

# ESTRATÉGIAS DE MODELAGEM DE DADOS LONGITUDINAIS: UMA APLICAÇÃO EM CRIANÇAS COM HIDRONEFROSE FETAL

Marcelo Militão Abrantes – [militaoma@uol.com.br](mailto:militaoma@uol.com.br)  
UFMG e Faculdade de Medicina de Barbacena (FAME)

## 1 INTRODUÇÃO SOBRE DADOS LONGITUDINAIS

A principal vantagem dos estudos longitudinais é que eles conseguem separar em contextos populacionais o que é efeito temporal do efeito individual. Os dados provenientes de estudos longitudinais se caracterizam pela seqüência de duas ou mais observações em cada indivíduo.

As etapas envolvidas na análise de efeitos aleatórios são as seguintes: análise exploratória dos dados, ajuste do modelo (efeitos fixos e efeitos aleatórios) e avaliação do ajuste do modelo (efeitos aleatórios e termos de interação) (EKUMA e LIX, 2004).

A análise exploratória dos dados tem como objetivos avaliar se existe correlação entre as medidas repetidas e se esta diminui ou aumenta com o tempo, se a tendência temporal é linear ou com outra forma, se existe variabilidade entre as medidas ao longo do tempo e a presença de valores extremos (“outliers”). As ferramentas utilizadas na análise exploratória são análise gráfica e descritiva. Quando os dados são balanceados (as medidas são obtidas em tempos igualmente espaçados), a análise exploratória pode ser feita por meio da matriz de correlação dos dados (avaliação das correlação das medidas), de gráficos longitudinais de médias (avaliação da tendência temporal), medidas-resumo e análise gráfica da média dos resíduos obtidos com regressão linear ajustada por mínimos quadrados (avaliação da heterogeneidade) e gráficos de perfis (identificação de valores extremos) (FAUSTO, 2005).

O objetivo deste trabalho é apresentar diferentes estratégias de modelagem para dados longitudinais na presença de dados desbalanceados. Na Seção 2 vamos apresentar brevemente os dois principais modelos para dados longitudinais e métodos inferenciais. O banco de dados de nefrologia pediátrica que motivou este trabalho é apresentado na Seção 3. As diversas estratégias de modelagem em dados longitudinais são apresentadas na Seção 4 e este material termina na Seção 5 com considerações finais e proposta de trabalhos futuros.

## 2 Modelos em Dados Longitudinais

Os dois modelos clássicos para dados longitudinais, marginal e de efeitos aleatórios, são apresentados nesta seção.

### 2.1 Modelo marginal

A principal característica deste modelo é que a média e variância de  $Y$  são modeladas separadamente. A interpretação de  $\beta$  é exatamente a mesma dos modelos para dados transversais (para cada mudança de uma unidade na variável independente a variável dependente aumenta  $\beta$ ). A estimação da média será abordada no item 2.3 deste texto. A equação geral do modelo é apresentada abaixo usando a notação apresentada no quadro 1:

$$Y_i = X_i' \beta + \varepsilon_i \quad (1)$$

$Y$  representa variável resposta,  $i$  serve para identificar medidas de um determinado indivíduo,  $X_i$  representa a(s) variável(is) independente(s) de um determinado indivíduo,  $\beta$  representa o coeficiente estimado pelo modelo e significa o valor que mudará na variável dependente quando a variável independente apresentar variação de uma unidade, controlada pelas demais variáveis do modelo.

Os erros de ajuste do modelo, representados por  $\varepsilon_i$  apresentam uma distribuição normal com média zero e variância  $V_i$  ( $\varepsilon_i \sim N(0, V_i)$ ). A variância ( $V_i$ ) pode ser modelada de três formas principais:

- a) não estruturada: nesta forma são consideradas a variância e covariância das medidas. Esta forma exige que os dados sejam balanceados, ou seja, com o mesmo número de medidas para cada indivíduo e com mesmo intervalo de tempo entre as medidas.
- b) correlação uniforme: considera-se que a variância é 1 e que a covariância entre as medidas é uniforme o tempo todo.
- c) auto-regressiva ou correlação serial: semelhante à uniforme porém a matriz é resultado da multiplicação da variância por um número de valor um para as variâncias e por um número representado pelo valor de um coeficiente de correlação elevado à diferença de tempo entre as medidas ( $\rho^{(ni-j)}$ ). A intenção é

que as informações mais próximas tendem a ser mais correlacionadas que as informações mais distantes pois à medida que aumenta a diferença entre os tempos a correlação diminui, ou seja, quanto maior a distância entre os tempos maior o coeficiente e menor a correlação. A vantagem é que a matriz é modelada em apenas dois parâmetros ( $1$  e  $\rho$ ) e  $0 < \rho < 1$ .

Uma limitação desta modelagem é que estes modelos não medem fatores individuais que afetem  $y$  e possam causar correlação no tempo. Quando se introduz um efeito latente para cada indivíduo pode-se medir quanto o intercepto de um indivíduo difere do restante. Pode-se assumir então que estes efeitos sejam fixos ou aleatórios. O modelo de efeitos fixos assume que os fatores individuais são fixos. As desvantagens é que estes modelos possuem um grande número de parâmetros quando  $n$  é grande. Tais parâmetros não são adequadamente estimados quando o número de observações no tempo é pequeno. Além disso, os resultados dizem respeito apenas à amostra e não para uma determinada população (VIEIRA, 2006). O modelo de efeitos aleatórios será explicado mais detalhadamente no item abaixo.

## **2.2 Modelo linear de efeitos mistos (efeitos aleatórios)**

O modelo marginal, apresentado no item 2.1, não considera a diferença entre os indivíduos o que ocorre no modelo de efeitos aleatórios. Esses modelos têm dois componentes: um intraindividual (uma mudança longitudinal intraindividual é descrita pelo modelo de regressão com um intercepto e inclinação populacional) e outro entre-indivíduos (variação no intercepto e inclinação individual). O modelo de efeitos aleatórios permite não somente descrever a tendência temporal levando em conta a correlação que existe entre medidas sucessivas como também estimar a variação na medida basal e a taxa de mudança ao longo do tempo. Neste modelo, as medidas dos indivíduos não precisam ser igualmente espaçadas e balanceadas e as análises podem ser conduzidas com os dados de indivíduos que foram perdidos de seguimento ou que apresentam ausência de informação em algum momento do estudo (FAUSTO, 2005).

No modelo linear de efeitos aleatórios, os coeficientes da regressão são denominados efeitos fixos e o conjunto de variâncias e covariâncias, efeitos aleatórios (GOLDSTEIN, BROWNE e RABASH, 2002). O uso do estimador de máxima verossimilhança restrita (EMVR) produz estimativas não viciadas da variância,

removendo o vício que existe na estimação da média (DIGGLE *et al.*, 1994) (descrito mais detalhadamente no item 2.3). De modo simplificado, os componentes da variância são obtidos somando os resíduos ao quadrado e dividindo esse somatório pelos graus de liberdade (DIGGLE *et al.*, 1994). Ao aplicar o método da máxima verossimilhança restrita em modelos com vários efeitos fixos e aleatórios, utiliza-se um sistema de equações e processo iterativo para a obtenção dos componentes da variância (BARACHO, 2003; FAUSTO, 2005).

Os efeitos aleatórios no intercepto representam a heterogeneidade natural entre os indivíduos decorrente de fatores (genéticos ou ambientais) não medidos (MOLENBERGHS e VERBEKE, 2001; DIGGLE *et al.*, 1994). Ao utilizar o modelo de efeitos aleatórios, o método selecionado para codificar a variável tempo deve ser levado em consideração no momento de interpretar o modelo. Uma das alternativas para a codificação da variável identificadora do tempo, utilizada neste estudo, é aquela em que a medida basal tem o valor zero (data do diagnóstico intra-útero) e as demais medidas aumentam sucessivamente. Utilizando este método, o intercepto representa o valor médio da variável dependente no início do estudo (EKUMA e LIX, 2004).

