14.1 Path Analysis

Structural Equation Modeling (SEM) is a collection of confirmatory statistical methods for pre-specified relations among variables. Most common SEM methods are path analysis, mediation analysis, and confirmatory factor analysis.

A **path analysis** is a method of studying direct and indirect effects of independent variables on a single dependent variable. The analysis essentially involves running several general linear regressions, in which the response variables should be normally distributed (or at least measured on an interval scale).

Prior to conducting a path analysis, researchers should present hypothesized relationships among variables in a graphical way termed an **input path diagram**. In a path diagram, researchers use arrows to show how different variables relate to each other. An arrow pointing from, say, Variable A to Variable B, shows that Variable A is hypothesized to influence Variable B. Variables that influence other variables are called **exogenous variables**. Variables that are influenced by other variables are referred to as **endogenous variables**. In path diagrams, variables with arrows pointing out are exogenous, and those with arrows pointing in are endogenous.

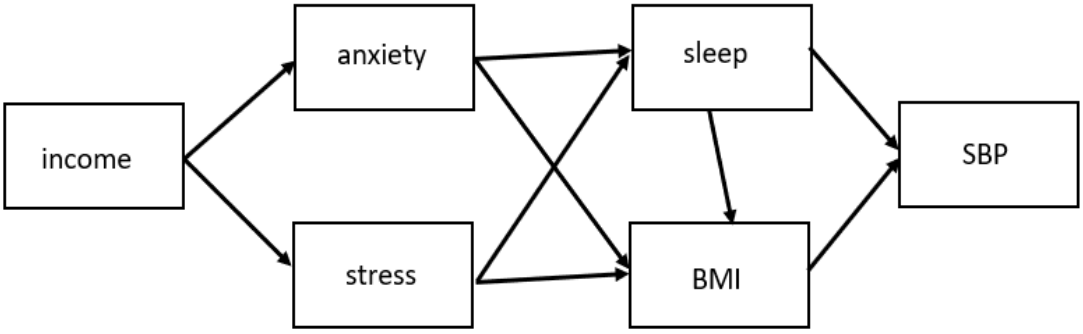
After the statistical analysis has been completed, a researcher would then construct an **output path diagram**, which illustrates the relationships as they actually exist, according to the analysis conducted. If the researcher's hypothesis is correct, the input path diagram and output path diagram will show the same relationships between variables.

There are several main requirements for path analysis:

- The causal flow is one-way, that is, all causal relationships between variables must go in one direction only (two variables cannot cause each other).
- All relations are linear and additive.
- Endogenous variables should be continuous. In case of ordinal data, the minimum number of categories should be five.

While path analysis is useful for evaluating causal hypotheses, this method cannot determine the direction of causality. It clarifies correlation and indicates the strength of a causal hypothesis, but does not prove direction of causation.

Example. Hypertension (systolic blood pressure above 140 mm Hg) is a major health problem worldwide due to the high prevalence rate and the number is associated with an increased risk of cardiovascular disease. Data were collected on family income (categories numbered 1 through 4), level of stress (10-point scale with 10 being the highest), level of anxiety (10-point scale with 10 being the highest), sleep quality (0 through 4 with 4 being the worst quality), body mass index (BMI, in kg/m^2), and systolic blood pressure (SBP, in mmHg). The data are given in the file "hypertension_data.csv". Researchers are interested in establishing causal relationships among the measured variables. They hypothesize that income influences anxiety and stress levels, which, in turn influence sleep and BMI. Sleep affects BMI, and both sleep and BMI are risk factors for elevated systolic blood pressure. The researchers sketch the input path diagram like the one in the picture.



Note that in this diagram, there are 5 exogenous variables (income, anxiety, stress, sleep, and BMI), and 5 endogenous variables (anxiety, stress, sleep, BMI, and SBP).

To verify the hypothesized path structure, a path analysis is conducted.

