‘Translating back’ common statistical test into their graphical causal language ancestors: adding practical statistical codes to the Jaccard & Jacoby ‘tests-causal’ model crosswalk

Emil Coman, PSTAT PhD, Health Disparities Institute, UConn Health

with *direct* assistance from
James Jaccard, Silver School of Social Work
New York University

UCONN HEALTH
HEALTH DISPARITIES INSTITUTE

NYU SILVER SCHOOL OF SOCIAL WORK

Statistics Talk, Friday, Feb. 1

InCHIP (Ryan Refectory)

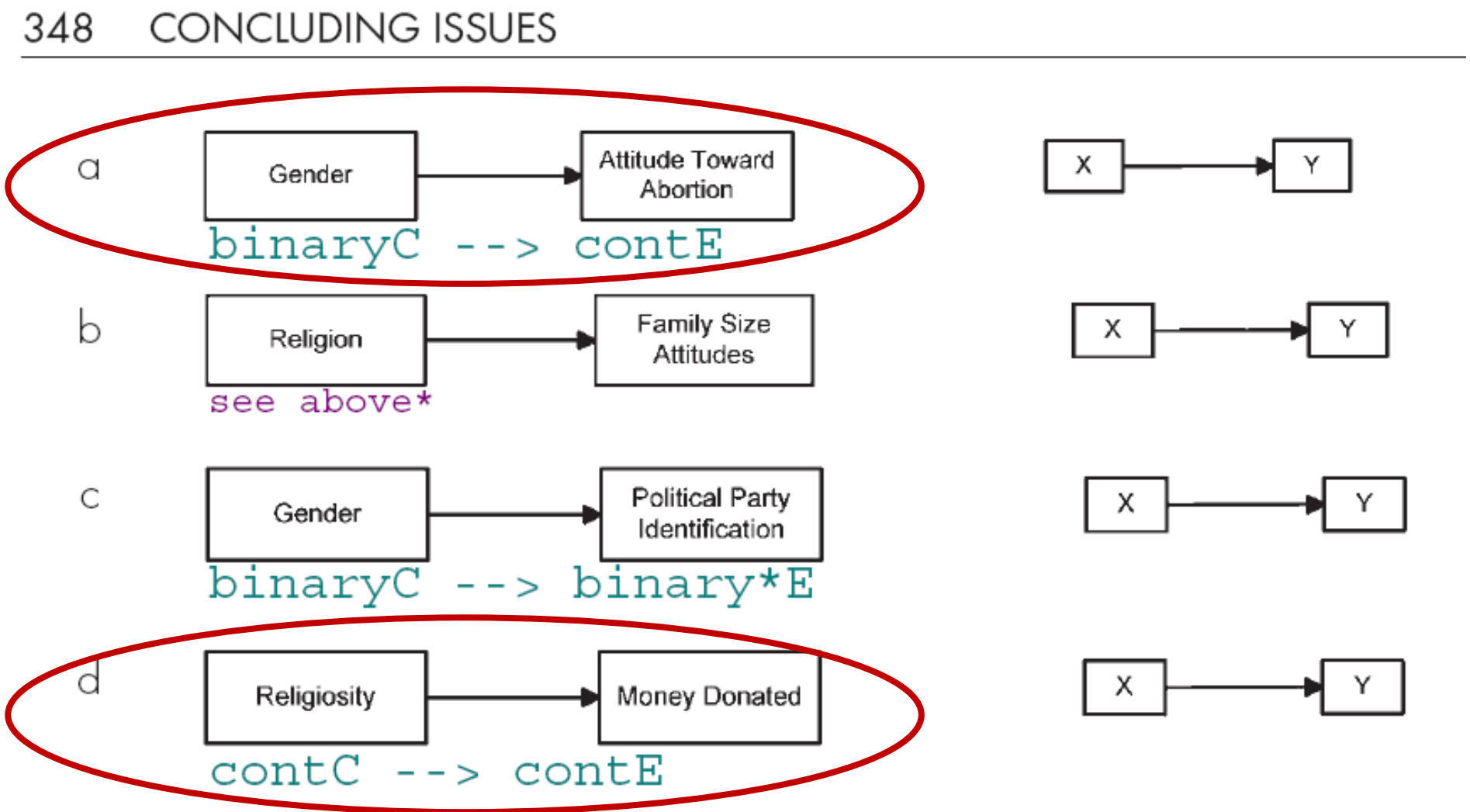
Promisory note:
I will add open source R code in *daggity* and *lavaan* and *Onyx* to give practical (and testable) meaning to theoretical causal models behind common statistical analyses (as shown in Jaccard & Jacoby, 2009: 1.i. Appendix**)**: **t-test**; [ANOVA; Pearson chi-square;] **regression**; **mediation;** [causal mediation, and more. We will tackle also how and where to 'add time' in causal models].

Actually: just 1 variable and 2 variable models today, sorry.

Refer for more to: Jaccard, J., & Jacoby, J. (2009). Theory construction and model-building skills: A practical guide for social scientists: Guilford Press.
1.i. APPENDIX: Inferring Theoretical Relationships from the Choice of Statistical Tests

This is an applied worksession (partly), so:
For R codes and data for illustrations go to:

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;



Main points

1. Statistical tests are *clueless* without a causal model;
   1. Research questions need: info about research design and on a working model of how data is generated by nature (with researcher's or policy maker's input sometimes).
   2. In Null Hypothesis Significance Testing (NHST) framework: one needs a model for the 'null' state.
2. Instead of mere NHST from statistical tests, the focus shifts to causally well-informed models and tests of differences in fit between alternative/nested models.
3. The easiest way to model is graphically: will show how with *lavaan* and *Ωnyx* parametrically (and with data), and with *dagitty* and *MIIVsem* nonparametrically (data-less).

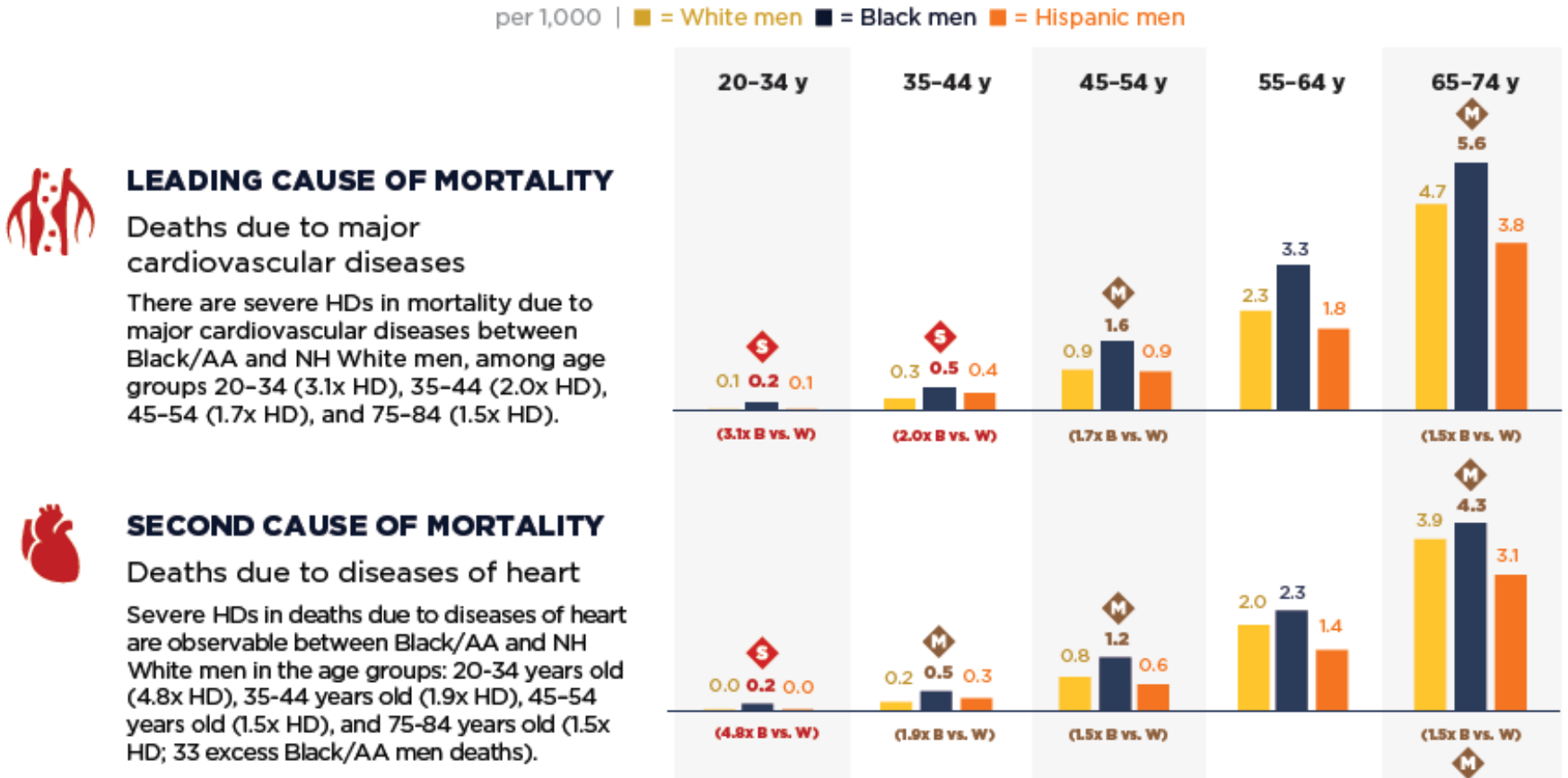
DAGitty - drawing and analyzing causal diagrams (DAGs)
Causal Inference In Statistics: A Companion for R Users – DAGitty

Who we are:

-HDI: UConn Health Disparities Institute

-BMoC: CT Report Card on Health Equity Among Boys And Men Of Color (BMoC) in CT: full report ; interview on WNPR with Dr. Wizdom Powell

-Biostatistics club – UConn Health formerly CICATS, now Connecticut Convergence Institute for Translation in Regenerative Engineering



Who: Modern Modeling Methods ‘perpetual student’; Storrs has seen the brightest statisticians; e.g.:
[1] p. 103 & 105:
But some authors have the ability to display complicated ideas with such force and simplicity that the development appears to be obvious in their exposition. Only upon reviewing what has been learned does the reader realize the great power of the results. Such an author was Jerzy Neyman. It is a pleasure to read his papers. The ideas evolve naturally, the notation is deceptively simple, and the conclusions appear to be so natural that you find it hard to see why no one produced these results long before.
Pfizer Central Research, where I worked for twenty-seven years, sponsors a yearly colloquium at the University of Connecticut. The statistics department of the university invites a major figure in biostatistical research to come for a day, meet with students, and then present a talk in the late afternoon. Since I was involved in setting up the grant for this series, I had the honor of meeting some of the great men of statistics through them. **Jerzy Neyman was one such invitee.**
[…]
My comments were directed to Neyman's presentation that day, as he had asked. In particular, I told how I had discovered the 1939 paper years before and revisited it in anticipation of this session. I described the paper, as best I could, showing enthusiasm when I came to the clever way Neyman had developed the meaning of the parameters of the distribution.”

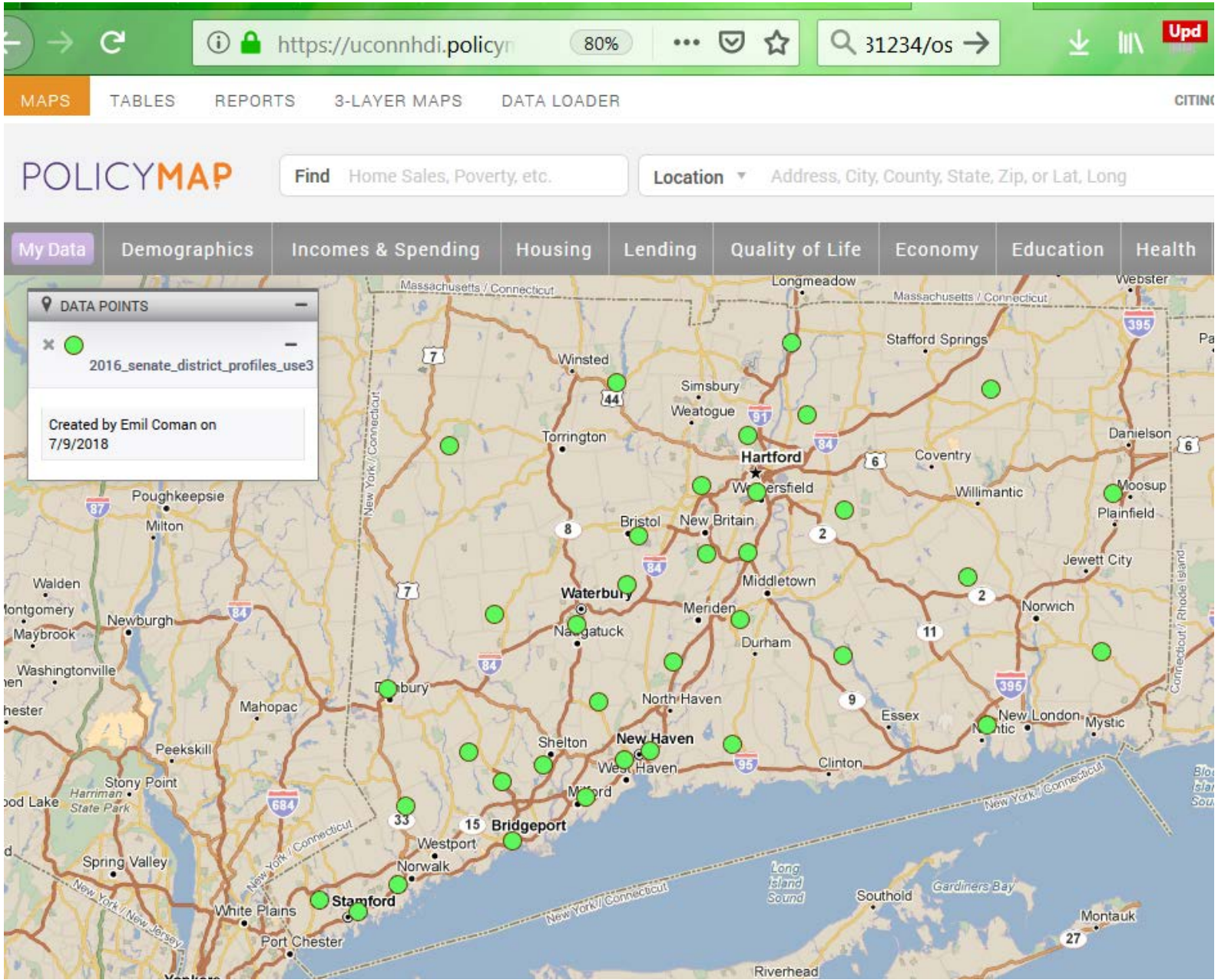
1. Salsburg, D. (2001). The lady tasting tea: How statistics revolutionized science in the twentieth century: Macmillan.
[he published one last year too: 2. Salsburg, D. (2017). Errors, Blunders, and Lies How to Tell the Difference.]

