



Universidade de Brasília
Departamento de Estatística

**Regressão Logística Multinível:
um estudo sobre o consumo de álcool
e inalantes entre crianças e adolescentes**

Paulo Henrique Fernandes de Araújo

Projeto apresentado para obtenção do título
de Bacharel em Estatística.

**Brasília
2017**

Paulo Henrique Fernandes de Araújo

**Regressão Logística Multinível:
um estudo sobre o consumo de álcool
e inalantes entre crianças e adolescentes**

Orientador:

Prof. Dr. **Luís Gustavo do Amaral Vinha**

Projeto apresentado para obtenção do título
de Bacharel em Estatística.

**Brasília
2017**

SUMÁRIO

1 INTRODUÇÃO	6
2 REFERENCIAL TEÓRICO	8
2.1 Regressão Logística	8
2.1.1 Regressão Logística Simples	10
2.1.2 Regressão Logística Múltipla	11
2.1.3 Estimação	12
2.2 Regressão Multinível	13
2.2.1 Regressão Linear Multinível	14
2.2.2 Regressão Logística Multinível	15
2.2.3 Construção do modelo	16
3 ANÁLISE DESCRITIVA	20
3.1 Variáveis da amostra	20
3.2 Características da amostra	21
3.3 Análise da associação	22
3.3.1 Álcool	22
3.3.2 Inalantes	26
4 RESULTADOS	30
4.1 Consumo de álcool	30
4.2 Consumo de inalantes	33
5 CONCLUSÃO	35
REFERÊNCIAS	36

1 INTRODUÇÃO

Assim como outros países em desenvolvimento, o Brasil ainda enfrenta diversos problemas em áreas importantes como a saúde pública. Um dos principais desafios dos governantes e uma das maiores preocupações da população brasileira é ter acesso a uma saúde de qualidade. Infelizmente, a falta de médicos, carência de hospitais e a má gestão de recursos financeiros ainda são desafios constantes. Por consequência, os cidadãos ficam à mercê da ineficiência do Estado. Um dos problemas que se encaixa na área da saúde pública é a inserção de crianças no mundo das drogas, que por sinal, não afeta apenas o usuário, mas também todas as pessoas próximas.

Os problemas decorrentes do uso abusivo de drogas podem incluir a ocorrência de diversas doenças tais como Hepatite B e C, contaminação pelo vírus do HIV/Aids, doenças sexualmente transmissíveis (DSTs); o desenvolvimento de transtornos mentais e comportamentais; mortalidade por intoxicação e violência (acidentes de trânsito e assassinatos); incapacitação dos indivíduos; e violência urbana (como roubos), tráfico, entre outros (Duarte, Stempluk, & Barroso, 2009). Devido a esses problemas é importante que as ações preventivas sejam iniciadas cedo, ainda na infância ou começo da adolescência. Para que essas ações preventivas sejam eficientes é de extrema importância estudar os fatores associados (sexo, idade, comportamento anti social, entre outros) ao uso de álcool, inalantes e outras drogas entre crianças e adolescentes.

Tendo em vista a importância do assunto, o governo brasileiro, recentemente, implementou três programas - Jogo Elos, #Tamojunto e Famílias Fortes - de prevenção ao uso de drogas em diversos municípios. Nesses programas são realizadas atividades preventivas com estudantes do ensino fundamental e familiares em ambiente escolar por meio da multiplicação de saberes de profissionais de saúde e assistência social. O Jogo Elos, direcionado a crianças entre seis e dez anos, se traduz em uma proposta lúdica na qual educandos, divididos em equipes, são convocados a seguir certas regras de convivência, combinadas coletivamente entre alunos e professor. O #Tamojunto tem como objetivo instrumentalizar adolescentes de 13 e 14 anos com habilidades sociais e habilidades da vida adquirindo conhecimento sobre drogas e suas consequências para a saúde. Similarmente, o programa Famílias Fortes é viabilizado por meio de reuniões semanais voltadas para o fortalecimento dos vínculos afetivos e de

cooperação entre pais, responsáveis e filhos entre 10 e 14 anos (Site MJC ¹).

O presente estudo faz parte de um projeto mais geral, a Avaliação da Efetividade, Qualidade de Implementação e Validade Social do Programa Famílias Fortes. Esse projeto tem como objetivo principal avaliar a efetividade do programa Famílias Fortes, versão adaptada para o Brasil, descrever a qualidade de sua implementação e avaliar sua validade social. Ele possui duração de 3 anos, a ser realizado por meio de ensaio clínico com crianças/adolescentes e pais ou responsáveis atendidos nas Unidades Básicas de Saúde-UBS e Centros de Referência em Assistência Social-CRAS. É prevista a participação de 2.860 crianças/adolescentes, dos quais 1.430 participarão de sete sessões do programa (Condição Experimental) e 1.430 receberão os serviços usuais da UBS e CRAS (Murta & Nobre-Sandoval, 2015). Esse programa social tem como foco crianças e adolescentes pertencentes a famílias em situação de vulnerabilidade social.

O objetivo geral deste trabalho é descrever e modelar o padrão de uso de álcool e inalantes por parte das crianças e adolescentes de 10 a 14 anos das escolas do Distrito Federal e Goiânia. Mais especificamente, o trabalho visa: a verificar a associação do consumo de álcool e inalantes com outras características de crianças e adolescentes e a modelar e descrever os dados utilizando o método de regressão logística multinível.

¹Para mais informações, acessar <http://www.justica.gov.br/noticias/mjc-coordena-programas-de-prevencao-ao-uso-de-drogas-no-ceara>

2 REFERENCIAL TEÓRICO

2.1 Regressão Logística

A análise de regressão é um tipo de modelagem onde são desenvolvidas expressões matemáticas que descrevem, em algum sentido, o comportamento de uma variável aleatória de interesse. Esta variável pode ser o preço do trigo no mercado mundial, o número de mortes por câncer de pulmão, o salário de funcionários de uma empresa, ou a presença ou ausência de uma certa doença. Em todos os casos, esta variável é chamada de variável resposta e é indicada com Y . Mais comumente, quando a variável Y é quantitativa a modelagem visa a descrever como a média da variável resposta $E(Y)$ se modifica de acordo com a variação de outros fatores (Jonh et al, 1988).

Antes de prosseguir para o modelo de regressão logística, será mostrado o porquê da regressão linear não ser um modelo adequado quando a variável resposta é dicotômica. Assim, considerando o modelo de regressão linear simples

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

uma vez que a variável resposta Y_i é dicotômica, assumindo valores 0 ou 1, a resposta esperada $E(Y_i)$ tem um significado especial: sempre representa a probabilidade de $Y_i = 1$ para os níveis da variável explicativa. Assim, desde que $E(e_i) = 0$, tem-se

$$E(Y_i) = \beta_0 + \beta_1 X_i. \quad (2)$$

Deve-se considerar que, utilizar o modelo de regressão simples quando a variável resposta é dicotômica, pode acarretar alguns problemas, tais como:

- não normalidade dos erros, logo um modelo que assume que os erros são normalmente distribuídos não é apropriado;
- variâncias heterogêneas dos erros, isso ocorre porque a variável resposta é dicotômica, portanto, $Var(Y_i) = \pi_i(1 - \pi_i)$;
- restrição na função resposta, como a resposta média representa a probabilidade de $Y_i = 1$ esse valor deve ser maior ou igual a 0 e menor ou igual a 1. Por essa razão, uma função linear não é adequada para esse tipo de modelagem.

Neste caso, quando a variável é dicotômica, a regressão logística é uma modelagem popular. Este modelo permite prever resultados a partir de um conjunto de variáveis explicativas que podem ser contínuas, discretas, dicotômicas ou uma mistura destas. Desse modo, a probabilidade de sucesso depende de outras variáveis.

A abordagem utilizada no problema das variáveis que não são normalmente distribuídas é incluir a transformação necessária e a escolha da distribuição de erros apropriada explicitamente no modelo estatístico. Este grupo de modelos estatísticos é chamado de modelos lineares generalizados e são definidos por três componentes (Hox, 2010):

- uma variável de resposta Y com uma distribuição de erro específica que tem média μ e variância σ^2 ;
- uma equação de regressão aditiva linear que produz um preditor não observado (latente) η da variável de resposta Y ;
- uma função de ligação que relaciona os valores esperados da variável de resposta Y aos valores preditos para $\eta : \eta = f(\mu)$.

