MAE 5776

ANÁLISE MULTIVARIADA

Júlia M Pavan Soler

pavan@ime.usp.br

IME - 1° Sem/2020

Análise Multivariada

$$Y_{n\times p} = \left(Y_{ij}\right) \in \Re^{n\times p}$$

Já vimos ☺

- Estatísticas Descritivas Multivariadas:
- Distribuição N_p, Distribuições Amostrais, Regiões de Confiança, MANOVA
- Análises Multivariadas Clássicas (n>p, *iid*): CP, AF, CoP, AC, AD, ACC, PLS
- Análises Multivariadas Esparsas (n<<p, *iid*): CP, AD, ACC

Lista 4

- **"Componentes Principais" em Observações Dependentes (dados de famílias)**
- "Componentes Principais" em Dados Heterogêneos

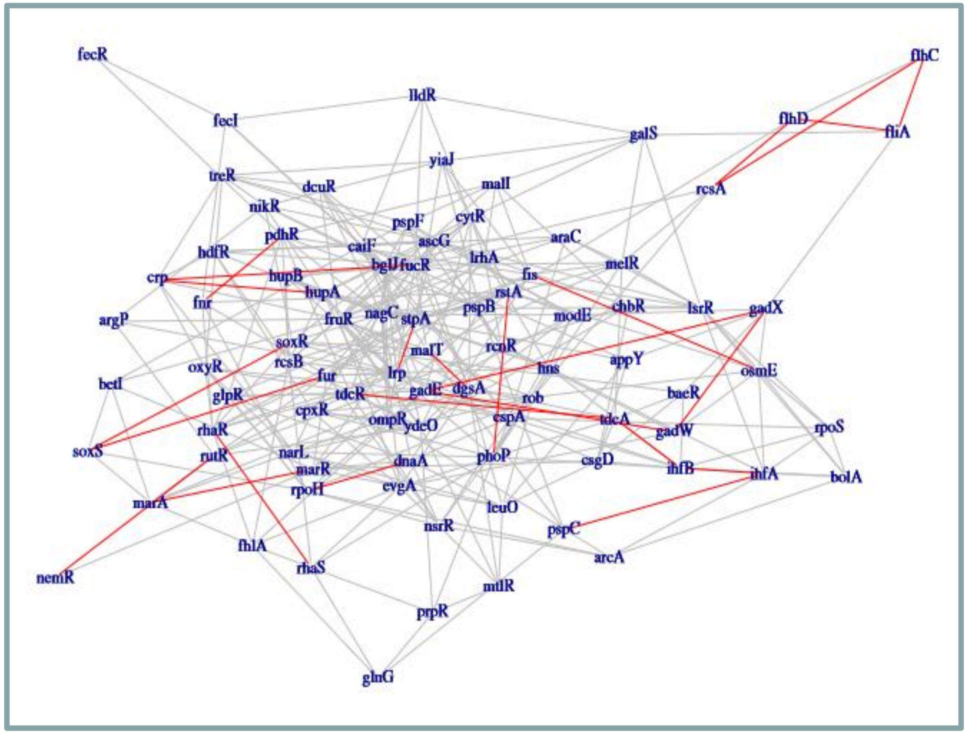
$\Rightarrow$ Aprendizado de Estruturas

- Modelos de Grafos Probabilísticos
- Modelos de Equações Estruturais
- Fatoração da Distribuição Conjunta

Aprendizado de Estruturas – Inferência Causal

$$Y_{n\times p} = \left(Y_{ij}\right) \in \Re^{n\times p}$$

Aprender (Estimar) a Estrutura de Dependência entre Variáveis $\Re^{pxp}$



Grafo Gaussiano de Independência Condicional
(p=87 variáveis em E. *coli.;* Drton and Maathuis, 2016)

Dados multivariados com observações independentes (n>p; p↑)

$$Y_i \in \Re^p \overset{iid}{\sim} \left(\mu;\Sigma\right)$$

Σ diagonal (variáveis independentes)	⇔	Σ Não Estruturada

Aprender (estimar) a estrutura de dependência entre as variáveis:

- Modelos de GRAFOS +
- Modelos de Equações Estruturais +
- Propriedades de Markov

Aprendizado de Estruturas
Correlação x Causa

Relação entre Variáveis: Correlação (é simétrica) *versus* Causa (é assimétrica)

$X \;—\; Y$

$X \rightarrow Y$
$X \leftarrow Y$

- "Correlação" é uma relação não-direcionada entre variáveis
  "Causa" é uma relação direcionada (uma variável é causa ou efeito de outra)

- Inferências "Causais" são obtidas de Ensaios Clínicos Controlados e Aleatorizados que são padrão ouro em Planejamento de Experimentos.
  Desafio: extrair inferência causal a partir de Estudos Observacionais

No Aprendizado de Estruturas é necessário estabelecer suposições que possam capturar o padrão de dependências entre as variáveis.

Sistemas de 2 variáveis (X e Y) não podem inferir estruturas causais.
Ao incluir uma terceira variável (X, Y e Z) é possível inferir estruturas causais, sob certas suposições (Haussman, 1984; Papineau, 1985)

Aprendizado de Estruturas – Inferência Causal

Como os sistemas respondem a intervenções ?