Neyman, J. (1939). On a new class of" contagious" distributions, applicable in entomology and bacteriology. *The Annals of Mathematical Statistics, 10(1), 35-57.*

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.1933.0009. *Phil. Trans. R. Soc. Lond. A, 231(694-706), 289-337.*

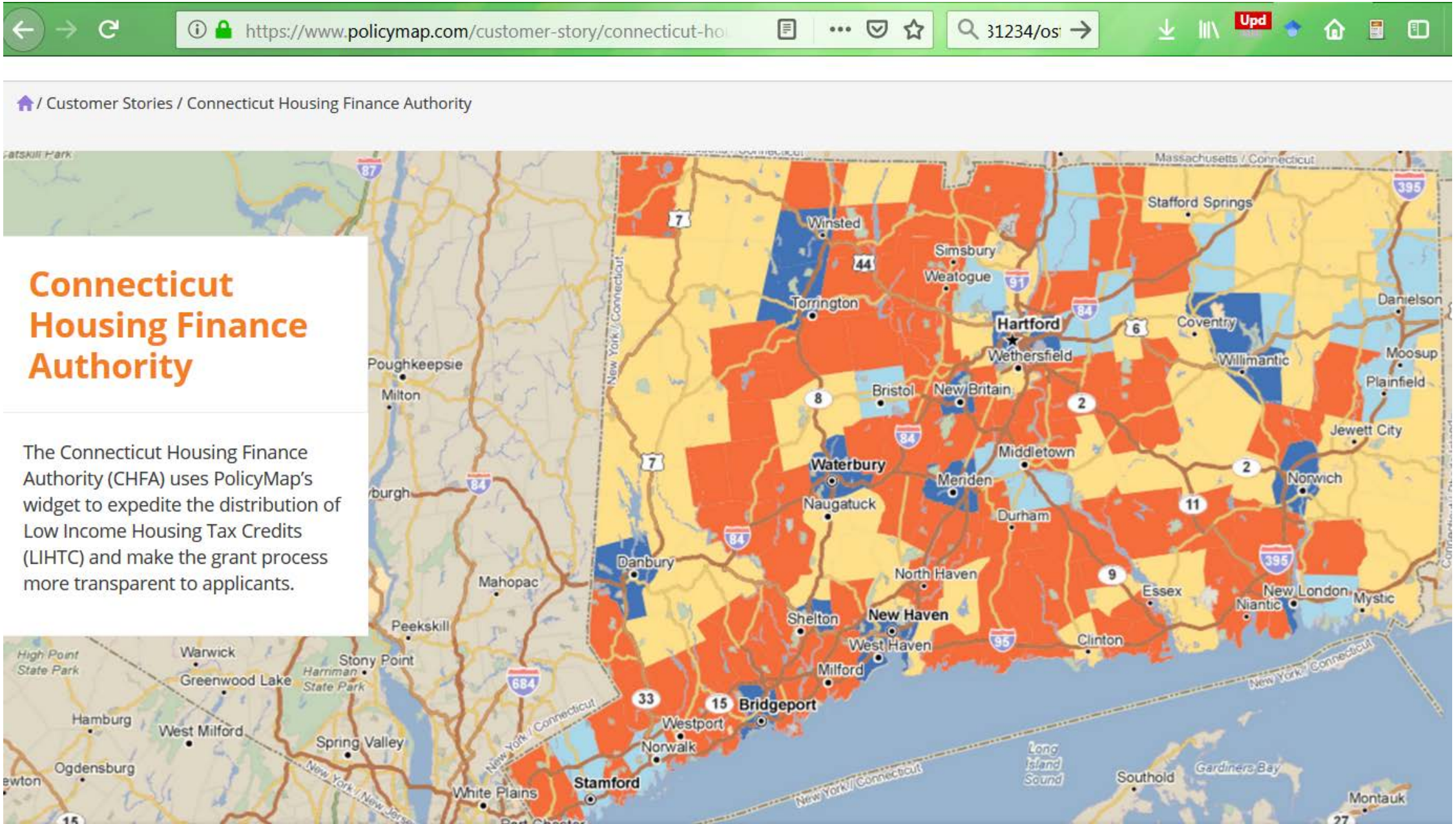
HDI: UConn Health Disparities Institute

PolicyMap https://uconnhdi.policymap.com



HDI: UConn Health Disparities Institute

PolicyMap https://uconnhdi.policymap.com



Example of what one can do with PolicyMap data

Spatial models reveal a significant relationship between **lower** *percent minority residents* and **higher** *percent of residents reporting very good health*, across CT state districts. Senate districts with more minorities have fewer residents reporting very good health. (BMoC report, HDI).

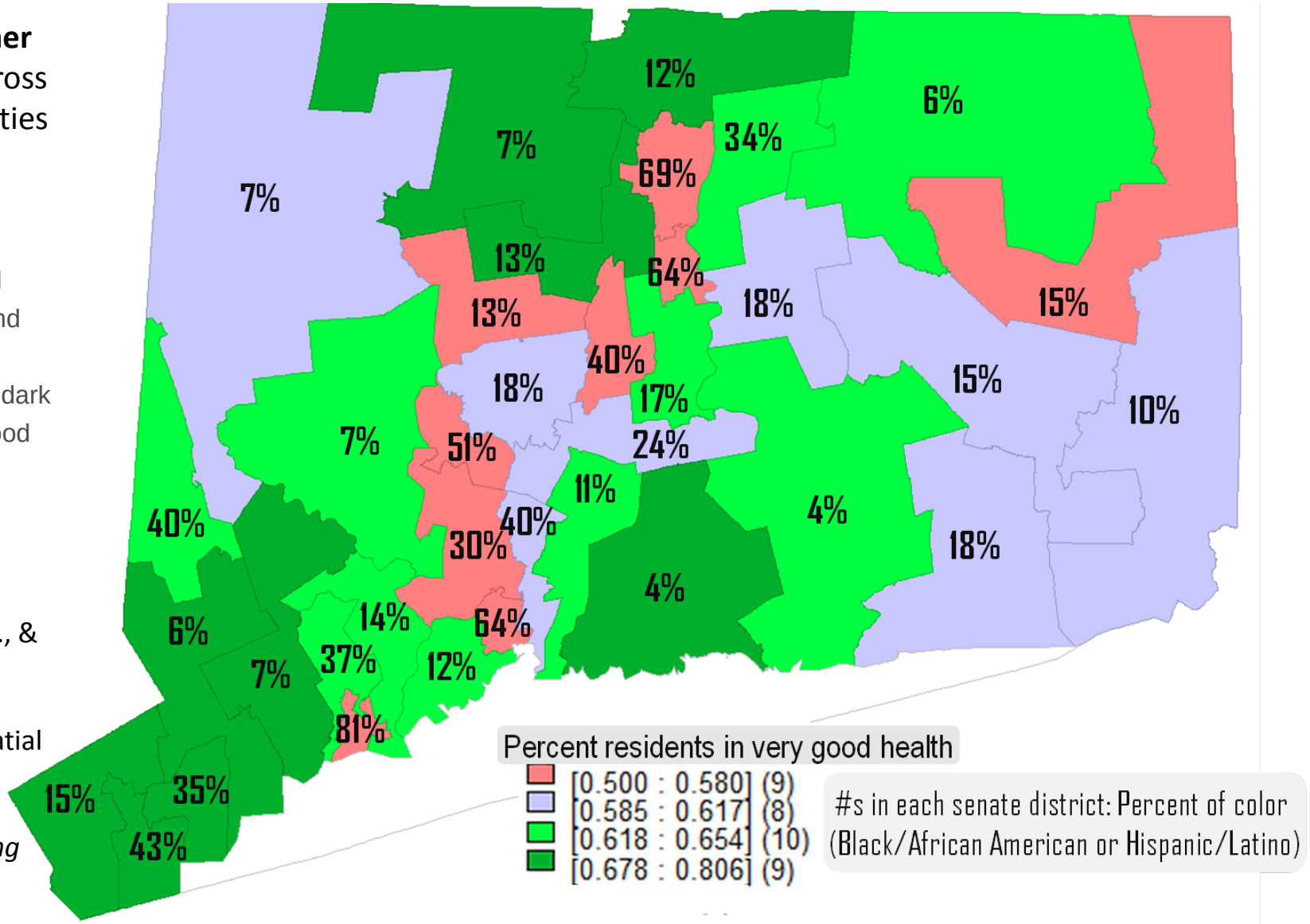
The 9 red districts have the least residents reporting very good health (<58%) , while the 8 blue districts have between 58% and 62% residents enjoying very good health. The 10 light green districts have 62%--65% residents with good health, and the 9 dark green districts are those with >68% residents reporting very good health. The data was combined from several sources by DataHaven [6], and is freely downloadable.

To analyze data further:

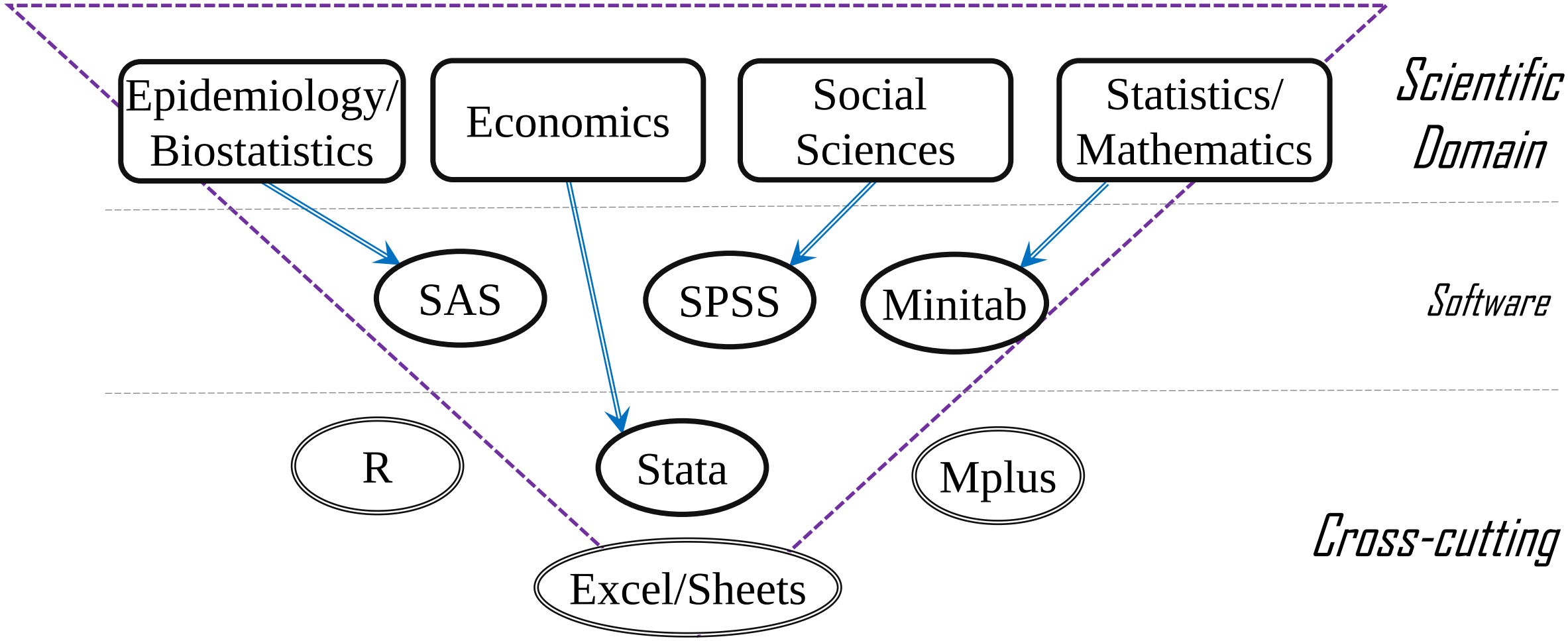
Lamb, E. G., Mengersen, K. L., Stewart, K. J., Attanayake, U., & Siciliano, S. D. (2014). Spatially explicit structural equation modeling. *Ecology, 95(9), 2434-2442.*

Liu, X., Wall, M. M., & Hodges, J. S. (2005). Generalized spatial structural equation models. *Biostatistics, 6(4), 539-557.*

Wall, M. M. (2012). Spatial Structural Equation Modeling. In R. Hoyle (Ed.), *Handbook of Structural Equation Modeling (pp. 674-689). New York: Guilford Press.*



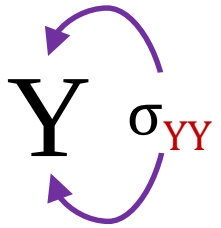
Statistical and software traditions - simplified



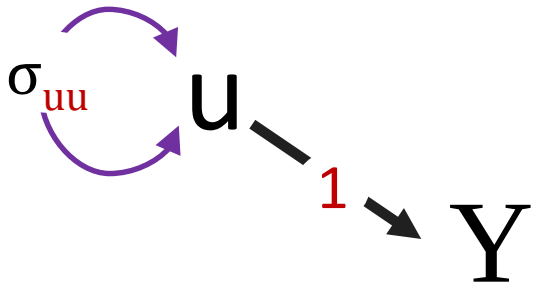
Some use combinations of code, or many at once.

‘Tracing rule’ powers at work with 1 variable

We add σ‘s, and run the ‘tracing’ tool along directed and bidirectional paths (except common effects along the way; and we can ‘go back in time) + grab the cov/ariance of the variable at the ‘root’ on the pathway.



$Cov(YY) = \sigma_{YY}$



An <u>exogenous</u> variable (to the model!) is a ‘big fat disturbance’ or structural error: all of its variance is error, or unexplained. So:

$Y = 1 \cdot u$

and therefore

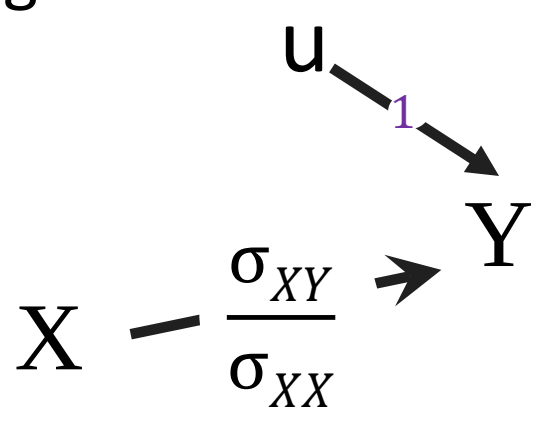
$\sigma_{YY} = \sigma_{uu}$

“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.” [:115]

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

Regression and how potential 'change' is deduced

Regression



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'.

$$Y_i = \beta_{YX} \cdot X_i + 1 \cdot u_i \qquad [\textit{easier if } \alpha_Y = 0]$$

Hence if one multiplies by X_i :

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

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

$$\frac{\sum_i^N X_i \cdot Y_i}{N} = \beta_{YX} \cdot \frac{\sum_i^N X_i \cdot X_i}{N} + \frac{\sum_i^N X_i \cdot u_i}{N}$$

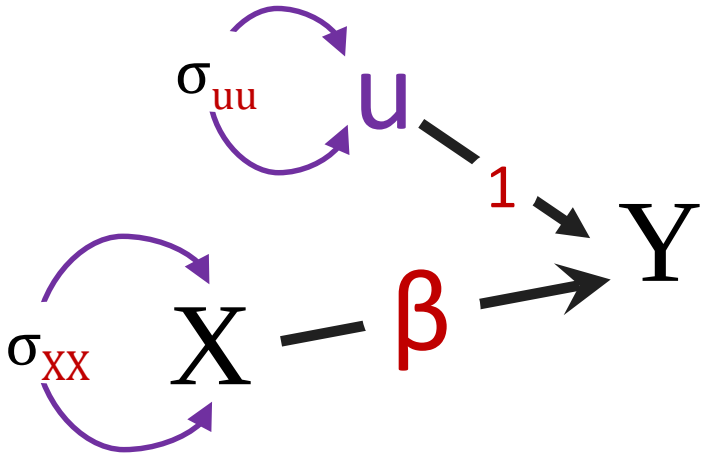
Hence :

$$\sigma_{YX} = \beta_{YX} \cdot \sigma_{XX} + \sigma_{Xu}$$

So:

$$\beta_{YX} = \frac{\sigma_{YX}}{\sigma_{XX}} - \sigma_{Xu} = \frac{Cov(Y,X)}{Cov(X,X)} - Cov(X,u)$$

‘Tracing rule’ powers at work



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

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

Hence:

$$\beta = \frac{\sigma_{XY}}{\sigma_{XX}}$$

Simpler?

