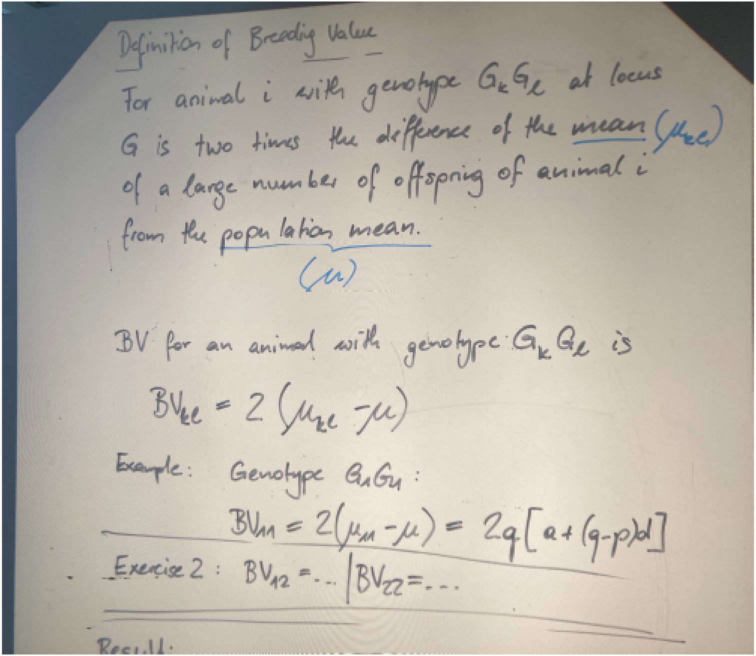
Overhead Pictures

Peter von Rohr

10/2/2020

Recap Definition



Definition of Breeding Value

For animal i with genotype G_kG_l at locus G is two times the difference of the mean (μ_{kl}) of a large number of offspring of animal i from the $\underbrace{\text{population mean.}}_{(\mu)}$

BV for an animal with genotype G_kG_l is

$$BV_{kl} = 2(\mu_{kl} - \mu)$$

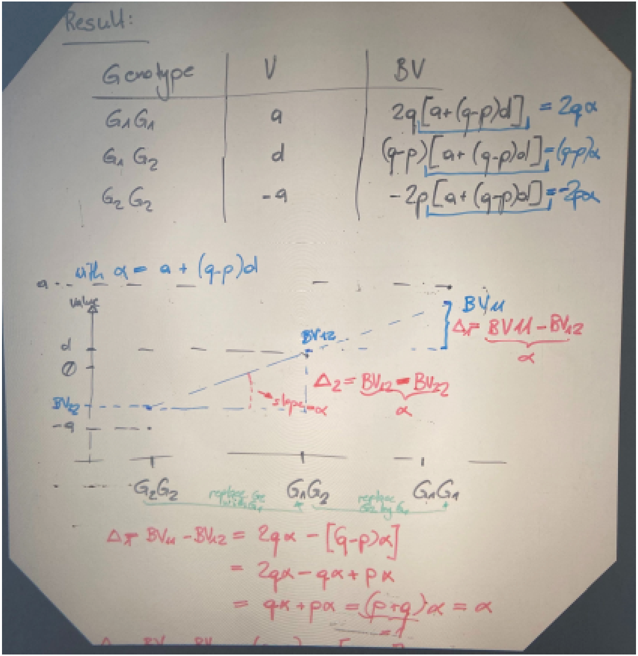
Example: Genotype G_1G_1:

$$BV_{11} = 2(\mu_{11} - \mu) = 2q[a + (q-p)d]$$

Exercise 2: $BV_{12} = \ldots$ | $BV_{22} = \ldots$

Result:

Result



Result:

Genotype	V	BV
G_1G_1	a	$2q\underbrace{[a+(q-p)d]} = 2q\alpha$
G_1G_2	d	$(q-p)\underbrace{[a+(q-p)d]} = (q-p)\alpha$
G_2G_2	$-a$	$-2p\underbrace{[a+(q-p)d]} = -2p\alpha$

