Path Analysis and its Application in Models using Twin Data

Elizabeth Prom-Wormley[1]

Special Thanks to

Frühling Rijsdijk[2], Michael Neale[1], and Hermine Maes[1]

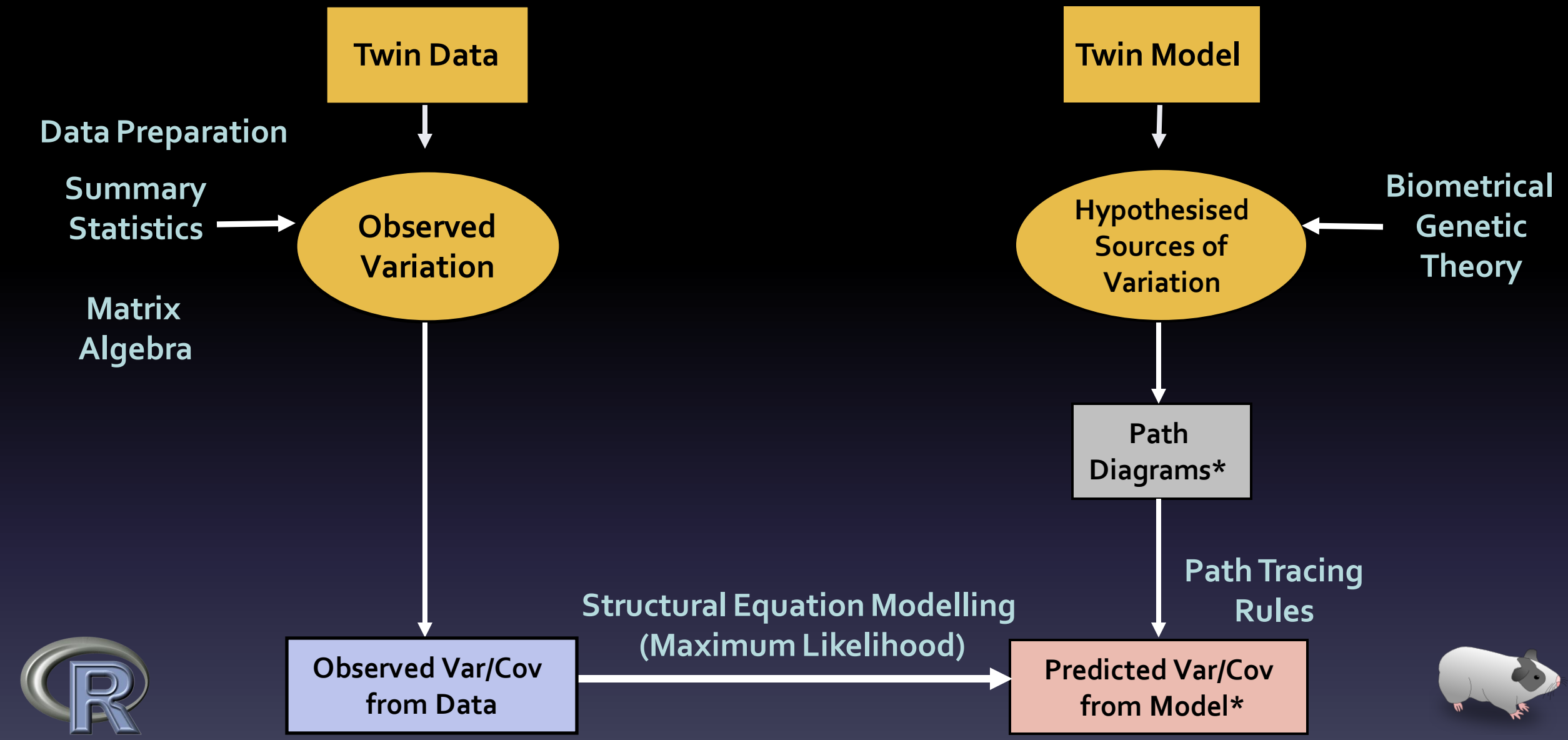
1- Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, USA

2- SGDP Centre Institute of Psychiatry, King's College London, UK

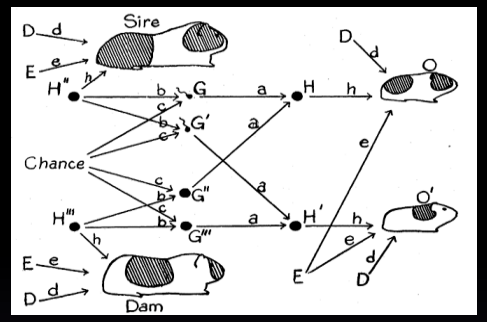
Session Objectives

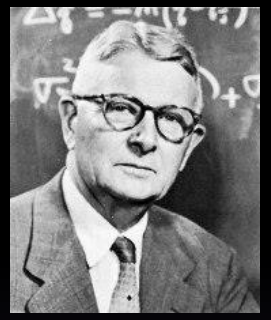
By the end of this session, you will be able to:

- Identify the advantages of applying path tracing rules and their use in structural equation models
- Summarize the basic path tracing rules
- Apply basic path tracing rules to derive the expected parameters generated from simple regression models of unrelated individuals as well as basic twin models


Twin Data
Twin Model
Data Preparation
Summary Statistics
Observed Variation
Hypothesised Sources of Variation
Biometrical Genetic Theory
Matrix Algebra
Path Diagrams*
Path Tracing Rules
Structural Equation Modelling (Maximum Likelihood)
Observed Var/Cov from Data
Predicted Var/Cov from Model*

A Brief Background on Path Analysis and SEM





- Developed around 1918 by Sewall Wright (see last slide for references)
- Guinea pigs: interrelationships of factors determining weight at birth and at weaning

- Path analysis combines our knowledge on causal relations with degree of observed correlations
- Path analysis and equation modelling (SEM) have one-to-one mathematical equivalence with simple matrix algebra expression
- SEM is a unified platform for path analysis, regression, factor and variance components models

Path Diagram Conventions

Observed Variable (Manifest)- Directly measured

Latent Variable- Not directly measured

Constant Variable (Intercept)

