

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP
RELATÓRIO DE CONSULTA

TÍTULO DO PROJETO: “Análise do sistema fonológico em crianças de sete a oito anos”

PESQUISADORA: Talita Consorti

ORIENTADORA: Haydée Fiszbein Wertzner

INSTITUIÇÃO: Faculdade de Medicina - USP

FINALIDADE DO PROJETO: Iniciação científica

PARTICIPANTES DA ENTREVISTA: Haydée Fiszbein Wertzner
Talita Consorti
Rinaldo Artes
Carmen Diva Saldiva de André
Frederico Zanqueta Poletto
Daniel Hayakawa Kamo
Daniel Kashiwamura Scheffer

DATA: 09/04/2002

FINALIDADE DA CONSULTA: Sugestões e orientação para análise de dados

RELATÓRIO ELABORADO POR: Daniel Hayakawa Kamo
Daniel Kashiwamura Scheffer

1. INTRODUÇÃO

O processo gradual de aquisição da fala é chamado de desenvolvimento do sistema fonológico. O estudo desse desenvolvimento possibilita uma melhor compreensão do que acontece com crianças que apresentam alterações na fala que caracterizam o distúrbio fonológico. Esse distúrbio é freqüentemente diagnosticado em crianças de até 7 anos. Após essa idade é comum que a criança passe a utilizar corretamente os fonemas que conseguiu produzir.

Dentre os distúrbios fonológicos, têm-se a simplificação de encontro consonantal (SEC) e a simplificação da consoante final (SCF). O distúrbio SEC ocorre quando a criança não pronuncia uma das consoantes de uma palavra que possui duas ou mais consoantes em sequência. O distúrbio SCF ocorre quando a consoante final da palavra não é pronunciada pela criança.

Este estudo foi realizado com a finalidade de estudar o desenvolvimento fonológico de crianças, com idade de 7 anos a 7 anos e 11 meses e de 8 anos a 8 anos e 11 meses, em relação aos distúrbios SEC e SCF.

2. DESCRIÇÃO DO ESTUDO

Fizeram parte do estudo 80 crianças: 40 com idade de 7 anos completos e 40 com idade de 8 anos completos, consideradas normais por seus familiares e professores (ou seja, não tiveram queixas de alteração de linguagem, fala e/ou audição, bem como alterações emocionais ou cognitivas). Além disso, todas elas eram estudantes de 5 escolas da região do bairro do Butantã (São Paulo), sendo 3 escolas públicas e 2 particulares. Foram formados quatro grupos, cada um com 20 crianças, de acordo com as idades (7 anos e 8 anos), sendo metade do sexo feminino e metade do sexo masculino. Dois grupos foram formados por crianças que freqüentavam escola pública e dois grupos escola privada.

Inicialmente, foi preenchido um questionário sobre o desenvolvimento da criança com o responsável pela mesma. Posteriormente, foram aplicadas provas de fonologia compostas por uma prova de nomeação e uma de imitação. A prova de nomeação

consistia na associação de objetos ou figuras aos seus respectivos nomes no menor intervalo de tempo possível. Na prova de imitação a criança tinha que repetir o vocábulo pronunciado pelo examinador.

3. DESCRIÇÃO DAS VARIÁVEIS

As variáveis explicativas utilizadas no estudo são :

- **idade** $\left\{ \begin{array}{l} 0, \text{ para crianças de 7 anos a 7 anos e 11 meses} \\ 1, \text{ para crianças de 8 anos a 8 anos e 11 meses} \end{array} \right.$
- **sexo** $\left\{ \begin{array}{l} 0, \text{ para masculino} \\ 1, \text{ para feminino} \end{array} \right.$
- **escola** $\left\{ \begin{array}{l} 0, \text{ para escola particular} \\ 1, \text{ para escola publica} \end{array} \right.$

As variáveis resposta do estudo são :

- **SEC** (número de processos fonológicos de simplificação de encontro consonantal)
- **SCF** (número de processos fonológicos de simplificação de consoante final)

Como a finalidade do estudo é verificar a presença ou não de um processo fonológico nas crianças, a variável Processo Fonológico foi transformada nas variáveis dicotômicas :

- **Presença de SEC** $\left\{ \begin{array}{l} 0, \text{ se a criança não apresenta processo SEC} \\ 1, \text{ se a criança apresenta algum processo SEC} \end{array} \right.$
- $\left\{ \begin{array}{l} 0, \text{ se a criança não apresenta processo SCF} \\ 1, \text{ se a criança apresenta algum processo SCF} \end{array} \right.$

- **Presença de SCF**

4. SITUAÇÃO DO PROJETO

A pesquisadora terminou a coletas dos dados e veio ao CEA em busca de orientação para a análise estatística do seu estudo.

