A Distribution-Free Asymptotic Method of Estimating, Testing, and Setting Confidence Limits for Heritability

Author(s): Lorraine Schwartz and Stanley Wearden

Source: *Biometrics*, Jun., 1959, Vol. 15, No. 2 (Jun., 1959), pp. 227-235

Published by: International Biometric Society

Stable URL: https://www.jstor.org/stable/2527671

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/2527671?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

A DISTRIBUTION-FREE ASYMPTOTIC METHOD OF ESTIMATING, TESTING, AND SETTING CONFIDENCE LIMITS FOR HERITABILITY*

Lorraine Schwartz and Stanley Wearden

Kansas State University
Manhattan, Kansas, U.S.A.

Introduction

The genetic parameter heritability indicates the extent to which phenotypic variation is controlled by gene action. Its detection is often one of the chief purposes of a genetic experiment. Many different procedures have been used to estimate heritability when the observable random variables can be assumed to be normally distributed. For example, when a linear regression model is used, the regression coefficient, or slope of the regression line, is taken to be a measure of heritability and normal theory is used to test hypotheses about this parameter and to find confidence regions for it. However, there are certain characteristics of interest to the geneticist which are known to be distributed in a non-normal fashion; for example, the social order of fowl has a uniform distribution. In this paper, the expectation of a rank order statistic closely related to the Mann-Whitney U statistic is suggested as a measure of heritability when the normal assumption is not appropriate and in particular when rank within a group is a meaningful measure of the trait under consideration.

Heritability

For experiments in which selection is conducted in opposite directions for the same trait, there is a distribution-free method for estimating heritability. The result may be described as the ratio of change per generation to the average selection differential, or

$$H = f \frac{\Delta o}{\Delta p} , \tag{1}$$

a constant f times the ratio of the difference between the arithmetic

means of the two groups of progeny (Δo) to the average selection differential in the parental generation (Δp). The constant f is unity if selection is performed on both parents. If selection is made with respect to only one of the parents, the ratio estimates only one half of the hereditary effect and must be doubled to give an unbiased (in the genetic sense) estimate of heritability.

However, when it is desired to test hypotheses or find confidence intervals for heritability [defined here by $E(H)$], then the distribution of H must be known. Hence, it is suggested that the actual observations be transformed to rank order values when such ranks still classify the data reliably. The rank order analog of (1), denoted below by H_r, is shown to be a function of the Mann-Whitney U statistic, and properties of this statistic are employed for the objectives mentioned above.

The Mann-Whitney U Statistic

Suppose the random variables X_1, X_2, $\cdots$, X_{n_1}; Y_1, $\cdots$, Y_{n_2} are completely independent and have continuous cumulative distribution functions $F(s) = P\{X_i < s\}$ and $G(s) = P\{Y_j < s\}$, where $i = 1$, $\cdots$, n_1; $j = 1$, $\cdots$, n_2 and F and G are unknown. Let the X's and Y's be arranged in order and let U count the number of times a Y variable precedes an X variable. The statistic T which is the sum of the ranks of the Y's among the ordered sequence of X's and Y's was proposed by Wilcoxon [1945] as a test criterion for testing the hypothesis that F and G are identical [$F(s) = G(s)$] against all alternatives of the form $G(s) \leq F(s)$ for all real s with strict inequality for some s (i.e. that the Y's are stochastically larger than the X's). Wilcoxon's test, which is to reject the hypothesis for "large" values of T, was proposed for the case when $n_1 = n_2$. Later, Mann and Whitney |1947] proposed essentially the same test for arbitrary n_1 and n_2 and found the distribution, in terms of a recursion formula, for the test statistic U which is related to T by

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T.$$

The test in the final form is, then, to reject the hypothesis of equality of distributions when the observed value of U is "too small," the critical values of U being determined by the level of significance.

Whatever the relationship of F to G may be, the expectation and variance of U are

$$E(U) = n_1 n_2 p \tag{2}$$

$$\sigma_U^2 = n_1 n_2 (n_1 - 1)\varphi^2 + (n_2 - 1)\gamma^2 + p(1 - p), \tag{3}$$

where $p = P\{Y_j < X_i\}$, $\varphi^2 = P\{X_i, X_k < Y_j\} - (1 + p)^2$, and $\gamma^2 = P\{Y_j, Y_k < X_i\} - p^2$, as shown by Mann and Whitney [1947]; the notation is that of Birnbaum and Klose [1957].

