aula modelagem adicionais e Introducao ao pacote lme4: parte 2

Prof. Caio Azevedo

Voltando a exemplos apresentados anteriormente

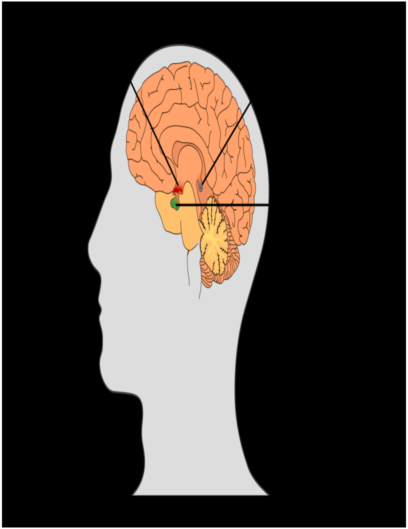
- Como temos visto, estruturas hierárquicas são uma constante na resolução de problemas aqui, aqui.
- Podemos ter tanto UAE's agrupadas em outras UAE's (Exemplos 1, 2, 3) bem como medidas aninhadas (Exemplos: 4 e 5) aqui.
- Estruturas hierárquicas podem estar associadas (induzir) heterocedasticidades, dependência, não normalidade das variáveis (respostas, explicativas), relação não lineares entre a(s) resposta(s) e as informações colateriais etc.
- Consideremos um conjunto de dados longitudinais.

Exemplo 5: Distância do centro da glândula pituitária para a fissura pterigomaxilar (Potthoff and Roy (1964))

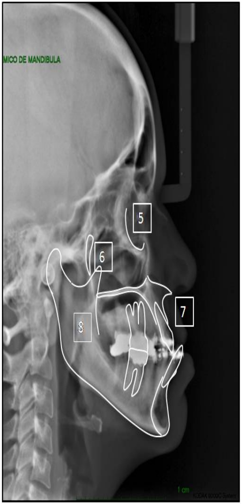
- Esta pesquisa está relacionada aos famosos dados de Potthoff-Roy, usados para demonstrar a utilização da MANOVA em análise de medidas repetidas (comparação entre grupos, embora comparação entre variáveis seja possível).
- O estudo considerou 16 meninos e 11 meninas, nos quais, nas idades 8, 10, 12 e 14 anos tiveram a distância (mm) do centro da glândula pituitária para a fissura pterigomaxilar medidas.

Exemplo 5 (Cont.)

- Mudanças nas distâncias pituitária-pterigomaxilar durante o crescimento são importantes na terapia ortodôntica.
- Os objetivos do estudo foram descrever a distância em função da idade e comparar esse desenvolvimento (temporal) entre sexos.
- Nível 1 (medida repetida ao longo dos anos), Nível 2 (crianças).
- Disponível no programa R, pacote "mice" sob o nome de "potthoffroy".



- 5 – Borda inferior da órbita
- 6 – Fissura pterigomaxilar
- 7 – Maxila
- 8 – Mandíbula
- 9 – Dentes (incisivos e 1º. molares)



Banco de dados (multivariado)

		idade			
Indivíduo	Sexo	8	10	12	14
1	Feminino	21,0	20,0	21,5	23,0
2	Feminino	21,0	21,5	24,0	25,5
⋮	⋮	⋮	⋮	⋮	⋮
11	Feminino	24,5	25,0	28,0	28,0
1	Masculino	26,0	25,0	29,0	31,0
2	Masculino	21,5	22,5	23,0	26,5
⋮	⋮	⋮	⋮	⋮	⋮
16	Masculino	22,0	21,5	23,5	25,0

Banco de dados (longitudinal/hierárquico)

Indivíduo	Sexo	Idade	Distância
1	Feminino	8	21,0
1	Feminino	10	20,0
1	Feminino	12	21,5
1	Feminino	14	23,0
⋮	⋮	⋮	⋮
16	Masculino	8	22,0
16	Masculino	10	21,5
16	Masculino	12	23,5
16	Masculino	14	25,0

Medidas resumo

sexo	ano	média	dp	var	cv(%)	min.	med.	max.	ca	curt.	n
Fem.	8	21,18	2,12	4,51	10,03	16,50	21,00	24,50	-0,56	3,49	11
	10	22,23	1,90	3,62	8,56	19,00	22,50	25,00	-0,14	1,97	11
	12	23,09	2,36	5,59	10,24	19,00	23,00	28,00	0,36	3,19	11
	14	24,09	2,44	5,94	10,12	19,50	24,00	28,00	-0,26	2,44	11
Masc.	8	22,88	2,45	6,02	10,72	17,00	23,00	27,50	-0,37	3,69	16
	10	23,81	2,14	4,56	8,97	20,50	23,50	28,00	0,45	2,41	16
	12	25,72	2,65	7,03	10,31	22,50	25,00	31,00	0,90	2,74	16
	14	27,47	2,09	4,35	7,59	25,00	26,75	31,50	0,65	2,20	16

