



**UFSM**

**Monografia de Especialização**

**ESTUDO SOBRE O CONSUMO DE CIGARROS EM  
ESCOLAS ESTADUAIS DE SANTA MARIA – RS**

---

**Ana Cláudia Antunes Trindade**

**CEEMQ**

**Santa Maria, RS, Brasil**

**2004**

**ESTUDO SOBRE O CONSUMO DE CIGARROS EM  
ESCOLAS ESTADUAIS DE SANTA MARIA – RS**

---

**por**

**Ana Cláudia Antunes Trindade**

Monografia apresentada ao Curso de Especialização em  
Estatística e Modelagem Quantitativa, da Universidade Federal de  
Santa Maria (UFSM, RS), como requisito parcial para obtenção  
do grau de  
**Especialista em Estatística e Modelagem Quantitativa.**

**CEEMQ**

**Santa Maria, RS, Brasil**

**2004**

**Universidade Federal de Santa Maria  
Centro de Ciências Naturais e Exatas  
Curso de Pós-Graduação em Estatística e Modelagem Quantitativa**

A Comissão Examinadora, abaixo assinada,  
aprova a Monografia de Especialização

**ESTUDO SOBRE O CONSUMO DE CIGARROS EM  
ESCOLAS ESTADUAIS DE SANTA MARIA – RS**

elaborada por  
**Ana Cláudia Antunes Trindade**

como requisito parcial para obtenção do grau de  
**Especialista em Estatística e Modelagem Quantitativa**

**COMISSÃO EXAMINADORA:**

---

**Roselaine Ruviaro Zanini**  
(Presidente/Orientador)

---

**Anaelena Bragança de Moraes Ethur**

---

**Danielle de Souza Saad**

Santa Maria, 29 de dezembro de 2004

## **AGRADECIMENTOS**

Ao término desta monografia, gostaria de registrar meus agradecimentos a todas as pessoas que colaboraram na sua construção:

À orientadora Roselaine Ruviano Zanini, pela amizade, apoio, disponibilidade e dedicação no desenvolvimento do trabalho.

A minha mãe e Luciane, pela compreensão e carinho.

A todos aqueles que, mesmo não tendo seus nomes citados, estiveram presentes da sua maneira, acreditando na realização deste trabalho.

## **SUMÁRIO**

<b>LISTA DE SIGLAS</b> .....	
<b>LISTA DE FIGURAS</b> .....	v
<b>LISTA DE TABELAS</b> .....	vii
<b>LISTA DE ANEXOS</b> .....	ix
<b>LISTA DE APÊNDICES</b> .....	viii
<b>RESUMO</b> .....	ix
<b>ABSTRACT</b> .....	xii
<b>1 INTRODUÇÃO</b> .....	01
1.1 Objetivos.....	04
1.1.1 Objetivo geral.....	04
1.1.2 Objetivos específicos.....	04
1.2 Justificativa.....	04
1.3 Estrutura do trabalho.....	05
<b>2 REVISÃO DE LITERATURA</b> .....	07
<b>3 CONCEITOS BÁSICOS</b> .....	14
<b>4 REGRESSÃO LOGÍSTICA</b> .....	19
<b>5 MATERIAL E MÉTODO</b> ...	32
<b>6 RESULTADOS</b> .....	35
<b>7 DISCUSSÃO</b> .....	56
<b>8 CONCLUSÃO</b> .....	59
<b>6 BIBLIOGRAFIA</b> .....	61
<b>APÊNDICE</b> .....	65

## LISTA DE SIGLAS

CEBRID – Centro Brasileiro de Informações Sobre Drogas Psicotrópicas;

EMV – Estimação de Máxima Verossimilhança;

EPA – Agência para Proteção do Meio Ambiente;

IC – Intervalo de Confiança;

INCA – Instituto Nacional do Câncer;

OR – Razão de *Odds*;

PNSN – Pesquisa Nacional sobre Saúde e Nutrição;

RR – Risco Relativo;

SE – Erro Padrão;

SQE – Soma de Quadrado dos Erros.

## LISTA DE FIGURAS

FIGURA 01 – Distribuição 2 x 2 .....	16
FIGURA 02 – Representação da função e do logit .....	26
FIGURA 03 – Motivos para fumar relatados peloos estudantes das escolas estaduais de Santa Maria – RS, em 2002 .....	41

## LISTA DE TABELAS

- TABELA 01 – Distribuição dos estudantes do ensino médio das escolas estaduais de Santa Maria – RS em 2002, em relação às características investigadas .....
- TABELA 02 – Prevalência de tabagismo entre estudantes do ensino médio das escolas estaduais de Santa Maria – RS, em 2002, em relação às características investigadas .....
- TABELA 03 – Prevalência de tabagismo entre estudantes do ensino médio das escolas estaduais de Santa Maria – RS, em 2002, em relação aos hábitos de fumar .....
- TABELA 04 – Distribuição dos estudantes do ensino médio das escolas estaduais de Santa Maria – RS em 2002, em relação ao hábito de fumar .....
- TABELA 05 – Variáveis candidatas a entrar no modelo de Regressão Logística multivariável ( $p < 25\%$ ) .....
- TABELA 06 – Resultados da Regressão Logística multivariável: modelo A .....
- TABELA 07 – Resultados da Regressão Logística multivariável: modelo B .....
- TABELA 08 – Resultados da Regressão Logística multivariável: modelo C .....
- TABELA 09 – Resultados da Regressão Logística multivariável: modelo D .....
- TABELA 10 – Resultados da Regressão Logística multivariável: modelo E .....
- TABELA 11 – Resultados da Regressão Logística multivariável: modelo F .....
- TABELA 12 – Resultados da Regressão Logística multivariável: modelo G .....
- TABELA 13 – Resultados da Regressão Logística multivariável: modelo H .....
- TABELA 14 – Resultados da Regressão Logística multivariável: modelo I .....
- TABELA 15 – Resultados das interações entre as variáveis .....



## **LISTA DE APÊNDICES**

APÊNDICE 01 –	Questionário respondido pelos estudantes .....	66
---------------	--	----

## **RESUMO**

**Monografia de Especialização  
Programa de Pós-Graduação em Estatística e Modelagem Quantitativa  
Universidade Federal de Santa Maria, RS, Brasil**

### **ESTUDO SOBRE O CONSUMO DE CIGARROS EM ESCOLAS ESTADUAIS DE SANTA MARIA – RS**

**AUTORA: ANA CLÁUDIA ANTUNES TRINDADE**

**ORIENTADORA: ROSELAINÉ RUVIARO ZANINI**

**CO-ORIENTADORA: ANAELENA BRAGANÇA DE MORAES ETHUR**

**Data e Local da Defesa: Santa Maria, 29 de dezembro de 2004.**

O consumo de cigarro é considerado a causa mais devastadora e evitável das doenças da história da humanidade, com muitas mortes prematuras. Segundo a Organização Mundial da Saúde, estima-se que existam, no mundo, cerca de 1,2 bilhão de fumantes. A expectativa é de que, em duas décadas, ocorram 10 milhões de mortes por ano. É importante priorizar as pesquisas voltadas aos jovens e adolescentes, pois é nesta fase que estes adquirem o hábito de fumar. Assim, em 2002, realizou-se um estudo transversal entre os estudantes de 2º grau da rede estadual de ensino de Santa Maria – RS, para determinar a prevalência e os fatores de risco associados ao tabagismo, obtendo um modelo logístico que descreve como as chances de ser fumante estão relacionadas com as variáveis investigadas. Foram entrevistados 459 alunos, observando-se uma prevalência de 18% para o tabagismo (consumo de cigarros), IC 95% (14,6% – 21,7%), sendo maior no sexo feminino (20,4%) do que no masculino (16,7%), apesar da diferença não-significativa ( $p = 0,666$ ). Os fatores de risco foram “se tem amigos fumantes”, OR = 4,370 ( $p = 0,000$ ); “renda familiar mensal, em salários mínimos”, OR = 2,038 ( $p = 0,013$ ) e “idade”, OR = 1,864 ( $p = 0,031$ ). Um estudante com mais de 19 anos, com renda familiar de 7 ou mais salários mínimos e que tem amigos fumantes apresentou a maior probabilidade de ser fumante. Os resultados deste estudo mostram que os estudantes começam a fumar precocemente, destacando-se a necessidade de se trabalhar, em termos de prevenção intensiva, diretamente no grupo de risco observado.

## **ABSTRACT**

**Monografia de Especialização  
Programa de Pós-Graduação em Estatística e Modelagem Quantitativa  
Universidade Federal de Santa Maria**

### **ESTUDO SOBRE O CONSUMO DE CIGARROS EM ESCOLAS ESTADUAIS DE SANTA MARIA – RS**

**AUTHOR: ANA CLÁUDIA ANTUNES TRINDADE**

**ADVISOR: ROSELAINÉ RUVIARO ZANINI**

**CO-ADVISOR: ANAELENA BRAGANÇA DE MORAES ETHUR**

**Date and place of defense: December, 2004, Santa Maria – RS.**

The use of cigarettes is one of the main causes of preventable diseases. A prevalence study on smoking habits was carried out in 2002 among 459 students from eight public high schools in Santa Maria, in the state of Rio Grande do Sul, Brazil. All adolescents were interviewed separately using a confidential questionnaire. The study was aimed at measuring smoking prevalence and related risk factors among adolescents. The overall smoking prevalence rate in the sample was 18% (95% CI 14,6% - 21,7%). The risk factors for smoking found in the multivariate logistic regression analysis were: “smoking friends” (OR = 4,37; p = 0,000), “family income” (OR = 2,04; p = 0,013) and “age” (OR = 1,86; p = 0,031). The results displayed that adolescents had begun the smoking habit very early at a mean age of 14 years. Its very important adopts actions in order to prevent adolescents to have access to smoking.

# 1 INTRODUÇÃO

Os prejuízos causados à saúde pelo consumo de tabaco são amplamente conhecidos, sendo o seu controle, considerado pela Organização Mundial da Saúde, como um dos maiores desafios da saúde pública no mundo atual (Álcool e Drogas, 2004).

O tabaco pode ser usado de diversas maneiras, de acordo com sua forma de apresentação: inalado (cigarro, cachimbo, charuto, cigarro de palha), aspirado (rapé) e mascado (fumo-de-rolô), porém sob todas as formas ele é maléfico à saúde (Medjunior, 2004).

O consumo de cigarro é tido como a causa mais devastadora e, ao mesmo tempo, evitável de doenças e mortes prematuras da história da humanidade (RADIOBRAS, 2002).

Segundo a Organização Mundial da Saúde, estima-se que existam no mundo cerca de 1,2 bilhão de fumantes (incluindo 200 milhões de mulheres), ou seja, um terço da população mundial com mais de 15 anos de idade fuma. Além disso, todos os anos, 4 milhões de pessoas morrem em razão de doenças relacionadas com o fumo. A expectativa é de que, em duas décadas, esse número atingirá 10 milhões de mortes/ano. Em 2020, o fumo vai matar mais gente do que a Aids, tuberculose, acidentes de carro, suicídios e homicídios somados (A Notícia, 2001).

O risco de surgir um tumor pulmonar para quem fuma 20 cigarros por dia é 24 vezes a de um não-fumante e 90% dos cânceres estão associados ao tabagismo (INCA, 2004).

No Brasil existem cerca de 30 milhões de fumantes, ou seja, um terço da população adulta fuma, sendo 11,2 milhões de mulheres e 16,7 milhões

de homens e, destes, 300 morrem por dia, vítimas de doenças provocadas pelo fumo. O total de mortes anuais, devido ao tabagismo, atingiu a marca de 4 milhões, ou seja, mais de 10 mil mortes por dia (INCA, 2004).

Noventa por cento dos fumantes ficam dependentes da nicotina entre os 5 e os 19 anos de idade. Atualmente, existem no país 2,8 milhões de fumantes nessa faixa etária. A maioria dos fumantes tem entre 20 e 49 anos de idade. Os homens fumam em maior proporção que as mulheres em todas as faixas etárias. Porém, a mulher vem aumentando sua participação no número de fumantes, sobretudo na faixa etária mais jovem.

O uso inicial de tabaco é bastante precoce na vida dos estudantes da rede pública de ensino, sendo que, entre os 10 e 12 anos de idade cerca de 11,6% já fizeram pelo menos uso experimental do cigarro, de acordo com o estudo realizado pelo Centro Brasileiro de Informações sobre Drogas Psicotrópicas (CEBRID). Nesse mesmo trabalho foi observado que 6% dos adolescentes consomem cigarros freqüentemente (INCA, 2004).

Fuma-se mais no Rio Grande do Sul, sendo Porto Alegre a detentora dos maiores índices conhecidos de câncer de pulmão no país, inclusive entre as mulheres. Na capital gaúcha, a prevalência entre os estudantes é 36% para o sexo masculino e 47% para o feminino, sendo que 90% iniciaram a fumar antes dos 19 anos (A Razão, 2004).

Nas últimas décadas, tem-se verificado uma preocupação marcante de vários setores nacionais no sentido de dotar o Brasil de dados científicos sobre o consumo de cigarros entre estudantes das escolas brasileiras (Doll, 1986).

São inúmeras as pesquisas feitas envolvendo o uso do tabaco e as conseqüências que este traz diretamente (aos que consomem) e

indiretamente (àqueles que convivem em ambientes onde o seu uso é comum), pois cada cigarro contém 4.700 substâncias tóxicas, sendo que somente a fumaça contém milhares de substâncias tóxicas irritantes, cancerígenas (como o benzopireno) e radioativas (CCOO, 2004). Deve-se ressaltar, que nenhuma poluição industrial é comparada àquela produzida pela fumaça do cigarro.

