Twin and Adoption Designs

Vinay Tummarakota

July 2025

Disclaimer:

- This document is designed to provide an introduction to the way that twin and adoption studies estimate the heritability of traits. It is not comprehensive, nor should the derivations shown be interpreted as formal mathematical proofs.
- While I tried to accompany each derivation with as much explanation as possible, I will admit that I sometimes skipped over certain explanations that I felt a reader with enough statistical background would be able to intuit. If there's anything confusing about the derivations, feel free to send me a DM on Twitter/X or leave a comment over at my substack "Unboxing Politics".
- The specific notation for the ACE model that I am using is from David Cesarini's paper "Essays on Genetic Variation and Economic Behavior" (specifically Section III).
- If you notice any issues, feel free to send me a DM on Twitter/X or leave a comment over at my substack "Unboxing Politics".

1 The ACE Model

Let P represent a phenotype of interest (e.g. IQ, BMI, height, personality) for a given individual. We assume that the phenotype P can be partitioned into three additive and independent components.

$$P = aA + cC + e\epsilon$$

These components have the following definitions:

1. aA represents the component of the phenotype caused by additive genetic effects.
2. cC represents the component of the phenotype caused by shared environmental effects (i.e. environmental variables which make a given pair of relatives more similar to one another with respect to the phenotype in question).

3. $e\epsilon$ represents the component of the phenotype caused by non-shared environmental effects (i.e. environmental variables that are uncorrelated between a given pair of relatives which impact the phenotype in question).

Note that all capitalized variables (i.e. the $PACE$ variables) are assumed to have a mean of 0 and variance of 1. This model of an individual's phenotype is called the "ACE" model and serves as the basis for the classical twin design and many adoption designs.

2 Heritability

Heritability is defined as the variation in a given phenotype that is driven by variation in genetics. In the ACE model, heritability is calculated like so:

$$\begin{aligned} h^2 &= \frac{Var(aA)}{Var(P)} \\ &= a^2 \cdot \frac{Var(A)}{Var(P)} \\ &= a^2 \cdot \frac{1}{1} \\ &= a^2 \end{aligned}$$

Because the ACE model assumes that all genetic effects are additive, the estimated heritability is described as *narrow-sense* heritability. This type of heritability contrasts with *broad-sense* heritability which encompasses non-additive genetic variation in addition to additive genetic variation.

3 Classical Twin Design

In the classical twin design, heritability is estimated like so:

$$h^2 = 2 \cdot (r_{MZ} - r_{DZ})$$

To understand the intuition behind this formula, I will walk through a derivation. To simplify the math behind this derivation, we will assume that the variance of the phenotype in question is one (i.e. $Var(P) = 1$).

3.0.1 The Monozygotic Twin Correlation

Let P_{MZ1} and P_{MZ2} denote the phenotype values for a given pair of monozygotic twins. Under the ACE model, the covariance between the phenotypes of monozygotic twins can be written like so:

$$Cov(P_{MZ1}, P_{MZ2}) = Cov(aA_{MZ1} + cC_{MZ1} + e\epsilon_{MZ1}, aA_{MZ2} + cC_{MZ2} + e\epsilon_{MZ2})$$

Under the assumption of no gene-environment covariance, we can simplify like so:

$$= Cov(aA_{MZ1}, aA_{MZ2}) + Cov(cC_{MZ1}, cC_{MZ2}) + Cov(e\epsilon_{MZ1}, e\epsilon_{MZ2})$$

Because monozygotic twins ...

- share 100% of their genomes, the additive genetic components of their phenotypes are exactly equal (i.e. $A_{MZ1} = A_{MZ2}$).
- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{MZ1}, \epsilon_{MZ2}) = 0$).

Thus, we can simplify the above equation like so:

$$\begin{aligned} &= Cov(aA_{MZ1}, aA_{MZ1}) + Cov(cC_{MZ1}, cC_{MZ2}) \\ &= a^2 Var(A_{MZ1}) + c^2 Cov(C_{MZ1}, C_{MZ2}) \\ &= a^2 + c^2 Cov(C_{MZ1}, C_{MZ2}) \end{aligned}$$

Note that the correlation in the phenotype between monozygotic twins can be expressed like so:

$$Corr(P_{MZ1}, P_{MZ2}) = \frac{Cov(P_{MZ1}, P_{MZ2})}{\sqrt{Var(P_{MZ1}) \cdot Var(P_{MZ2})}}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e. $Var(P_{MZ1}) = Var(P_{MZ2}) = 1$), the monozygotic twin correlation is equal to the monozygotic twin covariance shown above.