“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.*

‘Tracing rule’

“Tracing rules (or Wright’s rules) are simply a way to estimate the covariance between two variables by summing the appropriate connecting paths.” [1:23]

“Trace all paths between two variables (or a variable back to itself), multiplying all the coefficients along a given [open] path.

i. You can start by going backwards along a single-headed arrow, but once you start going forward along these arrows you can no longer go backwards.

ii. No loops! That is, you cannot go through the same variable more than once for a given path.

iii. At *maximum*, there can be one double-headed arrow included in a path.

iv. After tracing all the paths for a given relationship, sum all the paths.” [1:24]

+v. EC: A ‘collider (z_1--> w <-- z_2) is: not open/closed/disconnected

Cannot go forward THEN backwards ‘in time’

“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.” [2:115]

1. Beaujean, A. A. (2014). Latent variable modeling using R: A step-by-step guide: Routledge.
2. Wright, S. (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics, 6(2), 111.*

Original translational insight

Not clear which came first:

1. The decomposition idea
2. The regression->path conceptual leap.

kept unchanged, to the total standard deviation. A path coefficient differs from a coefficient of correlation in having direction.

The symbol $p_{X \cdot A}$ means the coefficient for the path of influence from A to X. In most cases in the present paper, however, it will be more convenient to represent the path coefficients by single letters.

It can be shown that the squares of the path coefficients measure the degree of determination by each cause. If the causes are independent of each other, the sum of the squared path coefficients is unity. If the causes are correlated, terms representing joint determination must be recognized. The complete determination of X in figure 1 by factor A and the correlated factors B and C can be expressed by the equation:

$$a^2 + b^2 + c^2 + 2\,bc\,r_{BC} = 1 \tag{1}$$

The squared path coefficients and the expressions for joint determination measure the portion of the squared standard deviation of the effect due to the causes singly and jointly, respectively.

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. In figure 1, X and Y are connected by four paths.

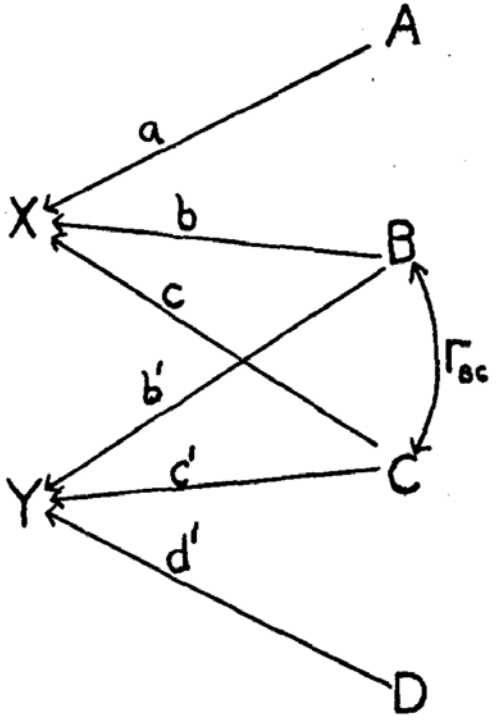


FIGURE 1.—A diagram illustrating the case of two variables (X and Y) determined in part by causes in common (B and C) which are correlated with each other.

$$r_{XY} = bb' + cc' + br_{BC}c' + cr_{BC}b' \tag{2}$$

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

First SEM model: 1920

"**The path coefficient,** measuring the importance of a given path of influence from cause to effect, is defined as the **ratio of the variability of the effect to be found when all causes are constant except the one in question, the variability of which is kept unchanged,** to the total variability.
Variability is measured by the standard deviation." [1]:329

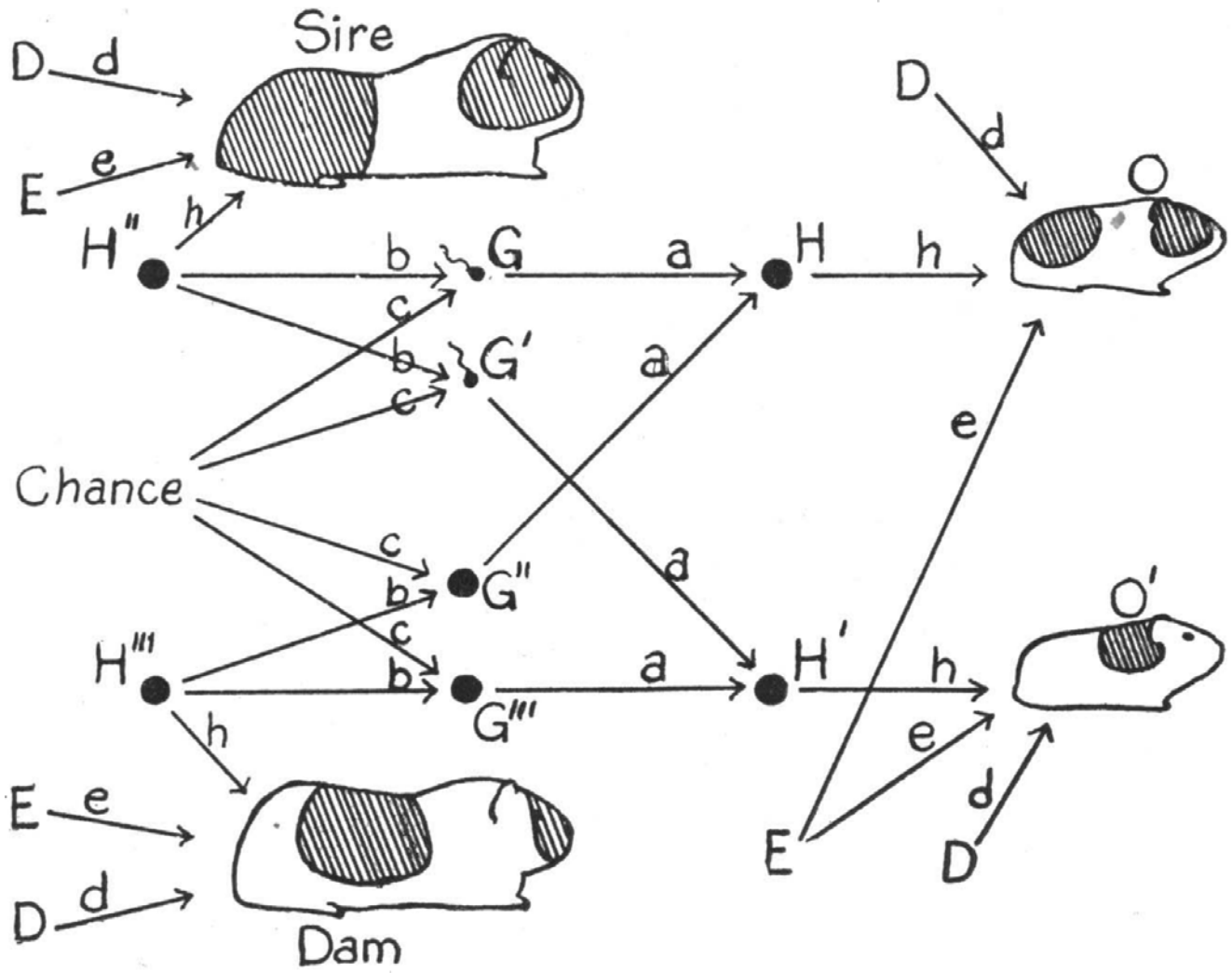


FIG. 5.

Diagram illustrating the casual relations between litter mates (O, O') and between each of them and their parents. H, H', H'', H,''' represent the genetic constitutions of the four individuals, G, G', G'', and G''' that of four germ cells. E represents such environmental factors as are common to litter mates. D represents other factors, largely ontogenetic irregularity. The small letters stand for the various path coefficients.

1. Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences, 6(6), 320-332.*

Original insight

The ‘multiply- along-the-trek’ rule may have come from the derivatives of composite functions formulas:

[3]Pearl (2013) calls them nested counterfactuals; the key insight Sewall Wright foresaw when proposing the path analytic method may have been that the change in Y in relation to the change in X (the slope $\delta Y/\delta X$), traced on the path through an intermediary M, is linked to the slopes $\delta M/\delta X$ and $\delta Y/\delta M$ following the composite function chain rule of derivatives: $\delta y/\delta x = \delta y/\delta m \cdot \delta m/\delta x$, which mirrors the Baron and Kenny $i = a \cdot b$. Adding the contributions of all such X-to-Y open paths yields the model predicted association between X and Y (see the “tracing rule,” Loehlin, 2004).

Frontiers in Psychology | www.frontiersin.org February 2017 | Volume 8 | Article 151

Coman, E. N., Thoemmes, F., & Fifield, J. (2017). Commentary: Causal Effects in Mediation Modeling: An Introduction with Applications to Latent Variables. *Frontiers in Psychology, 8(151). doi:10.3389/fpsyg.2017.00151*

Origins of the Bayesian Networks

“Networks employing Directed Acyclic Graphs (DAGs) have a long and rich tradition, starting with the geneticist Wright (1921). He developed a method called path analysis [Wright, 1934] which later on, became an established representation of causal models in economics [Wold, 1964], sociology [Blalock, 1971] and psychology [Duncan, 1975]. Influence diagrams represent another application of DAG representation [Howard and Matheson, 1981], [Shachter, 1988] and [Smith, 1987]. These were developed for decision analysis and contain both chance nodes and decision nodes (our definition of causal models excludes decision nodes). Recursive models is the name given to such networks by statisticians seeking meaningful and effective decompositions of contingency tables (Lauritzen, 1982), (Wermuth & Lauritzen, 1983], [Kiiveri et al, 1984]. Bayesian Belief Networks (or Causal Networks) is the name adopted for describing networks that perform evidential reasoning ((Pearl, 1986a, 1988]). This paper establishes a clear semantics for these networks that might explain their wide usage as models for forecasting, decision analysis and evidential reasoning.” *[1]*:136

1. Geiger, D. and J. Pearl, *On the logic of causal models.* arXiv preprint arXiv:1304.2355, 1988.

Causal chains: early origins

"The φ's may be thought of as quasiprobabilities of the events. They are defined successively as follows [2: 48]:

$\Phi_0 = I.$

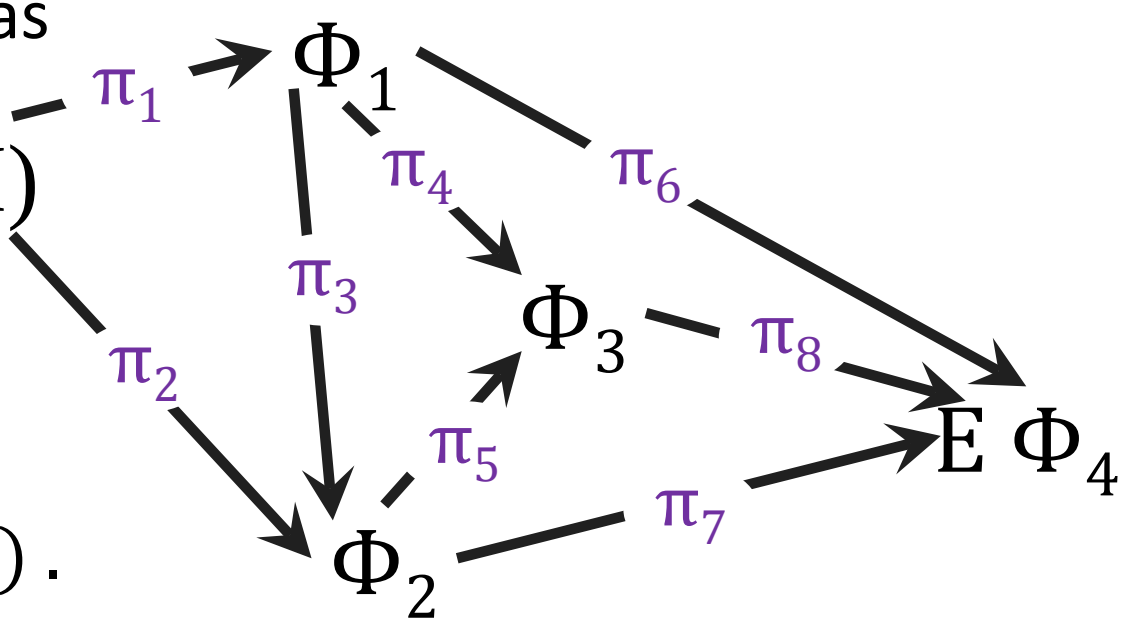
$\Phi_1 = \pi_1.$

$\Phi_2 = I - (I - \pi_2) \cdot (I - \Phi_1 \cdot \pi_3)\ .$

$\Phi_3 = I - (I - \Phi_1 \cdot \pi_4) \cdot (I - \Phi_2 \cdot \pi_5)\ .$

$\Phi_4 = I - (I - \Phi_1 \cdot \pi_6) \cdot (I - \Phi_2 \cdot \pi_7) \cdot (I - \Phi_3 \cdot \pi_8)\ .$

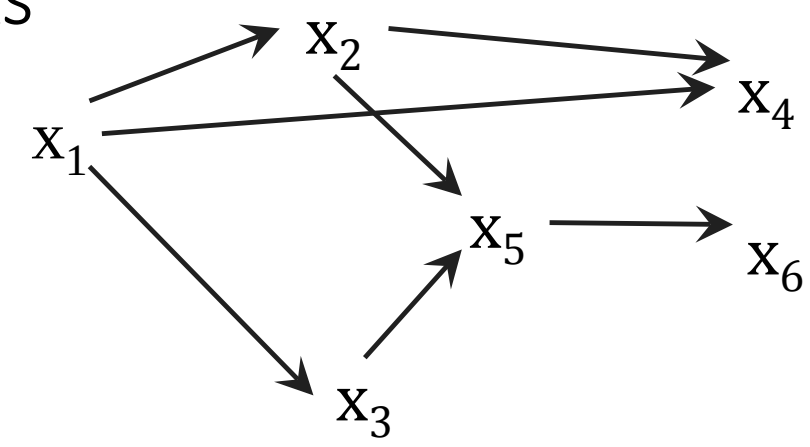
$F(\Phi_0 = I)$



$S(\pi) = Q(\Phi_4\ , o) = -\log\ (I - \Phi_4)\ .$

1. Good, I. J. (1961). A causal calculus (I). *The British Journal for the Philosophy of Science, 11(44), 305-318.*
2. Good, I. J. (1961). A causal calculus (II). *The British Journal for the Philosophy of Science, 12(45), 43-51.*

Origins of the Bayesian Networks



""[1] Probabilistic formulae of this kind are shorthand notation for the statement that for any instantiation i of the variables $x_1, x_2 \ldots\ldots x_n$, the probability of the joint event $(x_1 = i\ )$ & $(x_2 = i_2)$ &... & $(x_n = i_n)$ is equal to the product of the probabilities of the corresponding conditional events $(x_1 = i_1)$, $(x_2 = i_2$ if $x_1 = i_1)$, $(x_3 = i_3$ if $(x_2 = i_2$ & $x_1 = i_1)\ )\ \ldots\ldots$ For this expansion to be valid, we must require that $P(E) > 0$ for all conditioning events E." So, for example, the distribution corresponding to the graph of Fig. 1 can be written by inspection:

$P(x_1, x_2, x_3, x_4, x_5, x_6) =$
$= P(x_6|x_5) \cdot P(x_6|x_5) \cdot P(x_5|x_2, x_3) \cdot$
$P(x_4|x_1, x_2)\ P(x_3|x_1) \cdot P(x_2|x_1) \cdot P(x_1)$ (Pearl 1986):244-5

