Modelo Linear Hierárquico
Estatística II (22018)

Ricardo Neves Pires, 2001645
2001645@estudante.uab.pt

Resumo: Faz-se uma exposição das vantagens e desvantagens do uso de Modelos Lineares Hierarquicos ou modelos multinível. Em que contextos e estruturas estes devem ser usados e quando dar preferência por estes relativamente a outros modelos. Passa-se a uma exposição matemática do modelo, transitando do modelo de regressão linear, a uma regressão mutlinível (do modelo basal a 2 níveis), o caso mais simples em Modelos Lineares Hierárquico. Dá-se enfâse à extração de informação possível, graças a um gênero de modelo e a versatilidade com que este se adapta às proprias caracteristicas e estruturas dos dados. Demonstra-se as interações a vários níveis, sejam dentro do conglomerados (níveis) ou entre eles, que sucede neste tipo de modelos. É feita uma passagem breve por dois métodos de estimação e pelos diversos passos a seguir no momento de escolha do modelo dentro do espectro hierárquico, com o suporte de critérios de informação que serão sucintamente expostos. Utilizam-se os dados que constam em Hox (2018) como forma de ilustração e exemplificação.
Palavras-chave: Modelo Hierárquico, Regressão Multinível, Correlação Intraclass, Métodos de Estimação, Critérios de informação.

1. Introdução

Nas últimas décadas os modelos lineares hierárquicos têm ganho reconhecimento pela sua abrangente e eficiente aplicação em diversos campo do conhecimento. Com emergência nas ciencias sociais e humanas foram sendo adaptados noutras áreas, nomeadamente as ciencias da saúde, demografia, gestão organizacional e entre outras. A deconsideração da estrutura hierárquica dos dados conduz a uma leitura demasiado parcial e truncada da informação disponível. O que traduz e proporciona uma visão distorcida e fragmentada do que se possa pretender obter. A omissão das estruturas inerentes aos dados pode levar a uma resposta pouco clara e abrangente no momento de tomada de decisões ou conclusões. Na sua essência um Modelo Linear Hierárquico (MLH) permite captar nos dados, informações que em outros modelos seria difícil de conseguir. A regressão multinível tenta colmatar problemas e pressupostos existentes em modelos lineares. O termo multinível refere-se a uma estrutura de dados hierárquica ou aninhada (Marôco, 2021), geralmente sujeitos dentro de grupos. Contudo, o conceito aninhado pode também referir-se a medidas repetidas dentro de sujeitos. A análise multinível é usada para examinar as relações entre variáveis medidas em diferentes niveis da estrutura dos dados. Goldstein (2011), Hox (2018) e Gelman (2007) contextualizam e exemplificam de forma sublime quais os cuidados a ter durante de todo o processo de modelação e escolha do modelo. Mesmo trazendo vantagens em várias frentes, o nível de complexidade e intrepretação pode tornar-se desafiante. O poder computacional alcançado em recentes anos permitiu que fosse possivel modelar e analisar de forma flexivel e ágil, dados complexos e com estruturas multinivel. Os software´s com relevo e especificamente desenvolvidos para o efeito são o HLM e o MLwiN (Rasbash, 2014). Os software´s ditos convencionais, como o STAT, SPSS e SAS, tem módulos incorporados para a análise e processamento de MLH. A linguagem R, permite inclusivamente modelar e analisar MLH com o uso de bibliotecas dedicadas especificamente para o efeito, das quais por exemplo, lme4, nlme e lmer. A regressão multinivel permite analisar as interações ou interconecções entre o micro, niveis inferiores (e.g., nivel 1) e macro, niveis superiores (e.g., nivel 2), como as que ocorrem intra niveis. A regressão pelo método dos mínimo quadrados tem como pressuposto que os erros são independentes e em determinados dados este pressuposto é violado, o que leva à sub

estimação dos erros padrão. Os MLH foram desenvolvidos para tomar em consideração o agrupamento ou conglomeração e estimar a regressão sem inflacionar a possibilidade de erros do tipo I e o não cumprimento do pressuposto de independência dos erros. A vantagem que estes modelos trazem são as interações entre variaveis de diferentes niveis, sendo possivel compreender melhor a decomposição da variância dos erros nos diversos niveis. Torna-se assim possivel uma análise detalhada da correlação entre variáveis, dissemelhanças entre niveis, comportamento da variância, da subsequente decomposição da mesma e de efeitos cruzados. As estimativas dos coeficientes da regressão e dos valores como erros padrão e intervalos de confiança tendem a proporcionar estatisticas de qualidade. A existencia de correlação intraclasse é uma indicação sobre a proporção de variância existente no nivel 2. Os modelos multinível são versáteis pelo fato de permitirem entender a variabilidade relativa entre niveis, sujeitos ou grupos e entender onde se encontra alojada a componente estocástica.