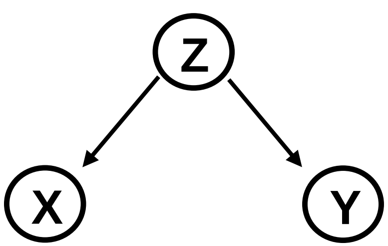
Ensaios Controlados Aleatorizados (padrão ouro de inferência causal) X Estudos observacionais (são mais fáceis de conduzir MAS podem conter muitos "ruídos, vícios")

Avaliar padrões de dependência entre trios de variáveis:

(X,Y) d-conectados dado Z: critério de orientação de arestas em um grafo

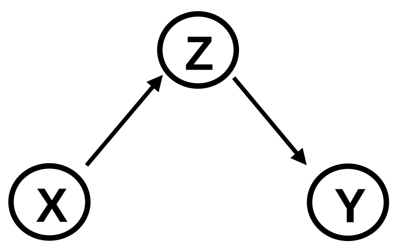
Caso 1

Z é causa comum
Z é confundidor



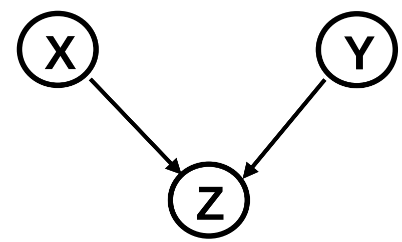
Caso 2

Z é efeito intermediário
X tem efeito indireto sobre Y



Caso 3

Z é efeito comum
V-Estrutura: colisão não conectada



$(X \not\perp Y)$ associados

$(X \perp Y \mid Z)$ independência condicional

Modelos Probabilísticos Equivalentes Grafos Diferentes

$(X \perp Y)$ independência

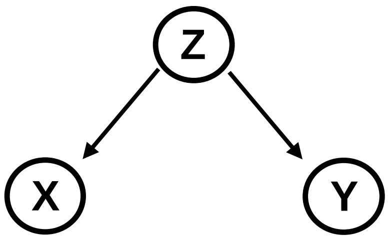
$(X \not\perp Y \mid Z)$ dependência condicional

Aprendizado de Estruturas

Alguns resultados!

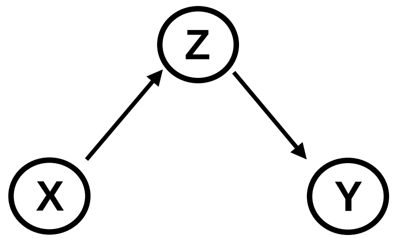
Caso 1

Z é causa comum
Z é confundidor



Caso 2

Z é efeito intermediário
X tem efeito indireto sobre Y



Análise de Independência condicional

$$\left(X \perp Y \mid Z\right)$$

Problemas: mesma relação de independência (mesma distribuição conjunta) MAS sob diferentes Grafos.

$$P(X \mid Z)\, P(Y \mid Z)\, P(Z) \qquad P(X)\, P(Z \mid X)\, P(Y \mid Z)$$

$$P(X, Y, Z)$$

$$P(X \mid Z)\, P(Y \mid Z)\, P(Z) = P(X, Z) P(Y \mid Z) = P(X) P(Z \mid X) P(Y \mid Z)$$

Como quebrar a equivalência de distribuições (geradas de “grafos” diferentes) ?

- Condições para que o modelo probabilístico determine unicamente o grafo:
  $\Rightarrow$ Teorema da Identificabilidade (ou da descoberta)

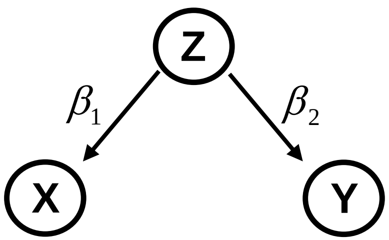
Aprendizado de Estruturas

Alguns resultados!

Correspondência entre Grafos e Modelos de Equações Estruturais (SEM)

Caso 1

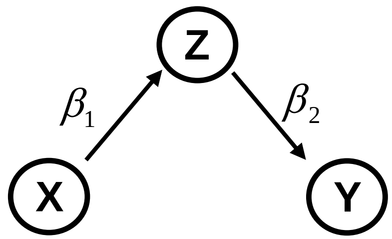
Z é causa comum
Z é confundidor



$$P(X \mid Z)\,P(Y \mid Z)\,P(Z)$$

Caso 2

Z é efeito intermediário
X tem efeito indireto sobre Y



$$P(X)\,P(Z \mid X)\,P(Y \mid Z)$$

Correspondência entre "Grafos" e "SEM"

Cada efeito é uma função arbitrária (linear ou não) de sua causa direta e de um termo de erro. A função não é tão importante quanto a independência dos termos de erro (Spirtes,1994)

Considere SEM lineares e com erros independentes:

Caso 1:

$$Y = \beta_2 Z + e_{Y|Z}$$

$$X = \beta_1 Z + e_{X|Z}$$

$$(X \perp Y \mid Z) \Leftrightarrow e_{Y|Z} \perp e_{X|Z}$$

Caso 2:

$$Y = \beta_2 Z + e_{Y|Z}$$

$$Z = \beta_1 X + e_{Z|X}$$

$$(X \perp Y \mid Z) \Leftrightarrow e_{Y|Z} \perp e_{Z|X}$$

Sob termos de erros independentes, tanto a representação SEM como a representação por Grafos permitem identificar a "independência condicional" entre as variáveis.