Causal Path

Covariance Path

Tracing Rules of Path Analysis

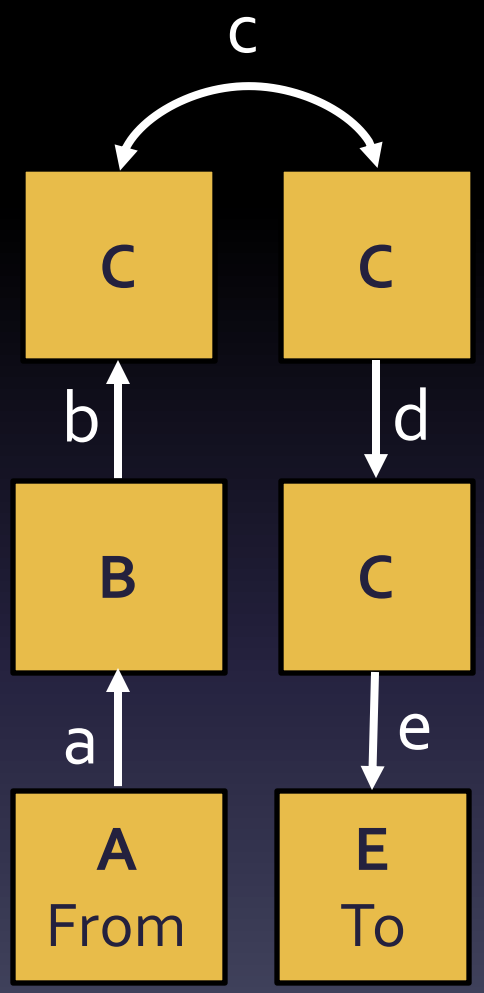
The correlation (covariance) between any two variables can be expressed as the sum of the compound paths connecting them.

1) To obtain a compound path:
   a) Go backwards along single-headed arrows
   b) Change direction at one and only one double-headed arrow
   c) Trace forwards along single-headed arrows
   d) DO NOT go forward then backwards between two single headed arrows

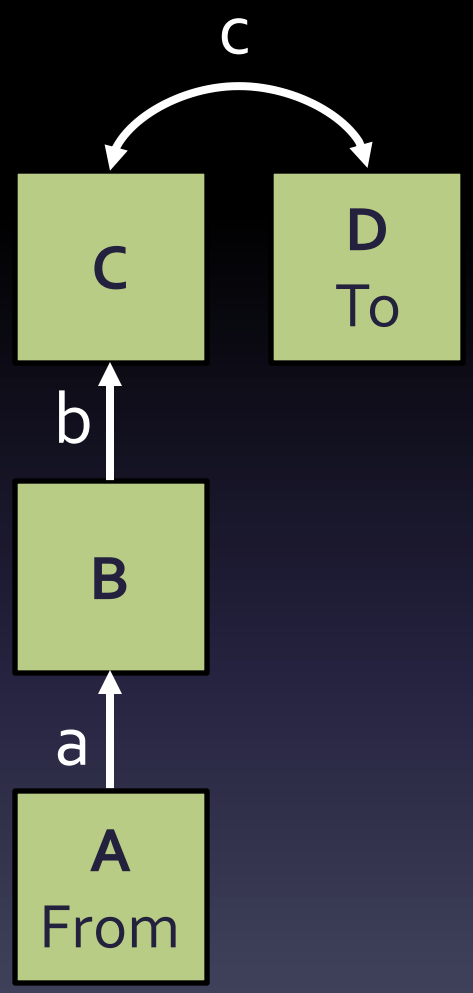
2) To obtain covariance:
   a) Multiply path coefficients in a compound path
   b) Sum all the distinct compound paths, where paths are considered distinct if they contain different coefficients, or encounter those coefficients in a different order.

3) For covariance of a variable with itself (Variance), compound paths are distinct if they have different paths or a different order

