Introduction to structural equation modelling - advanced modelling techniques

Frank Pennekamp

02.11.2021

Department of Evolutionary Biology and Environmental Sciences

University of Zurich

Overview

Advanced SEM topics:

- Latent variables

Overview

Advanced SEM topics:

- Latent variables
- Composite variables

Overview

Advanced SEM topics:

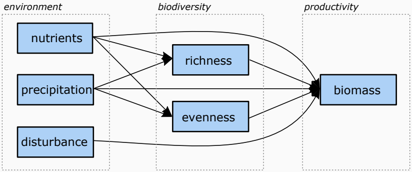
- Latent variables
- Composite variables
- Interactions

Overview

Advanced SEM topics:

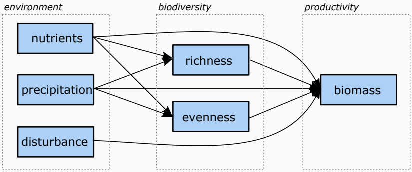
- Latent variables
- Composite variables
- Interactions
- Complex survey designs

Revisiting the meta-model



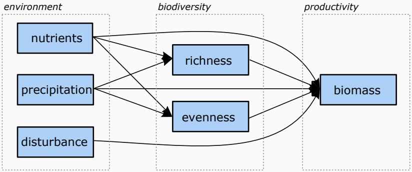
- A meta-model summarizes the concept behind a model and links it to theory.

Revisiting the meta-model



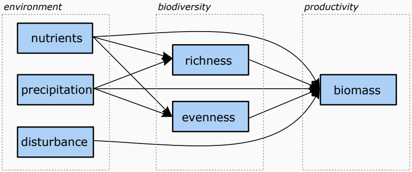
- A meta-model summarizes the concept behind a model and links it to theory.
- However, some of the parts of the model may be difficult to quantify and measure directly.

Revisiting the meta-model



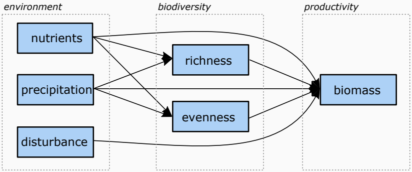
- A meta-model summarizes the concept behind a model and links it to theory.
- However, some of the parts of the model may be difficult to quantify and measure directly.
- This is because they represent an abstract, multifaceted concept (e.g., intelligence).

Revisiting the meta-model



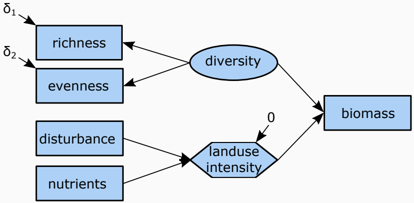
- A meta-model summarizes the concept behind a model and links it to theory.
- However, some of the parts of the model may be difficult to quantify and measure directly.
- This is because they represent an abstract, multifaceted concept (e.g., intelligence).
- Or, because we measure variables with error.

Revisiting the meta-model



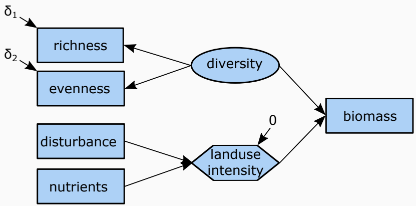
- A meta-model summarizes the concept behind a model and links it to theory.
- However, some of the parts of the model may be difficult to quantify and measure directly.
- This is because they represent an abstract, multifaceted concept (e.g., intelligence).
- Or, because we measure variables with error.
- This is where latent and composite variables are needed.

Latent variables



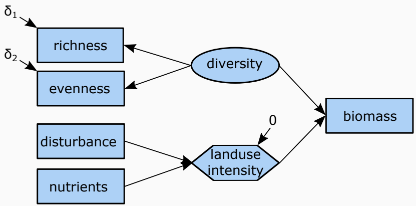
- Latent variables are variables that are unobserved, but whose influence can be summarized through one or more indicator variables.

Latent variables



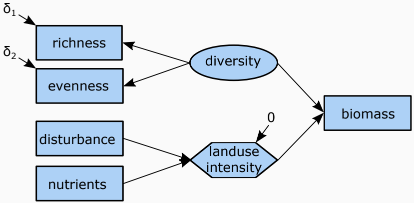
- Latent variables are variables that are unobserved, but whose influence can be summarized through one or more indicator variables.
- They can capture complex or conceptual properties of a system that are difficult to quantify or measure directly.

