PSYC40005 - 2018
ADVANCED DESIGN AND DATA ANALYSIS

Lecture 6:
Structural equation modelling 2:
Path analysis

Geoff Saw

Melbourne School of Psychological Sciences
University of Melbourne

The agenda for this lecture

1. Basic ideas in Structural Equation Modeling (SEM)
2. Multi-step regressions
3. Model fit
4. Path models with latent variables
5. Bootstrapping

GOALS OF THIS LECTURE

- To present SEM as a combination of regression and CFA
- To show how multi-stage regressions can be done simultaneously in an SEM framework
- To introduce measures of model fit
- To illustrate how to fit a latent variable model in AMOS
- To introduce Bootstrapping methods and to show how they can be conducted in AMOS

Section 1

BASIC IDEAS IN STRUCTURAL EQUATION MODELING (SEM)

Models & Equations

The regression model:

$$y_p = \beta_{p1} x_1 + \beta_{p2} x_2 + \dots \beta_{pk} x_k + e_p \qquad [1]$$

or in matrix form

$$\boldsymbol{y} = \boldsymbol{\beta x} + \boldsymbol{e}$$

where we have a dependent variable *y* being predicted by a series of independent variables ***x***.

The factor model:

An observed variable *x* is predicted by a series of unobserved factors [latent variables] ***f***.

$$x_p = \lambda_{p1} f_1 + \lambda_{p2} f_2 + \dots \lambda_{pk} f_k + u_p \qquad [2]$$

or in matrix form

$$\boldsymbol{x} = \lambda \boldsymbol{f} + \boldsymbol{u}$$

Combining these models

- Notice however that we have the observed variable *x* in both equations.
- We could build a composite by substituting for *x* in [1] from *x* in [2] and carry out the kind of predictive modelling we do in the regression model with the latent variables involved in equation [2].

An advantage of latent variable modelling

- Models the characteristic of interest, say depression, rather than scores on a depression test.
- How do the two differ?

Test theory: O = T + E

Observed score O, true score T, error E

Observed score O is *x* in equation 1 or 2

True score T is latent variable *f* in equation 2

Reliability is the proportion of observed score variance that is true score variance

$$\alpha = \sigma_T^2 / \sigma_O^2$$

Reliability never equals 1.0

Therefore our observed score never fully captures the information in the latent variable

Structural Equation Modelling

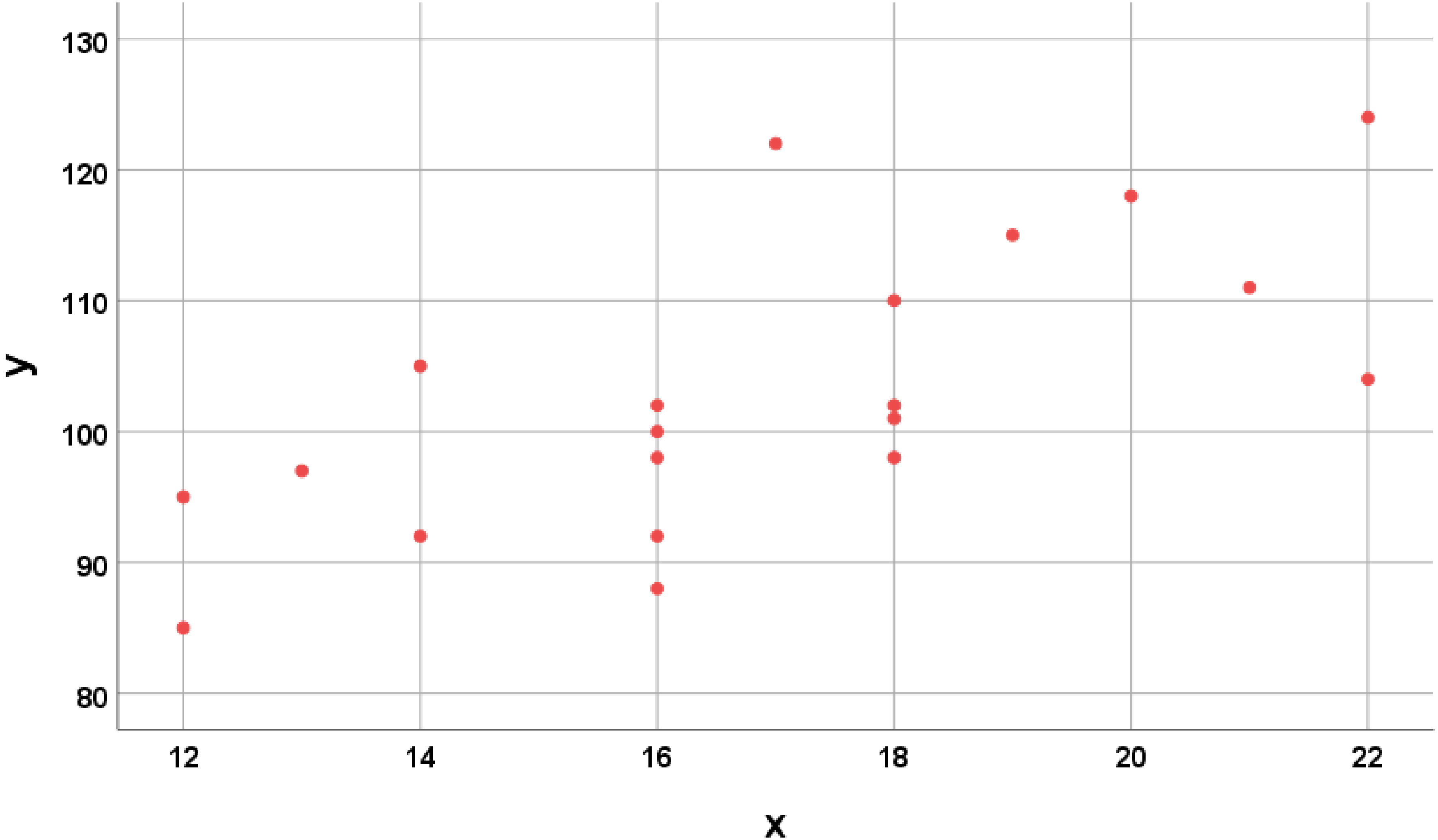
- The modelling of data by jointly by equations [1] and [2] is known as structural equation modelling.
- That aspect of the model concerned with equation [2] is often called the *measurement* model, and
- that part focussing on equation [1] is known as the *structural* model. It is what is known in regression terms, *path analysis.*

Correlation & Regression

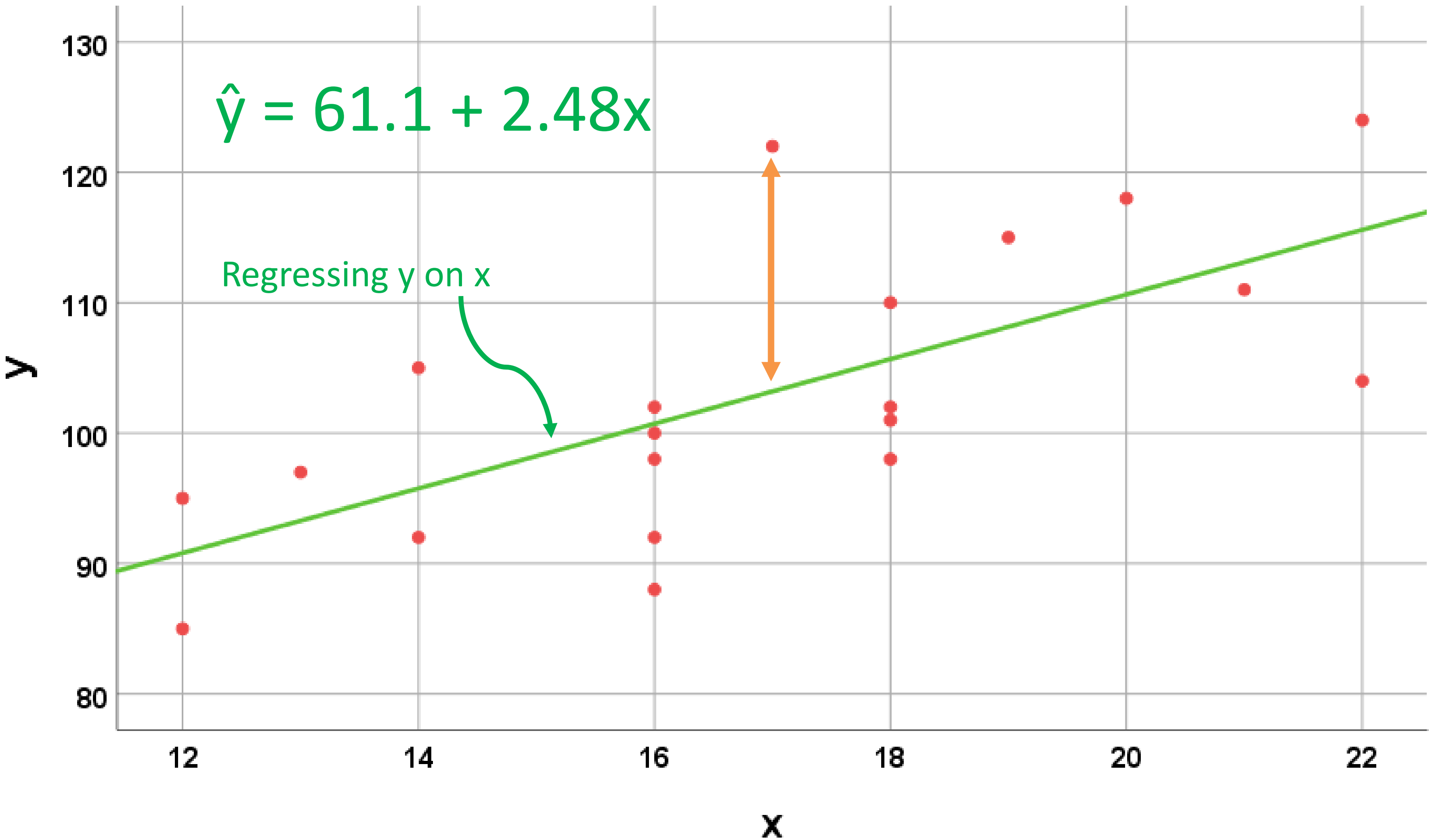
- We have always taught you that “correlation is not causation”
- A correlation coefficient is a symmetric measure, i.e. it doesn’t distinguish between predicting *a* from *b* and *b* from *a*.
- But regression does predict.
  - although this does not necessarily mean “cause”
- And regression & correlation are (almost) the same thing

Some example data

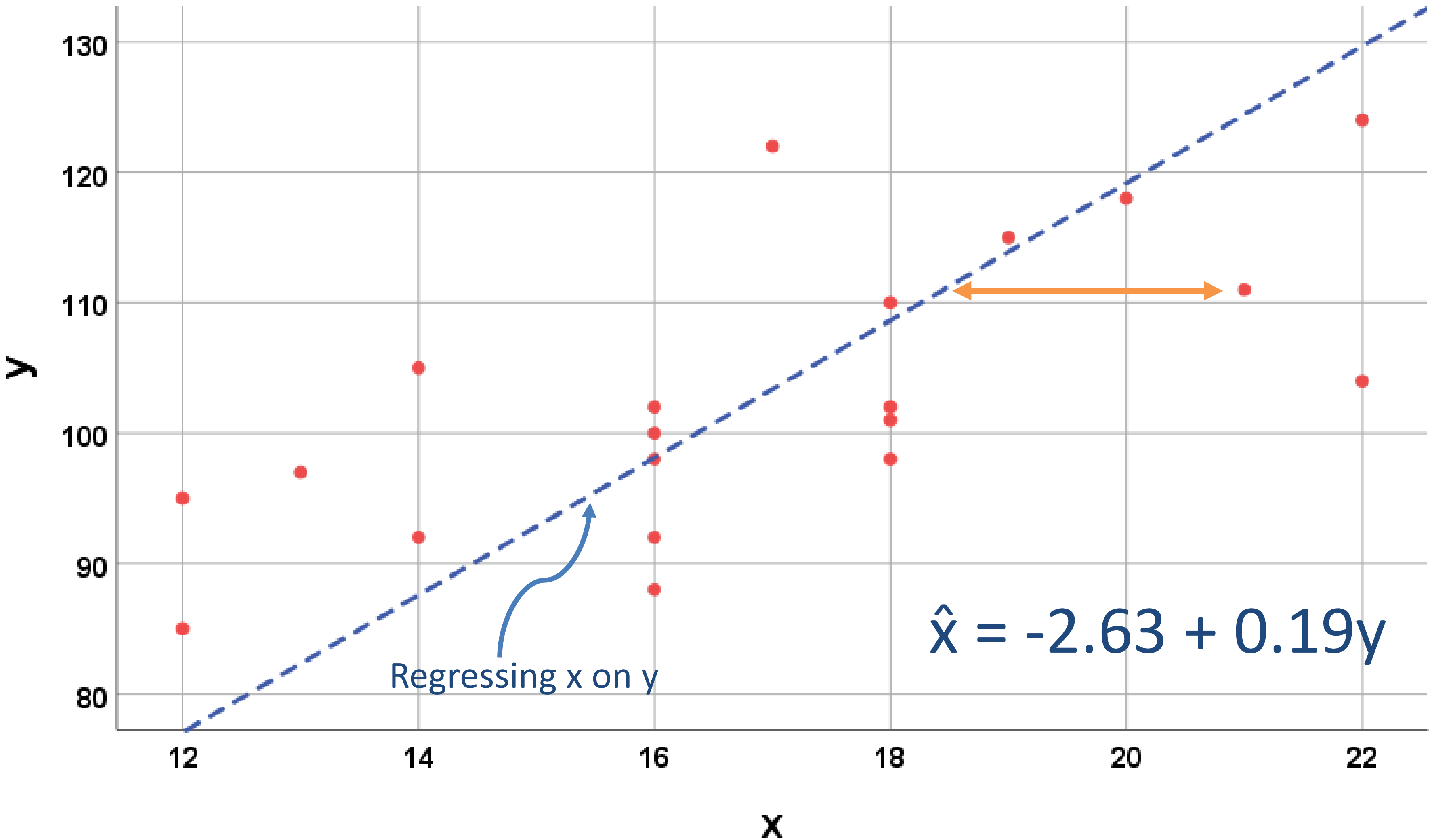
Pearson's $r = 0.686$



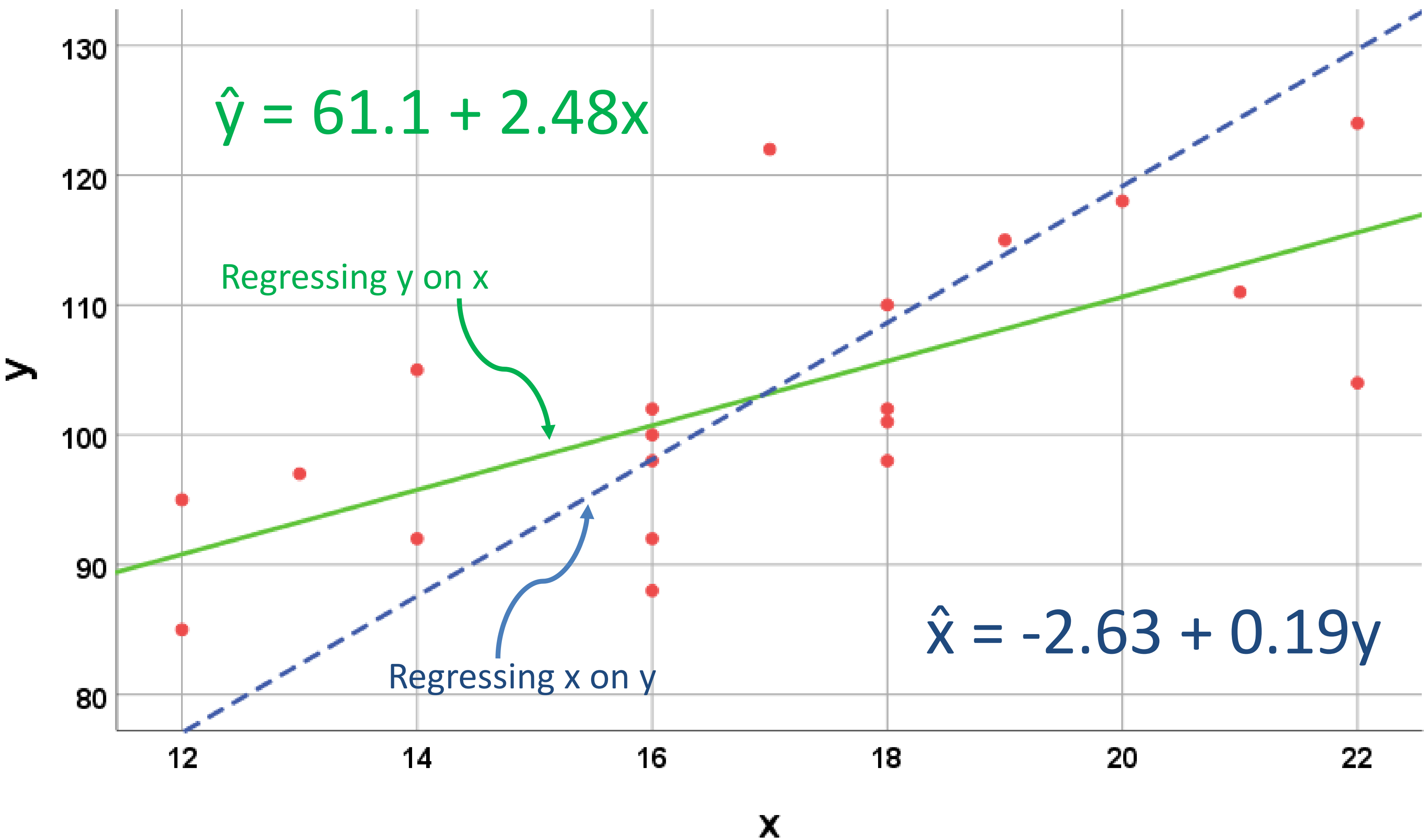
Minimizing the sum of squared vertical errors