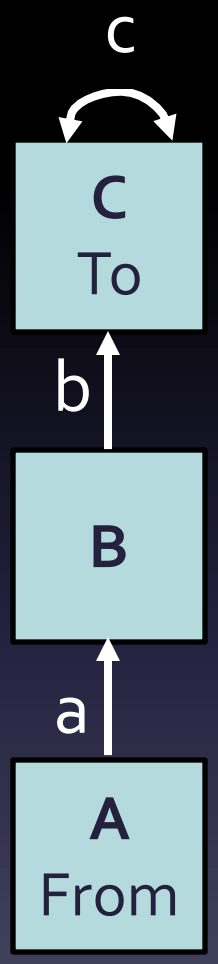
Path Analysis- A Little Practice



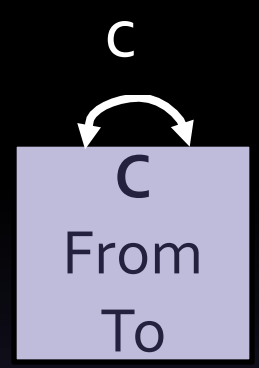
a*b*c*d*e



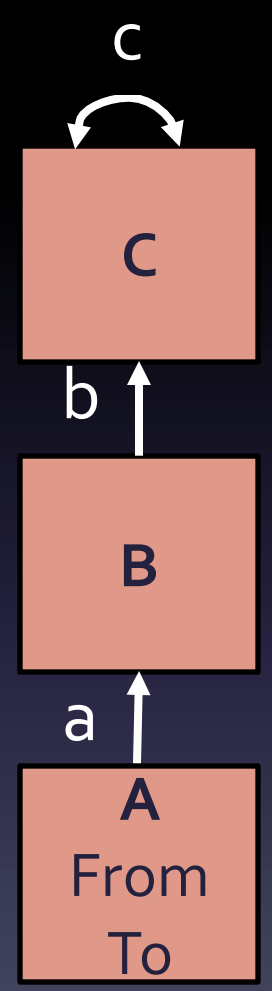
a*b*c



a*b*c

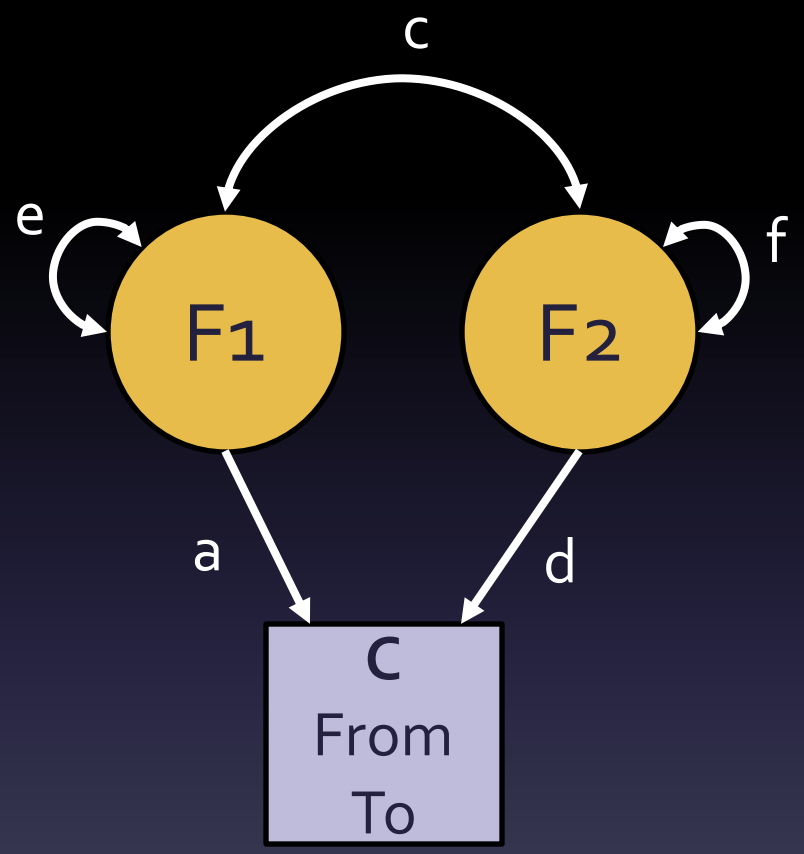


c



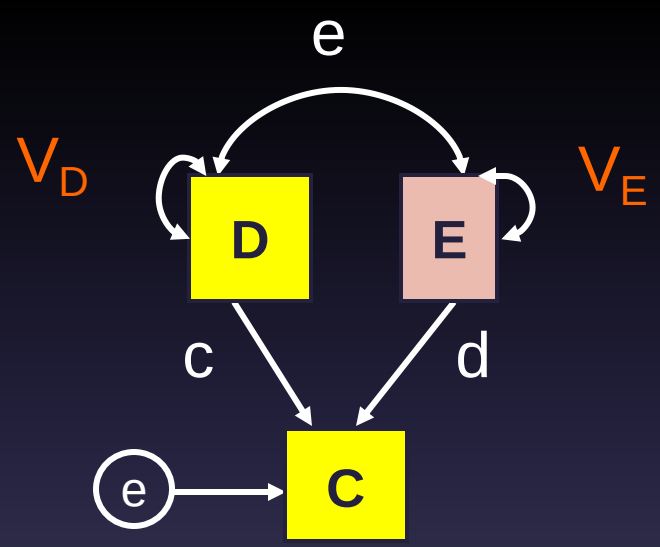
a*b*c*b*a

A Little More Practice- Variance



a*e*a + d*f*d + a*c*d + d*c*a

A Little More Practice
How Would We Calculate the Covariance between C and D?

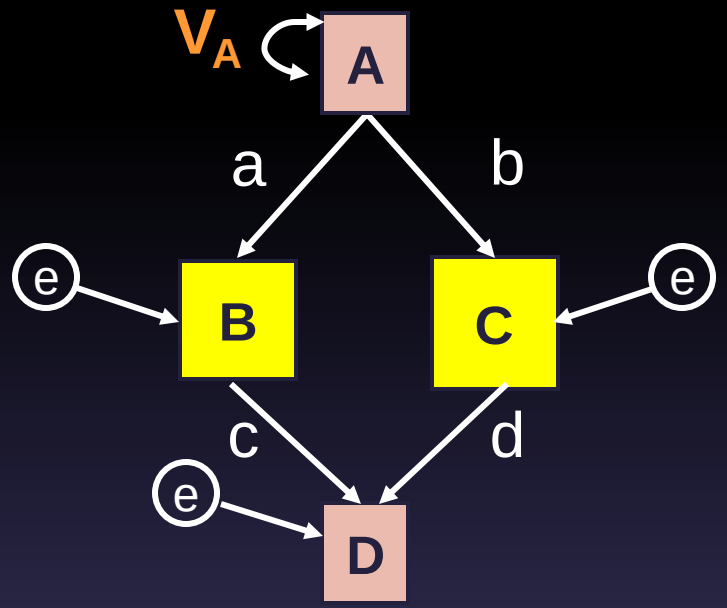


$Cov_{CD} : c*V_D + d*e$

NOT $d*V_E*e$

A maximum of one curved arrow per path. The double-headed arrow from the independent variable to itself is included, unless the chain includes another double-headed arrow (e.g., a correlation path)

How Would We Calculate the Covariance between B and C?



