Path Analysis

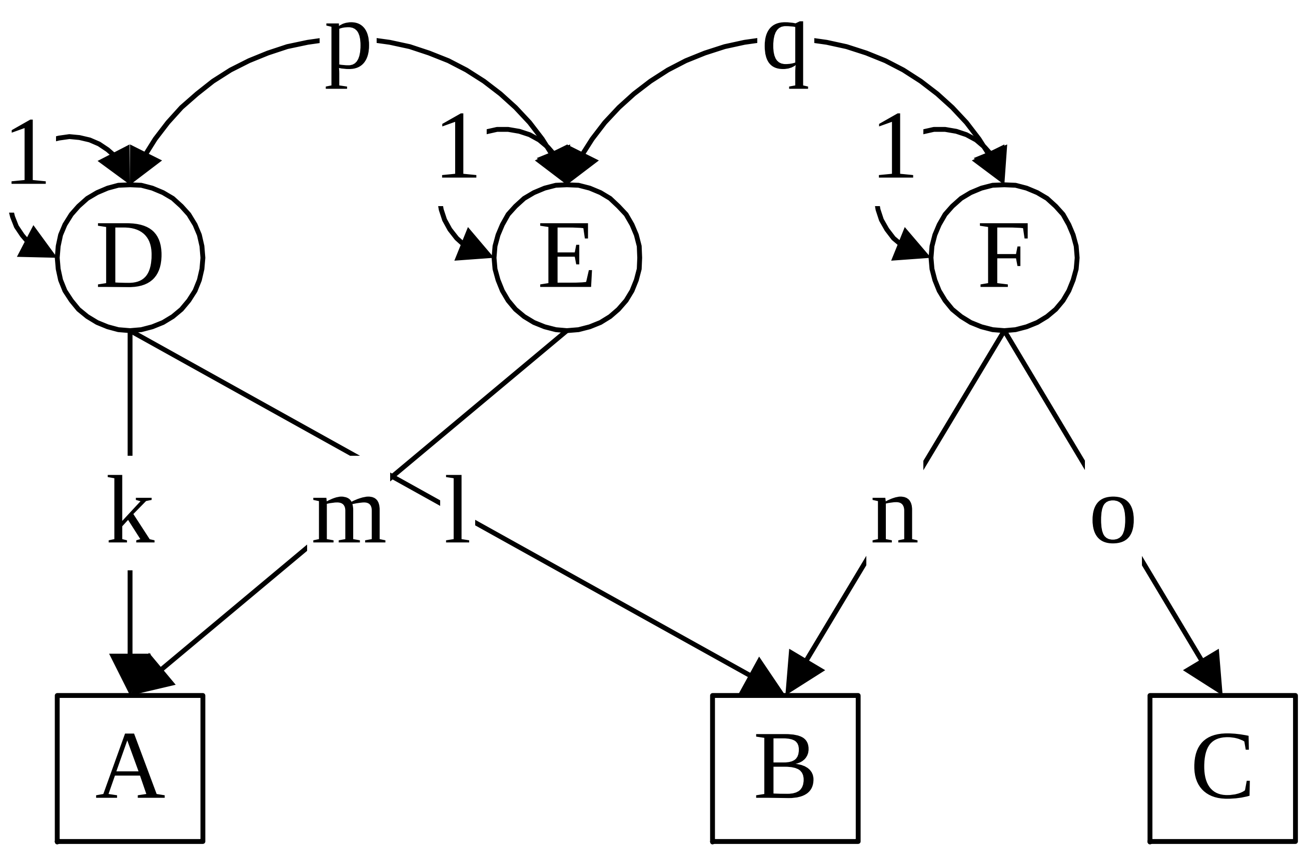
Danielle Dick

Boulder 2006

Path Analysis

- Allows us to represent linear models for the relationships between variables in diagrammatic form
- Makes it easy to derive expectation for the variances and covariances of variables in terms of the parameters proposed by the model
- Is easily translated into matrix form for use in programs such as Mx

Example



Conventions of Path Analysis

- Squares or rectangles denote observed variables.
- Circles or ellipses denote latent (unmeasured) variables.
  - (Triangle denote means, used when modeling raw data)

- Upper-case letters are used to denote variables.
- Lower-case letters (or numeric values) are used to denote covariances or path coefficients.

Conventions of Path Analysis

- Single-headed arrows or paths (–>) are used to represent causal relationships between variables under a particular model - where the variable at the tail is hypothesized to have a direct influence on the variable at the head.

  A –> B

- Double-headed arrows (<–>) are used to represent a covariance between two variables, which may arise through common causes not represented in the model. They may also be used to represent the variance of a variable.

  A <–> B

Conventions of Path Analysis

- Double-headed arrows may not be used for any variable which has one or more single-headed arrows pointing to it - these variables are called endogenous variables. Other variables are exogenous variables.

- Single-headed arrows may be drawn from exogenous to endogenous variables or from endogenous variables to other endogenous variables.

Conventions of Path Analysis

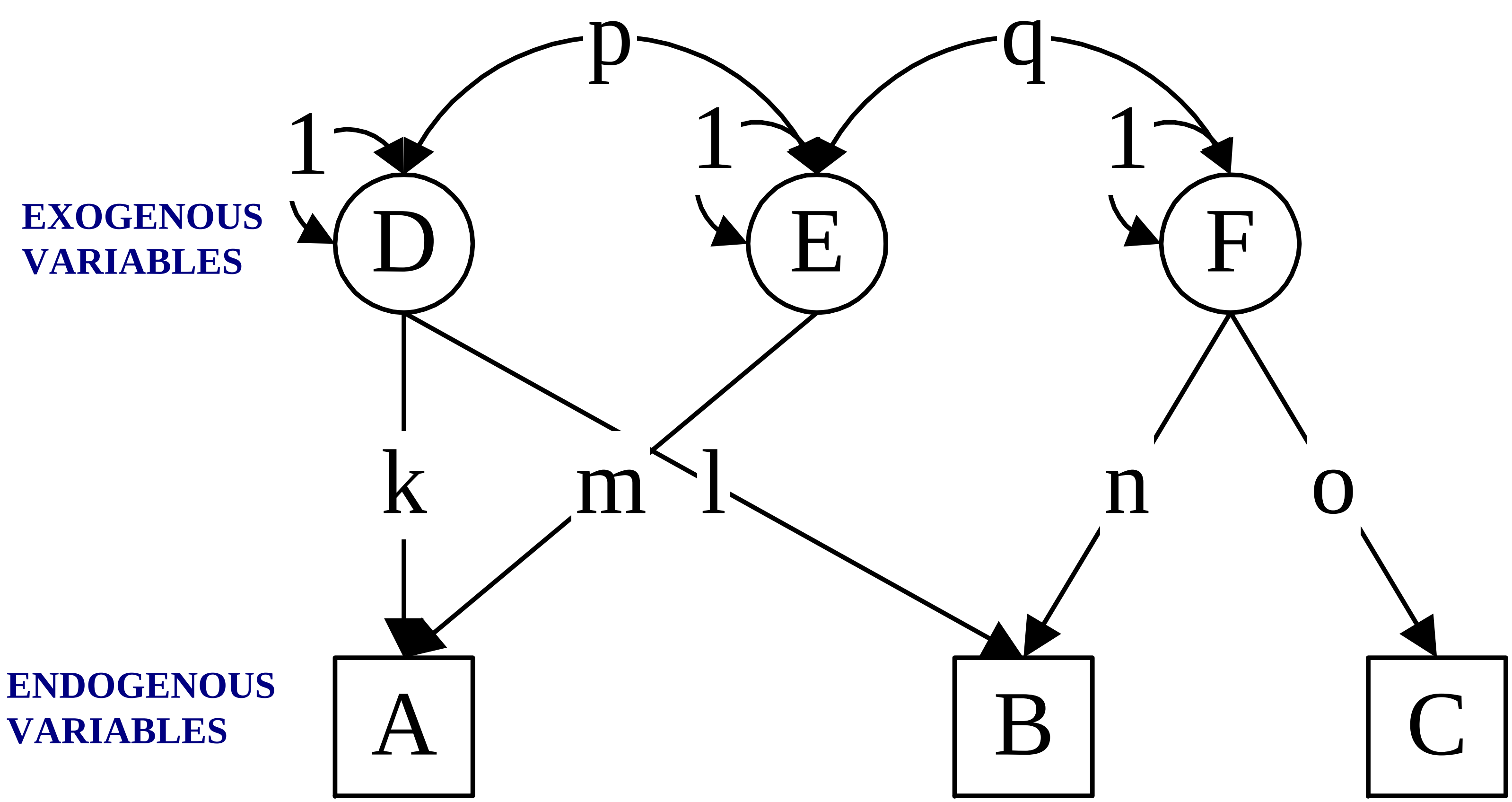
- Omission of a two-headed arrow between two exogenous variables implies the assumption that the covariance of those variables is zero (e.g., no genotype-environment correlation).

