

MAE 5776

ANÁLISE MULTIVARIADA

Júlia M Pavan Soler
pavan@ime.usp.br


Verão IME/2019

Análise Multivariada

$$Y_{n \times p} = (Y_{ij}) \in \mathfrak{R}^{n \times p}$$

- Estatísticas Descritivas Multivariadas
- Distribuição Normal Multivariada
- Distribuições Amostras
- Regiões de Confiança, Testes Multivariados, MANOVA, IC Simultâneos, Correções para Múltiplos Testes
- Análises Multivariadas Clássicas ($n > p$, iid): CP, CoP, AC, AF, AD, AAgr, CC
- Análises Multivariadas Esparsas ($n \ll p$, iid): CP, AD, CC, PLS

Já vimos 😊

- 
- Componentes Principais em Observações Correlacionadas
 - Aprendizado de Estruturas – Modelos de Grafos Probabilísticos (Modelos de Equações Estruturais)
Propriedades de Markov

Matriz de Dados

Unidades Amostras	Variáveis					
	1	2	...	j	...	p
1	Y_{11}	Y_{12}		Y_{1j}		Y_{1p}
2	Y_{21}	Y_{22}		Y_{2j}		Y_{2p}
...
i	Y_{i1}	Y_{i2}		Y_{ij}		Y_{ip}
...
n	Y_{n1}	Y_{n2}		Y_{nj}		Y_{np}

- Amostra Aleatória Simples de n-Vetores em \mathcal{R}^p

$$Y_{n \times p} \in \mathcal{R}^{n \times p}; \quad Y_i \in \mathcal{R}^p \stackrel{iid}{\sim} (\mu; \Sigma)$$

Já vimos vários resultados
sob a formalização de
observações independentes
(AASn)!

Casos
mais
gerais:

- Amostra Aleatória de G-Matrizes em $\mathcal{R}^{n_g \times p}$ ($g = 1, \dots, G; \sum_g n_g = n$)
(AAS_G)

Variáveis Aleatórias Multidimensionais

- Matriz aleatória (Gupta and Nagar, 2000):

Formulações alternativas

$$Y_{n \times p} = (Y_{ij}) \in \mathbb{R}^{n \times p}; \quad Y_{n \times p} \sim N_{n,p}(M; \Psi \otimes \Sigma); \quad \text{vec}(Y)_{np \times 1} \sim N_{np}(\text{vec}(M); \Psi \otimes \Sigma)$$

$$M_{n \times p} = \mathbf{1}_n \mu'_{p \times 1} : \text{matriz de médias}$$

$$\text{vec}(M)_{np \times 1} = \mathbf{1}_n \otimes \mu_{p \times 1} : \text{vetor de médias de } n \text{ observações em } p \text{ variáveis}$$

$$(\Psi_{n \times n} \otimes \Sigma_{p \times p})_{np \times np} : \text{matriz de covariâncias}$$

Formulação flexível para diferentes modelagens

Matrizes de covariância Estruturadas: entre indivíduos (Ψ) e entre variáveis (Σ)

$$\Psi = I_n; \quad \Sigma = I_p$$

Observações e variáveis independentes

$$\Psi = I_n; \quad \Sigma = (1 - \rho) I_p + \rho \mathbf{1}_p \mathbf{1}_p'$$

Observações independentes e correlação uniforme entre as variáveis

$$\Psi = \bigoplus_{g=1}^G \left[(1 - \rho) I_{n_g} + \rho \mathbf{1}_{n_g} \mathbf{1}_{n_g}' \right]; \quad \Sigma = (\sigma_{jl})$$

Correlação uniforme entre observações agrupadas em G grupos

Correlação não estruturada entre variáveis

Matriz Aleatória

- Amostra Aleatória de n -Vetores em \mathbb{R}^p

$$Y_{n \times p} \in \mathbb{R}^{n \times p}; \quad Y_i \in \mathbb{R}^p \stackrel{iid}{\sim} (\mu; \Sigma)$$

- Amostra Aleatória de n -Vetores em \mathbb{R}^p tal que $n = \sum_{g=1}^G n_g$

Já vimos!

Considere Amostra Estratificada MAS com
Correlação entre Observações dentro dos Estratos (Grupos)

Grupos	Unidades amostrais	Variáveis					
		1	2	...	j	...	p
1	1	Y_{11}	Y_{12}		Y_{1j}		Y_{1p}
	2	Y_{21}	Y_{22}		Y_{2j}		Y_{2p}
...
	i	Y_{i1}	Y_{i2}		Y_{ij}		Y_{ip}
G
	n	Y_{n1}	Y_{n2}		Y_{nj}		Y_{np}

$$Y_{n \times p} = \begin{pmatrix} Y_1 \\ \dots \\ Y_G \end{pmatrix}$$

$$Y_{g(n_g \times p)}$$

$$n = \sum_{g=1}^G n_g$$

$$Y_{n \times p} \in \mathbb{R}^{n \times p}; \quad Y_g = Y_{n_g \times p} \in \mathbb{R}^{n_g \times p} \stackrel{iid}{\sim} (\mu_g; \Omega_g) \left\{ \begin{array}{l} \mu_g (n_g \times p) = 1_{n_g} \otimes \mu'_{p \times 1} \\ \Omega_g (n_g \times p \times n_g \times p) = \Psi_g \otimes \Sigma_{p \times p} \end{array} \right.$$

Matriz Aleatória

- Amostra Aleatória de G-Matrizes em $\mathfrak{R}^{n_g \times p}$ ($g = 1, \dots, G; \sum_g n_g = n$)

$$Y_{n \times p} \in \mathfrak{R}^{n \times p}; \quad Y_g = Y_{n_g \times p} \in \mathfrak{R}^{n_g \times p} \stackrel{iid}{\sim} (\mu_g; \Omega_g) \left\{ \begin{array}{l} \mu_{g(n_g \times p)} = 1_{n_g} \otimes \mu'_{p \times 1} \\ \Omega_{g(n_g \times p \times n_g \times p)} = \Psi_g \otimes \Sigma_{p \times p} \end{array} \right.$$

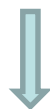


$$Y_{n \times p} \sim (\mu_{n \times p}; \Omega_{np \times np}) \left\{ \begin{array}{l} \mu_{n \times p} = \bigoplus_{g=1}^G (1_{n_g} \otimes \mu'_{p \times 1}) = \left(\bigoplus_{g=1}^G 1_{n_g} \right) \otimes \mu'_{p \times 1} \\ \Omega_{np \times np} = \bigoplus_{g=1}^G \Omega_g = \bigoplus_{g=1}^G (\Psi_g \otimes \Sigma_{p \times p}) = \left(\bigoplus_{g=1}^G \Psi_g \right) \otimes \Sigma_{p \times p} \end{array} \right.$$

Matriz Aleatórias – Componentes de Covariâncias em $\mathfrak{R}^{p \times p}$

- Amostra Aleatória de G-Matrizes em $\mathfrak{R}^{n_g \times p}$ ($g = 1, \dots, G; \sum_g n_g = n$)

$$Y_{n \times p} \in \mathfrak{R}^{n \times p}; \quad Y_g = Y_{n_g \times p} \in \mathfrak{R}^{n_g \times p} \stackrel{iid}{\sim} (\mu_g; \Omega_g) \begin{cases} \mu_{g(n_g \times p)} = 1_{n_g} \otimes \mu'_{p \times 1} \\ \Omega_{g(n_g p \times n_g p)} = \Psi_g \otimes \Sigma_{p \times p} \end{cases}$$



$$Y_{n \times p} \sim (\mu_{n \times p}; \Omega_{np \times np}) \begin{cases} \mu_{n \times p} = \left(\bigoplus_{g=1}^G 1_{n_g} \right) \otimes \mu'_{p \times 1} \\ \Omega_{np \times np} = \bigoplus_{g=1}^G (\Psi_g \otimes \Sigma_{p \times p}) = \left(\bigoplus_{g=1}^G \Psi_g \right) \otimes \Sigma_{p \times p} \end{cases}$$

Decompor $\mathfrak{R}^{p \times p}$ em
Componentes de Covariâncias



$$\Sigma_{p \times p} = \Sigma_B + \Sigma_W$$



$$\Omega_{np \times np} = \left(\bigoplus_{g=1}^G \Psi_g \right) \otimes \Sigma \Rightarrow \Omega = \left(\bigoplus_{g=1}^G \Psi_g \right) \otimes \Sigma_B + I_n \otimes \Sigma_W$$

Σ_B : componente da covariância entre as p-variáveis,
associado à covar. entre as n-observações

Componentes de Covariâncias em $\mathfrak{R}^{p \times p}$

$$Y_{n \times p} \sim (\mu_{n \times p}; \Omega_{np \times np}) \left\{ \begin{array}{l} \mu_{n \times p} = \bigoplus_{g=1}^G (1_{n_g} \otimes \mu'_{g \ p \times 1}) \\ \Omega_{np \times np} = \left(\bigoplus_{g=1}^G \Psi_g \right) \otimes \Sigma_B + I_n \otimes \Sigma_W \end{array} \right.$$

$$\Omega_{np \times np} = \begin{bmatrix} \boxed{\Psi_1} & & 0 \\ & \boxed{\Psi_2} & \\ 0 & & \dots \\ & & \boxed{\Psi_G} \end{bmatrix} \otimes \Sigma_B + \begin{bmatrix} I_{n_1} & & 0 \\ & I_{n_2} & \\ 0 & & \dots \\ & & I_{n_G} \end{bmatrix} \otimes \Sigma_W$$

A matriz de covariância Σ é decomposta em dois componentes, um associado à suposição de correlação entre observações do mesmo grupo (Σ_B : **matriz de covariância ENTRE grupos**) e outro associado à independência condicional entre observações dado o grupo (Σ_W : **matriz de covariância DENTRO de grupos**)

Componentes Principais em Modelos de Componentes de Covariâncias

$$\text{Cov}(Y_g) = \Omega_g = \Psi_g \otimes \Sigma_B + I_{n_g} \otimes \Sigma_W; \quad \Sigma_B + \Sigma_W = \Sigma$$

$$PC_B \Rightarrow \max_{\|a\|=1} \frac{a' \Sigma_B a}{a' a} : \text{direção com máxima variação entre grupos}$$

$$PC_W \Rightarrow \max_{\|a\|=1} \frac{a' \Sigma_W a}{a' a} : \text{direção com máxima variação dentro dos grupos}$$

$$PCH \Rightarrow \max_{\|\Sigma_W^{1/2} a\|=1} \frac{a' \Sigma_B a}{a' \Sigma a} = \max_{\|\Sigma_W^{1/2} a\|=1} \frac{a' \Sigma_B a}{a' \hat{\Sigma}_W a} : \text{direção com máxima variação entre e mínima variação dentro de grupos}$$



Obter estimativas dos Componentes de Covariância!

$$\hat{\Sigma}_B, \quad \hat{\Sigma}_W$$

ANOVA – Efeito Fixo e Aleatório

Como obter estimativas de componentes de variância?
Recordando o caso $p=1$

$$Y_{gi} = \mu_g + e_{gi}$$

$$= \mu + \tau_g + e_{gi}; \quad i = 1, \dots, n_g, \quad g = 1, \dots, G$$

Modelo de Efeitos Fixos

$$\sum_{g=1}^G \tau_g = 0$$

$$e_{gi} \stackrel{iid}{\sim} (0; \sigma^2)$$

$$\Rightarrow Y_{gi} \stackrel{iid}{\sim} (\mu_g; \sigma^2)$$

$$Cov(Y_{n \times 1}) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

Modelo de Efeitos Aleatórios

$$\tau_g \sim (0; \sigma_B^2) \quad \mu_g = \mu + \tau_g \sim (\mu; \sigma_B^2)$$

$$e_{gi} \sim (0; \sigma_W^2) \quad \tau_g \perp e_{gi}$$

σ_b^2 : é componente de variância e de covariância

$$\Rightarrow Y_{gi} \sim (\mu; \sigma^2 = \sigma_B^2 + \sigma_W^2)$$

$$Cov(Y_{gi}; Y_{g'i'}) = \begin{cases} \sigma_B^2 + \sigma_W^2; & i = i', g = g' \\ \sigma_B^2; & i \neq i', g = g' \\ 0; & i \neq i', g \neq g' \end{cases}$$

$$Cov(Y_{n \times 1}) = \begin{bmatrix} \sigma^2 & \sigma_B^2 & 0 & 0 \\ \sigma_B^2 & \sigma^2 & 0 & 0 \\ 0 & 0 & \dots & \sigma_B^2 \\ 0 & 0 & \sigma_B^2 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & \dots & \rho \\ 0 & 0 & \rho & 1 \end{bmatrix}$$

