

Homework 10: Separated at Birth

36-402, Advanced Data Analysis

Due at 10:30 am on Tuesday, 24 April 2012

In many species of plants and animals, the genetic causes of bodily traits or behavior can be studied by breeding, or experimentally creating, organisms with specific genetic characteristics, and comparing different “strains” or “lines” which differ only in their genes, in known ways. This is not allowed for people¹. To try to get around this, researchers in psychology and medicine have used what are called “twin designs”, which compare identical twins raised together to identical twins raised apart, hoping to indirectly isolate genetic influences on aspects of the body (e.g., height, weight, cholesterol) or behavior (e.g., extroversion, school grades, voting, belief in astrology)². This problem set will use graphical causal models to examine the basis of such studies.

Unless specifically noted otherwise, you can assume that all variables are continuous, with mean zero, and each variable is a sum of its causal parents plus noise. You can assume that the noise term is independent and mean-zero for each variable, but you cannot assume that they all have the same variance; write the noise variance for X as σ_X^2 .

Figure 1 shows the models which are usually assumed in such studies, for identical twins raised together and identical twins raised apart. The letters on the arrows are the names of the path coefficients.

1. *Paths and equations* (10)

- (a) (4) For the models in Figure 1, write the equation for the trait of twin A in terms of its causal parents, the path coefficients, and the noise.
- (b) (1) How must these equations be changed for twin B?
- (c) (5) Explain why we can assume the path coefficients a and b are both always 1.

2. *Genetic component of variance* (25) The **heritability** of a trait is the fraction of its total variance which is genetically caused³. We will focus

¹Even if it was ethical, it wouldn't be very practical, because it takes so long to raise a new generation.

²All of these are traits which have actually been studied with twin designs.

³There is also a “narrow” heritability, which has to do with genetic causation which adds up across genes without interactions; we won't try to deal with that.

on the genetically-caused component of variance, i.e., the numerator of the heritability.

- (a) (4) Find the variances of the trait of twin A in both models, in terms of the path coefficients and the variances of its causal parents.
 - (b) (1) Find the variances of the trait of twin B in both models.
 - (c) (5) Find the covariances between twin A and twin B in both models.
 - (d) (5) Find the variance of the genes in terms of the observable variances and covariances. (You may have to compare twins raised together to twins raised apart.)
 - (e) (5) Find an expression for the family-environment variance, in terms of the variances and covariances of the observable variables.
 - (f) (5) Can you find the heritability from the previous results? If so, what is it? If not, what else would you need to know?
3. *Not separated that far* (20) Twins “raised apart” are placed in families through the same process at the same time; in many documented cases they end up living with relatives of their parents, and sometimes grow up in the same town, attending the same schools. This is referred to as a **community effect**: see Figure 2.
- (a) (5) Find the variances of the twins’ traits in the new models, in terms of the variances of the traits causal parents and the path coefficients. Explain why the path coefficients can all be taken to be 1.
 - (b) (5) Find the covariance between twins raised together and the covariance between twins raised apart in the new models.
 - (c) (5) Is your expression from Problem 2d still equal to the variance of the genes? If not, how would using the sample covariance between twins raised apart lead to a biased estimate of genetic variance?
 - (d) (5) Is your expression from Problem 2e still equal to the family-environment variance? If not, how would using the sample version of your expression lead to a biased estimate?
4. *Family proxies* (20) Suppose we can imperfectly measure family environment.
- (a) (5) Modify the models in Figure 2 to include an observable variable which measures each family environment. Call its path coefficient m .
 - (b) (5) Does conditioning on the shared genes and the measurements of family environments d-separate the twins when they are raised apart? Does it d-separate them when they are raised together? (Show your work.)
 - (c) (10) Conditional on the measurements of family environments, find the variances and covariances of twins raised together and twins raised apart. For this problem only, you may assume all variables are Gaussian, if that helps. *Hint*: Chapter 14.

5. *Not separated before birth* (25) Twins share the same environment for nine months (or so) before birth. This is called the **maternal environment** or **maternal effect**.
- (a) (5) Draw graphical models including the maternal effect for both twins raised together and twins raised apart.
 - (b) (5) Find the variances of the twins' traits in the new models, in terms of the variances of the traits causal parents and the path coefficients. Explain why the path coefficients can all be taken to be 1.
 - (c) (5) Find the covariance between twins raised together and the covariance between twins raised apart in the new models.
 - (d) (5) Is your expression from Problem 2d still equal to the variance of the genes? If not, how would using the sample covariance between twins raised apart lead to a biased estimate of genetic variance?
 - (e) (5) Is your expression from Problem 2e still equal to the variance of the family environment? If not, how would using the sample version of your expression lead to a biased estimate?

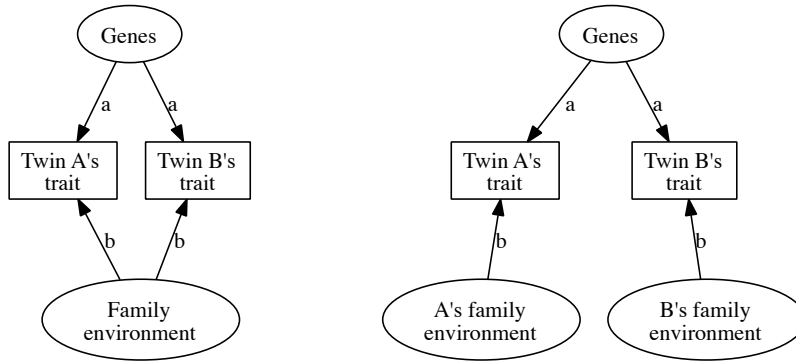


Figure 1: Standard models for twins raised together (left) and twins raised apart (right). Variables in boxes are observable, variables in ellipses are latent. Letters on edges indicate path coefficients.

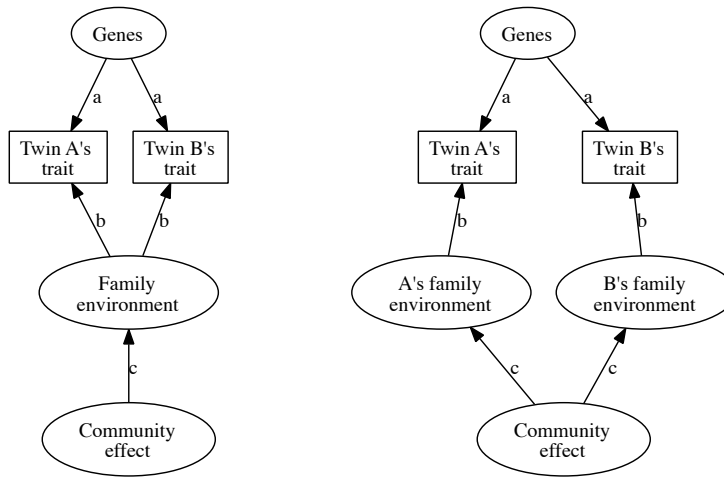


Figure 2: Models for twins raised together and apart, incorporating community effects.