Aprendizado de Estruturas

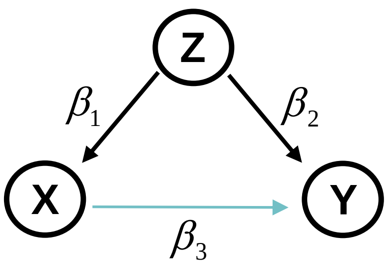
Alguns resultados!

Na correspondência entre Grafos e Modelos de Equações Estruturais é importante que o modelo probabilístico seja Fiel ao Grafo.

$$\left(X \perp Y \mid Z\right)$$

Caso 1

Z é causa comum (confundidor)



Efeito Z $\rightarrow$ Y = $\boxed{\text{Efeito direto + Efeito Indireto}}$

$$\left.\begin{aligned} Y &= \beta_2 Z + \beta_3 X + e_{Y|Z,X} \\ X &= \beta_1 Z + e_{X|Z} \end{aligned}\right\}$$

$$\begin{aligned} Y &= \beta_2 Z + \beta_3(\beta_1 Z + e_{X|Z}) + e_{Y|Z,X} \\ &= \boxed{(\beta_2 + \beta_1\beta_3)} Z + \beta_3 e_{X|Z} + e_{Y|Z,X} \end{aligned}$$

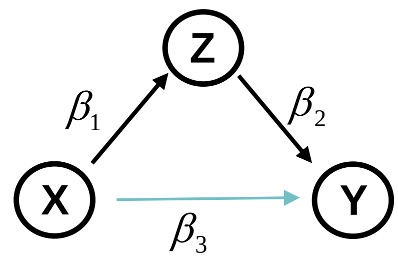
$\boxed{\beta_2 = -\beta_1\beta_3} \Rightarrow$ Efeito total nulo

O modelo probabilístico é infiel ao grafo

Caso 2

Z é efeito intermediário

X tem efeito indireto sobre Y



Efeito X $\rightarrow$ Y = Efeito direto + Efeito Indireto

$$\left.\begin{aligned} Y &= \beta_2 Z + \beta_3 X + e_{Y|Z,X} \\ Z &= \beta_1 X + e_{Z|X} \end{aligned}\right\}$$

$$\begin{aligned} Y &= \beta_2(\beta_1 X + e_{Z|X}) + \beta_3 X + e_{Y|Z,X} \\ &= (\beta_3 + \beta_1\beta_2) X + \beta_2 e_{Z|X} + e_{Y|Z,X} \end{aligned}$$

$\beta_3 = -\beta_1\beta_2 \Rightarrow$ Efeito total nulo

O modelo probabilístico é infiel ao grafo

Aprendizado de Estruturas – Inferência Causal

Componentes da Inferência e do Aprendizado de Estruturas envolvendo "p" variáveis:

1. Fatoração da Distribuição de probabilidades conjunta (Lauritzen, 1990, 1996)
   Independência condicional e
   Propriedades de Markov (M-Pares de Variáveis, M-Local, M-Global)

2. Modelos de Grafos Probabilísticos (Pear, 1989; Verma e Pear, 1990)
   d-separação (V-estrutura de colisão não-conectada: $X \rightarrow Y \leftarrow Z$)
   Fidelidade

3. Modelos de Equações Estruturais (SEM) (Boolen,1989)
   Equações de Mensuração e Equações das variáveis latentes
   Linearidade
   Erros independentes

Grafos: Diagramas de Representação “Causal”

Grafo é um conjunto de vértices (V) e arestas (E): $G=(V,E)$; $E \subset VxV$

V: variáveis (biométricas, sobrevivência, espectros de imagens, genótipos, etc)

E: arestas conectando V, representam “dependências” (não independência condicional) entre variáveis

- Grafo Não Dirigido (UDG): $X — Y — Z$

  Definem Vizinhanças: $nb_G(v)=\{w \in V; \{w,v\} \in E\}$

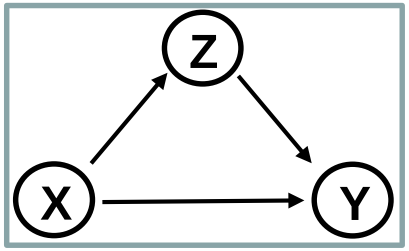
- Grafo Dirigido (DAG, acíclico): $X \rightarrow Y \rightarrow Z$

  Definem estruturas de Pais e Descendentes: $pa_G(v)=\{w \in V; \{w,v\} \in E\}$

  $de_G(v)=\{w \in V; w=v \text{ ou } v\rightarrow ... \rightarrow w \text{ em } G\}$

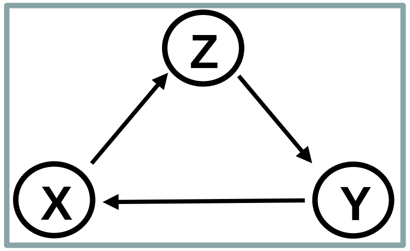
Grafos: Diagramas de Representação “Causal”

- Grafos dirigidos: G=(V,E); E $\subset$ VxV



- DAGs: Grafos Acíclicos Dirigidos

  SEM recursivos: com termos de erros não correlacionados e sem relacionamentos cíclicos
  Matriz de Adjacência “B” é triangular inferior (arestas conectando variáveis/Coluna causando Linha)



- DCGs: Grafos Cíclicos Dirigidos

  SEM não recursivos: matriz “B” geral mas sem elementos na diagonal (não ocorre self-loops)

Termos de “erro” e efeitos não são indicados no grafo (como acontece na representação de SEM).

