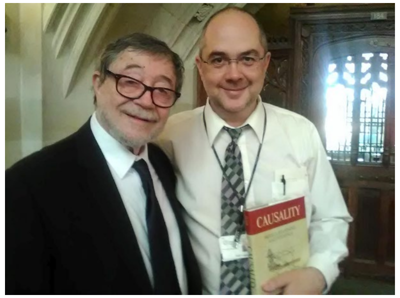
Restructuring Basic Statistical Curricula: Mixing Older Analytic Methods with Modern Software Tools in Psychological Research

Modern Modeling Methods – 2024, Storrs CT,
June 25-26, 2024 **Session 8B**

Slides at https://tinyurl.com/mmmintrostat



Emil Coman Pstat, SEMNET ‘moderator’; James Jaccard; Sabrina Uva;
Ana-Maria Cazan

comanus@gmail.com

tinyurl.com/agecause

knowledge as a human right

Humanity, Earth, Milky Way, Universe

Goals

Measurable Action Verbs for Writing Learning Objectives

"By the end of this course/module, you should be able to..."

Lower Order Thinking Skills → → → Higher Order Thinking Skills

I. Remembering	II. Understanding	III. Applying	IV. Analyzing	V. Evaluating	VI. Creating
Retrieve relevant knowledge from long-term memory.	Construct meaning from instructional messages, including oral, written, and graphic communication.	Carry out or use a procedure in a given situation.	Break material into its constituent parts and determine how the parts relate to-one another and to an overall structure or purpose.	Make judgments based on criteria and standards	Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure.

Verbs		Verbs		Verbs		Verbs		Verbs		Verbs	
Arrange	Show	Associate	Order	Apply	Deduce	Analyze	Deduce	Assess	Defend	Adapt	Design
Choose	Recall	Cite	Outline	Operate	Record	Categorize	Survey	Appraise	Support	Invent	Propose
Copy	Recite	Classify	Relate	Administer	Devise	Compare	Devise	Award	Determine	Build	Develop
Define	Record	Compare	Rephrase	Calculate	Show	Classify	Test	Judge	Validate	Maximize	Reframe
Duplicate	Relate	Contrast	Report	Perform	Demonstrate	Prioritize	Dissect	Justify	Disprove	Combine	Elaborate
Enumerate	Repeat	Discuss	Research	Collect	Sketch	Correlate	Distinguish	Convince	Dispute	Minimize	Rewrite
Find	Spell	Describe	Rewrite	Model	Determine	Simplify	Organize	Persuade	Estimate	Compile	Formulate
Label	State	Explain	Select	Compute	Teach	Contrast	Differentiate	Criticize	Evaluate	Modify	Simplify
List	Tabulate	Illustrate	Show	Prepare	Diagram			Critique	Influence	Compose	Improve
Locate	Tell	Indicate	Summarize	Connect	Utilize			Prioritize	Recommend	Construct	Solve
Name	Trace	Infer	Translate	Produce	Modify			Debate		Originate	Transform
Match	Select	Interpret		Construct	Solve			Prove		Create	
Memorize				Convert	Relate					Produce	

General points

Best to see and do it by yourself:
* Submitted paper http://**tinyurl.com/tracepath**
* All analyses results, instructions, and data are posted online http://**tinyurl.com/pathstats**

1. Models (as graphs, e.g.) happen before RQs and Hyp's
2. Research questions (RQ) or hypotheses (Hyp) are not p-hrased with statistical wording: no 'chi-square test' (or t-test) in them!
3. Statistics 'kicks in' after RQs and Hyp's are laid out.

1st step

Variables are :
1. Categorical
2. Continuous

i. "In this book, we will distinguish between two different types of variables. A categorical variable is a characteristic of an individual which can be broken down into different classes or categories.
Simple examples of a categorical variable are the eye color of a student, the political affiliation of a voter, the manufacturer of your current car, and the letter grade in a particular class.
Typically, a categorical variable is nonnumerical, although numbers are occasionally used in classification.
The social security number of a person is an example of a categorical variable, since its main purpose is to identify or classify individuals. Binary variables are categorical variables for which only two possible categories exist.
A measurement variable is a number associated with an individual that is obtained by means of some measurement.
Examples of a measurement variable include your age, your height, the weight of your car, and the distance that you traveled during your Thanksgiving vacation. A measurement variable will have a range of possible numerical values. A person's age, for example, ranges from 0 to approximately 100." (Albert & Rossman, 2001) p. 5
ii. "Throughout the text, I will use the phrase continuous for quantitative variables (even if they are not truly continuous in the sense of having all possible intermediate values between integers), and the phrase categorical for discrete, grouping variables (i.e., in which differences between specific levels are of interest, although those levels may or may not be ordered)." (Hoffman, 2015) p. 9

Albert, J. B., & Rossman, A. J. (2001). Workshop Statistics: Discovery with Data. A Bayesian approach https://drive.google.com/file/d/1ok2n3ju23wOenxws-g7hx7HGlZe9-X6f/view?usp=sharing: John Wiley & Sons.
Hoffman, L. (2015). Longitudinal analysis: Modeling within-person fluctuation and change: Routledge.

2nd step

Modeling variables, in the simplest manner

1. 2 only
2. 3

   + yes, more possibly

* Best way to proceed is the graphical view, which translates plain phrasing like ‘Gender → Education’ or ‘(Religiosity + Health.) → Anxiety’

  - Sewall Wright proposed the ‘chain rule’ called ‘path analysis’ to decompose relations into components:

1. Causal
2. Non-Causal

Wright, S. (1921). Correlation and causation. Part I Method of path coefficients. Journal of agricultural research, 20(7), 557-585.

