

MAT02010 - Tópicos Avançados em Estatística II

Diagramas causais: uma introdução

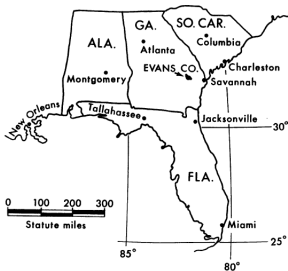
Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2019

Atividade 1

Doença coronariana e níveis de catecolaminas



- ▶ Com base em experimentos com animais, a função de emergência da medula adrenal foi estudada.
- ▶ O sistema nervoso simpático é ativado quando um indivíduo é desafiado.
- ▶ Via o hipotálamo, o estresse psicológico estimula a medula adrenal para liberar as duas catecolaminas, epinefrina (adrenalina) e norepinefrina (noradrenalina) na corrente sanguínea.
 - ▶ Esta reação rápida de defesa prepara o corpo para a batalha.
- ▶ No entanto, nos dias atuais o sistema simpático é desafiado por ameaças de natureza social ou mental.
 - ▶ Elevação da pressão arterial e da frequência cardíaca e da liberação de ácidos graxos livres e glicose na corrente sanguínea numa rotina mentalmente estressante, mas sedentária vai ser prejudicial para o organismo, principalmente para o sistema cardiovascular.

Doença coronariana e níveis de catecolaminas

- ▶ Com este estudo de caso teremos a oportunidade de comparar a incidência de doença coronariana (CHD) em dois grupos de exposição (níveis alto e baixo de catecolaminas).

P1: Esta relação é causal?

P2: Quais as suposições necessárias para o estabelecimento de uma relação causal?

P3: É necessário a análise dos dados para responder as perguntas acima?

Dados

- ▶ Os dados considerados neste estudo de caso são derivados de um estudo de coorte – *Evans County Heart Disease Study* – realizado entre 1960 e 1969.
- ▶ Estes dados dizem respeito a uma coorte de 609 indivíduos do sexo masculino, brancos, com idades entre 40 e 76 anos, livres de **doença coronariana** e residentes no distrito de Evans, Georgia, em 1960.
- ▶ Após sete anos, toda a coorte foi reexaminada e os novos casos de **doença coronariana** foram identificados.
- ▶ **Nível de catecolaminas** é a variável exposição de interesse.
- ▶ Todas as variáveis, com exceção de CHD, foram medidas na linha de base (1960).

Variáveis

<i>Nome da variável</i>	<i>Descrição da variável</i>	<i>Código da variável</i>
CHD	Ocorrência de doença coronariana	0 = não caso 1 = novo caso
CAT	Nível sérico de catecolaminas	0 = baixo 1 = alto
SMK	Tabagismo	0 = nunca fumou 1 = fumante
ECG	Alterações do eletrocardiograma	0 = ECG normal 1 = qualquer alteração
OCC	Tipo de ocupação	0 = não agricultor 1 = agricultor
MAR	Estado civil	0 = não casado 1 = casado

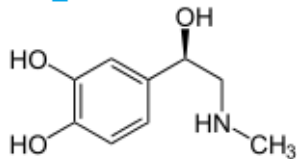
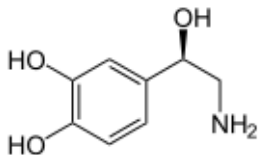
Variáveis

<i>Nome da variável</i>	<i>Descrição da variável</i>	<i>Código da variável</i>
AGE	Idade	Anos
CHL	Colesterol	mg/100 mL
SBP	Pressão arterial sistólica	mmHg
DBP	Pressão arterial diastólica	mmHg
QTI	Índice de Quételet	$100 \times [\text{peso (lb)}/\text{altura (in.)}^2]$
HEM	Hematócrito	Percentual
SES	Status socioeconômico	Índice de McGuire-White
PLS	Pulsção	Batidas/minuto

Dados (amostra grátis)

id	chd	cat	age	chl	smk	ecg	dbp	sbp	hpt	cc	ch
21	0	0	56	270	0	0	80	138	0	0	0
31	0	0	43	159	1	0	74	128	0	0	0
51	1	1	56	201	1	1	112	164	1	1	201
71	0	1	64	179	1	0	100	200	1	1	179
74	0	0	49	243	1	0	82	145	0	0	0
91	0	0	46	252	1	0	88	142	0	0	0

Conhecimento a priori



P: Faça uma discussão a respeito dos fatores psicossociais associados a liberação de catecolaminas. Quais os efeitos cardiovasculares são esperados destas catecolaminas?