Para facilitar a interpretação desses modelos é importante ter em mente o conceito de chance, usualmente chamado pelo nome em inglês, *odds*. Suponha que temos uma variável Y com duas categorias: sucesso, denotado por 1, ou fracasso, denotado por 0. Para a probabilidade de sucesso $P(Y = 1) = \pi$, a chance de sucesso é definida por

$$odds = \frac{\pi}{(1 - \pi)}. \quad (3)$$

Se a probabilidade de sucesso é maior que a de fracasso, $odds > 1$. Caso contrário, $odds < 1$. Se, por exemplo, $odds = 2$, a probabilidade de sucesso é duas vezes a probabilidade de fracasso. Por outro lado, se $odds = 0,5$, a probabilidade de sucesso é metade da probabilidade de fracasso. Isolando π na equação (3) obtemos que

$$\pi = \frac{odds}{(odds + 1)}, \quad (4)$$

sendo π a probabilidade de sucesso em função da *odds*. Com esse conceito fixado pode-se iniciar o estudo da regressão logística.

2.1.1 Regressão Logística Simples

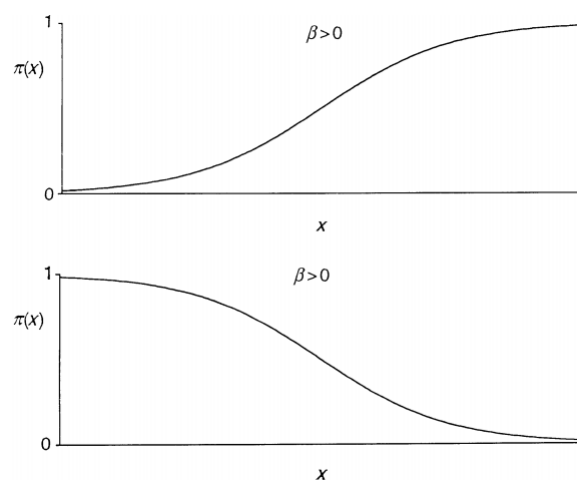
Um uso comum do modelo linear generalizado para dados dicotômicos é o modelo de regressão logística que é especificado por: distribuição de probabilidade binomial com média μ ; preditor linear, equação de regressão para η por exemplo, $\eta = \beta_0 + \beta_1 X_1$; função *logito* como função de ligação dada por $\eta = \text{logito}(\mu)$.

Assim, seja $E(Y|x) = \pi(x)$ a média condicional da variável resposta Y e x uma variável explicativa. Então, o modelo de regressão logística define $\pi(x)$ como

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}, \quad i = 1, \dots, n, \quad (5)$$

onde β_0 e β_1 são parâmetros do modelo. O fato da função logística variar entre 0 e 1 é o principal motivo pelo qual esse modelo é tão popular. A fórmula implica que $\pi(x)$ aumenta ou diminui como uma função em forma de "S" alongado (Figura 1). O parâmetro β_1 na Equação (5) determina a taxa de aumento ou diminuição da curva. O sinal de β_1 indica se a curva cresce ($\beta_1 > 0$) ou decresce ($\beta_1 < 0$) e a taxa de mudança aumenta quando $|\beta_1|$ aumenta. Quando $\beta_1 = 0$, o lado direito da Equação (5) passa a ser constante. Logo, $\pi(x)$ é idêntico para todo X , com isso, a curva se torna uma linha reta horizontal e conclui-se que a variável X é independente de Y (Agresti, 2007).

Figura 1 – Função Logística.



Fonte: Agresti (2007), página 71.

Em vez de usar o modelo apresentado na Equação 5, usa-se uma função de ligação $g(x)$, conhecida como *logito*, que permite que a variável resposta seja modelada como

$$\text{logito}(\pi(x)) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x, \quad (6)$$

isolando $\pi(x)$, obtém-se a probabilidade de sucesso

$$\begin{aligned} \left(\frac{\pi(x)}{1 - \pi(x)} \right) &= \exp(\beta_0 + \beta_1 x) \\ \pi(x) &= \left(\frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \right). \end{aligned} \quad (7)$$

A razão entre a chance do nível $X+1$ e a chance do nível X é conhecida como razão de chances (*odds ratio*) designada por Ψ , que é a base para a interpretação dos coeficientes do modelo de regressão logística. Assim, a *odds ratio* pode ser expressa como

$$\Psi = \left(\frac{\text{odds}(X+1)}{\text{odds}(X)} \right) = \frac{\left(\frac{\pi(X+1)}{1-\pi(X+1)} \right)}{\left(\frac{\pi(X)}{1-\pi(X)} \right)} = e^{\beta_1}. \quad (8)$$

A *odds ratio* é a medida de aumento ou redução da chance de sucesso quando a variável resposta x aumenta em uma unidade. Assim, depois de estimar os parâmetros, o efeito da variável explicativa na variável resposta pode ser medido através da *odds ratio*.

2.1.2 Regressão Logística Múltipla

Como visto até agora, foi introduzida a regressão logística no caso univariado. Porém, a força de uma técnica de modelagem consiste em modelar quantas variáveis forem necessárias, inclusive variáveis em diferentes escalas de mensuração. A abordagem de estimação e modelagem seguirá o mesmo procedimento usado na regressão logística simples. Considere o conjunto de p variáveis independentes descritas pelo vetor $\mathbf{x} = (x_1, x_2, \dots, x_p)$ e a probabilidade condicional de sucesso é dada por $\pi(\mathbf{x}) = P(Y = 1 \mid \mathbf{x})$. O *logito* do modelo de regressão logística é dado pela equação

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p, \quad (9)$$

tal que o modelo de regressão logística é expresso por

$$\pi(\mathbf{x}) = \frac{\exp(g(\mathbf{x}))}{1 + \exp(g(\mathbf{x}))}. \quad (10)$$

Deve-se considerar que, no caso da inclusão de variáveis qualitativas, onde, no registro dos dados as categorias são representadas por números, é inapropriado usá-las como se fossem variáveis quantitativas. Os números usados para representá-las não possuem significância numérica, eles são apenas identificadores. A maioria dos softwares estatísticos geram as variáveis identificadoras na presença de variáveis qualitativas. Em geral, se a variável qualitativa possui k categorias, será necessário o uso de $k-1$ variáveis indicadoras para a variável em estudo.

2.1.3 Estimação

Na regressão linear, os parâmetros do modelo podem ser estimados através do método dos mínimos quadrados. Na regressão logística, isso não é possível e, com isso, os parâmetros do modelo podem ser estimados através do método de máxima verossimilhança. Dado que a distribuição que está associada com a variável resposta dicotômica é a distribuição Bernoulli, que tem a seguinte forma

$$P(Y = y_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i}, \quad y_i = 0, 1, \quad (11)$$

sendo y_i a resposta do i -ésimo indivíduo da amostra com vetor de covariáveis $x_i = (x_{1i}, x_{2i}, \dots, x_{ki})$. A função de verossimilhança é dada por

$$L(\beta | Y) = \prod_{i=1}^n \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i}, \quad (12)$$

em que

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}}} \quad (13)$$

e $\beta_1, \beta_2, \dots, \beta_k$ são os coeficientes do modelo a serem estimados.

As estimativas dos parâmetros são valores que maximizam a função de verossimilhança $L(\beta | Y)$ para a amostra analisada. Estes estimadores são obtidos derivando a função em relação a cada parâmetro e igualando a zero. Para facilitar a maximização, é possível utilizar o logaritmo da função de verossimilhança. Porém, como essas equações são não-lineares nos parâmetros, resolvemos isso aplicando métodos numéricos computacionais, como o Newton-Raphson.

2.2 Regressão Multinível

A análise multinível é aplicável a uma ampla gama de situações em que a população apresenta estrutura hierárquica, ou seja, as unidades em um nível mais baixo estão aninhadas dentro de unidades em um nível mais alto. Considerando uma estrutura com apenas dois níveis, temos que o primeiro geralmente é referido como nível 1 e é esse nível em que a análise se concentra. O outro nível é referido como nível 2 e fornece o contexto para as unidades de nível 1. Por exemplo, unidades de nível 1 poderiam ser eleitores que estão aninhados em diferentes países (unidades de nível 2). A variável resposta é sempre medida para as unidades de nível 1, uma vez que este é o nível primário de análise.

O modelo de regressão multinível pode ser expresso como um sistema de equações de dois estágios. No primeiro estágio, temos uma equação que explica a variação individual. No segundo, uma equação que explica a variação entre grupos.

