

Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

# How to Test DAG Models

Turing Institute Causal Inference Interest Group, November 8th, 2021

Johannes Textor

Data Science, Radboud University  
Tumor Immunology, Radboudumc  
Nijmegen, The Netherlands  
[computational-immunology.org](http://computational-immunology.org)

## Acknowledgements

### RADBOUD U(MC)



Ankur Ankan



Inge Wortel

### UNIVERSITÄT ZU LÜBECK

Juliane Hardt  
Benito van der Zander  
Maciej Liskiewicz



### ETH ZÜRICH

Ema Perkovic  
Markus Kalisch  
Marloes Maathuis

### UNIVERSITY OF LEEDS

Peter Tennant  
Mark Gilthorpe  
George Ellison

### DIFE POTSDAM

Sven Knüppel

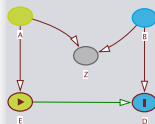
### FUNDING

Dutch Cancer Society (KWF)  
Dutch Research Council (NWO)  
Human Frontiers Science Program  
Hanarth Fonds



### How to Test DAG Models

Johannes Textor



Why Test DAG Models?

Categorical Data

Continuous Data

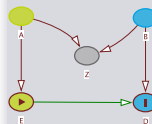
Mixed Data

# Overview

- 1 Why Test DAG Models?
- 2 Categorical Data
- 3 Continuous Data
- 4 Mixed Data

## How to Test DAG Models

Johannes Textor



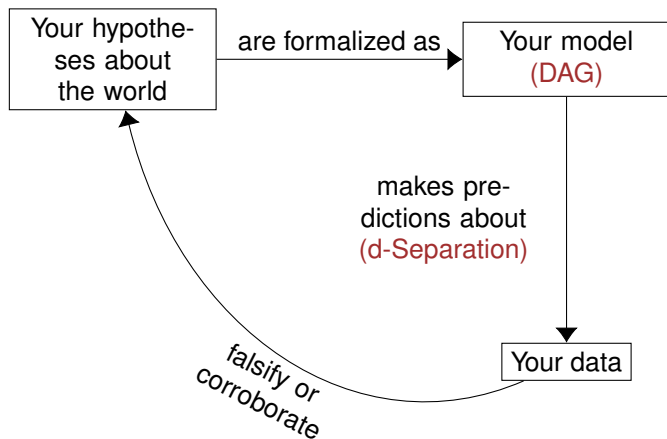
Why Test DAG Models?

Categorical Data

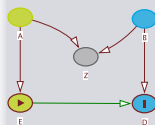
Continuous Data

Mixed Data

## The Model Building Cycle



This doesn't only apply to DAGs, but many other kinds of models as well.



### Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

## Why Test DAG Models?

- DAGs are **models** of variable relationships in a certain domain.
- DAGs encode certain **assumptions** about these relationships.
- Any **inference** drawn from the DAG (such as an adjustment set) hinges on the **assumptions** that are made.
- Incorrect assumptions may lead to **incorrect** inferences.

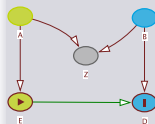
To guard against wrong inference, we can test our model before we use it.

### Caveats

- Model testing never guarantees a correct model! It can only refute, but never prove it.
- Bad actors can use model testing to iteratively tweak a model until it perfectly fits a given dataset.

## How to Test DAG Models

Johannes Textor



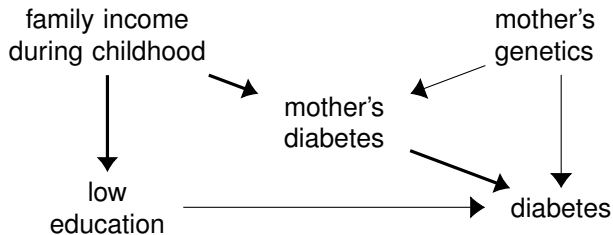
### Why Test DAG Models?

Categorical Data

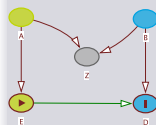
Continuous Data

Mixed Data

## How Can We Tell If This DAG Is Correct?



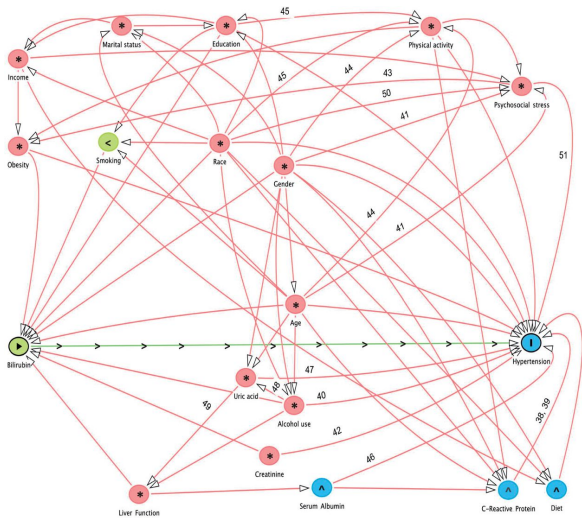
*Rothman, Greenland & Lash, Modern Epidemiology, 2008*