Uma desvantagem deste modelo é que o mesmo pressupõe que os erros não estão correlacionados com as variáveis dependentes consideradas no modelo. Este pressuposto pode ser testado pelo teste de Hausman. Como resultado as estimativas do coeficiente pode ficar viesadas. Este pressuposto pode ser violado como resultado de inconsistências no modelo, causadas por variáveis não mensuradas. Estes modelos podem ser usados quando a hipótese nula de independência é aceita, caso contrário, uma alternativa seria a adoção dos modelos de efeitos fixos (VIEIRA, 2006).

O modelo pode ser apresentado pela equação abaixo:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \quad (2)$$

Onde  $Z_i$  é usualmente um subconjunto de  $X_i$  e freqüentemente as covariáveis que formam  $Z_i$  são dependentes do tempo. Ainda neste modelo assume-se que  $b_i$  e  $\varepsilon_i$  são independentes e tem distribuição normal com média zero e variâncias apresentadas abaixo:

$b_i \sim N(0, D)$  onde  $D$  é diferente para cada indivíduo

$\varepsilon_i \sim N(0, \Sigma)$  onde  $\Sigma = \sigma^2 I$

### 2.3 Inferência para os modelos para dados longitudinais

O estimador de máxima verossimilhança (EMV) e o estimador de máxima verossimilhança restrita (EMVR) tem ambos o mesmo mérito de ser baseado no

princípio da verossimilhança que leva a propriedades úteis de consistência, normalidade assintótica e eficiência. EMV também fornece estimadores dos efeitos fixos enquanto EMVR não. Por outro lado, para modelos balanceados ANOVA mistos, o EMVR para variância são idênticos aos estimadores do tipo ANOVA obtidos pela equação que ajusta os quadrados médios com valor igual ao esperado. Isto implica propriedades de variância ótimas e mostra que EMVR naquele contexto não precisam se basear em nenhum critério de normalidade pois somente suposições de momentos estão envolvidas (VERBEKE, 2000). Quando  $n$  é grande os EMV e os EMVR são similares mas quando a amostra não é tão grande e o número de regressores (parâmetros) for grande deve ser dada preferência à EMVR.

#### **2.4 Dados balanceados e não balanceados**

São considerados dados balanceados quando o número de medidas e o tempo entre estas medidas, para cada indivíduo, é a mesma. Por outro lado, quando o número de medidas é diferente entre os indivíduos ou o tempo em que elas foram medidas é diferente os dados são denominados desbalanceados.

Com dados desbalanceados, a avaliação da correlação dos dados por meio de matriz de correlação só pode ser realizada com a utilização de uma variável categórica do tempo. A avaliação da estrutura de correlação dos dados é importante quando se trabalha com dados balanceados e se pretende avaliar o uso, por exemplo, de um modelo autoregressivo para modelar a covariância dos dados. O modelo autoregressivo exige que as medidas sejam obtidas em tempos igualmente espaçados, não sendo aplicável para dados desbalanceados (BELLOCO, 2001). Dados desbalanceados e com um número grande de observações para cada indivíduo também não permitem que se trabalhe com uma matriz de correlação não estruturada (CNAAN, LAIRD e SLASOR, 1997; FAUSTO, 2005). Na prática dados longitudinais são frequentemente desbalanceados no sentido de que um número diferente de medidas está disponível para todos os sujeitos e/ou que as medidas não são obtidas no mesmo momento (VERBEKE, 2000).

Assim, na análise de dados longitudinais desbalanceados o modelo marginal não pode ser utilizado porque não se consegue estruturar a matriz de variância e covariâncias. A opção recai então no uso do modelo de efeitos aleatórios.

### **3 SITUAÇÃO CLÍNICA E BANCO DE DADOS (HIDRONEFROSE FETAL)**

O aparelho urinário é estéril em todo seu trajeto, com exceção do terço terminal da uretra, onde se encontram bactérias oriundas do períneo. Quando microorganismos alcançam outras áreas do trato urinário (via ascendente ou via hematogênica) colonizando e invadindo os tecidos, desencadeia-se a infecção do trato urinário (ITU).