- Omission of a direct path from an exogenous (or endogenous) variable to an endogenous variable implies that there is no direct causal effect of the former on the latter variable.

Tracing Rules of Path Analysis

- Trace backwards, change direction at a double-headed arrow, then trace forwards.
  - This implies that we can never trace through double-headed arrows in the same chain.
- The expected covariance between two variables, or the expected variance of a variable, is computed by multiplying together all the coefficients in a chain, and then summing over all possible chains.

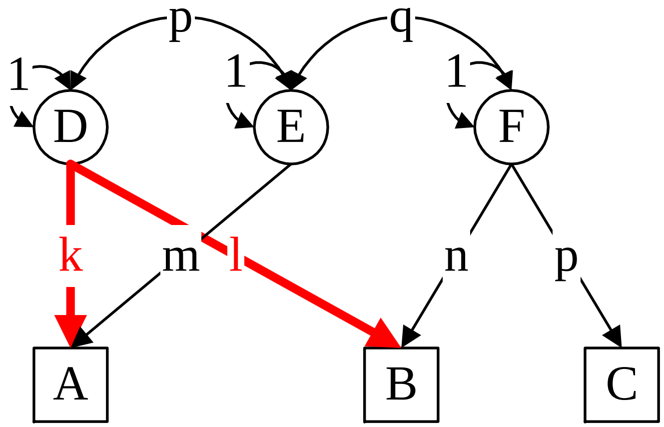
Example

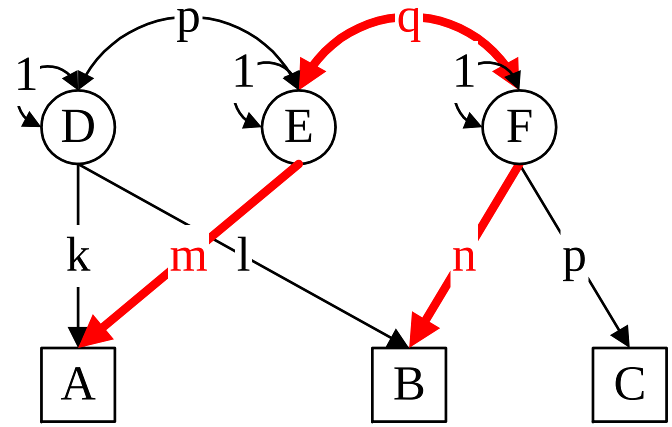

EXOGENOUS
VARIABLES
ENDOGENOUS
VARIABLES
p
q
1
1
1
D
E
F
k
m
l
n
o
A
B
C

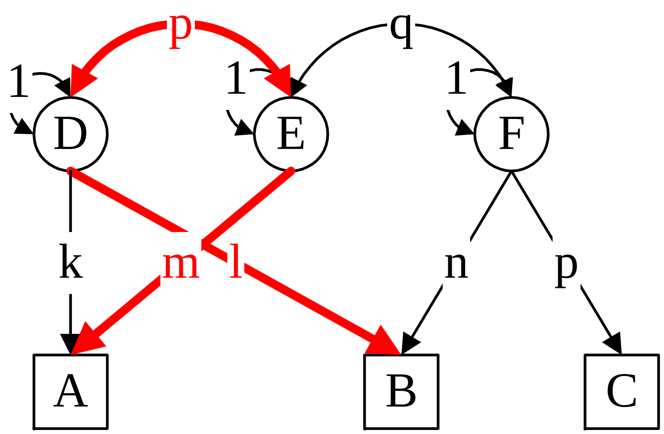
Exercises

- Cov AB =
- Cov BC =
- Cov AC =
- Var A =
- Var B =
- Var C =
- Var E

Covariance between A and B







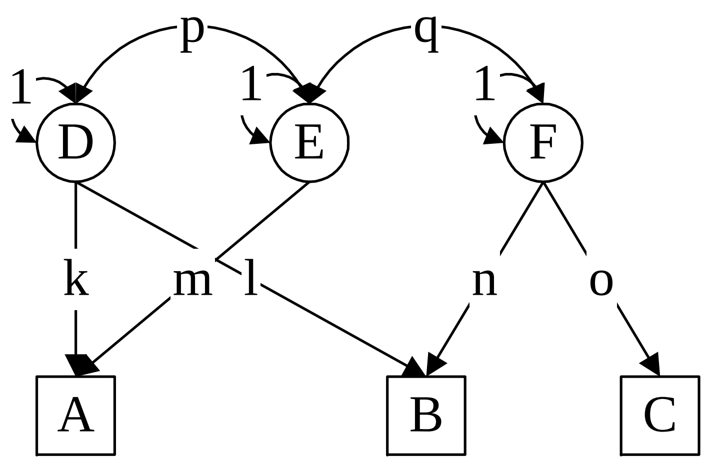
Cov AB = kl + mqn + mpl

Exercises

- Cov AB =
- Cov BC =
- Cov AC =
- Var A =
- Var B =
- Var C =
- Var E

Expectations

- Cov AB = kl + mqn + mpl
- Cov BC = no
- Cov AC = mqo
- Var A = $k^2 + m^2 + 2$ kpm
- Var B = $l^2 + n^2$
- Var C = o^2
- Var E = 1

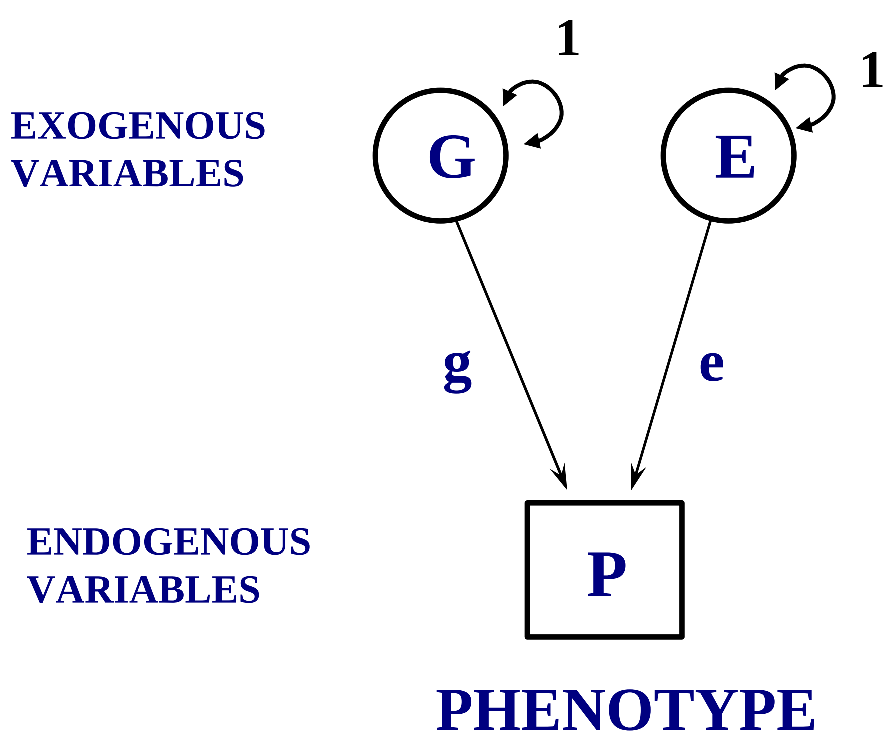


Quantitative Genetic Theory

- Observed behavioral differences stem from two primary sources: ***genetic*** and ***environmental***