### Why Test DAG Models?

- Categorical Data
- Continuous Data
- Mixed Data

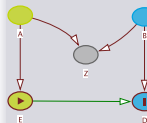
# How Can We Tell If This DAG Is Correct?



▶ Exposure; | Outcome; ◀ Ancestor of exposure; ◀ Ancestor of outcome; \* Ancestor of exposure and outcome; → Causal path; → Biasing path

## How to Test DAG Models

Johannes Textor



## Why Test DAG Models?

- Categorical Data
- Continuous Data
- Mixed Data

Wang and Bautista, *IJE* 2015, <https://doi.org/10.1093/ije/dyu242>

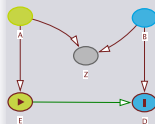
## Testing Graphical Causal Models Using the R Package “dagitty”

Ankur Ankan, Inge M. N. Wortel, Johannes Textor 

First published: 16 February 2021 | <https://doi.org/10.1002/cpz1.45>

Paper is available at:  
<https://doi.org/10.1002/cpz1.45>

Code is available at:  
<https://github.com/ankurankan/2020-dagitty-manual/>





## d-Separation in DAGs

### Definition

A **path** in a DAG is a sequence of variables (possibly with repetitions) that are connected by edges.

Note that this differs from the “intuitive” concept of a path because we allow to move against arrow directions:

$x \rightarrow m \leftarrow y$  is a path too!

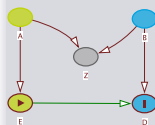
### d-Separation

A path is **d-separated** if

- It contains a collider  $x \rightarrow m \leftarrow y$  where  $m$  and all descendants of  $m$  **are not conditioned on**;
- It contains a non-collider  $x \rightarrow m \rightarrow y$ ,  $x \leftarrow m \rightarrow y$  or  $x \leftarrow m \leftarrow y$  where  $m$  **is conditioned on**.

This means that:

- A one-arrow path  $x \rightarrow y$  is never d-separated.
- Conditioning may close but also open paths.



### Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

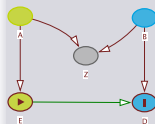
## How $d$ -Separation Links DAGs to Data

### Theorem (Verma & Pearl, 1984)

If conditioning on  $\mathbf{Z}$   $d$ -separates  $X$  and  $Y$  in a DAG  $G$ , then  $X \perp\!\!\!\perp Y \mid \mathbf{Z}$  in **every** probability density  $P$  that factorizes according to  $G$ .

DAG structure	What it claims about the data
$X \rightarrow Y$	Nothing
$X \rightarrow Y \rightarrow Z$	$X \perp\!\!\!\perp Y \mid Z$
$I \rightarrow X \rightarrow Y \rightarrow Z$	$I \perp\!\!\!\perp Y \mid X$
	$I \perp\!\!\!\perp Z \mid X$
	$I \perp\!\!\!\perp Z \mid Y$
	$X \perp\!\!\!\perp Z \mid Y$

- Larger DAGs **can** make more claims about the data.
- Each claim depends on at least one **missing arrow**.



### Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

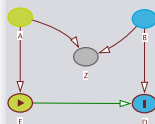
# How Do We Test DAG Implications?

It depends!

- The general **theory** of  $d$ -separation is **non-parametric**, but
- depending on the type of data involved, different **statistical tests** are needed.

We will cover a few example tests for the following cases:

- Categorical data only
- Continuous data only
- Mix of ordinal and continuous data



Why Test DAG Models?

Categorical Data

Continuous Data

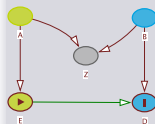
Mixed Data

## Main Types of Conditional Independence Tests

There are three main approaches to test a statement  $X \perp\!\!\!\perp Y \mid Z$ :

- **Stratification-based tests** split the dataset according to the values of  $Z$ , perform a simple independence test  $X \perp\!\!\!\perp Y$  in each part, and combine the results.
- **Variable importance-based tests** compare the fit of a model  $\hat{p}(X \mid Y, Z)$  to a simpler model  $\hat{p}(X \mid Z)$ . If  $X \perp\!\!\!\perp Y \mid Z$ , the simpler model should fit just as well.
- **Residualization-based tests** fit two (regression) models  $\mathbb{E}[X|Z]$  and  $\mathbb{E}[Y|Z]$ , and test for independence between the residuals.