A principal condição de defesa do trato urinário contra invasão, fixação e multiplicação bacteriana é o livre fluxo da urina, desde sua formação no parênquima renal até seu armazenamento temporário na bexiga, seguido da micção. A alteração desse fluxo livre, causada por fatores que a ele se interponham – uropatias congênicas ou outras alterações adquiridas – representa riscos de ITU, quase sempre de caráter recidivante. Por isso, reconhecê-los e tratá-los adequadamente são pontos importantes no manuseio desses pacientes. A investigação do trato urinário por imagens está indicada após o primeiro episódio bem documentado de infecção urinária em qualquer idade e para ambos os gêneros. A indicação dessa investigação está baseada na alta prevalência de malformações congênicas, especialmente o refluxo vesicoureteral, que tem sido identificado em 25% a 50% das crianças avaliadas após um surto de ITU.

Com a utilização crescente da ultra-sonografia obstétrica, aliada à melhoria tecnológica dos equipamentos, as malformações do trato urinário têm sido freqüentemente identificadas intra-útero. Entre essas, a hidronefrose é a alteração mais comumente diagnosticada, sendo relatada uma freqüência de 1:700 gestações. Grande parte delas pode apresentar uma resolução espontânea intra-útero ou no período neonatal, considerando-se, nesse caso, a característica da micção fetal, efetuada em ambiente fechado e de maior pressão – bolsa amniótica. Entretanto as que persistem podem ser indício de graves alterações, obstrutivas ou não, do trato urinário. É de muita importância a avaliação do líquido amniótico pela ultra-sonografia fetal. A diminuição dele pode significar uma baixa taxa de filtração glomerular fetal – como ocorre nas obstruções baixas (válvula de uretra posterior) ou altas (ureterais bilaterais), que indicam pior prognóstico funcional imediato. Dados da Unidade de Nefrologia Pediátrica do Hospital das Clínicas da UFMG (UNP – HC-UFMG) demonstram que as causas mais freqüentes, em 222 crianças portadoras de hidronefrose fetal persistente são: estenose pieloureteral (26%), hidronefrose idiopática (22%), refluxo vesicoureteral primário (15%), rim multicístico (13%), megaureter primário (10%) e válvula de uretra posterior (8%). A conduta no pós-natal imediato, visa à profilaxia da ITU. De uma

maneira geral, entre quatro e sete dias de vida, quando há incremento do ritmo de filtração glomerular do recém-nascido, reduzindo-se a incidência de resultados falso-negativos, faz-se a ultra-sonografia dinâmica do trato urinário. Conforme o que se comprovar nesse exame, a Uretrocistografia Miccional (UCM) com estudo radiológico da coluna lombos sacra deve ser o passo seguinte. A propedêutica de imagens, se necessário, pode ser complementada com a realização de cintilografia renal estática e dinâmica e, em alguns raros casos, com a de urografia excretora. O tratamento adotado vai depender da causa básica encontrada. O diagnóstico precoce tem permitido uma conduta conservadora na maioria dos casos, sendo a correção cirúrgica indicada nos casos de obstrução grave do fluxo urinário. O seguimento deve ser a longo prazo, mantendo-se quimioprofilaxia e avaliação periódica de imagens. Com esse protocolo, tem-se obtido sucesso na prevenção de infecções urinárias, sendo que apenas cinco das 222 crianças da casuística acima referida apresentaram infecções urinárias recidivantes, apesar da quimioprofilaxia. No entanto, nessa mesma casuística, 12% das crianças apresentaram um mau prognóstico, evoluindo para o óbito e/ou falência renal, apesar de todas as intervenções realizadas. Assim, a presença de hidronefrose fetal deve ser sempre investigada no período neonatal e mantendo-se a urina estéril e um fluxo urinário adequado (SILVA *et al.*, 2005). Definição de hidronefrose pelo MEDLINE é “Aumento anormal do rim ou inchaço devido à dilatação dos cálices renais e pelve renal. Está frequentemente associado com obstrução do ureter ou doença renal crônica que dificultam o fluxo urinário normal da urina até a bexiga”. A válvula de uretra posterior está associada com evolução para doença renal crônica (OLIVEIRA, 2002).