FIG. 1. A typical Bayesian network representing the distribution
$P(x_1 \ldots x_6) = P(x_6|x_5) \cdot P(x_5|x_2, x_3) \cdot$
$P(x_4|x_1, x_2)\ P(x_3|x_1) \cdot P(x_2|x_1) \cdot P(x_1)$

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence, 29(3), 241-288. doi:http://dx.doi.org/10.1016/0004-3702(86)90072-X*

Modern Bayesian Networks

“In words, the qualitative statistical dependencies shown in this small Bayesian network can be described as follows:
1. A recent trip to Asia (A) increases the chances of tuberculosis (T).
2. Smoking (S) is a risk factor for both lung cancer (L) and bronchitis (B).
3. The presence of either € tuberculosis or lung cancer can be detected by an X-ray result, but the X-ray alone cannot distinguish between them.
4. Dyspnoea (D) (shortness of breath) may be caused by bronchitis (B), or either € tuberculosis or lung cancer.
Each node represents a variable that can be in a discrete number of possible states. We write x_i for the variable representing the different possible states of the node i. In addition to the qualitative dependencies described by the Bayesian network graph, there are quantitative statistical relationships that we assign to each arrow in the graph. Associated with each arrow is a conditional probability: for example, we write $p(x_L| x_S)$ for the conditional probability of a patient having lung cancer given that he does or does not smoke. For this link, we say that the S node is the ‘parent’ of the L node because x_L is conditionally dependent on x_S. Some nodes like the D node might have more than one parent; thus we write $p(x_D| x_E, x_B)$ for the conditional probability of having dyspnoea. “ (Yedidia, Freeman et al. 2003):6

Fig. 1 The fictional “Asia” Bayesian network, taken from (Lauritzen and Spiegelhalter 1988)

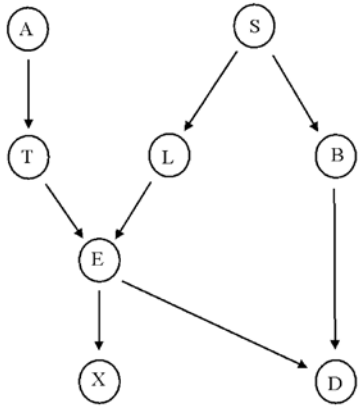
The over-all probability

$p([x]) = p (x_A, x_S, x_T, x_L, x_B, x_E, x_X, x_D)$

that the patient has some combination of symptoms, test results, and diseases is just the product of all probabilities of the parent nodes and all the conditional probabilities’:

$p([x]) = p(x_A) \cdot p(x_S) \cdot p(x_T|x_A) \cdot p(x_L|x_S) \cdot p(x_B|x_S) \cdot p(x_E|x_L, x_T) \cdot p(x_D|x_B , x_E) \cdot p(x_X|x_E)$

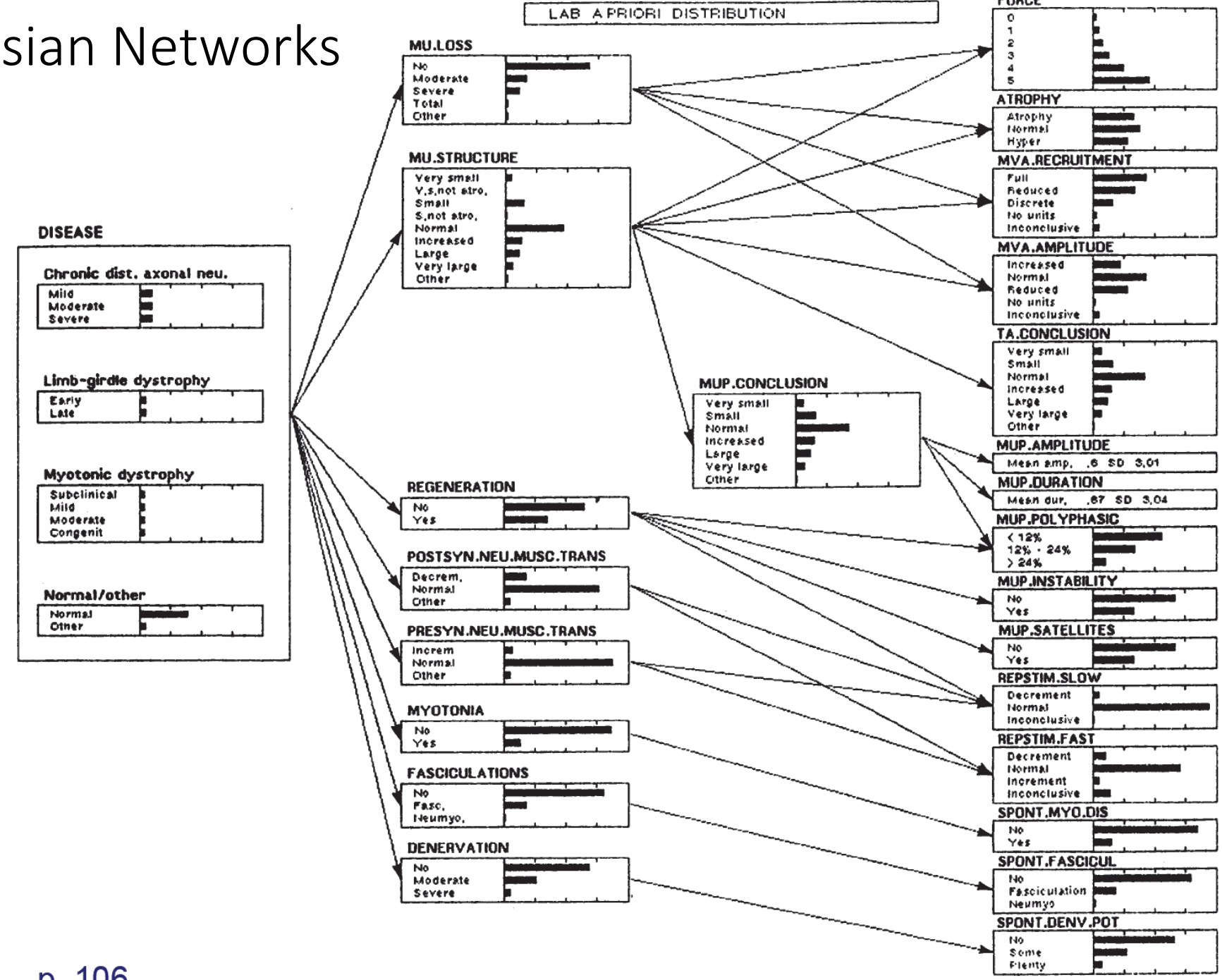
(1: p. 6)



1. Spiegelhalter, D. J., & Lauritzen, S. L. (1988). Statistical reasoning and learning in knowledge-bases represented as causal networks *Expert Systems and Decision Support in Medici*

Origins of the Bayesian Networks

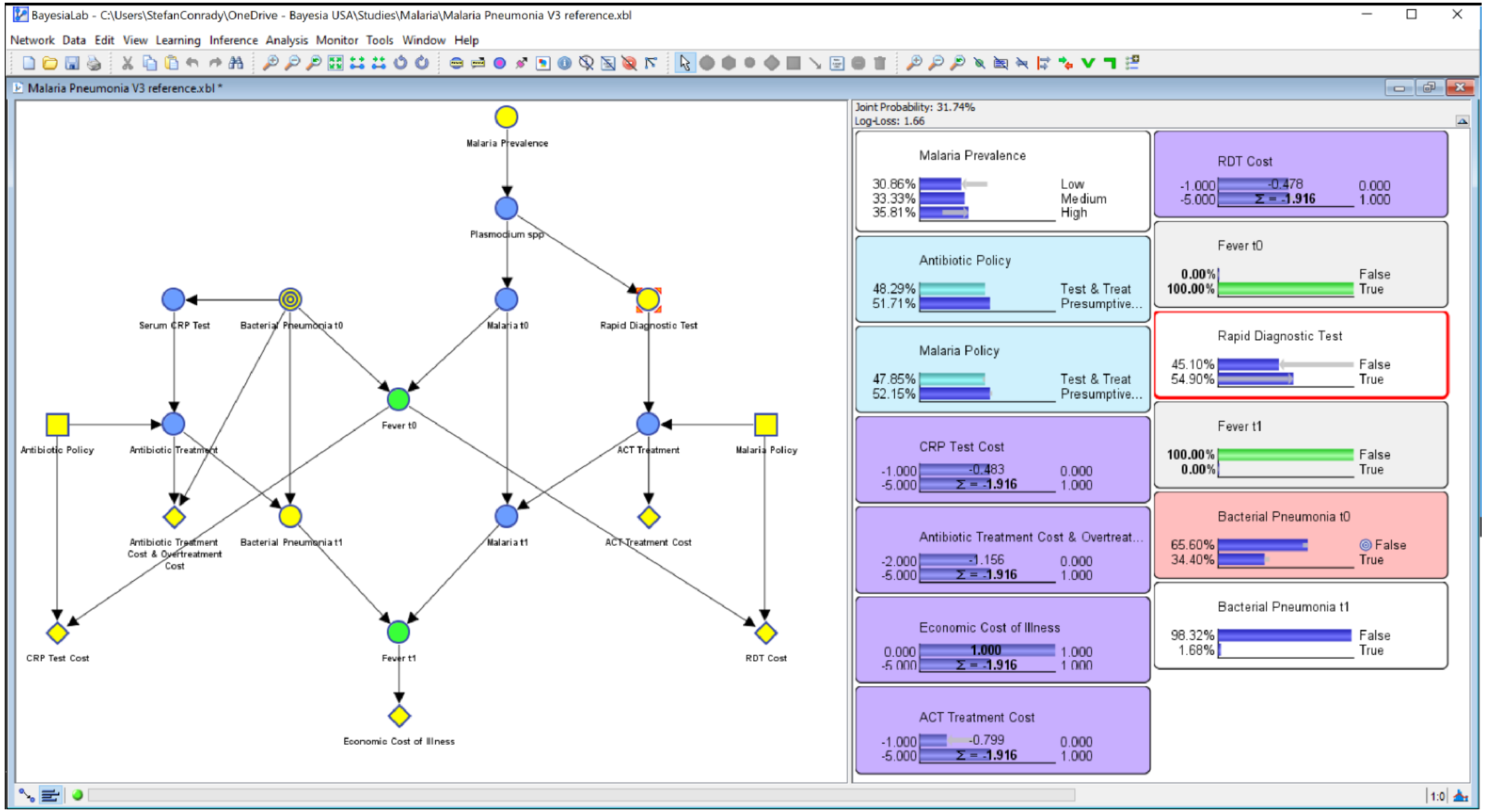
Figure 1 shows an example of such a diagram, derived from a prototype system MUNIN which forms part of a collaborative ESPRIT project to build an advisor in electromyography (EMG).



Spiegelhalter, D. J., & Lauritzen, S. L. (1988). Statistical reasoning and learning in knowledge-bases represented as causal networks *Expert Systems and Decision Support in Medicine (pp. 105-112): Springer.*

Not unlike the BayesiaLab interface!

Bayesian Networks for Health Economics and Public Policy Research slide 139



Bayesian network example

BayesiaLab likely

Epidemiology and Health 2011;33:e2011006

Nguefack-Tsague, G. (2011). Using Bayesian Networks to Model Hierarchical Relationships in Epidemiological Studies. Epidemiol Health, 33(0), e2011006-2011000. doi:10.4178/epih/e2011006

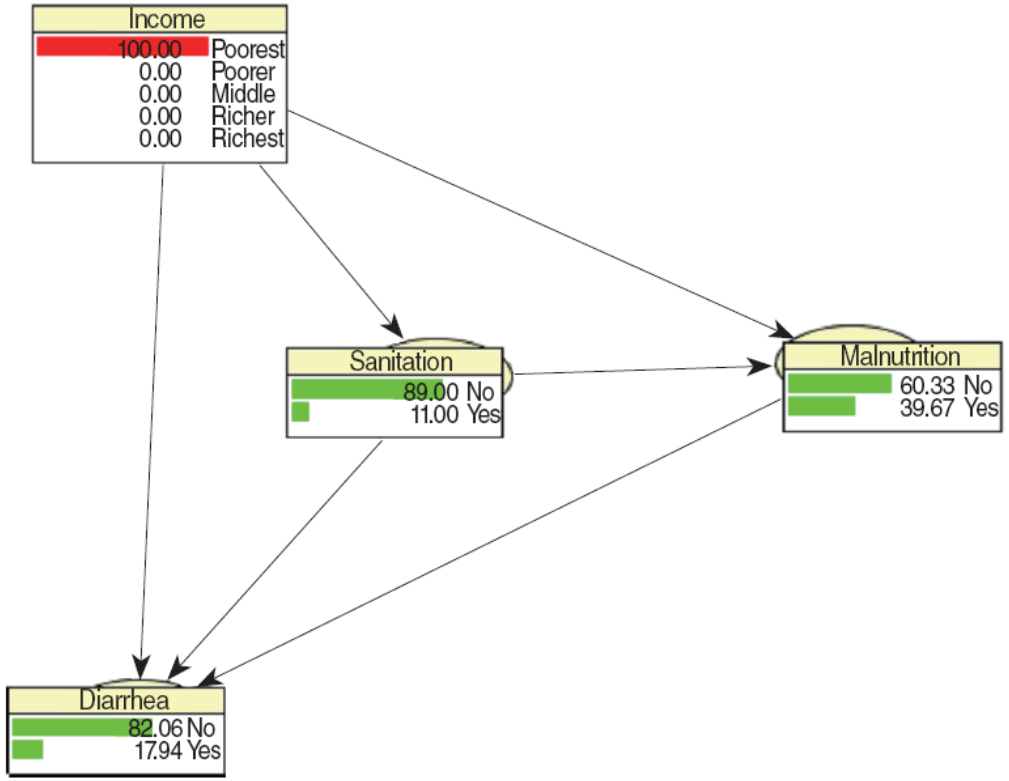


Figure 3. Frequency network showing posterior probabilities (%) when there is evidence that the child belongs to a family in the poorest quintile.

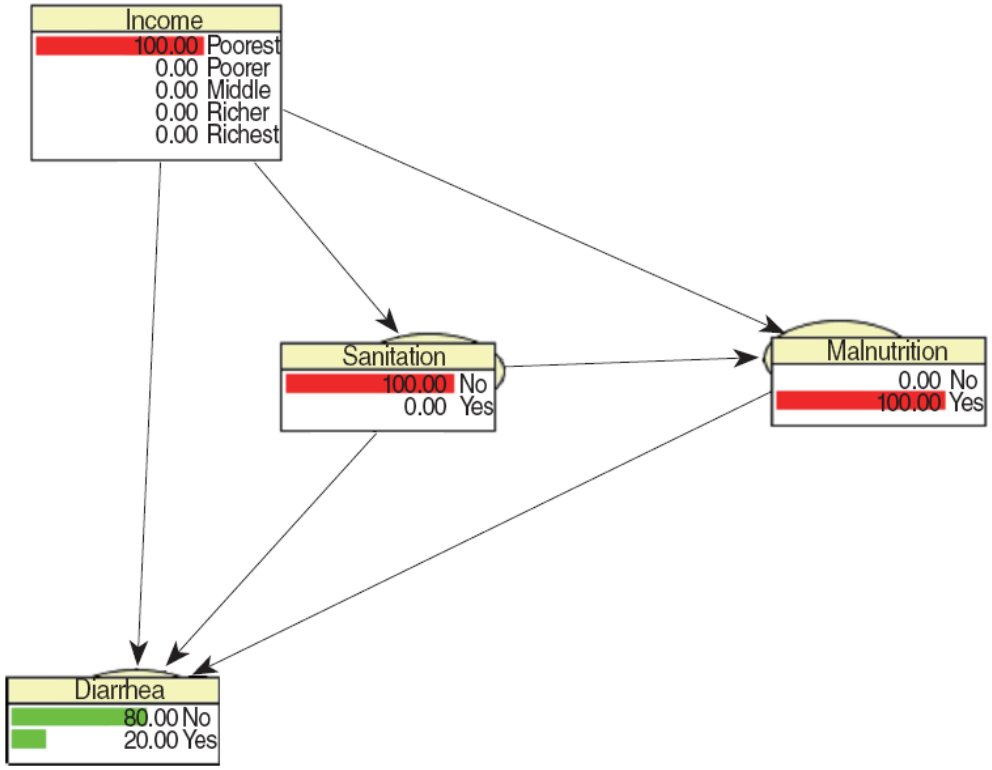


Figure 4. Frequency network showing posterior probabilities (%) of developing diarrhea when there is evidence that the child belongs to family in the poorest quintile, has poor sanitation conditions and is malnourished.