Aprendizado de Estruturas – Inferência Causal

Modelos de Equações Estruturais (SEM) - Equações de Mensuração

Variáveis (vértices)

B: conecta os vértices

Covariáveis

Erro

Premissas clássicas

$$Y_{i\,p\times1} = B_{p\times p}Y_{i\,p\times1} + X_{ip\times q}\beta_{q\times1} + e_{i\,p\times1};\quad e_i \overset{iid}{\sim} N_p\left(0;\boxed{\Sigma = I_p\sigma^2}\right)$$

$B_{p\times p}$: Matriz de Adjacência (0 e 1's; B_{ii}=0)

$|I-B| = 1$ para DAGs

Modelo reduzido

$$Y_i = \left(I_p - B\right)^{-1} X_i\beta + \boxed{\left(I_p - B\right)^{-1} e_i}$$

Mistura de erros independentes

Matriz de covariância imposta pelo SEM: $Cov\left(Y_i\right) = \Sigma_{YY} = (I-B)^{-1}\,\Sigma\ \,(I-B)^{-1\prime}$

Objetivo da inferência: Estimar B e β

Inferência: Algoritmo SML (Cai et al., 2013)

$$Y_i \overset{iid}{\sim} N_p\left(\left[I_p - B_{p\times p}\right]^{-1} X_i\beta\,;\sigma^2\left[I_p - B_{p\times p}\right]^{-1}\left[I_p - B_{p\times p})\right]^{-1\prime}\right)$$

Aprendizado de Estruturas – Inferência Causal

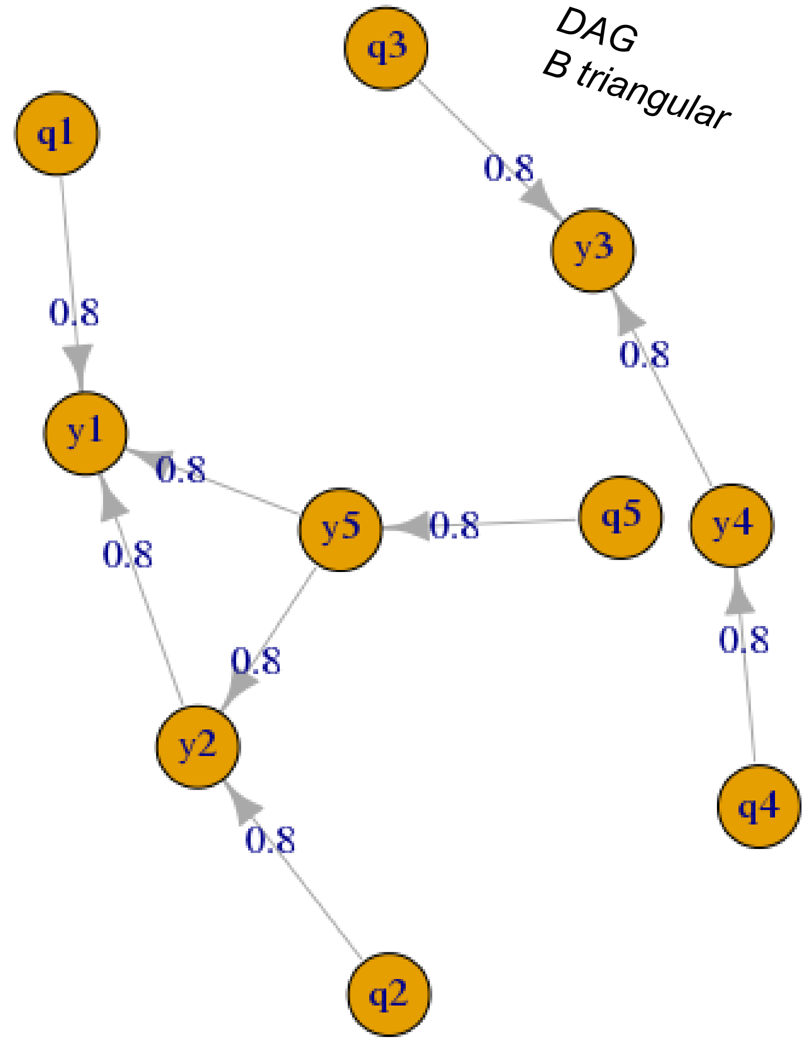
p= 5 variáveis “y”
B matriz de Adjacência: coluna causa linha
X matriz de covariáveis: cada variável “y” é afetada por um único Fator “q”

B =

	y1	y2	y3	y4	y5
y1	0.0	0.8	0.0	0.0	0.8
y2	0.0	0.0	0.0	0.0	0.8
y3	0.0	0.0	0.0	0.8	0.0
y4	0.0	0.0	0.0	0.0	0.0
y5	0.0	0.0	0.0	0.0	0.0

X =

	q1	q2	q3	q4	q5
y1	0.8	0.0	0.0	0.0	0.0
y2	0.0	0.8	0.0	0.0	0.0
y3	0.0	0.0	0.8	0.0	0.0
y4	0.0	0.0	0.0	0.8	0.0
y5	0.0	0.0	0.0	0.0	0.8