5. SUGESTÕES DO CEA

Para leitura adequada dos dados pelos pacotes de análise estatística mais comuns, sugerimos que a pesquisadora os organize conforme está apresentado na tabela abaixo. Essa formatação pode ser implementada utilizando programas como o Microsoft Excel.

Sujeitos	SEC	Presença de SEC	Idade	Sexo	Escola
1	0	0	0	0	1
2	3	1	0	1	1
3	1	1	1	0	0
4	0	0	0	0	1
5	0	0	1	0	1
6	0	0	0	1	1
...
80	0	0	1	1	0

Para descrever a relação entre as variáveis de controle **sexo**, **idade** e **escola** e as variáveis resposta **presença de SEC** e **presença de SCF** foi sugerida uma análise de regressão logística (Neter et al., 1996). Deverão ser feitas duas análises, pois nesse estudo há 2 processos fonológicos distintos a serem analisados (SEC e SCF).

No Apêndice A é apresentado o procedimento, passo a passo, de realização da análise de regressão logística no software Minitab. No Apêndice B é apresentada a

análise do processo fonológico SEC, com uma explicação sucinta de cada passo realizado.

Para comparar o número médio de processos em relação à idade e escola das crianças foi sugerida uma análise de variância (Neter et al., 1996). No apêndice C é apresentado o procedimento, passo a passo, de realização da análise de variância no software Minitab. No apêndice D apresentamos a análise poara a variável **SEC**.

6. REFERÊNCIA BIBLIOGRÁFICA

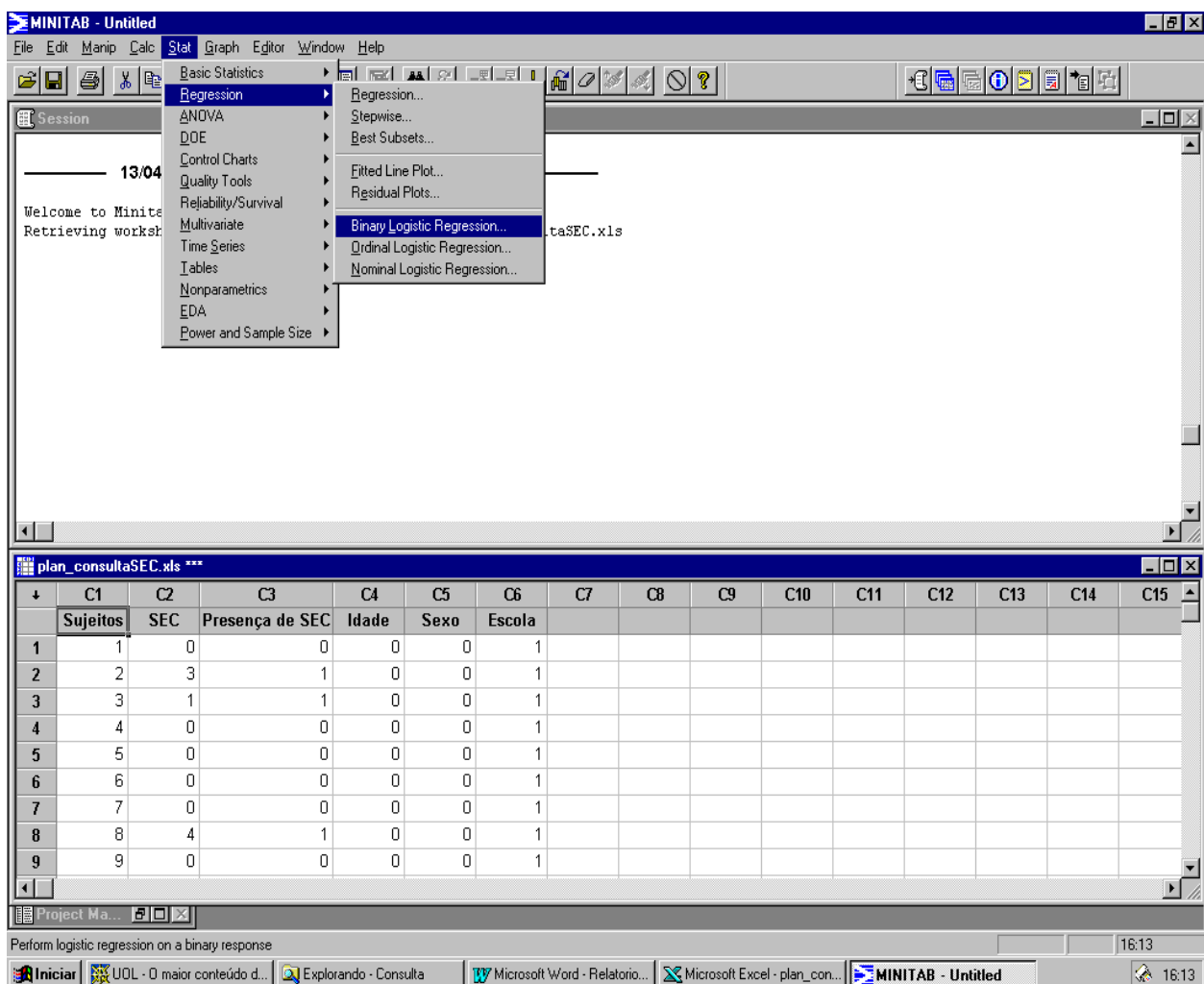
NETER, J., KUTNER, M. H., NACHTSHEIM, C. J. And WASSERMAN, W. (1996). **Applied Linear Statistical Models**. 4ed. Boston: Mc Graw Hill. 1408p.