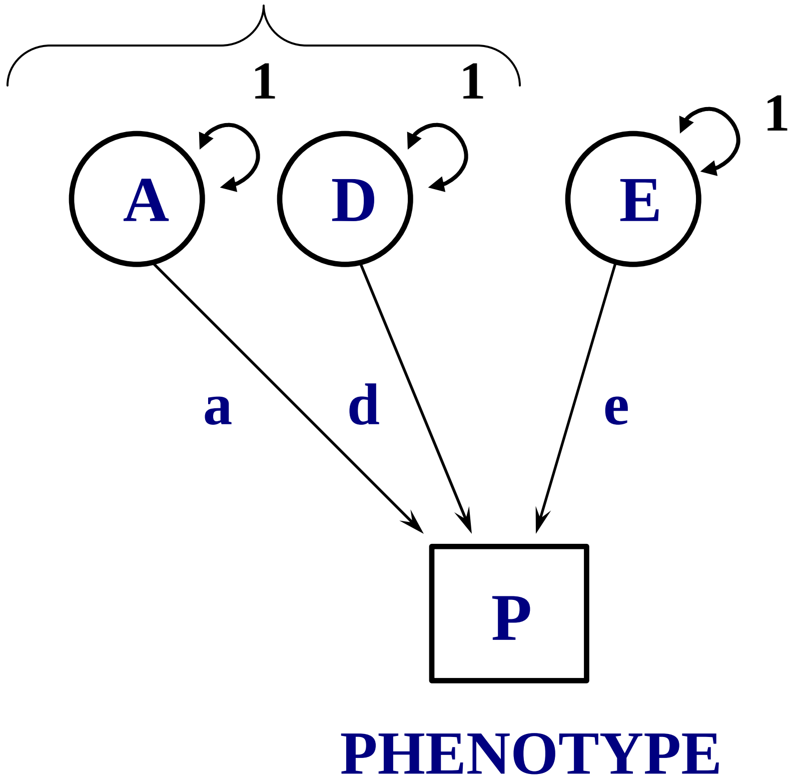
Quantitative Genetic Theory

- Observed behavioral differences stem from two primary sources: ***genetic*** and ***environmental***



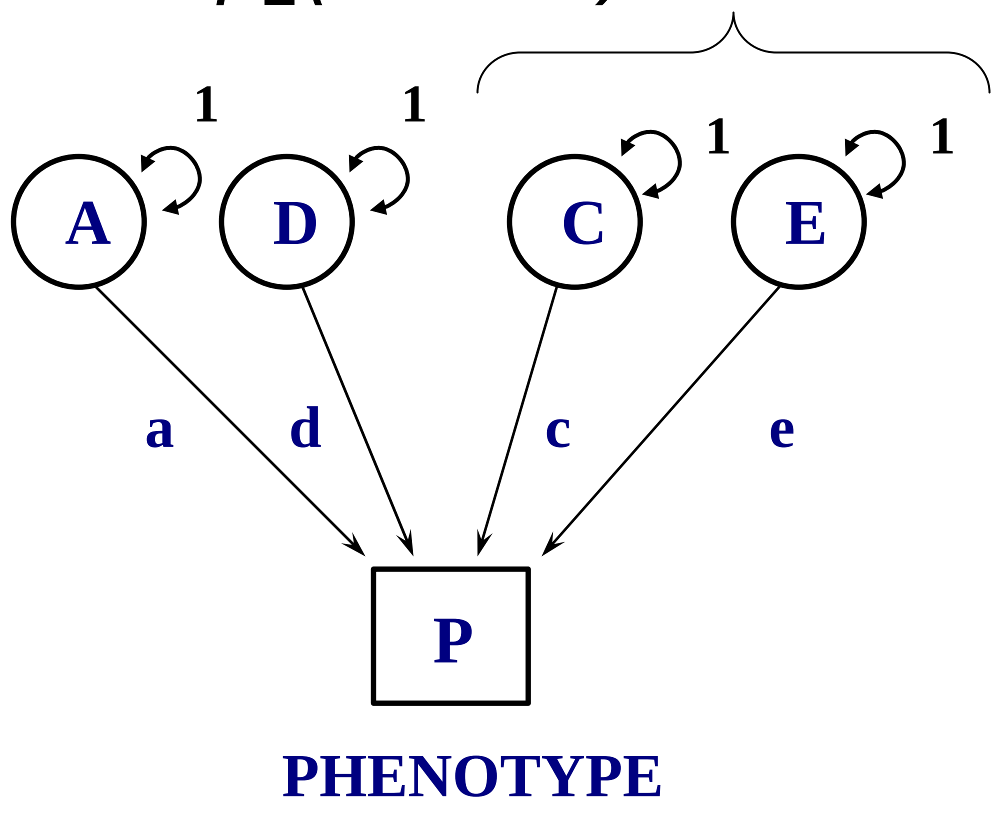
Quantitative Genetic Theory

- There are two sources of genetic influences: ***Additive*** and ***Dominant***



Quantitative Genetic Theory

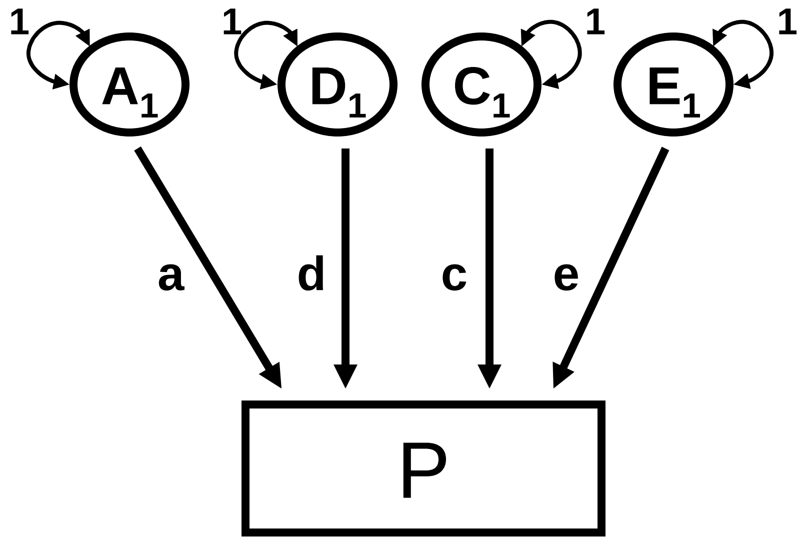
- There are two sources of environmental influences: ***Common (shared)*** and ***Unique (nonshared)***



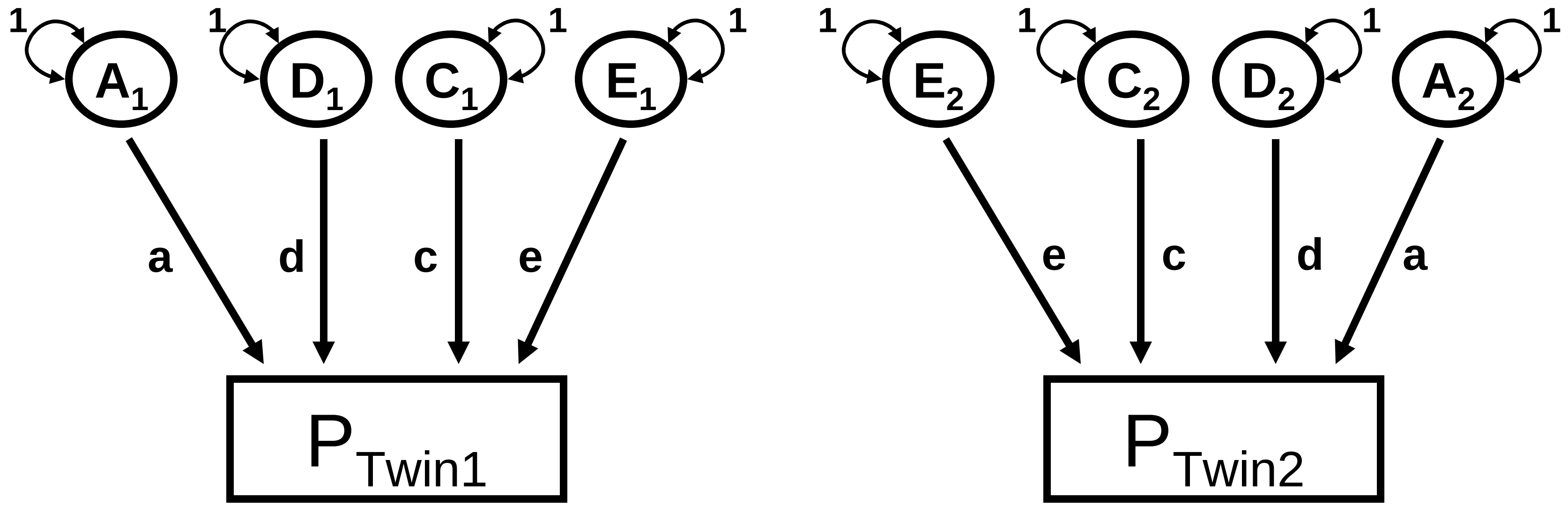
In the preceding diagram…

- A, D, C, E are *exogenous* variables
  - A = Additive genetic influences
  - D = Non-additive genetic influences (i.e., dominance)
  - C = Shared environmental influences
  - E = Nonshared environmental influences
  - A, D, C, E have variances of 1
- Phenotype is an *endogenous* variable
  - P = phenotype; the measured variable
- a, d, c, e are parameter estimates