Minimizing the sum of squared horizontal errors



Both together

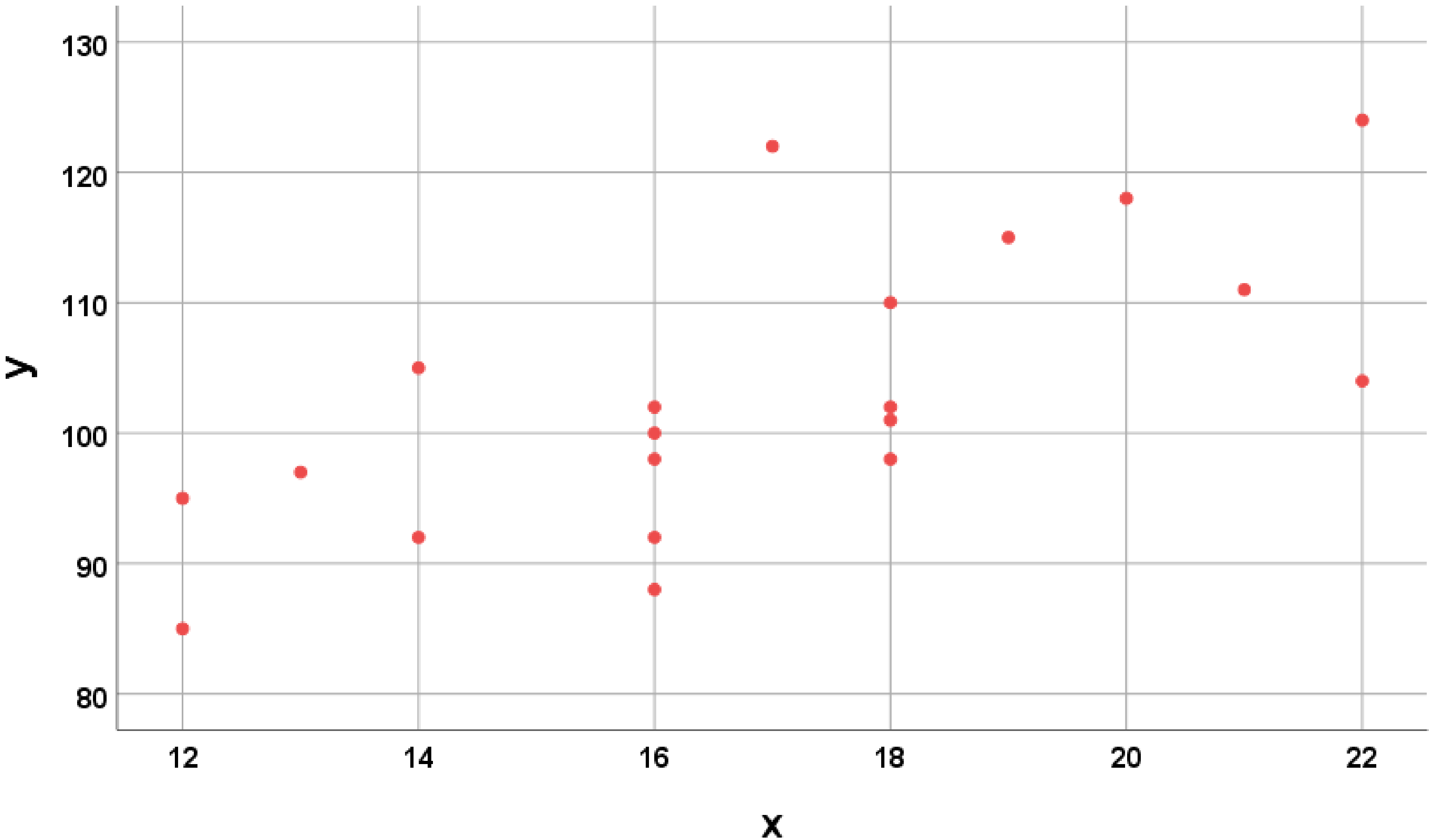


Regression vs Correlation

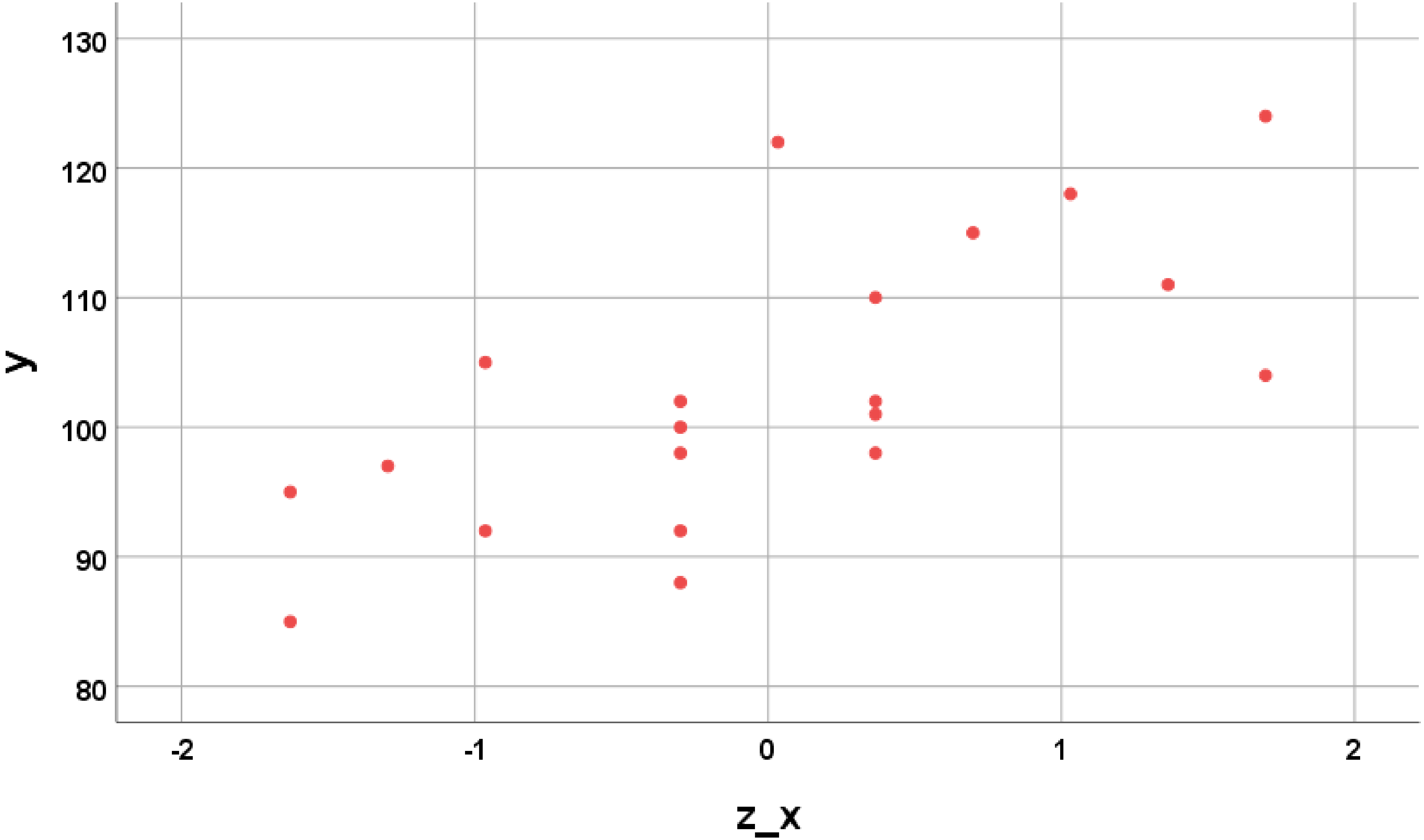
Regress y on x	Regress x on y
$\hat{\beta}_1 = \dfrac{Cov\,(y,x)}{Var(x)}$	$\hat{\beta}_1 = \dfrac{Cov\,(x,y)}{Var(y)}$

Correlate y with x	Correlate x with y
$r = \dfrac{Cov\,(y,x)}{SD(y)SD(x)}$	$r = \dfrac{Cov\,(x,y)}{SD(x)SD(y)}$

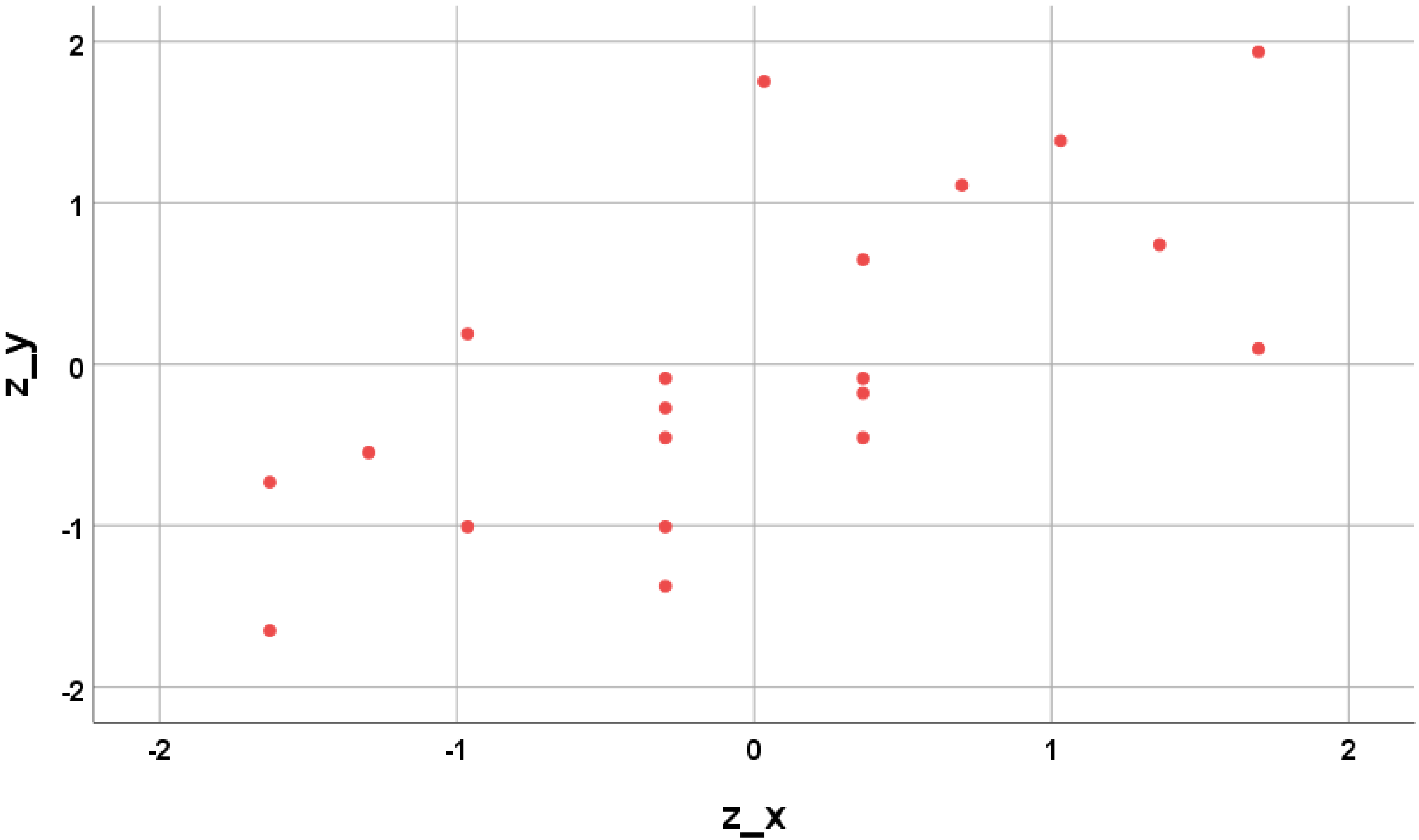
Back to the raw data



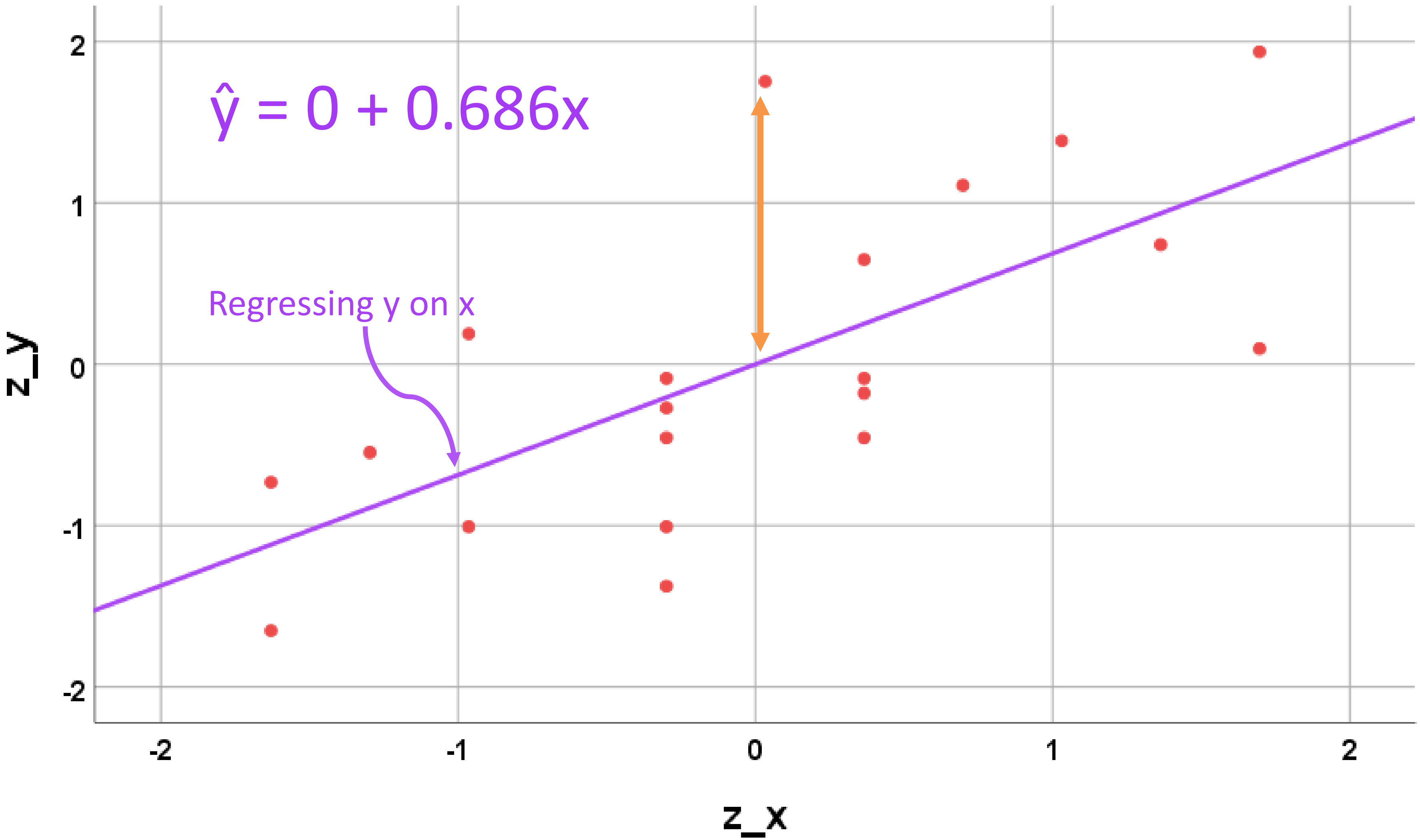
Standardizing the x-values



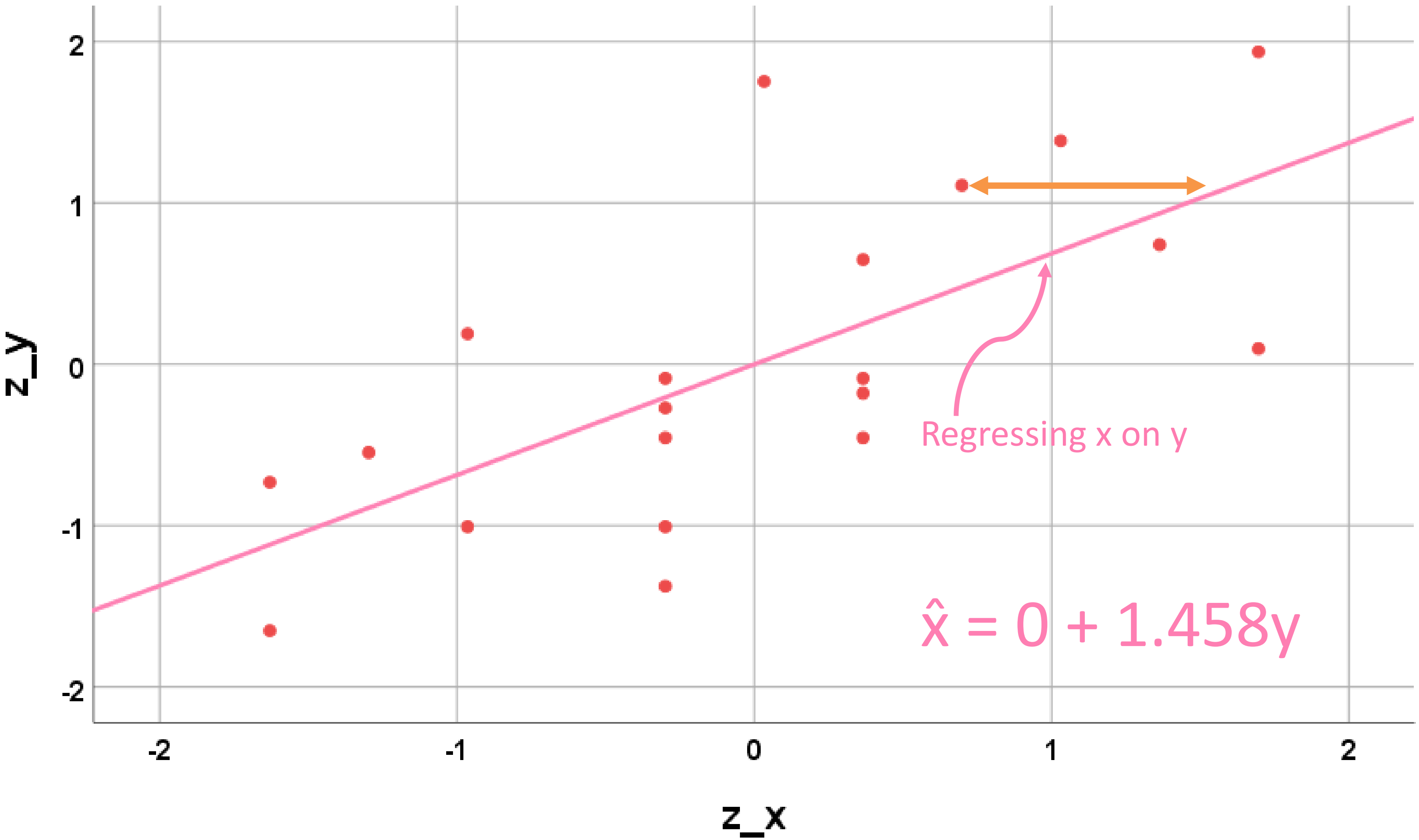
Standardizing the y-values as well



Minimizing the sum of squared vertical errors



Minimizing the sum of squared horizontal errors



Correlation & Causation: 1921

CORRELATION AND CAUSATION

By SEWALL WRIGHT

Senior Animal Husbandman in Animal Genetics, Bureau of Animal Industry, United States Department of Agriculture

PART I. METHOD OF PATH COEFFICIENTS

INTRODUCTION

The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated. Unfortunately, causes of variation often seem to be beyond control. In the biological sciences, especially, one often has to deal with a group of characteristics or conditions which are correlated because of a complex of interacting, uncontrollable, and often obscure causes. The degree of correlation between two variables can be calculated by well-known methods, but when it is found it gives merely the resultant of all connecting paths of influence.

present a method of measuring the

THE METHOD OF PATH COEFFICIENTS

By

SEWALL WRIGHT
Department of Zoology, The University of Chicago.

Fig. 1. Those variables which are treated as dependent are connected with those of which they are considered functions by arrows. The system of factors back of each variable may be made formally complete by the introduction of symbols representative of total residual determination (as V_0 in Fig. 1). A residual correlation between variables is represented by a double-headed arrow. It will be assumed that all relations are linear.[1] Thus each variable is related to those from which uni-

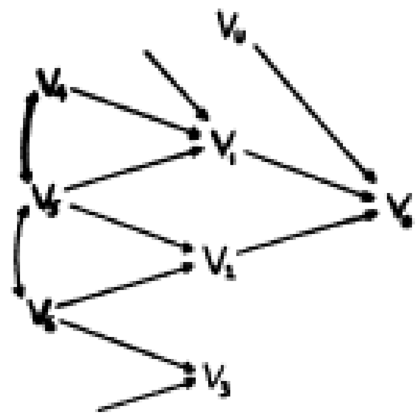


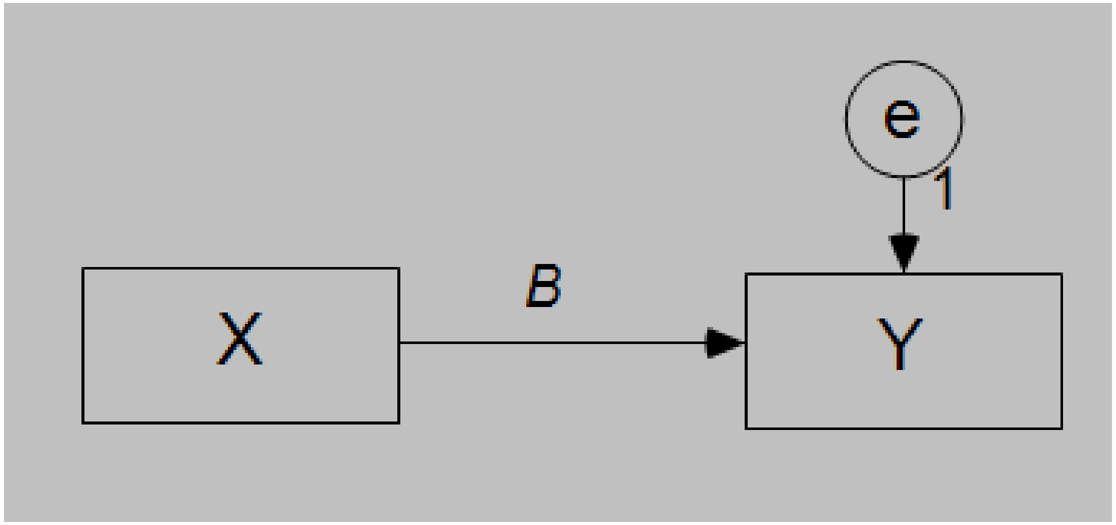
FIG. 1

Contemporary Path Models

- Path models are expressed as diagrams.
- The drawing convention is the same as in confirmatory factor analysis:
  - observed variables are drawn as rectangles,
  - unobserved variables as circles/ellipses and
  - relations are expressed as arrows;
    - straight single-headed arrows are used to indicate causal or predictive relationships and
    - curved double-headed arrows indicate a non-directional relationship such as a correlation or covariance.

Bivariate regression

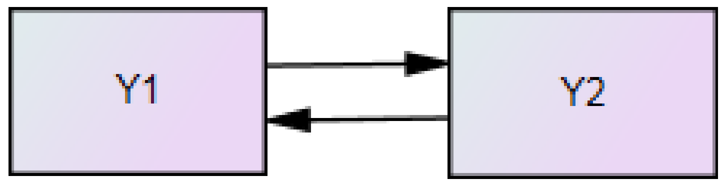
$$y = \beta x + e$$



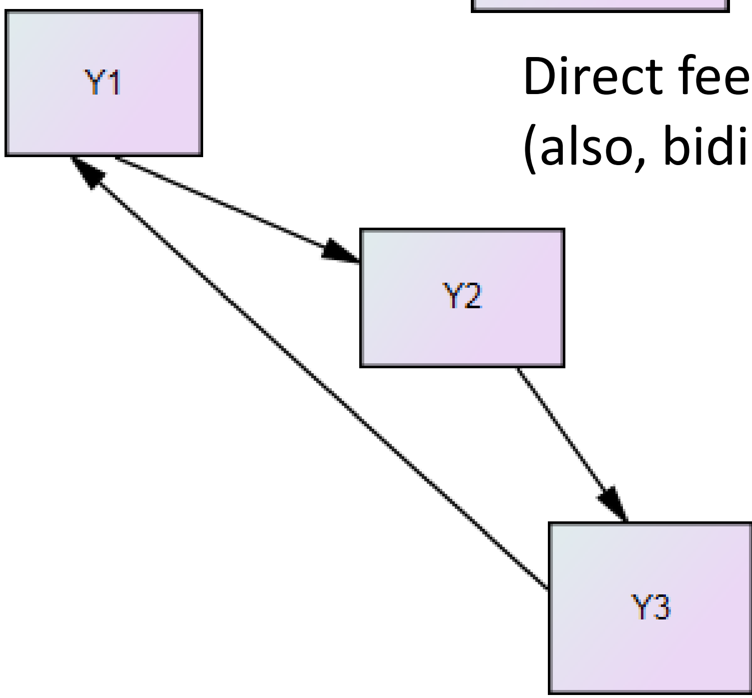
Two kinds of path models

- One is called recursive, the other non-recursive.
- Recursive models are simpler:
  - The paths are unidirectional and
  - the residual [error] terms are independent.
  - Such models can be tested with standard multiple regression.
- Non-recursive models can have
  - bidirectional paths,
  - correlated errors and
  - feedback loops.
  - Such models need structural equation software to fit them.

