions (5) and (11) for the variance V and the number n of components respectively. In the generalized situation with unequal variances σ_i^2 the formulae for V and n remain almost unchanged: σ_0^2 becomes $\overline{\sigma^2}$ and besides this slight modification we only have to introduce an additional term $V(\sigma_i^2)/(\overline{\sigma^2})^2$ which measures the effect of differences between the variances σ_i^2.

We suppose that the simplifying assumptions a) – d) mentioned above can be discussed in an analogous manner. We think this will result in similar generalizations by introducing some further terms with a maintenance of the basic formulae (5) and (11). But these theoretical investigations of a) – d) have not been performed until now.

Another generalization has been discussed by Hühn (1984): The model can be extended in such a way that it includes the analysis of competitive effects between the different components in the mixture. This generalized model, however, has been investigated in Hühn (1984) only for a deterministic approach. The numerical magnitude of the number of components was not altered too much by including the competitive effects (Hühn 1984).

References

Allard RW, Adams J (1969) Population studies in predominantly selfpollinating species. 13. Intergenotypic competition and population structure in barley and wheat. Am Nat 103:621–645

Clay RE, Allard RW (1969) A comparison of the performance of homogeneous and heterogeneous barley populations. Crop Sci 9:407–412

Harlan HV, Martini ML (1938) The effect of natural selection in a mixture of barley varieties. J Agric Res 57:189–199

Hühn M (1984) Theoretische Untersuchungen zur Problematik der Mindestklonzahl in Mehrklonsorten. Teil IV. Untersuchungen zur Beziehung zwischen Leistungsstabilität und Klonanzahl. Hessisches Ministerium für Landesentwicklung, Umwelt, Landwirtschaft und Forsten, Wiesbaden, S 1–103

Jennings PR, Herrera RM (1968) Studies on competition in rice. 2. Competition in segregating populations. Evolution 22:332–336

Kampmeijer P, Zadoks JC (1977) EPIMUL, a simulator of foci and epidemics in mixtures of resistant and susceptible plants, mosaics and multilines. Simulation Monographs, PUDOC, Wageningen

Luthra JK, Rao MV (1979) Multiline cultivars – how their resistance influence leaf rust diseases in wheat. Euphytica 28:137–144

Marshall DR, Brown AHD (1973) Stability of performance of mixtures and multilines. Euphytica 22:405–412

Marshall DR, Allard RW (1974) Performance and stability of mixtures of grain sorghum. Theor Appl Genet 44:145–152

Morgenstern D (1968) Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik. 2. Aufl, Die Grundlehren der mathematischen Wissenschaften, Band 124. Springer, Berlin Heidelberg New York

Murphy JP, Helsel DB, Elliott A, Thro AM, Frey KJ (1982) Compositional stability of an oat multiline. Euphytica 31:33–40

Østergaard H (1983) Predicting development of epidemics on cultivar mixtures. Phytopathology 73:166–172

Pfahler PL, Linskens HF (1979) Yield stability and population diversity in oats (Avena sp.). Theor Appl Genet 54:1–5

Rasch D (1978) Einführung in die mathematische Statistik, Band 1. Wahrscheinlichkeitsrechnung und Grundlagen der mathematischen Statistik. 2. Aufl, VEB Deutscher Verlag der Wissenschaften, Berlin

Schutz WM, Brim CA (1971) Intergenotypic competition in soybeans. 3. An evaluation of stability in multiline mixtures. Crop Sci 11:684–689

Suneson CA, Wiebe GA (1942) Survival of barley and wheat varieties in mixtures. J Am Soc Agron 34:1052–1056

Suneson CA (1949) Survival of four barley varieties in a mixture. Agron J 41:459–461

Suneson CA (1956) An evolutionary plant breeding method. Agron J 48:188–191

Suneson CA, Ramage RT (1962) Competition between near-isogenic genotypes. Crop Sci 2:249–250

Trenbath BR (1977) Interactions among diverse hosts and diverse parasites. Ann NY Acad Sci 287:124–150

Workman PL, Allard RW (1964) Population studies in predominantly self-pollinated species. 5. Analysis of differential and random viabilities in mixtures of competing pure lines. Heredity 19:181–189

Appendix 1. Proof of expressions (4) and (5) for the phenotypic variance

For the variance of the ratio of the two random variables Y and Z the following approximation holds (Rasch 1978):

$$V\left(\frac{Y}{Z}\right) \cong \frac{1}{\bar{z}^2}\left[V(Y) + \left(\frac{\bar{y}}{\bar{z}}\right)^2 \cdot V(Z) - 2\frac{\bar{y}}{\bar{z}}\operatorname{Cov}(Y,Z)\right] \qquad (1)$$

where: $\bar{y}$, $\bar{z}$ = mean of Y and Z; V(Y), V(Z) = variances of Y and Z and Cov (Y, Z) = covariance between Y and Z. Additionally, we use the symbol $\mathfrak{E}(Y)$ for the expectation of a random variable Y.

