

Análise de Dados Longitudinais

Modelos Lineares Mistos

Enrico A. Colosimo/UFMG

<http://www.est.ufmg.br/~enricoc/>

Modelo Linear Misto

- 1 Modelo de Efeitos Fixos: apresenta somente fatores fixos, exceto o termo do erro experimental (erro de medida).
- 2 Modelo Misto: apresenta tanto fatores fixos quanto aleatórios, além do erro experimental.

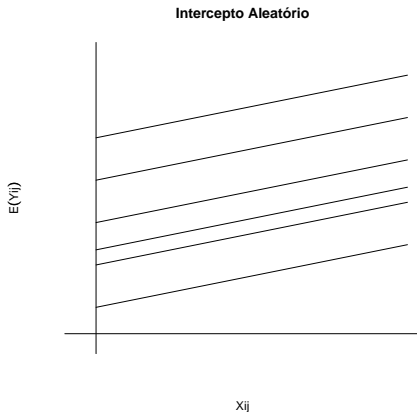
Ideia:

- Os parâmetros da regressão variam de indivíduo para indivíduo explicando as fontes de heterogeneidade da população.
- Cada indivíduo tem a sua própria trajetória média e um subconjunto dos parâmetros de regressão são tomados como aleatórios.
- Efeitos fixos são compartilhados por todos os indivíduos e os aleatórios são específicos de cada um.

Modelo Linear Misto

Exemplo: Intercepto aleatório

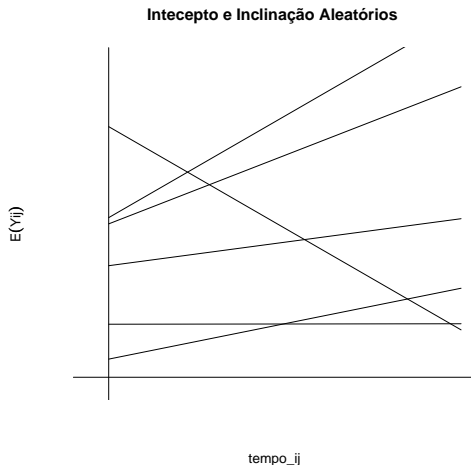
$$Y_{ij} = \beta_{0i} + \beta_1 t_{ij} + \varepsilon_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + \varepsilon_{ij}$$



Modelo Linear Misto

Exemplo: Intercepto e Inclinação aleatórios

$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \varepsilon_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i}t_{ij} + \varepsilon_{ij}$$



- **Características:**

- 1 Características populacionais β (fixos);
- 2 Características individuais β_i ou b_i (aleatórios).

- **Efeito:**

- 1 Média: $E(Y_i) = X_i\beta$
- 2 Estrutura de Covariância: Efeito aleatório induz $Var(Y_i)$.
Separa a variação entre indivíduos daquela intra indivíduos.
- 3 Permite obter estimativa de trajetórias individuais no tempo.

Modelo Linear Misto - Simetria Composta

Exemplo: $Y_{ij} = \beta_{0i} + \beta t_{ij} + \varepsilon_{ij}$ (Intercepto aleatório).

- $\beta_{0i} \sim N(\beta_0, \sigma_{\beta_0}^2)$.
- $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- β_{0i} e ε_{ij} são independentes.

- 1 $Var(Y_{ij}) = \sigma^2 + \sigma_{\beta_0}^2$.
- 2 $Cov(Y_{ij}, Y_{ij'}) = \sigma_{\beta_0}^2$

Modelo Linear Misto - Inclinação aleatória

Exemplo: $Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \varepsilon_{ij}$ (Intercepto e inclinação aleatórios).

- $\beta_{0i} \sim N(\beta_0, \sigma_{\beta_0}^2)$, $\beta_{1i} \sim N(\beta_1, \sigma_{\beta_1}^2)$, $Cov(\beta_{0i}, \beta_{1i}) = \sigma_{\beta_{01}}$.
- $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- $\beta' = (\beta_{0i}, \beta_{1i})$ e ε_{ij} são independentes.

❶ $Var(Y_{ij}) = \sigma_{\beta_0}^2 + \sigma_{\beta_1}^2 t_{ij}^2 + 2t_{ij}\sigma_{\beta_{01}} + \sigma^2.$

❷ $Cov(Y_{ij}, Y_{ij'}) = \sigma_{\beta_0}^2 + t_{ij}t_{ij'}\sigma_{\beta_1}^2 + (t_{ij} + t_{ij'})\sigma_{\beta_{01}}.$

- 1 Predizer trajetórias individuais

$$Y_{ij} = X_{ij}\beta + b_i + \varepsilon_{ij}$$

Resposta Média populacional:

$$E(Y_{ij}) = X_{ij}\beta$$

Resposta média para o i -ésimo indivíduo (trajetória):

$$E(Y_{ij}/b_i) = X_{ij}\beta + b_i.$$

- 2 Flexibilidade em acomodar estruturas não balanceadas

Forma Geral do Modelo Misto

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i$$

em que:

$(\beta)_{p \times 1}$: efeitos fixos;

$(b_i)_{q \times 1}$: efeitos aleatórios.

e,

$$b_i \sim N_q(0, \Sigma) \text{ e } \varepsilon_{ij} \sim N(0, \sigma^2)$$

Sendo b_i e ε_{ij} independentes.

$$q \leq p \Rightarrow Z_i \text{ é um subconjunto de } X_i$$

Incluimos efeitos aleatórios somente para as covariáveis que variam com o tempo.

