

## 4.4 Complex traits: I. Quantitative Genetics

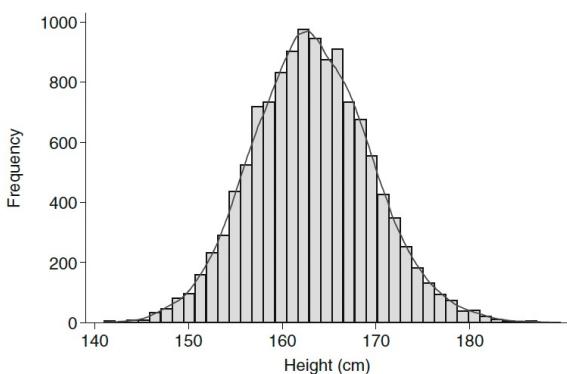
We all know families that are tall or athletic or musical. Your doctor may ask if you have a family history of cancer, or heart disease. What is the role of genetics – and what is the role of shared environment – in shaping each of these traits? Why do relatives often resemble one another?

So far we have been looking at the role of large-effect mutations in disease. But most traits are very different: they are influenced by thousands of SNPs spread across the genome, and by environmental factors. Such traits are referred to as **complex traits**, and they will be our focus in the upcoming chapters.<sup>701 a.</sup>

We start our study of complex traits by introducing a modeling framework known as **quantitative genetics**. Quantitative genetics describes the inheritance of complex traits at a macroscopic level, without explicit reference to the actual SNPs, genes or pathways that drive trait variation<sup>b</sup>. This might seem limiting, but these simplifications enable fundamental insights into the nature of inheritance.

Quantitative genetics is mainly concerned with traits that vary along a continuous measurement scale, known as **quantitative traits**. As we'll see, these models can also be extended to understand binary traits including diseases.

One classic example of a quantitative trait is human **height**, and we'll use height to illustrate our main themes throughout this chapter:



In this chapter we'll use the tools of quantitative genetics to tackle the following questions:

- Given that genetic variation is discrete, how do we explain the existence of *continuously varying phenotypes*?
- What controls the amount of variation (*the phenotypic variance*) in a quantitative phenotype?
- What controls the inheritance (*similarity between relatives*) of quantitative phenotypes?
- How does selection act on quantitative traits?

<sup>a</sup> Rough draft version of chapter. Figures to be redrawn.

<sup>b</sup> The role of quantitative genetics in human genetics is similar to the role of the Ideal Gas Law  $PV = nRT$  in Intro Chemistry. The Ideal Gas Law gives a powerful description of the macroscopic properties of a gas in a container, without modeling the movements of individual molecules.

Figure 4.31: Distribution of height among adult women in England. How does genetics contribute to variation in height?

Credit: Figure 1a Hazel Inskip et al (2017) [[Link](#)] CC BY 4



Figure 4.32: "The Two Sisters" (Théodore Chassériau, 1843). Why do relatives often have similar phenotypes? Louvre, Public Domain.

**Galton, and the battle of the Mendelians and Biometricalians.** Our story begins in the mid-19th Century with Charles Darwin's half-cousin, Francis Galton. One key gap in Darwin's work was that it was unclear what controls inheritance (Chapter 2.5). Recall that although Mendel's work on genetics in peas was published in 1866, it was largely unknown until the rediscovery of the work in 1900.

Darwin's work prompted Galton into a series of studies to better understand the nature of inheritance<sup>702</sup>. Galton's first innovation was to collect careful, precise, data on the inheritance of traits among relatives.

In a pioneering paper from 1886, Galton reported data on the heights of 930 adults, and their parents<sup>703</sup>. The raw data are shown here. Each data point shows the average height of two parents (x-axis) and their adult child (y-axis):

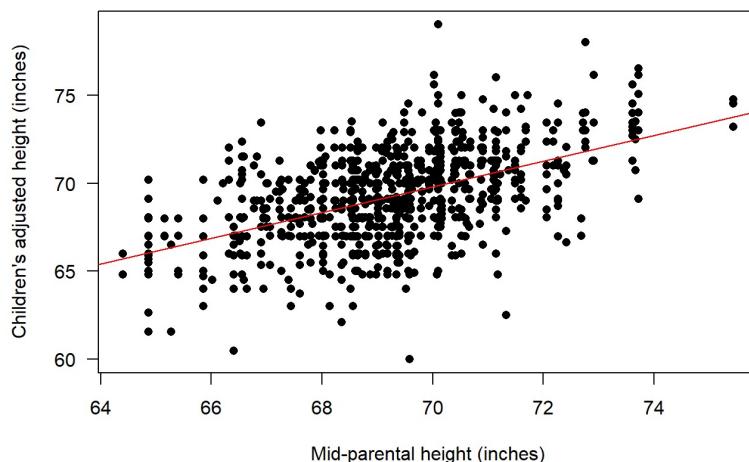


Figure 4.33: The height of adult children compared to their parents (data from Galton 1886). All female heights are multiplied by 1.08 in order to match the male and female averages. The red line shows the regression to the data. Credit: Figure by Yuk Tung Liu [[Link](#)] GPL 3.0 [redraw this!]

These data show something we probably expect intuitively: tall parents tend to have tall kids, and short parents have short kids. But Galton made a very interesting point: the best-fit line for the height of the children (the 'children' line, below) does *not* track the average height of their parents (the 'mid-parents' line, below). Instead, the children's line lies between the midparents line and the overall average:

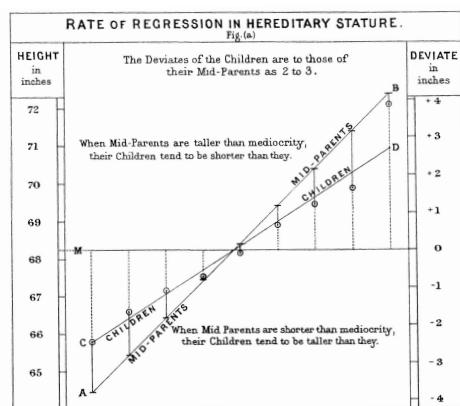


Figure 4.34: Regression Towards Mediocrity in Hereditary Stature (Galton, 1886). The axes are the same as above; the 'mid-parents' line shows the  $x=y$  diagonal. The "children" line shows the regression fit to the data. Credit: Francis Galton (1886) [[Link](#)]

As Galton put it, the children tend to regress<sup>c</sup> (or move back) towards the average (or 'mediocrity' in his terms). In other words, *children of tall*

<sup>c</sup> We now use Galton's term **regression** to mean fitting a line through the data, similar to what he did here.

parents tend to be slightly shorter than their parents, while children of short parents tend to be taller.

In the next few pages we'll show how quantitative genetics came to explain Galton's observations.

**The bitter fight between the biometricians and Mendelians.** Galton's work helped motivate a new field of quantitative studies of inheritance. Its proponents were known as **biometricians**, and they were chiefly interested in measuring phenotypic variation and its inheritance, but not especially in understanding mechanisms of transmission. (Galton's 1886 paper proposed an incorrect model of inheritance in which a person's phenotype was a weighted average over recent and distant ancestors<sup>704</sup>).

So while the rediscovery of **Mendel's work** in 1900 provided remarkable new insight into the nature of inheritance, the biometricians were uninterested. They believed that evolution was driven by small shifts in phenotypes, and viewed Mendel's observations of discrete traits and discrete inheritance of alleles, as representing some kind of extreme process that would not be relevant in nature.

In 1902, one of the leading Mendelians, William Bateson, bemoaning the lack of interest in Mendel's discoveries, likened Mendel's work to a new gospel and complained bitterly about the entrenched views of the biometricians: "Not lightly do men let their occupation go; small...wonder (that) we find the established prophet unconvinced"<sup>705</sup>.

Udny Yule fired back on behalf of the biometricians that "one cannot help feeling that (Bateson's) speculations would have had more value had he kept his emotions under better control"<sup>706</sup> <sup>707</sup>. Fiery stuff! It's left to the reader to speculate what they might have said if there was social media in 1902!

**Reconciliation: Fisher and the Infinitesimal Model.** The resolution of these schools was brought about in 1918 by a young mathematics student named Ronald Fisher, who went on to become a founding figure of both statistics and population genetics<sup>708</sup>. Fisher's paper, titled "The Correlation between Relatives on the Supposition of Mendelian Inheritance" showed that it was possible to derive the key results of Galton and the biometricians if one assumed that traits were controlled not by one, but by many Mendelian loci.

We can describe Fisher's idea in modern terms as follows. We model an individual  $i$ 's phenotype  $Y_i$  for a quantitative trait as a sum of the population mean phenotype  $\bar{y}$  plus random components based on their genotype ( $G_i$ ) and their environment ( $E_i$ ):

$$Y_i = \bar{y} + G_i + E_i \quad (4.16)$$

Here,  $E_i$  includes all types of non-genetic factors: anything from maternal effects during pregnancy, to nutrition, to household smoking, to infection history; as well as any other unexplained randomness in the phenotype such as measurement error.



Figure 4.35: "Research about plant hybrids" (Mendel, 1866). Mendel's revolutionary work was published in an obscure journal with an unrevealing title; the scientific world was largely unaware of the paper for 34 years. [[Link](#)]

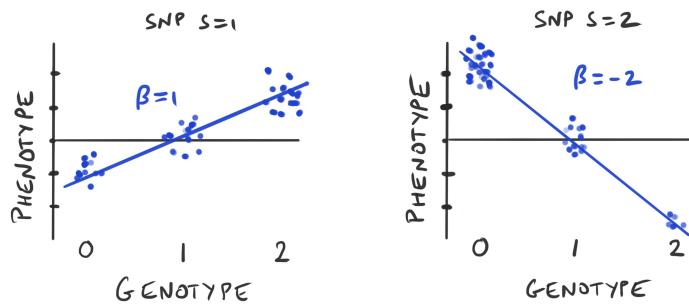
$$Y = \bar{y} + G + E$$

$$168\text{cm} = 162\text{cm} + 8\text{cm} + (-2\text{cm})$$

To model  $G_i$ , suppose that  $L$  sites in the genome can affect the trait (we'll refer to these sites as SNPs, but in practice these could represent any type of genetic variant). We'll use  $l$  to index the SNPs, where  $l$  is an integer between 1 and  $L$ .

For each SNP,  $g_{i,l}$  indicates the number of derived alleles<sup>709</sup> that individual  $i$  carries at SNP  $l$ , where  $g_{i,l}$  can be 0, 1, or 2.<sup>d</sup>

Next,  $\alpha_l$  measures the effect of SNP  $l$  on the trait<sup>710</sup>. Specifically,  $\alpha_l$  is the average change in the trait per copy of the derived allele:



If  $L$  SNPs across the gene can all affect the trait, how would they combine to produce the overall genotype effect  $G$ ? The simplest possible model – and one that works well in practice – is simply to add the individual SNP effects together<sup>e</sup>:

$$G_i^* = \sum_{l=1}^L g_{i,l} \alpha_l. \quad (4.17)$$

(In practice it's helpful to shift the distribution of genetic scores to mean zero. To do this, we use the genetic score  $G_i$ , defined  $G_i^* - E(G_i^*)$ .)

