Bio Vet Innovator Magazine
(Fueling The Future of Science...)
Volume 3 (Issue 5) MAY 2026
OPEN ACCESS
International Day for Biological Diversity – 22th May
Popular Article

Variance Components in Animal Breeding: Methods of Estimation

Shubham Kumar Singh[1*], Shalvi Srivastava[2]
1MVSc Scholar, Division of Animal Genetics and Breeding,
ICAR – National Dairy Research Institute, Karnal, 132001, Haryana, India
2MVSc Scholar, Division of Animal Nutrition,
ICAR – National Dairy Research Institute, Karnal, 132001, Haryana, India
***Corresponding Author:** drshubhsingh23@gmail.com

DOI: https://doi.org/10.5281/zenodo.20356014

Received: May 15, 2026
Published: May 20, 2026

Abstract:

The accurate estimation of variance components is a foundational step in modern animal breeding programs, as they are crucial for calculating the genetic parameters such as heritability and genetic correlations that underpin the prediction of breeding values. The core challenge lies in inferring the unobservable genetic merit (genotype) of an animal from its observed performance (phenotype). Variance components provide the essential, quantitative link between these two, allowing for the partitioning of the total phenotypic variance into its genetic and environmental components. Several statistical methods have been developed for estimation of variance components. The Analysis of Variance method, pioneered by R.A. Fisher, provides a straightforward approach by partitioning variance into between-group (e.g., between sires, representing genetics) and within-group (environmental) components. For more complex, real-world data, Henderson's methods, particularly his Method III which uses mixed model equations, became a significant step forward. Today, the gold standard is REML. While the Maximum Likelihood method finds the most probable values for the variance components, its estimates are statistically biased because it doesn't account for the fixed effects in the model. REML solves this by first filtering out these fixed effects, leading to unbiased and reliable estimates. Its ability to handle unbalanced and complex datasets makes it the preferred method in modern animal breeding programs. Other methods, like MINQUE & MIVQUE, offer unique properties but are less practical for routine use. In conclusion, the accurate estimation of variance components is the backbone of genetic evaluation. By quantifying the hidden sources of variation, breeders can make more informed decisions, ultimately leading to more efficient and sustainable genetic improvement in livestock populations.

Keywords: Variance Components, Animal Breeding, Genotype, REML, ANOVA, Henderson's Methods, MINQUE.

Introduction:

In the field of animal breeding, the central challenge is to make genetic improvements in livestock populations by selecting superior individuals to be parents of the next generation. The observable performance of an animal known as its phenotype (P), is a result of two fundamental forces: its genetic makeup (genotype G) and the environmental influences (environment E) it experiences. This relationship

Citation: Shubham Kumar Singh, Shalvi Srivastava. (2026). Variance Components in Animal Breeding: Methods of Estimation. In Bio Vet Innovator Magazine (Vol. 3, Issue 5, pp. 31–36). Bio Vet Innovator Magazine. https://doi.org/10.5281/zenodo.20356014

is classically expressed as, P = G + E. However, the core problem is the invisible genotype. We can only see the visible phenotype, which is a messy mix of genetics and environment. For instance, we can observe that Cow A produces 30 litres of milk per day while Cow B produces 25 litres. While it is clear that Cow A has a superior phenotype, the critical question for a breeder is: Why*?* Is this superiority due to genetically superior traits that can be passed to its offspring, or is it solely the result of a better diet, management, or climate? Relying on phenotype alone for selection can be misleading, as an animal with a mediocre genotype might perform well in an excellent environment, and vice-versa. To bridge this gap between the visible phenotype and the invisible genotype, breeders turn to the concept of variance components. Falconer and Mackay (1996) stated that when values are expressed as deviations from the population mean, variance represents the average of squared deviations and forms the cornerstone for genetic evaluation and heritability estimation. Variance component is a quantitative measure that partitions the total phenotypic variance of a trait into distinct parts attributable to different sources of variation. By partitioning the total phenotypic variance into its constituent parts, we can quantify the relative contributions of genetics and environment, enabling more accurate predictions of breeding values and more effective selection decisions.