Each approach has distinct advantages and disadvantages.



## The Chi-Square Test for Conditional Independence

To test conditional independence  $X \perp\!\!\!\perp Y \mid Z$ , we

- 1 split the dataset according to  $Z$ ;
- 2 perform a simple chi-square test of  $X \perp\!\!\!\perp Y$  in each part;
- 3 combine the results.

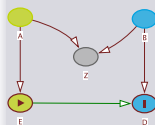
### Combining results of chi-square tests

If  $T_0 \sim \chi^2(a)$  and  $T_1 \sim \chi^2(b)$  are two chi-square distributed variables with  $a$  and  $b$  degrees of freedom, then

$$T_1 + T_2 \sim \chi^2(a + b)$$

```
attach( dagitty::simulateLogistic( "X <- Z -> Y" , b.default=1 ) )
T.Z0<-chisq.test(X[Z== -1],Y[Z== -1])
T.Z1<-chisq.test(X[Z== 1],Y[Z== 1])
chisq.combined <- T.Z0$statistic + T.Z1$statistic
df.combined <- T.Z0$parameter + T.Z1$parameter
c(chisq.combined,df.combined)
```

```
## X-squared      df
##      0.93      2.00
```



Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

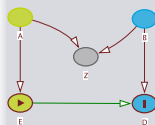
## The Root Mean Square Error of Approximation (RMSEA)

Instead of a p-value, which conflates information about dependence strength and sample size, an **effect size** is often more useful. For chi-square tests, various effect sizes can be defined. An important one is the RMSEA:

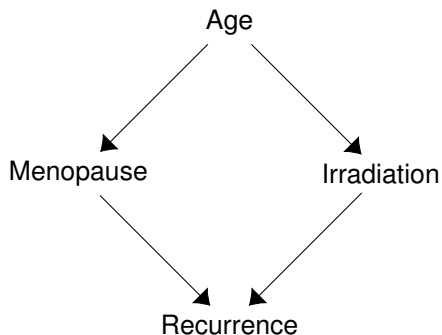
$$\text{RMSEA} = \sqrt{\frac{\chi^2/\text{df} - 1}{N - 1}}$$

### Properties of the RMSEA

- Under independence, the expected RMSEA is 0.
- Under dependence, the RMSEA converges to a constant positive value as  $N \rightarrow \infty$  (the p-value converges to 0).
- Higher RMSEA means stronger dependence.

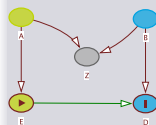


## Simulated Example: Breast Cancer Recurrence



Implications:

- Irradiation  $\perp\!\!\!\perp$  Menopause | Age
- Age  $\perp\!\!\!\perp$  Recurrence | Irradiation, Menopause



Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

## Testing the Assumed Model

Let's start with the first implied independence:

Irradiation  $\perp$  Menopause | Age

```
d <- read.table("https://git.io/JX4w5", header=TRUE)
dagitty::ciTest("Irradiation", "Menopause", "Age", data=d, type="cis.chisq")
```

##		rmsea	x2	df	p.value	rmsea	2.5%	rmsea	97.5%
##	Irrd _  _ Mnps   Age	0.029	2.7	4	0.61		0		0.31

There's no strong evidence against independence here.

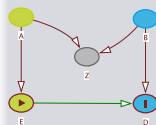
Now let's test the second implied independence:

Age  $\perp$  Recurrence | Irradiation, Menopause

```
dagitty::ciTest("Age", "Recurrence", c("Irradiation", "Menopause"),
  data=d, type="cis.chisq")
```

##		rmsea	x2	df	p.value	rmsea	2.5%	rmsea	97.5%
##	Age _  _ Rcrr   Irrd, Mnps	0.31	104	14	7.7e-16		0.16		0.48

The data strongly contradict this.



Why Test DAG Models?

