

## **Aviso ao usuário**

Para que o Sisvar 4.3 rode adequadamente use uma configuração de vídeo 600x800 pixels e fontes pequenas 16, True color (24 bits) ou High Color (16 bits). Mensagens de erros podem ser dadas ao usuário após instalação perfeita devido a má especificação da configuração de vídeo. Se o Drive de Vídeo não estiver bem instalado o usuário poderá ter problemas para executar o Sisvar. Certifique-se disso antes de usar o Sisvar.

Esse problema é devido a um objeto da Borland TDBGrid e não do Sistema Sisvar, que herdou tal propriedade desse objeto, que só se permite executar em vídeos de alta resolução e devidamente configurado.

Boa sorte ao usar o Sisvar! Espero que suas expectativas sejam plenamente satisfeitas.

**UNIVERSIDADE FEDERAL DE LAVRAS**  
**DEPARTAMENTO DE CIÊNCIAS**  
**EXATAS**

**Manual do sistema Sisvar para análises**  
**estatísticas**

**Responsável: Professor Dr. Daniel Furtado Ferreira**

**LAVRAS, MG**

**2000**

## SUMÁRIO

	Página
1. Estatística descritiva	1
1.1. Apresentação de dados	1
1.2. Medidas de posição e de dispersão	6
2. Cálculo de probabilidades	8
3. Estimação de parâmetros	17
3.1. Estimação de médias, diferenças de médias, proporções, coeficientes de variação, variâncias e desvios padrões	17
3.2. Estimação de diferenças de médias	22
3.3. Estimação de proporções	27
3.4. Estimação de coeficientes de assimetria e curtose	28
4. Teoria da decisão	30
4.1. Teste sobre média	30
4.2. Teste sobre diferença de médias	32
4.3. Teste sobre proporções	35
4.4. Teste sobre variâncias	36
5. Análise de modelos de regressão linear	38
6. Análise de variância	45
6.1. Delineamento inteiramente casualizado	45
6.1.1. Regressão na análise de variância	58
6.2. Delineamento em blocos casualizados	63

# 1. Estatística descritiva

A estatística é um ramo da matemática que se interessa em obter conclusões a partir de dados observados e nos métodos científicos para coleta, organização, resumo, apresentação, análise e interpretação dos dados amostrais. A estatística se divide em estatística descritiva e indutiva (ou inferência). A estatística descritiva preocupa-se com a coleta, organização e apresentação dos dados amostrais, sem inferir sobre a população; e a estatística indutiva preocupa-se com a análise e interpretação dos dados amostrais. A inferência não pode ser "absolutamente certa", daí a necessidade de se utilizar uma linguagem de probabilidade.

Na maioria dos problemas as leis de causa e efeito não são conhecidas pelo pesquisador, no entanto existe a necessidade de se obter uma solução objetiva. Foi esse objetivo, ou seja, o de apresentar tais soluções, é que a estatística se desenvolveu, face as incertezas oriundas da variabilidade dos dados provenientes das observações dos pesquisadores. Para que seja bem usada é necessário que se conheça os seus fundamentos e os seus princípios, e acima de tudo que o pesquisador desenvolva um espírito crítico sobre a pesquisa empreendida.

A utilização de softwares para atender tais necessidades é de suma importância, uma vez que necessita-se de realizar laboriosos cálculos na análise dos dados oriundos das pesquisas, bem como os resultados obtidos têm que ser exatos e acurados. No entanto, os programas de estatísticas disponíveis, necessitam ser bem usados, para cumprir adequadamente o seu papel. O mal uso dos recursos computacionais, bem como, da interpretação de seus resultados foi a motivação para a redação desse material. Nas próximas seções desse material serão enfatizados o programa de análise estatística Sisvar, mostrando suas qualidades e acima de tudo, as maneiras do uso correto de suas funções e comandos.

## 1.1. Apresentação de dados

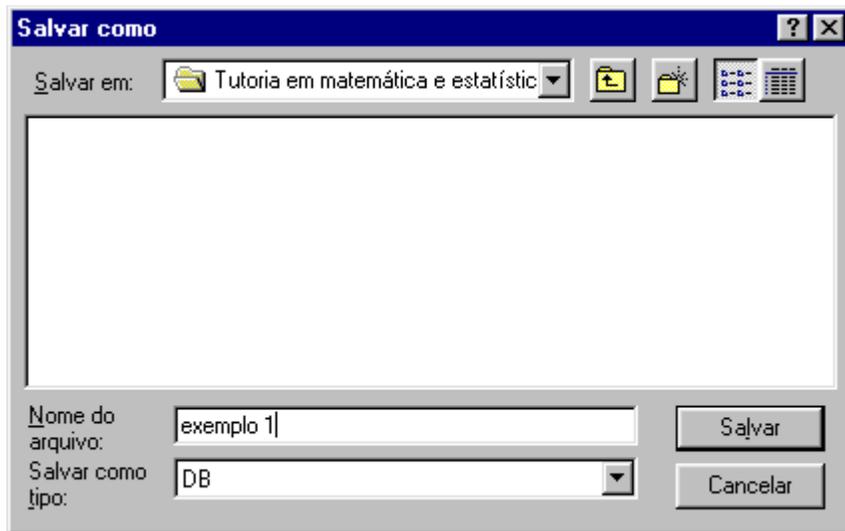
A observação de dados em experimentos ou em levantamentos amostrais se constitui numa das etapas da pesquisa científica. A principal característica dessas observações ou conjunto de dados é a variabilidade inerente aos mesmos, que é a principal razão do desenvolvimento da estatística como ciência. De posse desses dados o pesquisador poderá extrair uma série de informações úteis para a sua pesquisa por meio de uma avaliação exploratória desse conjunto de observações. Essa análise

exploratória depende do tipo e da natureza dos dados. Os dados podem ser qualitativos ou quantitativos, sendo que os quantitativos subdividem ainda em discretos, aqueles que são provenientes de contagens (valores inteiros), e em contínuos, ou seja, os que podem assumir quaisquer valores reais dentro de um certo intervalo. A este último tipo de dados, é que se dará maiores atenções, por se tratarem dos mais freqüentes nas pesquisas. A principal forma de se apresentar um grande conjunto de observações originárias da pesquisa é por meio das tabelas de distribuição de freqüência ou dos gráficos de barras conhecidos por histogramas, ou ainda, pelos gráficos conhecidos por polígono de freqüência. É possível extrair uma série de informações desse tipo de representação, que já foram vistas nos demais módulos desse curso. Será dado ênfase na forma de se obter tais representações por meio do Sisvar.

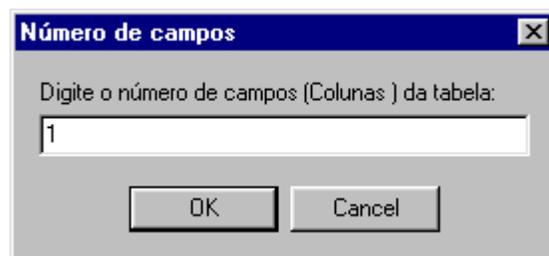
O primeiro exemplo que será usado para ilustrar trata-se da amostragem de 20 plantas de feijoeiro cultivar Carioca, nas quais se observou a severidade da doença cujo agente causal é o fungo *Colletotrichum lindemuthianum*. Os dados estão apresentados a seguir:

1,71	2,34	0,98	0,21
0,24	0,21	0,35	0,35
5,14	1,33	1,68	0,21
5,11	4,44	3,50	5,95
5,28	2,83	0,91	6,26

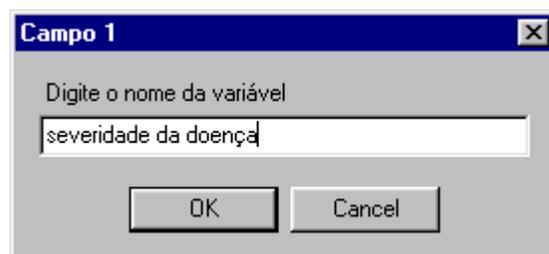
Para se utilizar o Sisvar é possível entrar com os dados de imediato via teclado, durante a execução da análise pretendida, ou então, via arquivo de dados. A segunda forma é a melhor maneira, uma vez que os dados ficam disponíveis para a realização de novas análises que venham a se tornar necessárias. Será descrito a seguir a forma de entrar com os dados via arquivo de dados. É conveniente chamar a atenção de que o Sisvar permite que dois tipos de campos (colunas) distintos no arquivo: alfanuméricas (texto) e as numéricas (inteiros ou reais). Dados de cada variável tomadas na pesquisa deverão ocupar um campo (coluna) do arquivo. Nesse exemplo tem-se apenas uma variável. Para criar o arquivo, escolhe a opção de manipulação de arquivos do programa e em seguida no menu principal escolhe-se a opção criar. Escolhe-se um nome para o mesmo e o diretório de trabalho conforme esquema apresentado a seguir:



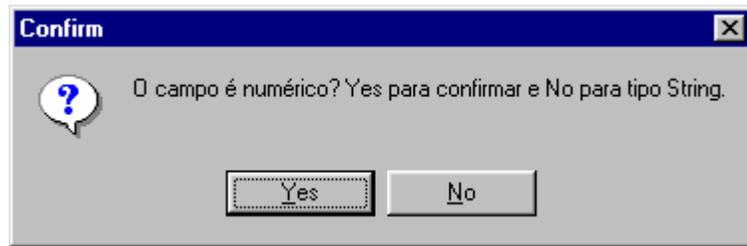
Nesse caso, escolheu-se o diretório de trabalho e o nome do arquivo: exemplo 1, cuja extensão é DB. O arquivo possui estrutura do tipo Paradox. Em seguida escolhe-se o número de campos do arquivo, que nesse caso será 1.



Em seguida escolhe-se o nome do campo 1, que nesse caso foi severidade da doença:



O próximo passo exige que se confirme o tipo do campo. O usuário responderá em sim (Yes) para campo numérico ou não para campo não numérico (texto). Nesse exemplo a resposta é afirmativa.



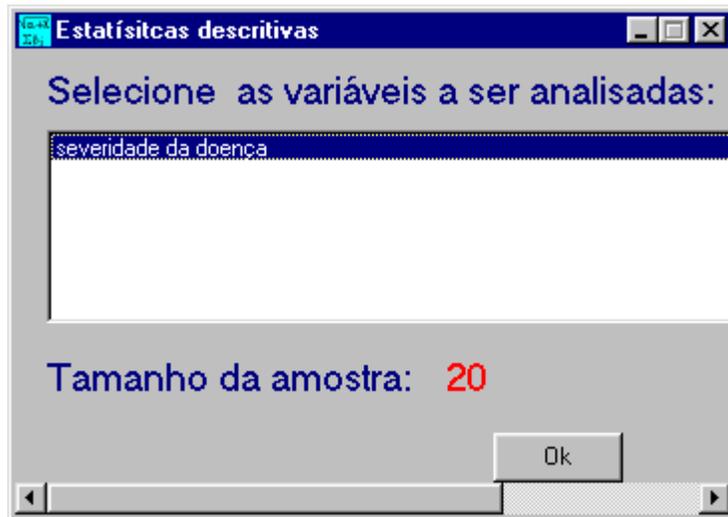
A planilha assim estrutura fica pronta para a digitação dos dados, que é feita na próxima etapa pelo usuário. Após a digitação o arquivo poderá ser analisado pelo programa. Qualquer outro arquivo assim criado poderá ser usado para análises estatísticas posteriores. A parte final do arquivo em questão está apresentado a seguir.



The image shows a software window titled "Arquivos" with a menu bar containing "Arquivo", "Importar", "Exportar", "Registros", and "Campos". Below the menu bar is a table with a single column header "severidade da doença". The table contains several rows of numerical values.

severidade da doença
3.5
0.91
0.21
0.35
0.21
5.95
8.26

Em seguida na tela (formulário) principal do Sisvar, escolhe-se o menu de análises e dentro do mesmo a opção de estatística descritiva. Responde-se adequadamente a forma de entrada dos dados, que nesse caso é via arquivo, e em seguida seleciona-se as variáveis que serão analisadas. É possível pressionando a tecla Ctrl selecionar múltiplas variáveis com o mouse. No exemplo apenas uma variável pode ser selecionada. Teclando Ok o resultado aparecerá em seguida, para que sejam feitas alterações no gráfico (histograma e polígono de frequência) gerado.



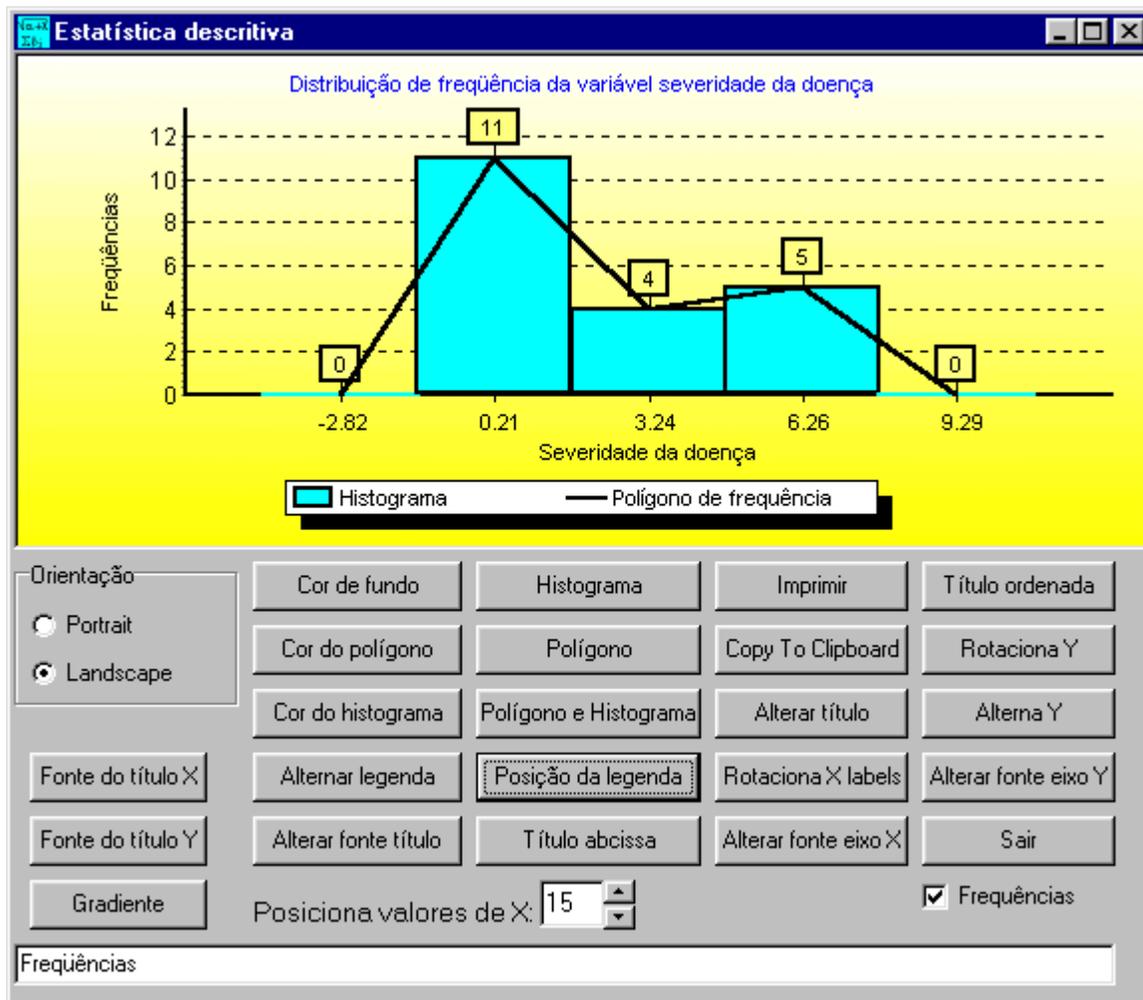
Com as opções do formulário a seguir é possível alterar as características do gráfico recém gerado. Essas propriedades incluem cor de fundo, gradiente, legendas, títulos dos eixos e do gráfico, cor e tipo de fonte dos títulos. É possível imprimir ou copiar para a área de transferência o gráfico, o qual poderá ser colado em um editor de texto apropriado.

A tabela de distribuição de frequência, construída pelo Sisvar, e usada na construção dos gráficos está apresentada a seguir. Destaca-se as classes e as frequências absolutas, percentuais ou relativas das mesmas.

**Saída 1 do Sisvar:**  
distribuição de frequências - Critério de Scott(1979)

LI	LS	Ponto Médio	FOi	Fri
-1.30250  ---	1.72250	0.21000	11.00000	0.55000
1.72250  ---	4.74750	3.23500	4.00000	0.20000
4.74750  ---	7.77250	6.26000	5.00000	0.25000

Em seguida apresenta-se o formulário com os gráficos (histograma e polígono) gerados.



Por esses resultados pode se concluir que se trata de uma distribuição com uma certa assimetria a direita, indicando que a grande maioria das plantas tem baixa severidade da doença.

## 1.2. Medidas de posição e de dispersão

Muitas vezes é necessário apresentar medidas que sintetizem as informações de uma grande massa de dados, de tal sorte que não se perca o valor dos dados originais. Essas medidas se dividem em medidas de posição (médias, medianas e modas, que dão idéias da posição central representativa do conjunto de dados, e em medidas de dispersão (variâncias, coeficientes de variação,

erro padrão da média) que fornecem informações da variabilidade dos dados em torno desse valor central. Inicialmente será apresentado os resultados do exemplo número 1, da severidade da antracnose em feijão, pelo Sisvar. Realizando os procedimentos para a obtenção dos gráficos, descritos anteriormente, é possível simultaneamente obter os resultados apresentados a seguir sobre as medidas de posição e dispersão da amostra relativa a severidade de doença.