Característica do Modelo

- 1 Média Populacional ou Marginal

$$E(Y_i) = X_i\beta.$$

- 2 Média condicional ou específica por indivíduo

$$E(Y_i|b_i) = X_i\beta + Z_ib_i.$$

- 3 Covariância Marginal

$$\text{Var}(Y_i) = Z_i \text{Var}(b_i) Z_i' + \sigma^2 I_{n_i}.$$

- 4 Podemos assumir que $\varepsilon_i \sim N(0, R_i)$ mas o usual é tomar $R_i = \sigma^2 I_{n_i}$ e interpreta-lo como covariância condicional. Ou seja,

$$\text{Var}(Y_i/b_i) = R_i = \sigma^2 I_{n_i}.$$

Inferência para o Modelo Misto

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i,$$

em que,

$$b_i \sim N_q(0, \Sigma(\alpha)) \text{ e } \varepsilon_{ij} \sim N(0, \sigma^2),$$

b_i e ε_{ij} independentes.

Desta forma tem-se:

p efeitos fixos e $\frac{q(q+1)}{2} + 1$ efeitos aleatórios.

Inferência Estatística para $\theta = (\beta, \alpha, \sigma^2)$;

- 1 Máxima Verossimilhança.
- 2 Máxima Verossimilhança Restrita.

Função de Verossimilhança

$$\begin{aligned}L(\theta/y) &= \prod_{i=1}^N p(y_i/\theta) \\&= \prod_{i=1}^N \int p(y_i, b_i/\theta) db_i \\&= \prod_{i=1}^N \int p(y_i/b_i, \theta) p(b_i/\theta) db_i\end{aligned}$$

em que,

$$p(y_i/b_i, \theta) \sim N_n(X_i\beta + Z_ib_i, \sigma^2 I)$$

e

$$p(b_i/\theta) \sim N_q(0, \Sigma)$$

Observações

- 1 EMV É obtido usando verossimilhança perfilada e iterações via algoritmo EM ou/e Newton-Raphson. Detalhes numéricos podem ser encontrados em Pinheiro e Bates (2000), Cap. 2.
- 2 O EMVR também pode ser obtido através de

$$l * (\theta) = l(\theta) + \textit{termo}.$$

- 3 A função lme do R fornece EMVR e EMV usando um enfoque híbrido (EM + Newton-Raphson). Esta função é de autoria de Pinheiro e Bates.
- 4 EMV e EMVR têm assintoticamente as propriedades usuais de um estimador de máxima verossimilhança (consistência e normalidade).

Avaliação dos Componentes de Variância

- 1 Número de componentes é igual a $\frac{q(q+1)}{2} + 1$ em que q é o número de efeitos aleatórios no modelo.
- 2 Muitas situações envolvem $q = 2$ (intercepto e inclinação aleatórios) e portanto:

$$\frac{2(2 + 1)}{2} + 1 = 4,$$

que permite termos heterogeneidade de variâncias e covariâncias pois ficam em função do tempo.

- 3 A escolha da "melhor" estrutura de variância-covariância pode ser realizada utilizando o teste da RMVR. Estes testes, usualmente, são na fronteira do espaço de parâmetros. Neste caso, a estatística da RMVR não tem, sob H_0 uma distribuição qui-quadrado.

Dist. da Estatística da RMVR sob H_0

- 1 A distribuição neste caso é uma mistura (50:50) de dist. qui-quadrado. Ou seja, por exemplo, para $H_0 : \sigma_{\beta_1} = 0$

$$RMVR \sim 0.5\chi_q + 0.5\chi_{q+1}$$

- 2 Exemplo
Modelo completo: $q=2$ (intercepto e inclinação aleatórios)
Modelo restrito: $q=1$ (somente intercepto aleatório)

Teste usual (errado): nível de significância: 5,99

Teste correto:

$$RMVR \sim 0,5\chi_1 + 0,5\chi_2$$

nível é 5,14 (Tabela, Apend. C, Fitzmaurice et al, 2004).

- 3 Proposta ad hoc: para testar a 0,05, use o nível de 0,10.

Transmissão Vertical - HIV

Estudo Longitudinal Desbalanceado: Avaliação longitudinal do crescimento de lactentes nascidos de mães infectadas com o HIV-1.

