Introdução à
Modelagem Multinível em *Avaliação Educacional*

Maria Eugénia Ferrão
Iuri da Costa Leite
Kaizô Iwakami Beltrão

PREFEITURA DA CIDADE DO RIO DE JANEIRO
Secretaria Municipal de Urbanismo
Instituto Municipal de Urbanismo Pereira Passos - IPP

ENCE
ESCOLA NACIONAL
DE CIÊNCIAS ESTATÍSTICAS

IBGE
Instituto Brasileiro de Geografia e Estatística

Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística – IBGE
Escola Nacional de Ciências Estatísticas

Curso
Avaliação Educacional:
Modelagem Multinível de Dados

Introdução à Modelagem Multinível em Avaliação Educacional

Maria Eugénia Ferrão
Iuri da Costa Leite
Kaizô Iwakami Beltrão

Rio de Janeiro
2001

Instituto Brasileiro de Geografia e Estatística – IBGE
Av. Franklin Roosevelt, 166 – Centro – 20021–120 – Rio de Janeiro, RJ - Brasil

Sumário

Apresentação

A Escola Nacional de Ciências Estatísticas – ENCE, do Instituto Brasileiro de Geografia e Estatística – IBGE, organizou a oficina **"Introdução à Modelagem Multinível em Avaliação Educacional"**, cuja primeira realização conta com a parceria do Instituto Municipal Pereira Passos – IPP, de **17 a 20 de dezembro**, no **IBGE (sala da ENCE – complexo B)**, localizado na Rua General Canabarro, 706 – Maracanã, Rio de Janeiro.

A oficina "Introdução à Modelagem Multinível em Avaliação Educacional" insere-se no projeto "Avaliação Educacional – uma abordagem quantitativa" e visa fomentar a implantação e disseminação de uma cultura de avaliação no Brasil. Este documento é um guia para esta oficina. É destinado a educadores, pesquisadores e técnicos de nível intermediário e superior que atuem em diferentes níveis do sistema de ensino e que tenham familiaridade com ambiente Windows.

Com o curso pretende-se, fundamentalmente, contribuir para fomentar e/ou consolidar uma avaliação educacional, alargando a discussão a todos os interessados, particularmente aos colegas que têm algum tipo de participação no Sistema Educativo.

Kaizô Iwakami Beltrão
Superintendente da ENCE

1 Introdução

No decorrer do curso serão abordados conceitos e definições tais como os de *valor agregado*, *eficácia da escola* e *efeito-escola*.

A componente laboratorial do curso será desenvolvida usando os dados do Sistema Nacional de Avaliação do Educação Básica (Brasil) e, sempre que possível, os dados próprios dos participantes.

O texto é composto por 4 seções. A primeira é composta por esta introdução. A segunda apresenta os tópicos selecionados de estatística e probabilidade que são necessários ao desenvolvimento da modelagem multinível, nomeadamente tipos de dados, distribuições de probabilidade e seus parâmetros, medida de qualidade do ajuste, teste de hipóteses e intervalo de confiança, além da regressão linear clássica. A seção 3 constitui uma introdução aos modelos de regressão multinível que é desenvolvida a partir da generalização da regressão linear clássica. São abordados os modelos de componentes de variância e de componentes aleatórias. É apresentado o coeficiente de correlação intra-escola como estimativa do *efeito-escola*. Segue-se, por fim, o tutorial do pacote estatístico usado para os modelos de regressão multinível.

2 Tópicos Elementares de Probabilidade e Estatística

A estatística pode ser dividida em dois grandes ramos: descritiva (ou exploratória) e confirmatória (ou inferência estatística). Na estatística descritiva o conjunto de dados é explorado de forma a ser classificado e a fornecer resumos de suas características e inter-relações. Uma vez escolhida uma descrição paramétrica (possivelmente uma distribuição ou um modelo descritivo) cabe à estatística confirmatória estimar e testar os parâmetros calculados nos modelos e distribuições.

Os dados estatísticos, por sua vez, podem ser classificados de acordo com a sua natureza:

- Dados quantitativos
  - Discretos
  - Contínuos (com e sem zero absoluto)

- Dados qualitativos
  - Nominais
  - Ordinais

Por exemplo, os dados de nível socioeconômico considerados neste curso são (apesar de estarem associados a números entre 1 e 7) dados qualitativos ordinais. São ordinais porque podemos ordená-los de uma forma única. O nível socioeconômico mais baixo é o 1 e o mais alto é o 7. Não podemos, porém, dizer que o nível 6 é duas vezes melhor que o nível 3, nem que a diferença entre os níveis 1 e 2 é igual a diferença entre os níveis 4 e 5. Os dados de dependência administrativa (estadual, municipal e privado), por sua vez, são dados qualitativos nominais – não existe uma forma única de ordená-los.

Alguns dados quantitativos são intrinsecamente discretos, usualmente resultado de contagens: número de alunos numa sala de aula, número de dias letivos, etc. Alguns outros são intrinsecamente contínuos tais como peso e altura de indivíduos, distâncias entre pontos e tempo decorrido entre eventos, mas são discretizados por motivos operacionais, já que são mensurados com alguma escala específica (gramas no caso de peso, centímetros no caso de altura e minutos para tempo decorrido) ainda que estas escalas tenham sempre subdivisões que poderiam também ser utilizadas (miligramas, milímetros, segundos, etc.).

Alguns dados são medidos em escalas onde o zero é um ponto arbitrário. Por exemplo o ano calendário da civilização ocidental tem um zero arbitrado no nascimento de Cristo. Este zero é diferente do calendário

judeu (onde zero é a criação do mundo), do japonês (onde o zero é o nascimento do imperador), ou do chinês. No entanto, todos os anos (solares) tem a mesma duração e o tempo de corrido entre 1900 e 1901 é igual ao tempo decorrido entre 1800 e 1801.

2.1 Distribuição de Probabilidades

Para descrever populações, os estatísticos criaram famílias de distribuições, usualmente caracterizadas por parâmetros. A distribuição mais simples é a Bernoulli. Numa distribuição de Bernoulli existem somente duas possíveis respostas do experimento (digamos 0 e 1, ou, sucesso e fracasso) e para cada resposta está associada uma probabilidade. Por exemplo ao lançarmos um dado existe uma certa probabilidade de obtermos a face com 5 pontos. Esta probabilidade é de 1/6 num dado não viciado. A probabilidade de a face conter qualquer outro número é de 5/6. No caso desta distribuição, o parâmetro definidor da mesma é a probabilidade **p** de acontecer o evento. Se repetirmos o lançamento do dado um certo número **n** de vezes, a distribuição do número de vezes em que o dado apresenta a face 5 é uma distribuição binomial com parâmetros n e 1/6, B(n, 1/6). Considerando-se **n** conhecido, o parâmetro desconhecido desta distribuição é novamente a probabilidade **p**=1/6. A Figura 1.1 apresenta estas probabilidades para n=10.

Vemos nesta figura que a probabilidade de, ao se jogar o dado 10 vezes, não tirarmos nenhuma vez o número 5 (k=0) é 16,15%; de tirarmos exatamente uma vez (k=1) é de 32,30%. Se continuarmos jogando o dado, digamos 100 vezes, o gráfico da probabilidade de se tirar k vezes o número 5 está descrito na figura 1.2. Note que o gráfico ficou mais simétrico ao se aumentar o número de jogadas.

Figura 1.1

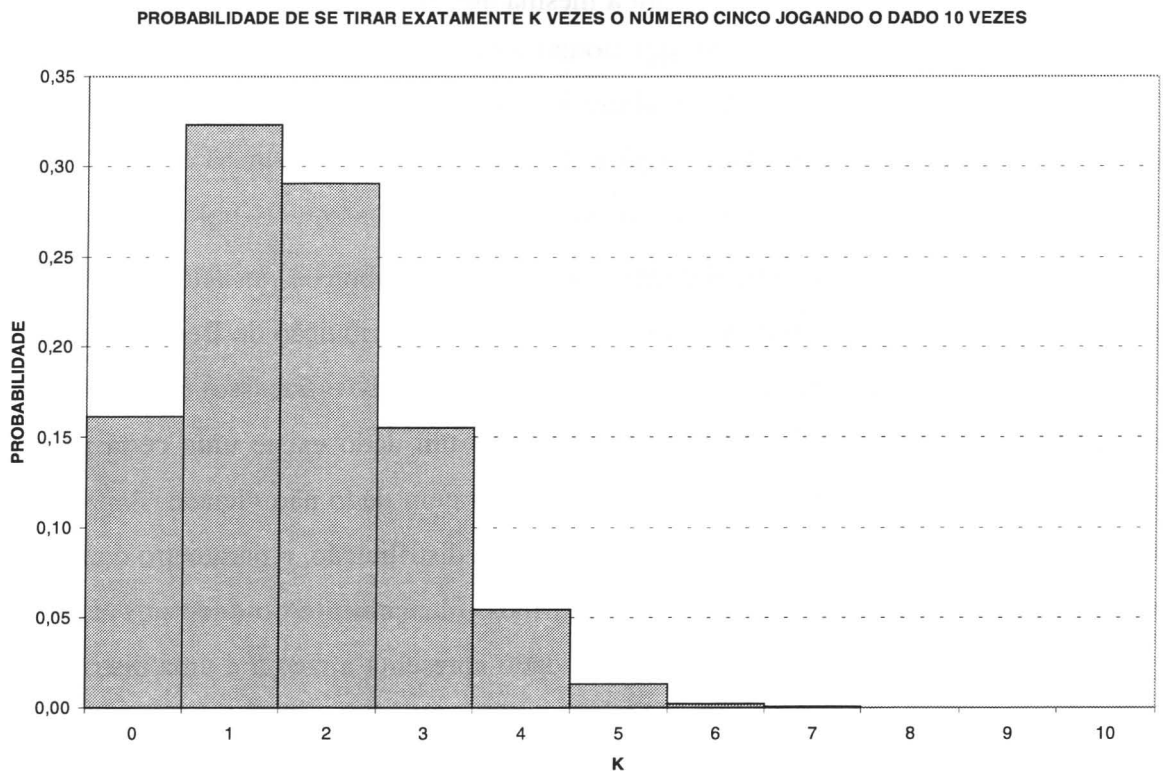
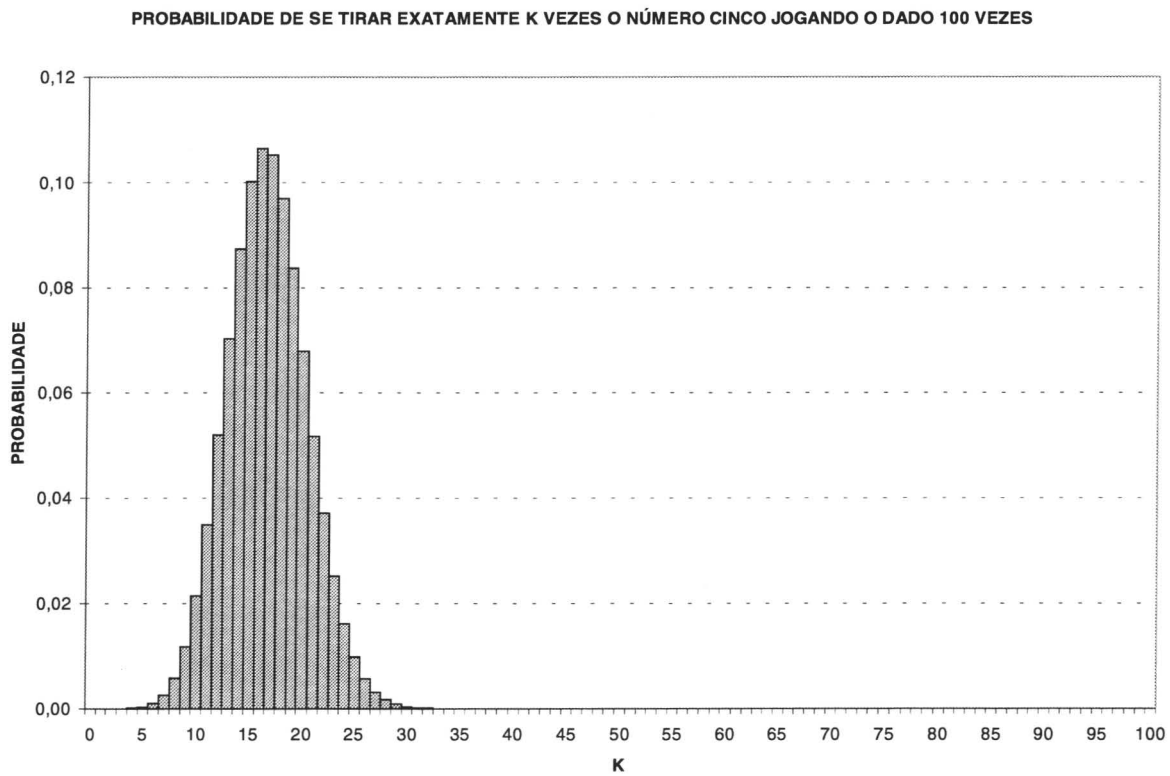


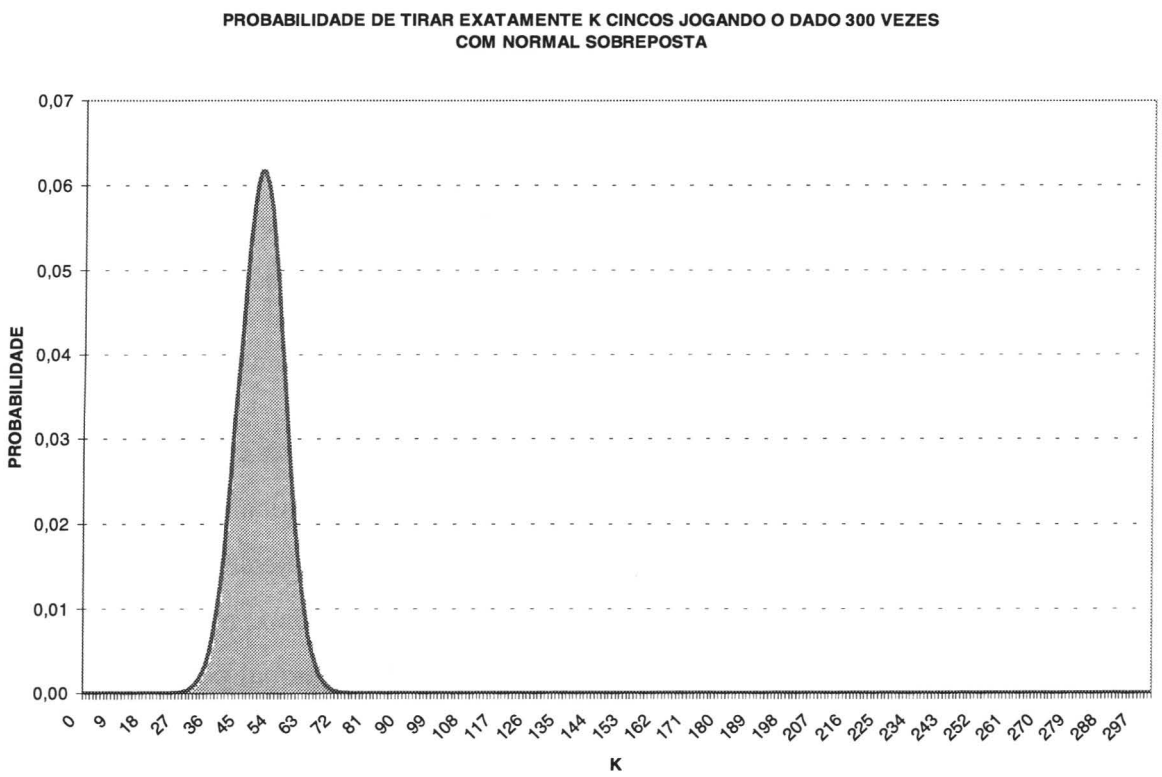
Figura 1.2



Se aumentarmos ainda mais o número de jogadas, digamos para 300, a distribuição fica ainda mais simétrica e aproxima-se de uma outra curva, a curva Normal. O Teorema dos Grandes Números[1], garante que, quando o número de repetições de um evento é grande a sua distribuição assemelha-se à de uma curva Normal.

A Figura 1.3 apresenta a distribuição das 300 jogadas do dado com a curva Normal sobreposta em vermelho. Note que as duas distribuições são muito semelhantes.

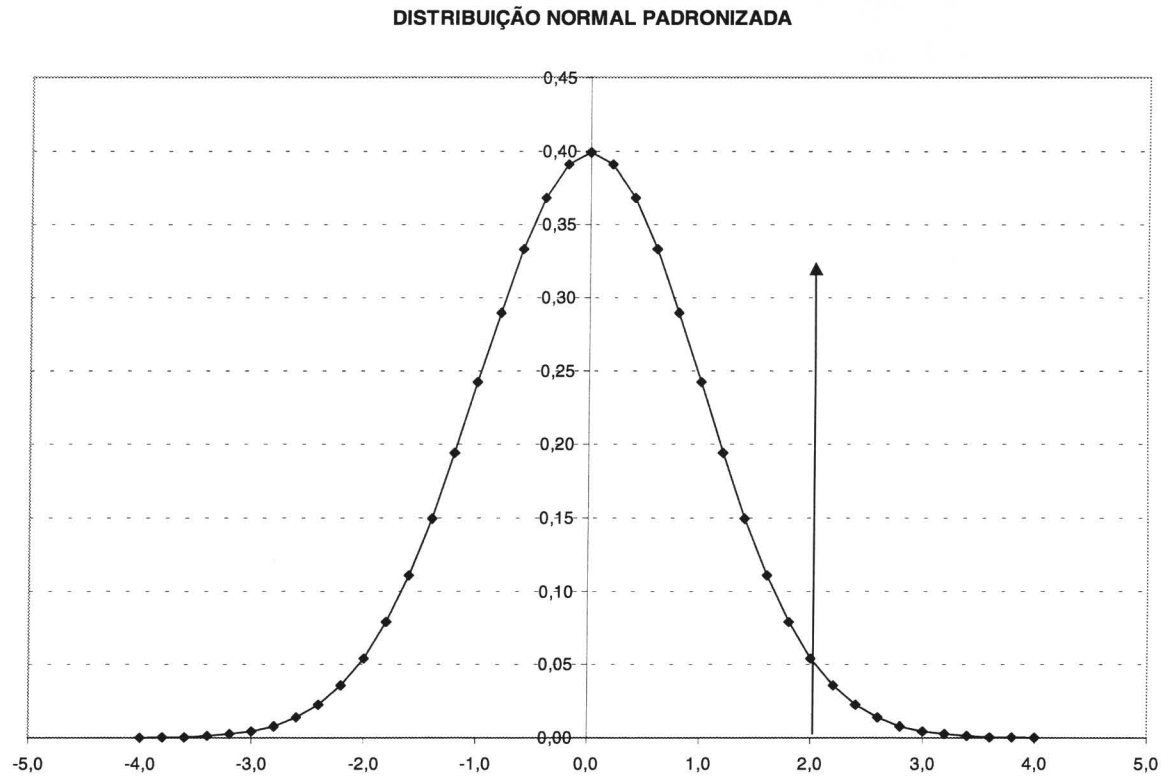
Figura 1.3



Neste caso, a Normal sobreposta tem a mesma média e desvio padrão da Binomial original. É comum trabalharmos com uma Normal padronizada, com média zero e desvio padrão 1, como representado na figura 1.4.