So, to summarize, we predict the genetic effect for individual  $i$  by looking up their genotype  $g_{i,l}$  at each relevant SNP, and multiply that by the relevant SNP-effect-on-trait  $\alpha_l$ ; these are added up to predict the total genetic effect,  $G_i$ .  $G_i$  is known as a **polygenic score**.

**$G_i$  converges to a normal distribution.** Fisher's fundamental insight was that the classic Mendelian segregation of discrete phenotypes occurs when a single variant contributes to trait variation. But if  $G$  is a sum of effects across many loci, then the phenotypic distribution quickly converges to a simple normal (i.e., bell-shaped) distribution.

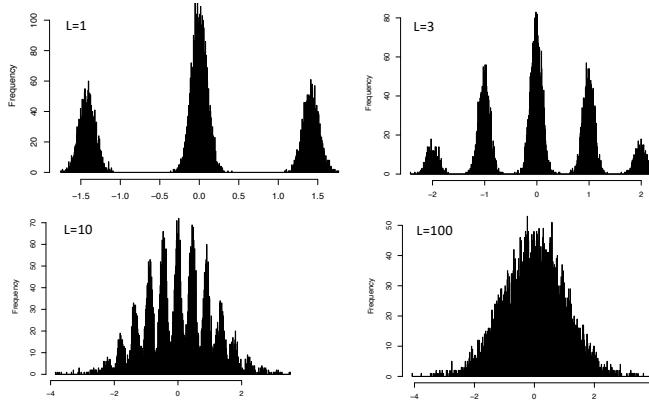
Here's a simulation of what this looks like under an additive model, for different numbers of SNPs:

Figure 4.36: We can model a person's height as the sum of the population average (female) height  $\bar{y}$  plus two independent random variables, representing genotype and environment. In this hypothetical example, this person's genotype  $G_i$  adds 8 cm to her expected height, and her environment (e.g., childhood nutrition) subtracts 2 cm. update subscripts

<sup>d</sup> Change notation to  $\alpha$  to match the next chapter.

Figure 4.37: Allelic effects on a trait. Here, each blue dot shows the observed phenotype for a different individual, plotted according to genotype at each of two SNPs.  $\beta_1$  is the average per-allele effect of SNP  $l$ , and corresponds to the slope of the regression line. fix snp s to 1 and  $\beta$  to  $\alpha$

<sup>e</sup> This is known as an **additive model**; we'll briefly discuss more complex models below.



**Figure 4.38: Theory: Polygenic models converge toward a continuous distribution.**

Each panel shows a simulated distribution of phenotypes, where the phenotypes are controlled by an increasing number of loci (from 1 to 100) and environmental noise. The simulations assume shared  $\beta$  and  $p = .5$  at all SNPs. Credit: Source unknown. [Redraw this figure. nb that L=3 is mislabeled]

In the last case here, with  $L = 100$ , the distribution of phenotypes is practically indistinguishable from a normal distribution.

This simulation suggests that instead of modeling inheritance in terms of the transmission dynamics of individual loci, we could instead model inheritance as a sum of normal distributions of genetic and environmental effects. As we shall see, this makes the modeling much simpler. Indeed, so long as  $L$  is “large enough” we cannot tell from the phenotypic data how many loci are contributing: the inheritance patterns with  $L$  SNPs are practical identical to the inheritance patterns with  $10 \times L$  SNPs, if each SNP contributes 1/10th as much variance. We now know that this model is relevant in practice, as most complex traits in humans are impacted by many thousands of SNPs<sup>711</sup>.

Over time, this model has come to be known as the **Infinitesimal Model** because we can formally derive the normal approximation to  $G$  by assuming infinitely many SNPs, each with each with infinitely small effects<sup>f</sup> 712. However, the term does not appear in Fisher’s 1918 paper, and its origin appears to be unknown 713.

**Modern data confirm the concept of polygenic scores.** The first direct confirmation of Fisher’s model came about 90 years later, with the advent of genome-wide association studies (GWAS)<sup>g</sup>.

In 2008, Michael Weedon and colleagues identified 20 SNPs across the genome that affect adult height<sup>714</sup>. For each SNP they described one allele as the ‘tall’ allele, on the basis that it is associated with increased height. Then, for each individual, they counted the number of ‘tall’ alleles across the 20 SNPs (possible range 0–40).

This distribution is shown in the histogram below. As you can see, individuals with  $>27$  tall alleles were around 5cm taller than individuals with  $<17$  tall alleles<sup>h</sup>:

<sup>f</sup> One main implication of the infinitesimal model is that we can understand inheritance without modeling each causal SNP individually.

<sup>g</sup> We’ll come back to GWAS in much more detail in the next chapter.

<sup>h</sup> In practice, we use a different  $\bar{y}$  for male and female height, though for brevity this is not emphasized in our notation.

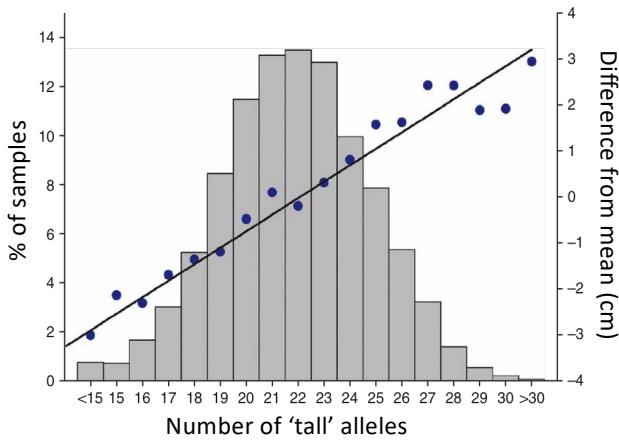


Figure 4.39: **Polygenic prediction of height (2008).** The histogram shows the numbers of tall alleles at carried by each individual. For each bin, the average height (relative to the population mean) is shown by a blue dot, and the black line shows the regression fit through the blue dots.

Note: You can think of the x-axis here as equivalent to  $G_i$  but setting all the  $\beta$ s to 1. Later versions of polygenic scores use the estimated  $\beta$ s for better accuracy. Credit: Figure 5 from Michael Weedon et al (2008) [Link]

This example illustrates how the predicted genotype using just a handful of SNPs quickly converges to a normal distribution. As we shall see, in fact height is hugely polygenic: a recent GWAS of 5.4 million people identified more than 12,000 SNPs that are independently associated with height (most with tiny effect sizes). Polygenic prediction of height has also improved greatly since 2008: a 2023 study reported a 23 cm difference between the predicted top and bottom quantiles<sup>715</sup>.

We're now ready to build a basic quantitative model of inheritance. We'll write this in terms that are used to measure statistical distributions: **expectation, variance, covariance, and correlation.** If you're not familiar with those, you should first read the short overview below:

**Optional interlude: A short primer on statistics.** Our goal in this section is to quantify phenotype distributions – e.g., students' heights – and the similarity between pairs of distributions – e.g., a sample of students' heights and their parents' heights. I'm keeping this short as there's loads of helpful material on the internet.

**Expectation.** The symbol  $E(y)$  is read as the “expectation of  $y$ ” or the “expected value of  $y$ ”. This corresponds to the average (mean) value of a variable – for example the average value of female height in a population. Suppose we have  $m$  individuals, and the height of the  $i$ th individual is  $y_i$ . Then we estimate  $E(y)$  as<sup>716</sup>:

$$E(y) = \frac{1}{m} \sum_{i=1}^m y_i \quad (4.18)$$

We'll sometimes shorten the notation by using  $\bar{y}$  in place of  $E(y)$

**Variance.**  $\text{Var}(y)$ , the “variance of  $y$ ”, tells us how spread out a variable is. It is computed as the average squared difference between a variable and its mean:

$$\text{Var}(y) = E[(y_i - \bar{y})^2] \quad (4.19)$$

The **standard deviation** is defined as the square root of the variance. The standard deviation and variance are sometimes written as  $\sigma_y$ , and  $\sigma_y^2$ , respectively.

Our goal in quantitative genetics is to understand the causes of phenotypic variation. For example, the standard deviation of people's heights is about 7 cm (within sex), and so the variance is about  $49 \text{ cm}^2$ . A useful rule of thumb is that about 95% of the population is within 2 standard deviations of the mean, so within  $\pm 14\text{cm}$  of the population average height.

**Covariance.**  $\text{Cov}(x, y)$ , the “covariance of  $x$  and  $y$ ” measures the relationship between two random variables. This is defined as

$$\text{Cov}(x, y) = E[(x_i - \bar{x})(y_i - \bar{y})] \quad (4.20)$$

$$= E(x_i y_i) - \bar{x} \cdot \bar{y} \quad (4.21)$$

where you can think of  $x_i$  and  $y_i$  as being two related measurements: for example the heights of a parent and child from the same family. An inherited trait such as height will tend to have a *positive covariance* between relatives. If one relative ( $x_i$ ) is higher than the mean, then the other relative ( $y_i$ ) is also likely to be higher than the mean (and the reverse if they are both less than the mean). This will result in positive values in the covariance formula above.

The covariance is closely related to the **correlation**:

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} \quad (4.22)$$

A bit of math shows that the correlation is mathematically required to fall between  $-1$  and  $1$ . Two variables with a strongly positive relationship will have a correlation near  $1$ .

One question we will ask in this chapter is: *How much of the population variance in height is due to genetic factors?* Another key goal is to model correlations between phenotypes of relatives – for example, *What is the correlation between the height of a parent and their child, and how does this depend on genetics?*

**A quantitative model of inheritance.** Recall that our basic additive model treats a person's phenotype  $Y_i$  as the sum of a polygenic score  $G_i$  and an environmental/random effect  $E_i$ :

$$Y_i = \bar{y} + G_i + E_i, \quad (4.23)$$

where  $\bar{y}$  is the population mean. You can think of both  $G_i$  and  $E_i$  as random draws from normal distributions with mean 0. Then the variance in the phenotype is a sum of the genetic and environmental variances (assuming independence of  $G_i$  and  $Y_i$ ):

$$\text{Var}(Y_i) = \text{Var}(G_i) + \text{Var}(E_i). \quad (4.24)$$

We define the **heritability**,  $h^2$ , of trait  $Y$  as

$$h^2 = \frac{\text{Var}(G_i)}{\text{Var}(Y_i)}. \quad (4.25)$$

In words, heritability measures the fraction of phenotypic variance that is due to additive genetic effects. By definition,  $h^2$  ranges between 0 (no genetic variance) and 1 (only genetic variance)<sup>1</sup>.

<sup>1</sup> This definition of heritability is sometimes known as **narrow-sense** to distinguish it from a related concept, **broad-sense heritability**,  $H^2$ , which we'll see shortly.

For example, the variance in height in European populations is around  $50 \text{ cm}^2$  (within sex), and the heritability is estimated at around 60–80%. This implies that additive genetic variance for height is around 30–40  $\text{cm}^2$ , while the remaining variation is driven mainly by environmental effects such as differences in childhood nutrition, random effects in development, and measurement error<sup>717</sup>.

