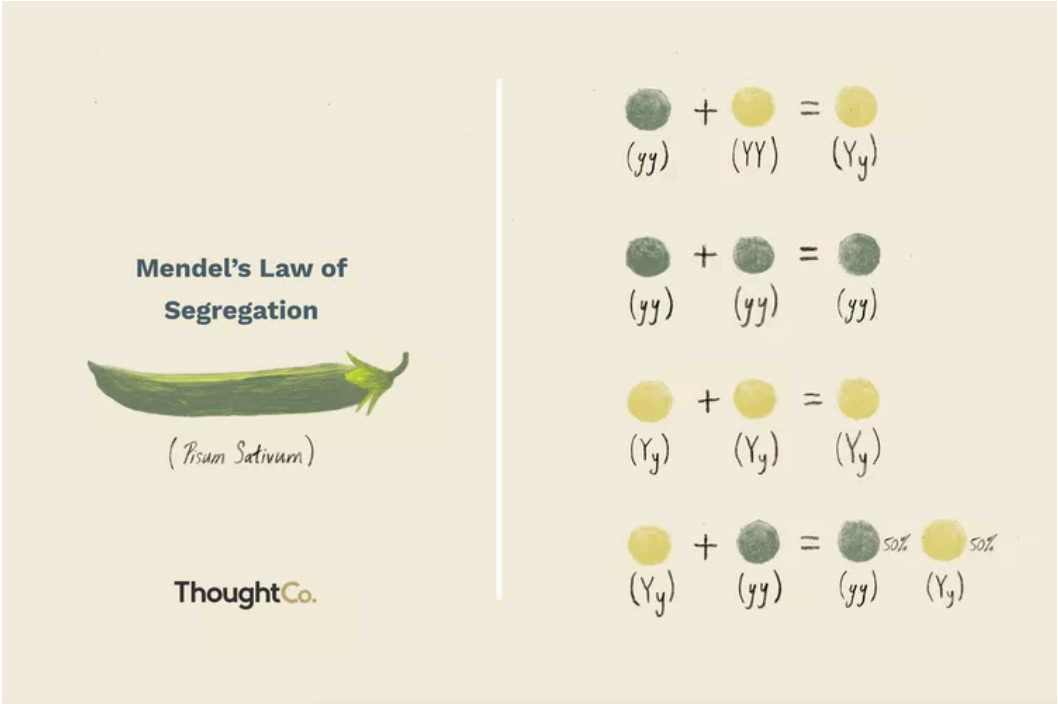
1. Mendelian Genetics

1. Mendel's First Law

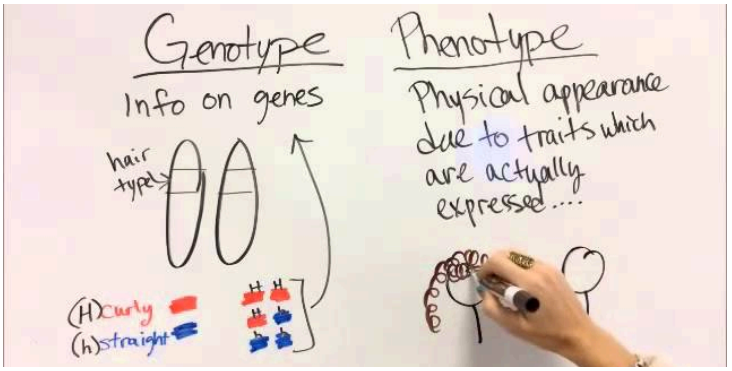
Mendel's First Law (also called the **Law of Segregation**) states that each organism has two alleles for each gene (one from each parent), and these alleles segregate (separate) during the formation of gametes. This means that each gamete will carry only one allele for a particular gene, and the offspring will inherit one allele from each parent.