Já foi demonstrada correlação entre o comprimento de rim displásico multicístico e parâmetros corporais como peso e altura com “modelo linear de efeito misto” (linear mixed effect model) (RABELO *et al.*, 2006). A involução do rim afetado foi mais intensa nos primeiros 30 meses com hipertrofia renal compensatória na unidade renal contralateral mais intensa nos primeiros 24 meses de idade (RABELO *et al.*, 2005-C). Variáveis associadas com involução são gênero, rim não palpável, comprimento renal à admissão (RABELO *et al.*, 2005-B; RABELO *et al.*, 2005-A).

O refluxo vesicoureteral é fator preditivo de evolução para doença renal crônica (SILVA *et al.*, 2006-B). Já foi demonstrada a melhora na taxa e evolução para doença crônica em séries histórias de pacientes com refluxo vesicoureteral após melhora na elucidação da sua fisiopatologia (SILVA *et al.*, 2006-A). Em 53 crianças com refluxo vesicoureteral diagnosticado intra-útero durante investigação de hidronefrose a evolução

foi relativamente benigna (SILVA *et al.*, 2006-C). Em pacientes com diagnóstico após dois anos de idade a prevalência de lesão renal foi maior (SILVA *et al.*, 2003).

A medida do diâmetro antero-posterior da pelve renal (DAP) apresenta sensibilidade adequada para identificar fetos com uropatia significativa bem como aqueles que precisarão de intervenção após o nascimento (BOUZADA *et al.*, 2004). A presença de oligohidrânio e megacistos no ultrassom pré-natal é altamente preditor de obstrução fetal uretral. A obstrução da junção ureteropélvica também pode ser causa de hidronefrose fetal e em 77 pacientes a intervenção cirúrgica foi benéfica naqueles com hidronefrose intensa e função renal diminuída por permitir a melhora ou preservação do parênquima renal. Entretanto, nos pacientes com dilatação moderada ou leve o tratamento conservador e seguimento clínico foram satisfatórios (APOCALYPSE *et al.*, 2003).

O presente estudo objetiva avaliar a evolução da pelve renal de crianças com diagnóstico de hidronefrose realizado ainda na fase intra-uterina (fetal). O banco de dados contém informações de 241 crianças com duas unidades renais cada, ou seja, 482 unidades renais. Todas as crianças tiveram uma avaliação basal intra-útero, outra na primeira semana após o nascimento e em seguida foram avaliadas em média a cada seis meses. As variáveis que serão consideradas no presente estudo são as apresentadas no quadro 2. Os dados do presente estudo são desbalanceados uma vez que o número de medidas e o intervalo entre as medidas para um mesmo indivíduo são diferentes.

A avaliação antropométrica será baseada somente na altura porque esta reflete variações ao longo do tempo, ou seja, variações crônicas. O peso por sua vez usualmente reflete alterações agudas que podem ser provocadas por diversas doenças da infância (GOULART *et al.*, 2005).

Quadro 2 – Variáveis do banco de dados

Variável	Classificação	Respostas possíveis
Dilatação da pelve	Contínua	Dependente
Lateralidade	Catagórica	Independente
Bilateral	Catagórica	Independente
Tratamento	Catagórica	Independente
Uropatia	Catagórica	Independente
Parto prematuro	Catagórica	Independente
Sexo	Catagórica	Independente
Infecção do trato urinário	Catagórica	Independente
Altura	Contínua	Independente com medidas repetidas
		Percentil ou escore-Z

## **4 ESTRATÉGIAS NA ANÁLISE DE DADOS LONGITUDINAIS**

No quadro 3 são apresentadas as estratégias de análise para dados balanceados. Para melhor compreensão destas estratégias são apresentados dados hipotéticos, com apenas uma variável e uma covariável no quadro 4.