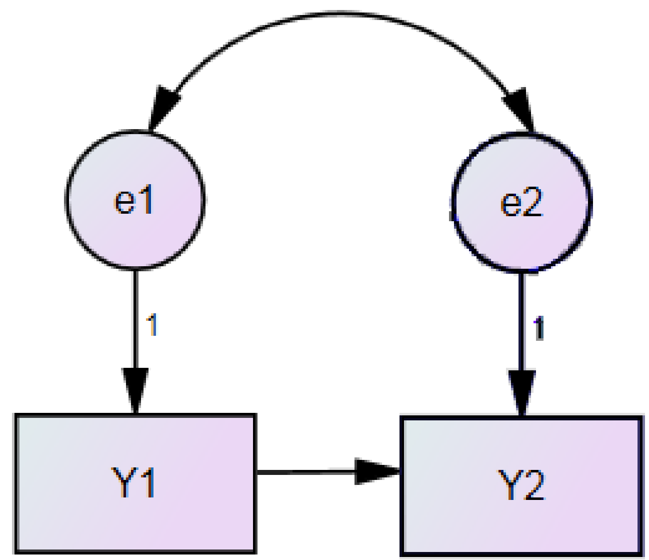
The terminology illustrated



Direct feedback loop
(also, bidirectional path)

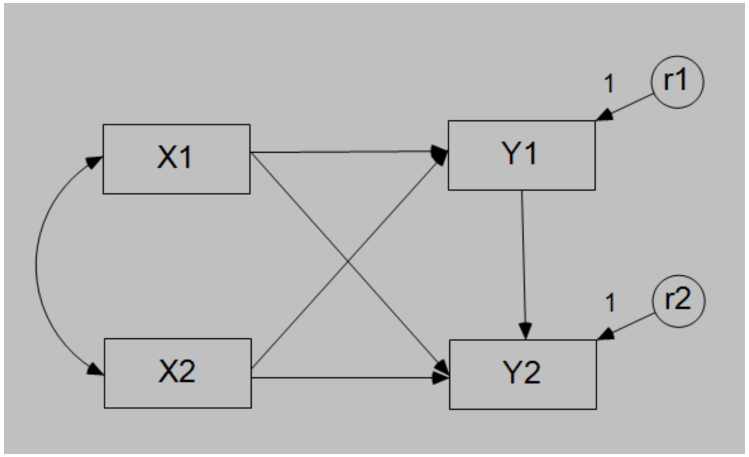
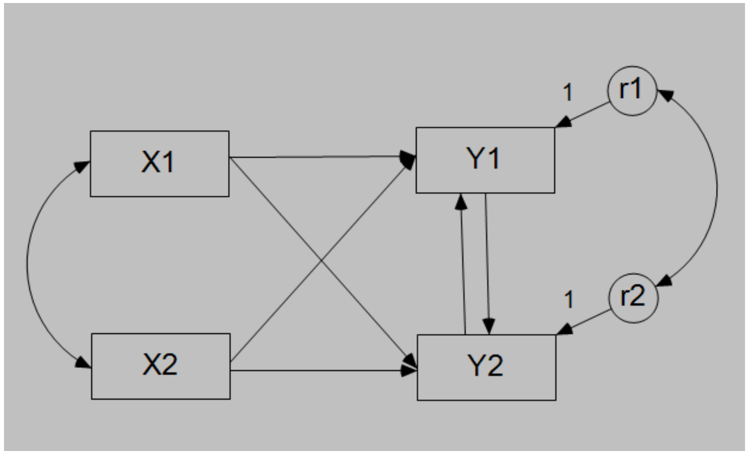


Indirect feedback loop



Correlated errors

The terminology illustrated

Recursive Model	Non-Recursive Model
	

Modelling data for path analysis

- Can be done via
  - Multiple regression
  - Structural Equation Modelling
- An example
  - some data from attitude modelling of factors affecting the perceived risk in genetically modified food
  - Scores on four attitude scales [measuring attitudes to technology, attitudes to nature, neophobia, and alienation] were used to predict scores on a perceived risk scale

ANOVA[b]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10338.569	4	2584.642	73.877	.000[a]
	Residual	16408.348	469	34.986		
	Total	26746.916	473			

a. Predictors: (Constant), Alienation, Technology, Neophobia, Nature

b. Dependent Variable: Risk

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.622[a]	.387	.381	5.9148809

a. Predictors: (Constant), Alienation, Technology, Neophobia, Nature

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.287	2.398		3.038	.003
	Technology	-.240	.052	-.175	-4.615	.000
	Nature	.371	.048	.297	7.774	.000
	Neophobia	.274	.043	.240	6.302	.000
	Alienation	.684	.083	.318	8.286	.000

a. Dependent Variable: Risk

Our model then would be

```
Predicted Risk = 7.287 – 0.240*Technology
                 + 0.371*Nature + 0.274*Neophobia
                 + 0.684*Alienation
```

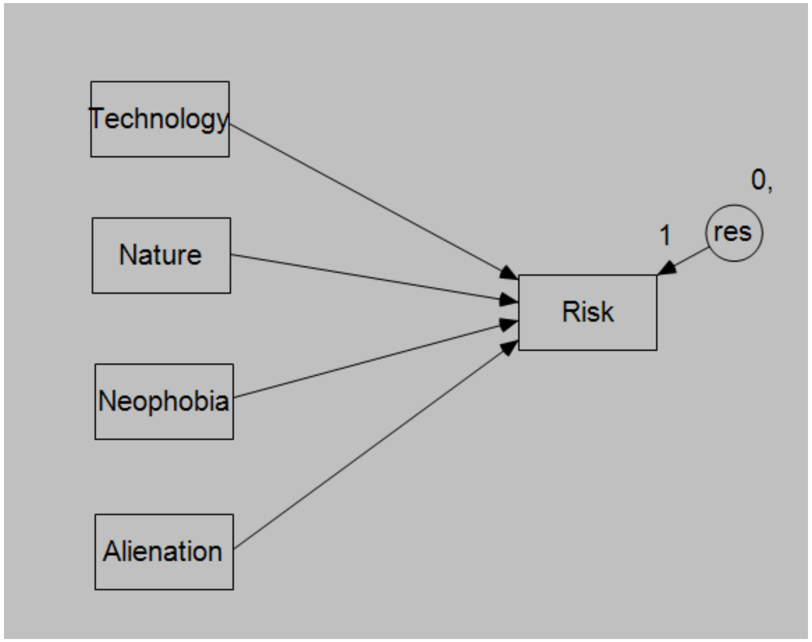
or in standard scores

```
Predicted Risk = -0.175*Technology
                 + 0.297*Nature
                 + 0.240*Neophobia + 0.318*Alienation
```

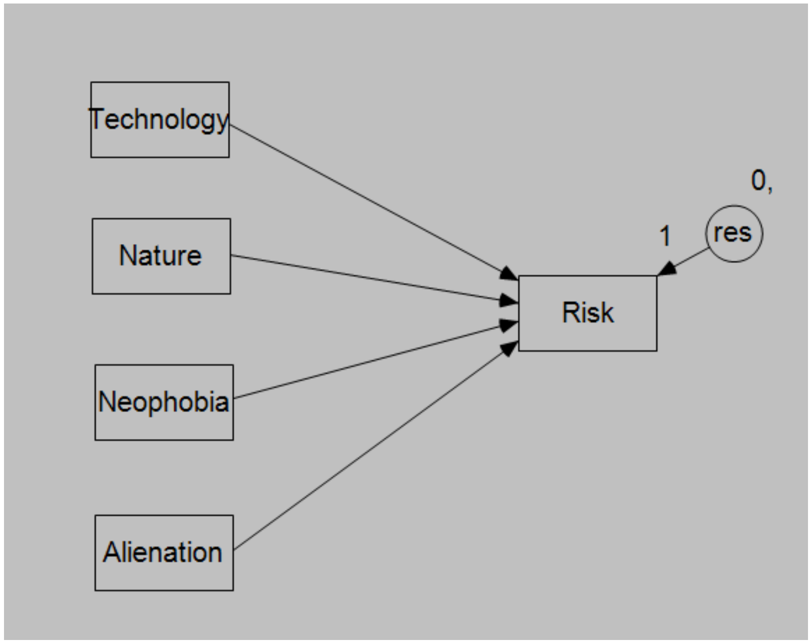
or in terms of actual risk scores

```
Risk = 7.287 – 0.240*Technology
                 + 0.371*Nature + 0.274*Neophobia
                 + 0.684*Alienation + residual
```

In Amos the multiple regression model would look like this:



In Amos the multiple regression model would look like this:



Squared Multiple Correlations:

			Estimate
Risk			.312

Standardized Regression Weights

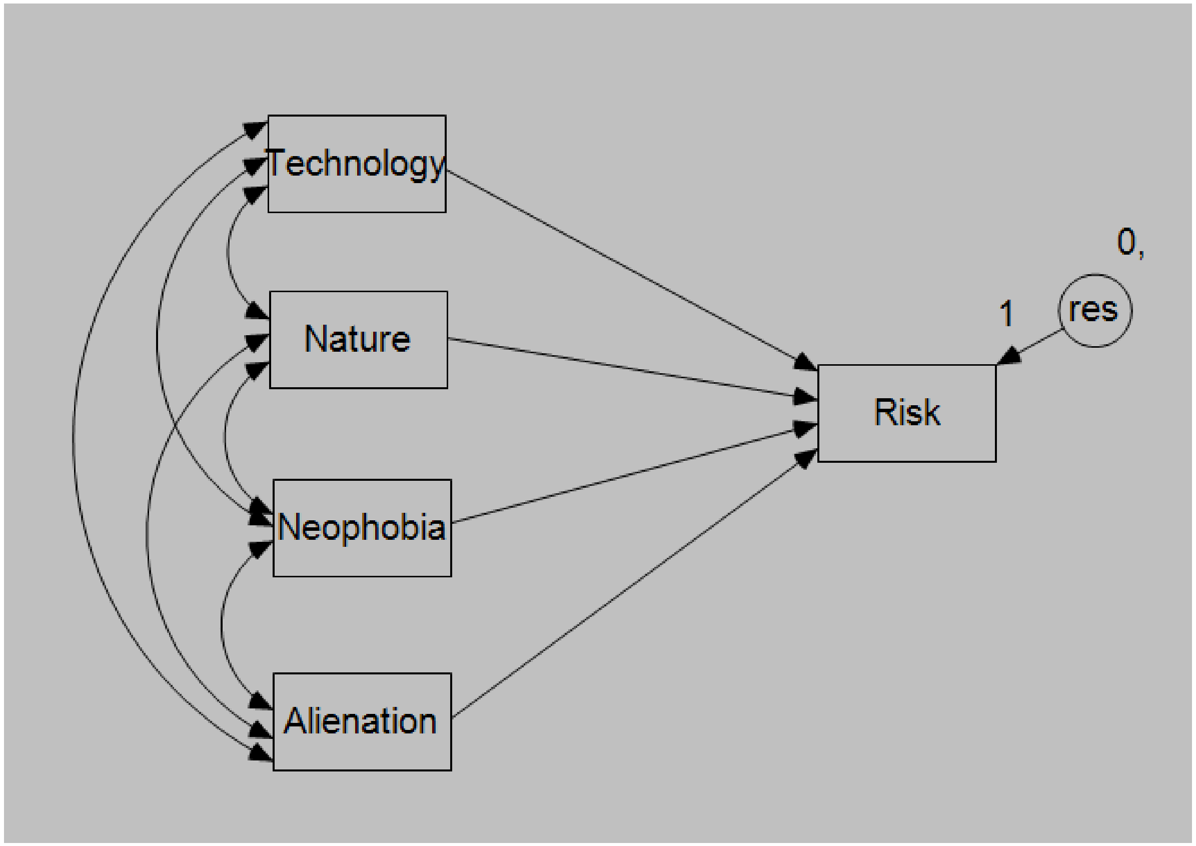
			Estimate
Risk	<---	Technology	-.186
Risk	<---	Nature	.314
Risk	<---	Neophobia	.255
Risk	<---	Alienation	.337

Regression Weights:

			Estimate	S.E.	C.R.	P	Label
Risk	<---	Technology	-.240	.049	-4.866	***	
Risk	<---	Nature	.371	.045	8.238	***	
Risk	<---	Neophobia	.274	.041	6.676	***	
Risk	<---	Alienation	.684	.077	8.828	***	

Comparing regression & SEM

- Regression weights agree perfectly, but
- standard errors differ
- standardized regression weights differ
- and the squared multiple correlation is rather less in Amos.
- and we did get a warning in Amos
  - (uncorrelated predictors)


Technology
Nature
Neophobia
Alienation
Risk
1
res
0,

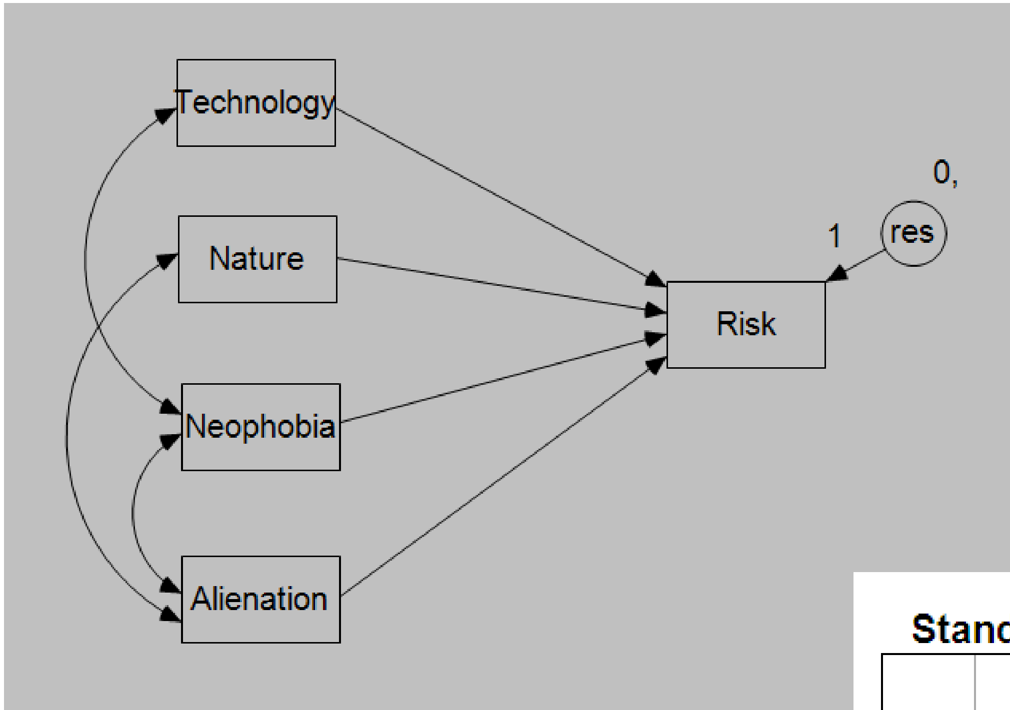
	Regression			Structural Equation Modelling					
				No Exogenous Correlation			Exogenous Correlations		
	B	S.E.	Beta	B	S.E.	Beta	B	S.E.	Beta
Technology	-.240	.052	-.175	-.240	.049	-.186	-.240	.052	-.175
Nature	.371	.048	.297	.371	.045	.314	.371	.048	.297
Neophobia	.274	.043	.240	.274	.041	.255	.274	.043	.240
Alienation	.684	.083	.318	.684	.077	.337	.684	.082	.318
SMC			.387			.312			.387

Conclusion

- Multiple regression must model the correlations among the independent variables, although this is not shown.
- A path analytic representation is thus a much more accurate representation.
  - And gives more information

For example,
in modelling the associations among the exogenous variables,
we can see that not all are significantly different from zero
and could be deleted from the model.

			Covariance	S.E.	C.R.	P		Correlation
Neophobia	<-->	Alienation	2.411	1.065	2.263	.024		.105
Nature	<-->	Neophobia	1.487	1.825	.815	.415		.037
Technology	<-->	Nature	-1.098	1.517	-.724	.469		-.033
Nature	<-->	Alienation	6.651	1.012	6.572	***		.317
Technology	<-->	Alienation	.167	.881	.189	.850		.009
Technology	<-->	Neophobia	-10.863	1.738	-6.249	***		-.300



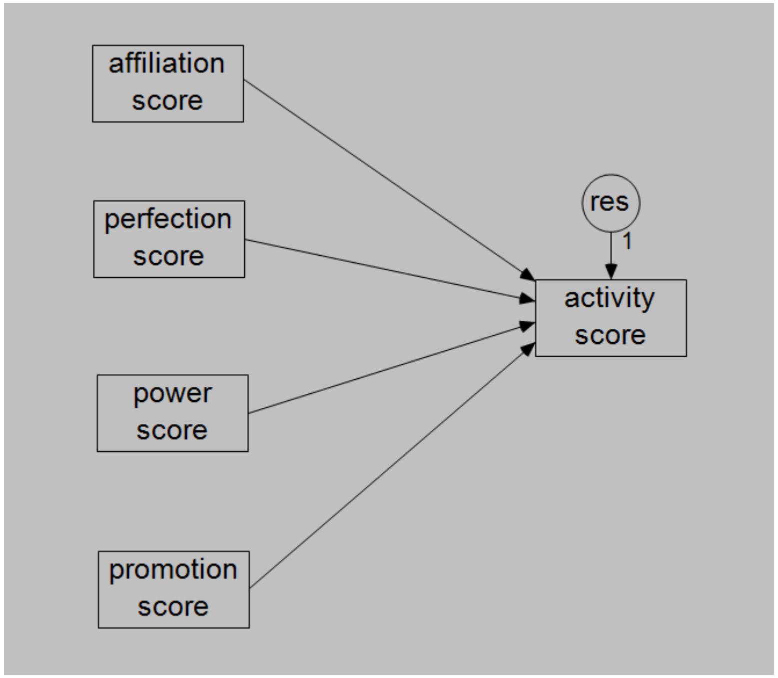
Standardized Regression Weights:

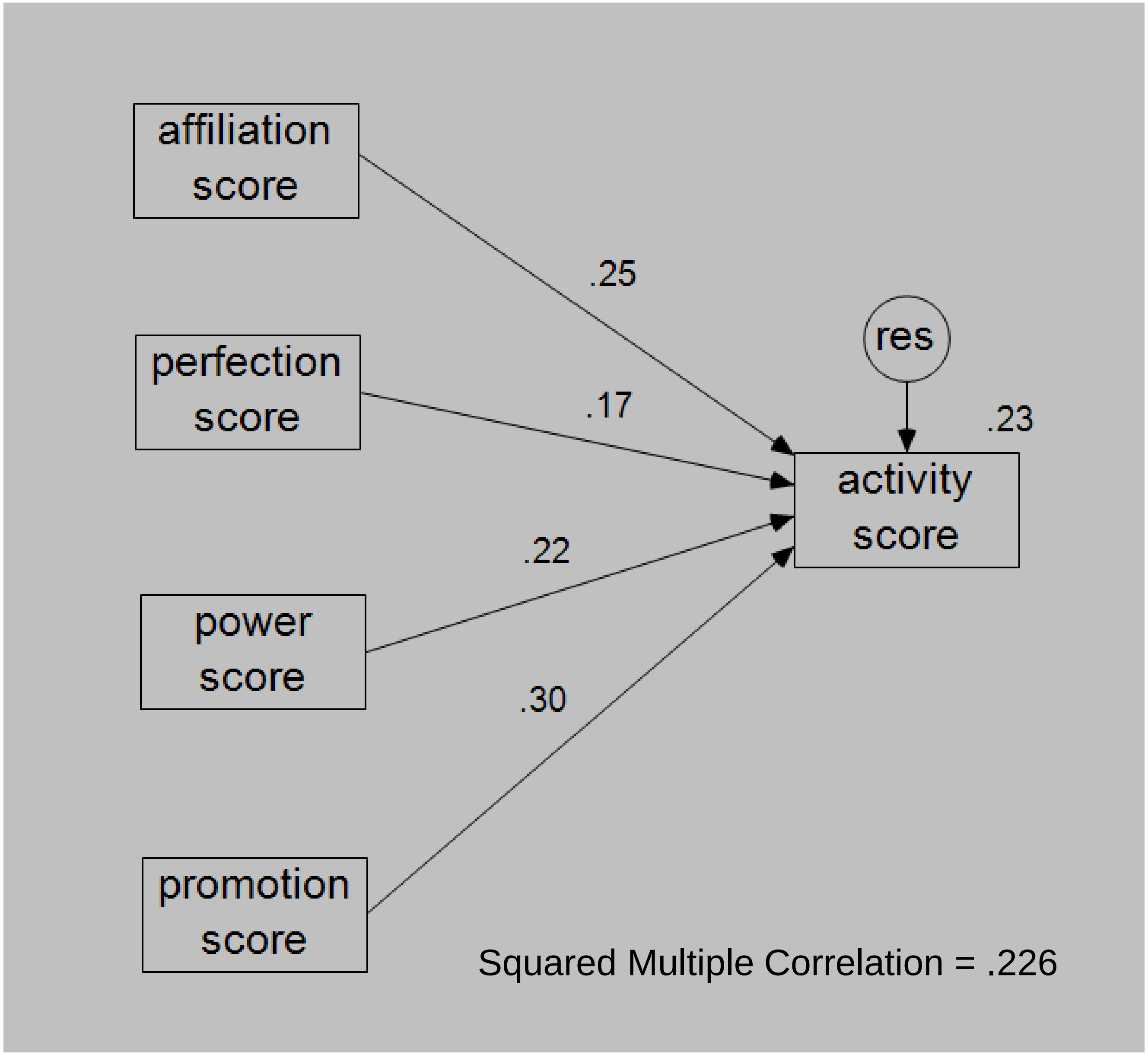
			Partly Correlated Estimate	Fully Correlated Estimate
Risk	<---	Technology	-.176	-.175
Risk	<---	Nature	.298	.297
Risk	<---	Neophobia	.242	.240
Risk	<---	Alienation	.319	.318