Path analysis and the power of the ‘tracing rule’

“The correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients along all of the paths by which the variables are connected.

[…] A path coefficient differs from a coefficient of correlation in having direction.” [1]:114-115

Wright, S. (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics, 6(2), 111.*

What flows through a path network? ASSOCIATION

Evidence 1. Felix Elwert: DAG workshop

Evidence 2. Kline:

Estimation and Local Fit Testing 251

An alternative definition comes from Chen and Pearl (2015): A valid tracing does not involve colliding arrowheads, such as

Recall that paths blocked by a collider do not convey a statistical association between the variables at either end of the path, if the collider is not included among the covariates.

Kline 2015 4th [Principles and Practice of Structural Equation Modeling]

Path analysis and the 'tracing rule'

Wright's rules

Briefly, Wright showed that if a situation can be presented as a proper path diagram, then the correlation between any two variables in the diagram can be expressed as the *sum of the compound paths connecting these two points*, where a compound path is a path along arrows that follows three rules:

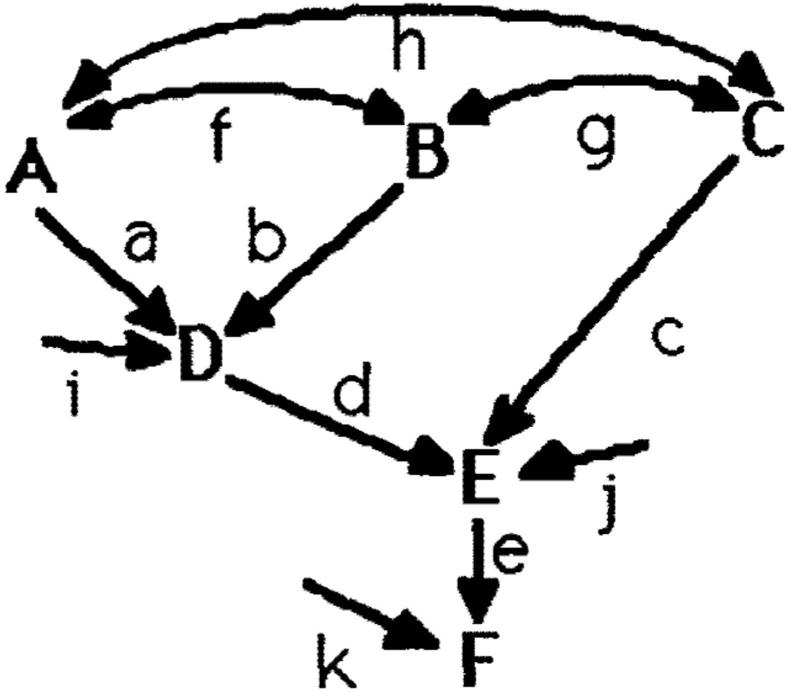
(a) no loops;
(b) no going forward then backward;
(c) a maximum of one curved arrow per path.

p. 8-9

Loehlin, J. C. (2004). Latent variable models: An introduction to factor, path, and structural equation analysis. Mahwah, NJ: Lawrence Erlbaum.

Path analysis and the ‘tracing rule’

Fig. 1.7 Examples of tracing paths in a path diagram.

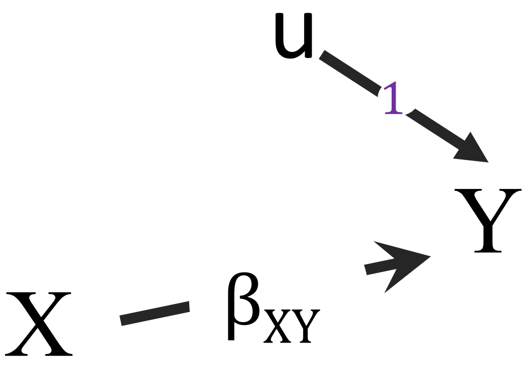


For example, what is the correlation between variables A and D in Fig. 1.7? Two paths are legal: *a* and *fb*. A path like *hgb* would be excluded by the rule about only one curved arrow, and paths going further down the diagram like *adcgb* would violate both the rules about no forward then backward and no loops. So the numerical value of r_{AD} can be expressed as $a + fb$. I hope that the reader can see that $r_{BD} = b + fa$, and that $r_{CD} = gb + ha$.

Loehlin, J. C. (2004). Latent variable models: An introduction to factor, path, and structural equation analysis. Mahwah, NJ: Lawrence Erlbaum.

How to get β for a ‘regression’ with 2 variables: X->Y

Regression



$$Y_i = \beta_{XY} \cdot X_i + 1 \cdot u_i$$

[*easier if* $\alpha_Y = 0$]

Hence if one multiplies by X_i :

$$X_i \cdot Y_i = \beta_{XY} \cdot X_i \cdot X_i + X_i \cdot u_i$$

Sum across N (sample cases) & divide by N:

$$\frac{\Sigma_i^N X_i \cdot Y_i}{N} = \beta_{XY} \cdot \frac{\Sigma_i^N X_i \cdot X_i}{N} + \frac{\Sigma_i^N X_i \cdot u_i}{N}$$

