

# Introdução aos modelos hierárquicos (multiníveis)

Prof. Caio Azevedo

# Introdução

- Estatística: área do conhecimento/Ciência que trata de metodologias (análise de dados) apropriadas para se coletar, organizar e analisar dados.
- A Estatística incorpora elementos de Probabilidade, Matemática, Computação e Ciência de dados (envolve outras formas de análise de dados), desenvolvendo (novas) metodologias.
- A Estatística é uma ferramenta muito importante na resolução de problemas levantados em diversas áreas: Biologia, Psicometria, Educação, Medicina, Física, Computação entre outras.
- É importante que o Estatístico participe de todas as etapas de um estudo (pesquisa/consultoria).

# Etapas para a resolução de um problema

- 1 Determinação do problema/objeto de estudo (incluindo a população de interesse).
- 2 Determinação dos objetivos (gerais e específicos).
- 3 Determinação do tamanho da amostra-delineamento amostral/experimental.
- 4 Levantamento dos dados: entrevistas, experimento, coleta de dados etc.
- 5 Análise Descritiva.
- 6 Análise Inferencial (Modelos de regressão).
- 7 Conclusões e elaboração dos relatórios/artigos/trabalhos pertinentes.

Pode-se retornar a pontos anteriores ou mesmo avançar,

desconsiderando-se alguns pontos, consoante a necessidade.



# Introdução

- Foram vistas no Bacharelado em Estatística, até o momento, diversas ferramentas de análise: descritiva, probabilística e inferencial.
- Estudaremos como analisar a influência de uma ou mais variáveis (variáveis explicativas, covariáveis, variáveis explanatórias) em uma variável de interesse (variável resposta ou resposta), dentro de uma estrutura de dados hierárquicos (multinível).
- Na disciplina de Análise de regressão (ME 613) foram discutidas metodologias mencionadas no item anterior, porém sem a presença (ou sem considerar) estruturas hierárquicas (multinível).

# Introdução

- Nos focaremos em extensões dos modelos modelos de regressão normal linear homocedásticos (simples/múltipla), vistos nas disciplinas de Análise de Regressão (ME 613).
  - Graduação: [página do curso de ME613](#).
  - Mestrado: [página do curso de MI406](#).

# Introdução

- Extensões dos modelos apresentados na disciplina de Planejamento e Pesquisa (ME623, [página do curso de ME623](#)) também serão consideradas.
- Técnicas de amostragem (ME430, [página do curso de ME430](#)) também serão discutidas. Com efeito, estruturas hierárquicas, via de regra, guardam relação com planejamentos amostrais complexos ([pdf](#), [bookdown do livro](#)).

# Pré-requisitos

- Cálculo diferencial e integral: [Cálculo I](#), [Cálculo II](#), [Cálculo III](#).
  - Probabilidade I : [página do curso de Probabilidade I](#)
  - Probabilidade II : [página do curso de Probabilidade II](#)
- Inferência (Graduação/Mestrado)
  - Graduação : [página do curso de ME 419/ME 420](#)
  - Mestrado: [página do curso de MI 402](#)
- Distribuição normal multivariada: [aula](#).

# Notações

- Variável aleatória: quantidade desconhecida e observável, representada por uma letra (latina) maiúscula:  $X, Y, Z$ . Respectivo valor observado, representado por uma letra (latina) minúscula:  $x, y, z$ .
- Parâmetro: objeto de interesse, quantidade desconhecida e não observável, representado por uma letra grega minúscula:  $\theta, \gamma, \beta$ .
- Vetor (negrito). Vetor aleatório  $\mathbf{X} = (X_1, \dots, X_n)'$ ; vetor paramétrico  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$  (vetores coluna).
- Tamanho da amostra:  $n$ . Tamanho da população:  $N$  (em geral, trabalharemos com populações de tamanho infinito ou infinitamente grande).



# Notações

- Número de parâmetros:  $p$  (eventualmente,  $p = 1$ ).
- Função de probabilidade (caso discreto) ou função densidade (de probabilidade), ou simplesmente densidade (caso contínuo) (fdp):  
 $f_X(.) \equiv f_X(., \theta) \equiv p_X(.) \equiv p_X(., \theta)$  (o subíndice pode ser, eventualmente suprimido).
- Função de distribuição acumulada (fda):  $F_X(.) \equiv F_X(., \theta)$
- O subíndice (em todas as quantidades pertinentes) pode ser, eventualmente suprimido.
- Quando do cálculo das duas quantidades acima, em algum valor de interesse, digamos “ $x$ ”, o símbolo “.” será substituído por este ( $x$ ).

# Revisão de Cálculo de probabilidades

- Sejam  $X$ ,  $Y$  e  $Z$  va's definidas em um mesmo espaço de probabilidade.

$$p(x|y) = \begin{cases} \frac{p(x, y)}{p(y)}, & \text{se } p(y) > 0, \\ 0, & \text{se } p(y) = 0 \end{cases}$$

$$p(x) = \begin{cases} \sum_y p(x, y), & \text{se } y \text{ for discreto,} \\ \int_{\Omega_y} p(x, y) dy, & \text{se } y \text{ for contínuo} \end{cases}$$

## Revisão de Cálculo de probabilidades

- $E_X(X) = E_Y(E_{X|Y}(X|Y))$  e  
 $V_X(X) = E_Y[V_{X|Y}(X|Y)] + V_Y[E_{X|Y}(X|Y)]$ .
- $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$
- X e Y são independentes se e somente se  $p(x, y) = p(x)p(y)$   
( $\rightarrow p(x|y) = p(x)$  e  $p(y|x) = p(y)$ ).
- X e Y são condicionalmente independentes dado Z se e somente se  
 $p(x, y|z) = p(x|z)p(y|z)$  ( $\rightarrow p(x|y, z) = p(x|z)$  e  
 $p(y|x, z) = p(y|z)$ ).

# Estrutura Estatística

- Considere um espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P}_\theta)$ , em que :
  - $\mathcal{X}$  : suporte do modelo estatístico ou da família de distribuições de probabilidade ( $\mathcal{P}_\theta$ ) -  $\mathcal{X} = \{\mathbf{x} \in \mathcal{R}^n, f_{\mathbf{X}}(\mathbf{x}; \theta) > 0\}$ .
  - $\mathcal{A}$  :  $\sigma$ -álgebra (conjunto que contém todos os subconjuntos de (interesse) de  $\mathcal{X}$ ).
  - $\theta$ : parâmetro de interesse (possivelmente um vetor, digamos:  $\boldsymbol{\theta}$ ).
  - $\mathcal{P}_\theta$  : família de medidas de probabilidade (família de distribuições de probabilidade)-  $\mathcal{P}_\theta : \mathcal{A} \rightarrow [0, 1], \forall \theta \in \Theta$ , em que  $\Theta \in \mathcal{R}^p$  é o espaço paramétrico (conjunto de possíveis valores que o parâmetro pode assumir).

# Continuação

- Suporte da fdp:
  - $\Omega = \{x : f_X(x; \theta) \geq 0\}$ ,  $f_X(\cdot; \theta)$  é a fdp associada à fda  $F_X(\cdot; \theta)$ .
- Amostra aleatória (aa):  $X_1, \dots, X_n$  independentes e distribuídas segundo  $X \sim F_X(\cdot; \theta)$ . Como trabalharemos com modelos de regressão, em geral, as v.a.'s de interesse não serão identicamente distribuídas.
- Trabalhar-se-á tanto com distribuições condicionais, e.g.,  $X|\theta$  quanto marginais,  $X$ . Ou seja, poderemos ter, p.e., aa  $X_i$ , como  $X_i|\theta$ ,  $i = 1, 2, \dots, n$ .

# Dados/modelos hierárquicos (multiníveis)

- Literatura Americana (Pan-Americana): dados ou modelo(s) hierárquico(s).
- Literatura Européia: dados ou modelo(s) multinível(is).
- Os termos (hierárquico e multinível) serão usados de forma indistinta.
- Literatura Anglo-Saxônica: multilevel, hierarhical, clustered (data).
- Dados hierárquicos : apresentam alguma estrutura (natural/intrínseca) de hierarquia (vários níveis).
  - Esta estrutura pode ser intrínseca ao problema ou induzida pelo planejamento amostral e/ou planejamento experimental.

# Dados/modelos hierárquicos (multiníveis)

- Em geral, principalmente com o advento do “Big data” e (res)surgimento da nomenclatura “Data Science”, os conjuntos de dados, das mais diversas áreas, costumam apresentar estruturas hierárquicas (por diversas razões, tais estruturas não são consideradas, ou o são de forma pouco apropriada).
- UA: unidade amostral - elementos selecionados de uma população, através de algum plano amostral.
- UE: unidade experimental - elementos não são selecionados (de uma população), através de algum plano amostral, mas sim, submetidos a algum experimento.

# Dados/modelos hierárquicos (multiníveis)

- UAE: unidade amostral e/ou experimental (pelo menos uma das duas características anteriores, UA/UE).
- Estrutura hierárquica: grupos de UAE's agrupadas em outras UAE's que podem ou não formar outros grupos (e assim por diante).
- Dados “não hierárquicos” apresentam apenas um nível. Exemplo: grupo de alunos selecionados de uma lista de chamada.
- Dados hierárquicos apresentam pelo menos dois níveis. Exemplo: grupos de escolas selecionadas a partir de um sistema de referência (base de dados) (nível 2) e, de cada uma delas selecionam-se alunos (nível 1).



# Dados/modelos hierárquicos (multiníveis)

- Note que o tipo de seleção anterior guarda semelhanças com os planejamentos amostrais por estratos e/ou por conglomerados ([link](#)).
- Em geral, quando se tem apenas um nível e  $n$  UAE's, usa-se  $UAE_1, UAE_2, \dots, UAE_n$ .
- Dois níveis:
  - Nível 2 (com  $J$  UAE's):  $UAE_1, UAE_2, \dots, UAE_J$ .
  - Nível 1 (com  $n_j$  UAE's, para cada UAE do nível 2):  $UAE_{j1}, UAE_{j2}, \dots, UAE_{jn_j}, j = 1, 2, \dots, J$ , ou seja  $UAE_{ji}, j = 1, \dots, J; i = 1, 2, \dots, n_j$ .

# Dados/modelos hierárquicos (multiníveis)

- Três níveis:
  - Nível 3 (com  $K$  UAE's):  $UAE_1, UAE_2, \dots, UAE_K$ .
  - Nível 2 (com  $m_k$  UAE's, para cada UAE do nível 3):  $UAE_{k1}, UAE_{k2}, \dots, UAE_{km_k}$ ,  $k = 1, 2, \dots, K$ . Ou seja,  $UAE_{kj}$ ,  $k = 1, \dots, K$ ;  $j = 1, 2, \dots, m_k$ .
  - Nível 1 (com  $n_{kj}$  UAE's, para cada UAE do nível 2, dentro de cada UAE do nível 3):  $UAE_{kj1}, UAE_{kj2}, \dots, UAE_{kjin_{kj}}$ ,  $j = 1, 2, \dots, m_k$ . Ou seja,  $UAE_{kji}$ ,  $k = 1, \dots, K$ ;  $j = 1, 2, \dots, m_k$  e  $i = 1, 2, \dots, n_{kj}$ .
- O nível 1 ocupa a última posição dentro da estrutura hierárquica, o nível 2 ocupa a penúltima posição, ..., e o último nível ocupa a primeira posição.

# Dados/modelos hierárquicos (multiníveis)

- Veremos exemplos ao longo do curso (e nos próximos slides). Nos slides seguintes cada quadrado representa uma UAE (de acordo com o nível indicado), com os respectivos índices.
- Como usual, nos modelos de regressão que serão vistos, consideraremos letras do alfabeto latino ( $Y, X, \dots$ ) para representar as variáveis (ao invés de usar “UAE”). Os parâmetros, via de regra, serão representados por letras gregas ( $\theta, \beta, \alpha, \psi \dots$ ).