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

Coeficiente de correlação intraclasse

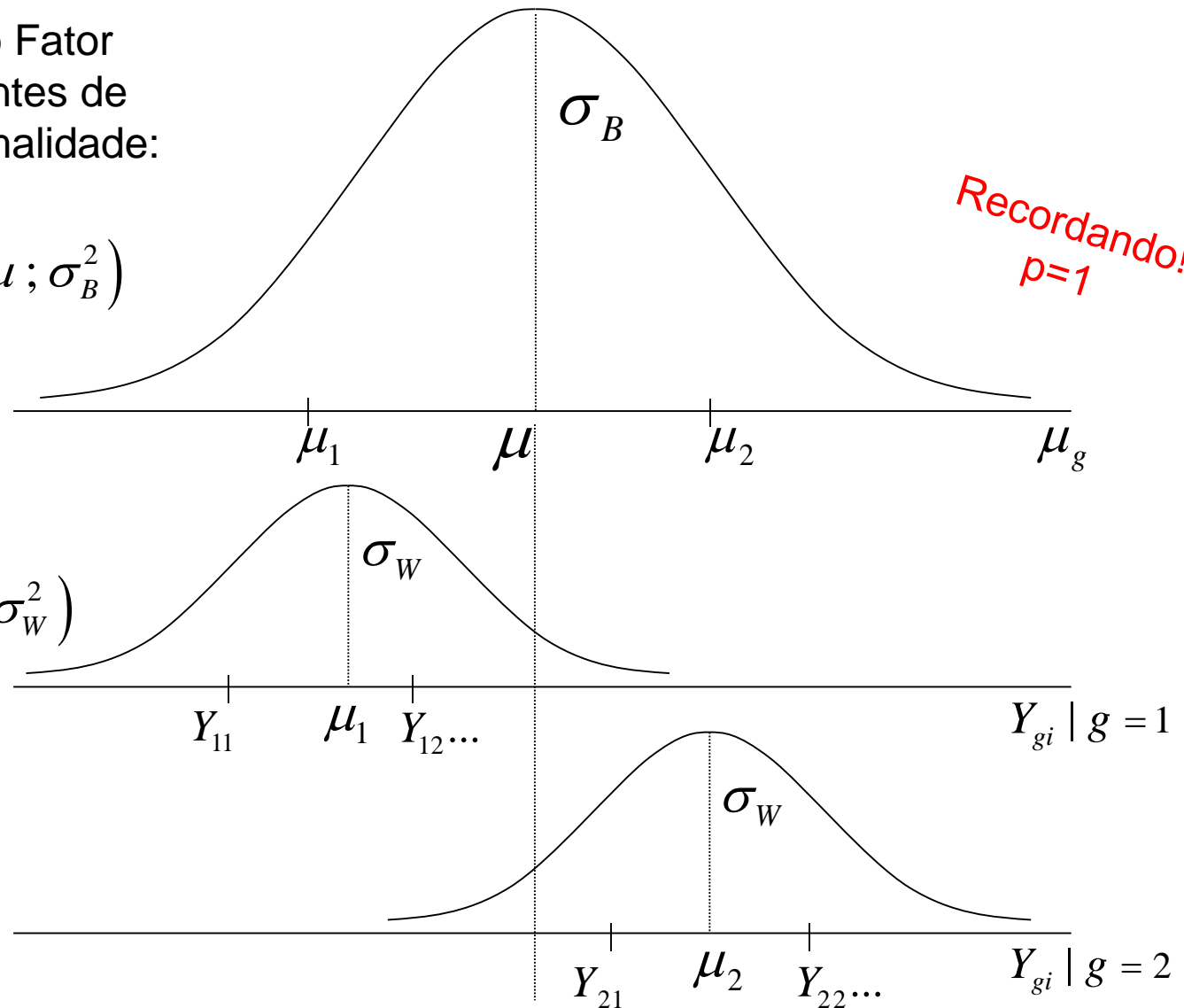
Modelo de um único Fator Aleatório (componentes de variância), sob Normalidade:

Recordando!
 $p=1$

$$\mu_g = \mu + \tau_g \sim N(\mu; \sigma_B^2)$$

Distribuição condicional de Y:

$$\Rightarrow y_{gi} | g \sim N(\mu_g; \sigma_W^2)$$



$$H_0 : \sigma_B^2 = 0 \Leftrightarrow H_0 : \mu_g = \mu$$

Sob H_0 , os modelos de ANOVA com um fator aleatório ou um fator fixo são equivalentes (mesma tabela ANOVA)

Tabela de ANOVA

Recordando o caso $p=1$.
Tabela ANOVA é
equivalente para os
modelos de um fator fixo
ou um fator aleatório!

$$H_0 : \sigma_B^2 = 0 \Leftrightarrow H_0 : \mu_g = \mu$$

F.V.	g.l.	SQ	QM	F
Grupo	G-1	$S_b = \sum r (\bar{Y}_g - \bar{Y})^2$	$QMGr = S_b / (G-1)$	$\frac{QMGr}{QM Res}$
Resíduo	n-G	$S_w = \sum_{gi} (Y_{gi} - \bar{Y}_g)^2$	$QM Res = S_w / (n-G)$	
TOTAL	n-1	$S_T = \sum_{gi} (Y_{gi} - \bar{Y})^2$		

$n_g = r$: Dados balanceados

$$E(QM Res) = \sigma_w^2 \Rightarrow \hat{\sigma}_w^2 = QM Res$$

$$E(QMTr) = \sigma_w^2 + r \sigma_B^2 \Rightarrow \hat{\sigma}_B^2 = \frac{QMTr - QM Res}{r}$$

$$F \sim F_{(G-1), (n-G)}$$

Componentes de Covariâncias em $\mathfrak{R}^{p \times p}$

Estruturas para Ψ em $\mathfrak{R}^{n \times n}$

- **Correlação Uniforme entre Observações** (Simetria Composta ou Equicorrelação)

(Konish and Rao, 1992)

$$Y_{n \times p} \in \mathfrak{R}^{n \times p}; \quad Y_g = Y_{n_g \times p} \in \mathfrak{R}^{n_g \times p} \stackrel{iid}{\sim} (\mu_g; \Omega_g) \left\{ \begin{array}{l} \mu_g (n_g \times p) = 1_{n_g} \otimes \mu'_{g \ p \times 1} \\ \Omega_g (n_g p \times n_g p) = \Psi_g \otimes \Sigma_{p \times p} \end{array} \right.$$

$$Cov(Y_{gi}, Y_{g'i'})_{(p \times p)} = \begin{cases} \Sigma = \Sigma_B + \Sigma_W & \text{se } g = g', i = i' \\ \Sigma_B & \text{se } g = g', i \neq i' \\ 0 & \text{se } g \neq g', i \neq i' \end{cases}$$

$$\begin{aligned} \Psi_g &= 1_{n_g} 1_{n_g}' \\ \Sigma &= \Sigma_B + \Sigma_W \end{aligned}$$

$$\Rightarrow Cov(Y_g)_{n_g p \times n_g p} = \Omega_g = (1_{n_g} 1_{n_g}') \otimes \Sigma_B + I_{n_g} \otimes \Sigma_W$$

$$\Rightarrow Y_g \stackrel{iid}{\sim} (1_{n_g} \otimes \mu'_g; \Omega_g = (1_{n_g} 1_{n_g}') \Sigma_B + I_{n_g} \otimes \Sigma_W)$$

$$\Rightarrow Y_{n \times p} \sim \left(\bigoplus_{g=1}^G (1_{n_g} \otimes \mu'_g); \Omega = \bigoplus_{g=1}^G \Omega_g \right)$$

Estimadores MANOVA

Equivalência analítica entre MANOVA com um Fator Fixo ou um Fator Aleatório (sob equicorrelação)

$$\Rightarrow Y_g \stackrel{iid}{\sim} \left(1_{n_g} \otimes \mu'_g; \Omega_g = (1_{n_g} 1_{n_g}') \Sigma_B + I_{n_g} \otimes \Sigma_W \right)$$

Tabela de MANOVA:

F.V.	g.l.	Matriz de SQPC
Trat	G-1	$H_{p \times p} = \sum_{g=1}^G n_g (\bar{Y}_g - \bar{Y})(\bar{Y}_g - \bar{Y})' = S_b$
Resíduo	n-G	$E_{p \times p} = \sum_{g=1}^G \sum_{i=1}^{n_g} (Y_{gi} - \bar{Y}_g)(Y_{gi} - \bar{Y}_g)' = S_w$
TOTAL	n-1	$H + E = \sum_{g=1}^G \sum_{i=1}^{n_g} (Y_{gi} - \bar{Y})(Y_{gi} - \bar{Y})'$

Sob $H_0 : \mu_g = \mu, \quad g = 1, \dots, G$

$$E\left(\frac{S_w}{n-G}\right) = \Sigma_W; \quad E\left(\frac{S_b}{G-1}\right) = \Sigma_W + n_0 \Sigma_B$$

$$n_0 = \frac{n - \left(\sum_g n_g^2 / n \right)}{G-1}$$

Componentes de Covariâncias em $\Re^{p \times p}$ Sob Correlação Uniforme entre Observações

$$\Rightarrow \hat{\Sigma}_W = \frac{S_w}{n-G}; \quad \hat{\Sigma}_B = n_0^{-1} \left\{ \frac{S_b}{G-1} - \frac{S_w}{n-G} \right\}$$

$$\Rightarrow \hat{\Sigma} = \hat{\Sigma}_W + \hat{\Sigma}_B = n_0^{-1} \left\{ \frac{S_b}{G-1} + \frac{(n_0-1)S_w}{n-G} \right\}$$

$$n_0 = \frac{n - \left(\sum_g n_g^2 / n \right)}{G-1}$$

Componentes de Covariâncias em $\mathbb{R}^{p \times p}$ Sob Correlação Uniforme entre Observações

Konishi and Rao, 1992; Oualkacha et al., 2012)

$$\Rightarrow Y_g \stackrel{iid}{\sim} \left(1_{n_g} \otimes \mu'_g; \Omega_g = (1_{n_g} 1_{n_g}') \Sigma_B + I_{n_g} \otimes \Sigma_W \right); \quad \hat{\Sigma}_{p \times p} = \hat{\Sigma}_{B_{p \times p}} + \hat{\Sigma}_{W_{p \times p}}$$

$$PC_g \Rightarrow \max_a \frac{a' \hat{\Sigma}_B a}{a' a}, \quad a' a = 1 \quad \text{Direção com máxima variação Entre grupos}$$

$$PC_e \Rightarrow \max_a \frac{a' \hat{\Sigma}_W a}{a' a}, \quad a' a = 1 \quad \text{Direção com máxima variação Dentro de grupos}$$

$$PC_T \Rightarrow \max_a \frac{a' \hat{\Sigma} a}{a' a} = \max_a \frac{a' [\hat{\Sigma}_B + \hat{\Sigma}_W] a}{a' a}, \quad a' a = 1 \quad \text{Direção com máxima variação Total}$$

$$PCH \Rightarrow \max_a \frac{a' \hat{\Sigma}_B a}{a' [\hat{\Sigma}_B + \hat{\Sigma}_W] a} = \max_a \frac{a' \hat{\Sigma}_B a}{a' \hat{\Sigma}_W a}, \quad a' \hat{\Sigma}_W a = 1 \quad \text{Direção com máxima variação Entre grupos e mínima variação Dentro}$$

$$\Rightarrow \hat{\Sigma}_W^{-1} \hat{\Sigma}_B \Rightarrow \hat{\Sigma}_W^{-1/2} \hat{\Sigma}_B \hat{\Sigma}_W^{-1/2}$$