Hence :

$$\sigma_{YX} = \beta_{YX} \cdot \sigma^2_{XX} + \sigma_{Xu}$$

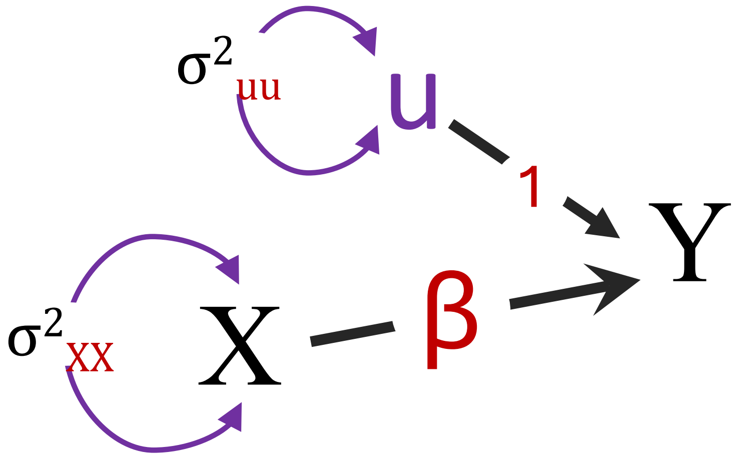
So:

$$\beta_{XY} = \frac{\sigma_{YX}}{\sigma^2_{XX}} - \sigma_{Xu} = \frac{Cov(Y,X)}{Cov(X,X)} - Cov(X,u)$$

With deviation scores one gets $\alpha_Y = 0$.

Notation: u is better here than ε because it represents ‘ignored-for-now-other-causes’, not just ‘error’.

Obtain β with Wright's tracing rule



Cov(YX) is "sum of products path/structural coefficients, of all open pathways from X to Y":

$$\text{Cov(YX)} \overset{notation}{\Longleftrightarrow} \sigma_{YX} \overset{Wright\ Tracing\ Rule}{\Longleftrightarrow} \sigma^2{}_{XX} \cdot \beta$$

Hence:

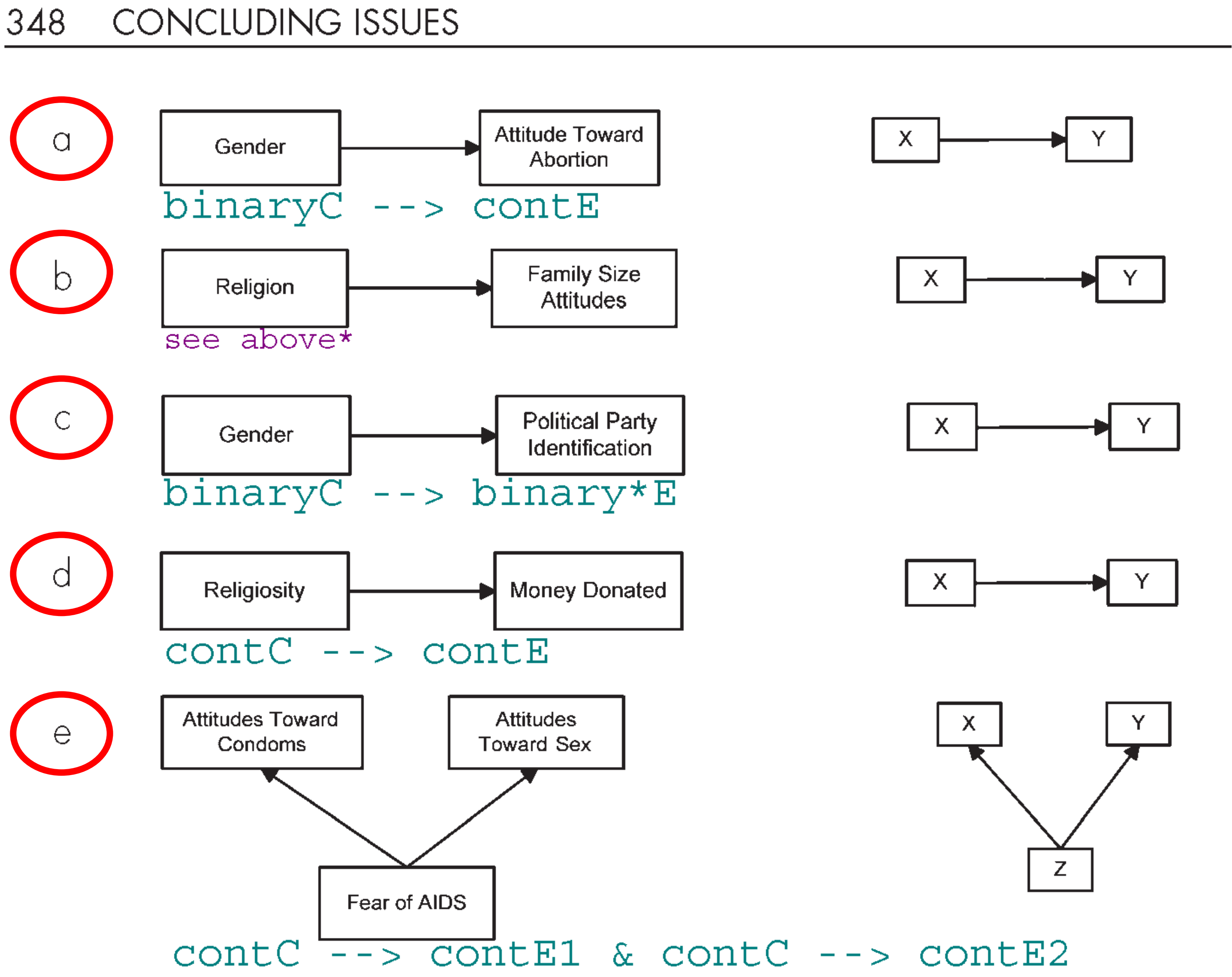
$$\beta = \frac{\sigma_{XY}}{\sigma^2{}_{XX}}$$

"The correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients along all of the paths by which the variables are connected." [Wright:115]

Wright, S. (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics, 6(2), 111.*

Inferring Theoretical Relationships from the Choice of Statistical Tests [1]

