

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA

ANÁLISE ESTATÍSTICA USANDO O
STATISTICA® 6.0

Prof. Paulo José Ogliari, Dr.
Eng. Juliano Anderson Pacheco

FLORIANÓPOLIS – SC
ABRIL DE 2004

SUMÁRIO

1. INTRODUÇÃO.....	3
2. GERENCIAMENTO DE ARQUIVOS.....	5
2.1. CRIAR UM NOVO ARQUIVO	6
2.2. IMPORTAÇÃO DE ARQUIVOS DO <i>EXCEL</i>	10
3. ANÁLISE EXPLORATÓRIA DE DADOS	13
3.1. DISTRIBUIÇÃO DE FREQÜÊNCIA DE VARIÁVEIS QUALITATIVAS	13
3.1.1. <i>Unidimensional</i>	13
3.1.2. <i>Bidimensional</i>	23
3.2. DISTRIBUIÇÃO DE FREQÜÊNCIA DE VARIÁVEIS QUANTITATIVAS	35
3.2.1. <i>Tabela de Distribuição de Freqüência para Variável Quantitativa Discreta</i>	36
3.2.2. <i>Representações Gráficas de Variáveis Quantitativas Discretas</i>	38
3.2.3. <i>Dados Agrupados em Classes</i>	40
3.2.4. <i>Polígono de Freqüências Múltiplos</i>	44
3.2.5. <i>O Modelo Normal</i>	49
3.3. ESTATÍSTICAS DESCRIPTIVAS	50
3.3.1. <i>Medidas de Tendência Central e Dispersão</i>	50
3.3.2. <i>A Média e o Desvio Padrão sob um Modelo Normal</i>	56
3.3.3. <i>Desenho Esquemático (Box Plot)</i>	59
3.3.4. <i>Diagrama de Dispersão</i>	62
3.3.5. <i>Coeficiente de Correlação</i>	64
3.3.6. <i>Ajuste da Equação de uma Reta</i>	66
4. OPÇÕES GRÁFICAS	72
5. TESTES ESTATÍSTICOS PARA COMPARAÇÃO DE DUAS MÉDIAS.....	79
5.1. TESTE DE DUAS MÉDIAS POPULACIONAIS COM VARIÂNCIAS DESCONHECIDAS	79
5.1.1. <i>Amostras Independentes com Variâncias Iguais e Desconhecidas</i>	79
5.1.2. <i>Amostras Independentes com Variâncias Desiguais e Desconhecidas</i>	85
5.1.3. <i>Amostras Dependentes (Dados Paresados)</i>	89
5.1.4. <i>Intervalo de Confiança para a Diferença de Duas Médias no Caso de Dados Paresados</i>	93
6. TESTE PARA DUAS OU MAIS PROPORÇÕES	96
6.1. O TESTE DO QUI-QUADRADO (<i>CHI-SQUARE</i>).....	96
7. ANÁLISE DE VARIÂNCIA.....	115
7.1. DELINEAMENTO INTEIRAMENTE CASUALIZADO.....	116
7.2. EXEMPLO	116
7.3. ANÁLISE ESTATÍSTICA	118
7.3.1. <i>Análise de Variância</i>	120
7.3.2. <i>Comparações Múltiplas de Médias</i>	122
7.3.3. <i>Contrastes Ortogonais</i>	124
7.3.4. <i>Diagnóstico do Modelo</i>	127

1. INTRODUÇÃO

O software *STATISTICA®* é um programa integrado para gerenciar Análise Estatística e Bases de Dados, caracterizando uma ampla seleção do processo analítico, do básico ao avançado, para as mais diversas áreas - Ciências Biomédicas, Biológicas, Engenharias, Ciências Sociais, Agrárias - permitindo a realização de *Data Mining* (Mineração de Dados).

O sistema não inclui somente procedimentos estatísticos e gráficos gerais, mas, também, módulos especializados (Análise de Regressão, Análise de Sobrevida, Séries Temporais, Análise Fatorial, Análise Discriminante e diversos outros módulos). Dificilmente, o usuário necessitará de todos os módulos oferecidos pelo *STATISTICA®* para a análise do seu trabalho, e será, a partir de sua criatividade e conhecimento teórico, que chegará as respostas dos seus objetivos.

Neste trabalho, foram desenvolvidos três assuntos: a) Análise exploratória de dados, b) Alguns testes de hipóteses sobre médias e proporções e c) Análise de variância com um fator (*one-way ANOVA*).

Esta apostila foi elaborada com o propósito de auxiliar os estudantes de graduação e pós-graduação e demais interessados, na iniciação do uso de um programa estatístico. Escolheu-se o programa *STATISTICA®*, por apresentar uma plataforma interativa com o usuário e, também, por ser ele dos programas mais completos da área existente na atualidade.

Pede-se que qualquer problema encontrado na apostila, entrar em contato preferencialmente para jap@inf.ufsc.br ou para ogliari@inf.ufsc.br.

Os capítulos 2 e 3 serão baseados no seguinte exemplo:

Quadro 1.1 - Resultados de um experimento de competição de híbridos de milho para a região preferencial I, com altitudes abaixo de 800m - safra: 1987/1988

Cultivar	Rendimento médio (kg/ha)	Ciclo (dias)	Altura planta (cm)	Altura espiga (cm)	Tipo de grão	Ferrugem* (escala)
1	6388	65	242	103	dentado	r
2	6166	65	258	134	semi-dentado	r
3	6047	65	240	104	semi-dentado	s
4	5889	66	243	108	semi-dentado	s
5	5823	69	257	128	dentado	ms
6	5513	68	241	108	semi-dentado	s
7	5202	64	235	108	dentado	r
8	5172	68	240	103	dentado	s
9	5166	69	253	123	dentado	ms
10	4975	70	250	117	semi-dentado	ms
11	4778	70	242	114	dentado	mr
12	4680	66	245	111	semi-duro	ms
13	4660	69	239	110	semi-duro	mr
14	5403	73	264	138	dentado	ms
15	5117	76	282	149	dentado	mr
16	5063	72	274	151	dentado	r
17	4993	71	279	134	semi-dentado	r
18	4980	72	274	140	dentado	ms
19	4770	73	244	140	dentado	r
20	4685	71	265	139	semi-duro	mr
21	4614	73	248	110	semi-dentado	r
22	4552	73	265	128	semi-dentado	r
23	3973	74	261	124	semi-dentado	mr
24	4550	71	259	129	semi-duro	s
25	5056	64	252	104	semi-duro	mr
26	4500	70	271	109	dentado	ms
27	4760	68	243	137	semi-duro	r
28	5110	66	252	141	semi-dentado	ms
29	4960	70	262	120	dentado	ms
30	4769	73	260	118	dentado	r
31	4849	74	250	119	semi-dentado	s
32	5230	71	255	138	semi-duro	s

Fonte: Boletim Técnico, número 39, EMPASC S. A.

* r = resistente, mr = moderadamente resistente, s = susceptível e ms = moderadamente susceptível.

2. GERENCIAMENTO DE ARQUIVOS

O STATISTICA® trabalha com variáveis categorizadas, qualitativas e quantitativas, permitindo a formação de grupos que serão analisados. Ou seja, utiliza uma técnica mais apropriada para resumir as informações, adaptando qualquer tipo de variável ao caso estudado e armazena as variáveis mediante notação dupla, isto é, utiliza os valores da variável de forma categorizada (texto) e na forma de números, que podem ser valores codificados (numéricos).

Por exemplo, a variável Ferrugem, apresentada no Quadro 1.1, foi dividida em categorias (quatro níveis): resistente (r), suscetível (s), moderadamente suscetível (ms) e moderadamente resistente (mr). O STATISTICA® usará os códigos numéricos internamente, mas o usuário pode sempre referenciar a ferrugem usando os textos respectivos, conforme apresentado na figura 2.1.

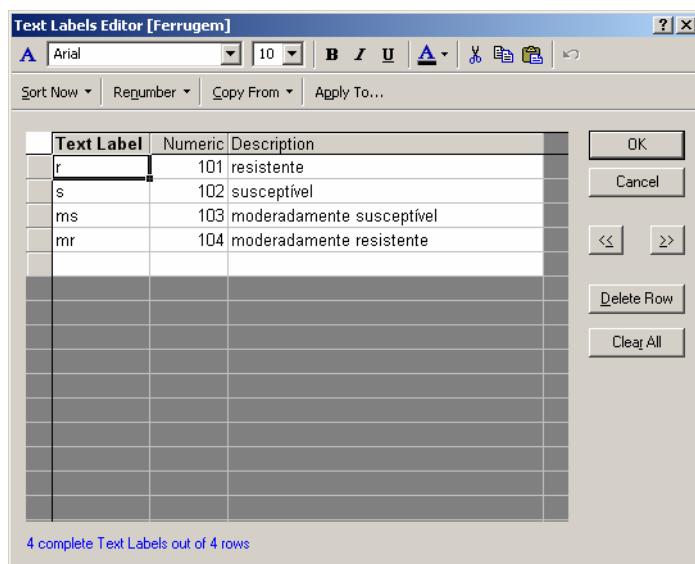


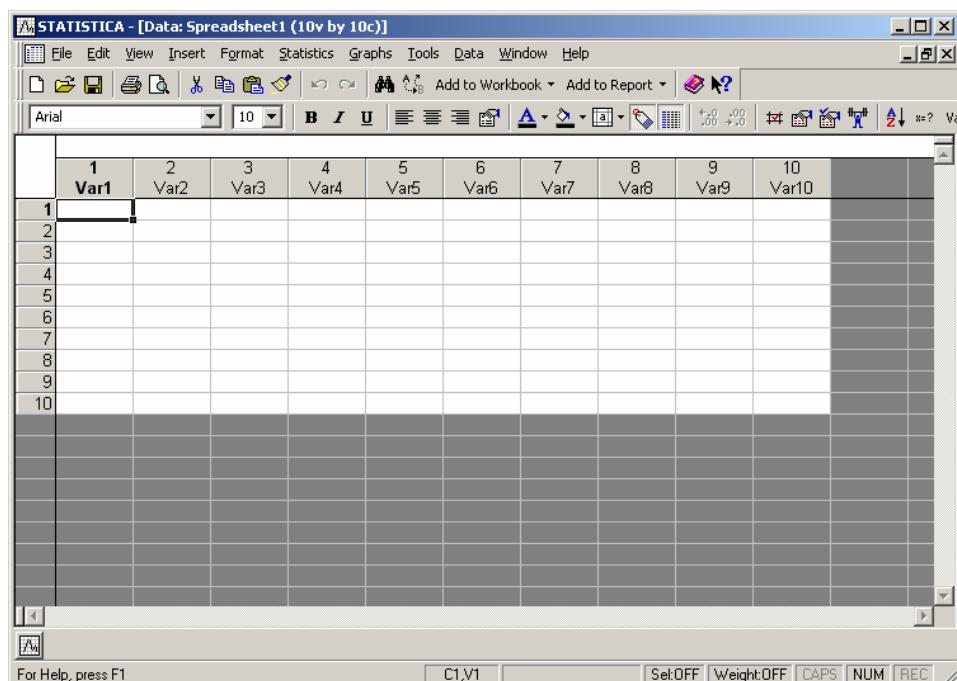
Figura 2.1 – Administrador de valores

Feita a coleta de dados, através de censos, de levantamentos por amostragem ou de experimentos, geralmente estes dados se apresentam de maneira desorganizada, ainda sem valor informativo sobre o fenômeno em estudo. Portanto, os mesmos devem ser organizados para possibilitarem a análise.

A matriz de dados pode ser criada ou importada. A seguir estão os procedimentos para criar um novo arquivo e importar uma planilha do *Excel*.

2.1. Criar um novo arquivo

- (1) Abra o programa STATISTICA®;
- (2) Surgirá um arquivo de dados vazio com 10 colunas (ou variáveis) e 10 linhas (ou casos), conforme apresentado na Figura 2.2:



The screenshot shows the STATISTICA software interface with a title bar 'STATISTICA - [Data: Spreadsheet1 (10v by 10c)]'. The menu bar includes File, Edit, View, Insert, Format, Statistics, Graphs, Tools, Data, Window, and Help. The toolbar below the menu bar contains various icons for file operations, data entry, and analysis. The main window is a spreadsheet grid with 10 columns labeled '1 Var1' through '10 Var10' and 10 rows labeled '1' through '10'. The cell at the top-left, '1 Var1', is highlighted with a black border. The rest of the grid is empty. The status bar at the bottom shows 'For Help, press F1', 'C1,V1', 'Sel:OFF', 'Weight:OFF', 'CAPS', 'NUM', and 'REC'.

Figura 2.2 – Arquivo de dados inicial

- (3) Geralmente é necessário alterar esse arquivo inicial, com relação ao número de variáveis ou de casos para permitir a digitação dos dados coletados
- (4) Esse arquivo será alterado para possibilitar a entrada dos dados apresentados no Quadro 1.1, logo é necessário excluir 3 variáveis e incluir mais 22 casos, gerando uma tabela com 7 colunas e 32 linhas
- (5) Para excluir as 3 colunas, selecione as três últimas colunas, clique com o botão direito do mouse e escolha a opção [DELETE VARIABLES], proceder conforme apresentado nas Figuras 2.3 e 2.4:

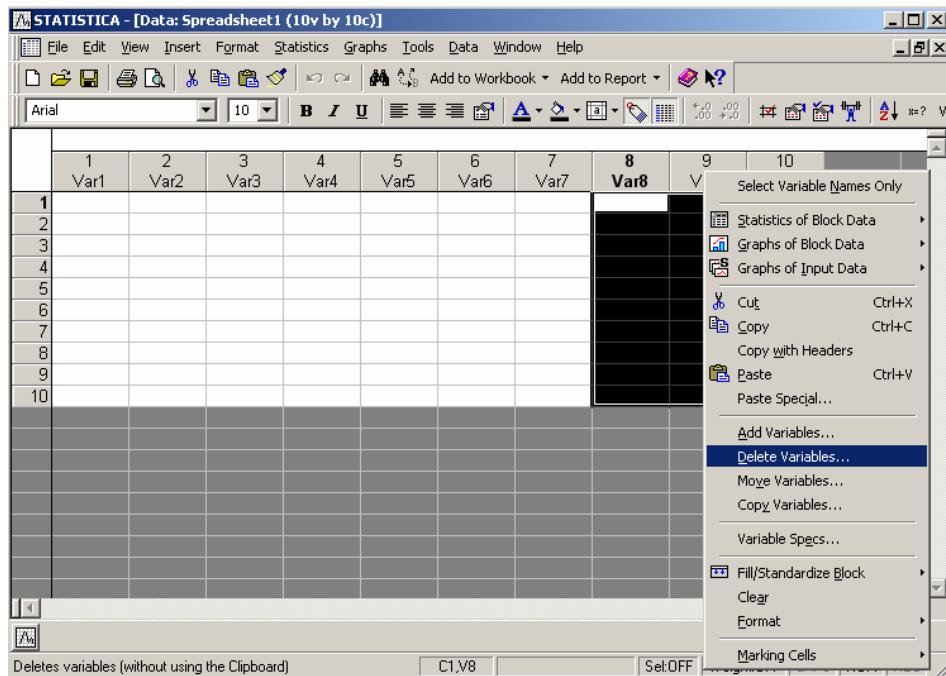


Figura 2.3 – Procedimento para excluir variáveis

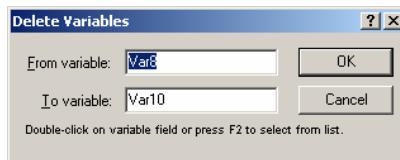


Figura 2.4 – Definição das variáveis a excluir

(6) Para incluir os 22 casos, selecione a última linha, clique com o botão direito do mouse e escolha a opção [ADD CASES], proceder conforme apresentado nas Figuras 2.5 e 2.6:

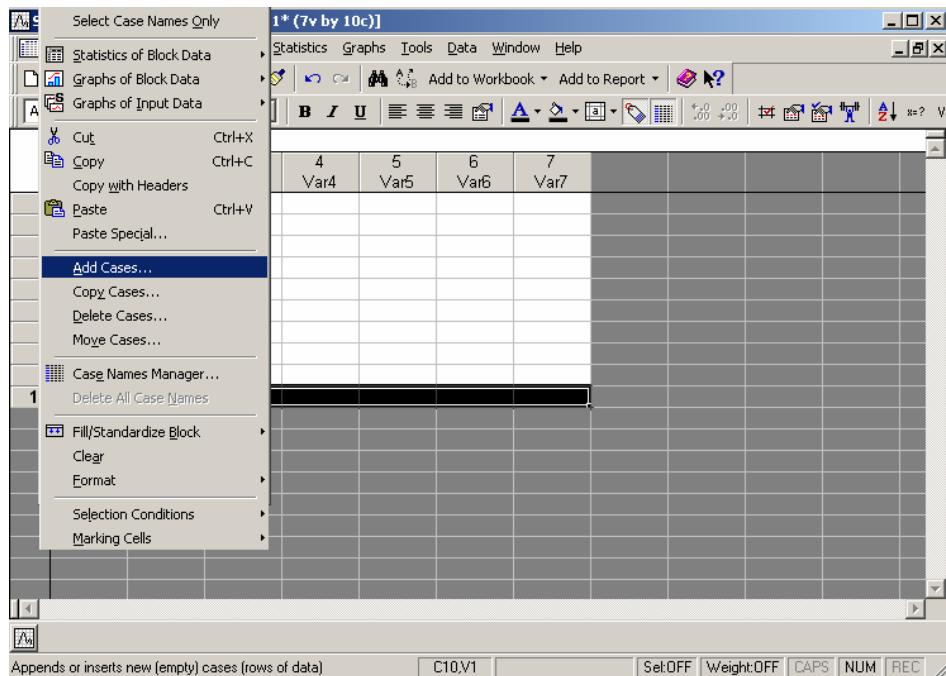


Figura 2.5 – Procedimento para incluir casos

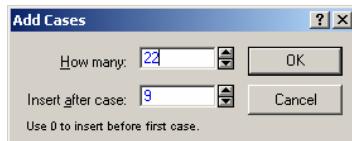


Figura 2.6 – Definição do número de casos a incluir

(7) o arquivo deve ser salvo, para isso no menu [FILE], escolha a opção [SAVE] e defina o nome do arquivo, conforme apresentado na Figura 2.7:

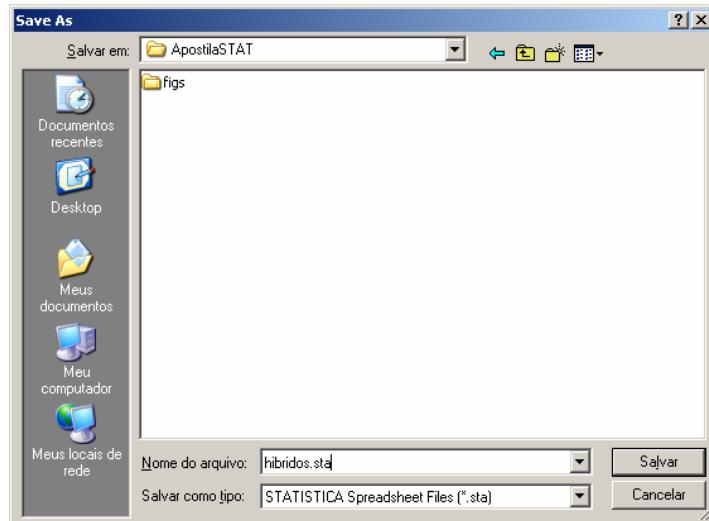


Figura 2.7 – Gravação do arquivo criado

(8) Neste momento dispomos de uma tabela possível para inserir os dados apresentados no Quadro 1.1, mas primeiramente é necessário definir os nomes das variáveis e os tipo de dados a serem inseridos

(9) Clique duas vezes em cima da primeira coluna e aparecerá a janela apresentada na Figura 2.8:

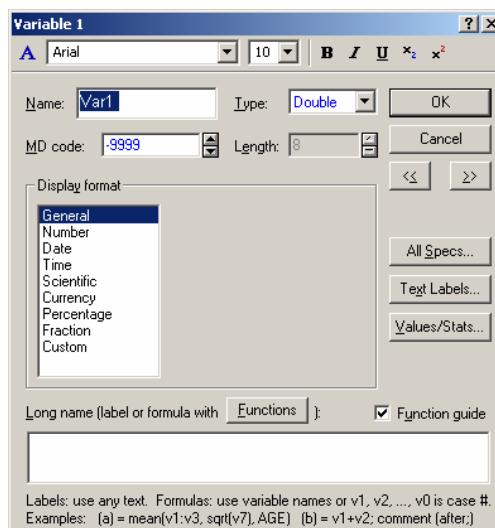


Figura 2.8 – Definição das variáveis

onde:

Name: corresponde ao nome da variável

Type: corresponde ao tipo da variável, onde os valores possíveis são: *Double* (valores com decimais), *Integer* (valores inteiros), *Text* (texto) e *Byte* (valores inteiros entre 0 e 255)

Display format: formato de apresentação dos dados

Observação: Para acessar a janela da Figura 2.1 e digitar os respectivos textos da variável Ferrugem, basta clicar no botão [Text Labels] da janela apresentada na Figura 2.8.

(10) No Quadro 2.1 estão as possíveis definições das 7 variáveis apresentadas no Quadro 1.1:

Quadro 2.1. Definições das variáveis

Variável	Name	Type	Display Format
Cultivar	Cultivar	Byte	Number
Rendimento médio (kg/ há)	Rendimento	Integer	Number
Ciclo (dias)	Ciclo	Byte	Number
Altura planta (cm)	Altura_planta	Integer	Number
Altura espiga (cm)	Altura_espiga	Integer	Number
Tipo de grão	Tipo_grao	Text	-
Ferrugem (escala)	Ferrugem	Text	-

(11) Agora é possível digitar os dados, conforme apresentado na Figura 2.9:

	1	2	3	4	5	6	7	
	Cultivar	Rendimento	Ciclo	Altura_planta	Altura_espiga	Tipo_grao	Ferrugem	
1	1	6398	65	242	103	dentado	r	
2	2	6166	65	258	134	semi-dentado	r	
3	3	6047	65	240	104	semi-dentado	s	
4	4	5889	66	243	108	semi-dentado	s	
5	5							
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								

Figura 2.9 – Digitação dos dados do Quadro 1

(12) Após a digitação dos dados, salvar novamente o arquivo.

2.2. Importação de Arquivos do Excel

Existem duas maneiras de importar arquivos no formato *Excel*. A primeira que será descrita é ideal no caso de planilhas pequenas, pois é fácil selecionar a matriz de dados usando somente o *mouse*. No caso de grandes arquivos, utilize a segunda opção.

1^a Opção:

(1) Abra o aplicativo *Excel*

(2) Carregue o arquivo de dados que se deseja transferir para o *STATISTICA®* e copie para a área de transferência a região da planilha que contém os dados, conforme apresentado na Figura 2.10:

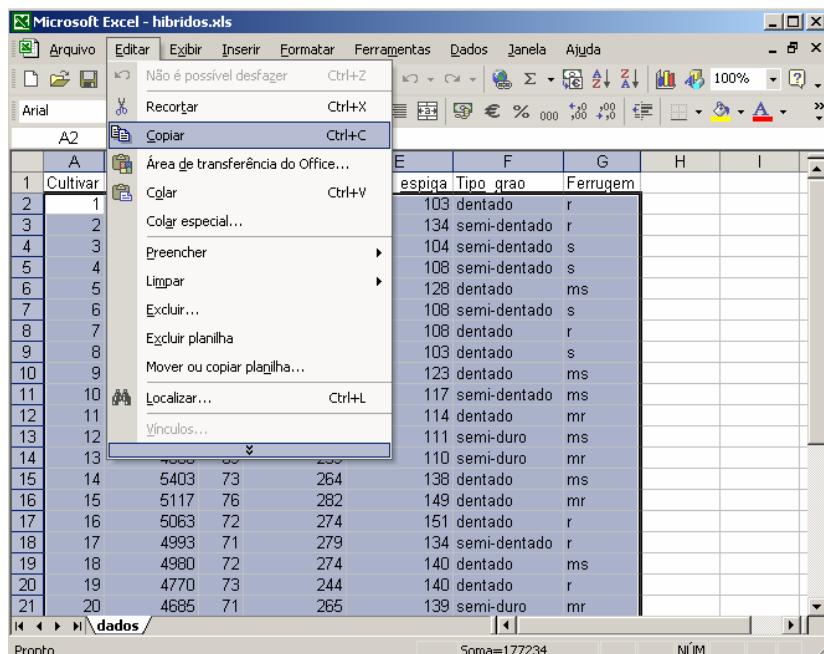
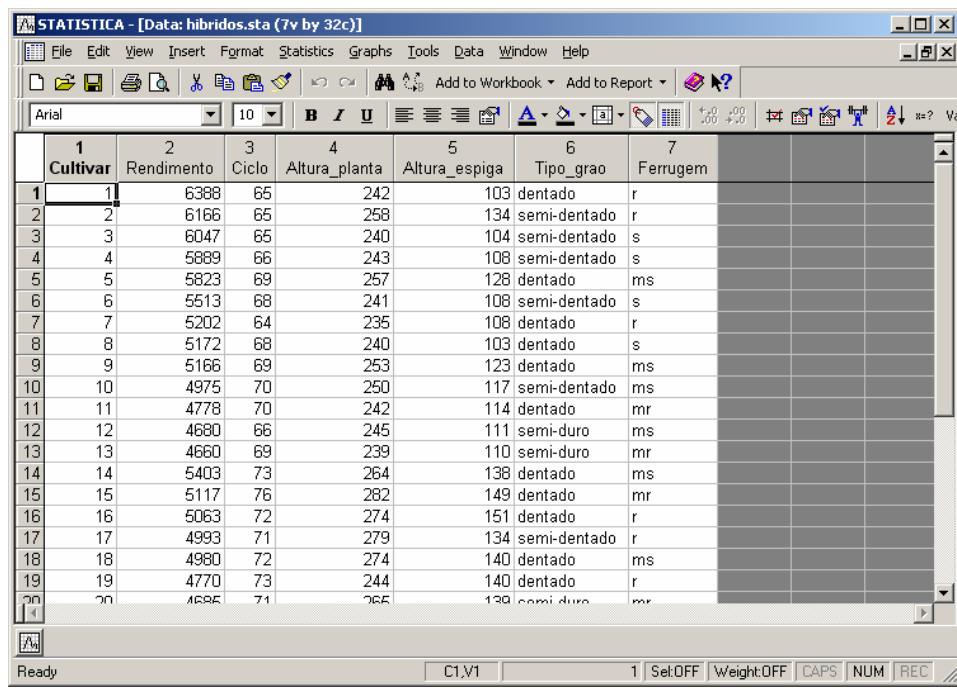


Figura 2.10 – Cópia dos dados para a área de transferência

(3) No *STATISTICA®*, crie um novo arquivo com o mesmo tamanho que a região dos dados no *Excel* (mesmo número de colunas e linhas), o item 2.1 ensina como criar um novo arquivo

(4) Criado o arquivo, selecione a primeira célula cole os dados da área de transferência, através do menu [EDIT], opção [PASTE] e o resultado está apresentado na Figura 2.11:



1	2	3	4	5	6	7
Cultivar	Rendimento	Ciclo	Altura_planta	Altura_espiga	Tipo_grao	Ferrugem
1	6388	65	242	103	dentado	r
2	6166	65	258	134	semi-dentado	r
3	6047	65	240	104	semi-dentado	s
4	5889	66	243	108	semi-dentado	s
5	5823	69	257	128	dentado	ms
6	5513	68	241	108	semi-dentado	s
7	5202	64	235	108	dentado	r
8	5172	68	240	103	dentado	s
9	5166	69	253	123	dentado	ms
10	4975	70	250	117	semi-dentado	ms
11	4778	70	242	114	dentado	mr
12	4680	66	245	111	semi-duro	ms
13	4660	69	239	110	semi-duro	mr
14	5403	73	264	138	dentado	ms
15	5117	76	282	149	dentado	mr
16	5083	72	274	151	dentado	r
17	4993	71	279	134	semi-dentado	r
18	4980	72	274	140	dentado	ms
19	4770	73	244	140	dentado	r
20	4655	71	265	139	semi duro	mr

Figura 2.11 – Resultado da cópia dos dados da planilha do Excel

(5) Salve o novo arquivo criado, através desse processo.

2^a Opção:

(1) Abrir o programa STATISTICA®;

(2) No menu [FILE], escolher a opção [Open], aparecerá a janela apresentada na Figura 2.12, onde o formato a ser escolhido será *Excel files (*.xls)*

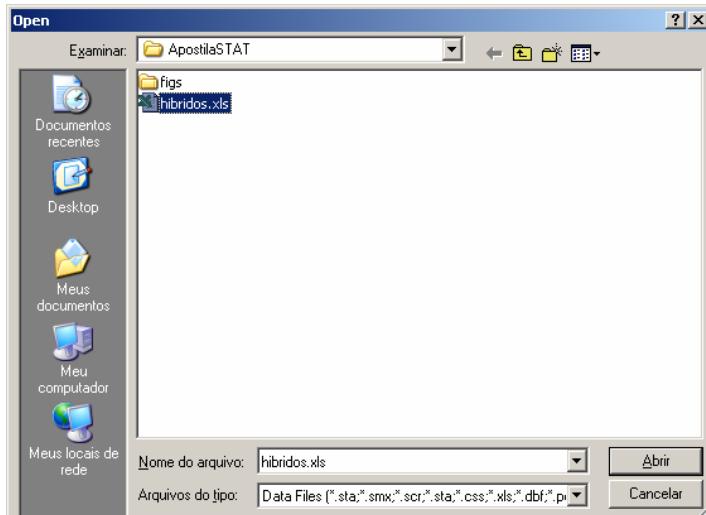


Figura 2.12 – Importação de arquivos

(3) Selecione a planilha a ser importada

(4) Será apresentada a janela mostrada na Figura 2.13, que representa um resumo da tabela de dados a ser importada;

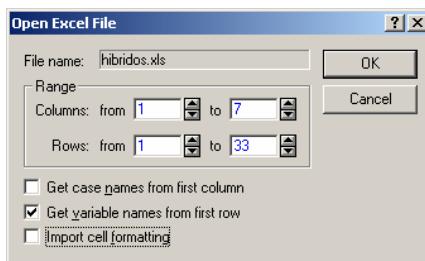


Figura 2.13 – Resumo da planilha a ser importada

onde:

Range Columns: representa o número de variáveis existentes na tabela de dados

Range Rows: representa o número de casos existentes na tabela de dados

Get case names from first column: atribui nomes aos casos conforme a primeira coluna da tabela de dados

Get variable names from first row: atribui nomes às variáveis conforme a primeira linha da tabela de dados

Import cell formatting: importa a tabela mantendo a formatação existente no *Excel*

(5) O resumo deve ser verificado quanto ao número de variáveis e casos existentes na tabela de dados, se o mesmo diferir dos valores corretos, deve-se fazer o seguinte processo: copiar a área dos dados, juntamente com os nomes das variáveis e casos se existirem, para um novo arquivo do *Excel*, salvar e repetir todos os itens anteriores

(6) Se o quadro resumo estiver correto, prossiga e o resultado está apresentado na Figura 2.14:

1	2	3	4	5	6	7
Cultivar	Rendimento	Ciclo	Altura_planta	Altura_espiga	Tipo_grao	Ferrugem
1	11	6388	65	242	103	dentado
2	2	6166	65	258	134	semi-dentado
3	3	6047	65	240	104	semi-dentado
4	4	5889	66	243	108	semi-dentado
5	5	5823	69	257	128	dentado
6	6	5513	68	241	108	semi-dentado
7	7	5202	64	235	108	dentado
8	8	5172	68	240	103	dentado
9	9	5166	69	253	123	dentado
10	10	4975	70	250	117	semi-dentado
11	11	4778	70	242	114	dentado
12	12	4680	66	245	111	semi-duro
13	13	4680	69	239	110	semi-duro
14	14	5403	73	264	138	dentado
15	15	5117	76	282	149	dentado
16	16	5063	72	274	151	dentado
17	17	4993	71	279	134	semi-dentado
18	18	4980	72	274	140	dentado
19	19	4770	73	244	140	dentado
20	20	4655	71	265	139	semi-duro

Figura 2.14 – Resultado da importação da planilha do Excel

(7) salve o novo arquivo criado, através desse processo.

3. ANÁLISE EXPLORATÓRIA DE DADOS

Os dados recolhidos que devem ser estudados, ou seja, organizados, sumarizados e descritos, serão pelo STATISTICA®, explorados. Isto é, **procuraremos tirar o máximo de informações de um conjunto de dados** e, principalmente, tentar indicar um modelo que será utilizado em uma fase posterior da análise, a Inferência Estatística.

3.1. Distribuição de Freqüência de Variáveis Qualitativas

O estudo de distribuições de freqüências nos permite conhecer a forma, ou seja, a maneira de como os valores de uma variável se comportam, e assim, podemos ter uma boa idéia global dos valores e da sua distribuição.

- **Variável qualitativa** é aquela que apresenta seu resultado em forma de uma qualidade ou um atributo. Podem ser **ordinais**, quando existe uma ordem nos possíveis resultados da mesma ou **nominais**, quando não existe uma ordenação.

3.1.1. Unidimensional

- Distribuição da freqüência de **uma única variável**.

Nesta seção, vamos construir a **Tabela de Distribuição de Freqüência Unidimensional** e o **Gráfico de Barras** para a variável qualitativa ordinal resistência à ferrugem. A ordem é o grau de ataque da doença.

3.1.1.1. Tabela de Distribuição de Freqüência Unidimensional

Exemplo: Análise da distribuição de freqüência da variável resistência à ferrugem.

(1) Abra o programa STATISTICA® e o arquivo *hibridos.sta*

(2) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 3.1, onde a opção a ser escolhida será *Frequency Table*:

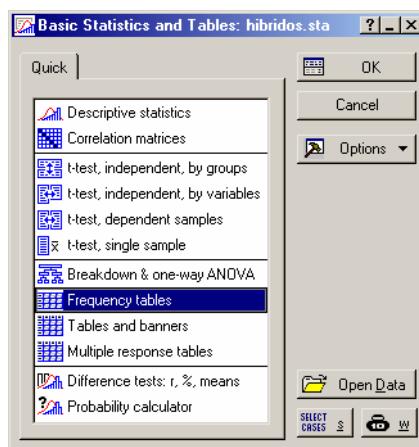


Figura 3.1 – Estatísticas básicas e tabelas

(3) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.2:

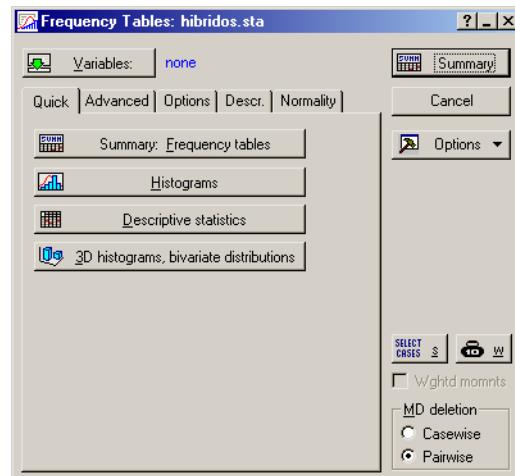


Figura 3.2 – Tabelas de freqüência

(4) Clique no botão [Variables] e selecione a variável Ferrugem, conforme apresentado na Figura 3.3. Se for mantido o botão esquerdo do *mouse* apertado, com o movimento de subida e descida é possível selecionar várias variáveis ao mesmo tempo, para pular alguma, mantenha apertado a tecla *CTRL* do teclado e com o botão esquerdo do *mouse* clique na variável de interesse:

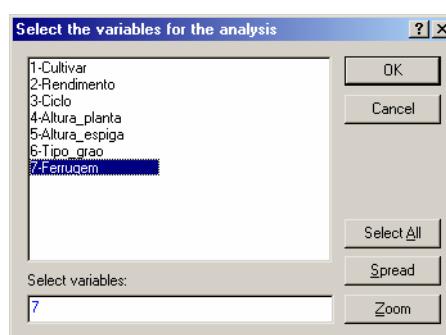


Figura 3.3. Janela para seleção das variáveis para a análise

(5) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.4:



Figura 3.4 – Tabela de freqüência com a variável ferrugem selecionada

(6) Clique na aba [Options] e defina as opções conforme apresentado na figura 3.5. Para desativar, basta clicar na opção correspondente, retirando assim, o símbolo . Para ativar, basta clicar na opção ainda não selecionada:

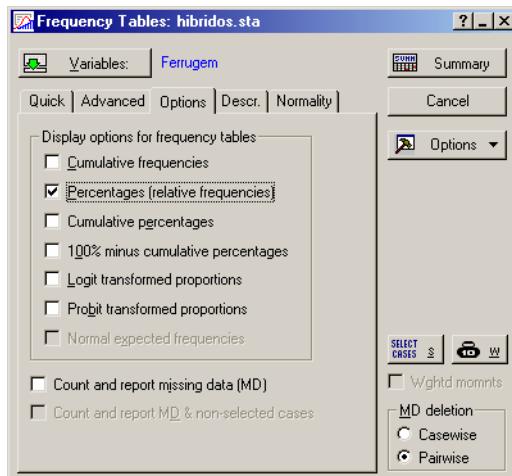


Figura 3.5 – Opções da tabela de freqüência

(7) Clique no botão [Summary] e o resultado está apresentado na Figura 3.6:

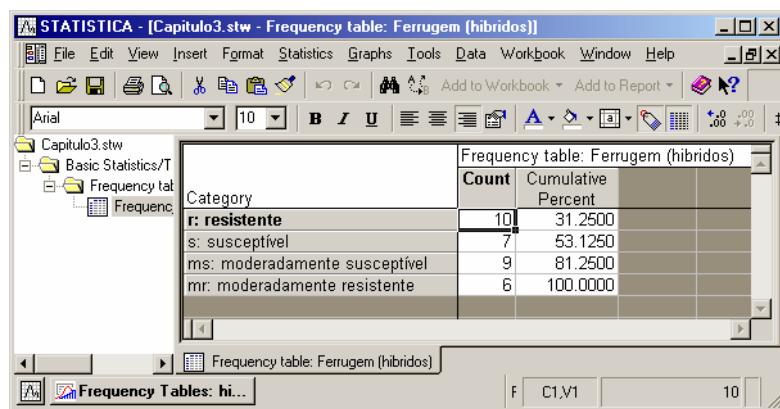


Figura 3.6 - Tabela de Distribuição de Freqüência Unidimensional da variável resistência à ferrugem.

Conclusão: a freqüência (ocorrência) da variável resistência à ferrugem na análise de 32 híbridos é:

- 10 resistentes (r) à ferrugem = 31,250%
- 7 susceptíveis (s) à ferrugem = 21,875%
- 9 moderadamente susceptíveis (ms) à ferrugem = 28,125%
- 6 moderadamente resistentes (mr) à ferrugem = 18,750%

Importante: a forma como o *STATISTICA®* organiza as janelas durante a análise é a seguinte:

- há o arquivo de dados com extensão **.sta**
- os resultados são gerados no *workbook*, estes podem ser gravados num arquivo com extensão **.stw**. No caso da Figura 3.6 foi gravado o arquivo **Capitulo3.stw**.
- após os resultados serem apresentados no *workbook*, a janela da análise é minimizada no canto inferior esquerdo. No caso da Figura 3.6 vê-se que a análise atual é *Frequency Tables*.

3.1.1.2. Tabela de Distribuição de Freqüência Unidimensional com fator estratificado

Conforme pode ser visto à direita da janela apresentada na Figura 3.5, existe o botão [Select Cases S] que permite filtrar os casos com base nas variáveis disponíveis. Abaixo estão dois exemplos da aplicação desse procedimento de filtragem.

(1) Clique no botão [Select Cases S] da janela apresentada na Figura 3.5. Aparecerá a janela apresentada na Figura 3.7:

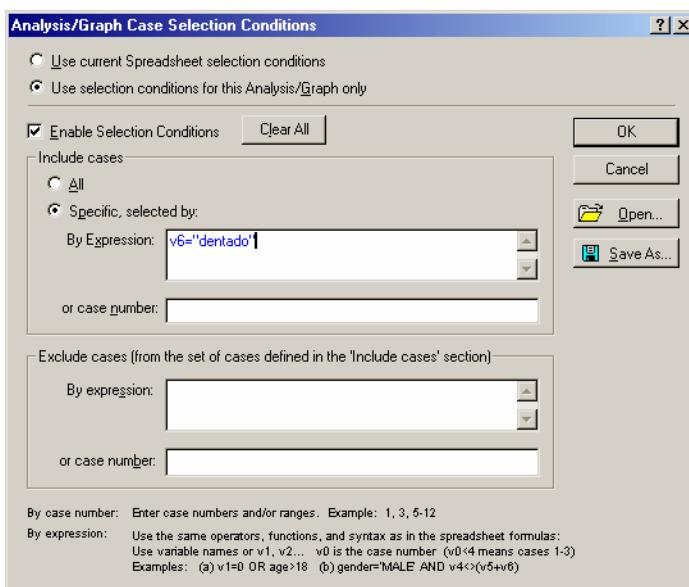


Figura 3.7 – Janela que permite especificar as condições

(2) Como demonstrado na Figura 3.7, serão incluídos somente os casos que a variável v6 (Tipo de grão) assume o valor “dentado”, que significará que a tabela construída mostrará somente a distribuição da resistência à ferrugem para tipo de grão dentados

(3) Para isso, clique em *Specific, selected by: By Expression*, digite a condição: v6=“dentado” e clique no botão [OK]

(4) Ao voltar para a janela da figura 3.5, clique novamente no botão [Summary] e o resultado está apresentado na Figura 3.8:

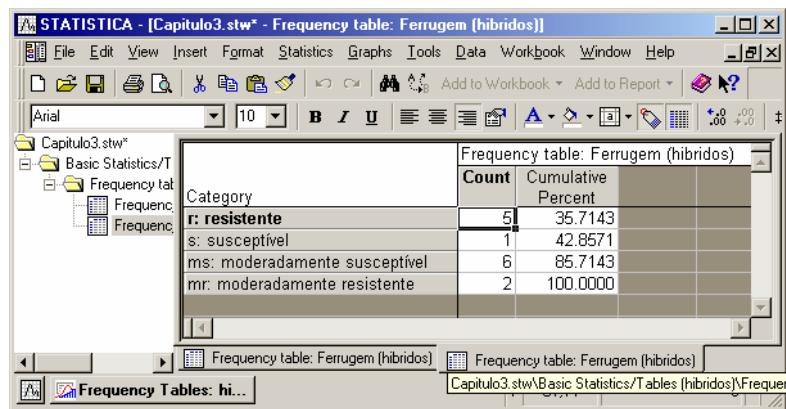


Figura 3.8 – Resultado da distribuição da freqüência da variável resistência à ferrugem para o tipo de grão dentados

(6) Para realizar um estudo para híbridos de ciclo curto, híbridos com ciclo de cultura entre 64 e 70 dias, altere a condição de filtragem conforme apresentado na Figura 3.9, onde a condição é: $v3>=64$ and $v3<=70$

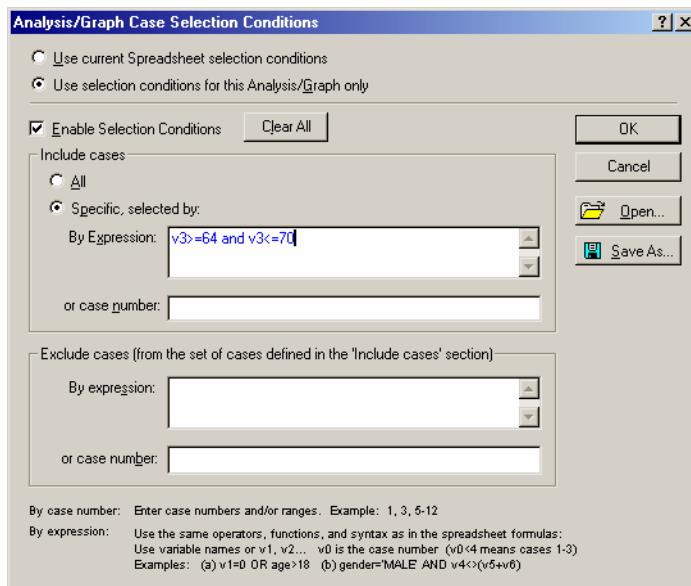


Figura 3.9 – Janela que permite especificar as condições

(7) Ao clicar no botão [OK] e voltar para a janela da figura 3.5, clique novamente no botão [Summary] e o resultado está apresentado na Figura 3.10:

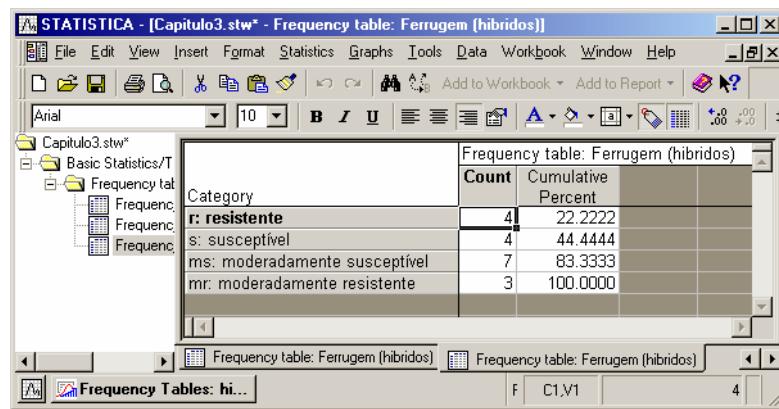


Figura 3.10 – Resultado da distribuição da freqüência da variável resistência à ferrugem para híbridos de ciclo curto

3.1.1.3. Gráfico de Colunas/Barras

O STATISTICA® é um programa que permite integrar um grande potencial gráfico com a análise estatística de dados num único aplicativo. A editoração gráfica é facilitada por meios de opções do sistema.