Apesar de existir várias abordagens diretas de analisar esses dados onde os indivíduos estão aninhados dentro de grupos, essas análises não levam em consideração a hierarquia. A primeira é ignorar a identidade do grupo e se concentrar exclusivamente na variação individual e em atributos de nível individual. Esta abordagem tem a inconveniência de ignorar a importância potencial dos atributos de nível de grupo na influência de resultados de nível individual. Além disso, se os resultados para indivíduos dentro de grupos estão correlacionados, a suposição de independência das observações é violada, o que resulta em erros padrão subestimados e, portanto, testes de significância que rejeitam a hipótese nula muito mais frequentemente do que o nível de significância sugere. Uma segunda abordagem é focar exclusivamente na variação entre grupos e em dados agregados ao nível do grupo, como resultado, perde-se muitas informações e poder da análise estatística. Uma terceira abordagem consiste em definir regressões separadas para cada grupo. Essa abordagem permite que coeficientes de regressão variem de grupo para grupo, mas não examina como as propriedades específicas do nível de grupo podem afetar resultados de nível individual ou interagir com variáveis de nível individual. Uma quarta abordagem é incluir características de grupo em equações de nível individual sob a forma de variáveis *dummy*. Essa abordagem é análoga ao planejamento de regressões separadas para cada grupo e não permite examinar exatamente quais características do grupo podem ser importantes para explicar o resultado. Além disso,

essa abordagem trata os grupos como não relacionados.

A análise multinível difere das abordagens mencionadas acima em diversos aspectos. Primeiro, permite que se analise de forma exclusiva os fatores de nível de grupo e os preditores de nível individual. Segundo, a não independência das observações dentro dos grupos é considerada. Terceiro, os grupos são tratados como relacionados e são vistos como provenientes de uma maior população de grupos. Por fim, utilizando a análise de regressão multinível pode-se examinar a variação individual e entre grupos, ou seja, permite lidar com essas variações simultaneamente (Hox 2010).

2.2.1 Regressão Linear Multinível

Segundo Hox (2010), o modelo mais básico para a compreensão da análise multinível é o que permite apenas a variação do intercepto entre grupos, mantendo constante os outros coeficientes. Então, considerando um modelo linear simples para os dados com estrutura hierárquica (com dois níveis) e com uma variável explicativa, tem-se

$$y_{ij} = \alpha_j + \beta_1 X_{ij} + e_{ij}, \quad (14)$$

onde y_{ij} é a variável resposta para o i -ésimo indivíduo no nível 1 dentro do j -ésimo grupo no nível 2, α_j é o intercepto para o grupo j , X_{ij} é a variável explicativa para o i -ésimo indivíduo e o j -ésimo grupo, β_1 é o efeito de X_{ij} e e_{ij} é o efeito aleatório de nível 1. O termo α_j é uma variável aleatória em vez de uma constante e pode ser escrita como

$$\alpha_j = \beta_0 + u_j, \quad (15)$$

onde β_0 é o intercepto (constante no nível 2) e u_j é um efeito aleatório que permite que o intercepto varie no nível 2. Combinando as equações (14) e (15), o modelo linear de dois níveis pode ser escrito como

$$y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + e_{ij}. \quad (16)$$

Na equação (16), u_j e e_{ij} são quantidades aleatórias que seguem distribuições normais com média zero e variâncias σ_u^2 e σ_e^2 , respectivamente. Além disso supõem-se que esses termos são não correlacionados, ou seja, $cov(u_j, e_{ij}) = 0$ e $Cov(u_j, u_{j'}) = 0$ para $j \neq j'$. Neste modelo, β_0 e β_1 são conhecidos como parâmetros fixos e σ_u^2 e σ_e^2 são conhecidos como

parâmetros aleatórios. A equação (16) também é conhecida como modelo de componentes de variância, uma vez que a variância da resposta, com os componentes fixos (β_0 e β_1), é

$$\text{var}(y_{ij} \mid \beta_0, \beta_1, x_{ij}) = \text{var}(u_j + e_{ij}) = \sigma_u^2 + \sigma_e^2 \quad (17)$$

que é a variação total obtida somando a variância de nível 1 e de nível 2. A covariância entre duas unidades de nível 1 (digamos, i_1, i_2) pode ser definida como

$$\text{cov}(u_j + e_{i_1j}, u_j + e_{i_2j}) = \sigma_u^2. \quad (18)$$

Usando este modelo, podemos definir a correlação intraclass ρ pela equação

$$\rho = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2)}. \quad (19)$$

A Equação 19 simplesmente destaca que a correlação intraclass é a proporção da variância no nível de grupo em comparação com a variância total. Essa correlação indica a proporção da variância explicada pelo agrupamento na população. Também, a correlação intraclass ρ pode ser interpretada como a correlação esperada entre duas unidades desenhadas aleatoriamente que estão no mesmo grupo.

2.2.2 Regressão Logística Multinível

Basicamente, o modelo logístico de dois níveis é equivalente ao modelo (6), exceto para a variável resposta. Seja y_{ij} a variável resposta binária associada ao indivíduo i de nível 1 aninhada dentro do grupo j de nível 2. Considerando que, na regressão logística, o $\pi(x)$ é modelado usando a função de ligação *logito*, o modelo de regressão logística de dois níveis pode ser escrito como

$$\ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 X_{ij} + u_j, \quad (20)$$

onde u_j é o efeito aleatório do nível 2. Sem u_j , a equação (20) pode ser considerada como um modelo padrão de regressão logística. Aqui, u_j é uma quantidade aleatória e segue distribuição $N(0, \sigma_u^2)$. O modelo (20) pode ser escrito dividindo-se em dois modelos, um para o nível 1 e outro para o nível 2:

$$\ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_{0j} + \beta_1 X_{ij}, \quad (21)$$

$$\beta_{0j} = \beta_0 + u_j. \quad (22)$$

Nesse caso, onde não temos a variância individual, deve-se usar a variância que é dada por $\pi^2/3 \approx 3,29$, onde $\pi \approx 3,14$. (Evans, Hastings e Peacock, 2000).

2.2.3 Construção do modelo

O princípio de que os modelos devem ser parcimoniosos, isto é, o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta indica que geralmente deve-se escolher o modelo mais simples. A fim de testar a significância dos coeficientes e para comparar os modelos foram utilizados o teste Wald e o índice AIC, respectivamente, que são especificados abaixo.

O teste de Wald é utilizado para avaliar se o parâmetro é estatisticamente significativo para um determinado nível de significância. Logo, as hipóteses a serem testadas são

$$H_0 : \beta = 0,$$

$$H_1 : \beta \neq 0.$$

A estatística do teste é obtida por meio da razão do coeficiente pelo seu respectivo erro padrão, tem-se assim, que essa estatística apresenta distribuição normal padrão.

O AIC, *Akaike's Information Criterion*, é um índice de ajuste geral para comparar modelos estatísticos. Para modelos de regressão multinível é convenientemente calculado a partir da *deviance* d e o número de parâmetros estimados q :

$$AIC = d + 2q.$$

O AIC é um índice de ajuste geral que pressupõe que estamos comparando modelos que são adequados ao mesmo conjunto de dados, usando o mesmo método de estimativa.

Para a escolha do modelo será utilizada a estratégia que começa com um modelo com todas as variáveis e retira-se aquelas que apresentam a menor relação com a variável resposta, iterativamente e uma por vez, sobrando apenas aquelas que são significativas. Em seguida, compara-se o modelo completo com o modelo reduzido e escolhe-se o que apresenta o melhor índice AIC. O procedimento começa pela parte fixa e depois pela parte aleatória

e, uma vez que temos tamanhos de amostras maiores no nível mais baixo, faz sentido iniciar a partir daí. Além disso, os parâmetros fixos são tipicamente estimados com muito mais precisão do que os aleatórios, começamos com os coeficientes de regressão fixos e adicionamos efeitos aleatórios em uma fase posterior. Os diferentes passos para a construção do modelo multinível são propostos por Hox (2010) e descritos a seguir.

Passo 1

Analisa-se o modelo sem variáveis explicativas. Este modelo, o modelo nulo, é dado por

$$Y_{ij} = \beta_{00} + u_{0j} + e_{ij}, \quad (23)$$

onde β_{00} é o intercepto da regressão, u_{0j} e e_{ij} são os resíduos do nível do grupo e do nível individual, respectivamente. Este modelo é útil porque fornece uma estimativa da correlação intraclasses ρ , dado pela Equação 19.