Saída 2 do Sisvar:	
Estatísticas descritivas	
n:	20
Média:	2.451500
Variância:	4.736319
Desvio padrão:	2.176308
Coeficiente variação:	88.774567
Erro padrão da média:	0.486637
Coef. de assimetria:	0.501291
Coef. de Curtose:	1.715168
Mínimo:	0.210000
Máximo:	6.260000
Amplitude total:	6.050000
Mediana:	1.447500
Moda:	0.546111

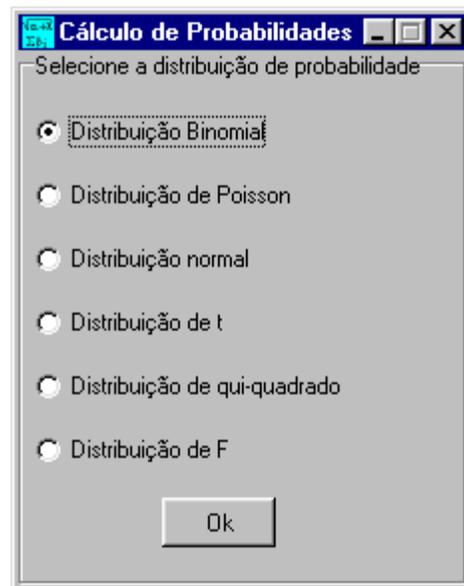
Dentre as estatísticas descritivas apresentadas na saída 2 do Sisvar, destacam-se também, as medidas de assimetria e de curtose da distribuição. Tomando-se por base a distribuição normal ou gaussiana, cujos coeficientes de assimetria e curtose são iguais a 0 e 3, verifica-se que a distribuição tem assimetria positiva ou a direita (cauda longa a direita) e é platicúrtica (mais achatada que a normal, pois o coeficiente é menor que 3). Distribuições com coeficiente de curtose iguais a 3 são ditas mesocúrticas e aquelas com coeficiente maiores que 3 são ditas leptocúrticas (mais afiladas que a distribuição normal). O erro padrão da média, por sua vez, é uma medida da precisão da estimativa desta. Finalmente, para distribuições assimétricas, muitos estatísticos optam pelo uso de medidas de posição que não seja a média, mas por aquelas que são mais adequadas para essas situações como a mediana e a moda.

## 2. Cálculo de probabilidades

Muitas vezes na pesquisa torna-se necessário obter valores probabilísticos de algumas distribuições ou até mesmo obter quantis superiores que são úteis para definir regiões críticas para os testes de hipóteses sobre parâmetros populacionais de interesse. Será descrito os procedimentos básicos do Sisvar para a obtenção de probabilidades de interesse para as distribuições Binomial, Poisson, Normal, t de Student, Qui-quadrado e F de Snedecor. Inicialmente se buscará por meio de alguns exemplos ilustrar as formas de obtenção dessas probabilidades ou desses quantis, que poderão ser extrapolados para situações diferentes das exemplificadas.

### A distribuição binomial

No Sisvar os cálculos de probabilidades são feitos escolhendo-se no menu a opção análises e em seguida a opção cálculo de probabilidades. Escolhe-se a distribuição de interesse no seguinte formulário:



Cálculo de Probabilidades

Selecione a distribuição de probabilidade

- Distribuição Binomial
- Distribuição de Poisson
- Distribuição normal
- Distribuição de t
- Distribuição de qui-quadrado
- Distribuição de F

Ok

Nesse caso escolhe-se a opção da distribuição binomial e em seguida defini-se os parâmetros dessa distribuição. Para o caso do nascimento de fêmeas em uma leitegada, sabe-se da genética que o probabilidade do sucesso em um único nascimento do suíno ser fêmea é de 50% ou  $p=0,5$ . Nesse caso pergunta-se qual é a probabilidade de no nascimento de 10 suínos exatamente 3 deles serem fêmeas? Para responder tal probabilidade basta preencher o formulário apresentado a

seguir com as opções nele apresentada e clicar em calcular. Nesse caso  $y=3$  é o evento de interesse do cálculo e  $n=10$  é o tamanho da amostra.

The image shows a software window titled "Distribuição binomial". It contains a section for selecting an event with three radio buttons:  $P(Y = y)$  (selected),  $P(Y > y)$ , and  $P(Y \leq y)$ . To the right, the word "Resultados" is displayed in red above a text box containing the value  $1.17187500000000E-0001$ . Below this is a "Calcular" button. At the bottom, there are three input fields: "Probabilidade de sucessos (P):" with the value 0.5, "Tamanho da amostra (n):" with the value 10, and "Número de sucessos (y):" with the value 3. A "Sair - Relatório" button is located at the bottom right.

A resposta para essa pergunta é 11,72%, ou seja, no nascimento de 10 leitões existe 11,72% de chance de exatamente 3 deles serem fêmeas. Outras probabilidades são disponíveis, bastando optar pelas opções dos botões de rádio no canto superior direito do formulário. O botão Sair-Relatório permite ao usuário ter acesso a todos os cálculos executados, editá-los e modificá-los.

## A distribuição Poisson

As probabilidades da distribuição de Poisson, que é a distribuição limite da binomial quando  $n$  tende a infinito e  $p$  tende a zero, mas o produto  $k=np$  continua finito, podem ser obtidas pelo Sisvar. O valor de  $k$  é o parâmetro (média) dessa distribuição. Para usar o Sisvar, é necessário escolher os mesmos passos da Binomial e marcar a opção Poisson quando essa aparecer no formulário de escolha das distribuições de probabilidades. No formulário a seguir defini-se o valor de  $k$  e a probabilidade desejada com a escolha de  $y$  e do evento,  $P(Y=y)$ ,  $P(Y\leq y)$  ou  $P(Y>y)$ . Para ilustrar, sabendo que uma doença em uma região tem média de 2 casos por ano, qual é a probabilidade de se encontrar exatamente 3 casos em um dado ano? Para obter tal probabilidade siga as instruções do formulário apresentado a seguir:

Distribuição de Poisson

Selecionar a opção

$P(Y = y)$

$P(Y > y)$

$P(Y \leq y)$

**Resultados:**

1.8044704431548E-0001

Calcular

Digite a média de Poisson (k): 2

Digite o valor de y: 3

Sair - Relatório

Verifica-se que o valor dessa probabilidade, ou seja, a chance de um dado ano ter exatamente 3 casos da doença, é de 18,04%.

## A distribuição Normal

A distribuição normal é de suma importância para a experimentação e para a inferência estatística como um todo. Para ilustrar os cálculos de probabilidade e os quantis dessa distribuição, será apresentado um exemplo. Supondo que a distribuição de pesos ao abate aos 90 dias de coelhos híbridos Norfolk é normal, e que a média é de 2,58kg e o desvio padrão de 0,0675kg, pergunta-se: a) qual a probabilidade de que um coelho supere o peso de 2,70kg. b) qual é o peso que deixa 2,5% de pesos acima dele? Para responder a essa pergunta-se, escolhe-se no Sisvar a opção normal e defini-se os parâmetros desta distribuição. Para responder a pergunta da letra (a) deve-se preencher o formulário como apresentado a seguir.

The image shows a software window titled "Distribuição normal" with a blue title bar. Inside the window, there is a section labeled "Escolha a opção" with two radio buttons. The first option, "Cálculo da probabilidade (p), dado xc => P(X > xc) = p", is selected. The second option is "Cálculo dos quantis (xc), dado p => P(X > xc) = p". Below this, there are three input fields: "Média:" with the value 2.58, "Variância:" with the value 0.00455625, and "Valor de xc (real):" with the value 2.70. A "Sair - Relatório" button is located above the "Média" field. Below the input fields, there is a "Resultados:" label and a text box containing the value 0.03772014. A "Calcular" button is at the bottom of the window.

Nesse caso a probabilidade solicitada é de 3,77%. É conveniente salientar que é necessário entrar com a variância (0,00455625) e não com o desvio padrão. Para responder a letra (b) escolhe-se na parte superior do formulário a opção cálculo dos quantis, e preenche o formulário como apresentado a seguir, e clica-se em calcular. O valor do quantil desejado é 2,7122kg.

**Distribuição normal**

Escolha a opção

Cálculo da probabilidade (p), dado  $x_c \Rightarrow P(X > x_c) = p$

Cálculo dos quantis (xc), dado  $p \Rightarrow P(X > x_c) = p$

Sair - Relatório

Média: 2.58

Variância: 0.00455625

Valor de p, tal que  $P(X > x_c) = p$  : 0.025

Resultados: 2.71229757

Calcular

## A distribuição t de Student

Para o cálculo das probabilidades da distribuição t de Student e dos quantis dessa distribuição de probabilidade é necessário especificar os graus de liberdade e os valores de t ou de probabilidades desejadas para a obtenção dos valores pretendidos. Para ilustrar supondo que se deseja calcular (a) a probabilidade de um valor de t superar o valor calculado de 2,10 em uma amostra de tamanho 10 (9 graus de liberdade); (b) o valor de t cuja probabilidade de superá-lo seja de 2,5%, ou seja, qual é o  $t_c$  cuja  $P(t > t_c) = 0,025$ .

**Distribuição de t**

Escolha a opção

Cálculo da probabilidade (p), dado  $t_c \Rightarrow P(T > t_c) = p$   
 Cálculo dos quantis ( $t_c$ ), dado  $p \Rightarrow P(T > t_c) = p$

Sair - Relatório

Graus de liberdade ( $\nu > 0$ ):

Valor de  $t_c$ , tal que  $P(T > t_c) = p$ :

**Resultados:**

Calcular

A probabilidade  $p$  obtida pelo Sisvar na letra (a) é de 3,26%. Para obter o segundo valor, basta seguir as instruções do formulário a seguir. O valor do quantil superior da distribuição de  $t$  que deixa uma probabilidade acima do mesmo de 2,5%, para 9 graus de liberdade, é de 2,26.

**Distribuição de t**

Escolha a opção

Cálculo da probabilidade (p), dado  $t_c \Rightarrow P(T > t_c) = p$   
 Cálculo dos quantis ( $t_c$ ), dado  $p \Rightarrow P(T > t_c) = p$

Sair - Relatório

Graus de liberdade ( $\nu > 0$ ):

Valor de p, tal que  $P(T > t_c) = p$ :

**Resultados:**

Calcular

## A distribuição F de Snedecor

Os valores de probabilidades da distribuição F e os quantis são obtidos conforme descritos para o t, a diferença básica é que dois graus de liberdade são considerados na primeira. No Sisvar para se (a) obter o valor da probabilidade de um valor F superar 3,45 com  $v_1=5$  e  $v_2=10$  graus de liberdade e (b) os quantis da distribuição F com  $v_1=5$  e  $v_2=10$  graus de liberdade, que deixam 5% de chance de obter um valor de F que supere esse quantil, são obtidos conforme os formulários apresentados a seguir.

a) probabilidade

Distribuição F

Escolha a opção

Cálculo da probabilidade (p), dado  $f_c \Rightarrow P(F > f_c) = p$

Cálculo dos quantis ( $f_c$ ), dado  $p \Rightarrow P(F > f_c) = p$

Sair - Relatório

Graus de liberdade do numerador (V1): 5

Graus de liberdade do denominador (V2): 10

Valor de  $f_c$ , tal que  $P(F > f_c) = p$ : 3.45

**Resultados:** 0.04524687

Calcular

Essa probabilidade é  $P(F > 3,45) = 4,52\%$

b) quantil

Distribuição F

Escolha a opção

Cálculo da probabilidade (p), dado  $f_c \Rightarrow P(F > f_c) = p$

Cálculo dos quantis ( $f_c$ ), dado  $p \Rightarrow P(F > f_c) = p$

Sair - Relatório

Graus de liberdade do numerador (V1): 5

Graus de liberdade do denominador (V2): 10

Valor de (p), tal que  $P(F > f_c) = p$ : 0.05

Resultados: 3.32583063

Calcular

Valor do quantil é 3,326.

## A distribuição de qui-quadrado ( $\chi^2$ )

Parecida com a distribuição t de Student, com relação a obtenção de probabilidades e quantis, a distribuição de qui-quadrado tem importante papel na estatística paramétrica e não paramétrica. Para ilustrar: (a) calcular a probabilidade de um valor de  $\chi^2$  superar o valor de 4,8 numa distribuição com 4 graus de liberdade?; (b) qual é o quantil que deixa 5% dos valores de  $\chi^2$  acima do mesmo, com 4 graus de liberdade?

Primeiramente será utilizado o Sisvar para obter as respostas da letra (a) e (b), respectivamente. No caso da primeira questão, basta preencher o formulário da seguinte forma:

**Distribuição de qui-quadrado**

Escolha a opção

Cálculo da probabilidade (p), dado  $x_{2c} \Rightarrow P(X^2 > X_{2c}) = p$   
 Cálculo dos quantis ( $x_{2c}$ ), dado  $p \Rightarrow P(X^2 > x_{2c}) = p$

Sair - Relatório

Graus de liberdade ( $\nu > 0$ ):

Valor de  $x_{2c}$ , tal que  $P(X^2 > x_{2c}) = p$ :

**Resultados:**

Calcular

Portanto a probabilidade desejada é de 30,84%. Já para a segunda questão, tem-se:

**Distribuição de qui-quadrado**

Escolha a opção

Cálculo da probabilidade (p), dado  $x_{2c} \Rightarrow P(X^2 > X_{2c}) = p$   
 Cálculo dos quantis ( $x_{2c}$ ), dado  $p \Rightarrow P(X^2 > x_{2c}) = p$

Sair - Relatório

Graus de liberdade ( $\nu > 0$ ):

Valor de p, tal que  $P(X^2 > x_{2c}) = p$ :

**Resultados:**

Calcular

Cujo valor de qui-quadrado (quantil superior de 5%) é 9,488.

### 3. Estimação de parâmetros

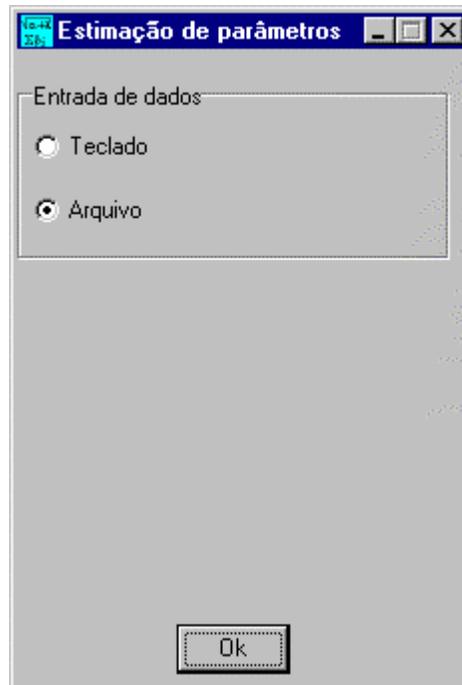
Na inferência estatística uma das principais técnicas usadas é o da estimação de parâmetros por intervalo. Baseado no estimador pontual de um determinado parâmetro é construído um intervalo, cuja probabilidade de que o mesmo contenha o valor real é de  $100(1-\alpha)\%$ , para um valor pré fixado do valor de  $\alpha$ . Em seguida serão apresentados para os parâmetros usuais da pesquisa científica os principais mecanismos de obtenção dos intervalos de confiança por meio do Sisvar. É conveniente salientar que os valores de  $\alpha$  não podem ser fixados o menor possível para aumentar a probabilidade do intervalo conter o valor real. Isso por que os resultados do intervalos podem fornecer valores esdrúxulos.

#### 3.1. Estimação de médias, coeficientes de variação, variâncias e desvios padrões

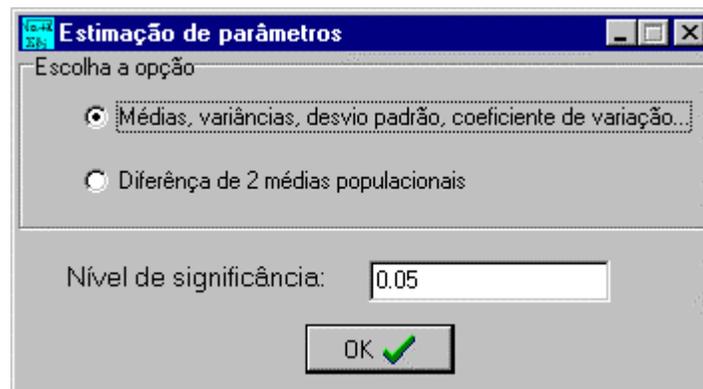
No Sisvar a estimação dos principais parâmetros pode ser feita por duas formas distintas. A primeira delas é feita baseada nos dados amostrais digitados em um arquivo, e a segunda é feita a partir da digitação das estimativas amostrais em um formulário próprio do Sisvar. Para ilustrar o primeiro caso, será usado o exemplo 1, da antracnose no feijoeiro. Para isso escolhe-se as opções do menu: análises e estimação de parâmetros, conforme apresentado a seguir.