A construção de gráficos está disponível para qualquer módulo do STATISTICA®, visto que existe na barra do menu, a opção *Graphs*. Clicando nela aparecem as opções gráficas do aplicativo e cabe ao usuário conhecimento teórico e intuitivo para que seja encontrada a melhor forma de representação do fenômeno.

Serão construídos gráficos em duas e três dimensões, como 1º exemplo, um simples histograma de freqüência (gráfico de colunas/barras).

DICA: a compatibilidade do STATISTICA® com o Windows é plena, isto significa que todas as operações comuns no Windows são reconhecidas e aceitas. Assim sendo, cortar e colar se combinam com *Screen Catcher* (Opção em *Edit*, na *Barra de ferramentas do menu principal do STATISTICA®* que permite a captura da janela ou de um quadro), *Alt + Print Screen* (no teclado) que copia, trabalhando em conjunto com algum editor de figuras, como o *Paint*; e mais, há a possibilidade de salvar tudo o que foi feito em formato *.rtf (rich text format)* através do uso da opção *Report*, que permite criar relatórios internamente no programa, tornando o STATISTICA® um editor de gráficos bem completo.

Exemplo: Construção do gráfico de colunas para a variável resistência à ferrugem.

(1) No menu [Graphs], escolha a opção [Histograms], clique na aba [Advanced] e aparecerá a janela apresentada na Figura 3.11:

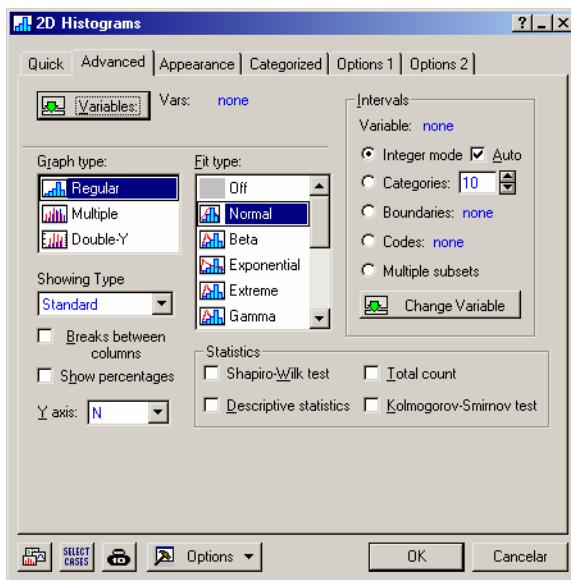


Figura 3.11 – Janela para a especificação do gráfico de colunas

(2) Selecione a variável, clicando no botão [Variables] e depois na variável que será estudada, conforme apresentada na figura a seguir:

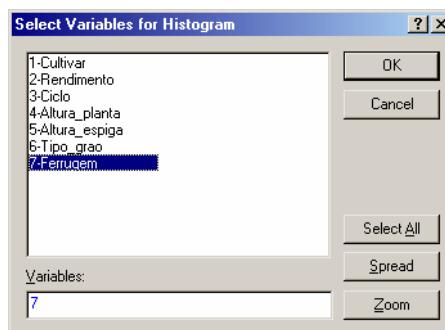


Figura 3.12 – Janela que mostra as variáveis que podem ser selecionadas

(3) Na janela da Figura 3.11, especifique:

- o tipo de gráfico (*Graph Type*) regular
 - o tipo de ajuste (*Fit Type*) off (sem)
 - ativar o *Breaks between Columns* (para que haja espaço entre as colunas)
- O resultado das especificações está apresentado na figura a seguir:

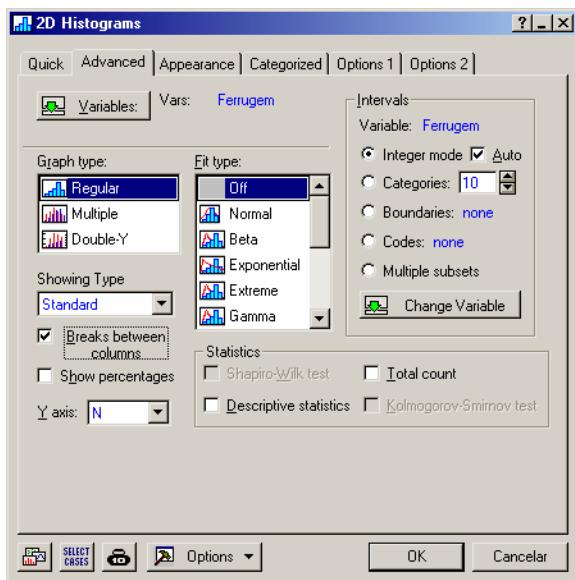


Figura 3.13 – Janela para a especificação do gráfico de colunas

(4) Clicar em OK e o gráfico resultante está apresentado na figura a seguir:

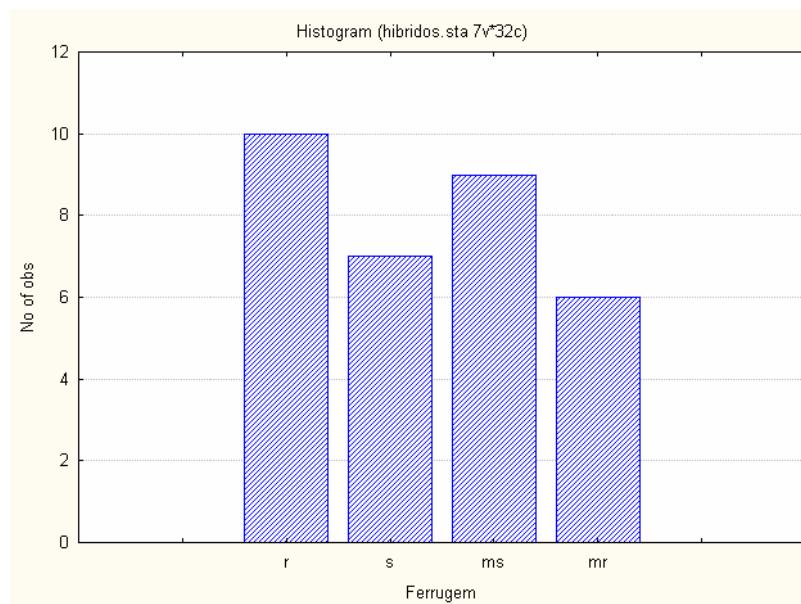


Figura 3.14 – Gráfico de colunas da variável resistência à ferrugem

(5) Para fazer um gráfico de barras, na janela da Figura 3.13, clique na aba [Options2], selecione na opção [X-Y Axis position] o valor *Reverse* e clique no botão [OK]. O gráfico resultante está apresentado na figura a seguir:

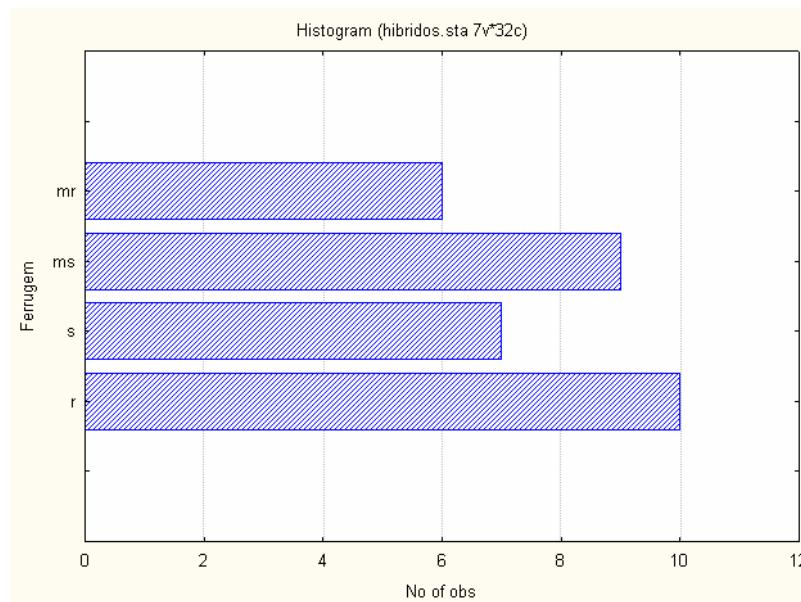


Figura 3.15 – Gráfico de barras da variável resistência à ferrugem

Nota: No capítulo 4 serão apresentadas as opções gráficas para modificação do layout dos gráficos.

Imagina-se que os cultivares de 1 a 11 sejam do município de Içara. Se desejarmos fazer um gráfico para este município, então:

(6) Volte na janela da Figura 3.13, clique no botão [Select Cases] e crie o filtro apresentado na figura a seguir:

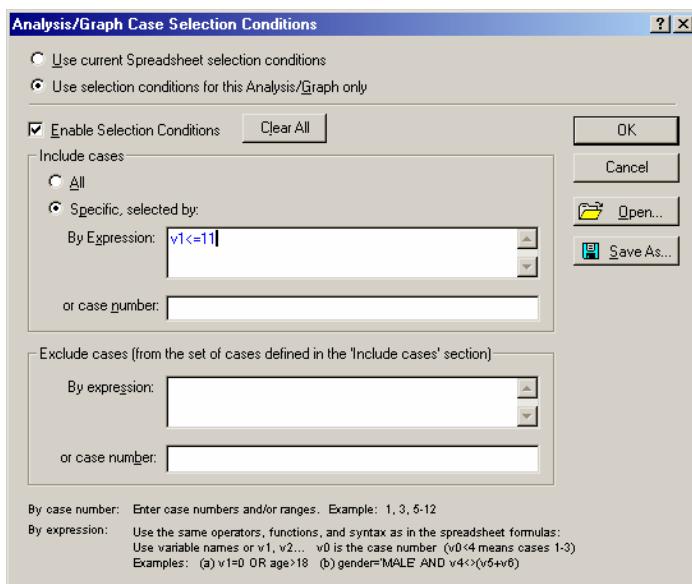


Figura 3.16 – Janela que permite especificar os cultivares

(7) O gráfico resultante está apresentado na figura a seguir:

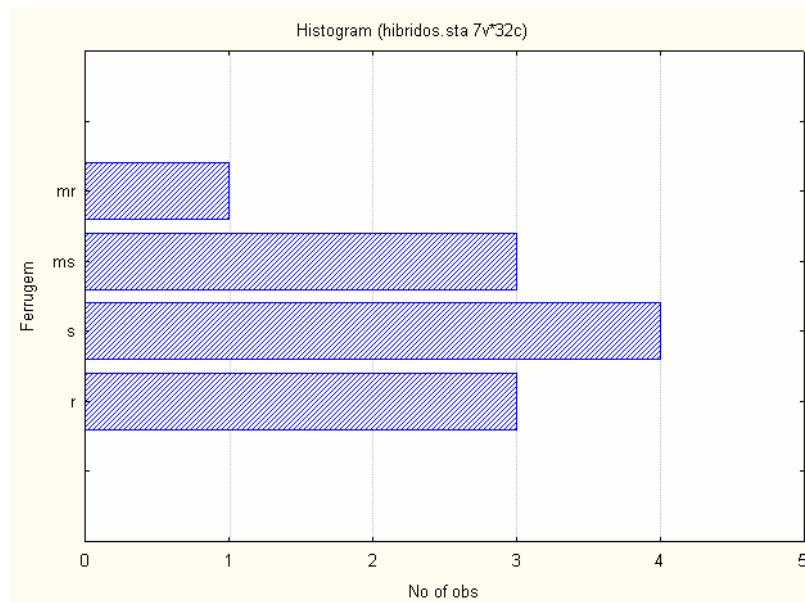


Figura 3.17 – Gráfico de barras da variável resistência à ferrugem para o município de Içara

Observação: Existe mais de uma maneira de gerar um gráfico, como o uso do menu flutuante (*pop-up menu*), para usá-lo, selecione a coluna da variável desejada e clique com botão direito do mouse. Aparecerá uma janela onde deve selecionar a opção *Graphs of Input Data* a seguir escolha o gráfico adequado, conforme apresentado na figura a seguir.

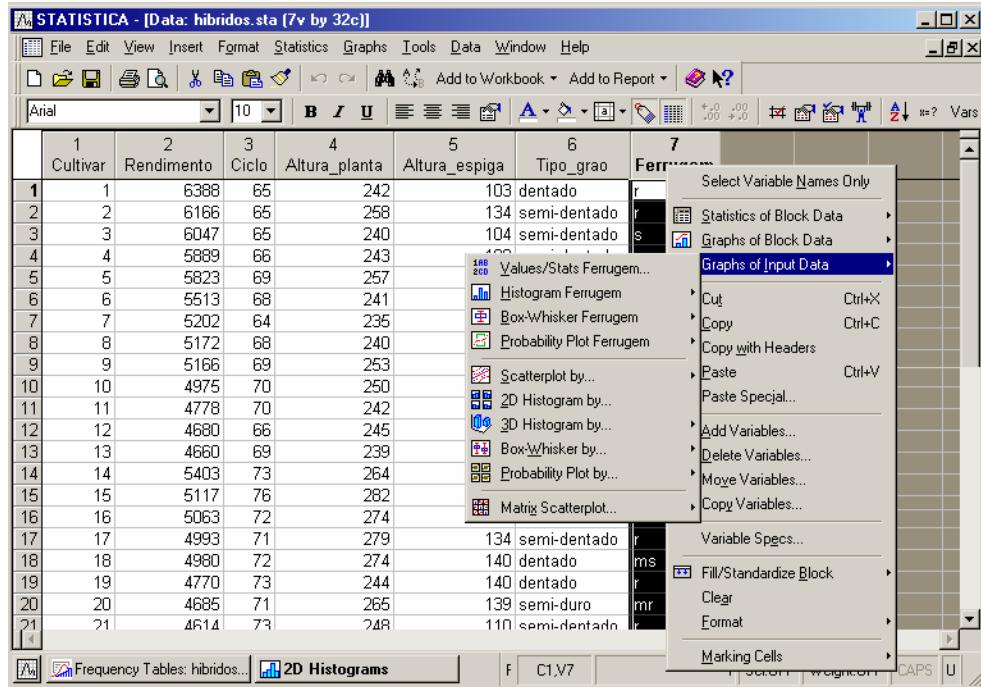


Figura 3.18 – Forma alternativa de gerar um Gráfico

3.1.2. Bidimensional

Com essa análise é possível criar:

- Distribuição da freqüência para **duas variáveis**;
- Tabelas de contingência: independência de variáveis qualitativas, uso do:
 - o Coeficiente de contingência de *Pearson*;
 - o Coeficiente de *Cramér*.

3.1.2.1. Tabela da Distribuição de Freqüência Bidimensional

Exemplo: Análise da distribuição de freqüência das variáveis resistência à ferrugem e tipo de grão.

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 3.19, onde a opção a ser escolhida será *Frequency Table*:

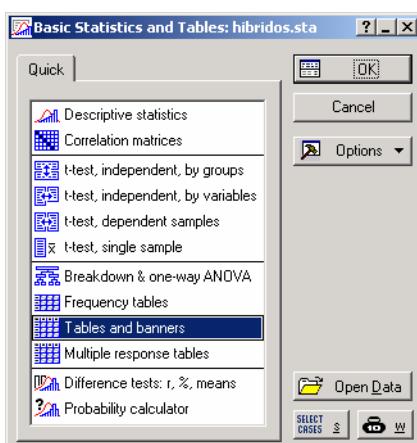


Figura 3.19 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.20:



Figura 3.20 – Janela para especificação de tabela

(3) Clique no botão [Specify Tables (select variables)]. Como a tabela é bidimensional, implica que devemos escolher duas variáveis conforme apresentado na Figura 3.21. As variáveis aparecerão na *List 1* e *List 2*:



Figura 3.21 – Janela com as variáveis selecionadas para a tabela bidimensional

(4) Clique nos botões [OK] da janela anterior e da Figura 3.20 e aparecerá a janela apresentada na Figura 3.22:

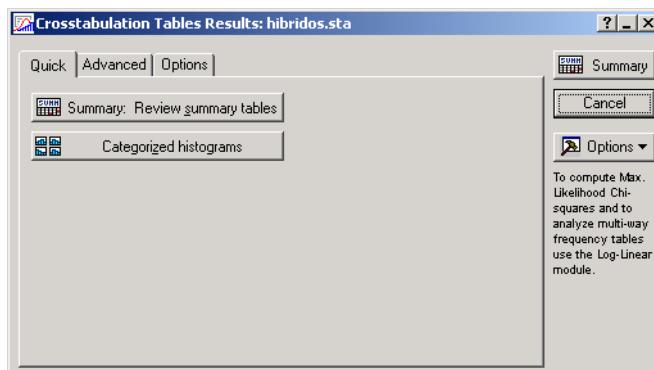


Figura 3.22 – Janela para a visualização do resultado

(5) Clique no botão [Summary] e o resultado está apresentado na Figura 3.23:

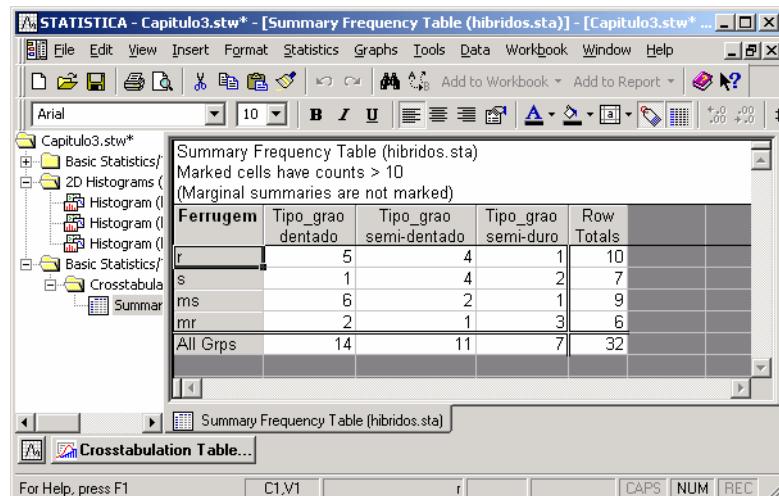


Figura 3.23 – Tabela de distribuição de freqüência conjunta do nº de híbridos, de acordo com a resistência à ferrugem e tipo de grão

A linha *All Grps* mostra as freqüências absolutas, ou seja, o total de híbridos para cada categoria do tipo de grão. Como os totais marginais são diferentes, dificuldades reais de interpretação são introduzidas.

Para uma melhor interpretação da tabela, o ideal é selecionar que a tabela mostre os resultados em termos de proporções, ou seja, podemos fixar o **total de linhas** e/ou o **total de colunas** e/ou o **total geral**.

(6) Na janela da Figura 3.22, clique na aba [Options] e defina as opções conforme apresentado na figura 3.24. Para desativar, basta clicar na opção correspondente, retirando assim, o símbolo . Para ativar, basta clicar na opção ainda não selecionada:

- Selecione *Percentages of total count* para a percentagem do total geral; ou
- selecione *Percentages of row counts* para a percentagem do total das linhas; ou
- selecione *Percentages of column counts* para a percentagem do total das colunas.

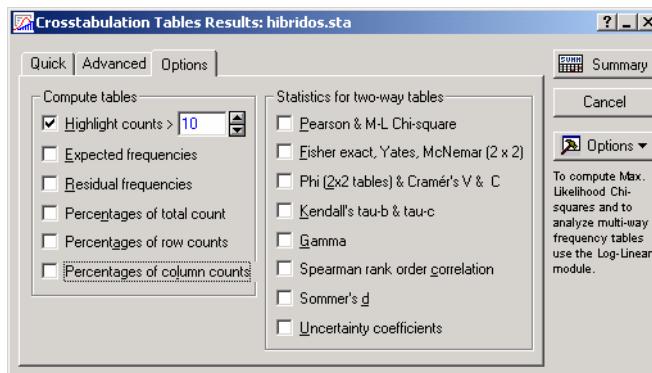


Figura 3.24 – Opções da tabela de freqüência conjunta

- Quando fixar o total de linhas e/ou o total de colunas e/ou o total geral?

A resposta deve ser de acordo com o objetivo da sua pesquisa. Fixe a variável que você quer dar mais ênfase. No exemplo, se for a ferrugem, fixe somente o total de linhas. Se for o tipo de grão, fixe somente o total de colunas.

(7) Escolha o tipo de percentagem e clique no botão [Summary]. Os resultados estão apresentados nas três tabelas a seguir:

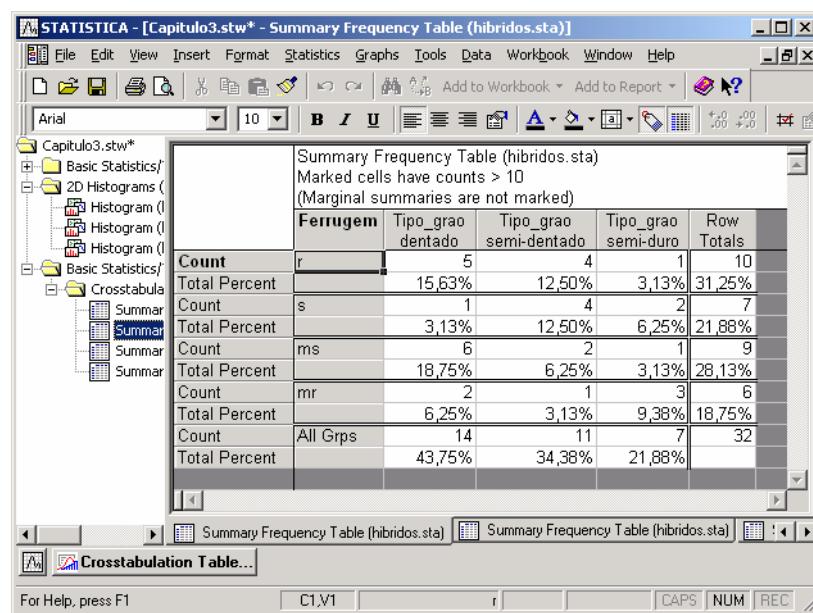


Figura 3.25 – Selecionando somente o **total geral**

Conclusão: observando a figura anterior, pode-se concluir que do total de híbridos, 5 ou 15,63% são resistentes e dentados.

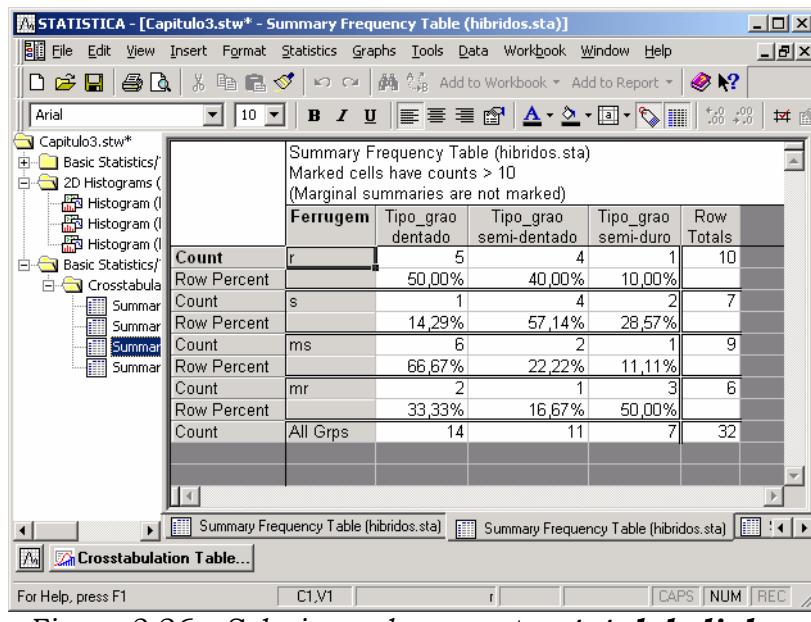


Figura 3.26 – Selecionando somente o **total de linhas**

Conclusão: observando a figura anterior, pode-se concluir que 4 ou 40% dos resistentes são semi-dentados pois a soma na linha é igual a 100%.

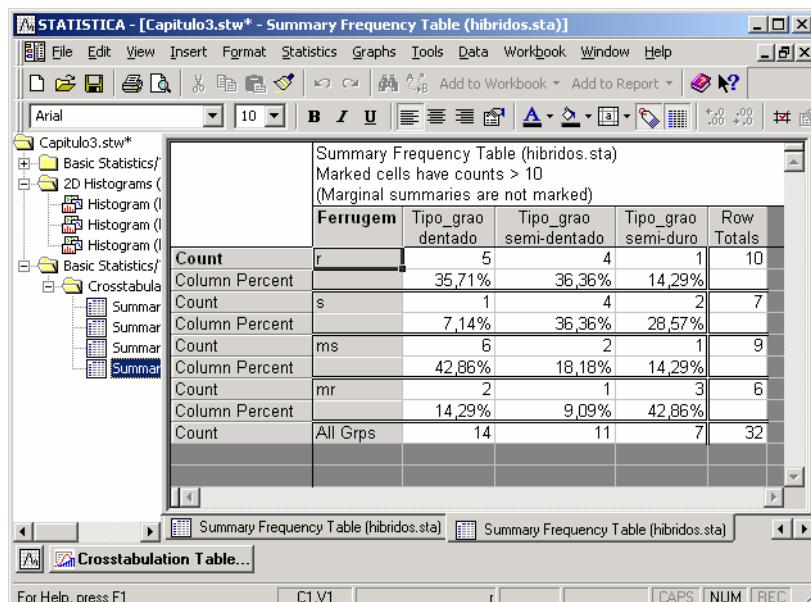


Figura 3.27 – Selecionando somente o **total de colunas**

Conclusão: observando a figura anterior, pode-se concluir que dos híbridos semi-dentados, 4 ou 36,36% são resistentes, ou ainda, semi-dentados e dentados tem maior porcentagem de resistência (36,36% e 35,71%, respectivamente contra 14,29% dos semi-duros).

As quatro figuras anteriores mostram o mesmo resultado, porém, repare que fixar os totais percentuais não modifica o resultado e sim, enfatiza que resistentes, são os semi-dentados e os dentados.

3.1.2.2. Gráfico de Barras Múltiplas

Para efetuar uma análise comparativa de várias distribuições, podemos construir um gráfico de barras múltiplo.

(1) No menu [Graphs], escolha a opção [Categorized Graphs], em seguida [Histograms] e aparecerá a janela apresentada na Figura 3.28:

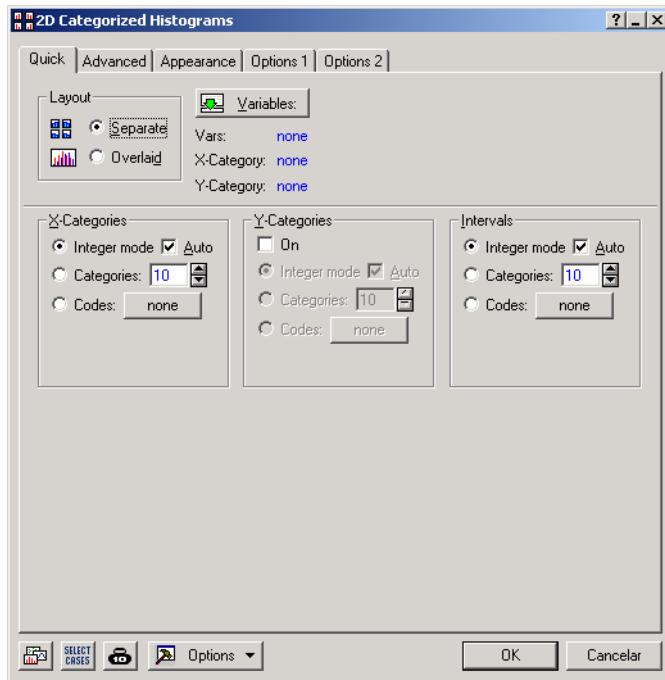


Figura 3.28 – Janela para a especificação do gráfico de barras múltiplas

(2) Selecione as variáveis, clicando no botão [Variables] e depois escolha a variável a ser estudada e a que será a categoria, conforme apresentada na figura a seguir:

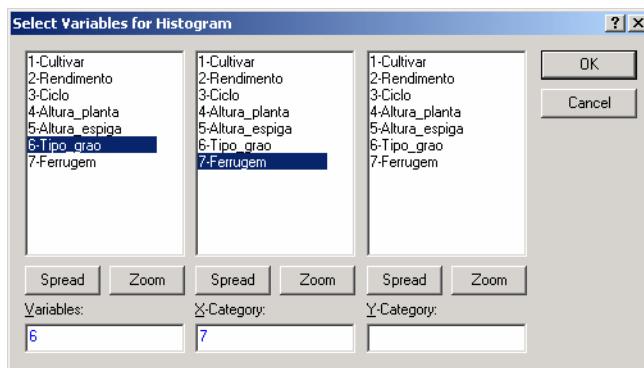


Figura 3.29 – Janela para seleção das variáveis

(3) Na janela da Figura 3.28, especifique:

- o tipo de gráfico (Layout:) *Overlaid*

O resultado das especificações está apresentado na figura a seguir:

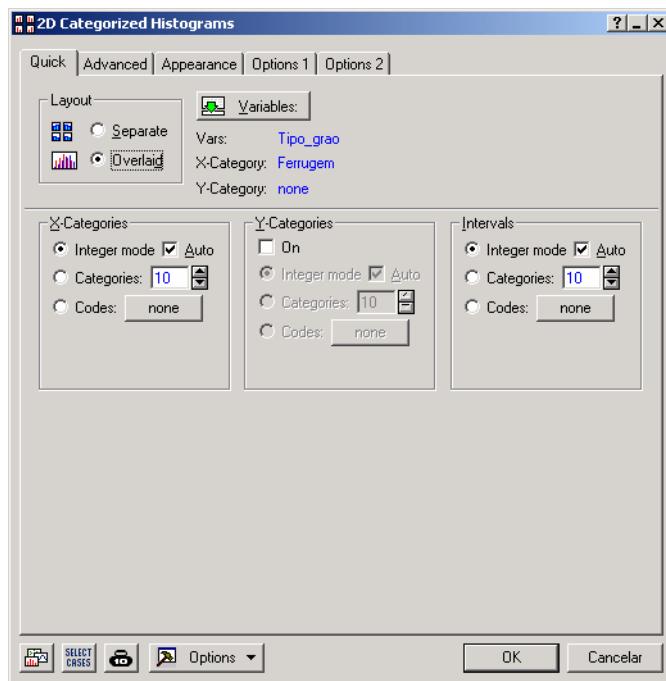


Figura 3.30 – Janela para a especificação do gráfico de barras multiplas

(5) Para fazer um gráfico de barras, na janela da figura anterior, clique na aba [Options2], selecione na opção [X-Y Axis position] o valor *Reverse* e clique no botão [OK]. O gráfico resultante está apresentado na figura a seguir:

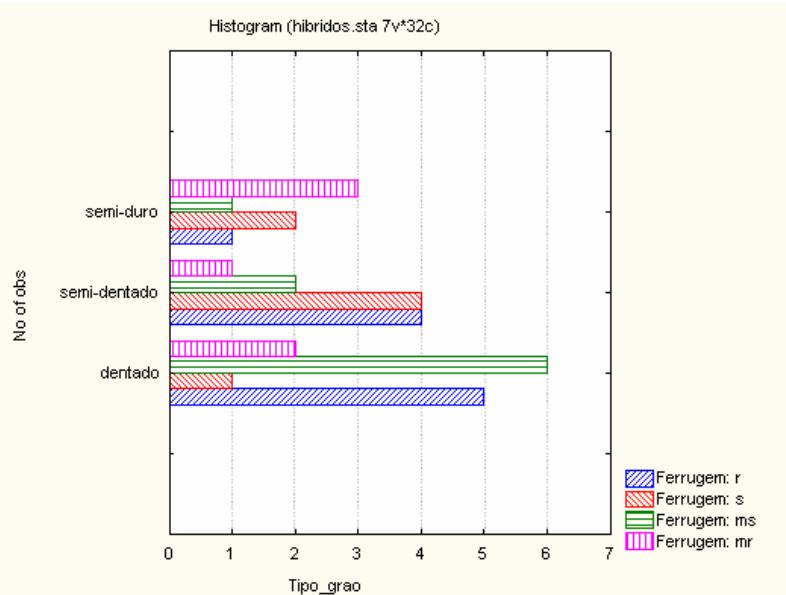


Figura 3.31 – Gráfico de barras multiplas, ferrugem (e seus quatro níveis: resistente, susceptível, moderadamente susceptível e moderadamente resistente) versus tipo de grão (e seus três níveis: semi-duro, semi-dentado e dentado)

3.1.2.3. Tabelas de Contingência: independência de variáveis

As tabelas construídas até aqui, são denominadas tecnicamente de tabelas de contingência e são úteis para estudar a dependência (ou independência) entre

variáveis. Vamos construir novamente a tabela bidimensional com as variáveis ferrugem e tipo de grão, onde os totais de linhas foram fixados em 100%.

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 3.32, onde a opção a ser escolhida será *Frequency Table*:

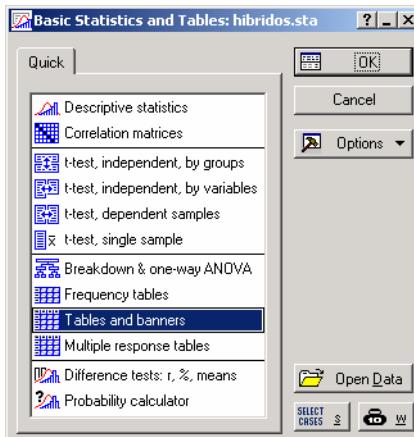


Figura 3.32 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.33:



Figura 3.33 – Janela para especificação de tabela

(3) Clique no botão [Specify Tables (select variables)]. Como a tabela é bidimensional, implica que devemos escolher duas variáveis conforme apresentado na Figura 3.34. As variáveis aparecerão na *List 1*: e *List 2*:

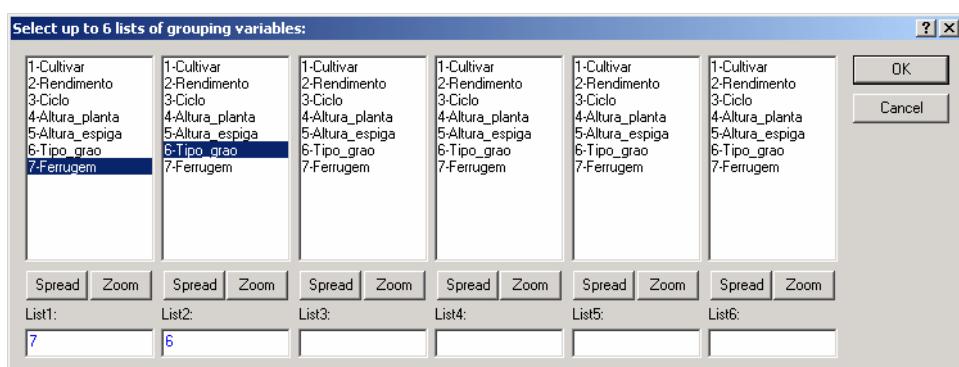


Figura 3.34 - Janela com as variáveis selecionadas para a tabela bidimensional

(4) Clique nos botões [OK] da janela anterior e da Figura 3.33 e aparecerá a janela apresentada na Figura 3.35:



Figura 3.35 – Janela para a visualização do resultado

(5) Na janela da figura anterior, clique na aba [Options] e defina a opção percentagem do total das linhas (*Percentages of row counts*) conforme apresentado na Figura 3.36.

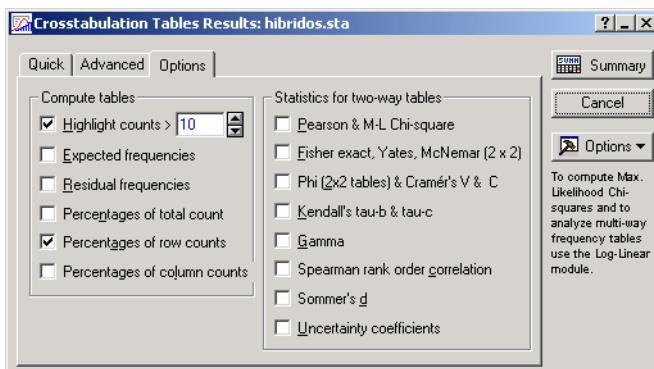


Figura 3.36 – Opções da tabela de freqüência conjunta

(6) Na janela da figura anterior, clique na aba [Advanced] e selecione a opção (*Display selected %'s in sep. tables*) conforme apresentado na Figura 3.37. Essa opção separa as freqüências dos valores percentuais em tabelas distintas.



Figura 3.37 – Opções da tabela de freqüência conjunta

(7) Clique no botão [Summary] e os resultados estão apresentado nas figuras 3.38 e 3.39:

Ferrugem	Tipo_grao dentado	Tipo_grao semi-dentado	Tipo_grao semi-duro	Row Totals
r	5	4	1	10
s	1	4	2	7
ms	6	2	1	9
mr	2	1	3	6
All Grps	14	11	7	32

Figura 3.38 – Tabela de distribuição de freqüência conjunta do nº de híbridos, de acordo com a resistência à ferrugem e tipo de grão – freqüências

Ferrugem	Tipo_grao dentado	Tipo_grao semi-dentado	Tipo_grao semi-duro	Row Totals
r	50,00%	40,00%	10,00%	31,25%
s	14,29%	57,14%	28,57%	21,88%
ms	66,67%	22,22%	11,11%	28,13%
mr	33,33%	16,67%	50,00%	18,75%
All Grps	43,75%	34,38%	21,88%	

Figura 3.39 – Tabela de distribuição de freqüência conjunta do nº de híbridos, de acordo com a resistência à ferrugem e tipo de grão - percentuais

Interpretação: se os semi-dentados fossem totalmente independentes da variável resistência a ferrugem, esperaríamos 34,38% (All Grps) para as categoria r, s, ms e mr. Na tabela observamos que a categoria s apresenta 57,14%, valor este acima dos 34,38%. Da mesma forma, observamos também que a categoria mr, com 16,67%, está abaixo dos 34,38%. Isto implica um distanciamento da independência entre as variáveis. Podemos deduzir que elas estão associadas.

A determinação dos coeficientes de continência de Pearson e Cramér, vão quantificar (forte, moderada ou fraca) a dependência.

(7) Na janela da Figura 3.36 selecione a opção (*Phi (2x2 tables) & Cramér's V & C*) conforme apresentado na Figura 3.40.

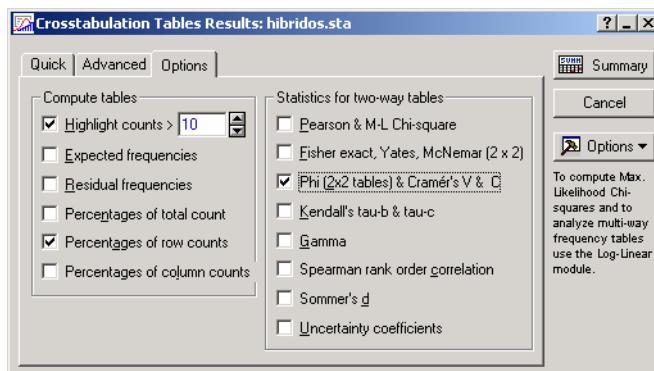


Figura 3.40 – Opções da tabela de freqüência conjunta

(8) Na janela da figura anterior, clique na aba [Advanced] e no botão [Detailed two-way tables] e o resultado está apresentado na Figura 3.41:

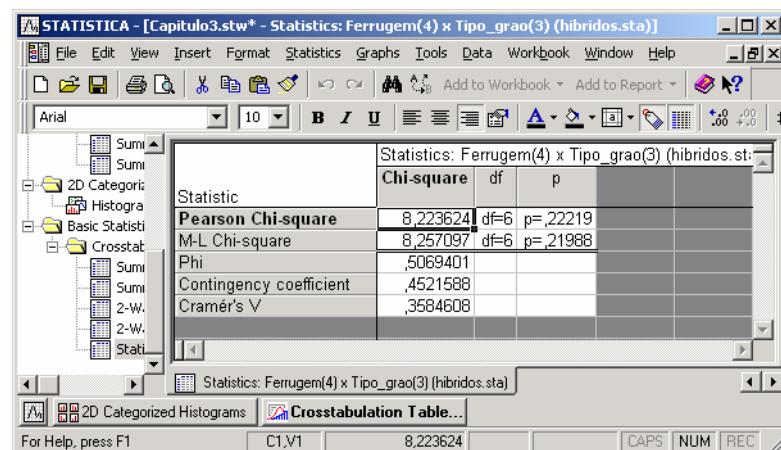


Figura 3.41 – Estatísticas para verificar a associação entre variáveis

Observação: O STATISTICA® não corrige o coeficiente de contingência (*Contingency coefficient*) = 0,4521588. A seguir está a fórmula para correção:

$$C^* = \frac{C}{\sqrt{(t-1)/t}} = \frac{0,4521588}{\sqrt{(3-1)/3}} = 0,5538$$

Como o coeficiente de contingência está entre 0 e 1 ($0 \leq C^* \leq 1$), temos uma associação moderada (média). Por outro lado, o coeficiente de Cramér (Cramér's V) - $0 \leq V \leq 1$, indica uma associação fraca.