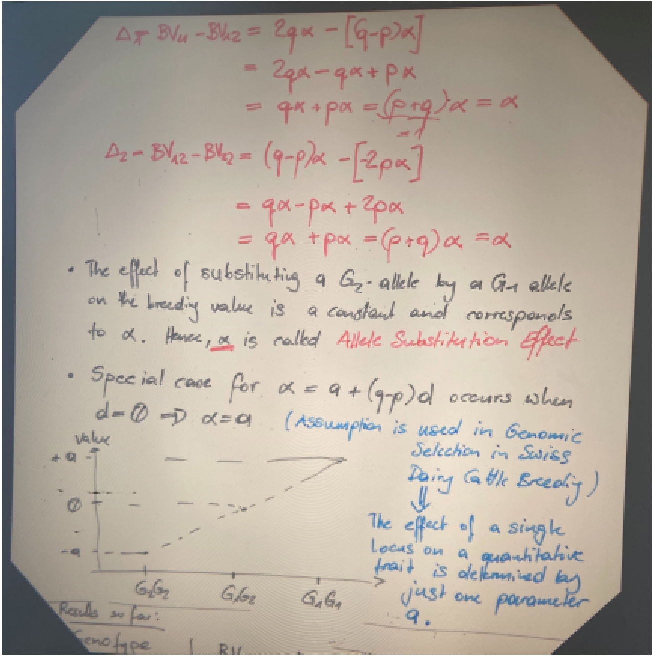
with $\alpha = a + (q-p)d$

$$\Delta_1 = \underbrace{BV_{11} - BV_{12}}_{\alpha}$$

$$\Delta_2 = \underbrace{BV_{12} - BV_{22}}_{\alpha}$$

$$\begin{aligned}\Delta_1 = BV_{11} - BV_{12} &= 2q\alpha - [(q-p)\alpha] \\ &= 2q\alpha - q\alpha + p\alpha \\ &= q\alpha + p\alpha = \underbrace{(p+q)}_{=1}\alpha = \alpha\end{aligned}$$

Allele Substitution



$$\Delta_1 = BV_{11} - BV_{12} = 2q\alpha - [(q-p)\alpha]$$
$$= 2q\alpha - q\alpha + p\alpha$$
$$= q\alpha + p\alpha = \underbrace{(p+q)}_{=1}\alpha = \alpha$$

$$\Delta_2 = BV_{12} - BV_{22} = (q-p)\alpha - [-2p\alpha]$$
$$= q\alpha - p\alpha + 2p\alpha$$
$$= q\alpha + p\alpha = (p+q)\alpha = \alpha$$

- The effect of substituting a G_2-allele by a G_1 allele on the breeding value is a constant and corresponds to α. Hence, α is called Allele Substitution Effect
- Special case for $\alpha = a + (q-p)d$ occurs when $d = 0 \Rightarrow \alpha = a$ (Assumption is used in Genomic Selection in Swiss Dairy Cattle Breeding)

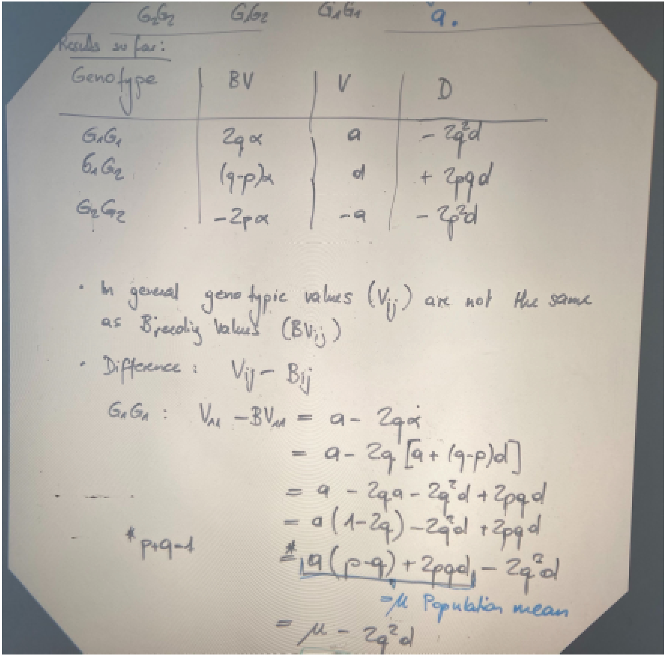
$\Downarrow$

The effect of a single Locus on a quantitative trait is determined by just one parameter a.