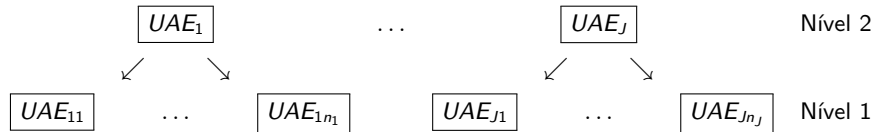
# Dados/modelos hierárquicos (multiníveis)

- Variável aleatória - letra maiúscula ( $Y, X, \dots$ ); respectivo valor observado - letra minúscula ( $y, x, \dots$ ), salvo quando explicitado o contrário (como veremos mais adiante, podemos usar letras maiúsculas para representar covariáveis, ainda que elas não sejam aleatórias).

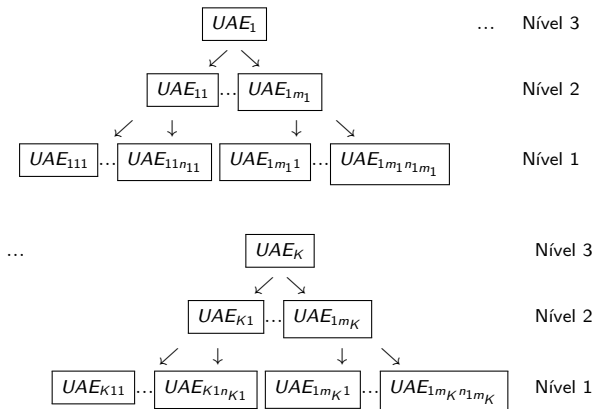
## Estrutura geral: um nível

$UAE_1$   $UAE_2$  ...  $UAE_n$  Único nível

## Estrutura geral: dois níveis



# Estrutura geral: três níveis (não coube na mesma “linha”)

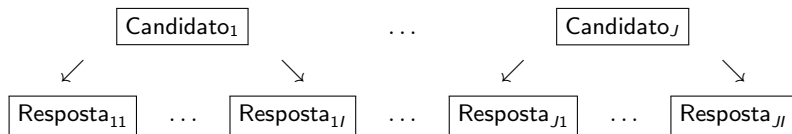


# Exemplos reais

- Exemplo dois níveis: Os candidatos ao vestibular da Unicamp em 2021 ([site da COMVEST](#)). As respostas deles (nível 1) aos itens (questões) das provas (em particular da prova da Fase 1) estão agrupados nos candidatos (nível 2). Ou seja, dados multivariados (mais de uma resposta) apresentam, naturalmente, uma estrutura hierárquica ([página do curso de Métodos em Análise Multivariada \(ME 731\)](#)).
- Veja próximo slide.



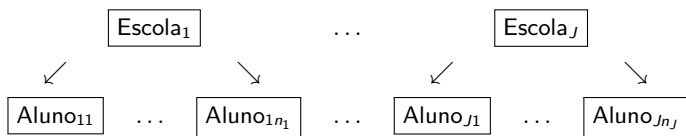
## Cont. slide anterior



# Exemplos reais

- Exemplo dois níveis: **SARESP**: Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo. Sua estrutura de amostragem é complexa mas, de uma forma simples, temos os alunos avaliados (nível 1), agrupados nas escolas (nível 2).
- Veja próximo slide.

## Cont. slide anterior

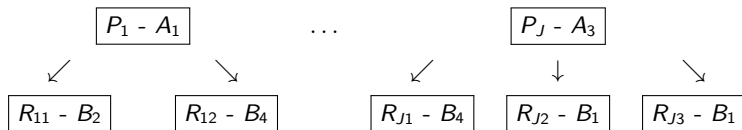


## Exemplos reais

- Exemplo dois níveis (parcela (sub)dividida/split-plot): Nos experimentos em parcela (sub)dividida (split-plot) o tratamento A (com  $a$  níveis,  $A_1, A_2, \dots, A_a$ ) é aleatoriamente distribuído a unidade experimental “maior”, denominada Parcela (nível 2).
- Cada Parcela, por sua vez, é, então, “repartida” e os níveis do tratamento B ( $b$  níveis,  $B_1, B_2, \dots, B_b$ ) são aleatoriamente distribuídos a unidades experimentais menores denominadas subparcelas (nível 1).
- Ou seja, para cada Parcela é atribuída uma combinação definida por um nível do tratamento A e pelo menos um nível do Tratamento B. Veja próximo slide.

## Cont. exemplo anterior

- P: Parcela, R: resposta.  $A_i$  e  $B_j$  representam, respectivamente, os níveis  $i$  do tratamento A e  $j$  do tratamento B.

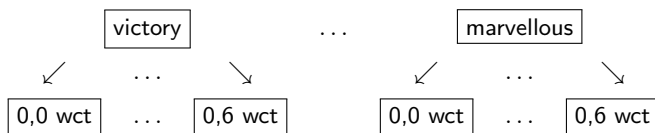


## Exemplos reais

- Exemplo real: No pacote MASS do R, há um banco de dados sob o nome “oats”. O rendimento da aveia em um ensaio de campo em parcelas subdivididas, usando três variedades (V: victory, golden rain, marvellous) e quatro níveis de tratamento com manjares (0,0 wct; 0,2 wct; 0,4 wct; 0,6 wct).
- O experimento foi desenvolvido em 6 blocos de 3 parcelas principais, cada uma dividida em 4 subparcelas. As variedades foram aplicadas nas parcelas principais e os tratamentos adubos nas subparcelas. Desconsiderando-se o bloco (fora um experimento em blocos completos casualizados).

## Exemplos reais

- Exemplo real (cont.): Temos, assim, uma estrutura de parcela dividida (excluindo-se os blocos):



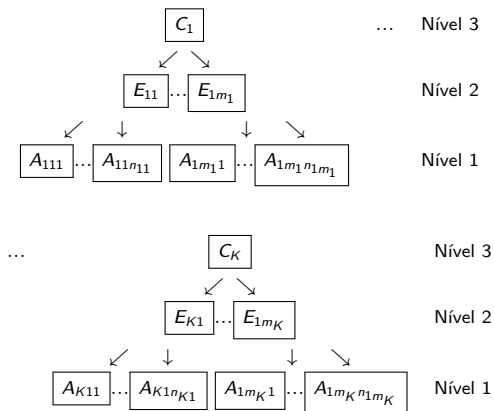
Neste caso, cada parcela (variedades: victory, golden rain, marvellous) comporta todos os níveis do outro fator (sub-parcela): manjares : 0,0 wct; 0,2 wct; 0,4 wct; 0,6 wct).

# Exemplos reais

- Exemplo três níveis: SARESP, novamente. Podemos também considerar, de uma forma simples, que temos os alunos (A) avaliados (nível 1), agrupados nas escolas (E) (nível 2) que, por sua vez, estão agrupados nas cidades (C) do estado de SP (nível 3).



# Exemplos reais



# Dados multiníveis (DM)/ Dados hierárquicos (DH)

- Características comuns (DM):
  - Heterocedasticidade.
  - Diferentes fontes de variabilidade.
  - Dependência entre as UAE's pertencentes aos mesmos grupos.
  - Relação com planos amostrais complexos.
  - As covariáveis podem estar relacionadas com a resposta de formas diferentes, consoante o nível da hierarquia.

# Modelos multiníveis (MM)/ Modelos hierárquicos (MH)

- Características comuns (MM):
  - Permitem incorporar heterocedasticidade.
  - Permitem incorporar e/ou separar diferentes fontes de variabilidade.
  - Permitem incorporar dependência entre as UAE's pertencentes aos mesmos grupos.
  - Permitem incorporar estruturas amostrais/experimentais complexas.
  - Permitem que os efeitos das covariáveis na resposta sejam considerados de forma diferente, consoante o nível da hierarquia.

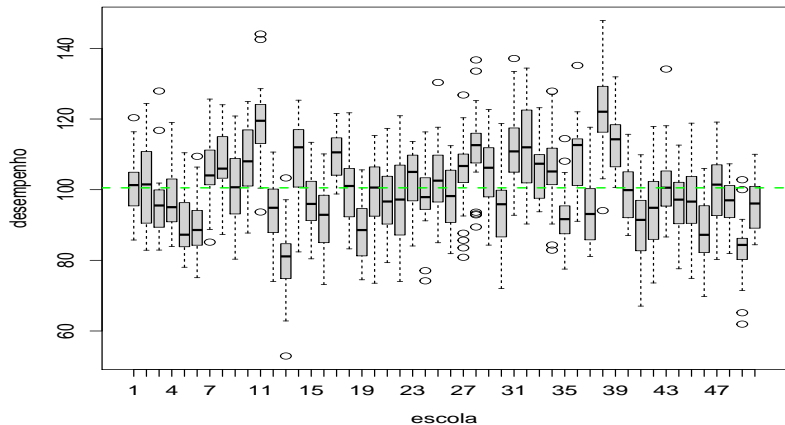
# Problemas em se ignorar a estrutura de dados multiníveis

- Má estimação dos erros-padrão associados aos parâmetros (de regressão) (em geral são subestimados).
- Má estimação dos p-valores associados a testes de hipótese de interesse. Pelo exposto no item anterior, há uma tendência de se identificar significância, quando não há.
- Limitações em se entender o comportamento da variável resposta em função de variáveis explicativas (uma mesma covariável pode influenciar de forma diferente, a resposta, consoante o nível em que se está medindo esse efeito).
- Estruturas de dependência não são estudadas/consideradas, o que pode limitar a compreensão do fenômeno em estudo.

## Exemplo artificial ( “Toy example” )

- Considere um conjunto de  $J = 50$  escolas, cada uma com  $n_j$  alunos,  $j = 1, 2, \dots, 50$  .
- Objetivo: comparar o desempenho dos alunos concludentes do 9º ano do Ensino Fundamental II, dessas escolas, com relação ao conhecimento em Matemática (esperado para tais concludentes).
- Aplicou-se uma prova para cada um deles e uma nota fora gerada (métodos psicométricos), em uma determinada escala. Importante, quanto maior a nota, maior o conhecimento!
- Pergunta: como atingir esse objetivo ?

# Box-plots das notas dos alunos por escola



# Comentários (descritivos) sobre o box-plot

- A linha verde tracejada representa a média do desempenho de todos os alunos.
- As distribuições das notas entre as escolas são diferentes (efeito de escola).
- Os alunos dentro de cada escola parecem apresentar desempenhos diferentes (variabilidade intra escola/efeito de aluno).
- A variabilidade das notas entre os alunos de cada escola parecem ser diferentes entre as escolas (variabilidade entre escolas).
- A maioria das distribuições das notas em cada escola parece ser (aproximadamente) simétrica, com a presença de poucos ou nenhum

# Algumas formas de resolver o problema

- Testes (paramétricos) de comparações múltiplas: Duncan, Dunnett, Scheffé etc.
  - Vantagens: menos suposições em relação aos modelos de regressão.
  - Desvantagens: limitação na incorporação das características dos dados, controle do nível de significância tem de ser feito de forma cuidadosa e as conclusões podem ser de difícil interpretação.
  - Veja [aula sobre testes de comparação múltipla](#) e [pacote do R: DescTools](#).



# Algumas formas de resolver o problema

- Testes de hipótese não-paramétricos: Wilcoxon, Kolmogorov-Smirnov, Cramer-Von Mises etc.
  - Vantagens: menos suposições em relação aos modelos de regressão e em relação aos testes paramétricos de comparação múltipla.
  - Desvantagens: limitação na incorporação das características dos dados, controle do nível de significância tem de ser feito de forma cuidadosa e as conclusões podem ser de difícil interpretação.
  - Veja [livro Nonparametric Statistical Methods Using R](#) e [pacote do R: nparcomp](#).