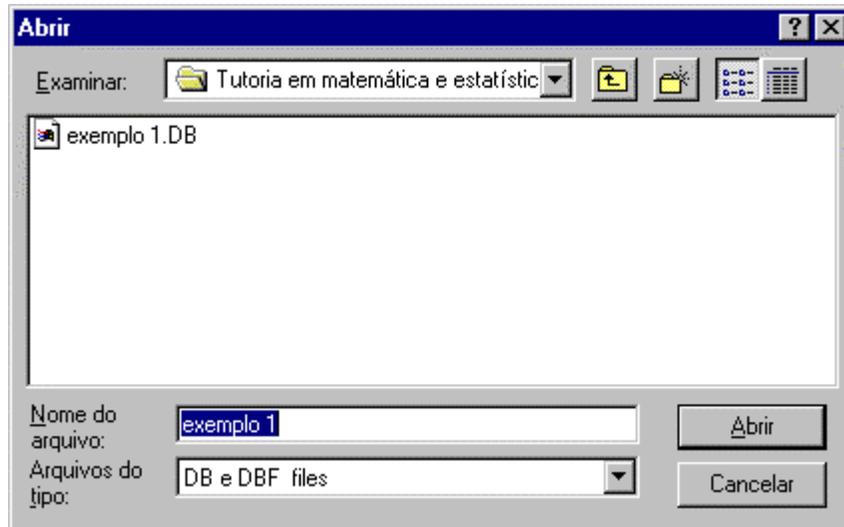
No formulário seguinte, escolhe-se a forma de entrada dos dados via arquivo e tecla-se em Ok.



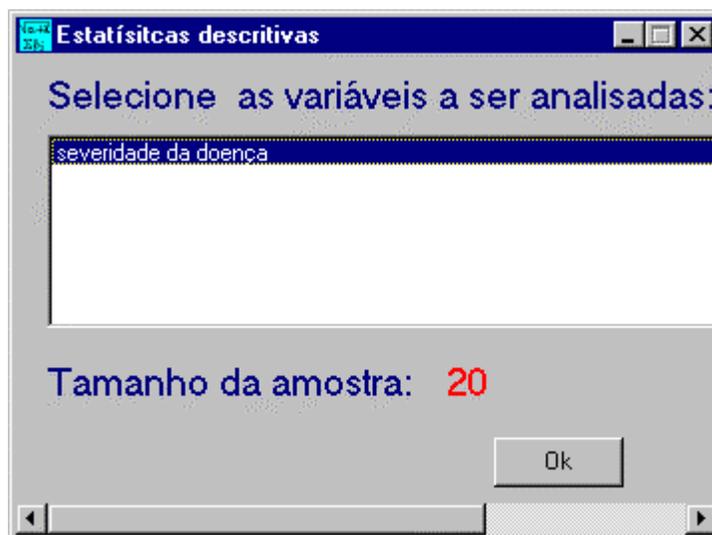
No passo seguinte escolhe-se no formulário apresentado a opção de estimação de médias, variâncias, etc. e fixa o valor de  $\alpha$ .



Escolhe-se o arquivo que será analisado no formulário de diálogo, conforme o exemplo apresentado a seguir:



E finalmente seleciona as variáveis que serão analisadas no próximo formulário. Para seleção múltipla de variáveis basta pressionando a tecla Ctrl usar o mouse para marcar as variáveis de interesse.



O resultado será lançado em um formulário que possui grandes recursos de um editor completo, compatível com diversos programas do Windows. A seguir serão apresentados os resultados fornecidos separados por parâmetros.

### Estimação de média

Média amostral:	2.451500
Tamanho da amostra:	20
Coefficiente de Confiança (%):	95.00
Intervalo de confiança: [LI; LS]	
[	1.432956; 3.470044]

O intervalo apresentado [1,43; 3,47] possui 95% de chance de conter o verdadeiro valor da média populacional.

### Estimação de variância

Variância amostral:	4.736319
Tamanho da amostra:	20
Coefficiente de Confiança (%):	95.00
Intervalo de confiança: [LI; LS]	
[	2.739229; 10.103844]

Esse intervalo é válido somente para uma amostra retirada de uma população normal.

### Estimação de desvio padrão

Desvio padrão amostral:	2.176308
Tamanho da amostra:	20
Coefficiente de Confiança (%):	95.00
Intervalo de confiança: [LI; LS]	
[	1.655062; 3.178654]

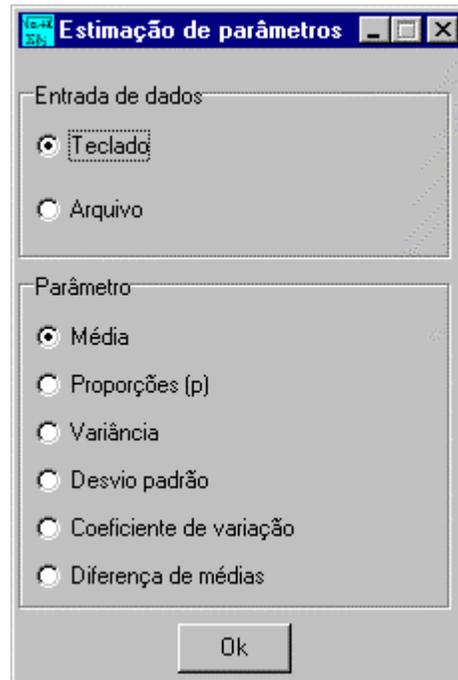
Pode-se facilmente observar que os limites desse intervalo referem-se a raiz quadrada dos limites do intervalo para variância.

### Estimação de coeficiente de variação

Coefficiente de variação(CV) (0, 1):	0.887746
Tamanho da amostra:	20
Coefficiente de Confiança (%):	95.00
Intervalo de confiança: [LI; LS]	
[	0.583548; 2.671331]

Será ilustrado também a opção do Sisvar, usando as estimativas amostrais por ponto, para estimar o IC. Apenas um exemplo será apresentado uma vez que trata-se de um procedimento auto

explicativo. Para isso será usado a estimação de médias. As estimativas por ponto da média é 2,4515 e da variância é 4,736319. Assim para usar o Sisvar, basta escolher a opção teclado para entrar com os dados e a opção média no formulário apresentado a seguir.



Preencher os campos do formulário que se apresentará a seguir e clicar no botão calcular e em seguida no de relatório para ver o resultado ou resultados dos intervalos solicitados. Nesse caso o resultado é idêntico ao obtido anteriormente, como deveria ser.

Média amostral:	2.451500
Tamanho da amostra:	20
Coeficiente de Confiança (%):	95.00
Intervalo de confiança: [LI; LS]	
[	1.432956; 3.470044]

Parâmetro	Valor
Média amostral	2.4515
Tamanho da amostra	20
Nível de significância	0.05
Variância da amostra	4.736319

### 3.2. Estimação de diferenças de médias

No Sisvar existem procedimentos para a estimação dos IC para diferenças de médias. Será apresentado no Sisvar os passos para obtenção do intervalo de confiança para diferença de médias populacionais. Para isso será apresentado um segundo exemplo em que o pesquisador tinha interesse em comparar duas progênies de *Eucalyptus camaldulensis*, referente ao volume de madeira por árvore em  $m^3 \times 10^4$ . Foram obtidas informações de dezoito plantas de cada progênie (Tabela 1).

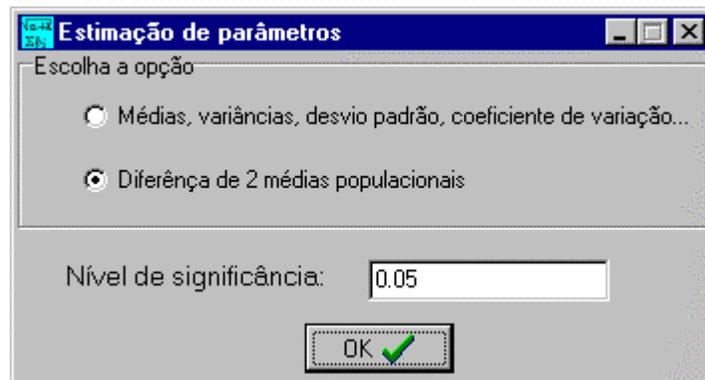
**Tabela 1. Volume de madeira por árvore, em  $m^3 \times 10^4$ , da avaliação de 2 progênies de *E. camaldulensis***

Progênie	ÁRVORES					
	I	II	III	IV	V	VI
1	55	96	212	289	140	142
2	124	230	108	111	46	111
Progênie	ÁRVORES					
	VII	VIII	IX	X	XI	XII
1	218	162	106	124	119	155
2	146	138	194	236	214	116
Progênie	ÁRVORES					
	XIII	XIV	XV	XVI	XVII	XVIII
1	105	38	124	119	59	58
2	218	207	63	146	212	192

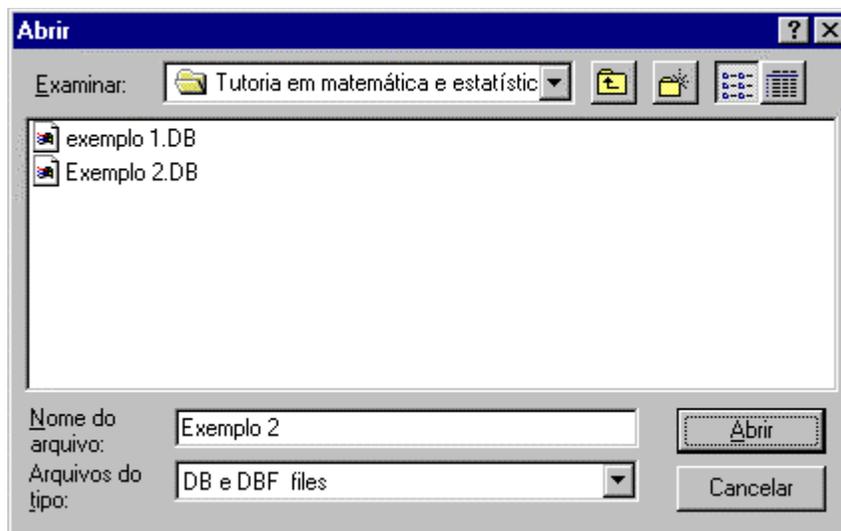
O próximo passo é construir o arquivo de dados, com duas colunas, uma que vai identificar as progênies (com níveis 1 e 2) e outra que vai receber os valores da produtividade das árvores. Parte do arquivo de dados está apresentado a seguir.

Progênies	Volume de madeira $m^3 \times 10^4$
1	38
1	124
1	119
1	59
1	58
2	124
2	230
2	108

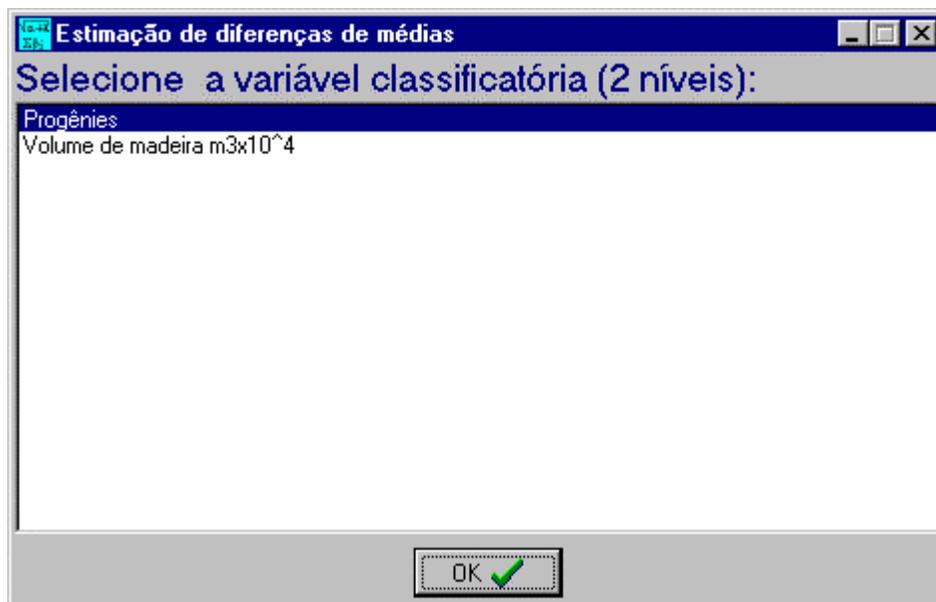
A primeira coluna do arquivo recebe progênies, pode ser alfanumérica ou numérica. No exemplo foi escolhida como numérica. O cuidado que se deve ter é que esta coluna deve ter apenas dois níveis. O próximo passo é escolher estimação de parâmetros, opção arquivo e diferença de médias no último formulário, que está apresentado a seguir.



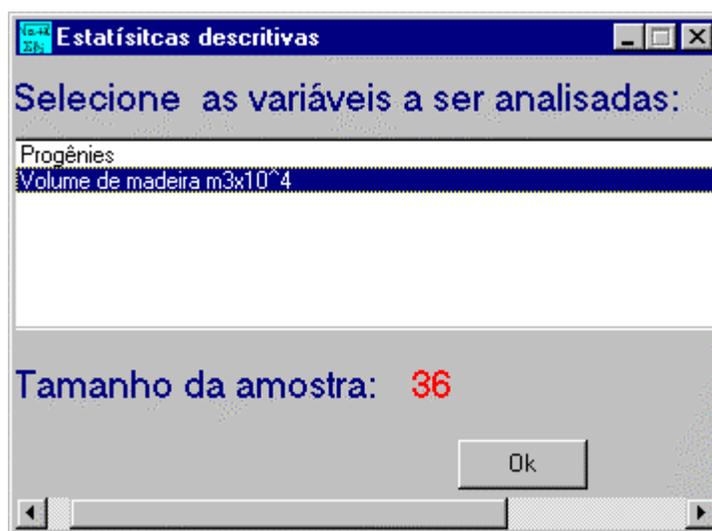
O próximo passo é selecionar o arquivo Exemplo 2.DB que foi criado:



Seleciona-se a variável classificatória (progênes) a qual deve possuir dois níveis e em seguida as variáveis que serão analisadas.



Variáveis para serem analisadas (use a tecla Ctrl para múltiplas seleções):



A primeira parte da saída do programa é um teste para verificar se as variâncias são ou não homogêneas. Esse teste, no exemplo em questão, afirma que as variâncias são homogêneas, uma

vez que a significância do teste F (0,36) é maior que  $\alpha/2$ . Assim, as estatísticas descritivas e o teste para igualdade das variâncias populacionais, estão apresentadas a seguir:

```

População 1 :1
População 2 :2
Variável classificatória Progênies
Variável Volume de madeira m3x10^4
-----
Estatísticas amostrais
-----
Pop.          Média          Variância  Tamanho da amostra
-----
1            128.944444      4035.702614      18
2            156.222222      3400.653595      18
-----
Teste da hipótese de igualdade das variâncias populacionais (1 e 2)
-----
Fc=          1.187 V1=          17 V2=          17
P(F* >      1.187)=          0.3641
Ao nível de 5% de significância as variâncias podem ser consideradas iguais
-----

```

Em seguida dois resultados são disponibilizados, os que consideram as variâncias homogêneas (intervalo exato) e os que as consideram heterocedásticas (intervalo usando a aproximação de Satterthwaite, 1946). Cabe ao usuário escolher a opção adequada, cuja decisão deve ser baseada no teste para variâncias apresentados anteriormente. No exemplo o leitor deve optar pelo intervalo que considera as variâncias iguais (homogêneas).

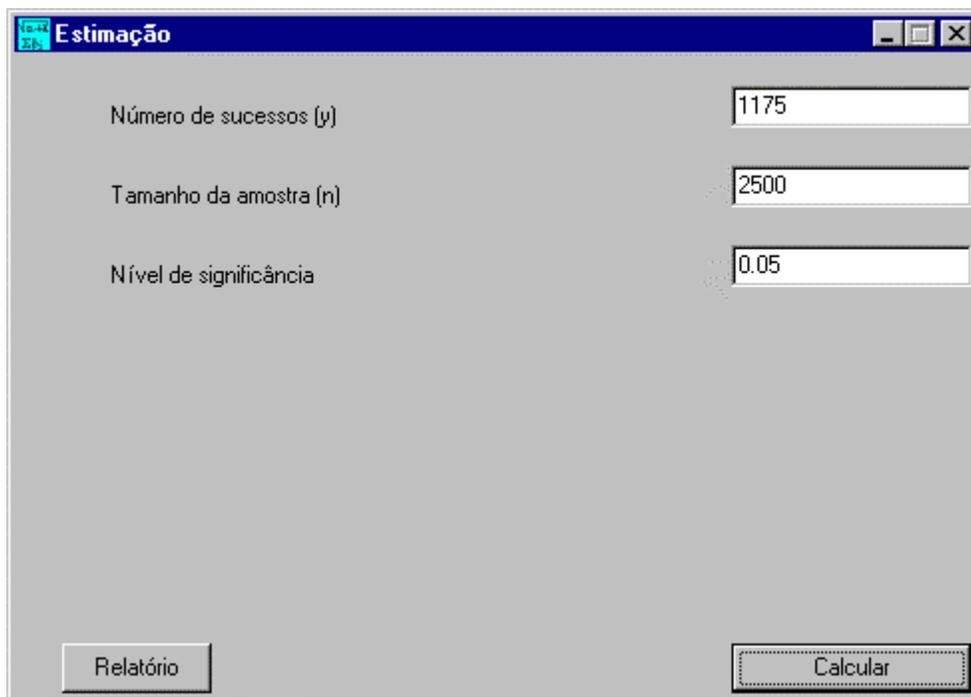
```

Situação em que as variâncias são consideradas homogêneas
-----
Estimativa por ponto de m1-m2:          -27.27778
Variância combinada (pooled):          3718.178105
Graus de liberdade:                     34
Coeficiente de confiança:              95.00000
Intervalo de confiança para m1-m2: [LI; LS]
[          -68.584412;          14.028856]
-----
Situação em que as variâncias não são consideradas homogêneas
-----
Estimativa por ponto de m1-m2:          -27.27778
Graus de liberdade:                     34
Coeficiente de confiança:              95.00000
Intervalo de confiança para m1-m2: [LI; LS]
[          -68.584412;          14.028856]
-----

```

### 3.3. Estimação de proporções

O Sisvar apresenta diversas alternativas para a estimação de proporções. Para ilustrar esse processo vamos imaginando que o IBOPE ao realizar uma pesquisa eleitoral em 2.500 eleitores, verificou que 1.175 deles eram favoráveis ao candidato Beltrano da Silva. Estime por intervalo, usando um coeficiente de confiança de 95%, a verdadeira proporção de eleitores favoráveis ao candidato Beltrano. Ecolhendo-se Estimação de parâmetros, Teclado e proporções, o seguinte formulário aparece. Esse formulário deve ser preenchido com os dados da pesquisa realizada, conforme a ilustração.