QUADRO 3 – Estratégias de análise para dados balanceados

Estratégia	Fórmula	Comentários sobre modelo
1 Modelo de efeitos aleatórios	$Peso_{it} = \alpha + \beta altura_{it} + \varepsilon_i$	Influência da altura no peso ao longo do tempo. Ao colocar t como covariável o tempo pode ser diferente entre as variáveis. A variação intra-indivíduo é parecida com a variação inter-indivíduo.
2 Modelo com variáveis de mudança nas covariáveis	$Peso_{it} = \alpha + \beta (\Delta altura)_t + \varepsilon_i$	Influência da variação da altura (em dois momentos) no peso ao longo do tempo. O peso no tempo t2 será função da mudança na altura de t1 a t2.
3 Modelo com variável dependente no tempo anterior como covariável	$Peso_{it} = peso_{it-1} + altura_{it} + \varepsilon_i$	O peso no tempo t2 será função da altura no mesmo momento e do peso no momento anterior.
4 Somente mudança na variável independente	$\Delta peso_{it} = \beta altura_{it} + \varepsilon_i$	A variação no peso, e não a medida do peso, será consequência da altura ao longo do tempo
5 Mudança na variável independente e nas covariáveis	$\Delta peso_{it} = \beta (\Delta altura_{it}) + \varepsilon_i$	A variação no peso, e não a medida do peso, será consequência da variação da altura ao longo do tempo

QUADRO 4 – Dados hipotéticos para exemplificar estratégias de análise

Indivíduo	Tempo	Resposta (peso)	Independente Altura	$\Delta$ altura	Peso-1	$\Delta$ peso
1	1	15	90	...	...	...
1	2	18	90	0	15	3
1	3	19	91	1	18	1
1	4	20	91	0	19	1
2	1	13	101	...	-	...
2	2	15	101	0	13	2
2	3	17	101	1	15	2
2	4	18	103	0	17	1
3	1	9	70	...	...	...
3	2	9	71	1	9	0
3	3	8	72	1	9	1
3	4	9	73	1	8	1

Podemos observar que os três indivíduos hipotéticos possuem quatro medidas cada e todas mensuradas com o mesmo intervalo de tempo. Neste exemplo o objetivo é prever a variação do peso pela variação da altura ao longo do tempo.

A estratégia 1 é o modelo de efeitos aleatório normalmente utilizado, onde o peso de um indivíduo ao longo do tempo (coluna 3 do quadro 4) é previsto por informações de altura também ao longo do tempo (coluna 4 do quadro 4).

Na estratégia 2 o peso ao longo do tempo (coluna 3 do quadro 4) é previsto pela variação da altura entre o momento atual e momento anterior (coluna 5 do quadro 4). Na estratégia 3 o peso é previsto pelo peso no momento anterior (coluna 6 do quadro 4) e

pela altura no momento atual (coluna 4 do quadro 4). A inclusão de uma variável endógena na equação ajuda a controlar para efeito das variáveis não medidas (MENARD, 1991 apud VIEIRA, 2006).

Nas estratégias 4 e 5 o objetivo é prever a variação de peso (coluna 7 do quadro 4). Na estratégia 4 a predição é baseada na altura (coluna 4 do quadro 4) e na estratégia 5 pela variação de altura (coluna 5 do quadro 4).

No quadro 5 são apresentadas as mesmas estratégias do quadro 3 porém com comentários sobre as implicações de aplicá-las na análise de dados desbalanceados. Como citado anteriormente o objetivo é prever a evolução da dilatação da pelve renal (que caracteriza a hidronefrose) medida repetidamente ao longo do tempo. Entretanto, o número de medidas por indivíduo e o intervalo entre as medidas não é o mesmo o que caracteriza dados desbalanceados.

A estratégia 1, modelo de efeitos aleatórios, não apresenta dificuldades de aplicação em dados desbalanceados. Já na estratégia 2 a variação de altura deve ser ajustada de alguma forma pois o intervalo entre as medidas não é uniforme. Uma solução é padronizar estas variações pelo intervalo de tempo.

Na estratégia 3, onde a medida da pelve renal no tempo anterior será considerada como uma das variáveis preditoras surge uma dúvida em como ajustar esta medida pois o tempo entre o valor “atual” e valor “anterior” é diferente. Como padronizar este valor? A imputação seria uma opção adequada?