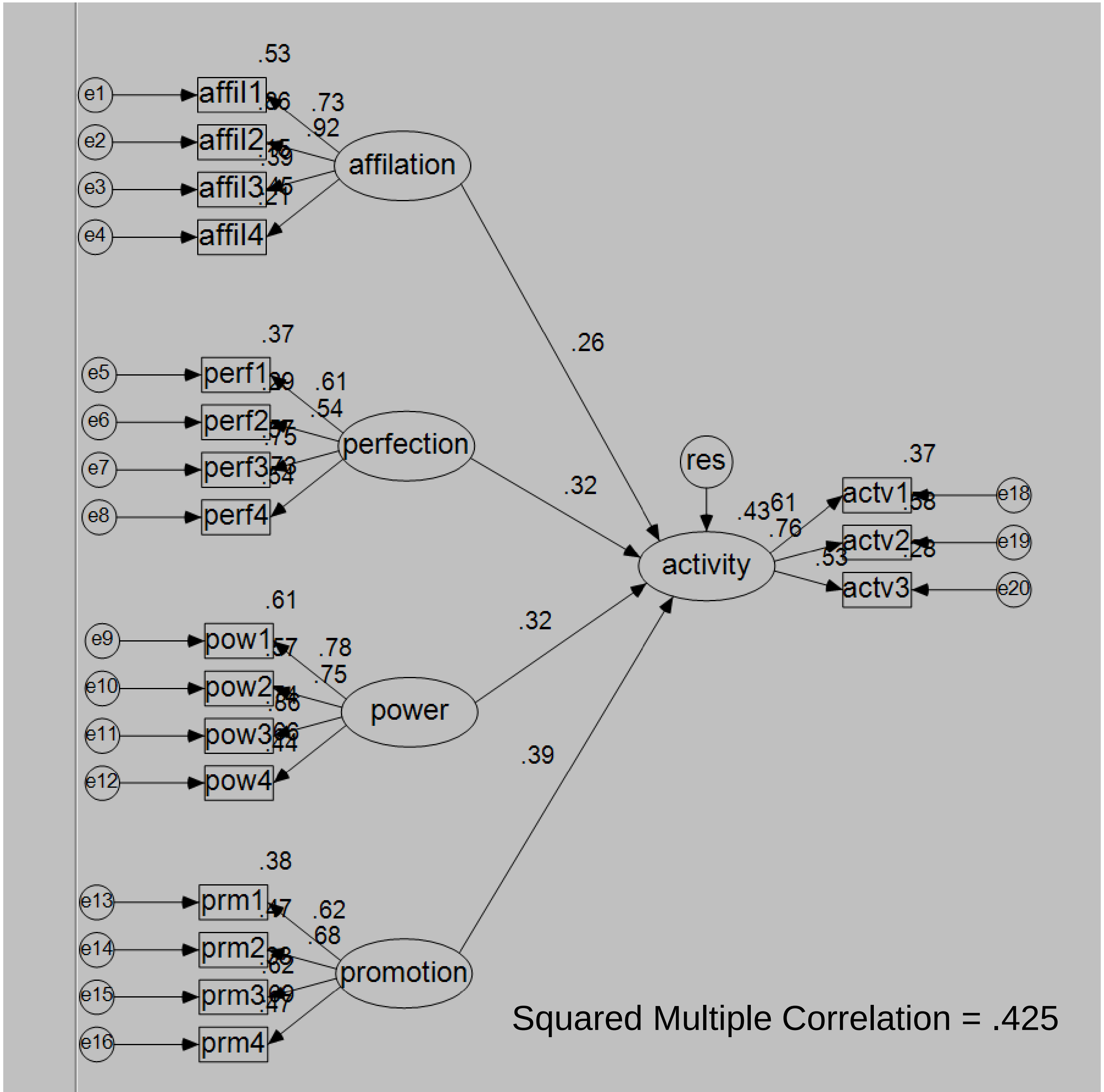
This model is ‘over-identified’ and thus can be tested for significance.

Latent Variable regression

Some real advantages come about when we have latent variables in the model ie a combined factor-regression analysis




affiliation score
perfection score
power score
promotion score
activity score
res
.25
.17
.22
.30
.23
Squared Multiple Correlation = .226


affil1
affil2
affil3
affil4
affilation
perf1
perf2
perf3
perf4
perfection
pow1
pow2
pow3
pow4
power
prm1
prm2
prm3
prm4
promotion
activity
res
actv1
actv2
actv3
e1
e2
e3
e4
e5
e6
e7
e8
e9
e10
e11
e12
e13
e14
e15
e16
e18
e19
e20
.53
.73
.92
.37
.61
.54
.61
.78
.75
.38
.62
.68
.26
.32
.32
.39
.43
.76
.37
Squared Multiple Correlation = .425

Comparison

	Manifest	Latent
affiliation	.25	.26
perfection	.17	.32
power	.22	.32
promotion	.30	.39
Variance Accounted for	.226	.425

Section 2

MULTI STEP REGRESSIONS

Multi-Step Path Analysis

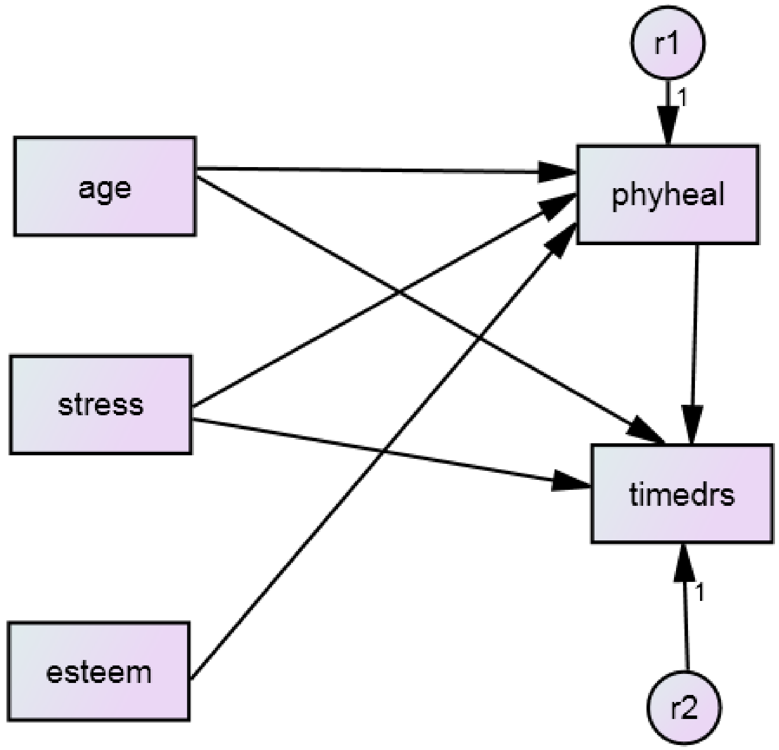
- Some variables are predictors [A]
- Some variables are predicted [B]
- Some variables are both predictors and predicted [M]
- A → M → B
- Such intervening variables [like M] may be considered as mediating the relationship between A and B.

A simple example: Women's health study

- Survey of 465 females on a variety of health issues (Hoffman & Fidell, 1979 – reported in Tabachnik & Fidell, 2013).
- One interest was in usage of health facilities (eg. Frequency of visits to health professionals – *TIMEDRS*).
- Other variables:
  - Age, Stress, Self esteem *(esteem*), Self reported Physical health (*phyheal*)

A simple example: Women's health study

Proposed theoretical model



Requires two regressions:

1. Regress *phyheal* on *age, stress, esteem*

2. Regress *timedrs* on *age, stress, phyheal*

A simple example: Women's health study

Two regressions:

1. Regress *phyheal* on *age, stress, esteem*
2. Regress *timedrs* on *age, stress, phyheal*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.345[a]	.119	.113	2.23799

a. Predictors: (Constant), esteem, age, stress

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.900	.549		3.460	.001
	age	.136	.049	.128	2.778	.006
	stress	.006	.001	.349	7.548	.000
	esteem	.073	.027	.122	2.764	.006

a. Dependent Variable: phyheal

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.476[a]	.227	.222	8.80548

a. Predictors: (Constant), phyheal, age, stress

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.439	1.404		-2.449	.015
	age	.055	.194	.012	.283	.777
	stress	.010	.004	.134	2.940	.003
	phyheal	1.761	.183	.419	9.629	.000

a. Dependent Variable: timedrs

But how do we combine them?

- Overall Fit:
  - No overall significance test
  - Can estimate 'generalized squared multiple correlation' $R_m^2 = 1 - \Pi(1 - R_i^2)$ where R_i^2 is the SMC for each of the *i* regression equations.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.345[a]	.119	.113	2.23799

a. Predictors: (Constant), esteem, age, stress

Model Summary

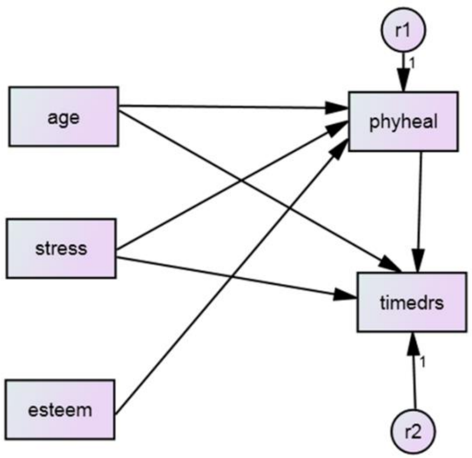
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.476[a]	.227	.222	8.80548

a. Predictors: (Constant), phyheal, age, stress

$$R_m^2 = 1 - (1 - .119) \times (1 - .227)$$
$$= 0.32$$

But how do we combine them?

- The impact of antecedent variables on the dependent variable:
  - Directly antecedent variables: β [called *Direct Effects*]
  - Indirect (multiple sequential arrows) = Π β (ie multiply the β's together) [called *Indirect Effects*]



Direct and indirect effects

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1.900	.549		3.460	.001
	age	.136	.049	.128	2.778	.006
	stress	.006	.001	.349	7.548	.000
	esteem	.073	.027	.122	2.764	.006

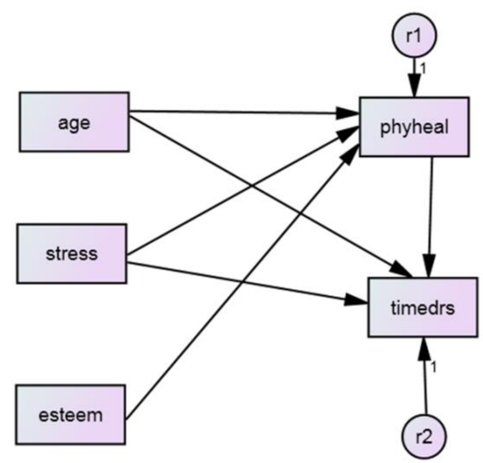
a. Dependent Variable: phyheal

Coefficients[a]

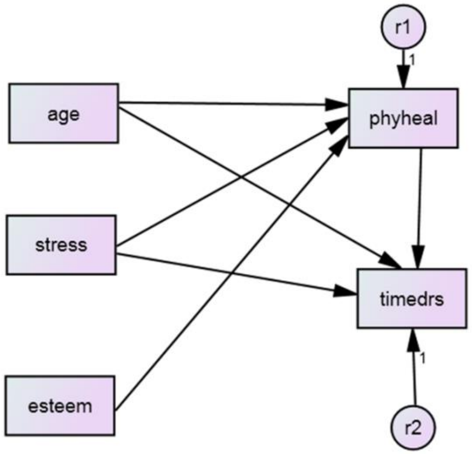
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-3.439	1.404		-2.449	.015
	age	.055	.194	.012	.283	.777
	stress	.010	.004	.134	2.940	.003
	phyheal	1.761	.183	.419	9.629	.000

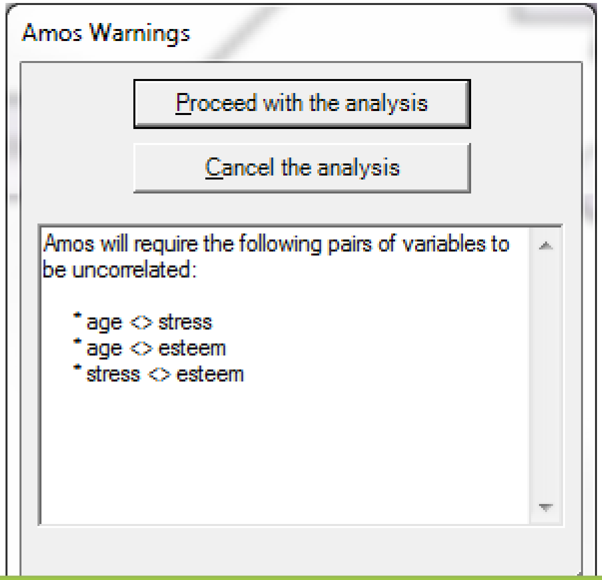
a. Dependent Variable: timedrs

- Direct Effects on Timedrs
  - $\beta_{age} = 0.012$
  - $\beta_{stress} = 0.134$
  - $\beta_{phyheal} = 0.419$
- Indirect effects
  - $\beta_{\text{age (via phyheal)}} = 0.128 \times 0.419 = 0.054$
  - $\beta_{\text{stress (via phyheal)}} = 0.349 \times 0.419 = 0.146$
  - $\beta_{\text{esteem (via phyheal)}} = 0.122 \times 0.419 = 0.051$



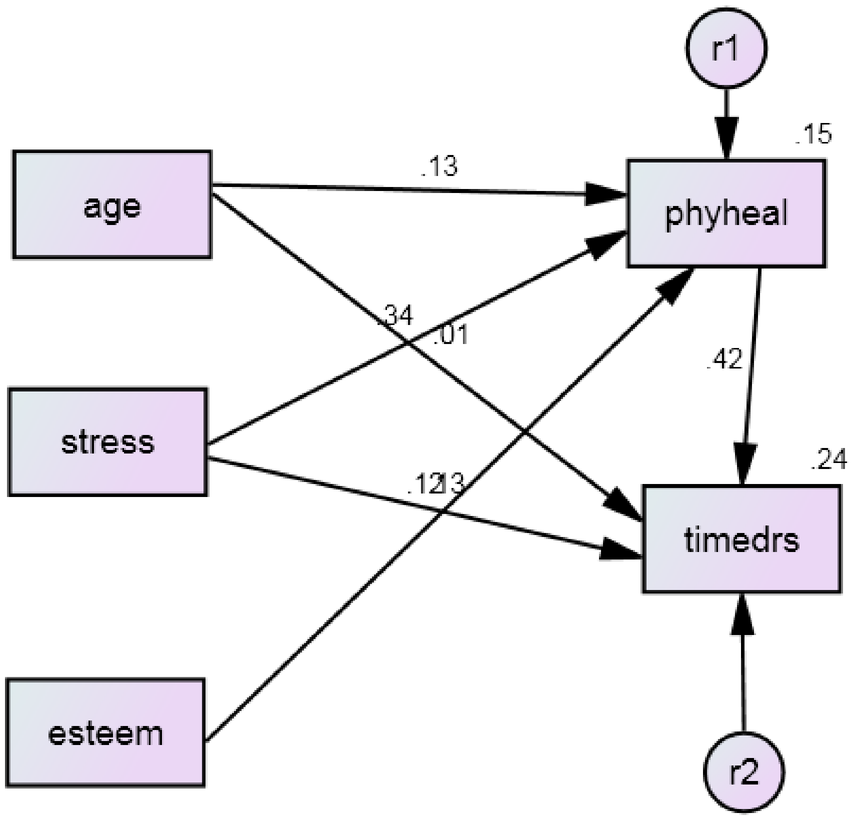
Structural equation model





AMOS is suspicious of uncorrelated predictors (exogenous)

Standardized estimates



Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1.900	.549		3.460	.001
	age	.136	.049	.128	2.778	.006
	stress	.006	.001	.349	7.548	.000
	esteem	.073	.027	.122	2.764	.006

a. Dependent Variable: phyheal

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-3.439	1.404		-2.449	.015
	age	.055	.194	.012	.283	.777
	stress	.010	.004	.134	2.940	.003
	phyheal	1.761	.183	.419	9.629	.000

a. Dependent Variable: timedrs

Unstandardized estimates

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.900	.549		3.460	.001
	age	.136	.049	.128	2.778	.006
	stress	.006	.001	.349	7.548	.000
	esteem	.073	.027	.122	2.764	.006

a. Dependent Variable: phyheal

Coefficients[a]

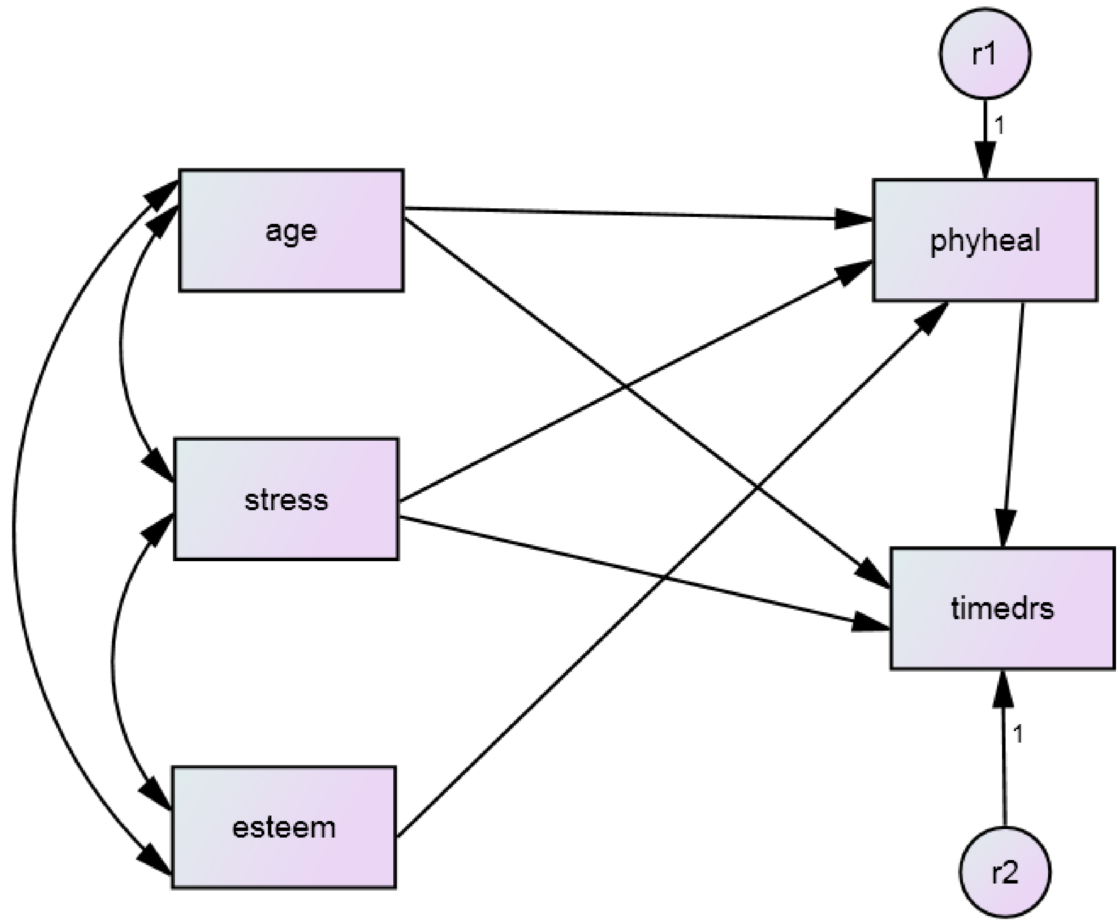
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.439	1.404		-2.449	.015
	age	.055	.194	.012	.283	.777
	stress	.010	.004	.134	2.940	.003
	phyheal	1.761	.183	.419	9.629	.000

a. Dependent Variable: timedrs

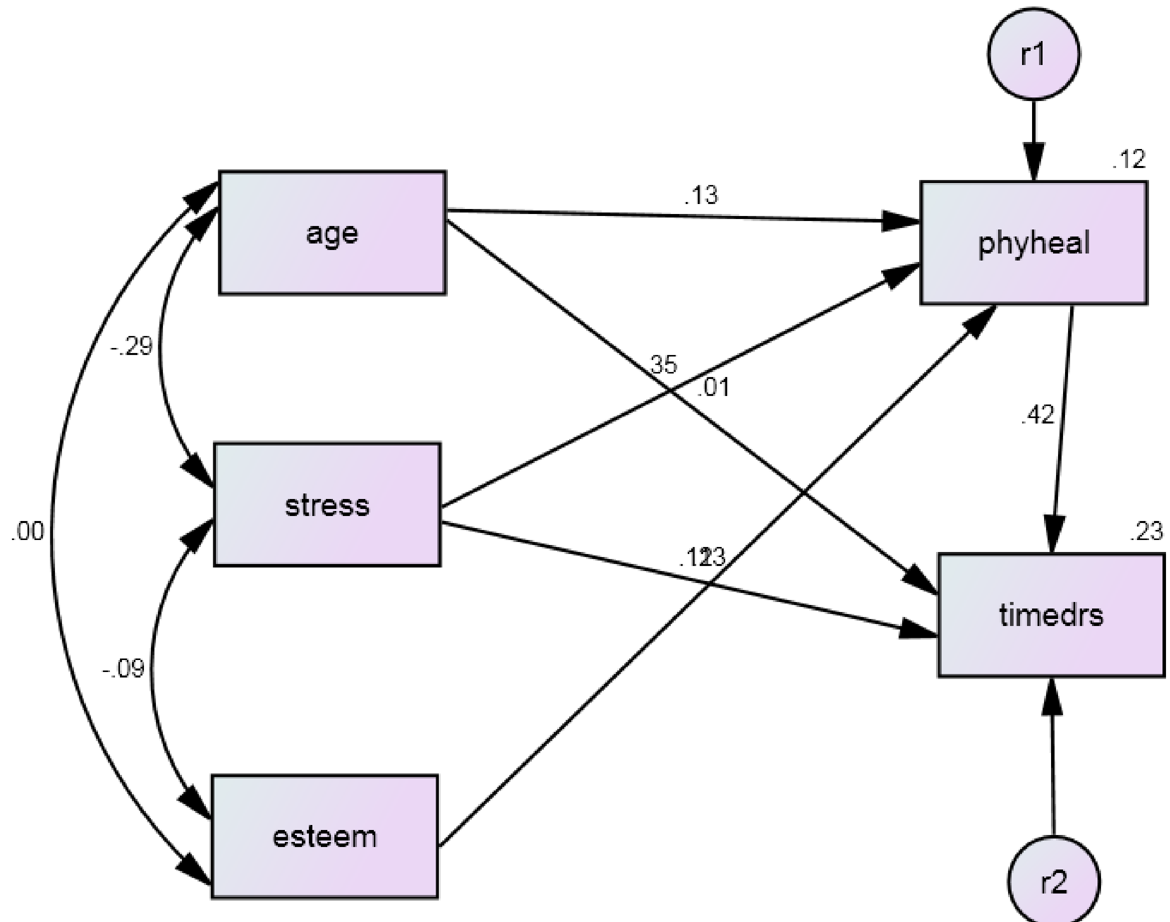
			Estimate	S.E.	C.R.	P	Label
phyheal	<---	age	.136	.047	2.913	.004	
phyheal	<---	stress	.006	.001	7.948	***	
phyheal	<---	esteem	.073	.026	2.785	.005	
timedrs	<---	phyheal	1.761	.182	9.661	***	
timedrs	<---	age	.055	.186	.297	.767	
timedrs	<---	stress	.010	.003	3.055	.002	

Squared Multiple Correlations:	Estimate
phyheal	.148
timedrs	.237

Include correlations among exogenous predictors



Include correlations among exogenous predictors



Squared Multiple Correlations: Estimate
phyheal .119
timedrs .227

Section 3

MODEL FIT

Assessing fit

"With respect to model fit, researchers do not seem adequately sensitive to the fundamental reality that there is no true model [...], that all models are wrong to some degree, even in the population, and that the best one can hope for is to identify a parsimonious, substantively meaningful model that fits observed data adequately well. At the same time, one must recognize that there may well be other models that fit the data to approximately the same degree. Given this perspective, it is clear that a finding of good fit does not imply that a model is correct or true, but only plausible. These facts must temper conclusions drawn about good-fitting models. (MacCallum & Austin 2000, p.218)

Model Test Statistics

- Model Test Statistics ask:
  "Is the variance-covariance matrix implied by your model sufficiently close to your observed variance-covariance matrix that the difference could plausibly be due to sampling error?"

