An Open-Source WYSIWYG Web Application for Drawing Path Diagrams of Structural Equation Models

Yujiao Mai[1], Ziqian Xu[*1], Zhiyong Zhang[†1], and Ke-Hai Yuan[1]

[1]Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA

Acknowledgments

This research is supported by grants from the Department of Education (R305D140037; R305D210023). We thank Meghan Cain, Han Du, Ge Jiang, Haiyan Liu, Agung Santoso and Miao Yang for their suggestions to improve the software.

[*]The author has equal contribution to the study as the first author.

[†]Corresponding author. Email: zzhang4@nd.edu

An Open-Source WYSIWYG Web Application for Drawing Path Diagrams of Structural Equation Models

Abstract

Structural equation modeling (SEM) is widely used in behavioral, social, and education research. Drawing publication-ready path diagrams for SEM is not a pleasant task with the existing software. In this study, we develop an open-source web-based graphical application, *semdiag*, for drawing WYSIWYG SEM path diagrams interactively. The application is developed using JavaScript and can be used in major web browsers, both online and offline. Several examples are provided to demonstrate how to use the application.

Keywords. structural equation modeling, path diagram, graphical interface, JavaScript

1 Motivation and Significance

Structural equation modeling (SEM) is a statistical technique that can be used to evaluate relations among observed and latent variables (Hoyle, 1995), where an observed variable can be measured directly but a latent variable has to be assessed indirectly using observed variables. The relations between two variables can be non-directional as correlation relationship or directional as regression relationship (Hoyle, 1995; Kline, 2011). SEM generalizes many commonly used statistical models such as confirmatory factor models, path models, latent regression models, and growth curve models (Kline, 2011). Over the past few decades, the use of SEM has been rapidly growing in many disciplines (Kline, 2011; Nachtigall, Kroehne, Funke, & Steyer, 2003; Westland, 2015), especially in education (Khine, 2013), psychology (MacCallum & Austin, 2000), management (Shook, Ketchen, Hult, & Kacmar, 2004), and marketing (Babin, Hair, & Boles, 2008).

A variety of software packages have been developed to conduct SEM analysis including commercial ones such as LISREL (Joreskog & Sorbom, 2022), EQS (Bentler, 2006, 2020), AMOS (Arbuckle, 2021), and Mplus (Muthén & Muthén, 2017, 2021) as well as free ones such as Ωnyx (von Oertzen, Brandmaier, & Tsang, 2015a, 2015b), R packages Lavaan (Rosseel, 2012; Rosseel, Jorgensen, & Rockwood, 2022), OpenMx (Boker et al., 2022), and sem (Fox, Nie, & Byrnes, 2022). Some general purpose statistical software such as SAS and STATA also provides routines to conduct SEM analysis. However, most of the above-mentioned software either does not allow the drawing of a diagram or only produces less-satisfactory diagrams. Some software such as Ωnyx and SEM Builder in STATA can produce high-equality diagrams but is not open-source. Ωnyx is the most promising software for easily drawing path diagrams for SEM, but it requires the JAVA runtime environment to use. Additionally, although the open-source R package semPlot (Epskamp, 2015, 2022) provides a systematic way to plot path diagrams and is compatible with syntax from other software, the diagram customization processes and options may not be the most satisfactory. The R package DiagrammeR could also be utilized for interactively constructing path diagrams with

many customization possibilities, but it may not be the most user-friendly given that the package is not designed solely for SEMs (Iannone, 2022; Mahr, 2015).

Therefore, the purpose of this study is to develop and provide an easy-to-use open-source tool, *semdiag*, for drawing SEM path diagrams. The tool is developed as a web application with the following features.

1. The application can be readily used within a web browser both online and offline, and therefore does not require the installation of special software.

2. The application allows creation of new path diagrams and modification of existing path diagrams through an interactive graphical interface, and the diagram is WYSIWYG (what you see is what you get).

3. The application allows saving path diagrams to a local computer and loading existing path diagrams from the local computer.

