A METHOD OF SETTING CONFIDENCE LIMITS
ON THE HERITABILITY RATIO

By

FRANK GARLAND MARTIN, JR.

Bachelor of Science

Oklahoma Agricultural and Mechanical College

Stillwater, Oklahoma

1954

Submitted to the faculty of the Graduate School of
the Oklahoma Agricultural and Mechanical College
in partial fulfillment of the requirements
for the degree of
MASTER OF SCIENCE
August, 1955

A METHOD OF SETTING CONFIDENCE LIMITS

ON THE HERITABILITY RATIO

Thesis Approved:

Franklin Graybill
Thesis Adviser

Herbert Scholz

Robert MacVicar
Dean of the Graduate School

PREFACE

Heritability of a trait is one of the important statistics that must be known if the rate of progress of a breeding program is to be correctly evaluated. There are two methods used to calculate estimates of heritability, the regression technique and the analysis of variance technique. Kempthorne (1) has shown that satisfactory confidence limits can be set on estimates calculated by the regression technique. Little work has been done on the setting of confidence limits when estimated by the analysis of variance technique. Osborne (2) has found an approximation of the standard error of the heritability ratio based on the assumption of normality. The purpose of this thesis is to determine a method of setting confidence limits on estimates found by the analysis of variance technique. Both the case of equal and unequal subclass numbers are considered.

Indebtedness is acknowledged to Dr. Franklin Graybill for suggesting this problem to me and for his helpful criticism given me in the preparation of this thesis. I should also like to acknowledge my indebtedness to Mrs. Edwin Titt for her work in calculating the data used in the empirical study.

TABLE OF CONTENTS

INTRODUCTION

Phenotypic differences between individuals in most traits are partly due to differences in heredity and partly due to the differences of the individual's environment. Each developed trait is the result of the action of genes, the action of the environment, and the interaction of the genes and the environment. Heritability is a quantitative description of the amount of hereditary variation in a trait.

It is important for the livestock breeder to know which traits have some degree of heritability if he wants to make any permanent improvement in his livestock. The only permanent changes in livestock quality are genetic changes brought about by a breeding program that will bring together favorable gene combinations. If the heritability of the trait is high, improvement will quickly follow a good breeding program. If on the other hand the heritability is low, then improvement is long in coming despite the quality of the breeding program.

It is for these reasons that it is desirable to set confidence limits on the heritability ratio, thus giving the livestock breeder an indication of what kind of program he should undertake to bring about improvement in his herd.

Consider the three-fold classification whose model is

$$(1.1) \qquad Y_{ijk} = \mu + a_i + b_{ij} + c_{ijk}$$

where $i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; and $k = 1, 2, \ldots, t$. It is assumed that the a_i are distributed normally with mean zero and variance σ_a^2. Simiarly the b_{ij} and the c_{ijk} are distributed normally with zero means and variances σ_b^2 and σ_c^2 respectively. It is further assumed that all the terms are uncorrelated. The analysis of variance is found in table 1.1.

TABLE 1.1

Analysis of Variance of the Three-fold Classification

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
A effect	$n_3 = r-1$	A_3	$\sigma_3^2 = \sigma_c^2 + t\sigma_b^2 + st\sigma_a^2$
B in A	$n_2 = r(s-1)$	A_2	$\sigma_2^2 = \sigma_c^2 + t\sigma_b^2$
C in B in A	$n_1 = rs(t-1)$	A_1	$\sigma_1^2 = \sigma_c^2$

The purpose of this thesis will be to set confidence limits on the ratio

$$(1.2) \qquad h^2 = \frac{2(\sigma_a^2 + \sigma_b^2)}{\sigma_a^2 + \sigma_b^2 + \sigma_c^2}\,.$$

This is the heritability ratio used in genetic studies to measure the genetic contribution of the sire and the dam to their offspring, where σ_a^2 is the contribution due to the sire, σ_b^2 is the contribution due to the dam, and σ_c^2 is the contribution due to the offspring or the environmental effect.

It is known that the sums of squares in the analysis of variance when divided by its expected mean square is distributed as chi-square. While it is known that the linear sum of independent chi-squares is distributed as a chi-square if and only if the coefficients are unity, it seems safe to assume that a linear combination of independent chi-square variates is well approximated by some chi-square curve.

The method presented in this thesis is somewhat patterned after the method of attack used by Satterthwaite (3) to set confidence limits on variance components. The method purposed in this thesis consists of equating the moments of a linear function

$$(1.3) \qquad Y = N\left[\frac{\alpha_2 A_2 + \alpha_3 A_3}{\sigma_4^2}\right] \quad \text{where} \quad \sigma_4^2 = \gamma(\sigma_a^2 + \sigma_b^2) + \sigma_c^2$$

which is independent of A_1 to the moments of a chi-square with N degrees of freedom. Equal coefficients must be chosen for σ_a^2 and σ_b^2 in order to derive h^2 from (1.3). The value of N is then determined in order to find the "best" agreement among the moments. If this linear combination is closely approximated by chi-square, then the ratio

$$\frac{Y}{N} \Big/ \frac{A_1}{\sigma_c^2}$$

will be approximately distributed as Snedecor's F and it will be possible to determine confidence limits on h^2.

DETERMINATION OF THE DEGREES OF FREEDOM

We now find the moment generating function of

$$(2.1) \qquad Y = N\left[\frac{\alpha_2 A_2 + \alpha_3 A_3}{\sigma_4^2}\right]$$

and determine N such that the first and second moments of (2.1) are equal to the first and second moments of a chi-square with N degrees of freedom where $\sigma_4^2 = \gamma(\sigma_a^2 + \sigma_b^2) + \sigma_c^2$. Equal coefficients are chosen for σ_a^2 and σ_b^2 so that it is possible to set confidence limits on (1.2).