Apêndice A
(Regressão Logística no Minitab)

Para esta análise, o software utilizado foi o Minitab (versão 13.1) por ser uma ferramenta de fácil utilização e que contém os recursos necessários.

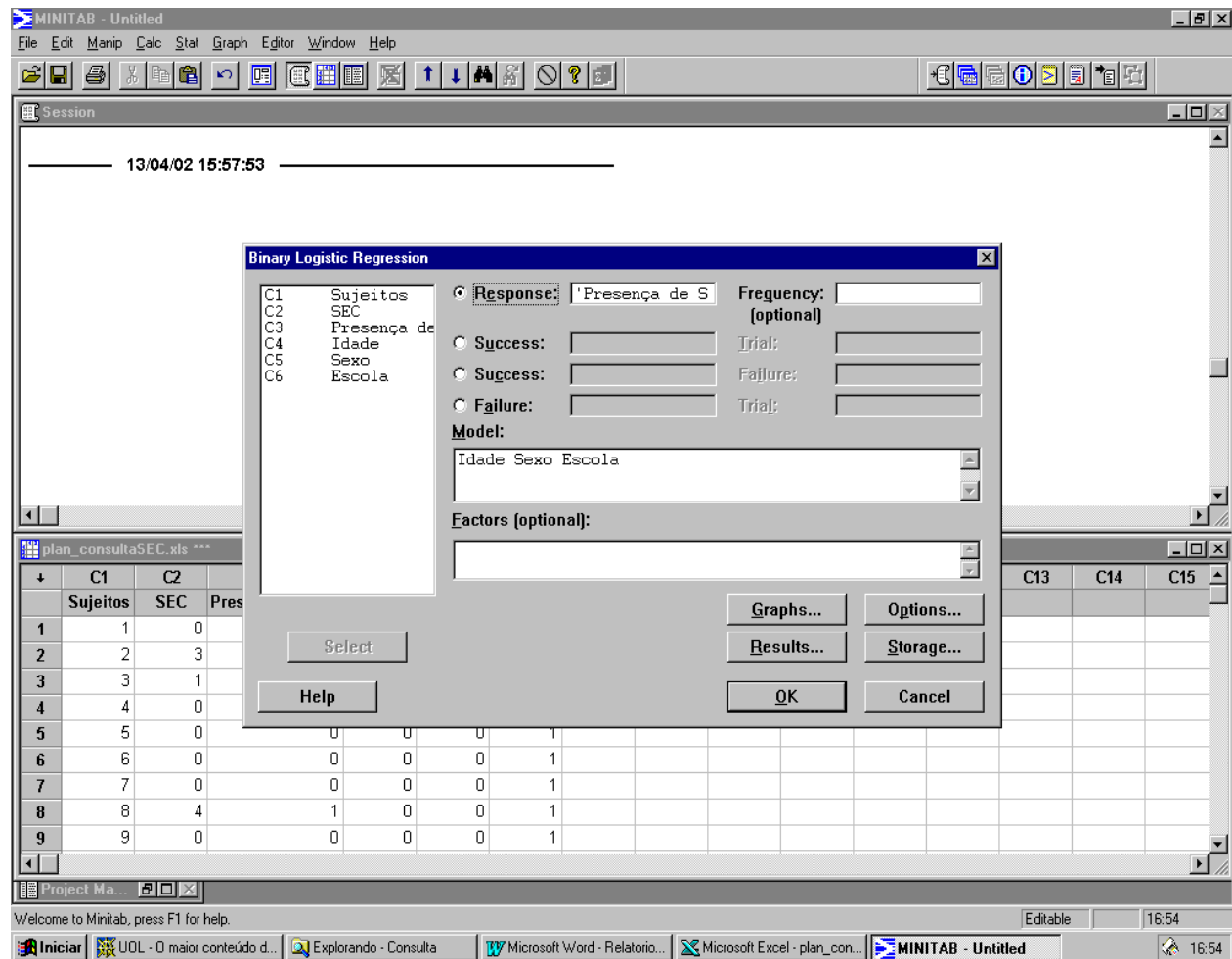
Insira os dados das variáveis explicativas e da variável resposta de interesse no estudo na planilha do Minitab de forma que na primeira linha (células cinzas) estejam os nomes das variáveis. Para efetuar a análise de regressão logística selecione dentro do menu Stat a opção Regression e depois Binary Logistic Regression, como indica a Figura A.1.

Figura A.1. Procedimento de inserção de dados e localização da ferramenta de análise de regressão logística no Minitab.



Após a seleção da opção desejada, uma janela será aberta na qual deve-se selecionar a variável resposta a ser analisada e as variáveis explicativas do estudo. Insira a variável resposta no campo Response e as variáveis explicativas no campo Model. Para inserir a variável, basta selecionar o campo desejado e escolher a(s) variável(is) desejada(s) listadas na caixa maior e clicar em Select, como mostra a Figura A.2.

Figura A.2. Procedimento de execução da análise de regressão logística no Minitab.



Clicando em OK os resultados aparecerão na janela Session localizada na parte superior da tela.

