Classical and Bootstrap Estimates of Heritability of Milk Yield in Zimbabwean Holstein Cows

PRINCIPAL NDLOVU
Department of Statistics
The Pennsylvania State University
University Park 16802

ABSTRACT

The h^2 of first lactation milk yield in Zimbabwean grade Holstein cows was estimated as the regression of daughter on the dams' adjusted milk yield, using least squares, least absolute value, bootstrap restricted least squares, and bootstrap restricted least absolute value estimators. The estimators were evaluated for efficiency and robustness in estimating h^2. The estimated h^2 were .272, .265, .283, and .290 by least squares, least absolute value, bootstrap restricted least squares, and bootstrap restricted least absolute value procedures, respectively. The classic 90% confidence intervals for h^2 were .030 to .514 by least squares and .056 to .474 by least absolute value procedures. The bootstrap restricted least squares 90% confidence interval for h^2 was .000 to .542, and the bootstrap restricted least absolute value 90% confidence interval was .000 to .615. The point estimates by the four procedures were comparable, but the bootstrap confidence intervals were more conservative than their classic counterparts. The least absolute value estimator was the most efficient and the second best in robustness. The least squares estimator was second best in efficiency and worst in robustness. The bootstrap restricted least absolute value estimator was the least efficient and the second best in robustness.
(**Key words**: heritability, milk yield, least absolute value, bootstrap least squares)

Abbreviation key: **LAV** = least absolute value, **LS** = least squares, **N(0, σ^2)** = normal distribution with mean 0 and variance σ^2, **RLAV** = restricted least absolute value, **RLS** = restricted least squares, **ZMRS** = Zimbabwe Milk Recording Scheme.

INTRODUCTION

The h^2 of milk yield is commonly estimated from the analysis of paternal half-sib milk yield using REML. The REML estimator of h^2 is derived from the REML variance component estimators obtained from the analysis of the mixed linear models of the form given by Arnold (1). The REML variance component estimators are always nonnegative compared with their unbiased counterparts and are less biased than their maximum likelihood counterparts (13). Harville (13) describes some iterative algorithms for variance component estimation using the REML method. The iterations of the algorithms start with estimated values of the variance components. When computing resources are limiting and when no prior information about the approximate values of the variance components for the population is available, the REML estimate of h^2 is difficult to obtain. Under these circumstances, the offspring on parent regression may be a simpler alternative for obtaining some crude estimates of the h^2 of milk yield.

The regression estimator of h^2 of milk yield is given by Falconer (9). The estimator is obtained from the regression analysis of the milk yield after adjustment for all nuisance factor effects, provided that such milk yield adjustment factors are available for the population under study. The estimator has not been widely used because it tends to be biased by maternal effects. The REML estimator is free of this bias. Furthermore, the estimator has unbounded influence (11) and can take on

Received January 6, 1992.
Accepted December 21, 1992.

values outside the range of h^2. However, the estimator is easily transformed to a biased estimator with values in the range of h^2, and robust and semiparametric estimation techniques are easier to incorporate to the analysis of the regression model to obtain robust and semiparametric regression estimates of h^2 than to the analysis of the mixed model. Robust and semiparametric estimators are desirable in cases for which the errors in the observed milk yield deviate from the Gaussian error structure.

The objective of the paper is to show that, in situations for which REML estimate of h^2 is impossible too obtain, regression may be used to obtain classical, robust, and semiparametric estimators of h^2 if approximate milk yield adjustment factors for nuisance factor effects can be found. This situation is illustrated by the h^2 of milk yield that were estimated for the Zimbabwean Holstein cows. The methods are likely to be applicable for the estimation of h^2 in unselected or noninbred populations in which the bias from maternal effects is assumed to be not too large.

Theory Overview

When the h^2 of milk yield is to be estimated as the regression of the daughter on the dam's milk yield, in theory, the linear model (9)

$$Y_i = B_0 + B_1X_i + E_i,$$

where i = 1,2, . . ,n can be assumed to hold for the observed adjusted milk yield, where the (Y_i,X_i) are independently, identically distributed daughter-dam adjusted milk yield pairs, B_0 is the intercept, $B_1 = h^2/2$ is the slope, and the E_i are errors that include random environmental effects, each of which has a conditional normal distribution with mean zero and variance σ^2 ($\mathbf{N(0,\sigma^2)}$) given the X_i.

The least squares (**LS**) estimator of h^2 is

$$\hat{h}^2 = \frac{2\sum_{i=1}^{n}(Y_i - \overline{Y}.)(X_i - \overline{X}.)}{\sum_{i=1}^{n}(X_i - \overline{X}.)^2}, \quad [1]$$

with a standard error given by

$$s(\hat{h}^2) = 2\left[\frac{(n-2)^{-1}\sum_{i=1}^{n}(Y_i - \hat{B}_0 - \hat{B}_1X_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X}.)^2}\right]^{.5}, \quad [2]$$

where $\hat{B}_0$ and $\hat{B}_1$ are the LS estimators of B_0 and B_1, respectively. The inferences about h^2 are then based on the distribution of the pivotal quantity

$$H = \frac{(\hat{h}^2 - h^2)}{s(\hat{h}^2)}.$$

The estimator can take on values that are outside the range of h^2. Because of this, the restricted LS (**RLS**) estimator given by

$$\hat{\hat{h}}^2 = \begin{cases} 0 \text{ if } \hat{h}^2 < 0 \\ \hat{h}^2 \text{ if } 0 \le \hat{h}^2 \le 1 \\ 1 \text{ if } \hat{h}^2 > 1 \end{cases} \quad [3]$$

may be preferred. The distribution of $\hat{\hat{h}}^2$ is more complicated than that of $\hat{h}^2$.

Arnold (1) argued that, as long as the X_i have finite variance and the E_i are independently and identically distributed with mean zero and finite variance, then the inferences about h^2 based on the distribution of H or $\hat{\hat{h}}^2$, as specified by the Gaussian distribution assumption, are insensitive to the violation of the assumption if the sample size n is large. For small n, the inferences based on the distributions of H and $\hat{\hat{h}}^2$ derived, assuming a Gaussian error structure, are sensitive to the violation of the assumption. In certain situations, the data can be transformed to meet the Gaussian distribution assumption. Non-Gaussian errors in milk yield may occur because the error-generating distribution in the population is non-Gaussian (e.g., the Cauchy), or the errors may be Gaussian but contaminated because of errors arising from improperly recorded milk yield, inclusion of milk yield from another population, and from measurement errors. Extreme errors in the Y_i (outliers) lead to unstable LS estimates, and those errors in the X_i (influential errors) tend to induce large positive or negative biases in the estimates (11).

One alternative estimator of h^2 that is not too sensitive to data contamination is the least