FIGURE 12.1. Causal models underlying statistical tests (text example on left, generic form on right). (a) Two Group/ Condition t-Test; (b) One-Way Analysis of Variance; (c) Chi-Square Test of Independence and Test of Proportions; (d) Pearson Correlation/ Linear Regression: Direct Cause Model; (e) Pearson Correlation: Common Cause or Spurious Effect Model

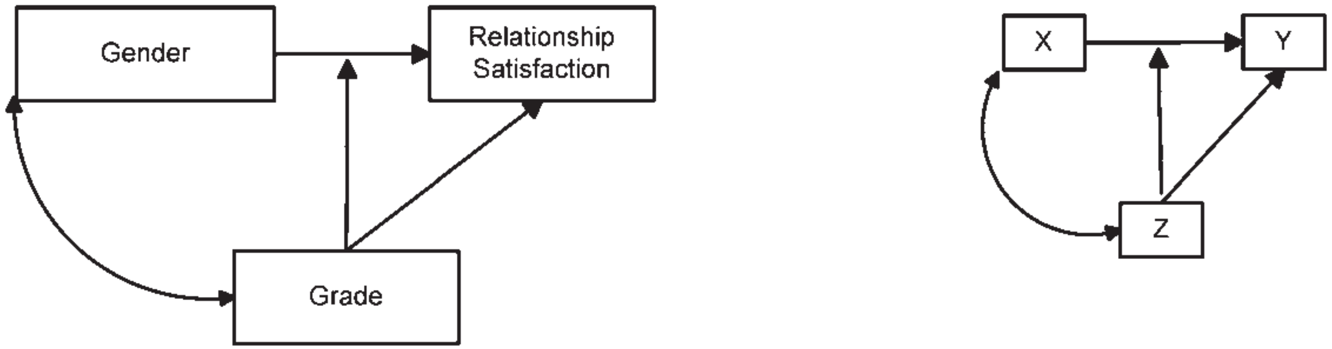


Jaccard, J., & Jacoby, J. (2009). Theory construction and model-building skills: A practical guide for social scientists https://www.guilford.com/books/Theory-Construction-and-Model-Building-Skills/Jaccard-Jacoby/9781462542437 : Guilford Press.

Inferring Theoretical Relationships from the Choice of Statistical Tests [2]

FIGURE 12.1. (cont.) (f) Two-Factor Analysis of Variance; (g) One-Way Analysis of Covariance: Mediation; (h) One-Way Analysis of Covariance: Independent Influence and Error Reduction; (i) Partial Correlation: Mediation.

f

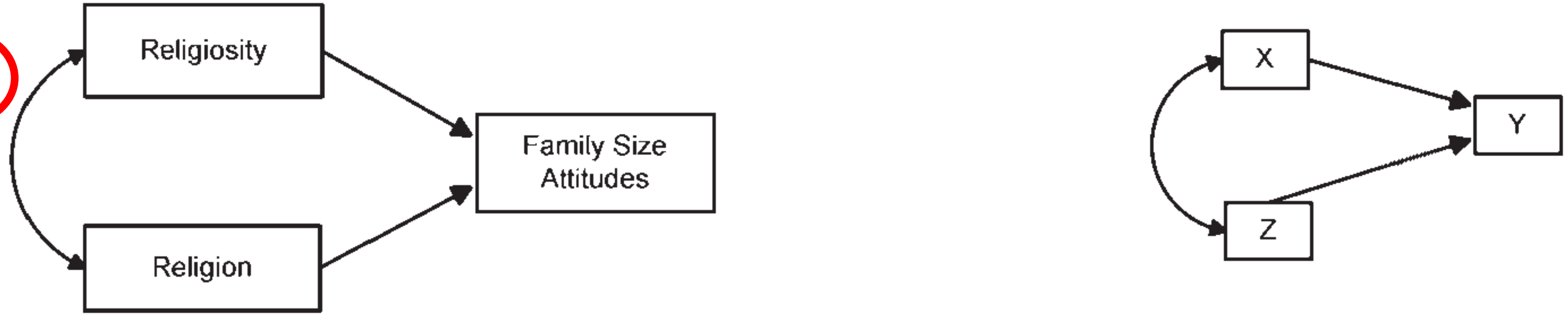



```
contC1 +contC2 + C1xC2--> contE & ContC1<->ContC2
```

