A (Brief) Introduction to SEM

DATA 695 Research Capstone Project

David Errington
Master of Data Science & Analytics Student
Faculty of Graduate Studies

June 13, 2025

UNIVERSITY OF CALGARY

Table of Contents

- **Territorial Acknowledgment**

- **Chapter 1: Introduction, Background, & Review**
  - Review of Simple/Multiple Regression
  - Linear Regression as a Structural Equation Model
  - Limitations of Multiple Regression

- **Chapter 2: Path Analysis**

- **Chapter 3: Confirmatory vs. Exploratory Factor Analysis**

- **Chapter 4: Structural Equation Models w/ Latent Variables**

- **References**

UNIVERSITY OF CALGARY

The University of Calgary, located in the heart of Southern Alberta, both acknowledges and pays tribute to the traditional territories of the peoples of Treaty 7, which include the Blackfoot Confederacy (comprised of the Siksika, the Piikani, and the Kainai First Nations), the Tsuut'ina First Nation, and the Stoney Nakoda (including Chiniki, Bearspaw, and Goodstoney First Nations). The City of Calgary is also home to the Métis Nation of Alberta (Districts 5 and 6).

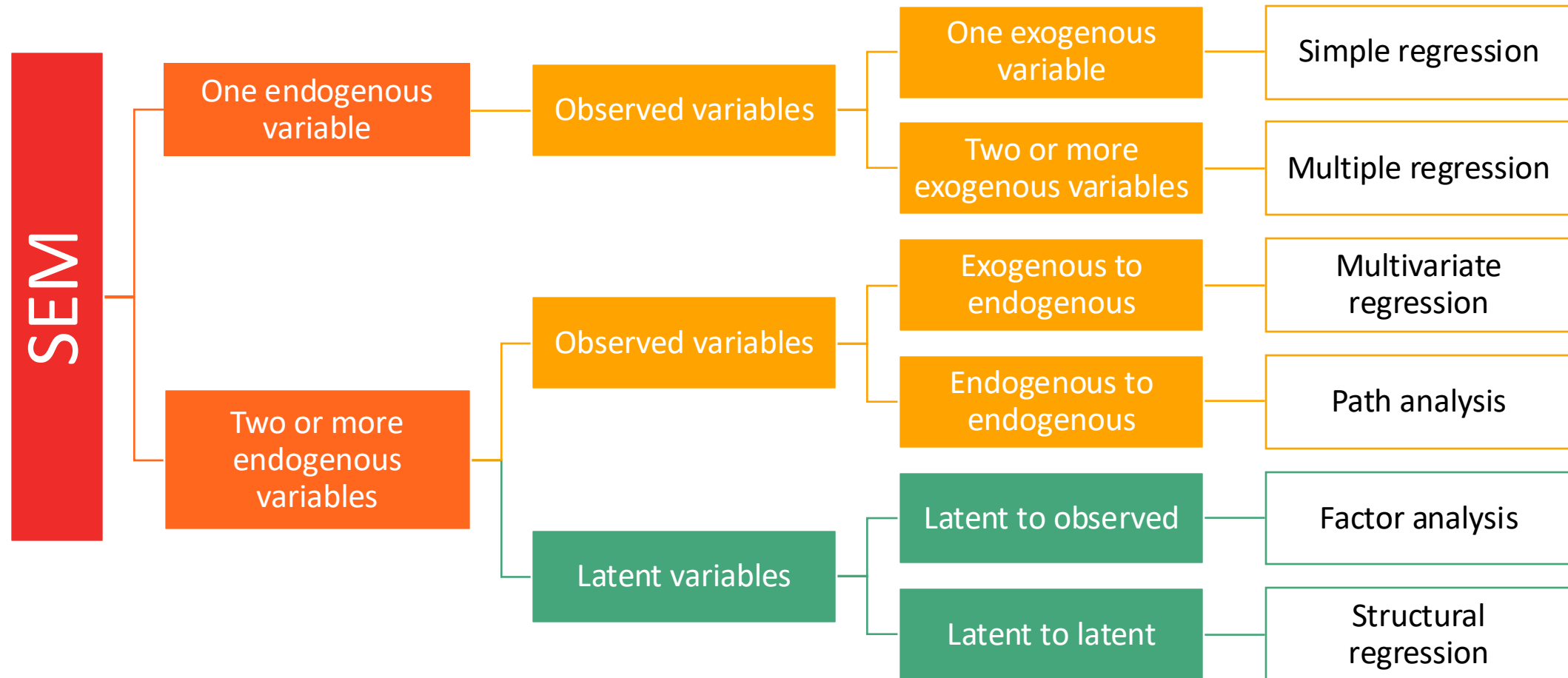# Chapter 1: Introduction, Background, & Review

UNIVERSITY OF
CALGARY

What is Structural Equation Modelling?