Unidades governamentais, familiares, entidades sociais e meios de comunicação estão, constantemente, envolvidos na luta a favor da diminuição do consumo cigarros, devido aos diversos males provocados.

Os principais fatores de risco, encontrados na literatura, que levam ao hábito de fumar são: sexo, idade, nível socioeconômico, tabagismo de familiares de primeiro grau e dos amigos, rendimento escolar, separação dos pais e trabalho (Malcon, 2003).

Portanto, é importante que se dê enfoque aos estudos feitos com jovens e adolescentes, buscando identificar as variáveis associadas com o objetivo de compreender este fenômeno, pois é, principalmente, nesta fase que estes experimentam o tabaco e adquirem o hábito de fumar.

Alguns recursos importantes, disponíveis para análise destas variáveis são: a Estatística Descritiva, a metodologia de Regressão Logística, a qual permite o tratamento de variáveis binárias, além de Técnicas Multivariadas, que são de grande importância para identificar combinações de categorias de variáveis, as quais caracterizam uma situação de risco do estudante ser fumante.

## **1. 1 Objetivos**

### 1. 1. 1 Objetivo geral

Desenvolver um estudo de base escolar entre estudantes de 2<sup>o</sup> grau de escolas estaduais, em Santa Maria – RS, para determinar a prevalência e os fatores de risco associados ao tabagismo.

### 1. 1. 2 Objetivos específicos

- estudar o consumo de cigarros por estudantes de 2<sup>o</sup> grau, em escolas estaduais de Santa Maria;
- determinar a idade na qual os estudantes começaram a fumar;
- descrever o consumo de cigarros;
- verificar as variáveis associadas ao consumo de cigarros;
- avaliar a associação do tabagismo familiar e dos amigos com o tabagismo dos estudantes;
- encontrar um modelo que descreva como a chance de ser fumante está relacionada com as variáveis investigadas;
- desenvolver os conceitos básicos sobre a técnica de modelagem de Regressão Logística.

## **1. 2 Justificativa**

Os males causados pelo consumo de cigarros são amplamente divulgados, porém o número de pessoas que fumam, ainda, é muito grande.

Muitas campanhas de conscientização e prevenção são realizadas em função deste hábito, combatido incessantemente pelos órgãos de saúde. Para um planejamento de ações eficientes é necessário que se conheça o perfil dos consumidores de cigarros. Assim, este trabalho busca levantar informações sobre este tema, analisando variáveis que envolvem o consumo de cigarros, principalmente entre os estudantes jovens e adolescentes, pois é nesta fase da vida que eles adquirem esse hábito.

Também, neste estudo, destaca-se o desenvolvimento da metodologia de Regressão Logística, que se faz importante sob vários aspectos, já que pode ser aplicada num grande número de casos, nas mais diversas áreas, principalmente, nas ciências sociais, humanas e da saúde, proporcionando simplicidade e facilidade no tratamento numérico dos dados.

### **1.3 Estrutura do trabalho**

Para atingir os objetivos propostos, este trabalho foi dividido nos seguintes capítulos.

O Capítulo 1 apresenta uma introdução ao tema e o Capítulo 2 mostra uma Revisão de Literatura, abordando os principais trabalhos e resultados obtidos sobre o consumo de cigarros.

No Capítulo 3 são apresentados alguns conceitos básicos utilizados para a análise da amostra estudada e o Capítulo 4 apresenta uma fundamentação teórica sobre Regressão Logística Binária.

No Capítulo 5 estão descritos o Material e Método, empregados no levantamento e na análise dos dados obtidos.



No Capítulo 6 são apresentados os resultados e no Capítulo 7 a discussão dos mesmos, comparando-os com os encontrados na literatura em geral.

O Capítulo 8 descreve as conclusões deste estudo. No Capítulo 9 são apresentadas as Bibliografias referenciadas nesta pesquisa.

## **2 REVISÃO DE LITERATURA**

Um estudo realizado em 1993 pela Agência para Proteção do Meio Ambiente (EPA, em inglês) concluiu que a fumaça de cigarro no ambiente é um carcinógeno do Grupo A, o mais perigoso. O relatório analisou exaustivamente os resultados de 30 estudos da fumaça produzidas pelo cigarro em repouso e da fumaça expelida depois de tragada.

A EPA diz que a inalação passiva da fumaça de cigarro é responsável pelo câncer de pulmão que mata 3.000 pessoas todo o ano nos Estados Unidos. A Associação Médica Americana confirmou essas conclusões, em junho de 1994, com a publicação de um estudo que revela que as mulheres que nunca fumaram, mas que inalam fumaça de cigarro no ambiente, correm um risco 30% maior de contrair câncer de pulmão do que outras pessoas que também nunca fumaram.

No caso das crianças pequenas, a fumaça de cigarro resulta em 150.000 a 300.000 casos anuais de bronquite e pneumonia. A fumaça agrava os sintomas de asma em 200.000 a 1.000.000 de crianças todo o ano nos Estados Unidos. A Associação Cardíaca Americana calcula que ocorram, todos os anos, 40.000 mortes por doenças cardiovasculares causadas pela fumaça de cigarro no ambiente (Fumo, 2004).

No Brasil foi desenvolvido um estudo de base populacional – a Pesquisa Nacional sobre Saúde e Nutrição (PNSN) – que teve como objetivo central a apuração de indicadores da situação nutricional da população brasileira, na qual, foi encontrada uma proporção de 1,1% de fumantes na idade de 10 a 14 anos e 15,6% na idade de 15 a 19 anos na zona urbana (INAN, 1989).

Barbosa et al, em 1989, analisaram o uso de tabaco em estudantes de 1º e 2º graus em 10 capitais brasileiras – Belém, Brasília, Belo Horizonte, Curitiba, Fortaleza, Porto Alegre, Recife, Rio de Janeiro, Salvador, São Paulo (1987), com um total de 16.149 estudantes da rede estadual. A prevalência do uso de tabaco na vida foi 19,5% e, no último mês, 10,5%. Os alunos que freqüentam o curso noturno, têm pais que fumam, estão defasados nos estudos, trabalham e são maiores de 15 anos apresentaram chance maior de serem fumantes quando comparados aos colegas.

Analisando uma amostra de 1.918 alunos de 1º e 2º graus, regularmente matriculados em escolas públicas e particulares da zona urbana da cidade de Araraquara, São Paulo, foi encontrada uma proporção de 6,7% de fumantes. Foi constatado que a maior prevalência de fumantes quanto ao início do hábito de fumar, esteve na faixa etária de 12 a 14 anos, com 46,8% do sexo feminino e 43,5% do sexo masculino. Entre os que fumavam mais de 16 cigarros por dia, a maior percentagem (17,7%) foi para os meninos do que para as meninas (7,8%) (Simões, 1990).

Um estudo transversal encontrou a prevalência de fumantes de 8,4% entre estudantes de 2º grau, da zona urbana de Porto Alegre, para a faixa de idade de 14 a 20 anos. O critério para ser fumante, neste estudo, foi ter fumado um cigarro por dia nos últimos seis meses (Schio et al., 1992).

Com o objetivo de avaliar a prevalência de tabagismo em Porto Alegre, RS e identificar os fatores associados foi realizado um estudo observacional, de delineamento transversal e base populacional, por amostragem aleatória proporcional em estágios múltiplos e conglomerados. Foram selecionados 1.091 indivíduos, a partir de 18 anos, que responderam a um questionário, em entrevista domiciliar. Aferiu-se o hábito de fumar

através de perguntas dirigidas ao tipo, frequência e tempo de exposição ao fumo. A prevalência geral foi de 34,9%, sendo de 41,5% entre os homens e 29,5% entre as mulheres. Analisaram-se as associações através de regressão logística, incluindo-se no modelo sexo, idade, educação, renda, qualificação profissional e consumo de álcool. O hábito de fumar foi mais freqüente entre os homens, os indivíduos de menor nível socioeconômico, na faixa de 30 a 39 anos e entre os usuários de bebidas alcoólicas. Concluiu-se que o tabagismo é freqüente em Porto Alegre, constituindo-se problema de saúde pública similar ao referido pela literatura (Lolio et al, 1993).

Foi estudada uma amostra aleatória representativa e proporcional de 2.967 escolares de educação básica e média, no Chile, de acordo com o grau de estudo, tipo de escola, sexo e área geográfica na qual residem. A prevalência de tabagismo encontrada (10,6%) aumentou, significativamente, com a idade de 1,3% nos escolares menores de 13 anos a 15,4 a 36,9% nos adolescentes entre 13 a 15 anos e com 16 anos ou mais, respectivamente ( $p < 0,001$ ), sendo maior nos escolares de sexo feminino. O impacto relativo às variáveis independentes sobre o tabagismo foi diferente de acordo com a idade, sexo e nível socioeconômico (Ivanovic et al., 1997).

Em 1997, realizou-se um estudo transversal, de base populacional, com uma amostragem em múltiplos estágios dos adolescentes com idade entre 12 e 18 anos completos, residentes na zona urbana da cidade de Pelotas, RS. O objetivo foi descrever a prevalência e os fatores de risco para o tabagismo. Foram entrevistados 632 adolescentes, sendo que 11,1% destes eram fumantes, 6,8% eram ex-fumantes e 82,1% nunca haviam fumado regularmente. A prevalência de tabagismo foi diretamente

relacionada com a idade do adolescente. Mesmo após controle para possíveis fatores de confusão, aqueles adolescentes que não estavam estudando, que eram repetentes, cujos pais estavam separados ou que relataram terem abusado de bebidas alcoólicas no último mês apresentaram uma maior razão de *odds* para tabagismo (Horta, 2001).

A partir de um delineamento transversal de base populacional, com o objetivo de medir a prevalência e estudar fatores de risco associados ao tabagismo nos adolescentes, foi estudada uma amostra de 1.187 adolescentes de 10 a 19 anos, da zona urbana de Pelotas, no sul do Brasil. Todos os adolescentes da amostra, em cada domicílio, foram entrevistados por meio de questionário pré-codificado, individual e confidencial. A prevalência de tabagismo na amostra foi de 12,1%. As prevalências foram similares para os sexos femininos e masculinos. Os fatores de risco identificados por regressão logística, foram: maior idade, OR = 28,7, com IC 95% (11,5 – 71,4), irmãos mais velhos fumantes, OR = 2,4, com IC 95% (1,5 – 3,8), três ou mais amigos fumantes, OR = 17,5, com IC 95% (8,8 – 34,8) e baixa escolaridade, OR = 3,5, com IC 95% (1,5 – 8,0). A prevalência de tabagismo na adolescência mostrou-se alta na cidade de Pelotas (Malcon et al., 2003).

Foi desenvolvido um estudo do tipo corte transversal de caráter exploratório, com o objetivo de estimar a prevalência do tabagismo entre os adolescentes da 8ª série do ensino fundamental à 3ª série do curso médio, em escolas de Salvador, nos quais foram aplicados 3.500 questionários em alunos matriculados em cinco escolas da região metropolitana. A prevalência do tabagismo encontrada foi de 9,6%, considerando-se os 3.180 questionários válidos, sendo maior no sexo masculino (14%) do que

no feminino (6%). À medida que aumentava a idade, elevava-se essa prevalência.

A média de idade do início do tabagismo foi de  $14 \pm 2$  anos. Dentre adolescentes, 46% experimentaram o cigarro e 20% destes continuaram fumando. A experimentação e a influência dos pais foram associadas ao tabagismo nos adolescentes (Machado-Neto, 2003).

Um levantamento feito pela equipe do pneumologista brasileiro José Rosember, avaliou os efeitos do tabagismo na saúde de 15 mil crianças entre zero e 1 ano. Nas famílias nas quais o pai fumava, cerca de 25% das crianças apresentaram problemas respiratórios. Quando a mãe era fumante o número passava para 49%, já que esta tem mais contato com os filhos (Fumo, 2004).

Algumas constatações encontradas em cerca de 50.000 estudos e que preocupam os pesquisadores com relação ao fumo e à saúde (Fumo, 2004) foram:

- **câncer de pulmão:** 87% das mortes por câncer de pulmão ocorrem entre os fumantes;
- **doenças cardíacas:** os fumantes correm um risco 70% maior de apresentar doenças cardíacas;
- **câncer de mama:** as mulheres que fumam 40 ou mais cigarros por dia têm uma probabilidade 74% maior de morrer de câncer de mama,
- **deficiências auditivas:** os bebês de mulheres fumantes têm maiores dificuldades em processar sons;
- **complicações da diabetes:** os diabéticos que fumam ou que mascam tabaco correm maior risco de ter graves complicações renais e apresentam retinopatia (distúrbios da retina) de evoluções mais rápidas;

- **câncer de cólon:** dois estudos com mais de 150.000 pessoas mostram uma relação clara entre o fumo e o câncer de cólon;
- **asma:** a fumaça pode piorar a asma em crianças;
- **predisposição ao fumo:** as filhas de mulheres que fumaram durante a gravidez têm quatro vezes mais probabilidade de fumar também;
- **leucemia:** suspeita-se que o fumo cause leucemia mielóide;
- **contusões em atividades físicas:** segundo um estudo do Exército dos Estados Unidos, os fumantes têm maior probabilidade de sofrer contusões em atividades físicas;
- **memória:** doses altas de nicotina podem reduzir a destreza mental em tarefas complexas;
- **depressão:** psiquiatras estão investigando evidências de que há uma relação entre o fumo e a depressão profunda, além da esquizofrenia;
- **suicídio:** um estudo feito entre enfermeiras mostrou que a probabilidade de cometer suicídio era duas vezes maior entre as enfermeiras que fumavam;
- **outros perigos a acrescentar à lista:** câncer da boca, laringe, garganta, esôfago, pâncreas, estômago, intestino delgado, bexiga, rins e colo do útero; derrame cerebral, ataque cardíaco, doenças pulmonares crônicas, distúrbios circulares, úlceras pépticas, diabetes, infertilidade, bebês com baixo peso, osteoporose e infecções dos ouvidos. Pode-se acrescentar ainda o perigo de incêndios, já que o fumo é a principal causa de incêndios em residências, hotéis e hospitais.