Componente
Principal de
Herdabilidade

Componentes Principais em Modelos de Componentes de Covariâncias

$$\Rightarrow Y_g \stackrel{iid}{\sim} \left(1_{n_g} \otimes \mu'_g; \Omega_g = (1_{n_g} 1_{n_g}') \Sigma_B + I_{n_g} \otimes \Sigma_W \right)^{(**)}$$

$$PCH \Rightarrow \max_a \frac{a \hat{\Sigma}_B a}{a \hat{\Sigma}_W a}$$

: Componentes Principais de Herdabilidade
(direção com máxima variação Entre e
mínima variação Dentro de grupos)

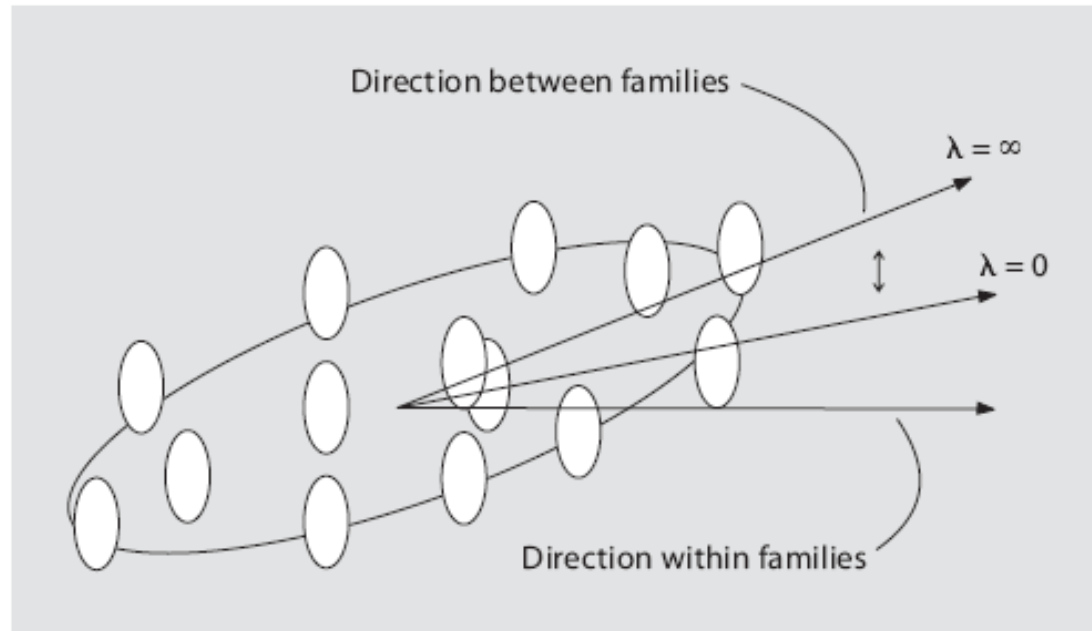
$$\Rightarrow \hat{\Sigma}_W^{**} = \frac{S_w}{n-G}; \quad \hat{\Sigma}_B = n_0^{-1} \left\{ \frac{S_b}{G-1} - \frac{S_w}{n-G} \right\} \quad n_0 = \frac{n - \left(\sum_g n_g^2 / n \right)}{G-1}$$

Os PCH de dados correlacionados (sob **) correspondem às direções da Análise Discriminante sob obs. independentes!

Obter os autovalores e autovetores (V) de $\hat{\Sigma}_W^{-1} \hat{\Sigma}_B \Rightarrow PCH = YV$

Componentes Principais-Componentes de Covariância

Soluções Regularizadas



(Wang, 2007)

Elipse vertical: corresponde à variabilidade dentro dos grupos (famílias)

Elipse maior: corresponde à variação entre grupos (famílias)

$$PCH_{\lambda} \Rightarrow \max_{a, \lambda > 0} \frac{a' \hat{\Sigma}_B a}{a' \hat{\Sigma}_W a + \lambda \|a\|^2} = \max_{a, \lambda > 0} \frac{a' \hat{\Sigma}_B a}{a' [\hat{\Sigma}_W + \lambda I_p] a}$$

Solução
Regularizada
 $n \ll p$

$\lambda=0$: solução não regularizada do PCH

$\lambda=\infty$: solução do PCH próxima à solução para Σ_B . (maximização entre grupos)

Componentes Principais-Componentes de Covariância Soluções Regularizadas

Algoritmo de obtenção do parâmetro de regularização λ


Passo 1: Partição dos grupos em dois sub-grupos: Grupo 1 e Grupo 2.
(Repetir L=50 vezes)

Passo 2: Grupo1: para $\lambda=0.01$, obter o j-ésimo autovetor $V_{j\lambda}^{(1)l}$ de $\left(\hat{\Sigma}_W + \lambda I_p\right)^{-1/2}$.

Será necessário substituir os autovalores negativos por “0” (Amemiya, 1985).

Grupo2: obter as estimativas $\hat{\Sigma}_B^{(2)l}$, $\hat{\Sigma}_W^{(2)l}$;

Passo 3. Repetir para $\lambda = 0.01, 2, 4, \dots, 1000$. O parâmetro de regularização é estimado como:

$$\lambda_{CV} = \arg \max_{\lambda} \frac{1}{L} \sum_{l=1}^L \frac{V_{j\lambda}^{(1)l'} \hat{\Sigma}_B^{(2)l} V_{j\lambda}^{(1)l}}{V_{j\lambda}^{(1)l'} \hat{\Sigma}_W^{(2)l} V_{j\lambda}^{(1)l}}; \quad PCH_{\lambda_{CV}} = Ya'; \max_a \frac{a' \hat{\Sigma}_B a}{a' [\hat{\Sigma}_W + \lambda_{CV} I_p] a}$$


Componentes Principais-Componentes de Covariâncias

Correlação Familiar

Matriz de covariância mais geral entre obs.!

$Y_{n \times p}$ Observações correlacionadas devido à estrutura familiar (grau de parentesco entre indivíduos)

Modelo de Componentes de Covariância

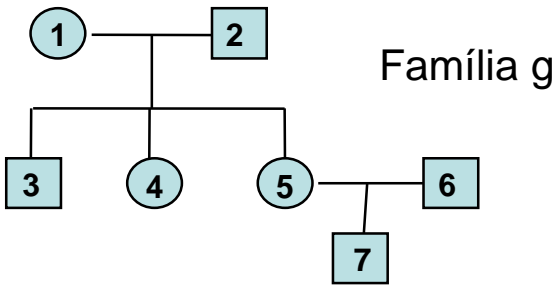
Família	Unidade Amostral	Y_1	Y_2	...	Y_p
1	1	Y_{111}	Y_{112}		Y_{11p}
1	2	Y_{121}	Y_{122}		Y_{12p}
...	...				
1	n_1	Y_{1n11}	Y_{1n12}		Y_{1n1p}
Médias da Família 1		\bar{Y}_{11}	\bar{Y}_{12}		\bar{Y}_{1p}
...					
G	1	Y_{G11}	Y_{G12}		Y_{G1p}
G	2	Y_{G21}	Y_{G22}		Y_{G2p}
...	...				
G	n_G	Y_{GnG1}	Y_{GnG2}		Y_{GnGp}
Médias da Família G		\bar{Y}_{G1}	\bar{Y}_{G2}		\bar{Y}_{Gp}
Vetor de Médias Geral		$\bar{Y}_{.1}$	$\bar{Y}_{.2}$		$\bar{Y}_{.p}$

$\Rightarrow Cov(Y_g) = \Omega_g = \Psi_g \otimes \Sigma_B + I_{n_g} \otimes \Sigma_W$

matriz de parentesco

$\Psi_g =$

	1	2	3	4	5	6	7
1	1	0	1/2	1/2	1/2	0	1/4
2	0	1	1/2	1/2	1/2	0	1/4
3	1/2	1/2	1	1/2	1/2	0	1/4
4	1/2	1/2	1/2	1	1/2	0	1/4
5	1/2	1/2	1/2	1/2	1	0	1/2
6	0	0	0	0	0	1	1/2
7	1/4	1/4	1/4	1/4	1/2	1/2	1



Componentes Principais-Componentes de Covariâncias

Correlação Familiar

Estimadores MANOVA do Modelo de Componentes de Covariância: são funções lineares de S_b e S_w (Oualkacha et al., 2012)

$$\hat{\Sigma}_B = \frac{S_b / (G-1) - S_w / (n-G)}{(\tau_c - \tau_b / n) / (G-1) - (\tau_a - \tau_c) / (n-G)}$$

$$\hat{\Sigma}_W = \frac{1}{(n-G)} S_w - \frac{(\tau_a - \tau_c)}{(n-G)} \hat{\Sigma}_B$$

$$n = \sum_{g=1}^G n_g, \quad \tau_a = \sum_{g=1}^G \tau_{a_g}, \quad \tau_b = \sum_{g=1}^G \tau_{b_g}, \quad \tau_c = \sum_{g=1}^G \frac{1}{n_g} \tau_{b_g}$$

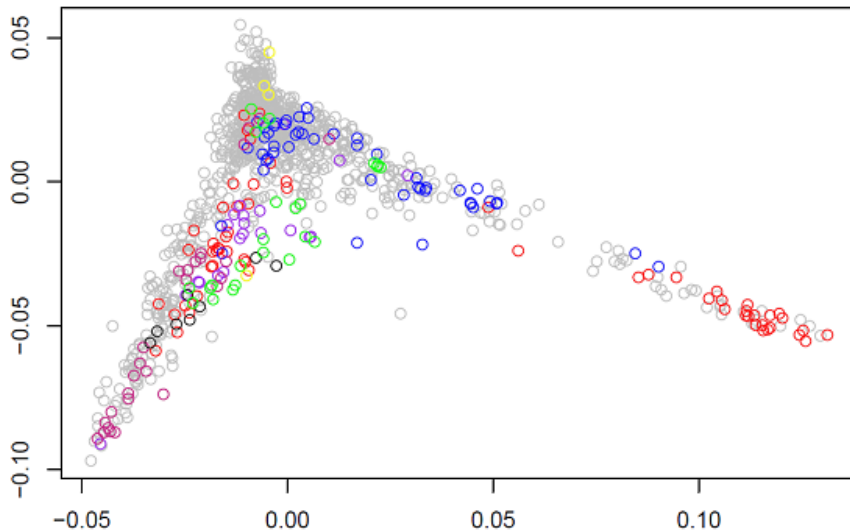
$$\tau_{a_g} = 2\text{Trace}[\Phi_g], \quad \tau_{b_g} = 2 \sum_{\substack{i'=1 \\ i'>i}}^{n_g} \sum_{i=1}^{n_g} (\Phi_g)_{ii'}$$