Categorical Data

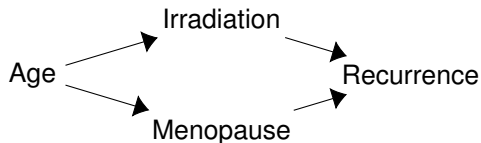
Continuous Data

Mixed Data



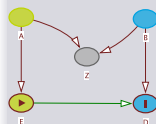
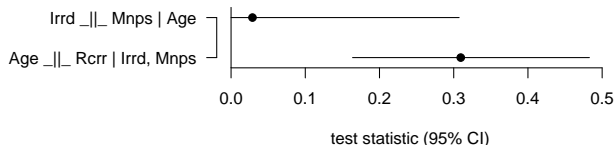
## Summary of Test Results

- We assumed the following DAG:



- We could not refute the first claim:  
 $\text{Irradiation} \perp\!\!\!\perp \text{Menopause} \mid \text{Age}$ .
- We refuted the second claim:  
 $\text{Age} \perp\!\!\!\perp \text{Recurrence} \mid \text{Irradiation, Menopause}$ .

```
r <- dagitty::localTests("Age->{Irradiation Menopause}->Recurrence",  
  d, type="cis.chisq")  
dagitty::plotLocalTestResults( r, pch=19, ylim=c(0.8, 2.2) )
```

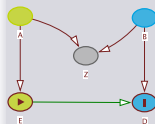
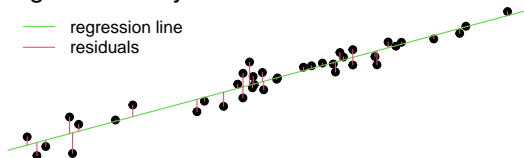


## Residual Correlation Tests for Continuous Data

For continuous data, stratification-based tests are not generally used. Instead, the residualization-based approach is more common:

- Regress  $X$  on  $\mathbf{Z}$  and get residuals  $r_X = X - E[X | \mathbf{Z}]$ .
- Regress  $Y$  on  $\mathbf{Z}$  and get residuals  $r_Y = Y - E[Y | \mathbf{Z}]$ .
- If  $r_X$  and  $r_Y$  correlate, then  $X \perp\!\!\!\perp Y | \mathbf{Z}$  does not hold.

This approach can be used with many types of regression (Gaussian process, Kernel ridge, ...). However, we will focus on **linear** regression only.



Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

# Example: Protein Signalling Network

## RESEARCH ARTICLE

# Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data

Karen Sachs<sup>1,\*</sup>, Omar Perez<sup>2,\*</sup>, Dana Pe'er<sup>3,\*</sup>, Douglas A. Lauffenburger<sup>1,†</sup>, Garry P. Nolan<sup>2,†</sup>

+ Author Affiliations

† To whom correspondence should be addressed. E-mail: [lauffen@mit.edu](mailto:lauffen@mit.edu) (D.A.L.); [gnolan@stanford.edu](mailto:gnolan@stanford.edu) (G.P.N.)

\* These authors contributed equally to this work.

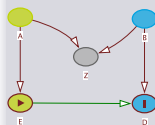
Science 22 Apr 2005:  
Vol. 308, Issue 5721, pp. 523-529  
DOI: 10.1126/science.1105809

Sachs *et al.* were interested in the interactions between a set of 11 proteins. They had measured the expression of these proteins on single cells using flow cytometry.

Their data is continuous, but flow cytometry data needs to be transformed for linearity to be reasonable.

## How to Test DAG Models

Johannes Textor



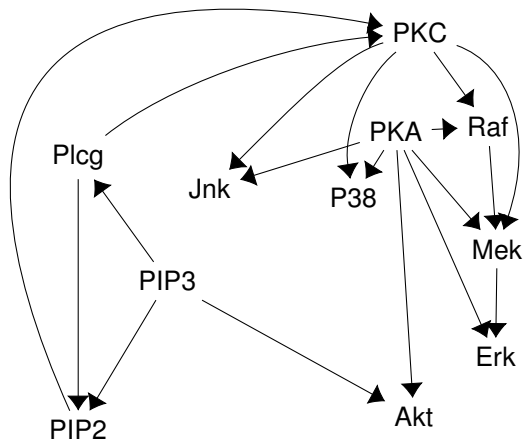
Why Test DAG Models?

Categorical Data

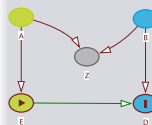
Continuous Data

Mixed Data

# The "Consensus" Protein Signalling Network



According to Sachs *et al.*, this network represents the expert consensus at their time of writing. Let's test this!



Why Test DAG Models?