Segundo pesquisas científicas (Fumo, 2004), os principais benefícios adquiridos ao parar de fumar são:

- vinte minutos depois de deixar o cigarro a pressão arterial e os batimentos cardíacos retornam ao normal;
- um dia depois de largar o vício as chances de infarto começam a reduzir;
- após três dias há um aumento da capacidade respiratória;
- de 2 a 12 semanas a circulação sanguínea melhora;
- no intervalo de 1 a 9 meses a tosse e as infecções das vias aéreas vão cessando, melhorando a capacidade física;
- em um ano diminui em 50% o risco de doença coronariana. Em dez anos caem as chances do aparecimento de câncer;
- no período de 10 a 15 anos o perigo de desenvolver problemas cardíacos se iguala ao de uma pessoa que nunca fumou.



## 3 CONCEITOS BÁSICOS

Neste capítulo serão apresentados alguns conceitos estatísticos que foram utilizados para descrever a amostra em estudo.

### 3.1 Tipos de variáveis

Mensuração de variáveis é o processo que consiste em atribuir números a objetos ou observações. A teoria da mensuração consiste em um conjunto de teorias distintas ou separadas, cada uma delas relacionada com um nível distinto de mensuração (Siegel, 1975). As variáveis mensuradas podem ser classificadas em:

**a) Variáveis quantitativas:** são aquelas cujos valores numéricos expressam quantidades, por exemplo, estatura das pessoas, número de filhos, etc. São separadas em variáveis quantitativas discretas ou contínuas. As **contínuas** têm intervalos quantificados em uma escala infinita de valores. O número de valores possíveis para o peso corporal, por exemplo, é apenas limitado pela sensibilidade do equipamento usado para a medição. Diz-se que as variáveis contínuas são muito informativas. As **discretas** têm uma escala com um número finito de intervalos, por exemplo, número de cigarros fumados por dia ou por mês.

**b) Variáveis qualitativas** (categóricas ou de atributos): fenômenos que não são facilmente quantificáveis podem ser classificados em categorias. Variáveis categóricas que assumem apenas dois possíveis valores (fumante ou não-fumante) são denominadas **dicotômicas**. As que podem assumir mais de duas categorias são classificadas nominais ou ordinais. As

**nominais** se apresentam em categorias não-ordenadas; por exemplo, o sangue tipo 0 não é mais nem menos que o tipo B. Estas variáveis têm um caráter qualitativo e absoluto que as torna simples de medir. Já as **ordinais**, são distribuídas em categorias ordenadas, como por exemplo, dores fortes, moderadas ou leves.

### 3.2 Níveis de mensuração das variáveis

Nível de mensuração corresponde à escala na qual foi medida a variável investigada. O nível de mensuração de cada variável nos dados usados é a mais elementar informação que um pesquisador precisa ter antes de selecionar a técnica estatística que será aplicada ao conjunto de dados.

Os quatro níveis de mensuração são definidos como nominal, ordinal, intervalar e de razão:

**a) Escala nominal ou classificadora:** a mensuração em seu mais baixo nível existe quando números ou outros símbolos são usados para classificar um objeto, pessoa ou característica. Quando números ou outros símbolos são usados para identificar os grupos a que vários objetos pertencem, esses números ou símbolos constituem uma escala nominal ou classificadora. Então, consiste na contagem ou enumeração de uma variável em suas diversas categorias, as quais são mutuamente exclusivas.

**b) Escala ordinal ou escala por postos:** pode ocorrer que os elementos em uma categoria de cada escala não sejam apenas diferentes dos elementos em outras categorias da mesma escala, mas que guardem certo tipo de relação entre eles. Isto é, a variável em estudo é partida em categorias

ordenadas em grau convencionados havendo uma relação entre categorias do tipo maior do que.

**c) Escala intervalar:** quando a escala tem todas as características de uma escala ordinal e, além disso, se conhecem as distâncias entre dois números quaisquer da escala, se consegue uma mensuração consideravelmente mais forte que a ordinal. Atribui-se à variável um número real, uma unidade constante e comum de mensuração. A unidade de mensuração e o ponto zero são arbitrários. Essa é a primeira escala, verdadeiramente, quantitativa.

**d) Escala de razões:** quando uma escala tem todas as características de uma escala de intervalos e, além disso, tem verdadeiro ponto zero como origem, é chamada escala de razões (Siegel, 1975).

### 3.3 Risco relativo (RR)

Suponha que os indivíduos de uma determinada população sejam classificados segundo um fator com duas possíveis respostas, por exemplo, F (fumante) e  $\bar{F}$  (não-fumante). Ainda, suponha a ocorrência ou não de uma certa característica, por exemplo, doente (D) e não-doente ( $\bar{D}$ ). A distribuição da amostra pode ser representada, num quadro ou tabela 2 x 2, conforme a Figura 01:

Fator	Característica	
	D	$\bar{D}$
F	a	B
$\bar{F}$	c	D

FIGURA 01 – Distribuição 2 x 2

Assim, as seguintes quantidades podem ser definidas:

- $R_1 = a/(a + b)$ : proporção de fumantes (F) com a doença (D);
- $R_2 = c/(c + d)$ : proporção de não-fumantes ( $\bar{F}$ ) com a doença (D).

A razão entre essas medidas foi denominada por Cornfield (1951), como sendo o risco relativo para doença, de acordo com o fator de risco de exposição em estudo, no caso ser ou não fumante.

$$RR = \frac{R_1}{R_2} = \frac{a/(a + b)}{c/(c + d)}$$

Cornfield (1951) também descreveu que, se a doença for rara ( $a \ll b$  e  $c \ll d$ ) a quantidade RR pode ser simplificada, pelo cálculo da medida chamada de razão de *odds* (OR), dada por:

$$OR = \frac{a d}{b c}$$

Resumindo, pode-se definir o risco relativo como sendo a proporção de pessoas expostas ao fator de risco que desenvolveram a doença sobre a proporção de pessoas não expostas que desenvolveram a doença. O risco relativo vale 1, se as pessoas expostas e não-expostas desenvolveram a doença na mesma proporção.

### 3.4 Razão de *odds* (OR - *Odds ratio*)

Define-se a razão de *odds* como sendo a chance de se observar casos expostos ao fator de risco sobre a chance de se observar não-casos expostos ao fator de risco. Se a exposição ao fator de risco for a mesma para casos e não-casos, a razão de *odds* vale 1. Essa medida é dada por:

$$\text{OR} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

Os coeficientes ( $\beta$ ) estimados na análise de Regressão Logística, quando exponenciados ( $e^\beta$ ), correspondem à razão de *odds* (OR) para cada variável independente (covariável) incluída no modelo multivariável.

## 4 REGRESSÃO LOGÍSTICA

Os métodos de regressão têm sido usados como uma ferramenta auxiliar a qualquer análise que investiga o relacionamento entre a variável resposta e uma ou mais variáveis explicativas. É freqüente o caso em que a variável de saída é discreta, com dois ou mais possíveis valores, segundo Hosmer e Lemeshow (1989). Neste caso os modelos tradicionais de regressão não podem ser aplicados diretamente aos dados binários, existe a necessidade de se construir modelos especiais, como o modelo de Regressão Logística Binária.

A análise de Regressão Logística aparece como um dos principais métodos de modelagem estatística, cuja resposta de interesse (variável dependente) é do tipo binária ou dicotômica e que, geralmente, assume o valor 1 para o acontecimento de interesse (sucesso) e o valor 0 para o acontecimento complementar (insucesso).

Essa técnica é regida pelas mesmas diretrizes da análise de regressão linear, apesar de seus modelos não terem as suposições de normalidade e variância constante.

O objetivo da Regressão Logística é prever o desfecho usando um modelo parcimonioso. Nesse sentido, a Regressão Logística estima a probabilidade de certo evento ocorrer, utilizando o método de estimação de máxima verossimilhança, depois de transformar a variável dependente numa variável logit, ou seja, o logaritmo natural da *odds* da variável dependente dicotômica (SYLLABUS, 2004).

A análise de Regressão Logística é usada para prever ou explicar a variável dependente binária em função de uma ou mais variáveis independentes (covariáveis), que podem ser categóricas ou contínuas.

Alguns pesquisadores têm dicotomizado respostas não-binárias, de modo que, a probabilidade de ocorrência do evento possa ser modelada com o uso dessa técnica. O principal objetivo da análise de dados através deste critério é a estatística razão de *odds*.

Observa-se que, na Regressão Logística, os coeficientes estimam a mudança no log (*odds*) da variável dependente, não estimando a mudança dela mesma, como ocorre na Regressão Linear.

Embora a Regressão Logística seja conhecida desde os anos 50, foi através de Cox (1972) que a mesma tornou-se popular entre os usuários da Estatística.

Atualmente, são utilizados outros modelos, como por exemplo, a Regressão Logística Ordinal (mais de dois grupos ordenados), a Regressão Logística Multinomial (mais de dois grupos não-ordenados) e a Regressão Logística Condicional (dados pareados).

Este estudo dará enfoque apenas ao modelo de Regressão Logística Binária, buscando explicar o consumo de cigarros entre os estudantes das escolas estaduais de Santa Maria em função das variáveis investigadas.

Considerando uma distribuição 2 x 2, do tipo:

<b>Fator</b>	<b>Evento</b>	
	<b>Sim</b>	<b>Não</b>
<b>Presente</b>	A	b
<b>Ausente</b>	C	d

$$Odds_1 = a/b \quad e \quad Odds_2 = c/d$$

$$OR = \left( \frac{Odds_1}{Odds_2} \right)^h$$

Ainda, seja a transformação logit ou log (*odds*), designada por z:

$$z = \ln \left( \frac{Odds_1}{Odds_2} \right)$$

Na Regressão Logística, diferentemente da função linear, define-se:

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

A probabilidade de que o evento ocorra pode ser obtida, usando-se a função logística:

$$P(Y / X_1, X_2, \dots, X_k) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

### **Interpretação dos coeficientes logit**

Quando a variável  $X_k$  muda por uma unidade:

- z muda por  $\beta_k$ ;
- o  $\ln(\text{odds})$  muda por  $\beta_k$ , e
- a razão de *odds* (OR) muda por  $e^{\beta_k}$ , com  $e = 2,71828\dots$

O modelo de Regressão Logística é muito utilizado como um modelo de probabilidade, interpretado como o risco de que um indivíduo contraia



uma doença durante algum período de tempo especificado durante o qual esteve exposto a certa condição conhecida e associada à doença (fator de risco).

O valor da variável independente indica o estado do sujeito em relação à presença ou ausência de algum fator de risco; o da variável dependente (desfecho), usualmente, indica se o sujeito contraiu ou não a doença ou se morreu ou não.

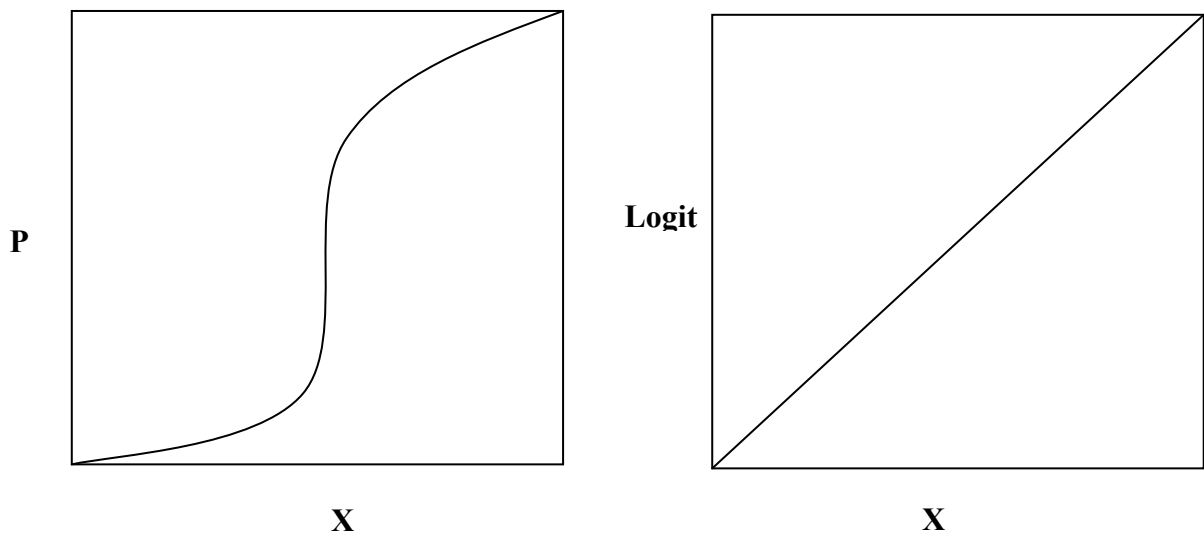


FIGURA 02 – Representação da função e do logit

$$P(Y / X) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1)}}$$

$$\text{logit} = \log \left[ \frac{P}{1 - P} \right] = \alpha + \beta_1 X_1$$

## **Pressuposições do modelo de Regressão Logística**

- a) não assume linearidade no relacionamento entre as variáveis independentes e a variável dependente, mas entre o log (*odds*) ou logit;
- b) a variável dependente não precisa ser normalmente distribuída (mas assume-se que a distribuição pertence à família exponencial, tal como a normal, a Poisson, a binomial ou a gama;
- c) a variável dependente não precisa ser homocedástica para cada nível das variáveis independentes, isto é, não há pressuposição de homogeneidade de variância;
- d) os erros não são assumidos normalmente distribuídos;
- e) não requer que as variáveis independentes sejam intervalares nem ilimitadas.