*"Structural equation modeling is a **linear model framework** that models both simultaneous regression equations with latent variables."*

— Johnny Lin, PhD (2024)

UNIVERSITY OF CALGARY

What is Structural Equation Modelling?

```
SEM
├── One endogenous variable
│   └── Observed variables
│       ├── One exogenous variable ──── Simple regression
│       └── Two or more exogenous variables ──── Multiple regression
│
└── Two or more endogenous variables
    ├── Observed variables
    │   ├── Exogenous to endogenous ──── Multivariate regression
    │   └── Endogenous to endogenous ──── Path analysis
    │
    └── Latent variables
        ├── Latent to observed ──── Factor analysis
        └── Latent to latent ──── Structural regression
```

UNIVERSITY OF
CALGARY

Definitions / Terminology

- ***Exogenous*** (a.k.a. ***independent***) *variables*:
    - Variables that are **not** expressed as a function of other variables; they exist "**outside**" the system of variables under study.
    - Notation:
        - x (single); or
        - x_i (multiple)

- ***Endogenous*** (a.k.a. ***dependent***) *variables*:
    - Variables that are expressed as a function of one or more other variables; they exist "**inside**" the system of variables under study.
    - Notation:
        - y (single); or
        - y_i (multiple)

Definitions / Terminology

- ***Observed** variable(s)*:
    - Variables that can be directly measured or "observed".
    - Examples: Height, weight, age, etc.

- ***Latent** variable(s)*:
    - Variables that (usually) cannot be directly measured; often "inferred".
    - Example: Intelligence.
    - Notation:
        - η "eta" (single); or
        - η_i (multiple)

UNIVERSITY OF
CALGARY

Linear Regression as a Structural Equation Model

- Step 1: Start with linear regression equation.

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

UNIVERSITY OF
CALGARY

Linear Regression as a Structural Equation Model

- Step 2: Add subscript i to each variable.

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon$$

Linear Regression as a Structural Equation Model

- Step 3: Take the intercept…

$$y_i = \boxed{\beta_0} + \beta_1 x_{1i} + \varepsilon$$

Intercept

UNIVERSITY OF
CALGARY

Linear Regression as a Structural Equation Model

- Step 3: And replace it with α.

$$y_i = \boxed{\alpha} + \beta_1 x_{1i} + \varepsilon$$

Intercept

UNIVERSITY OF
CALGARY

Linear Regression as a Structural Equation Model

- Step 4: Take the regression coefficient(s)...

$$y_i = \boxed{\alpha} + \boxed{\beta_1} x_{1i} + \varepsilon$$

Intercept

Coefficient

UNIVERSITY OF
CALGARY

Linear Regression as a Structural Equation Model

- Step 4: And replace it(them) with γ.

$$y_i = \boxed{\alpha} + \boxed{\gamma_1} x_{1i} + \varepsilon$$

Intercept

Coefficient

UNIVERSITY OF
CALGARY

Linear Regression as a Structural Equation Model

- Step 5: Take the residual error term…

$$y_i = \boxed{\alpha} + \boxed{\gamma_1}x_{1i} + \boxed{\varepsilon}$$

Error

Intercept

Coefficient

UNIVERSITY OF CALGARY

Linear Regression as a Structural Equation Model

- Step 5: And replace it with ζ_i.

$$y_i = \boxed{\alpha} + \boxed{\gamma_1} x_{1i} + \boxed{\zeta_i}$$

Error

Intercept

Coefficient

UNIVERSITY OF CALGARY

References

- Lin, J. (2024). *Introduction to Structural Equation Modeling (SEM) in R with lavaan.* OARC Stats – Statistical Consulting Web Resources | UCLA. https://stats.oarc.ucla.edu/r/seminars/rsem/