Categorical Data

Continuous Data

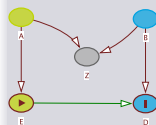
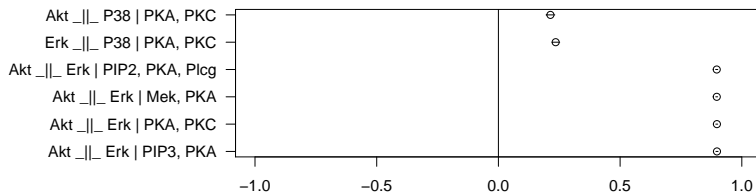
Mixed Data

# Testing the Consensus Network

```
# Download preprocessed protein signaling data
data <- read.csv( "https://git.io/JX4jh" )
# Residual correlation with bootstrapped CIs
r <- localTests( x=model, data=data, type="cis", R=50 )
r <- r[order( abs( r$estimate ) ),] # Worst tests last
tail( r )
```

##		estimate	std.error	2.5%	97.5%
##	Akt _  _ P38   PKA, PKC	0.21	0.0095	0.20	0.23
##	Erk _  _ P38   PKA, PKC	0.24	0.0076	0.22	0.25
##	Akt _  _ Erk   PIP2, PKA, Plcg	0.90	0.0025	0.89	0.90
##	Akt _  _ Erk   Mek, PKA	0.90	0.0025	0.89	0.90
##	Akt _  _ Erk   PKA, PKC	0.90	0.0021	0.89	0.90
##	Akt _  _ Erk   PIP3, PKA	0.90	0.0026	0.89	0.90

```
plotLocalTestResults( tail( r ), xlim=c(-1,1) )
```



Why Test DAG Models?

Categorical Data

Continuous Data

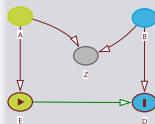
Mixed Data

## A Latent Variable Approach for Heterogeneous Datasets

We'll now discuss a “mixed” approach for datasets containing both discrete and continuous data, which is applicable if all of your variables are:

- 1 continuous; or
- 2 binary; or
- 3 ordinal (=categorical with an ordering).

Non-ordinal categorical can be incorporated via “dummy coding”, but this works well only if they have no parents in the DAG.



Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

## Can We “Unbin” Data?

Let's consider this example dataset on schooling performance:

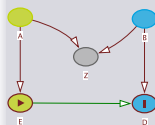
```
hsb2 <- read.table(header=T, sep="," ,  
"https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2-2.csv")  
head(hsb2)
```

##	id	female	race	ses	schtyp	prog	read	write	math	science	socst
## 1	70	0	4	1	1	1	57	52	41	47	57
## 2	121	1	4	2	1	3	68	59	53	63	61
## 3	86	0	4	3	1	1	44	33	54	58	31
## 4	141	0	4	3	1	3	63	44	47	53	56
## 5	172	0	4	2	1	2	47	52	57	53	61
## 6	113	0	4	2	1	2	44	52	51	63	61

```
table(hsb2$ses)
```

```
##  
## 1 2 3  
## 47 95 58
```

The “socioeconomic status” (SES) is an ordinal composite measure based on multiple indicators. We can think of this as an underlying continuous measure that has been “binned”. Is there a way to “un-bin” it?



Why Test DAG  
Models?

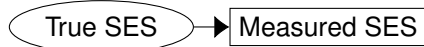
Categorical Data

Continuous Data

Mixed Data

## Threshold Functions

Let's assume that SES is standard normally distributed, and the binned values are obtained by applying cutoffs.



The most likely cutoffs are those where the Gaussian probability density mass is split into the same fractions as those we observe in the data.

```
y <- table(hsb2$ses)/nrow(hsb2)

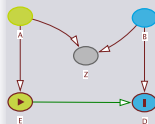
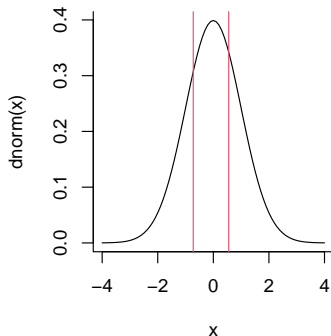
##
## 1 2 3
## 0.23 0.47 0.29

y <- cumsum( y )

## 1 2 3
## 0.23 0.71 1.00

y <- qnorm( y )

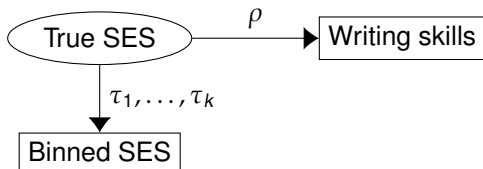
## 1 2 3
## -0.72 0.55 Inf
```





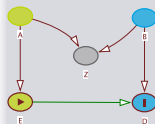
## Polyserial Correlations

A **polyserial correlation** is an estimated correlation between one continuous variable with another that's given as binned measurements.



