



[CLIQUE AQUI PARA ACESSAR O MATERIAL DO CURSO](#)



Regressão multinível OU HIERÁRQUICA

Professora Naiara Sandi de Almeida Alcantara
E-mail: nayarasandy@ufpa.br

**ESCOLA AMAZÔNICA
DE FORMAÇÃO METODOLÓGICA**



➤ **Diferentes Denominações:**

- ✓ Modelos Lineares Multiníveis na Sociologia
 - ✓ Mixed-effects Models na Biometria
 - ✓ Random-coefficient Regression Models na Econometria
 - ✓ Covariance Componentes Models na Estatística
-
- ✓ Modelos Lineares Hierárquicos: adotado por Raudenbush e Brick por focalizar a forma de estruturação dos dados.
 - ✓ O termo foi criado por Lindley e Smith (1972) como parte de uma importante contribuição sobre estimação Bayesiana de modelos lineares.



Estimação de Componentes de Variância

- Dempster, Laird e Rubin (1977) desenvolveram um algoritmo que tornou possível a aplicação do conceito de Lindley e Smith à um conjunto mais ampliado de casos.
- Dempster (1981) demonstrou que a técnica de ECV poderia ser aplicada a estruturas de dados hierarquizadas.
- Em 1982 e 1983 os primeiros trabalhos com aplicações dessa técnica foram publicados.
- Em 1988 foram apresentadas as primeiras ferramentas para ajustes desses modelos:

Ferramentas para Ajustes de Modelos:

- **VARCL (1988)**: Uma das primeiras ferramentas desenvolvidas para ajustar modelos de ECV.
- **MIXOR (1996)**: Outra ferramenta que surgiu mais tarde, usada para modelos de ECV, especialmente em dados categóricos.
- **SAS Proc Mixer (1996)**: Procedimento no software SAS que permite ajustar modelos de ECV.
- **HLM (2000)**: Ferramenta usada para ajustar modelos hierárquicos lineares, que são uma forma de modelos de ECV.

Introdução a regressão multinível

 É uma técnica estatística para analisar dados estruturados hierarquicamente

Muitas pesquisas em Ciências Sociais envolvem dados estruturados de forma hierárquica

- ✓ Exemplo: Estudos envolvendo escolaridade
 - ✓ **Unidades de Análises:** Escolas, salas de aula e alunos
 - ✓ **Três níveis de análise:** Macro, intermediário e micro/ individual
 - ✓ **Dados Agrupados:** As unidades de análise que compõem o nível micro estão relacionadas diretamente aos demais níveis

Introdução a regressão multinível



Estruturas de Três Níveis: em pesquisas na área educacional são comuns estruturas de dados hierarquizadas em três níveis.

Ex. 1: Identificação dos efeitos de uma nova técnica didático-pedagógica **sobre o rendimento dos alunos ao longo do tempo.**

- ✓ Nível micro: dados sobre o rendimento dos alunos coletados trimestralmente durante três anos.
- ✓ Nível intermediário: as múltiplas observações dos rendimentos estão agrupadas no nível dos alunos, que passam a ser um nível hierarquicamente superior nesse caso.
- ✓ Nível superior: as escolas

• Problema na análise de estruturas hierárquicas

Apesar dessa recorrência em pesquisas sociais até recentemente os pesquisadores não produziam análises adequadas a essa estrutura hierarquizada de dados.

- Limitações das técnicas convencionais de estimação de modelos lineares com estruturas agrupadas, gerando nas Ciências Sociais problemas de viés de agregação, erros de precisão e confusões envolvendo as unidades de análises.
 - Empobrecimento: essa limitação tem conduzido a uma postura reducionista na formulação de hipóteses, que se concentram apenas em um nível de análise quando a realidade empírica exige claramente uma abordagem multinível.



Falácia ecológica

- **Testes precedentes da técnica**

- **Regressão e Análise de Variância:** a exposição dos fundamentos dessa técnica indica que se trata de algo derivado dos princípios presentes nas técnicas de regressão e de análise de variância.
- Revisão desses princípios (Regressão e ANOVA)

Essas técnicas são fundamentais em estatística e são usadas para investigar a relação entre variáveis.

A regressão geralmente analisa como uma ou mais variáveis independentes (preditoras) afetam uma variável dependente (resultado). ANOVA, por outro lado, é usada para comparar as médias de diferentes grupos e determinar se as diferenças observadas são estatisticamente significativas.