If the distributions F and G are identical, then $p = \frac{1}{2}$ and $\varphi^2 = \gamma^2 = 1/12$. Thus, under the hypothesis of equality of distributions,

$$E(U) = n_1 n_2/2 \text{ and } \sigma_U^2 = n_1 n_2(n_1 + n_2 + 1)/12. \tag{4}$$

For this special situation, the distribution of U for small values of n_1 and n_2 may be found in tables calculated by Mann and Whitney [1947], while for larger values, it is closely approximated by the normal distribution.

In general, the expectation and variance of U depend on three parameters, p, φ^2, and γ^2. However, van Dantzig [1951] and Birnbaum and Klose [1957] have obtained strict bounds for the variance, as given in (3), which depend only on p, and hence only on $E(U)$. These are

$$\sigma_U^2 \leqq n_1 n_2 p(1 - p) \max (n_1, n_2) \tag{5}$$

and

$$\sigma_U^2 \geqq \begin{cases} n_1 n_2 \left[\mu r(1 - r) - \dfrac{(\mu - 1)^2}{12(\nu - 1)} \right], & \text{if } \dfrac{\mu - 1}{\nu - 1} \leqq 2r \qquad (6) \\[3mm] n_1 n_2 \left[\dfrac{4}{3} r \sqrt{2(n_1 - 1)(n_2 - 1)r} - (n_1 + n_2 - 2)r^2 + p(1 - p) \right], \\[3mm] \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \dfrac{\mu - 1}{\nu - 1} \geqq 2r \end{cases}$$

where $\mu = \min (n_1, n_2)$, $\nu = \max (n_1, n_2)$, and $r = \min (p, 1 - p)$. They also provided even sharper bounds in the case where Y is known to be stochastically larger than X. However, these will not be used in this paper.

Lehmann [1950] proved that

$$\hat{p} = U/n_1 n_2 \tag{7}$$

is the uniformly minimum variance unbiased estimate of p. Other results of Lehman [1951] include a proof of the asymptotic normality of $\sqrt{n_2} (\hat{p} - p)$, no matter what F and G may be, when two conditions are satisfied: i) n_1/n_2 remains constant as $n_1, n_2 \to \infty$, and ii) $p \neq 0, 1$. The approach to normality is most rapid when p is near $\frac{1}{2}$ and becomes less rapid the farther p deviates from $\frac{1}{2}$.

Since, for large sample sizes, the distribution of U is approximately normal, it is suggested [Birnbaum, 1956] that the normal approximation be used with σ_U replaced by its upper bound in (5). The effect of sub-

stitution of the upper bound for σ_U is to "widen" the confidence interval obtained. More precisely, if $1 - \alpha$ is the assigned level of confidence and M is the square root of the upper bound for σ_U^2, then

$$P\left\{ \left| \frac{U - E(U)}{M} \right| \leqq u_\alpha \right\} \geqq P\left\{ \left| \frac{U - E(U)}{\sigma_U} \right| \leqq u_\alpha \right\} = 1 - \alpha. \qquad (8)$$

For large n_1 and n_2, u_α may be determined from tables of the normal distribution with zero expectation and unit variance.

The inequality in the first bracket in (8) may be written as

$$-u_\alpha \leqq \frac{U - n_1 n_2 p}{\sqrt{n_1 n_2 p(1 - p) \max (n_1, n_2)}} \leqq u_\alpha$$

$$\text{or} \quad p_L \leqq p \leqq p_U$$

where the lower bound

$$p_L = \frac{\dfrac{2U}{n_1 n_2} + a - \sqrt{\left(\dfrac{2U}{n_1 n_2} + a\right)^2 - 4\left(\dfrac{U}{n_1 n_2}\right)^2 (1 + a)}}{2(1 + a)} \qquad (9)$$

and the upper bound

$$p_U = \frac{\dfrac{2U}{n_1 n_2} + a + \sqrt{\left(\dfrac{2U}{n_1 n_2} + a\right)^2 - 4\left(\dfrac{U}{n_1 n_2}\right)^2 (1 + a)}}{2(1 + a)} \qquad (10)$$

with $a = (u_\alpha^2/n_1 n_2) \max (n_1, n_2)$.