2. Métodos

As ideias centrais que fundamentam os MLH já abordadas são agora especificadas em termos matemáticos. Iremos focar na diferença entre efeitos aleatórios e fixos. Discutir-se-á os fundamentos da estimação de parâmetros com foco nos dois métodos comumente utilizados, a máxima verossimilhança e a máxima verossimilhança restrita. Inicia-se com uma breve revisão da regressão linear sendo o ponto de partida para a transição para MLH.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (1)$$

Temos a variável dependente Y expressa em função da variavel independente X com os devidos coeficientes β_0 e β_1 e o termo de erro aleatório de sujeito para sujeito. Com isto é de notar que com este modelo existe uma ordenada, β_0 comum a todas as observações ou sujeitos perante a população de interesse em estudo. Porém, caso as observações estejam conglomeradas (em cluster) ou em estruturar de níveis, existe potencialmente uma ordenada para cada conglomerado ou grupo, isto caso não exista um efeito conglomerado o modelo (1) dentro dos pressupostos e as necessidades, poderá ser o adequado. O modelo mais simple dos MLH é o modelo residual. Este modelo não contém variáveis explicativas, assume que a variavel dependente é estimada por um valor médio que varia por grupo e por um erro aleatório de cada sujeito em cada grupo. Isto é, os efeitos fixos são as ordenadas β_0 e os efeitos aleatórios são as variâncias/covariâncias, ϵ_{ij}.

$$Y_{ij} = \beta_{0j} + \epsilon_{ij} \quad (2)$$

ij - representam o i-ésimo sujeito no j-ésimo grupo, com $i = 1, \dots, n_j$ para o nível micro (sujeito) e $j = 1, \dots, k$ para o nível macro (grupo). O cerne deste modelo é a da decomposição da variância de Y_{ij} numa componente ao nível micro e macro, com intuito de auferir a variância entre grupos e sujeitos relativamente à variância total. A informação retirada deste modelo permite entender a estrutura dos dados, ter uma base para acrescento de complexidade caso assim seja necessário, comparação entre modelos (de nível 1 ou 2) e fornece estimativas para $\sigma^2_{u_{0j}}$, variância entre grupos e $\sigma^2_{e_{ij}}$, variância dos erros dos sujeitos, que por sua vez são úteis para estimar o coeficeinte de corrrelação intracalsses (ICC). Este pretende medir o grau de agrupamento dentro dos grupos.

$$ICC = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}} + \sigma^2_{e_{ij}}} \quad (3)$$

Com a possibilidade da ordenada diferir entre grupos, introduz-se aleatoriedade que é expressa do seguinte modo:

$$\beta_{0j} = \gamma_{00} + u_{0j} \text{ (4)}$$

em que a grande média γ_{00}(a média da média dos grupos) como efeito fixo, mais um efeito variável/aleatório entre grupos, u_{0j}(desvio da grande média). Inserindo a equação β_{0j} em Y_{ij}, obtém-se o modelo residual completamente expresso, com a parte fixa e a parte aleatória totalmente descriminada.

$$Y_{ij} = \gamma_{00} + u_{0j} + \epsilon_{ij} \text{ (5)}$$

Após a validação através do modelo residual da relevância em prosseguir com um MLH, inicia-se com a inclusão de preditores com a possibilidade de vir a explicar a variância da observações dos respetivos niveis. O modelo residual é assim facilmente extendido. Caso se considere incluir uma variável explicativa X_{ij} ao nível micro (nivel 1), têm-se:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + u_{0j} + \epsilon_{ij} \text{ (6)}$$

em que X_{ij} a variavel explicativa e ϵ_{ij} são os erros ou residuos do modelo. Este modelo pode ser expresso em dois niveis separados. Obtém-se assim as seguintes esquações:

Ao nivel 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X + \epsilon_{ij} \text{ (7)}$$