Since A_2 and A_3 are independent, the moment generating function of the sum is equal to the product of the moment generating functions. Hence

$$(2.2) \qquad M_y(t) = \left[M_{\theta_2}(t)\right]\left[M_{\theta_3}(t)\right]$$

where $\theta_i = \alpha_i A_i N/\sigma_4^2$, $i = 2, 3$. It is known that $\theta_i' = n_i A_i/\sigma_i^2$ is distributed as chi-square with n_i degrees of freedom and its moment generating function is $(1 - 2t)^{-n_i/2}$. Since $A_i = \sigma_i^2 \theta_i'/n_i$, it follows that

$$M_{A_i}(t) = \left[1 - \frac{2\sigma_i^2 t}{n_i}\right]^{-n_i/2}$$

and finally we obtain

$$(2.3) \qquad M_{\theta_i}(t) = \left[1 - \frac{2\alpha_i \sigma_i^2 N t}{n_i \sigma_4^2}\right]^{-n_i/2} .$$

Let $B_i = \alpha_i \sigma_i^2 N / n_i \sigma_4^2$, then (2.2) becomes

$$(2.4) \qquad M_y(t) = (1 - 2B_2 t)^{-n_2/2} (1 - 2B_3 t)^{-n_3/2}.$$

Expanding (2.4) we obtain

$$(2.5) \qquad M_y(t) = 1 + (n_2B_2 + n_3B_3)t + \left[n_2(n_2+2)B_2^2 + 2n_2n_3B_2B_3 + n_3(n_3+2)B_3^2\right]t^2/2!$$

$$+ \left[n_2(n_2+2)(n_2+4)B_2^3 + 3n_2(n_2+2)n_3B_2^2B_3 + 3n_2n_3(n_3+2)B_2B_3^2 + n_3(n_3+2)(n_3+4)B_3^3\right]t^3/3!$$

$$+ \ldots + \left[\sum_{p=2}^{3} \prod_{j=0}^{k-1} B_p(n_p+2j) + \sum_{i=1}^{k-1} \binom{k}{i} \prod_{j=0}^{i-1} B_2(n_2+2j) \prod_{m=0}^{k-i-1} B_3(n_3+2m)\right] t^k/k!$$

$$+ \ldots$$

The moment generating function of a chi-square variate with N degrees of freedom is $(1 - 2t)^{-N/2}$. Expanding into an infinite series we get

$$(2.6) \qquad M_{\chi^2}(t) = 1 + Nt + N(N+2)t^2/2! + N(N+2)(N+4)t^3/3! + \ldots + \prod_{i=0}^{k-1} (N+2i)t^k/k! + \ldots$$

If these two moment generating functions are to be equivalent they must have the same set of moments, i.e., equation (2.5) must be identically equal to (2.6) for all k.

Equating the first moments of (2.5) and (2.6) we find that

$$N = n_2B_2 + n_3B_3.$$

Substituting for the B's in the above relation we find that

$$(2.7) \qquad \frac{\alpha_2\sigma_2^2 + \alpha_3\sigma_3^2}{\sigma_4^2} = 1 .$$

It is now possible to determine the values of α_2, α_3, and γ. It follows from (2.7) and table 1.1 that

$$\gamma(\sigma_a^2 + \sigma_b^2) + \sigma_c^2 = \alpha_2(\sigma_c^2 + t\sigma_b^2) + \alpha_3(\sigma_c^2 + t\sigma_b^2 + st\sigma_a^2)$$

and that the following relations must hold if (2.7) is to be true:

1. $\gamma = st\alpha_3$,
2. $\gamma = (\alpha_2 + \alpha_3)t$,
3. $1 = \alpha_2 + \alpha_3$.

The values which satisfy these conditions are $\gamma = t$, $\alpha_2 = (s-1)/s$, and $\alpha_3 = 1/s$.

By equating the second moments of (2.5) and (2.6) it is possible to determine the value of N. Equating the second moments we obtain

$$(2.8) \qquad N^2 + 2N = n_2(n_2+2)B_2^2 + 2n_2n_3B_2B_3 + n_3(n_3+2)B_3^2.$$

Substituting for the B's and dividing by N^2, we get

$$1 + \frac{2}{N} = \frac{(n_2+2)\alpha_2^2\sigma_2^4}{n_2\sigma_4^4} + \frac{2\alpha_2\alpha_3\sigma_2^2\sigma_3^2}{\sigma_4^4} + \frac{(n_3+2)\alpha_3^2\sigma_3^4}{n_3\sigma_4^4}$$

$$= \frac{\alpha_2^2\sigma_2^4 + 2\alpha_2\alpha_3\sigma_2^2\sigma_3^2 + \alpha_3^2\sigma_3^4}{\sigma_4^4} + \frac{2\alpha_2^2\sigma_2^4}{n_2\sigma_4^4} + \frac{2\alpha_3^2\sigma_3^4}{n_3\sigma_4^4}$$

$$(2.9) \qquad = \left[\frac{\alpha_2\sigma_2^2 + \alpha_3\sigma_3^2}{\sigma_4^2}\right]^2 + \frac{2(n_3\alpha_2^2\sigma_2^4 + n_2\alpha_3^2\sigma_3^4)}{n_2n_3\sigma_4^4} .$$

But the first term of the right-hand side of (2.9) is unity by (2.7), hence

$$N = \frac{n_2 n_3 \sigma_4^4}{n_3 \alpha_2^2 \sigma_2^4 + n_2 \alpha_3^2 \sigma_3^4} . \qquad (2.10)$$

