In Real dataset:

- The number of predictors can be large
  - In ChatGPT: 10^{12}
- In Livestock datasets, there might be 20 - 100 predictors, for environmental factors
- Genomic datasets: Genotypes at 150K positions.

Animal	Body Weight	SNP_1	SNP_2	...	$SNP_{150'000}$	...
1						
⋮						
N						

Linear Model:

$$y_i = b_0 + b_{SNP_1} \cdot x_{i,1} + b_{SNP_2} \cdot x_{i,2} + \ldots + b_{SNP_{150'000}} \cdot x_{i,150'000} + \ldots + e_i$$

$$\underline{b} = \begin{bmatrix} b_0 \\ b_{SNP_1} \\ b_{SNP_2} \\ b_{SNP_{100}} \\ b_{SNP_{150'000}} \\ b_{BC} \\ \vdots \end{bmatrix}$$

important

Least Squares Solution can no longer be used because $(X'X)^{-1}$ cannot be computed.