Partitioning of Phenotypic Variance:

The total phenotypic variance is the starting point for all genetic analyses. The foundational step is to partition this total variance (V_P) into its two major components, Genotypic variance (V_G) & Environmental variance (V_E) i.e., $\mathbf{V_P = V_G + V_E}$. Genotypic variance is the portion of the total phenotypic variance in a population that is attributable to the differences in the **genotypic values** among individuals whereas environmental variance is the portion of the total phenotypic variance in a trait that is caused by non-genetic (environmental) factors affecting individuals

This genotypic variance is then further partitioned into additive genetic variance, dominance variance and epistatic variance as $\mathbf{V_G = V_A + V_D + V_I}$. Additive genetic variance is the variance due to the additive effects of genes—the sum of the individual effects of alleles that are passed from parents to offspring. It is the most crucial component for animal breeders, as it determines the response to selection. Dominance variance arises from interactions between alleles at the same locus (e.g., dominance) and epistatic or interaction variance results from interactions between genes at different loci.

Similarly, the environmental variance can be partitioned to account for different types of non-genetic effect such as permanent ($\mathbf{V_{PE}}$) and temporary environmental variances ($\mathbf{V_{TE}}$), $\mathbf{V_E = V_{PE} + V_{TE}}$. Permanent environmental variance represents non-genetic effects that permanently influence an individual's performance whereas temporary environmental variance represents temporary, non-repeatable effects. Therefore, the complete and detailed partition of the total phenotypic variance becomes: $\mathbf{V_P = V_A + V_D + V_I + V_{PE} + V_{TE}}$

Citation: Shubham Kumar Singh, Shalvi Srivastava. (2026). Variance Components in Animal Breeding: Methods of Estimation. In Bio Vet Innovator Magazine (Vol. 3, Issue 5, pp. 31–36). Bio Vet Innovator Magazine. https://doi.org/10.5281/zenodo.20356014

The primary goal of variance component estimation in animal breeding is to accurately estimate these components, particularly V_A, as it allows for the calculation of key parameters like heritability ($h^2 = V_A / V_P$), which measures the proportion of phenotypic variance due to additive genetic effects likely to be transferred to the next generation.

Why Variance Components are Needed ?

Variance components are not merely abstract statistical concepts. They are essential in animal breeding because they form the quantitative link between phenotype and genotype. While phenotypes are directly observable, genotypes are not. Hence, breeders rely on variance component estimation to infer genetic potential from observed performance.

- **Predicting Breeding Values:** The accuracy of estimating an animal's genetic merit (its Breeding Value) depends entirely on the estimates of additive genetic variance. These values are used in mixed model equations to compute Best Linear Unbiased Predictions (BLUP).
- **Genetic Evaluation:** National genetic evaluation programs for dairy, beef, swine, and poultry rely on accurate variance components to rank animals across different herds and environments.
- **Informing Selection Decisions:** Knowing the heritability of a trait tells a breeder how responsive that trait is likely to be to selection. Low heritability traits require different selection strategies (e.g., family selection) compared to high heritability traits.
- **Understanding Trait Architecture:** The relative magnitudes of additive genetic variance, dominance & interaction variance provide insights into the genetic architecture of a trait, which is valuable for both breeders and geneticists.

Methods of Estimating Variance Components:

A variety of statistical methods have been developed to estimate variance components. The choice of method depends on the complexity of the data, the model, and computational resources.

1. Analysis of Variance (ANOVA) Method:

One of the earliest and most straightforward methods is the Analysis of Variance (ANOVA). R.A. Fisher published his first major application of ANOVA in 1921, and the technique became widely known after being included in his book, *Statistical Methods for Research Workers*, published in 1925. It is particularly useful for balanced data from simple experimental designs, such as half-sib families where each sire is mated to several dams. In this design, the total variance is partitioned into "Between Sire" and "Within Sire" components. The key principle is that sires pass on half of their genes to each offspring. Therefore, the similarity between half-sibs is a function of the sire's additive genetic value.