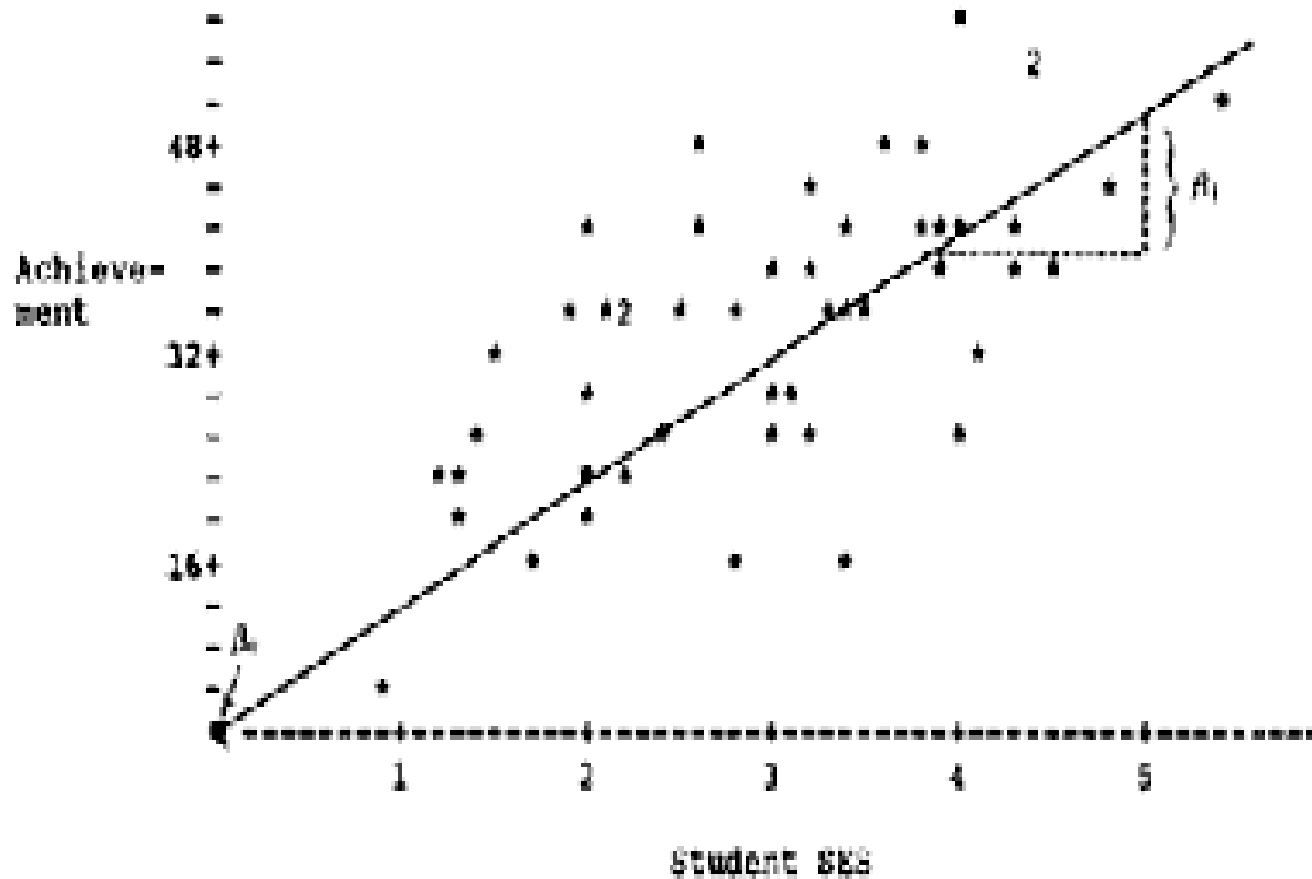
Testes precedentes da técnica

Exemplo de Estudo Educacional:

- Estudo Educacional: vamos seguir os passos propostos por Raudenbush e Brik (2002) - especialistas em modelos hierárquicos- e usar como exemplo um estudo sobre a relação entre Situação Socioeconômica (SS) e Nota em Matemática (NM).
- 1. Variável Dependente:** Nota de Matemática (NM) - Essa é a variável de interesse, cujo comportamento estamos tentando explicar.
 - 2. Único Preditor:** Situação Socioeconômica (SS) - Essa é a variável que se acredita influenciar a Nota de Matemática.
 - 3. Nível 1:** Ambas as variáveis (NM e SS) se originam no nível "aluno", o que significa que estão associadas a cada estudante individualmente (nível micro).
 - 4. Uma Escola:** O modelo inicialmente investiga a relação entre essas variáveis em uma única escola. Esse é um cenário simplificado que facilita a análise inicial antes de expandir para múltiplas escolas ou níveis hierárquicos adicionais.

Testes precedentes da técnica

Diagrama de Dispersão que descreve o relacionamento entre NM e SS



Nesse diagrama de dispersão podemos verificar esse relacionamento em uma única escola.

Intercepto e Inclinação: a linha reta representa a equação da regressão e parece representar bem o relacionamento.

$$Y_i = \beta_0 + \beta_1 X_i + r_i.$$

$$Y = NM$$

$$X = SS$$

β_0 = NM esperada para um aluno com 0 em SS.