g


```
contC(X)->contE(M) & contC(M)-> contE(Y)
```

h




```
contC1 -> contE & contC2 -> contE
```

i

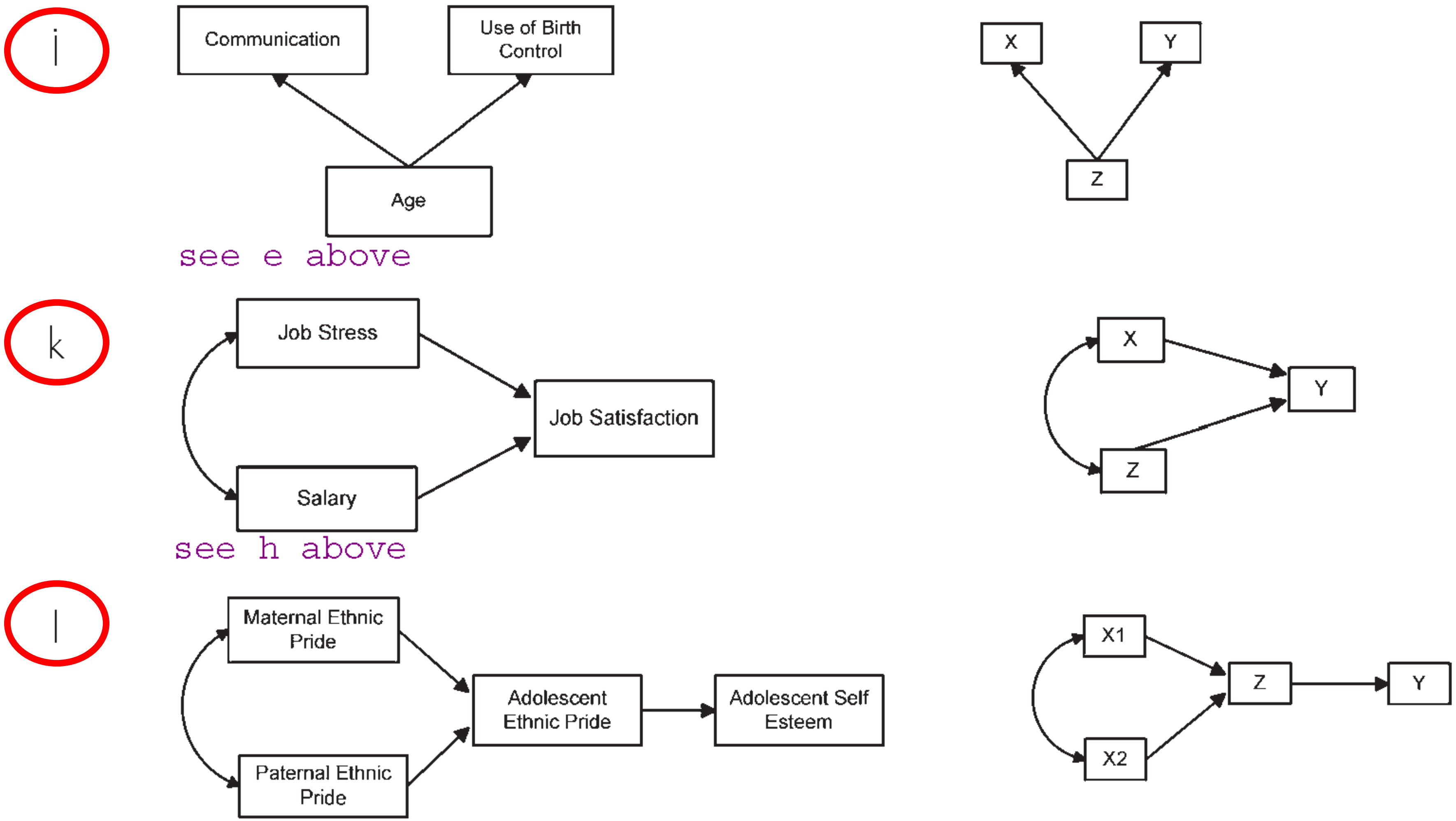
see g above

cont.

Jaccard, J., & Jacoby, J. (2009). Theory construction and model-building skills: A practical guide for social scientists https://www.guilford.com/books/Theory-Construction-and-Model-Building-Skills/Jaccard-Jacoby/9781462542437 : Guilford Press.

Inferring Theoretical Relationships from the Choice of Statistical Tests [3]

FIGURE 12.1. (cont.)
(j) Partial Correlation: Common Cause or Spurious Effect Model; (k) Multiple Regression; (l) Hierarchical Multiple Regression—Mediation.



see e above

see h above

contC1 +contC2-> contE(M) & ContC(M)<->ContC2E(Y)

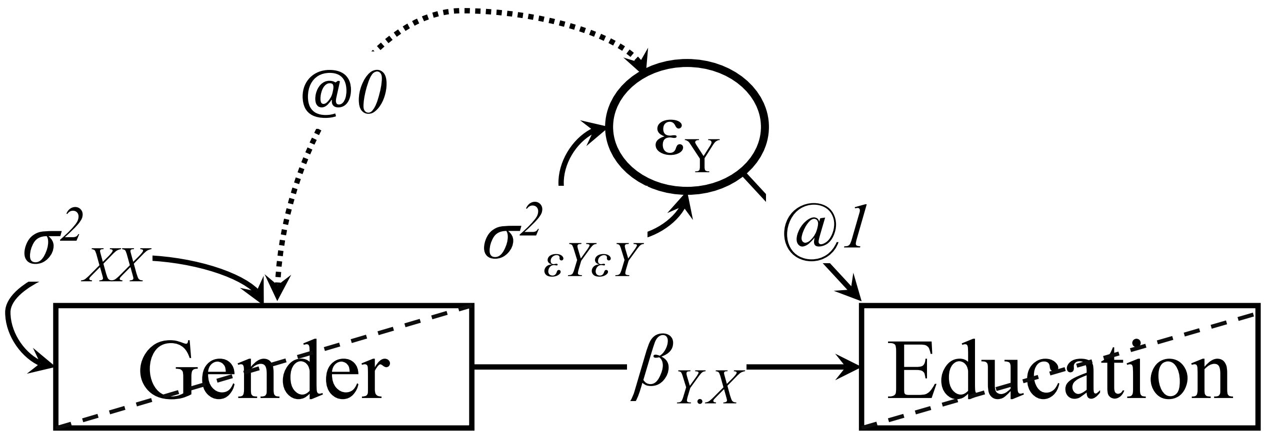
Jaccard, J., & Jacoby, J. (2009). Theory construction and model-building skills: A practical guide for social scientists https://www.guilford.com/books/Theory-Construction-and-Model-Building-Skills/Jaccard-Jacoby/9781462542437 : Guilford Press.