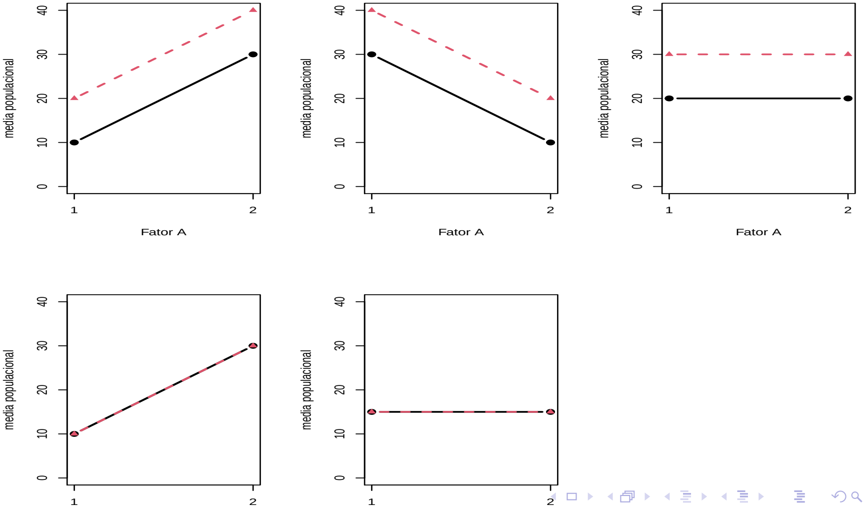
Um pouco sobre interação

- Em muitas situações, o pesquisador tem interesse em como dois ou mais fatores afetam o comportamento da variável resposta.
- Nem todos os fatores são, necessariamente, de interesse. Contudo, em princípio, todos devem ser controlados de alguma forma.
- Em nosso exemplo temos um fator intra-unidades (ano) e um entre-unidades (sexo).
- Pode haver fatores que funcionam como "bloco" (como na literatura de planejamento de experimentos).
- Nos próximos dois slides: linha preta - nível 1, fator B, linha vermelha - nível 2 fator B.

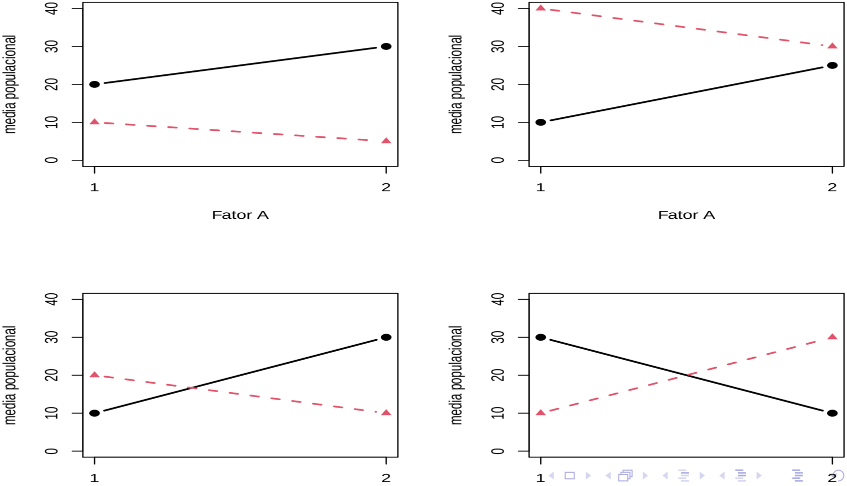
Exemplo

- Fator A: possui a níveis.
- Fator B: possui b níveis.
- Um deles pode ser quantitativo (não fator).
- Conceito importante: interação entre os fatores.
- Interação: a diferença entre as médias da resposta, entre dois níveis do Fator A, são iguais ao longo dos níveis do Fator B (vice-versa).
- Se uma das covariáveis não for um fator (as curvas em relação aos níveis do fator têm de ser paralelas).

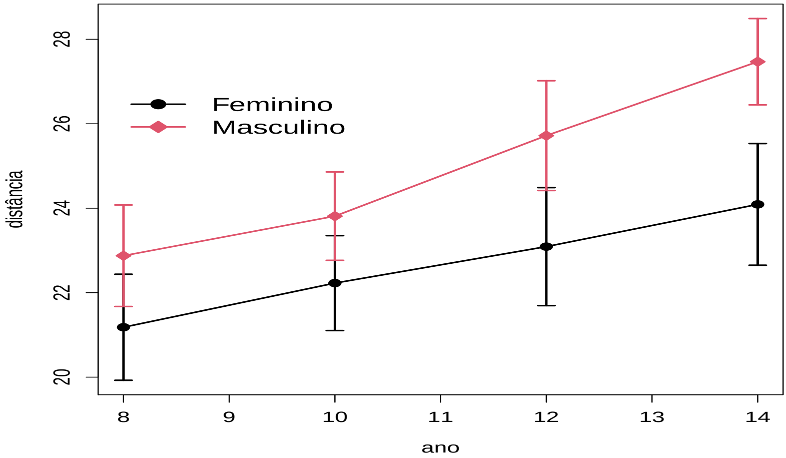
Perfis médios hipotéticos: ausência de interação



