STRUCTURAL EQUATION MODELLING WITH LAVAAN IN R

ABOUT THE PRESENTER

- Data Scientist at Business Data Laboratory
- Statistician at Federal Teaching Hospital Gombe
- Udemy Instructor

CONTENTS OF TODAY'S PRESENTATION

- Introduction to SEM
- Key Concepts in SEM
- Overview of the lavaan Package
- Model Interpretation
- Case Study
- Summary
- Q/A

INTRODUCTION TO SEM

OVERVIEW OF SEM

- Multivariate statistical analysis describing relationship between latent and observed variables.
- Extension of multiple regression analysis and factor analysis.
- Used in social science to test models where theoretical constructs such as attitudes, beliefs and values are measured directly through multiple items.
- Used in psychology, where latent variables such as personality traits or intelligence are measured through questionnaires.

BENEFITS OF SEM

- Used to model complex relationships with inter related variables.
- Provides estimates for direct and indirect effects.
- Ability to handle missing data and provide model parameter estimates in the presence of missing data.

LIMITATIONS OF SEM

- Large sample size to obtain reliable estimates of model parameters
- Requires a well specified theoretical model
- Difficult to learn, with familiarity of statistical software and advanced statistical concepts needed.

APPLICATIONS OF SEM

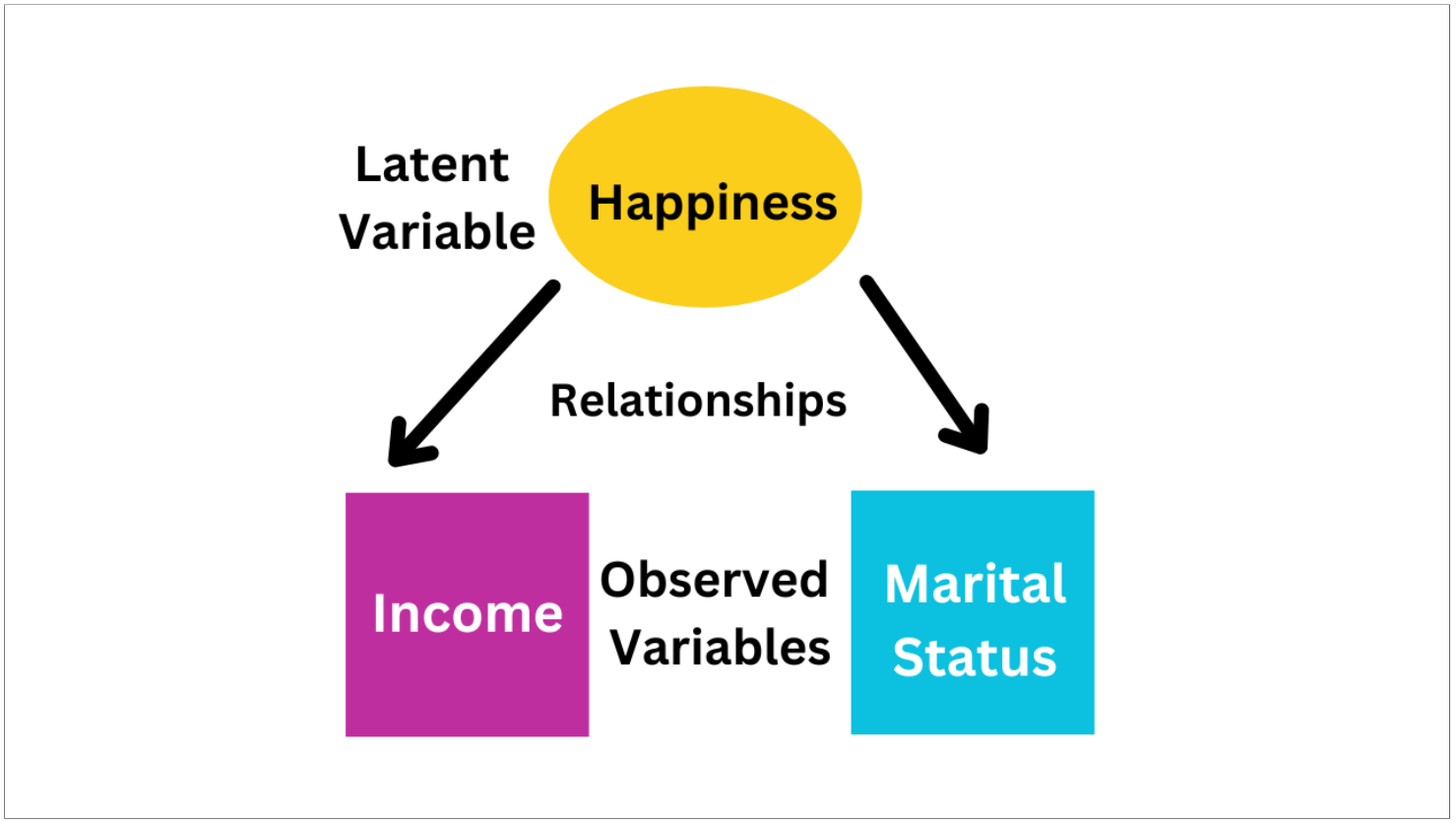
- Psychology
- Social Sciences
- Education
- Business