For each component i, i = 1, 2, ..., n, we have:

$$\mathfrak{E}(u_i) = \overline{u_i} \quad \text{and} \quad V(u_i) = \sigma_i^2 . \qquad (2)$$

Skewness and kurtosis of the distribution of u_i will be denoted by γ_i and δ_i respectively. With this notations we obtain:

$$\mathfrak{E}\left(\sum_{i=1}^n u_i\right) = \sum_{i=1}^n \overline{u_i}$$

$$\mathfrak{E}\left(\sum_{i=1}^n u_i^2\right) = \sum_{i=1}^n \sigma_i^2 + \sum_{i=1}^n \overline{u_i}^2$$

$$V\left(\sum_{i=1}^n u_i\right) = \sum_{i=1}^n \sigma_i^2 \qquad (3)$$

$$V\left(\sum_{i=1}^n u_i^2\right) = \sum_{i=1}^n [2\sigma_i^4 + 4\overline{u_i}^2\sigma_i^2 + 4\overline{u_i}\,\sigma_i^3\gamma_i + \sigma_i^4\delta_i]$$

$$\operatorname{Cov}\left(\sum_{i=1}^n u_i^2, \sum_{i=1}^n u_i\right) = \sum_{i=1}^n [2\overline{u_i}\,\sigma_i^2 + \sigma_i^3\gamma_i] .$$

(Assumption: variables u_i are uncorrelated among each other – that means: Cov $(u_i, u_j) = 0$ for $i \neq j$.)

The parameters γ_i and δ_i are contained in the terms $\sum_{i=1}^n \overline{u_i}\sigma_i^3\gamma_i$, $\sum_{i=1}^n \sigma_i^4\delta_i$ and $\sum_{i=1}^n \sigma_i^3\gamma_i$. Because of the different signs of γ_i and δ_i for different components these effects may be partially cancelled out. Since we are only interested in approximate results the simplification $\gamma_i = \delta_i = 0$ for each i may be justified.

Putting the explicit expressions (3) into (1) it follows:

$$V = V\left[\frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n u_i}\right] = \frac{\left(2\sum_{i=1}^n \sigma_i^4 + 4\sum_{i=1}^n \overline{u_i}^2\sigma_i^2\right) + \left(\frac{\sum_{i=1}^n \sigma_i^2 + \sum_{i=1}^n \overline{u_i}^2}{\sum_{i=1}^n \overline{u_i}}\right)^2 \cdot \sum_{i=1}^n \sigma_i^2 - 2\frac{\sum_{i=1}^n \sigma_i^2 + \sum_{i=1}^n \overline{u_i}^2}{\sum_{i=1}^n \overline{u_i}}\,2\sum_{i=1}^n \overline{u_i}\,\sigma_i^2}{\left(\sum_{i=1}^n \overline{u_i}\right)^2} \qquad (4)$$

Finally, we will assume:

$$\mathrm{Cov}\,(\overline{u_i}, \sigma_i^2) = \mathrm{Cov}\,(\overline{u_i}^2, \sigma_i^2) = 0\,. \tag{5}$$

Furthermore we introduce the following denotions: $\lambda = \sum_{i=1}^{n} \overline{u_i}/n$ (mean of the $\overline{u_i}$), $\sigma_u^2 =$ variance of the $\overline{u_i}$ and $\overline{\sigma^2} = \sum_{i=1}^{n} \sigma_i^2/n$ (mean of the σ_i^2). Applying (5) to $\sum_{i=1}^{n} \overline{u_i}^2 \sigma_i^2$ and $\sum_{i=1}^{n} \overline{u_i}\,\sigma_i^2$ and putting into (4) gives:

$$V = \frac{2\sum_{i=1}^{n} \sigma_i^4 + \left(\frac{\overline{\sigma^2} + \sigma^2 + \lambda^2}{\lambda}\right)^2 n\,\overline{\sigma^2} - 4\,n\,(\overline{\sigma^2})^2}{n^2\,\lambda^2}\,. \tag{6}$$