[1]• Para maiores detalhes sugere-se Breiman, L. (1969). Probability and Stochastic Processes: with a view toward applications. Boston, Hougthon Mifflin Company.

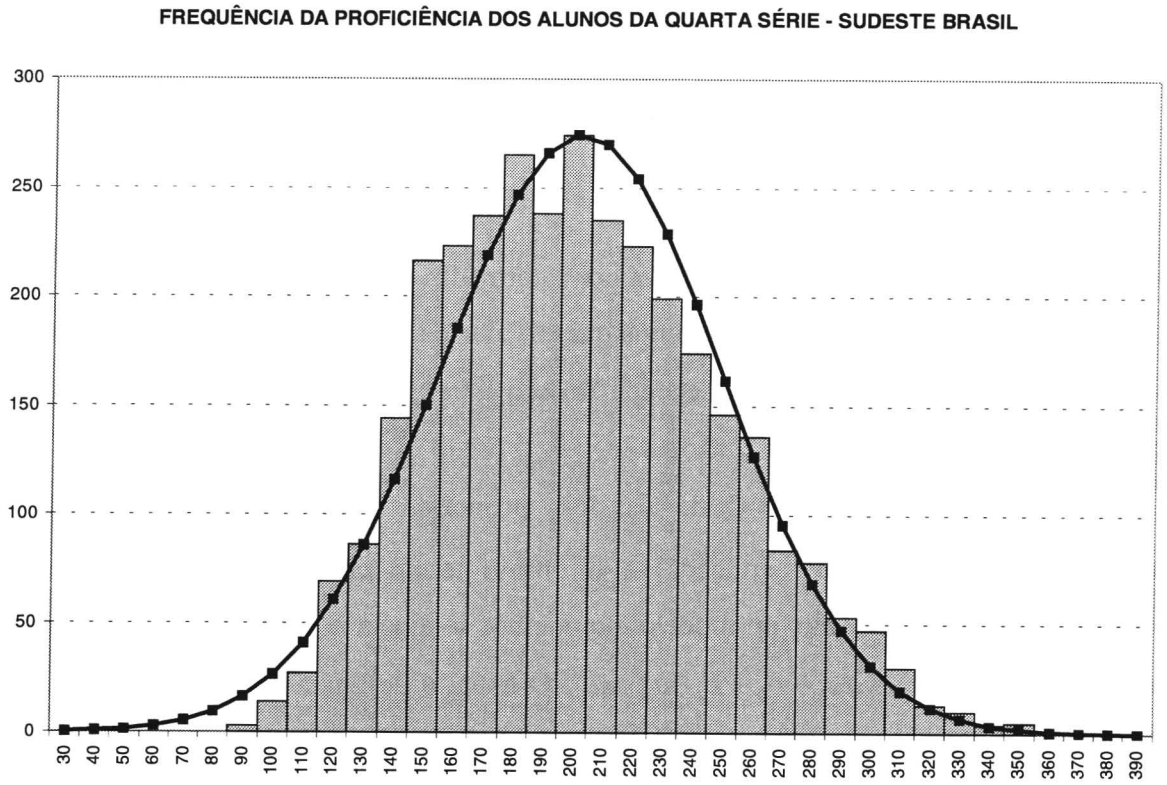
Figura 1.4



A área embaixo da curva é sempre igual a 1. Para uma dada abcissa (por exemplo x=2) é possível calcular a área à direita daquele ponto (no caso, 0,0275). Isto quer dizer que a probabilidade de termos um valor superior a 2 numa Normal padrão é 0,0275. É mais comum termos o problema inverso, ou seja, dada uma probabilidade querermos saber qual é a abcissa correspondente. O valor de x, para o qual a área à direita é 0,025 é 1,96, ou seja a probabilidade de valores de uma distribuição normal padronizada serem maiores do que 1,96 é 0,025. Considerando-se que a curva Normal é simétrica em torno de 0, decorre daí que a probabilidade de obtermos um valor cujo módulo é maior do que 1,96, é 0,05 (2x0,025).

Os dados da proficiência são distribuídos aproximadamente como uma Normal, como pode ser visto na figura 1.5 que apresenta a distribuição dos dados de proficiência e de uma curva Normal ajustada sobreposta. Nesta figura pode verificar-se que há aproximadamente 200 alunos com classificação entre 230 e 240 pontos. A diferença por falta (curva normal maior) é mais notável na cauda inferior e na metade central direita, sendo o excesso compensado na cauda superior e na metade central esquerda. Ao assumirmos que os dados de proficiência seguem uma distribuição Normal, os parâmetros a serem estimados são a média e o desvio padrão. Neste caso a média e o desvio padrão estimados para o conjunto de alunos são, respectivamente, 196,63 e 47,00. Estas estimativas de parâmetros tem associados imprecisões e medidas de erro. O desvio padrão da estimativa da média é 0,826631.

Figura 1.5



Alguém poderia levantar a hipótese de que a distribuição das proficiências deveria ter um dado valor para a quarta série, digamos 200. Considerando que os dados utilizados provêem de uma amostra é possível testar se a população original tem realmente a média igual a 200. A chamada estatística de teste é o escore z, definido como:

$$\frac{|200 - média|}{desvio} = \frac{|200 - 196{,}63|}{0{,}826631} = 4{,}072$$

Numa distribuição Normal a probabilidade de termos valores iguais ou maiores do que 4,072 é 0,002%, ou seja, muito pouco provável. Em estatística diríamos que devemos rejeitar a hipótese de que a média da população é igual a 200. *Stricto Sensu*, a distribuição da estatística calculada acima é de uma variável t de Student (razão de uma variável normal e da raiz se soma de quadrados de normais), porém, dado o número de variáveis envolvido na soma de quadrados, a aproximação Normal é perfeita.

Para cada teste existe um intervalo de confiança equivalente. O usual é definir algum nível de confiança (por exemplo 95%) e calcular qual o valor da variável que teria probabilidade de 95% de cair dentro do

intervalo (ou de 5% de cair fora). Neste nosso exemplo o intervalo de confiança de 95% é igual a média +- 1,96*devio padrão, ou seja, (195,01;198,25). Podemos dizer que em 95% dos casos a média da população envolvida estaria contida neste intervalo, seria maior do que 195,01 e menor do que 198,25.

2.2 Modelos e Qualidade de Ajuste

Poderíamos então tentar explicar a proficiência por intermédio de variáveis explicativas (por exemplo o nível socio-econômico). Neste modelo diríamos que o valor da proficiência de um dado aluno é igual a uma constante (β_0) mais um fator que depende do seu nível socio-econômico. Este modelo pode ser escrito, para o i-ésimo aluno, como:

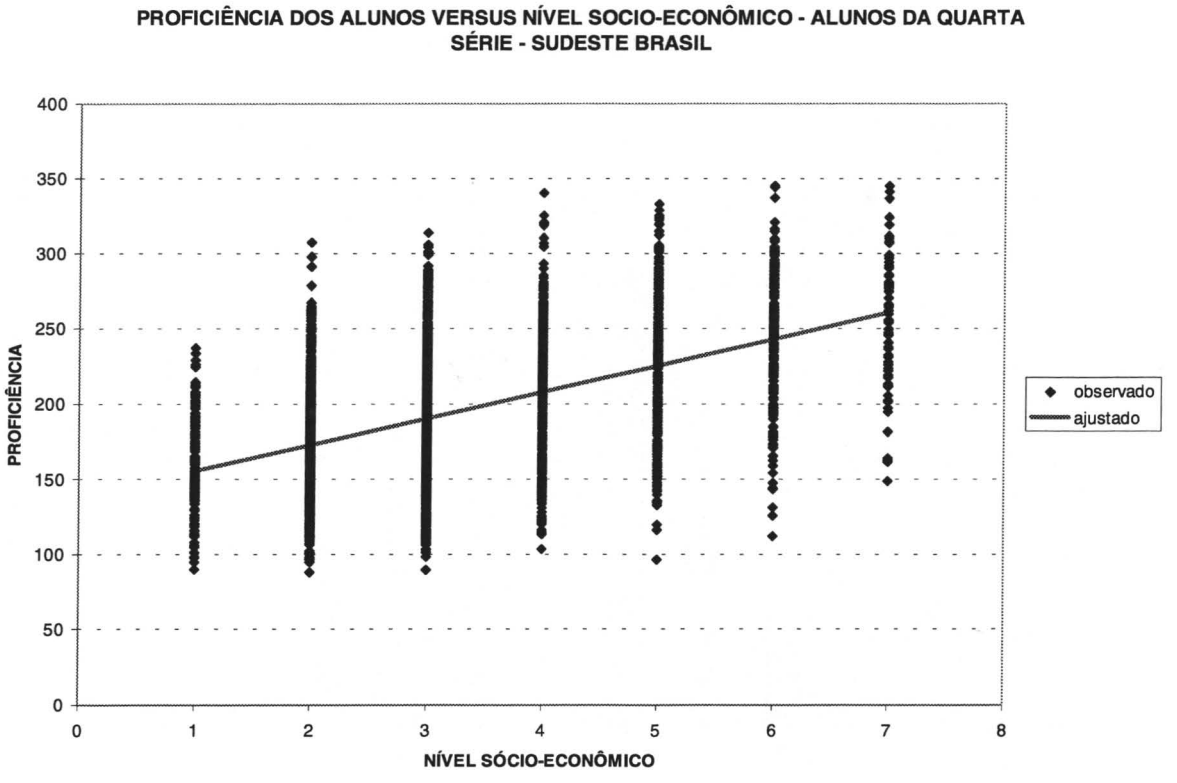
$proficiência_i = \beta_0 + \beta_1 * a\_nse_i + e_i$. Onde os dois primeiros termos do lado direito da equação correspondem ao modelo (parte explicada) e o último termo corresponde ao erro ou resíduo. A determinação do "melhor" modelo, ou modelo que melhor se ajusta aos dados tem a ver com os valores dos resíduos $res_i = proficiência_i - \beta_0 + \beta_1 * a\_nse_i$. O usual é definir a qualidade do ajuste a partir de alguma estatística, por exemplo a soma de quadrados dos resíduos e escolher o modelo correspondente a menor soma de quadrados (este é o chamado estimador de mínimos quadrados usado pela maior parte dos pacotes estatísticos). Neste caso escolhemos os parâmetros do melhor modelo $\hat{\beta}_0$ e $\hat{\beta}_1$ como os que minimizam $\sum_i (res_i)^2 = \sum_i (proficiência_i - \beta_0 + \beta_1 * a\_nse_i)^2$.

Da mesma forma que associado ao estimador da média obtinha-se um desvio padrão, a estes estimadores correspondem, também, desvios-padrão. Para os nossos dados os parâmetros estimados e os respectivos desvios-padrão são:

Parâmetro	Estimativa	Desvio-padrão
$\hat{\beta}_0$	137,7482	1,88267
$\hat{\beta}_1$	17,5286	0,51895

A Figura 1.6 apresenta um gráfico com os dados originais de proficiência plotados contra o nível socioeconômico, bem como a reta ajustada do modelo.

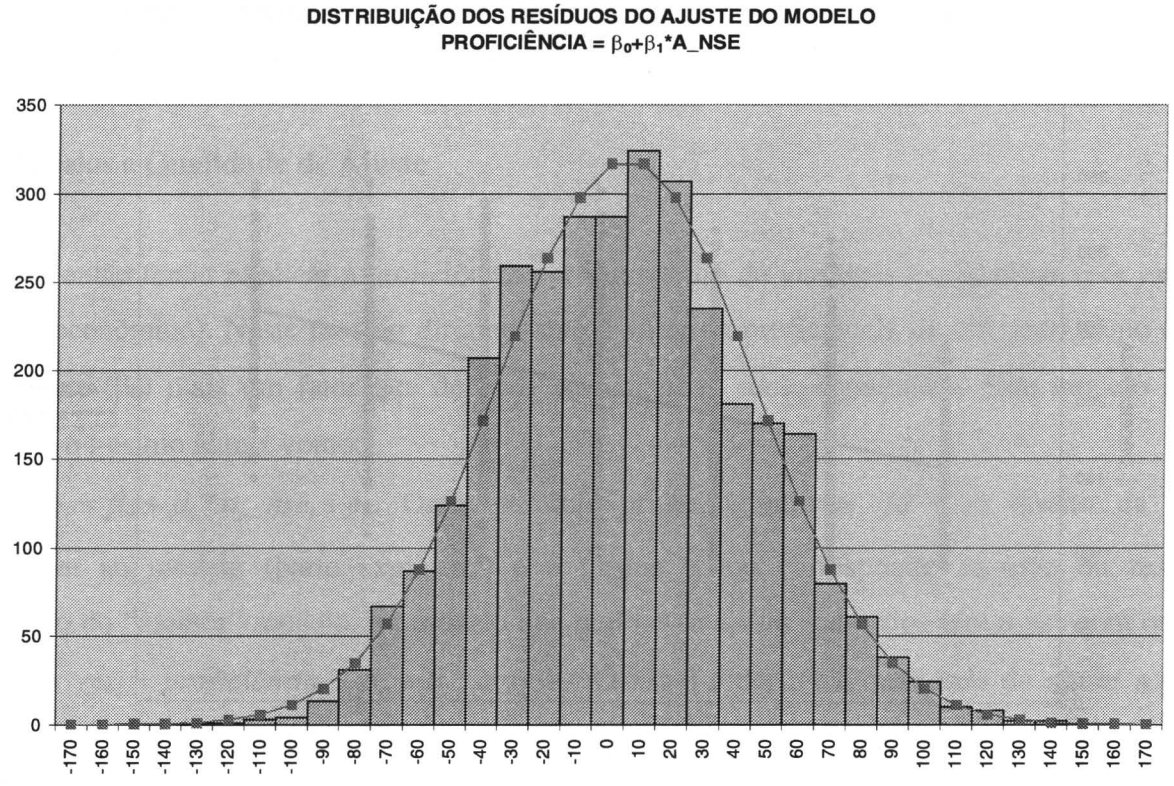
Figura 1.6



Poderíamos, também, testar algo como "existe efeito de nível socioeconômico?". Neste caso a estatística de teste envolveria o parâmetro $\hat{\beta}_1$ e a comparação seria com o valor zero (nível socioeconômico não tem efeito) e seria calculado como $\frac{|0 - média|}{desvio} = \frac{|17{,}5286|}{0{,}51895} = 33{,}7773$ e a probabilidade de um tal valor acontecer aleatoriamente é para todos os efeitos nulo e não podemos rejeitar a hipótese de que nível socioeconômico tem efeito.

O passo seguinte é verificar se a hipótese da distribuição Normal dos resíduos é satisfeita. A Figura 1.7 apresenta a distribuição dos resíduos e uma curva Normal ajustada superposta.

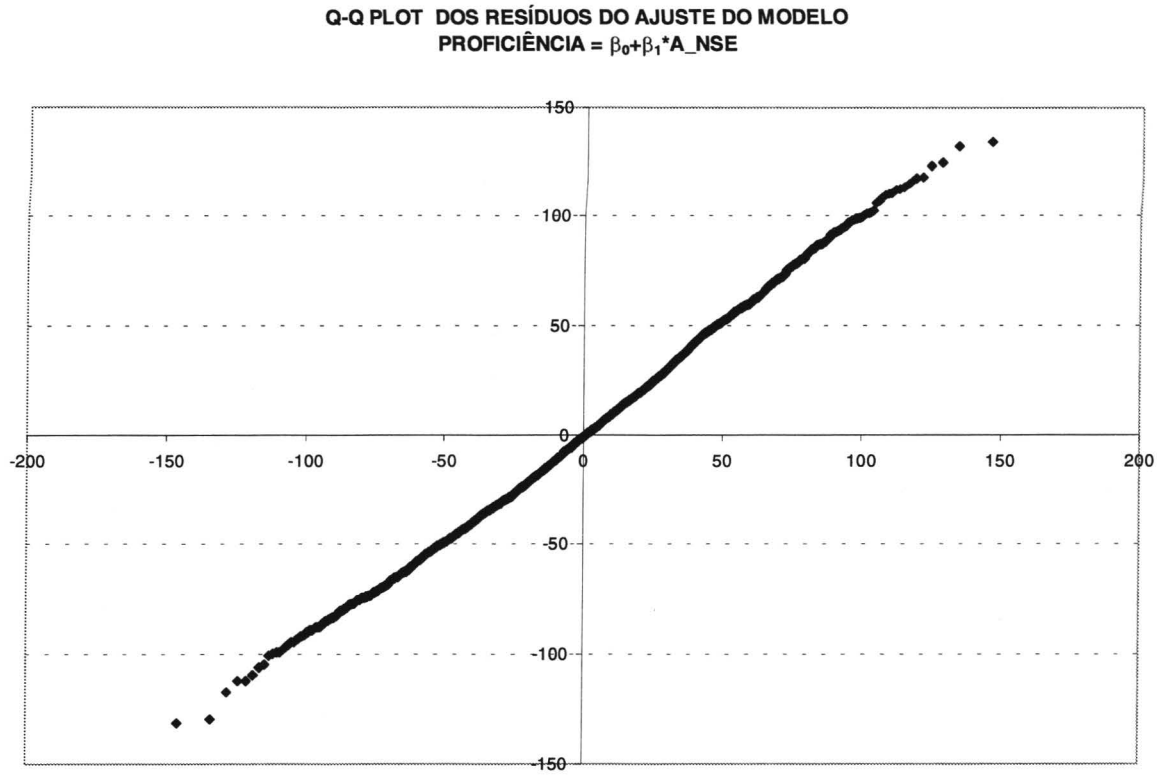
Figura 1.7



Podemos notar que nesta figura a curva ajustada é muito mais perto da distribuição do que no caso da proficiência.