4. The application eases the creation of diagrams by simplifying the drawing of the needed nodes and arrows.

5. The diagrams can be saved into SVG, PNG, PDF, and HTML formats and can be readily used in academic publications.

6. The application is free and open-source, and accepts contribution from others.

7. The application allows the use and display of mathematical symbols and formulas through LaTeX.

2 Software Framework

2.1 Framework of SEM Path Diagram

2.1.1 Components of SEM path diagram

A SEM path diagram consists of two types of graphical components: variables and paths. The variables are also called vertices/nodes and the paths are also called edges in graph theory.

Variables. Three types of variables are allowed in a SEM path diagram: observed, latent and constant variables. An observed variable is represented by a rectangle, a latent variable is represented by an ellipse, and a constant variable is represented by a triangle. The constant variable is a special variable with the value of 1. It is used to represent a mean or an intercept in the model. We will call the nodes by their respective geometric names (i.e., ellipse, triangle, and rectangle nodes) in this paper.

Paths. Two types of paths, a directed one and an undirected one, are allowed to define the relationship of the variables in a model or diagram. A directed path, represented by a single-headed arrow (directed arrow), is used to indicate that one variable can predict another variable. In statistical language, it typically means a regression relationship. A directed path points from the predicting variable toward the outcome variable. More specifically in SEM, a directed path from an ellipse

(a latent variable) to a rectangle (an observed variable) represents a factor loading; otherwise, it represents a regression coefficient.

An undirected path is represented by a double-headed arrow (undirected arrow). If an undirected path connects two variables, it is the covariance between the two variables. An undirected path can start from one variable and end on the same variable. Such a path can represent either a variance or a residual variance in a structural equation model. For a variable with such a path, if there is no directed path pointing to it, the path represents a variance, otherwise, a residual variance. Similar to the nodes, we will also call paths by their geometric names (i.e., directed and undirected arrows) in this paper.

Both nodes and arrows can be labeled or named. The labels can be as meaningful as possible. For arrows, numbers can be used to denote the values of coefficients in a model.

2.1.2 Basic rules for drawing SEM path diagram

To ensure consistency and improve usability, the following rules are imposed by the application in constructing path diagrams. First, no more than one triangle node can be created in one diagram. Second, if existing, a directed arrow can only be drawn starting from the triangle node. Third, either a directed or an undirected arrow can be drawn between two rectangle and ellipse nodes. Fourth, an undirected arrow is automatically created for a node that is predicted by one or more nodes. The arrow is also automatically removed when the involved node loses its predicting nodes. Finally, when a node is removed, all arrows connecting to it will be removed automatically.

2.2 Software Architecture

The interface is designed as an interactive web application. It consists of two modules. First, the nodes and arrows of the path diagrams are defined in the scalable vector graphics (SVG) format (Ferraiolo, Jun, & Jackson, 2000). Second, the client-side behavior is implemented using JavaScript to interact with the SVG graphs. Specifically, the library D3.js (Bostock, 2013; Myatt & Johnson, 2011) is applied to realize the data-driven SVG diagramming while the library jQuery (Resig et al., 2012) is used to create the dynamic dialogs. The file operations for saving and exporting diagrams are also completed with JavaScript. In particular, conversions with PDFs are done through the jsPDF and svg2pdf libraries (Hall, 2020; yWorks, 2019). Further, the software supports LaTeX notations via the use of the MathJax library (American Mathematical Society, 2017; Cervone, 2012). As a web application, the interface can be used within a web browser on a computer, a tablet, a phone, or other devices that support modern web browsers. The web application is best supported by the Chrome web browser but also works in Microsoft Edge, Firefox, and Safari although features such as selecting multiple elements in the web application and saving the diagram to desired output formats can behave differently depending on the browsers used.

2.3 Software Functionalities

The major functionalities of the interface include creating, editing, saving, and loading path diagrams, which are implemented through different buttons. A brief introduction of the functionalities is included in Table 1 (see also the left panel of Figure 1).

2.3.1 Create a new diagram

A rectangle node or an ellipse node can be created by clicking their corresponding buttons in the interface. A new arrow can be drawn by first selecting the start node, then clicking the arrow button ("Directed" or "Undirected" button), and finally clicking the end node. Another way to create an arrow is to first select the arrow, and then drag from the start node to the end node. Both the directed and undirected arrows can be drawn in these ways. For the arrows representing self links, the start node and the end node are the same.