Passo 2

Analisa-se o modelo com todas as variáveis explicativas do nível mais baixo excluindo, uma por vez, aquelas que não são significativas. Isso significa que os componentes de variância correspondentes aos coeficientes são fixados em zero. Este modelo é escrito como

$$Y_{ij} = \beta_{00} + \beta_{p0}X_{pij} + u_{0j} + e_{ij}, \quad (24)$$

onde o X_{pij} são as p variáveis explicativas no nível mais baixo. Nesta etapa, avalia-se a contribuição de cada variável explicativa de nível individual e a significância de cada preditor. Para comparar a melhoria do modelo final escolhido compara-se o AIC deste modelo e do modelo anterior (o modelo nulo).

Passo 3

Adiciona-se as variáveis explicativas de nível mais alto

$$Y_{ij} = \beta_{00} + \beta_{p0}X_{pij} + \beta_{0q}Z_{qj} + u_{0j} + e_{ij}, \quad (25)$$

onde o Z_{qj} são as q variáveis explicativas do nível do grupo. Este modelo permite examinar se as variáveis explicativas de nível de grupo ajudam a explicar a variável resposta. Novamente,

usa-se o teste de Wald e o AIC para testar a melhoria do ajuste. Os modelos nos passos 2 e 3 são frequentemente designados como modelos de componentes de variância, porque eles decompõem a variância do intercepto em diferentes componentes de variância, para cada nível hierárquico. Em um modelo de componente de variância, o intercepto de regressão varia em todos os grupos, mas as inclinações de regressão são assumidas fixas.

Passo 4

Avalia se algum dos coeficientes de qualquer uma das variáveis explicativas tem um efeito aleatório significativo entre os grupos. Este modelo, o modelo de coeficiente aleatório, é dado por

$$Y_{ij} = \beta_{00} + \beta_{p0}X_{pij} + \beta_{0q}Z_{qj} + \mu_{pj}X_{pij} + u_{0j} + e_{ij}, \quad (26)$$

onde os μ_{pj} são os coeficientes aleatórios das variáveis explicativas X_{pij} . O teste para a variação da inclinação aleatória é melhor quando feito com uma variável por vez. Quando inclui-se todos os termos de variância possíveis em um modelo, o resultado é provavelmente uma super parametrização do modelo, com sérios problemas de estimativa, como problemas de convergência ou cálculos extremamente lentos. As variáveis que foram omitidas no passo 2 podem ser analisadas novamente neste passo. Uma variável explicativa que não tenha uma inclinação média significativa de regressão (conforme testado no passo 2) pode ter efeito aleatório significativo para essa inclinação. Depois de decidir qual dos coeficientes tem uma variância significativa entre os grupos adiciona-se todos esses efeitos aleatórios simultaneamente em um modelo final.

Passo 5

Adiciona-se interações entre variáveis explicativas em nível de grupo e aquelas de nível individual que apresentaram variação significativa da inclinação no passo 4. Isso leva ao modelo completo:

$$Y_{ij} = \beta_{00} + \beta_{10}X_{ij} + \beta_{01}Z_j + \beta_{11}X_{ij}Z_j + \mu_{1j}X_{1ij} + u_{0j} + e_{ij}. \quad (27)$$

Tabela de classificação

Para resumir o poder preditivo da regressão escolhida, podemos fazer uma tabela de classificação. Esta tabela classifica a resposta verdadeira de y com sua predição. A predição é $\hat{y} = \text{sim}$ quando $\hat{\pi}_i > \pi_0$ e $\hat{y} = \text{não}$ quando $\hat{\pi}_i \leq \pi_0$, para algum ponto de corte π_0 .

Dois resumos úteis do poder preditivo são a *sensibilidade* $= P(\hat{y} = 1 \mid y = 1)$ e a *especificidade* $= P(\hat{y} = 0 \mid y = 0)$. Outro resumo do poder de predição da tabela de classificação é a proporção de classificações corretas. Logo, $P(\text{classificação correta}) = P(y = 1, \hat{y} = 1) + P(y = 0, \hat{y} = 0) = P(\hat{y} = 1|y = 1)P(y = 1) + P(\hat{y} = 0|y = 0)P(y = 0)$, que é uma média ponderada da sensibilidade e especificidade.

3 ANÁLISE DESCRITIVA

3.1 Variáveis da amostra

Nesse estudo, os dados foram obtidos a partir de aplicação de questionário para as crianças e adolescentes de algumas escolas do Distrito Federal e Goiânia, que participam de algum programa social. Na Tabela 1 será mostrado as variáveis utilizadas nesse estudo e uma breve descrição.

Tabela 1 – Descrição das variáveis

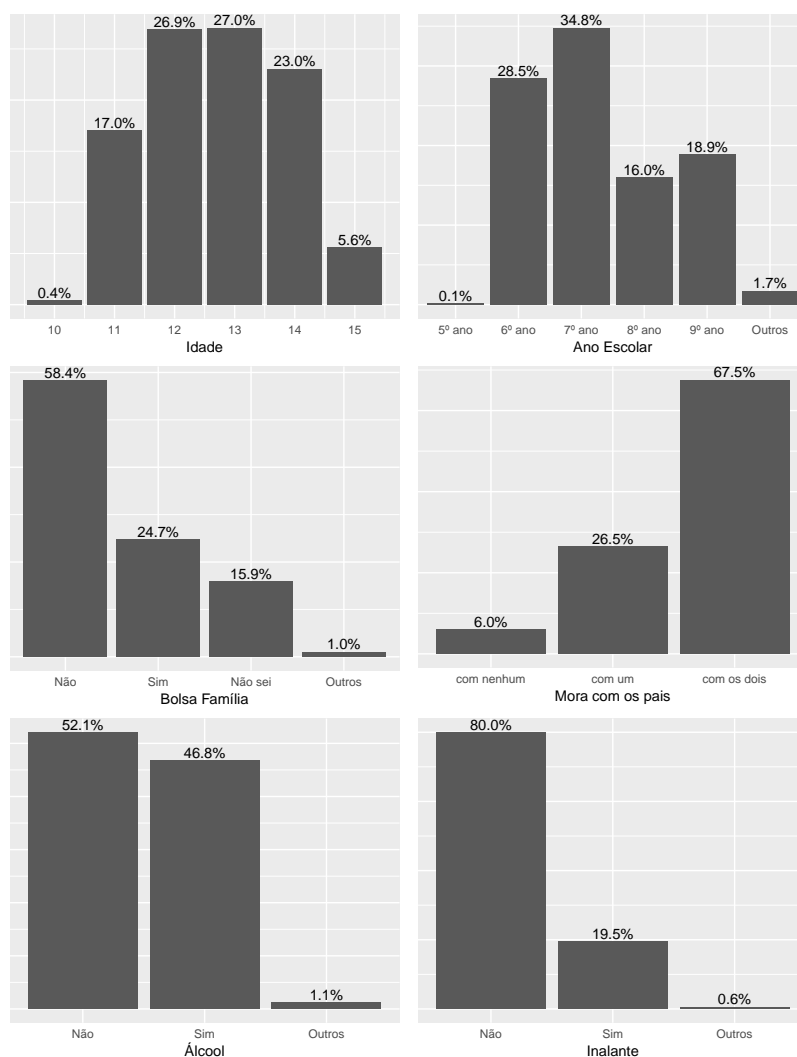
Características	Variável	Descrição
Sociodemográficas	Sexo	1 = menina; 2 = menino.
	Idade	Idade dos participantes em anos.
	Ano escolar	Ano escolar no momento da entrevista.
	Bolsa Família	Participação da família no programa (0 = não, 1 = sim, 2 = não sei).
	Mora com os pais	2 = sim, com os dois; 1 = sim, com um; 0 = não, com nenhum.
	Escolaridade da mãe	0 = não possui ensino médio; 1 = possui ensino médio; 2 = não sabe.
Uso de droga	Álcool	Utilizou alguma bebida alcoólica; 1=não, 2=sim.
	Inalantes	Utilizou algum inalantes (loló, lança, cola, éter, removedor de tinta, gasolina, entre outros), exceto cocaína. 1=não, 2=sim.
Autoeficácia ¹		Fatores associados a escala de autoeficácia. Exemplos de itens:
	Aprendizagem autorregulada	"Estudar mesmo quando há outras coisas interessantes para fazer".
	Autoassertiva	"Manter-me firme com alguém que está pedindo para que eu faça alguma coisa absurda ou inconveniente".
	Suporte social	"Pedir aos meus pais, ou responsáveis, que me ajudem com um problema".
Práticas parentais ¹		Fatores associados a escala de práticas parentais. Exemplo de itens:
	Apoio emocional	"Encontra um tempo para estar comigo e fazermos juntos algo agradável".
	Intromissão	"Dá palpite em tudo que eu faço".
	Incentivo autonomia	"Me estimula para que eu tome decisões por conta própria".
	Supervisão	"Procura saber aonde vou quando saio de casa".
	Cobrança de responsabilidade	"Exige que eu vá bem na escola".
	Controle punitivo	"Me proíbe de fazer algo que gosto quando eu faço alguma coisa errada".
Comportamento social ¹	Anti social	Fator associado a escala de comportamento anti social, por exemplo: "Xingou pessoas".