If we let $K = \sigma_a^2 / (t\sigma_b^2 + \sigma_c^2)$, then

$$\frac{\sigma_2^4}{\sigma_4^4} = \left[\frac{\sigma_c^2 + t\sigma_b^2}{t\sigma_a^2 + t\sigma_b^2 + \sigma_c^2}\right]^2 = \left[\frac{1}{1 + tK}\right]^2 \quad \text{and}$$

(2.11)

$$\frac{\sigma_3^4}{\sigma_4^4} = \left[\frac{\sigma_c^2 + t\sigma_b^2 + st\sigma_a^2}{t\sigma_a^2 + t\sigma_b^2 + \sigma_c^2}\right]^2 = \left[\frac{1 + stK}{1 + tK}\right]^2 .$$

It is then easily verified that (2.10) simplifies into

$$N = \frac{n_2 n_3 (1 + tK)^2}{n_3 \alpha_2^2 + n_2 \alpha_3^2 (1 + stK)^2} .$$

Finally substituting for n_2, n_3, α_2, and α_3 the formula for N becomes

$$N = \frac{rs^2(r-1)(1 + tK)^2}{(r-1)(s-1) + r(1 + stK)^2} . \qquad (2.12)$$

In this thesis we will assume that $\sigma_a^2 = \sigma_b^2$. It seems feasible to assume that in random matings the contribution of the sire and the dam to the genetic make-up of their offspring is equal since each will contribute one-half of the offspring's genes. It is also known that h^2 as defined by (1.2) is bounded by zero and one. If we let w equal σ_c^2/σ_a^2, then $h^2 = 4/(w+2)$. Then $0 \leq 1/(w+2) \leq 0.25$ and it follows that $(t + 2) \leq (t + w) \leq \infty$. Also $0 \leq \frac{1}{t + w} \leq \frac{1}{t + 2}$. But $K = \frac{1}{t + w}$, therefore

$$(2.13) \qquad 0 \leq tK \leq \frac{t}{t+2} .$$

By using the inequality (2.13) we can determine the minimum and maximum values of N as defined by (2.12) such that the first two moments of (2.5) and (2.6) are identically equal.

When $tK = 0$ (2.12) obtains its maximum value since the denominator obtains its minimum value. In this case (2.12) equals

$$(2.14) \qquad \frac{rs^2(r-1)}{rs - s + 1} .$$

When we substitute $tK = t/(t+2)$ into (2.12) we find the minimum value that N can take. Substituting into (2.12) we get

$$N = \frac{rs^2(r-1)(2t+t)^2}{(r-1)(s-1)(t+2)^2 + r(st+t+2)^2}$$

$$= \frac{4rs^2(r-1)(t+1)^2}{(rs-s+1)(t+2)^2 + rst(st+2t+4)}$$

$$= \frac{4rs^2(r-1)(t+1)^2}{4rs(t+1)^2 + rst^2(s-1) - (s-1)(t+2)^2}$$

$$(2.15) \qquad = \frac{4rs^2(r-1)(t+1)^2}{4rs(t+1)^2 + (s-1)\left[rst^2 - (t+2)^2\right]} .$$

We now must show that N is a monotonic decreasing function of K for the values $0 \leq K \leq 1/(t+2)$. Taking the partial derivative of N with respect to K we get

$$\frac{\partial N}{\partial K} = \frac{2rs^2t(1+tK)\left[(r-1)(s-1) + r(1+stk)^2\right] - 2r^2s^3t(r-1)(1+tK)^2(1+stK)}{\left[(r-1)(s-1) + r(1+stK)^2\right]^2} .$$

If N is a monotonic decreasing function of K, then the partial derivative must be non-positive. Hence

$$\frac{\partial N}{\partial K} \leq 0 \qquad \text{for} \quad 0 \leq K \leq \frac{1}{t+2}\,.$$

Setting $\partial N/\partial K \leq 0$ and simplifying we get

$$1 - s + rstK - rs^2tK \leq 0.$$

Since $s \geq 1$ and $rs^2tK \geq rstK$ the partial derivative is always non-positive and N is a monotonic decreasing function of K for all values of $K \geq 0$. Therefore

$$\text{(2.16)} \qquad \frac{4rs^2(r-1)(t+1)^2}{4rs(t+1)^2 + (s-1)\left[rst^2 - (t+2)^2\right]} \leq N \leq \frac{rs^2(r-1)}{rs - s + 1}\,.$$

FURTHER JUSTIFICATION OF THE USE OF CHI-SQUARE

In order for the ratio

$$\frac{Y}{N} \Big/ \frac{A_1}{\sigma_c^2}$$

to be distributed as F the following conditions must hold:

1. Y/N be distributed as $\chi^2_{(N)}/N$,

2. A_1/σ_c^2 be distributed as $\chi^2_{(n_1)}/n_1$, and

3. Y/N and A_1/σ_c^2 must be independent.

Conditions (2) and (3) are immediately satisfied. We now need to find how closely the moments of Y/N approximate the moments of χ^2/N where chi-square has N degrees of freedom.