To create a triangle node, one needs to first select a rectangle or ellipse node and then click on the "Triangle" button. A directed arrow will also be added automatically from the triangle node to the node selected. There can only be one triangle node per diagram. However, additional arrows can be added between the triangle node and other ellipse or rectangle nodes.

Nodes and arrows can be duplicated easily. To duplicate one node, simply select it by clicking it and then click the "Clone" button. To copy and paste multiple nodes, one can select all the nodes as well as arrows then click the "Clone" button. There are two ways to select multiple elements in the diagram: One can directly drag over multiple elements or one can hold the "control" key and click on each element to be selected.

Both the nodes and arrows can be named. When first created, the nodes are labeled with default names whereas the arrows are unnamed. One can change them in a pop-up dialog by double clicking a node or an arrow.

A text can be added anywhere in the diagram. To add a text, first click the "Text" button and then click on the canvas where you want to put the text. Mathematical equations can be used in the text using LaTeX notations. For example, `$\alpha$` will be rendered as α after clicking the "Render" button.

2.3.2 Edit an existing diagram

A node within a path diagram can be moved freely by dragging it. To move multiple nodes and arrows, one can select the target elements and then drag them to a desired location. Nodes can also be moved using the "Move" buttons. The properties of a selected node or arrow can be modified by clicking the properties buttons such as "Color", "Font Size", and "Line Width." To delete a node or arrow, simply select it and then click the "Delete" button. The size of nodes can be adjusted using the "Width" and "Height" buttons. Labels on arrows can be turned on or off using the "Show Label" and "Hide Label" buttons. Line types of arrows and nodes can be toggled using the "Dotted" and "Solid" buttons.

If multiple nodes are selected, one can also align them using the "Left", "Right", "Top", and "Bottom" buttons. The "Expand Width" button can be used to expand width of a group of selected nodes to be the same as the node with the largest width. The "Shrink Width", "Expand Height", and "Shrink Height" buttons work similarly.

After clicking on a variance or residual variance arrow, its location relative to the nodes can be rotated using the "Rotate Variance" buttons. The "Variance Size" button can be used to change sizes of the variance and residual variance arrows.

2.3.3 Save and load a diagram

By clicking the "Save" button, a path diagram will be saved in a file on the local computer as a text file. The extension name of the file is ".diag". The file can then be loaded for further edits by clicking the "Open" button.

2.3.4 Generate publication-ready path diagrams

The path diagram can be exported to a SVG graph, a PNG, an HTML, or a PDF using the "SVG", "PNG", "HTML", or "PDF" button. A dialog box is prompted for customizing the desired file name. Saving LaTeX notations is currently only supported when saving the path diagrams to HTML files. If one needs to save the path diagram with mathematical symbols or equations to PDF, we recommend first saving the diagram into HTML and then printing the HTML to PDF.

3 Illustrative Examples

To illustrate how to use the web application for drawing path diagrams, we provide several examples here. Four path diagrams for popular SEM models are provided here. A Youtube video tutorial is available showing how to draw a path diagram in action (`https://youtu.be/rSc4nbpWRh4`).

3.1 A Factor Model

Figure 1 shows the path diagram for a factor model. There are two latent factors (variables) as portrayed by the two ellipse nodes labeled as $F1$ and $F2$. The 6 rectangle nodes represent the observed variables and are labeled as $X1$ to $X6$, respectively. The directed arrows show the connection (factor loading) between the factors and the observed variables. Through the diagram, it is clear that $F1$ is measured by $X1$, $X2$, and $X3$, while $F2$ is measured by $X4$, $X5$, and $X6$. In editing labels in this model, LaTeX is also utilized. Labels can be added to arrows and nodes. The arrows such as the correlation between $F1$ and $F2$ can also be curved. Note that at the bottom left of the software panel, the corresponding start node name, type of element, whether the element is an arrow, end node name, and label given to the element are displayed for each element of the path diagram. This can be used in other SEM software for connecting our web application to the model estimation methods. All the labels can be hidden by clicking on the "Hide Label" button.