- Comparar longitudinalmente altura de lactentes infectados e não-infectados nascidos de mães infectadas pelo HIV.
- Uma coorte aberta acompanhada no ambulatório de AIDS pediátrica do Hospital das Clínicas da Universidade Federal de Minas Gerais.
- Período: 1995 a 2003.
- Inclusão: primeiros três meses de vida.
- Grupos: (1) não-infectados: 97; (2) infectados: 42.
- Controlado por sexo.

Perfis médio por grupo

Gráfico para os Grupos

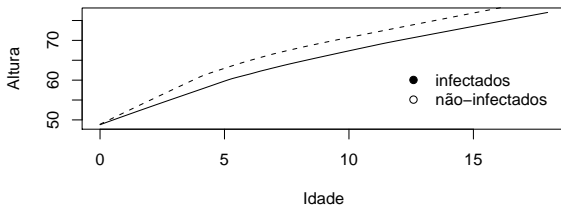
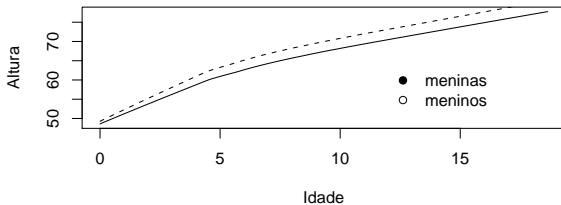


Gráfico para Meninos e Meninas



Modelo para a Média Populacional

Modelo Polinomial:

$$\begin{aligned} E(Y_{ij}) &= \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 \text{sexo}_i + \beta_4 \text{status}_i \\ &+ \beta_5 t_{ij} * \text{sexo}_i + \beta_6 t_{ij} * \text{grupo}_i + \beta_7 t_{ij}^2 * \text{sexo}_i + \beta_8 t_{ij}^2 * \text{grupo}_i \end{aligned}$$

Crianças - Transmissão Vertical - Modelo Marginal (gls)

- Modelo quadrático para a média com termos de interação.
- Algumas formas para a $Var(Y_i)$: exponencial, simetria composta.
- Resultados para os quatro termos da interação.

Coeficiente	Indep. (Incorreto)		Exponencial		Simetria Composta	
	Est.	EP	Est.	EP	Est.	EP
Idade:grupo	-0,164	0,041	-0,160	0,057	-0,142	0,027
Idade2:grupo	0,020	0,008	0,017	0,008	0,018	0,005
Idade:sexo	0,046	0,037	0,165	0,052	0,100	0,025
Idade2:sexo	-0,014	0,007	-0,020	0,008	-0,015	0,005

Crianças - Transmissão Vertical - GEE

- Modelo quadrático para a média com termos de interação.
- Algumas formas para a $Var(Y_i)$: exponencial, simetria composta.
- Modelo para média com 9 termos (interceptos diferentes)
- Resultados para os quatro termos da interação.

Coeficiente	Independente		Simetria Composta	
	Est.	EP	Est.	EP
Idade:grupo	-0,164	0,059	-0,142	0,057
Idade2:grupo	0,020	0,011	0,018	0,008
Idade:sexo	0,046	0,050	0,100	0,047
Idade2:sexo	-0,014	0,009	-0,015	0,007

Obs. As estimativas são as mesmas do gls-normal e o erro-padrão fica inflacionado.

Crianças - Transmissão Vertical - Modelos Misto

- Modelo quadrático para a média com termos de interação.
- Termos aleatórios: somente intercepto, intercepto + linear, intercepto + liner + quadrático.
- Resultados para os quatro termos da interação.

Coeficiente	Interc.		Inter + linear		três termos	
	Est.	EP	Est.	EP	Est.	EP
Idade:grupo	-0,142	0,027	-0,161	0,049	-0,148	0,050
Idade2:grupo	0,018	0,005	0,022	0,004	0,026	0,007
Idade:sexo	0,100	0,025	0,131	0,045	0,114	0,046
Idade2:sexo	-0,015	0,005	-0,009	0,004	-0,013	0,007

Formulação em dois Estágios do Modelo Linear Misto

1 Estágio 1

Medidas Longitudinais no i -ésimo indivíduo são modeladas como:

$$Y_i = Z_i\beta_i + \varepsilon_i$$

em que Z_i covariáveis intra-indivíduo (tempo dependente) e

$$\varepsilon_i \sim N(0, \sigma^2 I_k).$$

2 Estágio 2

β_i : aleatório (variando de indivíduo para indivíduo) tal que:

$$E(\beta_i) = A_i\beta$$

em que A_i contém somente covariáveis que variam entre indivíduos (não dependente do tempo) e

Formulação em dois Estágios do Modelo Linear Misto

$$\text{Var}(\beta_i) = \Sigma.$$

Desta forma,

$$\begin{aligned} Y_i &= Z_i(A_i\beta + b_i) + \varepsilon_i \\ &= X_i\beta + Z_ib_i + \varepsilon_i \end{aligned}$$

Ou seja, sob a restrição que

$$X_i = Z_iA_i$$

obtém-se o modelo de efeitos aleatórios.