3.1.2.4. Gráfico de Setores (Pie Charts)

(1) No menu [Graphs], escolha a opção [2D Graphs], em seguida [Pie Charts] e aparecerá a janela apresentada na Figura 3.42:

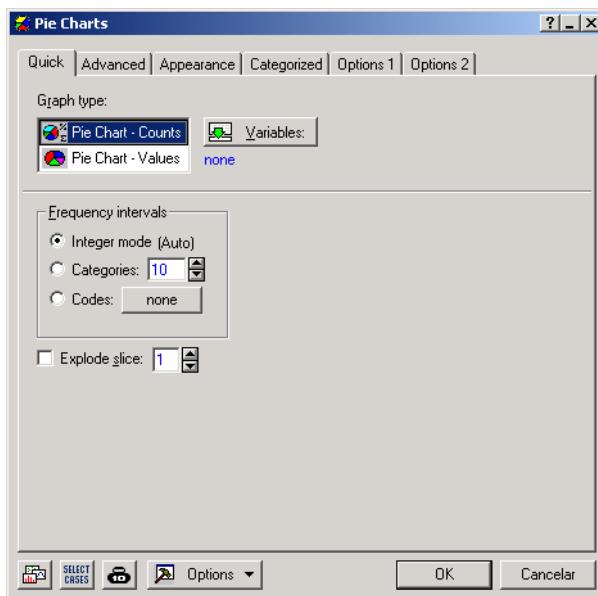


Figura 3.42 – Janela para especificação do gráfico de setores

(2) Selecione a variável, clicando no botão [Variables] e depois escolha a variável a ser estudada, conforme apresentada na figura a seguir:

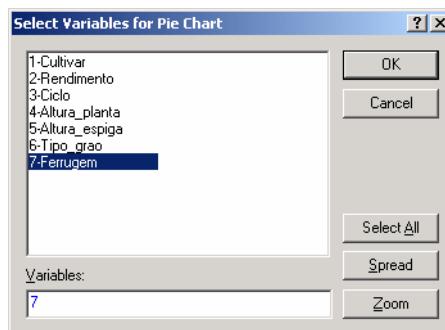


Figura 3.43 – Janela para seleção das variáveis

(3) Na janela da Figura 3.42, clique na aba [Advanced] e especifique:

- o tipo de legenda (*Pie legend*): *Text and Percent* (mostra o texto (níveis) e as porcentagens)
- o tipo de visualização (*Type*): *3D* (tridimensional)
- a forma (*Shape*): *Ellipse* (elipse)

O resultado das especificações está apresentado na figura a seguir:

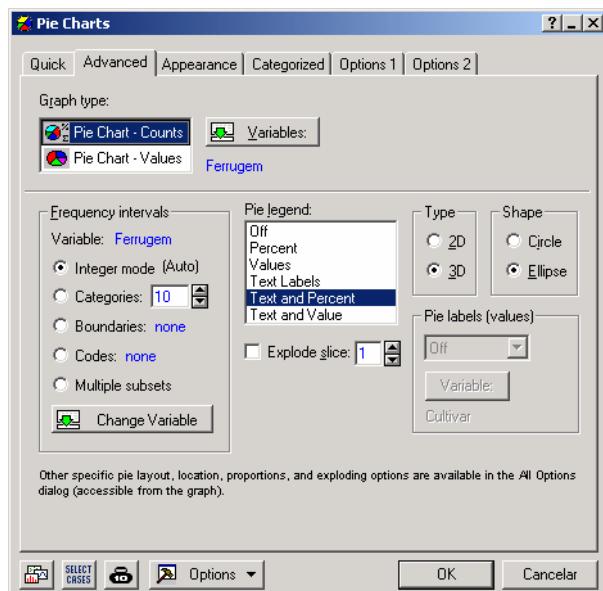


Figura 3.44 – Janela para a especificação do gráfico de setores

(4) Clique no botão [OK] e o gráfico resultante está apresentado na figura a seguir:

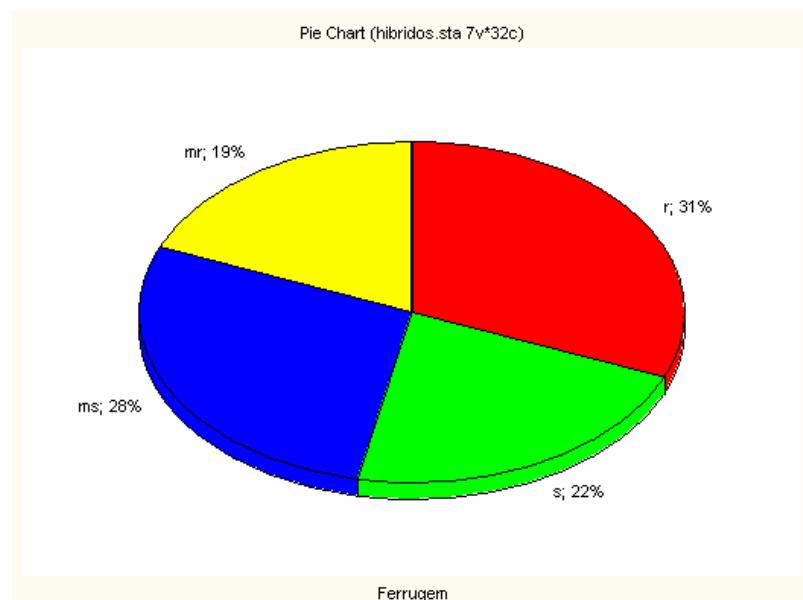


Figura 3.45 – Gráfico de setores para a ferrugem

(5) Para que os setores fiquem separados, clique com o botão direito do mouse em cima do gráfico e escolha a opção (*Graph Properties (All Options)*). Na janela que abrirá, clique na aba [Plot: Pie], ative o *Exploded* dos quatro setores, conforme apresentado na Figura 3.46:

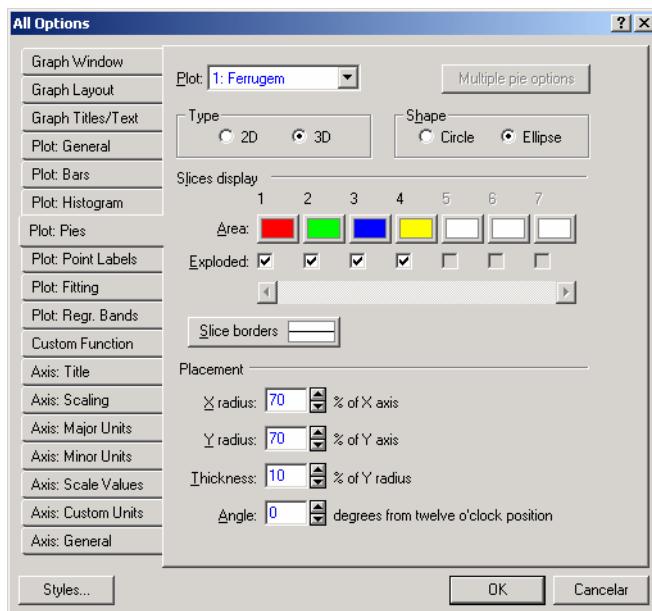


Figura 3.46 – Opções gráficas do Pie Charts

(6) Clique no botão [OK] e o gráfico resultante está apresentado na figura a seguir:

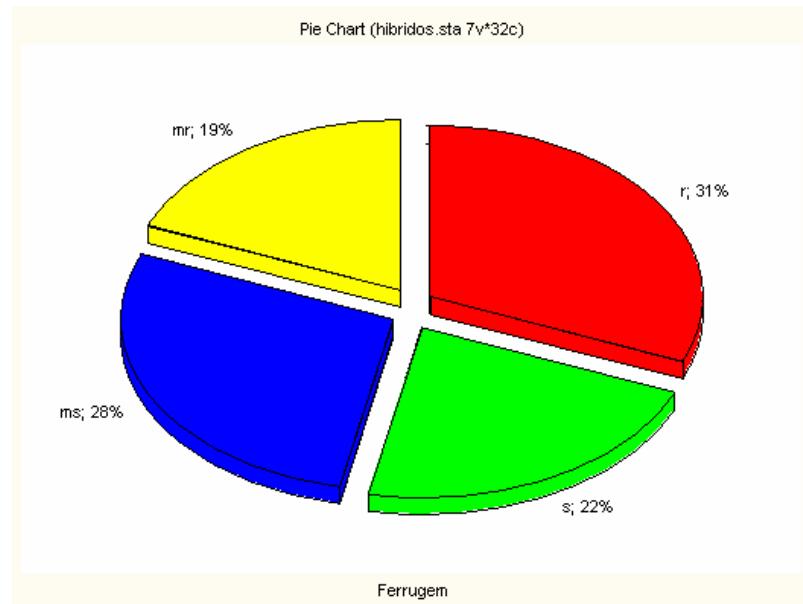


Figura 3.47 – Gráfico de setores para a ferrugem

3.2. Distribuição de Freqüência de Variáveis Quantitativas

Variável quantitativa é aquela que apresenta como resultado, medidas e contagens. Podem ser **contínuas** e **discretas**. As contínuas, são todas as que, pelo menos em teoria, possam assumir qualquer valor dentro de um intervalo; enquanto as discretas, normalmente resultam de contagens, são números inteiros e só assumem certos valores.

Neste caso podemos estabelecer duas situações:

1^{a)} Variável contínua - classes de ocorrência

Observação: quando temos poucas observações, até aproximadamente 25, a distribuição pode ser representada por meio de um **diagrama de pontos**. O programa **STATISTICA®** não faz este diagrama.

2^{a)} **Variável discreta** - distribuição de freqüência feita de maneira idêntica às variáveis qualitativas (categorizadas). Também pode ser representada por um diagrama de pontos.

3.2.1. Tabela de Distribuição de Freqüência para Variável Quantitativa Discreta

Se os resultados da variável discreta não são muitos, podemos dar o mesmo enfoque dado as variáveis qualitativas.

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 3.50, onde a opção a ser escolhida será *Frequency Table*:

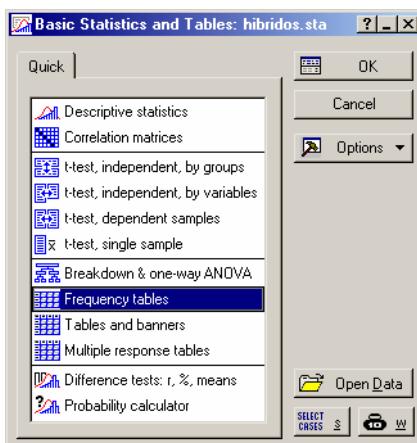


Figura 3.50 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.51:

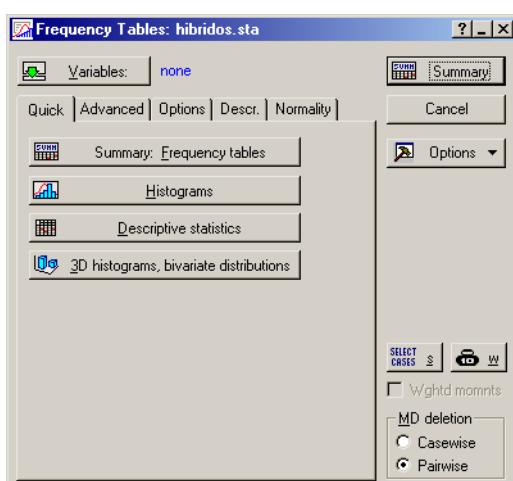


Figura 3.51 – Tabelas de freqüência

(3) Clique no botão [Variables] e selecione a variável Ciclo, conforme apresentado na Figura 3.52:

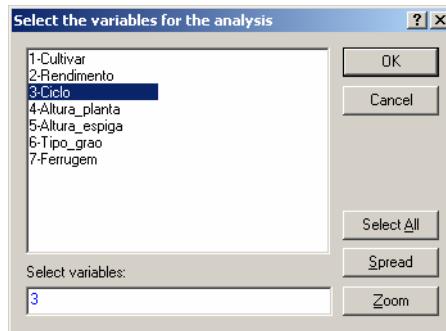


Figura 3.52. Janela para seleção das variáveis para a análise

(4) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.53:

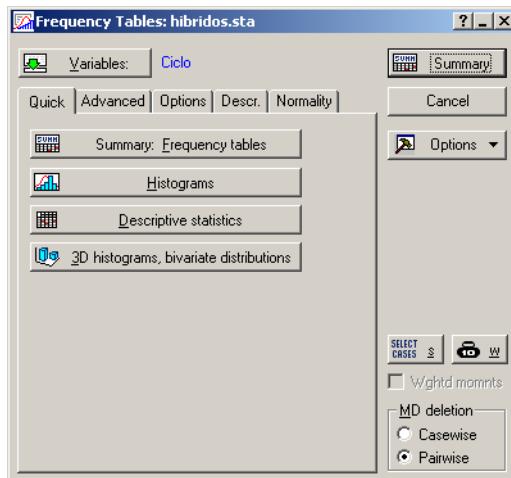


Figura 3.53 – Tabela de freqüência com a variável Ciclo selecionada

(5) Clique na aba [Options] e defina as opções conforme apresentado na figura 3.54:

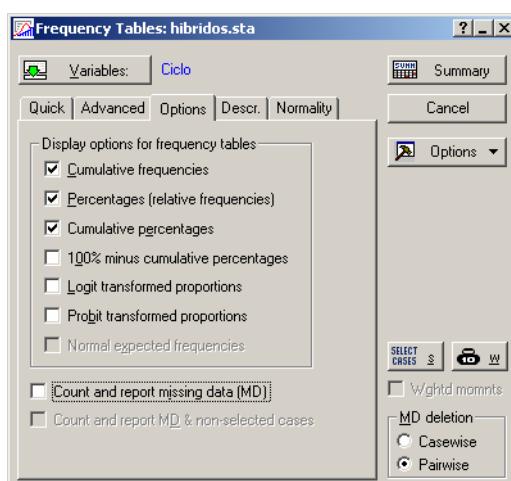


Figura 3.54 – Opções da tabela de freqüência

(6) Clique no botão [Summary] e o resultado está apresentado na Figura 3.55:

Category	Count	Cumulative Count	Percent	Cumulative Percent
64	2	2	6.25000	6.2500
65	3	5	9.37500	15.6250
66	3	8	9.37500	25.0000
68	3	11	9.37500	34.3750
69	3	14	9.37500	43.7500
70	4	18	12.50000	56.2500
71	4	22	12.50000	68.7500
72	2	24	6.25000	75.0000
73	5	29	15.62500	90.6250
74	2	31	6.25000	96.8750
76	1	32	3.12500	100.0000

Figura 3.55 – Tabela de Distribuição de Freqüência da variável Ciclo

3.2.2. Representações Gráficas de Variáveis Quantitativas Discretas

Os gráficos de barras justapostas, conhecidos como histogramas, são indicados no caso de variáveis contínuas. Podemos então, construir um GRÁFICO DE ORDENADAS para o caso de uma variável discreta, como por exemplo, a variável ciclo da cultura.

O procedimento no STATISTICA® é o mesmo que construir um gráfico de barras já descrito anteriormente no item 3.1.1.3. Entretanto, o gráfico de ordenadas será construído para a variável Ciclo, portanto, siga os passos do exemplo das figuras 3.11 a 3.14, não esquecendo de alterar no item (2) para a variável Ciclo.

(1) O gráfico resultante está apresentado na Figura 3.56:

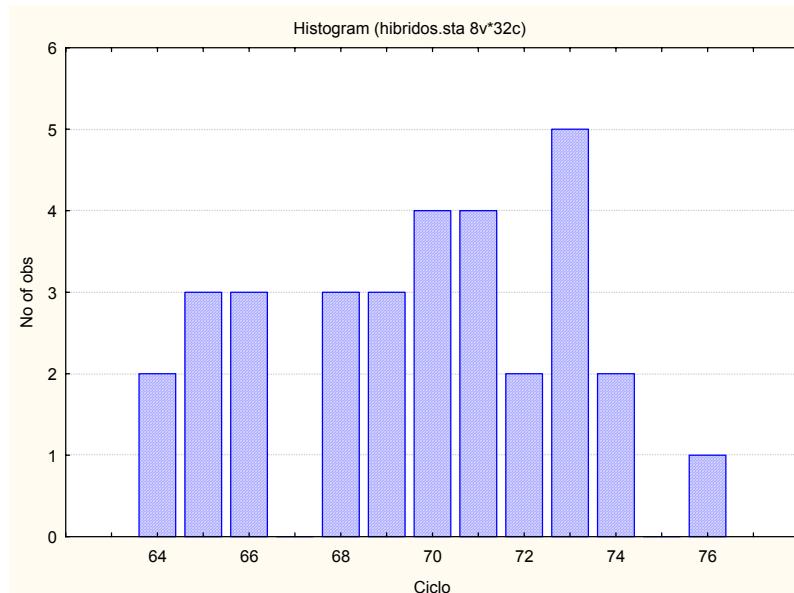


Figura 3.56 – Gráfico de ordenadas da variável Ciclo

(2) Clique com o botão direito do *mouse* em cima do gráfico e escolha a opção (*Graph Properties (All Options)*). Na janela que abrirá, clique na aba [Plot: Bars] e modifique a opção *Type* para *Lines*, conforme apresentado na Figura 3.57:

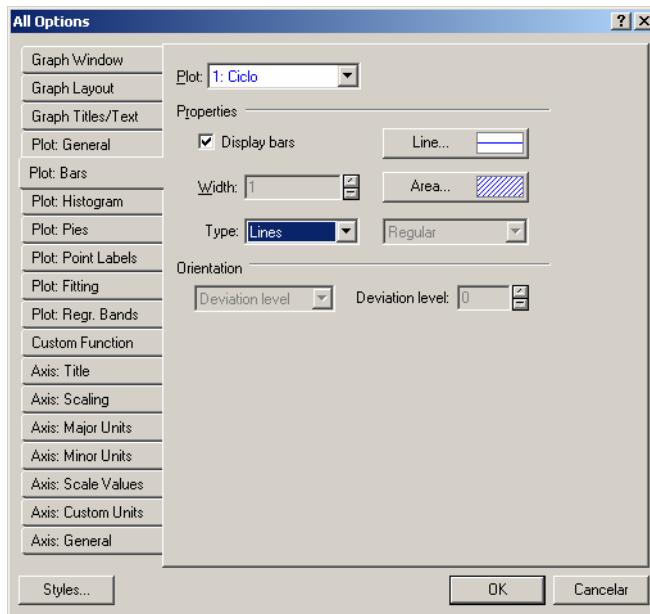


Figura 3.57 – Opções gráficas do 2D Histograms

(3) Clique no botão [OK] e o gráfico resultante está apresentado na figura a seguir:

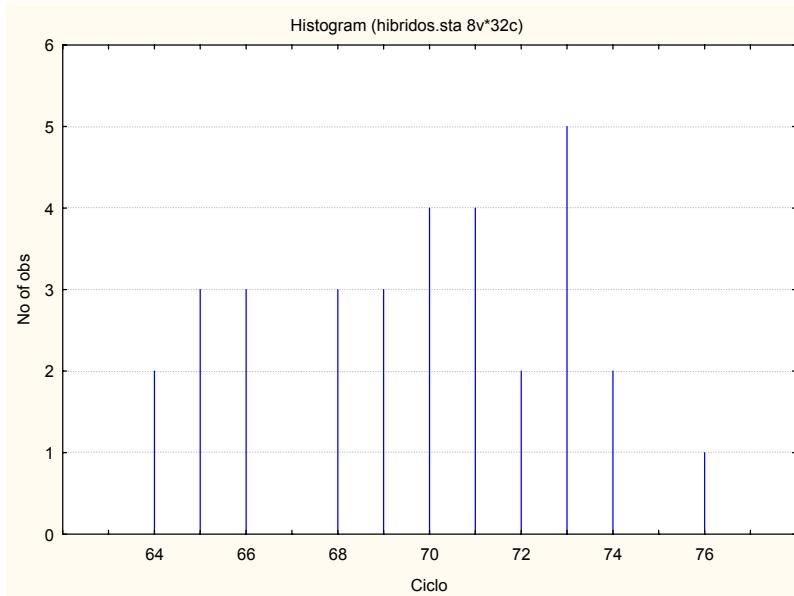


Figura 3.58 – Gráfico de ordenadas da variável ciclo

Quando temos muitas observações de uma variável em estudo (acima de 25), é recomendável a formação de intervalos de valores, isto é, construir classes de ocorrências.

O STATISTICA® pode agrupar os dados em quantas classes quisermos, e aí, construir tabelas e gráficos em função dos intervalos. Usar este artifício é uma maneira de encontrar uma melhor representação do fenômeno em estudo.

Os intervalos ou são por ele agrupados (dados agrupados em classes), divididos de maneira igual conforme um nº previamente estipulado (*nº of exact intervals*), ou, o pesquisador define estas classes (amplitude de casas desiguais).

3.2.3. Dados Agrupados em Classes

Quantos intervalos escolher?

O número de classes deve ser o suficiente para nos dar uma boa idéia do fenômeno, devendo por em evidência a regularidade do fenômeno.

O compromisso é de cada pesquisador. Construindo sucessivamente diferentes histogramas, alterando o número de intervalos, é possível identificar a melhor opção, e assim, deve escolher o que melhor representa a distribuição do fenômeno por ele estudado.

Recomenda-se que o número não seja muito grande (maior que 15) e nem muito pequeno (menor que 5) e também, sempre que for possível, que as classes tenham a mesma amplitude.

Procedimento:

- (1) Repetir os passos (1) a (5) do item 3.2.1 (Tabela de Distribuição de Freqüência para Variável Quantitativa Discreta), porém altere a variável para rendimento.
- (2) Clique na aba [Advanced] e defina a opção *No. of exact intervals* igual a 6, conforme apresentado na figura 3.59:

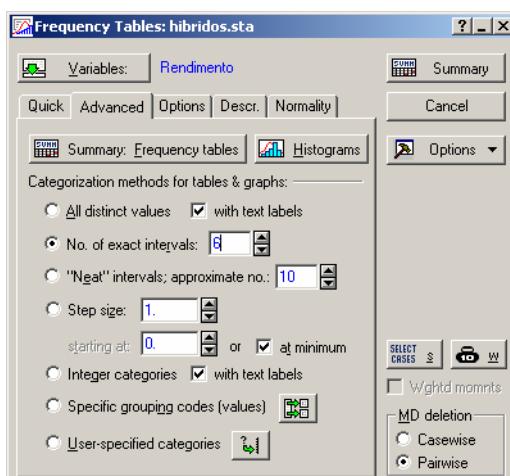
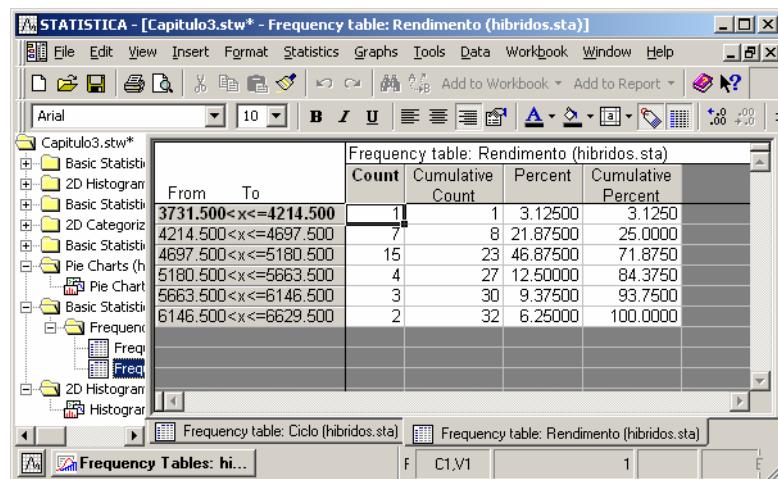


Figura 3.59 – Opções da tabela de freqüência

Iniciaremos o exemplo com um número de classes igual a 6. Posteriormente reduziremos para 5 e finalizaremos com 15 classes. O propósito é baseado em uma verificação que vai possibilitar que visualizemos a realidade do fenômeno aonde certos grupos irão então aparecer destacados.

- (3) Clique no botão [Summary] e o resultado está apresentado na Figura 3.60:



From	To	Count	Cumulative Count	Percent	Cumulative Percent
3731.500 < x <= 4214.500		11	11	3.12500	3.1250
4214.500 < x <= 4697.500		7	18	21.87500	25.0000
4697.500 < x <= 5180.500		15	33	46.87500	71.8750
5180.500 < x <= 5663.500		4	37	12.50000	84.3750
5663.500 < x <= 6146.500		3	40	9.37500	93.7500
6146.500 < x <= 6629.500		2	42	6.25000	100.0000

Figura 3.60 – Tabela de Distribuição de Freqüência com 6 classes da variável rendimento de grãos

Observação: Neste caso, de dados agrupados em classes, repare que os intervalos de classes são de tamanhos iguais.

(4) Para a construção do histograma, clique no botão [Histograms] da Figura 3.59. O resultado está apresentado a seguir:

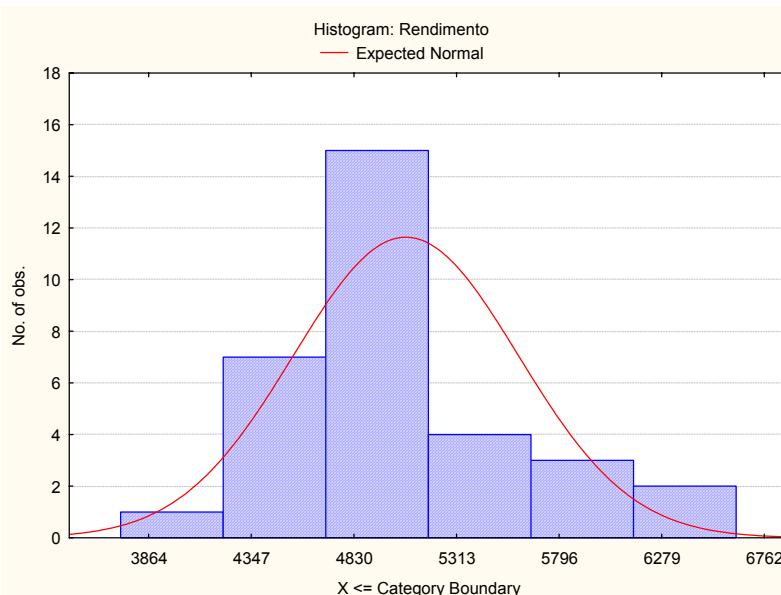


Figura 3.61 – Distribuição de freqüência em 6 classes da variável Rendimento

(5) A seguir, apresenta-se o histograma com o nº de classes igual a 5.

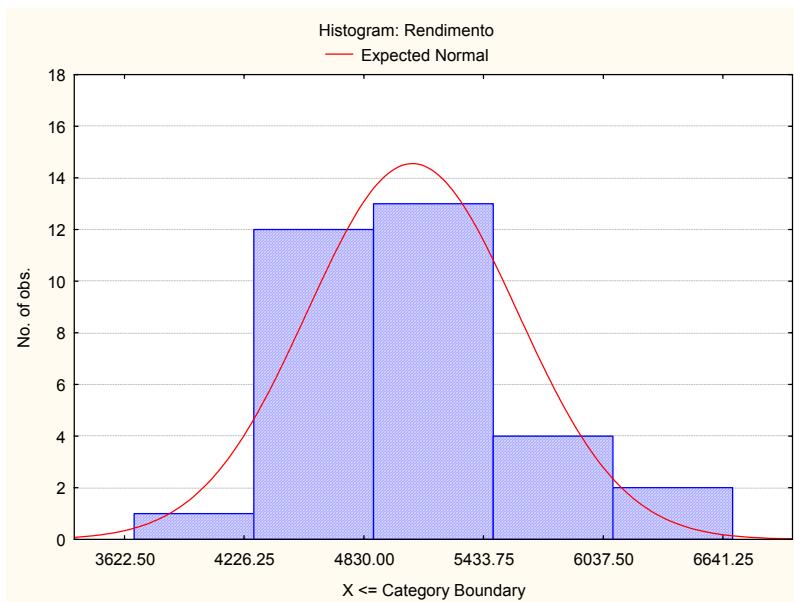


Figura 3.62 – Distribuição de freqüência em 5 classes da variável Rendimento

(6) No último gráfico, a divisão em 15 classes permite que fiquem destacados os grupos de híbridos que possuem maior rendimento e o híbrido com baixo rendimento.

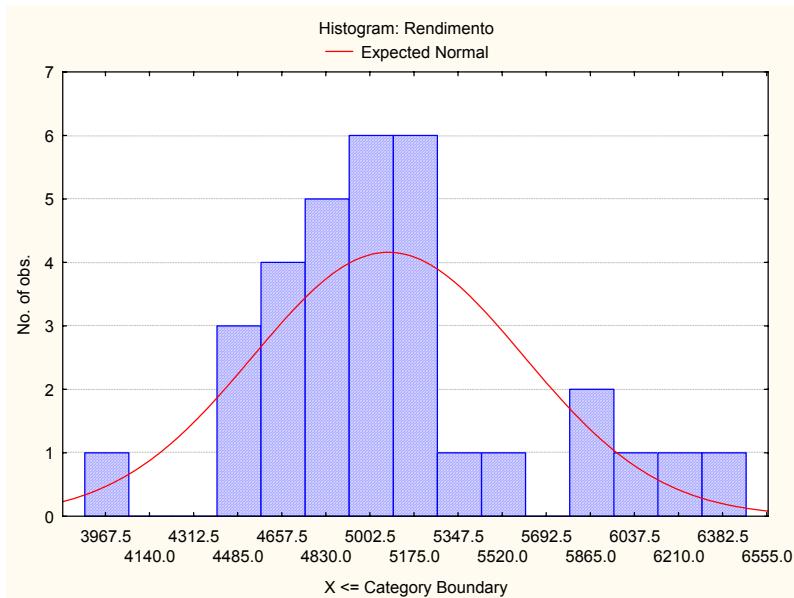


Figura 3.63 – Distribuição de freqüência em 15 classes da variável Rendimento

Note ainda que as linhas nos gráficos representam o modelo normal (*Karl Gauss*) em forma de sino. Este aspecto teórico representa a maioria dos fenômenos naturais e será discutido mais adiante.

Amplitude de classes desiguais

Neste caso o STATISTICA® **não constrói gráficos** (histogramas). Ele pode apenas construir uma tabela de freqüência com as classes definidas. Dando continuidade ao procedimento:

(7) Na janela da Figura 3.59, na aba [Advanced], defina a opção *User-specified categories* e digite os intervalos das classes de cada categoria conforme apresentado na figura a seguir:

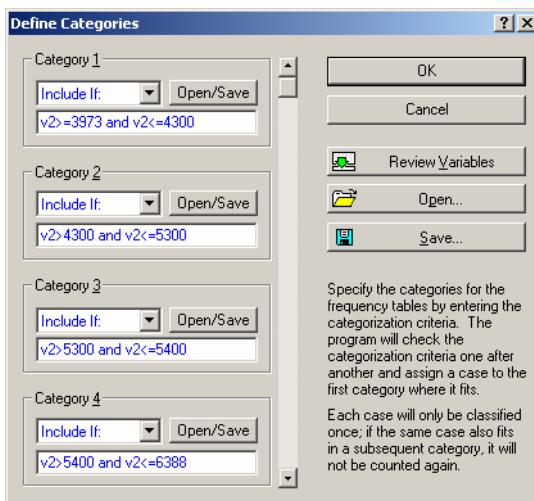


Figura 3.64 – Janela onde são definidas as categorias

(8) Clique no botão [Summary] e o resultado está apresentado na Figura 3.65:

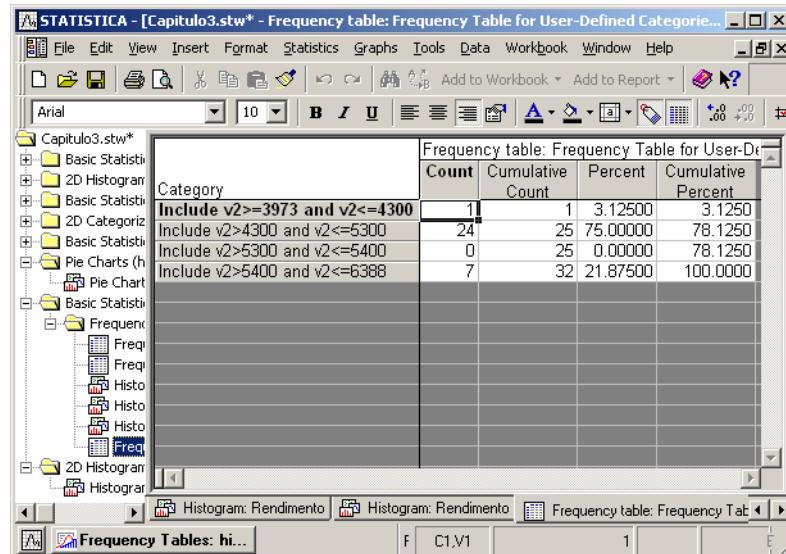


Figura 3.65 – Tabela de distribuição de freqüência usando categorias definidas pelo usuário

Observações:

- A variável deve ser indicada por seu número, ou por sua sigla. No exemplo, vamos trabalhar com a variável Rendimento que pode ser identificada como v2 ou como RENDIMENTO.

3.2.4. Polígono de Freqüências Múltiplos

Com esse tipo de gráfico pode-se verificar múltiplas distribuições de freqüência num mesmo gráfico. Segue a seqüência de passos para a construção do mesmo.

(1) No menu [INSERT], escolha a opção [Add Variables], conforme apresentado na Figura 3.66:

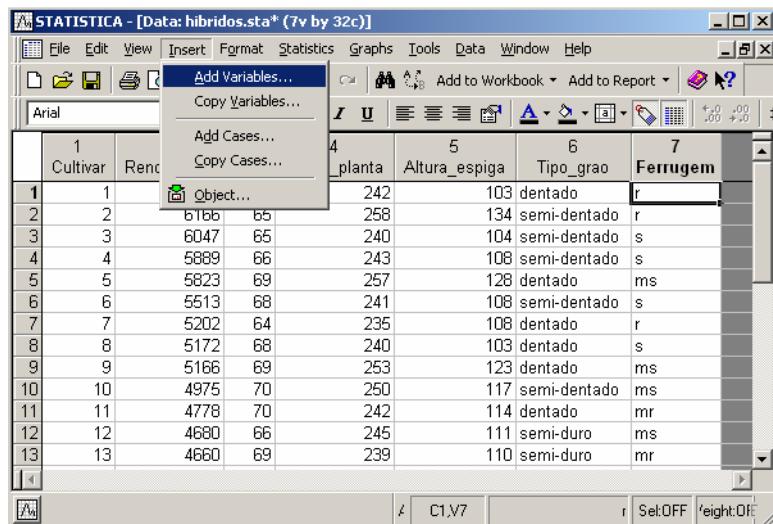


Figura 3.66 – Criação de novas variáveis

(2) Crie uma variável após a variável Ferrugem, conforme definido na Figura 3.67:

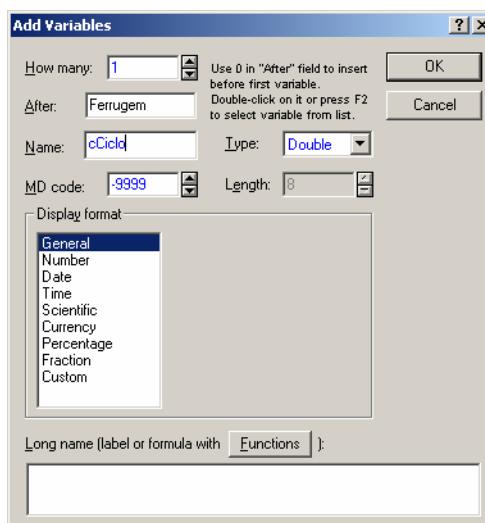


Figura 3.67 – Janela para a criação de novas variáveis

(3) A nova variável cCiclo corresponderá a categorização da variável Ciclo, para isso basta posicionar o cursor na nova variável e no menu [DATA], escolher a opção [Recode], conforme apresentado na Figura 3.68:

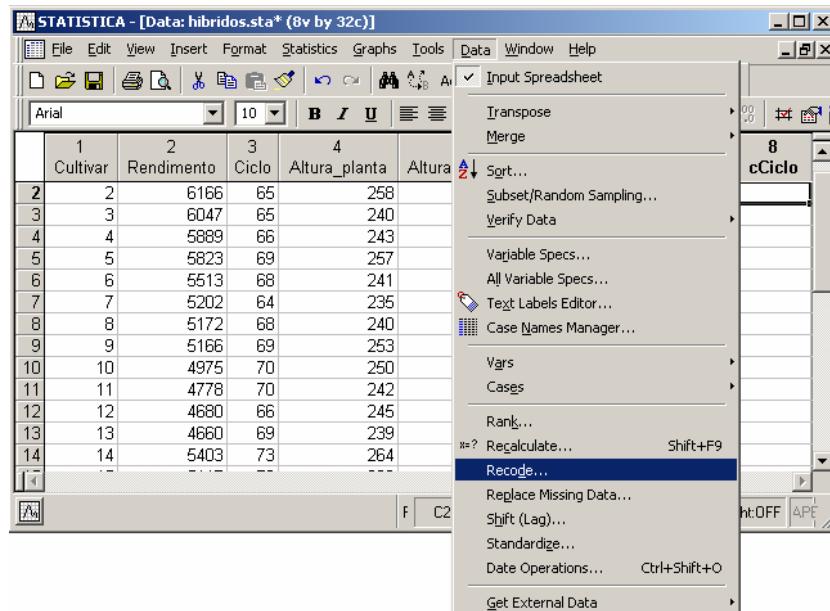


Figura 3.68 – Acesso à janela para a categorização da variável Ciclo

(4) A categorização da variável Ciclo será feita conforme apresentado na Figura 3.69:

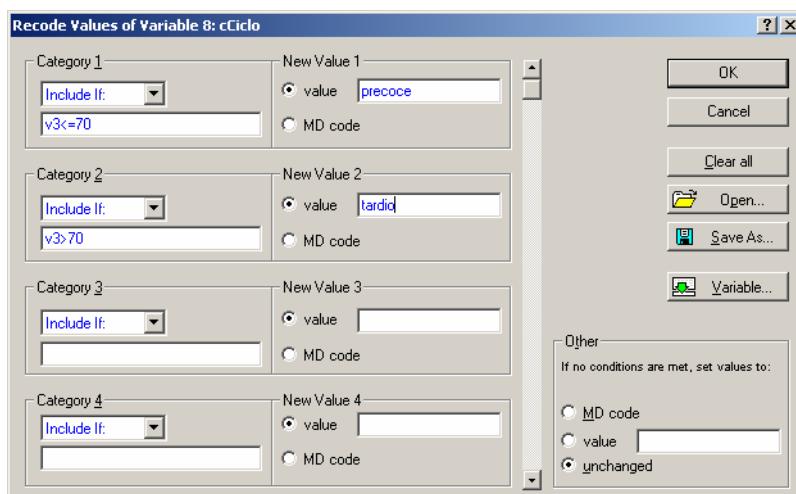


Figura 3.69 – Janela para a categorização da variável Ciclo

(5) No menu [Graphs], escolha a opção [Categorized Graphs], em seguida [Histograms] e aparecerá a janela apresentada na Figura 3.70:

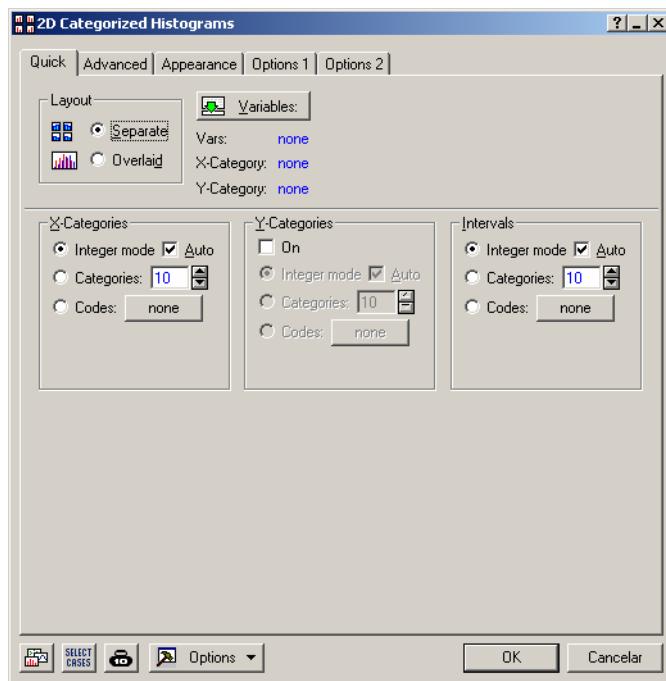


Figura 3.70 – Janela para a especificação do gráfico de colunas multiplas

(6) Selecione as variáveis, clicando no botão [Variables] e depois escolha a variável a ser estudada e a que será a categoria, conforme apresentado na Figura 3.71:

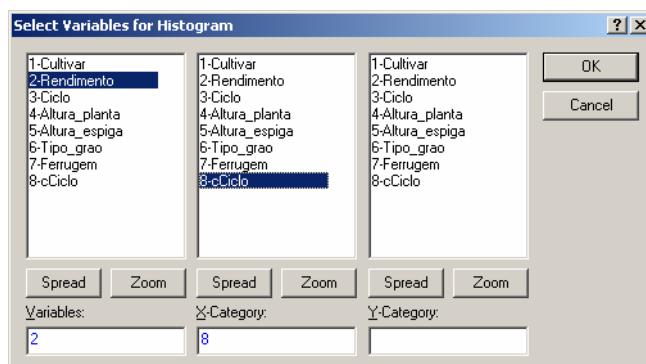


Figura 3.71 – Janela para seleção das variáveis

(7) Na janela da Figura 3.70, especifique:

- o tipo de gráfico (Layout:) *Overlaid*

O resultado das especificações está apresentado na Figura 3.72:

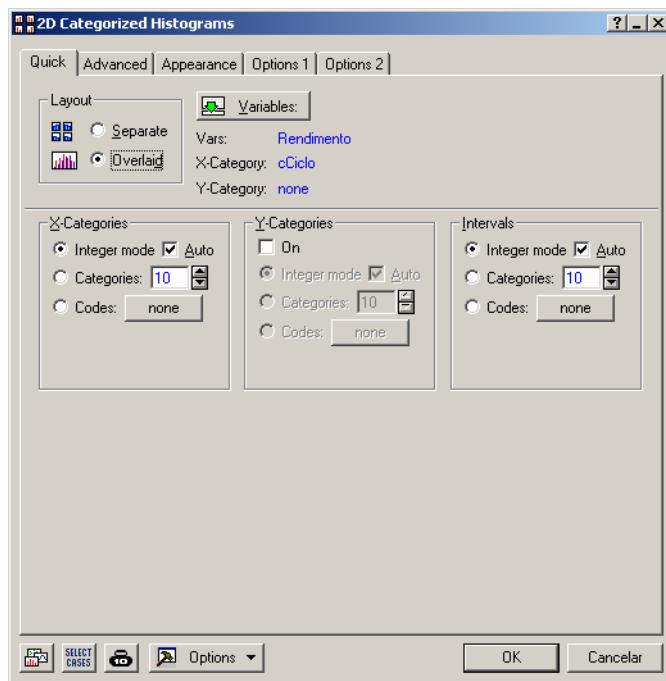


Figura 3.72 – Janela para a especificação do gráfico de colunas multiplas

(8) Clique no botão [OK] e o gráfico resultante está apresentado na Figura 3.73:

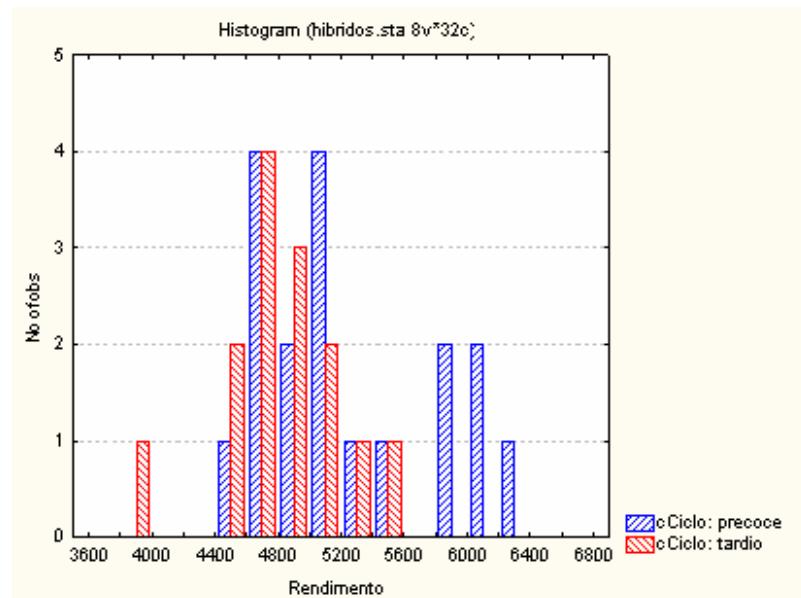


Figura 3.73 – Gráfico de colunas multiplas, rendimento versus tipo de ciclo (precoce ou tardio)

(9) Para o gráfico de freqüências multiplas, clique com o botão direito do *mouse* em cima do gráfico e escolha a opção (*Graph Properties (All Options)*). Na janela que abrirá, clique na aba [Plot: General], ative o *Multiple Lines* conforme apresentado na Figura 3.74:

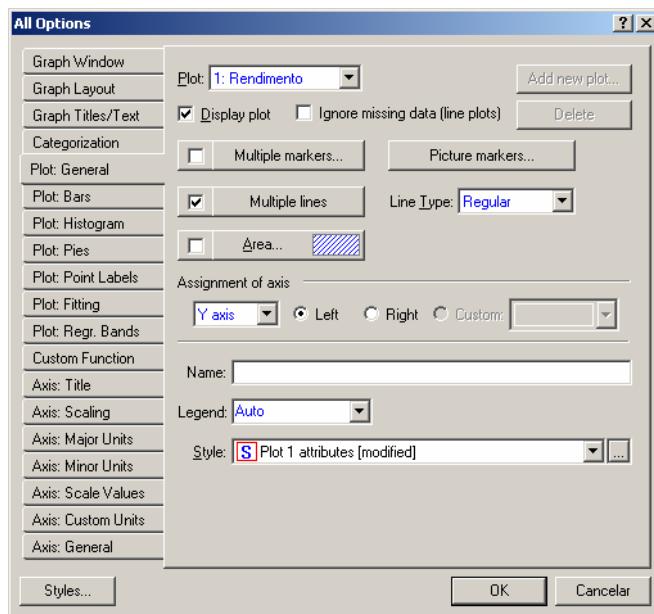


Figura 3.74 – Opções gráficas

(10) Na janela da Figura 3.74, clique na aba [Plot: Bars], desative o *Display Bars* conforme apresentado na Figura 3.75:

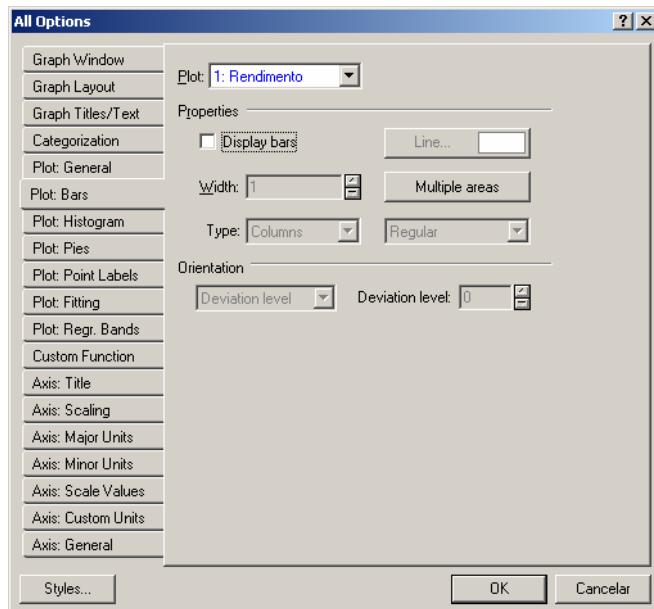


Figura 3.75 – Opções gráficas

(11) Clique no botão [OK] e o gráfico resultante está apresentado na Figura 3.76:

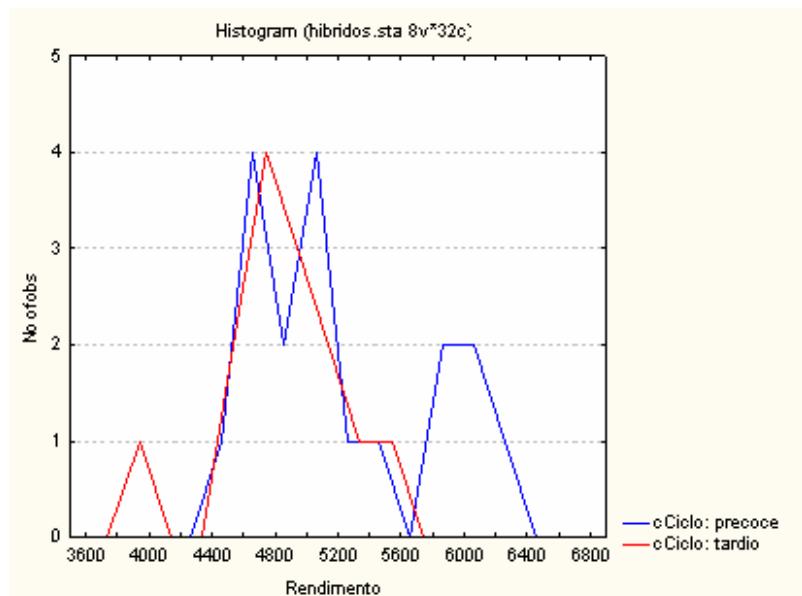


Figura 3.76 – Gráfico polígono de freqüências múltiplas, rendimento versus tipo de ciclo (precoce ou tardio)

3.2.5. O Modelo Normal

Uma distribuição de freqüência muito importante em estatística surge quando os dados tendem a se concentrarem simetricamente em torno de um valor central. Essa distribuição é conhecida como a distribuição normal ou *Gaussian*a e sua forma é a de sino.