3.0.2 The Dizygotic Twin Correlation

Let P_{DZ1} and P_{DZ2} denote the phenotype values for a given pair of dizygotic twins. Under the ACE model, the covariance between the phenotypes of dizygotic twins can be written like so:

$$Cov(P_{DZ1}, P_{DZ2}) = Cov(aA_{DZ1} + cC_{DZ1} + e\epsilon_{DZ1}, aA_{DZ2} + cC_{DZ2} + e\epsilon_{DZ2})$$

Because dizygotic twins ...

- share 50% of their genomes, the covariance between the additive genetic components of their phenotypes is equal to one-half (i.e. $Cov(A_{DZ1}, A_{DZ2}) = \frac{1}{2}$).
- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{DZ1}, \epsilon_{DZ2}) = 0$).

Thus, we can simplify the above equation like so:

$$\begin{aligned}
&= Cov(aA_{DZ1}, aA_{DZ1}) + Cov(cC_{DZ1}, cC_{DZ2}) \\
&= a^2 Cov(A_{DZ1}, A_{DZ2}) + c^2 Cov(C_{DZ1}, C_{DZ2}) \\
&= \frac{1}{2}a^2 + c^2 Cov(C_{DZ1}, C_{DZ2})
\end{aligned}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e. $Var(P_{DZ1}) = Var(P_{DZ2}) = 1$), the dizygotic twin correlation is equal to the dizygotic twin covariance shown above.

3.0.3 Estimating Heritability

With the monozygotic and dizygotic twin correlations specified, we can now estimate heritability like so:

$$\begin{aligned}
h^2 &= 2 \cdot (r_{MZ} - r_{DZ}) \\
&= 2 \cdot (a^2 + c^2 Cov(C_{MZ1}, C_{MZ2}) - (\frac{1}{2}a^2 + c^2 Cov(C_{DZ1}, C_{DZ2}))) \\
&= 2 \cdot (\frac{1}{2}a^2 + c^2 Cov(C_{MZ1}, C_{MZ2}) - c^2 Cov(C_{DZ1}, C_{DZ2}))
\end{aligned}$$

By the Equal Environments Assumption, the covariance between shared environment components of the phenotype for monozygotic twins is equal to the covariance between the shared environment components of the phenotype for dizygotic twins (i.e. $Cov(C_{MZ1}, C_{MZ2}) = Cov(C_{DZ1}, C_{DZ2})$). Thus, we can simplify the equation to yield heritability like so:

$$\begin{aligned}
&= 2 \cdot (\frac{1}{2}a^2) \\
&= a^2
\end{aligned}$$

3.0.4 Assortative Mating Bias

The $2 \cdot (r_{MZ} - r_{DZ})$ formula assumes that the covariance between the additive genetic components of the phenotype for dizygotic twins is one-half (i.e. $Cov(A_{DZ1}, A_{DZ2}) = \frac{1}{2}$). However, this assumption is violated in the presence of a type of assortative mating known as *genetic homogamy* (where people mate with partners who are genetically similar to themselves)

To understand the impact of assortative mating bias, we must first understand what impact assortative mating has on the covariance between the additive genetic components of non-identical siblings.

To start, let us suppose that the additive genetic components of the *parent's* phenotypes are correlated at m (i.e. $Corr(A_{MOM}, A_{DAD}) = m$). Then, the covariance between the additive genetic components of the mother's and father's

phenotype can be expressed like so:

$$\begin{aligned} Cov(A_{MOM}, A_{DAD}) &= Corr(A_{MOM}, A_{DAD}) \cdot \sqrt{Var(A_{MOM}) \cdot Var(A_{DAD})} \\ &= m \cdot \sqrt{Var(A_{MOM}) \cdot Var(A_{DAD})} \\ &= m \cdot \sqrt{1 \cdot 1} \\ &= m \end{aligned}$$

Next, note that the additive genetic component of the child's phenotype is equal to an average of the mother and father's additive genetic components plus a random noise term which reflects the random inheritance of genetic variants from mom vs dad:

$$A_{CHILD} = \frac{1}{2}(A_{MOM} + A_{DAD}) + \epsilon$$

We can now estimate the covariance of the additive genetic component of the phenotype between two non-identical children of the same parents like so:

$$Cov(A_{CHILD1}, A_{CHILD2}) = Cov(\frac{1}{2}(A_{MOM} + A_{DAD}) + \epsilon_1, \frac{1}{2}(A_{MOM} + A_{DAD}) + \epsilon_2)$$