Número de sucessos (y)	1175
Tamanho da amostra (n)	2500
Nível de significância	0.05

Relatório      Calcular

Clicar em calcular e relatório, para visualização dos resultados. O Sisvar além de apresentar o intervalo de confiança exato apresenta diversas aproximações para o intervalo, devido a aspectos didáticos do programa. A melhor delas, mesmo para pequenos valores de  $n$  e  $p$  afastado de 0,5, é a de Pratt (1968). Os resultados do IC exato e dos aproximados estão apresentados a seguir. Com base no IC exato pode afirmar com 95% de confiança que Beltrano da Silva teria na eleição, se ela fosse realizada naquele dia da pesquisa, uma proporção de votos entre 45,03% e 48,98%.

Estimação de proporções	
-----	
Intervalo exato	
-----	
Tamanho da amostra:	2500
Número de sucessos(y):	1175
Estimativa por ponto:	0.47000000
Coefficiente de Confiança(%):	95.00
Intervalo de confiança: [LI; LS]	
[	0.450281; 0.489790]
-----	
Aproximação normal	
-----	
Intervalo de confiança: [LI; LS]	
[	0.450436; 0.489564]
-----	
Aproximação normal com convergência ArcSen	
-----	
Intervalo de confiança: [LI; LS]	
[	0.450464; 0.489582]
-----	
Aproximação de Poisson	
-----	
Intervalo de confiança: [LI; LS]	
[	0.443507; 0.497662]
-----	
Aproximação de Blyth (1986)	
-----	
Intervalo de confiança: [LI; LS]	
[	0.450297; 0.489796]
-----	
Aproximação Hall (1982)	
-----	
Intervalo de confiança: [LI; LS]	
[	0.450271; 0.489799]
-----	
Aproximação Molenaar (1973)	
-----	
Intervalo de confiança: [LI; LS]	
[	0.450281; 0.489790]
-----	
Aproximação Pratt (1968)	
-----	
Intervalo de confiança: [LI; LS]	
[	0.450280; 0.489790]
-----	

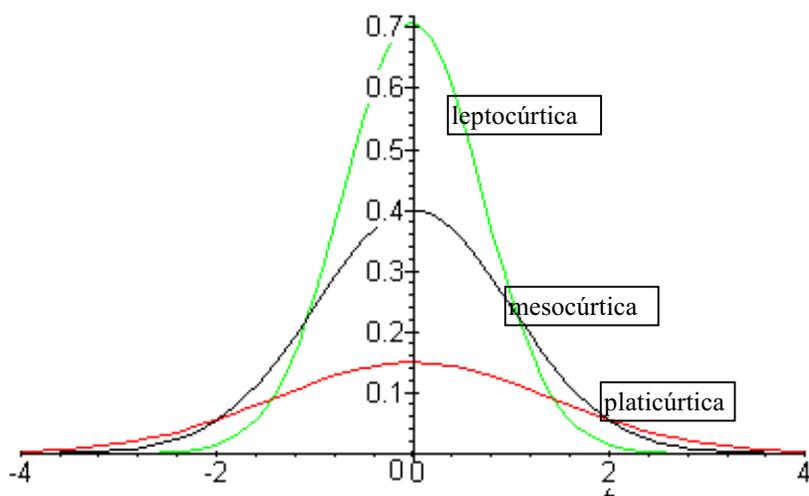
### 3.4. Estimação de coeficientes de assimetria e curtose

Seguindo os passos para o Sisvar, descritos na seção 3.1, dentre os intervalos obtidos são apresentados os dos coeficientes de assimetria e curtose. Esses intervalos são baseados em resultados assintóticos descritos em outros módulos desse curso. Os intervalos de confiança para esses coeficientes estão apresentados a seguir.

### Estimação dos coeficientes de assimetria e curtose pelo Sisvar

Coeficiente de assimetria - Normal=0:	0.501291
Tamanho da amostra:	20
Coeficiente de Confiança (%):	95.00
Intervalo de confiança: [LI; LS]	
[	-0.425510; 1.428091]
-----	
Coeficiente de curtose- normal=3:	1.715168
Tamanho da amostra:	20
Coeficiente de Confiança (%):	95.00
Intervalo de confiança: [LI; LS]	
[	0.509200; 3.492564]
-----	

Verifica-se que ambos os intervalos abrangem os valores paramétricos da distribuição normal. Baseados nesses resultados pode-se afirmar com 95% de confiança que não existem evidências suportadas pelos dados de incidência da doença no feijoeiro de que exista algum desvio de assimetria ou de curtose. Conclui-se portanto que os dados são provenientes de uma distribuição normal, sendo simétrica e mesocúrtica, conforme o gráfico assinalado a seguir:

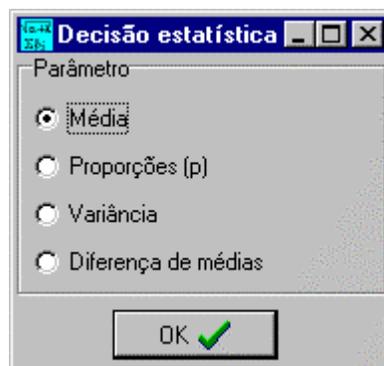


## 4. Teoria da decisão

Na inferência estatística uma das principais objetivos é o teste de hipóteses sobre os principais parâmetros populacionais. Baseando nos resultados amostrais, na teoria probabilística e nas distribuições amostrais dos estimadores é possível testar hipóteses sobre parâmetros de interesse. Os testes de hipóteses estão sujeitos a dois tipos de erro. O erro tipo I, que aquele em que se comete ao rejeitar uma hipótese de fato verdadeira, cuja probabilidade de cometê-lo é  $100\alpha\%$ , e o erro tipo II, que é aquele que se comete ao aceitar uma hipótese falsa como sendo verdadeira, cuja probabilidade é  $100\beta\%$ . Serão apresentados a seguir os testes de hipóteses sobre os principais parâmetros da pesquisa científica por meio dos mecanismos propiciados pelo Sisvar.

### 4.1. Testes sobre média

Para se testar uma hipótese de interesse sobre a média de uma população, será apresentado os procedimentos no Sisvar. Supondo que o pesquisador tenha interesse em testar uma hipótese formulada sobre a incidência da doença no feijoeiro. Supondo que a hipótese seja  $H_0: \mu=20\%$ , ou seja, que os danos causados pelo fungo são de 20%. Para se testar essa hipótese pelo Sisvar, basta escolher as opções do menu principal: Análises e Testes de hipóteses, selecionando o parâmetro a ser testado no formulário que se apresenta em seguida.



Defini-se os valores das estimativas amostrais nos campos apropriados e responde-se na caixa de edição o valor da hipótese  $H_0$  que se pretende testar (20%). Clicar no botão calcular e relatório para verificar a saída.

**Testes de hipóteses**

Média amostral: 2.451500

Tamanho da amostra: 20

Nível de significância: 0.05

Variância da amostra: 4.736319

Ho: Parâmetro igual a: 20

Relatório      Calcular

O resultado do teste é:

Hipóteses sobre médias	
Obs. Esse teste é bilateral! Para obter a significância do correspondente unilateral, divida o valor calculado do nível de significância por 2	
Tamanho da amostra:	20
Média da amostra:	2.451500
Variância da amostra:	4.736319
Teste da hipótese: Ho: m =	20.000000
t calculado: tc =	-36.060731
Probabilidade: P(t> tc =	0.000000
Rejeita-se Ho ao nível de 5.00% de probabilidade	

O Sisvar calcula o valor da estatística t de Student calculada (-36,06), a probabilidade exata da significância ( $P(t > |t_c|) = 0.0000$ ), e uma conclusão estatística para o resultado do teste considerando o valor da significância escolhido pelo usuário no formulário de opções. Nesse caso a

hipótese foi rejeitada a 5% de probabilidade e como a incidência amostral foi menor do que o valor hipotético, pode-se afirmar que a doença tem incidência menor do que 20% no feijoeiro.

## 4.2. Teste sobre diferença de médias

Quando se tem duas populações amostradas independentemente, ou dois tratamentos, o pesquisador deseja inferir qual delas é a melhor em termos médios. Para isso, é possível realizar o teste de hipótese de que as médias das duas populações são iguais. Esse teste é equivalente a testar a nulidade da diferença entre elas. No entanto, se as variâncias são heterogêneas, o teste de t é apenas aproximado. Para melhorar tal aproximação recomenda-se o procedimento de Satterthwaite ou de Cochran e Cox, que são equivalentes.

Para realizar o teste no Sisvar, escolhe-se as opções do menu principal: Análises, Testes de hipóteses e diferenças de médias no formulário que será apresentado. Em seguida basta preencher o campo do novo formulário que se apresenta. Para ilustrar o teste será apresentado o exemplo 2, com o interesse do pesquisador de verificar se os volumes médios das duas progênies são idênticos. Isso equivale ao teste da hipótese  $H_0: \mu_1 - \mu_2 = 0$ . Assim, o formulário ficará da seguinte forma:

**Testes de hipóteses**

Média da amostra 1: 128.944444

Tamanho da amostra 1: 18

Média da amostra 2: 156.222222

Tamanho da amostra 2: 18

Variância da amostra 1: 4035.702614

Variância da amostra 2: 3400.653595

Nível de significância: 0.05

H<sub>0</sub>: Diferença de médias igual a: 0

Relatório      Calcular

E parte dos resultados que deverão ser consultados em primeiro lugar estão apresentados a seguir. Esse resultado refere-se ao teste de hipótese de igualdade das variâncias das 2 populações. Verificou-se que elas são iguais como já discutido no capítulo 3.

Hipóteses sobre diferenças de médias			
-----			
Estatísticas amostrais			
-----			
Pop.	Média	Variância	Tamanho da amostra
-----			
1	128.944444	4035.702614	18
2	156.222222	3400.653595	18
-----			
Teste da hipótese de igualdade das variâncias populacionais (1 e 2)			
-----			
F <sub>c</sub> =	1.187	V <sub>1</sub> =	17
		V <sub>2</sub> =	17
P(F* >	1.187)=	0.3641	
Ao nível de 5% de significância as variâncias podem ser consideradas iguais			

O próximo passo é escolher dentre as duas opções de teste, aquela em que as variâncias são consideradas iguais ou homogêneas. Esses resultados estão apresentados a seguir.

Situação em que as variâncias são consideradas homogêneas	
Estimativa por ponto de $m_1 - m_2$ :	-27.27778
Variância combinada (pooled):	3718.178105
Graus de liberdade:	34
Nível de significância:	5.00000%
Teste da hipótese: $H_0: m_1 - m_2 =$	0.000000
t calculado: $t_c =$	-1.342039
Probabilidade: $P(t >  t_c ) =$	0.188477
Não existe evidências para rejeitar $H_0$ ao nível de 5.00% de probabilidade	
Situação em que as variâncias não são consideradas homogêneas	
Estimativa por ponto de $m_1 - m_2$ :	-27.27778
Graus de liberdade:	34
Nível de significância:	5.00000
Teste da hipótese: $H_0: m_1 - m_2 =$	0.000000
t calculado: $t_c =$	-1.342039
Probabilidade: $P(t >  t_c ) =$	0.188477
Não existe evidências para rejeitar $H_0$ ao nível de 5.00% de probabilidade	

A primeira situação, refere-se ao teste para o caso das variâncias iguais, sendo a opção adequada para a escolha do usuário. O valor de  $t_c = -1,34$  e a sua significância (0,1885) indicam que a hipótese de que as progênies tem a mesma performance não pode ser falseada com 95% de confiança.

É conveniente chamar a atenção para o teste da hipótese de igualdade das variâncias populacionais. Nesse teste a maior variância amostral é sempre colocada no numerador da estatística  $F'$ . Assim, esse teste é um teste bilateral, feito de forma unilateral. Esse fato faz com que a probabilidade calculada (valor da significância) deva ser multiplicada por 2. No Sisvar, é necessário que a verificação se dê com o valor  $\alpha/2$ , ou seja, 2,5%.

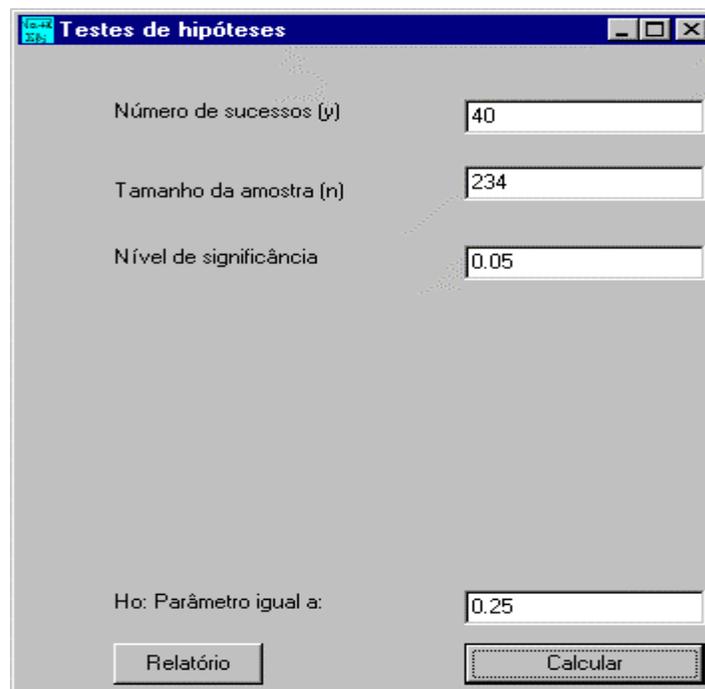
Uma outra situação da pesquisa que merece ser destacada é o caso das comparações pareadas, ou seja, aquela em que medidas são feitas antes e após a aplicação de um medicamento, ração, produto químico, etc. Assim, o pesquisador quer avaliar se existe efeito do tratamento aplicado. Nesse caso, basta obter as diferenças dos resultados após e antes do tratamento, e então aplicar um teste de hipótese sobre a média da diferença. Esse procedimento pode ser feito conforme descrito na seção 4.1.

### 4.3. Teste sobre proporções

O Sisvar apresenta somente com o uso da aproximação normal. Uma alternativa para o teste de hipóteses de interesse sobre a binomial poderia ser realizada pela equivalência entre a teoria da estimação (intervalos de confiança) e a teoria da decisão (teste de hipótese). Se os valores hipotéticos estiverem no intervalo de confiança a hipótese é aceita, caso contrário deve ser rejeitada. Com essa opção o IC exato pode ser usado, conforme descrição feita no capítulo 3.

O processo para usar o Sisvar no teste de hipótese, será descrito usando-se um exemplo. Um geneticista, ao estudar o controle genético de um caráter, levantou a hipótese dele ser controlado por um único gene, com dois alelos, com dominância total. O alelo recessivo em homozigose conferia a cor branca as flores da espécie que ele estudava. De acordo com sua hipótese e com as leis da genética, numa geração  $F_2$  seriam esperadas  $\frac{1}{4}$  de plantas com flores brancas se a mesma for verdadeira. Assim ele realizou um experimento e observou 234 plantas constatando que 40 delas tinham a cor branca. Existe evidências de que a hipótese formulada seja suportada pelos dados?

Para responder a essa pergunta, basta testar a hipótese de que a proporções de plantas com flores brancas seja de  $\frac{1}{4}$ , ou seja,  $H_0: P = 0,25$ . Assim, escolhe-se as opções do menu principal: Análises, Testes de hipóteses e proporções no formulário que será apresentado. Em seguida basta preencher o campo do novo formulário, como ilustrado a seguir.



Número de sucessos (y)	40
Tamanho da amostra (n)	234
Nível de significância	0.05
Ho: Parâmetro igual a:	0.25

Relatório      Calcular

Consultando os resultados obtidos tem-se:

Hipóteses sobre proporções	
Obs. Esse teste é bilateral! Para obter a significância do correspondente unilateral, Divida o valor calculado do nível de significância por 2	
Tamanho da amostra:	234
Obs. Hipóteses testadas via aproximação normal	
Proporção amostral:	0.170940
Número de sucessos:	40
Teste da hipótese: Ho: p =	0.250000
Z calculado. zc =	-3.212542
Probabilidade: P(z> zc )=	0.001316
Rejeita-se Ho ao nível de 5.00% de probabilidade	

O valor de  $Z_c = -3,21$  e de sua significância ( $0P<,0013$ ) indicam que a hipótese deve ser rejeitada. Portanto, o geneticista estava errado em supor que apenas um gene controla a cor de flor dessa espécie.