SEM KEY CONCEPTS

LATENT AND OBSERVED VARIABLES

- Observed variables are measured directly
- Latent variables are not directly measured and are inferred from observed variables.
- Latent variables are also known atent constructs.

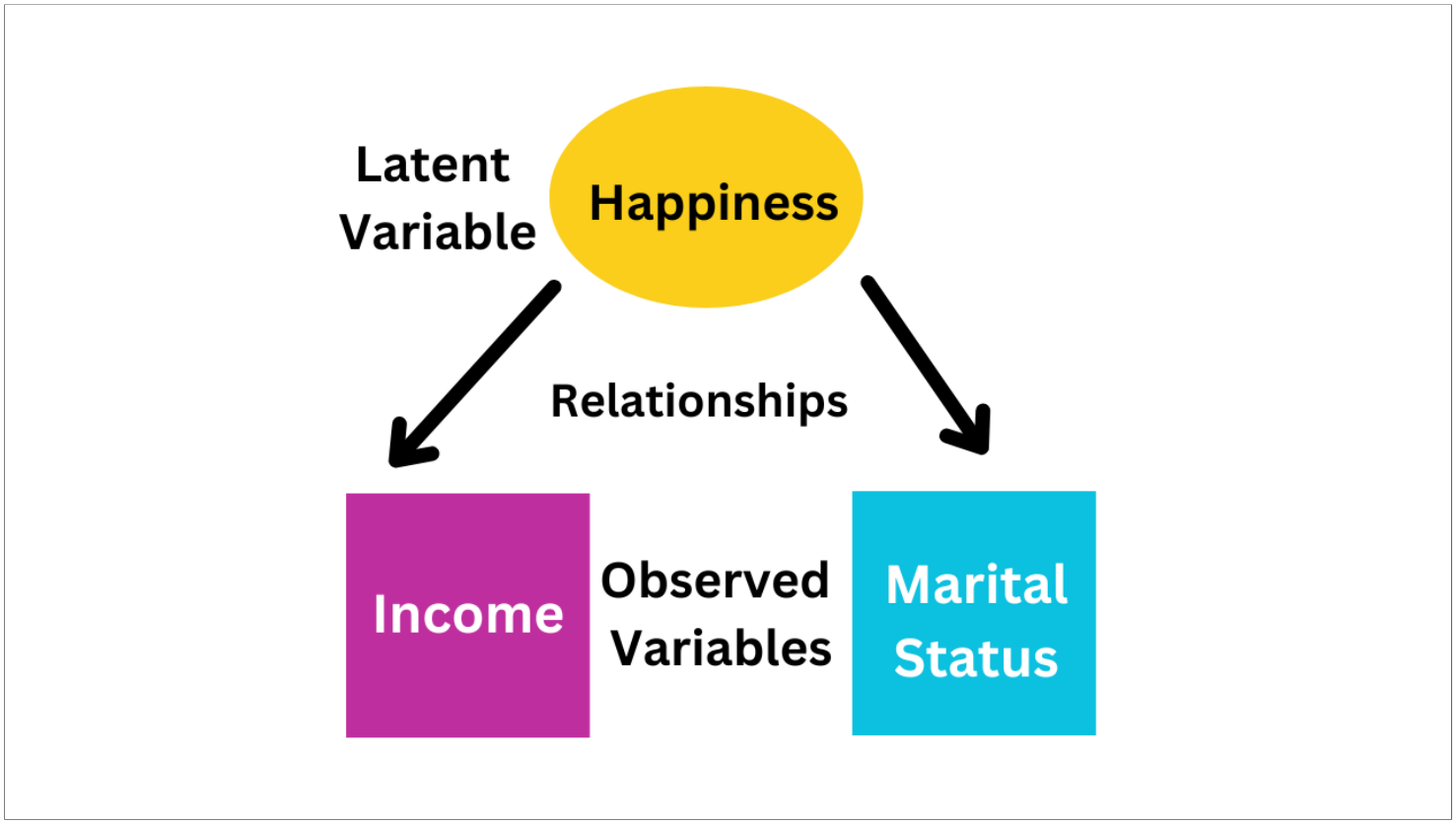
LATENT AND OBSERVED VARIABLES



PATH DIAGRAMS AND NOTATION

- Graphical representations of the relationships between variables in a SEM model
- Observed variables are represented by squares or rectangles
- Latent variables are represented by circles or ovals.
- Arrows between variables represents hypothesized relationships between them either direct or bidirectional.