- **Model Chi-Square Test (**χ^2 test**)**
  - $\chi^2 = 0$ with perfect model fit, and increases as model misspecification increases. $p = 1$ with perfect model fit, and decreases as model misspecification increases.
  - Non-significant Chi-Square test indicates the model is consistent with the observed variance-covariance matrix.
  - Significant Chi-Square insufficient in itself to determine whether a model should be rejected, but can be treated as a 'smoke alarm' in relation to model fit.
  - χ^2 is strongly affected by multivariate non-normality, correlation size, the size of unique variances, and sample size.

Approximate Fit Indices

- Approximate Fit Indices ignore the issue of sampling error and take different perspectives on providing a continuous measure of model-data correspondence.

- Three main flavours available under ML estimation:

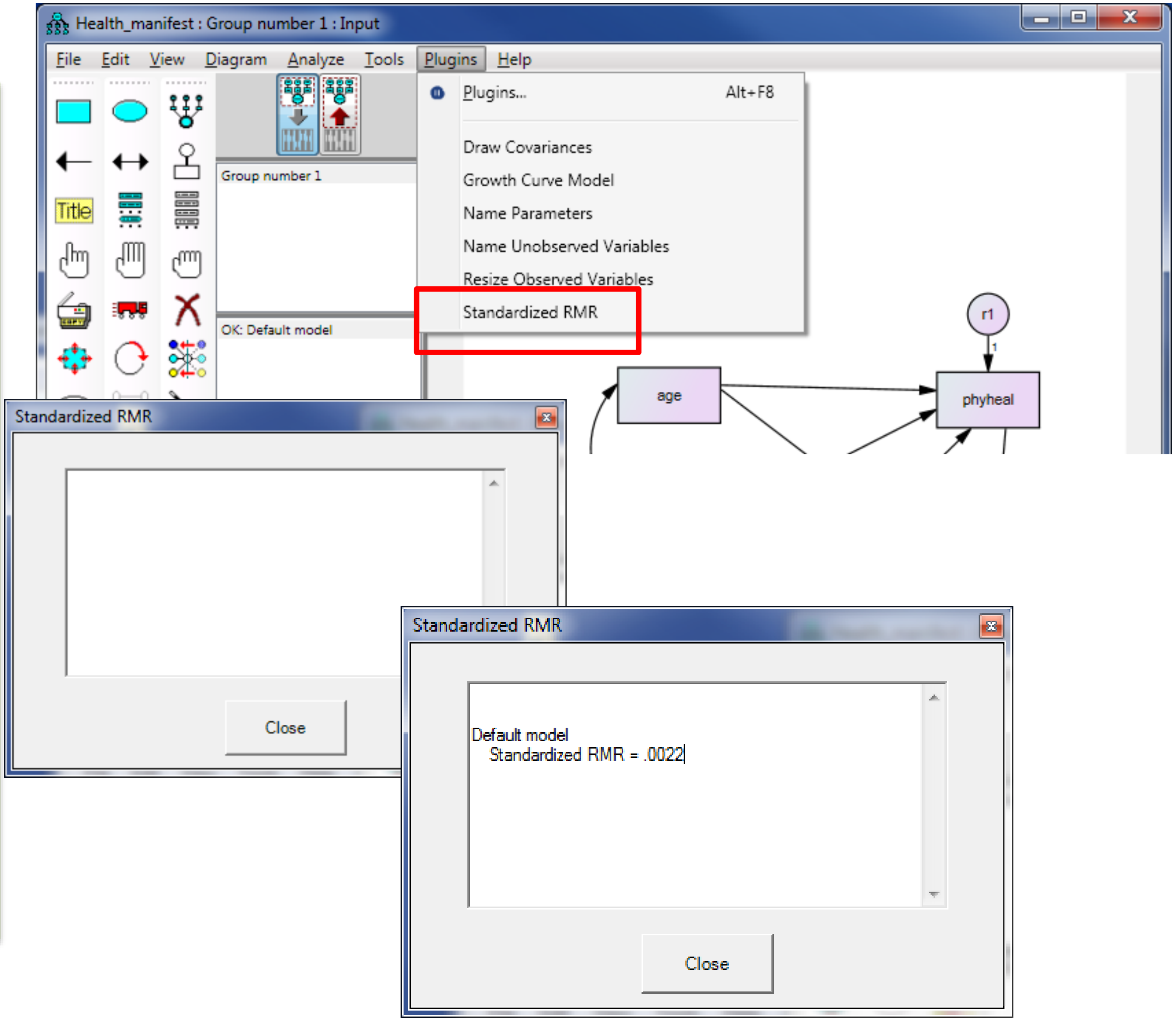
  1. Absolute Fit Indices (proportion of the observed variance-covariance matrix explained by model), e.g. SRMR.
  2. Comparative Fit Indices (relative improvement in fit compared to a baseline), e.g. CFI.
  3. Parsimony-adjusted-indices (compare model to observed data but penalise models with greater complexity) e.g. RMSEA.

Absolute Fit Indices

- RMR: Root mean square residual
  - Average differences between observed and model-estimated covariance matrices
  - Small indicates good fit, 0 = perfect fit
  - Hard to interpret because RMR's range depends on the range of the observed variables
- SRMR: Standardized RMR
  - Transforms the sample and model-estimated covariance matrices into correlation matrices.
  - Ranges from 0 to 1, with 0 = perfect fit.
  - SRMR < 0.08 regarded as good fit by Hu and Bentler, 1999 – beware that this average value can hide some big residuals.

To get SRMR in Amos

- Run the analysis
- Then click 'Plugins' on Menu Bar to get
- Select "Standardized RMR"
- Click Run
- The SRMR appears in the window



Comparative Fit Indices

- GFI: Goodness of fit index.
  - Estimates the proportion of covariances in the observed data that are explained by the model
  - Analogous to R^2 in a regression, and is meant to range from 0 (worst model fit) to 1 (best).
  - Limitations: expected values vary with sample size, doesn't always stick to the range 0-1.
- CFI: Bentler's Comparative Fit Index
  - Similar rationale to GFI, but stays normed to 0-1.
  - High values (> 0.95) regarded by Hu and Bentler (1999) as indicating good fit – again this has been much criticized.

Parsimony-adjusted indices

- RMSEA: Root Mean Square Error of Approximation.

$$RMSEA = \sqrt{\frac{Chisquare - df}{(N-1)df}}$$

  - Acts to ‘reward’ models analysed with larger samples, and models with more degrees of freedom.
  - Badness of fit statistic – lower is better and zero is best.
  - If it turns out to be less than zero, treat it as zero.

RMSEA less than 0.05 – close fit
RMSEA between 0.05 and 0.08 – fair fit
RMSEA between 0.08 and 0.10 – mediocre fit
RMSEA over 0.10 – unacceptable fit.
Browne & Cudeck (1993)

Limitations of fit statistics

Kline (2016) lists six main limitations of fit statistics:

1. They test only the average/overall fit of a model.

2. Each statistic reflects only a specific aspect of fit.

3. They don't relate clearly to the degree/type of model misspecificiation.

4. Well-fitting models do not necessarily have high explanatory power.

5. They cannot indicate whether results are theoretically meaningful.

6. Fit statistics say little about person-level fit.

Overfitting

An example from my life: predicting house prices in Brunswick. I want to predict the results of future auctions from information gained from past auctions.

Price ($m)	Size (m^2)	Street number	Red front door	Phone number /10,000,000
2.0	280	4	1	8.5340100
1.2	142	72	0	8.5353242
1.6	172	180	0	8.5349900
1.9	202	9	0	8.5445452
2.1	350	22	1	8.5392342

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients
		B	Std. Error	Beta
1	(Constant)	-2788.958	.000	
	Size	-.026	.000	-6.064
	StreetNum	.012	.000	2.414
	RedDoor	5.640	.000	8.471
	PhoneNum	327.228	.000	3.906

a. Dependent Variable: Price

$R^2 = 1$! But this model is severely overfit and won't generalise well to new auctions.

Model Summary

Model	R	R Square
1	1.000[a]	1.000

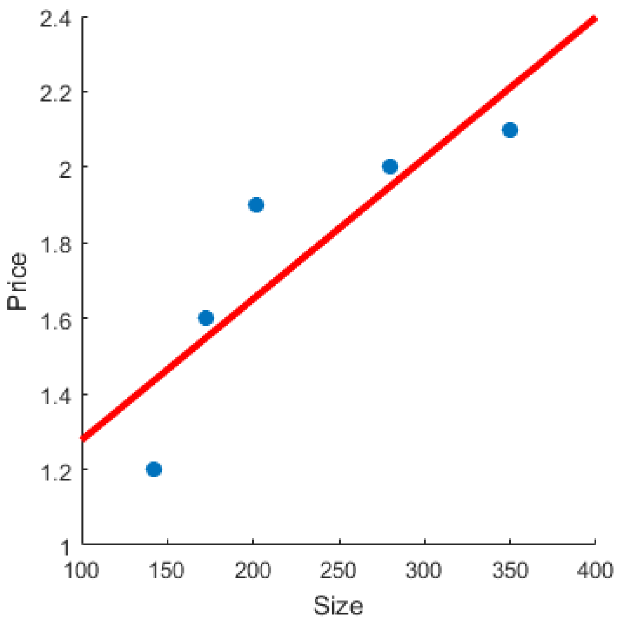
a. Predictors: (Constant), PhoneNum, Size, StreetNum, RedDoor

Overfitting vs Underfitting

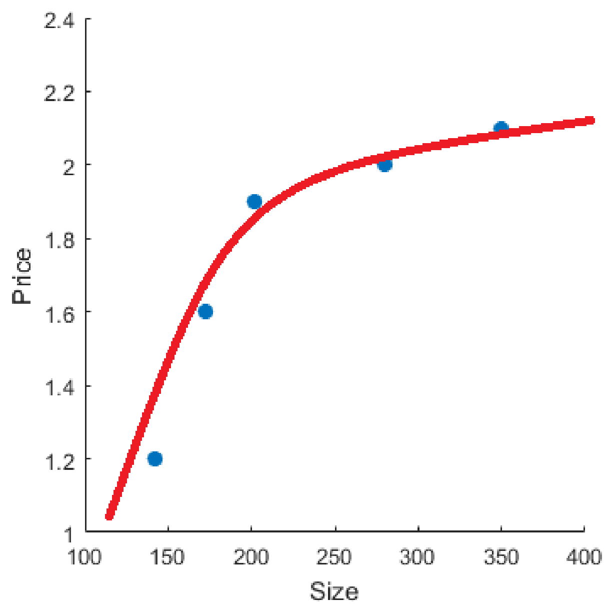
Price ($m)	Size (m^2)
2.0	280
1.2	142
1.6	172
1.9	202
2.1	350

Underfitting occurs when the model is overly simple.

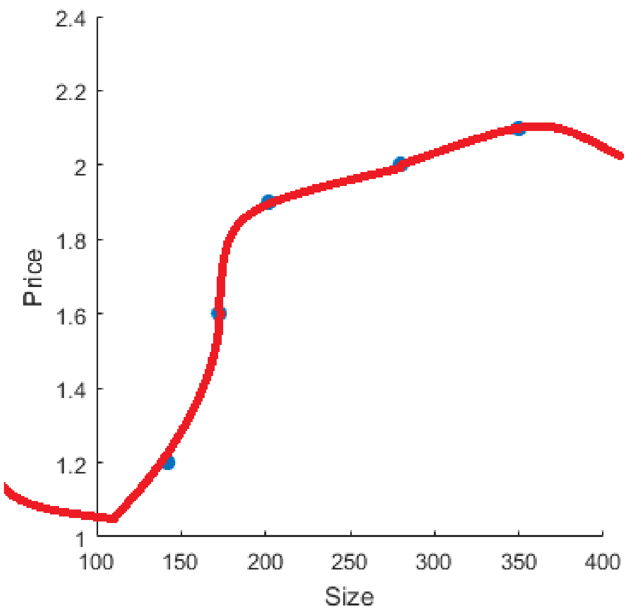
Overfitting occurs when the model is overly complex.



$\theta_0 + \theta_1 x$
(Probably underfitting)



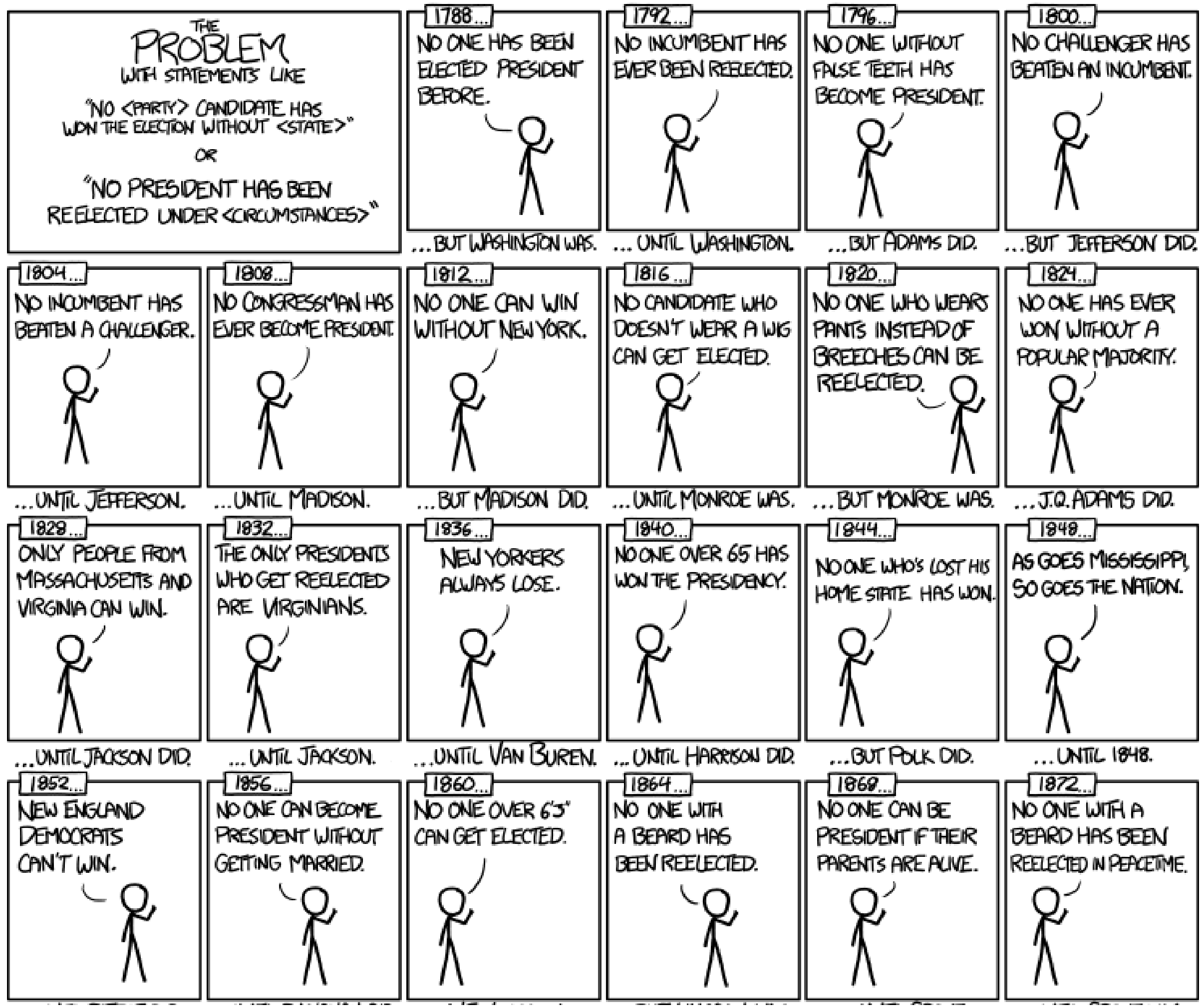
$\theta_0 + \theta_1 x + \theta_2 x^2$
(Looks good)



$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$
(Overfitting)

Overfitting and Electoral Precedent

- An example of how overfitting can creep into everyday conversation
- Full cartoon can be viewed at https://xkcd.com/1122/