In SAS:

```
proc import out=hypertension datafile="./hypertension_data.csv"
dbms=csv replace;
run;

/*PROC CALIS=Covariance Analysis and Linear Structural Equations*/
proc calis plots=pathdiagram;
lineqs
```

```
anxiety = b1 income + e1,
stress = b2 income + e2,
sleep = b3 anxiety + b4 stress + e3,
BMI = b5 anxiety + b6 stress + b7 sleep + e4,
SBP = b8 sleep + b9 BMI + e5;
run;
```

Here b's stand for regression coefficients and e's denote random errors.

Linear Equations

anxiety	=		-0.9589	(**)	income	+	1.0000		e1							
stress	=		-0.8718	(**)	income	+	1.0000		e2							
sleep	=		0.1967	(**)	anxiety	+	0.2565	(**)	stress	+	1.0000		e3			
BMI	=		0.1282	(ns)	anxiety	+	0.4335	(**)	stress	+	3.2736	(**)	sleep	+	1.0000	e4
SBP	=		6.7650	(**)	sleep	+	-0.1365	(ns)	BMI	+	1.0000		e5			

Effects in Linear Equations

Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
anxiety	income	b1	-0.95888	0.22440	-4.2731	<.0001
stress	income	b2	-0.87183	0.22789	-3.8256	0.0001
sleep	anxiety	b3	0.19669	0.04122	4.7721	<.0001
sleep	stress	b4	0.25653	0.04132	6.2077	<.0001
BMI	anxiety	b5	0.12820	0.19812	0.6471	0.5176
BMI	stress	b6	0.43346	0.21279	2.0371	0.0416
BMI	sleep	b7	3.27358	0.46517	7.0374	<.0001
SBP	sleep	b8	6.76502	1.13716	5.9490	<.0001
SBP	BMI	b9	-0.13654	0.22647	-0.6029	0.5466

Estimates for Variances of Exogenous Variables						
Variable Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Observed	income	_Add1	0.87199	0.13455	6.4807	<.0001
Error	e1	_Add2	3.68844	0.56914	6.4807	<.0001
	e2	_Add3	3.80416	0.58699	6.4807	<.0001
	e3	_Add4	0.62379	0.09625	6.4807	<.0001
	e4	_Add5	11.33801	1.74949	6.4807	<.0001
	e5	_Add6	51.33736	7.92153	6.4807	<.0001

From the output, the effect of anxiety on BMI, and that of BMI on SBP are not statistically significant at the 5% level.

Further, since these coefficients are unstandardized, we cannot judge relative importance of each factor. SAS also outputs standardized coefficients and standardized variances of the error terms.

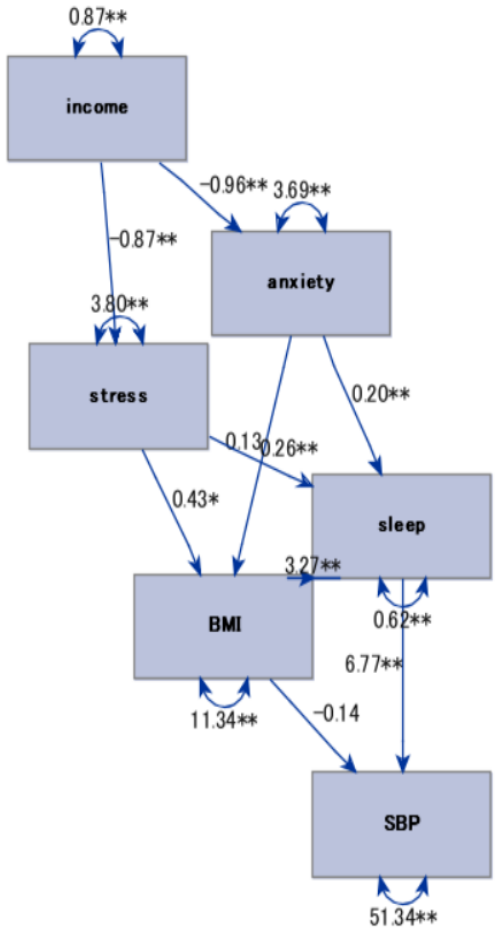
Standardized Effects in Linear Equations						
Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
anxiety	income	b1	-0.42256	0.08963	-4.7147	<.0001
stress	income	b2	-0.38520	0.09292	-4.1455	<.0001
sleep	anxiety	b3	0.38615	0.07737	4.9908	<.0001
sleep	stress	b4	0.50231	0.07278	6.9022	<.0001
BMI	anxiety	b5	0.05012	0.07747	0.6470	0.5176
BMI	stress	b6	0.16903	0.08280	2.0414	0.0412
BMI	sleep	b7	0.65194	0.08199	7.9518	<.0001
SBP	sleep	b8	0.74189	0.10974	6.7603	<.0001
SBP	BMI	b9	-0.07519	0.12462	-0.6033	0.5463