Latent variables



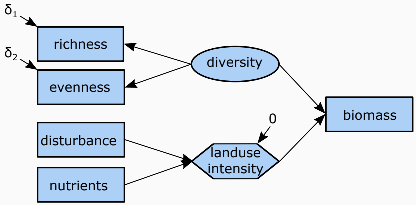
- Latent variables are variables that are unobserved, but whose influence can be summarized through one or more indicator variables.
- They can capture complex or conceptual properties of a system that are difficult to quantify or measure directly.
- Graphically, latent variables are often represented by an oval node shapes (ellipses).

Latent variables



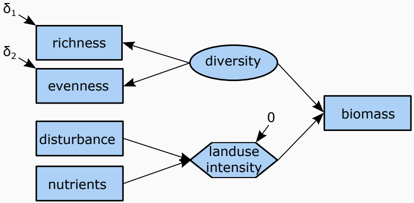
- Direction of causality reversed: from latent variables to the observed variable (reflective indicators).

Latent variables



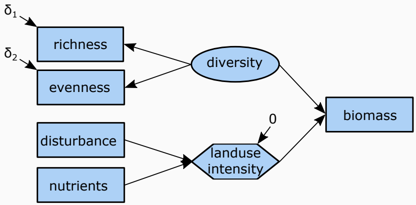
- Direction of causality reversed: from latent variables to the observed variable (reflective indicators).
- Indicator variables are emergent manifestation of the underlying phenomenon represented by the latent variable.

Latent variables



- Direction of causality reversed: from latent variables to the observed variable (reflective indicators).
- Indicator variables are emergent manifestation of the underlying phenomenon represented by the latent variable.
- Indicator variables should be positively correlated to the latent variable (i.e., driver). Recode to achieve this.

Latent variables



- Direction of causality reversed: from latent variables to the observed variable (reflective indicators).
- Indicator variables are emergent manifestation of the underlying phenomenon represented by the latent variable.
- Indicator variables should be positively correlated to the latent variable (i.e., driver). Recode to achieve this.
- Latent variable free of random or systematic measurement errors in contrast to its indicators.

Latent variables

- Both endogenous and exogenous latent variables possible.

Latent variables

- Both endogenous and exogenous latent variables possible.
- *Measurement model* focuses solely on relating indicators to latent variables.

Latent variables

- Both endogenous and exogenous latent variables possible.
- *Measurement model* focuses solely on relating indicators to latent variables.
- *Structural model* is one with directed paths between latent variables.

- Multi-indicator latent variables test hypothesis that multiple indicator variables are generated by the same underlying process.

Confirmatory factor analysis

- Multi-indicator latent variables test hypothesis that multiple indicator variables are generated by the same underlying process.
- Test that latent variable has given rise to emergent properties that, by virtue of a common cause, are correlated.

Confirmatory factor analysis

- Multi-indicator latent variables test hypothesis that multiple indicator variables are generated by the same underlying process.
- Test that latent variable has given rise to emergent properties that, by virtue of a common cause, are correlated.
- Precursor to evaluation of any structural models in which the latent variables appear.

Heywood cases

- Improper solutions:

Heywood cases

- Improper solutions:
    - Negative variances

Heywood cases

- Improper solutions:
    - Negative variances
    - Correlations larger than 1

Heywood cases

- Improper solutions:
    - Negative variances
    - Correlations larger than 1
- Reasons for the current warning could be

Heywood cases

- Improper solutions:
  - Negative variances
  - Correlations larger than 1
- Reasons for the current warning could be
  1) slightly negative error estimates (what would not really be a problem)

Heywood cases

- Improper solutions:
    - Negative variances
    - Correlations larger than 1
- Reasons for the current warning could be
    1) slightly negative error estimates (what would not really be a problem)
    2) the need to code indicators for the latent variable to be positively related

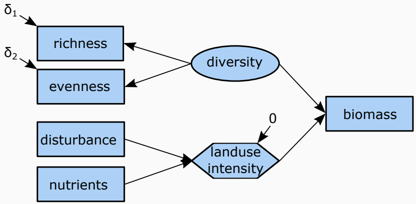
Heywood cases

- Improper solutions:
  - Negative variances
  - Correlations larger than 1
- Reasons for the current warning could be
  1) slightly negative error estimates (what would not really be a problem)
  2) the need to code indicators for the latent variable to be positively related
  3) local non-identification