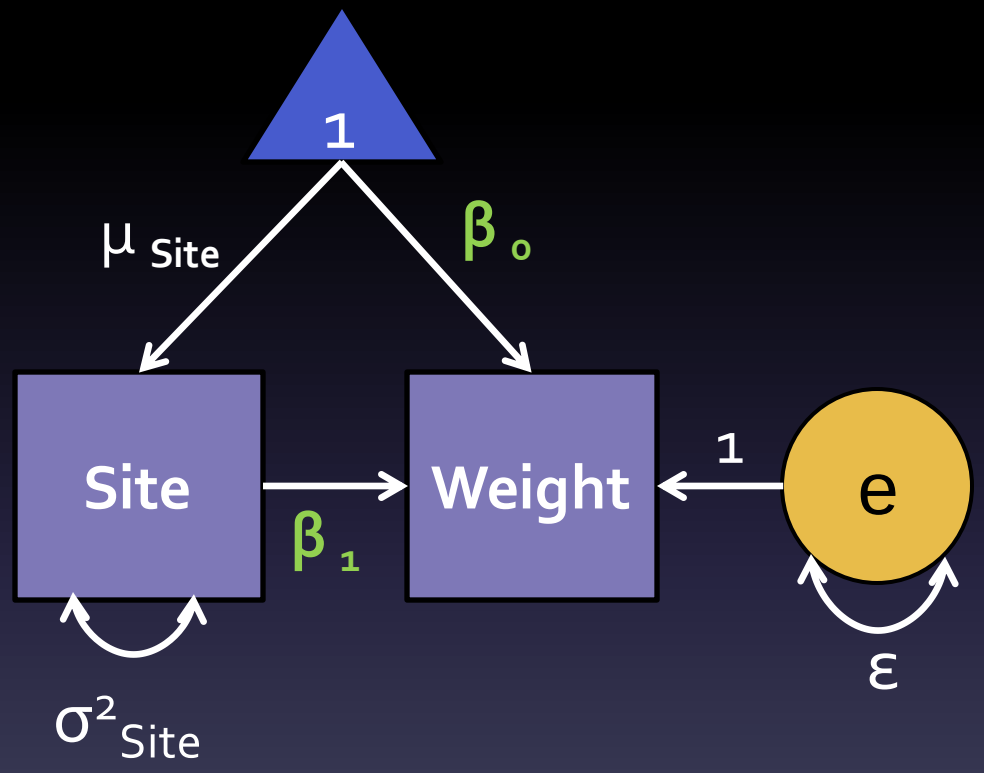
(i) Trace backward, then forward, or simply forward from one variable to another. NEVER forward then backward. Include double-headed arrows from the independent variables to itself.
These variances will be 1 for standardized variables

Cov_{BC} : $a*V_A*b$

NOT

c*d

Linear Regression as a Path Diagram



Weight = $\beta_0 + \beta_1$Site + ε

Squares or rectangular boxes: observed (manifest) variables of Weight and Site

Single-headed arrows: linear relationship between two variables. Starts from an independent variable and ends on a dependent variable. Site is influencing Weight

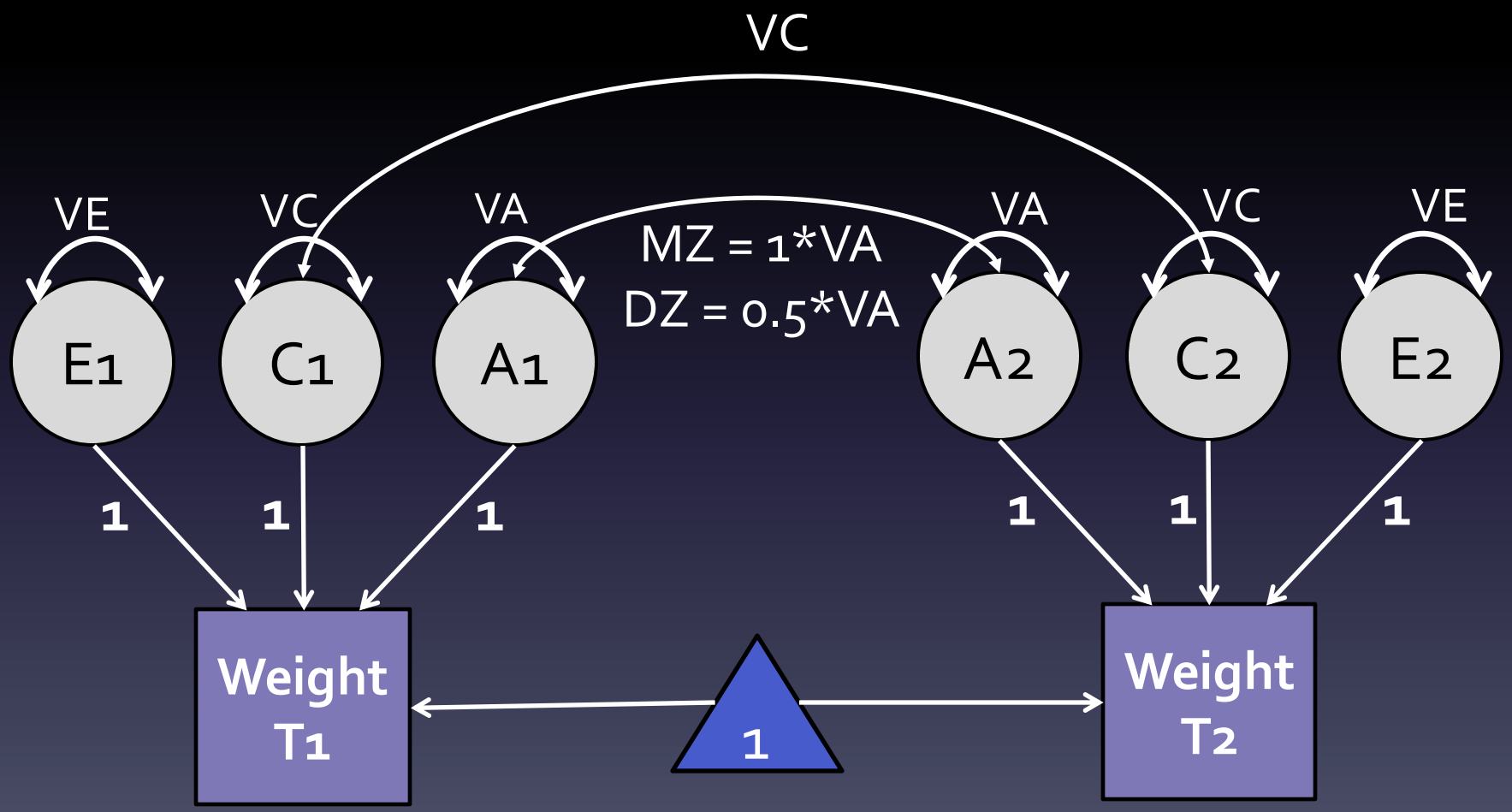
Double-headed arrows: variance of a variable or covariance between 2 variables. Variance of Site and and Weight.