Because the noise terms are uncorrelated with one another, they drop out of the equation and we can simplify like so:

$$\begin{aligned} &= \frac{1}{4}Cov(A_{MOM}, A_{MOM}) + \frac{1}{4}Cov(A_{DAD}, A_{DAD}) + \frac{1}{2}Cov(A_{MOM}, A_{DAD}) \\ &= \frac{1}{4}Var(A_{MOM}) + \frac{1}{4}Var(A_{DAD}) + \frac{1}{2}Cov(A_{MOM}, A_{DAD}) \\ &= \frac{1}{2} + \frac{m}{2} \\ &= \frac{1+m}{2} \end{aligned}$$

Thus, genetic homogamy of m increases the covariance between the additive genetic components of the phenotype for dizygotic twins from $\frac{1}{2}$ to $\frac{1+m}{2}$. Importantly, this result means that our heritability estimation formula is downwardly biased!

$$\begin{aligned} h^2 &= 2 \cdot (r_{MZ} - r_{DZ}) \\ &= 2 \cdot (a^2 + c^2 Cov(C_{MZ1}, C_{MZ2}) - (\frac{1+m}{2}a^2 + c^2 Cov(C_{DZ1}, C_{DZ2}))) \\ &= 2 \cdot (\frac{1-m}{2}a^2 + c^2 Cov(C_{MZ1}, C_{MZ2}) - c^2 Cov(C_{DZ1}, C_{DZ2})) \\ &= 2 \cdot (\frac{1-m}{2}a^2) \\ &= (1-m)a^2 \end{aligned}$$

Instead of yielding heritability equal to a^2, the formula yields $(1-m)a^2$. To eliminate this assortative mating bias, we can divide the estimated heritability by $1-m$ like so:

$$h^2 = \frac{h^2_{twin}}{1-m}$$

3.0.5 Non-Additivity Bias

The $2 \cdot (r_{MZ} - r_{DZ})$ formula assumes that the phenotype of interest can be additively separated into three independent components: an additive genetic component (A), a shared environmental component (C), and a non-shared environmental component (E). However, this assumption is violated when genetic effects are non-additive. To deal with non-additivity, we must extend the ACE model into include a fourth component D which captures dominance and a fifth component $A \times A$ which captures epistasis (i.e. gene-gene interactions). Please excuse the bastardized notation used to denote epistasis.

$$P = aA + cC + dD + (a \times a)(A \times A) + e\epsilon$$

Because monozygotic twins fully share non-additive genetic effects, the monozygotic twin correlation can be written like so:

$$r_{MZ} = a^2 + c^2 Cov(C_{MZ1}, C_{MZ2}) + d^2 + (a \times a)^2$$

Because dizygotic twins share only *one-fourth* of non-additive genetic effects, the dizygotic twin correlation can be written like so:

$$r_{DZ} = \frac{1}{2}a^2 + c^2 Cov(C_{DZ1}, C_{DZ2}) + \frac{1}{4}d^2 + \frac{1}{4}(a \times a)^2$$

The $2 \cdot (r_{MZ} - r_{DZ})$ formula yields the following heritability estimate:

$$\begin{aligned}
h^2 &= 2 \cdot (r_{MZ} - r_{DZ}) \\
&= 2 \cdot (a^2 + c^2 Cov(C_{MZ1}, C_{MZ2}) + d^2 + (a \times a)^2 - (\frac{1}{2}a^2 + c^2 Cov(C_{DZ1}, C_{DZ2}) + \frac{1}{4}d^2 + \frac{1}{4}(a \times a)^2)) \\
&= 2 \cdot (\frac{1}{2}a^2 + \frac{3}{4}d^2 + \frac{3}{4}(a \times a)^2) \\
&= a^2 + \frac{3}{2}d^2 + \frac{3}{2}(a \times a)^2
\end{aligned}$$

Instead of yielding a *broad-sense* heritability estimate equal to $a^2 + d^2 + (a \times a)^2$, the formula yields $a^2 + \frac{3}{2}d^2 + \frac{3}{2}(a \times a)^2$. In other words, the formula over-estimates non-additive genetic variation by 50%!

3.0.6 De Novo Variant Bias

The $2 \cdot (r_{MZ} - r_{DZ})$ formula assumes that the covariance between the additive genetic components of the phenotype is one for monozygotic twins and one-half

for dizygotic twins (i.e. $Cov(A_{MZ1}, A_{MZ2}) = 1$ and $Cov(A_{DZ1}, A_{DZ2}) = \frac{1}{2}$). However, this assumption is violated when de novo variants that are unshared between twins impact the phenotype of interest.

To understand the impact of this de novo variant bias, we can split the additive genetic component of the phenotype into three parts:

- *MZ-Shared De Novo Variants*: These are de novo variants that are shared by MZ twins but not shared by DZ twins.
- *Non-Shared De Novo Variants*: These are de novo variants that are not shared by both MZ and DZ twins.
- *All Other Variants*: These are all other genetic variants.