#### 4.4. Teste sobre variâncias

Para se testar hipóteses sobre variâncias é necessário supor que a população amostrada possui distribuição normal, conforme já visto em outros módulos desse curso. Assim, para ilustrar o teste será apresentado o seguinte exemplo: suponha que um melhorista só atuará em uma população se ela tiver variância igual a  $30\text{kg}^2$ . Realizando uma amostra de tamanho 50 dessa mesma população ele estimou uma variância igual a  $27\text{kg}^2$ . Qual deverá ser a decisão do melhorista? Para responder a essa questão é necessário que se teste a hipótese  $H_0: \sigma^2 = 30$ . Para isso, usando o Sisvar, basta escolher no menu principal as opções: Análises, Testes de hipóteses e marcar variâncias no formulário que será apresentado em seguida. Preenche-se o formulário com as informações amostrais e hipotéticas, conforme a figura apresentada a seguir.

De acordo com essas informações, após ser clicado o botão calcular e relatório, o Sisvar apresenta os seguintes resultados.

Hipóteses sobre variâncias	
Tamanho da amostra:	50
Variância amostral:	27.000000
Teste da hipótese: Ho: V =	30.000000
Contra Ha: V diferente de	30.000000
Qui-quadrado calculado: X <sup>2</sup> <sub>c</sub> =	44.100000
Limite da região inferior de rejeição:	31.554917
Limite da região superior de rejeição:	70.222413
Não existe evidências para rejeitar Ho ao nível de 5.00% de probabilidade	

Portanto, como a hipótese não foi rejeitada, pois o valor de qui-quadrado calculado (44,1) pertence a região de aceitação de  $H_0$  ([31,6, 70,2]), o melhorista deve atuar nessa população.

## 5. Modelos de regressão linear dentro da análise de variância

Uma importante ferramenta na pesquisa é formulação de modelos empíricos que relacionam duas ou mais variáveis. Esses modelos tem seus coeficientes desconhecidos pelo pesquisador e precisam ser calculados para se estabelecer as relações pretendidas. Usando-se a linguagem estatística, pode-se dizer que os coeficientes desse modelo são os parâmetros e o seus cálculos são os estimadores para obter as estimativas, que devem ser baseadas em resultados amostrais ou experimentais.

Para exemplificar a utilização do procedimento de regressão, um exemplo com uma população de *Araucaria angustifolia* com 10 anos de idade será utilizado. Deseja-se explicar o desenvolvimento de volume/acre (Y) em função da área basal  $X_1$ , e altura em pés  $X_2$ .

Y	$X_1$	$X_2$
65	41	35
78	71	53
82	90	64
86	80	59
87	93	66
90	90	64
93	87	62
96	95	67
104	100	70
113	101	71

O procedimento regressão foi idealizado para implementar regressões lineares múltiplas de acordo com o modelo:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon_i$$

o qual relaciona o comportamento da variável dependente Y com uma função linear de um grupo de variáveis independentes  $X_1, X_2, \dots, X_m$ . Os  $\beta$ 's são parâmetros estimados por mínimos quadrados através das equações normais (EN):

$$X'X\hat{\beta} = X'Y$$

Uma solução para este sistema é:

$$\hat{\beta}=(X'X)^{-1}X'Y$$

A notação de redução (R) é usada para diferenciar os dois tipos de somas de quadrados fornecidas pelo procedimento REG.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i \quad (1)$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_i \quad (2)$$

então,

$$R(\beta_3/\beta_0, \beta_1, \beta_2) = \text{SQmodelo}(2) - \text{SQmodelo}(1)$$

A soma de quadrado de um dado modelo é obtida por:

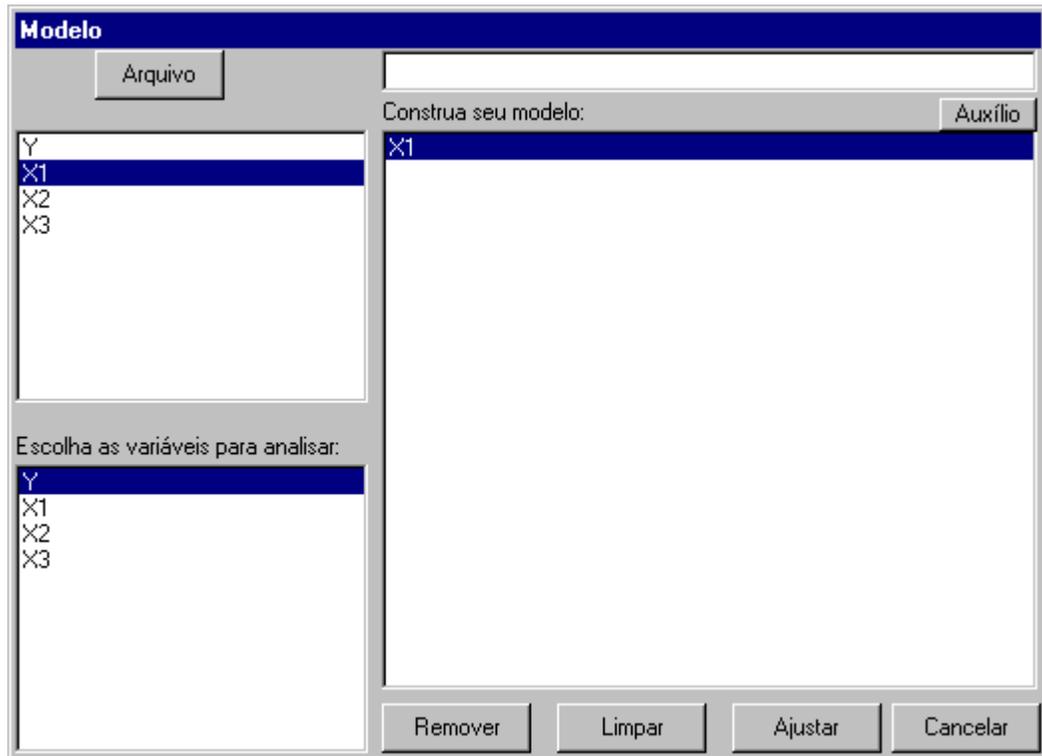
$$\text{SQmodelo} = \hat{\beta}' X' Y$$

Conhecendo-se o funcionamento da notação R, pode-se agora apresentar os dois tipos de somas de quadrados do procedimento Regressão. As somas de quadrados tipo I (seqüencial) refere-se a reduções consecutivas do modelo e a tipo II refere-se a reduções de um parâmetro por vez a partir do modelo completo, conforme ilustra a Tabela 2, para o exemplo em questão.

**Tabela 2.** Somas de quadrados tipo I e Tipo II do procedimento regressão do programa Sisvar.

FV	tipo I (seqüencial)	tipo II (parcial)
X <sub>1</sub>	R(β <sub>1</sub> /β <sub>0</sub> )	R(β <sub>1</sub> /β <sub>0</sub> , β <sub>2</sub> , β <sub>3</sub> )
X <sub>2</sub>	R(β <sub>2</sub> /β <sub>0</sub> , β <sub>1</sub> )	R(β <sub>2</sub> /β <sub>0</sub> , β <sub>1</sub> , β <sub>3</sub> )
X <sub>3</sub>	R(β <sub>3</sub> /β <sub>0</sub> , β <sub>1</sub> , β <sub>2</sub> )	R(β <sub>3</sub> /β <sub>0</sub> , β <sub>1</sub> , β <sub>2</sub> )

Para ilustrar a utilização do procedimento regressão linear realizou-se a análise de regressão de uma variável dependente (Y) com a variável independente X<sub>1</sub>, usando os seguintes escolhas no formulário de modelagem do Sisvar 4.2.



Esse programa está ajustando um modelo de  $y=f(x_1)$  dado por:

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i \quad (1)$$

O modelo ajustado bem como a análise de variância estão apresentados na saída Sisvar de regressão:

### Saída Sisvar de regressão:

Variável analisada: Y

Variáveis do modelo e codificação usada

b( 1): X1

Análise de variância

FV	GL	SQ	QM	Fc	Pr>Fc
Modelo	1	1220.388793949	1220.38879395	24.1654	0.0012
Erro	8	404.011206051	50.50140076		
Total corrigido	9	1624.400000000			
Total não corrigido	81548				
Média	89.40000000	Raiz do QME	7.10643376		
R^2	0.75128589	R^2 ajustado	0.72019662		
C.V. (%)	7.94903105				

Estimativas dos parâmetros

Variável	GL	Estimativa dos parâmetros	EP	t para H0: parâmetro = 0	Pr> t
b ( 0)	1	33.96344025774	11.498868269	2.953633302	0.0183
b ( 1)	1	0.65373301583	0.132985113	4.915836087	0.0012

Concluí-se uma alta significância do modelo de regressão ( $P < 0,0012$ ),  $R^2$  relativamente alto (0,75) e modelo de regressão estimado dado por:

$$\hat{Y} = 33,9634 + 0,6537 X_1$$

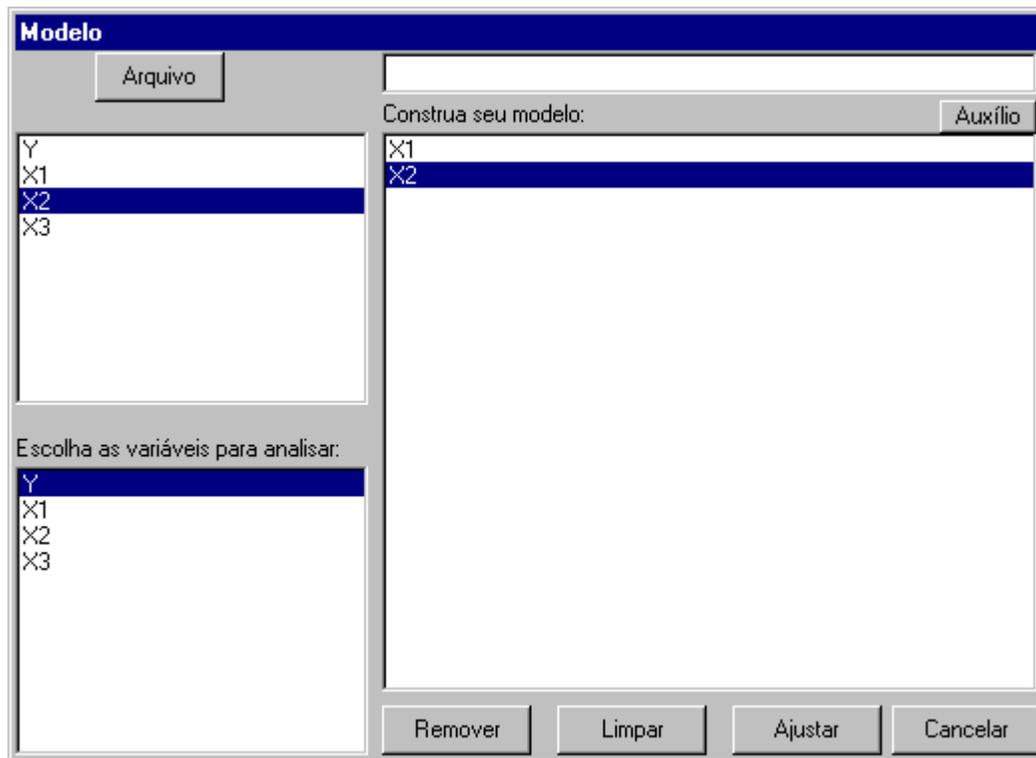
Muitas opções são possíveis no procedimento Regressão e serão introduzidas nos momentos convenientes.

Bons modelos possuem SQ de regressão representando grandes frações da SQTotal (Soma de quadrados total).

Considere agora o modelo:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$$

Use os seguintes especificações no Sisvar para obter esse modelo ajustado.



A saída (output): 1.  $R^2$  ajustado: refere-se ao  $R^2$  ajustado para o número de parâmetros do modelo de acordo com a seguinte fórmula:

$$R_{Ajustado}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - m - 1}$$

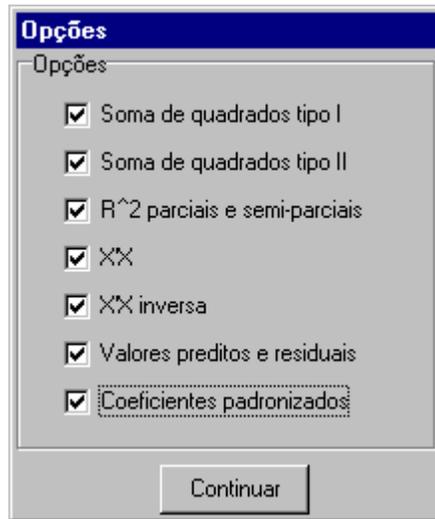
Em que,  $n$  é o número de observações;  $m$  é número de parâmetros, com exceção do intercepto.

Este valor deve ser preferido, pelo fato do  $R^2$  crescer com o simples incremento de variáveis no modelo, mesmo que o benefício provocado por elas seja pequeno.

Um objetivo da regressão é calcular valores preditos:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_m X_m$$

No procedimento de regressão a marca na opção valores preditos possibilita tais cálculos. Usando o exemplo anterior e escolhendo a opção mencionada como mostrado a seguir, é possível obter tais valores.



O erro padrão do valor predito é calculado por:

$$S(\hat{Y}) = \sqrt{z' (X' X)^{-1} z QME}$$

onde,  $z$  é o vetor de coeficientes.

A opção cria o intervalo de confiança para média da sub-população específica, através de:

$$\hat{Y} \pm t_{\alpha/2} S(\hat{Y})$$

Para se obter as somas de quadrados seqüencial (tipo I) e parcial (tipo II), apresentadas na Tabela 2, basta selecionar tais a opções, como ilustrado anteriormente.

Muitas vezes as somas de quadrados do modelo de regressão apresentam F altamente significativo. No entanto, algumas, ou até mesmo todas as variáveis, não apresentam F parcial significativo. A primeira vista, isto parece uma incoerência ou até mesmo uma inconsistência do método. Mas ao se analisar com maior profundidade, percebe-se que não se trata de nenhum paradigma. Este fato se simplesmente porque algumas variáveis possuem informações redundantes da variação que ocorre na variável dependente  $Y$ . Dessa forma a variação adicional explicada por essa variável, na presença de outra ou outras, não é suficientemente importante para ser detectada naquele

nível de precisão, devendo ser descartada do modelo. Para esta finalidade alguns métodos existem. Os mais comuns são: StepWise, BackWard e ForWard, os quais serão apresentados a seguir.

**Nível de significância de entrada:** Nível de significância para entrada de variáveis no modelo. Deve ser usada no STEPWISE e no FORWARD.

**Nível de significância de permanência:** Nível de significância para saída ou permanência de variáveis no modelo. Deve ser usada no STEPWISE e no BACKWARD.

## EXERCÍCIOS PROPOSTOS

- 1) Use os dados apresentados nesse capítulo e faça a seguinte regressão usando os processos de seleção:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_1 * X_2 + \beta_5 1/X_2 + \varepsilon_i$$

Tire todas as conclusões de interesse e interprete os resultados.

- 2) Teste todos os programas apresentados no capítulo e interprete os resultados, usando o conjunto de dados apresentados no capítulo.

## 6. Análise de variância

A análise de variância é a mais utilizada entre as técnicas estatísticas. O pesquisador ao estudar um problema se vê elaborando um projeto que envolve um modelo linear, cujos efeitos sobre o objeto de estudo ele deseja medir e comparar. Diversos são os delineamentos estatísticos e esquemas experimentais apropriados para isso. Felizmente, os recursos computacionais são abundantes e eficientes, uma vez que é grande quantidade de cálculos envolvido, bem como, é necessário elevada precisão e segurança nos cálculos e testes efetuados. Será descrito a seguir alguns delineamentos fundamentais e alguns dos recursos do Sisvar para analisá-los.

### 6.1. Delineamento inteiramente casualizado

O delineamento inteiramente casualizado (DIC) é aquele em que os tratamentos são aleatorizados pela área ou pelo material experimental. A área ou material experimental devem ter a característica de serem o mais homogêneos possíveis. Caso essa condição não ocorra, esse material experimental deve ser quebrado em frações mais uniformes, que são chamadas de blocos.

Para ilustrar a análise de variância de um delineamento, será apresentado um exemplo de um experimento com suínos em que cada animal constituiu-se uma parcela, e sobre os quais aplicou-se cinco rações com fontes de proteínas diferentes. Usou-se quatro animais para cada ração (4 repetições) casualizados de um grupo de vinte suínos bastante uniformes quanto ao peso e características genéticas. Os tratamentos foram:

- A: Fonte de proteína vegetal da firma X;
- B: Fonte de proteína vegetal da firma Y;
- C: Fonte de proteína animal da firma X;
- D: Fonte de proteína animal da firma Y;
- E: Fonte de proteína animal composta de  $\frac{1}{2}$  da firma X +  $\frac{1}{2}$  da firma Y;

Os dados de ganho de peso desse experimento estão apresentados na Tabela 4.