Uma outra forma de fazermos esta verificação é com um gráfico q-q. Para este gráfico ordenamos todos os resíduos e plotamos contra o que seria o valor esperado caso estes fossem Normais. A Figura 1.8 apresenta o dito gráfico. Caso os resíduos sejam realmente Normais, estes devem estar dispostos em uma linha reta. No nosso exemplo vemos que são poucos os pontos que se afastam desta tendência, e mesmo assim são pontos extremos (ou muito grandes, ou muito pequenos). Neste caso colocamos nas abcissas os valores esperados da Normal com a média e desvio padrão ajustados, mas o resultado seria o mesmo se utilizássemos uma Normal padrão (com média zero e desvio-padrão igual a unidade).

Figura 1.8


Q-Q PLOT DOS RESÍDUOS DO AJUSTE DO MODELO
PROFICIÊNCIA = $\beta_0+\beta_1$*A_NSE
150
100
50
0
-50
-100
-150
-200
-150
-100
-50
0
50
100
150
200

3 Modelo Multinível para a Proficiência[2]

Avaliação educacional é usualmente a área de aplicação clássica dos modelos multinível ou hierárquicos. A estrutura hierárquica é facilmente identificada, pois os alunos são agrupados em turmas e as turmas agrupadas em escolas e, por conseguinte, os dados carregam a mesma estrutura da população onde são coletados.

Até ao final da década de 80, a grande maioria da modelagem realizada não levava em conta a estrutura organizativa dos dados e, portanto, dos sistemas em estudo. Em parte, isto devia-se à falta de métodos e softwares que viabilizassem a generalização da abordagem multinível ou hierárquica. Assim sendo, o analista tinha que escolher a unidade sobre a qual o seu estudo iria incidir. Num estudo de aferição educacional, por exemplo, no qual o desempenho do sistema é aferido pelo desempenho dos alunos, recolhem-se dados sobre os alunos e sobre as escolas, tais como desempenho dos alunos em uma ou mais disciplinas, idade, sexo, cor de pele ou grupo étnico, localização da escola (urbana ou rural), tipo de rede da escola (pública ou particular), dimensão da escola, etc. Muitas análises consideram como unidade de análise a escola e, para tal, tomam o desempenho da escola como a média do desempenho dos seus alunos, fazem o mesmo em relação à idade e tomam a proporção de alunos por sexo e por etnia ou cor de pele. Os dados referentes à unidade de observação micro, o aluno, são resumidos através de médias, proporções ou outras estatísticas, para a unidade de observação macro, a escola. Com esta abordagem perde-se muita informação relativa à variabilidade intra-escola. Por outro lado, poderia ser escolhido o aluno como unidade principal de observação. Neste caso, o estudo do impacto de variáveis relativas a escola implicaria em replicar o dado de escola tantas vezes quantos são os alunos. Isto conduz a estimativas de erro padrão das estimativas dos parâmetros incorretas com implicações nas inferências e conclusões decorrentes.

A literatura mostra-nos alguns exemplos do tipo de erros que podem ocorrer. Por exemplo, Bennett[3] pretende verificar qual o método de ensino que trazia melhores resultados escolares – o método

[2] Extraído na íntegra de Ferrão, M. E., Beltrão, K. I. e Fernandes, C. (2001). Aprendendo sobre Escola Eficaz Evidências do SAEB99, INEP/MEC (no prelo).

[3] Bennet, N. (1976). Teaching styles and pupil progress. Open Books, London.

tradicional (formal) de ensino ou o método progressivo. Na sua abordagem ele conclui que o primeiro era melhor. Posteriormente, a analise foi refeita considerando a hierarquia dos dados e conclui-se que não havia evidência de que algum dos métodos fosse preferível[4].

Como levar em conta a estrutura de agrupamento dos dados? Considere-se um modelo de regressão clássico sem variáveis explicativas (modelo nulo), para modelar os dados do desempenho dos alunos de 3 escolas. Bastará ajustar um intercepto separadamente para cada escola e, tanto a unidade aluno como escola, serão consideradas. Se o foco de interesse do estudo estiver centrado nas 3 escolas e nos alunos que as freqüentam, esta poderá constituir uma solução viável, mas esgota-se à medida que o numero de escolas aumenta. Além disso, na maioria dos estudos, estamos interessados em colher evidências a partir do comportamento global do sistema e não duma determinada escola.

O modelo multinível ou hierárquico respeita a estrutura de agrupamento da população em estudo duma forma parcimoniosa. Além das vantagens acima enunciadas, o modelo multinível também permite estabelecer listas comparativas de desempenho institucional que vêm ganhando relevo em áreas como a educação e saúde; a abordagem multinível fornece a 'ordenação' intrínseca das instituições condicionada às variáveis de controle fundamentais em cada área de estudo.

3.1 Diferença entre Regressão Clássica e Multinível através de exemplo

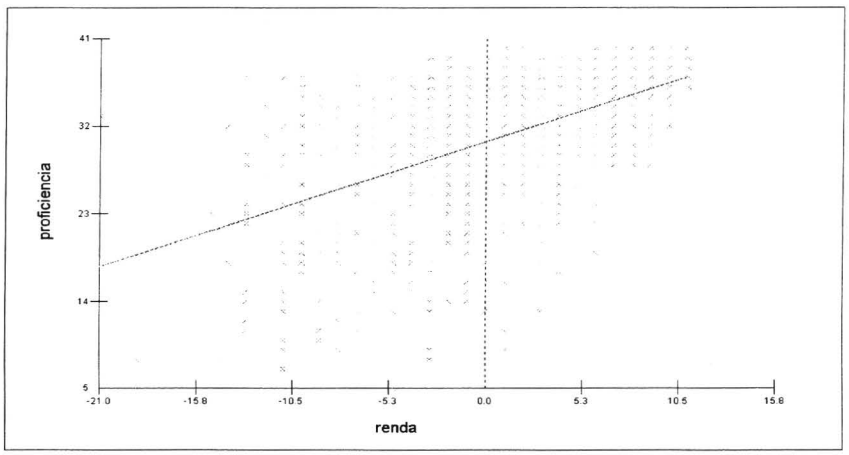
Para melhor ilustrar as diferenças entre a regressão clássica e a multinível, considere um conjunto de dados hipotéticos sobre os resultados escolares dos alunos de 14 escolas e a respectiva renda familiar. Pretende-se saber se existe relação entre a renda familiar e a proficiência do aluno.

3.1.1 Modelo de Regressão Clássica

A figura 1, representa o que seria a reta de regressão clássica para este exemplo, a qual ignora a alocação heterogênea dos alunos às escolas.

[4] Aitkin, M., Anderson, D. e Hinde J., (1981). Statistical modelling of data on teaching styles (with discussion), *Journal of the Royal Statistical Society* A 144 (4): 419-461.
Aitkin, M., Bennet, N. e Hesketh, J. (1981). Teaching styles and pupil progress: a reanalysis, *British Journal of Educational Psychology* 51 (2): 170-186.

Figura 1 – Reta ajustada do modelo de regressão clássico

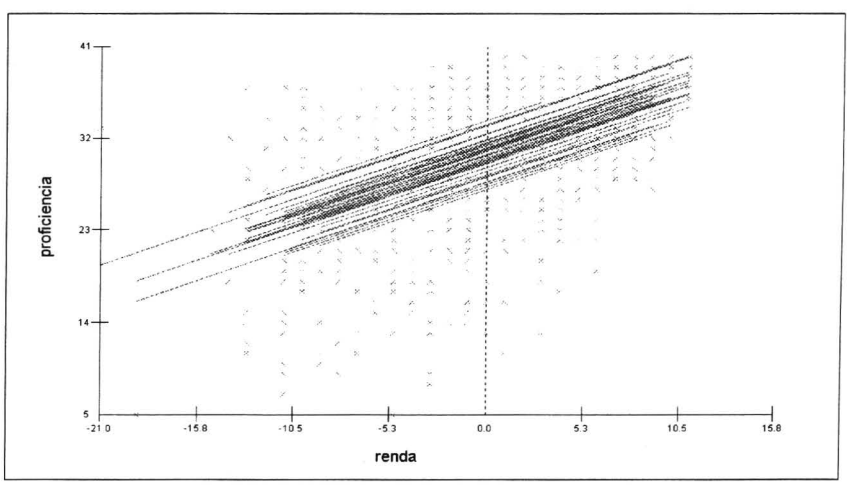


No eixo dos Xs esta representada a renda[5] e no eixo dos Ys a proficiência. Os pontos representam, os valores observados de renda e proficiência para cada aluno. A reta mostra que, em média, alunos com renda familiar mais alta também têm melhores resultados escolares.

3.1.2 Modelo Multinível de Componentes de Variância

A figura 2 representa as retas ajustadas sob um modelo multinível, onde cada uma delas está associada a uma escola. Esta figura considera o intercepto aleatório (um para cada escola) e a inclinação fixa, ou seja, o efeito da renda familiar do aluno na sua proficiência não varia entre as escolas.

Figura 2 – Retas ajustadas do modelo multinível com intercepto aleatório



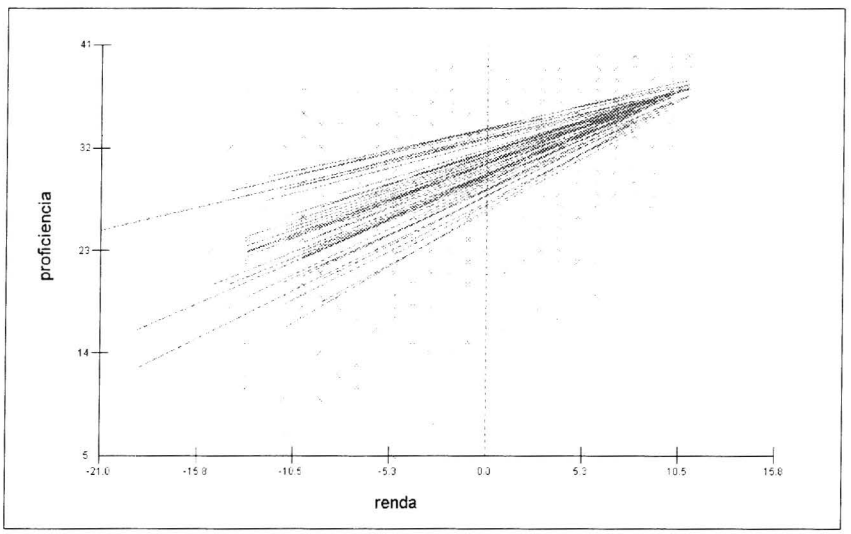
[5]Variável centrada na média. Nesta situação o intercepto é interpretado como o valor médio da proficiência de um aluno cuja renda é igual a média da renda dos alunos de todas as escolas.

A análise do gráfico mostra que a média da proficiência varia de escola para escola, o que da origem a várias retas. No entanto, perante este modelo, o impacto da renda familiar na proficiência do aluno é igual em todas as escolas, ou seja, a inclinação da reta independe da escola.

3.1.3 Modelo Multinível de Coeficientes Aleatórios

A figura 3, representa uma generalização do modelo multinível acima apresentado, onde tanto o intercepto como o coeficiente de inclinação são aleatórios, isto é, ambos variam entre as escolas.

Figura 3 - Retas ajustadas do modelo multinível com intercepto e coeficiente de inclinação aleatórios



Pode verificar-se por meio da análise do gráfico que, efetivamente, alunos com renda familiar mais elevada atingem melhores resultados escolares, mas que o efeito da renda na proficiência depende da escola que o aluno freqüenta. As retas têm inclinação diferentes (o coeficiente associado com a renda familiar é aleatório). Podemos observar que há escolas em que a inclinação da reta é muito tênue – escolas que promovem a equidade social, onde o impacto da renda no desempenho escolar do aluno é pequeno. Por outro lado, outras escolas têm reta com grande inclinação – são escolas onde o nível econômico do agregado familiar influencia fortemente os resultados escolares.

Em resumo, a escola que procuramos é aquela que tem elevado intercepto (proficiência média alta) e coeficiente de inclinação tão próximo de zero quanto possível (renda não influencia os resultados escolares).

3.2 Especificação Formal do Modelo Multinível

Em seguida, mostrar-se-á como é que o modelo multinível acomoda a estrutura hierárquica presente nos dados. Considere-se um modelo com dois níveis: alunos e escolas, e suponha que se pretende explicar o desempenho escolar dos alunos (proficiência) através de duas variáveis explicativas: uma medida no nível do aluno (por exemplo, renda familiar) e a outra no nível da escola (por exemplo, rede administrativa da escola - pública ou privada). Os alunos são identificados pelo índice i e as escolas pelo índice k. O índice k varia de 1 a K (sendo K o número total de escolas em estudo) e o índice i varia de 1 a n_k (sendo n_k o número de alunos que pertence à escola k). A variável resposta do modelo é a proficiência do aluno i pertencente à escola k, $proficiência_{ik}$, e a variável explicativa associada a este aluno, é a respectiva renda familiar, $renda\_familiar_{ik}$. O modelo de regressão clássico apresentado na equação (1) especifica a relação entre estas duas variáveis e é o modelo que está subjacente ao gráfico apresentado na figura 1.

$$proficiência_{ik} = \beta_0 + \beta_1 * renda\_familiar_{ik} + e_{ik} \qquad (1)$$

onde β_0 e β_1, são o intercepto e o coeficiente de inclinação, respectivamente. Estes parâmetros são desconhecidos e devem ser estimados a partir dos dados. O intercepto β_0 pode ser interpretado como o valor esperado da proficiência para os alunos que têm valor nulo de renda[6]. O coeficiente de inclinação, β_1, representa o impacto da renda familiar no desempenho escolar do aluno. Assim, por cada unidade adicional na renda familiar, *cœteris paribus*, a média do desempenho do aluno observará uma variação de β_1 unidades. O termo e_{ik} é o distúrbio aleatório ou erro do modelo, associado aos efeitos individuais do aluno não captados pela componente determinística do modelo, e o pressuposto usual é que tenha uma distribuição normal com média nula e variância, σ_e^2, constante entre os grupos, e que sejam não correlacionados entre si, isto é, $e_{ik} \sim NID(0, \sigma_e^2)$.

No modelo de dois níveis (alunos e escolas) tanto o intercepto como o coeficiente de inclinação podem ser considerados variáveis aleatórias que variam de escola para escola. Na equação (2) apresenta-se o modelo em que apenas o intercepto varia aleatoriamente entre as escolas. Este é o modelo subjacente ao gráfico apresentado na figura 2.

[6]Variável centrada na média, descrita na nota anterior

$$proficiência_{ik} = \beta_{0k} + \beta_1 * renda\_familiar_{ik} + e_{ik} \quad (2)$$
$$\beta_{0k} = \gamma_{00} + u_{0k}$$
$$e_{ik} \sim NID(0, \sigma_e^2)$$
$$u_{0k} \sim NID(0, \sigma_{u0}^2)$$

A primeira característica a ser observada é que neste modelo o parâmetro do intercepto, β_{0k} , tem o índice *k*, indicando a existência de um parâmetro para cada escola, conforme enunciado previamente. Ou seja, o valor médio da proficiência está dividido na contribuição da escola (β_{0k}) e no desvio (e_{ik}) de cada estudante à contribuição da escola. Na segunda linha de (2) pode observar-se que a contribuição de cada escola foi decomposta na média global da proficiência (envolvendo todas as escolas), γ_{00}, e no afastamento de cada escola, u_{0k}, a essa média global. Este é o efeito individual da escola *k* (componente aleatória do nível 2 associada ao intercepto).

Os parâmetros desconhecidos do modelo: $\beta_1, \gamma_{00}, \sigma_{u0}^2$ *e* σ_e^2 são estimados a partir dos dados, sendo os primeiros dois parâmetros designados por parâmetros fixos e os dois últimos por parâmetros aleatórios. A componente aleatória associada ao intercepto tem variância σ_{u0}^2, representando a variabilidade do intercepto entre escolas. O erro de nível 1, e_{ik}, tem variância σ_e^2 e representa a variabilidade intra-escola.

O modelo especificado em (3) (correspondente ao gráfico 3) além do intercepto aleatório também tem o coeficiente de inclinação aleatório.

$$proficiência_{ik} = \beta_{0k} + \beta_{1k} * renda\_familiar_{ik} + e_{ik} \quad (3)$$
$$\beta_{0k} = \gamma_{00} + u_{0k}$$
$$\beta_{1k} = \gamma_{10} + u_{1k}$$
$$e_{ik} \sim NID(0, \sigma_e^2)$$
$$u_{0k} \sim NID(0, \sigma_{u0}^2)$$
$$u_{1k} \sim NID(0, \sigma_{u1}^2)$$

Similarmente à inclusão do parâmetro do intercepto, a inclusão do parâmetro de inclinação específico para cada escola indica que a relação entre a proficiência e o nível socioeconômico varia de escola para escola.

Até ao momento, considerou-se apenas uma variável explicativa, renda do agregado familiar do aluno, medida no nível 1. Na seqüência, apresentar-se-á o modelo (2) acrescentando uma variável explicativa medida ao nível da escola. Esta variável é a rede administrativa, $rede\_escolar_k$, uma variável binária que designa se a escola é pública ou particular. A sua inclusão no nível 2 do modelo dá-se substituindo a equação (2.a) na segunda linha do modelo (2),

$$\beta_{0k} = \gamma_{00} + \gamma_{01} rede\_escolar_k + u_{0k} \qquad (2.a)$$

$$\begin{aligned} proficiência_{ik} &= (\gamma_{00} + \gamma_{01} rede\_escolar_k + u_{0k}) + \beta_1 * renda\_familiar_{ik} + e_{ik} \\ &= (\gamma_{00} + \beta_1 renda\_familiar_{ik} + \gamma_{01} rede\_escolar) + (e_{ik} + u_{0k}) \end{aligned} \qquad (4)$$

Finalmente, substituindo a equação (2.a) na primeira linha da equação (2), podemos identificar duas componentes distintas no modelo. A componente determinística ou sistemática do modelo é dada pela expressão $(\gamma_{00} + \beta_1 renda\_familiar_{ik} + \gamma_{01} rede\_escolar)$, enquanto que a componente aleatória ou estocástica é dada por $(e_{ik} + u_{0k})$. Não é demais reforçar a idéia de que a parte aleatória ou estocástica do modelo representa numerosos efeitos aleatórios que impactam a proficiência do aluno, atuando tanto ao nível do aluno como ao nível da escola, e que não são captados pela parte determinística do modelo.

