Quantitative Genetics and Quantitative Traits

Introduction

In the chapter on Mendel and Morgan, we saw how the transmission of genes from one generation to another follows precise mathematical formula. The traits discussed in that chapter, however, were discrete traits—peas are either yellow or green, someone either has a disorder or does not have the disorder. But many behavioral traits are not like these clear-cut, have-it-or-don't-have-it phenotypes. People vary from being quite shy to very outgoing. But is shyness a discrete trait or merely a descriptive adjective for one end of a continuous distribution? In this chapter, we will discuss the genetics of quantitative, continuously distributed distribution.

Let us note first that genetics has made important—albeit not well recognized—contributions to quantitative methodology in the social sciences. The concept of regression was initially developed by Sir Francis Galton in his attempt to predict offspring's phenotypes from parental phenotypes; it was later expanded and systematized by his colleague, Karl Pearson[1], in the context of evolutionary theory. The analysis of variance was formulated by Sir Ronald A. Fisher[2] to solve genetic problems in agriculture. Finally, the famous American geneticist Sewell Wright developed the technique of path analysis which is now used widely in psychology, sociology, anthropology, and other social sciences.

Basic Tools: Three Statistics

To understand quantitative genetics, it is important to understand the meaning of three basic statistics—the *mean*, the *variance*, and the *correlation coefficient*. The mathematics behind each of these three statistics is much less important that the conceptual issues of what they measure.

The mean is a *measure of central tendency or location* and answers the question, "Around which number do the scores tend to cluster?" The mean is the arithmetic average and is computed by summing all the scores and dividing by the number of observations.

The variance of a collection of scores is a measure of *individual differences around the mean.* It is a measure of the degree to which the scores are dispersed away from the mean. A variance can range from 0 to a large positive number. A variance of 0 signifies that there is no dispersion around the mean—every score is the same and every score equals the mean. The larger the variance, the more the scores are scattered around the mean.

An important feature of variance is that it can be partitioned. As we will see, the variance in the phenotype can be partitioned into a portion due to genetic variance and another portion due to environmental variance. This partitioning helps geneticists to answer two important questions—to what extent are observable individual differences

---

[1] After whom is named the Pearson product moment correlation.

[2] After whom the *F* statistic is named.

due to individual differences in genotype and to what extent are observable individual differences due to individual differences in the environment?

A correlation coefficient is a measure of *the extent to which one scores on one variable can predict scores on a second variable*. Mathematically, a correlation coefficient can range from -1.0 to 1.0. There are two important attributes of a correlation coefficient. The first is the *sign* of the correlation. A positive sign (i.e., a correlation between 0 and 1.0) denotes a direct relationship. Here, high scores on the first variable predict high scores on the second variable, and conversely low scores on the second variable predict low scores on the second variable. The correlation between height and weight is positive. People who are taller than average tend to (but do not necessary *have* to) weigh more than the average person, and people smaller than average in height tend to weigh less than the average person. A negative sign (i.e., a correlation coefficient less than 0) denotes an inverse relationship. In this case, high scores on one variable predict low scores on the second variable, and conversely low scores on the first variable predict high scores on the second variable. The correlation between the amount of time spent partying and grades is negative. Students who spend a very large amount of time partying tend to get lower than average grades while students who spend little time at parties tend to receive higher grades.

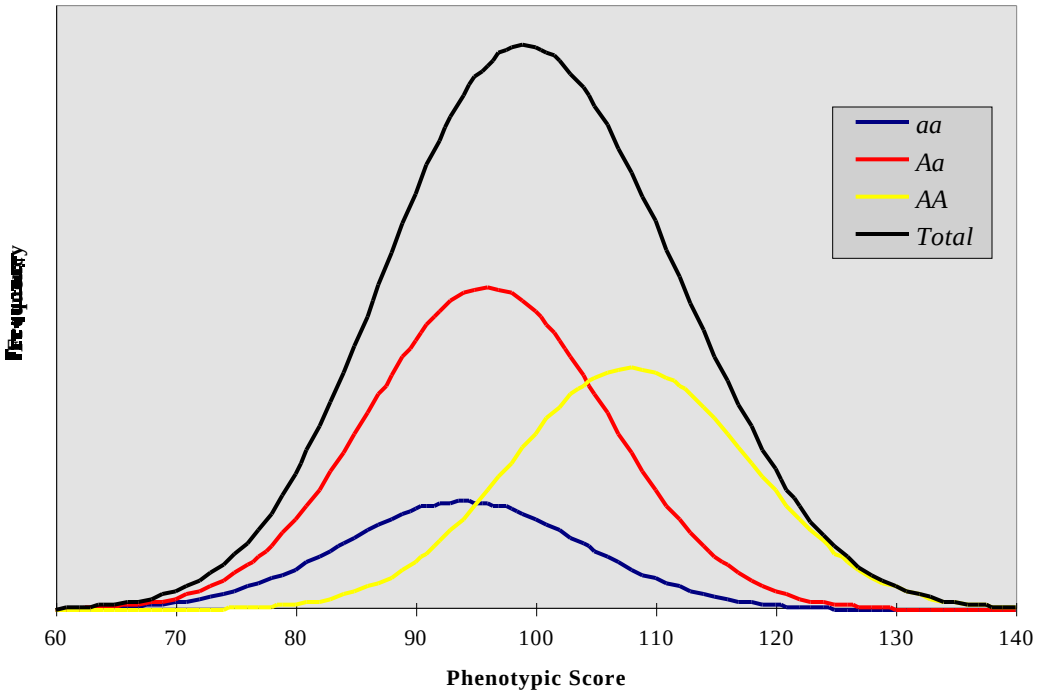
The second important attribute of the correlation is the square of the correlation coefficient. Because a correlation can range between -1.0 and 1.0, the square of the correlation must range between 0 and 1.0. The square of the correlation is a measure of the *amount of predictability* between the two variables. Statistically speaking, the correlation squared gives *the proportion of variance in one variable that is predicable from the other variable.* Because variance is a measure of individual differences, another way of stating the previous statement is that the correlation squared is a measure of the extent to which *individual differences in one variable are predictable from the second variable*. If the correlation squared is 0, then there is no predictability—the two variables are not related to each other. If the squared correlation is 1.0, then we can perfectly predict scores on one variable by knowing scores on the second variable.

Continuous Variation and the Single Locus

Let us begin the development of a quantitative model by considering a single gene with two alleles, *a* and *A*. Define the *genotypic value* (a.k.a. *genetic value*) for a genotype as *the average phenotypic value for all individuals with that genotype*. For example, suppose that the phenotype were IQ, and we measured IQ on a very large number of individuals. Suppose that we also genotyped these individuals for the locus. The genotypic value for genotype *aa* would be the average IQ of all individuals who had genotype *aa*. A hypothetical example is presented in Figure X.1.

The first point to notice about Figure X.1 is the variation in IQ around each of the three genotypes, *aa*, *Aa*, and *AA*. Not everyone with genotype *aa*, say, has the same IQ. The reasons for this variation within each genotype is unknown. It would include environmental variation as well as the effects of loci other than the one genotyped.

Figure X.1. Example and notation for the quantitative genetic analysis of a single locus.



A second important feature about Figure 1 is that the means of the distributions for the three genotypes differ. The mean IQ (i.e., the genotypic value) for *aa* is 94, that for *Aa* is 96, and the mean for *AA* is 108. This implies that the locus has some influence on individual differences in IQ.