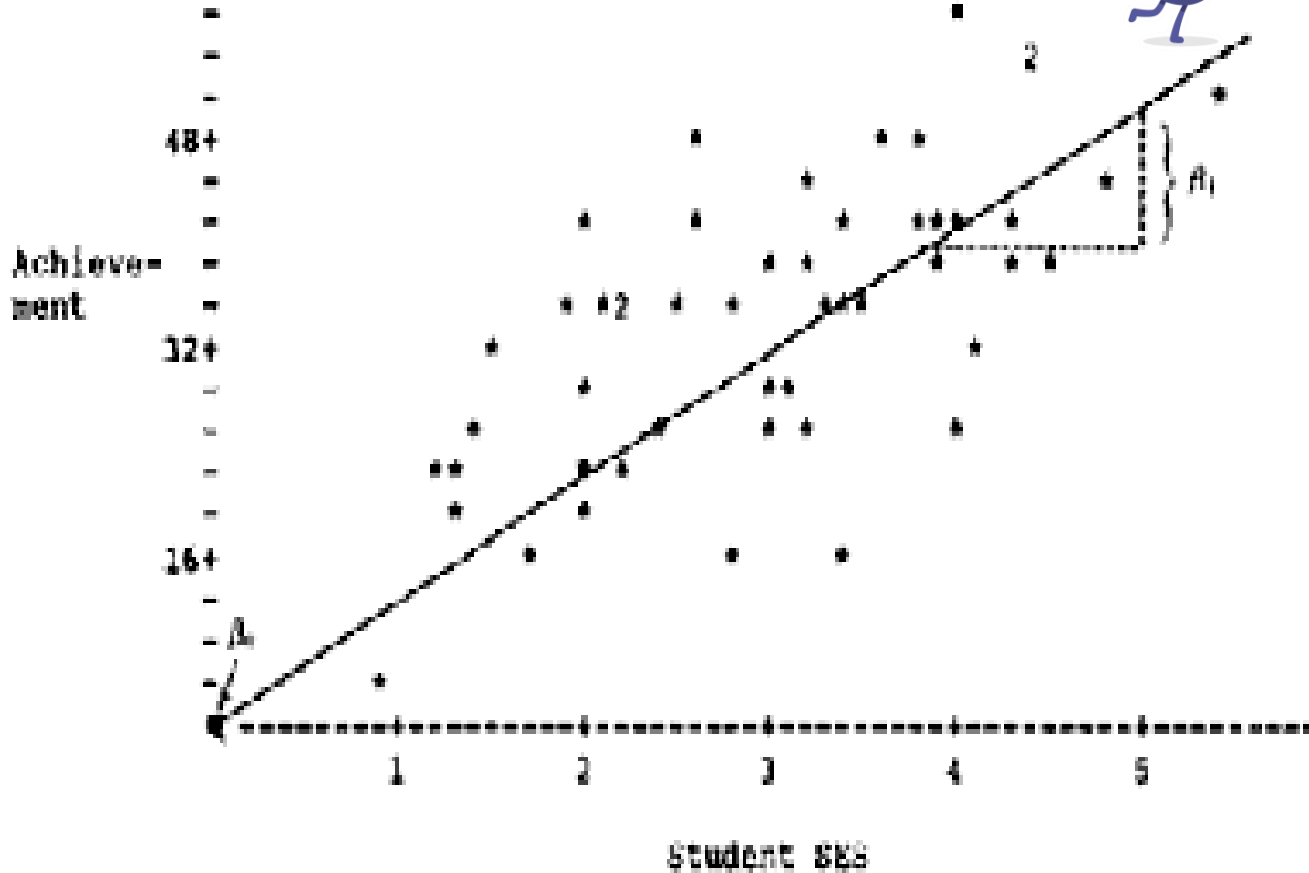
β_1 = A inclinação da linha (a "inclinação" ou "slope") indica como a Nota de Matemática muda com o aumento na Situação Socioeconômica.

Especificamente, β_1 indica a mudança esperada na Nota de Matemática para cada unidade de aumento em SS.

r_i = o termo de erro que representa a diferença entre o valor observado e o valor predito pela linha de regressão para o aluno i.

Testes precedentes da técnica

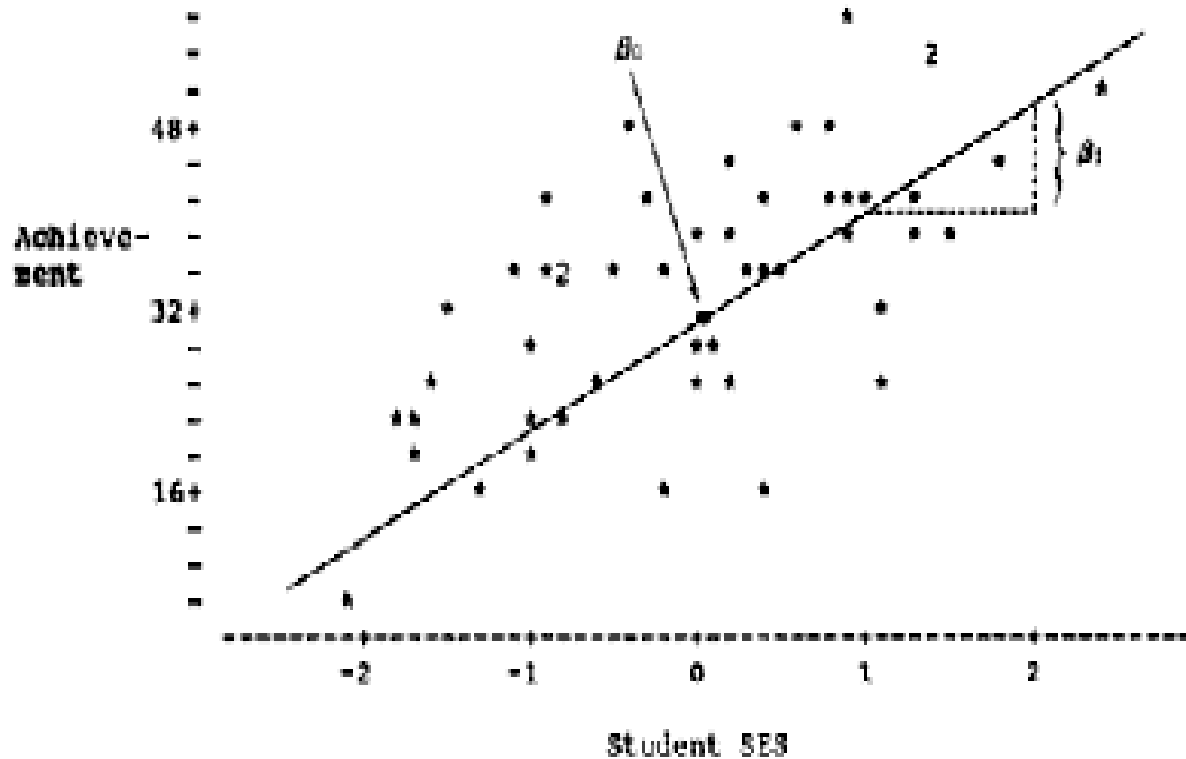
Diagrama de Dispersão que descreve o relacionamento entre NM e SS



Este diagrama ajuda a visualizar como a Nota de Matemática (NM) tende a variar em função da Situação Socioeconômica (SS). Se a linha de regressão se ajusta bem aos dados (os pontos estão próximos da linha), isso indica uma forte relação linear entre NM e SS. A inclinação da linha mostra a direção e a força dessa relação: uma inclinação positiva indicaria que notas mais altas em matemática estão associadas a uma situação socioeconômica mais alta, enquanto uma inclinação negativa indicaria o contrário.