Predição dos Efeitos Aleatórios

Objetivo: prever perfis individuais ou identificar indivíduos acima ou abaixo do perfil médio.

Obs.: não dizemos estimar os efeitos pois os mesmos são aleatórios. Dizemos prever os efeitos aleatórios.

Deseja-se:

$$\hat{Y}_i = \hat{E}(Y_i/b_i) = X_i\hat{\beta} + Z_i\hat{b}_i$$

e para tal necessita-se de \hat{b}_i , o chamado Estimador BLUP ("Best Linear Unbiased Predictor") de b_i .

Predição dos Efeitos Aleatórios

No modelo linear misto,

- Y_i e b_i tem uma distribuição conjunta normal multivariada.
- Usando conhecidas propriedades da normal multivariada, temos que

$$E(b_i / Y_i, \hat{\beta}) = \Sigma Z_i' \text{Var}(Y_i)^{-1} (Y_i - X_i \hat{\beta})$$

- Usando os estimadores MVR dos componentes de variância,

$$\hat{b}_i = \hat{\Sigma} Z_i' \widehat{\text{Var}}(Y_i)^{-1} (Y_i - X_i \hat{\beta})$$

o BLUP de b_i .

Predição dos Efeitos Aleatórios

Desta forma obtemos:

$$\begin{aligned}\hat{Y}_i &= X_i\hat{\beta} + Z_i\hat{b}_i \\ &= X_i\hat{\beta} + Z_i\hat{\Sigma}Z_i'\widehat{Var}(Y_i)^{-1}(Y_i - X_i\hat{\beta}) \\ &= X_i\hat{\beta} + (Z_i\hat{\Sigma}Z_i' + \hat{R}_i - \hat{R}_i)\widehat{Var}(Y_i)^{-1}(Y_i - X_i\hat{\beta}) \\ &= (\hat{R}_i\widehat{Var}(Y_i)^{-1})X_i\hat{\beta} + (I_{n_i} - \hat{R}_i\widehat{Var}(Y_i)^{-1})Y_i\end{aligned}$$

em que $Var(\varepsilon_i) = R_i$.

Interpretação: média ponderada entre a média populacional $X_i\hat{\beta}$ e o i -ésimo perfil observado. Isto significa que o perfil predito é encolhido na direção da média populacional.

Interpretação dos Efeitos Aleatórios Preditos

Ou seja,

- R_i : variação intra-indivíduo:
 - $Var(Y_i)$: variação total.
-
- R_i grande, mais peso em $X_i\hat{\beta}$;
 - $Var(b_i)$ grande, mais peso em Y_i ;
 - n_i pequeno mais peso em $X_i\hat{\beta}$.

Pontos Principais:

- A análise de dados longitudinais não fica completa sem a examinação dos resíduos. Ou seja, a verificação das suposições impostas ao modelo e ao processo de inferência.
- As ferramentas usuais de análise de resíduos para a regressão convencional (com observações independentes) podem ser estendidas para a estrutura longitudinal.

Suposições dos Modelos

- Estrutura da média: forma analítica, linearidade dos β 's.
- Normalidade (resposta e efeitos aleatórios).
- Estrutura de Variância-Covariância: Homocedasticidade e correlação das medidas do mesmo indivíduo.

- Defina o vetor de resíduos para cada indivíduo

$$r_i = Y_i - X_i \hat{\beta}, \quad i = 1, \dots, N,$$

que é um estimador para o vetor de erros

$$\epsilon_i = Y_i - X_i \beta, \quad i = 1, \dots, N.$$

- Tratando-se de dados longitudinais, sabemos que os componentes do vetor de resíduos r_i são correlacionados e não necessariamente têm variância constante.

Gráficos:

- Gráfico de r_{ij} vs \hat{Y}_{ij} : é útil para identificar alguma tendência sistemática (por exemplo, presença de curvatura) e presença de pontos extremos ("outliers"). O modelo corretamente especificado não deve apresentar nenhuma tendência neste gráfico.

Limitação: este gráfico não tem necessariamente uma largura constante. Ou seja, cuidado ao interpretar este gráfico com relação a homocedasticidade.

- Gráfico de r_{ij} vs t_{ij} : é também útil para identificar alguma tendência sistemática da média no tempo.

Solução: Examinar resíduos transformados

- Há muitas possibilidades para transformar os resíduos.
- A transformação deve ser realizada de forma que os resíduos “imitem” aqueles da regressão linear padrão.
- Os resíduos r_i^* definidos a seguir são não-correlacionados e têm variância unitária:

$$r_i^* = L_i^{-1} r_i,$$

em que L_i é a matriz triangular superior resultante da decomposição de Cholesky da matriz de covariâncias estimada $\widehat{Var}(Y_i)$, ou seja, $\widehat{Var}(Y_i) = L_i L_i'$.

