

# Modèles d'équations structureaux

## Contents

<b>Introduction</b>	<b>1</b>
<b>Contenu du cours</b>	<b>1</b>
<b>Types of variables and relationships in a structural equation model:</b>	<b>1</b>
From theoretical model to statistical model: . . . . .	5
<b>Model Fitting in Lavaan</b>	<b>7</b>
Du modèle théorique au modèle statistique: . . . . .	13
<b>Ajustement du modèle sur Lavaan</b>	<b>17</b>
Références . . . . .	21

## Introduction

Structural equation models belong to a family of models consisting of a set of mathematical equations and assumptions about a studied system. These assumptions stem from our prior knowledge or assumptions about how the system operates. In statistics, a system is a set of variables or phenomena that are studied within the framework of an analysis or study. These variables may be related by complex relationships. Our aim with the analysis is to understand how these variables interact with each other or how they influence a specific outcome or phenomenon.

## Contenu du cours

- Types of variables and relationships in a structural equation model;
- From theoretical model to statistical model;
- Model fitting in lavaan.

## Types of variables and relationships in a structural equation model:

The theoretical structure of a structural equation model encompasses several types of variables, defining their characteristics and roles within the model.

Based on their nature, variables can be classified into 1) latent variables and 2) observed variables. A latent variable is a variable that is not directly measured but represents concepts or traits that do not have a clear unit of measurement. This type of variable is often used in psychology (intelligence, satisfaction, etc.). Observed variables are measured or collected using established methods within the discipline. In ecology, we often work with observed variables.

Based on the role of variables in the theoretical model, variables can be exogenous or endogenous. Exogenous variables are independent variables that influence other variables in the model but are not influenced by any other variable in return. They represent the drivers of changes in our system. Endogenous variables, on the

other hand, are variables influenced by exogenous variables or other variables in the model, and typically represent the core of our system and the outcomes of the processes we are describing with our model.

Depending on how variables are conceptualized in the model, variables may have moderator and mediator status. A moderator is a variable that influences the strength and direction of the relationship between two variables. A moderator does not explain the “causes” of the relationship but only intervenes in the quantitative aspects of the relationship between two variables. A mediator is a variable that explains the relationship between an independent variable and a dependent variable. A variable acts as a mediator when it represents the causal mechanism linking two variables.

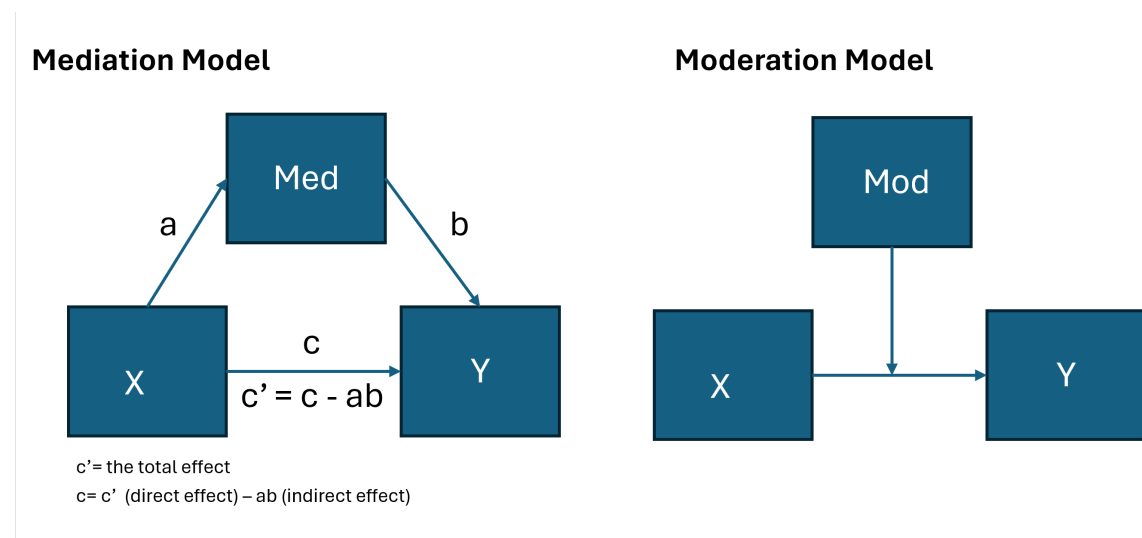


Figure 1: Models for mediation analysis and moderator analysis

In the figure, the arrows represent paths. In a structural equation model, a path represents a relationship that connects two variables, which can be either direct or indirect, and may include mediated effects. As you can see in the image, mediated effects refer to a situation where the impact of an independent variable on a dependent variable passes through another variable. In other words, the mediating variable transmits or mediates the effect of the independent variable on the dependent variable.

In a structural equation model, mediation and moderation analysis are integrated to gain a deeper understanding of the relationships between our variables, considering both mediating processes and the effects of moderators. However, mediation and moderation analyses can be conducted separately if we want to test relationships among three variables using the Psych package:

The ‘mediate’ function in the ‘psych’ package allows you to conduct a mediation analysis with the ‘mediate’ function. The mediating variable must be enclosed in parentheses to inform the function of its role in the model. In the example we will use to illustrate the function, we use the ‘mediate’ function to test the direct and indirect effects of temperature on the width of growth rings. Indeed, in the northern hemisphere and in cold environments where temperature is a limiting factor for growth, the width of growth rings increases with temperature. However, the width of the growth ring is closely related to the number of wood cells that make up the ring, which is, in turn, related to temperature. Thus, a portion of the total effect of temperature on the width of the ring is indeed mediated by the number of cells. If the direct effect of temperature on the number of cells is greater than that on the width of the ring, temperature changes will be more related to changes in the number of cells than to the width of the ring, and therefore the latter will be more easily predictable.

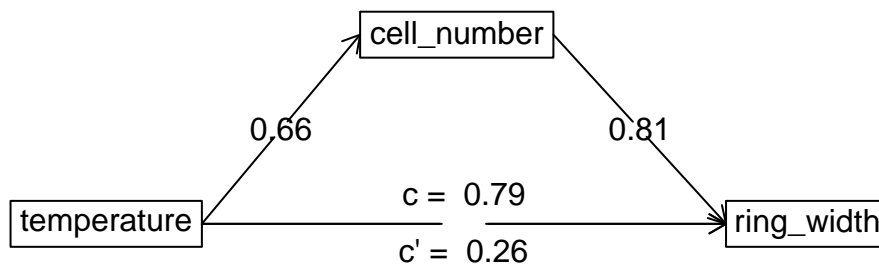
```
require(psych)
```

```
## Le chargement a nécessité le package : psych
```

```
## Warning: le package 'psych' a été compilé avec la version R 4.3.3
ringdatacell <- read.csv("C:/Users/buttoval/Documents/ECL8202/donnees//simulatedsemring.csv")

medanalysis<-mediate( ring_width ~ temperature + (cell_number) , data=ringdatacell )
```

## Mediation



“In the diagram, it is clear that the total effect of temperature is  $c = 0.79$ , but the direct effect of temperature on growth ring is much smaller,  $c' = 0.26$ .”