Nas estratégias 4 e 5 tanto a variável dependente como as variáveis preditoras devem ser padronizadas pelo intervalo de tempo.

QUADRO 5 Estratégias de análise no presente estudo

	Fórmula*	Dificuldade em dados desbalanceados
1	$Pelverenal = \beta_0 + \beta_1 tempo + \beta_2 altura$	Sem problemas
2	$Pelverenal = \beta_0 + \beta_1 tempo + \beta_2 (\Delta altura)$	Como o intervalo do tempo é diferente entre as medidas uma solução será a padronização, ou seja, dividir a variação entre dois momentos pelo tempo transcorrido entre eles.
3	$Pelverenal = \beta_0 + \beta_1 pelverenal_{(j-1)} + \beta_2 tempo + \beta_3 altura$	Como usar esta estratégia em dados desbalanceados pois o intervalo de tempo entre a medida atual e medida anterior não será constante. Padronizar a variação entre medida atual e medida anterior pelo intervalo do tempo? Fazer imputação?
4	$\Delta Pelverenal = \beta_0 + \beta_1 tempo + \beta_2 altura$	Usar padronização pelo intervalo do tempo como descrito na estratégia 3
5	$\Delta Pelverenal = \beta_0 + \beta_1 (\Delta tempo) + \beta_2 (\Delta altura)$	Usar padronização pelo intervalo do tempo como descrito na estratégia 3

\* em todos os modelos as variáveis  $\beta_{\text{sexo}} + \beta_{\text{ITU}} + \beta_{\text{uropatia}} + \beta_{\text{bilateral}}$  estarão presentes mas não são apresentadas porque são constantes

Neste estudo não será considerada ao mesmo tempo a variável idade e a variável tempo ao contrário de estudos em que se procura identificar o efeito histórico (VIEIRA, 2006).

## 5 CONSIDERAÇÕES FINAIS

O próximo passo deste trabalho é manipular o banco de dados e ajustar os modelos apresentados na seção 4. Em seguida tais modelos serão interpretados e comparados com a intenção de verificar influência da modelagem nos coeficientes das variáveis preditoras. Serão propostas formas de ajuste de dados desbalanceados como padronização pelo intervalo de tempo e por fim discutir os benefícios e desvantagens de cada estratégia.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

APOCALYPSE GT, OLIVEIRA EA, RABELO EAS, DINIZ JS, MARINO VS, PEREIRA AK, SIMAL CJ, GAZOLLA LP, FAGUNDES TA. Outcome of apparent ureteropelvic obstruction identified by investigation of fetal hydronephrosis. *Int Urol Nephrol* 2003; 35 (4): 441-448.

BARACHO SMLN. Tratamento de dados ausentes em estudos longitudinais. Dissertação de mestrado, Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

BELLOCO R. *Analysis of longitudinal data in Stata, Splus and SAS*. Department of Medical Epidemiology, Karolinska Institutet. Stockholm, Sweden. March 12, 2001. E-mail: [rino@mep.ki.se](mailto:rino@mep.ki.se)

BOUZADA MC, OLIVEIRA EA, PEREIRA AK, LEITE HV, RODRIGUES AM, FAGUNDES LA, GONÇALVES RP, PARREIRAS R. Diagnostic accuracy of fetal renal pelvis anteroposterior diameter as a predictor of uropathy: a prospective study. *Pediatr Radiol*. 2004; 34 (10): 798-804.

DIGGLE PJ, LIANG KY, ZEGER SL. *Analysis of longitudinal data*. Clarendon Press: Oxford, 1994.

EKUMA O, LIX L. *Random effects models for longitudinal data – continuous data*. Manitoba Centre for Health Policy. May, 2004.

FAUSTO MA. Avaliação longitudinal do crescimento de lactentes nascidos de mães vivendo com HIV/AIDS da coorte de Belo Horizonte [tese de Doutorado]. Belo Horizonte: UFMG; 2005.

GOLDSTEIN H, BROWNE W, RABASH J. Multilevel modelling of medical data. *Statist. Med.*, 2002.