The variances are estimated as Environmental Variance (σ^2e) = Mean square within sires; Genetic Variance (σ^2s) = Difference between Mean square between sires & mean square within sire divided by K (number of progenies per sire).

Citation: Shubham Kumar Singh, Shalvi Srivastava. (2026). Variance Components in Animal Breeding: Methods of Estimation. In Bio Vet Innovator Magazine (Vol. 3, Issue 5, pp. 31–36). Bio Vet Innovator Magazine. https://doi.org/10.5281/zenodo.20356014

Advantages of this method is that, it is simple, computationally easy for balanced data and provides a clear, intuitive understanding of variance partitioning. Disadvantages is that, it is not suitable for complex pedigrees or unbalanced data common in field records. It cannot easily account for multiple fixed effects (e.g., herd, year, season) & can produce negative variance estimates, which are biologically nonsensical.

2. Henderson's Methods:

C.R. Henderson in 1953, developed a series of three methods that extended variance component estimation to more complex, unbalanced data sets using principles of least squares.

❖ Henderson's Method I:

This method is essentially the ANOVA approach applied to unbalanced data. It uses the sums of squares from fitting a least-squares model with fixed effects to equate to their expected values and solve for the variance components. It is unbiased but can be inefficient for complex models.

❖ Henderson's Method II:

Method II was an attempt to improve upon Method I by adjusting for the fixed effects in the model before estimating the variance components. However, it has significant demerits like it cannot effectively analyse models that include interactions between fixed and random factors. Also, it can produce negative estimates of variance components.

❖ Henderson's Method III:

This is the most widely used of Henderson's methods and is also known as the "fitting constants method." It is based on using reductions in sums of squares due to fitting sub-models. The general linear mixed model is:

$$\mathbf{Y = X\beta + Zu + e}$$

where, Y is the vector of observations; β is the vector of fixed effects; u is the vector of random effects; X and Z are incidence matrices for fixed & random effects respectively; **e** is the vector of random residuals.

The method involves "absorbing" the fixed effects (using a projection matrix $\mathbf{M = I - X(X'X)^{-1}X'}$) to reduce the equations to $\mathbf{Z'MZ\,\hat{u} = Z'MY}$. The solutions from these equations are then used to compute sums of squares, which are equated to their expectations to solve for the variance components.

Merits include its ability to handle unbalanced data and provide unbiased estimates. However, demerits are that it is not a single, uniform procedure and can be computationally intensive; furthermore, it can still yield negative estimates..

3. Maximum Likelihood (ML) Methods

Maximum Likelihood was introduced in its modern form by R. A. Fisher in 1922. This method represent a paradigm shift from the methods of moments (like ANOVA and Henderson's methods) to a more general and powerful framework.

Citation: Shubham Kumar Singh, Shalvi Srivastava. (2026). Variance Components in Animal Breeding: Methods of Estimation. In Bio Vet Innovator Magazine (Vol. 3, Issue 5, pp. 31–36). Bio Vet Innovator Magazine. https://doi.org/10.5281/zenodo.20356014

Principle of Maximum Likelihood (ML):

The ML procedure finds the values of the parameters (VC) that are most likely, given the observed data. It assumes a specific distribution for the data, typically the normal distribution. A likelihood function L(β, σ^2A, σ^2e | Y) is formulated, which expresses the probability of observing the data Y given specific values for the parameters.

The goal is to find the parameter values that maximize this likelihood function. This is typically achieved through complex, iterative computational algorithms (e.g., Newton-Raphson, EM algorithm) that systematically try different values for σ^2A and σ^2e until the combination that gives the highest possible value of L is found.