Heywood cases

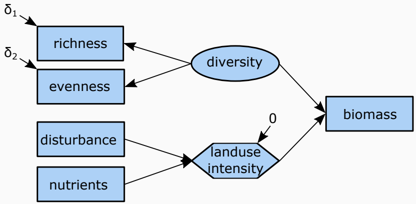
- Improper solutions:
  - Negative variances
  - Correlations larger than 1
- Reasons for the current warning could be
  1) slightly negative error estimates (what would not really be a problem)
  2) the need to code indicators for the latent variable to be positively related
  3) local non-identification
  4) general misspecification

Composite variables



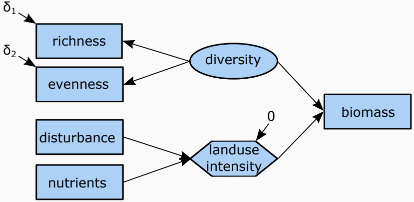
- Composite variables specify the influences of collections of other variables (examples).

Composite variables



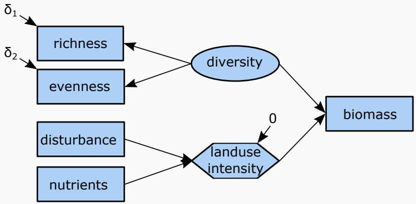
- Composite variables specify the influences of collections of other variables (examples).
- In comparison to latent variables, they arise from the indicators.

Composite variables



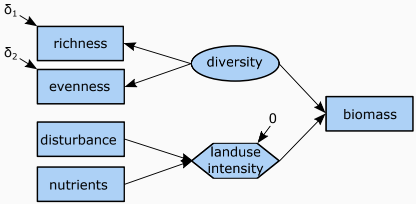
- Composite variables specify the influences of collections of other variables (examples).
- In comparison to latent variables, they arise from the indicators.
- they are visualized by arrows pointing from the indicators to the composite (formative or causal indicators).

Composite variables



- Composite variables specify the influences of collections of other variables (examples).
- In comparison to latent variables, they arise from the indicators.
- they are visualized by arrows pointing from the indicators to the composite (formative or causal indicators).
- since its values are determined by its causes (indicators), the error variance is specified to be 0.

Composite variables



- Composite variables specify the influences of collections of other variables (examples).
- In comparison to latent variables, they arise from the indicators.
- they are visualized by arrows pointing from the indicators to the composite (formative or causal indicators).
- since its values are determined by its causes (indicators), the error variance is specified to be 0.
- Composites are often shown as hexangular shapes or ellipses (latent composites).

Key differences between latent and composite variables

- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process represented by the latent variable.

Key differences between latent and composite variables

- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process represented by the latent variable.
- For composite variables, the flow of causation is reversed and the indicators are independent entities.

Key differences between latent and composite variables

- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process represented by the latent variable.
- For composite variables, the flow of causation is reversed and the indicators are independent entities.
- Rule of thumb 1: if the indicators are redundant, they likely belong to a latent variable.

Key differences between latent and composite variables

- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process represented by the latent variable.
- For composite variables, the flow of causation is reversed and the indicators are independent entities.
- Rule of thumb 1: if the indicators are redundant, they likely belong to a latent variable.
- Rule of thumb 2: If the meaning of the construct changes after dropping one of the indicators, the construct likely is a composite.

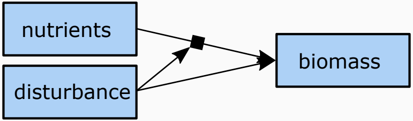
Key differences between latent and composite variables

- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process represented by the latent variable.
- For composite variables, the flow of causation is reversed and the indicators are independent entities.
- Rule of thumb 1: if the indicators are redundant, they likely belong to a latent variable.
- Rule of thumb 2: If the meaning of the construct changes after dropping one of the indicators, the construct likely is a composite.
- As a latent variable represents an underlying, data-generating process, its indicators need to be correlated to each other.