# Algumas formas de resolver o problema

- Modelos de regressão: essencialmente, essa será a abordagem considerada neste curso.
  - Vantagens: incorpora mais características dos dados e mais informação, interpretação mais fácil e, dependendo do procedimento adotado, há um controle maior sobre o nível de significância dos testes e dos resultados obtidos.
  - Desvantagens: requer mais suposições e métodos de inferência mais sofisticados.
  - Em linhas gerais, à medida que as suposições dos modelos forem “mais” válidas, os modelos de regressão apresentam melhor desempenho (dependendo do critério) do que os outros dois métodos.



# Modelo 1

- Modelo de regressão normal linear homecedástico (MRNLH) (ME 613), parametrização casela de referência (ME 623).

$$Y_{ji} = \mu + \alpha_j + \xi_{ji}, j = 1, 2, \dots, 50 \text{ (escola)}; i = 1, 2, \dots, n_j \text{ (aluno)}$$

- Erros  $\xi_{ji} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ ,  $\mu, \alpha_j$  são não aleatórios.
- $\mathcal{E}_{\xi_{ji}}(Y_{ji}) \equiv \mathcal{E}_{\xi_{ji}}(Y_{ji}) = \mu_j, \mathcal{V}_{\xi_{ji}}(Y_{ji}) \equiv \mathcal{V}(Y_{ji}) = \sigma^2$ .
- $\mu_j = \mu + \alpha_j$  : média populacional do desempenho da j-ésima escola, em que  $\alpha_1 = 0$  (a escola 1 é a casela de referência). Assim,  $\mu = \mu_1$ .

# Modelo 1

- $\alpha_j = \mu_j - \mu_1$  é o incremento/diferença (positivo(a) ou negativo(a)), entre a média do desempenho da  $j$ -ésima escola e da escola 1.

Assim, se for positivo (negativo) indica que a escola  $j$  possui média maior (menor) do que a escola 1.

- Portanto, temos que:  $Y_{ji} \stackrel{ind.}{\sim} N(\mu + \alpha_j, \sigma^2)$ .
- Defina  $\beta = (\mu, \alpha_2, \dots, \alpha_{50})'$ .

# Modelo 1

- Este modelo assume:
  - Homocadasticidade [nesse caso, iguais variâncias (em relação as combinações) alunos e escolas]. Em particular, entre as escolas e entre os alunos.
  - Normalidade (dos desempenhos dos alunos de cada escola)
  - Independência entre as observações (mesmo para alunos da mesma escola).
- Além disso, ele possui muitos parâmetros ( $51 = 50$  parâmetros de regressão ( $\beta$ ) + 1 parâmetro de dispersão ( $\sigma^2$ )).

# Modelo 1

- Para se comparar os desempenhos dos alunos entre as escolas (desempenhos das escolas) pode ser necessário realizar vários testes de hipótese.

- Uma hipótese (estatística) inicial, de interesse é

$$H_0 : \alpha_2 = \alpha_3 = \dots = \alpha_{50} = 0$$

$$H_1 : \text{há pelo menos um } \alpha_j, j = 2, 3, \dots, 50 \text{ diferente de zero}$$

- Note que, pela estrutura do modelo, se não rejeitarmos  $H_0$ , equivale a dizer que distribuições das notas entre as escolas, não são diferentes (afirmação forte!). Ou seja, a variabilidade dos dados deve-se a outros fatores não considerados (por exemplo, características dos alunos).

# Modelo 1

- Em se rejeitando  $H_0$ , podemos tentar identificar padrões (grupos) de escola que não sejam diferentes entre si, através de metodologias para testar hipóteses do tipo:

$$H_0 : \mathbf{C}_{(q \times p)} \boldsymbol{\beta}_{(p \times 1)} = \mathbf{0}_{(q \times 1)} \text{ vs } H_1 : \mathbf{C}_{(q \times p)} \boldsymbol{\beta}_{(p \times 1)} \neq \mathbf{0}_{(q \times 1)},$$

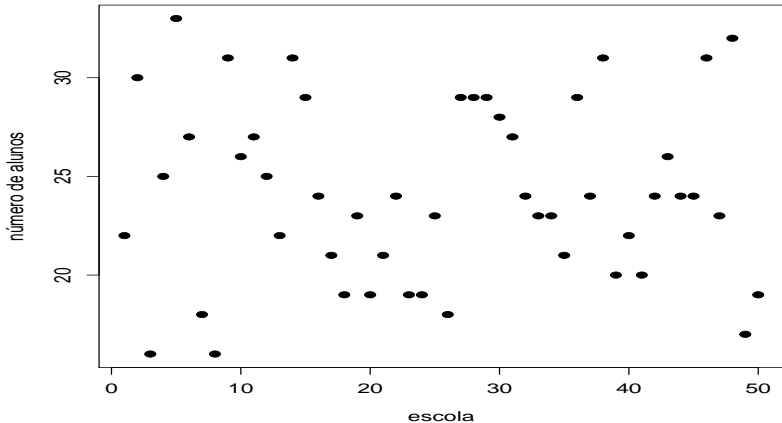
em que, via de regra,  $q \leq p$  e  $\mathbf{C}$  é conhecida e não aleatória. Além disso, as linhas da matriz  $\mathbf{C}$  têm de ser linearmente independentes (do contrário, estar-se-ia testando a(s) mesma(s) hipótese(s) mais de uma vez), que podem ser testadas utilizando metodologias como essa: [Testes do tipo Wald](#).

# Modelo 1

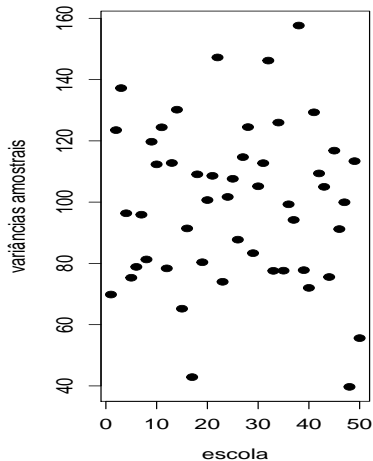
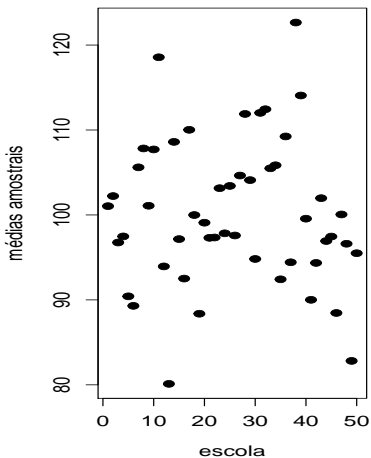
- Tentar identificar todas as diferenças possíveis (comparando todos os pares de escolas) não é apropriado (interpretação dos resultados + nível de significância).
- Contudo, como mencionado, a primeira etapa, depois de se ajustar o modelo aos dados (estimar seus parâmetros), é necessário verificar a qualidade do ajuste.



Número de alunos por escola ( $n = \sum_{j=1}^J n_j = 1207$  total)



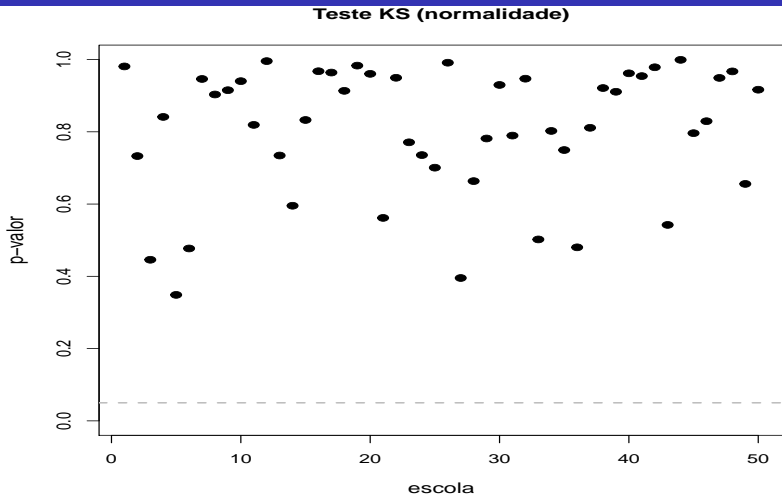
# Médias e variâncias amostrais das notas das escolas



# Médias e variâncias amostrais das notas das escolas

- As médias e as variâncias amostrais sugerem que as respectivas quantidade populacionais sejam diferentes.
- Contudo o teste de Levene para a igualdade das variâncias populacionais indica que elas não são diferentes:  $F = 0,7105$  ( $p = 0,9343$ ). Ou seja, parece que a hipótese de homocedasticidade é razoável.
- Pelo comportamento das variâncias (populacionais), sabemos que mesmo que elas não sejam diferentes, as variâncias amostrais podem apresentar diferenças razoáveis.

# Normalidade dos desempenhos das escolas?



## Cont.

- Não rejeitamos a normalidade para nenhuma das escolas.
- Assim, a hipótese de normalidade (dos erros) parece ser razoável.
- Um ponto interessante a ser explorado (que também tem a ver com possíveis estruturas de dependência dos dados) é o coeficiente de correlação intraclasse (CCI).
- O CCI é uma medida de particular importância no contexto de DH e sua definição (parâmetro) geral é

$$\rho_I = \frac{\psi}{\psi + \sigma^2} \quad (1)$$

em que  $\psi$  é a variância (variabilidade) entre conglomerados (grupos) e  $\sigma^2$  é a variabilidade intra grupos.

## Cont.

- Pode-se estimar  $\psi$  e  $\sigma^2$  de várias formas; por exemplo através dos quadrados médios do modelo (QMM) e residual (QMR) da ANOVA para um Fator ([veja aqui](#)), respectivamente.
- Pode-se ainda definir, e conseqüentemente estimar, o CCI de outras formas (diferente da Equação (1)), veja, por exemplo: ([discussão no contexto de amostragem](#)).
- Para este exemplo o CCI fora estimado (definido) usando três métodos ([link1](#), [link2](#)).

## Cont.

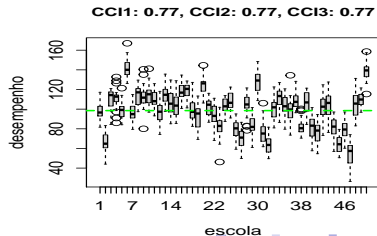
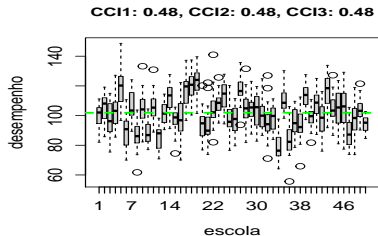
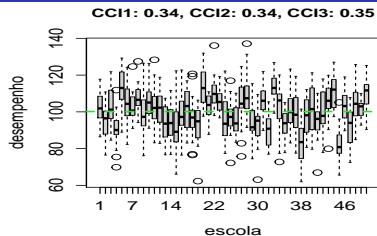
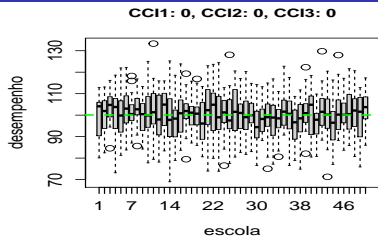
- O primeiro ( $CCI_1$ ): é definido como sendo o coeficiente de correlação de Pearson calculado justapondo-se todos os pares de unidades distintas possíveis (formadas dentro de cada conglomerado) ([veja aqui a partir da página 31](#)).
- O segundo ( $CCI_2$ ):  $1 - \frac{QMR}{QMR+QMM}$ .
- O terceiro ( $CCI_3$ ):  $\frac{QMM-QMR}{QMM+(m-1)QMR}$ , e  $m$  é o tamanho médio ajustado dos conglomerados (veja as referências anteriores).
- Os resultados foram:  $CCI_1 = 0,417$ ;  $CCI_2 = 0,417$ ;  $CCI_3 = 0,422$ .