A third feature of note in Figure X.1 is that the genotypic value of heterozygote is not equal to the average of the genotypic values of the two homozygotes. The average value of genotypes *aa* and *AA* is (94 + 108)/2 = 101, but the actual genotypic value of *Aa* is 96. This indicates a certain degree of *dominant gene action* for allele *a*. Allele *a* is not completely dominant; otherwise, the genotype value for *Aa* would equal that of *aa*. Hence, the degree of dominance is incomplete.

A fourth feature of importance is that the curves for the three genotypes do not achieve the same height. This is due to the fact that the three genotypes have different frequencies. In the calculations used to generate the figure, it was assumed that the allele frequency for *a* was .4 and the frequency of *A* was .6. Hence, the frequency of genotype *aa* would be .16, the frequency of *Aa* would be .48, and the frequency of *AA* would be .36. Consequently, the curve for *Aa* has the highest peak, the one for *AA* has the second highest peak, and that for *aa* has the smallest peak.

A final feature of note is that the phenotypic distribution of IQ in the general population (the black line in Figure X.1) looks very much like a normal distribution. The phenotypic distribution is simply the sum of the distributions for the three genotypes. For example, the height of the black curve when IQ equals 90 is the distance from the

horizontal axis to the yellow line when IQ is 90 plus the distance from the horizontal axis to the blue line plus the distance from the horizontal axis to the red line. Often social scientists mistakenly conclude that the phenotypic distribution must be trimodal because it is the sum of three different distributions.[3]

The gene depicted in Figure X.1 is currently termed a *QTL* for *Q*uantitative *T*rait *L*ocus. Behavioral genetic research devotes considerable effort towards uncovering QTLs for many different traits—intelligence, reading disability, various personality traits, and psychopathology. The mathematical models that quantify the extent to which a QTL contributes to trait variance are not necessary for us to know. The interesting reader may consult the Advanced Topics section at the end of this chapter.

Continuous Variation and Multiple Loci

Suppose that we could genotype people at another locus for IQ, say the *B* locus with its two alleles, *b* and *B*. We would now have nine genotypic values as illustrated in Table X.1. Once again, we would compute the mean IQ score for all those with a genotype of *aabb* and then enter this mean in the appropriate cell of the table.

Table X.1 Genotypic values for two loci.

	bb	*Bb*	*BB*	Mean
AA	*AAbb* 101	*AABb* 106	*AABB* 111	108
Aa	*Aabb* 89	*AaBb* 94	*AaBB* 99	96
aa	*aabb* 87	*aaBb* 92	*aaBB* 97	94
Mean	93	98	103	100

We could also draw curves for each genotype analogous to the curves depicted in Figure X.1. This time, however, there would be nine normal curves, one for each genotype. We could continue by adding a third locus with two alleles. This would give 27 different genotypes and 27 curves. If we could identify each and every locus that contributes to IQ, then we would probably have a very large number of curves. The variation within each curve would be due to the environment.

This model is known as the *polygenic* model—*poly* for many and *genic* for genes. A special case of polygenic transmission occurs when there are only a few QTLs that contribute to a trait. This is called *oligogenic* transmission (*oligo* for few). There is not an exact number of loci that distinguish oligogenic from polygenic transmission. Indeed, the mathematics behind oligogenic and polygenic models are identical except for the number of loci. At the present time, there is no empirical evidence that gives even a remote glimpse at the number of genes that may be responsible for human quantitative traits. It may turn out that only five or six QTLs are needed to explain all but a small

[3] The phenotypic distribution *may* be trimodal, but it will be so only when the means for the three genotypes are very, very different. When single genes exert only a small influence on a phenotype, then the phenotypic distribution can appear quite smooth as the present example suggests.

amount of genetic variance in some phenotypes. However, the number of genes expressed in the mammalian central nervous system is estimated in the tens of thousands, so many traits may involve hundreds of loci.

Heritability and Environmentability

The concepts of heritability and environmentability of polygenic traits are central to quantitative analysis in behavioral genetics. Instead of providing formal definitions of these terms, let us begin with a simple thought experiment and then discover the definitions through induction.

Imagine that scores on the behavioral trait of impulsivity are gathered on a population of individuals. These observed scores will be called the *phenotypic values* of the individuals. Assume that there was a futuristic genetic technology that could genotype all of the individuals in this population for all the loci that contribute to impulsivity. One could then construct a *genotypic value* for each individual. Just as with one or two loci, the genotypic value for a polygenic trait is defined as the mean phenotypic value of all those individuals with that genotype in the population. For example, if Wilbur Waterschmeltzer's genotype for impulsivity is AaBBCCddEeff and the mean impulsivity score for all individuals in the population who have genotype AaBBCCddEeff is 43.27, then Wilbur's genotypic value is 43.27.

Imagine another technical advance that would permit us to calculate and quantify all the environmental experiences in a person's life that could contribute to the person's level of impulsivity. This would be the *environmental value* for an individual. We would now have a very large set of data, the initial part of which would resemble Table X.2.

Table X.2. Hypothetical data set containing the phenotypic, genetic, and environmental values for individuals.

Observation	Genetic Value = ***G***	Environ-mental Value = ***E***	Phenotypic Value = ***P***
Abernathy Abercrombie	113	96	107
Beulah Bellingwacker	92	74	77
…	.	.	.
Zelda Zorkminder	118	104	118

We would compute a correlation coefficient between the genotypic values and the phenotypic values. Recall that the square of the correlation coefficient between two variables gives the proportion of variance in one variable attributable to (i.e., predicted by) the other variable. Consequently, if we square the correlation coefficient between the genotypic values and the phenotypic values, we would arrive at the proportion of phenotypic variance predicted by (or attributable to) genetic variance. This quantity, the square of

the correlation coefficient between genotypic values and phenotypic values, is called *heritability*[4].

Thus, heritability is a quantitative index of the importance of genetics for individual differences in a phenotype. Strictly defined, heritability is *the proportion of phenotypic variance attributable to or predicted by genetic variance*. Because heritability is a proportion, it will range from 0 to 1.0. A heritability of 0 means that there is no genetic influence on a trait, whereas a heritability of 1.0 mean that trait variance is due solely to heredity. A less technical view would define heritability as *a measure, ranging from 0 to 1.0, of the extent to which observed individual differences can be traced in any way to genetic individual differences.*

Just as we could compute a correlation between genetic values and phenotypic values, we could also compute correlations between environmental values and phenotypic values. Squaring this correlation would give us the *environmentability* of the trait. Environmentability has the same logical meaning as heritability but applies to the environment instead of the genes. *Environmentability is the proportion of phenotypic variance attributable or predicted by environmental variance*. It is also *a quantitative index, ranging from 0 to 1.0, of the extent to which environmental individual differences underlie observable, phenotypic individual differences.*

Estimating Heritability and Environmentability

Family Correlations

Even with the marvelous technology of modern genetics, it is not possible to directly measure genotypic values for polygenic traits. And it stretches imagination to suppose that we can measure environmental values for all those varying factors that influence a trait. Instead, we observe only phenotypes in relatives.

Table X.3. Organization of data for computing the correlation between parent and offspring.

Family	Parent IQ	Child IQ
Athabaska	107	104
Bottomwinger	77	98
.		
Zakmeister	118	102

Table X.3 illustrates the type of data that behavioral geneticists gather. The family is the unit of observation and the phenotypic scores for the different classes of relatives are the variables. For the data in Table X.3, we would compute the correlation between the variables "parental IQ" and "child IQ" giving a parent-offspring correlation[5].

[4] Two assumptions are necessary to define heritability (and later, environmentability) this way. First, it is assumed that the genotypic values are uncorrelated with the environmental values. Second, there is no statistical interaction between genotypic values and environmental values. These assumptions will be discussed later int he handout.