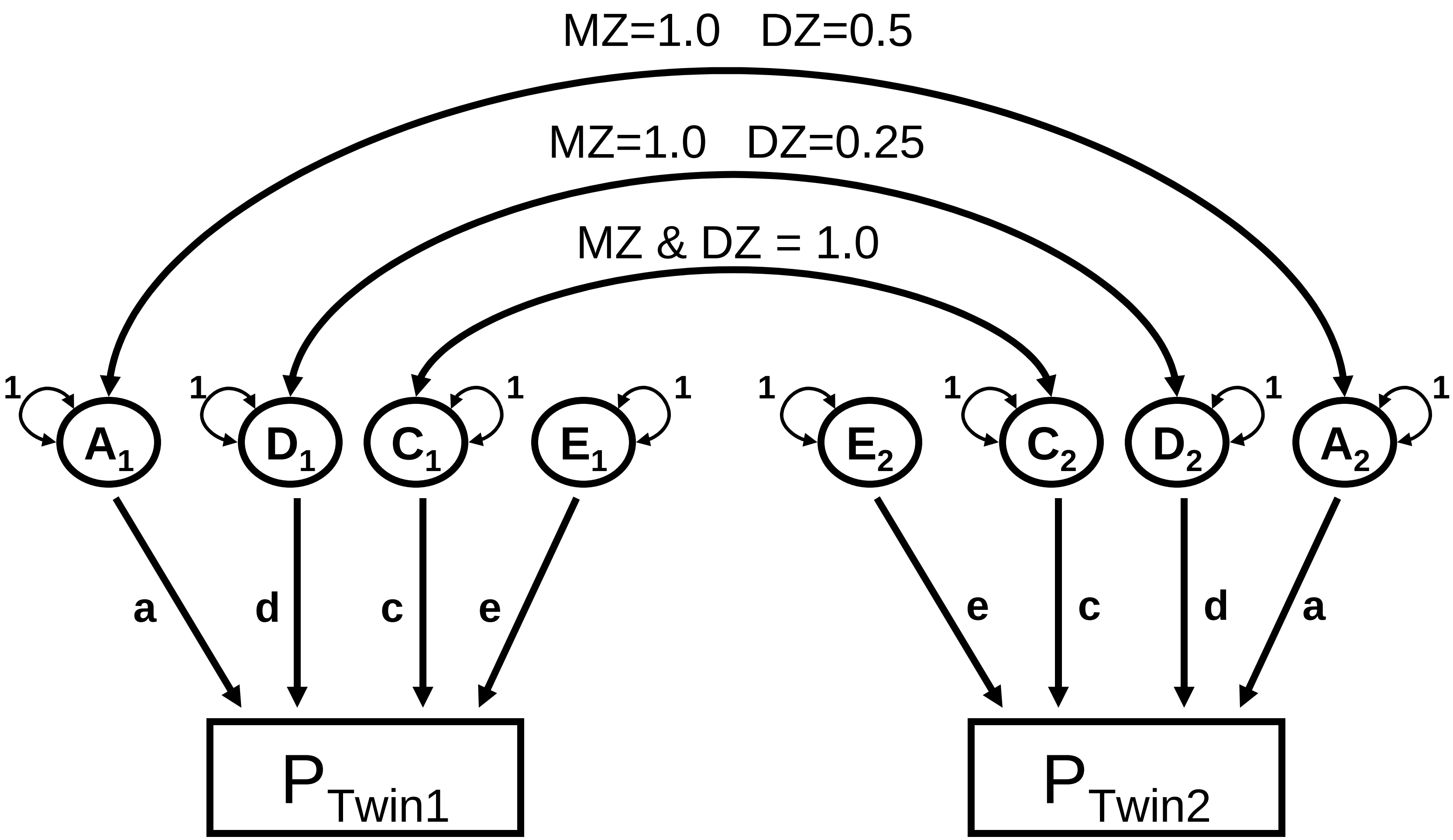
Univariate Twin Path Model



Univariate Twin Path Model



Univariate Twin Path Model



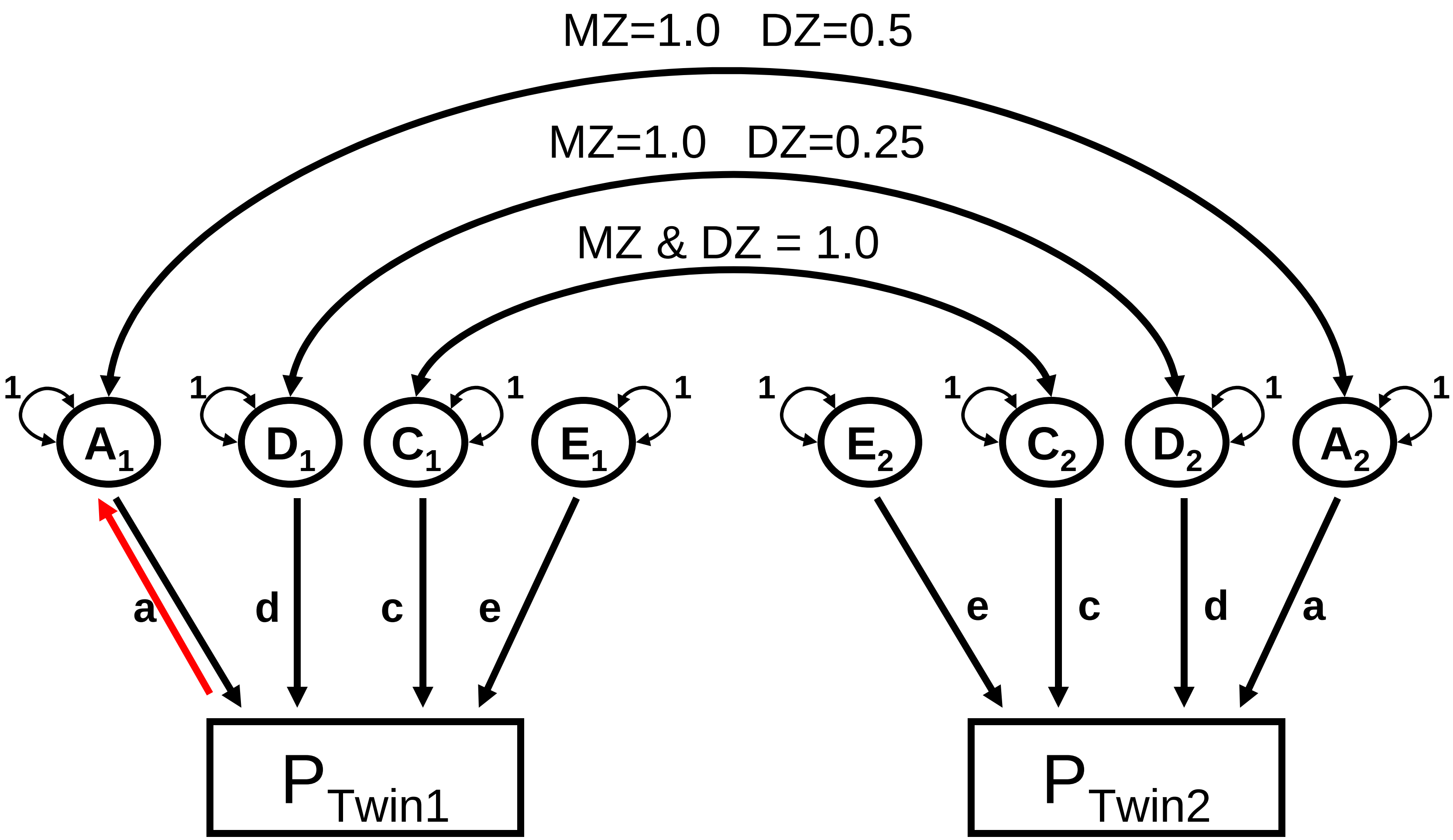
Assumptions of this Model

- All effects are linear and additive (i.e., no genotype x environment or other multiplicative interactions)
- A, D, C, and E are mutually uncorrelated (i.e., there is no genotype-environment covariance/correlation)
- Path coefficients for $Twin_1$ = $Twin_2$
- There are no reciprocal sibling effects (i.e., there are no direct paths between P_1 and P_2