3.2 A Mediation Model

Figure 2 depicts a mediation model where the observed variable M in the rectangle node is a mediator. Both the input variable $F1$ and the outcome variable $F2$ are latent variables. In the diagram, a, b, and c represent the direct effects from $F1$ to M, M to $F2$, and $F1$ to $F2$, respectively. These labels can be used to create new parameters or calculate new effects. For example, the indirect effect from $F1$ to $F2$ can be constructed as $a \times b$, the multiplication of a and b. Arrows such as the direct effects among $F1$, $F2$, and M can be colored red, black, green, or blue.

3.3 A Latent Growth Model

Figure 3 is the path diagram of a latent growth curve model. In the model shown, the error variances on the outcome variables are constrained to be the same. Note that the label for an arrow can be a number. This typically means that the path takes a fixed value in the model. If the same labels are used for multiple arrows, it means the paths take the same value. In addition, different colors can be applied to nodes.

3.4 A Multiple Group Model

Figure 4 shows the path diagram of a two group mediation model. For each of the two groups, predictor X, mediator M, and outcome Y are considered. Whereas the mediation effects a and b are constrained to be the same across groups 1 and 2, the direct effects c_1 and c_2 are different across the two groups.

4 Conclusions

In this study, we developed a web application for drawing SEM path diagrams. We showed how to use the web application through its interactive interface. Using the web application, a researcher can easily draw publication-ready path diagrams. Currently, the web application does not provide statistical analysis, and several functionalities behave differently in different browsers. In the future, we hope to connect the web application with existing SEM software for statistical data analysis and upgrade the compatibility across browsers. More diagram customization options will also be added in the web application.

5 Software Information

The web application can be accessed via `https://johnnyzhz.github.io/semdiag/`. The source code can be downloaded via the GitHub repository `https://github.com/johnnyzhz/semdiag`. Suggestions for improvements are welcomed.

References

American Mathematical Society. (2017). *Mathjax (version 2.7.1).* Retrieved from `https://www.mathjax.org/`

Arbuckle, J. L. (2021). *Amos (version 26.0).* Chicago, IL: SPSS. Retrieved from https://www.ibm.com/products/structural-equation-modeling-sem

Babin, B. J., Hair, J. F., & Boles, J. S. (2008). Publishing research in marketing journals using structural equation modeling. *The Journal of Marketing Theory and Practice*, *16*(4), 279–286.

Bentler, P. M. (2006). Eqs 6 structural equations program manual.

Bentler, P. M. (2020). *Eqs (version 6.4).* Encino: Multivariate Software, Inc. Retrieved from https://mvsoft.com/eqs/

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., & Fox, J. (2022). *Openmx: Extended structural equation modelling (version 2.20.6).* Retrieved from https://openmx.ssri.psu.edu/

Bostock, M. (2013). *D3.js - data-driven documents (version 3).* Retrieved from http://d3js.org/

Cervone, D. (2012). Mathjax: a platform for mathematics on the web. *Notices of the AMS*, *59*(2), 312–316.

Epskamp, S. (2015). semplot: Unified visualizations of structural equation models. *Structural Equation Modeling: a multidisciplinary journal*, *22*(3), 474–483.

Epskamp, S. (2022). *semplot: Path diagrams and visual analysis of various sem packages' output (version 1.1.5).* Retrieved from https://cran.r-project.org/web/packages/semPlot/index.html

Ferraiolo, J., Jun, F., & Jackson, D. (2000). *Scalable vector graphics (svg) 1.0 specification.* iuniverse.

Fox, J., Nie, Z., & Byrnes, J. (2022). *sem: Structural equation models (version 3.1-15).* Retrieved from http://CRAN.R-project.org/package=sem

Hall, J. (2020). *jspdf: Client-side javascript pdf generation for everyone (version 2.1.1).* Retrieved from https://github.com/parallax/jsPDF

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications.* Sage Publications.