A componente aleatória $(e_{ik} + u_{0k})$ do modelo está decomposta no erro de nível 1, e_{ik}, e no erro de nível 2, u_{0k}. As estimativas destes erros são os resíduos (o que no modelo fica por explicar). Assim, a variância residual do modelo é dada por $(\sigma_e^2 + \sigma_{u0}^2)$. Com a decomposição da variância residual do modelo torna-se fácil avaliar o impacto de cada variável (seja ela medida ao nível do aluno ou da escola) na explicação da proficiência. Além disso, no modelo nulo (modelo sem variáveis explicativas), é a proporção da estimativa da variância entre escolas, σ_{u0}^2, face à variância total (variância entre-escolas e variância intra-escolas), que sinaliza a presença do "**efeito-escola**" no desempenho escolar do aluno. A pesquisa em avaliação educacional deverá estar orientada à investigação das características das escolas (características intra-escolares) passíveis de intervenção que contribuem positivamente para esse efeito. Maior refinamento deste assunto é desenvolvido na seção 3.4.

Finalmente apresenta-se o modelo (3) incluindo a variável explicativa *rede_escolar* na equação do coeficiente de inclinação. A equação resultante é a seguinte (5):

$$proficiência_{ik} = (\gamma_{00} + \gamma_{01} rede\_escolar_k + u_{0k}) + (\gamma_{10} + \gamma_{11} rede\_escolar_k + u_{1k}) * renda\_familiar_{ik} + e_{ik}$$
$$= \gamma_{00} + \gamma_{10} renda\_familiar_{ik} + \gamma_{01} rede\_escolar + \gamma_{11} rede\_escolar_k * renda\_familiar_{ik} +$$
$$+ u_{1k} * renda\_familiar + e_{ik} + u_{0k} \qquad (5)$$

3.3 Coeficiente de Correlação intra-escola e o efeito-escola

Uma das questões de mais interesse é estudar o tamanho de σ^2_{u0} . Se, relativamente à variância total, é pequeno então podemos concluir que a escola tem pouco efeito, ou, dito de outro modo, saber qual é a escola onde o aluno estuda não ajuda a explicar os resultados escolares atingidos pois eles poderiam tê-los atingido em qualquer outra escola.

O coeficiente de correlação intra-escola é uma estatística que permite aferir sobre a magnitude do efeito-escola. Assumindo que u_{0k} e e_{ik} variam independentemente, o coeficiente de correlação intra-escola define-se em (5):

$$\rho = \frac{\sigma^2_{u0}}{\sigma^2_e + \sigma^2_{u0}} \qquad (5)$$

No modelo nulo (modelo sem variáveis explicativas)[7], o coeficiente de correlação representa o tamanho relativo da variância entre escolas[8]. O coeficiente varia de 0 a 1. Quando o seu valor é nulo significa que os alunos estão homogeneamente distribuídos entre as escolas e que o desempenho do aluno independe da escola que ele freqüenta. Nesta situação hipotética, $\sigma_{u0}{}^2$ seria estatisticamente igual a zero, significando que toda a variância da proficiência seria devido à variabilidade entre alunos e, por conseguinte, a quota de responsabilidade da escola nos resultados atingidos pelos alunos, ou o efeito-escola, seria inexistente.

Na situação extrema a esta, quando o coeficiente de correlação intra-escola toma o valor 1, toda a variabilidade no desempenho dos alunos deve-se à diferença entre as escolas e, nesta situação hipotética, as características individuais do aluno em nada afetariam o seu desempenho escolar ficando este a dever-se inteiramente às características da escola que ele freqüenta.

[7] Adiante mostrar-se-á que julgar o efeito-escola com base num modelo tão simples como o modelo nulo pode ser prematuro.
[8] Também é a correlação da proficiência entre dois alunos da mesma escola.

4 Tutorial

Este tutorial constitui-se numa introdução prática à modelagem multinível com base no software MLwiN. Destacam-se os procedimentos básicos utilizados para a especificação de um modelo multinível, estimação de parâmetros, uso de inferências e análise gráficas.

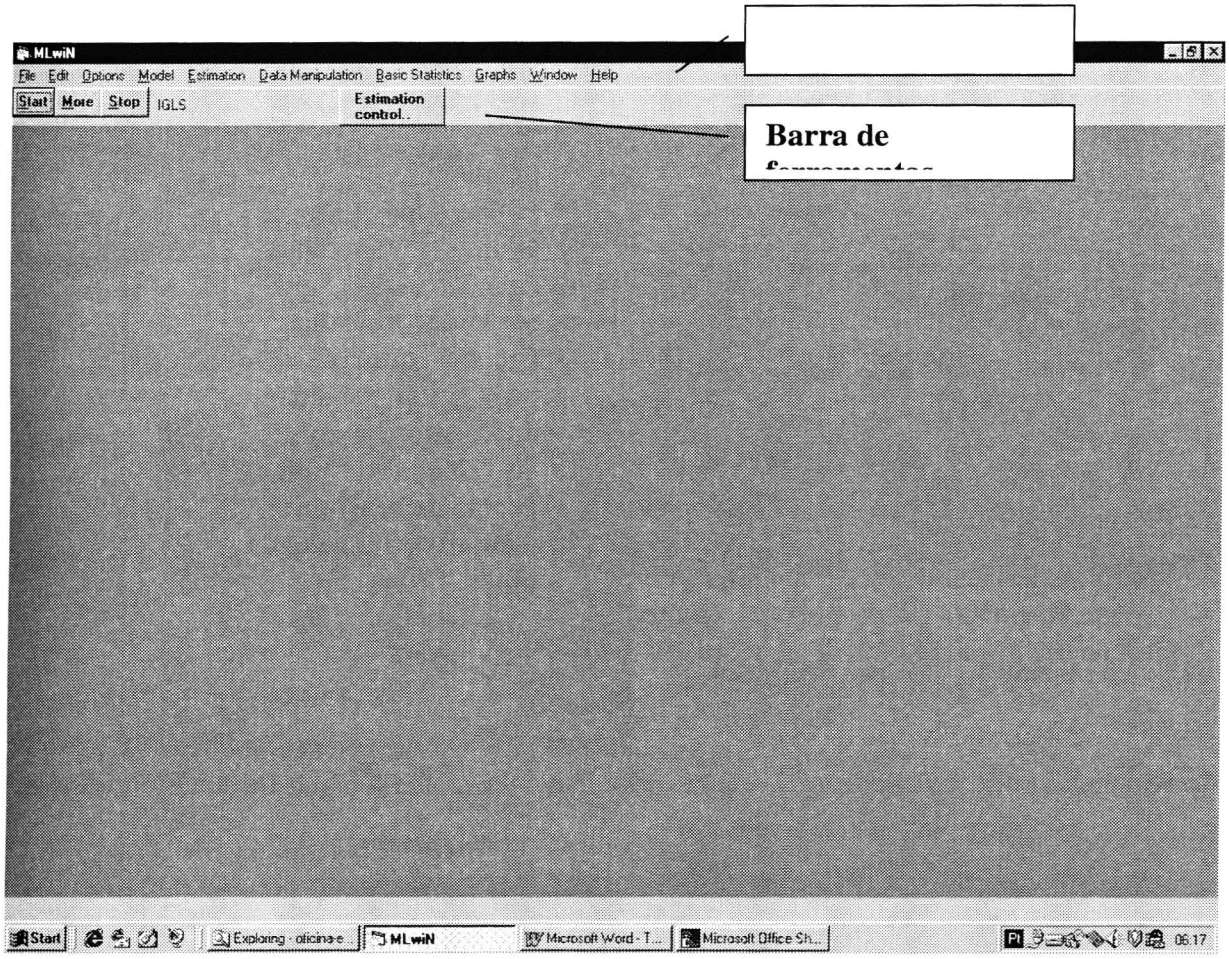
Para ilustração, usamos um conjunto de dados educacionais cuja planilha é fornecida e se descreve adiante. No início da análise, o usuário terá que criar a planilha pela introdução direta dos dados ou, alternativamente, ler os dados de um outro arquivo criado. As facilidades para conseguir isto são mencionadas no fim deste documento.

O banco de dados a ser utilizado foi criado a partir das informações do Sistema Nacional de Avaliação da Educação Básica (SAEB) do Brasil e faz parte do projeto de investigação do efeito-escola, nas cinco macro regiões do Brasil, desenvolvido por Ferrão, Beltrão e Fernandes (2001)[9]. Este banco, com 3223 observações, refere-se aos alunos da 4ª série do ensino fundamental, residentes na região Sudeste, submetidos ao exame de matemática.

4.1 Abrir a planilha e examinar os dados

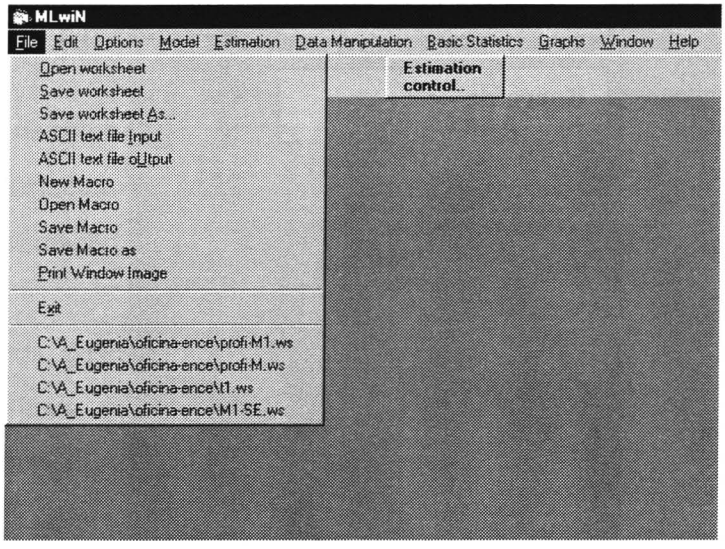
Ao iniciar o *MLwiN* aparece a janela principal e imediatamente abaixo do título do programa, aparece a barra de menu e a barra de ferramentas, tal como se segue:

[9]* Sendo o efeito-escola estatisticamente significativo, o estudo procura identificar alguns dos fatores que tornam umas escolas mais eficazes do que outras na promoção da aprendizagem e desenvolvimento dos alunos, tendo em consideração as diversas características dos alunos da 4a. série do Ensino Fundamental.



Na barra de ferramentas encontram-se os comandos referentes aos procedimentos de estimação dos modelos. Estes procedimentos serão descritos em detalhes mais adiante, na seção 4.5. A região central da janela é denominada área de trabalho, e na borda inferior têm-se a barra de Status, que nos permite monitorar o progresso do programa.

Usaremos a opção **File** na barra de menus para abrir a planilha. Dê um clique em **File** para obter a seguinte lista de operações:

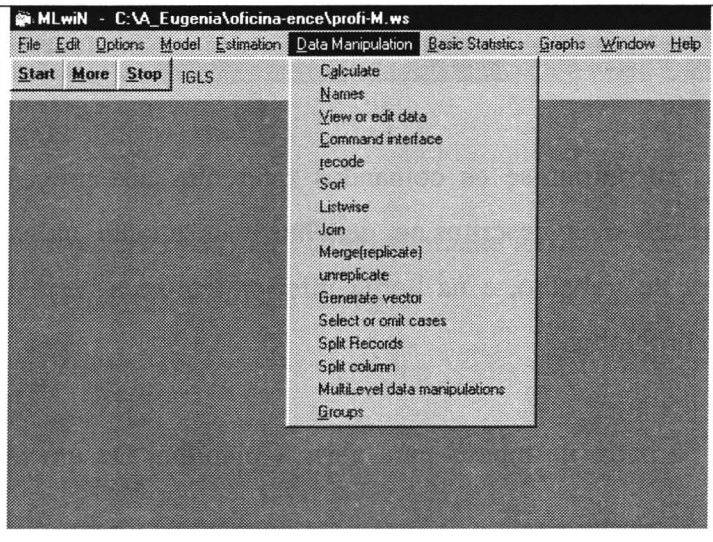


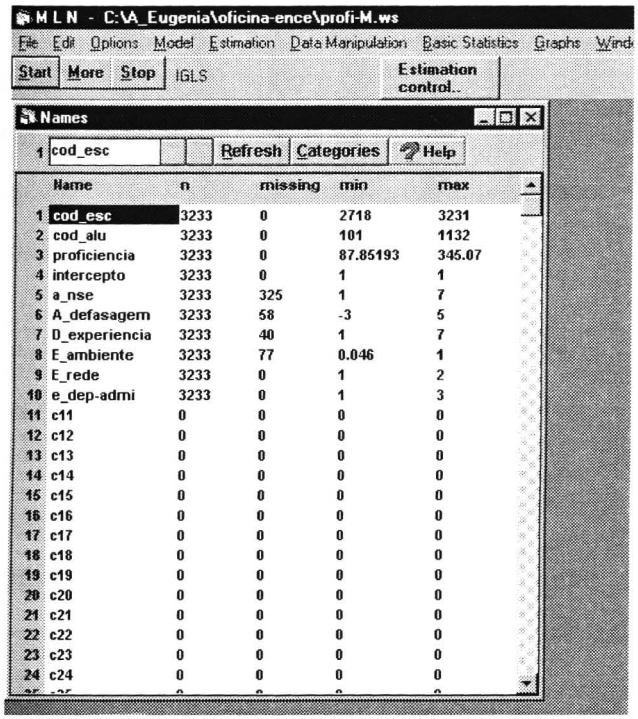
Clique em **Open worksheet** para visualizar todas as planilhas armazenadas e assim escolher aquela em que pretende trabalhar. Escolha **tutorial.ws** com um clique duplo sobre o nome do arquivo ou, depois de selecionar o arquivo, clique em **Open**.

4.2 Visualização e Edição da Base de Dados

Já com o arquivo aberto poderá visualizar o seu conteúdo. Para tal escolha o menu **Data Manipulation.**

As opções oferecidas estão associadas a operações realizadas sobre a base de dados. Selecione a opção **Names** e visualizará algumas características da base de dados.





	Name	n	missing	min	max
1	cod_esc	3233	0	2718	3231
2	cod_alu	3233	0	101	1132
3	proficiencia	3233	0	87.85193	345.07
4	intercepto	3233	0	1	1
5	a_nse	3233	325	1	7
6	A_defasagem	3233	58	-3	5
7	D_experiencia	3233	40	1	7
8	E_ambiente	3233	77	0.046	1
9	E_rede	3233	0	1	2
10	e_dep-admi	3233	0	1	3
11	c11	0	0	0	0
12	c12	0	0	0	0
13	c13	0	0	0	0
14	c14	0	0	0	0
15	c15	0	0	0	0
16	c16	0	0	0	0
17	c17	0	0	0	0
18	c18	0	0	0	0
19	c19	0	0	0	0
20	c20	0	0	0	0
21	c21	0	0	0	0
22	c22	0	0	0	0
23	c23	0	0	0	0
24	c24	0	0	0	0

Na primeira coluna encontra-se o nome das variáveis, seguida pelo número de registos (casos) na base de dados. A coluna seguinte mostra o número de casos com *valores omissos* e, finalmente, os valores mínimo e máximo para cada variável.

Para dar um novo nome à variável, posicione o cursor no nome que deseja alterar e escreva o novo nome. Depois de pressionar 'enter' verificar se o nome da variável foi alterado.

Se alguma das variáveis é categórica, por exemplo "E_rede", você poderá definir o nome das respectivas categorias. Para isso, selecione a variável com um clique duplo e aperte depois em **categories**.

Nesta janela, defina 1 como sendo o nível da variável associada à rede pública e 2 à rede particular. Depois de escrever publica e particular, clique **Apply** e **Quit**.

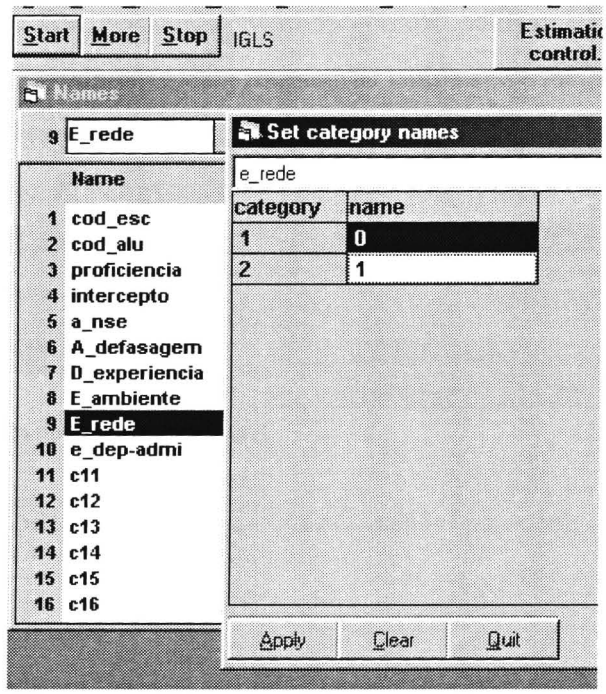
Como exercício, repita para a variável referente à dependência administrativa, "E_dep-admi".

Salve as alterações efetuadas sobre a base de dados até ao momento. Selecione:

File

Save Worksheet as...

E dê um novo nome ao arquivo, por exemplo, avalia-mat.



As variáveis designadas por "cod_esc" e "cod_alu" contêm os códigos identificadores de escola e de alunos, respectivamente. A variável "proficiência" é a classificação obtida no exame de matemática, por cada aluno da 4ª série, na região Sudeste. A variável "intercepto" é uma coluna de 1's[10] e as outras variáveis são relativas aos alunos e as escolas ("A_nse", é o nível sócio econômico; "a_defasagem", é o

[10] Usada como variável explicativa.

número de anos que o aluno se encontra defasado face à idade adequada para a série; "d_experiência", é a experiência do diretor mensurada em termos de número de anos na direção da escola; "E_ambiente" é um indicador da qualidade do ambiente e clima da escola; "E_rede" indica se a escola pertence a rede pública ou particular e "E_dep_admini" indica qual a dependência administrativa da escola (estadual, municipal ou particular).