Apêndice B
(Resultados obtidos no ajuste da Regressão Logística)

Seguindo as instruções do Apêndice A, fizemos a análise considerando o modelo com todas as variáveis explicativas e interações possíveis e com a variável resposta **presença de SEC**. O modelo de regressão logística é:

$$P(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_2 X_3 + \beta_6 X_1 X_3 + \beta_7 X_1 X_2 X_3)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_2 X_3 + \beta_6 X_1 X_3 + \beta_7 X_1 X_2 X_3)}$$

onde:

Y_i : **presença de SEC**

X_1 : **idade** (0 = 7 anos completos e 1 = 8 anos completos)

X_2 : **sexo** (0 = masculino e 1 = feminino)

X_3 : **escola** (0 = particular e 1 = pública)

β_i : parâmetros desconhecidos do modelo, $i = 0, 1, \dots, 7$

A interação, quando presente no modelo, significa que o comportamento médio da variável resposta é influenciada pelas outras variáveis contidas na interação. Por exemplo, a interação entre **idade** e **escola** significa que o comportamento médio da probabilidade da **presença de SEC** para cada tipo de escola é influenciado pela idade das crianças.

Os resultados obtidos pelo Minitab estão apresentados nos quadros a seguir:

Quadro B.1. Saída da regressão logística do Minitab para o modelo completo.