This split yields the following extended ACE model:

$$P = aA + m_{MZ-shared}M_{MZ-Shared} + m_{Non-Shared}M_{Non-Shared} + cC + \epsilon E$$

From the above model, it is easy to see that the covariance between the phenotypes of monozygotic and dizygotic twins should be equal to the following:

$$\begin{aligned} Cov(P_{MZ1}, P_{MZ2}) &= a^2 + m^2_{MZ-Shared} + c^2 \\ Cov(P_{DZ1}, P_{DZ2}) &= \frac{1}{2}a^2 + c^2 \end{aligned}$$

Importantly, this result means our heritability estimation formula is biased!

$$\begin{aligned} h^2 &= 2 \cdot (r_{MZ} - r_{DZ}) \\ &= 2 \cdot (a^2 + m^2_{MZ-Shared} + c^2 - \frac{1}{2}a^2 + c^2) \\ &= a^2 + 2 \cdot m^2_{MZ-Shared} \end{aligned}$$

Instead of yielding heritability equal to $a^2 + m^2_{MZ-Shared} + m^2_{Non-Shared}$, the formula yields $a^2 + 2 \cdot m^2_{MZ-Shared}$. In other words, we have over-estimated the variation driven by MZ-Shared de novo variants and under-estimated the variation driven by Non-Shared de novo variants. To eliminate this bias, we would need to know the percentage of variation in the trait driven by MZ-Shared and Non-Shared de novo variants.

4 Adoption Designs

Adoption designs estimate heritability in different ways depending on the relative pairs under consideration. We will consider how assortative mating impacts each design separately. All calculations will assume the ACE model and that the variance of the phenotype is one (i.e. $Var(P) = 1$).

4.1 Birth Parent - Adoptive Child Formulation

In the Birth Parent-Adoptive Child Formulation, heritability is estimated like so:

$$h^2 = 2 \cdot r_{BP-AC}$$

4.1.1 The Birth Parent - Adoptive Child Correlation

Let P_{BP} and P_{AC} denote the phenotype values for the birth parent and their child who was adopted away, respectively. Under the ACE model, the covariance between the phenotypes of the birth parent and the adoptive child can be written like so:

$$Cov(P_{BP}, P_{AC}) = Cov(aA_{BP} + cC_{BP} + e\epsilon_{BP}, aA_{AC} + cC_{AC} + e\epsilon_{AC})$$

Because birth parents and their adopted-away children:

- share 50% of their genomes, the covariance between the additive genetic components of their phenotypes is equal to one-half (i.e. $Cov(A_{BP}, A_{AC}) = \frac{1}{2}$).
- do not live in the same environment, the covariance between the shared environment components of their phenotype is assumed to be equal to zero (i.e. $Cov(C_{BP}, C_{AC}) = 0$).
- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{BP}, \epsilon_{AC}) = 0$).

Thus, we can simplify the above equation like so:

$$\begin{aligned} &= Cov(aA_{BP}, aA_{AC}) \\ &= a^2 Cov(A_{BP}, A_{AC}) \\ &= \frac{1}{2}a^2 \end{aligned}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e $Var(P_{BP}) = Var(P_{AC}) = 1$), the birth parent-adoptive child correlation is equal to the birth parent-adoptive child covariance shown above.

4.1.2 Estimating Heritability

With the birth parent-adoptive child correlation specified, we can now estimate heritability like so:

$$\begin{aligned} h^2 &= 2 \cdot r_{BP-AC} \\ &= 2 \cdot \frac{1}{2}a^2 \\ &= a^2 \end{aligned}$$

4.1.3 Assortative Mating Bias

The $2 \cdot r_{BP-AC}$ formula assumes that the covariance between the phenotype of the birth parents and their adopted-away children is one-half times the heritability (i.e. $Cov(P_{BP}, P_{AC}) = \frac{1}{2}a^2$). However, this assumption is violated in the presence of one of two types of assortative mating: *genetic homogamy* (where people mate with partners who are genetically similar to themselves) and *social-genetic homogamy* (where environmentally-advantaged individuals mate with genetically-advantaged partners).