Mendelian Genetics is the foundation of classical genetics, based on the work of Gregor Mendel in the 19th century. He discovered how traits are inherited through **dominant** and **recessive** alleles. Key principles include:

- **Law of Segregation**: Each individual has two alleles for each gene (one from each parent), and these alleles segregate during gamete formation. Each gamete receives one allele.



Genotype vs. Phenotype: The genotype is the genetic constitution (the specific alleles an individual carries), while the phenotype is the physical expression of the genotype (traits like eye color, flower color, etc.).

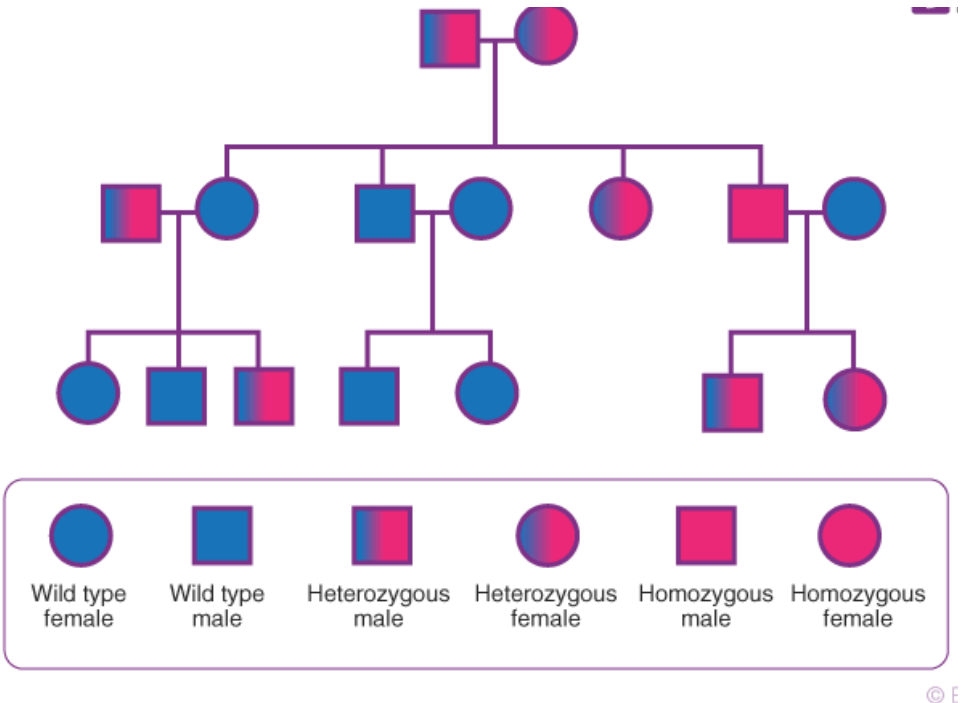


2. Variations to Mendel's First Law

While Mendel's First Law generally holds true, there are **exceptions** where genes may not behave in the simple dominant/recessive way he proposed. These exceptions include **incomplete dominance** (when the heterozygote shows an intermediate phenotype), **codominance** (where both alleles are expressed equally), and **multiple alleles** (where there are more than two possible alleles for a gene).