## Cont.

- Essencialmente, o CCI é uma medida (que varia, em geral entre  $-1/(ng-1)$  e 1 ou entre 0 e 1, dependendo de como ele seja definido, em que  $ng$  é o número de grupos) que quantifica o quanto os grupos são homogêneos entre si. Em geral, quanto mais próximo de 1, mais os elementos de cada grupo são parecidos entre si, o que implica que os grupos tendem a ser (bem) diferentes entre eles.
- Os resultados indicam a existência de uma variabilidade razoável entre os grupos (o que pode indicar uma dependência/similaridade entre os elementos dentro de cada grupo!). Note que o modelo 1 assume independência entre todas as observações.



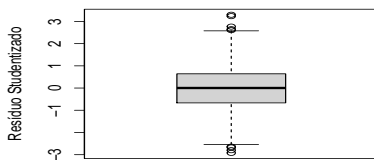
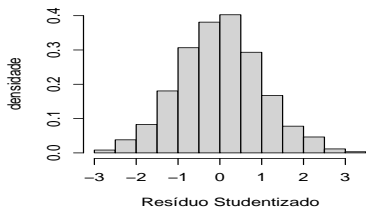
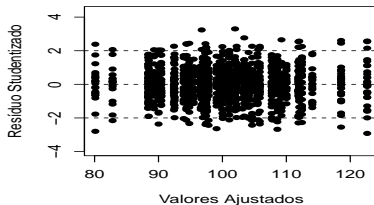
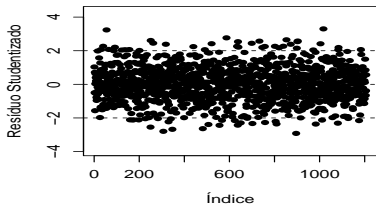
## Situações hipotéticas acerca do CCI (linha verde: média das obs.)



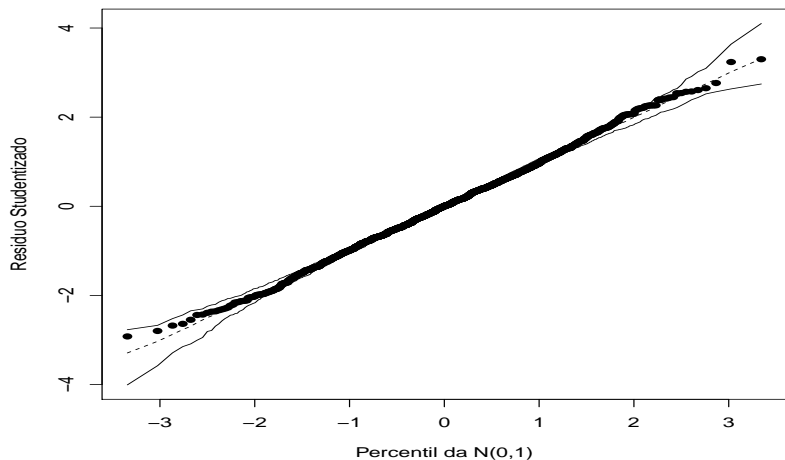
## Cont.

- O modelo 1 foi ajustado através do método de mínimos quadrados ([metodologia](#), [Wikibook sobre o ajuste do modelos de regressão lineares](#)), usando a função `lm` ([manual](#)) (e outras funções, veja o site do curso!) do pacote R e uma análise de diagnóstico a partir do resíduo studentizado ([material sobre resíduo studentizado](#)) fora realizada.

# Gráficos de diagnóstico: resíduo studentizado



## Gráficos de envelopes: resíduo studentizado



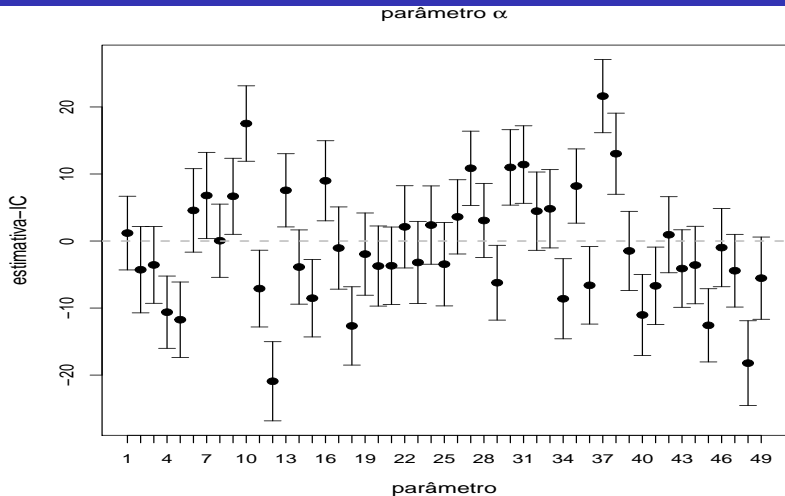
# Sobre o ajuste do modelo 1

- As suposições de normalidade, homocedasticidade e independência parecem estar sendo satisfeitas.
- A distribuição dos resíduos parece ser simétrica com um percentual muito pequeno de (outliers).
- Ausência de padrões (tendência) que acusem heterocedasticidade.
- Ausência de padrões (tendência) que acusem correlação entre as observações.

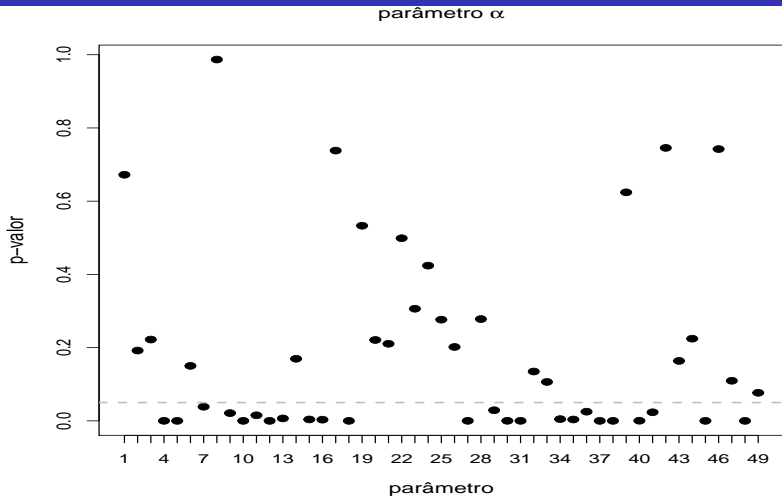
# Sobre o ajuste do modelo 1

- Os resíduos (gráfico de envelopes) estão dispersos, aleatoriamente, dentro das bandas de confiança, embora alguma (level) tendência pode ser observada nas caudas.
- Em geral, o modelo parece estar bem ajustado. Pergunta: será que, realmente, não há nenhum problema?

# Intervalos de confiança para os $\alpha$ 's



# p-valor para os testes de nulidade dos $\alpha$ 's

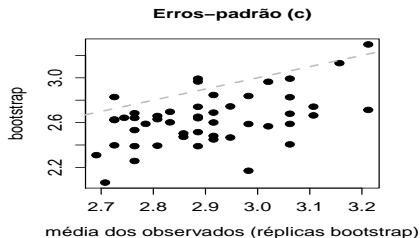
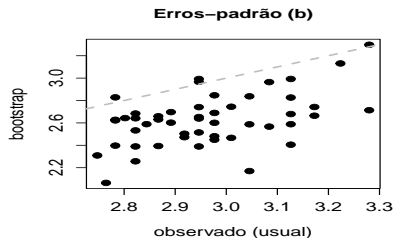
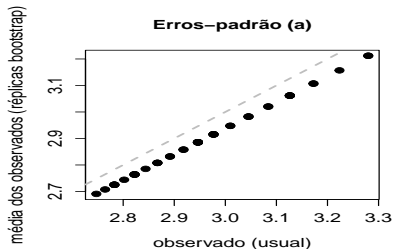




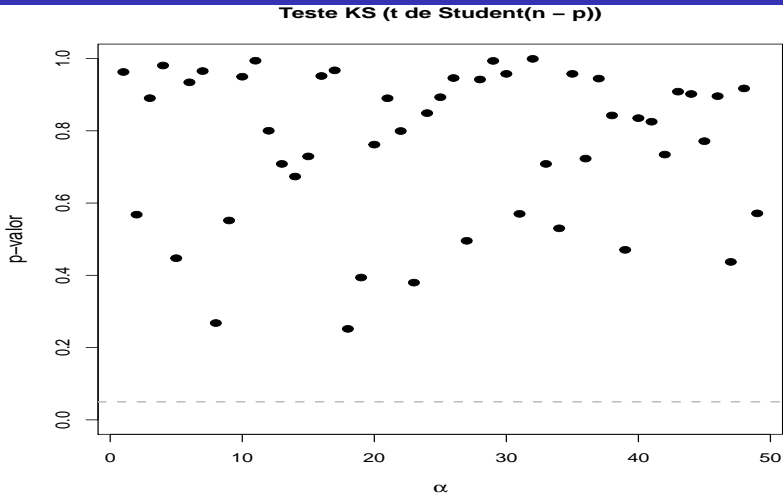
# Comentários sobre os resultados inferenciais

- Podemos notar que alguns parâmetros não são diferentes de zero (p-valor).
- Podemos notar que alguns (grupos) de parâmetros parecem não ser diferentes entre si.
- Vamos reajustar o modelo utilizando Bootstrap não paramétrico ([curso sobre métodos computacionais em Estatística](#), [várias referências](#)), para estudar o comportamento dos estimadores de mínimos quadrados.
- Utilizou-se  $B = 1.000$  amostras Bootstrap, com reposição e de mesmo tamanho (quantidade de alunos e escolas) em relação aos dados originais. Os resultados seguem:

# Erros-padrão: observados e via bootstrap



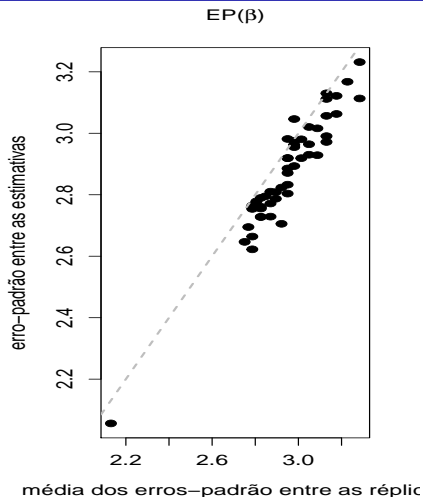
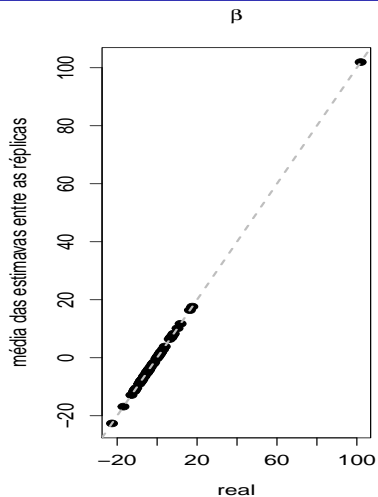
# Teste de aderência dos $\beta$ 's a distribuição $t_{(n-p)}$



# Exercício de simulação

- Para ilustrar melhor as limitações do modelo 1, desenvolveu-se o seguinte exercício de simulação:
  - Simulou-se  $R = 1.000$  réplicas (conjuntos de dados) de modo semelhante ao que fora feito para o exemplo em questão.
  - Para cada réplica ajustou-se o modelo 1, do mesmo modo ao que fora feito.
  - Armazenou-se, para cada uma das réplicas, dado o respectivo ajuste, as estimativas dos parâmetros de regressão e os respectivos erros-padrão.