## **Tipos de questões respondidas pela Regressão Logística**

- quanto da variável dependente binária é explicado pelo modelo ajustado?
- quanto cada variável preditora contribui na probabilidade de um caso ser de um grupo ou de outro da variável dependente?
- qual a probabilidade de que um caso esteja num grupo ou noutro da variável dependente? (Friel, 2004).

## **Algumas aplicações usando Regressão Logística**

### **a) Área da saúde**

- fatores que influenciam na sobrevivência de pacientes;
- fatores de risco na ocorrência de doenças cardíacas;
- estudo de variáveis pré-operatórias para verificar as que são boas preditoras de envolvimento nodal em câncer de próstata;
- estudo de mortalidade fetal e exame pré-natal;
- taxas de mortalidade;
- fatores de risco para hemorragias;
- presença ou não de câncer de pele e exposição ao sol;
- ocorrência ou não de câncer pulmonar em indivíduos com o hábito de fumar;
- a sobrevivência ou não de um paciente após receber uma dose de certa droga.

#### **b) Empresas**

- planos mensais de produção: encomenda e estoque;
- relacionamento entre tempo de execução de tarefas, tempo de experiência e habilidade dos executores;
- quebra ou não de um componente resistente após determinado teste de impacto.

#### **c) Pesquisas sociais**

- pesquisas realizadas a respeito do uso de drogas psicotrópicas;
- estudos sobre desigualdades sociais;
- estudos sobre distribuição de empregos.

## Seleção de variáveis no modelo de Regressão Logística

Vários métodos são utilizados para a seleção das variáveis independentes ou preditoras.

Uma vez definido o conjunto de covariáveis a ser incluído num modelo logístico, busca-se a melhor maneira de encontrar um modelo reduzido que inclua apenas as covariáveis e interações mais importantes para explicar a probabilidade de ocorrência do evento  $\pi(x)$ .

Esse problema poderia ser resolvido pelos métodos usuais de seleção de variáveis em regressão normal linear. Contudo, a questão de interpretação dos parâmetros é crucial num modelo logístico, implicando que uma forma puramente mecânica de seleção pode levar a um modelo sem sentido e de difícil interpretação (Hosmer e Lemeshow, 1989).

Particularmente, a inclusão de certas interações impõe a permanência, no modelo, de seus respectivos efeitos principais de ordem inferior, na ótica do princípio hierárquico. Um aspecto importante é que, muitas vezes, variáveis consideradas biologicamente importantes não devem ser deixadas de lado pela sua falta de significância estatística. Assim, a seleção de um modelo logístico deve ser um processo conjugado de seleção estatística de variáveis e de bom senso (Hosmer e Lemeshow, 1989).

Um dos métodos mais conhecidos de seleção de covariáveis é o método *stepwise*, o qual se baseia num algoritmo misto de inclusão e eliminação de covariáveis segundo a importância das mesmas de acordo com algum critério estatístico. Esse grau de importância pode ser avaliado, por exemplo, pelo nível de significância do teste da razão de verossimilhança entre os modelos que incluem ou excluem as covariáveis

em questão. Quanto menor for esse nível de significância tanto mais importante será considerada a covariável.

Segundo Hosmer e Lemeshow (1989), o processo de seleção deve iniciar com uma análise univariada cuidadosa, para cada uma das variáveis preditoras do estudo, com o cálculo dos valores das razões de *odds*, intervalos de confiança e testes de associação (verificando a pressuposição da linearidade do logit, para as variáveis de escala contínua).

Posteriormente, serão selecionadas as variáveis para a análise multivariável. Nesse caso, qualquer variável cujo teste, na análise univariada, forneceu  $p < 0,25$  deveria ser considerada como uma candidata para o modelo multivariável, juntamente com todas aquelas de conhecida importância biológica.

Uma vez que todas essas variáveis tenham sido identificadas, procedese à análise incluindo todas elas num modelo inicial.

Segundo Hosmer e Lemeshow (1989) *apud* Mickey & Greenland (1989), o nível de significância usual de 0,05 freqüentemente é ineficiente para identificar variáveis reconhecidamente importantes.

Seguindo o ajuste do modelo multivariável, deve-se verificar a importância de cada variável incluída. Isto pode ser feito por:

- análise da estatística de Wald;
- comparação de cada coeficiente estimado com o coeficiente do modelo univariado, contendo apenas essa variável.

Variáveis que não contribuem com o modelo, com base nesses critérios, deveriam ser eliminadas, para assim, ser ajustado um novo modelo, o qual deverá ser comparado com o anterior, procedendo-se o teste

da razão de verossimilhança. O processo de excluir, reajustar e verificar é repetido até que todas as variáveis importantes tenham sido incluídas no modelo.

Depois de obtido um modelo preliminar, verifica-se a necessidade de se incluir termos de interações entre as variáveis selecionadas.

A decisão final de incluir ou não uma interação no modelo deve ser baseada na análise estatística, bem como nas considerações práticas, ou seja, considerando uma perspectiva biológica.

### **Estimação dos parâmetros**

A *likelihood* é uma probabilidade, especificamente, a probabilidade de que valores observados da variável dependente podem ser preditos por valores observados das variáveis independentes. Assim como qualquer probabilidade, a *likelihood* varia de 0 a 1 (SYLLABUS, 2004).

A Estimação de Máxima Verossimilhança (EMV) é o método usado para calcular os coeficientes logit da Regressão Logística, o qual maximiza o log *likelihood*. Isso se reflete em como, provavelmente, os valores observados da variável dependente podem ser preditos em função dos valores das covariáveis.

Este método calcula o valor do log *likelihood* usando um algoritmo iterativo, o qual inicia com uma suposta estimativa (arbitrária) para os valores dos coeficientes logit, determinando a direção e o incremento dos mesmos, os quais aumentarão o valor do log *likelihood*. Depois dessa função inicial ser estimada, os resíduos são testados e uma nova estimativa é obtida, com a aplicação de uma função melhorada.

O processo é repetido até convergir, isto é até que o valor do log *likelihood* não sofra mudanças significativas. Existem vários critérios alternativos para convergência, conforme o programa computacional utilizado.

A log *likelihood* varia de 0 a menos infinito. Essa medida é a base dos testes estatísticos para o modelo de Regressão Logística (SYLLABUS, 2004).

### **Teste de significância para os parâmetros estimados (Wald)**

$$H_0: \beta_k = 0;$$

$$H_1: \beta_k \neq 0.$$

A estatística de Wald (W) é, comumente, usada para testar a significância individual dos coeficientes da Regressão Logística (coeficientes logit), para cada variável independente, isto é, testa-se a hipótese nula de que um particular coeficiente logit é zero. A estatística de Wald é dada por:

$$W = \left( \frac{\beta_k}{SE(\beta_k)} \right)^2 \sim \chi_{\delta}^2, \text{ com } \delta = \text{graus de liberdade.}$$

onde:  $\delta = 1$ , se a variável é contínua;

$\delta = m - 1$ , se a variável não é contínua, sendo  $m$  o número de categorias da mesma;

$\beta_k$  é o coeficiente logit;

$SE(\beta_k)$  é o erro padrão do coeficiente logit.

## Teste da razão de verossimilhança

Analogamente à soma de quadrados dos erros (SQE) da Regressão Linear, pode-se avaliar a significância do modelo de Regressão Logística usando a estatística G, dada por:

$$G = -2 \ln \left[ \frac{\text{likelihood sem a variável}}{\text{likelihood com a variável}} \right] \sim \chi_{\delta}^2$$

onde:  $\delta$  = diferença entre o número de variáveis em cada modelo.

## Teste de Hosmer e Lemeshow

O teste estatístico de Hosmer e Lemeshow é utilizado para avaliar o ajuste do modelo. São testadas as seguintes hipóteses:

$H_0$ : não há diferença entre os valores observados e os valores preditos pelo modelo;

$H_1$ : há diferença entre os valores observados e os valores preditos pelo modelo.

Sendo assim, a decisão do teste considera:

1. se a estatística calculada for menor que 0,05, rejeita-se  $H_0$ ;



2. se a estatística calculada for maior que 0,05, como se quer para o bom ajuste do modelo, aceita-se  $H_0$ , o que implica que as estimativas do modelo ajustam os dados num nível aceitável. Isso não significa que o modelo, necessariamente, explica muita da variância da variável dependente, mas somente que muito ou pouco dessa explicação é significativa. Observa-se que, quanto maior o tamanho da amostra, maior o poder do teste para detectar diferenças na hipótese nula.

### **Avaliando o ajuste do modelo**

Nesta etapa podem-se utilizar várias medidas estatísticas, como:

- a) o valor da log da verossimilhança (L): menores valores dessa medida indicam melhor ajuste do modelo;
- b)  $R_{CS}^2$  de Cox-Snell: permite uma interpretação similar do  $R^2$  múltiplo, no qual maiores valores dessa medida indicam melhor adequação do modelo. Entretanto, seu valor máximo pode ser, e usualmente é, menor do que 1, mesmo que o ajuste seja perfeito, dificultando a interpretação. É calculado por:

$$R_{CS}^2 = 1 - \left( \frac{L_o}{L} \right)^{\frac{2}{N}}$$

onde:  $L_o$  é a *likelihood* do modelo nulo;

$L$  é a *likelihood* do modelo final;

$N$  é o número de observações.

c)  $R_N^2$  de Nagelkerke: é uma modificação do  $R_{CS}^2$  de Cox-Snell, porém seu valor pode variar de 0 a 1. Esse coeficiente é calculado por:

$$R_N^2 = \left( \frac{R_{CS}}{R_{\max}} \right)^2$$

onde:  $R_{\max}^2 = 1 - (L_o)^{\frac{2}{N}}$

O valor de  $R_N^2$  expressa a proporção da variância que é explicada pelas covariáveis incluídas no modelo.

A análise de resíduos também é muito utilizada nesta etapa.

## **5 MATERIAL E MÉTODO**

A cidade de Santa Maria – RS está localizada na região central do Rio Grande do Sul, com cerca de 250 mil habitantes e é conhecida como cidade cultura, devido à concentração de estudantes de todo o país, que buscam formação superior na Universidade Federal de Santa Maria, referência de ensino no interior do estado.

Em maio de 2002, realizou-se um estudo de corte transversal, de base escolar, com o objetivo de avaliar o tabagismo (somente consumo de cigarros) entre os estudantes das escolas estaduais de ensino médio do município de Santa Maria.

Para tanto foi selecionada uma amostra aleatória e representativa da população de estudantes, distribuídos em 8 escolas estaduais, em todas as séries do ensino médio, incluindo o Magistério e o Supletivo, nos turnos: diurno e noturno. Informações complementares foram obtidas na 8<sup>a</sup> Coordenadoria de Ensino de Santa Maria.

O tamanho da amostra foi determinado tendo como base uma estimativa de que a prevalência de consumo de cigarros, entre os estudantes, seria em torno de 20% (Barbosa et al., 1989). O nível de confiança estabelecido foi de 95%, com margem de erro de 3,5 pontos percentuais, estimando-se uma amostra de 459 estudantes.

De acordo com a proporcionalidade da amostra, pré-estabelecida, foram sorteados, no local da entrevista, as turmas e os alunos que participariam da pesquisa. O questionário foi aplicado em sala de aula, após autorização do diretor da escola e na presença do professor, não identificando o respondente, garantindo o sigilo em relação às respostas.

A coleta de informações foi realizada por meio de questionário padronizado de autopreenchimento, no qual os estudantes responderam a questões demográficas e sociais (sexo, idade, renda familiar, se trabalha, série, turno e defasagem nos estudos) e a questões específicas sobre seus hábitos, de sua família e de seus amigos, em relação ao tabagismo (se fuma, número de cigarros consumidos por dia, a importância do cigarro na sua vida, os motivos que os levam a fumar, se o pai fuma, se a mãe fuma, se ambos fumam, se os amigos fumam, se têm vários familiares que fumam e se algum amigo o influenciou a fumar).

Foi definido como sendo fumante aquele aluno que estava fumando regularmente, e como ex-fumante aquele que relatou que já havia fumado regularmente no passado. Para o pai, a mãe e os amigos foram considerados o consumo atual de cigarros.

Foram considerados os questionários que estavam em bom estado de conservação, cujas respostas eram coerentes e desconsiderados aqueles devolvidos em branco ou com respostas incoerentes. Por exemplo: o aluno respondeu que não era fumante e, posteriormente, referiu o número de cigarros consumidos por dia.

Foi realizada uma análise descritiva dos dados, com a construção de tabelas e gráficos, cálculo de média, desvio-padrão, proporções e intervalos de confiança de 95%.

Uma análise bivariada, utilizando-se testes do Qui-quadrado de Pearson e de tendência linear, permitiu verificar a associação entre a variável desfecho (consumo de cigarros entre os estudantes) e as variáveis independentes.

Também foram calculados as razões de odds e os respectivos intervalos de confiança de 95%, utilizando-se uma análise de regressão univariada para escolher as variáveis candidatas à inclusão no modelo multivariável, considerando-se  $p < 25\%$  (Hosmer & Lemeshow, 1989).