Nguefack-Tsague, G. (2011). Using Bayesian Networks to Model Hierarchical Relationships in Epidemiological Studies. *Epidemiol Health, 33(0), e2011006-2011000. doi:10.4178/epih/e2011006*

How inventions ‘happen’

“Jamie Robins (Figure 9.9), a pioneering statistician and epidemiologist at Harvard University who, together with Sander Greenland at the University of California, Los Angeles, is largely responsible for the widespread adoption of graphical models in epidemiology today. We collaborated for a couple of years, from 1993 to 1995, [and] he got me thinking about the problem of sequential intervention plans, which was one of his principal research interests.

After Jamie flew out to California to meet me on hearing about the “napkin problem” (Chapter 7), he was keenly interested in applying graphical methods to the sequential treatment plans that were his métier. Together we came up with a sequential back-door criterion for estimating the causal effect of such a treatment stream. I learned some important lessons from this collaboration. In particular, he showed me that two actions are sometimes easier to analyze than one because an action corresponds to erasing arrows on a graph, which makes it sparser.” JP p. 328 & 330

“In 1976 Terry Speed invited me to Perth, Australia where he conducted a research seminar exploring relations between statistics and statistical physics. Among other things we studied the relation between the notion of interaction as used in contingency table analysis and in thermodynamics. To our delight they were formally the same […]” (Lauritzen 1996)

Pearl, J., & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect: Hachette UK. My notes

Lauritzen, S. L. (1996). Graphical models: Clarendon Press.

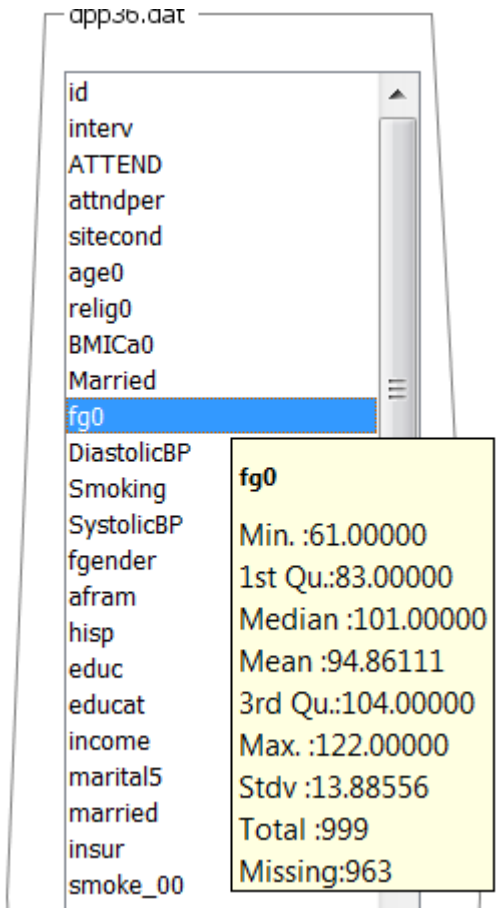
Ωnyx added value

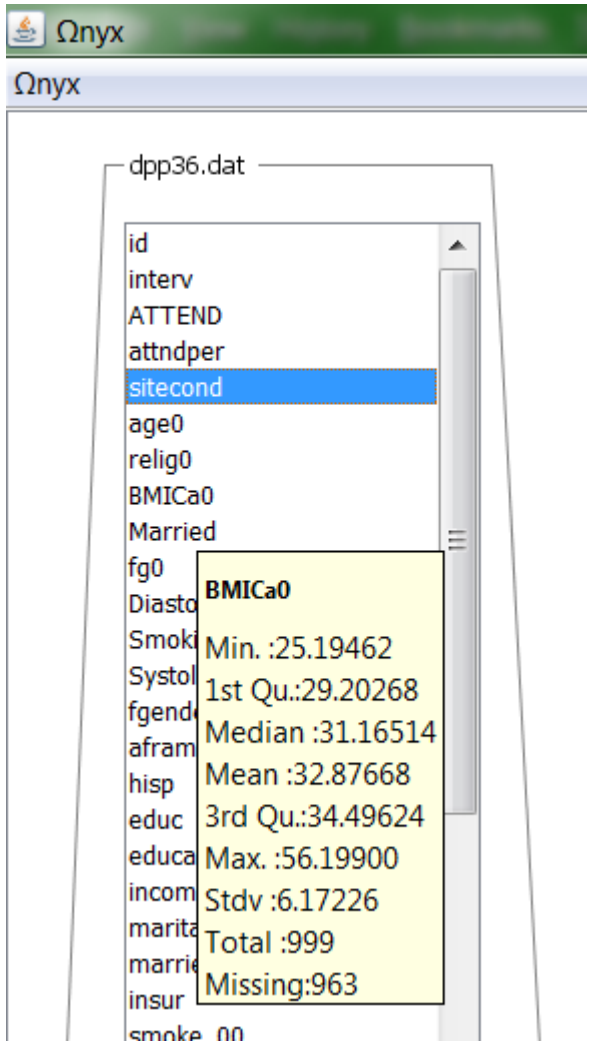
"Structural Equation Modeling is a frequently used multivariate analysis technique in the behavioral and social sciences. SEM are linear models of both observed and latent variables and their relationships. The maximum-likelihood-framework allows estimation of structural parameters even on the latent level by modeling the covariances and expectations of the observed variables.
There are various text books that cover the essentials of SEM, for example, Bollen(1989).
SEM can be conceived of as a unification of several multivariate analysis techniques under a single framework. Particularly, linear regression, ANOVA, correlation, path analysis, factor analysis, autoregression, and growth curve modeling can be considered special cases of SEM."

Ωnyx documentation from: http://onyx.brandmaier.de/documentation/

Robin Beaumont Ωnyx Youtube

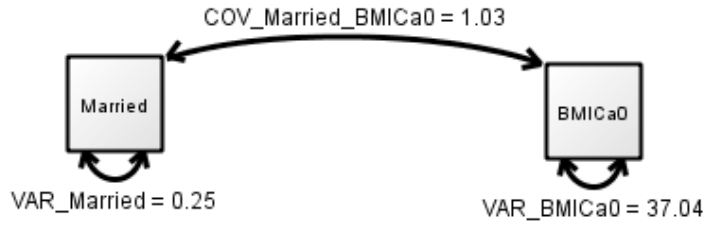
Timo von Oertzen, Brandmaier, A. M., & Tsang, S. (2015). Structural Equation Modeling With Ωnyx. *Structural Equation Modeling: A Multidisciplinary Journal, 22(1), 148-161. doi:10.1080/10705511.2014.935842*





Robin Beaumont Ωnyx Youtube

Unnamed Model * - Maximum Likelihood Estimate (best)

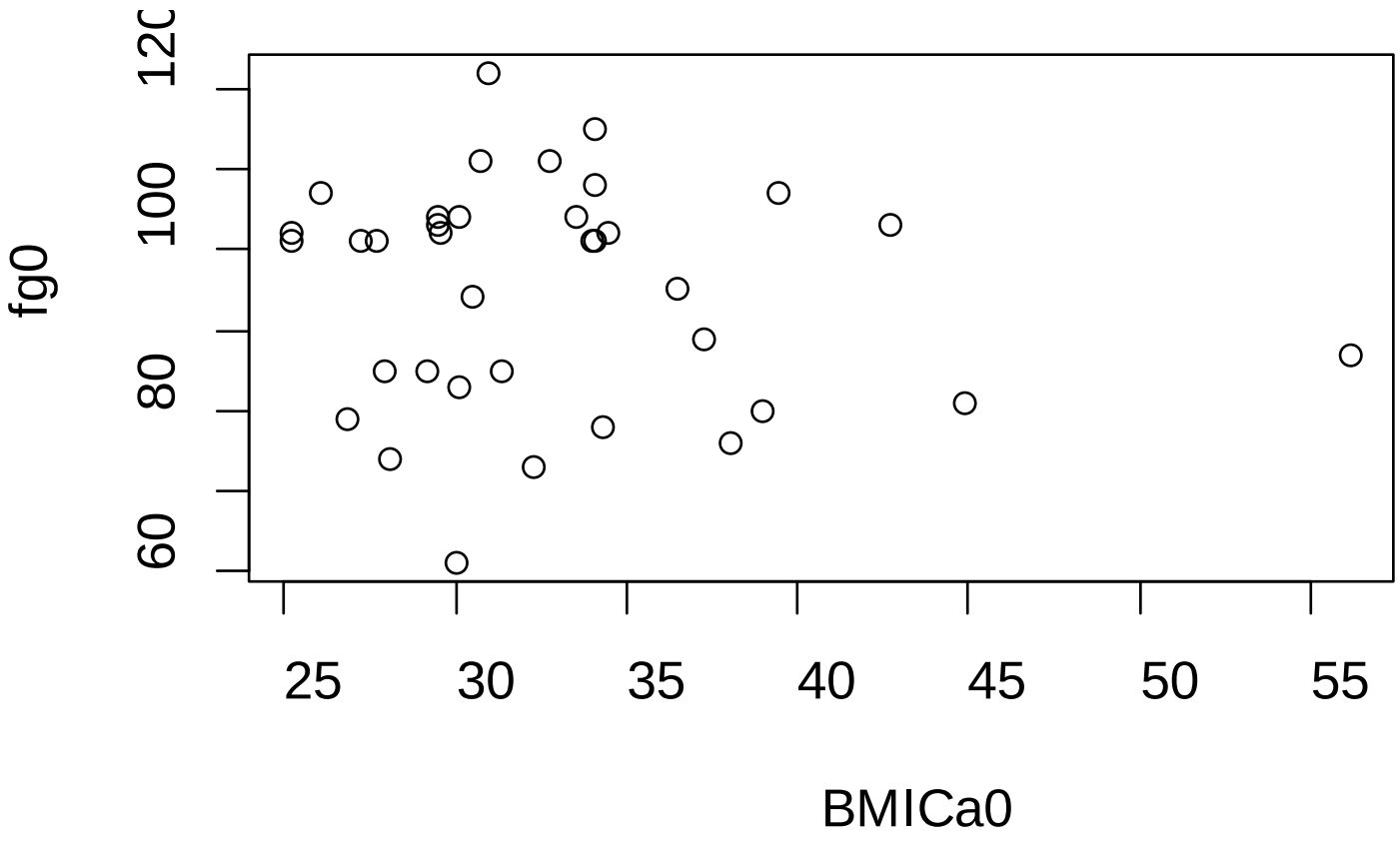


Estimate Summary

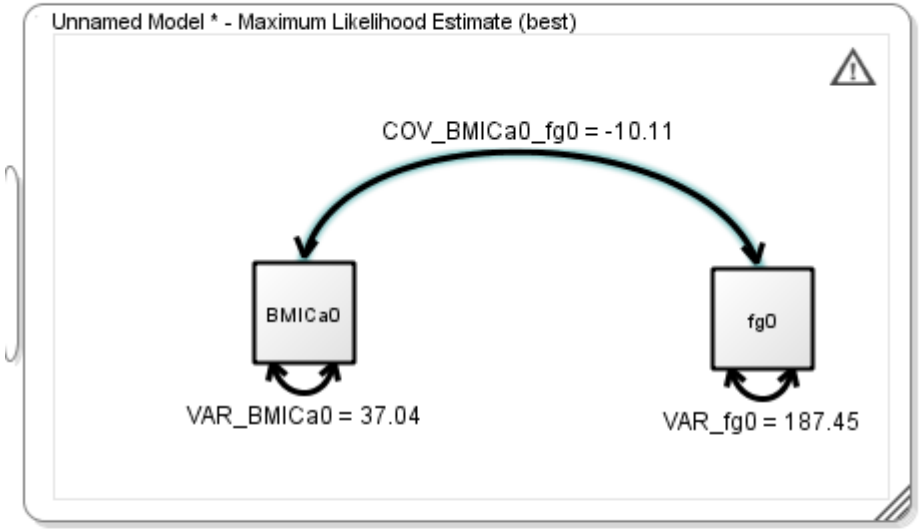
```
1st Qu.:0.00000   1st Qu.:29.20268
Median :1.00000   Median :31.16514
Mean   :0.55556   Mean   :32.87668
3rd Qu.:1.00000   3rd Qu.:34.49624
Max.   :1.00000   Max.   :56.19900
Stdv   :0.50395   Stdv   :6.17226
Total  :36        Total  :36
Missing:0         Missing:0

#|              name|          From / To|Estimate|Std.Error|lbound|rbound
-+------------------+-------------------+--------+---------+------+------
0|       VAR_Married|Married <-> Married| 0.24691|  0.05820|      |
1|        VAR_BMICa0|  BMICa0 <-> BMICa0|37.03850|  8.73006|      |
2|COV_Married_BMICa0| BMICa0 <-> Married| 1.02590|  0.53223|      |

Observed Statistics           : 3
Estimated Parameters          : 3
Non-Missing Ratio             : 0.036
Number of Observations        : 36
Minus Two Log Likelihood      : 279.602
Log Likelihood                : -139.801
Independent -2LL              : 284.004
Saturated -2LL                : 279.602
χ²                            : 0.0
Restricted Degrees of Freedom : 0
```

R 2 variable ‘model’

```
plot(fg0~BMICa0, data =  dpp_36males_Hartford_fewer)
> with (dpp_36males_Hartford_fewer, cov(fg0,BMICa0))
[1] -10.39484
> mycorr <- with (dpp_36males_Hartford_fewer, cor(fg0,BMICa0, method="pearson"))
> mycorr
[1] -0.121286
> rsquared <- mycorr^2
> rsquared
[1] 0.0147103

> mycorr <- with (dpp_36males_Hartford_fewer,
cor.test(fg0,BMICa0, alternative="two.sided",
method="pearson"))
> mycorr
        Pearson's product-moment correlation
data:  fg0 and BMICa0
t = -0.71247, df = 34, p-value = 0.481
alternative hypothesis: true correlation is
not equal to 0
95 percent confidence interval:
 -0.4325846  0.2158507
sample estimates:
      cor
-0.121286
```