Nota bene: The ‘mediate’ function provides standardized coefficients and centered means by default. To avoid this, use the ‘zero=FALSE’ and ‘std=FALSE’ arguments.”

```
summary(medanalysis)
```

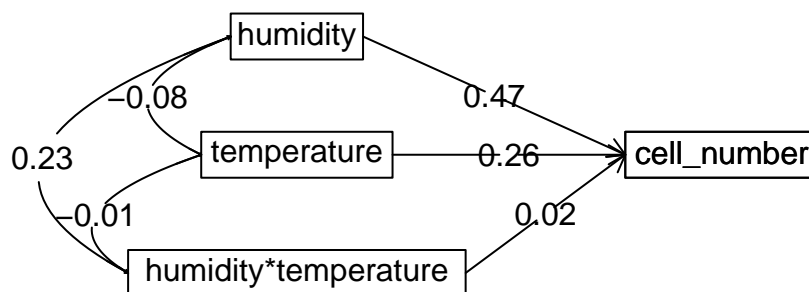
```
## Call: mediate(y = ring_width ~ temperature + (cell_number), data = ringdatacell)
##
## Direct effect estimates (traditional regression)      (c') X + M on Y
##           ring_width    se      t  df      Prob
## Intercept          1010.98 13.36 75.67 197 1.47e-147
## temperature           0.26  0.19  1.38 197  1.69e-01
## cell_number           0.81  0.06 13.12 197  1.14e-28
##
## R = 0.7 R2 = 0.49    F = 95.47 on 2 and 197 DF    p-value:  1.03e-29
##
## Total effect estimates (c) (X on Y)
##           ring_width    se      t  df      Prob
## Intercept          1179.31  5.08 232.16 198 3.47e-243
```

```
## temperature      0.79 0.25   3.18 198  1.72e-03
##
## 'a' effect estimates (X on M)
##           cell_number  se      t  df      Prob
## Intercept           207.84 4.28 48.53 198 6.92e-112
## temperature          0.66 0.21  3.14 198  1.95e-03
##
## 'b' effect estimates (M on Y controlling for X)
##           ring_width  se      t  df      Prob
## cell_number          0.81 0.06 13.12 197 1.14e-28
##
## 'ab' effect estimates (through all mediators)
##           ring_width boot  sd lower upper
## temperature          0.53 0.53 0.18  0.18  0.91
```

In a moderation analysis, we examine how the effect of an independent variable on a dependent variable may be modified by another variable, known as a moderator. To assess this interaction, we typically include an interaction term in the multiple regression model. This interaction term allows us to test whether the effect of the independent variable on the dependent variable varies depending on the levels of the moderator. In summary, a moderation analysis is a multiple regression with an interaction. We can obtain a moderation model using the 'mediate' function, but without specifying the effect of the moderator

```
mod_analysis<-mediate( cell_number ~ humidity + humidity*temperature + temperature,
                      data=ringdatacell,std=TRUE)
```

## Moderation model



```
summary(mod_analysis)
```

```
## Call: mediate(y = cell_number ~ humidity + humidity * temperature +
```

```
##      temperature, data = ringdatacell, std = TRUE)
##
## No mediator specified leads to traditional regression
##           cell_number    se    t   df    Prob
## Intercept                0.00 0.06 0.00 196 1.00e+00
## humidity                  0.47 0.06 7.56 196 1.55e-12
## temperature               0.26 0.06 4.20 196 4.13e-05
## humidity*temperature      0.02 0.06 0.26 196 7.92e-01
##
## R = 0.52 R2 = 0.27   F = 24.77 on 3 and 196 DF   p-value: 1.25e-13
```

## From theoretical model to statistical model:

When deciding to conduct a SEM, a well-defined hypothesis represents the best investment to leverage this analysis. Therefore, we will classify our variables according to the typology defined in the previous paragraph, and then construct our a priori model. This model represents our understanding of the system based on the scientific evidence we have gathered in our study and contains our hypotheses in the form of links between variables.

In a SEM, we establish a theoretical model based on postulated relationships between variables, and then test this model with real data to see if it fits these data well. The objective is to determine if the theoretical model is statistically valid and can be generalized to real data. Model validation thus requires not rejecting the null hypothesis that the relationships between variables as specified in the theoretical model are also present in the real data, and that any observed difference between the theoretical model and real data is due to chance or measurement errors.

In a SEM, symbols can be used to represent links and variables in a diagram:



Figure 2: Symbols and diagrams

There are several packages available in R that allow you to perform a SEM. Here, we will use Lavaan, which has its own syntax for defining the variables of the model and their links.

Formula and Definition	Operator	Meaning
Latent variable	$\sim$	is obtained from
Covariate	$\sim\sim$	is correlated with
intercept	$\sim 1$	intercept

For an example, we will use a simulated dataset containing the following information:

Temperature: The average temperature in degrees Celsius recorded during the growing season. Humidity: The average percentage of relative humidity recorded during the growing season. Stem Size: The average stem size of the plant, in cm. Cell Number: The total number of cells observed in each growth ring. Ring Width: The average width of growth rings, a measure of annual tree growth.