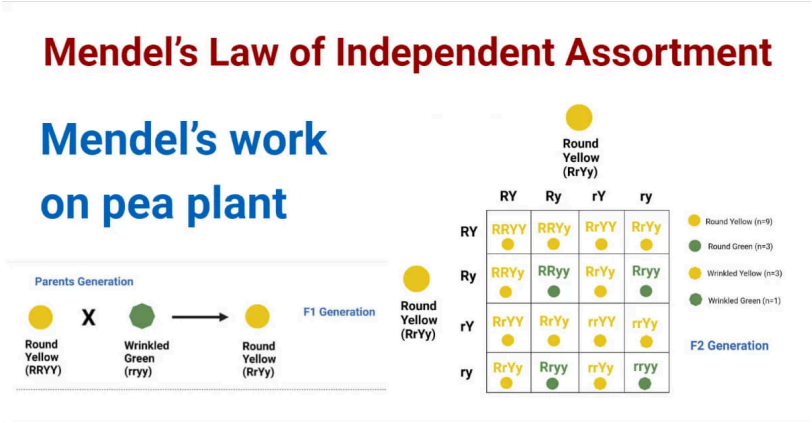
3. Pedigree Analysis

Pedigree analysis is a family tree diagram used to track the inheritance of traits across generations. It helps determine whether a genetic trait is dominant, recessive, or X-linked, and can help identify carriers of recessive genetic disorders.

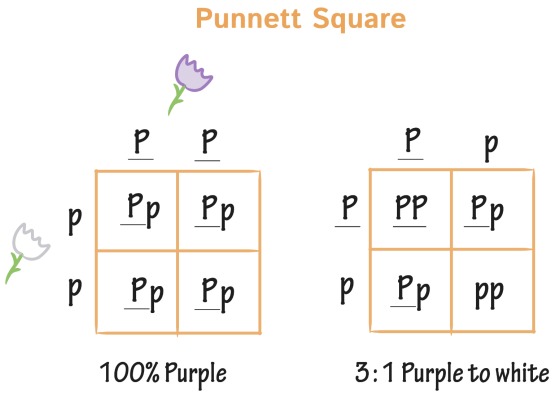


4. Mendel's Second Law

- **Law of Independent Assortment**: Genes for different traits are inherited independently of each other, provided they are on different chromosomes.



Mendelian genetics is often explained using **Punnett squares**, a diagram used in genetics to predict the probability of different genotypes and phenotypes in offspring. It's a visual representation of how alleles combine during fertilization.



The gene that controls flower color has two alleles: purple and white.
- ✓ P = dominant purple allele
- ✓ p = recessive white allele

Phenotype = purple
Genotype = PP or Pp

Phenotype = white
Genotype = pp

Example:

- **Monohybrid cross**: A cross between two individuals with a single trait being studied, like plant height (tall vs. short).
- **Dihybrid cross**: A cross involving two traits, such as seed shape and color.

Monohybrid cross

Mother is heterozygous for a particular trait (*Aa*).

Father is also heterozygous for the same trait (*Aa*).

Homozygous dominant (*AA*) = 1/4
Heterozygous (*Aa*) = 1/2
Homozygous recessive (*aa*) = 1/4

♀ \ ♂	*A*	*a*
A	*AA*	*Aa*
a	*Aa*	*aa*

Dihybrid cross (gene linkage)

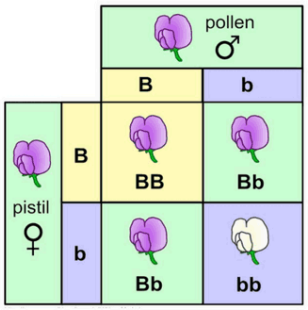
A and *a* represent one trait, and *B* and *b* represent a different trait that is linked to inheritance of *A* or *a*.

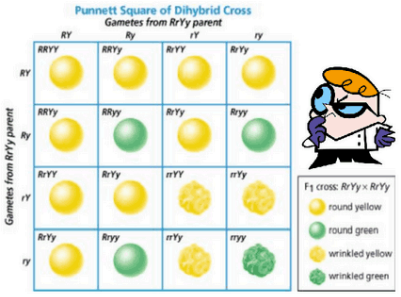
	AB	*Ab*	*aB*	*ab*
AB	*AABB*	*AABb*	*AaBB*	*AaBb*
Ab	*AABb*	*AAbb*	*AaBb*	*Aabb*
aB	*AaBB*	*AaBb*	*aaBB*	*aaBb*
ab	*AaBb*	*Aabb*	*aaBb*	*aabb*

Dominant for *A* and *B* = 9/16
Recessive for *a*, dominant for *B* = 3/16
Dominant for *A*, recessive for *b* = 3/16
Recessive for *a*, recessive for *b* = 1/16

Monohybrid vs. Dihybrid Crosses

AP BIO





1:2:1 and 9:3:3:1 Phenotypic Ratio

- The **9:3:3:1 ratio** shows how traits are inherited independently, which is **Mendel's Law of Independent Assortment**.
- The inheritance of one trait (like plant height) does **not** affect the inheritance of the other trait (like flower color).