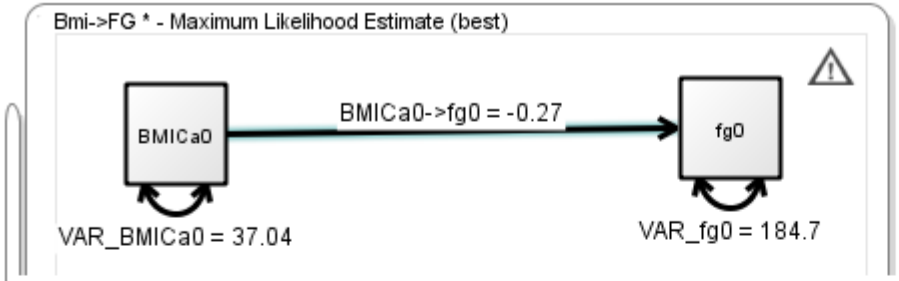
Robin Beaumont Ωnyx Youtube



Estimate Summary

```
BMICa0              fg0
Min.   :25.19462    Min.   :61.00000
1st Qu.:29.20268    1st Qu.:83.00000
Median :31.16514    Median :101.00000
Mean   :32.87668    Mean   :94.86111
3rd Qu.:34.49624    3rd Qu.:104.00000
Max.   :56.19900    Max.   :122.00000
Stdv   :6.17226     Stdv   :13.88556
Total  :36          Total  :36
Missing:0           Missing:0

#|          name|        From / To| Estimate|Std.Error|lbound|rbound
-+--------------+-----------------+---------+---------+------+------
0|    VAR_BMICa0|BMICa0 <-> BMICa0| 37.03850|  8.73006|      |
1|       VAR_fg0|      fg0 <-> fg0|187.45293| 44.18308|      |
2|COV_BMICa0_fg0|   fg0 <-> BMICa0|-10.10610| 13.98919|      |

Observed Statistics            : 3
Estimated Parameters           : 3
Non-Missing Ratio              : 0.036
Number of Observations         : 36
Minus Two Log Likelihood       : 522.231
Log Likelihood                 : -261.116
Independent -2LL               : 522.765
Saturated -2LL                 : 522.231
χ²                             : 0.0
```

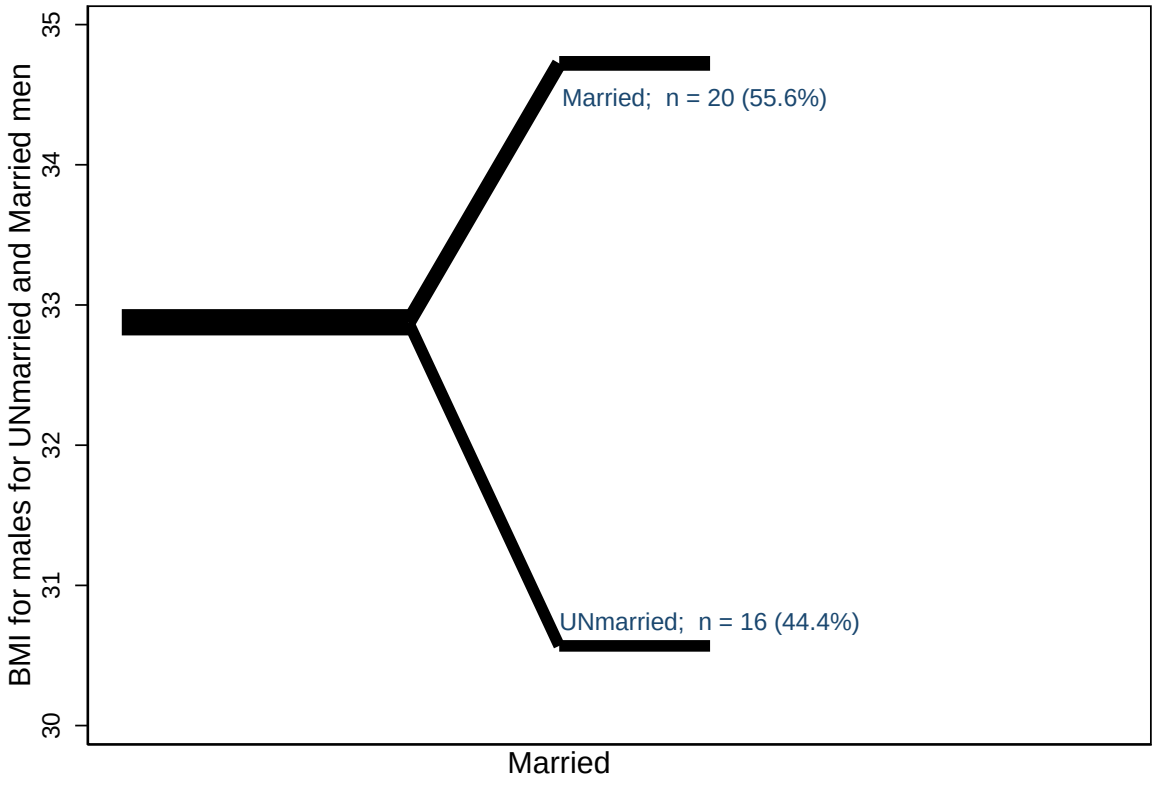
Robin Beaumont Ωnyx Youtube

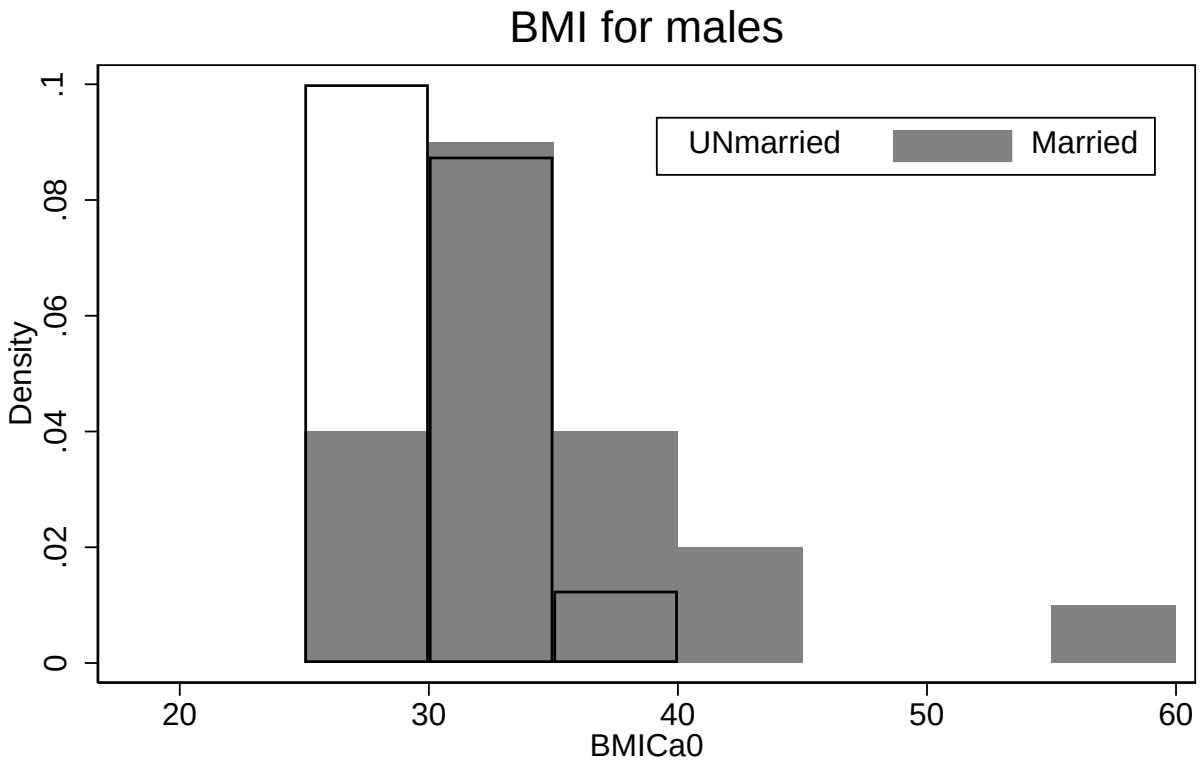


Estimate Summary

```
#|       name|        From / To| Estimate|Std.Error|lbound|rbound
-+-----------+-----------------+---------+---------+------+------
0| VAR_BMICa0|BMICa0 <-> BMICa0| 37.03850|  8.73006|      |
1|    VAR_fg0|      fg0 <-> fg0|184.69545| 43.53314|      |
2|BMICa0->fg0|   BMICa0 --> fg0| -0.27285|  0.37218|      |

Observed Statistics           : 3
Estimated Parameters          : 3
Non-Missing Ratio             : 0.036
Number of Observations        : 36
Minus Two Log Likelihood      : 522.231
Log Likelihood                : -261.116
Independent -2LL              : 522.765
Saturated -2LL                : 522.231
χ²                            : 0.0
Restricted Degrees of Freedom : 0
```

BMI of males: by marital status





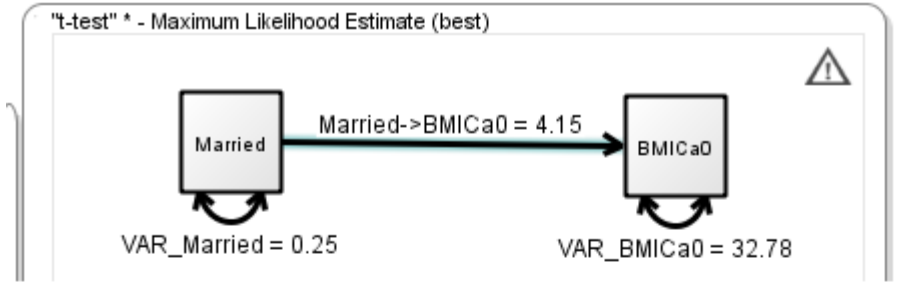
	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Skew	-.0148387	.4194667	-0.04	0.972	-.8369783	.8073009
kurtosis	1.515828	.410243	3.69	0.000	.7117664	2.319889

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Skew	1.539829	.5361251	2.87	0.004	.4890433	2.590615
kurtosis	5.160547	1.927927	2.68	0.007	1.381881	8.939214

Marriage -> BMI of males

Z value for the Married->BMI is 4.15488/1.92024 = 2.16373

```
> zis<-4.15488/1.92024
Zis
[1] 2.16373

t.test
t = -2.1028, df = 34, p-value = 0.04296
sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329

Welch Two Sample t-test
t = -2.2389, df = 30.278, p-value = 0.03266

sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329
```



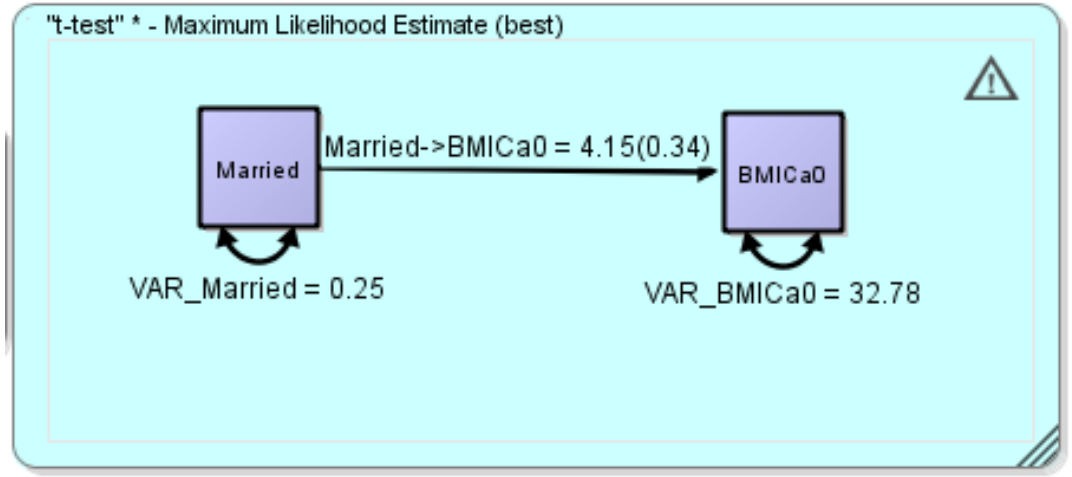
Estimate Summary

#	name	From / To	Estimate	Std.Error
0	VAR_BMICa0	BMICa0 <-> BMICa0	32.77603	7.72538
1	VAR_Married	Married <-> Married	0.24691	0.05820
2	Married->BMICa0	Married --> BMICa0	4.15488	1.92024

Observed Statistics	: 3
Estimated Parameters	: 3
Non-Missing Ratio	: 0.036
Number of Observations	: 36
Minus Two Log Likelihood	: 279.602
Log Likelihood	: -139.801
Independent -2LL	: 284.004
Saturated -2LL	: 279.602
χ^2	: 0.0
Restricted Degrees of Freedom	: 0
AIC	: 285.602
AICc	: 285.602
BIC	: 290.353
BIC (sample-size adjusted)	: 290.353
Kulback-Leibler to Saturated	: 0.0
χ^2 from independent	: 4.401
Degrees of Freedom (indep.)	:1
RMSEA (df corrected)	: 0.0
RMSEA (Kulback Leibler)	: 0.0

RAM notation

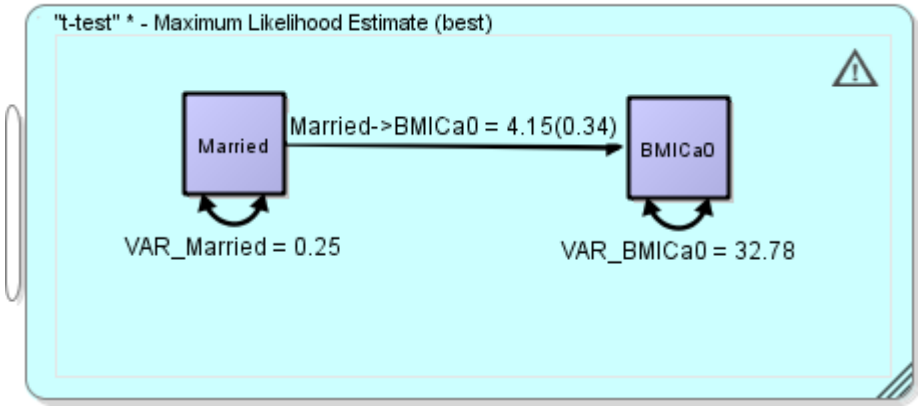
1. McArdle, J. J., & Boker, S. M. (1990). RAMpath: Path diagram software: Data Transforms.
2. McArdle, J. J. (2005). The development of the RAM rules for latent variable structural equation modeling. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald (pp. 225-273).*
3. Zhang, Z., Hamagami, F., Grimm, K. J., & McArdle, J. J. (2015). Using R Package RAMpath for Tracing SEM Path Diagrams and Conducting Complex Longitudinal Data Analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 22(1), 132-147.*



RAM Matrices

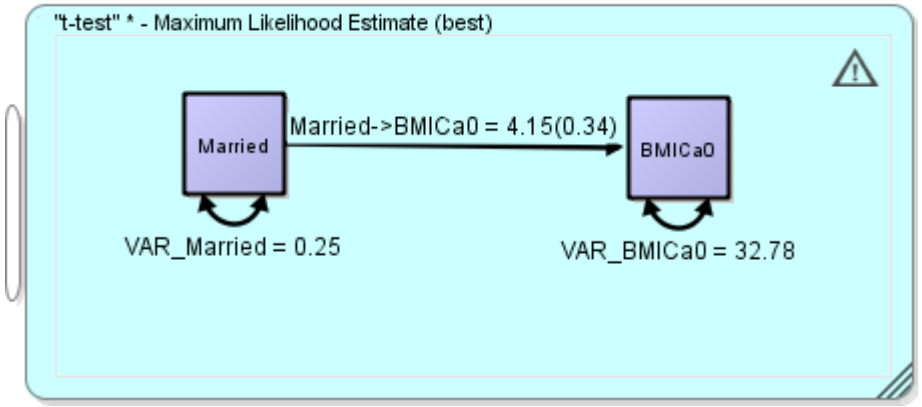
```
Variables: BMICa0, Married

m = (  0.00          0.00         )

A = /  0.00          Marr..0=4.15  \
    \  0.00          0.00          /


S = /  VAR_..0=32.78 0.00          \
    \  0.00          VAR_..d=0.25  /


F = /  1             0             \
    \  0             1             /


Model Covariance matrix = F (I-A)^{-1} S (I-A)^{-T} F^T
Model Mean Vector       = F (I-A)^{-1} m
```