The primary advantage of this approach is its high degree of generality, which allows it to be applied to highly complex models and diverse data structures. Furthermore, it produces estimates characterized by strong statistical properties, such as consistency and efficiency. However, a significant drawback is that the resulting estimates are biased downward. This occurs because ML fails to account for the degrees of freedom lost when estimating the fixed effects (β). This bias can become particularly substantial in cases where the number of fixed effects is large relative to the overall sample size.

4. Restricted Maximum Likelihood (REML):

REML was officially introduced and formalized in 1971 by Patterson and Thompson. It was developed to correct the bias inherent in the standard ML method. Instead of maximizing the likelihood based on the original data, REML maximizes the likelihood of linearly independent error contrasts—transformations of the data that are invariant to the fixed effects. In essence, REML filters out the fixed effects before estimating the variance components. This correction for the loss of degrees of freedom makes REML the preferred and most widely used method in modern animal breeding for variance component estimation.

REML offers several significant advantages, most notably providing unbiased estimates of variance components. It is a highly flexible approach capable of handling complex, unbalanced data and sophisticated model structures, while generally ensuring that variance estimates remain non-negative. However, a primary disadvantage is that the procedure is computationally more demanding than traditional methods of moments, though this has become less of a practical concern with the advent of modern computing power.

5. MINQUE & MIVQUE:

MINQUE & MIVQUE are another approach developed by C.R. Rao in 1970-71.

- **Quadratic:** The estimator is a quadratic function of the observations.
- **Unbiased:** The expected value of the estimator equals the true parameter value.
- **Minimum Norm:** Among all quadratic unbiased estimators, it has the smallest possible variance (or norm).

Citation: Shubham Kumar Singh, Shalvi Srivastava. (2026). Variance Components in Animal Breeding: Methods of Estimation. In Bio Vet Innovator Magazine (Vol. 3, Issue 5, pp. 31–36). Bio Vet Innovator Magazine. https://doi.org/10.5281/zenodo.20356014

A key feature of MIVQUE is that it does not require an assumption of normality. However, to achieve the "minimum variance" property, it requires that the true values of the variance components be known in advance to construct an optimal weighting matrix. Since the true values are unknown, prior guesses (or estimates from a previous analysis) must be used.

When these prior values are chosen optimally, the method is referred to as MIVQUE (Minimum Variance Quadratic Unbiased Estimation). In practice, MIVQUE is MINQUE with an optimal prior specification. If the prior values are correct, MIVQUE provides the best quadratic unbiased estimates. However, the dependence on prior values is a significant drawback, making REML a more robust and commonly preferred choice.

Conclusion:

The accurate estimation of variance components is a cornerstone of modern animal breeding. It allows us to quantitatively dissect the total phenotypic variation into its genetic and environmental sources, providing the essential link between the observable phenotype and the unobservable genotype.

The journey of estimation methods has evolved from simple, intuitive techniques like ANOVA for balanced data to more robust methods like Henderson's Method III for unbalanced data. The field has now largely converged on likelihood-based approaches, with REML standing out as the gold standard. REML's ability to provide unbiased estimates from complex, unbalanced datasets—common in field records—while correctly accounting for the loss of degrees of freedom from fixed effects, makes it the preferred method for national genetic evaluations and research worldwide.

While other methods like MINQUE/MIVQUE have their theoretical merits, REML strikes the best balance between statistical properties, practical applicability, and robustness. The continuous refinement of these methods ensures that animal breeders can make increasingly accurate predictions of breeding values, leading to more efficient and sustainable genetic improvement of livestock populations.

References:

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Longman.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics, 9*(2), 226–252.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Associates.

Mrode, R. A. (2014). *Linear models for the prediction of animal breeding values* (3rd ed.). CABI.

Rao, C. R. (1971). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis, 1*(4), 445–456.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. Wiley.

Citation: Shubham Kumar Singh, Shalvi Srivastava. (2026). Variance Components in Animal Breeding: Methods of Estimation. In Bio Vet Innovator Magazine (Vol. 3, Issue 5, pp. 31–36). Bio Vet Innovator Magazine. https://doi.org/10.5281/zenodo.20356014