¹Escala de 1 a 5 construída a partir de itens, exemplificado na descrição.

A autoeficácia se caracteriza como a percepção que o indivíduo possui sobre sua capacidade de realizar com sucesso determinada tarefa; o que pode vir a afetar suas decisões e desempenhos profissionais e escolares (Barros e Batista-dos-Santos, 2010). As práticas parentais podem ser entendidas como conjunto de comportamentos emitidos pelos pais ao educar ou socializar os filhos que levam a um objetivo comum (Teixeira, Oliveira & Wottrich, 2006). O comportamento antissocial é aquele viola normas e interesses sociais, além de ser uma ação prejudicial contra outros. (Bringas et. al 2006)

3.2 Características da amostra

Figura 2 – Frequências relativas.



Participaram do estudo 699 participantes, sendo 379 meninas (54,5%), 305 meninos (43,9%) e 15 (1,6%) que não informaram o sexo. As demais características dos participantes estão apresentadas na Figura 2. Nesse trabalho, foram classificados como "outros" aqueles que não responderam, deram respostas incoerentes e erros de digitação.

Os participantes tinham entre 10 e 15 anos de idade, com média etária de 12,72 anos ($DP=1,17$) e a maioria estava entre 11 e 14 anos. Os dados revelaram que a maioria dos participantes cursava o 6º ano (28,5%) e o 7º ano (34,8%) do ensino fundamental e 58,4% das famílias dos participantes não recebiam o incentivo do Bolsa Família (Figura 2 (c)). Grande parte (67,5%) moravam com os pais. Com a Figura 2 (e) percebe-se que 46,8% já consumiram alguma bebida alcoólica. Além disso, 19,5% disseram ter consumido algum tipo de inalantes ao menos uma vez na vida.

3.3 Análise da associação

3.3.1 Álcool

Para testar a associação das variáveis qualitativas com o consumo de álcool foi utilizado o teste Qui-quadrado (χ^2) (Tabela 2).

Na amostra utilizada no presente estudo verifica-se que o consumo de álcool é diferente entre os alunos do 6º ano (26,5%) e os alunos do 8º ano (69,3%), além disso, o teste rejeita a hipótese de não associação entre o ano escolar e o consumo de álcool ($p\text{-valor}<0,001$). As variáveis Bolsa Família e escolaridade da mãe também apresentaram associação com o consumo de álcool (Tabela 2). A variável mora com os pais apresentou diferença no consumo de álcool entre os que não moravam com os pais (40,4%) e os que moravam com os pais (48,2%), apesar disso, esta variável não se mostrou significativamente associada ao consumo de álcool, dessa forma não se pode afirmar que existe diferença na população. Percebe-se que o consumo é pouco diferente na variável sexo, logo, o teste não rejeita a hipótese de não associação.

Tabela 2 – Frequências relativas das variáveis por uso de álcool e teste Qui-quadrado

Variável		Álcool		Total	χ^2	$p - valor$
		Não	Sim			
Sexo	Menina	209 (55,2%)	169 (44,7%)	378	2,549	0,110
	Menino	148 (48,8%)	155 (51,1%)	303		
Ano Escolar	6º ano	144 (73,4%)	52 (26,5%)	192	76,517	<0,001
	7º ano	137 (56,6%)	105 (43,3%)	242		
	8º ano	34 (30,6%)	77 (69,3%)	111		
	9º ano	43 (33,3%)	86 (66,6%)	129		
Bolsa Família	Não	186 (46,1%)	217 (53,8%)	403	15,774	<0,001
	Sim	105 (61,4%)	66 (38,5%)	171		
	Não sabe	68 (61,8%)	42 (38,1%)	110		
Mora com os pais	Nenhum	25 (59,5%)	17 (40,4%)	42	1,018	0,601
	Apenas um	98 (53,5%)	85 (46,4%)	183		
	Com os dois	241 (51,7%)	225 (48,2%)	466		
Escolaridade da mãe	Não concluiu o ensino médio	95 (56,8%)	72 (43,1%)	167	15,167	<0,001
	Terminou o ensino médio ou mais	156 (45,0%)	190 (54,9%)	346		
	Não sabe	101 (62,3%)	61 (37,6%)	162		

O estudo da associação entre a variável álcool e as variáveis quantitativas são dados pela Figura 3 e Tabela 3. Nota-se que visualmente há diferença entre os grupos das variáveis da Figura 3. Utilizando o teste *t-student* para comparar as médias de idade entre os que consumiram álcool e os que não consumiram obtivemos médias de 12,29 anos (não consumiram álcool) e 13,18 (consumiram). Além disso, as médias populacionais dessa variável foram consideradas significativamente diferentes. Na Tabela 3, encontram-se as demais comparações.

Figura 3 – Boxplot das variáveis em relação ao consumo de álcool.