[5] The reader familiar with data analysis should realize that because families do not have the same number of offspring, family data is usually not "rectangular." There are methods to take care of such data sets but they are too advanced for this text. The interested reader should consult Neale and Cardon (199x).

But correlations among the relationships in ordinary nuclear families cannot be used to estimate heritability. Behavioral similarity between, say, parents and offspring, may be due to any of three factors: (1) shared genes; (2) shared family environment; and (3) some combination of shared genes and shared family environment. Consequently, behavioral scientists usually study two special types of relatives to tease apart the influence of shared genes from that of shared environment. These two special populations are twins and adoptees. Each is discussed in turn.

The Twin Method: Rationale

Monozygotic (MZ) or *identical* twins are the result of the fertilization of a single egg. The cells from this zygote[6] divide and divide, but early in the course of development, some cells physically separate and begin development as an independent fetus. The reasons for the separation are currently unknown. Because the two individuals start out with the same genes, they are effectively genetic clones of each other. Any differences between the members of an identical twin pair must be due to the environment. Included in the environment is the fact that one twin may have developed from more cells than the other since it is suspected that the original separation is seldom an equal 50-50 split. MZ twins look so alike that they are often confused by people who do not know them well.

Dizytogic (DZ) or *fraternal* twins result when a woman double ovulates and each egg is independently fertilized. Genetically, DZ twins are as alike as ordinary siblings, sharing on average 50% of their genes, and look alike as ordinary brothers and sisters Differences between the members of a fraternal twin pair will be due to both the environment and also to the different alleles that each member inherits.

Consequently, the logic of the twin method is quite simple. If genes contribute to a trait, then MZ twins should be more similar to each other than DZ twins. Thus, the striking physical similarity of MZ twins in terms of height, facial features, body shape, hair color, eye color, etc. suggests that genes influence individual differences in these traits because fraternal twins are as alike in their physical features as ordinary siblings.

The Adoption Method: Rationale

The logic of the adoption method is as simple as the logic of the twin method, provided that nonfamilial adoptions are used. When parents adopt and raise a child to whom they are not genetically related, then any similarity between the parents and child must have something to do with the environment. Similarly, when there are two adoptive children raised in the same family, then sibling resemblance between the two must also be environmental in nature.

When children are adopted at shortly after their birth, then shared genes are the only reason why they would show similarity with their biological relatives. Thus, correlations between adoptees and their genetic relatives give evidence for heritability.

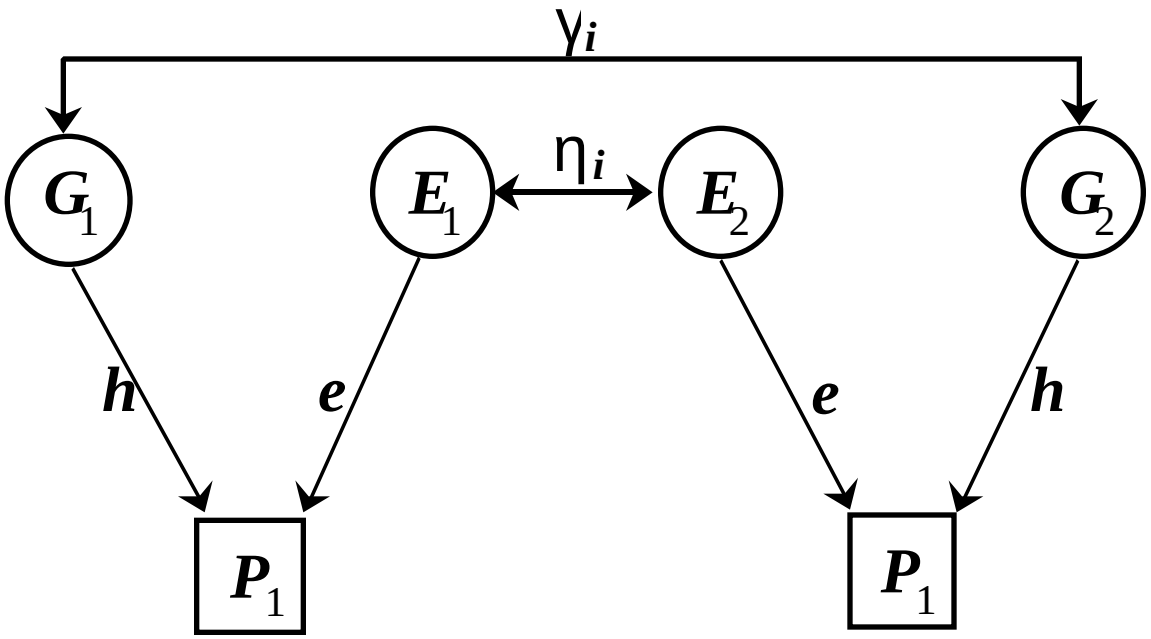
---

[6] A *zygote* is scientificese for a fertilized egg.

The Quantitative Model

A model for the similarity for any of pair of relatives is depicted in Figure X.2. In this figure, *G* denotes genotypic value, *E* denotes environmental value, and *P* stands for phenotypic value[7]. Subscripts 1 and 2 denote respectively, the first and the second relative. If the relatives were siblings, then G_1 denotes the genotypic value for sib 1, E_2 stands for the environmental value of sib 2, etc. If the relatives were parent and offspring, then E_1 would represent the environmental values of parents, P_2 would denote the phenotypic values of offspring, etc.

The straight, single-headed arrows (or *paths*) originating in the *G*s and entering the *P*s denotes the possibility that genotypic values predict phenotypic values. The *h*s on these two arrows quantify this effect. Strictly speaking, *h* is the correlation between genotypic and phenotypic values[8]. Similarly, the path between the *E*s and the *P*s denotes the prediction of phenotypic values from environmental values, and *e* is the correlation between *E* and *P*.



The double-headed arrow connecting the G_1 to G_2 denotes that fact that the genotypic values of relatives may be correlated. The quantity γ_i gives this correlation. Similarly, the double-headed arrow connecting E_1 to E_2 allows for the possibility that the environmental values of the two relatives are correlated, and η_i (Greek lowercase eta) denotes this correlation. Both γ and η have the subscript i attached to them. This

---

[7] Technically, this figure is a path diagram. Observed variables are denoted by rectangles. Because we measure phenotypes, the two *P*s are encased in rectangles. Unobserved or latent variables are denoted by circles or ellipses. Because we cannot measure genotypic values and environmental values, the *G*s and *E*s are enclosed in circles.

[8] In general, *h* is the standardized regression coefficient when phenotypic values are regressed on genotypic values. This equals the correlation in the present case because it is assumed that *G* is not correlated with *E*.

Table X.4. Values of γ and η for different types of genetic and adoptive relationships. Under the equal environments assumptions, η for MZ twins will equal η for DZ twins. With no selective placement, γ for adoptive relatives and η for genetic relatives will be 0.

Relationship	**Notation**	γ	η
MZ twins together	mzt	1.0	η_{mzt}
MZ twins apart	mza	1.0	η_{mza}
DZ twins together	dzt	.5	η_{dzt}
DZ twins apart	dza	.5	η_{dza}
Siblings together	sibst	.5	η_{sibst}
Siblings apart	sibsa	.5	η_{sibsa}
Parent-offspring together	pot	.5	η_{pot}
Parent-offspring apart	poa	.5	η_{poa}
Grandparent-grandchild	gg	.25	η_{ggt}
Uncle/aunt-nephew/niece	uann	.25	η_{uannt}
Cousins	cous	.125	η_{cous}
Adoptive parent-offspring	apo	γ_{apo}	η_{apo}
Adoptive siblings	asibs	γ_{asibs}	η_{asibs}

denotes the ith type of relationship. The values of γ and η for different types of relationships are given in Table X.4. (The values of γ in this table are derived from a simple, additive genetic model. Assumptions and difficulties with this model are noted in the text box titled "The Problem with γ." You should also carefully read the text box on the problems with η.)