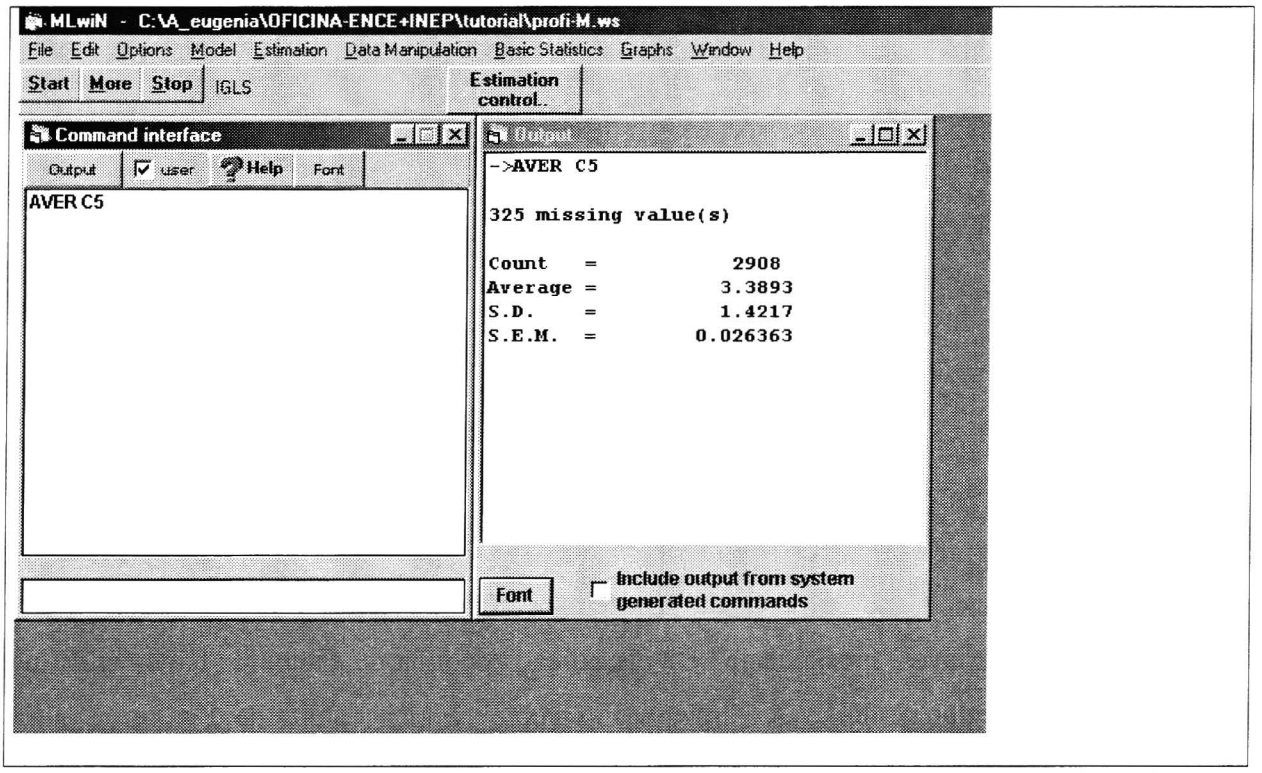
Para ver o valor das variáveis e poder editá-las, clique em:

Data Manipulation

View or Edit data

A janela mostra as três primeiras variáveis da base de dados. Clique em **view** para selecionar outras colunas que deseje visualizar e, eventualmente, editá-las. A opção de **Command Interface** abre uma linha para que comandos possam ser executados por meio de sua digitação. O sistema de ajuda lista exaustivamente os comandos possíveis e a respectiva sintaxe. Usar-se-á um exemplo de aplicação do Command Interface: iremos *centrar a variável* referente ao status socioeconômico do aluno e também criar a respectiva *variável contextual* para a escola.

Para centrar a variável na média execute o comando **AVERage c5** para conhecer o valor médio da variável.



Verifica-se a existência de 325 registos com valores omissos e 2.908 casos válidos. A média é 3,3893 e para centrar a variável na média teremos que lhe subtrair esse valor. Execute os comandos:

calc c11=c5-3.3893
name c11 "A_nse-c"

Para criar a variável contextual de escola a partir do nível sócio econômico dos alunos precisamos calcular a média do nível sócio-econômica dos alunos de cada escola. Para isto, digite:

MLAVER c1 c11 c12
Name c12 "E_nse-c"

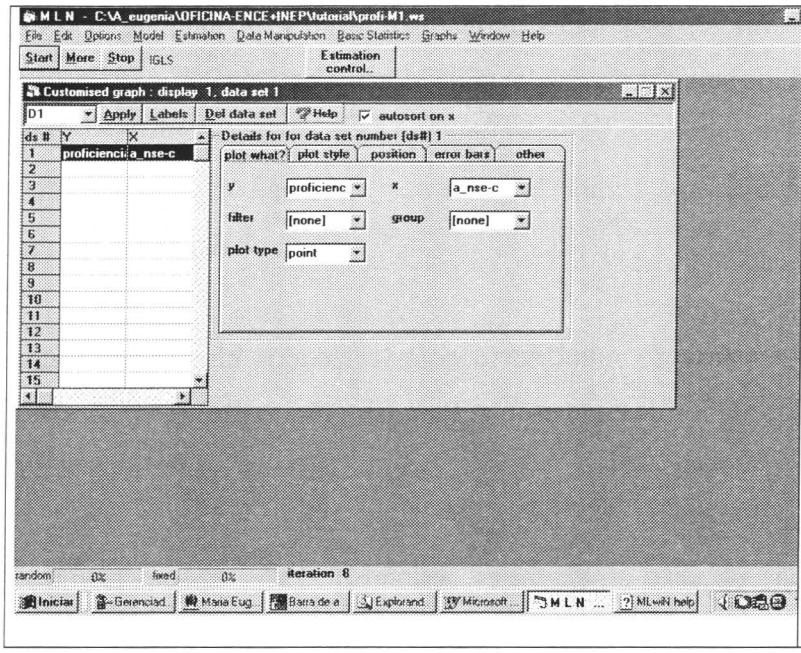
O comando **MLAVER** calcula a média da variável armazenada em C11 para cada escola cujo código é definido em C1 e o resultado é depositado na variável c12 que, com o comando **NAME,** designaremos por 'A_nse-c'.

As demais opções do comando **Data Manipulation** referem-se às operações sobre a base de dados, tais como ordenação, criação de um vetor, recodificação de variáveis, etc.

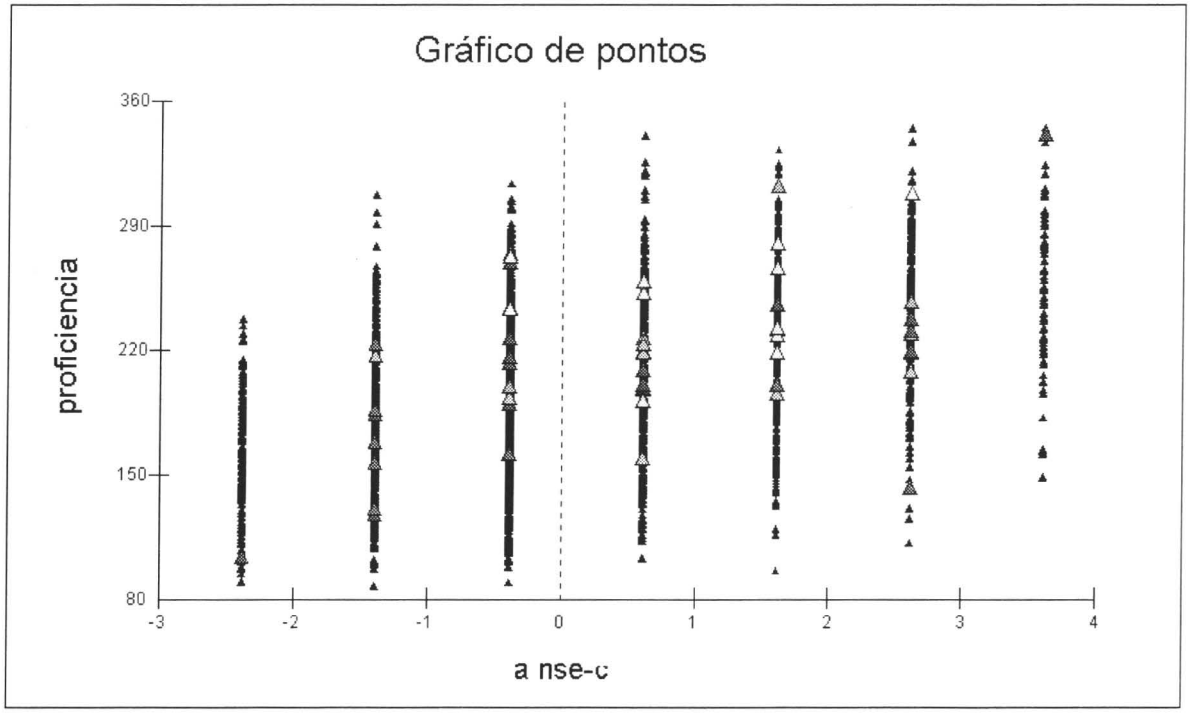
A seguir abordaremos brevemente as facilidades gráficas do *MlwiN.*

4.3 Traçar Gráficos

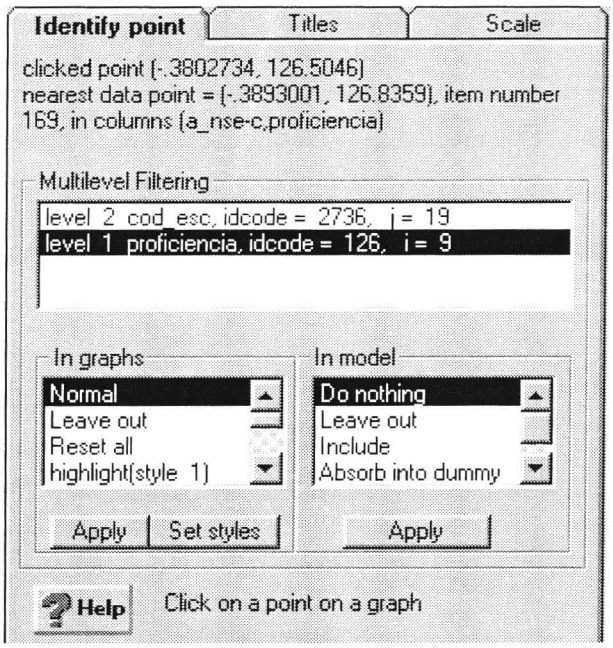
Usando o menu **Graphs,** selecione a opção **Customised graphs** e a seguinte janela aparecerá:



Existe um número considerável de opções disponíveis para traçar gráficos, mas por enquanto, vamos apenas demonstrar como traçar um gráfico de pontos (scatter plot) e de linhas. Neste gráfico, a classificação do exame (PROFICIÊNCIA) será inserida no eixo dos Y, e o nível socioeconômico do aluno (A_NSE-C) no eixo dos X. Para tal, vamos colocar os nomes destas variáveis nas células associadas a Y e X, respectivamente, no retângulo de **settings for data set(ds) .** Depois de pressionar e **Apply,** o gráfico é produzido mostrando a relação existente entre as variáveis envolvidas.



Com o cursor posicionado em cima de qualquer ponto do gráfico, dê um clique e quadro referente às opções do gráfico aparecerá:

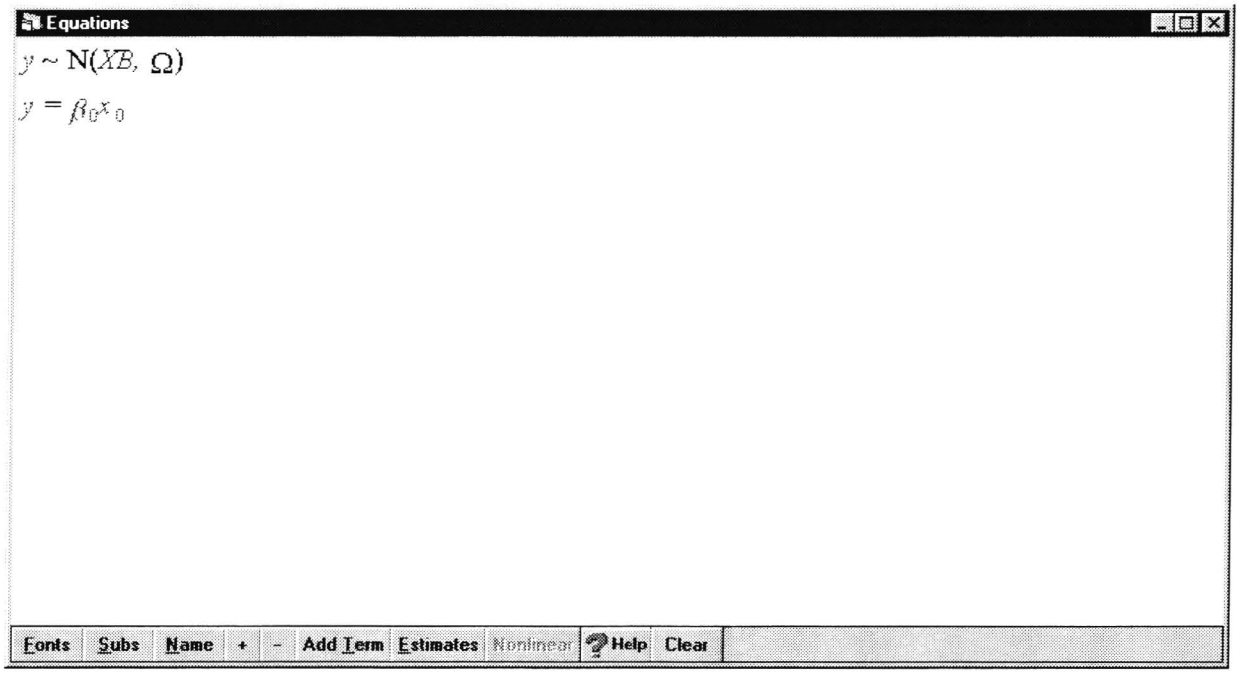


Com a opção de **Identity point**, você poderá visualizar as coordenadas desse ponto (isto é, o valor exato das variáveis 'proficiência' e 'a_nse-c'), bem como a identificação do ponto mais próximo. Na opção **Titles** poderá definir os títulos e os nomes dos eixos para o gráfico. Em **Scale** poderá ajustar a escala do gráfico.

Vários comandos já foram executados e, seria prudente salvar a planilha pois se algum problema ocorrer com o MLWin, todo trabalho já realizado será perdido. Para isso, no menu **File** escolha a opção **Save worksheet As.** Digite o nome do arquivo. Mais uma vez se aconselha a usar um nome diferente do anterior. Pressione em **Save**.

4.4 Especificação do modelo de componentes de variância

Selecione no menu **Model** a opção **Equations** e, no canto superior esquerdo, poderá observar algo como:



A primeira linha especifica a distribuição assumida, no caso é a distribuição Normal, com os respectivos parâmetros. O vector resposta tem média definida, em notação matricial, pela *componente fixa* $X\beta$, e a *componente aleatória* é descrita pela matriz de variância e covariância Ω.

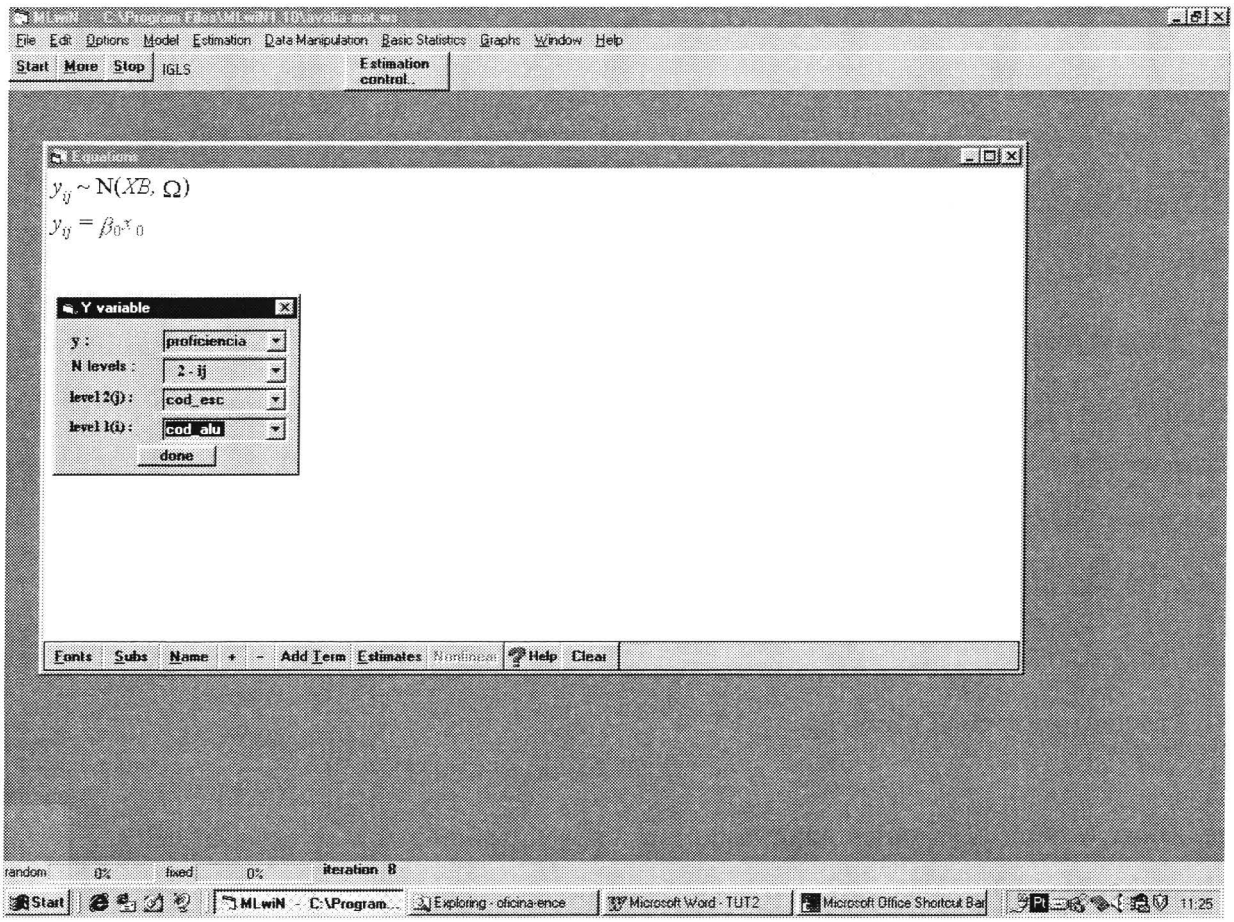
A barra de ferramentas da janela permite ao usuário definir fontes diferentes para a notação – ferramenta **Fonts**; **Name** substitui as variáveis resposta e explicativas pelo respectivo nome após terem sido especificadas; + e – permite que o modelo seja ou não completamente mostrado na tela; **Add Term** inclui novas variáveis no modelo; **Estimates** mostra as estimavas do modelo; **Nonlinear** é uma opção para dados cuja variável resposta não apresenta distribuição Normal – assunto que não é abordado nesta fase do curso. O usuário ainda tem à disposição mais um botão para aceder ao sistema de ajuda, **help**, e, finalmente, a opção de **Clear** permite reinicializar qualquer modelo especificado.