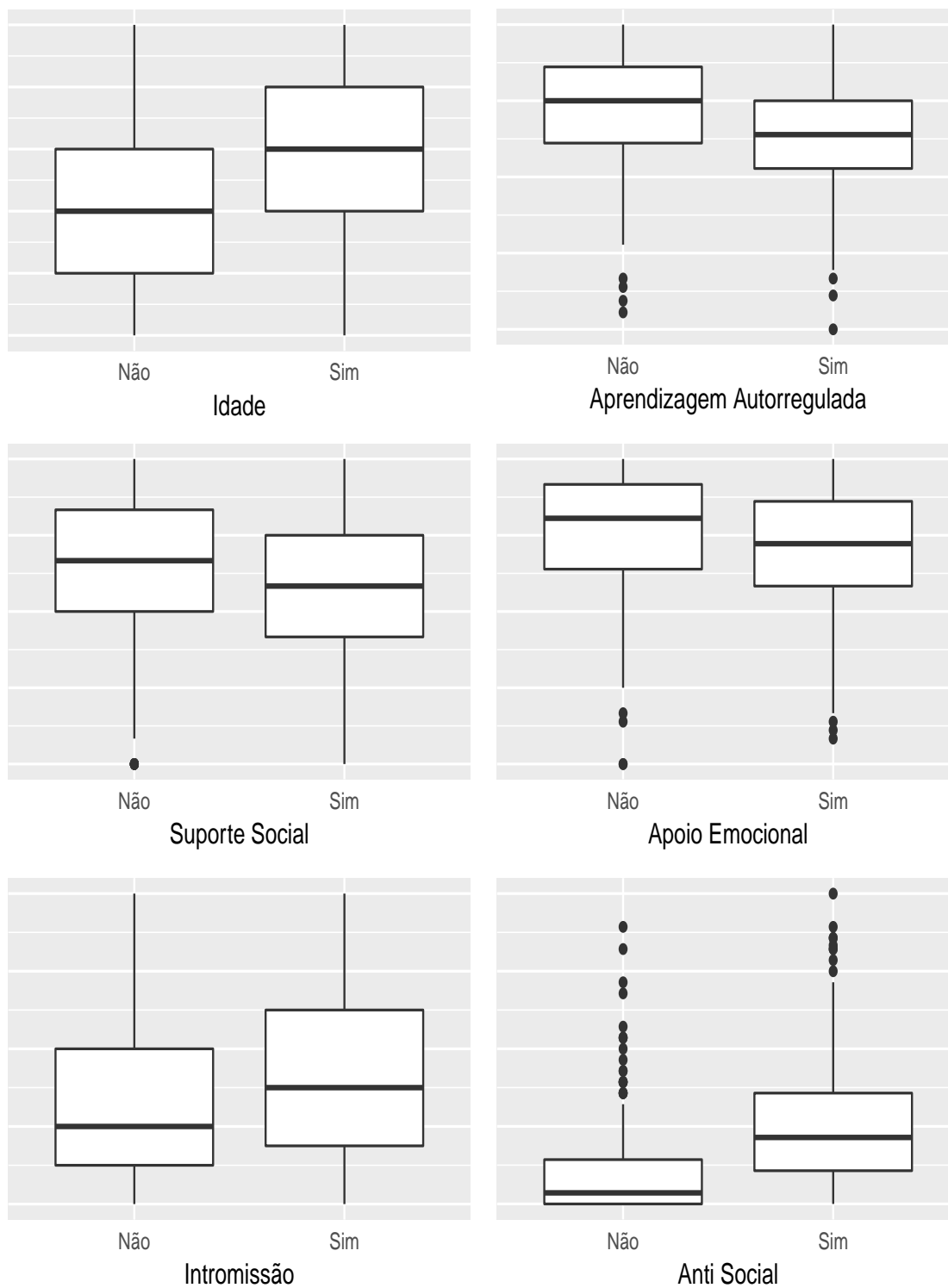


Tabela 3 – Médias das variáveis por uso de álcool e teste *t-student*

Característica	Variável	Uso de álcool	Média	<i>T</i>	p-valor
Auto eficácia	Aprendizagem autorregulada	Não	3,89	7,03	<0,001
		Sim	3,54		
	Auto assertiva	Não	3,94	1,26	0,206
		Sim	3,87		
	Suporte social	Não	3,70	5,75	<0,001
		Sim	3,27		
Práticas parentais	Apoio emocional	Não	4,05	3,92	<0,001
		Sim	3,80		
	Intromissão	Não	2,26	-5,36	<0,001
		Sim	2,69		
	Incentivo autonomia	Não	3,14	-2,37	0,017
		Sim	3,34		
	Supervisão	Não	4,07	0,29	0,769
		Sim	4,05		
	Cobrança responsabilidade	Não	4,38	1,50	0,133
		Sim	4,28		
	Controle punitivo	Não	3,29	-0,70	0,483
		Sim	3,35		
Comportamento	Anti social	Não	1,39	-11,51	<0,001
		Sim	1,99		
Sociodemográfica	Idade	Não	12,29	-10,83	<0,001
		Sim	13,18		

De acordo com a Tabela 3, os participantes que relatam não ter usado álcool apresentaram aprendizagem autorregulada e suporte social significativamente maior, quando comparados aos que relataram uso. Por outro lado, considerando a eficácia auto assertiva, apesar da tendência de médias maiores para os participantes que não relataram uso de álcool, não houve diferenças significativas entre os dois grupos.

Os participantes que relataram não ter feito uso de álcool apresentaram média de apoio emocional significativamente maior, quando comparados aos que relataram ter experimentado álcool. Por outro lado, ter feito uso de álcool, mostrou estar significativamente associado à maior média de intromissão e incentivo a autonomia. Supervisão do comportamento, cobrança de responsabilidade e controle punitivo não apresentaram diferenças significativas entre os grupos, sugerindo que essas práticas não influenciam no comportamento de experimentar ou não álcool.

O comportamento anti social apresentou a maior diferença entre as médias, o que resulta em uma maior associação. Ou seja, aqueles que possuem um maior comportamento anti social tem mais chances de consumir álcool.

3.3.2 Inalantes

Para testar a associação das variáveis qualitativas com o consumo de inalantes foi utilizado o teste Qui-quadrado (χ^2) (Tabela 4).

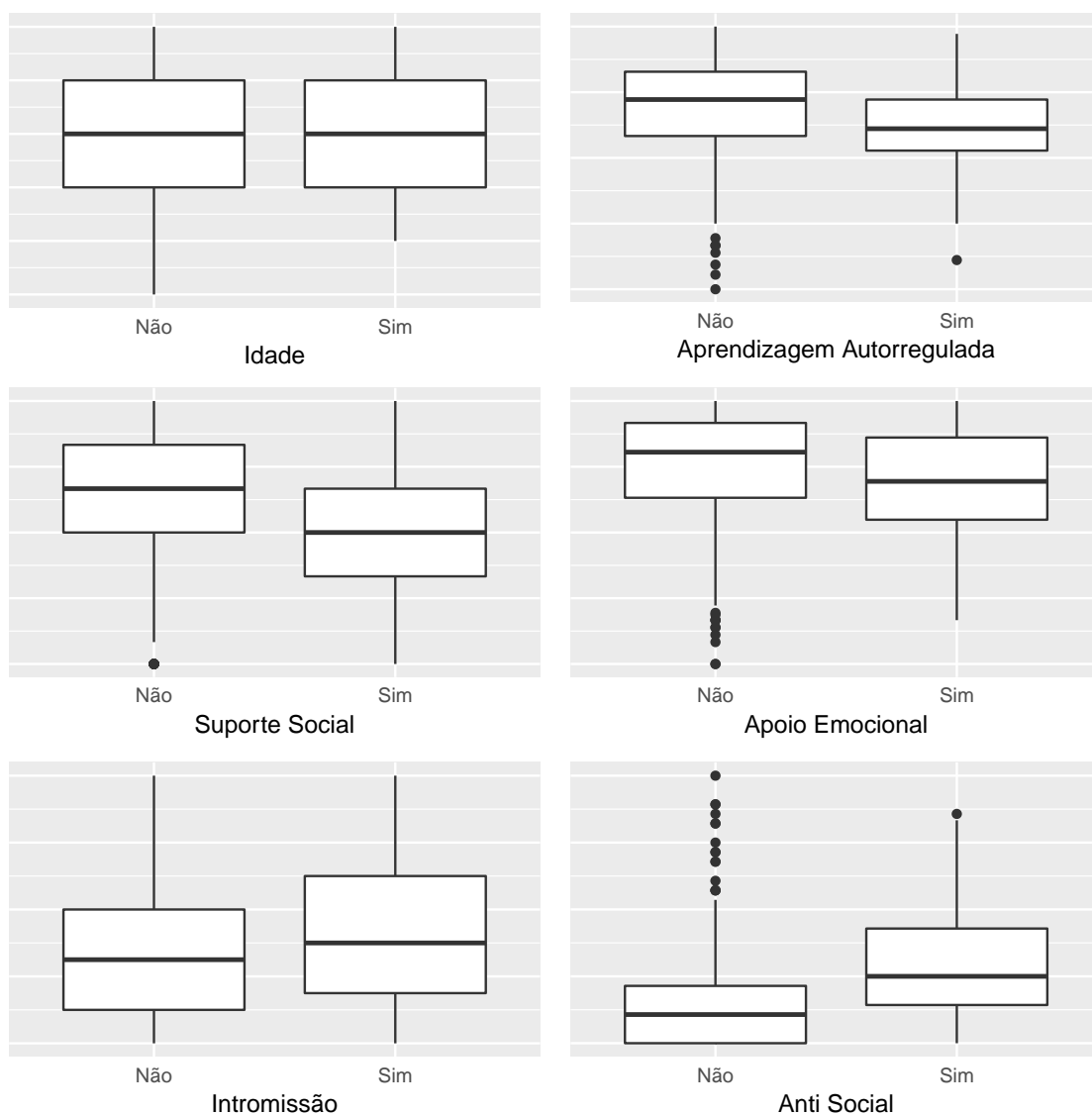
Tabela 4 – Frequências relativas das variáveis por uso de inalantes e teste Qui-quadrado

Características		Inalantes		Total	χ^2	p-valor
		Não	Sim			
Sexo	Menina	305 (80,4%)	74 (19,5%)	379	<0,001	>0,999
	Menino	246 (80,6%)	59 (19,3%)	305		
Ano Escolar	6º ano	162 (82,2%)	35 (17,7%)	197	2,397	0,494
	7º ano	195 (80,2%)	48 (19,7%)	243		
	8º ano	84 (75,6%)	27 (24,3%)	111		
	9º ano	109 (82,5%)	23 (17,4%)	132		
Bolsa Família	Não	322 (79,1%)	85 (20,8%)	407	1,271	0,529
	Sim	140 (82,3%)	30 (17,6%)	170		
	Não sabe	92 (82,8%)	19 (17,1%)	111		
Mora com os pais	Nenhum	32 (76,1%)	10 (23,8%)	42	0,515	0,772
	Apenas um	149 (80,5%)	36 (19,4%)	185		
	Com os dois	378 (80,7%)	90 (19,2%)	468		
Escolaridade da mãe	Não concluiu o ensino médio	133 (80,1%)	33 (19,8%)	166	11,067	0,003
	Terminou o ensino médio ou mais	263 (76,0%)	83 (24,0%)	346		
	Não sabe	147 (88,5%)	19 (11,4%)	166		

Na amostra utilizada no presente estudo verifica-se que o consumo de inalantes é diferente entre os alunos do 8º ano (24,3%) e os alunos do 9º ano (17,4%). No entanto, segundo o resultado do teste, expresso na Tabela 4, a variável ano escolar não mostrou ser

significativamente associada ao consumo de inalantes, dessa forma não se pode afirmar que existe diferença na população. O mesmo ocorre com as variáveis Bolsa Família e mora com os pais, apesar de terem diferenças nas porcentagens daqueles que já consumiram inalantes, o teste mostrou não ser uma diferença significativa. Apenas a variável escolaridade da mãe mostrou ser significativamente diferente. Percebe-se que o consumo é praticamente igual na variável sexo.