Testes precedentes da técnica

Diagrama de Dispersão que descreve o relacionamento entre NM e SS



Centralização de X (SS): nessa equação temos informações úteis sobre o quanto elevações na escala de SS impacta NM, entretanto, podemos tornar o Intercepto mais informativo se centralizarmos a variável independente.

Centralização da Variável Independente (SS):

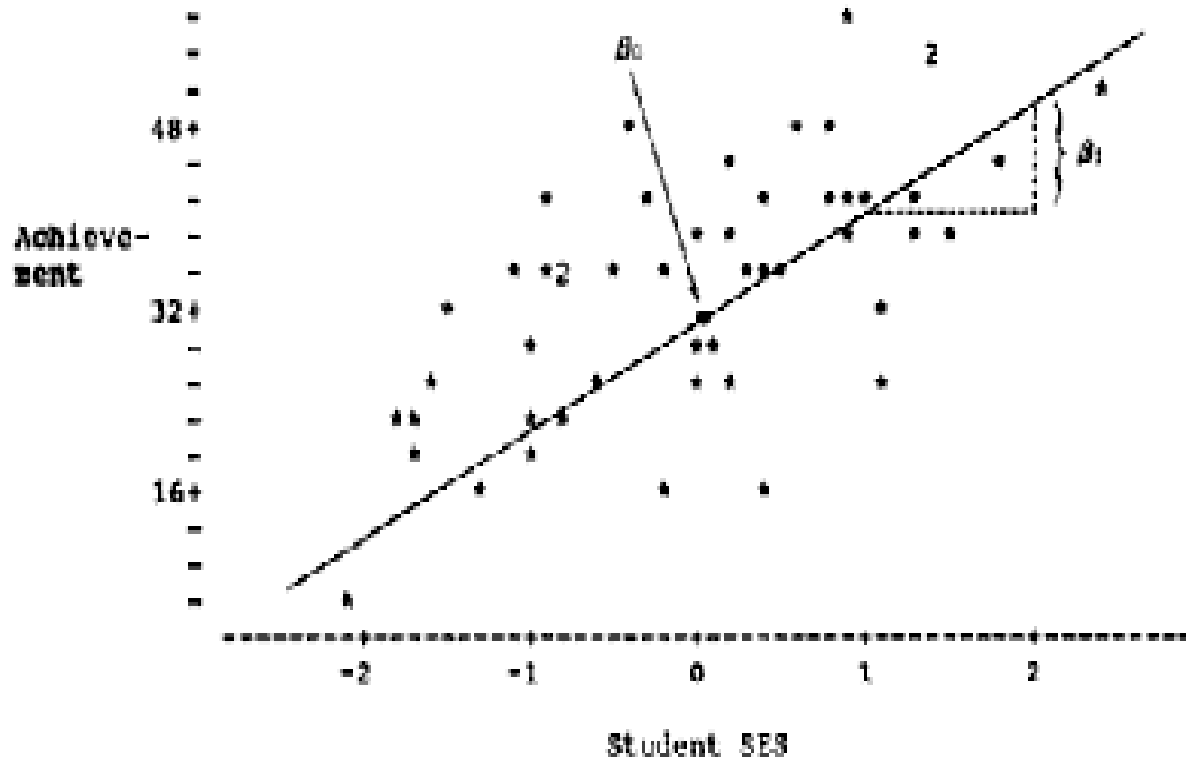
Propósito: A centralização é uma técnica onde se subtrai a média da variável de cada valor observado. Isso não altera a inclinação da reta de regressão, mas muda o intercepto (β_0) de forma que ele se torna mais significativo e interpretável.

Procedimento: Para centralizar SS, você subtrai a média de SS da escola (vamos chamar de \bar{x}) de cada valor de SS de cada aluno (X_i):

$$x \text{ centralizado} = x_i - \bar{x}$$

Testes precedentes da técnica

Diagrama de Dispersão que descreve o relacionamento entre NM e SS



Médias: ao invés da pontuação na escala de SS teremos informação sobre a média de SS na escola e sobre a distância de cada aluno em relação a esse patamar.