Results so far:

Genotype | BV

Summary BV



G_2G_2 G_1G_2 G_1G_1 $a.$

Results so far:

Genotype	BV	V	D
G_1G_1	$2q\alpha$	a	$-2q^2d$
G_1G_2	$(q-p)\alpha$	d	$+2pqd$
G_2G_2	$-2p\alpha$	$-a$	$-2p^2d$

- In general genotypic values (V_{ij}) are not the same as Breeding values (BV_{ij})
- Difference : $V_{ij} - B_{ij}$

G_1G_1 : $V_{11} - BV_{11} = a - 2q\alpha$

$= a - 2q[a + (q-p)d]$

$= a - 2qa - 2q^2d + 2pqd$

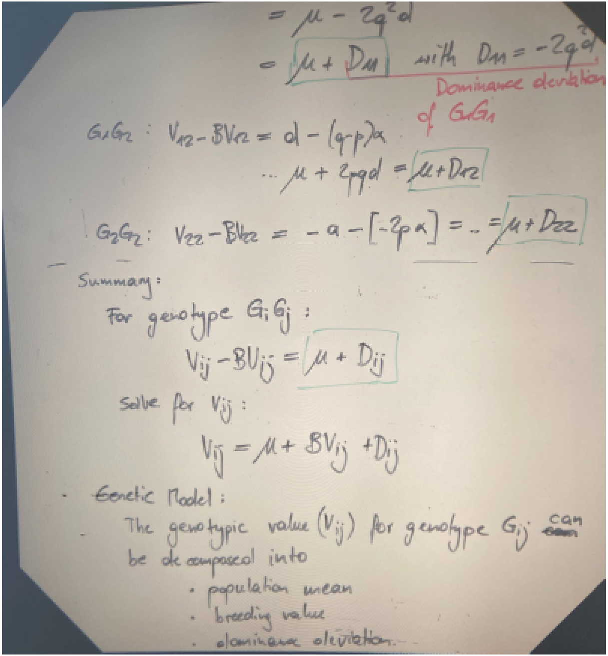
$= a(1-2q) - 2q^2d + 2pqd$

$* p+q=1$

$\overset{*}{=} \underbrace{a(p-q) + 2pqd}_{=\mu \text{ Population mean}} - 2q^2d$

$= \mu - 2q^2d$

Dominance



$$= \mu - 2q^2 d$$

$$= \boxed{\mu + D_{11}} \quad \text{with } D_{11} = -2q^2 d$$

Dominance deviation of G_1G_1

G_1G_2 : $V_{12} - BV_{12} = d - (q-p)\alpha$

$$\ldots \mu + 2pqd = \boxed{\mu + D_{12}}$$

G_2G_2 : $V_{22} - BV_{22} = -a - [-2p\alpha] = \ldots = \boxed{\mu + D_{22}}$

Summary:

For genotype G_iG_j :

$$V_{ij} - BV_{ij} = \boxed{\mu + D_{ij}}$$

Solve for V_{ij} :

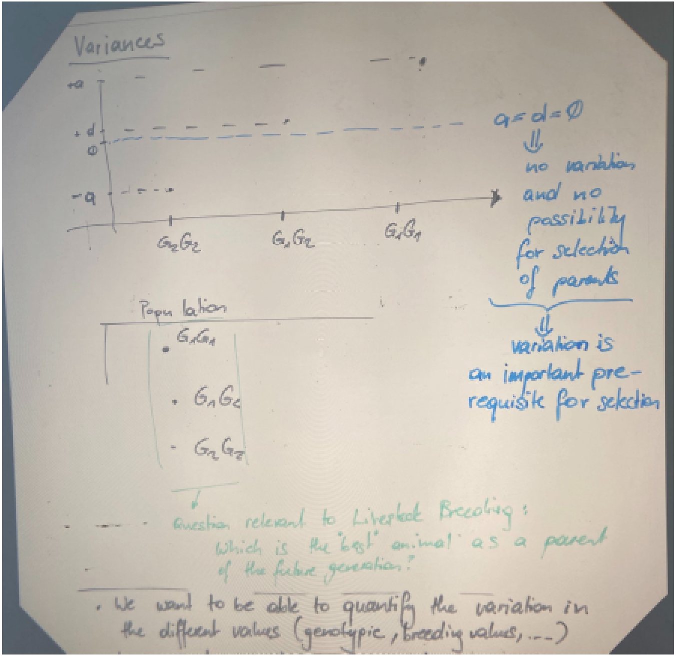
$$V_{ij} = \mu + BV_{ij} + D_{ij}$$

Genetic Model:

The genotypic value (V_{ij}) for genotype G_{ij} can be decomposed into

- population mean
- breeding value
- dominance deviation.

Variance



$a = d = 0$

⇓

no variation and no possibility for selection of parents

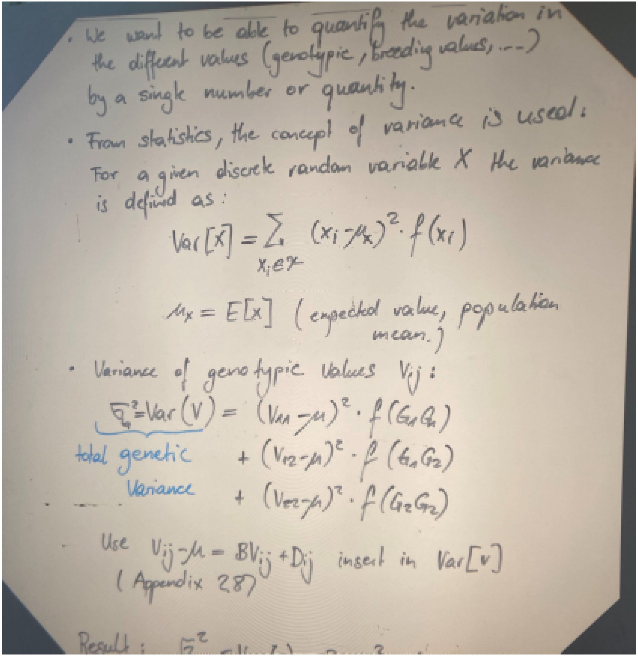
⇓

variation is an important pre-requisite for selection

Question relevant to Livestock Breeding:
Which is the "best" animal as a parent of the future generation?

- We want to be able to quantify the variation in the different values (genotypic, breeding values, ...)

Variance II



- We want to be able to quantify the variation in the different values (genotypic, breeding values, ...) by a single number or quantity.
- From statistics, the concept of variance is used:
  For a given discrete random variable X the variance is defined as:

$$Var[X] = \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 \cdot f(x_i)$$

$$\mu_X = E[X] \text{ (expected value, population mean.)}$$

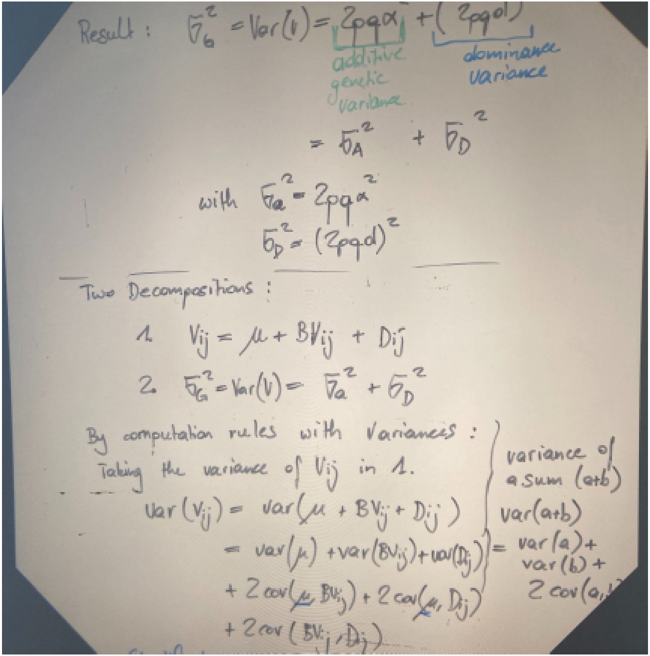
- Variance of genotypic values V_{ij}:

$$\underbrace{\sigma_G^2 = Var(V)}_{\text{total genetic variance}} = (V_{11} - \mu)^2 \cdot f(G_1G_1) + (V_{12} - \mu)^2 \cdot f(G_1G_2) + (V_{22} - \mu)^2 \cdot f(G_2G_2)$$