A maioria dos fenômenos naturais têm esta distribuição, porém, apesar desta forma ser a mais esperada, o aspecto é teórico e cabe ao pesquisador estudá-lo.

O STATISTICA®, por *default*, quando constrói o histograma, também representa o modelo normal. Assim, o pesquisador pode ter uma idéia se a sua amostra é oriunda de uma distribuição aproximadamente normal.

Normal Probability Plot (O Gráfico Normal de Probabilidades)

- (1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], a opção a ser escolhida será *Frequency Table*. Selecione a variável Rendimento.
- (2) Clique na aba [Descr.] conforme apresentado na Figura 3.77:



Figura 3.77 – Opções da tabela de freqüência

(3) Clique no botão [Normal probability plots (2)] e o resultado está apresentado na Figura 3.78:

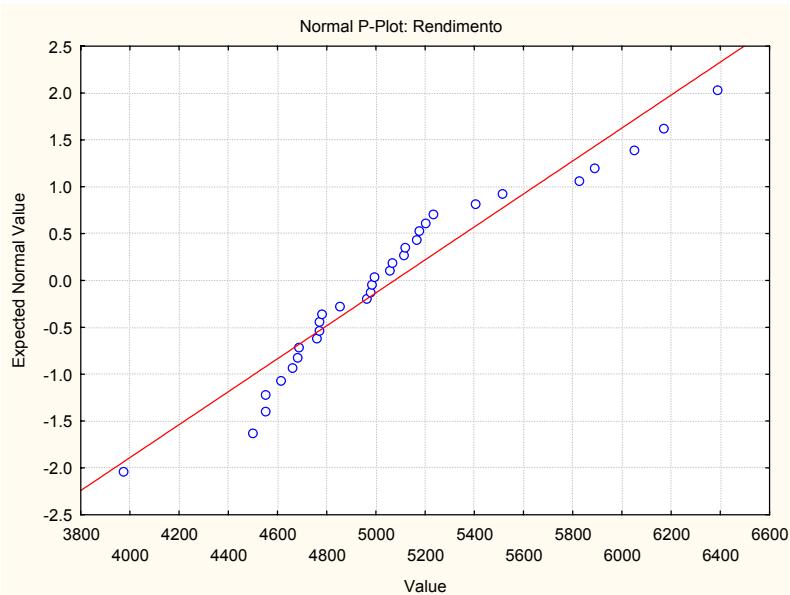


Figura 3.78 – Gráfico Normal de Probabilidade da variável Rendimento

Interpretação: como os pontos estão dispostos próximos à linha reta, conclui-se que a nossa amostra é oriunda de uma distribuição aproximadamente normal.

3.3. Estatísticas Descritivas

3.3.1. Medidas de Tendência Central e Dispersão

A variável Rendimento é adotada como exemplo para se obter as estatísticas descritivas.

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 3.79, onde a opção a ser escolhida será *Descriptive statistics*:

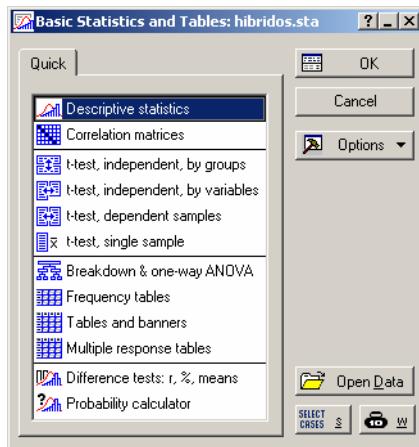


Figura 3.79 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.80:

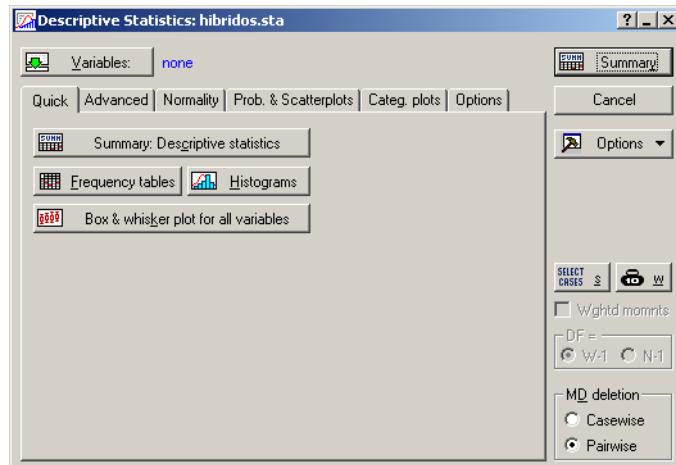


Figura 3.80 – Estatísticas descritivas

(3) Clique no botão [Variables] e selecione a variável Rendimento.

(4) Clique na aba [Advanced] conforme apresentado na Figura 3.81:

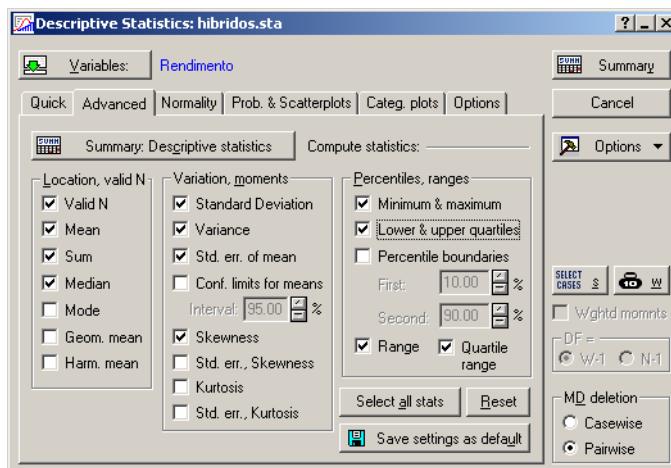


Figura 3.81 – Opções de estatísticas descritivas

Na janela anterior, selecionou-se as estatísticas relevantes para o caso, conforme apresentado a seguir:

- *Valid N* - tamanho da amostra
- *Mean* – média
- *Sum* – soma
- *Median* – mediana
- *Standard Deviation* - desvio padrão
- *Variance* – variância
- *Std. err. of mean* - erro padrão da média
- *Skewness* – assimetria
- *Minimum & maximum* - mínimo e máximo
- *Lower & upper quartiles* - primeiro quartil (Q1) e terceiro quartil (Q3)
- *Range* – amplitude
- *Quartile range* - desvio interquartílico

(4) Clique no botão [Summary] e os resultados estão apresentados nas figuras 3.82 e 3.83:

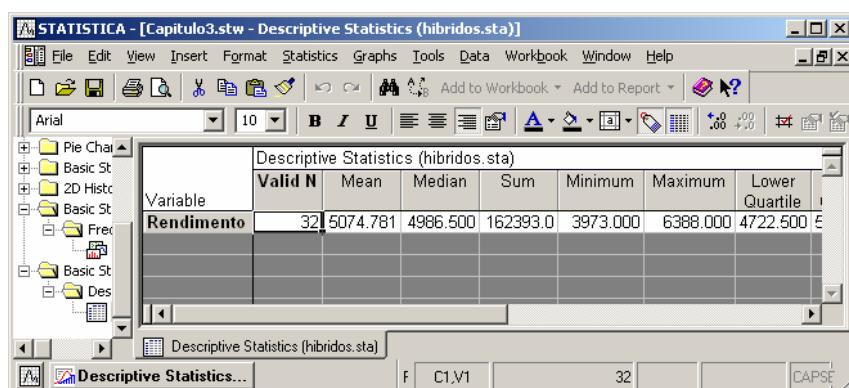


Figura 3.82 – Estatísticas descritivas

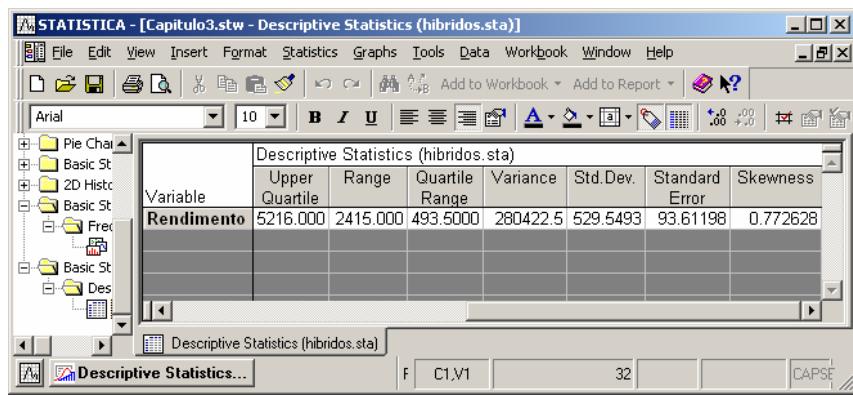


Figura 3.83 – Estatísticas descritivas

Outro exemplo:

Esta análise será baseada em um outro arquivo: **bezerros.sta**, que deve ser criado conforme dados apresentados no Quadro 3.1.

Quadro 3.1 - Resultado experimental do peso e comprimento, para duas raças de bezerros

Raça	Peso (Kg)	Comprimento (cm)
A	45	102
A	46	98
A	47	89
A	49	91
A	50	110
A	50	81
A	51	96
A	51	108
A	52	85
A	53	104
B	40	86
B	43	79
B	44	82
B	46	90
B	48	72
B	51	69
B	54	93
B	55	88
B	56	79
B	57	83

Como primeira análise, serão determinadas as estatísticas básicas da variável Peso para cada tipo de raça de bezerros (A e B) apresentados no quadro anterior, para isso basta seguir o procedimento abaixo:

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 3.84, onde a opção a ser escolhida será *Breakdown & one-way ANOVA*:

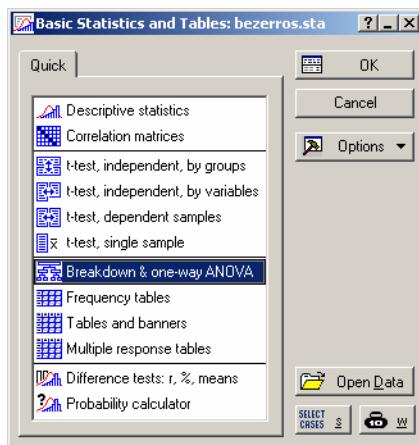


Figura 3.84 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.85:

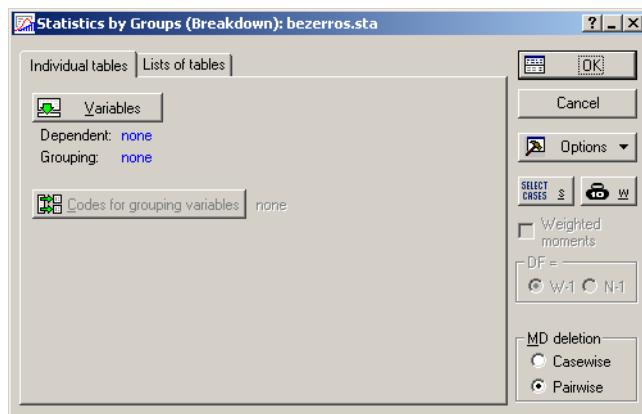


Figura 3.85 – Estatísticas descritivas por variável agrupadora

(3) Clique no botão [Variables] e selecione a variável Peso como dependente (*Dependent variables*) e a variável raça como agrupadora (*Grouping variables*), conforme apresentado na Figura 3.86.

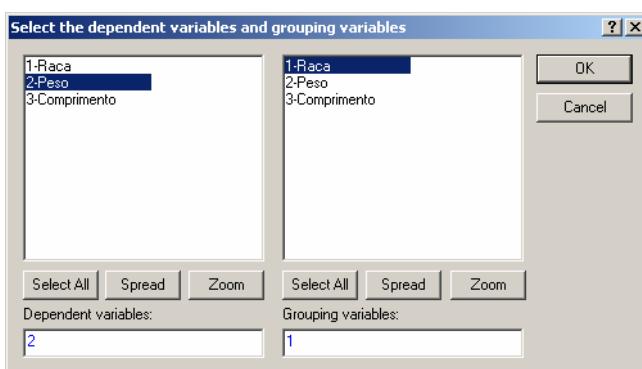


Figura 3.86 – Janela para seleção das variáveis para a análise

(4) Clique nos botões [OK] das duas figuras anteriores e aparecerá a janela apresentada na Figura 3.87, após o clique na aba [Descriptives]:

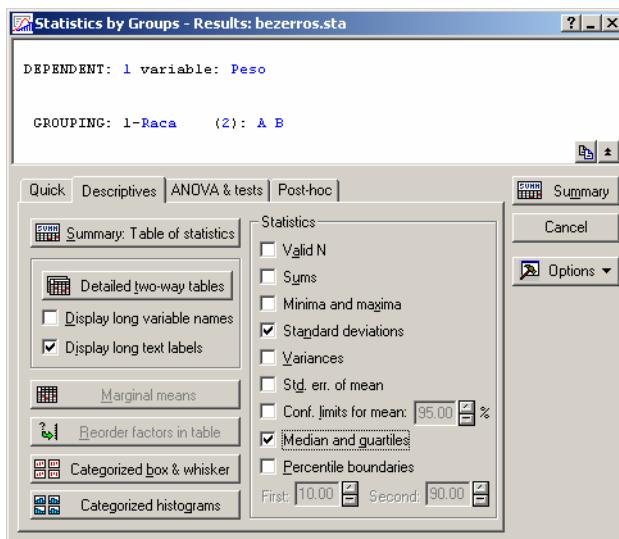


Figura 3.87 – Opções de estatísticas descritivas

Na janela anterior, selecionou-se as estatísticas relevantes para o caso, sendo que a média (*mean*) vem automaticamente, conforme apresentado a seguir:

- *Standard Deviation* - desvio padrão
- *Median & quartiles* – mediana, primeiro quartil (Q1) e terceiro quartil (Q3)

(5) Clique no botão [Summary] e os resultados estão apresentados na Figura 3.88:

Raca	Peso Means	Peso Std.Dev.	Peso Q25	Peso Median	Peso Q75
A	49.40000	2.633122	47.00000	50.00000	51.00000
B	49.40000	6.040603	44.00000	49.50000	55.00000
All Grps	49.40000	4.535242	46.00000	50.00000	52.50000

 The 'Statistics by Groups' button is visible at the bottom left of the window."/>

Figura 3.88 - Estatísticas descritivas da variável Peso por tipo de raça

A análise estatística descritiva é utilizada como uma descrição de um conjunto de valores.

CUIDADO !!! A média e o desvio padrão são afetados por valores extremos, altos ou baixos, e a estatística torna-se irreal. A **média** (*Mean*) representará bem apenas em casos onde existe simetria, pois a média de um modelo assimétrico não reflete a realidade, e neste caso, a **mediana** (*Median*) torna-se uma melhor maneira de representação.

O **desvio padrão** (*Standard deviation*) e o **erro padrão da média** (*Std. err. of mean*) são medidas calculadas em torno da média e a intenção é quantificar a variabilidade dos dados em torno da média. Por ser, então, uma medida relativa, deve existir uma referência para que façamos a interpretação. Um pesquisador

com experiência já espera um determinado valor da variabilidade, caso contrário, a análise deve ser baseada em resultados de outros trabalhos.

Nos **quartis** temos a distribuição dividida em quatro partes iguais.



3.3.2. A Média e o Desvio Padrão sob um Modelo Normal

Nessa análise, será utilizada a ferramenta do **STATISTICA®** de cálculo de probabilidade, para isso basta seguir o procedimento abaixo:

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 3.89, onde a opção a ser escolhida será *Probability calculator*:

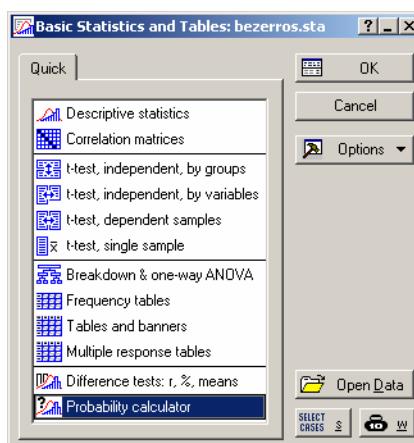


Figura 3.89 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.90:

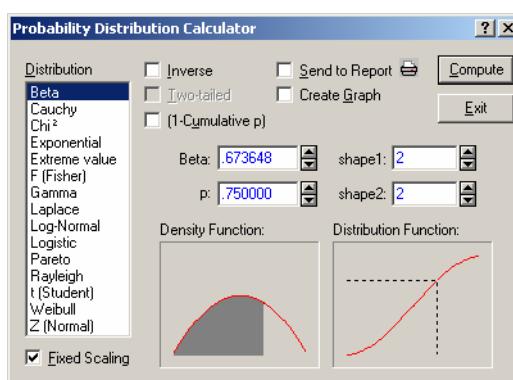


Figura 3.90 – Janela para cálculo de probabilidade

(3) Na janela anterior, especificar:

- *Distribution: Z (Normal)*
- *Two-tailed* (bilateral) e *Create Graph* (criar o gráfico)

- X : digite o valor 1

Na figura a seguir estão apresentados os resultados da especificação.

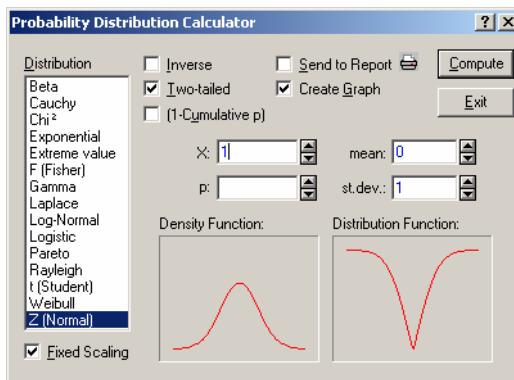


Figura 3.91 – Janela para cálculo de probabilidade

(4) Clique no botão [Compute], aparecerá o valor da probabilidade (p) e os gráficos das funções densidade e distribuição de probabilidade, apresentados respectivamente nas figuras 3.92 e 3.93:

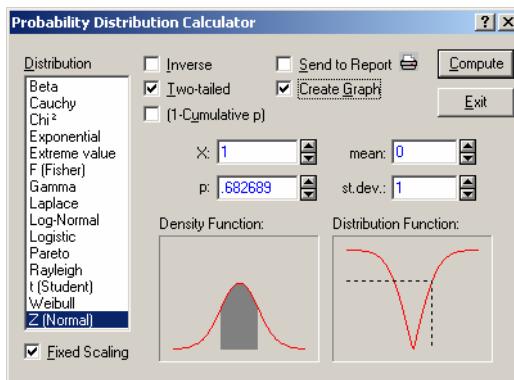


Figura 3.92 – Janela para cálculo de probabilidade

Observação: Para ter acesso ao gráfico é necessário sair da janela de cálculo de probabilidade, para isto clique no botão [Exit].

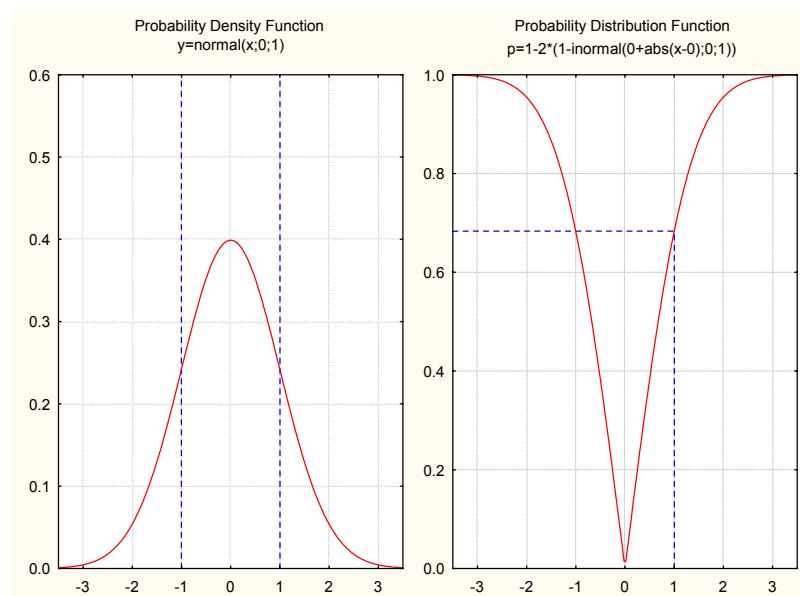


Figura 3.93 - Distribuição normal padrão: média mais ou menos 1 desvio padrão ($\bar{x} \pm 1s$)

Interpretação: Observar na Figura 3.81, que no intervalo $\bar{x} \pm 1s$, temos exatamente 68,2689% dos valores ($p=0,682689$).

(4) Repetir os itens (3) e (4) alterando o valor de X para 2 e 3. Os resultados estão apresentados nas figuras a seguir:

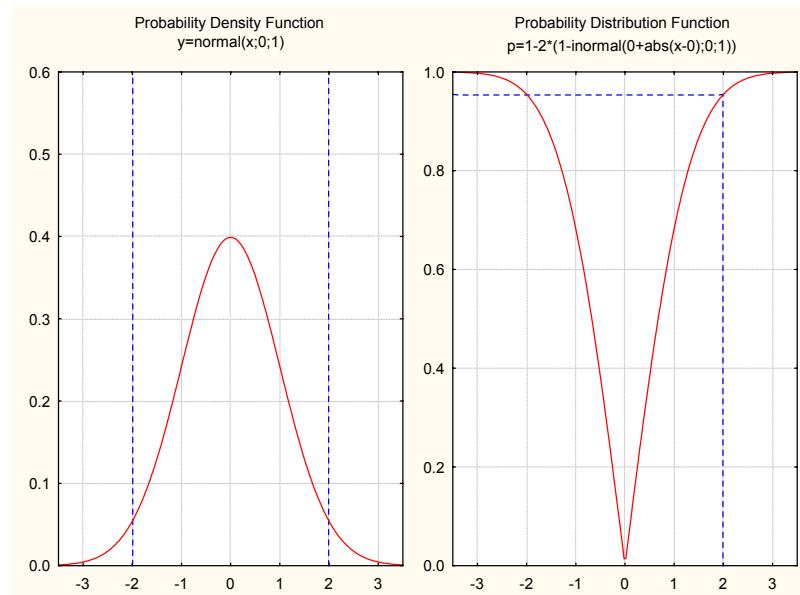


Figura 3.94 - Distribuição normal padrão: média mais ou menos 2 desvios padrões ($\bar{x} \pm 2s$)

Interpretação: Temos agora, 95,45% dos valores ($p=0,9545$).

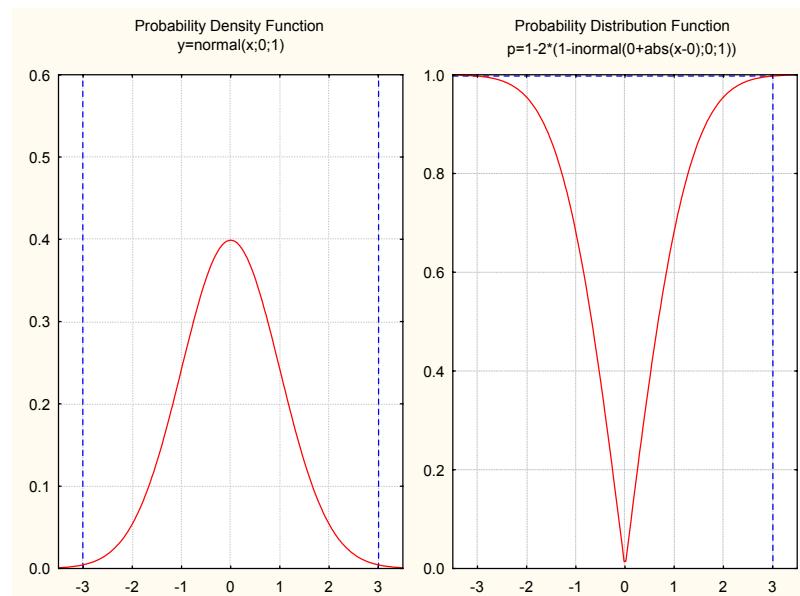


Figura 3.95 - Distribuição normal padrão: média mais ou menos 3 desvios padrões ($\bar{x} \pm 3s$)

Interpretação: Temos agora, 99,73% dos valores ($p=0,9973$).

3.3.3. Desenho Esquemático (Box Plot)

O uso da mediana, quartis e extremos é importante para obtermos informações sobre a forma, valor representativo, dispersão e valores discrepantes da distribuição. A análise será realizada sobre o arquivo **bezerros.sta**.

(1) No menu [Graphs], escolha a opção [2D Graphs], em seguida [Box Plots] e aparecerá a janela apresentada na Figura 3.96:

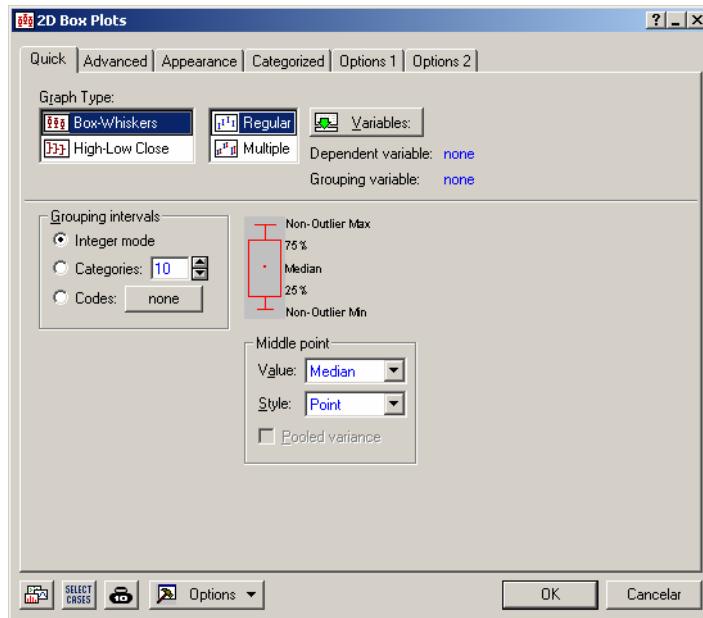


Figura 3.96 – Janela para especificação do desenho esquemático

(2) Selecione a variável, clicando no botão [Variables] e depois escolha a variável Peso como a variável a ser analisada (dependent variable), clique no botão [Ok] e aparecerá a janela apresentada na figura a seguir:

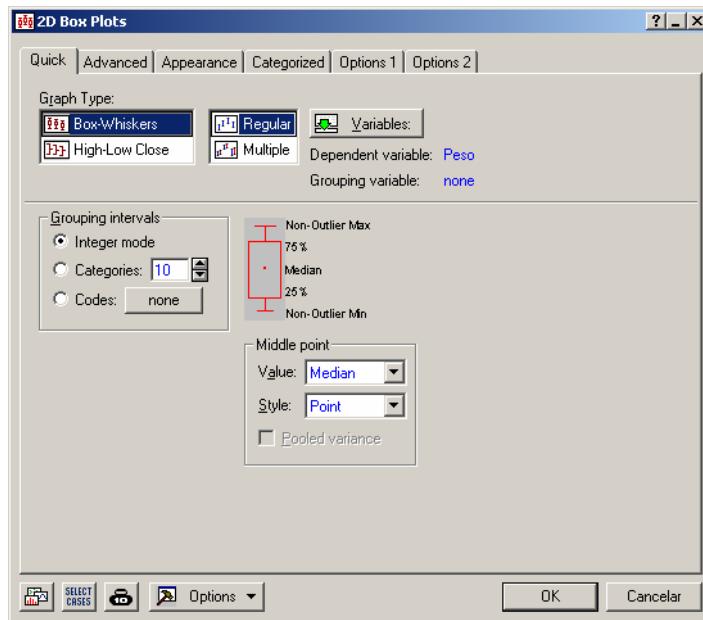


Figura 3.97 – Janela para especificação do desenho esquemático

(3) Clique no botão [OK] e o gráfico resultante está apresentado na figura a seguir:

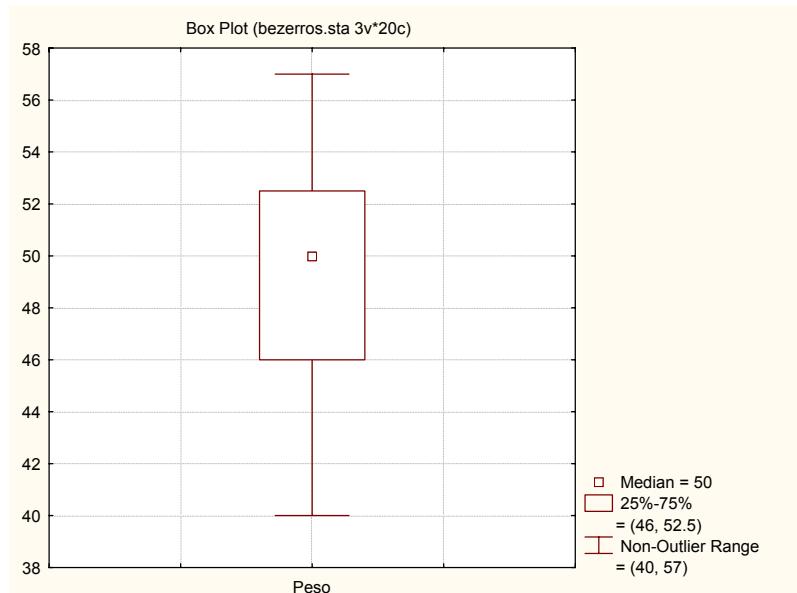


Figura 3.98 – Desenho esquemático em duas dimensões para o peso

Para saber se existe diferença na distribuição entre as duas raças, pode-se construir um gráfico *box plot* categorizado, conforme apresentado no procedimento a seguir:

(4) Na janela da Figura 3.97, clique na aba [Advanced] e especifique:

- as variáveis: independente (*Independent variables*): PESO e agrupadora (*Grouping variables*): RACA
- o tipo de gráfico (*Graph Type*): *Box – Whiskers* e *Regular*
- o intervalo de agrupamento (*Grouping intervals*): *Integer mode*
- o ajuste (*FIT*): *off* (sem linha)
- o ponto central (*Middle Point*): *Value=Median* e *Style=Point*
- a caixa (*Box*): *Value=Percentiles* e *Coefficient=25*
- o limite (*Whisker*): *Value=Non Outlier Range* e *Coefficient=1 (default)*
- os pontos discrepantes (*Outliers*): *Value=Outl & Extremes* e *Coefficient=1,5*

O resultado das especificações está apresentado na figura a seguir:

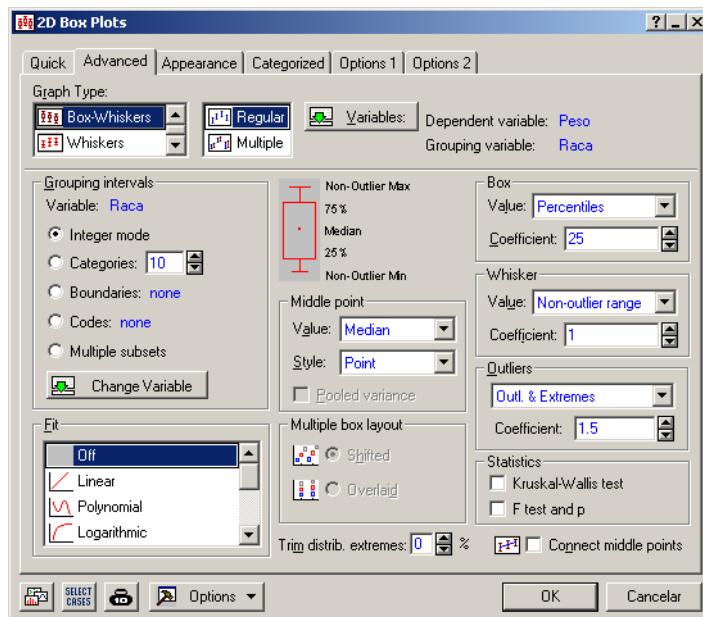


Figura 3.99 – Janela para a especificação do desenho esquemático

(5) Clique no botão [OK] e o gráfico resultante está apresentado na figura a seguir:

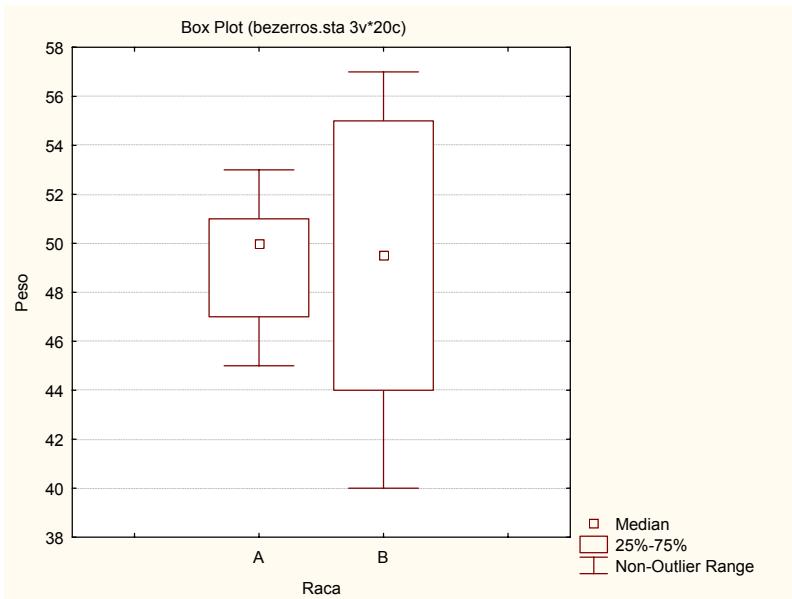


Figura 3.100 – Desenho esquemático para as raças A e B

Se diminuirmos o coeficiente para 0,4 vão aparecer *outliers* no desenho, conforme apresentado na Figura 3.101. O padrão do STATISTICA® e de muitos livros de estatística é usar o critério coeficiente (*Coefficient*) igual a 1,5, porém, o pesquisador tem autonomia para alterá-lo.

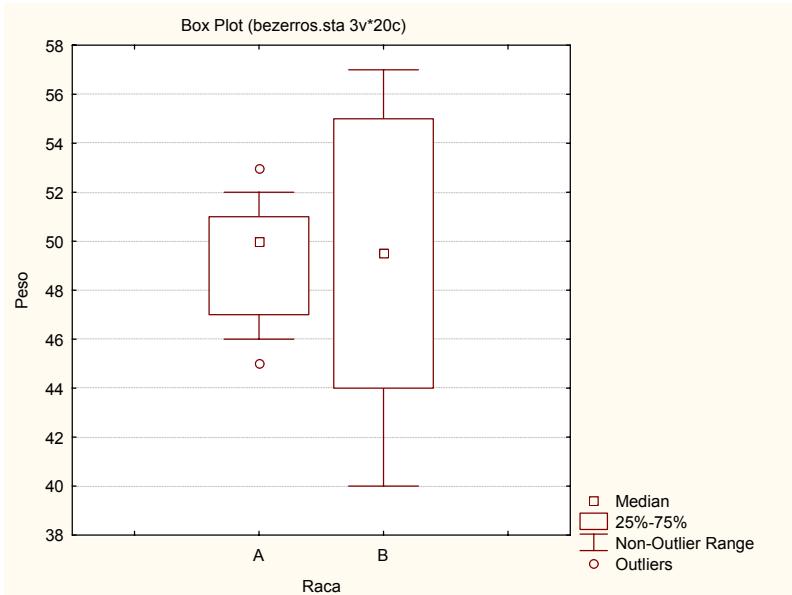


Figura 3.101 – Desenho esquemático para as raças A e B, usando o coeficiente igual a 0,4

3.3.4. Diagrama de Dispersão

Para verificar o relacionamento entre duas variáveis pode-se utilizar o diagrama de dispersão. A análise será realizada sobre o arquivo **híbridos.sta**.

(1) No menu [Graphs], escolha a opção [2D Graphs], em seguida [Scatterplots] e aparecerá a janela apresentada na Figura 3.102:

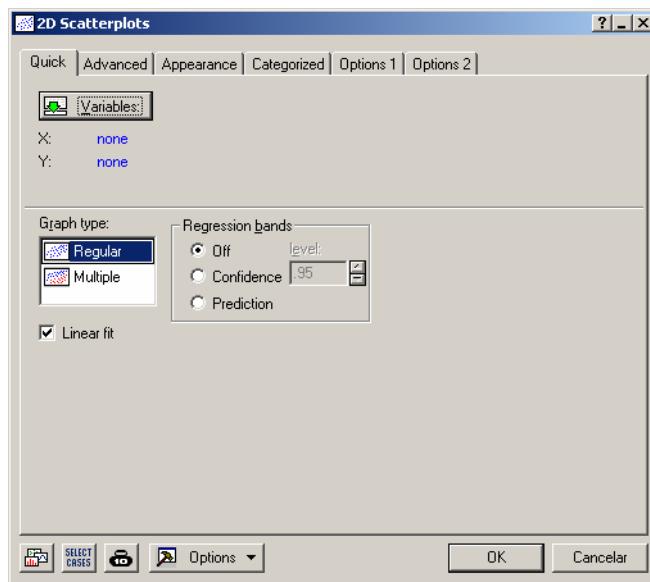


Figura 3.102 – Janela para especificação do diagrama de dispersão

(2) Na janela da Figura anterior, especifique:

- as variáveis: X: Altura_planta e Y: Altura_espiga
- o tipo de gráfico (*Graph Type*): *Regular*
- o ajuste (*Linear fit*): sem seleção

O resultado das especificações está apresentado na figura a seguir:

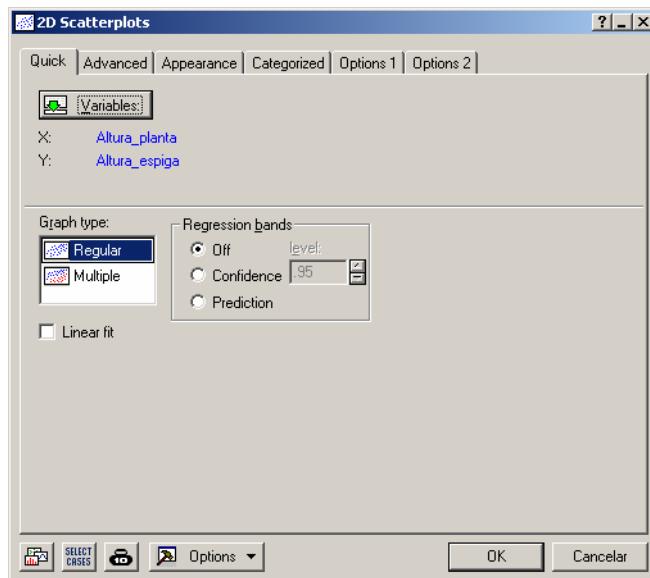


Figura 3.103 – Janela para a especificação do diagrama de dispersão

(5) Clique no botão [OK] e o gráfico resultante está apresentado na figura a seguir:

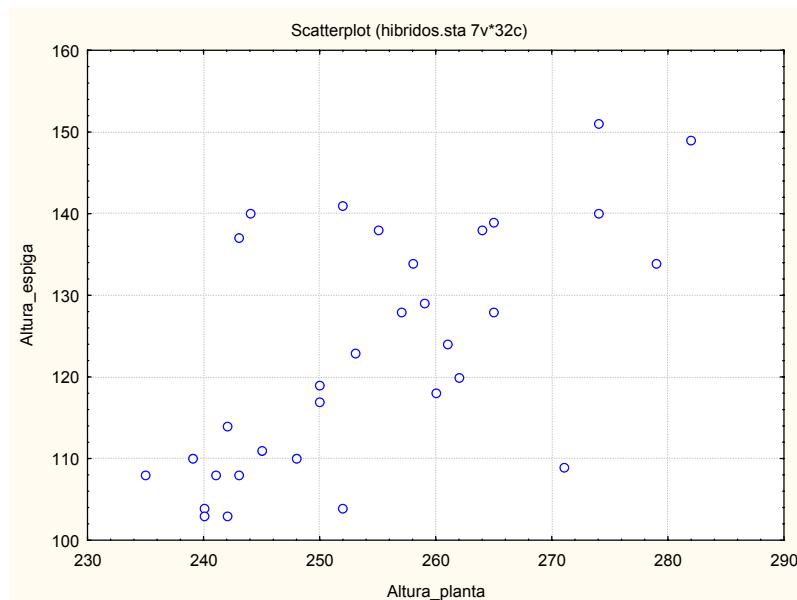


Figura 3.104 – Diagrama de dispersão para as variáveis quantitativas, altura de espiga e altura de planta.