4.5 Definição da variável resposta e estrutura hierárquica

Pressione em ***y***. Note que inicialmente ***y*** surge em vermelho, indicando que a variável resposta ainda não foi definida. Associe ***y*** a variável proficiência e especifique o modelo como tendo 2 níveis. Duas novas células surgirão para que as variáveis associadas aos níveis sejam especificadas. Sendo assim, digite

"***cod_esc***" na célula referente ao segundo nível (***índice j***) e "***cod_alu***" na célula referente ao primeiro (índice i).

Se se considerasse o nível de agrupamento de turma, como seria definida a estrutura hierárquica do modelo? Comente sobre as vantagens.



Agora pressione em **done** para que todas as operações tenham efeito.

4.5.1 Definição das variáveis explicativas

Agora pressione em x_0 (que está vermelho) e selecione **intercepto** como a primeira variável explicativa. Como o conteúdo de **intercepto** é uma constante igual a 1, estamos assim a definir o intercepto na componente fixa do modelo. Até agora especificamos a componente fixa do modelo. Para definir a componente aleatória, pressione em β_0 e deverá assinalar quais os níveis para os quais pretende o coeficiente aleatório. No caso, deverá assinalar tanto o nível 1 (i) como o nível 2 (j). Por fim, pressione **done.**

Para visualizar o modelo especificado, clique na tecla '+' na barra de ferramentas situada na parte inferior da janela. Pressionado-se a tecla **Subs** visualizará no índice das variáveis os códigos associados com os níveis de agrupamento.

Este modelo multinível especificado apenas com o intercepto, sem outras variáveis explicativas, é designado por modelo nulo.

Segue-se a sua especificação:

$$\text{proficiencia}_{cod\_alu,\ cod\_esc} \sim N(XB,\ \Omega)$$

$$\text{proficiencia}_{cod\_alu,\ cod\_esc} = \beta_{0cod\_alu,\ cod\_esc}\text{Cons}$$

$$\beta_{0cod\_alu,\ cod\_esc} = \beta_0 + u_{0cod\_esc} + e_{0cod\_alu,\ cod\_esc}$$

$$\left[u_{0cod\_esc}\right] \sim N(0,\ \Omega_u) \ :\ \Omega_u = \left[\sigma^2_{u0}\right]$$

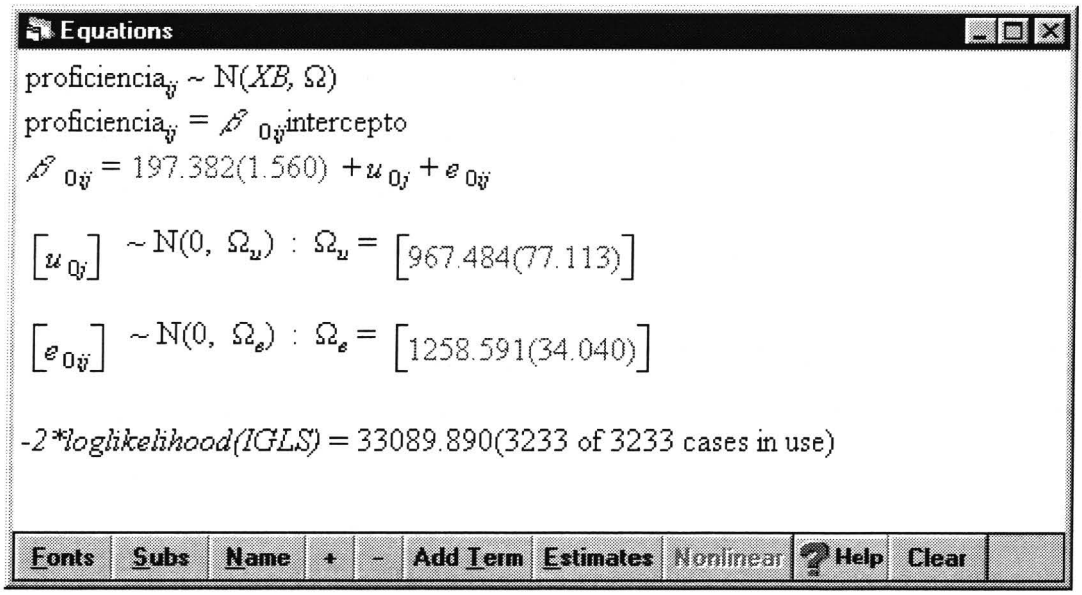
$$\left[e_{0cod\_alu,\ cod\_esc}\right] \sim N(0,\ \Omega_e) \ :\ \Omega_e = \left[\sigma^2_{e0}\right]$$

4.6 Procedimento de Estimação

Se pressionar o botão **Estimates** aparecerão em azul os parâmetros a serem estimados. Para ajustar o modelo especificado, basta pressionar o botão **Start** no menu de ferramentas. Caso pretenda interromper o processo de estimação pressione a opção **Stop**. Neste exemplo, o procedimento de estimação termina depois de 4 iterações e os parâmetros ficam verdes sinalizando que a convergência foi atingida.
Pressionando uma vez mais em **Estimates** podemos ver o valor das estimativas, para os parâmetros fixos e aleatórios, seus respectivos erros-padrão, assim como o valor de *-2 log(verosimilhança)* que permite calcular a estatística designada por *deviance*[11], usada adiante.

As estimativas resultantes estão assinaladas em verde e são as seguintes:

[11] A *deviance* é uma estatística que permite seleccionar o melhor dentre dois modelos. É considerada uma estatística de qualidade do ajuste, e segue a distribuição de qui-quadrado com tantos graus de liberdade quanto a diferença de parâmetros nos dois modelos avaliados.



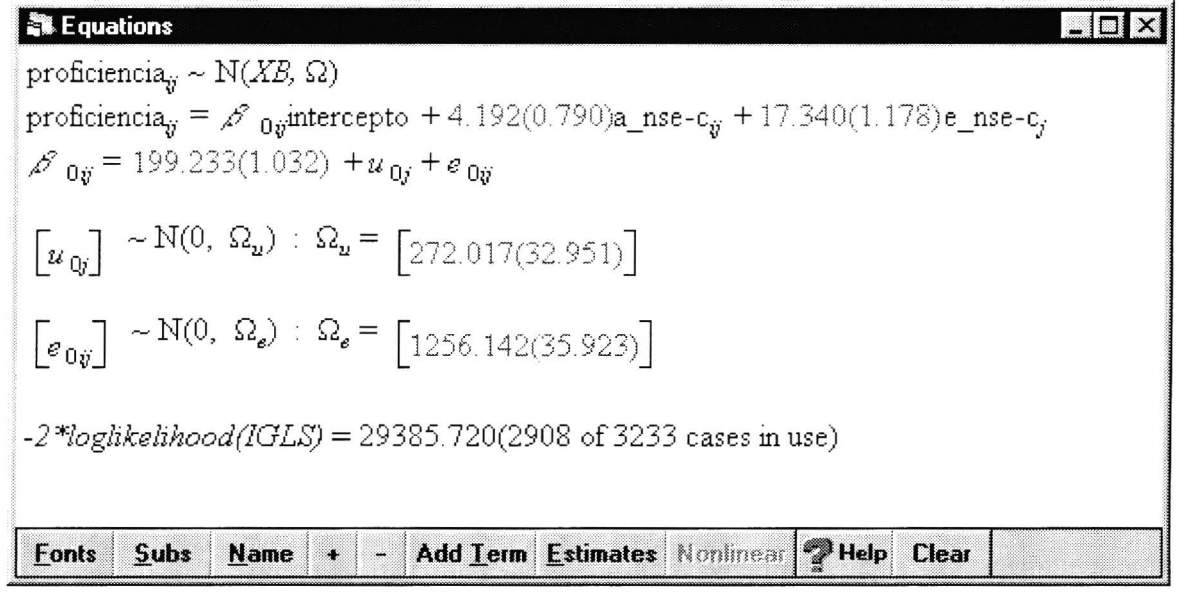
O modelo ajustado contém apenas o intercepto e os parâmetros aleatórios que medem a variação entre escolas e indivíduos respectivamente. De acordo com o modelo, a proficiência média em matemática é aproximadamente 197,38. A estimativa da variância entre escolas é 967.48 e a da variância entre alunos é 1258,59. Com estas duas componentes pode-se calcular o coeficiente de correlação intra-classe, (*intra-escola*) que mede o quanto da variação total é explicada pela unidade de segundo nível, ou seja, a escola. O coeficiente de correlação intra-classe é dado por:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \frac{967{,}48}{967{,}48 + 1258{,}59} = 0{,}4346 \text{ ou } 43{,}46\%$$

De acordo com o resultado do coeficiente de correlação intra-classe, 43,46% da variação total é devido à variação entre escolas. O procedimento usado para a estimação dos parâmetros (mínimos quadrados iterativos (IGLS/RIGLS)) é equivalente ao procedimento de máxima verosimilhança quando o pressuposto de Normalidade está subjacente. Sendo assim, podemos utilizar o valor da estimativa de **–2*loglikelihood** (33089.89) para implementar o teste da Razão de Verossimilhança e, assim, avaliar se os parâmetros aleatórios são estatisticamente significativos. Para isto, é preciso se calcular o valor de **–2*loglikelihood** no modelo sem o parâmetro aleatório, ou seja, o modelo com apenas o intercepto fixo. Há duas formas de se implementar este modelo. No primeiro, especifica-se a constante como aleatória apenas no primeiro nível. A segunda forma é implementada a partir da criação de um vetor de zeros que é associado com o segundo nível. Nos dados analisados, a estimativa de **–2*loglikelihood** para o modelo fixo foi igual a 34069,160. A estatística teste é a diferença entre os valores de **–2*loglikelihood** (34069,160-33089.890=979,270) que deve ser comparada com o valor de obtido de uma distribuição qui-

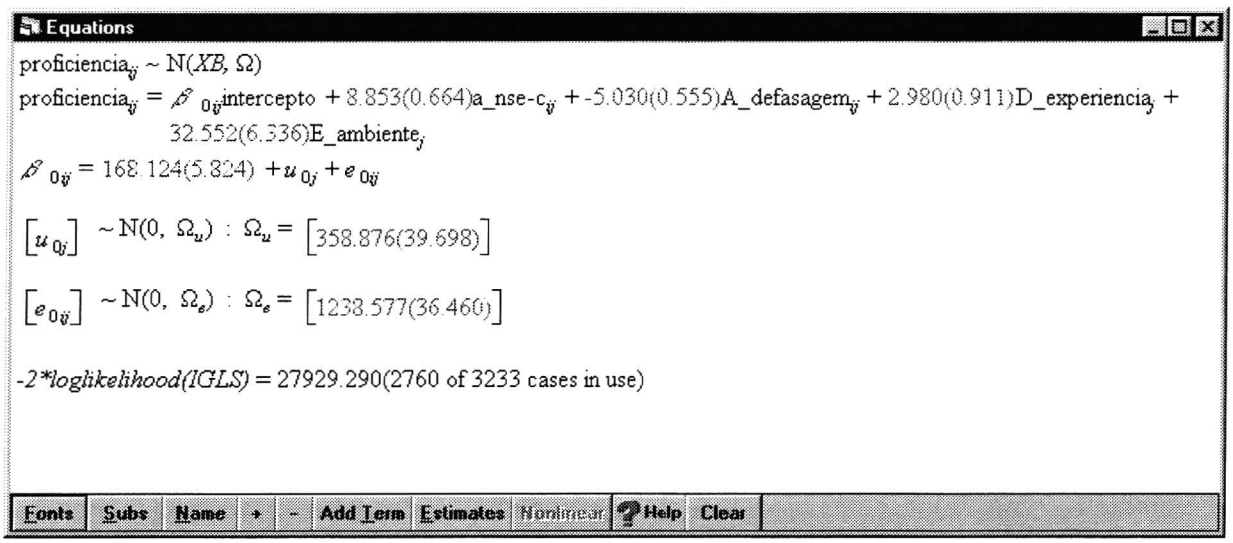
quadrado com número de graus de liberdade igual à diferença no número de parâmetros nos dois modelos (no exemplo considerado, qui-quadrado com 1 grau de liberdade que é igual a 3,84). Conforme pode ser visto, o parâmetro aleatório é estatisticamente significativo. Outra forma de se avaliar o efeito do parâmetro aleatório é utilizando-se uma estatística teste disponível no ***MLwiN*** que será apresentada posteriormente. Esta estatística também é utilizada para testar a significância estatística dos parâmetros fixos.

Guiados pela definição do efeito-escola e de coeficiente de correlação intra-escola dados na seção 3.4, podemos ser levados a pensar que as percentagens resultantes do exercício constituem a quota de responsabilidade da escola no desempenho acadêmico dos seus alunos. Isto não é de todo verdade, já que a alocação dos alunos às escolas não é aleatória e que o nível socioeconômico das famílias contribui em muito na escolha da escola que o aluno vai freqüentar. Assim, o efeito-escola deverá ser expurgado dessa componente extra-escolar. Para isso deve-se acrescentar ao modelo as variáveis do nível socioeconômico - do aluno e a variável contextual da escola. Na barra inferior de opções, pressione em **add term** e depois em x_1 para lhe associar às variáveis explicativas **a_nse-c** e **e_nse-c**, respectivamente. Pressione **More** para reestimar os parâmetros. Os resultados obtidos são apresentados abaixo:



Note que apenas o índice *j* aparece associado com a variável **e_nse-c**, indicando que a mesma pertence ao segundo nível. Pode-se ver que houve uma redução substancial no coeficiente de correlação intra-classe que passou de 43,5% para 17,8% .

Ao longo do processo de modelagem dos dados, decidiu-se retirar a variável contextual, nível socio-econômico médio, pois esta é extremamente correlacionada com a variável ambiente, que foi criada a partir de várias características da escola. O modelo abaixo contém assim quatro variáveis: nível socio-econômico do aluno centrado na média geral, defasagem do aluno, a experiência do diretor e, a que descreve o ambiente da escola.



De acordo com o modelo, a cada unidade de variação na variável status socio-econômico centrada resulta em um aumento de aproximadamente 9 na proficiência em matemática. Por outro lado, cada ano de defasagem do aluno reduz a proficiência em 5. Um aluno cujo diretor tem 3 anos de experiência teria, em média, sua proficiência elevada em 9. A variável ambiente tem um efeito forte sobre a proficiência do aluno. Conforme ressaltado anteriormente esta variável está restrita ao intervalo [0, 1]. Numa escola, com um ambiente de alta qualidade (valor igual a 1) a proficiência de seus alunos é acrescida de 32,5.

Tabelas de Resultados

Para visualizar as estimativas em forma de tabela, basta selecionar o comando **Estimate tables** que se encontra dentro do menu **Model** e obtemos uma tabela com os resultados do ajuste, tal como se segue:

Estimates

FIXED PART | S E S P C N | Help

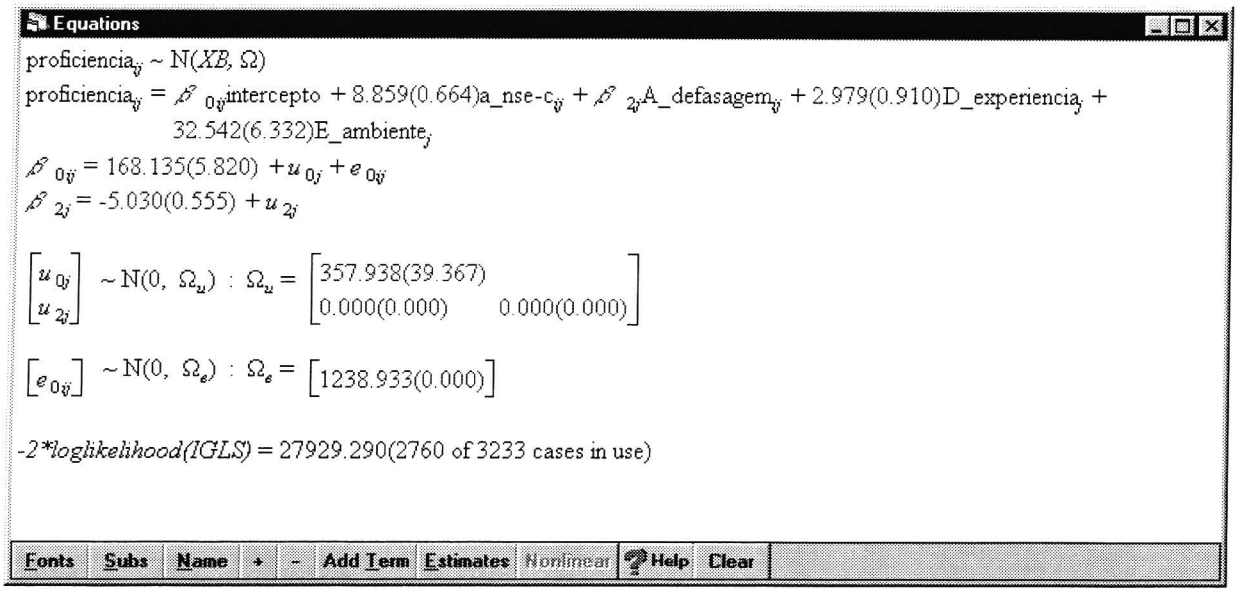
intercepto	a_nse-c	A_defasagem	D_experiencia	E_ambiente
β_0	β_1	β_2	β_3	β_4
168.115	8.852	-5.029	2.980	32.561
(5.824)	(0.664)	(0.555)	(0.911)	(6.336)
~~168.092~~	~~8.843~~	~~-5.028~~	~~2.982~~	~~32.584~~

Conforme pode ser visto, apenas as estimativas dos parâmetros fixos do modelo são apresentadas. O conteúdo desta tabela pode, entretanto, ser modificado. Assim, os botões + e – à esquerda permitem a apresentação simultânea de diversas tabelas, e as opções S,E,S,P,C,N, sinalizam, respectivamente, que o nome do parâmetro deve aparecer, a estimativa corrente, o erro padrão, a estimativa prévia, a correlação, e o numero de iterações necessárias para a convergência ser atingida.