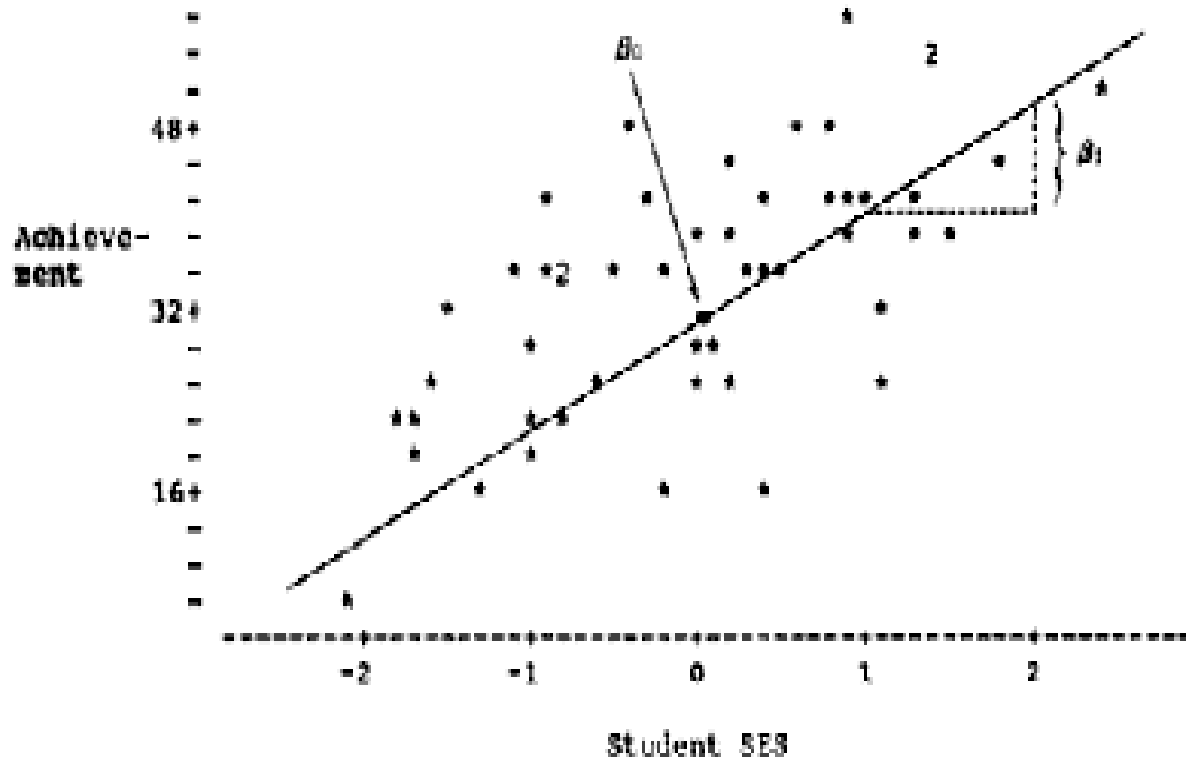
Mais realismo: com isso temos um intercepto que pode ser interpretado de maneira mais realista.

Procedimento: subtração da pontuação em SS de cada aluno pela média de SS da escola.

NM como função de SS centralizado: ao dispormos essa variável centralizada verificamos que o Intercepto (β_0) passa agora a representar a média de SS da escola e a inclinação (β_1) continua a mesma.

Testes precedentes da técnica

Diagrama de Dispersão que descreve o relacionamento entre NM e SS



Médias: ao invés da pontuação na escala de SS teremos informação sobre a média de SS na escola e sobre a distância de cada aluno em relação a esse patamar.

Mais realismo: com isso temos um intercepto que pode ser interpretado de maneira mais realista.

Procedimento: subtração da pontuação em SS de cada aluno pela média de SS da escola.

NM como função de SS centralizado: ao dispormos essa variável centralizada verificamos que o Intercepto (β_0) passa agora a representar a média de SS da escola e a inclinação (β_1) continua a mesma.

Testes precedentes da técnica

Duas Escolas: consideramos agora essa relação entre SS e NM em duas escolas hipotéticas.

DIFERENÇAS ENTRE AS ESCOLAS:

As duas escolas se distinguem em alguns aspectos, como
A média de NM na escola A é maior que na escola B

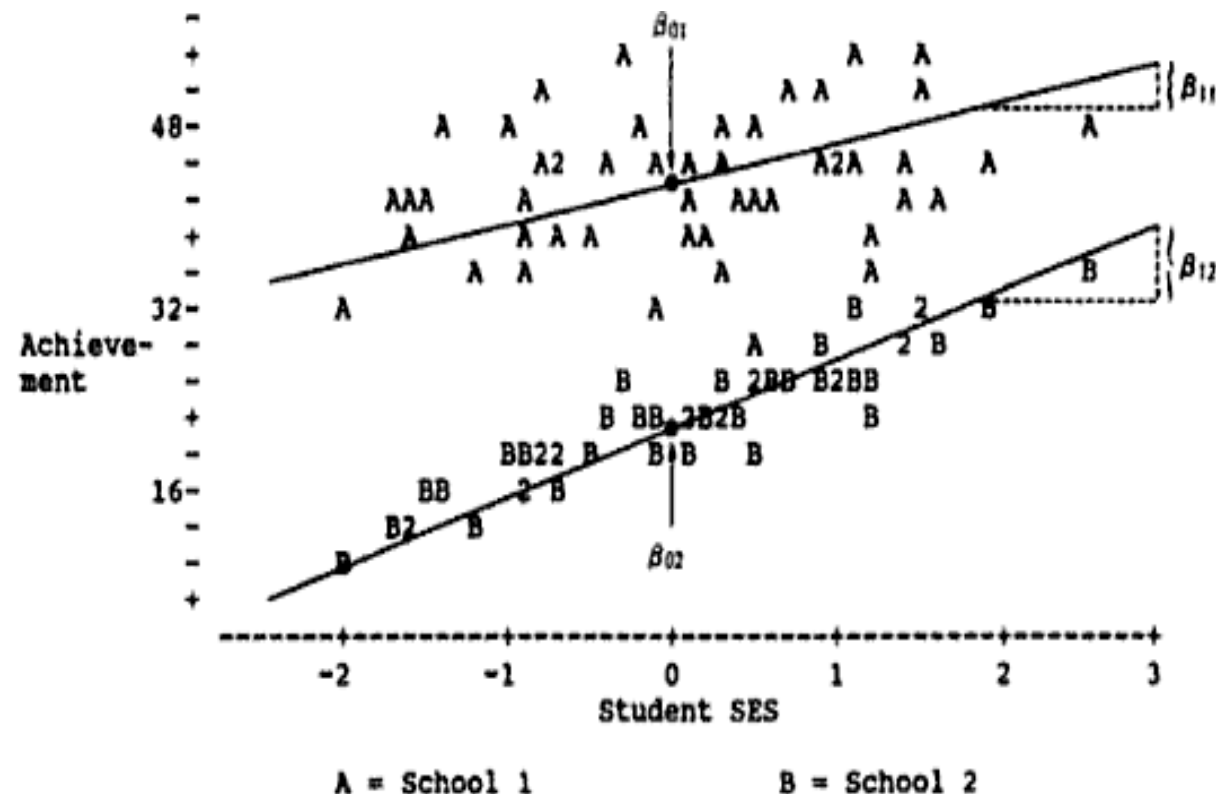
Essa diferença se reflete nos interceptos ($B_{01} > B_{02}$)