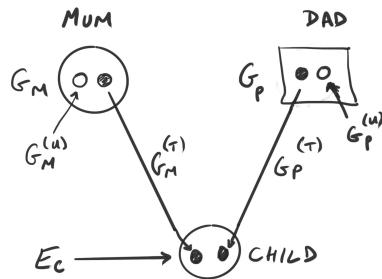
It's often easier to think about standard deviations of a trait than variances: in this case, the standard deviation of phenotypic variance is about 7 cm (remember, standard deviation is the square root of variance<sup>718</sup>). The genetic effects  $G_i$  have a standard deviation around 5.5–6.5 cm. For a normal distribution, about 95% of the distribution lies within 2 standard deviations of the mean, so we can interpret this to imply that 95% of people have  $G_i$  for height between about -12 cm and +12 cm.

**Modeling the similarity between relatives.** I promised you earlier that we'd be able to model the similarity in phenotype among relatives. Let's take a crack at that now.

First we ask:

*What's the correlation between the phenotypes of parents and their child?*

**Optional details:** To model this we need a bit of notation. We'll use subscripts  $m$ ,  $p$ , and  $c$  to indicate maternal, paternal, and child, respectively. Next, for each parent, we will denote the half of their genome that is *transmitted* to the child as  $G^{(t)}$ , and the half that is *untransmitted* to the child as  $G^{(u)}$ .



**Box Figure: Modeling phenotypes in parents and their child.** Conceptually, we split the genome of each parent into the part that is transmitted to a particular child, and the part that is untransmitted. So for example, the polygenic score of the mother  $G_m$  is a sum of two haploid components  $G_m^{(t)} + G_m^{(u)}$ . (The superscripts  $(t)$  and  $(u)$  are used here as labels not exponents.)

Then the phenotypes of the three family members are

$$\begin{aligned} Y_m &= \bar{y} + G_m^{(t)} + G_m^{(u)} + E_m \\ Y_p &= \bar{y} + G_p^{(t)} + G_p^{(u)} + E_p \\ Y_c &= \bar{y} + G_m^{(t)} + G_p^{(t)} + E_c \end{aligned} \tag{4.26}$$

We're now ready to compute the covariance between the mother (or father's) height and the child. This is:

$$\text{Cov}(Y_m, Y_c) = \text{Cov}(G_m^{(t)} + G_m^{(u)} + E_m, G_m^{(t)} + G_p^{(t)} + E_c) \tag{4.27}$$

Assuming that all the random terms are independent of one another, there's one shared term between the mother and child ( $G_m^{(t)}$ ), and the other terms have covariance 0:

$$\begin{aligned}\text{Cov}(Y_m, Y_c) &= \text{Var}(G_m^{(t)}) \\ &= \text{Var}(G_m)/2.\end{aligned}\tag{4.28}$$

For expanded derivations for this section see endnote <sup>719</sup>.

We can now compute the *correlation between a parent and the child*, which gives a pleasing connection to the heritability  $h^2$  that we defined above:

$$\text{Corr}(Y_m, Y_c) = \frac{h^2}{2}\tag{4.29}$$

What if we want to predict the child from the average ('midpoint') of the two parents, as in Galton's analysis of height? We'll write the midpoint of the two parents as  $Y_{mid}$ . Now

$$\text{Corr}(Y_{mid}, Y_c) = \frac{h^2}{\sqrt{2}}.\tag{4.30}$$

An easier way to think about this is in terms of the slope of Galton's plot. What is the slope when we plot the midparent phenotype on the x-axis and child's phenotype on the y-axis? A short calculation shows that

$$\text{Slope of } Y_{mid} \text{ vs } Y_c = h^2.\tag{4.31}$$

**This last expression reveals a remarkable result: the plots that Galton made in 1886 are directly related to a fundamental measure on inheritance: the heritability  $h^2$ !** If we make plots in the style of Galton's height plot, the slope of the plot is the heritability. Here are simulated data for traits with different heritabilities:

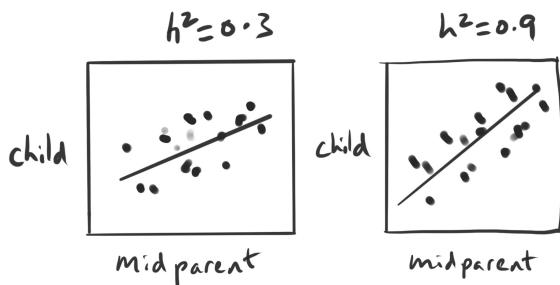


Figure 4.40: Simulated phenotypes for parent-child trios. Note that the slope is always  $\leq 1$ , and that there is always unpredictable scatter around the line even when  $h^2 = 1$ .

These results explain two intriguing features of Galton's analysis:

- Galton noted the 'regression to the mean', in which offspring tend to be closer to the mean than their parents. Kids with tall parents tend to be shorter than their parents, and vice versa. We can understand this as follows: parents who are extremely tall probably have both high  $G$  AND high  $E$ . Only  $G$  is transmitted to the child, while the child gets a random

$E$  with mean zero. So most children with tall parents will have a smaller  $E$  than their parents (and hence be shorter than their parents) <sup>720</sup>.

- Second, there is a great deal of variation around the line. This occurs in the model even when  $h^2 = 1$ . This is because each parent carries a mixture of trait-increasing and trait-decreasing variants. Even if the parent has a preponderance of trait-increasing alleles, there is still randomness in the precise fraction of trait-increasing alleles each parent transmits to the child.

*All the theory that we've developed so far was developed in the first half of the 20th Century, long before the modern genome era. And the first main applications came not in human genetics but in agriculture. So in the next section we'll take a bit of a detour to see how these principles have shaped our livestock, our meat, and the plants we eat.*

**Quantitative genetics in agriculture.** Suppose we want to breed from a particular individual – What is the expected phenotype of the offspring? For example, the heritability for lifetime success of racehorses has been estimated at around 0.3–0.4. Should a breeder pay extra to breed from a successful horse? Breeders certainly think so: it regularly costs \$100,000 or more to breed a mare with a top stallion <sup>721</sup>.

Similarly, in dairy cattle, the ‘quality’ of bulls is measured by the milk production of their daughters. One famous Holstein bull, named Carlin-M Ivanhoe Bell born in 1974, had such good genetics that he was used to sire more than 80,000 offspring. Due to selective breeding of bulls, more than 99% of the US Holstein herd of around 9 million animals is descended from just two males <sup>722</sup>!

We can model this using the theory above. What’s the expected phenotype of the child given the phenotype of either one parent, or both parents? This is known as the **breeding value** of a parent. We can read the breeding value directly off the plot below:

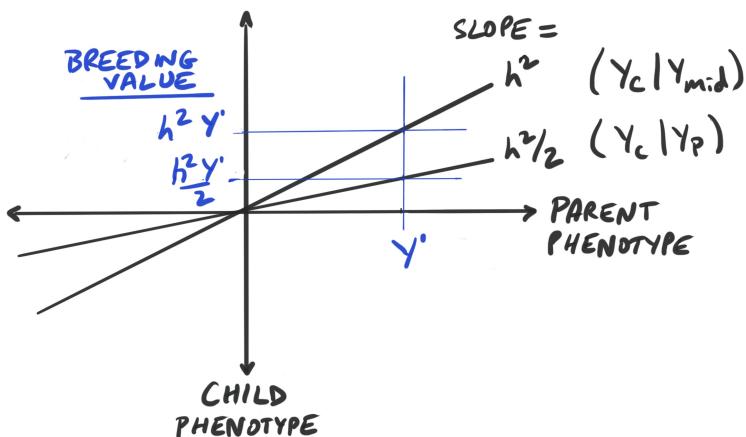


Figure 4.41: ‘Justify’ winning the 2018 Belmont Stakes. Stud fees for Justify the following year were set at \$150,000 per live foal [\[Link\]](#).

Credit: Mike Luzzi [\[Link\]](#) CC BY SA 2.0

Figure 4.42: Illustration of Breeding Value. The two lines show the expected child phenotype as a function of either the midparent or a single parent phenotype. The values in blue show the expected outcome given a specific parental phenotype  $Y'$ . To simplify the plot,  $\bar{y}$  has been set to 0.

The breeding values are:

$$E[Y_c|Y_{mid}] = \bar{y} + h^2(Y_{mid} - \bar{y}) \quad (4.32)$$

$$E[Y_c|Y_p] = \bar{y} + \frac{h^2}{2}(Y_p - \bar{y}) \quad (4.33)$$

In words, the expected phenotype of the child shifts partway toward that of their parent(s) by an amount proportional to the heritability.

**Selective breeding on a quantitative trait.** This result about breeding values allows us to derive the most important equation in quantitative genetics<sup>723!</sup> Imagine an animal breeder who wants to improve the properties of their herd, for example to increase meat or dairy production or decrease time to maturity. Suppose that they only breed from animals that exceed some threshold. What is the expected value of the phenotype in the next generation?

To answer this, suppose that the mean value in the herd is  $\bar{y}$ . The breeder only selects animals for breeding with high phenotype values – we'll use  $\bar{y} + S$  to indicate the mean phenotype among those animals selected to breed. Now the midpoint phenotype varies across pairs of parents, but since the overall mean phenotype is  $\bar{y} + S$ , the mean parent midpoint must also be  $\bar{y} + S$ . Then using the breeding-value equation above, we can compute the average phenotype in the next generation ( $E[Y']$ ):

$$E[Y'] = \bar{y} + h^2((\bar{y} + S) - \bar{y}). \quad (4.34)$$

This is traditionally written in terms of the *response to selection R*, where  $R$  is the change in mean phenotype in a single generation due to selection ( $E[Y'] - \bar{y}$ ). This produces the most famous result in quantitative genetics, **The Breeder's Equation:**

$$R = h^2S \quad (4.35)$$

Here's a sketch of this:

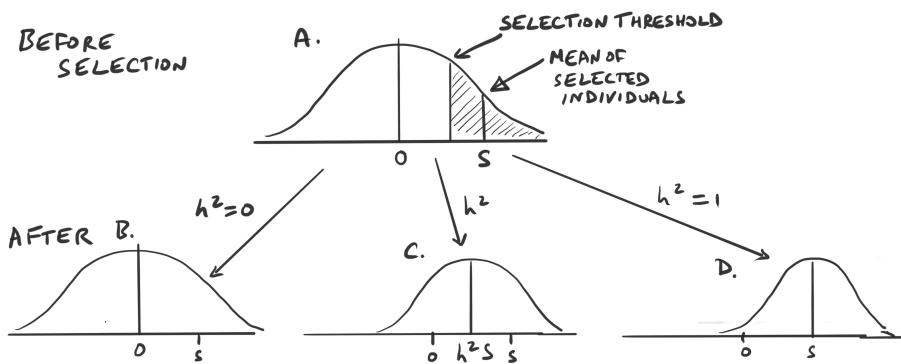


Figure 4.43: The Breeder's Equation.

A. Distribution of phenotypes before selection. Individuals above the threshold reproduce to form the next generation. B., C., D. After selection the new population is shifted by  $h^2S$  to the right. [define R on plot; change 0 to  $\bar{y}$ ?]

