

## OHP Picture 1

Recap: 2023-03-13

- Predictors with discrete fixed levels such as "breed", the different levels are converted to numeric codes

In R: Assignment of factor levels to numbers is done according to the alphabetic order of the factor levels

- This sort of predictors ('x') are also called factors. The possible values that a factor can take are called levels

- Eg: "Breed" is a factor  
"Angus", "Limousin", "Simmental" are the levels

---

Model: - Show influence of breed on expected

## OHP Picture 2

Model : - Show influence of breed on expected body weight:  
- For animal  $i$ , the expected body weight

$$E(y_i) = b_0 + b_1 x_i$$

where  $x_i$  is the numeric code for the breed of animal  $i$

Consequence of Model : Consider data

Animal 1:  $E(y_1) = b_0 + b_1 \cdot 1$   
3:  $E(y_3) = b_0 + b_1 \cdot 3$   
10:  $E(y_{10}) = b_0 + b_1 \cdot 2$

} dependent on assignment of codes to breeds

$$E(y_{3i}) - E(y_{10i}) = 2 \cdot b_1$$

$$E(y_{3i}) - E(y_{1i}) = b_1$$

Exception : Regression on discrete Factors can be used in Marker effect estimation

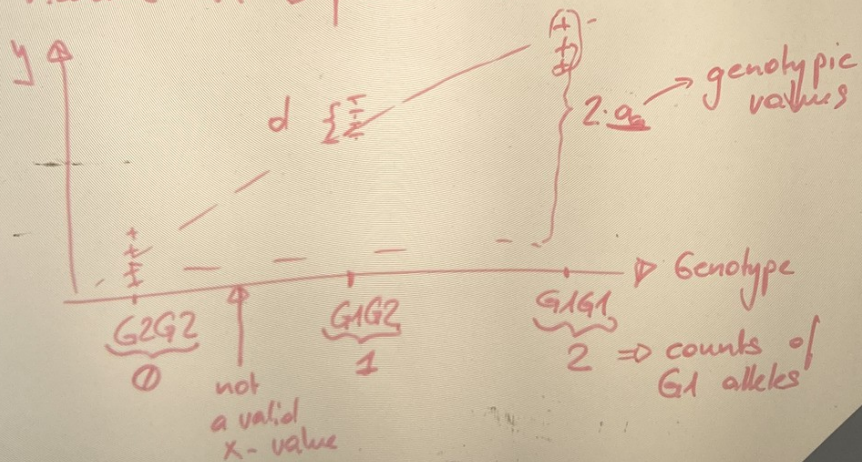
OHP Picture 3

Exception: Regression on discrete Factors can be used in Marker effect estimation

□ Data set:

Animall	Locus G	Observation
1	G1G2	471
2	G2G2	463
⋮		
N	G1G1	541

□ Assume G1 is positive Allele:



OHP Picture 4

In single-locus Genetic Model with 3 genotypes

Genotype	genotypic Value
$G_1G_1$	$+ a$
$G_1G_2$	$+ d$
$G_2G_2$	$- a$

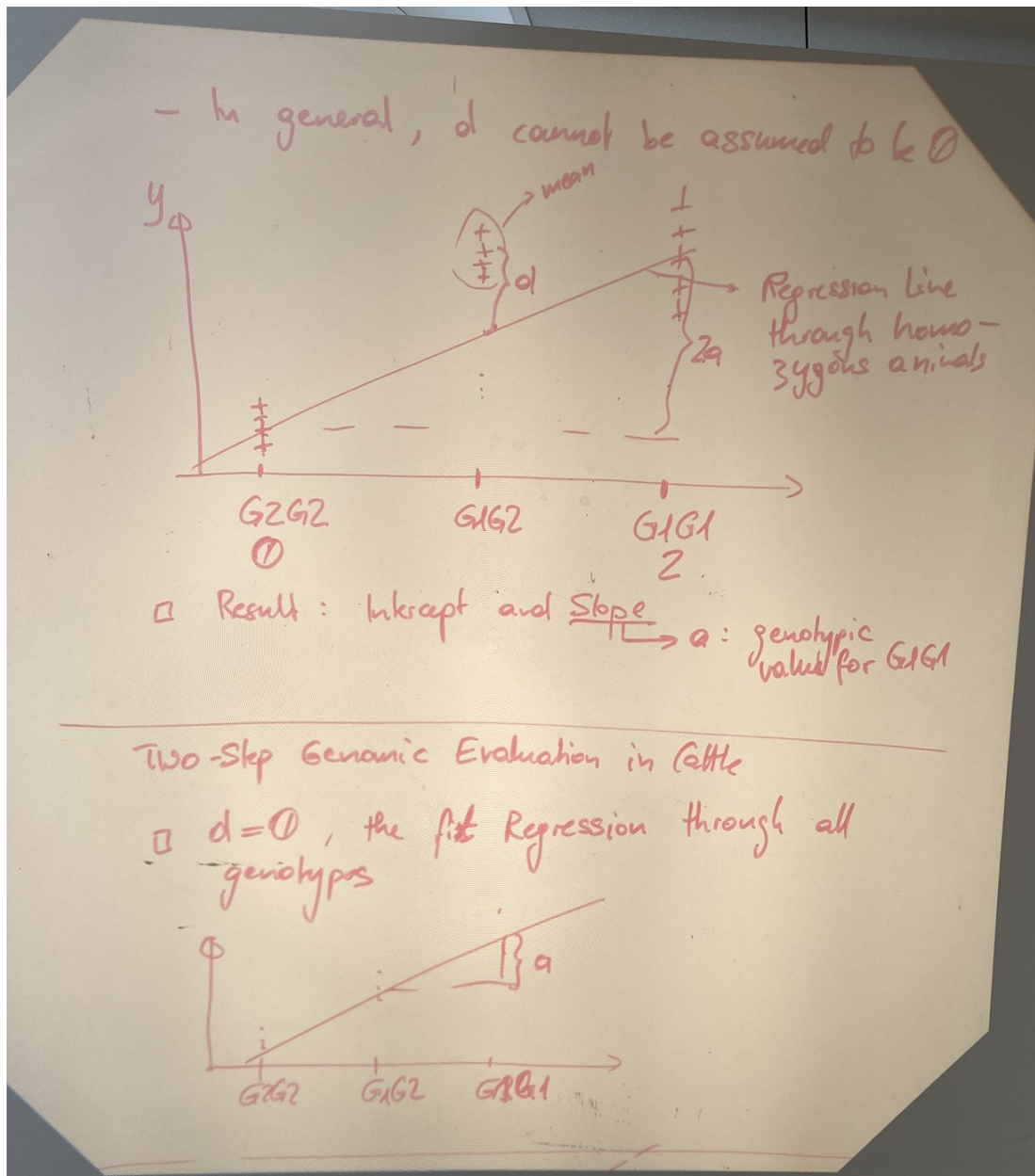
□  $a$  and  $d$  are parameters to be estimated from data

□ Data

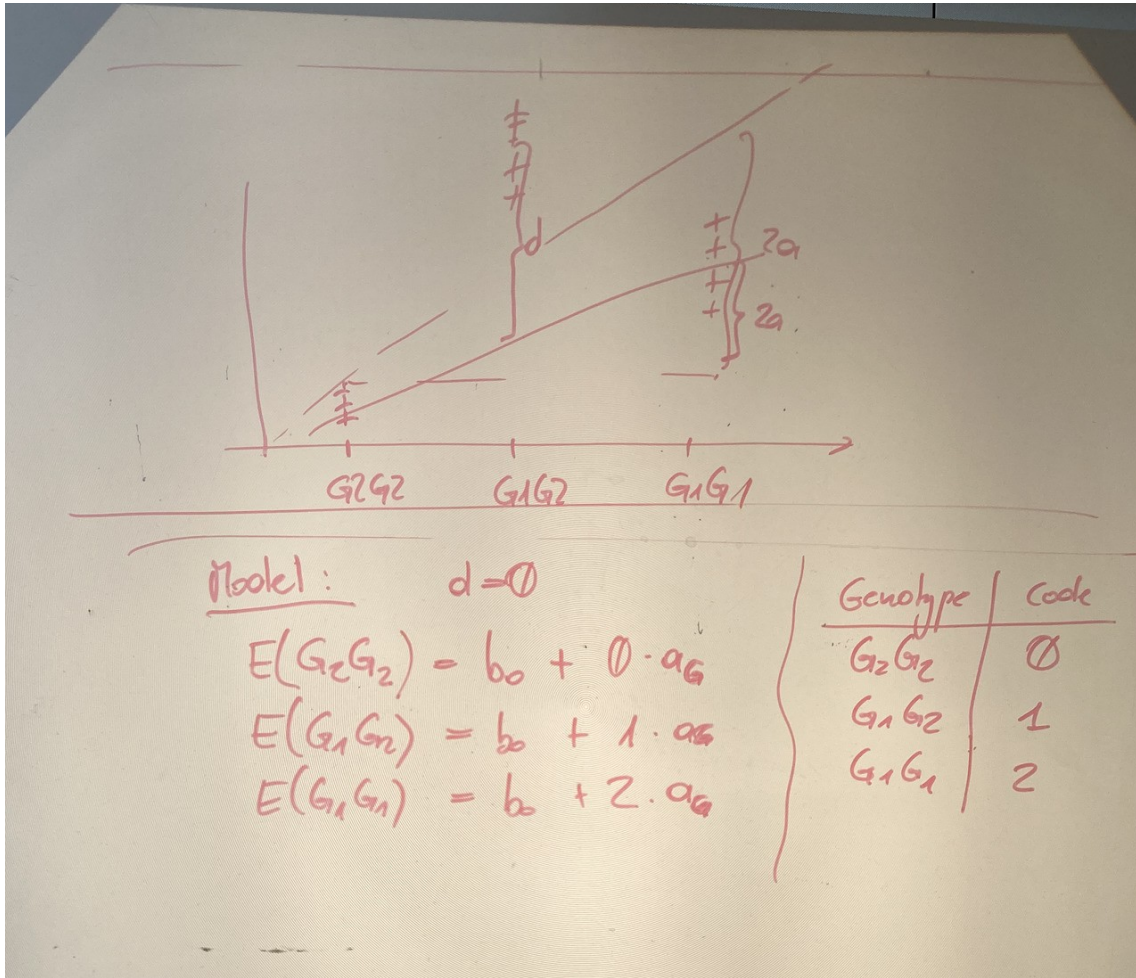
Animals	Genotype	Observations
1	$G_1G_2$	4.71
2	$G_2G_2$	4.63
⋮	⋮	⋮
$N$	$G_1G_1$	5.41