Lavaan code for lazy folk: From Ωnix



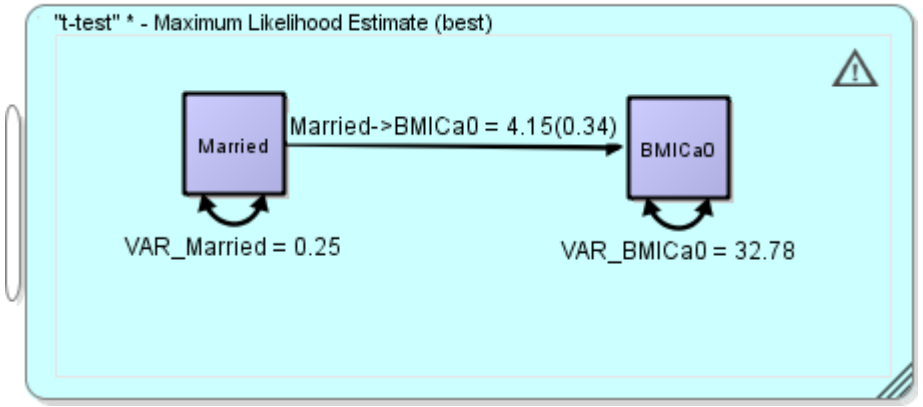

```
# This model specification was automatically generated by Onyx
#
library(lavaan);
modelData <- read.table(DATAFILENAME, header = TRUE) ;
 model<-"
! regressions
   BMICa0 ~ Married__BMICa0*Married
! residuals, variances and covariances
   BMICa0 ~~ VAR_BMICa0*BMICa0
   Married ~~ VAR_Married*Married
! observed means
   BMICa0~1;
   Married~1;
";
result<-lavaan(model, data=modelData, fixed.x=FALSE, missing="FIML");
summary(result, fit.measures=TRUE);
```

sem code for lazy folk: From Ωnix



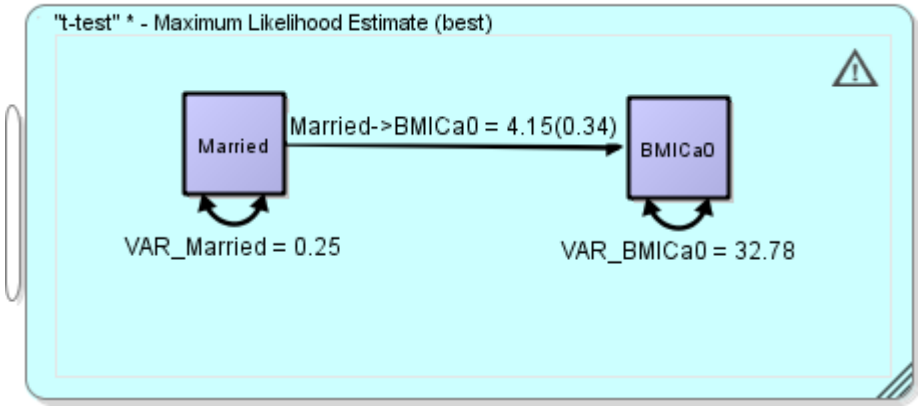

```
# This model specification was automatically generated by Onyx
#
require("sem");
modelData <- read.table(DATAFILENAME, header = TRUE)
paths <- c("BMICa0 <-> BMICa0", "Married <-> Married", "Married -> BMICa0")
parameter <- c("VAR_BMICa0", "VAR_Married", "Married__BMICa0")
values <- c("32.776030306921605", "0.24691358028835825", "4.154875879839016")
model <- array(c(paths, parameter, values), dim = c(3,3))
colnames(model) <- c("col1","col2","col3")
result <- sem(model = model, data = modelData)
summary(result)
```

Lavaan code for lazy folk: From Ωnix




```
# This model specification was automatically generated by Onyx
#
library(lavaan);
modelData <- read.table(DATAFILENAME, header = TRUE) ;
 model<-"
! regressions
   BMICa0 ~ Married__BMICa0*Married
! residuals, variances and covariances
   BMICa0 ~~ VAR_BMICa0*BMICa0
   Married ~~ VAR_Married*Married
! observed means
   BMICa0~1;
   Married~1;
";
result<-lavaan(model, data=modelData, fixed.x=FALSE, missing="FIML");
summary(result, fit.measures=TRUE);
```

OpenMx code for lazy folk: From Ωnix



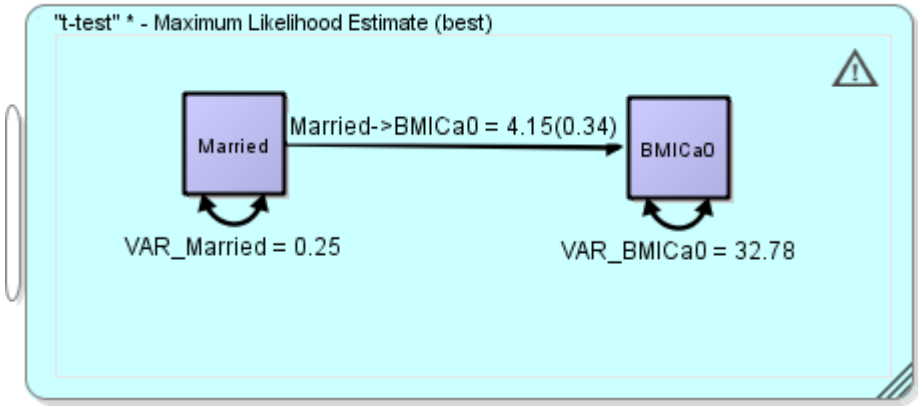

```
# This model specification was automatically generated by Onyx
require("OpenMx");
modelData <- read.table(DATAFILENAME, header = TRUE)
manifests<-c("BMICa0","Married")
latents<-c()
model <- mxModel("_t_test_",
type="RAM",
manifestVars = manifests,
latentVars = latents,
mxPath(from="Married",to=c("BMICa0"), free=c(TRUE), value=c(1.0) , arrows=1, label=c("Married__BMICa0")
),
mxPath(from="BMICa0",to=c("BMICa0"), free=c(TRUE), value=c(1.0) , arrows=2, label=c("VAR_BMICa0") ),
mxPath(from="Married",to=c("Married"), free=c(TRUE), value=c(1.0) , arrows=2, label=c("VAR_Married") ),
mxPath(from="one",to=c("BMICa0","Married"), free=F, value=0, arrows=1),
mxData(modelData, type = "raw")
);
result <- mxRun(model)
summary(result)
```

Appendix: a 2 versions of t tests

```
> t.test(dpp_36males_Hartford_fewer$BMICa0~dpp_36males_Hartford_fewer$married,var.equal=TRUE)

        Two Sample t-test

data:  dpp_36males_Hartford_fewer$BMICa0 by dpp_36males_Hartford_fewer$married
t = -2.1028, df = 34, p-value = 0.04296
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.1703995 -0.1393522
sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329

vs. Welch t-test
Welch Two Sample t-test
data:  dpp_36males_Hartford_fewer$BMICa0 by dpp_36males_Hartford_fewer$married
t = -2.2389, df = 30.278, p-value = 0.03266
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.9434686 -0.3662831
sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329
```

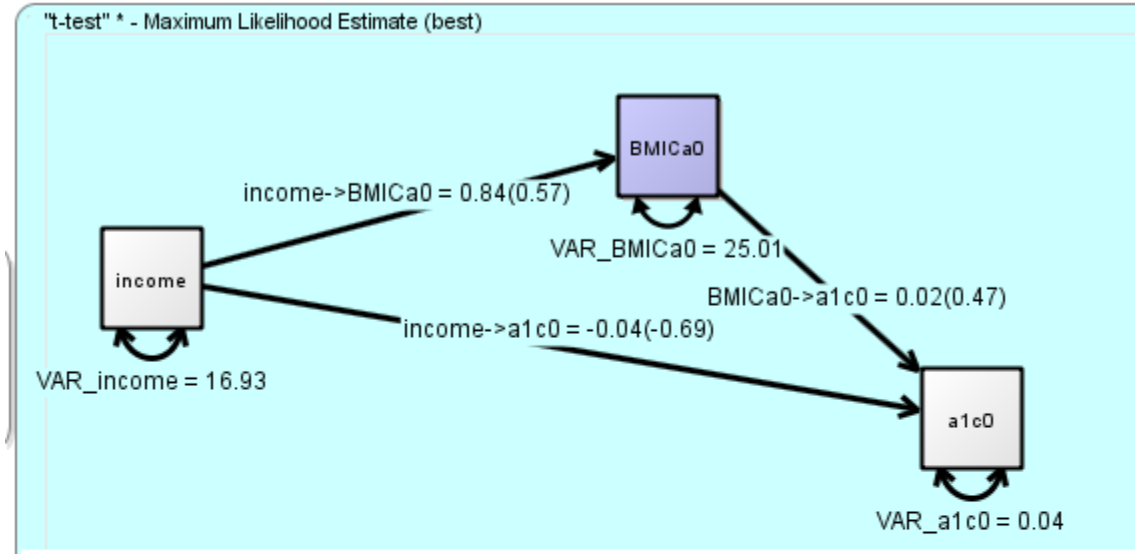
Mplus code for lazy folk: From Ωnix

```
Variables: BMICa0, Married
37.0385 1.0259
1.0259  0.2469

!This model specification was automatically created by Onyx
TITLE:
   "t-test"
DATA:
   FILE IS "DATAFILENAME";
VARIABLE:   NAMES ARE BMICA0 MARRIED ;
   USEVARIABLES ARE BMICA0 MARRIED ;
MODEL:
! regressions of latents on manifest
! regressions of manifest on manifest
   BMICA0 ON MARRIED*1.0;
! regressions of latents on latents or manifests
! residuals, variances and covariances
   BMICA0*1.0;
   MARRIED*1.0;
ANALYSIS:
   TYPE = general;
   ESTIMATOR = ml;
OUTPUT:
   sampstat;
```



A more informed model: Ωnix

There is likely an indirect effect too here.



Estimate Summary

Missing:963 Missing:990 Missing:973

#	name	From / To	Estimate	Std.Error	lbound	rbound
0	VAR_BMICa0	BMICa0 <-> BMICa0	25.01032	8.61107		
1	VAR_income	income <-> income	16.93030	6.92982		
2	income->BMICa0	income --> BMICa0	0.84288	0.29039		
3	VAR_a1c0	a1c0 <-> a1c0	0.03995	0.01396		
4	BMICa0->a1c0	BMICa0 --> a1c0	0.01900	0.01112		
5	income->a1c0	income --> a1c0	-0.04072	0.01791		

Appendix: a

```
fit <- sem(SEMJaccApp_d, data = dpp_36males_Hartford_fewer)
> summary(fit, rsq = T)
lavaan (0.5-23.1097) converged normally after  17 iterations
  Number of observations                            36
  Estimator                                         ML
  Minimum Function Test Statistic                0.000
  Degrees of freedom                                 0
Parameter Estimates:
  Information                                 Expected
  Standard Errors                             Standard
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
 BMICa0 ~
    married           4.155    1.920    2.164    0.030    4.155    0.339
Variances:
                  Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .BMICa0           32.776    7.725    4.243    0.000   32.776    0.885
[only 1 variance estimated: but we have 2 groups…]
R-Square:
                  Estimate
    BMICa0           0.115
vs. Welch t-test
Welch Two Sample t-test
data:  dpp_36males_Hartford_fewer$BMICa0 by dpp_36males_Hartford_fewer$married
t = -2.2389, df = 30.278, p-value = 0.03266
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.9434686 -0.3662831
sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329
```

Appendix: a

```
SEMJaccApp_a2 <-   '
BMICa0~1
'
 fit <- sem(SEMJaccApp_a2,
+            data = dpp36,
+            group = "married")
Estimator                                         ML
  Model Fit Test Statistic                       0.000
  Degrees of freedom                                 0
Number of observations per group
  0                                                 16
Group 1 [0]:
Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0           30.568    0.929   32.917    0.000
Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0           13.798    4.878    2.828    0.005
vs. Welch t-test
Welch Two Sample t-test
t = -2.2389, df = 30.278, p-value = 0.03266
95 percent confidence interval:-7.9434686 -0.3662831
sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329
```

```
lavaan 0.6-3 ended normally after 15 iterations
 Optimization method                           NLMINB
  Number of free parameters                         4
Chi-square for each group:
  0                                             0.000
  1                                             0.000
Parameter Estimates:
  Information                                 Expected
  Information saturated (h1) model          Structured
  Standard Errors                             Standard
Number of observations per group
  1                                                20
Group 2 [1]:
Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0           34.723    1.549   22.424    0.000
Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0           47.958   15.166    3.162    0.002
vs. Welch t-test
Welch Two Sample t-test
t = -2.2389, df = 30.278, p-value = 0.03266
alternative hypothesis: true difference in means is
not equal to 0
mean in group 0 mean in group 1
       30.56842        34.72329
```

Appendix: a

```
> anova(fitdif,fiteq )
Chi Square Difference Test
       Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fitdif  0 236.19 239.36 0.0000
fiteq   1 232.45 237.20 4.8911     4.8911       1      0.027 *
```

```
> fiteq <- sem(SEMJaccApp_a2,
data = dpp36, group = "married",
group.equal = c("intercepts"))
lavaan 0.6-3 ended normally after 15 iterations
 Optimization method                           NLMINB
  Number of free parameters                         4
 Estimator                                         ML
  Model Fit Test Statistic                       4.891
  Degrees of freedom                                 1
  P-value (Chi-square)                           0.027
Number of observations per group
  0                                                 16
Group 1 [0]:
Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0   (.p1.)   31.576    0.837   37.703    0.000
Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0            14.813    5.237    2.828    0.005
vs. Welch t-test
Welch Two Sample t-test
t = -2.2389, df = 30.278, p-value = 0.03266
95 percent confidence interval:
 -7.9434686 -0.3662831
sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329
```

```
Chi-square for each group:
 0                                                 1.135
 1                                                 3.756
Parameter Estimates:
  Information                                   Expected
  Information saturated (h1) model            Structured
  Standard Errors                               Standard
Number of observations per group
  1                                                   20
Group 2 [1]:
Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0   (.p1.)   31.576    0.837   37.703    0.000
Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    BMICa0            57.866   18.299    3.162    0.002
vs. Welch t-test
Welch Two Sample t-test
t = -2.2389, df = 30.278, p-value = 0.03266
alternative hypothesis: true difference in means is
not equal to 0
sample estimates:
mean in group 0 mean in group 1
       30.56842        34.72329
```

Appendix a : Compare means in 2 independent groups

Classic

Model is… a t-test

Welch Two Sample t-test

t(df=1) = -2.2389, p = 0.033

Modern

Model is a 2 group 1 effect variable (no cause)

[$ContinuousEffect_{Group1}$ & $ContinousEffect_{Group2}$]

Chi Square Difference Test

χ^2diff(df=1) = 4.8911 , p= 0.027

Pre-data: in DAG world

```
# 2 DAG
DagJaccApp_a2 <- dagitty('dag {
Married[pos="1,2"]
BMI [pos="2,2"]
Married-> BMI
}')

plot (DagJaccApp_a2)

#then can technically simulate data DID NOT WORK
DagJaccApp_a2 <- simulateSEM( g, .2, .3)
coef( summary( lm( BMI ~ Married, DagJaccApp_a2 ) ) )
```