Equating the third moments of (2.5) and (2.6) we obtain

$$N(N+2)(N+4) = n_2(n_2+2)(n_2+4)B_2^3 + 3n_2(n_2+2)n_3B_2^2B_3 + 3n_2n_3(n_3+2)B_2B_3^2 + n_3(n_3+2)(n_3+4)B_3^3 \,. \tag{3.1}$$

Substituting for the B's in (3.1) and dividing by N^3 we get

$$1 + \frac{6}{N} + \frac{8}{N^2} = \frac{(n_2+2)(n_2+4)\alpha_2^3\sigma_2^6}{n_2^2\sigma_4^6} + \frac{3(n_2+2)\alpha_2^2\alpha_3\sigma_2^4\sigma_3^2}{n_2\sigma_4^6} + \frac{3(n_3+2)\alpha_2\alpha_3^2\sigma_2^2\sigma_3^4}{n_3\sigma_4^6} + \frac{(n_3+2)(n_3+4)\alpha_3^3\sigma_3^6}{n_3^3\sigma_4^6} \,.$$

Expanding the right-hand side and grouping terms we get

$$1+\frac{6}{N}+\frac{8}{N^2}=\frac{\alpha_2^3\sigma_2^6+3\alpha_2^2\alpha_3\sigma_2^4\sigma_3^2+3\alpha_2\alpha_3^2\sigma_2^2\sigma_3^4+\alpha_3^3\sigma_3^6}{\sigma_4^6}+\frac{8}{\sigma_4^6}\left[\frac{\alpha_3^3\sigma_3^6}{n_3^2}+\frac{\alpha_2^3\sigma_2^6}{n_2^2}\right]$$

$$+\frac{6}{\sigma_4^6}\left[\frac{\alpha_2^3\sigma_2^6}{n_2}+\frac{\alpha_2\alpha_3^2\sigma_2^2\sigma_3^4}{n_3}+\frac{\alpha_2^2\alpha_3\sigma_2^4\sigma_3^2}{n_2}+\frac{\alpha_3^3\sigma_3^6}{n_3}\right]$$

$$=\left[\frac{\alpha_2\sigma_2^2+\alpha_3\sigma_3^2}{\sigma_4^2}\right]^3+\frac{6\alpha_2\sigma_2^2}{\sigma_4^6}\left[\frac{\alpha_2^2\sigma_2^4}{n_2}+\frac{\alpha_3^2\sigma_3^4}{n_3}\right]+\frac{6\alpha_3\sigma_3^2}{\sigma_4^6}\left[\frac{\alpha_2^2\sigma_2^4}{n_2}+\frac{\alpha_3^2\sigma_3^4}{n_3}\right]$$

$$+\frac{8}{\sigma_4^6}\left[\frac{n_3^2\alpha_2^3\sigma_2^6+n_2^2\alpha_3^3\sigma_3^6}{n_2^2n_3^2}\right]$$

(3.2) $$=\left[\frac{\alpha_2\sigma_2^2+\alpha_3\sigma_3^2}{\sigma_4^2}\right]^3+6\left[\frac{n_3\alpha_2^2\sigma_2^4+n_2\alpha_3^2\sigma_3^4}{n_2n_3\sigma_4^4}\right]\left[\frac{\alpha_2\sigma_2^2+\alpha_3\sigma_3^2}{\sigma_4^2}\right]$$

$$+8\left[\frac{n_3^2\alpha_2^3\sigma_2^6+n_2^2\alpha_3^3\sigma_3^6}{n_2^2n_3^2\sigma_4^6}\right].$$

The first term of the right-hand member of equation (3.2) is unity by (2.7) as is the second factor of the second term. The first factor of the second term by (2.10) is $1/N$. Therefore by the third moment

(3.3) $$\frac{1}{N^2}=\frac{n_3^2\alpha_2^3\sigma_2^6+n_2^2\alpha_3^3\sigma_3^6}{n_2^2n_3^2\sigma_4^6}.$$

But by using the value of N as determined by the second moment

(3.4) $$\frac{1}{N^2}=\frac{n_3^2\alpha_2^4\sigma_2^8+2n_2n_3\alpha_2^2\alpha_3^2\sigma_2^4\sigma_3^4+n_2^2\alpha_3^4\sigma_3^8}{n_2^2n_3^2\sigma_4^8}.$$

Using the common denominator $n_2^2n_3^2\sigma_4^8$ and subtracting (3.4) from

(3.3) the difference in the third moments is (working with the numerator only) equal to

$$8(n_3^2\alpha_2^3\sigma_2^6\sigma_4^2 + n_2^2\alpha_3^3\sigma_3^6\sigma_4^2 - n_3^2\alpha_2^4\sigma_2^8 - 2n_2n_3\alpha_2^2\alpha_3^2\sigma_2^4\sigma_3^4 - n_2^2\alpha_3^4\sigma_3^8)$$

$$= -8\left[n_3^2\alpha_2^3\sigma_2^6(\alpha_2\sigma_2^2 - \sigma_4^2) + 2n_2n_3\alpha_2^2\alpha_3^2\sigma_2^4\sigma_3^4 + n_2^2\alpha_3^3\sigma_3^6(\alpha_3\sigma_3^2 - \sigma_4^2)\right] .$$

Now $$\alpha_2\sigma_2^2 - \sigma_4^2 = \frac{s-1}{s}(\sigma_c^2 + t\sigma_b^2) - t(\sigma_a^2 + \sigma_b^2) - \sigma_c^2$$

$$= \frac{-(\sigma_c^2 + t\sigma_b^2 + st\sigma_a^2)}{s}$$

$$= -\alpha_3\sigma_3^2 .$$

Similarly $$\alpha_3\sigma_3^2 - \sigma_4^2 = -\alpha_2\sigma_2^2 .$$

Hence the numerator of the difference equals

$$8\left[n_3^2\alpha_2^3\alpha_2\sigma_2^6\sigma_3^2 - 2n_2n_3\alpha_2^2\alpha_3^2\sigma_2^4\sigma_3^4 + n_2^2\alpha_2\alpha_3^3\sigma_2^2\sigma_3^6\right]$$

$$= 8\alpha_2\alpha_3\sigma_2^2\sigma_3^2(n_3^2\alpha_2^2\sigma_2^4 - 2n_2n_3\alpha_2\alpha_3\sigma_2^2\sigma_3^2 + n_2^2\alpha_3^2\sigma_3^4)$$

$$= 8\alpha_2\alpha_3\sigma_2^2\sigma_3^2(n_3\alpha_2\sigma_2^2 - n_2\alpha_3\sigma_3^2)^2 .$$