When the Breeder's Equation is iterated across successive generations, it can result in remarkable changes in phenotype. In agriculture, this process is known as **selective breeding**.

Selective breeding is one of the most enabling processes in human history. Starting around 9,000 years ago, the invention of agriculture allowed higher population densities and the formation of towns and cities. Even early farmers implemented forms of selective breeding, greatly increasing the productivity of a wide range of domesticated plants and animals. Modern agriculture has greatly accelerated selective breeding at industrial scales, to feed a global population of 8 billion people.

One example, for **maize**, is shown at the right. Maize is derived from a bushy grass called **teosinte** that has small ears of grains. Teosinte was first domesticated in Mexico around 9,000 years ago, and spread throughout the Americas from there. Early selective breeding changed the form of the plant to a single main stem and distinct cobs similar to modern maize. Over the following thousands of years, selective breeding drove dramatic increases in the sizes of the cobs and kernels.<sup>724</sup>

Selective breeding continues to be a major focus of agriculture, and most agricultural crops and animals have seen enormous increases in performance in the last century.

For example, starting in the late 1930s, maize became the focus of selective breeding programs by the leading seed companies; since then maize yields have increased around 5-fold.<sup>725</sup> About half of these yield gains are due to genetic improvement (G), and half to improvements in agricultural practice such as better pest control and soil management (E).<sup>726</sup><sup>727</sup>

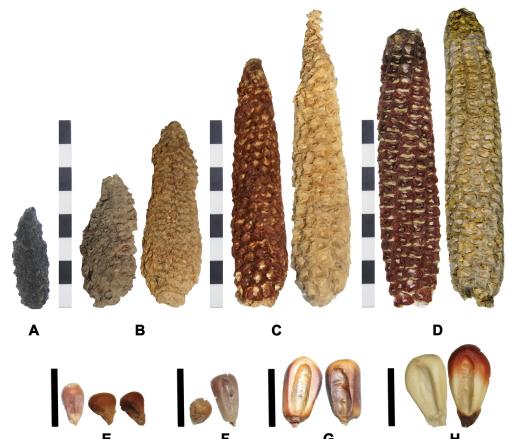
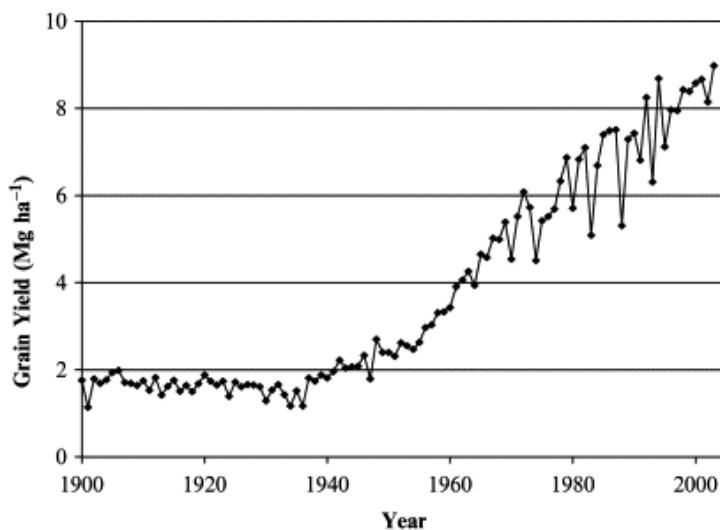


Figure 4.44: **Selective breeding of maize over 2000 years.** Maize cobs and kernels from archaeological sites in Chile ranging from 2000 to 500 years old, as well as contemporary maize (left to right).

Credit: Figure 2 from Ale Vidal Elgueta et al (2019) [[Link](#)] CC BY 4.0

Figure 4.45: **Maize yield increases during the last century (US).** Industrial breeding programs were implemented starting in the mid-1930s. Yield in metric tons per hectare. Credit: Figure 1 from Donald Duvick (2005) [[Link](#)] Request Permission.

**Genomic selection.** For nine thousand years, farmers have conducted selective breeding by choosing individuals with desirable *phenotypes*. But this is inefficient because the parental phenotypes also reflect the random contributions of environment (E). This is why – as Galton noticed – the phenotypes of children tend to regress back toward the mean.

But what if we could actually select animals based on their genetic scores (G)? This would be *much* more efficient! In fact, if we could compute G precisely, then the breeding response each year can be predicted from the

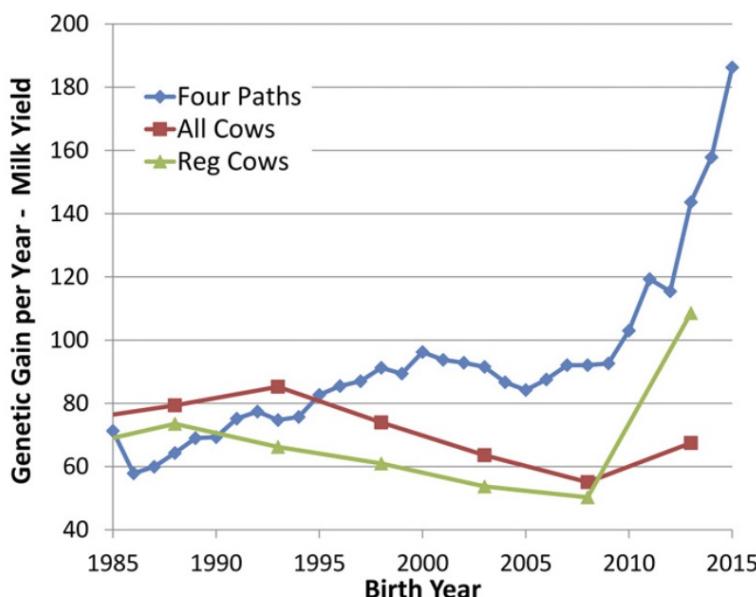
Breeder's Equation with  $h^2 = 1$  (Equation 4.35).

A highly influential 2002 paper, by Meuwissen et al, proposed that it would be possible to use genetic markers to build genetic predictors of  $G$  for cattle and other agricultural species<sup>728</sup><sup>j</sup>.

Estimating  $G$  also solves a major practical issue for breeders of long-lived animals such as pigs and cattle, which is that they can predict phenotype at birth, instead of having to wait until the animal reaches adulthood to assess its value. This problem is even more profound for identifying top dairy bulls: in this case the breeder actually needs to measure the milk production of his *daughters!* Before the use of genomic selection, it took around 7 years to determine which bulls had high quality genetics for milk production, at a cost of \$500,000 per selected bull<sup>729</sup>.

And another related advantage is that genomic selection can greatly speed up the breeding response by reducing generation time, because animals can be bred from as soon as they reach sexual maturity, without needing to wait for phenotypic measurement<sup>730</sup>.

High density genotyping arrays for cattle became available in 2008, and genomic selection rapidly became an essential breeding tool. The plot below shows estimated genetic improvement for milk production in different groups of US Holstein cattle<sup>731</sup><sup>732</sup>. Look at the increased rate of improvement since 2008:



<sup>j</sup> Genomic prediction in animals is conceptually similar to prediction from polygenic scores in humans (see Figure 4.39, above, and Chapter 4.6).

Figure 4.46: Estimated genetic gain per year (milk) increases after the onset of genomic selection in 2008. Estimated  $G$  for milk production is improving every year across the plot, but shows a marked upward elbow starting in 2008. The three lines correspond to different subgroups of cattle. Credit: Figure 3 Adrian Garcia-Ruiz et al (2016) [Link] Reuse permitted

I hope you have enjoyed this brief detour on quantitative genetics in agriculture. We can now return to your regular scheduled programming.

**Limitations of the basic additive model.** It's now time for me to confess that some of the assumptions we made above may not always hold in real settings<sup>k</sup>. The first two are especially relevant for humans.

- **Environment is rarely independent between relatives.** In humans, key

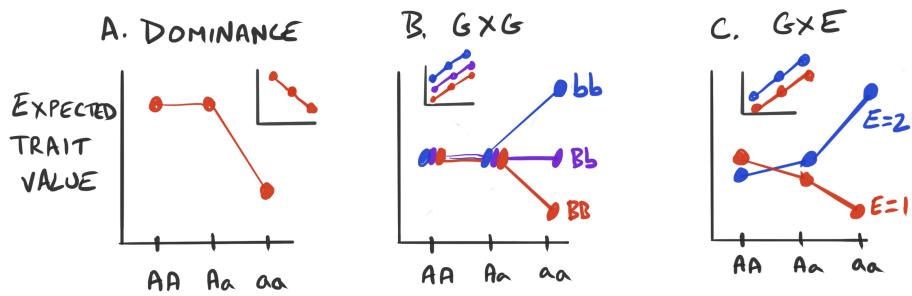
<sup>k</sup> Even with limitations, quantitative genetics is still an essential tool for understanding inheritance, but we should always consider which assumptions are violated in any particular setting, and how this may affect outcomes.

aspects of environment such as household wealth, education, diet, and behavior are highly shared within families. *Shared environment increases the phenotypic correlation among relatives, sometimes to a huge degree.*

- **Parental genotypes may not be independent.** We often choose partners who are similar to ourselves along various axes including height, weight, behaviors, and education. This is called **assortative mating**. To the extent that these traits are heritable, this creates positive correlations between  $G_m$  and  $G_p$ . Moreover, assortative mating in previous generation(s) creates positive correlations between the transmitted and untransmitted genomes within each parent. In summary, *assortative mating increases the phenotypic variance in the population, and increases the correlation between parents and children.*

- **Interactions: dominance, gene-by-gene, gene-by-environment.**<sup>1</sup> So far we've been focusing on a relatively simple model where allelic effects combine as a sum of individual effects to generate the genetic score  $G_i$ .

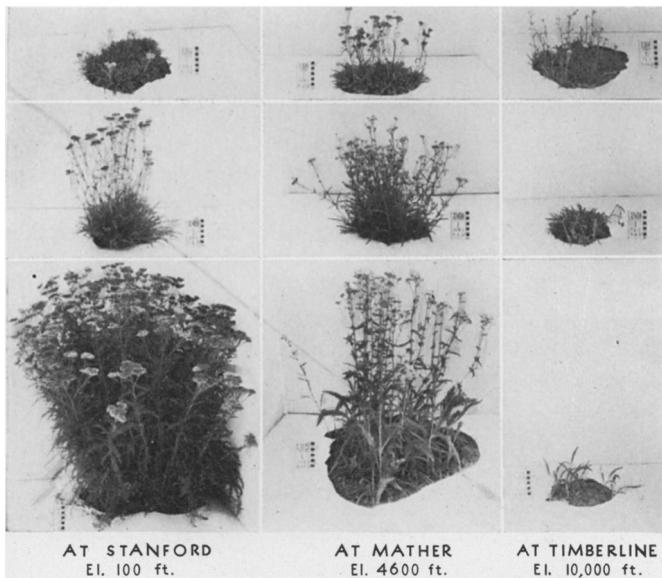
This means that we ignore a variety of possible interaction effects, including between alleles (called dominance effects), between loci (called  $G \times G$  effects, or **epistasis**), and between genotype and environment (called  $G \times E$  effects):



Dramatic examples of  $G \times E$  were found in pioneering studies of native plants by a team based at Stanford in the 1940s: Clausen, Keck and Hiesner. They took advantage of the fact that the elevation in California ranges from sea level in the west to more than 3,000 meters in the mountains in the east of the state. They collected wild plants of particular species from natural habitats at different elevations, and then grew them in shared gardens. One example is shown below: three cultivars of the same species show clear evidence of local adaptation, growing better close to their original source elevations:

<sup>1</sup> Interaction effects refer to any type of combination of allele/genotype/environment effects on the expected trait value that cannot be fit by sums of independent effects. In a deplorable clash of notation, by convention the  $G$  in ' $G \times G$ ' and ' $G \times E$ ' refers to 'gene', and is different from the genetic score  $G_i$  that we use elsewhere in this chapter.