SS explica menos a NM na escola A do que na B

Essa diferença se reflete no coeficiente de regressão ($B_{11} < B_{12}$)

Efetividade e Igualdade: considerando que os alunos são aleatoriamente distribuídos entre as duas escolas, esses resultados indicam que

✓ Escola A é mais efetiva no ensino, uma vez que possui a melhor média e é mais igualitária, na medida em que o efeito de SS é menor sobre NM.



Pré-requisitos

- 1-Inicialmente realize o teste de modelo nulo ou vazio, de dois níveis, que nos permite estimar as probabilidades da sua variável dependente, no caso do exemplo, de tolerar politicamente os homossexuais, sem a inclusão dos preditores.
- 2-Em seguida, utilize o teste de diferenças de médias, ANOVA, a fim de identificar em que medida os diferentes níveis de análise são relevantes. Este teste nos apresenta o valor do Coeficiente de Correlação Intraclassa (ICC), isto é, o quanto a variabilidade da tolerância se deve aos fatores localizados no nível 2.
- 3-Após a realização dos testes acima mencionados, são organizados os modelos de regressão linear: o primeiro modelo deve conter apenas as variáveis de nível 1, dessa forma será possível, em seguida, avaliar os ganhos da inserção das variáveis de nível 2.
- 4-Na sequência, devemos especificar um modelo cuja inclinação possa variar aleatoriamente entre as unidades nacionais, configurando assim, um teste de hipótese que poderá justificar a pertinência das interações cross-level (individual e nacional). Faremos isso, através de três testes de diferenças de médias (Anova), um para cada dimensão da religiosidade (comportamento, pertencimento e intensidade), e os compararemos aos equivalentes sem randomização para identificar qual das dimensões das religiosidades possui resultado mais significativo. **Na sua pesquisa essas dimensões podem ser substituídas pelas categorias da principal variável independente, no caso de variável categórica.**
- 5-Por fim, realizamos os testes com as interações entre a variável religiosa (que apresentou maior significância) e a variável de nível 2, que mensura as mudanças na legislação dos países. A análise cross-level trata-se de uma importante técnica de análise no estudo de populações que estão inseridas em um mesmo contexto e, por isso, possuem características semelhantes. Nesses casos, tanto o nível menor (ou individual) quanto o nível maior (ou contextual) apresentam parte da explicação.

Após descrever a técnica, apresentaremos o material empírico, bem como as bases de dados utilizadas.

Dados: Rodada de 2023 do LAPOP

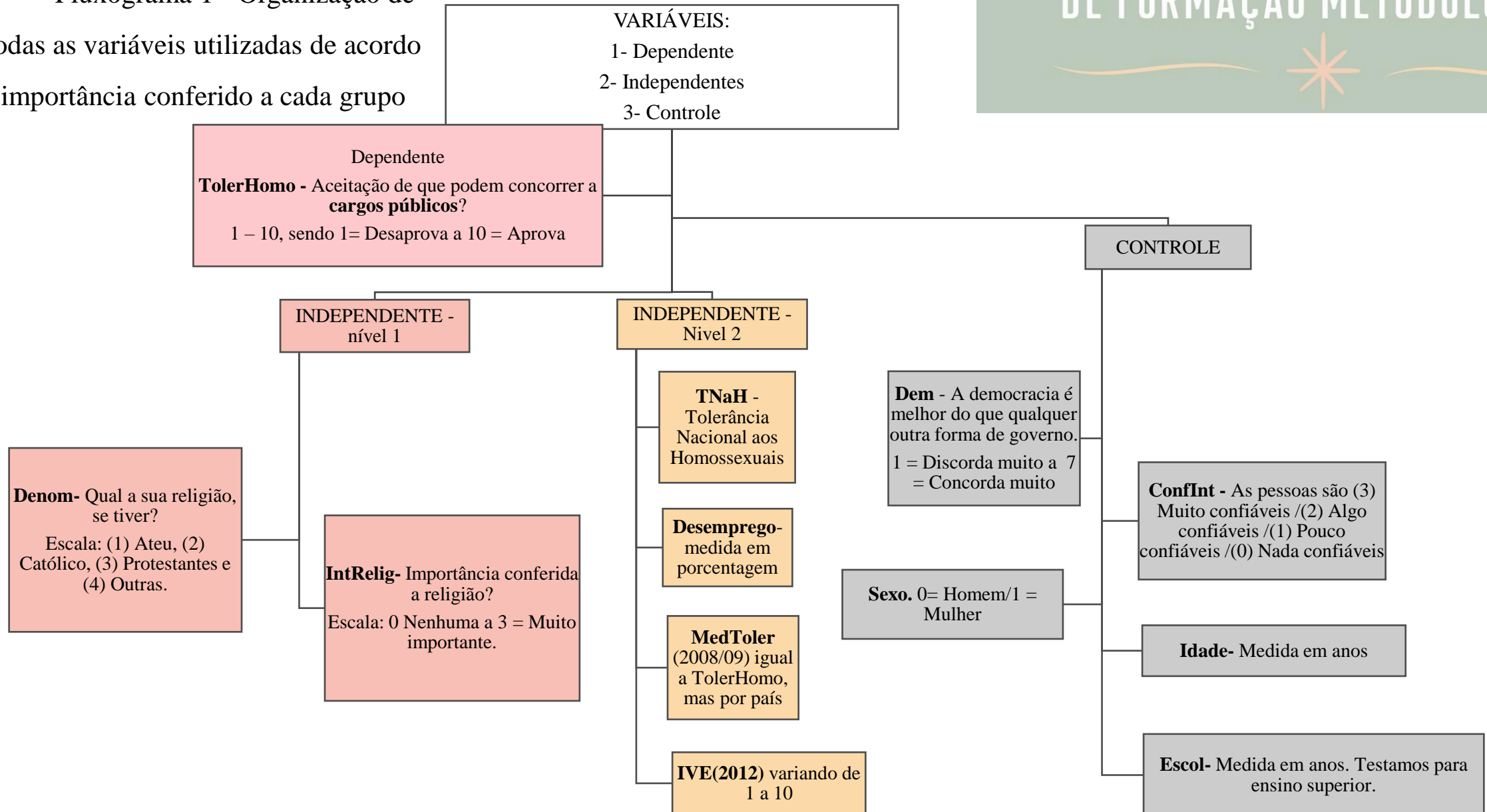
Tabela 1- Tamanho da amostra do Lapop por país – 2023