Binary Logistic Regression: Presença de SEC versus Idade. Sexo. Escola								
Response Information								
Variable	Value	Count						
Presença	1	15	(Event)					
	0	65						
	Total	80						
Logistic Regression Table								
Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI		
Constant	-22	10395	-0,00	0,998				
Idade	20	10395	0,00	0,998	7,34E+08	0,00		*
Sexo	-0	14701	-0,00	1,000	1,00	0,00		*
Escola	21	10395	0,00	0,998	1,26E+09	0,00		*
Idade*Sexo	-20	18005	-0,00	0,999	0,00	0,00		*
Sexo*Escola	1	14701	0,00	1,000	2,33	0,00		*
Idade*Escola	-21	10395	-0,00	0,998	0,00	0,00		*
Idade*Sexo*Escola	20	18005	0,00	0,999	5,40E+08	0,00		*
Test that all slopes are zero: G = 18,899. DF = 7. P-Value = 0,009								

Analisando o Quadro B.1, notamos que o modelo com todas as interações apresenta valores elevados para os níveis descritivos (acima de 0,998), o que indicaria que todos os termos são não significantes, ou seja, que todas as variáveis juntas têm pouca importância na explicação da probabilidade da **presença de SEC**. Porém, podemos observar que o teste em que todos os coeficientes são iguais a zero (P-valor = 0,009) indica que pelo menos um dos termos é diferente de zero, ou seja, pelo menos um dos termos é significativo. Como no modelo existem termos de interação não significantes devemos retirá-los um a um (sendo que em cada retirada de um termo realiza-se a análise novamente) e verificar se ainda restam interações não significantes. Caso existam retira-se outro termo de interação e realiza-se a análise mais uma vez. Nesse estudo, não havia efeito de interação no modelo – realizando o procedimento descrito anteriormente, todas se mostraram não significantes.

Retirando todas as interações, restou o modelo com as três variáveis explicativas, que é apresentado no Quadro B.2.

Quadro B.2. Saída da regressão logística do Minitab para o modelo com as três variáveis explicativas.

Binary Logistic Regression: Resposta versus Idade; Sexo; Escola

Response Information

Variable	Value	Count	
Presença	1	15	(Event)
	0	65	
	Total	80	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-2,9524	0,8461	-3,49	0,000			
Idade	-0,1879	0,6138	-0,31	0,760	0,83	0,25	2,76
Sexo	0,1879	0,6138	0,31	0,760	1,21	0,36	4,02
Escola	2,2184	0,8010	2,77	0,006	9,19	1,91	44,19

Analisando o Quadro B.2 notamos que os altos valores dos níveis descritivos (P-valor=0,760) indicam que as variáveis **idade** e **sexo** têm pouco efeito na explicação da probabilidade da **presença de SEC**. Assim, optamos por retirar uma dessas variáveis a fim de verificar se a retirada da variável escolhida influencia no comportamento das restantes, ou seja, se existe um efeito conjunto entre a variável que está sendo retirada e as demais.

Quadro B.3. Saída da regressão logística do Minitab excluindo a variável **sexo**.

Binary Logistic Regression: Presença de SEC versus Idade; Escola

Response Information

Variable	Value	Count
Presença	1	15 (Event)
	0	65
	Total	80

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-2,8546	0,7781	-3,67	0,000			
Idade	-0,1876	0,6133	-0,31	0,760	0,83	0,25	2,76
Escola	2,2160	0,8006	2,77	0,006	9,17	1,91	44,04

Observando o Quadro B.3 notamos que os coeficientes relacionados às variáveis restantes não apresentaram uma alteração muito grande (em relação ao Quadro B.2) com a retirada da variável **sexo**. No entanto, a variável **idade** continua apresentando um alto nível descritivo. Pelo mesmo motivo que nos levou a retirar a variável **sexo** do modelo, optamos também pela retirada da variável **idade**.

Quadro B.4. Saída da regressão logística do Minitab excluindo as variáveis **sexo** e **idade**.