We can now use the rules of path analysis developed by Sewall Wright to derive two central equations for the quantitative model. The first important equation is that for the phenotypic variance. Because the quantitative model is expressed in terms of standardized variables[9], the variance of the phenotype will equal 1.0. The equation for this variance is

$$1.0 = h^2 + e^2.$$

The second equation expressed the correlation for any type of relative pair in terms of the unknowns in Figure X.2 (i.e., γ_i, η_i, h and e). Let R_i denote the correlation for the ith type of relationship. Then,

$$R_i = \gamma_i h^2 + \eta_i e^2.$$

This equation may be used to find the correlation for any type of relationship listed in Table X.4. Simply take the relationship that you want and substitute the appropriate γ_i and η_i from Table X.4. For example, the correlation for MZ twins raised together will be

[9] In this case standardized variables have means of 0 and standard deviations of 1. Because the variance is the square of the standard deviation, the variance of standardized variables will also be 1.

The Problem with γ.

Here, a small digression is in order because the quantity γ in the path model requires some explanation. This quantity is the correlation between the genotypic values of relatives. If the relatives are identical twins, then γ = 1.0 because the twins have identical genotypes. For fraternal twins and for ordinary siblings, the precise mathematical value of γ is not known. If the world of genetics were a simple place where each allele merely added or subtracted a small value from the phenotype and there were no assortative mating for the trait, then γ would equal .50. This value of γ is often assumed in the analysis of actual data, more for the sake of mathematical convenience than for substantive research demonstrating that the assumptions for choosing this value are valid.

If gene action is not simple and additive, then the value of γ will be something less than .50. The two classic types of nonadditive gene action are *dominance* and *epistasis*. Dominance, of course, occurs when the phenotypic value for a heterozygote is not exactly half way between the phenotypic values of the two homozygotes. Epistasis occurs when there is a statistical interaction between genotypes. Both dominance and epistasis create what is termed *nonadditive genetic variance*. For technical reasons, nonadditive genetic variance reduces the correlation between relatives to something less than .50. The extent of the reduction depends upon the type of relatives.

Assortative mating, on the other hand, will tend to increase the value of γ. When parents are phenotypically similar and when there is some heritability, then the genotypes of parents will be correlated. The effect of this is to increase the genetic resemblance of their offspring over and above what it would be under random mating.

What should be done under such complexities? The typical strategy of setting γ equal to .50 is not a bad place to start. If a trait shows strong assortative mating, then more elaborate mathematical models can be developed to account for the effects of nonrandom mating. The real problem occurs with nonadditive genetic variance. When this is present, then the techniques described above can overestimate heritability. This is another reason why heritability estimates should not be interpreted as precise, mathematical quantities.

$$R_{mzt} = h^2 + \eta_{mzt} e^2,$$

and the correlation for sibs raised apart will be

$$R_{sibsa} = .5h^2 + \eta_{sibsa} e^2.$$

This cumbersome notation has been used deliberately because later it will reveal to us some important assumptions about the twin and the adoption method.

The Problem with η.

A small digression is useful here to explore the meaning of η. This parameter[1] is usually interpreted as a measure of *family environment* but the term "family environment" has a rather strange and esoteric meaning. It denotes much more than the physical surroundings within a household and the interpersonal dynamics of the family members. The :family environment" may actually exclude interpersonal events between family members! The family environment in this context is defined *as all those factors, both inside and outside of the physical household, that make a pair of relatives similar on the phenotype being studied*. An example will clarify this admittedly vague definition.

Suppose that relatives were pairs of young sibs. Siblings live in the same neighborhood, usually attend the same schools, and often have friends in common. If neighborhood, quality of school, and peers influence a phenotypic like achievement motivation, then they are part of the "family environment" for siblings, even though they are not physically located within the household.

Suppose that future research found that parents subtlety treat their children differently by, say, encouraging the sib with the higher grades in school to study more and take academics more seriously. This parental action will make pairs of siblings *different*, not similar. Hence, it would *not* be considered a family environmental factor even though from a psychological perspective it involves social interaction between parents and their offspring.

An astute reader may question why anyone would regard η as a measure of "family environment" when family environment is defined in such an odd way. There is indeed considerable merit to this perspective, but the sad fact is that this definition of η and the family environment has been used so much in the literature that it is almost carved in stone. At the risk of offending many colleagues, I suggest a vigorous sandblasting of that stone. Let us begin to view η for what it really is—the correlation between the environments of relatives. It is an index of the environmental similarity of relatives and measures the extent to which relatives are correlated because they have some environmental factors in common. Family factors can make η high or can make η low, just as factors outside the family can influence η.

Now let us look at the typical twin method that gathers data on MZ and DZ twins raised together. There will be three equations for these type of data. The first is the equation for the phenotypic variance,

$$1.0 = h^2 + e^2.$$

The second and third are the equations for the MZ and DZ correlations, or

$$R_{mzt} = h^2 + \eta_{mzt} e^2,$$

and

$$R_{dzt} = .5h^2 + \eta_{dzt}e^2.$$

Because we have gathered data, we will have actual numbers for R_{mzt} and for R_{dzt}. Now recall some high school algebra. We have three equations with actual numbers on the left hand side but four unknowns (h^2, e^2, η_{mzt}, and η_{dzt}) on the right hand side. Can we solve for four unknown when there are only three equations? No. There must be at least as many equations as there are unknowns to get a solution.

Behavioral geneticists resolve this problem by making the *equal environments assumption*. This assumption states that *the correlation between the environments for MZ twins equals the correlation between the environments for DZ twins*, or in mathematical terms, $\eta_{mzt} = \eta_{dzt}$. We will discuss this assumption in detail later on. For now, we concentrate on the quantitative model. Let us make the equal environments assumption, and let us denote the correlation between twins' environments as η_{tt}, the subscripts standing for *t*wins raised *t*ogether. Then the three simultaneous equations are

$$1.0 = h^2 + e^2,$$
$$R_{mzt} = h^2 + \eta_{tt}e^2,$$

and

$$R_{dzt} = .5h^2 + \eta_{tt}e^2.$$

Now we have three equations in three unknowns (h, e, and η_{tt}). The algebraic is left to the reader. The solution is

$$h^2 = 2(R_{mzt} - R_{dzt}),$$
$$e^2 = 1 - h^2,$$

and

$$\eta_{tt} = \frac{R_{mzt} - h^2}{e^2}.$$

For example, suppose that we gathered data on the personality trait of sociability and found that $R_{mzt} = .60$ and $R_{dzt} = .45$. Then we would estimate heritability as

$$h^2 = 2(R_{mzt} - R_{dzt}) = 2(.60 - .45) = .30.$$

We would conclude that 30% of phenotypic variance is attributed genetic variance or, using more common sense language, that 30% of all the observable individual differences in sociability can be traced in some way to genetic individual differences.

Because e is the correlation between environmental values and phenotypic values, then e^2 is the environmentability or the proportion of phenotypic variance attributable to environmental variance. For the hypothetical sociability data,

$$e^2 = 1 - h^2 = 1 - .30 = .70.$$

Here, we would conclude that the environmentability is .70, so 70% of observed individual differences in sociability may be traced in some way to environmental individual differences among people.

Finally, we can compute the correlation between the environments of the twins.

$$\eta_{tt} = \frac{R_{mzt} - h^2}{e^2} = \frac{.60 - .30}{.70} = .43.$$

This value for η_{tt} suggests that there is an important correlation between the twin's environments. Perhaps, then, factors such as being raised in the same family, having the same friends, etc. are important in producing environmental similarity among siblings for sociability.