Key differences between latent and composite variables

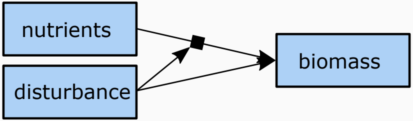
- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process represented by the latent variable.
- For composite variables, the flow of causation is reversed and the indicators are independent entities.
- Rule of thumb 1: if the indicators are redundant, they likely belong to a latent variable.
- Rule of thumb 2: If the meaning of the construct changes after dropping one of the indicators, the construct likely is a composite.
- As a latent variable represents an underlying, data-generating process, its indicators need to be correlated to each other.
- In the case of a composite variable, there is no assumption about the relation between the indicators. Thus, correlation among indicators are uninformative about the direction of the causal flow, but a lack of such contraindicates that the construct is a latent

Interactions (Moderation)



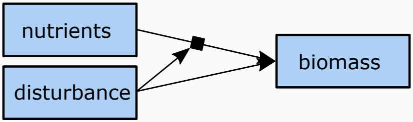
- In nature, things often are contingent on each other.

Interactions (Moderation)



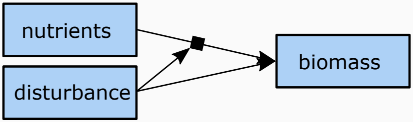
- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.

Interactions (Moderation)



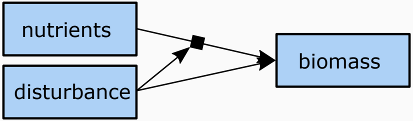
- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.
- Such a behaviour is called an interaction, which indicates that the effect of the two main effects are different when combined.

Interactions (Moderation)



- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.
- Such a behaviour is called an interaction, which indicates that the effect of the two main effects are different when combined.
- Both positive and negative interactions are possible.

Interactions (Moderation)



- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.
- Such a behaviour is called an interaction, which indicates that the effect of the two main effects are different when combined.
- Both positive and negative interactions are possible.
- In regression, the interaction is represented by a coefficient that estimates the effect of the product of the two predictors.

Interactions can be modelled in different ways in lavaan:

1) Multiple groups

Interactions (Moderation)

Interactions can be modelled in different ways in lavaan:

1) Multiple groups
2) Composites

- We will first use multigroup fitting.

```
mod <- sem(model, group = "age_class", data = dd)
```

Interactions (Moderation)

- We will first use multigroup fitting.
- This allows coefficients to vary among groups.

```
mod <- sem(model, group = "age_class", data = dd)
```

Interactions (Moderation)

- We will first use multigroup fitting.
- This allows coefficients to vary among groups.
- Lavaan offers the “group” argument to specify for which groups coefficients should be estimated.

```
mod <- sem(model, group = "age_class", data = dd)
```

Interactions (Moderation)

- We will first use multigroup fitting.
- This allows coefficients to vary among groups.
- Lavaan offers the “group” argument to specify for which groups coefficients should be estimated.
- Importantly, groups have to be of categorical nature.

```
mod <- sem(model, group = "age_class", data = dd)
```

Interactions (Moderation)

Lavaan allows to introduce equality constraints on various aspects via the `group.equal` argument:

```
mod <- sem(model, group = "age_class", group.equal =
c("regressions"), data = dd)
```

Additional constraints could be:

```
group.equal=c(
"intercepts",
"means",
"regressions",
"residuals",
"residual.covariances")
```

- Even more control by having the same name for different parameters:

```
model <- '
y ~ c("b1", "b1") * x1 + c("b2", "b2") * x3
x2 ~ c("b3", "b4") * x1
x3 ~ c("b5", "b5") * x2
'
```

Same coefficients for all but the effect of x1 on x2.

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.
- As such, data violates the principle of being i.i.d. (independent and identically distributed).

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.
- As such, data violates the principle of being i.i.d. (independent and identically distributed).
- Necessary to account for this structure in the data in the model.

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.
- As such, data violates the principle of being i.i.d. (independent and identically distributed).
- Necessary to account for this structure in the data in the model.
- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.
- As such, data violates the principle of being i.i.d. (independent and identically distributed).
- Necessary to account for this structure in the data in the model.
- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.
- As such, data violates the principle of being i.i.d. (independent and identically distributed).
- Necessary to account for this structure in the data in the model.
- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:
    1) initialize the design

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.
- As such, data violates the principle of being i.i.d. (independent and identically distributed).
- Necessary to account for this structure in the data in the model.
- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:
    1) initialize the design
    2) post-process the lavaan object and compute the adjusted results.

Complex sampling structure

- Often, the data is nested within sites or contains groups with non-random differences such as sexes or lifestages.
- As such, data violates the principle of being i.i.d. (independent and identically distributed).
- Necessary to account for this structure in the data in the model.
- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:
    1) initialize the design
    2) post-process the lavaan object and compute the adjusted results.
    3) Result is a corrected lavaan object.

Questions?