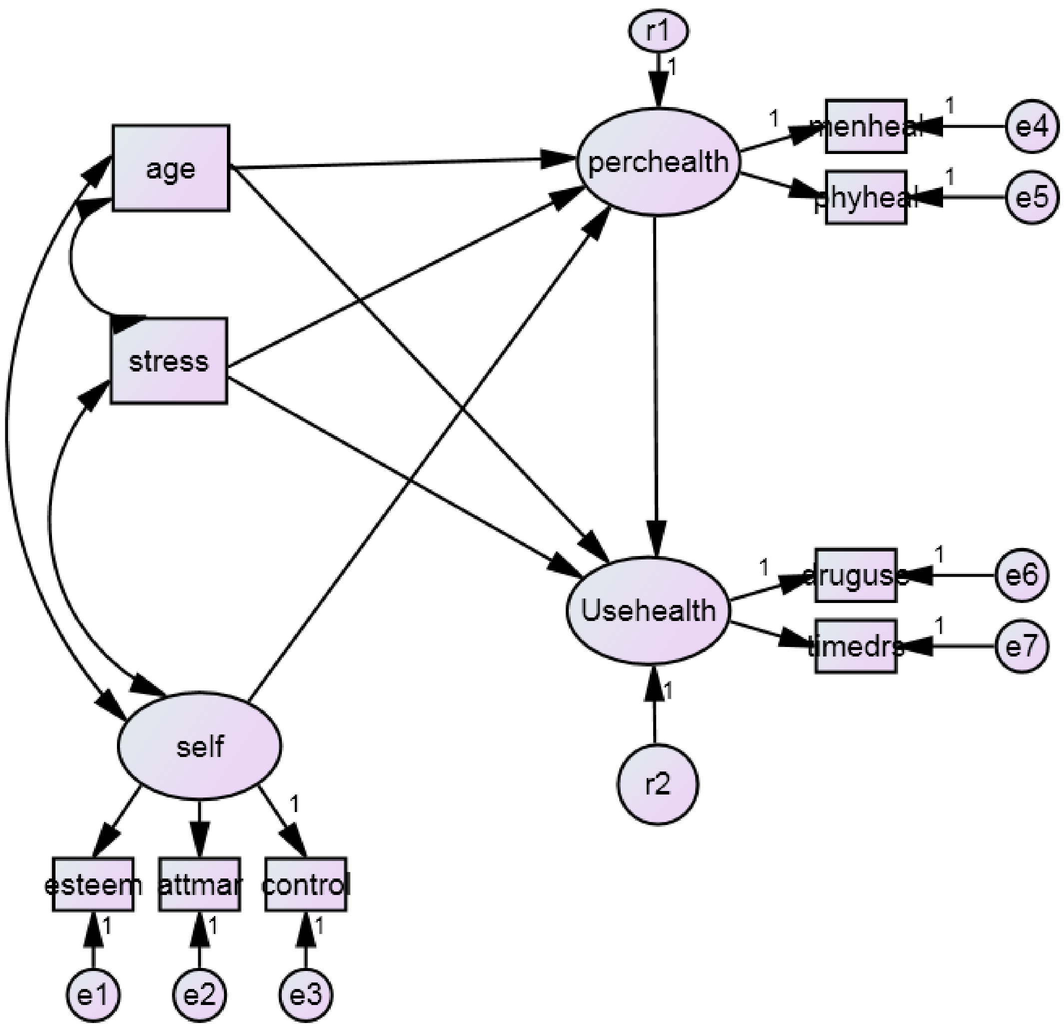
Section 4

PATH MODELS WITH LATENT VARIABLES

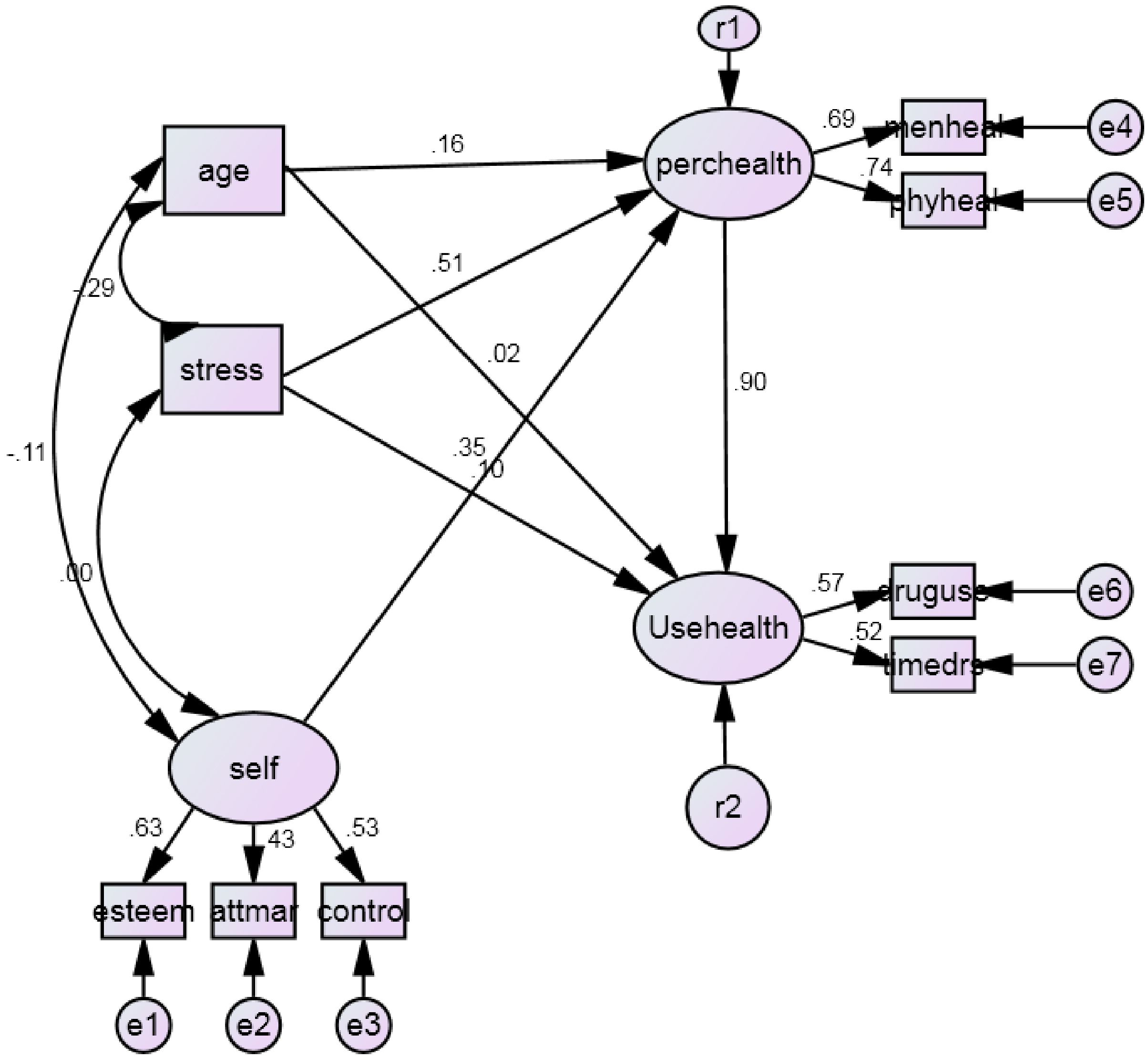
Our example:
Women's health study

- Survey of 465 females on a variety of health issues (Hoffman & Fidell, 1979 – reported in Tabachnik & Fidell, 2013).
- The research question involves the usage of health facilities.
- Now we will assume latent constructs: sense of self (*self*), Perception of health (*perchealth*), Use of health facilities (*Usehealth*)
- Measured by:
  - *Self: Self* esteem (*esteem*); Attitudes to marriage (*attmar*); Locus of control (*control*)
  - *Perchealth*: Self reported mental health problems (*menhealth*); physical health problems (*phyhealth*)
  - *Usehealth:* use of medicines (*druguse*); visits to health professionals (*timedrs*)
- Control variables:
  - Age, Stress will be used as manifest control variables

The theoretical model



Standardized estimates



Model Fit

Chi-square = 99.942
Degrees of freedom = 20
Probability level = .000

SRMR is OK, but CFI and RMSEA do not look great, So this is not a well fitting model.

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.858	.745	.883	.785	.881
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.093	.076	.112	.000
Independence model	.202	.189	.215	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	149.942	151.058	253.168	278.168
Saturated model	90.000	92.009	275.807	320.807
Independence model	723.531	723.933	760.693	769.693

Default model
Standardized RMR = .0592

Modification indices: Improving the model?

Direct paths

			M.I.	Par Change
Usehealth	<---	self	9.150	-1.739
timedrs	<---	phyheal	5.822	.425
timedrs	<---	menheal	4.980	-.223
druguse	<---	self	5.300	-1.663
druguse	<---	esteem	9.948	-.297
esteem	<---	age	5.661	.179
esteem	<---	stress	9.344	-.004
esteem	<---	Usehealth	4.921	-.083
esteem	<---	druguse	12.110	-.064
attmar	<---	stress	7.072	.008
attmar	<---	menheal	6.072	.231
control	<---	age	4.241	-.051
control	<---	menheal	8.036	.037
phyheal	<---	timedrs	12.471	.031
phyheal	<---	attmar	5.844	-.024
menheal	<---	age	4.202	-.147
menheal	<---	stress	4.220	.003
menheal	<---	self	21.841	1.442
menheal	<---	timedrs	8.436	-.046
menheal	<---	esteem	12.954	.145
menheal	<---	attmar	13.633	.066
menheal	<---	control	20.775	.574

Largest MIs often involve *self* and the observed measures

Theoretically, the most interesting MI here is whether *self* directly predicts *Usehealth*.

Modification indices: Improving the model?

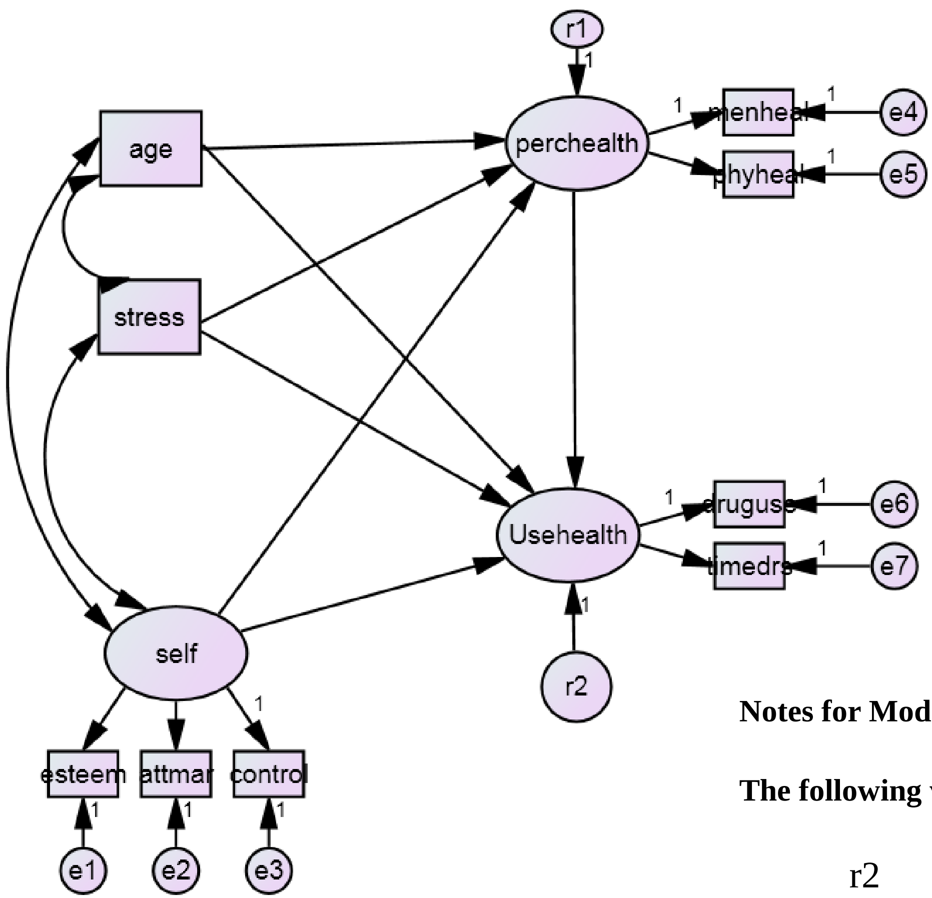
Correlations

			M.I.	Par Change
r2	<-->	self	9.357	-.792
e6	<-->	self	5.456	-.760
e1	<-->	stress	6.086	-51.327
e1	<-->	r2	8.195	-3.043
e1	<-->	e6	9.548	-4.125
e2	<-->	stress	6.058	119.742
e5	<-->	r2	12.908	1.934
e5	<-->	e7	17.629	3.264
e5	<-->	e2	4.391	-1.549
e4	<-->	self	20.120	.623
e4	<-->	r1	8.500	-1.295
e4	<-->	r2	6.694	-2.545
e4	<-->	e7	12.352	-4.958
e4	<-->	e1	4.062	1.151
e4	<-->	e2	6.261	3.343
e4	<-->	e3	11.686	.636

Indications here that some important correlations among measures might be missing.

Correlations among the residuals of measures induces correlations among latent outcome variables

Improve the model? Try *self->usehealth*



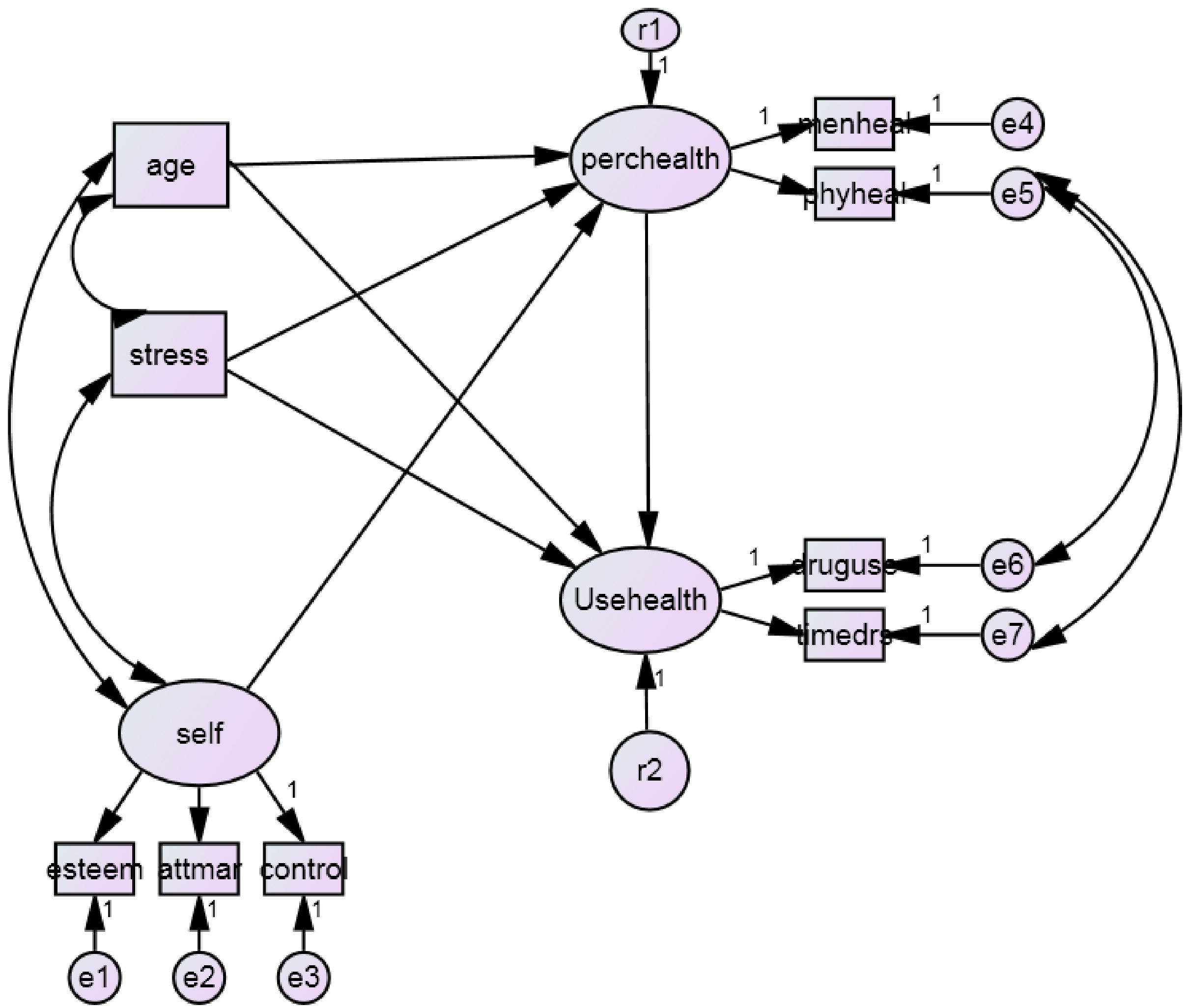
Notes for Model (Group number 1 - Default model)

The following variances are negative. (Group number 1 - Default model)

r2
-3.262

Uh-oh! We don't want negative variances, this model is NOT good!

One solution provided in T&F: Correlations among residuals



Model fit

Chi-square = 44.721
Degrees of freedom = 18
Probability level = .000

Previous chi-square = 99.9
Improvement of 55.2 at cost of 2 df
This is highly significant!
This is a much better fitting model.
This type of comparison can done for nested models.

SRMR, CFI and RMSEA are all now acceptable.
AIC has decreased from 150 to 99.

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.937	.873	.961	.920	.960
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

RMSEA

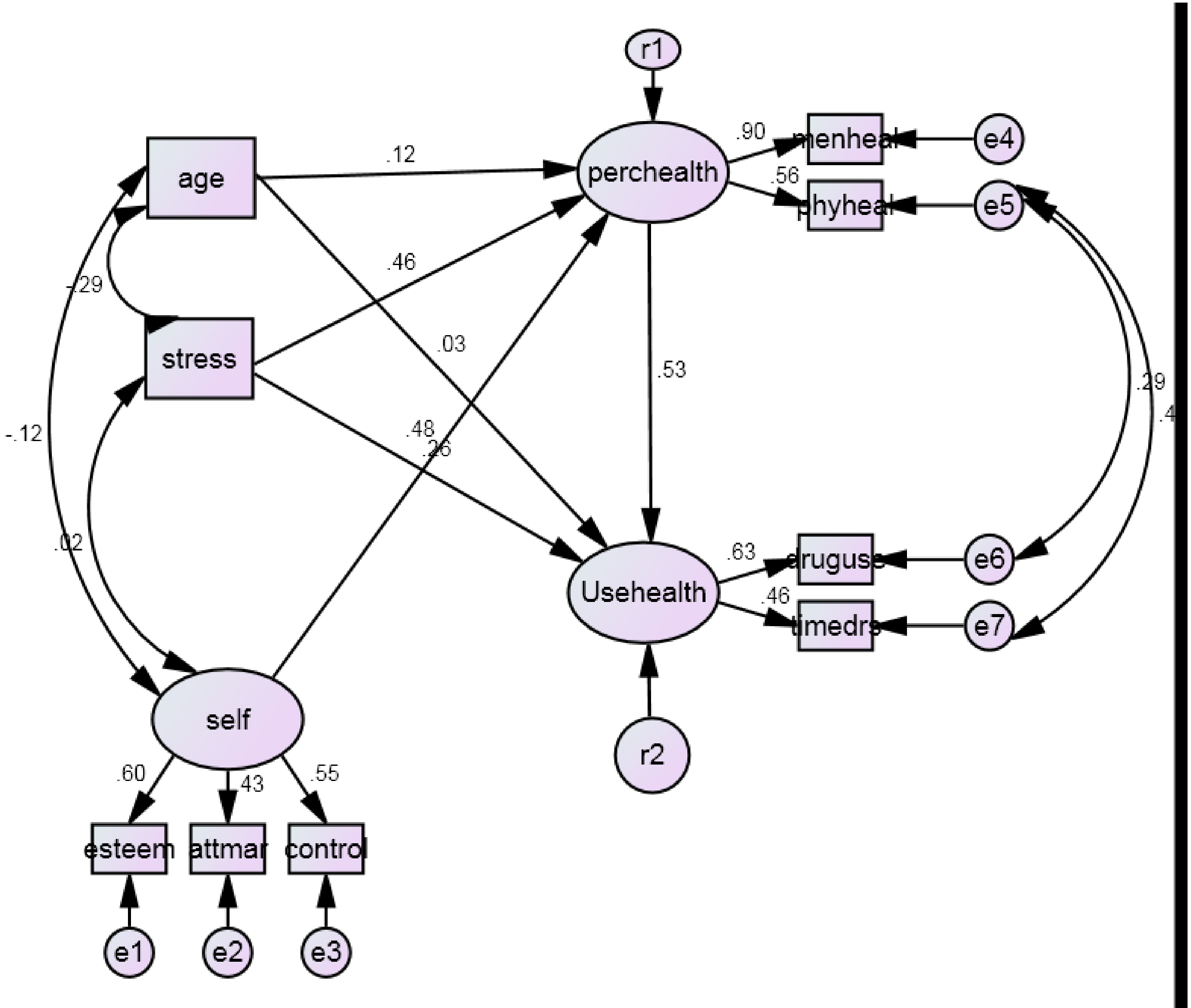
Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.057	.036	.078	.268
Independence model	.202	.189	.215	.000

AIC

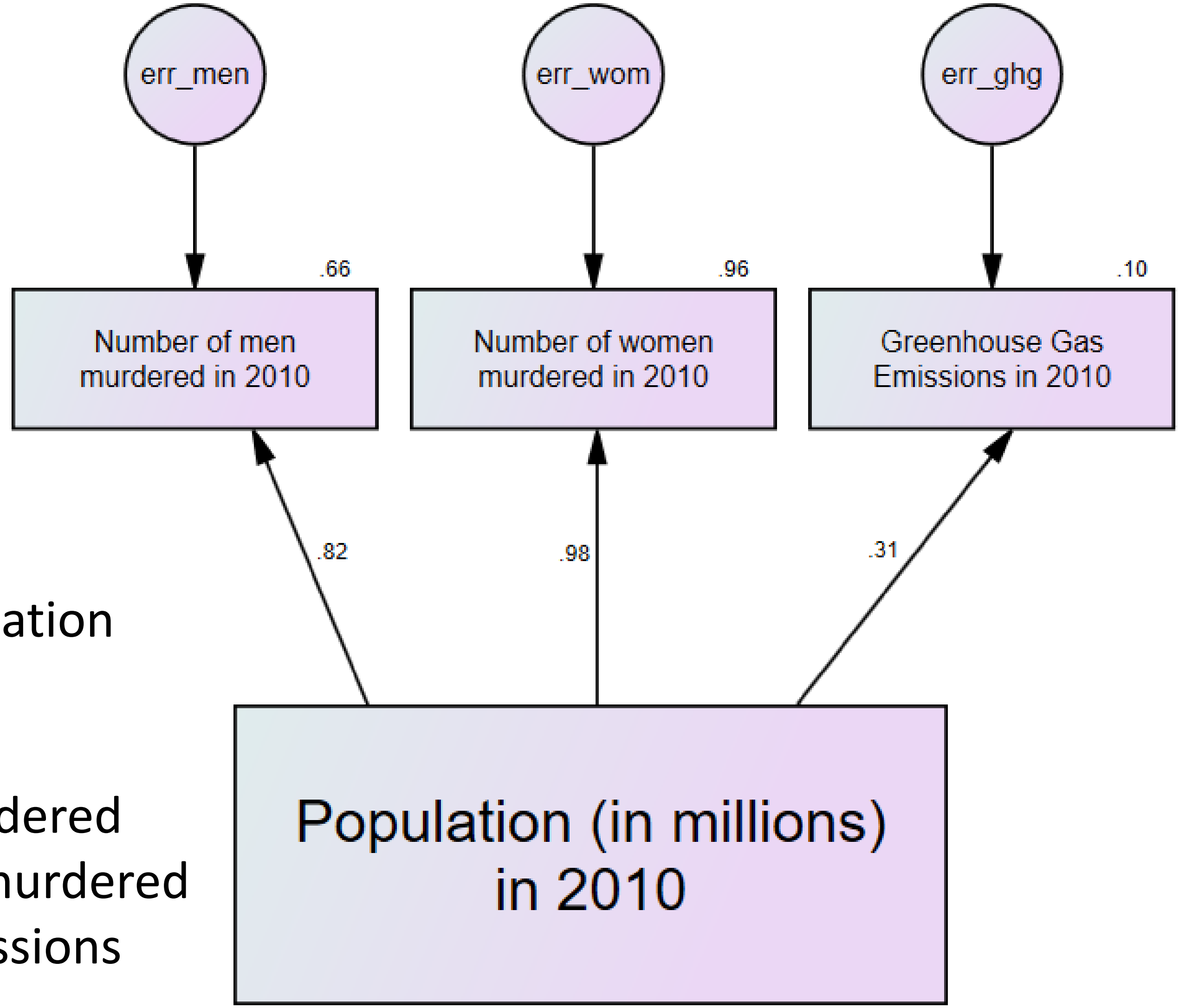
Model	AIC	BCC	BIC	CAIC
Default model	98.721	99.926	210.205	237.205
Saturated model	90.000	92.009	275.807	320.807
Independence model	723.531	723.933	760.693	769.693

Default model
Standardized RMR = .0442

Standardized parameter estimates



When might it make sense to correlate errors?