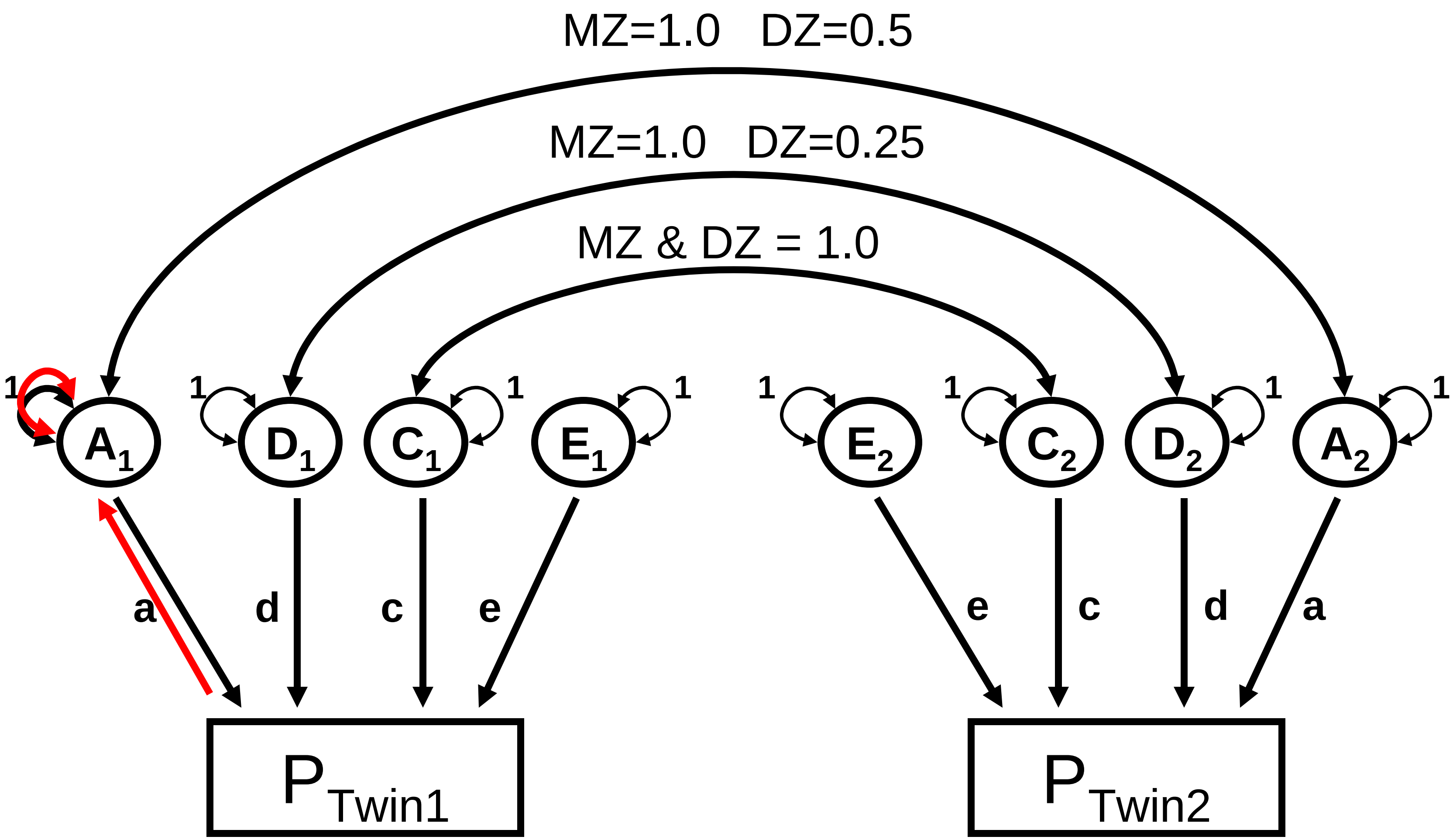
Tracing Rules of Path Analysis

- Trace backwards, change direction at a double-headed arrow, then trace forwards.
  - This implies that we can never trace through double-headed arrows in the same chain.
- The expected covariance between two variables, or the expected variance of a variable, is computed by multiplying together all the coefficients in a chain, and then summing over all possible chains.