Therefore

$$d = \frac{8\,\alpha_2\alpha_3\sigma_2^2\sigma_3^2(n_3\alpha_2\sigma_2^2 - n_2\alpha_3\sigma_3^2)^2}{n_2^2n_3^2\sigma_4^8} . \quad (3.5)$$

Substituting for the constants in (3.5), using (2.11), and simplifying we get

$$d = \frac{8(s-1)(1 + stK)(1 + rstK)^2}{r^2(r-1)^2s^4(1+tK)^4} . \quad (3.6)$$

In heritability studies we have seen that tK can vary between zero and $t/(t+2)$. By using this fact we can find the maximum and minimum difference between the third moments. When $tK = 0$ the difference is a minimum and is equal to

$$(3.7) \qquad \frac{8(s-1)}{s^4 r^2 (r-1)^2}$$

and the difference is less than $8/s^3(r-1)^4$ for all r and s greater than one. When $tK = t/(t+2)$ we have the greatest difference between the two moments. This difference is equal to

$$(3.8) \qquad \frac{(s-1)(t+2)(st+t+2)(rst+t+2)^2}{2r^2(r-1)^2 s^4 (t+1)^4}$$

and the difference is less than $2/(r-1)^2$ for all r and s greater than one.

It can be shown that (3.6) is a monotonic increasing function of K for $0 \leq K \leq 1/(t+2)$, hence

$$(3.9) \qquad \frac{8(s-1)}{s^4 r^2 (r-1)^2} \leq d \leq \frac{(s-1)(t+2)(st+t+2)(rst+t+2)^2}{2r^2(r-1)^2 s^4 (t+1)^4} .$$

There is exact agreement between the two moments of (2.5) and (2.6) when the difference is equal to zero. Setting (3.6) equal to zero we find that

$$8(s-1)(1+stK)(1+rstK)^2 = 0$$

which implies that $(1+stK) = 0$ and/or $(1+rstK) = 0$ since s must be greater than one. This in turn implies that K must be negative. But K was defined as the ratio of positive quantities which seems to indicate that the best possible agreement that one could hope to attain between the third moments is when K is as small as possible.

THE CONFIDENCE INTERVAL

We now need to determine two points, a and b, such that

$$(4.1) \qquad P(a \leq h^2 \leq b) = 1 - 2\alpha$$

where h^2 is defined by (1.2) and $1 - 2\alpha$ is the probability that the interval covers the population value of h^2. The value 2α is some positive quantity less than one.

We have shown that (2.1) is approximately distributed as chi-square with N degrees of freedom. We know that $rs(t-1)A_1/\sigma_c^2$ is distributed as chi-square with $rs(t-1)$ degrees of freedom. These two quantities are independent since they are functions of independent quantities. The ratio of two independent chi-squares divided by their degrees of freedom is distributed as F. Hence the ratio

$$\frac{Y}{N} \Big/ \frac{rs(t-1)A_1}{n_1\sigma_c^2} = F_o$$

is approximately distributed as F with N and n_1 degrees of freedom, where $n_1 = rs(t-1)$. We will assume that this ratio is distributed as F in order to determine the confidence limits. The ratio then simplifies to

$$F_o = \frac{\sigma_c^2(\alpha_2A_2 + \alpha_3A_3)}{A_1(t\sigma_a^2 + t\sigma_b^2 + \sigma_c^2)}$$

$$(4.2) \qquad = \frac{C\sigma_c^2}{t(\sigma_a^2 + \sigma_b^2) + \sigma_c^2} \qquad \text{where} \qquad C = \frac{(s-1)A_2 + A_3}{sA_1} .$$

Let F_1 and F_2 be values of F such that $P(F_1 \leq F \leq F_2) = 1 - 2\alpha$ where $\int_0^{F_1} F\, dF = \alpha$ and $\int_{F_2}^{\infty} F\, dF = \alpha$. If we denote $F_{n_1,N}$ by F^{-1}, then the points F_1 and F_2 can be determined by $\int_0^{f} F^{-1}\, dF^{-1} = 1 - \alpha$ and $\int_0^{F_2} F\, dF = 1 - \alpha$, where $f = 1/F_1$, and read from Snedecor's F tables.

Then to determine the points a and b we know that

$$P(F_1 \leq F_o \leq F_2) = P\left[F_1 \leq \frac{C\sigma_c^2}{t(\sigma_a^2 + \sigma_b^2) + \sigma_c^2} \leq F_2\right]$$

$$= P\left[\frac{C}{F_2} \leq \frac{t(\sigma_a^2 + \sigma_b^2) + \sigma_c^2}{\sigma_c^2} \leq \frac{C}{F_1}\right]$$

$$= P\left[\frac{C - F_2}{tF_2} \leq \frac{\sigma_a^2 + \sigma_b^2}{\sigma_c^2} \leq \frac{C - F_1}{tF_1}\right]$$

$$= P\left[\frac{tF_1}{C - F_1} + 1 \leq \frac{\sigma_a^2 + \sigma_b^2 + \sigma_c^2}{\sigma_a^2 + \sigma_b^2} \leq \frac{tF_2}{C - F_2} + 1\right]$$

$$= P\left[\frac{2(C - F_2)}{C + (t-1)F_2} \leq \frac{2(\sigma_a^2 + \sigma_b^2)}{\sigma_a^2 + \sigma_b^2 + \sigma_c^2} \leq \frac{2(C - F_1)}{C + (t-1)F_1}\right]. \tag{4.3}$$

Hence the points a and b are determined and (4.3) is an $(1 - 2\alpha)\%$ confidence interval on h^2.