A cheesy example

Using a country's population to predict:

1. Number of men murdered
2. Number of women murdered
3. Greenhouse gas emissions

The model fits very poorly

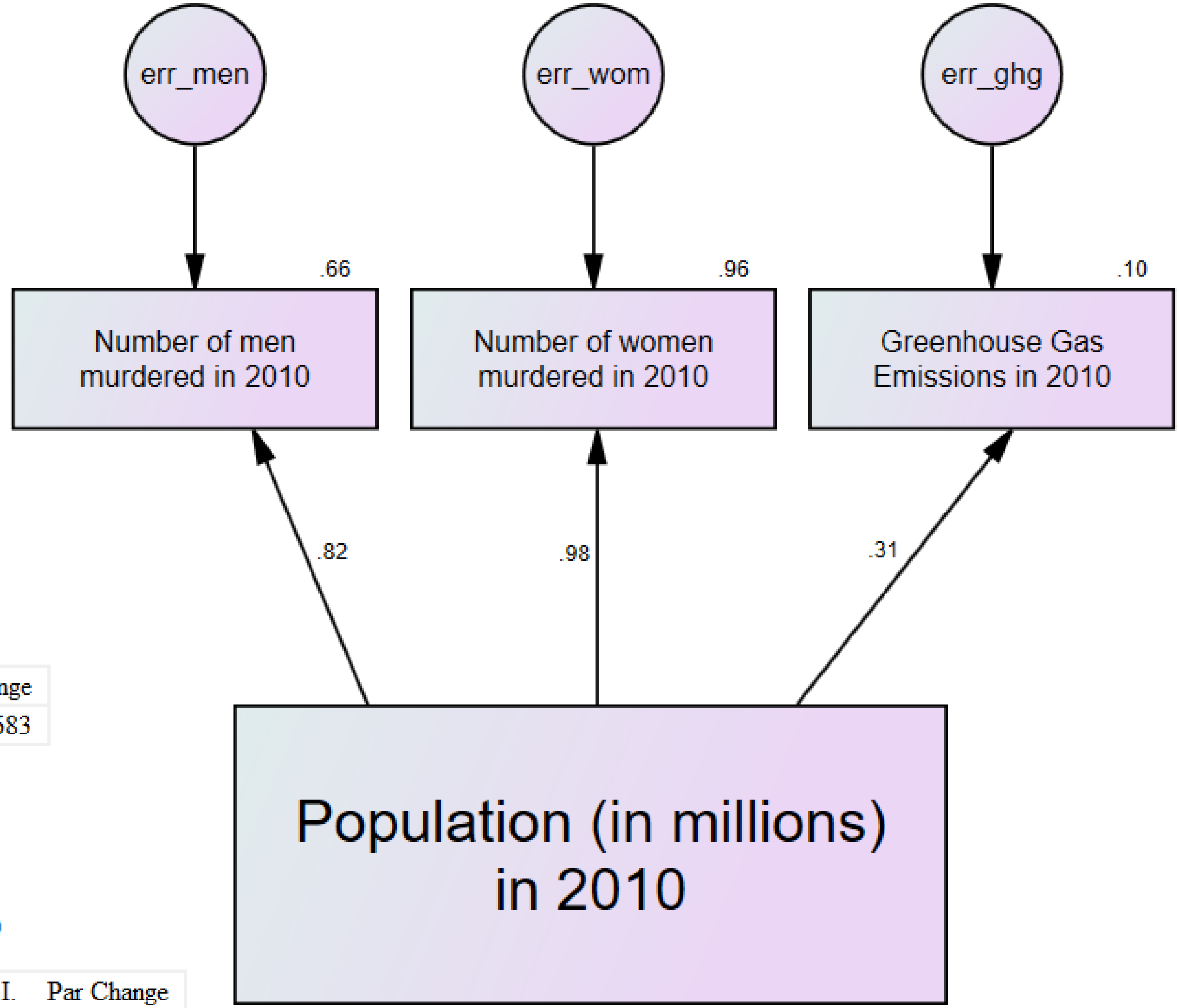
Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	10
Number of distinct parameters to be estimated:	7
Degrees of freedom (10 - 7):	3

Result (Default model)

Minimum was achieved
Chi-square = 36.185
Degrees of freedom = 3
Probability level = .000



Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

	M.I.	Par Change
err_men <--> err_wom	23.756	599369.683

Variances: (Group number 1 - Default model)

	M.I.	Par Change

Regression Weights: (Group number 1 - Default model)

	M.I.	Par Change
women_murdered <--- men_murdered	7.958	.038

A somewhat better fitting model

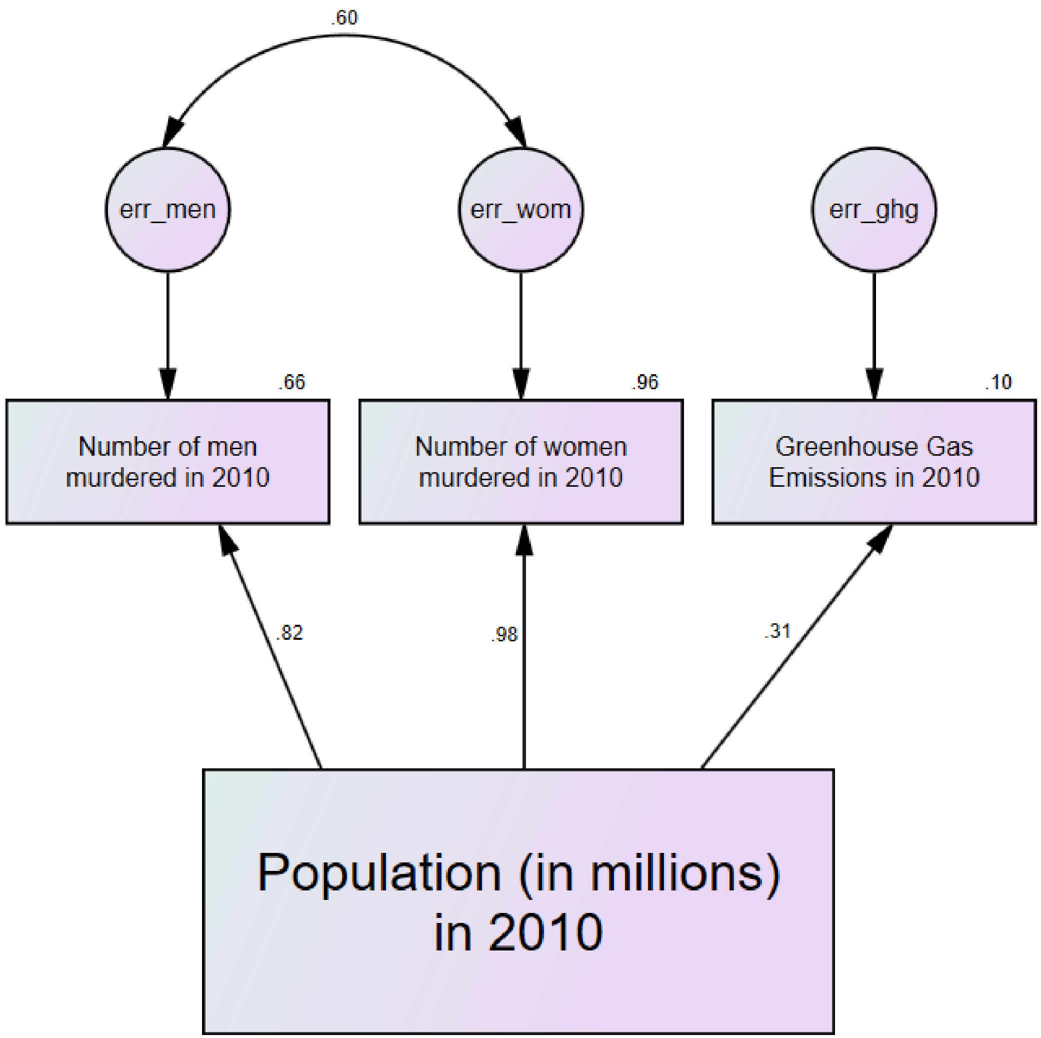
Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	10
Number of distinct parameters to be estimated:	8
Degrees of freedom (10 - 8):	2

Result (Default model)

Minimum was achieved
Chi-square = 6.736
Degrees of freedom = 2
Probability level = .034



What caused the difference?

If the model's prediction is too high for males murdered, it will also tend to be too high for females murdered.

If the model's prediction is too low for males murdered, it will also tend to be too low for females murdered.

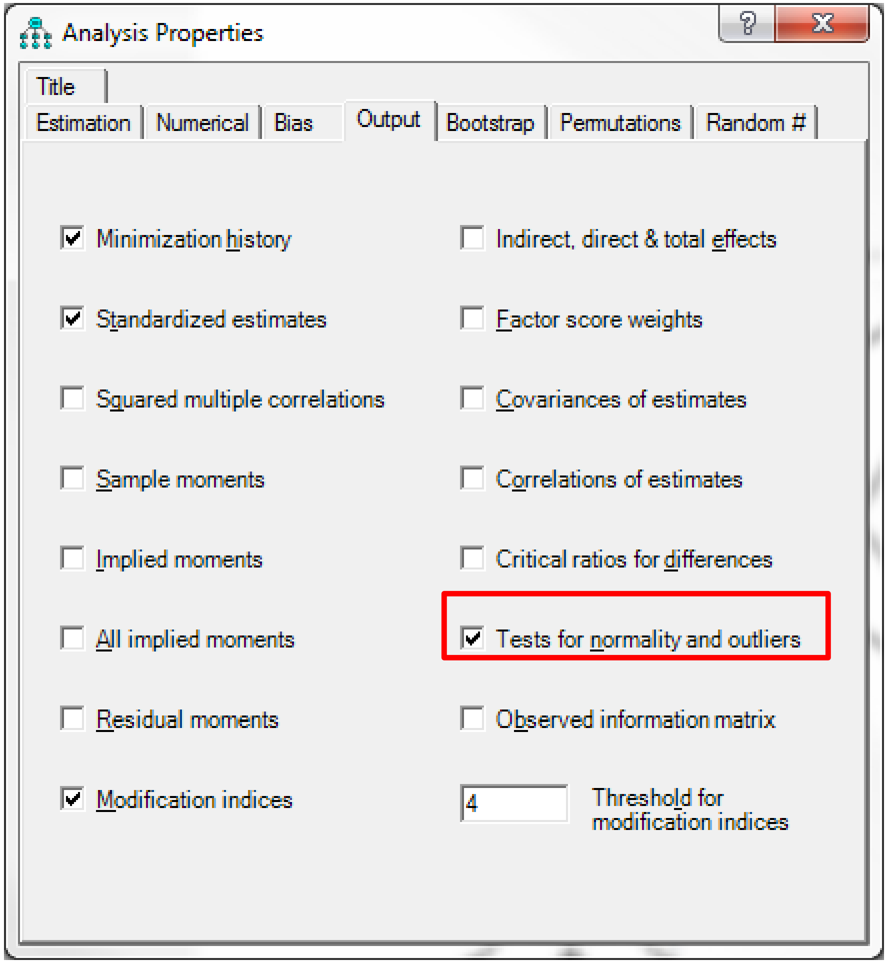
The point is illustrated by comparing Australia (AUS) and South Africa (RSA)

Nation	Population (Millions)	Males murdered	Males predicted to be murdered	Residual (Males)	Females Murdered	Females predicted to be murdered	Residual (Females)
AUS	22.30	141	890.44	-749.44	89	207.98	-118.98
RSA	49.99	13182	1498.10	11683.90	2758	598.23	2159.77

Section 5

BOOTSTRAPPING

Let’s check normality assumptions



The critical statistic is kurtosis. According to https://stat.utexas.edu/software-faqs/amos

> “Practically, very small multivariate kurtosis values (e.g., less than 1.00) are considered negligible while values ranging from one to ten often indicate moderate non-normality. Values that exceed ten indicate severe non-normality.”

Normality checks

Assessment of normality (Group number 1)

Variable	min	max	skew	c.r.	kurtosis	c.r.
age	.000	8.000	.037	.326	-1.162	-5.084
stress	.000	643.000	.764	6.680	.244	1.065
timedrs	.000	60.000	2.904	25.397	9.997	43.718
druguse	.000	42.000	1.268	11.092	1.062	4.644
esteem	8.000	29.000	.487	4.260	.282	1.234
attmar	.000	58.000	.794	6.942	.867	3.791
control	5.000	10.000	.491	4.296	-.398	-1.740
phyheal	2.000	15.000	1.059	9.265	1.227	5.367
menheal	.000	18.000	.617	5.401	.261	1.139
Multivariate					23.723	18.060

Some big problems with normality, including multivariate kurtosis

Observations farthest from the centroid (Mahalanobis distance) (Group

Observation number	Mahalanobis d-squared	p1	p2
167	44.010	.000	.001
277	42.301	.000	.000
205	39.212	.000	.000
273	38.552	.000	.000
39	37.792	.000	.000
247	36.555	.000	.000
364	33.057	.000	.000
369	31.957	.000	.000
385	31.359	.000	.000
112	30.606	.000	.000
387	30.362	.000	.000
162	29.330	.001	.000
192	28.439	.001	.000
211	26.447	.002	.000
217	25.589	.002	.000
245	24.358	.004	.000
275	23.522	.005	.000
[illegible]	[illegible]	.005	.000
[illegible]	[illegible]	.006	.000
[illegible]	[illegible]	.006	.000
[illegible]	[illegible]	.009	.000
[illegible]	[illegible]	.014	.000
[illegible]	[illegible]	.015	.000
[illegible]	[illegible]	.018	.000
[illegible]	[illegible]	.021	.000
[illegible]	[illegible]	.022	.000
[illegible]	[illegible]	.026	.000

Mahalanobis distance
p1 … probability that any observation could be so far out
– want this small
p2 … probability that this particular case should be so far out
– want this large
– here lots of observations with small p2, so lots of outliers

What happens when assumptions not met?

- Model can be incorrectly rejected as not fitting
- Standard errors will be smaller than they really are (*i.e.,* parameters may seem significant when they are not)
- Solve these problems through bootstrapping

Bootstrap

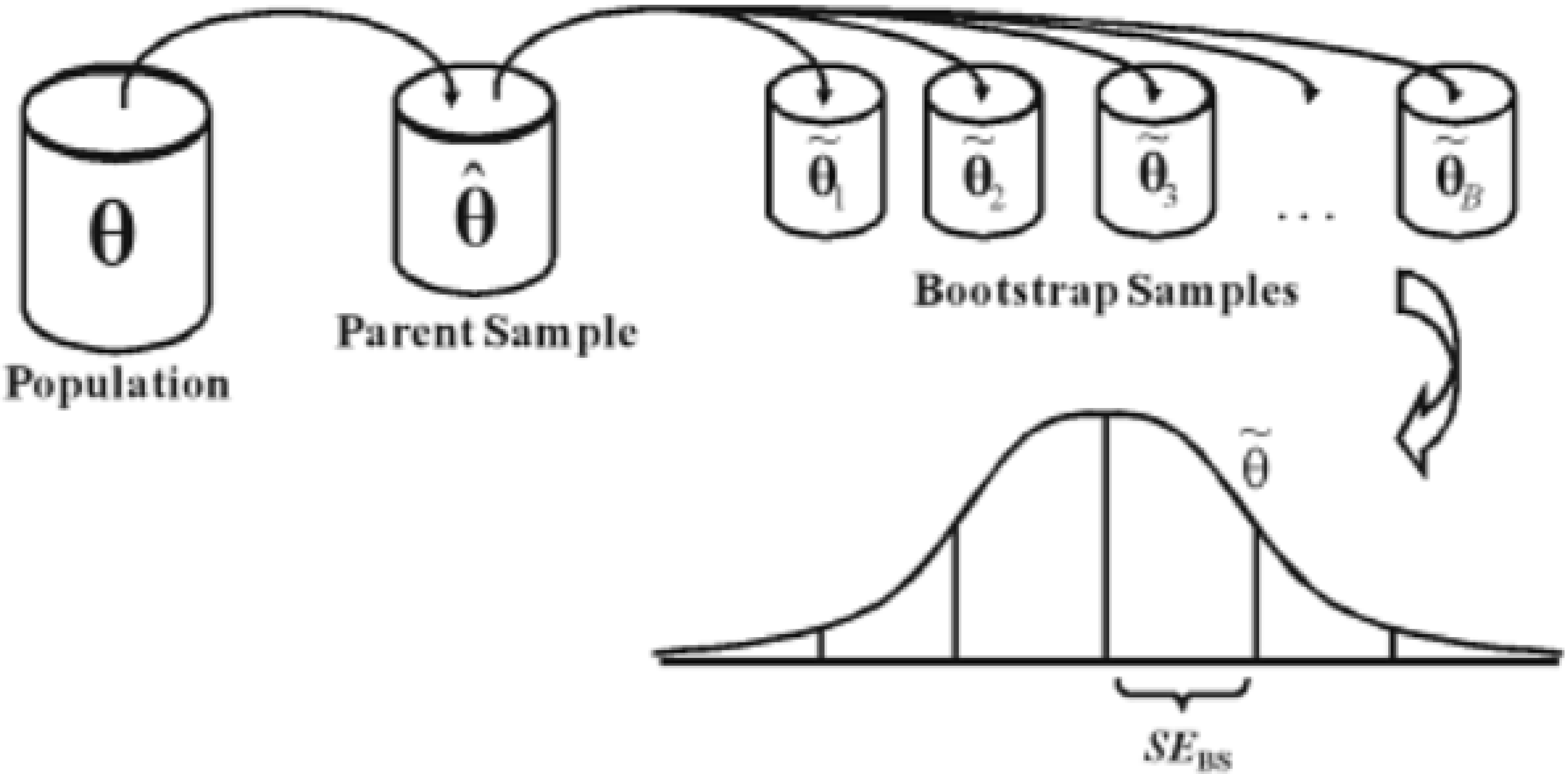
- Takes repeated samples of the data
- Allow each observation to be included more than once in any sample
- Each observation could represent any number of similar cases in the population.
- This can be thought of as ‘sampling with replacement’.
- Calculate the statistic for each boot-strapped sample to produce a bootstrap distribution.
- Calculate a standard error as the standard deviation of the bootstrap distribution.

Using the bootstrap in Amos

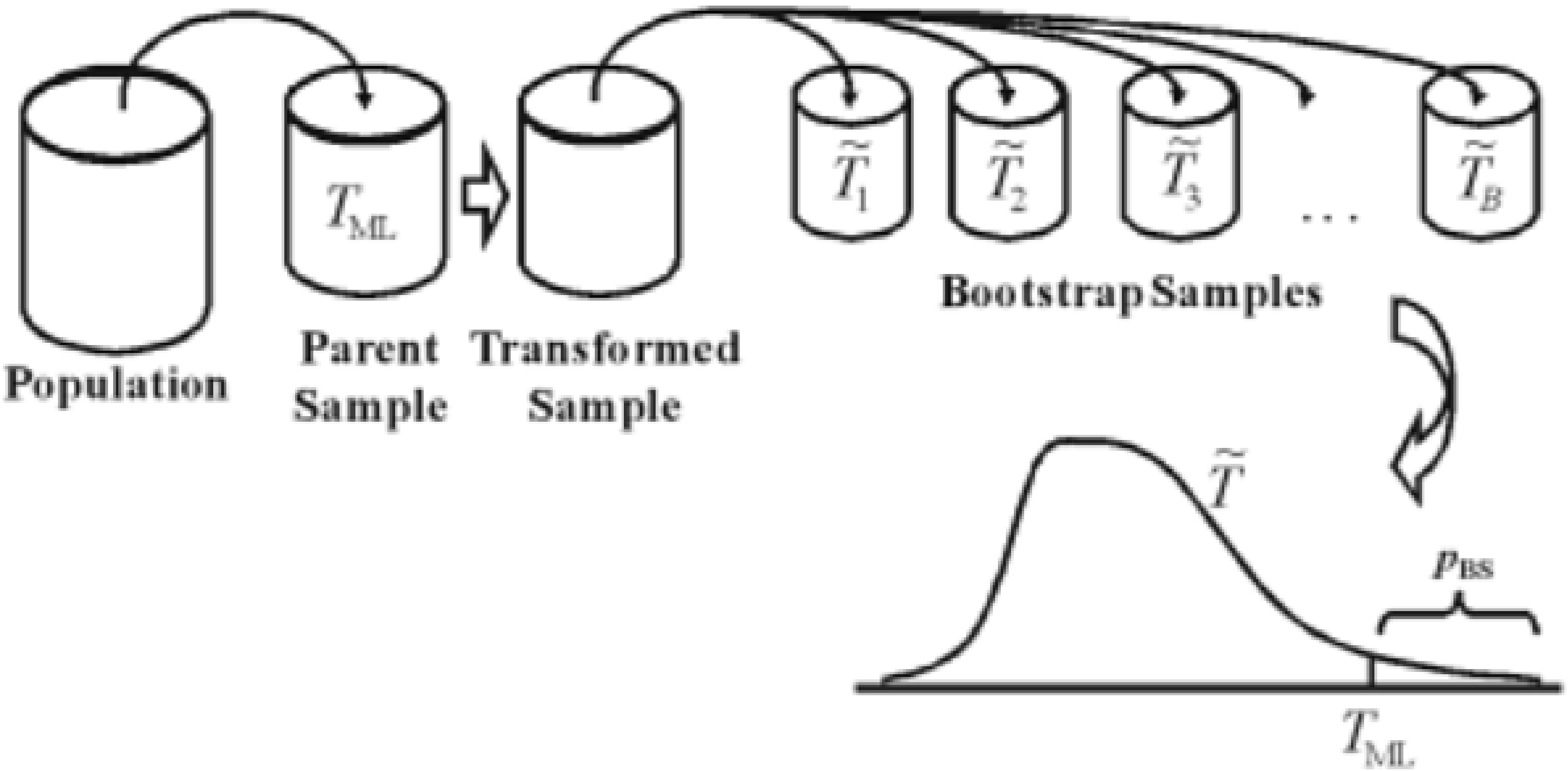
- When distribution not normal
  - To assess overall fit: Bollen-Stine test (Bollen & Stine, 1992)
  - To obtain accurate standard errors
- Unfortunately cannot be done at the same time
  - Need to do two analysis runs