Johannes Textor, Andrew Forney, and Judea Pearl (2016). Causal Inference In Statistics: A Companion for R Users
Banack, H. R., & Kaufman, J. S. (2015). From bad to worse: collider stratification amplifies confounding bias in the “obesity paradox”. *European journal of epidemiology, 1-4.*

Pre-data: in MIIVsem world

```
##1 MIIVsem
model.MiivJaccApp_a2 <- '
BMI ~ Married
+ '

miivs(model.MiivJaccApp_a2)
Model Equation Information

 LHS RHS     MIIVs
 BMI Married Married
```

CRAN - Package MIIVsem
GitHub - zackfisher/MIIVsem
Bollen, K. A., & Fisher, Z. (2017). *Model Implied Instrumental Variable (MIIV) Methods using MIIVsem: An R Package for Structural Equation Models (SEMs). Paper presented at the Modern Modeling Methods (M3) Conference, Storrs, CT.*
https://miivs.shinyapps.io/miivs/
Bollen, K. A. (2018). Model Implied Instrumental Variables (MIIVs): An Alternative Orientation to Structural Equation Modeling. *Multivariate Behavioral Research, 1-16.*

Example 2: obesity paradox” **Pre-data: in DAG world**

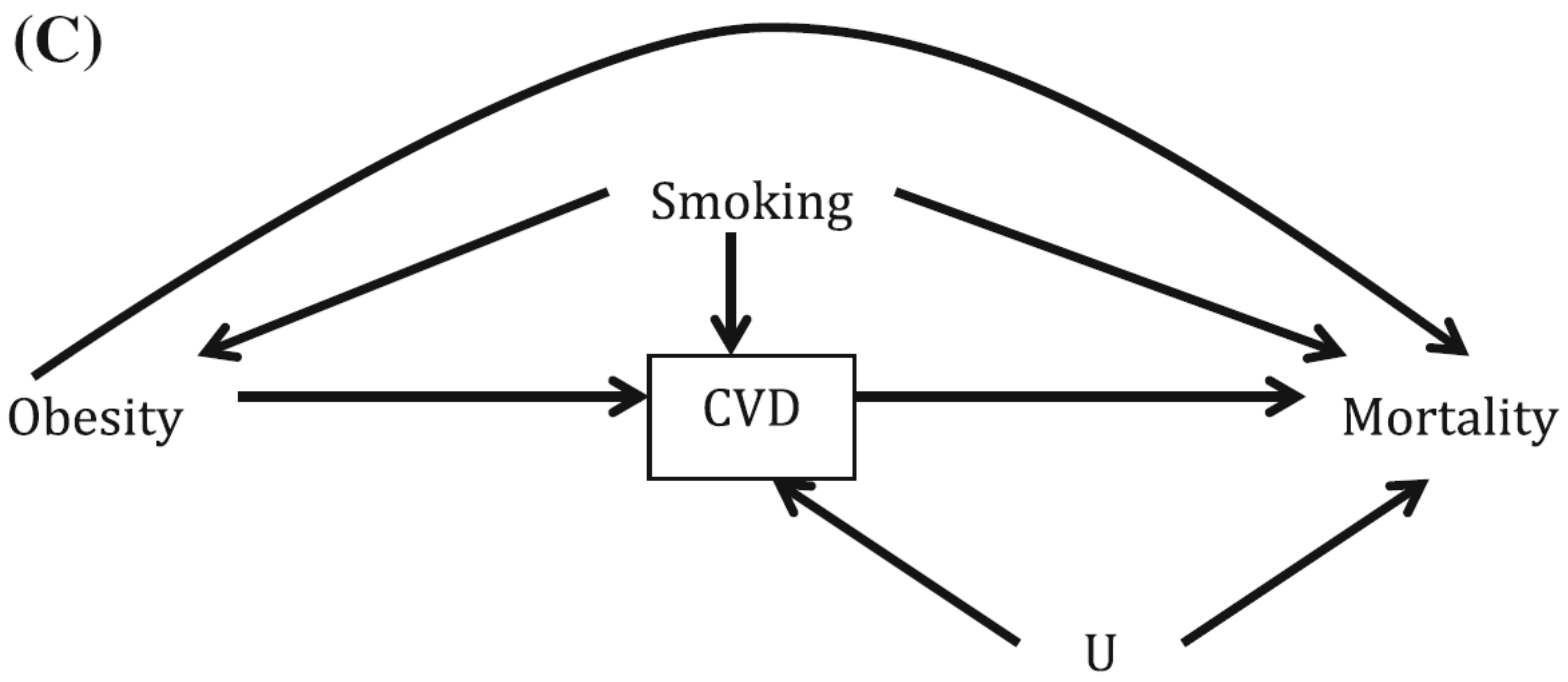
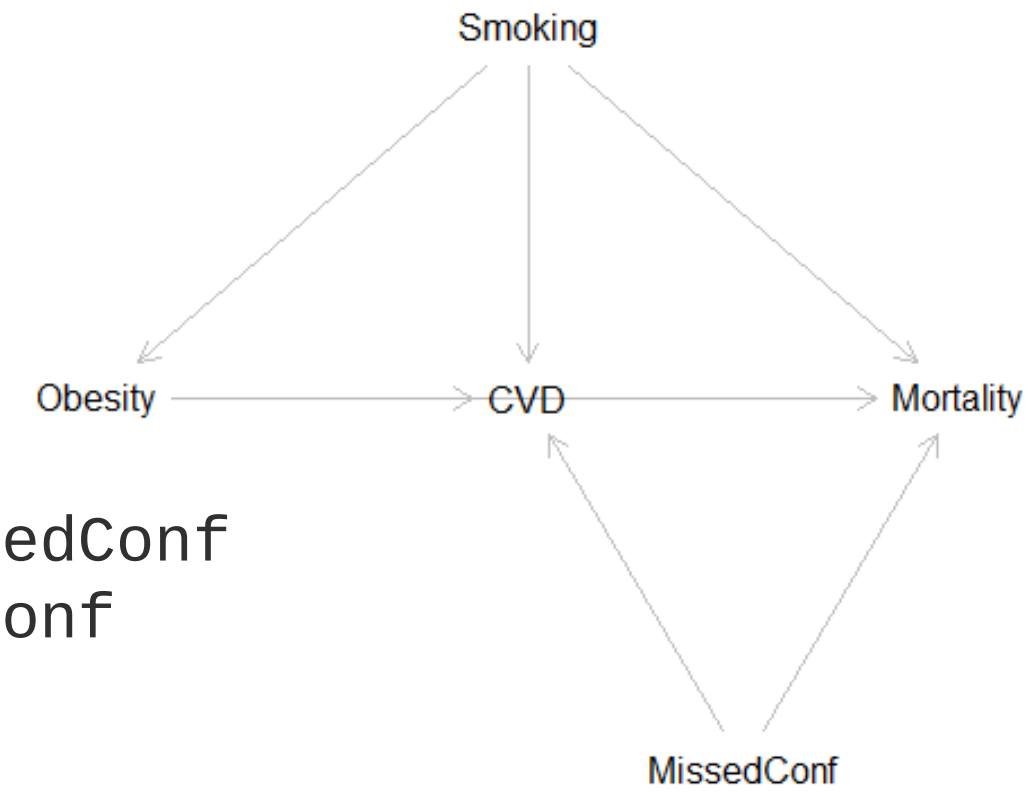


Fig. 1 a–c Directed acyclic graphs representing causal relations between obesity, cardiovascular disease (CVD), and mortality. Smoking and unmeasured factors (U) are included as confounders in (**b**, **c**), respectively

Banack, H. R., & Kaufman, J. S. (2015). From bad to worse: collider stratification amplifies confounding bias in the “obesity paradox”. *European journal of epidemiology, 1-4.*

Obesity paradox” **Pre-data: in DAG world**

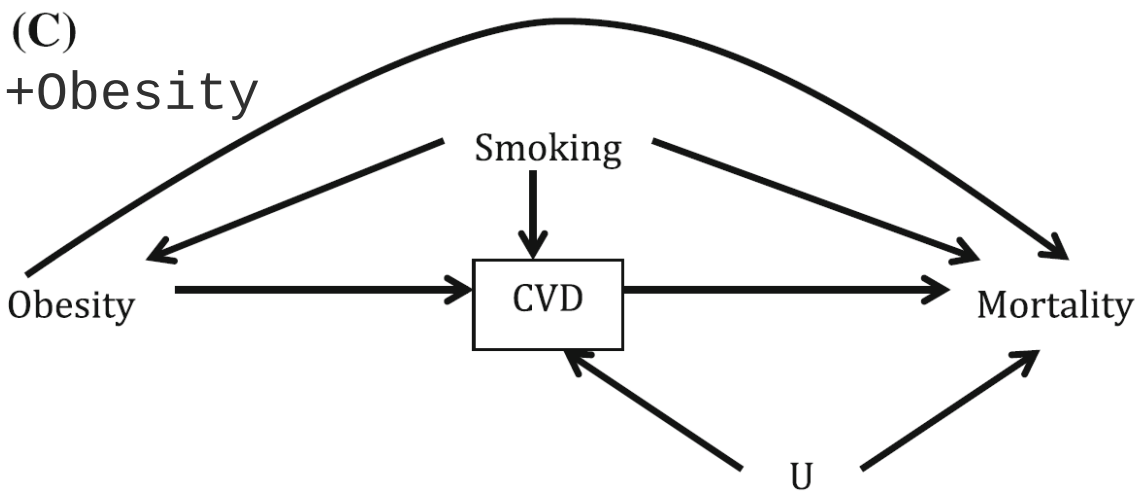
```
g <- dagitty('dag {
Obesity [pos="0,1"]
CVD [pos="1,1"]
Mortality [pos="2,1"]
Smoking [pos="1,0"]
MissedConf [pos="1.5,2"]
Obesity  -> CVD  -> Mortality  <- MissedConf
Obesity  <- Smoking -> CVD  <- MissedConf
Smoking -> Mortality
Obesity-> Mortality
}')
plot(g)
```



Johannes Textor, Andrew Forney, and Judea Pearl (2016). Causal Inference In Statistics: A Companion for R Users
Banack, H. R., & Kaufman, J. S. (2015). From bad to worse: collider stratification amplifies confounding bias in the “obesity paradox”. *European journal of epidemiology, 1-4.*

Obesity paradox” **Pre-data: in MIIVsem**

```
model.obesparadox <- '
 Mortality  ~ CVD  +MissedConf + Smoking +Obesity
CVD   ~ Obesity  + MissedConf + Smoking
Obesity  ~ Smoking
 '
```




```
miivs(model.obesparadox)
Model Equation Information

 LHS        RHS                                  MIIVs
 Mortality MissedConf, Smoking, CVD, Obesity CVD, Obesity, MissedConf, Smoking
 CVD        MissedConf, Smoking, Obesity        Obesity, MissedConf, Smoking
 Obesity    Smoking                             MissedConf, Smoking
```

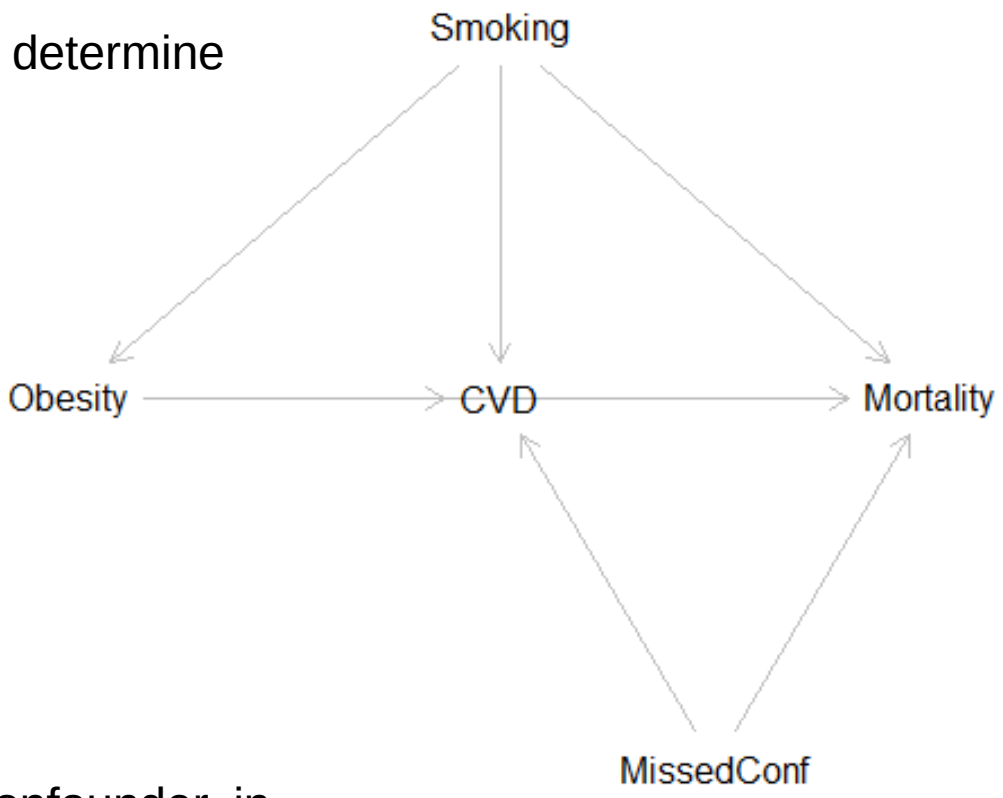
GitHub - zackfisher/MIIVsem

Obesity paradox” Which adjustments are required

“List all of the sets of variables that satisfy the backdoor criterion to determine the causal effect of Obesity on *Mortality*.”

```
adjustmentSets( g, "Obesity",
"Mortality", type="all" )
```

```
{ Smoking }
{ MissedConf, Smoking }
```



This is the cherry on the cake:
One **HAS TO to measure** smoking, OR smoking and the Missed Confounder, in order to identify the causal effect of Obesity on Mortality.

Johannes Textor, Andrew Forney, and Judea Pearl (2016). Causal Inference In Statistics: A Companion for R Users
Banack, H. R., & Kaufman, J. S. (2015). From bad to worse: collider stratification amplifies confounding bias in the “obesity paradox”. *European journal of epidemiology, 1-4.*

Why bother

Modern

1. It allows one to think in causal modeling manner

2. It is more flexible modeling: allows for fewer ‘unrelaxable’ assumptions, and can test assumptions along the way

3. The reality operates as different effects in different populations, e.g. health disparities: the effect in income on health differs for races/ethnic groups

4. It open the doors for formal causal inquiries: when and how can one recover causal effects with simple regressions, e.g.

What you may have learned

1. R is not that scary: nor is causal modeling
2. Causal calculus' has been implemented and is easy to use on dagitty
3. Several traditions have been merged in designing *MIIVsem* (economics, latent variables)
4. One can better teach/train by using graphical models: Onyx is there to help.
5. Answering 'what if' questions require solid causal footing:
the science of cause and effect is here with us

Next:

The science of cause and effect is 'complete': Elias Bareinboim will come to Storrs to prove it.

Grab (and read) and go

Beaujean, A. A. (2014). Latent variable modeling using R: A step-by-step guide: Routledge.
Nagarajan, R., Scutari, M., & Lèbre, S. (2013). Bayesian networks in R. Springer, 122, 125-127.
Lauritzen, S. L. (1996). Graphical models: Clarendon Press.
Lauritzen, S. L. (1979). Lectures on contingency tables: Inst. of mathematical statistics, University of Copenhagen.
Lauritzen, S. L. (2001). Causal inference from graphical models. In D. Cox & C. Kluppelberg (Eds.), Complex stochastic systems (pp. 63-107).
Lauritzen, S. L. (2001). Causal inference from graphical models. PPT