- **QTL (Quantitative Trait Loci)** are genomic regions associated with variation in a quantitative trait — like plant height, yield, or weight. Unlike simple Mendelian traits (like flower color or pea shape), quantitative traits are:
- Influenced by multiple genes (polygenic)
  Affected by gene-gene interactions (epistasis)
  Often influenced by environmental factors

5. Chi-Square Test

The **Chi-square test** is used to compare observed genetic ratios in a population (such as the results of a genetic cross) with the expected ratios based on Mendelian inheritance. It helps determine if the differences are due to random chance or if some other factor is influencing the results.

The **Chi-Square Test** is used to determine if your observed data matches an expected Mendelian ratio (like 9:3:3:1). It helps you figure out whether the differences between observed and expected outcomes are due to chance or some other factor.

Steps to Perform a Chi-Square Test:

1. **Calculate Expected Values**:
    - First, you need to calculate the **expected values** for each group in your experiment. This is done by multiplying the total number of observations by the ratio expected by Mendelian inheritance.
      We're checking if the observed seed traits fit a **9:3:3:1 ratio** (as expected from Mendel's dihybrid cross).

Step 2: List observed and expected values

Trait	Observed	Expected (based on 9:3:3:1 ratio and total = 556)
Round, Yellow	315	(9/16) × 556 = 312.75
Round, Green	108	(3/16) × 556 = 104.25
Wrinkled, Yellow	101	(3/16) × 556 = 104.25
Wrinkled, Green	32	(1/16) × 556 = 34.75

Total	**556**	**556**

Step 3: Calculate degrees of freedom (df)

- Number of categories (**n**) = 4
- **df = n - 1 = 4 - 1 = 3**

Step 4: Calculate chi-square value (χ^2)

Use the formula for each category:
(Observed - Expected)² / Expected

Then add them all:

- Round, Yellow: (315 - 312.75)² / 312.75 = 0.016
- Round, Green: (108 - 104.25)² / 104.25 = 0.135
- Wrinkled, Yellow: (101 - 104.25)² / 104.25 = 0.101
- Wrinkled, Green: (32 - 34.75)² / 34.75 = 0.218

Total χ^2 = 0.016 + 0.135 + 0.101 + 0.218 = 0.47 (Expected)

Step 5: Compare with chi-square table

Look at the **df = 3** row in the chi-square table:

A Chi-Square Table

	Probability				
Degrees of Freedom	**0.9**	**0.5**	**0.1**	**0.05**	**0.01**
1	0.02	0.46	2.71	3.84	6.64
2	0.21	1.39	4.61	5.99	9.21
3	0.58	2.37	6.25	7.82	11.35
4	1.06	3.36	7.78	9.49	13.28
5	1.61	4.35	9.24	11.07	15.09

In statistical hypothesis testing, a p-value of 0.05 (or less) is typically taken as the threshold for declaring a result as statistically significant. This means there's a 5% or less chance of observing the results if the null hypothesis is true. If the p-value is less than or equal to 0.05, the null hypothesis is rejected, and the alternative hypothesis is favored.

p-value	0.9	0.5	0.1	0.05	0.01
χ^2	0.58	2.37	6.25	**7.82**	11.35

Our value **0.47 (expected) < 7.82 (observed)**, so it is **less** than the critical value at p = 0.05.

Step 6: Make a conclusion

Because **0.47 is less than 7.82**, we **accept the null hypothesis**. **Null Hypothesis:**

"There is no significant difference between the observed and expected results."
✅ This means the observed results **do fit** the 9:3:3:1 expected ratio.