Figure 4.47: Examples of hypothetical interaction effects. A. The plot shows trait values as a function of genotype at one locus. In this example, A is dominant with respect to a. B. Trait values for 9 combinations of genotypes at two different loci. Here, the BB/Bb/bb genotype has affects the trait only when it is coupled with aa at the other locus. C. Genotype effects conditional on whether an individual inhabits environment 1 or 2. Notice that the genotype-phenotype relationship is opposite depending on the environment. The insets show examples of additive models (i.e., without interactions).



**Figure 4.48: GxE interaction (Clausen et al 1941).** Plants of the species *Achillea lanulosa* cultivated from different elevations (rows) were grown in common gardens at three elevations (columns). The high-elevation cultivar (top row) grew poorly at sea level (left) but better at medium and high elevation. The low-elevation cultivar (bottom row) was delightfully happy at sea level but a failure in the mountains. Cultivar source elevations (top to bottom): 10,700 ft, 7,100 ft, 4,675 ft. Credit: Figure 6, Jens Clausen et al 1941. [[Link](#)].

We refer to this as a  $G \times E$  interaction, because we cannot model the phenotype (plant size) as a sum of independent effects from  $G$  and from  $E$ . Here ‘ $G$ ’ likely comprises the effects of local adaptation at many genes, rather than a single variant as shown in the sketch above. *Importantly, it is much harder to find clear examples of  $G \times E$  in humans, and there is active debate about how widespread this is in our own species.*

**Variance partitioning and broad sense heritability.** We can formalize some of these ideas into a more explicit model that **partitions the variance** into different types of genetic and environmental effects:

We previously wrote a simple model (Equation 4.16) for an individual  $i$ ’s phenotype as

$$Y_i = \bar{y} + G_i + E_i \quad (4.36)$$

in which  $G_i$  was a sum of additive effects. Let’s now expand the concept of  $G_i$  to include non-additive effects as follows:

$$Y_i = \bar{y} + \underbrace{A_i + D_i}_{\text{Genetic Effects}} + \underbrace{\{G \times G\}_i + \{G \times E\}_i}_{G \times E} + E_i. \quad (4.37)$$

where  $A_i$ ,  $D_i$ ,  $\{G \times G\}_i$ , and  $\{G \times E\}_i$  represent contributions from additive, dominance, gene-gene, and gene-environment interactions, respectively. You can think of these new terms as helping us to fit residuals in the genotype→phenotype model that are not captured by the basic additive model.

We previously defined the **narrow-sense heritability**,  $h^2$  as the fraction of total phenotypic variance explained by additive genetic effects; in the expanded notation this is:

$$h^2 = \frac{\text{Var}(A_i)}{\text{Var}(Y_i)} \quad (4.38)$$

It would also be natural to ask: *What fraction of phenotypic variance is due to any kind of genetic effect?* This quantity is called **broad-sense heritability**, denoted  $H^2$ , and is usually written as<sup>733</sup>:

$$H^2 = \frac{\text{Var}(\text{total genetic effects})}{\text{Var}(Y_i)} \quad (4.39)$$

$$= \frac{\text{Var}(A_i) + \text{Var}(D_i) + \text{Var}(\{G \times G\}_i)}{\text{Var}(Y_i)} \quad (4.40)$$

Broad sense heritability provides a more complete measure of the role of genetics in determining phenotype, but *narrow sense heritability is usually the correct measure for understanding transmission of phenotypes among relatives, and the response to selection.* This is because we inherit recombined haploid genomes from our parents (and not diploid genotypes), so interaction effects are not very transmissible.

### Closing Remarks.

1. *Despite all these caveats, the additive model provides a surprisingly good approximation for complex traits in humans.* There are few clear examples of dominance or GxG among common variants – which drive the bulk of the heritability for complex traits<sup>734 735</sup>.
2. Similarly, *there are few clear examples of G×E effects in humans*<sup>736</sup>. This is partly because we lack power to detect weak G×E signals, since we cannot perform experimental manipulations in humans, and perhaps partly because large G×E effects are quite limited in our species<sup>737</sup>.
3. In contrast, interaction effects, including dominance, G×G, and G×E are often much more evident in other species, both in the wild (as for the California wildflowers above) and in agricultural systems.

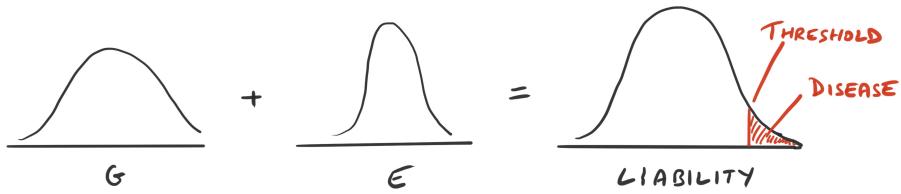
**Yet Another Optional Interlude: The Liability Threshold model for binary traits.** So far, we've been focusing on continuous models of quantitative traits. But in human genetics we're often interested in so-called binary outcomes, particularly for disease: Does an individual have diabetes, or not? Do they have heart disease? Do they have schizophrenia (see Table 4.2 for data on co-occurrence of schizophrenia in different types of relatives.)?

All these diseases have a significant genetic component; how can we model this?

We won't go into great detail, but we can sketch out the basic idea. Let's modify our basic model for inheritance, Equation 4.16, to output a quantity that we call **liability**. Individual  $i$  will have a liability  $L_i$ :

$$L_i = \bar{l} + G_i + E_i. \quad (4.41)$$

The key idea is that liability captures some unobserved quantitative phenotype that determines whether someone will get the disease. Specifically, we imagine that there is some threshold  $T$  such that everyone above  $T$  has the disease, and everyone below  $T$  does not. Here's a sketch of this:



**Box Figure: The Liability Threshold model.** In this model, individuals have an unobserved quantitative phenotype called 'liability', which follows the usual inheritance of quantitative traits. Individuals whose liability is above a particular threshold are diagnosed with disease.

For example, the clinical definition of obesity is a body mass index (BMI)  $> 30$ . So we can think of BMI as measuring the liability scale, and 30 serving as the threshold for obesity.

But, for most diseases, you should think of liability as a theoretical construct that contains a mixture of known and unknown drivers of disease from genetics plus environment and chance<sup>738</sup>.

The last key point is that, because liability is assumed to be a standard quantitative trait, all the standard quantitative genetics models about inheritance between relatives work for liability. But the modeling of disease inheritance is a bit more complicated because we need some math to account for the thresholding effect. That's beyond our scope here.

**Estimating heritability.** Now that we've developed the main mathematical principles, we can turn our attention to how to estimate heritability in real life – and what heritability actually measures.

We saw above that, for our basic model, the phenotype correlations between relatives depend in a simple way on the narrow sense heritability,  $h^2$  (Equation 4.31).

But when we derived this, we ignored the fact that **environment is also usually shared among relatives**. Not only do you share genes with your family, but you probably also share a household, diet, social and economic factors. So if we simply measure the correlation among relatives, genes and environment are hopelessly confounded with each other. For this reason, *there has been a huge amount of work on how to get reliable estimates of heritability that can separate inheritance of genetics and environment, but this remains a difficult and often controversial problem*.

**The Twin Design.** The classic approach to estimating heritability uses twins as a way to try to untangle genetics and environment. The central concept in a twin design is to compare the phenotypic correlation of "identical" (MZ) twins, with fraternal (DZ) twins. MZ twins have essentially identical genomes, while DZ twins share 50% of their genomes, like ordinary siblings. Crucially, both types of twins are born at the same time and grow up together. **So we might plausibly assume that MZ share twice as much genome as DZ twin – but the same amount of environment.**

If we make this assumption, we can derive a famous result known as **Falconer's Equation**, which estimates  $h^2$  from twin data as:

$$\hat{h}^2 = 2(\rho_{MZ} - \rho_{DZ}). \quad (4.42)$$



Susan and Angela Tooby

Figure 4.49: What role does genetics play in the phenotypic similarity of relatives? The "identical" twin Tooby sisters were both British Olympians in 1988 (at 10,000m and the marathon). Susan's son Jake Wightman would become the 2022 world champion at 1500m.

Credit: Athletics Weekly, January 14th, 1984.

Falconer's Equation is pleasingly simple, but it's important to note that this estimator can be biased if (1) there are non-additive genetic effects<sup>739</sup>; (2) parents treat MZ twins more similarly than DZ twins<sup>740</sup>; (3) if there is assortative mating<sup>741</sup>.

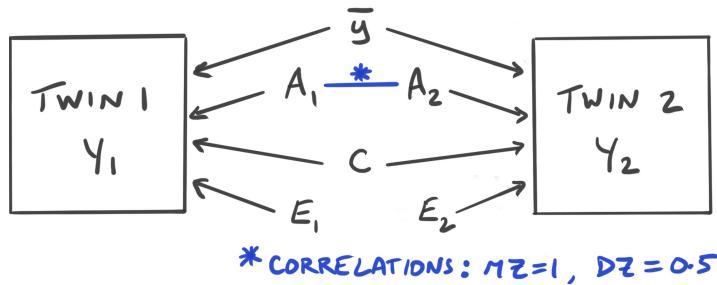
But despite these limitations, for some 50 years this was our best tool for measuring heritability. In a minute I'll show how this can be checked against a genetic approach.

**Optional Derivation of Falconer's Equation.** Let's write a simple model for the phenotypes  $Y_1$  and  $Y_2$  of two twins. This will be similar to Equation 4.37, but with three random components:  $A$ ,  $C$ , and  $E$ , hence the name **ACE model**:

$$Y_1 = \bar{y} + A_1 + C + E_1 \quad (4.43)$$

$$Y_2 = \bar{y} + A_2 + C + E_2 \quad (4.44)$$

Here,  $A$  is the additive genetic component (this model ignores non-additive components). We split the environmental contribution into parts:  $C$ , which is identical between the twins, and  $E$ , which is independent. The model is illustrated below:



**Box Figure. The ACE Model of phenotypic effects in twins.**  $A$ ,  $C$ , and  $E$  are random components that determine each twin's phenotype. The extra similarity of MZ twins compared to DZ twins is due to the complete sharing of  $A_1$  and  $A_2$  in MZ twins.