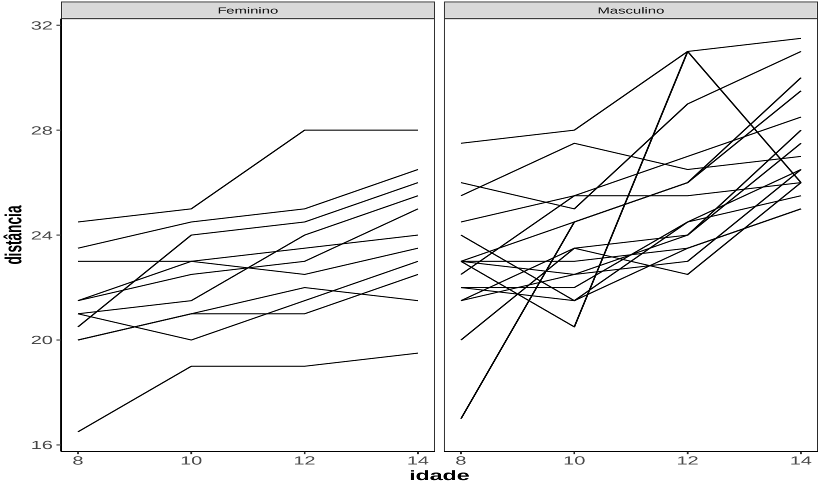
Perfis médios hipotéticos: presença de interação



Perfis médios observados



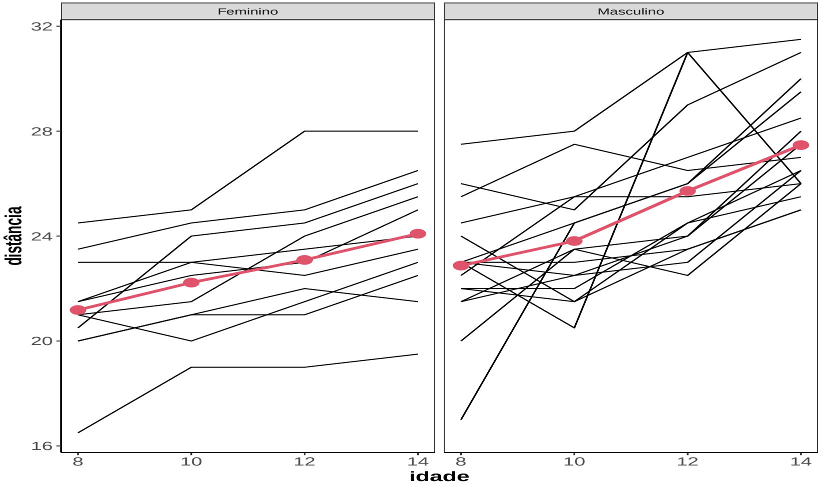
Perfis individuais observados



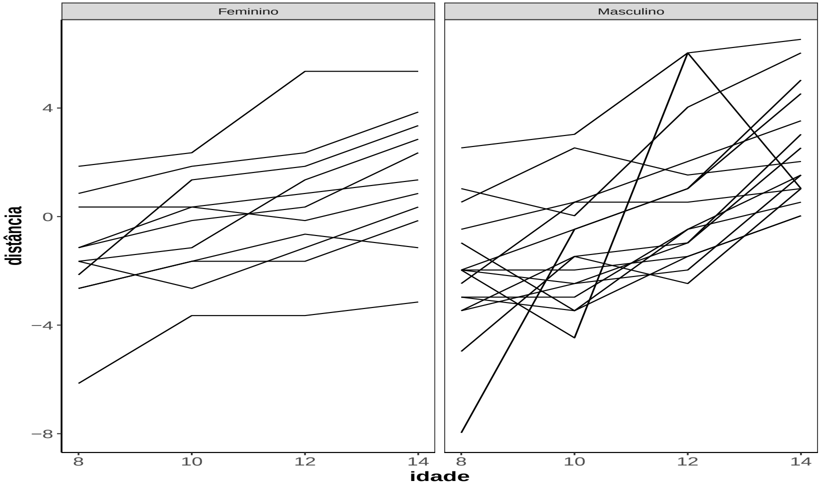
Perfis individuais observados (seperados)