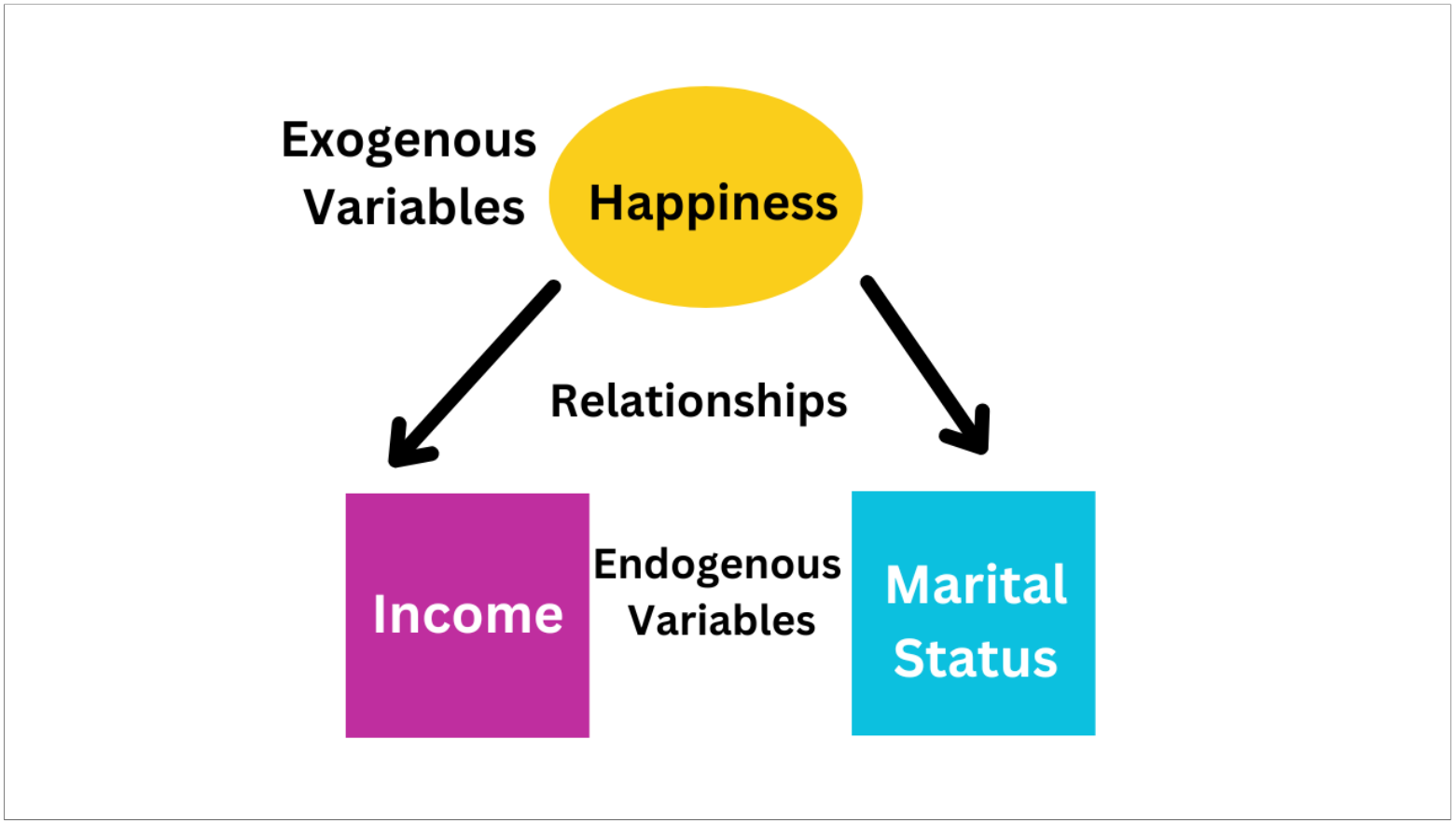
PATH DIAGRAMS AND NOTATION



ENDOGENOUS AND EXOGENOUS VARIABLES

- Endogenous variables are variables that are hypothesized to be caused by other variables in the model.
- Exogenous variables are not influenced by other variables in the model.
- In a path diagram, endogenous variables have arrows pointing to them while exogenous don't.