Descriptives of the analyzed variables

Variable	Counts	%s
Females[A]	378	73.8%
College education[B]	338	66.7%
Among males	87	65.9%
Among females	250	67.4%
	Means	SDs
Females	0.738	0.440
College education	0.667	0.470
Religiosity	27.30	8.13
Anxiety	1.25	0.45
Age	53.90	12.00
Chronic diseases	0.76	0.89
‘Church-to see people’[C]	2.21	1.27
Health rating	7.52	1.63

Notes: Valid *N ranges between 495-536;* [A]: vs. males; [B]: vs. less than college; [C]: ‘I go to church mainly because I enjoy seeing people I know there’ religiosity question; the Gender-Education covariance is .003, and the correlation.014; the Religiosity-Anxiety covariance is -.337, and the correlation -.091.

Two categorical variables causal model: Gender → Education



$$\beta = \frac{\sigma_{XY}}{\sigma^2_{XX}} = -.337 / (8.13*8.13) = -0.00510$$

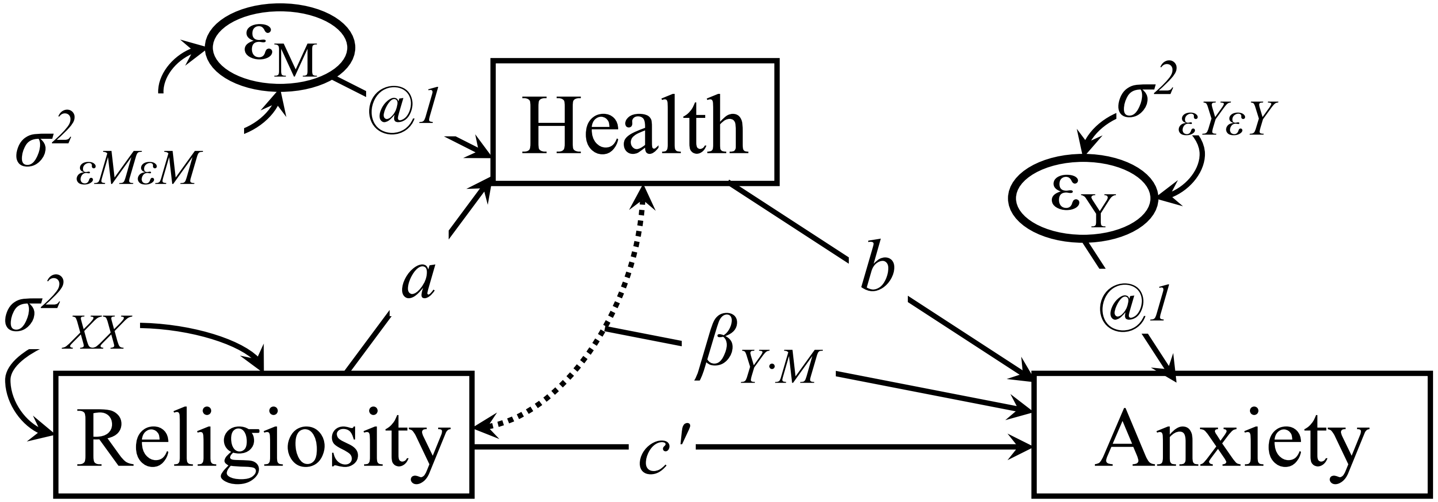
$$\rho = \beta \frac{\sigma_{XX}}{\sigma_{YY}} = -0.00510 *8.130/0.453 = -0.09150$$

Notes: X = Main cause, Y = Outcome; is the residual error; σ^2's are variances; the $\beta_{Y.X}$ parameter represent the X (Gender) -> Y (Education) effect; the interrupted line depicts the possibility of a correlation between predictor and residual error (forced to 0: @0); the model tests the hypothesis: $\overline{\text{College}}_{\text{Males}} = \overline{\text{College}}_{\text{Females}}$; this is the one group Female -> College model setup; a two-group path model is possible, which allows for inclusion of group specific variances (and covariates too), which can allow for additionally testing whether $\sigma^2_{\text{Education.Females}} = \sigma^2_{\text{Education.Males}}$; the binary variables are shown with an inside interrupted line.

Descriptives of analyzed variables

Model & Statistical test	Coefficient	Test Statistic	p value
Correlational Model [A]:	Gender ↔ Education		
Chi-square [B]	---	0.096	.757
t-test [C]	---	-0.309	.757
Log-linear model [D]	0.066	0.310	.757
Correlation between observables	0.014 $^{\rho}$	---	.757
Correlation –corrected for attenuation [E]	0.015	---	---
True Correlation between latents [E]	0.015	0.310	.757
Cause-Effect Model:	Gender → Education		
Tracing rule	0.014 [F]	---	---
Regression/path analysis - observables	0.014	0.310	.757
Logistic regression – observables	0.066 [G]	0.310	.757
Path analysis - latent Gender [E]	0.016	0.310	.757
IV estimation – health as instrument	0.431	1.380	.168

A three continuous variables causal model



Notes: The a-b-c' notation follows the classic Barron-Kenny labels; σ^2's are variances; the parameter represent the interaction (moderation) term effect;