    - By convention, we use **0.05** as a cutoff (meaning a 5% chance that the difference is due to random chance). If your Chi-Square value is **greater** than the table value for 0.05, you **reject** the hypothesis. If it's **less**, you **accept** the hypothesis.

- Use the **Chi-Square test** to see if your observed data matches the expected Mendelian ratios.
  If the Chi-Square value is **below** the critical value from the table (at the 0.05 level), the data **fits** the expected ratio.
  If the Chi-Square value is **above** the critical value, you **reject** the hypothesis.

Gene Interactions

Genes don't work in isolation—they interact in the same cell, influencing traits together.

1. Epistasis

It's when **one gene masks or modifies the effect of another gene**.

Types of Epistasis (with examples and simplified ratios):

A. Duplicate Gene Action (15:1 Ratio)

- **Example**: Wheat kernel color
  Explanation: Either gene A or B can produce the colored kernel. Only double recessive (aabb) gives white.

Phenotype Ratio: 15 colored : 1 white

B. Complementary Gene Action (9:7 Ratio)

- **Example**: Sweet pea flower color
  Explanation: Both genes (C and P) must be present for color. One defective gene = no color.
  Phenotype Ratio: 9 colored : 7 white

C. Dominant Epistasis (12:3:1 Ratio)

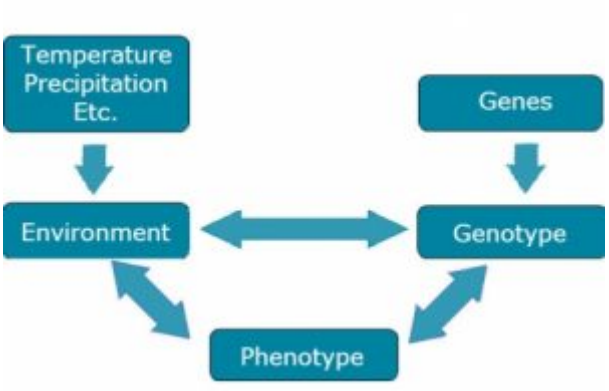
- **Example**: Squash fruit color
  Explanation: Dominant W allele blocks color expression. Yellow/green show only if ww.
  Phenotype Ratio: 12 white : 3 yellow : 1 green

D. Dominant Suppression (13:3 Ratio)

- **Example**: Malvidin pigment in Primula
  Explanation: Dominant D suppresses pigment gene K. Only K_dd shows color.
  Phenotype Ratio: 13 no color : 3 color

Key Terms:

- **Epistasis**: Interaction between genes affecting a single trait.
  Suppressor Gene: Prevents another gene from expressing.

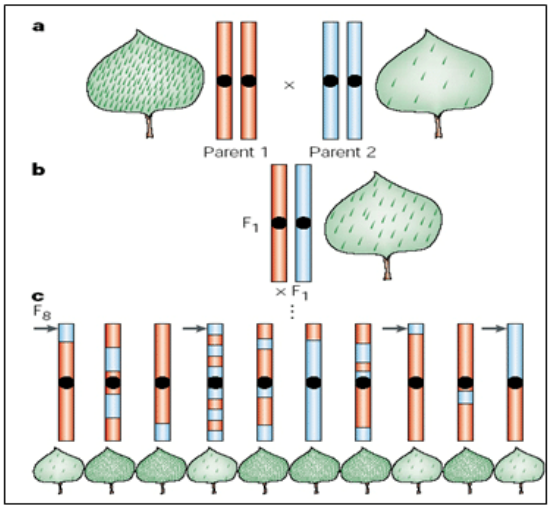


Over the years, plant breeders have developed many statistical models to study **genotype × environment interaction (GE)**. These models help assess how stable important traits are and predict how new genotypes will perform in different environments.