To understand the impact of these biases, we must first understand what impact assortative mating has on the covariance between the phenotypes of birth parents and their adopted-away children. We can express this covariance like so:

$$\begin{aligned} Cov(P_{BP}, P_{AC}) &= Cov(aA_{BP} + cC_{BP} + e\epsilon_{BP}, aA_{AC} + cC_{AC} + e\epsilon_{AC}) \\ &= Cov(aA_{BP} + cC_{BP}, aA_{AC} + cC_{AC}) \\ &= Cov(aA_{BP}, aA_{AC}) + Cov(aA_{BP}, cC_{AC}) + Cov(cC_{BP}, aA_{AC}) + Cov(cC_{BP}, cC_{AC}) \end{aligned}$$

Because birth parents and their adopted-away children do not live in the same environment:

- The covariance between the shared environment components of their phenotype is assumed to be equal to zero (i.e. $Cov(C_{BP}, C_{AC}) = 0$).
- The covariance between the genetic component of the parent's phenotype and the shared environment component of the child's phenotype is assumed to be equal to zero (i.e. $Cov(A_{BP}, C_{AC}) = 0$). For example, we assume that birth parents who are genetically predisposed to the trait of interest do not have sufficient time to cultivate an environment which also predisposes their adopted-away child to the trait of interest.

Thus, we can simplify the above equation like so:

$$\begin{aligned} &= Cov(aA_{BP}, aA_{AC}) + Cov(cC_{BP}, aA_{AC}) \\ &= a^2 Cov(A_{BP}, A_{AC}) + caCov(C_{BP}, A_{AC}) \end{aligned}$$

Note that we still keep the second term $Cov(cC_{BP}, aA_{AC})$ due to social-genetic homogamy. Under social-genetic homogamy, the non-genetic component of one parent's phenotype is correlated with the genetic component of the other parent's phenotype. Because the genetic component of the other parent's phenotype is inherited by the child, the non-genetic component of the first parent's phenotype also becomes correlated with genetic component of the child's phenotype.

Let us suppose that the additive genetic component of the *birth parents'* phenotypes are correlated at m (i.e. $Corr(A_{BPM}, A_{BPD}) = Cov(A_{BPM}, A_{BPD}) = m$) where BPM denotes the birth mother and BPD denotes the birth father.

Next, note that the additive genetic component of the child's phenotype is equal to an average of the mother and father's additive genetic components plus a random noise term which reflects the random inheritance of genetic variants from mom vs dad:

$$A_{CHILD} = \frac{1}{2}(A_{BPM} + A_{BPD}) + \epsilon$$

We can now estimate the covariance of the additive genetic component of the phenotype between birth parents and their adopted-away children. Without loss of generality, assume that the birth parent in question is the mother.

$$Cov(A_{BPM}, A_{AC}) = Cov(A_{BPM}, \frac{1}{2}(A_{BPM} + A_{BPD}) + \epsilon_{AC})$$

Because the noise term is random, it drops out of the equation and we can simplify like so:

$$\begin{aligned} &= \frac{1}{2}Cov(A_{BPM}, A_{BPM}) + \frac{1}{2}Cov(A_{BPM}, A_{BPD}) \\ &= \frac{1}{2} + \frac{m}{2} \\ &= \frac{1+m}{2} \end{aligned}$$

Thus, genetic homogamy of m increases the covariance between the additive genetic components of the phenotype for birth parents and their adopted-away offspring from $\frac{1}{2}$ to $\frac{1+m}{2}$. Importantly, this result means that our heritability estimation formula is upwardly biased!

$$\begin{aligned} h^2 &= 2 \cdot r_{BP-AC} \\ &= 2 \cdot (\frac{1+m}{2}a^2 + caCov(A_{BP}, C_{AC})) \\ &= (1+m)a^2 + 2caCov(A_{BP}, C_{AC}) \end{aligned}$$

Instead of yielding heritability equal to a^2, the formula yields $(1+m)a^2 + 2caCov(C_{BP}, A_{AC})$. To eliminate this assortative mating bias, we have to subtract the social-genetic homogamy component and then divide the estimated heritability by $1+m$ like so:

$$h^2 = \frac{h^2_{adoption} - 2caCov(C_{BP}, A_{AC})}{1+m}$$

This adjustment is quite challenging to implement in practice because estimating the social-genetic homogamy component is non-trivial. Consequently, geneticists in practice sometimes assume a specific type of assortative mating known as *direct phenotypic assortment* (where people mate directly on the phenotype of interest without respect to the underlying causes of the phenotype).

Under direct phenotypic assortment, the true heritability can be expressed as a function of the *phenotypic* correlation between partners (denoted r):

$$h^2 = \frac{h^2_{adoption}}{1+r}$$

This adjustment can be derived like so:

$$\begin{aligned} Cov(P_{BPM}, P_{AC}) &= Cov(P_{BPM}, A_{AC}) \\ &= Cov(P_{BPM}, \frac{1}{2}(A_{BPM} + A_{BPD}) + \epsilon_{AC}) \\ &= Cov(P_{BPM}, \frac{1}{2}A_{BPM}) + Cov(P_{BPM}, \frac{1}{2}A_{BPD}) \end{aligned}$$