Interpretação: de modo geral observamos uma tendência linear nos dados, isto é, quanto mais altas são as plantas, mais altas as alturas das espigas.

3.3.5. Coeficiente de Correlação

(1) No menu [STATISTICS], escolha a opção [Basic Statistics/Tables], aparecerá a janela apresentada na Figura 3.105, onde a opção a ser escolhida será *Correlation matrices*:

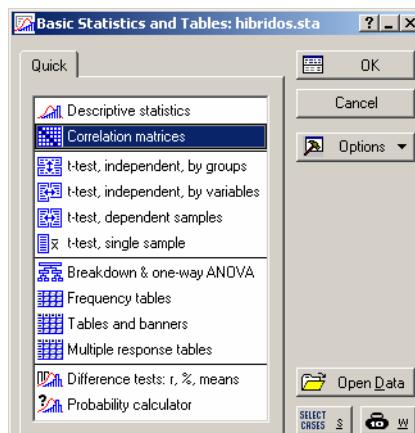


Figura 3.105 – Janela para o cálculo do coeficiente de correlação

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.106:

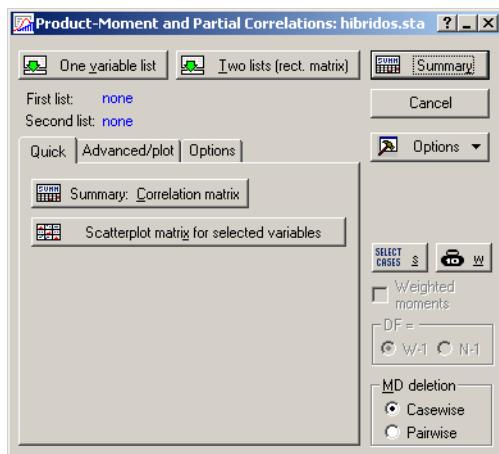


Figura 3.106 – Janela para seleção das variáveis para a análise

(3) Clique no botão [Two lists (rect. matrix)] e selecione a variável Altura_planta (*First variable list*) e a variável Altura_espiga (*Second variable list (optional)*), conforme apresentado na Figura 3.107.

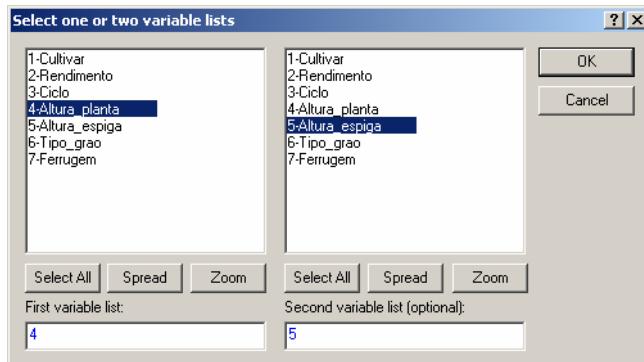


Figura 3.107 – Janela para seleção das variáveis para a análise

(4) Clique no botão [OK] e aparecerá a janela apresentada na Figura 3.108:

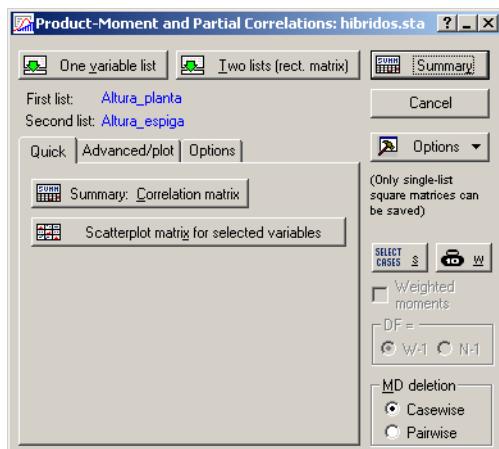


Figura 3.108 – Janela para seleção das variáveis para a análise

(5) Clique no botão [Summary] da figura anterior e o valor da correlação estará apresentado na Figura 3.109:

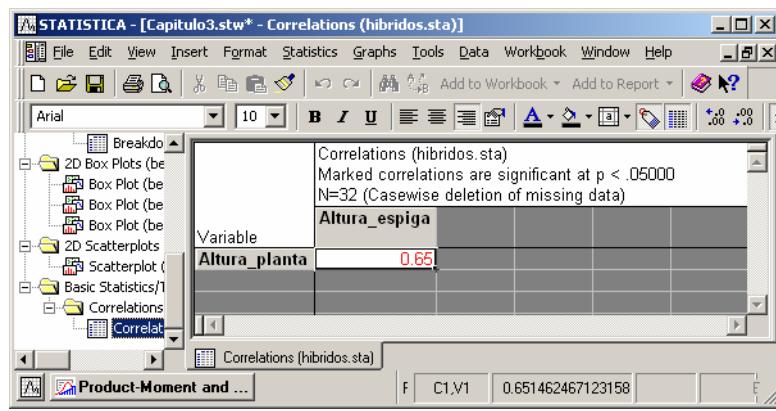


Figura 3.109 – Valor do coeficiente de correlação

Interpretação: Podemos concluir que as variáveis estão correlacionadas positivamente, ou seja, quanto maior a altura da planta maior é a altura da espiga.

3.3.6. Ajuste da Equação de uma Reta

O ajuste de um modelo linear simples é como resultado uma equação matemática que descreve o relacionamento entre duas variáveis. Para ajustar um modelo linear simples entre a altura da espiga e altura da planta, basta repetir o procedimento do item 3.3.4. habilitando o ajuste linear (*Linear fit*), conforme apresentado na Figura 3.110.

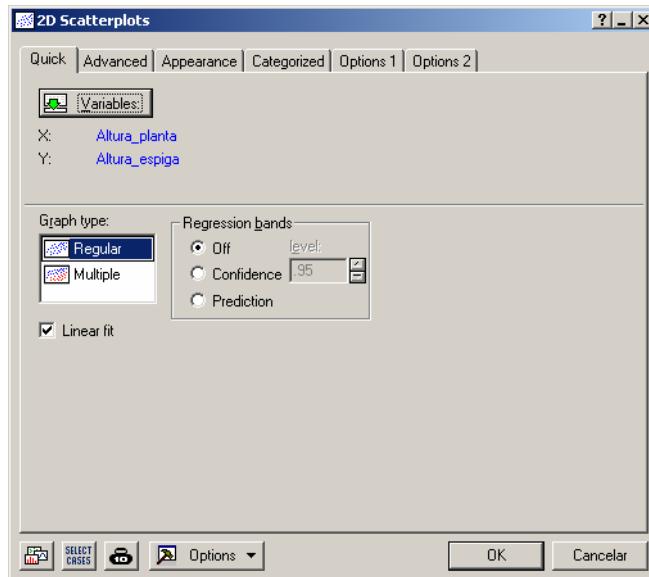


Figura 3.110 – Janela para a especificação do diagrama de dispersão

(1) Clique no botão [OK] e o gráfico resultante está apresentado na figura a seguir:

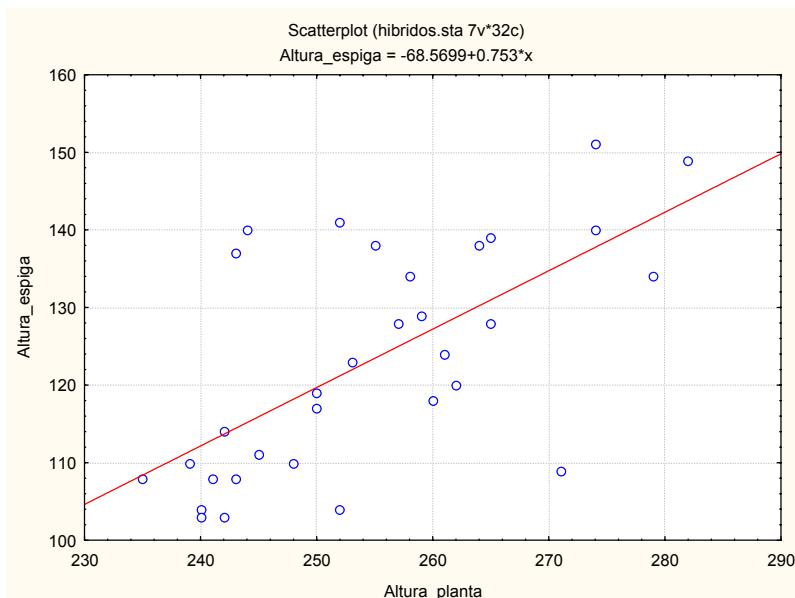


Figura 3.111 – Valores observados e a equação de regressão ajustada

O modelo linear simples ajustado é:

$$\hat{\text{Altura_espiga}} = -68,5699 + 0,753 \cdot \text{Altura_planta}$$

Interpretação: o aumento de 1 cm no valor da altura da planta faz com que a altura da espiga aumente em 0,753 cm.

Análise dos Resíduos

Essa análise é feita para a variável dependente e permite verificar se o modelo ajustado é adequado para os dados. Os resíduos consistem da diferença entre os valores observados e os estimados pelo modelo ajustado, conforme equação a seguir:

$$e_i = Y_i - \hat{Y}_i$$

onde:

e_i = resíduo da i -ésima observação

Y_i = valor observado da variável resposta da i -ésima observação

\hat{Y}_i = valor estimado da variável resposta da i -ésima observação

O modelo será considerado adequado se a distribuição dos resíduos for normalmente distribuída com média nula e variância constante (σ^2). Uma das formas de verificação é através de análise gráfica, explorada neste tópico. Na seqüência estão os passos para a análise dos resíduos.

(1) No menu [INSERT], escolha a opção [Add Variables], conforme apresentado na Figura 3.112:

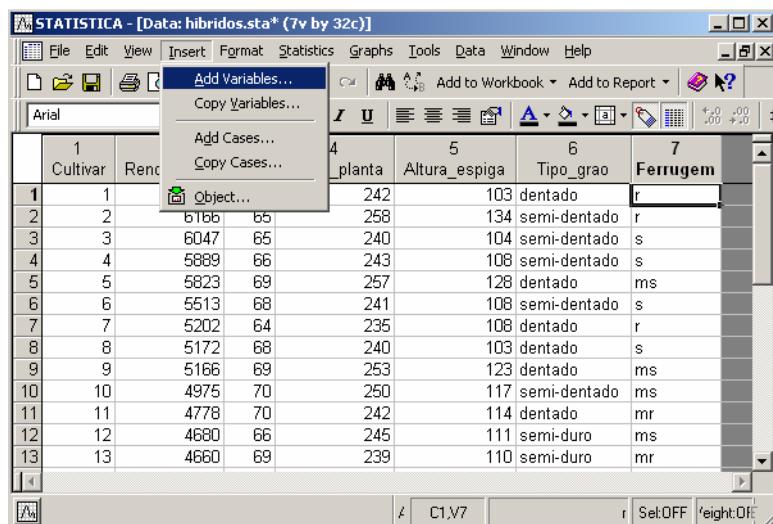


Figura 3.112 – Criação de novas variáveis

(2) Crie duas variáveis após a variável Ferrugem, conforme apresentado na Figura 3.113:

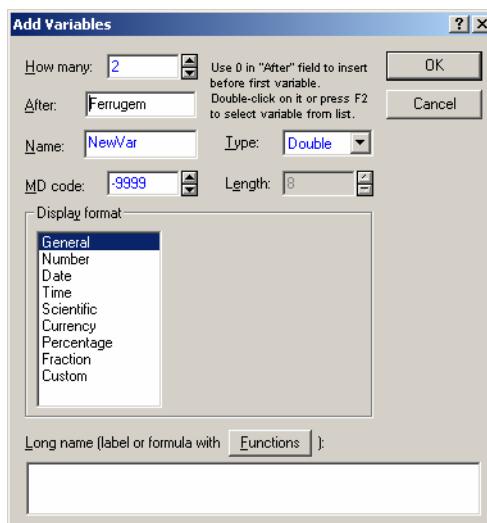


Figura 3.113 – Janela para a criação de novas variáveis

(3) Defina as variáveis, uma será os valores estimados e a outra os resíduos, conforme apresentado nas figuras 3.114 e 3.115, respectivamente:

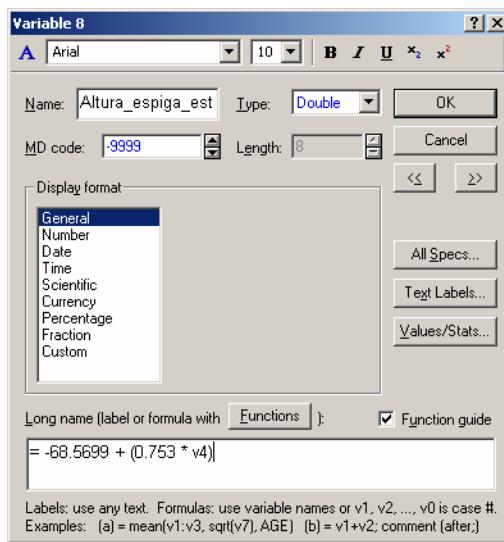


Figura 3.114 – Janela para a criação dos valores estimados

Observação: na janela anterior foi inserido o modelo linear ajustado ($= -68.5699 + (0.753 * v4)$), onde v4 corresponde à variável Altura_planta.

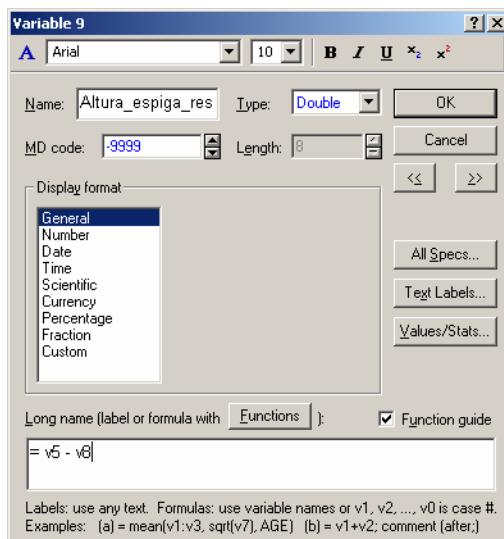


Figura 3.115 – Janela para a criação dos resíduos

Observação: na janela anterior foi inserido o cálculo dos resíduos ($= v5 - v8$), onde v5 corresponde à variável Altura_espiga e v8 à Altura_espiga_est (valores estimados).

(4) Para verificar se os resíduos são normalmente distribuídos, é construído o Gráfico normal de probabilidade normal dos resíduos, conforme apresentado na Figura 3.116, cujo resultado está apresentado na Figura 3.117:

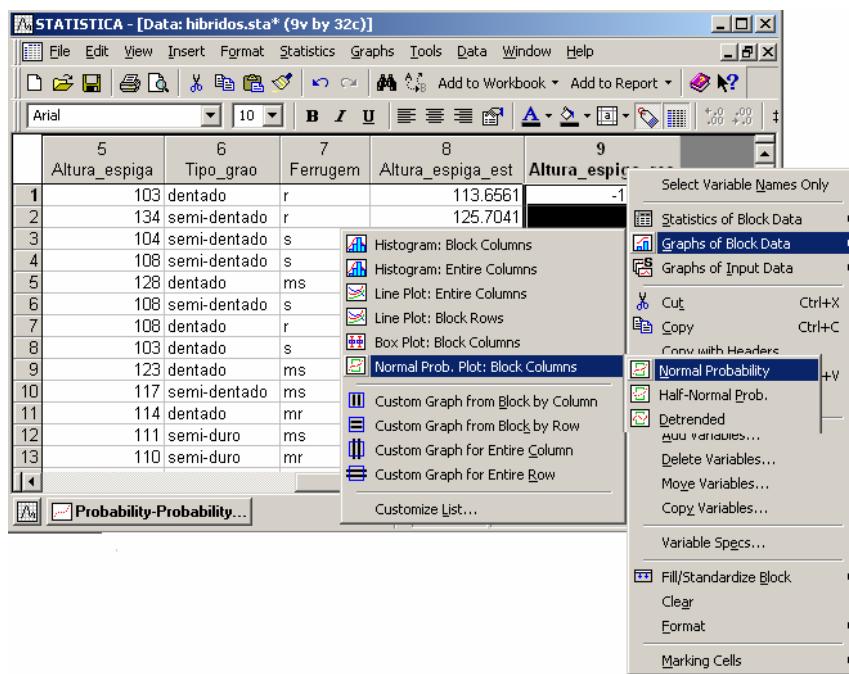


Figura 3.116 – Janela para a criação do gráfico de probabilidade normal dos resíduos

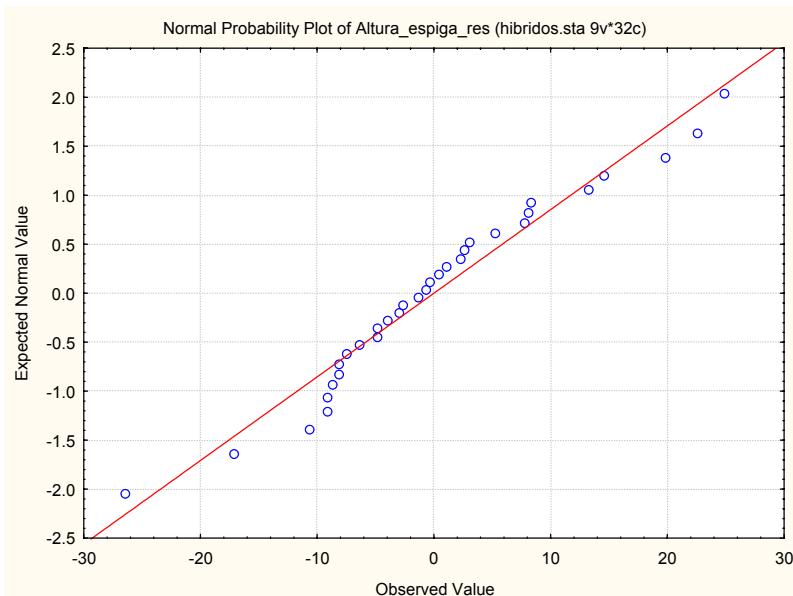


Figura 3.117 – Gráfico de probabilidade normal dos resíduos

Interpretação: o gráfico anterior indica que os resíduos possuem uma distribuição normal.

(5) Para verificar se os resíduos possuem variância constante, é construído um gráfico de dispersão (scatterplot) entre os resíduos e a variável independente ou preditora (Altura_planta), conforme apresentado na Figura 3.118, cujo resultado está apresentado na Figura 3.119:

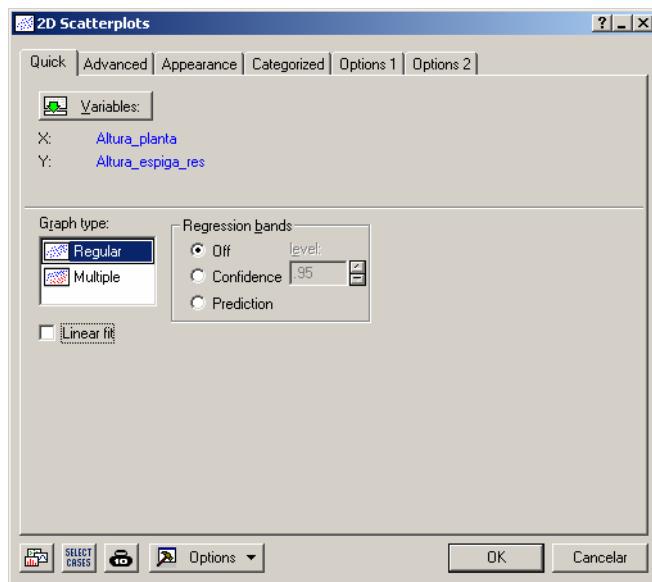


Figura 3.118 – Janela para a criação do gráfico de dispersão entre os resíduos e a variável independente ou preditora (Altura_planta)

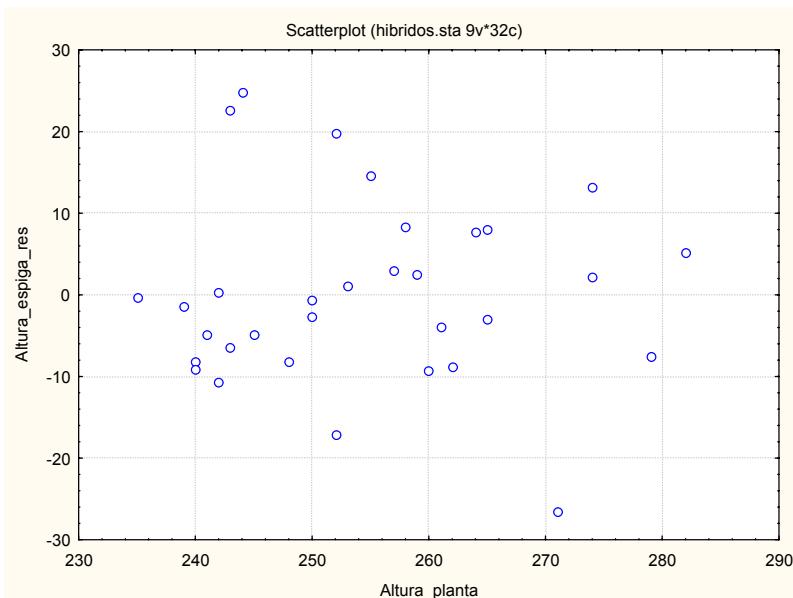


Figura 3.119 – Gráfico de dispersão entre os resíduos e a variável independente ou preditora (Altura_planta)

Interpretação: o gráfico anterior indica que a distribuição dos resíduos é aleatória e que a variância pode ser considerada constante.

4. OPÇÕES GRÁFICAS

Para editar o aspecto geral do gráfico, basta clicar duas vezes com o botão esquerdo do *mouse* sobre a área de fora do gráfico propriamente dita (área onde estão as palavras de legenda, título e as escalas). Uma outra forma é clicar com o botão direito do mouse e escolher a opção *Graph Properties (All Options)*, conforme apresentado na Figura 4.1:

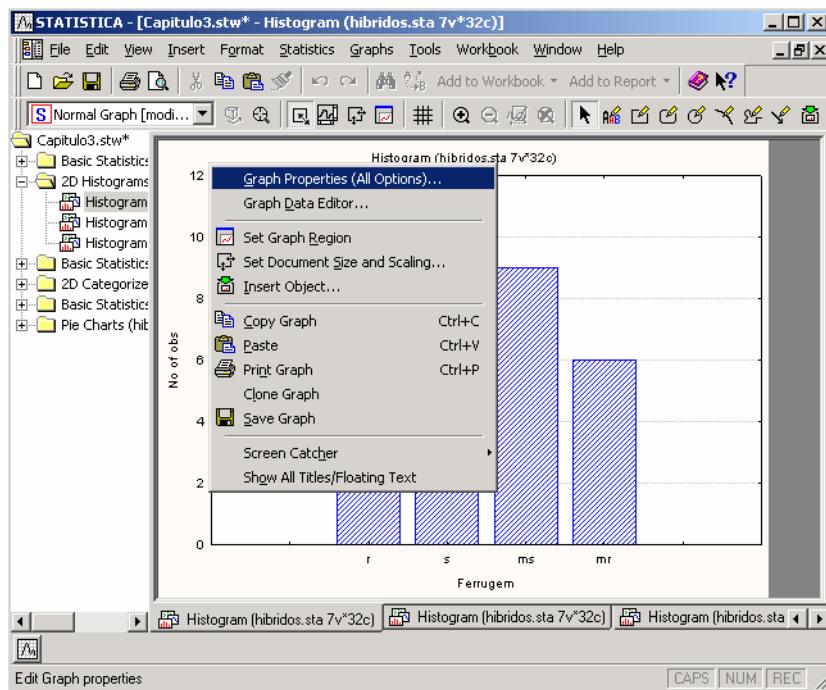


Figura 4.1 – Edição das propriedades do gráfico

Ambos os procedimentos abrem a janela de edição geral do gráfico apresentada na Figura 4.2, porém a segunda forma permite acessar outras opções, como por exemplo: copiar (*Copy Graph*) o gráfico ou capturar uma região do mesmo (*Screen Catcher*) para um editor de texto, imprimir (*Print Graph*).

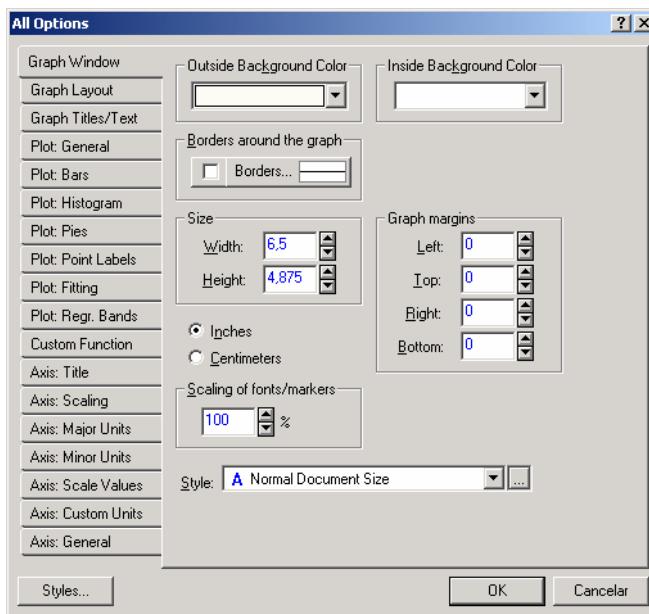


Figura 4.2 – Janela das propriedades do gráfico

Com a janela apresentada na Figura 4.2 é possível, por exemplo:

- alterar as propriedades da janela (*Graph Window*), como a cor do fundo (*Outside Background Color*), a borda (*Borders around the graph*) e o tamanho (*Size*) do mesmo;
- alterar o layout (*Graph Layout*) como a disposição dos eixos (*Axis positions*);
- modificar títulos e subtítulos (*Graph Titles/Text*), como o tipo, tamanho e cor da letra;
- editar os eixos (*Axis Scaling*), como os valores início (*Minimum*), fim (*Maximum*) e o passo (*Step Size*) da escala.

Existem outras opções que são específicas para cada tipo de gráfico, como:

- *Plot: Bar* – gráfico de barras;
- *Plot: Histogram* – histogramas;
- *Plot: Pies* – gráfico de setores;
- *Plot: Point Labels* – gráfico de dispersão.

Uma forma mais rápida para alterar a formatação de um gráfico é clicar duas vezes em cima do que se quer mudar e alterar conforme padrão do trabalho que se está fazendo ou para traduzir textos para o Português, por exemplo. Para exemplificar, será utilizado o gráfico de colunas construído para a variável resistência à ferrugem reapresentado na Figura 4.1. A seguir, está uma seqüência de procedimentos que visa exemplificar o potencial gráfico que está disponível no programa STATISTICA ®.

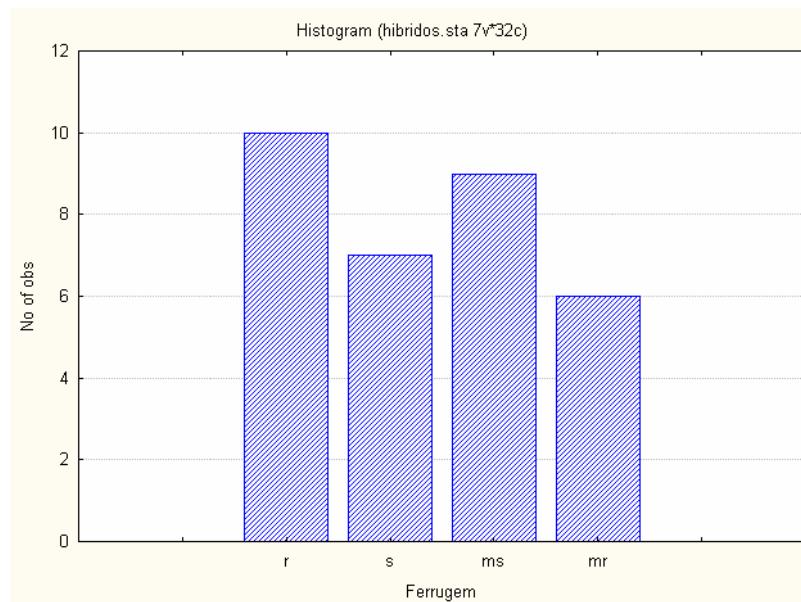


Figura 4.3 – Gráfico de colunas da variável resistência à ferrugem

(1) Clicar duas vezes em cima do título do gráfico, aparecerá a janela apresentada na Figura 4.4:

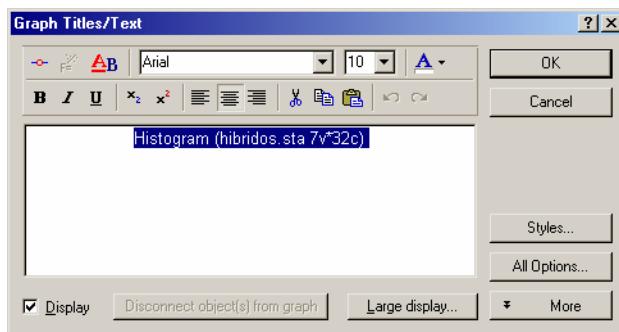


Figura 4.4 – Janela das propriedades do título

(2) Alterar o título, sua fonte, tamanho e cor, conforme seu padrão adotado. Na Figura 4.5 está um exemplo:

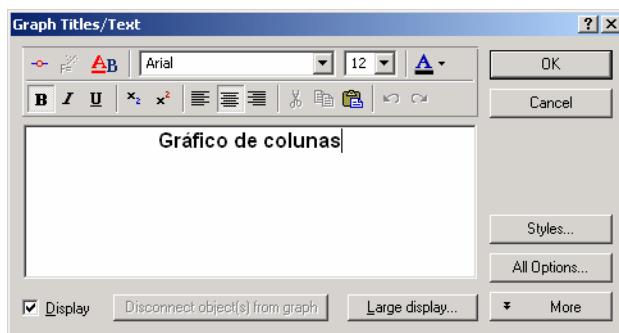


Figura 4.5 – Janela das propriedades do título

(3) Clicar duas vezes em cima do fundo do gráfico, cujo padrão é a cor amarela clara, trocando-se a cor conforme apresentado na Figura 4.6:

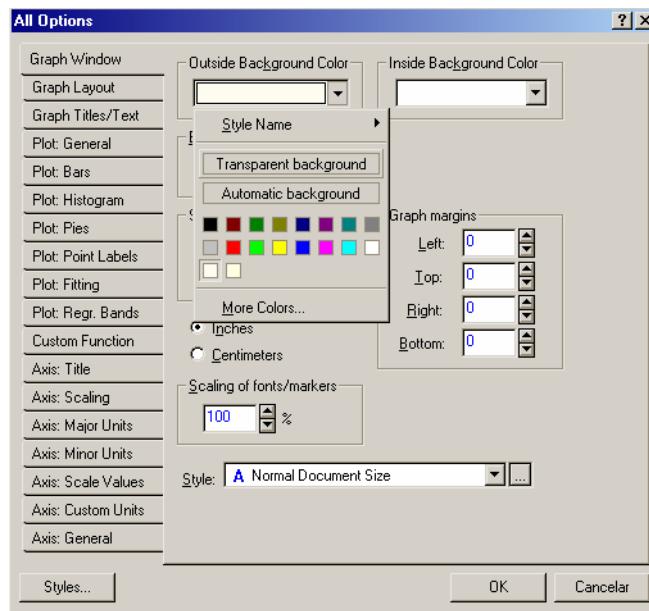


Figura 4.6 – Alteração das propriedades do fundo

(4) Clicar duas vezes no título do eixo que se deseja modificar, no exemplo o Y, apresentado na Figura 4.7:

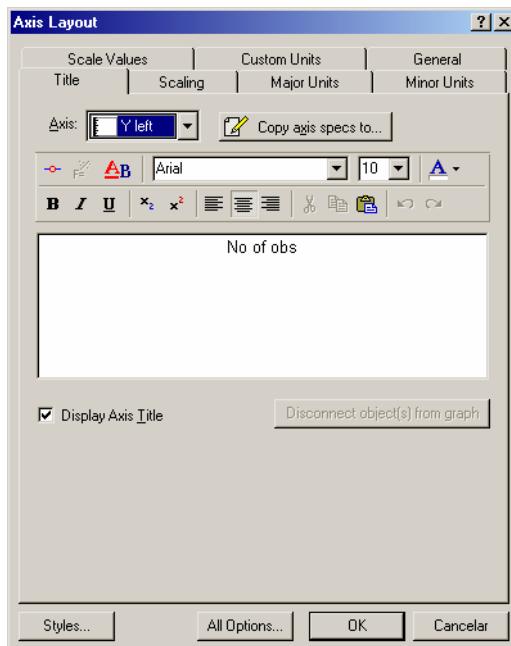


Figura 4.7 – Janela das propriedades do eixo Y

(5) Alterar o título do eixo, conforme apresentado na Figura 4.8:

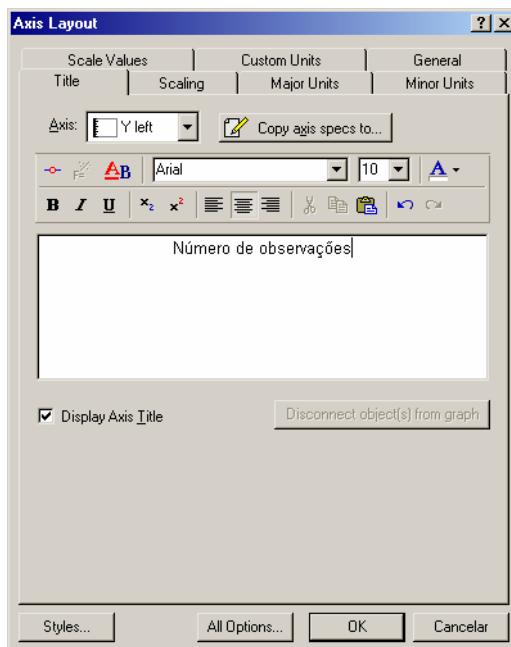


Figura 4.8 – Janela das propriedades do eixo Y

(6) Para alterar o eixo X, basta alterar a caixa de texto da opção *Axis* na Figura 4.8 para *X* e modificar o título deste, conforme apresentado na Figura 4.8:

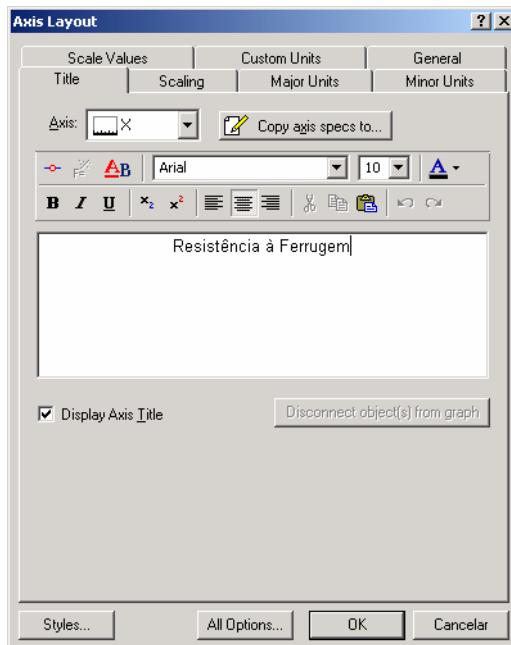


Figura 4.9 – Janela das propriedades do eixo Y

(7) Para modificar o preenchimento (área) e a cor das barras, clicar com o botão da direita em cima de uma das colunas, apresentado na Figura 4.10:

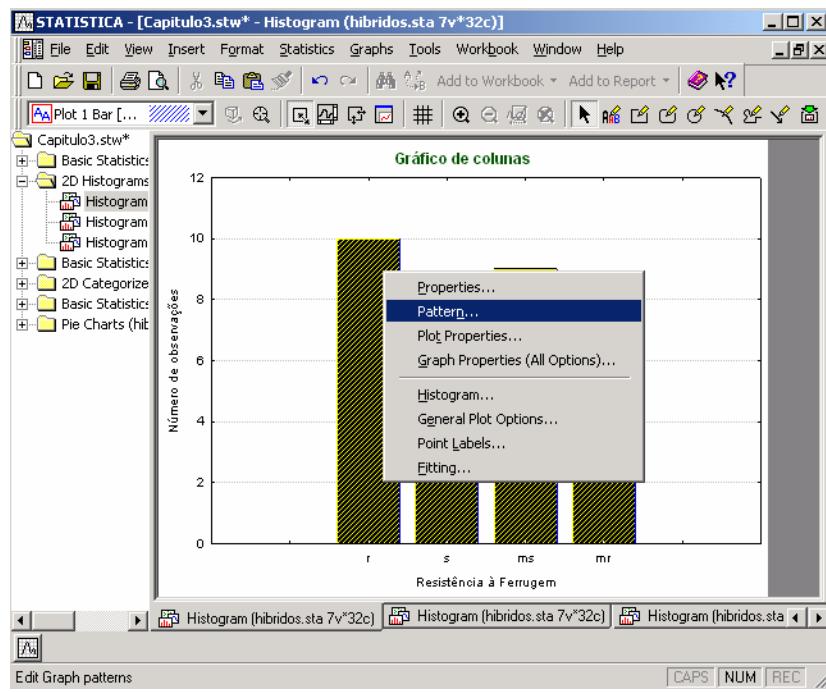


Figura 4.10 – Alteração do o preenchimento da coluna

(8) Escolher a cor e o padrão através da janela apresentada na Figura 4.11:

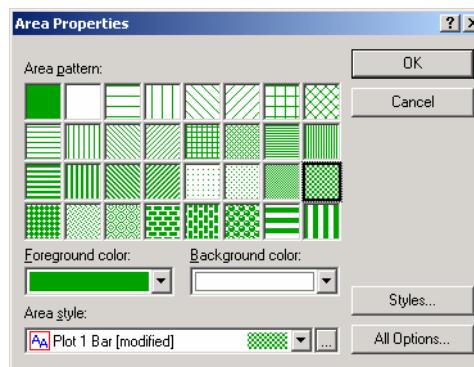


Figura 4.11 – Janela para escolha do preenchimento da coluna

(9) O gráfico resultante está apresentado na Figura 4.12:

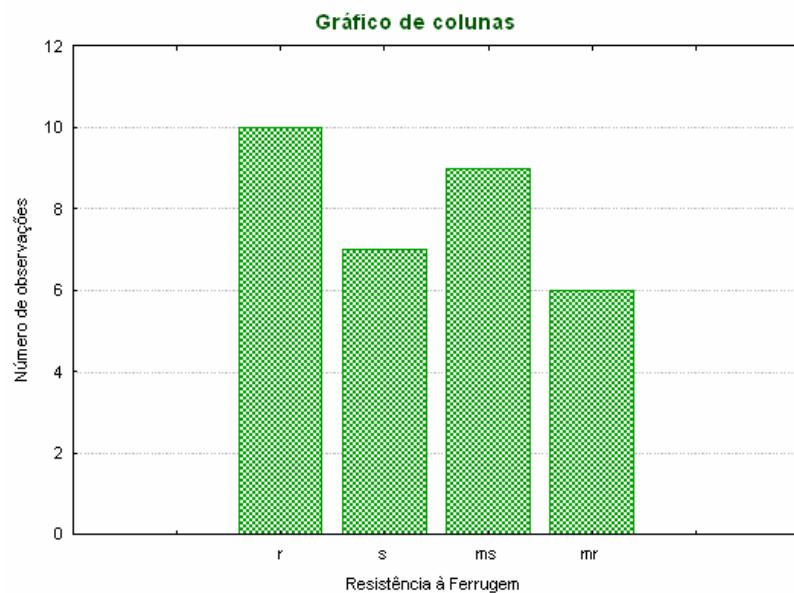


Figura 4.12 – Gráfico de colunas da variável resistência à ferrugem

Observação: Ressalta-se que foram exploradas somente algumas das opções gráficas disponíveis no programa STATISTICA ®, cabendo ao leitor explorar as diversas formas de edição e possíveis padronizações de seus gráficos.

5. TESTES ESTATÍSTICOS PARA COMPARAÇÃO DE DUAS MÉDIAS

5.1. Teste de Duas Médias Populacionais com Variâncias Desconhecidas

- a) Dados pareados ou amostras dependentes
- b) Dados não pareados ou amostras independentes
 - b.1) Variâncias homogêneas (iguais)
 - b.2) Variâncias heterogêneas (desiguais)

O método de análise selecionado para o **teste-t** deve ser previamente estudado para que não implique em resultados falsos. Isto significa que precisamos testar a homogeneidade das variâncias e verificar o planejamento da pesquisa, e baseado nesta conclusão, aplicar então o teste correto, que pode ser:

1. Amostras independentes - variâncias homogêneas - (*T-Test for Independent Samples (Groups)*)
2. Amostras independentes - variâncias heterogêneas - (*T-Test for Independent Samples (Groups)* aplicado o *t-test with separate variances estimates*)
3. Amostras dependentes (Correlacionadas) - (*T-Test for Dependent (Correlated) Samples*)

5.1.1. Amostras Independentes com Variâncias Iguais e Desconhecidas

Exemplo 1: Criar o arquivo **solvente.sta** conforme dados apresentados no Quadro 5.1.