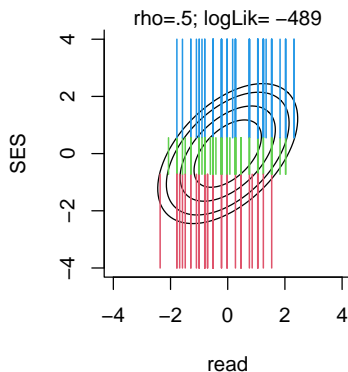
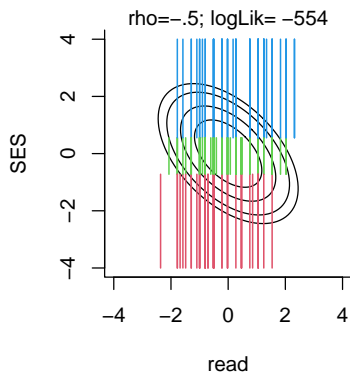
The polyserial correlation  $\rho$  can be estimated in two steps:

- 1 Infer thresholds for the discrete latent variable.
- 2 Optimize the value for the latent correlation to maximize the likelihood of the data conditional on the thresholds.



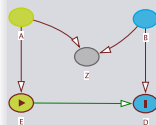
## Polyserial Correlations

In the optimization, we determine the latent correlation that maximizes the sum of the shown line integrals of the resulting 2D Gaussian distribution.



### How to Test DAG Models

Johannes Textor



Why Test DAG Models?

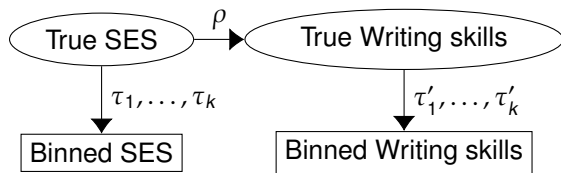
Categorical Data

Continuous Data

Mixed Data

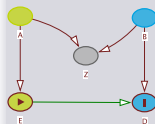
## Polychoric Correlations

A **polychoric correlation** is an estimated correlation between two latent variables that are both given as binned measurements.



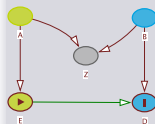
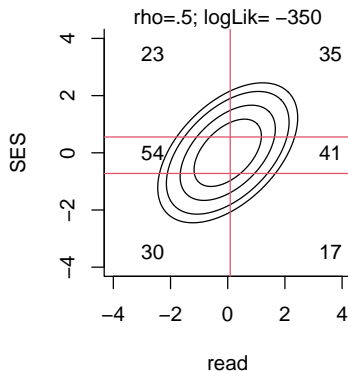
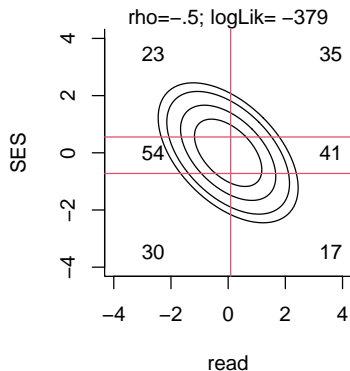
The polychoric correlation  $\rho$  can be estimated in two steps:

- 1 Infer thresholds for the discrete latent variables.
- 2 Choose the value for the latent correlation that maximizes the likelihood of the data given the correlation and the thresholds.



## Polychoric Correlations

In the optimization, we determine the latent correlation that maximizes the likelihood to obtain the shown number of observations in each rectangle.



Why Test DAG Models?

Categorical Data

Continuous Data

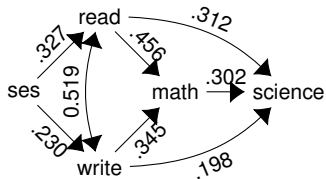
Mixed Data

## Testing and Fitting Models Using Polychoric Correlations

By appropriate pre-processing of our data, we can generate a correlation matrix in which the ordinal variables are represented by polychoric correlations.

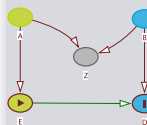
We can then apply standard SEM methodology to this matrix, since all SEM testing and fitting can be performed based on only the correlation matrix.

```
d <- hsb2[,c("ses", "read", "write",  
            "science", "math")]  
d$ses <- ordered(d$ses, levels=1:3)  
M <- lavaan::lavCor(d)  
g <- dagitty("dag{ses  
-> {read <-> write}  
-> {math->science}}")
```



```
localTests( g, sample.cov=M, sample.nobs=nrow(d) )
```

```
##  
## math _||_ ses | read, writ    estimate p.value    2.5% 97.5%  
## scnc _||_ ses | read, writ    0.11    0.13 -0.0330 0.24  
## scnc _||_ ses | read, writ    0.14    0.05 -0.0003 0.27
```



Why Test DAG Models?