# Alguns resultados



## Modelo 2

- Modelo multinível (hierárquico) de dois níveis

$$Y_{ji} = \beta_{0j} + \xi_{ji}, \text{ (nível 1 - aluno)}$$

$$j = 1, 2, \dots, 50 \text{ (escola)}, i = 1, 2, \dots, n_j \text{ (aluno)}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}, \text{ (nível 2 - escola)}$$

$$\xi_{ji} \stackrel{iid}{\sim} N(0, \sigma^2), u_{0j} \stackrel{iid}{\sim} N(0, \psi), \xi_{ji} \perp u_{0j}$$

- $\gamma_{00}$  é um efeito fixo,  $\beta_{0j}$ ,  $j = 1, 2, \dots, 50$  são efeitos aleatórios e  $\sigma^2$  e  $\psi$  são os componentes de variância (parâmetros de dispersão).

## Modelo 2

- $\mathcal{E}(Y_{ji}|u_{0j}) = \gamma_{00} + u_{0j}$  (média condicional do desempenho do i-ésimo aluno da j-ésima escola). Ou seja, valor esperado do desempenho de um aluno da j-ésima escola.
- $\mathcal{E}(Y_{ji}) = \gamma_{00}$  (média marginal do desempenho do i-ésimo aluno da j-ésima escola). Ou seja, valor esperado do desempenho de um aluno considerando-se todas as escolas (média de todas as escolas).
- $\mathcal{V}(Y_{ji}|u_{0j}) = \sigma^2$  (variância condicional do i-ésimo aluno da j-ésima escola). Ou seja, variância do desempenho entre os alunos da j-ésima escola.

## Modelo 2

- $\mathcal{V}(Y_{ji}) = \psi + \sigma^2$  (variância marginal do  $i$ -ésimo aluno da  $j$ -ésima escola). Ou seja, variância do desempenho entre os todos os alunos.
- Também, temos que:  $\text{Cov}(Y_{ji}, Y_{j'j'}) = \psi$  e  $\text{Corre}(Y_{ji}, Y_{j'j'}) = \frac{\psi}{\psi + \sigma^2}$ ,  $\forall j \neq j'$  (Exercício)
- Além disso,  $Y_{ji}|b_{0j} \stackrel{ind.}{\sim} N(\gamma_{00} + u_{0j}, \sigma^2)$  e  $Y_{ji} \sim N(\gamma_{00}, \sigma^2 + \psi)$  (estes últimos, **não** são independentes) (Exercício)
- Neste caso, as hipóteses de interesse são:  $H_0 : \psi = 0$  vs  $H_1 : \psi > 0$ . Assim, não rejeitar  $H_0$  implica em afirmar que as escolas tem a mesma distribuição de desempenho. Pergunta: como testar essas hipóteses? ([aqui](#), [aqui](#)), entre outras referências ([ref. 1](#)), ([ref. 2](#)).



## Modelo 2

- De uma forma mais geral, temos que

$\mathbf{Y}_j = (Y_{j1}, Y_{j2}, \dots, Y_{jn_j})' \sim N_{n_j}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ ,  $j = 1, 2, \dots, J$ , em que

$$\boldsymbol{\mu}_j = \gamma_{00} \mathbf{1}_{n_j}, \mathbf{1}_{n_j} = (1, 1, \dots, 1)'_{(n_j \times 1)}, \text{ e}$$

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} \sigma^2 + \psi & \psi & \dots & \psi \\ \psi & \sigma^2 + \psi & \dots & \psi \\ \vdots & \vdots & \ddots & \vdots \\ \psi & \psi & \dots & \sigma^2 + \psi \end{bmatrix}_{n_j \times n_j}$$

## Modelo 2

- Podemos notar, assim, que:
  - Os desempenhos médios entre as escolas diferem em termos do efeito aleatório  $u_{0j}$ . Se seu valor for positivo, significa que a escola  $j$  tem um desempenho médio superior a média de todas as escolas. Se negativo, tem um desempenho médio inferior.
  - A variância do desempenho de cada escola é menor do que a variância do desempenho de todas as escolas.
  - As observações (dentro da mesma escola) podem ser correlacionadas (nesse caso, dependentes - devido à normalidade multivariada).

# Parâmetros, efeitos fixos e efeitos aleatórios

- Visão frequentista:
  - Parâmetros: são quantidades fixas (não aleatórias), desconhecidas e não observáveis. Podem ser variáveis (não fixas) mas, não aleatórias.
  - Efeitos fixos: parâmetros para os quais se deseja fazer inferência apenas para os níveis observados. Exemplo: observam-se as temperaturas  $20^{\circ}C$ ,  $25^{\circ}C$  e  $30^{\circ}C$  e tem-se, somente, o interesse em se estudar esses níveis.
  - Efeitos aleatórios: parâmetros para os quais se deseja fazer inferência para os níveis observados e não observados. Exemplo: observam-se 50 escolas, de um total de 200, mas tem-se o interesse em inferir o comportamento para as outras 150.

# Parâmetros, efeitos fixos e efeitos aleatórios

- Visão bayesiana:
  - Parâmetros: são quantidades desconhecidas e não observáveis para as quais se assume um caráter aleatório (mensuração da incerteza acerca do conhecimento sobre eles). à rigor, podem ou não ser variáveis.
  - Efeitos fixos e efeitos aleatórios desempenham papel semelhante ao caso frequentista.
- A distinção entre os efeitos fixos e aleatórios pode se dar pela escolha do analista, consultante, modelagem/metodologia escolhida(s) etc) entre outros fatores (ou seja, não necessariamente pelas características acima).

# Parâmetros, efeitos fixos e efeitos aleatórios

- Trataremos parâmetro (a menos se mencionado ao contrário) como quantidades de interesse desconhecidas, não observáveis, que podem ou não estarem “fixas”.
- Parâmetro “fixo”: altura média da população brasileira neste momento. Parâmetro “aleatório” (não fixo): proporção de pessoas que votarão no candidato “A” nas próximas eleições para prefeito de Campinas.
- Efeitos aleatórios serão tratados como descrito anteriormente (variáveis aleatórias não observáveis).

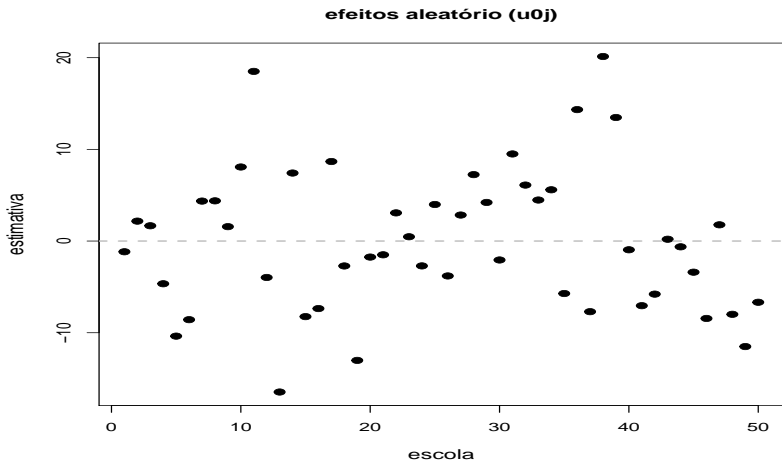
## Ajuste do modelo 2

- Veremos, mais adiante, os detalhes inferenciais (estimação, verificação da qualidade do ajuste, predição, comparação de modelos) necessários sobre a classe de modelos hierárquicos (da qual o modelo 2 faz parte).
- Ajustou-se-o utilizando o pacote “lme4” do programa R.

## Ajuste do modelo 2

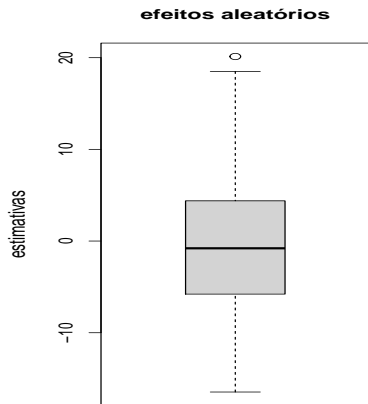
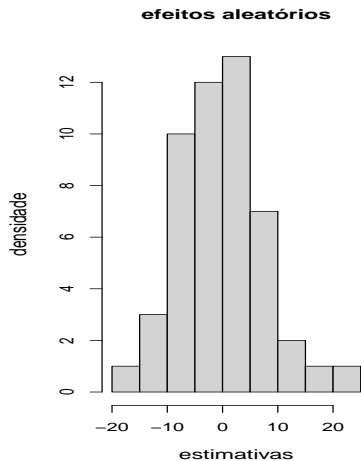
- Uma breve comparação entre os modelos resultou em: Modelo 1 (AIC = **9030,41**; BIC = 9290,30) e Modelo 2 (AIC = 9121,37; BIC = **9136,66**)(maiores detalhes serão apresentados posteriormente). Isso indica que o modelo 2 é preferível ao modelo 1 (contudo, análises de resíduos são necessárias, para o modelo 2, a fim de ser subsídusos suficientes para tal afirmação).
- Comentaremos, em aulas vindouras, sobre seleção/comparação de modelos hierárquicos. Para uma revisão sobre tais metodologias para MRNLH veja : [link](#).
- Mais resultados podem ser encontrados nos próximos slides.

# Predição dos efeitos aleatórios

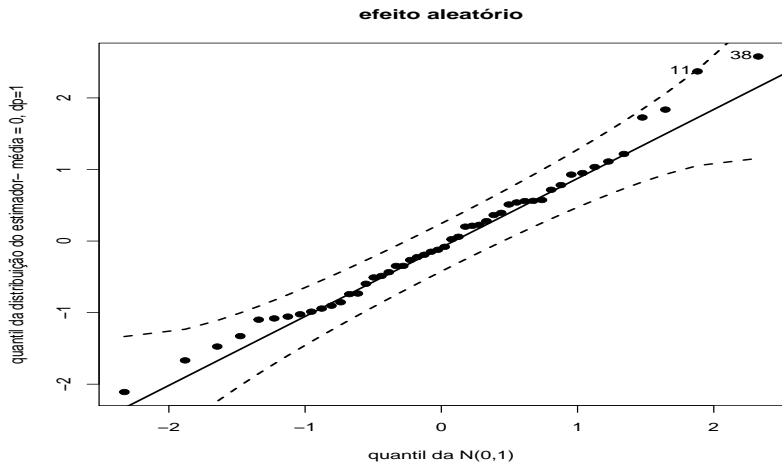




# Histograma e box-plot dos efeitos aleatórios



# Gráfico de quantil-quantil dos efeitos aleatórios



# Comentários sobre os efeitos aleatórios

- Pelos três conjuntos anteriores de gráficos, podemos perceber:
  - Que as escolas apresentam (distribuições de) desempenhos diferentes ( $\psi > 0$ ) (eventualmente podemos agrupar escolas com desempenhos equivalentes).
  - Que temos escolas que apresentam um desempenho médio maior (menor) em torno de 20 pontos em relação à média geral das escolas.
  - Que os efeitos aleatórios ( $u_{0j}$ ) apresentam distribuição (aproximadamente) normal.

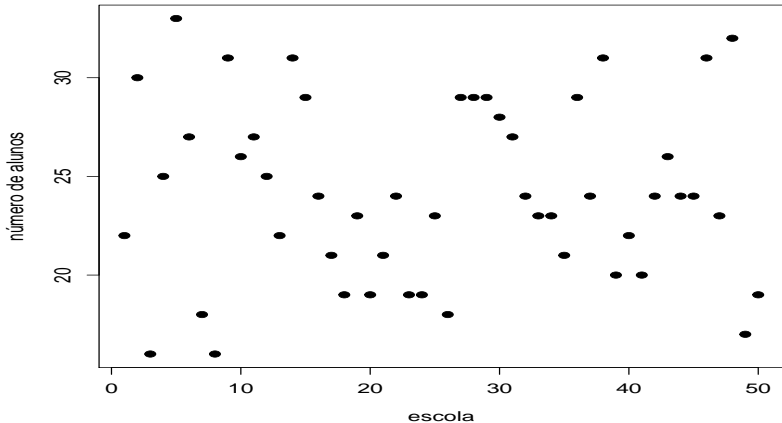
## Exemplo artificial 2 (“Toy example 2”)

- Considere um conjunto de  $J = 50$  escolas, cada uma com  $n_j$ ,  $j = 1, 2, \dots, 50$  alunos.
- Objetivos:
  - Comparar o desempenho (doravante, “desempenho”) dos alunos concludentes do 9º ano do Ensino Fundamental II, dessas escolas, com relação ao conhecimento em Matemática (esperado para tais concludentes).
  - Avaliar o impacto do número (médio) de horas semanais de estudo dos alunos (doravante, “horas”), nos seus desempenhos, entre as escolas.