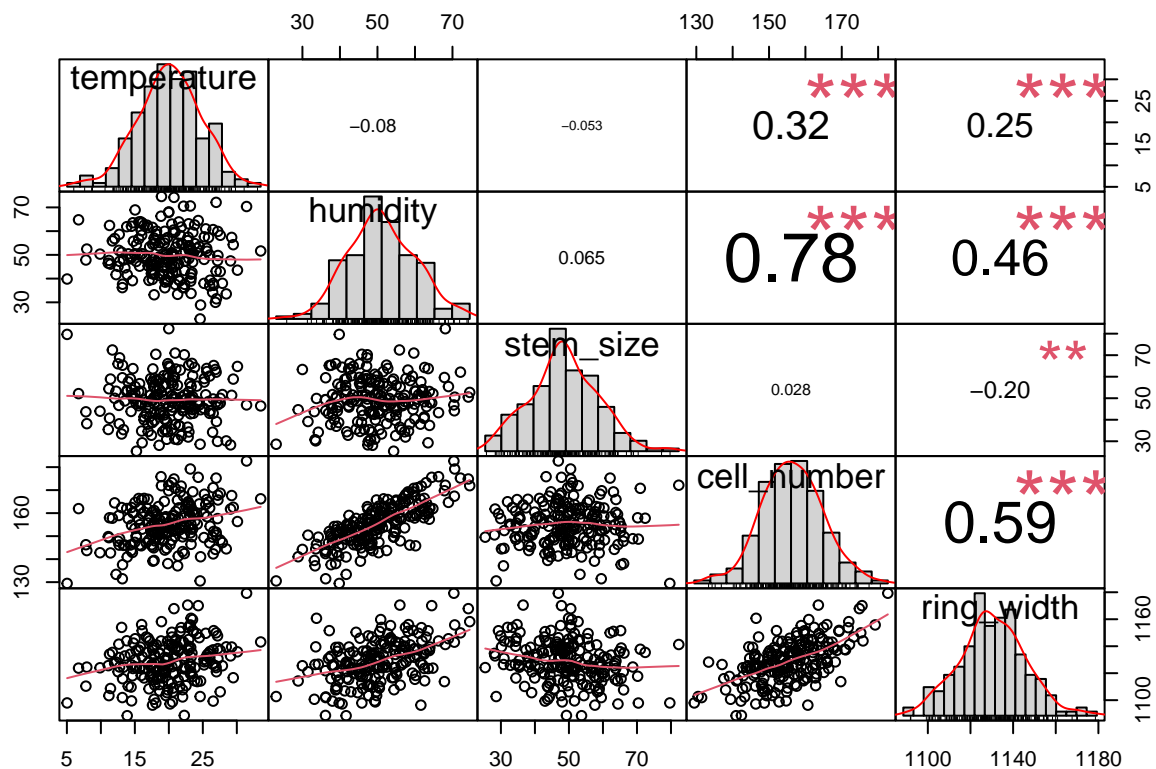
```
simulated_data1 <- read.csv("C:/Users/buttoval/Documents/ECL8202/donnees/simulatedsemring1.csv")
```

Before adjusting a SEM model it is often useful to visualize a correlation matrix between the variables:

```
library(PerformanceAnalytics)
```

```
## Le chargement a nécessité le package : xts
```

```
## Le chargement a nécessité le package : zoo
##
## Attachement du package : 'zoo'
## Les objets suivants sont masqués depuis 'package:base':
##
##   as.Date, as.Date.numeric
##
## Attachement du package : 'PerformanceAnalytics'
## L'objet suivant est masqué depuis 'package:graphics':
##
##   legend
chart.Correlation(simulated_data1, histogram = TRUE, method = "pearson")
```



The correlation matrix displays high correlations between most of our variables, but it does not tell us anything about the relationships among them.

A theoretical model based on the literature is proposed to explain the relationships between our variables and the underlying process of growth:

On lavaan, we can translate the model using the following syntax:

```
require(lavaan)
```

```
## Le chargement a nécessité le package : lavaan
```

```
## Warning: le package 'lavaan' a été compilé avec la version R 4.3.3
```

## Sem structure – Theoretical model

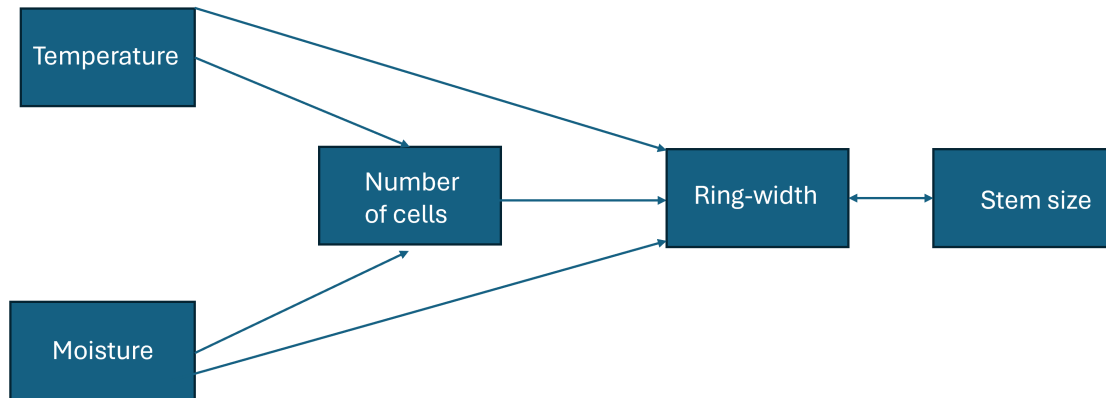


Figure 3: Relationship between environmental factors and growth: each link is supported by literature sources

```
## This is lavaan 0.6-17
## lavaan is FREE software! Please report any bugs.
##
## Attachement du package : 'lavaan'
## L'objet suivant est masqué depuis 'package:psych':
##
##      cor2cov
myModel <- '
# regressions
ring_width ~ temperature + humidity+ cell_number
cell_number ~ temperature + humidity
ring_width ~ ~ stem_size
'

fit <- sem(model = myModel,
           data = simulated_data1)
```

The summary of the `sem` function provides us with regression coefficients that indicate the strength and direction of the relationship between our variables. Covariance coefficients also measure the strength and direction of the correlation between variables. These coefficients can be interpreted like any coefficients in a linear regression. However, it can be very useful to standardize the coefficients if we want to compare the effect of different variables on our response variable.

## Model Fitting in Lavaan

We consider a SEM to adequately represent our data distribution and natural phenomenon when the p-value (chi-square) of the model is not significant. This is because we aim to observe a correspondence between the observed data and the predictions of our model, which should not exhibit significant differences. In this case, the p-value is 0.713, which reassures us about the model's ability to represent our phenomenon. It is also

useful to check other model fit measures for a more comprehensive evaluation of model adequacy. Here are the commonly used indicators, as summarized by Joreskog, K., & Sorbom, D. (1993):