Triangle: a constant variable, usually a vector of ones

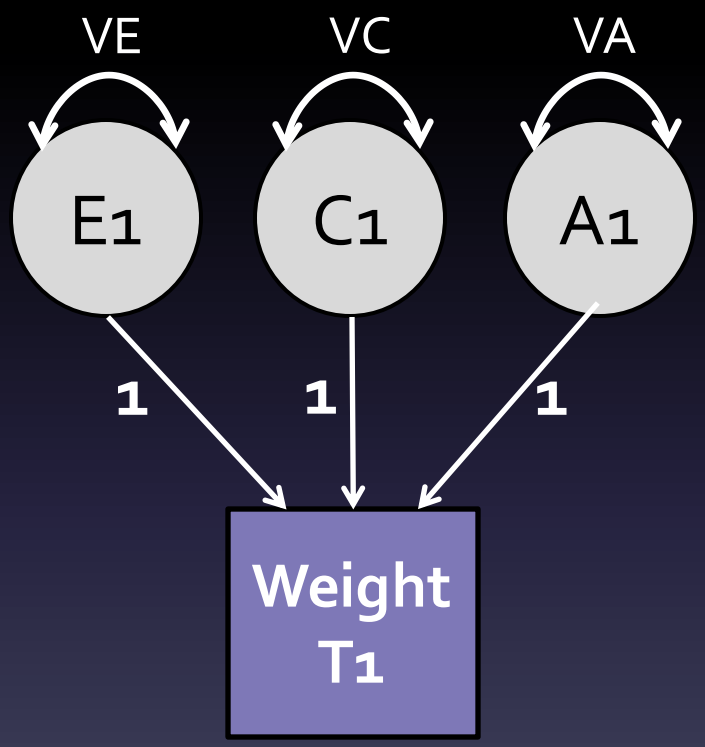
Circles or ovals: error terms, factors, latent variables

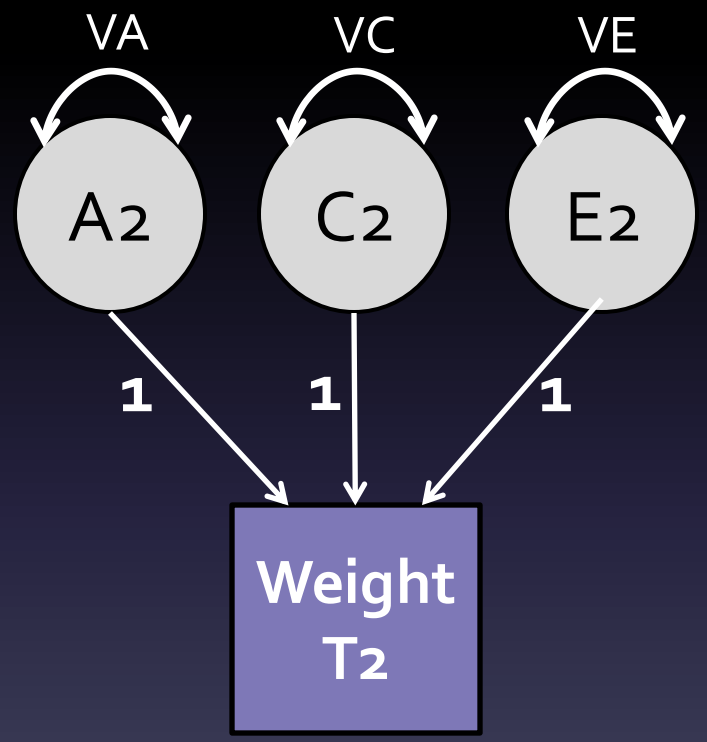
Applying Path Analysis to the Twin Study Design

Path Diagram Representation of a Classical ACE Model Using Twin Data

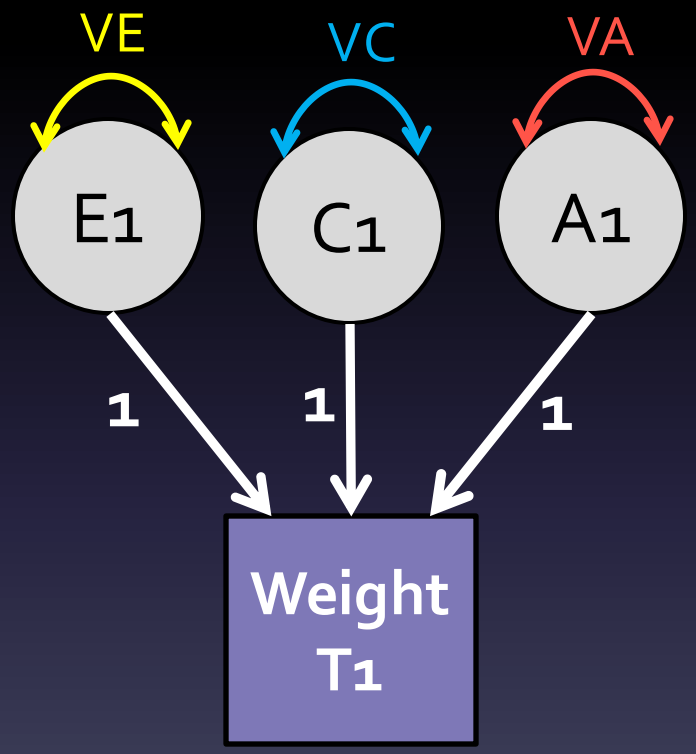


Individual Members of a Twin Pair
How Do We Calculate the Variance of Weight?





Individual Members of a Twin Pair
How Do We Calculate the Variance of Weight (T1/T2)?