Relationship of the Statistic H_r to the Mann-Whitney U Statistic

If selection is performed on the parental generation in order to divide it into the two groups, one of which can be rated as being above the median rank for the trait under consideration and the other as below median rank for the same trait, then the estimator (H_r) of heritability which depends on rank is

$$H_r = f \frac{b_2 - b_1}{a_2 - a_1}. \qquad (11)$$

In (11), $(a_2 - a_1)$ is the difference in mean rank between the high parental group and the low, while $(b_2 - b_1)$ is the difference in the mean ranks of the progeny of the high parental group and those of the low.

The statistic H_r is seen to be a function of the ranks, among all the progeny, of those whose parents had rank above the median rank (the high parental group), so that (11) may be written in terms of the Mann-

Whitney U statistic. If N is the number of parents, the difference between the mean ranks of the high and the low parental groups is

$$a_2 - a_1 = \frac{2}{N} \sum_{k=(N/2)+1}^{N} k - \frac{2}{N} \sum_{k=1}^{N/2} k = \frac{N}{2}. \qquad (12)$$

Let R_i be the rank, among all the progeny, of the ith offspring from the low parental group, and S_i be the rank, among all the progeny, of the ith offspring from the high parental group. Then

$$b_2 - b_1 = \frac{1}{n_2} \sum_{i=1}^{n_2} S_i - \frac{1}{n_1} \sum_{i=1}^{n_1} R_i$$

$$= \frac{n_1 + n_2}{n_1 n_2} \sum_{i=1}^{n_2} S_i - \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2 n_1}, \qquad (13)$$

where n_1 is the number of offspring from the low parental group, and n_2 is the number of offspring from the high parental group. Since $\sum_{i=1}^{n_2} S_i$, which is the Wilcoxon T, is equal to $n_1 n_2 + [n_2(n_2 + 1)/2] - U$, the estimator

$$H_r = 2f \frac{n_1 + n_2}{N} \left(\frac{1}{2} - \frac{U}{n_1 n_2} \right). \qquad (14)$$

Point Estimate and Confidence Interval for Heritability

Heritability will be assumed to be measured by $E(H_r) = \theta$. Since by (14), $\theta = 2f\,(n_1 + n_2)(\frac{1}{2} - p)/N$ where p is defined as in (2), the point estimate for heritability is

$$\hat{\theta} = 2f \left(\frac{n_1 + n_2}{N} \right) \left(\frac{1}{2} - \hat{p} \right). \qquad (15)$$

The limits of the confidence interval for θ are obtained by (9), (10), and (14). The lower limit is

$$\theta_L = 2f \left(\frac{n_1 + n_2}{N} \right) \left(\frac{1}{2} - p_U \right) \qquad (16)$$

and the upper limit is

$$\theta_U = 2f \left(\frac{n_1 + n_2}{N} \right) \left(\frac{1}{2} - p_L \right). \qquad (17)$$

Test of the Hypothesis that there is no Heritability

The test of the hypothesis (H_0) that θ is zero against the specific alternative hypothesis that θ is greater than zero will be precisely the test of the hypothesis that $p = \frac{1}{2}$ against the alternatives $p < \frac{1}{2}$. By