THE CASE OF UNEQUAL SUBCLASS NUMBERS

A more practical case for us to consider in a breeding program is the case of unequal subclass numbers. While we may be able to mate each sire to the same number of dams, it may be impossible for us to obtain an equal number of offspring from each dam. We will now set confidence limits on the heritability ratio as defined by (1.2) for the case of unequal subclass numbers.

Our model is

$$(5.1) \qquad Y_{ijk} = \mu + a_i + b_{ij} + c_{ijk}$$

where $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s_i$, $k = 1, 2, \ldots, t_{ij}$, and $\sum_{i,j} t_{ij} = n$. The assumptions made for the case of equal subclasses still hold. The analysis of variance is shown in table 5.1.

TABLE 5.1

Analysis of Variance of the Three-fold Classification, Unequal Subclass Numbers

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
A effect	$n_3 = r-1$	A_3	$\sigma_3^2 = \sigma_c^2 + k_1\sigma_b^2 + k_2\sigma_a^2$
B in A	$n_2 = \sum_i s_i - r$	A_2	$\sigma_2^2 = \sigma_c^2 + k_0\sigma_b^2$
C in B in A	$n_1 = n - \sum_i s_i$	A_1	$\sigma_1^2 = \sigma_c^2$

Ganguli (4) presents the formulas for the expected values of the mean squares when there are unequal numbers in the various subclasses. They are as follows:

$$k_1 = \sum_i \sum_j \frac{t_{ij}^2\,[(1/t_i) - (1/t)]}{n_3}, \qquad k_2 = \sum_i \frac{t_i^2\,[(1/t_i) - (1/t)]}{n_3},$$

and $k_0 = \sum_i \sum_j \frac{t_{ij}^2\,[(1/t_{ij}) - (1/t_i)]}{n_2}$ where t_{ij} is the number of offspring of sire i and dam j, t_i is the total number of offspring of sire i, and t is the total number of offspring.

Let

$$(5.2) \qquad Y = N\left[\frac{\alpha_2 A_2 + \alpha_3 A_3}{\sigma_4^2}\right]$$

where $\sigma_4^2 = \gamma(\sigma_a^2 + \sigma_b^2) + \sigma_c^2$ as was done for equal subclasses. As the fact that there are unequal class numbers does not effect the distribution of the sum of squares or the fact that they are independent, then the moment generating function of (5.2) is the same as for equal subclasses. Hence the moment generating function of (5.2) is given by (2.5).

As before we desire to see if the constants γ, α_2, α_3, and a value for N can be determined such that the first and second moments of (5.2) are equal to the first and second moments of a chi-square with N degrees of freedom.

Equating the first moments of (2.5) and (2.6) we get

$$N = n_2 B_2 + n_3 B_3$$

and substituting for the B's, where $B_i = N\sigma_i^2\alpha_i / n_i\sigma_4^2$, we find that (2.7) still holds, i.e.,

$$\frac{\alpha_2\sigma_2^2 + \alpha_3\sigma_3^2}{\sigma_4^2} = 1$$

or

$$\gamma(\sigma_a^2 + \sigma_b^2) + \sigma_c^2 = \alpha_2(\sigma_c^2 + k_0\sigma_b^2) + \alpha_3(\sigma_c^2 + k_1\sigma_b^2 + k_2\sigma_a^2)$$

$$= \alpha_3 k_2\sigma_a^2 + (\alpha_2 k_0 + \alpha_3 k_1)\sigma_b^2 + (\alpha_2 + \alpha_3)\sigma_c^2.$$

The constants are then defined by the following conditions which must hold if (2.7) is true. The conditions are:

1. $\gamma = \alpha_3 k_2 = (\alpha_2 k_0 + \alpha_3 k_1)$,

2. $1 = \alpha_2 + \alpha_3$.

Solving the pair of equations

$$k_2\alpha_3 = k_0\alpha_2 + k_1\alpha_3$$

$$1 = \alpha_2 + \alpha_3$$

for α_2 and α_3 we find

$$\alpha_2 = \frac{k_2 - k_1}{k_0 - k_1 + k_2},$$

(5.3) $$\alpha_3 = \frac{k_0}{k_0 - k_1 + k_2}, \qquad \text{and}$$

$$\gamma = \frac{k_0 k_2}{k_0 - k_1 + k_2}.$$

Now by equating the second moments we can determine the value for N that will make the first two moments of (5.2) equal to those of chi-square with N degrees of freedom. Equating second moments we get

$$1 + \frac{2}{N} = \frac{(n_2+2)\alpha_2^2\sigma_2^4}{n_2\sigma_4^4} + \frac{2\alpha_2\alpha_3\sigma_2^2\sigma_3^2}{\sigma_4^4} + \frac{(n_3+2)\alpha_3\sigma_3^4}{n_3\sigma_4^4}$$

after substituting for the B's. Simplifing we find that

(5.4) $$N = \frac{n_2 n_3\sigma_4^4}{n_3\alpha_2^2\sigma_2^4 + n_2\alpha_3^2\sigma_3^4}.$$

Substituting for the constants in (5.4) we get

$$(5.5) \qquad N = \frac{(r-1)(\Sigma s_i - r)\left[\frac{k_0 k_2}{k_0-k_1+k_2}(\sigma_a^2 + \sigma_b^2) + \sigma_c^2\right]^2}{\frac{(r-1)(k_2-k_1)^2}{(k_0-k_1+k_2)^2}(\sigma_c^2 + k_0\sigma_b^2)^2 + \frac{(\Sigma s_i - r)k_0^2}{(k_0-k_1+k_2)^2}(\sigma_c^2 + k_1\sigma_b^2 + k_2\sigma_a^2)^2}.$$