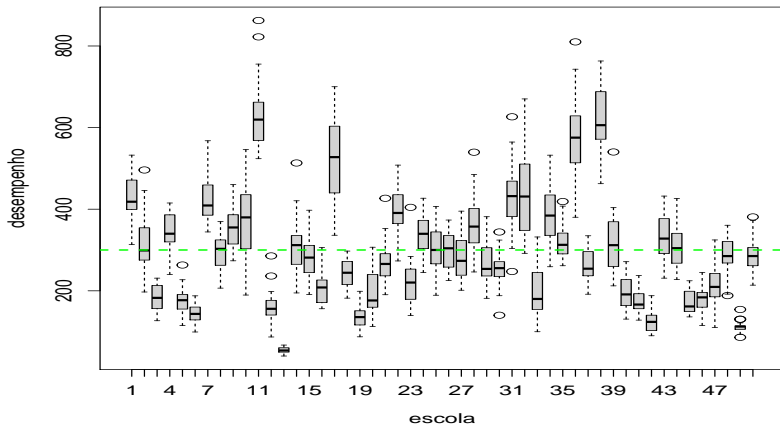
## Exemplo artificial 2 (“Toy example 2”), cont.

- Aplicou-se uma prova para cada um deles e uma nota fora gerada (métodos psicométricos), em uma determinada escala. Importante, quanto maior a nota, maior o conhecimento!
- Pergunta: como atingir esse objetivo ?
- Temos uma situação bastante similar ao EA1 (Exemplo artificial 1). Contudo, neste caso, temos uma covariável (número de horas de estudo semanais).
- De fato, essencialmente, todos os comentários feitos para o EA1 se aplicam a este exemplo.

Número de alunos por escola ( $n = \sum_{j=1}^J n_j = 1207$  total)



# Box-plots do desempenho, por escola

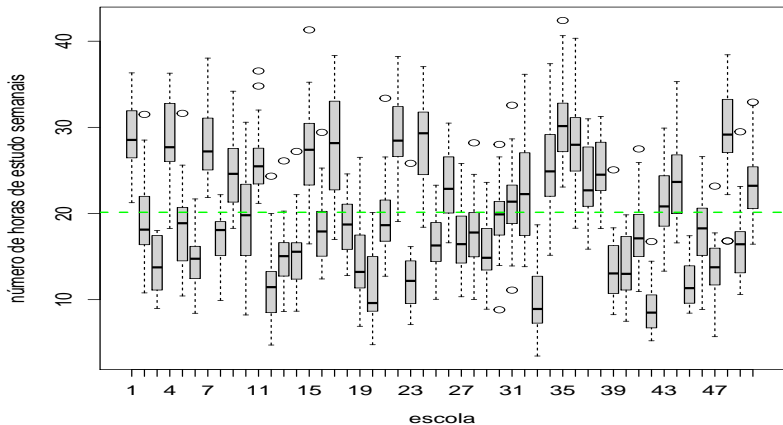


# Comentários (descritivos) sobre o box-plot (desempenho)

- A linha verde tracejada representa a média de todos os alunos.
- As distribuições das notas entre as escolas são diferentes (efeito de escola).
- Os alunos dentro de cada escola parecem apresentar desempenhos diferentes (variabilidade intra escola/efeito de aluno).
- A variabilidade das notas entre os alunos de cada escola parecem ser diferentes entre as escolas (variabilidade entre escolas).
- A maioria das distribuições das notas em cada escola parece ser (aproximadamente) simétrica, com a presença de poucos ou nenhum outlier.



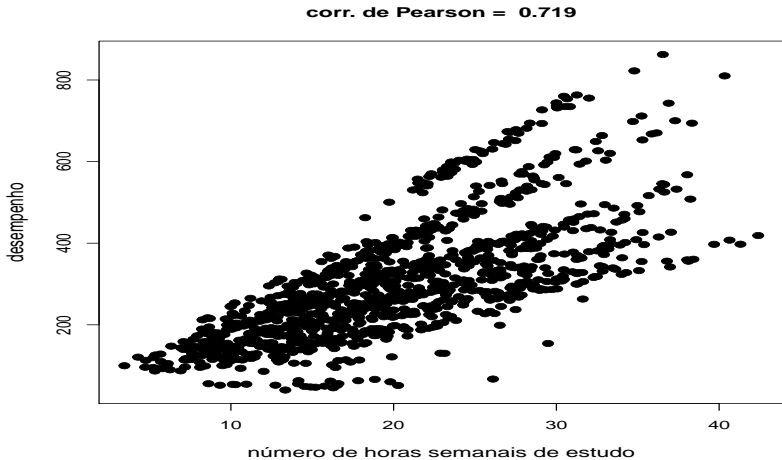
# Box-plots das horas, por escola



# Comentários (descritivos) sobre o box-plot (horas)

- A linha verde tracejada representa a média de todos os alunos.
- As distribuições das horas entre as escolas são diferentes.
- Os alunos dentro de cada escola parecem apresentar horas diferentes (variabilidade intra escola).
- A variabilidade das horas entre os alunos de cada escola parecem ser diferentes entre as escolas (variabilidade entre escolas).
- A maioria das distribuições das notas em cada escola parece não ser simétrica, com a presença de poucos ou nenhum outlier.

# Dispersão entre desempenho e horas (global)



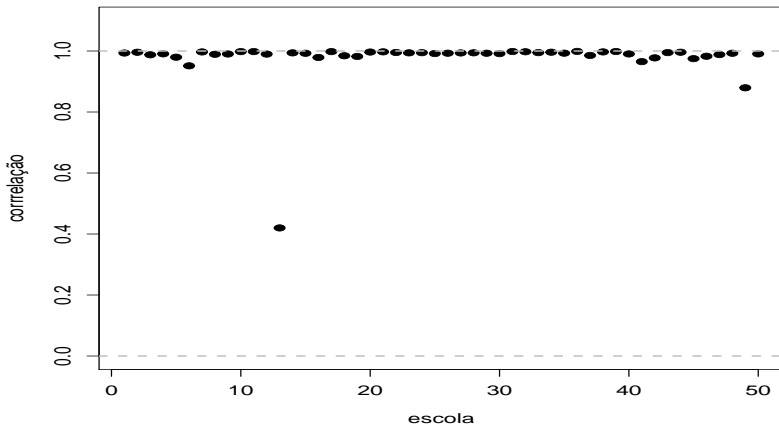
# Comentários (descritivos) sobre o gráfico de dispersão

- Parece haver uma relação linear crescente entre horas e desempenho.
- O gráfico sugere a presença de grupos (“clusters”), pois há vários “padrões de dispersão”.
- Com efeito, os gráficos a seguir endossam essa suspeita (em função das escolas).

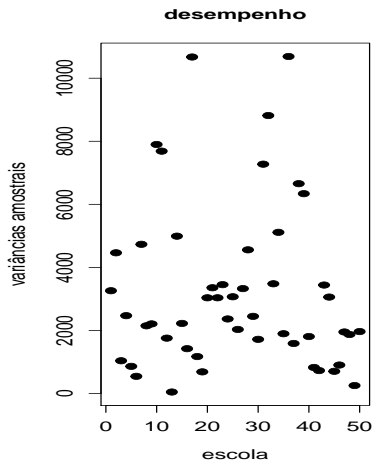
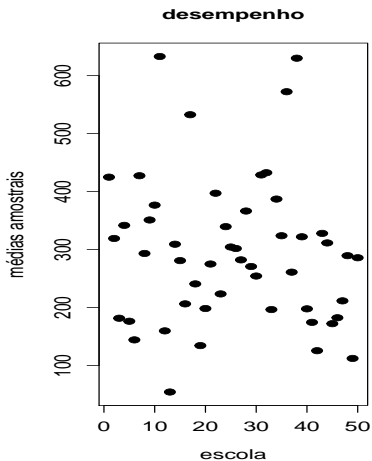
# Dispersão entre desempenho e horas (por escola)

ExemArt2DXYgrupo.pdf

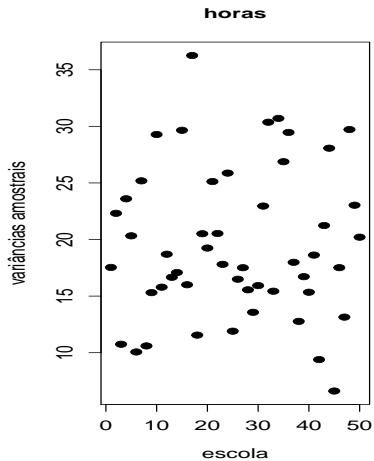
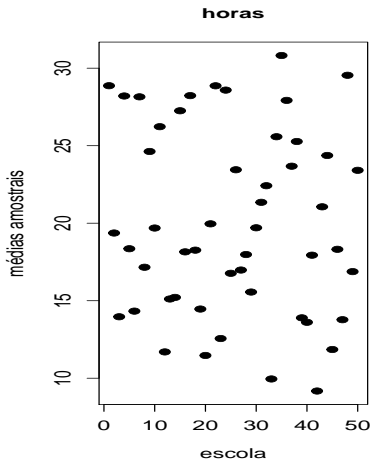
## Corr. de Person entre desempenho e horas (por escola)



# Médias e variâncias do desempenho por escola



# Médias e variâncias das horas por escola





# Comentários

- Note que temos duas covariáveis: escola (qualitativa nominal) e horas (quantitativa contínua).
- Nos modelos de regressão modela-se a distribuição condicional da resposta em função das combinações de todos os valores das covariáveis.
- Devido as duas características acima, torna-se inviável (ou não útil) realizar as análises (descritivas) sobre normalidade, CCI, homocedasticidade etc. Tais análises devem ser feitas de forma inferencial, através de análise de resíduos, análise de influência etc, a partir do ajuste de modelos de regressão apropriados.

# Comentários

- Pelas análises descritivas realizadas, podemos notar:
  - Uma possível diferença na distribuição do desempenho e das horas, entre as escolas.
  - Em particular, as médias e as variâncias do desempenho e das horas parecem ser diferentes, entre as escolas.
  - Parece existir uma relação linear (positiva) entre desempenho e horas, diferente entre as escolas: tanto em relação ao aumento no desempenho para o aumento em horas, quanto em relação ao desempenho esperado para zero horas de estudo.
  - Quase todas as correlações (entre desempenho e horas) são maiores do que 0,90.

# Comentários

- Tanto no contexto de um nível (modelo 1 do EA1) quanto no de dois níveis (modelo 2 do EA1), temos várias opções de modelagem: interceptos iguais ou diferentes e/ou coeficientes angulares iguais ou diferentes (pensando em retas), resultando num total de quatro (grupos de modelo).
- Por simplicidade, e para fazer um paralelo com o EA1, utilizaremos o modelo em que os interceptos são iguais e os coeficientes angulares podem ser diferentes.

# Modelos lineares (retas) possíveis

		Parâmetro	
		Intercepto	Coefficiente angular
Estrutura	Iguais	Iguais	Iguais
	Diferentes	Diferentes	Iguais
	Iguais	Iguais	Diferentes
	Diferentes	Diferentes	Diferentes

# Modelo 1 (EA2)

- Modelo de regressão normal linear homecedástico (MRNLH) (ME 613), parametrização casela de referência (ME 623).

$$Y_{ji} = \alpha + (\beta + \beta_j)x_{ji} + \xi_{ji}, i = 1, 2, \dots, 50 \text{ (escola)}; j = 1, 2, \dots, n_j \text{ (aluno)}$$

- Erros  $\xi_{ji} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ ,  $\mu, \alpha_j$  são não aleatórios.
- $\mu_i = \mathcal{E}_{\xi_{ji}}(Y_{ji}) \equiv \mathcal{E}(Y_{ji}) = \alpha + (\beta + \beta_j)x_{ji}$ ,  $\mathcal{V}_{\xi_{ji}}(Y_{ji}) \equiv \mathcal{V}(Y_{ji}) = \sigma^2$ .