Aprendizado de Estruturas – Inferência Causal

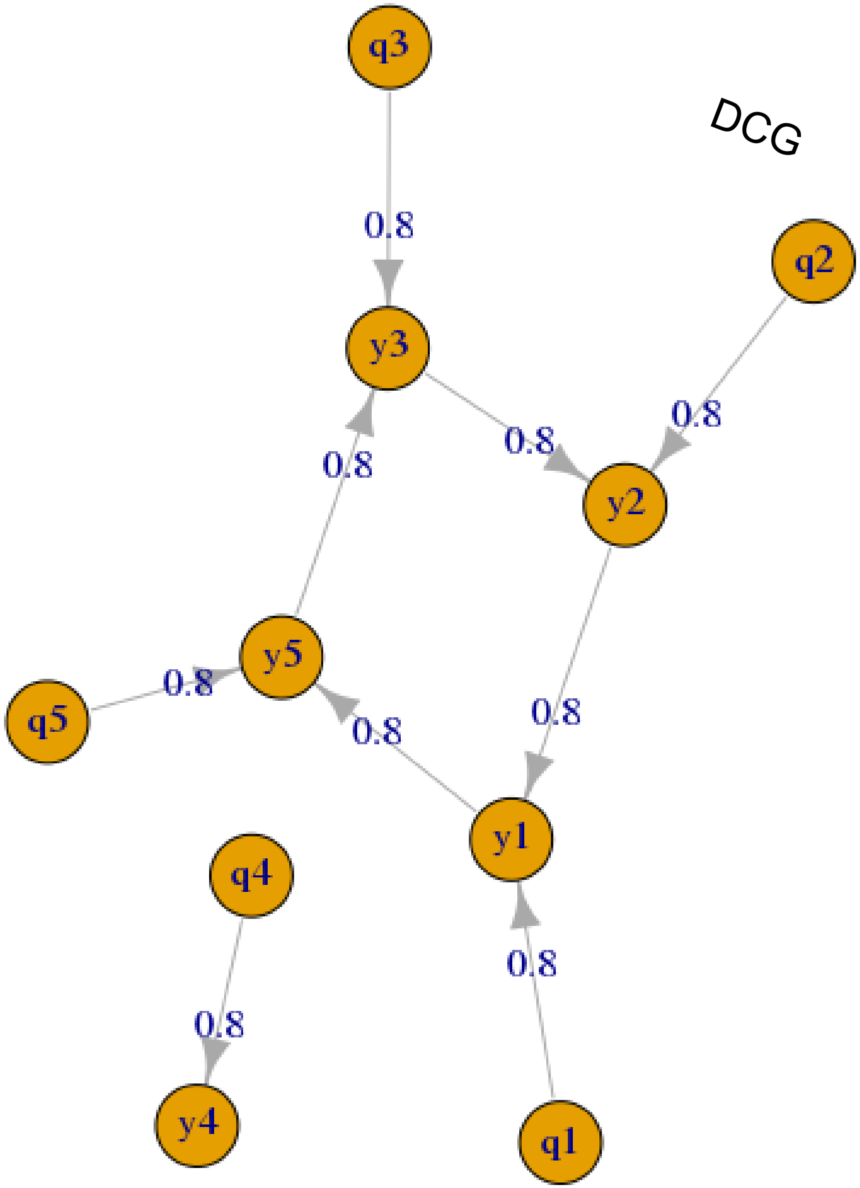
p= 5 variáveis
B matriz de Adjacência: coluna causa linha
X matriz de covariáveis: cada variável “y” é afetada por um único Fator “q”

B =

	y1	y2	y3	y4	y5
y1	0.0	0.8	0.0	0.0	0.0
y2	0.0	0.0	0.8	0.0	0.0
y3	0.0	0.0	0.0	0.0	0.8
y4	0.0	0.0	0.0	0.0	0.0
y5	0.8	0.0	0.0	0.0	0.0

X =

	q1	q2	q3	q4	q5
y1	0.8	0.0	0.0	0.0	0.0
y2	0.0	0.8	0.0	0.0	0.0
y3	0.0	0.0	0.8	0.0	0.0
y4	0.0	0.0	0.0	0.8	0.0
y5	0.0	0.0	0.0	0.0	0.8



Aprendizado de Estruturas – Inferência Causal

Os **Algoritmos de Aprendizado de Estruturas** são principalmente baseados em Testes de Independência Condicional.

- **Passo 1.** Grafo Completo

- **Passo 2.** Aprendizado do Esqueleto (**UDG** - relações simétricas) Eliminar arestas

Independência condicional de **Pares de variáveis dado as restantes**

- **Passo 3.** Aprendizado do Grafo Direcionado (**DAG** – relações de causa/efeito)

Orientar as arestas: encontrar V-estruturas $X \rightarrow Z \leftarrow Y$

Realizar testes de Independência Condicional de Pares de variáveis dado todo possível subconjunto condicionante.