Use $V_{ij} - \mu = BV_{ij} + D_{ij}$ insert in $Var[V]$
(Appendix 28)

Result :

Decompositions



Result: $\sigma_G^2 = Var(V) = \underbrace{2pq\alpha^2}_{\text{additive genetic variance}} + \underbrace{(2pqd)^2}_{\text{dominance variance}}$

$$= \sigma_A^2 + \sigma_D^2$$

with $\sigma_a^2 = 2pq\alpha^2$

$\sigma_D^2 = (2pqd)^2$

Two Decompositions:

1. $V_{ij} = \mu + BV_{ij} + D_{ij}$
2. $\sigma_G^2 = Var(V) = \sigma_a^2 + \sigma_D^2$

By computation rules with Variances:

Taking the variance of V_{ij} in 1.

$$Var(V_{ij}) = var(\mu + BV_{ij} + D_{ij})$$

$$= var(\mu) + var(BV_{ij}) + var(D_{ij})$$

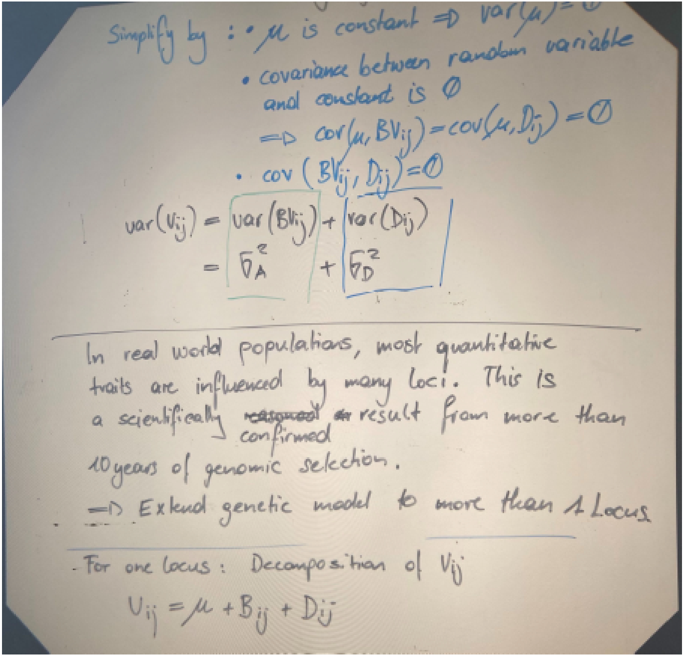
$$+ 2cov(\mu, BV_{ij}) + 2cov(\mu, D_{ij})$$

$$+ 2cov(BV_{ij}, D_{ij})$$

variance of a sum (a+b)