Posteriormente, realizou-se uma análise multivariável, verificando o efeito conjunto das variáveis independentes no desfecho, elaborando-se um modelo de Regressão Logística que permite estimar a chance de que um estudante de escola estadual de 2<sup>o</sup> grau de Santa Maria seja fumante, em função das variáveis que se mostraram significativas ao nível de 5% e que foram incluídas no modelo ajustado.

A significância estatística entre os modelos ajustados foi avaliada pelo teste da razão de verossimilhança.

As análises estatísticas foram realizadas usando-se os programas computacionais Statistica 5.0 (Statsoft) e SPSS 10.0.

## 6 RESULTADOS

Para a realização deste trabalho foram entrevistados 459 alunos da rede estadual de ensino, na cidade de Santa Maria – RS, em 2002.

Com base nos dados observados, verificou-se que a prevalência de tabagismo entre os estudantes foi de 18,1%, IC 95% (14,6% – 21,7%), sendo 375 (81,9%) não-fumantes, incluindo 61 (13,3%) ex-fumantes e 1 (0,2%) que não informou.

A Tabela 01 apresenta uma descrição da amostra de acordo com algumas características investigadas. Observou-se um predomínio do sexo feminino de 277 (60,5%) mulheres, sendo que a faixa etária de maior concentração foi de 14 a 19 anos, com 329 (73,8%) estudantes.

A idade média encontrada foi de 19,7 anos, com desvio padrão de 6,9 e IC 95% (19,1 – 20,4) anos, sendo a idade mínima de 14 e a máxima de 66 anos.

A maioria deles 213 (49,5%) apresentou renda familiar menor do que 4 salários mínimos e, a minoria 55 (12,8%) apresentou renda de 10 salários mínimos ou mais.

Em relação à ocupação profissional, constatou-se que 272 (59,5%) dos alunos não trabalham. Quanto à defasagem nos estudos, 288 (64,5%) responderam afirmativamente, sendo que 75 (16,8%) estavam atrasados mais de 1 ano, 79 (17,7%) mais de 2 anos e 134 (30,0%) apresentavam, no mínimo, 3 anos de atraso.

TABELA 01 - Distribuição dos estudantes do ensino médio das escolas estaduais de Santa Maria - RS em 2002, em relação a características investigadas.

<b>Variáveis</b>	<b>n<sup>1</sup></b>	<b>%</b>
<b>Sexo</b>		
Masculino	181	39,5
Feminino	277	60,5
<b>Idade (anos)</b>		
14 a 19	329	73,8
20 a 25	62	13,9
26 a 31	21	4,7
32 a 37	15	3,4
38 ou mais	19	4,2
<b>Renda familiar mensal (salários mínimos)</b>		
1 a 3	213	49,5
4 a 6	99	23,0
7 a 9	63	14,7
10 ou mais	55	12,8
<b>Trabalha</b>		
Não	272	59,5
Sim	185	40,5
<b>Série</b>		
1 <sup>a</sup>	195	42,7
2 <sup>a</sup>	119	26,0
3 <sup>a</sup>	105	23,0
Supletivo do ensino médio	38	8,3
<b>Turno</b>		
Diurno	230	50,2
Noturno	229	49,8
<b>Defasagem nos estudos</b>		
Não está atrasado	158	35,5
1 ano	75	16,8
2 anos	79	17,7
3 anos ou mais	134	30,0

<sup>1</sup>O total (n) para cada variável difere devido à ocorrência de dados faltantes.

De acordo com a Tabela 02, a prevalência de tabagismo foi maior para o sexo feminino 37 (20,4%) do que para o masculino 46 (16,7%), porém esta diferença não foi estatisticamente significativa ( $p = 0,666$ ). O sexo não se mostrou associado com o tabagismo ( $p = 0,306$ ).

A idade dos estudantes também não se mostrou estatisticamente associada ( $p = 0,058$ ). Porém, como este valor é muito próximo do nível de significância de 5% e várias pesquisas científicas mostraram que esta variável foi significativa (Malcon, 2003), ela será incluída, posteriormente, no modelo multivariável.

A renda familiar mensal, em salários mínimos, apresentou associação com tendência linear ( $p = 0,035$ ) com o tabagismo, ou seja, quanto maior a renda, maior a razão de *odds* (OR) e a prevalência para o tabagismo entre os estudantes.

As outras variáveis mostradas na Tabela 02 não foram estatisticamente significativas ao nível de 5%.



TABELA 02 – Prevalência de tabagismo entre estudantes do ensino médio das escolas estaduais de Santa Maria - RS em 2002, em relação a características investigadas.

Variáveis	Fumante (%)		p <sup>1</sup>
	Sim	Não	
<b>Sexo</b>			
Masculino	46 (16,7)	230 (83,3)	0,306
Feminino	37 (20,4)	144 (79,6)	
<b>Idade (anos)</b>			
14 a 19	53 (16,1)	276 (83,9)	0,058
20 a 25	18 (29,0)	44 (71,0)	
26 a 31	3 (14,3)	18 (85,7)	
32 a 37	5 (33,3)	10 (66,7)	
38 ou mais	2 (10,5)	17 (89,5)	
<b>Renda familiar mensal (salários mínimos)</b>			
1 a 3	36 (17,0)	176 (83,0)	0,035
4 a 6	13 (13,1)	86 (86,9)	
7 a 9	11 (17,5)	52 (82,5)	
10 ou mais	18 (32,7)	37 (67,3)	
<b>Trabalha</b>			
Não	47 (17,3)	225 (82,7)	0,535
Sim	36 (19,6)	148 (80,4)	
<b>Série</b>			
1 <sup>a</sup>	32 (16,4)	163 (83,6)	0,137
2 <sup>a</sup>	30 (25,2)	89 (74,8)	
3 <sup>a</sup>	15 (14,3)	90 (85,7)	
Supletivo do ensino médio	6 (16,2)	31 (83,8)	
<b>Turno</b>			
Diurno	43 (18,7)	187 (81,3)	0,766
Noturno	40 (17,6)	187 (82,4)	
<b>Defasagem nos estudos</b>			
Não está atrasado	31 (19,6)	127 (80,4)	0,145
1 ano	8 (10,7)	67 (89,3)	
2 anos	13 (16,5)	66 (83,5)	
3 anos ou mais	31 (23,3)	102 (76,7)	

<sup>1</sup> Significância do teste de associação do Qui-quadrado.

Quando todos os estudantes foram questionados em relação a hábitos do tabagismo, 166 (38,2%) responderam que apenas o pai fuma, 112 (24,8%) que a mãe fuma e 231 (51,3%) que ambos fumam. Ainda, 184 (40,3%) relataram que têm vários fumantes na família.

Observou-se que 314 estudantes (70,6%) têm vários amigos fumantes e 82 (23,5%) deles foram influenciados por algum amigo para começar a fumar. Além disto, 24 (11,9%) disseram que o cigarro é importante na sua vida, sendo 1 deles não-fumante.

Observou-se que as variáveis: amigos fumam ( $p = 0,000$ ), algum amigo o influenciou a fumar ( $p = 0,026$ ) e fumar é importante na sua vida ( $p = 0,000$ ) foram estatisticamente significativas ao nível de 5%.

A Tabela 03 mostra algumas informações relativas aos hábitos de fumar dos estudantes.

TABELA 03 – Prevalência de tabagismo entre estudantes do ensino médio das escolas estaduais de Santa Maria - RS em 2002, em relação aos hábitos de fumar.

Variáveis	Fumante (%)		p <sup>1</sup>
	Sim	Não	
<b>Pai fuma</b>			
Não	44 (16,4)	224 (83,6)	0,169
Sim	36 (21,7)	130 (78,3)	
<b>Mãe fuma</b>			
Não	56 (16,5)	283 (83,5)	0,165
Sim	25 (22,3)	87 (77,7)	
<b>Ambos fumam</b>			
Não	37 (16,9)	182 (83,1)	0,409
Sim	46 (19,9)	185 (80,1)	
<b>Algum familiar fuma</b>			
Não	42 (15,4)	231 (84,6)	0,061
Sim	41 (22,3)	143 (77,7)	
<b>Amigos fumam</b>			
Não	9 (6,9)	122 (93,1)	0,000
Sim	74 (23,6)	239 (76,4)	
<b>Amigo influenciou a fumar</b>			
Não	56 (21,0)	211 (79,0)	0,026
Sim	27 (32,9)	55 (67,1)	
<b>Fumar é importante na vida</b>			
Não	59 (33,3)	118 (66,7)	0,000
Sim	23 (95,8)	1 (4,2)	

<sup>1</sup> Significância do teste de associação do Qui-quadrado.

A Figura 03 mostra alguns motivos que os levam a fumar:

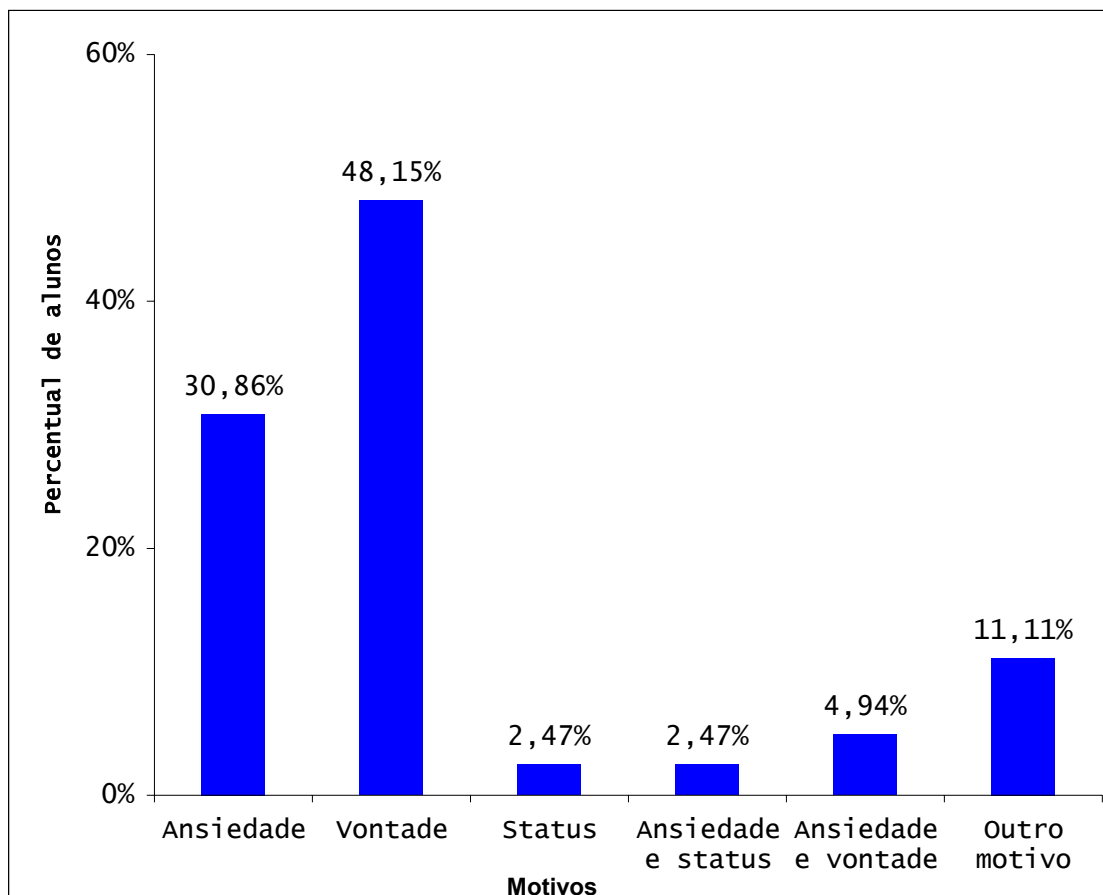


FIGURA 03 – Motivos para fumar relatados pelos estudantes das escolas estaduais de Santa Maria – RS, em 2002.

Observou-se que os fumantes começaram a fumar, em média, aos 14 anos, sendo o desvio padrão de 0,3 e IC 95% (13,4 – 14,6) anos, com idade mínima de 8 e máxima de 25 anos.

A maioria 61 (74,4%) dos fumantes disse que o fazem há mais de 1 ano e consomem, em média, 9,3 cigarros/dia, com desvio padrão de 7,4 e IC 95% (7,6 – 11,0) cigarros/dia.

Algumas informações relativas a este hábito estão na Tabela 04.

TABELA 04 - Distribuição dos estudantes do ensino médio das escolas estaduais de Santa Maria - RS em 2002, em relação ao hábito de fumar.

<b>Variáveis</b>	<b>n<sup>1</sup></b>	<b>%</b>
<b>Início do hábito (anos)</b>		
13 ou menos	41	49,4
14 a 19	40	48,2
20 a 25	2	2,4
<b>Tempo que fuma (anos)</b>		
1 ano ou menos	21	25,6
Mais de 1 ano	61	74,4
<b>Nº de cigarros fumados por dia</b>		
10 ou menos	52	67,5
11 a 20	21	27,3
21 a 30	4	5,2

<sup>1</sup> O total (n) para cada variável difere devido à ocorrência de dados faltantes.

A Tabela 05 apresenta as variáveis que se mostraram, na análise de regressão univariada, associadas estatisticamente ao hábito de fumar dos estudantes, considerando-se o valor de  $p < 25\%$  (Hosmer & Lemeshow, 1989), além daquelas que, de acordo com a literatura revisada são potenciais fatores de risco. Ainda, são apresentados os resultados das razões de *odds* (OR) e seus respectivos intervalos de confiança de 95%.