**Tabela 4. Ganho de peso dos animais em kg (valores do período experimental) para um delineamento inteiramente casualizado com 4 repetições e 5 tratamentos (fontes protéicas)**

Progênie	REPETIÇÕES			
	I	II	III	IV
A	12	9	7	10
B	7	4	5	4
C	7	9	8	7
D	7	5	6	4
E	11	10	9	10

## Sisvar

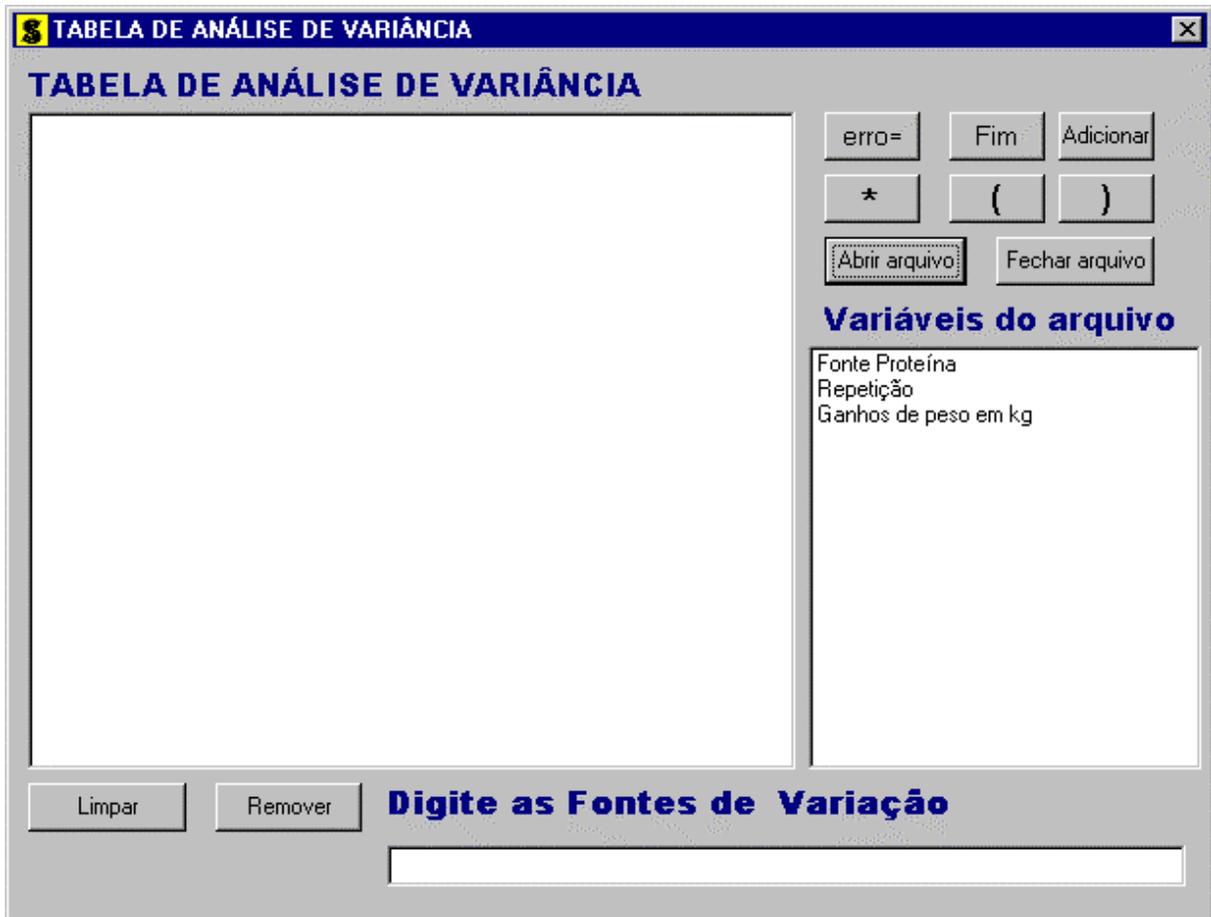
Será descrito os principais passos para a análise de variância usando o Sisvar. Será aproveitada a oportunidade para realizar análise de variância e apresentar as formas de comparações múltiplas entre tratamentos. Nesse caso, a estrutura dos tratamentos sugere que seja aplicado contrastes entre os efeitos dos mesmos. O primeiro passo é criar o arquivo, conforme os passos demonstrados no capítulo 1. Uma parte desse arquivo está apresentado a seguir:

Fonte Proteína	Repetição	Ganhos de peso em kg
C	4	7
D	1	7
D	2	5
D	3	6
D	4	4
E	1	11
E	2	10
E	3	9
E	4	10

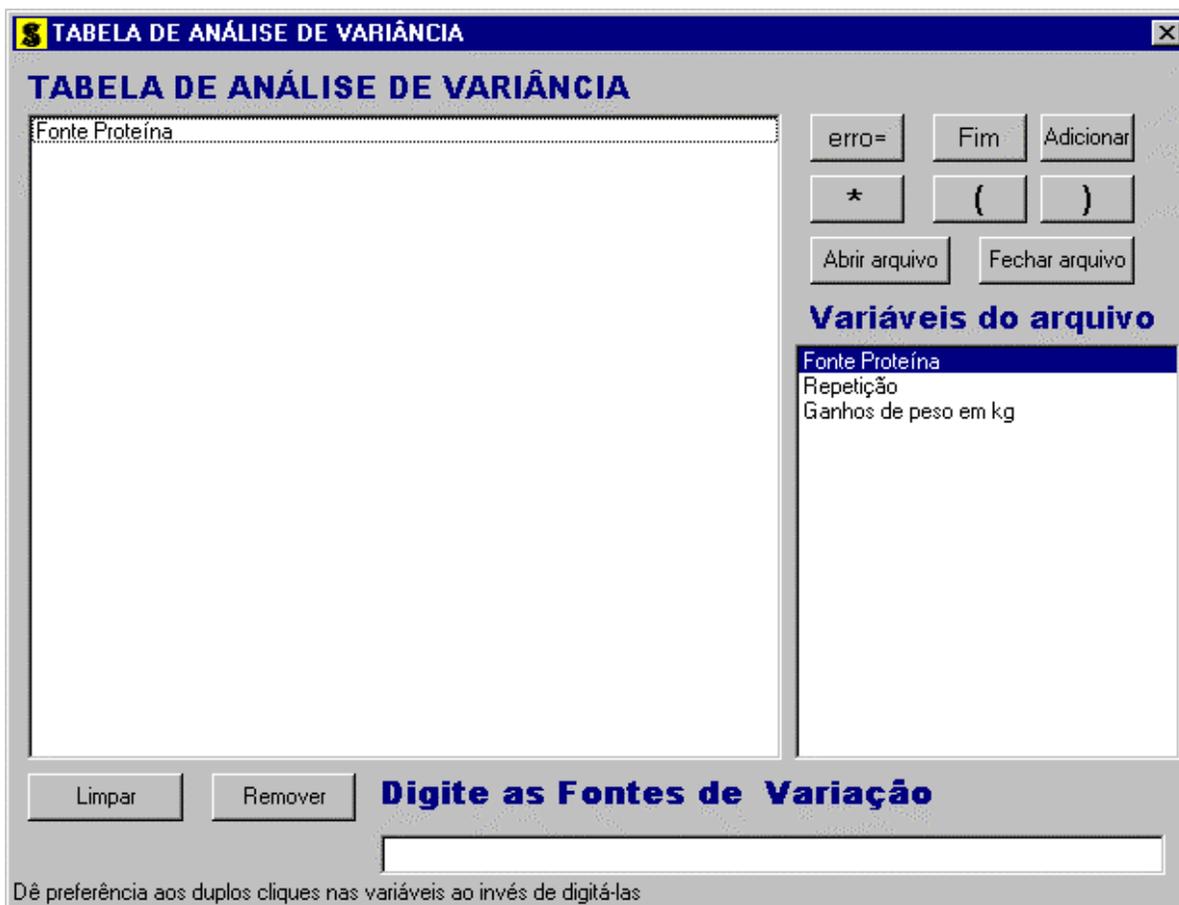
Em seguida abandona-se o formulário de manipulação de arquivos e escolhe-se no menu principal do Sisvar “Análise de variância” conforme formulário apresentado a seguir:



Em seguida ao usuário será apresentado um formulário em que o quadro de análise de variância deverá ser construído. No canto superior direito do formulário é apresentado o quadro de análise de variância. No canto inferior esquerdo o quadro onde serão listadas as variáveis do arquivo a ser aberto. Em baixo uma caixa de edição para se montar fontes de variação da análise de variância envolvendo interações e efeitos aninhados (ou hierarquizados). O primeiro passo é procurar o arquivo desejado e realizar a análise de variância. Esse arquivo foi denominado de exemplo3.db. Assim, ao abrir o arquivo usando-se o botão “abrir arquivo”, a listagem de suas variáveis aparece no canto inferior esquerdo do formulário. A tabela de análise de variância desse modelo irá conter as fontes de variação fonte protéica (tratamentos) e o erro experimental. O Sisvar exige que todas as fontes de variação, com exceção do **último** erro experimental, sejam acrescentadas a essa tabela. Nesse exemplo a única fonte de variação diferente do erro experimental é a relativa aos tratamentos. Existem no formulário, além desses objetos descritos, os botões de asteriscos e de parênteses, para se criarem fontes de variações com interações e efeitos hierárquicos. Existem o botão de limpar o quadro de análise de variância e o de remover uma fonte de variação mal especificada. Além desses, existe ainda, um botão para entrada de erros intermediários em modelos mais complexos, tais como parcelas subdivida, parcela sub-subdividida, parcelas subdividida no tempo ou no espaço, esquemas de faixas, entre outros. O formulário com as variáveis do arquivo aberto, está apresentado a seguir.

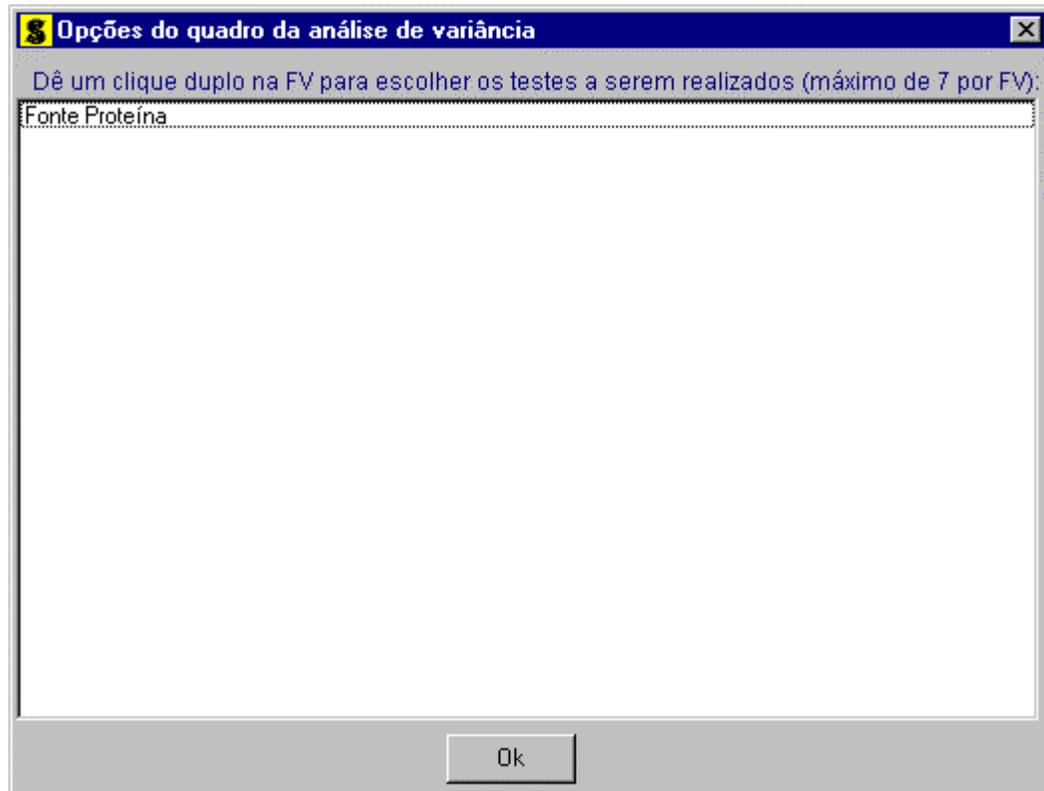


O próximo passo é clicar nas fontes de variação que deverão compor o quadro de análise de variância, montá-las na janela de edição das mesmas (na parte de baixo do formulário) e adicioná-las ao quadro de análise de variância por meio do botão adicionar. Nesse caso o usuário, deverá clicar na variável Fonte de proteína, ela deverá ser enviada para a caixa de edição, e em seguida clicar no botão adicionar. O próximo passo é clicar no botão “Fim”, uma vez que o erro (resíduo) da análise de variância é obtido por diferença pelo Sisvar. O formulário com a análise de variância construída dessa forma está apresentado a seguir.

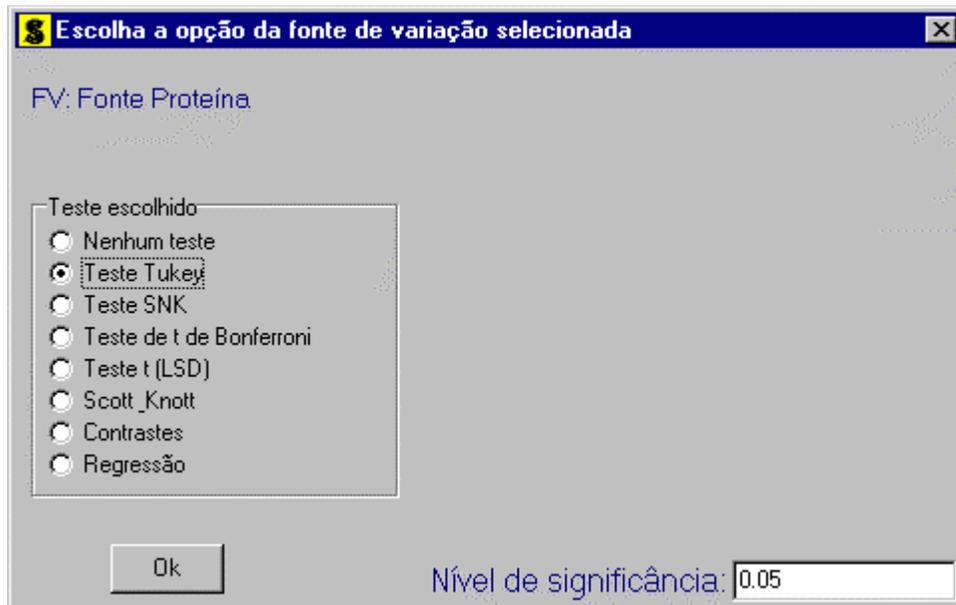


Pressionando-se em fim aparecerá um novo formulário para opções de testes de comparações múltiplas entre os tratamentos. Nessa fase de plano da análise de variância, o usuário escolhe todos os testes e comparações que foram planejadas em seu projeto de pesquisa. Se a análise de variância não acusar diferenças entre os tratamentos, os testes escolhidos devem ser eliminados do relatório de saída. Para escolher as opções de teste basta que o usuário dê um clique na fonte de variação de interesse (nesse caso só tem uma). Pode-se fazer até sete testes para cada fonte de variação. Nesse exemplo, será escolhido três testes: Tukey, Scott e Knott e contrastes de interesse. No caso dos contrastes, toda vez que os tratamentos sugerem isso, eles devem ser os testes preferidos. A finalidade de apresentar os testes de Tukey e de Scott e Knott é para demonstrar como proceder e como interpretar determinadas características dos mesmos. Nesse caso, os contrastes de interesse do pesquisador eram: i) comparar as fontes protéicas vegetal e animal,  $Y_1=3A+3B-2C-2D-2E$ ; ii) comparar as duas firmas quanto as fontes de proteína vegetal,  $Y_2=A-B$ ; iii) comparar a fonte proteína animal combinada contra as não combinadas,  $Y_3=C+D-2E$ ; iv) comparar as fontes de proteína

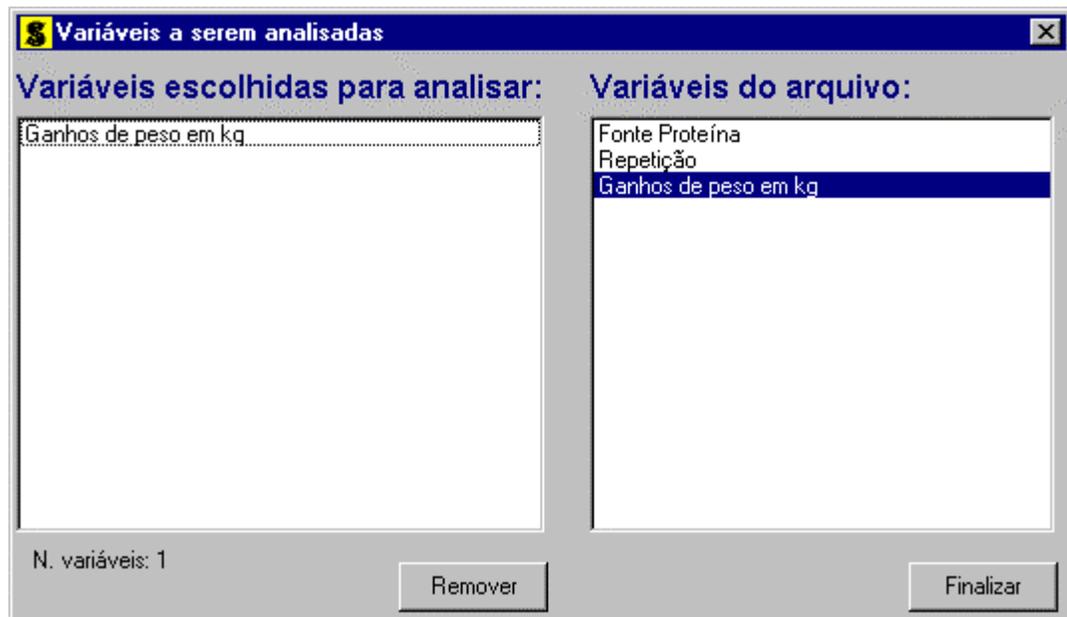
animal,  $Y_4=C-D$ . Observe que os coeficientes dos contrastes sempre somam zero, a soma dos produtos dos coeficientes correspondentes de todos os pares de contrastes somam zero (contrastos ortogonais) e que 4 contrastes foram realizados, correspondendo ao número de graus de liberdade dos tratamentos.