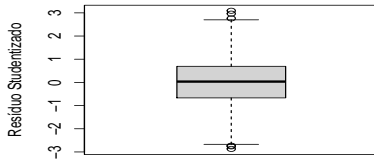
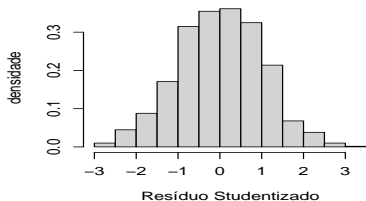
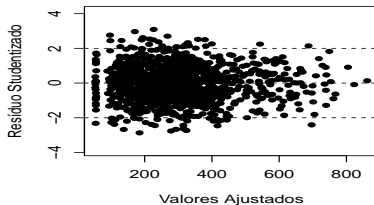
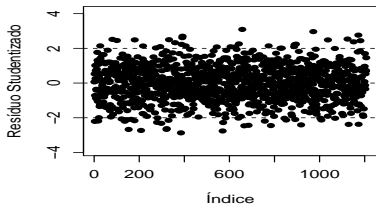
## Modelo 1 (EA2) Cont.

- $\mu_j = \alpha + (\beta + \beta_j)x_{ij}$ , em que  $\beta_1 = 0$ . Assim,  $\beta$  é o incremento (positivo ou negativo) esperado no desempenho, para o aumento em 1 hora nas horas semanais de estudo, de alunos da escola 1.
- $\beta_j$  é o quanto o incremento (mencionado no item anterior) da  $j$ -ésima escola difere do incremento da escola 1. Se for positivo (negativo), este incremento é maior (menor) para a escola  $i$  em relação à escola 1. Se for igual a 0, então os incrementos entre a escola  $i$  e a escola 1 são iguais.

## Modelo 1 (EA2) Cont.

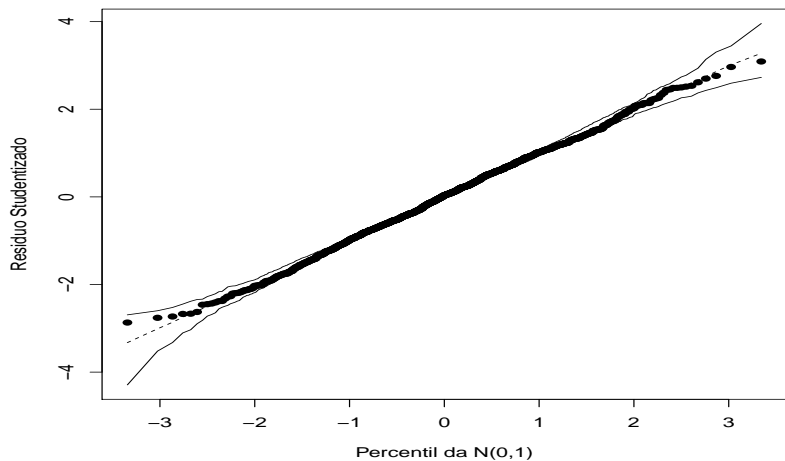
- O modelo 1 foi ajustado através do método de mínimos quadrados ([metodologia](#), [Wikibook sobre o ajuste do modelos de regressão lineares](#)), usando a função `lm` ([manual](#)) (e outras funções, veja o site do curso!) do pacote R e uma análise de diagnóstico a partir do resíduo studentizado ([material sobre resíduo studentizado](#)) fora realizada.

# Gráficos de diagnóstico: resíduo studentizado





## Gráficos de envelopes: resíduo studentizado



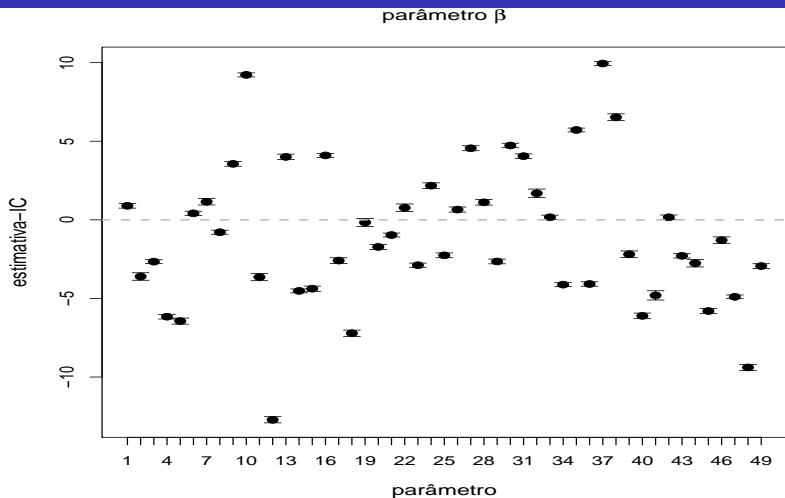
# Sobre o ajuste do modelo 1

- As suposições de normalidade e independência parecem estar sendo satisfeitas.
- A distribuição dos resíduos parece ser simétrica com um percentual muito pequeno de (outliers).
- Possível indicações de heterocedasticidade (diferença nos padrões de variabilidade dos resíduos, em função do valor da covariável horas).
- Ausência de padrões (tendência) que acusem correlação entre as observações.

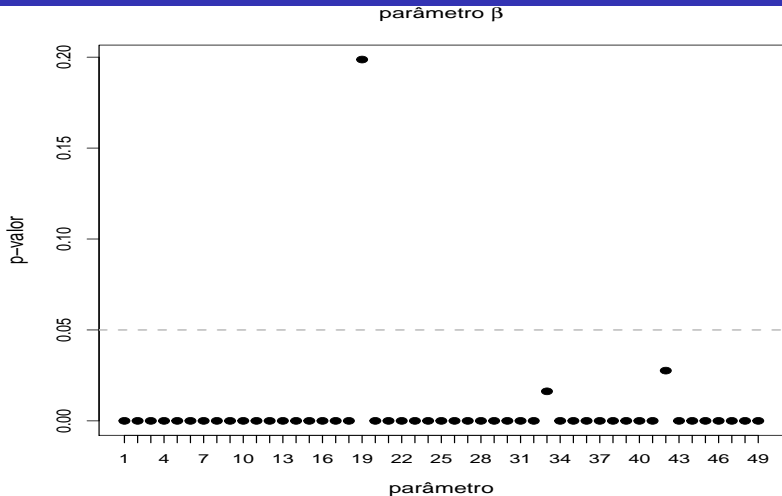
# Sobre o ajuste do modelo 1

- Os resíduos (gráfico de envelopes) estão dispersos, aleatoriamente, dentro das bandas de confiança, embora alguma tendência pode ser observada nas caudas.
- Em geral, o modelo parece estar apresentar um ajuste razoável (a menos da heterocedasticidade). Pergunta: será que, realmente, não há nenhum problema?

# Intervalos de confiança para os $\beta_j$ 's



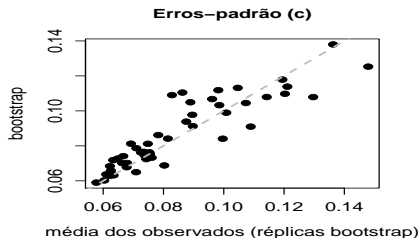
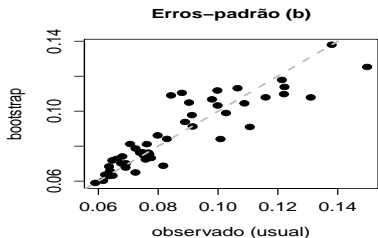
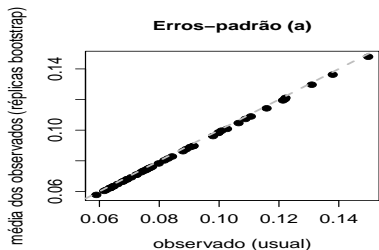
# p-valor para os testes de nulidade dos $\beta_j$ 's



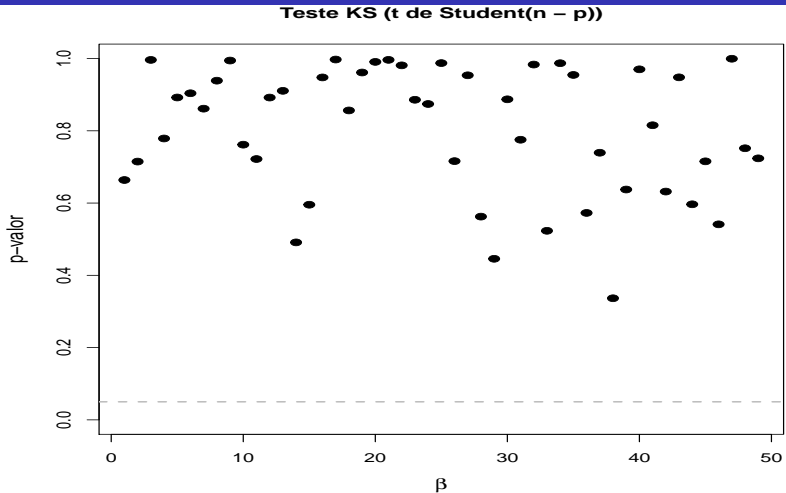
# Comentários sobre os resultados inferenciais

- Podemos notar que quase todos os parâmetros são diferentes de zero (p-valor).
- Podemos notar que alguns (grupos) de parâmetros parecem não ser diferentes entre si.
- Vamos reajustar o modelo utilizando Bootstrap não paramétrico ([curso sobre métodos computacionais em Estatística](#), [várias referências](#)), para estudar o comportamento dos estimadores de mínimos quadrados.
- Utilizou-se  $B = 1.000$  amostras Bootstrap, com amostras (com reposição) de mesmo tamanho (quantidade de alunos e escolas) em relação aos dados originais. Os resultados seguem:

# Erros-padrão: observados e via bootstrap



# Teste de aderência dos $\beta'_j$ s à distribuição $t_{(n-p)}$





## Modelo 2 - EA2

- Modelos multinível (hierárquico) de dois níveis

$$Y_{ji} = \beta_0 + \beta_{1j}x_{ji} + \xi_{ji}, \text{ (nível 1 -aluno)}$$

$$j = 1, 2, \dots, 50 \text{ (escola); } i = 1, 2, \dots, n_j \text{ (aluno)}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \text{ (nível 2 - escola),}$$

$$\xi_{ji} \stackrel{iid}{\sim} N(0, \sigma^2), u_{1j} \stackrel{iid}{\sim} N(0, \psi), \xi_{ji} \perp u_{1j}$$

- $\gamma_{10}$  é um efeito fixo,  $\beta_{1j}(u_{1j})$ ,  $j = 1, 2, \dots, 50$  são efeitos aleatórios, e  $\sigma^2$  e  $\psi$  são os componentes de variância (parâmetros de dispersão).

## Modelo 2

- $\mathcal{E}(Y_{ji}|u_{1j}) = \beta_0 + (\gamma_{10} + u_{1j})x_{ij}$  (média condicional do desempenho do  $i$ -ésimo aluno da  $j$ -ésima escola). Ou seja, valor esperado do desempenho de um aluno da  $j$ -ésima escola, dado que ele despense por semana  $x_{ji}$  horas. Se  $x_{ji} = 0$  então  $\mathcal{E}(Y_{ji}|u_{1j}) = \beta_0$  (ou seja, para todo e qualquer aluno com zero horas de estudo semanal, seu desempenho esperado é o mesmo, independentemente da escola).