Resíduos transformados

- Podemos aplicar a mesma transformação ao vetor de valores preditos \hat{Y}_i , ao vetor da variável resposta Y_i e à matriz de covariáveis \mathbf{X}_i :

$$\hat{Y}_i^* = L_i^{-1} \hat{Y}_i$$

$$Y_i^* = L_i^{-1} Y_i$$

$$\mathbf{X}_i^* = \hat{L}_i^{-1} \mathbf{X}_i$$

e então todos os diagnósticos de resíduos usuais para a regressão linear padrão podem ser aplicados para r_i^* .

Gráficos de Adequação

- Gráfico de dispersão dos resíduos transformados r_{ij}^* versus os valores preditos transformados \hat{Y}_{ij}^* : não deve apresentar nenhum padrão sistemático para um modelo corretamente especificado. Ou seja, deve apresentar um padrão aleatório em torno de uma média zero. Útil para verificar homocedasticidade.
- Gráfico de dispersão dos resíduos transformados r_{ij}^* versus covariáveis transformadas X_{ij}^* (em especial, idade ou tempo): verificar padrões de mudança na resposta média ao longo do tempo;
- QQ-plot de r_i^* : verificar normalidade e identificar outliers.

Semi-variograma

- O semi-variograma, denotado por $\gamma(h_{ijk})$, é dado por:

$$\gamma(h_{ijk}) = \frac{1}{2}E(r_{ij} - r_{ik})^2,$$

em que $h_{ijk} = t_{ij} - t_{ik}$.

- O semi-variograma pode ser utilizado como uma ferramenta para verificar a adequação do modelo selecionado para a estrutura de covariância dos dados.

Semi-variograma

- Como os resíduos têm média zero, o semi-variograma pode ser reescrito como:

$$\begin{aligned}\gamma(h_{ijk}) &= \frac{1}{2}E(r_{ij} - r_{ik})^2 \\ &= \frac{1}{2}E(r_{ij}^2 + r_{ik}^2 - 2r_{ij}r_{ik}) \\ &= \frac{1}{2}\text{Var}(r_{ij}) + \frac{1}{2}\text{Var}(r_{ik}) - \text{Cov}(r_{ij}, r_{ik}).\end{aligned}$$

- Quando o semivariograma é aplicado aos resíduos transformados, r_{ij}^* , a seguinte simplificação é obtida:

$$\gamma(h_{ijk}) = \frac{1}{2}(1) + \frac{1}{2}(1) - 0 = 1.$$

Semi-variograma

- Logo, se o modelo é corretamente especificado para a matriz de covariâncias, o gráfico do semi-variograma amostral $\hat{\gamma}(h_{ijk})$ dos resíduos transformados versus h_{ijk} deveria flutuar aleatoriamente em torno de uma linha horizontal centrada em 1.
- O semi-variograma é muito sensível a outliers.

Estudo de caso: Influência da menarca nas mudanças do percentual de gordura corporal

- Estudo prospectivo do aumento de gordura corporal em uma coorte de 162 garotas.
- Sabe-se que o percentual de gordura nas garotas tem um aumento considerável no período em torno da menarca (primeira menstruação).
- Parece que este aumento continua significativo por aproximadamente quatro anos depois da menarca, mas este comportamento ainda não foi devidamente estudado.
- As meninas foram acompanhadas até quatro anos depois da menarca.

- Há um total de 1049 medidas, com uma média de 6,4 medidas por menina.
- Variáveis de interesse:
 - Resposta: Percentual de gordura corporal;
 - Covariáveis: Tempo em relação à menarca (Idade da menina no instante observado menos Idade quando teve a menarca) - pode ser positivo ou negativo.

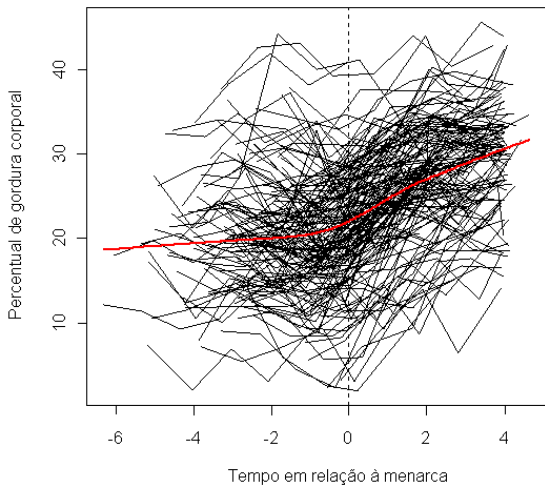


Figura: Gráfico de perfis com curva alisada

- O modelo inicialmente proposto considera que cada garota tem uma curva de crescimento spline linear com um knot no tempo da menarca.
- Ajustou-se o seguinte modelo linear de efeitos mistos:

$$E(Y_{ij}|b_i) = \beta_1 + \beta_2 t_{ij} + \beta_3 (t_{ij})_+ + b_{1i} + b_{2i} t_{ij} + b_{3i} (t_{ij})_+,$$

em que

$$(t_{ij})_+ = \begin{cases} t_{ij} & \text{se } t_{ij} > 0 \\ 0 & \text{se } t_{ij} \leq 0. \end{cases}$$