In the adoption design, the typical assumption made by behavioral geneticists is that *the environments of these relatives raised apart are uncorrelated and the genotypic values for genetically unrelated relatives are uncorrelated*. Formally, this assumption is called the *absence of selective placement*. Once again, discussion of selective placement will be deferred to concentrate on the quantitative model.

Consider now the equation for MZ twins raised apart. In Table X.4, the γ_i value for this relationship is 1.0 and the η_i value is η_{mza}, giving the equation

$$R_{mza} = h^2 + \eta_{mza} e^2.$$

If there is no selective placement, then $\eta_{mza} = 0$, so the equation reduces to

$$R_{mza} = h^2.$$

In English, this means that the correlation between twins raised apart is a direct estimate of heritability.

Let us now examine a typical adoption design. Here, we often observe two correlations—that for biological parent and adoptive offspring and that for adoptive parent and adoptive offspring. Using Table X.3, these correlations are

$$R_{bpo} = .5h^2 + \eta_{bpo} e^2$$

and

$$R_{apo} = \gamma_{apo} h^2 + \eta_{apo} e^2.$$

Added to these, we have the third equation for the phenotypic variance

$$1.0 = h^2 + e^2,$$

giving us three equations in five unknowns (h^2, e^2, η_{bpo}, γ_{apo}, and η_{apo}). If we make the assumption of no selective placement, then $\eta_{bpo} = 0$ and $\gamma_{apo} = 0$. The three equations are now

$$R_{bpo} = .5h^2,$$
$$R_{apo} = \eta_{apo} e^2.$$

and

$$1.0 = h^2 + e^2.$$

There are now three equations in three unknowns. Once again, the algebraic solution is left to the reader. The results are

$$h^2 = 2R_{bpo},$$
$$e^2 = 1 - h^2,$$

and

$$\eta_{apo} = \frac{R_{apo}}{e^2}.$$

For example, suppose that we gathered adoption data on interest in "blood sports" (e.g., interest in watching boxing events, hunting, etc.) and found that the correlation between

biological parents and their adoptees was .19 and the correlation between adoptive parents and their adoptive children was .11. Then we would estimate heritability as

$$h^2 = 2R_{bpo} = 2(.19) = .38.$$

Environmentability would be estimated as

$$e^2 = 1 - h^2 = 1 - .38 = .62.$$

And the correlation between the environments of adoptive parents and their adoptive offspring would be

$$\eta_{apo} = \frac{R_{apo}}{e^2} = \frac{.11}{.62} = .163.$$

The Quantitative Model: Overall Perspectives

Any student reading trying to cram this material before an exam is bound to be bewildered. There are a large number of esoteric symbols, a giddy array of equations, and quite a bit of hidden algebra. If you feel confused at this point, then you have lost the forest for the trees. Let us step backwards for a moment to examine the big picture.

Step 1: quantitative behavioral genetics starts with a mathematical model that gives equations for the correlation between various classes of relatives. Step 2: for any actual data set, there are a limited number of relationships; hence, the number of unknowns usually exceeds the number of equations given in step 1. Step 3: to solve this problem, certain simplifying assumptions are made. Step 5: after these assumptions are made, the problem becomes mathematically solvable.

Now, it takes little insight to recognize that the strength of any mathematical model depends upon the extent to which the *assumptions of that model are robust.* The term "robust" was deliberately used. A "robust" assumption is one that might actually be violated, but the effect of violating the assumption is so small that substantive results will not be altered. Consequently, to critically examine whether any of the above makes practical sense, we must determine whether the two major assumptions—the equal environments assumptions and the no selective placement assumption—are robust.

Assumptions in Behavioral Genetics Research

The Twin Method

The central assumption of the twin method is often called the *equal environments assumption*. This assumption states that *environmental factors do not make MZ twins more similar than they make DZ twins similar*. In terms of the mathematics of the model in Figure X.2 and Table X.4, this assumption implies that $\eta_{mzt} = \eta_{dzt}$. To violate this assumption, two very important phenomenon must *both* occur: (1) environmental factors must treat MZ twins more similarly than DZ twins; and (2) that similarity in treatment *must* make a difference in the *phenotype under study*. An example can help to illustrate. Parents often dress identical twin children in similar outfits. All of us have seen a pair of identical twin girls outfitted in the same dress or a pair of young MZ boys both wearing a sailor suit. Parents frequently dress their DZ twins in identical attire but not nearly with the frequency of parents with MZ twins. Consequently, if the phenotype under study

were "fashion in young children," then the equal environments assumption would be violated and we should not use the twin method to estimate heritability.

Let us take this example a bit further. Suppose that the phenotype under study was adult shyness. The first facet of the equal environments is violated because the MZ twins in the sample will have been dressed more alike as children than the DZ twins. However, the second facet of the equal environments assumption—that similarity in treatment makes a difference in the phenotype under study—would probably not be true. If it were true, then being dressed as a child in, say, a cowboy outfit as opposed to a sailor suit would have an important influence on adult shyness. Hence, for the phenotype of childhood fashion, the equal environments assumption would be violated, but for the phenotype of adult shyness, the assumption may be valid.

Empirical data on the equal environments assumption suggests that the assumption is very robust. That is, for most substantive human behaviors studies thus far, the effects of violating the assumption are very minor. It is quite true that as children MZ twins are often called by rhyming or alliterative names (e.g., Johnnie and Donnie), that they are dressed alike more frequently than DZ twins, and that in general parents treat them more as a unit that they do fraternal twins. However, several different types of data suggest that this treatment does not influence substantive traits.

The first line of evidence is that *actual* zygosity predicts behavioral similarity better than *perceived* zygosity. In the past, many parents of twins were misinformed or make erroneous conclusions on their own part about the zygosity of their offspring[10]. Consequently some parents raised their DZ twins as MZ twins while other treated their MZ offspring as DZ pairs. The behavioral similarity of these twins is better correlated with their biological zygosity rather than their rearing zygosity ( ).

A second line of evidence relies that even though *on average* parents of MZ children treat them more alike than parents of DZ children, there is still strong variability in the way parents of MZ pairs treat their children. Some parents accentuate their MZ offspring's similarity by making certain that they have the same hairstyle, clothing, brand of bicycle, etc. Other parents will actually go out of their way to avoid treated their MZ children as a unit and deliberately try to "individualize" them. However, those MZ twins were treated as a unit were no more similar in their adolescent and adult behavior than those who were deliberately individualized ( ).

The final and best line of evidence comes from studies of twins raised apart. These twins are not raised in completely random environments, but they are certainly not subject to the subtle treatments of being dressed alike as twins who are raised day-in and day-out in the same household for all their childhood and early adolescence. As one scholar of twins raised apart, James Shields, put it, "The importance of studying separated twins is to demonstrate that the microenvironment of daily living in the same

[10] A persistent myth, held even by some MDs, was that identical twins have one afterbirth while fraternal twins have two afterbirths. What is true is that DZ twins always have two chorions (a sac enclosing the anmion and amnionic fluid) while MZ twins may have either one or two chorions. Either type of twins can have one or two afterbirths.

household is not solely responsible for the great similarity observed in twins raised together" (Shield, personal communication, 1976).