link

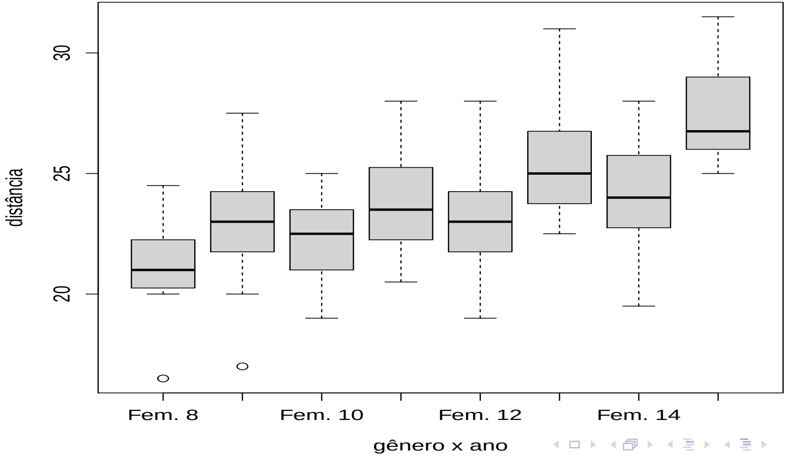
Perfis individuais e médios observados



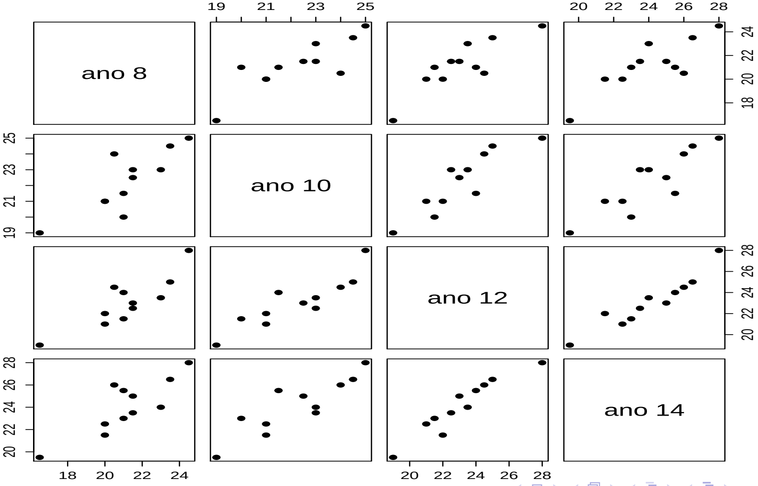
Perfis individuais centrados observados



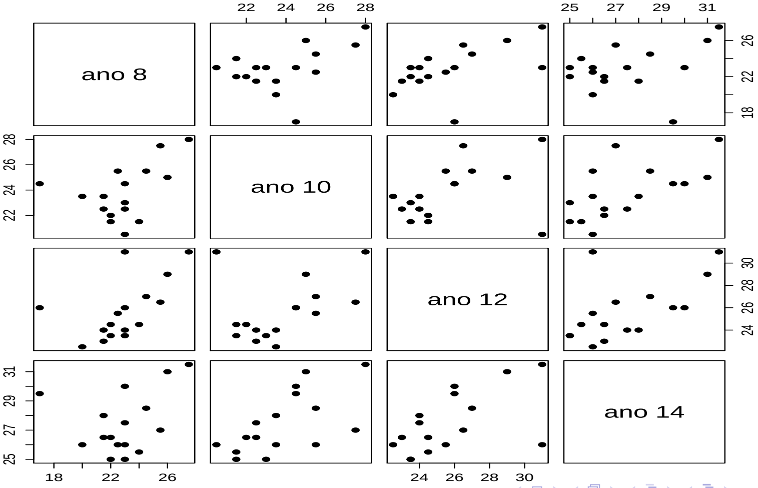
Box plot



Matriz de diagramas de dispersão: sexo feminino



Matriz de diagramas de dispersão: sexo masculino



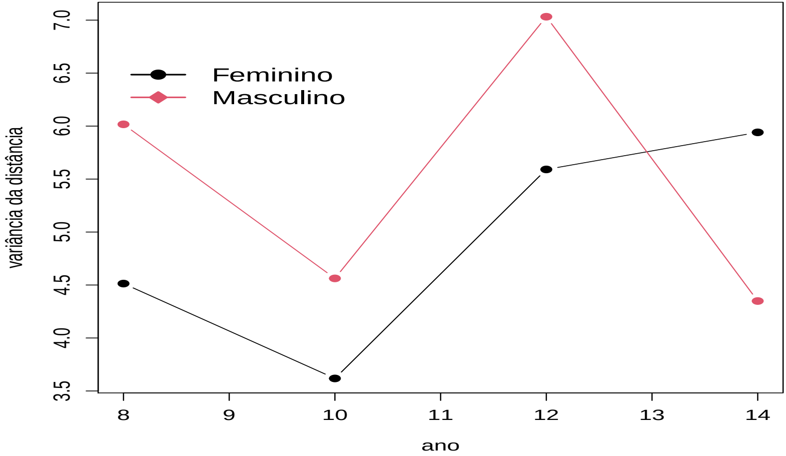
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo): sexo feminino