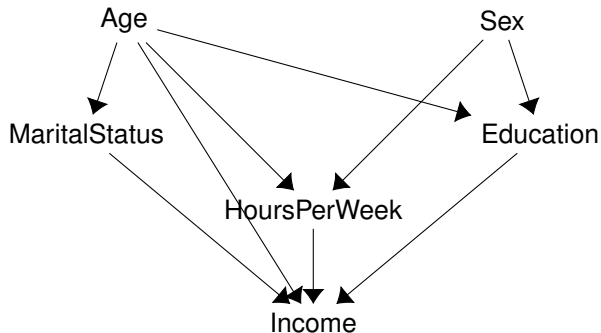
Categorical Data

Continuous Data

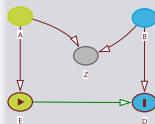
Mixed Data

## A More Complex Example

Let's consider a hypothetical (incorrect) causal diagram for a part of the widely known 1994 "adult census income" dataset:



Note the conspicuous absence of the gender pay gap in this model.



Why Test DAG Models?

Categorical Data

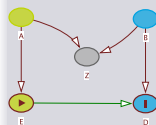
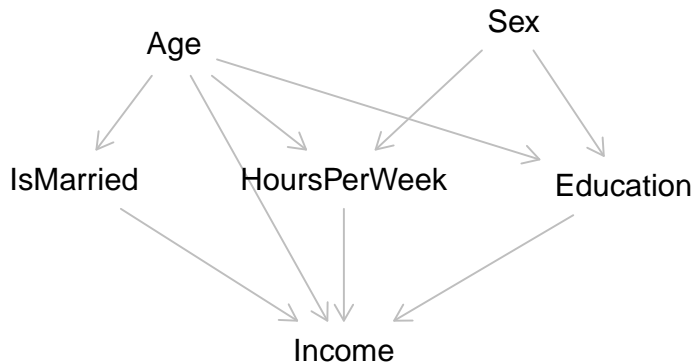
Continuous Data

Mixed Data

## Putting Everything Together

Step 1: We build our DAG.

```
g <- dagitty('Age [pos="-2.5,-2"] Education [pos="2,0.2"]  
HoursPerWeek [pos="-0.8,0.1"] Income [pos="-0.8,2.8"]  
IsMarried [pos="-3.5,0.1"] Sex [pos="0.9,-2.4"]  
Age -> { Education HoursPerWeek Income IsMarried }  
{ Education HoursPerWeek IsMarried } -> Income  
Sex -> { Education HoursPerWeek }')  
plot(g)
```



Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data

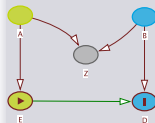
# Putting Everything Together

Step 2: We load and preprocess our data.

```
d <- read.csv("adult.csv")
d$Age <- ordered( d$Age, levels=c("<20", "20-34", "35-49", "50-65", ">65") )
d$Education <- ordered( d$Education, levels=c("Non-HS-Grad", "HS-grad",
"College-Associate", "Academic-Degree") )
d$Income <- ordered( d$Income, c("<=50K", ">50K") )
d$HoursPerWeek <- ordered( d$HoursPerWeek, c("<20", "20-39", "40", ">40") )
d$Sex <- as.numeric( ordered( d$Sex, c("Male", "Female") ) )
d$IsMarried <- as.integer( d$MaritalStatus == "Is-Married" )

d <- d[,c("Age", "Education", "Income", "HoursPerWeek", "Sex", "IsMarried")]
```

We need to make sure that our analysis is aware of the ordering between the categories (where meaningful).



Why Test DAG Models?

Categorical Data

Continuous Data

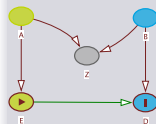
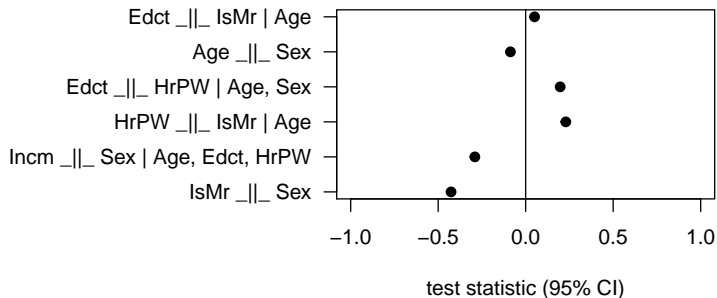
Mixed Data



## Putting Everything Together

Step 3: We test our DAG.

```
plotLocalTestResults(localTests(  
  g, sample.cov=M, sample.nobs=nrow(d) ), xlim=c(-1,1), pch=19 )
```



Why Test DAG Models?