1*VA*1 = VA

+

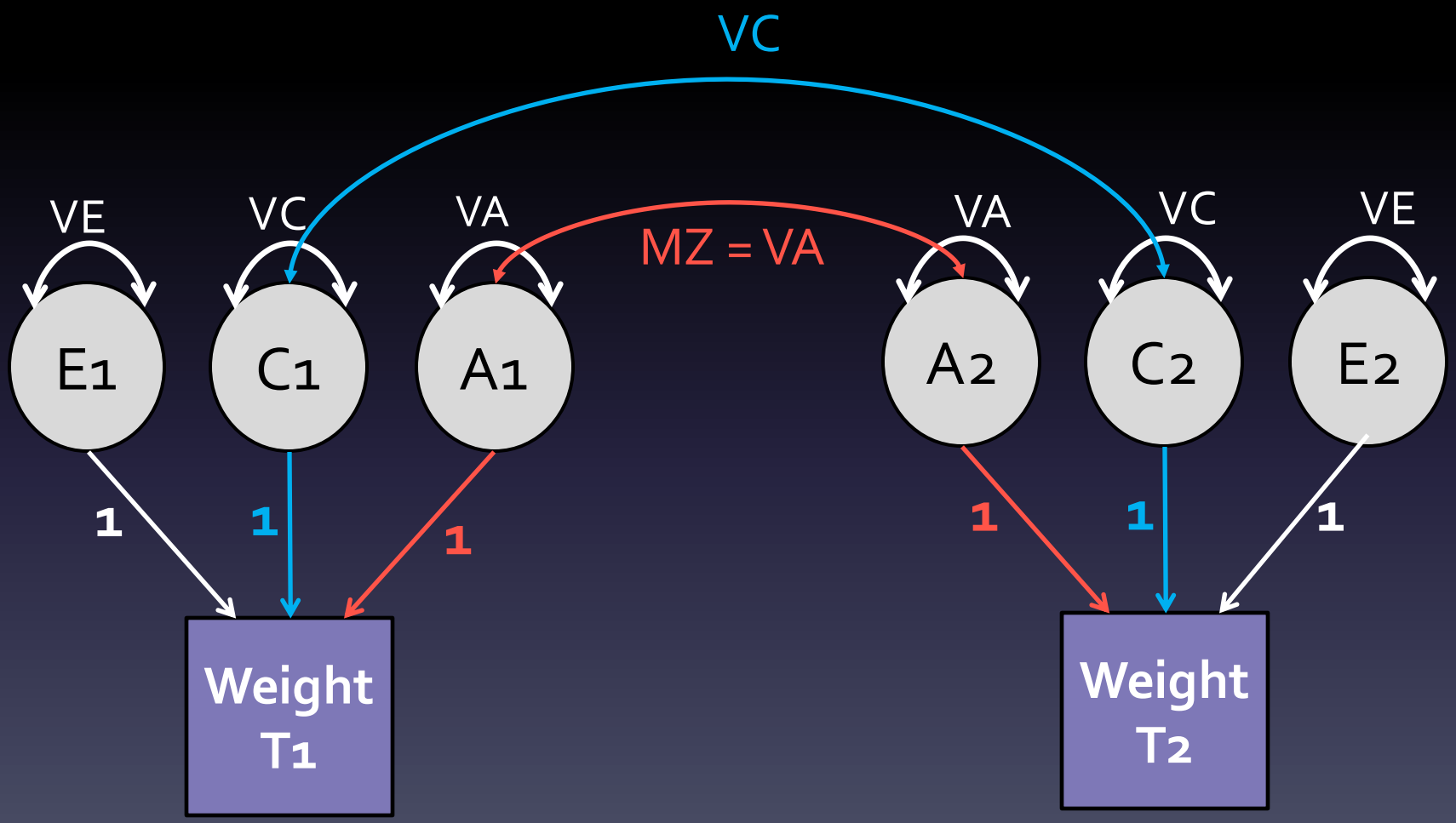
1*VC*1 = VC

+

1*VE*1 = VE

Total Variance = VA + VC + VE

How Do We Represent the Covariances of Weight between T1 and T2 for MZ Pairs?