Binary Logistic Regression: Presença de SEC versus Escola

Response Information

Variable	Value	Count	
Presença	1	15	(Event)
	0	65	
Total		80	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
Constant	-2,9444	0,7255	-4,06	0,000			
Escola	2,2136	0,8002	2,77	0,006	9,15	1,91	43,90

Através do Quadro B.4, percebemos que novamente não houve alterações nos valores dos coeficientes. Notamos também que a variável restante, **escola**, apresentou um nível descritivo baixo. Então o modelo final é :

$$\hat{P}_i = \frac{\exp(-2,9444 + 2,2136 \cdot X_3)}{1 + \exp(-2,9444 + 2,2136 \cdot X_3)}$$

Assim, concluímos que o modelo com apenas a variável **escola** é suficiente para explicar a probabilidade da **presença de SEC** e, dessa forma, mantivemos apenas essa variável no modelo. Então, baseado na estimativa da razão de chances (Odds Ratio) desse quadro, podemos concluir que a chance de uma criança que estuda em

escola pública apresentar o processo SEC é 9,15 vezes a chance de uma criança que estuda em escola particular.

A razão de chances é calculada dividindo o quociente entre a probabilidade da criança apresentar processo SEC dado que ela estuda em escola pública e a probabilidade da criança não apresentar processo SEC dado que ela estuda em escola pública pelo quociente entre a probabilidade da criança apresentar processo SEC dado que ela estuda em escola particular e a probabilidade da criança não apresentar processo SEC dado que ela estuda em escola particular. Ou seja:

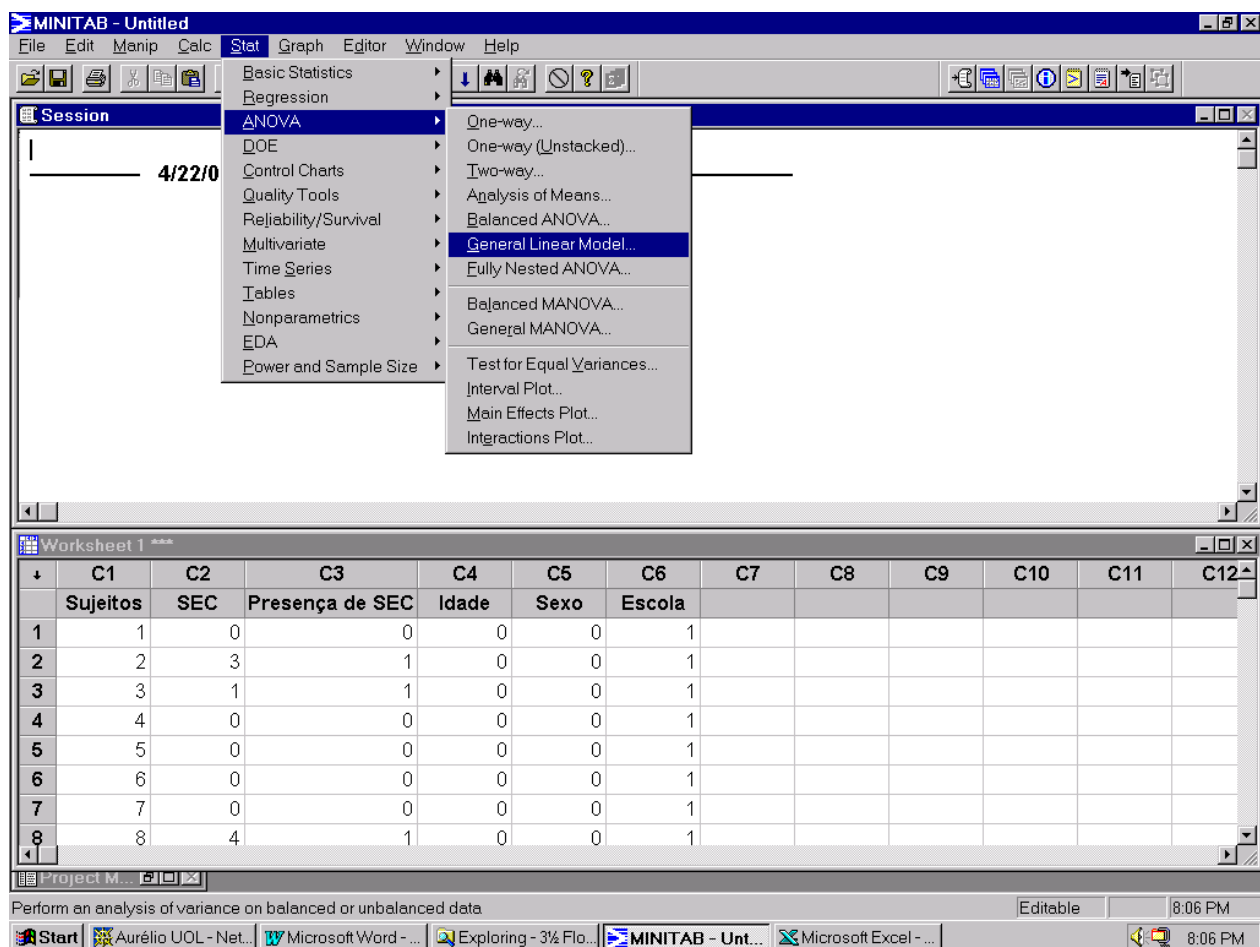
$$\text{Razão de chances} = \frac{\frac{P(\text{apresenta SEC} | \text{estuda em escola pública})}{1 - P(\text{apresenta SEC} | \text{estuda em escola pública})}}{\frac{P(\text{apresenta SEC} | \text{estuda em escola particular})}{1 - P(\text{apresenta SEC} | \text{estuda em escola particular})}}$$