Ao nível 2:

$$\beta_{0j} = \gamma_{00} + u_{0j} \text{ (8)}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \text{ (9)}$$

em que β_{0j} representa a média da variavel Y no grupo j, β_{1j}representa a média da variavel Y no j-ésimo grupo. Ao inserir em (7) as equações (8) e (9) obtém-se:

$$Y_{ij} = \gamma_{00} + u_{0j} + (\gamma_{10} + u_{1j})X_{ij} + \epsilon_{ij} \text{ (10)}$$

Reescrevendo o modelo (10), tem-se:

$$Y_{ij} = \gamma_{00} + u_{0j} + \gamma_{10}X_{ij} + u_{1j}X_{ij} + \epsilon_{ij} \text{ (11)}$$

Temos assim a componente fixa do modelo, $\gamma_{00} + \gamma_{10}X_{ij}$ e a componente aleatória $u_{0j} + u_{1j}X_{ij} + \epsilon_{ij}$. Pelo modelo, pode-se afirmar que existe uma interação entre grupos e X, de tal modo que a relação de X e Y não é constante entre grupos. O erro total do modelo e a variancia dependem também dos valores da variavel indpendente/explicativa. Desta feita mais nos aproximamos do contexto do problema e melhor o conseguimos explicar (Hox, 2018). Focamo-nos agora na parte da estimação, que é a técnica tendo em consideração valores observadores numa amostra e produzir um valor que é o melhor valor para um parâmetro do qual não se conhece, de determinada população.

Os dois métodos de estimação usuais em modelos multiníveis são a Máxima Verosimilhança (ML) e a Máxima Verosimilhança Restrita (REML). Ambas produzem resultados semelhantes perante amostras grandes e número de grupos elevados. Os métodos nesses casos são assintoticamente equivalentes. Porém, com amostras pequenas ou número de grupos reduzidos o REML produz boas estimativas para efeitos fixos e aleatório. O método ML pode representar uma vantagem em relação a estimadores baseados na minimização da soma dos quadrados dos residuos. A especificação deste método reside na distribuição condicional de uma variavel endógena, y, dado um conjunto de variáveis explicativas. Conhecida a densidade ou probabilidade condicional de y dado x, em que o vetor θ são os parâmetros desconhecidos, o método ML atribui a θ o valor que máximiza a função densidade conjunta. O método tem como

propriedade originar um estimador assintoticamente não enviesado. Sucede que existe enviesamento neste método no contexto multinivel para os parâmetros aleatórios. Devido à perda de graus de liberdade da estimação que resulta da estimação dos parâmetros fixos e que o método ML não considera. Por forma a colmatar o problema, o REML tem em conta o ajuste do número de graus de liberdade e máximiza separadamente as funções de verosimilhança (L) dos efeitos fixos e aleatórios.

Segundo Hox (2018) a estratégia mais apropriada para os modelos multinível é *step-up.* Como o nome indica, acrescenta-se complexidade ao modelo passo a passo. Inicia-se com um modelo simples, como o modelo residual (referência) e progride-se para modelos mais complexos até ao momento em que o ajuste não seja significativamente melhor que o modelo anterior. A forma de testar o ajustamento de um modelo e comparar a qualidade de um modelo para o outra é através de critérios de informação do modelo. São considerados dois critérios, AIC (Akaike, 1974) e BIC (Schwarz, 1978), estes tendem a ser os mais utilizados. O critério de Akaike ou AIC, desenvolvido por Hirotugu Akaike é definido do seguinte modo:

$$AIC = -2 \times log(L_{MV}) + 2p \quad (12)$$

em que p é o número de parâmetros estimados no modelo. Este critério penaliza modelos com número de parâmetros elevados .

O critério Bayesiano ou BIC desenvolvido por Gideon Schwarz é definido da seguinte forma:

$$BIC = -2 \times log(L_{MV}) + p \times log(N) \quad (13)$$

em que p tem o mesmo significado que no critério anterior e N é a dimesnão da amostra. Ambos os critérios são usados simultaneamento, pois nenhum deles é superior ao outro. O melhor modelo de entre os comparados, é o que apresenta menor valor do critério de informação, sejam ele AIC ou BIC.