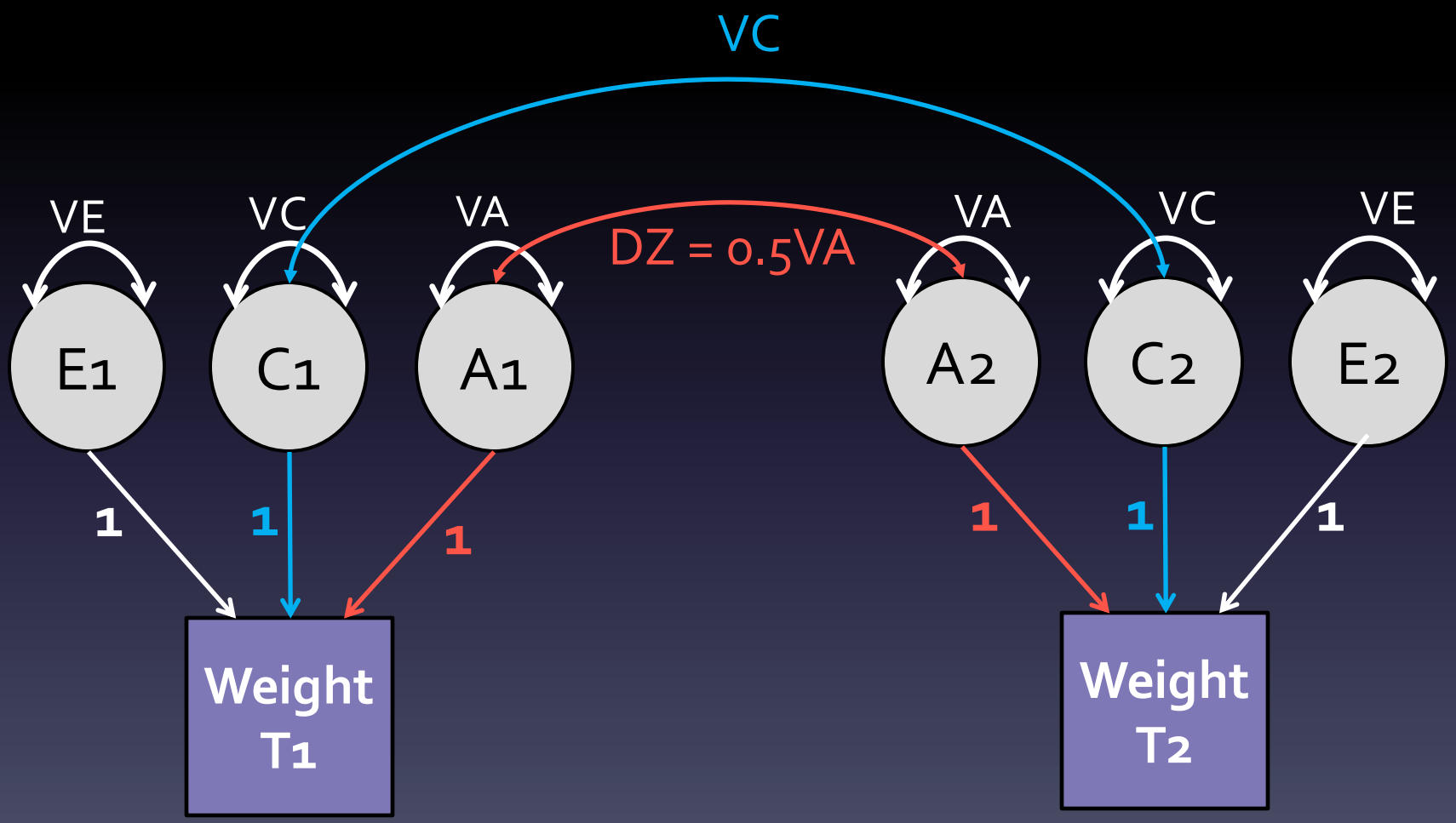
$1 * VA * 1 = VA$

+

$1 * VC * 1 = VC$

$MZ_{cov} = VA + VC$

How Do We Represent the Covariances of Weight between T1 and T2 for DZ Pairs?



$1 * 0.5VA * 1 = 0.5VA$

+

$1 * VC * 1 = VC$

$DZ_{cov} = 0.5VA + VC$

Putting it All Together: Predicted Variance and Covariance Matrices

MZ Twins

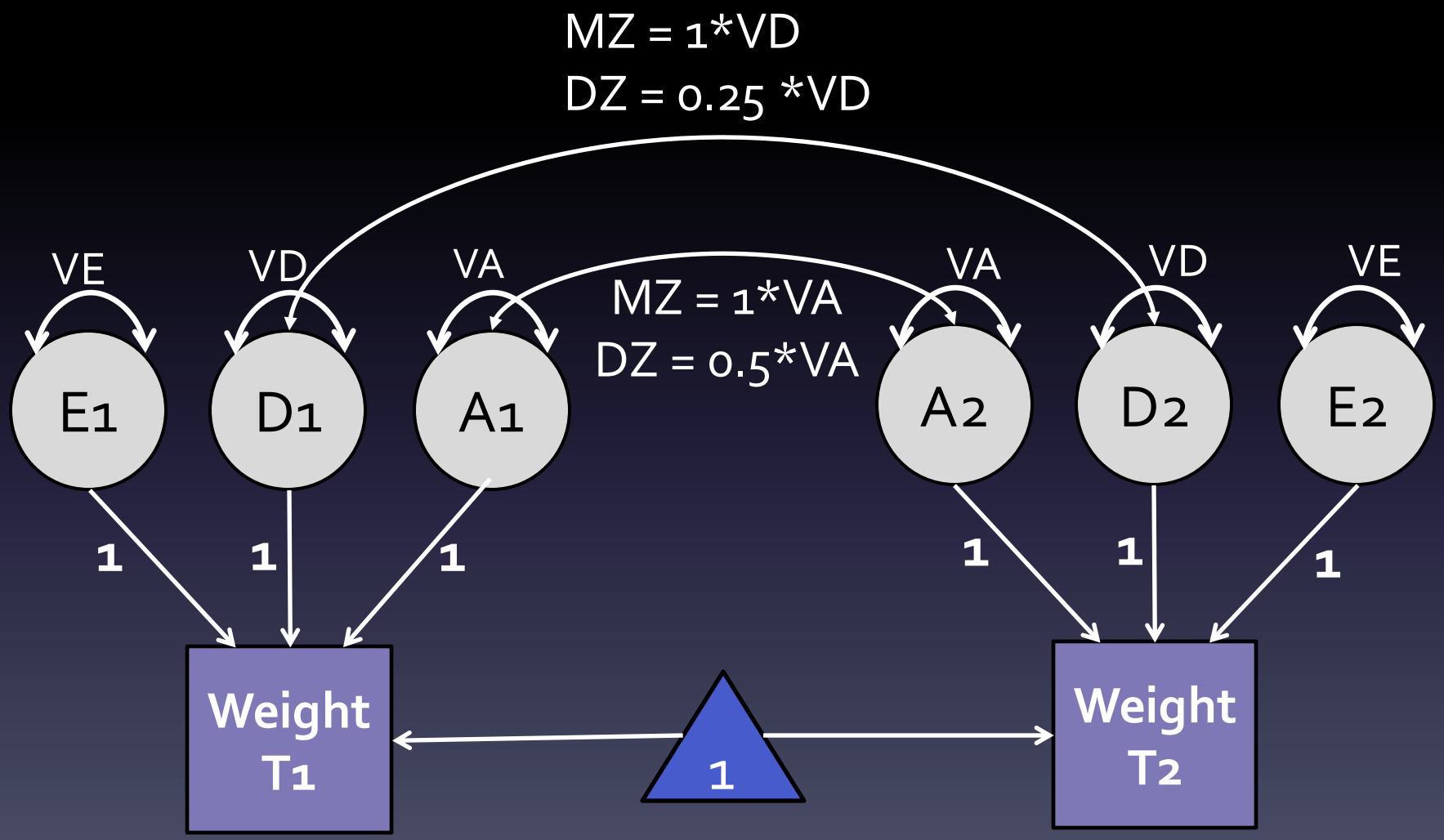
	T1	**T2**
T1	VA + VC +VE	VA + VC
T2	VA + VC	VA + VC+VE