Because the ACE components within the same individual are independent of one another, the covariance between the birth mother's phenotype and her genotype is solely due to the genotype. In other words, we can simplify the first term like so:

$$= \frac{1}{2}a^2 + Cov(P_{BPM}, \frac{1}{2}A_{BPD})$$

To simplify the second term, we need to assume direct phenotypic assortment. In mathematical terms, we assume that the covariance between the birth mother's phenotype and the birth father's *genotype* is completely explained by the birth father's *phenotype* (i.e. $Cov(P_{BPM}, A_{BPD}|P_{BPD})$. This assumption enables us to claim the following:

$$\begin{aligned} Cov(P_{BPM}, A_{BPD}) &= \frac{Cov(P_{BPM}, P_{BPD}) \cdot Cov(P_{BPD}, A_{BPD})}{Var(P_{BPD})} \\ &= r \cdot \frac{Cov(P_{BPD}, A_{BPD})}{Var(P_{BPD})} \\ &= r \cdot a^2 \end{aligned}$$

We can then simplify our original expression like so:

$$\begin{aligned} &= \frac{1}{2}a^2 + \frac{r}{2}a^2 \\ &= \frac{1+r}{2}a^2 \end{aligned}$$

Note that the estimated adoption heritability is simply two times the above expression. Thus, under the assumption of direct phenotypic assortment, the true heritability is equal to $\frac{h^2_{adoption}}{1+r}$.

4.2 Biological Child - Adoptive Child Formulation

In the Biological Child - Adoptive Child Formulation, heritability is estimated like so:

$$h^2 = 2 \cdot (r_{RP-BC} - r_{RP-AC})$$

4.2.1 The Rearing Parent-Biological Child Correlation

Let P_{RP} and P_{BC} denote the phenotype values for the rearing parent and their biological child, respectively. Under the ACE model, the covariance between the phenotypes of the rearing parent and the biological child can be written like so:

$$Cov(P_{RP}, P_{BC}) = Cov(aA_{RP} + cC_{RP} + e\epsilon_{RP}, aA_{BC} + cC_{BC} + e\epsilon_{BC})$$

Because rearing parents and their biological children:

- share 50% of their genomes, the covariance between the additive genetic components of their phenotypes is equal to one-half (i.e. $Cov(A_{RP}, A_{BC}) = \frac{1}{2}$).
- live in the same environment, the covariance between the shared environment components of their phenotype is assumed to be equal to one (i.e. $Cov(C_{RP}, C_{BC}) = 1$).
- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{RP}, \epsilon_{BC}) = 0$).

Thus, we can simplify the above equation like so:

$$\begin{aligned}
&= Cov(aA_{RP}, aA_{BC}) + Cov(cC_{RP}, cC_{BC}) \\
&= a^2 Cov(A_{BP}, A_{AC}) + c^2 Cov(C_{RP}, C_{BC}) \\
&= \frac{1}{2}a^2 + c^2 Cov(C_{RP}, C_{BC})
\end{aligned}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e $Var(P_{RP}) = Var(P_{BC}) = 1$), the rearing parent-biological child correlation is equal to the rearing parent-biological child covariance shown above.

4.2.2 The Rearing Parent-Adoptive Child Correlation

Let P_{RP} and P_{AC} denote the phenotype values for the rearing parent and their adoptive child, respectively. Under the ACE model, the covariance between the phenotypes of the rearing parent and the adoptive child can be written like so:

$$Cov(P_{RP}, P_{AC}) = Cov(aA_{RP} + cC_{RP} + e\epsilon_{RP}, aA_{AC} + cC_{AC} + e\epsilon_{AC})$$

Because rearing parents and their adoptive children:

- are assumed to share 0% of their genomes, the covariance between the additive genetic components of their phenotypes is assumed to be equal to zero (i.e. $Cov(A_{RP}, A_{AC}) = 0$).
- live in the same environment, the covariance between the shared environment components of their phenotype is assumed to be equal to one (i.e. $Cov(C_{RP}, C_{AC}) = 1$).

- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{RP}, \epsilon_{AC}) = 0$).

Thus, we can simplify the above equation like so:

$$\begin{aligned} &= Cov(cC_{RP}, cC_{AC}) \\ &= c^2 Cov(C_{RP}, C_{AC}) \end{aligned}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e $Var(P_{RP}) = Var(P_{AC}) = 1$), the rearing parent-adoptive child correlation is equal to the rearing parent-adoptive child covariance shown above.