Dados

ESCOLA AMAZÔNICA DE FORMAÇÃO PSICOLÓGICA	
País	Amostra-2023
Argentina	1540
Bolívia	1706
Brasil	1526
Chile	1653
Costa Rica	1527
Equador	1604
El Salvador	1516
Guatemala	1556
Haiti	2221
Honduras	1602
México	1622
Nicarágua	1547
Panamá	1532
Paraguai	1524
Peru	1535
República Dominicana	1596
Uruguai	1517
Venezuela	1558
Total	28882

Variáveis

Fluxograma 1 - Organização de todas as variáveis utilizadas de acordo importância conferido a cada grupo



Dados

Tabela 2- medidas de análise do desemprego e da TNaH por país

Fonte: Autores, a partir dos dados do ILGA e do Actualitix, 2020. Observação: Não Houve Alteração no Status da variável TNaH entre as bases (2018-19 e 2023).

Países	Desemprego (%)	TNaH
Argentina	8	2
Bolívia	4	0
Brasil	13	2
Chile	7	0
Costa Rica	8	0
El Salvador	7	0
Equador	5	0
Guatemala	2	0
Haiti	41	0
Honduras	6	0
México	3	1
Nicarágua	6	0
Panamá	6	0
Paraguai	6	0
Peru	5	0
República Dominicana	23	0
Uruguai	8	2
Venezuela	27	0

Tabela 3- Média geral da variável dependente Tolerância política por país (TolerHomo), renomeada como MedToler, 2008 e 2009. Índice de Valores Emancipatórios, 2012

Países	Medida de Tolerância 2008-2009	IVE2012
Argentina	7,29	6,7
Bolívia	4,25	5,4
Brasil	5,77	6,5
Chile	5,49	6,0
Costa Rica	5,09	5,4
El Salvador	3,51	4,7
Equador	4,32	5,1
Guatemala	3,53	5,6
Haiti	2,27	5,3
Honduras	4,15	4,7
México	5,17	5,9
Nicarágua	4,45	5,3
Panamá	4,74	5,6
Paraguai	3,57	5,2
Peru	4,25	5,3
República Dominicana	3,27	5,0
Uruguai	6,94	7,2
Venezuela	4,44	5,4

Tabela 4- Modelo ANOVA para Tolerância Política, América Latina 2023

Variação de tolerância política entre os países		
	Variância	Desvio Padrão
(Intercept) τ_{00}	2.707924	1.645577
Residual σ^2	9.880173	3.143274
ICC ($\tau_{00}/(\tau_{00} + \sigma^2)$) 0.2151178(21,5%)		

Tabela 5- Ajustes de modelo com efeitos fixos e aleatórios
2023

	Df	AIC	LogLik	Razão de probabilidade	P
2023					
Modelo sem inclinações randômicas de preditores individuais				Não se aplica	
Denominação religiosa	28	129131.5	-64537.76		
Intensidade religiosa	14	129191.3	-64581.67		
Modelo com inclinação randômica da variável denominação					
Denominação religiosa	14	129191.3	-64581.67	87.8248	0
Intensidade religiosa	28	129148.2	-64546.10	71.14292	0

RESUMO

- **Intercepto:** A tolerância média à homossexualidade é 5.384.
- **Efeitos Aleatórios:** Há variação significativa na tolerância à homossexualidade entre os países (StdDev: 1.645577), e também dentro dos países (StdDev: 3.143274).
- **ICC:** Aproximadamente 21,4% da variação na tolerância à homossexualidade é explicada por diferenças entre países.