**Table 1: Fit indices and their acceptable thresholds**

Fit Index	Acceptable Threshold Levels	Description
<i>Absolute Fit Indices</i>		
Chi-Square $\chi^2$	Low $\chi^2$ relative to degrees of freedom with an insignificant $p$ value ( $p > 0.05$ )	
Relative $\chi^2$ ( $\chi^2/\text{df}$ )	2:1 (Tabachnik and Fidell, 2007) 3:1 (Kline, 2005)	Adjusts for sample size.
Root Mean Square Error of Approximation (RMSEA)	Values less than 0.07 (Steiger, 2007)	Has a known distribution. Favours parsimony. Values less than 0.03 represent excellent fit.
GFI	Values greater than 0.95	Scaled between 0 and 1, with higher values indicating better model fit. This statistic should be used with caution.
AGFI	Values greater than 0.95	Adjusts the GFI based on the number of parameters in the model. Values can fall outside the 0-1.0 range.
RMR	Good models have small RMR (Tabachnik and Fidell, 2007)	Residual based. The average squared differences between the residuals of the sample covariances and the residuals of the estimated covariances. Unstandardised.
SRMR	SRMR less than 0.08 (Hu and Bentler, 1999)	Standardised version of the RMR. Easier to interpret due to its standardised nature.
<i>Incremental Fit Indices</i>		
NFI	Values greater than 0.95	Assesses fit relative to a baseline model which assumes no covariances between the observed variables. Has a tendency to overestimate fit in small samples.
NNFI (TLI)	Values greater than 0.95	Non-normed, values can fall outside the 0-1 range. Favours parsimony. Performs well in simulation studies (Sharma et al, 2005; McDonald and Marsh, 1990)
CFI	Values greater than 0.95	Normed, 0-1 range.

Figure 4: Model adjustment

it is possible to compute some of these metrics with the function `fitMeasures()`

```
fitMeasures(fit, c("chisq", "df", "pvalue", "cfi", "rmsea"))
```

```
## chisq    df pvalue    cfi  rmsea
## 1.334 3.000 0.721 1.000 0.000
```

A combination of these metrics can be used to validate the model:

It's possible to visualize the standardized coefficients for the relationships: very useful if one wants to avoid the size effect, as well as the determination coefficients of each regression, very important for interpreting the model fit on each variable

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



**Table 2: Hu and Bentler's Two-Index Presentation Strategy (1999)**

Fit Index Combination	Combinational Rules
NNFI (TLI) and SRMR	NNFI of 0.96 or higher and an SRMR of .09 or lower
RMSEA and SRMR	RMSEA of 0.06 or lower and a SRMR of 0.09 or lower
CFI and SRMR	CFI of .96 or higher and a SRMR of 0.09 or lower

Figure 5: Acceptable combinations of diagnostic indicators

```
## lavaan 0.6.17 ended normally after 32 iterations
##
## Estimator ML
## Optimization method NLMINB
## Number of model parameters 9
##
## Number of observations 200
##
## Model Test User Model:
##
## Test statistic 1.334
## Degrees of freedom 3
## P-value (Chi-square) 0.721
##
## Parameter Estimates:
##
## Standard errors Standard
## Information Expected
## Information saturated (h1) model Structured
##
## Regressions:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## ring_width ~
## temperature 0.251 0.230 1.092 0.275 0.251 0.076
## humidity 0.127 0.179 0.711 0.477 0.127 0.075
## cell_number 0.883 0.190 4.638 0.000 0.883 0.512
## cell_number ~
## temperature 0.742 0.068 10.991 0.000 0.742 0.386
## humidity 0.801 0.035 23.050 0.000 0.801 0.810
##
## Covariances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .ring_width ~~
## stem_size -35.300 9.797 -3.603 0.000 -35.300 -0.263
##
## Variances:
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all
## .ring_width 166.464 16.646 10.000 0.000 166.464 0.643
## .cell_number 21.391 2.139 10.000 0.000 21.391 0.245
## stem_size 107.827 10.783 10.000 0.000 107.827 1.000
```

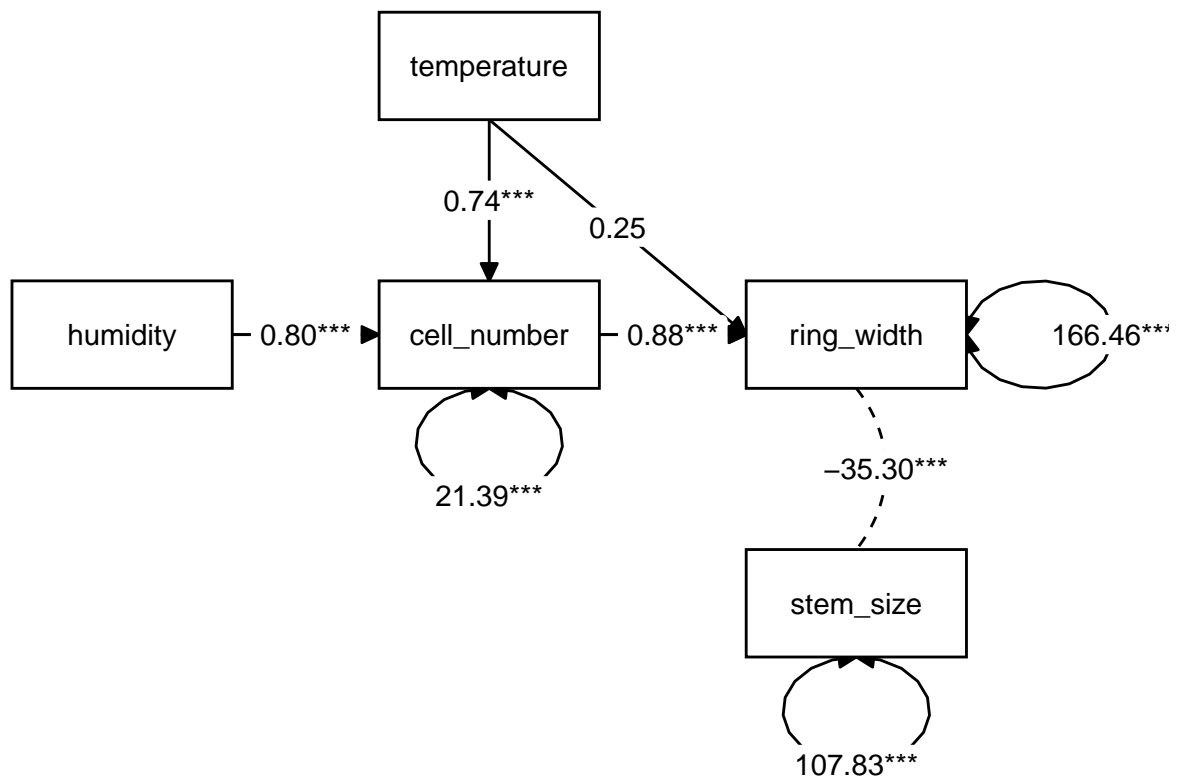
```
##
## R-Square:
##           Estimate
##   ring_width    0.357
##   cell_number   0.755
```