Ano			
8	10	12	14
4,51	0,83	0,86	0,84
3,35	3,62	0,90	0,88
4,33	4,03	5,59	0,95
4,36	4,08	5,47	5,94

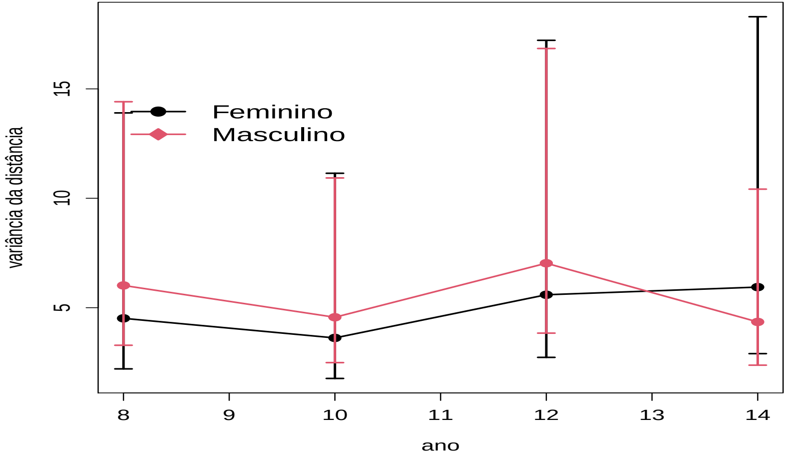
Variâncias (diagonal), correlações (acima) e covariâncias (abaixo): sexo masculino

Ano			
8	10	12	14
6,02	0,44	0,56	0,32
2,29	4,56	0,39	0,63
3,63	2,19	7,03	0,59
1,61	2,81	3,24	4,35

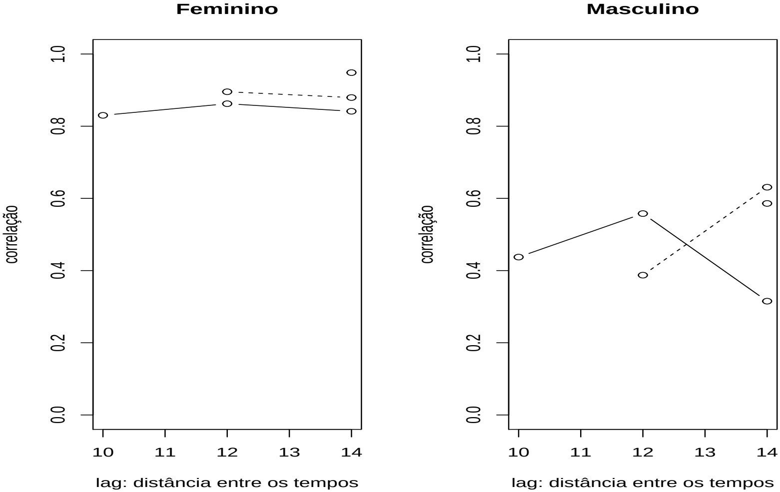
Variâncias ao longo dos anos



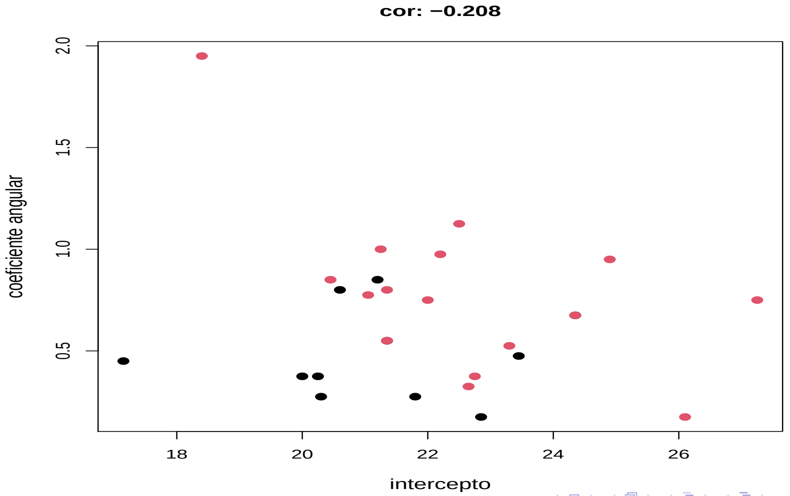
Variâncias em cada condição com intervalos de confiança