[CLIQUE AQUI PARA ACESSAR OS MODELOS](#)

Resultados

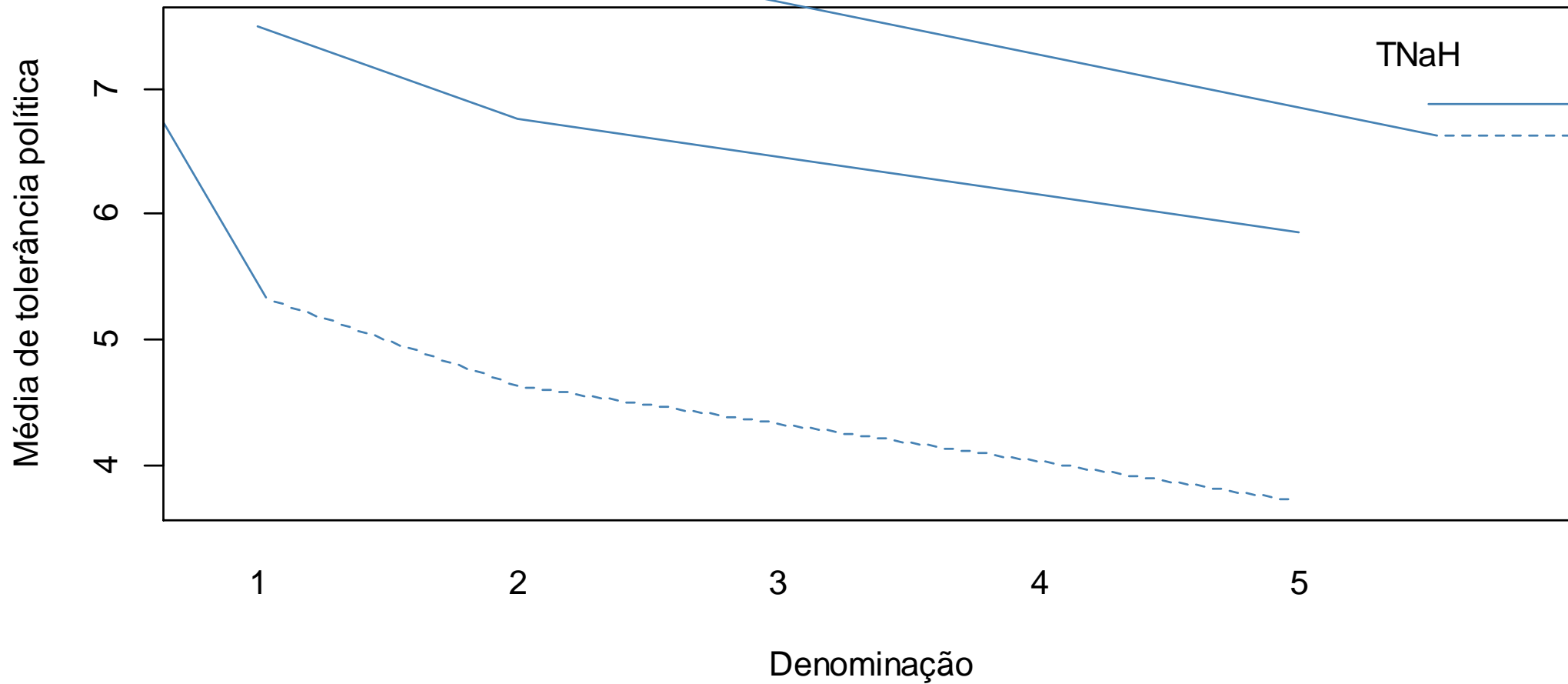
Quadro 1 – VALORES PREDITOS

ESCOLA AMAZÔNICA DE FORMAÇÃO METODOLÓGICA

PERFIL		VALORES PREDITOS
1)	Homem de 40 anos, ateu, que confia na democracia e nos outros cidadãos e que está em um contexto em que os homossexuais não possuem direito a união e adoção.	5,0
2)	Esse mesmo homem, de 40 anos, ateu, que confia na democracia e nos outros cidadãos, mas está inserido em um contexto em que os homossexuais têm direitos.	7,3
3)	Homem de 40 anos, católico, ativo religiosamente, que considera a religião importante, confia na democracia e nos outros cidadãos, mas está em um país que não possui nenhum direito para os homossexuais.	4,6
4)	Esse mesmo homem de 40 anos, católico, ativo religiosamente, que considera a religião importante, confia na democracia e nos outros cidadãos e está em um país que permite que homossexuais se relacionem, se casem e adotem crianças.	6,9
5)	Homem de 40 anos, evangélico pentecostal, ativo religiosamente, que considera a religião importante, confia na democracia e nos outros cidadãos, mas está em um país que não possui nenhum direito para os homossexuais.	3,4
6)	Esse mesmo homem de 40 anos, evangélico pentecostal, ativo religiosamente, que considera a religião importante, confia na democracia e nos outros cidadãos, mas está em um país que permite que homossexuais se relacionem, se casem e adotem crianças.	5,8
7)	Homem de 40 anos, protestante tradicional, ativo religiosamente, que considera a religião importante, confia na democracia e nos outros cidadãos, mas está em um país que não possui nenhum direito para os homossexuais.	3,8
8)	Esse mesmo homem de 40 anos, protestante tradicional, ativo religiosamente, que considera a religião importante, confia na democracia e nos outros cidadãos, mas está em um país que permite que homossexuais se relacionem, se casem e adotem crianças.	6,1
9)	Homem de outra religião, também com 40 anos, ativo religiosamente, que considera a religião importante, confia na democracia e nos outros cidadãos, mas está em um país que não possui nenhum direito para os homossexuais.	3,9

Resultados

Interação entre Denominações religiosas e Direitos dos Homossexuais (TNaH)



ESCOLA AMAZÔNICA DE FORMAÇÃO METODOLÓGICA

Obrigado

[CLIQUE AQUI PARA ACESSAR O MATERIAL DO CURSO](#)