$var(a+b) = var(a) + var(b) + 2cov(a,$

More Loci



Simplify by : • μ is constant $\Rightarrow var(\mu) = 0$

• covariance between random variable and constant is 0

$\Rightarrow cov(\mu, BV_{ij}) = cov(\mu, D_{ij}) = 0$

• $cov(BV_{ij}, D_{ij}) = 0$

$$var(V_{ij}) = var(BV_{ij}) + var(D_{ij})$$
$$= \sigma_A^2 + \sigma_D^2$$

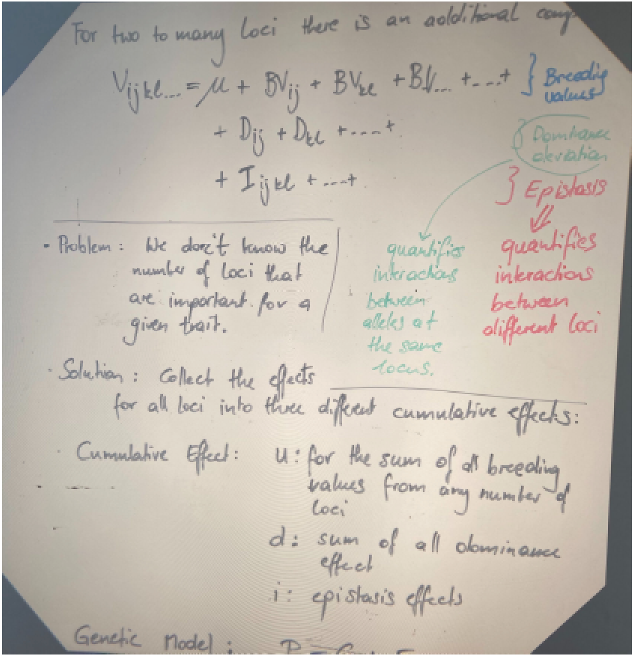
In real world populations, most quantitative traits are influenced by many loci. This is a scientifically ~~[illegible]~~ confirmed ~~[illegible]~~ result from more than 10 years of genomic selection.

$\Rightarrow$ Extend genetic model to more than 1 Locus

- For one locus : Decomposition of V_{ij}

$$V_{ij} = \mu + B_{ij} + D_{ij}$$

Two Loci



For two to many loci there is an additional comp

$$V_{ijkl\ldots} = \mu + BV_{ij} + BV_{kl} + BV_{\ldots} + \ldots + \Big\}\ \text{Breedig values}$$

$$+ D_{ij} + D_{kl} + \ldots + \Big\}\ \text{Dominance deviation}$$

$$+ I_{ijkl} + \ldots + \Big\}\ \text{Epistasis}$$

quantifies interactions between alleles at the same locus.

quantifies interactions between different loci

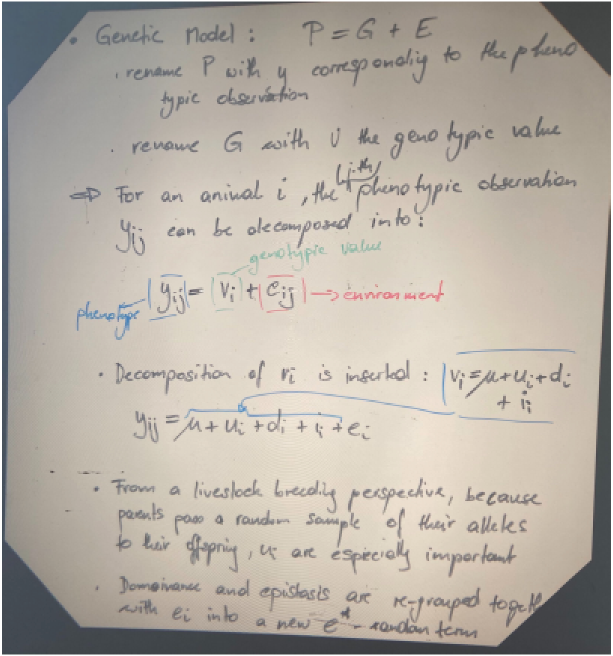
- Problem: We don't know the number of loci that are important for a given trait.

- Solution: Collect the effects for all loci into three different cumulative effects:

- Cumulative Effect:
  - u: for the sum of all breeding values from any number of loci
  - d: sum of all dominance effect
  - i: epistasis effects

Genetic model: P = G + E

Genetic Model



- Genetic model : $P = G + E$
  - rename P with y correspondig to the phenotypic observation
  - rename G with v the genotypic value

⇒ For an animal i, the (j-th) phenotypic observation y_{ij} can be decomposed into:

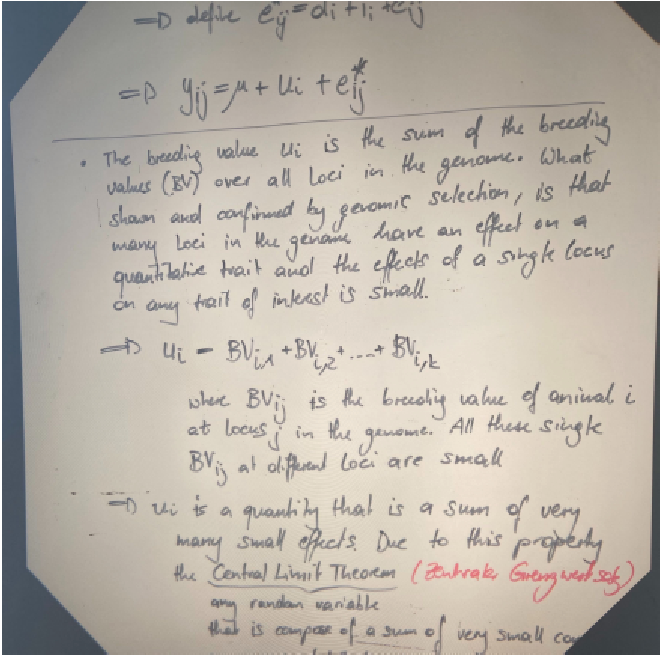
$$\underbrace{\boxed{y_{ij}}}_{\text{phenotype}} = \overbrace{\boxed{v_i}}^{\text{genotypic value}} + \boxed{e_{ij}} \rightarrow \text{environment}$$

- Decomposition of v_i is inserted : $\boxed{v_i = \mu + u_i + d_i + i_i}$

$$y_{ij} = \mu + u_i + d_i + i_i + e_i$$

- From a livestock breedig perspective, because parents pass a random sample of their alleles to heir offsprig, u_i are especially important
- Dominance and epistasis are re-grouped togeth with e_i into a new e^* - random term

Breeding Value



⇒ define $e^{*}_{ij} = d_i + l_i + e_{ij}$

⇒ $y_{ij} = \mu + u_i + e^{*}_{ij}$

- The breeding value u_i is the sum of the breeding values (BV) over all loci in the genome. What shown and confirmed by genomic selection, is that many loci in the genome have an effect on a quantitative trait and the effects of a single locus on any trait of interest is small.

⇒ $u_i = BV_{i,1} + BV_{i,2} + \dots + BV_{i,k}$

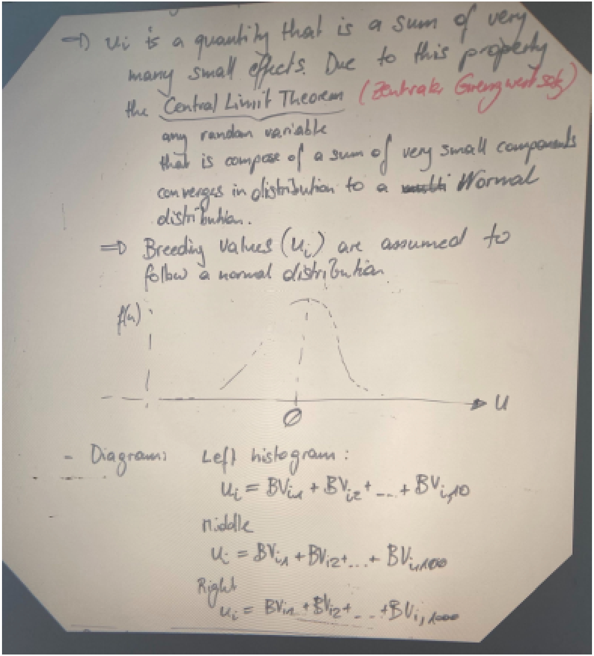
where BV_{ij} is the breeding value of animal i at locus j in the genome. All these single BV_{ij} at different loci are small

⇒ u_i is a quantity that is a sum of very many small effects. Due to this property the Central Limit Theorem (Zentraler Grenzwertsatz)

any random variable

that is compose of a sum of very small co

CLT



→ u_i is a quantity that is a sum of very many small effects. Due to this property the Central Limit Theorem (Zentrale Grenzwertsatz)

any random variable that is compose of a sum of very small components converges in distribution to a ~~multi~~ Normal distribution.