Quadro 5.1 – Resultados das absorbâncias para dois tipos de solventes

Tratamento	Repetição	Absorbância
1	1	0,6286
1	2	0,6143
1	3	0,5826
1	4	0,7498
1	5	0,6060
2	1	0,4748
2	2	0,4321
2	3	0,4309
2	4	0,5010
2	5	0,4094

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 5.1, onde a opção a ser escolhida será *t-test, independent, by groups*:

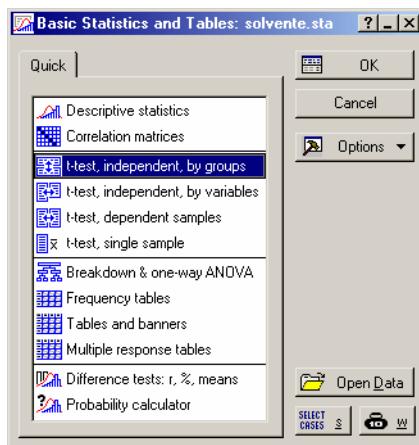


Figura 5.1 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 5.2:

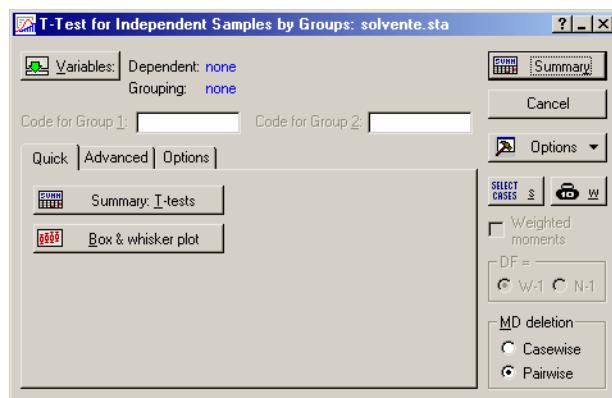


Figura 5.2 – Teste-t para amostras independentes

(3) Clique no botão [Variables] e selecione as variáveis para a análise (*Grouping variable* = Tratamento e *Dependent variable* = Absorbancia), conforme apresentado na Figura 5.3:

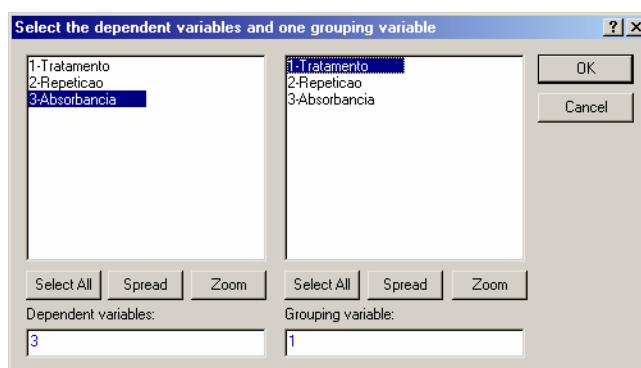


Figura 5.3 – Janela de seleção das variáveis para a análise

(4) Clique no botão [OK] e aparecerá a janela apresentada na Figura 5.4:

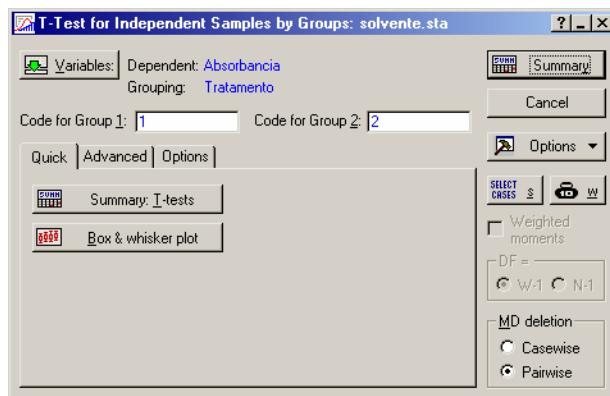


Figura 5.4 – Teste-t para amostras independentes

(5) Clique no botāo [Summary] e o resultado estā apresentado na Figura 5.5:

T-tests; Grouping: Tratamento (solvente.sta)														
Variable	Group 1: 1		Group 2: 2		Mean	t-value	df	p	Valid N	Valid N	Std.Dev.	Std.Dev.	F-ratio	p
	1	2	1	2										
Absorancia	0.636260	0.449640	5.530084	8	0.000554		5	5	0.065626	0.037246	3.104595	0.298289		

Figura 5.5 – Resultado do teste-t para os dois solventes

Interpretação: Da figura anterior conclui-se que:

- Médias amostrais: $\bar{x}_1=0,63626$

$$\bar{x}_2=0,44964$$

- Desvios padrão amostrais: $s_1=0,06563$

$$s_2=0,03725$$

1º) Teste de homogeneidade de variâncias

As hipóteses a serem testadas estāo apresentadas a seguir:

H₀: $\sigma^2_1 = \sigma^2_2 \Rightarrow$ significa que há homogeneidade de variância

H_a: $\sigma^2_1 > \sigma^2_2 \Rightarrow$ significa que não há homogeneidade de variância

Tipo do teste: unilateral

* **Importante:** O STATISTICA ® sempre faz um teste bilateral. No exemplo, sendo um teste unilateral, deve-se dividir o valor p por dois. Assim, $p=0,2983/2=0,1491$. Sendo assim, não se pode rejeitar a hipótese nula e concluir-se que as variâncias podem ser consideradas homogêneas ($p=0,1491$).

2º) Teste de igualdade de médias

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: $\mu_1 = \mu_2 \Rightarrow$ significa que não há diferença entre as médias de absorbância com relação ao tipo de solvente

H_a: $\mu_1 \neq \mu_2 \Rightarrow$ significa que há diferença entre as médias de absorbância com relação ao tipo de solvente

Tipo do teste: bilateral

Da Figura 5.5, $t=5,5301$ ($p=0,0006$), ou seja, é a probabilidade de $t \geq 5,5301$ ocorrer devido ao acaso. Sendo assim, rejeita-se a hipótese nula e conclui-se que existe diferença estatisticamente significativa entre as médias ao nível de significância de 0,06% ($0,0006 \times 100$).

Outra maneira de calcular o valor p:

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 5.6, onde a opção a ser escolhida será *Probability calculator*:

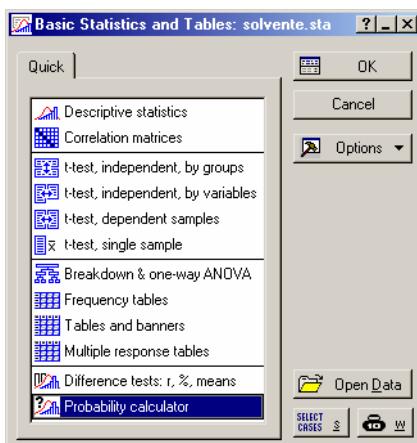


Figura 5.6 – Estatísticas básicas e tabelas

(2) Selecione as seguintes opções:

- *Distribution: t (Student);*
- *Two-tailed* (teste bilateral);
- *(1-Cumulative p);*
- $t=5,530084$ (valor encontrado pelo *t-test for Independent Samples*)
- $df = 8$ (graus de liberdade)

(3) Clique no botão [Compute] e o resultado está apresentado na Figura 5.7:

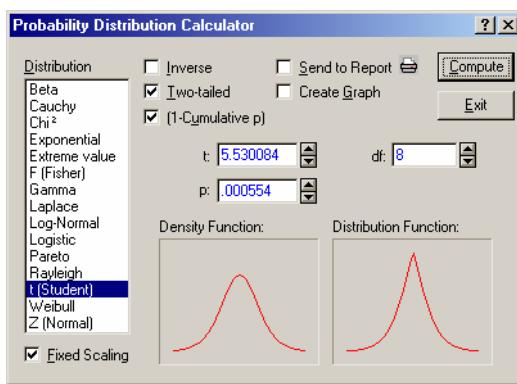


Figura 5.7 – Janela para cálculo do valor p do teste

Note que encontra-se o valor p exatamente igual ao descrito pelo *T-Test for Independent Samples*, conforme apresentado na Figura 5.5.

Diagrama de caixas (BOX PLOT)

(1) Na janela da Figura 5.4, para construir o diagrama de caixas, clique no botão [Box & whisker plot] e o gráfico resultante está apresentado na Figura 5.8:

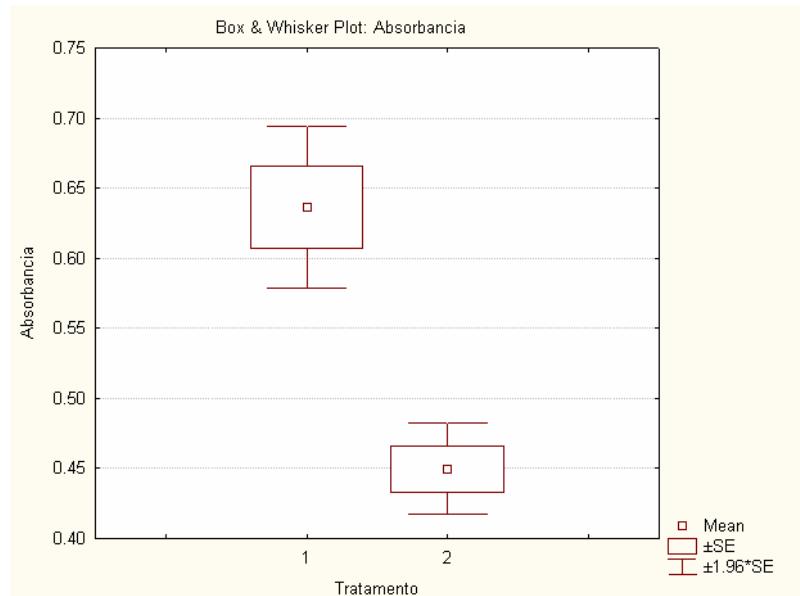


Figura 5.8 – Diagrama de caixas para os dois tipos de solventes

Interpretação: Pela Figura 5.8, pode-se considerar que há homogeneidade de variância e que a média da absorbância no solvente 1 é maior com relação ao solvente 2.

Gráfico normal de probabilidades por tratamento

(1) Na janela da Figura 5.4, clique na aba [Advanced], aparecerá a janela apresentada na Figura 5.9:



Figura 5.9 – Teste-t para amostras independentes

(2) Na janela da Figura 5.9, clique no botão [Categorized normal plot] e o gráfico resultante está apresentado na Figura 5.10:

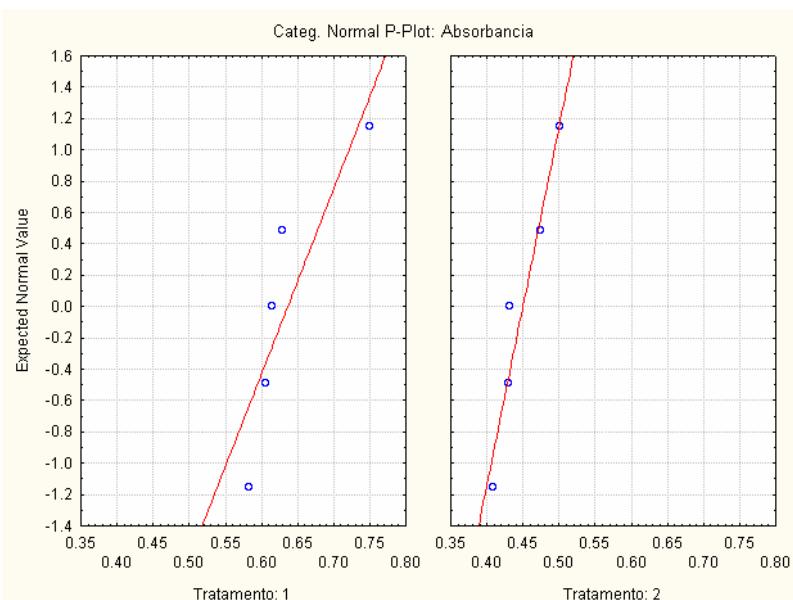


Figura 5.10 – Gráfico normal de probabilidades por tratamento

Exemplo 2: Criar o arquivo **tomate.sta** conforme dados apresentados no Quadro 5.2.

Quadro 5.2 – Resultados de produção de dois tipos de adubos em tomateiro

Tratamento	Repetição	Produção
1	1	29,9
1	2	11,4
1	3	25,3
1	4	16,5
1	5	21,1
2	1	26,6
2	2	23,7
2	3	28,5
2	4	14,2
2	5	17,9
2	6	24,3

Fonte: Box, Hunter & Hunter

(1) Repetir os cinco primeiros passos do **Exemplo 1**, porém selecionando as seguintes variáveis para a análise: *Grouping variable* = Tratamento e *Dependent variable* = Producao. O resultado do teste está apresentado na Figura 5.11:

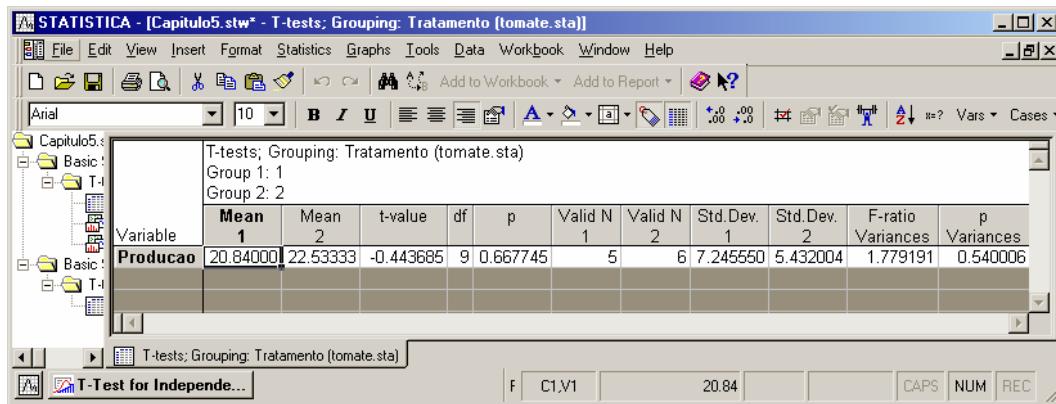


Figura 5.11 – Resultado do teste-t para a produção dos dois diferentes tipos de adubos

1º) Teste de homogeneidade de variâncias

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: $\sigma^2_1 = \sigma^2_2 \Rightarrow$ significa que há homogeneidade de variância

H_a: $\sigma^2_1 > \sigma^2_2 \Rightarrow$ significa que não há homogeneidade de variância

Tipo do teste: unilateral

Da Figura 5.11, $F=1,7792$ ($p=0,5400/2=0,2700$) significa que não se pode rejeitar a hipótese nula, ou seja, as variâncias podem ser consideradas homogêneas.

2º) Teste de igualdade de médias

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: $\mu_1 = \mu_2 \Rightarrow$ significa que não há diferença entre as médias de produção com relação ao tipo de adubo

H_a: $\mu_2 > \mu_1 \Rightarrow$ significa que a média da produção do adubo tipo 1 é menor que a média da produção do adubo tipo 2

Tipo do teste: unilateral

Da Figura 5.5, $F=3,1046$ ($p=0,6677/2=0,3339$) significa que não se pode rejeitar a hipótese nula, ou seja, não existe diferença estatisticamente significativa entre as médias.

5.1.2. Amostras Independentes com Variâncias Desiguais e Desconhecidas

Exemplo 3: Criar o arquivo **variedade.sta** conforme dados apresentados no Quadro 5.3.

Quadro 5.3 – Resultados do rendimento em kg/ha de duas variedades de milho

Varied_A	Varied_B
1300	1800
1350	1600
1250	1900
1400	1850
1200	1750
	2500

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 5.12, onde a opção a ser escolhida será *t-test, independent, by variables*:

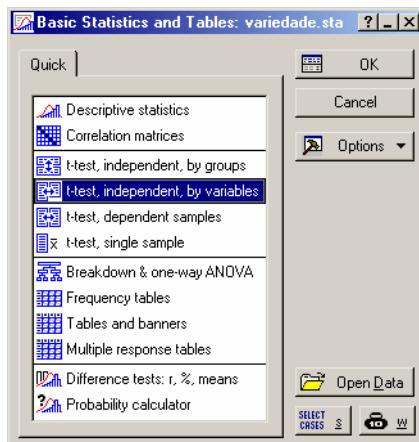


Figura 5.12 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 5.13:

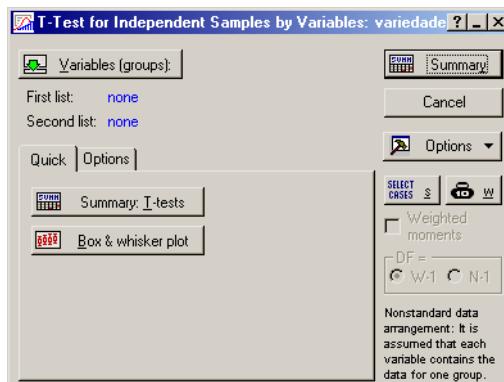


Figura 5.13 – Teste-t para amostras independentes

(3) Clique no botão [Variables] e selecione as variáveis para a análise (*First variable (group) list* - Varied_A e *Second variable (group) list* - Varied_B), conforme apresentado na Figura 5.14:

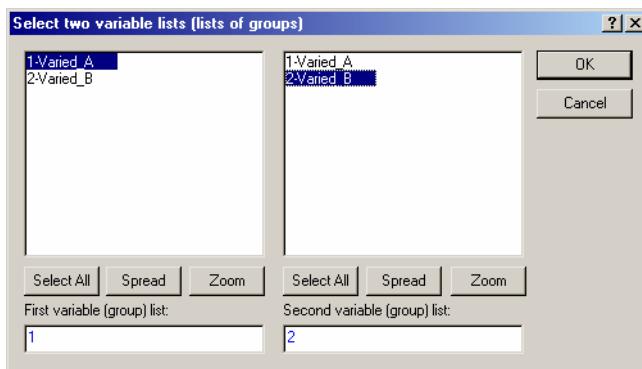


Figura 5.14 – Janela de seleção das variáveis para a análise

(4) Clique no botão [OK] e aparecerá a janela apresentada na Figura 5.15:

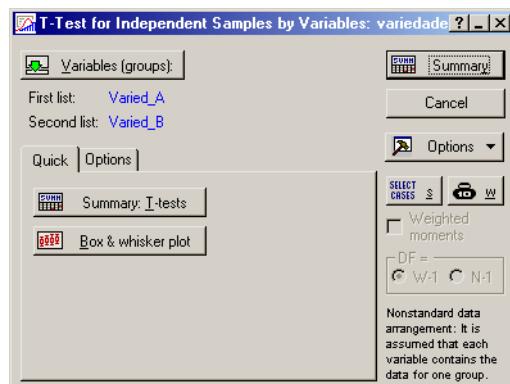


Figura 5.15 – Teste-t para amostras independentes

(5) Clique no botão [Summary] e o resultado está apresentado na Figura 5.16:

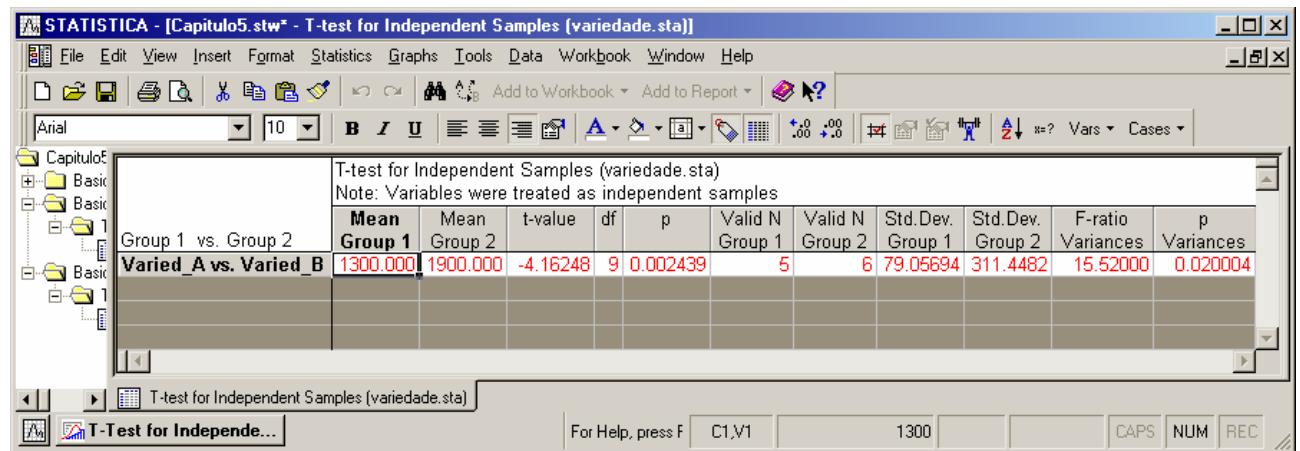


Figura 5.16 – Resultado do teste-t para as duas variedades

1º) Teste de homogeneidade de variâncias

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: $\sigma^2_A = \sigma^2_B \Rightarrow$ significa que há homogeneidade de variância

H_a: $\sigma^2_B > \sigma^2_A \Rightarrow$ significa que não há homogeneidade de variância

Tipo do teste: unilateral

Da Figura 5.16, $F=15,5200$ ($p=0,0200/2=0,0100$) significa que rejeita-se a hipótese nula, ou seja, as variâncias não podem ser consideradas homogêneas.

Neste caso, como as variâncias são heterogêneas, o teste-t deve ser calculado com variâncias separadas.

(6) Na janela da Figura 5.15, clique na aba [Options] e selecione a opção *t-test with separate variance estimates*, conforme apresentado na Figura 5.17:

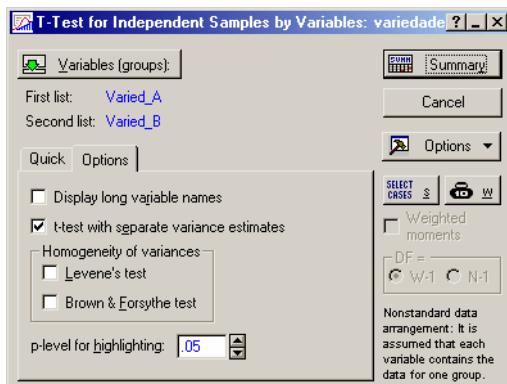


Figura 5.17 – Teste-t para amostras independentes

(7) Clique no botão [Summary] e o resultado está apresentado na Figura 5.18:

T-test for Independent Samples (variedade.sta)							
Note: Variables were treated as independent samples							
	Mean Group 1	Mean Group 2	t-value	df	p	t separ. var.est.	df
Group 1 vs. Group 2							
Varied_A vs. Varied_B	1300.000	1900.000	-4.16248	9	0.002439	-4.54641	5.760043

Figura 5.18 – Resultado do teste-t para as duas variedades com variâncias desiguais

2º) Teste de igualdade de médias

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: $\mu_A = \mu_B \Rightarrow$ significa que não há diferença entre as médias de produção com relação ao tipo de variedade

H_a: $\mu_A \neq \mu_B \Rightarrow$ significa que há diferença entre as médias de produção com relação ao tipo de variedade

Tipo do teste: bilateral

Da Figura 5.18, $t=-4,5464$ ($p=0,0043$), ou seja, rejeita-se a hipótese nula e conclui-se que existe diferença estatisticamente significativa entre as médias de produção com relação ao tipo de variedade.

(8) Na janela da Figura 5.15, para construir o diagrama de caixas, clique no botão [Box & whisker plot] e o gráfico resultante está apresentado na Figura 5.19:

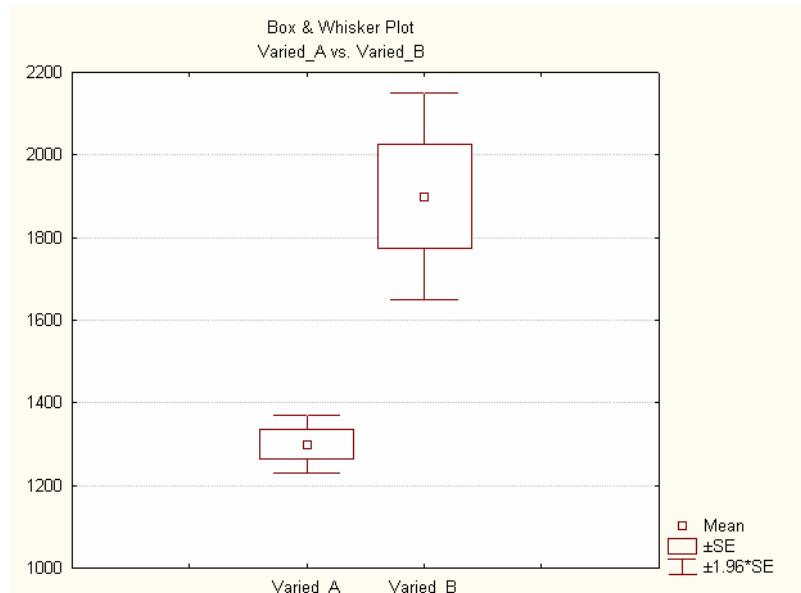


Figura 5.19 – Diagrama de caixas para as duas variedades de milho

Interpretação: Pela Figura 5.19, é visível que a variância da Variedade A é menor que a da Variedade B e que a média na Variedade B é maior com relação à Variedade A.

5.1.3. Amostras Dependentes (Dados Pareados)

Exemplo 4: Criar o arquivo **solas.sta** conforme dados apresentados no Quadro 5.4.

Quadro 5.4 – Resultados do desgaste de solas de sapatos, confeccionadas com dois tipos diferentes de materiais

Criança	Material_A	Material_B
1	13,2	14,0
2	8,2	8,8
3	10,9	11,2
4	14,3	14,2
5	10,7	11,8
6	6,6	6,4
7	9,5	9,8
8	10,8	11,3
9	8,8	9,3
10	13,3	13,6

Neste experimento, cada criança usou um tipo de material diferente de solado em cada um de seus sapatos, caracterizando a dependência entre as medidas, pois numa mesma criança são realizadas duas medidas de desgaste nos dois tipos de materiais.

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: $\mu_A = \mu_B \Rightarrow$ significa que não há diferença entre as médias de desgaste com relação ao tipo de material

$H_a: \mu_{JB} > \mu_A \Rightarrow$ significa que a média de desgaste do material A é menor que a média de desgaste do material B

Tipo do teste: unilateral

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 5.20, onde a opção a ser escolhida será *t-test, dependent samples*:



Figura 5.20 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 5.21:

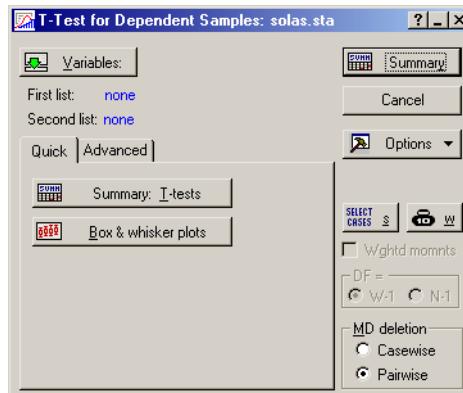


Figura 5.21 – Teste-t para amostras dependentes

(3) Clique no botão [Variables] e selecione as variáveis para a análise (*First variable list* - Material_A e *Second variable list (optional)* - Material_B), conforme apresentado na Figura 5.22:

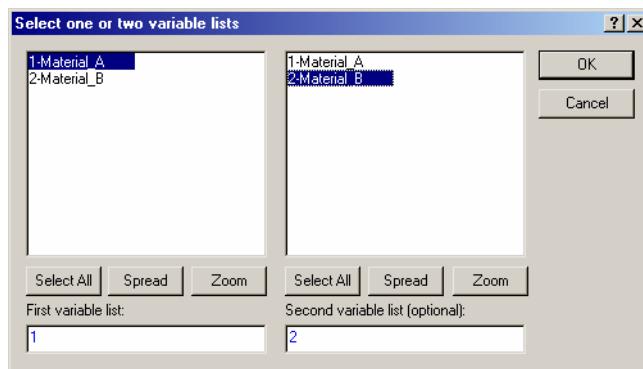


Figura 5.22 – Janela de seleção das variáveis para a análise

(4) Clique no botão [OK] e aparecerá a janela apresentada na Figura 5.23:

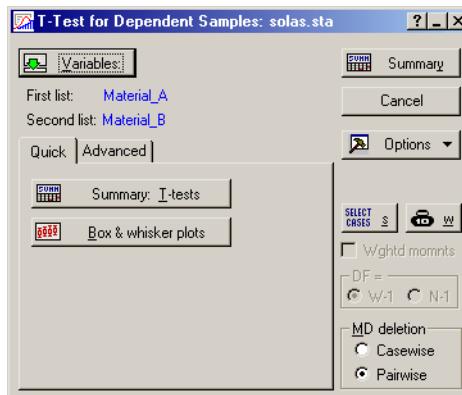


Figura 5.23 – Teste-t para amostras dependentes

(5) Clique no botão [Summary] e o resultado está apresentado na Figura 5.24:

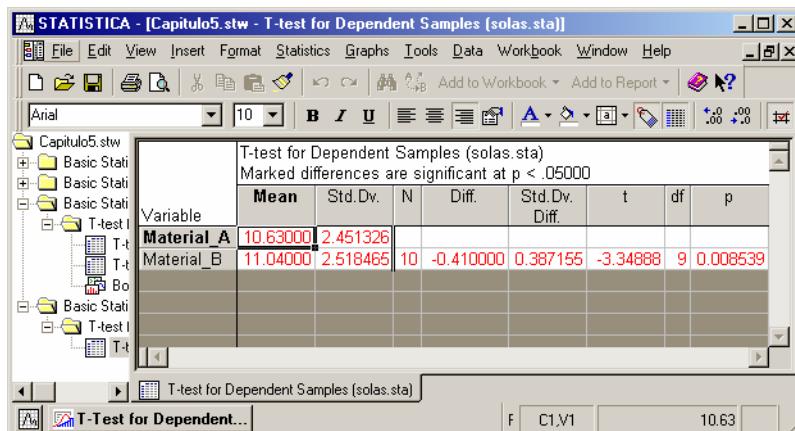


Figura 5.24 – Resultado do teste-t para amostras dependentes

Conclusão: Da Figura 5.24, $t=-3,3489$ ($p=0,0085/2=0,0043$), ou seja, rejeita-se a hipótese nula e conclui-se que a média de desgaste do material A é menor que a média de desgaste do material B.

Selecionando o método incorreto

- (1) Refazer o teste anterior da mesma forma que no Exemplo 3, escolhendo a opção para amostras independentes (*t-test, independent, by variables*). Selecione as variáveis conforme mostrado na Figura 5.25:

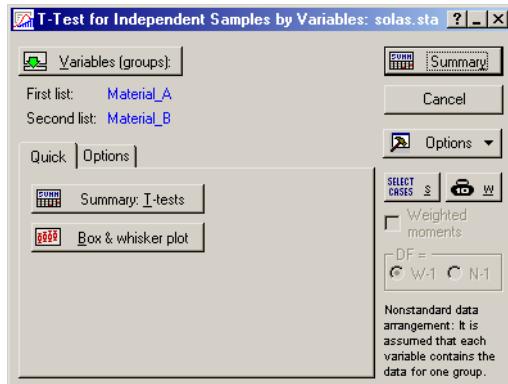


Figura 5.25 – Teste-*t* para amostras independentes

- (2) Clique no botão [Summary] e o resultado está apresentado na Figura 5.26:

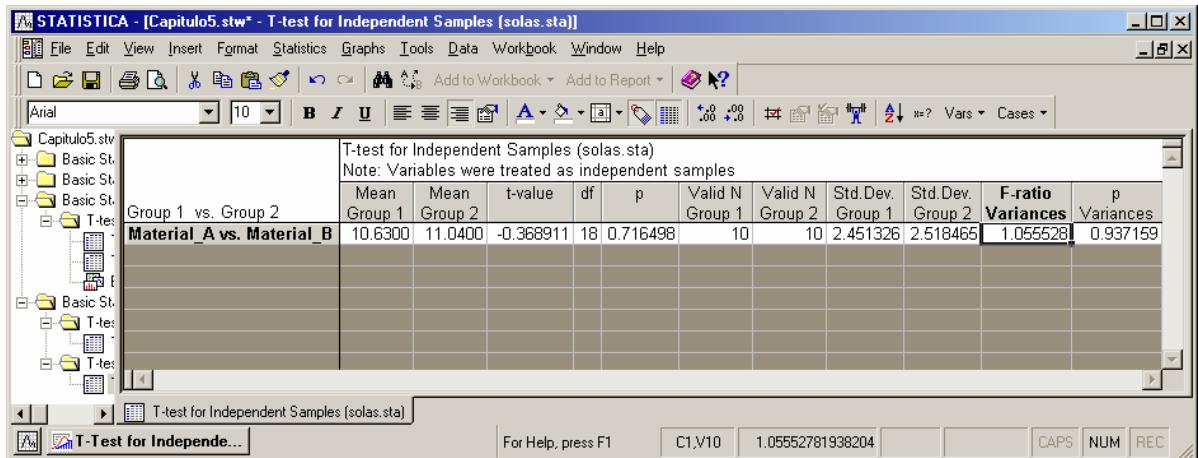


Figura 5.26 – Resultado do teste-*t* para amostras independentes

1º) Teste de homogeneidade de variâncias

Da Figura 5.26, $F=1,0555$ ($p=0,9372/2=0,4686$) significa que não se pode rejeitar a hipótese nula, ou seja, as variâncias podem ser consideradas homogêneas.

2º) Teste de igualdade de médias

Da Figura 5.26, $t=-0,3689$ ($p=0,7165/2=0,3582$) significa que não se pode rejeitar a hipótese nula, ou seja, a média de desgaste do material A não é menor que a média de desgaste do material B.

Sendo assim, não se rejeita a hipótese nula, só que, ESTA CONCLUSÃO NÃO É VERDADEIRA. Por isto, estudamos anteriormente o delineamento da pesquisa.

A seguir, o gráfico mostra como as médias estão próximas, pois aqui não foi retirado o efeito de meninos.

(3) Na janela da Figura 5.25, para construir o diagrama de caixas, clique no botão [Box & whisker plot] e o gráfico resultante está apresentado na Figura 5.27:

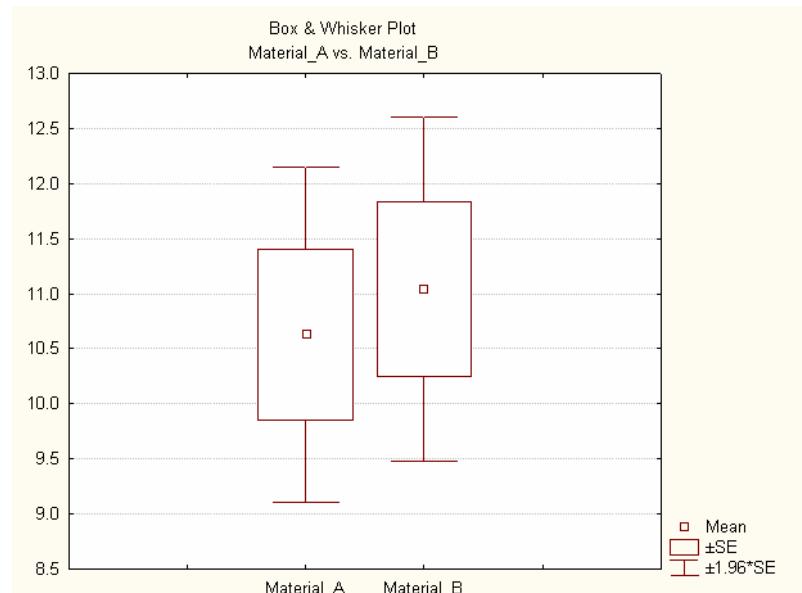


Figura 5.27 – Diagrama de caixas para os dois tipos de materiais

5.1.4. Intervalo de Confiança para a Diferença de Duas Médias no Caso de Dados Pareados

(1) Criar uma nova variável (DifB_A) no arquivo **solas.sta**, que representa a diferença entre os valores de desgastes dos dois materiais. Para isso, clique com o botão da direita em cima da variável Material_B, selecione a opção [Add variable] e defina a nova variável conforme apresentado na Figura 5.28:

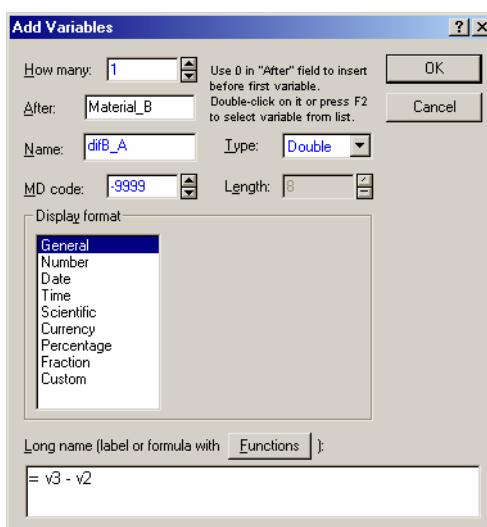


Figura 5.28 – Janela de especificação da variável DifB_A

(2) Clique no botāo [OK] e se aparecer a expressāo “*Expression OK. Recalculate the variable now?*” clique no botāo [SIM]. O arquivo resultante estā apresentado no Quadro 5.5:

Quadro 5.5 – Resultados do desgaste de solas de sapatos, confeccionadas com dois tipos diferentes de materiais

Criança	Material_A	Material_B	DifB_A
1	13,2	14,0	0,8
2	8,2	8,8	0,6
3	10,9	11,2	0,3
4	14,3	14,2	-0,1
5	10,7	11,8	1,1
6	6,6	6,4	-0,2
7	9,5	9,8	0,3
8	10,8	11,3	0,5
9	8,8	9,3	0,5
10	13,3	13,6	0,3

(3) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 5.29, onde a opção a ser escolhida será *Descriptive statistics*:

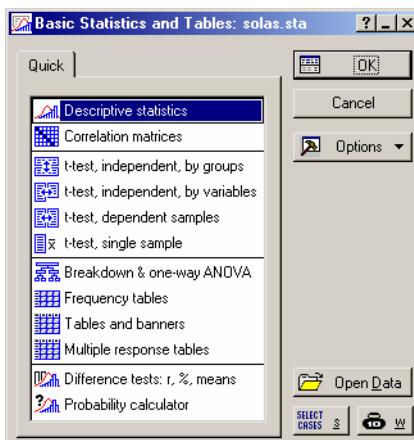


Figura 5.29 – Estatísticas básicas e tabelas

(4) Clique no botāo [OK], selecione a nova variável criada (DifB_A) clicando no botāo [Variables] e na aba [Advanced] selecione as opções conforme apresentado na Figura 5.30:

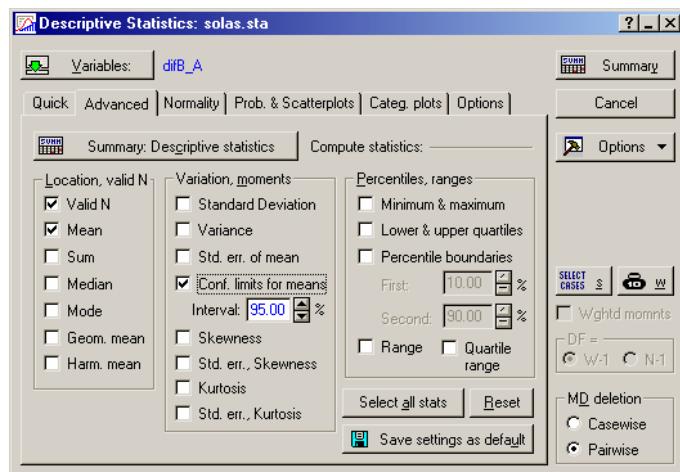


Figura 5.30 – Estatísticas descritivas

(5) Clique no botão [Summary] e o resultado está apresentado na Figura 5.31:

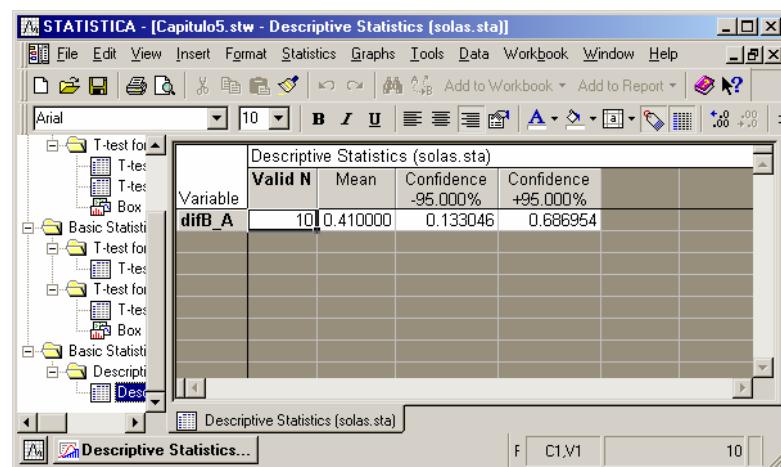


Figura 5.31 – Resultado do intervalo de confiança para a variável DifB_A

Interpretação: Da Figura 5.31 estimasse que a verdadeira diferença média do desgaste das solas entre os dois materiais está no intervalo $0,1330 \leq \mu \leq 0,6870$

6. TESTE PARA DUAS OU MAIS PROPORÇÕES

6.1. O Teste do Qui-Quadrado (CHI-SQUARE)

O exemplo trata de um experimento fatorial 2^2 (dois fatores com dois níveis cada) que são: duas épocas de plantio (fora e na primavera) e duas formas de corte (longo e curto). A resposta é a sobrevivência dos enxertos dada em contagens.

Exemplo 1: Criar o arquivo **ameixeira.sta** conforme dados apresentados no Quadro 6.1.

Quadro 6.1 – Dados de sobrevivência de enxertos de ameixeira

Época	Forma	Sobrevive	Observação	Tratamento
fora	longo	sim	156	fl
fora	curto	sim	107	fc
na	longo	sim	84	nl
na	curto	sim	31	nc
fora	longo	não	84	fl
fora	curto	não	133	fc
na	longo	não	156	nl
na	curto	não	209	nc

Nesta seção realizaremos cinco análises:

1º) Verificar se existe diferença significativa entre os quatro tratamentos.

Observação: Se for constatado que não existe diferença, encerra-se a análise por aqui. Caso exista diferença, prosseguir com a seguinte análise.

2º) Testar se a interação é significativa.

Observação: Se a interação for significativa, far-se-á o desdobramento da mesma. Caso contrário, prosseguir com a seguinte análise.

3º) Testar se existe efeito de época.

4º) Testar se existe efeito de forma.

5º) Supor que a interação é significativa e testar os efeitos (desdobramentos). Forma (Época) - forma dentro de época e Época (Forma) - época dentro de forma.