3. Aplicação

Usa-se como exemplo de aplicação uma base de dados de dois mil alunos de cem escolas. O objetivo é ofecer um exemplo de análise de regressão multinivel. Temos como variável dependente, a popularidade do aluno (Y_{ij}), com classficação numa escala de 1 a 10. Ter em conta que os dados foram recolhidos através de um procedimento sociométrico (Bonito, 2018). Por norma estes procedimentos solicitam ao inquerido, nesta caso os alunos de uma turma (grupo) que avaliem todos os outros alunos. Seguidamente é atribuida uma classificação média de popularidade para cada aluno. Neste procedimento o efeito de grupo aparenta ter um variância do nivel superior forte. Uma segunda variável de relevo entra na equação, a popularidade do aluno avaliada pelo seu professor em que a escala é de 1 a 10. As variaveis explicativas são o sexo do aluno (X_{1ij}), se o aluno é extrovertido (X_{2ij}) e a experiencia do professor em anos (Z_j). O software SPSS é usado como ferramenta de análise e os outputs constam em anexo. Iniciamos o construção do modelo do seguinte modo:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \epsilon_{ij} \quad (14)$$

O modelo assume que cada grupo tem uma ordenada e declives diferentes ao contrário do que sucede com a regressão linear, tal como foi demonstrado acima no texto. Uma turma (grupo, o nosso nivel 2) com valores de ordenadas mais elevados prevê que os alunos sejam mais populares do que numa turma com ordenadas baixas. A variação na ordenada alteram o valor médio para toda a turma (sejam do sexo feminino ou masculino). Quanto aos declives das retas dos coeficientes para sexo e extrovertido, o modelo sugere que a relação entre esses preditores e a popularidade do aluno não são os mesmo em todas as turmas. Turmas com coeficentes em que os declives sejam mais acentuados, por exemplo o sexo, as diferenças entre feminino e masculino são grandes. O contrário tambem é possivel. Coeficientes para sexo (β_{1j}) com

declives pouco acentuados em determinadas turmas sugere que o gênero do aluno tem pouco efeito na popularidade e/ou que a diferença entre feminino e masculino é pequena. A intrepretação é similar para o coeficiente extrovertido. A variação nas inclinações dos declives ou coeficientes tem impacto na diferença entre as ordenadas. O modelo pretende explicar a variação dos coeficientes da regressão e para isso necessita-se introduzir variáveis explicativas ao nível do grupo.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (15)$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (16)$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_j + u_{2j} \quad (17)$$

em que Z_jé a variavel explicativa experiência do professor. Para a equação (15), a média da popularidade é dada pela experiência do professor isto é, quando γ_{01} é negativo a média da popularidade é mais baixa em turmas com professores mais experientes. As restantes equações tem intrepretação e significado semelhantes sobre a popularidade relativamente à dependencia da experiencia do professor seja para o sexo (16) ou extrovertido (17). A variável experiência do professor tem inlfuência na relação entre popularidade e sexo ou extrovertido. Os erros u são ao nível de turma (nível superior) e ϵ_{ij} é o termo de erro no modelo ao nível dos alunos.
Ao substituir as equações (15), (16) e (17) em (14) temos,

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j + u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + \epsilon_{ij}$$

a componente fixa do modelo é, $\gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j$ e a componente aleatória é, $u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + \epsilon_{ij}$. Os termos $X_{1ij}Z_j$ e $X_{2ij}Z_j$ são interações que derivam da influencia da variação de β_j, ao nível dos alunos com o nivel de grupos da variavel Z_j, dito de efeito cruzado. Os coeficientes da regressão para as três variáveis explicativas são significativas, ao nivel de significância de 5%. De notar que os efeitos de interações não são expostos e somente faz-se a intrepretação do modelo dos efeitos principais tal como os outputs em anexo sugerem. A variância da popularidade entre turmas é significativa ($\sigma^2_{u_{0j}} = 1.252; Z_w = 4.152; p - value < 0.001$) e o efetio de extrovertido sobre a popularidade varia significativamente entre turmas ($\sigma^2_{u_{1j}} = 0.339; Z_w = 4.028; p - value < 0.001$) Uma variancia signficativa do declive associado a extrovertido indica que este varia significativamente entre turmas, a covariância da relação entre extrovertido e popularidade e turma é fraca e significativa (Cov = -0.184; r = - 0.28 $Z_w = 4.152; p - value < 0.001$) inidcando que o efeito de extrovertido sobre a popularidade depende da turma. A conclusão sobre o modelo poderia diferir com alguma significancia com as interações as serem tomadas em conta. Não indiciando que o modelo sem interação esteja errado, o que realmente sucede é que talvez nem toda a informação da estrutura dos dados seja captada. A estrategia do modelo poderia-se prolongar com o nível de complexidade. Seria possivel retirar a aleatoriedade nos declives e manter interação entre variáveis. Todo um conjunto de cenários são possiveis, porém não devendo esquecer o peso dos critérios da qualidade de ajustamento do modelo aos dados.