An excellent webpage on bootstrapping in Amos is at:
https://stat.**utexas**.edu/software-**faqs/amos**

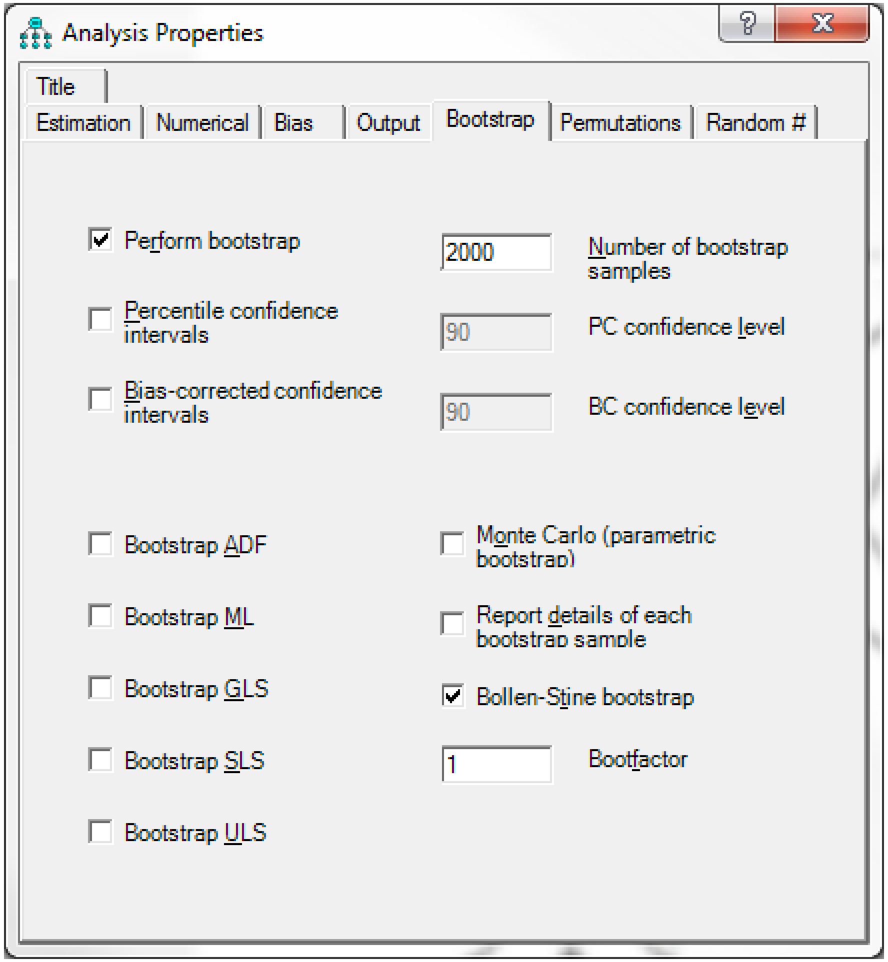
Naïve Bootstrapping



Bollen-Stine bootstrapping



1. Bollen-Stine bootstrap for overall fit



Summary of Bootstrap Iterations (Default model)

(Default model)

Iterations	Method 0	Method 1	Method 2
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	3
5	0	0	5
6	0	0	4
7	0	3	0
8	0	39	0
9	0	138	0
10	0	319	0
11	0	386	0
12	0	392	0
13	0	280	0
14	0	188	0
15	0	124	0
16	0	63	0
17	0	30	0
18	0	12	0
19	0	14	0
Total	0	1988	12

0 bootstrap samples were unused because of a singular covariance matrix.
0 bootstrap samples were unused because a solution was not found.
2000 usable bootstrap samples were obtained

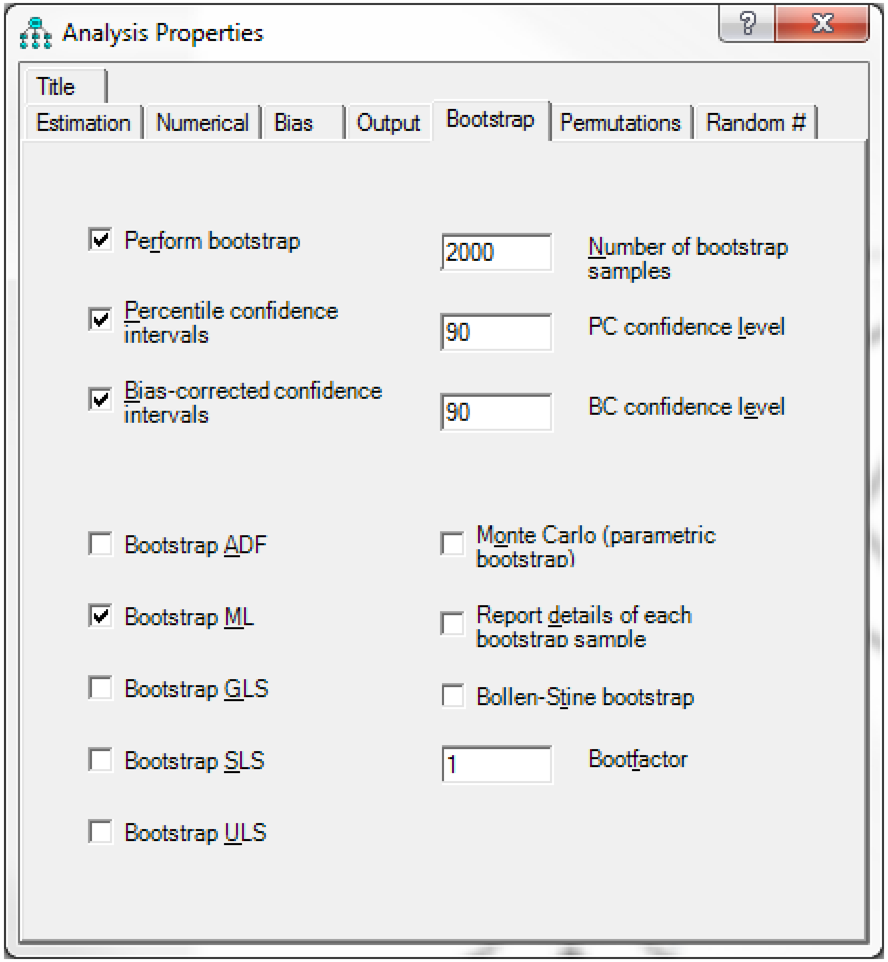
some diagnostic information.
Method 1 is the standard
Method 2 is used if method 1 fails.
Method 0 does not exist

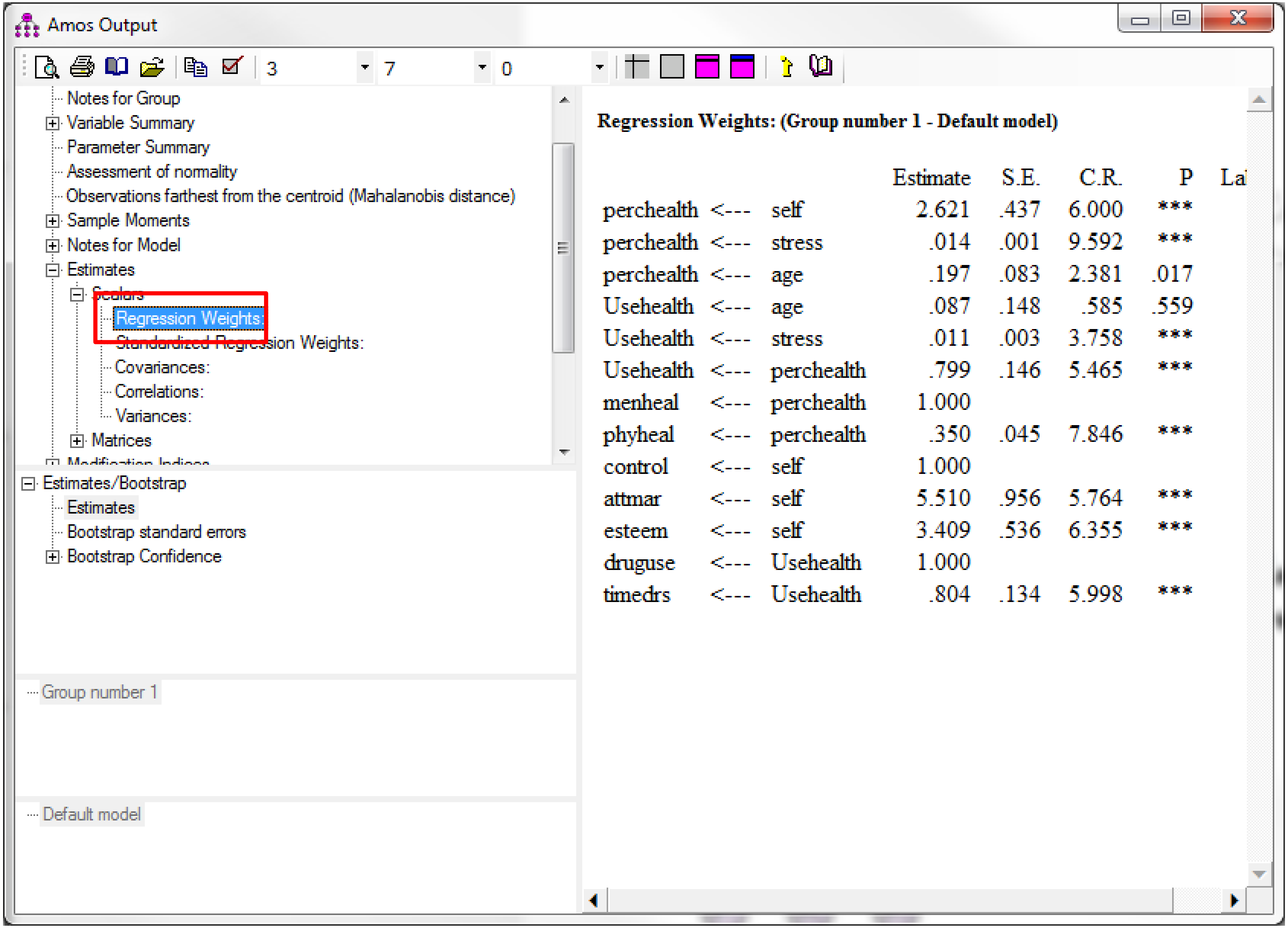
Bollen-Stine Bootstrap (Default model)

The model fit better in 1991 bootstrap samples.
It fit about equally well in 0 bootstrap samples.
It fit worse or failed to fit in 9 bootstrap samples.
Testing the null hypothesis that the model is correct, Bollen-Stine bootstrap $\boxed{p = .005}$

Uh-oh! Fit isn't as good as we thought!

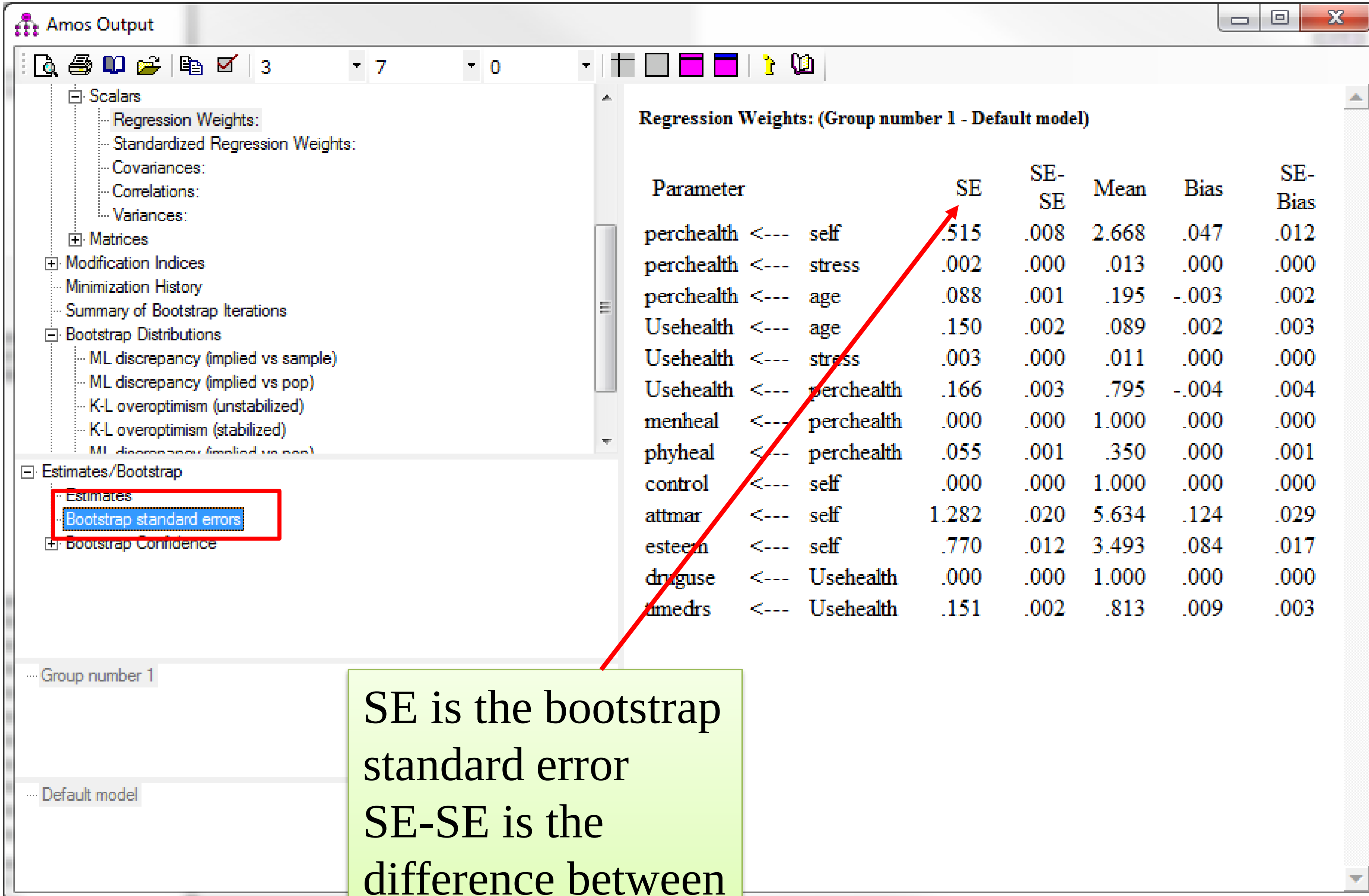
2. ML bootstrap for path estimate confidence intervals





Regression Weights: (Group number 1 - Default model)

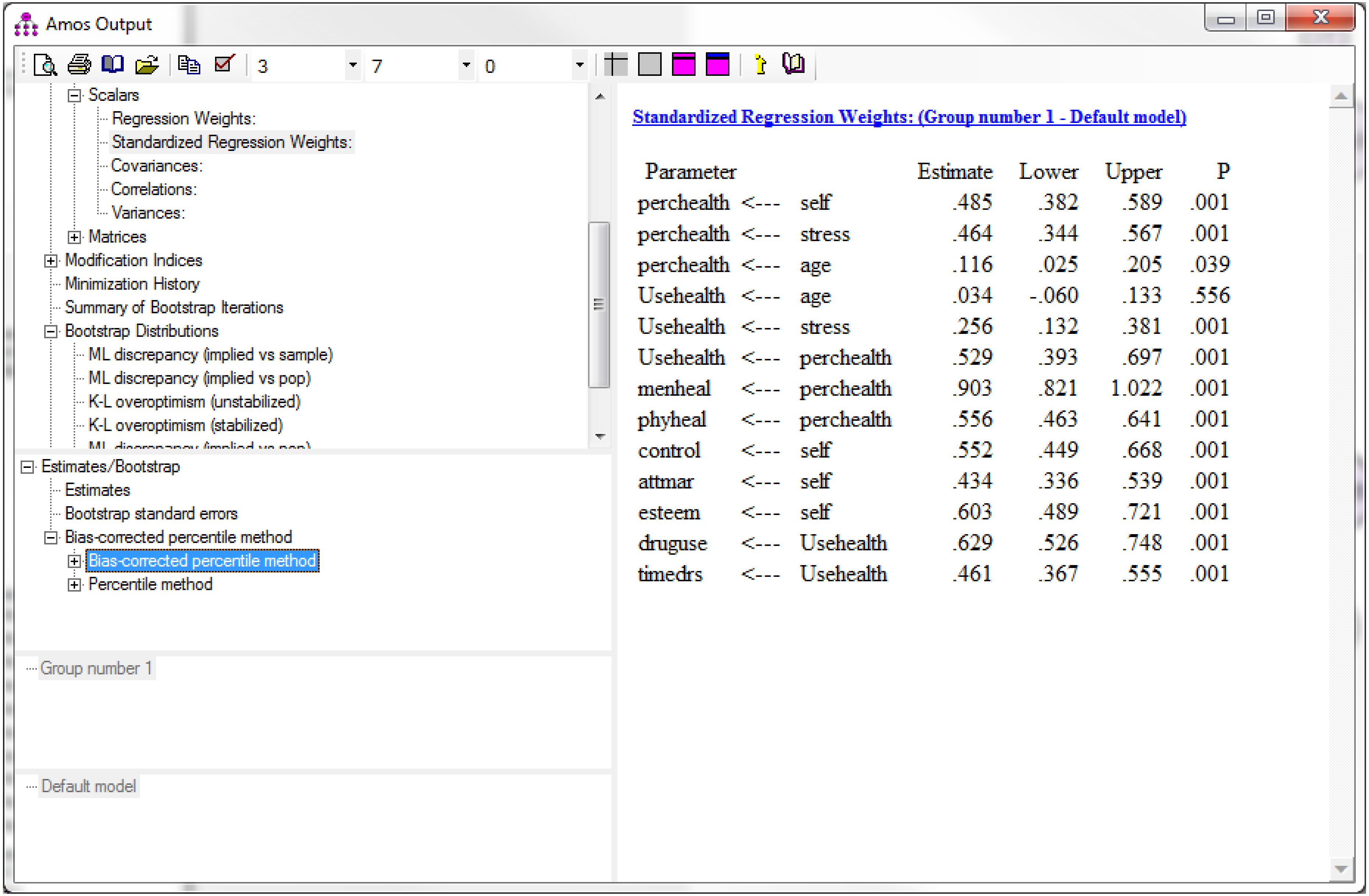
			Estimate	S.E.	C.R.	P	La
perchealth	<---	self	2.621	.437	6.000	***	
perchealth	<---	stress	.014	.001	9.592	***	
perchealth	<---	age	.197	.083	2.381	.017	
Usehealth	<---	age	.087	.148	.585	.559	
Usehealth	<---	stress	.011	.003	3.758	***	
Usehealth	<---	perchealth	.799	.146	5.465	***	
menheal	<---	perchealth	1.000				
phyheal	<---	perchealth	.350	.045	7.846	***	
control	<---	self	1.000				
attmar	<---	self	5.510	.956	5.764	***	
esteem	<---	self	3.409	.536	6.355	***	
druguse	<---	Usehealth	1.000				
timedrs	<---	Usehealth	.804	.134	5.998	***	



Regression Weights: (Group number 1 - Default model)

Parameter			SE	SE-SE	Mean	Bias	SE-Bias
perchealth	<---	self	.515	.008	2.668	.047	.012
perchealth	<---	stress	.002	.000	.013	.000	.000
perchealth	<---	age	.088	.001	.195	-.003	.002
Usehealth	<---	age	.150	.002	.089	.002	.003
Usehealth	<---	stress	.003	.000	.011	.000	.000
Usehealth	<---	perchealth	.166	.003	.795	-.004	.004
menheal	<---	perchealth	.000	.000	1.000	.000	.000
phyheal	<---	perchealth	.055	.001	.350	.000	.001
control	<---	self	.000	.000	1.000	.000	.000
attmar	<---	self	1.282	.020	5.634	.124	.029
esteem	<---	self	.770	.012	3.493	.084	.017
druguse	<---	Usehealth	.000	.000	1.000	.000	.000
timedrs	<---	Usehealth	.151	.002	.813	.009	.003

SE is the bootstrap standard error
SE-SE is the difference between bootstrap & usual



Standardized Regression Weights: (Group number 1 - Default model)

Parameter			Estimate	Lower	Upper	P
perchealth	<---	self	.485	.382	.589	.001
perchealth	<---	stress	.464	.344	.567	.001
perchealth	<---	age	.116	.025	.205	.039
Usehealth	<---	age	.034	-.060	.133	.556
Usehealth	<---	stress	.256	.132	.381	.001
Usehealth	<---	perchealth	.529	.393	.697	.001
menheal	<---	perchealth	.903	.821	1.022	.001
phyheal	<---	perchealth	.556	.463	.641	.001
control	<---	self	.552	.449	.668	.001
attmar	<---	self	.434	.336	.539	.001
esteem	<---	self	.603	.489	.721	.001
druguse	<---	Usehealth	.629	.526	.748	.001
timedrs	<---	Usehealth	.461	.367	.555	.001

IN THIS LECTURE, you learnt

- how SEM is akin to a combination of regression and CFA
- how to conduct multi-stage regressions (path analyses) in AMOS
- about the different types of SEM indices of model fit
- why overfitting is evil, and a bit about how to detect it
- how to fit a latent variable model in AMOS, and how to use modification indices to improve the model.
- about bootstrapping methods that can be applied in AMOS when assumptions of normality are not met.

References

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, 21(2), 205-229.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Newbury Park, CA Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford Press.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–236.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics.* Pearson.

Wright, S. (1921). Correlation and causation. *Journal of agricultural research*, *20*, 557-585.

Wright, S. (1934). The method of path coefficients. *The annals of mathematical statistics*, *5*, 161-215.