As variáveis: idade e renda familiar foram dicotomizadas para facilitar a análise multivariável, já que os níveis de significância encontrados foram praticamente os mesmos, quando comparados com as mesmas variáveis distribuídas em faixas ( $p = 0,058$ ) e ( $p = 0,036$ ), respectivamente.

TABELA 05 – Variáveis candidatas a entrar no modelo de Regressão Logística multivariável ( $p < 25\%$ ).

<b>Variáveis</b>	<b>Razão de odds</b>	<b>IC (95%)</b>	<b>p<sup>1</sup></b>
<b>Idade</b>			
14 a 19	1,0	–	
20 ou mais	1,6	0,9 – 2,7	0,061
<b>Renda familiar mensal</b>			
6 ou menos	1,0	–	
7 ou mais	1,7	1,1 – 2,9	0,034
<b>Série</b>			
1 <sup>a</sup>	1,0	–	
2 <sup>a</sup>	1,0	0,4 – 2,6	
3 <sup>a</sup>	1,7	0,7 – 4,6	
Supletivo do ensino médio	0,9	0,3 – 2,4	0,144
<b>Defasagem nos estudos</b>			
Não está atrasado	1,0	–	
1 ano	0,5	0,2 – 1,1	
2 anos	0,9	0,4 – 1,6	
3 anos ou mais	1,3	0,7 – 2,2	0,156
<b>Pai fuma</b>			
Não	1,0	–	
Sim	1,4	0,9 – 2,3	0,170
<b>Mãe fuma</b>			
Não	1,0	–	
Sim	1,5	0,9 – 2,5	0,167
<b>Vários familiares fumam</b>			
Não	1,0	–	
Sim	1,6	1,0 – 2,5	0,062
<b>Amigos fumam</b>			
Não	1,0	–	
Sim	4,2	2,0 – 8,7	0,000
<b>Amigo influenciou a fumar</b>			
Não	1,0	–	
Sim	1,8	1,1 – 3,2	0,027
<b>Fumar é importante na vida<sup>2</sup></b>			
Não	1,0	–	
Sim	45,8	6,0 – 346,0	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

<sup>2</sup> Variável excluída da análise multivariável.

A variável: fumar é importante na sua vida será excluída da análise multivariável, pois o intervalo de confiança para a razão de *odds* (OR) foi muito amplo. Isto se deve ao fato de que há um desequilíbrio na distribuição das freqüências das caselas da tabela 2x2.

Este desequilíbrio afeta o erro-padrão da estimativa da razão de *odds* (OR) que é dado por:  $SE(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$ . Se houver valores muito baixos, o erro padrão aumentará, tornando o intervalo de confiança mais amplo:  $IC(OR) = \exp[\ln OR \pm z_{\alpha} SE(\ln OR)]$ .

Foi o caso da distribuição nesta amostra, onde apareceram as freqüências, em cada casela, iguais a: 118, 59, 1 e 23, quando se fez o cruzamento das variáveis: você fuma e fumar é importante na sua vida, de acordo com a Tabela 03.

Segundo Hosmer & Lemeshow (1989), o principal problema de se colocar mais variáveis explicativas do que o necessário no modelo é o fato de que este pode produzir estimativas instáveis (*overfitting*).

Nesse caso o modelo caracteriza-se por estimar coeficientes e erros-padrão irrealistas e muito grandes. Isto pode ocorrer nos casos em que o número de variáveis é grande em relação ao número de parâmetros e/ou a proporção de respondentes é 0 ou 1, como ocorreu nesta amostra.

A análise univariada mostrou que as variáveis: sexo ( $p = 0,307$ ), se trabalha ( $p = 0,535$ ), pai e mãe fumam ( $p = 0,410$ ) e turno ( $p = 0,766$ ) não foram estatisticamente significativas, ao nível de 25%, seguindo o critério adotado, sendo desconsideradas na análise inicial.

Posteriormente, para selecionar as variáveis candidatas à inclusão em cada modelo ajustado, considerou-se o nível de significância de 5%.

Sendo assim, foi construído o modelo A incluindo as variáveis apresentadas na Tabela 06.

TABELA 06 – Resultados da Regressão Logística multivariável: modelo A

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística de Wald	Razão de Odds	IC (95 %)	p <sup>1</sup>
<b>Idade</b>	0,275	0,447	0,379	1,317	0,548 – 3,161	0,538
<b>Renda</b>	0,622	0,327	3,617	1,863	0,981 – 3,538	0,057
<b>Série</b>						
1 <sup>a</sup>	–	–	3,275	–	–	0,351
2 <sup>a</sup>	0,528	0,352	2,247	1,695	0,850 – 3,381	0,134
3 <sup>a</sup>	-0,081	0,421	0,037	0,922	0,404 – 2,103	0,847
Supletivo do ensino médio	-0,115	0,597	0,037	0,891	0,277 – 2,873	0,847
<b>Defasagem nos estudos</b>						
Não está atrasado	–	–	3,476	–	–	0,324
1 ano	-0,913	0,517	3,121	0,401	0,146 – 1,105	0,077
2 anos	-0,210	0,427	0,243	0,811	0,351 – 1,870	0,622
3 anos ou mais	0,073	0,465	0,024	1,075	0,432 – 2,677	0,876
<b>Pai fuma</b>	0,383	0,309	1,541	1,467	0,801 – 2,687	0,214
<b>Mãe fuma</b>	0,219	0,341	0,411	1,244	0,637 – 2,429	0,522
<b>Vários familiares fumam</b>	0,143	0,309	0,215	1,154	0,629 – 2,117	<b>0,643</b>
<b>Amigos fumam</b>	1,616	0,472	11,726	5,034	1,996 – 12,696	0,001
<b>Amigo influenciou a fumar</b>	0,504	0,320	2,486	1,656	0,885 – 3,100	0,115
<b>Constante</b>	-3,159	0,541	34,041	0,042	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo A forneceu os seguintes resultados:

- 1)  $-2LL_0 = 289,885$ .
- 2) Teste de Hosmer e Lemeshow:  $\chi^2 = 8,27$ ; g. l. = 8;  $p = 0,408$ .
- 3)  $R^2$  (Cox-Snell) = 0,122.
- 4)  $R^2$  (Nagelkerke) = 0,182.



Observando os resultados da Tabela 06, deve-se verificar qual das variáveis apresentou o maior p na análise conjunta. No caso das variáveis categóricas deve-se considerar o valor p da categoria de referência, ou seja, a primeira categoria.

No modelo A, observou-se pela estatística de Wald, que a variável: vários familiares fumam apresentou o maior valor  $p = 0,643$ , e será excluída do modelo B, apresentado na Tabela 07.

TABELA 07 – Resultados da Regressão Logística multivariável: modelo B

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	$p^1$
<b>Idade</b>	0,293	0,445	0,434	1,340	0,561 – 3,205	0,510
<b>Renda</b>	0,639	0,325	3,856	1,894	1,001 – 3,582	0,050
<b>Série</b>						
1 <sup>a</sup>	–	–	3,374	–	–	0,338
2 <sup>a</sup>	0,538	0,351	2,347	1,713	0,860 – 3,410	0,126
3 <sup>a</sup>	-0,078	0,420	0,034	0,925	0,406 – 2106	0,853
Supletivo do ensino médio	-0,107	0,597	0,032	0,899	0,279 – 2,897	0,858
<b>Defasagem nos estudos</b>						
Não está atrasado	–	–	3,360	–	–	0,339
1 ano	-0,890	0,514	3,003	0,411	0,150 – 1,124	0,083
2 anos	-0,212	0,426	0,249	0,809	0,351 – 1,863	0,618
3 anos ou mais	0,076	0,464	0,027	1,079	0,434 – 2,680	0,870
<b>Pai fuma</b>	0,414	0,302	1,879	1,512	0,837 – 2,732	0,170
<b>Mãe fuma</b>	0,245	0,336	0,531	1,278	0,661 – 2,471	0,466
<b>Amigos fumam</b>	1,629	0,471	11,982	5,099	2,027 – 12,827	0,001
<b>Amigo influenciou a fumar</b>	0,517	0,319	2,631	1,677	0,898 – 3,131	0,105
<b>Constante</b>	-3,145	0,541	33,821	0,043	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo B forneceu os seguintes resultados:

1)  $-2LL_0 = 290,099$ .

2) Teste da razão de verossimilhança:

$G = (290,099 - 289,885) = 0,214$  (Qui-quadrado) e g. l. =  $(13 - 12) = 1$ , o qual corresponde ao valor de  $p = 0,644$ , sugerindo que: vários familiares fumam contribuiu muito pouco com o modelo, quando as outras variáveis foram incluídas.

Os valores 13 e 12, que compõem os graus de liberdade, correspondem ao número de coeficientes ( $\beta$ ) estimados no modelo A e B, respectivamente, desconsiderando a constante.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 11,136$ ; g. l. = 8;  $p = 0,194$ .

4)  $R^2$  (Cox-Snell) = 0,121.

5)  $R^2$  (Nagelkerke) = 0,181.

Observou-se, pela estatística de Wald, que a variável idade apresentou o maior valor  $p = 0,510$ , tendo sido excluída do modelo B. Porém, de acordo com a literatura consultada e com os testes realizados, verificou-se que a idade contribuiu significativamente ( $p = 0,001$ ), quando incluída no modelo, para  $G = (299,957 - 290,099) = 9,858$  e g. l. =  $(12 - 11) = 1$ .

Por este motivo, a idade será mantida e a variável: mãe fuma, que apresenta o segundo maior valor de  $p = 0,466$  será excluída no modelo C, apresentado na Tabela 08.

TABELA 08 – Resultados da Regressão Logística multivariável: modelo C

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	p <sup>1</sup>
<b>Idade</b>	0,343	0,440	0,607	1,409	0,595 – 3,339	0,436
<b>Renda</b>	0,679	0,322	4,452	1,972	1,049 – 3,706	0,035
<b>Série</b>						
1 <sup>a</sup>	–	–	3,037	–	–	0,386
2 <sup>a</sup>	0,520	0,350	2,198	1,681	0,846 – 3,341	0,138
3 <sup>a</sup>	0,013	0,412	0,001	1,013	0,452 – 2,272	0,974
Supletivo do ensino médio	-0,195	0,594	0,108	0,823	0,257 – 2,637	0,742
<b>Defasagem nos estudos</b>						
Não está atrasado	–	–	3,434	–	–	0,329
1 ano	-0,864	0,511	2,858	0,421	0,155 – 1,148	0,091
2 anos	-0,172	0,424	0,164	0,842	0,367 – 1,933	0,685
3 anos ou mais	0,162	0,459	0,124	1,176	0,478 – 2,893	0,725
<b>Pai fuma</b>	0,485	0,295	2,695	1,623	0,910 – 2,895	0,101
<b>Amigos fumam</b>	1,508	0,442	11,668	4,520	1,902–10,740	0,001
<b>Amigo influenciou a fumar</b>	0,497	0,317	2,455	1,643	0,883 – 3,058	0,117
<b>Constante</b>	-3,057	0,523	34,125	0,047	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo C forneceu os seguintes resultados:

1)  $-2LL_0 = 294,127$ .

2) Teste da razão de verossimilhança:

$$G = (294,127 - 290,099) = 4,028 \text{ e g. l.} = (12 - 11) = 1.$$

O valor de  $p = 0,045$  indica que a variável: mãe fuma tem significância limítrofe, tendo pouca contribuição no modelo. Por este motivo ela será excluída, neste momento, da análise, retornando posteriormente.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 12,046$ ; g. l. = 8;  $p = 0,149$ .

4)  $R^2$  (Cox-Snell) = 0,118.

5)  $R^2$  (Nagelkerke) = 0,175.

Observou-se, pela estatística de Wald, que a variável: idade apresentou o maior valor  $p = 0,436$  e será excluída no modelo D, apresentado na Tabela 09.

TABELA 09 – Resultados da Regressão Logística multivariável: modelo D

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	$p^1$
<b>Renda</b>	0,689	0,315	4,794	1,992	1,075 – 3,692	0,029
<b>Série</b>						
1 <sup>a</sup>	–	–	4,090	–	–	0,252
2 <sup>a</sup>	0,604	0,344	3,089	1,830	0,933 – 3,588	0,079
3 <sup>a</sup>	0,017	0,407	0,002	1,018	0,458 – 2,259	0,966
Supletivo do ensino médio	-0,138	0,576	0,058	0,871	0,282 – 2,693	0,810
<b>Defasagem nos estudos</b>						
Não está atrasado	–	–	4,621	–	–	0,202
1 ano	-6,95	0,487	2,037	0,499	0,192 – 1,296	0,154
2 anos	-1,176	0,423	0,174	0,839	0,366 – 1,920	0,677
3 anos ou mais	0,386	0,371	1,084	1,471	0,711 – 3,043	0,298
<b>Pai fuma</b>	0,443	0,290	2,328	1,557	0,882 – 2,750	0,127
<b>Amigos fumam</b>	1,420	0,419	11,472	4,136	1,189 – 9,404	0,001
<b>Amigo influenciou a fumar</b>	0,531	0,311	2,912	1,700	0,924 – 3,129	0,088
<b>Constante</b>	-2,988	0,501	35,542	0,050	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo D forneceu os seguintes resultados:

1)  $-2LL_0 = 303,910$ .

2) Teste da razão de verossimilhança:

$$G = (303,910 - 294,127) = 9,783 \text{ e g. l.} = (11 - 10) = 1.$$

O valor de  $p = 0,002$  sugere que a variável: idade contribuiu, significativamente, quando as outras variáveis foram incluídas. Portanto, posteriormente, ela retornará ao modelo.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 10,383$ ; g. l. = 8; p = 0,239.