Componentes Principais-Componentes de Covariância

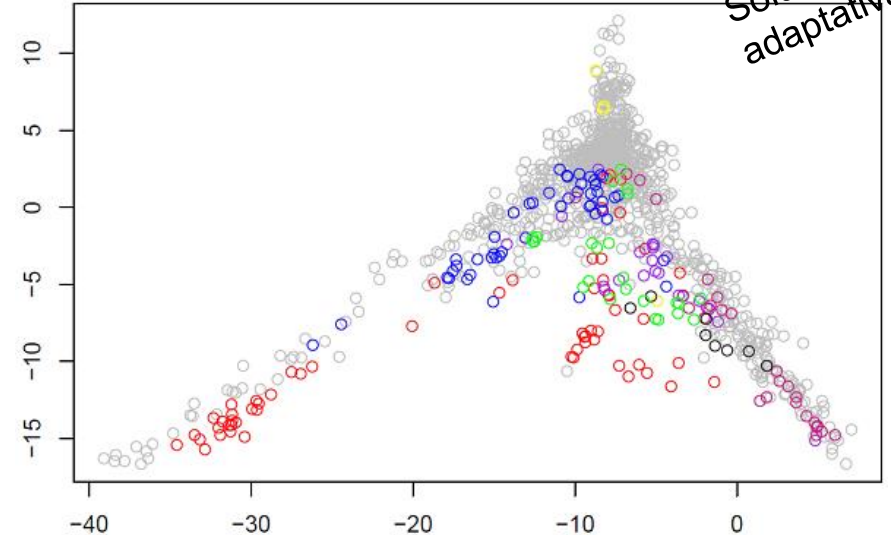
Correlação Familiar (de Andrade et al., 2015)

Aplicação: Projeto Corações de Baependi (MG)
 n=1.109 indivíduos de G=80 famílias e p=8.764 variáveis genéticas (SNPs)

PC sob Independência: $PC \Rightarrow \max_{\|a\|=1} \frac{a'Ra}{a'a}$



PCH $\Rightarrow \max_{\|R_W^{1/2}a\|=1} \frac{a'R_Ba}{a'R_Wa}$



Solução mais adaptativa

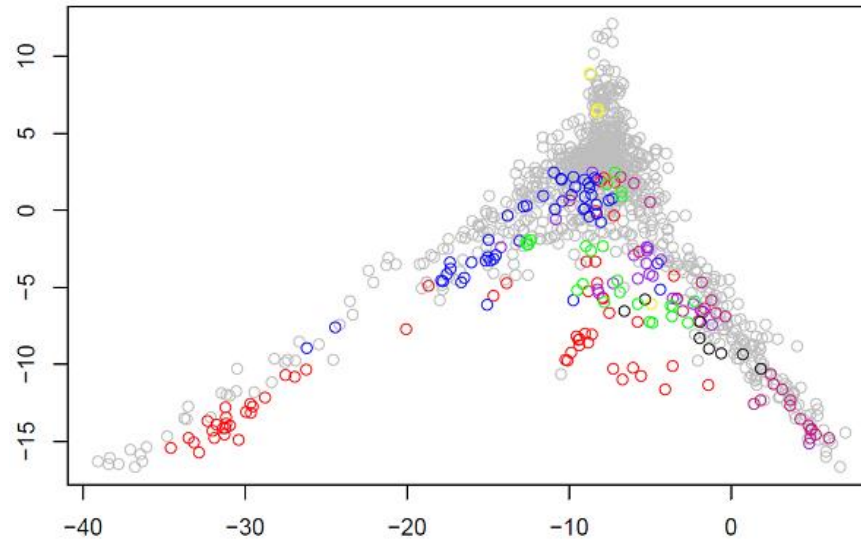
Proporção da variância explicada pelos PC1 e PCH1

CP	1	2	3	4	5	6	7	8	9	10
R	0.022	0.014	0.008	0.0069	0.0068	0.0061	0.0059	0.0055	0.0053	0.005
$R_W^{-1}R_B$	0.086	0.070	0.035	0.031	0.028	0.026	0.0255	0.0249	0.0246	0.0239

Componentes Principais-Componentes de Covariância

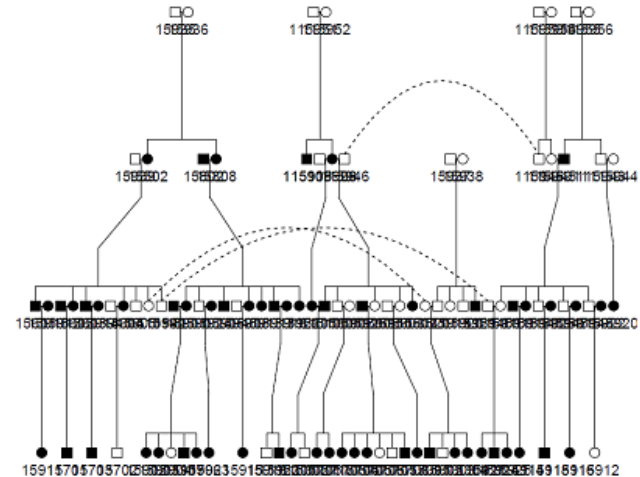
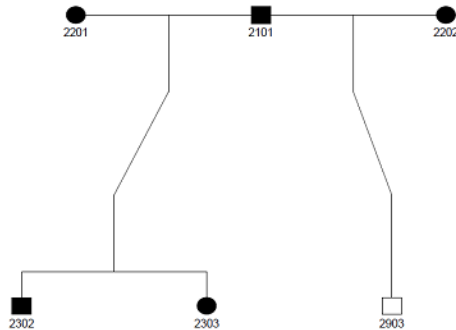
Correlação Familiar

PCH1 x PCH2



○ Família: 5 membros (homogênea)

○ Família: 60 membros (heterogênea)



Componentes Principais-Componentes de Covariâncias

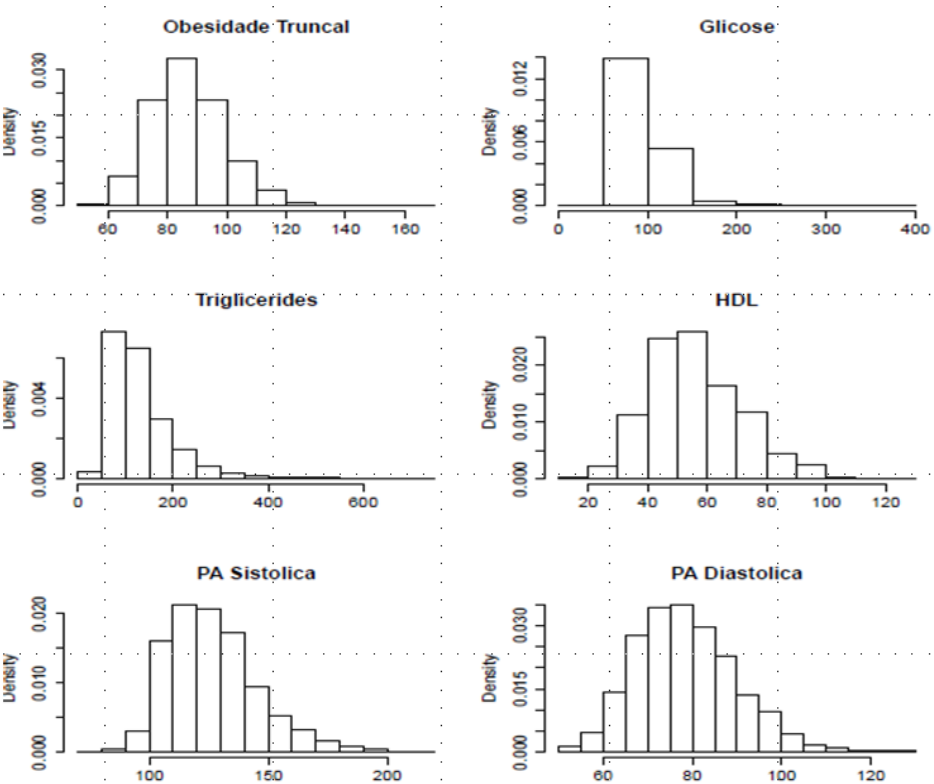
Correlação Familiar

Aplicação: Projeto Corações de Baependi (MG)
 Componentes Principais da Síndrome Metabólica (doença multifatorial: p=6)

Calcular o PCH1 (variável latente da SM)

$$Y_g \sim \left(1_{n_g} \mu_g'; \Psi_g \otimes \Sigma_B + I_n \otimes \Sigma_W\right)$$

$$\max_a \frac{a' \hat{\Sigma}_B a}{a' \hat{\Sigma}_W a} \quad PCH = a' Y_{6 \times 1}$$



Var.	ObTr	Glic	Triglic	HDL	SBP	DBP	PCH1
ρ	0,16	0,12	0,35	0,30	0,18	0,13	0,36
$\rho(Y,PCH1)$	-0,23	-0,34	-0,93	-0,28	-0,35	-0,39	

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

Coef. de correlação intraclasse

Componentes Principais em Dados de Famílias

Minicurso: Oficina - R

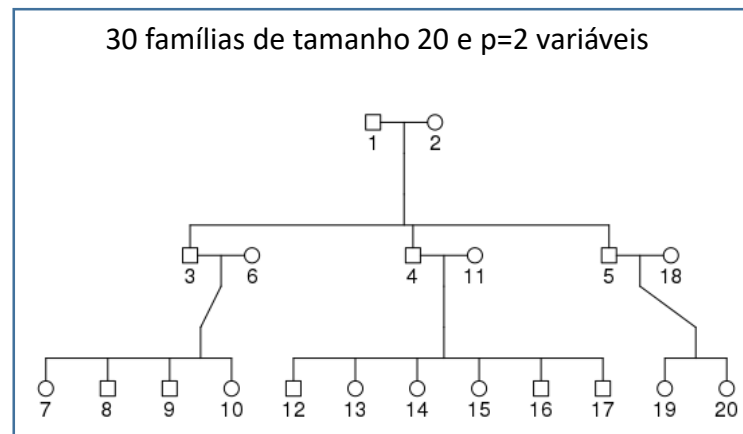
Visualização
do PCH!

- Gerar dados com **estrutura familiar** aleatória e **vetor de médias** e **matrizes de covariância** conhecidas ($p=2$)

$$Y_g \sim \left(1_{n_g} \mu_g'; \Psi_g \otimes \Sigma_B + I_n \otimes \Sigma_W \right)$$

- Obter as estimativas das matrizes de covariância e os seguintes Componentes Principais ($a'Y$):

$$\max_a \frac{a' \hat{\Sigma}_B a}{a' \hat{\Sigma}_W a}$$



Exemplo 1: Σ_g e Σ_g com correlações positivas (0,30 e 0,25, respectivamente) e fenótipos com herdabilidade moderada (0,40 a 0,70)

Exemplo 2: Σ_g e Σ_g com correlações de sinais opostos (0,30 e -0,25, respectivamente) e fenótipos com herdabilidade moderada (0,40 a 0,70)

Exemplo 3: Σ_g e Σ_g com correlações negativas (-0,90 e -0,80, respectivamente) e fenótipos com herdabilidade moderada (0,40 a 0,70)

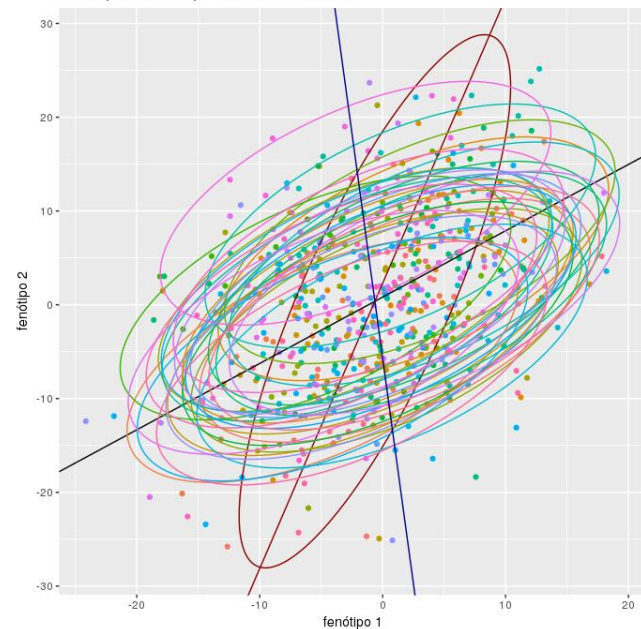
Componentes Principais em Dados de Famílias