4. Conclusão

O propósito deste documento é a introdução de alguns conceitos base dos Modelos Lineares Hierárquicos e fazer a sua exposição teórica com a transição para um exemplo prático. Tornar-se-ia dificil ser completamente exaustivo sobre o tema, pois a possibilidade do grau complexidade nestes modelos é ampla, tal como foi visto no caso prático em que os efeitos cruzados não foram exposto pela necessidade de ter uma base teorica do assunto em estudo forte. Mesmo assim o modelo na parte de aplicação já ganha algumas proporções de

complexidade sendo que o analista deve ter os devidos cuidados nas intrepretações. Caso os dados utilizados fossem modelados por uma regressão linear, muita da variância e informação não seria fragmentada da forma que foi e seria transposta toda ela para o termo de erro. Com isto, este texto pretende ser uma directriz bastante geral pelos passos essenciais a seguir no processo multinível. Inicia-se com o modelo residual como alicerce para a construção e verificação do uso de modelos subsequentes e mais complexos, caso assim o seja necessário. Os modelos multínivel são uma alternativa bastante atrativa e versátil em relação a modelos lineares comuns. Porém, a adequação e uso deste modelos deve sempre ter em consideração a estrutura dos dados.

5. Referências

Akaike, H. (1974). *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19 (6): 716–723, DOI: 10.1109/TAC.1974.1100705

Bonito, J. (2018). *Sociometria*. Évora, Portugal: Universidade de Évora. URL: http://hdl.handle.net/10174/26119

Gelman, A. (2006). *Multilevel (Hierarchical) Modeling: What It Can and Cannot Do*, Technometrics, 48:3, 432-435, DOI: https://doi.org/10.1198/004017005000000661

Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Goldstein, H. (2011). *Multilevel Statistical Models*. 4th Edition, John Wiley & Sons

Hox,J., Moerbeek,M., Schoot, R. (2018). *Multilevel Analysis: Techniques and Applications*. 3rd Edition, Routledge

Marôco, J. (2021). *Análise Estatística com o SPSS STATISTICS.* 8ª Edição, Pêro Pinheiro, Portugal: Report Number

Rasbash, J., Steele, F., Browne, W.J. and Goldstein, H. (2014). *A User's Guide to MLwiN v2.31*. Centre for Multilevel Modelling, University of Bristol. UK

Schwarz, Gideon E. (1978). *Estimating the dimension of a model*. *Annals of Statistics*. 6 (2): 461–464, DOI: 10.1214/aos/1176344136

Snijders, T.A.B., & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* 2nd Edition. London, UK: Sage.

ANEXO

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	extrav	1		1	
	sex	1		1	
	texp	1		1	
Random Effects	Intercept + extrav[b]	2	Unstructured	3	class
Residual				1	
Total		6		8	

a. Dependent Variable: popular.

b. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

Information Criteria[a]

-2 Log Likelihood	4812.8
Akaike's Information Criterion (AIC)	4828.8
Hurvich and Tsai's Criterion (AICC)	4828.9
Bozdogan's Criterion (CAIC)	4881.6
Schwarz's Bayesian Criterion (BIC)	4873.6

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: popular.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	185.12	14.285	.000
extrav	1	98.499	341.88	.000
sex	1	1915.7	1174.0	.000
texp	1	103.75	111.71	.000

a. Dependent Variable: popular.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	.73771	.19518	185.12	3.780	.000	.35264	1.1228
extrav	.45264	.02448	98.499	18.490	.000	.40406	.50122
sex	1.2525	.03655	1915.7	34.264	.000	1.1808	1.3242
texp	.09087	.00860	103.75	10.569	.000	.07382	.10792

a. Dependent Variable: popular.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		.55180	.01838	30.017	.000	.51692	.58904
Intercept + extrav [subject = class]	UN (1,1)	1.2805	.30842	4.152	.000	.79866	2.0531
	UN (2,1)	-.1847	.04779	-3.864	.000	-.2784	-.0910
	UN (2,2)	.03393	.00840	4.038	.000	.02088	.05512

a. Dependent Variable: popular.