3.1. X e Y não conectados: $(X \perp Y)$
3.2. Pesquisar por Z, tal que $(X \not\perp Y \mid Z)$
3.3. Se existir Z $\Rightarrow$ $\boxed{X \longrightarrow Z \longleftarrow Y}$

Algoritmo IC
Algoritmo PC

Implementar um teste de independência condicional
Oráculo da d-separação

Independência Condicional – Coeficiente de Correlação Parcial

Passo 2. Aprendizado do Grafo não dirigido (***UDG***) é baseado em **Testes de Correlação Parcial Nula**

$$S = V \setminus \{X, Y\}, \qquad X \perp Y \mid S \quad \Leftrightarrow \quad \rho(X, Y \mid S) = 0$$

Modelo Normal Multivariado: $V_i \overset{iid}{\sim} N_p\left(\mu_{p\times 1}; \Sigma_{p\times p}\right) \quad \Rightarrow \quad \Sigma_{p\times p} = \begin{pmatrix} \sigma_{XX} & \sigma_{XY} & \Sigma_{XS} \\ \sigma_{YX} & \sigma_{YY} & \Sigma_{YS} \\ \Sigma_{SX} & \Sigma_{SY} & \Sigma_{SS} \end{pmatrix}$

Distribuição Condicional:

$$\Sigma_{XY.S} = \begin{pmatrix} \sigma_{XX.S} & \sigma_{XY.S} \\ \sigma_{YX.S} & \sigma_{YY.S} \end{pmatrix}$$
$$= \begin{pmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{YX} & \sigma_{YY} \end{pmatrix} - \begin{pmatrix} \Sigma_{XS} \\ \Sigma_{YS} \end{pmatrix} \Sigma_{SS}^{-1} \begin{pmatrix} \Sigma_{SX} & \Sigma_{SY} \end{pmatrix}$$

$$\rho(X, Y \mid S) = \frac{\sigma_{XY.S}}{\sqrt{\sigma_{XX.S}\sigma_{YY.S}}} = \frac{-k_{XY}}{\sqrt{k_{XX}k_{YY}}}$$

Matriz de Precisão:

$$\Sigma^{-1} = \left(k_{XY}\right)$$

Testes de Independência Condicional

Passo 2. Aprendizado do Grafo não dirigido (***UDG***)

Testes da Correlação Parcial Nula

Alternativa 1: Aprendizado de um UDG via **Matriz de Precisão**

$$V \overset{iid}{\sim} N_p\left(\mu_{p\times 1};\Sigma_{p\times p}\right); \quad \boxed{\Sigma^{-1} = \left(k_{XY}\right)} \quad \rho\left(X,Y \mid S\right) = \frac{-k_{XY}}{\sqrt{k_{XX}k_{YY}}}$$

Teste da Correlação Parcial Nula via a Estatística *z* de Fisher

$$\begin{cases} H_0 : \rho_{XY.Z} = 0 \\ H_1 : \rho_{XY.Z} \neq 0 \end{cases} \qquad z_{XY.S} = \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_{XY.S}}{1-\hat{\rho}_{XY.S}}\right); \quad \sqrt{n-|S|-3}\;\left|z_{XY.S}\right| \overset{H_0}{\sim} N(0;1)$$

Testes de Independência Condicional

Alternativa 2: Aprendizado de um **UDG** via Ajuste de Modelos Univariados

Modelos Univariados:

- $Y_{n\times 1} = \mu_{Y|S} + \beta_S^Y{}'S + e_{Y|S}; \quad e_{Y|S} \sim N\left(0; \sigma^2_{Y|S}\right)$
- $X_{n\times 1} = \mu_{X|S} + \beta_S^X{}'S + e_{X|S}; \quad e_{X|S} \sim N\left(0; \sigma^2_{X|S}\right)$
- $Y_{n\times 1} = \mu_{Y|X,S} + \beta_{X|S}^Y{}'X + \beta_{S|X}^Y{}'S + e_{Y|X,S}; \quad e_{Y|X,S} \sim N\left(0; \sigma^2_{Y|X,S}\right)$

$$\boxed{\begin{aligned}\rho(X,Y\mid S) &= \frac{\sigma_{XY.S}}{\sqrt{\sigma_{XX.S}\sigma_{YY.S}}} = \frac{-k_{XY}}{\sqrt{k_{XX}k_{YY}}} \\ &= \frac{Cov\left(e_{X|S}; e_{Y|S}\right)}{\sigma_{X|S}\sigma_{Y|S}} = \beta_{X|S}^Y \frac{\sigma_{X|S}}{\sigma_{Y|S}}\end{aligned}}$$

$$\begin{cases} H_0 : \rho_{XY.Z} = 0 \Leftrightarrow \beta_{X|S}^Y = 0 \\ H_1 : \rho_{XY.Z} \neq 0 \end{cases}$$

Método das Vizinhanças(Meinshausen and Buhlmann, 2006)

Encontrar a vizinhança de cada vértice: $nb\left(V_j\right) = \left\{V_i; \beta_{V_{i|S}}^{V_j} \neq 0,\ j \neq i\,(j,i=1,\dots,p)\right\}$

Incluir arestas se: $\hat{E} = \left\{\left(X,Y\right); X \in nb\left(Y\right) \text{ and } Y \in nb\left(X\right)\right\}$