Iannone, R. (2022). *Diagrammer: Graph/network visualization (version 1.0.9).* Retrieved from https://cran.r-project.org/web/packages/DiagrammeR/index.html

Joreskog, K. G., & Sorbom, D. (2022). *Lisrel (version 11.0.3.2).* Chicago, IL: Scientific Software International. Retrieved from https://ssicentral.com/index.php/products/lisrel/

Khine, M. S. (2013). *Application of structural equation modeling in educational research and practice.* Springer.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual review of psychology*, *51*(1), 201–226.

Mahr, T. (2015). *Sem path diagrams with diagrammer.* Retrieved from https://rpubs.com/tjmahr/sem_diagrammer

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

Muthén, L. K., & Muthén, B. O. (2021). *Mplus (version 8.7).* Los Angeles, CA: CA: Muthén and Muthén. Retrieved from `http://www.statmodel.com/`

Myatt, G. J., & Johnson, W. P. (2011). *Making sense of data iii: A practical guide to designing interactive data visualizations.* John Wiley & Sons.

Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). should we use sem? pros and cons of structural equation modeling. *Methods of Psychological Research Online*, *8*(2), 1–22.

Resig, J., et al. (2012). *jquery (version 1.7.2).* Retrieved from `http://jquery.com`

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rosseel, Y., Jorgensen, T. D., & Rockwood, N. (2022). *lavaan: Latent variable analysis (version 0.6-11).* Retrieved from `https://cran.r-project.org/web/packages/lavaan/index.html`

Shook, C. L., Ketchen, D. J., Hult, G. T. M., & Kacmar, K. M. (2004). An assessment of the use of structural equation modeling in strategic management research. *Strategic management journal*, *25*(4), 397–404.

von Oertzen, T., Brandmaier, A. M., & Tsang, S. (2015a). *ωnyx (version 1.0).* Retrieved from `https://onyx-sem.com/`

von Oertzen, T., Brandmaier, A. M., & Tsang, S. (2015b). Structural equation modeling with ωnyx. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 148–161.

Westland, J. C. (2015). *Structural equation models: From paths to networks* (Vol. 22). Springer.

yWorks. (2019). *svg2pdf.js (version 1.5.0).* Retrieved from `https://github.com/yWorks/svg2pdf.js`

Table 1: Buttons and their functionalities in the interface.

Name	Button	Functionality
Open		Open a path diagram from a .diag file
New		Start a new path diagram
Save		Save a path diagram to a .diag file
SVG		Save a path diagram to a .svg file
PNG		Save a path diagram to a .png file
PDF		Save a path diagram to a .pdf file
HTML		Save a path diagram to a .html file
Render		Render LaTeX in a diagram
Rectangle		Draw a rectangle node
Ellipse		Draw an ellipse node
Triangle		Draw a triangle node
Directed		Draw a directed arrow
Undirected		Draw an undirected arrow
Text		Add or edit texts or labels
Delete		Delete elements
Clone		Clone nodes
Black, Red, Blue, Green		Color nodes or arrows
Dotted		Change solid lines to dotted
Solid		Change dotted lines to solid
Hide Label, Show Label		Hide or show labels
Grid		Show or hide assistance grids of the canvas
Left, Right, Top, Bottom		Align nodes
Expand Width		Expand width to be the same as the largest node
Shrink Width		Shrink width to be the same as the smallest node
Expand Height		Expand height to be the same as the largest node
Shrink Height		Shrink height to be the same as the smallest node
Height		Adjust height
Width		Adjust width
Font Size		Adjust font size
Line Width		Adjust line width
Variance Size		Adjust variance arrow size
Rotate Variance		Rotate variance arrow direction
Move		Move elements

Figure 1: A factor model.