Using the variance $V(\sigma_i^2)$ of the variances σ_i^2 the following relation holds:

$$\sum_{i=1}^{n} \sigma_i^4 = n\,(\overline{\sigma^2})^2 + n\,V(\sigma_i^2) \tag{7}$$

(6) together with (7) gives:

$$\begin{aligned} V &= \frac{\overline{\sigma^2}}{n}\,[1 + x^2 + y^2 + 2\,x\,y + 2\,y] + \frac{2\,V(\sigma_i^2)}{n\,\lambda^2} \\ &= \frac{\overline{\sigma^2}}{n}\left[1 + x^2 + y^2 + 2\,y + 2\,x\left(y + \frac{V(\sigma_i^2)}{(\overline{\sigma^2})^2}\right)\right] \end{aligned} \tag{8}$$

where the abbreviations $x = \overline{\sigma^2}/\lambda^2$ and $y = \sigma_u^2/\lambda^2$ have been used.

In the special case of equal variances σ_i^2 for the different components: $\sigma_i^2 = \sigma_0^2$ for each i, $i = 1, 2, \ldots, n$, formula (8) reduces to:

$$V = \frac{\sigma_0^2}{n}\,(1 + x^2 + y^2 + 2\,y + 2\,x\,y)\,. \tag{9}$$

Appendix 2. Proof of expression (9) for the expectation of the total yield

For random variables Y and Z with $\left|\frac{Y}{Z}\right| \leqq C$, the following inequality holds:

$$\left|\mathfrak{E}\left(\frac{Y}{Z}\right) - \frac{\mathfrak{E}(Y)}{\mathfrak{E}(Z)}\right| \leqq \frac{C\,\sqrt{V(Z)}}{|\mathfrak{E}(Z)|} \tag{1}$$

(see, for example: Morgenstern 1968).

If we identify $Y \equiv \sum_{i=1}^{n} u_i^2$ and $Z \equiv \sum_{i=1}^{n} u_i$, inequality (1) gives together with the explicit terms for $\mathfrak{E}\left(\sum_{i=1}^{n} u_i\right)$, $\mathfrak{E}\left(\sum_{i=1}^{n} u_i^2\right)$ and $V\left(\sum_{i=1}^{n} u_i\right)$ from appendix 1:

$$\left|\mathfrak{E}\begin{pmatrix}\text{total}\\ \text{yield}\end{pmatrix} - \frac{\sum_{i=1}^{n} \overline{\sigma_i^2} + \sum_{i=1}^{n} \overline{u_i}^2}{\sum_{i=1}^{n} \overline{u_i}}\right| \leqq \frac{C\sqrt{\sum_{i=1}^{n} \sigma_i^2}}{\left|\sum_{i=1}^{n} \overline{u_i}\right|} \tag{2}$$

Because of $0 \leqq \sum_{i=1}^{n} u_i^2 \Big/ \sum_{i=1}^{n} u_i \leqq 1$ we can use $C = 1$ and from (2) it follows:

$$\left|\mathfrak{E}\begin{pmatrix}\text{total}\\ \text{yield}\end{pmatrix} - \frac{\overline{\sigma^2} + \sigma_u^2 + \lambda^2}{\lambda}\right| \leqq \frac{\sqrt{\overline{\sigma^2}}}{\lambda\sqrt{n}}\,. \tag{3}$$

Applying the abbreviations $x = \overline{\sigma^2}/\lambda^2$ and $y = \sigma_u^2/\lambda^2$ we obtain:

$$\left|\mathfrak{E}\begin{pmatrix}\text{total}\\ \text{yield}\end{pmatrix} - (1 + x + y)\,\lambda\right| \leqq \sqrt{\frac{x}{n}}\,. \tag{4}$$

For sufficiently large n we therefore get the approximate result:

$$\mathfrak{E}\begin{pmatrix}\text{total}\\ \text{yield}\end{pmatrix} \cong \frac{1 + x + y}{\sqrt{x}}\sqrt{\overline{\sigma^2}}\,. \tag{5}$$