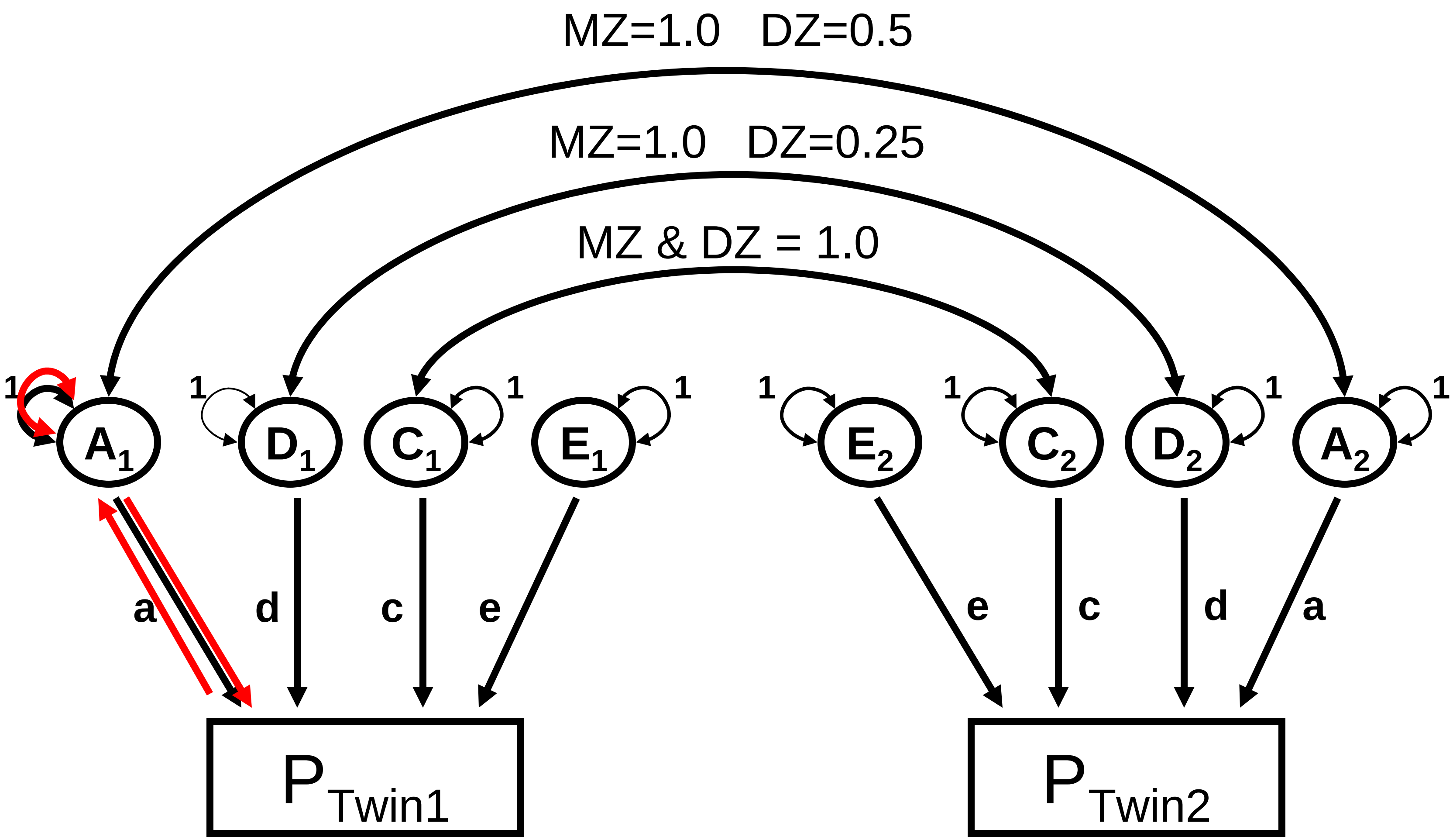
Calculating the Variance of P_1



Calculating the Variance of P_1

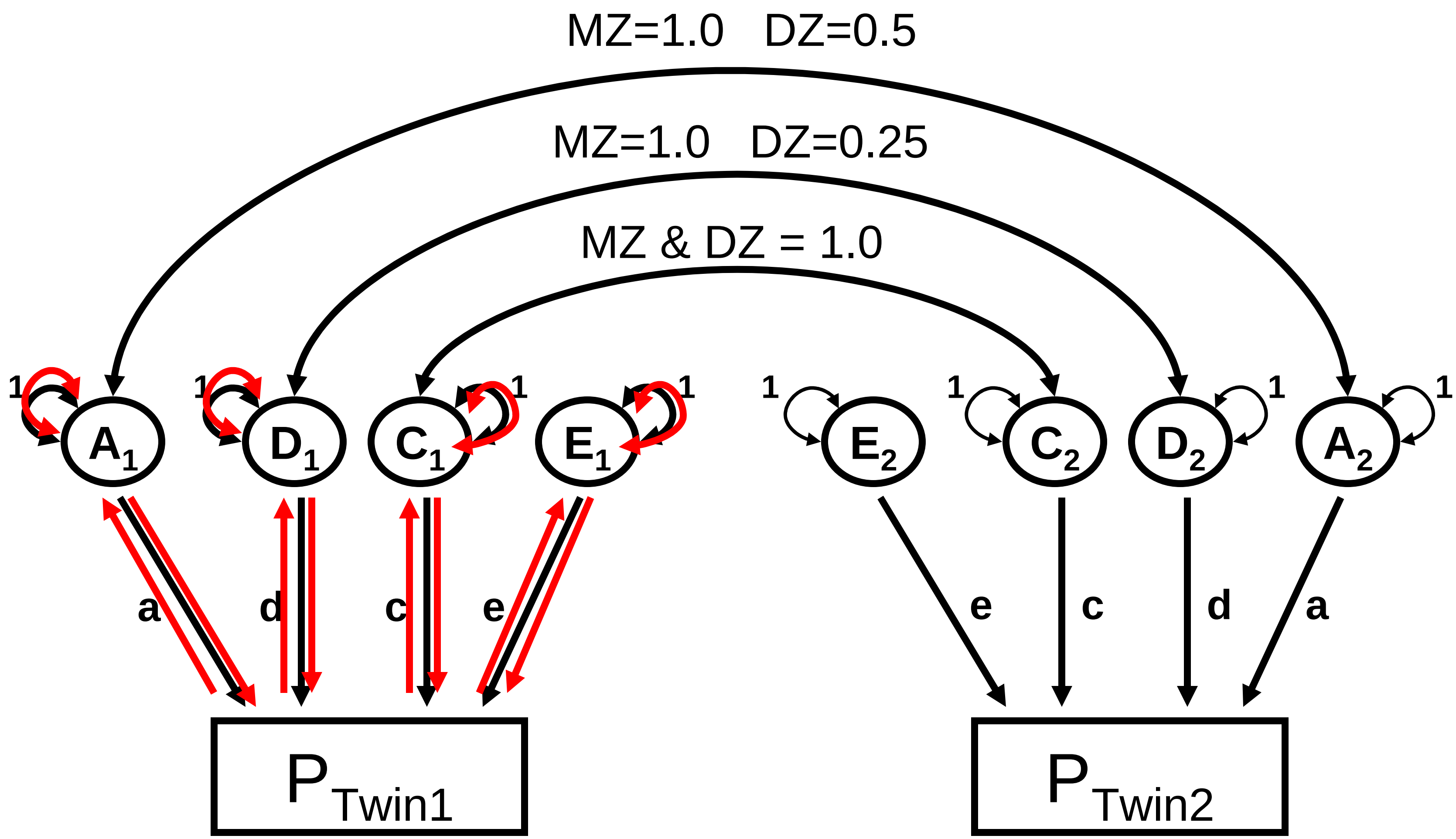


Calculating the Variance of P_1



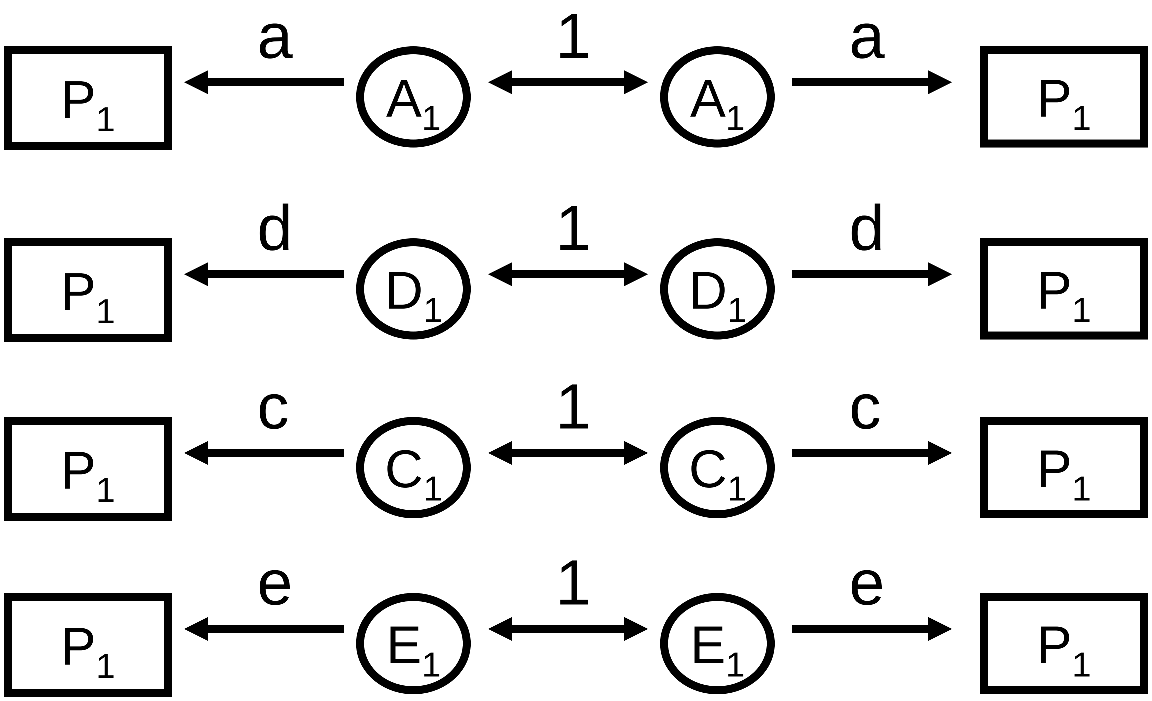
P=a(1)a

Calculating the Variance of P_1



P=a(1)a + d(1)d + c(1)c + e(1)e

Calculating the Variance of P_1



$= 1a^2$

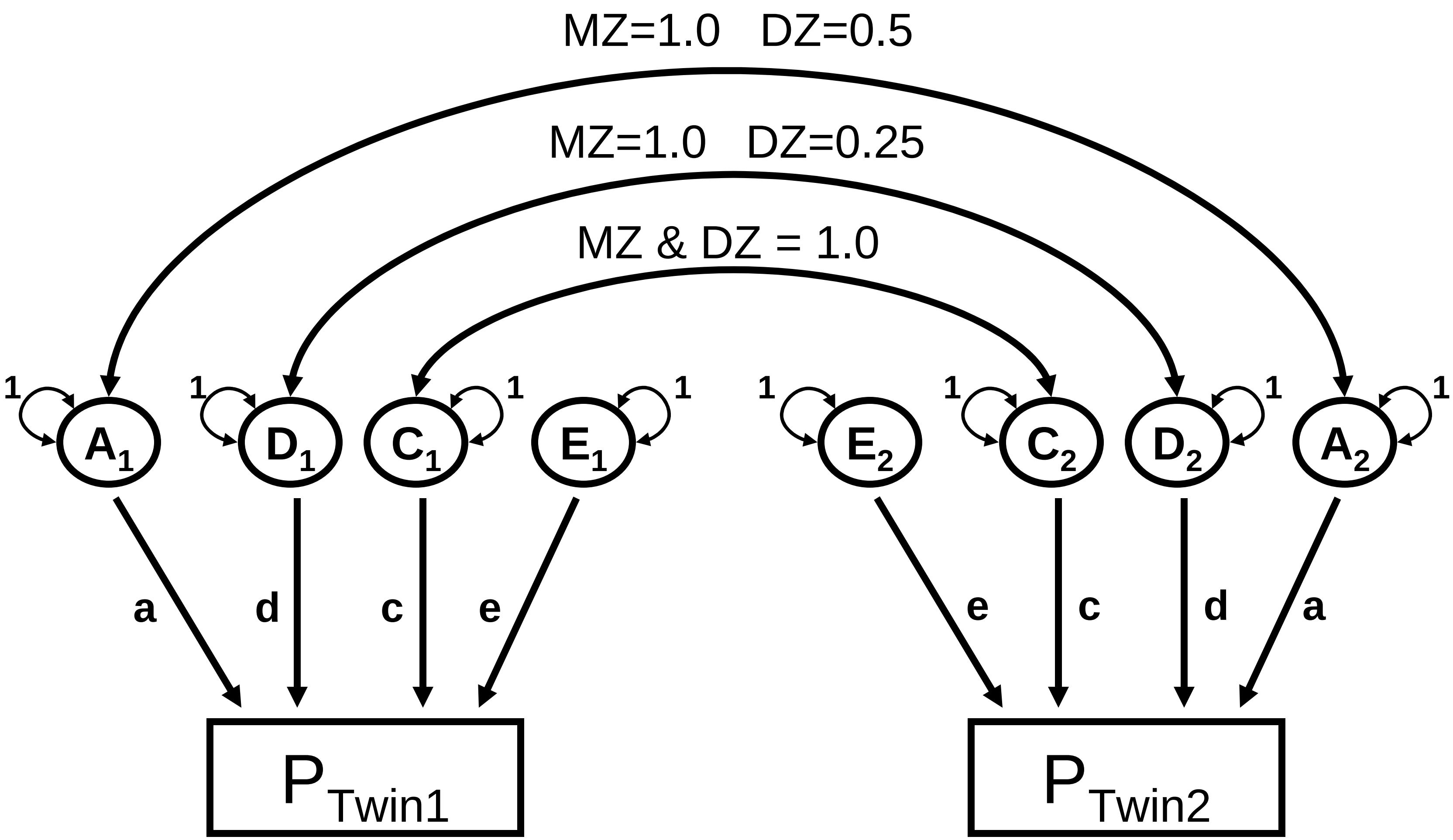
$= 1d^2$

$= 1c^2$

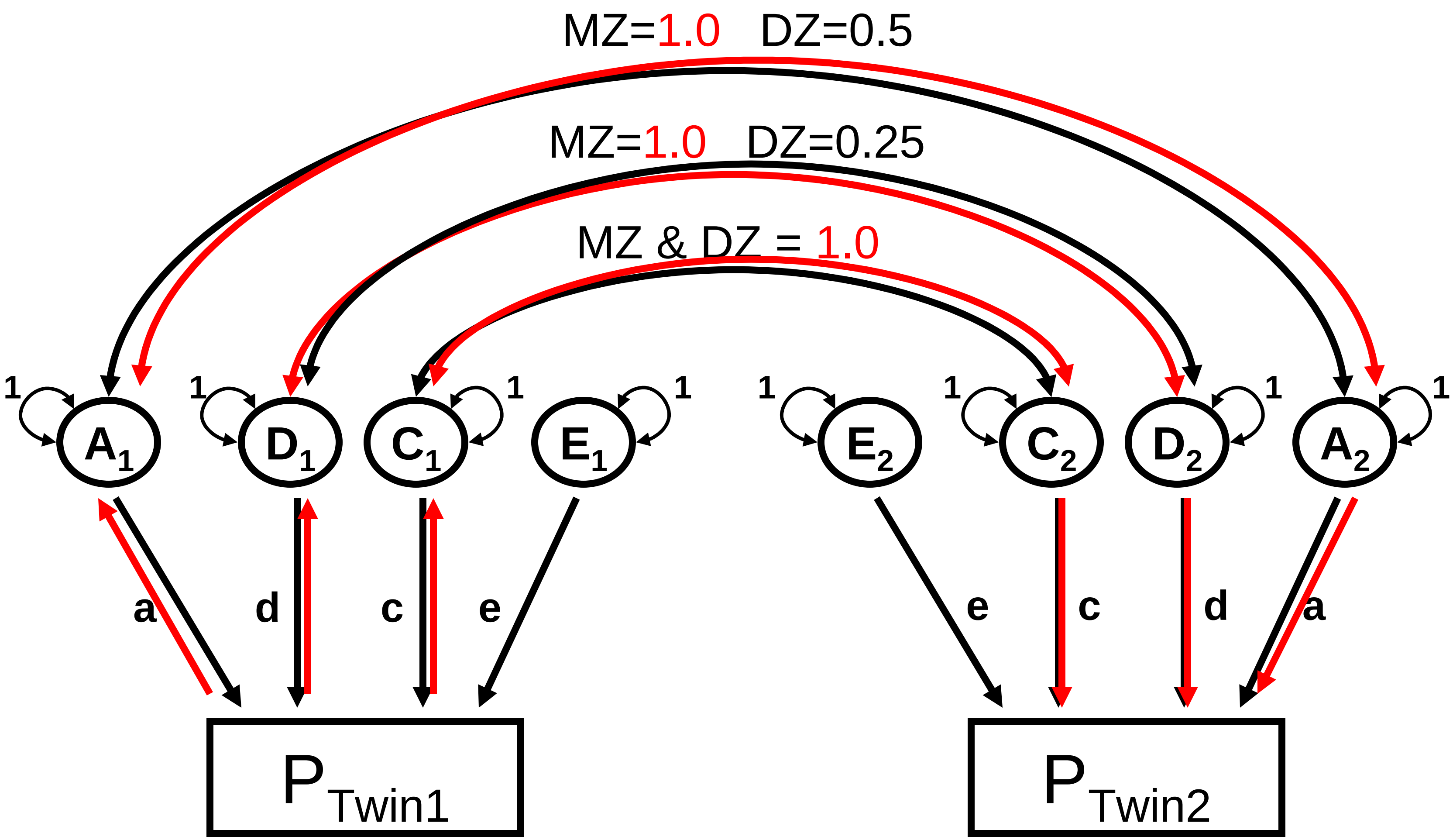