Apêndice C
(Análise de Variância no Minitab)

Para esta análise, o software utilizado foi o Minitab (versão 13.1), pois ele contém todos os recursos necessários.

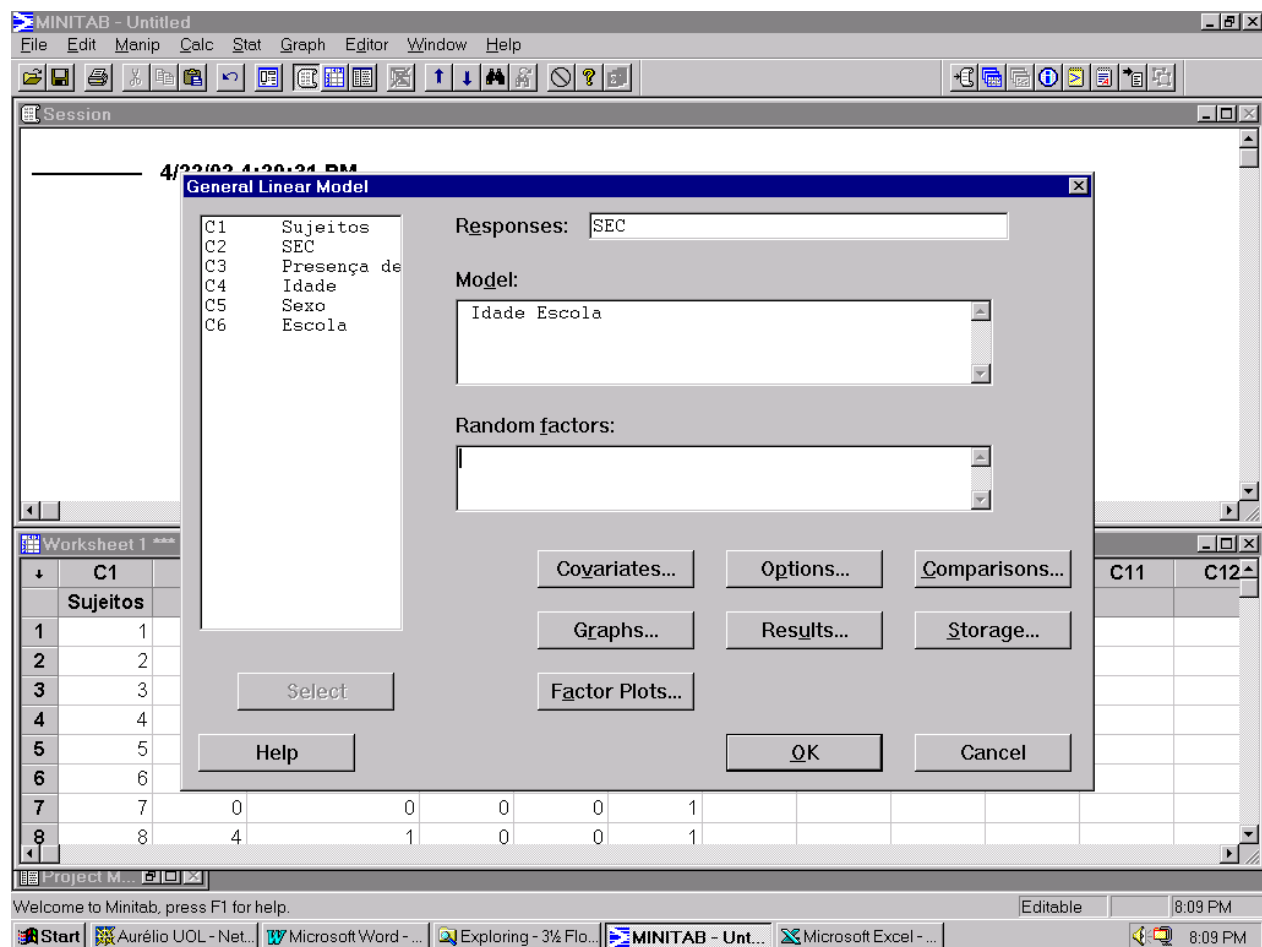
Insira os dados das variáveis explicativas e da variável resposta de interesse no estudo na planilha do Minitab de forma que na primeira linha (células cinzas) estejam os nomes das variáveis. Para efetuar a análise de variância para modelos lineares gerais selecione dentro do menu Stat a opção ANOVA e depois General Linear Model, como indica a Figura C.1.