4)  $R^2$  (Cox-Snell) = 0,114.

5)  $R^2$  (Nagelkerke) = 0,169.

Observou-se, pela estatística de Wald, que a variável: série apresentou o maior valor p = 0,252 e será excluída no modelo E, apresentado na Tabela 10.

TABELA 10 – Resultados da Regressão Logística multivariável: modelo E

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	p <sup>1</sup>
<b>Renda</b>	0,748	0,303	6,075	2,112	1,165 – 3,827	0,014
<b>Defasagem nos estudos</b>						
Não está atrasado	–	–	4,968	–	–	0,174
1 ano	-6,96	0,483	2,072	0,499	0,193 – 1,286	0,150
2 anos	-2,43	0,411	0,350	0,784	0,351 – 1,754	0,554
3 anos ou mais	0,350	0,338	1,071	1,420	0,731 – 2,756	0,301
<b>Pai fuma</b>	0,386	0,286	1,824	1,471	0,840 – 2,574	0,177
<b>Amigos fumam</b>	1,363	0,414	10,857	3,909	1,737 – 8,797	0,001
<b>Amigo influenciou</b>						
<b>A fumar</b>	0,563	0,307	3,368	1,756	0,962 – 3,203	0,066
<b>Constante</b>	-2,754	0,457	36,272	0,064	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo E forneceu os seguintes resultados:

1)  $-2LL_0 = 309,028$ .

2) Teste da razão de verossimilhança:

$$G = (309,028 - 303,910) = 5,118 \text{ e g. l.} = (10 - 7) = 3.$$

O valor de p = 0,163 sugere que a variável: série não contribuiu, significativamente, quando as outras variáveis foram incluídas.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 4,172$ ; g. l. = 8; p = 0,841.

4)  $R^2$  (Cox-Snell) = 0,101.

5)  $R^2$  (Nagelkerke) = 0,51.

Observou-se, pela estatística de Wald, que a variável: pai fuma apresentou o maior valor  $p = 0,177$  e será excluída no modelo F, apresentado na Tabela 11.

TABELA 11 – Resultados da Regressão Logística multivariável: modelo F

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	$p^1$
<b>Renda</b>	0,712	0,298	5,715	2,037	1,137 – 3,651	0,017
<b>Defasagem nos estudos</b>						
Não está atrasado	–	–	6,252	–	–	0,100
1 ano	-7,86	0,478	2,700	0,456	0,179 – 1,164	0,100
2 anos	-2,53	0,395	0,410	0,777	0,358 – 1,685	0,522
3 anos ou mais	0,370	0,329	1,262	1,447	0,759 – 2,759	0,261
<b>Amigos fumam</b>	1,305	0,390	11,192	3,688	1,717 – 7,922	0,001
<b>Amigo influenciou a fumar</b>	0,501	0,301	2,765	1,651	0,914 – 2,980	0,096
<b>Constante</b>	-2,495	0,404	38,196	0,082	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo F forneceu os seguintes resultados:

1)  $-2LL_0 = 326,633$ .

2) Teste da razão de verossimilhança:

$$G = (326,633 - 309,028) = 17,605 \text{ e g. l.} = (7 - 6) = 1.$$

O valor de  $p = 0,000$  sugere que a variável: pai fuma contribuiu, significativamente, quando as outras variáveis foram incluídas. Ela será testada, posteriormente, no modelo.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 4,663$ ; g. l. = 8;  $p = 0,793$ .

4)  $R^2$  (Cox-Snell) = 0,091.

5)  $R^2$  (Nagelkerke) = 0,137.

Observou-se, pela estatística de Wald, que a variável: defasagem nos estudos apresentou o maior valor  $p = 0,100$  e será excluída no modelo G, apresentado na Tabela 12.

TABELA 12 – Resultados da Regressão Logística multivariável: modelo G

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	$p^1$
<b>Renda</b>	0,521	0,284	3,358	1,684	0,964 – 2,940	0,067
<b>Amigos fumam</b>	1,340	0,383	12,270	3,819	1,804 – 8,083	0,000
<b>Amigo influenciou a fumar</b>	0,535	0,295	3,280	1,707	0,957 – 3,045	0,070
<b>Constante</b>	-2,513	0,375	44,964	0,081	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo G forneceu os seguintes resultados:

1)  $-2LL_0 = 334,643$ .

2) Teste da razão de verossimilhança:

$$G = (334,643 - 326,633) = 8,01 \text{ e g. l.} = (6 - 3) = 3.$$

O valor de  $p = 0,046$  sugere que a variável: defasagem nos estudos contribuiu, significativamente, quando as outras variáveis foram incluídas. Ela será testada, posteriormente, no modelo.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 0,557$ ; g. l. = 4;  $p = 0,968$ .

4)  $R^2$  (Cox-Snell) = 0,070.

5)  $R^2$  (Nagelkerke) = 0,105.

Observou-se, pela estatística de Wald, que a variável: algum amigo o influenciou a fumar apresentou o maior valor  $p = 0,070$  e será excluída no modelo H, apresentado na Tabela 13.

TABELA 13 – Resultados da Regressão Logística multivariável: modelo H

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	$p^1$
<b>Renda</b>	0,613	0,272	5,069	1,846	1,083 – 3,149	0,024
<b>Amigos fumam</b>	1,399	0,374	13,980	4,049	1,945 – 8,429	0,000
<b>Constante</b>	-2,757	0,363	57,826	0,064	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo H forneceu os seguintes resultados:

1)  $-2LL_0 = 379,090$ .

2) Teste da razão de verossimilhança:

$$G = (379,090 - 334,643) = 44,447 \text{ e g. l.} = (3 - 2) = 1.$$

O valor de  $p = 0,000$  sugere que a variável: algum amigo o influenciou a fumar contribuiu, significativamente, quando as outras variáveis foram incluídas. Ela será testada, posteriormente, no modelo.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 0,111$ ; g. l. = 2;  $p = 0,946$ .

4)  $R^2$  (Cox-Snell) = 0,053.

5)  $R^2$  (Nagelkerke) = 0,086.

Após várias outras análises realizadas, nas quais foram incluídas e excluídas as variáveis estatisticamente significativas citadas anteriormente e aquelas, de importância biológica comprovada, de acordo com a literatura revisada, ajustou-se o modelo final I, apresentado na Tabela 14.



TABELA 14 – Resultados da Regressão Logística multivariável: modelo I

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	p <sup>1</sup>
<b>Idade</b>	0,623	0,289	4,645	1,864	1,058 – 3,284	0,031
<b>Renda</b>	0,712	0,287	6,137	2,038	1,160 – 3,581	0,013
<b>Amigos fumam</b>	1,475	0,394	13,998	4,370	2,018 – 9,463	0,000
<b>Constante</b>	-3,044	0,401	57,572	0,048	–	0,000

<sup>1</sup> Significância do teste de Wald para os coeficientes.

O modelo I forneceu os seguintes resultados:

1)  $-2LL_0 = 363,225$ .

2) Teste da razão de verossimilhança:

$$G = (379,090 - 363,225) = 15,865 \text{ e g. l.} = (3 - 2) = 1.$$

O valor de  $p = 0,000$  sugere que a variável: idade contribuiu, significativamente, quando as outras variáveis foram incluídas.

3) Teste de Hosmer e Lemeshow:  $\chi^2 = 0,650$ ; g. l. = 4;  $p = 0,957$ .

4)  $R^2$  (Cox-Snell) = 0,066.

5)  $R^2$  (Nagelkerke) = 0,107.

A Tabela 15 mostra os resultados obtidos das interações que foram testadas para o modelo, na qual verificou-se que não existe interação significativa entre as variáveis, ou seja, o efeito de uma não modifica a outra em relação ao desfecho.

Não foi identificada nenhuma interação significativa entre as outras variáveis analisadas.

TABELA 15 – Resultados das interações entre as variáveis

Variáveis	Coefficiente regressão ( $\beta$ )	Erro padrão	Estatística De Wald	Razão de Odds	IC (95%)	p <sup>1</sup>
Idade x renda	-0,665	0,706	0,887	0,514	0,129 – 2,052	0,346
Idade x amigos fumam	0,866	0,795	1,188	2,378	0,501 – 11,290	0,276
Renda x amigos fumam	-2,53	0,763	0,110	0,776	0,174 – 3,461	0,740

<sup>1</sup> Significância do teste de Wald para os coeficientes.

Após a análise final obteve-se o modelo, cujas variáveis preditivas para o tabagismo entre estudantes de escolas estaduais de Santa Maria são:  $X_1$  = idade ( $p = 0,031$ ),  $X_2$  = renda familiar ( $p = 0,013$ ) e  $X_3$  = se os amigos fumam ( $p = 0,000$ ), a seguir:

$$z = \ln odds (\text{fumante}) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$z = \ln odds (\text{fumante}) = -3,044 + 0,623 X_1 + 0,712 X_2 + 1,475 X_3$$

Em termos de probabilidade, tem-se:

$$P(Y / X_1, X_2, X_3) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}}$$

Substituindo-se os coeficientes estimados na Regressão Logística multivariável, obteve-se o modelo:

$$P(Y / X_1, X_2, X_3) = \frac{1}{1 + e^{-(-3,044 + 0,623 X_1 + 0,712 X_2 + 1,475 X_3)}}$$

As estimativas dos coeficientes e as principais estatísticas deste modelo são mostradas na Tabela 14. O resultado do teste de Hosmer e Lemeshow indicou que não há problema de ajuste ( $p = 0,957$ ).

Assim, por exemplo, a estimativa da razão de *odds* para um estudante com menos de 19 anos ( $X_1 = 0$ ), renda familiar de 7 salários mínimos ou mais ( $X_2 = 1$ ) e que tem amigos fumantes ( $X_3 = 1$ ) é igual a:

$$z = \ln \text{odds} (\text{fumante}) = -3,044 + 0,623 (0) + 0,712 (1) + 1,475 (1).$$

$$z = \ln \text{odds} (\text{fumante}) = -0,857.$$

$$\text{odds} (\text{fumante}) = e^z = e^{-0,857} = 0,42.$$

A probabilidade de que um estudante das escolas estaduais de Santa Maria que apresente estas características seja fumante é dada por:

$$P(Y = 1 / X_1, X_2, X_3) = \frac{1}{1 + e^{0,857}} = 29,79\%.$$

Ou ainda, se for considerado que este mesmo estudante não tenha amigos fumantes, a probabilidade diminui para 8,85%, sendo esta a variável mais importante no modelo.

## 7 DISCUSSÃO

Este estudo selecionou uma amostra aleatória e representativa dos estudantes das escolas estaduais de Santa Maria, em 2002, possibilitando estimar a prevalência e identificar fatores de risco associados ao tabagismo.

No entanto, estudos de base escolar podem fornecer estimativas subestimadas dos valores reais da prevalência de tabagismo, uma vez que não incluem os alunos evadidos e que, segundo Segat et al (1998), apresentam a maior prevalência para o fumo.

A prevalência encontrada neste estudo foi de cerca de 18%, maior do que a observada por Segat et al (1998) também em Santa Maria. Entretanto, o estudo foi realizado entre estudantes de 10 a 19 anos, das escolas de 1<sup>o</sup> e 2<sup>o</sup> graus. Barbosa et al (1989) observaram um consumo de tabaco de 19,5% entre estudantes de 10 capitais brasileiras.

Embora alguns estudos, como Barbosa et al (1989), tenham observado que as prevalências de tabagismo são maiores para o sexo masculino, neste estudo a diferença não foi estatisticamente significativa ( $p = 0,666$ ), sendo maior para o sexo feminino (20,4%) do que para o masculino (16,7%), concordando com os resultados obtidos por Ivanovic (1997).

A maior parte dos estudos analisados foi realizada após aplicação de questionários, cujas respostas não passaram por nenhum processo de validação, o que poderia ser feito, em alguns casos, usando-se marcadores biológicos, segundo Malcon et al (2003). Neste caso, foi adotado procedimento semelhante. O sigilo das respostas foi garantido, com os questionários sendo respondidos de forma consentida, individual e sem

identificação. Porém, deve-se considerar a possibilidade de que algum estudante tenha omitido seu hábito de fumar.

Algumas limitações ocorreram, quando se tentou comparar as prevalências com outras publicações, devido ao fato de que são utilizadas várias definições para a variável desfecho (consumo de cigarros). Neste estudo foram incluídos aqueles estudantes que responderam que estavam fumando regularmente.

Este estudo observou que os estudantes começam a fumar, em média, aos 14 anos, sendo que a idade mínima declarada foi 8 e a máxima 25 anos, o que coincide com alguns resultados encontrados na literatura, como Schio et al (1992). Ainda, 97,6% começaram a fumar antes dos 19 anos, sendo que apenas 2,4% começaram depois de adulto. Esse resultado coincide com a maioria dos estudos analisados, como Machado-Neto et al (2003) e Malcon et al (2003). A maioria dos estudantes (74,4%) declarou que estava fumando há mais de 1 ano e que consumia, em média, 9,3 cigarros por dia.

Neste estudo, assim como na maioria dos trabalhos referenciados, a prevalência de tabagismo aumentou com a idade, porém a significância foi limiar ( $p = 0,058$ ).

Os resultados obtidos para a associação de fatores socioeconômicos com o tabagismo não se mostraram consistentes na literatura. Neste estudo foi observada associação significativa para a variável: renda familiar, em salários mínimos, ( $p = 0,035$ ) e não-significativa para a variável: se o estudante trabalha ( $p = 0,535$ ). Schio et al (1992) observaram que o trabalho é fator de risco para o tabagismo em jovens.