$= 1e^2$

$$\text{Var } P_1 = a^2 + d^2 + c^2 + e^2$$

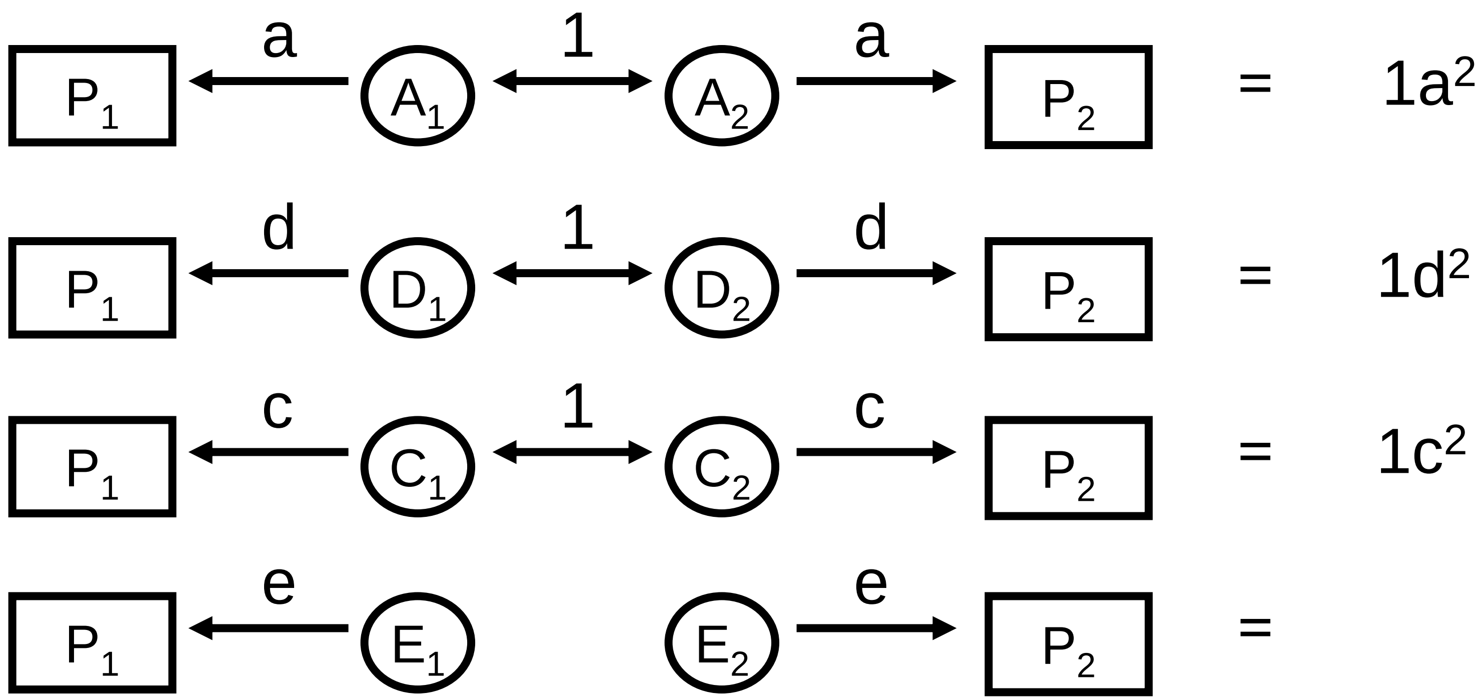
Calculating the MZ Covariance



Calculating the MZ Covariance

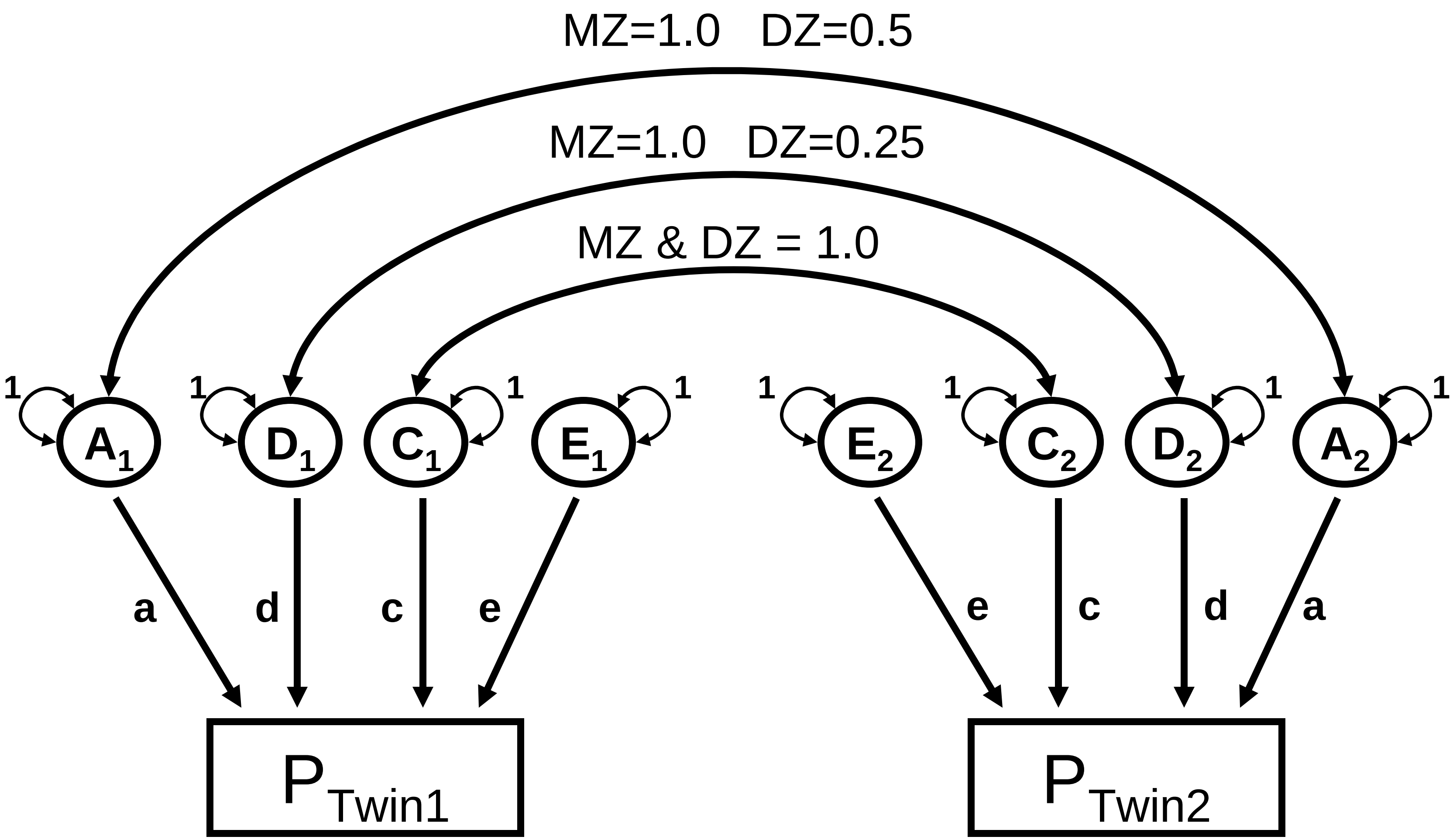


Calculating the MZ Covariance



$$Cov_{MZ} = a^2 + d^2 + c^2$$

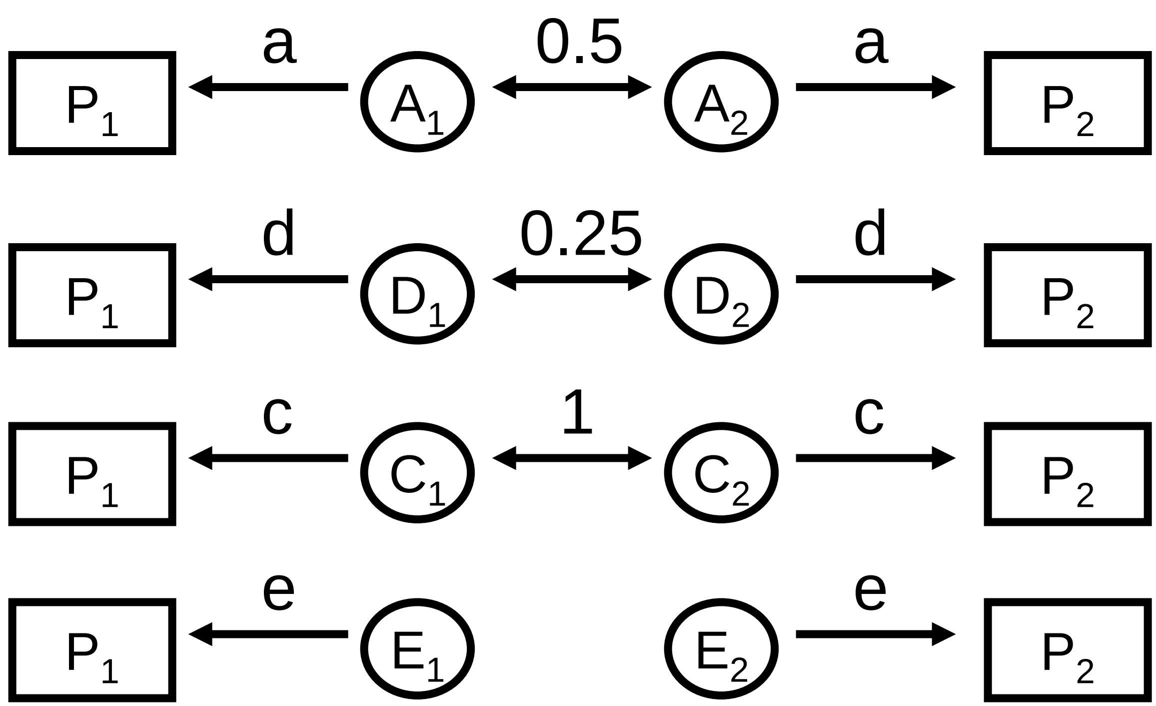
Calculating the DZ Covariance



Calculating the DZ covariance

$Cov_{DZ} = ?$

Calculating the DZ covariance



= $0.5a^2$

= $0.25d^2$

= $1c^2$

=

$$Cov_{DZ} = 0.5a^2 + 0.25d^2 + c^2$$

Twin Variance/Covariance

$$\begin{bmatrix} Var_{Twin1} & Cov_{12} \\ Cov_{21} & Var_{Twin2} \end{bmatrix}$$

$$\mathbf{MZ} = \begin{bmatrix} a^2 + d^2 + c^2 + e^2 & a^2 + d^2 + c^2 \\ a^2 + d^2 + c^2 & a^2 + d^2 + c^2 + e^2 \end{bmatrix}$$

$$\mathbf{DZ} = \begin{bmatrix} a^2 + d^2 + c^2 + e^2 & 0.5a^2 + 0.25d^2 + c^2 \\ 0.5a^2 + 0.25d^2 + c^2 & a^2 + d^2 + c^2 + e^2 \end{bmatrix}$$