Prática

- ▶ Construa o **diagrama causal** (DAG) para avaliar o efeito da exposição na resposta de interesse.
- ▶ Utilizando o critério *back-door* (**apresente os passos utilizados**) responda:
 1. Existe confundimento? O que isso implica em termos de suposições de identificação do efeito causal?
 2. Se existe confundimento, é possível controlar este confundimento?
 3. Ainda, se existe confundimento, quais são as variáveis do conjunto suficiente para controle/ajuste?



Relembando

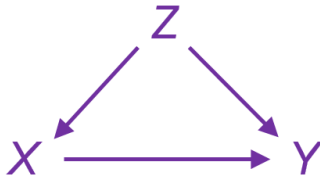
Como duas variáveis podem estar associadas?



- ▶ X e Y serão **associadas na população** se:
 - ▶ X causa Y.
 - ▶ Y causa X.
 - ▶ existe uma Z que é causa comum de X e Y.
- ▶ X e Y serão **associadas em subpopulações definadas por Z** se Z é um **efeito** de X e Y.

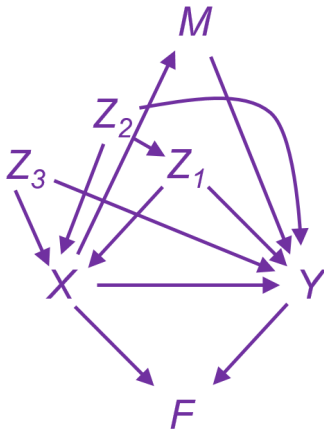
Como duas variáveis podem estar associadas?

- O que podemos concluir do diagrama abaixo?



Como duas variáveis podem estar associadas?

- E neste outro caso? O que podemos concluir?

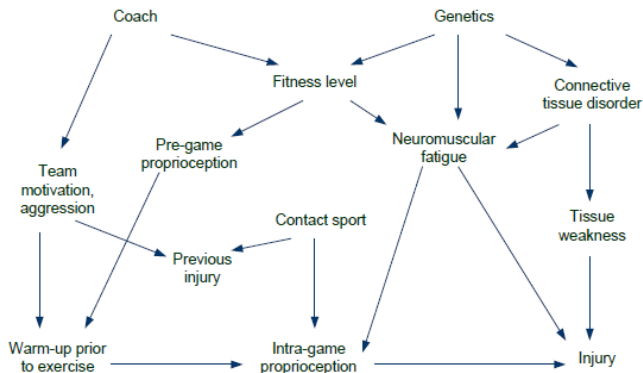


Grafos acíclicos dirigidos

- ▶ Os **grafos acíclicos dirigidos** (DAGs), ou **diagramas causais**, oferecem uma linguagem para especificação de relações (causais) entre as variáveis do quadro conceitual.
- ▶ **Critério back-door**: (1) verifica a existência de **viés de confusão**; (2) em caso afirmativo, verifica a existência de um conjunto de variáveis **suficiente para o controle** (ajuste) do viés de confusão.

Ferramentas computacionais para a construção de DAGs

Um exemplo



- A análise de DAGs pode ser tediosa na prática, e se presta bem à automatização por um programa de computador.

DAGitty

Welcome to DAGitty!

Launch



[Launch DAGitty
online in your
browser](#)

Download



[Download
DAGitty's source
for offline use](#)

Learn



[Learn more about
DAGs and
DAGitty](#)

Code



The R package
"dagitty" is
available on
[CRAN](#) or [github](#)

DAGitty

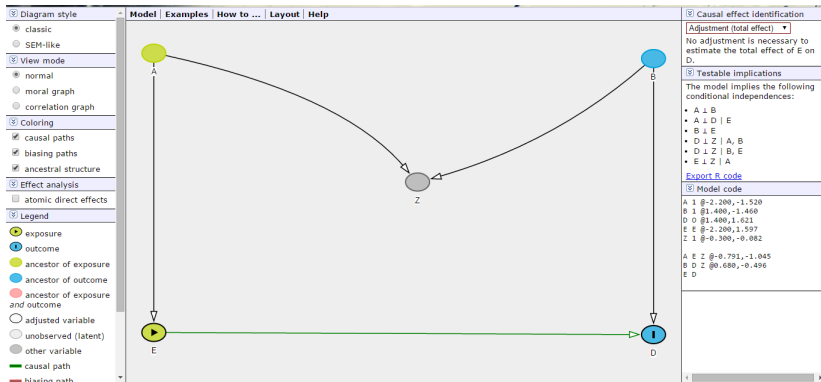
- ▶ O DAGitty é um ambiente baseado em navegador para criar, editar e analisar modelos causais (DAGs).
 - ▶ O foco está no uso de diagramas causais para minimizar o viés em estudos empíricos em epidemiologia e outras disciplinas.