With tidySEM, it's possible to obtain a graphical representation of our model. In this case, it's important to pay attention to how we interpret the different elements because the `graph_sem` function uses its own code to identify the elements of the model. For example, the dashed lines represent covariances, while the small circle with arrows represents the residual variances of each variable. You may have already seen this term in the lavaan summary with the estimation error and p-values. Significant variances indicate that the variable has a significant influence on the model results

```
require(tidySEM)
```

```
## Le chargement a nécessité le package : tidySEM
## Warning: le package 'tidySEM' a été compilé avec la version R 4.3.3
## Le chargement a nécessité le package : OpenMx
## Warning: le package 'OpenMx' a été compilé avec la version R 4.3.3
##
## Attachement du package : 'OpenMx'
## L'objet suivant est masqué depuis 'package:psych':
##
##      tr
##
## Registered S3 method overwritten by 'tidySEM':
##   method      from
##   predict.MxModel OpenMx
```

```
graph_sem(fit)
```



## ## References

Revelle, 2024: How to use the psych package for regression and mediation analysis: <https://cran.r-project.org/web/packages/psychTools/vignettes/mediation.pdf>

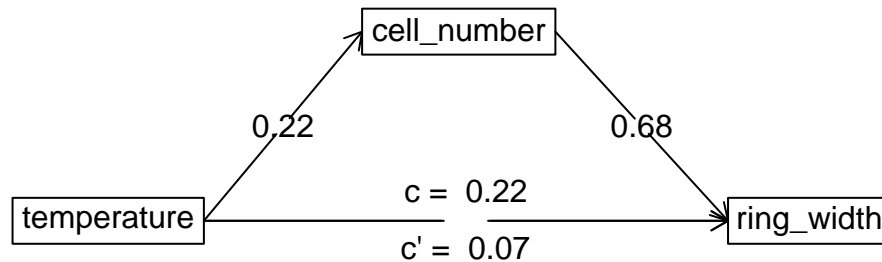
Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/> <https://lavaan.ugent.be/>

Joreskog, K., & Sorbom, D. (1993). Structural equation modelling: Guidelines for determining model fit. NY: University Press of America.

*Nota bene* la fonction “mediate” nous donne les coefficients standardisées par défaut, ainsi que les moyennes centrées. pour éviter cela il faut utiliser les arguments “zero=FALSE” et std=FALSE.

```
medanalysis<-mediate( ring_width ~ temperature + (cell_number) , std=TRUE, data=ringdatacell)
```

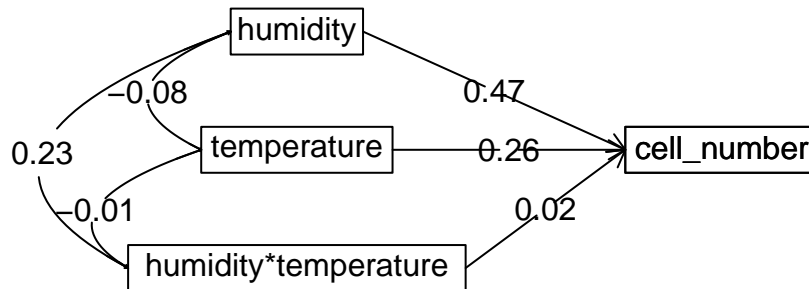
## Mediation



Dans une analyse de modération, on examine comment l'effet d'une variable indépendante sur une variable dépendante peut être modifié par une autre variable, appelée modérateur. Pour évaluer cette interaction, on inclut généralement un terme d'interaction dans le modèle de régression multiple. Ce terme d'interaction permet de tester si l'effet de la variable indépendante sur la variable dépendante varie en fonction des niveaux du modérateur. En résumé, une analyse de modération est une régression multiple avec une interaction. On peut obtenir un modèle de modération en utilisant la fonction `mediate`, mais sans spécifier l'effet du modérateur.

```
mod_analysis<-mediate( cell_number ~ humidity + humidity*temperature + temperature,
                        data=ringdatacell,std=TRUE)
```

## Moderation model



```
summary(mod_analysis)
```

```
## Call: mediate(y = cell_number ~ humidity + humidity * temperature +  
##           temperature, data = ringdatacell, std = TRUE)  
##  
## No mediator specified leads to traditional regression  
##  
##           cell_number    se    t  df    Prob  
## Intercept              0.00 0.06 0.00 196 1.00e+00  
## humidity                0.47 0.06 7.56 196 1.55e-12  
## temperature             0.26 0.06 4.20 196 4.13e-05  
## humidity*temperature    0.02 0.06 0.26 196 7.92e-01  
##  
## R = 0.52 R2 = 0.27    F = 24.77 on 3 and 196 DF    p-value: 1.25e-13
```

### Du modèle théorique au modèle statistique:

Lorsqu'on décide de réaliser une SEM, une hypothèse bien définie représente le meilleur investissement pour valoriser cette analyse. Nous allons donc classer nos variables selon la typologie définie dans le paragraphe précédent, puis construire notre modèle **a priori**. Ce modèle représente notre compréhension du système basée sur les preuves scientifiques que nous avons recueillies dans notre étude, et contient nos hypothèses sous forme de liens entre les variables.

Dans une SEM, on établit un modèle théorique basé sur des relations postulées entre les variables, puis on teste ce modèle avec des données réelles pour voir s'il correspond bien à ces données. L'objectif est de déterminer si le modèle théorique est statistiquement valide et peut être généralisé aux données réelles. La validation du modèle nécessite donc de ne pas rejeter l'hypothèse nulle selon laquelle les relations entre les variables telles qu'elles sont spécifiées dans le modèle théorique sont également présentes dans les données.

réelles, et que toute différence observée entre le modèle théorique et les données réelles est due au hasard ou à des erreurs de mesures.

Dans un SEM, on peut utiliser des symboles peuvent être utilisés pour représenter les liens et les variables dans un diagramme:



Figure 6: Symboles pour les diagrammes

Il y a plusieurs packages qui peuvent vous permettre de réaliser un modèle SEM sur R. Ici nous allons utiliser Lavaan, qui utilise sa propre syntaxe pour définir les variables du modèle et leur liens.