TABLE 1
Percentages of Flock Dominated by Low and High Dams

Low Dams	Per Cent of Flock Dominated	Daughters of Low Dams	Per Cent of Flock Dominated	Rank (R_i)	High Dams	Per Cent of Flock Dominated	Daughters of High Dams	Per Cent of Flock Dominated	Rank (S_i)
A_1	4.8	a_{11}	31.2	15.0	B_1	52.9	b_{11}	87.4	42.0
A_1	4.8	a_{12}	6.2	3.0	B_2	55.8	b_{21}	97.9	47.0
A_2	8.8	a_{21}	8.3	4.0	B_2	55.8	b_{22}	81.1*	39.5
A_3	11.8	a_{31}	79.0	38.0	B_3	61.7	b_{31}	54.1	26.0
A_4	14.7	a_{41}	18.7	9.0	B_4	64.7	b_{41}	37.4	18.0
A_4	14.7	a_{42}	95.7	46.0	B_4	64.7	b_{42}	20.8	10.0
A_5	17.6	a_{51}	22.9	11.0	B_5	67.6	b_{51}	14.6	7.0
A_6	20.6	a_{61}	39.5	19.0	B_6	70.6	b_{61}	68.4	33.0
A_7	23.5	a_{71}	66.6	32.0	B_7	73.5	b_{71}	33.2	16.0
A_7	23.5	a_{72}	58.3	28.0	B_8	76.4	b_{81}	50.0	24.0
A_8	26.5	a_{81}	2.1	1.0	B_8	76.4	b_{82}	12.5	6.0
A_9	29.4	a_{91}	35.4	17.0	B_9	79.4	b_{91}	10.4	5.0
A_{10}	33.8	$a_{10,1}$	62.4	30.0	B_{10}	82.3	$b_{10,1}$	60.3	29.0
A_{11}	33.8	$a_{11,1}$	41.6	20.0	B_{10}	82.3	$b_{10,2}$	56.2	27.0
A_{11}	33.8	$a_{11,2}$	25.0	12.0	B_{11}	85.3	$b_{11,1}$	93.6	45.0
A_{12}	35.3	$a_{12,1}$	43.7	21.0	B_{12}	88.2	$b_{12,1}$	89.4	43.0
A_{13}	38.2	$a_{13,1}$	29.1	14.0	B_{12}	88.2	$b_{12,2}$	74.9	36.0
A_{13}	38.2	$a_{13,2}$	16.6	8.0	B_{13}	91.1	$b_{13,1}$	100.0	48.0
A_{14}	41.2	$a_{14,1}$	85.3	41.0	B_{14}	94.1	$b_{14,1}$	27.0	13.0
A_{15}	44.0	$a_{15,1}$	52.0	25.0	B_{14}	94.1	$b_{14,2}$	91.5	44.0
A_{16}	47.0	$a_{16,1}$	45.8	22.0	B_{15}	97.0	$b_{15,1}$	81.1*	39.5
A_{16}	47.0	$a_{16,2}$	47.8	23.0	B_{15}	97.0	$b_{15,2}$	77.0	37.0
A_{17}	48.5	$a_{17,1}$	4.2	2.0	B_{16}	100.0	$b_{16,1}$	72.8	35.0
A_{18}	50.0	$a_{18,1}$	70.7	34.0					
A_{18}	50.0	$a_{18,2}$	64.5	31.0					
		$n_1 = 25$		$\sum R_i =$ 506.0			$n_2 = 23$		$\sum S_i =$ 670.0

*Individuals b_{22} and $b_{15,1}$ are tied for the 39th and 40th ranks, so an average of the two ranks is assigned to each.

means of the correspondence of tests to confidence sets, a one-sided test for $p = \frac{1}{2}$, against $p < \frac{1}{2}$, may be obtained by taking as the rejection region the set of values of U for which $p_U < \frac{1}{2}$, or equivalently when

$$U < \frac{n_1 n_2}{2} - u_\alpha \sqrt{\frac{1}{4} n_1 n_2 \max (n_1, n_2)}.$$

The test is, then, to reject H_0 whenever

$$H_r > f \frac{u_\alpha (n_1 + n_2)}{N} \sqrt{\frac{\max (n_1, n_2)}{n_1 n_2}}. \tag{18}$$

Example

Table 1 represents the kind of information the geneticist might obtain on the social order within a flock of hens. It is desired to determine the extent to which a dam's social position is inherited by her daughter(s). Figuratively, the maternal flock can be divided into two groups, a "high" group within which each member dominates over one-half of her contemporaries, and a "low" group in which no member dominates more than one-half of the other birds in the maternal flock.