Then the correlations between the phenotypes of MZ, and DZ twins, respectively, are<sup>742</sup>:

$$\rho_{MZ}(Y_1, Y_2) = [1 \cdot \text{Var}(A) + \text{Var}(C) + 0 \cdot \text{Var}(E)]/\text{Var}(Y) \quad (4.45)$$

$$\rho_{DZ}(Y_1, Y_2) = [0.5 \cdot \text{Var}(A) + \text{Var}(C) + 0 \cdot \text{Var}(E)]/\text{Var}(Y) \quad (4.46)$$

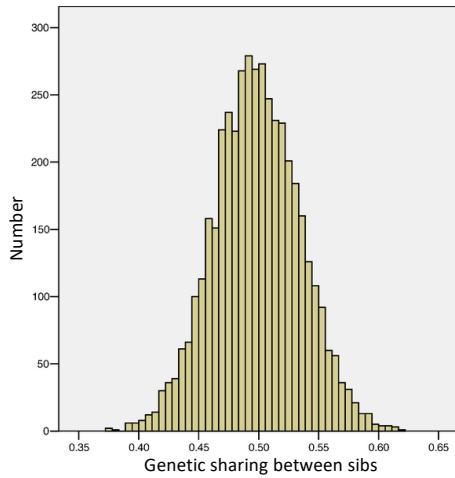
By subtraction we get

$$\rho_{MZ} - \rho_{DZ} = 0.5 \cdot \frac{\text{Var}(A)}{\text{Var}(Y)} \quad (4.47)$$

The right-hand side is simply  $0.5 \times h^2$ , which then suggests Falconer's estimator:  $\hat{h}^2 = 2(\rho_{MZ} - \rho_{DZ})$ .

**Genetic-based estimation of  $h^2$  in relatives.** By the early 2000s, there were many hundreds of twin studies, but no clear way to assess the accuracy of the heritability estimates.

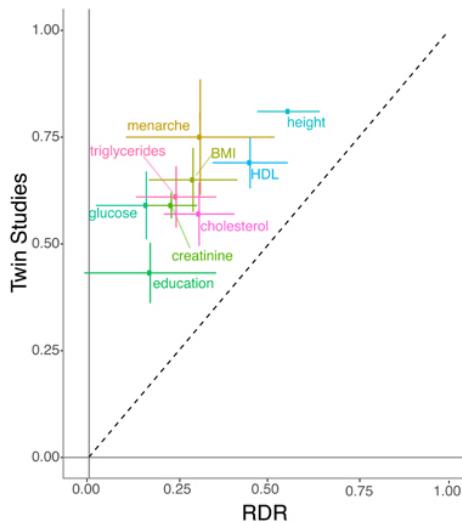
But in 2006, Peter Visscher and colleagues developed a clever trick to use genetic data from siblings to get an entirely different estimate of heritability<sup>743</sup>. We know that, on average, siblings share 50% of their genomes. But due to the vagaries of meiosis, some sibling pairs actually share slightly more than 50%, and some share slightly less. Here's the estimated distribution of genetic sharing between sibs, from the 2006 paper:



**Figure 4.50: Distribution of genetic sharing between siblings.** The actual amount of shared genome between two siblings is 50% on average, but varies from around 40–60% due to the randomness of meiosis. Credit: Figure 1 from Peter Visscher et al (2006) [Link] CC BY

The key idea is that *even though siblings share environment, the precise amount of shared environment should be independent of genetic sharing*. And if a trait is heritable, siblings with higher genetic sharing should have more similar phenotypes. This logic suggests a simple estimator of heritability based on regressing the squared difference in sib phenotype against proportion of genome sharing<sup>744</sup>.

The initial application was a bit underpowered, but in 2019, Alex Young and colleagues extended this idea to consider more types of relatives, in a larger sample<sup>745</sup>. The plot below from their paper compares their estimates (RDR) versus classical estimates from twin studies for a range of traits:



**Figure 4.51: Estimated heritabilities using IBD sharing between relatives versus twin studies.** This analysis finds that twin studies systematically overestimate heritability. For many complex traits, current estimates of  $h^2$  are ~0.05–0.7. Credit: Figure 2d from Alexander Young et al (2019) [Link]

Replace this with free biorxiv version:  
[Link]

As you can see above, the RDR estimates are generally much lower than

twin-based estimates. Converging lines of evidence suggest that the RDR estimates and that *twin studies consistently over-estimate heritability*<sup>746</sup>.

These results also give us a sense of the heritability for a range of complex traits. These results (and other recent estimates) find that heritability for typical complex traits range from near zero to around 0.65 for height. Height is among the most heritable of all complex traits; the heritability of height is about twice that of body mass index (BMI), which I think is pretty intuitive given that body weight is modifiable throughout life depending on diet and exercise.

Traits that measure human behavior – anything from years-of-education to personality measures to hours-of-sleep-per night – typically have low, but non-zero, heritability, ranging from around 0.05 to 0.3 depending on the trait and study.

In Chapter 4.6, we'll see a third type of approach for estimating heritability, which is based on the SNP effects that can be measured in GWAS.

Early estimates of SNP heritability were much lower than people expected, leading to a controversy known as *The Mystery of The Missing Heritability* – as we'll see, that discrepancy has now been largely resolved (but you'll have to keep reading to learn exactly how!) <sup>m</sup>.

<sup>m</sup> *The biggest open question now is not "Where is the missing heritability?" but "Why are MZ twins so similar?", a question Sasha Gusev has termed the missing environment problem [Link].*

**The use and misuse of heritability.** *Heritability* is an essential parameter in quantitative genetics: it predicts the similarity among relatives; it determines the phenotypic response to selection in evolution and in agriculture; it places an upper bound on the future accuracy of polygenic prediction for complex traits. For all these reasons, there has been great interest in estimating heritability for many traits <sup>n</sup>.

But the concept of heritability has also been pulled into broader social debates. What is the role of genetics and environment in determining who we are (the “**nature versus nurture**” debate)? And even debates like: Can estimates of heritability inform us about the value of implementing enhanced educational programs for young children? Unfortunately, people often misinterpret and misuse concepts of heritability to try to score points in political debates.

So, in this section I'll provide some thoughts on what heritability can, and cannot teach us <sup>o</sup>:

**1. Heritability does not test whether a trait is controlled by genes, but measures the extent to which variation in a trait is driven by variation in genetics.**

For example, whether or not one speaks Hungarian is almost entirely determined by environment (i.e., whether one grew up around Hungarian speakers), and should therefore have heritability near zero. At the same time, the *ability* to learn human languages is genetically encoded – other animals simply cannot learn Hungarian, regardless of exposure.

**2. Heritability is a property of a sample.** Heritability is implicitly defined with respect to a particular group in a particular environment, so changing either of these can change heritability. For example, in animal breed-

<sup>n</sup> *Heritability,  $h^2$ , plays a role in quantitative genetics much like effective population size,  $N_e$ , in population genetics. Both  $h^2$  and  $N_e$  are essential in modeling, but the estimates are hard to interpret and subject to misuse.*

<sup>o</sup> *Heritability should not be thought of as a fundamental parameter like the speed of light, or even a biological parameter like the human mutation rate; instead heritability is highly context-specific, and can vary greatly due to non-biological factors.*

ing, heritability tends to go down over time due to loss of genetic variance.

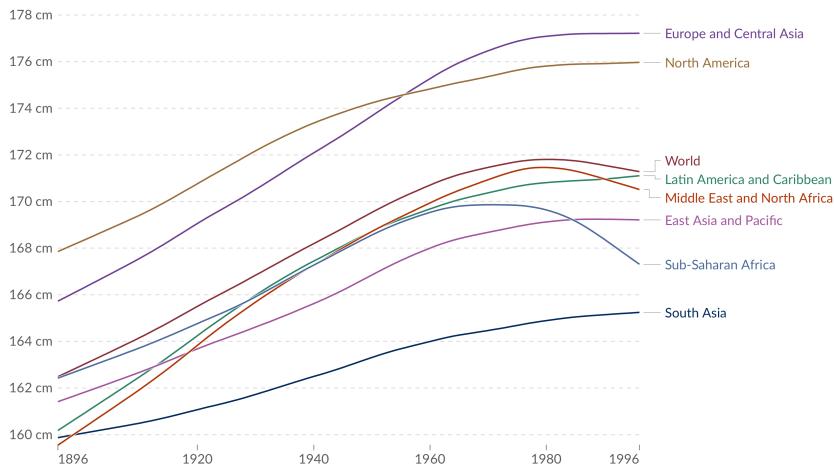
**3. Heritability depends on environment.** Since heritability measures the fraction of the phenotypic variance due to genetics, heritability is heavily influenced by the degree of variability in the environment. So for example, we can expect to find lower heritability in a highly varied environment (or for a trait with high measurement error) than in a very uniform environment.

**4. Even highly heritable traits can be changed dramatically by changing the average environment.** All human complex traits are influenced by environment, to a greater or lesser extent. Remember our basic model where

$$Y_i = \bar{y} + G_i + E_i.$$

So if we change the average environment  $E$  in this model, we'll have a direct impact on the average phenotype  $Y$ .

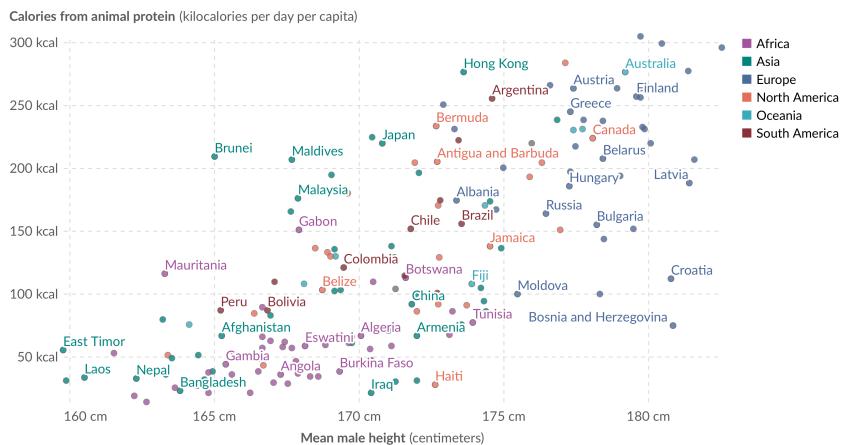
For example, height is the most heritable of all well-studied complex traits, with heritability around 0.6–0.8. But height is also strongly influenced by childhood nutrition, and changes in nutrition can have dramatic impacts on average height. As you can see, average height increased globally by around 5% on all continents in the 20th century <sup>747</sup>:



**Figure 4.52: Global changes in male adult height: birth years 1896–1996.** Data shown here are for males; females show similar increases. Credit: Max Roser et al, Our World in Data (2013, updated 2024) [[Link](#)] CC BY

These dramatic increases in height were likely driven by global gains in general health, caloric intake, and nutritional quality – i.e., access to nutritionally balanced diets that enhance childhood growth.