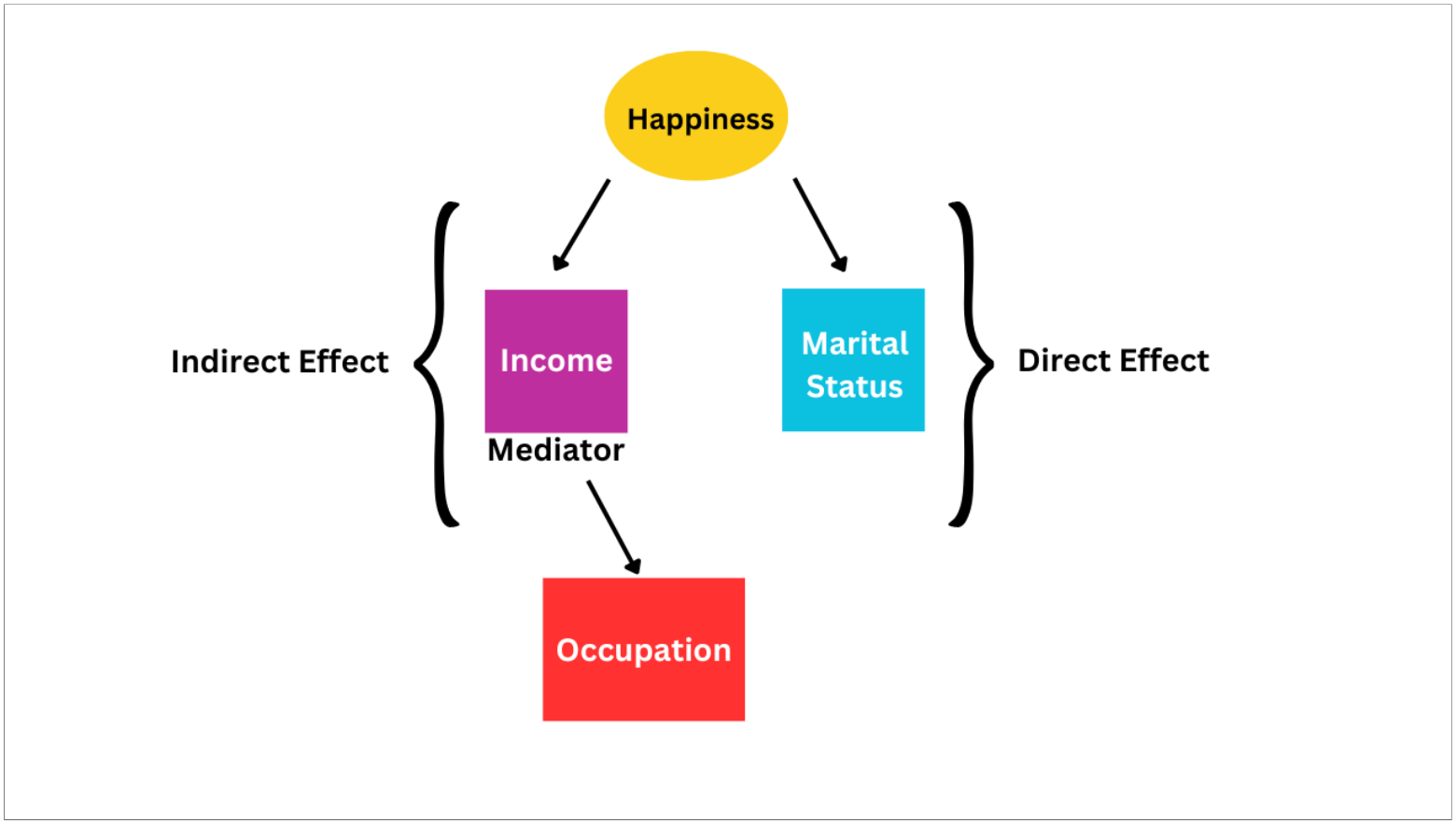
ENDOGENOUS AND EXOGENOUS VARIABLES



DIRECT AND INDIRECT EFFECTS

- A direct effect is the direct relationship between two variables.
- An indirect effect is a relationship that is mediated by one or more other variables in the model.

DIRECT AND INDIRECT EFFECTS



LAVAAN OVERVIEW

WHAT IS LAVAAN?

- R package for Structural Equation Modelling (SEM) analysis.
- Lavaan stands for Latent Variable Analysis

INSTALLING AND LOADING THE PACKAGE

To install the package

```
install.packages("lavaan")
```

To load the package

```
library(lavaan)
```

CREATING A MODEL OBJECT

- Model object specifies relationship between observed and latent variables

```
1 model <- 'y =~ x1 + x2'
```

- ~= specify observed variables to latent variable relationship
- ~ specifies regression between outcome and predictor variables.
- ~~ specifies covariance between variables.

FITTING THE MODEL

```
fit <- sem(model, data = mydata)
```

EXTRACTING RESULTS FROM THE MODEL

```
summary(fit, standardized = TRUE)
```

MODEL INTERPRETATION

INTERPRETING PARAMETER ESTIMATES AND STANDARD ERRORS

- Parameter estimates are values of the coefficients calculated for each predictor variable in the model
- Standard errors are measure of uncertainty in the estimation of these coefficients.

INTERPRETING FIT INDICES AND GOODNESS-OF-FIT TESTS

- Fit indices and goodness-of-fit Measures the goodness-of-fit of the model
- Fit indices includes
  - Chi-square
  - Root Mean Square Error of Approximation(RMSEA)
  - Comparative Fit Index(CFI)
  - Tucker-Lewis Index(TLI)
  - Standardized Root Mean Square Residual(SRMR)

BENCHMARK VALUES FOR FIT INDICES

- A non-significant chi-square value indicates a good fit
- A RMSEA value of 0.05 or less is considered a good fit, while a value between 0.05 and 0.08 indicates a reasonable fit. Values above 0.1 are considered a poor fit.
- A CFI value of 0.90 or above indicates a good fit, while a value above 0.95 indicates an excellent fit.
- A TLI value of 0.90 or above indicates a good fit, while a value above 0.95 indicates an excellent fit.
- A SRMR value of 0.08 or less is considered a good fit, while a value between 0.08 and 0.10 indicates a reasonable fit. Values above 0.10 are considered a poor fit.