Oficina - R

*Visualização
do PCH!*

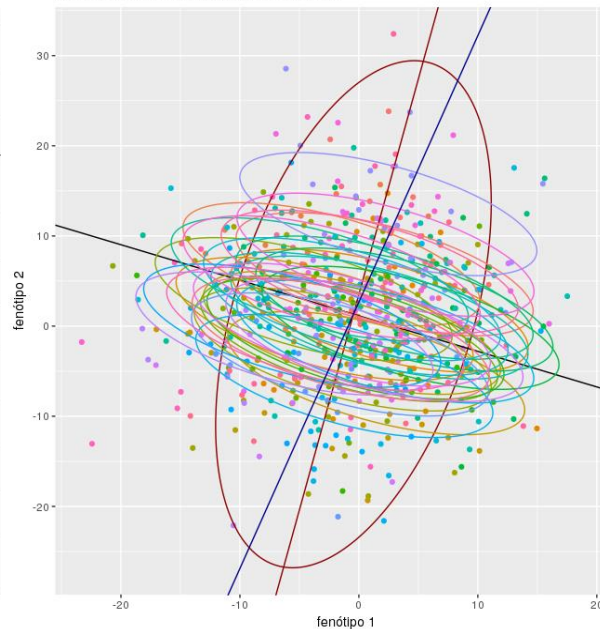
Exemplo 1

Fenótipos com dependência de família



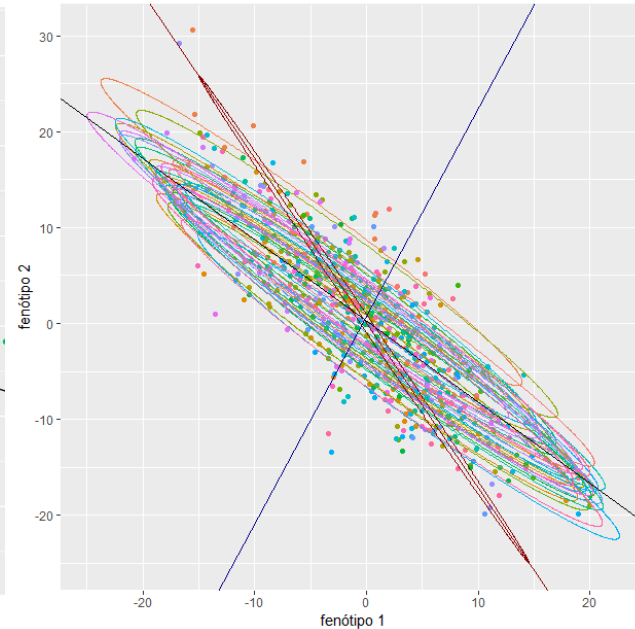
Exemplo 2

Fenótipos com dependência de família



Exemplo 3

Fenótipos com dependência de família



Preto: reta de MQO

Vermelho: CP maximizando a variabilidade ENTRE famílias

Azul: Componente Principal de Herdabilidade (PCH)

Componentes Principais em Dados Correlacionados

Diferentes propostas na literatura!

- **Alternativa 1:** Componentes Principais Discriminantes (PCH)

$$Y_g \stackrel{iid}{\sim} (1_{n_g} \mu'_g; \Omega_g); \quad \Omega_g = \Psi_g \otimes \Sigma_B + I_{n_g} \otimes \Sigma_W; \quad \max_{V_j=a} \frac{a' \hat{\Sigma}_B a}{a' \hat{\Sigma}_W a} \Rightarrow PCH = YV$$

- **Alternativa 2:** Componentes Principais sob Transformação Espectral em $\mathbb{R}^{n \times n}$

$$Y_{n \times p} \sim (\mu_{n \times p}; \Omega_{np \times np} = \Psi_{n \times n} \otimes \Sigma_{p \times p}); \quad \Psi_{n \times n} = U_n \Lambda_n^{1/2} U_n' \Rightarrow Z = U_n' Y$$

$$\text{Cov}(\Lambda_n^{1/2} U_n' Y_{n \times p}) = I_n \otimes \Sigma$$

transformar Y em n
observações
independentes



Obter os Componentes
Principais Clássicos de $\hat{\Sigma}$

n obs correlacionadas
transformadas em n
novas variáveis
independentes!

Análise Multivariada

$$Y_{n \times p} = (Y_{ij}) \in \mathfrak{R}^{n \times p}$$

Já vimos ☺

- Estatísticas Descritivas Multivariadas:
- Distribuição Normal Multivariada:
- Distribuições Amostrais:
- Regiões de Confiança, Testes Multivariados, MANOVA, IC Simultâneos, Correções para Múltiplos Testes
- Análises Multivariadas Clássicas ($n > p$, *iid*): CP, AF, CoP, AC, AD, ACC
- Análises Multivariadas Esparsas ($n \ll p$, *iid*): CP, AD, ACC

Formulações de AD
e ACC em dados
correlacionados!!!

✓ Componentes Principais em Observações Correlacionadas

- Aprendizado de Estruturas – Modelos de Grafos Probabilísticos
Modelos de Equações Estruturais
Propriedades de Markov

Aprendizado de Estruturas

Correlação x Causa

padrão-ouro

- Inferências “Causais” : obtidas de Ensaios Clínicos Controlados e Aleatorizados
- Desafios: - Extrair Inferência Causal de Estudos Observacionais

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

$$X \text{ — } Y$$

$$\begin{array}{l} X \rightarrow Y \\ X \leftarrow Y \end{array}$$



Estabelecer suposições que possam capturar a assimetria das estruturas causais.

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 ?

Ensaio Controlado Aleatorizado
(padrão ouro de inferência causal)

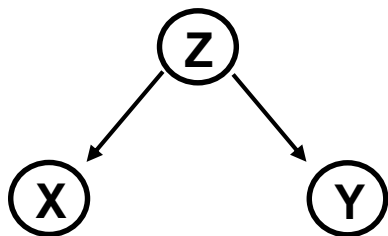
X

Estudos observacionais
(podem conter muitos “ruídos/vícios”)

Avaliar padrões de dependência entre trios de variáveis:

Caso 1

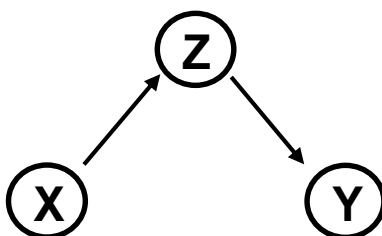
Z é causa comum
Z é confundidor



$(X \perp Y)$ associados

Caso 2

Z é efeito intermediário
X tem efeito indireto sobre Y

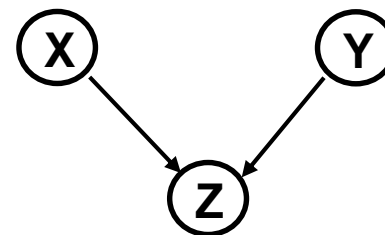


**Modelos Probabilísticos
Equivalentes Grafos Diferentes**

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

Caso 3

Z é efeito comum
(colisão não conectada)
(unshielded collider)



$(X \perp Y)$ independência

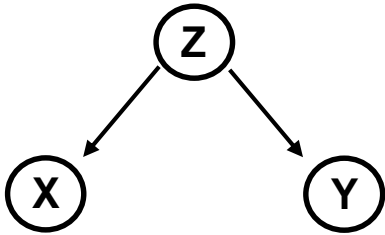
$(X \perp Y | Z)$ dependência condicional

*(X, Y) d-conectados dado Z:
critério de orientação de
arestas*

Aprendizado de Estruturas – Trios de Variáveis

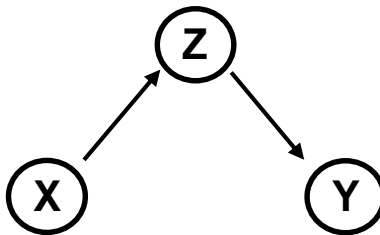
Caso 1

Z é causa comum
Z é confundidor



Caso 2

Z é efeito intermediário
X tem efeito indireto sobre Y



$$(X \perp Y | Z)$$

Independência condicional

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

$$P(X | Z) P(Y | Z) P(Z)$$

$$P(X) P(Z | X) P(Y | Z)$$

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

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

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

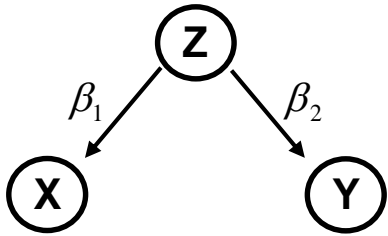
- Condições para que o modelo probabilístico determine unicamente o grafo:
⇒ Teorema da Identificabilidade (ou da descoberta)

Aprendizado de Estruturas – Trios de Variáveis

Correspondência entre Grafos e SEM (Modelos de Equações Estruturais)

Caso 1

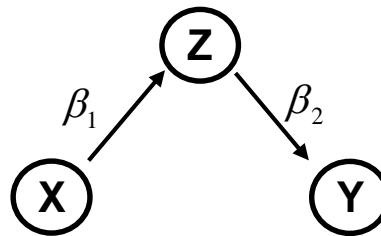
Z é causa comum
Z é confundidor



$$P(X | Z) P(Y | Z) P(Z)$$

Caso 2

Z é efeito intermediário
X tem efeito indireto sobre Y



$$P(X) P(Z | X) P(Y | Z)$$

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:

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

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

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

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

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

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

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

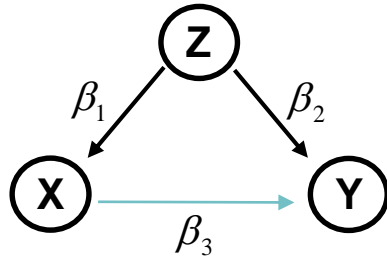
Aprendizado de Estruturas – Trios de Variáveis

Na **correspondência entre SEM e Grafos** é importante que o modelo probabilístico seja **Fiel** ao Grafo.

$$(X \perp Y | Z)$$

Caso 1

Z é causa comum (confundidor)



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

$$Y = \beta_2 Z + \beta_3 X + e_{Y|Z,X}$$

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

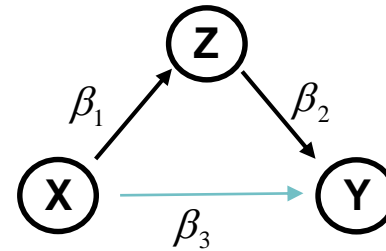
$$Y = \beta_2 Z + \beta_3 (\beta_1 Z + e_{X|Z}) + e_{Y|Z,X}$$

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

$\beta_2 = -\beta_1 \beta_3 \Rightarrow$ 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

$$Y = \beta_2 Z + \beta_3 X + e_{Y|Z,X}$$

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


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

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

$\beta_3 = -\beta_1 \beta_2 \Rightarrow$ 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 Pearl, 1990)
d-separação (estrutura de colisão não-conectada: $X \rightarrow Y \leftarrow Z$)
Fidelidade
 3. **Modelos de Equações Estruturais (SEM)** (Boalen, 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)**: $G=(V,E)$; $E \subset V \times V$

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

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

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

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

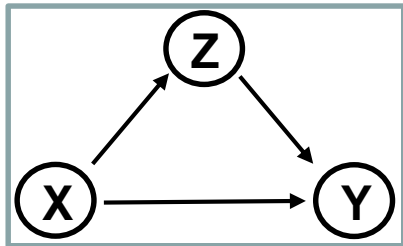
- **Grafo Dirigido (DAG, acíclico):** $X \rightarrow Y \rightarrow Z$

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

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

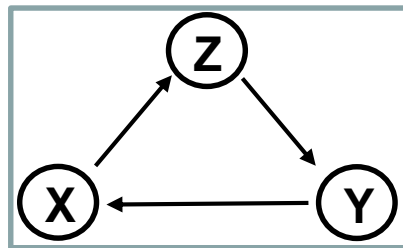
Grafos: Diagramas de Representação “Causal”

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



- 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

- SEM para variáveis “observadas” de um sistema (não há variáveis latentes).