- In general,  $d$  cannot be assumed to be 0

OHP Picture 5



OHP Picture 6



## Regression on Dummy Variables

□ Not one predictor (x-variable) per factor (breed) but an additional predictor per level

□ Factor breed

Levels	Predictor
Angus	$x_1$
Limousin	$x_2$
Simmental	$x_3$

$\left. \begin{array}{l} b_1 \\ b_2 \\ b_3 \end{array} \right\}$  In matrix-vector notation additional columns in X-matrix  
 $\downarrow$   
 unknowns

$$\begin{aligned} \text{Animal 1: } y_{11} &= b_0 + 1 \cdot b_1 + 0 \cdot b_2 + 0 \cdot b_3 + e_{11} \\ \text{Animal 2: } y_{21} &= b_0 + 1 \cdot b_1 + 0 \cdot b_2 + 0 \cdot b_3 + e_{21} \\ \text{3: } y_{33} &= b_0 + 0 \cdot b_1 + 0 \cdot b_2 + 1 \cdot b_3 + e_{33} \\ \text{10: } y_{102} &= b_0 + 0 \cdot b_1 + 1 \cdot b_2 + 0 \cdot b_3 + e_{102} \end{aligned}$$

In Matrix-Vector Notation } Group animals according to

OHP Picture 8

In Matrix - Vector Notation :

vector  $y = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{41} \\ y_{72} \\ y_{82} \\ y_{92} \\ y_{102} \\ y_{13} \\ y_{63} \end{bmatrix} = \begin{bmatrix} 471 \\ 463 \\ 470 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 491 \end{bmatrix}$

Group animals according to breeds  
All Angus Animals first, then all Limousin animals, last all Simmental animals

vector:  $b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$

- $b_0$  → intercept
- $b_1$  → effect of Angus on BW
- $b_2$  → effect of Limousin on BW
- $b_3$  → effect of Simmental on BW

vector  $e = \begin{bmatrix} e_{11} \\ e_{21} \\ e_{41} \\ \cdot \\ e_{63} \end{bmatrix}$

---

Matrix  $X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$



Matrix  $X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}$  design matrix  
incidence

Model :  $\underbrace{y}_{\text{Known}} = \underbrace{X}_{\text{Known}} \cdot \underbrace{b}_{\text{Unknown, estimated from data}} + \underbrace{e}_{\text{Unknown}}$

Estimates of  $b$  using least squares

- Find  $\hat{b}$  such that  $e^T e$  is minimal
- Result :  $X^T X b^{(0)} = X^T y$  (NEq)

Req :  $\hat{b} = (X^T X)^{-1} X^T y$   
is only possible, if  $X$  has full rank



OHP Picture 11

For  $Ax = y$  with  $G$  being a generalized  
inverse of  $A \Leftrightarrow AGA = A$   
 $x = Gy$  is a solution of  $Ax = y$ .

Pre-multiply with  $\Rightarrow Ax = AGy$   
replace  $y$  with  $Ax \Rightarrow Ax = \underbrace{AGA}_{\neq A}x$

---

Normal equations:  $(X^T X) b^0 = X^T y$

$G$  being a generalized inverse of  $(X^T X)$   
 $\Rightarrow b^0 = GX^T y$