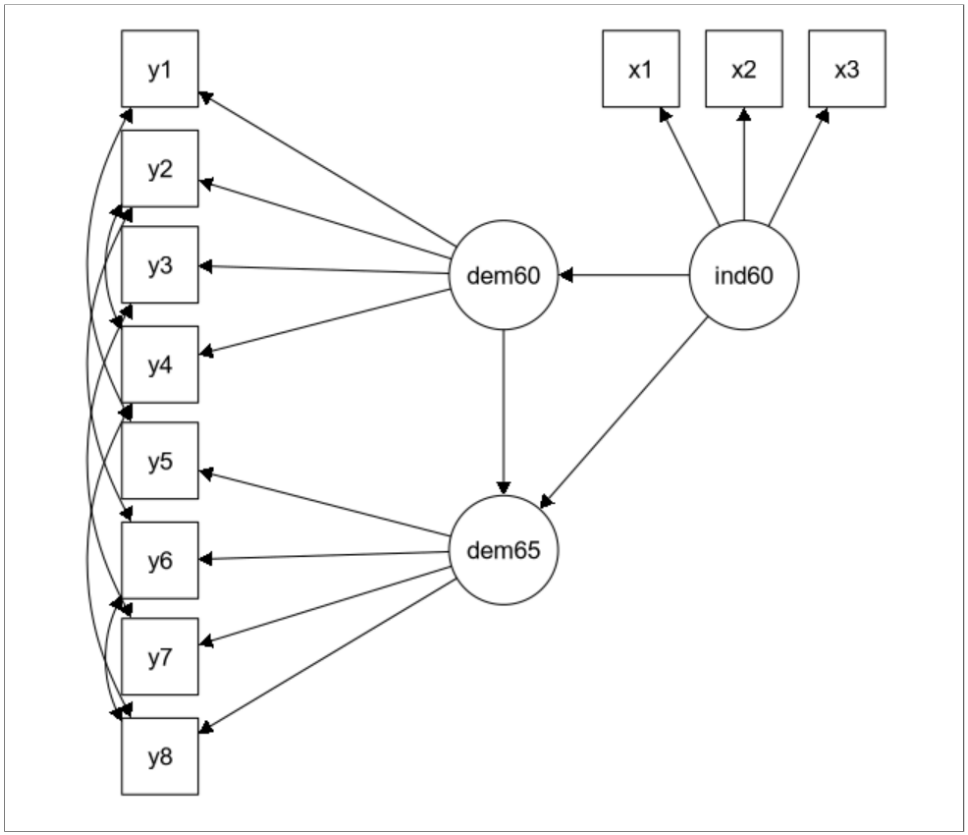
CASE STUDY

POLITICAL DEMOCRACY DATASET

The dataset contains various measures of political democracy and industrialization in developing countries.

- Political Democracy in 1960
  - y1 - Expert ratings of the freedom of the press in 1960
  - y2 - The freedom of political opposition in 1960
  - y3 - The fairness of elections in 1960
  - y4 - The effectiveness of the elected legistlature in 1960
- Political Democracy in 1965
  - y5 - Expert ratings of the freedom of the press in 1965
  - y6 - The freedom of political opposition in 1965
  - y7 - The fairness of elections in 1965
  - y8 - The effectiveness of the elected legistlature in 1965
- Level of Industrialization in 1960
  - x1 - The gross national product (GNP) per capita in 1960
  - x2 - The inanimate energy consumption per captita in 1960
  - x3 - The percentage of the labour force in industry in 1960

HYPOTHESIZED RELATIONSHIP



SUMMARY

KEY TAKEAWAYS

- Structural equation modeling (SEM) is a statistical technique used to analyze the relationships between observed and latent (unobserved) variables.
- Lavaan is an R package for SEM that allows users to specify their models using syntax that is similar to standard regression equations.
- When using lavaan, it's important to specify the model correctly, including specifying the relationships between the variables and the measurement model for the latent variables.
- Lavaan allows users to test the fit of their model using several fit indices, such as chi-square, RMSEA, CFI, and TLI.
- When interpreting the results of a lavaan analysis, it's important to consider both the statistical significance of the estimates and their practical significance.
- Finally, SEM is a powerful tool that can help researchers better understand the complex relationships between variables, but it's important to use it appropriately and with caution.

RESOURCES

- Chapter 6 Structural Equation Modeling, *Introduction to R for Data Science*
- Chapter 6 Structural Equation Modeling, *Using R For Social Work Research*
- The Lavaan Project

Q&A SESSION

Error