1º teste: Teste do efeito dos quatro tratamentos

- Fora/Longo (fl)
- Fora/Curto (fc)
- Na/Longo (nl)
- Na/Curto (nc)

(1) No menu [STATISTICS], escolha a opção [Basics Statistics/Tables], aparecerá a janela apresentada na Figura 6.1, onde a opção a ser escolhida será *Tables and banners*:

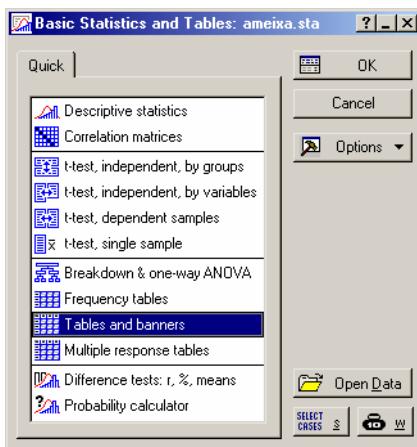


Figura 6.1 – Estatísticas básicas e tabelas

(2) Clique no botão [OK] e aparecerá a janela apresentada na Figura 6.2:

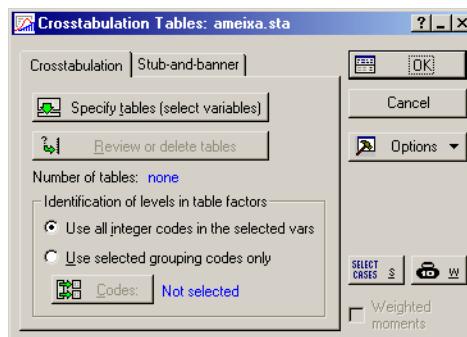


Figura 6.2 – Janela para especificação da tabela de contingência

(3) Clique no botão [Specify Tables (select variables)]. Como a tabela é tridimensional, implica que devemos escolher três variáveis conforme apresentado na Figura 6.3:

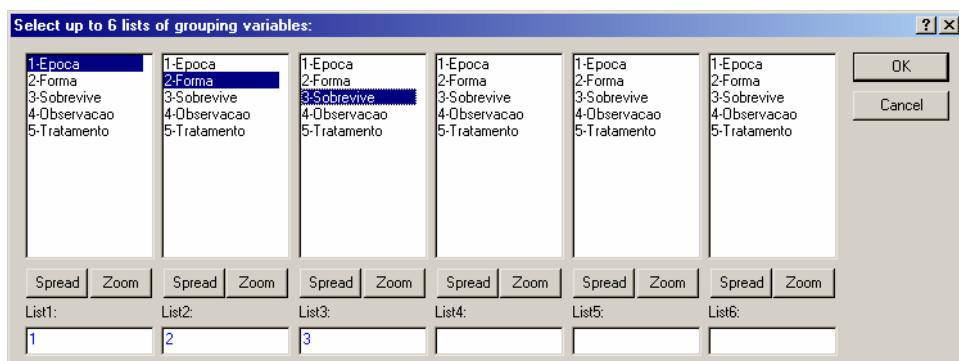


Figura 6.3 – Janela com as variáveis selecionadas para a tabela tridimensional

(4) Clicar no botão [W] (weight = peso ou ponderação) existente no canto inferior direito da janela da Figura 6.2 e defina as opções conforme apresentado na Figura 6.4;

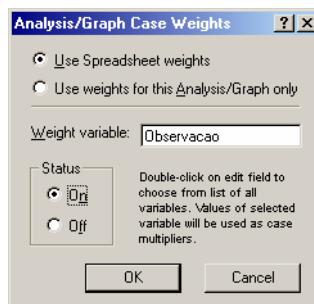


Figura 6.4 – Janela para definição da ponderação

(5) Clique nos botões [OK] da janela anterior e da Figura 6.2 e aparecerá a janela apresentada na Figura 6.5:

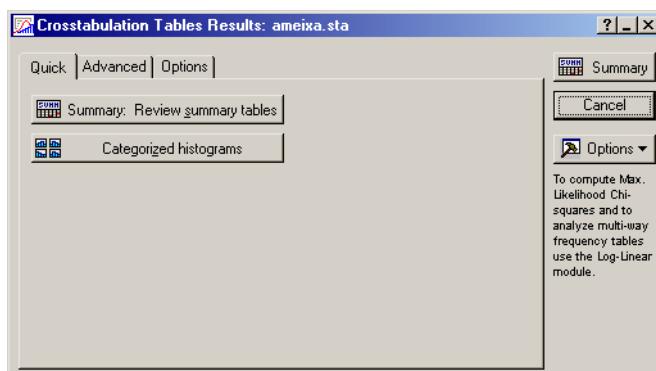


Figura 6.5 – Janela para a visualização do resultado

(6) Clique no botão [Summary] e o resultado está apresentado na Figura 6.6:

Epoca	Forma	Sobrevisa sim	Sobrevisa não	Row Totals
fora	longo	156	84	240
	curto	107	133	240
Total		263	217	480
na	longo	84	156	240
	curto	31	209	240
Total		115	365	480
Column Total		378	582	960

Figura 6.6 – Tabela de contingência da sobrevida dos enxertos

Observação: Como o STATISTICA ® não calcula a estatística qui-quadrado de uma tabela com três variáveis. Para contornar este problema foi criada a variável Tratamento. Sendo assim, agora, vamos especificar uma nova tabela de contingência.

(7) Retorne a janela de especificação das variáveis (item 3) e selecione as mesmas, conforme apresentado na Figura 6.7:



Figura 6.7 – Janela com as variáveis selecionadas para a tabela bidimensional

(8) Clique nos botões [OK] da janela anterior e da Figura 6.2, sendo que na janela que aparecer, clique na aba [Options] e selecione as opções conforme apresentado na Figura 6.8:

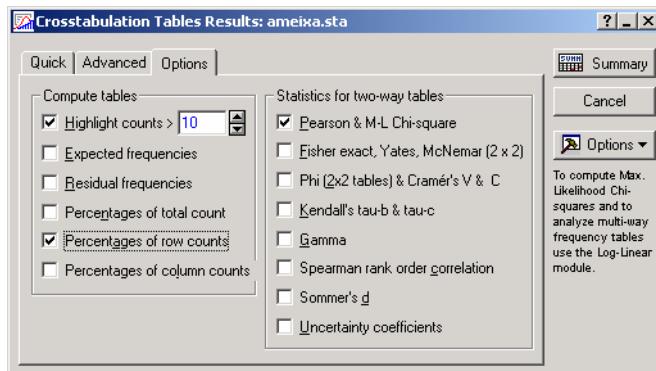


Figura 6.8 – Janela para a visualização do resultado

(9) Clique na aba [Advanced] da janela anterior conforme apresentado na Figura 6.9:

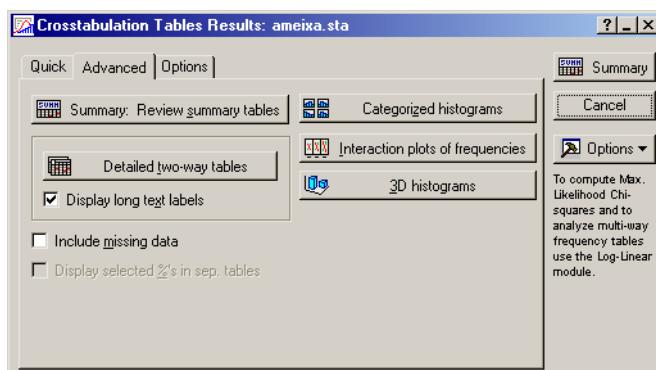


Figura 6.9 – Janela para a visualização do resultado

(10) Clique no botão [Detailed two-way tables] e os resultados estão apresentados nas figuras 6.10 e 6.11:

2-Way Summary Table: Observed Frequencies (ameixa.sta)

Marked cells have counts > 10

Tratamento	Sobrevive		Row Totals
	sim	não	
fl	156	84	240
Row %	65,00%	35,00%	
fc	107	133	240
Row %	44,58%	55,42%	
nl	84	156	240
Row %	35,00%	65,00%	
nc	31	209	240
Row %	12,92%	87,08%	
Totals	378	582	960

Figura 6.10 – Tabela de contingência da sobrevivência dos enxertos

Statistics: Tratamento(4) x Sobrevive(2) (ameixa.sta)

Statistic	Chi-square	df	p
Pearson Chi-square	141,0527	df=3	p=0,0000
M-L Chi-square	151,0193	df=3	p=0,0000

Figura 6.11 – Resultados do teste de qui-quadrado para os tratamentos

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: a proporção de sobreviventes é a mesma para os quatro tratamentos

H_a: a proporção de sobreviventes não é a mesma para os quatro tratamentos

Tipo do teste: bilateral

Interpretação: considerando a estatística qui-quadrado de Pearson, cujo valor foi de 141,05 (p<0,0001), rejeita-se a hipótese nula, ou seja, existe diferença estatisticamente significativa entre as proporções de sobreviventes com relação aos quatro tratamentos.

2º teste: Teste do efeito da Interação

Para testar esse efeito, há necessidade de criar a variável Interação, conforme apresentado no Quadro 6.2:

Quadro 6.2 – Critério para criação da variável interação

Época	Forma	Interação Época*Forma
fora	+1	longo +1
fora	+1	curto -1
na	-1	longo +1
na	-1	curto -1
fora	+1	longo +1
fora	+1	curto -1
na	-1	longo +1
na	-1	curto -1

(1) Crie a nova variável (Interacao) no arquivo **ameixa.sta**, conforme apresentado na Figura 6.12 e preencha conforme os valores apresentados na terceira coluna do Quadro 6.2:

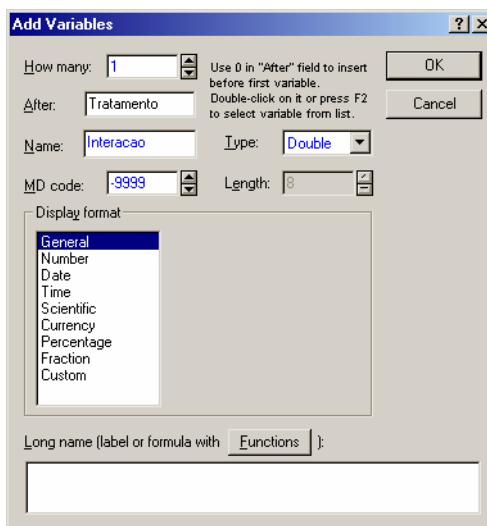


Figura 6.12 – Janela de criação da variável Interacao

(2) Repita os três primeiros itens do 1º teste, selecionando as variáveis, conforme apresentado na Figura 6.13:

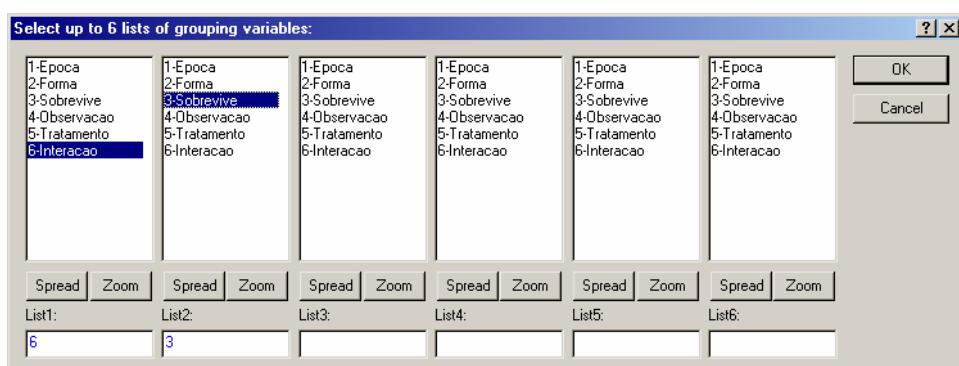


Figura 6.13 – Janela com as variáveis selecionadas para a tabela bidimensional

(3) Clique nos botões [OK] da janela anterior e da Figura 6.2, sendo que na janela que aparecer, clique na aba [Options] e selecione as opções conforme apresentado na Figura 6.14:

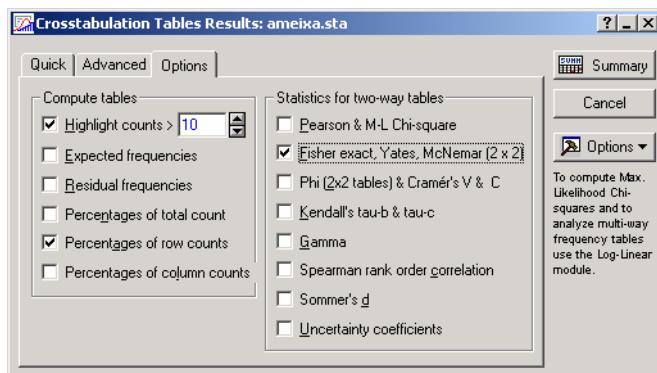


Figura 6.14 – Janela para a visualização do resultado

Observação: A tabela resultante será 2x2 e neste caso recomenda-se utilizar a correção de continuidade de Yates.

(4) Clique na aba [Advanced] da janela anterior, no botão [Detailed two-way tables] e os resultados estão apresentados nas figuras 6.15 e 6.16:

2-Way Summary Table: Observed Frequencies (ameixa.sta)			
Interacao	Sobrevive sim		Row Totals
	191	289	
-1	39,79%	60,21%	
1	38,96%	61,04%	
Totals	378	582	960

Figura 6.15 – Tabela de contingência da sobrevivência dos enxertos

Statistics: Interacao(2) x Sobrevive(2) (ameixa.sta)			
Statistic	Chi-square	df	p
Pearson Chi-square	,0698195	df=1	p=.79160
M-L Chi-square	,0698205	df=1	p=.79160
Yates Chi-square	,0392734	df=1	p=.84291
Fisher exact, one-tailed			-----
two-tailed			
McNemar Chi-square (A/D)	21,07645	df=1	p=.00000
(B/C)	21,43067	df=1	p=.00000

Figura 6.16 – Resultados do teste de qui-quadrado para a Interação

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: o efeito interação não é significativo

H_a: o efeito interação é significativo

Tipo do teste: bilateral

Interpretação: considerando a estatística qui-quadrado com correção de Yates, cujo valor foi de 0,0393 ($p=0,8429$), não rejeita a hipótese nula, ou seja, o efeito interação não é significativo.

Análise Gráfica

A seguir, será construído um gráfico que permite verificar que não existe interação entre época e forma.

(1) Repita os três primeiros itens do 1º teste, selecionando as variáveis, conforme apresentado na Figura 6.17:

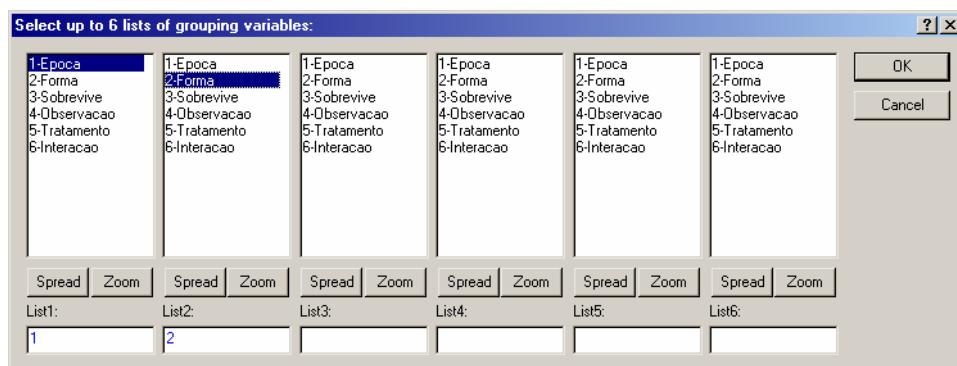


Figura 6.17 – Janela com as variáveis selecionadas para a tabela bidimensional

(2) Clicar no botão [Select cases] existente no canto inferior direito da janela da Figura 6.2 e defina as opções conforme apresentado na Figura 6.18:

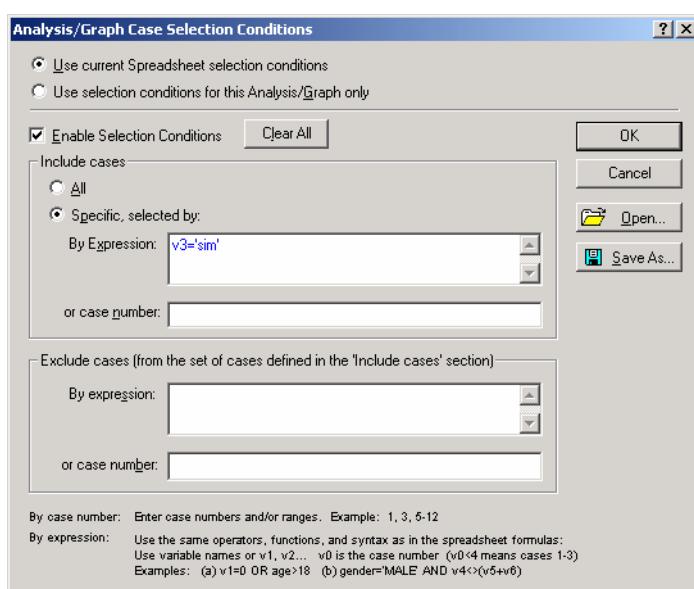


Figura 6.18 – Janela para especificação da condição selecionada

(3) Clique nos botões [OK] da janela anterior e da Figura 6.2. Na janela que aparecer, clique na aba [Advanced] e no botão [Interaction plots of frequencies] e o gráfico resultante está apresentado na Figura 6.19:

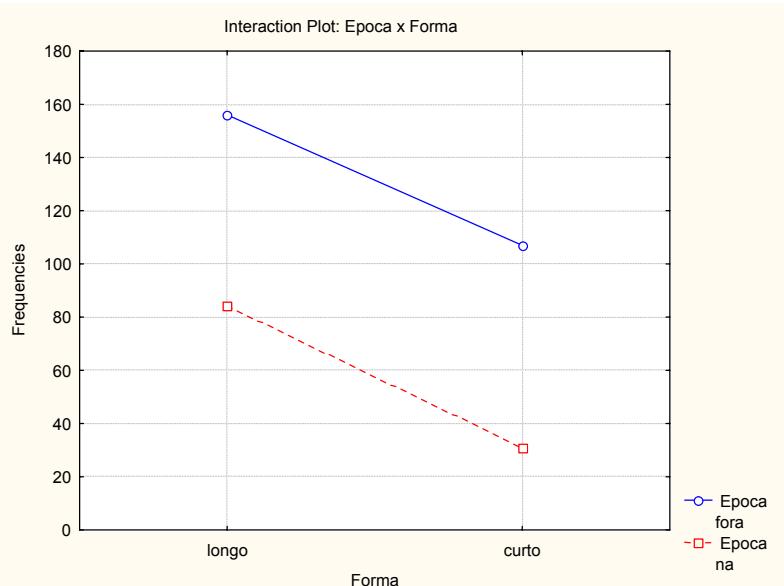


Figura 6.19 – Gráfico da interação forma de corte e época de plantio para sobreviventes.

Interpretação: dado que as linhas do gráfico não se cruzam, conclui-se que não há interação entre os tratamentos.

3º teste: Teste do efeito Época

(1) Repita os três primeiros itens do 1º teste, selecionando as variáveis, conforme apresentado na Figura 6.20:

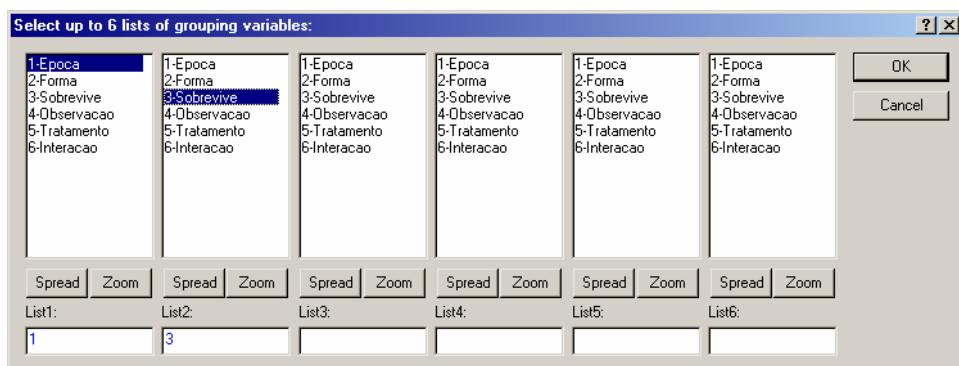


Figura 6.20 – Janela com as variáveis selecionadas para a tabela bidimensional

(2) Clique no botão [Select Case] e desative a condição de seleção, conforme apresentado na Figura 6.21:

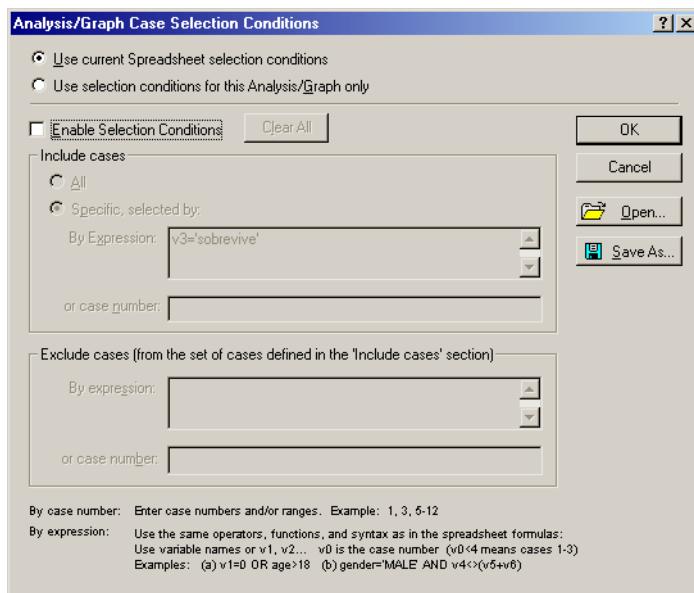


Figura 6.21 – Janela para desabilitar a seleção

(3) Clique nos botões [OK] da janela anterior e da Figura 6.2, sendo que na janela que aparecer, clique na aba [Options] e selecione as opções conforme apresentado na Figura 6.22:

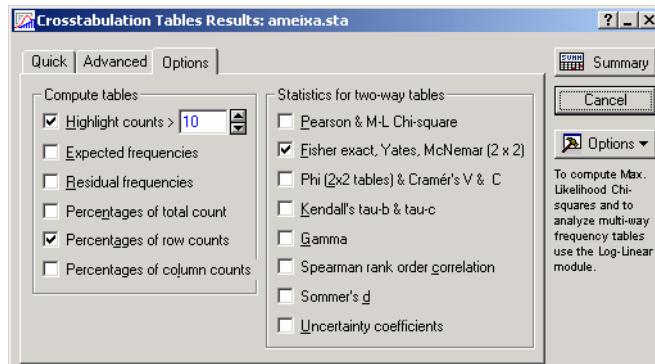


Figura 6.22 – Janela para a visualização do resultado

(4) Clique na aba [Advanced] da janela anterior, no botão [Detailed two-way tables] e os resultados estão apresentados nas figuras 6.23 e 6.24:

		Sobrevida		Row	Totals
		sim	não		
Epoca	fora	263	217	480	
	na	115	365	480	
Row %	fora	54,79%	45,21%		
	na	23,96%	76,04%		
Totals		378	582	960	

Figura 6.23 – Tabela de contingência da sobrevivência dos enxertos

Statistic	Chi-square	df	p
Pearson Chi-square	95,58283	df=1	p=0,0000
M-L Chi-square	97,57890	df=1	p=0,0000
Yates Chi-square	94,29553	df=1	p=0,0000
Fisher exact, one-tailed			----
two-tailed			
McNemar Chi-square (AVD)	16,24363	df=1	p=.00006
(B/C)	30,72590	df=1	p=.00000

Figura 6.24 – Resultados do teste de qui-quadrado para efeito Época

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: a proporção de sobreviventes na e fora da primavera é a mesma.

H_a: a proporção de sobreviventes na e fora da primavera é diferente.

Tipo do teste: bilateral

Interpretação: considerando a estatística qui-quadrado com correção de Yates, cujo valor foi de 94,2955 ($p<0,0001$), rejeita-se a hipótese nula, ou seja, a melhor época de plantio é fora da primavera (54,79% de sobreviventes), conforme pode ser visto na Figura 6.23.

4º teste: Teste do efeito Forma

(1) Repetir o procedimento do teste para o efeito Época (3º teste), selecionando as variáveis, conforme apresentado na Figura 6.25:

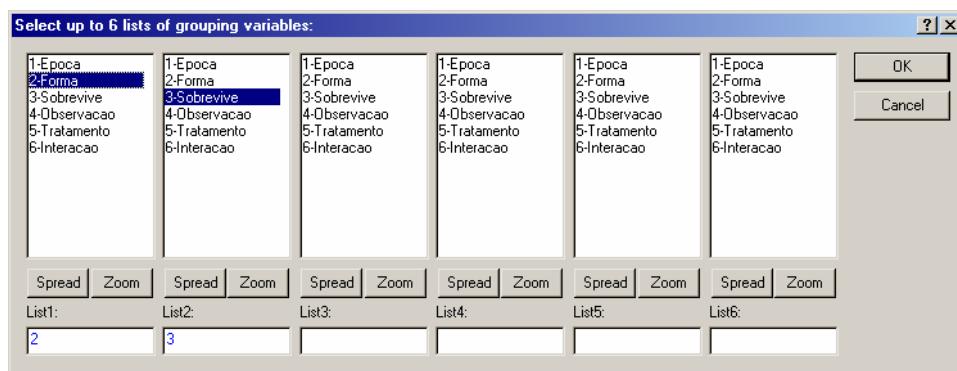


Figura 6.25 – Janela com as variáveis selecionadas para a tabela bidimensional

(2) Os resultados estão apresentados nas figuras 6.26 e 6.27:

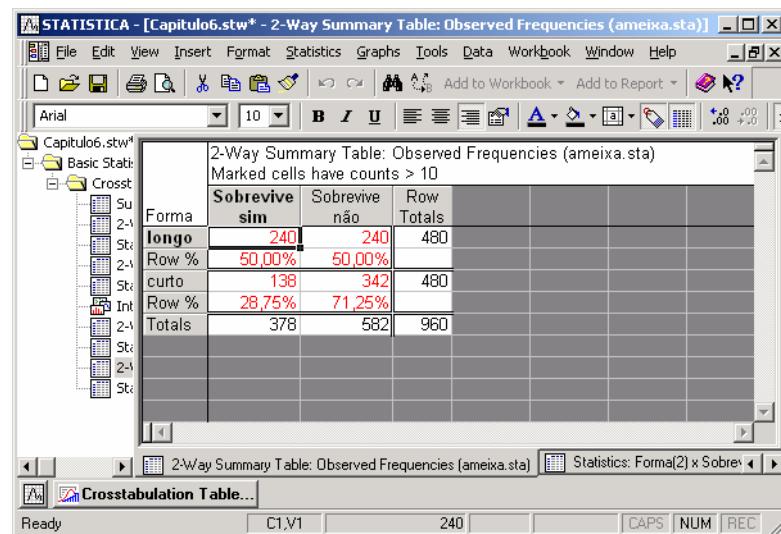


Figura 6.26 – Tabela de contingência da sobrevivência dos enxertos

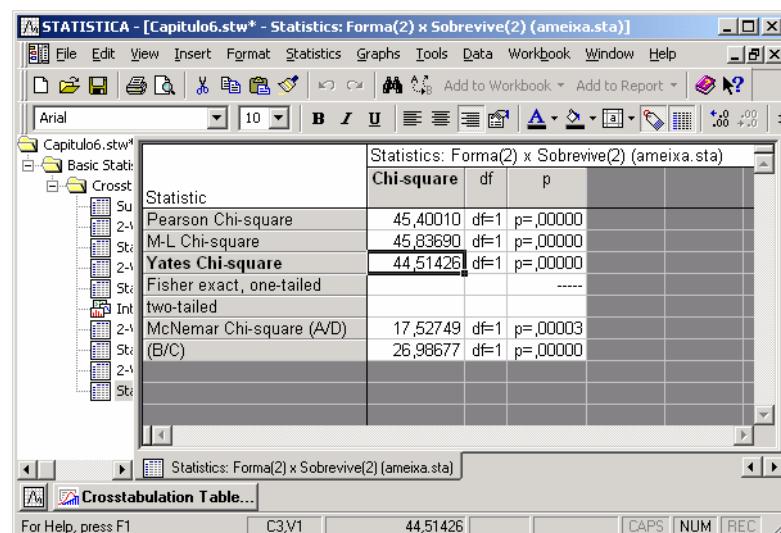


Figura 6.27 – Resultados do teste de qui-quadrado para efeito Forma

As hipóteses a serem testadas estão apresentadas a seguir:

H₀: a proporção de sobreviventes entre cortes longos e curtos é a mesma.

H_a: a proporção de sobreviventes entre cortes longos e curtos é diferente.

Tipo do teste: bilateral

Interpretação: considerando a estatística qui-quadrado com correção de Yates, cujo valor foi de 44,5143 ($p<0,0001$), rejeita-se a hipótese nula, ou seja, a melhor forma de corte é o longo (50,00% de sobreviventes), conforme pode ser visto na Figura 6.26.

5º teste: Estudo da interação Época*Forma (Desdobramentos)

Existem dois casos:

A. Época (Forma) - Época dentro de Forma:

- Época (Forma = longo)
- Época (Forma = curto)

(1) Repita os três primeiros itens do 1º teste, selecionando as variáveis, conforme apresentado na Figura 6.28:

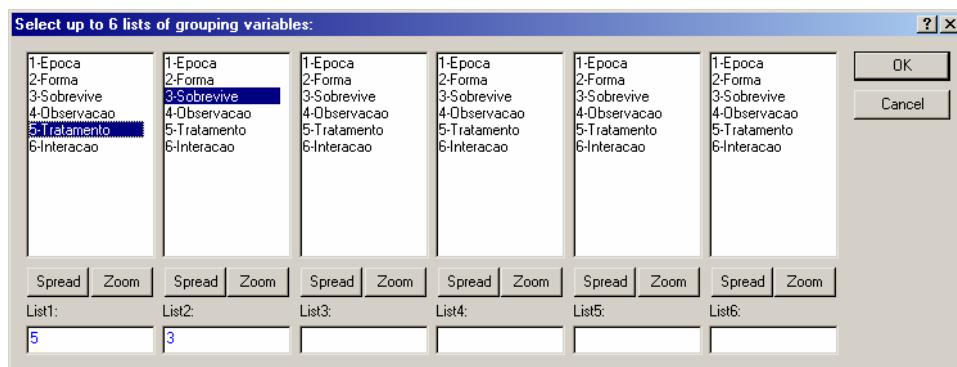


Figura 6.28 – Janela com as variáveis selecionadas para a tabela bidimensional

(2) Clicar no botão [Select cases] existente no canto inferior direito da janela da Figura 6.2 e defina as opções conforme apresentado na Figura 6.29, o filtro é construído para estudar a forma longo:

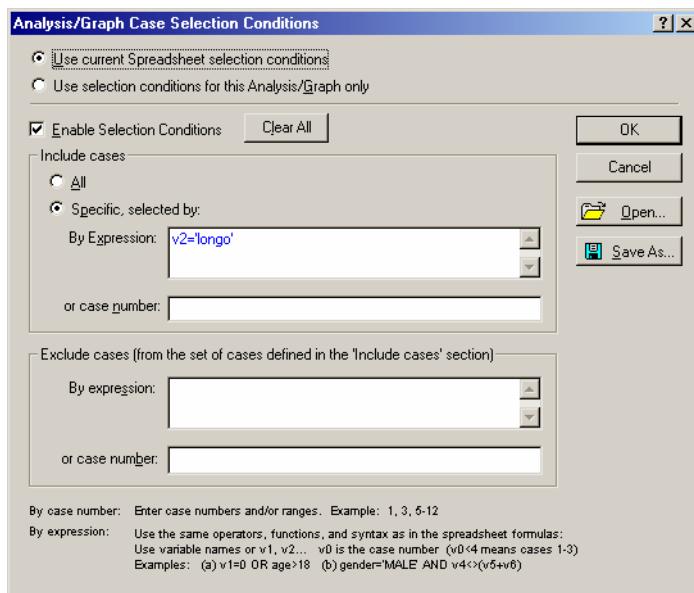


Figura 6.29 – Janela para especificação da condição selecionada

(3) Clique nos botões [OK] da janela anterior e da Figura 6.2, sendo que na janela que aparecer, clique na aba [Options] e selecione as opções *Fisher exact*, *Yates*, *McNemar (2 x 2 Tables)* e *Percentages of column count*, conforme apresentado na Figura 6.22.

(4) Clique na aba [Advanced] da janela da Figura 6.22, no botão [Detailed two-way tables] e os resultados estão apresentados nas figuras 6.30 e 6.31:

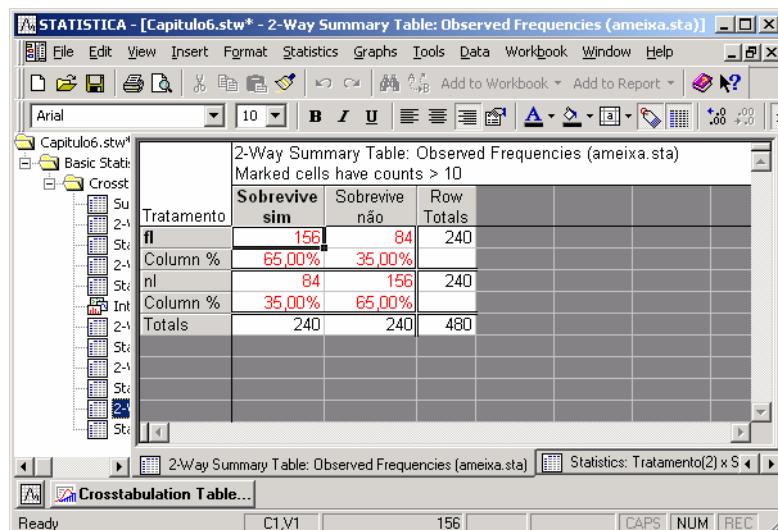


Figura 6.30 – Tabela de contingência da sobrevivência dos enxertos

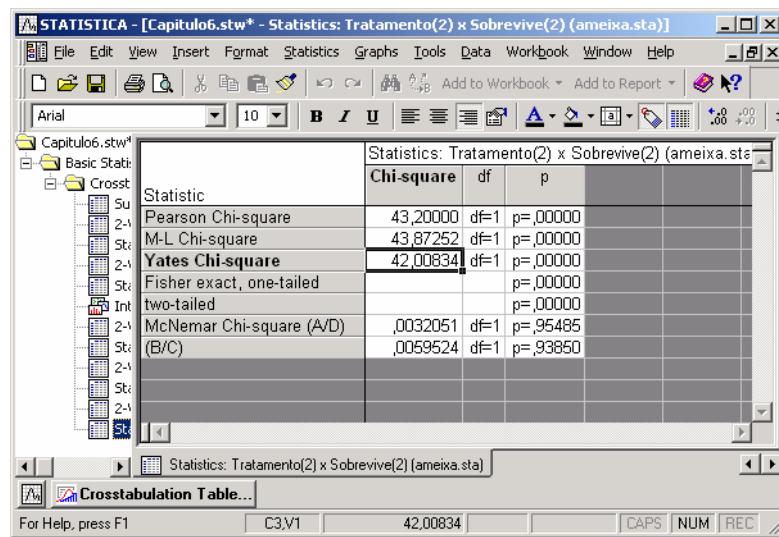


Figura 6.31 – Resultados do teste de qui-quadrado para o efeito da interação da época dentro da forma longo

Interpretação: considerando a estatística qui-quadrado com correção de Yates, cujo valor foi de 42,0083 (p<0,0001), a melhor época de plantio, dentro da forma longo, é fora da primavera (65,00% dos sobreviventes), conforme pode ser visto na Figura 6.30.

(5) Repetir os quatro primeiros itens, sendo que o filtro é construído para estudar a forma curto, conforme apresentado na Figura 6.32. Os resultados estão apresentados nas figuras 6.33 e 6.34:

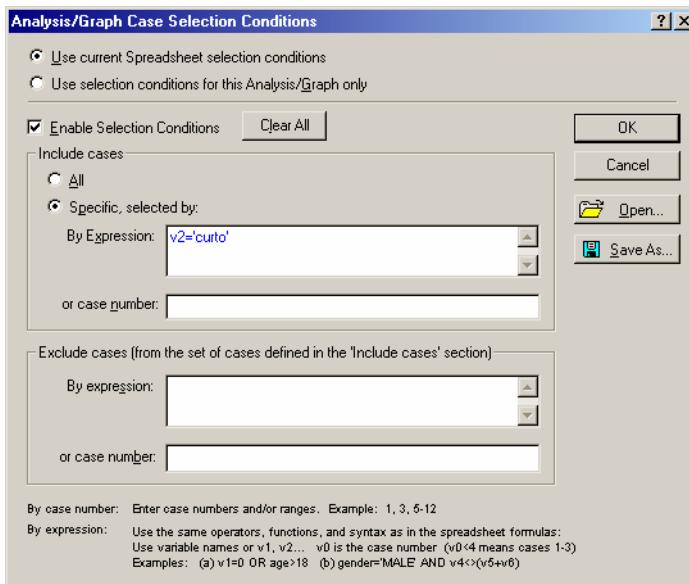


Figura 6.32 – Janela para especificação da condição selecionada

		2-Way Summary Table: Observed Frequencies (ameixa.sta)	
		Sobrevive sim	Sobrevive não
Tratamento	fc	107	133
	Column %	77,54%	22,46%
Tratamento	nc	31	209
	Column %	61,11%	22,46%
Totals		138	342
		240	
		480	

Figura 6.33 – Tabela de contingência da sobrevivência dos enxertos

Statistic	Chi-square	df	p
Pearson Chi-square	58,74396	df=1	p=.00000
M-L Chi-square	61,30989	df=1	p=.00000
Yates Chi-square	57,20824	df=1	p=.00000
Fisher exact, one-tailed			p=.00000
two-tailed			p=.00000
McNemar Chi-square (A/D)	32,28165	df=1	p=.00000
(B/C)	62,20122	df=1	p=.00000

Figura 6.34 – Resultados do teste de qui-quadrado para o efeito da interação da época dentro da forma curto

Interpretação: considerando a estatística qui-quadrado com correção de Yates, cujo valor foi de 50,2081 ($p<0,0001$), a melhor época de plantio, dentro da forma curto, é fora da primavera (77,54% dos sobreviventes), conforme pode ser visto na Figura 6.33.

Observação: note que independente da forma, a melhor época de plantio é fora da primavera, confirmando a não existência do efeito interação.

B. Forma (Época) - Forma dentro de Época:

- Forma (Época = na)
- Forma (Época = fora)

(6) Repetir os quatro primeiros itens, sendo que o filtro é construído para estudar a época na primavera, conforme apresentado na Figura 6.35. Os resultados estão apresentados nas figuras 6.36 e 6.37:

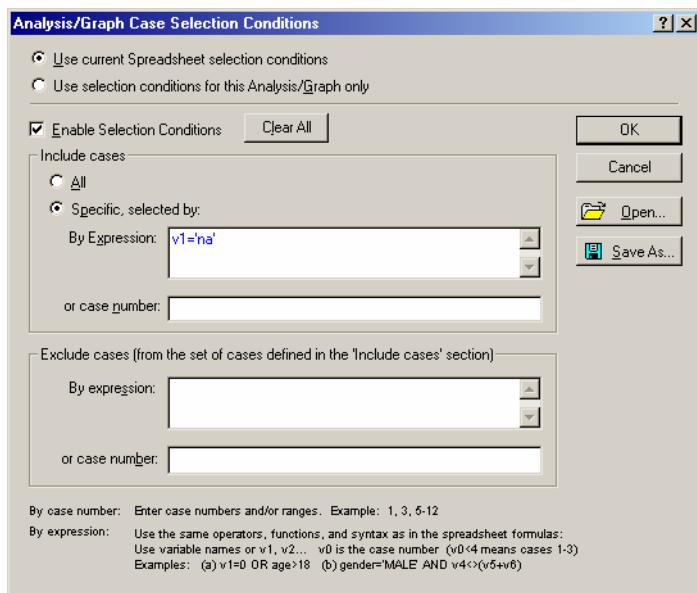


Figura 6.35 – Janela para especificação da condição selecionada

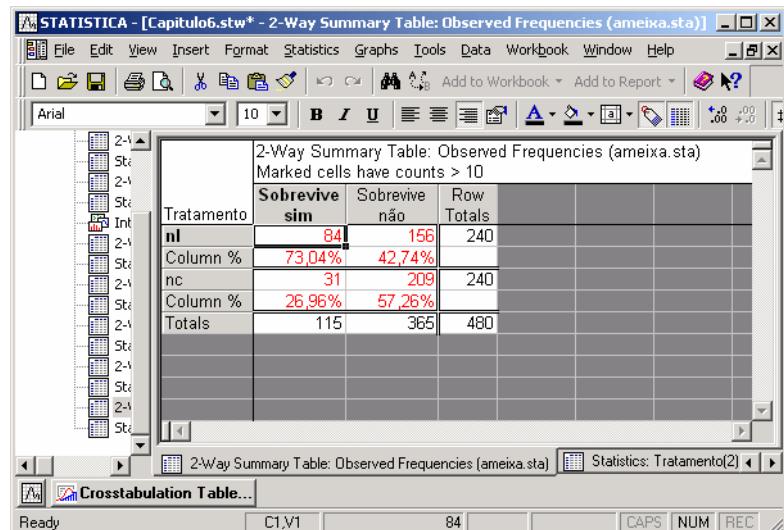


Figura 6.36 – Tabela de contingência da sobrevivência dos enxertos

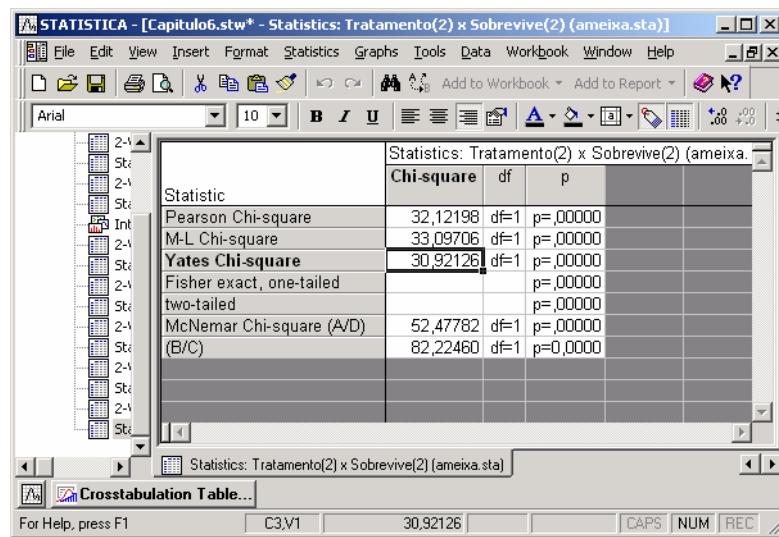


Figura 6.37 – Resultados do teste de qui-quadrado para o efeito da interação da forma dentro da época na primavera

Interpretação: considerando a estatística qui-quadrado com correção de Yates, cujo valor foi de 30,9213 (p<0,0001), a melhor forma de plantio, dentro da época na primavera, é o longo (17,50% de sobreviventes), conforme pode ser visto na Figura 6.36.

(7) Repetir os quatro primeiros itens, sendo que o filtro é construído para estudar a época fora da primavera, conforme apresentado na Figura 6.38. Os resultados estão apresentados nas figuras 6.36 e 6.37:

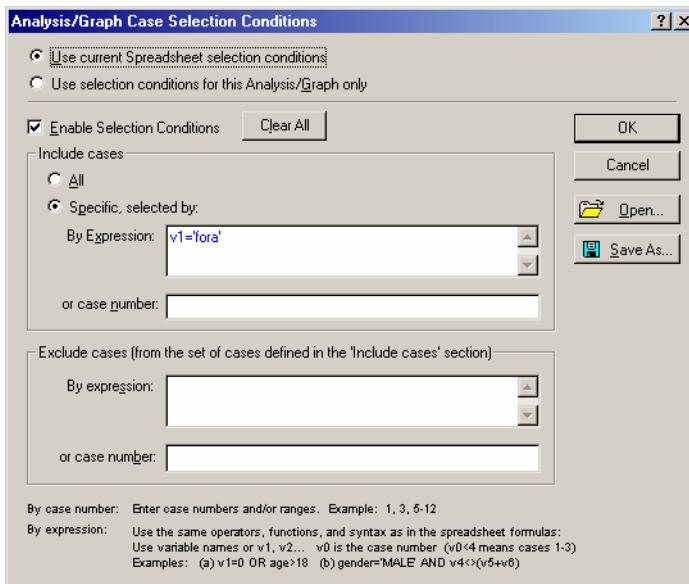


Figura 6.38 – Janela para especificação da condição selecionada

2-Way Summary Table: Observed Frequencies (ameixa.sta)
 Marked cells have counts > 10

Tratamento	Sobrevive sim		Row Totals
	Sobrevive não		
1	156	84	240
Column %	59,32%	38,71%	
fc	107	133	240
Column %	40,68%	61,29%	
Totals	263	217	480

Figura 6.39 – Tabela de contingência da sobrevivência dos enxertos

Statistics: Tratamento(2) x Sobrevive(2) (ameixa.sta)

Statistic	Chi-square	df	p
Pearson Chi-square	20,19379	df=1	p=.00001
M-L Chi-square	20,34336	df=1	p=.00001
Yates Chi-square	19,37797	df=1	p=.00001
Fisher exact, one-tailed			p=.00001
two-tailed			p=.00001
McNemar Chi-square (A/D)	1,674740	df=1	p=.19563
(B/C)	2,534031	df=1	p=.11142

Figura 6.40 – Resultados do teste de qui-quadrado para o efeito da interação da forma dentro da época fora da primavera

Interpretação: considerando a estatística qui-quadrado com correção de Yates, cujo valor foi de 19,3780 ($p < 0,0001$), a melhor forma de plantio, dentro da época fora da primavera, é o longo (32,50% de sobreviventes), conforme pode ser visto na Figura 6.36.