Equações de Mensuração

A correlação entre as p-variáveis está modelada em B

$$Y_{i \ p \times 1} = B_{p \times p} Y_{i \ p \times 1} + X_{i \ p \times q} \beta_{q \times 1} + e_{i \ p \times 1}; \quad e_i \stackrel{iid}{\sim} N_p \left(0; \Sigma = I_p \sigma^2 \right)$$

$B_{p \times p}$: **Matriz de Adjacência** (0 e 1's; $B_{ii}=0$)
Efeitos causais diretos entre as variáveis
 $|I - B| = 1$ para DAGs

*B: Representação
Matricial de Grafos
(coluna causando linha ou
vice-versa)*

Objetivo da inferência: Estimar B e β
Modelo na forma reduzida:

$$Y_i (I - B) = X_i \beta + e_i; \quad Y_i = (I_p - B)^{-1} X_i \beta + (I_p - B)^{-1} e_i$$

*Inferência penalizada da
matriz B – Algoritmo **SML**
(Cai et al., 2013)*

$$\arg \max_{B, \beta} \left\{ n \ln |I - B| - \frac{np}{2} \ln (2\pi\sigma^2) - \frac{1}{2\sigma^2} \|Y(I - B) - X\beta\|_2^2 - \lambda \|B\|_1 \right\}$$

Aprendizado de Estruturas – Inferência Causal

- SEM para variáveis “observadas” de um sistema (não há variáveis latentes).

Equações de Mensuração

$$Y_{i \ p \times 1} = B_{p \times p} Y_{i \ p \times 1} + X_{i \ p \times q} \beta_{q \times 1} + e_{i \ p \times 1}; \quad e_i \stackrel{iid}{\sim} N_p \left(0; \Sigma = I_p \sigma^2 \right)$$

$B_{p \times p}$: **Matriz de Adjacência** (0 e 1's; $B_{ii}=0$)
 $|I - B| = 1$ para DAGs

Modelo
reduzido



$$Y_i = (I_p - B)^{-1} X_i \beta + \boxed{(I_p - B)^{-1} e_i} \quad \text{Mistura de erros independentes}$$

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

Objetivo da inferência: Estimar B e β

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

Aprendizado de Estruturas – Inferência Causal

$p = 5$ variáveis

B matriz de Adjacência: coluna causa linha
coluna \rightarrow linha

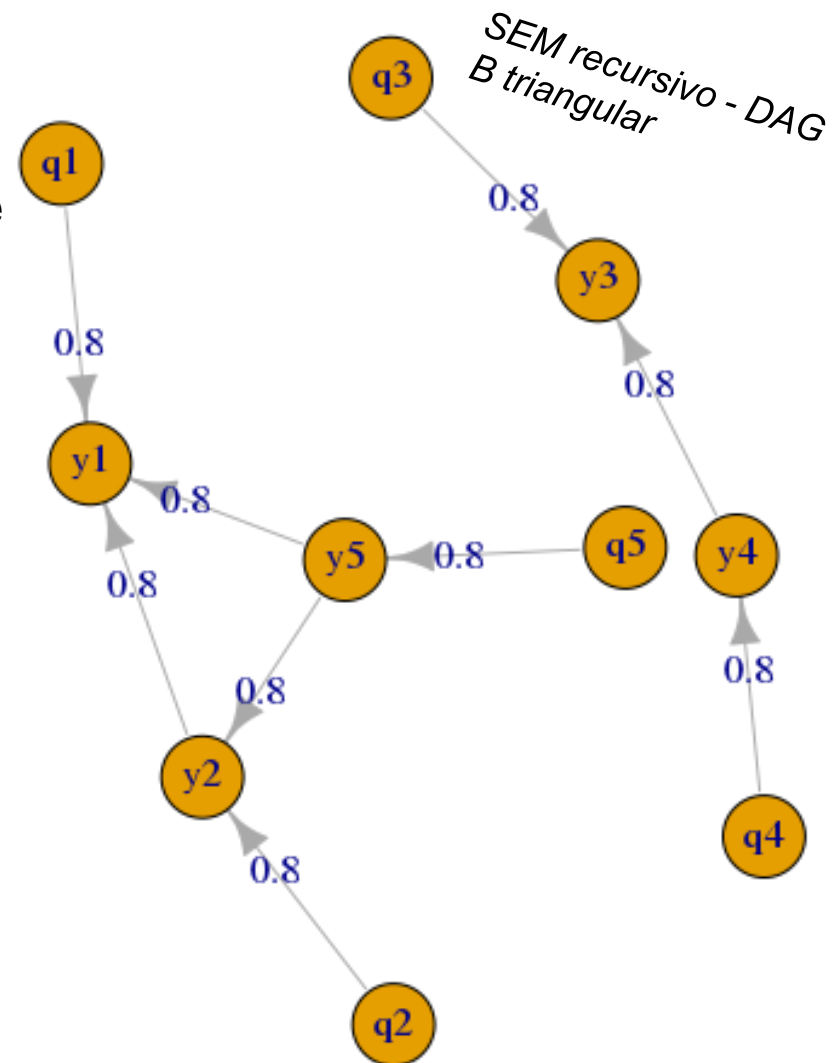
$X=Q$ matriz de efeitos genéticos: cada variável é afetada por um único “Fator Genético”

$$B =$$

	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

$$Q =$$

	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



Algoritmos de Geração de Grafos Aleatórios

- Erdős–Rényi (1959): atribui as arestas aleatoriamente
- Barabási-Albert (1999): segue modelos biológicos

Aprendizado de Estruturas – Inferência Causal

$p=5$ variáveis

B matriz de Adjacência: coluna causa linha
coluna \rightarrow linha

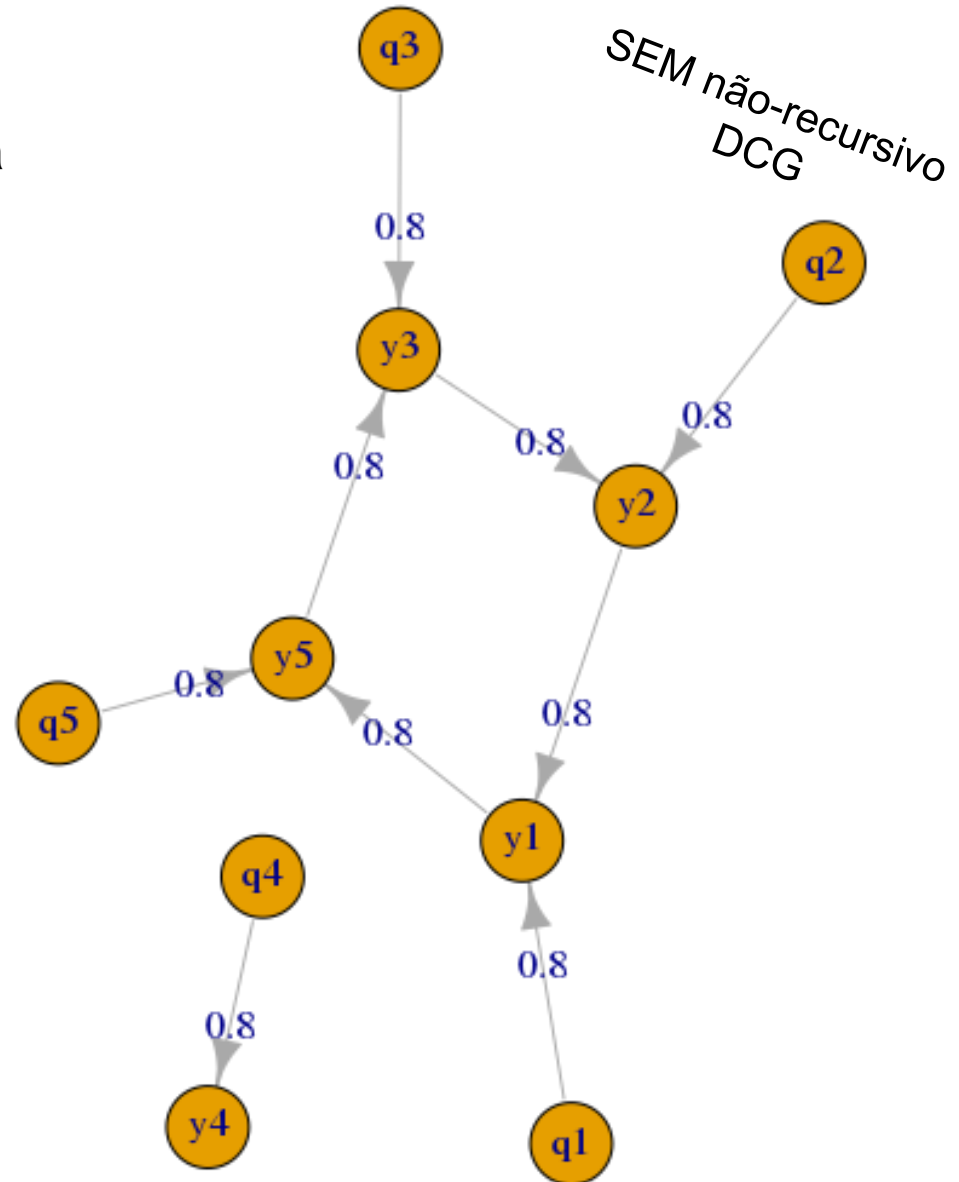
X=Q matriz de efeitos genéticos: cada
variável é afetada por um único “gene”

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

Q =

	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



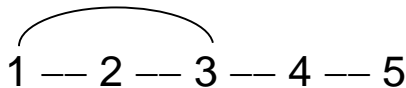
Propriedades de Markov

V : Vértices, $nb(v)$: vizinhança de v , A , B e C subconjuntos disjuntos em V

Independência Condicional para Pares de variáveis: $(X_v \perp X_w \mid X_{V \setminus \{v,w\}})$

Independência Condicional Local: $(X_v \perp X_{V \setminus \{nb(v) \cup v\}} \mid X_{nb(v)})$

Independência Condicional Global: $(X_A \perp X_B \mid X_C)$



M-Pares: $X_1 \perp X_4 \mid (X_2, X_3, X_5)$ $X_1 \perp X_5 \mid (X_2, X_3, X_4)$

$X_2 \perp X_4 \mid (X_1, X_3, X_5)$ $X_2 \perp X_5 \mid (X_1, X_3, X_4)$ $X_3 \perp X_5 \mid (X_1, X_2, X_4)$

M-Local: $X_4 \perp (X_1, X_2) \mid (X_3, X_5)$ $X_5 \perp (X_1, X_2, X_3) \mid X_4$

M-Global: $(X_4, X_5) \perp (X_1, X_2) \mid X_3$ $X_4 \perp X_2 \mid (X_1, X_3)$

*Modelos Gaussianos:
avaliação de Σ^{-1}*

*Modelos Gaussianos: a variável
pode ser predita de seus
vizinhos (regressão LASSO)*

*Distribuições conjuntas
decomponíveis: avaliar fatorações
da densidade*

Grafos Dirigidos (DAGs)

Conceito de d-Separação

- Definição: Critério de d-separação (Pearl, 1988, 2000)

$(X \perp Y | Z)_G$: X e Y estão d-separados dado Z se para qualquer caminho (aresta) U entre X e Y,

- ✓ U não contém colisão não conectada, e então $Z \in U$ (Z é efeito intermediário ou causa comum), ou
- ✓ U contém uma colisão não-conectado, e então nem o colisor nem seus descendentes $\in Z$.