⇒ Breeding values (u_i) are assumed to follow a normal distribution

- Diagrams: Left histogram:

$u_i = BV_{i1} + BV_{i2} + \dots + BV_{i,10}$

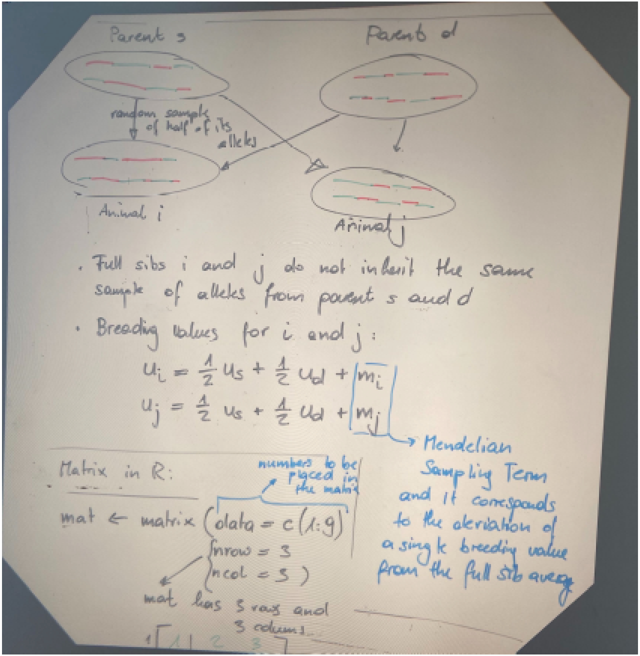
middle

$u_i = BV_{i1} + BV_{i2} + \dots + BV_{i,100}$

Right

$u_i = BV_{i1} + BV_{i2} + \dots + BV_{i,1000}$

Parent - Offspring



- Full sibs i and j do not inherit the same sample of alleles from parent s and d
- Breeding values for i and j:

$$u_i = \frac{1}{2} u_s + \frac{1}{2} u_d + \boxed{m_i}$$
$$u_j = \frac{1}{2} u_s + \frac{1}{2} u_d + \boxed{m_j}$$

→ Mendelian Sampling Term and it corresponds to the deviation of a single breeding value from the full sib average

Matrix in R:

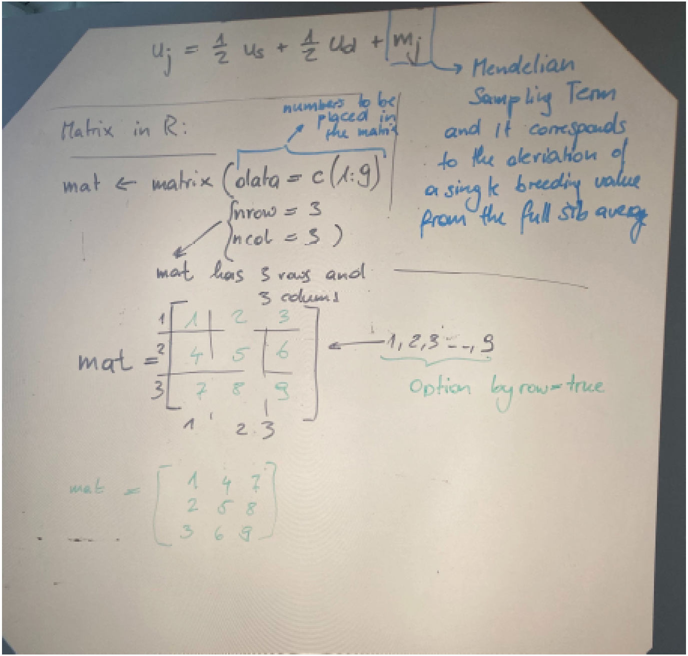
numbers to be placed in the matrix

```
mat ← matrix (data = c(1:9)
              nrow = 3
              ncol = 3 )
```

mat has 3 rows and 3 columns

1 [1 2 3]

Matrix in R



$$u_j = \frac{1}{2} u_s + \frac{1}{2} u_d + \boxed{m_j}$$

→ Mendelian Sampling Term and it corresponds to the deviation of a single breeding value from the full sib averag

Matrix in R:

```
mat <- matrix(data = c(1:9),
              nrow = 3,
              ncol = 3 )
```

numbers to be placed in the matrix

mat has 3 rows and 3 columns

$$\text{mat} = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 1 & 2 & 3 \\ 2 & 4 & 5 & 6 \\ 3 & 7 & 8 & 9 \end{array} \leftarrow \underbrace{1, 2, 3 \ldots, 9}$$

Option byrow=true

$$\text{mat} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$