Modeling and Simulating Phenotypes in Quantitative Genetics

Introduction

In quantitative genetics, a phenotype is understood as the observable characteristics of an individual, resulting from the interplay of its genetic makeup and environmental influences. Mathematical models are essential tools for simulating these complex relationships, allowing researchers to test hypotheses and understand the architecture of traits.

Below, we describe two simulation models. The first is a foundational model for a single phenotype, and the second is an extended model for two phenotypes with a causal relationship and confounding factors.

SpatialTranscriptomicsSimulation
CausalPhenotypeSimulation

Simulating a Single Phenotype with a Defined Heritability

This model simulates a single phenotype by combining genetic and environmental effects.

The Model

The phenotype (P) of an individual is modeled as the sum of its genetic component (G) and an environmental component (E):

$$P = G + E$$

- P **(Phenotype):** The final, observable trait value.
- G **(Genotype):** The aggregate effect of all relevant genes on the phenotype. In a simulation context, this value can be derived from real genetic data, such as a polygenic risk score (PRS). We assume its variance, $\mathrm{Var}(G)$, is known or can be calculated from the data.
- E **(Environment):** The sum of all non-genetic factors that influence the phenotype. This term also includes measurement error. We assume the environmental effects are random and follow a normal distribution with a mean of 0.

Deriving Environmental Variance from Heritability

A key parameter in genetics is **narrow-sense heritability (h^2)**, which represents the proportion of the total phenotypic variance ($\mathrm{Var}(P)$) that is due to additive genetic variance ($\mathrm{Var}(G)$).

$$h^2 = \frac{\mathrm{Var}(G)}{\mathrm{Var}(P)}$$

To simulate a phenotype with a specific heritability, we need to determine the appropriate variance for the environmental component, $\mathrm{Var}(E)$. We can derive this as follows, assuming the genetic and environmental effects are independent ($\mathrm{Cov}(G, E) = 0$):

1. The total phenotypic variance is the sum of the genetic and environmental variances:

$$\mathrm{Var}(P) = \mathrm{Var}(G) + \mathrm{Var}(E)$$

2. From the definition of heritability, we can express the total phenotypic variance as:

$$\mathrm{Var}(P) = \frac{\mathrm{Var}(G)}{h^2}$$

3. By equating the two expressions for $\mathrm{Var}(P)$, we can solve for $\mathrm{Var}(E)$:

$$\mathrm{Var}(G) + \mathrm{Var}(E) = \frac{\mathrm{Var}(G)}{h^2}$$
$$\mathrm{Var}(E) = \frac{\mathrm{Var}(G)}{h^2} - \mathrm{Var}(G)$$

4. Factoring out $\mathrm{Var}(G)$ gives the final formula for the environmental variance:

$$\mathrm{Var}(E) = \mathrm{Var}(G)\left(\frac{1}{h^2} - 1\right) = \mathrm{Var}(G)\frac{1 - h^2}{h^2}$$

With this formula, you can calculate the necessary environmental variance to achieve a desired heritability for a given genetic variance.