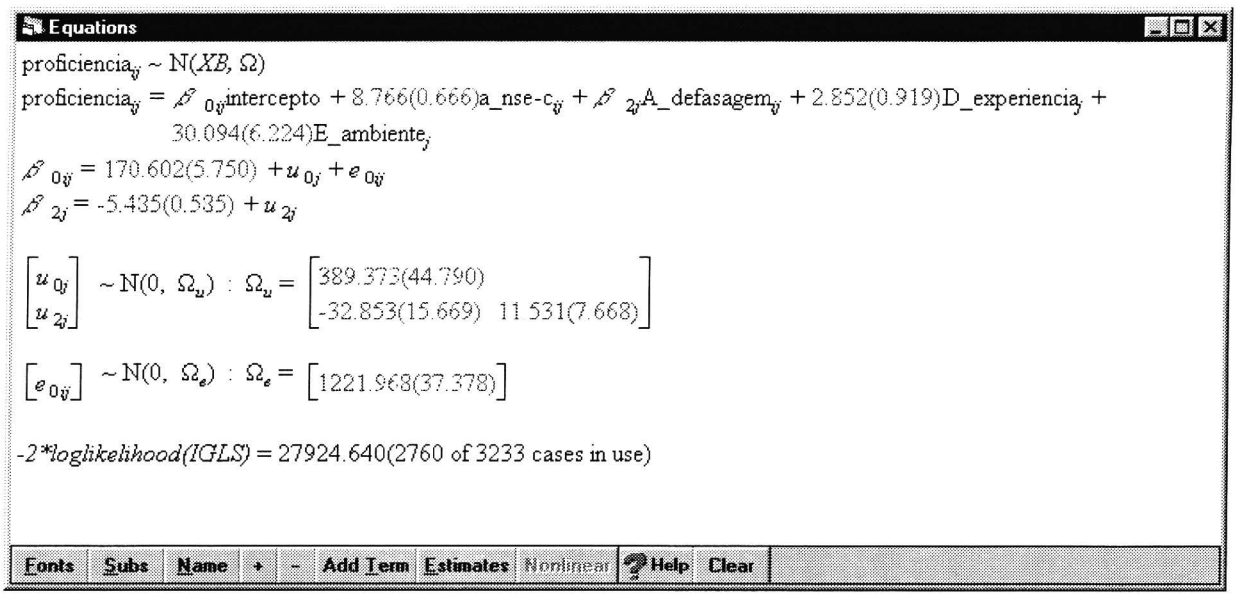
Até agora vimos como especificar e estimar um modelo multinível relativamente simples. Daqui para a frente veremos como lidar com modelos mais complexos, concretamente aqueles que envolvem coeficientes aleatórios, estudaremos como proceder a análise de resíduos e usar o modelo para fins de predição.

4.7 Modelo de Coeficientes Aleatórios

O modelo de componentes de variância ajustado assume que a relação entre a variável defasagem do aluno (a_defasagem) e a proficiência do mesmo independe da escola na qual o aluno estuda. Na análise multinível é comum encontrarmos variáveis explicativas com efeitos diferenciados por escola, isto é, cada escola tem seu próprio coeficiente de inclinação. Neste caso, denominamos o coeficiente como sendo aleatório ao nível da escola. Para especificar um modelo como este, precisamos definir que o coeficiente associado com a variável "a_defasagem" é aleatório. Para isto, clique em "a_defasagem" e assinale no retângulo "j(cod_esc)". Pode-se ver que o coeficiente β2 aparece apenas com o índice j (β2j) indicando que este coeficiente varia entre as unidades de nível 2, isto é, escolas. Assim, o coeficiente é formado por uma componente fixa ou valor médio entre as escolas (β2), e pela componente aleatória, que tem média zero e variância σ^2_{u2}. O termo σ_{u02} é a covariância de u_{2j} com o termo aleatório associado ao intercepto, u_{0j}. O modelo descrito está especificado abaixo. Note que os parâmetros aparecem em azul indicando que os mesmos precisam ser reestimados.



Ao pressionar em **MORE** e deixar que o procedimento de estimação atinja a convergência, nós obtemos as seguintes estimativas.



Os parâmetros do modelo anterior não se alteraram significativamente. As variâncias do primeiro e segundo nível sofreram uma pequena redução. O teste da razão de verossimilhança pode ser utilizado para testar se os parâmetros aleatórios incluídos no modelo são estatisticamente significativos.

A estatística do teste da razão de verossimilhança (a diferença entre os valores estimados para – ***2*loglikelihood***) é igual a 4,65 que é menor do que o valor tabelado de distribuição qui-quadrado com 2 graus de liberdade, ao nível de significância de 5%. Isto significa que os dois parâmetros aleatórios

incluídos no modelo não são estatisticamente significativos, isto é, que o efeito da variável "a_defasagem" não é diferenciado por escola. Ressalta-se aqui que os valores das estimativas de – ***2*loglikelihood*** são armazenados seqüencialmente na coluna 91 (***c91***).

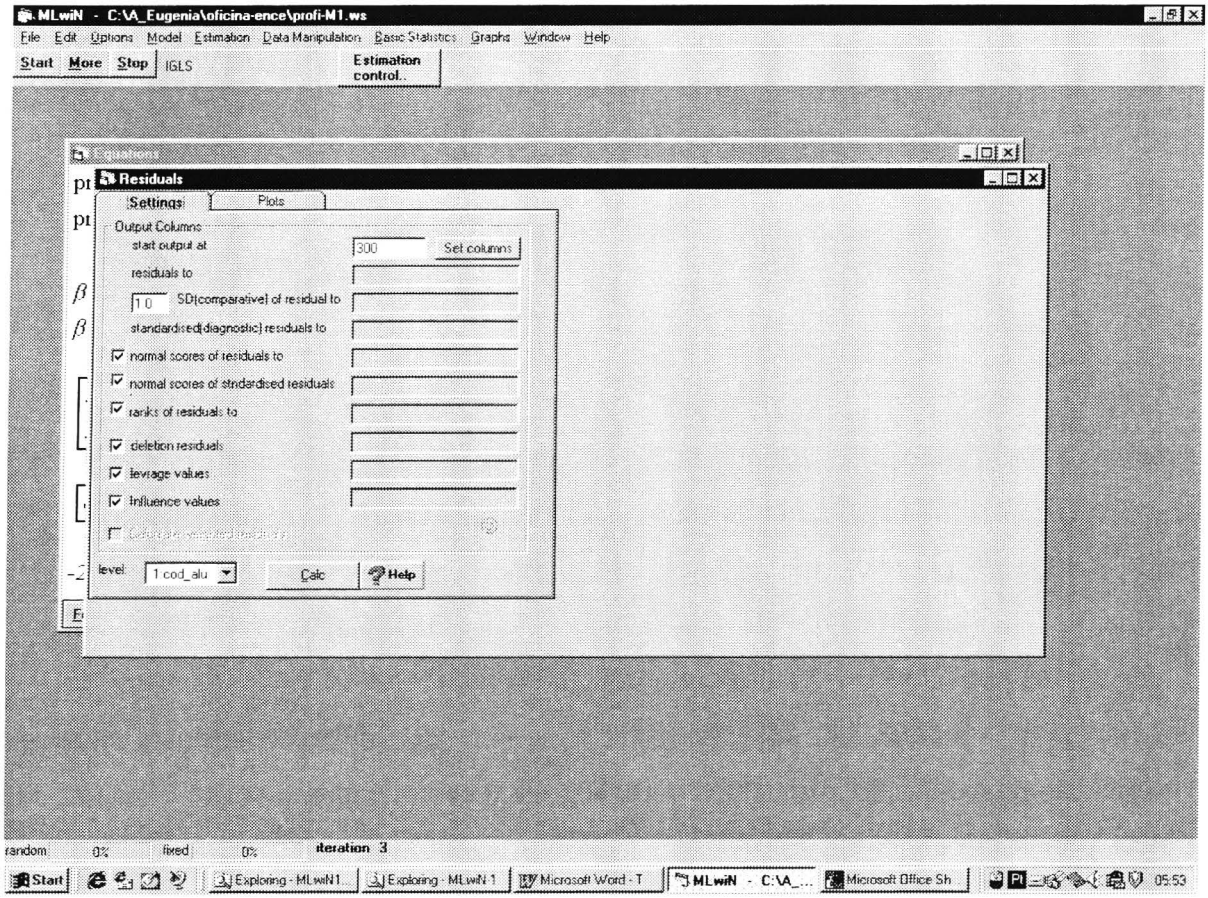
Como os novos parâmetros são estatisticamente significativos ao nível de 10%, decidimos não excluí-los do modelo. Sendo assim, podemos calcular o coeficiente de correlação entre o intercepto e coeficiente relativo a variável "***a_defasagem***". Esta estatística é dada por:

$$r(\mu_{0j}, \mu_{2j}) = \frac{-32{,}853}{\sqrt{389{,}373 \times 11{,}531}} = -0{,}49.$$

Este coeficiente indica que a correlação entre o intercepto e o coeficiente associado com a variável "***a_defasagem***" é negativo.

4.8 Análise de Resíduos

Na seção anterior introduzimos o termo *resíduos* para representar os efeitos aleatórios no modelo. Estudaremos em seguida como obter as estimativas destas quantidades aleatórias desconhecidas. Podemos pensar nestas quantidades como os valores preditos, dado o valor observado para a variável resposta e o seu valor estimado a partir do modelo. O *MlwiN* permite-nos fazer isto para qualquer nível do modelo, fornecendo igualmente os erros padrão das estimativas. Para fazer análise de resíduos precisamos usar **Model** e **Residuals**.

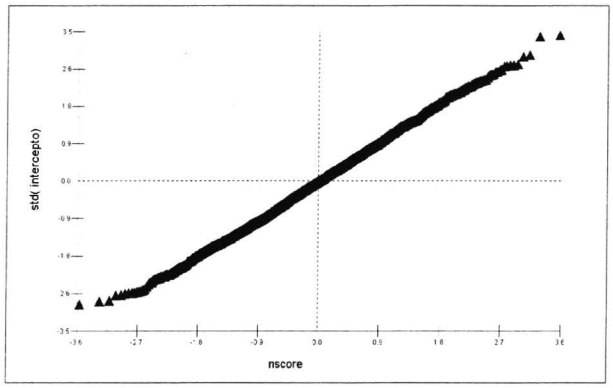


A tela aparece ativa em **Settings** e a função dela é especificar o nível para o qual os resíduos vão ser calculados e determinar a função para o fazer. Note que o nível 1 está definido na célula do canto inferior esquerdo da tela, mas o utente poderá alterar para qualquer nível. Vamos primeiro calcular os resíduos padronizados para o nível 1.

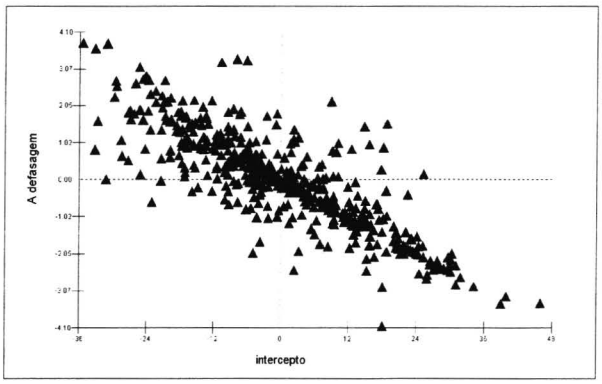
À esquerda da janela poderá observar que algumas opções já aparecem definidas. O numero das colunas mencionado em **set columns** indica a posição da base de dados onde o MlwiN irá colocar os resíduos e respectivas funções calculadas. No caso em estudo, as colunas c300 até c308 serão ocupadas.

Pressionando **calc** os resíduos serão calculados.

Pressionando em **Plots** e poderá verificar que é possível traçar diversos tipos de gráficos. A escolha de qual será traçado depende se pretende a análise gráfica para fins de diagnóstico ou comparação. Poderá fazer um de cada vez. De seguida faremos o gráfico dos resíduos padronizados contra os respectivos valores da distribuição Normal. Para tal, selecione **standardised residual x normal scores** e clique em **apply.** Obterá um gráfico semelhante ao que se apresenta de seguida, indicando que o pressuposto de Normalidade se verifica.



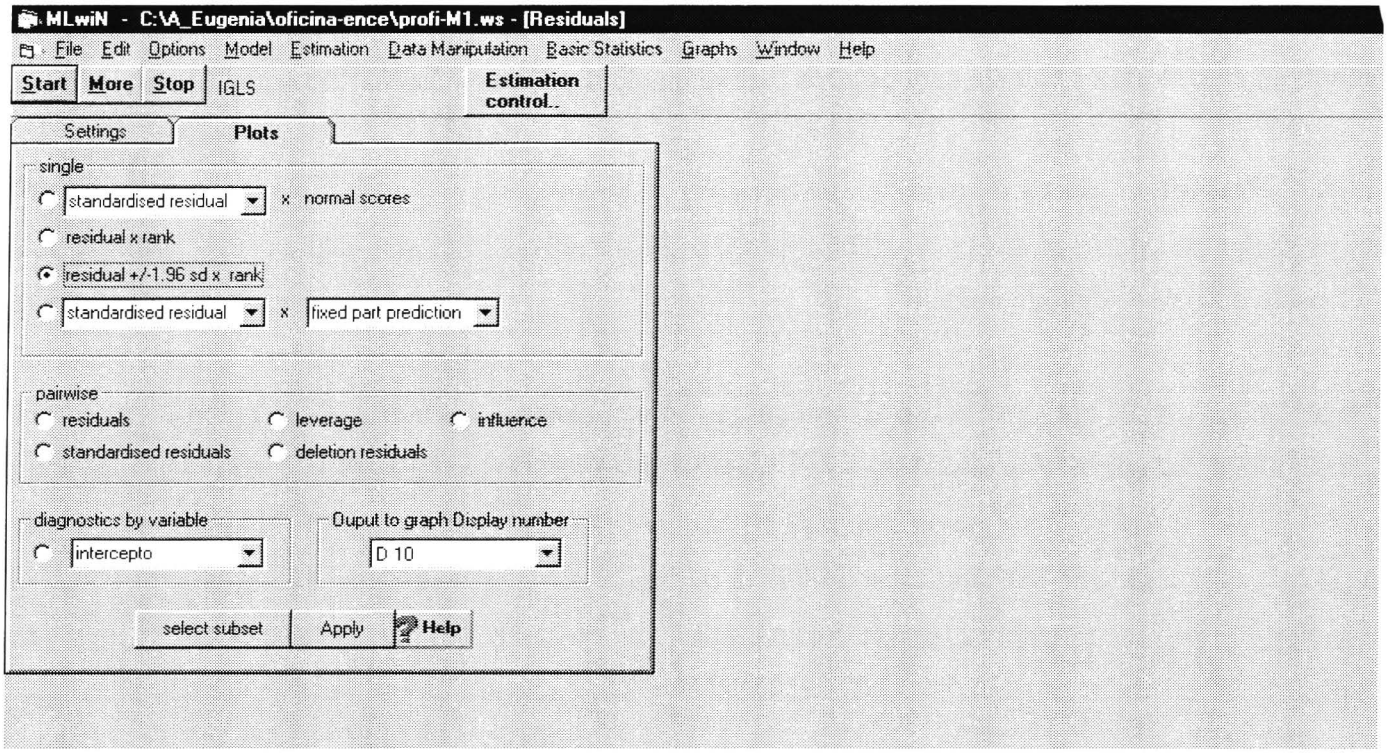
Poderá experimentar as outras opções que estão à disposição, incluindo as mencionadas por 'pairwise' que permitem traçar o gráfico do intercepto e resíduos LRT ou os resíduos padronizados para o nível 2. Selecione **Settings** na janela referente aos resíduos, e selecione o nível 2, **cod_esc** , no canto inferior esquerdo da tela e na célula em frente de **start output at** coloque 310, para que o MlwiN coloque os resíduos e funções daí para a frente. Clique em **set columns.** Coloque 1.96 no multiplicador de **SD(comparative)**. Clique em **Calc**. Já que existem dois conjuntos de resíduos – um para intercepto e outro para o coeficiente de inclinação, poderemos traçar o gráfico de um contra o outro. Selecione **Plots, residuals** na parte **pairwise** e clique em **Apply**.



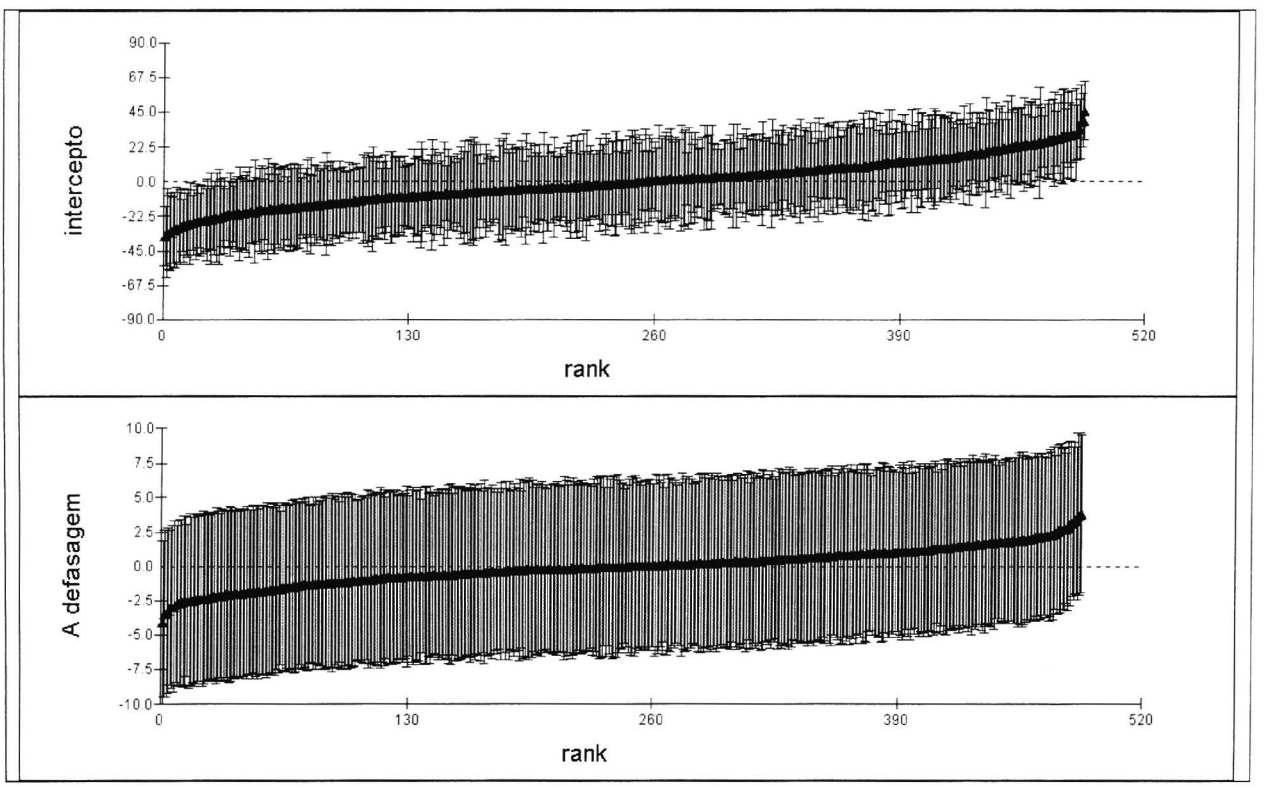
Isto reflete a correlação entre o coeficiente de inclinação e o intercepto. No entanto, chama-se a atenção de que estes resíduos estimados são os designados por resíduos 'shrunken'[12] e por isso eles apresentam menor variância do que os resíduos verdadeiros, por isto, a correlação calculada para os resíduos estimados, tem valor mais elevado.