Figura C.1. Procedimento de execução da análise de variância para modelos lineares gerais no Minitab.



Após a seleção da opção desejada, uma janela será aberta na qual deve-se selecionar a variável resposta a ser analisada e as variáveis explicativas do estudo. Insira a variável resposta no campo Response e as variáveis explicativas no campo Model. Para inserir a variável, basta selecionar o campo desejado e escolher a(s) variável(is) desejada(s) listadas na caixa maior e clicar em Select, como mostra a Figura C.2.

Figura C.2. Procedimento de execução da análise de variância para modelos lineares gerais.



Apêndice D
(Resultados obtidos da Análise de Variância)

No estudo, queremos comparar o número médio de processos SEC em relação à idade das crianças e escola em que elas estudavam. Para essa verificação utilizou-se um modelo de análise de variância que incluiu um termo de interação entre as variáveis explicativas **idade** e **escola**. A interação significa que o comportamento médio da variável resposta SEC para cada tipo de escola é influenciado pela idade das crianças. Através do Quadro D.1 notamos que a interação não é significativa (P-valor = 0,110), possibilitando o ajuste de um modelo mais simples.

Quadro D.1. Saída da análise de variância do Minitab com os fatores idade e escola considerando o termo de interação entre eles.

General Linear Model: SEC versus Idade, Escola						
Factor	Type	Levels	Values			
Idade	fixed	2	0	1		
Escola	fixed	2	0	1		
Analysis of Variance for SEC, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Idade	1	0.8000	0.8000	0.8000	1.16	0.285
Escola	1	7.2000	7.2000	7.2000	10.44	0.002
Idade*Escola	1	1.8000	1.8000	1.8000	2.61	0.110
Error	76	52.4000	52.4000	0.6895		
Total	79	62.2000				

Quadro D.2. Saída da análise de variância do Minitab com os fatores idade e escola.

General Linear Model: SEC versus Idade e Escola						
Analysis of Variance for SEC, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Idade	1	0,8000	0,8000	0,8000	1,14	0,290
Escola	1	7,2000	7,2000	7,2000	10,23	0,002
Error	77	54,2000	54,2000	0,7039		
Total	79	62,2000				

O teste F da análise de variância indica se há evidências suficientes para considerar que todas as médias são iguais.

Através do Quadro D.2, notamos que não há diferença entre número médio de processos SEC em crianças de 7 anos em relação às de 8 anos (P-valor = 0,290) considerando um nível de significância de 5%. Notamos também que há diferença entre o número médio de processos em crianças que estudam em escolas particulares em relação às que estudam em escolas públicas (P-valor = 0,002).