If sibling resemblance were due mainly to the environment and if violation of the equal environments assumption was the only major reason why MZ twins raised together correlated higher than DZ twins raised together, then two predictions can be made about separated twins. First, the correlation for separated twins should be small and close to 0; it should certainly be less than the correlation for siblings raised together. Second, the correlation for MZ twins raised apart should be no different than the correlation for DZ twins raised apart. The available data on twins raised apart are inconsistent with both of these predictions ( ). First, for almost all traits that have been studied, the correlations for twins raised apart have been substantial and significant. Indeed, MZ twins raised apart are consistently *more* similar than biological siblings and DZ twins who are raised together. Second, MZ twins raised apart correlate higher than DZ twins raised apart.

Taken together, all these lines of evidence suggest that the equal environments assumption is indeed robust. The term "robust" was used quite deliberately. It means that if there is any violation of the equal environments assumption, then the quantitative effect of that violation is quite small and does not compromise the study of twins raised together.

The Adoption Method

There are two critical assumptions about the adoption method—the absence of selective placement and the representativeness of the adoptive families. Selective placement occurs most often when adoption agencies deliberately try to place adoptees with adoptive parents who resemble the adoptee's genetic parents. Like the equal environments assumption in the twin method, the critical issue is not whether selective placement occurs—it does—but whether the selective placement influences the trait in question.

Most contemporary adoption studies report strong selective placement for race/ethnicity and for religion ( ). Placement for religion is seldom done deliberately. It is mostly a secondary consequence of different religious denominations supporting their own adoption agencies. Catholic Social Services, for example, deal mostly with Catholic unwed mothers and place children into Catholic homes. Similar venues occur for other religiously affiliated agencies.

There is moderate selective placement for certain physical characteristics, especially height. The rationale here is to avoid placing a child into home where the child might "stick out like a sore thumb." The empirical evidence suggests that for behavioral traits selective placement is very small or nonexistent. However, one must be cautious in interpreting adoption data on phenotypes like attitudes toward abortion that may correlate strongly with religious affiliation.

The second assumption about the adoption method concerns the representativeness of the adoptive families. Adoptive parents are screened—sometimes intensively—on issues of positive mental and physical health, the ability to financially support a child, and the probability of providing a safe and secure home for the child.

Researchers mistakenly assume that the screening process is *for* wealth, *for* positive mental health, etc. Instead, the process is *against* extreme poverty and *against* serious psychopathology. As a result, mean income of adoptive families is not very different from average income in the general population—it is just that the lower tail of the income distribution is missing.

Selection against psychopathology is a more serious matter. Parental alcoholism, criminal behavior, psychosis, drug abuse, and several other factors exclude a parent from adoption. As a result, there may be a restriction in range in the environments provided by adoptive parents, making it very difficult to detect a correlation between adoptees and their adoptive relatives. Hence, one should be cautious in interpreting low correlations among adoptive relatives as evidence for a lack of family environmental influence on the trait. Once again, one must consider restriction in range on a trait by trait basis. It may be very important for phenotypes like antisocial behavior, but rather weak for personality traits.

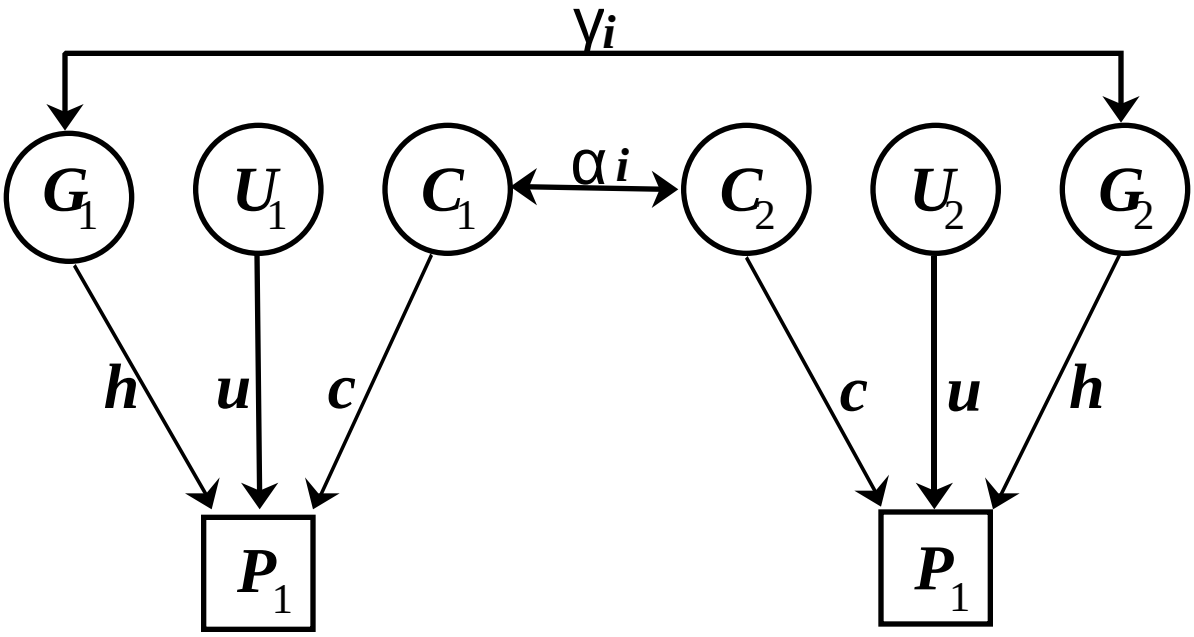
The Twin Method: Another Quantitative Model

There is a second quantitative model used for twin data. It has the same assumptions as the model given above, but it expresses the information in a different way. This model subdivides environmental values into two parts, *common environmental values* and *unique environmental values.* Common environmental values are *all those environmental factors that influence the trait of interest and at the same time make siblings raised together similar on the trait of interest*. Being raised in the same home, going to the same school, having friends in common are all factors that would be potentially included in the common environment. Be careful not to confuse common environment with simple events and circumstances that siblings share when they are raised together. Sharing is necessary but it is not sufficient to be in the common environment. The shared *environment must also influence the trait and make siblings similar*. For example, siblings raised together live in the same neighborhood. If neighborhood is a factor influencing the phenotype of juvenile delinquency, then neighborhood would be a common environmental factor for that phenotype. On the other hand, if neighborhood has no influence on the development of schizophrenia, then neighborhood is *not* a common environmental factor for the phenotype of schizophrenia. Although sibs share neighborhood, their shared neighborhood does *not* influence schizophrenia and does *not* make them similar for schizophrenia.

The unique environment is defined as *all those environmental factors that influence the trait and make siblings different from each other on the trait of interest*. Individual learning experiences, having different friends, being treated differently by parents are all aspects of the unique environment of a phenotype provided that they influence that phenotype. Note that when phenotypes are not measured with perfect accuracy (something that happens with virtually every behavioral phenotype), then measurement error is included in the unique environment.

In this model, h remains the correlation between genotypic values and phenotypic values, c is the correlation between common environmental values and phenotypic values,

and *u* is the correlation between unique environmental values. The meaning of γ_i remains



the same—the correlation between the genotypes of sibs. The quantity γ_i is the correlation between the common environments of sibs. For sibs raised together, $\alpha_i = 1.0$ but for sibs raised apart $\alpha_i = 0$.