Observação: note que independente da época, a melhor forma de corte é o longo, confirmando a não existência do efeito interação.

7. ANÁLISE DE VARIÂNCIA

Deseja-se verificar se existem diferenças entre as respostas médias de tratamentos; o procedimento utilizado para inferir se tais diferenças realmente existem é chamado de Análise de Variância (ANOVA).

Idéia - Decomposição da Variação (Variâncias) em:

- Variação atribuída às diferenças entre as unidades experimentais (QMR - quadrado médio do resíduo)
- Variação atribuída às diferenças entre as unidades experimentais e atribuída às diferenças causadas pelos tratamentos (diferenças entre as médias) (QMTrat - quadrado médio dos tratamentos)

O teste baseia-se em duas hipóteses:

- H_0 : as médias dos tratamentos são todas iguais
- H_a : as médias dos tratamentos não são todas iguais

A primeira hipótese H_0 , também chamada de hipótese de nulidade, admite que não existem diferenças significativas entre as médias. Já a segunda hipótese, H_a , também chamada de hipótese alternativa, considera que existe diferença significativa entre as médias e, portanto, elas não são todas iguais.

O Teste F:

$$F = \frac{QMTrat}{QMR} \quad (7.1)$$

- H_0 não é rejeitada, o valor de F é pequeno
- H_0 é rejeitada, o valor de F é grande

Portanto, podemos fazer inferência sobre as médias de tratamentos (se elas diferem), usando esta razão de estimadores de variância. Vamos usar esta razão como Teste Estatístico para verificar H_0 versus H_a .

Resultado Estatístico:

Se H_0 não é rejeitada, a estatística F tem distribuição $F_{(t-1), t(r-1)}$ como sua distribuição amostral.

$$F \sim F_{(t-1), t(r-1)} \quad (7.2)$$

onde:

t: número de tratamentos

r: número de repetições

Vamos usar este resultado para testar hipóteses sobre diferenças entre médias de tratamentos. Realizaremos a Análise de Variância para podermos concluir, com um certo nível de confiança, que os nossos tratamentos (resultados) são ou não são significativamente diferentes entre si.

7.1. Delineamento Inteiramente Casualizado

Modelo:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (7.3)$$

onde:

$i = 1, 2, \dots, t$

$j = 1, 2, \dots, r$

μ : média geral sem tratamento

τ_i : mudança na média (com relação à μ) associada com o tratamento i

ε_{ij} : erro associado ao tratamento i da j -ésima repetição

t : número de tratamentos

r_i : número de repetições no tratamento i

Média dos Tratamentos:

$$\mu_i = \mu + \tau_i \quad (7.4)$$

onde:

μ_i : média da população que descreve a resposta das unidades experimentais que receberam o tratamento i

7.2. Exemplo

O exemplo que usaremos foi extraído de *Sokal and Rohlf, 1981, Biometry*, p.219-221, e, trata de um estudo de crescimento em cultura de tecido com a presença de auxinas (arquivo *oneway.sta*).

- **Objetivo:** testar o efeito da adição de diversos açúcares no crescimento.
- **Variável Resposta:** comprimento de seções de sementes (x 0,1114 mm).
- **Tratamentos:** Plantas foram aleatoriamente designadas aos cinco tratamentos:
 1. controle (sem adição de açúcar)

2. 2% glicose
3. 2% de frutose
4. 1% de glicose + 1% de frutose
5. 2% de sacarose

• **Unidades Experimentais:** plantas individuais para os quais os tratamentos foram aplicados.

• **Unidade de Amostragem:** plantas.

• **Repetições:** foram feitas dez repetições.

Na tabela abaixo estão os dados coletados no experimento.

Tabela 7.1 – Dados coletados no experimento

Tratamento	Tamanho											
1	75	67	70	75	65	71	67	67	76	68		
2	57	58	60	59	62	60	60	57	59	61		
3	58	61	56	58	57	56	61	60	57	58		
4	58	59	58	61	57	56	58	57	57	59		
5	62	66	65	63	64	62	65	65	62	67		

Para realizar a análise, os dados devem ser construídos sempre da forma matricial, conforme apresentado no quadro abaixo.

Quadro 7.1 – Matriz de dados

Açúcar	Repetição	Tamanho
1	1	75
1	2	67
1	3	70
1	4	75
1	5	65
1	6	71
1	7	67
1	8	67
1	9	76
1	10	68
2	1	57
2	2	58
2	3	60
2	4	59
2	5	62
2	6	60
2	7	60
2	8	57
2	9	59
2	10	61
3	1	58
3	2	61
3	3	56

3	4	58
3	5	57
3	6	56
3	7	61
3	8	60
3	9	57
3	10	58
4	1	58
4	2	59
4	3	58
4	4	61
4	5	57
4	6	56
4	7	58
4	8	57
4	9	57
4	10	59
5	1	62
5	2	66
5	3	65
5	4	63
5	5	64
5	6	62
5	7	65
5	8	65
5	9	62
5	10	67

7.3. Análise Estatística

Antes de iniciarmos a ANOVA, iremos construir o diagrama de dispersão e o diagrama de caixas. Ressalta-se que todos os gráficos apresentados neste capítulo foram alterados e padronizados conforme as opções gráficas exploradas no Capítulo 4.

• Diagrama de Dispersão

Para construir o gráfico do diagrama de dispersão, primeiramente abra o arquivo *oneway.sta* e siga os passos definidos abaixo:

(1) No menu [Graphs], escolha [2D Graphs] [Scatterplots]

(2) Selecione as variáveis: X = ACUCAR

 Y = TAMANHO

(3) Defina as opções: *Graph type: Regular*

Linear Fit: OFF

Regression bands: OFF

(4) Clique no botão [OK]

(5) O diagrama de dispersão está apresentado na figura 7.1:

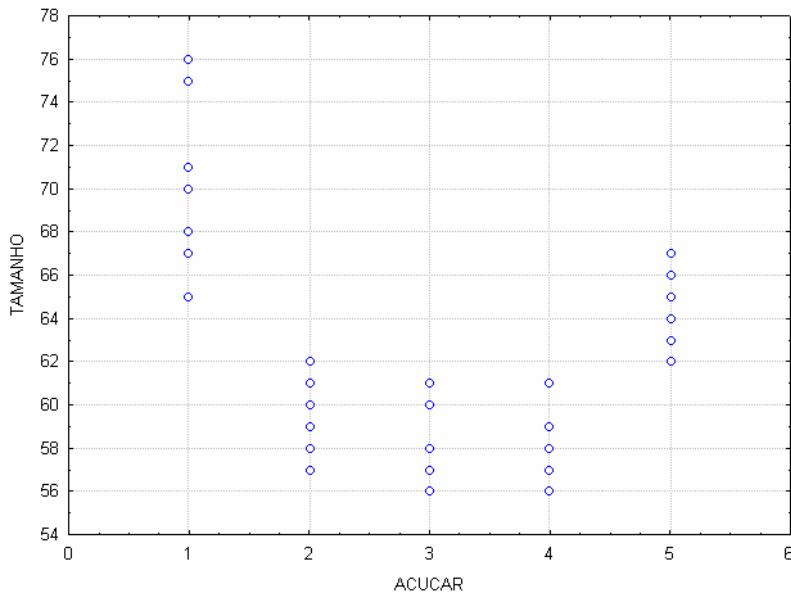


Figura 7.1 – Diagrama de dispersão do crescimento nos cinco tratamentos

• Diagrama de Caixas

Para construir o gráfico do diagrama de caixas siga os passos definidos abaixo:

- (1) No menu [Graphs], escolha [2D Graphs] [Box Plots]
- (2) Selecione as variáveis: *Dependent Variable* = ACUCAR
Grouping Variable = TAMANHO
- (3) Defina as opções: *Graph type*: Box-Whiskers, Regular
Grouping intervals: Integer mode
- (4) Clique no botão [OK]
- (5) O diagrama de caixas está apresentado na Figura 7.2:

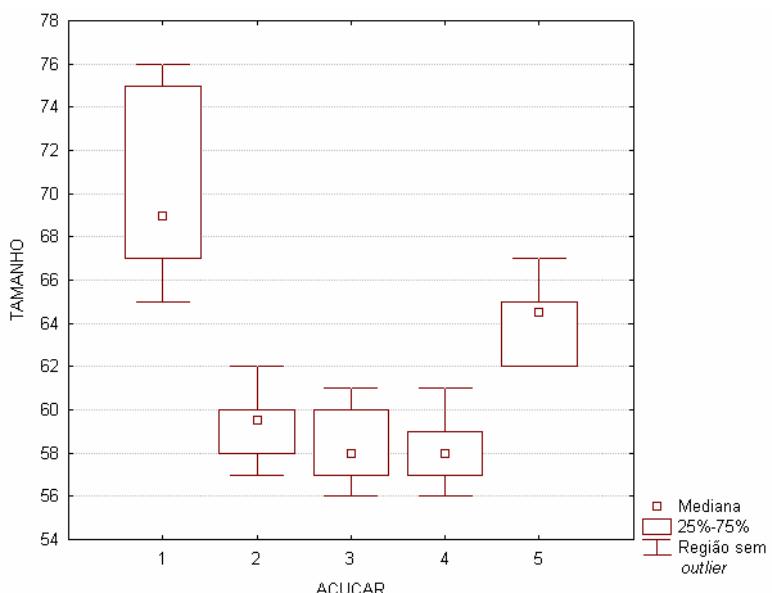


Figura 7.2 – Diagrama de caixas do crescimento nos cinco tratamentos

Observação: note que independente da época, a melhor forma de corte é o longo, confirmando a não existência do efeito interação.

7.3.1. Análise de Variância

Para realizar a análise de variância, siga os passos abaixo:

(1) Abra o módulo ANOVA, acesse o menu [STATISTICS], opção [ANOVA] e aparecerá a janela apresentada na Figura 7.3:

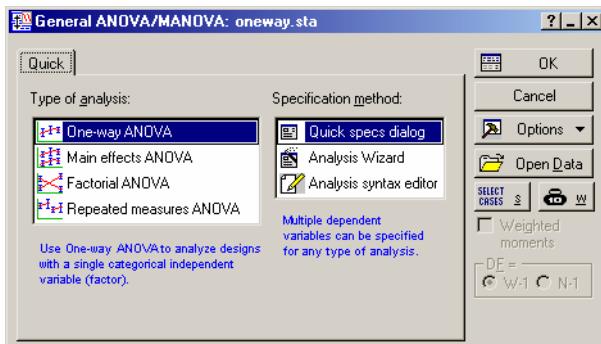


Figura 7.3 – Tela inicial da ANOVA

(2) Escolha: *Type of analysis: One-way ANOVA*

Specification method: Quick specs dialog

(3) Aparecerá a janela apresentada na Figura 7.4:

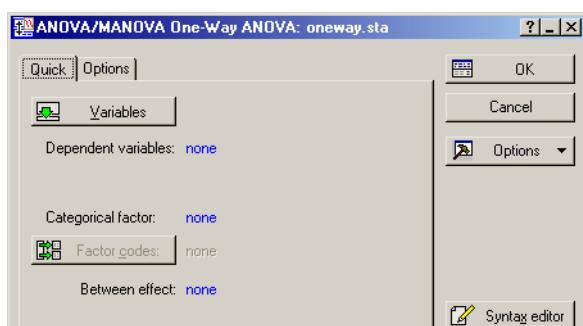


Figura 7.4 – Escolha das opções da ANOVA

(4) Clique no botão [Variables] e escolha as seguintes opções:

Dependent Variables: TAMANHO

Categorical factor: ACUCAR

(5) O resultado das escolhas está apresentado na Figura 7.5:

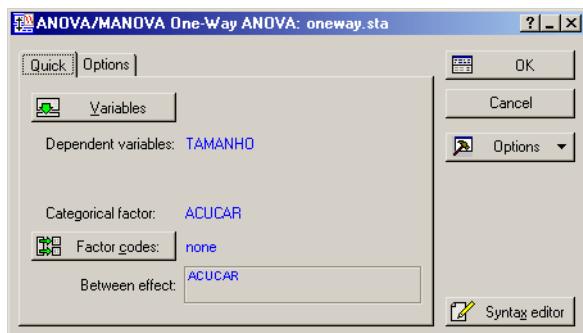


Figura 7.5 – Resultado das escolhas

- (6)** Clique na aba [Options], conforme a Figura 7.6, e escolha a opção:
Sums of squares: Type I (sequential)

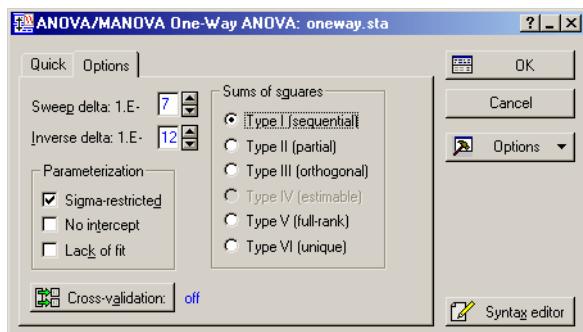


Figura 7.6 – Escolha do tipo de soma de quadrados

- (7)** Clique no botão [OK] e aparecerá a janela apresentada na Figura 7.7:

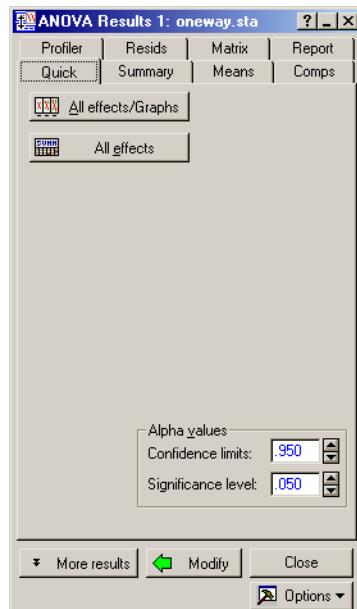


Figura 7.7 – Resultados da ANOVA

- (8)** Clique em [All effects] e o resultado está apresentado na Figura 7.8:

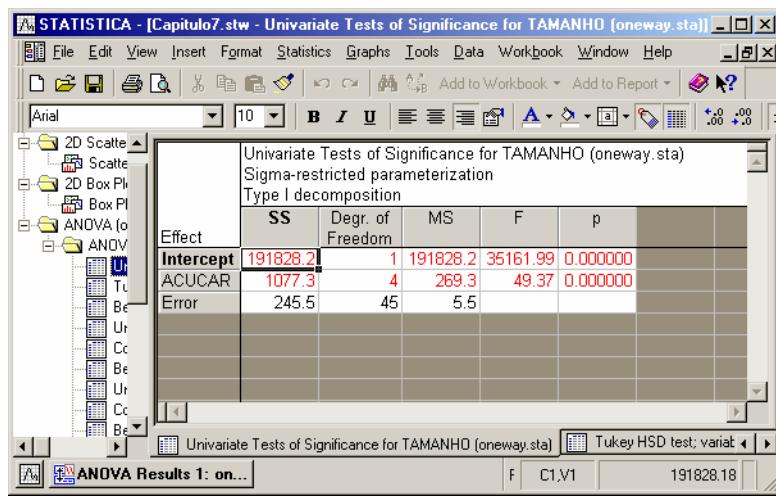


Figura 7.8 – Resultados da Análise de Variância do experimento

Conclusão: Existe diferença significativa entre as médias dos tratamentos ao nível de significância de $p < 0,0001$.

7.3.2. Comparações Múltiplas de Médias

No item 7.1.1., testamos as médias dos cinco tratamentos e concluímos que existe diferença entre as mesmas. Porém, não sabemos ainda se a média do tratamento 1 é igual ou diferente de 2, de 3, de 4 ou de 5, ou seja, existe diferença, mas não podemos dizer entre quais médias. O seguinte teste é necessário quando existe diferença, pois completa a análise, comparando e mostrando então, o resultado de todas as combinações de pares de médias.

Para realizar as comparações múltiplas de médias, siga os passos abaixo:

(1) Na janela da Figura 7.8, clique no botão [More Results], aparecerá a janela da Figura 7.9:

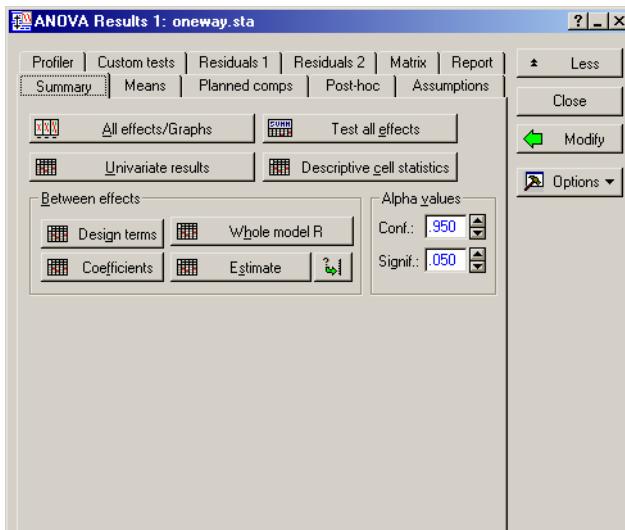


Figura 7.9 – Todos os resultados da ANOVA

(2) Clique na aba [Post-hoc], conforme apresentado na Figura 7.10:

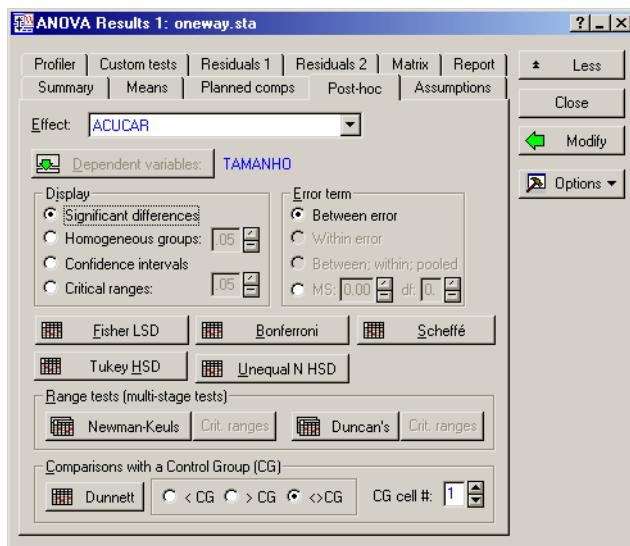
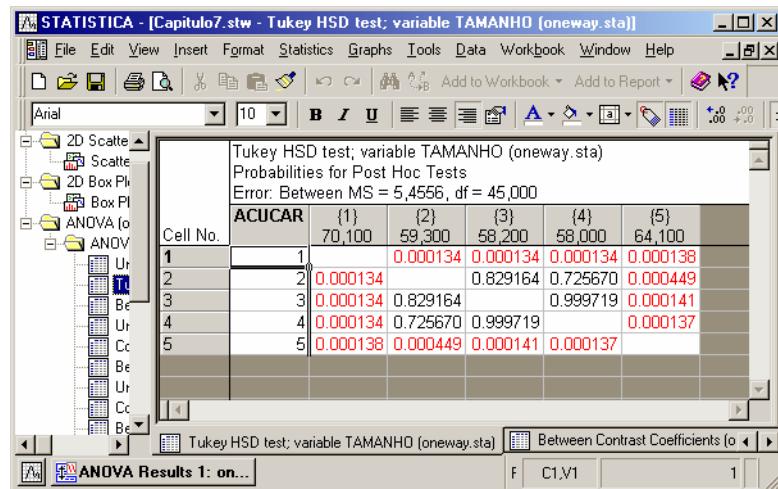


Figura 7.10 – Testes para comparações múltiplas de médias

(3) Clique no botão [Tukey HSD] e o resultado aparecerá conforme apresentado na Figura 7.11:

Figura 7.11 – Valores *p* para todos os pares de médias

(4) Na Figura 7.10, na região *Display*, selecione a opção *Homogeneous groups*, clique no botão [Tukey HSD] e o resultado aparecerá conforme apresentado na Figura 7.12:

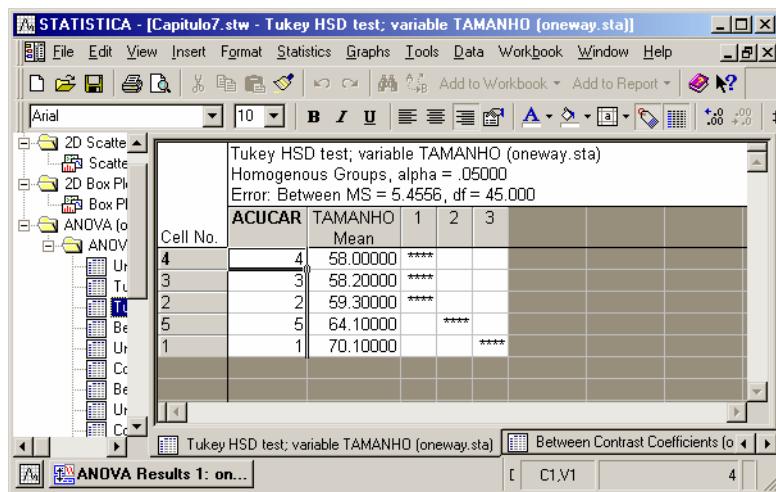


Figura 7.12 – Grupos homogêneos do teste de Tukey

Conclusão: A partir da Figura 7.12, construi-se a Tabela 7.2:

Tabela 7.2 – Resultados do teste de Tukey			
Grupo Tukey	Tratamentos	Médias	
a	1	70,1	
b	5	64,1	
c	2	59,3	
c	3	58,2	
c	4	58,0	

Observe na Tabela 7.1, que o tratamento 1 (a) é superior e diferente significativamente dos demais. Os tratamentos 2, 3 e 4 (c), são considerados iguais entre si, ou seja, as médias não diferem significativamente entre elas. Estes tratamentos são os de médias inferiores. O tratamento 5 (b) difere dos demais e é o segundo melhor tratamento, ou seja, a segunda melhor média, inferior ao 1 e superior ao 2, 3 e 4.

7.3.3. Contrastos Ortogonais

Consiste de uma combinação linear onde a soma dos coeficientes é nula.

Os contrastes de interesse nesse estudo são para verificar se:

- (A) o tratamento controle 1 difere significativamente dos demais tratamentos;
- (B) verificar se o tratamento 5 difere dos tratamentos 2, 3 e 4; e
- (C) verificar se o tratamento 5 difere do tratamento 1.

Os respectivos coeficientes para verificar os contrastes estão apresentados na Tabela 7.3.

Tabela 7.3 – Coeficientes dos contrastes

Contrastes	Tratamentos				
	1	2	3	4	5
A	4	-1	-1	-1	-1
B	0	1	1	1	-3
C	1	0	0	0	-1

Para realizar os contrastes de interesse, siga os passos abaixo:

(1) Clique na aba [Planned comps], conforme apresentado na Figura 7.13:

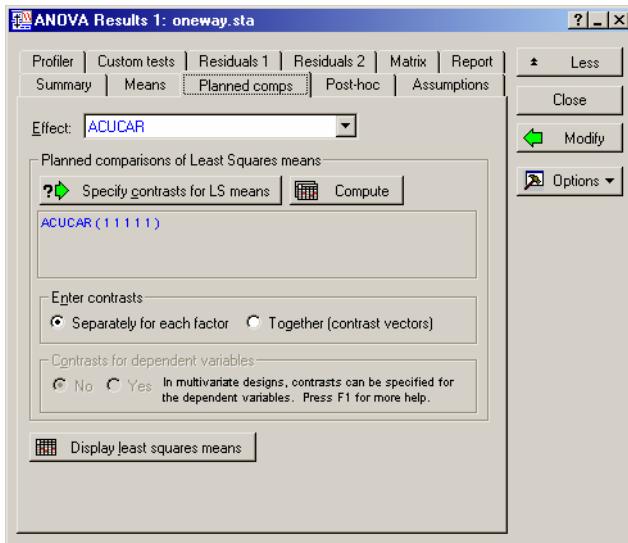


Figura 7.13 – Testes de Contrastos

(2) Clique no botão [Specify contrasts for LS means], e defina os coeficientes do contraste conforme apresentado na Figura 7.14:

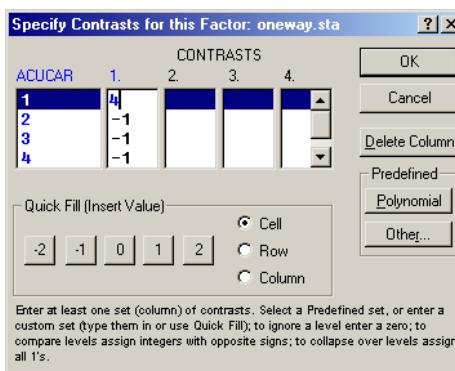


Figura 7.14 – Definição dos coeficientes dos contrastes

(3) Digite os coeficientes do contraste como mostra a Figura 7.14, clicando duas vezes com o botão esquerdo do mouse nos espaços da coluna 1.

(4) Clique no botão [OK]

(5) Note que o vetor ACUCAR mudou para $(4 -1 -1 -1 -1)$, conforme apresentado na Figura 7.15:

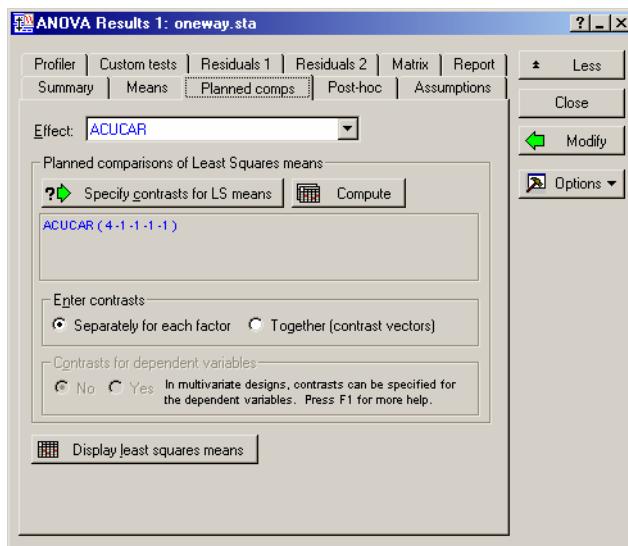


Figura 7.15 – Definição dos coeficientes do contraste A

(6) Clique no botão [Compute] da Figura 7.15

(7) O resultado do teste do contraste A está apresentado na Figura 7.16:

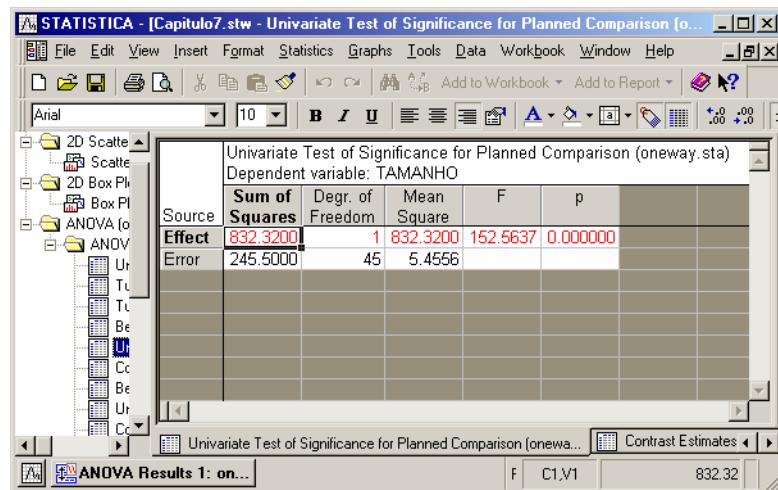


Figura 7.16 – Teste F do contraste A

Repetir os passos (2) a (6) para os outros dois contrastes, sendo que os resultados estão apresentados nas Figuras 7.17 e 7.18:

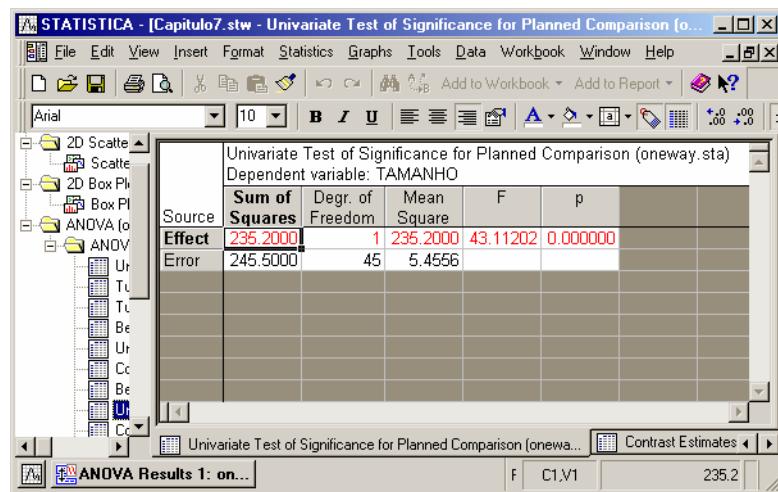


Figura 7.17 – Teste F do contraste B

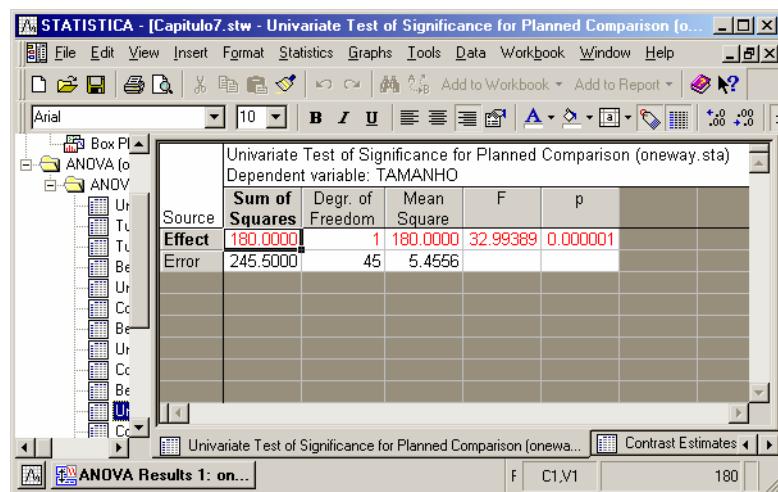


Figura 7.18 – Teste F do contraste C

7.3.4. Diagnóstico do Modelo

7.3.4.1. Ajuste do Modelo

A não significância do modelo geral está relacionada com a violação de algum pressuposto. É importante verificar as suposições de normalidade, independência e homogeneidade de variâncias (homocedasticidade). O ajuste do modelo é verificado pelo teste F da Análise de Variância, obtido, selecionando-se a opção [All effects].

No Delineamento Inteiramente Casualizado, o teste F coincide com o efeito principal.

7.3.4.2. Normalidade dos Resíduos

(1) Clique na aba [Assumptions], conforme apresentado na figura 7.19;

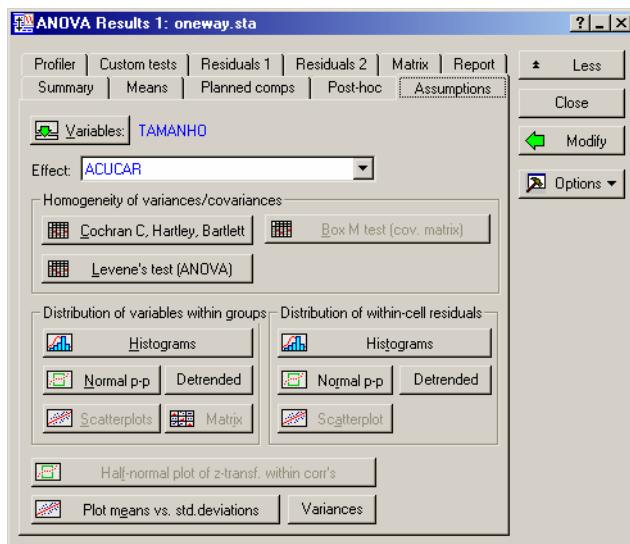


Figura 7.19 – Normalidade dos resíduos

- (2)** Na opção *Distribution of within-cell residuals*, clique no botão [Normal p-p]
(3) O gráfico normal de probabilidades dos resíduos está apresentado na Figura 7.20:

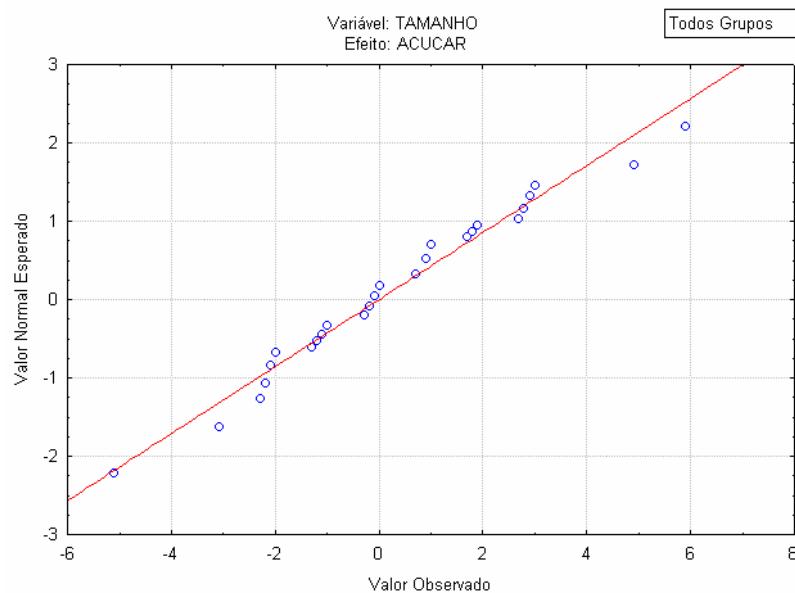


Figura 7.20 – Gráfico Normal de Probabilidades dos Resíduos

Conclusão: Os valores aproximam-se da reta normal e, portanto, a suposição de normalidade dos dados é atendida.

7.3.4.3. Independência dos Resíduos

Correlações positivas entre os erros de diferentes unidades experimentais de um mesmo tratamento podem surgir se o pesquisador não tiver cuidado ao conduzir um experimento. A independência pode ser assegurada por um dos princípios básicos da experimentação, que é a casualização.

De acordo com as condições do experimento em estudo, assume-se que os resultados são independentes.

7.3.4.4. Homogeneidade das Variâncias por Tratamento

(1) Conforme Figura 7.19, na opção *Homogeneity of variances/covariances*, clique no botão [Cochran C, Hartley, Bartellet]

(2) O resultado do teste está apresentado na Figura 7.21:

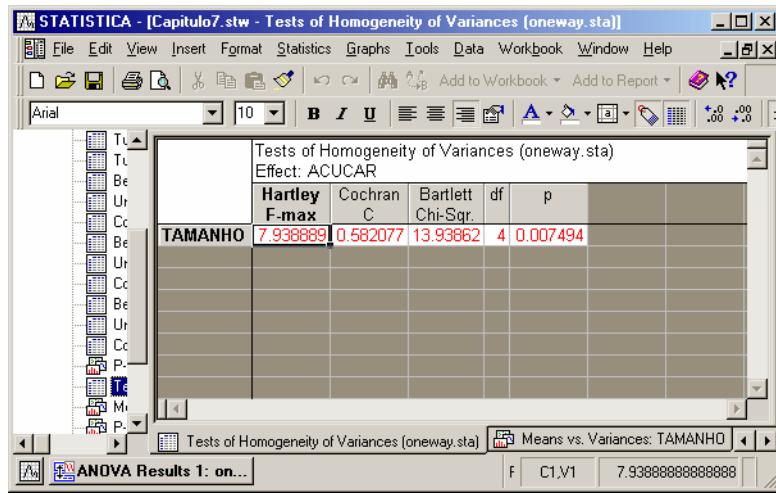


Figura 7.21 – Resultado do teste de homogeneidade de variâncias

Interpretação do teste de Cochran:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_a: \sigma_i^2 \neq \sigma_j^2 \text{ para } i \neq j$$

Rejeitamos a hipótese nula ao nível de significância de 0,0075 ou 0,75%. Portanto, a suposição de homogeneidade de variâncias não é válida.

Uma outra forma, através da qual podemos "visualizar" a heterogeneidade de variâncias, é fazer um gráfico das médias versus as variâncias, conforme procedimento abaixo:

(3) Conforme Figura 7.19, clique no botão [Variances]

(4) O gráfico resultante está apresentado na Figura 7.22:

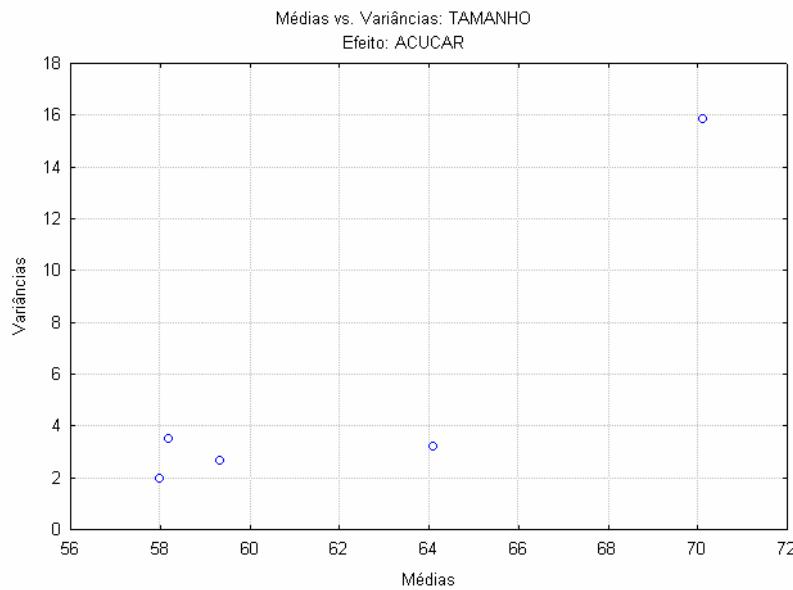


Figura 7.22 – Gráfico das médias versus as variâncias dos cinco tratamentos

Conclusão: O tratamento controle 1 apresenta média elevada e variância elevada, ou seja, apresentam-se relacionadas. Isto implica em heterocedasticidade (heterogeneidade de variância).

Observação: Se os dados não se afastam muito da reta (gráfico normal de probabilidades) podemos usar qualquer teste de homogeneidade de variância. Se existir falta de normalidade devemos usar o teste de *Levene*.

(5) Conforme Figura 7.19, na opção *Distribution of variables within groups*, clique no botão [Normal p-p]

(6) Escolha os grupos conforme apresentado na Figura 7.23:

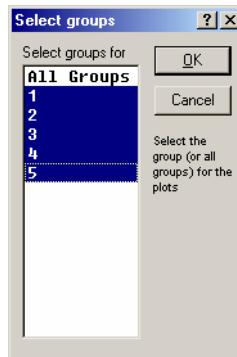


Figura 7.23 – Escolha dos grupos

(7) Os gráficos normais de probabilidades dos resíduos para cada tratamento estão apresentados na Figura 7.24:

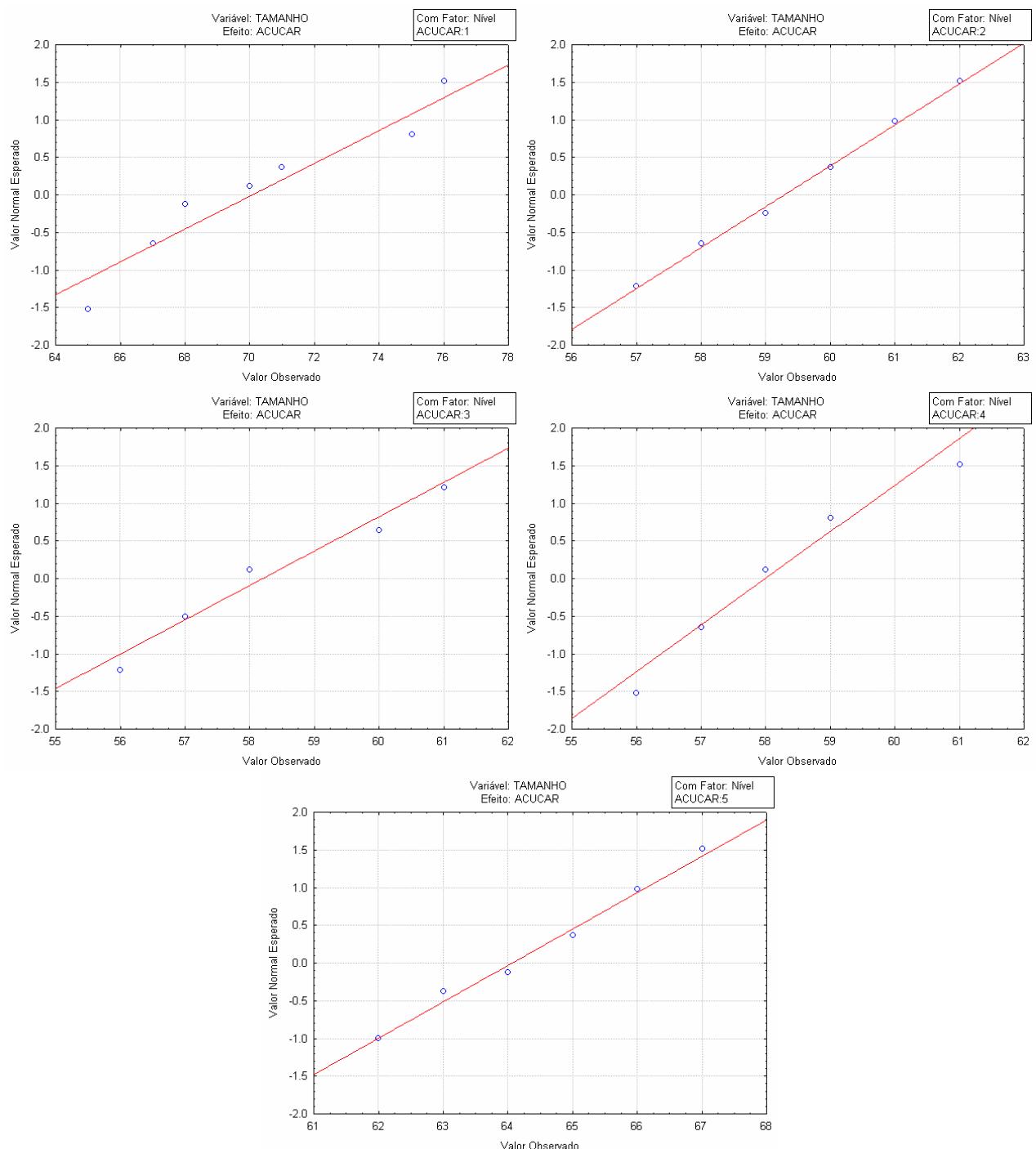


Figura 7.24 – Gráficos normais de probabilidades dos resíduos para cada tratamento

Conclusão: as cinco amostras são oriundas de distribuição aproximadamente normal.