In the past decade, using a small number of molecular markers made it possible to identify chromosome regions tied to trait variation—this is called **QTL mapping**. QTL mapping has

helped with **marker-assisted selection** for simple traits but has been less effective for complex traits that involve many genes.

Recently, the availability of cheaper and more numerous markers allows us to densely cover the genome and use this data to **predict genomic breeding values**. This increases the accuracy of genetic value predictions compared to traditional pedigree-based methods. Genomic data can also help us understand how gene effects vary across environments.



🔬 QTL Mapping – Simplified

1. **Start with two parent plants** that differ in a trait — for example, one has **lots of trichomes (tiny hairs on leaves)**, and the other has **few trichomes**.

2. **Cross them** to make F_1 plants, which show a mix (medium number of trichomes).

3. **Self the F_1** to produce an F_2 generation, and then self those for **six generations** to get **Recombinant Inbred Lines (RILs)**. Each RIL has a unique mix of the parent genes but is genetically stable.

4. Test each RIL for:

    - **DNA markers** (which parts came from which parent)
      The trait (trichome density)

If RILs that inherited a specific chromosome segment from the **low-trichome parent** also have **low trichome density**, then that segment likely contains a **QTL (Quantitative Trait Locus)** controlling the trait.

In plant breeding, **genotype × environment interaction (GE)** is common in multi-environment trials (MET). It appears either as changes in genotype rankings across environments or as differences in how much genotypes vary without a rank change.

Here's a simple explanation with examples to help you understand Genotype × Environment interaction (GE) and how linear, bilinear, and linear-bilinear models help analyze it in multi-environment trials (MET):

🌱 What is GE interaction?

In plant breeding, a genotype may perform differently in different environments.

Genotype	Environment 1 (Yield)	Environment 2 (Yield)
A	4.5 tons/ha	3.0 tons/ha
B	3.8 tons/ha	4.2 tons/ha

Genotype A performs better in Env 1
Genotype B performs better in Env 2

🔹 Linear Models

These assume the effect of genotype and environment are additive (no interaction).

Example:

Yield = Genotype effect + Environment effect

But in real life, this is too simple and often doesn't capture changes in ranking of genotypes across environments.

🔹 Bilinear Models

These model the interaction between genotypes and environments using multiplicative terms (like PCA).
A famous example: AMMI (Additive Main effects and Multiplicative Interaction)

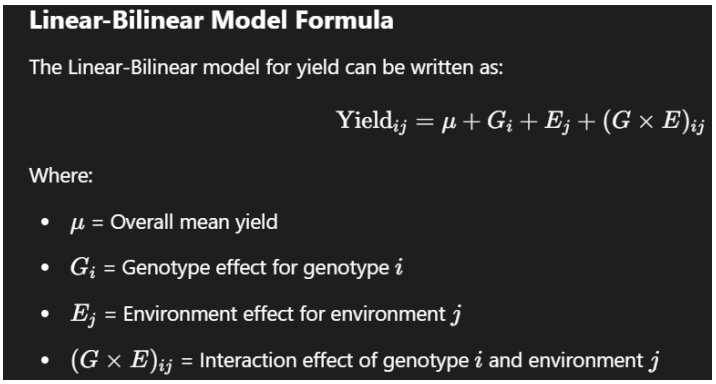
If Genotype A is specially adapted to a drought-prone environment and B to high-rainfall, a bilinear model helps explain why A is better in some places and B in others — it's not just additive.

Sample Data for Linear-Bilinear Model (Additive + Multiplicative Effects)

Consider a scenario where you are studying the yield of 3 genotypes (G1, G2, G3) in 4 environments (E1, E2, E3, E4). Here's a sample dataset:

Genotype \ Environment	E1	E2	E3	E4
G1	8.0	7.0	5.0	6.5
G2	6.0	6.5	8.5	7.5
G3	7.5	8.0	6.0	6.0

Linear-Bilinear Model Formula



AMMI Analysis in R

In R, you can perform AMMI analysis to analyze the additive main effects and multiplicative interaction components. AMMI combines ANOVA (Additive effects) with Principal Component Analysis (PCA) for interaction terms.