- Lembremos que no modelo linear de efeitos mistos, a matriz de variância-covariância de Y_i é dada por:

$$\text{Var}(Y_i) = Z_i \Sigma Z_i' + \sigma^2 I_{n_i},$$

em que Z_i é a matriz de covariáveis relacionadas aos efeitos aleatórios, Σ é a matriz de covariância dos efeitos aleatórios e n_i é o número de observações do i -ésimo indivíduo, $i = 1, \dots, N$.

- Logo, os resíduos transformados neste caso podem ser obtidos a partir da decomposição de Cholesky da matriz estimada

$$\widehat{\text{Var}}(Y_i) = Z_i \hat{\Sigma} Z_i' + \hat{\sigma}^2 I_{n_i}.$$

- Resultados do ajuste:

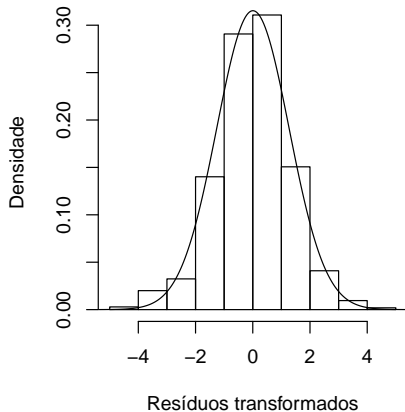
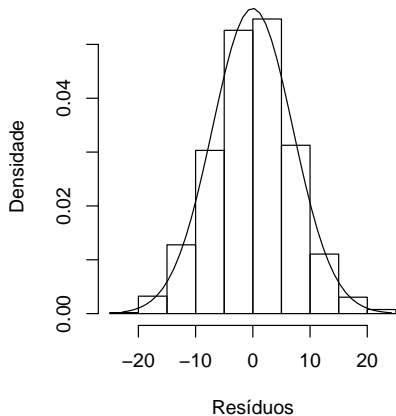
Tabela: Coeficientes de regressão estimados (efeitos fixos) e erros padrões

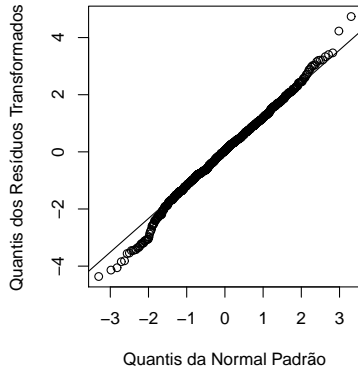
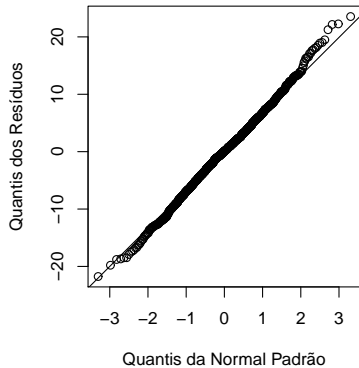
Variável	Estimativa	EP	t	p-valor
Intercepto	21,3614	0,5646	37,8400	0,0000
Tempo	0,4171	0,1572	2,6500	0,0081
(Tempo) ₊	2,0471	0,2280	8,9800	0,0000

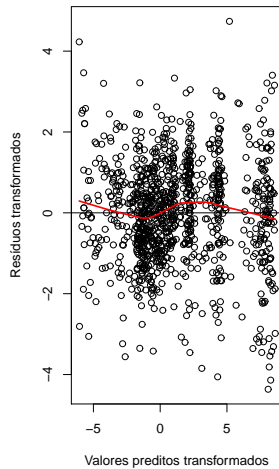
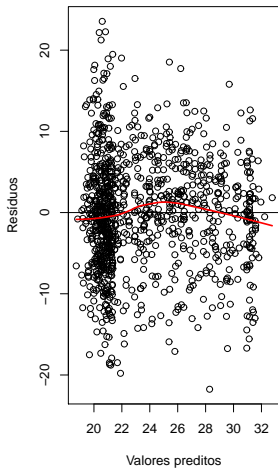
Tabela: Covariâncias estimadas para os efeitos aleatórios (\hat{G}) e variância estimada para os erros ($\hat{\sigma}^2$)