Indeed, we can get hints at the importance of environmental factors by noticing that, across countries, there is a clear relationship between measures of nutritional quality, as well as other correlates of economic development, and adult height <sup>748</sup>:



**Figure 4.53: Relationship between adult male height and consumption of animal proteins.** Note that other measures of economic development also show strong correlations with adult height, so this is not meant to imply that consumption of animal proteins, *per se*, is the (or the only) causal factor. Data shown here are for males born in 1996. Credit: Max Roser et al, Our World in Data (2013, updated 2024) [\[Link\]](#) CC BY

## 5. Humans are not an experimental organism, and some types of questions are fundamentally difficult to answer.

First, it is generally difficult to determine whether phenotypic differences between human groups are due to genetics or environment (and to what extent)<sup>749</sup>. We might notice that the residents of Sweden, say, are taller on average, than those of Bangladesh. Why?

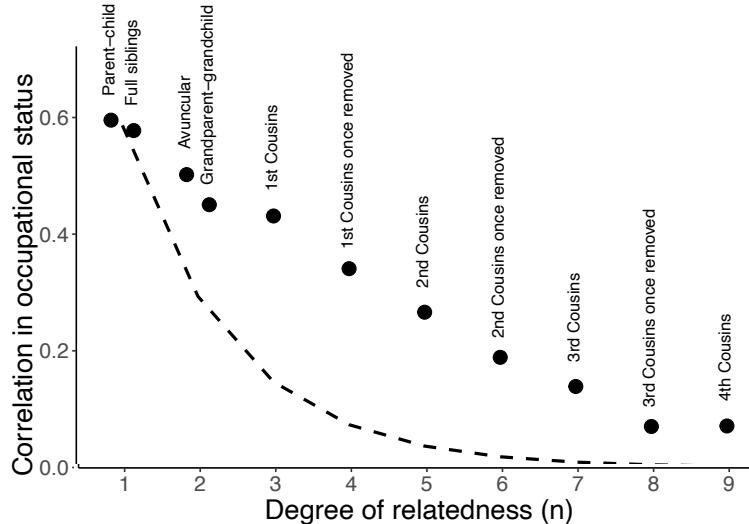
It may be disappointing to learn that these kinds of questions are generally very difficult to answer. Humans are not an experimental species, so we cannot take people from different populations and grow them in a single shared environment – as we might do for plant seeds collected in different locations. Instead, we can only address this type of question through indirect arguments. How heritable is the trait? Do relevant environmental factors differ substantially across populations? Does the average genetic effect,  $G$ , differ meaningfully across populations?

It turns out that even for a trait as well-studied as height, it is difficult to pin down exactly which environmental factors are causal, let alone to measure them precisely. And on the genetics side, there are serious technical challenges in comparing polygenic scores (i.e., estimates of  $G$ ) across populations and determining whether these are meaningfully different across populations. For these reasons it is still unresolved how much of the global variation in height comes from genetics – and we know more about the genetics of height than any other trait<sup>750</sup>.

*Second, some components of environment are passed down within families in a way that is extremely difficult to separate from genetics, even within populations.* This is especially true for **behavioral traits** such as educational attainment, intelligence, occupation and income – I think it's intuitive that most of us are strongly shaped by the family environments that we grow up in<sup>751</sup>.

Two recent studies of educational attainment estimated heritability at rather low levels, namely 8% and 17% that used RDR-style methods<sup>752</sup>; however, other estimates vary widely depending on the methods of controlling for shared environment<sup>753</sup>.

To show why we need to work so hard to control for culture that is passed on within families, consider the analysis below, showing how occupational status can be inherited across large familial distances – extending to 2nd, 3rd, or even 4th cousins – in historical data from England:



**Figure 4.54: Long-range persistence of occupational status in historical data from England.** The data points show pairwise correlations in occupational status between different types of relatives, ordered by degree of relatedness. The dashed line shows a basic heritability model starting from the observed correlation in first-degree relatives. Credit: Figure kindly provided by Arbel Harpak, based on Figure 3 of John Benning et al (2023) [[Link](#)] CC BY Ask arbel for corrected version

The observed patterns are incompatible with sensible genetic models, and instead reflect the long-term stability and inheritance of social class and culture within families<sup>754</sup>.

The study of behavioral genetics is not hopeless, but it is technically far more difficult than any other trait, while at the same time being more freighted with possibilities for misinterpretation and misuse<sup>755</sup>. It thus behooves us to treat these topics with care. The situation is summarized nicely in a 2019 article by Nick Barton and colleagues<sup>756</sup>:

*"(A)lthough quantitative genetics has proved highly successful in plant and animal breeding, it should be remembered that this success has been based on large pedigrees, well-controlled environments, and short-term prediction. ... Once we leave an experimental setting, we are effectively skating on thin ice, and whether the ice will hold depends on how far out we skate."*

Well done! You have completed your tour of quantitative genetics! We now turn our attention to genetic studies of complex traits, starting with introducing the GWAS paradigm.

## Notes and References.

<sup>701</sup>In preparing this chapter I benefited greatly from teaching slide decks by Guy Sella and Ziyue Gao. Thanks to the people who commented on earlier drafts of this chapter: Hakhamanesh Mostafavi and Molly Przeworski. Many thanks to Karl Kremling and Jesse Hoff for helpful discussions on selective breeding in agriculture. As always, any errors are my own.

<sup>702</sup>We cover Galton's work on inheritance because it's important in the history of genetics but Galton, and Fisher whom we'll get to shortly, also have troubled legacies as originators of eugenics.

<sup>703</sup>Galton 1886

<sup>704</sup>Galton's explanation for regression to the mean (Galton 1866, p252): "The explanation of it is as follows. The child inherits partly from his parents, partly from his ancestry. Speaking generally, the further his genealogy goes back, the more numerous and varied will his ancestry become, until they cease to differ from any equally numerous sample taken at haphazard... Or, to put the same fact into another form, the most probable value of the mid-ancestral deviates in any remote generation is zero."

<sup>705</sup>Bateson 1902 [Link]. Bateson had a pretty good career, coining the term 'genetics', co-discovering linkage, and originating part of the Bateson-Muller-Dobzhansky incompatibility model of speciation

<sup>706</sup>As cited by Visscher and Goddard 2019 [Link], which is a good reference source for this section.

<sup>707</sup>Recall also that Hardy's landmark paper on the Hardy-Weinberg principle was rejected from Nature in 1906 for being "tainted" with Mendelism (Chapter 1.3)

<sup>708</sup>Barton et al 2017

<sup>709</sup>As usual, any of our familiar allele-labeling system works here, eg reference/non-reference, major/minor, alternate/derived. Flipping the labeling changes the sign of  $\beta$  so  $g\beta$  is unaffected by the labeling.

<sup>710</sup>There are a bunch of challenges to implementing this in practice, including that we don't know in advance which SNPs have causal effects, nor the true  $\beta$ s. We'll address these issues in the upcoming chapters.

<sup>711</sup>Although it was known for a long time that many human traits are polygenic, it was a surprise to learn in the 2010s that many traits are influenced by 10,000+ SNPs; Boyle 2017

<sup>712</sup>Technically this arises as  $S \rightarrow \infty$  while holding  $\text{Var}(G_A)$  constant.

<sup>713</sup>Barton 2017.

<sup>714</sup>Weedon et al 2008 [Link]

<sup>715</sup>This represents the difference between the highest and lowest of 12 equal-sized bins of polygenic score. Figure 2, Abdellaoui 2023 [Link]

<sup>716</sup>To be more precise,  $E(y)$  is defined as

$$\int_{-\infty}^{\infty} y \Pr(y) dy. \quad (4.48)$$

<sup>717</sup>An additional source of variance is that height is subject to strong assortative mating as tall individuals are more likely to have tall partners, and conversely for short individuals. This increases the population variance. This calculation also ignores any possible role from non-additive genetic effects such as dominance or epistasis.

<sup>718</sup>Remember that standard deviations are simply the square root of variances. Variances are easier to work with for mathematical calculations because variances can often be computed as sums of variances and covariances, but it's usually easier to interpret standard deviations as they use the original measurement units instead of using squared units.

<sup>719</sup>We're now ready to compute the covariance between the mother (or father's) height and the child. This is:

$$\text{Cov}(Y_m, Y_c) = \text{Cov}(G_m^{(t)} + G_m^{(u)} + E_m, G_m^{(t)} + G_p^{(t)} + E_c) \quad (4.49)$$

Assuming that all the random terms are independent of one another, this reduces to:

$$\begin{aligned} \text{Cov}(Y_m, Y_c) &= \text{Cov}(G_m^{(t)}, G_m^{(t)}) \\ &= \text{Var}(G_m^{(t)}) \\ &= \text{Var}(G_m)/2, \end{aligned} \quad (4.50)$$

noting that the variance of  $G$  for the transmitted half of the genome is one half the variance of  $G$  for the total genome. Finally, we compute the correlation of parent and child by dividing by the standard deviations of  $Y_m$  and  $Y_c$ , per the definition of correlation:

$$\text{Corr}(Y_m, Y_c) = \frac{\text{Var}(G_m)/2}{\sqrt{\text{Var}(Y_m)} \sqrt{\text{Var}(Y_c)}}. \quad (4.51)$$

Since  $Y_m$  and  $Y_c$  have the same variance we can rewrite this as

$$\begin{aligned} \text{Corr}(Y_m, Y_c) &= \frac{\text{Var}(G_m)}{2 \cdot \text{Var}(Y)} \\ &= \frac{h^2}{2} \end{aligned} \quad (4.52)$$

We can extend this theory to an arbitrary relative that shares an expected fraction  $r$  of their genome with a specific individual  $i$ . For example, two cousins share  $r = 1/8$ th of their genomes, on average. It's easy to show by a similar argument that

$$\text{Corr}(Y_r, Y_i) = rh^2. \quad (4.53)$$

Next, what if we want to predict the child from the average ('midpoint') of the two parents, as in Galton's analysis of height? We'll write the midpoint of the two parents as  $Y_{mid}$ . Now

$$\text{Cov}(Y_{mid}, Y_c) = \text{Cov}((G_m^{(t)} + G_m^{(u)} + E_m + G_p^{(t)} + G_p^{(u)} + E_p)/2, \quad (4.54)$$

$$G_m^{(t)} + G_p^{(t)} + E_c) \quad (4.55)$$

Assuming that all the random terms are independent of one another, this reduces to:

$$\begin{aligned} \text{Cov}(Y_{mid}, Y_c) &= \text{Cov}((G_m^{(t)} + G_p^{(t)})/2, G_m^{(t)} + G_p^{(t)}) \\ &= \text{Var}(G_m^{(t)} + G_p^{(t)})/2 \\ &= \text{Var}(G)/2. \end{aligned} \quad (4.56)$$

Then

$$\text{Corr}(Y_{mid}, Y_c) = \frac{\text{Var}(G)/2}{\sqrt{\text{Var}(Y_{mid})} \sqrt{\text{Var}(Y_c)}} \quad (4.57)$$

$$\begin{aligned} &= \frac{\text{Var}(G)/2}{\sqrt{\text{Var}(Y)^2/2}} \\ &= \frac{h^2}{\sqrt{2}} \end{aligned} \quad (4.58)$$

where we use the fact that  $\text{Var}(Y_{mid}) = \text{Var}(Y)/2$ .