$X \rightarrow Z \rightarrow Y$ (efeito intermediário)

$$(X \perp Y | Z)_G$$

X e Y estão “d-separados” dado Z

$$X \rightarrow Z \leftarrow Y$$

(efeito comum, colisão não conectada)

$$(X \not\perp Y | Z)_G$$

X e Y estão “d-conectados” dado Z

Aprendizado de Estruturas – Inferência Causal

- Notação:

$(X \perp Y | Z)_P$: X é **condicionalmente independente** de Y dado Z sob P

$$P(X = x | Y = y, Z = z) = P(X = x | Z = z); \quad P(Y = y, Z = z) > 0$$

$(X \perp Y | Z)_L$: X é **condicionalmente independente** de Y dado Z em um SEM
(linear com erros iid)

$(X \perp Y | Z)_{UDG}$: X é **separado** de Y dado Z em um Grafo não direcionado UDG

$(X \perp Y | Z)_G$: X é **d-separado** de Y dado Z em um Grafo direcionado G (DAG ou DCG)

$(X \perp Y Z)_G$	M-Global \Rightarrow \Leftarrow Fidelidade	$(X \perp Y Z)_P ; (X \perp Y Z)_L$ L satisfazendo $\rho_{XY Z} = 0$
---------------------	---	--

Resultados bem estabelecidos para os Modelos de Grafos Gaussianos
(sob distribuição conjunta Normal Multivariada e observações independentes,
estruturas podem ser aprendidas via testes de Independência Condicional)

Aprendizado de Estruturas – Inferência Causal

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

- **Passo 1.** Iniciar com arestas entre todas as variáveis
- **Passo 2.** Aprendizado do Grafo Não Direcionado (**UDG** - relações simétricas)

Eliminar
arestas

Testes de independência condicional entre **Pares de variáveis dado as restantes**

$$X \perp Y | V \setminus \{X, Y\}$$



Inferir o Esqueleto do Grafo

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

No UDG algumas variáveis podem estar d-conectadas devido à **V-estrutura** $X \rightarrow Z \leftarrow Y$

Avaliar todos os trios de variáveis e orientar os casos de colisão

Passo 2 →

Modelos de Grafos Gaussianos

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

Testes da Correlação Parcial Nula

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

$$\text{Modelo Normal Multivariado: } V_i \stackrel{iid}{\sim} N_p(\mu_{p \times 1}; \Sigma_{p \times p})$$

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

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

Teste da Correlação Parcial Nula: Estatística z de Fisher

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

Situações mais gerais: **n?**

Modelos de Grafos Gaussianos

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

$$\Sigma_{p \times p} = \begin{pmatrix} \sigma_{XX} & \sigma_{XY} & \Sigma_{XS} \\ \sigma_{YX} & \sigma_{YY} & \Sigma_{YS} \\ \Sigma_{SX} & \Sigma_{SY} & \Sigma_{SS} \end{pmatrix} \xrightarrow{\text{Distr. Condicional}} \Sigma_{XY.S} = \begin{pmatrix} \sigma_{XX.S} & \sigma_{XY.S} \\ \sigma_{YX.S} & \sigma_{YY.S} \end{pmatrix}$$

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

Modelos Univariados:

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

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

$$= \frac{\text{Cov}(e_{X|S}; e_{Y|S})}{\sigma_{X|S} \sigma_{Y|S}} = \beta_{X|S}^Y \frac{\sigma_{X|S}}{\sigma_{Y|S}}$$

Método das Vizinhanças(Meinshausen and Buhlmann, 2006)

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

Incluir arestas se: $\hat{E} = \{(X, Y); X \in nb(Y) \wedge \vee Y \in nb(X)\}$ Critério AND/OR

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

3.2. Pesquisar por Z; $(X \perp Y|Z)$

3.3. Se não existir Z \Rightarrow $X \rightarrow Z \leftarrow Y$



Algoritmo IC
Algoritmo PC

} Implementar um teste de independência condicional
Oráculo da d-separação

Simulação de Grafos: “UDG”

Adèle Ribeiro, 2018

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

Uso do R

```
#Gerar dados N3
```

```
#Aprender a estrutura
```

```
#Estimar as arestas via Método das Vizinhanças
```

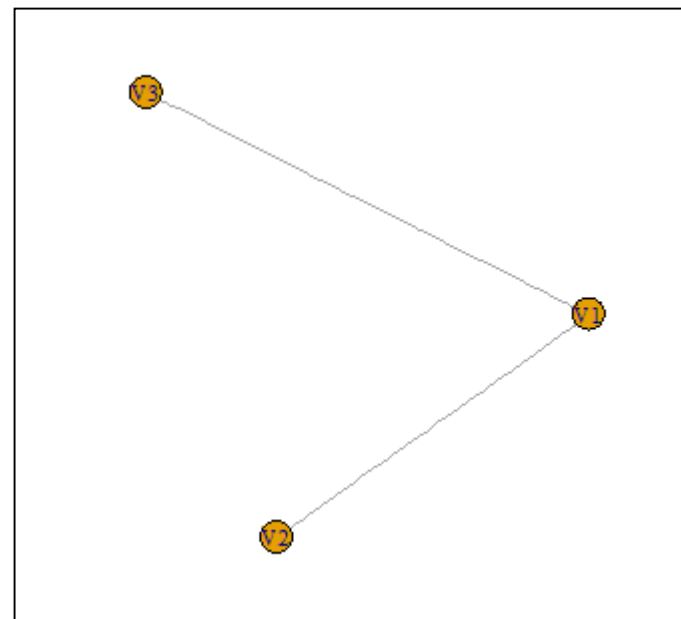
```
#Teste z de Fisher
```

```
> 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



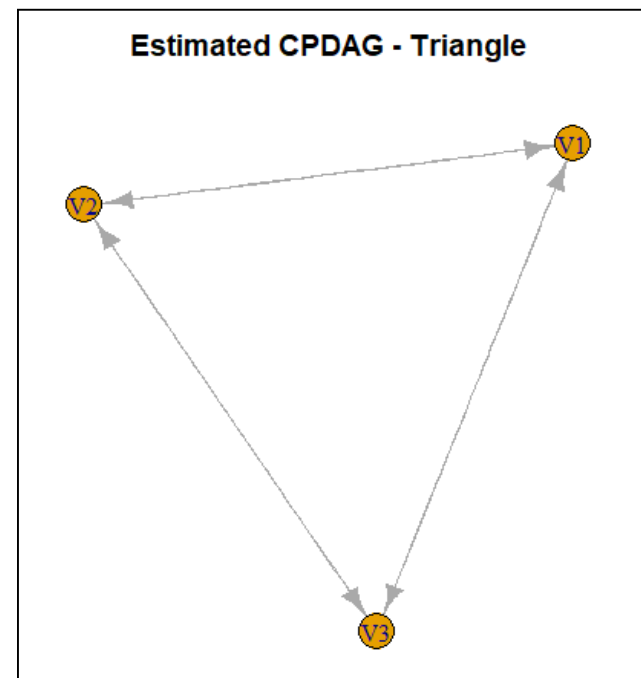
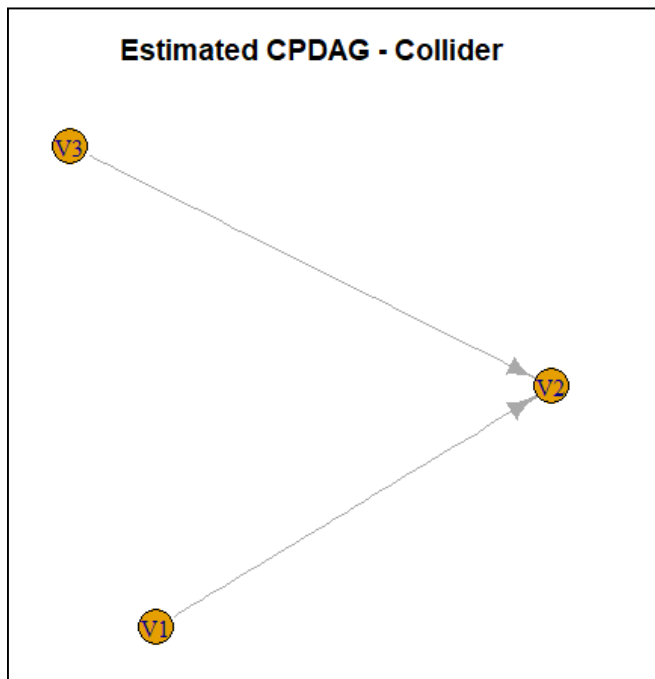
Simulação de Grafos: “DAG”

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
```

```
#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
```



Simulação de Grafos: DAG

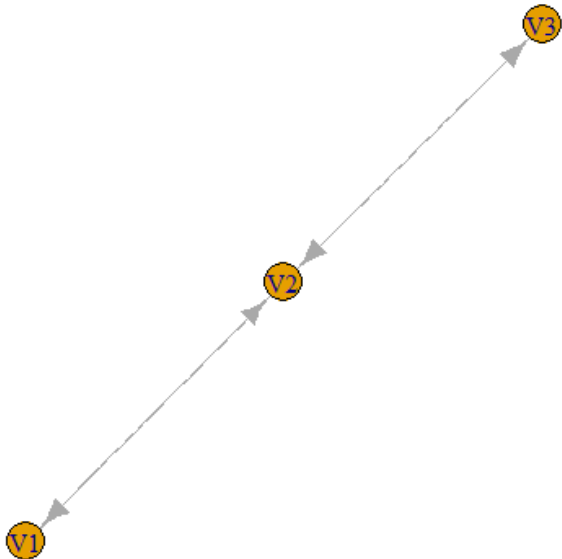
p=3 Variáveis

Adèle Ribeiro, 2018

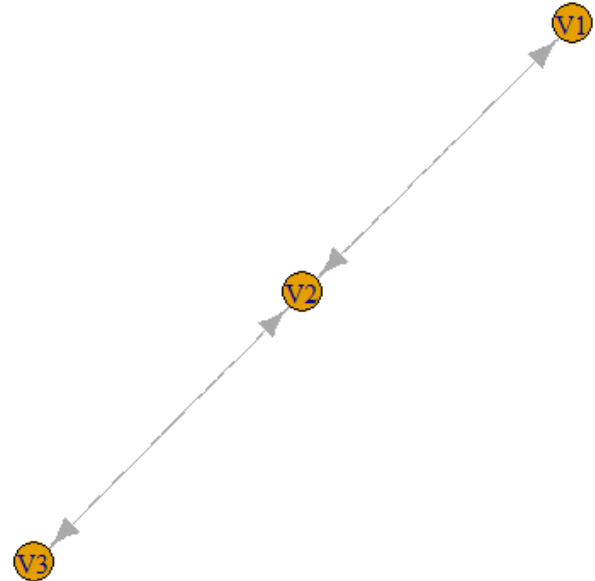
```
# Possible types:  
# - chain: V1 -> V2 -> V3  
# - collider: V1 -> V2 <- V3  
# - fork: V1 <- V2 -> V3  
# - triangle: V1 -> V2 <- V3 <- V1  
# - independent: V1 V2 V3
```

Chains and forks have the same joint distribution
they cannot be statistically distinguished from each other.