If we let $K = \sigma_a^2 / (T\sigma_b^2 + \sigma_c^2)$ where $T = k_0 k_2 / (k_0-k_1+k_2)$ and if we again assume that we have random matings such that $\sigma_a^2 = \sigma_b^2$, then the following relations can be derived:

(5.6)

1. $\dfrac{\sigma_a^2}{\sigma_4^2} = \dfrac{\sigma_b^2}{\sigma_4^2} = \dfrac{K}{1 + TK}$,

2. $\dfrac{\sigma_c^2}{\sigma_4^2} = \dfrac{1 - TK}{1 + TK}$,

3. $\dfrac{\sigma_2^2}{\sigma_4^2} = \dfrac{1 + (k_0 - T)K}{1 + TK}$, and

4. $\dfrac{\sigma_3^2}{\sigma_4^2} = \dfrac{1 + (k_1 + k_2 - T)K}{1 + TK}$.

Letting $k = (k_0-k_1+k_2)$ and using the relations (5.6) N becomes equal to

$$(5.7) \qquad \frac{(r-1)(\Sigma s_i - r)k^2(1 + TK)^2}{(r-1)(k_2-k_1)^2\left[1 + (k_0-T)K\right]^2 + (\Sigma s_i - r)k_0^2\left[1 + (k_1+k_2-T)K\right]^2}.$$

Since we are assuming that the contribution of the sire and the dam to their offspring is equal, it can be shown that K is bounded by zero and $1/(T+2)$ by an argument similar to that used in Chapter II. By using this inequality we can determine the minimum and maximum values of N when N is defined by (5.7). When $k = 0$ (5.7) obtains its maximum value and the minimum value of N is determined when $K = 1/(T+2)$. When $K = 0$ (5.7) becomes

$$(5.8) \qquad \frac{(r-1)(\Sigma s_i - r)k^2}{(r-1)(k_2-k_1)^2 + (\Sigma s_i - r)k_0^2}$$

and when $K = 1/(T+2)$ (5.7) becomes

$$(5.9) \qquad \frac{4(r-1)(\Sigma s_i - r)k^2(1+T)^2}{(r-1)(k_2-k_1)^2(k_0+2)^2 + (\Sigma s_i - r)k_0^2(k_1+k_2+2)^2} \,.$$

We now need to determine the points a and b such that $P(a \leq h^2 \leq b) = 1 - 2\alpha$ where h^2 is defined by (1.2), $1 - 2\alpha$ is the probability that the interval covers the population value of h^2, and 2α is some positive quantity less than one.

The ratio

$$\frac{Y}{N} \Big/ \frac{(n-\Sigma s_i)A_1}{n_1\sigma_c^2} = F^*$$

is approximately distributed as F with N and $n-\Sigma s_i$ degrees of freedom. We will assume that this ratio is distributed as F in order to determine the confidence limits. The ratio simplifies to

$$F^* = \frac{\sigma_c^2(\alpha_2 A_2 + \alpha_3 A_3)k}{A_1\left[k_0 k_2(\sigma_a^2 + \sigma_b^2) + k\sigma_c^2\right]} \,.$$

Let $C = \dfrac{(k_2-k_1)A_2 + k_0 A_3}{A_1}$ and $k^* = k_0 k_2$, then F^* simplifies to

$$\frac{C\sigma_c^2}{k^*(\sigma_a^2 + \sigma_b^2) + k\sigma_c^2} .$$

Let F_1 and F_2 be values of F such that $P(F_1 \leq F \leq F_2)$ equals $1 - 2\alpha$ where $\int_0^{F_1} F\,dF = \alpha$ and $\int_{F_2}^{\infty} F\,dF = \alpha$. Then to determine the points a and b we know that

$$P(F_1 \leq F^* \leq F_2) = P\left[F_1 \leq \frac{C\sigma_c^2}{k^*(\sigma_a^2 + \sigma_b^2) + k\sigma_c^2} \leq F_2\right]$$

$$= P\left[\frac{C}{kF_2} \leq \frac{k^*(\sigma_a^2 + \sigma_b^2) + k\sigma_c^2}{k\sigma_c^2} \leq \frac{C}{kF_1}\right]$$

$$= P\left[\frac{C - kF_2}{k^*F_2} \leq \frac{\sigma_a^2 + \sigma_b^2}{\sigma_c^2} \leq \frac{C - kF_1}{k^*F_1}\right]$$

$$= P\left[\frac{k^*F_1}{C - kF_1} + 1 \leq \frac{\sigma_a^2 + \sigma_b^2 + \sigma_c^2}{\sigma_a^2 + \sigma_b^2} \leq \frac{k^*F_2}{C - kF_2} + 1\right]$$

$$(5.10) \qquad = P\left[\frac{2(C - kF_2)}{C + (k^*-k)F_2} \leq \frac{2(\sigma_a^2 + \sigma_b^2)}{\sigma_a^2 + \sigma_b^2 + \sigma_c^2} \leq \frac{2(C - kF_1)}{C + (k^*-k)F_1}\right] .$$

Hence the points a and b are determined and (5.10) is an $(1 - 2\alpha)\%$ confidence interval on h^2.

SOME EMPIRICAL RESULTS

A population of 100,000 items from a normal population with mean zero and variance one was used to construct a genetic model similar to the three-fold classification of sire, dam, and offspring that is found in a breeding population. From this population was constructed eight different combinations of sample sizes. The eight cases are listed in table 6.1.