Formule et définition	opérateur	signification
variable latente	$=\sim$	obtenue à partir de
Covariable	$\sim\sim$	est corrélé avec
intercepte	$\sim 1$	intercepte

Pour un exemple, nous allons utiliser un jeu de données simulées contenant les informations suivantes:

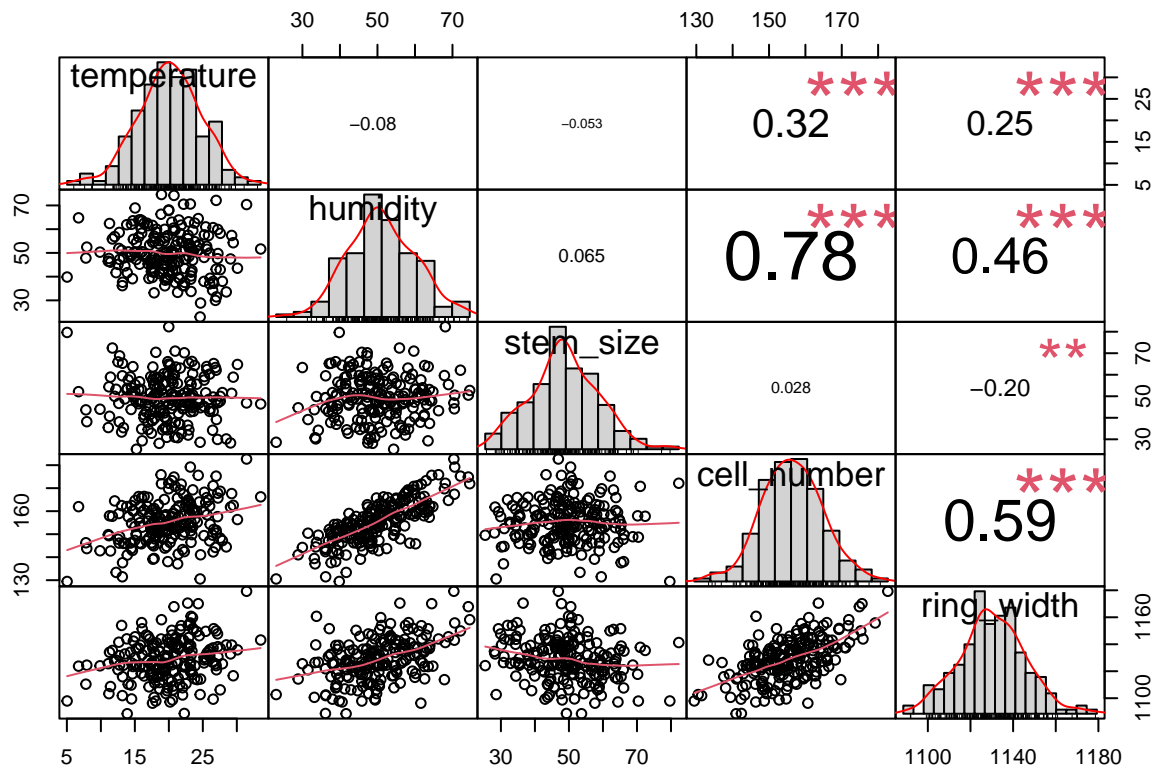
Température : La température moyenne en degrés Celsius enregistrées pendant la saison de croissance.  
Humidité : Le pourcentage moyen d'humidité relatif enregistré pendant la saison de croissance.  
Taille de la tige : La taille moyenne de la tige de la plante, en cm  
nombre de cellules : le nombre total de cellules observées dans chaque cerne de croissance  
Largeur des cernes : La largeur moyenne des anneaux de croissance, une mesure de la croissance annuelle des arbres.

```
simulated_data1 <- read.csv("C:/Users/buttoval/Documents/ECL8202/donnees/simulatedsemring1.csv")
```

Avant d'ajuster un modèle SEM, il est souvent utile de visualiser une matrice de corrélation entre les variables

```
library(PerformanceAnalytics)
```

```
chart.Correlation(simulated_data1, histogram = TRUE, method = "pearson")
```



La matrice de corrélation montre des corrélations élevées entre la plupart de nos variables, mais elle ne nous dit rien sur les relations entre elles. Un modèle théorique basé sur la littérature est proposé pour expliquer les liens entre nos variables et le processus sous-jacent de la croissance:

### Sem structure – Theoretical model

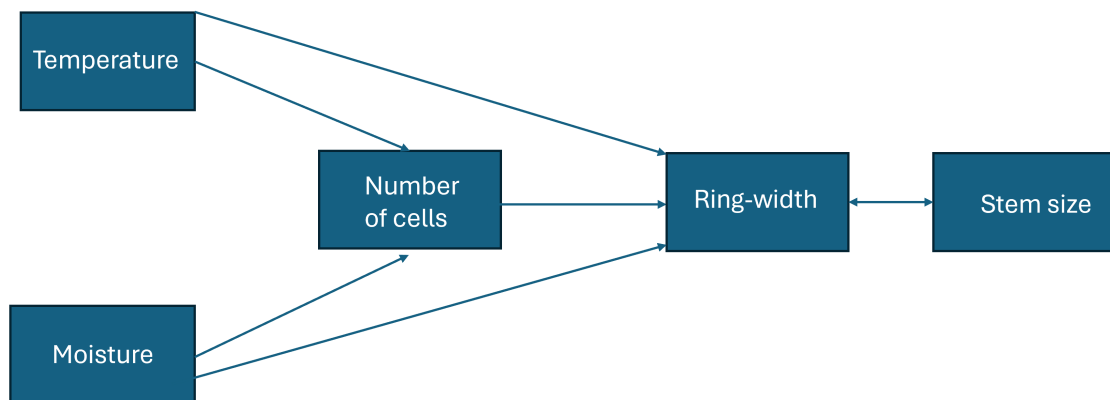


Figure 7: Relations entre les facteurs environnementaux et la croissance: chaque lien est appuyé par des sources de littérature

Sur lavaan, nous pouvons traduire le modèle avec la syntax suivante:

```

require(lavaan)

myModel <- '

# regressions
ring_width ~ temperature + humidity+ cell_number
cell_number ~ temperature + humidity
ring_width ~ ~ stem_size

'

fit <- sem(model = myModel,
           data = simulated_data1)

summary(fit)

## lavaan 0.6.17 ended normally after 32 iterations
##
##      Estimator                      ML
##      Optimization method          NLMINB
##      Number of model parameters          9
##
##      Number of observations          200
##
## Model Test User Model:
##
##      Test statistic                  1.334
##      Degrees of freedom                 3
##      P-value (Chi-square)             0.721
##
## Parameter Estimates:
##
##      Standard errors                  Standard
##      Information                      Expected
##      Information saturated (h1) model  Structured
##
## Regressions:
##              Estimate  Std.Err  z-value  P(>|z|)
##   ring_width ~
##     temperature      0.251    0.230    1.092    0.275
##     humidity         0.127    0.179    0.711    0.477
##     cell_number      0.883    0.190    4.638    0.000
##   cell_number ~
##     temperature      0.742    0.068   10.991    0.000
##     humidity         0.801    0.035   23.050    0.000
##
## Covariances:
##              Estimate  Std.Err  z-value  P(>|z|)
##   .ring_width ~~
##     stem_size       -35.300    9.797   -3.603    0.000
##
## Variances:
##              Estimate  Std.Err  z-value  P(>|z|)
##   .ring_width      166.464   16.646   10.000    0.000

```



```
##      .cell_number      21.391      2.139      10.000      0.000
##      stem_size      107.827      10.783      10.000      0.000
```

Le résumé de la fonction `sem` nous fournit ainsi les coefficients de régression qui indiquent la force et la direction de la relation entre nos variables. Les coefficients de covariance mesurent également la force et la direction de la corrélation entre les variables. On peut interpréter ces coefficients comme n'importe quels coefficients d'une régression linéaire. Cependant, il peut être très utile de standardiser la valeur des coefficients si l'on souhaite comparer l'effet des différentes variables sur notre variable réponse.