- ▶ O DAGitty é desenvolvido e mantido por **Johannes Textor** (Tumor Immunology Lab and Institute for Computing and Information Sciences, Radboud University Nijmegen).
- ▶ <http://dagitty.net/>

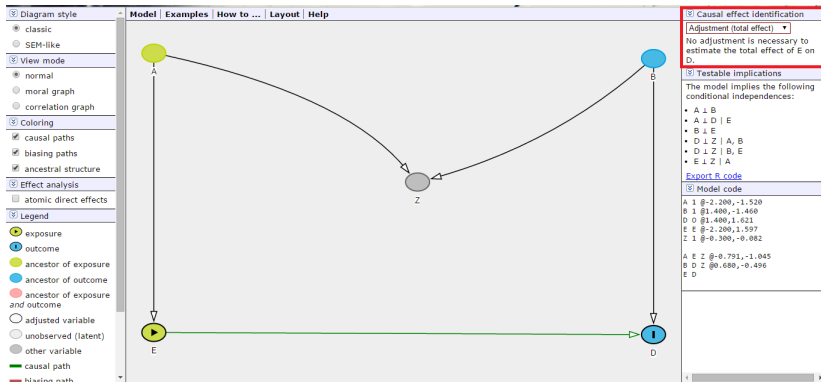
DAGitty

Uma visão geral



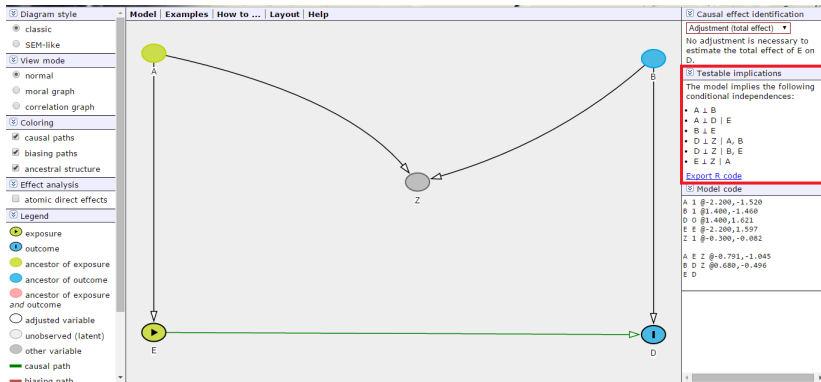
DAGitty

Uma visão geral



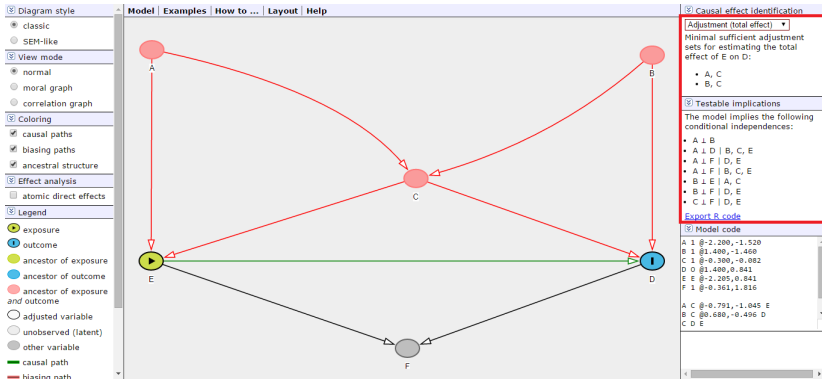
DAGitty

Uma visão geral



DAGitty

Uma visão geral



Criando um DAG no DAGitty

- ▶ No menu **Model**, clique em **New model**.
- ▶ O DAGitty irá solicitar o nome da variável de **exposição**, e logo em seguida o nome da variável de **desfecho**.
 - ▶ Estas variáveis serão criadas no grafo com o caminho $E \rightarrow D$.
- ▶ Para acrescentar uma nova variável dê um **duplo-clique** na área do grafo e dê um nome para esta variável.
- ▶ Para especificar uma relação entre duas variáveis, dê um duplo-clique na **variável de origem** e um duplo-clique na **variável de destino**.
 - ▶ O mesmo procedimento serve para remover uma relação entre duas variáveis já existente.
- ▶ Para remover uma variável do grafo, clique sobre a variável pressionando a **tecla D**.
- ▶ Para renomear uma variável, clique sobre a variável pressionando a **tecla R**.
- ▶ Na dúvida, consulte os menus **How to ...** e **Help**.