Vamos agora ilustrar como usar os resíduos comparativos do nível 2 para comparar as diferenças entre as escolas. Na janela **Residuals** selecione **residual +/- 1.96 sd x rank** .

[12] Ver em Goldestein, H. (1995). Multilevel Statistical Models. London, Edward Arnold; New York, Halstead Press .



Pressione **Apply** e obterá gráficos semelhantes aos que se seguem:



Cada ponto estimado está cercado pelo respectivo intervalo de confiança. Assim, poderemos afirmar que duas escolas têm resíduos significativamente diferentes (ao nível de 5%) se e só se as suas barras de erro

não forem sobrepostas. A comparação de instituições, nem sempre é fácil pois as estimativas têm normalmente elevados valores para os erros padrão que lhes estão associados[13].

4.9 Predições a partir do modelo ajustado

Considere agora o seguinte modelo

$$\text{proficiencia}_{ij} \sim N(XB,\ \Omega)$$

$$\text{proficiencia}_{ij} = \beta_{0ij}\text{intercepto} + \beta_{1j}\text{a\_nse}_{ij} + 2.990(0.979)\text{D\_experiencia}_j + 37.340(6.095)\text{E\_ambiente}_j$$

$$\beta_{0ij} = 128.070(5.384) + u_{0j} + e_{0ij}$$

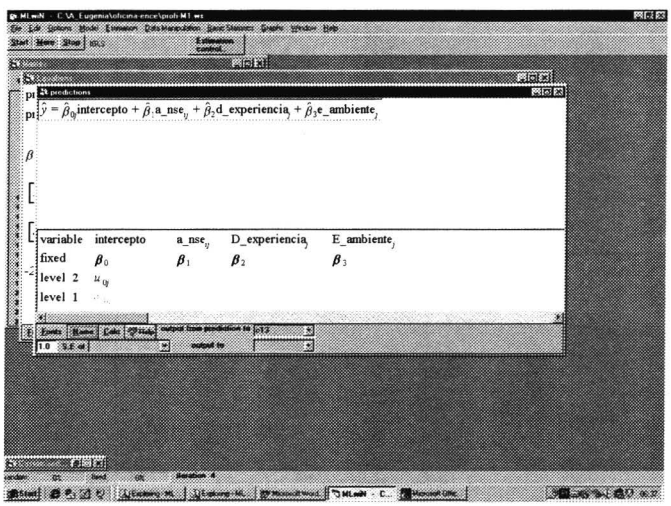
$$\beta_{1j} = 9.322(0.730) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0,\ \Omega_u) : \Omega_u = \begin{bmatrix} 339.346(159.991) & \\ -38.689(44.903) & 25.747(13.052) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0,\ \Omega_e) : \Omega_e = \begin{bmatrix} 1246.186(37.733) \end{bmatrix}$$

*-2*loglikelihood(IGLS)* = 28391.490(2799 of 3233 cases in use)

Se pressionar **Model** e depois **Predictions,** obterá a seguinte janela:



Temos cinco colunas de variáveis, sendo uma para cada variável explicativa com os respectivos coeficientes fixos e as componentes aleatórias consideradas no modelo. Bastará pressionar cada uma das delas para ser incorporada na equação de predição.

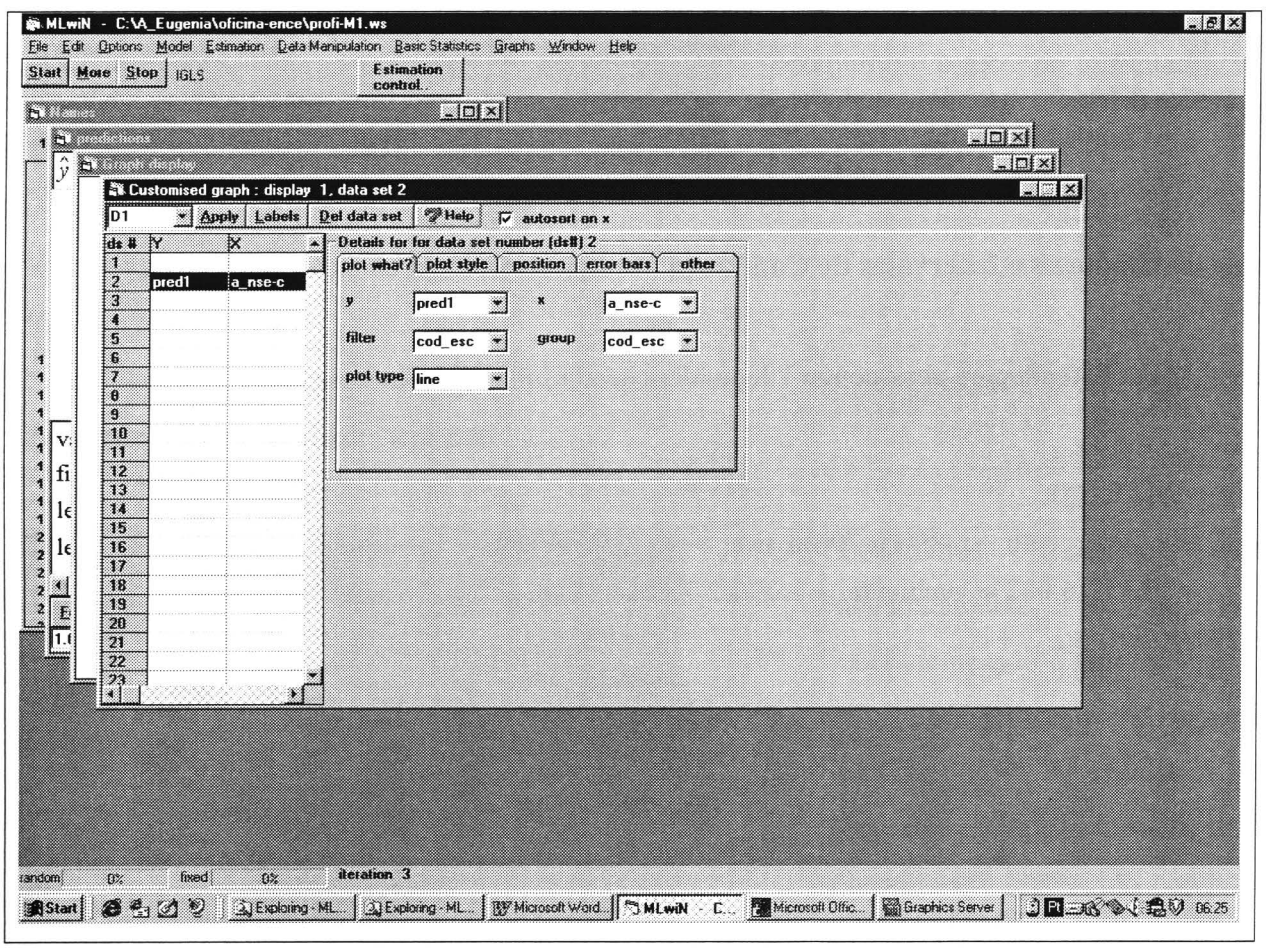
Suponha que desejamos fazer predições a partir da componente fixa do modelo juntando-lhe os resíduos associados quer ao intercepto, quer ao coeficiente de inclinação. Pressione em $\beta_0, \beta_1, \beta_2, \beta_3, u_{0j}$ e escolha a coluna C13, onde serão depositados os resultados, na célula de **output from prediction to.** A

[13] Para maiores detalhes ver em Goldestein, H. e Spiegelhater, D. J. (1996). League tables and their limitations: satistiscal issues in comparisons of institutional performance (with discussion), *Journal of the Royal Statistical Society* A 159: 385-444.

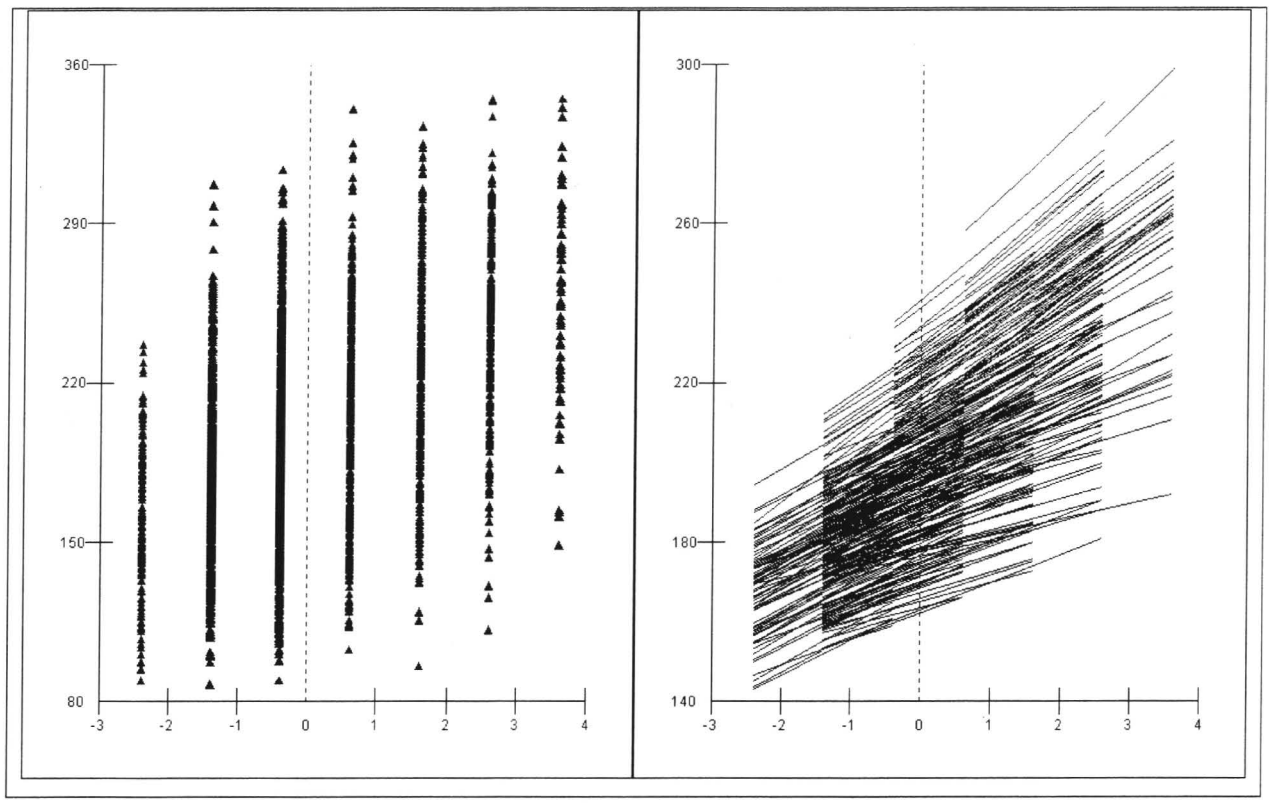
coluna C13 deverá ser designada como *PRED1* (usando **Names**). Para que o *MLwiN* proceda ao cálculo da predição pressione em **CALC**.

Façamos agora o gráfico das retas preditoras para cada escola contra o A_nse-c. Pressione em **Graphs**. Em **position,** estipule que dois gráficos devem ser traçados simultaneamente, assinalando a posição em que deverão ser mostrados.

Ainda para o conjunto de gráfico D1, selecione o novo conjunto de dados tornando ativa a linha 2 de **ds#** (data set #2) – para isso deve pressionar sobre a segunda linha. Verá que os 'settings' à direita ficarão vazios e poderá assim definir um novo gráfico. Associe PRED1 ao eixo de Y e STANDLRT ao eixo X. Defina o tipo de gráfico em **Plot what?** como gráfico de linhas, selecionando 'line' na célula **plot type**; na célula **group** selecione 'cod_esc' para que seja traçada uma reta preditora para cada escola. Depois de todas estas alterações o seu monitor deverá estar semelhante ao que seguidamente se apresenta.



Para que os gráficos sejam traçados de acordo com estas especificações, pressione **apply.** Aparecerá um gráfico como este:



As escalas foram definidas usando as opções de **user defined scale,** e os títulos dos gráficos atribuídos em **titles**.

4.10 Teste de Hipóteses e Intervalos de Confiança

Usar **Model** e **Main Effects** para acrescentar "dependência administrativa" ao modelo.

$$\text{proficiencia}_{ij} \sim N(XB,\ \Omega)$$

$$\text{proficiencia}_{ij} = \beta_{0ij}\text{intercepto} + 6.890(0.706)\text{a\_nse}_{ij} + -0.725(0.907)\text{D\_experiencia}_{j} + 20.472(6.083)\text{e\_ambiente}_{j} + 1.096(2.721)\text{estadual}_{j} + 33.675(3.229)\text{particular}_{j}$$

$$\beta_{0ij} = 146.227(5.889) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} 277.334(33.863) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} 1252.758(36.526) \end{bmatrix}$$

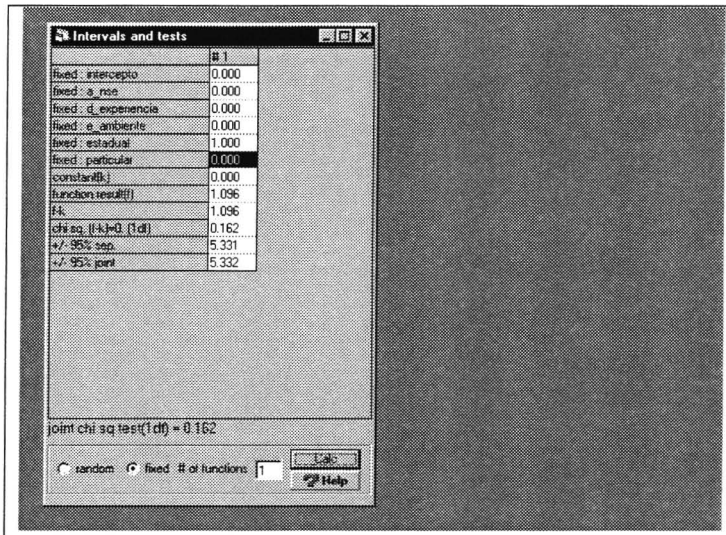
$$-2*loglikelihood(IGLS) = 28282.390(2799 \text{ of } 3233 \text{ cases in use})$$

Abra a janela **Model** e **Intervals and tests** e selecione **Fixed** na barra de ferramentas no fundo da janela.

Vamos testar a hipótese nula

$H_0: \beta_4=0$

β_5 é o coeficiente associado a variável 'estadual' . Coloque 1 na célula frente a 'estadual' e 0 em todas as restantes. A célula designada por **constant(k)** deve ter o valor que esta do lado direito de $\beta_5=0$, no caso 0. Pressione agora **calc**.

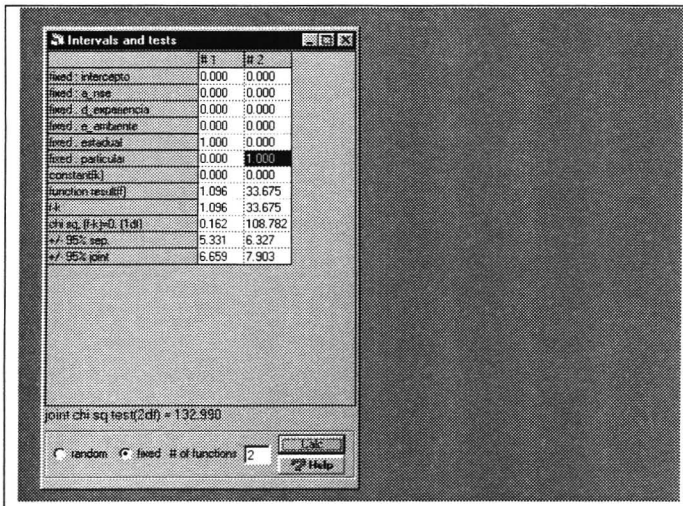


MLwiN calculou a diferença, **f-k,** entre o valor estimado e o valor da hipótese, calculou o respectivo valor de teste qui-quadrado, e os metade da amplitude dos intervalos de confiança de 95%. De notar que estes intervalos incluem o zero, deixando a suspeita de que o coeficiente pode ser estatisticamente não significante.
O intervalo de confiança de 95% conjunto e o teste do qui-quadrado conjunto tem os mesmos valores dos separados, tal como era de esperar.

Vamos agora fazer o teste de hipóteses conjunto:

$H_0: \beta_5=\beta_6=0$

Na célula **# of functions** coloque 2; na segunda coluna, na linha respectiva a 'particular' coloque 1. Pressione **calc**. O resultado será este:



Poderemos verificar que, separadamente, o primeiro coeficiente não é significante ao nível de 5%, mas que o segundo coeficiente é. A estatística de teste de qui-quadrado conjunto (usualmente conhecido por teste de Wald) é significante ao nível de 5%.