## Ajustement du modèle sur Lavaan

On considère qu'un SEM représente bien notre distribution de données et notre phénomène naturel lorsque le p-value (chi carré) du modèle n'est pas significatif. Cela arrive parce que nous souhaitons observer une correspondance entre les données observées et les prédictions de notre modèle, lesquelles ne doivent pas présenter de différences significatives. Dans ce cas, le p-value est de 0.713, ce qui nous rassure quant à la capacité du modèle à représenter notre phénomène. Il est également utile de vérifier d'autres mesures d'ajustement du modèle, pour obtenir une évaluation plus complète de l'adéquation du modèle. Voici les indicateurs plus couramment utilisés. Voici une synthèse présentée par Joreskog, K., & Sorbom, D. (1993):

il est possible de trouver ces métriques d'ajustement pour notre modèle grâce à la fonction: `fitMeasures()`

```
fitMeasures(fit, c("chisq", "df", "pvalue", "cfi", "rmsea"))
```

```
##  chisq      df pvalue      cfi  rmsea
##  1.334  3.000  0.721  1.000  0.000
```

Il existe des standards pour nous naviguer dans la présentation de ces indicateurs:

Il est possible de visualiser les coefficients standardisés pour les relations : très utiles si l'on veut éviter l'effet taille, ainsi que les coefficients de détermination de chaque régression, très importants pour interpréter l'ajustement du modèle sur chaque variable.

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

```
## lavaan 0.6.17 ended normally after 32 iterations
##
##      Estimator                      ML
##      Optimization method          NLMINB
##      Number of model parameters          9
##
##      Number of observations          200
##
## Model Test User Model:
##
##      Test statistic          1.334
##      Degrees of freedom          3
##      P-value (Chi-square)          0.721
##
## Parameter Estimates:
##
##      Standard errors          Standard
##      Information          Expected
##      Information saturated (h1) model          Structured
##
## Regressions:
##
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##      ring_width ~
```

**Table 1: Fit indices and their acceptable thresholds**

Fit Index	Acceptable Threshold Levels	Description
<i>Absolute Fit Indices</i> Chi-Square $\chi^2$	Low $\chi^2$ relative to degrees of freedom with an insignificant $p$ value ( $p > 0.05$ )	
Relative $\chi^2$ ( $\chi^2/\text{df}$ )	2:1 (Tabachnik and Fidell, 2007) 3:1 (Kline, 2005)	Adjusts for sample size.
Root Mean Square Error of Approximation (RMSEA)	Values less than 0.07 (Steiger, 2007)	Has a known distribution. Favours parsimony. Values less than 0.03 represent excellent fit.
GFI	Values greater than 0.95	Scaled between 0 and 1, with higher values indicating better model fit. This statistic should be used with caution.
AGFI	Values greater than 0.95	Adjusts the GFI based on the number of parameters in the model. Values can fall outside the 0-1.0 range.
RMR	Good models have small RMR (Tabachnik and Fidell, 2007)	Residual based. The average squared differences between the residuals of the sample covariances and the residuals of the estimated covariances. Unstandardised.
SRMR	SRMR less than 0.08 (Hu and Bentler, 1999)	Standardised version of the RMR. Easier to interpret due to its standardised nature.
<i>Incremental Fit Indices</i> NFI	Values greater than 0.95	Assesses fit relative to a baseline model which assumes no covariances between the observed variables. Has a tendency to overestimate fit in small samples.
NNFI (TLI)	Values greater than 0.95	Non-normed, values can fall outside the 0-1 range. Favours parsimony. Performs well in simulation studies (Sharma et al, 2005; McDonald and Marsh, 1990)
CFI	Values greater than 0.95	Normed, 0-1 range.

Figure 8: Ajustement du modèle

**Table 2: Hu and Bentler's Two-Index Presentation Strategy (1999)**

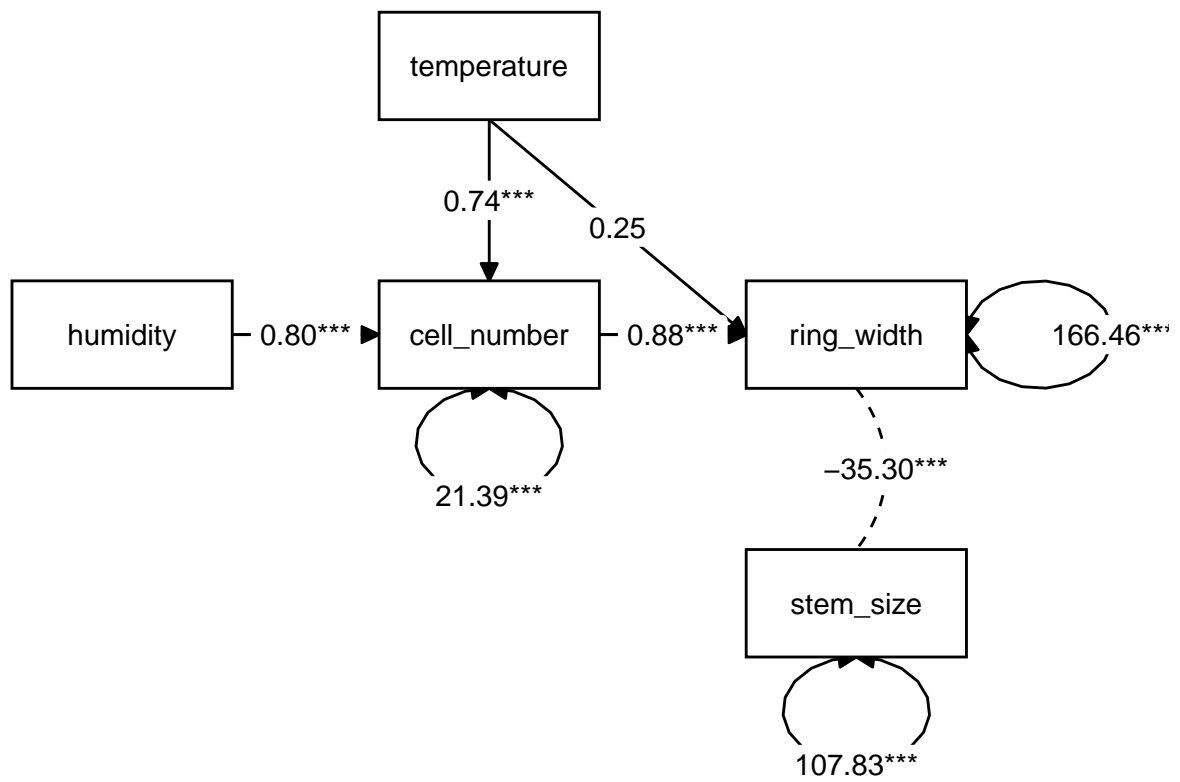
Fit Index Combination	Combinational Rules
NNFI (TLI) and SRMR	NNFI of 0.96 or higher and an SRMR of .09 or lower
RMSEA and SRMR	RMSEA of 0.06 or lower and a SRMR of 0.09 or lower
CFI and SRMR	CFI of .96 or higher and a SRMR of 0.09 or lower