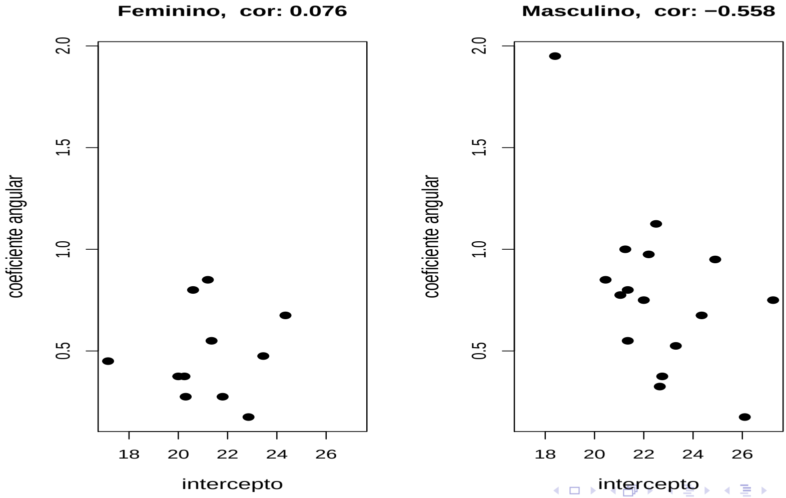
Gráficos dos perfis das linhas da matriz de correlações



Interceptos e coeficientes angulares (regressões individuais)



Interceptos e coeficientes angulares (regressões individuais)



Um modelo geral de dois níveis

- Neste caso, o modelo (doravante, modelo 1) é dado por:

$$
\begin{aligned}
Y_{jki} &= \beta_{0jk} + \beta_{1jk}(x_{jki} - 8) + \xi_{jki}, j = 1, 2, ..., n_k, \text{(indivíduos)} \\
& \text{(sexo)} k = 1\text{(feminino)}, 2\text{(masculino)}; \\
& i = 1, 2, 3, 4 \, \text{(nível 1 - medidas repetidas)} \\
\beta_{0jk} &= \gamma_{000} + \gamma_{00k} + u_{0jk} \, , \gamma_{001} = 0 \, \text{(nível 2 - indivíduos)} \\
\beta_{1jk} &= \gamma_{10k} + u_{1jk} \, \text{(nível 2 - indivíduos)}
\end{aligned}
$$

- Erros e efeitos aleatórios: $\xi_{jki} \overset{iid}{\sim} N(0, \sigma^2)$, $\boldsymbol{u}_{jk} = (u_{0jk}, u_{1jk})' \overset{iid}{\sim} N_2(0, \boldsymbol{\Psi})$, $\xi_{jki} \perp \boldsymbol{u}_{jk}, \forall i, j$ e $(\gamma_{00k}, \gamma_{10k})'$ são não aleatórios, $\boldsymbol{\Psi} = \begin{bmatrix} \psi_{00} & \psi_{01} \\ \psi_{01} & \psi_{11} \end{bmatrix}$

Modelo 1

- Y_{jki} : é a distância do indivíduo j, do sexo k no ano (instante) i.
- $x^*_{jki} = x_{jki} - 8$: é o valor do ano (idade) no instante i, do indivíduo j do sexo k, menoso valor 8 (ano basal).
- (momentos condicionais)
  $\mathcal{E}(Y_{jki}|\boldsymbol{u}_{jk}) = \gamma_{000} + \gamma_{00k} + u_{0jk} + (\gamma_{10k} + u_{1jk})x^*_{jki}$ - valor esperado da distância para indivíduo j, do sexo k no ano i com idade igual a x_{jki}. $\mathcal{V}(Y_{jki}|\boldsymbol{u}_{jk}) = \sigma^2$.
- (momentos marginais) $\mathcal{E}(Y_{jki}) = \gamma_{000} + \gamma_{00k} + \gamma_{10k}x^*_{jki}$ - valor esperado da distância para indivíduos do sexo k no ano i com idade igual a x_{jki}. $\mathcal{V}(Y_{jki}) = \psi_{00} + \psi_{11}x^{*2}_{jki} + \psi_{01}x^*_{jki} + \sigma^2$.

Modelo 1

- $Cov(Y_{jki}, Y_{jki'}) = \psi_{00} + \psi_{11}x^*_{jki}x^*_{jki'} + \psi_{01}x^*_{jki} + \psi_{01}x^*_{jki'}$, $i \neq i'$.
- $Corre(Y_{jki}, Y_{jki'}) = \frac{\psi_{00}+\psi_{11}x^*_{jki}x^*_{jki'}+\psi_{01}x^*_{jki}+\psi_{01}x^*_{jki'}}{\sqrt{\psi_{00}+\psi_{11}x^{*2}_{jki}+\psi_{01}x^*_{jki}+\sigma^2}\sqrt{\psi_{00}+\psi_{11}x^{*2}_{jki'}+\psi_{01}x^*_{jki'}+\sigma^2}}$
- γ_{000} : distância esperada para indivíduos do sexo feminino no ano 8.
- γ_{002} : diferença da distância esperada entre indivíduos do sexo masculino e do sexo feminino no ano 8.
- γ_{101} : incremento (positivo ou negativo na distância) na distância, para indivíduos do sexo feminino. para o aumento em uma no da idade.
- γ_{102} : incremento (positivo ou negativo na distância) na distância, para indivíduos do sexo masculino, para o aumento em uma no da idade.