GOULART EMA, CORRÊA EJ, LEÃO E, XAVIER CC, ABRANTES MM. Avaliação do crescimento. In: Leão E, Corrêa EJ, Viana MB, Mota JAC. *Pediatria Ambulatorial*. 4ª edição. Belo Horizonte: COOPMED. 2005.

MOLENBERGHS G, VERBEKE G. *A Review on Linear Mixed Models for Longitudinal Data, Possibly Subject to Dropout*. 2001. URL: [http:// biostatistik.uibk.ac.at/roes/papers/RoES\\_2001\\_Molenberghs\\_Verbeke\\_Paper.pdf](http://biostatistik.uibk.ac.at/roes/papers/RoES_2001_Molenberghs_Verbeke_Paper.pdf) (consultado em 30 de julho de 2003)

OLIVEIRA EA, RABELO EA, PEREIRA AK, DINIZ JSS, CABRAL AC, LEITE HV, SILVA JMP, FAGUNDES TA. Prognostic factors in prenatally-detected posterior urethral valves: a multivariate analysis. *Pediatr Surg Int*. 2002; 18 (8): 662-667.

RABELO EAS, OLIVEIRA EA, DINIZ JS, SILVA JMP, FILGUEIRAS MT, PEZZUTI IL, TATSUO ES. Natural history of multicystic kidney conservatively managed: a prospective study. *Pediatr Nephrol*. 2005; 20 (5): 690-691.

RABELO EAS, OLIVEIRA EA, SILVA GS, PEZZUTI IL, TATSUO ES. Predictive factors of ultrasonographic involution of prenatally detected multicystic dysplastic kidney. *BJU Int*. 2005; 95 (6): 868-871.

RABELO EAS, OLIVEIRA EA, SILVA JM, BOUZADA MC, SOUSA BC, ALMEIDA MN, TATSUO ES. Conservative management of multicystic dysplastic kidney: clinical course and ultrasound outcome. *Jornal de Pediatria (Rio J)* 2005; 81(5): 400-404.

RABELO EAS, OLIVEIRA EA, SILVA JMP, OLIVEIRA DS, COLOSIMO EA. Ultrasound progression of prenatally detected multicystic dysplastic kidney. *Urology*. 2006; 68 (5): 1098-1102.

SILVA JMP, DINIZ JS, OLIVEIRA EA, CARDOSO LS, MARINO VS, PIMENTA MR, MATOS CC, VIEIRA SB. Features of primary vesicoureteral reflux and renal damage in children at a single institution in Brazil from 1969 to 1999. *Int Urol Nephrol*. 2003; 35 (2): 161-168.

SILVA JMP, DINIZ JSS, OLIVEIRA EA, CARDOSO LSB, MARINO VS. Infecção do trato urinário. In: LEÃO E, CORRÊA EJ, MOTA JAC, VIANA MB. *Pediatria Ambulatorial*. 4ª edição COOPMED. Belo Horizonte. 2005: 600-609

SILVA JMP, DINIZ JSS, MARINO VS, LIMA EM, CARDOSO LS, VASCONCELOS MA, OLIVEIRA EA. Clinical course of 735 children and adolescents with primary vesicoureteral reflux. *Pediatric Nephrology* 2006; 21(7): 981-988.

SILVA JMP, DINIZ JSS, SILVA AC, AZEVEDO MV, PIMENTA MR, OLIVEIRA EA. Predictive factors of chronic kidney disease in severe vesicoureteral reflux. *Pediatric Nephrology*. 2006; 21 (9): 1285-1292.

SILVA JMP, OLIVEIRA EA, DINIZ JS, BOUZADA MC, VERGARA RM, SOUZA BC. Clinical course of prenatally detected primary vesicoureteral reflux. *Pediatric Nephrology* 2006; 21 (1) 86-91.

VERBEKE G, MOLENBERGHS G. *Linear mixed models for longitudinal data*. Springer. 2000. New York.

VIEIRA MDT. Evaluating different longitudinal modelling strategies: an application to data from the BHPS. Seminário no Departamento de Estatística da UFMG. Belo Horizonte: 15 de dezembro de 2006.