TABLE 6.1

Combinations of Samples Drawn for the Empirical Study

Case Number	Number of Sires	Number of Dams per Sire	Number of Offspring per Dam
1	2	2	2
2	3	2	2
3	5	2	2
4	10	2	2
5	2	4	2
6	3	4	2
7	5	4	2
8	10	4	2

For each of the combinations a sample of size 500 was drawn from the known population and h^2 calculated by the analysis of variance technique. Thus for each sample size combination there were calculated 500 h^2's by the use of (1.2). The estimates for the variance components were found from the analysis of variance. Then 95% confidence limits were set on each of the h^2's by the method presented in this thesis. To give an empirical check of the results of this method the number of intervals that contained the population value for h^2 were counted. Since it was found that N varied according to the value of K, 95% confidence limits were calculated using the minimum, maximum, and average value of N.

The population value of h^2 in this empirical study was 4/3. The data for this study was available from another study and this is the reason that h^2 was greater than one. It was easier to calculate the quantities needed using this data than to construct another population where h^2 was less than one.

To faciliate the calculation of the variance components the actual calculations were done by I.B.M. machines. The items were taken from a set of random normal deviates published by the Rand Corporation. The values of h^2 were also calculated by I.B.M. equipment.

As the confidence limits were calculated by the use of hand calculators, a method was devised to speed up the determination of the interval. A quantity

$$H_o = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_c^2}$$

was calculated and confidence limits were set on H_o such that the interval on h^2 would not exceed the 95% confidence region. This is easily seen by examing the confidence interval (4.3) and expressing it in terms of the variance components. Making the necessary substitutions into (4.3) we find that

$$(6.1) \qquad P\left[\frac{2(tH_o + 1 - F_2)}{tH_o + 1 + (t-1)F_2} \leq h^2 \leq \frac{2(tH_o + 1 - F_1)}{tH_o + 1 + (t-1)F_1}\right] = 1 - 2\alpha .$$

Since every quantity is constant for a fixed sample size and for a fixed confidence level, and since the population value of h^2 is known, then the inequalities of (6.1) can be solved to determine the maximum and minimum values of H_o that will give a 95% confidence interval on h^2. The results of this empirical study are listed in table 6.2 for maximum, minimum, and average N.

TABLE 6.2

Percentage of Confidence Limits Containing the Population Heritability Ratio 4/3 for Maximum, Minimum, and Average Values of N

Case Number	Percentage of Confidence Limits Containing Population Value for		
	Maximum N	Minimum N	Average N
1	94.8	97.6	97.6
2	91.6	93.6	93.6
3	96.2	96.8	96.8
4	95.8	96.6	96.4
5	91.6	96.8	93.8
6	91.0	95.6	93.0
7	89.6	93.0	90.4
8	92.6	94.6	94.2

CONCLUSIONS

This thesis proposes a method of setting confidence limits on estimates of heritability determined by the analysis of variance technique. The maximum and minimum values of N along with the formulas for determining N are presented. This value was determined so that the linear combination of the mean squares (2.1) is best approximated by a chi-square variate with N degrees of freedom. In this thesis it was assumed that the contribution of the sire and the dam to the genetic make-up of their offspring is equal in random matings.

Since the value of N that is used to set confidence limits is a function of K, one might well estimate K from the analysis of variance to decide which value of N should be used. If the estimate, $\hat{K}$, lies between 0 and $1/(t+2)$, then the value of N can be determined from (2.12). The use of an average value of N will also give good results in this case. If $\hat{K}$ falls outside the possible range for K then the nearest possible value for K should be used.

In the case of unequal subclasses a similar procedure may be used with N being determined by (3.7). Instead of calculating the coefficients k_0, k_1, and k_2 they may be estimated by averages. In this case k_0 and k_1 are estimated by the average number of offspring for each sire and dam and k_2 is estimated by the average number of offspring for each sire. This should give fairly good results if the subclass numbers are not too divergent.

Once N is determined the values of F can be found from Snedecor's F tables and confidence limits set on h^2.

BIBLIOGRAPHY

(1) Kempthorne, O., and O. B. Tandon. "The Estimation of Heritability by Regression of Offspring on Parent". Biometrics, 9 (March, 1953), 90-98.

(2) Osborne, R., and W. S. B. Paterson. "On the Sampling Variance of Heritability Estimates Derived from Variance Analysis." Proceedings of the Royal Society of Edinburgh, B, LXIV (1952), 456-461.

(3) Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components". Biometrics Bulletin, 2 (December, 1946), 110-114.

(4) Ganguli, M. "A Note on Nested Sampling." Sankhyā, 5 (1941), 449-452.

VITA

Frank Garland Martin, Jr.
candidate for the degree of
Master of Science

Thesis: A METHOD OF SETTING CONFIDENCE LIMITS ON THE HERITABILITY RATIO

Major: Mathematics

Minor: Statistics

Biographical:

Born: October 9, 1932 at New Orleans, Louisiana

Undergraduate Study: Oklahoma Agricultrual and Mechanical College, 1950-54

Graduate Study: Virginia Polytechnic Institute, 1954; Oklahoma Agricultrual and Mechanical College, 1954-55

Experiences: Night watchman, 1952-54; Computer in the Statistical Laboratory, 1954; Graduate teaching assistant in the Mathematics Department, 1954-55

Member of Phi Eta Sigma, Pi Mu Epsilion, Phi Kappa Phi, Omicron Delta Kappa, The American Statistical Association, and Associate Member of The Society of Sigma Xi.

Date of Final Examination: August, 1955

THESIS TITLE: A METHOD OF SETTING CONFIDENCE LIMITS
ON THE HERITABILITY RATIO

AUTHOR: Frank Garland Martin, Jr.

THESIS ADVISER: Dr. Franklin A. Graybill

The content and form have been checked and approved by the author and thesis adviser. The Graduate School Office assumes no responsibility for errors either in form or content. The copies are sent to the bindery just as they are approved by the author and faculty adviser.

TYPIST: Frank G. Martin, Jr.