Modelo 1

- $u_{0jk} = \mathcal{E}(Y_{jki}|\boldsymbol{u}_j, x_{jki} = 8) - \mathcal{E}(Y_{jki}|x_{jki} = 8)$ é a diferença entre o valor esperado para o indivíduo j, do sexo k, no instante i e a esperança global (todos os indivíduos do sexo k no instante i), com idade $x_{jki} = 8$. Outra interpretação: é o quanto o intercepto do indivíduo j se difere do intercepto comum a todas os indivíduos (do sexo k no instante i).

Modelo 1

- $u_{1jk} = \frac{1}{x^*_{jki}} \left(\mathcal{E}(Y_{jki}|u_{1jk}) - \mathcal{E}(Y_{jki}) \right), \forall x^*_{jki} \neq 0$, é a diferença entre o valor esperado da distância para o indivíduo j do sexo k no instante i, em relação a distribuição de u_{0jk} e a esperança global da distância (para indivíduos do sexo k no instante i) (ponderado pelo valor x^*_{jki}). Outra interpretação: é o quanto o coeficiente angular do indivíduo j do sexo k se difere do coeficiente angular comum a todas os indivíduos do sexo k.

Casos pariticulares

- Coeficientes independentes/não correlacionados (modelo 2): $\psi_{01} = 0$
- Somente coeficientes angulares aleatórios (modelo 3): $\psi_{00} = 0$
- Somente interceptos aleatórios (modelo 4): $\psi_{11} = 0$

Sintaxe para o ajuste dos cinco modelos

- M4: lmer(Distancia ~ Sexo+I(Ano-8):Sexo + (1|Individuo),data=dadosl)
- M3: lmer(Distancia ~ Sexo+I(Ano-8):Sexo + (-1+I(Ano-8)|Individuo),data=dadosl)
- M2: lmer(Distancia ~ Sexo+I(Ano-8):Sexo +(1|Individuo) + (0+I(Ano-8)|Individuo),data=dadosl)
- M1: lmer(Distancia ~ Sexo+I(Ano-8):Sexo + (1+I(Ano-8)|Individuo),data=dadosl)

Estatísticas de Comparação de Modelo (link)

Estatística	modelo			
	4	3	2	1
AIC	445,76	477,87	446,62	448,58
BIC	461,85	493,97	465,39	470,04
AICC	434,59	466,70	433,74	434,04
HQIC	452,28	484,40	454,23	457,28
CAIC	467,85	499,97	472,39	478,04
SABIC	442,89	475,01	443,27	444,76

Estimativas: efeitos fixos e componentes de variância

Efeitos fixos

Parâmetro	Estimativa	EP	IC(95%)	Estat. z	p-valor
γ_{000}	21,21	0,65	[19,94 ; 22,48]	32,64	$< 0,0001$
γ_{002}	1,41	0,84	[-0,25 ; 3,06]	1,67	$< 0,0956$
γ_{101}	0,48	0,09	[0,30 ; 0,66]	5,13	$< 0,0001$
γ_{102}	0,78	0,08	[0,63 ; 0,94]	10,12	$< 0,0001$

Componentes de variância

ψ	σ^2	CCI($\psi/(\psi+\sigma^2)$)
3,30	1,92	0,63

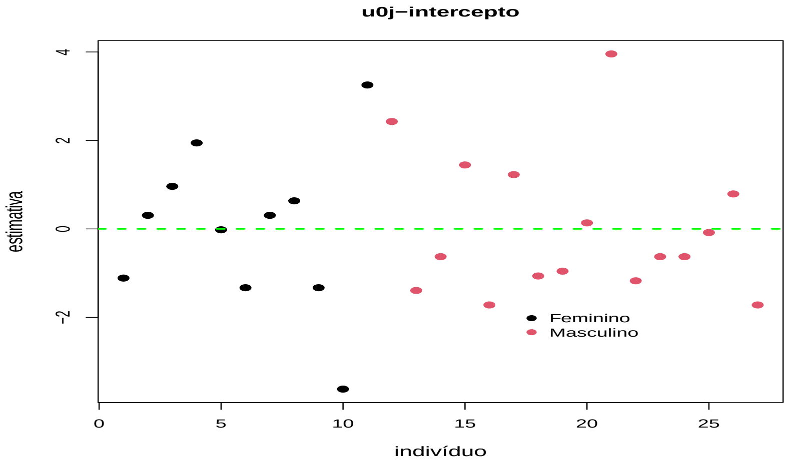
Cont.