Code Example:

AMMI and GGE Biplot Analysis:

1. Stable Genotypes: From both the AMMI and GGE biplots, you can determine which genotypes perform consistently well across environments. For instance, if Genotype G1 lies close to the origin on both plots, it can be classified as stable across all

environments.

2. Best Genotypes for Specific Environments: You can identify which genotype performs best in each specific environment. For example, Genotype G2 may perform the best in Environment E3, as visualized in the GGE Biplot.

3. Interaction Effects: The interaction component in both methods helps to understand how genotype and environment affect each other, revealing if certain genotypes respond differently to changes in environmental conditions.

How to Use in Plant Breeding:

- Selection of Stable Genotypes: Using AMMI and GGE Biplots, breeders can select genotypes that are stable across diverse environments or target specific environments for different genotypes.

- Environment-specific Adaptation: The analysis helps in breeding environment-specific genotypes where performance is maximized for that environment.

More recently, **linear mixed models** have become the preferred choice in breeding trials. They offer advantages like:

- Handling missing data
  Modeling error variance and spatial correlations
  Flexibly estimating genetic relationships using pedigree or marker data
  Clustering environments and genotypes to reduce **crossover interactions (COI)**

In terms of **predictive accuracy**, studies like **Cornelius and Crossa** used cross-validation to compare fixed-effect models. They showed that **shrinkage-based linear-bilinear models** often perform as well or better than traditional **BLUP**. However, most of this research only looked at prediction within environments, not across different environments.

🔍 What Is Genomic BLUP?

Genomic BLUP (GBLUP) is a statistical method used in **genomic selection** to predict how good an individual (like a plant hybrid or animal) will be based on its DNA (SNP markers). It replaces traditional pedigree data with a **genomic relationship matrix**.

What does BLUP stand for?

- **Best**: Among all predictors, BLUP gives the **lowest prediction error** (difference between true and predicted breeding values).

- **Linear**: The predictions are made using **linear combinations** of the data.

- **Unbiased**: On average, the predicted breeding value equals the true breeding value.
- **Prediction**: It estimates **random effects**, like the animal's breeding value, not just fixed effects like sex or diet.

Model Setup

The mixed model looks like:

y = Xb + Za + e

This says: the phenotype = fixed effects + genetic effects + residuals

---

This function solves the **Mixed Model Equations (MME)** for a basic animal model with:

- **Fixed effects** (`X`)
  Random animal effects (`Z`, `A`)
  Phenotype vector (`y`)
  Ratio of variances (`alpha = σ²_e / σ²_a`)
- **Objective**: We want to estimate the **Best Linear Unbiased Estimates (BLUPs)** and **Estimated Breeding Values (EBVs)** for each calf. These values will help us understand the genetic potential of each calf for the trait of interest (e.g., weight, milk yield, or growth rate).

1. **What are we predicting?**:
   - **Breeding Value (EBV)**: The EBV represents the genetic contribution of an individual calf to the trait being studied. It quantifies the calf's genetic potential based on its genetic effects and how it differs from the population average. This is what we are primarily trying to predict.
   - **Phenotype**: We are also using the **phenotype** (observed trait value, e.g., weight or growth) as an outcome, which we want to model using a combination of the fixed environmental effects and the random genetic effects. In the model, the phenotype is treated as a combination of the environmental influences (fixed effects) and the genetic contributions (random effects).
2. **Modeling Goal**:
   - By fitting a **mixed-effects model** (in this case, using the `lmer` function), we are predicting the genetic contribution of each calf to the observed phenotype, while accounting for the fixed environmental effects (like feeding or weather conditions).