Standardized Results for Variances of Exogenous Variables						
Variable Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Observed	income	_Add1	1.00000			
Error	e1	_Add2	0.82144	0.07575	10.8448	<.0001
	e2	_Add3	0.85162	0.07158	11.8968	<.0001
	e3	_Add4	0.53543	0.07735	6.9223	<.0001
	e4	_Add5	0.38599	0.06515	5.9245	<.0001
	e5	_Add6	0.52995	0.07882	6.7233	<.0001

From this result, if we put the standardized coefficients in descending order, we can see that the highest effect sleep has on SBP, then sleep on BMI, then stress on sleep.

An output path diagram graphically summarizes the findings. Here the unstandardized estimates of the coefficients are placed on each arrow between

two variables, and the unstandardized estimated variances are depicted on self-arrows for each variable.



In R:

```
hypertension.data<- read.csv("./hypertension_data.csv", header=TRUE, sep=",")

library(lavaan)#lavaan=latent variable analysis

path<- 'anxiety ~ income
stress ~ income
sleep ~ anxiety + stress
BMI ~ anxiety + stress + sleep
SBP ~ sleep + BMI'
```

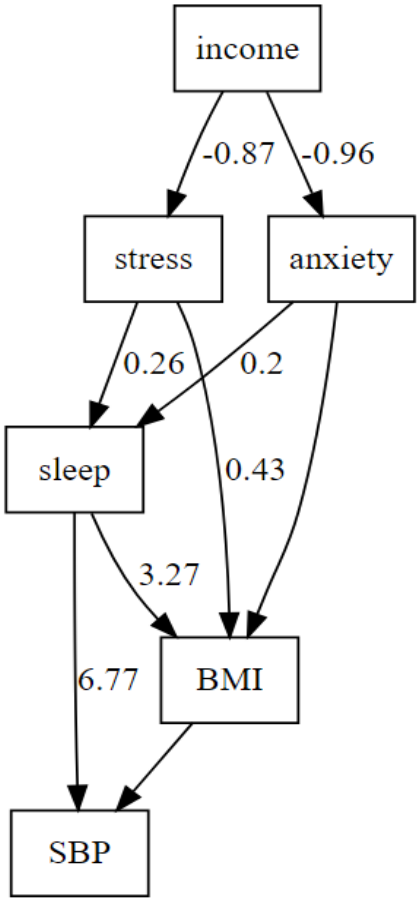
```
fitted.path<- sem(path,hypertension.data)
summary(fitted.path)

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  anxiety ~
    income           -0.959    0.223   -4.298    0.000
  stress ~
    income           -0.872    0.227   -3.848    0.000
  sleep ~
    anxiety           0.197    0.041    4.800    0.000
    stress            0.257    0.041    6.245    0.000
  BMI ~
    anxiety           0.128    0.197    0.651    0.515
    stress            0.433    0.212    2.049    0.040
    sleep             3.274    0.462    7.079    0.000
  SBP ~
    sleep             6.765    1.130    5.984    0.000
    BMI              -0.137    0.225   -0.606    0.544

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .anxiety           3.645    0.559    6.519    0.000
   .stress            3.759    0.577    6.519    0.000
   .sleep             0.616    0.095    6.519    0.000
   .BMI              11.205    1.719    6.519    0.000
   .SBP              50.733    7.782    6.519    0.000

library(lavaanPlot)

lavaanPlot(model = fitted.path, coefs = TRUE, stand = FALSE, sig = 0.05)
```


income
-0.87
-0.96
stress
anxiety
0.26
0.2
sleep
0.43
3.27
6.77
BMI
SBP

□