Estimated CPDAG - Fork



Estimated CPDAG - Chain



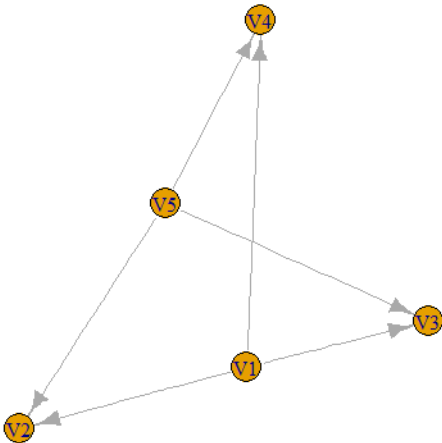
Simulação de Grafos: “DAG”

p=5 Variáveis

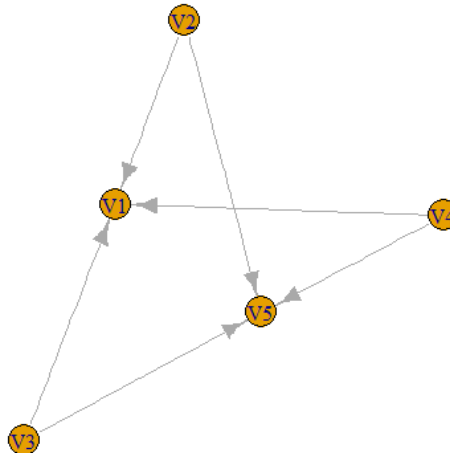
Adèle Ribeiro, 2018

```
# 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.
```

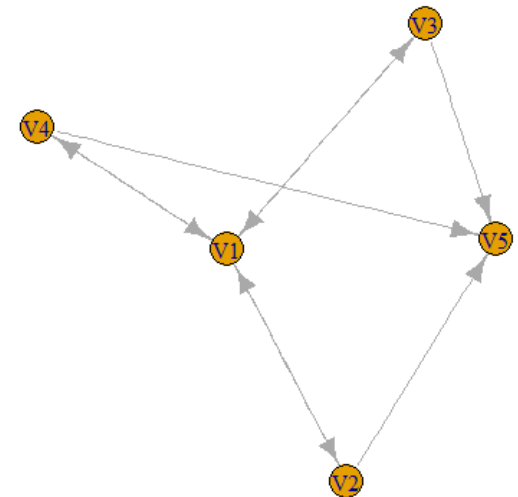
Estimated CPDAG - Multicollider



Estimated CPDAG - Multifork



Estimated CPDAG - Multichain



Aprendizado de Estruturas em Observações Correlacionadas

$$\Rightarrow Y_{np}; Y_{g \ n_g \times p} \sim \left(1_{n_g} \otimes \mu'_g; \Psi_{g \ n_g \times n_g} \otimes \Sigma_{B_{p \times p}} + I_{n_g} \otimes \Sigma_{W_{p \times p}} \right); \quad \Sigma_T = \Sigma_B + \Sigma_W$$

Aprendizado de **COMPONENTES de Covariância do Grafo**

Passo 2. Aprendizado dos Grafos UDG (não orientados)

Alternativa 2.1. Testes de Correlação Parcial Nula a partir do Ajuste do Modelo de Componentes de Covariância

$\Sigma_T = \Sigma_B + \Sigma_W$; $\hat{\Sigma}_T, \hat{\Sigma}_B, \hat{\Sigma}_W$ Estimadores MANOVA
(Oualkacha et al., 2012)

Como obter testes válidos de correlação parcial? $\rho(X, Y | S) = \frac{\sigma_{XY.S}}{\sqrt{\sigma_{XX.S} \sigma_{YY.S}}} = \frac{-k_{XY}}{\sqrt{k_{XX} k_{YY}}}$

Matrizes de Precisão

Grafo do Efeito de Grupos $\rightarrow \hat{\Sigma}_B^{-1}$

Grafo Residual (Dentro de Grupo) $\rightarrow \hat{\Sigma}_W^{-1}$

Grafo Total (Efeito Combinado) $\rightarrow \hat{\Sigma}_T^{-1}$

Limitação: Estabelecer o tamanho amostral efetivo (n_e) na formulação da estatística z de Fisher.

Solução proposta: uso do denominador das estimativas MANOVA de cada matriz de covariância.

Aprendizado de Estruturas em Observações Correlacionadas

Alternativa 2.2. Teste de Correlação Parcial Nula a partir do Ajuste de Modelos de Componentes de Covariâncias

$$Y_{n \times 1} = \mu + \beta_S' S + \boxed{g + e}; \quad g \sim N_n(0; \Psi_g \sigma_B^2); \quad e \sim N_n(0; I_n \sigma_W^2);$$

Efeitos
aleatórios

$$\text{Cov} \begin{bmatrix} g \\ e \end{bmatrix} = \begin{pmatrix} \Psi_{n \times n} \sigma_B^2 & 0_{n \times n} \\ 0_{n \times n} & I_n \sigma_W^2 \end{pmatrix}; \quad \Psi = I_G \otimes \Psi_g$$

Ajuste de Modelos Univariados de Componentes de variância

Como obter testes válidos de correlação parcial?

$$\rho(X, Y | S) = \frac{\sigma_{XY.S}}{\sqrt{\sigma_{XX.S} \sigma_{YY.S}}} = \frac{\text{Cov}(e_{X|S}; e_{Y|S})}{\sigma_{X|S} \sigma_{Y|S}}$$

$$Y_{n \times 1} = \mu_{Y|S} + \beta_S^Y' S + g_{Y|S} + e_{Y|S}; \quad g_{Y|S} \sim N(0; \Psi \sigma_{B Y|S}^2); \quad e_{Y|S} \sim N(0; \Psi \sigma_{W Y|S}^2)$$

$$X_{n \times 1} = \mu_{X|S} + \beta_S^X' S + g_{X|S} + e_{X|S}; \quad g_{X|S} \sim N(0; \Psi \sigma_{B X|S}^2); \quad e_{X|S} \sim N(0; \Psi \sigma_{W X|S}^2)$$

Aprendizado de Estruturas em Observações Correlacionadas

Alternativa 2.2. Teste de Correlação Parcial Nula a partir do Ajuste de Modelos de Componentes de Covariâncias

$$Y_{n \times 1} = \mu + \beta_S' S + g + e; \quad g \sim N_n(0; \Psi_g \sigma_B^2); \quad e \sim N_n(0; I_n \sigma_W^2);$$

$$\text{Cov} \begin{bmatrix} g \\ e \end{bmatrix} = \begin{pmatrix} \Psi_{n \times n} \sigma_B^2 & 0_{n \times n} \\ 0_{n \times n} & I_n \sigma_W^2 \end{pmatrix}; \quad \Psi = I_G \otimes \Psi_g$$

Como obter testes válidos de correlação parcial para cada componente (e e g)?

Ajuste de Modelos Univariados de Componentes de variância

- Erro Marginal: $\xi = Y - E(Y) = g + e \longrightarrow \hat{\rho}(X, Y | S)_T = \frac{\text{Cov}(\hat{\xi}_S^X; \hat{\xi}_S^Y)}{\sqrt{\text{Var}(\hat{\xi}_S^X) \text{Var}(\hat{\xi}_S^Y)}}$
(Talluri and Shete, 2014)
- Erro Condicional: $e = Y - E(Y | g) \longrightarrow \hat{\rho}(X, Y | S)_W = \frac{\text{Cov}(\hat{e}_S^X; \hat{e}_S^Y)}{\sqrt{\text{Var}(\hat{e}_S^X) \text{Var}(\hat{e}_S^Y)}}$
(Valente et al., 2010 sob SEM)
- Efeito Aleatório: $g = E(Y | g) - E(Y) \longrightarrow \hat{\rho}(X, Y | S)_B = \frac{\text{Cov}(\hat{g}_S^X; \hat{g}_S^Y)}{\sqrt{\text{Var}(\hat{g}_S^X) \text{Var}(\hat{g}_S^Y)}}$

Aprendizado de Estruturas em Observações Correlacionadas

Como obter testes válidos de correlação parcial para cada Componente (e e g)?

- Componente de Erro Condicional:

$$\hat{\rho}(X, Y | S)_w = \frac{Cov(\hat{e}_S^X; \hat{e}_S^Y)}{\sqrt{Var(\hat{e}_S^X)Var(\hat{e}_S^Y)}}$$

- Efeito Aleatório:

$$\hat{\rho}(X, Y | S)_B = \frac{Cov(\hat{g}_S^X; \hat{g}_S^Y)}{\sqrt{Var(\hat{g}_S^X)Var(\hat{g}_S^Y)}}$$

Limitação: Resíduo condicional $(\hat{e}_S^X, \hat{e}_S^Y)$ e o efeito aleatório predito $(\hat{g}_S^X, \hat{g}_S^Y)$ **estão confundidos.**

Solução para reduzir o confundimento nas estimativas: uso do procedimento proposto por Loy and Hoffman(2015) → obter uma aproximação de baixa dimensão, para os efeitos aleatórios rotacionados bem como para os resíduos condicionais, que minimize o confundimento.

Grafo-Entre Grupos

Grafo Residual

Grafo Total

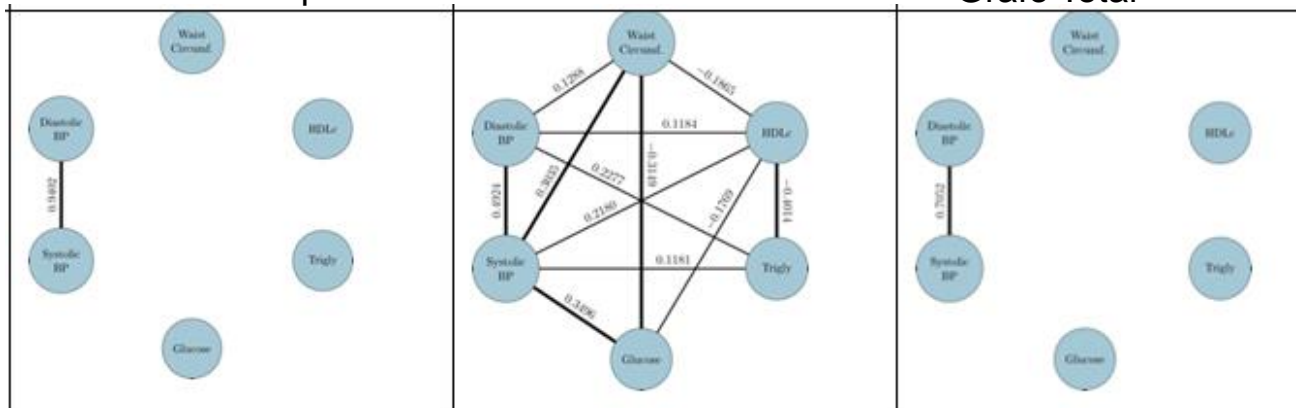
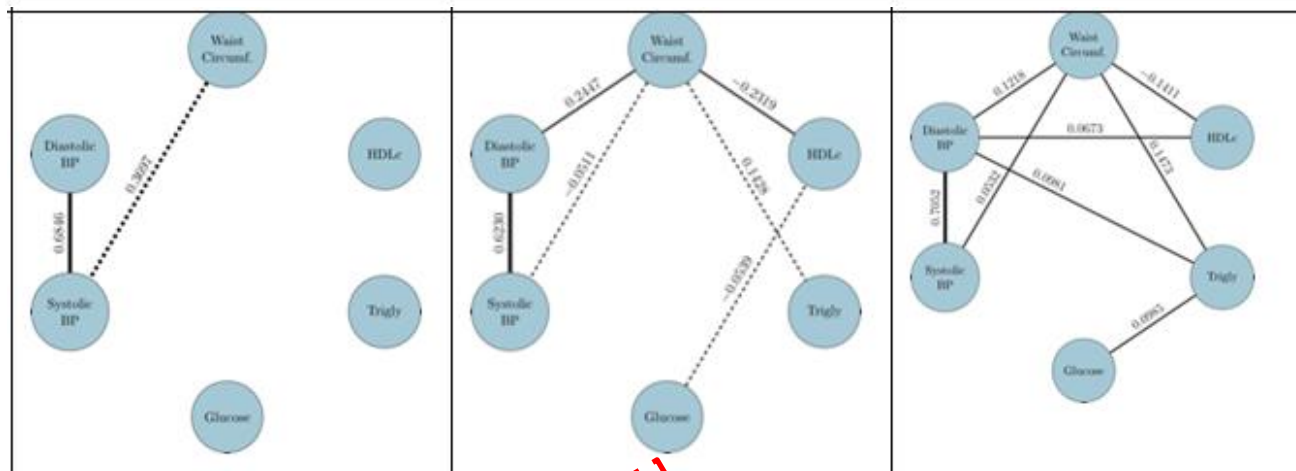
Aplicação

Projeto Corações de
Baependi (MG)

SubGrafos (**UDG** e
DAG) das variáveis da
Síndrome Metabólica:

Circunferência Abdominal
Glicose
Triglicérides
HDL
SBP: pressão sistólica
DBP: pressão diastólica

Relação causal conhecida
(validada): Circunferência
abdominal como causa
direta da pressão

Modelo Multivariado
Matriz de PrecisãoModelo Univariados
Soluções Rotacionadas

Algoritmo PC