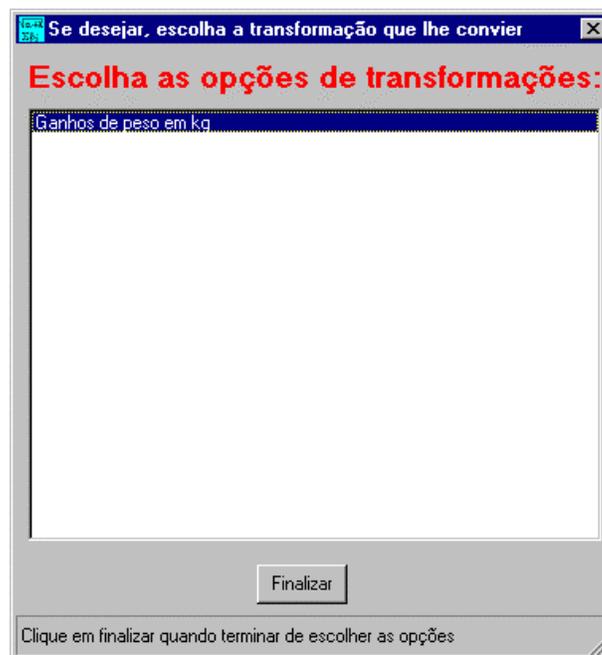
Com o mouse seleciona-se a fonte de variação de interesse para se escolher os testes de desejados. Faz-se isso três vezes repetidamente. Os formulário de opções dos testes é apresentado a seguir. As opções disponíveis são: teste de Tukey, SNK, LSD, LSD protegido por Bonferroni, Scott e Knott, contrastes e regressão. Regressão deve ser escolhida para tratamento cujos níveis são quantitativos. No exemplo em questão, foi escolhida a opção do teste de Tukey. O processo foi repetido escolhendo-se o teste de Scott e Knott e contrastes. Fixou-se o valor da significância em 5%, que é o default do Sisvar.



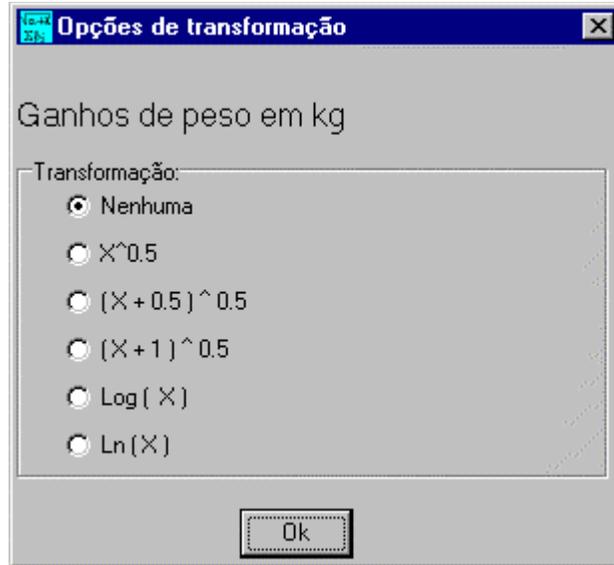
O próximo passo é selecionar as variáveis que serão analisadas. O Sisvar não permite que sejam analisadas variáveis que foram fontes de variação do modelo de análise de variância. Também, não se tem interesse e muito menos sentido em analisar repetição. Assim, no formulário que se apresenta o usuário tem o quadro da sua direita que possui as variáveis do arquivo, e a sua esquerda o quadro que recebe as variáveis selecionadas para ser analisada. É possível remover uma variável que por engano foi selecionada pressionando-se na mesma e em seguida no botão remover. Para selecionar uma variável basta clicar na mesma que ela será imediatamente colocada no quadro a sua esquerda. No exemplo, apenas a variável relativa aos ganhos de peso será analisada, sendo a única selecionada. É possível realizar algumas transformações simples no Sisvar. Ao clicar em Finalizar, aparece a listagem das variáveis que foram selecionadas para análise (variáveis dependentes), se o usuário desejar alguma transformação basta clicar sobre a variável e escolher a transformação marcando nos botões de opções adequados.



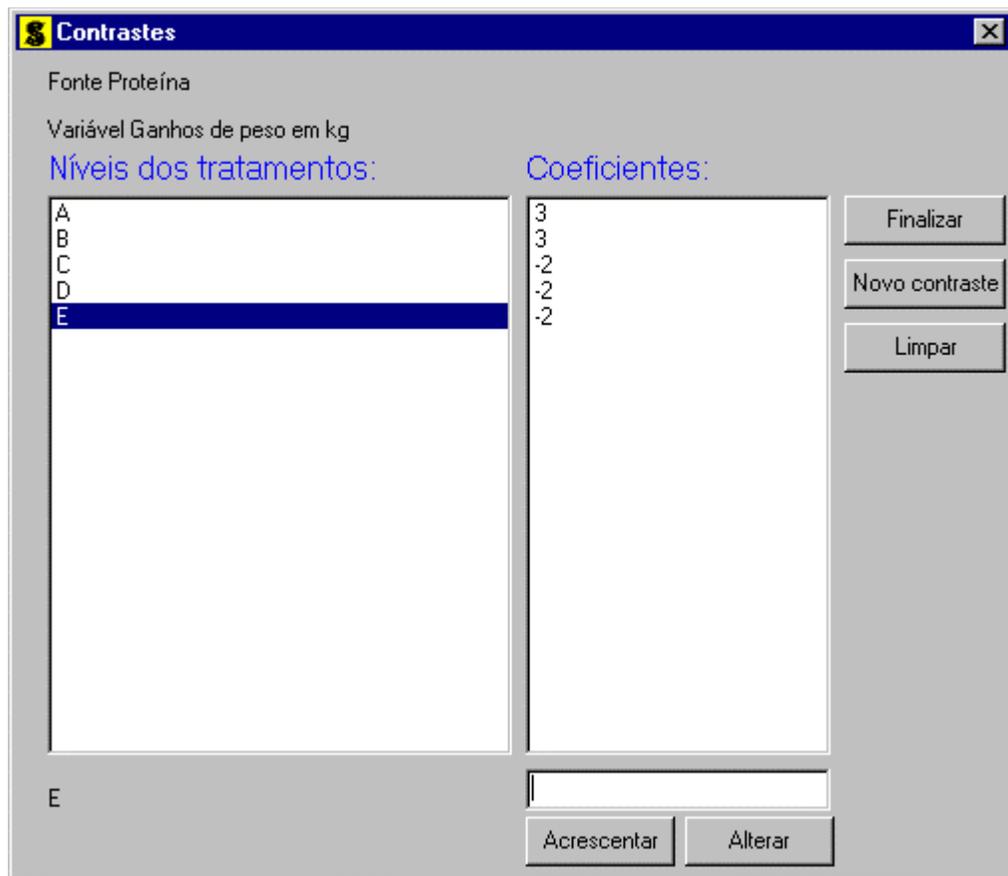
Após ter finalizado o seguinte formulário aparece para o usuário escolher alguma opção de transformação. Nesse caso, nenhuma transformação será escolhida.



Caso tenha-se optado por alguma transformação, que não é o caso, o formulário de opções aparece para o usuário. Até 5 opções podem ser escolhidas pelo mesmo.



O próximo passo é a realização da análise pelo Sisvar. Caso haja opções de contrastes selecionadas, então o Sisvar apresentará um formulário com para cada contraste que deve ser formulado. O usuário deverá digitar os coeficientes dos contrastes para cada nível dos tratamentos de acordo com seu planejamento feito anteriormente. Para o primeiro contraste formulado,  $Y_1=3A+3B-2C-2D-2E$ , o formulário com os coeficientes do mesmo está apresentado a seguir. É importante observar que os níveis dos tratamentos estão em uma janela ordenada, e que os coeficientes que estão sendo digitados estão em outra. Após cada contraste ser digitado o botão “novo contraste” deve ser pressionado até que o último seja digitado. Então, pressiona-se o botão “Finalizar”. Convém chamar a atenção de que para contrastes envolvendo os primeiros tratamentos apenas, não é necessário completar com zeros os níveis dos tratamentos não contemplados. Já para contrastes envolvendo os últimos tratamentos, os níveis daqueles que os precedem devem ser completados com zeros. Ao terminar o Sisvar apresentará os relatórios de análise de acordo com o plano de análise realizado.



Serão apresentados os resultados em etapas para maior clareza, embora o Sisvar o libere em um único formulário de resultados, sendo possível editorá-los, imprimi-los ou salvá-los. A análise de variância está apresentada a seguir.

```
Arquivo analisado:
D:\daniel\Tutoria em matemática e estatística\Exemplos\exemplo3.DB

Variável analisada: Ganhos de peso em kg
Opção de transformação: Variável sem transformação ( Y )
```

TABELA DE ANÁLISE DE VARIÂNCIA					
FV	GL	SQ	QM	Fc	Pr>Fc
Fonte Proteína	4	82.200000	20.550000	10.722	0.0003
erro	15	28.750000	1.916667		
Total corrigido	19	110.950000	5.839474		
CV (%) =	18.34				
Média geral:	7.5500000	Número de observações:	20		

Como já é do conhecimento dos leitores, o teste F é apropriado para a averiguação da hipótese de igualdade de médias entre tratamentos. Como o valor de F (10,722) foi significativo ( $P < 0,003$ ) indica que pelo menos uma das fontes protéicas difere das demais. Outras informações, como a média geral, coeficiente de variação, entre outras estatísticas, estão disponíveis na saída do Programa. O próximo passo são as saídas do teste de Tukey e Scott e Knott.

Teste Tukey para a FV Fonte Proteína			Teste Scott-Knott (1974) para Fonte Proteína		
DMS: 3.02384372256167 NMS: 0.05			NMS: 0.05		
Média harmonica do número de repetições (r): 4 Erro padrão: 0.692218655243173			Média harmonica do número de repetições (r): 4 Erro padrão: 0.692218655243173		
Tratamentos	Médias	Resultados do teste	Tratamentos	Médias	Resultados do teste
B	5.000000	a1	B	5.000000	a1
D	5.500000	a1	D	5.500000	a1
C	7.750000	a1 a2	C	7.750000	a2
A	9.500000	a2	A	9.500000	a2
E	10.000000	a2	E	10.000000	a2

Pode-se observar que os testes diferem nos seus resultados. O ponto chave dessas diferenças é que o teste de Tukey, como a maioria dos testes de comparações múltiplas, possuem o que é denominado de ambigüidades de resultados. Para mostrar isso, será tomado um exemplo: os tratamentos B e C são estatisticamente considerados iguais pelo teste de Tukey por possuírem a mesma letra (a1). Os tratamentos C e A também o são por terem a mesma letra (a2). Porém, os tratamentos B e A diferem entre si. Matematicamente é impossível explicar tal resultado. Os estatísticos convivem com esse fato, mas possuem muitas dificuldades ao ministrar um aula, ou ao tentar explicar aos pesquisadores resultados desses testes. Já o teste de Scott e Knott é um teste isento de ambigüidades. Seus resultados são mais claros e objetivos. Assim, o teste em questão afirma que os tratamentos B e D formam um grupo de menor ganho de peso e o os tratamentos C, A e E são estatisticamente iguais e superiores aos demais (B e D).

A melhor abordagem para esse exemplo, como já foi comentado, é por meio da realização dos contrastes de interesse. Os resultados do Sisvar para os quatro contrastes são apresentados a seguir. Primeiro o Sisvar apresenta o contraste formulado, e em seguida diversos testes para o mesmo.

Nível dessa Fonte de Variação		Coefficientes
A		3.0000
B		3.0000
C		-2.0000
D		-2.0000
E		-2.0000

Obs. Valores dos coeficientes positivos foram divididos por 6 e os negativos por 6

Estimativa	:	-0.50000000
DMS Scheffé	:	2.20915995
NMS:	:	0.05
Variância	:	0.39930556
Erro padrão	:	0.63190629
t para H0: Y = 0	:	-0.791
Pr> t	:	0.441
F para H0: Y = 0	:	0.626
Pr>F	:	0.441
Pr exata Scheffé	:	0.957

Nível dessa Fonte de Variação		Coefficientes
A		1.0000
B		-1.0000

Obs. Valores dos coeficientes positivos foram divididos por 1 e os negativos por 1

Estimativa	:	4.50000000
DMS Scheffé	:	3.42241589
NMS:	:	0.05
Variância	:	0.95833333
Erro padrão	:	0.97894501
t para H0: Y = 0	:	4.597
Pr> t	:	0.000
F para H0: Y = 0	:	21.130
Pr>F	:	0.000
Pr exata Scheffé	:	0.007

Concluí-se pelo primeiro contraste que as fontes de proteínas não diferem entre si quanto ao seus efeitos no ganho de peso dos animais. O Sisvar apresenta três opções para o teste da significância do contraste: o teste t, de seu equivalente F e o teste Scheffé. O teste Scheffé nesse caso apresenta uma novidade, que se trata da significância exata. Também apresenta o critério da DMS, para o valor nominal de  $\alpha$  escolhido. Se a estimativa em módulo supera a DMS diz-se que o contraste é significativo, caso contrário os dois grupos contrastantes não diferem entre si. Esse critério é o mais rigoroso deles, e pode apresentar resultados não coincidentes com os demais, como por exemplo aconteceu no último contraste. O segundo contraste foi altamente significativo, indicando que o

tratamento A supera em média o B em 4,5 kg/animal (estimativa do contraste). Além dessas estatísticas, o Sisvar apresenta a estimativa da variância e do erro padrão do contraste. Os dois últimos contrastes apresentados a seguir também foram significativos ( $P < 0,05$ ) pelo teste t ou F, indicando que combinar as fontes de proteína animal é melhor do que fornecê-las isoladamente e que a fonte de proteína animal C é melhor em média do que a D.

Nível dessa Fonte de Variação	Coefficientes
A	0.0000
B	0.0000
C	1.0000
D	1.0000
E	-2.0000

Obs. Valores dos coeficientes positivos foram divididos por 2 e os negativos por 2

Estimativa	:	-3.37500000
DMS Scheffé	:	2.96389910
NMS:	:	0.05
Variância	:	0.71875000
Erro padrão	:	0.84779125
t para H0: Y = 0	:	-3.981
Pr> t	:	0.001
F para H0: Y = 0	:	15.848
Pr>F	:	0.001
Pr exata Scheffé	:	0.022

Nível dessa Fonte de Variação	Coefficientes
A	0.0000
B	0.0000
C	1.0000
D	-1.0000

Obs. Valores dos coeficientes positivos foram divididos por 1 e os negativos por 1

Estimativa	:	2.25000000
DMS Scheffé	:	3.42241589
NMS:	:	0.05
Variância	:	0.95833333
Erro padrão	:	0.97894501
t para H0: Y = 0	:	2.298
Pr> t	:	0.036
F para H0: Y = 0	:	5.283
Pr>F	:	0.036
Pr exata Scheffé	:	0.307

No Sisvar o valor da estimativa do contraste A e B vs C, D e E é -0,5 com erro padrão de 0,63, o que corresponde a um t de -0,79. Observe que uma correção é feita no Sisvar para dar um valor real para a estimativa do efeito do contraste, ou seja, comparando a média dos dois grupos de tratamentos contrastados, fornecendo um significado mais prático ao pesquisador.

### 6.1.1. Regressão na análise de variância

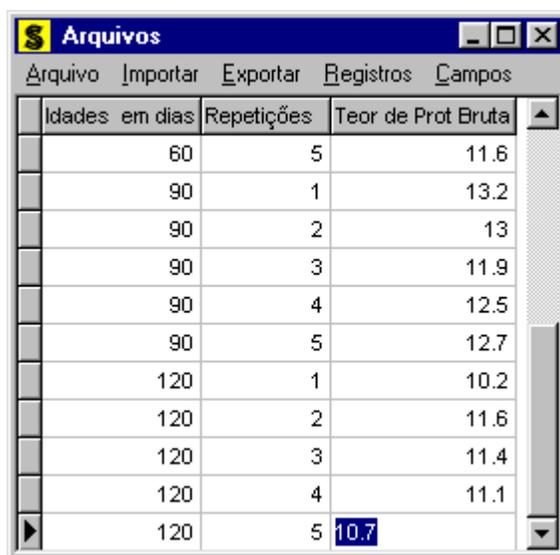
Muitas das pesquisas lidam com dosagens de produtos químicos, adubos, irrigação entre outros tipos de tratamentos. Esse tipo de estrutura dos tratamentos, em que seus níveis são quantitativos, requer dos pesquisador uma abordagem diferenciada dos testes de comparações múltiplas. Os tratamentos são modelados como variáveis explicativas do fenômeno (variável) sob estudo. Essa modelagem é feita pelo estabelecimentos de modelos de regressão entre a variável resposta e os níveis dos tratamentos. Como os tratamentos são controlados, repetidos em um experimento, com ou sem controle local, é possível explorar os modelos de regressão de uma forma bastante eficiente. São considerados para isso testes para o desvio de regressão, para a nulidade dos parâmetros (coeficientes) do modelo de regressão,  $R^2$ , valores estimados e preditos. É possível usar o procedimento de regressão e obter todos os testes comentados anteriormente. Para isso, porém, é preciso usar de artifícios e cálculos extras usando calculadoras ou planilhas. No caso, de modelos mais complexos não contemplados por esse módulo do Sisvar, o uso do procedimento de regressão linear pode ser imprescindível.