4.2.3 Estimating Heritability

With the parent-biological child and rearing parent-adoptive child correlations specified, we can now estimate heritability like so:

$$\begin{aligned} h^2 &= 2 \cdot (r_{RP-BC} - r_{RP-AC}) \\ &= 2 \cdot (\frac{1}{2}a^2 + c^2 Cov(C_{RP}, C_{BC}) - (c^2 Cov(C_{RP}, C_{AC}))) \end{aligned}$$

By the Equal Environments Assumption, the covariance between shared environment components of the phenotype for rearing parents and their biological children is equal to the covariance between the shared environment components of the phenotype for rearing parents and their adoptive children (i.e. $Cov(C_{RP}, C_{BC}) = Cov(C_{RP}, C_{AC})$). Thus, we can simplify the equation to yield heritability like so:

$$\begin{aligned} &= 2 \cdot (\frac{1}{2}a^2) \\ &= a^2 \end{aligned}$$

4.2.4 Assortative Mating Bias

From the derivation in Section 4.1.3, it is easy to see that covariance between the phenotypes for rearing parents and their biological children should be equal to the following in the presence of assortative mating.

$$Cov(P_{RP}, P_{BC}) = \frac{1+m}{2}a^2 + caCov(C_{RP}, A_{BC}) + c^2 Cov(C_{RP}, C_{BC})$$

Importantly, this result means that our heritability estimation formula is upwardly biased!

$$\begin{aligned} h^2 &= 2 \cdot r_{RP-BC} - r_{RP-AC} \\ &= 2 \cdot (\frac{1+m}{2}a^2 + caCov(C_{RP}, A_{BC}) + c^2 Cov(C_{RP}, C_{BC}) - (c^2 Cov(C_{RP}, C_{AC}))) \\ &= 2 \cdot (\frac{1+m}{2}a^2 + caCov(C_{RP}, A_{BC})) \\ &= (1+m)a^2 + 2caCov(C_{RP}, A_{BC}) \end{aligned}$$

Instead of yielding heritability equal to a^2, the formula yields $(1+m)a^2 + 2caCov(C_{BP}, A_{AC})$. To eliminate this assortative mating bias, we have to subtract the social-genetic homogamy component and then divide the estimated heritability by $1+m$ like so:

$$h^2 = \frac{h^2_{adoption} - 2caCov(C_{BP}, A_{AC})}{1+m}$$

As discussed previously, this adjustment is quite challenging to implement in practice because estimating the social-genetic homogamy component is non-trivial. Thus, geneticists may decide to assume direct phenotypic assortment and estimate the true heritability as a function of the phenotypic correlation between mates (denoted r):

$$h^2 = \frac{h^2_{adoption}}{1+r}$$

4.3 Biological Sibling - Adoptive Sibling Formulation

In the Biological Sibling - Adoptive Sibling Formulation, heritability is estimated like so:

$$h^2 = 2 \cdot (r_{BS} - r_{AS})$$

4.3.1 The Biological Sibling Correlation

Let P_{BS1} and P_{BS2} denote the phenotype values for a given pair of biological siblings. Under the ACE model, the covariance between the phenotypes of monozygotic twins can be written like so:

$$Cov(P_{BS1}, P_{BS2}) = Cov(aA_{BS1} + cC_{BS1} + e\epsilon_{BS1}, aA_{BS2} + cC_{BS2} + e\epsilon_{BS2})$$

Because biological siblings ...

- share 50% of their genomes, the covariance between the additive genetic components of their phenotypes is one-half (i.e. $Cov(A_{BS1}, A_{BS2}) = \frac{1}{2}$).
- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{BS1}, \epsilon_{BS2}) = 0$).

Thus, we can simplify the above equation* like so:

$$\begin{aligned}
&= Cov(aA_{BS1}, aA_{BS1}) + Cov(cC_{BS1}, cC_{BS2}) \\
&= \frac{1}{2}a^2 Var(A_{BS1}) + c^2 Cov(C_{BS1}, C_{BS2}) \\
&= \frac{1}{2}a^2 + c^2 Cov(C_{BS1}, C_{BS2})
\end{aligned}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e. $Var(P_{BS1}) = Var(P_{BS2}) = 1$), the biological sibling correlation is equal to the biological sibling covariance shown above.

4.3.2 The Adoptive Sibling Correlation

Let P_{AS1} and P_{AS2} denote the phenotype values for a given pair of adoptive siblings (i.e. sibling pairs where at least one member is adopted). Under the ACE model, the covariance between the phenotypes of adoptive siblings can be written like so:

$$Cov(P_{AS1}, P_{AS2}) = Cov(aA_{AS1} + cC_{AS1} + e\epsilon_{AS1}, aA_{AS2} + cC_{AS2} + e\epsilon_{AS2})$$

Because adoptive siblings ...

- are assumed to share 0% of their genomes, the covariance between the additive genetic components of their phenotypes is assumed to be equal to zero (i.e. $Cov(A_{AS1}, A_{AS2}) = 0$).
- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{AS1}, \epsilon_{AS2}) = 0$).