DZ Twins

	T1	**T2**
T1	VA + VC + VE	0.5VA + VC
T2	0.5VA + VC	VA + VC + VE

Bonus!
Test Your Skill in Path Analysis of Basic Twin Models

ADE Model


MZ = 1*VD
DZ = 0.25 *VD
VE
VD
VA
MZ = 1*VA
DZ = 0.5*VA
VA
VD
VE
E1
D1
A1
A2
D2
E2
1
1
1
1
1
1
Weight T1
1
Weight T2

ADE Model
Predicted Variance and Covariance Matrices

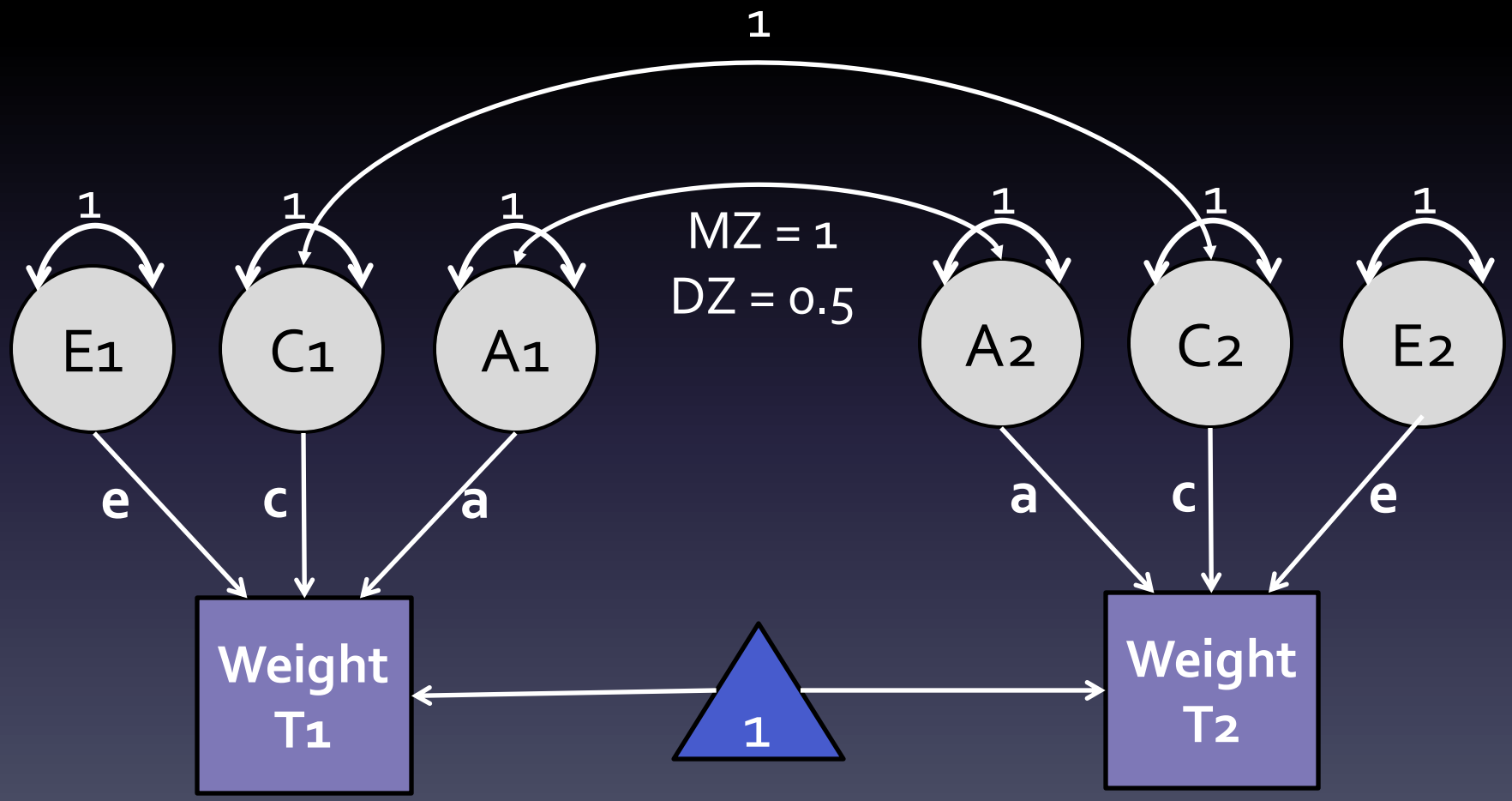
MZ Twins

	T1	**T2**
T1	VA + VD +VE	VA + VD
T2	VA + VD	VA + VD+VE

DZ Twins

	T1	**T2**
T1	VA + VD + VE	0.5VA + 0.25VD + VE
T2	0.5VA + 0.25VD + VE	VA + VD + VE

A Path Coefficient Twin Model (Classic Approach)



Predicted Variance and Covariance Matrices from a Path Coefficient Model

MZ Twins

	T1	**T2**
T1	$a^2 + c^2 + e^2$	$a^2 + c^2$
T2	$a^2 + c^2$	$a^2 + c^2 + e^2$

DZ Twins

	T1	**T2**
T1	$a^2 + c^2 + e^2$	$0.5a^2 + c^2$
T2	$0.5a^2 + c^2$	$a^2 + c^2 + e^2$

Summary

- The goal of path analysis is to work out the predicted variances and covariances of variables, given a specified model (e.g., ACE Model)

- This allows us to fit predicted variances/covariances to observed variances/covariances of the data to estimate the model parameters (e.g., regression coefficients, correlations) using statistical packages such as OpenMx

Thank You!

Questions?
Elizabeth.Prom-Wormley@vcuhealth.org

References

Wright, S. (1921). *Correlation and causation*. J. Agricultural Research 20: 557–585

Wright, S. (1920). *The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs*. Proceedings of the National Academy of Sciences 6: 320–332.