Para ilustrar o uso do Sisvar para efetuar análise de regressão será apresentado um modelo cujos tratamentos foram ensaiados em um delineamento inteiramente casualizado com cinco repetições. Esses tratamentos referem-se a idades de corte em dias (30, 60, 90 e 120) de uma gramínea forrageira. Avaliou-se o teor de proteína bruta em % de matéria seca após cada corte. Os resultados experimentais estão apresentados na Tabela 5.

**Tabela 5. Proteína bruta de uma determinada gramínea forrageira em porcentagem para um delineamento inteiramente casualizado com 5 repetições e 4 tratamentos (idade de corte)**

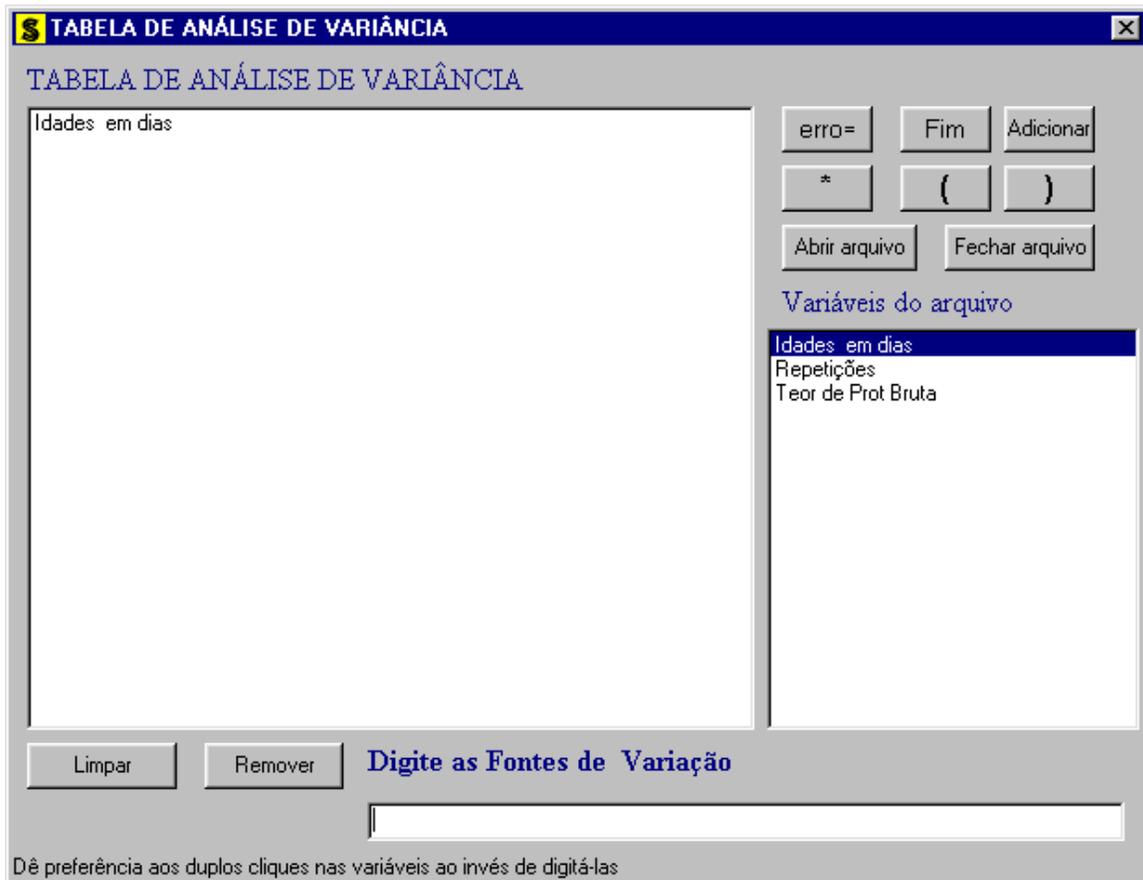
Idades de corte	REPETIÇÕES				
	I	II	III	IV	V
30	9,2	7,4	9,0	12,0	9,4
60	12,1	11,4	10,8	12,3	11,6
90	13,2	13,0	11,9	12,5	12,7
120	10,2	11,6	11,4	11,1	10,7

Uma parte do arquivo de dados do Sisvar está apresentado a seguir:



Idades em dias	Repetições	Teor de Prot Bruta
60	5	11.6
90	1	13.2
90	2	13
90	3	11.9
90	4	12.5
90	5	12.7
120	1	10.2
120	2	11.6
120	3	11.4
120	4	11.1
120	5	10.7

A tela de análise de variância com o quadro e as variáveis desse arquivo exemplo4.db, estão apresentadas a seguir. Nota-se que praticamente é um esquema análogo ao da seção anterior. Na tela seguinte que se seguirá a essa, de opções do quadro de análise, deve se clicar em Idades em dias e seleccionar opção de regressão. Se o processo for repetido, pode-se escolher novamente regressão, tantas vezes quanto convier ao usuário, até o limite de sete vezes. Esse procedimento tem a finalidade de se optar por diferentes modelos de regressão. O Sisvar disponibiliza 24 termos de regressão para que usuário formule o seu próprio modelo dentre os que julgar interessante. Para o caso de regressões polinomiais, seis termos (até grau 6) estão disponibilizados. É conveniente chamar a atenção de que o método de análise dos modelos de regressão é o de quadrados mínimos, executado via álgebra matricial. Dessa forma **nenhuma tipo de preocupação** deve passar pela mente do pesquisador, tal como o a da exigência de níveis equidistantes para se usar os modelos de regressão polinomial.



Selecionada a opção de regressão nos formulários apropriados o Sisvar apresenta aquele próprio para escolha do modelo de regressão. Nesse caso, foi escolhido um modelo de regressão polinomial. Como o total de tratamentos é igual a 4, com três graus de liberdade disponíveis, é possível selecionar um modelo com no máximo 2 termos, para que haja pelo menos 1 grau de liberdade para se testar o desvio de regressão. Nesse exemplo escolheu-se o termo X e X<sup>2</sup>, para o ajuste do seguinte modelo:

$$\bar{Y}_i = b_0 + b_1X_i + b_2X_i^2 + \delta_i + \varepsilon_i$$

em que, b<sub>i</sub>'s são os parâmetros a ser estimados,  $\delta_i$  é o desvio de regressão e  $\varepsilon_i$  o erro experimental suposto normal e independentemente distribuído.



Os resultados do Sisvar serão apresentados a seguir, separando-se a análise de variância da análise de regressão, para uma discussão mais simples.

Arquivo analisado:  
 C:\daniel\Tutoria em matemática e estatística\Exemplos\exemplo4.DB  
 Variável analisada: Teor de Prot Bruta

TABELA DE ANÁLISE DE VARIÂNCIA

FV	GL	SQ	QM	Fc	Pr>Fc
Idades em dias	3	28.013500	9.337833	10.202	0.0005
erro	16	14.644000	0.915250		
Total corrigido	19	42.657500	2.245132		
CV (%) =	8.56				
Média geral:	11.1750000	Número de observações:	20		

Pelo teste F da análise de variância detecta-se um efeito altamente significativo ( $P < 0,01$ ) da idade no teor de proteína bruta dessa gramínea. Como próximo passo, é necessário

explorar esse efeito por meio de um modelo de regressão, cuja escolha recaiu em um modelo polinomial de segundo grau. Os resultados dessa análise estão apresentados a seguir:

Regressão para a FV Idades em dias					
-----					
Média harmônica do número de repetições (r): 5					
Erro padrão da média geral dessa FV: 0.427843429305625					
-----					
b1 : X					
b2 : X^2					
-----					
Modelos reduzidos seqüenciais					
-----					
-----					
Parâmetro	Estimativa	SE	t para H0: Par=0	Pr> t	
-----					
b0	9.720000	0.52399905	18.550	0.0000	
b1	0.019400	0.00637791	3.042	0.0078	
-----					
R^2 = 30.23%					
Valores da variável independente					
	Médias observadas	Médias estimadas			
-----					
30.000000	9.400000	10.302000			
60.000000	11.640000	10.884000			
90.000000	12.660000	11.466000			
120.000000	11.000000	12.048000			
-----					
-----					
Parâmetro	Estimativa	SE	t para H0: Par=0	Pr> t	
-----					
b0	4.845000	1.19106570	4.068	0.0009	
b1	0.181900	0.03621958	5.022	0.0001	
b2	-0.001083	0.00023769	-4.558	0.0003	
-----					
R^2 = 98.10%					
Valores da variável independente					
	Médias observadas	Médias estimadas			
-----					
30.000000	9.400000	9.327000			
60.000000	11.640000	11.859000			
90.000000	12.660000	12.441000			
120.000000	11.000000	11.073000			
-----					
Somas de quadrados seqüenciais - Tipo I (Type I)					
-----					
Causas de Variação	G.L.	S.Q.	Q.M.	Fc	Prob.<F
-----					
b1	1	8.468100	8.468100	9.252	0.008
b2	1	19.012500	19.012500	20.773	0.000
Desvio	1	0.532900	0.532900	0.582	0.457
Resíduo	16	14.644000	0.915250		
-----					

Pela análise desses resultados pode-se verificar que existe efeitos linear e quadrático altamente significativos ( $P < 0,01$ ) e desvio de regressão não significativo. Portanto, o modelo

quadrático parece ser plausível para explicar a variação no teor de proteína bruta dessa gramínea em função da idade de corte. Como esse modelo apresentou um  $R^2$  de 98,1%, pode-se concluir que esse modelo possui todas as características desejáveis de um modelo ajustado, ou seja, desvio de regressão não significativo, todos os termos do modelo explicando uma fração significativa ( $P < 0,05$ ) da variação total da variável dependente, além de apresentar um elevado valor de  $R^2$ . O resultado do Sisvar para a regressão envolve sempre um ajuste sequencial dos termos nele especificado. Cabe ao usuário fazer a escolha baseado em fundamentações teóricas de qual deles é o mais apropriado. Nesse caso o modelo linear simples deve ser desprezado. O modelo ajustado é portanto:

$$\hat{Y}_i = 4,845 + 0,1819 X_i - 0,001083 X_i^2$$

## 6.2. Delineamento em blocos casualizados

Um dos mais importantes e mais usados delineamentos da pesquisa científica em todo o mundo é delineamento em blocos casualizados. Esse delineamento tem o apelo de eliminar da variação residual variações de natureza de heterogeneidade do material experimental. Isso é feito subdividindo o material ou a área experimental em frações mais uniformes e aplicando em cada uma delas todos os tratamentos. Essas frações uniformes são conhecidas por controle local ou simplesmente por blocos. Em cada um desses blocos os tratamentos são aleatorizados, sendo portanto, cada bloco correspondente a uma repetição dos tratamentos. A sua modelagem é feita adicionando-se simplesmente o efeito de blocos ao modelo. Nenhuma dificuldade adicional surge na realização das análises pelos programas de análise estatística, no que se refere as suas interações com os usuários. Internamente, esses programas executam muitos outros cálculos que ao usuário não é relevante, e nem é objetivo desse material, o seu destaque. O modelo geral de análise desse delineamento é:

$$Y_i = \mu + t_i + b_j + \varepsilon_{ij}$$

em que,  $t_i$  é o efeito do tratamento  $i$ ;  $b_j$  é o efeito do bloco  $j$  e  $\varepsilon_{ij}$  é o efeito do erro da parcela que recebeu o tratamento  $i$  do bloco  $j$ , supostos normal e independentemente distribuídos.

Um exemplo será apresentado para demonstrar como modelar esse esquema de análise no Sisvar. Um ensaio de ganho de pesos em animais para testar 4 rações. Os animais foram blocados

quanto a categorias de peso especificadas pelo pesquisador. Seis blocos foram usados para o delineamento em questão. A Tabela 6 apresenta os dados desse experimento.

**Tabela 6. Ganho de peso dos bovinos de corte em kg obtidos em um delineamento em blocos casualizados com 6 repetições.**

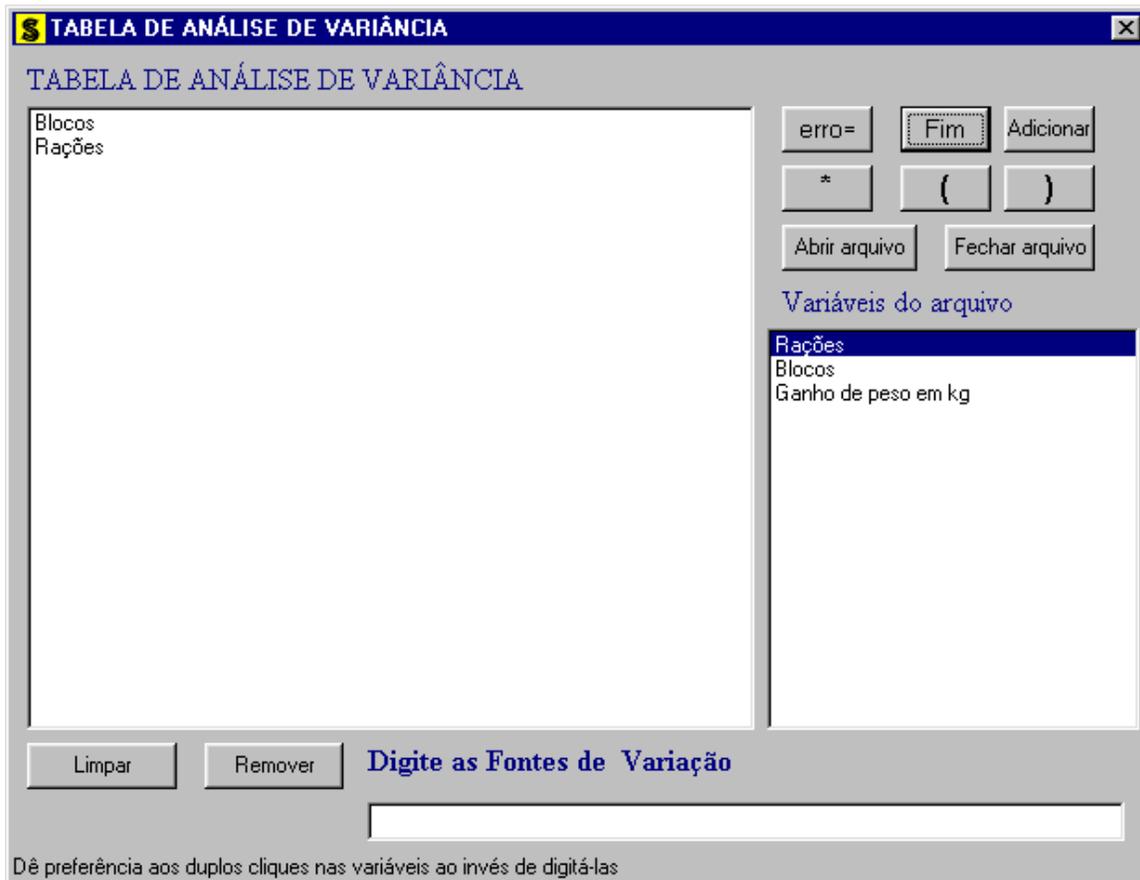
Rações	BLOCOS					
	I	II	III	IV	V	VI
A	31	19	25	14	21	22
B	21	19	22	15	18	19
C	15	18	16	14	17	16
D	24	25	21	23	22	23

## SISVAR

Os dados após serem digitados no Sisvar apresentarão a estrutura resumida apresentada a seguir.

Rações	Blocos	Ganho de peso em kg
C	2	18
C	3	16
C	4	14
C	5	17
C	6	16
D	1	24
D	2	25
D	3	21
D	4	23
D	5	22
D	6	23

Ao realizar as análises de variância o seguinte modelo deve ser especificado na fase de planejamento da análise. Para realizar essa análise específica basta seguir o exemplo apresentado no formulário que se segue.



Com esse esquema e escolhendo-se a opção de Scott e knott para o teste para rações, o seguinte resultado é fornecido pelo Sisvar.

```
Arquivo analisado:
C:\daniel\Tutoria em matemática e estatística\Exemplos\exemplo5.DB
Variável analisada: Ganho de peso em kg
```

TABELA DE ANÁLISE DE VARIÂNCIA					
FV	GL	SQ	QM	Fc	Pr>Fc
Blocos	5	84.500000	16.900000	1.958	0.1440
Rações	3	180.000000	60.000000	6.950	0.0037
erro	15	129.500000	8.633333		
Total corrigido	23	394.000000	17.130435		
CV (%) =	14.69				
Média geral:	20.0000000	Número de observações:	24		

A análise de variância desse modelo resultou na rejeição da hipótese de igualdade dos efeitos de rações pelo teste F, ao valor nominal de significância de 1% ( $P < 0,0037$ ).

Teste Scott-Knott (1974) para a FV Rações		
-----		
NMS: 0.05		
-----		
Média harmonica do número de repetições (r): 6		
Erro padrão: 1.19953694769644		
-----		
Tratamentos	Médias	Resultados do teste
-----		
C	16.000000	a1
B	19.000000	a1
A	22.000000	a2
D	23.000000	a2
-----		

O teste de Scott e Knott agrupou as rações em dois grupos distintos: o primeiro dos menores ganhos de peso formado pelas rações C e B, e o segundo dos maiores ganhos de peso formado pelas rações A e D.

O teste de Tukey apresenta a ambiguidade e em geral tem baixíssimo poder, embora controle adequadamente as taxas de erro tipo I por experimento.