The equation for the phenotypic variance is

$$1.0 = h^2 + c^2 + u^2 ,$$

where h^2 s the heritability, c^2 is the common environmentability, and u^2 is the unique environmentability. The correlation for MZ twins raised together is

$$R_{mz} = h^2 + c^2$$

and the correlation for DZ twins raised together is

$$R_{dz} = .5h^2 + c^2.$$

Again, we have three equations in three unknowns, so with some algebra, it can be shown that we can estimate heritability, common environmentability, and unique environmentability as

$$h^2 = 2(R_{mz} - R_{dz}),$$
$$c^2 = 2R_{dz} - R_{mz},$$

and

$$u^2 = 1 - R_{mz}.$$

In the example of sociability used above heritability is

$$h^2 = 2(R_{mz} - R_{dz}) = 2(.60 - .46) = .30$$

which is the same answer that we calculated before. Indeed, this is as it should be because the equation for h^2 is the same in the two models. The estimate of common environmentability is

$$c^2 = 2R_{dz} - R_{mz} = 2(.45) - .60 = .30 .$$

This means that 30% of the individual differences in observed sociability are due in some way to those environmental factors that sibs share *and* make sibs similar in sociability. The estimate of unique environmentability is

$$u^2 = 1 - R_{mz} = 1 - .60 = .40 .$$

Thus, 40% of observed individual differences in sociability may be traced in some way to the unique environment—all those unique, idiosyncratic experiences that influence sociability *and* make sibling different, including errors of measurement.

There is no substantive difference between the current model and the one outlined above. Indeed, the two models extract the same information but express it in different ways. With some algebra (not shown here), it can be shown that

$$e^2 = c^2 + u^2,$$

$$\eta = \frac{c^2}{c^2 + u^2},$$

and

$$c^2 = \eta e^2.$$

This is reasonable because the current model merely separates the total environmentability (e^2) into two components ($c^2 + u^2$) and expresses the information about the environment in terms of those two components.

Advanced Topics:

In this section, mathematical models are developed for the computation of different types of genetic variance. Several substantive points about genetic variance components and their effect on the analysis of behavioral data are also made. The reader uninterested in the mathematics can read the text boxes to gain the substantive conclusions.

Quantitative geneticists partition total genetic variance into three types—*additive*, *dominance*, and *epistatic* variance. Additive genetic variance measures the extent to which phenotypic individual differences are predictable from the additive effects of allelic substitutions. Dominance genetic variance is variance associated with dominant gene action—the fact that the genetic value for a heterozygote is not exactly the average of the genetic value of the two homozygotes. Epistatic genetic variance is the variance associated with the statistical interaction among loci—gene by gene interaction as it is often called..

Additive and dominance variance may be illustrated by examining a single locus, designated here as M locus, with two alleles M_1 and M_2. The additive effect of allele M_2 is the average change in genotypic values seen by substituting an M_2 allele for an M_1 allele. To find this effect, simply construct a new variable, called X_1 here, that equals the number of M_2 alleles for the individual's genotype. For genotype M_1M_1, $X_1 = 0$; for M_1M_2, $X_1 = 1$; and for M_2M_2, $X_1 = 2$. To account for dominance, construct another new variable, X_2, with values of 0, 1, and 0 for, respectively, genotypes M_1M_1, M_1M_2, and M_2M_2. Table

X.X provides hypothetical data set up with the new variables[11]. It is assumed that the phenotype is scaled so that the population mean is 100 and the population standard deviation is 15.

Table X.X. Hypothetical data for estimating genetic variance components at a single locus.

			Numerical Codes:	
Genotype	Frequency	Genotypic Value	Additive $= X_1$	Dominance $= X_2$
M_1M_1	.16	94	0	0
M_1M_2	.48	96	1	1
M_2M_2	.36	108	2	0

If we had actual data on individuals we would calculate the additive and dominance effects and their variance components by performing two regression. In the first, we would regress the phenotypic score, noted as Y herein, on X_1. The squared multiple correlation from this regression equals the *additive heritability* (h^2_A), the proportion of phenotypic variance associated with additive gene action at this locus. For the data in Table X.X, $R^2 = .137 = h^2_A$, so 13.7% of phenotypic variance is predicted from additive gene effects at this locus. The regression line for this equation will be the strait line of best fit that through the three genotypic values for the three genotypes. It is illustrated in Figure X.X.

The second regression model would be of the form

$$Y = b_0 + b_1 X_1 + b_2 X_2 \ .$$

The intercept, b_0, will equal the genotypic value for M_1M_1. The regression coefficient b_1 equals the average effect of substituting allele M_2 for M_1 in any genotype. Finally, the coefficient b_2 equals the genotypic value of the heterozygote less the average of the genotypic values of the two homozygotes. This measures dominant gene action. For the present example, $b_0 = 94$, $b_1 = 7$, and $b_2 = -5$. The value of coefficient b_1 informs us that on average one M_2 allele increases phenotypic values by 7 units. Because the value of b_2 is not 0, we can conclude that there is some degree of dominance. Because $b_2 = -5$, we can conclude that there is partial dominance for allele M_1 so that the genotypic value of the heterozygote is moved 5 units away from the midpoint of the two homozygotes and toward genotype M_1M_1.

[11] This coding system can be used to account for any number of alleles at a locus. For example, to model additive effects of allele M_3, construct another variable giving the number of M_3 alleles in a genotype. There would then be three dominance variables, one for the heterozygote M_1M_2, a second for the heterozygote M_1M_3, and the third for the heterozygote M_2M_3. For each of the three dominance variables, the appropriate heterozygote would have a value of 1 and all other genotypes would be assigned a value of 0.

The multiple correlation from this model equals the additive heritability plus the dominance heritability ($h_A^2 + h_D^2$). For the present example, R^2 = .162. Dominance heritability can be found by subtracting the R^2 from the first regression model from this value: $h_D^2 = .162 - .137 = .025$.

The regression coefficients can also be used to calculate additive and dominance heritability. Let p_1 denote the frequency of allele M_1 and p_2, the frequency of M_2. Then,

$$h_A^2 = 2p_1p_2[b_1 + (p_1 - p_2)b_2]^2 ,$$

and

$$h_D^2 = (2p_1p_2b_2)^2 .$$

To examine epistasis, consider the N locus with alleles N_1 and N_2. Just as we created two new variables for the M locus, we could also create two new variables to model the additive effect and the dominance effect at this locus. Call these variables X_3 and X_4. For genotypes N_1N_1, N_1N_2,and N_2N_2, the respective values for X_3 will be 0, 1, and 2; the respective values for X_4 would be 0, 1, and 0. The coding for the additive and the dominance effects at both the M and the N loci are given in Table X.X.

Table X.X. An example of numerical coding to calculate genetic variance components for two loci with two alleles at each locus.

		Numerical Codes:							
						Interactive			
		***M* Locus:**		***N* Locus:**		A*A	A*D	D*A	D*D
Genotypes:		Add.	Dom.	Add.	Dom.	X_1*X_3	X_1*X_4	X_2*X_3	X_2*X_4
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
$M_1M_1N_1N_1$		0	0	0	0	0	0	0	0
$M_1M_1N_1N_2$		0	0	1	1	0	0	0	0
$M_1M_1N_2N_2$		0	0	2	0	0	0	0	0
$M_1M_2N_1N_1$		1	1	0	0	0	0	0	0
$M_1M_2N_1N_2$		1	1	1	1	1	1	1	1
$M_1M_2N_2N_2$		1	1	2	0	2	0	2	0
$M_2M_2N_1N_1$		2	0	0	0	0	0	0	0
$M_2M_2N_1N_2$		2	0	1	1	2	2	0	0
$M_2M_2N_2N_2$		2	0	2	0	4	0	0	0

In regression, an interaction between two predictor variables is modeled by creating a new variable that is the product of the two predictor variables and entering this variable. Genetic epistasis is modeled in the same way. Multiplying the additive variable for the *M* locus (X_1) by the additive variable for the *N* locus (X_3) gives a new variable (X_5 in Table X.X) that geneticists call *additive by additive epistasis*. The variance associated with this is termed *additive by additive epistatic variance*.