Criando um DAG no DAGitty

- ▶ No menu **Model** é possível:
 - ▶ exportar o grafo
 - ▶ publicar o grafo
- ▶ Uma vez publicado, o grafo pode ser:
 - ▶ modificado
 - ▶ apagado

Atividade 2

Atividade 2

- ▶ Utilize o DAGitty para construir o DAG do “**exemplo das Catecolaminas**”.
- ▶ Verifique se as conclusões obtidas concordam com as que você obteve com o critério back-door “feito a mão”.

DAGitty encontra o R

```
# Instala o pacote dagitty
# install.packages("dagitty")

# Carrega o pacote dagitty
library(dagitty)

# Carrega o DAG a partir do DAGitty
dag1 <- downloadGraph(x = "dagitty.net/mBYpOXW")
```

DAGitty encontra o R

```
plot(dag1)
```



DAGitty encontra o R

```
dag1.bd <- backDoorGraph(dag1)  
plot(dag1.bd)
```



DAGitty encontra o R

```
print(adjustmentSets(dag1,  
                     effect = "total"))
```

```
## { ESTRESSE, IDADE, TABAGISMO }
```

```
# impliedConditionalIndependencies(dag1,  
#                                max.results = 2)
```

ggdag: dagitty encontra o ggplot2

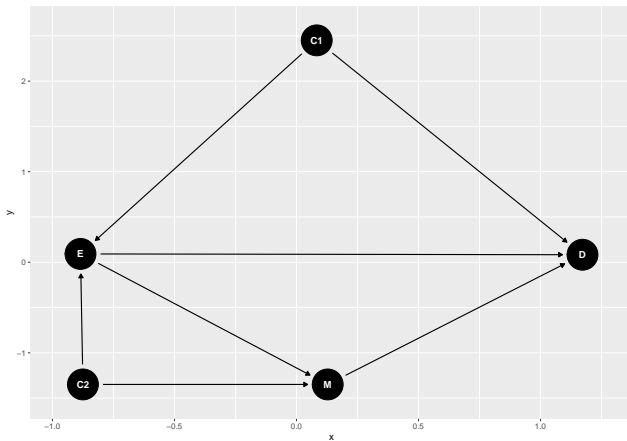
```
# Instala o pacote ggdag
# install.packages("ggdag")

# Carrega o pacote ggdag
library(ggdag)

# Carrega o DAG a partir do DAGitty
dag2 <- downloadGraph(x = "dagitty.net/mQLajCg")
```

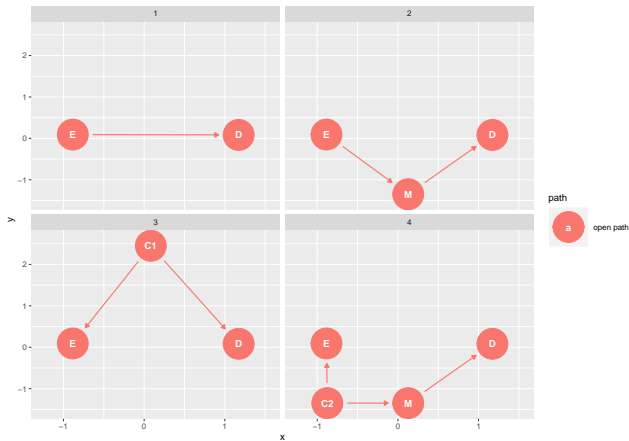

ggdag: dagitty encontra o ggplot2

```
ggdag(dag2)
```



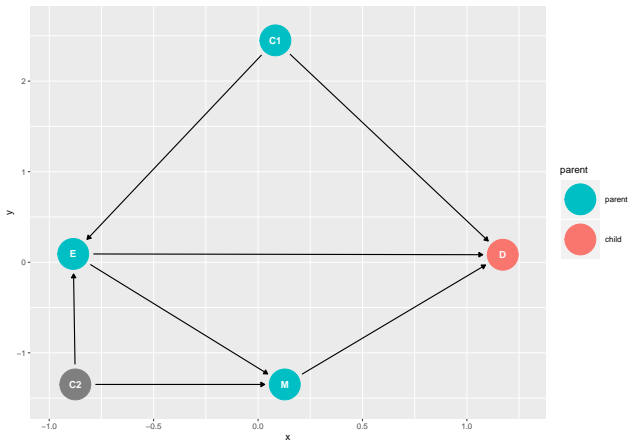
ggdag: dagitty encontra o ggplot2

```
ggdag_paths(dag2)
```