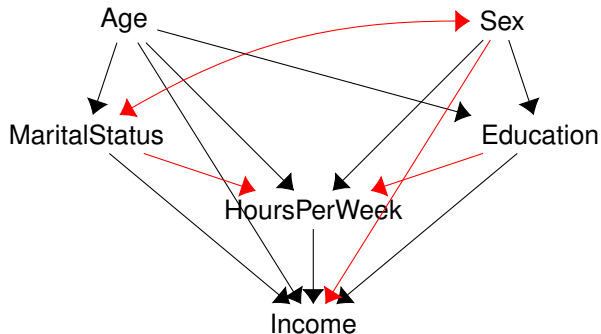
Categorical Data

Continuous Data

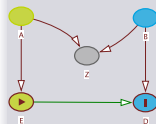
Mixed Data

## Updated Model

From the test results, we decide to add 4 edges that could be missing here.



Note that the bi-directed edge does not represent a causal link, but is just meant to “absorb” the sample correlation between Sex and Marital Status that should really not be there.



Why Test DAG Models?

Categorical Data

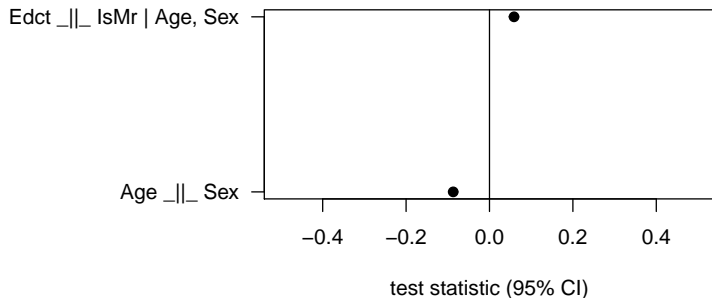
Continuous Data

Mixed Data

## Putting Everything Together

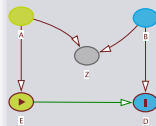
Step 2: We test our DAG (again).

```
plotLocalTestResults(localTests(  
  g, sample.cov=M, sample.nobs=nrow(d) ), xlim=c(-.5,.5), pch=19 )
```



### How to Test DAG Models

Johannes Textor



Why Test DAG Models?

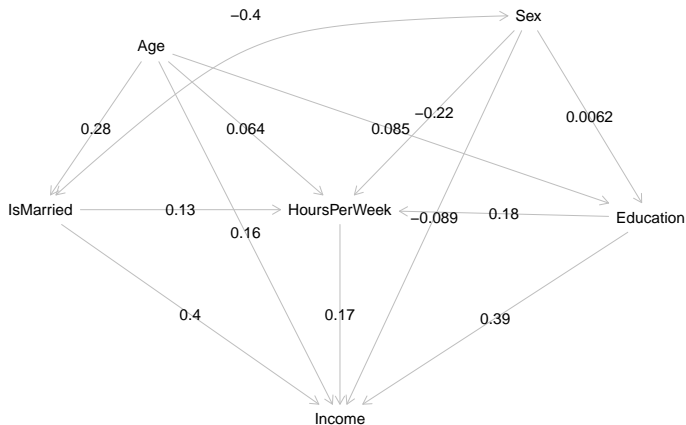
Categorical Data

Continuous Data

Mixed Data

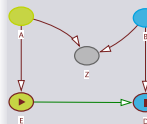
# Using lavaan to Estimate Path Coefficients

```
fit <- lavaan::sem( toString(g,"lavaan"),  
  sample.cov=M, sample.nobs=nrow(d) )  
fg <- lavaanToGraph(fit, digits=2)  
coordinates(fg) <- coordinates(g); plot(fg, show.coefficients=TRUE)
```



## How to Test DAG Models

Johannes Textor



Why Test DAG Models?


Categorical Data

Continuous Data


Mixed Data

## Summary

- 1 DAGs imply conditional independence constraints on compatible probability distributions.
- 2 The  $d$ -separation criterion allows to read off these constraints from the graphical model structure.
- 3 The dagitty package implements methodology for (human-focused) DAG model testing.

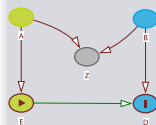
 Ankur Ankan, Inge Wortel, Johannes Textor:  
Testing Graphical Causal Models Using the R Package "dagitty".

*Current Protocols* 1(2), 2021. doi: 10.1002/cpz1.45

 Felix Thoemmes, Yves Rosseel, Johannes Textor:  
Local Fit Evaluation of Structural Equation Models Using Graphical Criteria.

*Psychological Methods* 23(1):27-41, 2018.

<https://psyarxiv.com/2kqxr/>



Why Test DAG Models?

Categorical Data

Continuous Data

Mixed Data