Embora o tabagismo dos pais tenha sido identificado como fator de risco em alguns estudos da literatura revisada, aqui não foi encontrada

associação significativa ( $p = 0,409$ ), concordando com o que foi observado em Segat et al (1989) e discordando com Machado-Neto e Cruz (2003). Deve-se considerar que há diferenças entre as faixas etárias e o tipo de amostragem adotado entre os estudos. Levando em conta que apenas um deles seja fumante, também não se observou associação significativa para as variáveis: pai fuma ( $p = 0,165$ ) e mãe fuma: ( $p = 0,169$ ).

O tabagismo dos familiares se mostrou com significância estatística limiar ( $p = 0,061$ ). Segat et al (1989), encontrou que o risco relativo para os estudantes que convivem com irmãos fumantes é de 2,29 ( $p < 0,001$ ), em Santa Maria.

Em relação a ter amigos fumantes, a associação obtida foi significativa, com OR = 4,2 ( $p = 0,000$ ). O estudo de Segat et al (1989) encontrou risco para o tabagismo igual a 5,24 ( $p = 0,001$ ). Outros estudos também encontraram resultados significativos, como Ivanovic et al (1997) e Schio et al (1992).

As variáveis: série ( $p = 0,137$ ) e turno ( $p = 0,766$ ) não se mostraram associadas com o tabagismo dos estudantes. A defasagem nos estudos não foi observada como fator de risco para o tabagismo ( $p = 0,145$ ). Schio et al (1992) encontraram associação significativa, considerando apenas estudantes adolescentes.

Algumas limitações foram relatadas neste estudo, entretanto procurou-se contornar os problemas que poderiam ser decorrentes das mesmas, buscando apresentar informações concisas e reais a respeito de um assunto relevante e que merece sempre a atenção de pesquisadores e responsáveis pela saúde pública.

## 8 CONCLUSÃO

Neste estudo foram entrevistados 459 alunos da rede estadual de ensino, na cidade de Santa Maria – RS, em 2002, entre os quais observou-se uma prevalência de 18,1%, para o tabagismo (consumo de cigarros), IC 95% (14,6% – 21,7%), sendo maior no sexo feminino (20,4%) do que no masculino (16,7%), embora esta diferença não tenha se mostrado significativa ( $p = 0,666$ ).

A idade dos estudantes, na análise univariada, mostrou significância limiar ( $p = 0,058$ ). Na análise multivariável, a mesma permaneceu no modelo, pois se mostrou como fator de risco, com OR = 1,864 ( $p = 0,031$ ).

A renda familiar mensal, em salários mínimos, apresentou associação com tendência linear ( $p = 0,035$ ) com o tabagismo, ou seja, quanto maior seu valor, maior a razão de *odds* e a prevalência para o tabagismo entre os estudantes. Na análise multivariável encontrou-se OR = 2,038 ( $p = 0,013$ ).

A variável que se mostrou como o maior fator de risco foi: se tem amigos fumantes, OR = 4,370 ( $p = 0,000$ ). Este mesmo resultado também foi encontrado na análise univariada.

Neste estudo, não se encontrou evidência significativa, ao nível de 5%, de que as outras variáveis analisadas: sexo, se trabalha, série, turno, defasagem nos estudos, se o pai fuma, se a mãe fuma, se ambos fumam, se têm familiares que fumam, se algum amigo o influenciou a fumar e se fumar é importante na sua vida sejam fatores de risco para o tabagismo dos estudantes.

Sendo assim, este trabalho possibilitou, de forma prática e econômica, levantar questões importantes em relação ao tabagismo entre os alunos de

escolas estaduais de Santa Maria – RS, estabelecendo-se um modelo que permite estimar a probabilidade de um estudante ser fumante em função das variáveis que se mostraram estatisticamente significativas na análise de regressão Logística multivariável: idade, renda familiar e amigos fumam.

A maior probabilidade de ser fumante (29,79%) acontece na seguinte situação: um estudante de escola estadual com mais de 19 anos, cuja renda familiar é de 7 ou mais salários mínimos e tem amigos fumantes.

Se considerarmos que este estudante não tem amigos fumantes percebe-se que este valor cai para 8,85%, mostrando a importância desta variável.

Apesar de muitas campanhas, a Organização Mundial de Saúde prevê que, se nada for feito, em 2020 o vício do cigarro levará mais de 10 milhões de pessoas à morte, anualmente (Fumo, 2004).

Os resultados deste e de outros estudos mostram que os estudantes começam a fumar precocemente, destacando-se a necessidade de se trabalhar, em termos de prevenção intensiva, diretamente com adolescentes do grupo de risco observado.

Deixam-se como sugestões: um estudo de base populacional para estimar a prevalência entre os que não estão frequentando a escola e também a inclusão de outras variáveis, como: pais separados, uso e abuso de bebidas alcoólicas, prática de esportes e religião.

Além disto, sugere-se o levantamento de uma amostra maior de forma que se possa fazer uma análise estratificada por sexo e por faixa etária.



## 9 BIBLIOGRAFIA

1. **ÁLCOOL E DROGAS SEM DISTORÇÃO**. Acesso em: 24 maio 2004.  
Disponível em: <http://www.einstein.br/alcooledrogas>.
2. **A NOTÍCIA**. Acesso em: 20 agosto 2004. Disponível em:  
<http://an.uol.com.br/2001/jun/03/0opi.htm>
3. BARBOSA, M. T. S., CARLINI-COTRIM, B., SILVA-FILHO, A. R.  
O uso de tabaco por estudantes de primeiro e segundo graus em dez capitais brasileiras: possíveis contribuições da Estatística Multivariada para a compreensão do fenômeno. **Revista de Saúde Pública**. v.23(5), p.401-409, 1989.
4. CCOO – Centro de Cultura Oriente Ocidente. Acesso em: 23 agosto 2004. Disponível em:  
[http://www.falconi.com.br/artigos\\_sosfumo/artigofumo12.htm](http://www.falconi.com.br/artigos_sosfumo/artigofumo12.htm).
5. CORNFIELD, J. A. Method of estimating comparative rates from clinical data: Applications to cancer of the lung, breast and cervix. **Journal of the National Cancer Institute**. v.11, p.1269-1275, 1951.
6. Cox, D. R. **The Analysis of Binary Data**. Chapman and Hall, 1970.
7. DOLL, R. **Tobacco: an overview of health effects**. In: D. Zaridze & R. Peto, eds. Tobacco: a major International Health Hazard. Lyon: International Agency for Research on Cancer. p.11-22, 1986.
8. FUMO. Acesso em: 20 agosto 2004. Disponível em:  
<http://www.sesmt.com.br/fumo.htm>.
9. FRIEL, C. M. **Logistical Regression**. Acesso em: 20 novembro 2004.  
Disponível em: [http://www.shsu.edu/~icc\\_cmf/cj\\_742/logit.doc](http://www.shsu.edu/~icc_cmf/cj_742/logit.doc).

10. HORTA, B. L., CALHEIROS, P., PINHEIRO, R. T., TOMASI, E. et al. Tabagismo em adolescentes de área urbana na região Sul do Brasil. **Revista de Saúde Pública**. v.35, n.2, p.159-164, 2001. Acesso em: 21 maio 2002. Disponível em:  
[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-89102001000200009&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102001000200009&lng=pt&nrm=iso).
11. HOSMER, D. W., LEMESHOW, S. **Applied logistic regression**. New York: John Wiley, 1989.
12. INAN - Instituto Nacional de Alimentação e Nutrição - PNSN: estatísticas sobre hábitos de fumo no Brasil. Brasília (DF), 1989.
13. INCA. Pare de fumar: atualidades sobre o tabagismo. Acesso em: 24 maio 2004. Disponível em:  
<http://www.inca.gov.br/tabagismo/atualidades>.
14. IVANOVIC, D. M., CASTRO, C. G., IVANOVIC, R. M. Factors affecting chilean elementary and high school children's smoking. **Revista de Saúde Pública**. v.31(1), p.30-43, 1997.
15. JORNAL A RAZÃO. Edição: 25/05/2004. Santa Maria – RS.
16. LOLIO, C. A., SOUZA, J. M. P. D., SANTO, A. H., BUCHALLA, C. M. Prevalência de tabagismo em localidade urbana da região sudeste do Brasil. **Revista de Saúde Pública**. v.27(4), p.262-265, 1993.
17. MACHADO-NETO, A. S., CRUZ, A. A. Tabagismo em amostra de adolescentes escolares de Salvador – Bahia. **Jornal de Pneumologia**. v.29 (5), 2003.
18. MALCON, M. C., MENEZES, A. M. B., CHATKIN, M. Prevalência e fatores de risco para tabagismo em adolescentes. **Revista de Saúde Pública**. v.37(1), p.1-7, 2003.

19. MALCON, M. C., MENEZES, A. M. B., MAIA, M. F. S., CHATKIN, M. et al. Prevalência e fatores de risco para tabagismo em adolescentes na América do Sul: uma revisão sistemática da literatura. **Revista Panamericana de la Salud Pública**. v.13 (4), p.222-228, 2003.
20. MEDJUNIOR. Acesso em: 22 setembro 2004. Disponível em: <http://medjunior.vilabol.uol.com.br/>.
21. MUZA G. M., BETTIOL, H. MUCCILLO, G. BARBIERI, M. A. Consumo de substâncias psicoativas por adolescentes escolares de Ribeirão Preto, SP (Brasil): II – Distribuição por classe social. **Revista de Saúde Pública**. v. 31, p.163-170, 1997.
22. PASQUALOTTO, A. C., PASQUALOTTO, G. C., SANTOS, R. P., SEGAT, F. M. et al. Relação entre o adolescente e o tabaco: estudo de fatores sócio-demográficos de escolares em Santa Maria, RS. **Pediatria**. v.24(1/2), p.11-16, 2002.
23. PREGIBON, D. Logistic regression diagnostics. **The annals of Statistics**. v.9(4), p.705-724, 1981.
24. RADIOBRAS. Acesso em: 09 setembro 2004. Disponível em: [http://www.radiobras.gov.br/ct/2002/materia\\_300802\\_4.htm](http://www.radiobras.gov.br/ct/2002/materia_300802_4.htm).
25. SCHIO, C., REVERBEL, E., FERNANDES, E., GUGEL, F. et al. O tabagismo entre estudantes secundaristas da zona urbana de Porto Alegre. **Rev. HCPA**. v.12(2), p.117-120, 1992.
26. SEGAT, F. M., SANTOS, R. P., GUILLANDE, S., PASQUALOTTO, A. C. et al. Fatores de risco associados ao tabagismo em adolescentes. **Adolescência Latinoamericana**. v.1(3), p.163-169, 1998.
27. SIEGEL, S. **Estatística não-paramétrica** (para as ciências do comportamento). São Paulo: McGraw-Hill, 1975.

28. SIMÕES, M. J. S. Tabagismo entre estudantes do primeiro e segundo graus em Araraquara, SP, Brasil, 1988. **Medicina**. Ribeirão Preto, SP. v.23(4), p.223–231, 1990.
29. SOARES, J. F., BARTMAN, F. C. **Métodos estatísticos em Medicina e Biologia**. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 14, 1983, MG.
30. SYLLABUS. **Logistic Regression**. Acesso em: 02 setembro 2004.  
Disponível em: <http://www2.cahss.ncsu.edu/garson/pa765/logistic.htm>.

## **APÊNDICE**

---

APÊNDICE 01– Questionário respondido pelos estudantes.

**UNIVERSIDADE FEDERAL DE SANTA MARIA**  
**DEPARTAMENTO DE ESTATÍSTICA**  
**ESPECIALIZAÇÃO EM ESTATÍSTICA E MODELAGEM QUANTITATIVA**

Caro aluno, responda às questões abaixo e ajude a traçar o perfil dos alunos que utilizam o tabaco (cigarros).

- 1 – Sexo:  
 feminino  
 masculino
- 2 – Idade: \_\_\_\_\_ (anos)
- 3 – Trabalha:  
 sim  
 não
- 4 – Renda familiar mensal:  
 1 a 3 salários mínimos  
 4 a 6 salários mínimos  
 7 a 9 salários mínimos  
 10 ou mais salários mínimos
- 5 – Série: \_\_\_\_\_
- 6 – Turno:  
 manhã  
 tarde  
 noite
- 7 – Pai fuma:  
 sim  
 não  
 não tem conhecimento
- 8 – Mãe fuma:  
 sim  
  
 não  
 não tem conhecimento
- 9 – Na sua família, várias pessoas fumam?  
 sim  
 não
- 10 – Você fuma?  
 sim  
 não
- 11 – Com que idade você fumou pela primeira vez? \_\_\_\_\_(anos)
- 12 - Há quanto tempo você fuma?  
 duas semanas  
 aproximadamente 2 meses  
 1 ano  
 há mais tempo
- 13 - Quantos cigarros você fuma durante o dia?  
\_\_\_\_\_
- 14 - O uso do cigarro é importante na sua vida?  
 sim  
 não
- 15 - Qual o motivo que leva você a fumar?  
 ansiedade  
 vontade  
 status  
 outro motivo: \_\_\_\_\_
- 16 – Você tem muitos amigos fumantes?  
 sim  
 não
- 17 – Algum amigo influenciou você a começar a fumar?  
 sim  
 não
- 18 – Tempo de atraso nos estudos:  
 não está atrasado  
 1 ano  
 2 anos  
 3 anos ou mais

Orientadora: Roselaine Ruviaro Zanini