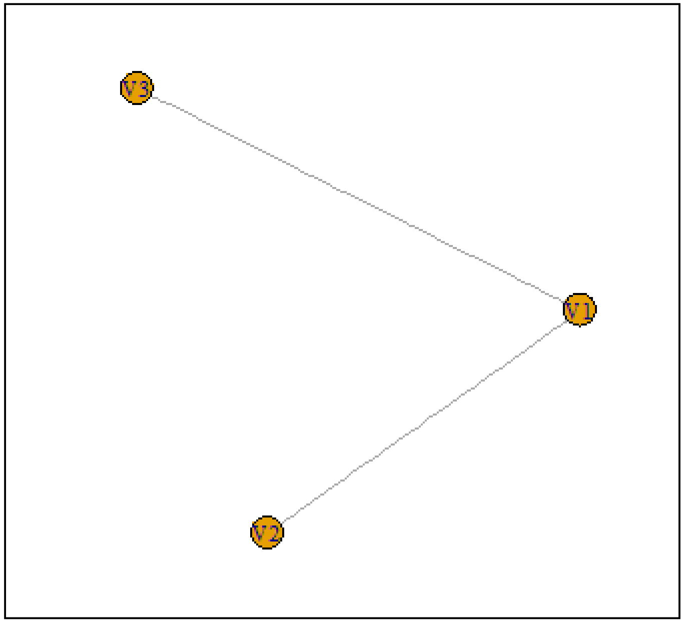
Critério AND/OR

Simulação de Grafos: "UDG"

Adèle Ribeiro, 2018,
IME-USP

Uso do R

```
Entrada: Matriz de Correlação Parcial
> data.pCor
      [,1]  [,2] [,3]
[1,]   1.0   0.6  -0.5
[2,]   0.6   1.0   0.0
[3,]  -0.5   0.0   1.0
```



#Gerar dados N_3
#Aprender a estrutura
#Estimar as arestas via Método das Vizinhanças
#Teste z de Fisher

```
> nei_out$pCor$estimates
    V1          V2          V3
V1   NA         0.64450169  -0.48553853
V2   0.6445017  NA           0.04922369
V3  -0.4855385  0.04922369   NA

> nei_out$pCor$p.values
   V1              V2              V3
V1 NA              1.825967e-118   2.817058e-60
V2 1.825967e-118   NA              1.198066e-01
V3 2.817058e-60    1.198066e-01    NA
```

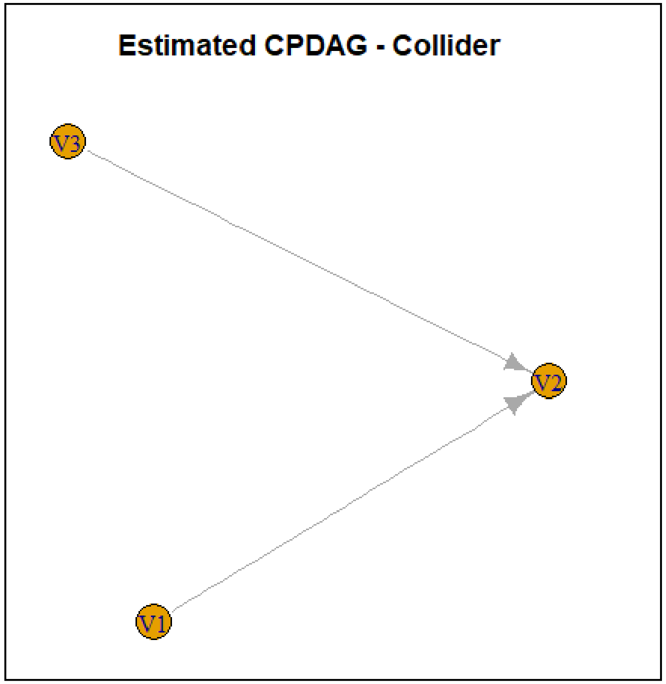
Não rejeitar a correlação
parcial nula entre (V2,V3) | V1

Simulação de Grafos: “DAG” p=3 Variáveis

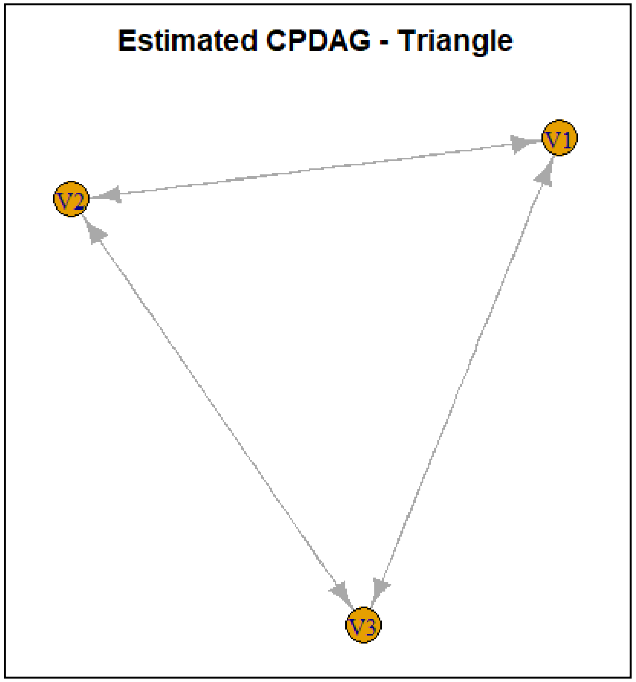
Adèle Ribeiro, 2018,
IME/USP

```
# Possible types:
#   - chain: V1 -> V2 -> V3
#   - collider: V1 -> V2 <- V3
#   - fork: V1 <- V2 -> V3
#   - triangle: V1 -> V2 <- V3 <- V1
#   - independent: V1  V2  V3
```

#Gerar dados via Modelos de Mensuração (SEM)
#Entrar com os Betas: no exemplo, Beta=0 e 0.8
#Testes de Independência Condicional com
p-valor=0.01 e Correção de Bonferroni



A V-estrutura (colisor) pode ser aprendida, as demais nem sempre!.