Figura 4 – Boxplot das variáveis em relação ao consumo de inalantes.



Percebe-se que visualmente há diferença entre os grupos das variáveis da Figura 4, exceto para a variável idade. O teste *t-student* foi utilizado para comparação de médias

entre os que consumiram inalantes e os que não consumiram (5). Comparando a idade com o consumo de inalantes obtivemos médias de 12,70 anos (não consumiram inalantes) e 12,79 (consumiram) e essas médias da população não foram consideradas significativamente diferentes. Na

Tabela 5 – Médias das variáveis por uso de inalantes e teste *t-student*

Característica	Variável	Uso de inalantes	Média	<i>T</i>	p-valor
Auto eficácia	Aprendizagem autorregulada	Não	3,78	4,95	<0,001
		Sim	3,49		
	Auto assertiva	Não	3,91	0,31	0,756
		Sim	3,88		
	Suporte social	Não	3,58	4,96	<0,001
		Sim	3,10		
Práticas parentais	Apoio emocional	Não	3,98	3,60	<0,001
		Sim	3,71		
	Intromissão	Não	2,41	-2,63	0,008
		Sim	2,68		
	Incentivo autonomia	Não	3,21	-1,59	0,111
		Sim	3,37		
	Supervisão	Não	4,07	0,84	0,397
		Sim	3,99		
	Cobrança responsabilidade	Não	4,32	-1,23	0,218
		Sim	4,41		
	Controle punitivo	Não	3,32	-0,10	0,913
		Sim	3,34		
Comportamento	Anti social	Não	1,55	-7,87	<0,001
		Sim	2,17		
Sociodemográfica	Idade	Não	12,70	-0,81	0,417
		Sim	12,79		

De acordo com a Tabela 5, os participantes que relatam não ter usado inalantes apresentaram aprendizagem autorregulada e suporte social significativamente maior, quando comparados aos que relataram uso. Por outro lado, considerando a auto assertiva, apesar de

apresentar maior média para os participantes que não relataram uso de inalantes, não houve diferenças significativas entre os dois grupos.

Os participantes que relataram não ter feito uso de inalantes apresentaram média de apoio emocional significativamente maior, quando comparados aos que relataram ter experimentado inalantes. Por outro lado, ter feito uso de inalantes, mostrou estar significativamente associado à maior média de intromissão. Incentivo à autonomia, supervisão do comportamento, cobrança de responsabilidade e controle punitivo não apresentaram diferenças significativas entre os grupos, sugerindo que essas práticas não influenciam no comportamento de experimentar ou não inalantes.

O comportamento anti social apresentou a maior diferença entre as médias, o que resulta em uma maior associação. Ou seja, aqueles que possuem um maior comportamento anti social tem mais chances de consumir inalantes.

4 RESULTADOS

Considerando as variáveis definidas na análise descritiva, foi realizado o ajuste do modelo para as variáveis respostas consumo de álcool e consumo de inalantes, separadamente. O modelo será construído a partir dos passos definidas por Hox (2010), que foram apresentados no referencial teórico deste trabalho. A variável reposta assume valor 1 quando houve o consumo de álcool ou de inalantes por parte do participante, além disso, os locais onde foram realizadas as coletas foram considerados como os agrupamentos.

4.1 Consumo de álcool

Passo 1

A Tabela 6 apresenta os resultados do modelo nulo.

Tabela 6 – Modelo nulo - Álcool

	Valor	Erro padrão	P-valor
Intercepto	-0,168	0,178	0,347
σ_u^2	0,290	-	-
AIC	934,7	-	-

A partir da Tabela 6 tem-se que a variância entre grupos σ_u^2 foi igual a 0,29, resultando em um coeficiente de correlação intra-classe dado por

$$\rho = \frac{\sigma_u^2}{(\sigma_u^2 + 3,29)} = \frac{0,29}{(0,29 + 3,29)} = 0,081. \quad (28)$$

Ou seja, 8,1% da variância total é explicada pelo local. Como esse valor indica a utilização da regressão logística multinível, seguiremos as passos de escolha do modelo.

Passo 2

Após o ajuste do modelo nulo, inicia-se o passo de seleção das variáveis explicativas no nível mais baixo, usando o teste de significância de Wald e o índice AIC. Com isso, o modelo final desse passo é descrito na Tabela 7.

Tabela 7 – Modelo passo 2 - Álcool

	Valor	Erro padrão	P-valor	Odds ratio	IC[95%]
Intercepto	-7,709	1,340	<0,001		
Idade	0,582	0,089	<0,001	1,790	[1,508;2,142]
Bolsa Família(sim)	-0,536	0,228	0,018	0,584	[0,372;0,913]
Bolsa Família(não sei)	-0,542	0,263	0,039	0,581	[0,345;0,971]
Aprendizagem autorregulada	-0,483	0,144	<0,001	0,616	[0,463;0,817]
Intromissão	0,193	0,090	0,033	1,213	[1,015;1,451]
Anti social	1,022	0,166	<0,001	2,780	[2,026;3,893]
σ_u^2	0,087	-	-		
AIC	736,7	-			

Nota-se que restaram apenas cinco variáveis: idade, Bolsa Família, aprendizagem autorregulada, intromissão e anti social. Além disso, todas são significativas devido ao p-valor ser inferior ao nível de significância de 5%. Em comparação com o modelo nulo, a variância do nível da turma teve uma redução, assim, o coeficiente de correlação intraclasse passou a ser igual a 2,5%. O índice AIC deste modelo é igual a 736,7, o que representa uma diminuição de 198 em relação ao anterior. Esse resultado significa que este modelo se ajusta melhor aos dados que o modelo nulo.

Passo 3

Analisa-se o modelo incluindo, separadamente, cada uma das variáveis explicativas fixas do nível mais agregado, que é o nível do grupo. Como nos dados estudados não havia variável no nível do grupo, as variáveis bolsa família e escolaridade da mãe foram agregadas. Entretanto, nenhuma das variáveis apresentou o coeficiente significativo, apesar da diminuição do índice AIC. Assim, o modelo permanece o mesmo definido no passo anterior.

Passos 4 e 5

No passo 4, avalia-se se algum dos coeficientes das variáveis explicativas tem um efeito aleatório significativo entre os grupos. No passo 5, adiciona-se interações de nível

cruzado entre variáveis explicativas em nível de grupo e aquelas de nível individual que apresentaram variação significativa da inclinação no passo 4. Verificou-se nessas etapas que não houve variáveis explicativas de efeito aleatório e efeito de interação cruzado significativos.

Tabela de classificação

A Tabela 8 apresenta a tabela de classificação com $\pi_0 = 0,50$.

Tabela 8 – Tabela de classificação

Resposta	Predito, $\pi_0 = 0.50$		Total
	Não	Sim	
Não	278	75	353
Sim	88	234	322

Com os dados da Tabela 8, obtemos a sensibilidade estimada $= 234/322 = 0,726$ e a especificidade $= 278/353 = 0,787$. Além disso, a proporção de classificações corretas é $(278 + 234) / 675 = 0,758$.