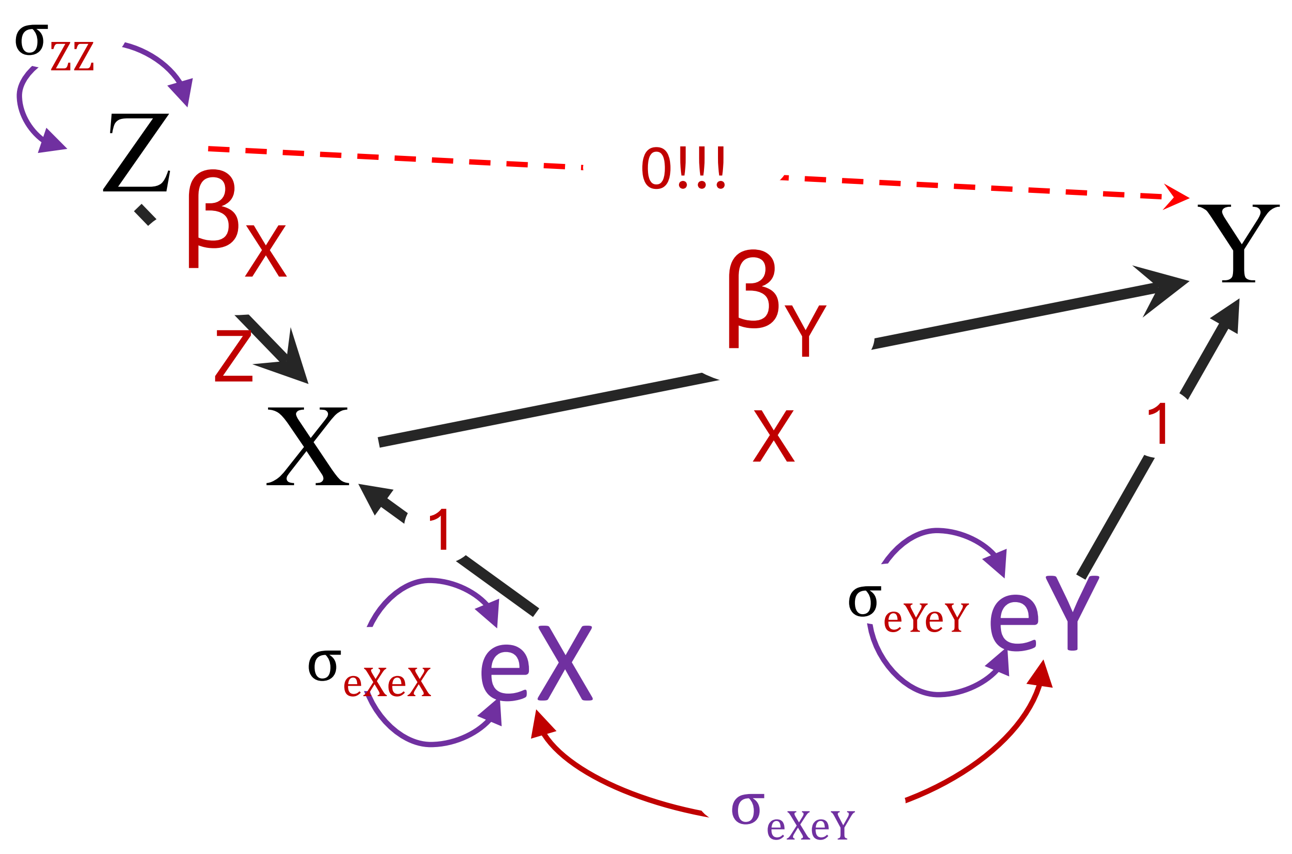
IV - total IV->X->Y mediation

If you think such instances are impossible, they are not:

A priest reads a prayer before the blessing ceremony. Parents are standing next to their vehicle, and the child safety seats are in place.

- Blessing of car seats (in Latino communities): great example of indirect effect of religious blessing on say car accident deaths: no direct effect however: divinity acts only through human agency.
- Effects of prayer on one’s own behavior.

'Tracing rule' in IV' = total IV->X->Y mediation



$\text{cov}(X,Y) = \sigma_{XY} = \beta_{Y.X} + \sigma_{eXeY}$ (

[trek Y -> X + trek Y-> e_Y -> e_X -> X]

[this is not needed in fact]

$\text{cov}(Z,X) = \sigma_{ZX} = \beta_{X.Z}\,\sigma_{ZZ}$ (2)

[trek X -> 'Z'-> σ_{ZZ} -> Z]

$\text{cov}(Z,Y) = \sigma_{ZY} = \beta_{Y.X}\,\beta_{X.Z}\,\sigma_{ZZ}$ (3)

[trek Y -> X -> 'Z'-> σ_{ZZ} -> Z, 2nd trek-0

Bring in from 2nd $\beta_{X.Z}\,\sigma_{ZZ}$

So $\sigma_{ZY} = \beta_{Y.X}\,\sigma_{ZX}$

Therefore $\beta_{Y.X} = \sigma_{ZY} / \sigma_{ZX}$

Simpler?

"Sewall Wright (1925 [3]) used instrumental variables to estimate the coefficients of a multiple equation model of corn and hog cycles." [1]

[1] Stock, J. H., & Trebbi, F. (2003). Retrospectives: Who invented instrumental variable regression? The Journal of Economic Perspectives, 17(3), 177-194.

[2]. Wright, S. (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics, 6(2), 111.*

[3]/ Wright, S., & Mcphee, H. C. (1925). An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees https://naldc.nal.usda.gov/download/IND43966972/PDF. Journal of agricultural research, 31(4), 377-383.