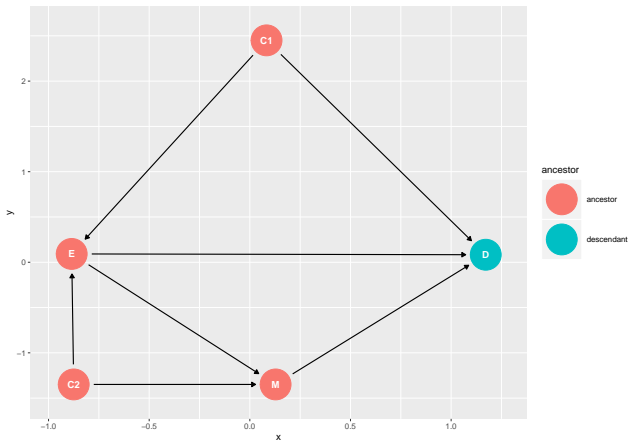
ggdag: dagitty encontra o ggplot2

```
ggdag_parents(dag2, "D")
```



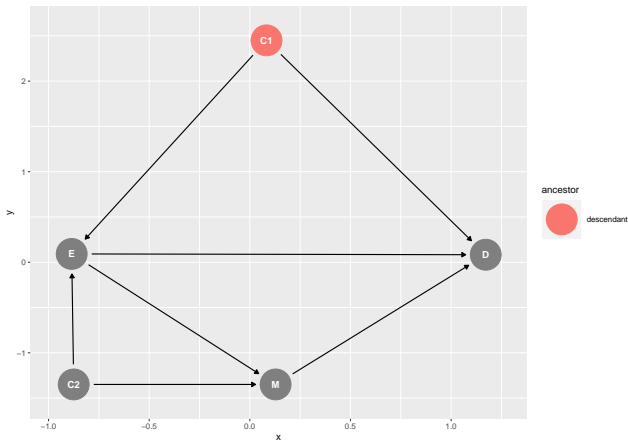
ggdag: dagitty encontra o ggplot2

```
ggdag_ancestors(dag2, "D")
```



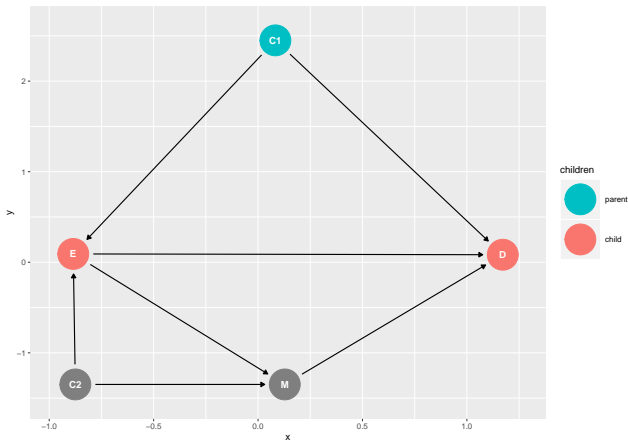
ggdag: dagitty encontra o ggplot2

```
ggdag_ancestors(dag2, "C1")
```



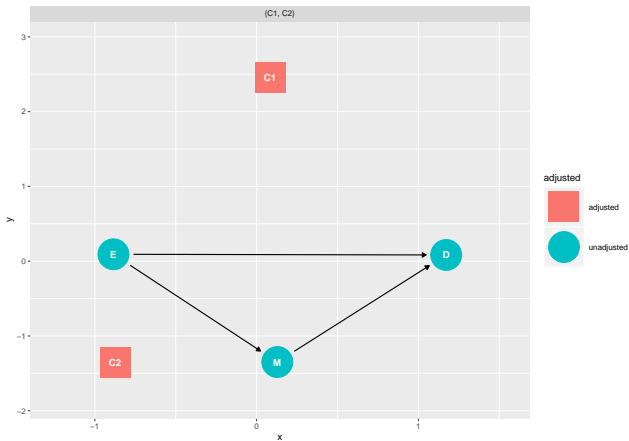
ggdag: dagitty encontra o ggplot2

```
ggdag_children(dag2, "C1")
```



ggdag: dagitty encontra o ggplot2

```
ggdag_adjustment_set(dag2)
```



Atividade 3

DAGitty + ggdag

- ▶ Pense em uma **questão causal** que você tem interesse em estudar.
- ▶ Utilize um ou mais recursos computacionais para construir um DAG (diagrama causal) do modelo causal relacionado à sua questão de pesquisa.

Avisos

Avisos

- ▶ **Próxima semana (13/11): métodos de pareamento/escores de propensão** para estimação de efeitos causais.
- ▶ **Para casa:** Ler artigos de Miguel Hernán.

Por hoje é só!