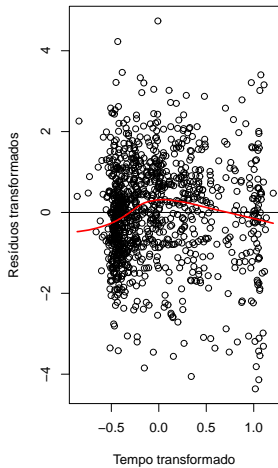
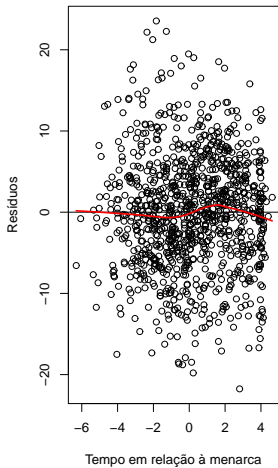
Parâmetro	Estimativa	Parâmetro	Estimativa
$Var(b_{1i}) = g_{11}$	45,9407	$Cov(b_{1i}, b_{2i}) = g_{12}$	2,5275
$Var(b_{2i}) = g_{22}$	1,6309	$Cov(b_{1i}, b_{3i}) = g_{13}$	-6,1141
$Var(b_{3i}) = g_{33}$	2,7496	$Cov(b_{2i}, b_{3i}) = g_{23}$	-1,7513
$Var(e_i) = \sigma^2$	9,4734		

- Análise de resíduos:









- Da figura anterior (Resíduos vs Tempo), observa-se uma tendência quadrática no período após a menarca.
- Refinando o modelo anterior, consideraremos agora que cada garota tem uma curva de crescimento spline linear-quadrática com um knot no tempo da menarca.
- Ajustou-se o seguinte modelo linear de efeitos mistos:

$$\begin{aligned} E(Y_{ij}|b_i) &= \beta_1 + \beta_2 t_{ij} + \beta_3 (t_{ij})_+ + \beta_4 (t_{ij})_+^2 + b_{1i} + b_{2i} t_{ij} + \\ &+ b_{3i} (t_{ij})_+ + b_{4i} (t_{ij})_+^2, \end{aligned}$$

em que

$$(t_{ij})_+^2 = \begin{cases} t_{ij}^2 & \text{se } t_{ij} > 0 \\ 0 & \text{se } t_{ij} \leq 0. \end{cases}$$

- Resultados do ajuste:

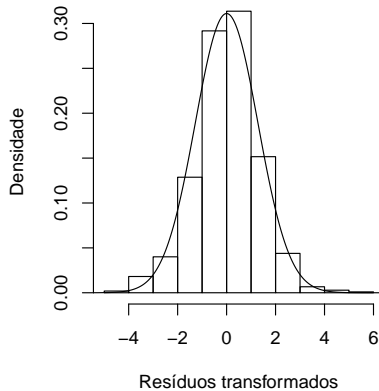
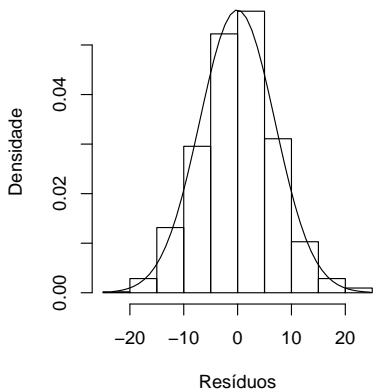
Tabela: Coeficientes de regressão estimados (efeitos fixos) e erros padrões

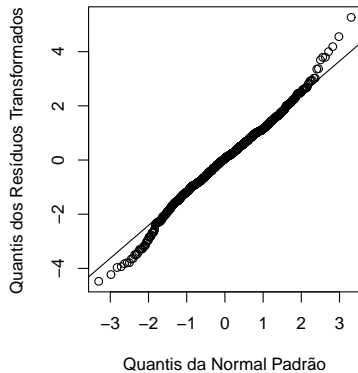
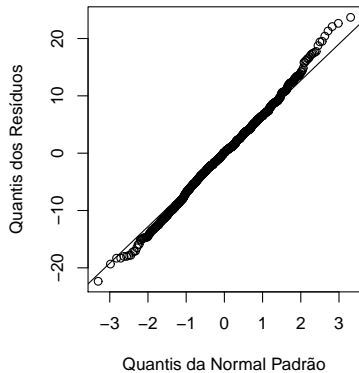
Variável	Estimativa	EP	t	p-valor
Intercepto	20,4201	0,5817	35,1032	0,0000
Tempo	-0,0155	0,1612	-0,0962	0,9234
(Tempo) ₊	4,8439	0,4055	11,9446	0,0000
(Tempo) ₊ ²	-0,6469	0,0772	-8,3842	0,0000