Figure 9: Combinaisons d'indicateurs d'ajustement

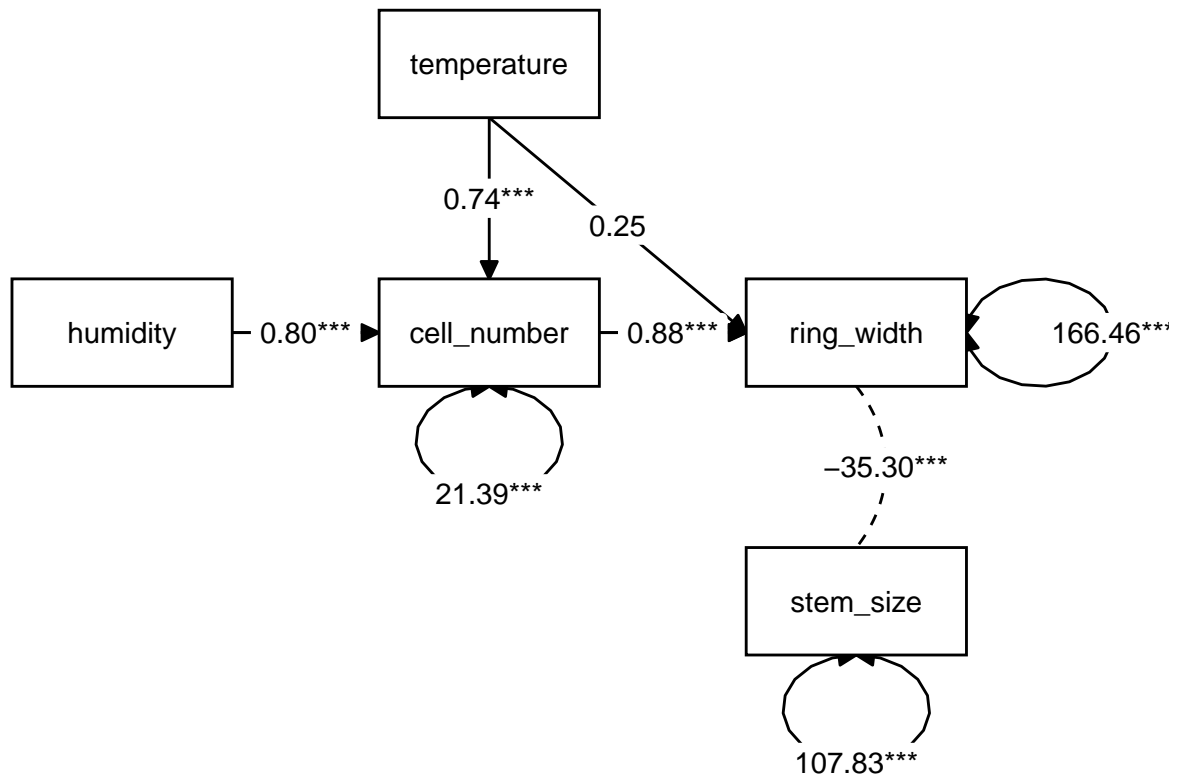
```
##      temperature      0.251    0.230    1.092    0.275    0.251    0.076
##      humidity         0.127    0.179    0.711    0.477    0.127    0.075
##      cell_number      0.883    0.190    4.638    0.000    0.883    0.512
##      cell_number ~
##      temperature      0.742    0.068   10.991    0.000    0.742    0.386
##      humidity         0.801    0.035   23.050    0.000    0.801    0.810
##
## Covariances:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##      .ring_width ~~
##      stem_size     -35.300    9.797   -3.603    0.000  -35.300   -0.263
##
## Variances:
##      Estimate Std.Err z-value P(>|z|) Std.lv Std.all
##      .ring_width     166.464   16.646   10.000    0.000  166.464    0.643
##      .cell_number     21.391    2.139   10.000    0.000   21.391    0.245
##      stem_size       107.827   10.783   10.000    0.000  107.827    1.000
##
## R-Square:
##      Estimate
##      ring_width      0.357
##      cell_number     0.755
```

Avec tidySEM, il est possible d'obtenir une représentation graphique de notre modèle. Dans ce cas, il est important de faire attention à la manière dont nous interprétons les différents éléments car la fonction `graph_sem` utilise son propre code pour identifier les éléments du modèle. Par exemple, la ligne pointillée représente ici des covariances, tandis que le petit cercle avec des flèches représente les variances résiduelles de chaque variable. Vous avez peut-être déjà rencontré ce terme dans le résumé de lavaan avec l'erreur d'estimation et les valeurs p. Des variances significatives indiquent que la variable a une influence significative sur les résultats du modèle

```
require(tidySEM)
graph_sem(fit)
```



```
require(tidySEM)
graph_sem(fit)
```



## Références

Revelle, 2024: How to use the psych package for regression and mediation analysis: <https://cran.r-project.org/web/packages/psychTools/vignettes/mediation.pdf>

Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/> <https://lavaan.ugent.be/>

Joreskog, K., & Sorbom, D. (1993). Structural equation modelling: Guidelines for determining model fit. NY: University Press of America.