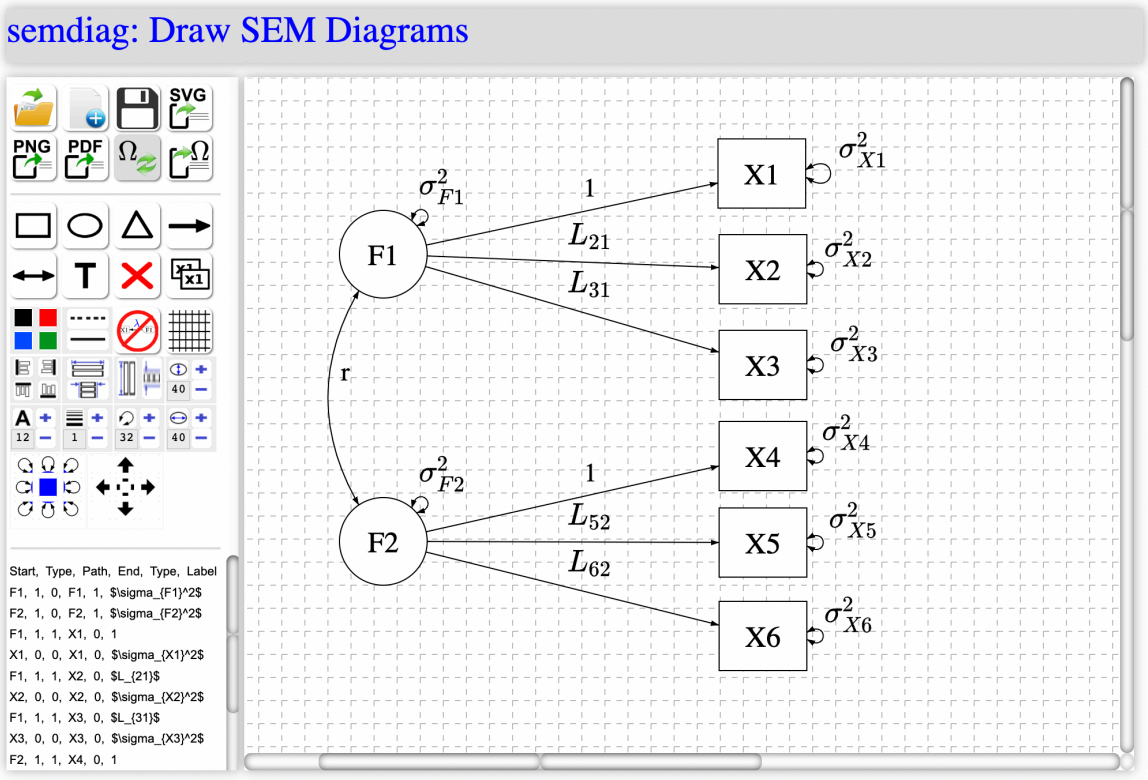

Figure 2: A mediation model.