What is the slope when we regress  $Y_c$  against  $Y_{mid}$ ? In linear regression of variable  $y$  against  $x$ , the slope is given by  $\text{corr}(x, y)\sigma_x/\sigma_y$ . Here  $\sigma_c/\sigma_{mid} = \sqrt{2}$ . Hence

$$\text{Slope of } (Y_{mid} \text{ vs } Y_c) = h^2. \quad (4.59)$$

Lastly much variation is there around the slope? In other words, how well can we predict the child's phenotype from the parents? This is called the coefficient of determination ( $r^2$ ) and equals the correlation squared. When we are predicting  $Y_c$  from  $Y_{mid}$  this is

$$\text{Coefficient of Determination for } Y_c \text{ given } Y_{mid} = \frac{h^4}{2}. \quad (4.60)$$

Notice that even for a trait with  $h^2 = 1$  we predict considerable unpredictability among the children ( $r^2 = 0.5$ ). This arises because due to randomness in which alleles are transmitted from each parent.

<sup>720</sup>Here's an interesting article on how regression to the mean affects the interpretation of data on whether installation of traffic cameras can help reduce road accidents: [Link].

<sup>721</sup>Thiruvenkadan 2009, measures include life-time earnings [Link]

<sup>722</sup>To be precise, male Holsteins carry only two distinct Y chromosomes [Link] More recently, starting in 2003, the bull O-Bee Manfred Justice-ET sired over 100,000 daughters through artificial insemination, and had 425 sons who themselves served as sires (Garcia-Ruiz et al 2016 www.pnas.org/cgi/doi/10.1073/pnas.1519061113)

<sup>723</sup>Hill 2014: <https://PMC3872177/>

<sup>724</sup>A single origin of maize: [Link]; timing of selection on 3 critical genes [Link]

<sup>725</sup>For a nice essay on maize improvement, with figures at the bottom see Bob Nielsen (2023). One key factor was the widespread adoption of hybrid lines, starting in 1937, and major investments in selective breeding programs by the leading seed companies [Link]. See also Duvick 2022

<sup>726</sup>Duvick 2022: [Link]

<sup>727</sup>Similar gains have been achieved for another globally critical crop, wheat. As for maize, wheat improvements reflect the impact of both selective breeding as well as agronomic improvements. For a nice visualization of this see CITE Fischer 2022

<sup>728</sup>Meuwissen et al (2002) built on a previous body of work pioneered by Charles Henderson, in the mid-20th Century, as follows. Suppose we want to estimate the breeding value of a bull (for example, for meat or milk production). The estimate for that individual is inherently noisy (especially for a female phenotype like milk production), but we could use the phenotypes of relatives to estimate of this bull's breeding value. Henderson developed mixed model approaches to obtain BLUP (best linear unbiased predictions) that provide appropriate weighting of the information from different relatives at different degrees of relatedness.

Meuwissen et al showed that one can improve these estimates by using QTL mapping to measure SNP-level effects (this is done in a mixed model framework reminiscent of the BLUP approach, to account for high and variable relatedness of sampled animals). The QTL estimates are then used to predict  $G$ . Some approaches blend the QTL-based prediction with traditional BLUP predictions. [M2002]; [Link], [Link]

<sup>729</sup>Taylor/Taylor/Decker 2016

<sup>730</sup>The concept of *velogenetics* proposes to speed this up even further by aspirating oocytes from females at birth, fertilizing them with the sperm of 10-month-old bulls, and raising the next generation by in vitro fertilization (Taylor et al, 2016).

<sup>731</sup>The different data tracks correspond to the following: "a four-path model of genetic improvement in which genetic progress occurs with differing selection dynamics, partitioned into improvement due to genetic changes in sire(s) of bulls (SB), sire(s) of cows (SC), dam(s) of bulls (DB), and dam(s) of cows (DC)" (Garcia-Ruiz et al 2016).

<sup>732</sup>You might (like me) worry that there is some circularity in showing increases in *estimated G* each year, when that is the quantity being selected for. In principle it might be preferable to plot the realized phenotypic gains, but I have not been able to find good data plots of these. Taylor et al (2016) reassure us that "By 2011, GEBV reliabilities for yield traits had reached 75%, only slightly less than for 7-y-old progeny tested bulls."

<sup>733</sup>For more detail see Box 1 of Visscher/Hill/Wray 2008 NRG. It's conceptually ambiguous whether one should consider  $G \times E$  in the numerator of  $H^2$ , but most descriptions of  $H^2$  do not.

<sup>734</sup>Non-additive interactions are more common among large-effect alleles, which are usually very rare, and thus combinations are hard to find. Of course there are many examples of dominance for rare disease alleles: for example cystic fibrosis is driven by recessive mutations in CFTR. One example of an epistatic interaction is the FUT1 gene, which is essential for processing the cell-surface proteins that mark red blood cells as type A or B. Thus individuals with FUT1 mutations have Type O blood regardless of their genotype at the ABO locus.

<sup>735</sup>Even though  $G \times G$  effects among common variants are very hard to find, it is hard to rule out the possibility that a sum of many small  $G \times G$  interactions may contribute significantly to the total heritability estimated in twin/sib studies.

<sup>736</sup>eg pigmentation and cancer; Mostafavi eLife; smoking example. Point out that it's a bit hard to measure subtle  $G \times E$  in humans, but if it were as strong as in other species we would surely know this.

<sup>737</sup>One main way that  $G \times E$  can arise is as a result of local adaptation. But humans are a relatively homogeneous species, with fairly limited local adaptation. Additionally common genetic variants in humans have almost universally small effects and this tends to dampen non-linear interaction effects at the level of individual variants.

<sup>738</sup>In some cases diseases are defined on the basis of cutoffs of quantitative measurements, including hypertension (high blood pressure), obesity (high BMI), and diabetes (high HbA1c). Other diseases, such as heart disease, result from combinations of factors, some of which we can measure, including cholesterol levels. Finally, for diseases such as schizophrenia or bipolar disorder we still have rather limited knowledge of what features would contribute to a liability phenotype.

<sup>739</sup> Let's add dominance into the model:

$$Y_1 = \bar{y} + A_1 + C + D_1 + E_1 \quad (4.61)$$

$$Y_2 = \bar{y} + A_2 + C + D_2 + E_2 \quad (4.62)$$

For MZ twins (who have identical genomes), the genetic components  $A$ , and  $D$ , are entirely shared (correlation=1). DZ twins share half their genomes on average, and so the correlation of  $A_1$  and  $A_2$  is 0.5. And in one quarter of their genomes they share both chromosomes (which is required to share dominance effects), so the correlation of  $D_1$  and  $D_2$  is 0.25. For both types of twins, environment  $C$  is shared and environment  $E$  is not. Then the correlation between twin phenotypes is:

$$\rho_{MZ}(Y_1, Y_2) = [1 \cdot \text{Var}(A) + \text{Var}(C) + 1 \cdot \text{Var}(D)]/\text{Var}(Y) \quad (4.63)$$

$$\rho_{DZ}(Y_1, Y_2) = [0.5 \cdot \text{Var}(A) + \text{Var}(C) + 0.25 \cdot \text{Var}(D)]/\text{Var}(Y) \quad (4.64)$$

Computing the difference between the equations eliminates  $\text{Var}(C)$ :

$$2(\rho_{MZ} - \rho_{DZ}) = [\text{Var}(A) + 1.5 \cdot \text{Var}(D)]/\text{Var}(Y) \quad (4.65)$$

In this case Falconer's estimator exceeds not only the narrow sense, but even the broad sense heritability. The modeling is a bit more complicated for GxG and GxE but the intuition is similar.

<sup>740</sup>This will mean that  $C$  is bigger for MZ twins than for DZ twins and tend to inflate estimated  $h^2$

<sup>741</sup>Another important potential bias comes from assortative mating – which we know occurs for traits including height and education/social status. This tends to cause downward bias in the estimated heritability because DZ twins are genetically more similar than expected.

<sup>742</sup>Assuming independence of the variance components.

<sup>743</sup>Visscher 2006: [\[Link\]](#)

<sup>744</sup> Specifically, let  $\alpha$  and  $\beta$  be the regression intercept and slope, respectively, in the following regression of squared phenotypic differences in sib pairs against sib pair IBD sharing:

$$\frac{(Y_{i1} - Y_{i2})^2}{2\sigma_p^2} = \alpha + \beta\pi_i, \quad (4.66)$$

where  $Y$  is phenotype, subscript  $i$  indexes the  $i$ th family,  $\sigma_p^2$  is the phenotypic variance, and  $\pi_i$  is the IBD sharing in family  $i$ . Then we can estimate narrow sense heritability as

$$\hat{h}^2 = \beta \quad (4.67)$$

[rewritten slightly from Equations 12 and 14 of Visscher et al (2006)].

<sup>745</sup>Young et al 2019 [\[Link\]](#)

<sup>746</sup>The RDR-style approach should be much less susceptible to bias than the classic twin studies. Assortative mating can be a downward-biasing factor for traits such as height and years-of-education. For another paper in this space see Kemper et al 2021 [\[Link\]](#). See also Wainschtain and the Gusev [\[Link\]](#) and Paige Harden [\[Link\]](#) blogs.

<sup>747</sup>There's a notable leveling-off in some locations starting in the 1970s, although some parts of the world have continued to see height increases very recently (*Lancet*). For a great essay on demographics of height see: Max Roser et al, Our World in Data (2013, updated 2024) [\[Link\]](#). For fantastic data on very recent changes in height and BMI see [\[Link\]](#)

<sup>748</sup>I'm not implying here that high meat consumption is necessary for healthy growth, but rather that this is a marker of access to high quality, balanced, and nutritionally diverse diets.

<sup>749</sup>Doc Edge paper

<sup>750</sup>Berg and Moshaal references; discuss portability, with REFS

<sup>751</sup>A second huge analysis challenge is that people tend to choose partners with similar traits (known as assortative mating) which breaks the default assumptions of most our models.

<sup>752</sup>Bear in mind that RDR methods are as close as we have right now to a gold-standard method. For estimates see Gusev blog [\[Link\]](#); 8% from table 2 Beauchamp preprint [\[Link\]](#); 17% from table S4 Young et al [\[Link\]](#)

<sup>753</sup>e.g a study of SNP-heritability found estimates ranging from 34%–50%, depending on covariates; meanwhile, an unadjusted pedigree-based estimate is as high as 75% [figure s2 of Wainschtain 2025]

<sup>754</sup>These data come originally from a 2023 study by Gregory Clark that used a database of genealogical records to measure the inheritance of social status from 1600 onwards in England [[Link](#)]. Clark explained these data with a purely genetic model of inheritance, involving a combination of high assortative mating and implausibly high heritability of behavioral traits. For a valuable commentary on the Clark paper see Benning et al (2023) [[Link](#)].

<sup>755</sup>link to some background on this point

<sup>756</sup>[[Link](#)]