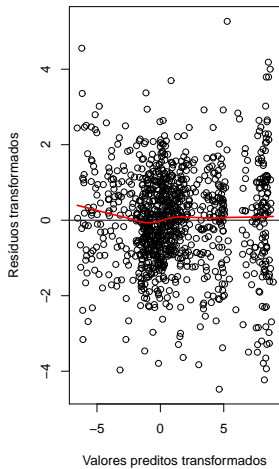
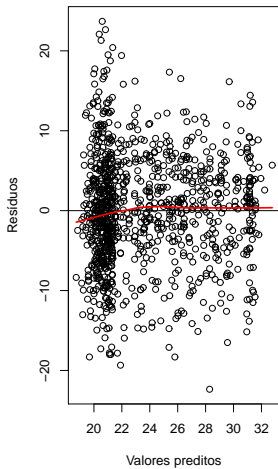
Tabela: Covariâncias estimadas para os efeitos aleatórios (\hat{G}) e variância estimada para os erros ($\hat{\sigma}^2$)

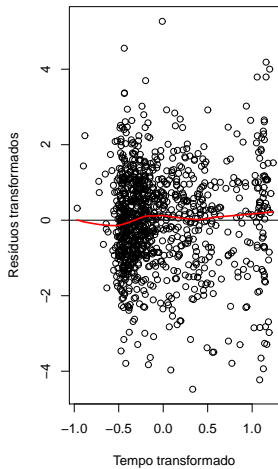
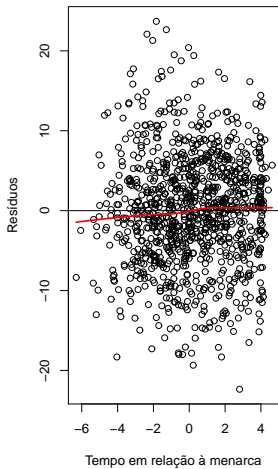
Parâmetro	Estimativa	Parâmetro	Estimativa
$Var(b_{1i}) = g_{11}$	48,0586	$Cov(b_{1i}, b_{3i}) = g_{13}$	-9,5900
$Var(b_{2i}) = g_{22}$	1,7326	$Cov(b_{1i}, b_{4i}) = g_{14}$	0,6479
$Var(b_{3i}) = g_{33}$	5,3693	$Cov(b_{2i}, b_{3i}) = g_{23}$	-1,5342
$Var(b_{4i}) = g_{44}$	0,1172	$Cov(b_{2i}, b_{4i}) = g_{24}$	-0,1735
$Cov(b_{1i}, b_{2i}) = g_{12}$	3,0295	$Cov(b_{3i}, b_{4i}) = g_{34}$	-0,4395
$Var(e_i) = \sigma^2$	8,0274		

- Análise de resíduos:









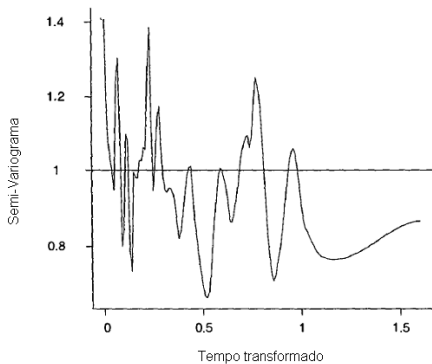


Figura: Semi-variograma empírico para os resíduos transformados

- Gráficos de dispersão não apresentam mais nenhuma tendência acentuada.
- Semi-variograma está oscilando aleatoriamente em torno da linha horizontal 1.
- Pela análise de resíduos, confirmamos a adequação do segundo modelo proposto.

O que fazer frente a violação de suposições?

- Verificar a estrutura da média.
- Transformar a resposta.
- Propor outra estrutura de Variância-Covariância para os erros (Modelo Marginal)
- Modelar a estrutura variância-covariância do erro intra-indivíduo (erro de medida, Modelo de Efeito Aleatórios).

Verificar a Estrutura da Média

- Existe alguma proposta teórica da área?
- Perfis, especialmente os alisados, são as principais ferramentas.
- Propostas Empíricas: splines (com um ou no máximo dois knots), modelos lineares ou quadráticos. Possivelmente algo como decaimento exponencial.

Transformar a resposta

- Vantagens quando temos distribuição assimétrica para a resposta. Por exemplo: custo. Utilizar transformação logarítmica.
- Desvantagem: interpretação dos resultados.

Propor outra estrutura de Variância-Covariância para os erros (Modelo Marginal)

- Utilizar a não-estruturada em delineamentos balanceados quando o número de tempos medidos não for excessivo.
- Incluir heterocedasticidade quando possível.

Modelar a variância-covariância do erro Intra Indivíduos (Modelo de Efeito Aleatórios)

- Suposição: $Var(\varepsilon_i) = \sigma^2 I$.

- Podemos estruturar a

$$Var(\varepsilon_i).$$

Isso pode ser feito inclusive em termos de covariáveis.

- O R ajusta alguns tipos de estrutura.