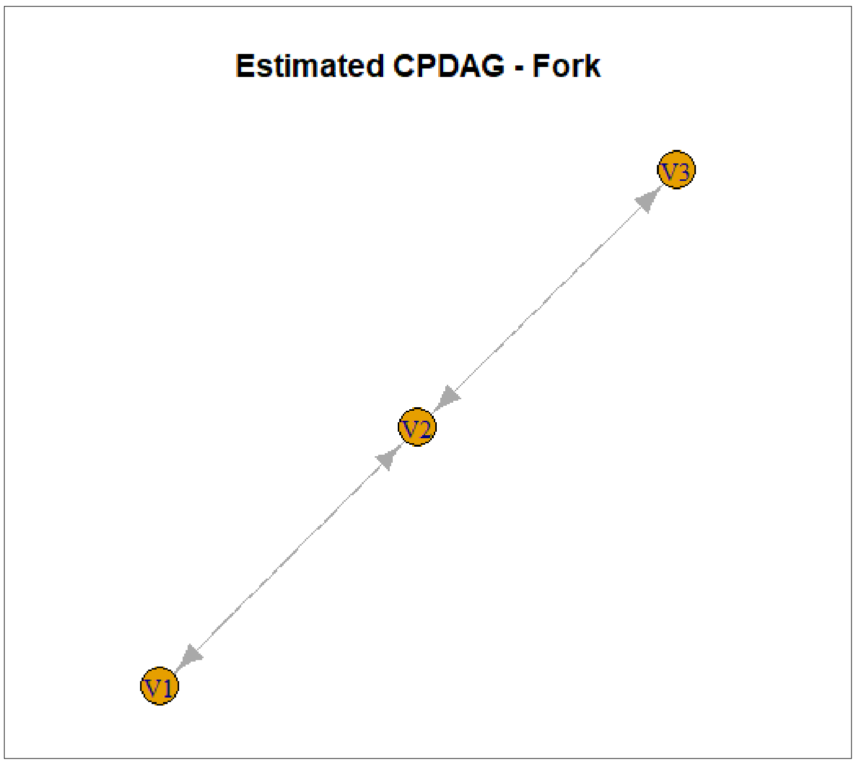


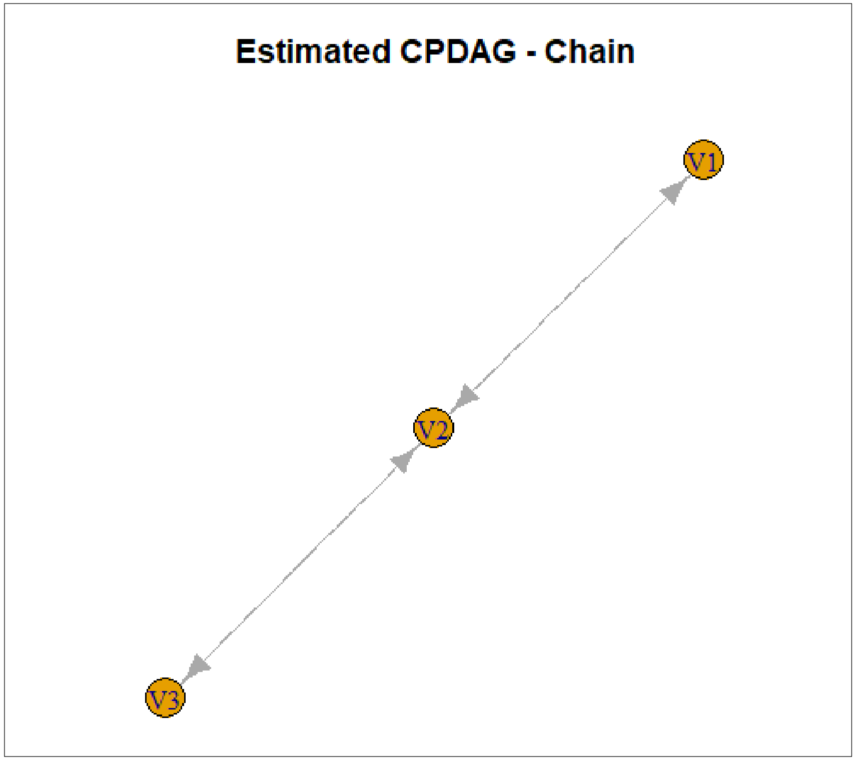
Simulação de Grafos: DAG p=3 Variáveis

Adèle Ribeiro, 2018

```
# Possible types:
# - chain: V1 -> V2 -> V3
# - collider: V1 -> V2 <- V3
# - fork: V1 <- V2 -> V3
# - triangle: V1 -> V2 <- V3 <- V1
# - independent: V1  V2  V3
```

Estruturas *Chain* e *fork* têm a mesma distribuição conjunta e não podem ser estatisticamente distinguíveis.





Simulação de Grafos: “DAG”
p=5 Variáveis

Adèle Ribeiro, 2018

```
# Possible types:
# - multichain
# - multicollider: V2, V3, and V4 are colliders and sink/terminal vertices;
#                  V1 and V5 are source nodes;
# - multifork: V2, V3, and V4 are source nodes;
#              V1 and V5 are colliders and sink/terminal vertices.
```