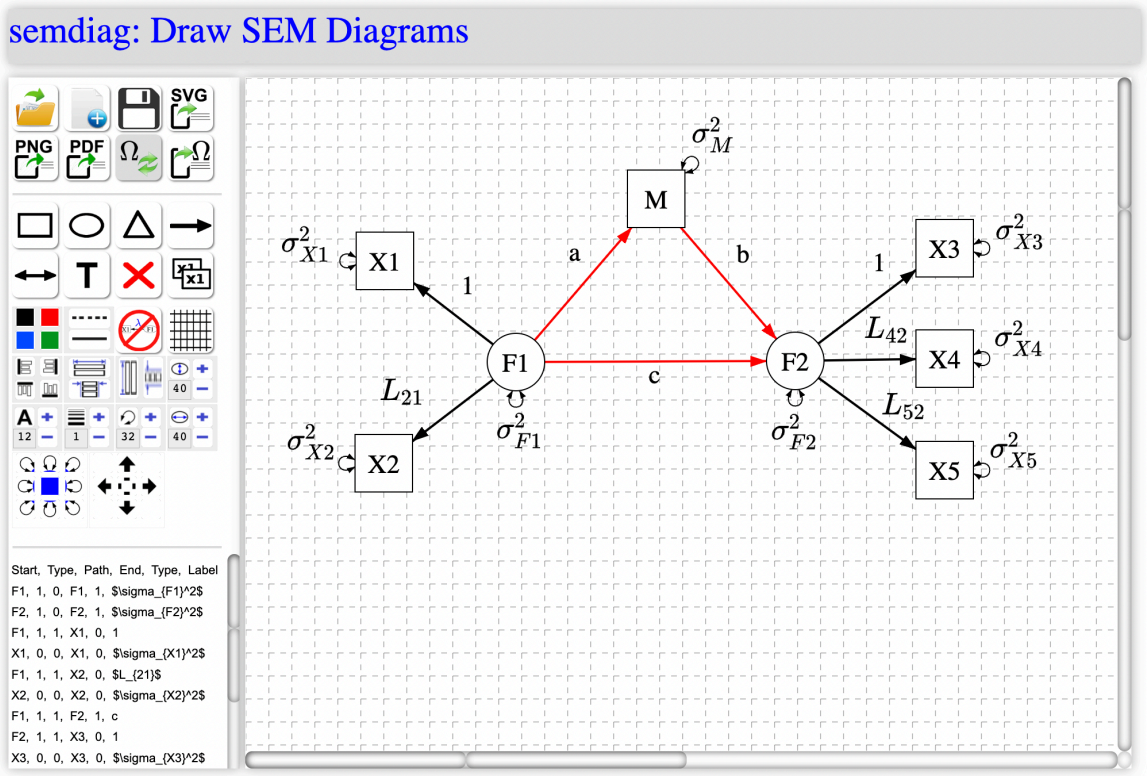

Figure 3: A latent growth model.

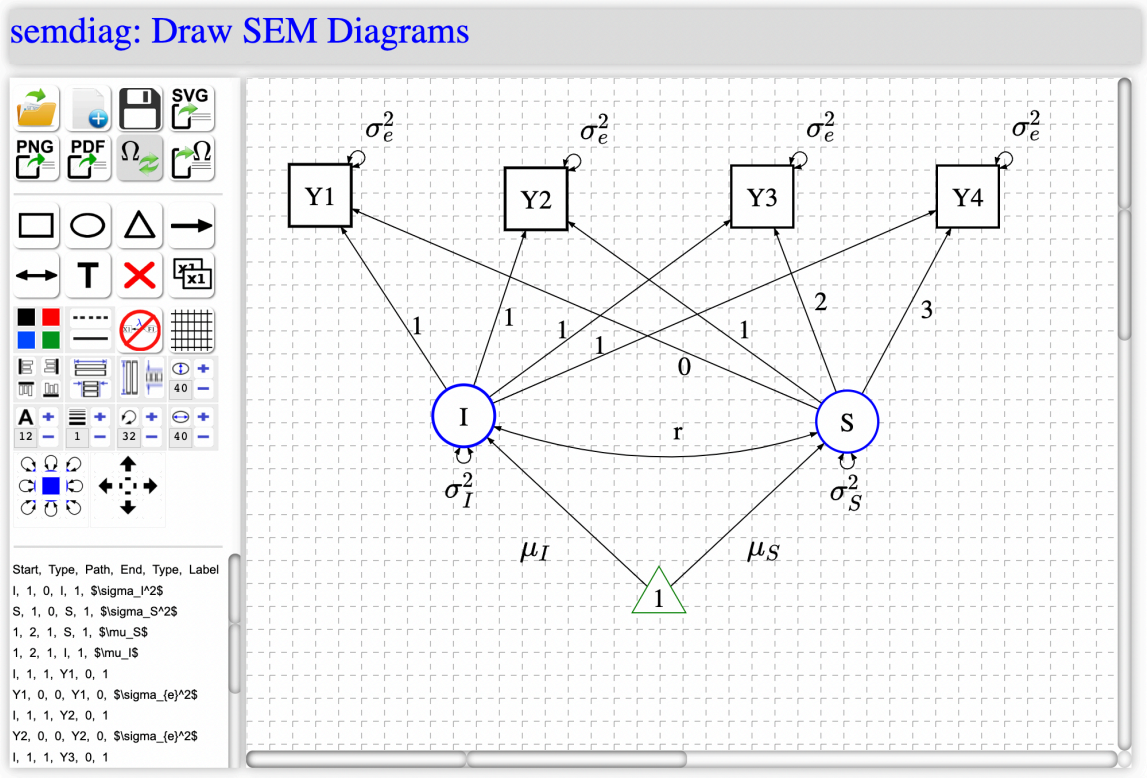

Figure 4: A multiple group model.