The actual number of dams in Table 1 is thirty-four, but the effective number, because of differential reproductive rates, is forty-eight. For example, a dam such as A_{18}, which dominates fifty per cent of the flock and produces two daughters, would be expected to have the same effect on the next generation as two dams, each of which dominates fifty per cent of the flock and produces only one daughter. Thus, the value of N in this example will be forty-eight.

Using the information in Table 1, the calculated value of U is

$$(25)(23) + \frac{23(23 + 1)}{2} - 670 = 181.$$

Utilizing this value, the estimate of heritability may be calculated. However, information has been obtained only on the females. Assuming that the same cock (or socially equivalent cocks) sired all the progeny, the ratio in (11) would estimate only the genetic effect of the hen, or $\frac{1}{2} \theta$; consequently f is equal to 2. Using (7),

$$\hat{p} = \frac{1}{575} (181) = .315$$

and by (15) the estimate of heritability

$$\hat{\theta} = (2)(2) \frac{25 + 23}{48} (.500 - .315) = .740.$$

In order to test whether this estimate is significantly different from zero, let the rejection level (α) be .10; thus, u_α will be $+ 1.28$. According to (18), the test of significance is to reject the hypothesis that $\theta = 0$ when H_r is greater than

$$(2) \quad \frac{1.28(25 + 23)}{48} \sqrt{\frac{25}{(25)(23)}} = .534.$$

Since the estimate ($\hat{\theta} = .740$) is greater than .534, it is significantly greater than zero at the .10 level of probability.

If one wishes to place 2-sided confidence limits on θ, the upper and lower limits of p must first be obtained. By (9) and (10),

$$p_L = \frac{2(.315) + .071 - \sqrt{(.630 + .071)^2 - 4(.099)(1 + .071)}}{2(1 + .071)} = .206,$$

and

$$p_U = \frac{2(.315) + .071 + \sqrt{(.630 + .071)^2 - 4(.099)(1 + .071)}}{2(1 + .071)} = .448.$$

Thus the 80 per cent (or greater) confidence limits for θ will be

$$\theta_L = 2(2) \frac{25 + 23}{48} (.500 - .448) = .208$$

and

$$\theta_U = 2(2) \frac{25 + 23}{48} (.500 - .206) = 1.176.$$

It is impossible, of course, for more than 100 per cent of the variability in a trait to be due to genetic causes; thus, the upper limit of θ is inordinately large from a genetic standpoint. However, it is to be recalled that an upper bound for σ_U^2 is used to determine the points p_U and p_L; therefore, these are *at least* 80 per cent confidence limits. Furthermore, information only on the female parent is included in this example; thus, only the dam's genetic effect, or $\frac{1}{2} \theta$, is estimated. Since any error in the estimate is doubled, the confidence interval will be larger than one based on an estimate which utilizes information on both parents of each individual.

Summary

A method is proposed for the estimation of heritability under the situation where the experimenter may not be able to assume a normal distribution for the variate being studied. It requires that the individuals in the parental and the F_1 generation can be ranked on the basis

of the character under study. The heritability of rank can then be estimated, the estimate tested for significance, and confidence limits placed on the parameter.

ACKNOWLEDGMENTS

REFERENCES

Birnbaum, Z. W. [1956]. On the use of the Mann-Whitney statistic. Proceedings of the Third Berkeley Symposium on Math. Stat. and Prob., Vol. 1, pp. 13–17. Univ. of Calif. Press.

Birnbaum, Z. W. and Klose, Orval M. [1957]. Bounds for the variance of the Mann-Whitney statistic. *Ann. Math. Stat. 28*:933.

Lehmann, E. L. [1950]. Notes on the theory of estimation. Mimeographed lectures, pp. 3–23, 3–24. University of California, Berkeley.

Lehmann, E. L. [1951]. Consistency and unbiasedness of certain nonparametric tests. *Ann. Math. Stat. 22*:165.

Mann, H. B. and Whitney, D. R. [1947]. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat. 18*:50.

van Dantzig, D. [1951]. On the consistency and the power of Wilcoxon's two-sample test. *Koninklijke Nederlandse Akademie Van Wetenschappen, Proceedings*, Series A, *54*:1.

Wilcoxon, Frank [1945]. Individual comparisons by ranking methods. *Biometrics Bull. 1*:80.