## Modelo 2

- $\mathcal{E}(Y_{ji}) = \beta_0 + \gamma_{10}x_{ji}$  (média marginal do desempenho do i-ésimo aluno da j-ésima escola). Ou seja, valor esperado do desempenho de um aluno, dado que ele despende por semana  $x_{ji}$  horas, considerando todas as escolas (note que as médias só são diferentes se o valor das horas o for). Ademais, se  $x_{ji} = 0$  então  $\mathcal{E}(Y_{ji}) = \mathcal{E}(Y_{ji}|u_{1j}) = \beta_0$ . O que isto significa?
- $\mathcal{V}(Y_{ji}|u_{1j}) = \sigma^2$  (variância condicional do i-ésimo aluno da j-ésima escola). Ou seja, variância do desempenho entre os alunos da j-ésima escola.

## Modelo 2

- $\mathcal{V}(Y_{ji}) = \psi x_{ji}^2 + \sigma^2$  (variância marginal do desempenho do  $i$ -ésimo aluno da  $j$ -ésima escola). Ou seja, variância do desempenho entre todos os alunos (note que se tem uma estrutura heterocedástica).
- Também, temos que,  $Cov(Y_{ji}, Y_{ji'}) = x_{ji}x_{ji'}\psi$  e  $Corre(Y_{ji}, Y_{ji'}) = \frac{x_{ji}x_{ji'}\psi}{\sqrt{x_{ji}^2\psi + \sigma^2}\sqrt{x_{ji'}^2\psi + \sigma^2}}$ ,  $\forall i \neq i'$  (Exercício). Note que se pelo menos um dos alunos apresentar zero horas, as observações são não correlacionadas (independentes) (o que isso significa?)
- $u_{1j} = \frac{1}{x_{ji}} (\mathcal{E}(Y_{ji}|u_{1j}) - \mathcal{E}(Y_{ji}))$ ,  $x_{ji} \neq 0$  é a diferença entre o valor esperado do desempenho do aluno  $i$  da escola  $j$  e o desempenho geral, para alunos com horas iguais a  $x_{ji}$  (ponderado por esta).

## Modelo 2

- Além disso,  $Y_{ji}|b_{0j} \stackrel{ind.}{\sim} N(\beta_0 + \gamma_{10}x_{ji} + u_{0j}, \sigma^2)$  e  
 $Y_{ji} \sim N(\beta_0 + \gamma_{10}x_{ji}, x_{ji}^2\psi + \sigma^2)$  (**não** são independentes) (Exercício)
- Neste caso, as hipóteses de interesse são:  $H_0 : \psi = 0$  vs  $H_1 : \psi > 0$ .  
Assim, não rejeitar  $H_0$  implica em afirmar que as escolas tem a mesma distribuição de desempenho. Pergunta: como testar essas hipóteses? ([veja](#), [veja](#)), entre outras referências ([ref. 1](#)), ([ref. 2](#)).

## Modelo 2

- De uma forma mais geral, temos que

$$\mathbf{Y}_j = (Y_{j1}, Y_{j2}, \dots, Y_{jn_j})' \sim N_{n_j}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), j = 1, 2, \dots, J, \text{ em que}$$

$$\boldsymbol{\mu} = \beta_0 \mathbf{1}_{n_j} + \gamma_{10} \mathbf{x}_j, \mathbf{1}_{n_j} = (1, 1, \dots, 1)'_{n_j \times 1}, \mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jn_j})', \text{ e}$$

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} \sigma^2 + x_{j1}^2 \psi & x_{j1} x_{j2} \psi & \dots & x_{j1} x_{jn_j} \psi \\ x_{j1} x_{j2} \psi & \sigma^2 + x_{j2}^2 \psi & \dots & x_{j2} x_{jn_j} \psi \\ \vdots & \vdots & \ddots & \vdots \\ x_{j1} x_{jn_j} \psi & x_{j2} x_{jn_j} \psi & \dots & \sigma^2 + x_{jn_j}^2 \psi \end{bmatrix}$$

## Modelo 2

- Podemos notar, assim, que:
  - Os desempenhos (médios) entre as escolas diferem em termos do efeito aleatório  $u_{0j}$  ( $\psi > 0$ ). Ou seja, para um mesmo número horas, os desempenhos médios tendem a ser maiores quanto maior o valor do efeito aleatório.
  - Além disso, se o valor de  $u_{0j}$  for positivo (negativo), significa que, para alunos da escola  $j$ , as horas tem um impacto maior (menor) no desempenho médio.
  - A variância do desempenho de cada escola é menor do que a variância do desempenho de todas as escolas.

## Modelo 2 (cont.)

- Cont.:
  - As observações (dentro da mesma escola) podem ser correlacionadas (nesse caso, dependentes - devido à normalidade multivariada). Além disso, a magnitude da dependência depende do valor das horas.
  - A variabilidade (marginal) no desempenho é diretamente proporcional ao número de horas semanais de estudo (quanto maior o valor das horas, maior a variabilidade do desempenho).



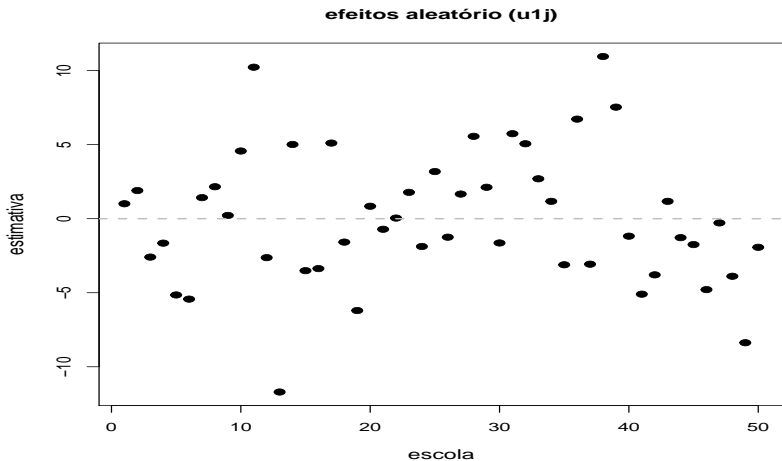
## Ajuste do modelo 2

- Veremos, mais adiante, os detalhes inferenciais (estimação, verificação da qualidade do ajuste, predição, comparação de modelos) necessários sobre a classe de modelos hierárquicos (da qual o modelo 2 faz parte).
- Ajustou-se-o utilizando o pacote “lme4” do programa R.

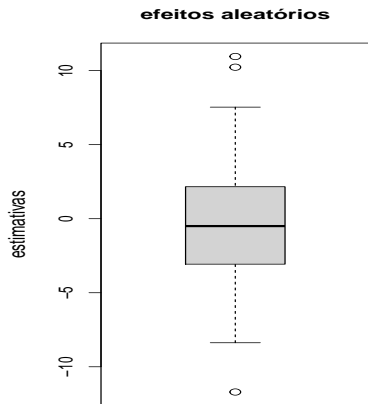
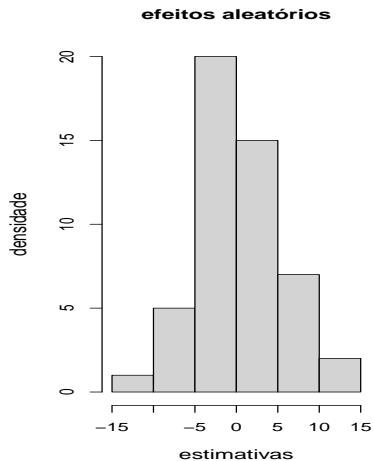
## Ajuste do modelo 2

- Uma breve comparação entre os modelos resultou em: Modelo 1 (AIC = 7913,35; BIC = 8178,34) e Modelo 2 (AIC = **7806,89**; BIC = **8071,88**)(maiores detalhes serão apresentados posteriormente). Isso indica que o modelo 2 é preferível ao modelo 1 (contudo, análises de resíduos são necessárias, para o modelo 2, a fim de ser subsídusos suficientes para tal afirmação).
- Comentaremos, em aulas vindouras, sobre seleção/comparação de modelos hierárquicos. Para uma revisão sobre tais metodologias para MRNLH veja : [link](#).
- Mais resultados podem ser encontrados nos próximos slides.

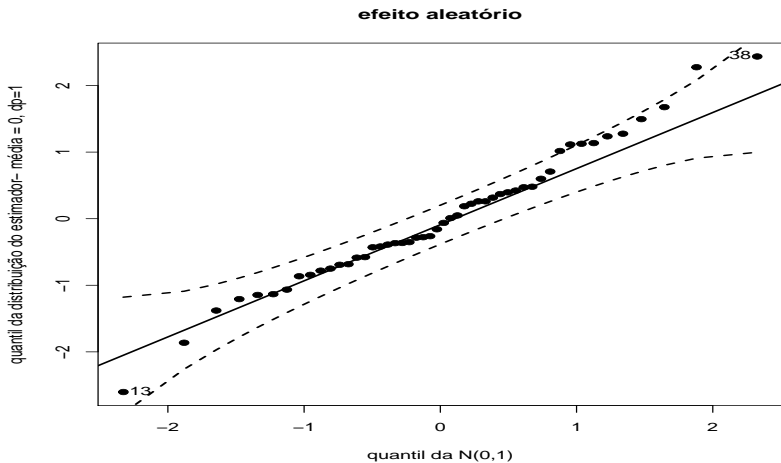
# Predição dos efeitos aleatórios



# Histograma e box-plot dos efeitos aleatórios



# Gráfico de quantil-quantil dos efeitos aleatórios



# Comentários sobre os efeitos aleatórios

- Pelos três conjuntos anteriores de gráficos, podemos perceber:
  - Que as escolas apresentam (distribuições de) desempenhos diferentes ( $\psi > 0$ ) (eventualmente podemos agrupar escolas com desempenhos equivalentes).
  - Que temos escolas que apresentam um impacto das horas no desempenho razoavelmente maior (+10) e razoavelmente menor (-10), em relação à escola 1.
  - Que os efeitos aleatórios ( $u_{0j}$ ) apresentam distribuição (aproximadamente) normal.

# Comentários

- Notamos que é muito importante considerar a estrutura hierárquica dos dados, sob pena de obtermos resultados menos confiáveis (errados).
- Nos dois exemplos anteriores, mesmo se  $\sigma^2 \equiv 0$ , o modelo hierárquico não seria um caso geral do (respectivo) modelo de um nível (EA1 se  $\alpha_j = 0, \forall j$ , EA2 se  $\beta_j = 0, \forall j$ ). Ou seja, nesses casos, o modelo multinível não é um caso geral do respectivo modelo usual.
- Devemos tomar cuidado especial com a questão da identificabilidade ([aula sobre identificabilidade e parametrizações](#)), nos modelos hierárquicos.

# Comentários

- A análise residual tem de ser feita de forma ainda mais elaborada do que nos modelos de regressão sem efeitos aleatórios ([referência 1](#), [referência 2](#), [artigo](#)). Veremos em aulas vindouras.
- A seleção/comparação de modelos (na presença de efeitos aleatórios) deve ser de forma ainda mais cautelosa ([artigo](#), [artigo](#), [artigo](#), [referências](#)). Veremos em aulas vindouras.
- Verificações de suposições acerca dos efeitos aleatórios necessitam de cautela adicional ([artigo](#), [livro](#)).



# Objetivos

- Desenvolvimento, por parte do aluno, da compreensão e desenvolvimento Estatístico de metodologias e das necessárias ferramentas matemáticas.
- Desenvolvimento de habilidades computacionais.
- Aprendizado de modelagem e resolução de problemas com dados hierárquicos.
- O quanto possível, estimular a interação com o responsável pelo problema, a fim de buscar as melhores soluções. Neste curso, analisar-se-á conjuntos de dados reais para os quais, basicamente, nunca, não se disporá de um consultante (o Prof. tentará fazer esse papel).