- The random genetic effects (represented by **calf_id**) capture the calf's unique genetic traits. The model estimates how each calf deviates from the average population level for the trait of interest due to its genetics.

3. **BLUP**: This is a **prediction** of the calf's genetic merit, accounting for both the fixed environmental effects and the random genetic effects. BLUPs are particularly useful for predicting the future performance of calves or their potential as breeding animals. In essence, we are trying to **predict the genetic potential** of each calf for a specific trait (based on observed data) and determine how much of that potential is due to its genetics versus environmental factors. This prediction helps in **breeding decisions** and **genetic selection** to improve traits in future generations.

	calf_id	fixed_effect	genetic_effect	phenotype
1	1	3.2	1.1	4.5
2	2	3.2	-0.5	2.8
3	3	3.2	0.3	3.5
4	4	3.2	0.8	4.0
5	5	3.2	-0.1	3.0
6	6	3.2	1.2	4.1
7	7	3.2	-0.6	2.9
8	8	3.2	0.5	3.7
9	9	3.2	-0.2	3.1
10	10	3.2	0.7	4.2

$$\text{Phenotype} = \text{Fixed Effect} + \text{Random Effect} + \text{Residual Error}$$

So, for this specific example:

$$4.5 = 3.2 + 1.1 + \text{Residual Error}$$

Calf	Sex	WWG
4	male	4.5
5	female	2.9
6	female	3.9
7	male	3.5
8	male	5.0

And the cooresponding pedigree would be

Calf	Sire	Dam
1	0	0
2	0	0
3	0	0
4	1	0
5	3	2
6	1	2
7	4	5
8	3	6

Formula for WWG (Daily Gain):

$$\text{WWG} = \frac{\text{Weaning Weight} - \text{Birth Weight}}{\text{Number of Days from Birth to Weaning}}$$

If the trait being evaluated is **WWG** (Weaning Weight Gain), the interpretation of the **EBVs (Best Linear Unbiased Estimates)** would focus on how each calf's genetic potential for weight gain compares to the average in the population.

Here's how to interpret the example:

Interpretation of EBVs for WWG:

1. **EBV = -0.0760 (Calf 3)**:
    - This indicates that **Calf 3** has a **genetic potential for weight gain** that is **slightly below average** when compared to the population's average WWG.
    - **Negative EBV** suggests that this calf is expected to have a slightly **lower than average** weaning weight gain based on its genetics, even though environmental factors (such as feeding, care, etc.) are kept constant.
    - The value of -0.0760 reflects the calf's deviation from the population mean for WWG. The closer the EBV is to 0, the closer the calf's performance is to the population average.
2. **EBV = 0.0706 (Calf 4)**:
    - This indicates that **Calf 4** has a **genetic potential for weight gain** that is **slightly above average** when compared to the population's average WWG.
    - **Positive EBV** suggests that this calf is expected to have a **higher than average** weaning weight gain due to its genetic traits, again assuming environmental factors are held constant.
    - The value of 0.0706 means that this calf is genetically predisposed to gain more weight by weaning compared to the average of the population.

General EBV Interpretation for WWG:

- **EBV for WWG** tells you how much more (or less) weight gain a calf is expected to have compared to the population average, based on its genetic makeup.
- A **positive EBV** means the calf has genetics that favor **greater weaning weight gain**, and a **negative EBV** suggests the calf may have **genetics that result in lower weight gain**.

- The **larger the absolute value of the EBV**, the greater the deviation from the average weight gain.

For instance:

- **EBV of 0.0706 (Calf 4)** means Calf 4 is expected to gain **more** weight at weaning compared to the average calf.

- **EBV of -0.0760 (Calf 3)** means Calf 3 is expected to gain **slightly less** weight at weaning compared to the average calf.

In practice, these EBVs help in selecting animals with the best genetic potential for desired traits, in this case, for increasing WWG (which is a desirable trait in livestock breeding for improving growth rates).