Thus, we can simplify the above equation* like so:

$$\begin{aligned} &= Cov(aA_{AS1}, aA_{AS1}) + Cov(cC_{AS1}, cC_{AS2}) \\ &= c^2 Cov(C_{AS1}, C_{AS2}) \end{aligned}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e. $Var(P_{AS1}) = Var(P_{AS2}) = 1$), the adoptive sibling correlation is equal to the adoptive sibling covariance shown above.

4.3.3 Estimating Heritability

With the biological and adoptive sibling correlations specified, we can now estimate heritability like so:

$$\begin{aligned} h^2 &= 2 \cdot (r_{BS} - r_{AS}) \\ &= 2 \cdot (\frac{1}{2}a^2 + c^2 Cov(C_{BS1}, C_{BS2}) - c^2 Cov(C_{AS1}, C_{AS2})) \end{aligned}$$

By the Equal Environments Assumption, the covariance between shared environment components of the phenotype for biological siblings is equal to the covariance between the shared environment components of the phenotype for adoptive siblings (i.e. $Cov(C_{BS1}, C_{BS2}) = Cov(C_{AS1}, C_{AS2})$). Thus, we can simplify the equation to yield heritability like so:

$$\begin{aligned} &= 2 \cdot (\frac{1}{2}a^2) \\ &= a^2 \end{aligned}$$

4.3.4 Assortative Mating Bias

The $2 \cdot (r_{BS} - r_{AS})$ formula assumes that the covariance between the additive genetic components of the phenotype for biological siblings is one-half (i.e. $Cov(A_{DZ1}, A_{DZ2}) = \frac{1}{2}$). However, this assumption is violated in the presence of assortative mating.

By the derivation in Section 3.0.4, the covariance between the additive genetic components of biological siblings is equal to $\frac{1+m}{2}$. Importantly, this result means that our heritability estimation formula is upwardly biased!

$$\begin{aligned} h^2 &= 2 \cdot (r_{BS} - r_{AS}) \\ &= 2 \cdot (\frac{1+m}{2}a^2 + c^2 Cov(C_{BS1}, C_{BS2}) - c^2 Cov(C_{AS1}, C_{AS2})) \\ &= 2 \cdot (\frac{1+m}{2}a^2) \\ &= (1+m)a^2 \end{aligned}$$

Instead of yielding heritability equal to a^2, the formula yields $(1+m)a^2$. To eliminate this assortative mating bias, we can divide the estimated heritability by $1+m$ like so:

$$h^2 = \frac{h^2_{adoption}}{1+m}$$

4.4 Monozygotic Twins Reared Apart Formulation

In the Monozygotic Twins Reared Apart Formulation, heritability is estimated like so:

$$h^2 = r_{MZA}$$

4.4.1 The Monozygotic Twins Reared Apart Correlation

Let P_{MZA1} and P_{MZA2} denote the phenotype values for a given pair of monozygotic twins reared apart. Under the ACE model, the covariance between the phenotypes of monozygotic twins reared apart can be written like so:

$$Cov(P_{MZA1}, P_{MZA2}) = Cov(aA_{MZA1}+cC_{MZA1}+e\epsilon_{MZA1}, aA_{MZA2}+cC_{MZA2}+e\epsilon_{MZA2})$$

Because monozygotic twins reared apart...

- share 100% of their genomes, the additive genetic components of their phenotypes are exactly equal (i.e. $A_{MZA1} = A_{MZA2}$).
- do not live in the same environment, the covariance between the shared environment components of their phenotype is assumed to be equal to zero (i.e. $Cov(C_{MZA1}, C_{MZA2}) = 0$).

- experience uncorrelated non-shared environmental influences, the covariance between the non-shared environment components of their phenotypes is equal to zero (i.e. $Cov(\epsilon_{MZA1}, \epsilon_{MZA2}) = 0$.

Thus, we can simplify the above equation like so:

$$\begin{aligned} &= Cov(aA_{MZA1}, aA_{MZA1}) \\ &= a^2 Var(A_{MZA1}) \\ &= a^2 \end{aligned}$$

Because we have assumed that the variance in the phenotype is equal to one (i.e. $Var(P_{MZA1}) = Var(P_{MZA2}) = 1$), the monozygotic twin correlation is equal to the monozygotic twin covariance shown above.

4.4.2 Estimating Heritability

As demonstrated in the previous section, heritability is equal to the correlation between monozygotic twins reared apart.

4.4.3 Assortative Mating Bias

Because it is not possible to increase the genetic similarity of monozygotic twins beyond 100%, assortative mating has no impact on the resulting heritability estimate.