- O teste Z (distribuição normal) para a hipótese $H_0 : \gamma_{002} = 0$ resultou em um p-valor $0,0956$.
- Além disso, o teste de Wald (link), para testar a igualdade entre os coeficientes angulares ($H_0 : \gamma_{101} = \gamma_{102}$), resultou em: 6,30 (0,0121).
- Por resultar em significâncias marginais, com p-valores próximo de 0,10 e 0,01, respectivamente e pela análise descritiva, optou-se por rejeitar tais igualdades.

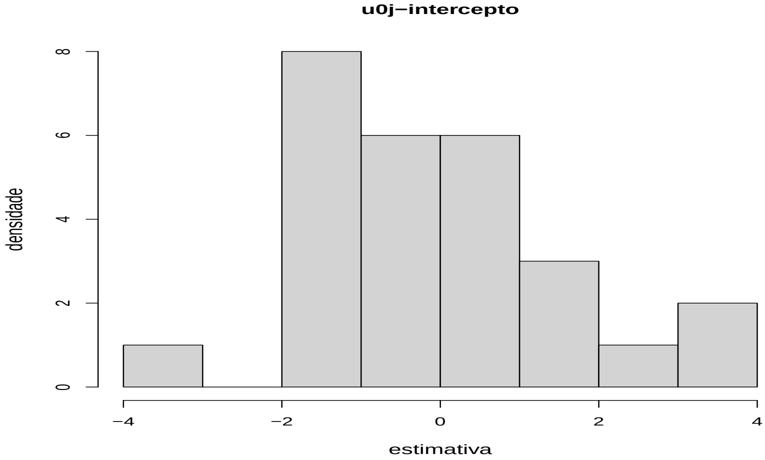
Cont.

- Ademais, ajustou-se um modelo reduzido, o qual corresponde ao modelo 4 com intercepto comum ($\gamma_{002} = 0$). Os critérios de informação foram: AIC = 455,00; BIC = 465,73; AICC = 447,40; HQIC = 459,3; CAIC = 469,73; SABIC = 453,09. Tais valores depoem a favor do modelo 4, original.
- Sintaxe do modelo reduzido: lmer(Distancia $\sim$ I(Ano-8) + (1|Individuo),data=dadosl)

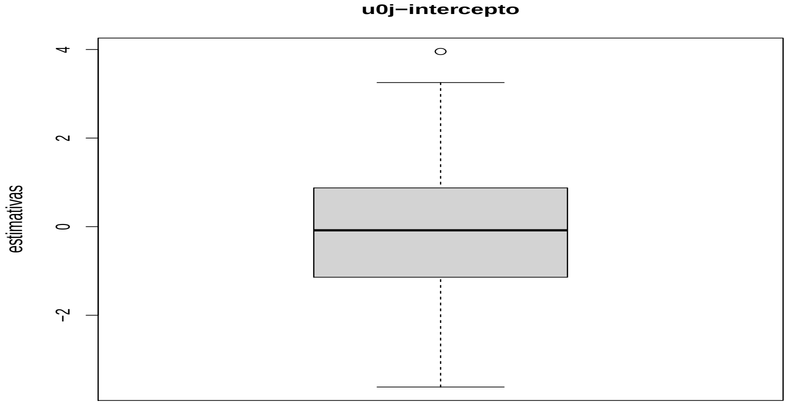
Estimativas dos efeitos aleatórios



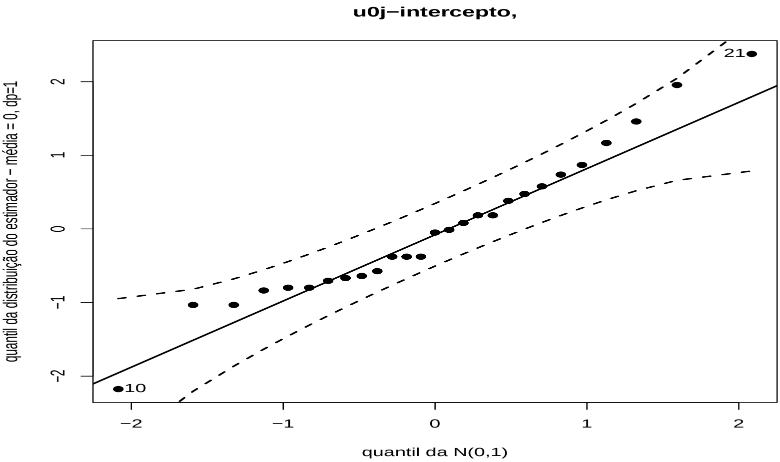
Histograma das estimativas dos efeitos aleatórios



Boxplot das estimativas dos efeitos aleatórios



QQ plot das estimativas dos efeitos aleatórios



Perfis médios observados e ajustados