There are two different ways to model the interaction between an additive effect at one locus and a dominance effect at a second locus. First, we could multiply the additive variable for the *M* locus by the dominance variable for the *N* locus. This new variable is given as X_6 in Table X.X. The second way is to multiply the dominance variable for *M* by the additive variable for *N*, giving variable X_7 in Table X.X. Together variables X_6 and X_7 model *additive by dominance epistasis* and the variance associated with these two variables is called *additive by dominance epistatic variance*.

The final interactive term is the product of the two sominance variables for the *M* and *N* loci. It is given as X_8 in Table X.X. This is *dominance by dominance epistasis* and it associated variance is *dominance by dominance epistatic variance*.

Estimation of epistatic variance components proceeds in the hierarchical manner described previously for additive and dominance variance at a single locus. The regression models and their associated variance components are listed in Table X.X. To calculate the proportion of phenotypic variance associated with any single effect, one simply takes the R^2 for the model and subtracts from it the R^2 of the model above it in Table X.X.

Table X.X. Regression models for estimating heritability components.

	Model:	**Heritability Component (R^2):**
1	$Y = X_1 + X_3$	Additive = h^2_A
2	$Y = X_1 + X_2 + X_3 + X_4$	Additive + Dominance = $h^2_A + h^2_D$
3	$Y = X_1 + X_2 + X_3 + X_4 + X_5$	Additive + Dominance + A*A Epistasis = $h^2_A + h^2_D + h^2_{AA}$
4	$Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$	Additive + Dominance + A*A Epistasis + A*D Epistasis = $h^2_A + h^2_D + h^2_{AA} + h^2_{AD}$
5	$Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$	Additive + Dominance + A*A Epistasis + A*D Epistasis + D*D Epistasis = $h^2_A + h^2_D + h^2_{AA} + h^2_{AD} + h^2_{DD}$

For a numerical example consider the genotypic values presented in Table X.X measured on a scale such as IQ with a population mean of 100 and a population standard deviation of 15. In calculating these numbers, the frequency of allele M_2 was set to .60 and the frequency of N_2 was .70. Variables X_1 through X_8 were constructed as in Table X.X,

Table X.X. Genotypic values for two genotypes.

	N_1N_1	N_1N_2	N_2N_2	Mean
M_2M_2	93.00	103.78	114.37	108.00
M_1M_2	93.00	95.00	97.41	96.00
M_1M_1	93.00	94.00	94.18	94.00
Mean	93.00	98.00	103.00	100.00

and the regression models in Table X.X were fitted to the data. The results of the regression models and the heritability components given in Table X.X. In this table, the heritability due to dominance variance (h^2_D) equals the R^2 for model 2 less the R^2 for model 1 or .20881 - .18321 = .0256. The heritability due to additive by additive epistasis (h^2_{AA}) equals the R^2 for model 3 less the R^2 for model 2 or .23854 - .20881 = .0297, and so on.

Table X.X. Components of heritability for the data in Table X.X.

	Model:	R^2	**Heritability Component**
1	$Y = X_1 + X_2$	.18321	h^2_A = .1832
2	$Y = X_1 + X_2 + X_3 + X_4$	.20881	h^2_D = .0256
3	$Y = X_1 + X_2 + X_3 + X_4 + X_5$	.23854	h^2_{AA} = .0297
4	$Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$	.24309	h^2_{AD} = .0046
5	$Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$	.24314	h^2_{DD} = .0001

Notice how the hierarchical decomposition of genetic variance tends to extract large amounts for the additive variance and progressively smaller amounts for the dominance and epistatic variance.

TEXT BOX: Gene Action and Genetic Variance Components.

Gene action is requires for a variance component. For example, without some degree of dominant gene action, there can be no dominance variance. However, genetic variance components are a function of *both* gene action and genotypic frequencies. Consequently, one can have strong gene action, but if the genotypic frequencies are just right, then the variance component associated with it can be very small. To illustrate this consider the two equations given in the text to compute the additive and dominance variance at a single locus from the regression parameters,

$$h_A^2 = 2p_1p_2[b_1 + (p_1 - p_2)b_2]^2 ,$$

and

$$h_D^2 = (2p_1p_2b_2)^2 .$$

Let us assume that allele M_2 shows complete dominance to allele M_1. In this case, $b_2 = b_1$. Let us substitute b_1 for b_2 in the above equations and derive the ratio of additive to dominance variance,

$$\frac{h_A^2}{h_D^2} = \frac{2p_1p_2[b_1 + (p_1 - p_2)b_1]^2}{(2p_1p_2b_1)^2}$$

which reduces to

$$\frac{h_A^2}{h_D^2} = 2\frac{p_1}{p_2} .$$

Even though we have modeled a completely dominant gene, this equation tells us that the ratio of additive to dominance variance depends *only* on the allele frequencies! When allele frequencies are even (i.e., $p_1 = p_2$) then the ratio is 2 and we will have twice as much additive variance as dominance variance. As the frequency of the recessive allele increases, p_1 becomes larger and larger and there is more and more additive variance relative to dominance variance. This tells us that rare dominant alleles have large additive variance relative to their dominance variance.

As p_1 becomes smaller and smaller relative to p_2, the ratio will get less than 1 and approach 0. Thus, rare recessive alleles have large dominance variance relative to their additive variance.

TEXT BOX: Hierarchical Decomposition of Genetic Variance Components

Return to the single locus with two alleles. In performing the regression to calculate additive genetic variance, we fitted a regression line that minimizes the squared differences between the observed genotypic values and those predicted by the regression line. This procedure is deliberately geared to maximize additive genetic variance (the variance associated with the regression line) and minimize residual genetic variance (the dominance genetic variance).

The hierarchical decomposition of genetic variance components follows this maximization/minimization algorithm. In the polygenic case, after additive variance has been extracted the procedure will try to maximize dominance genetic variance and minimize the residual genetic variance (epistatic variance). After additive and dominance variances have been extracted, the regression will maximize the additive by additive epistatic variance and minimize the other epistatic variance components.

As a consequence, additive genetic variance tends to be the largest component with continually smaller and smaller components following. There is no mathematical guarantee that this will *always* happen, but the pattern is almost always expressed in biologically plausible models of gene action. The major exception to this rule is the phenotype due to rare recessive alleles at a single locus.

Any human behavioral trait is are probably influenced by several different genes. It is unlikely that nonadditive gene action (dominance and epistasis) are completely absent at all loci that contribute to behavior. Would anyone care to bet that the genotypic value of the heterozygote lies exactly at the midpoint of the two homozygotes for every single locus operating in the central nervous system? But the hierarchical decomposition of variance for polygenic traits is likely to generate considerable additive variance with relatively small dominance and epistatic variance. The net result is that the typical assumption used in fitting genetic models to human behavioral data—that all genetic variance is additive—will give the wrong answer but will not give a substantively misleading answer.

To illustrate, let us examine what twin correlations would look like using the variance components given in Table X.X. The identical twin correlation would be

$$R_{mz} = h^2_A + h^2_D + h^2_{AA} + h^2_{AD} + h^2_{DD} = .243 ,$$

and the fraternal twin correlation would be

$$R_{dz} = \frac{1}{2}h^2_A + \frac{1}{4}h^2_D + \frac{1}{4}h^2_{AA} + \frac{1}{8}h^2_{AD} + \frac{1}{16}h^2_{DD} = .106 .$$

(See Kempthorne ()for the coefficients for nonadditive effects for relatives other than MZ twins.) If we estimated heritability with the traditional formula that assumes no additive genetic variance, we would have

$$\hat{h}^2 = 2(R_{mz} - R_{dz}) = 2(.243 - .106) = .274 .$$

Although the estimate of .274 is not correct, it is not very different from the total heritability of .243.