Two variable models: effects between Religiosity (**Rel.**) and Anxiety (**Anx.**)

Models and effect estimates	Unst.	SE	p	Stand.
Regression/path direct effects Model:	Rel. → Anx.			
Tracing rule Religiosity → Anxiety	-0.005	---	. ---	-0.092
Religiosity → Anxiety	-0.005	0.002	.038	-0.091
Latent Religiosity [A] → Anxiety	-0.006	0.003	.038	-0.102
'True' Religiosity [B] → Anxiety	-0.031	0.024	.188	-0.070
Bi-directional total effects [C] Model:	Re. → Anx. & Anx. → Rel.			
Religiosity → Anxiety [T]	0.023	0.010	.023	0.422
Anxiety → Religiosity [T]	-3.946	1.162	.001	-0.219
'True' Religiosity [B] → Latent Anxiety [,AT]	0.232	0.146	.112	4.187
Latent Anxiety [A] → 'True' Religiosity [B,T]	-0.284	0.125	.023	-0.016
'Instrumental variable' effects [D] Model:	IV → Rel. → Anx.			
Tracing rule Religiosity → Anxiety	0.036	---	---	---
Religiosity → Anxiety	0.036	0.024	.134	0.403
True' Religiosity [B] → Anxiety [A]	0.577	0.304	.057	0.544

Three variable models: effects from Religiosity (**Rel.**) to Anxiety (**Anx.**), modified by and through self-rated Health (**Hlth.**, Mediator, or Moderator)

Models and effect estimates	Unst.	SE	p	Stand.
Co-predictors Model:	(Rel. + Hlth.) → Anx.			
Rel. → Anx.	-0.007	0.004	0.086	-0.079
Hlth. → Anx.	0.000	0.001	0.902	0.006
Moderation/interaction Model:	(Rel. + Hlth.+ Rel.*Hlth.) → Anx.			
Rel. → Anx.	-0.035	0.017	0.045	-0.420
Hlth. → Anx.	-0.011	0.007	0.106	-0.391
Rel.*Hlth. → Anx.	0.002	0.013	0.902	0.006
Mediation Model:	Rel. → Hlth. & (Rel. + Hlth.) → Anx.			
DE c': Rel. → Anx.	-0.007	0.004	0.086	-0.079
IE a*b: Rel. → Anx.	0.000	0.000	0.903	0.000
TE c: Rel. → Anx.	-0.007	0.004	0.086	-0.079
a: Rel. → Hlth.	0.131	0.140	0.352	0.043
b: Hlth. → Anx.	0.000	0.001	0.902	0.006
Mediation & moderation[A] Model:	(Rel. + Hlth. + Rel.*Hlth.) → Anx. & Rel. → Hlth.			
Rel.* Hlth.→ Anx.	0.000	0.000	0.095	0.472
tDE c': Rel. → Anx.	-0.035	0.017	0.045	-0.367
pIE a*b: Rel. → Anx.	-0.001	0.002	0.420	-0.015
TE c: Rel. → Anx.	-0.036	0.018	0.047	-0.382
a: Rel. → Hlth.	0.013	0.014	0.352	0.043
b: Hlth.→ Anx.	-0.107	0.066	0.106	-0.342

Walk through the applied examples

1. Start with the model, then estimate the parameters:
 i. Using the tracing rule
 ii. Using free software Onyx and Jamovi
 - R\lavaan logic in Jamovi makes clear these 2 are distinct operations:

I. Model specification, e.g. “X → Y”
II. Estimation: ‘fit’ the model unto the data

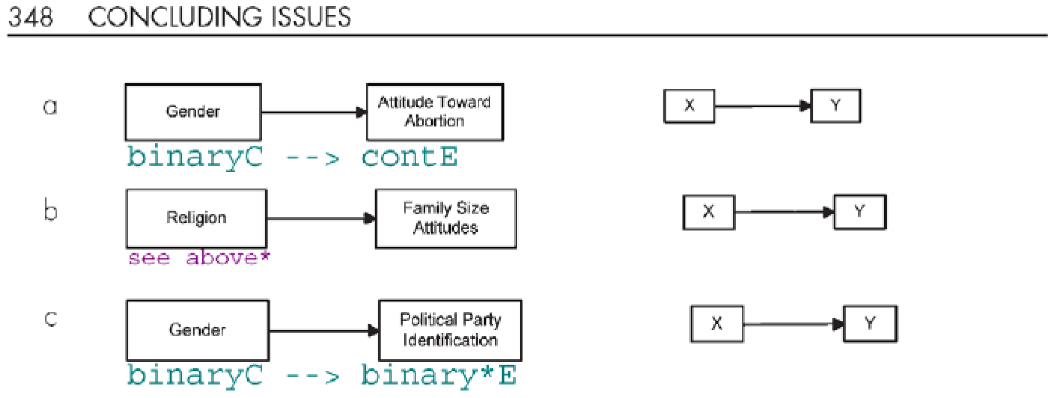
http://tinyurl.com/pathstats

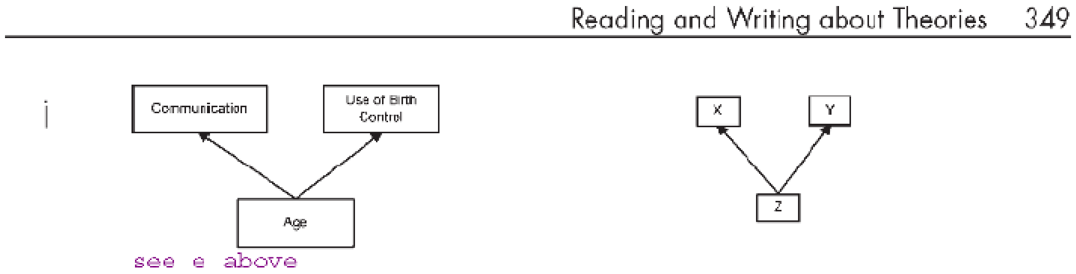
Walk through the applied examples

http://tinyurl.com/pathstats

Content:

A1. Annotated appendix of models vs statistical tests, Jaccard & Jacobi





Conclusions

1. Introducing learners to statistics can be done using modelling logic: graphic view helps

2. The ‘tracing rule’ set of simple rules allows one to estimate model parameters visually in a graph.

3. Simple software can make the mechanics more visible and intuitive (Excel-based intros are even better).