Interpretação

Assim, o modelo final é aquele definido no passo 3, Tabela 7. Como o coeficiente da variável aprendizagem autorregulada é negativo, quanto maior for essa variável menor a chance de consumir álcool. Como a variável Bolsa Família é qualitativa foram construídas duas variáveis dicotômicas e a referência foi não participar do Bolsa Família. As variáveis idade, prática parental relacionada a intromissão e comportamento anti social são diretamente proporcionais ao consumo de álcool, ou seja, quanto maior for essas variáveis maior será a chance de consumir álcool. Além disso, tem-se que a *odds ratio* para a variável comportamento anti social é de 2,780 (Tabela 7), com intervalo de 95% confiança de 2,026 a 3,893. Assim, controlando as outras variáveis, para cada unidade na variável anti social, a chance de consumir álcool aumenta em 2,780 vezes. Por exemplo, a chance de um adolescente com score de comportamento anti social igual a 3 consumir álcool é aproximadamente 2,780 vezes a chance de um adolescente com score de comportamento anti social igual a 2 consumir, controlando as outras variáveis. A interpretação das outras variáveis é análoga.

4.2 Consumo de inalantes

Passo 1

A Tabela 9 apresenta os resultados do modelo nulo, considerando o consumo de inalantes como resposta.

Tabela 9 – Modelo nulo - Inalantes

Modelo	Valor	Erro padrão	P-valor
Intercepto	-1,539	0,169	<0,001
σ_u^2	0,147	-	-
AIC	688,3	-	

Tem-se que a variância entre grupos σ_u^2 é igual a 0,147, resultando em um coeficiente de correlação intra-classe de 4,2%. Assim, a abordagem multinível não será considerada para essa droga devido ao ICC ser muito baixo. Neste caso deve-se utilizar o modelo de regressão logística múltipla, desconsiderando a estrutura hierárquica.

Regressão logística

A Tabela 10 apresenta o modelo logístico escolhido para a variável consumo de inalantes. Esse modelo foi obtido a partir do critério de seleção *stepwise*.

Tabela 10 – Modelo logístico - Inalantes

	Valor	Erro padrão	P-valor	Odds ratio	IC[95%]
Intercepto	-3,048	0,721	<0,001		
Suporte Social	-0,369	0,105	<0,001	0,691	[0,561;0,849]
Cobrança de responsabilidade	0,289	0,140	0,038	1,335	[1,023;1,774]
Anti social	0,889	0,130	<0,001	2,433	[1,891;3,159]
AIC	608,36	-			

Apenas as variáveis suporte social, cobrança de responsabilidade e comportamento anti social foram significativas, além disso o AIC foi de 608,336 (Tabela 10).

Tabela de classificação

A Tabela 11 apresenta a tabela de classificação do modelo logístico com $\pi_0 = 0,50$.

Tabela 11 – Tabela de classificação

Resposta	Predito, $\pi_0 = 0.50$		Total
	Não	Sim	
Não	534	14	548
Sim	118	17	135

Com os dados da 11, obtemos a sensibilidade estimada $= 17/135 = 0,125$ e especificidade $= 534/548 = 0,974$. Além disso, a proporção de classificações corretas é $(534 + 17) / 683 = 0,806$.

Interpretação

As variáveis cobrança de responsabilidade e comportamento anti social são diretamente proporcionais ao consumo de inalantes, ou seja, quanto maior for o nível dessas variáveis maior será a chance de consumir inalantes. Além disso, tem-se que a *odds ratio* para a variável anti social é de 2,433 (Tabela 10), com intervalo de 95% confiança de 1,891 a 3,159. Assim, controlando as outras variáveis, para cada unidade na variável comportamento anti social, a chance de consumir inalantes aumenta em 2,433 vezes. Por exemplo, a chance de um participante com score de comportamento anti social igual a 3 consumir inalantes é aproximadamente 2,58 vezes a chance de um participante com score de comportamento anti social igual a 2 consumir, controlando as outras variáveis. A interpretação das outras variáveis é análoga.

5 CONCLUSÃO

Com este trabalho, é possível perceber que o modelo de regressão logística multinível é adequado para modelar o consumo de álcool entre os participantes, pois 8,1% da variância total pode ser atribuída ao local que eles pertencem. No caso do consumo de inalantes, apenas 4,2% da variância total é devido ao local que eles pertencem. Logo, as características individuais dos participantes é que tendem a definir se eles consumiram inalantes. Por essa razão a abordagem multinível não é adequada para esse caso, assim, aplicou-se o método usual de regressão logística.

O consumo de álcool é explicado, principalmente, pelas variáveis idade, Bolsa Família, auto eficácia relacionada a aprendizagem autorregulada, prática parental relacionada a intromissão e comportamento anti social. Entretanto, variáveis de nível de grupo, efeitos aleatórios e interações não foram adicionadas. O modelo logístico multinível, neste caso, apresentou um bom ajuste nos dados, alcançando uma taxa de 75,8% de classificação correta.

O consumo de inalantes é explicado, principalmente, pelas variáveis auto eficácia relacionada ao suporte social, prática parental relacionada a cobrança de responsabilidade e comportamento anti social. Além disso, obteve uma taxa de classificação correta de 80,6%.

Uma sugestão é que seja dada continuidade a este trabalho quando todos os dados forem coletados, pois esse trabalho foi baseado na amostra piloto. Assim, será possível obter mais grupos e melhorar as estimativas dos modelos.

É importante destacar que os resultados desse trabalho não podem ser generalizados para toda a população brasileira, pois, a amostra é composta apenas de crianças e adolescentes de algumas escolas do Distrito Federal e da cidade de Goiânia.

REFERÊNCIAS

AGRESTI, A. (2007). **An Introduction to Categorical Data Analysis**. Second Edition. Hoboken, New Jersey: John Wiley & Sons.

BARROS, M.; BATISTA-DOS-SANTOS, A. C. **Por dentro da autoeficácia: um estudo sobre seus fundamentos teóricos, suas fontes e conceitos correlatos**. Revista Espaço Acadêmico, v.10, n. 112, p. 1-9, Setembro, 2010.

BRINGAS, C., HERRERO, F. J., CUESTA, M. & RPDRÍGUEZ, F. J. (2006). **La conducta antisocial en adolescentes no conflictivos: Adaptación del Inventário de Conductas Antisociales (ICA)**. Revista Eletrónica de Metodología Aplicada, 11(2), 1-10.

CARLINI, E. L., NOTO, A. R., SANCHEZ, Z. M., CARLINI-COTRIM, M., LOCATELLI, D. P., ABEID, L. R., . MOURA, Y. G. (2010). **VI Levantamento Nacional sobre o Consumo de Drogas Psicotrópicas entre Estudantes do Ensino Fundamental e Médio das Redes Pública e Privada de Ensino nas 27 Capitais Brasileiras**. Retrieved from <http://www.obid.senad.gov.br>

DUARTE, P. C. A. V., STEMPLIUK, V. A., & BARROSO, L. P. (2009). **Relatório Brasileiro sobre Drogas**. Brasília, DF: Secretaria Nacional de Políticas sobre Drogas.

HASTINGS, N.; PEACOCK, B.; EVANS, M.; FORBES, C.(2011). **Statistical Distributions**. Fourth Edition. Hoboken, New Jersey: John Wiley & Sons.

HEIM, J., & ANDRADE, A. G. (2008). **Efeitos do uso do álcool e das drogas ilícitas no comportamento de adolescentes de risco: uma revisão das publicações científicas entre 1997 e 2007**. Rev. Pesiq. Clín., 35(supl 1), p. 61-64.

HOX, J. J.(2010). **Multilevel analysis: techniques and applications**. Second Edition. Great Britain: Routledge.

LAROS, J. A.; MARCIANO, J.L.P.(2008). **Análise multinível aplicada a dados do NELS:88**. Estudos em Avaliação Educacional, v. 19, p. 263-278.

MALTA, D. C., PORTO, D. L., MELO, F. C. M., MONTEIRO, R. A., SARDINHA, L. M. V., & LESSA, B. H. (2011). **Família e proteção ao uso de tabaco, álcool e drogas em adolescentes, Pesquisa Nacional de Saúde dos Escolares**. Rev Bras Epidemiol, 14 (1) supl, p. 166.

MURTA, S. G., NOBRE-SANDOVAL, L. A., IGLESIAS, F., VINHA, L. G. A., ROCHA, V. P., MIRANDA, A. A. V., SOUZA, A. S. S. (2015). **Avaliação da Efetividade, Qualidade de Implementação e Validade Social do Programa Fortalecendo Famílias**. Brasília, Universidade de Brasília: Relatório Não Publicado.

TEIXEIRA, M. A. P., OLIVEIRA, A. M., WOTTRICH, S. H. (2006). **de Práticas Parentais (EPP): Avaliando dimensões de práticas parentais em relação a adolescentes**. Psicologia: Reflexão e Crítica, 19 (3), 433-441.